Exploiting Traded Volume for Forecasting Volatility of Financial Assets

by

Yasin Şimşek

A Dissertation Submitted to the
Graduate School of Social Sciences and Humanities
in Partial Fulfillment of the Requirements for
the Degree of

Master of Arts

in

Economics

KOÇ ÜNİVERSİTESİ

June 30, 2021

Exploiting Traded Volume for Forecasting Volatility of Financial Assets

Koç University

Graduate School of Social Sciences and Humanities

This is to certify that I have examined this copy of a master's thesis by

Yasin Şimşek

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

_____

Dr. Cem Çakmaklı

_____

Prof. Kamil Yılmaz

_____

Dr. Umut Gökçen

Date: _____

[To my family]

ABSTRACT

Exploiting Traded Volume for Forecasting Volatility of Financial Assets
Yasin Şimşek
Master of Arts in Economics
June 30, 2021

In this paper, we propose a novel GARCH specification which embraces traded volume in a statistically coherent and computationally efficient way. To capture the dependence between traded volume and asset return volatility, the Mixture of Distribution Hypothesis (MDH) is used. We model the time evolution of the latent variable(return variance) ,which drives the bivariate relation in MDH, in the spirit of Generalized Autoregressive Score modelling technique. The resulting volume enhanced GARCH model has quite parsimonious yet flexible structure. Since the peculiar form of score driven models, the likelihood can be obtained in closed form and hence estimation procedure is computationally inexpensive. In an empirical application with IBM, PG and MSFT stock data, we estimate our novel model with its 3 variants. Moreover, standard GARCH models from the literature are considered for comparison. Our results highlight the superiority of volume enhanced models both in-sample estimates and out-of-sample predictions among competing models. Particularly, when we consider returns as t distributed, the gain from traded volume becomes pronounced in out-of-sample prediction study.

ÖZETÇE

Finansal Varlıkların Oynaklık Tahmininde Hacim Kullanımı
Yasin Şimşek
Ekonomi, Yüksek Lisans
30 Haziran 2021

Bu makalede, işlem hacmi verisini içeren, istatiksel olarak mantıklı ve hesaplama açısından pahalı olmayan bir GARCH modeli türevi öne sürdük. Hacim verisi ve getiri oynaklığı arasındaki ilişkiyi yakalayabilmek için Mixture of Distribution Hypothesis (MDH) modelini kullandık. Hacim ve oynaklık arasındaki ikili ilişkiyi yöneten bilgi oranı (belirlenen sabit bir güne göre) verisinin zaman içerisindeki evrimi Generalized Autoregressive Score (GAS) modellerine öykünerek yazıldı. Ortaya çıkan hacim ile geliştirilmişGARH modeli basit ancak esnek bir yapıya sahiptir. GAS modellerinin özel formu sayesinde, modelin olabilirlik fonksiyonu kapalı bir formda elde edilebilir ve tahmin prosedürü yoğun hesaplamalar gerektirmez. Öne sürülen modeli ve 2 varyantını IBM, PG and MSFT verileriyle çalıştık. Buna ek olarak literaturden standard GARCH modelleri karşılaştırma yapmak için analize dahil edildi. Sonuçlarımız, hacim ile geliştirilmiş GARCH modellerinin örneklem içindeki ve dışındaki tahminlerinin üstünlüğünü gösteriyor. Özellikle, t dağılımı getiri verisi dikkate alındığında, hacim verisi kullanımın faydaları belirginleşiyor.

ACKNOWLEDGMENTS

First and foremost, I would like to express my sincere thanks to my advisor Dr.Cem Çakmaklı for giving me invaluable support and guidance in every phases of this research. Especially, I want to say many thanks since he did not bothered with my midnight messages.

I am grateful to my family for their prayers, support and love. Without their support this thesis may not be possible. Also, I am exteremely thankful to my dear Zeynep for her understanding and continuous support. Last but not the least I thank my friends from master class Ekin, Elif, Sündüz, Talha, İrfan, Yakup, Orhan and Okan. They make my years in İstanbul enjoyful. My special thanks goes to my roommate Onur. I am very grateful to him because of productive and endless midnight discussions.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

NOMENCLATURE

ACD	Autoregressive Conditional Duration
ACI	Autoregressive Conditional Intensity
GAS	Generalized Autoregressive Score
GARCH	Generalized Autoregressive Conditional Heteroscedasticity
MDH	Mixture of Distributions Hypothesis
NBER	National Bureau of Economic Research
RMSFE	Root Mean Squared Forecast Error

Chapter 1

INTRODUCTION

Modelling the stock return volatility has been a growing research area in the econometrics literature. Understanding the evolution of volatility has importance from different perspectives. For instance, volatility estimates can be used in option pricing or portfolio allocation. Moreover, financial asset volatility could be an indicator of economic environment. Consequently, both non-parametric and parametric methods, such as GARCH, Stochastic Volatility, Realized Volatility etc., have been developed for modelling volatility of financial assets. Specifically, GARCH model family has attracted intensive attention thanks to quite flexible yet parsimonious structure and ability of capturing commonly observed features of the return data. There are abundant variants of GARCH model including extensions with multivariate data sets, threshold constraints, regime switching, leverage, semi-parametric methods etc.

It is empirically well documented that return volatility and traded volume are positively related with each other. While GARCH models are studied in very broad perspective, introducing volume to volatility or let say variance equation could not reach an end beyond some trivial extensions. Not surprisingly, GARCH model can be augmented by relating traded volume with volatility through variance equation in a linear fashion. In case of using contemporaneous volume, this model becomes subject to the endogeneity problem. Even if we circumvent this issue by choosing lags of traded volume, still we are ending up with serious limitations. First, we omit possible non-linear relationships. Second, since we do not characterize the stochastic structure of traded volume, an out-of-sample forecasting exercise may not be statistically coherent. Another strand of the literature

documents positive relationship between volatility and volume. To account for this empirical finding, a bivariate specification with a stochastic volatility representation named as Mixture of Distribution Hypothesis (MDH) is offered. In contrast to former naive attempts, MDH puts stochastic structure on volume and relates its distribution with volatility. Moreover, it considers information arrivals as the source of volume-volatility co-movement. Despite remarkable theoretical innovations, MDH fails to capture high persistence in volatility which is commonly observed in the data. In addition to this drawback, to the best of our knowledge, it is not used to build a GARCH model coupled with traded volume.

In this paper, we propose a novel GARCH specification based on MDH in which traded volume is embraced in a statistically coherent way. Particularly, the distributions of traded volume and stock return and dependence structure are specified in line with MDH which proved to be useful in modelling volume-volatility relationship, see for example [Tauchen and Pitts, 1983], [Andersen, 1996], [Darolles et al., 2015], [Darolles et al., 2017], among others. The novelty of this paper is that the variance equation is modelled in the spirit of Generalized Autoregressive Score (GAS) Models unlike existing studies. GAS models family is such a flexible and general modelling framework that many conventional models like GARCH, ACD, ACI can be contained as a special case. This framework offers a quite intuitive yet useful updating mechanism for time varying volatility depending on the score of local likelihood. Since the score indicates the most ascending direction of predictive likelihood, the time evolution of volatility seeks to maximize local fit of the model. The success of GAS models both in-sample and out-of-sample shown in the comparative study of [Koopman et al., 2016] and many empirical papers like [Creal et al., 2014], [Harvey and Sucarrat, 2014], [Lucas et al., 2014], [Salvatierra and Patton, 2015], [Opschoor et al., 2018]. This study combines the strengths of GAS framework and MDH specification in a statistically consistent way. Indeed, this creates a GARCH model enhanced by traded volume variable. While this model considers stochastic processes of return and volume simultaneously, it pays significant attention to obtain valuable information for volatility from the density of traded volume via score driven variance equation. In our novel

model, we do not stick with one particular specification. Instead, relaxing of distributional assumptions and different parametrizations are studied to ensure the validity of the models in various environments. Specifically, we alter the distribution of returns from normal distribution to t distribution to address heavy tails. Comparing to benchmark model, the variance equation gives more importance to traded volume once we switch to t distribution. Therefore, our model controls the volatility spikes via heavier tails of t distribution and, at the same time, implies more valuable information content of volume for stock return volatility. Moreover, with an additional specification we include leverage effect to our study. In total, we propose a simple model with its 2 extensions and call them as *volume enhanced models*.

The empirical application is conducted with IBM, PG, MSFT stock series between 4 November 2001 and 31 December 2019. We include 3 GARCH models in our analysis to compare them with volume enhanced specifications. Additionally, a GARCH model of [Lamoureux and Lastrapes, 1990], i.e. VA-GARCH, and its extension with t distribution are examined. First, our findings show that introducing traded volume is statistically significant and it is robust to changes in distributional assumptions and re-parametrizations. Second, the volatility persistence implied by volume enhanced models is estimated closely to the persistence levels provided by standard GARCH models. Furthermore, the log-likelihood of stock return in volume enhanced models are close to the GARCH counterparts when we consider the full sample. That is, while volume enhanced models are performing relatively well for IBM and MSFT series, GARCH counterparts outperforms the others. Going one step forward, we divide the whole sample into 2 different periods based on Global Financial Crisis, namely pre-crisis and post-crisis periods. The sub-sample analysis reveals similiar results as in the full sample case.

When we consider out-of-sample forecasting exercise, our results show the merits of volume enhancing models. We compute 1 year rolling window predictions for $h = 1, ..., 10$ step ahead for all models. The evaluation period is 4 November 2002 until 31 December 2019. For comparison, root mean squared forecast error (RMSFE) is computed while taking 5 min realized volatility as a volatility proxy.

Our results indicate that, volume enhanced models‘ forecasts can be outperformed by other models‘ predictions when we consider stock returns as normal. However, the volume enhanced models significantly outperforms remaining models when we switch to t distribution case.

The rest of the paper is organized as follows. Chapter 2 reviews the related works about volume-volatility dynamics and score models. Chapter 3 provides theoretical backgrounds of MDH and GAS models. We describe the volume enhancing models in Chapter 4 and discuss the empirical results in Chapter 5. Finally, Chapter 6 concludes.

Chapter 2

RELATED WORK

2.1 Volume-Volatility Dynamics

The relationship between volume and volatility is extensively studied in econometrics literature for a long time. The early studies [Karpoff, 1987] and [Gallant et al., 1992] empirically document the fact that price changes and volume positively related. The Mixture of Distribution Hypothesis model of [Tauchen and Pitts, 1983] provides an appealing explanation for volume-volatility co-movement. In this specification, the information arrivals to the market are considered as the latent factor which drives the traded volume and return volatility simultaneously. Several papers provide mixed evidences regarding MDH introduced in [Tauchen and Pitts, 1983]. For instance, [Richardson and Smith, 1994] conducts GMM tests whether the unconditional moments implied by MDH conform with the data or not. They conclude that the supportive findings are less strong comparing to previous studies. [Lamoureux and Lastrapes, 1994] relaxes the assumption of serially uncorrelated information arrival process of [Tauchen and Pitts, 1983]. In a seminal paper, [Andersen, 1996] proposes modified MDH model which derives a conditional Poisson distribution for traded volume instead of normal distribution. Similar to the previous papers, he finds that modified MDH outperforms standard MDH model. The superior performance of [Andersen, 1996] is resulted from the fact that he uses statistically coherent distributional specifications and allows serial correlation in information arrivals. [Liesenfeld, 1998] shows that modified MDH cannot account for high persistence in volatility which is commonly observed feature of the data. Consequently, [Liesenfeld, 2001] builds a MDH model with time varying information sensitiveness of traders. In a sense, this paper proposes a new latent factor which drives the volume-volatility relations. Indeed,

the findings demonstrate that MDH model can capture the high persistence in stock return volatility when an additional mixing variable is introduced. [Li and Wu, 2006] study on modified MDH and allows separate effects of different parts of volume. Their results suggest that the positive relationship between volatility and volume driven by the informed part. Also, when the informed volume is controlled, the positive relationship turns out to be a negative. More recently, [Darolles et al., 2015] and [Darolles et al., 2017] assert that one latent variable is not sufficient to accommodate both the short and long run dynamics of volume-volatility relationship. They consider a new second mixing variable: liquidity. While the information variable is responsible for long term dynamics, the liquidity is governing the short term effects. Since the short term effects are supposed to be absorbed within the trading day, the stock return remains unchanged after liquidity variable is introduced. On the other hand, [Darolles et al., 2015] supposes that the liquidty can affect the level of trading volume. They assume that latent variables are independent and serially uncorrelated. However, [Darolles et al., 2017] builds a similar model which permits information arrivals and liquidity to have time persistence. Both papers show that two latent variable specification outperforms all competing models.

2.2 Score Driven Models

Since the family of score driven models introduced by [Creal et al., 2013] and [Harvey, 2013], it has gained notable popularity in time series econometrics literature. Score models provides a general specification which fits well in wide range of empirical models. In addition to this flexibility, the likelihood function can be obtained in a closed form which means low computational cost. Considering these advantages, there is a flourishing strand of the literature regarding their theoretical aspects and empirical applications, despite the very recent discovery. For instance, [Creal et al., 2011] considers a multivariate heavy tailed asset returns model for time varying volatility and correlations using dynamic score structure. [Harvey and Luati, 2014] study the t-location distribution with an application to

the rail travel data by allowing the model parameter update itself with scores. [Creal et al., 2014] proposes a mixed measurement and a mixed frequency score driven model. Many observations coming from various family of distributions and different frequencies are supposed to be shared a common latent factor which has an autoregressive score structure. [Lucas et al., 2014] models Euro Area CDS spreads with score driven skewed t distribution to capture salient features of the data. [Salvatierra and Patton, 2015] builds a dynamic copula model coupled with realized high frequency measures to study stock return volatility dynamics. Correspondingly, [Opschoor et al., 2018] develops a dynamic score model for joint dynamics of fat tailed realized covariance matrix and daily returns. In [Lucas et al., 2017], a dynamic copula model based on GAS framework is developed to asses the conditional and joint default risk for financial sector firms. Similarly, [Oh and Patton, 2018] focuses on systemic risk among US firms by using a high dimensional copula model. More recently, [Thiele, 2020] studies the implications of asymmetric tailed asset return distributions for volatility in a dynamic score model. In addition to macro-finance applications, GAS models can be utilized in many different settings. For example, [Cakmakli and Simsek, 2020], among others, integrates an epidemiological model to GAS framework which brings about a novel dynamic disease model.

Although autoregressive score model family has a diversified empirical application portfolio, it is hard to study its theoretical aspects because of nonlinear dynamics resulted from score functions. [Blasques et al., 2014] establishes strong consistency and asymptotic normality result for maximum likelihood estimators of time varying parameter models driven by score function. [Blasques et al., 2015] show the information theoretic optimality of score driven updates in terms of approximating series of conditional densities of observations. Also, [Koopman et al., 2016] find that GAS models have similar predictive performance comparing to parameter driven state space models. Recently, [Blasques et al., 2020] examines the finite sample optimality properties of score driven models.

Chapter 3

BACKGROUND

The model proposed in this study is built on Mixture of Distribution Hypothesis(MDH) and Generalized Autoregressive Score(GAS) models. To understand the model, it is important to figure out underlying mechanism of MDH and fundamentals of GAS models. In this regard, Section 3.1 is devoted to express MDH and its derivations with an emphasis on market microstructure ideas behind the hypothesis. Then, Section 3.2 describes GAS models starting with its very general form and expressing some special cases at the end.

3.1 Mixture of Distributions Hypothesis

In this study, we closely follow the modified MDH model of [Andersen, 1996]. Assume that each information arrival induces a price discovery phase followed by an equilibrium phase. Let there are J_t information arrival at time t. The associated equilibrium prices are denoted by $P_{j,t}$ for $j = 1, ..., J_t$. Thus, the daily return, R_t, is given by,

$$1 + R_t = \prod_{j=1}^{J_t - 1} \frac{P_{j+1,t}}{P_{j,t}} = \frac{P_{J_t,t}}{P_{1,t}} \tag{3.1}$$

where $P_{1,t}$ denotes the first transaction price. By approximation we can state the following,

$$\begin{aligned} R_t &= \sum_{j=1}^{J_t} \log\left(\frac{P_{j+1,t}}{P_{j,t}}\right) \\ &= \sum_{j=1}^{J_t} \eta_{j,t} \end{aligned} \tag{3.2}$$

where $\eta_{j,t} \sim i.i.d.(0, \sigma_\eta^2)$. Although the number of information arrival, J_t, is assumed to be very large, it may display substantial variation across time. There-

fore, we choose a reference day with J information arrival and let $J_t = K_t J$ for any t where K_t is positive scaling factor and stands for intensity of information arrival relative to benchmark day. Now suppose that $\sigma^2 = J\sigma_\eta^2$ and $\eta_{j,t} = \sigma\epsilon_{j,t}J^{-\frac{1}{2}}$ where $\epsilon_{j,t} \sim i.i.d.(0,1)$. Then, daily return becomes,

$$R_t = \sigma K_t^{1/2}\left[\frac{1}{(JK_t)^{1/2}}\sum_{j=1}^{JK_t}\epsilon_{j,t}\right]. \tag{3.3}$$

If we apply Lindeberg-Levy CLT for the terms in square brackets, we get the following parsimonious yet quite intuitive conditional distribution for daily return variable.

$$R_t \mid K_t \sim N(0, \sigma^2 K_t) \tag{3.4}$$

Let us denote the daily trading volume by V_t. It has two components, $V_t = IV_t + NV_t$, which are named informed and uninformed volume respectively. Uninformed trading volume(or noisy trade, or liquid trade) is assumed to be stochastic process driven by constant Poisson arrivals. That is, $NV_t \sim \text{Poisson}(v_0)$ where v_0 is the arrival intensity per day. Consider a single information arrival to the market. Suppose that I denotes the maximum number of informed traders that may make transaction according to that information. Also assume that $p_{j,t}$ be the probability of trading at jth price discovery phase and at time t. Therefore, at price discovery phase j, the resulting trading volume is a binomially distributed random variable, Binomial$(I, p_{j,t})$. For very large I, it can be considered as Poisson random variable with parameter $Ip_{j,t}$, Poisson$(Ip_{j,t})$. It implies the below specification for informed traded volume in a day with J_t information arrival;

$$IV_t \mid K_t \sim \sum_{j=1}^{J_t} Poisson\left(Ip_{j,t}\right) = Poisson\left(I\sum_{j=1}^{J_t} p_{j,t}\right) \tag{3.5}$$

where the informed volume is independent across information arrivals. Let us examine the expression in parenthesis. For large J, it can be said that

$$\begin{aligned} \sum_{j=1}^{J_t} p_{j,t} &= JK_t \left[\frac{1}{JK_t} \sum_{j=1}^{JK_t} p_{j,t} \right] \\ &= (J \,\mathrm{E}\,(p_{j,t}))K_t \end{aligned} \tag{3.6}$$

in which $J \,\mathrm{E}\,(p_{j,t})$ stands for the expected number of trades in a day with J information arrivals, or benchmark day. Denote this quatity by μ. Therefore, we obtain $IV_t \mid K_t \sim \text{Poisson}(I\mu K_t)$. Consequently, the total trading volume is given by,

$$V_t \mid K_t \sim \text{Poisson}(v_0 + v_1 K_t) \tag{3.7}$$

where $v_1 = I\mu$. In words, we could say v_1 measures how much the traders are sensitive to news and v_0 stands for the liquid trades. Hence, the characterization of traded volume which stems from market microstructure takes quite simple form.

3.2 Generalized Autoregressive Score Models

Generalized autoregressive score models can be classified as observation driven time varying parameters models. The time evolution of parameters depends on the score function of predictive likelihood. The score function is such a proper choice that many celebrated econometrics models like GARCH, ACD and ACI models can be derived with appropriate distributions and scaling of the score function. Since this framework belongs to observation driven models family, the likelihood is available in closed form and then the estimation does not require expensive computation. Moreover, the time evolution structure utilizes the whole density of observations thanks to the presence of the score function.

3.2.1 Basic Model

Let y_t be a $N \times 1$ vector of dependent variable, f_t be the time varying parameter vector and θ be vector of static parameters. For simplicity, we assume there is no

exogenous variable. Also, let Ω_t denotes the information set available up to time t. Specifically, it can be defined as, for any $t = 1, ..., N$,

$$\Omega_t \equiv \{y_1, ..., y_{t-1}, f_0, ..., f_{t-1}\}. \tag{3.8}$$

Suppose that $y_t \sim p(\cdot|f_t, \Omega_t; \theta)$ in which $p(\cdot|\cdot)$ is the observation density. The key ingredient of this framework is the updating mechanism of f_t. Assume that it is given by the following autoregressive equation,

$$f_{t+1} = \omega + \sum_{i=1}^{p} A_i s_{t-i+1} + \sum_{j=1}^{q} B_j f_{t-j+1} \tag{3.9}$$

where ω is a vector of constants, A_i, B_j's are the coefficient matrices. We can consider ω, A_i, B_j in static parameter vector θ. For $i = 1, ..., p$, s_{t-i+1} be the score function scaled by a scaling factor. Particularly, it can be expressed as,

$$\begin{aligned} s_t &= S_t \nabla_t \\ \nabla_t &= \frac{\partial \log(p(y_t|f_t, \Omega_t; \theta))}{\partial f_t} \end{aligned} \tag{3.10}$$

where S_t is the scaling matrix and numerator is the predictive local likelihood. The system described above is called as GAS model with order p and q, i.e. GAS(p,q). The scaled score function, s_t, indicates the direction in which local likelihood increases most. Hence, it provides a natural updating mechanism for f_t. So far, we do not specify the choice of scaling factor, S_t. Essentially, it brings an additional flexibility to the GAS framework. In the literature, there are 3 common form used for S_t. First one is the inverse of Fischer information matrix,

$$S_t = I_{t|t-1}^{-1}, \quad I_{t|t-1} = \mathrm{E}\left(\nabla_t \nabla_t' \mid \Omega_{t-1}\right) \tag{3.11}$$

where $\mathrm{E}(\cdot \mid \Omega_{t-1})$ denotes the conditional expectation in terms of observation density $p(y_t \mid f_t, \Omega_t; \theta)$. For this particular choice of scaling matrix, GAS(p, q) model encompasses the GARCH model of [Engle, 1982] and [Bollerslev, 1986], the ACD model of [Engle and Russell, 1998] and also Poisson count model of [Davis et al.,

2003]. Another option for S_t could be the Cholesky factor of inverse information matrix. That is to say,

$$S_t = J_{t|t-1}, \quad J'_{t|t-1} J_{t|t-1} = I^{-1}_{t|t-1}. \tag{3.12}$$

In this specification, s_t have a unit variance and this makes the dynamics of the model more tractable. A third common choice is the identity matrix, $S_t = I$. With this option, GAS(p, q) model can capture, for example, ACM model of [Russell and Engle, 2005]. The parameters of the model can be estimated by maximum likelihood estimation method regardless of the choice of scaling factor and observation density. The details are discussed in [Creal et al., 2013] with special applications.

3.2.2 Parametrization

In various settings, we are often interested in functions of time varying parameters rather than itself. For instance, consider $f_t = \sigma_t^2$ which can be deemed as time varying variance. It implies that $f_t \geq 0$ for any t. An alternative would be describing the time evolution in terms of $\log(f_t)$. This is very useful transformation in the sense that we are no longer worry about the nonnegativity restriction as before. The GAS models can internalize these transformations automatically. In general, suppose that we are interested in $\tilde{f}_t = h(f_t)$ where $h(\cdot)$ is differentiable and invertible function. Let, $(s_t, S_t, \nabla_t, I_{t|t-1})$ and $(\tilde{s}_t, \tilde{S}_t, \tilde{\nabla}_t, \tilde{I}_{t|t-1})$ be variables according to f_t and $\tilde{f}_t$ respectively. Also, the scaling factor assumed to be inverse information matrix. Then,

$$\begin{aligned}
\tilde{s}_t &= \tilde{S}_t \tilde{\nabla}_t \\
&= \tilde{I}^{-1}_{t|t-1} \tilde{\nabla}_t \\
&= E((\dot{h}'_t)^{-1} \nabla_t \nabla'_t (\dot{h}_t)^{-1} \mid \Omega_{t-1})^{-1} (\dot{h}'_t)^{-1} \nabla_t \\
&= \dot{h}_t E(\nabla_t \nabla'_t \mid \Omega_{t-1})^{-1} (\dot{h}'_t)(\dot{h}'_t)^{-1} \nabla_t \\
&= \dot{h}_t I^{-1}_{t|t-1} \nabla_t \\
&= \dot{h}_t s_t
\end{aligned} \tag{3.13}$$

where $\dot{h}_t = \frac{\partial h(f_t)}{\partial f_t}$. Thus, the parametrization of time varying parameter only changes the score function in a tractable way. The resulting updating equation can be stated as,

$$h(f_{t+1}) = \omega + \sum_{i=1}^{p} A_i \dot{h}_{t-i+1} s_{t-i+1} + \sum_{j=1}^{q} B_j h(f_{t-j+1}). \tag{3.14}$$

Chapter 4

MODEL DESCRIPTION

As in Section 3.2, let R_t and V_t be stock return and traded volume at time t respectively. According to the Mixture of Distribution Hypothesis, the conditional distributions of variables could be described as,

$$\begin{aligned} R_t^2|K_t &\sim \Gamma(\tfrac{1}{2}, \tfrac{1}{2\sigma^2 K_t}) \\ V_t|K_t &\sim \text{Poisson}(v_0 + v_1 K_t) \end{aligned} \tag{4.1}$$

where K_t is information arrival rate and $\Gamma(\cdot)$ stands for Gamma distribution. Although Section 3.1 states the distribution of returns instead of squared return, we opt to express the latter. This specification implies that volatility of stock return is driven by information arrivals. Moreover, traded volume is a Poisson count process so that its mean is linearly related to the information arrivals. We suppose that conditional on information arrivals, K_t, traded volume and stock return are independently distributed. Hence, the conditional joint distribution becomes product of individual distributions in Equation 4.1.

So far, we do not have any words on the dynamics of information arrivals. Several considerations might be convincing when introducing a dynamic representation of K_t. First, it takes time to completely reveal itself for any new information. The details of events may be available on next few days. Second, regarding the capabilities of GARCH process for the stock return volatility models, the information arrival rate which drives the return volatility may have similar dynamic structure. Indeed, [Lamoureux and Lastrapes, 1994], [Andersen, 1996], [Darolles et al., 2015] and [Darolles et al., 2017] consider a first order autoregressive stochastic volatility representation for information arrivals. In this study, we formulate the dynamics of K_t in the spirit of Generalized Autoregressive Score models. There are two

main reasons of selecting score driven representation instead of stochastic volatility structure. First, GAS models bring an intuitive and natural updating mechanism to time varying parameters. Indeed, they are forced to move towards a direction in which likelihood increases most. Not surprisingly, this translates into better in-sample fit and superior out-of-sample performance. Second, the formulations of previous studies require expensive computations in which introducing non-linear and/or non-Gaussian dynamics might be quite burdensome. Despite the non-linearities of the model, score models can be estimated directly with likelihood maximization since score models are an observation driven time varying parameters model.

Throughout the paper we assume $\sigma = 1$ due to identification problems. Score driven updating in information arrival is stated as

$$K_t = \omega + \alpha s_{t-1} + \beta K_{t-1} \tag{4.2}$$

where (α, β, ω) are static model parameters and s_t is the scaled score of predictive joint likelihood. The inverse information matrix is selected for scaling factor. The explicit form of s_t is displayed in Equation 4.3 and derived in Appendix A.

$$s_t = \left[\frac{v_0 + v_1 K_t}{v_0 + v_1 K_t + 2v_1^2 K_t^2}\right](R_t^2 - K_t) + \left[\frac{2v_1^2 K_t^2}{v_0 + v_1 K_t + 2v_1^2 K_t^2}\right]\left(\frac{V_t - (v_0 + v_1 K_t)}{v_1}\right) \tag{4.3}$$

Although Equation 4.3 contains nonlinear and puzzling terms, it provides several insights about the dynamics of information arrival rate. First, the second parenthesis of each component can be regarded as mean deviations, put differently prediction errors, for random variables R_t^2 and V_t respectively. That is to say that, the model extract some valuable information about K_t from the most recent prediction errors of random variables of interest. Moreover if we were modelling individual densities rather than joint density of squared return and traded volume, we would obtain the same expressions with the ones in normal parenthesis. Since the fractions in square brackets sum up to one, we can say that the scaled score

for joint density is a weighted average of scaled scores for individual densities. Another significant point is that the weights written in square brackets are determined by the variances of score functions for random variables R_t^2 and V_t. That is, for instance, the first square brackets is the variance of score for squared return density divided by the sum of score function variances for traded volume and squared return densities. As a result, the scaled score implied by our model is the weighted average of prediction errors in which weights are set according to variance of individual score functions. Furthermore, it is shown in Appendix B that $\mathrm{E}(s_t) = 0$ and $\mathrm{Cov}(s_t, s_{t-j}) = 0, \forall t, j$. That means, s_t can be regarded as zero mean and serially uncorrelated innovations.

As noted before, while v_0 could stand for liquid or noisy trades, the parameter v_1 might be interpreted as the sensitiveness of traders to information arrivals. It is noteworthy to look at comparative statistic of s_t with respect to v_0 and v_1. If v_0, i.e. liquid trades, increases then the weight on prediction error of traded volume decreases and consequently the weight on squared return prediction error increases. In case of an increase in v_1, we observe the reverse effects comparing to the former case. That is, an increase in information sensitiveness leads to decrease in the weight on squared return prediction error and increase in traded volume counterpart. More rigorously, we can express these findings as,

$$\lim_{v_0 \to \infty} s_t = R_t^2 - K_t \quad \text{and} \quad \lim_{v_1 \to 0} s_t = R_t^2 - K_t \tag{4.4}$$

which mean that our model boils down to GARCH(1, 1) process in the limit. Put differently, trade volume becomes more valuable for volatility when we have small v_0 (less liquid trades) and/or large v_1 (more sensitive to information arrivals). In this case, we may expect a better predictive performance comparing to models which do not utilize traded volume.

Furthermore, to account for extreme observations we suppose t distribution for returns instead of normal distribution. The resulting scaled score of the model is

formulated in Equation 4.5.

$$s_t = \left[\frac{\tilde{v}(v_0 + v_1 K_{t-1})}{\tilde{v}(v_0 + v_1 K_{t-1}) + 2K_{t-1}^2 v_1^2}\right]\left(\frac{R_{t-1}^2 - K_{t-1}}{\tilde{v} + \frac{R_{t-1}^2}{(v+3)K_{t-1}}}\right) + \left[\frac{2K_{t-1}^2 v_1^2}{\tilde{v}(v_0 + v_1 K_{t-1}) + 2K_{t-1}^2 v_1^2}\right]\left(\frac{V_{t-1} - (v_0 + v_1 K_{t-1})}{v_1}\right) \quad (4.5)$$

where $\tilde{v} = \frac{v}{v+3}$ and v is the degrees of freedom. The details of the derivation can be found in Appendix A. The new updating equation resembles the normal distribution case together with minor differences. The previous arguments on the weighted average of individual scaled scores apply here as well. Therefore, the changes in the parameters v_0 and v_1 have the same effects on the model formation. However, the presence of degrees of freedom parameter in the scaled score function may have interesting implications. As it is expected, when $v \to \infty$ the model boils down to former case. If v is considered as fixed, we observe that t distribution assumption lowers the weight on individual scaled score of return variable and consequently rises the weight on the scaled score of traded volume. In other words, heavier tails of return distribution gives more importance on the prediction error of traded volume. [Giot et al., 2010] show that the positive relation between volume and volatility is driven by the continuous part of the volatility. Moreover, the sign of relation becomes reversed once we consider the jump component of the volatility. In this regard, t distribution enables us to capture the extreme realizations of variance, which might be a jump, through additional parameter v. Parallel to this, our model with t distributed returns establishes a stronger link between volume and volatility as reflected in the form of Equation 4.5.

The financial econometrics literature documents an important stylized fact about stock returns that while positive returns decreases the volatility, negative returns brings high future volatility. This asymmetry is called as leverage effect and it has a couple of explanations. One is that when a stock price declines unexpectedly, the financial leverage ratio raises. Indeed, it translates into high volatility in the future. A detailed discussion can be found in recent paper [Anatolyev and Kobotaev, 2018]. As a result, we consider the leverage effect in the context of scaled

score innovations. [Harvey and Sucarrat, 2014] proposes a specification which encompasses leverage with a score driven model. Following the framework of [Harvey and Sucarrat, 2014], the new updating equation is written in terms of $\log(K_t)$ as

$$\log(K_t) = \omega + \alpha \tilde{s}_{t-1} + \beta \log(K_{t-1}) + \gamma \operatorname{sgn}(R_t)(\tilde{s}_{t-1} + 1) \tag{4.6}$$

where $\operatorname{sgn}(R_t)$ denotes $+1$ for positive returns and -1 for negative returns.

4.1 Statistical Properties

The unconditional second moments implied by the model are similar to [Andersen, 1996]. In that regard, we have the following identities,

$$\begin{aligned} \operatorname{Cov}(R_t^2, V_t) &= \sigma^2 v_1 \operatorname{Var}(K_t) \\ \operatorname{Cov}(V_t, V_{t+1}) &= v_1^2 \operatorname{Cov}(K_t, K_{t+1}) \\ \operatorname{Cov}(R_t^2, R_{t+1}^2) &= \sigma^4 \operatorname{Cov}(K_t, K_{t+1}). \end{aligned} \tag{4.7}$$

First line of Equation 4.7 refers that the contemporaneous correlation between squared return and volume is driven by how traders are sensitive to news coming to the market and the variance of information arrival process. Moreover, the autocovariances of R_t^2 and V_t are resulted from the serial correlation of K_t.
It is empirically well documented that the stock return series display excess kurtosis, i.e. $\text{Kurtosis}(R_t) > 3$. We know that standard GARCH(1, 1) model accounts for that property. In Appendix B we show that the kurtosis of return variable implied by our model is greater than 3.

Chapter 5

EMPIRICAL APPLICATION

In Chapter 4, we provide 3 different dynamic representation for information arrival process, put differently we have 3 dynamic versions of MDH. The benchmark model, Equation 4.1 and 4.2, is called as VE-GARCH model as it resembles a GARCH process by incorporating traded volume. The second specification focuses on the stock return with t distribution. We call this model as VE-t-GARCH. The last model considers the leverage effect together with the t distribution specification. Hence, we name it as VE-t-EGARCH referring EGARCH model of [Nelson, 1991]. These 3 models constitute a broad perspective for financial asset volatility by melting salient features of data and traded volume in the same pot. Thanks to the features of score driven models, we can obtain the likelihood function in closed form. Hence, the estimation of the models does not require expensive computations. For each model, in sample estimations and out of sample predictions are computed. Moreover, 3 standard GARCH models (GARCH, t-GARCH and t-EGARCH) and VA-GARCH model of [Lamoureux and Lastrapes, 1990] with its t distribution extension are examined both in sample and out of sample in order to demonstrate the value of introducing traded volume via score and MDH structure in volatility prediction. Note that GARCH, t-GARCH and t-EGARCH are corresponding GARCH models of volume enhanced models of VE-GARCH, VE-t-GARCH and VE-t-EGARCH respectively. The models studied in this paper are summarized in Table C.1.

5.1 Data

We use IBM, PG and MSFT stock data to analyse volume enhanced models and compare their performance with counterpart GARCH models. The source of the

data is CRSP database. The sample time period is between 04 November 2001 and 31 December 2019 in which the starting point is set according to NBER business cycle dates. The selected sample provides quite rich characteristics so that it includes both turbulent times like Global Financial Crisis and mild periods. The evolution of the return and volume data are displayed in Figures D.1, D.2 and D.3.

Stock return data is in terms of percentage points and dividend corrected. Following [Andersen, 1996], [Darolles et al., 2017], among others, turnover ratio (in percentage) is selected for traded volume variable. Turnover ratio could be defined as the number of shares traded in a day divided by the total number of shares available in the market. The descriptive statistics of the data set is reported in Table C.2, C.3 and C.4 for full sample together with 2 sub-samples. The sub-samples are determined based on Global Financial Crisis period. Not surprisingly, we call the expansion and recession period before crisis as pre sample and the expansion period after the crisis as post sample.

Furthermore, as a justification of our study we consider sample correlation between squared return and traded volume. We report time varying correlations between squared return and traded volume with 3 year rolling window in Figures D.4, D.5 and D.6. For all stocks, we observe that the correlation is not stable around a certain value. The important point is that the sub-sample average is higher in post sample than in pre sample. This may mean that post sample traded volume series can provide more information about volatility.

5.2 In-sample Estimation

Tables C.5, C.6, C.7, C.8, C.9, C.10, C.11, C.12 and C.13 show the in-sample estimation results of competing models for all 3 stocks and all sub-samples. In addition to parameter estimates we report the standard errors and model selection criteria LL, log-likelihood of stock return. Our results imply important findings. First, according to stock return log-likelihood, either VE-t-GARCH or t-EGARCH outperforms remaining models although the values are very close to

each other. For instance, let us consider IBM stock. The best performing model is VE-t-GARCH for full and post sample. But, if we consider pre sample, this becomes t-EGARCH model. Specifically, among the models with Gaussian returns, i.e. VE-GARCH, VA-GARCH, GARCH, the highest log-likelihood is provided by VA-GARCH model for full sample. When heavy tails are accounted for, VE-t-GARCH outperforms VA-t-GARCH and t-GARCH. Put differently, volume enhanced models are the models which benefits from t distribution most. Yet, this finding is similar for other stocks and sub-samples. In addition to the heavy tails, the result does not change when we account for leverage effect in both volume enhanced model and standard GARCH model. Thus, this finding is robust to leverage effect.

[Andersen, 1996] and [Liesenfeld, 1998] state that MDH model together with stochastic volatility representation is not able to produce highly persistent volatility estimates. Nonetheless, volume enhanced models can capture the persistence in return volatility similar to univariate specifications such as GARCH model of [Bollerslev, 1986] [1]. We think that this improvement stems from quite peculiar yet natural updating mechanism of proposed models. In the literature, e.g. [Liesenfeld, 2001] and [Darolles et al., 2017], MDH model is augmented by an additional mixing variable so that it can produce highly persistent volatility estimates. We satisfy this empirical observation by score driven updating rules instead of a second latent variable.

Not surprisingly, for each stock and selected sample, the parameters v_0 and v_1 are estimated significantly. This shows the validity of our specifications and this result is robust across sub-samples and distributional choices. However, the levels of these parameters are not stable for different sub-samples and return distribution. For example, when we consider IBM stock and VE-GARCH model, v_0 and v_1 are estimated as 0.323 and 0.068 for full sample. In pre sample, we observe an increase in liquid trades, v_0, and a decrease in information sensitiveness, v_1. However, when we switch post sample, we estimate v_1 notably higher comparing

[1]Note that, due to the model building, while β stands for volatility persistence for volume enhanced models, standard GARCH models‘ implied persistence is captured by $\beta + \alpha$.

to previous samples. Also, v_0 is estimated significantly lower than other cases. The pre sample period contains Global Financial Crisis days in which the occurrence of jumps in volatility could be higher relative to normal times. Motivated by the fact that traded volume does not relate with volatility jumps, we can expect that the post sample should suggest a higher information sensitiveness since post sample exclude Global Financial Crisis. Indeed, these results are consistent with our observations in time varying sample correlation of different stocks at the beginning of this section. Also, it accompanies the results presented in [Giot et al., 2010]. This level shift between sub-samples is robust to all chosen stocks and different volume enhanced models, i.e. heavy tails and leverage effect. To address the chages in the level of v_1, we illustrate the time evolution of information sensitiveness by estimating VE-GARCH, VE-t-GARCH and VE-t-EGARCH models with IBM data and 5 year rolling estimation window [2]. Figure D.7, D.8 and D.9 show the time evolution of information sensitiveness for VE-GARCH, VE-t-GARCH and VE-t-EGARCH respectively. All 3 figures have similar shape. That is, we see a gradual increase in v_1 before Global Financial Crisis. But after crisis hit the markets, a large drop in v_1 is observed. When the effects of crisis looses its impact, or after 5 years when crisis observations are excluded from estimation sample, again v_1 gradually increases and reaches pre crisis levels.

5.3 Out-of-sample Forecast

As a real time forecasting exercise, we forecast volatility at time $t + h$ with available information between time $t - w + 1$ and t for $h = 1, ..., 10$ and $w = 250$. That is to say that, we compute 1 year rolling window volatility predictions motivated by the level shifts in the model parameters presented in previous section. Since the return volatility is unobserved it is hard to asses predictive power of the models. In this regard, we consider realized volatility as a volatility proxy [3]. We

[2] 5 year window, or 1250 observations, is selected because we want to reduce finite sample bias by taking more observations into the estimation.

[3] Refer to [Patton, 2011] for detailed discussion on the different proxies with various loss functions.

compute 5 minute realized variance for the trade data [4] which is taken from TAQ database. Before we proceed forecast evaluation, a bias correction step is implemented. The presence of the bias may have several explanations. First, the realized volatility is computed from prices formed in 6.5 hours of a day. As a result, we omit the price variation of non trading hours, e.g. overnight volatility. Second, the bias may stem from our imperfect sub-sampling methods. Third, the autocorrelated intraday prices may induce biased estimates of conditional variance [5]. Hence, we scale realized measures with a constant so that sample mean of squared return and realized volatility are equal to each other.

We take the period 11 November 2002 - 31 December 2019 as evaluation sample in which full sample is almost covered. The RMSFE of competing models presented in Table C.14, C.15, C.16. We report the numbers relative to GARCH results to facilitate the comparison. Moreover, Diebold-Mariano test statistics results of best performing model and its strongest rival are displayed via $*$ signs. For all stocks, volume enhanced models, either VE-t-GARCH or VE-t-EGARCH, outperforms remaining models. In forecasting exercise with IBM and MSFT, Diebold-Mariano test shows the difference is statistically significant at 5% level. However, for PG data, we could not reject the null hypothesis of the test in which the lowest RMSFE is given by VE-t-GARCH and its closest rival is t-GARCH. Taking a further step, we evaluate the models‘ forecasts in the post sample period. The results are shown in Tables C.17, C.18 and C.19. Similarly, we obtain statistically significant lowest RMSFE from volume enhanced models for IBM and MSFT data. Nevertheless, the significance is provided after 6 step forecast horizon for PG data. Consistent with the in-sample estimation study, the post sample period offers a better volatility predictions after traded volume is considered in a GARCH environment.

When we look at the RMSFE values by grouping by return distributions, we can see interesting patterns. For example, in the study with IBM data, VA-GARCH model outperforms GARCH and VE-GARCH. However, if we introduce t distri-

[4]The trade and quote data produce similar results, see [Barndorff-Nielsen et al., 2009]

[5]Refer to [Hansen and Lunde, 2005] for a decent discussion.

bution for these models, the best performing model becomes VE-t-GARCH. After introducing leverage effect in addition to the t distirbution, the results does not chenage. This finding is similar for other stock data with exceptions in some forecast horizons.

Chapter 6

CONCLUSION

In this paper, we propose a novel variant of GARCH model which embraces traded volume in a statistically coherent and computationally efficient way by combining MDH and GAS models. Based on this new simple model, we create two different models to account for heavy tails and leverage effect. Competing our models with conventional GARCH models which do not utilize traded volume together with VA-GARCH model that linearly relates traded volume to volatility reveal crucial findings.

Our results emphasize the significance of traded volume for modelling volatility in both in-sample estimations and out-of-sample predictions. Although we obtain humble improvements in model fit in-sample,if any, the out-of-sample forecasting exercise produce significant advancement in terms of RMSFE. In-sample estimation study suggests significant v_0 and v_1 parameters for all stocks and specifications. However, their level differ across sub-samples. A 5 year rolling window estimation of v_1 for VE-GARCH, VE-t-GARCH and VE-t-EGARCH reveals that, v_1 or information sensitiveness drops significantly in Global Financial Crisis. But after and before crisis gradual increase in v_1 is observed. Moreover, contrary to previous studies, volume enhanced models imply volatility estimates as persistent as standard GARCH models‘ estimates. In the out-of-sample prediction exercise, volume enhanced models with t distribution outperform other models and Diebold-Mariano test shows the statistical significance of this result. Also, it is robust to leverage effect. The crucial result is that volume enhanced models‘ achievements become more notable when we specify returns with t distribution. Our results suggest that the score driven updating structure and MDH specification can benefit most from controlling heavy tails(or extreme observations) both in-sample and out-of-sample.

This research may have several implausible sides. First, although we consider the informed volume as time varying with a particular dynamic structure, we do not pay attention to the dynamics of liquid trades. A future research can address this point in the context of score driven updates. Second, the conditional Poisson assumption leads to suppose the equality of conditional mean and variance. [Andersen, 1996] solves this problem by introduing a sscaling parameter. Unfortunately, this is not identifiable in the context of maximum likelihood estimation. Instead, to get rid of this presumption, more general specifications can be used such as Negative Binomial distribution. Last but not the least, the information sensitiveness can be considered as a time varying parameter rather than fixed. This extension may improve volatility predictions considerably. But estimation procedures might require expensive computations.

BIBLIOGRAPHY

[Anatolyev and Kobotaev, 2018] Anatolyev, S. and Kobotaev, N. (2018). Modeling and forecasting realized covariance matrices with accounting for leverage. *Econometric Reviews*, 37(2):114–139.

[Andersen, 1996] Andersen, T. G. (1996). Return volatility and trading volume: An information flow interpretation of stochastic volatility. *The Journal of Finance*, 51(1):169–204.

[Barndorff-Nielsen et al., 2009] Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2009). Realized kernels in practice: Trades and quotes.

[Blasques et al., 2014] Blasques, F., Koopman, S. J., and Lucas, A. (2014). Maximum likelihood estimation for score-driven models.

[Blasques et al., 2015] Blasques, F., Koopman, S. J., and Lucas, A. (2015). Information-theoretic optimality of observation-driven time series models for continuous responses. *Biometrika*, 102(2):325–343.

[Blasques et al., 2020] Blasques, F., Lucas, A., and van Vlodrop, A. (2020). Finite sample optimality of score-driven volatility models: Some monte carlo evidence. *Econometrics and Statistics*.

[Bollerslev, 1986] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327.

[Cakmakli and Simsek, 2020] Cakmakli, C. and Simsek, Y. (2020). Bridging the covid-19 data and the epidemiological model using time varying parameter sird model. *arXiv preprint arXiv:2007.02726*.

[Creal et al., 2011] Creal, D., Koopman, S. J., and Lucas, A. (2011). A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. *Journal of Business & Economic Statistics*, 29(4):552–563.

[Creal et al., 2013] Creal, D., Koopman, S. J., and Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, 28(5):777–795.

[Creal et al., 2014] Creal, D., Schwaab, B., Koopman, S. J., and Lucas, A. (2014). Observation-driven mixed-measurement dynamic factor models with an application to credit risk. *Review of Economics and Statistics*, 96(5):898–915.

[Darolles et al., 2015] Darolles, S., Le Fol, G., and Mero, G. (2015). Measuring the liquidity part of volume. *Journal of Banking & Finance*, 50:92–105.

[Darolles et al., 2017] Darolles, S., Le Fol, G., and Mero, G. (2017). Mixture of distribution hypothesis: Analyzing daily liquidity frictions and information flows. *Journal of econometrics*, 201(2):367–383.

[Davis et al., 2003] Davis, R. A., Dunsmuir, W. T., and Streett, S. B. (2003). Observation-driven models for poisson counts. *Biometrika*, 90(4):777–790.

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

[Engle and Russell, 1998] Engle, R. F. and Russell, J. R. (1998). Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica*, pages 1127–1162.

[Gallant et al., 1992] Gallant, A. R., Rossi, P. E., and Tauchen, G. (1992). Stock prices and volume. *The Review of Financial Studies*, 5(2):199–242.

[Giot et al., 2010] Giot, P., Laurent, S., and Petitjean, M. (2010). Trading activity, realized volatility and jumps. *Journal of Empirical Finance*, 17(1):168–175.

[Hansen and Lunde, 2005] Hansen, P. R. and Lunde, A. (2005). A forecast comparison of volatility models: does anything beat a garch (1, 1)? *Journal of applied econometrics*, 20(7):873–889.

[Harvey and Luati, 2014] Harvey, A. and Luati, A. (2014). Filtering with heavy tails. *Journal of the American Statistical Association*, 109(507):1112–1122.

[Harvey and Sucarrat, 2014] Harvey, A. and Sucarrat, G. (2014). Egarch models with fat tails, skewness and leverage. *Computational Statistics & Data Analysis*, 76:320–338.

[Harvey, 2013] Harvey, A. C. (2013). *Dynamic models for volatility and heavy tails: with applications to financial and economic time series*, volume 52. Cambridge University Press.

[Karpoff, 1987] Karpoff, J. M. (1987). The relation between price changes and trading volume: A survey. *Journal of Financial and quantitative Analysis*, 22(1):109–126.

[Koopman et al., 2016] Koopman, S. J., Lucas, A., and Scharth, M. (2016). Predicting time-varying parameters with parameter-driven and observation-driven models. *Review of Economics and Statistics*, 98(1):97–110.

[Lamoureux and Lastrapes, 1990] Lamoureux, C. G. and Lastrapes, W. D. (1990). Heteroskedasticity in stock return data: Volume versus garch effects. *The journal of finance*, 45(1):221–229.

[Lamoureux and Lastrapes, 1994] Lamoureux, C. G. and Lastrapes, W. D. (1994). Endogenous trading volume and momentum in stock-return volatility. *Journal of Business & Economic Statistics*, 12(2):253–260.

[Li and Wu, 2006] Li, J. and Wu, C. (2006). Daily return volatility, bid-ask spreads, and information flow: Analyzing the information content of volume. *The Journal of Business*, 79(5):2697–2739.

[Liesenfeld, 1998] Liesenfeld, R. (1998). Dynamic bivarsate mixture models: Modeling the behavior of prices and trading volume. *Journal of Business & Economic Statistics*, 16(1):101–109.

[Liesenfeld, 2001] Liesenfeld, R. (2001). A generalized bivariate mixture model for stock price volatility and trading volume. *Journal of Econometrics*, 104(1):141–178.

[Lucas et al., 2014] Lucas, A., Schwaab, B., and Zhang, X. (2014). Conditional euro area sovereign default risk. *Journal of Business & Economic Statistics*, 32(2):271–284.

[Lucas et al., 2017] Lucas, A., Schwaab, B., and Zhang, X. (2017). Modeling financial sector joint tail risk in the euro area. *Journal of Applied Econometrics*, 32(1):171–191.

[Nelson, 1991] Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, pages 347–370.

[Oh and Patton, 2018] Oh, D. H. and Patton, A. J. (2018). Time-varying systemic risk: Evidence from a dynamic copula model of cds spreads. *Journal of Business & Economic Statistics*, 36(2):181–195.

[Opschoor et al., 2018] Opschoor, A., Janus, P., Lucas, A., and Van Dijk, D. (2018). New heavy models for fat-tailed realized covariances and returns. *Journal of Business & Economic Statistics*, 36(4):643–657.

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

[Richardson and Smith, 1994] Richardson, M. and Smith, T. (1994). A direct test of the mixture of distributions hypothesis: Measuring the daily flow of information. *Journal of Financial and Quantitative Analysis*, pages 101–116.

[Russell and Engle, 2005] Russell, J. R. and Engle, R. F. (2005). A discrete-state continuous-time model of financial transactions prices and times: The autoregressive conditional multinomial–autoregressive conditional duration model. *Journal of Business & Economic Statistics*, 23(2):166–180.

[Salvatierra and Patton, 2015] Salvatierra, I. D. L. and Patton, A. J. (2015). Dynamic copula models and high frequency data. *Journal of Empirical Finance*, 30:120–135.

[Tauchen and Pitts, 1983] Tauchen, G. E. and Pitts, M. (1983). The price variability-volume relationship on speculative markets. *Econometrica: Journal of the Econometric Society*, pages 485–505.

[Thiele, 2020] Thiele, S. (2020). Modeling the conditional distribution of financial returns with asymmetric tails. *Journal of Applied Econometrics*, 35(1):46–60.

Appendix A

DERIVATIONS OF UPDATING RULES

In this appendix, we show the derivation of updating equations of VE-GARCH model and its variants in full details.

A.1 Gaussian Returns

Assuming the independence of R_t^2 and V_t conditional on K_t, the joint probability density function takes the form of

$$\begin{aligned} f\left(R_t^2, V_t \mid K_t\right) &= f_1(R_t^2 \mid K_t) f_2\left(V_t \mid K_t\right) \text{ where,} \\ f_1(R_t^2 \mid K_t) &= \frac{1}{\sqrt{2\pi\sigma^2 K_t R_t^2}} \exp\left(-\frac{R_t^2}{2\sigma^2 K_t}\right) \text{ and} \\ f_2\left(V_t \mid K_t\right) &= \frac{\exp(-(v_0+v_1 K_t))}{V_t!}(v_0+v_1 K_t)^{V_t} \end{aligned} \tag{A.1}$$

Thus, the joint pdf can be written explicitly as follows,

$$f\left(R_t^2, V_t \mid K_t\right) = \exp\left(-\frac{R_t^2}{2K_t\sigma^2} - (v_0+v_1K_t)\right)\left(\frac{(v_0+v_1K_t)^{V_t}}{V_t!\sqrt{2\pi\sigma^2 K_t R_t^2}}\right). \tag{A.2}$$

Accordingly, the conditional log-likelihood function at time t could be expressed up to some constant. Set $\sigma = 1$ and we have,

$$\log\left(f\left(R_t^2, V_t \mid K_t\right)\right) \propto -\frac{R_t^2}{2K_t} - (v_0+v_1K_t) + V_t \log(v_0+v_1K_t) - \frac{1}{2}\log(K_t R_t^2) \tag{A.3}$$

The score function of the log-likelihood is defined as $\nabla_t = \frac{\partial \log(f(R_t^2, V_t|K_t))}{\partial K_t}$. In explicit terms, it takes the form of

$$\nabla_t = \frac{R_t^2 - K_t}{2K_t^2} + v_1\left(\frac{V_t - (v_0+v_1K_t)}{v_0+v_1K_t}\right). \tag{A.4}$$

We know that $\text{Var}(R_t^2|K_t) = 2K_t^2$ and $\text{Var}(V_t|K_t) = v_0 + v_1 K_t$. Therefore, the score function is the weighted sum of mean deviations divided by variances of squared return and traded volume random variables. The score function is assumed to be scaled with its variance, i.e inverse Fisher information matrix. Note that, $\text{E}[\nabla_t|K_t] = 0$ hence $\text{Var}(\nabla_t|K_t) = \text{E}[\nabla_t^2|K_t]$. From the independence assumption, we can write,

$$\text{E}[\nabla_t^2|K_t] = \frac{\text{E}[(R_t^2 - K_t)^2]}{4K_t^4} + v_1^2 \frac{\text{E}[(V_t - (v_0 + v_1 K_t))^2]}{(v_0 + v_1 K_t)^2} \tag{A.5}$$

where numerators could be interpreted as the conditional variances of R_t^2 and V_t. This implies that,

$$\text{Var}(\nabla_t|K_t) = \frac{1}{2K_t^2} + \frac{v_1^2}{v_0 + v_1 K_t} \tag{A.6}$$

where the first term is the contribution of the squared returns to the variance of the score function and the second one is the part of the variance comes from traded volume. We describe the scaled score as $s_t = \frac{\nabla_t}{\text{Var}(\nabla_t|K_t)}$ which takes the explicit form of equation A.7.

$$s_t = \left(\frac{v_0 + v_1 K_t}{v_0 + v_1 K_t + 2v_1^2 K_t^2}\right)(R_t^2 - K_t) + \left(\frac{2v_1^2 K_t^2}{v_0 + v_1 K_t + 2v_1^2 K_t^2}\right)\left(\frac{V_t - (v_0 + v_1 K_t)}{v_1}\right) \tag{A.7}$$

Indeed, the updating equation of VE-GARCH model could be expressed as follows:

$$\begin{aligned} K_t = \omega + \beta K_{t-1} + \alpha \left(\frac{v_0 + v_1 K_{t-1}}{v_0 + v_1 K_{t-1} + 2v_1^2 K_{t-1}^2}\right)(R_{t-1}^2 - K_{t-1}) + \\ \alpha \left(\frac{2v_1^2 K_{t-1}^2}{v_0 + v_1 K_{t-1} + 2v_1^2 K_{t-1}^2}\right)\left(\frac{V_{t-1} - (v_0 + v_1 K_{t-1})}{v_1}\right) \end{aligned} \tag{A.8}$$

Now, let us consider the form of the updating equation in case of log transformation of latent process, $\log(K_t)$. We name the model with log transformation as VE-EGARCH which refers to standard exponential GARCH model of [Nelson,

1991]. It is easy to see that $\tilde{s}_t = \frac{s_t}{K_t}$ where $\tilde{s}_t$ is the scaled score with respect to $\log(K_t)$ [1]. Hence, VE-EGARCH model has the updating equation as below.

$$\log(K_t) = \omega + \beta \log(K_{t-1}) + \alpha \left(\frac{v_0 + v_1 K_{t-1}}{v_0 + v_1 K_{t-1} + 2v_1^2 K_{t-1}^2} \right) \left(\frac{R_{t-1}^2 - K_{t-1}}{K_{t-1}} \right) + \alpha \left(\frac{2v_1^2 K_{t-1}^2}{v_0 + v_1 K_{t-1} + 2v_1^2 K_{t-1}^2} \right) \left(\frac{V_{t-1} - (v_0 + v_1 K_{t-1})}{v_1 K_{t-1}} \right) \quad \text{(A.9)}$$

Since the leverage play a key role in [Nelson, 1991], we let the leverage appear in our model with log transformation. We follow the methodology of [Harvey and Sucarrat, 2014]. The resulting dynamic equation is

$$\log(K_t) = \omega + \beta \log(K_{t-1}) + \alpha \tilde{s}_{t-1} + \gamma \operatorname{sgn}(R_t)(\tilde{s}_{t-1} + 1) \quad \text{(A.10)}$$

where $\operatorname{sgn}(R_t)$ is the sign function which take +1 or -1 for negative and positive returns respectively. The parameter γ refers to the leverage parameter.

A.2 t Distribution Case

As a further step, we can consider the model when returns are not normal. It is known that the empirical distribution of an asset return has heavier tails comparing to the normal distribution. Now, assume that $R_t \sim t(0, \sigma^2 K_t, v)$ in which v stands for degrees of freedom. Its mean and variance are 0 and $\frac{v}{v-2}\sigma^2 K_t$ respectively. Assuming $\sigma = 1$, the probability density function of R_t^2 is

$$f_1(R_t^2 | K_t) = \mathrm{B}\left(\frac{1}{2}, \frac{v}{2}\right)^{-1} \frac{1}{\sqrt{v K_t R_t^2}} \left(1 + \frac{R_t^2}{v K_t}\right)^{-\left(\frac{v+1}{2}\right)}$$

where $B(\cdot)$ denotes the Beta function. The density function for V_t described in equation A.1 also applies here. Hence, the joint conditional probability density function can be formed as a product of $f_1(\cdot)$ and $f_2(\cdot)$ by exploiting the conditional independence assumption. The joint log-likelihood function at time t is written by retaining the kernel as

[1] For details refer to [Creal et al., 2013].

$$\log(f(R_t^2, V_t|K_t)) \propto -\frac{1}{2}\log(vK_tR_t^2) - \left(\frac{v+1}{2}\right)\log\left(1+\frac{R_t^2}{vK_t}\right) - (v_0+v_1K_t) + V_t\log(v_0+v_1K_t). \tag{A.11}$$

We derive the scaled score by following the same steps with appendix A.1. When we assume t distribution the score of log-likelihood slightly differs from the previous normal case. Apparently, the latter case converges to former one as $v \to \infty$.

$$\nabla_t = \frac{R_t^2 - K_t}{2K_t(K_t + \frac{R_t^2}{v})} + v_1\left(\frac{V_t - (v_0 + v_1K_t)}{v_0 + v_1K_t}\right) \tag{A.12}$$

Define $Z_t = \frac{R_t^2}{R_t^2 + vK_t}$. Note that the random variable $Z_t \sim \text{Beta}\left(\frac{1}{2}, \frac{v}{2}\right)$ where Beta$(\cdot)$ stands for Beta distribution. The variance of the score turns out to be,

$$\text{Var}(\nabla_t|K_t) = \frac{\text{E}[(Z_t(v+1)-1)^2|K_t]}{4K_t^2} + \frac{\text{E}[(V_t - (v_0+v_1K_t))^2|K_t]}{v_0+v_1K_t} \tag{A.13}$$

where numerators could be replaced by the variances of $Z_t(v+1)$ and V_t conditional on K_t. Hence, we formulate $\text{Var}(\nabla_t|K_t)$ explicitly as follows.

$$\text{Var}(\nabla_t|K_t) = \left(\frac{v}{v+3}\right)\frac{1}{2K_t^2} + \frac{v_1^2}{v_0+v_1K_t} \tag{A.14}$$

When the model accounts for t distribution the scaled score function and the updating equation take similar analytic expressions. Nonetheless, we have slight modifications in the part of the score which comes from squared return series. For instance, the score component is driven by the mean deviations as before. Whereas the weights of the mean deviations in scaled score function differ from the benchmark model. We name the model which embraces this new dynamic updating equation as VE-t-GARCH. It is clear that as $v \to \infty$, the return variable becomes normal and as a result the updating equation converges to the one which

derived for VE-GARCH. The resulting dynamics equation can be displayed

$$K_t = \omega + \beta K_{t-1} + \alpha \left(\frac{\tilde{v}(v_0 + v_1 K_{t-1})}{\tilde{v}(v_0 + v_1 K_{t-1}) + 2K_{t-1}^2 v_1^2} \right) \left(\frac{R_{t-1}^2 - K_{t-1}}{\tilde{v} + \frac{R_{t-1}^2}{(v+3)K_{t-1}}} \right) + \left(\frac{2K_{t-1}^2 v_1^2}{\tilde{v}(v_0 + v_1 K_{t-1}) + 2K_{t-1}^2 v_1^2} \right) \left(\frac{V_{t-1} - (v_0 + v_1 K_{t-1})}{v_1} \right). \quad \text{(A.15)}$$

where $\tilde{v} = \frac{v}{v+3}$ As in appendix A.1, the log transformation of information arrival rate and leverage could be further steps. Following the same methodology, the dynamic part of the model can be written in terms of time varying parameter $\log(K_t)$.

$$\log(K_t) = \omega + \beta \log(K_{t-1}) + \alpha \left(\frac{\tilde{v}(v_0 + v_1 K_{t-1})}{\tilde{v}(v_0 + v_1 K_{t-1}) + 2K_{t-1}^2 v_1^2} \right) \left(\frac{R_{t-1}^2 - K_{t-1}}{\tilde{v} K_{t-1} + \frac{R_{t-1}^2}{v+3}} \right) + \left(\frac{2K_{t-1}^2 v_1^2}{\tilde{v}(v_0 + v_1 K_{t-1}) + 2K_{t-1}^2 v_1^2} \right) \left(\frac{V_{t-1} - (v_0 + v_1 K_{t-1})}{v_1 K_{t-1}} \right) \quad \text{(A.16)}$$

We name the model with this dynamic updating rule as VE-t-EGARCH. In this regard, we introduce the leverage as in [Harvey and Sucarrat, 2014] and it has exactly the same form with former cases.

$$\log(K_t) = \omega + \beta \log(K_{t-1}) + \alpha \bar{s}_{t-1} + \gamma \operatorname{sgn}(R_t)(\bar{s}_{t-1} + 1) \quad \text{(A.17)}$$

where $\bar{s}_t$ is the scaled score of the model VE-t-EGARCH.

Appendix B

STATISTICAL PROPERTIES

In this appendix, we provide the derivation of kurtosis implied by VE-GARCH model. The kurtosis of R_t can be formulated as

$$Kurtosis(R_t) = \frac{\mathrm{E}(R_t^4)}{\mathrm{E}(\mathrm{R_t^{'}})^2}. \tag{B.1}$$

Since scaled score, s_t, has 0 conditional mean, we have $\mathrm{E}(R_t^2) = \frac{\omega}{1-\beta}$. For $\mathrm{E}(R_t^4)$ we have, by normality of returns,

$$\begin{aligned} \mathrm{E}(\mathrm{R_t^4}|\mathrm{K_t}) &= 3(\mathrm{E}(R_t^2|K_t))^2 \\ &= 3(\omega + \beta K_{t-1} + \alpha s_{t-1})^2 \\ &= 3(\omega^2 + \beta^2 K_{t-1}^2 + \alpha^2 s_{t-1}^2 + 2\omega\beta K_{t-1} + 2\omega\alpha s_{t-1} + 2\beta\alpha K_{t-1} s_{t-1}) \end{aligned} \tag{B.2}$$

Note that $\mathrm{E}(K_{t-1}s_{t-1}|K_{t-1})$ and so $\mathrm{E}(K_{t-1}s_{t-1})$. Also, by normality, $\mathrm{E}(R_t^4) = 3\,\mathrm{E}(K_t^2)$. These imply that,

$$\mathrm{E}(R_t^4) = \frac{3}{1-\beta^2}(\omega^2 + \alpha^2\,\mathrm{E}(s_{t-1}^2) + \frac{2\omega^2\beta}{1-\beta}) \tag{B.3}$$

and after some algebra,

$$Kurtosis(R_t) = 3(1 + \frac{\alpha^2(1-\beta)}{\omega^2(1+\beta)}\,\mathrm{E}(s_{t-1}^2)). \tag{B.4}$$

Clearly, $Kurtosis(R_t) > 3$, whenever $\beta < 1$.

Appendix C

TABLES

Table C.1: The models studied in this paper

Model Name	Source/Explanation
VE-GARCH	Proposed in this paper
VE-t-GARCH	VE-GARCH with t distribution(heavy tails)
VE-t-EGARCH	VE-t-GARCH with leverage effect
VA-GARCH	[Lamoureux and Lastrapes, 1990]
VA-t-GARCH	VA-GARCH with t distribution(heavy tails)
GARCH	[Bollerslev, 1986]
t-GARCH	GARCH with t distribution(heavy tails)
t-EGARCH	[Nelson, 1991] with t distribution(heavy tails)

Table C.2: Summary statistics of IBM data

	Stock Return (R_t)			Traded Volume (V_t)		
	Full	Pre	Post	Full	Pre	Post
Mean	0.000	-0.002	0.001	0.467	0.485	0.453
Std	1.445	1.697	1.233	0.240	0.243	0.238
Skewness	0.079	0.337	-0.406	2.976	2.003	3.745
Kurtosis	10.056	9.087	9.193	18.935	9.731	26.719
Min	-10.133	-10.133	-8.294	0.082	0.082	0.126
Max	11.500	11.500	8.850	3.272	2.391	3.272

Table C.3: Summary statistics of PG data

	Stock Return (R_t)			Traded Volume (V_t)		
	Full	Pre	Post	Full	Pre	Post
Mean	0.000	-0.006	0.004	0.345	0.327	0.358
Std	1.086	1.246	0.956	0.201	0.182	0.212
Skewness	0.022	-0.030	0.112	6.179	3.453	7.366
Kurtosis	10.087	9.714	8.793	82.855	25.532	102.647
Min	-7.929	-7.929	-6.599	0.040	0.040	0.081
Max	10.182	10.182	8.766	4.635	2.406	4.635

Table C.4: Summary statistics of MSFT data

	Stock Return (R_t)			Traded Volume (V_t)		
	Full	Pre	Post	Full	Pre	Post
Mean	0.000	-0.049	0.035	0.580	0.699	0.495
Std	1.700	2.008	1.439	0.308	0.316	0.273
Skewness	0.309	0.480	0.011	3.162	3.951	3.089
Kurtosis	11.957	11.383	9.221	30.892	44.699	23.382
Min	-11.763	-11.763	-11.450	0.096	0.110	0.096
Max	18.554	18.554	10.402	5.812	5.812	3.790

Table C.5: Full sample estimation of competing models with IBM stock data

	α	β	ω	v_0	v_1	γ	v	LL
VE-GARCH	0.155	0.923	0.164	0.323	0.068			-7676.32
	(0.027)	(0.014)	(0.023)	(0.027)	(0.016)			
VE-t-GARCH	0.101	0.972	0.032	0.351	0.096		4.667	-7309.74
	(0.012)	(0.005)	(0.005)	(0.013)	(0.014)		(0.313)	
VE-t-EGARCH	0.073	0.976	-0.003	0.364	0.085	-0.034	4.536	**-7300.29**
	(0.014)	(0.008)	(0.001)	(0.036)	(0.035)	(0.012)	(0.318)	
VA-GARCH	0.215	0.476	0.000		1.331			-7603.23
	(0.057)	(0.197)	(0.000)		(0.694)			
VA-t-GARCH	0.040	0.900	0.004		0.060		4.507	-7340.88
	(0.006)	(0.013)	(0.002)		(0.016)		(0.297)	
GARCH	0.110	0.834	0.113					-7680.68
	(0.006)	(0.008)	(0.007)					
t-GARCH	0.068	0.918	0.032				4.426	-7343.62
	(0.008)	(0.010)	(0.008)				(0.266)	
t-EGARCH	0.138	0.986	0.005			-0.060	4.592	-7301.96
	(0.015)	(0.003)	(0.002)			(0.010)	(0.274)	

Table C.6: Full sample estimation of competing models with PG stock data

	α	β	ω	v_0	v_1	γ	v	LL
VE-GARCH	0.120	0.937	0.074	0.238	0.091			-6443.79
	(0.020)	(0.014)	(0.014)	(0.024)	(0.025)			
VE-t-GARCH	0.096	0.964	0.024	0.237	0.156		4.791	-6197.36
	(0.007)	(0.006)	(0.004)	(0.030)	(0.059)		(0.414)	
VE-t-EGARCH	0.060	0.978	-0.013	0.257	0.125	-0.040	4.864	-6172.81
	(0.011)	(0.007)	(0.005)	(0.017)	(0.028)	(0.008)	(0.319)	
VA-GARCH	0.107	0.822	0.045		0.104			-6438.34
	(0.057)	(0.072)	(0.065)		(0.122)			
VA-t-GARCH	0.054	0.869	0.019		0.021		4.612	-6204.69
	(0.014)	(0.024)	(0.005)		(0.007)		(0.333)	
GARCH	0.102	0.843	0.063					-6442.31
	(0.006)	(0.010)	(0.005)					
t-GARCH	0.094	0.874	0.041				4.596	-6205.53
	(0.012)	(0.015)	(0.008)				(0.320)	
t-EGARCH	0.132	0.981	-0.002			-0.086	4.816	**-6166.81**
	(0.015)	(0.004)	(0.002)			(0.011)	(0.336)	

Table C.7: Full sample estimation of competing models with MSFT stock data

	α	β	ω	v_0	v_1	γ	v	LL
VE-GARCH	0.112	0.951	0.142	0.422	0.054			-8412.42
	(0.049)	(0.028)	(0.088)	(0.034)	(0.012)			
VE-t-GARCH	0.087	0.985	0.026	0.440	0.084		4.622	-8000.11
	(0.013)	(0.006)	(0.009)	(0.022)	(0.014)		(0.343)	
VE-t-EGARCH	0.076	0.981	0.005	0.449	0.079	-0.027	4.577	**-7994.88**
	(0.011)	(0.006)	(0.003)	(0.021)	(0.013)	(0.006)	(0.342)	
VA-GARCH	0.183	0.622	0.000		1.000			-8407.45
	(0.007)	(0.033)	(0.000)		(0.090)			
VA-t-GARCH	0.036	0.923	0.008		0.019		4.316	-8031.2
	(0.004)	(0.008)	(0.001)		(0.005)		(0.223)	
GARCH	0.054	0.926	0.055					-8408.66
	(0.003)	(0.004)	(0.004)					
t-GARCH	0.071	0.922	0.031				4.342	-8033.68
	(0.008)	(0.008)	(0.008)				(0.254)	
t-EGARCH	0.142	0.988	0.009			-0.047	4.441	-8002.04
	(0.014)	(0.003)	(0.003)			(0.010)	(0.256)	

Table C.8: Pre sample estimation of competing models with IBM stock data

	α	β	ω	v_0	v_1	γ	v	LL
VE-GARCH	0.124	0.980	0.062	0.339	0.051			-3374.75
	(0.035)	(0.013)	(0.029)	(0.017)	(0.009)			
VE-t-GARCH	0.081	0.991	0.017	0.343	0.075		6.078	-3297.26
	(0.011)	(0.005)	(0.005)	(0.022)	(0.015)		(0.809)	
VE-t-EGARCH	0.040	0.994	0.002	0.353	0.068	-0.031	6.497	-3281.55
	(0.012)	(0.002)	(0.001)	(0.014)	(0.009)	(0.006)	(0.899)	
VA-GARCH	0.085	0.893	0.000		0.129			-3372.01
	(0.024)	(0.031)	(0.000)		(0.057)			
VA-t-GARCH	0.034	0.943	0.002		0.021		6.126	-3299.72
	(0.005)	(0.007)	(0.001)		(0.005)		(0.814)	
GARCH	0.067	0.924	0.026					-3373.95
	(0.006)	(0.006)	(0.003)					
t-GARCH	0.052	0.944	0.013				6.100	-3300.05
	(0.009)	(0.009)	(0.006)				(0.704)	
t-EGARCH	0.094	0.994	0.003			-0.062	6.853	**-3278.85**
	(0.018)	(0.002)	(0.002)			(0.011)	(0.898)	

Table C.9: Pre sample estimation of competing models with PG stock data

	α	β	ω	v_0	v_1	γ	v	LL
VE-GARCH	0.077	0.970	0.041	0.219	0.071			-2854.72
	(0.018)	(0.008)	(0.009)	(0.012)	(0.009)			
VE-t-GARCH	0.077	0.985	0.016	0.223	0.100		5.961	-2782.52
	(0.023)	(0.007)	(0.007)	(0.010)	(0.012)		(0.893)	
VE-t-EGARCH	0.050	0.989	-0.002	0.226	0.096	-0.041	6.239	-2769.41
	(0.011)	(0.003)	(0.000)	(0.009)	(0.010)	(0.007)	(0.822)	
VA-GARCH	0.052	0.923	0.014		0.060			-2854.96
	(0.012)	(0.009)	(0.011)		(0.033)			
VA-t-GARCH	0.043	0.920	0.005		0.029		5.622	-2781.15
	(0.013)	(0.015)	(0.004)		(0.013)		(0.770)	
GARCH	0.052	0.929	0.025					-2856.49
	(0.005)	(0.007)	(0.003)					
t-GARCH	0.070	0.920	0.017				5.569	-2781.66
	(0.013)	(0.014)	(0.007)				(0.645)	
t-EGARCH	0.114	0.989	0.000			-0.087	5.815	**-2767.3**
	(0.019)	(0.004)	(0.002)			(0.016)	(0.669)	

Table C.10: Pre sample estimation of competing models with MSFT stock data

	α	β	ω	v_0	v_1	γ	v	LL
VE-GARCH	0.076	0.981	0.071	0.583	0.029			-3718.82
	(0.044)	(0.013)	(0.048)	(0.018)	(0.005)			
VE-t-GARCH	0.075	0.996	0.012	0.594	0.042		4.958	-3532.91
	(0.013)	(0.005)	(0.003)	(0.012)	(0.006)		(0.639)	
VE-t-EGARCH	0.064	0.992	0.005	0.598	0.040	-0.024	4.886	**-3531**
	(0.041)	(0.016)	(0.014)	(0.055)	(0.048)	(0.114)	(0.911)	
VA-GARCH	0.060	0.926	0.053		0.000			-3718.36
	(0.001)	(0.009)	(0.011)		(0.000)			
VA-t-GARCH	0.031	0.945	0.007		0.000		4.618	-3554.69
	(0.004)	(0.008)	(0.002)		(0.000)		0.563	
GARCH	0.060	0.926	0.053					-3718.36
	(0.005)	(0.007)	(0.007)					
t-GARCH	0.055	0.944	0.012				4.661	-3555.11
	(0.010)	(0.010)	(0.007)				(0.409)	
t-EGARCH	0.110	0.997	0.003			-0.020	4.872	-3538.61
	(0.017)	(0.002)	(0.003)			(0.013)	(0.421)	

Table C.11: Post sample estimation of competing models with IBM stock data

	α	β	ω	v_0	v_1	γ	v	LL
VE-GARCH	0.221	0.673	0.522	0.115	0.216			-4256.92
	(0.038)	(0.139)	(0.217)	(0.117)	(0.075)			
VE-t-GARCH	0.110	0.934	0.053	0.290	0.206		4.138	-3997.51
	(0.014)	(0.017)	(0.013)	(0.024)	(0.036)		(0.358)	
VE-t-EGARCH	0.083	0.928	-0.027	0.328	0.158	-0.045	4.034	**-3988.58**
	(0.012)	(0.018)	(0.008)	(0.027)	(0.038)	(0.010)	(0.320)	
VA-GARCH	0.027	0.000	0.000		3.087			-4139.73
	(0.011)	(0.000)	(0.000)		(0.154)			
VA-t-GARCH	0.026	0.089	0.000		1.560		4.483	-4008.87
	(0.005)	(0.020)	(0.000)		(0.092)		(0.366)	
GARCH	0.097	0.763	0.221					-4278.86
	(0.011)	(0.025)	(0.026)					
t-GARCH	0.091	0.848	0.104				3.864	-4021.95
	(0.018)	(0.029)	(0.027)				(0.282)	
t-EGARCH	0.177	0.939	0.019			-0.090	3.970	-3996.85
	(0.029)	(0.015)	(0.007)			(0.019)	(0.286)	

Table C.12: Post sample estimation of competing models with PG stock data

	α	β	ω	v_0	v_1	γ	v	LL
VE-GARCH	0.178	0.839	0.157	0.159	0.210			-3570.25
	(0.044)	(0.050)	(0.049)	(0.052)	(0.065)			
VE-t-GARCH	0.121	0.920	0.040	0.144	0.429		4.248	-3400.28
	(0.020)	(0.022)	(0.012)	(0.041)	(0.093)		(0.302)	
VE-t-EGARCH	0.092	0.936	-0.051	0.167	0.381	-0.036	4.271	**-3389.38**
	(0.018)	(0.024)	(0.019)	(0.038)	(0.090)	(0.008)	(0.351)	
VA-GARCH	0.161	0.561	0.060		0.594			-3564.27
	(0.003)	(0.020)	(0.004)		(0.024)			
VA-t-GARCH	0.063	0.738	0.035		0.110		4.183	-3409.36
	(0.018)	(0.111)	(0.024)		(0.097)		(0.368)	
GARCH	0.162	0.684	0.152					-3575.13
	(0.014)	(0.030)	(0.019)					
t-GARCH	0.120	0.785	0.099				4.121	-3412.45
	(0.023)	(0.039)	(0.025)				(0.360)	
t-EGARCH	0.159	0.953	-0.009			-0.091	4.325	-3390.69
	(0.027)	(0.013)	(0.005)			(0.018)	(0.384)	

Table C.13: Post sample estimation of competing models with MSFT stock data

	α	β	ω	v_0	v_1	γ	v	LL
VE-GARCH	0.185	0.829	0.368	0.034	0.219			-4661.93
	(0.021)	(0.076)	(0.167)	(0.007)	(0.011)			
VE-t-GARCH	0.124	0.955	0.051	0.099	0.357		4.272	-4455.74
	(0.013)	(0.008)	(0.009)	(0.081)	(0.076)		0.380	
VE-t-EGARCH	0.101	0.947	0.001	0.138	0.322	-0.028	4.271	-4447.43
	(0.047)	(0.019)	(0.001)	(0.156)	(0.177)	(0.027)	(0.905)	
VA-GARCH	0.087	0.000	0.000		3.942			-4607.6
	(0.042)	(0.000)	(0.000)		(0.243)			
VA-t-GARCH	0.053	0.765	0.018		0.300		4.350	-4459.96
	(0.007)	(0.042)	(0.005)		(0.122)		(0.593)	
GARCH	0.130	0.729	0.312					-4673.8
	(0.013)	(0.026)	(0.036)					
t-GARCH	0.080	0.884	0.085				4.182	-4468.9
	(0.014)	(0.020)	(0.024)				(0.327)	
t-EGARCH	0.166	0.953	0.029			-0.094	4.312	**-4444.36**
	(0.025)	(0.011)	(0.008)			(0.018)	(0.326)	

Table C.14: Full sample RMSFE of competing models with IBM data

h	VE-GARCH	VE-t-GARCH	VE-t-EGARCH	VA-GARCH	VA-t-GARCH	GARCH	t-GARCH	t-EGARCH
1	1.020	0.863	**0.848*****	0.902	0.898	1.000	0.909	1.035
2	0.995	0.887	**0.849*****	0.919	0.917	1.000	0.932	1.018
3	0.979	0.880	**0.841*****	0.913	0.909	1.000	0.920	0.992
4	0.973	0.891	**0.848*****	0.928	0.922	1.000	0.927	0.978
5	0.964	0.879	**0.839*****	0.917	0.908	1.000	0.910	0.957
6	0.959	0.883	**0.847*****	0.921	0.910	1.000	0.916	0.953
7	0.946	0.871	**0.838*****	0.906	0.898	1.000	0.902	0.934
8	0.944	0.876	**0.842*****	0.911	0.901	1.000	0.908	0.924
9	0.929	0.857	**0.824*****	0.891	0.881	1.000	0.891	0.896
10	0.922	0.855	**0.823*****	0.889	0.879	1.000	0.892	0.895

Note: The values are reported relative to GARCH column. *,**,*** indicate 10%, 5% and 1% significance levels respectively.

Table C.15: Full sample RMSFE of competing models with PG data

h	VE-GARCH	VE-t-GARCH	VE-t-EGARCH	VA-GARCH	VA-t-GARCH	GARCH	t-GARCH	t-EGARCH
1	1.032	**0.952**	0.954	0.986	0.958	1.000	0.960	1.101
2	0.940	**0.874***	0.883	0.955	0.914	1.000	0.911	1.007
3	0.967	**0.911**	0.933	0.970	0.963	1.000	0.941	1.055
4	0.972	**0.926**	0.941	0.977	0.976	1.000	0.953	1.045
5	0.982	**0.942**	0.961	0.978	0.979	1.000	0.957	1.050
6	0.998	**0.953**	0.973	0.983	0.994	1.000	0.966	1.048
7	1.002	**0.955**	0.986	0.988	0.996	1.000	0.969	1.046
8	1.001	**0.959**	1.001	0.992	1.000	1.000	0.970	1.039
9	1.009	**0.967**	1.035	0.986	0.990	1.000	0.969	1.033
10	1.015	0.977	1.046	0.983	0.985	1.000	**0.971**	1.024

Note: The values are reported relative to GARCH column. *,**,*** indicate 10%, 5% and 1% significance levels respectively.

Table C.16: Full sample RMSFE of competing models with MSFT data

h	VE-GARCH	VE-t-GARCH	VE-t-EGARCH	VA-GARCH	VA-t-GARCH	GARCH	t-GARCH	t-EGARCH
1	0.945	**0.871*****	0.875	0.936	0.958	1.000	0.924	0.996
2	0.950	**0.882*****	0.888	0.960	0.914	1.000	0.925	0.977
3	0.944	**0.880*****	0.884	0.946	0.963	1.000	0.921	0.962
4	0.941	0.879	**0.877****	0.941	0.976	1.000	0.916	0.940
5	0.937	**0.875*****	0.877	0.928	0.979	1.000	0.913	0.945
6	0.929	**0.867****	0.871	0.921	0.994	1.000	0.905	0.906
7	0.918	**0.857****	0.863	0.906	0.996	1.000	0.895	0.898
8	0.909	**0.846****	0.852	0.896	1.000	1.000	0.888	0.916
9	0.908	**0.849****	0.853	0.891	0.990	1.000	0.888	0.911
10	0.895	**0.831*****	0.832	0.874	0.985	1.000	0.874	0.883

Note: The values are reported relative to GARCH column. *,**,*** indicate 10%, 5% and 1% significance levels respectively.

Table C.17: Post sample RMSFE of competing models with IBM data

	VE-GARCH	VE-t-GARCH	VE-t-EGARCH	VA-GARCH	VA-t-GARCH	GARCH	t-GARCH	t-EGARCH
1	1.007	0.811	**0.809****	0.854	0.866	1.000	0.883	1.047
2	0.957	0.819	**0.819****	0.844	0.859	1.000	0.890	1.031
3	0.921	0.819	**0.821****	0.847	0.855	1.000	0.889	1.025
4	0.895	0.812	**0.810*****	0.845	0.849	1.000	0.874	0.998
5	0.878	0.807	**0.801*****	0.841	0.841	1.000	0.866	0.981
6	0.863	0.798	**0.787*****	0.828	0.828	1.000	0.857	0.966
7	0.850	0.791	**0.780*****	0.820	0.820	1.000	0.848	0.953
8	0.839	0.778	**0.769*****	0.808	0.805	1.000	0.838	0.933
9	0.819	0.758	**0.746*****	0.788	0.785	1.000	0.819	0.889
10	0.802	0.737	**0.726*****	0.767	0.764	1.000	0.801	0.875

Note: The values are reported relative to GARCH column. *,**,*** indicate 10%, 5% and 1% significance levels respectively.

Table C.18: Post sample RMSFE of competing models with PG data

h	VE-GARCH	VE-t-GARCH	VE-t-EGARCH	VA-GARCH	VA-t-GARCH	GARCH	t-GARCH	t-EGARCH
1	1.057	0.900	**0.900**	0.997	0.901	1.000	0.920	1.225
2	0.854	0.753	**0.753**	0.902	0.758	1.000	0.798	0.992
3	0.929	**0.844**	0.849	0.930	0.850	1.000	0.869	1.092
4	0.954	**0.874**	0.876	0.953	0.880	1.000	0.898	1.109
5	0.982	0.909	0.911	0.969	**0.898**	1.000	0.927	1.136
6	0.997	0.925	**0.924**	0.988	0.927	1.000	0.953	1.138
7	1.004	0.929	**0.927***	0.993	0.953	1.000	0.952	1.123
8	1.013	0.934	**0.930***	1.001	0.960	1.000	0.963	1.115
9	1.016	0.936	**0.932***	0.995	0.962	1.000	0.968	1.109
10	1.015	0.937	**0.930***	0.995	0.961	1.000	0.971	1.099

Note: The values are reported relative to GARCH column. *,**,*** indicate 10%, 5% and 1% significance levels respectively.

Table C.19: Post sample RMSFE of competing models with MSFT data

h	VE-GARCH	VE-t-GARCH	VE-t-EGARCH	VA-GARCH	VA-t-GARCH	GARCH	t-GARCH	t-EGARCH
1	0.978	**0.900*****	0.904	0.958	0.914	1.000	0.938	1.035
2	0.979	**0.925*****	0.928	0.984	0.938	1.000	0.956	1.042
3	0.977	0.928	**0.927****	0.985	0.944	1.000	0.956	1.043
4	0.989	0.945	**0.939****	0.999	0.961	1.000	0.970	1.048
5	0.976	0.935	**0.930****	0.986	0.946	1.000	0.957	1.040
6	0.978	0.930	**0.928****	0.991	0.948	1.000	0.956	0.968
7	0.975	0.927	**0.924****	0.985	0.946	1.000	0.955	0.963
8	0.982	0.930	**0.928****	0.988	0.949	1.000	0.960	1.054
9	0.978	0.927	**0.921****	0.976	0.937	1.000	0.957	1.031
10	0.975	0.925	**0.917****	0.972	0.936	1.000	0.955	1.017

Note: The values are reported relative to GARCH column. *,**,*** indicate 10%, 5% and 1% significance levels respectively.

Appendix D

FIGURES

Figure D.1: Stock Return and Traded Volume series for IBM

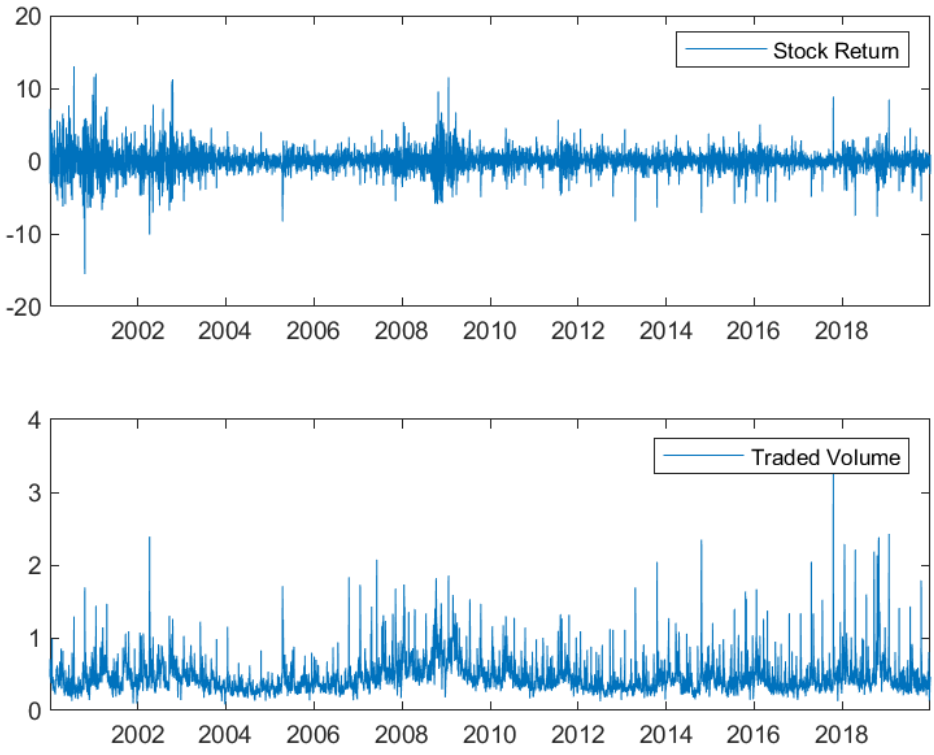


Figure D.2: Stock Return and Traded Volume series for PG

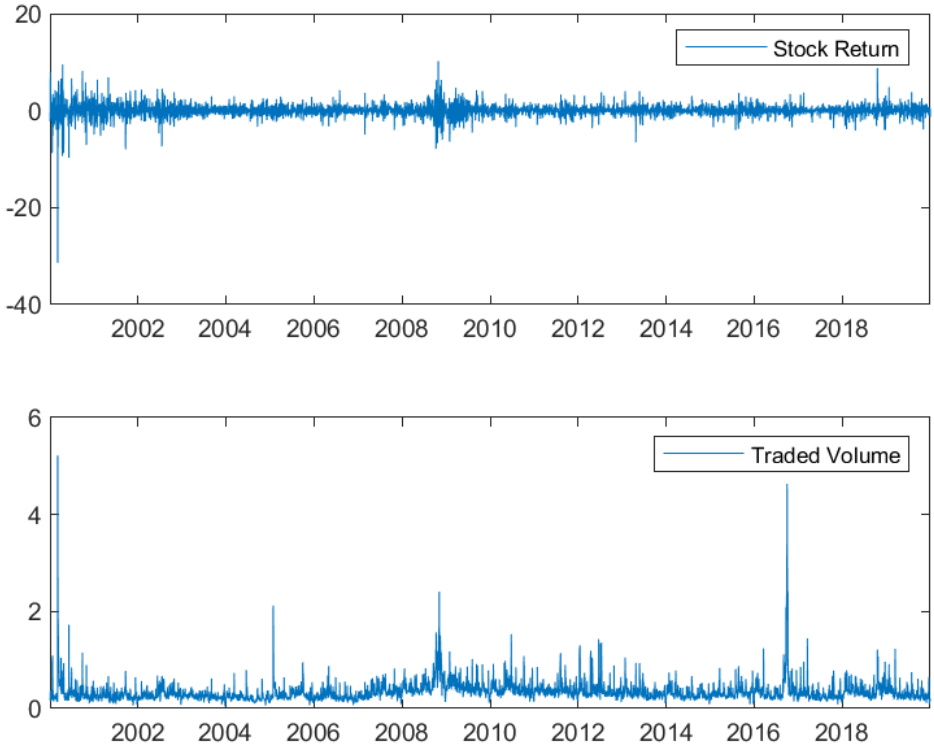


Figure D.3: Stock Return and Traded Volume series for MSFT

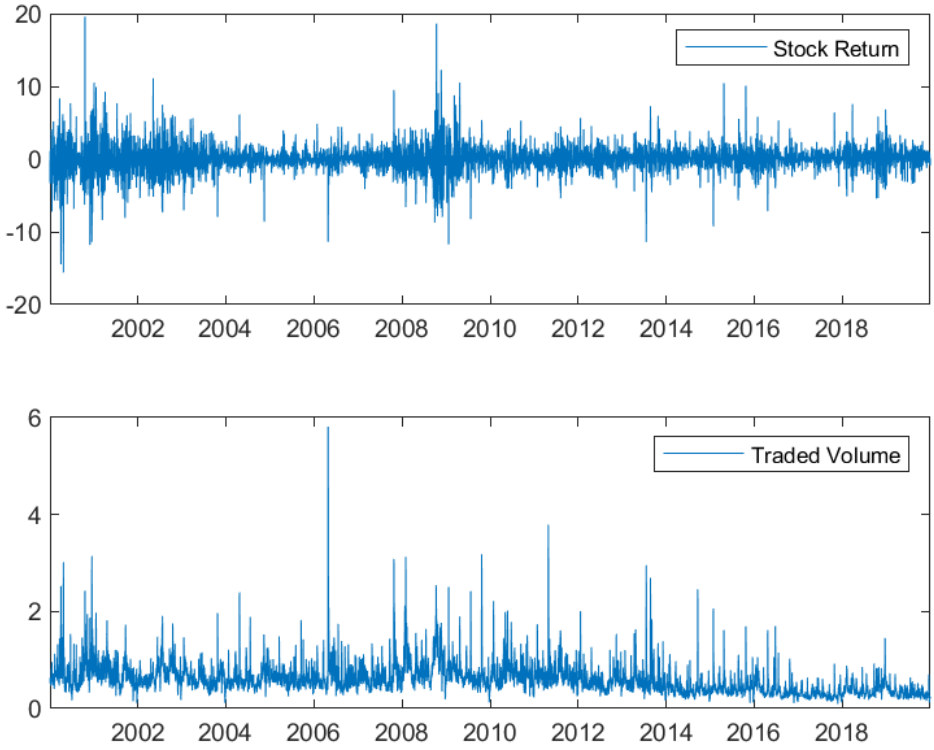
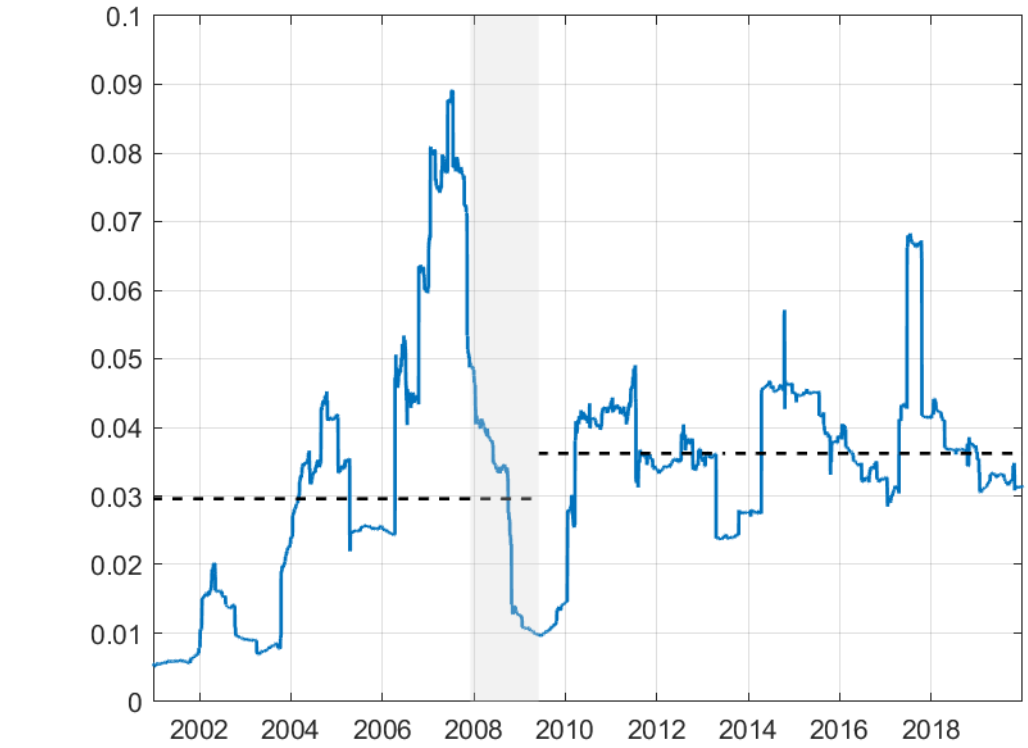
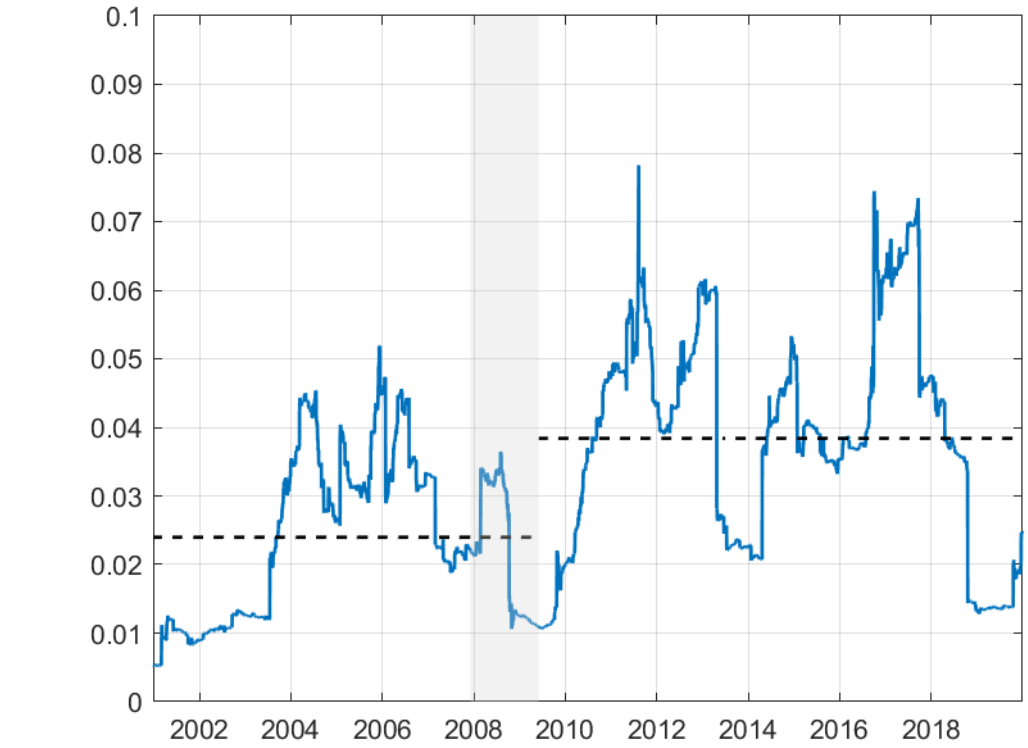


Figure D.4: Time varying regression coefficient of V_t on R_t^2 for IBM (1 years rolling window)



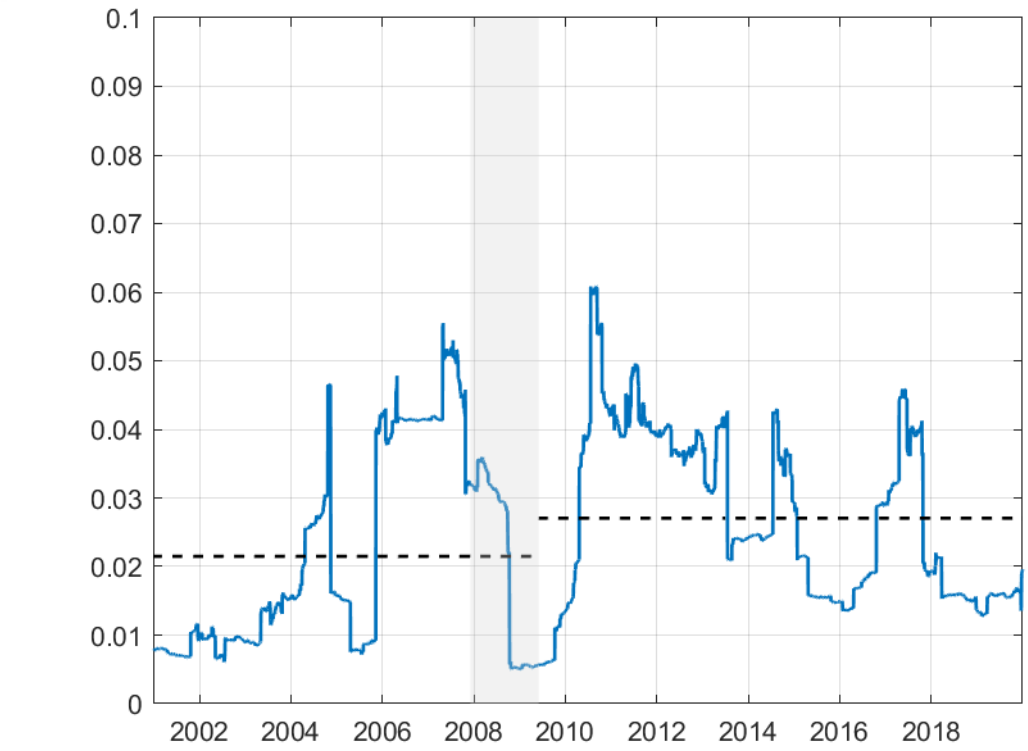
Note: Shaded area indicates Global Financial Crisis. Dashed lines are the average of time varying correlations for pre and post sample respectively.

Figure D.5: Time varying regression coefficient of V_t on R_t^2 for PG (1 years rolling window)



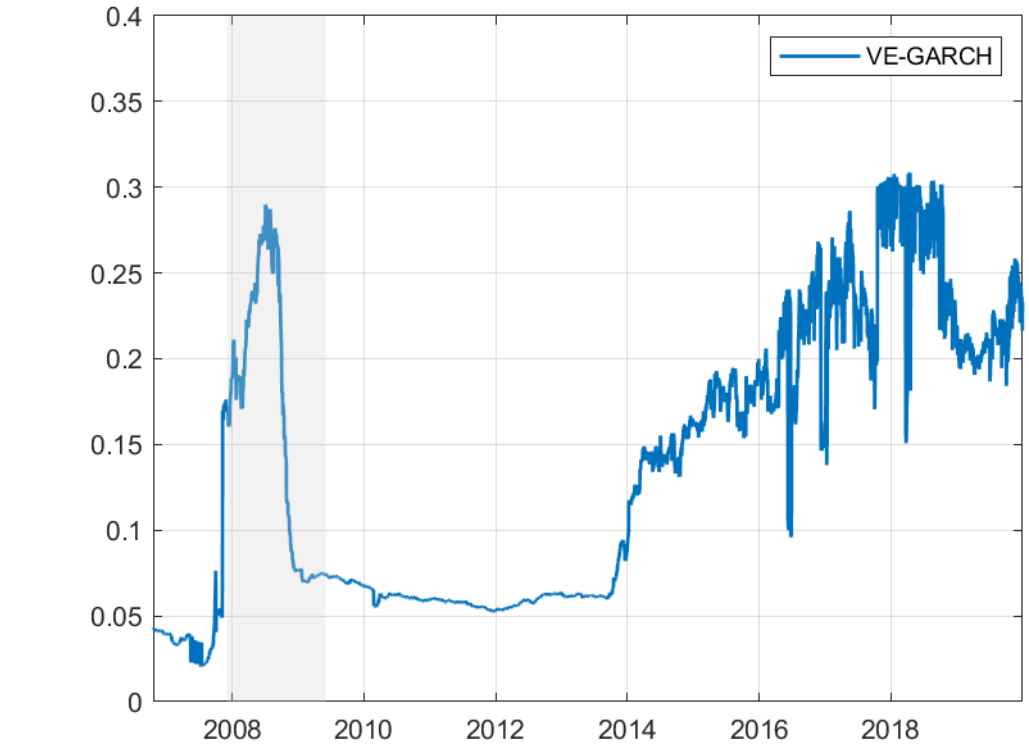
Note: Shaded area indicates Global Financial Crisis. Dashed lines are the average of time varying correlations for pre and post sample respectively.

Figure D.6: Time varying regression coefficient of V_t on R_t^2 for MSFT (1 years rolling window



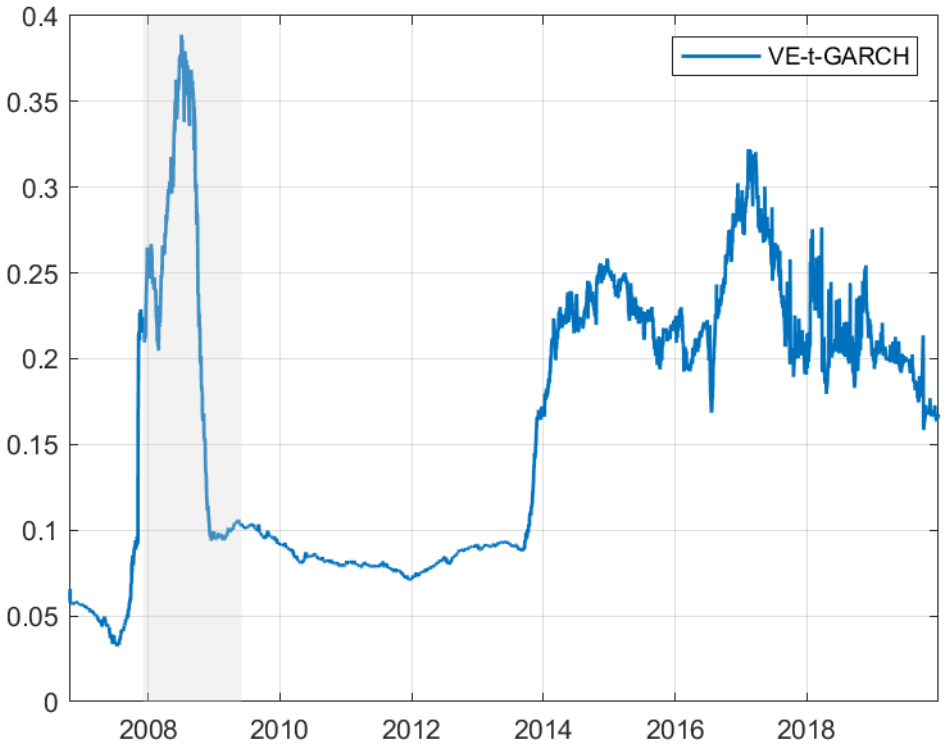
Note: Shaded area indicates Global Financial Crisis. Dashed lines are the average of time varying correlations for pre and post sample respectively.

Figure D.7: Time varying v_1 of VE-GARCH model with IBM data (5 years rolling window)



Note: Shaded area indicates Global Financial Crisis.

Figure D.8: Time varying v_1 of VE-t-GARCH model with IBM data (5 years rolling window)



Note: Shaded area indicates Global Financial Crisis.

Figure D.9: Time varying v_1 of VE-t-EGARCH model with IBM data (5 years rolling window)

Note: Shaded area indicates Global Financial Crisis.