Brunel University London

Department of
Economics and Finance

Working Paper No. 2102

Economics and Finance Working Paper Series

Helmut Herwartz and Fang Xu

Structural transmissions among investor attention, stock market volatility and trading volumes

February 2021

http://www.brunel.ac.uk/economics

Structural transmissions among investor attention, stock market volatility and trading volumes

Helmut Herwartz[a], Fang Xu[*,b]

[a]*Georg-August-University of Goettingen, Humboldtallee 3, 37073 Goettingen, Germany*
[b]*Brunel-University of London, UB8 3PH, Uxbridge, United Kingdom*

Abstract

We employ data-based approaches to identify the transmissions of structural shocks among investor attention measured by Google search queries, realized volatilities and trading volumes in the US, the UK and the German stock market. The two identification approaches adopted for the structural VAR analysis are based on independent component analysis and the informational content of disproportional variance changes. Our results show robust evidence that investors' attention affects both volatilities and trading volumes contemporaneously, whereas the latter two variables lack immediate impacts on investors' attention. Some movements in investors' attention can be traced back to market sentiment.

Key words: Search engine data, realised volatility, structural VAR.

JEL Classification: G10, G14

[*]Corresponding author
Email addresses: hherwartz@uni-goettingen.de (Helmut Herwartz), fang.xu@brunel.ac.uk (Fang Xu)

1. Introduction

The Google search engine has become an integral tool to find information for more and more people around the globe. The aggregate search frequency in Google provides a direct measure of individual/retail investor's attention (Da, Engelberg, and Gao, 2011). When an individual searches for DOW in Google, she/he certainly pays attention to it. Empirical analysis has shown that stock return volatilities and trading volumes are positively associated with Google search queries (Vlastakis and Markellos, 2012). Moreover, changes in search queries today can help to explain future changes in return volatility. Dimpfl and Jank (2016) show that search queries Granger-cause volatility, and including search queries in models of realized volatility improves volatility forecasts out-of-sample. It is, however, unclear what is the propagation mechanism of the shocks. Does a volatility shock trigger search queries (investors' attention), or/and is it the increased investor's attention (reflected in a positive shock in search queries) that triggers more trading and thereby higher volatility?

On the one hand, there are several theoretical underpinnings for the impact of investors' attention on volatility. If investors pay more attention, new information is quickly incorporated into prices and, thus, can induce high return volatility (Andrei and Hasler, 2015). Moreover, since retail investors are often regarded as uninformed noise traders, their trading can lead to excessive volatilities of asset prices according to the noise trader model (De Long et al., 1990). Similarly, exogenous shocks of the fundamental prices can be interpreted by noise traders as a potential future trend in agent-based models (Lux and Marchesi, 1999). When there is a large fraction of noise-trader agents in the market, the volatility of the stock becomes larger. Thus, the higher the volume of the search queries, the more likely it is that retail investors are actively trading, and the larger are the volatilities of the relevant stocks. On the other hand, volatile movements in the stock markets have been frequently featured in the news, specially in downturn periods. This could attract retail investors' attention and increase the count of search queries for the stock indices. The recursive structural vector regressive model (SVAR) in Dimpfl and Jank (2016), for example, builds upon the hierarchical assumption of an immediate impact of volatility on search queries.

This paper contributes to the literature by estimating the contemporaneous relationship between search queries and return volatilities. For such a purpose, ad-hoc impositions of triangular structures (e.g., in terms of lower triangular Cholesky factors) for SVAR models leave no room for the data to object against the model implied hierarchy. In this paper, we use data-driven identification approaches, which let data determine the latent structural relationships. Our analysis is based on daily Google search queries for US, UK and German stock market indices from 2006 to 2011. Our data exhibit both deviations from a conditionally multivariate Gaussian model and conditional changes in the covariance structure. Therefore, we exploit the uniqueness of independent structural shocks (Matteson and Tsay, 2017), and the informational content of disproportional variance changes of the model implied structural shocks (Normadin and Phaneuf, 2004; Bouakez and Normadin, 2010; Lanne and Saikkonen, 2007) for SVAR identification.

Results from both identification approaches and the three markets point to the same evidence — shocks in Google search queries affect return volatilities immediately, while shocks in volatilities exert an only mild (if any) instantaneous effect on search queries. Therefore, what underlies the positive correlation observed among search queries and volatility is the increased investor's attention which triggers more trading and thus higher volatility. Introducing the trading volume as a third variable into the SVARs confirms that search queries also affect the trading volume simultaneously. The results from this paper also provide a guide for the order of the variables in the recursive SVARs, namely search queries as the first variable and realized volatility as the second in trivariate SVARs. Using a different ordering of the variables, the model implied impulse responses are at the risk to provide a misleading perspective on impact and dynamic relations within the triad of search queries, volatilities and volumes.

After highlighting the contemporaneous impact of retail investors' attention on stock market volatility, we further explore if market sentiment is partially behind movements in investors' attention. We estimate bi-variate SVARs using Goggle search queries on DOW and the FEARS sentiment index from Da et al. (2015), the latter of which is available for the US. The results show that changes in the market sentiment have a significant contemporaneous impact on variations in retail investors' attention. De Long et al. (1990) demonstrate that changes in investors' sentiment can lead to more noise trading and excess volatility, if uninformed noise traders base their trading decision on sentiment. Da et al. (2015) confirm the positive contemporaneous relationship between sentiment and the market volatility empirically. Our results are in line with the view that the retail investor's attention could be part of this transmission channel.

The remainder of this paper is organized as follows: The next Section provides a brief formalization of the structural VAR model and sketches the data-based identification schemes. Section 3 introduces the data and provides some preliminary empirical analyses. Section 4 addresses structural and dynamic empirical relationships in the triad of search queries, realized volatilities and trading volumes. Section 5 looks at the relationship between search queries market sentiment. Section 6 summarizes our main findings and concludes.

2. Data-based identification of SVARs

This section provides an outline of the VAR model in its reduced form and in SVAR representation. The identification problem is described, and subsequently we sketch two alternative data-based identification schemes.

2.1. The structural VAR

Consider a p-th order autoregressive model for the K-dimensional system of random variables y_t, i.e.,

$$\begin{aligned} y_t &= \nu + A_1 y_{t-1} + \ldots + A_p y_{t-p} + u_t, && (1)\\ &= \nu + A_1 y_{t-1} + \ldots + A_p y_{t-p} + B\varepsilon_t, && (2)\\ \Leftrightarrow \quad A(L) y_t &= \nu + B\varepsilon_t, \quad t = 1, \ldots, T, && (3) \end{aligned}$$

with vector-valued intercept terms ν, $K \times K$ parameter matrices A_i, the backshift operator L such that $Ly_t = y_{t-1}$, $A(L) = I_K - A_1 L - \ldots - A_p L^p$, and I_K denoting the identity matrix. By assumption, the model is causal, i.e., $\det(A(z)) \neq 0$ for all $|z| \leq 1$.

The representations in (1) and (2) differ in terms of their stochastic model components. *Reduced-form residuals* u_t in (1) are of mean zero ($E(u_t) = 0$) and subject to contemporaneous correlation with covariance Σ. Residuals ε_t in (2) are orthogonal with $E(\varepsilon_t) = 0$ and $\text{Cov}[\varepsilon_t] = I_K$. By implication, the covariance matrix Σ aligns with the decomposition $\Sigma = BB'$, where B is a nonsingular $K \times K$ parameter matrix. Orthogonalized residuals ε_t qualify as 'structural shocks' if an identified parameter matrix B benefits from a sound theoretical underpinning of its implied effect structure, which is typically summarized in the form of impulse response functions (IRFs). Unlike reduced form residuals u_t, the structural shocks ε_t cannot be recovered uniquely by means of OLS estimation. Since the space of potential covariance decompositions $\Sigma = BB'$ is infinite, it deserves further assumptions to identify the structural parameters in B.

To design a space of alternative covariance decompositions, let Q denote a rotation matrix ($Q \neq I_K, QQ' = I_K$) which is parameterized with (vector-valued) rotation angle(s) θ, i.e. $Q = Q(\theta)$. Moreover, D is the lower triangular Choleski factor of Σ, such that $\Sigma = DD'$. Then, a space of covariance decompositions results from rewriting the residual form covariance $\Sigma = DQ(\theta)Q'(\theta)D'$ as

$$\mathcal{B} = \{B | BB' = \Sigma, B = DQ(\theta)\}, \tag{4}$$

where the representation $B(\theta) = DQ(\theta)$ points to $B(\theta)$ as a specific member of $\mathcal{B}$. The prime aim of SVAR analysis is to identify a particular structural matrix $B = B(\theta)$, since this matrix formalizes the instantaneous impacts of the structural shocks ε_t on the observable variables in u_t (or y_t). Henceforth, the dependence of the structural parameter matrix B on θ is suppressed whenever appropriate.

2.2. Marginal effects

Typical parameters of the structural matrix B, denoted b_{ij}, quantify the instantaneous impact of shocks ε_{jt} on reduced form residuals u_{it} (or y_{it}). A reformulation of model (2) allows for an explicit formalization of the implied contemporaneous linkages among the variables in y_t, i.e.,

$$B^{-1}y_t = B^{-1}\nu + A_1^* y_{t-1} + A_2^* y_{t-2} + \ldots + A_p^* y_{t-p} + \varepsilon_t, \tag{5}$$

where $A_i^* = B^{-1}A_i$, $i = 1, 2, \ldots, p$. Unlike the model in (2), the left hand side of (5) is explicit on the marginal effect patterns that involve the variables in y_t contemporaneously. To provide these effects in normalized form, define Ω_t as the information set comprising the process information up to time t, i.e. $\Omega_t = \{y_t, y_{t-1}, \ldots\}$, and let $b^{(ij)}$ denote a typical element of B^{-1}. Conditional on Ω_{t-1}, nonlinear transformations of the elements in B^{-1} describe the marginal causal effects. Consider, for instance, the bivariate case $K = 2$. From (5) and

$$B^{-1}y_t = \begin{pmatrix} b^{(11)} & b^{(12)} \\ b^{(21)} & b^{(22)} \end{pmatrix} \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix}$$

we obtain after normalization

$$E[y_{1t}|y_{2t} \cup \Omega_{t-1}] = -\frac{b^{(12)}}{b^{(11)}} y_{2t} + \omega_{t-1}^{(1)} \text{ and } E[y_{2t}|y_{1t} \cup \Omega_{t-1}] = -\frac{b^{(21)}}{b^{(22)}} y_{1t} + \omega_{t-1}^{(2)}. \tag{6}$$

Estimates of $\omega_{t-1}^{(1)}$ and $\omega_{t-1}^{(2)}$ in (6) are obtained from the VAR parameters and information in Ω_{t-1}. If B could be observed or properly identified, one can directly recover causal effects that are typically in the focus of single equation regression models. Conditional on Ω_{t-1}, the marginal effect of y_{2t} on y_{1t} is measured by $-b^{(12)}/b^{(11)}$, and the marginal effect of y_{1t} on y_{2t} is measured by $-b^{(21)}/b^{(22)}$.

2.3. Identification

For purposes of structural analysis, triangular-model structures (i.e. Cholesky factors) have become popular (Sims, 1980). However, to justify the choice $B = D$ one has to make a-priori assumptions to motivate a particular hierarchical model. Kilian and Lütkepohl (2017) provide an up-to-date textbook treatment of identification schemes in SVAR analysis. In this work we avoid the setting of a-priori triangular model structures, and use instead data-based identification schemes to uncover structural transmission patterns.

Two specific statistical properties of ε_t have become prominent in identifying structural information, namely (i) independently distributed and non-Gaussian shocks, and (ii) shocks with informative (i.e. non-proportional) changes in variance. These properties provided the motivation for two classes of data-based identification. In this work, we identify structural models by means of one representative of each class. First, we exploit the independence property following Matteson and Tsay (2017). Second, we focus on the assumption of generalized autoregressive conditionally heteroskedastic structural shocks (GARCH) as suggested, e.g., by Normadin and Phaneuf (2004), Bouakez and Normadin (2010) and Lanne et al. (2017).[2] We next brief outlines of the two alternative (or complementary) identification schemes employed in this work.

(i) Independent components

Independence based identification builds upon a fundamental result of Comon (1994) stating that a vector of independent components ε_t allows the unique recovery of B from reduced form residuals u_t, if at most one independent component ε_{it} exhibits a Gaussian distribution. For an intuitive illustration, let's assume that a false member of the class of covariance decompositions $\mathcal{B}$ in (4), denoted $\widetilde{B}$, is used for the structural analysis. Then, the corresponding structural shocks read as

$$\tilde{\varepsilon}_t = \widetilde{B}^{-1} u_t = \widetilde{B}^{-1} B \varepsilon_t = J \varepsilon_t \neq \varepsilon_t,$$

where $u_t = B\varepsilon_t$ and $J = \widetilde{B}^{-1} B \neq I_N$. By implication, if the elements in ε_t are independent and non-Gaussian by assumption, the elements $\tilde{\varepsilon}_{k,t}$ in $\tilde{\varepsilon}_t$ obtain as linear combinations of $\varepsilon_{k,t}$, and therefore lack independence. Put differently, to recover the true independent, non-Gaussian shocks ε_t from reduced form residuals u_t, it is essential to employ the correct structural matrix B. Based on the uniqueness of independent non-Gaussian components, several approaches have been suggested for SVAR identification. These approaches differ with respect to the investigated space of structural matrices (e.g. the comparison of alternative recursive structures in Moneta et al., 2013) and the rigidity of underlying parametric model assumptions (e.g. the imposition of distributional assumptions to enable

[2]For the implementation of data-based identification and further computations we use the R-package 'svars' (*https://CRAN.R-project.org/package=svars* Lange et al., 2019, see also Lange et al. (2020)). We use the modules 'id.dc' (independence of non-Gaussian shocks) and 'id.garch' (conditionally heteroskedatic shocks). The module 'id.dc' builds upon the function *steadyICA* from the `R` package `steadyICA` (Risk et al., 2015).

(pseudo) ML estimation in Gouriéroux et al., 2017; Lanne et al., 2017). As our daily data provide huge sample information and exhibit heterogeneous second-order properties, we refrain from parametric pseudo ML estimation. Instead we pursue a semi-parametric estimation by targeting at implied shocks ε_t which provide weakest evidence against the null hypothesis of independence in terms of a suitable test statistic. More specific, we follow Matteson and Tsay (2017) who suggest to obtain an estimate of the structural parameters from solving the minimization problem

$$\widehat{B} = \text{argmin}_{B \in \mathcal{B}} \{\text{Joint dependence of elements in } \varepsilon_t, \text{ with } \varepsilon_t = B^{-1}(\theta) u_t\}, \qquad (7)$$

where the degree of dependence is quantified in terms of the so-called distance covariance statistic of Székely et al. (2007). Henceforth, we denote the structural matrix estimates based on an independence assumption as B_{DC}.

(ii) Heteroskedastic shocks

Rigobon (2003) has pioneered the identification of heteroskedastic structural shocks.[3] Going beyond the stylized covariance break model of Rigobon (2003), the second identification scheme that we employ in this study builds upon patterns of generalized autoregressive conditional heteroskedasticity (GARCH, Bollerslev, 1987) as proposed by Normadin and Phaneuf (2004), Lanne and Saikkonen (2007) and Bouakez and Normadin (2010). In this framework, time varying covariances are formulated as

$$\begin{aligned} E(u_t u_t' | \Omega_{t-1}) &= \Sigma_t \\ &= B \Lambda_t B', \end{aligned} \qquad (8)$$

[3]For simplicity of demonstration of this identification scheme, assume that there are two distinct covariances (denoted as $\Sigma^{(1)}$ and $\Sigma^{(2)}$) that characterize the reduced form residuals u_t across two disjoint subsamples. From these subsamples, one can estimate two sets of $N(N+1)/2$ empirical (co)variances. The two covariance matrices allow for a reparameterization as $\Sigma^{(1)} = BB'$ and $\Sigma^{(2)} = B\Lambda B'$, where Λ is a diagonal matrix. This representation comprises $N^2 + N = N(N+1)$ unknown parameters that formalize the structural model. Hence, it becomes possible to map the estimated (co)variances one-to-one into the parameter space of the structural model. For a unique mapping, however, it is essential that the diagonal elements in Λ are distinct from each other. Otherwise single structural shocks remain unidentified.

where $\Lambda_t = \text{diag}(s^2_{1,t}, ..., s^2_{K,t})$ is a diagonal matrix and $s^2_{k,t}$ denotes GARCH-type conditional variance processes capturing the conditional second order properties of the structural shocks. Assuming a parsimonious GARCH(1,1) specification and noticing that $E[\varepsilon^2_t] = \text{Var}[\varepsilon_{kt}] = 1\,\forall k$, by assumption, the individual conditional variances $\text{Var}[\varepsilon_{kt}|\Omega t-1] = s^2_{k,t}$ exhibit a dynamic structure as

$$s^2_{k,t} = (1-\gamma_k - g_k) + \gamma_k \varepsilon^2_{k,t-1} + g_k s^2_{k,t-1}, \qquad k = 1,..,K. \tag{9}$$

Under suitable distributional and parametric restrictions ($\gamma_k > 0$, $g_k \geq 0$ and $\gamma_k + g_k < 1$), the GARCH processes $\varepsilon_{k,t}$ are covariance stationary (Milunovich and Yang, 2013). Sentana and Fiorentini (2001) have shown that the structural parameters in B can be determined uniquely by means of (quasi) ML estimation, if at least $K-1$ structural shocks exhibit dynamic GARCH-type variance patterns. Henceforth, we denote the structural matrix estimates based on changes of variance as B_{GARCH}.

Irrespective of an applied identification scheme, any identified matrix B is unique only up to the ordering and signs of its columns. With $b_{\cdot i}$ denoting a typical column of B it is immediate to observe that the reduced form covariance (Σ) allows for a representation $\Sigma = BB' = \sum_{i=1}^{K} b_{\cdot i} b'_{\cdot i}$. In the bivariate case, for example, the matrices $[b_{\cdot 1}\ b_{\cdot 2}]$, $[-b_{\cdot 1}\ b_{\cdot 2}]$ or $[-b_{\cdot 2}\ b_{\cdot 1}]$ imply the same covariance $BB' = \Sigma$. Hence, the explicit exposition of structural parameter estimates deserves a suitable guideline. We follow the convention to document impact effects of positive shocks. In case that a particular structural matrix candidate obtained from SVAR estimation has a negative diagonal element (i.e., $b_{ii} < 0$), $b_{\cdot i}$ is multiplied with minus one. To achieve a unique column ordering, we choose from the set of alternative column orderings the which implies the largest sum of diagonal elements of B (Lütkepohl and Netšunajev, 2017). Hence, an identified structural shock is supposed to exert the strongest effect on the particular variable to which it is primarily associated.

3. Data and preliminary analysis

In this section, we introduce the data and conduct a preliminary analysis with recursive SVARs.

Table 1: Descriptive statistics

	DOW			DAX			FTSE		
	SQ	RV	VO	SQ	RV	VO	SQ	RV	VO
	Raw data								
Min.	0.294	0.272	52.64	0.437	0.447	1.593	0.523	0.703	66.77
Max.	11.27	767.3	673.0	8.675	689.6	494.0	8.257	1261.6	2757.3
Mean	0.999	18.24	234.0	1.000	17.52	136.5	1.000	20.70	1282.1
S.D.	0.458	1.202	0.344	0.318	1.014	0.373	0.318	1.052	0.385
Skewness	5.206	8.299	1.474	6.802	9.201	2.155	5.967	12.94	0.283
Kurtosis	48.14	113.02	6.089	66.66	131.04	11.00	54.29	250.74	2.525
	Data in natural logarithms								
Min.	-1.225	-8.210	3.963	-0.829	-7.713	0.466	-0.648	-7.260	4.201
Max.	2.422	-0.265	6.512	2.160	-0.372	6.203	2.111	0.232	7.922
Mean	-0.131	-4.841	5.395	-0.069	-4.666	4.849	-0.067	-4.592	7.091
S.D.	0.458	1.202	0.344	0.318	1.014	0.373	0.318	1.051	0.385
Skewness	1.248	0.415	0.163	2.374	0.460	-1.016	2.300	0.585	-1.215
Kurtosis	5.690	3.114	3.810	12.86	3.601	18.58	11.24	3.793	7.768

Notes: Data for RV is multiplied with 1000 to facilitate visualization.

Our analysis focuses on the US stock market index (the Dow Jones Industrial Average - DJIA), the German stock market index (DAX) and the UK stock market index (FTSE 100). The daily realized volatilities for the three indices are obtained from Oxford-Man Institute.[4] It is the sum of the squared intraday log-price changes of the index over 10 minute intervals (Andersen et al., 2001; Barndorff-Nielsen and Shephard, 2002). Daily data on trading volumes have been obtained from Datastream. It is the total number of shares traded from the underlying stocks of the corresponding index per day in millions. With regard to data on Google search queries, those for the keywords "DOW" (US search queries) from the 3rd July 2006 to the 30th December 2011 are from Dimpfl and Jank (2016). Search queries for the keywords "FTSE" (UK stock market) and "DAX" (German stock market) from the 3rd July 2006 to the 30th June 2011 are from Dimpfl and Jank (2011).[5] These are the longest time series of search query daily data that we

[4]We downloaded the data from https://realized.oxford-man.ox.ac.uk/ and multiplied them by 100 so that the scale is comparable to the one from the realized volatilities in Dimpfl and Jank (2016).

[5]We were not able to use the Google search queries for the keywords "CAC" (French search queries) as daily data on trading volumes for CAC 40 is not available in Datastream for the time period considered.

can obtain for each market and that were directly downloaded from Google trend.[6] The time periods of available data on search queries determine the sample period of the SVAR models in the analysis.

Table 1 presents descriptive statistics for the search queries (SQ), realized volatilities (RV), and trading volumes (VO). For realized volatility, the raw data are heavily skewed and have excessive kurtosis. Applying the log transformation, however, reduces the skewness close to zero and the kurtosis close to three. The log transformation also helps to reduce the high skewness and larger kurtosis of search queries. We use all variables in natural logarithms in our analysis. The estimated AR coefficient matrices in the (reduced form) VAR models are similar to those in Dimpfl and Jank (2011) and Dimpfl and Jank (2016). They are not reported here for space considerations.

The linear relation between structural shocks and observable variables is unique if the former are independent and non-Gaussian distributed. Hence, the applicability of the identification of independent components relies on the testable assumption of non-Gaussianity. Unreported diagnostic results from Jarque-Bera tests of the null hypothesis of joint Gaussianity of the reduced form model residuals (i.e. p-values below 0.01%) indicate that model residuals are at odds with the assumption of joint Gaussianity.[7] Since alternative recursive models imply quite distinct hierarchical/causal patterns of shock transmission, it is interesting to unravel in how far these transmissions apply to independent shocks or only to orthogonalized model residuals. Accordingly, independence diagnosis provides indicative information on alternative variable orderings in the VAR model. We use distance covariance statistics to assess the dependence of orthogonalized model residuals implied by lower triangular covariance factors under two and six alternative variable orderings in bi- and trivariate VARs, respectively.[8] Given strong evidence against mul-

[6]We are grateful to Thomas Dimpfl and Stephan Jank for providing us with these data. They downloaded the data from Google trend directly when it was possible to download daily data for a period of more than 270 days. From October 2018, daily Google trends data are only provided for a 270-day period and a direct combination of each 270-day data set is not feasible, as each data set is standardized with regard to a particular reference date.

[7]An implicit indication of leptokurtic model residuals and hence of non-Gaussianity obtains from the estimation of GARCH-type variance processes in the context of identification via heteroskedasticity. As displayed in Table 7 below, all GARCH processes fitted to respective structural shocks indicate clear evidence for GARCH, and, hence, against both marginal and joint Gaussianity.

[8]For testing independence of orthogonalized model residuals, we use the default implementation of the 'gmul-

Table 2: Diagnostic results

	(SQ,RV)	(RV,SQ)	(SQ, RV,VO)	(SQ, VO,RV)	(RV, VO,SQ)	(RV, SQ,VO)	(VO, RV,SQ)	(VO, SQ,RV)
				DOW				
stat.	0.244	3.715	0.149	0.702	2.849	2.488	3.247	2.030
p-value	0.500	0.100	4.900	0.100	0.100	0.100	0.100	0.100
				DAX				
stat.	0.027	2.239	0.127	0.800	1.831	1.700	2.274	1.537
p-value	24.9	0.100	7.20	0.100	0.100	0.100	0.100	0.100
				FTSE				
stat.	0.431	3.576	0.653	0.510	3.025	3.103	2.858	0.804
p-value	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100

Notes: This table documents diagnostic results for orthogonalized model residuals from lower triangular models applied to distinct variable orderings. Orderings are indicated in the first line. Distance covariance test statistics and p-values are multiplied with 100.

tivariate Gaussian models, it is not surprising to see that most variations of variable orderings in lower triangular models obtain orthogonalized model residuals which lack independence.

As can be seen in Table 2 for all markets, orderings with the search queries not in the first position obtain strong evidence against the null hypothesis of independence (p-values below 1%). For the few cases where the null hypothesis of independence cannot be rejected, the search query variable is in the first position throughout, and the realized volatility variable is ordered second. Hence, from the set of potential hierarchical models the particular order where shocks to search queries have an immediate impact on the remaining variable(s) of the dynamic system seems best in line with the assumption of independent shocks. Specifically, for such triangular covariance decompositions the p-values for the German market with the order - (SQ, RV) - in the bivariate VAR and the order - (SQ, RV, VO) - in the trivariate VAR are in excess of 10% and 5%, respectively. The p-value for the US market with the order - (SQ, RV, VO) - is about 5%. With these exceptions, however, hierarchical models generally fail to yield structural shocks which can be reasonably considered as independent. Hence, it is of further interest to investigate if unrestricted covariance decompositions allow the retrieval of unique independent structural shocks.

tidcov' function in the R package steadyICA (Risk et al., 2015). Respective p-values are determined by means of a permutation test with 999 replications ('permTest').

4. Data-driven SVARs

In this section we discuss the model selection, present structural parameter estimates and results from the data-driven SVARs. We also show some further model diagnostics which underpin the informational content of independent components and disproportional covariance changes for model identification.

4.1. Model selection

Empirical estimates from the two alternative identification approaches (B_{DC} and B_{GARCH}) are quite similar for all markets under scrutiny. For all markets, bivariate specifications identified by the two alternative data-based identification schemes obtain shocks with almost complete correlation, i.e., correlation estimates which are beyond 0.99 throughout. With regard to the trivariate specifications, even the smallest correlations between shocks implied by B_{DC} and B_{GARCH} are 0.9768, 0.9453 and 0.9991 for the DOW Jones, DAX and FTSE market, respectively. Moreover, two structural shocks retrieved from trivariate models correlate strongly with their counterparts obtained from bivariate specifications. In this regard, the smallest (out of two) correlation statistics are 0.995, 0.983 and 0.971 for the Dow Jones, DAX and FTSE market, respectively. This supports the robustness of the underlying shocks identified by means of SVARs of alternative dimensions $K = 2, 3$.

Since the identified structural parameter matrices B_{DC} and B_{GARCH} are quite similar for all markets and VAR dimensions, the following discussion of empirical results refers mainly to the implications of B_{DC}. We provide further diagnostic outcomes subsequent to the discussion of estimation results.

4.2. Structural implications of identified models

The estimation of the structural relations confirms that retail investors' attention (measured by Google search queries) affect stock return volatility contemporaneously, while effects of volatility on search queries are negligible. This evidence is robust and can be found for all three stock indices and both identification schemes. Table 3 summarises the estimated structural relations for shocks in search queries and realized volatilities. All matrices are close to a lower triangular

Table 3: Estimated structural relations in bivariate SVARs

		DC		GARCH	
DOW	SQ	**0.1614**	−0.0010	**0.1719**	**0.0173**
		(0.0002)	(0.0072)	(0.0220)	(0.0074)
	RV	**0.3006**	**0.6124**	**0.2412**	**0.6485**
		(0.0272)	(0.0138)	(0.0468)	(0.0200)
DAX	SQ	**0.1442**	−0.0070	**0.1581**	−0.0076
		(0.0003)	(0.0056)	(0.0225)	(0.0082)
	RV	**0.1873**	**0.5124**	**0.2015**	**0.5099**
		(0.0201)	(0.0072)	(0.0409)	(0.0213)
FTSE	SQ	**0.1501**	**−0.0243**	**0.1496**	**−0.0253**
		(0.0011)	(0.0061)	(0.0184)	(0.0083)
	RV	**0.3154**	**0.5016**	**0.3173**	**0.4964**
		(0.0201)	(0.0131)	(0.0410)	(0.0205)

Notes: The table documents the estimates of the structural parameters in B (standard errors in parentheses) determined by means of Distance Covariance criteria (DC, left hand side) and GARCH method (GARCH, right). The search query (SQ) is the first and the realized volatility (RV) the second variable in the analysed series y_t. Parameter estimates with 5% significance are highlighted.

matrix. Numerical values of the upper right estimates are very close to zero albeit statistically significant in some cases. This result confirms the adopted variable ordering for a recursive SVAR model, as indicated by the preliminary analysis.

Our results are consistent with the existing theories. In the noise trader model of De Long et al. (1990), when noise traders are present, asset prices become excessively volatile such that they move more than can be explained by changes of fundamental values. In the agent-based model of Lux and Marchesi (1999), exogenous shocks of the fundamental prices can be interpreted by noise traders as a potential future trend. If there is a large fraction of noise-trader agents in the market, the volatility of the stock increases. In the model of Andrei and Hasler (2015), when investors pay more attention to news, new information is quickly incorporated into prices and thus induces high return volatility. All these theoretical models are supporting our results. Google search queries approximate the retail investors' (noise traders') attention. The higher the volume of the search queries, the more interest (retail) investors show. As more (retail) investors trade, the volatility increases.

Introducing the trading volume as a third variable into our system confirms that shocks in search queries affect the trading volume on impact. Table 4 summarises the estimated structural relations (B matrices) of the trivariate SVARs comprising search queries, realized volatility and trading volumes. The relationships between the first two variables (search queries and realized volatility) documented in Table 4 are very close to those characterizing the bivariate systems (documented in Table 3). While shocks in search queries affect realized volatilities contemporaneously (see estimates of b_{21}), shocks in realized volatilities exert only weak impacts on search queries (see estimates of b_{12}). Now consider impacts on the trading volume. Shocks in search queries affect trading volumes significantly in all three markets (see estimates of b_{31}). Moreover, there is some evidence of significant impacts of shocks in realized volatilities on trading volumes for Dow Jones and DAX as implied by B_{DC} (see estimates of b_{32}).

Now consider the impacts of shocks to the trading volume (see estimates in $b_{\cdot 3}$). In this regard, we do not find any significant impact from shocks in the trading volume on search queries. This result is intuitive. Information about the trading volume is not a popular topic on mass media, as such changes in the trading volume would not draw the attention of the retail/noise investors immediately, and thereby affect the search queries. Shocks in trading volumes show significant impacts only on the realized volatilities of FTSE. The weak indications of impacts of the trading volume on realized volatilities are consistent with the evidence from the literature that information on trading volume does not improve the accuracy of volatility forecasts (e.g. Brooks, 1998).

While estimates of the structural matrix B demonstrate contemporaneous instantaneous effects of structural shocks on the variables of a dynamic system, their numerical interpretations are limited. In contrast, the model-implied marginal effects as displayed in (6) allow for a direct interpretation of effects among the variables conditional on the history Ω_{t-1}. Table 5 summarizes the model-implied marginal effects for all markets and models.[9] As all variables are measured in natural logarithms, the documented estimates allow for an interpretation as elasticities conditional on Ω_{t-1}. Changes in search queries (SQ) have almost a doubled effect on the realized volatility (RV). When the Google search volume on the stock index (relative to the

[9]The marginal representations involve a rescaling of the elements of structural shock vectors $\varepsilon_{k,t}$ by the structural volatilities.

Table 4: Estimated structural relations in trivariate SVARs

		DC			GARCH		
DOW	SQ	**0.1610**	−0.0026	0.0054	**0.1736**	0.0146	0.0134
		(0.0004)	(0.0083)	(0.0043)	(0.0247)	(0.0138)	(0.0098)
	RV	**0.3070**	**0.6083**	0.0169	**0.2433**	**0.6359**	0.1352
		(0.0313)	(0.0163)	(0.0342)	(0.0554)	(0.2279)	(0.1561)
	VO	**0.0928**	**0.0731**	**0.2042**	**0.0800**	0.0465	**0.2153**
		(0.0072)	(0.0115)	(0.0050)	(0.0249)	(0.0509)	(0.0801)
DAX	SQ	**0.1439**	−0.0071	0.0022	**0.1494**	−0.0080	−0.0037
		(0.0004)	(0.0056)	(0.0067)	(0.0317)	(0.0265)	(0.0077)
	RV	**0.1872**	**0.5055**	−0.0104	**0.1948**	0.4818	0.1503
		(0.0196)	(0.0103)	(0.0576)	(0.0642)	(0.2592)	(0.1532)
	VO	**0.0902**	**0.0914**	**0.2416**	**0.1007**	0.0174	0.2557
		(0.0108)	(0.0275)	(0.0112)	(0.0587)	(0.0659)	(0.1372)
FTSE	SQ	**0.1496**	**−0.0235**	0.0099	**0.1430**	**−0.0281**	0.0118
		(0.0011)	(0.0061)	(0.0069)	(0.0195)	(0.0107)	(0.0127)
	RV	**0.3074**	**0.4895**	**0.1232**	**0.3068**	**0.4754**	**0.1445**
		(0.0204)	(0.0138)	(0.0321)	(0.0410)	(0.1116)	(0.0505)
	VO	**0.0489**	0.0190	**0.2324**	**0.0447**	0.0103	**0.2337**
		(0.0100)	(0.0147)	(0.0029)	(0.0190)	(0.0136)	(0.0537)

Notes: Trading volumes are ordered in the third position of y_t. For further notes see Table 3.

total search volume) increases by 1%, the realized volatilities increase by 1.3% to 2.1% depending on the market. This result is found in both bivariate and trivariate SVARs. In addition, when the relative search volume increases by 1%, the trading volume of the corresponding index increases mildly by about 0.2% to 0.4%.

As the next, we look at the long-term impact of the identified contemporaneous structural relationships. This can be observed through IRFs, which trace the effects of the identified structural shocks on the variables of the system over time. Figure 1 shows the IRFs for the trivariate SVAR model as implied by B_{DC} for the Dow Jones.[10] A shock in search queries has a significant impact on realized volatilities up to around 90 days and on trading volumes up to around 35 days. The magnitude of the impact on realized volatilities is almost five times larger than the one on trading volumes. A shock in realized volatilities also has a significant impact on the trading volume lasting for about 10 days. There are no further significant IRFs among

[10]IRFs from other market indices and the GARCH identification method lead to the same conclusion that the contemporaneous relationship among the variable dominates the IRFs.

Table 5: Estimated marginal effects

		Bivariate SVARs		Trivariate SVARs					
		$a_{RV \leftarrow SQ}$	$a_{SQ \leftarrow RV}$	$a_{RV \leftarrow SQ}$	$a_{VO \leftarrow SQ}$	$a_{SQ \leftarrow RV}$	$a_{VO \leftarrow RV}$	$a_{SQ \leftarrow VO}$	$a_{RV \leftarrow VO}$
DOW	DC	**1.862**	−0.0016	**1.8876**	**0.3446**	−0.0075	**0.1216**	0.0269	0.033
		(0.17)	(0.012)	(0.2318)	(0.0451)	(0.0137)	(0.0176)	(0.0213)	(0.1672)
	GARCH	**1.4033**	**0.0267**	1.1452	0.3706	0.0193	0.0646	0.0503	0.5563
		(0.1936)	(0.0117)	(5.478)	(0.2549)	(0.0858)	(0.1296)	(0.4323)	(0.8617)
DAX	DC	**1.2993**	−0.0138	**1.3352**	**0.3847**	−0.0155	**0.1861**	0.0086	−0.0556
		(0.141)	(0.0111)	(0.2394)	(0.1133)	(0.0129)	(0.0563)	(0.0277)	(0.2597)
	GARCH	**1.2742**	−0.0148	0.8992	0.6136	−0.0165	0.0464	−0.0046	0.6005
		(0.2891)	(0.0174)	(12.5775)	(1.9496)	(0.0714)	(1.0361)	(0.3779)	(0.6333)
FTSE	DC	**2.1013**	**−0.0484**	**1.9087**	**0.2247**	**−0.0508**	0.0496	**0.0696**	**0.4484**
		(0.1497)	(0.0136)	(0.16)	(0.088)	(0.0138)	(0.0267)	(0.0321)	(0.1518)
	GARCH	**2.1218**	**−0.0509**	**1.9829**	0.2362	−0.0609	0.0357	0.088	**0.5183**
		(0.2514)	(0.0188)	(0.3565)	(0.6992)	(0.0428)	(0.128)	(0.894)	(0.1652)

Notes: This table documents estimates of marginal effects as defined in (6). More specific, let $\omega_{t-1}^{(j)}$, $j = 1, \ldots, 5$, denote quantities that are available from sample information Ω_{t-1} and estimated reduced form model parameters. Then, bivariate SVARs have the following structure

$$
\begin{aligned}
SQ_t &= a_{SQ \leftarrow RV} RV_t + \omega_{t-1}^{(1)} \\
RV_t &= a_{RV \leftarrow SQ} SQ_t + \omega_{t-1}^{(2)}
\end{aligned}
$$

Structural relations in trivariate SVARs are of the form:

$$
\begin{aligned}
SQ_t &= a_{SQ \leftarrow RV} RV_t + a_{SQ \leftarrow VO} VO_t + \omega_{t-1}^{(3)} \\
RV_t &= a_{RV \leftarrow SQ} SQ_t + a_{RV \leftarrow VO} VO_t + \omega_{t-1}^{(4)} \\
VO_t &= a_{VO \leftarrow SQ} SQ_t + a_{VO \leftarrow RV} RV_t + \omega_{t-1}^{(5)}.
\end{aligned}
$$

Significant parameters at the 5% level are highlighted in bold face. Standard errors are in parentheses. For further notes see Table 3.

other variations of the pairing of the variables. This evidence confirms that the contemporaneous relationships among the variables dominate the subsequent dynamics (IRFs).

It is then not surprising to see that a recursive SVAR model with a different variable sequence than the one suggested by the data-driven approach produces different IRFs, which can be misleading. Figure 2 shows the IRFs from a recursive SVAR using the ordering (RV,SQ,VO). This structure implies that realized volatilities have an impact on the other two variables and search queries have an impact on trading volumes. Indeed, the IRFs (Row 1 Column 2) show that realized volatilities have a lasting significant impact on the search queries up to 40 days, which are entirely different from the corresponding ones in Figure 1 showing no significant impacts. This result is purely driven by the assumption that realized volatility affects search queries in the recursive structure. Considering the IRFs of realized volatilities to shocks in search queries, the initial zero response (due to the assumed recursive structure) gives way to significant responses from 10 days onwards which is due to lagged impacts of search queries on realized volatilities (significant AR coefficients). Also the exclusion of contemporaneous impacts of search queries on realized volatilities seems to weaken the magnitude of the IRFs compared with results from the unrestricted structural model evaluation displayed in Figure 1.

4.3. Further Diagnostics

This subsection provides further diagnostics which highlight the informational content of the adopted data-driven SVAR models. We first discuss results from independence tests applied to model implied shocks, and subsequently comment on the informational content of estimated GARCH models as given in (9).

Considering the results of independence tests in Table 6, it turns out that the estimation of structural shocks in a data-based manner results in independent shocks for most cases. Unlike the lower triangular models, structural models implied by B_{DC} obtain for four out of six specifications p-values of the distance covariance in excess of 10%. Although one should be careful in interpreting these supremum p-values in the usual way as evidence in favour of the null hypothesis, it seems that identified shocks are not only orthogonal but also independent and unique in this sense. Subjecting the structural shocks identified by means of patterns of

Table 6: Independence diagnostics for identified shocks.

	Bivariate SVARs			Trivariate SVARs		
	DOW	DAX	FTSE	DOW	DAX	FTSE
	Distance covariance					
stat	0.240	-0.029	-0.064	0.122	0.080	0.046
p-value	0.300	62.8	87.7	8.00	14.7	26.9
	GARCH					
stat	0.501	-0.020	-0.062	0.465	0.362	0.068
p-value	0.100	57.7	84.4	0.100	0.700	19.0

Notes: Statistics and p-values are multiplied with 100.

conditional heteroskedasticity to independence testing is largely in line with the outcomes for the independence-based identification. For three systems (bivariate: DAX and FTSE; trivariate: FTSE) we find that the hypothesis of independent structural shocks cannot be rejected with 10% significance.

Sentana and Fiorentini (2001) have shown that assuming conditionally heteroskedastic structural shocks allows the full identification of the structural model if at least $(K-1)$ processes $\varepsilon_{k,t}$ are well described by (G)ARCH processes. For the emergence of stylized patterns of volatility clustering, it is essential that the news response parameter (i.e., γ in (9)) is positive (and significant). To diagnose if the SVAR is fully identified under conditionally heteroskedastic shocks, Table 7 documents GARCH parameter estimates and respective standard errors. Since all documented estimates $\hat{\gamma}$ are significant at conventional levels, it follows that the respective SVARs are fully identified for both dimensions ($K = 2, 3$) and all markets.

5. Market sentiment

The previous analysis has highlighted the contemporaneous impact of retail investors' attention on stock market volatility. This section explores what might trigger changes in retail investors' attention. De Long et al. (1990) demonstrate that changes in investors' sentiment can lead to more noise trading and excess volatility, if uninformed noise traders base their trading decisions on sentiment. Da et al. (2015) confirm the positive contemporaneous relationship between sentiment and the market volatility empirically. This section shows that the retail investor's attention can be part of this transmission channel. Changes in the market sentiment

Table 7: GARCH parameter estimates.

	DOW		DAX		FTSE	
	$\hat{\gamma}$	$\hat{g}$	$\hat{\gamma}$	$\hat{g}$	$\hat{\gamma}$	$\hat{g}$
	Bivariate SVARs					
ε_1	0.1331	0.8007	0.2227	0.7099	0.1358	0.8027
	(4.6E-04)	(1.1E-03)	(3.6E-04)	(6.5E-04)	(4.3E-04)	(1.0E-03)
ε_2	0.0167	0.9764	0.0366	0.9456	0.0086	0.9750
	(9.2E-05)	(2.7E-04)	(2.3E-04)	(7.5E-04)	(5.0E-05)	(5.9E-04)
	Trivariate SVARs					
ε_1	0.1591	0.7690	0.2139	0.7167	0.1284	0.8081
	(5.5E-04)	(1.2E-03)	(3.2E-04)	(5.8E-04)	(3.7E-04)	(9.5E-04)
ε_2	0.1378	0.0010	0.1532	0.6920	0.3616	0.2133
	(8.4E-04)	(2.4E-02)	(6.3E-04)	(2.3E-03)	(1.9E-03)	(8.1E-03)
ε_3	0.0124	0.9842	0.0415	0.9404	0.0120	0.9656
	(3.0E-05)	(5.9E-05)	(3.2E-04)	(9.7E-04)	(7.3E-05)	(7.4E-04)

Notes: GARCH estimates do not comprise an unrestricted intercept as the unconditional variance of the structural shocks is normalized to unity by assumption.

have a significant impact on variations in retail investors' attention.

We adopt the FEARS sentiment index of Da et al. (2015) in SVAR models. This index is more transparent compared with market-based measures and available at a higher frequency compared with survey-based sentiment measures. It reveals market-level sentiment by aggregating the internet search volume of queries related to households' sentiment about the economic conditions. Search queries with strongest historical correlations with the market — such as gold prices, recession, GDP, bankruptcy, unemployment — are used to construct the index. Data for this index is only available for the US, however, and can be obtained from Joseph Engelberg's website.[11]

We use the FEARS index based on the top twenty-five, thirty and thirty-five search terms, which are denoted as FEARS25, FEARS30 and FEARS35, respectively. It is calculated as the sum of daily log changes of top search terms, each of which is adjusted (winsorized, deseasonalized and standardized, see Da et al. (2015) for details). We winsorized and desasonalized the daily changes of search queries on DOW in the same way,[12] to obtain an adjusted SQ growth, denoted SQG. Table 8 provides the descriptive statistics of the series. The various FEARS indices have

[11] https://rady.ucsd.edu/faculty/directory/engelberg/pub/portfolios/research.htm

[12] We do not standardize this series, noticing that standardization is used in Da et al. (2015) to make their series of search terms comparable. As we have only one search term regarding a stock index (SQ), this step is not necessary.

Table 8: Descriptive statistics for adjusted SQG and FEARS indices

Var.	Min.	Max.	Mean	S.D.	Var.	Min.	Max.	Mean	S.D.
SQG	-0.333	0.435	0.000	0.134	Fears30	-1.530	2.978	0.001	0.335
Fears25	-1.655	2.977	0.002	0.346	Fears35	-1.597	2.924	0.001	0.330

Table 9: Estimates of marginal effects among FEARS and SQG

	DC		GARCH	
	$a_{FEARS\leftarrow SQG}$	$a_{SQG\leftarrow FEARS}$	$a_{FEARS\leftarrow SQG}$	$a_{SQG\leftarrow FEARS}$
FEARS30	0.0246	**0.1167**	0.1770	**0.0815**
	(0.1867)	(0.0277)	(0.1226)	(0.0196)
FEARS25	-0.0472	**0.1227**	0.1351	**0.0851**
	(0.1990)	(0.0274)	(0.1297)	(0.0213)
FEARS35	0.0477	**0.1124**	0.1606	**0.0831**
	(0.1711)	(0.0262)	(0.1234)	(0.0192)

Notes: This table documents estimates of marginal effects as defined in (6). More specific, let $\omega_{t-1}^{(j)}$, $j = 1, \ldots, 2$, denote quantities that are available from sample information Ω_{t-1} and estimated reduced form model parameters. Then, bivariate SVARs have the following structure

$$\begin{aligned} SQG_t &= a_{SQG\leftarrow FEARS} FEARS_t + \omega_{t-1}^{(1)} \\ FEARS_t &= a_{FEARS\leftarrow SQG} SQG_t + \omega_{t-1}^{(2)} \end{aligned}$$

Significant parameters at the 5% level are highlighted in bold face. Standard errors are in parentheses.

similar distributions with a minimum around -1.6 and maximum around 3. The SQG series has a minimum around -0.33 and a maximum about 0.43. Bivariate SVARs with SQG and FEARS are then estimated, and the estimated marginal effects (see equation (6)) are shown in Table 9.

We have robust evidence for the impact of FEARS on SQG. The estimated marginal effects are significant and vary between 0.08 to 0.12 depending on the FEARS index (FEARS25, FEARS30, or FEARS35) and the identification method (DC or GARCH). As both variables are growth rates of internet search terms, this indicates that a 1% increase in aggregated search terms revealing sentiment leads to around 0.1% increase in search on DOW in the US. In contrast, impacts of SQG on FEARS are subject to high estimation uncertainty and lack significance. The evidence from the bivariate SVARs confirms the contemporaneous impact of market sentiment on retail investors' attention.

6. Conclusion

This paper fills the gap of literature on the relationship between investor attention and stock market activities by identifying the underlying structural transmission among Google search queries, realized volatilities and trading volumes in the US, German and UK markets. We adopt data-based approaches to structural VAR identification. Unlike the a-priori imposition of triangular (i.e. hierarchical) model structures, the data-based identification allows to estimate the structural model parameters in an unrestricted manner. We consider the two identification strategies to provide complementary information. One is identification through the independence of non-Gaussian structural shocks, and the other is identification via conditionally heteroskedastic structural shocks. Our results show the important role of the investor attention in stock markets. While shocks in investor attention affect volatilities and trading volumes immediately, shocks in volatilities and trading volumes do not exert an instant impact on investor attention. Our results are largely robust across the three markets, with alternative identification schemes and using bivariate or trivariate SVARs. While our analysis does not fully support the assumption of a hierarchical model, our results provide important guidance on the hierarchical structure of the variables if a recursive SVAR were used. Finally, our bivariate SVARs with FEARS indices in the US and growth rates of search queries on DOW support the view that market sentiment has an impact on retail investor's attention.

References

Andersen, T., Bollerslev, T., Diebold, F., Labys, P., 2001. The distribution of exchange rate volatility. Journal of the American Statistical Association 96, 42–55.

Andrei, D., Hasler, M., 2015. Investor attention and stock market volatility. Review of Financial Studies 38, 33–72.

Barndorff-Nielsen, O. E., Shephard, N., 2002. Econometric analysis of realised volatility and its use in estimating stochastic volatility models. Journal of the Royal Statistical Society, Series B 64, 253–280.

Bollerslev, T., 1987. A conditional heteroscedastic time series model for speculative prices and rates of return. Review of Economics and Statistics 69, 542–547.

Bouakez, H., Normadin, M., 2010. Fluctuations in the foreign exchange market: How important are monetary policy shocks? Journal of International Economics 81, 139–153.

Brooks, C., 1998. Predicting stock index volatility: Can market volume help? Journal of Forecasting 17, 59–80.

Brüggemann, R., Jentsch, C., Trenkler, C., 2016. Inference in VARs with conditional heteroskedasticity of unknown form. Journal of Econometrics 191, 69–85.

Comon, P., 1994. Independent component analysis, a new concept? Signal Processing 36, 287–314.

Da, Z., Engelberg, J., Gao, P., 2011. In search of attention. Journal of Finance 66, 1461–1499.

Da, Z., Engelberg, J., Gao, P., 2015. The sum of all fears investor sentiment and asset prices. Review of Financial Studies 28, 1–32.

De Long, J., Shleifer, A., Summers, L., Waldmann, R., 1990. Noise trader risk in financial markets. Journal of Political Economy 98, 703–738.

Dimpfl, T., Jank, S., 2011. Can internet search queries help to predict stock market volatility? CFR Working Paper 11-15, CFR.

Dimpfl, T., Jank, S., 2016. Can internet search queries help to predict stock market volatility? European Financial Management 22, 171–192.

Gouriéroux, C., Monfort, A., Renne, J., 2017. Statistical inference for independent component analysis: Application to structural VAR models. Journal of Econometrics 196, 111–126.

Kilian, L., Lütkepohl, H., 2017. Structural vector autoregressive analysis. Themes in Modern Econometrics. Cambridge University Press.

Lange, A., Dalheimer, B., Herwartz, H., Maxand, S., 2019. Svars: Data-driven identification of SVAR models. R package version 1.3.6.

Lange, A., Maxand, S., Dahlheimer, B., Herwartz, H., 2020. Svars: Data-driven identification of svar models. Journal of Statistical Software, forthcoming.

Lanne, M., Meitz, M., Saikkonen, P., 2017. Identification and estimation of non-Gaussian structural vector autoregressions. Journal of Econometrics 196, 288–304.

Lanne, M., Saikkonen, P., 2007. A multivariate generalized orthogonal factor GARCH model. Journal of Business & Economic Statistics 25, 61–75.

Lütkepohl, H., Netšunajev, A., 2017. Structural vector autoregressions with smooth transition in variances. Journal of Economic Dynamics and Control 84, 43–57.

Lux, T., Marchesi, M., 1999. Scaling and criticality in a stochastic multi-agent model of a financial market. Nature 397, 498–500.

Matteson, D. S., Tsay, R. S., 2017. Independent component analysis via distance covariance. Journal of the American Statistical Association 112, 623–637.

Milunovich, G., Yang, M., 2013. On identifying structural VAR models via ARCH effects. Journal of Time Series Econometrics 5, 117–131.

Moneta, A., Entner, D., Hoyer, P. O., Coad, A., 2013. Causal inference by independent component analysis: Theory and applications. Oxford Bulletin of Economics and Statistics 75, 705–730.

Normadin, M., Phaneuf, L., 2004. Monetary policy shocks: Testing identification conditions under time-varying conditional volatility. Journal of Monetary Economics 51, 1217–1243.

Rigobon, R., 2003. Identification through heteroskedasticity. The Review of Economics and Statistics 85, 777–792.

Risk, B. B., James, N. A., Matteson, D. S., 2015. steadyICA: ICA and tests of independence via multivariate distance covariance. R package version 1.0.

Sentana, E., Fiorentini, G., 2001. Identification, estimation and testing of conditionally heteroskedastic factor models. Journal of Econometrics 102, 143–164.

Sims, C. A., 1980. Macroeconomics and reality. Econometrica 48, 1–48.

Székely, G. J., Rizzo, M. L., Bakirov, N. K., 2007. Measuring and testing dependence by correlation of distances. Annals of Statistics 35, 2769–2794.

Vlastakis, N., Markellos, R. N., 2012. Information demand and stock market volatility. Journal of Banking and Finance 36, 1808–1821.

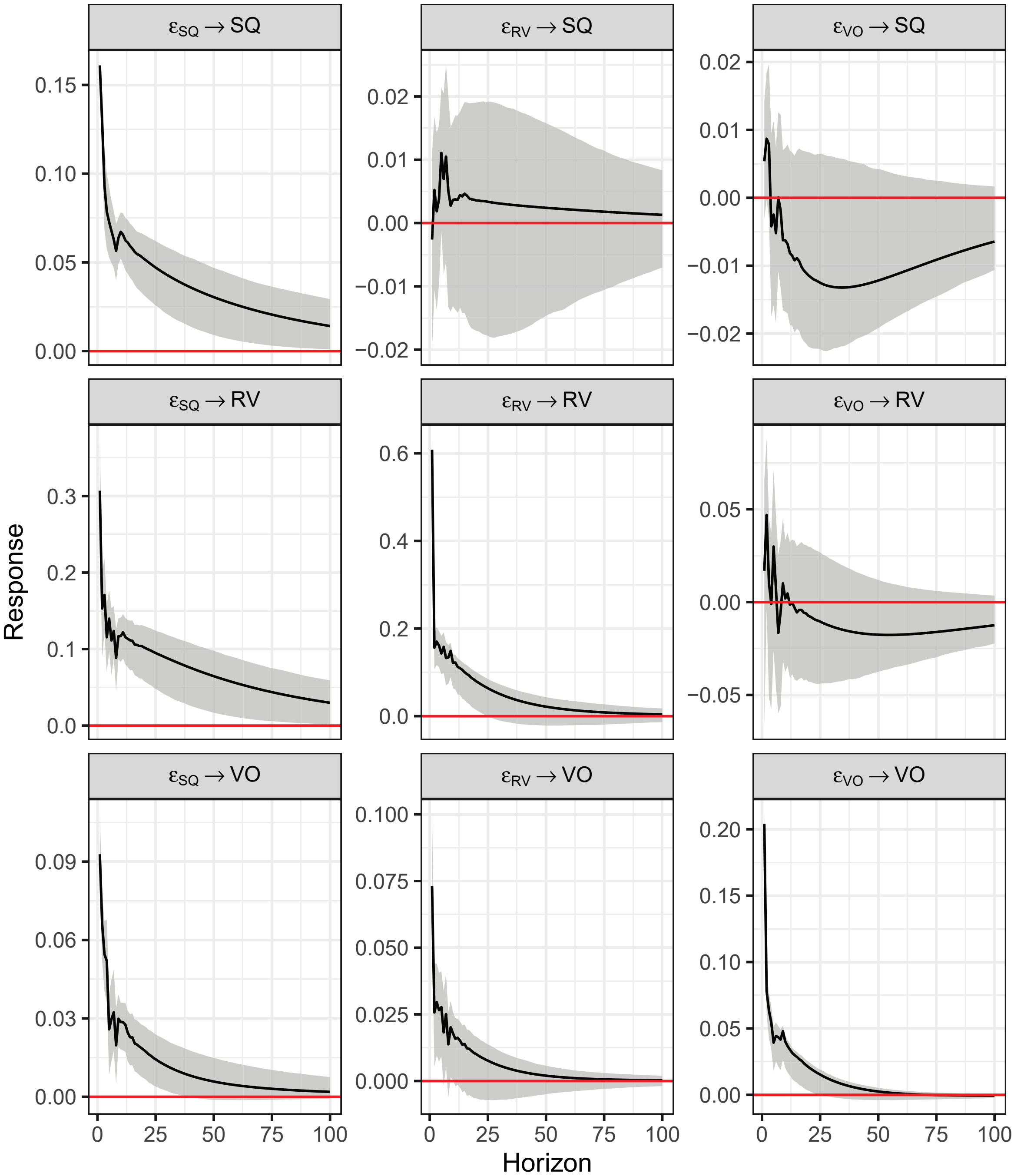

Figure 1: IRFs are displayed joint with 95% bootstrap confidence intervals. The R-package 'svars' supports the analyst with tools of bootstrap inference which are commonly used in (structural) VAR analysis. We opt for a recursive design moving block bootstrap approach. Brüggemann et al. (2016) have shown the asymptotic validity of moving block bootstrap designs for inferential analysis in SVARs. The chosen block length is 30 which is between 2.17% (Dow Jones) and 2.38% (FTSE) of the overall available sample information. The number of bootstrap replications is 1999.

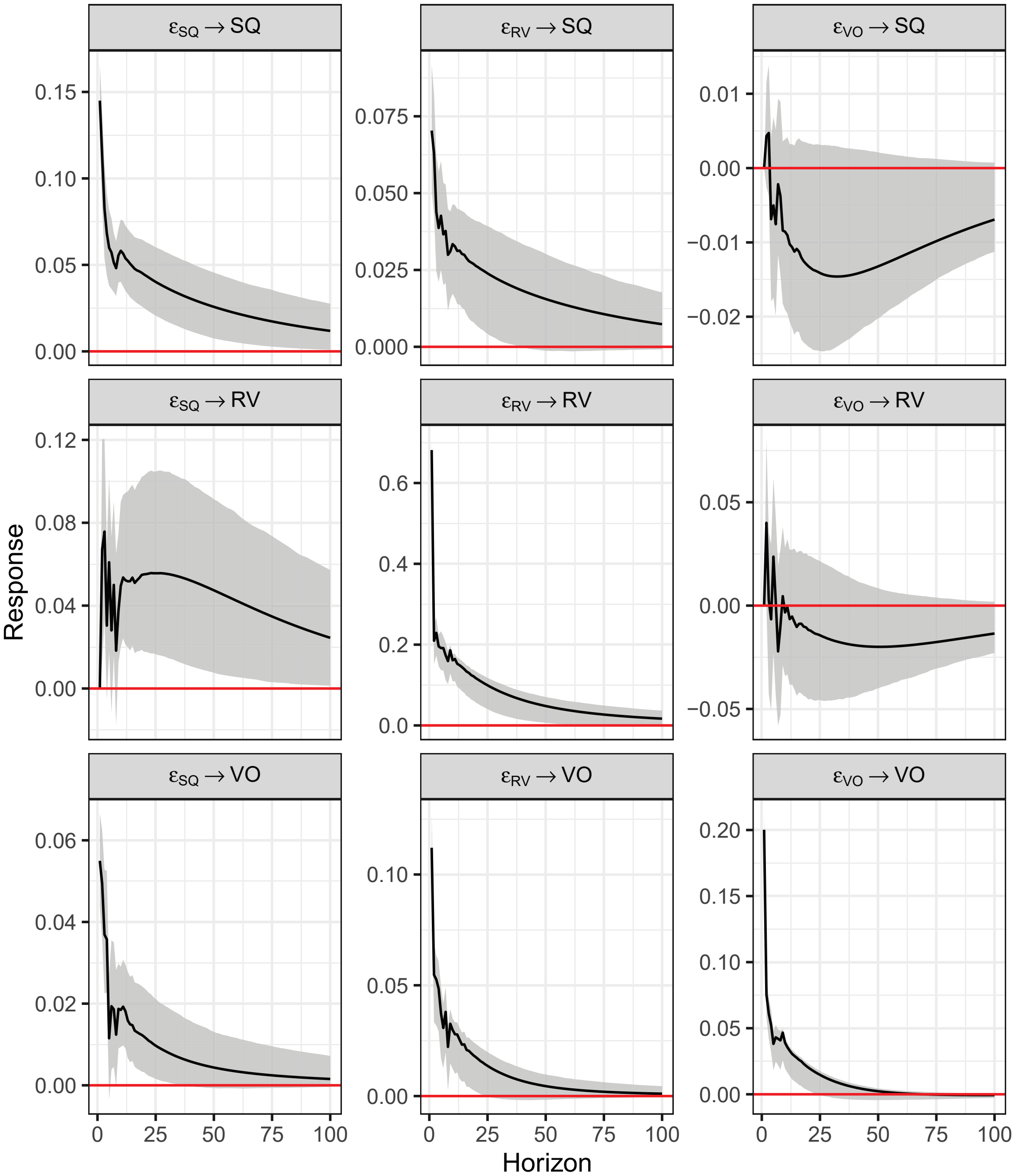

Figure 2: For the purpose of a facilitated comparison, we use the same variable ordering as indicated in Figure (i.e. search queries and then realized volatilities in the bivariate system; search queries, realized volatilities and then trading volumes in the trivariate system). For further notes see Figure 1.