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 a[illegible]

1. Introduction

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.,

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

The representations in (1) and (2) di□er 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

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 e￿ect patterns that involve the variables in y_t contemporaneously. To provide these e￿ects 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 e￿ects. Consider, for instance, the bivariate case $K = 2$. From (5) and

$$B^{-1}y$$

tion for two classes of data-based identification. In this work, we identify structural models by (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

$$\hat{B} = \text{argmin}$$

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 $E[\varepsilon^2_t] = \text{Var}[\varepsilon_{kt}] = 1 \;\; k$, 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_{GC} Q@97222220009722222036B970222220360al 04()18(8()10010)07(al)4588()03(a)09(q97

3. Data and preliminary analysis

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

Table 1: Descriptive statistics

can obtain for each market and that were directly downloaded from Google trend.

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 stll.4(98.6(s)-0.6(e)-0.4(l)-0.8

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. W28.6(e(at)-0.9gh)-(ou4(IW)-0.8(0TJETQe)-0

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

Table 5: Estimated marginal e[illegible]ects

Bivariate SVARs		Trivariate SVARs	
$a_{RV \leftarrow SQ}$	$a_{SQ \leftarrow RV}$	$a_{RV \leftarrow SQ}$	a

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

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)

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 aect 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.

Barndor-Nielsen, O. E., Shephard, N., 2002. Econometric analysis of realised volatility and its

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.

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 e￿ects. 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.