1 Introduction

This paper investigates the relationship between stock-market return, volatility and trading volume. We estimate an AR(10) to use the residuals for further analysis. The residuals are used and applied to a TGARCH(1,1) model to examine the effects of surprise trading volume on conditional variance. This paper finds that the model containing only positive surprise trading volume shocks is the better than one containing both negative and positive. Furthermore, we estimate a news impact curve to show the impact of "good" and "bad" news, and end up finding that bad news has an effect on the volatility. Our results end up aligning with the results of Wagner and Marsh (2005).

2 Description of data

The data has been downloaded from https://finance.yahoo.com/ for the period November 1, 1999 to November 22, 2021. The data set contains daily changes for the stock market index consisting of blue chip German companies trading on the Frankfurt Stock Exchange (DAX). We define the most relevant variables as:

- $S_t = \log Vol_t - \log VolMA50_t$
- $D\log DAX = 100 \cdot (\log DAX_t - \log DAX_{t-1}) = \sigma_t^2$

Where Vol_t is the trading volume and $VolMA50$ is the 50 days moving average of $\log Vol_t$.
Figure 1 shows a graph of S_t and $D\log DAX$. As seen in table 1, we have kurtosis > 3, which means the distribution has fatter tails than a Gaussian distribution, which indicate ARCH-effects.

	Mean	Variance	Skewness	Kurtosis
S_t	0.0025	0.0915	0.2563	4.5535

Table 1: Descriptive statistics for S_t

3 Econometric Theory

3.1 GARCH-models

To model the conditional variance of our time series data for the German stock market, we consider modeling an autoregressive conditional heteroskedasticity (ARCH) model.

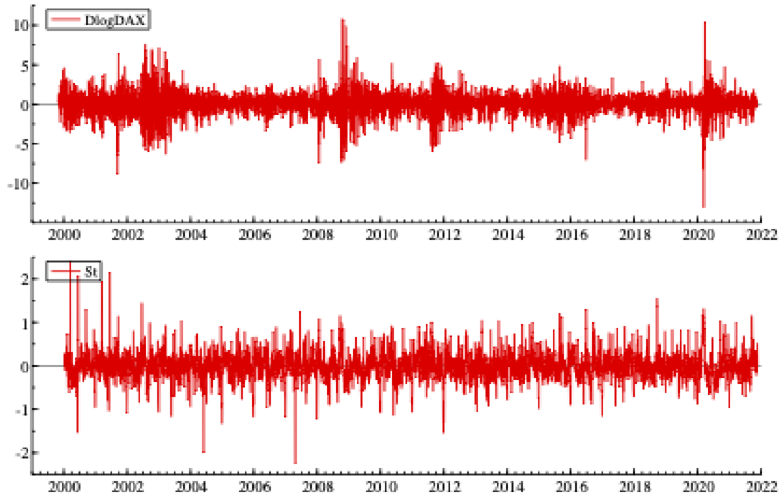

Figure 1: Logarithmic transformations of the time series

The ARCH model often requires many lags to be significant, so we instead choose to model the general ARCH (GARCH) model. A general form of the GARCH(p,q) model can be written as

$$\sigma_t^2 = \bar{w} + \sum_{j=1}^{p} \alpha_j \epsilon_{t-j}^2 + \sum_{j=i}^{q} \beta_j \sigma_{t-j}^2$$

The stationarity condition of the standard GARCH-model is given by:

$$\sum_{j=1}^{p} a_j + \sum_{j=1}^{q} \beta_j < 1$$

With a Threshold-GARCH(1,1) (TGARCH) being:

$$\sigma_t^2 = \varpi + \alpha\epsilon_{t-1}^2 + \kappa\epsilon_{t_1}^2 \mathbb{1}(\epsilon_{t-1} < 0) + \beta\sigma_{t-1}^2$$

The stationarity condition is:

$$\alpha + \beta + \frac{1}{2}\kappa < 1$$

The Threshold effect allows the conditional variance to differ when there are positive or negative shocks. The TGARCH model only contains the unknown parameters. When we estimate the model we need $\alpha, \beta, \kappa \geq 0$ in order for σ_t^2 to be non-negative. ϖ also needs to be non-negative for this to hold.
Our estimator of choice will be the QMLE, as the Student's t(v)-distribution will converge to the Gaussian as $v \to \infty$ and the non-normal distribution of errors prevents the use of MLE.

3.2 Testing for ARCH-effects and misspecification tests

To test for ARCH-effects we apply a Breuch-Pagan LM test for no-heteroskedasticity, where the auxiliary regression is:

$$\hat{\epsilon}_t = \beta_0 + \beta_1 \epsilon_{t-1}^2 + ... + \beta_p \hat{\epsilon}_{t-p}^2 + \text{error}$$

The residuals $\{\hat{\epsilon}_t\}_{t=1}^T$ are estimated from the linear regression. The null hypotheses is $H_0 : \beta_p = 0$ with the statistic:

$$\zeta_{\text{ARCH}} = T * R^2 \xrightarrow{d} \chi^2(p) \qquad \text{under } \text{H}_0$$

We also test for autocorrelation with a Portmanteau-test with a χ^2-distribution and for normality with a Goodness-of-fit test based on a Jarque-Bera test, which uses the same distribution.

3.3 Surprise Trading Volume

The surprise trading volume is both an expression of the private informations sets and the differences between different agent's information sets. Traders on a market share a public information set, but also possess a private derived from market signals. To find the surprise trading volume at a given time, we need to find the deviation from the trend, which is given by:

$$S_t = \log \text{Vol}_t - \log \text{VolMA50}_t$$

"$\log \text{Vol}_t$" is the logarithm of the number of stocks traded, while "$\log \text{VolMA50}_t$" is the 50 day moving average of "$\log \text{Vol}_t$". The surprise trading volume is given by the error term, f_t, in the following simple univariate time series given by an AR(k)-model:

$$S_t = \delta + \sum_{i=1}^{k} \phi_i S_{t-i} + \gamma_1 \text{FirstTrDay}_t + \gamma_2 \text{LastTrDay}_t + \kappa_1 \text{isJan}_t + ... + \kappa_{11} \text{isNov}_t + f_t \quad (1)$$

This way, we filter out the predictable part and we are left with the unpredictable part, $\hat{f}_t$. To capture the effect of negative and positive surprise trading volumes, we create the following two variables:

$$f_t^{pos} = \hat{f}_t \cdot \mathbb{1}(\hat{f}_t > 0)$$

$$f_t^{neg} = \hat{f}_t \cdot \mathbb{1}(\hat{f}_t < 0)$$

$\mathbb{1}(\hat{f}_t > 0)$ is 1 if $\hat{f}_t > 0$ and 0 elsewhere. The opposite is true for $\mathbb{1}(\hat{f}_t < 0)$.

4 Empirical analysis

4.1 Model selection

We estimate an AR(10) model including lags 1, 2, 3, 4, 5, 7 and 10 as given in (1). This model suffers from ARCH-effects, normality errors and heteroskedasticity, but it does not suffer from autocorrelation. The ARCH-effects are presented below, which are quite significant in the $\chi^2(1)$ distribution:

$$\zeta_{\text{ARCH}} = T * R^2 = 5537 * 0.3627 = 2008.26$$

This is why we do further analysis using the GARCH model. We do not interpret the estimates of equation (1) but only use the residuals $\hat{f}_t$ for further analysis. Therefore, normality errors and heteroskedastity are not of high importance.
To be able to compare our model with Wagner and Marsh, we write up the following TGARCH(1,1) model with positive lagged surprise trading volume included:

$$\sigma_{1t}^2 = \varpi + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + \omega_1 f_t^{pos} + \omega_2 f_{t-1}^{pos} + \kappa_1 \epsilon_{t-1}^2 \cdot \mathbb{1}[\epsilon_{t-1} < 0] \tag{2}$$

This equation is comparable to model M7 in the article, which Wagner and Marsh considers to be a quite robust model. We also consider another specification in which we have a negative surprise trading volume as well as a positive one, which we write up as:

$$\sigma_{2t}^2 = \varpi + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + \omega_1 f_t^{pos} + \omega_3 f_t^{neg} + \kappa_1 \epsilon_{t-1}^2 \cdot \mathbb{1}[\epsilon_{t-1} < 0] \tag{3}$$

In both of our models we also include threshold effects, since they are included in Wagner and Marsh. This also allows negative and positive shocks to affect our model in different ways. We impose the restriction that $\alpha_1, \beta_1 \geq 0$.

4.2 Estimation Results

As seen in table 2, the models indicate both normality problems and ARCH-effects. Therefore we use the QMLE as estimator. The QMLE is still consistent when there are ARCH effects, however it's not efficient. The estimates of the two models are given in table 2.

The coefficients of the shared variables of the two models shows similar results. When there is a positive surprise trading volume in period t, it affects the conditional mean negatively, while it affects the conditional variance positively. As mentioned earlier, the surprise trading volume is serving as a proxy for private information flow. When the private information flow is large in a day, there will also be a large amount of trading in the market in the same day. Contrary to the conditional variance, the

	$((2), f^{pos})$	$((3), f^{pos} \& f^{neg})$
σ^2_{t-1} (Y)	-0.06523 (0.0121)	-0.02204 (0.0127)
Constant (X)	0.1091 (0.0137)	0.1014 (0.0171)
ω_3 (X)	.	-0.02324 (0.0709)
ω_1 (X)	-1.224 (0.135)	-0.9986 (0.134)
ω_2 (X)	-0.05967 (0.0786)	.
ϖ (H)	$6.278 * 10^{-11}$ (-0.000)	$2.346 * 10^{-08}$ $(5.15 * 10^{-08})$
α_1 (H)	$7.989 * 10^{-09}$ (-0.000)	0.0000 (-0.000)
β_1 (H)	0.9311 (0.00624)	0.8960 (0.0162)
ω_3 (H)	.	0.1468 (0.0259)
ω_1 (H)	5.315 (0.357)	0.5651 (0.102)
ω_2 (H)	-4.906 (0.340)	.
κ (H)	0.09167 (0.00971)	0.1688 (0.0264)
student-t df	45.94 (24.5)	11.35 (1.62)
alpha(1)+beta(1)	0.931	0.896
Log-lik.	-8409.177	-8669.328
AIC	3.042	3.135
HQ	3.047	3.140
SC/BIC	3.055	3.149
Portmanteau, 1-74	[0.61]	[0.63]
No ARCH(1)	[0.00]	[0.00]
Normality	[0.00]	[0.00]
T	5536	5537
Sample start	2000-01-28	2000-01-27
Sample end	2021-11-22	2021-11-22

Table 2: The table shows estimates of the model in equation (X) with various restrictions imposed. Standard errors in $(\cdot)$ and p-values in $[\cdot]$ for misspecification tests.

conditional mean is affected negatively by a positive shock to the surprise trading volume. It appears that more private information result in a higher activity in where agents on the market sell their stocks, which in turn makes the mean stock price fall.

If we compare model (2) and (3), model (2) is the better model according to the Akaike Information Criterion (AIC). The AIC obtains the maximized log-likelihood function and penalize it for the numbers of variables in the model. According to model (2), the effect on the conditional variance from the positive surprise trading volume on the same day is large and highly significant. Contrary, a one day lagged positive surprise trading volume has a large negative effect on the conditional variance. A high surprise trading volume will result in high volatility in one day and low volatility the next day.
Even though model (2) appears more robust than model (3) according to the AIC, model (3) can still be useful for estimating the effects of negative surprise trading volume. The effects on the conditional variance are small but significant in the model, while the effect on the conditional mean is insignificant. Therefore, the negative surprise trading volume is nowhere near as important as the positive trading volume.

4.3 News impact curves

The effect from a positive shock in model (2) is given by $\alpha_1 = 7.989 * 10^{-09}$, which is very close to zero. Therefore, a positive market shock is too small to have an actual effect on the conditional variance. We instead check the effect from a negative shock, which is given by $\alpha_1 + \kappa = 7.989 * 10^{-09} + 0.09167 = 0.09167$. Negative shocks on the market has an effect on the conditional variance. This indicates that "bad news" has an effect on the volatility of the stock market while "good news" has no measurable effects. These two effects can be seen by the news impact curve given in figure 2. The news impact curve shows the effect of a shock ϵ_{t-1} on the conditional variance.

5 Discussion and concluding remarks

When specifying our model, we could have opted for many different GARCH models. One of the models we could have used is the EGARCH model. The EGARCH model is an exponential model which does not put any restrictions on the parameters. However, if we estimate our model without any restriction on α_1 and β_1, we find that β_1 is still positive and significant while α_1 still very close to zero.
We could also consider expanding our model to include both lagged values of f_t^{pos} and f_t^{neg}, but it was not possible to find any convergence for those models. At last, we could consider including another lag for our σ^2-variable - such as a fifth lag which

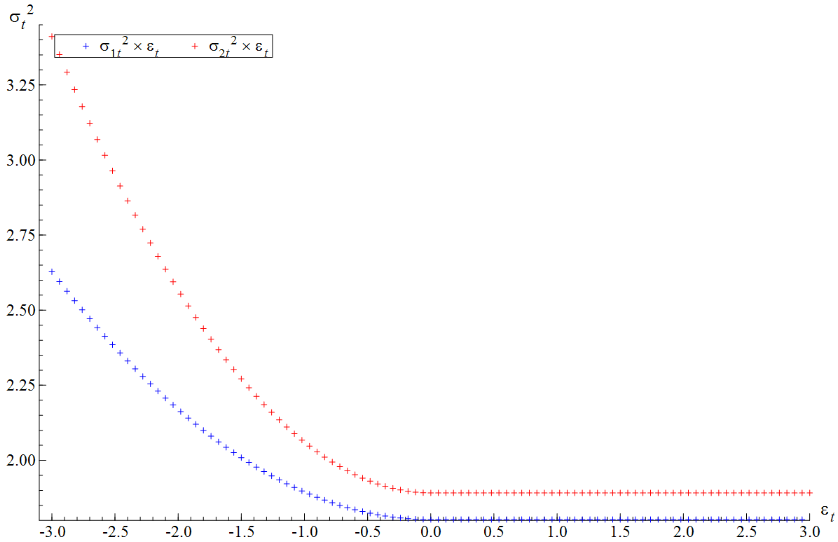

Figure 2: News impact curves

might be significant, because then we would also examine the effect from the same weekday. However, this would make it difficult to compare it to Wagner and Marsh.

Overall our results indicate the same as Wagner and Marsh (2005), where surprise trading volume has a large effect on the volatility of the market and thus conditional variance. However, it should be noted as we opted the for same GARCH-model and examine the same data, and we therefore expect the same effects compared to their study. Nonetheless, the effects from surprise trading volume are significantly larger in this paper compared to theirs.