UNIVERSITY OF LJUBLJANA
FACULTY OF ECONOMICS

MASTER'S THESIS

AN ANALYSIS OF THE OIL SHOCKS USING THREE - PASS REGRESSION FILTER FACTOR ESTIMATION WITHIN THE FAVAR

Ljubljana, September 2018

MIHA BREZNIKAR

AUTHORSHIP STATEMENT

The undersigned Miha Breznikar, a student at the University of Ljubljana, Faculty of Economics, (hereafter: FELU), author of this written final work of studies with the title An Analysis of the oil shocks using three - pass regression filter factor estimation within the FAVAR, prepared under supervision of Professor, Igor Masten PhD.

DECLARE

1. this written final work of studies to be based on the results of my own research;
2. the printed form of this written final work of studies to be identical to its electronic form;
3. the text of this written final work of studies to be language-edited and technically in adherence with the FELU's Technical Guidelines for Written Works, which means that I cited and / or quoted works and opinions of other authors in this written final work of studies in accordance with the FELU's Technical Guidelines for Written Works;
4. to be aware of the fact that plagiarism (in written or graphical form) is a criminal offence and can be prosecuted in accordance with the Criminal Code of the Republic of Slovenia;
5. to be aware of the consequences a proven plagiarism charge based on the this written final work could have for my status at the FELU in accordance with the relevant FELU Rules;
6. to have obtained all the necessary permits to use the data and works of other authors which are (in written or graphical form) referred to in this written final work of studies and to have clearly marked them;
7. to have acted in accordance with ethical principles during the preparation of this written final work of studies and to have, where necessary, obtained permission of the Ethics Committee;
8. my consent to use the electronic form of this written final work of studies for the detection of content similarity with other written works, using similarity detection software that is connected with the FELU Study Information System;
9. to transfer to the University of Ljubljana free of charge, non-exclusively, geographically and time-wise unlimited the right of saving this written final work of studies in the electronic form, the right of its reproduction, as well as the right of making this written final work of studies available to the public on the World Wide Web via the Repository of the University of Ljubljana;
10. my consent to publication of my personal data that are included in this written final work of studies and in this declaration, when this written final work of studies is published.

Ljubljana, September 17th 2018 Author's signature: ___________________

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

ang. – English

3PRF – Three-Rass Regression Filter

CFTC – U.S. Commodity Futures Trading Commission

CPI – Consumer Price Index

EIA – Energy Information Administration

FAVAR – Factor Augmented Vector Autoregression

$FAVAR_{3PRF}$ – Factor Augmented Vector Autoregression with Three-Pass Regression Filter Factors

$FAVAR_{PCA}$ – Factor Augmented Vector Autoregression with Principal Components Analysis Factors

FED – Federal Reserve Bank

FEVD – Forecast Error Variance Decomposition

FRED-MD – Monthly Database for Macroeconomic Research

GDP – Gross Domestic Product

ISM – Institute of Supply Management

MSPE – Mean Squared Prediction Error

OECD – Organisation for Economic Co-operation and Development

OLS – Ordinary Least Squares

OPEC – Organization of the Petroleum Exporting Countries

PCA – Principal Component Analysis

PCR – Principal Component Regression

PLS – Partial Least Squares

SFAVAR – Structural Factor Augmented Vector Autoregression

USD – U.S. Dollars

VAR – Vector Autoregression

WTI – West Texas Intermediate

INTRODUCTION

In the last decades changes in oil prices have been regarded as an indispensable source of macroeconomic fluctuations. Hamilton (1983) showed that a spike in oil prices preceded all U.S. recessions, except one since World War II. Many researchers found out that oil price shocks have a substantial and negative impact on the economy. Even though many alternative sources of energy have been developed in recent years, oil and its derivatives are still the essential energy source for industry and transportation. However, development of alternative energy sources, better counter-inflation policy, shorter duration of shocks, lower oil shocks intensity and changes in the source of oil price shocks contribute to the fact that the effects of oil shocks are today diminished.

In evaluating the impacts of oil prices on economy, the standard approach was to treat all oil price shocks as exogenous innovations in oil supply. However, recent researches have shed some light on the fact that sources of oil price fluctuations matter. Kilian (2009) initially identified three different shocks to the global crude oil market. These three oil shocks are divided into crude oil supply shock, a shock to the global demand for all industrial commodities and an oil specific demand shock. Results of his model suggest that demand factors have historically driven oil prices and that effects of shocks depend crucially on the origin of the shock. In the last years, speculation in oil market was also found as a relevant source of oil price fluctuation as shown in Juvenal and Petrella (2015), Kilian and Lee (2014), Kilian and Murphy (2014) and others. Speculative shocks were in most cases addressed only through oil inventories, the physical part of the oil market. However, I include an additional variable for speculative shock from financial market to control for the effect of increasing share of speculators in oil futures market, as also proposed by Medlock III (2013).

Additionally, I consider the effects of monetary policy within the oil market model following Aastveit (2014) to check the responses of monetary policy to oil shocks. As proposed by Frankel (2008), U.S. monetary policy may also affect commodity prices and therefore also other oil market variables.

Effects of various shocks in oil markets on the economy are commonly estimated with vector autoregressive (hereinafter: VAR) models proposed by Sims (1980). The main problem of VAR is the lack of informativeness, because of the limited number of variables included in the model. Bernanke, Boivin and Eliasz (2005) proposed factor-augmented VAR (hereinafter: FAVAR) as a suitable alternative to eliminate deficiency of standard VAR model. Even though FAVAR has already been proven as a suitable way for adding information to standard VAR only a few studies regarding oil shocks were presented. Aastveit (2014) and Juvenal and Petrella (2015) presented models where various oil shocks are combined by factors obtained by principal component analysis (hereinafter: PCA).

However, recently Kelly and Pruitt (2015) proposed a three-pass regression filter (hereinafter: 3PRF) which is a new method for relevant factor estimation. The main difference between PCA and 3PRF is that 3PRF calculates relevant factors from a set of variables. In this thesis, I combine 3PRF and FAVAR to estimate the effects of various oil shocks on U.S. economy. The potential improvement upon the traditional FAVAR model is the use of relevant information for the variable of interest.

The purpose of this thesis is the estimation of the oil market-related FAVAR model with 3PRF relevant factors. This approach enables me to use only relevant part of information in the structural analysis. The traditional FAVAR model with PCA contains information irrelevant to the variable of interest, and this deficiency may lead to contamination of structural shocks. 3PRF factors are constructed in a way to capture only the information relevant to the particular variable of interest and therefore bypasses the limitation of PCA factors in the structural analysis. Since the 3PRF factors are target oriented, the reestimation of the model for each variable of interest is required which means higher computational complexity due to changes in factors within the FAVAR. Because the inclusion of 3PRF factors directly in the FAVAR model is a new approach, I compare the results with two well-established procedures, namely baseline VAR model and the FAVAR model with PCA factors.

I hypothesize that the estimated FAVAR models with factors derived from 3PRF or PCA outperform basic VAR model concerning the sensibility of impulse responses, due to the inclusion of additional information. I am also interested in potential differences between estimated impulse responses of models and reasons for them. I also hypothesize that there are considerable differences in the response of the economy to the different oil price shocks and that oil shock source also determines monetary policy response. A further question to answer is related to reverse causality or the impact of monetary policy shock on the oil market.

Through the thesis, I discuss and test potential advantages of the proposed model, especially regarding informational sufficiency, the sensibility of obtained results and stronger relation to less prevalent sectors of the economy, which are not sufficiently represented by factors obtained from PCA. The efficient modeling of the response of less prevalent economic sectors is essential since one may be interested in how particular industries are affected by an oil shock.

1 HISTORY OF OIL SHOCKS

Oil is probably the most important commodity since it is crucial in transportation and also for gathering other sources of energy. A sudden increase in oil prices can therefore crucially increase the costs of companies and consequently prices, savings, and investments. Historically many recessions were somehow related to oil price hikes even though it was not so economically important as it is today. Strong relation to economy is also the main reason for the massive interest of economists. In this section, I will present the main oil shocks in the history and reasons for them to emphasize the importance of oil prices in the economic history.

1.1 Oil as an illuminant

An ascent of crude oil was initiated in the second half of the 19th century when it was used for illumination. Oil was produced by treatment of coal, asphalt, coal-tars or shale and because that kind of production of oil was reasonably expensive, prices of oil reached 80 U.S. dollars (hereinafter: USD) per barrel or approximately 1,900 USD expressed in 2009 dollars. However, in 1859 a new era was begun by Edwin Drake who successfully drilled for oil. As obtaining oil by drilling was much cheaper, oil production suddenly became a very profitable activity. Consequently, oil production quickly increased and the price of oil dropped to 0.1 USD by the end of 1861 (Hamilton, 2011, p. 240).

The first oil shock occurred between 1862 and 1864. The U.S. Civil War generally increased commodity prices and demand for commodities. The oil market was additionally affected by the limited supply of turpentine from the south and introduction of a very high tax on alcohol, which almost completely eliminated alcohol from the market of illuminants. Even as demand for oil clearly grew, oil production declined because of the initially low price. Consequently, the oil price increased by similar scale as during the 1970s (Hamilton, 2011, p. 241).

After the war, oil prices fell significantly due to the decreased demand and also increased production in new areas of Pennsylvania. Till 1890 oil production was five times the production in 1870 and oil prices again reached historically low levels. Because of increased production between 1870 and 1890, on Pennsylvanian oilfields, oil drilling became all the more demanding. Oil production fell, and even with new technologies it never reached the level seen in 1891 (Caplinger, 1997). Williamson and Daum (1959, p. 577) also suggested, that decline in oil production was the main reason for another oil price shock in 1895 even though supply shortages are not predominant reason for oil price spikes in 1895 (Hamilton, 2011, p. 243).

1.2 Oil in industry and transportation

Since crude oil was predominantly used for illumination, its price depended crucially on prices of alternative illuminants. However, in the 20th century electric lighting emerged and replaced crude oil in that respect. However, in the same period, crude oil gained importance in heating, as well as in transportation.

1.2.1 The West Coast Gasoline Famine

The West Coast Gasoline Famine in 1920 was the first oil shock in the 20th century. Use of oil derivatives for transport was on the rise at the time. Between 1918 and 1920 gasoline demand increased significantly, especially in the western market. The reason for increasing demand was in the increasing number of cars, tractors and other machines in farming. Because of scarcity of crude oil in California at the time, supply could not follow the increasing demand for gasoline. Even though, the supply and demand were not balanced, the effect on gasoline price was only negligible due to uniform pricing policy among retailers. However, limited gasoline supply disabled transportation and use of modern farming machinery (Olmstad & Rhode, 1985, p. 1046). Effects on economy were still negligible since oil and gasoline were not so widely used in production.

The fast growth of crude oil production quickly eliminated shortages of 1920 and a decade of declining crude oil prices began. Oil prices were additionally diminished by declining demand which was a consequence of great depression (Hamilton, 2011, p. 246).

1.2.2 Postwar supply shocks

From 1945 to 1948, demand for oil products increased drastically because of transition to the automotive era. Colossal demand hike led to 80 % higher price between 1945 and 1947, but it was insufficient to prevent crude oil shortages. Problems regarding crude oil supply and reserves in that period were mainly a consequence of **postwar dislocations**. In the period from 1950 to 1953, the price of oil was initially frozen during **Korean War** as the Office of Price Stabilization ordered it. Supply disruptions began in summer 1951 by removing 19 million barrels of monthly Iranian production from the world crude oil markets as a response to the nationalization of Iran's oil industry. Additionally, many U.S. refineries were closed down after a strike of refinery workers. Consequently, oil shortages emerged again and civilian use of crude oil needed to be cut again. After lifting of price controls in June 1953, oil prices rose by approximately 10 % and after a month the second postwar recession emerged (Hamilton, 2011, p. 248).

In 1956 the Suez Crisis emerged, following the nationalization of the Suez Canal by Egypt. Britain and France unsuccessfully tried to regain control of the canal. During the conflict, the canal was blocked, and oil transportation was diminished. Additionally, pumping stations for the Iraq Petroleum Company's pipelines were also sabotaged. The world crude oil production consequently dropped by 10.1 %. The impact of supply disruption was especially dramatic in Europe, which was more dependent on oil from the Middle East. The gap in oil supplies was quickly filled in by oil producers outside the Middle East, and also the Middle East production returned to the pre-crisis level by June 1957. Because of crises in Europe and other countries, U.S. export started to fall in 1957, and it was one of the factors contributing to the third postwar U.S. recession in August of 1957 (Hamilton, 2011, p. 249-251).

1.2.3 The age of OPEC

Allowed production levels in the U.S. were rapidly increased by the Texas Railroad Commission in the late sixties and then also conservation restrictions were omitted by 1972. U.S. oil production peaked in 1972, and it declined after that despite the substantial price increases which should have positively affected oil production. Crude oil shortages in the United States could be replaced by oil from the Middle East, but the transition from world petroleum market in the Gulf of Mexico to the one centered in the Persian Gulf was not an easy task (Hamilton, 2011, p. 252).

Even though oil prices skyrocketed in the early 1970s, many other factors affected oil price along with shortages in supply. The end of Bretton Woods system caused the dollar depreciation, and it increased the price of all commodities traded in U.S. dollars, nominal yield on 3-month Treasury Bills was below the realized CPI inflation, and that additionally increased commodity prices (Barsky & Kilian, 2001).

In October 1973 Arab members of Organization of the Petroleum Exporting Countries (hereinafter: OPEC) announced an embargo on oil exports to countries viewed as supporters of Israel and OPEC production dropped. Oil supply shortages implied huge price hike to almost 12 USD per barrel (Hamilton, 2011, p. 252). However, behind embargo also economic motivations were hidden. Arab producers discussed embargo before the war with Israel, and an embargo was lifted even before political objectives were achieved. As the embargo started during the financial crisis in the United States, it had an even more severe effect on the economy (Barsky & Kilian, 2004, p. 169). Baumeister and Kilian (2016) stated that oil producing countries were not even affected during the war, which took place mostly in Syria, Israel, and Egypt, also suggesting that the embargo had an economic background.

The Arab-Israeli War was only the beginning of a turbulent period in the Middle East and also in the oil market. Iran experienced massive public protests in 1978, which brought Iranian oil production down. Since Iran crucially increased production during the 1974 embargo, cuts in Iranian production additionally depressed global oil production. Until 1979, Iranian production returned to one half of level before the revolution, but production dropped in 1980 again, because of the war with Iraq. Oil price almost doubled during that period (Hamilton, 2011, p. 255-256).

The period between 1981 and 1986 is known as **the great price collapse**. World oil production would needed many years to achieve the pre-war level of production. However, demand responded to price hikes in previous years, and oil consumption dropped quite significantly. Even though oil production was additionally decreased, it was not enough to prevent initial price decline. Saudi Arabia abandoned its effort of price recovery and began to raise production again in 1986. Consequently, oil price collapsed to as low as 12 USD per barrel. Even though oil was considerably cheaper for consumers, this was a massive shock to oil producers, and therefore many oil-producing states in the United States experienced regional recessions in the 1980s (Hamilton, 2011, p. 256-257).

1.2.4 New Industrial Age

The considerable increase in oil consumption stigmatizes the last period in oil price shock history. It was a consequence of industrialization, change in living standards and development of newly industrialized economies. Chinese petroleum consumption, for example, had approximately 6 % annual growth rate. Even though the growth rate of newly industrialized economies was remarkable already two decades, these countries could influence oil market only after they achieve a certain stage of development (Hamilton, 2009, p. 229). That transition also meant that oil supply shocks, which were important oil price drivers in the first half of the century following World War II, were replaced by demand shocks, which became a more critical oil price determinant.

At the end of the 20th century, Asian countries achieved the fastest growth, but their contribution to world oil demand was initially modest. However, the belief of continuous growth of Asian tigers was probably a factor boosting oil prices. After the financial crisis in Asian countries, investors lost their belief in increasing oil demand, and consequently, also oil price returned to the lowest level since 1972 (Hamilton, 2011, p. 258). During this period, oil prices were already crucially driven by expectations and trading of futures contracts on crude oil was already well developed which induced the oil price to become more volatile.

Growing demand and stagnant supply characterized the last decade and a half. Global economic growth and consequent growth of demand were strongly persistent during 2004 and 2005. Demand pressures were also the main reason for oil price increase after excess oil inventories were dried out. Even after the price increase, oil production did not grow after 2005, and supply deficit was not a consequence of geopolitical events. As oil production did not follow the increasing demand, consumption had to decrease despite growing incomes. Since oil price elasticity has never been very high, prices needed to rise intensively to achieve market stability (Hamilton, 2011, p. 261). West Texas Intermediate (hereinafter: WTI) price consequently increased from January 2005 to June 2008 from 46.84 to 133.93 USD per barrel. Important contributing factors for the described price hike were also a speculative bubble in the price of oil through commodity futures markets and negative interest rates in August 2007. In 2008 oil prices achieved the highest level in the history, and in December 2007 the latest recession began. Even though the latest recession was not caused solely by oil price shock, but primarily by bubbles in financial markets, oil price shock was once again correlated with the recession.

2 EMPIRICAL LITERATURE ON MACROECONOMIC EFFECTS OF OIL SHOCKS

Due to the strong correlation between oil price shocks and economic recessions, oil price shocks gain much interest in economic literature. The empirical literature on oil supply shocks origins in the 1970s by increasing interest on effects of oil price on the economy, which was a consequence of many oil supply shocks after 1973. In this section, I present the evolution of empirical literature on oil price shocks.

2.1 Early studies

Initial research leaned towards oil supply shocks which are mainly the consequence of conflicts in the Middle East, and are therefore exogenous. Oil shocks were seen as a permanent price shock, and the economy was meant to adjust to new environment. Another popular question was about the size of the effect of an oil shock on the economy, and to what extent oil prices, government policies and other effects are responsible for recession (Jones & Leiby, 1996, p. 4).

The first econometric study of effects of oil shocks on the economy was written by Darby (1982). In his study, the causes of 1973-1975 recession were carefully observed. His hypothesis was, that along to the oil supply shock after the OPEC oil embargo, the removal of Nixon's price control regime, the breakdown of Bretton Woods exchange regime, and a restrictive monetary policy could as well be reasons for the recession. Using linear regression, he concluded that oil supply shock is the most probable reason for U.S. recession since other coefficients were not statistically significant (Labonte, 2004, p. 4).

Hamilton (1983) published a study where he stressed that all but one postwar recession had been preceded by oil price shocks. He has also shown statistically that oil price shocks solely, Granger causes the recessions. He also proved that oil prices were not Granger caused by any other economic variable. His results suggested that after a 10 % oil price shock, gross national product (GNP) decreases by 0.04 percentage points in the first quarter, then 0.07 percentage points in the next quarter, 0.5 percentage points after three quarters, and 0.6 percentage points after a year. Since oil price increased by 20 % in some of the quarters and shocks occurred several quarters in a row, the estimated effect of oil prices on the economy was quite substantial. Hamilton was also the first who noted that oil shocks affect the economy with lags.

2.2 A faded link between oil prices and economy

Over the 1980s and early 1990s, the link between oil prices and macroeconomy was broken down (Labonte, 2004, p. 5). Mork (1989) found out that after the extension of data through the year 1988 and controlling for other macroeconomic factors, the causal relationship between oil price shocks and economy disappeared since the effect of oil price change on

GNP growth was low and statistically insignificant. The reasons were probably price decreases in the 1980s which did not have a significant positive effect on economic growth. Similar conclusion regarding the effect of oil prices on the economy was made by Lee, Ni and Ratti (1995) who also showed that causal linear relationship was broken down. They argued that it is more important how the pattern of oil price changes. In an environment where oil price is always volatile, oil shock effects are diminished. They, therefore, stressed the importance of accounting for oil price volatility and this view was also supported by Ferderer (1996), who statistically proved that oil price volatility affects economic growth. He concluded that oil price changes explain around 12.1 % of industrial production variability, whereas volatility explains additional 13.9 % of the variation. These effects are substantial since his estimated explained variation by monetary policy was only 11.8 %.

Hooker (1996) found out that oil price causality on numerous U.S. macroeconomic variables is no longer present after 1973. He tested three possible reasons, namely sample stability issues, endogeneity of oil prices and misrepresentation of the form of oil price interaction. The data supported none of the tested hypotheses. Hamilton (1996) immediately responded to Hookers ”question” by the explanation that after 1986 all price increases were a consequence of even more substantial oil price decreases. He also represented a net oil price increase variable to control for oil price decrease effect, but he failed to prove that this variable Granger causes economic growth.

Hamilton (2003) stressed that linear relation between oil prices and the economy is no longer present because the relation is in fact non-linear. Results suggest that the impact of oil price on the economy is much stronger after one accounts for non-linearity. Hamilton also noted that oil price shocks could no longer be treated as exogenous since they can in principle be supply or demand driven.

2.3 Model extension

While the typical approach in empirical literature was to treat oil shocks as exogenous disruptions in supply, Barsky and Kilian (2004) in their seminal paper stressed that demand shocks are at least as significant as supply shocks, and they also have different effects on the economy. Later Kilian (2009) decomposed oil price shocks to three different shocks. Oil supply shock, demand shock as a consequence of increased economic activity and precautionary oil demand shock. He proposed a structural vector autoregressive (SVAR) model of the global crude oil market. His results suggest that different sources of oil price shock affect gross domestic product (hereinafter: GDP) and consumer price index (hereinafter: CPI) differently. The essential deficiency of the initial model was the absence of any macroeconomic variables, not even monetary policy indicators. This deficiency causes biased results due to an omitted variable. However, this paper was an essential breakthrough for future research of oil shocks.

After the decomposition of the oil price shock on three primary sources, a whole new stream of research emerged. There was much interest in the effects of various oil shocks on other economic variables after only GDP and CPI were initially examined. The model proposed by Kilian (2009) was extended by variables of interest. Kilian and Park (2009) added the U.S. stock market to observe the response of stock prices on oil shock and vice versa. Kilian and Lewis (2011) studied the response of monetary policy on various oil price shocks and they found no evidence of endogenous monetary policy response and also no evidence which would support the hypothesis, that monetary policy caused substantial fluctuations in the U.S. economy.

In the last five years, another important concept on oil market emerged. Studies leaned towards the role of inventories and speculative trading on the oil market and the role of speculative shock. Hamilton (2009) proposed that speculation with futures on oil could become a dominant oil price determinant. To asses the role of speculative shock, Kilian and Murphy (2014) decomposed oil price shocks on four sources. This decomposition again allowed researchers to test responses of macroeconomic variables on shock from speculative oil trading.

3 EXPLANATION OF OIL PRICE FLUCTUATION

Modeling of oil price fluctuation and effects of oil shocks on economy broadly emerged after oil supply shocks in the 1980s. At first, oil supply shocks were seen to be a predominant source of oil price fluctuation and this reasoning was natural since oil price commonly raised after supply cuts. Later Kilian (2009) concluded that oil price was historically primarily driven by demand shocks. In his model he distinguished commodity demand shock defined as a rapid improvement of world economic activity and precautionary demand shock. Kilian (2009), however, wholly neglected the role of inventories even though they are an essential part of oil market since oil is a storable commodity. Oil reserves can also be seen as a policy instrument, as proposed by Medlock III (2013) because the government is willing to mitigate oil shocks to the economy. The inclusion of oil reserves is also useful for identification of specific demand shocks as presented in Kilian and Lee (2014), Juvenal and Petrella (2015), Kilian and Murphy (2014), among others. Even though they defined a rapid change in oil inventories as a speculative shock, this may not be the only source of it. By the emergence of crude oil derivative contracts, speculative shocks reflected on the oil price can also be driven by speculation in the financial market as argued by Parsons and Espinasa (2010) who suggested, that oil price hike during the period from 2003 and 2008 was a speculative bubble.

The inclusion of interest rate is not evident on a first sight, but since it is one of the monetary policy instruments, it is, similarly to oil inventories, used to buffer the effects of oil shocks on the economy. As explained in Bruno and Shin (2015) monetary policy also has a risk-taking channel which may also be present for increased risk in the oil market. Oil production is also affected by interest rates since the restrictive monetary policy increases cost of oil production financing, and oil production decreases according to the theory.

Whereas oil supply shocks are easily observed and included in the model by only one variable, demand shocks can be further divided into three different demand shocks which also have different effects on the economy. The identification of oil supply shock and shock of global demand for industrial commodities is straightforward. However, a division of residual oil price fluctuation on the speculative component from financial market and oil specific demand shock is non-trivial. The first demand shock is defined as increased demand due to increased economic activity. The second demand shock is a consequence of precautionary oil demand by retail companies, refineries or a change in the strategic policy of the government. The third shock is defined as a speculative oil demand shock from the financial market, where oil derivative contract is bought from the sole anticipation of the higher future price. Since the emergence of futures contracts, speculative shocks gained importance and they have to be considered in the analysis of oil market. Speculative demand shock on financial market and oil specific demand shock are different from global commodity demand shock because it is driven by expectations in the crude oil market.

In my econometric model, I represent the oil market by five oil specific variables, where I follow Kilian and Murphy (2014), Kilian and Lee (2014) and Juvenal and Petrella (2015), among others. Whereas the listed authors transformed oil inventories to flow variable by first differencing, I kept it as a stock variable to consider a stock-flow model following Medlock III (2013). The variable choice is aimed towards the identification of four oil market shocks: oil supply shock, oil-specific demand shock, shock in global demand for industrial commodities and speculative oil demand shock on the financial market. Additionally, I considered the monetary policy shock to observe the reaction of oil-related variables to contractionary monetary policy.

Medlock III (2013) assumed that oil market is composed of two markets, namely an inventory (stock) market and a flow market and the price is determined simultaneously on both markets. In the stock market, the inventories are valued according to the current demand for future supply. The stock market is therefore a link between the expected future price and the current price. In the flow market, the current market clearing price is formed so that oil supply and oil demand are met. This partial market equilibrium or a short-term equilibrium would also be a final market equilibrium price, under the condition, that oil would not be a storable commodity. Long-term equilibrium price will, however, depend also on inventory adjustment and finally, stock and flow markets are cleared at the same price.

3.1 Oil supply shock

Oil supply shock is defined as a disruption in the global production of crude oil. Global crude oil production is a flow variable since it represents a continuous change of underground inventories to above-ground inventories. These shocks were, historically speaking, mostly caused by political disruptions in the Middle East. However, production cut may also be driven by the expectation of higher price in the future. Juvenal and Petrella (2015) refer to such shocks as a speculative supply shock.

The most dramatic oil supply shock occurred in 1973-74 when the nominal price of oil quadrupled in half a year. Even in this period, supply shock manages to explain at most 25 % of the observed price increase in the period. The rest of price increase was driven by demand factors as explained in Barsky and Kilian (2001). Global demand boom in all industrial commodity markets was a consequence of business cycle peak in the U.S., Europe and Japan for the first time in postwar history. When prices of raw materials and metals increased by around 95 %, the real price of oil increased by 125 % as a consequence of supply cut in the Middle East (Kilian, 2014, p. 137-138). Oil supply shocks therefore occur in times of economic activity peak because oil producers want to increase the price additionally.

Kilian (2008) analyzed exogenous oil supply shocks and found out, that historically, oil supply shocks were of lesser importance for oil price formation. However, the expectations about substantial supply shocks could have a significant effect, like in the Persian Gulf war episode. He also points out, that if substantial oil supply shocks would occur, they could have an enormous effect on the economy.

3.2 Shock in global commodity demand

Shock in global commodity demand is defined as a rapid increase in global economic activity, hence, an increase in demand for crude oil and other industrial commodities. According to Kilian (2014), flow demand is the demand for oil to be consumed immediately by producing oil derivatives. By expansion of the global economy, flow demand for oil increases because it is a necessary production factor. Since the measure of economic activity is carefully determined by an index proposed by Kilian (2009) reflecting economic activity relevant for commodity market analysis, these shocks proved to be an essential driver of oil prices. Ever since Kilian (2009) identified, that demand shocks are the most critical drivers of oil price, researchers considerably leaned to the demand side of the oil market.

Further analysis of Kilian (2009), where the detrended index of OECD industrial production is used for a measure of economic activity instead of Baltic Dry Cargo Index, suggest that unexpected high growth of oil demand was not caused by growth of OECD countries, but primarily by unexpected high growth of non-OECD countries (Kilian & Hicks, 2013, p. 385). Since the economy in those countries was in a developing state, the process of industrialization caused even stronger pressure on oil price. There is the correlation between the economic growth and oil price, and this relation is also plausible since entering the industrial stage of development has the most substantial effect on oil price. Further transition to service stage of development contributes very little to oil consumption, mostly through increased purchase power of residents and not so much for industrial needs. Consequently, the demand pressure moved from OECD countries to non-OECD countries which entered industrial development stage by a lag (Medlock III, 2013, p. 18-19).

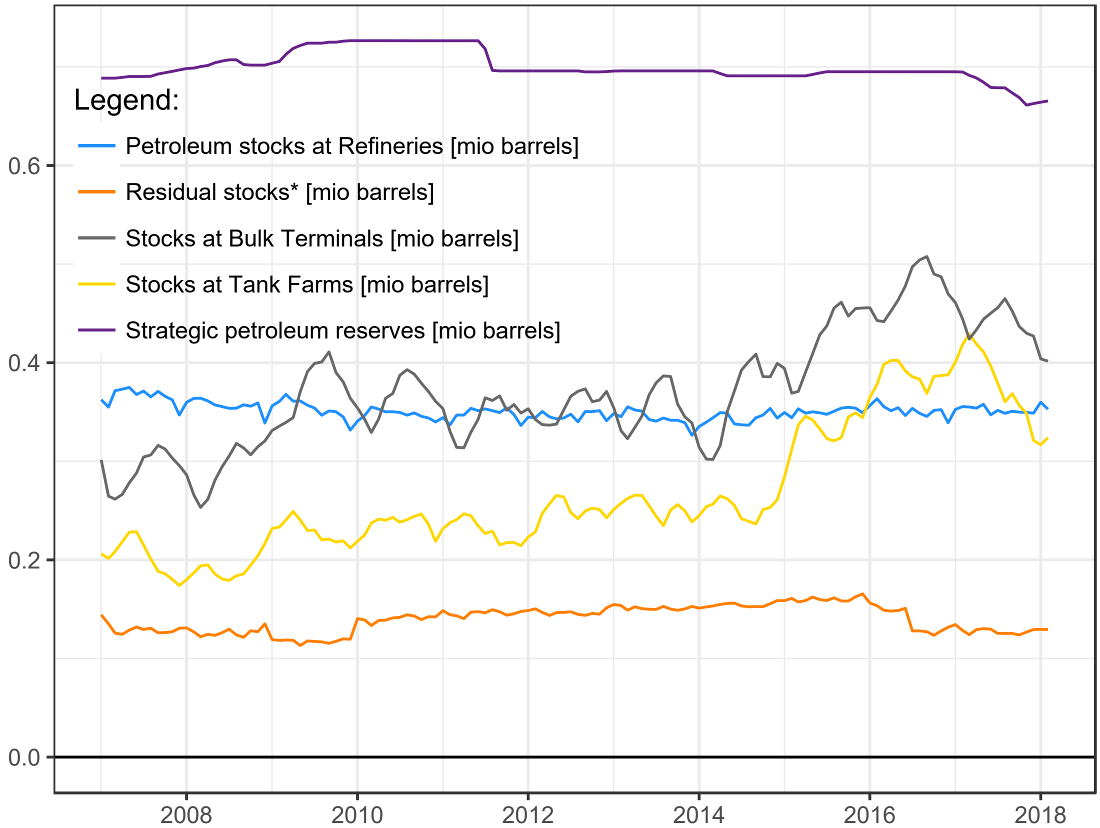
3.3 Oil specific demand shock

Oil specific demand shock or oil inventories shock is in my setting assumed to be a negative shock to a **stock** of oil inventories. Since many OECD countries have governmentally determined inventories of oil, the stock of oil returns to the prior shock quantities. This is a different set up as it was proposed in Kilian and Murphy (2014), Kilian and Lee (2014), Kaufmann (2011), where they use change in oil inventories as a variable included in the model and increased oil inventories were assumed to be a speculative shock. Another stream of studies Teisberg (1981), Considine (1997), Roekchamnong, Pornchaiwiseskul and Chiarawongse (2014), Hubbard and Weiner (1986), Kucher and Kurov (2012) considered oil inventories as a buffer to oil shocks.

In general, inventories and their size are determined by market forces. Producers, consumers, and speculators arbitrage prices, so that it is profitable to hold inventories and moderate fluctuations in the balance between production and consumption. Jaffe and Soligo (2002) noted that in case of critical industrial commodities, inventories could not be left solely to market forces. Oil is an essential input for the overall economy and is also an essential strategic resource in time of war. After the oil price shocks at the beginning of the 1980s, industrialized oil importing countries adopted the policy of strategic reserves to create a buffer for possible future oil supply shortages. While the definition of supply shortage made sense back then, nowadays supply disruptions caused by OPEC should almost immediately reflect in higher oil price and not in the unavailability of oil. Strategic reserves are still important since they can be released to mitigate oil price increase. Governments intervention of setting up oil inventories is therefore justified by the protection of social good (Jaffe & Soligo, 2002, p. 403-404).

I decide not to interpret the change of inventories solely as a speculative shock, but as an oil specific demand shock, which may be a transmission of demand from futures market through the arbitrage channel or a strategic decision of a government, oil retailers or refineries. Oil inventories have a much broader meaning since they allow oil retailers and refiners to lower their costs since they are allowed to order larger quantities and to enter into a binding contract with the supplier. For example, if commercial banks decide to underwrite a hedging contract for any quantity, retailers are allowed to hedge their open position and hold more oil in stock. Consequently, global oil inventories increase even without expectations of a higher price in the future. Even though retailers usually buy and sell refined oil products like gasoline, diesel, kerosene and other oil derivatives, they are essential in oil inventories data, since refined products are also included in OECD oil inventories.

Figure 1: Structure of U.S. oil inventories



*Note: residual stocks comprise stocks in transit from Alaska, stock in pipelines and stocks at leases

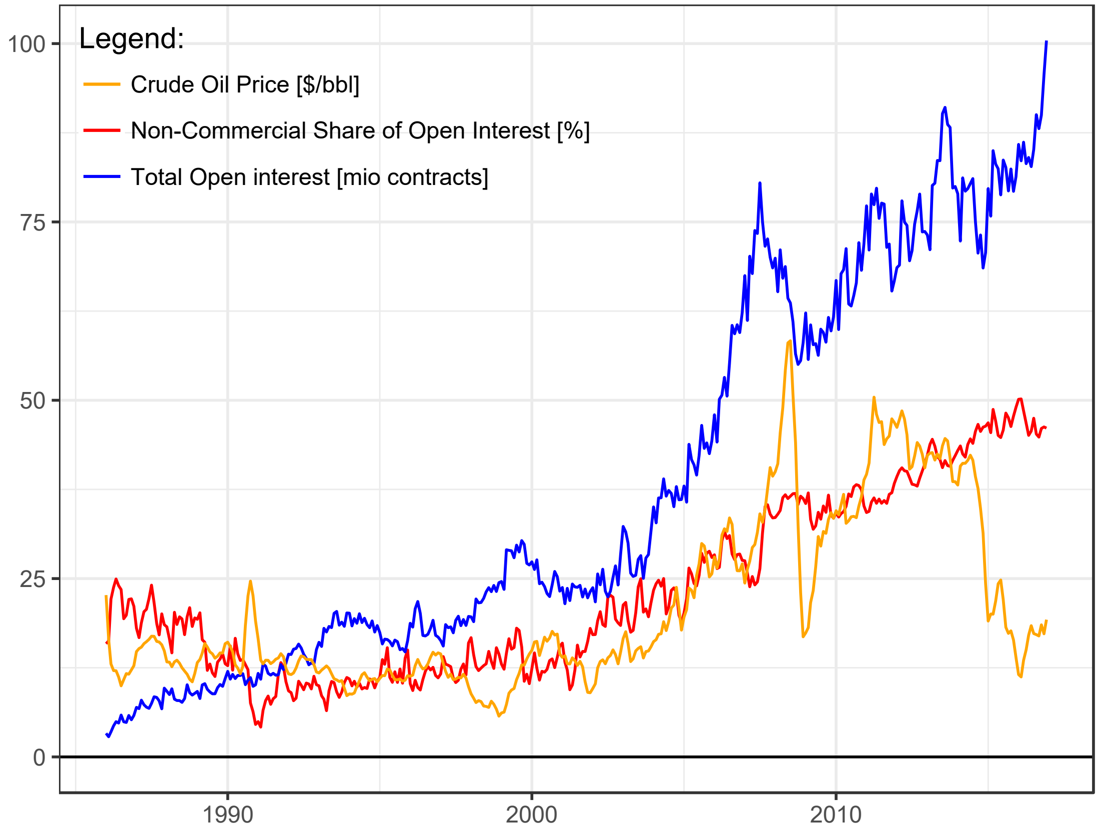
Source: Energy Information Administration (no date a).

In Figure 1, the structure of total U.S. inventories is presented. Residual stocks contain stocks in transit (on ships) from Alaska, stocks in pipelines and stocks at leases. The structure of U.S. oil inventories changed in the last decade. The inventories held by oil companies at tank farms and bulk terminals increased, and this means that the proportion of refined oil products in total oil inventories increased. The reason for an increased share of refined products in total inventories may be in the increasing interest of financial institutions to hedge smaller quantities held by smaller oil retail companies. This allowed them to order and store substantial amounts of refined oil products which lead to lower costs. U.S. oil inventories are therefore very diversified, and it is difficult to make a conclusion which factor mostly drives them. Of course, arbitrage after a speculative activity on the financial market may be one of the sources of oil inventory fluctuation but it is not the only one and a modeling speculative shock with a change in oil inventories, is therefore problematic for this reason.

3.4 Speculative shock from financial part of oil market

Another oil shock that I consider is the speculative shock from the financial part of oil market which is defined as a sudden increase of speculative pressure proxy. This means, that the difference in long non-commercial open interest increases over short non-commercial open interest and that speculators expect oil price to increase in the future.

Figure 2: Open Interest, Market Structure and Price



Source: Energy Information Administration (no date b), CFTC (no date).

The financialization of oil market has become increasingly important in recent years. It rose a question of the effect of massive money inflow through over the counter and exchange-traded futures contracts. The primary question is, whether or not the financial part of crude oil market has an impact on the physical crude oil market and crude oil price. The oil price movement between 2008 and 2009 caused a great debate about the regulation of speculative activity in oil futures market, launched in the U.S. by the Commodity Futures Trading Commission. The concern has risen from the fact that the share of non-commercial traders open interest rose significantly from around 10 % in the middle 1990s to 50 % in 2016 as presented in figure 2. Whereas commercial traders use financial contracts to offset risk in their financial position and allows them to hold more inventories, non-commercial traders are willing to open their position to speculate on favorable price change (Medlock III, 2013, p. 22-24).

After the oil price surge from 2003 to 2008, increasing share of non-commercial traders on financial market for crude oil was accused of causing increasing oil prices. The assumption of causality was based on correlation observed in Figure 2. According to Fattouh, Kilian and Mahadeva (2013), the broadest definition of speculation is the changing position of the asset without changing the actual consumption. This definition can be simply divided into the financial and physical market, where both markets are linked through arbitrage. In this definition of speculation, there is no negative connotation. Physical accumulation of oil is usually justified by production smoothing and market makers on financial market are also essential to allow oil companies to hedge their position. The speculation that causes distress in the market is the excessive speculation. According to Fattouh, Kilian and Mahadeva (2013), excessive speculation is the additional speculation, not needed for proper function of the oil market. Even though commodity markets attracted the speculators from financial markets, as shown in Alquist and Kilian (2010) and Hamilton and Wu (2014), the reason was also the excess liquidity in financial markets as a consequence of low interest rates. The remaining question is the effect of excessive speculation on the oil price.

In Figure 2, oil price, all traders open interest and share of non-commercial traders open interest are plotted. The massive inflow of non-commercial traders completely changed the composition of financial part of the crude oil market. It transformed from risk-taking channel to the new source of price volatility. Non-commercial traders also contributed to the overall increase of open interest presented by the blue line in Figure 2. Even though, correlation does not mean causation, the high correlation of total open interest and share of non-commercial traders with crude oil price rises concern about the impact of speculation on the physical market. This effect is questionable since oil is a storable commodity and any speculative pressure should reflect solely on inventory adjustment through arbitrage in the long run. There is, however, no theoretical reason to immediately reject short-run effect of speculative activity on the physical side of crude oil market (Medlock III, 2013, p. 23-27).

Hamilton (2009) discussed the importance of the oil forward market in the determination of future and spot oil price. By affection of future price ($E_t P_{t+1}$), the incentives of producers may change, and consequently, the supply side of the market is affected. For example, if commodity traders take a long position in an oil futures contract at a price F_t, and sell it before expiration at the higher price P_{t+1}. If expectations are such that $E_t P_{t+1} > F_t$, on the physical side of the market, producers will hold oil back from the market and accumulate their inventories (above and underground) to sell oil at a higher price (Juvenal & Petrella, 2015, p. 15-16).

Since the speculation can be formed on the physical market or with financial contracts, total inventories do not necessarily increase since the government may mitigate the price shock by decreasing the strategic part of inventories. Within my model, with separate treatment of inventories, the speculative pressure is oriented only to the financial part of the crude oil market. The residual shock is a consequence of other shocks, like weather, new technologies and similar.

3.5 Monetary policy shock

The last shock I consider within this model is the monetary policy shock. In inclusion of monetary policy in the oil market model, I follow Aastveit (2014) who noted that monetary policy reacts to consequences of oil price shocks and is essential in the model setting. The inclusion of monetary policy shock in oil market model is also advocated by Bodenstein, Guerrieri and Kilian (2012), who noted that the response of monetary policy is crucial for the understanding of the effects of oil shocks on the economy. In this model, monetary policy shock is defined as a rapid increase in shadow federal funds rate, the monetary policy indicator proposed by Wu and Xia (2016). Including monetary policy shock within the oil market model is useful for at least two reasons. The first advantage is an observation of effects of increased cost of financing and lower purchase power on oil market variables. The second interest is to check the presence of ”price puzzle” after 3PRF factor estimation. Monetary policy is a response of the government to oil shocks and is meant to mitigate responses of various variables to oil shocks, and is therefore useful to explain economy response after oil shocks.

Balancing higher inflation and higher unemployment was historically problematic after oil price movements since oil shocks cause economic slowdown and higher inflation. Following the idea of Bohi (1991), Bernanke, Gertler, Watson, Sims and Friedman (1997) identified the cause of economic recessions, and they suggested, that monetary policy follow the goal of keeping inflation low at the cost of economic slowdown. The cause of the economic slowdown is in their opinion on the side of monetary policy makers who act anti-inflationary. In their paper, they did not take into account the endogeneity of oil price of a different source of the shock. Their conclusions were challenged by Hamilton and Herrera (2004) who point out infeasibility of Federal Reserve Bank (hereinafter: FED) policy to implement such a striking policy, and that estimated effect of oil shocks by Bernanke, Gertler, Watson, Sims and Friedman (1997) was too low. Bachmeier (2008) analysed the effect of oil shocks on the stock market, and he concluded that systematic monetary policy was not so effective as proposed by Bernanke, Gertler, Watson, Sims and Friedman (1997). Kilian and Lewis (2011) in contrast to other studies took into account endogeneity of oil prices and they also allow policy response to depend on the different cause of oil price shock. They found out that monetary policy responds directly to oil price shock and not to inflation as a consequence of a shock.

Even though most literature is focused to the response of monetary policy to oil price shocks, the reversal causality is not irrelevant. Frankel (1986) was the first who proposed that monetary conditions could in principle affect commodity prices. Barsky and Kilian (2001, 2004) suggest, that oil shocks in 1970s were at least partially caused by monetary policy shocks. They concluded that monetary policy affects commodity prices mainly through indirect channel of expectations about growth and inflation in the future. Frankel (2008) listed additional transmission channels of monetary policy to commodity market. First channel he mentioned is so called **inventory channel**, which relates to higher opportunity cost of inventory holding in case of restrictive monetary policy. The second channel, also related to opportunity costs, is **supply channel** since also higher production and holding less underground inventories is necessary in high interest rate environment. The last channel he mentioned is the **financial channel** through which the speculators in financial market affect the spot price by arbitrage. Again, in environment of high interest rates, speculative position is more expensive and therefore short position is expected to prevail. Anzuini, Lombardi and Pagano (2013) also found out that U.S. monetary policy actions affect commodity prices, especially oil prices. They explained that loose monetary policy leads to higher commodity prices also through channels proposed by Frankel (2008) and not only through expectations.

4 ECONOMETRIC MODEL

The empirical framework considered in the thesis is based on factor augmented vector autoregressive model (FAVAR), proposed by Bernanke, Boivin and Eliasz (2005). The most important novelty regarding basic VAR is the inclusion of factors, which represent the general economic condition. Whereas factors in baseline FAVAR are estimated with the PCA, factors in this model are estimated according to Kelly and Pruitt (2015) definition of the 3PRF. To identify structural shocks, the structural FAVAR model (hereinafter: SFAVAR) is considered.

Following Aastveit (2014) who made a comparison of impulse responses of FAVAR model with two alternative models, I compare FAVAR model with estimated factors according to 3PRF (hereinafter: $FAVAR_{3PRF}$) with baseline FAVAR model with PCA factors (hereinafter $FAVAR_{PCA}$) and baseline VAR model, where only five oil related variables and shadow federal funds rate are included.

4.1 Factor Augmented VAR

Typical VAR models include a limited number of variables, due to limited degrees of freedom. Policy institutions, along with oil producers and other economic agents, consider big datasets for investment planning, policy actions, and production planning. It is therefore essential to include all relevant information in the model. Otherwise, it became informationally deficient and estimated responses to shock would be distorted by omitted variable bias. Detection of relevant variables which should be included in the model is another difficult task because the variables for which data are available could be different from those included in theoretical models and therefore the most suitable proxy needs to be considered. The inclusion of additional variables to VAR may also be motivated by examination of the impact of particular shock at a more disaggregated level. One may be interested in the response of specific sectors of the economy, and therefore, the specific sectors need to appropriately represented in the model. VAR model is not well suited for situations described above, because the inclusion of many variables undermines the precision of the model estimates in small samples. A number of variables included in VAR can easily become binding since the number of parameters in a VAR model increases with the square of the number of variables included. The FAVAR models relax the binding number of variables because the model includes one or more factors in addition to observable variables. The inclusion of factors is a perfect solution to informationally augment an existing VAR model, where also identification methods are conventional to existing methods. Alternatively, one could consider the class of dynamic factor models, which does not include any observable variables since all observable variables are expressed as a weighted average of factors. In case of dynamic factor models, the identification of shocks is based on the restriction of responses of observed variables to a structural shock (Kilian & Lütkepohl, 2017, p. 535-536).

4.1.1 Dynamic Factor Models

FAVAR model is a subset of a broader spectrum of models, called factor models. Factor models are in general divided into exact factor models and approximate factor models, where at the latter, some of the assumptions are relaxed. Exact dynamic factor model can be represented by equation:

$$\mathbf{X}_t = \Lambda^*(L)\mathbf{F}_t + \xi_t, \tag{1}$$

where $\mathbf{F}_t$ are dynamic factors, $\mathbf{X}_t$ is the data set with the dimension $(N \times T)$, and ξ_t represents an idiosyncratic component, which can be represented as $\xi_t = \delta(L)\xi_{t-1} + v_t$. The $\Lambda^*(L)$ is a lag polynomial, also called the dynamic factor loading. The assumptions of exact factor models are, that idiosyncratic disturbances and factors are uncorrelated at all leads and lags, $E(\mathbf{F}_t, \xi_{is}) = 0$ for all i, t, s and also that the error terms are uncorrelated at all leads and lags, such that $E(\xi_{it}, \xi_{j,s}) = 0$ for all i, j, t, s where $i \neq j$. The main difference between approximate and exact dynamic factor model is the relaxation of the second assumption in case of approximate factor model (Stock & Watson, 2005, p. 5-6).

The model (2) can become an exact dynamic factor model by the following representation:

$$\mathbf{X}_t = \Lambda(L)\mathbf{F}_t + \delta(L)\mathbf{X}_{t-1} + v_t, \tag{2}$$

where $\Lambda(L) = (1 - \delta(L)L)\Lambda^*(L)$. The complete representation of dynamic factor model consists from equation (3) and the equation describing the dynamics of factors:

$$\mathbf{F}_t = \Gamma(L)\mathbf{F}_{t-1} + \eta_t, \tag{3}$$

where $\Gamma(L)$ is a matrix lag polynomial and η_t is an error term (Stock & Watson, 2005, p. 6-7).

The basic idea of dynamic factor models is the observation that a dataset $\mathbf{X}_t$ with many time series variables is driven by only a few factors ($\mathbf{F}_t$) and an error term (v_t) (Stock & Watson, 2011, p. 3). If a big data set $\mathbf{X}_t$ is indeed related to only a few factors ($\mathbf{F}_t$) and if both idiosyncratic errors (v_t and η_t) are Gaussian, any variable in $\mathbf{X}_t$ can be efficiently estimated by factors and own lags. This property is also the reason for the inclusion of factors in FAVAR model since factors can informationally improve the VAR model in a very efficient way. The forecast is based only on R factors, which are assumed to contain most of the information from N variables in $\mathbf{X}_t$, and $R << N$. Forecast of variable x_{it} one period ahead can, therefore, be obtained as:

$$\begin{aligned}
E[x_{it+1}|\mathbf{X}_t, \mathbf{F}_t, \mathbf{X}_{t-1}, \mathbf{F}_{t-1}, ...] &= E[\Lambda_i(L)\mathbf{F}_{t+1} + v_{it+1}|\mathbf{X}_t, \mathbf{F}_t, \mathbf{X}_{t-1}, \mathbf{F}_{t-1}, ...] \\
&= E[\Lambda_i(L)\mathbf{F}_{t+1}|\mathbf{X}_t, \mathbf{F}_t, ...] + E[v_{it+1}|\mathbf{X}_t, \mathbf{F}_t, ...] \\
&= E[\Lambda_i(L)\mathbf{F}_{t+1}|\mathbf{F}_t, \mathbf{F}_{t-1}, ...] + E[v_{it+1}|v_{it}, v_{it-1}, ...] \\
&= \alpha(L)\mathbf{F}_t + \delta(L)x_{it}.
\end{aligned} \tag{4}$$

This vital fact about the use of factor models for forecast arises directly from equations (2) and (3) and assumptions regarding dynamic factor models (Stock & Watson, 2011, p. 4). The error of the forecast equation (4) depends crucially on the method of factor estimation, where factors that contain more information relevant to x_{it} would produce lower forecast error.

4.1.2 Structural FAVAR

The most straightforward way of obtaining FAVAR model is to augment set of observed variables $\mathbf{z}_t$, in VAR model by unobserved factors, $\mathbf{F}_t$, extracted from a broad set of observed variables $\mathbf{X}_t$ that does not include $\mathbf{z}_t$ (Kilian & Lütkepohl, 2017, p. 551).

As an ordinary VAR model, also FAVAR can be represented in structural form since it is only an extension of VAR. Proposed structural FAVAR model for $\mathbf{C}_t = (\mathbf{F}_t', \mathbf{z}_t')'$ can therefore be represented as:

$$\mathbf{B}(L)\begin{bmatrix}\mathbf{F}_t \\ \mathbf{z}_t\end{bmatrix} = w_t, \tag{5}$$

where the vector of structural shocks $\mathbf{w}_t$ is $(R + M)$ dimensional white noise, $\mathbf{B}(L) = (\mathbf{B}_0 + \mathbf{B}_1 L + ... + \mathbf{B}_p L^p)$ is a $(R + M) \times (R + M)$ matrix operator. $\mathbf{F}_t$ is a vector of R unobserved common factors that are related to $N \times 1$ vector of informational variables x_t by the observation equation

$$\mathbf{X}_t = \begin{bmatrix}\Lambda^F & \Lambda^z\end{bmatrix}\begin{bmatrix}\mathbf{F}_t \\ \mathbf{z}_t\end{bmatrix} + e_t, \tag{6}$$

where Λ^F is the $N \times R$ matrix of factor loadings and Λ^z is an $N \times M$ matrix of coefficients. Equation (7) is basically the same as FAVAR model proposed by Bernanke, Boivin and Eliasz (2005). This is a standard representation of VAR model, except that the variables in $\mathbf{F}_t$ are unobservable and usually capture information of some structural shocks that are important to the economy but cannot be represented by specific macroeconomic aggregate (Kilian & Lüthepohl, 2017, p. 551).

Since the model represented by equation (7) cannot be fully identified without additional restrictions, Bai, Li and Lu (2016) determined the number of needed restrictions on factor representation. They found that $R^2 + RM$ restrictions are needed for identification, and they also considered three sets of identification restrictions.

Since the FAVAR(p) process (6) can be rewritten as:

$$\mathbf{C}_t = \Phi_1 \mathbf{C}_{t-1} + \Phi_2 \mathbf{C}_{t-2} + ... + \Phi_p \mathbf{C}_{t+p} + w_t \tag{7}$$

or in more detail as:

$$\begin{bmatrix} \mathbf{F}_t \\ \mathbf{z}_t \end{bmatrix} = \Phi_1 \begin{bmatrix} \mathbf{F}_{t-1} \\ \mathbf{z}_{t-1} \end{bmatrix} + \Phi_2 \begin{bmatrix} \mathbf{F}_{t-2} \\ \mathbf{z}_{t-2} \end{bmatrix} + ... + \Phi_p \begin{bmatrix} \mathbf{F}_{t-p} \\ \mathbf{z}_{t-p} \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ v_t \end{bmatrix}, \tag{8}$$

where Φ_p represents a matrix of coefficients at a particular lag.

From there, it follows that Ω and Δ are defined as:

$$\Omega = E(w_t w_t') = \begin{bmatrix} E(\epsilon_t \epsilon_t') & E(\epsilon_t v_t') \\ E(v_t \epsilon_t') & E(v_t v_t') \end{bmatrix} = \begin{bmatrix} \Omega_{\epsilon\epsilon} & \Omega_{\epsilon v} \\ \Omega_{v\epsilon} & \Omega_{vv} \end{bmatrix}, \tag{9}$$

$$\Delta = E(\mathbf{C}_t \mathbf{C}_t') = \begin{bmatrix} E(\mathbf{F}_t \mathbf{F}_t') & E(\mathbf{F}_t \mathbf{z}_t') \\ E(\mathbf{z}_t \mathbf{F}_t') & E(\mathbf{z}_t \mathbf{z}_t') \end{bmatrix} = \begin{bmatrix} \Delta_{\mathbf{FF}} & \Delta_{\mathbf{Fz}} \\ \Delta_{\mathbf{zF}} & \Delta_{\mathbf{zz}} \end{bmatrix}, \tag{10}$$

where ϵ_t and v_t represent the innovations corresponding to $\mathbf{F}_t$ and $\mathbf{z}_t$.

Stock and Watson (2005) present identification schemes, which are imposed on a vector moving average system and are therefore an extension of identification schemes of baseline VAR. The identification restrictions they considered, are shortly presented below.

- Timing restrictions involve the Wold causal chain, where a variable listed first causes all variables in the system contemporaneously, whereas the second variable only causes the first variable in the next period. The structure is imposed on residuals through the Cholesky decomposition of a variance-covariance matrix of the residual.
- Long-run restrictions are based on identification scheme of Blanchard and Quah (1989).
- Restrictions imposed on factor structure are imposed directly to factor loadings $\Lambda(L)$ in the first step when dynamic factor model is estimated.
- Sign restriction scheme proposed by Uhlig (2005) is an approach, where the restrictions about the sign are imposed on the evolution of impulse responses.

The alternative set of identification restrictions considered in Bai, Li and Lu (2016) are restrictions imposed on matrices Ω and Δ and defined as listed bellow.

- Parameters satisfy the conditions that $\Omega_{\epsilon\epsilon} = I_{R\times R}$, $\Omega_{\epsilon v} = 0$ and $\frac{1}{N}\Lambda^{F\prime}\Sigma_{ee}^{-1}\Lambda^F = Q$, where Q is a diagonal matrix with distinct decreasingly arranged elements.

– Parameters satisfy the conditions that $\Omega_{\epsilon\epsilon} = I_{R\times R}$, $\Omega_{\epsilon v} = 0$ and Λ^F is a lower triangular matrix, where Λ^F is the upper $R \times R$ submatrix of Λ.
– Parameters satisfy the conditions that $\Omega_{\epsilon v} = 0$ and $\Lambda^F = I_{R\times R}$, where Λ^F is the upper $R \times R$ submatrix of Λ.

FAVAR models allow for many different identification restrictions, which become much more complex in cases, when shocks also enter the model through factors. In case of the model presented in this thesis, the identification scheme is a basic Wold causal chain, thus contemporaneous timing restriction.

In principle, there are multiple procedures available to estimate factors included in FAVAR. As already mentioned, the primary purpose of factor inclusion is data dimension reduction. In most studies, factors within FAVAR were components obtained from principal component analysis, where the aim is the maximization of explained variance contained in the dataset. Examples of such papers are the study of Bernanke, Boivin and Eliasz (2005), oil shock studies Aastveit (2014), Juvenal and Patrella (2015) and many studies of monetary policy effects. In the next section, I present an alternative factor estimation procedure, where factors aim to capture relevant information for the variable of interest and not the maximum amount of information in data set.

4.2 Three-Pass Regression Filter

The three-pass regression filter (3PRF) has been proposed by Kelly and Pruitt (2015) as a new estimator for factors, which are meant to forecast a single time series efficiently. Ordinary factors (components) are usually obtained directly from data matrix **X**, such that they capture as much variability in the data as possible. Then these factors are used to forecast y. 3PRF is based on the idea that the factors, that are **relevant**, to y are a strict subset of all the factors driving **X**. Accounting for y in data estimation procedure identifies factors relevant to the target and discards factors irrelevant for y even though this variable may be an essential driver of variation in **X** (Kelly & Pruitt, 2015, p. 294).

The aim of 3PRF is also to provide an improvement upon PCA. PCA condenses the cross-section according to **covariance within the predictors** (variables in **X**), whereas 3PRF condenses the cross-section according to **covariance with the forecast target** (y). Consequently, factors obtained by PCA drive the dataset **X** and some of them are **irrelevant** for the target. Another disadvantage of principal components regression is a compulsory estimation of all common factors to achieve consistency, including irrelevant ones. Since the 3PRF need only to estimate the relevant factors, a number of factors are always less or equal to PCA factors (Kelly & Pruitt, 2015, p. 295). This is especially advantageous within FAVAR models since a number of included variables is limited.

4.2.1 The Estimator

The three-pass regression filter is defined as a sequence of ordinary least squares (hereinafter: OLS) regressions. To use three-pass regression filter, we first need to determine the target variable y. There also exist many predictors contained in data set **X** which potentially contain relevant information for predicting the target. Kelly and Pruitt (2015) assumed additional variables called proxies **Z**, which are used to make a forecast. Proxies are variables assumed to be driven by factors (target relevant factors), and are always available from the target and predictors themselves or can alternatively be determined from economic theory (Kelly & Pruitt, 2015, p. 295-296). Along with other minor differences, the inclusion of proxies in factor estimation theory distinguishes 3PRF from similar target oriented factor techniques like partial least squares (hereinafter: PLS) which is proven to be a particular case of 3PRF. With proxy related factor estimation, the model with 3PRF factors can still be related to theory.

Factor construction according to 3PRF is well described as a sequence of three OLS regression. The general procedure proposed by Kelly and Pruitt (2015) is described in Table 1 below.

Table 1: Three-Pass Regression Filter

Pass	*Description*
1.	Run time series regression of $\mathbf{x}_i$ on $\mathbf{Z}$ for $i = 1...N$, $x_{i,t} = \phi_{0,i} + \mathbf{z}_t'\phi_i + \epsilon_{i,t}$, retain slope estimate $\hat{\phi}_i$
2.	Run cross section regression of $\mathbf{x}_t$ on $\hat{\phi}_i$ for $t = 1...T$, $x_{i,t} = \phi_{0,t} + \hat{\phi}_i'\mathbf{F}_t + \eta_{i,t}$, retain slope estimate $\hat{\mathbf{F}}_t$
3.	Run time series regression of y_{t+1} on predictive factors $\hat{\mathbf{F}}_t$, $y_{t+1} = \beta_0 + \hat{\mathbf{F}}_t'\beta + \mu_{t+1}$, delivers forecast $\hat{y}_{t+1}$

Source: Kelly & Pruitt (2015, p. 296).

As described in the table, in the **first pass** involves going through N dimension of dataset **X** and running N separate **time series** regressions, where the predictor x_i is the dependent variable and proxy variables in **Z** are treated as regressors. Obtained coefficients $\hat{\phi}_i$ describe the sensitivity of the predictor to factors represented by the proxies. The **second pass** runs through T dimension of the matrix **X** and involves T separate **cross section** regressions. Rows in the matrix **X** are the dependent variables in the regression, while the coefficients estimated in the first pass $\hat{\phi}_i$ are the regressors. Coefficients estimated in the first stage map the cross-sectional distribution of predictors to latent factors. In the second stage, estimates of the factors are backed out according to the map from the first stage. The mapping would be straightforward if coefficients estimated in the first stage would be observable. However, since loadings are not truly observable, the estimated loadings are used instead (Kelly & Pruitt, 2015, p. 295-296).

Estimated factors $\hat{\mathbf{F}}_t$ are carried forward to the **third pass**, where a time series forecasting regression is performed. At this stage, target variable y_{t+1} is regressed on factors estimated in the second stage. The fitted value in third stage regression is the 3PRF forecast (Kelly & Pruitt, 2015, p. 296).

An alternative representation of 3PRF in its one-step representation can be described by the equation:

$$\hat{\mathbf{y}} = \mathbf{i}_T\bar{y} + \mathbf{J}_T\mathbf{X}\mathbf{W}_X\mathbf{Z}(\mathbf{W}'_{XZ}\mathbf{S}_{XX}\mathbf{W}_{XZ})^{-1}\mathbf{W}'_{XZ}\mathbf{s}_{Xy}, \tag{11}$$

where $\mathbf{J}_T \equiv \mathbf{I}_T - \frac{1}{T}\mathbf{i}_T\mathbf{i}'_T$ with $\mathbf{I}_T$ being a T-dimensional identity matrix and $\mathbf{i}_T$ being T-vector of ones. $\mathbf{J}_N \equiv \mathbf{I}_N - \frac{1}{N}\mathbf{i}_N\mathbf{i}'_N$, $\bar{y} = \mathbf{i}'_T\mathbf{y}/T$, $\mathbf{W}_{X,Z} \equiv \mathbf{J}_N\mathbf{X}'\mathbf{J}_T\mathbf{Z}$, $S_{XX} \equiv \mathbf{X}'\mathbf{J}_T\mathbf{X}$ and $\mathbf{s}_{X,y} \equiv \mathbf{X}'\mathbf{J}_T\mathbf{y}$. All $\mathbf{J}$ matrices enter the formula due to constants included in OLS regressions presented above. If the matrix $\mathbf{X}$ is standardized, and proxies $\mathbf{Z}$ are defined as a target variable, the constant is the only difference between 3PRF and PLS (Kelly & Pruitt, 2015, p. 296).

Forecast can be rewritten as:

$$\begin{aligned}
\hat{\mathbf{y}} &= \mathbf{i}_T\bar{y} + \hat{\mathbf{F}}\hat{\beta} \\
\hat{\mathbf{F}}' &= \mathbf{S}_{ZZ}(\mathbf{W}'_{XZ}\mathbf{S}_{XZ})^{-1}\mathbf{W}'_{XZ}\mathbf{X}' \\
\hat{\beta} &= \mathbf{S}_{ZZ}\mathbf{W}'_{XZ}\mathbf{S}_{XZ}(\mathbf{W}'_{XZ}\mathbf{S}_{XX}\mathbf{W}_{XZ})^{-1}\mathbf{W}'_{XZ}\mathbf{s}_{Xy}
\end{aligned} \tag{12}$$

where $\mathbf{S}_{XZ} \equiv \mathbf{X}'\mathbf{J}_T\mathbf{Z}$. In this setting $\hat{\mathbf{F}}$ is interpreted as a predictive factor and $\hat{\beta}$ the estimated coefficient on that factor. Since the number of factors within $\hat{\mathbf{F}}$ is R and the number of columns of matrix $\mathbf{X}$ is N, where $R << N$, the 3PRF significantly reduces the dimensionality of forecasting problem. To show the relation of 3PRF to the regular OLS forecast can be rewritten as:

$$\begin{aligned}
\hat{\mathbf{y}} &= \mathbf{i}\bar{y} + \mathbf{J}_T\mathbf{X}\hat{\alpha} \\
\hat{\alpha} &= \mathbf{W}_{XZ}(\mathbf{W}'_{XZ}\mathbf{S}_{XX}\mathbf{W}'_{XZ})^{-1}\mathbf{W}'_{XZ}\mathbf{s}_{Xy}
\end{aligned} \tag{13}$$

interpreting $\hat{\alpha}$ as the predictive coefficient on individual predictors after inclusion of them through 3PRF factors. In the simple OLS estimation, the projection coefficient $\hat{\alpha}$ is $(\mathbf{S}_{XX}^{-1}\mathbf{s}_{Xy}$. The 3PRF can therefore be interpreted as a constrained version of least squares.

4.2.2 Relation to PCA

Principal components analysis (PCA) and principal component regression (hereinafter: PCR) was a predominant type of analysis and forecasting in the big data environment. Both, 3PRF and PCR can be calculated instantaneously for optional N and T. Minor difference emerges with missing data and unbalanced panels, which are well handled with 3PRF and very problematic for PCR since the components could not be obtained.

The main difference between both methods is the way of condensation of the panel of predictors. PCR condensed the data set according to **covariance within the predictors**. This identifies the factors driving the panel of predictors in decreasing order according to the amount of explained volatility within **X**. Identified factors drive the panel of predictors, some of which may be irrelevant for the dynamics of the variable of interest. Moreover, these factors are then used to forecast. 3PRF, on the other hand, condenses the data set according to **covariance with the forecast target** and this means that factors are by definition relevant for forecast target (Kelly & Pruitt, 2015, p. 295). Factors obtained by PCA may, therefore, be relevant for some applications, but if interest is directed to more specific sectors of the economy, principal components do not contain all relevant information for the analysis. If a variable irrelevant to the target but the important driver of variation in the **X** is added to original dataset **X**, 3PRF factors will not change due to zero loading to this additional variable but PCA factors will change significantly because factors are not target-oriented but are leaning towards maximal explanation of variability in **X**. Boivin and Ng (2006) found out that factors estimated by PCA can lead to rather different results if additional variables, irrelevant for the target, are considered. This is not the case with 3PRF factors, which should in principle have zero loading on all irrelevant explanation variables.

In the 3PRF original paper, Kelly and Pruitt (2015) performed several applications of 3PRF, and they also compared it to PCR. In the first example, they considered the forecastability of macroeconomic aggregates using quarterly data. Target variable was removed from the data set before factor estimation. They found that even conventional PCR with only one principal component performs very well. However, they found out that the 3PRF provides the best forecast in eight of thirteen series, where for two of these, its outperformance was statistically significant. This indicates that PCR performance strongly depends on the target and 3PRF is more flexible in this respect.

In their second example, they considered forecasting of market returns. The extent of market return predictability was estimated by consideration of 25 log price-dividend ratios of portfolios sorted by market equity and book-to-market ratio. They assumed that the predictors take the form $pd_{i,t} = \phi_{i,0} + \phi_i'\mathbf{F}_t + \epsilon_{i,t}$, while the target has the form $r_{t+1} = \beta_0^r + \mathbf{F}_t'\beta^r + \mu_{t+1}^r$. In this example, 3PRF achieved strong out of sample performance and it also significantly outperformed PCR. They have also shown in the example, that only one to two 3PRF factors were needed to achieve informational sufficiency, whereas Bayesian information criteria proposed four to five factors for principal components or factor model (Kelly & Pruitt, 2015, p. 300-302).

4.3 The Model

The model that I consider in the thesis is based on a factor-augmented vector autoregressive (FAVAR) model. The main feature of FAVAR model is the exploitation of all information available in a large data set. In this model, the information included in the model is response-dependent and thus estimated by 3PRF. Additionally, the structure is imposed on the model, and thus SFAVAR model is estimated. Variables considered in the model are chosen so that four possible oil market shocks were considered along with a monetary policy shock and factors which contain information form a larger dataset. In this setting I was able to estimate responses of variables to **oil supply shock**, **demand shock related to all commodities**, **oil specific demand shock**, **speculative oil demand shock** and **residual oil demand shock**. The sixth shock considered is the **monetary policy shock**.

The reasons for inclusion of shadow federal funds rate (monetary policy instrument) are twofold. Monetary policy could affect oil market through various channels, described explicitly in Frankel (2008), and it also reacts differently to various shocks in the oil market. By inclusion of monetary policy in the oil market model, I followed Aastveit (2014) who obtained interesting results regarding monetary policy reactions to oil shocks. However, his model did not include oil inventories, and therefore he did not distinguish between oil-specific demand shock and speculative demand shock.

The $\text{FAVAR}_{\text{3PRF}}$ is a new approach for inclusion of information available in a large data set. Such model includes ”shock variables” which are observable and latent factors, which are ”response-dependent” since they are leaned towards the particular variable of interest. Response dependency is also the main difference between $\text{FAVAR}_{\text{3PRF}}$ and ordinary $\text{FAVAR}_{\text{PCA}}$. In this setting I compare $\text{FAVAR}_{\text{3PRF}}$ with ordinary $\text{FAVAR}_{\text{PCA}}$ and also with informationally insufficient standard VAR and check how different the results are.

Another difference between the model presented in this thesis and other oil market models (Juvenal & Petrella, 2015; Aastveit, 2014; Kilian & Murphy, 2014) is consideration of stock-flow model following Medlock III (2013). In this setting, oil inventories are considered as a stock variable whereas oil supply and commodity demand are flow variables. In this setting, I also consider a slightly different identification of oil shocks, as I will present later.

4.3.1 Factor Estimation Procedure

As already noted, factor estimation is based on the three-pass regression filter. To calculate response dependent factors (specific factors for every variable considered), I considered automatic proxy selection, a special case of 3PRF, where the variable of interest y is chosen to be a proxy z. Choosing y being a proxy is an obvious choice since the factor has to summarize the variation in $\mathbf{X}$ relevant for forecast target (Guérin, Leiva-Leon & Marcellino, 2017). The resulting factors are therefore leaned towards forecasting target y. Target variable itself satisfies necessary assumptions for proxy selection when there is only one factor ($R = 1$). If the true number of factors is more than one ($R > 1$) the 3PRF procedure does not extract enough relevant information to forecast the target efficiently. Therefore additional one that depends only on relevant factors needs to be chosen. This additional proxy can be residual from the target 3PRF forecasts since those also have a non-zero loading on relevant factors (following from insufficiency of the target only proxy), have by definition zero loading on irrelevant factors and are linearly independent of the first factor (Kelly & Pruitt, 2015, p. 299).

Before Kelly and Pruitt (2015), Bai and Ng (2008) recognized the importance of omitting irrelevant information. They considered two types of threshold rules, to determine which variables are relevant for forecast target. After variables filtering, the factors are estimated according to PCA, and these factors are also relevant to the forecast target. The factor mapping is still the same, except that factors are estimated on the subsample of **X**. According to the 3PRF methodology, the first step is skipped, and an alternative mapping is considered.

Proxy construction, therefore, proceeds iteratively for any number of desired factors. In the first step, target variable is considered as a proxy in the 3PRF algorithm. After factor estimation, the forecast of the target is performed by the first 3PRF factor. Residual of the forecast is then considered as a proxy for estimation of the second factor. In this way R automatic proxy 3PRF factors can be iteratively estimated as explained in Table 2 below.

Table 2: Automatic Proxy-Selection Algorithm

Step	*Description*
0.	Initialize $r_0 = y$ For $k = 1, ..., R$:
1.	Define the k^{th} automatic proxy to be r_{k-1}. Stop if $k = R$ otherwise proceed.
2.	Compute the 3PRF for target y using cross section **X** using statistical proxies 1 through k. Denote the resulting forecast $\hat{y}_k$.
3.	Calculate $r_k = y - \hat{y}_k$, advance k and go to step 1.

Source: Kelly & Pruitt (2015, p. 299).

3PRF factors estimated according to automatic proxy selection algorithm are meant to be relevant to the forecast target. Within the $\text{FAVAR}_{\text{3PRF}}$ model, this means, factors have to be relevant for the analyzed variable. I, therefore, estimate new factors and also a new $\text{FAVAR}_{\text{3PRF}}$ model for each observed variable.

4.3.2 Estimation

Estimation of a $\text{FAVAR}_{\text{3PRF}}$, obtained by automatic proxy selection algorithm consists of three steps. In the first step I choose the variable of interest (target variable), then in the second step, I estimate R relevant factors, according to the procedure described in Table 2. Then I use estimated factors in step number three, where I estimate $\text{FAVAR}_{\text{3PRF}}$ model with those factors, where factors are treated like any other variable. The procedure is computationally demanding because $\text{FAVAR}_{\text{3PRF}}$ model needs to be estimated D times (D is a number of variables of interest). Estimation of a new model for every impulse response function is necessary, since the 3PRF factors are target oriented, whereas PCA components are general for the dataset **X**. The $\text{FAVAR}_{\text{3PRF}}$ function implemented in R language for statistical computing is available in Appendix 2. The function includes the library vars implemented by Pfaff (2008).

I assume that the state of economy is captured by a few common components, represented by the vector C_t. As already explained I include **change in oil production** ($\Delta prod_t$), **stock of oil inventories** (inv_t), **real economic activity measure** (rea_t) proposed by Kilian (2009), **speculative pressure** (sp_t) proxy to account for speculation from financial market, **real price of oil** (rpo_t), **shadow federal funds rate** (sr_t) proposed by Wu and Xia (2016) as a measure of monetary policy. Additionally, I include R 3PRF factors to achieve informational sufficiency. I assume that the dynamics of presented common components can be modelled as VAR of the following form:

$$C_t = A_0 + \Phi(L)C_{t-1} + u_t, \tag{14}$$

where

$$C_t = \left[\Delta prod_t, inv_t, rea_t, sp_t, rpo_t, \mathbf{F}_t', sr_t\right]', \tag{15}$$

and $\Phi(L)$ is a finite order lag polynomial. The error term u_t is assumed to be independent and identically distributed with zero mean. A represents a constant since some of variables in C_t have non-zero mean. Equation (15) is a VAR in C_t where $\mathbf{F}_t$ is a vector of factors. An important distinction from standard dynamic factor model presented in section 5.1.1 is an assumption that some of the factors which drive the economy are observable.

4.3.2.1 Model Specification

For efficient estimation of FAVAR model, the correct specification is essential. The first important decision is about the autoregressive order of the model, where multiple criteria have to be considered to prevent over-fitting and to obtain meaningful estimates (De Waele & Broersen, 2003, p. 427). After choosing the optimal lag order, the number of factors included in the model needs to be considered. After the emergence of FAVAR models, the necessary number of factors to achieve informational sufficiency was a crucial question to answer (Forni & Gambetti , 2014, p. 124). In a VAR model with many variables, it is even more important to include only as many factors as necessary, because of the limited number of free parameters. Finally, after the specification of FAVAR model, the theoretically sensible structure has to be imposed to the model, to obtain structural shocks.

4.3.2.1.1 Determining the Autoregressive Order

The choice of autoregressive order is based on multiple order selection criteria. I allow the highest lag order to be 15 and then I estimate a VAR by OLS. The model has the following form:

$$\tilde{C}_t = A_0 + A_1\tilde{C}_{t-1} + ... + A_p\tilde{C}_{t-p} + \tilde{u}_t, \tag{16}$$

where

$$\tilde{C}_t = \Big[\Delta prod_t, inv_t, rea_t, sp_t, rpo_t, sr_t\Big]'. \tag{17}$$

Model (17) is a constrained version of model represented by equation (15), since it does not include factors. The reason for that is the factor estimation procedure, where the number of lags is an essential input for determining the number of factors in the model. I therefore first determine an optimal number of lags for the constrained model (without factors) and then check if the number of lags would be different for model (15).

In equation (17) $\tilde{C}_t$ is a $M \times 1$ vector of endogenous variables and $\tilde{u}_t$ is a disturbance term of the same dimension. The coefficient matrices $A_1, ..., A_p$ are of dimension $M \times M$. For every lag number from 1 to 15 the model is estimated, and the following information criteria are computed:

$$AIC(p) = ln\ det(\tilde{\Sigma}_u(p)) + \frac{2}{T}pM^2, \tag{18}$$

$$HQ(p) = ln\ det(\tilde{\Sigma}_u(p)) + \frac{2ln(ln(T))}{T}pM^2, \tag{19}$$

$$SBC(p) = ln\ det(\tilde{\Sigma}_u(p)) + \frac{ln(T)}{T}pM^2, \tag{20}$$

$$FPE(p) = \left(\frac{T+n^*}{T-n^*}\right)^M det(\tilde{\Sigma}_u(p)). \tag{21}$$

$\tilde{\Sigma}_u(p) = T^{-1}\sum_{t=1}^{T}\hat{u}_t\hat{u}_t'$ and n^* is the total number of the parameters in each equation, and n assigns the lag order.

Table 3: Proposed Number of Lags by Four Information Criteria

Information criterion	*Number of lags (p)*
Akaike Information Criterion - $AIC(p)$	5
Hannan-Quinn Information Criteria - $HQ(p)$	2
Schwarz-Bayes criterion Criterion - $SC(p)$	2
Final Prediction Error Criterion - $FPE(p)$	5

Source: Own work.

As presented in Table 3, Akaike information criteria (AIC) and Akaike's final prediction error criterion (FPE) propose the fifth order of the model, whereas Schwarz-Bayes criterion

(SBC) and Hannan-Quinn criterion (HQ) propose the second order of the VAR. AIC and FPE criteria asymptotically overestimate the order with positive probability, whereas HQ and SBC criteria estimate the order consistently under quite general conditions if the actual data generating process has a finite VAR order and the maximum order is larger than the true order (Lütkepohl & Krätzig, 2004, p. 110). However, since other authors use more lags, Kilian (2009) uses 24 lags, Aastveit (2014) uses 13 lags, I decided to follow AIC and FPE and select the fifth order of the model. I later check for robustness of results after using fewer lags.

4.3.2.1.2 Sufficient Information and the Choice of Factors

The idea behind structural VAR and FAVAR models is the founding of structural economic shocks as a linear combination of the residuals. A requirement for the sensibility of such analysis is that the variables in VAR or FAVAR contain all relevant information. Informational sufficiency is thus implicitly assumed in SVAR and SFAVAR application. Any identification scheme can provide the correct structural shocks and impulse response functions if the VAR is informationally deficient (Forni & Gambetti, 2014, p. 125).

Bai and Ng (2002) proposed an alternative test to determine the number of factors. Their method, however, does not account for information already available within the VAR model. This is an essential drawback since the VAR models with many variables can in principle contain enough information and additional factors would not necessarily lead to better results.

To test for informational sufficiency and to determine the number of factors I follow Forni and Gambetti (2014) who proposed a test based on Granger causality. The intuition behind their procedure is the following: if dataset **X** contains information that Granger cause the variables included in the VAR then the VAR was informationally insufficient, and FAVAR has to be considered. In the construction of test statistics, I considered multivariate out-of-sample Granger causality test proposed in Gelper and Croux (2007), which is well suited for determination of the number of 3PRF factors. The reasoning of the procedure is that factor $\mathbf{F}_t$ Granger causes $\tilde{C}_t$ if it contains additional power in forecasting $\tilde{C}_t$ after controlling for the past of $\tilde{C}_t$. To establish Granger causality two models have to be compared. The full model with factors and the restricted model without factors. The question then is whether the **forecast** of full model is significantly better than the one of the restricted model (Gelper & Croux, 2007, p. 3320).

At the within-sample test, the risk of overfitting the data is substantial. If a too complex model is estimated, significant effects can be found to be spurious. As shown in Clark (2004) for univariate series, spurious effects can be avoided by out-of-sample testing (Gelper & Croux, 2007, p. 3327).

The full model has the following representation:

$$\begin{bmatrix} \tilde{C}_t \\ \mathbf{F}_t \end{bmatrix} = A_1 \begin{bmatrix} \tilde{C}_{t-1} \\ \mathbf{F}_{t-1} \end{bmatrix} + ... + A_p \begin{bmatrix} \tilde{C}_{t-p} \\ \mathbf{F}_{t-p} \end{bmatrix} + \epsilon_{f,t}, \tag{22}$$

where $\epsilon_{f,t}$ is a multivariate independently and identically distributed (iid) sequence with mean zero and the covariance matrix Σ_f, and the index t runs from $\pi + 1$ to T. It is worth noting that the ability of this model to forecast $\mathbf{F}_t$ is irrelevant. $\tilde{C}_t$ is of dimension M and F_t of dimension R. Test statistics is therefore based on bottom R rows of matrices $A_1, ...A_p$ and first M columns (bottom left part of matrices). Denoting by $\psi_1, ..\psi_p$ the $R \times l$ matrices of effect of F_t on $\tilde{C}_t$, the null hypothesis, stating that $\mathbf{F}_t$ does not Granger cause $\tilde{C}_t$ corresponds to:

$$H_0 : \psi_1 = \psi_2 = ... = \psi_p = 0. \tag{23}$$

Therefore under the null, the model (23) reduces to:

$$\tilde{C}_t = B_1\tilde{C}_{t-1} + ... + B_p\tilde{C}_{t-p} + \epsilon_{r,t}, \tag{24}$$

where $\epsilon_{r,t}$ is a multivariate independently and identically distributed sequence with mean zero and covariance matrix Σ_r. The restricted model (25) is compared to model (23) to test for Granger causality (Gelper & Croux, 2007, p. 3322). The out-of-sample test is conducted in three steps described in the table below.

Table 4: Testing Procedure

Step	*Description*
0.	Set $\tilde{C}_t = \left[\Delta prod_t, inv_t, rea_t, sp_t, rpo_t, sr_t\right]'$ For R in $1, ..., 6$:
1.	Divide $\tilde{C}_t$ on learning window $(1, ..., \pi)$ and forecasting window $(\pi + 1, ..., T)$.
2.	Forecast observations $\pi + 1, ..., T$ using a recursive scheme and retain forecast errors of full model (u_f) and restricted model (u_r).
3.	Compare forecasting performance of full model and restricted model by comparison of u_f and u_r. Add R factors to $\tilde{C}_t$.
4.	If $R < 6$ return to step 1.

Source: Gelper & Croux (2007, p. 3322).

I compare forecast error matrices according to the approach of Harvey, Leybourne, Newbold (1998), who describe how no Granger causality corresponds to zero correlation between $u_{r,t}$ and $u_{r,t} - u_{f,t}$. If we look for the best forecast combination of $\hat{C}_t^{OPT}$ by combination of $\hat{C}_{r,t}$ and $\hat{C}_{f,t}$:

$$\hat{C}_t^{OPT} = (1 - \gamma)\hat{C}_{r,t} + \gamma\hat{C}_{f,t}. \tag{25}$$

If γ equals zero, additional predictors included in the full model are irrelevant for the prediction and there is no Granger causality, all information in full model is therefore also contained in the restricted model. If I define the error of combined forecast as $e_t = \tilde{C} - \hat{C}_t^{OPT}$, and after using definitions of $u_{r,t}$ and $u_{f,t}$ from expression (26), it follows that:

$$u_{r,t} = \gamma(u_{r,t} - u_{f,t}) + e_t. \tag{26}$$

Testing whether $\gamma = 0$ therefore means zero correlation between $u_{r,t}$ and $u_{r,t} - u_{f,t}$ and implies rejection of Granger causality (Gelper & Croux, 2007, p. 3323-3324).

As mentioned, 3PRF factors are target oriented. I therefore run the procedure described in Table 5, M times (for each variable in $\tilde{C}_t$) to construct error matrices $u_{r,t}$ and $u_{f,t}$ respectively.

The procedure of obtaining PCA factors is simpler, and the out-of-sample forecast testing procedure in Table 4 runs only once. All variables in $\tilde{C}_t$ are forecasted simultaneously, and errors of full model and restricted model are written in matrices $u_{r,t}^{PCA}$ and $u_{f,t}^{PCA}$. I performed the test for PCA factors to compare results with other papers and with results for 3PRF factors.

The test statistics is based on direct OLS estimation of the regression model (27). Under the null, the expected value of estimated γ is close to zero. The hypothesis $H_0 : \gamma = 0$ can be tested according to a likelihood ratio test as:

$$LR = P(ln(|\mathbf{u}_r'\mathbf{u}_r|) - ln(|\hat{\epsilon}'\hat{\epsilon}|)), \tag{27}$$

where $\hat{\epsilon}$ is the $(J \times M)$ residual matrix obtained from regression (26) and J is the length of forecast window $(T - (\pi + 1))$ (Gelper & Croux, 2007, p. 3327).

The asymptotic distribution of LR test statistics is difficult to obtain in a multivariate setting, since the calculation is based on the month ahead forecast rather than original data. The critical values are therefore calculated by a residual based bootstrap (Gelper & Croux, 2007, p. 3325). The distribution of test statistics is bootstrapped under the assumption that H_0 holds true and therefore $\mathbf{u}_r = \mathbf{u}_f$. Since all residuals are identically and independently distributed, random resampling of both matrices can be formed such that random elements in $\mathbf{u}_r$ are replaced with random elements in $\mathbf{u}_f$ and vice versa. By iterating this procedure and computing LR test, I obtain the bootstrapped distribution of test statistic, and approximate p-value is then calculated as:

$$\hat{p} \approx B^{-1} \sum_b 1(|LR^*| > |LR^*_{obs}|), \tag{28}$$

where B denotes the number of occurrences that test statistics after resampling is higher than the observed test statistics. In the numerator is the number of bootstrap iterations (MacKinnon, 2009, p. 184).

Test results are presented in Table 5 and indicate, that in case of 3PRF estimation of relevant factors, one factor would be enough to achieve informational sufficiency and the forecast with the first factor was significantly better than the forecast without the factor. However, since the contribution to forecasting quality of the second factor is also weakly significant, I decided to include two factors in the model. Results for PCA factors are similar except that the fourth factor almost significantly improves the model. It is worth noting that 3PRF forecast oriented factors provide a more accessible choice of the number of factors since the relevant information for the model contained in sequential factors is decreasing by construction.

Table 5: Results of Test for Sufficient Information (p-values)

Factor	*3PRF*	*PCA*
1	0.028	0.029
2	0.051	0.044
3	0.155	0.203
4	0.163	0.054
5	0.199	0.127
6	0.205	0.219

Source: Own work.

I finally choose two factors for both models, $\text{FAVAR}_{\text{3PRF}}$ and for ordinary $\text{FAVAR}_{\text{PCA}}$ to make a relevant comparison of both models. Other authors, however, include more factors in similar models, for example, Aastveit (2014) arbitrarily included 5 PCA factors whereas Juvenal and Petrella (2015) performed a similar test and included four factors in the model. They, however, did not include monetary policy and specific variable for speculative pressure from the financial market in the model, and they used quarterly data.

4.3.2.2 Structural Analysis

After the establishment of a data generation process, the dynamic interaction between variables can be analyzed. Analysis of this sort is usually referred to as an impulse response analysis since the effects of shock in one variable to other variables are considered. VAR and FAVAR models in their basic form have no economic interpretation since they only summarise the dynamics in the data and are subject to the Lucas critique (Lütkepohl & Krätzig, 2004, p. 159).

Since baseline VAR models were often criticised and also their use in economic research was limited, Sims (1980, 1986), Bernanke (1986), Shapiro and Watson (1988) proposed an alternative model class, based on VAR models. Proposed structural VAR (SVAR) model focuses on residuals of VAR model and interprets them as a linear combination of exogenous shocks. The innovation in one structural shock is then tracked down to observe responses of other variables in the system (Lütkepohl & Krätzig, 2004, p. 159). Similar structure can be imposed to FAVAR models. According to Stock and Watson (2005), the identification scheme presented below is a natural extension of the Wold causal chain, also considered in Aastveit (2014).

It is worth noting that in this thesis I concentrate mainly on oil shocks and, therefore, the structural shocks from factors are not considered. Therefore the economic interpretation of 3PRF and PCA factors, proposed by Belviso and Milani (2006) is not considered. The procedure of 3PRF factor estimation by a response relevant factors also differs substantially from the concept of the factor as the source of the shock.

The equation (15) can alternatively be represented as

$$B_0C_t = A + B_1C_{t-1} + B_2C_{t-2} + ... + B_pC_{t-p} + w_t, \tag{29}$$

and the error term is consequently expressed as

$$w_t = B(L)C_t - A, \tag{30}$$

where $B(L) = B_0 - B_1L - B_2L^2 - ... - B_pL^p$. The normalized variance-covariance matrix of the error term is calculated as $\Sigma_w = E(w_tw_t') = I_K$, meaning that a maximal number of shocks considered cannot exceed the length of C_t and that structural shocks are uncorrelated. These properties are however not sufficient for model (15) to be SFAVAR since shocks must have theoretic interpretation. Going backward in expressing (15) as a simple transformation of (30) yields:

$$\begin{aligned} B_0^{-1}B_0C_t &= B_0^{-1}A + B_0^{-1}B_1C_{t-1} + ... + B_0^{-1}B_pC_{t-p} + B_0^{-1}w_t \\ C_t &= A_0 + A_1C_{t-1} + ... + A_pC_{t-p} + u_t. \end{aligned} \tag{31}$$

This means that $u_t = B_0^{-1}w_t$ and therefore u_t is a weighted average of w_t. Parameters $A_0, ..., A_p$ and Σ_u can be consistently estimated by OLS, but since B_0 is unknown, w_t cannot be reconstructed. However, since $u_t = B_0^{-1}w_t$ holds true by construction, the variance of u_t is

$$\Sigma_u = E(u_tu_t') = B_0^{-1}E(w_tw_t')B_0^{-1'} = B_0^{-1}\Sigma_wB_0^{-1'} = B_0^{-1}B_0^{-1'}. \tag{32}$$

Σ_u is a system of nonlinear equations which can be solved by numerical methods. Since it is a symmetric matrix, it contains only $K(K+1)/2$ independent equations, and this is precisely the maximal number of parameters in B_0 that can still be solved numerically. The only way to solve the system of equations is therefore by imposing restrictions on B_0^{-1}.

Structural innovations w_t are thus obtained from u_t by orthogonalization of reduced form errors, which makes errors mutually uncorrelated. Since the shocks are uncorrelated, the effects of isolated shocks can be considered. The natural way of achieving orthogonalization is by Cholesky decomposition of the matrix Σ_u such that $PP' = \Sigma_u$, where P is lower triangular Cholesky decomposition of Σ_u. From (32) it follows that $P = B_0^{-1}$ is obvious solution for the problem of recovering w_t. This identification scheme makes the models ”just identified” as opposed to over-identified models which include more restrictions than necessary to identify the system. According to the famous critique by Sims (1980), over-identified models are incredible, and therefore most SVAR models only have a necessary number of restrictions.

Since B_0^{-1} is a lower triangular matrix, B_0 is a lower triangular too. Consequently, the resulting SFAVAR is a recursive and causal relationship imposed to the system from a theoretical perspective rather than observed in the data. The ordering of variables in C_t is, therefore, crucial for sensible shock identification. This identification scheme is called Wold causal chain system after Wold (1960) or also timing scheme. In this system, the shocks successively enter the system such that variable ordered first in vector C_t affects all other variables instantaneously, whereas the variable ordered last affects other variables only in the next period.

Structural FAVAR model has, therefore, the following form:

$$\begin{bmatrix} u_t^{\Delta prod} \\ u_t^{rea} \\ u_t^{inv} \\ u_t^{sp} \\ u_t^{rpo} \\ u_t^{F_1} \\ u_t^{F_2} \\ u_t^{sr} \end{bmatrix} = \begin{bmatrix} S_{11} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ S_{21} & S_{22} & 0 & 0 & 0 & 0 & 0 & 0 \\ S_{31} & S_{32} & S_{33} & 0 & 0 & 0 & 0 & 0 \\ S_{41} & S_{42} & S_{43} & S_{44} & 0 & 0 & 0 & 0 \\ S_{51} & S_{52} & S_{53} & S_{54} & S_{55} & 0 & 0 & 0 \\ S_{61} & S_{62} & S_{63} & S_{64} & S_{65} & S_{66} & 0 & 0 \\ S_{71} & S_{72} & S_{73} & S_{74} & S_{75} & S_{76} & S_{77} & 0 \\ S_{81} & S_{82} & S_{83} & S_{84} & S_{85} & S_{86} & S_{87} & S_{88} \end{bmatrix} \begin{bmatrix} \epsilon_t^{OS} \\ \epsilon_t^{CD} \\ \epsilon_t^{OD} \\ \epsilon_t^{SP} \\ \epsilon_t^{OP} \\ \epsilon_t^{F_1} \\ \epsilon_t^{F_2} \\ \epsilon_t^{MP} \end{bmatrix}. \tag{33}$$

This identification scheme enables me to identify five oil-related shocks and monetary policy shock. ϵ_t^{OS} represents an oil supply shock, a massive cut in oil production. Oil supply shock is the first in the system, due to the assumption of slow adjustment of supply to demand shocks. This assumption is plausible because of high costs of production adjustment as argued by Kilian (2009) and because of shallow values of estimated short-run elasticity of oil supply estimated in Kilian and Murphy (2014). Flow demand shock ϵ_t^{CD} represents a rapid increase in demand for oil and other commodities as a consequence of economic growth. This shock was recognized as the most important determinant of oil prices, by Kilian (2009).

ϵ_t^{OD} represents an oil specific demand shock caused by rapid change in demand for inventories. According to the efficient market hypothesis, oil inventories would increase immediately after expectations about higher price in the future due to arbitrage. However, since world oil inventories include reserves held by government and refined products held by retailers, the relation of inventories to financial market is only limited and long-term applicable. Since oil inventories act as a balance between supply and demand, I ordered oil-specific demand shock on the third place to react as a buffer to supply shocks as argued in Jaffe and Soligo (2002). The speculative shock from financial market ϵ_t^{SP} is in this model defined as a rapid increase in the extended position over the short position of non-commercial traders indicating the beliefs of "bull" market. Speculative shock is ordered before oil price, both factors, and interest rate because it is assumed that monetary policy reacts contemporaneously to speculative pressure and also macroeconomy represented by factors reacts to speculative shocks within the same month.

ϵ_t^{OP} is a residual oil price shock which captures changes in oil price that cannot be explained by other variables in the model. Monetary policy shock ϵ_t^{MP} is ordered last because it is assumed that monetary policy reacts immediately after the shock to mitigate effects of shocks. Such ordering of monetary policy is adopted from Aastveit (2014), Bernanke, Boivin and Eliasz (2005). The identification scheme only restricts the contemporaneous effects. After the first period, all variables are allowed to react on all shocks.

4.3.3 Auxiliary Models

In this subsection, I shortly present two alternative oil market models with monetary policy. One model is a basic VAR model without factors and the second model is $\text{FAVAR}_{\text{PCA}}$, as proposed by Aastveit (2014) and Juvenal and Petrella (2015). All models have the same number of lags ($p = 5$) and the same ordering of variables. The purpose of comparison between $\text{FAVAR}_{\text{3PRF}}$ and alternative $\text{FAVAR}_{\text{PCA}}$ is the discovery of potential differences in results and to check whether results obtained from 3PRF factors are more sensible. Comparison of both FAVAR models to VAR model tests the importance of factor inclusion. Compared to $\text{FAVAR}_{\text{3PRF}}$ model, basic VAR is its restricted form, presented by equations (17) and (18). As noted in paragraph 5.3.2.1.2, such model suffers from significant informational insufficiency and should, therefore, perform much worse than both FAVAR models.

The relevant benchmark to $\text{FAVAR}_{\text{3PRF}}$ is FAVAR model that differs only in factor estimation step. This model incorporates principal components analysis to extract maximal amount of common variation from entire dataset **X**. There is, however, no guarantee that this information is relevant for analysis. Estimated factors are, therefore, not target oriented, and thus the model needs to be estimated only once after factors are obtained. The estimation procedure is composed of two steps, factor estimation and estimation of the model by using factors from the first step.

The FAVAR model is represented as:

$$C_t^{PCA} = A_0^{PCA} + \Phi(L)C_{t-1}^{PCA} + u_t^{PCA}, \tag{34}$$

where $\Phi(L)$ is a lag polynomial and $C_t^{PCA} = [\Delta prod_t, inv_t, rea_t, sp_t, rpo_t, \mathbf{F}_t^{PCA}, sr_t]'$.

The first factor is loaded to variables representing economic performance and the second factor is related to prices in the economy. My hypothesis is, therefore, that such model will perform well only in forecasting responses of oil-related variables and variables related to prices and economic performance. After estimation of equation (35), the structure (34) is imposed on residuals to obtain structural shocks.

5 DATA

Whereas ordinary VAR is limited to only few time series, the FAVAR model does not face those limitations. To cover as large spectrum of U.S. economy as possible, I use a dataset of 136 macroeconomic indicators, mostly obtained from a Economic Research Federal Reserve Bank of St. Louis (no date) prepared by McCracken and Ng (2016) (hereinafter: FRED-MD), which is produced specifically for econometric studies where big data is required. Additional data is obtained from the Institute of Supply Management (no date) (hereinafter: ISM), U.S. Commodity Futures Trading Commission (hereinafter: CFTC) and Energy Information Administration (hereinafter: EIA). The dataset contains 17 indicators of **output and income**, 30 indicators of **labour market**, 10 **housing** indicators, 19 indicators of **consumption, orders and inventories**, 14 indicators of **money and credit**, 21 indicators of **interest and exchange rates**, 20 indicators of **prices** and 5 indicators of **stock market**. The estimation period is on a monthly basis and spans from February 1986 to December 2016. All data series were transformed to ensure stationarity and tested with Dickey-Fuller test (Greene, 2012, p. 988-997). A detailed description of transformations and sources of variables is provided in the Appendix 3.

Additionally, I included five oil specific variables to capture dynamics specific to oil market and shadow federal funds rate, to describe the interaction of oil market with monetary policy.

The first oil specific variable is the **world crude oil production**. The variable was obtained from Energy Information Administration. It was further expressed as a deviation from the trend to ensure stationarity. Another oil specific variable is **real crude oil price**. It was calculated as average monthly WTI crude oil spot price deflated by U.S. consumer price index for all items.

Real economic activity index was chosen in a way to be a good representation of global demand for all commodities. For this index, I follow Kilian (2009) who constructed an index based on a global index of dry cargo single voyage freight rates. A good measure of global demand for industrial commodities is essential for identification of the demand side of the oil market.

There are multiple reasons for choosing Kilian's index of real economic activity over other representations of global economic activity. The first reason is in the availability of monthly value-added data for all countries. Monthly data is essential because delay restrictions are only plausible on a monthly basis. The second reason is the unavailability of proper weight of contribution to the global real economic activity for each country. As already mentioned, the third reason is that value-added is not the most appropriate measure of real economic activity for understanding industrial commodity markets. Since many countries faced the transition from manufacturing to service economy, increasing value added was not related to commodity market (Kilian, 2009, p. 1055-1056).

A measure of economic activity is based on Drewry Shipping Consultants Ltd. data of single voyage dry cargo freight rate. The original data series was not continuous for the entire sample period, and simple averages would ignore the existence of different fixed effects for different routes, commodities and ship sizes. Kilian eliminated fixed effects by computation of period to period growth rates for all time series in the data frame, and then he took the equal-weighted average of calculated growth rates and cumulated these growth rates, when January 1968 was normalized to unity. Finally, the index was deflated by U.S. CPI (Kilian, 2009, p. 1056-1058).

Fourth oil related variable is a proxy for **world crude oil inventories**, where I follow the methodology proposed by Hamilton (2009). Data on oil inventories is crucial for identification of the speculative component in crude oil price. Since the data for world crude oil inventories are scarce, I calculated the ration of the Organisation for Economic Co-operation and Development (hereinafter: OECD) petroleum stocks over U.S. petroleum stocks. The scale factor ranged from 2.23 to 2.61 in my sample. For the period where the data for OECD petroleum stocks were not available, I scaled U.S. petroleum stocks with an average factor to extend time series. With this methodology, strategic reserves are included in time series, and this is useful for identification of oil specific demand shock since a sudden drop in oil inventories needs to be replaced from the strategic point of view, and this creates an oil specific demand.

Another oil related indicator describes the **speculative pressure from financial part of the crude oil market**. This variable, therefore, captures only speculative pressure from financial markets and thus not included through above-ground inventories. In the construction of speculative pressure proxy, I follow Haase, Zimmermann and Zimmermann (2017) who propose the proxy calculated from short, long and spread position of non-commercial traders (SS_t, SL_t and SSP_t). As explained by U.S. Commodity Trading Commission, commercial traders use derivative contracts to hedge their position, whereas non-commercial traders do not have a physical position to hedge. Hence they are willing to accept the risk because they speculate about the future price. Non-commercial traders are therefore speculators in crude oil financial market. Speculative pressure proxy sp_t is according to Haase, Zimmermann and Zimmermann (2017) calculated as:

$$sp_t = \frac{(SL_t + SSP_t) - (SS_t + SSP_t)}{(SL_t + SSP_t) + (SS_t + SSP_t)} = \frac{SL_t - SS_t}{SL_t + SS_t + 2SSP_t}, \tag{35}$$

where SL_t represents speculative long open interest, SS_t is speculative short open interest and SSP_t is speculative spread open interest. A similar approach in identification of speculative shock, specifically form financial market was presented by Fueki et al. (2016). The main difference between their approach and approach of Haase, Zimmermann and

Zimmermann (2017) is that they use changes in net position, whereas Haase, Zimmermann and Zimmermann (2017) use normalized net position.

To model the interaction of oil market with monetary policy, it is useful to include monetary policy variable. The obvious choice of monetary policy measure is effective federal funds rate. However, since 2009, federal funds rate has been stuck at the zero lower bound, and this measure does no longer provide information regarding monetary policy response. A new policy rate was proposed by Wu and Xia (2016). They combined the federal funds rate before 2009 and the estimated shadow federal funds rate since 2009. The estimated federal funds rate is below zero for the period of the last financial crisis. The time series for shadow federal funds rate was obtained from Federal Reserve Bank of Atlanta (no date).

6 RESULTS

In this section, the results of the structural $\text{FAVAR}_{\text{3PRF}}$ model ($\text{SFAVAR}_{\text{3PRF}}$) are presented. Since the estimated parameters of $\text{FAVAR}_{\text{3PRF}}$ model do not have much economic interpretation, impulse responses of SFAVAR are presented along with variance decomposition of $\text{SFAVAR}_{\text{3PRF}}$ model and historical variance decomposition of $\text{SFAVAR}_{\text{3PRF}}$ model.

I perform the robustness check by estimation of some additional models. I extend the observed time span, after discarding speculative pressure proxy, to begin in February 1972. Estimated impulse responses are very similar, but it was impossible to identify the shock from the financial market and consequently this shock is hidden within other shocks. I also estimate several models with the different lag order. Estimated results after employing 2, 8 or 13 lags also do not exhibit changes in main conclusions. There are only minor differences in smoothness and convergence of impulse responses. I also try different orderings of variables, but only sensible ones. I therefore always order oil production first and shadow federal funds rate last. Additionally, I assume successive ordering of 3PRF factors. Variable permutations also do not cause drastic changes in results.

6.1 Structural Impulse Responses

After establishment of structural shocks w_t I can derive the response of each variable in $C_t = (C_{1,t}, ..., C_{K,t})'$ to an impulse in $w_t = (w_{1,t}, ..., w_{K,t})'$ scaled to one standard deviation of variable to which the shock is related. Response of C_t to a structural shock in w_t is defined as

$$\frac{\partial C_{t+i}}{\partial w_t'} = \Theta_i, \quad i = 0, 1, ...H \tag{36}$$

where Θ_i is a matrix of responses of C_t to a shock in w_t for $i = 0, 1, ..., H$. Θ is thus a $K \times K$ matrix of elements

$$\theta_{jk,i} = \frac{\partial C_{j,t+i}}{\partial w_{k,t}}. \tag{37}$$

By assuming covariance stationarity, C_t can in principle be expressed by a combination of current and past shocks weighted by Ω_i with values decreasing as $t \to 0$. Moving average representation of C_t is

$$C_t = \sum_{i=0}^{\infty} \Omega_i u_{t-i} = \sum_{i=0}^{\infty} \Omega_i B_0^{-1} B_0 u_{t-i} = \sum_{i=0}^{\infty} \Theta_i w_{t-i}, \tag{38}$$

where $\Theta_i \equiv \Omega_i B_0^{-1}$.

Under the assumption of stationarity it also holds that

$$\frac{\partial C_t}{\partial w'_{t-i}} = \frac{\partial C_{t+i}}{\partial w'_t} = \Theta_i, \tag{39}$$

and all later responses are then computed by multiplication of Ω_i with B_0^{-1}. The impulse response is initiated as $\Theta_0 = \Omega_0 B_0^{-1} = I_K B_0^{-1} = B_0^{-1}$ and then it continues as $\Theta_1 = \Omega_1 B_0^{-1}$, $\Theta_2 = \Omega_2 B_0^{-1}$ for $i = 0, 1, ...48$ (Kilian & Lütkepohl, 2017, p.108-111).

As already noted, an essential advantage of FAVAR$_{\text{3PRF}}$ over FAVAR$_{\text{PCA}}$ is a better forecast of the response of variables in dataset **X**. These variables are indirectly included in the model through factors and their response to shocks could thus not be modeled within FAVAR model. However, the estimated response of variable x_m can be obtained through the forecast of x_m by the model (15) and autoregressive model of x_m. The forecast model is a multivariate linear regression of the following form:

$$x_{m,t} = A_0 + \boldsymbol{\beta_1} C'_t + \boldsymbol{\beta_2} C'_{t-1} + ... + \boldsymbol{\beta_6} C'_{t-5} + \gamma_1 x_{m,t-1} + ... + \gamma_5 x_{m,t-5} + \zeta_t. \tag{40}$$

Each variable x_m is thus allowed to react contemporaneously to each shock without additional restrictions (Aastveit, 2014, p. 12).

After estimation of this model on the whole available data set, the response of x_m is calculated for H periods ahead with H iterations. Basically, the potential values of x_m are estimated, and responses of C_t are treated as inputs. In the first step, instantaneous response of x_m to a shock is estimated. Instantaneous response $x_{m,0}$ depends only on responses of variables in vector C_t to a shock. In the second period, the estimated instantaneous response enters equation (40) as a lag of x_m. In this way, the response of x_m is iteratively computed 48 months ahead.

Statistical inference of impulse responses is based on bootstrap techniques since the confidence intervals obtained in this way are more reliable and accurate for small samples. Bootstrap methods are also more general than the methods based on a standard asymptotic theory which assumes FAVAR errors being identically and independently distributed and Gaussian (Lütkepohl & Krätzig, 2004, p. 177). The bootstrap method that I considered is a residual based recursive design method. According to this method, the only assumption regarding residuals is that $E(u_t) = 0$ and that u_t has finite moments of an appropriate order. The parametric family of distribution of residuals is unknown, and its estimation is unnecessary, since bootstrapping residuals u_t^* can be drawn with replacement form the set of residuals of estimated model $\{\hat{u}_t\}_{t=1}^T$. Kilian (1998) compared parametric and non-parametric bootstrap, and he found out that there is no significant advantage of assuming parameters of distributions, even if the assumption is true. Wrong assumptions can, on the other hand, lead to the undermined accuracy of bootstrap inference (Kilian & Lütkepohl, 2017, p. 336).

The procedure that I considered has the following mechanics. I retain residuals from estimated $\text{FAVAR}_{\text{3PRF}}$ model (15) $\{\hat{u}_t\}_{t=1}^T$. For each bootstrap replication $r = 1, ..., \rho$ I randomly draw elements from $\{\hat{u}_t\}_{t=1}^T$ with replacement. This procedure ensures that resulting ρ series $\{u_t^*\}_{t=1}^T$ are independent and identically distributed. From obtained random samples u_t^*, the sequences $\{C_t^*\}_{t=-p+1}^T$ are generated. From this step, I proceed the same as before by fitting the model (15) to estimate the parameters and variance-covariance matrix. Then I impose the structure to the model and obtain ρ implied impulse responses $\hat{\theta}_{jk,i}^{*r}$. After bootstrapped approximation of the distribution of structural impulse response $\hat{\theta}_{jk,i}$, the percentile confidence interval is constructed according to Runkle (1987) who proposed the percentile interval

$$\left[\hat{\theta}_{jk,i,\gamma/2}^{*r}, \hat{\theta}_{jk,i,1-\gamma/2}^{*r}\right], \tag{41}$$

where $\hat{\theta}_{jk,i,\gamma/2}^{*r}, \hat{\theta}_{jk,i,1-\gamma/2}^{*r}$ are critical points defined by the $\gamma/2$ and $1 - \gamma/2$ quantiles of distribution of $\hat{\theta}_{jk,i}^{*r}$ (Kilian & Lütkepohl, 2017, p. 337-339).

It is worth noting that factors are not reestimated in this bootstrap procedure, and this could lead to underestimation of error. However, Yamamoto (2016) noted, that in cases, when N is sufficiently large, the factor estimation uncertainty is negligible.

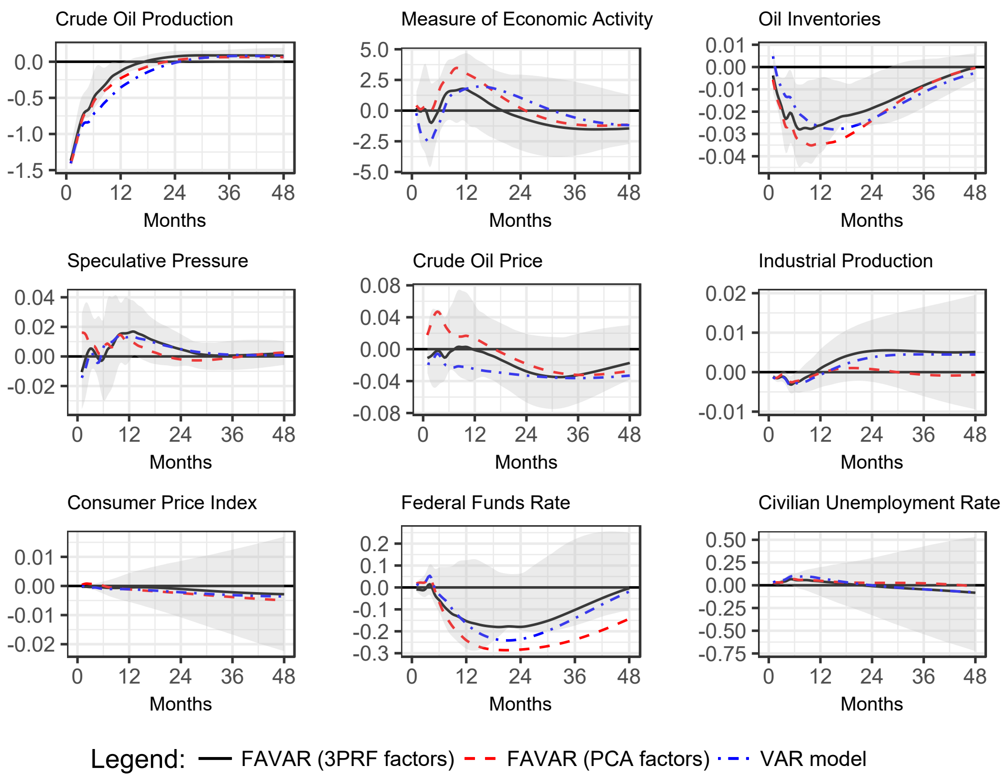
Bellow, I present the impulse responses to oil shocks and monetary policy shock, estimated by $\text{FAVAR}_{\text{3PRF}}$, $\text{FAVAR}_{\text{PCA}}$ and baseline VAR model. The impulse responses were estimated for 21 variables of interest to observe the effects of shocks on the U.S. economy.

6.1.1 Oil Supply Shock

The responses to one standard deviation oil supply shock are presented in Figures 3 and 4. Unexpected oil supply shock causes a significant decrease in crude oil production for approximately 1.5 million barrels per day, and it takes two years for the production to return to the quantities before the shock. This adjustment is fast in comparison to other authors who found out a very slow adjustment to supply shocks, Aastveit (2014) confirmed only partial convergence after four years, Kilian and Murphy (2014) found out no adjustment after 15 months. Economic activity decreases insignificantly on impact, increases after that and returns to negative territory after 12 months. Oil inventories start to decline on impact to mitigate the effect of limited supply and decreases as long as crude oil production returns to the level before the shock. Because of oil inventories adjustment, oil price does not respond significantly. Hence U.S. CPI also remains untouched. The estimated change of crude oil price by a $\text{FAVAR}_{\text{PCA}}$ is positive on impact, whereas other two models estimate immediate negative oil price response. A negative response is logical since oil inventories fully offset the flow market gap which is a consequence of low production.

It seems that non-commercial traders do not respond to oil supply shock since the response is insignificant through the entire observed period. The only possible indication of speculative pressure occurs one year after the shock as a response to oil inventories adjustment. Strong expansionary monetary policy response, though insignificant, is surprising, but is estimated with all three models. Monetary policy reaction can be explained by response to the decreasing industrial production and increased unemployment since the federal funds rate responses approximately in the period when also industrial production fell. The strong expansionary reaction of monetary policy to oil supply shock may also elevate oil production due to lower costs of financing. Similar results were proposed by Aastveit (2014) who also found an adverse effect of high interest rate on oil production. Bernanke, Gertler, Watson, Sims and Friedman (1997) similarly found a positive effect of high interest rate on real oil price measure. Expansionary monetary policy response is also advocated by Bodenstein, Guerrieri and Kilian (2012) who stress out, that monetary policy should react to falling industrial production and increasing unemployment, to pursue the optimal response from the social point of view.

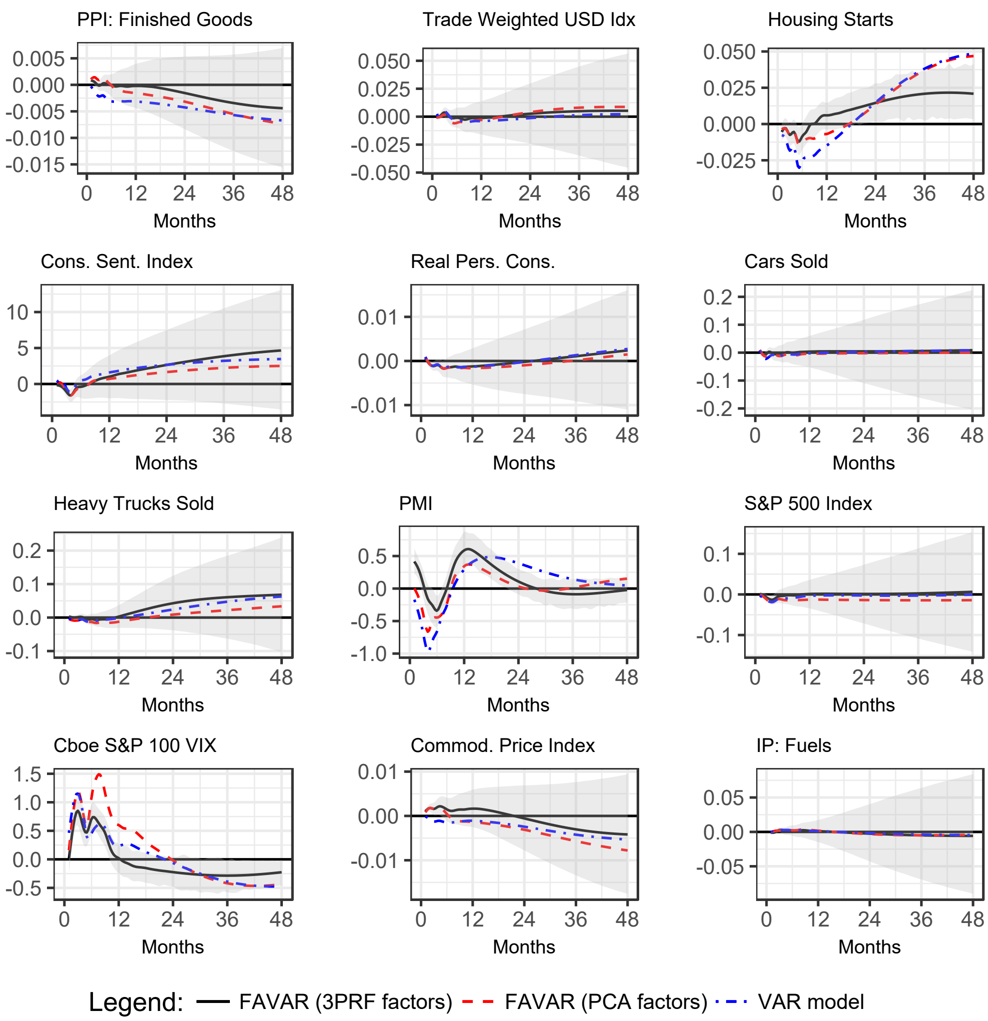
Figure 3: Responses of Main Variables to Oil Supply Shock



Source: Own work.

Surprisingly, U.S. industrial production responses differently than the measure of world economic activity. The reason is probably the transition of oil demand from OECD countries to Asia as also noted by Aastveit, Bjørnland, and Thorsrud (2015) and thus an improvement in U.S. industrial production does not mean improvement in crude oil commodity demand. The negative response of economic activity measure could also be a direct consequence of negative response of oil price since economic activity measure proposed by Kilian (2009) is based on freight rates, which also depend on oil prices. Lower oil prices decrease costs of freight and hence the freight rate decreases. Unemployment rate slightly increases as estimated by FAVAR$_{3PRF}$, however, the effect is statistically insignificant.

Figure 4: Responses of Additional Variables to Oil Supply Shock



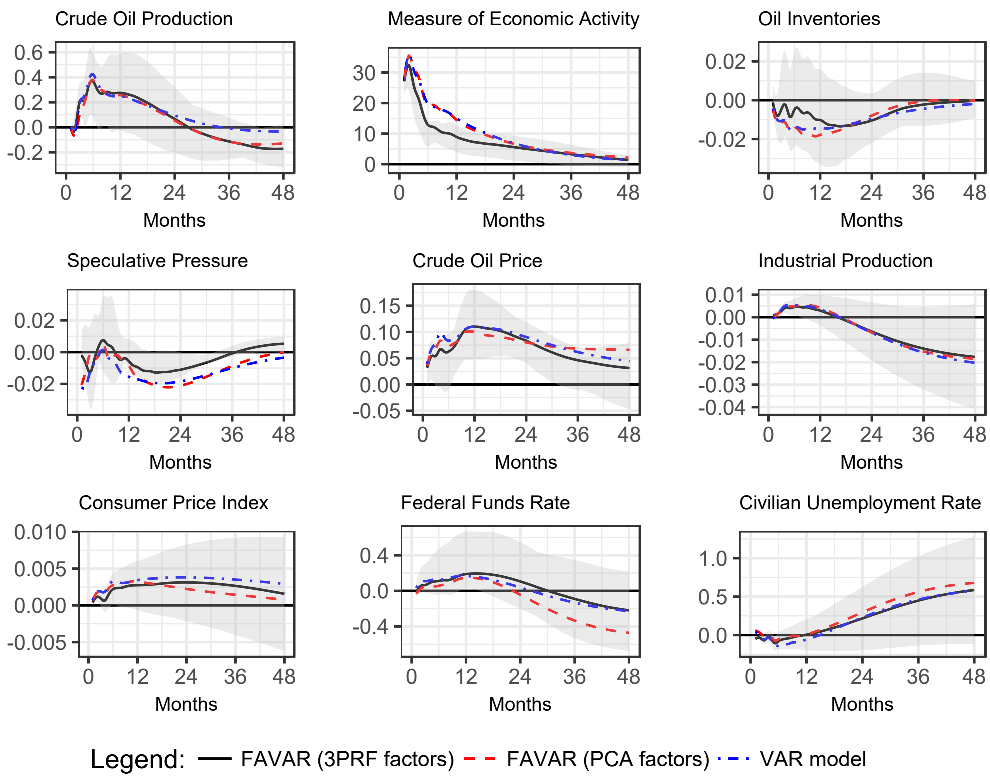
Source: Own work.

The effects of oil supply shock on additional macroeconomic variables are presented in Figure 4. Even though most responses are statistically insignificant, results generally indicate favorable U.S. economic conditions. Consumer sentiment index slowly increases six months after the shock just like housing starts and sale of heavy trucks. The response of stock market is negligible. Purchasing Managers' Index significantly increases on impact, then it quickly drops to negative territory and then turns back to definite indicating economic improvement. All three models indicate increased volatility of stock prices on impact and volatility remains elevated for almost a year. The estimated responses of all models are very similar, except the response of housing starts and volatility index, where $FAVAR_{3PRF}$ estimates more modest response since it accounts only for relevant information. Stock market response to oil supply shock is insignificant. In contrast, Cunado and de Gracia (2014) suggest that supply shocks have the most substantial effect on stock prices. They, however, estimated the effects of oil shocks on European countries and because Europe imports more significant share of consumed oil then the U.S., the impact of oil shocks on stock prices could be more significant.

6.1.2 Shock in Global Aggregate Demand for Commodities

The second shock that I consider is a standard deviation shock in global demand for industrial commodities. Flow demand slowly converges to the level before the shock, which is achieved after four years. Due to increasing demand, oil inventories insignificantly decreases on impact by 15 million barrels in 1.5 years. As after oil supply shock, also in case of a demand shock, inventories decrease and mitigate the effect on oil price. Real crude oil price increases on impact by 5 % and the impact builds up in 12 months to almost 15 % price increase. After that, the price slowly approaches the initial level.

Figure 5: Responses of Main Variables to Shock in Global Commodity Demand

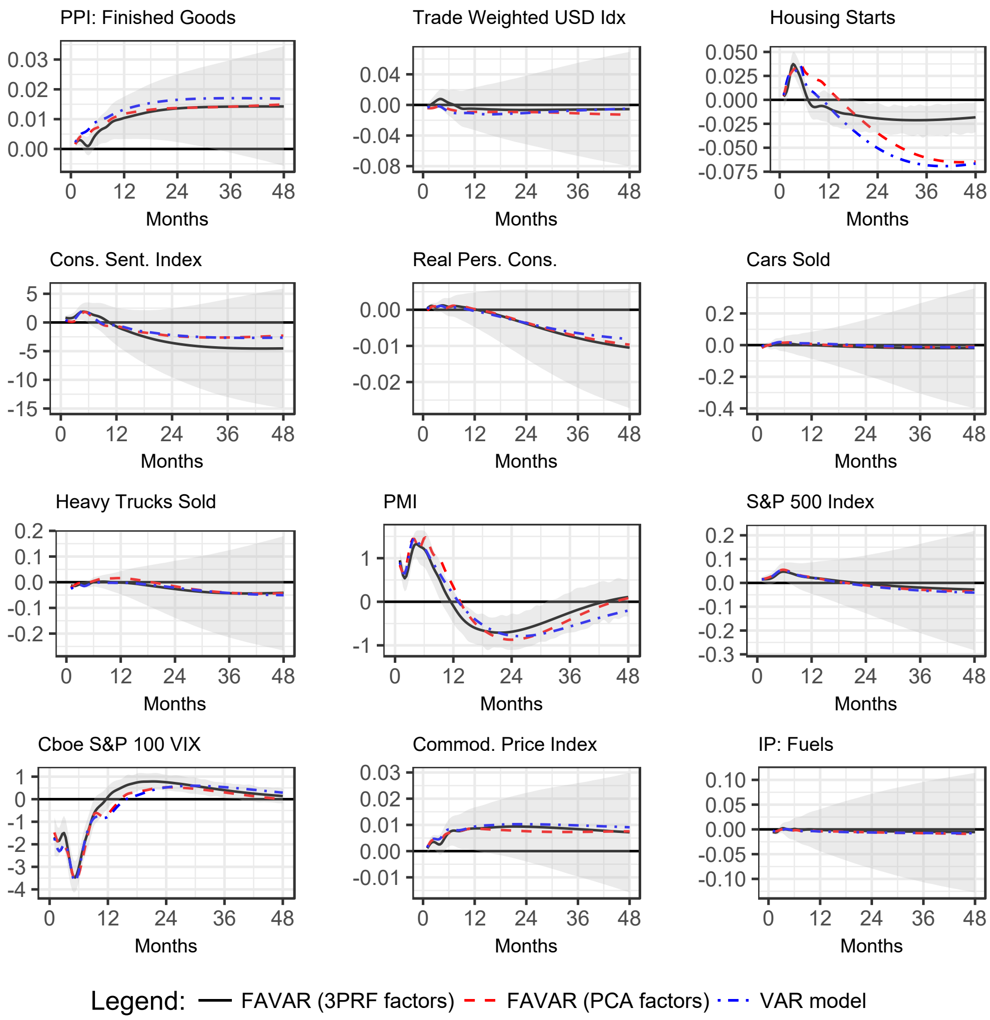


Source: Own work.

Monetary policy response is restrictive but insignificant. Its effect is however seen in consumer price index which remains almost unaffected even though crude oil price increases. The estimated response of monetary policy is consistent with estimations of Aastveit (2014). Crude oil production increases in a few months after the demand shock as a response to elevated prices.

Macroeconomic conditions in the U.S. worsen in one year after the shock, since industrial production decreases and unemployment rate increases, with both effects being insignificant. Speculators on the financial market do not react initially, but after 1.5 years the short position and anticipation of "bear" market prevail. Crude oil price may therefore also respond to speculative pressure and not only to increased supply.

Figure 6: Responses of Additional Variables to Shock in Global Commodity Demand



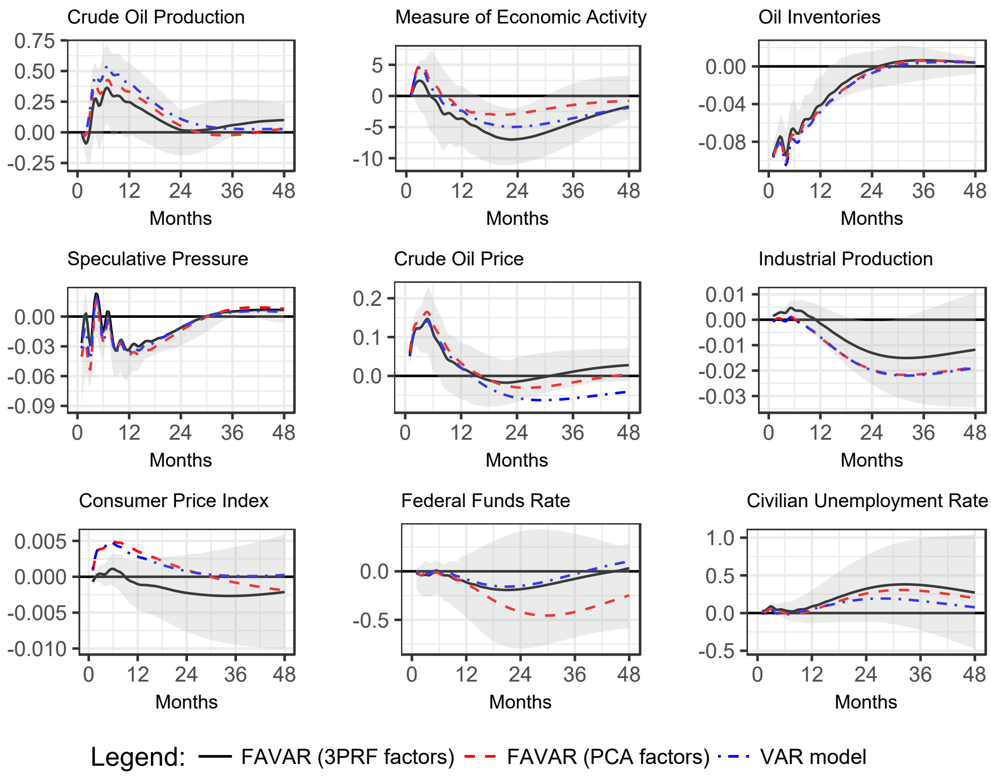
Source: Own work.

The responses of other economic variables to a shock in commodity demand in general exhibit temporary improvement in the U.S. economy followed by unfavorable macroeconomic conditions. All three models propose that stock prices increase through the first year as indicated by the response of S&P 500 index, in next three years the stock market remains in declining state. Initial improvement in stock prices is logical, because of stock market responses to improved world industrial conditions. The volatility of stock prices decreases on impact and returns to positive territory after one year. All three models estimate the negative long-run response of housing starts, but the response estimated by $\text{FAVAR}_{3\text{PRF}}$ is much less severe. The estimated response of consumer sentiment index by $\text{FAVAR}_{3\text{PRF}}$ is, on the other hand, stronger than the response of other two models, yet all responses are statistically insignificant.

6.1.3 Oil Specific Demand Shock

Oil specific demand shock represents a rapid increase in the demand for oil inventories. This demand shift may be a reflection of expectations of a higher price in the future, therefore speculations in the physical market. As argued by Kilian (2009) inventories increase could also be induced by a precautionary component of oil demand. The testing environment of oil specific demand shock requires the assumption of a negative shock to the level of oil inventories.

Figure 7: Responses of Main Variables to Oil Specific Demand Shock



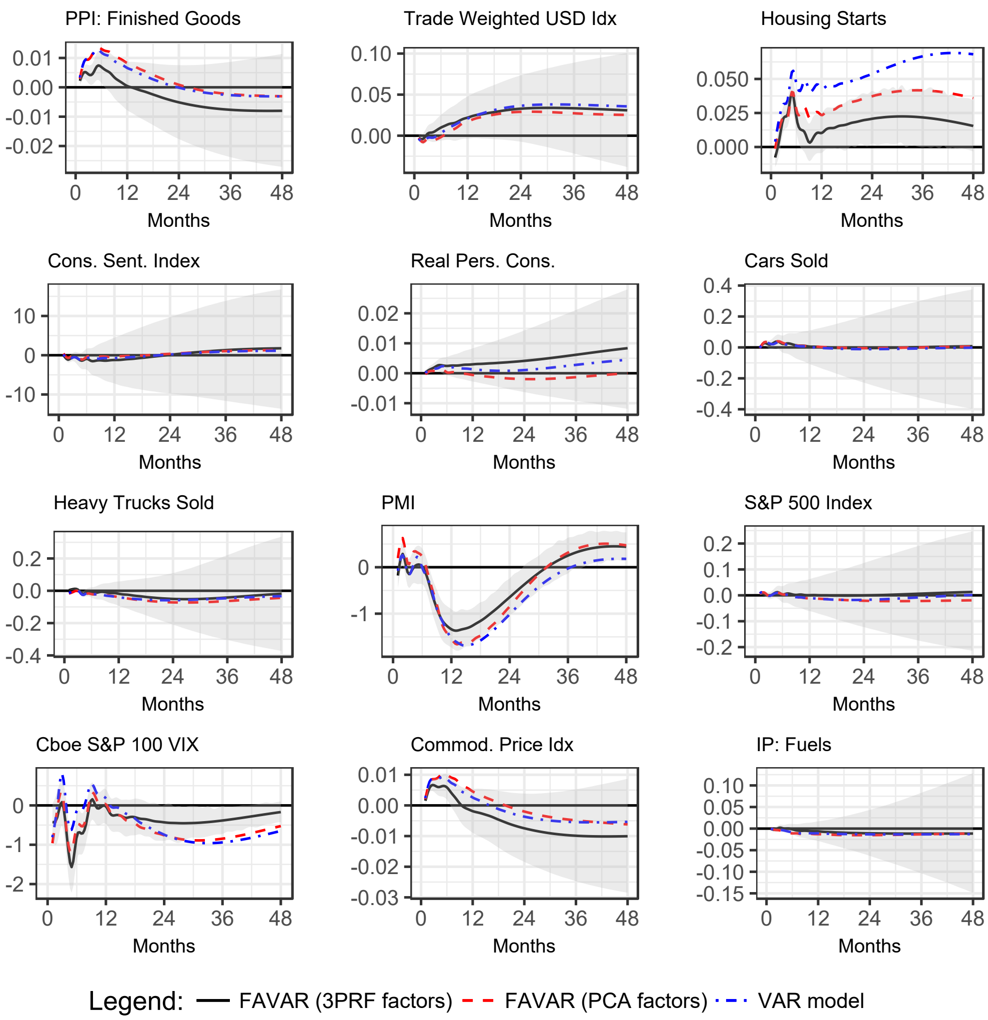
Source: Own work.

Since other oil shocks do not justify the drop in oil inventories, inventories level has to be adjusted to the pre-shock value which represents the oil specific demand. In other words, if the desired level of inventories is held initially, inventories will increase after a negative shock to inventory stock to close the gap between actual and desired stock of oil inventories.

World oil inventories decrease by approximately 100 million barrels on impact and return to the value before the shock in approximately two years.

Effect on oil price is estimated to be positive, as expected. Oil price increases on impact by 5 % and by additional 10 percentage points in the next six months, and then slowly converges to a pre-shock value which is achieved after one year. Consumer prices significantly increase, due to pressure on production costs. Interestingly, in contrast to Aastveit (2014) and Juvenal and Petrella (2015), world economic activity measure and U.S. industrial production both increase in first few months after the shock and then become negative. The estimated decrease of U.S. industrial production by $\text{FAVAR}_{3\text{PRF}}$ is much smaller than the estimates of other two models, whereas the negative response of global economic activity is the highest.

Figure 8: Responses of Additional Variables to Oil Specific Demand Shock



Source: Own work.

Crude oil production increases soon after the shock and reaches the peak increase of 0.3 million barrels per day after six months. To only fill the inventories deficit, approximately one year would be enough, but increased economic activity slows down the process of inventories adjustment. Increasing oil inventories do not depend much on speculative activity as it was proposed by Kilian and Murphy (2014), Kilian and Lee (2014) and Juvenal and Petrella (2015) who interpreted increasing oil inventories solely as a speculative shock in the oil market. If inventories were a good representation of speculative pressure, there would be a significant delay in response to oil production since oil producers would anticipate higher prices in the future. On the financial part of the oil market, non-commercial traders mostly speculate on higher price by taking a long position, and after approximately three quarters they speculate on price decrease, which may also be the reason for the decreasing oil price. Monetary policy reaction estimated by $FAVAR_{3PRF}$ and SVAR is smaller and insignificant, whereas the response estimated by $FAVAR_{PCA}$ proposes strong expansionary monetary policy by estimated 500 basis points decrease in the federal funds rate.

In Figure 8 the impulse responses of additional variables are presented. The implied volatility of stock prices exhibits a decidedly non-stationary pattern, but in general, volatility responds negatively. Purchase managers' index does not react on impact and drops significantly after 6 months. It stabilizes at the level before the shock in two years. Housing starts to react positively to an oil-specific demand shock, and the response remains in the positive territory even after four years and this finding is surprising since economic conditions in the U.S. worsen after oil-specific demand shock. However, in the long term, $FAVAR_{3PRF}$ becomes insignificant, and the response is also smaller than responses of other models, and this may be the privilege of using only relevant information. Both commodity price index and producer price index respond similarly to the response of consumer price index, whereas other variables do not respond significantly to an oil-specific demand shock.

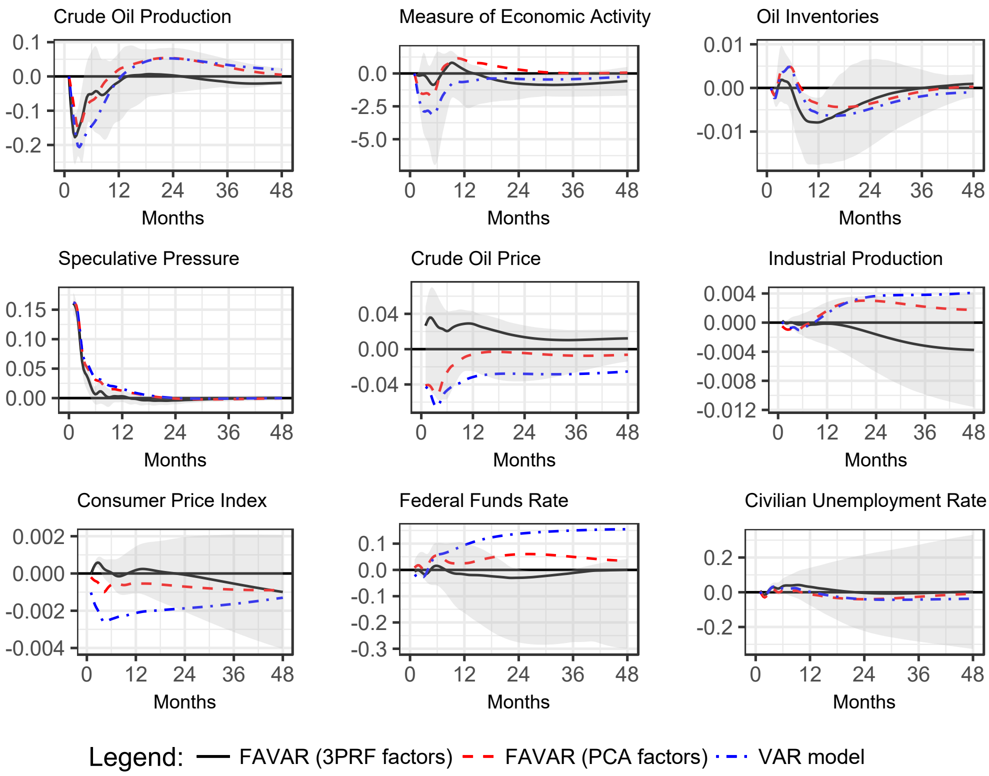
6.1.4 Speculative Shock in Financial Part of Oil Market

The speculative shock from the financial part of oil market represents an increase of long open interest over the short open interest of non-commercial traders. In other words, it is assumed that speculators expect a ”bull” market.

As presented in Figure 9, the response of speculative pressure on its shock converges to zero in only six months, as expected, since the gap is quickly closed in the financial market and the effects of shocks on financial market are relevant only on the short run. Crude oil production reacts negatively and converges to pre-shock value in 12 months. The negative response of oil production was also discovered by Juvenal and Petrella (2015). This phenomenon was also noted by Hamilton (2009), who claims that speculative shock affects oil market also through supply channel since producers are willing to hold underground inventories as they also expect the price hike.

Oil inventories respond insignificantly and negatively after six months. Some studies, like Kilian and Murphy (2014), Juvenal and Petrella (2015), use sign restrictions identification

Figure 9: Responses of Main Variables to Speculative Shock from Financial Market



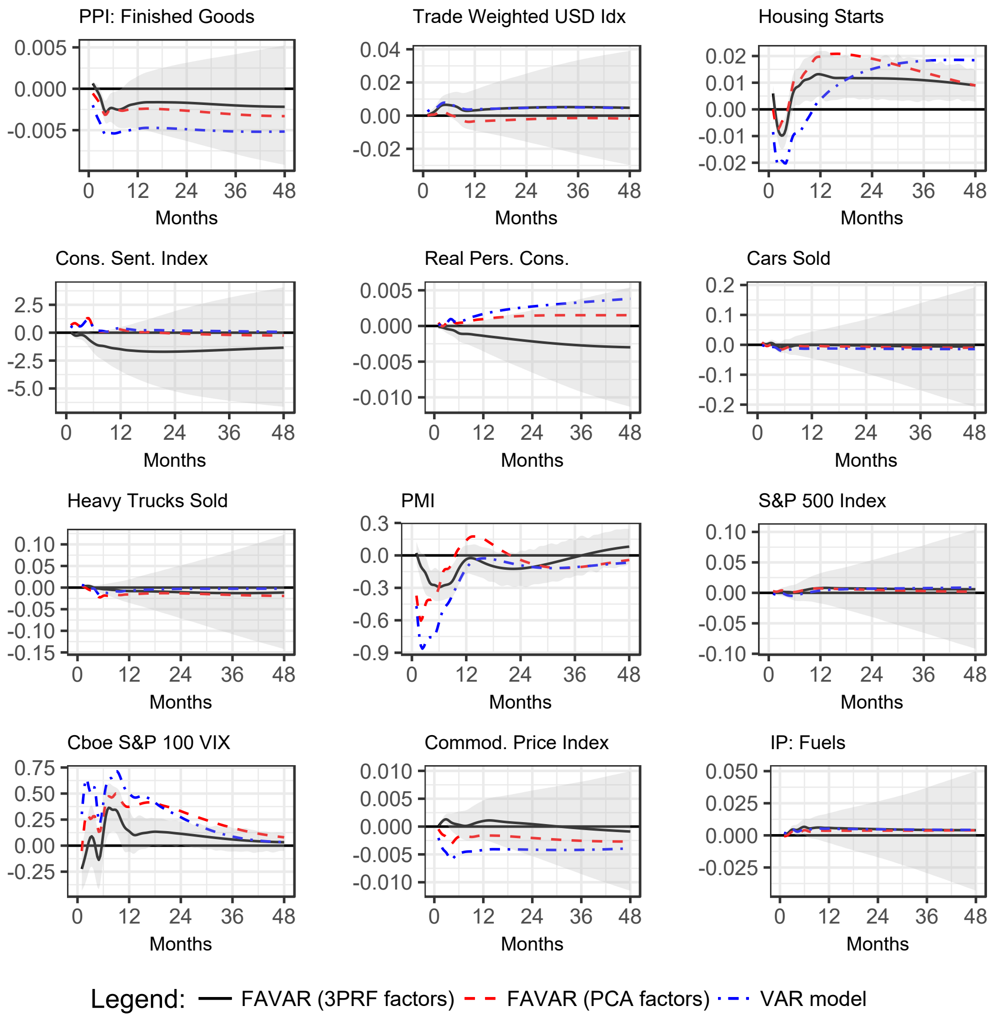
Source: Own work.

scheme, where they assume positive response of oil inventories. Even though the response is positive initially, it is small and quickly turns to the negative territory. Results, therefore, confirm my assumption that oil reserves are not held primarily for speculative reasons, but mostly to mitigate oil shocks.

Serious informational deficiency or use of irrelevant information is observed in the response of crude oil price. Expectations about the price increase, should together with lower production lead to a higher price. This is not confirmed by $\text{FAVAR}_{\text{PCA}}$ and SVAR which exhibit a price drop. The response of U.S. industrial production is also uncommon, since both, $\text{FAVAR}_{\text{PCA}}$ and VAR, estimated increased industrial production, where the response should be negative. The last irregularity is the strong restrictive response of monetary policy, estimated especially with VAR model. Such policy response would further depress the economy and is therefore unexpected.

Responses of additional variables of interest are presented in Figure 10. The response of producer price index is insignificant and negligible. Substantial differences between models are observed in impulse responses of consumer sentiment index and real personal consumption, where $\text{FAVAR}_{\text{3PRF}}$ results are more consistent with the theory. The number of housing starts decreases on impact and turns to positive 1 % increase in housing starts after six months. The purchase manager's index responds negatively to speculative shock as

Figure 10: Responses of Additional Variables to Speculative Shock from Financial Market



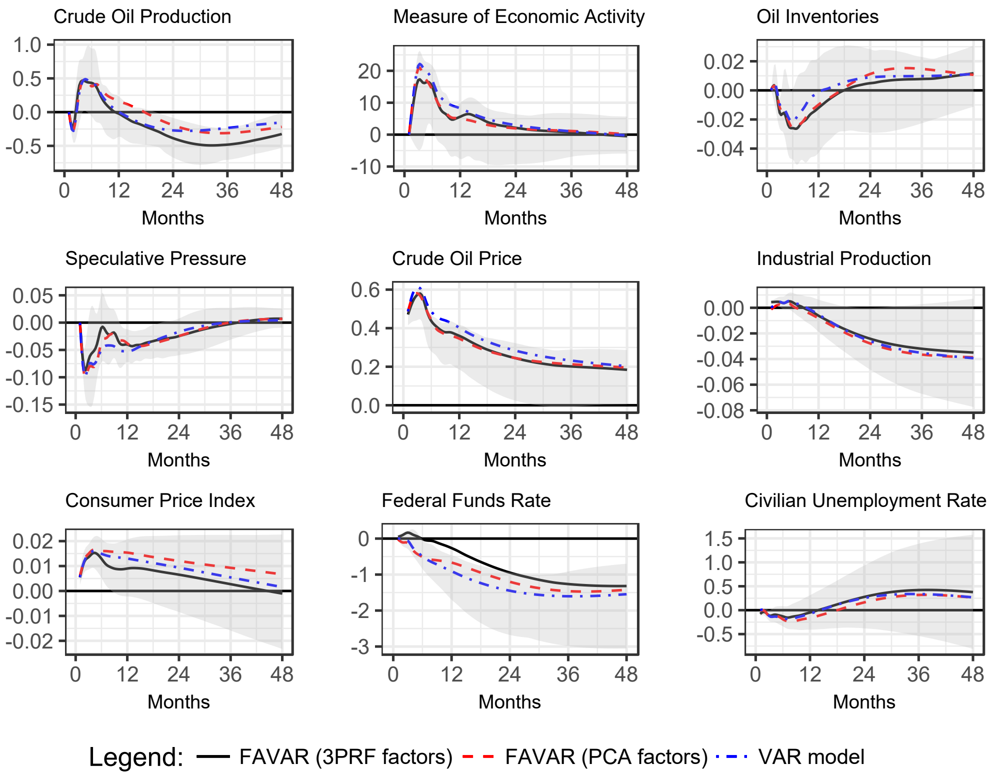
Source: Own work.

expected. The volatility of S&P 100 index significantly increases after six months, roughly in the period when speculative pressure already fades out. Increasing volatility, however, does not affect S&P 500 stock price index. In contrast to global oil production, U.S. industrial production of fuels increases on impact, but the effect is statistically insignificant. It is interesting that all models found an insignificant but negative response in the car and heavy trucks sale, which confirms worsened macroeconomic conditions in the U.S., proposed by $\text{FAVAR}_{\text{3PRF}}$.

6.1.5 Residual Oil Price Shock

The last oil shock that I consider is the residual oil price shock, which represents the sudden oil price increase which cannot be explained neither by speculative pressure, supply disruptions nor demand shocks. Residual oil price shock is a consequence of weather shocks, changes in inventory technology or imperfect measures of global economic activity

Figure 11: Responses of Main Variables to Residual Oil Price Shock

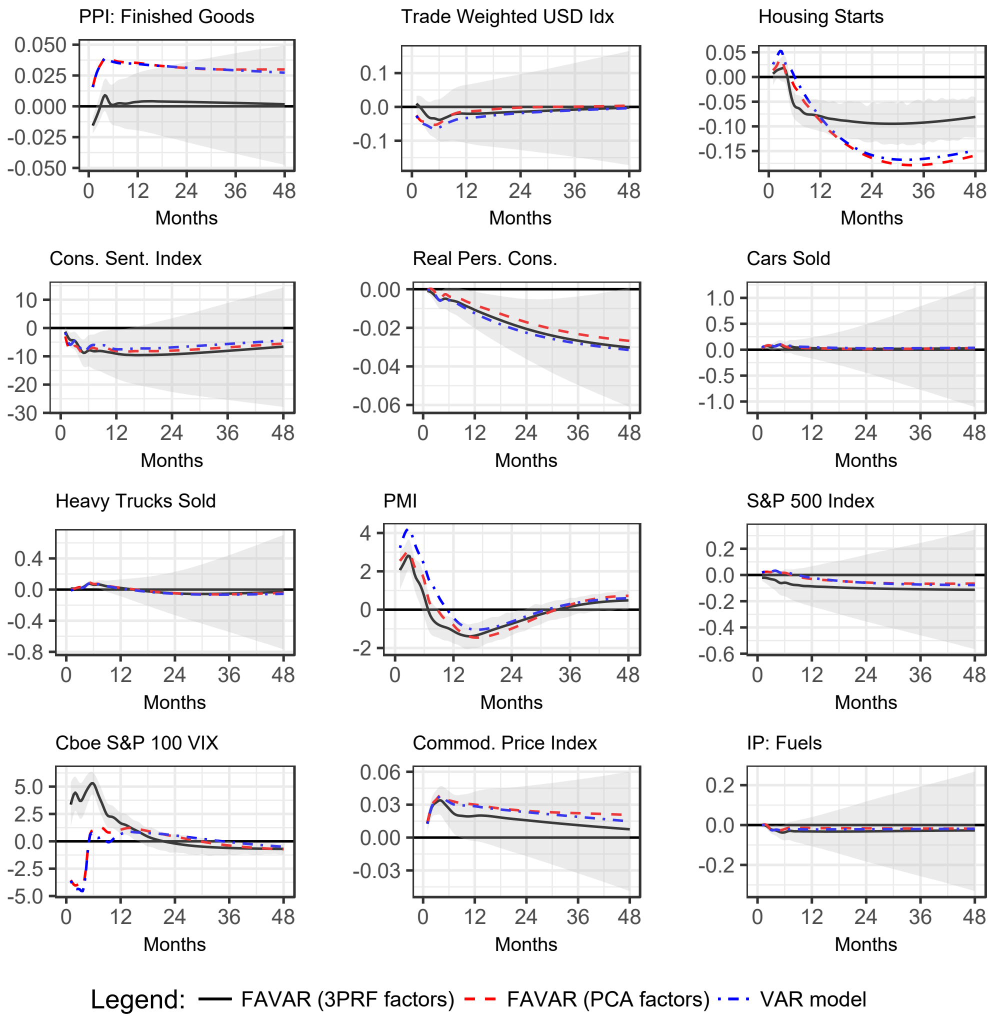


Source: Own work.

and speculation pressure as a consequence of limited information. Residual oil price shock could also be caused by personal consumption in countries outside the United States, since this shock is not included in the model. However, estimated impulse responses are very similar to those after oil-specific demand shock or flow demand shock which leads me to the conclusion, that the residual shock probably represents the non-industrial consumption of consumers outside the U.S.

Impulse responses after the unexpected one standard deviation oil price shock are depicted in Figure 11. Unexpected oil price hike causes the net short position of non-commercial traders since they expect that current price is unsustainable and will decrease soon. Oil producers react by increasing production due to expectations of lower future price. Oil inventories decrease on impact and slowly return to the pre-shock value in one year. Proxy for economic activity significantly increases after the shock. This phenomenon is explained by the construction of proxy since freight rates are directly linked to oil prices which represent an essential part of freight's price. U.S. monetary policy responses are entirely different than in the case of other demand shocks. It seems that in case of residual oil demand shock Federal Reserve concentrates on economic growth and lets oil market forces to decrease crude oil price. Moreover, low interest rates could also stimulate oil production through lower costs of financing.

Figure 12: Responses of Additional Variables to Residual Oil Price Shock



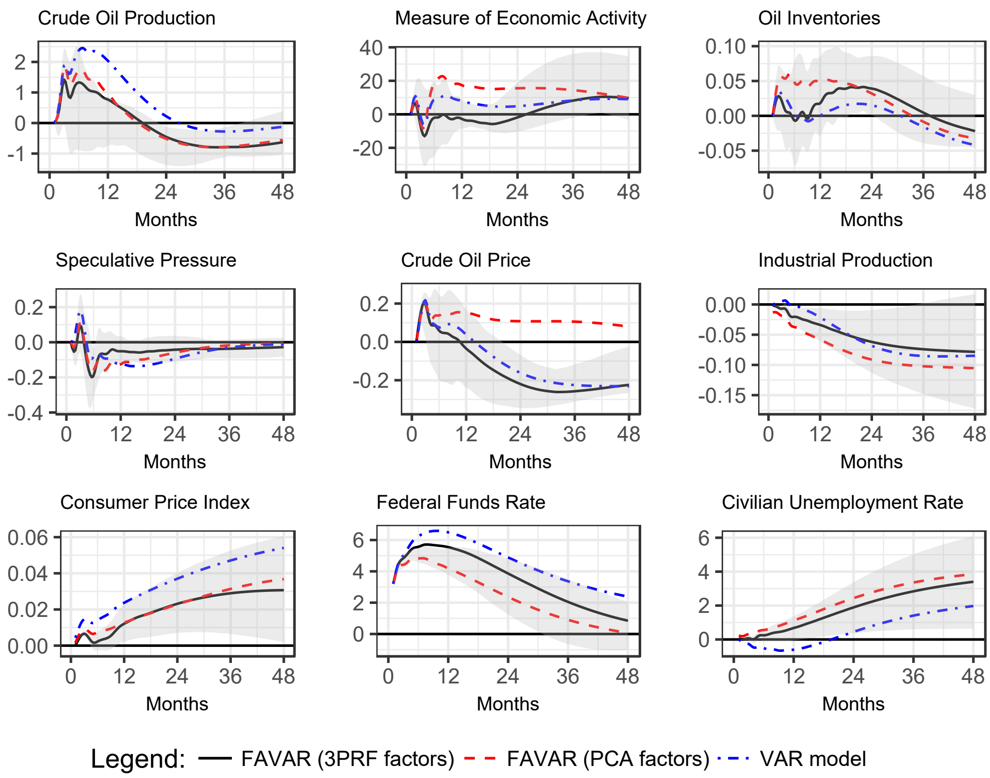
Source: Own work.

The responses of other U.S. macroeconomic variables of interest, presented in Figure 12 are different from the responses to other two demand shocks since the responses are much worse for U.S. economy. $\text{FAVAR}_{\text{3PRF}}$ estimates a significant and robust increase of stock price volatility index and insignificant negative response of stock prices. The comparison of models once again reflects the informational superiority of $\text{FAVAR}_{\text{3PRF}}$ over other models because volatility index is strongly related only to relevant variables, thus labor market and income. All models on the other hand estimate negative response of consumer sentiment index, real personal consumption and housing starts.

6.1.6 Monetary Policy Shock

Even though, this master thesis is about the effects of oil shocks on U.S. economy in interaction with monetary policy, I additionally consider the effects of monetary policy shock. I am especially interested in responses, if any, of oil market variables. Another reason for the exploration of responses to monetary policy shock was the comparison of results with other similar FAVAR models like Bernanke, Boivin and Eliasz (2005), Boivin, Giannoni and Mihov (2009) and Aastveit (2014), who also considered monetary policy within the oil market environment.

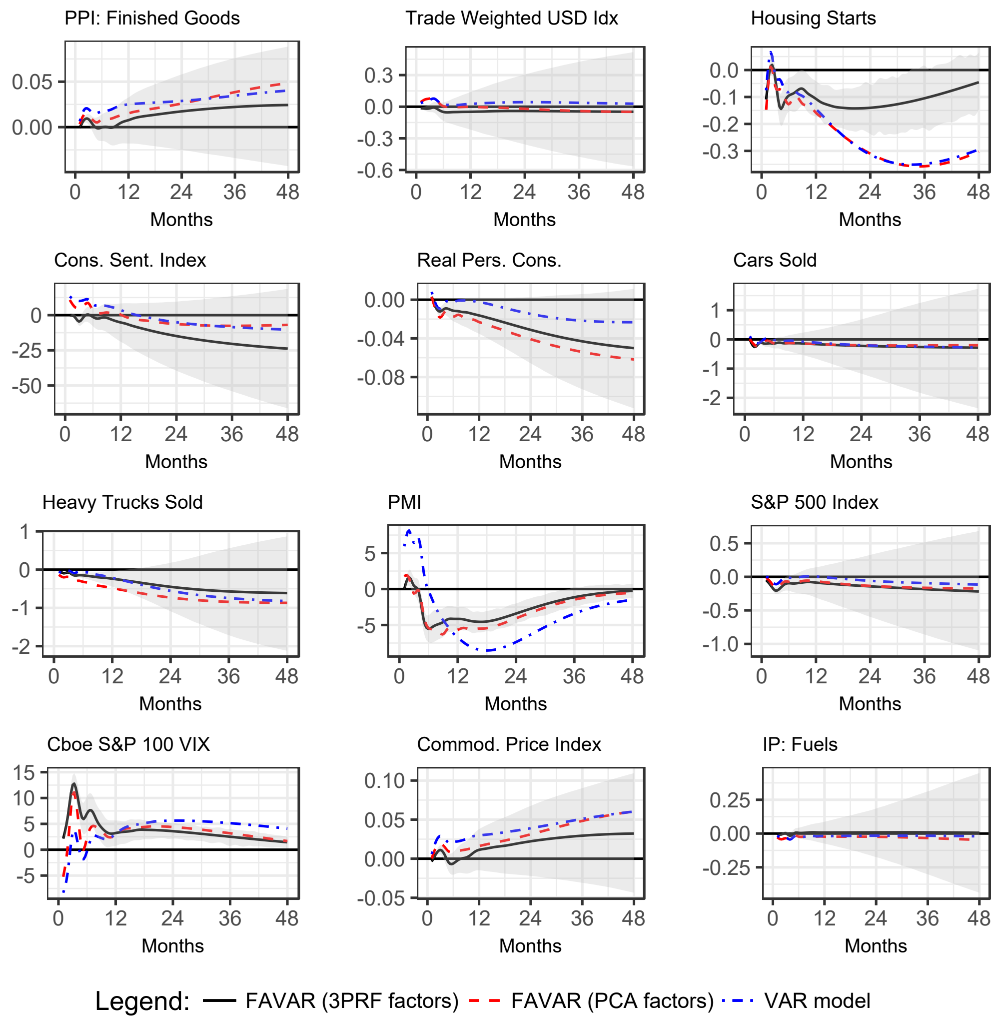
Figure 13: Responses of Main Variables to Monetary Policy Shock



Source: Own work.

In the Figure 13, the responses of main variables to monetary policy shock are presented. A 250 basis points increase in shadow federal funds rate, significantly affects the U.S. macroeconomic variables. Federal funds rate reacts positively to own 250 basis point shock, reaching 500 basis points interest rate hike in roughly 10 months and converges to the pre-shock value in 4 years. The responses of industrial production and civilian unemployment rate are in line with expectations, whereas the response of consumer price index exhibits an unconventional response, which has already been broadly discussed and is commonly referred to as the "price puzzle". Prices should have fallen after a restrictive monetary policy shock because of decreased demand. However, structural FAVAR and VAR models estimate a positive response. However, $FAVAR_{3PRF}$ estimates the lower positive effect on consumer price index and this is a small improvement, even though the positive response is statistically significant. The FAVAR model, initially proposed by Bernanke, Boivin and Eliasz (2005) was meant just for dealing with this implausibility of classic VAR models. However, $FAVAR_{3PRF}$ and $FAVAR_{PCA}$ both estimate positive but lower positive response in consumer price index, whereas VAR model shows a stronger positive response. Described responses of consumer price index and unemployment rate strongly support the hypothesis of informational insufficiency of basic VAR model and suggests the use of FAVAR models.

Figure 14: Responses of Additional Variables to Monetary Policy Shock



Source: Own work.

The responses of oil market variables are very interesting. Crude oil production increases by 1.5 million barrels per day in three months after the shock and decreases reaching the bottom at 1 million barrels per day lower production after three years. The response differs from zero significantly only at longer horizons. $\text{FAVAR}_{\text{PCA}}$ estimates the substantial positive impact on oil inventories, whereas $\text{FAVAR}_{\text{3PRF}}$ and VAR estimate a very mild response. Crude oil price increases on impact because of initially positive oil demand, but decreases after three months and becomes statistically significant and negative approximately two years after the shock. $\text{FAVAR}_{\text{PCA}}$ model on the other hand estimates a positive response of oil price through the entire period, which is not consistent with economic theory and other empirical studies, because restrictive monetary policy should have a negative impact on commodity prices, as proposed by Hamilton (2009), Barsky and Kilian (2001, 2004) and Frankel (2008). After six months, the speculators respond by net short position which also drives down the oil price.

Figure 14 shows the responses of U.S. macroeconomic variables to a monetary policy shock. The producer price index and commodity price index are both subject to the "price puzzle", but FAVAR$_{3PRF}$ estimates the lowest price increase and therefore performs better. The effects on other variables are consistent with the theory, the stock returns are negative, consumer sentiment index decreases and purchase manager's index drops as well. The response of housing starts estimated by FAVAR$_{3PRF}$ is much milder in the long run, compared to other two models. The estimated purchase manager's index response by VAR model is also strange since it exhibits economic improvement in the first six months.

6.2 Forecast Error Variance Decompositions

Forecast error variance decomposition (hereinafter: FEVD), estimates the amount of mean squared prediction error (hereinafter: MSPE) of C_t explained by particular shock w_{kt} at horizon $h = 0, 1, ..., H$. Under the assumption of stationarity, forecast error variance decomposition tends to variance decomposition of C_t as $h \to \infty$ because $MSPE \to \Sigma_C$. In integrated systems, MSPE diverges as $h \to \infty$ but FEVD remains valid up to finite horizon H (Kilian & Lütkepohl, 2017, p. 111).

The only necessary input for calculation of variance decomposition, are the matrices Θ_i, calculated already in the previous subsection. For a FAVAR process, the h-step ahead forecast error is

$$C_{t+h} - C_{t+h|t} = \sum_{i=1}^{h-1} \Phi_i u_{t+h-i} = \sum_{i=1}^{h-1} \Theta_i w_{t+h-i}. \tag{42}$$

Mean squared prediction error at horizon h is defined as

$$\begin{aligned} MSPE(h) \equiv E[(C_{t+h} - C_{t+h|t})(C_{t+h} - C_{t+h|t})'] &= \sum_{i=1}^{h-1} \Phi_i \Sigma_u \Phi_i' \\ &= \sum_{i=1}^{h-1} \Theta_i \Sigma_w \Theta_i'; \quad \Sigma_w = I_K \\ &= \sum_{i=1}^{h-1} \Theta_i \Theta_i'. \end{aligned} \tag{43}$$

From here, I can calculate the contribution of shock j to total MSPE of variable k at horizon h as:

$$MSPE^k(h) = \sum_{j=1}^{K} MSPE_j^k(h) = \sum_{j=1}^{K} (\theta_{jk,0}^2 + ... + \theta_{jk,h-1}^2). \tag{44}$$

Dividing elements in summation operator by $MPSE^k(h)$ yields the fraction of contribution of each shock to the forecast error variance of variable k (Kilian & Lütkepohl, 2017, p. 112-114).

The amount of the explained variance of oil market variables by variables included in SFAVAR $_{3PRF}$ are reported in the tables below. The question that I am trying to answer is, how important are particular variables in explanation of World Crude oil production, World economic activity, oil inventories, oil price and monetary policy.

In Table 6, the proportions of the explained variance of World Crude oil production are presented. As expected, the crude oil production variation is mostly explained by the internal process and with both 3PRF relevant factors. However, as already indicated in the previous section, in the long run, the impacts of other shocks could also have the effects on oil production. Especially oil-specific demand shocks explain a significant share of oil production variance since oil producers follow higher demand. Speculative pressure (sp_t) on the other hand explains only around 3 % of oil production variance which states that speculative pressure from financial markets is not a crucial determinant of oil production. This conclusion differs from Juvenal and Petrella (2015) who found out that the explained part of oil production variance by speculative shocks is around 20 %. They, however, use different identification scheme, and their speculative shock is defined as the shock on the physical market, whereas I modeled shocks from financial market separately.

Table 6: Variance Decomposition of World Crude Oil Production

$Horizon$	$\Delta prod$	rea	inv	sp	rpo	F_1	F_2	sr
1	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.932	0.003	0.010	0.032	0.008	0.008	0.006	0.002
8	0.822	0.008	0.050	0.031	0.020	0.043	0.005	0.020
12	0.752	0.016	0.073	0.030	0.024	0.070	0.012	0.023
24	0.653	0.030	0.084	0.027	0.024	0.118	0.041	0.021
48	0.599	0.031	0.078	0.028	0.061	0.122	0.051	0.030

Source: Own work.

Monetary policy shocks also explain approximately 3 % of total variation in the long run and surprisingly, commodity demand shocks also explain only 3 % of the variation in oil production. The results suggest that oil producers closely observe only the demand for oil and they do not respond much to changes in world economic activity.

Table 7: Variance Decomposition of the Real Economic Activity

$Horizon$	$\Delta prod$	rea	inv	sp	rpo	F_1	F_2	sr
1	0.001	0.999	0.000	0.000	0.000	0.000	0.000	0.000
4	0.002	0.872	0.002	0.006	0.069	0.038	0.010	0.001
8	0.009	0.688	0.013	0.017	0.104	0.080	0.088	0.001
12	0.022	0.589	0.021	0.031	0.107	0.103	0.126	0.001
24	0.032	0.468	0.066	0.031	0.113	0.144	0.145	0.002
48	0.032	0.434	0.106	0.028	0.104	0.148	0.143	0.005

Source: Own work.

Given the relative importance of world economic activity proxy for the crude oil market, I present the shares of explained variance of world real economic activity proxy in Table 7. In the long run, real economic activity is strongly driven by shocks in both $3PRF$ factors which explain almost 30 % of variation after four years.

Since the factors are based on the U.S. macroeconomic data, this implicitly implies that the U.S. macroeconomic shocks significantly affect the world economy. Residual oil price shocks (rpo) are the second most important driver of economic activity since they explain approximately 10 % of economic activity variation in almost all horizons. Oil specific demand shocks on the other side affect the real economic activity only by a lag of one year and explain 2.1 % of the variation and the explained share increases to 10.6 % after three years. The explained variation by speculative pressure explains at most 3.1 % of the variation. Oil supply shocks ($\Delta prod$) explain at most 3.2 % of variation but only at larger horizons. The effect of the U.S. monetary policy is negligible.

Table 8: Variance Decomposition of World Oil Inventories

$Horizon$	$\Delta prod$	rea	inv	sp	rpo	F_1	F_2	sr
1	0.016	0.000	0.984	0.000	0.000	0.000	0.000	0.000
4	0.092	0.000	0.864	0.010	0.005	0.019	0.007	0.003
8	0.134	0.004	0.746	0.017	0.019	0.046	0.032	0.002
12	0.177	0.006	0.674	0.014	0.027	0.044	0.055	0.002
24	0.262	0.027	0.565	0.018	0.025	0.036	0.065	0.002
48	0.284	0.038	0.532	0.022	0.024	0.034	0.062	0.004

Source: Own work.

The variance decomposition of oil inventories (inv) is especially crucial for comparison with other studies, where different shock identification scheme is applied. In my setting, oil inventories are defined as a stock which interacts between supply and demand, whereas in Kilian and Murphy (2014), Juvenal and Petrella (2015) oil inventories are assumed to be the source of speculation. The results of inv variance decomposition are available in Table 8. The only shock that notably affects oil inventories is an oil supply shock as a consequence of production smoothing process. This result was also confirmed by Juvenal and Petrella (2015). Estimated shares of explained variance by speculative shocks (sp) and global commodity demand shocks (rea) are much less prevalent as proposed by Juvenal and Petrella (2015). My results, therefore, suggest that speculative shocks are mostly not reflected in oil inventories, meaning that speculators hold only a negligible part of oil reserves.

The most important question that I would like to answer in this subsection is the importance of speculative shocks from the financial part of oil market for the oil price. To answer this question, I present oil price variance decomposition in Table 9. An interesting finding is, that in the short run, speculative shocks explain more variation than shocks in real economic activity or oil-specific demand shocks, whereas, in the long run, the demand shocks gain importance. U.S. macroeconomic shocks and monetary policy shocks explain a very limited part of the variation in oil price.

Table 9: Variance Decomposition of Real Crude Oil Price

$Horizon$	$\Delta prod$	rea	inv	sp	rpo	F_1	F_2	sr
1	0.011	0.018	0.029	0.130	0.813	0.000	0.000	0.000
4	0.003	0.035	0.097	0.105	0.722	0.003	0.025	0.008
8	0.005	0.051	0.108	0.107	0.699	0.003	0.021	0.006
12	0.004	0.078	0.086	0.117	0.684	0.007	0.019	0.005
24	0.010	0.119	0.056	0.119	0.641	0.032	0.013	0.010
48	0.038	0.117	0.041	0.104	0.587	0.066	0.012	0.035

Source: Own work.

Reduced ability of oil supply shocks ($\Delta prod$) in explaining the variation in oil price is already well established in oil market literature as it was already noted by Kilian (2009) and confirmed by Aastveit (2014), Juvenal and Petrella (2015), among others. According to obtained results, I can conclude, that speculative shocks have a non-negligible effect on oil price and have to be explicitly considered in oil market models.

Table 10: Variance Decomposition of the Shadow Federal Funds Rate

$Horizon$	$\Delta prod$	rea	inv	sp	rpo	F_1	F_2	sr
1	0.002	0.010	0.039	0.004	0.037	0.116	0.048	0.743
4	0.002	0.003	0.051	0.007	0.068	0.207	0.069	0.593
8	0.018	0.002	0.054	0.008	0.058	0.247	0.108	0.505
12	0.027	0.002	0.074	0.009	0.049	0.240	0.126	0.473
24	0.047	0.002	0.129	0.006	0.030	0.207	0.156	0.423
48	0.057	0.007	0.140	0.004	0.082	0.180	0.173	0.357

Source: Own work.

Finally, for the sake of completeness, I present the forecast error variance decomposition of shadow federal funds rate. If a particular variable or factor has the power to explain variation in the primary monetary policy instrument, this implies that monetary authority reacts to that information. As expected, the variability of shadow federal funds rate can mostly be explained by relevant factors based on U.S. macroeconomy. However, monetary policy reacts to residual oil price shocks, and to oil supply shocks and oil-specific demand shocks in the long run. Global commodity demand shocks have no explanatory power for shadow federal funds rate, indicating that FED does not react to economic depressions outside the U.S. Results also show, that monetary policy also does not react to speculative shock in the financial market. This conclusion neglects the response of monetary policy to increased volatility, as it was suggested by Bekaert, Hoerova and Duca (2013) who analyzed the response to the stock market volatility.

6.3 Historical Decomposition of Oil Price

Historical decompositions are oriented to the question, which shock primarily caused the observed significant fluctuations in C_t. Historical decomposition is, therefore, aimed at actual movements in the data and not towards unconditional expectations. The interest is, therefore, in the cumulative effect of a particular structural shock to each variable at every given point in time. Such analysis is, therefore, well suited for analysis how much effect the speculative shock has on oil price fluctuation not only on average but also the monthly effect for the last two decades (Kilian & Lütkepohl, 2017, p.114). Historical variance decomposition therefore enables me to answer whether speculative shocks have caused oil price variation in the last 16 years.

Historical decomposition is based on the fact that C_t can be rewritten as a combination of initial conditions and the cumulative effect of structural shocks. Historical decomposition is therefore calculated such that for any t:

$$\begin{aligned} C_t &= \sum_{s=0}^{t-1} \Theta_s w_{t-s} + \sum_{s=t}^{\infty} \Theta_s w_{t-s} \\ &= \sum_{s=0}^{t-1} \Theta_s w_{t-s} + IC. \end{aligned} \tag{45}$$

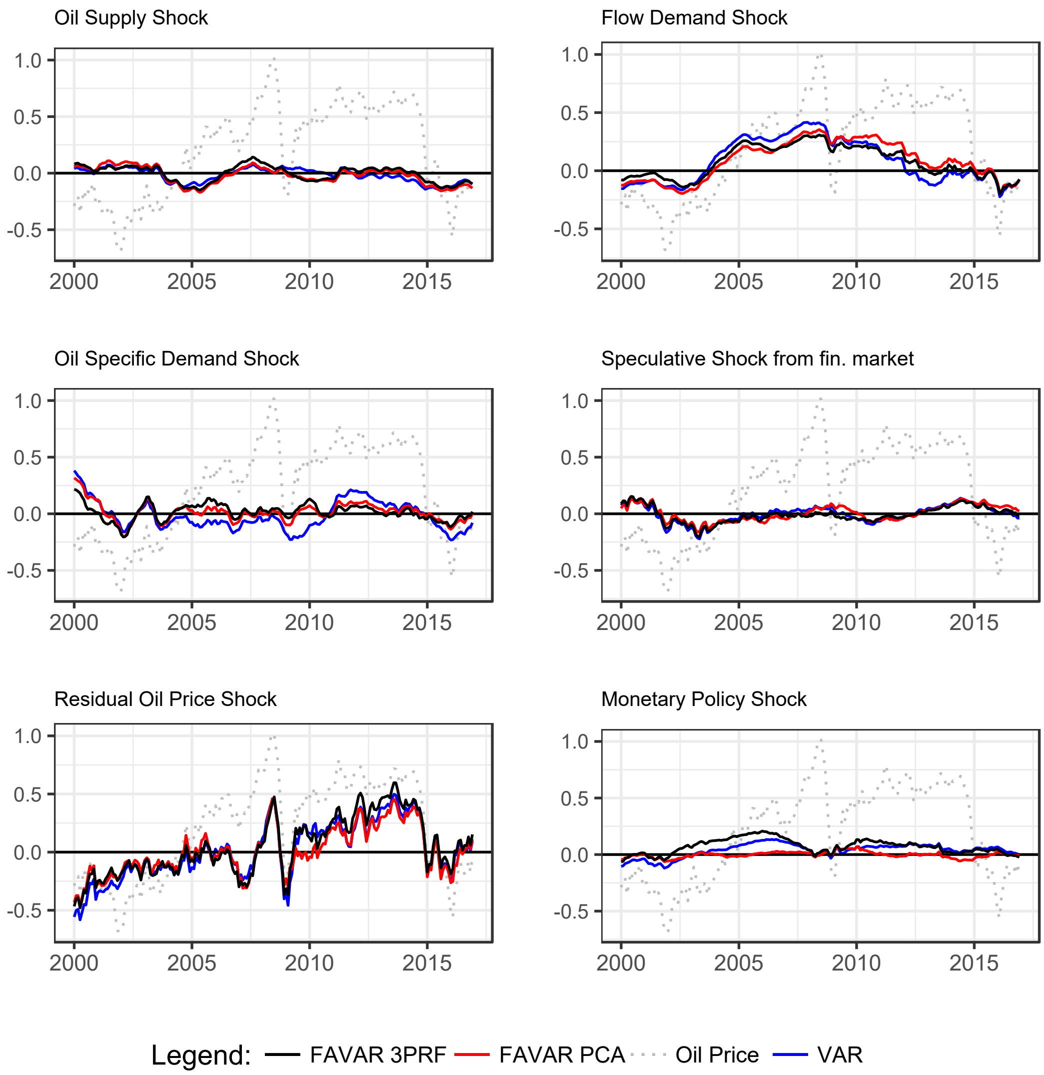
The second term of equation (45) is referred to as initial conditions since it represents shocks before the beginning of the sample and thus cannot be estimated. C_t can thus only be estimated by dropping initial conditions, such that $\hat{C}_t = \sum_{s=0}^{t-1} \Theta_s w_{t-s}$ (Kilian & Lütkepohl, 2017, p.114).

Calculation involves structural moving average coefficient matrices Θ_i and structural shocks $w_t = B_0 u_t \ t = 1, ..., T$. Each structural shock j is matched with the appropriate impulse response weight to form fitted time series $\hat{C}_K^{(j)}$ representing the time series of cumulative isolated response of variable K to the shock in j.

$$\hat{C}_{rpo,t}^{(j)} = \sum_{i=0}^{t-1} \theta_{rpo,j,i} w_{j,t-i}, \tag{46}$$

where $\theta_{rpo,j,i} w_{j,t-i}$ represents the response of real price of oil (rpo) to shock in j at horizon i and $w_{j,t}$ represents structural shock in j at time t. $\hat{C}_{rpo,t}^{(j)}$ represents the cumulative contribution of shock j on rpo over time. By construction, it also holds that $\hat{C}_{rpo,t} = \sum_{j=1}^{K} \hat{C}_{rpo,t}^{(j)}$. This approximation approaches demeaned rpo as $t \to \infty$ and therefore approximation for the last two decades in the sample is only trivially different than true real price of oil (Kilian & Lütkepohl, 2017, p.115-117).

Figure 15: Historical Decomposition of Oil Price



Source: Own work.

In Figure 15 the historical decomposition of oil price for the period between January 2000 and December 2016 is presented. Grey dashed line represents demeaned logarithm of the actual real oil price in the observed period, whereas other three curves represent the oil price if only particular shock would determine the price, estimated by three models, presented previously.

The results of the historical decomposition suggest that global commodity demand shock is the most important driver of oil price. This result is consistent with the evidence from other studies and with the rest of the results presented in this thesis. Supply shocks only have some negative effect on price in the period from 2003 to 2006 when the price would increase further without additional production and it also partially contributed to the price drop in 2015. Speculative shocks from the financial part of oil market do not affect oil prices significantly. However, the speculative shocks significantly negatively affected oil price in the period between 2002 and 2005. The positive effect on oil price is estimated for the period between 2013 and 2015, meaning, that high oil prices prior to 2015 were at least partially driven by speculation. Economou, Agnolucci, Fattouh and De Lipis (2017) also obtained a weak affection of oil price by speculative shock. The obtained results are consistent with Alquist and Gervais (2013), who also proposed that speculation from the financial market does not cause the oil price surge. According to their study, prices are predominantly driven by demand factors, which are especially important due to production constraints. Residual oil price shock is also an essential driver of oil price especially in the period from 2010 onward when other shocks do not exhibit almost any price increase. This finding suggests that additional shocks significantly affect oil price in this period and that speculation from the financial market was not the reason for increasing prices observed in the period. During the period from 2010 to 2015, the oil price is not well explained by structural oil shocks, because the residual oil price variation is considerable and the oil price was driven with some other factors, which are not included in this model and presumably arise from other non-U.S. countries.

Oil specific demand shocks contributed to price decrease in 2002 and also to price increase from 2011 to 2014. However, the effects of oil specific demand are limited, because these shocks mainly occur as strategic decisions, which are meant to affect oil market in the least possible way. Monetary policy shocks significantly affected the oil price in the period between 2002 and 2007 as estimated by FAVAR_{3PRF} and VAR model, but the effect is not confirmed by FAVAR_{PCA} model.

CONCLUSION

The FAVAR_{3PRF} is a new method for FAVAR analysis, where factors summarize the information in a large dataset conditional on the target variable. Because the factors are leaned towards the particular variable of interest, the response of a specific economic sector can be modeled more efficiently than with alternative methods, where factors do not depend on the target. This model is, therefore, well suited for analysis of oil shocks since irrelevant information from a large data set could lead to misleading conclusions.

Oil price shocks are an essential determinant of macroeconomic performance due to the importance of oil in the economy. Even though oil has become a less critical source of energy in recent years, significant price movements can still cause serious economic distortions due to its relation to other commodities and transportation costs. The importance of the source of oil price shock, initially proposed by Kilian (2009), is well established nowadays. In this thesis, I consider many different oil shocks, namely, oil supply shock, shock in global commodity demand, oil-specific demand shock, speculative shock from the financial part of oil market and residual oil price shock. I found out that every type of shock hits the economy differently, even though the similarities of responses to different types of demand shocks are considerable.

Monetary policy reaction to oil shocks depends on the underlying source of the oil shock. The reaction to oil supply shock is estimated to be expansionary, while results suggest that there is almost no reaction to shock in global demand for industrial commodities, oil-specific demand shock and speculative shock from the financial part of the crude oil market. Strong expansionary monetary policy response is also estimated to residual oil demand shock. This shock probably reflects non-industrial consumption of non-U.S. countries and monetary authority mitigates the effect on U.S. industrial production by expansionary policy even on the cost of higher inflation.

The important question to address is the potential reverse impact of the monetary policy on the oil market, which was only modestly studied. Aastveit (2014) in a similar study controlled for effects of monetary policy and noted the ability of Federal Reserve to considerably affect oil market, especially oil production. My results also suggest the importance of monetary policy for the oil market and are in line with Frankel (2008). Restrictive monetary policy shock causes higher oil prices in the first few months and by approximately 20 % lower prices in the long run due to increased supply, lower demand, and net short speculative open interest.

The share of the explained variability of real oil price by speculative pressure from the financial market is estimated at approximately 10 %. This amount of explained variance is substantial, and due to increasing part of non-commercial traders in the financial part of the oil market, the importance of speculation for crude oil determination could increase in the future. In the period from 2000 to 2017, speculative pressure positively affected oil price only from 2010 to 2015. Results of historical oil price decomposition are in line with other studies, suggesting that the most important drivers of oil price are demand shocks for industrial commodities.

The test for information sufficiency rejected the null hypothesis that the baseline VAR model is informationally sufficient. The test results indicate that two factors are required to eliminate informational deficiency of the VAR model. The estimated impulse responses also show that VAR model suffers from a severe informational deficiency because some of the estimated impulse responses are highly inconsistent with economic theory, which is eliminated by addition of factors in the model. Another interesting finding that steams from the test for informational sufficiency is directly related to the comparison of $\text{FAVAR}_{\text{PCA}}$ and $\text{FAVAR}_{\text{3PRF}}$ methods. When in case of $\text{FAVAR}_{\text{3PRF}}$ one factor is enough to achieve informational sufficiency, $\text{FAVAR}_{\text{PCA}}$ needs at least two factors, and the forecasting usability is not strictly decreasing in the order of factor since fourth PCA factor improves the forecast more than third PCA factor. The test for informational deficiency therefore in a way proves, the outperformance of 3PRF method of factor estimation.

The choice of factor estimation procedure is not of minor importance. Whereas PCA factors mapping depends on the maximal amount of information available in the data set, 3PRF factors contain only relevant information for a particular variable of interest. The results suggest that in most cases $\text{FAVAR}_{\text{PCA}}$ and $\text{FAVAR}_{\text{3PRF}}$ provide similar results. $\text{FAVAR}_{\text{3PRF}}$

provides a way to efficiently estimate responses of variables also in cases when the shock or response variable do not represent a significant part of the total variation in the dataset. An example is the response of oil price to the speculative shock from the financial market, where $FAVAR_{3PRF}$ estimates sensible (positive) response, whereas $FAVAR_{PCA}$ and VAR estimate the negative response of oil price, presumably due to informational irrelevance or deficiency.

$FAVAR_{3PRF}$, therefore, seems to improve traditional FAVAR models. The drawbacks of the models are reflected mainly in the computationally exhaustive estimation of multiple models, one model for each response variable of interest. This drawback is an obvious consequence of response specific factors included in the model, and therefore re-estimation of the model is required for every response variable. On the other hand, the results are theoretically more sensible, and the $FAVAR_{3PRF}$ can track down also the responses of less prevalent sectors of the economy, which is a desirable property of the descriptive macroeconomic model.

REFERENCE LIST

1. Aastveit, K. A. (2014). Oil price shocks in a data-rich environment. *Energy Economics, 45*, 268-279.
2. Aastveit, K. A., Bjørnland, H. C. & Thorsrud, L. A. (2015). What drives oil prices? Emerging versus developed economies. *Journal of Applied Econometrics, 30*(7), 1013-1028.
3. Alquist, R. & Gervais, O. (2013). The role of financial speculation in driving the price of crude oil. *The Energy Journal*, 35-54.
4. Alquist, R. & Kilian, L. (2010). What do we learn from the price of crude oil futures? *Journal of Applied Econometrics, 25*(4), 539-573.
5. Anzuini, A., Lombardi, M. & Pagano, P. (2013). The Impact of Monetary Policy Shocks on Commodity Prices. *International Journal of Central Banking, 9*(3), 125-150.
6. Bachmeier, L. (2008). Monetary policy and the transmission of oil shocks. *Journal of Macroeconomics, 30*(4), 1738-1755.
7. Bai, J., Li, K. & Lu, L. (2016). Estimation and inference of FAVAR models. *Journal of Business & Economic Statistics, 34*(4), 620-641.
8. Bai, J. & Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica, 70*(1), 191-221.
9. Bai, J. & Ng, S. (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics, 146*(2), 304-317.
10. Barsky, R. B. & Kilian, L. (2001). Do we really know that oil caused the great stagflation? A monetary alternative. *NBER Macroeconomics annual, 16*, 137-183.
11. Barsky, R. B. & Kilian, L. (2004). Oil and the Macroeconomy since the 1970s. *Journal of Economic Perspectives, 18*(4), 115-134.
12. Baumeister, C. & Kilian, L. (2016). Forty Years of Oil Price Fluctuations: Why the Price of Oil May Still Surprise Us. *The Journal of Economic Perspectives, 30*(1), 139-160.

13. Bekaert, G., Hoerova, M. & Duca, M. L. (2013). Risk, uncertainty and monetary policy. *Journal of Monetary Economics, 60*(7), 771-788.
14. Belviso, F. & Milani, F. (2006). Structural Factor-Augmented VARs (SFAVARs) and the Effects of Monetary Policy. *The BE Journal of Macroeconomics, 6*(3), 1-46.
15. Bernanke, B. S. (1986). Alternative explanations of the money-income correlation. *Carnegie-Rochester conference series on public policy, 25*, 49-99.
16. Bernanke, B. S., Boivin, J. & Eliasz, P. (2005). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach. *The Quarterly journal of economics, 120*(1), 387-422.
17. Bernanke, B. S., Gertler, M., Watson, M., Sims, C. A. & Friedman, B. M. (1997). Systematic monetary policy and the effects of oil price shocks. *Brookings papers on economic activity, 1997*(1), 91-157.
18. Blanchard, O. J. & Quah, D. (1989). The Dynamic Effects of Aggregate Demand and Supply Disturbances. *The American Economic Review, 79*(4), 655-673.
19. Bodenstein, M., Guerrieri, L. & Kilian, L. (2012). Monetary policy responses to oil price fluctuations. *IMF Economic Review, 60*(4), 470-504.
20. Bohi, D. R. (1991). On the macroeconomic effects of energy price shocks. *Resources and Energy, 13*(2), 145-162.
21. Boivin, J., Giannoni, M. P. & Mihov, I. (2009). Sticky prices and monetary policy: Evidence from disaggregated US data. *American economic review, 99*(1), 350-84.
22. Boivin, J. & Ng, S. (2006). Are more data always better for factor analysis? *Journal of Econometrics, 132*(1), 169-194.
23. Bruno, V. & Shin, H. S. (2015). Capital flows and the risk-taking channel of monetary policy. *Journal of Monetary Economics, 71*, 119-132.
24. Caplinger, M. W. (1997). Allegheny oil heritage project: a contextual overview of crude oil production in Pennsylvania. *HAER No. PA-436*. Retrieved June 5, 2018, from http://www.as.wvu.edu/ihtia/ANF%20Oil%20Context.pdf
25. Clark, T. E. (2004). Can out-of-sample forecast comparisons help prevent overfitting?. *Journal of forecasting, 23*(2), 115-139.
26. Considine, T. J. (1997). Inventories under joint production: An empirical analysis of petroleum refining. *Review of Economics and Statistics, 79*(3), 493-502.
27. Cunado, J. & de Gracia, F. P. (2014). Oil price shocks and stock market returns: Evidence for some European countries. *Energy Economics, 42*, 365-377.
28. De Waele, S. & Broersen, P. M. (2003). Order selection for vector autoregressive models. *IEEE Transactions on Signal Processing, 51*(2), 427-433.
29. Economic Research, Federal Reserve Bank of St. Louis. (no date). *FRED-MD and FRED-QD: Monthly and Quarterly Databases for Macroeconomic Research*. Retrieved February 2, 2018, from https://research.stlouisfed.org/econ/mccracken/fred-databases/
30. Economou, A., Agnolucci, P., Fattouh, B. & De Lipis, V. (2017). A Structural Model of the World Oil Market: The Role of Investment Dynamics and Capacity Constraints. *Oxford Institute for Energy Studies, 75*.

31. Energy Information Administration. (no date a). *U.S. Petroleum & Other Liquids Stocks by Stock Type*. Retrieved February 3, 2018, from https://www.eia.gov/dnav/pet/pet_stoc_typ_c_nus_EP00_mbbl_m.htm
32. Energy Information Administration. (no date b). *WTI Spot Prices*. Retrieved February 3, 2018, from https://www.eia.gov/dnav/pet/pet_pri_spt_s1_m.htm
33. Fattouh, B., Kilian, L. & Mahadeva, L. (2013). The role of speculation in oil markets: What have we learned so far?. *The Energy Journal*, 7-33.
34. Federal Reserve Bank of Atlanta. (no date). *Wu-Xia Shadow Federal Funds Rate*. Retrieved March 7, 2018, from https://www.frbatlanta.org/cqer/research/shadow_rate.aspx
35. Ferderer, J. P. (1996). Oil price volatility and the macroeconomy. *Journal of macroeconomics, 18*(1), 1-26.
36. Forni, M. & Gambetti, L. (2014). Sufficient information in structural VARs. *Journal of Monetary Economics, 66*, 124-136.
37. Frankel, J. A. (1986). Expectations and commodity price dynamics: The overshooting model. *American Journal of Agricultural Economics, 68*(2), 344-348.
38. Frankel, J. A. (2008). The Effect of Monetary Policy on Real Commodity Prices. *Asset Prices and Monetary Policy*, 291-333
39. Fueki, T., Higashi, H., Higashio, N., Nakajima, J., Ohyama, S. & Tamanyu, Y. (2016). *Identifying Oil Price Shocks and their Consequences Role of Expectations and Financial Factors in the Crude Oil Market* (Working Paper No. 16-E-17). Tokyo: Bank of Japan.
40. Gelper, S. & Croux, C. (2007). Multivariate out-of-sample tests for Granger causality. *Computational statistics & data analysis, 51*(7), 3319-3329.
41. Greene, W. H. (2012). *Econometric Analysis* (7th ed.). London: Prentice Hall.
42. Guérin, P., Leiva-Leon, D. & Marcellino, M. (2017). *Markov-switching three-pass regression filter* (Working Paper No. 1748). Madrid: Banco de España.
43. Haase, M., Zimmermann, Y. S. & Zimmermann, H. (2017). Commodity returns and their volatility in relation to speculation: a consistent empirical approach. *The Journal of Alternative Investments, 20*(2), 76-91.
44. Hamilton, J. D. (1983). Oil and the macroeconomy since World War II. *Journal of political economy, 91*(2), 228-248.
45. Hamilton, J. D. (1996). This is what happened to the oil price-macroeconomy relationship. *Journal of Monetary Economics, 38*(2), 215-220.
46. Hamilton, J. D. (2003). What is an oil shock?. *Journal of econometrics, 113*(2), 363-398.
47. Hamilton, J. D. (2009). Causes and Consequences of the Oil Shock of 2007–08. *Brookings Papers on Economic Activity, 2009*(1), 215-261.
48. Hamilton, J. D. (2011). *Historical oil shocks*. In Parker, R. E., & Whaples, R. M. (Eds.), *Routledge Handbook of Major Events in Economic History* (p. 239-265). New York: Routledge.

49. Hamilton, J. D. & Herrera, A. M. (2004). Oil shocks and aggregate macroeconomic behavior: The role of monetary policy: A comment. *Journal of Money, Credit, and Banking, 36*(2), 265-286.
50. Hamilton, J. D. & Wu , J. C. (2014). Risk premia in crude oil futures prices. *Journal of International Money and Finance, 42*, 9-37
51. Harvey, D. S., Leybourne, S. J., & Newbold, P. (1998). Tests for forecast encompassing.*Journal of Business & Economic Statistics, 16*(2), 254-259.
52. Hooker, M. A. (1996). What happened to the oil price-macroeconomy relationship?. *Journal of monetary Economics, 38*(2), 195-213.
53. Hubbard, R. G. & Weiner, R. J. (1986). Oil supply shocks and international policy coordination. *European economic review, 30*(1), 91-106.
54. Institute of Supply Management. (no date). *ISM Report*. Retrieved Retrieved March 7, 2018, from https://www.instituteforsupplymanagement.org
55. Jaffe, A. M. & Soligo, R. (2002). The role of inventories in oil market stability. *The quarterly review of economics and finance, 42*(2), 401-415.
56. Jones, D. W. & Leiby, P. N. (1996). The macroeconomic impacts of oil price shocks: a review of the literature and issues. *Oak Ridge National Laboratory*, 33.
57. Juvenal, L. & Petrella, I. (2015). Speculation in the oil market. *Journal of Applied Econometrics, 30*(4), 621-649.
58. Kaufmann, R. K. (2011). The role of market fundamentals and speculation in recent price changes for crude oil. *Energy Policy, 39*(1), 105-115.
59. Kelly, B. & Pruitt, S. (2015). The three-pass regression filter: A new approach to forecasting using many predictors. *Journal of Econometrics, 186*(2), 294-316.
60. Kilian, L. (1998). Small-sample confidence intervals for impulse response functions. *Review of economics and statistics, 80*(2), 218-230.
61. Kilian, L. (2008). Exogenous oil supply shocks: how big are they and how much do they matter for the US economy?. *The Review of Economics and Statistics, 90*(2), 216-240.
62. Kilian, L. (2009). Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. *American Economic Review, 99*(3), 1053-69.
63. Kilian, L. (2014). Oil Price Shocks: Causes and Consequences. *Annual Review of Resource Economics, 6*(1), 133-154.
64. Kilian, L. & Hicks, B. (2013). Did unexpectedly strong economic growth cause the oil price shock of 2003–2008?. *Journal of Forecasting, 32*(5), 385-394.
65. Kilian, L. & Lee, T. K. (2014). Quantifying the speculative component in the real price of oil: The role of global oil inventories. *Journal of International Money and Finance, 42*, 71-87.
66. Kilian, L. & Lewis, L. T. (2011). Does the Fed respond to oil price shocks? *The Economic Journal, 121*(555), 1047-1072.
67. Kilian, L. & Lütkepohl, H. (2017). *Structural vector autoregressive analysis*. Cambridge: Cambridge University Press.
68. Kilian, L. & Murphy, D. P. (2014). The role of inventories and speculative trading in the global market for crude oil. *Journal of Applied Econometrics, 29*(3), 454-478.

69. Kilian, L. & Park, C. (2009). The impact of oil price shocks on the US stock market. *International Economic Review, 50*(4), 1267-1287.
70. Kucher, O. & Kurov, A. (2014). Business cycle, storage, and energy prices. *Review of Financial Economics, 23*(4), 217-226.
71. Labonte, M. (2004). The effects of oil shocks on the economy: a review of the empirical evidence. *Congressional Research Service, Library of Congress*.
72. Lee, K., Ni, S. & Ratti, R. A. (1995). Oil shocks and the macroeconomy: the role of price variability. *The Energy Journal*, 39-56.
73. Lütkepohl, H. & Krätzig, M. (2004). *Applied time series econometrics*. New York: Cambridge University Press.
74. MacKinnon, J. G. (2009). Bootstrap hypothesis testing. *Handbook of Computational Econometrics*, 183-213.
75. McCracken, M. W. & Ng, S. (2016). FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics, 34*(4), 574-589.
76. Medlock III, K. B. (2013). Speculation, Fundamentals, and the Price of Crude Oil. *James A. Baker III Institute for Public Policy Rice University*
77. Mork, K. A. (1989). Oil and the macroeconomy when prices go up and down: an extension of Hamilton's results. *Journal of political Economy, 97*(3), 740-744.
78. Olmstead, A. L. & Rhode, P. (1985). Rationing without government: The west coast gas famine of 1920. *The American economic review, 75*(5), 1044-1055.
79. Parsons, J. E. & Espinasa, R. (2010). Black Gold and Fool's Gold: Speculation in the Oil Futures Market [with Comment]. *Economia, 10*(2), 81-116.
80. Pfaff, B. (2008). VAR, SVAR and SVEC models: Implementation within R package vars. *Journal of Statistical Software, 27*(4), 1-32.
81. Roekchamnong, W., Pornchaiwiseskul, P., & Chiarawongse, A. (2014). The effects of uncertainties on inventory management of petroleum products: a case study of Thailand. *International Journal of Energy Economics and Policy, 4*(3), 380-390.
82. Runkle, D. E. (1987). Vector autoregressions and reality. *Journal of Business & Economic Statistics, 5*(4), 437-442.
83. Shapiro, M. D. & Watson, M. W. (1988). Sources of business cycle fluctuations. *NBER Macroeconomics annual*, 3, 111-148.
84. Sims, C. A. (1980). Macroeconomics and reality. *Econometrica: Journal of the Econometric Society*, 1-48.
85. Sims, C. A. (1986). Are forecasting models usable for policy analysis?. *Quarterly Review, (Win)*, 2-16.
86. Stock, J. H. & Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American statistical association, 97*(460), 1167-1179.
87. Stock, J. H. & Watson, M. W. (2005). *Implications of dynamic factor models for VAR analysis* (Working Paper No. w11467). Cambridge: National Bureau of Economic Research.

88. Stock, J. H. & Watson, M. (2011). *Dynamic Factor Models.* In Clements, M. J. and Hendry, D. F. (Eds.), *Oxford Handbook on Economic Forecasting* (p. 35-59). Oxford: Oxford University Press.
89. Teisberg, T. J. (1981). A dynamic programming model of the US strategic petroleum reserve. *The Bell Journal of Economics*, 526-546
90. Uhlig, H. (2005). What are the effects of monetary policy on output? Results from an agnostic identification procedure. *Journal of Monetary Economics, 52*(2), 381-419.
91. U.S. Commodity Futures Trading Commission. (no date). *Commitment of Traders Reports*. Retrieved March 7, 2018, from https://www.cftc.gov/MarketReports CommitmentsofTraders/index.htm
92. Williamson, H. F. & Daum, A. R. (1959). *The American Petroleum Industry: 1859-1899. The Age of Illumination.* Evanston: Northwestern University Press.
93. Wold, H. O. (1960). A generalization of causal chain models (Part III of a Triptych on Causal Chain Systems). *Econometrica: Journal of the Econometric Society*, 443-463.
94. Wu, J. C. & Xia, F. D. (2016). Measuring the macroeconomic impact of monetary policy at the zero lower bound. *Journal of Money, Credit and Banking, 48*(2-3), 253-291.
95. Yamamoto, Y. (2016). *Bootstrap Inference for Impulse Response Functions in Factor-Augmented Vector Autoregressions* (Working Paper No. HIAS-E-26). Tokyo: Hitotsubashi Institute for Advanced Study, Hitotsubashi University.

APPENDICES

Appendix 1: Povzetek v slovenskem jeziku

ANALIZA NAFTNIH ŠOKOV S FAKTORJI, OCENJENIMI Z REGRESIJSKIM FILTROM S TREMI PREHODI ZNOTRAJ FAVAR

Uvod

Od razvoja industrije in transporta v smeri večje uporabe fosilnih goriv velja cena nafte za enega od glavnih vzrokov za makro-ekonomsko nestanovitnost. Že Hamilton (1983) je dokazal, da so vse večje recesije sledile obdobjem visokega povišanja cen nafte. Ta povezava je logična, saj so nafta in naftni derivati kljub razvoju drugih virov energije še vedno ključen energent za delovanje gospodarstva. Vseeno pa se vplivi naftnih šokov v zadnjem času vse bolj zmanjšujejo, kar je predvsem posledica boljše politike prilagajanja šokom, njihovega krajšega trajanja in njihove manjše intenzitete.

Vplive naftnih šokov na gospodarstvo sem proučil z uporabo vektorsko avtoregresijskega modela razširjenega s faktorji (ang. Factor Augmented Vector Autoregression, v nadaljevanju: FAVAR), kjer so faktorji ocenjeni po principu regresijskega filtra s tremi prehodi (ang. Three-Pass Regression Filter, v nadaljevanju: 3PRF), katerega sta predstavila Kelly in Pruitt (2015). Prednosti omenjenega modela se kažejo predvsem v uporabi dodatne informacije, ki izvira iz velikega števila makroekonomskih spremenljivk za Združene države Amerike in v uporabi zgolj relevantnega dela informacije, kar ta model loči od predhodno predstavljenih FAVAR, kjer so faktorji ocenjeni na podlagi analize poglavitnih komponent (ang. Principal Components Analysis, v nadaljevanju: PCA). Za testiranje smiselnosti in uporabnosti 3PRF faktorjev, ocene FAVAR modela s 3PRF faktorji (v nadaljevanju: $\text{FAVAR}_{\text{3PRF}}$) primerjam z ocenami FAVAR modela s PCA faktorji ($\text{FAVAR}_{\text{PCA}}$) in najbolj osnovnim vektorsko avtoregresijskim (ang. Vector Autoregressive, v nadaljevanju: VAR) modelom.

Temeljna hipoteza, ki jo potrjujem v magistrskem delu, je, da osnovni VAR model ne vsebuje vse relevantne informacije, kar pomeni, da bodo impulzni odzivi ocenjeni s FAVAR modeloma boljši od tistih ocenjenih z VAR modelom. Nadalje ugotavljam razlike med ocenjenimi impulznimi odzivi med $\text{FAVAR}_{\text{3PRF}}$ in $\text{FAVAR}_{\text{PCA}}$ modeloma in razloge za le-te. Druga hipoteza se nanaša na razlike med odzivi gospodarstva na različne izvore naftnih šokov in na različne odzive monetarne politike na posamezne šoke. Predpostavljam namreč, da igra izvor šoka pomembno vlogo tako pri odzivu gospodarstva kot tudi pri odzivu monetarne politike. Zadnje raziskovalno vprašanje pa je, kakšna je povratna odvisnost naftnega trga na šoke monetarne politike Združenih držav Amerike. Predstavljeni FAVAR model z ocenitvijo faktorjev po principu 3PRF je zanimiva kombinacija dveh modelov, zato je opis prednosti in slabosti modela še toliko bolj pomemben.

1 Podatki

Kot že omenjeno, so omejitve glede števila uporabljenih spremenljivk v modelu odpravljene z uporabo faktorjev, ki povzemajo variacijo celotne matrike podatkov. Analiza vplivov naftnih šokov temelji na 135 časovnih vrstah makroekonomskih spremenljivk na mesečni resoluciji, za obdobje od februarja 1986 do decembra 2016. Večina spremenljivk se nanaša na Združene države Amerike, nekaj indikatorjev pa se nanaša na svetovni trg nafte. Ti indikatorji so: cena sodčka surove nafte West Texas Intermediate na promptnem trgu (rpo_t), sprememba svetovne proizvodnje surove nafte v sodčkih na dan ($\Delta prod_t$), ocena svetovnih zalog naftnih derivatov in surove nafte (inv_t), cenilka svetovne gospodarske aktivnosti, katero je definiral Kilian (2009) in temelji na cenah ladijskih prevozov (rea_t). Dodatna spremenljivka je ocena špekulativnega pritiska s finančnega trga surove nafte (sp_t), ocenjena po metodologiji Haase, Zimmermann in Zimmermann (2017), kot delež nakupov v celotnem odprtem interesu nekomercialnih trgovcev na New Yorški borzi surovin. Dodatna posebna spremenljivka, ki jo upoštevam v modelu, je senčna ključna obrestna mera zvezne centrale banke Združenih držav Amerike, katero sta ocenila Wu in Xia (2016). Ta identifikacijska obrestna mera je zanimiva, saj v času krize, ko je dejanska obrestna mera omejena navzdol, kaže, kakšna je ocenjena vrednost in s tem kaže tudi na smer delovanja monetarne politike preko drugih instrumentov.

Vse uporabljene časovne vrste so pretvorjene v stacionarno obliko z uporabo temeljnih transformacij in testirane za nestacionarnost z uporabo Dickey- Fuller testa (Greene, 2012, p. 988-997).

2 Glavni vzroki za nestanovitnost cene nafte

Samo modeliranje naftnega trga in ugotavljanje vplivov na gospodarstvo se je v bolj poglobljeni obliki začelo po ponudbenih naftnih šokih v 1980 letih. Posledično je v začetku raziskovanja naftnih šokov veljalo, da so ponudbeni šoki glavni vir spremenljivosti cen nafte. Pozneje je Kilian (2009) ugotovil, da so cene nafte v največji meri določene s povpraševanjem po nafti. V svoji študiji je šoke povpraševanja po nafti ločil v dve skupini, in sicer na šok, ki izvira iz povečane svetovne gospodarske aktivnosti, in na šok previdnostnega povpraševanja, ki je posledica strahu pred pomanjkanjem nafte v prihodnosti. Kilian (2009) je povsem zanemaril pomen zalog nafte, ki so zelo pomemben del naftnega trga, saj je nafto mogoče skladiščiti. Zaloge nafte so tako lahko blažilec naftnih šokov, hkrati pa predstavljajo šok povpraševanja po nafti, ki ni spodbujen s takojšnjo potrošnjo, torej s povečano gospodarsko aktivnostjo.

Čeprav lahko nenadno povečanje zalog nafte interpretiramo kot špekulativni šok, kot so to predpostavili Kilian in Murphy (2013), Juvenal in Petrella (2015) in Kilian in Lee (2013), pa to gotovo ni edini špekulativni šok na naftnem trgu. Z razvojem finančnih pogodb za ščitenje fizične pozicije na naftnem trgu, se je odprl nov kanal za špekulacije. V zadnjih letih se je znatno povečal delež nekomercialnih trgovcev na New Yorški borzi surovin, kar kaže na povečanje pomena špekulacij s finančnega trga. Parsons in Espinasa (2010) celo trdita, da je bila hitra rast cene nafte med leti 2003 in 2008 posledica špekulativnega mehurčka s finančnega trga.

V model je poleg že omenjenih naftnih šokov vključena tudi senčna ključna obrestna mera ameriške zvezne centralne banke (ang. shadow federal funds rate), ki predstavlja indikator delovanja monetarne politike, ki je relevanten tudi za okolje, ko je obrestna mera na spodnjem ničelnem nivoju, kot sta ta indikator predstavila Wu in Xia (2016). Razlog za vključitev monetarne politike v model naftnih šokov je v odzivnosti monetarne politike, saj le ta deluje kot blažilec vpliva naftnih šokov na gospodarstvo. Nenazadnje pa obstaja tudi možnost povratnega vpliva ameriške monetarne politike na naftni trg preko stroškov financiranja, oportunitetnih stroškov držanja zalog in preko arbitraže z izvedenimi finančnimi inštrumenti, vezanimi na ceno nafte, kot je kanale vpliva opredelil Frankel (2008).

Medtem ko je naftne šoke na ponudbeni strani lažje opaziti in vključiti v model preko dejanske proizvodnje, pa je delitev šokov povpraševanja na vsaj tri pojasnjene vire nekoliko bolj kompleksna. Z uporabo indeksa svetovne gospodarske aktivnosti, katerega sem prevzel po Kilianu (2009), je že mogoče identificirati enega od virov šokov povpraševanja, tako imenovan šok povpraševanja po industrijskih surovinah, vendar pa je preostanek šokov povpraševanja, ki se glede na ugotovitve Kiliana (2009) nanaša na pričakovanja, nekoliko težje deliti glede na izvor. Ena možna delitev se nanaša na del trga, kjer posamezen šok nastane, torej na šok s fizičnega trga nafte in na šok s finančnega trga nafte. Drugi šok povpraševanja je torej mogoče opredeliti kot previdnostno povpraševanje s strani prodajalcev naftnih derivatov in rafinerij, ki si želijo zagotoviti blago, lahko pa gre tudi za povečanje rezerv s strani vlade, z vidika zagotavljanja le-te v primeru pomanjkanja na trgu. Tretji šok se nanaša na špekulativno povpraševanje, ki prihaja s finančnega trga nafte, kjer nekomercialni trgovci sprejemajo tveganje, ker pričakujejo ugodno gibanje cene. Špekulativni šok s finančnega trga in šok povpraševanja po zalogah nafte sta drugačna od drugih šokov, saj odražata vpliv pričakovanj na trgu.

V ekonomertičnem modelu, ki je podrobneje predstavljen spodaj, predstavljam naftni trg na podlagi petih naftnih spremenljivk, pri izbiri katerih sem v grobem sledil Kiliana in Murphyja (2014), Kiliana in Leeja (2014) ter Juvenala in Petrello (2015). Medtem ko so našteti avtorji v modelu uporabili spremembe zalog nafte, v mojem modelu zaloga nafte ostaja kot spremenljivka stoga, saj tudi sam nivo zaloge vpliva na prenos posameznih šokov na gospodarstvo, kot je opisal Medlock III (2013). Izbira spremenljivk torej omogoča identifikacijo štirih specifičnih naftnih šokov: ponudbeni šok, šok v povpraševanju po vseh industrijskih surovinah , šok v povpraševanju po nafti, špekulativni šok s finančnega trga. Dodaten šok, ki tudi vpliva na naftni trg ter na prenos naftnih šokov v gospodarstvo pa je šok monetarne politike, ki je pomemben dejavnik pri modeliranju trga surovin, kot so to izpostavili Frankel (2008) in Aastveit (2014).

3 Regresijski Filter s tremi prehodi

Regresijski filter s tremi prehodi sta predstavila Kelly in Pruitt (2015), kot cenilko faktorjev namenjenih učinkoviti napovedi ene časovne vrste. Navadno so faktorji oziroma poglavitne komponente usmerjeni k zajetju največjega deleža variance, ki obstaja v neki matriki podatkov (**X**). Tako ocenjeni faktorji so torej namenjeni predvsem zmanjšanju dimenzionalnosti napovednega modela. 3PRF model temelji na ideji, da so spremenljivke relevantne za napoved ciljne spremenljivke y striktna podmnožica matrike **X**. Z upoštevanjem ciljne spremenljivke y v samem postopku ocenjevanja faktorja je tako mogoče izločiti irelevantne spremenljivke v matriki **X** (Kelly & Pruitt, 2015, str. 294).

Medtem ko metoda PCA za kombiniranje spremenljivk v faktorje uporablja kovarianco med spremenljivkami v matriki **X**, metoda 3PRF uporablja kovarianco med spremenljivkami v **X** in ciljno spremenljivko *y*. Posledično so vsaj nekateri PCA faktorji irelevantni za napoved *y*, kar pomeni, da je vedno potrebnih več PCA faktorjev za zajetje relevantne informacije, kot pa 3PRF faktorjev (Kelly & Pruitt, 2015, str. 295-296).

Ocena 3PRF poteka v treh korakih in v vsakem koraku se izvede linearna regresija ocenjena z metodo najmanjših kvadratov. Pred samo oceno faktorja je potrebno izbrati proxy spremenljivko (**Z**), torej spremenljivko, h kateri tendirajo faktorji. Za proxy spremenljivke velja, da so v veliki meri določene s faktorji in so relevantne za ciljno spremenljivko, katero napovedujemo (Kelly & Pruitt, 2015, str. 298).

Ocena 3PRF faktorjev poteka v treh korakih, ki so predstavljeni v tabeli spodaj.

Tabela 1: Regresijski Filter s tremi prehodi

Korak	*Opis*
1.	Regresija časovne vrste $\mathbf{x}_i$ na $\mathbf{Z}$ za vse $i = 1...N$, $x_{i,t} = \phi_{0,i} + \mathbf{z}_t'\phi_i + \epsilon_{i,t}$, shrani regresijske koeficiente $\hat{\phi}_i$
2.	Regresija presečnih podatkov $\mathbf{x}_t$ na $\hat{\phi}_i$ za vse $t = 1...T$, $x_{i,t} = \phi_{0,t} + \hat{\phi}_i'\mathbf{F}_t + \eta_{i,t}$, shrani regresijske koeficiente $\hat{\mathbf{F}}_t$
3.	Regresija časovne vrste y_{t+1} na napovedne faktorje $\hat{\mathbf{F}}_t$, $y_{t+1} = \beta_0 + \hat{\mathbf{F}}_t'\beta + \mu_{t+1}$, poda napoved $\hat{y}_{t+1}$

Vir: Kelly & Pruitt (2015, p. 296).

V prvem koraku se torej izvede N (število stolpcev matrike **X**) regresij kjer posamezne časovne vrste v matriki **X**, torej $\mathbf{x}_i$ nastopajo kot odvisne spremenljivke, *proxy* spremenljivka pa nastopa kot regresor. Ocenjeni regresijski koeficienti se shranijo v vektor $\hat{\phi}_i$ in podajajo informacijo o občutljivosti prediktorjev $\mathbf{x}_i$ od *proxy* spremenljivke **Z**.

V drugem koraku izvedemo T (število vrstic matrike $\mathbf{X}$) regresij presečnih podatkov, kjer regresijski koeficienti $\hat{\phi}_i$ nastopajo kot regresorji, posamezne vrstice matrike $\mathbf{X}$ pa nastopajo kot odvisne spremenljivke. Rezultat tega postopka je vektor $\hat{\mathbf{F}}$, ki predstavlja 3PRF faktor. Koeficienti $\hat{\phi}_i$ torej določajo razporeditev prediktorjev na neopazovane faktorje.

V tretjem koraku naredimo napoved ciljne spremenljivke y z uporabo faktorjev $\hat{\mathbf{F}}$. Ta napoved je, glede na karakteristike faktorjev pridobljenih s 3PRF, konsistentna.

4 Metodologija

Za oceno modela za proučevanje vplivov naftnih šokov sem uporabil kombinacijo FAVAR modela in 3PRF, novo metodo za oceno faktorjev. FAVAR model so predstavili Bernanke, Boivin in Eliasz (2005) in v svoji osnovi vključuje faktorje ocenjene z metodo poglavitnih komponent. Vključitev faktorjev je najpomembnejša ločnica med VAR in FAVAR modeli, saj faktorji v model vključijo gospodarske pogoje v državi, kar pomeni, da model vsebuje več informacije, kar v teoriji vodi do nepristranskih ocen koeficientov. Nadaljnji razvoj FAVAR modelov gre predvsem v smer upoštevanja relevantnega dela podatkov, pri čimer pa se metode izbora relevantnih spremenljivk razlikujejo. Uporaba 3PRF faktorjev rešuje prav dilemo o relevantnosti informacije, saj sam postopek pridobitve faktorja zahteva podajanje ciljne spremenljivke, za katero se potem določijo relevantni faktorji. V okolju FAVAR to pomeni, da bo ocena posameznega impulznega odziva zahtevala novo ocenitev modela, s faktorji relevantnimi za to spremenljivko.

$\text{FAVAR}_{\text{3PRF}}$ torej omogoča upoštevanje relevantne informacije iz velike podatkovne matrike. Takšen model vključuje ”spremenljivke šoka”, ki so opazovane in neopazovane faktorje, ki so ”odvisni od odzivne spremenljivke”, saj to spremenljivko eksplicitno modeliramo. Odvisnost faktorjev oziroma modela od spremenljivke odziva na šok pa je tudi največja razlika glede na $\text{FAVAR}_{\text{PCA}}$.

Ocena $\text{FAVAR}_{\text{3PRF}}$ sestoji iz treh korakov. V prvem koraku izberem ciljno spremenljivko, za katero želimo oceniti odziv na šoke. V drugem koraku ocenim R relevantnih faktorjev po postopku 3PRF, natančneje po algoritmu avtomatske izbire *proxy*-jev, kot je podrobneje opisano v nadaljevanju. Potem pa v tretjem koraku uporabim R ocenjenih faktorjev v FAVAR modelu skupaj z drugimi spremenljivkami, kjer faktorje tretiram enako kot druge spremenljivke. Čeprav ocena enega računsko ni zahtevna, pa postane obseg izračunov večji, če upoštevamo, da nas navadno zanima D spremenljivk, ki se odzovejo na šoke, in tako je treba postopek ocenitve izpeljati D-krat.

Pod predpostavko, da lahko naftni trg predstavim z vektorjem C_t, lahko FAVAR model predstavim kot $C_t = A_0 + \Phi(L)C_{t-1} + u_t$, kjer je $\Phi(L)$ končni polinom odlogov, u_t pa napaka modela, za katero predpostavljam, da je neodvisno identično porazdeljna z ničelnim povprečjem. Vektor C_t pa ima obliko $C_t = [\Delta prod_t, inv_t, rea_t, sp_t, rpo_t, F_t', sr_t]'$, kjer F_t' predstavlja vektor R faktorjev in R_t senčno ključno obrestno mero, določeno s strani zveznih rezerv (ang. Federal Reserve).

4.1 Ocena Faktorjev

Ocena faktorjev je narejena na podlagi algoritma avtomatske izbire *proxy*-jev, kjer je izbira spremenljivke, h kateri tendirajo ocenjeni faktorji, trivialna, saj je izbrana spremenljivka kar spremenljivka, katero napovedujemo (y). Ocenjeni faktorji so posledično namensko ocenjeni tako, da iz podatkovne matrike **X** dobijo čim več relevantne informacije za napoved ali pojasnitev y. Kelly in Pruitt (2015) sta dokazala, da ciljna spremenljivka zadostuje vsem kriterijem za *proxy*, prav tako pa pogoje izpolnjujejo tudi residuali, izračunani po napovedi ciljne spremenljivke s 3PRF faktorjem, saj imajo utež na relevantne faktorje različno od nič, hkrati pa so neodvisni od prvega faktorja.

Metodologija izračuna faktorjev na podlagi algoritma avtomatske izbire *proxy*-jev se lahko uporabi za poljubno število faktorjev in je predstavljena v tabeli spodaj, za primer izračuna R faktorjev.

Tabela 2: Algoritem avtomatske izbire proxy-jev

Korak	*Opis*
0.	Začni tako, da velja $r_0 = y$. Za vsak $k = 1, R$:
1.	določi k-ti avtomatski *proxy* na vrednost r_{k-1}. Prenehaj, če valja, da je $k = R$.
2.	Izračunaj 3PRF za ciljno spremenljivko y na podlagi presečnih podatkov **X** z uporabo statističnih *proxy*-jev od 1 do k. Končno napoved označi kot $\hat{y}_k$.
3.	Izračunaj $r_k = y - \hat{y}_k$, povečaj k za 1 in se vrni na korak 1.

Source: Kelly & Pruitt (2015, str. 299).

V tabeli 2 zgoraj je opisan postopek ocene faktorjev, primernih za v FAVAR model. Glede na opisani postopek je mogoče izbrati poljubno število faktorjev, ki pa so medsebojno pravokotni oziroma neodvisni, hkrati pa so relevantni za ciljno spremenljivko.

4.2 Specifikacija modela

Specifikacija modela temelji na osnovnem modelu brez faktorjev, saj je potrebno točno število faktorjev še oceniti. Osnovni model ima obliko $\tilde{C}_t = A_0 + A_1\tilde{C}_{t-1} + ... + A_p\tilde{C}_{t-p} + \tilde{u}_t$, kjer je $\tilde{C}_t = [\Delta prod_t, inv_t, rea_t, sp_t, rpo_t, sr_t]'$.

V prvem koraku specifikacije modela je potrebno določiti **število odlogov**, ki modela ne prilagaja pretirano, hkrati pa prinaša smiselne ocene, saj odlogov ni premalo. Na podlagi Akaikevega informacijskega kriterija in popravljenega Akaikevega kriterija sem se tako odločil za uporabo petih odlogov(Lütkepohl & Krätzig, 2004, str. 100).

V drugem koraku je izveden še bolj pomemben del specifikacije FAVAR modela, in sicer določitev *števila faktorjev*. Pri določitvi števila faktorjev sem sledil pristopu Fornija in Gambettija (2014), ki sta pripravila test za informacijsko zadostnost VAR in FAVAR modelov. Nobena identifikacijska shema namreč ne omogoča pravilne ocenitve impulznih odzivov, če model informacijsko ni zadosten. Alternativen pristop za določitev števila faktorjev sta predlagala Bai in Ng (2002), vendar ta pristop ne upošteva informacije, ki je v modelu že upoštevana, kar pa lahko vodi do velike napake, sploh pri VAR modelih velikega obsega.

Test informacijske zadostnosti modela temelji na ideji, da dodatna informacija iz podatkovne matrike **X** ne more izboljšati napovedne moči VAR modela samo v primeru, ko ta model že vsebuje vso relevantno informacijo, z drugimi besedami, podatkovna matrika **X** Grangerjevo ne povzroča spremenljivk v VAR modelu, kar bi pomenilo, da je model informacijsko zadosten. Če se po drugi strani izkaže, da VAR model ne vsebuje vse relevantne informacije, potem je smiselno razmisliti o FAVAR modelu. Učinkovit test za število faktorjev sta predstavila Galper in Croux (2007), ki sta k problemu testiranja kavzalnosti pristopila s testiranjem Grangerjeve kavzalnosti izven vzorca. Pristop temelji na pomisleku, da faktor F_t Grangerjevo povzroča $\tilde{C}_t$ v primeru, ko v model, ki vsebuje samo pretekle vrednosti $\tilde{C}_t$, doprinese napovedno moč. Vprašanje pa je torej, če je napoved s faktorjem razširjenega modela statistično značilno boljša od osnovnega modela (Gelper & Croux, 2007, str. 3320).

Test informacijske zadostnosti in identifikacija zahtevanega števila faktorjev torej potekata vzporedno. Postopek je sestavljen tako, da se faktorji v model dodajajo tako dolgo, da dodaten faktor ne zmanjša napake napovedi spremenljivk v $\tilde{C}_t$. Po iterativni primerjavi napak napovedi modelov z dodatnimi faktorji in modelov z manj faktorji pa se na podlagi testa verjetnostnega razmerja (ang. likelihood ratio) kar poda oceno o številu faktorjev, ki jih je smiselno vključiti v VAR model (Gelper & Croux, 2007, str. 3323-3325).

Rezultati testa kažejo, da je osnovni VAR model informacijsko nezadosten in ga je potrebno razširiti s faktorji, torej oceniti FAVAR model. Test tudi kaže na to, da sta dva 3PRF faktorja dovolj za dosego informacijske zadostnosti, prav tako pa zadostujeta tudi dva PCA faktorja. Pomembna razlika pa je, da se v primeru PCA faktorjev tudi četrti faktor izkazuje kot šibko statistično značilen, česar ni opaziti pri 3PRF faktorjih, kjer test deluje bolje in lažje določi potrebno število faktorjev.

4.3 Strukturna analiza in test robustnosti

Po uspešni specifikaciji modela se lahko analizirajo dinamične povezave med spremenljivkami. Primer takšne analize je analiza impulznih odzivov, kjer opazujemo vpliv šoka ene spremenljivke v sistemu na vse druge spremenljivke. Rezultati VAR in FAVAR modelov v osnovi nimajo ekonomske interpretacije, temveč zgolj povzemajo dinamiko podatkov, zaradi tega pa so podvrženi Lukasovi kritiki (Lütkepuhl & Krätzig, 2004, str. 159).

Za analizo vplivov posameznih ekonomskih šokov je potrebno VAR in FAVAR modelom predpisati smiselno strukturo, kot so predlagali Sims (1980, 1986), Bernanke (1986) ter Shapiro in Watson (1988). Strukturni modeli se, namesto na spremenljivke same, osredotočajo na reziduale modela, ki so interpretirani kot linearne kombinacije eksogenih šokov. Vpliv spremembe enega od rezidualov se tako rekurzivno preračuna na druge spremenljivke, kar privede do izračunanih impulznih odzivov (Lütkepuhl & Krätzig, 2004, str. 159).

Identifikacija strukturnih šokov je izvedena na podlagi Woldovega sistema (Wold, 1960), kjer se identifikacija doseže z Cholesky dekompozicijo variančno-kovariančne metrike napak modela. S takšno identifikacijsko shemo se doseže, da je model "ravno identificiran" , kar pomeni, da ima ravno toliko omejitev, da je ocena modela mogoča. Kavzalna identifikacijska shema pomeni, da je kavzalnost spremenljivk modela podana z identifikacijo, ne pa iz podatkov, kar pomeni, da mora biti vrstni red spremenljivk v C_t osnovan na teoretičnih temeljih. Če je vrstni red spremenljivk $\tilde{C}_t = [\Delta prod_t, inv_t, rea_t, sp_t, rpo_t, sr_t]'$, to pomeni, da lahko ponudbeni naftni šok $(\delta prod_t)$ še v istem obdobju vpliva na vse ostale spremenljivke, saj je na prvem mestu, po drugi strani pa lahko šok monetarne politike (sr_t) reagira na vse šoke v istem obdobju, hkrati pa na druge spremenljivke vpliva šele v naslednjem obdobju. (Lütkepuhl & Krätzig, 2004, str. 162-163).

Vsi predstavljeni rezultati so preverjeni in primerjani s podobnimi modeli, z namenom preverjanja robustnosti ugotovitev. V enem od preverjanj sem razširil opazovano obdobje, tako da se je začelo že v februarju 1972, hkrati pa sem zaradi nerazpoložljivosti podatkov iz analize izločil spremenljivko špekulativnega pritiska s finančnega trga. Ocenjeni impulzni odzivi so bili zelo podobni, vendar je bil šok s finančnega trga prisoten med drugimi spremenljivkami. Ocenil sem tudi modele z drugačnim redom odlogov, in sicer model z 2, 8 in 13 odlogi. Modeli z različnim številom odlogov se niso bistveno razlikovali od začetnega modela s 5 odlogi, opazne so bile zgolj manjše razlike v konvergenci in teku impulznih odzivov. Preveril sem še rezultate modelov z drugačnim zaporedjem spremenljivk v vektorju C_t, kjer pa sem upošteval zaporedno razvrstitev faktorjev ter razvrstitev spremembe proizvodnje na prvo mesto in senčne obrestne mere na zadnje mesto. Tudi alternativno zaporedje spremenljivk ni bistveno spremenilo rezultatov.

5 Ugotovitve in zaključek

Naftni šoki imajo, kljub razvoju drugih virov energije, še vedno pomemben vpliv na gospodarstvo. Cena nafte lahko gospodarstvu povzroča resne gospodarske težave preko močne povezave s cenami drugih surovin in zaradi direktnega vpliva na transportne stroške. Čeprav je dolgo veljalo, da je vzrok za visoko ceno nafte nepomemben, saj ga lahko gledamo kot eksogeni šok na neko gospodarstvo, pa je Kilian (2009) dokazal nasprotno, in dejstvo, da je vir naftnega šoka izjemnega pomena, je danes v literaturi že splošno sprejeto. Skladno z ugotovitvami drugih znanstvenikov sem v magistrskem delu obravnaval več tipov naftnih šokov:

– ponudbeni naftni šok,
– šok v povpraševanju po industrijskih surovinah,
– šok v povpraševanju po naftnih rezervah,
– špekulativni šok s finančnega trga nafte in
– preostali šok v ceni nafte.

Rezultati analize jasno kažejo na razlike v odzivu gospodarstva na različne naftne šoke, vendar pa so odzivi gospodarstva na šoke povpraševanja bolj podobni.

Tudi odziv vršilca monetarne politike je odvisen od vira šoka. Reakcija monetarne politike na ponudbeni naftni šok je spodbujevalna, medtem ko rezultati kažejo, da monetarna politika praktično ne reagira na šok v povpraševanju po industrijskih surovinah, na šok v povpraševanju po naftnih rezervah ali na špekulativni šok s finančnega trga surove nafte. Odziv monetarne politike je ekspanziven tudi v primeru nepojasnjenega šoka v ceni nafte. Ta šok po vsej verjetnosti odraža neindustrijsko potrošnjo v državah zunaj Združenih držav Amerike in izvajalci monetarne politike z ekspanzivno monetarno politiko ublažijo predvsem negativen vpliv na industrijsko proizvodnjo, tudi če to na drugi strani pomeni višjo stopnjo inflacije.

Pomembno vprašanje je, ali ima tudi monetarna politika v Združenih državah Amerike vpliv na trg nafte. To vprašanje je redko predmet raziskav, izjema pa je Aastveit (2014), ki je v svoji študiji proučil tudi ta učinek. Ugotovil je, da ciljna obrestna mera, določena s strani zvezne centralne banke Združenih držav Amerike, statistično značilno vpliva na naftni trg, predvsem na proizvodnjo nafte. Tudi rezultati modela, predstavljenega v magistrskem delu, so na strani velikega pomena monetarne politike in so podobni vplivom, katere je opredelil Frankel (2008). Restriktivni šok monetarne politike po ocenah povzroči višje cene nafte v prvih nekaj mesecih, potem pa se cena nafte v povprečju zniža za približno 20 % zaradi povečane ponudbe, nižjega povpraševanja in neto kratkega špekulativnega pritiska.

Ocenjen delež, s špekulativnimi pritiski s finančnega trga pojasnjene nestanovitnosti cene nafte, brez upoštevanja vpliva inflacije, znaša približno 10 %. Tak ocenjen delež je precejšen in bi lahko v prihodnosti še naraščal, če se bo nadaljeval trend povečevanja deleža nekomercialnih trgovcev na finančnem trgu surove nafte. V obdobju od leta 2000 pa do leta 2017 so špekulativni pritiski pozitivno vplivali na ceno, predvsem v obdobju od leta 2010 do leta 2015. Sicer pa so rezultati zgodovinske razčlenitve gibanja cene nafte usklajeni z drugimi študijami, ki ugotavljajo, da je povpraševanje po industrijskih surovinah najpomembnejši vir nestanovitnosti cene nafte.

Test informacijske zadostnosti je statistično značilno zavrnil ničelno hipotezo, da je VAR model ustrezen za proučevanje vpliva naftnih šokov. Rezultati testa namreč kažejo na to, da sta dva faktorja dovolj za dosego informacijske zadostnosti VAR modela. Informacijska nezadostnost VAR modela se kaže tudi pri oceni impulznih odzivov, saj so nekatere ocene impulznih odzivov v nasprotju s teorijo, kar pa ne velja za FAVAR modela.

Tudi izbira metode za izračun faktorjev ni povsem nepomembna. Medtem ko so PCA faktorji zasnovani tako, da pojasnijo največji delež variabilnosti v podatkovni matriki, 3PRF faktorji zajamejo samo informacijo, ki je relevantna za ciljno spremenljivko. Rezultati kažejo, da $\text{FAVAR}_{\text{3PRF}}$ učinkovito oceni tudi odzive spremenljivk in šokov, ki ne predstavljajo pomembnega dela variacije podatkovne matrike. Sicer pa so rezultati $\text{FAVAR}_{\text{3PRF}}$ in $\text{FAVAR}_{\text{PCA}}$ zelo podobni.

Čeprav rezultati kažejo, da vpeljava 3PRF faktorjev nekoliko izboljša FAVAR model, ima model težavo, ki se kaže predvsem v računski intenzivnosti ocenjevanja večjega števila modelov (enega za vsako spremenljivko, katere odziv nas zanima). Ta težava je direktna posledica posebne metodologije izračuna faktorjev, ki eksplicitno upošteva spremenljivko, na katero se impulzni odziv nanaša. Po drugi strani so ocenjeni impulzni odzivi teoretično bolj smiselni, hkrati pa lahko $\text{FAVAR}_{\text{3PRF}}$ oceni tudi odzive manjših sektorjev gospodarstva, kar je zelo zaželena lastnost makroekonomskega modela.

Appendix 2: R function: FAVAR Model with 3PRF Factors

```
FAVAR_3PRF <- function(VARvars, response, df, lag = NULL, start,
                       end, nuFac, type="none", FACTOR_loc){
library(vars) # (Pfaff, 2008)
  # VARvars -> data frame of variables included in FAVAR within C
  # example VARvars <- as.data.frame(cbind(OILPRICE,OILPROD,
    # ECONACT, INTRATE,OIL_INVENTORIES, SP))
  # response -> variable of response to the shock
  # df -> stationary data frame from which we identify
    # 3PRF factors
  # lag -> number of lags; if not specified - lag order
    # chosen according to AIC
  # start -> format c(YYYY,MM)
  # end -> format c(YYYY, MM)
  # type -> type of var model c("trend", "constant", "none"),
    # "none" is  the default
  # nuFac -> number of factors
  # FACTOR_loc -> the location of the first factor in the C()
    # to prepare var model for Wold causal chain identification

  r <- response
df_mat <- t(as.matrix(df[,2:length(df)]))
r_vec <- as.vector(r)
index <- which(apply(df_mat, 1, function(x)
  return(all(x == r_vec))))
df_BR <- df
if (length(index) != 0){df_BR <- df[-c(index+1)]}

regcoeff <- matrix(ncol = 1, nrow = (length(df_BR)-1))
colnames(regcoeff) <- c("coeff")

regcoeff_2 <- matrix(ncol = (nuFac +1), nrow = (nrow(df_BR)))

for (m in (1:nuFac)){
  for (i in (2:length(df_BR))){
    OLS_1 <- lm(df_BR[1:nrow(df_BR),i] ~ r, data = df_BR)
    regcoeff[(i-1),1] <- OLS_1$coefficients[2]

  }
  for (j in (1:nrow(df_BR))){
    mdata <- as.numeric(df_BR[j,(2:length(df_BR))])
    OLS_2 <- lm(mdata ~ regcoeff[,1])
```

```
      regcoeff_2[j,1] <- df_BR[j,1]
      regcoeff_2[j,(m+1)] <- OLS_2$coefficients[2]
    }
    OLS_3 <- lm(r ~ as.numeric(regcoeff_2[,(2+m-1)]))
    r <- as.numeric(OLS_3$residuals)
  }
  Proxies <- matrix(ncol = nuFac, nrow=nrow(df))
  for(f in (1:nuFac)){
    assign(paste("P",f, sep = ""),as.numeric(regcoeff_2[,(f+1)]))
    Proxies[,f] <- get(paste("P", f, sep=""))
  }
  Proxies <- as.data.frame(Proxies)

  if(FACTOR_loc == 1){
    tz <- cbind(Proxies,VARvars[,FACTOR_loc:length(VARvars)])}
  if(FACTOR_loc > 1 & FACTOR_loc < length(VARvars)){
    tz <- cbind(VARvars[,1:(FACTOR_loc-1)],
                Proxies,VARvars[,FACTOR_loc:length(VARvars)])}
  if(FACTOR_loc == length(VARvars)){
    tz <- cbind(VARvars[,(FACTOR_loc-1)],Proxies)}

  ind = 0
  tzn <- matrix(ncol = length(names(VARvars))+nuFac, nrow = 1)
  tzn[ind:(FACTOR_loc-1)]=names(VARvars)[ind:(FACTOR_loc-1)]
  ind = ind + FACTOR_loc
  tzn[ind:(nuFac + ind -1)]=as.character(1:nuFac)
  ind = ind + nuFac
  tzn[ind:(length(VARvars)+nuFac)]=
    names(VARvars)[FACTOR_loc:length(VARvars)]
    names(tz) <- tzn

    tz<- ts(tz, start=start,end=end,frequency=12)

    VARselection <- VARselect(tz, lag.max = 20, type=type)
    if(is.null(lag)){lag = as.numeric(VARselection$selection[1])}
    var3 <- VAR(tz, p = lag, type=type)
    return(var3)
  }
```

Appendix 3: Data description

The data set is transformed according to the column TCODE to achieve stationarity. The transformations of x_t are denoted by: (1) for no transformation, (2) Δx_t, (3) $\Delta^2 x_t$, (4) $ln(x_t)$, (5) $\Delta ln(x_t)$, (6) $\Delta^2 ln(x_t)$. The column Variable gives mnemonics, followed by short description. SA stands for seasonal adjustment of time series.

Table 2: Description of the Data Set

ID	TCODE	Variable	Description	SA	Units	Source
			Group 1: Output and income			
1	5	RPI	Real Personal Income	Y	Billions of 2009 Dollars	FRED-MD
2	5	W875RX1	Real personal income excluded transfer receipts	Y	Billions of 2009 Dollars	FRED-MD
3	5	INDPRO	IP Index	Y	Index 2012=100	FRED-MD
4	5	IPFPNSS	IP: Final Products and Nonindustrial Supplies	Y	Index 2012=100	FRED-MD
5	5	IPFINAL	IP: Final Products (Market Group)	Y	Index 2012=100	FRED-MD
6	5	IPCONGD	IP: Consumer Goods	Y	Index 2012=100	FRED-MD
7	5	IPDCONGD	IP: Durable Consumer Goods	Y	Index 2012=100	FRED-MD
8	5	IPNCONGD	IP: Nondurable Consumer Goods	Y	Index 2012=100	FRED-MD
9	5	IPBUSEQ	IP: Business Equipment	Y	Index 2012=100	FRED-MD
10	5	IPMAT	IP: Materials	Y	Index 2012=100	FRED-MD
11	5	IPDMAT	IP: Durable Materials	Y	Index 2012=100	FRED-MD
12	5	IPNMAT	IP: Nondurable Materials	Y	Index 2012=100	FRED-MD
13	5	IPMANSICS	IP: Manufacturing (SIC)	Y	Index 2012=100	FRED-MD
14	5	IPB51222s	IP: Residential Utilities	Y	Index 2012=100	FRED-MD
15	5	IPFUELS	IP: Fuels	Y	Index 2012=100	FRED-MD
16	1	NAPMPI	ISM Manufacturing: Production Index	N	Index	ISM
17	2	CUMFNS	Capacity Utilization: Manufacturing	Y	Percent of capacity	FRED-MD
18	1	ECONACT	Measure of economic activity (Kilian)	N	Index	L. Kilian (2009)
19	5	WCOP	World Crude Oil production	N	million barrels per day	EIA (Monthly Energy Review)
			Group 2: Labor market			
20	5	CLF16OV	Civilian Labor Force	Y	Thousads of Persons	FRED-MD
21	5	CE16OV	Civilian Employment	Y	Thousads of Persons	FRED-MD
22	2	UNRATE	Civilian Unemployment Rate	Y	Percent	FRED-MD
23	2	UEMPMEAN	Average Duration of Unemployment	Y	Weeks	FRED-MD
24	5	UEMPLT5	Civilians Unemployed - Less Than 5 Weeks	Y	Thousads of Persons	FRED-MD
25	5	UEMP5TO14	Civilians Unemployed for 5-14 Weeks	Y	Thousads of Persons	FRED-MD
26	5	UEMP15OV	Civilians Unemployed - 15 Weeks & Over	Y	Thousads of Persons	FRED-MD
27	5	UEMP15T26	Civilians Unemployed for 15-26 Weeks	Y	Thousads of Persons	FRED-MD
28	5	UEMP27OV	Civilians Unemployed for 27 Weeks and Over	Y	Thousads of Persons	FRED-MD
29	5	CLAIMSx	Initial Claims	Y	Number	FRED-MD
30	5	PAYEMS	All Employees: Total nonfarm	Y	Thousads of Persons	FRED-MD
31	5	USGOOD	All Employees: Goods-Producing Industries	Y	Thousads of Persons	FRED-MD
32	5	CES1021000001	All Employees: Mining and Logging: Mining	Y	Thousads of Persons	FRED-MD
33	5	USCONS	All Employees: Construction	Y	Thousads of Persons	FRED-MD
34	5	MANEMP	All Employees: Manufacturing	Y	Thousads of Persons	FRED-MD
35	5	DMANEMP	All Employees: Durable goods	Y	Thousads of Persons	FRED-MD
36	5	NDMANEMP	All Employees: Nondurable goods	Y	Thousads of Persons	FRED-MD
37	5	SRVPRD	All Employees: Service-Providing Industries	Y	Thousads of Persons	FRED-MD
38	5	USTPU	All Employees: Trade, Transportation & Utilities	Y	Thousads of Persons	FRED-MD
39	5	USWTRADE	All Employees: Wholesale Trade	Y	Thousads of Persons	FRED-MD
40	5	USTRADE	All Employees: Retail Trade	Y	Thousads of Persons	FRED-MD
41	5	USFIRE	All Employees: Financial Activities	Y	Thousads of Persons	FRED-MD
42	5	USGOVT	All Employees: Government	Y	Thousads of Persons	FRED-MD
43	1	CES0600000007	Avg Weekly Hours : Goods-Producing	Y	Hours	FRED-MD
44	2	AWOTMAN	Avg Weekly Overtime Hours : Manufacturing	Y	Hours	FRED-MD
45	1	AWHMAN	Avg Weekly Hours : Manufacturing	Y	Hours	FRED-MD
46	1	NAPMEI	Manufacturing: Employment Index	N	Index	ISM
47	6	CES0600000008	Avg Hourly Earnings : Goods-Producing	Y	Dollars per Hour	FRED-MD
48	6	CES2000000008	Avg Hourly Earnings : Construction	Y	Dollars per Hour	FRED-MD
49	6	CES3000000008	Avg Hourly Earnings : Manufacturing	Y	Dollars per Hour	FRED-MD

continues on the next page

Table 2: Description of the Data Set (cont.)

ID	TCODE	Variable	Description	SA	Units	Source
			Group 3: Housing			
50	4	HOUST	Housing Starts: Total New Privately Owned	N	Thousads of Units	FRED-MD
51	4	HOUSTNE	Housing Starts, Northeast	N	Thousads of Units	FRED-MD
52	4	HOUSTMW	Housing Starts, Midwest	N	Thousads of Units	FRED-MD
53	4	HOUSTS	Housing Starts, South	N	Thousads of Units	FRED-MD
54	4	HOUSTW	Housing Starts, West	N	Thousads of Units	FRED-MD
55	4	PERMIT	New Private Housing Permits	Y	Thousads of Units	FRED-MD
56	4	PERMITNE	New Private Housing Permits, Northeast	Y	Thousads of Units	FRED-MD
57	4	PERMITMW	New Private Housing Permits, Midwest	Y	Thousads of Units	FRED-MD
58	4	PERMITS	New Private Housing Permits, South	Y	Thousads of Units	FRED-MD
59	4	PERMITW	New Private Housing Permits, West	Y	Thousads of Units	FRED-MD
			Group 4: Consumption, orders, and inventories			
60	5	DPCERA3M086SBEA	Real personal consumption expenditures	Y	Index 2009=100	FRED-MD
61	1	NAPM	ISM : PMI Composite Index	N	Index	ISM
62	1	NAPMNOI	ISM : New Orders Index	N	Index	ISM
63	1	NAPMSDI	ISM : Supplier Deliveries Index	N	Index	ISM
64	1	NAPMII	ISM : Inventories Index	N	Index	ISM
65	5	AMDMNOx	New Orders for Durable Goods	Y	Milions of Dollars	FRED-MD
66	5	ANDENOx	New Orders for Nondefense Capital Goods	Y	Milions of Dollars	FRED-MD
67	5	AMDMUOx	Unfilled Orders for Durable Goods	Y	Milions of Dollars	FRED-MD
68	5	BUSINVx	Total Business Inventories	Y	Milions of Dollars end of Period	FRED-MD
69	2	ISRATIOx	Total Business: Inventories to Sales Ratio	Y	Ratio	FRED-MD
70	2	UMCSENTx	Consumer Sentiment Index	N	Index 1966: Q1=100	FRED-MD
71	5	CARSALE	Total new cars sale U.S.	Y	Thousads of Units	FRED
72	5	LIGHTTRUCKSALE	Total new light trucks sale U.S.	Y	Thousads of Units	FRED
73	5	HEAVYTRUCKSALE	Total new heavy trucks sale U.S.	Y	Thousads of Units	FRED
74	2	NCLONG	NonComm_Positions_Long_All	N	Millions of contracts	CFTC (COT Report)
75	2	NCSHORT	NonComm_Positions_Short_All	N	Millions of contracts	CFTC (COT Report)
76	2	NCSPREAD	NonComm_Postions_Spread_All	N	Millions of contracts	CFTC (COT Report)
77	2	CLONG	Comm_Positions_Long_All	N	Millions of contracts	CFTC (COT Report)
78	2	CSHORT	Comm_Positions_Short_All	N	Millions of contracts	CFTC (COT Report)
			Group 5: Money and credit			
79	6	M1SL	M1 Money Stock	Y	Billions of Dollars	FRED-MD
80	6	M2SL	M2 Money Stock	Y	Billions of Dollars	FRED-MD
81	5	M2REAL	Real M2 Money Stock	Y	Billions of 1982-84 Dollars	FRED-MD
82	6	AMBSL	St. Louis Adjusted Monetary Base	Y	Billions of Dollars	FRED-MD
83	6	TOTRESNS	Total Reserves of Depository Institutions	N	Millions of Dollars	FRED-MD
84	6	NONBORRES	Reserves Of Depository Institutions	N	Millions of Dollars	FRED-MD
85	6	BUSLOANS	Commercial and Industrial Loans	Y	Billions of Dollars	FRED-MD
86	6	REALLN	Real Estate Loans at All Commercial Banks	Y	Billions of Dollars	FRED-MD
87	6	NONREVSL	Total Nonrevolving Credit	Y	Billions of Dollars	FRED-MD
88	2	CONSPI	Nonrevolving consumer credit to Personal Income	Y	Index	FRED-MD
89	6	MZMSL	MZM Money Stock	Y	Billions of Dollars	FRED-MD
90	6	DTCOLNVHFNM	Consumer Motor Vehicle Loans Outstanding	N	Millions of Dollars	FRED-MD
91	6	DTCTHFNM	Total Consumer Loans and Leases Outstanding	N	Millions of Dollars	FRED-MD
92	6	INVEST	Securities in Bank Credit at All Commercial Banks	Y	Billions of Dollars	FRED-MD

continues on the next page

Table 2: *Description of the Data Set (cont.)*

ID	TCODE	Variable	Description	SA	Units	Source
			Group 6: Interest and exchange rates			
93	2	CP3Mx	3-Month AA Financial Commercial Paper Rate	N	Percent	FRED-MD
94	2	TB3MS	3-Month Treasury Bill:	N	Percent	FRED-MD
95	2	TB6MS	6-Month Treasury Bill:	N	Percent	FRED-MD
96	2	GS1	1-Year Treasury Rate	N	Percent	FRED-MD
97	2	GS5	5-Year Treasury Rate	N	Percent	FRED-MD
98	2	GS10	10-Year Treasury Rate	N	Percent	FRED-MD
99	2	AAA	Moody's Seasoned Aaa Corporate Bond Yield	N	Percent	FRED-MD
100	2	BAA	Moody's Seasoned Baa Corporate Bond Yield	N	Percent	FRED-MD
101	1	COMPAPFFx	3-Month Commercial Paper Minus FEDFUNDS	N	Percent	FRED-MD
102	1	TB3SMFFM	3-Month Treasury C Minus FEDFUNDS	N	Percent	FRED-MD
103	1	TB6SMFFM	6-Month Treasury C Minus FEDFUNDS	N	Percent	FRED-MD
104	1	T1YFFM	1-Year Treasury C Minus FEDFUNDS	N	Percent	FRED-MD
105	1	T5YFFM	5-Year Treasury C Minus FEDFUNDS	N	Percent	FRED-MD
106	1	T10YFFM	10-Year Treasury C Minus FEDFUNDS	N	Percent	FRED-MD
107	1	AAAFFM	Moody's Aaa Corporate Bond Minus FEDFUNDS	N	Percent	FRED-MD
108	1	BAAFFM	Moody's Baa Corporate Bond Minus FEDFUNDS	N	Percent	FRED-MD
109	5	TWEXMMTH	Trade Weighted U.S. Dollar Index: Major Currencies	N	Index Jan 1997=100	FRED-MD
110	5	EXSZUSx	Switzerland / U.S. Foreign Exchange Rate	N	Swiss Francs to One U.S. Dollar	FRED-MD
111	5	EXJPUSx	Japan / U.S. Foreign Exchange Rate	N	Japanese Yen to One U.S. Dollar	FRED-MD
112	5	EXUSUKx	U.S. / U.K. Foreign Exchange Rate	N	U.S. Dollars to One British Pound	FRED-MD
113	5	EXCAUSx	Canada / U.S. Foreign Exchange Rate	N	Canadian Dollars to One U.S. Dollar	FRED-MD
114	1	SHADOWFFR	Wu and Xia (2016) Shadow rate	N	Percent	Wu & Xia (2016)
			Group 7: Prices			
115	5	WPSFD49207	PPI: Finished Goods	Y	Index 1982=100	FRED-MD
116	5	WPSFD49502	PPI: Finished Consumer Goods	Y	Index 1982=100	FRED-MD
117	5	WPSID61	PPI: Intermediate Materials	Y	Index 1982=100	FRED-MD
118	5	WPSID62	PPI: Crude Materials	Y	Index 1982=100	FRED-MD
119	5	PPICMM	PPI: Metals and metal products:	Y	Index 1982=100	FRED-MD
120	1	NAPMPRI	ISM Manufacturing: Prices Index	N	Index	ISM
121	5	CPIAUCSL	CPI : All Items	Y	Index 1982-1984=100	FRED-MD
122	5	CPIAPPSL	CPI : Apparel	Y	Index 1982-1984=100	FRED-MD
123	5	CPITRNSL	CPI : Transportation	Y	Index 1982-1984=100	FRED-MD
124	5	CPIMEDSL	CPI : Medical Care	Y	Index 1982-1984=100	FRED-MD
125	5	CUSR0000SAC	CPI : Commodities	Y	Index 1982-1984=100	FRED-MD
126	5	CUSR0000SAD	CPI : Durables	Y	Index 1982-1984=100	FRED-MD
127	5	CUSR0000SAS	CPI : Services	Y	Index 1982-1984=100	FRED-MD
128	5	CPIULFSL	CPI : All Items Less Food	Y	Index 1982-1984=100	FRED-MD
129	5	CUSR0000SA0L2	CPI : All items less shelter	Y	Index 1982-1984=100	FRED-MD
130	5	CUSR0000SA0L5	CPI : All items less medical care	Y	Index 1982-1984=100	FRED-MD
131	6	PCEPI	Personal Cons. Expend.: Chain Index	Y	Index 2009=100	FRED-MD
132	6	DDURRG3M086SBEA	Personal Cons. Exp: Durable goods	Y	Index 2009=100	FRED-MD
133	6	DNDGRG3M086SBEA	Personal Cons. Exp: Nondurable goods	Y	Index 2009=100	FRED-MD
134	6	DSERRG3M086SBEA	Personal Cons. Exp: Services	Y	Index 2009=100	FRED-MD
135	4	USIRACCO	US Imported Refiner Acquisition Cost of Crude Oil	N	Dollars per Barrel	EIA
136	1	OIL_INV	Ending Stocks of Crude Oil and Petroleum Products	Y	Billion Barrels	EIA
			Group 8: Stock market			
137	5	S&P 500	S&P's Common Stock Price Index: Composite	N	Index	FRED-MD
138	5	S&P: indust	S&P's Common Stock Price Index: Industrials	N	Index	FRED-MD
139	2	S&P div yield	S&P's Composite Common Stock: Dividend Yield	N	Index	FRED-MD
140	5	S&P PE ratio	S&P's Composite Common Stock: Price-Earnings Ratio	N	Index	FRED-MD
141	1	VXOCLSx	VXO	N	Index	FRED-MD

Source: Own work.