ERASMUS UNIVERSITEIT ROTTERDAM

Erasmus School of Economics

Master thesis in Econometrics and Management Science

Estimation of a time-varying parameter GARCH model based on Google Trends

Author: Fabian Thomassen

July 31, 2018

Supervised by Dr. Michel van der Wel

Co-reader: Dr. X. (Xiao) Xiao

Abstract

In this paper, I propose to include a 'bad news' index based on Google search volume in a GARCH model. The resulting specification allows its parameters to vary with the level of the index via a logistic transition function. In a simulation study, I then illustrate the model's dynamics and justify the use of maximum likelihood estimation, as well as the validity of a Likelihood Ratio Test to verify the explanatory power of the exogenous information. On a set of S&P 500 returns the test indicates that including Google search volume improves the fit significantly for the periods during and after the global financial crisis in 2008. A comparison to other GARCH specifications, including an extended GARCH and spline-GARCH, shows that the proposed model yields accurate Value-at-Risk predictions but the difference of its variance forecasts to a realized variance proxy rank behind a standard GARCH(1,1) model in both an in-sample and out-of-sample setting.

Keywords: GARCH model, varying parameters, Google Trends, Likelihood Ratio Test, Realized Variance, Value-at-Risk

Contents

1 Introduction

Correctly measuring and forecasting the volatility of asset returns as the underlying of various securities, for portfolio management or as the basis for risk metrics has become increasingly important in financial markets. Across markets and asset classes, large swings in returns are more likely to be followed by large swings and vice-versa. Also, periods of large and small variations alternate. This phenomenon is known as 'volatility clustering', which implies a certain degree of predictability. Moreover, 'bad news' tend to have a bigger impact on the volatility level than 'good news'. This response is commonly known as the 'leverage effect' (Black, 1976). However, political or economic shocks may change the dynamics of the volatility series over longer periods of time and a model should account for it.

In the financial industry the General Conditional Heteroskedasticity (GARCH) model of Bollerslev (1986) is still established as the 'workhorse'. In its simplest form the standard GARCH is able to explain the high persistence of the volatility or variance and provides a good first impression of the dynamics of the series under investigation. Nonetheless, the model requires the variance to be stationary over time. Although this assumption might hold for shorter time spans, the long-run properties of volatility clustering may change over time. Therefore, Teräsvirta and Amado (2008) propose a GARCH specification, whose parameters change smoothly through time via a logistic transition function. In this way, the time-varying GARCH (TV-GARCH) relaxes the assumption of stationarity and is able to capture trends. However, also this specification expects that tomorrow's variance can be predicted from today's variance and return without any uncertainty. But, there must also be an exogenous component that goes beyond the variance's own history, because news are still flowing in after markets close.

Amongst others, Vlastakis and Markellos (2012) therefore suggest to incorporate internet search volume as a proxy for investors' demand for news. When the market is closed, investors do not cease to search for information on financially relevant topics. As the internet has become the primary source of information, the relative popularity of a specific search keyword on the Google search engine is direct measure of investors' attention (Da, Engelberg, & Gao, 2011). Thus, a sudden rise in the search volume of keywords with bad financial connotation, such as ‘inflation’, can be an indicator for ‘bad news’ arrivals.

In this paper, I amend the TV-GARCH of Teräsvirta and Amado (2008) by a ‘bad news’ index created from Google search volumes of keywords with a bad financial connotation. Thus, the proposed model incorporates news also arriving, when the market is closed, and can account for non-stationarity. The component that depends on the external variable is added to a standard GARCH model and multiplied by a logistic transition function. This property lets the model parameters vary with the level of the Google search volume and implicitly with time. Consequently, the base variance level, the influence of the past returns and the autoregressive component all fluctuate depending on the search interest of investors.

This approach is different from the popular method, in which the exogenous information is added directly to the GARCH equation as a covariate, the so-called GARCH-X (Brenner, Harjes, & Kroner, 1996). By offering more degrees of freedom, the proposed method allows the search volume to influence the behavior of the conditional variance more flexibly. While this property might be beneficial, it also comes with the risks of potential misspecification and increased estimation noise. Furthermore, Engle and Rangel (2008) propose a specification, in which the variance is multiplicatively decomposed in a stationary and a non-stationary component. In their spline-GARCH model, the non-stationary part follows an exponential quadratic spline, which is fitted to the data before estimation and is able to incorporate exogenous information. In this way, Engle and Rangel are able to capture slow-moving trends or seasonalities.

In a simulation study, I verify that the theoretical properties of a Likelihood Ratio Test (LRT) hold, such that it can be applied to test the explanatory power of the Google search volume on the variance. If the test statistic does not exceed the critical value, the difference in likelihood compared to a standard GARCH is insignificant, so the likelihood added by the exogenous variable is to neglect. Moreover, the simulations show that the model parameters are estimated accurately by maximum likelihood, if the smoothness level of the transition function is chosen in advance. In the case of a misspecification, the estimates that depend on the external information become more noisy, but are still consistent.

For the empirical part of this paper, I focus on U.S data. In contrast to the European market, where most of the income is deposited in savings accounts and investments are managed by institutions, American retail investors have a more pronounced equity culture (Skypala, 2017). Because Google has become one of the primary sources of information for retail investors, I focus on U.S. search volumes, which is publicly available via Google's service 'Google Trends'. Moreover, I download a set of S&P 500 returns. This stock index is often seen as a gauge of the U.S. equity market. For the evaluation of the variance estimates, I employ a realized variance measure based on high-frequency data.

After testing the proposed specification on the data, the LRT suggests that including the bad news index into the varying parameter specification improves the likelihood compared to a standard GARCH model significantly. Hence, the Google search volume for keywords with a negative financial connotation helps to explain the magnitude of the price swings in the U.S. stock market. It also improves the out-of-sample predictions of the quantiles of the conditional distribution in form of Value-at-Risk forecasts. Nevertheless, due to outliers in the realized variance proxy during the financial crisis the accuracy of the variance estimates falls behind the alternative GARCH specifications.

The rest of this paper is organized as follows. Section 2 discusses the relevant literature about the development of different GARCH specifications as well as the use of Google Trends in a financial context. In section 3, I describe the S&P 500 dataset alongside the variance proxy and explain, which Google search keywords constitute the bad news index. Next, I specify the model that allows its parameters to vary with the level of the Google Trends index, as well as the other GARCH specifications that include exogenous information for a later comparison. Then, I conduct a simulation study in section 5 to confirm the consistency and accuracy of the estimated parameters and the properties of the Likelihood Ratio Test to determine, if there is a significant relationship between the S&P 500 returns and the Google Trends data. Lastly, I interpret the empirical results of Section 6 in a more general setting and suggest further improvements.

2 Literature Review

Since the introduction of the Autoregressive Conditional Heteroskedasticity (ARCH) model by Engle (1982) and its generalized extension by Bollerslev (1986), several more extensions have been developed, which can be summarized in two main categories. The first category of models adjusts the standard specification such that the model provides a better fit to the characteristics and statistical properties of the series under investigation. One important example is the threshold GARCH (T-GARCH) model by Zakoian (1994), which takes the leverage effect into account by treating the influence of positive and negative returns differently.

Another line of research in this context has focused on relaxing the assumption of the constant unconditional variance of the original specification and introduced time-varying ARCH (Dahlhaus, Rao, et al., 2006) and time-varying GARCH (TV-GARCH) (Teräsvirta & Amado, 2008) models. In both models the parameters are allowed to change smoothly through time by the use of transition functions.

The second kind of GARCH extensions tries to capture the dynamics of the conditional volatility by including exogenous information. Han and Kristensen (2014) verify that, if a covariate is directly added to the conditional variance equation, the asymptotic properties of the parameter estimates still hold, even if this variable is not stationary. The choices for exogenous variables cover a wide range of economic and financial indicators such as trading volumes and interest rate spreads.

Since Google search data has been made publicly available more than a decade ago, researchers have found vast empirical evidence for a relation between internet search volume and stock market returns. Preis, Moat, and Stanley (2013) have successfully implemented a trading strategy based on Google search volume and Da et al. (2011) consider the Google Trends series a good measure of investors' intention. Therefore, Da, Engelberg, and Gao (2014) also create an index out of search keywords with a negative connotation in an economic sense to sum up investors' fears and subsequently discover that an increase in the index coincides with a temporary increase in volatility. Also Vlastakis and Markellos (2012) find evidence for a significant positive relation between internet search volume and various volatility measures, such as backward-looking historical volatility and forward-looking implied volatility of option prices. Smith (2012) then extends their research from a firm level to the market of exchange rates. In contrast to the previous research, the methods in this paper propose a new way to include Google Trends data in a GARCH model to test its explanatory power on a major U.S. financial indicator.

A short experiment by Taleb and Goldstein (2007) shows that because volatility itself is a latent variable, even experts in financial markets make the mistake of interpreting the standard deviation of stock returns as a physical measure. This is one of the reasons why simply taking the squared or absolute daily return leads to a noisy proxy. Since its formulation by Andersen, Bollerslev, Diebold, and Labys (2001), the idea of using high-frequency data to calculate 'realized variance' has been growing in popularity, because realized variance is an unbiased and efficient estimator which is less prone to noise in the price measurement than squared returns.

Nonetheless, realized variance, or volatility is still an imperfect proxy. When comparing different volatility forecasts, this can lead to wrong rankings. Therefore, Patton (2011) defines a set of loss functions that is robust to noise in the proxy and homogeneous in data transformations, which includes the Mean Squared Error (MSE), but not the Mean Absolute Error (MAE). Hence, I will prefer the MSE over the MAE, if the rankings based on these measures contradict each other, even though the MSE is prone to outliers.

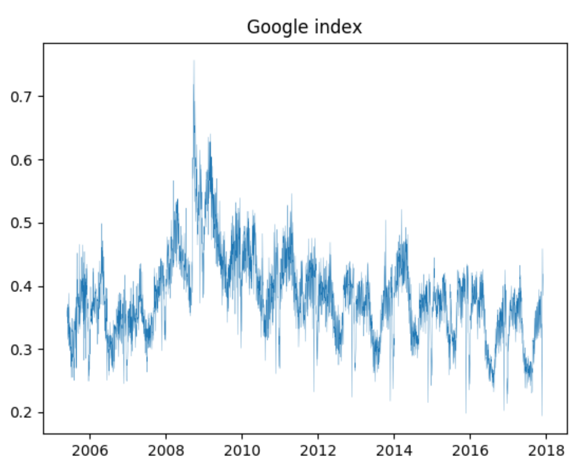
3 Data

3.1 Google Search Volume

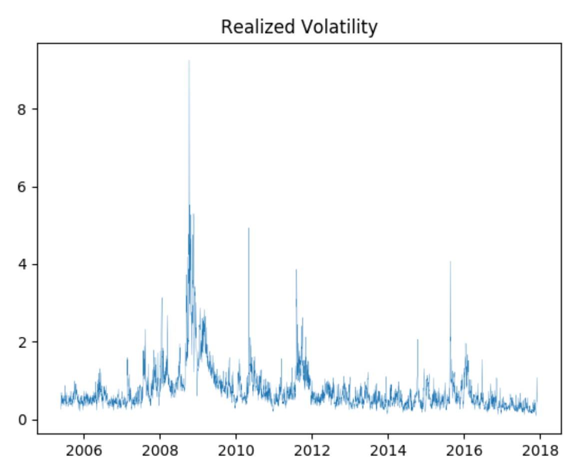
Google provides publicly available data on its search engine activity dating back to 1^{st} January, 2004 via 'Google Trends' (http://www.google.com/trends). The time series obtained from the service are an unbiased and random sample that indicate the searches for a specific keyword as a proportion of all searches on all topics on Google at certain points in time. The scale of observations ranges from zero to 100 and value of 100 marks the time of the biggest relative search interest. The service also lets the user obtain the search frequencies based on the location of the queries. To facilitate the analysis of the U.S. economy, I therefore restrict the data to searches coming from the U.S. only and thereby avoid including undesired search interest coming from worldwide search activity.

For periods smaller than 90 days Google provides the data in daily frequency and periods larger than 90 days and up to five years are provided in weekly frequency. Any date range that is longer than five years is only available in monthly frequency. For the analysis of a sample of daily stock returns that is longer than five years, I thus need to obtain multiple smaller samples of 90 days. I then rescale these daily values based on the monthly values of the entire sample period. Finally, I divide the trend data by 100 to avoid computational problems. As a last step, I discard the first observations from 1^{st} January, 2004 until 26^{th} May, 2005 due to missing values in the starting phase of the Google Trends service, and all non-trading days such that the final sample starts on 27^{th} May, 2005 and constitutes of 3146 observations.

As search keywords I consider financially relevant terms with connotations of bad news for the economy based on a measure created by Preis et al. (2013) and the Financial and Economic Attitudes Revealed by Search (FEARS) index of Da et al. (2014). Preis et al. (2013) propose to measure relevance by the number of occurrences of a word in the Financial Times newspaper normalized by the amount of hits on Google, while Da et al. (2014) identify keywords that are negatively related to asset returns. Thus, I include those search terms that are in the intersection of the 'FEARS' index and the keywords of the analysis of Preis et al. (2013). These terms are 'crisis', 'default', 'economy', 'inflation' and 'unemployment'. Finally, I take the average of over all keywords at each point in time to create my 'bad news' index.

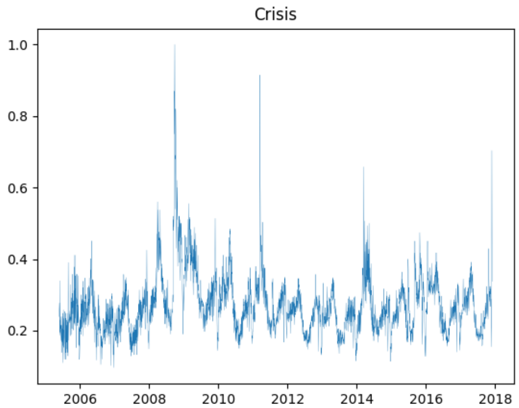


(a) Google Trends index

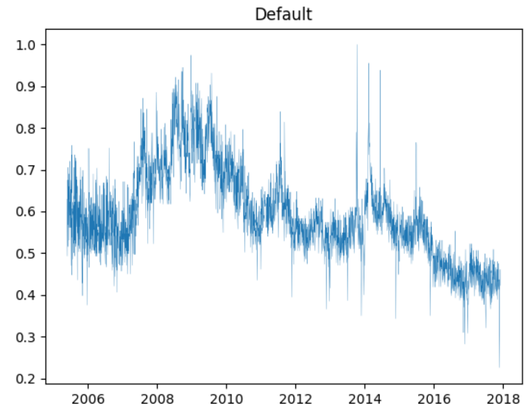


(b) Realized volatility

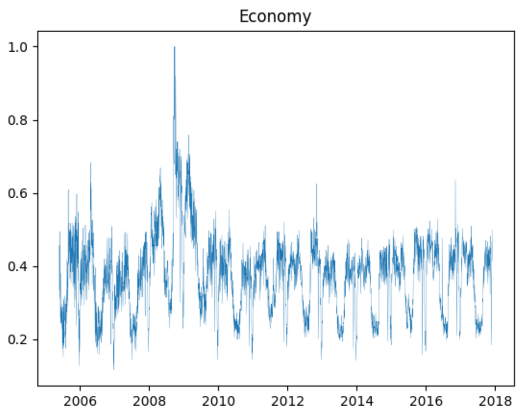
Figure 1: Comparison of the 'bad news' index and the realized volatility proxy



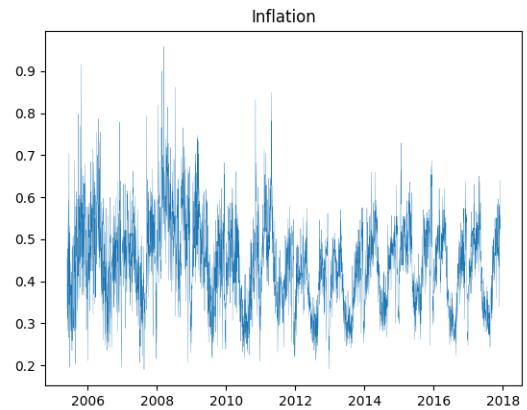
(a) Crisis



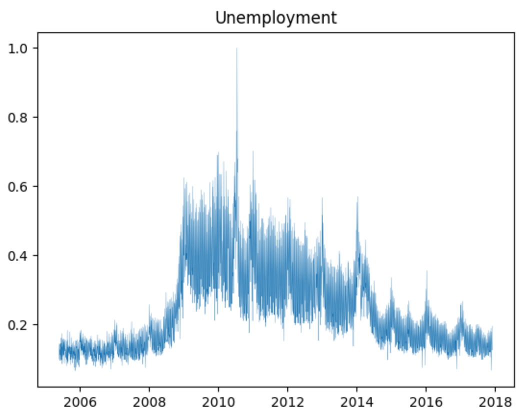
(b) Default



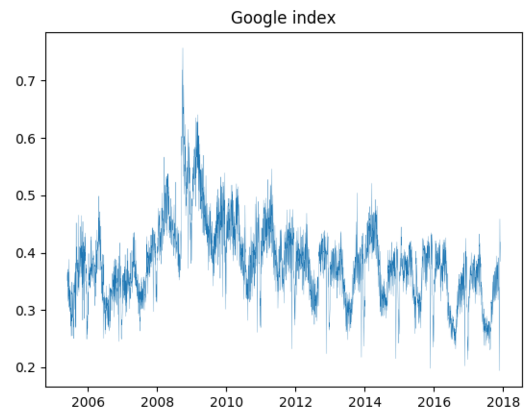
(c) Economy



(d) Inflation



(e) Unemployment



(f) Google Trends index

Figure 2: Individual search volume series next to Google Trends index

Figure 1 illustrates how the index level develops over time. After fluctuating around its mean in the beginning of the sample, the search interest reaches its peak after the global financial crisis of 2008. Hence, the peak of the Google Trends index is at the same time as the most volatile period of the S&P 500. In addition, Figure 2 displays the constituents of the index individually. Aside from 'inflation', all series exhibit a sharp increase around the crisis. The search volume of 'crisis' experiences a few smaller peaks every time more news about the European sovereign debt crisis emerge. Also the series of 'unemployment' follows a different pattern than the other constituents. After the sudden rise in search interest, the series returns to its pre-crisis level only slowly. This is an indication for the fact after workers lost their jobs at the beginning of the financial crisis continue to look for the term 'unemployment' until the economy picks up again and they find a new job.

Table 1: Summary statistics of the Google Trends data

	Crisis	Default	Economy	Inflation	Unemployment	Google index
Mean	0.271	0.594	0.377	0.447	0.221	0.384
Std deviation	0.084	0.110	0.104	0.113	0.119	0.069
Max	1.000	1.000	1.000	1.000	1.000	0.757
Min	0.000	0.226	0.117	0.000	0.000	0.194
Skewness	1.746	0.518	0.589	0.337	1.340	0.673
Kurtosis	8.328	0.114	1.764	0.839	1.936	1.339
T	3146	3146	3146	3146	3146	3146

3.2 Stock Data

I obtain daily returns of the S&P 500 index from the Realized Library of the Oxford-Man Institute of Quantitative Finance. As the index is often perceived as an accurate gauge of the U.S. equity market, modeling its volatility helps to quantify the current uncertainty in the U.S. economy and financial markets. The dataset consists of 3146 observations from 27^{th} May, 2005 to 29^{th} December, 2017 and includes periods of high volatility around the financial crisis in 2008 in the first half as well as less volatile periods in the second half. I report the summary statistics of the entire sample and the separate halves in the Appendix.

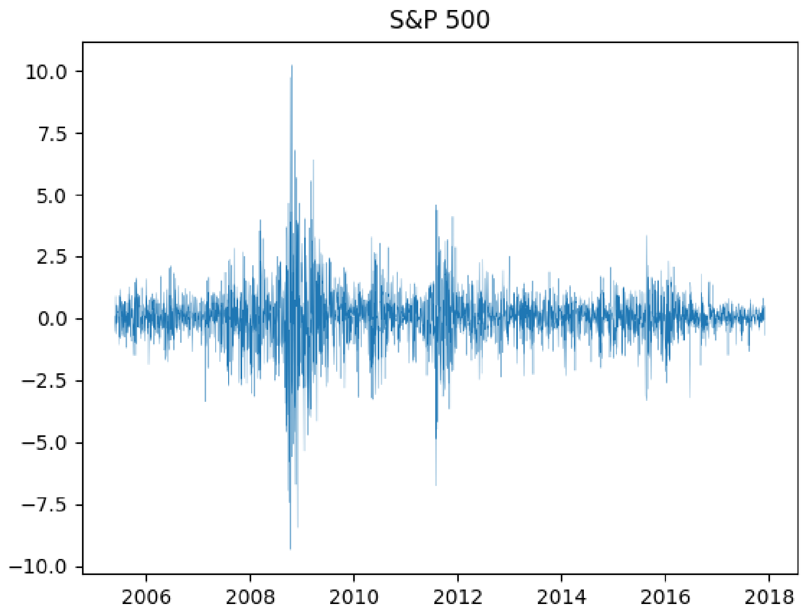


Figure 3: Daily S&P500 returns

Conditional variance itself is a latent variable. Therefore, I need to substitute a proxy for model evaluation purposes. Following Liu, Patton, and Sheppard (2015), I employ realized variance based on 5-minute high frequency data, which is also available in the Realized Library. In general, I only use the realized variance to compute the loss functions and not as a model input.

4 Model

As a starting point for all models presented in this paper I specify the following process for the stock returns: Let r_t be the daily return at time t and I assume that

$$r_t = \mu + \epsilon_t,$$

where $\mathbb{E}[\epsilon_t|\mathcal{I}_{t-1}] = 0$, and $\mathbb{E}[\epsilon_t^2|\mathcal{I}_{t-1}] = \sigma_t^2$ and $\mathcal{I}_{t-1} = \{r_{t-1}, r_{t-2}, ...\}$ is the information set available at time $t-1$. Furthermore, let

$$\epsilon_t = z_t \sigma_t, \tag{1}$$

with $z_t \sim GWN(0, 1)$, such that $\mathbb{E}[r_t|\mathcal{I}_{t-1}] = \mu$ and $Var[r_t|\mathcal{I}_{t-1}] = \mathbb{E}[(r_t - \mu)^2|\mathcal{I}_{t-1}] = \sigma_t^2$. The standard GARCH(1,1) specification introduced by Bollerslev (1986) is already able to model volatility clustering and determines the conditional variance σ_t^2 as

$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2. \tag{2}$$

This process is stationary, if $\alpha + \beta < 1$ and to ensure a positive conditional variance $\alpha, \beta > 0$. In some papers Equation (2) is referred to as a 'filter' and only uses backward-looking information.

4.1 Varying parameter specification

To relax the assumption of a stationary variance, assume that σ_t^2 is varying and measurable by an additive structure

$$\sigma_t^2 = h_t + g_t, \tag{3}$$

where h_t is a stationary process and follows the GARCH(1,1) model

$$h_t = \omega + \alpha\epsilon_{t-1}^2 + \beta h_{t-1} \tag{4}$$

and g_t induces non-stationarity in (3). Teräsvirta and Amado (2008) define g_t as

$$g_t = (\omega_* + \alpha_*\epsilon_{t-1}^2 + \beta_* h_{t-1})G(\cdot), \tag{5}$$

where $G(\cdot)$ is a non-negative, continuous function, bounded between zero and one and the standard specification of Equation (2) is nested in (3) if $g_t = 0$. However, instead of letting $G(\cdot)$ depend on $t^* = \frac{t}{T}$, with T being the total number of observations, so the GARCH parameters vary smoothly over time, I deviate from their approach and let $G(\cdot)$ depend on the exogenous variable x_t. Then, the parameters fluctuate with the level of the exogenous variable and I can combine (3) - (5) to

$$\sigma_t^2 = \omega(x_{t-1}) + \alpha(x_{t-1})\epsilon_{t-1}^2 + \beta(x_{t-1})\sigma_{t-1}^2, \tag{6}$$

with $\omega(x_t) = \omega + \omega_* G(x_t)$, $\alpha(x_t) = \alpha + \alpha_* G(x_t)$, and $\beta(x_t) = \beta + \beta_* G(x_t)$ and to ensure positivity of the conditional variance $\omega(x_t), \alpha(x_t), \beta(x_t) > 0$. Moreover, for estimation I employ the condition $\alpha(x_t) + \beta(x_t) < 1$, which establishes second order stationarity.

According to Teräsvirta and Amado (2008), a reasonable choice for $G(x_t)$ is the general logistic function defined as

$$G(x_t; \gamma, c) = (1 + e^{-\gamma(x_t - c)})^{-1}. \tag{7}$$

When $x_t = c$, $G(x_t; \gamma, c) = 0.5$ and I select c to be the sample mean of x_t. The parameter γ controls, how smoothly the function transitions from zero to one. Figure 4 illustrates, how the transition function evolves for different values of γ. More specifically, if $\gamma = 1$ the function is a straight line, while it approaches the step function for large values of $\gamma > 100$. Then, the transition becomes steep and transforms the model to some extent into a switching model whose dynamics change immediately, as soon as the exogenous variable deviates from its mean level.

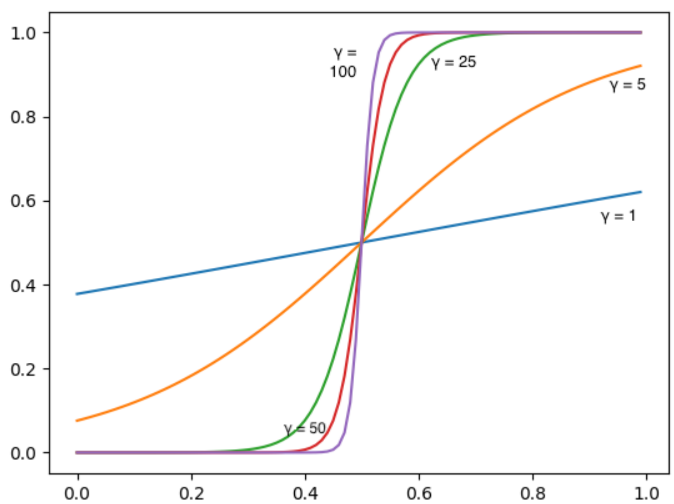


Figure 4: Plots of the transition function with location parameter $c = 0.5$ and smoothness level $\gamma = 1, 5, 25, 50, 100$.

4.2 Constant parameter specification

A popular method to include exogenous variables into the volatility dynamics is to add the additional regressors directly to the variance in (2), such that

$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \pi x_{t-1}. \tag{8}$$

The model needs the same restrictions as the standard GARCH specification, namely $\omega, \alpha, \beta > 0$ and $\alpha + \beta < 1$.

Additionally, $\pi > 0$ and the exogenous variable is often squared to ensure positivity, but as the Google Trends series are bounded between zero and 100, a transformation is not necessary in this case. Compared to the standard GARCH specification in Equation (2), the extended-GARCH (GARCH-X) in Equation (8) introduces only one more parameter to be estimated, but the parameters are not allowed to vary and the external information enters linearly. Even though, the GARCH-X is therefore not able to model non-stationarity in the conditional variance, according to Han (2015) it still provides additional explanations to the stylized facts of returns that the standard GARCH(1,1) cannot capture. Yet, this property depends on the explanatory power of the covariate.

4.3 Spline-GARCH

Another approach to include exogenous information is to follow Engle and Rangel (2008) who separate slow-moving trends and seasonality in volatility from the standard GARCH process by fitting an exponential quadratic spline. Their spline-GARCH model is defined as follows:

$$\epsilon_t = \sqrt{\tau_t \sigma_t^2} z_t, \tag{9}$$

where σ_t follows the standard GARCH specification as in (2) and $z_t \sim GWN(0, 1)$. The spline τ_t is defines as

$$\tau_t = c \exp(w_0 t + \sum_{i=1}^{k} w_i \max(t - t_{i-1}, 0)^2 + \pi x_t), \tag{10}$$

with exogenous information x_t and k knots at $t_1, t_2, \ldots t_k$. The number of knots is chosen by minimizing the Bayesian Information Criterion (BIC) and there is an equal distance between each knot. The difference to the previous specifications is that in this case the exogenous information enters multiplicatively. Furthermore, the spline is fitted before estimation, so the parameters are determined conditional on the position of the knots. In that sense, the model also incorporates forward-looking information. Through the smooth spline this specification is able to handle non-stationarity, but in highly volatile periods of small datasets the spline might be too smooth to react to shocks in the variance.

4.4 Estimation

Under the assumption that the error terms are normally distributed, estimates of the parameter vector θ for all model specifications are obtained by maximizing the log-likelihood function

$$\ell(\theta) = \sum_{t=1}^{T} -\frac{1}{2} ln(2\pi) - \frac{1}{2}\log[\sigma_t^2(\theta)] - \frac{\epsilon_t^2}{2\sigma_t^2(\theta)} \tag{11}$$

with respect to θ. The fact that ϵ_t is assumed to be Gaussian does not directly imply that the unconditional distribution of the returns is Gaussian. Under this assumption the unconditional distribution can still present the excess kurtosis that is often observed in stock returns. In this paper the normal distribution is chosen for simplicity and to increase computational speed. Nonetheless, I provide an example with t-distributed error terms in Section 6.5 as a Sensitivity Analysis.

Further, I compute the standard errors around the parameter estimates using the so-called 'sandwich' form by Bollerslev and Wooldridge (1992), which is defined as

$$SE(\theta) = \text{diag}((\mathcal{J}^{-1}\mathcal{I}\mathcal{J}^{-1})^{\frac{1}{2}}), \tag{12}$$

where $\mathcal{J}$ is the expected Hessian and $\mathcal{I}$ is the covariance matrix of the log-likelihood scores. Both are approximated numerically (Sheppard, 2012).

In the case of varying parameters, $\sigma_t^2(\theta) = h_t(\theta_1) + g_t(\theta_2)$ with $\theta_1 = (\omega, \alpha, \beta)'$ being the standard GARCH parameters and $\theta_2 = (\omega_*, \alpha_*, \beta_*, \gamma)'$, while I choose the smoothness level γ before estimation. If γ is estimated alongside the the other parameters, its standard error is large (a multiple of 10). However, this is not a sign of its insignificance, but rather an indication that θ_2 and the log-likelihood are insensitive to small changes in γ and an exact estimate is not necessary. Therefore, it suffices to search for the neighborhood of the true γ

on a fairly coarse grid by choosing it before estimation and taking the value which yields the biggest likelihood.

For GARCH models with time-varying parameters Dahlhaus et al. (2006) showed that the time-varying model can be locally approximated by stationary models such that the maximum likelihood estimates are consistent and have asymptotic properties. In the case of this analysis, where the model parameters vary with the level of the exogenous variable, I verify the model and the estimation accuracy by conducting a simulation study.

5 Simulation Study

In this section I conduct a simulation experiment to assure that all parameters of the proposed specification in Section 4.1 are estimated correctly by maximum likelihood and evaluate, how the dynamics of the varying parameter specification enter in a model with constant parameters. Further, I verify that a Likelihood Ratio Test (LRT) can be applied to detect, whether adding an external variable improves the fit significantly. All calculations and simulations are run in Python 3.5. For the numerical maximization of the log-likelihood function I employ the Sequential Least Squares ("SLSQP") (Kraft, 1988) algorithm from the SciPy package.

Following the additive structure of Equation (3) in Section 4.1, I consider two data generating processes ('DGP'). In the first case

$$DGP \quad 1: \quad \sigma_t^2 = h_t + g_t, \tag{13}$$

where h_t again follows the GARCH(1,1) model

$$h_t = \omega + \alpha\epsilon_{t-1}^2 + \beta h_{t-1}$$

and

$$g_t = (\omega_* + \alpha_*\epsilon_{t-1}^2 + \beta_* h_{t-1})G(x_{t-1}).$$

In the second case

$$DGP \quad 2: \quad \sigma_t^2 = h_t, \tag{14}$$

with $g_t = 0$, so there is no influence of external factors on the conditional variance. The parameters for both simulations are shown in Table 2. I choose the values of the DGP parameters after a test estimation on empirical data. In this estimation, γ was computed by maximum likelihood alongside the other parameters. Results for different values for γ are reported in the Appendix. In DGP 1, the past returns lose their influence, when the exogenous variable increases, while at the same time the process becomes more persistent. On the other hand, DGP 2 is a standard GARCH process with constant parameters.

Table 2: Simulation parameters

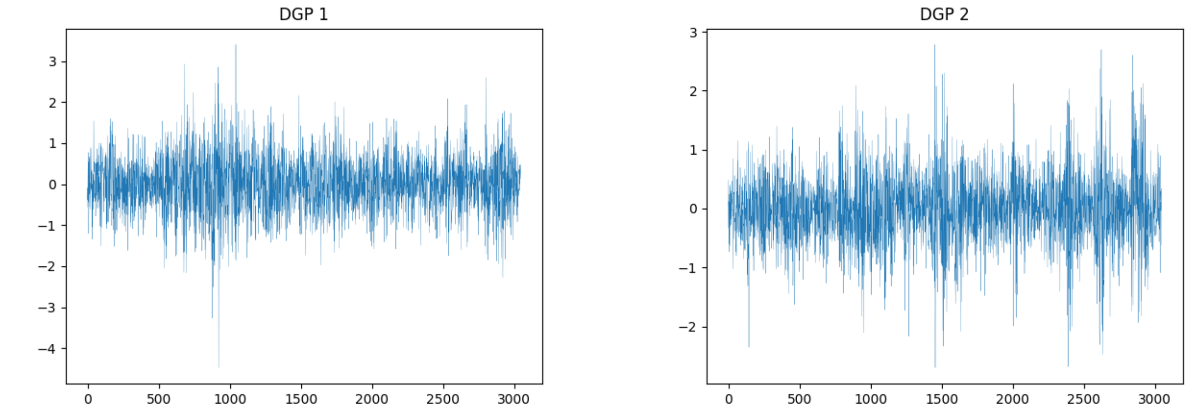
	ω	α	β	ω_*	α_*	β_*	γ
DGP 1	0.012	0.12	0.85	0.09	-0.11	0.14	77
DGP 2	0.012	0.12	0.85	-	-	-	-

I consider a sample size of $T = 3146$ and discard the first 100 observations to reduce the initialization effect, and $N = 5000$ repetitions. Figure 5(a) and (b) show examples of return series based on DGP 1 and DGP 2, respectively. Both use the same random seed, but only Figure 5(a) includes the effects of the exogenous variable. Therefore, the period in which the graph exhibits the highest volatility coincides with the period of the bad news index reaching its peak and the largest variations in return cluster around the 1000th observation. In Figure 5(b) on the other hand, the exogenous information has no influence and the volatility clustering appears randomly.

Figure 6 (a) depicts the first 100 autocorrelations of absolute returns created by DGP 1 (blue) and DGP 2 (orange). The horizontal lines represent the 95% and 99% confidence bounds. The degree of persistence created by the first process is higher than the one of a standard GARCH process and therefore closer to the behavior of the S&P 500 dataset (Figure 6(b)) and also consistent with slowly declining autocorrelations of absolute returns in other datasets.

For the remainder of the simulation study I simulate the exogenous information x_t alongside the returns to improve its reliability for other datasets and assume x_t to follow an autoregressive process of order 1, such that

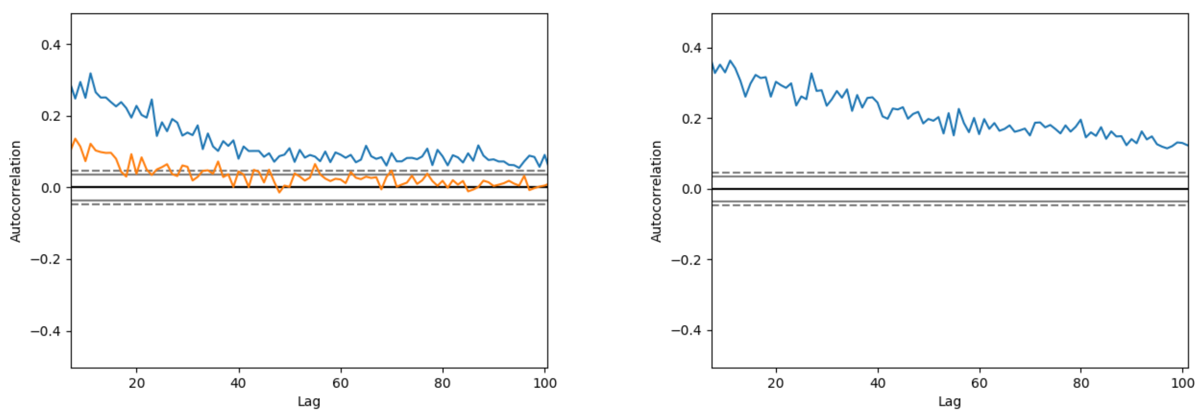
$$x_t = 0.38 + 0.85x_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, 0.04^2).$$



(a) Varying parameter specification

(b) standard GARCH specification

Figure 5: Simulated returns of GARCH processes with and without varying parameters



(a) Varying parameter specification (blue) and standard GARCH specification (orange)

(b) S&P 500

Figure 6: The first 100 autocorrelations of absolute returns with 95% (and 99%) confidence intervals.

The coefficients of the process correspond to the parameters estimates of an AR(1) model fitted to the Google Trends index.

5.1 Estimation accuracy

Table 3 displays the true values plus the mean, standard deviation, median, 2.5% and 97.5% percentiles of the sample of estimated parameters. I choose γ in advance, so it is not estimated alongside the other parameters. The numbers in Table 3 show the case, when γ is indeed set to its true value. The other cases of γ being misspecified as either too high or too low are reported in the Appendix in Table 15. The means and medians of the h_t parameters θ_1 in Panel A of Table 3 do not significantly differ from their true values. In contrast, the parameter vector θ_2 of g_t is not estimated as precisely and both mean and median indicate that the magnitude of the coefficients is underestimated. Especially the estimation of β_* is relatively inaccurate and has the largest error. Its true value of 0.14 is undershot by both mean (0.051) and median (0.073). Nonetheless, the true value lays within one standard deviation from the mean. In general, the 95% confidence interval, given by the percentiles, includes the true values in all cases. If γ is misspecified, the errors around all estimates increase and the means of θ_2 are further off their true values. Yet, the model still estimates θ_1 accurately and for both θ_1 and θ_2 the percentiles contain the true values.

The results for θ_1 of DGP 2 in Panel B of Table 3 correspond to the results for DGP 1. Again, both means and medians are close to the true values of θ_1 and the confidence intervals are small. Because θ_2 is unspecified in DGP 2, the standard deviation around α_* and β_* increase and the means shrink closer to zero. Yet, with a value of 0.014 ω_* is close to the mean of ω (0.012), although ω_* is unspecified. From the simulation results in Panel B I cannot definitely conclude that $g_t = 0$. Therefore, I specify a test in Section 5.3 to investigate, if g_t is indeed insignificant, even though not all elements of the parameter set θ_2 are estimated to be zero.

Table 3: Simulation results of the varying parameter GARCH for 5000 replications

Panel A							
DGP 1	Parameter	True value	Mean	Std.	Median	2.50%	97.50%
θ_1	ω	0.012	0.014	0.005	0.014	0.006	0.024
	α	0.12	0.121	0.015	0.120	0.093	0.152
	β	0.85	0.857	0.017	0.857	0.822	0.888
θ_2	ω_*	0.09	0.032	0.041	0.024	-0.020	0.130
	α_*	-0.11	-0.080	0.034	-0.082	-0.139	-0.008
	β_*	0.14	0.051	0.086	0.073	-0.183	0.156
	γ	77	77				
	LL		-3408.94				

Panel B							
DGP 2	Parameter	True value	Mean	Std.	Median	2.50%	97.50%
θ_1	ω	0.012	0.012	0.004	0.012	0.005	0.020
	α	0.12	0.123	0.017	0.122	0.094	0.158
	β	0.85	0.847	0.019	0.849	0.808	0.880
θ_2	ω_*	-	0.014	0.028	0.008	-0.018	0.084
	α_*	-	-0.006	0.042	-0.006	-0.087	0.082
	β_*	-	-0.045	0.099	-0.026	-0.299	0.089
	γ	-	77				
	LL		-2627.49				

Notes: True values plus the mean, standard deviation, median, 2.5% and 97.5% percentiles of the sample of estimated parameters of the varying parameter specification on DGP 1 and DGP 2. γ is chosen in advance, so it is not estimated alongside the other parameters by maximum likelihood.

5.2 Constant versus varying parameter specification

Further, I investigate, if a model with a constant parameter specification, where the exogenous variable enters linearly is able to grasp the nonlinear effects induced by g_t. Table 4 includes the summary statistics of the sample of GARCH-X parameters estimates based on 5000 replications for DGP 1 and DGP 2. The mean of π in Panel A indicates that the linear model is able to pick up the external influence to a certain extent, but the means of ω and β deviate from their true values with differences of 0.005 and 0.013, respectively. The mean of ω is also more than two standard deviations away from its true value. So, the constant specification does not collect all of the influence of g_t in π and distributes it across the constant and the autoregressive component. On the other hand, the model captures the influence of the squared returns accurately, as the mean of α (0.112) almost matches the true value of 0.12. In general, the model's average log-likelihood is smaller than the one of the varying parameter specification and when both are estimated on the same simulated returns, the varying parameter model yields a better fit in 97% of the cases.

In Panel B both mean and median of π rightfully decrease and converge to zero with values of 0.006 and 0.000, respectively. Also the means of ω, α and β are closer to their true values. In fact the mean of ω matches the true value of 0.012 exactly and its standard deviation decreases by a factor of 4.5 compared to the estimates in Panel A. On the other hand, the noise around β increases slightly. Essentially, this analysis indicates that a model with constant parameter specification is not able to identify the two different constants ω and ω_* induced by DGP 1. As a consequence for the application on empirical data, large standard errors around the estimates of ω and π of the GARCH-X suggest that in this case the exogenous information does not enter linearly and the varying parameter specification will provide a better fit to the data.

Table 4: Simulation results of the GARCH-X for 5000 replications

			Panel A				
DGP 1	Parameter	True value	Mean	Median	Std.	2.50%	97.50%
	ω	0.012	0.007	0.006	0.018	0.001	0.016
	α	0.12	0.112	0.112	0.013	0.087	0.137
	β	0.85	0.863	0.863	0.015	0.832	0.891
	π	g_t	0.026	0.026	0.032	0.002	0.056
	LL		-3411.52				

			Panel B				
DGP 2	Parameter	True value	Mean	Median	Std.	2.50%	97.50%
	ω	0.012	0.011	0.012	0.004	0.002	0.019
	α	0.12	0.119	0.119	0.014	0.093	0.149
	β	0.85	0.845	0.845	0.017	0.811	0.876
	π	0	0.006	0.000	0.010	0.000	0.032
	LL		-2629.11				

Notes: True values plus the mean, standard deviation, median, 2.5% and 97.5% percentiles of the sample of estimated parameters of the GARCH-X on DGP 1 and DGP 2.

5.3 Likelihood Ratio Test

Finally, to test, whether there is a relationship between the external variable and the conditional variance, and including it by means of the varying parameter specification improves the fit, I propose to use a Likelihood Ratio Test (LRT). More specifically, the standard GARCH model with parameters θ_1 is nested in the the varying parameter specification, if $g_t = 0$ and consists of four parameters less. The LRT compares the goodness of fit of both models with the null hypothesis that the models do not differ significantly, such that the more parsimonious model is preferred. The statistic of the test in this case is defined as

$$D = -2\frac{\mathcal{L}(\hat{\theta}_1)}{\mathcal{L}((\hat{\theta}_1, \hat{\theta}_2)')} \tag{15}$$

and follows a chi-squared distribution with four degrees of freedom (Casella & Berger, 2002). If the null hypothesis is rejected, the varying parameter specification, with parameters θ_1 and θ_2, improves the fit significantly. Otherwise the external variable has no meaningful influence on the conditional variance.

To examine the validity of the LRT in this context, I simulate again from both DGPs. When applying the test combined with a 5% significance level on N simulations from DGP 1 that include the external effects, I expect a rejection rate of 95%. Equivalently, when simulating from DGP 2, I expect a rejection of the test in 5% of the cases. Nevertheless, this also requires that the statistic is indeed chi-squared distributed.

Table 5: Rejection rates of the LRT for 100 and 1000 simulations

DGP 1			DGP 2		
N	Rejections	Rate	N	Rejections	Rate
100	94	0.94	100	3	0.03
1000	920	0.92	1000	25	0.025

Notes: Theoretically true rejection rates: 0.95 for DGP1; 0.05 for DGP2

Table 5 contains the results of this analysis. In both cases, the rejection rates are slightly lower than they should be. However, they are still close enough to their theoretically true values, such that the test is valid when applied on empirical data and the chi-squared distribution is a legitimate choice.

6 Empirical Results

6.1 Setup

In this section, I illustrate the relation between the bad news index based on Google search volume and the conditional variances of S&P 500 returns. I examine the degree if nonlinearity of this relationship by choosing the smoothness parameter γ such that the resulting varying parameter specification (vpx-GARCH) provides the best fit. Then, I compare the performance of this specification against a standard GARCH model, the GARCH-X and the spline-GARCH with regard to goodness of fit and accuracy and employ the Likelihood Ratio Test (LRT) to determine, whether the bad news index carries explanatory power. Besides the log-likelihood, I report the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Both criteria are based on the likelihood, but penalize models for including extra parameters. For ranking the models in terms of accuracy, I follow the approach of Patton (2011) and compute the Mean Squared Error (MSE), which is robust against noise in the realized variance proxy. Additionally, I report the Mean Absolute Error (MAE) as a second measure, which is less prone to outliers, but at the same time less robust to noise in the proxy to investigate, if both measures imply the same ranking.

Secondly, I split the sample in two halves to investigate, if the effect of including the bad news index differs for the period during the financial crisis and the more stable times afterwards. In addition to the in-sample estimates of the full sample, I generate 300 out-of-sample one-step rolling window forecasts. For each model this involves repeatedly placing the last return and estimate of the rolling sample in the respective GARCH filters after the estimation of the parameters and I set the cutoff point at the three quarters of each sample.

Then, I use these out-of-sample forecasts to calculate 300 daily 95% Value-at-Risk (VaR) estimates for both halves and determine how often the actual daily losses exceed the VaR predictions. Ultimately, I employ the Kupiec test (Kupiec, 1995), which can reject a model for too many or too few violations.

Before each estimation, I de-mean the return series to assure that I only capture the effect of the external information on the conditional second moments. Again, I run all models on Python 3.5 and use the SLSQP algorithm (Kraft, 1988) for the numerical minimization of the negative log-likelihood function of all model specifications. I initialize the optimization multiple times at random starting values that satisfy all constraints and save the estimates that yield the most optimal log-likelihood value.

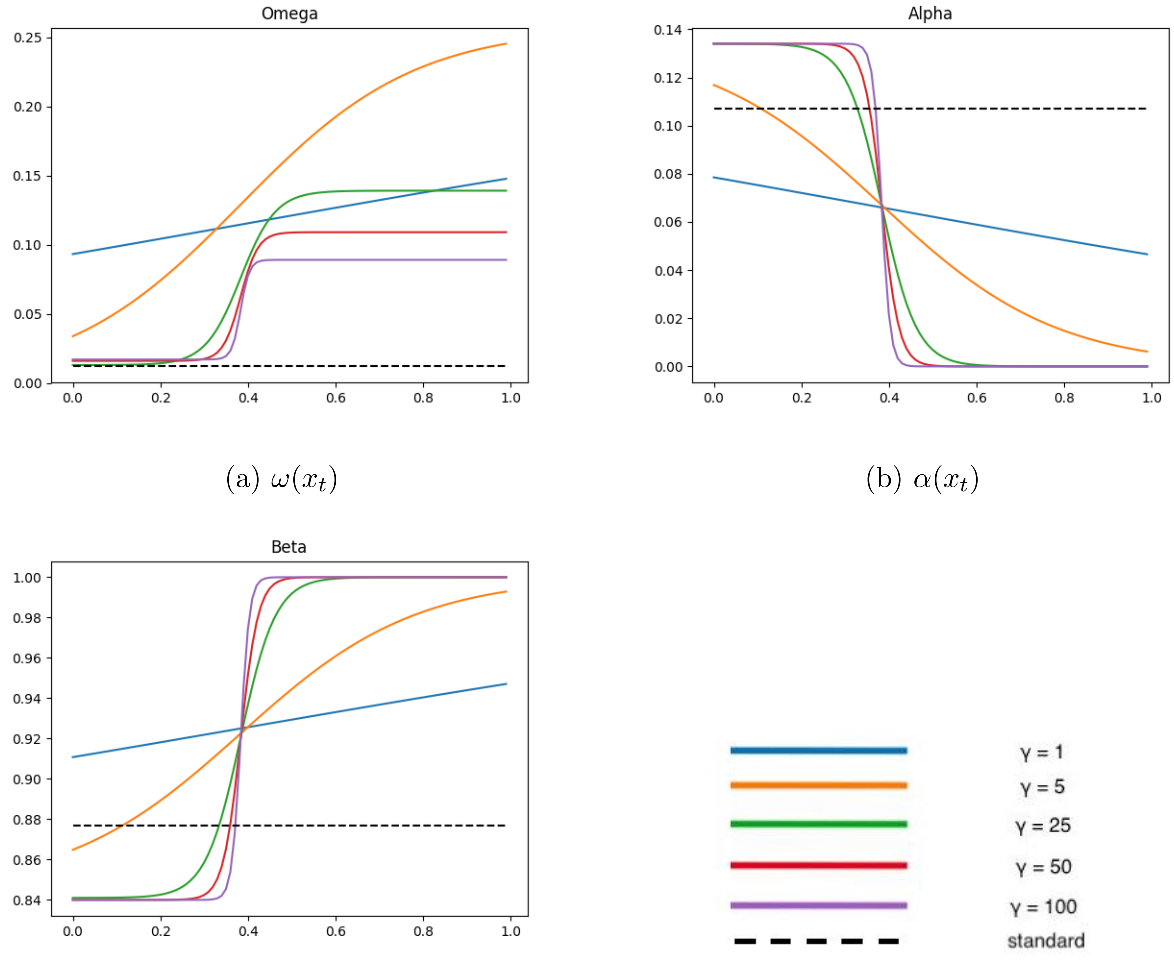
6.2 Full sample results

Table 6: Estimates and log-likelihoods for different values of γ

Full sample	γ	ω	α	β	ω_*	α_*	β_*	LL
	1	0.002	0.132	0.850	0.225	-0.132	0.150	-3855.94
		(0.001)	(0.033)	(0.043)	(0.080)	(0.047)	(0.158)	
	5	0.001	0.134	0.845	0.256	-0.134	0.155	-3851.39
		(0.001)	(0.033)	(0.044)	(0.085)	(0.046)	(0.156)	
	25	0.013	0.134	0.841	0.126	-0.134	0.159	-3845.69
		(0.005)	(0.021)	(0.022)	(0.066)	(0.033)	(0.113)	
	50	**0.016**	**0.134**	**0.840**	**0.093**	**-0.134**	**0.160**	**-3844.49**
		(0.006)	(0.020)	(0.022)	(0.068)	(0.032)	(0.108)	
	100	0.017	0.134	0.840	0.072	-0.134	0.160	-3844.72
		(0.007)	(0.020)	(0.023)	(0.069)	(0.032)	(0.106)	

Notes: Estimates of the varying parameter specification on the full sample of S&P 500 returns. γ is chosen before estimation. Numbers in brackets are standard errors as in Bollerslev and Wooldridge (1992) based on numerical derivatives developed by Sheppard (2012). Bold numbers highlight the parameter set with the highest log-likelihood.

Table 6 contains the parameter estimates and log-likelihood for different values of γ and Figure 7 illustrates the corresponding transition functions of the parameters for different smoothness levels. The dotted line in the graph represents the constant estimates of a standard GARCH estimates (Table 18 in the Appendix). The intercept $\omega(x_t)$ is rather sensitive to changes in γ as ω increases and ω_* decreases, when γ increases. From $\gamma = 1$ to 100, ω increases by 8.5 times and ω_* at $\gamma = 100$ is only a third of its value at $\gamma = 1$. The opposite effect is visible for $\beta(x_t)$ but the relative differences are smaller. Finally, $\alpha(x_t)$ is the least sensitive and raising γ only steepens the transition.



(a) $\omega(x_t)$

(b) $\alpha(x_t)$

(c) $\beta(x_t)$

(d) Legend

Figure 7: Transition functions of the parameters for different values of γ

Both $\omega(x_t)$ and $\beta(x_t)$ are increasing in x_t, which means that a high level of the news index raises the base variance level and makes the dynamics more persistent. Contrarily, $\alpha(x_t)$ is a decreasing function and even falls to zero for in extreme cases. So, the dynamics of the conditional variance are solely driven by the autoregressive component and past shocks lose their influence entirely, as soon as the Google search volume index is above its mean.

For all values of γ the standard errors around the parameters that are multiplied by the transition function are higher than their standard counterparts θ_1. For instance, the standard error around β_* is about five times larger than around its counterpart β. Except for ω, all standard errors are decreasing functions of γ.

The model fits best for $\gamma = 50$ and therefore I select this specification for the remainder of the empirical analysis. However, the difference in likelihood to even higher values of γ is only minor. At a log-likelihood of -3844.49 the vpx-GARCH including the Google Trends index provides a better fit than the standard GARCH specification, given in Table 7. Furthermore, the LRT statistic introduced in the previous section then exceeds the critical values at 95% (9.49) and 99% (13.28) with a value of 40.20. The same specification including a series of random draws between zero and one instead of the Google Trends index, yields a log-likelihood of only -3863.24 and leads to a rejection of the model. Thus, adding the bad news index as exogenous information improves the likelihood significantly and carries explanatory power for the conditional second moment.

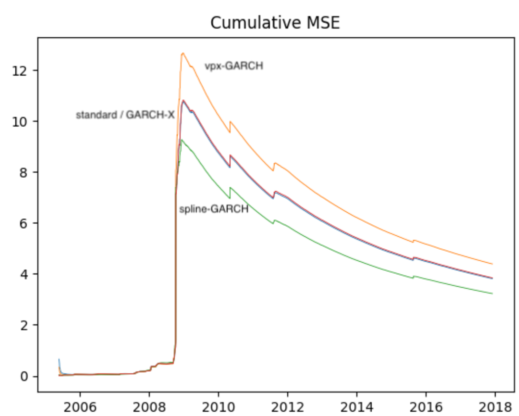
Based on the other fit criteria provided in Table 7, the varying parameter specification also beats the more parsimonious GARCH-X, because both its AIC and BIC are smaller. Nonetheless, the spline method, which uses forward-looking information, still provides the best fit out of all specifications with a log-likelihood of -3808.87. The spline-GARCH beats the other specifications even with regards to AIC and BIC, although it involves the estimation of 12 parameters with seven knots.

The differences between the realized variance proxy and the estimated conditional variances contradict these results. Although the spline-GARCH also produces the lowest mean absolute and mean squared errors, the varying parameter specification ranks last in both measures and the standard GARCH is in the second place. The bad performance of the vpx-GARCH can be explained by its underestimation of a few huge outliers in the proxy. In October 2008, the graph of the realized variance displays a spike. This spike is so large in magnitude that its underestimation makes up a large portion of both MSE and MAE. At this point in time, the bad news index is above its mean, so the process is highly persistent, but neglects the influence of the past return. Therefore, the model cannot react to the outlier. Figure 8 shows the cumulative MSE and MAE of all four models. The spike of the orange line in both subplots hence suggests that the estimate of the vpx-GARCH misses the outliers in 2008 to a greater extent than the other specifications. After the spike, the orange line actually depreciates at a higher rate, which is a reason for the high likelihood of the vpx-GARCH.

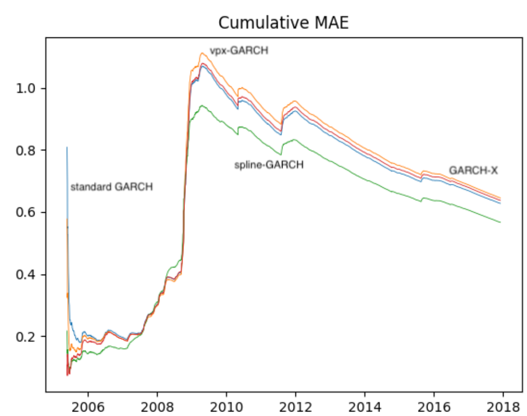
Table 7: Fit and errors of different specifications based on the full sample

	Fit			in-sample	
	LL	AIC	BIC	MSE	MAE
standard GARCH	-3864.59	7735.18	7747.83	3.813	0.628
vpx-GARCH	-3844.49	7702.98	7732.49	4.390	0.645
GARCH-X	-3860.14	7728.27	7745.13	3.845	0.638
spline-GARCH	**-3808.87**	**7641.73**	**7692.32**	**3.227**	**0.567**

Notes: Bold numbers represent the value, which ranks first for a certain criterion. Specifications: spline-GARCH, 7 knots; vpx-GARCH $\gamma = 50$



(a) Cumulative MSE



(b) Cumulative MAE

Figure 8: Cumulative MSE and MAE of in-sample fitted values

6.3 Time-varying effects

Table 8: Split sample estimates for $\gamma = 50$

Sample	γ	ω	α	β	ω_*	α_*	β_*
Full sample	50	0.016	0.134	0.840	0.093	-0.134	0.160
		(0.006)	(0.020)	(0.022)	(0.068)	(0.032)	(0.108)
27.05.2005 - 02.09.2011	50	0.018	0.105	0.874	0.092	-0.105	0.126
		(0.011)	(0.017)	(0.021)	(0.134)	(0.041)	(0.114)
06.09.2011 - 04.12.2017	50	0.017	0.241	0.755	0.128	-0.202	-0.094
		(0.007)	(0.041)	(0.042)	(0.058)	(0.088)	(0.110)

Notes: Estimates of the varying parameter specification on the full sample and both sample halves of the S&P 500 returns. γ is set to 50 before estimation. Numbers in brackets are standard errors as in Bollerslev and Wooldridge (1992) based on numerical derivatives developed by Sheppard (2012).

The parameter estimates of the full sample are mainly characterized by the effect of the global financial crisis. The signs of the estimates in Table 8 that are based only on the first half of the dataset, which includes the crisis, are equal to the estimates of the entire sample and they only differ in magnitude. As soon as the bad news index is above its mean level, the past returns lose their influence and the process becomes highly persistent in the autoregressive component $\beta(x_t)$.

On the other hand, estimating the model on the second, more stable half makes $\beta(x_t)$ a decreasing function, such that for this sample, the model sets the conditional variance mainly by shifting the constant and the past shocks never lose their influence entirely, as $\alpha + \alpha_* > 0$. In this half, the standard error around β_* is with a value of 0.110 larger in an absolute sense than the coefficient of -0.094. Hence, it is likely that $\beta(x_t) = \beta$ and therefore a constant.

In both halves, including the bad news index in the model improves the likelihood significantly, as the LRT statistic exceeds the critical values at 95% (9.49) and 99% (13.28) with values of 16.92 in the first half and 16.55 in the second one. However, the estimates

Table 9: Fit, in-sample and out-of-sample errors based on two halves of the entire sample

	Fit			in-sample		out-of-sample	
27.05.2005 - 02.09.2011	LL	AIC	BIC	MSE	MAE	MSE	MAE
standard GARCH	-2242.33	4490.67	4501.24	**7.217**	**0.919**	**2.359**	**0.567**
vpx-GARCH	**-2233.87**	**4481.75**	4506.41	8.201	0.959	2.418	0.632
GARCH-X	-2236.94	4481.87	**4495.97**	7.544	0.966	2.374	0.603
spline-GARCH	-2802.00	5628.00	5670.28	26.697	3.718	15.330	3.731

	Fit			in-sample		out-of-sample	
06.09.2011 - 04.12.2017	LL	AIC	BIC	MSE	MAE	MSE	MAE
standard GARCH	-1612.46	3230.92	3241.49	0.473	0.362	0.123	0.230
vpx-GARCH	-1604.19	3222.37	3247.03	0.411	0.335	0.178	0.253
GARCH-X	-1612.46	3232.92	3247.01	0.473	0.362	**0.123**	**0.228**
spline-GARCH	**-1570.65**	**3165.30**	**3207.57**	**0.382**	**0.315**	0.572	0.720

Notes: Bold numbers highlight the best model in each category. Specification: For both halves, $\gamma = 50$ in the vpx-GARCH and 7 knots in the spline-GARCH.

of the GARCH-X (Table 16 in the Appendix) would suggest that the Google Trends have no effect in the second half, because the coefficient π is 0. Therefore, both GARCH-X and standard GARCH also yield the same likelihood and predictions in the bottom panel of Table 9. Nonetheless, the simulation study indicates that the LRT fails to reject slightly more often than it should. This consideration gives the rejection of the null in the second half more weight and suggests that the exogenous information enters nonlinearly and does have explanatory power.

With regards to out-of-sample MSEs and MAEs, the more parsimonious standard GARCH and the GARCH-X outperform the vpx-GARCH and the spline-GARCH in both halves. Especially during the financial crisis, the spline seems to be too smooth to cope with the high variances and the errors explode. Its MSE of 15.330 and MAE of 3.731 are far off the second to last vpx-GARCH with an MSE and MAE of 2.418 and 0.632, respectively. Because the standard GARCH model outranks the models that include the bad news index in both halves, it seems that the Google Trends data lacks out-of-sample predictive power.

Regarding in-sample errors, the vpx-GARCH performs well in the second half compared to its high errors in the full sample. Because this sample half does not include the outliers around the financial crisis, the model ranks second with an MSE of 0.411 and MAE of 0.335 and is only beaten by the spline-GARCH. In this case the rankings according to the goodness of fit measured by the likelihood and the in-sample errors actually coincide. However, in the first half the outliers during the financial crisis distort this ranking again. Even though the vpx-GARCH provides the best fit in terms of likelihood its MSE of 8.201 is affected by the outliers in the variance proxy and only ranks third behind the standard GARCH (7.212) and the GARCH-X (7.544). Consequently, the distorted ranking of the full sample in Table 7 is also caused by the effects of the financial crisis.

6.4 Value-at-Risk violations

With the 300 one-step out-of-sample forecasts of each sample as a basis I estimate the daily Value-at-Risk (VaR) at time t with probability level p as

$$VaR_p(r_t) = \mu - \sigma_t \Phi^{-1}(p), \tag{16}$$

where $\Phi^{-1}(p)$ is the quantile of a standard normal distribution, μ is the mean return and σ_t is the predicted conditional volatility coming from the GARCH model. The VaR is a risk measure that describes the worst possible loss of an investment, excluding the worst outcomes in the tail of the return distribution that have a probability of $1 - p$. At a probability level $p = 0.95$, the daily loss is expected to exceed the VaR estimate in 5% of the cases. Given the 300 predictions, a model's VaR predictions should be violated exactly 15 times.

Table 10 contains VaR failures of 300 out-of-sample predictions and the corresponding Kupiec statistic. The test follows a chi-squared distribution with one degree of freedom. Hence, a model is rejected for too high or low failures on a 95% significance level, if the statistic exceeds the critical value of 3.84. At first glance the spline-GARCH seems to perform well, because it yields the lowest number of violations with zero in the first half and four in the

Table 10: 95% VaR violations and Kupiec (1995) test statistics

95% VaR	Violations		Kupiec test statistic	
	first half	second half	first half	second half
standard GARCH	17	8	**0.27**	4.11
GARCH-X	17	8	**0.27**	4.11
vpx-GARCH	16	9	**0.07**	**2.93**
spline-GARCH	0	4	30.77	11.85

Notes: All numbers are based on 300 out-of-sample forecasts. Given this sample size, a model's predictions should be violated 15 times at the probability level of 95%. Bold values do not exceed the critical value of 3.84 (95%) and the model is not rejected for too many / too few violations.

second one. Nonetheless, the Kupiec test rejects the model for overestimating the VaR both times. Moreover, the standard specification and the GARCH-X underestimate the quantile in the second half. Consequently, the only model, which is not rejected in both samples is the vpx-GARCH and therefore yields the most accurate predictions of the quantile.

6.5 Sensitivity Analysis

In this section, I turn back to the entire sample to evaluate how my empirical results change, if I (1) change the conditional distribution of the returns and (2) use Principal Component Analysis ('PCA') as a more sophisticated method to create the Google Trends index. Table 11 is the replication of Table 7 of Section 6.2. The difference is that all numbers are computed under the assumption of t-distributed error terms. Following the approach of (Egan, 2007), I fit a standardized t-distribution with a mean of 0 and a variance of 1 to the return data. The estimated degrees of freedom of 3.5 lead the new distribution to have considerably heaver tails than a standard Gaussian. As a result, all models provide a better fit compared to their counterparts in Table 7 and also the in-sample errors are smaller. However, changing the conditional distribution does not change the rankings in terms of fit and in-sample errors.

Only in this case the log-likelihood of the standard GARCH and the GARCH-X are equal, so there is no difference between the models with regard to the goodness of fit.

Looking at the vpx-GARCH, its difference in likelihood to the standard GARCH is still significant as the LRT statistic of 25.05 exceeds the critical value of the confidence level at 99%. Hence, changing the conditional distribution does not change the significance of the Google Trends index. Unfortunately, the optimizer does not converge after changing the distribution for the spline-GARCH model. Therefore, it is uncertain whether this specification still provides the best fit under a different residual distribution.

Table 11: Full sample results with t-distributed residuals

	Fit			in-sample	
	LL	AIC	BIC	MSE	MAE
standard GARCH	-3781.80	7569.60	7582.25	3.538	0.452
vpx-GARCH	-3769.27	7552.55	7582.06	3.914	0.461
GARCH-X	-3781.65	7588.16	7571.29	3.538	0.450
spline-GARCH	-	-	-	-	-

Notes: In-sample fit and errors of the different specifications using a standardized t-distribution with 3.5 degrees of freedom as residual distribution. Due to convergence problems, '-' are reported for the spline-GARCH

In a second effort, I perform a PCA on all five Google Trends index constituents. PCA is a method to transform a set of seemingly correlated variables into a set of linearly uncorrelated principal components. The aim of this transformation is that the first principal component accounts for most of the variation in the original data. In this case, the first component explains 43% of the variance and is displayed in Figure 9 in the Appendix. However, taking the component as the Google Trends index instead of computing the simple average, does not improve the likelihood of the vpx-GARCH on the full sample. A comparison of Table 12 to Table 6 in Section 6.2 illustrates that the bad news index based on the simple average

Table 12: vpx-GARCH with PCA index

Full sample	γ	ω	α	β	ω_*	α_*	β_*	LL
	1	0.002	0.132	0.849	0.226	-0.132	0.151	3855.19
		(0.002)	(0.029)	(0.024)	(0.069)	(0.076)	(0.296)	
	5	0.003	0.133	0.845	0.228	-0.133	0.155	3850.26
		(0.007)	(0.025)	(0.023)	(0.118)	(0.055)	(0.230)	
	25	0.014	0.133	0.843	0.085	-0.133	0.157	3846.32
		(0.006)	(0.020)	(0.022)	(0.073)	(0.033)	(0.109)	
	50	0.016	0.133	0.843	0.069	-0.133	0.157	3846.32
		(0.006)	(0.020)	(0.022)	(0.067)	(0.029)	(0.103)	
	100	0.016	0.133	0.843	0.065	-0.133	0.157	3846.86
		(0.006)	(0.020)	(0.021)	(0.063)	(0.026)	(0.100)	

Notes: Estimates of the varying parameter specification on the full sample of S&P 500 returns including the first principal component as bad news index. γ is chosen before estimation. Numbers in brackets are standard errors as in Bollerslev and Wooldridge (1992) based on numerical derivatives developed by Sheppard (2012).

yields a larger likelihood at all smoothness levels except for $\gamma = 1$.

7 Discussion & Conclusion

In this paper, I amend the time varying parameter GARCH model of Teräsvirta and Amado (2008) by a bad news index based on Google search volume data of keywords with a negative financial connotation. The result is a model, whose parameters vary with the level of the bad news index and implicitly with time. A simulation study then verifies the accuracy of the estimated maximum likelihood parameters and that a Likelihood Ratio Test (LRT) against a standard GARCH(1,1) model can be applied to confirm the explanatory power of the exogenous regressor.

As a consequence of the promising simulation results, I put the varying parameter model to test by applying it to daily returns of the S&P 500. This stock index usually acts as gauge for the U.S. stock market and economy. On a sample ranging from May, 2005 until December, 2017 the LRT suggests a significant increase of the log-likelihood, which means that the bad news index adds explanatory power. The same is true, when estimating the model on both halves of the sample separately. However, the effect of the external variable differs for both cases and the impact of the global financial crisis 2008 seems to dominate the dynamics of the conditional variance.

In sum, including a bad news index based on Google Trends in a GARCH model via a logistic transition function improves the likelihood significantly. When the investors' interest for topics with a negative financial connotation, such as inflation, rises, the stock market volatility increases. Thus, additionally to its explanatory ability for stock returns as in (Preis et al., 2013) and (Da et al., 2011, 2014), Google Trends data also improves the goodness of fit of models for the returns' second moment.

To some, this might be a surprising outcome, because usually retail investors rather than financial institutions use Google as one of their primary sources of information before making financial decisions. On the other hand, institutional investors rely on professional services like Bloomberg or Reuters. For the most part, the actions of retail investors do not impact the market remarkably, but because they tend to 'panic sell', the increased volatility, when Google search volume rises, might be probable. Additionally, breaking news, which lead both investor groups to act in the market, appear in the professional news services and the Google feed almost simultaneously and Google search behavior might therefore also explain the actions of institutional investors.

Nonetheless, the predictive ability of the Google Trends data seems to be limited. Even though, including the bad news index in the varying parameter specification, an extended-GARCH, and the spline-GARCH by Engle and Rangel (2008) beats a standard GARCH model in terms of fit, this most parsimonious model still yields the most accurate out-of-sample predictions of the conditional variance. This drawback could be attributed to the rather large standard errors around the estimates that depend on the external information. On the other hand, given the outcomes of the Kupiec (1995) test, the varying parameter specification is the only model whose out-of-sample Value-at-Risk predictions are not rejected in both sample halves.

Throughout the entire paper, I work with the assumption of normally distributed stock returns for simplicity and computational speed. An extension to a distribution which takes the leptokurtosis of returns into account and goes beyond the sensitivity analysis is easily possible and would probably improve the viability for non-academic purposes. Further, the bad news index could be extended by more search keywords, which would make it smoother, but could also erase its explanatory power. Also the creation of a 'good news' index appears possible and it would be interesting to see how the dynamics would change. Finally, the analysis of other markets is the next logical step and would further uncover the explanatory power of Google Trends as an economic and financial indicator.

References

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2001). The distribution of realized exchange rate volatility. *Journal of the American statistical association*, *96*(453), 42–55.

Black, F. (1976). Studies of stock price volatility changes. In *Proceedings of the 1976 meeting of the business and economic statistics section.* Washington, D.C: American Statistical Association.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327.

Bollerslev, T., & Wooldridge, J. M. (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric reviews*, *11*(2), 143–172.

Brenner, R. J., Harjes, R. H., & Kroner, K. F. (1996). Another look at models of the short-term interest rate. *Journal of financial and quantitative analysis*, *31*(1), 85–107.

Casella, G., & Berger, R. L. (2002). Statistical inference. In (Vol. 2). Duxbury Pacific Grove, CA.

Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention. *The Journal of Finance*, *66*(5), 1461–1499.

Da, Z., Engelberg, J., & Gao, P. (2014). The sum of all FEARS investor sentiment and asset prices. *The Review of Financial Studies*, *28*(1), 1–32.

Dahlhaus, R., Rao, S. S., et al. (2006). Statistical inference for time-varying ARCH processes. *The Annals of Statistics*, *34*(3), 1075–1114.

Egan, W. J. (2007). The distribution of S&P 500 index returns. *SSRN Electronic Journal*.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*, 987–1007.

Engle, R. F., & Rangel, J. G. (2008). The spline-GARCH model for low-frequency volatility and its global macroeconomic causes. *The Review of Financial Studies*, *21*(3), 1187–1222.

Han, H. (2015). Asymptotic properties of GARCH-X processes. *Journal of Financial Econometrics*, *13*(1), 188-221.

Han, H., & Kristensen, D. (2014). Asymptotic theory for the QMLE in GARCH-X models with stationary and nonstationary covariates. *Journal of business & economic statistics*, *32*(3), 416–429.

Kraft, D. (1988). A software package for sequential quadratic programming. *Forschungsbericht- Deutsche Forschungs- und Versuchsanstalt fur Luft- und Raumfahrt*.

Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives*, *3*(2), 73–84.

Liu, L. Y., Patton, A. J., & Sheppard, K. (2015). Does anything beat 5-minute RV? a comparison of realized measures across multiple asset classes. *Journal of Econometrics*, *187*(1), 293–311.

Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, *160*(1), 246–256.

Preis, T., Moat, H. S., & Stanley, H. E. (2013). Quantifying trading behavior in financial markets using Google Trends. *Scientific reports*, *3*.

Sheppard, K. (2012). Introduction to Python for econometrics, statistics and data analysis. *Self-published, University of Oxford, version*, *2*.

Skypala, P. (2017). Investment culture in europe stunts growth in etf market. *Financial Times*.

Smith, G. P. (2012). Google internet search activity and volatility prediction in the market for foreign currency. *Finance Research Letters*, *9*(2), 103–110.

Taleb, N. N., & Goldstein, D. (2007). We don't quite know what we are talking about when we talk about volatility. *Journal of Portfolio Management*, *33*(4).

Teräsvirta, T., & Amado, C. (2008). *Modelling conditional and unconditional heteroscedasticity with smoothly time-varying structure* (Tech. Rep.). SSE/EFI, Working Paper Series in Economics and Finance.

Vlastakis, N., & Markellos, R. N. (2012). Information demand and stock market volatility. *Journal of Banking & Finance*, *36*(6), 1808–1821.

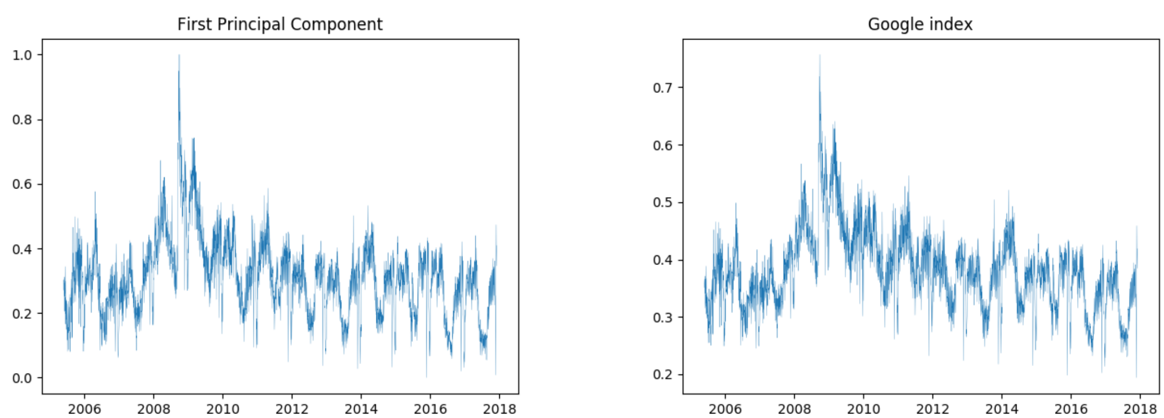
Zakoian, J.-M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control*, *18*(5), 931–955.

Appendices

Tables

Table 13: Summary statistics S&P 500

S&P 500	27.05.2005 - 04.12.2017	27.05.2005 - 02.09.2011	06.09.2011 - 04.12.2017
Mean	0.02	0.01	0.04
Std Dev.	1.13	1.39	0.78
Max	10.22	10.22	4.11
Min	-9.35	-9.35	-3.66
Skewness	-0.30	-0.26	-0.24
Kurtosis	11.75	9.21	3.00
No. Obs.	3146	1573	1573



(a) First Principal Component (b) Average

Figure 9: First Principal Component vs simple average as index

Table 14: Simulation results for $\gamma = 5$

	Parameter	True value	Mean	Std.	Median	2.50%	97.50%
θ_1	ω	0.01	0.012	0.020	0.010	0.000	0.030
	α	0.12	0.129	0.023	0.130	0.090	0.170
	β	0.85	0.850	0.027	0.850	0.810	0.880
θ_2	ω_*	0.09	0.112	0.114	0.100	-0.030	0.310
	α_*	-0.11	-0.112	0.038	-0.120	-0.170	-0.020
	β_*	0.14	-0.027	0.184	0.040	-0.430	0.180
	γ	5	5				
	LL		-3447.13				

Notes: The number of simulations is N=5000

Definitions

The Kupiec (1995) test statistic is defined as

$$LR_K = 2 * \ln[(1 - \frac{N}{T})^{T-N}(\frac{N}{T})^N] - 2\ln[(1-p)^{T-N}p^N],$$

where p is the expected rejection frequency, N is the number of violations and T is the number of predictions. The test is chi-squared distributed with one degree of freedom with the null hypothesis $p = \frac{N}{T}$.

Table 15: Simulation results with falsely specified smoothness parameter γ

	Panel A						
	Parameter	True value	Mean	Std.	Median	2.50%	97.50%
θ_1	ω	0.01	0.012	0.010	0.011	0.000	0.032
	α	0.12	0.126	0.019	0.124	0.094	0.166
	β	0.85	0.851	0.019	0.852	0.811	0.886
θ_2	ω_*	0.09	0.131	0.567	0.101	-0.032	0.322
	α_*	-0.11	-0.110	0.034	-0.116	-0.162	-0.020
	β_*	0.14	0.017	0.161	0.088	-0.355	0.177
	γ	5	75				
	LL		-3481.41				

	Panel B						
	Parameter	True value	Mean	Std.	Median	2.50%	97.50%
θ_1	ω	0.01	0.012	0.014	0.011	0.000	0.032
	α	0.12	0.126	0.020	0.125	0.093	0.167
	β	0.85	0.849	0.033	0.851	0.814	0.883
θ_2	ω_*	0.09	0.043	0.587	0.100	-0.032	0.319
	α_*	-0.11	-0.112	0.033	-0.118	-0.159	-0.031
	β_*	0.14	0.014	0.168	0.093	-0.368	0.177
	γ	75	5				
	LL		-3328.29				

Notes: Panel A contains the results for the case, when the true $\gamma = 5$ and for estimation γ is falsely set to 75. Panel B contains the results for the opposite case. The number of simulations is $N = 5000$.

Table 16: GARCH-X estimates

GARCH-X	ω	α	β	π
27.05.2005 - 04.12.2017	0.003	0.110	0.875	0.034
27.05.2005 - 02.09.2011	0.002	0.092	0.897	0.039
06.09.2011 - 04.12.2017	0.021	0.195	0.791	0.000

Table 17: standard GARCH estimates

standard GARCH	ω	α	β
27.05.2005 - 02.09.2011	0.014	0.107	0.877
27.05.2005 - 02.09.2011	0.017	0.091	0.894
06.09.2011 - 04.12.2017	0.021	0.195	0.791

Table 18: Full sample estimates with random data

Full sample	γ	ω	α	β	ω_*	α_*	β_*	LL
	1	0.018	0.128	0.851	-0.018	-0.128	0.149	3860.72
		(0.051)	(0.027)	(0.025)	(0.750)	(0.079)	(0.352)	
	5	0.017	0.116	0.866	-0.017	-0.056	0.045	3863.19
		(0.007)	(0.020)	(0.021)	(0.078)	(0.064)	(0.119)	
	25	0.015	0.108	0.876	-0.015	0.005	-0.013	3863.28
		(0.006)	(0.018)	(0.020)	(0.057)	(0.056)	(0.100)	
	50	0.015	0.107	0.878	-0.015	0.014	-0.013	3863.25
		(0.006)	(0.018)	(0.020)	(0.057)	(0.063)	(0.097)	
	100	0.015	0.106	0.878	-0.015	0.016	-0.013	3863.24
		(0.005)	(0.018)	(0.020)	(0.053)	(0.061)	(0.089)	

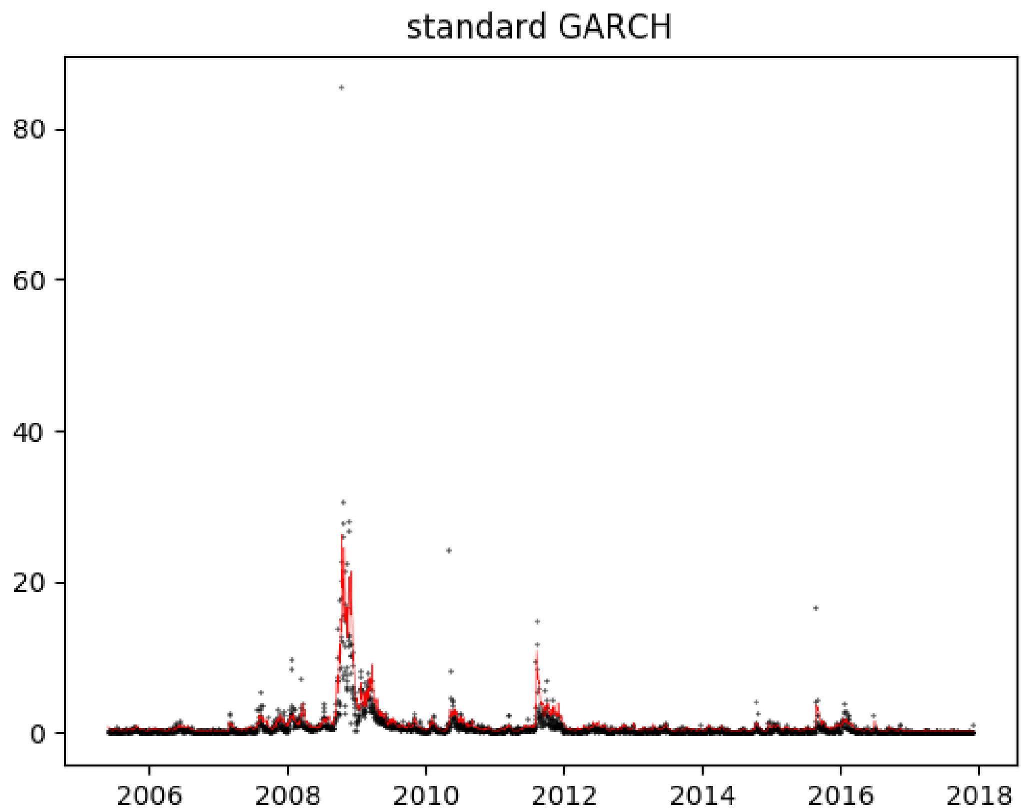


Figure 10: standard GARCH fit (red line) vs realized variance (dots)

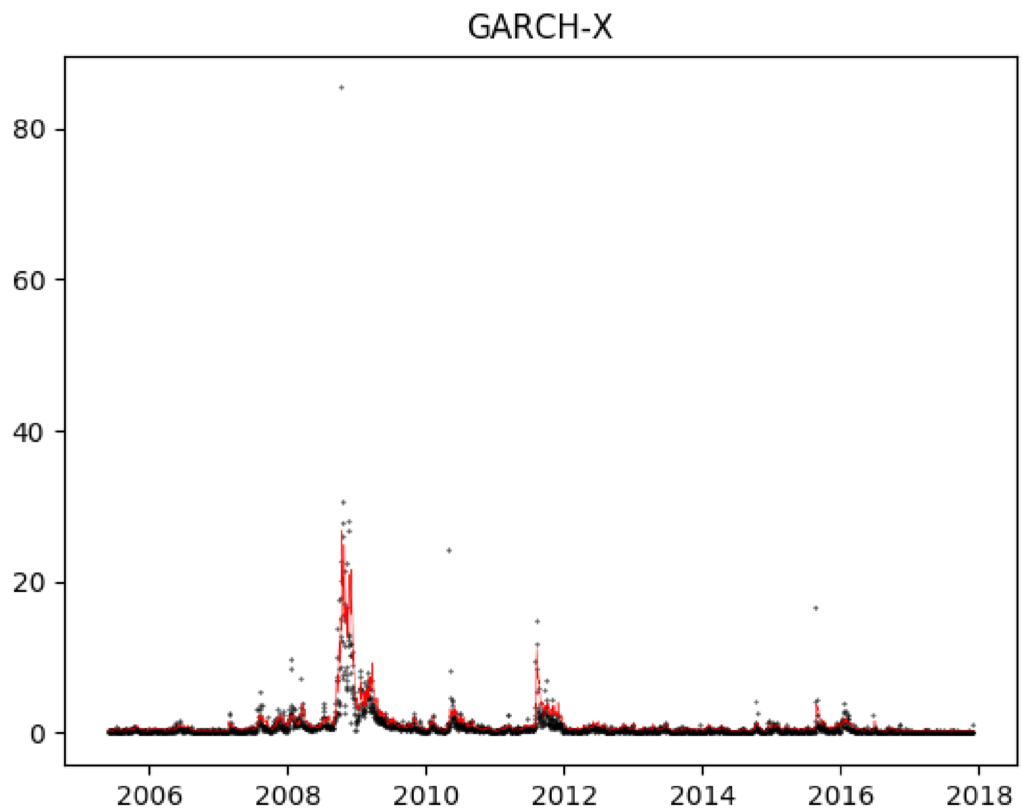


Figure 11: GARCH-X fit (red line) vs realized variance (dots)

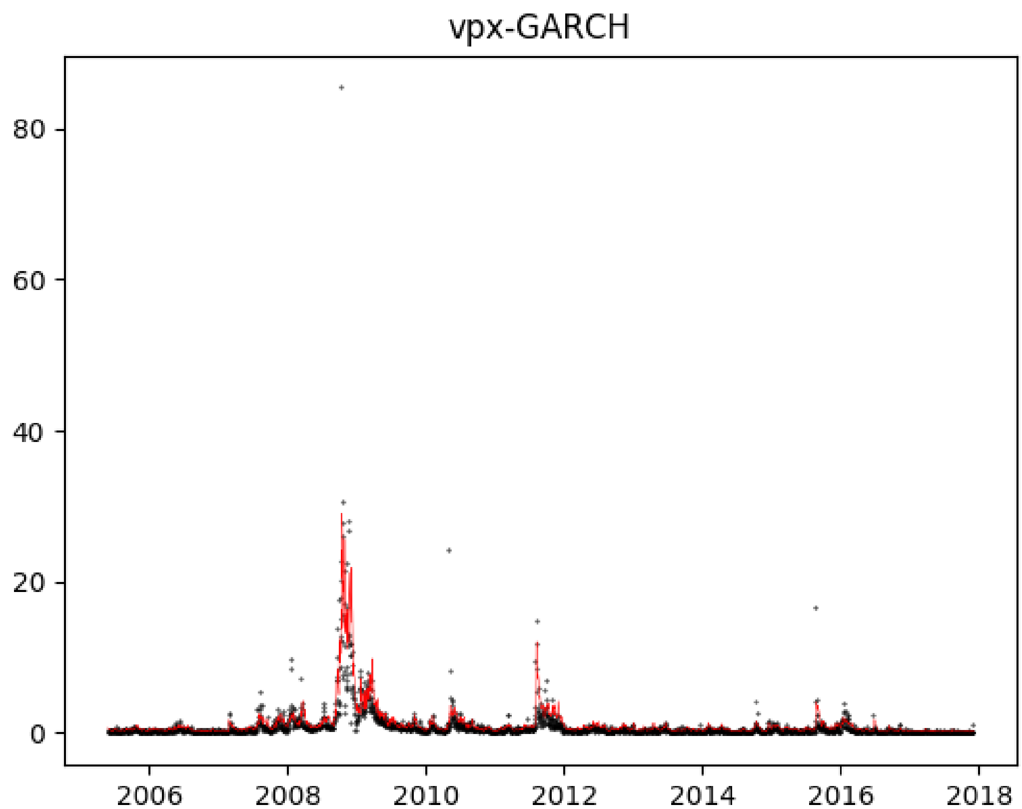


Figure 12: vpx-GARCH fit (red line) vs realized variance (dots)

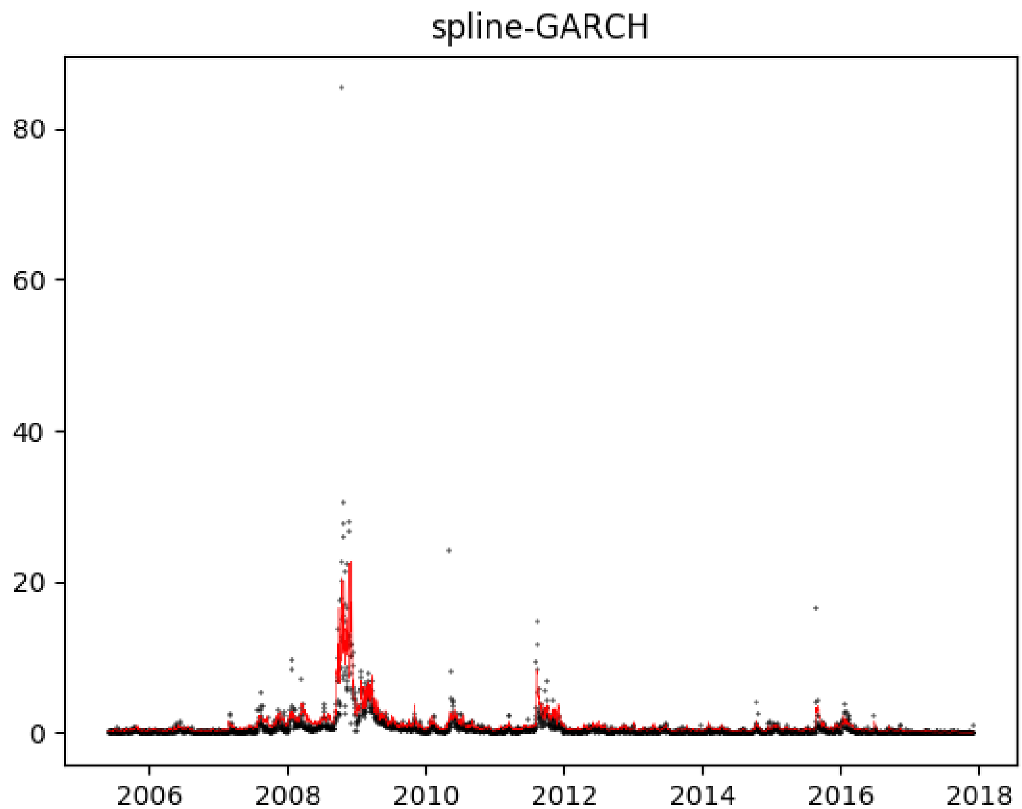


Figure 13: spline-GARCH fit (red line) vs realized variance (dots)