Oil and the Stock Market Revisited: A mixed functional VAR approach*

Hilde C. Bjørnland† Yoosoon Chang‡ Jamie L. Cross§

July 31, 2024

Abstract

This paper proposes a new mixed vector autoregression (MVAR) model to examine the relationship between aggregate time series and functional variables in a multivariate setting. The model facilitates a re-examination of the oil-stock price nexus by estimating the effects of demand and supply shocks from the global market for crude oil on the entire distribution of U.S. stock returns since the late 1980s. We show that the MVAR effectively extracts information from the returns distribution that is more relevant for understanding the oil-stock price nexus beyond simply looking at the first few moments. Using novel functional impulse response functions (FIRFs), we find that oil market demand shocks tend to increase returns, while both demand and supply shocks reduce volatility, and have an asymmetric effect on the returns distribution as a whole. In a value-at-risk (VaR) analysis we also find that the oil market contains important information that reduces expected loss, and that the response of VaR to the oil market demand and supply shocks has changed over time.

JEL-Classification: C13, C32, C58, Q41, Q43
Keywords: Oil market, stock market, oil-stock price nexus, functional VAR.

*The authors gratefully acknowledge comments and suggestions from the editor James D. Hamilton, three anonymous referees, Efrem Castelnuovo, Nida Çakır Melek, Haroon Mumtaz, Joon Y. Park, Francesco Ravazzolo, Leif Anders Thorsrud and conference participants at the NBER Summer Institute 2023, the 3rd Dolomiti Macro Meetings 2023, the 30th Annual SNDE Symposium 2023, the Padova Macro Talks 2022 and the CAMP 2022 workshop on Energy and Climate. We are also grateful to Sangmyung Ha for his superb research assistance. This paper is part of the research activities at the Centre for Applied Macroeconomics and Commodity Prices (CAMP) at the BI Norwegian Business School. The views expressed are those of the authors and do not necessarily reflect those of Norges Bank.

†BI Norwegian Business School and Norges Bank
‡Indiana University
§Melbourne Business School, University of Melbourne and BI Norwegian Business School

1 Introduction

A large body of literature on the oil-stock price nexus has shown that oil price shocks are an important driver of stock markets both in the US and around the world.[1] Within this literature, it is common to examine the effects of oil price shocks on stock market returns and volatility in isolation. This is potentially problematic because a large body of research in empirical finance has shown strong evidence of mutual causality between these variables.[2] On the one hand, an increase in the volatility of an equity may induce a *risk premium effect* that results in a positive correlation between returns and volatility. On the other hand, the *leverage effect* associated with negative returns implies that the correlation is negative. This suggests that quantifying the true effects of oil market shocks on the stock market requires us to go beyond an individual moment-by-moment assessment, towards a holistic approach that can account for such distributional effects.

In this paper, we revisit the oil-stock price nexus by investigating the effects of demand and supply shocks from the global crude oil market on the entire distribution of U.S. stock returns. To capture the distributional effects of economic shocks, we propose a new type of functional vector autoregression model that facilitates the joint specification of functional and aggregate time series. This is done by extending the recently proposed functional VAR in Chang et al. (2021b) to also have aggregate time series à la Sims (1980). Since the model utilizes information from both aggregate and functional variables we refer to it as a Mixed Vector Autoregression (MVAR) model.[3]

A key advantage of the MVAR model over conventional VARs is that it facilitates the use of functional impulse response functions (FIRFs). FIRFs enable the researcher to assess how shocks from the aggregate time series impact the distribution of interest, and vice-versa. In our application, this facilitates a novel investigation into how the entire stock return distribution responds to fundamental shocks underlying the global market for crude oil. This feature allows us to move beyond a simple examination of the effects of oil market shocks on average returns and volatility (which are easily extracted from the stock returns

[1] A recent survey paper is provided by Degiannakis et al. (2018).

[2] Seminal contributions include French et al. (1987) and Baillie and DeGennaro (1990).

[3] We call it MVAR instead of the natural MFVAR since it is common to refer to mixed frequency VAR models using the acronym MFVAR.

distribution), towards a complete examination of the oil-stock price nexus.

To make the link between the MVAR and conventional VARs as simple as possible, we start our analysis by comparing the MVAR model with the structural VAR (SVAR) model popularized by Kilian and Park (2009). Building on the seminal work of Kilian (2009), they posit that the global real oil price is driven by three fundamental shocks: (1) shocks to the physical availability of crude oil (*flow-supply shocks*), (2) shocks to the demand for crude oil driven by unanticipated fluctuations in the global business cycle (*flow-demand shocks*), and (3) a residual shock designed to capture all other unanticipated changes in the demand for oil (*oil-specific demand shocks*), such as shifts in the demand for above-ground oil inventories arising from forward-looking behavior, or politically motivated changes in the Strategic Petroleum Reserve.[4] The stock market is represented by a single time series of stock returns. Specified in this manner, their SVAR model identifies demand and supply shocks underlying the global market for crude oil and examines their dynamic effects on stock returns. The model has since been used to study the effects of oil price shocks on US stock market volatility (Bastianin and Manera, 2018), and stock returns in various international markets (Güntner, 2014).

For both the conventional VAR and the MVAR models, the oil market is represented by three series: world oil production, international industrial production, and West Texas Intermediate. In the conventional VAR model, the stock market is represented by a specific moment of the S&P 500 returns distribution, e.g., mean returns (first-moment), volatility (second-moment), or skewness (third-moment). In the MVAR model, the stock market is represented by approximating the entire returns distribution using functional principle components analysis (FPCA). Using this method, we find that almost 98% of the variation in the returns distribution is summarized by three leading functional principle components. We also find that FPCA provides a superior in-sample fit compared to other methods such as quantiles or moments. This suggests that our approach more effectively extracts information from the returns distribution that is more relevant for understanding the oil-stock price nexus than a sole examination of the first few moments of the returns distribution.

After demonstrating the econometric advantages of the MVAR approach, we compare

[4]We highlight that the names of these shocks have changed over the past decade and our choice here is in line with current terminology in the literature (see, e.g. Cross et al. (2022) and references therein).

the associated structural model IRFs and FIRFs following a conventional one standard deviation oil market demand and supply shocks on the US stock market. The main result is that the oil market shocks tend to elicit a persistent positive response in the mean of the stock price distribution while reducing the variance. An important feature of the MVAR approach is that it jointly determines the dynamic responses of the mean and variance of the stock returns distribution to the various oil market shocks as opposed to focusing solely on a given moment. For instance, we also find that the flow-demand shock reduces the kurtosis of the returns distribution in the first few months after the shock, and also has an asymmetric effect on the distribution; eliciting a relatively larger positive response in the returns distribution. Since volatility is a commonly used measure of risk, we also identify the effects of a volatility-maximizing functional shock (referred to as a *Var Max* shock). We find that the Var Max shock significantly decreases returns, while increasing skewness, and kurtosis. These results highlight the importance of our functional approach and controlling for distributional dynamics when studying the oil-stock price nexus.

To highlight the potential policy applications of the MVAR approach we also demonstrate how it can be used to conduct a functional value-at-risk (VaR) analysis. The VaR is defined as the maximum expected loss on an investment, over a given time period at a specified degree of confidence. It is therefore a widely used measure of risk within academia and by private and institutional investors, as well as financial market regulators, see for instance Wiener (1999) for a textbook introduction, or Cabedo and Moya (2003) and Tiwari et al. (2020) for applications of VaR for oil price risk quantifications. Using a functional VaR analysis, our results show that, with the exception of the Great Recession, the estimated VaR from the MVAR model closely tracks the VaR implied by the data distribution. This suggests that the oil market variables contain useful information that may allow portfolio managers to reduce their risks when holding the S&P 500 index. In line with the previous result that demand and supply shocks have very different effects on the returns distribution, we also find that the VaR responds differently to each of the oil market shocks. On average, flow-demand and oil-specific demand shocks have a short-run impact, while flow-supply shocks tend to have no effect. However, there is also evidence of significant time variations in these effects. For instance, the flow supply shock had no impact on the VaR

in 1998:01, but had a significant and long lasting effect in 2018:01. The MVAR model may therefore also be a useful tool for financial regulators and institutional investors who each rely on VaR measures when formulating their decisions.

Our research is related to several strands of literature. From a methodological perspective, our MVAR model contributes towards a recently emerging literature that analyses the distributional effects of macroeconomic shocks (e.g., Chang et al., 2021a; Inoue and Rossi, 2021; Meeks and Monti, 2022; Chang et al., 2022). Much of this literature utilizes existing methods in functional data analysis (see, e.g., Bosq, 2000; Ramsay and Silverman, 2005, 2007; Paparoditis, 2018, and Kokoszka and Reimherr (2017) for a recent non-technical review). For instance, Meeks and Monti (2022) estimate a New Keynesian Phillips curve where the inflation expectations distribution is used as an exogenous covariate using a functional principal component regression. Kokoszka et al. (2018) analyses the cross-section of returns using a functional regression model, thereby complementing the Fama–French approach based on portfolio construction.[5] Chang et al. (2022) take a different approach and consider a functional autoregression where the inflation expectations distribution can be impacted by aggregate economic shocks. Inoue and Rossi (2021) consider a VAR in which aggregate series can be impacted by functional shocks. In contrast to these approaches, our MVAR considers the joint dynamics of functional variables and aggregate time series in response to both functional and conventional shocks. In this sense, it is most similar to Chang et al. (2021a), who specify a nonlinear state-space model to capture the joint dynamics of aggregate macroeconomic time series and the earnings distribution. The main difference between our approach and theirs is that we view the functional variables as observable as opposed to unobservable states. This means that we do not require the use of simulation-based estimation or filtering methods but can instead estimate the model parameters using Ordinary Least Squares (OLS). Doing so, we can identify both aggregate and functional shocks, and produce functional impulse response functions.

From an empirical perspective, our results build on a long line of research that seeks to quantify the relationship between shocks to oil prices and stock markets by focusing on

[5] While the cross-sectional approach to returns is undoubtedly valuable for financial research, our paper's contribution lies in utilizing the MVAR model to uncover deeper insights into the nuanced effects of oil market dynamics on stock market behavior over time.

either stock returns or volatility in isolation (see e.g., Jones and Kaul, 1996; Huang et al., 1996; Sadorsky, 1999; Park and Ratti, 2008; Kilian and Park, 2009; Bjørnland, 2009; Miller and Ratti, 2009; Güntner, 2014; Degiannakis et al., 2014; Alsalman and Herrera, 2015; Boldanov et al., 2016; Bastianin and Manera, 2018; Thorbecke, 2019, among many others), by detailing the importance of controlling for distributional dynamics. We also find that the MVAR model reproduces a well-known result in the empirical oil market literature that oil market demand and supply shocks have very different effects on the real price of oil (Kilian, 2009).

Empirically, we also relate to several papers that have examined financial risk related to a portfolio using a measure such as VaR, c.f. Cabedo and Moya (2003) and Tiwari et al. (2020) among others. We contribute by providing a novel functional approach to quantify downside risks using VaRs computed from the entire stock return distribution, and by examining the effects of oil market shocks on VaR (as opposed to aggregate price shocks).

While we have focused on the oil-stock price nexus in this paper, there is clearly a wide scope of possible avenues for future research with the MVAR approach, including direct extensions examining the effects of oil price shocks on equity prices across sectors, or other areas at the frontier of research in macroeconomics and finance such as drivers of income inequality, the effects of shocks on the distribution of firms' revenue, and understanding the distributional dynamics of climate change.

The rest of the paper is structured as follows. In Section 2 we present the data and econometric methodology. Section 3 contains the results. Finally, we conclude in Section 4.

2 The Data and Econometric Methodology

In this section, we discuss the data, provide a short recap of VAR models of aggregate variables currently used in the literature, formally introduce the MVAR model, explain how to estimate the MVAR using an approximate VAR representation and show how to identify the underlying structural shocks and construct functional impulse response functions.

2.1 Data

Our empirical analysis requires data that is representative of the global market for crude oil and the U.S. stock market from January 1988 to January 2018. We will show below, that by applying a robust structural identification strategy, we can disentangle the distinct effects of oil market shocks from those influences stemming from stock price fluctuations, effectively addressing potential confounding issues. In particular, building on paper such as Kilian (2009), Kilian and Park (2009) and Baumeister and Hamilton (2019), the global oil market is represented using time series on global oil production (WOP), economic activity (IPI), and the real price of oil (WTI). All three series are measured at a monthly frequency. For global oil production, we use the percentage change of the global crude oil production data made available from the U.S. Department of Energy. For real economic activity, we use the percentage change of OECD+6 industrial production (IP) index as proposed in Baumeister and Hamilton (2019) and made available from Christiane Baumeister's personal website.[6] Finally, the monthly real price of oil is given by the West Texas Intermediate deflated by the U.S. consumer price index (CPI). The oil price series is calculated as monthly averages from from daily data by the U.S. Energy Information Agency (EIA), by taking an unweighted average of the daily closing spot prices over the specified time period. Both WTI and CPI are available from the Federal Reserve Economic Data (FRED) online database. As mentioned above, both global oil production and the real economic activity are differenced to be stationary in the model. The real oil price, however, is included in levels, in line with previous literature. Although it is more persistent than the growth rates of the other variables, it is still stationary over a long sample (results can be obtained at request). More importantly, the VAR model itself is stable.[7] Finally, note also that all series are demeaned prior to estimation. The transformed time series are shown in Figure 1.

The US stock market is represented using intra-month distributions of stock returns of the S&P 500 index. For each month, we use capitalization-weighted S&P 500 index returns at the ten-minute frequency to estimate densities for the intra-month distributions.[8] The

[6] Website link: https://sites.google.com/site/cjsbaumeister/research

[7] Alternatively, we could have included the non-stationary variables in a vector error correction model, see for instance Chang et al. (2023) for a different application.

[8] The density estimates are obtained using the normal kernel with the most commonly used bandwidth given by $h = 1.06\hat{\sigma}n^{-1/5}$, where n and $\hat{\sigma}$ are the number and the sample variance, respectively, of ten-minute

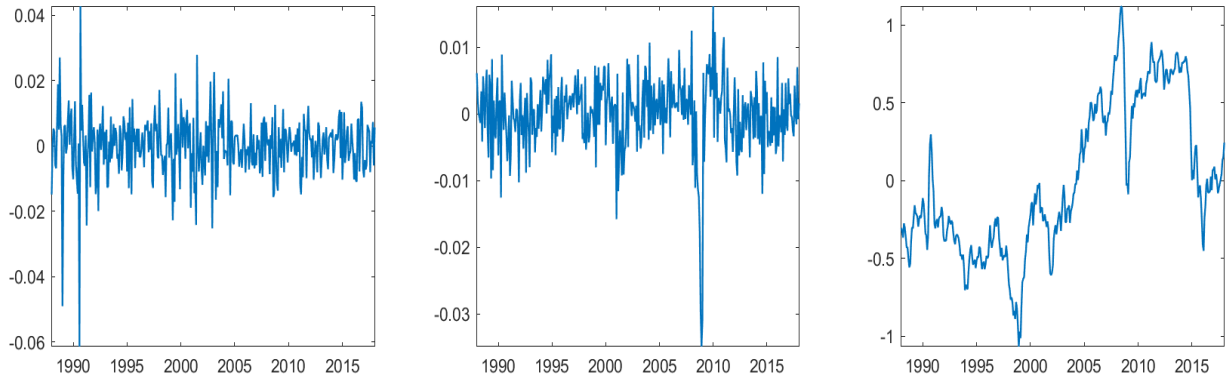


Figure 1: Transformed oil market time series: global oil production (left), global industrial production (center) and real price of crude oil (right) over the period 1988:02-2018:01. See the main text for details about the transformation.

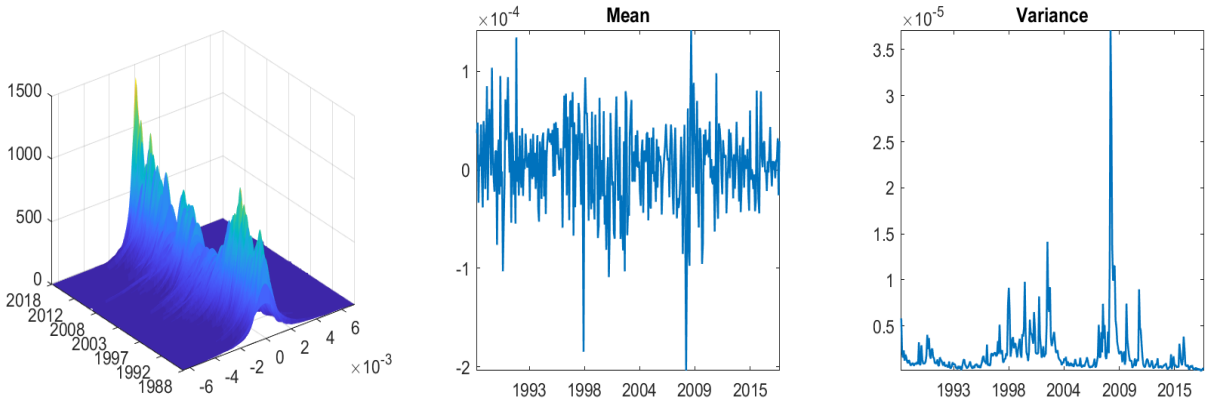


Figure 2: Kernel estimates of stock return densities obtained using high-frequency returns and associated sample mean and variance over the period 1988:02-2018:01.

returns data are obtained from Tick Data Inc and are truncated at 0.50% and 99.5% percentiles prior to estimating the state densities. Figure 2 contains the time series of the estimated densities for intra-month distributions (left panel), along with the associated sample mean (center panel) and sample variance (right panel).[9] It is clear that the moments of the distribution vary across time. This suggests that it may be necessary to account for the entire stock price distribution in order to fully understand the oil-stock price nexus.

returns available for a given month.

[9]When estimating our model we use the demeaned density of the stock return distribution in Figure 2.

2.2 SVAR model

In order to make the links between the MVAR and conventional SVARs as transparent as possible, we first provide a short recap of the foundational SVAR model currently used in the literature. Let $z_t = (x_t, r_t)'$, in which $x_t = (q_t, y_t, p_t)$ where q_t denotes global oil production, y_t global economic activity, p_t the real price of oil and r_t an aggregate time series representation of the stock market such as returns or volatility. The *reduced form* VAR model is given by

$$z_t = \sum_{i=1}^{p} A_i z_{t-i} + \varepsilon_t, \tag{1}$$

where A_i are matrices of autoregressive coefficients, z_{t-i} are vectors of lags, $i = 1, \ldots, p$, and ε_t is a white noise error with mean zero and variance Σ. For simplicity, we exclude the constant term. The VAR model links the oil and stock markets via two blocks of equations. The first block is a global crude oil market model which consists of the first three equations. The second block is a single series that represents the stock market.

It is well known that the structural representation of (1) is underidentified. To see this let e_t denote the vector of serially and mutually uncorrelated structural innovations. It then follows that $\varepsilon_t = Be_t$, where B is the structural impact matrix and that $\Sigma = BB'$. It is easily verified that B contains K^2 unique elements, while the covariance matrix Σ only contains $K(K+1)/2$ unique elements. Exact identification is therefore achieved by restricting $K(K-1)/2$ elements to be zero.

Following Kilian and Park (2009) the identification problem can be solved by assuming that the structural impact matrix has the following recursive structure

$$B = \begin{pmatrix} b_{11} & 0 & 0 & 0 \\ b_{21} & b_{22} & 0 & 0 \\ b_{31} & b_{32} & b_{33} & 0 \\ b_{41} & b_{42} & b_{43} & b_{44} \end{pmatrix}. \tag{2}$$

The exclusion restrictions in the oil market block are consistent with a vertical short-run supply curve and a downward sloping demand curve, as introduced in Kilian (2009).[10]

[10] Since the seminal paper of Hamilton (1983), a growing body of literature has suggested to use SVAR models to identify the effects of oil market shocks on aggregate activity, see for instance Burbidge and

These assumptions are motivated as follows: (1) crude oil supply is slow to adjust due to production costs and uncertainty about the current and future state of the economy, (2) global real economic activity is sluggish and shocks to the real price of oil will take time to propagate, (3) shocks to factors other than supply or aggregate demand for industrial commodities must be demand shocks that are specific to the oil market. This final shock includes factors such as financial speculation, precautionary demand associated with uncertainty surrounding the real price of oil, changes to storage technology, or politically motivated changes in the Strategic Petroleum Reserve. Following the literature, we refer to these three oil market shocks as *flow-supply* (FS) shocks, *flow-demand* (FD) shocks, and *oil-specific demand* (OSD) shocks respectively. The final assumption asserts that the US stock market responds to all three oil market shocks on impact, while only affecting the oil market variables with a delay of at least one month. This assumption is consistent with the notion that innovations to the price of oil are predetermined with respect to the US economy. This set of assumptions has also been used by Bastianin and Manera (2018) to study the effects of demand and supply shocks from the global oil market on US stock market volatility, and by Güntner (2014) to examine the effects of the oil market shocks on returns in various international markets.

2.3 Mixed Vector Autoregression (MVAR) model

Now, let $z_t = (x_t, f_t)$ in which x_t denotes an ℓ-dimensional vector of aggregate variables, and f_t denote a functional variable taking values in a Hilbert space. In our application x_t is the same 3-dimensional vector of oil market variables discussed in the previous section, while f_t is the stock return density. Using a similar notation to the previous section, the Mixed Autoregression (MVAR) model is given by

$$z_t = \sum_{i=1}^{p} A_i z_{t-i} + \varepsilon_t, \tag{3}$$

however in this case A_i are bounded linear operators (as opposed to matrices), and ε_t is a functional (as opposed to a real-valued vector) error, which is assumed to be a functional

Harrison (1984), Ahmed et al. (1988), Bernanke et al. (1997), Bjørnland (2000), Hamilton (2003), Hamilton (2009) and Kilian (2009) for some initial studies using different types of identifying restrictions.

white noise with zero mean.

A constant term is not included in our presentation of MVAR in (3), as in VAR in (1), to avoid a lengthy discussion required to define it rigorously. Instead, we follow a common convention and use the demeaned version of the aggregate and functional variables introduced above for z_t in our analysis. Under the convention, in particular, our functional variable f_t is given not as a density of stock returns, but as a demeaned density of stock returns which has zero integral. This is important to note, since the set of densities itself cannot be defined properly as a Hilbert space. Throughout the paper, we let H be a Hilbert space given by

$$H = \left\{ f : \mathbb{R} \to \mathbb{R} \middle| \int f(r)dr = 0, \int f(r)^2 dr < \infty \right\}, \tag{4}$$

which is a proper Hilbert space as a subspace of the Hilbert space of square integrable functions.[11]

The main difference between the MVAR and VAR is the introduction of the functional variable f_t taking values in H. This implies that z_t in (3) is defined as a time series of random elements in $R^\ell \times H$, which itself is defined as a Hilbert space endowed with the inner product $\langle \cdot, \cdot \rangle$ and the norm $\| \cdot \|$.[12] This is in contrast to (1) in which z_t is defined as a real-valued vector in $R^{(\ell+1)}$. This difference creates a key computational challenge in how to estimate the MVAR model. In what follows we first show that the MVAR model can be approximated using an appropriately specified finite-dimensional VAR model. We then discuss how to identify shocks for structural analysis. Finally we present three key practical challenges that users need to resolve when specifying MVAR models to address their research questions.

2.3.1 Reduced form approximation and estimation

In this section, we show how to approximate the MVAR using a finite dimension VAR model, and then estimate it using standard ordinary least squares (OLS) methods. Our approach builds on recent results in Chang et al. (2021b) — who show that functional VARs (without aggregate time series) can be approximated using finite VAR models — to

[11] Note that any bounded density is square-integrable, and therefore, the assumption of square-integrability is very mild.

[12] A more formal definition of the MVAR model is given in Appendix A.

a framework that includes both functional variables and aggregate time series.[13] The key idea is to fix an orthonormal basis of $R^\ell \times H$ such that the approximation error term goes to zero at an appropriate rate. For notational simplicity we here work with a first-order MVAR since the theory easily extends to the case of arbitrary lag length.[14]

Approximation The MVAR can be approximated with a finite-dimensional VAR by fixing an orthonormal basis (v_i) of $R^\ell \times H$. Let $R^\ell \times V$ be the subspace of $R^\ell \times H$ spanned by a sub-basis $(v_i)_{i=1}^n$ with $n > \ell$, and denote by P the Hilbert space projection on the subspace $R^\ell \times V$ so that $Pz = \sum_{i=1}^n \langle v_i, z\rangle v_i$. The MVAR in (3) can then be approximated as

$$z_t = APz_{t-1} + A(1-P)z_{t-1} + \varepsilon_t \approx APz_{t-1} + \varepsilon_t, \tag{5}$$

where $1-P$ is an operator on $R^\ell \times H$ defined as $(1-P)z = z - Pz = \sum_{i=n+1}^\infty \langle v_i, z\rangle v_i$ for all $z \in R^\ell \times H$. Note that the approximation error term $\big(A(1-P)z_{t-1}\big)$ is asymptotically negligible under suitable regularity conditions if we set $n \to \infty$ as $T \to \infty$ at an appropriate rate.

Next, to represent this approximate MAR as a finite-dimensional VAR model, let

$$\pi : z \mapsto (z) = \begin{pmatrix} \langle v_1, z\rangle \\ \vdots \\ \langle v_n, z\rangle \end{pmatrix} \tag{6}$$

for $z \in R^\ell \times H$, and

$$\pi : A \mapsto (A) = \begin{pmatrix} \langle v_1, Av_1\rangle & \cdots & \langle v_1, Av_n\rangle \\ \vdots & \vdots & \vdots \\ \langle v_n, Av_1\rangle & \cdots & \langle v_n, Av_n\rangle \end{pmatrix} \tag{7}$$

[13]Since our MVAR extends the functional model in Chang et al. (2021b) by adding aggregate time series, the conditions for consistency of our estimators and bootstrap confidence bands are all extended straightforwardly to those required in our paper. Our method allows us to compare the functional approach with the conventional approach as shown in Appendix B, as well as to use the standard VAR mthodology for MVARs.

[14]This is analogous to the fact that any VAR(p) can be written as a VAR(1), thus allowing theoretical derivations involving VAR models to be done with the VAR(1) process.

for $A \in L(R^\ell \times H)$. Then, using our notation in (6) and (7), we may readily represent the approximate MVAR in (3) as

$$(z_t) \approx (A)(z_{t-1}) + (\varepsilon_t), \tag{8}$$

a conventional n-dimensional VAR. Our empirical analysis and statistical inference in the paper are all based on the approximate VAR in (8).

Functional Principle Components Analysis We emphasize that the VAR representation in (8) may be obtained for any choice of an orthonormal basis (v_i) of $R^\ell \times H$. In this paper we make use of the standard basis for R^ℓ and the functional principle components basis for H.[15] In Appendix B we show that this basis possesses a theoretical optimality property that makes it a natural choice.

To introduce the basis (v_i^*) we use in our analysis explicitly, we formally define the underlying Hilbert space $R^\ell \times H$ as $R^\ell \oplus H$. Let

$$v_i^* = e_i$$

for $i = 1, \ldots, \ell$, where (e_i) is the standard basis for R^ℓ, so that $x_t = \big(\langle v_1^*, z_t\rangle, \ldots, \langle v_\ell^*, z_t\rangle\big)'$ for all t. Furthermore, we define

$$v_{i+\ell}^* = e_i(\Gamma)$$

for $i = 1, 2, \ldots$, where $e_i(\Gamma)$ denotes the eigenfunction of the sample variance operator of (f_t)

$$\Gamma = \frac{1}{T}\sum_{t=1}^{T}(f_t \otimes f_t) \tag{9}$$

associated with its i-th largest eigenvalue.[16] Note that $\big(e_i(\Gamma)\big)$ is the (normalized) functional principal component, which is widely used in the functional data analysis literature including Bosq (2000), Ramsay and Silverman (2005), Hall and Horowitz (2007) and Park

[15] Since functional principle components analysis consists of finding the eigenfunctions of the sample functional autocovariance operator, it is also commonly referred to as an eigenbasis.

[16] If a function is represented as a long vector, as is usually done in functional data analysis, the FPCA reduces to the principal component analysis (PCA). In our analysis, we represent the estimated densities of stock returns as 1024-dimensional vectors to obtain FPCs. The dimension 1024 is non-consequential and does not affect our results unless it is set to be too small.

and Qian (2012), among many others.

Estimation Having approximated the infinite-dimensional MVAR with a finite-dimensional VAR, we now derive the estimator of the autoregressive operator $\widehat{A}$ of A and the fitted values reduced form errors $\widehat{\varepsilon}_t$ of the reduced form errors ε_t.[17] Since the restricted versions of π's in (6) and (7) are one-to-one, their inverses exist and are well defined and given by[18]

$$\pi^{-1}\big((z)\big) = Pz \quad \text{and} \quad \pi^{-1}\big((A)\big) = PAP.$$

Let $\widehat{(A)}$ be the OLS estimator of the autoregressive coefficient matrix (A) and $\widehat{(\varepsilon_t)}$ be the fitted values of the residuals (ε_t) obtained from the approximate VAR in (8). This implies that the estimate $\widehat{(A)}$ of the autoregressive coefficient matrix and the fitted residuals $\widehat{(\varepsilon_t)}$, obtained from the approximate VAR in (8), allow us to derive the associated estimate $\widehat{A}$ and the fitted functional residuals $\widehat{\varepsilon}_t$ as

$$\widehat{A} = \pi^{-1}\big(\widehat{(A)}\big) \quad \text{and} \quad \widehat{\varepsilon}_t = \pi^{-1}\big(\widehat{(\varepsilon_t)}\big),$$

where $\widehat{A}$ is a linear operator on $R^\ell \times V$ and $\widehat{\varepsilon}_t$ is a functional time series taking values in $R^\ell \times V$.

2.3.2 Structural analysis

Having shown how to estimate the reduced form MVAR model, we now show how to identify the underlying structural shocks and construct functional impulse response functions.

Identifying an aggregate shock Once the functional variable f_t in (3) is approximated using an m-dimensional basis, z_t becomes a $n \times 1$ vector, where $n = \ell + m$, standard identification strategy from the conventional SVAR literature can be applied to the approximate

[17] Our VAR in (8) may be regarded as a FAVAR if FPCs are defined as factors. However, even in this case, the factors in the VAR formulation of our approximated MVAR are very different from the usual factors constructed from a panel and used in a FAVAR. The factors in our VAR are smooth functions taking real values over a completely ordered subset of R.

[18] Note that although the π's in (6) and (7) are defined for any z in $R^\ell \times H$ and for any linear operator in $R^\ell \times H$, we can interpret them as their restricted versions on $R^\ell \times V$ whenever necessary.

VAR in (8).[19] To see this, first note that the VAR representation of the MVAR model in (8) will have a covariance matrix of size n. Identifying the structural shocks therefore involves recovering the structural impact matrix, B, from the relationship

$$(\varepsilon_t) = Be_t, \tag{10}$$

where $e_t = (e_{1t}, \ldots, e_{nt})'$ is the $n \times 1$ vector of structural innovations. Each column of B contains contemporaneous responses of the ℓ-dimensional vector of aggregate variables and the m-dimensional vector of loadings on the m functional principal components.

To remain close to the key literature discussed in Section 2.2, we identify B by positing a recursive relationship between the variables of interest. Under this assumption, the structural impact matrix B is obtained uniquely from

$$BB' = \widehat{(\Sigma)} = \frac{1}{T}\sum_{t=1}^{T} \widehat{(\varepsilon_t)}\widehat{(\varepsilon_t)}', \tag{11}$$

where $\widehat{(\varepsilon_t)}$, $t = 1, \ldots, T$, are the fitted residuals from the reduced form VAR in (8).

Identifying a functional shock The above procedure shows that the ℓ structural shocks associated with the aggregate time series, i.e., $e_{1t}, e_{2t}, \ldots, e_{\ell t}$, can be identified as in the standard VAR in the literature discussed in Section 2.2. In particular, we will identify the three shocks related to the oil market (*flow-supply* (FS) shocks, *flow-demand* (FD) shocks, and *oil-specific demand* (OSD) shocks) as above. However, the functional shocks, $e_{(\ell+1)t}, \ldots, e_{(\ell+m)t}$, are not meaningfully identified in any structural sense. We therefore refer to these shocks as *semi-structural* and seek to identify a single economically interpretable functional shock. In principle, we may identify any functional structural shock as long as it can be defined as a linear combination of these semi-structural shocks

$$e_t^f = w_1 e_{(\ell+1)t} + \cdots + w_m e_{(\ell+m)t}, \tag{12}$$

[19] In our application the oil market block consists of three aggregate time series implying $\ell = 3$.

where denoted e_t^f denotes the functional shock of interest, and $w_1, \ldots, w_m$ denote the weights with normalization restriction $\sum_{i=1}^m w_i^2 = 1$ imposed to make e_t^f have unit variance. To identify any functional shock of interest, we should therefore find the weights that define it as a linear combination of the semi-structural shocks. We return to this when discussing practical considerations in the next sub-section.

Functional IRFs Impulse response functions of $(z_t) = (x_t', (f_t)')'$ to any of the structural and semi-structural shocks $e_{it}, k = 1, \ldots, \ell + m$, identified by (10) and (11), can be easily obtained from the approximate VAR in (8) as in the standard SVAR. Computing the responses of the aggregate variable x_t requires no extra work. For the functional variable f_t, however, we need to convert the responses of (f_t) to those of the original functional variable f_t, since the goal of our analysis is to assess the impact of structural shocks on f_t, not on (f_t). It is indeed simple and straightforward to get the responses of f_t, once we obtain the responses of (f_t), to these structural and semi-structural shocks. To recover impulse responses of the original functional variable f_t from those of (f_t) to each structural and semi-structural shock, we may just reverse the mapping implied by the isometry π in Section 2.3.1. Since π is a one-to-one mapping from an m-dimensional subspace V of the function space H to R^m, the response of f_t to each of structural and semi-structural shocks aggregate shock e_{kt} is given by $\pi^{-1}\big(r(e_{kt})\big)$, where $r(e_{kt})$ is the impulse response function of (f_t) to the structural or semi-structural shock e_{kt}, $k = 1, \ldots, \ell + m$.

Note that $\pi^{-1}\big(r(e_{kt})\big)$ is a function in the subspace V of the original function space H. Visually, this results in a response of the three-dimensional response surface of the stock return distribution to each aggregate structural shock, as opposed to the usual two-dimensional lines provided by conventional impulse response functions.[20] Since these new responses map the structural shocks to the functional variable of interest, we refer to this tool as a functional impulse response function (FIRF).

[20] A graphical presentation of these functional responses is given in our application reported in Figure 6 in the following section. The three panels in the bottom row show the functional responses of the stock return distribution to the three aggregate shocks at impact and over the next twelve month horizons.

2.4 Practical Considerations

When using the MVAR model the researcher must consider three key practical considerations: (1) how to construct the functional variable? (2) which basis to choose for the reduced form estimation? (3) which aggregate functional shock to identify? We now discuss each of these considerations in the context of our application.

Constructing Functional Variables The first challenge that researchers must overcome when estimating such models is that the functional variable may not be continuously observed at the same frequency as the aggregate time series of interest. In our application this is overcome by constructing a series of estimated stock return densities that are observable at the same monthly frequency as the oil market variables (see Section 2.1). In our subsequent analysis, the errors incurred by using estimated densities, rather than true densities, of stock returns are ignored. This seems to be reasonable, given the large number of high-frequency observations we use to estimate stock return densities.

The density of stock returns is used as a functional variable in our model.[21] It is possible to use a transformation of density, instead of density itself, such as log-density in Chang et al. (2021a), distribution function in Horta and Ziegelmann (2018), and the log quantile density transformation in Petersen and Müller (2016) and Petersen et al. (2022), among many others.[22] The main reason to use one of these transformations is to avoid the two restrictions density must satisfy: the unit-integral and non-negativity restrictions. In our analysis based on demeaned-densities, these restrictions matter only when we reconstruct a counterfactual stock return density by superimposing a fitted demeaned-density on any base density. The unit-integral restriction is satisfied for any stock return density obtained from arbitrary linear combination of approximated demeaned-densities because the FPC basis consists of zero-integral functions. The non-negative restriction, on the other hand, is generally violated.[23] This, however, does not cause any problem in interpreting our results

[21] See Kneip and Utikal (2001) for the use of density in performing inference on density families using FPCA.

[22] All the papers cited here use the standard Hilbert space norm. Note, however, that they use different measures of the distance between two distributions depending upon their choice of transformation. In contrast, Tsay (2016) uses the Hellinger distance for his autoregressive model of density functions.

[23] If log-densities are used, the unit-integral restriction is generally violated though the non-negativity restriction is always satisfied. The use of distribution functions generally yields functions that are not

which are focused on the *change* of the stock return density in response to a structural shock. We therefore use the stock return density directly as a functional variable in our model.[24]

Selecting a Basis The second challenge pertains to the choice of an appropriate basis to estimate the VAR in (8). Basis functions can be broadly categorized as either *a priori* fixed bases or data-driven bases. As the name suggests, *a priori* fixed bases are chosen independently of the data. We instead take the second route and seek an optimal (in the least squares sense) empirical orthonormal basis that is data-driven. While a fixed (spline) basis is used in the analysis of Chang et al. (2021a), data-driven bases have been used in other recent macroeconomic analyses involving the distribution of inflation expectations (Meeks and Monti, 2022; Chang et al., 2022). In Appendix B we show that the data-driven functional principal component (FPC) basis possesses a theoretical optimality property that makes it a natural choice. However, in the spirit of being completely data-driven, we also compare the in-sample fit of the FPC basis with three alternative basis functions: the moment basis, the quantile basis, and the histogram basis, in terms of functional R-squared, integrated variance, and trace statistics. To conserve space, we defer the details of this procedure to Appendix B.

The main result is that we find the FPC basis to be the best way to summarize the temporal variation of the returns distribution, and we use this as our basis throughout the paper. Letting the number of functional principal component series be denoted by m, then (z_t) in (3) becomes an $(\ell + m) \times 1$ vector with $\ell = 3$ oil market variables. Given these inputs, the MVAR representation in (8) can then be estimated using standard OLS methods. Results for optimal selection of m are provided in Section 3.1.

Identifying a Functional Shock The final consideration pertains to the identification of the function shock. As discussed previously, we may identify any functional structural shock as a linear combination of the underlying semi-structural shocks. Various functional shocks are of practical interest. In the context of the oil-stock price nexus, for instance,

proper distribution functions.

[24] As shown in Appendix B, the use of density has an important advantage of making it possible to compare the results from our MAR model with those from more conventional models.

a financial portfolio manager may be interested in analyzing the impacts of shocks that minimize the value-at-risk at a given horizon. An oil market analyst may alternatively prefer to examine shocks that maximize the cumulative response of one of the aggregate variables (world economic activities, for example) some years after the impact date.

Since volatility is a commonly used measure of risk, we here identify a volatility-maximizing functional shock. This is done by finding the linear combination of m semi-structural functional shocks so that the resulting combined functional shock has the largest positive effect on stock return volatility at impact. We refer to this shock as a *Var Max shock* and defer the details to Appendix C.

3 Results

The results are presented across four subsections. First, we investigate how to best represent the stock returns density using functional principal component analysis (FPCA). Second, we examine the dynamic effects of the various oil market demand and supply shocks on the stock market using functional impulse response functions. Third, we quantifies the distributional effects of the aggregate and functional shocks on the stock return distribution. This is illustrated by investigating histograms with responses at deciles to the various shocks. Fourth, we present a value-at-risk (VaR) analysis to demonstrate the potential policy applications of the MVAR approach. The VaR is a widely used measure of risk within academia and by private and institutional investors, as well as financial market regulators.

3.1 Functional principal component analysis

To select an appropriate number of components to summarize the S&P 500 returns density we perform functional principal component analysis (FPCA). The number of components used for the analysis is selected using the cumulative scree plot in Figure 3. The results show that the first three functional principal components (FPCs): v_k, $k = 1, 2, 3$, can explain around 98 percent of the variance in the S&P 500 returns distribution, with subsequent FPCs providing marginal contributions. We therefore set the number of FPCs equal to

three in our analysis, but analyse robustness to chosing two FPCs in Appendix D.[25]

Figure 4 presents the results from applying the FPCA to the stock return distributions. The figure displays the first three functional principle components (FPCs) (left), along with the time series of their respective loadings (center) and mean distributional dynamics (right).[26] Defined as linear combinations of f_t, the FPCs are functions in the Hilbert space H introduced in (4), and therefore, they are all integrated to zero. The gray shaded areas in the middle column indicate the NBER recession periods.

We note in general that the three FPCs, presented in the middle column, show salient time-invariant features of the stock return distributions reflecting financial market conditions, while their loadings, presented in the left column, inform the direction and magnitude of their effects on the distribution over time. Note that FPCs are functions defined over the range of stock returns while their loadings evolve over time. In particular, we find that the first FPC, displayed in the top row, places the majority of the mass on the center of the distribution, and reduces the mass in the tails (see the top left panel). It can therefore be interpreted as a *central tendecy factor*. As shown in Figure 3, this factor explains about 89% of the total variation in the returns distribution over the sample. Positive values of this component are associated with a narrower range of stock returns as shown in the red line in the top right panel in Figure 4, while negative values of the component lead to a density with larger dispersion as reflected in the yellow line there. We note that positive values of loadings largely overlap with the stable economic conditions (non-recession times), shown as non-shaded areas in the top middle panel, while negative values of loadings coincide with the recession periods, shown as shaded areas in the same panel.

In contrast to the first FPC, the second FPC, shown in the middle row of Figure 4, places much more mass in the tails, with lesser mass in the center of the distribution. Since it highlights tail behavior it can be viewed as a *dispersion factor* and explains approximately 7% of the total variation in the returns distribution. We see from the middle panel that the loading increases during the financial crisis in 2008-09, but also in the years following

[25]For the choice of m, we simply followed the convention in the literature, c.f., Section 3.2 of Horváth and Kokoszka (2012). Alternatively, we could have selected m by using the leave-one-out cross-validation (LOOCV) method on minimizing the one-period ahead forecast error variances.

[26]The mean distributional dynamics are computed by taking the mean of the principle component and then multiplying it with the max and min of the associated factor loading time series.

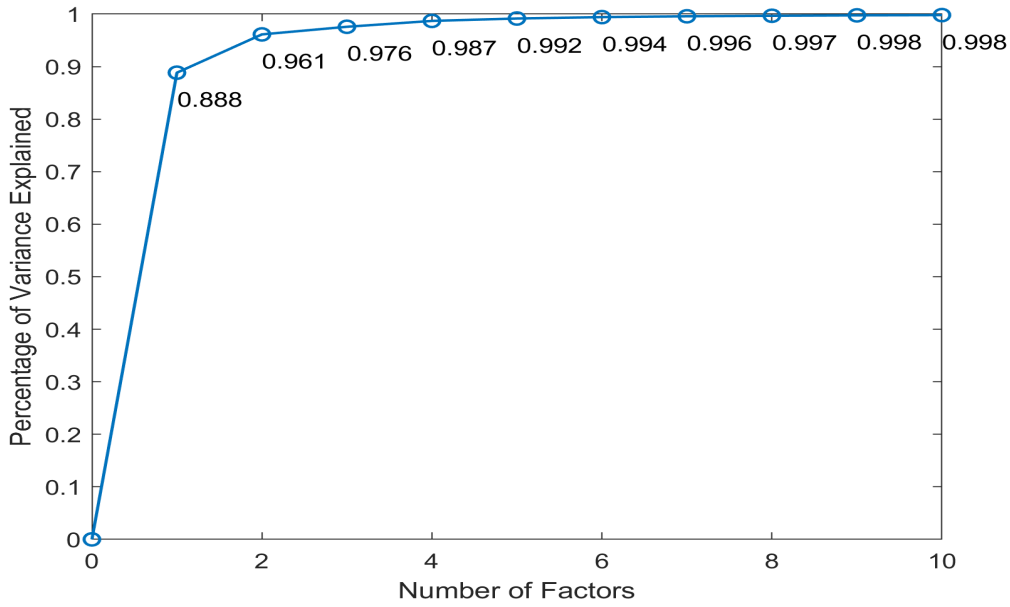


Figure 3: Cumulative scree plot for S&P 500 returns distribution

the oil price collapse in 2014-16, which is consistent with this being related to dispersion. Finally, the third FPC shown in the bottom row of Figure 4, places more mass on the positive side of the returns distributions and can therefore be viewed as a *skew factor*. This factor explains approximately 2% of the total variation in the returns distribution, and does not seem to be related to any specific recessionary period.

3.2 Functional impulse response functions

How do oil market shocks impact the US stock market? Using data on aggregate real stock returns in the US, Kilian and Park (2009) find that unanticipated disruptions in crude oil production have no statistically significant impact, while flow demand shocks have a positive effect and oil-specific demand shocks have a negative effect on stock returns. Their central conclusion is that the stock market may react very differently depending on the underlying oil market shock. This broad conclusion is also found by Bastianin and Manera (2018) in the case of returns volatility. Using the Kilian and Park (2009) SVAR model with realized stock return volatility in place of the returns series, they find that flow-supply shocks have no impact on volatility, while flow-demand shocks reduce volatility, and oil-specific demand increase volatility after a delay of about 5 months. Building on this study

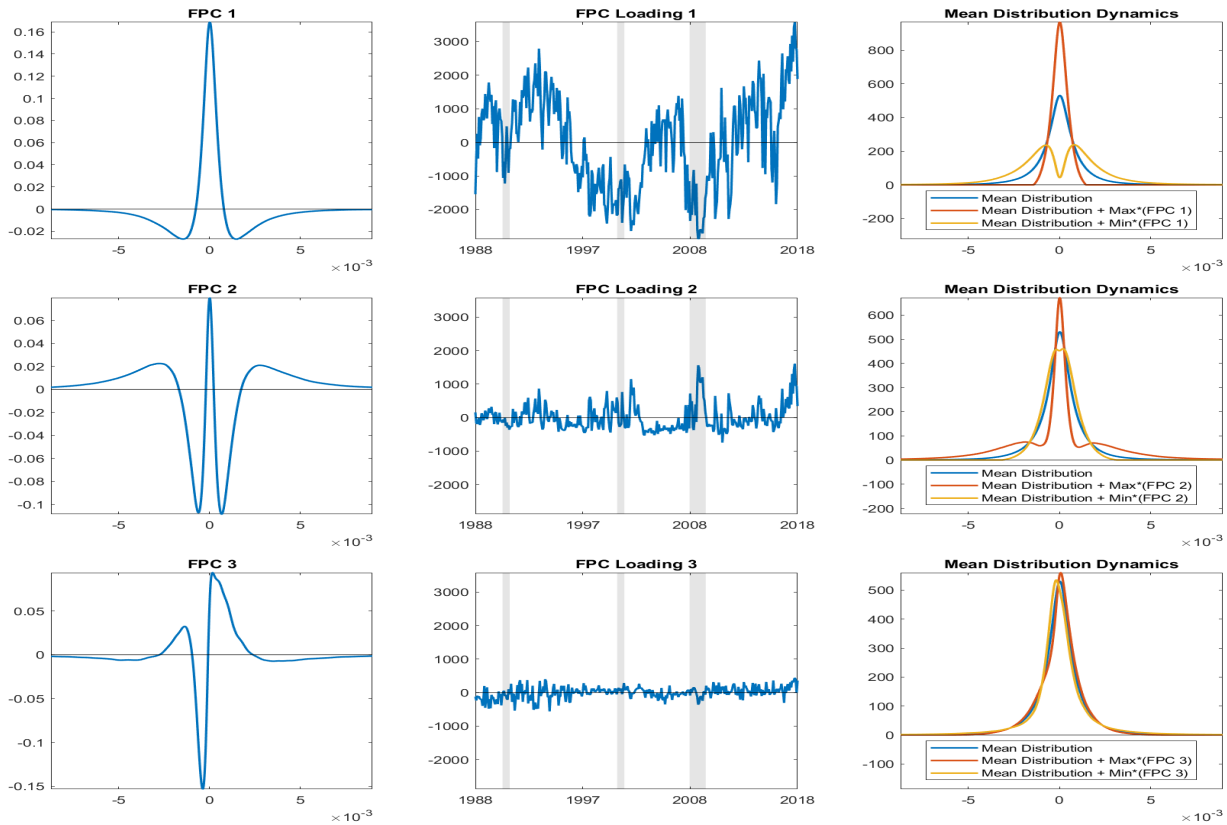


Figure 4: Functional Principal Components of S&P 500 returns distribution.

Notes: Functional principal components (left), time series of their loadings (center) and mean distributional dynamics (right). The top, middle, and bottom figures refer to characteristics for the first, second, and third functional principal component respectively.

we here re-estimate the effects of the three oil market shocks on realized volatility.

To see that our analysis is directly related to Kilian and Park (2009) and Bastianin and Manera (2018) cited above, note that

$$\int r^2 f_t(r)dr \approx \frac{1}{N_t}\frac{1}{39}\sum_{j=1}^{N_t}\sum_{k=1}^{39} r_{j_k,t}^2 = \frac{1}{N_t}\frac{1}{39}\mathrm{RV}_t,$$

where f_t is the stock return density estimated using 10-minute returns for month t, $r_{j_k,t}$ is the k-th 10-minute return in day j of month t having N_t trading days, the number 39 reflects that we have 39-ten minute returns in each trading day and RV_t is the realized variance in month t measured at the 10-minute frequency. The time series analysis of f_t

thus encompasses that of RV_t.[27] We expect that our definition of RV_t is not sensitive to the frequencies of observations on returns, since

$$\sum_{j=1}^{N_t}\sum_{k=1}^{39} r_{j_k,t}^2 \approx \sum_{j=1}^{N_t}\left(\sum_{k=1}^{39} r_{j_k,t}\right)^2 \approx \sum_{j=1}^{N_t} r_{j,t}^2,$$

where $r_{j,t}$ denotes the return in day j of month t as in Bastianin and Manera (2018).

The impulse response functions from the conventional SVAR model are shown in Figure 5. The darker and lighter bands signify 68% and 90% confidence bands, respectively. The first three rows reflect responses in the oil market block (oil production, economic activity and real oil prices), while the fourth row show responses to realized volatility. First focusing on the oil market block, we find that: (1) a positive flow-supply (FS) shock displayed in the left column, leads to a reduction in the price of oil, resulting in a global economic expansion; (2) a positive flow-demand (FD) shock displayed in the middle column, increases the real price of oil which causes producers to increase crude oil production; (3) a positive oil-specific demand (OSD) shock displayed in the right column increases real oil prices, but has a negligible impact of global economic activity and production. Next, turning our attention to the effects on realized volatility displayed in the bottom row, we find that all three shocks decrease realized volatility, although results for FS shocks are not statistically significant. With exception of the effects of the OSD shocks on realized volatility, these results are broadly consistent with earlier work in Bastianin and Manera (2018). For the OSD shock, however, Bastianin and Manera (2018) find realized volatility to decline at first, but then the effect becomes positive before dying out.

We now address the same question with an MVAR approach using functional impulse response functions (FIRFs). That is, Figure 6 now graphs impulse response functions for the oil market variables in the first three rows as in Figure 5, and then FIRFs as 3-D surfaces of the stock returns distribution (DSP) in the bottom row.[28] We note that each response surface presented in the last row of Figure 6 plots probability densities of return

[27] Strictly, this is not true since N_t is not constant and varies between 19 and 24 in our sample. However, the calendar effect generated by N_t is not systematic and can be ignored.

[28] These response functions are responses of demeaned densities. The corresponding responses of densities may be obtained by superimposing them on any density of reference. Note that the resulting counterfactual densities may have negative parts, since no restriction is imposed to ensure them to be non-negative.

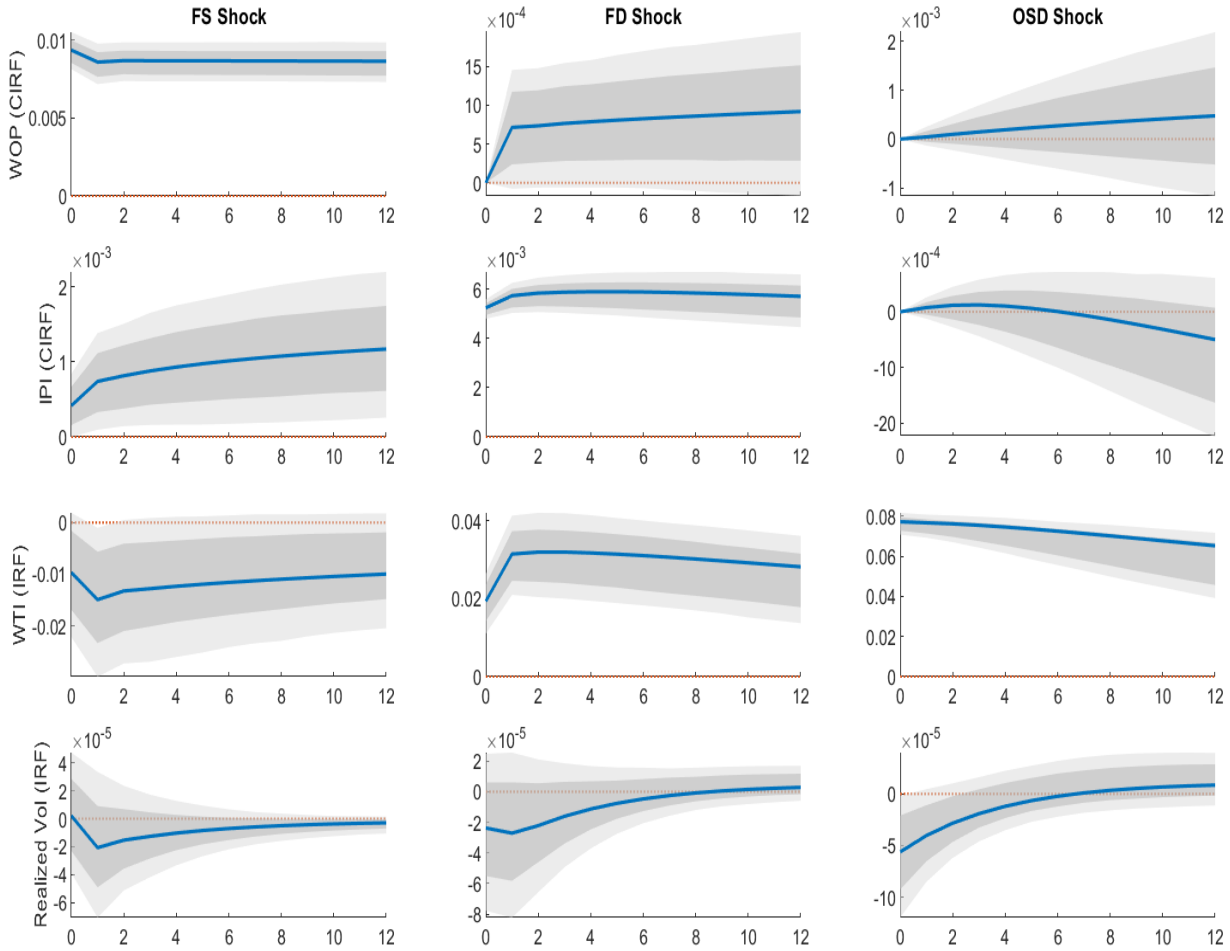


Figure 5: Impulse response functions (IRFs) from conventional SVAR model.

Notes: Each plot shows horizons (months) in the x-axis. The darker and lighter bands signify 68% and 90% confidence bands, respectively. The shocks are flow-supply (FS), flow-demand (FD) and oil-specific demand (OSD).

distributions in the z-axis, along the horizons (months) in the x-axis, as in the conventional response functions of the aggregate variables in the top three rows, and over the ranges of return densities in the y-axis. Finally, the first three columns show the effect of the three oil market shocks, while the fourth column now shows effects of the Var Max shock.

First, we note that the responses in the oil market block are broadly consistent with those from the conventional SVAR model. An important feature of the MVAR approach in this application is that it can jointly determine the dynamic responses of the mean and variance of the (demeaned) stock returns distribution to the various oil market shocks as

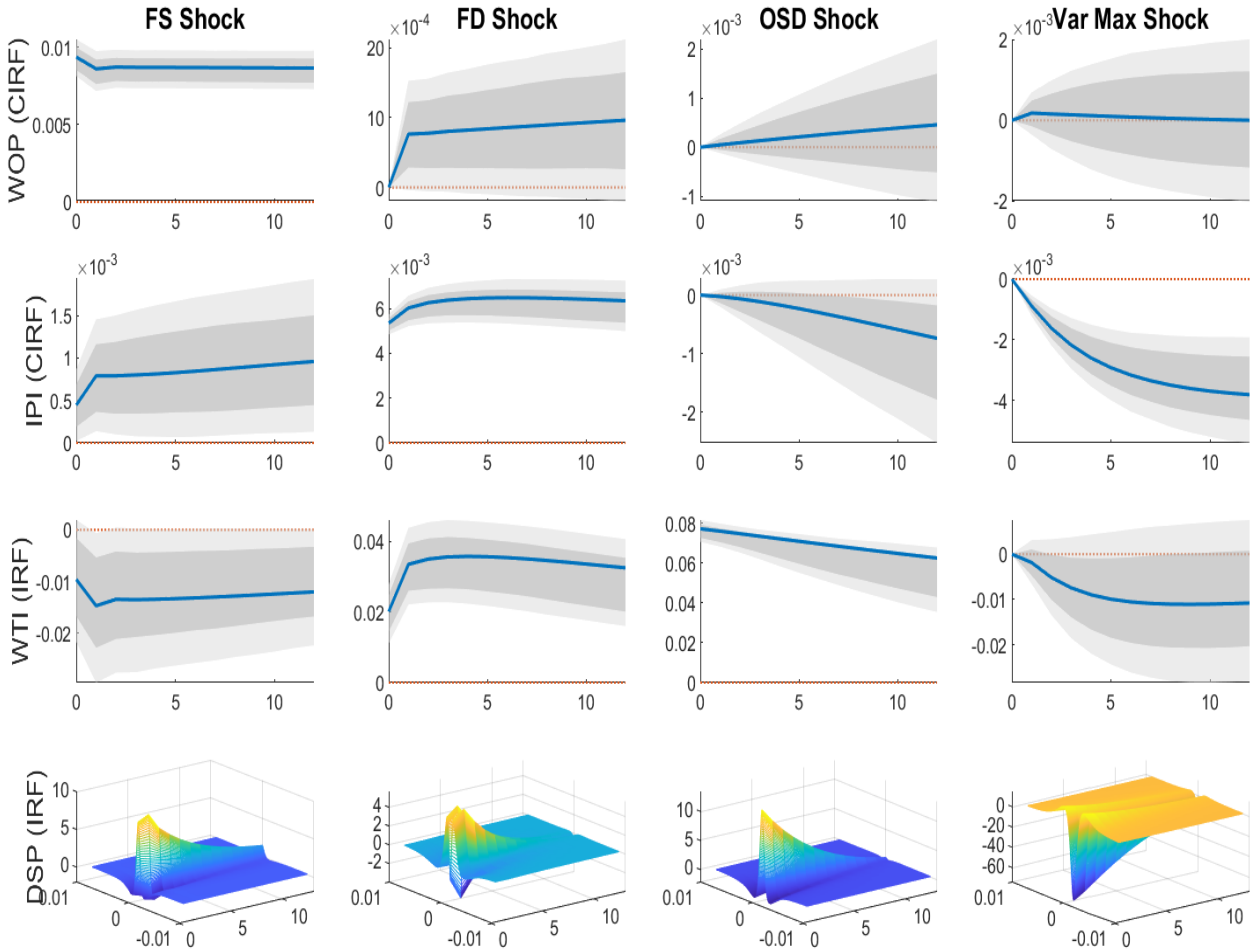


Figure 6: Full set of of impulse response functions and surfaces from MVAR model

Notes: The figure shows conventional impulse response functions in the first three rows, and functional impulse response functions (FIRFs) as 3-D surfaces in the bottom row. Each response surface presented in the last row plots probability densities of return distributions in the z-axis, along the horizons (months) in the x-axis, as in the conventional response functions of the aggregate variables in the top three rows, and over the ranges of return densities in the y-axis. The shocks are flow-supply (FS), flow-demand (FD), oil-specific demand (OSD) and Var Max (see section 2.4 for details). The 68% (dark) and 90% (light) confidence bands (colored areas) are calculated using a bootstrapping method.

opposed to focusing solely on a given moment. Focusing on the distributions in the final row, we observe that both flow-supply and oil-specific demand shocks elicit a persistent positive response in the mean of the DSP. In contrast, the flow-demand shock tends to have an asymmetric effect on the distribution. We also find that the Var Max shock, which is normalized to increase stock return volatility on impact, decreases the return distribution.

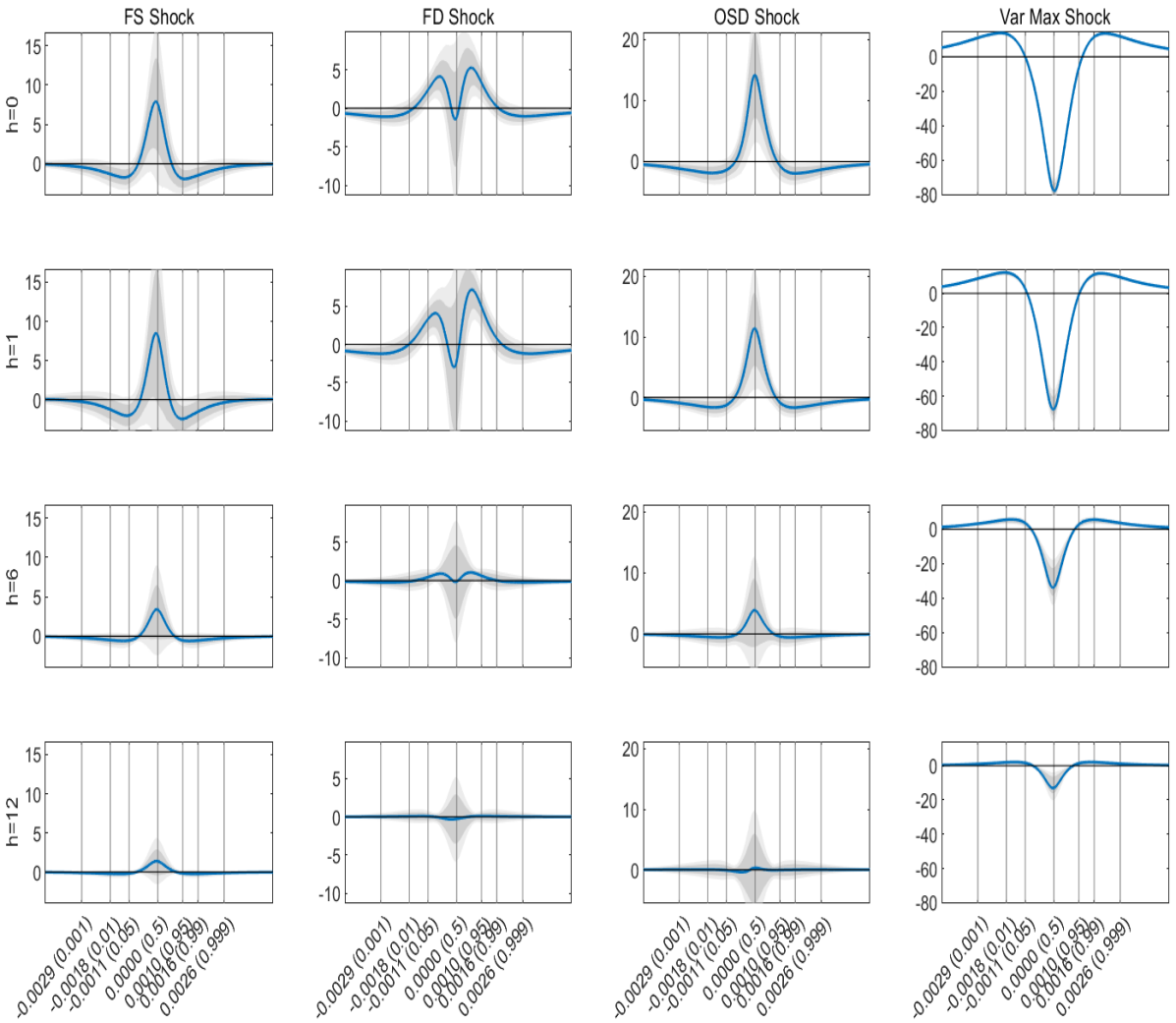


Figure 7: Functional impulse response functions (FIRFs) for the S&P 500 returns distribution at selected horizons.

Notes: The shocks are flow-supply (FS), flow-demand (FD), oil-specific demand (OSD) and Var Max (see section 2.4 for details). The 68% (dark) and 90% (light) confidence bands (colored areas) are calculated using a bootstrapping method. The solid line is the sample response.

It also has a quite persistent negative effect on economic activity, and to a certain extents also on oil prices.

To gauge the statistical significance of these results, we present the FIRFs for the returns distribution in response to the various shocks at selected impulse horizons in Figure 7. We find all shocks to have a significant effect on the distribution in the short run (h=0), and FD and OSD shocks also at h=1. This suggests that there are important distributional

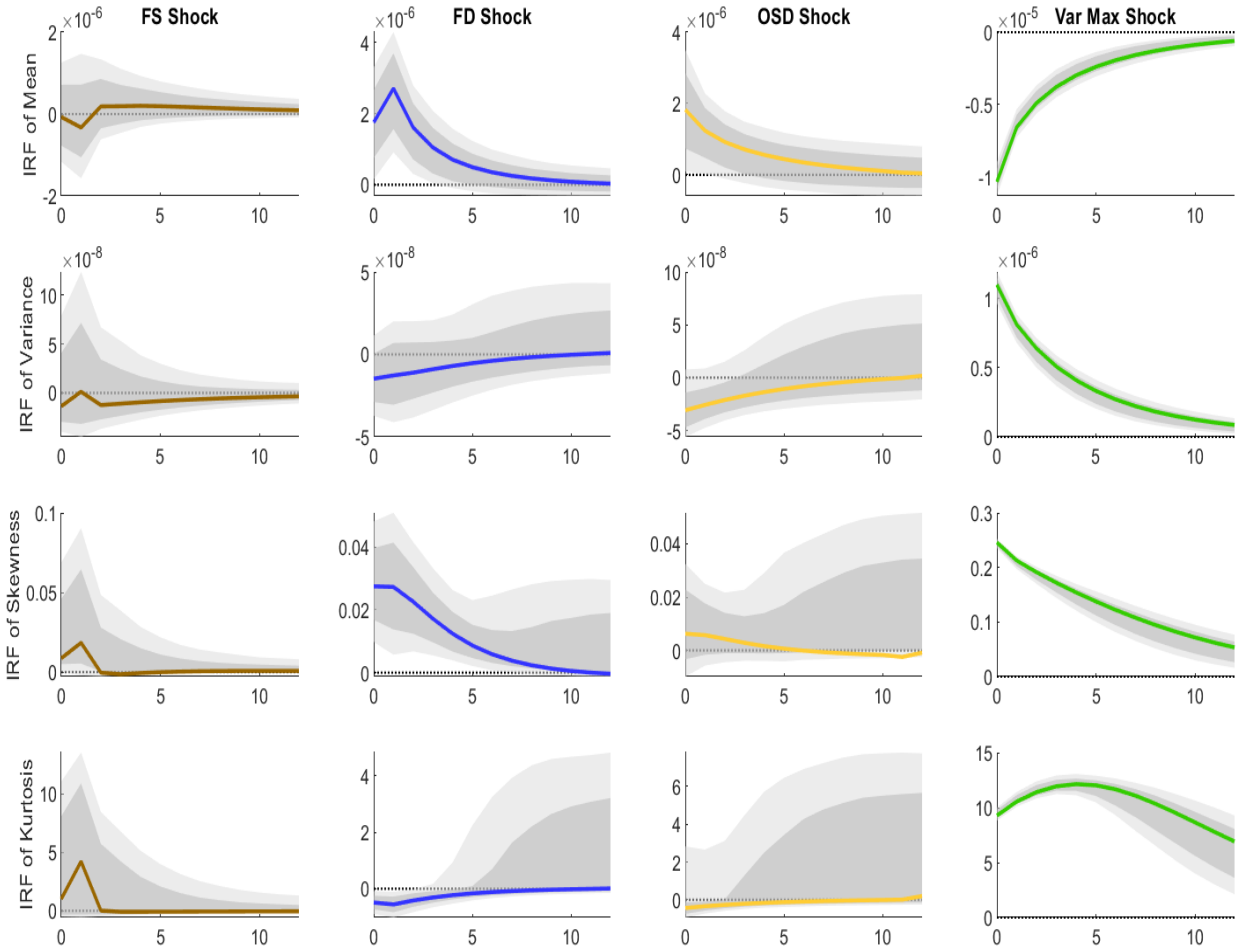


Figure 8: Impulse response functions for the four moments: mean, variance, skewness and kurtosis of the S&P 500 returns distribution.

Notes: The shocks are flow-supply (FS), flow-demand (FD), oil-specific demand (OSD) and Var Max (see section 2.4 for details).

effects that needs to be taken into account when analysing the oil price - stock nexus.

To further highlight the dynamic responses of the moments of the stock returns distribution to the various oil market shocks, Figure 8 shows the impulse responses for the first four moments of the stock returns distribution: mean, variance, skewness, and kurtosis. Starting from the top row, the response of mean returns shows that unanticipated expansions of crude oil production (FS shock) have no significant effect on stock returns. In contrast, an unexpected increase in the global demand for industrial commodities driven by increased global real economic activity (FD shock) causes a short-term increase in U.S.

stock returns. Both of these results are in line with those presented in Kilian and Park (2009). In contrast to their results, however, we find that an oil-specific demand (OSD) shock elicits a brief positive response in mean returns.

Moving to the second row we observe that all three shocks tend to elicit a negative response in stock price volatility: however only the OSD shock is statistically significant in the first few months after the shock. These results are broadly consistent with the IRFs from the SVAR in Figure 5. As discussed also there, results are broadly consistent with Bastianin and Manera (2018), except that realized volatility now decline systematiclly following an OSD shock that increases oil prices.

Combining the mean and variance results suggest that the oil market shocks tend to elicit a negative correlation between volatility and returns, thereby reflecting a financial leverage effect within the stock market. These findings are novel in the literature, which usually focuses on specific moments in time. The results pertaining to oil-specific shocks, however, align well with recent studies suggesting a shift in the relationship between oil and macroeconomic variables in recent years. In particular, Bodenstein et al. (2013) and Datta et al. (2021) argue that lower interest rates (zero lower bound) may have mitigated the negative impact of higher oil prices, leading to a boost in economic activity. Additionally, Bjørnland and Skretting (2024) suggest more positive spillovers to non-oil industries from the shale oil boom. The idea is that if the oil market specific demand shock now also contribute to positive spillovers to some non-oil industries, this can lead to a more optimistic outlook for corporate earnings and economic growth, hence boosting overall stock prices. At the same time, the overall economic stability and growth prospects reduce the uncertainty, thereby lowering volatility. As the aforementioned studies have not been able to focus on the entire returns distribution, our research presents new results in the literature.

To the best of our knowledge, responses relating to the third and fourth moments of the stock return distribution have not been discussed elsewhere in the literature. Focusing on the third row, we find that skewness increases after all three oil market shocks, and for flow supply and demand, significantly so. This suggests that oil market shocks lead to a more asymmetric shape of stock returns. Finally, results in the fourth row show that the FD shock decreases kurtosis in the short-run, which can be interpreted as an increase in the

tail behavior, while the effects of the FS and OSD shocks are not statistically significant. Thus, while the latter two moments are less explored in the literature compared to the first and second moments, the fact that they exhibit statistically significant responses suggests that accounting for these moments is crucial for accurately quantifying the effects of oil market shocks on the stock market.

3.3 Quantifying Distributional Effects

Another feature of the MVAR model is that it facilitates the quantification of the distributional effects of the aggregate and functional shocks. This is given by the contemporaneous functional response of the stock return distribution. These results are illustrated by the histograms in Figures 9 and 10. Figure 9 shows the at-impact responses of the stock return distribution in each decile to the various oil market shocks, while Figure 10 displays similar at-impact responses following the Var Max shock. In both figures, the blue bars represent the 10-th percentile of the stock return distribution prior to the shock. Note that the model is linear, so that results will be the opposite if we change the sign.

Starting with the shocks to the oil market, Figure 9 shows that, on impact, all three oil market shocks decrease extreme (very low or high) returns, that is, the return at the tail end of the distribution. The effect on the middle return, however, varies somewhat. In particular, both the flow supply (FS) and the oil-specific demand (OSD) shocks increase returns between the second and eighth deciles. This pattern is most pronounced for the OSD shock: as demand for oil drives up the price for oil, this leads to higher revenues and profits for oil companies, with potential spillovers to other sectors c.f. Bjørnland and Skretting (2024). This demand increase has a more substantial impact on the oil market than flow supply shocks. The effects of the flow demand (FD) shock, however, are more complex. Like the two other oil market shocks, the FD shock decreases extreme returns at the top and bottom deciles. However, in contrast to the other shocks, it also decreases the returns between the fourth and sixth deciles. For the remaining deciles, it increases returns as the two other shocks: for moderately low and high (second and ninth decline) returns, the FD shock has noticeably larger effects than the FS and OSD shocks, but for mildly low and high (third and seventh-eights decile) returns, the FD shock has smaller effect relative

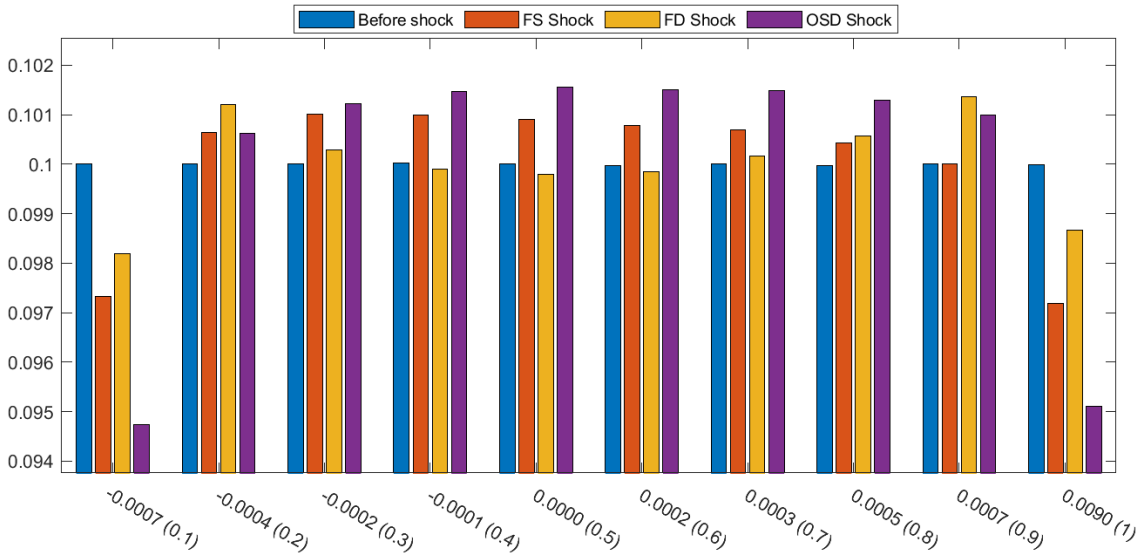


Figure 9: Response of S&P Distribution to oil market shocks (h=0)
Note: The shocks are flow-supply (FS), flow-demand (FD), oil-specific demand (OSD)

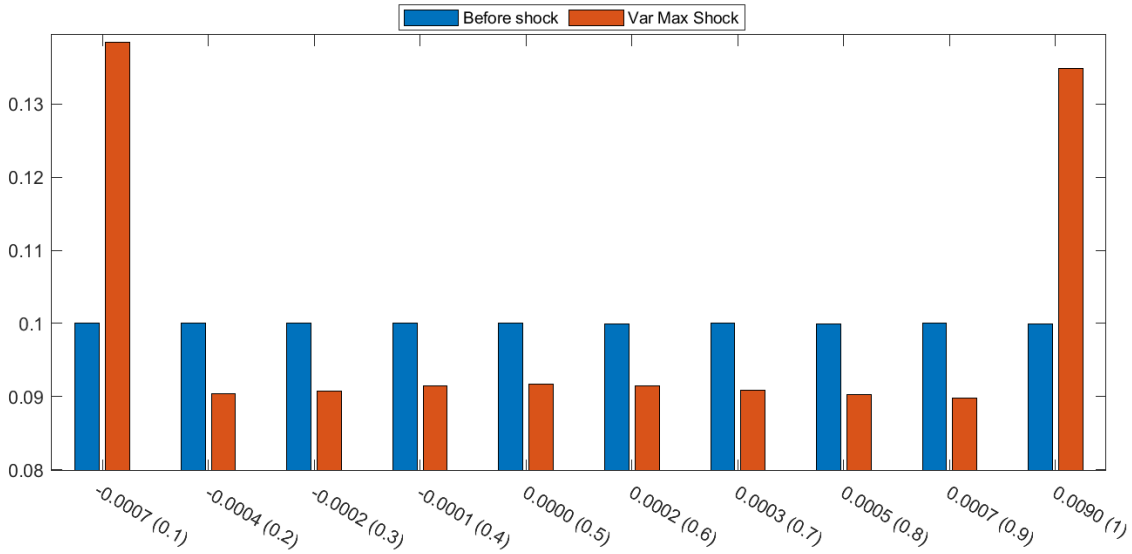


Figure 10: Response of S&P Distribution to Var Max Shock (h=0)

to the two other oil market shocks.

In clear contrast to the oil market shocks, Figure 10 shows that the Var Max shock substantially increases extreme returns (both top and bottom deciles) while decreasing the returns in the mid-range of the distribution. Hence, at impact, the Var Max shock is clearly affecting the tails of the stock distribution relative to the middle returns.

In summary, the results suggest that each of the shocks have very different effects on the

range of the stock return distribution at impact. The Var Max shock primarily increases the extreme returns, while the oil market shocks mostly increase the mid-range returns.

3.4 Value at Risk analysis

Another advantage of the MVAR approach is that it allows us to examine the effects of the structural shocks on the value-at-risk (VaR). The VaR is defined as the maximum expected loss on an investment, over a given time period at a specified degree of confidence. It is therefore a widely used measure of risk within academia and by private and institutional investors, and financial market regulators, see for instance Wiener (1999) for a textbook treatment or Cabedo and Moya (2003) and Tiwari et al. (2020) for some papers examining the effect of oil prices on VaR. In contrast to these papers, we examine the effects of oil market shocks on VaR (as opposed to aggregate price shocks), and will provide a novel functional approach to quantify downside risks using VaRs computed from the entire stock return distribution.

To first gauge the accuracy of the FPCA approach in approximating the VaR dynamics, we plot the implied 5% VaR associated with the distributions of 10 minute returns from the data and first three functional PCs from the FPCA analysis in Figure 11.[29] The vertical axis are in fractional form. For instance, our 5% VaR of -2×10^{-3} means that there is a 5% chance of a 0.2% loss over a 10 minutes trading during that month. The most notable feature of the VaR time series behavior is that the estimated VaR is mostly consistent with that of the original distribution. This suggests that FPCA approach does a good job of capturing the distributional dynamics of the stock returns distribution. The notable exception is during the Great Recession where the VaR associated with the MVAR model is around half the size of the VaR implied by the returns distribution implied by the data alone.

To investigate which of the oil market shocks are important for determining the VaR, we conduct a counterfactual in which we report the VaR when only one of the structural shocks

[29] We may obtain the VaR over a month from the VaR over a 10 minute interval under the i.i.d. assumption. If Gaussian distribution, or more generally, a stable distribution with thicker tails is assumed, then we may compute a monthly VaR directly from a 10-minute VaR for each month. Otherwise, we need to do bootstrap by randomly drawing 10-minute returns and obtain monthly returns by compounding them over a given month.

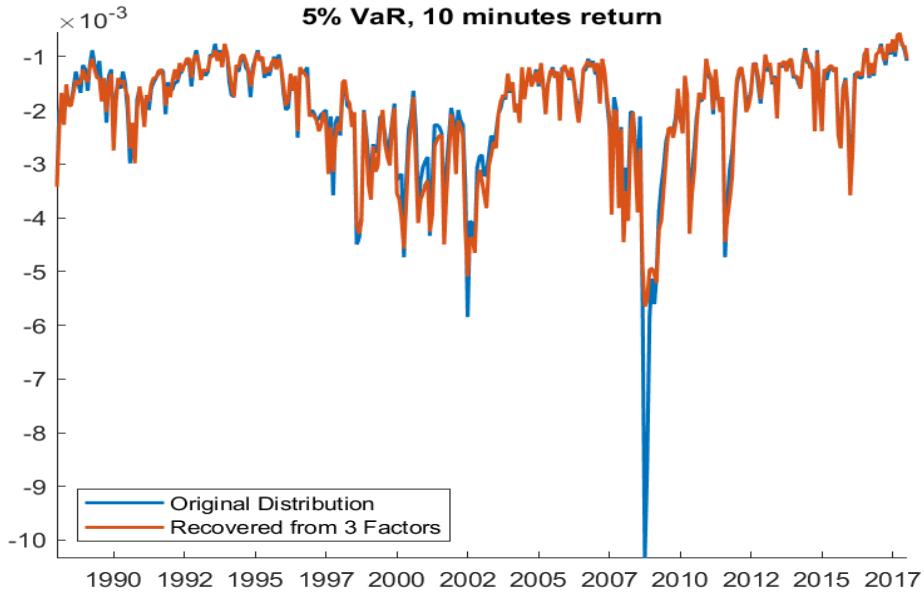
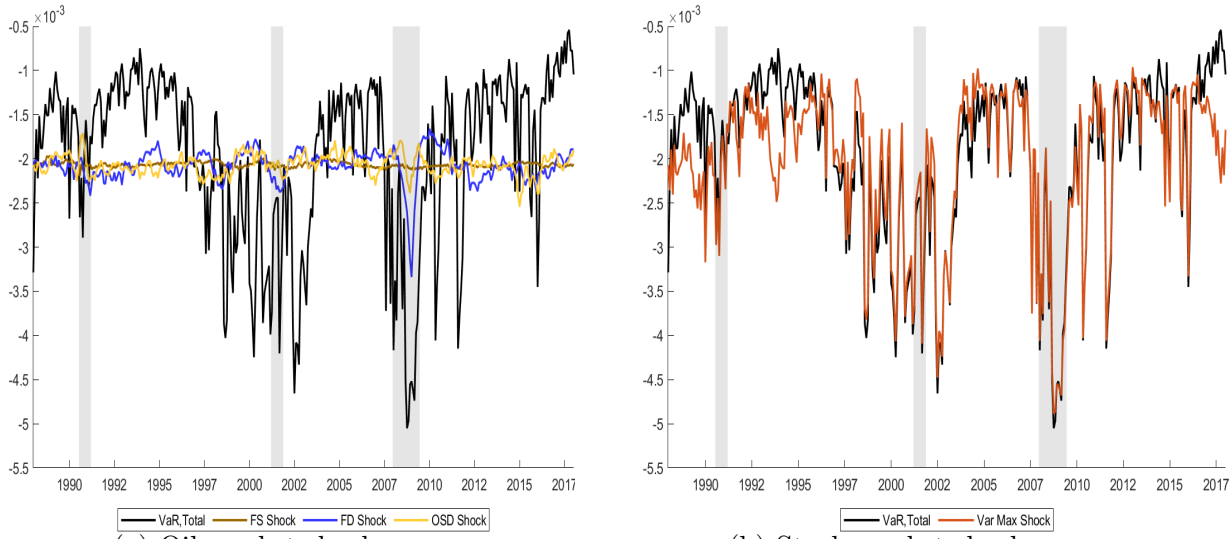


Figure 11: Value at risk (VaR) for the S&P 500 returns distribution over the period 1988:02-2018:01.



(a) Oil market shocks (b) Stock market shocks

Figure 12: Counterfactual: Value at risk (VaR) for the S&P 500 returns distribution over the sample 1988:01-2018:01 when only one shock is realized.

Notes: The shocks are flow-supply (FS), flow-demand (FD), oil-specific demand (OSD) and maximum variance (Var Max).

is realized. The results are provided in Figure 12. It is immediately clear that the Var Max shock closely tracks the VaR total throughout the sample period. Perhaps unsurprisingly,

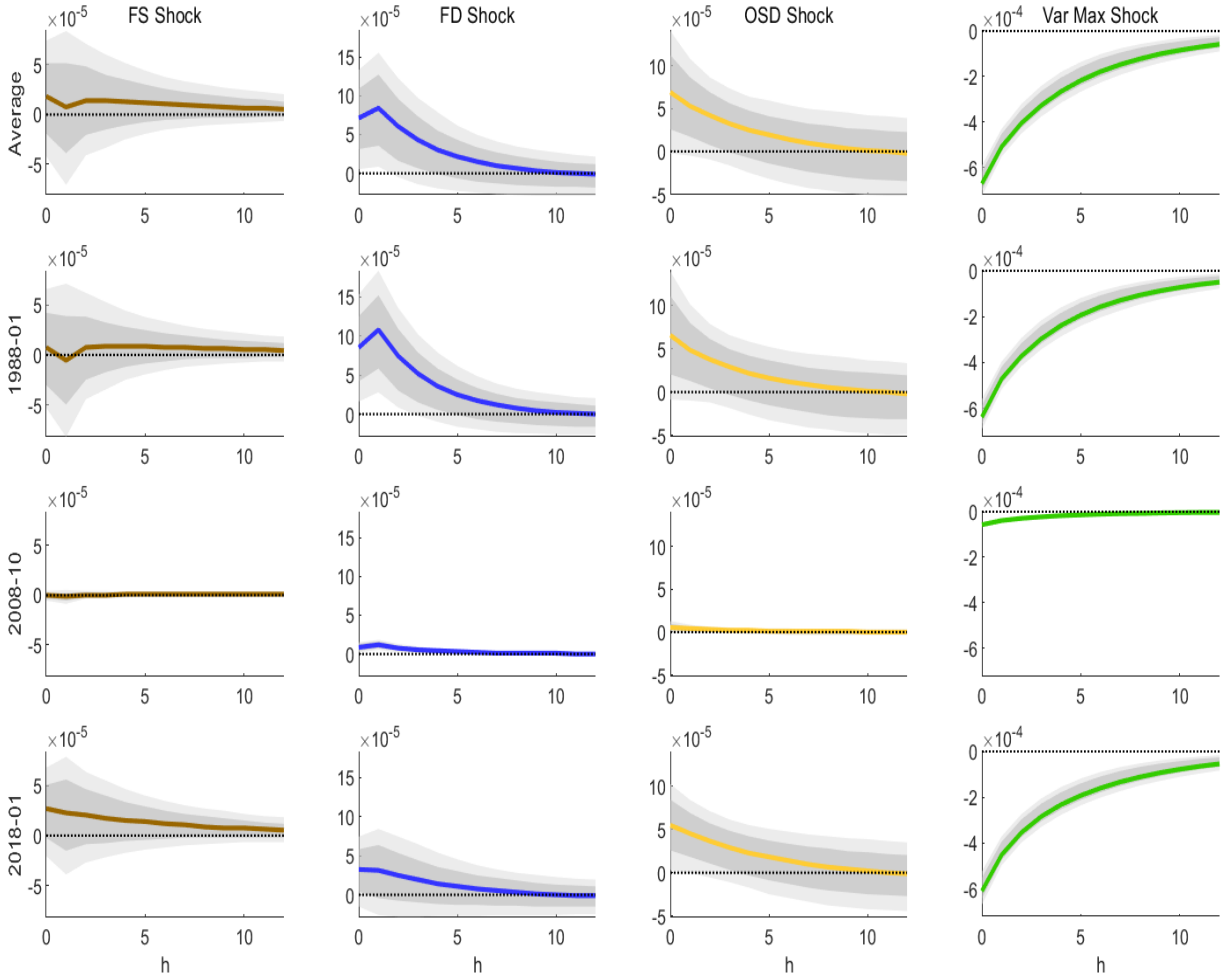


Figure 13: Impulse response functions for the value-at-risk (VaR) for the S&P 500 returns distribution to the oil market demand and supply shocks using the MVAR model.

Notes: The shocks are flow-supply (FS), flow-demand (FD), oil-specific demand (OSD) and Var Max

this suggests that most of the variation in the VaR is driven by forces underlying the US stock market rather than the global market for crude oil. However, this does not imply that the oil market shocks are unimportant. For instance, oil market shocks are found to have (absolutely) increased the VaR during periods of economic turmoil, such as the crisis of the early 1990s and 2000s, as well as the 2007-09 Great Recession. Such shocks also had sizable impacts on the VaR during other periods, such as the Iraqi invasion of Kuwait of 1990, OPEC's announcement to increase its production ceiling in 1997-98, and the oil price decline of 2014.

Next, to determine how the VaR responds to the various oil market shocks, Figure 13

provides FIRFs for the VaR to each of the oil market shocks both across the entire sample (first row) and in selected periods (rows 2-4). First focusing on the top row, it is immediately obvious that the effects of the oil market demand and supply shocks on the VaR are extremely short-lived, often disappearing after 1-3 months, and distinct. On average, we find that flow-supply shocks tend to elicit no response in the VaR over the sample period. In contrast, flow-demand shocks have a positive and significant effect on impact and in the short run, before reverting to zero. Finally, the oil-specific demand shock immediately increases the VaR and reverts to zero after three months. The fact that the two demand shocks increase the VaR in the short-run implies that they tend to make the value-at-risk (absolutely) smaller, thereby reducing tail risk.

To further examine the effects of these shocks over the sample, rows 2-4 present the VaR responses in selected periods. We observe that the oil market shocks impact the VaR differently over time. For instance, results in row 2 show that the two demand shocks increase the VaR at the start of the sample, thereby reducing the tail risk in the returns distribution. In contrast, we find that the oil market had no impact on the VaR in 2008. Finally, the bottom row shows that both supply shocks and oil-specific demand shocks generated persistent increases in the VaR, while the flow demand shock had no effect.

Taken together, the results suggest that the impact of oil market demand and supply shocks on the VaR is distinct and changes over time. When they do have an impact, the effects tend to be extremely short-lived, often disappearing after 1-3 months, however there are episodes where the effects are more persistent. Such information may prove useful for market participants when formulating the effects of oil price shocks on portfolio risks.

4 Conclusion

We have reexamined the oil-stock price nexus using a novel VAR model that allows for a mixture of aggregate and functional variables in a multivariate setting. The major innovation of this mixed vector autoregression (MVAR) approach is that it allows us to jointly examine the dynamic response of aggregate and functional time series to conventional and distributional shocks. In our application, this facilitated an examination of how oil price

shocks impact the entire stock return distribution, instead of focusing on a single representative series, such as returns or realized volatility. Using this new framework we yielded new insights into the oil-stock price nexus from three perspectives.

First, from an econometric modelling perspective, we showed how the reduced form MVAR model can be estimated using a combination of functional principle components analysis and ordinary least squares, how to identify both aggregate and functional shocks, and how to produce and interpret functional impulse response functions. Theoretical and quantitative evidence was also provided for the use of the FPC basis over three alternative data-driven basis functions: the moment basis, the quantile basis, and the histogram basis. Empirically, we found that the FPC basis clearly outperforms the other bases across three measures of in-sample fit. The first three leading functional principal components were found to effectively extract around 98% of the temporal variation of the intra-month S&P 500 returns distribution.

Next, from an economic perspective, we used functional IRFs (FIRFs) to show the importance of accounting for the changes in the entire stock return distribution when assessing the effects of oil shocks on the stock market. As a preliminary diagnostic, we showed that the FIRFs reproduce the general result from IRFs that demand and supply shocks have very different effects on the real price of oil. We then shifted focus to the dynamic response of the returns distribution to the oil market shocks and found that both of the demand shocks elicit a persistent positive response in the mean of the DSP while also reducing the variance of the distribution. In contrast, the flow-supply shock tends to have no statistically significant impact on the first two moments. Another intriguing result was that the flow-demand shock increases the skewness and decreases the kurtosis of the returns distribution in the first few months after the shock; however, the other two oil market shocks had no statistically significant impact on these higher-order moments. Moreover, the Var Max shock significantly decreases returns, while increasing the skewness and kurtosis of the distribution. These results highlight the importance of controlling for distributional dynamics when studying the oil-stock price nexus.

Finally, from a policy perspective, we demonstrated how the MVAR approach can be used to conduct a functional value-at-risk (VaR) analysis. The results showed that our

approach does a good job of reflecting the VaR implied by the returns distribution implied by the data alone. We also provided the novel result that oil market demand and supply shocks may have very different effects on the VaR, and that these effects change over time. Market participants may benefit from such information when determining how oil price shocks affect portfolio risk.

While we have concentrated on the oil-stock price nexus in the US, there is clearly a wide scope of possible avenues for future research with the MVAR approach. Natural direct extensions of our empirical study include examining the effects of oil price shocks on individual equity prices, equity prices across sectors, stock markets in other countries, and problems of optimal portfolio choice. The MVAR approach may also prove to be useful in other areas at the frontier of research in macroeconomics and finance such as identifying drivers of income inequality, analysing the effects of shocks on the distribution of firms' revenue, and understanding the distributional dynamics of climate change.

References

Ahmed, E., Rosser Jr, J. B., and Sheehan, R. G. (1988). A global model of oecd aggregate supply and demand using vector autoregressive techniques. *European Economic Review*, 32(9):1711–1729.

Alsalman, Z. and Herrera, A. M. (2015). Oil price shocks and the US stock market: do sign and size matter? *The Energy Journal*, pages 171–188.

Baillie, R. T. and DeGennaro, R. P. (1990). Stock returns and volatility. *Journal of Financial and Quantitative Analysis*, 25(2):203–214.

Bastianin, A. and Manera, M. (2018). How does stock market volatility react to oil price shocks? *Macroeconomic Dynamics*, 22(3):666–682.

Baumeister, C. and Hamilton, J. D. (2019). Structural interpretation of vector autoregressions with incomplete identification: Revisiting the role of oil supply and demand shocks. *American Economic Review*, 109(5):1873–1910.

Bernanke, B. S., Gertler, M., Watson, M., Sims, C. A., and Friedman, B. M. (1997). Systematic monetary policy and the effects of oil price shocks. *Brookings papers on economic activity*, 1997(1):91–157.

Bjørnland, H. C. (2000). The dynamic effects of aggregate demand, supply and oil price shocks—a comparative study. *The Manchester School*, 68(5):578–607.

Bjørnland, H. C. (2009). Oil price shocks and stock market booms in an oil exporting country. *Scottish Journal of Political Economy*, 56(2):232–254.

Bjørnland, H. C. and Skretting, J. (2024). The shale oil boom and the US economy: Spillovers and time-varying effects. *Journal of Applied Econometrics*.

Bodenstein, M., Guerrieri, L., and Gust, C. J. (2013). Oil shocks and the zero bound on nominal interest rates. *Journal of International Money and Finance*, 32:941–967.

Boldanov, R., Degiannakis, S., and Filis, G. (2016). Time-varying correlation between oil and stock market volatilities: Evidence from oil-importing and oil-exporting countries. *International Review of Financial Analysis*, 48:209–220.

Bosq, D. (2000). *Linear Processes in Function Spaces*. Springer-Verlag.

Burbidge, J. and Harrison, A. (1984). Testing for the effects of oil-price rises using vector autoregressions. *International Economic Review*, pages 459–484.

Cabedo, J. D. and Moya, I. (2003). Estimating oil price 'Value at Risk'using the historical simulation approach. *Energy economics*, 25(3):239–253.

Chang, M., Chen, X., and Schorfheide, F. (2021a). Heterogeneity and aggregate fluctuations. Technical report, National Bureau of Economic Research. Working Paper 28853.

Chang, Y., Gómez-Rodríguez, F., and Hong, M. G. H. (2022). *The Effects of Economic Shocks on Heterogeneous Inflation Expectations*. International Monetary Fund.

Chang, Y., Miller, I. J., and Park, J. Y. (2023). Shocking climate: Identifying economic damages from anthropogenic and natural climate change. Mimeo, Indiana University.

Chang, Y., Park, J. Y., and Pyun, D. (2021b). From functional autoregressions to vector autoregressions. Working Paper, Indiana University.

Cross, J. L., Nguyen, B. H., and Tran, T. D. (2022). The role of precautionary and speculative demand in the global market for crude oil. *Journal of Applied Econometrics*, 37(5):882–895.

Datta, D. D., Johannsen, B. K., Kwon, H., and Vigfusson, R. J. (2021). Oil, equities, and the zero lower bound. *American Economic Journal: Macroeconomics*, 13(2):214–253.

Degiannakis, S., Filis, G., and Arora, V. (2018). Oil prices and stock markets: A review of the theory and empirical evidence. *The Energy Journal*, 39(5).

Degiannakis, S., Filis, G., and Kizys, R. (2014). The effects of oil price shocks on stock market volatility: Evidence from European data. *The Energy Journal*, 35(1).

French, K. R., Schwert, G. W., and Stambaugh, R. F. (1987). Expected stock returns and volatility. *Journal of Financial Economics*, 19(1):3–29.

Güntner, J. H. (2014). How do international stock markets respond to oil demand and supply shocks? *Macroeconomic Dynamics*, 18(8):1657–1682.

Hall, P. and Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression. *Annals of Statistics*, 35:70–91.

Hamilton, J. D. (1983). Oil and the macroeconomy since World War II. *Journal of Political Economy*, 91(2):228–248.

Hamilton, J. D. (2003). What is an oil shock? *Journal of Econometrics*, 113(2):363–398.

Hamilton, J. D. (2009). Causes and consequences of the oil shock of 2007-08. Technical report, National Bureau of Economic Research.

Horta, E. and Ziegelmann, F. (2018). Dynamics of financial returns densities: A functional approach applied to the bovespa intraday index. *International Journal of Forecasting*, 34:75–88.

Horváth, L. and Kokoszka, P. (2012). *Inference for functional data with applications*, volume 200. Springer Science & Business Media.

Huang, R. D., Masulis, R. W., and Stoll, H. R. (1996). Energy shocks and financial markets. *Journal of Futures Markets*, 16(1):1–27.

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

Jones, C. M. and Kaul, G. (1996). Oil and the stock markets. *The Journal of Finance*, 51(2):463–491.

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.

Kilian, L. and Park, C. (2009). The impact of oil price shocks on the US stock market. *International Economic Review*, 50(4):1267–1287.

Kneip, A. and Utikal, K. J. (2001). Inference for density families using functional principal component analysis. *Journal of the American Statistical Association*, 96:519–542.

Kokoszka, P., Miao, H., Reimherr, M., and Taoufik, B. (2018). Dynamic functional regression with application to the cross-section of returns. *Journal of Financial Econometrics*, 16(3):461–485.

Kokoszka, P. and Reimherr, M. (2017). *Introduction to Functional Data Analysis*. CRC Press.

Meeks, R. and Monti, F. (2022). Heterogeneous beliefs and the Phillips curve. *Available at SSRN*.

Miller, J. I. and Ratti, R. A. (2009). Crude oil and stock markets: Stability, instability, and bubbles. *Energy economics*, 31(4):559–568.

Paparoditis, E. (2018). Sieve bootstrap for functional time series. *Annals of Statistics*, 46:3510–3538.

Park, J. and Ratti, R. A. (2008). Oil price shocks and stock markets in the US and 13 European countries. *Energy Economics*, 30(5):2587–2608.

Park, J. Y. and Qian, J. (2012). Functional regression of continuous state distributions. *Journal of Econometrics*, 167:397–412.

Petersen, A. and Müller, H.-G. (2016). Functional data analysis for density functions by transformation to a Hilbert space. *Annals of Statistics*, 44:183–218.

Petersen, A., Zhang, C., and Kokoszka, P. (2022). Modeling probability density functions as data objects. *Econometrics and Statistics*, 21:159–178.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer.

Ramsay, J. O. and Silverman, B. W. (2007). *Applied Functional Data Analysis: Methods and Case Studies*. Springer.

Sadorsky, P. (1999). Oil price shocks and stock market activity. *Energy Economics*, 21(5):449–469.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, pages 1–48.

Thorbecke, W. (2019). Oil prices and the US economy: Evidence from the stock market. *Journal of Macroeconomics*, 61:103137.

Tiwari, A. K., Trabelsi, N., Alqahtani, F., and Raheem, I. D. (2020). Systemic risk spillovers between crude oil and stock index returns of G7 economies: Conditional value-at-risk and marginal expected shortfall approaches. *Energy Economics*, 86:104646.

Tsay, R. (2016). Some methods for analyzing big dependent data. *Journal of Business and Economic Statistics*, 34(4):673–688.

Wiener, Z. (1999). Introduction to VaR (value-at-risk). In *Risk Management and Regulation in Banking: Proceedings of the International Conference on Risk Management and Regulation in Banking (1997)*, pages 47–63. Springer.

Appendices

Appendix A Extended Hilbert Space

To define our MVAR more precisely, we need to introduce some basic concepts of Hilbert space. In (3), z_t is formally defined as a time series of random elements in $R^\ell \times H$, which itself is also defined as another Hilbert space. The new Hilbert space $R^\ell \times H$ is endowed with the inner product

$$\langle z, w \rangle = \langle x, y \rangle + \langle f, g \rangle$$

for all $z = (x, f)$ and $w = (y, g)$ with $x, y \in R^\ell$ and $f, g \in H$, where $\langle \cdot, \cdot \rangle$ is used as a generic notation for the inner product in R^ℓ, H and $R^\ell \times H$.[30] The standard inner products in R^ℓ and H are given by

$$\langle x, y \rangle = x'y \quad \text{and} \quad \langle f, g \rangle = \int f(r)g(r)dr,$$

respectively, for all $x, y \in R^\ell$ and $f, g \in H$. We also use the same notation $\|\cdot\|$ for the norm induced by the inner product $\langle \cdot, \cdot \rangle$ in R^ℓ, H and $R^\ell \times H$, and therefore, $\|z\|^2 = \langle z, z \rangle = \langle x, x \rangle + \langle f, f \rangle = \|x\|^2 + \|f\|^2$ for all $z = (x, f) \in R^\ell \times H$.

Note that the autoregressive operator A in (3) is an element of the space $L(R^\ell \times H)$ of linear operators on $R^\ell \times H$ and may be written more explicitly as the matrix of operators

$$A = \begin{pmatrix} A_{RR} & A_{RH} \\ A_{HR} & A_{HH} \end{pmatrix},$$

where $A_{RR} : R^\ell \to R^\ell$, $A_{RH} : H \to R^\ell$, $A_{HR} : R^\ell \to H$ and $A_{HH} : H \to H$. If we let $w = Az$ for $z = (x, f)'$ and $w = (y, g)'$ with $x, y \in R^\ell$ and $f, g \in H$, we may represent A_{RR} as an $\ell \times \ell$ matrix such that $y = A_{RR}x$,

$$A_{RH} = \begin{pmatrix} \langle h_1, \cdot \rangle \\ \vdots \\ \langle h_\ell, \cdot \rangle \end{pmatrix} \quad \text{with} \quad y = A_{RH}f = \begin{pmatrix} \langle h_1, f \rangle \\ \vdots \\ \langle h_\ell, f \rangle \end{pmatrix},$$

and $A_{HR} = (k_1, \ldots, k_\ell)'$ with $g = A_{HR}x = (k_1, \ldots, k_\ell)'x$, for some $h_1, \ldots, h_\ell \in H$ and $k_1, \ldots, k_\ell \in H$, and A_{HH} is a bounded linear operator on H for which $g = A_{HH}f$.

We also need to introduce the notion of tensor product in H and $R^\ell \times H$. For $f, g \in H$,

[30] This minor abuse of notation is done for notational brevity and to improve the overall readability.

their tensor product $f \otimes g$ is defined as a linear operator on H satisfying

$$(f \otimes g)h = \langle h, g \rangle f$$

for all $h \in H$. We may define the tensor product $z \otimes w$ of $z = (x, f), w = (y, g) \in R^\ell \times H$ similarly as a linear operator on $R^\ell \times H$ given by

$$(z \otimes w)v = \langle w, v \rangle z = \big(c'y + \langle h, g \rangle\big)(x, f)$$

for all $v = (c, h) \in R^\ell \times H$. When $H = R^m$, we have $f \otimes g = fg'$ and $z \otimes w = zw'$, and they become outer products of two vectors, contrastingly with their inner products given by $\langle f, g \rangle = f'g$ and $\langle z, w \rangle = z'w$.

The functional error (ε_t) in (3) is assumed to be white noise. That is, $\mathbb{E}(\varepsilon_t) = 0$, and

$$\mathbb{E}(\varepsilon_t \otimes \varepsilon_s) = \begin{cases} \Sigma & \text{for} \quad t = s \\ 0 & \phantom{\text{for}} \quad t \neq s \end{cases},$$

where $\mathbb{E}(\varepsilon_t \otimes \varepsilon_s)$ is the covariance operator of ε_t and ε_s, which is defined as a linear operator on $R^\ell \times H$ such that $\langle z, \mathbb{E}(\varepsilon_t \otimes \varepsilon_s)w \rangle = \mathbb{E}\langle z, \varepsilon_t \rangle \langle w, \varepsilon_s \rangle$ for all $z, w \in R^\ell \times H$. In short, (ε_t) is a serially uncorrelated functional time series with mean zero and a common variance operator, as a white noise vector time series in the finite-dimensional case.

Appendix B Choice of Basis

The VAR representation in (8) may be obtained for any choice of an orthonormal basis (v_i) of $R^\ell \times H$. The effectiveness of the resulting approximation, however, depends crucially on the choice of basis. We here show this from both theoretical and applied perspectives.

Theory Let $(v^*_{i+\ell})_{i=1}^m$ be a sub-basis of $(v_{i+\ell})$, which generates the subspace V^* of H, so that $(v^*_i)_{i=1}^n$ with $n = \ell + m$ spans the subspace $R^\ell \times V^*$ of $R^\ell \times H$. In contrast, we denote by (v_i) a basis of $R^\ell \times H$ such that $v_i = e_i$ for $i = 1, \ldots, \ell$ and $(v_{i+\ell})_{i=1}^m$ spans an arbitrary subspace V of H, and therefore, its sub-basis $(v_i)_{i=1}^n$ generates $R^\ell \times V$. In what follows, we let Q and Q^* be the Hilbert space projections on V and V^*, respectively, and P and P^* be the Hilbert space projections on $R^\ell \times V$ and $R^\ell \times V^*$, respectively. By definition, we have

$$\sum_{t=1}^T \|Q^* f_t\|^2 \geqslant \sum_{t=1}^T \|Q f_t\|^2,$$

which implies that $(Q^* f_t)$ has the maximum temporal variation. It follows that

$$\sum_{t=1}^{T} \|P^* z_t\|^2 \geqslant \sum_{t=1}^{T} \|P z_t\|^2,$$

since

$$\sum_{t=1}^{T} \|P z_t\|^2 = \sum_{t=1}^{T} \|x_t\|^2 + \sum_{t=1}^{T} \|Q f_t\|^2 \quad \text{and} \quad \sum_{t=1}^{T} \|P^* z_t\|^2 = \sum_{t=1}^{T} \|x_t\|^2 + \sum_{t=1}^{T} \|Q^* f_t\|^2.$$

This shows that the approximation of A given by $P^* A P^*$ restricts A to the subspace $R^\ell \times V^*$ of $R^\ell \times H$, where (z_t) has the maximum temporal variation and A is most strongly identified. In this sense, the basis (v_i^*) provides the most effective approximation of A among its restrictions on an n-dimensional subspace $R^\ell \times V$ of $R^\ell \times H$ spanned by $(v_i^*)_{i=1}^n$.

Application To compare the consequences of using $(v_i)_{i=1}^n$ and $(v_i^*)_{i=1}^n$ more explicitly, we consider the functional R-squared (FR-squared) of an arbitrary sub-basis $(v_{i+\ell})_{i=1}^m$ given by

$$\mathrm{FR}^2 = \frac{\sum_{t=1}^{T} \|Q f_t\|^2}{\sum_{t=1}^{T} \|f_t\|^2},$$

which represents the proportion of the total variation of (f_t) explained by its projection $(Q f_t)$ on the subspace V spanned by $(v_{i+\ell})_{i=1}^m$. If we denote by FR_*^2 the FR-squared of $(v_{i+\ell}^*)_{i=1}^m$, then

$$\mathrm{FR}_*^2 \geqslant \mathrm{FR}^2,$$

since by definition $(Q^* f_t)$ has the maximum temporal variation.

Before moving on, we note that the π's in (6) and (7) are not just one-to-one mappings. They are *isometries*. This is seen from the fact that $\|z\|^2 = \big\|(z)\big\|^2$ for $z \in R^\ell \times V$, and $\|A\| = \big\|(A)\big\|$ for $A \in L(R^\ell \times V)$.[31] These isometric properties of π hold for several different norms in $R^\ell \times V$ and $L(R^\ell \times V)$ with the corresponding norms in the space R^n of n-tuple of numbers and the space $R^{n \times n}$ of $n \times n$ matrices. It follows that

$$\sum_{t=1}^{T} \|Q f_t\|^2 = \operatorname{trace}\left(\sum_{t=1}^{T} (f_t)(f_t)'\right),$$

since $\|Q f_t\|^2 = \|(f_t)\|^2$ for $t = 1, \ldots, T$.

[31] Here we use as elsewhere in the paper the same generic notation $\|\cdot\|$ to denote distinctive norms in various spaces. This convention is also adopted throughout the paper.

We may make the FR-squared as large as we want simply by increasing the truncation number m in any basis $(v_{i+\ell})_{i=1}^m$. However, this does not come at no cost. As m gets large, the variance of the estimator $\widehat{A}$ for the autoregressive operator A is expected to increase. It increases often very sharply in many practical applications, and therefore, we also need to examine how fast the variance of $\widehat{A}$ increases as m gets large. Here, we focus on the integrated variance of $\widehat{A}_{HH}$

$$\text{trace}\left[(\widehat{A}_{HH} - \mathbb{E}\widehat{A}_{HH})'(\widehat{A}_{HH} - \mathbb{E}\widehat{A}_{HH})\right]$$

conditional on all other components of $\widehat{A}$ introduced in Section 2.3, which can be consistently estimated by

$$\left[\text{trace}(\widehat{\Sigma})\right]\left[\text{trace}\left(\sum_{t=1}^{T}(f_t)(f_t)\right)^{-1}\right].$$

See Chang et al. (2021b) for more details. We let

$$\text{IVAR} = \text{trace}\left(\sum_{t=1}^{T}(f_t)(f_t)\right)^{-1}$$

in what follows.

To show how important it is in practice to choose a basis, we obtain and compare the FR-squared's and the IVARs based on our basis $(v^*_{i+\ell})_{i=1}^m$ and other bases. As an alternative to our basis $(v^*_{i+\ell})_{i=1}^m$, we consider three other bases given by the orthonormalized moments, histograms, and quantiles, which will be referred to as the *moment basis*, *histogram basis* and *quantile basis*, respectively. The moment basis $(v_{i+\ell})_{i=1}^m$ is obtained by the Gram-Schmidt orthogonalization procedure from the pre-basis defined as $u_i(r) = r^i$ for $i \geqslant 1$ over the interval $[p, q)$ with p and q representing the minimum and maximum values of stock returns in the sample. We call (u_i) the moment basis, since

$$\langle u_i, f_t\rangle = \int r^i f_t(r)dr$$

and $\big(\langle u_i, f_t\rangle\big)$ represents the i-th moments of the stock return distributions given by the densities (f_t) for $i \geqslant 1$.

The histogram basis $(v_{i+\ell})_{i=1}^m$ is given by

$$v_i(r) = \frac{1}{\sqrt{q_i - p_i}}1\{p_i \leqslant r < q_i\},$$

m	FPC Basis	Histogram Basis	Quantile Basis	Moment Basis
1	0.8884	0.0021	0.0015	0.0001
2	0.9613	0.0311	0.6476	0.0048
3	0.9758	0.0158	0.7084	0.0052
4	0.9870	0.1783	0.7568	0.0183
5	0.9915	0.0565	0.7919	0.0190
6	0.9942	0.3632	0.8165	0.0380
7	0.9958	0.1281	0.8401	0.0390
8	0.9967	0.5313	0.8573	0.0590
9	0.9975	0.2202	0.8728	0.0602
10	0.9980	0.6445	0.8848	0.0779
11	0.9984	0.3203	0.8965	0.0793
12	0.9987	0.7203	0.9048	0.0933
13	0.9989	0.4163	0.9126	0.0948
14	0.9991	0.7684	0.9195	0.1053
15	0.9993	0.5099	0.9255	0.1068
16	0.9994	0.8010	0.9315	0.1148
17	0.9995	0.5896	0.9359	0.1163
18	0.9996	0.8251	0.9399	0.1228
19	0.9996	0.6476	0.9437	0.1245
20	0.9997	0.8389	0.9471	0.1341

Table 1: FR2 of four different choices of basis including the functional principal component (FPC) basis, histogram basis, quantile basis and moment basis

where $\big([p_i, q_i)\big)$ is a partition of the support $[p, q)$ of the densities (f_t). As before, we let p and q represent the minimum and maximum values of stock returns in the sample, and obtain the $(m+1)$-number of sub-intervals $\big([p_i, q_i)\big)$ of equal length, from which we take only m indicators as a basis ignoring the first sub-interval. This is because the $(m+1)$ indicators over the $(m+1)$-number of sub-intervals $\big([p_i, q_i)\big)$ are linearly dependent. The quantile basis $(v_{i+\ell})_{i=1}^m$ is defined similarly as indicators over a different set of partition $\big([p_i, q_i)\big)$. The $(m+1)$-sub-intervals $\big([p_i, q_i)\big)$ in the partition are obtained with (q_i) defined as the $i/(m+1)$-th sample quantiles of entire observations for $i = 1, \ldots, m+1$. Similarly, as for the histogram basis, we only include m indicators in the quantile basis.

Table 1 presents the FR-squared's relying on four different choices of basis including the functional principal component (FPC) basis, histogram basis, quantile basis and moment basis. We may clearly see that the FPC basis effectively represents the temporal variation of S&P returns distribution even with the truncation number m as small as $m = 1$. For $m = 2$, it already captures more than 95% of the total temporal variation of the distribution. This is in sharp contrast with all other bases. For any of these bases, the explained proportion of the temporal variation in the distribution of S&P 500 is considerably smaller for each

m	FPC Basis	Histogram Basis	Quantile Basis	Moment Basis
1	0.030	8.279	48.783	2972.226
2	0.407	20546.529	97.532	3567.099
3	2.314	64533.382	136.168	58298.927
4	4.768	104624.287	184.432	63136.865
5	10.756	199842.404	262.499	192798.466
6	21.025	470187.220	321.191	415766.908
7	38.371	771540.927	573.885	556171.467
8	66.274	1059776.323	683.730	762275.109
9	100.808	1372883.503	1506.889	1014200.719
10	155.767	1801998.806	2252.585	1484067.909
11	223.570	2375284.232	11526.641	2083498.294
12	330.760	3305582.818	13066.491	2587962.679
13	448.905	4747248.143	102530.873	3139667.558
14	596.729	6535095.372	97171.602	5811565.107
15	767.026	8027593.697	781137.212	8534069.734
16	991.775	9736105.847	423075.122	15878248.786
17	1248.442	10540587.105	11373474.535	32448718.652
18	1576.804	11474309.474	3068505.013	49585806.513
19	1956.788	11980101.620	225546554.257	84824379.415
20	2472.949	12695040.536	56923577.212	174783112.168

Table 2: IVARs of four different choices of basis including the functional principal component (FPC) basis, histogram basis, quantile basis and moment basis

m and increases at a noticeably slower rate as m. Out of the three, the quantile basis performs relatively better, although it is still not comparable at all with the FPC basis. The moment basis, which is often used to represent distributional dynamics in practice is especially ineffective. It captures less than 5% of the total variation of S&P returns distribution over time even for $m = 20$. The FR2 is expected to strictly increase as m gets large. However, this is not the case for the histogram basis, since it is defined differently for different values of m.

Table 2 provides the IVARs resulting from the choice of the four bases.[32] As explained, the IVAR is asymptotically proportional to the integrated variances of the autoregressive operator estimator $\widehat{A}_{HH}$ conditional on all other components of $\widehat{A}$. The IVARs from the FPC basis are by far smaller than the IVARs from the other three bases. They are simply not comparable. The IVARs based on the FPC basis increases as m gets large, but only at a moderate rate. In contrast, the IVARs of all other bases explode as m increases. For

[32]IVAR only provides relative precision of the estimator for autoregressive functional operator relying on any given basis. In particular, it may not be a most relevant measure if we focus on some specific features like VaR. Unfortunately, however, we do not know how we may systematically measure the effectiveness of each basis to capture fluctuations of such specific features.

instance, if the moment basis is used, the IVARs increase as large as 1.7×10^8 when we set $m = 20$ to explain less than 5% of the temporal variation of S&P 500 distribution. We may compare these figures with the use of the FPC basis with our choice of $m = 3$, which explains approximately 98% with the IVAR given by 2.314. It is clear that using other bases is not an option. Our demonstration here shows that the choice of basis is critically important in studying distributional dynamics. It also suggests that researchers should be cautious when interpreting any empirical evidence of distributional dynamics based on ineffective representations such as moments and indicator functions.

Number of basis	PC Basis	Histogram Basis	Quantile Basis	Moment Basis
1	0.0309	12.9389	18.5492	266.3455
2	0.4072	372.5639	34.7502	313.9543
3	2.3138	765.6190	48.9964	1836.7246
4	4.7670	1793.7837	64.0437	1912.1885
5	10.7594	3425.5253	105.7411	4187.0666
6	21.0604	5021.6921	123.0412	5010.2841
7	38.4103	6172.5796	334.7554	8839.2092
8	66.4541	7369.8165	394.5109	13115.4658
9	101.0445	8719.2907	1365.7257	18943.4063
10	156.5246	10523.1538	1960.5110	27794.6129
11	224.3866	12914.7477	10377.1900	43395.1710
12	333.2454	15558.3464	13910.8308	58669.7040
13	451.5645	18713.6332	65855.4328	77731.8587
14	599.6033	23052.4111	202746.4866	110320.5740
15	772.3732	27374.7554	350429.9574	166160.5353
16	997.5695	31979.2814	1124993.5561	224750.0189
17	1256.5014	37421.7444	3797757.7395	316133.8735
18	1585.2403	43640.6818	12005833.7353	424393.0530
19	1970.1756	51410.0966	47657331.1725	561742.4970
20	2487.9453	61377.5021	233152327.5670	771527.4828

Table 3: $trace(\Gamma_{f \cdot x})$ for each basis

Appendix C Identifying a functional shock

To identify the Var Max shock, denoted e_t^{DST*}, we seek to find the set of weights $w_1^*, w_2^*, \ldots, w_m^*$ that generate the largest increase in stock return volatility at-impact. To that end, let $\psi_1, \ldots, \psi_m$ be the at-impact functional responses to the semi-structural functional shocks $(e_{(\ell+1)t}), \ldots, (e_{(\ell+m)t})$, respectively, and define

$$\psi(w) = w_1\psi_1 + \cdots + w_m\psi_m$$

with $w = (w_1, \ldots, w_m)'$. Then $w^* = (w_1^*, w_2^*, \ldots, w_m^*)'$ is defined as the maximizer of

$$\begin{aligned}\sigma^2(w) &= \int_{-\infty}^{\infty} x^2\big(\phi + \psi(w)\big)(x)dx \\ &= \int_{-\infty}^{\infty} x^2\phi(x)dx + \sum_{k=1}^{m} w_k \int_{-\infty}^{\infty} x^2\psi_k(x)dx\end{aligned}$$

with respect to w subject to $\sum_{k=1}^m w_k^2 = 1$, where ϕ is a reference density which could be the temporal mean density $\overline{f}$ or the observed density f_T at the end of the sample period.

Without loss of generally, suppose that the densities of interest are temporally demeaned, i.e., deviations from the temporal mean density. Then, the functions $\psi_1, \ldots, \psi_m$ are not densities on their own, and solving this maximization problem is straightforward. Let

$$\tau_k = \int_{-\infty}^{\infty} x^2(\psi_k)(x)dx$$

for $k = 1, \ldots, m$, then $w^* = (w_1^*, w_2^*, \ldots, w_m^*)'$ is given by

$$w_k^* = \frac{\tau_k}{\sqrt{\tau_1^2 + \cdots + \tau_m^2}}$$

for $k = 1, \ldots, m$. The Var-Max shock may therefore be defined explicitly.

In our empirical study with $m = 3$ and $\phi = f_T$, we find that the Var Max shock described above is identified with the weights $w^* = (w_1^*, w_2^*, w_3^*) = (-0.72, 0.66, -0.20)$ normalized by its norm $\|w^*\|$ so that the distributional shock (e_t^{DST*}) also has the unit variance.

Appendix D Additional Results

As a robustness check, we repeat the analysis with two FPCs instead of three, i.e., $m = 2$. All qualitative conclusions remain unchanged, while minor quantitative differences occur. One notable difference is in the IRFs of skewness and kurtosis to the FS shock in Figure D.3. For these responses, we find a smaller average response with two FPCs compared to three. However, the remaining results in this figure are very similar to those with $m = 3$.

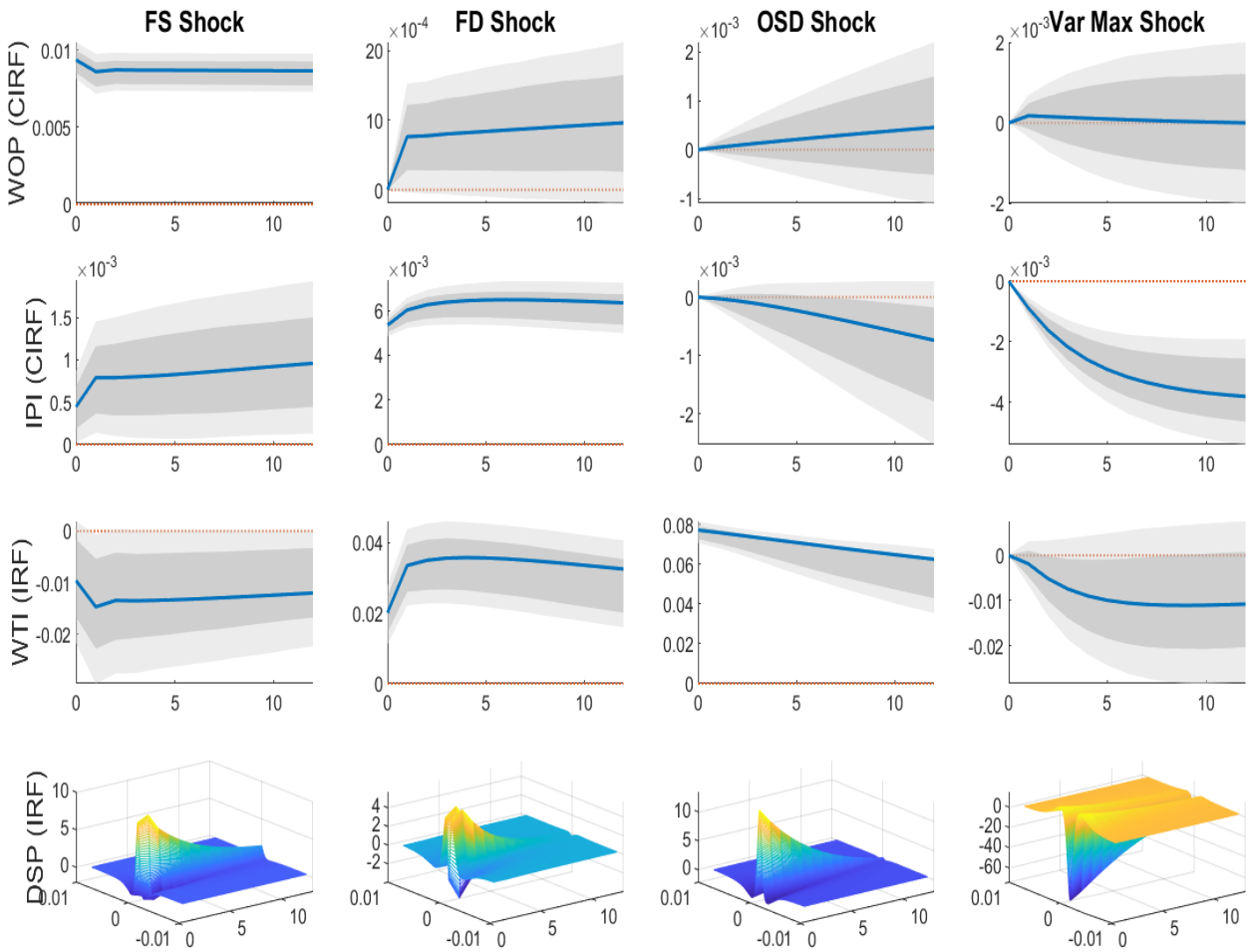


Figure D.1: Full set of of impulse response functions and surfaces from MVAR model.

Notes: The shocks are flow-supply (FS), flow-demand (FD), oil-specific demand (OSD) and Var Max (see section 2.4 for details). The 68% (dark) and 90% (light) confidence bands (colored areas) are calculated using a bootstrapping method.

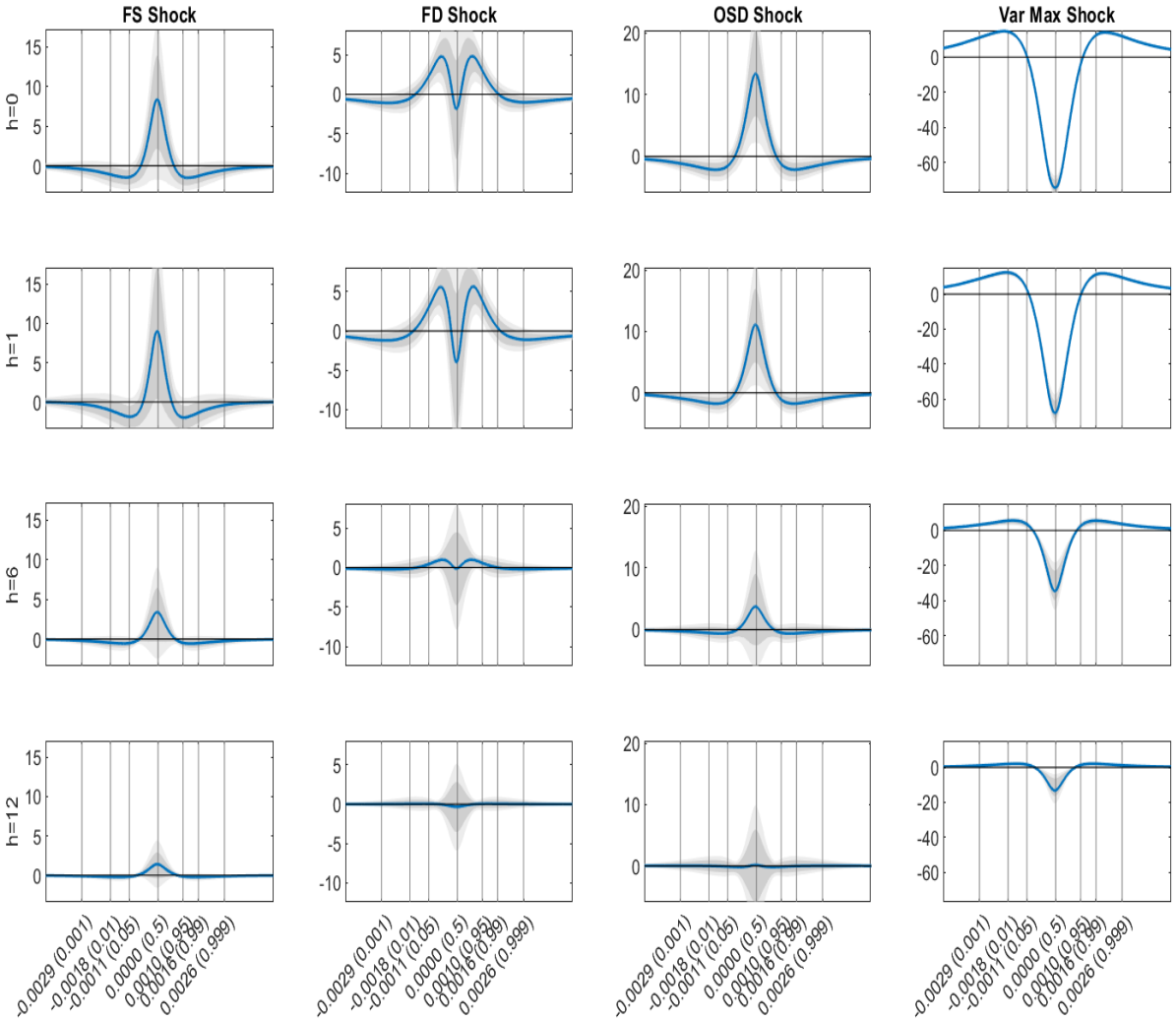


Figure D.2: Functional impulse response functions (FIRFs) for the S&P 500 returns distribution at selected horizons.

Notes: The shocks are flow-supply (FS), flow-demand (FD), oil-specific demand (OSD) and Var Max (see section 2.4 for details). The 68% (dark) and 90% (light) confidence bands (colored areas) are calculated using a bootstrapping method. The solid line is the sample response.

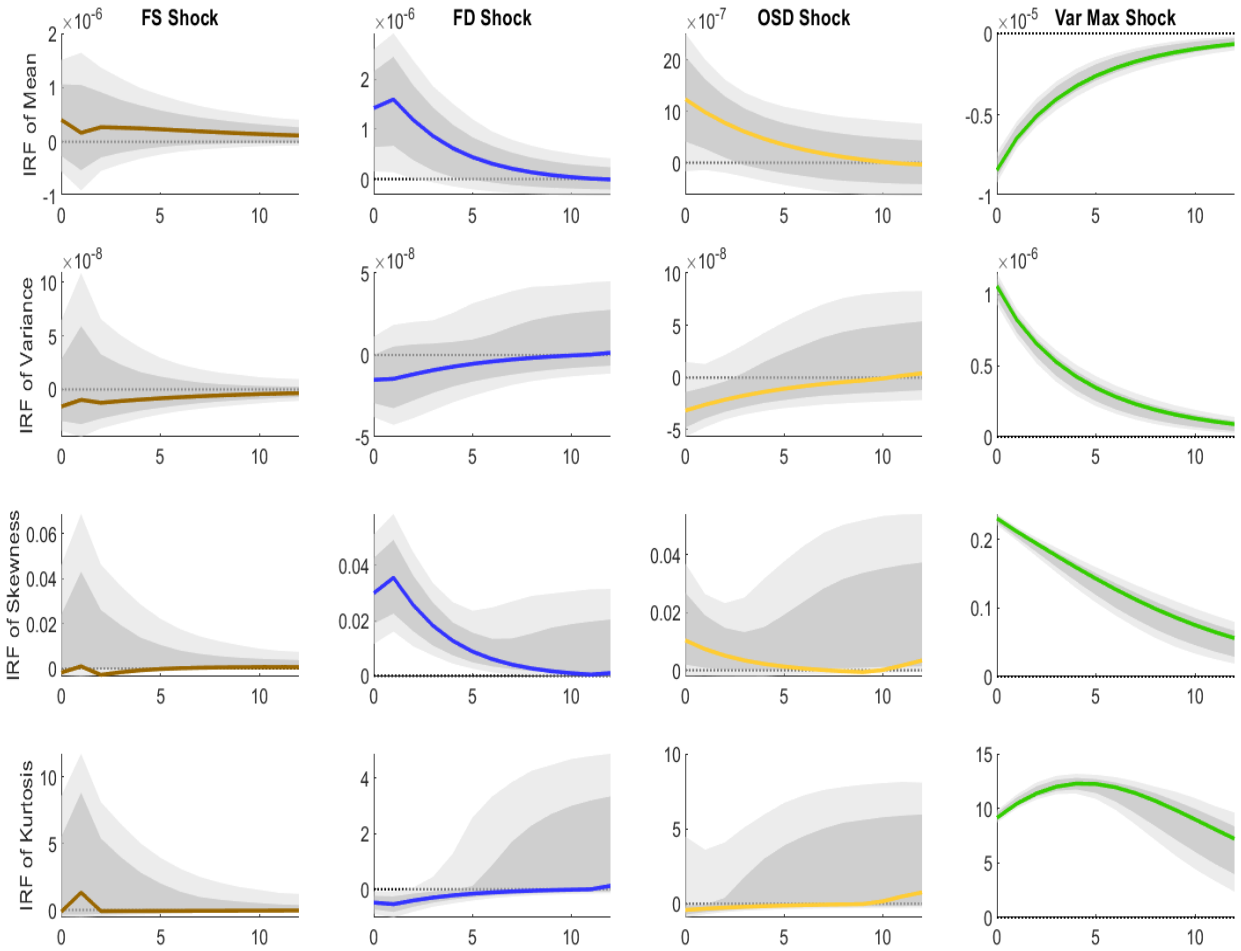


Figure D.3: Impulse response functions for the four moments: mean, variance, skewness and kurtosis of the S&P 500 returns distribution.

Notes: The shocks are flow-supply (FS), flow-demand (FD), oil-specific demand (OSD) and Var Max (see section 2.4 for details).

Quantifying Distributional Effects

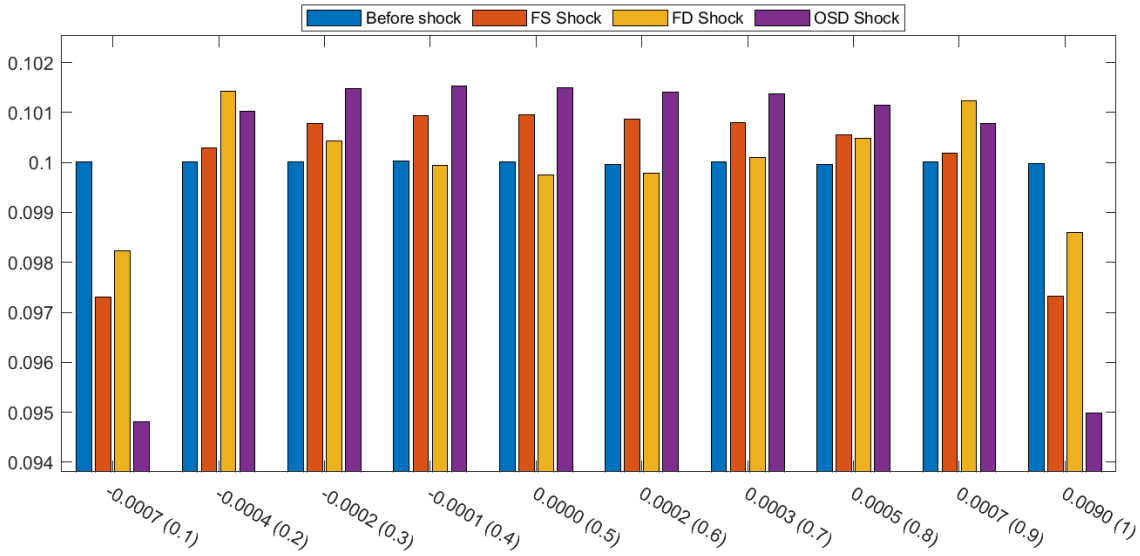


Figure D.4: Response of S&P Distribution to oil market shocks (h=0)
Note: The shocks are flow-supply (FS), flow-demand (FD), oil-specific demand (OSD)

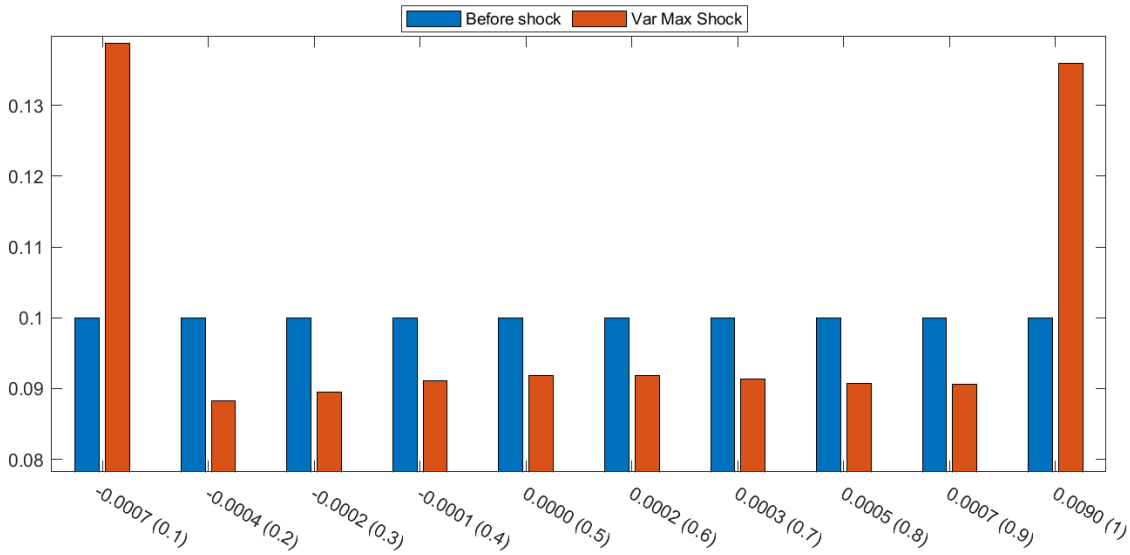


Figure D.5: Response of S&P Distribution to Var Max Shock (h=0)

Value at Risk analysis

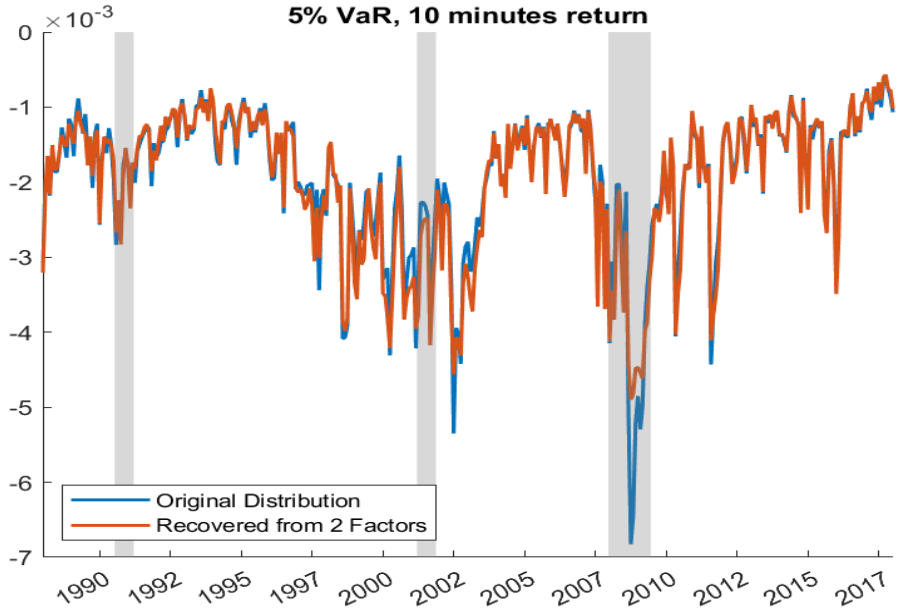
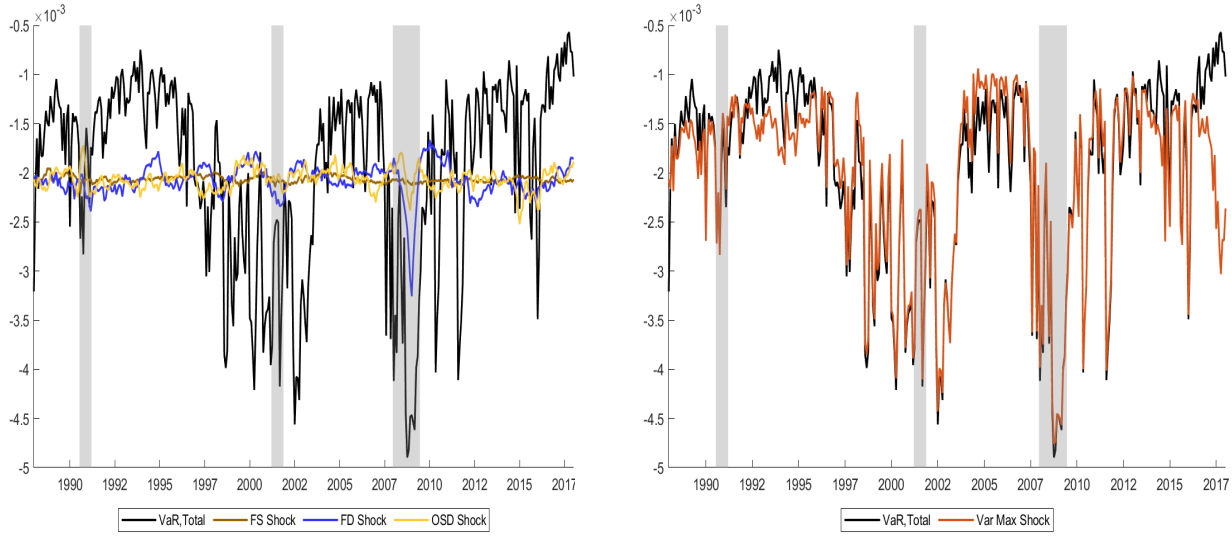


Figure D.6: Value at risk (VaR) for the S&P 500 returns distribution over the period 1988:02-2018:01.



(a) Oil market shocks (b) Stock market shocks

Figure D.7: Counterfactual: Value at risk (VaR) for the S&P 500 returns distribution over the sample 1988:01-2018:01 when only one shock is realized.

Notes: The shocks are flow-supply (FS), flow-demand (FD), oil-specific demand (OSD) and maximum variance (Var Max).

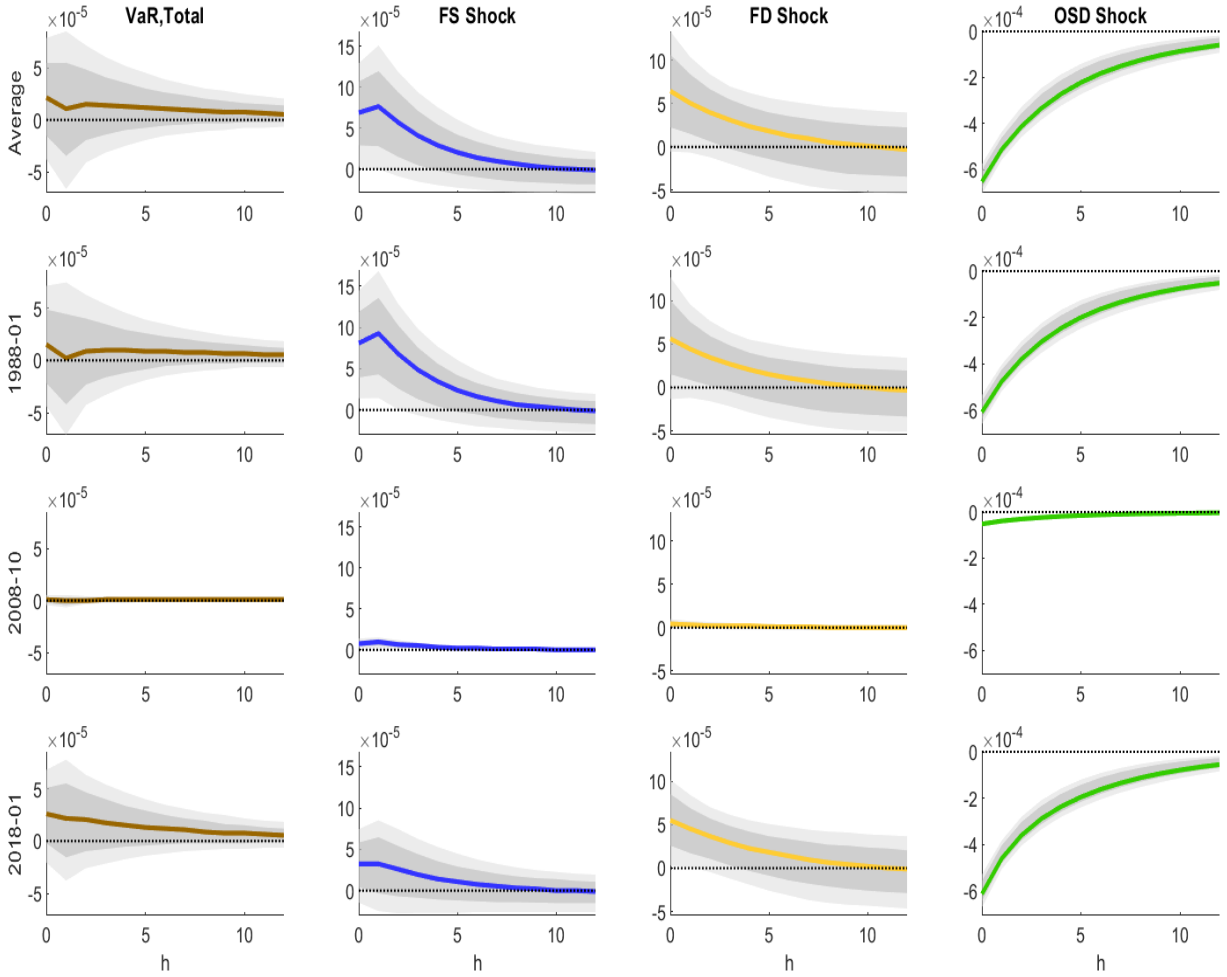


Figure D.8: Impulse response functions for the value-at-risk (VaR) for the S&P 500 returns distribution to the oil market demand and supply shocks using the MVAR model

Notes: The shocks are flow-supply (FS), flow-demand (FD), oil-specific demand (OSD) and Var Max