FEB22009(Q/X/S)-23: Introductory Seminar Case Studies
Specialisation: Quantitative Finance & Econometrics
Volatility: I need a VIX 'cause I'm going down.

Rutger-Jan Lange, **Gustavo Freire**, and **Simon Donker van Heel**
Department of Econometrics, Erasmus University Rotterdam

1 Introduction

In 1968, the Beatles sang "I need a fix 'cause I'm going down". Due to the Corona outbreak, the war in Ukraine, and, more recently, a spell of high inflation, financial markets have also been going down. During the covid crisis, in particular, the VIX shot up, reaching levels not seen since the global financial crisis of 2007-2008 (see Figure 1 below). Last year, however, has seen one of the strongest bull markets on record, with low VIX levels.

The 'VIX' is short-hand for the volatility index constructed by the Chicago Board Options Exchange (CBOE). It is constructed from option prices and intends to measure the expectation among investors of the (annualised) standard deviation of the S&P500 return over the next month. For a more detailed explanation of the VIX, click here. The annualisation simply means that the (monthly) standard deviation is multiplied by $\sqrt{12}$ to get an 'annual' figure. Its level of around 80 at the height of the Corona crisis meant that investors were expecting the standard deviation of the next month's return to be $80/\sqrt{12} = 23\%$. Typically, the standard deviation of S&P500 returns, on an *annual* basis, is around 20%. The fact that more volatility was expected in a single month than normally would be expected in an entire year illustrates the magnitude of this crisis.

Typically, when anticipated uncertainty (e.g. the VIX) goes up, the market index (e.g. the S&P500) goes down. Figure 2 displays daily changes in the VIX on the horizontal axis versus S&P500 close-to-close log returns on the vertical axis. Clearly, the index goes down when the VIX goes up, and vice versa. Unfortunately for investors, the relation is contemporaneous, i.e., there is no lead-lag relationship. For any given day, movements of the VIX and the market index are uncertain, but they will probably move in opposite directions.

The VIX is a 'forward-looking' volatility measure, in the sense that it captures investors' expectations regarding the magnitude of changes in the S&P500 in the next month. This is in contrast with 'realised variance' measures, which are based directly on (historical) movements of the index itself. Roughly speaking,

Figure 1: The VIX from 4 January 2000. The data are from from Yahoo Finance.

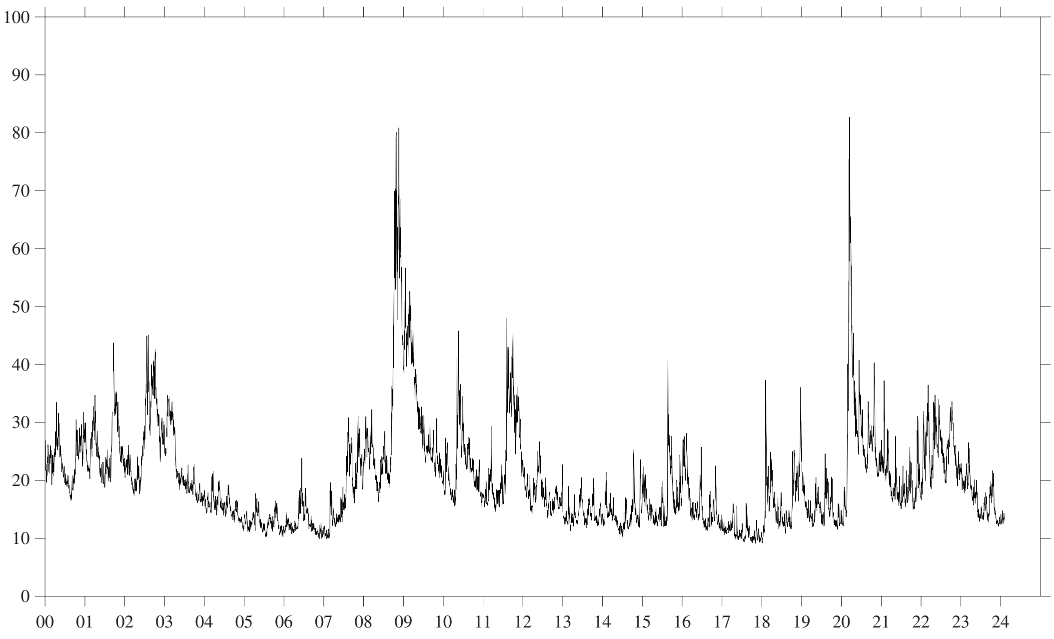
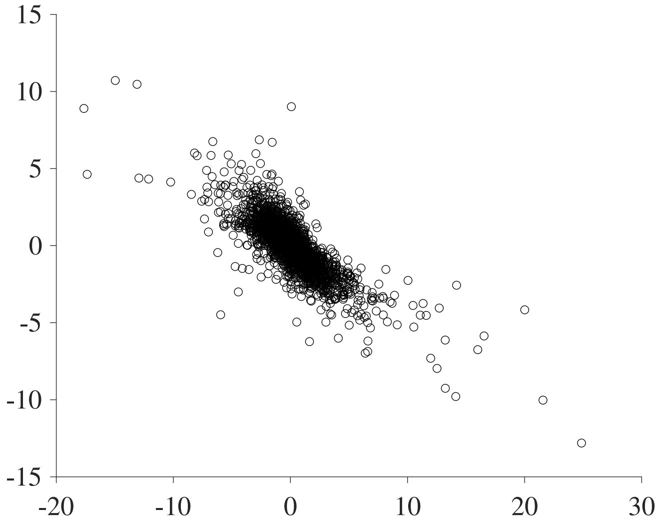


Figure 2: Changes in the VIX from 4 January 2000 (horizontally) versus close-to-close returns on the S&P500 (vertically).



realised variance is a measure of variability obtained by summing the squares of intra-day returns. In this assignment, you will use the variable rv5, which is a realised variance based on summing squares of returns measured on 5-minute intervals. Realised variance is essentially a 'backward-looking' volatility measure in the sense that it can only be computed in hindsight.

Given the crucial role of volatility in financial markets and portfolio management, volatility modelling started as early as 1982 (see Engle, 1982). Researchers have since proposed a variety of models, such as the autoregressive conditional heteroskedasticity (ARCH) and generalized ARCH (GARCH) models of Bollerslev (1986), the latter of which will be used in this assignment. These models account for the fact that volatility is persistent but time-varying. That is, there are prolonged periods of large swings in the market, as well as extensive periods of stability. This phenomenon is known as 'volatility clustering'. While volatility changes over time, it does so gradually, such that periods of high volatility cluster together (see e.g. Figure 1).

To estimate such volatility models, researchers have historically had access to daily returns, such as those available from Yahoo Finance.[1] Nowadays, researchers additionally have access to realised variance measures based on high-frequency data, as well as market expectations of volatility given by the VIX. Given this abundance of new, high-quality data, is it still useful to estimate 'classic' volatility models, such as GARCH? And what about more recent contenders, such as the Beta-t-EGARCH model by Harvey (2013)? A separate strand of the literature proposes to model the realised variance directly, i.e. without a GARCH structure. The heterogeneous autoregressive realised volatility (HAR-RV) of Corsi (2009) is an important example of this approach, and will be used as a benchmark model in this assignment. These considerations lead to the following research questions:

1. Can the volatility of the S&P500 index over a given period be better predicted using classic GARCH models or more recently developed Beta-t-EGARCH models?

2. Can either of these models beat our two benchmark models: market expectations of volatility, as provided by the VIX, and the HAR-RV model of Corsi (2009)?

The (actual) variance of the S&P500 index over some period, which we hope to predict, can be approximated either by summing squared close-to-close log returns, or by summing daily realised variances. As such, we have two possible 'target variables' that we hope to predict, using two advanced models (GARCH and Beta-t-EGARCH) and two simple benchmark models (HAR-RV and the VIX). We are interested in making predictions over a horizon of one trading day, five trading days (corresponding to one calendar week), or 21 trading days (corresponding to one calendar month). To compare the quality of different forecasts, you may consider the robust loss function proposed in equation (24) of Patton (2011), important special cases of which are the mean-squared error (MSE) and so-called 'QLIKE' loss functions.[2] The purpose of these loss functions is to compare our predictions against the target and assign some 'loss value' when they are different. In sum, you will need to compute (at least) 4 models $\times$ 2 targets $\times$ 3 horizons $\times$ 2 loss functions = 48 loss values. It's up to you to report these in a meaningful way and draw economic conclusions.

[1] For example, click here for the S&P500.

[2] For alternative loss functions that could also be considered, see section 3 of Hansen and Lunde (2005).

Creativity is important, so you may also add other models and/or use different performance measures. Essentially, you may deviate from the suggested approach in any way you like; the point in academic work is that you present compelling arguments for whatever approach you choose. Of course the supervisors will judge how compelling they find your approach to be. Deviating from the suggested approach is therefore riskier, being associated with more downside as well as upside. For this reason it may be good to check your own ideas with the supervisors during the feedback sessions.

Data: To answer these research questions, you are provided with an Excel sheet titled `data.xlsx`, which contains six series related to the S&P500 index: (1) daily dates, (2) opening price of S&P500, (3) closing price of S&P500, (4) close-to-close log returns, (5) realised variance "rv5" computed from intra-day returns, and (6) closing level of the VIX. For your convenience, these data have been pre-loaded in the Matlab file `data.mat`.

Demo code: The Matlab file `demo.m` that you have been provided with plots the data, and contains an implementation of a simple GARCH model with Student's t distributed shocks.

Sections 2 though 5 provide background information that should help you understand the models and how to work with them. Section 6 provides a concrete roadmap with step-by-step suggestions for how to answer the research questions.

2 Classic volatility modelling: GARCH

Let the r_t denote the close-to-close log return on the index, that is $r_t = 100 \times \log(y_t/y_{t-1})$, where y_t is the level of the index at the end of day t. These returns have already been constructed for you in the data set provided. Many models for time-varying volatility assume that r_t is generated by a model like this:

$$r_t \;=\; \mu \,+\, \sigma_{t|t-1}\,\varepsilon_t, \qquad 1 \le t \le T, \tag{2.1}$$

where T denotes the sample size, and the $\varepsilon_t's$ are independently and identically distributed (i.i.d.) 'shocks' with $\mathrm{E}[\varepsilon_t] = 0$, $\mathrm{V}[\varepsilon_t] = 1$. People often use the word 'volatility' to mean either $\sigma_{t|t-1}$ or $\sigma^2_{t|t-1}$, depending on the context. If μ and $\sigma_{t|t-1}$ are known, then the variance of returns is $\mathrm{V}[r_t] = \sigma^2_{t|t-1}$. The phenomenon of 'volatility clustering' implies that $\sigma_{t|t-1}$ is time-varying (as indicated by the subscript) but also persistent, which means that $\sigma_{t+1|t} \approx \sigma_{t|t-1}$. In other words, $\sigma_{t|t-1}$ is changing over time, but not too much.

Throughout this assignment, we assume that the shocks ε_t are Student's t distributed with $\nu > 2$ degrees of freedom, such that the probability density function (p.d.f.) of each ε_t reads

$$p(\varepsilon_t) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\nu/2)\sqrt{\pi(\nu-2)}} \left(1 + \frac{\varepsilon_t^2}{\nu-2}\right)^{-(\nu+1)/2}, \qquad \nu > 2. \tag{2.2}$$

This p.d.f. is especially formulated to ensure $\mathrm{V}[\varepsilon_t] = 1$; as such, it may look slightly different to what you are used to. The parameter $\nu > 2$ is *not* related to the number of observations; it need not even be integer. Rather, it's just a (continuous) parameter that you will estimate along with all the other parameters in the model. Typically, it's estimated at $\sim$6.

For the purpose of prediction, our estimate of $\sigma_{t|t-1}$ should be based on the information up to (and including) time $t-1$. That is, we cannot use r_t to predict the distribution of r_t! Instead, we only use the information set $\mathcal{I}_{t-1}$, which is essentially short-hand for the returns contained in the vector $\{r_1, \ldots, r_{t-1}\}$. We write $\sigma_{t|t-1}$ to emphasize that the estimate is made conditional on the econometrician's knowledge at time $t-1$. In other words, $\sigma_{t|t-1}$ is really a one-step-ahead prediction (made at time $t-1$) of the volatility at time t. A rule for constructing $\sigma_{t|t-1}$ is called a 'filter'. Any filter allows us to compute the implied shock, or implied residual, ε_t, as follows:

$$r_t \;=\; \mu \,+\, \sigma_{t|t-1}\,\varepsilon_t, \qquad \Leftrightarrow \qquad \varepsilon_t \;=\; (r_t - \mu)\,/\,\sigma_{t|t-1}. \tag{2.3}$$

The implied residual will differ from the true shock ε_t if the filtered volatility $\sigma_{t|t-1}$ is inaccurate. That is, we must employ a 'reasonably good' filter. The next subsection discusses a particularly famous filter introduced by Bollerslev (1986). Our treatment below covers only the basics; for more detail, you are encouraged to read Chapter 7 of Franses et al. (2014).

2.1 GARCH

The industry workhorse for filtering volatility is the generalized autoregressive conditional heteroskedasticity (GARCH) model of order $(1,1)$ introduced by Bollerslev (1986), which reads

$$\begin{aligned} \sigma^2_{t+1|t} \;&=\; \omega \,+\, \alpha\,(r_t - \mu)^2 \,+\, \beta\,\sigma^2_{t|t-1}, \\ &=\; \omega \,+\, \alpha\,\sigma^2_{t|t-1}\,\varepsilon_t^2 \,+\, \beta\,\sigma^2_{t|t-1}, \end{aligned} \tag{2.4}$$

where the second line follows from (2.3). Here, ω, α, β are parameters. Importantly, everything on the right-hand side is evaluated at time t, such that $\sigma^2_{t+1|t}$ can be computed at time t, as indicated by the conditional bar in the subscript. To guarantee that $\sigma^2_{t|t-1} \geq 0$ for all t, we impose $\omega > 0$, $\alpha \geq 0$ and $\beta \geq 0$. The GARCH equation ensures that volatility changes gradually over time, such that $\sigma^2_{t+1|t}$ will be close to $\sigma^2_{t|t-1}$. Note that the unconditional variance of r_t is given by

$$\mathrm{E}[(r_t-\mu)^2] \;=\; \mathrm{E}[\sigma^2_{t|t-1}\,\varepsilon_t^2] \;=\; \mathrm{E}[\sigma^2_{t|t-1}]\,\mathrm{E}[\varepsilon_t^2] \;=\; \sigma^2,$$

where $\mathrm{E}[\sigma^2_{t|t-1}] = \sigma^2$ denotes the unconditional expectation of $\sigma^2_{t|t-1}$. By taking expectations on both sides of the GARCH equation (2.4), it follows that $\mathrm{E}[\sigma^2_{t+1|t}] = \omega + \alpha\,\mathrm{E}[\sigma^2_{t|t-1}]\,\mathrm{E}[\varepsilon_t^2] + \beta\,\mathrm{E}[\sigma^2_{t|t-1}]$. In the steady state (for large t), we expect $\mathrm{E}[\sigma^2_{t+1|t}] = \mathrm{E}[\sigma^2_{t|t-1}] = \sigma^2$, which would imply

$$\sigma^2 = \frac{\omega}{1-\alpha-\beta},$$

as long as $\alpha+\beta<1$. Using $\omega = (1-\alpha-\beta)\sigma^2$, GARCH equation (2.4) can equivalently be written as

$$\begin{aligned}
\sigma^2_{t+1|t} \;&=\; (1-\alpha-\beta)\,\sigma^2 + \alpha\,\sigma^2_{t|t-1}\,\varepsilon_t^2 + \beta\,\sigma^2_{t|t-1},\\
&=\; \sigma^2 + \alpha\,(\sigma^2_{t|t-1}\,\varepsilon_t^2 - \sigma^2) + \beta\,(\sigma^2_{t|t-1} - \sigma^2),\\
&=\; \sigma^2 + \alpha\,\sigma^2_{t|t-1}(\varepsilon_t^2 - 1) + (\alpha+\beta)\,(\sigma^2_{t|t-1} - \sigma^2),
\end{aligned} \tag{2.5}$$

where we note that the term $\varepsilon_t^2 - 1$ has expectation zero. This simple reformulation demonstrates that the parameter combination $0 \leq \alpha+\beta < 1$ controls how quickly or slowly $\sigma^2_{t|t-1}$ mean-reverts to σ^2, while the parameter $\alpha \geq 0$ controls how sensitive the filter is to each individual observation r_t. It is common to set $\sigma^2_{1|0} = \sigma^2$, i.e. start off at the unconditional mean.

2.2 Asymmetric GARCH

Typically, positive shocks ε_t affect tomorrow's volatility differently than negative shocks. For this purpose, we consider the following asymmetric GARCH model by to Glosten et al. (1993):

$$\begin{aligned}
\sigma^2_{t+1|t} \;&=\; \omega \;+\; (\alpha_{\text{pos}}\,1_{\varepsilon_t\geq 0} \;+\; \alpha_{\text{neg}}\,1_{\varepsilon_t<0})\;\sigma^2_{t|t-1}\,\varepsilon^2_{t|t-1} \;+\; \beta\,\sigma^2_{t|t-1},\\
&=\; \omega + \alpha_{\text{pos}}\left[\varepsilon_t^2\,1_{\varepsilon_t\geq 0} - \frac{1}{2}\right]\sigma^2_{t|t-1} + \alpha_{\text{neg}}\left[\varepsilon_t^2\,1_{\varepsilon_t<0} - \frac{1}{2}\right]\,\sigma^2_{t|t-1} + \left(\beta + \frac{\alpha_{\text{pos}}}{2} + \frac{\alpha_{\text{neg}}}{2}\right)\sigma^2_{t|t-1},\\
&=\; \sigma^2 + \alpha_{\text{pos}}\left[\varepsilon_t^2\,1_{\varepsilon_t\geq 0} - \frac{1}{2}\right]\sigma^2_{t|t-1} + \alpha_{\text{neg}}\left[\varepsilon_t^2\,1_{\varepsilon_t<0} - \frac{1}{2}\right]\,\sigma^2_{t|t-1} + \left(\beta + \frac{\alpha_{\text{pos}}}{2} + \frac{\alpha_{\text{neg}}}{2}\right)(\sigma^2_{t|t-1} - \sigma^2),
\end{aligned} \tag{2.6}$$

where the indicator function 1_A equals one if the event A happens and zero if it does not. Here, the coefficients α_{pos} and α_{neg} describe the sensitivity of volatility with respect to the positive and negative shocks. Note that both terms in square brackets have expectation zero. The last line can be viewed as a definition of σ^2, implying $\sigma^2 := \omega/(1-\beta-\alpha_{\text{pos}}/2-\alpha_{\text{neg}}/2)$. The quantity σ^2 is still the unconditional mean of $\sigma^2_{t|t-1}$ as long as $\beta+\alpha_{\text{pos}}/2+\alpha_{\text{neg}}/2<1$. It is common to set $\sigma_{1|0} = \sigma^2$. As you should verify for yourself, classic (symmetric) GARCH is recovered if $\alpha_{\text{pos}} = \alpha_{\text{neg}} = \alpha$.

2.3 News impact curve

At the end of day t, when the close-to-close log return r_t has been received, we can compute the implied residual ε_t using equation (2.3). Using the asymmetric GARCH equation (2.6), we can compute $\sigma^2_{t+1|t}$, which will generally be different from $\sigma^2_{t|t-1}$. The difference between $\sigma^2_{t+1|t}$ and $\sigma^2_{t|t-1}$ can be viewed as being caused by the 'news' received on day t. To quantify this effect, we define the 'news impact curve' (NIC) as the 'unexpected return' on the GARCH-implied variance. That is, we take the unexpected part of $\sigma^2_{t+1|t} - \sigma^2_{t|t-1}$ and divide by $\sigma^2_{t|t-1}$ as follows:

$$\begin{aligned}
\mathrm{NIC}_t \;&:=\; \frac{\sigma^2_{t+1|t} - \sigma^2_{t|t-1} - \mathrm{E}_{t-1}[\sigma^2_{t+1|t} - \sigma^2_{t|t-1}]}{\sigma^2_{t|t-1}} = \frac{\sigma^2_{t+1|t} - \mathrm{E}_{t-1}\sigma^2_{t+1|t}}{\sigma^2_{t|t-1}},\\
&=\; \alpha_{\text{pos}}\left[\varepsilon_t^2\,1_{\varepsilon_t\geq 0} - \frac{1}{2}\right] + \alpha_{\text{neg}}\left[\varepsilon_t^2\,1_{\varepsilon_t<0} - \frac{1}{2}\right],
\end{aligned} \tag{2.7}$$

where $\mathrm{E}_t[\cdot] := \mathrm{E}[\cdot|\mathcal{I}_t]$ denotes an expectation conditional on the knowledge at time t, and where we have used equation (2.6) in the second line. The news impact curve is typically plotted (on the vertical axis) as a function of the implied residual ε_t (on the horizontal axis).

2.4 Predictions

Having investigated how the GARCH-filtered quantity $\sigma^2_{t|t-1}$ changes from day to day, we are also interested in d-day-ahead volatility forecasts. Because the terms in square brackets in equation (2.6) have expectation zero for any t, it follows that the d-day ahead prediction of the variance is given by

$$\mathrm{E}_t[\sigma^2_{t+d|t+d-1}] = \sigma^2 + \left(\frac{\alpha_{\text{pos}}}{2} + \frac{\alpha_{\text{neg}}}{2} + \beta\right)^{d-1} (\sigma^2_{t+1|t} - \sigma^2), \quad d \geq 1, \tag{2.8}$$

where E_t is again an expectation conditional on the knowledge at time t. You should check for yourself that this equation holds for $d = 1$ or $d = 2$. This equation says that the d-day-ahead forecast mean-reverts to the unconditional variance σ^2 at the geometric rate $\alpha_{\text{pos}}/2 + \alpha_{\text{neg}}/2 + \beta < 1$. To obtain the predicted volatility (PVol) for *the sum of days* up to (and including) some horizon h, we compute

$$\begin{aligned}
\mathrm{PVol}_t := \mathrm{E}_t \sum_{d=1}^{h} r^2_{t+d} &= \sum_{d=1}^{h} \mathrm{E}_t \left(\mu + \sigma_{t+d|t+d-1}\varepsilon_{t+d}\right)^2 = h\mu^2 + \sum_{d=1}^{h} \mathrm{E}_t[\sigma^2_{t+d|t+d-1}], \\
&= h\mu^2 + \sum_{d=1}^{h} \left[\sigma^2 + \left(\frac{\alpha_{\text{pos}}}{2} + \frac{\alpha_{\text{neg}}}{2} + \beta\right)^{d-1} (\sigma^2_{t+1|t} - \sigma^2)\right], \\
&= h(\mu^2 + \sigma^2) + \frac{1 - (\frac{\alpha_{\text{pos}}}{2} + \frac{\alpha_{\text{neg}}}{2} + \beta)^h}{1 - \frac{\alpha_{\text{pos}}}{2} - \frac{\alpha_{\text{neg}}}{2} - \beta} (\sigma^2_{t+1} - \sigma^2).
\end{aligned} \tag{2.9}$$

Here, the first line uses (2.3) and the fact that the shocks ε_t are i.i.d. with unit variance. The second line uses (2.8) above, while the last line uses the fact $\sum_{d=1}^{h} z^{d-1} = (1 - z^h)/(1 - z)$ for $0 \leq z < 1$ with $z = \alpha_{\text{pos}}/2 + \alpha_{\text{neg}}/2 + \beta$.

The variable PVol_t can be viewed as being a 'pseudo out-of-sample' prediction if the whole data set is used to estimate the constant parameters $(\mu, \omega, \alpha_{\text{pos}}, \alpha_{\text{neg}}, \beta, \nu)$. Conversely, prediction (2.9) can be viewed as a genuine out-of-sample prediction if the constant parameters are estimated using only data that would have been available at time t.

3 Recent advances in volatility modelling: Beta-t-EGARCH

To ensure that σ^2_t remains positive, an alternative approach is to set $\sigma_{t|t-1} = \exp(\lambda_{t|t-1})$, where $\lambda_{t|t-1}$ for $t = 1, \ldots, T$ is unrestricted. Here, the variable $\sigma_{t|t-1}$ is guaranteed to remain positive because of the exponential formulation. This reasoning leads to exponential GARCH models, also known as EGARCH models. The Beta-t-EGARCH model first formulated in Harvey and Chakravarty (2008) and later in Harvey (2013) is closely related to the following specification:

$$\begin{aligned}
&\text{close-to-close log return:} & r_t &= \mu + \exp(\lambda_{t|t-1})\,\varepsilon_t, & (3.1)\\
&\text{dynamic scale:} & \lambda_{t+1|t} &= \lambda\,(1 - \phi) + \phi\,\lambda_{t|t-1} + \kappa\,u_t + \tilde{\kappa}\,v_t, & (3.2)\\
&\text{robust variance measure:} & u_t &:= \frac{\sqrt{\nu + 3}}{\sqrt{2\nu}} \left(\frac{\nu + 1}{\nu - 2 + \varepsilon^2_t}\varepsilon^2_t - 1\right), & (3.3)\\
&\text{robust location measure:} & v_t &:= \frac{\sqrt{(\nu - 2)(\nu + 3)}}{\sqrt{\nu(\nu + 1)}} \frac{\nu + 1}{\nu - 2 + \varepsilon^2_t}\,\varepsilon_t, & (3.4)
\end{aligned}$$

where ε_t follows the p.d.f. (2.2) with $\nu > 2$ degrees of freedom. When taking $\tilde{\kappa} = 0$, the model above resembles that in Harvey and Lange (2017); see their equations (1) and (2). The six constant model parameters are μ, λ, ϕ, κ, $\tilde{\kappa}$ and ν. The dynamic parameter $\lambda_{t|t-1}$ is driven by the robust variance and location measures u_t and v_t, respectively, which will be shown to have mean zero (i.e. $\mathrm{E}_{t-1}u_t = \mathrm{E}_{t-1}v_t = 0$). We assume $|\phi| < 1$, such that λ is the unconditional mean of $\lambda_{t|t-1}$. Typically we set $\lambda_{1|0} = \lambda$; i.e., we initialise the time-varying parameter using the unconditional mean.

Before we delve into the intuition, we try to understand the model mechanically. Suppose the constant parameters and the initialisation $\lambda_{1|0} = \lambda$ are given. Given first close-to-close log return, r_1, we can compute the first implied residual, ε_1 via $\varepsilon_t = (r_t - \mu)\exp(-\lambda_{t|t-1})$ for $t = 1$. Given ε_1, we can compute the robust variance and location measures, u_1 and v_1. Given these measures, we can compute the new volatility measure $\lambda_{2|1}$. Next, when the observation r_2 is received, we can repeat this process and sequentially compute ε_2, u_2, v_2, $\lambda_{3|2}$, and so on. Thus, given the observations r_1 through r_T, given a starting point $\lambda_{1|0}$, and given the constant parameters, the model above allows us to recursively compute a whole sequence volatility measures $\exp(\lambda_{t|t-1})$ for all $t = 1, \ldots, T$.

Before motivating further why $\lambda_{t|t-1}$ should be driven by the quantities u_t and v_t, we note that they have the following attractive properties:

1. Mean zero ($\mathrm{E}_{t-1}u_t = \mathrm{E}_{t-1}v_t = 0$), standardised ($\mathrm{E}_{t-1}u_t^2 = \mathrm{E}_{t-1}v_t^2 = 1$) and orthogonal ($\mathrm{E}_{t-1}[v_t u_t] = 0$).
2. For $\nu < \infty$, u_t and v_t are both bounded when viewed as functions of ε_t; they are 'robust'.
3. As $\nu \to \infty$, we have $u_t \to \varepsilon_t^2 - 1$, while $v_t \to \varepsilon_t$. As such, u_t is a robust measure of variability, while v_t is a robust measure of location.

Having stated the Beta-t-EGARCH model, the next subsection explains *why* $\lambda_{t|t-1}$ should be driven by the quantities u_t and v_t.

3.1 Justification for using Beta-t-EGARCH

A large and growing body of literature on dynamic conditional score (DCS) models, also known as generalised autoregressive score (GAS) models, has suggested that time-varying quantities such as $\lambda_{t|t-1}$ should be driven by the *score* of the predictive destiny with respect to these quantities.[3] See, for example, Harvey (2013) and Creal et al. (2013).

The score is defined as the derivative of the relevant log-likelihood function with respect to the parameter of interest. In our case, we are interested in the predictive distribution of the return r_t conditional on the information set $\mathcal{I}_{t-1}$, which is given by

$$f(r_t|\mathcal{I}_{t-1}) \;=\; \frac{1}{\sigma_{t|t-1}}\, p\left(\frac{r_t-\mu}{\sigma_{t|t-1}}\right), \tag{3.5}$$

where $\sigma_{t|t-1} = \exp(\lambda_{t|t-1})$ and $p(\cdot)$ is given by (2.2). The DCS/GAS literature suggests that $\lambda_{t|t-1}$ should be driven by the score of this distribution with respect to $\lambda_{t|t-1}$. As you can verify for yourself, the score with respect to $\lambda_{t|t-1}$ is

$$\frac{\mathrm{d}\log f(r_t|\mathcal{I}_{t-1})}{\mathrm{d}\lambda_{t|t-1}} = \frac{\mathrm{d}}{\mathrm{d}\lambda_{t|t-1}}\left[-\lambda_{t|t-1} - \frac{\nu+1}{2}\log\left(1+\frac{\varepsilon_t^2}{\nu-2}\right)\right] = \frac{\nu+1}{\nu-2+\varepsilon_t^2}\varepsilon_t^2 - 1,$$

where, as always, $\varepsilon_t = (r_t - \mu)\exp(-\lambda_{t|t-1})$(which depends on $\lambda_{t|t-1}$) denotes the implied residual. The score with respect to $\lambda_{t|t-1}$ has mean zero (scores always do, for a general proof click here) and variance $2\nu/(\nu+3)$; you do not need to verify this. The DCS/GAS literature suggests it is useful to construct 'normalised' score variables by dividing the score by its standard deviation to obtain

$$u_t := \frac{\sqrt{\nu+3}}{\sqrt{2\nu}}\left(\frac{\nu+1}{\nu-2+\varepsilon_t^2}\varepsilon_t^2 - 1\right), \tag{3.6}$$

such that $\mathrm{E}u_t = 0$ and $\mathrm{E}u_t^2 = 1$. The fact that u_t is related to a random variable following a beta distribution explains why the resulting model is called Beta-t-EGARCH. The t relates to the distribution of the ε_t's, while the E in EGARCH refers to the exponential formulation $\sigma_{t|t-1} = \exp(\lambda_{t|t-1})$.

To allow the sign of implied residuals to affect volatility, we also compute the score with respect to the location parameter μ as follows:

$$\frac{\mathrm{d}\log f(r_t|\mathcal{I}_{t-1})}{\mathrm{d}\mu} = \frac{\nu+1}{\nu-2+\varepsilon_t^2}\,\frac{\varepsilon_t}{\exp\lambda_{t|t-1}}.$$

The variance of this score equals $\exp(-2\lambda_{t|t-1})\nu(\nu+1)/(\nu-2)/(\nu+3)$; you do not need to verify this. It follows that the normalised score with respect to μ reads

$$v_t := \frac{\sqrt{(\nu-2)(\nu+3)}}{\sqrt{\nu(\nu+1)}}\,\frac{\nu+1}{\nu-2+\varepsilon_t^2}\,\varepsilon_t. \tag{3.7}$$

Based on the DCS/GAS literature, it is natural that u_t should be used to drive the dynamics of $\lambda_{t|t-1}$. However, u_t depends on ε_t^2 but not ε_t, such that it takes into account the (absolute) size but not the sign of the shock ε_t. For this reason, we also allow v_t to drive the dynamics of $\lambda_{t|t-1}$.

3.2 News impact curve

To quantify the effect of shocks ε_t on $\lambda_{t+1|t}$, we define the 'news impact curve' (NIC) for Beta-t-EGARCH as follows. At the end of day t, we can compute the *unexpected* part of the difference between $\log(\sigma_{t+1|t})$ and

[3] For more details on DCS and GAS models, click here and here, respectively.

$\log(\sigma_{t|t-1})$ as follows:

$$\begin{aligned}
\mathrm{NIC}_t &:= \log(\sigma^2_{t+1|1}) - \log(\sigma^2_{t|t-1}) - \mathrm{E}_{t-1}\left[\log(\sigma^2_{t+1|1}) - \log(\sigma^2_{t|t-1})\right], \\
&= \log(\sigma^2_{t+1|1}) - \mathrm{E}_{t-1}\log(\sigma^2_{t+1|1}), \\
&= 2\lambda_{t+1|t} - 2\mathrm{E}_{t-1}\lambda_{t+1|t}, \\
&= 2\kappa u_t + 2\tilde{\kappa} v_t, \qquad (3.8)
\end{aligned}$$

where E_{t-1} denotes an expectation conditional on the knowledge at time $t-1$, and where we have used equation (3.2).

3.3 Predictions

Similar to (asymmetric) GARCH models, Beta-t-EGARCH may be used to compute the predicted volatility (PVol) over an h-day horizon as follows:

$$\mathrm{PVol}_t := \mathrm{E}_t \sum_{d=1}^{h} r^2_{t+d} \approx h\mu^2 + \sum_{d=1}^{h} \exp\left\{2\lambda + 2\phi^{d-1}(\lambda_{t+1|t} - \lambda) + 2(\kappa^2 + \tilde{\kappa}^2)\frac{1-\phi^{2(d-1)}}{1-\phi^2}\right\}, \qquad (3.9)$$

where E_t denotes an expectation conditional on the knowledge at time t. You can take the above expression as given; we explain it further in Appendix A. As before, the variable PVol_t can be viewed as being a 'pseudo out-of-sample' prediction if the whole data set is used to estimate the constant parameters, while it is a genuine out-of-sample prediction if the constant parameters are estimated using only data that were available at time t.

4 Estimation of volatility models

This section discusses the estimation of both the GARCH and Beta-t-EGARCH models using maximum likelihood (ML). For both models, the observation r_t is distributed according to the probability density function (p.d.f.) $f(r_t|\mathcal{I}_{t-1})$, where

$$f(r_t|\mathcal{I}_{t-1}) = \frac{1}{\sigma_{t|t-1}}\, p\left(\frac{r_t-\mu}{\sigma_{t|t-1}}\right) = \frac{1}{\sigma_{t|t-1}}\, p\left(\varepsilon_t\right), \qquad (4.1)$$

where $p(\cdot)$ is the p.d.f. of the Student's t distribution given in (2.2). For Beta-t-EGARCH, recall that $\sigma_{t|t-1} = \exp(\lambda_{t|t-1})$. To estimate the constant parameters, consider the GARCH filter (2.6), in which case the vector of constant parameters reads $\boldsymbol{\theta} = (\mu, \omega, \alpha_{\text{pos}}, \alpha_{\text{neg}}, \beta, \nu)'$. The likelihood function $L(\boldsymbol{\theta})$ is the joint p.d.f. of the returns, that is

$$L(\boldsymbol{\theta}) = f(r_1, r_2, \ldots, r_T; \boldsymbol{\theta}) = \prod_{t=1}^{T} f(r_t|\mathcal{I}_{t-1}; \boldsymbol{\theta}).$$

ML estimation involves finding the parameter vector $\boldsymbol{\theta}$ that maximises (the logarithm of) this function, that is

$$\begin{aligned}
\widehat{\boldsymbol{\theta}}_{\mathrm{ML}} = \mathrm{argmax}_{\boldsymbol{\theta}}\, L(\boldsymbol{\theta}) &= \mathrm{argmax}_{\boldsymbol{\theta}} \sum_{t=1}^{T} \log f(r_t|\mathcal{I}_{t-1}; \boldsymbol{\theta}), \\
&= \mathrm{argmax}_{\boldsymbol{\theta}} \sum_{t=1}^{T} \left[-\log\left(\sigma_{t|t-1}\right) + \log p\left(\frac{r_t-\mu}{\sigma_{t|t-1}}\right)\right], \\
&= \mathrm{argmax}_{\boldsymbol{\theta}} \sum_{t=1}^{T} \left[-\log\left(\sigma_{t|t-1}\right) + \log p\left(\varepsilon_t\right)\right]. \qquad (4.2)
\end{aligned}$$

The same procedure is used to estimate Beta-t-EGARCH, where $\sigma_{t|t-1} = \exp(\lambda_{t|t-1})$ and $\lambda_{t|t-1}$ evolves according to equation (3.2). As equation (4.2) makes clear, we aim to find constant parameters such that the log likelihood of the implied shocks $\log p(\varepsilon_t)$ is maximised, without making $\sigma_{t|t-1}$ excessively large, as this term appears with a minus sign.

In general, no closed-form or analytic solutions are available, but 'standard' numerical optimisation procedures can be used to obtain $\widehat{\boldsymbol{\theta}}_{\mathrm{ML}}$. We have provided you with a very basic Matlab demo code, which estimates the simplest GARCH(1,1) model given by equations (2.3), (2.4) and (4.2), with Student's t distributed shocks ε_t's. This demo code uses the Matlab function `fmincon` to minimise the negative log likelihood function. Because Matlab likes to minimise rather than maximise, we work with the negative log likelihood. In your report, make sure to report the actual log likelihood. This actually matters for the interpretation: for the negative log likelihood, lower is better, while for the (actual) log likelihood, higher is better. You may use this demo code as your starting point for estimating (i) the asymmetric GARCH model (2.6) and (ii) Beta-t-EGARCH. No further coding help will be available from the supervisors.

5 Evaluating model forecasts

To evaluate the accuracy of GARCH predictions (2.9) and Beta-t-EGARCH predictions (3.9), we consider two 'target variables' as follows:

1. **Target variable 1: Target variance (TV) computed from daily returns**. Since we have access to daily returns, we can compute the target variance over the next d days when viewed from day t as follows:

$$\mathrm{TV}_{t,d} = \sum_{i=1}^{d} r_{t+i}^2, \tag{5.1}$$

   where r_{t+i} is the close-to-close log return on day $t+i$. While the (forward-looking) target variable $\mathrm{TV}_{t,d}$ is not available on day t, it represents our *target* in the sense that we would like to predict it on day t.

2. **Target variable 2: Target variance (TV) computed from intra-day returns**. Alternatively, we may use daily realised-variance measures obtained from intra-day returns. While this measure excludes overnight returns, experience suggests that around seventy percent of the total (daily and overnight) variance is realised during the day. For this reason, we 'scale up' the Oxford realised variance by a factor $1/0.7 \approx 1.4$.[4] The target variance over a d-day horizon can thus be defined as

$$\mathrm{TV}_{t,d} = 1.4 \times \sum_{i=1}^{d} \mathrm{rv5}_{t+i}, \tag{5.2}$$

   where $\mathrm{rv5}_{t+i}$ is realised variance of intra-day returns on day $t+i$.

The advantage of target variable 1, which uses close-to-close log returns, is that it includes both daily and overnight volatility. The disadvantage is that much information is potentially ignored by using only closing prices. The advantage of target variable 2, which uses intra-day returns, is that it accurately estimates the volatility during the opening hours of the market. The disadvantage is that overnight volatility is only artificially included via the *ad hoc* multiplicative factor of 1.4.

We now have two models (GARCH and Beta-t-EGARCH) and two target variables. Moreover, we are interested in three horizons, namely $d = 1$, $d = 5$ and $d = 21$ trading days, which correspond to roughly to one calendar day, one calendar week and one calendar month, respectively. As our benchmark models, we consider the VIX and a simple autoregressive model for realised volatility, as discussed next.

5.1 Benchmark model 1: The VIX

The VIX intends to predict the (annualised) standard deviation of the return of the S&P500 index in the next month. Based on the VIX, the predicted variance at time t over a d-day horizon is

$$\mathrm{PV}_{t,d} = d/250 \times \mathrm{VIX}_t^2. \tag{5.3}$$

As the VIX squared can be interpreted as an annualised variance, we multiply by $d/250$ to account for the d-day horizon. The factor of 250 originates from the fact that there are roughly 250 days per year on which the market is open.

5.2 Benchmark model 2: The HAR-RV model

Corsi (2009) proposes an heterogeneous autoregressive realised variance (HAR-RV) model that directly models the target variable that we are interested in. He models the target variable as a linear function of several one-period lagged realised variances, which are measured at different (i.e., heterogeneous) frequencies, explaining the name of the model.

Specifically, the realised variances in the HAR-RV model correspond to the daily, weekly and monthly frequencies. The daily realised variance, denoted $\mathrm{RV}_t^{(d)}$, is simply $\mathrm{rv5}_t$. The weekly and monthly realised variances, denoted $\mathrm{RV}_t^{(w)}$ and $\mathrm{RV}_t^{(m)}$, respectively, equal the *average* realised variance over these periods, i.e.

$$\text{weekly:} \qquad \mathrm{RV}_t^{(w)} = \frac{1}{5}\left(\mathrm{RV}_t^{(d)} + \mathrm{RV}_{t-1}^{(d)} + \ldots + \mathrm{RV}_{t-4}^{(d)}\right),$$

$$\text{monthly:} \qquad \mathrm{RV}_t^{(m)} = \frac{1}{21}\left(\mathrm{RV}_t^{(d)} + \mathrm{RV}_{t-1}^{(d)} + \ldots + \mathrm{RV}_{t-20}^{(d)}\right).$$

[4] This is consistent with equation (1) in Hansen and Lunde (2005), because the factor 1.4 brings the average of target variable 2 in line with that of target variable 1.

The HAR-RV model is implemented by regressing via ordinary least squares (OLS) the target variable $\mathrm{TV}_{t,d}$ onto the three realised-variance components as follows:

$$\mathrm{TV}_{t,d} = c_d + \beta_d^{(d)}\mathrm{RV}_t^{(d)} + \beta_d^{(w)}\mathrm{RV}_t^{(w)} + \beta_d^{(m)}\mathrm{RV}_t^{(m)} + e_t, \tag{5.4}$$

where the error e_t satisfies the appropriate assumptions. The OLS output (estimates of $c_d, \beta_d^{(d)}, \beta_d^{(w)}, \beta_d^{(m)}$) will differ for different d-day-ahead prediction horizons, as indicated by the subscripts. If we use data before date t to estimate the constant parameters, we can compute the out-of-sample predicted variance (PV) over a d-day horizon as follows:

$$\mathrm{PV}_{t,d} = \hat{c}_d + \hat{\beta}_d^{(d)}\mathrm{RV}_t^{(d)} + \hat{\beta}_d^{(w)}\mathrm{RV}_t^{(w)} + \hat{\beta}_d^{(m)}\mathrm{RV}_t^{(m)}, \tag{5.5}$$

where hats denote OLS estimates based on data that would have been available at time t.

In sum, you now have four predicted variances (PVs), given by equations (2.9), (3.9), (5.3), and (5.5), two target variances (TVs), given in equations (5.1) and (5.2), and three horizons ($d = 1, 5, 21$). For any given loss function (e.g. MSE or QLIKE), this allows you compute $4 \times 2 \times 3 = 24$ loss values.

6 Roadmap

This section contains a logical sequence of modelling steps that you may consider. Note, however, that this is *not* how you should write your report. It is up to you to decide which analyses to perform and how to report the output. When you have completed (some or all of) the modelling steps presented below, you will have to decide which results to report and which to leave out.

1. **In-sample analysis.** Estimate four models by maximum likelihood (ML) using the entire data set:
   (a) Symmetric GARCH, i.e. set $\alpha = \alpha_1 = \alpha_2$. You can use the demo code for this purpose
   (b) Asymmetric GARCH, i.e. allow $\alpha_1 \neq \alpha_2$
   (c) Symmetric Beta-t-EGARCH, i.e. set $\tilde{\kappa} = 0$
   (d) Asymmetric Beta-t-EGARCH, i.e. allow $\tilde{\kappa} \neq 0$

   In a single table, report parameter estimates for all models and corresponding log likelihoods. For parameter estimates, report 3 decimal places. For log likelihoods, one decimal place is enough. You may include additional statistics if you want. Not that a more flexible model should always have a *higher* log-likelihood value.

2. Of the two GARCH models, which do you think is best? Of the two Beta-t-EGARCH models, which do you think is best? Use statistical arguments. From now on, consider only your favourite GARCH model and your favourite Beta-t-EGARCH model.

3. For your two favourite models, discuss the difference in parameter estimates. As much as possible, draw economic (not econometric) conclusions.

4. For your two favourite models, plot implied residuals ε_t (on the horizontal axis) against the estimated news impact curves (2.7) and (3.8) (on the vertical axis) Discuss asymmetry, shape and robustness.

5. For your favourite GARCH and Beta-t-EGARCH models, and all prediction horizons, perform Mincer-Zarnowitz (MZ) regressions[5] using target variable 1 or 2 as your dependent variable. Report constants and slopes with their standard errors, root mean squared errors and R^2 values. Interpret your results.

6. **Out-of-sample analysis.** Use only the first x% of the data to estimate the constant parameters, e.g. 50%. Alternatively, you may also use an expanding window or moving window to estimate the constant parameters. In the out-of-sample period, construct volatility predictions for $d = 1$, $d = 5$ and $d = 21$ days ahead, for the HAR-RV, VIX and your favourite (realised) (E)GARCH models, leading to $4 \times 3 = 12$ series of predictions.

7. Compare your 12 series of predictions against both possible targets using the MSE and QLIKE robust loss functions of Patton (2011). This means you have 4 models (GARCH, Beta-t-EGARCH, VIX, HAR-RV), 3 horizons, 2 targets, and at least 2 loss functions, leading to at least $4 \times 3 \times 2 \times 2 = 48$ numbers. Organise your results in such a way that they can be easily read and interpreted. Pretty tables and figures are important; it will be detrimental to your grade if assessors have trouble understanding them.

8. **Bonus question:** Does any model *significantly* outperform the others in making predictions?

[5]For an explanation of MZ regressions, refer to "Chapter 3 - Basic concepts" of the course "Time Series Analysis".

9. **Conclusions.** In the conclusion of your report, you may address the provocative question raised by Hansen and Lunde (2005): "Does anything beat a GARCH(1,1)?". They conclude that the answer is essentially "no" as long as you allow for asymmetric GARCH as in equation (2.6). Do you agree with their assessment?

10. *Note for the final version:* Your reference list should be carefully edited. Please see the references at the end of this case study: each reference has a journal name with appropriate capitalisation, as well as volume, issue and page numbers. Book titles are capitalised differently from journal titles. The year goes after the authors. Authors have initials, not full names. References are listed alphabetically, not numerically. Hyperlinks are useful but not mandatory. Points will be deducted for erroneous or inconsistent reference lists.

7 Econometric software

You may use any software package, such as EVIEWS, Matlab, R, etc. To encourage you to write your own code, a basic demo code for estimating a simple GARCH(1,1) model with Student's t distributed errors is provided in Matlab. No additional coding help will be available from the supervisors.

8 Proposal

The research proposal should be no more than two pages and should contain at least:

1. Output for steps 1 through 5 of the Roadmap in Section 6.
2. Output tables that you would *like* to produce for the remaining steps in the Roadmap. These tables may still be *empty*, i.e. contain no numbers, but the rows and columns should be clearly labelled, so that supervisors get a sense of the output you intend to produce.

Use the proposal to sketch your intended output, *not* to summarise the case (as the supervisors already know the case). Proposals that do not contain the minimum requirements will receive minimal feedback. You can also present your own (possibly creative) ideas and ask for feedback.

9 Final report

The final report consists of **at most 5,000 words or 12 pages**, whichever comes first. Both limits include the cover page and references, but not the appendices. Any part of the main report that stretches beyond either of the above limits will lead to a deduction of points. Appendices (e.g. page 13 and beyond) are allowed but should not contain important information. Essentially, you should assume the examiners will stop reading at the end of page 12 (or after 5,000 words).

The report should focus on what you have done, not on summarizing the case; e.g. you need *not* replicate all formulas or derivations. Importantly, the report should *not* refer to this case study, let alone to the roadmap above. Rather, your report should be a *completely stand-alone document* that is understandable to other econometrics students unfamiliar with this case study. Hence, you should only refer to publicly available sources and all modelling choices should be clearly motivated.

Reports that fail to mention the group number will not be graded. Neither will late submissions. Reports must be written in Dutch or English for FEB22009 students and in English for FEB22009X, FEB22009Q and FEB22009S students. In addition to the group number, the cover page should contain the title of the report, the names of all team members together with their student numbers and two word counts: one for the main documents (at most 5,000 words) and one for the abstract (at most 150 words). Do not include a table of contents and/or list of tables or figures. Reports should be in line with common rules and practice in scientific publishing as taught in the skills course. When in doubt, follow the notation in this case study. You are expected to write your report using LaTeX.

References

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

Corsi, F., 2009, A simple approximate long-memory model of realized volatility, *Journal of Financial Econometrics* 7, 174–196.

Creal, D., S. J. Koopman, and A. Lucas, 2013, Generalized autoregressive score models with applications, *Journal of Applied Econometrics* 28, 777–795.

Engle, R. F., 1982, Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica: Journal of the Econometric Society* 50, 987–1007.

Franses, P. H., D. van Dijk, and A. Opschoor, 2014, *Time series models for business and economic forecasting* (Cambridge University Press).

Glosten, L. R., R. Jagannathan, and D. E. Runkle, 1993, On the relation between the expected value and the volatility of the nominal excess return on stocks, *The Journal of Finance* 48, 1779–1801.

Hansen, P. R., and A. Lunde, 2005, A forecast comparison of volatility models: Does anything beat a GARCH(1,1)?, *Journal of Applied Econometrics* 20, 873–889.

Harvey, A., and R.-J. Lange, 2017, Volatility Modeling with a Generalized t Distribution, *Journal of Time Series Analysis* 38, 175–190.

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

Harvey, A. C., and T. Chakravarty, 2008, Beta-t-(E)GARCH, *Unpublished manuscript* .

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

A Derivation of equation (3.9)

Although you are not required to understand the following derivation, equation (**??**) can be derived as follows:

$$\begin{aligned}
\mathrm{E}_t \sum_{d=1}^{h} r_{t+d}^2 &= h\mu^2 + \mathrm{E}_t \sum_{d=1}^{h} (r_{t+d} - \mu)^2 && \text{(A.1)} \\
&= h\mu^2 + \mathrm{E}_t \sum_{d=1}^{h} \exp(2\lambda_{t+d|t+d-1})\,\varepsilon_{t+d}^2, \\
&= h\mu^2 + \sum_{d=1}^{h} \mathrm{E}_t \exp(2\lambda_{t+d|t+d-1}), && \text{as } \varepsilon_t \text{ is i.i.d. with mean zero and variance one,} \\
&\approx h\mu^2 + \sum_{d=1}^{h} \exp\left(2\mathrm{E}_t[\lambda_{t+d|t+d-1}] + 2\mathrm{V}_t[\lambda_{t+d|t+d-1}]\right), && \text{log normal approximation} \\
&= h\mu^2 + \sum_{d=1}^{h} \exp\left(2\lambda + 2\phi^{d-1}(\lambda_{t+1|t} - \omega) + 2(\kappa^2 + \tilde{\kappa}^2)\frac{1-\phi^{2(d-1)}}{1-\phi^2}\right). && \text{(A.2)}
\end{aligned}$$

The approximation in the third line pretends that $2\lambda_{t+d|t+d-1}$ is normally distributed with mean $2\mathrm{E}_t[\lambda_{t+d|t+d-1}]$ and variance $4\mathrm{V}_t[\lambda_{t+d|t+d-1}]$. Under this approximation we can use the well known fact that the expectation of a log-normal random variable equals $\mathrm{E}\exp(\mathrm{N}(m, s^2)) = \exp(m + s^2/2)$. The last line follows from

$$\mathrm{V}_t[\lambda_{t+d|t+d-1}] = (\kappa^2 + \tilde{\kappa}^2)\frac{1-\phi^{2(d-1)}}{1-\phi^2}, \tag{A.3}$$

where $\mathrm{V}_t[\cdot] = \mathrm{V}[\cdot|\mathcal{I}_t]$ denotes the variance conditional on the information at time t. To complete the proof, we must show that equation (A.3) holds. To this end, we consider

$$\begin{array}{llll}
d = 1 & \mathrm{V}_t[\lambda_{t+1|t}] = 0, & & \\
d = 2 & \mathrm{V}_t[\lambda_{t+2|t+1}] = \mathrm{V}_t[\phi\lambda_{t+1|t} + \kappa u_t + \tilde{\kappa} v_t] & = \kappa^2 + \tilde{\kappa}^2, & \\
d = 3 & \mathrm{V}_t[\lambda_{t+3|t+2}] = \phi^2 \mathrm{V}_t[\lambda_{t+2|t+1}] + \kappa^2 + \tilde{\kappa}^2 & = (\kappa^2 + \tilde{\kappa}^2)(1 + \phi^2), & \\
d = 4 & \mathrm{V}_t[\lambda_{t+4|t+3}] = \phi^2 \mathrm{V}[\lambda_{t+3|t+2}|\mathcal{I}_t] + \kappa^2 + \tilde{\kappa}^2 & = (\kappa^2 + \tilde{\kappa}^2)(1 + \phi^2 + \phi^4), & \\
\vdots & \vdots & \vdots & \\
d = d & \mathrm{V}_t[\lambda_{t+d|t+d-1}] & = (\kappa^2 + \tilde{\kappa}^2)\displaystyle\sum_{i=1}^{d-1} \phi^{2(i-1)}. & \text{(A.4)}
\end{array}$$

Finally, to confirm (A.3), note that the above computation implies

$$\mathrm{V}_t[\lambda_{t+d|t+d-1}] = (\kappa^2 + \tilde{\kappa}^2)\sum_{i=1}^{d-1} \phi^{2(i-1)} = (\kappa^2 + \tilde{\kappa}^2)\frac{1-\phi^{2(d-1)}}{1-\phi^2}, \tag{A.5}$$

where, in the second equality, we have used $\sum_{i=1}^{d-1} z^{i-1} = (1 - z^{d-1})/(1-z)$ for $|z| < 1$ with $z = \phi^2$.