Realising the future: forecasting with high frequency based volatility (HEAVY) models*

Neil Shephard
Oxford-Man Institute, University of Oxford,
Eagle House, Walton Well Road, Oxford OX2 6EE, UK
&
Department of Economics, University of Oxford
neil.shephard@economics.ox.ac.uk

Kevin Sheppard
Department of Economics, University of Oxford,
Manor Road Building, Manor Road, Oxford, OX1 3UQ, UK
&
Oxford-Man Institute, University of Oxford
kevin.sheppard@economics.ox.ac.uk

July 10, 2009

Abstract

This paper studies in some detail a class of high frequency based volatility (HEAVY) models. These models are direct models of daily asset return volatility based on realized measures constructed from high frequency data. Our analysis identifies that the models have momentum and mean reversion effects, and that they adjust quickly to structural breaks in the level of the volatility process. We study how to estimate the models and how they perform through the credit crunch, comparing their fit to more traditional GARCH models. We analyse a model based bootstrap which allow us to estimate the entire predictive distribution of returns. We also provide an analysis of missing data in the context of these models.

Keywords: ARCH models; bootstrap; missing data; multiplicative error model; multistep ahead prediction; non-nested likelihood ratio test; realised kernel; realised volatility.

1 Introduction

This paper analyses the performance of some predictive volatility models built to exploit high frequency data. This is carried out through the development of a class of models we call high frequency based volatility (HEAVY) models, which are designed to harness high frequency data to make multistep ahead predictions of the volatility of returns. These models allow for both mean reversion and momentum. They are somewhat robust to certain types of structural breaks and

*We thank Tim Bollerslev, Rob Engle, Nathaniel Frank, Giampiero Gallo, Andrew Patton and Natalia Sizova for various helpful suggestions. We are responsible for any errors.

adjust rapidly to changes in the level of volatility. The models are run across periods where the level of volatility has varied substantially to assess their ability to perform in stressful environments.

Our approach to inference will be based on the use of the "OMI's realised measure library" of historical volatility statistics, constructed using high frequency data (OMI denotes the Oxford-Man Institute). Such statistics are based on a variety of theoretically sound non-parametric estimators of the daily variation of prices. In particular it includes two estimators of interests to us. The first is realised variance, which was systematically studied by Andersen, Bollerslev, Diebold, and Labys (2001) and Barndorff-Nielsen and Shephard (2002). The second, which has some robustness to the effect of market microstructure effects, is realised kernel, which was introduced by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008a). Alternatives to the realised kernel include the multiscale estimators of Zhang, Mykland, and Aït-Sahalia (2005) and Zhang (2006) and the preaveraging estimator of Jacod, Li, Mykland, Podolskij, and Vetter (2009)[1].

The focus of this paper is on predictive models, rather than on non-parametric measurement of past volatility. Torben Andersen, Tim Bollerslev and Frank Diebold, with various coauthors, have carried out important work on looking at predicting volatility using realised variances. Typically they fit reduced form time series models of the sequence of realised variances — e.g. autoregressions or long memory models on the realised volatilities or their logged versions. Examples of this work include Andersen, Bollerslev, Diebold, and Labys (2001), Andersen, Bollerslev, Diebold, and Labys (2003), Andersen, Bollerslev, Diebold, and Ebens (2001) and Andersen, Bollerslev, and Diebold (2007).

The approach we follow in this paper is somewhat different. We build models out of the intellectual insights of the ARCH literature pioneered by Engle (1982) and Bollerslev (1986), but bolster them with high frequency information. The resulting models will be called HEAVY models. These models also use ideas generated by Engle (2002), Engle and Gallo (2006) and Cipollini, Engle, and Gallo (2007) in their work on pooling information across multiple volatility indicators and the paper by Brownlees and Gallo (2009) on risk management using realised measures. Our analysis can be thought of as taking a small subset of some of the Engle et al. models and analysing them in depth for a specific purpose looking at their performance over many assets. Our model structure is very simple which allows us to cleanly understand its general features, strengths and potential weaknesses. We provide no new contribution to estimation theory, simply using existing results on quasi-likelihoods. We show that when we marginalise out the effect of the realised measures that simple HEAVY models of squared returns have some similarities with the component GARCH model of Engle and Lee (1999, equation (3.2)). However, HEAVY models are much easier

[1]See also the work by Bandi and Russell (2008), Bandi and Russell (2006), Andersen, Bollerslev, and Meddahi (2006), Hansen and Lunde (2006), Corradi and Distaso (2006), and Christensen and Podolskij (2007).

to estimate as they bring two sources of information to identify the longer term component of volatility. We further find that the additional information in the realized measure generates out-of-sample gains, which are particularly strong when the parameters of the model are estimated to match the prediction horizon, using so-called "direct projection."

The structure of this paper is the following. In Section 2 we will define HEAVY models, which use realised measures as the basis for multi-period ahead forecasting of volatility. We provide a detailed analysis of these models. In Section 3 we detail the main properties of "OMI's realised measures library" which we use throughout the paper. In Section 4 we fit the HEAVY models to the data and compare their predictions to those familiar from GARCH processes. Section 5 gives a discussion of possible extensions. Section 6 draws some conclusions.

2 HEAVY models

2.1 Assumed data structure

Our analysis will be based on daily financial returns

$$r_1, r_2, ..., r_T,$$

and a corresponding sequence of daily realised measures

$$RM_1, RM_2, ..., RM_T.$$

Realised measures are theoretically sound high frequency, nonparametric based estimators of the variation of the price path of an asset during the times at which the asset trades frequently on an exchange. Realised measures ignore the variation of prices overnight and sometimes the variation in the first few minutes of the trading day when recorded prices may contain large errors. The background to realised measures can be found in the survey articles by Andersen, Bollerslev, and Diebold (2009) and Barndorff-Nielsen and Shephard (2007).

The simplest realised measure is realised variance

$$RM_t = \sum_{0 \leq t_{j-1,t} < t_{j,t} \leq 1} x_{j,t}^2, \quad x_{j,t} = X_{t+t_{j,t}} - X_{t+t_{j-1,t}} \tag{1}$$

where $t_{j,t}$ are the normalised times of trades or quotes (or a subset of them) on the t-th day. The theoretical justification of this measure is that if prices are observed without noise then as $\min_j |t_{j,t} - t_{j-1,t}| \downarrow 0$ it consistently estimates the quadratic variation of the price process on the t-th day. It was formalised econometrically by Andersen, Bollerslev, Diebold, and Labys (2001) and Barndorff-Nielsen and Shephard (2002). In practice market microstructure noise plays an important part and the above authors use 1-5 minute return data or a subset of trades or quotes (e.g. every

15th trade) to mitigate the effect of the noise. Hansen and Lunde (2006) systematically study the impact of noise on realised variance. If a subset of the data is used with the realised variance, then it is possible to average across many such estimators each using different subsets. This is called subsampling. When we report RV estimators we always subsample them to the maximise degree possible from the data as this averaging is always theoretically beneficial especially in the presence of modest amounts of noise.

Three classes of estimators which are somewhat robust to noise have been suggested in the literature: preaveraging (Jacod, Li, Mykland, Podolskij, and Vetter (2009)), multiscale (Zhang (2006) and Zhang, Mykland, and Aït-Sahalia (2005)) and realised kernel (Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008a))[2]. Here we focus on the realised kernel in the case where we use a Parzen weight function. It has the familiar form of a HAC type estimator (except there is no adjustment for mean and the sums are not scaled by their sample size)

$$RM_t = \sum_{h=-H}^{H} k\left(\tfrac{h}{H+1}\right)\gamma_h, \quad \gamma_h = \sum_{j=|h|+1}^{n} x_{j,t}x_{j-|h|,t}, \tag{2}$$

where $k(x)$ is the Parzen kernel function

$$k(x) = \begin{cases} 1-6x^2+6x^3 & 0 \leq x \leq 1/2 \\ 2(1-x)^3 & 1/2 \leq x \leq 1 \\ 0 & x > 1. \end{cases}$$

It is necessary for H to increase with the sample size in order to consistently estimate the increments of quadratic variation in the presence of noise. We follow precisely the bandwidth choice of H spelt out in Barndorff-Nielsen, Hansen, Lunde, and Shephard (2009), to which we refer the reader for details. This realised kernel is guaranteed to be non-negative, which is quite important as some of our time series methods rely on this property.

2.2 Definitions

We will write a sequence of daily returns as r_1, r_2, ..., r_T, while we will use $\mathcal{F}_{t-1}^{LF}$ to denote low frequency past data. A benchmark model for time-varying volatility is the GARCH model of Engle (1982) and Bollerslev (1986) where we assume that

$$\text{Var}\left(r_t|\mathcal{F}_{t-1}^{LF}\right) = \sigma_t^2 = \ \omega_G + \alpha_G r_{t-1}^2 + \beta_G \sigma_{t-1}^2.$$

This can be extended in many directions, for example allowing for statistical leverage. The persistence of this model, $\alpha_G + \beta_G$, can be seen through the representation

$$\sigma_t^2 = \ \mu_G + \alpha_G\left(r_{t-1}^2 - \sigma_{t-1}^2\right) + \left(\alpha_G + \beta_G\right)\sigma_{t-1}^2,$$

[2]See also the important work of Fan and Wang (2007) on the use of wavelets in this context.

since $r_t^2 - \sigma_t^2$ is a martingale difference with respect to $\mathcal{F}_{t-1}^{LF}$.

Our focus is on additionally using some daily realised measures. The models we will analyse will be called "HEAVY models" (High frEquency bAsed VolatilitY models) and are made up of the system

$$\left\{ \begin{array}{c} \mathrm{Var}\left(r_t|\mathcal{F}_{t-1}^{HF}\right) \\ \mathrm{E}\left(RM_t|\mathcal{F}_{t-1}^{HF}\right) \end{array} \right\}, \quad t = 2, 3, ..., T,$$

where $\mathcal{F}_{t-1}^{HF}$ is used to denote the past of r_t and RM_t, that is the high frequency dataset. The most basic example of this is the linear model

$$\mathrm{Var}\left(r_t|\mathcal{F}_{t-1}^{HF}\right) = h_t = \omega + \alpha RM_{t-1} + \beta h_{t-1}, \quad \omega, \alpha \geq 0, \quad \beta \in [0,1], \tag{3}$$

$$\mathrm{E}\left(RM_t|\mathcal{F}_{t-1}^{HF}\right) = \mu_t = \omega_R + \alpha_R RM_{t-1} + \beta_R \mu_{t-1}, \quad \omega_R, \alpha_R, \beta_R \geq 0, \alpha_R + \beta_R \in [0,1], \tag{4}$$

These semiparametric models could be extended to include on the right hand side of both equations the variable r_{t-1}^2 (see the discussion above (5) in a moment) but we will see these variables typically test out. Hence it is useful to focus directly on the above model. Other possible extensions include adding a more complicated dynamic to (4), such as a component structure with short and long term components, a fractional model, allowing for statistical leverage type effects, or a Corsi (2009) type approximate long-memory model.

Remark 1 *(3) models the close-to-close conditional variance, while (4) models the conditional expectation of the open-to-close variation.*

It will be convenient to have labels for the two equations in the HEAVY model. We call (3) the HEAVY-r model and (4) the HEAVY-RM model. Econometrically it is important to note that GARCH and HEAVY models are non-nested.

It is helpful to solve out explicitly the HEAVY-r model and GARCH models as

$$\mathrm{Var}\left(r_t|\mathcal{F}_{t-1}^{HF}\right) = \frac{\omega}{1-\beta} + \alpha \sum_{j=0}^{\infty} \beta^j RM_{t-1-j}, \quad \mathrm{Var}\left(r_t|\mathcal{F}_{t-1}^{LF}\right) = \frac{\omega_G}{1-\beta_G} + \alpha_G \sum_{j=0}^{\infty} \beta_G^j r_{t-1-j}^2.$$

In applied work we will typically estimate β to be around 0.6 to 0.7 and ω to be small. So the HEAVY-r's conditional variance is roughly a small constant plus a weighted sum of very recent realised measures. In estimated GARCH models in our later empirical work β_G is usually around 0.91 or above, so has much more memory and so averages more data points.

We can quantify this differential weighting in the following way. Let us focus on the number of days j which have weights β^j or β_G^j above 0.1. In the GARCH case with $\beta_G = 0.91$ this would be around 24.4 days. In the HEAVY-r model with $\beta = 0.7$ it amounts to 4.5 days. Hence, roughly, GARCH models average a month or so of recent squared returns and HEAVY-r models average a week of realised measures.

Notice that unlike GARCH models, the HEAVY-r model has no feedback and so the conditional variance of returns is entirely determined by the sequence of realised measures. Hence the properties of the realised measures determine the properties of $\text{Var}\left(r_t|\mathcal{F}_{t-1}^{HF}\right)$.

The predictive model for the times series of realised measures is not novel. The work of Andersen, Bollerslev, Diebold, and Labys (2001), Andersen, Bollerslev, Diebold, and Labys (2003), Andersen, Bollerslev, Diebold, and Ebens (2001) and Andersen, Bollerslev, and Diebold (2007) typically looked at using least squares estimators of autoregressive cousins discussed in (4) or their logged transformed versions. These authors also emphasised the evidence for long memory in these time series and studied various ways of making inference for those types of processes. Some of this work uses the model of Corsi (2009) which is easy to estimate and mimics some aspects of long memory.

Engle (2002) estimated GARCHX type models, which specialise to (3), for foreign exchange data using realised variances computed using 5 minute returns. He found the coefficient on r_{t-1}^2 to be small. He also fitted models like (4) but again including lagged square daily returns. He argues that the squared daily return helps forecast the realised variance although there is some uncertainty over whether the effect is statistically significant (see his footnote 2). He did not, however, express (3)-(4) as a simple basis for a multistep ahead forecasting system.

Engle and Gallo (2006) extended Engle (2002) to look at multiple volatility indicators, trying to pool information across many indicators including daily ranges — rather than focusing solely on theoretically sound high frequency based statistics. They then relate this to the VIX. In that paper they do study multistep ahead forecasting using a trivariate system which has daily absolute returns, daily range and realised variance (computed using 5 minute returns for the S&P500, so using a very small sample size). Their estimated models are quite sophisticated with again daily returns playing a large role in predicting each series. These results are at odds with our own empirical experience expressed in section 4. Some clues as to why this might be the case can be seen from their Table 1 which shows realised volatility having roughly the same average level as absolute returns and daily range but realised volatility being massively more variable and having a very long right hand tail. It perhaps suggests their realised measures were quite poor which distracted from the power and simplicity of using realised measures in HEAVY type models.

Brownlees and Gallo (2009) look at risk management in the context of exploiting high frequency data. Their model, in Section 5 of their paper, links the conditional variance of returns to an affine transform of the predicted realised measure. In particular their model has a HEAVY type structure but instead of using $h_t = \omega + \alpha RM_{t-1} + \beta h_{t-1}$ they model $h_t = \omega_B + \alpha_B \mu_t$. That is they place in the HEAVY-r equation a smoothed version μ_t of the lagged realised measures where the smoothing

is chosen to perform well in the HEAVY-RM equation, rather than the raw version which is then smoothed through the role of the momentum parameter β (which is optimally chosen to preform well in the HEAVY-r equation). Although these models are distinct, they have quite a lot of common thinking in their structure. Maheu and McCurdy (2009) have similarities with Brownlees and Gallo (2009), but focusing on an even more tightly parameterised model working with open-to-close daily returns (i.e. ignoring overnight effects) where realised variance captures much of the variation of the asset price. Giot and Laurent (2004) looks at some similar types of models.

Finally for some data the realised measure is not enough to entirely crowd out the lagged squared squared daily returns. In that case it makes sense to augment the HEAVY-r model into its extended version

$$\mathrm{Var}(r_t|\mathcal{F}_{t-1}^{HF}) = h_t = \omega_X + \alpha_X RM_{t-1} + \beta_X h_{t-1} + \gamma_X r_{t-1}^2, \quad \beta_X + \gamma_X < 1. \tag{5}$$

This could be thought of as a GARCHX type model, but that name suggests that it is the squared returns which drives the model, where in fact in our empirical work it is the lagged realised measure which does almost all the work at moving around the conditional variance even in the rare occasions that γ_X is estimated to be positive. There seems little point in extending the HEAVY-RM model in the same way.

2.3 Representations and dynamics

2.3.1 Multiplicative representation

The vector multiplicative representation of HEAVY models rewrites (3) and (4) as

$$\begin{pmatrix} r_t^2 \\ RM_t \end{pmatrix} = \begin{pmatrix} \varepsilon_t h_t \\ \eta_t \mu_t \end{pmatrix} = \begin{pmatrix} h_t \\ \mu_t \end{pmatrix} + \begin{pmatrix} h_t(\varepsilon_t - 1) \\ \mu_t(\eta_t - 1) \end{pmatrix}, \quad \text{where} \quad \mathrm{E}\left\{ \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} |\mathcal{F}_{t-1}^{HF} \right\} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Such representations are the key behind the work of Engle (2002) and Engle and Gallo (2006). They are powerful as $(\varepsilon_t, \eta_t)' - (1,1)'$ is a martingale difference[3] with respect to $\mathcal{F}_{t-1}^{HF}$.

The dynamic structure of the bivariate model can be gleaned from writing

$$\begin{aligned} \begin{pmatrix} h_t \\ \mu_t \end{pmatrix} &= w + \begin{pmatrix} \beta & 0 \\ 0 & \beta_R \end{pmatrix} \begin{pmatrix} h_{t-1} \\ \mu_{t-1} \end{pmatrix} + \begin{pmatrix} \alpha & 0 \\ 0 & \alpha_R \end{pmatrix} RM_{t-1}, \quad w = \begin{pmatrix} \omega \\ \omega_R \end{pmatrix}, \\ &= w + B \begin{pmatrix} h_{t-1} \\ \mu_{t-1} \end{pmatrix} + \begin{pmatrix} \alpha & 0 \\ 0 & \alpha_R \end{pmatrix} (RM_{t-1} - \mu_{t-1}), \quad B = \begin{pmatrix} \beta & \alpha \\ 0 & \alpha_R + \beta_R \end{pmatrix}. \end{aligned}$$

Hence this process is driven by a common factor $RM_t - \mu_t$, which is itself a martingale difference sequence with respect to $\mathcal{F}_{t-1}^{HF}$.

The memory in the HEAVY model is governed by

$$\begin{pmatrix} \beta & \alpha \\ 0 & \alpha_R + \beta_R \end{pmatrix}.$$

[3] A stronger set of assumptions, which is useful in inspiring a quasi-likelihood, is that jointly $(\varepsilon_t, \eta_t) \sim i.i.d.$, over the subscript t. We will not make the latter assumption unless we explicitly say so.

This has two eigenvalues (e.g. Golub and Van Loan (1989, p. 333)): β which we call a momentum parameter (a justification for this name will be given shortly) and $\alpha_R+\beta_R$ which is the persistence parameter of the realised measure. In empirical work we will typically see β to be around 0.6 and the persistence parameter being close to but slightly less than one so $\alpha_R+\beta_R$ governs the implied memory of r_t^2 at longer lags. The persistence parameter will be close to that seen for estimated $\alpha_G+\beta_G$ for GARCH models.

The role of β is interesting. In typical GARCH models the main feature is that the current value of conditional variance monotonically mean reverts to the long run average value as the forecast horizon increases. In HEAVY models this is not the case because of β.

2.3.2 Dynamics of the r_t^2 process

The HEAVY model can be solved out to imply the autocovariance function of the squared returns. This seems of little practical interests but allows some theoretical in sights.

Theorem 1 *Assume that α_R, β_R, $\beta\in[0,1)$ and $\alpha_R+\beta_R<1$. Define $u_t=r_t^2-h_t$, $u_{Rt}=RM_t-\mu_t$, which under the model is are martingale difference sequences with respect to $\mathcal{F}_{t-1}^{HF}$. Then writing L as the lag operator, we can write out the marginal process for the r_t^2 from a HEAVY model as*

$$\left\{1-\left(\alpha_R+\beta_R\right)L\right\}\left(1-\beta L\right)r_t^2=\left\{1-\left(\alpha_R+\beta_R\right)\right\}\ \omega+\alpha\ \omega_R+\xi_t, \tag{6}$$

where

$$\begin{aligned}\xi_t &= \left(1-\beta_R L\right)u_{Rt-1}+\left\{1-\left(\alpha_R+\beta_R\right)L\right\}\left(1-\beta L\right)u_t\\ &= u_t+\left\{u_{Rt-1}-\left(\alpha_R+\beta_R+\beta\right)u_{t-1}\right\}-\left\{\beta_R u_{Rt-2}+\left(\alpha_R+\beta_R\right)\beta u_{t-2}\right\}.\end{aligned}$$

If we assume that

$$\mathrm{Var}\begin{pmatrix}u_t\\ u_{Rt}\end{pmatrix}=\begin{pmatrix}\sigma_u^2 & \sigma_{u,R}\\ \sigma_{u,R} & \sigma_R^2\end{pmatrix}$$

exists then ξ_t has a zero mean weak $MA(2)$ representation and r_t^2 is weak $GARCH(2,2)$ in the sense of Drost and Nijman (1993). The autoregressive roots of r_t^2 are β and $\alpha_R+\beta_R$, so are real and positive. A biproduct of the derivation of these results is the VARMA(1,1) representation

$$\begin{pmatrix}r_t^2\\ RM_t\end{pmatrix}=\begin{pmatrix}\omega\\ \omega_R\end{pmatrix}+\begin{pmatrix}\beta & \alpha\\ 0 & \alpha_R+\beta_R\end{pmatrix}\begin{pmatrix}r_{t-1}^2\\ RM_{t-1}\end{pmatrix}+\begin{pmatrix}\left(1-\beta L\right)u_t\\ \left(1-\beta_R L\right)u_{Rt}\end{pmatrix},$$

and the equilibrium correction form (see Hendry (1995))

$$\Delta r_t^2=\omega+\alpha\left(RM_{t-1}-\gamma r_{t-1}^2\right)+\left(1-\beta L\right)u_t,\quad \textit{where}\quad \gamma=\frac{1-\beta}{\alpha}. \tag{7}$$

Proof. Clearly

$$r_t^2 = h_t + u_t = \frac{\omega}{1-\beta L} + \frac{\alpha RM_{t-1}}{1-\beta L} + u_t, \quad \text{where} \quad u_t = r_t^2 - h_t,$$

where L is the lag operator. So

$$(1-\beta L)\, r_t^2 = \omega + \alpha RM_{t-1} + (1-\beta L)\, u_t.$$

Likewise

$$\{1-(\alpha_R+\beta_R)\, L\}\, RM_t = \omega_R + (1-\beta_R L)\, u_{Rt}, \quad u_{Rt} = RM_t - \mu_t.$$

Combining delivers the result. The rest is trivial. □

An important aspect of the above result is that the memory parameters in the $MA(2)$ depend upon the covariance matrix of (u_t, u_{Rt}).

Remark 2 *The weak $GARCH(2,2)$ representation is quite like that discussed in the component model of Engle and Lee (1999, equation (3.2)) which models*

$$\begin{aligned} \text{Var}\left(r_t|\mathcal{F}_{t-1}^{LF}\right) &= \sigma_t^2 = q_t + \alpha_C\left(r_{t-1}^2 - q_{t-1}\right) + \beta_C\left(\sigma_{t-1}^2 - q_{t-1}\right), \quad \textit{where} \\ q_t &= \omega_C + \rho_C q_{t-1} + \varphi_C\left(\sigma_{t-1}^2 - h_{t-1}\right). \end{aligned}$$

The q_t process is called the long-term component and $\sigma_{t-1}^2 - q_{t-1}$ the transitory component of the conditional variance. Thus we expect ρ_C to be close to one and $\alpha_C + \beta_C$ to be substantially less than one. This model has autoregressive roots $(\alpha_C + \beta_C + \rho_C)$ and $-\rho_C(\alpha_C + \beta_C)$. These play exactly the opposite role to the ones here (their parameter ρ_C is like the HEAVY-r model's β).

Remark 3 *An importance aspect of the marginal r_t^2 process is that*

$$r_t^2 = (\alpha_R + \beta_R + \beta)\, r_{t-1}^2 - \beta(\alpha_R + \beta_R)\, r_{t-2}^2 + \{1-(\alpha_R+\beta_R)\}\ \omega + \alpha\ \omega_R + \xi_t. \tag{8}$$

This makes plain the role of β in generating momentum. It can push $\alpha_R + \beta_R + \beta$ above one heightening significant moves in the volatility while $\alpha_R + \beta_R < 1$ causes it to mean revert. If $\beta = 0$ then r_t^2 becomes a weak $GARCH(1,2)$ and has no momentum even though the realised measures still pushes the volatility around. The component model of Engle and Lee (1999, equation (3.2)) is also a weak $GARCH(1,2)$ if $\rho_C = 0$.

Remark 4 *If $\beta_R = \beta$ then*

$$\begin{aligned} \{1-(\alpha_R+\beta_R)\, L\}\,(1-\beta_R L)\, r_t^2 &= \{1-(\alpha_R+\beta_R)\}\ \omega + \alpha\ \omega_R + \xi_t, \\ \xi_t &= (1-\beta_R L)\, u_{Rt-1} + \{1-(\alpha_R+\beta_R)\, L\}\,(1-\beta_R L)\, u_t, \end{aligned}$$

so we can divide through by $(1-\beta_R L)$ *to produce*

$$\begin{aligned}\{1-(\alpha_R+\beta_R)L\}\, r_t^2 &= \frac{\{1-(\alpha_R+\beta_R)\}}{(1-\beta_R)}\,\omega + \frac{\alpha}{(1-\beta_R)}\,\omega_R + \xi_t,\\ \xi_t &= u_{Rt-1} + \{1-(\alpha_R+\beta_R)L\}\, u_t.\end{aligned}$$

Hence under that constraint the r_t^2 *is a weak GARCH(1,1) model.*

2.3.3 Integrated HEAVY models

The marginal process (8) can be rewrite in equilibrium correction form as

$$\Delta r_t^2 = -\{(1-\beta)(1-\alpha_R-\beta_R)\}\, r_{t-1}^2 + \beta(\alpha_R+\beta_R)\,\Delta r_{t-1}^2 + \{1-(\alpha_R+\beta_R)\}\;\omega + \alpha\,\omega_R + \xi_t,$$

where Δ is the difference operator. In practice the coefficients on the level and difference are likely to be slightly negative and close to β, respectively.

Clements and Hendry (1999) have argued that most economic forecasting failure is due to shifts in long run relationships and so this can be mitigated by imposing unit roots on the model. In this context this means setting $(1-\beta)(1-\alpha_R-\beta_R)$ to be zero. In order to avoid β being set to zero, this is achieved by setting $\alpha_R+\beta_R=1$, and killing the intercept ω_R (otherwise the intercept becomes a trend slope). The resulting forecasting model would then be based around

$$\Delta r_t^2 = \beta \Delta r_{t-1}^2 + \xi_t,$$

which has momentum but no mean reversion. This type of model would not be upset by structural changes in the level of the process. Imposing the unit root in GARCH type models is usually associated with the work of RiskMetrics, but that analysis does not have any momentum effects. Hence such a suggestion looks novel in the context of volatility models. It would imply using a HEAVY model of the type, for example, of

$$\begin{aligned}\mathrm{Var}\left(r_t|\mathcal{F}_{t-1}^{HF}\right) &= h_t = \omega + \alpha RM_{t-1} + \beta h_{t-1}, \quad \omega,\alpha \geq 0, \quad \beta\in[0,1), &(9)\\ \mathrm{E}\left(RM_t|\mathcal{F}_{t-1}^{HF}\right) &= \mu_t = \alpha_R RM_{t-1} + (1-\alpha_R)\,\mu_{t-1}, \quad \alpha_R\in[0,1). &(10)\end{aligned}$$

We call this the "Integrated HEAVY model". We will see later that this very simple model can generate excellent and reliable multiperiod forecasts.

Remark 5 *Suppose at time t onwards the volatility of the asset goes to exactly zero (an extreme structural break), which implies that* $r_{t+s}=0$ *and* $RM_{t+s}=0$ *for all* $s\geq 0$. *Then* $\sigma_t^2 = \omega_G + \beta_G\sigma_{t-1}^2$, *and* $h_t = \omega+\beta h_{t-1}$. *In typical empirical work both* ω_G *and* ω *are estimated to be very small. The speed of adjustment is determined by* β_G *and* β. *In empirical work we observe that* β_G *is usually above* 0.9 *while* β *is typically* 0.6. *So HEAVY models have put little weight on past data*

beyond a week; GARCH models look back around a month. This challenge for GARCH models has been recognised for some time and has prompted the development of component model by Engle and Lee (1999) (see also Christensen, Jacobs, and Wang (2008)).

2.3.4 Iterative multistep ahead forecasts

Multistep ahead forecasts of volatility are very important for asset allocation or risk assessment is usually carried out over multiple days. For one step ahead forecasts of volatility we only need (3), but for the multistep equation (4) plays a central role.

For $s \geq 0$,from the martingale difference representation, we have that

$$\begin{pmatrix} \text{Var}\left(r_{t+s}|\mathcal{F}_{t-1}^{HF}\right) \\ \text{E}\left(RM_{t+s}|\mathcal{F}_{t-1}^{HF}\right) \end{pmatrix} = \begin{pmatrix} h_{t+s|t-1} \\ \mu_{t+s|t-1} \end{pmatrix} = (I + B + ... + B^s)w + B^{s+1} \begin{pmatrix} h_{t-1} \\ \mu_{t-1} \end{pmatrix}. \tag{11}$$

Write $\vartheta = (\alpha_R + \beta_R)$. It has two roots β and $\alpha_R + \beta_R$. Further

$$B^J = \begin{pmatrix} \beta^J & \alpha\left(\vartheta^{J-1} + \vartheta^{J-2}\beta + ... + \beta^{J-1}\right) \\ 0 & \vartheta^J \end{pmatrix}, \quad J = 1, 2, 3, ...$$

Of course of interest is the integrated variance prediction $\text{Var}\left(r_t + r_{t+1} + ... + r_{t+s}|\mathcal{F}_{t-1}^{HF}\right)$. We will assume this can be simplified to

$$\text{Var}\left(r_t + r_{t+1} + ... + r_{t+s}|\mathcal{F}_{t-1}^{HF}\right) = \sum_{j=0}^{s} \text{Var}\left(r_{t+j}|\mathcal{F}_{t-1}^{HF}\right)$$

which would mean (11) could be used to compute it.

The forecasting performance of the HEAVY model can be assessed at distinct horizons by comparing the performance using the QLIK loss function

$$\text{Loss}\left(r_{t+s}^2, \widetilde{\sigma}_{t+s|t-1}^2\right) = \frac{r_{t+s}^2}{\widetilde{\sigma}_{t+s|t-1}^2} - \log\left(\frac{r_{t+s}^2}{\widetilde{\sigma}_{t+s|t-1}^2}\right) - 1, \quad s = 0, 1, ..., S, \tag{12}$$

where r_{t+s}^2 is the proxy used for the time $t+s$ (latent) variance and $\widetilde{\sigma}_{t+s|t}^2$ is some predictor made at time $t-1$. This loss function has been shown to be robust to certain types of noise in the proxy in Patton (2009) and Patton and Sheppard (2009a). It will later be used to compare the forecast performance of non-nested volatility models.

Also important is the cumulative loss function, which we take as

$$\text{Loss}\left(\sum_{j=0}^{s} r_{t+j}^2, \sum_{j=0}^{s} \widetilde{\sigma}_{t+j|t-1}^2\right) = \frac{\sum_{j=0}^{s} r_{t+j}^2}{\sum_{j=0}^{s} \widetilde{\sigma}_{t+j|t-1}^2} - \log\left(\frac{\sum_{j=0}^{s} r_{t+j}^2}{\sum_{j=0}^{s} \widetilde{\sigma}_{t+j|t-1}^2}\right) - 1, \quad s = 0, 1, ..., S,$$

which is distinct from the cumulative sum of (12). This uses the s-period realised variance as the observations.

2.3.5 Tracking reparameterisation

In the case of a stationary HEAVY model there are some advantages in reparameterising the equations in the HEAVY model so the intercepts are explicitly related to the unconditional mean of squared returns and realised measures. In the HEAVY-RM model this is easy to do as

$$\begin{aligned}\mu_t &= \omega_R + \alpha_R RM_{t-1} + \beta_R \mu_{t-1}, \quad \alpha_R, \beta_R \geq 0, \quad \alpha_R + \beta_R < 1, \\ &= \mu_R (1 - \alpha_R - \beta_R) + \alpha_R RM_{t-1} + \beta_R \mu_{t-1}, \end{aligned} \tag{13}$$

so that $\mathrm{E}(RM_t) = \mu_R$. For the HEAVY-r equation it is less clear since the realized measure is likely to be a biased downward measure of the daily squared return (due to overnight effects). Writing $\mu = \mathrm{E}(r_t^2)$ then we can set

$$\begin{aligned} h_t &= \omega_r + \alpha RM_{t-1} + \beta h_{t-1} \\ &= \mu (1 - \alpha\kappa - \beta) + \alpha RM_{t-1} + \beta h_{t-1}, \quad \kappa = \frac{\mu_R}{\mu} \leq 1. \end{aligned} \tag{14}$$

Taken together we call (14) and (13) the "tracking parameterisation" for the HEAVY model.

This parameterisation of the HEAVY model has the virtue that it is possible to use the estimators[4]

$$\widehat{\mu}_R = \frac{1}{T}\sum_{t=1}^{T} RM_t, \quad \widehat{\mu} = \frac{1}{T}\sum_{t=1}^{T} r_t^2, \quad \widehat{\kappa} = \frac{\widehat{\mu}_R}{\widehat{\mu}},$$

of μ_R, μ and κ. Thus this reparameterisation is the HEAVY extension of variance tracking introduced by Engle and Mezrich (1996).

2.4 Inference for HEAVY based models

2.4.1 Quasi-likelihood estimation

Inference for HEAVY models is a simple application of multiplicative error models discussed by Engle (2002) who uses standard quasi-likelihood asymptotic theory.

The HEAVY model has two equations

$$\begin{aligned} \mathrm{Var}\left(r_t | \mathcal{F}_{t-1}^{HF}\right) &= h_t = \omega + \alpha RM_{t-1} + \beta h_{t-1}, \\ \mathrm{E}\left(RM_t | \mathcal{F}_{t-1}^{HF}\right) &= \mu_t = \omega_R + \alpha_R RM_{t-1} + \beta_R \mu_{t-1}. \end{aligned}$$

We will estimate each equation separately, which makes optimisation straightforward. No attempt will be made to pool information across the two equations, although more information is potentially available if this was attempted (see the analysis of Cipollini, Engle, and Gallo (2007)).

[4]There may be advantages in truncating the estimator of κ to insist it is weakly less than one but we have not done that in this paper.

The first equation will be initially estimated using a Gaussian quasi-likelihood

$$\log Q_1(\omega,\psi)=\sum_{t=2}^{T} l_t^r, \quad \text{where} \quad l_t^r=-\frac{1}{2}\left(\log h_t + r_t^2/h_t\right), \quad \psi=(\alpha,\beta)', \tag{15}$$

where we take $h_1 = T^{-1/2}\sum_{t=1}^{\lfloor T\rfloor^{1/2}} r_t^2$.

The second equation will be estimated using the same structure with

$$\log Q_2(\omega_R,\psi_R)=\sum_{t=2}^{T} l_t^{RM} \quad \text{where} \quad l_t^{RM}=-\frac{1}{2}\left(\log \mu_t + RM_t/\mu_t\right), \quad \psi_R=(\alpha_R,\beta_R)', \tag{16}$$

where we take $\mu_1 = T^{-1/2}\sum_{t=1}^{\lfloor T\rfloor^{1/2}} RM_t$.

In inference we will regard the parameters as having no link between the HEAVY-r and HEAVY-RM models, i.e. (ω,ψ) and $(\ \omega_R,\psi_R)$ are variation free (e.g. Engle, Hendry, and Richard (1983)), which we will see in the next subsection is important for inference. For now we just note that it is valid to carry out optimisation equation by equation to find the maximum of the quasi-likelihoods, rather than jointly. This is convenient as existing GARCH type code can simply be used in this context (see the remarks in Engle (2002)). We will write $\theta=\left(\omega,\psi',\ \omega_R,\psi_R'\right)'$ and the resulting maximum of the quasi-likelihoods as $\widehat{\theta}$.

The alternative tracking parameterisation has

$$\begin{aligned} h_t &= \mu\left(1-\alpha\kappa-\beta\right)+\alpha RM_{t-1}+\beta h_{t-1}, \quad \kappa=\frac{\mu_R}{\mu}\leq 1, \\ \mu_t &= \mu_R\left(1-\alpha_R-\beta_R\right)+\alpha_R RM_{t-1}+\beta_R\mu_{t-1}, \quad \alpha_R+\beta_R<1, \end{aligned}$$

so that $\mathrm{E}(RM_t)=\mu_R$ and $\mathrm{E}(r_t^2)=\mu$. This has the virtue that we can employ a two-step approach, first setting

$$\widehat{\mu}=\frac{1}{T}\sum_{t=1}^{T} r_t^2 \quad \text{and} \quad \widehat{\mu}_R=\frac{1}{T}\sum_{t=1}^{T} RM_t,$$

and then we compute

$$\widehat{\psi}=\underset{\psi}{\arg\max}\log Q_1(\widehat{\mu},\widehat{\mu}_R,\psi) \quad \text{and} \quad \widehat{\psi}_R=\underset{\psi_R}{\arg\max}\log Q_2(\widehat{\mu}_R,\psi_R).$$

This has the advantage that it reduces the dimension of the optimisations by one each time, it has the disadvantage that the two equations are no longer variation-free which complicates the asymptotic distribution.

2.4.2 Quasi-likelihood based asymptotic distribution

Inference using robust standard errors is standard in this context of (15) and (16). We stack the scores so that

$$\sum_{t=2}^{T} m_t(\widehat{\theta})=0, \quad \text{where} \quad m_t(\theta)=\left(\frac{\partial l_t^r}{\partial\lambda'},\frac{\partial l_t^{RM}}{\partial\lambda_R'}\right)', \quad \lambda=\left(\omega,\psi'\right)', \quad \lambda_R=\left(\ \omega_R,\psi_R'\right)',$$

where $\theta = \left(\lambda', \lambda_R'\right)'$. Then if we denote the point in the parameter space where the model (3) and (4) holds as θ^* then under the model

$$\mathrm{E}\left\{m_t(\theta^*)|\mathcal{F}_{t-1}^{HF}\right\} = 0,$$

that is $m_t(\theta^*)$ is a martingale difference sequence with respect to $\mathcal{F}_{t-1}^{HF}$. Under standard quasi-likelihood conditions we have that

$$\sqrt{T}\left(\widehat{\theta} - \theta^*\right) \xrightarrow{d} N(0, \mathcal{J}^{-1}\mathcal{I}\mathcal{J}^{-1\prime}),$$

where the Hessian is

$$\mathcal{J} = p\lim_{T\to\infty} \widehat{\mathcal{J}}_T, \quad \text{where} \quad \widehat{\mathcal{J}}_T = -\frac{1}{T}\begin{pmatrix} \sum_{t=2}^{T} \frac{\partial^2 l_t^r}{\partial\lambda\partial\lambda'} & 0 \\ 0 & \sum_{t=2}^{T} \frac{\partial^2 l_t^{RM}}{\partial\lambda_R\partial\lambda_R'} \end{pmatrix}, \tag{17}$$

and

$$\mathcal{I} = p\lim_{T\to\infty} \widehat{\mathcal{I}}_T, \quad \text{where} \quad \widehat{\mathcal{I}}_T = \frac{1}{T}\sum_{t=2}^{T} m_t(\widehat{\theta})m_t(\widehat{\theta})'. \tag{18}$$

The block diagonality of (17) is due to the variation freeness of the blocks, while it is not necessary to use a HAC estimator in (18) due to the martingale difference features of the stacked scores. This is a straightforward application of quasi-likelihood theory and can be viewed as an extension of Bollerslev and Wooldridge (1992) and is already discussed extensively in Cipollini, Engle, and Gallo (2007).

The most important implication of the block diagonality of the Hessian (17) is that the equation by equation standard errors for the HEAVY-r and HEAVY-RM are correct, even when viewing the HEAVY model as a system. This means that standard software can be used to compute them.

When the two step approach is used on the tracking parameterisation then the moment conditions change to

$$m_t(\theta_E) = \left\{\frac{1}{T}\left(r_t - \mu\right), \frac{\partial l_t^r}{\partial\psi'}, \frac{1}{T}\left(MR_t - \mu_R\right), \frac{\partial l_t^{RM}}{\partial\psi_R'}\right\}', \quad \theta_E = \left(\mu, \psi', \mu_R, \psi_R'\right)'.$$

The moment conditions are no longer martingale difference sequences, but they do have a zero mean for all values of t at the true parameter point

$$\widehat{\mathcal{J}}_T = -\frac{1}{T}\begin{pmatrix} T & \sum_{t=2}^{T} \frac{\partial^2 l_t^r}{\partial\mu\partial\psi'} & \sum_{t=2}^{T} \frac{\partial^2 l_t^r}{\partial\mu_R\partial\psi'} & 0 \\ 0 & \sum_{t=2}^{T} \frac{\partial^2 l_t^r}{\partial\psi\partial\psi'} & 0 & 0 \\ 0 & 0 & T & \sum_{t=2}^{T} \frac{\partial^2 l_t^{RM}}{\partial\mu_R\partial\psi_R'} \\ 0 & 0 & 0 & \sum_{t=2}^{T} \frac{\partial^2 l_t^{RM}}{\partial\psi_R\partial\psi_R'} \end{pmatrix},$$

while $\widehat{\mathcal{I}}_T$ needs to be a HAC estimator applied to the time series of $m_t(\theta_E)$.

2.4.3 Non-tested tests

One natural way to assess the forecasting power of the HEAVY model is to compare it to that generated by the GARCH model. We will do this by comparing QLIK loss, following the discussion given in Section 2.3.4.

The temporal average $(s+1)$-step ahead relative loss will be

$$\widehat{L}_s = \frac{1}{T-s}\sum_{t=s+1}^{T} L_{t,s}, \quad s = 0, 1, ..., S,$$

where

$$\begin{aligned} L_{t,s} &= \text{Loss}\left(r_{t+s}^2, h_{t+s|t-1}\right) - \text{Loss}\left(r_{t+s}^2, \sigma_{t+s|t-1}^2\right), \quad s = 0, 1, ..., S \\ &= \left\{\frac{r_{t+s}^2}{h_{t+s|t-1}} + \ln\left(h_{t+s|t-1}\right)\right\} - \left\{\frac{r_{t+s}^2}{\sigma_{t+s|t-1}^2} + \ln\left(\sigma_{t+s|t-1}^2\right)\right\} \\ &= -2\log\frac{f(r_{t+s}|0, h_{t+s|t-1})}{f(r_{t+s}|0, \sigma_{t+s|t-1}^2)}. \end{aligned}$$

Here $h_{t+s|t-1}$ is the forecast from the HEAVY model, $\sigma_{t+s|t}^2$ is the corresponding GARCH forecast and $f(x|\mu, \sigma^2)$ denotes a Gaussian density with mean μ and variance σ^2, evaluated at x. The framework will allow both the HEAVY and GARCH model to be estimated using QML techniques. The HEAVY model will be favoured if $\widehat{L}_s$ is negative.

$\widehat{L}_s$ estimates $L_s = \text{E}(L_{t,s})$, $s = 0, 1, ..., S$, for each s, the unconditional average likelihood ratio between the two models. The HEAVY model will be favoured at s-steps if $L_s < 0$ and the GARCH model if $L_s > 0$. We will say that the HEAVY model forecast-dominates the GARCH model if $L_s < 0$ for all $s = 1, 2, ..., S$. Weakly forecast-dominates means that $L_s \leq 0$ for all $s = 1, 2, ..., S$ with at least one of the $\leq$ relationships being a strict inequality. This approach follows the ideas of Cox (1961b) on non-nested testing using the Vuong (1989) and Rivers and Vuong (2002) implementation[5].

The above scheme can be implemented simply if $L_{t,s}$ (evaluated at their pseudo-true parameter values) is sufficiently weakly dependent to allow the parameter estimates of the HEAVY and GARCH models to obey a standard Gaussian central limit theorem (e.g. Rivers and Vuong (2002)). Then

$$\sqrt{T}\left(\widehat{L}_s - L_s\right) \xrightarrow{d} N(0, V_s),$$

where V_s is the long-run variance of the $L_{t,s}$. The scale V_s has to be estimated by a HAC estimator (e.g. Andrews (1991)).

[5] In the context of forecasting this is related to Diebold and Mariano (1995). As well as an elegant implementation, Vuong (1989) has the virtue of being valid even if neither model is correct. It just assesses which is better in terms of the unconditional average likelihood ratio.

2.4.4 Horizon tuned estimation and evaluation

Having multistep ahead loss functions suggests separately estimating the model at each forecast horizon by minimising expected loss at that horizon. This way of tuning the model to produce multistep ahead forecasts is called "direct forecasting" and has been studied by, for example, Marcellino, Stock, and Watson (2006) and Ghysels, Rubia, and Valkanov (2009). The former argue direct forecasting may be more robust to model misspecification than iterating one-period ahead models, although they find iterative methods more effective in forecasting for macroeconomic variables in practice. Direct forecasting dates at least to Cox (1961a). Marcellino, Stock, and Watson (2006) provide an extensive discussion of the literature.

Minimising the QLIK multistep ahead loss can be thought of as maximising a distinct quasi-likelihood for each value of s

$$\begin{aligned}\log Q_{1,s}(\omega_s,\psi_s) &= \sum_{t=2}^{T} l_{t,s}^{r}, \quad \text{where} \quad l_{t,s}^{r} = -\frac{1}{2}\left(\log h_{t+s|t-1} + \frac{r_{t+s}^2}{h_{t+s|t-1}}\right), \psi_s = (\alpha_s,\beta_s)', \\ \log Q_{2,s}(\omega_{R,s},\psi_{R,s}) &= \sum_{t=2}^{T} l_{t,s}^{RM} \quad \text{where} \quad l_{t,s}^{RM} = -\frac{1}{2}\left(\log \mu_{t+s|t-1} + \frac{RM_{t+s}}{\mu_{t+s|t-1}}\right), \psi_{R,s} = (\alpha_{R,s},\beta_{R,s})',\end{aligned}$$

where the quasi-likelihood is the Gaussian likelihood based on multistep ahead forecasts. This delivers the sequence of horizon tuned estimators $\widehat{\omega}_s$, $\widehat{\psi}_s$, $\widehat{\omega}_{R,s}$, $\widehat{\psi}_{R,s}$, whose standard errors can be computed using the usual theory of quasi-likelihoods. In practice, because of the structure of our HEAVY model, by far the most important of these equations is the second one, which allows horizon tuning for the HEAVY-RM forecasts[6]. The same exercise can be carried out for a GARCH model.

2.4.5 Bootstrapping

Like GARCH models, a significant drawback of HEAVY models is that they only specify the conditional means of r_t^2 and RM_t given $\mathcal{F}_{t-1}^{HF}$. It is sometimes helpful to produce the entire forecast distributions

$$F(r_{t+s}|\mathcal{F}_{t-1}^{HF}), \quad s = 0,1,2,..., \tag{19}$$

or

$$F\left(r_t + r_{t+1} + ... + r_{t+s}|\mathcal{F}_{t-1}^{HF}\right). \tag{20}$$

A simple way of carrying this out is via a model based bootstrap. To do this we use the representation $r_t = \zeta_t h_t^{1/2}$, $RM_t = \eta_t \mu_t$, $\mathrm{E}(\zeta_t^2|\mathcal{F}_{t-1}^{HF}) = 1$, $\mathrm{E}(\eta_t|\mathcal{F}_{t-1}^{HF}) = 1$ and then assume that

[6]If we condition on the lagged realised measure the additional memory in the HEAVY-r model is modest.

$(\zeta_t, \eta_t)' \overset{i.i.d.}{\sim} F_{\zeta,\eta}$. Typically these bivariate variables will be contemporaneously correlated. In particular for equities we would expect a sharp negative correlation reflecting the statistical leverage effect. If we had knowledge of $F_{\zeta,\eta}$ it would be a trivial task to carry out model based simulation from (19) or (20).

We can estimate the joint distribution function $F_{\zeta,\eta}$ by simply taking the filtered $(h_t, \mu_t)'$ and computing the devolatilised[7]

$$\widehat{\zeta}_t = r_t/h_t^{1/2}, \quad \widehat{\eta}_t = \left(RM_t/\mu_t\right)^{1/2}, \quad t = 2, 3, ..., T, \tag{21}$$

and computing the empirical distribution function $\widehat{F}_{\zeta,\eta}$. It is a simple matter to sample with replacement pairs from this population[8], which can then be used to drive a simulated joint path of the pair $(r_t, RM_t)'$, $(r_{t+1}, RM_{t+1})'$, ..., $(r_{t+s}, RM_{t+s})'$. By discarding the drawn realised measures gives us paths of daily returns r_t, r_{t+1}, ..., r_{t+s}. Carrying out this simulation many times approximates the predictive distributions.

2.5 Statistical leverage effect

We can parametrically model statistical leverage effects, where falls in asset prices are associated with increases in future volatility, by adding a new equation for a realised semivariance (RM_t^*). Realised semivariances were introduced by Barndorff-Nielsen, Kinnebrouck, and Shephard (2009) and further emphasised in empirical work by Patton and Sheppard (2009b). Now our model becomes

$$\begin{aligned}
\text{Var}\left(r_t|\mathcal{F}_{t-1}^{HF}\right) &= h_t = \omega + \alpha RM_{t-1} + \alpha^* RM_{t-1}^* + \beta h_{t-1}, \quad \alpha^* \geq 0, \\
\text{E}\left(RM_t|\mathcal{F}_{t-1}^{HF}\right) &= \mu_t = \omega_R + \alpha_R RM_{t-1} + \beta_R \mu_{t-1}, \\
\text{E}\left(RM_t^*|\mathcal{F}_{t-1}^{HF}\right) &= \mu_t^* = \omega_R^* + \alpha_R^* RM_{t-1}^* + \beta_R^* \mu_{t-1}^*, \quad \alpha_R^*, \beta_R^* \geq 0, \quad \alpha_R^* + \beta_R^* < 1.
\end{aligned}$$

The expansion of the model to allow for the appearance of realised semivariances raises no new issues (allowing lags of RM_t^* to appear in the dynamic of RM_t could potentially help to, but we will not discuss that here).

The paper by Engle and Gallo (2006) suggests an alternative approach. Let $i_t = 1_{r_t<0}$ then they extend models by interacting i_t with volatility measures, following the tradition of the GARCH literature. If one does this to the HEAVY model it becomes

$$\text{Var}\left(r_t|\mathcal{F}_{t-1}^{HF}\right) = h_t = \omega + \alpha RM_{t-1} + \alpha^* i_{t-1} RM_{t-1} + \beta h_{t-1}, \quad \alpha^* \geq 0, \tag{22}$$

[7]We work with the $RM_t^{1/2}$, rather than the original RM_t as volatilities (as opposed to variance type objects) are easier to interprete later, but this choice has little impact here and the same exercise could be carried out based on the RM_t.

[8]There maybe some advantages in using a block sampling scheme for the innovations (ζ_t, η_t) as they are not expected to be exactly temporally independent, although they should be temporally uncorrelated. But we have not explored that here.

$$\mathrm{E}\left(RM_t|\mathcal{F}_{t-1}^{HF}\right) \quad = \quad \mu_t = \ \omega_R + \alpha_R RM_{t-1} + \alpha i_{t-1} RM_{t-1} + \beta_R \mu_{t-1}, \quad \alpha_R^* \geq 0.$$

This model is easy to estimate, for i_{t-1} is in $\mathcal{F}_{t-1}^{HF}$. However, to make two step ahead forecasts we run into trouble for we do not know $i_t RM_t$ or have a forecast of it.

One approach to this is to assume that

$$i_{t+h} \perp\!\!\!\perp RM_{t+h}|\mathcal{F}_{t-1}^{HF}, \quad h = 0, 1, 2, ...$$

where $A \perp\!\!\!\perp B$ denotes A and B are statistically independent. This would imply

$$\mathrm{E}(i_{t+h} RM_{t+h}|\mathcal{F}_{t-1}^{HF}) = \mathrm{E}(i_{t+h}|\mathcal{F}_{t-1}^{HF})\mathrm{E}(RM_{t+h}|\mathcal{F}_{t-1}^{HF}).$$

Typically we would assume that $\mathrm{E}(i_{t+h}|\mathcal{F}_{t-1}^{HF}) = \mathrm{E}(i_{t+h})$, which is likely to be very close to 1/2. This would allow multistep ahead forecasts to be computed analytically and straightforwardly.

Perhaps more wisely we could use a bootstrap to simulate the empirical distribution of $\widehat{\zeta}_t, \widehat{\eta}_t$ from (21) and this allows to simulate through (22). This method of dealing with statistical leverage has the virtue is that it also delivers an estimator of the multistep ahead prediction distribution, and so may reveal the long left hand tail of the asset prices often induced by statistical leverage even though $\widehat{\zeta}_t$ is marginally relatively symmetric.

3 OMI's realised measure library 0.1

3.1 A list of assets and data cleaning

In this section we will discuss the database used in this paper. The database is called the "OMI's realised measure library" version 0.1 and has been produced by Heber, Lunde, Shephard, and Sheppard (2009)[9].

The version 0.1 of the library currently starts on the 2nd January 1996 and finishes 27th March 2009. Some of the series are available throughout this period, but quite a number start after 1996, as detailed in Table 1. In total the database covers 34 different assets. Some of these series are indexes computed by MSCI. Others are traded assets or indexes computed by other data providers computed in real time. For each asset we give in Table 1 the basic features of the data used to compute the library, indicating the frequency of the base data used in the calculations of realised measures.

For each asset the library currently records daily returns, daily subsampled realised variances and daily realised kernels. In this paper we use the daily returns and realised kernels in our modelling. If the market is closed or the data is regarded as being of unacceptably low quality

[9]The library cannot be used for commerical purposes without the written permission of the Oxford-Man Institute, but it can be used for academic research as long as it is quoted appropriately.

Asset	Med dur	Start Dates	T	Asset	Med dur	Start Dates	T
Dow Jones Industrials	2	2-1-1996	3,278	MSCI Australia	60	2-12-1999	2,323
Nasdaq 100	15	2-1-1996	3,279	MSCI Belgium	60	1-7-1999	2,442
S&P 400 Midcap	15	2-1-1996	3,275	MSCI Brazil	60	4-10-2002	1,587
S&P 500	15	2-1-1996	3,284	MSCI Canada	60	12-2-2001	2,013
Russell 3000	15	2-1-1996	3,279	MSCI Switzerland	60	9-6-1999	2,434
Russell 1000	15	2-1-1996	3,279	MSCI Germany	60	1-7-1999	2,448
Russell 2000	15	2-1-1996	3,281	MSCI Spain	60	1-7-1999	2,423
CAC 40	30	2-1-1996	3,322	MSCI France	60	1-7-1999	2,455
FTSE 100	15	20-10-1997	2,862	MSCI UK	60	8-6-1999	2,451
German DAX	15	2-1-1996	3,317	MSCI Italy	60	1-7-1999	2,437
Italian MIBTEL	60	3-7-2000	2,194	MSCI Japan	15	2-12-1999	2,240
Milan MIB 30	60	2-1-1996	3,310	MSCI South Korea	60	3-12-1999	2,263
Nikkei 250	60	5-1-1996	3,177	MSCI Mexico	60	4-10-2002	1,612
Spanish IBEX	5	2-1-1996	3,288	MSCI Netherlands	60	1-7-1999	2,454
S&P TSE	15	31-12-1998	2,546	MSCI World	60	11-2-2001	2,101
British Pound	2	3-1-1999	2,584				
Euro	1	3-1-1999	2,600				
Swiss Franc	3	3-1-1999	2,579				
Japanese Yen	2	3-1-1999	2,599				

Table 1: A description of the "OMI's realised measures library," version 0.1. The Table shows how each measure is built and the length of time series available denoted T. "Med dur" denotes the median duration in seconds between price updates during September 2008 in our database. All data series stop on 27th March 2009.

for that asset then the database records it as missing, except for days when all the markets are simultaneously closed in which case the day is not recorded in the database. An example of this is that Saturdays are never present in the library. Summary features of the library will be discussed in the next subsection.

Realised variances (1) are computed by first calculating 5 minute returns (using the last tick method) and subsampling this statistic using every 30 seconds[10]. Realised kernels are computed in tick time using every available data point, after cleaning.

The library is based on underlying high frequency data, which we obtain through Reuters. We are not in a position to make available this base data, or its cleaned version, for commercial reasons as Reuters owns the copyright to it. Although the raw data is of high quality it does need to be cleaned so it is suitable for econometric inference. Cleaning is an important aspect of computing realised measures. Although realised kernels are somewhat robust to noise, experience suggests that when there are misrecordings of prices or hit large amounts of turbulence at the start of a trading day then they may sometimes give false signals. Barndorff-Nielsen, Hansen, Lunde, and Shephard (2009) have studied systematically the effect of cleaning on realised kernels, using cleaning methods which build on those documented by Falkenberry (2002) and Brownlees and Gallo (2006). Our data has more variation in structure than that dealt with in Barndorff-Nielsen, Hansen, Lunde,

[10]For our MSCI index data we only have raw returns at the 1 minute level, which meant that when we subsampled at the 30 second level we produce the same RV twice (this has no impact as we divide everything by two).

Asset	r_t^2			RV			KV		
	Avol	sd	acf_1	Avol	sd	acf_1	Avol	sd	acf_1
Dow Jones Industrials	19.4	4.81	.125	15.2	1.94	.663	15.0	1.95	.655
Nasdaq 100	28.1	8.35	.180	17.8	2.22	.664	18.7	2.52	.646
S&P 400 Midcap	21.7	5.68	.260	13.5	1.90	.800	13.7	1.96	.799
S&P 500	20.8	5.46	.209	15.5	2.09	.699	15.9	2.14	.701
Russell 3000	20.3	5.32	.127	14.3	1.86	.694	14.5	1.90	.697
Russell 1000	20.4	5.38	.125	14.7	1.91	.692	14.9	1.94	.695
Russell 2000	23.3	6.02	.313	13.2	1.85	.715	13.4	1.96	.720
CAC 40	23.7	5.95	.236	18.1	2.18	.662	18.3	2.21	.669
FTSE 100	20.7	4.66	.229	15.2	1.62	.645	15.6	1.74	.620
German DAX	25.1	6.57	.163	21.1	3.10	.659	21.3	3.22	.626
Italian MIBTEL	20.1	5.07	.218	13.1	1.34	.665	13.7	1.52	.662
Milan MIB 30	23.2	5.69	.214	16.5	1.84	.624	17.0	1.99	.615
Nikkei 250	24.9	6.96	.241	16.0	1.37	.691	16.5	1.48	.668
Spanish IBEX	23.7	6.57	.295	16.7	1.76	.639	16.5	1.73	.655
S&P TSE	20.9	5.54	.292	14.1	1.82	.785	14.3	1.89	.774
MSCI Australia	16.4	3.05	.229	8.8	0.53	.763	9.1	0.57	.749
MSCI Belgium	23.4	10.5	.159	16.4	1.66	.718	16.1	1.84	.684
MSCI Brazil	43.7	24.3	.155	28.5	6.30	.796	29.6	7.21	.749
MSCI Canada	19.5	5.05	.320	12.6	1.67	.819	13.1	1.88	.761
MSCI Switzerland	20.6	5.25	.330	14.5	1.44	.727	14.5	1.56	.700
MSCI Germany	25.7	6.94	.163	21.1	3.10	.677	20.8	2.99	.692
MSCI Spain	24.0	6.08	.225	17.5	1.84	.690	17.6	1.92	.676
MSCI France	23.9	6.29	.238	18.2	2.23	.682	18.4	2.32	.669
MSCI UK	20.0	4.95	.233	15.6	1.84	.615	15.7	1.89	.649
MSCI Italy	21.4	5.35	.247	16.0	1.82	.672	16.2	1.93	.670
MSCI Japan	23.7	6.40	.273	14.2	1.27	.746	14.4	1.26	.755
MSCI South Korea	32.0	9.63	.131	21.6	2.61	.700	21.9	2.80	.682
MSCI Mexico	29.6	11.8	.144	16.3	2.59	.675	17.5	2.87	.678
MSCI Netherlands	23.9	6.14	.281	17.7	2.09	.733	17.9	2.25	.716
MSCI World	17.7	4.22	.250	13.1	1.44	.766	13.6	1.68	.691
British Pound	9.2	0.75	.215	9.8	0.51	.876	9.4	0.51	.879
Euro	10.4	0.79	.103	11.1	0.45	.668	10.5	0.45	.658
Swiss Franc	11.0	0.91	.133	11.6	0.39	.690	10.8	0.38	.650
Japanese Yen	10.9	1.32	.134	11.6	0.64	.698	11.2	0.63	.696

Table 2: Calculations use 100 times differences of the log price (ie roughly % changes). Avol is the square root of the mean of 252 times either squared returns or the realised measure. It is the approximate annualised volatility. The sd is the daily standard deviation of % daily returns or realised measure. Same data is used to compute the acfs (serial correlations) at 1 lag.

and Shephard (2009) and so we discuss how our methods use their rules.

Most of the datasets we use are based on indexes, which are updated at distinct frequencies. Some indexes, such as the DAX and Dow Jones index, are updated every second or a couple of seconds. Most are updated every 15 or 60 seconds. The only data cleaning we applied to this was that applied to all datasets, called P1, given below.

All data

P1. Delete entries with a time stamp outside the interval when the exchange is open.

Quote data for the exchange rates is very plentiful and has the virtue of having no market closures. We use four rules for this, given below as Q1-Q4. Q1 is by far the most commonly

used.

Quote data only

Q1. When multiple quotes have the same timestamp, we replace all these with a single entry with the median bid and median ask price.

Q2. Delete entries for which the spread is negative.

Q3. Delete entries for which the spread is more that 50 times the median spread on that day.

Q4. Delete entries for which the mid-quote deviated by more than 10 mean absolute deviations from a rolling centered median (excluding the observation under consideration) of 50 observations (25 observations before and 25 after).

In addition we have made various manual edits in the library when the results we unsatisfactory. Some of these were due to rebasing of indexes, which had their biggest effects on daily returns. It is the hope of the editors of the library that as it develops then the degree of manual edits will decline.

3.2 Summary statistics for the library

Table 2 gives summary statistics for the realised measures and squared daily returns for each asset. The Table is split into three sections, which are raw indexes, MSCI indexes and exchange rates all quoted against the US Dollar.

The Avol number takes either squared returns or the realised measure and multiplies them by 252 and then averages the value over the sample period. We then square root the result and report it. This is so that the Avol number is on the scale of an annualised volatility, which is familiar in financial economics. It shows the raw common indexes have annualised volatility for returns of usually just over 20%, with the corresponding results for the realised variance measures typically being around 16% and the realised kernels around the same level. Of course the realised measures miss out on the overnight return, which accounts for their lower level. The MSCI indexes have more variation in their Avol levels, sometimes going into the 30s and in one case the 40s. The overnight effects are large again. In the exchange rate case the Avols are lower for squared returns and in this case the realised measures have roughly the same average level — presumably as there is no overnight effect. The Avol for realised kernels is typically a little higher than for the realised variance, but the difference is very small.

The sd figures are standard deviations of percentage daily squared movements or realised measures, not scaled to present annualised quantities (as this would make them inelegantly large).

They show much higher standard deviations for squared returns than for their realised measure cousins. This is the expected result.

The acf figures are the serial correlation coefficients at one lag. It shows the modest degree of serial correlation of squared returns and much higher numbers of the realised variances and realised kernels. This is the result we expected from the econometric literature on realised measures.

4 Empirical analysis with a large cross-section

4.1 Estimated models

In this section we will take each univariate series of returns and realised measures and fit a HEAVY model together with the tracking GARCH

$$\sigma_t^2 = \mu_G (1 - \alpha_G - \beta_G) + \alpha_G r_{t-1}^2 + \alpha_G \sigma_{t-1}^2,$$

and the non-tracking GARCHX models. The HEAVY models are setup in their tracking parameterisation

$$\begin{aligned} \mu_t &= \mu_R (1 - \alpha_R - \beta_R) + \alpha_R RM_{t-1} + \beta_R \mu_{t-1}, \quad \alpha_R + \beta_R < 1, \\ h_t &= \mu (1 - \alpha\kappa - \beta) + \alpha RM_{t-1} + \beta h_{t-1}, \quad \kappa = \frac{\mu_R}{\mu} \leq 1, \quad \alpha + \beta < 1. \end{aligned}$$

In the GARCH and HEAVY cases they are estimated using a two step approach, using unconditional empirical moments for μ_G, μ_R and μ and then maximising the Gaussian quasi-likelihoods for (α_G, β_G), (α_R, β_R) and (α, β). The same estimation strategy is used for the GARCH model, but for the GARCHX model optimisation of the quasi-likelihood is used for all the parameters in the model.

For multistep ahead forecasts there are some arguments which favour imposing a unit root on the HEAVY-RM model, in which case we model

$$\begin{aligned} \mu_t &= \alpha_R RM_{t-1} + (1 - \alpha_R) \mu_{t-1}, \quad \alpha_R < 1, \\ h_t &= \omega + \alpha RM_{t-1} + \beta h_{t-1}, \quad \alpha + \beta < 1, \end{aligned} \tag{23}$$

which means it has no tracking features at all. It would seem illogical to want to impose tracking on HEAVY-r at the same time as using an integrated model for realised measures.

The results are presented in some detail in Table 3 for the dynamic parameters. In the HEAVY-r model the momentum parameter β is typically in the range from 0.6 to 0.75, but there are exceptions which are typically exchange rates where there is very considerable memory. The HEAVY-RM models show a very large degree of persistence in the series with α_R being typically in the region of 0.35 to 0.45, and $\alpha_R + \beta_R$ being close to one. For currencies using realised measures improves the fit of the model, but the improvement is modest.

Asset	HEAVY-r		GARCHX			GARCH		HEAVY-RM		Integrated	
	α	β	α_X	β_X	γ_X	α_G	β_G	α_R	β_R	α_{IG}	α_{IR}
Dow Jones Industrials	.407	.737	.407	.737	.000	.082	.912	.411	.567	.062	.336
Nasdaq 100	.730	.658	.439	.744	.051	.081	.916	.428	.567	.063	.349
S&P 400 Midcap	.848	.641	.270	.794	.083	.100	.886	.392	.603	.073	.333
S&P 500	.378	.773	.378	.773	.000	.076	.918	.417	.564	.054	.340
Russell 3000	.448	.747	.448	.747	.000	.081	.911	.403	.574	.059	.313
Russell 1000	.397	.768	.397	.768	.000	.078	.916	.402	.577	.057	.315
Russell 2000	.949	.678	.244	.812	.102	.106	.885	.387	.622	.077	.322
CAC 40	.526	.674	.526	.674	.000	.081	.917	.417	.573	.067	.350
FTSE 100	.613	.656	.613	.656	.000	.105	.892	.441	.556	.085	.369
German DAX	.447	.673	.447	.673	.000	.093	.903	.457	.536	.075	.376
Italian MIBTEL	.806	.630	.806	.630	.000	.107	.889	.512	.486	.080	.436
Milan MIB 30	.496	.748	.342	.779	.047	.102	.895	.484	.518	.075	.417
Nikkei 250	.508	.772	.508	.772	.000	.079	.905	.346	.641	.065	.295
Spanish IBEX	.640	.669	.481	.713	.035	.113	.885	.393	.603	.084	.343
S&P TSE	.643	.692	.637	.693	.002	.067	.930	.362	.635	.054	.324
Index's median	.526	.678	.447	.744	.000	.082	.905	.411	.573	.067	.340
MSCI Australia	.214	.645	.976	.668	.043	.098	.894	.324	.670	.069	.292
MSCI Belgium	.769	.568	.374	.692	.093	.143	.854	.399	.608	.105	.359
MSCI Brazil	.662	.652	.661	.653	.001	.096	.876	.433	.536	.071	.375
MSCI Canada	.515	.765	.485	.769	.009	.074	.914	.364	.630	.060	.329
MSCI Switzerland	.699	.638	.699	.638	.000	.131	.860	.474	.508	.093	.425
MSCI Germany	.568	.592	.568	.592	.000	.107	.885	.461	.529	.083	.388
MSCI Spain	.589	.659	.589	.659	.000	.090	.907	.417	.579	.067	.365
MSCI France	.596	.628	.596	.628	.000	.090	.908	.453	.543	.074	.386
MSCI UK	.582	.616	.582	.616	.000	.110	.886	.456	.543	.086	.393
MSCI Italy	.583	.659	.583	.659	.000	.100	.896	.537	.462	.075	.467
MSCI Japan	.741	.720	.741	.720	.000	.088	.902	.459	.533	.075	.387
MSCI South Korea	.765	.661	.765	.661	.000	.071	.928	.432	.564	.059	.392
MSCI Mexico	.872	.711	.723	.725	.032	.095	.885	.364	.624	.068	.328
MSCI Netherlands	.538	.678	.538	.678	.000	.105	.889	.453	.541	.084	.396
MSCI World	.339	.798	.339	.798	.000	.084	.910	.377	.610	.068	.340
MSCI's median	.596	.659	.589	.661	.000	.096	.894	.433	.543	.074	.386
British Pound	.162	.810	.162	.810	.000	.042	.950	.283	.699	.035	.264
Euro	.055	.936	.034	.947	.013	.030	.969	.247	.746	.028	.223
Swiss Franc	.046	.948	.045	.947	.002	.027	.971	.239	.748	.024	.220
Japanese Yen	.173	.772	.173	.772	.000	.048	.934	.398	.552	.035	.341
Currency's median	.109	.873	.104	.879	.001	.036	.959	.265	.722	.031	.244

Table 3: Fit of GARCH and HEAVY models for various indexes and exchange rates. The cross-sectional median takes the median of the parameter estimates for the indexes. GARCH and HEAVY-RM models are esimated using the tracking parameterisation. Integrated models are IGARCH and Int-HEAVY-RM.

When we allow for realised measures in the GARCH model, that is we specify the GARCHX model, typically the γ_X parameter is estimated to be on its boundary at exactly zero. There are eight exceptions to this, but the use of robust standard errors (not reported here) suggest only a couple of them are statistically significant. These two are the S&P 400 Midcap and Russell 2000. In those cases the realised kernel may not have dealt correctly with the dependence in their high frequency data induced by the staleness of the prices for some of the components of the index.

	Compare to GARCHX		Impose unit root			Compare to GARCHX		Impose unit root	
Asset	HEAVY-r	GARCH	GARCH	HEAVY-RM	MSCI Index	HEAVY-r	GARCH	GARCH	HEAVY-RM
DJI	0.0	-199.5	-48.4	-19.5	Australia	-6.6	-96.6	-31.2	-3.9
Nasdaq 100	-15.9	-108.5	-31.1	-14.4	Belgium	-22.7	-66.2	-60.2	-4.1
S&P 400 Midcap	-64.6	-61.8	-61.4	-11.0	Brazil	0.0	-60.2	-35.5	-7.1
S&P 500	0.0	-211.1	-50.6	-17.9	Canada	-0.4	-75.0	-22.9	-4.4
Russell 3000	0.0	-187.3	-49.8	-21.1	Switzerland	0.0	-153.4	-65.8	-9.1
Russell 1000	0.0	-186.3	-45.3	-20.0	Germany	0.0	-136.9	-45.0	-10.
Russell 2000	-163.2	-64.9	-57.4	-13.3	Spain	0.0	-106.7	-31.5	-7.5
CAC 40	0.0	-149.1	-30.8	-14.5	France	0.0	-158.3	-27.7	-9.4
FTSE 100	0.0	-125.5	-32.4	-12.3	UK	0.0	-134.3	-37.1	-9.3
German DAX	0.0	-153.4	-47.0	-16.0	Italy	0.0	-154.7	-38.3	-8.7
Italian MIBTEL	0.0	-141.2	-40.5	-9.9	Japan	0.0	-111.8	-33.7	-6.2
Milan MIB 30	-16.5	-100.7	-48.3	-13.0	South Korea	0.0	-118.6	-15.1	-4.1
Nikkei 250	0.0	-116.5	-64.5	-9.9	Mexico	-3.4	-61.2	-36.5	-3.5
Spanish IBEX	-9.3	-113.9	-59.0	-12.1	Netherlands	0.0	-117.8	-40.8	-7.6
S&P TSE	-0.0	-120.8	-17.3	-5.6	World	0.0	-92.9	-25.6	-6.3
Index's median	0.0	-125.5	-48.3	-13.3	Median	0.0	-111.8	-35.5	-7.1
British Pound	0.0	-50.4	-16.0	-1.8					
Euro	-2.7	-18.5	-6.0	-1.6					
Swiss Franc	-0.1	-33.0	-5.9	-1.7					
Japanese Yen	0.0	-67.4	-38.6	-8.4					
Currency's median	0.0	-41.7	-11.0	-1.8					

Table 4: Twice the likelihood change by imposing restrictions on the model. Left hand side shows twice the the likelihood change compared to GARCHX model. The right hand side compares the unconstrained GARCH and HEAVY-RM models with those which impose a unit root.

Also given in the Table is the median of the estimators for three blocks of the assets which provides a guide to the typical behaviour.

Finally, the Table also records the estimate value of α_R for the integrated HEAVY model. This does not change very much from the estimated HEAVY model, but there are small drops in the estimates are typical.

Table 4 shows the change in the log-likelihood function by moving to the HEAVY-r and GARCH models from the nesting GARCHX model. In the GARCH case the changes are always very large, in the HEAVY-r case the changes are usually zero. However, there are a couple of cases where the reduction in likelihood is quite large. The Table also shows the impact on the likelihood by imposing unit roots on the GARCH and HEAVY-RM models. The effect on the HEAVY-RM model is more modest than in the GARCH case.

Table 5 shows the HEAVY's model's average in sample iterated multistep ahead QLIK loss compared to the GARCH model, using the methodology discussed in Section 2.3.4. Here the parameters are estimated using the Gaussian quasi-likelihood, which means they are tuned to perform best at one-step ahead forecasting. The forecast horizon varies over 1, 2, 3, 5, 10 and 22 lags. Two models are fitted. The left hand side shows the result for the standard HEAVY model

which is estimated using a tracking parameterisation. The right hand side shows the corresponding result for the "integrated HEAVY" model, which is discussed in (23). Recall that negative values t-statistics indicate a statistically significant preference for HEAVY models.

The results are striking. It shows that in sample and pointwise the standard HEAVY model forecast-dominates the GARCH model, but that the outperformance gets weaker as the forecast horizon increases. At short lags the integrated HEAVY model performs more poorly than the unconstrained HEAVY model, but its forecast performance at higher lags is much better than GARCH and the degree of outperformance does not get worse as the forecast horizon increases. In other words, in terms of longer term forecasting, the HEAVY model benefits from imposing a unit root on the dynamic for the realised measures.

This picture is remarkably stable across assets with two counter examples. There are two cases where there was worse performance and that was the mid-cap series Russell 2000 and the S&P 400 Midcap. These have lower quasi-likelihoods and this underperformance continues when applied at multistep ahead periods.

4.2 Direct forecasting

The above estimation strategy fixes the parameters at the MQLE values and uses these to iterate through the multistep ahead forecast formula to produce multistep ahead forecasts and corresponding estimated losses. We now move on to a second approach, which allows different parameters to be used at different forecast horizon, maximising the multistep step ahead forecast quasi-likelihood for the HEAVY-RM model. Recall this is called the direct parameter estimator.

We first focus on the estimated parameters which come out from this approach, highlighting results from the Dow Jones Industrials example. The left of Figure 1 shows a plot of the estimated memory in the HEAVY-RM and GARCH models

$$(\alpha_R + \beta_R)^{s+1}, \quad \text{and} \quad (\alpha_G + \beta_G)^{s+1} \tag{24}$$

plotted against s when we use the quasi-likelihood, which is tuned to perform well at one-step. We see although the estimated values of these parameters are not very different, at long lags the difference becomes magnified. By the time we are one month out the HEAVY-RM model wants to give around a half the weight on recent past data and half the weight on the unconditional mean. In the GARCH model the figures are very different, it wants around 85% of the weight to come from the recent data and only 15% to come from the unconditional mean. GARCH is stronger in this aspect.

The top left of Figure 1 also shows the profile of (24) now for the directly estimated parameters, tuning each estimator to the appropriate forecast horizon. When we do this the persistence of

t-statistic for non-nested LR tests for iterative forecasts

	Horizon $h = s + 1$: HEAVY model						Horizon $h = s + 1$: Int HEAVY model					
Asset	1	2	3	5	10	22	1	2	3	5	10	22
Dow Jones Industrials	-5.72	-3.79	-3.07	-2.98	-2.16	0.78	-5.65	-3.71	-3.02	-2.75	-2.40	0.03
Nasdaq 100	-2.49	-0.46	-0.34	-0.72	1.03	-0.42	-2.47	-0.46	-0.33	-0.57	1.25	-0.02
S&P 400 Midcap	0.07	1.19	1.14	0.16	0.25	-0.41	0.16	1.21	1.15	0.38	0.72	0.64
S&P 500	-6.12	-4.50	-3.98	-4.14	-1.92	0.90	-6.01	-4.43	-3.91	-3.89	-1.51	0.81
Russell 3000	-5.69	-3.97	-3.25	-4.01	-1.82	-0.12	-5.52	-3.82	-3.20	-3.87	-1.75	-0.29
Russell 1000	-5.40	-3.88	-3.25	-3.88	-1.65	0.33	-5.25	-3.74	-3.20	-3.74	-1.67	0.06
Russell 2000	1.70	2.32	2.24	1.28	1.45	0.41	1.73	2.24	2.12	1.35	1.54	0.89
CAC 40	-4.43	-3.04	-2.32	-0.78	-0.17	1.56	-4.38	-2.96	-2.15	-0.70	-0.36	0.88
FTSE 100	-5.18	-3.34	-2.61	-1.71	-0.17	-0.10	-5.08	-3.19	-2.39	-1.62	-0.27	0.11
German DAX	-5.15	-3.40	-2.79	-1.10	-0.92	-0.47	-5.23	-3.40	-2.65	-0.68	-0.61	0.34
Italian MIBTEL	-4.13	-3.20	-3.22	-1.73	-0.86	-0.86	-4.02	-2.91	-2.66	-1.24	-0.14	-0.89
Milan MIB 30	-1.89	-0.98	-0.91	-0.17	-0.05	-0.08	-1.88	-0.89	-0.71	0.09	0.51	0.05
Nikkei 250	-3.87	-2.55	-2.06	-0.56	0.32	0.53	-3.63	-2.38	-1.75	-0.19	0.77	1.76
Spanish IBEX	-2.81	-2.51	-1.37	-0.63	-1.13	-0.61	-2.81	-2.46	-1.18	-0.53	-0.98	0.07
S&P TSE	-5.17	-4.44	-3.57	-2.23	-0.89	-0.23	-5.16	-4.40	-3.49	-2.04	-0.59	0.22
MSCI Australia	-3.14	-1.94	-2.57	-1.87	-2.35	-2.89	-3.14	-1.93	-2.54	-1.80	-1.70	-2.05
MSCI Belgium	-1.21	-1.21	-1.08	-1.75	-2.05	-2.14	-0.85	-1.04	-0.94	-1.59	-1.62	-0.58
MSCI Brazil	-3.54	-2.19	-1.40	-1.22	-1.35	-0.22	-3.31	-2.01	-1.01	-0.84	-0.49	0.45
MSCI Canada	-3.90	-3.15	-3.11	-2.47	-1.73	-1.03	-3.91	-3.14	-3.07	-2.34	-1.42	-0.43
MSCI Switzerland	-4.33	-3.01	-2.23	-1.94	-0.37	-1.50	-4.15	-2.87	-2.12	-1.88	0.13	0.50
MSCI Germany	-5.31	-4.50	-3.90	-2.45	-1.15	-1.45	-5.33	-4.43	-3.54	-1.64	-0.56	-0.07
MSCI Spain	-3.71	-2.59	-2.05	-1.22	-0.39	-0.55	-3.44	-2.36	-1.74	-1.06	-0.14	-1.05
MSCI France	-5.67	-4.56	-3.33	-1.69	-0.64	-0.06	-5.52	-4.31	-2.96	-1.33	-0.46	-0.08
MSCI UK	-5.54	-3.98	-3.20	-2.30	-0.42	-0.48	-5.17	-3.59	-2.92	-2.19	-0.47	-0.24
MSCI Italy	-5.38	-3.78	-3.32	-2.71	-1.02	-0.36	-5.29	-3.48	-2.96	-2.23	-0.63	-0.79
MSCI Japan	-5.30	-3.06	-2.28	-0.61	-0.09	0.62	-5.08	-2.90	-2.00	-0.25	0.31	1.44
MSCI South Korea	-4.79	-2.61	-2.29	-2.32	-0.49	2.74	-4.73	-2.53	-2.23	-2.25	-0.34	2.18
MSCI Mexico	-2.47	-1.79	-1.80	-1.21	-1.96	-1.26	-2.43	-1.73	-1.68	-1.03	-1.72	-1.04
MSCI Netherlands	-4.81	-3.34	-2.33	-2.14	-1.39	-1.46	-4.40	-3.06	-2.06	-1.79	-0.93	-0.57
MSCI World	-5.57	-4.37	-3.39	-2.02	-1.26	-0.37	-5.04	-3.97	-3.00	-1.41	-1.16	-0.10
British Pound	-3.33	-2.99	-2.06	-1.81	-1.44	-2.25	-3.36	-2.99	-2.02	-1.72	-1.16	-1.45
Euro	-1.14	-0.75	-0.63	-0.36	-0.22	-0.16	-1.11	-0.71	-0.59	-0.29	-0.16	0.10
Swiss Franc	-2.55	-2.82	-2.81	-2.08	-2.18	-2.32	-2.54	-2.82	-2.79	-2.00	-2.05	-1.86
Japanese Yen	-2.97	-2.35	-1.30	-0.25	-0.79	0.65	-2.88	-2.20	-1.16	-0.32	-0.64	0.12

Table 5: In-sample likelihood ratio tests where the loss generated by HEAVY and GARCH models. Negative values favour HEAVY models. Both models are estimated using the quasi-likelihood, ie tuned to one-step ahead predictions.

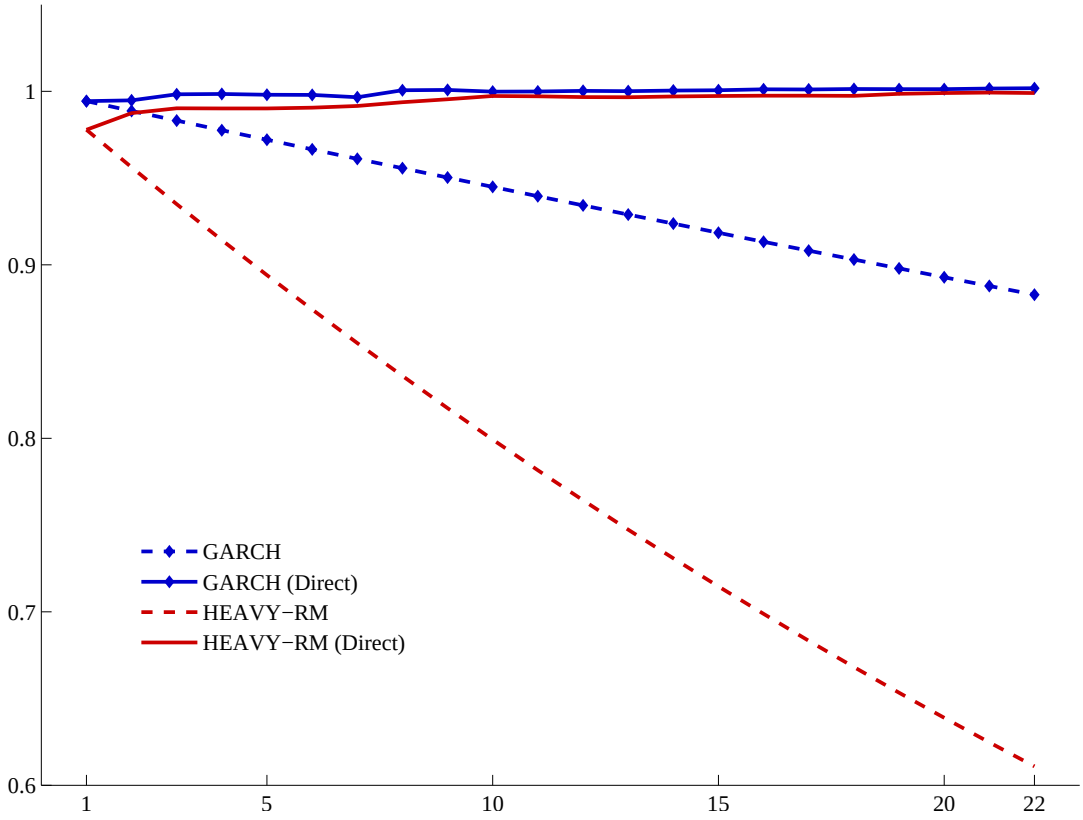

Figure 1: Direct method. Estimates of $(\alpha_R + \beta_R)^{s+1}$ and $(\alpha_G + \beta_G)^{s+1}$ drawn against forecast horizon $s + 1$.

the HEAVY-RM model jumps up beyond the level of the GARCH model. This is caused by a reduction in α_R from around 0.4 for small numbers of periods ahead to around 0.2 for longer periods ahead. As α_R decreased β_R increased even more so leading to an increase in the estimated value of $\alpha_R + \beta_R$ for large s. The increase in the level of the curve for the GARCH model in comparison is similar.

When we compare the forecast performance of the directly estimated GARCH and HEAVY models using the QLIK loss functions we see in Table 6 that the HEAVY models are systematically much better. This improvement is now sustained at quite long horizons and holds for standard HEAVY models and integrated versions.

An important question is how well we forecast the variance of the sum of s period returns. Again the forecast outperformance of HEAVY models appears for nearly all assets and forecast horizons. The results are given in Table 6.

	Pointwise Comparison						Cumulative Comparison					
	Direct HEAVY vs. Direct GARCH			Direct Int. HEAVY vs. Direct GARCH			Direct HEAVY vs. Direct GARCH			Direct Int. HEAVY vs. Direct GARCH		
Asset	1	10	22	1	10	22	5	10	22	5	10	22
Dow Jones Industrials	-5.72	-3.34	-0.95	-5.65	-3.50	-0.30	-4.40	-4.32	-3.60	-4.48	-4.52	-3.71
Nasdaq 100	-2.49	-0.51	-0.54	-2.47	-0.23	0.12	-0.88	-0.17	-0.45	-0.79	0.03	0.02
S&P 400 Midcap	0.07	0.55	0.24	0.16	0.81	1.00	0.78	0.80	0.54	0.89	1.06	1.15
S&P 500	-6.12	-4.52	-0.24	-6.01	-4.63	0.25	-5.43	-4.95	-2.84	-5.46	-5.12	-2.88
Russell 3000	-5.69	-4.24	-1.15	-5.52	-4.17	-0.27	-4.86	-4.61	-3.63	-4.78	-4.55	-3.30
Russell 1000	-5.40	-4.11	-0.69	-5.25	-4.17	-0.00	-4.75	-4.44	-3.09	-4.72	-4.52	-2.94
Russell 2000	1.70	1.56	0.81	1.73	1.66	1.17	2.01	1.97	1.59	1.99	1.99	1.76
CAC 40	-4.43	-0.98	-0.34	-4.38	-0.91	0.59	-2.98	-1.87	-1.40	-2.88	-1.79	-0.69
FTSE 100	-5.18	-1.97	-1.32	-5.08	-1.81	0.28	-3.46	-2.44	-2.35	-3.25	-2.08	-1.09
German DAX	-5.15	-1.18	-1.48	-5.23	-0.72	0.64	-3.70	-2.84	-2.84	-3.47	-2.04	-0.92
Italian MIBTEL	-4.13	-1.61	-1.52	-4.02	-1.08	-0.35	-3.17	-2.23	-2.32	-2.85	-1.84	-1.37
Milan MIB 30	-1.89	-0.37	-1.51	-1.88	-0.25	0.20	-1.07	-0.97	-1.73	-0.96	-1.08	-0.96
Nikkei 250	-3.87	-0.11	0.51	-3.63	0.18	1.03	-2.07	-1.05	-0.05	-1.84	-0.70	0.50
Spanish IBEX	-2.81	-0.90	-0.73	-2.81	-1.02	-0.11	-2.02	-2.27	-1.82	-1.96	-1.94	-0.85
S&P TSE	-5.17	-2.37	-1.83	-5.16	-2.24	-1.10	-4.14	-2.95	-2.50	-4.04	-2.92	-2.55
MSCI Australia	-3.14	-2.17	-2.84	-3.14	-2.09	-1.69	-2.51	-2.62	-3.42	-2.47	-2.45	-2.73
MSCI Belgium	-1.21	-1.89	-1.67	-0.85	-1.68	0.07	-1.60	-1.93	-2.23	-1.37	-1.68	-1.43
MSCI Brazil	-3.54	-1.56	0.03	-3.31	-1.04	0.74	-2.91	-2.03	-0.96	-2.47	-1.37	-0.24
MSCI Canada	-3.90	-2.41	-1.69	-3.91	-2.30	-1.02	-3.47	-2.71	-2.40	-3.41	-2.55	-2.20
MSCI Switzerland	-4.33	-1.95	-1.44	-4.15	-1.74	0.67	-3.10	-2.19	-2.14	-2.98	-1.71	-0.49
MSCI Germany	-5.31	-2.27	-1.50	-5.33	-1.51	0.48	-4.83	-3.03	-2.80	-4.37	-2.15	-1.13
MSCI Spain	-3.71	-1.30	-1.50	-3.44	-1.12	-0.73	-2.62	-1.82	-1.84	-2.38	-1.69	-1.37
MSCI France	-5.67	-1.61	-1.23	-5.52	-1.22	0.21	-4.25	-2.58	-2.20	-3.93	-2.05	-1.04
MSCI UK	-5.54	-2.43	-1.65	-5.17	-2.27	0.09	-3.84	-2.96	-2.54	-3.57	-2.59	-1.35
MSCI Italy	-5.38	-2.86	-2.19	-5.29	-2.43	-0.58	-4.10	-3.47	-3.72	-3.85	-3.52	-2.60
MSCI Japan	-5.30	-0.72	0.27	-5.08	-0.38	0.75	-2.88	-2.21	-1.17	-2.55	-1.65	-0.26
MSCI South Korea	-4.79	-2.30	1.21	-4.73	-2.13	1.06	-3.46	-2.71	-0.33	-3.39	-2.51	0.05
MSCI Mexico	-2.47	-1.47	-1.56	-2.43	-1.45	-1.27	-1.95	-2.12	-2.19	-1.90	-2.07	-2.27
MSCI Netherlands	-4.81	-2.14	-2.99	-4.40	-1.81	-0.83	-3.29	-2.59	-2.90	-2.99	-2.19	-1.81
MSCI World	-5.57	-2.25	-0.86	-5.04	-1.93	0.04	-4.05	-3.16	-2.69	-3.60	-2.86	-2.10
British Pound	-3.33	-1.74	-2.20	-3.36	-1.69	-1.44	-2.60	-2.18	-2.65	-2.59	-2.12	-2.34
Euro	-1.14	0.05	-0.14	-1.11	0.09	0.11	-0.60	-0.24	-0.40	-0.55	-0.19	-0.26
Swiss Franc	-2.55	-1.65	-2.50	-2.54	-1.61	-2.30	-2.66	-2.58	-3.17	-2.64	-2.56	-3.08
Japanese Yen	-2.97	-1.78	-1.25	-2.88	-1.71	0.05	-2.22	-2.24	-2.36	-2.16	-1.98	-1.42

Table 6: In-sample t-stat based LR tests comparing losses generated by the HEAVY and GARCH models. Negative values favour the HEAVY model. The left columns of each panel compared HEAVY and GARCH models using horizon tuned parameters and the right columns compare Integrated HEAVY against a standard GARCH model using horizon tuned parameters.

4.3 Out-of-Sample Performance

An out-of-sample exercise was conducted to assess the performance of HEAVY models in a more realistic scenario. All models were estimated using a moving window with a width of 4 years (1,008 observation) and parameters were updated daily. Forecasts were then produced for 1- through 22-steps ahead. Table 7 contains the results of this exercise based on two comparisons. The first comparison is based on direct estimation of both the HEAVY-RM model and its GARCH competitor. In both cases parameters were optimized to fitting the realized measure (HEAVY-RM) or squared return (GARCH) at the forecasting horizon. All HEAVY models used the same HEAVY-r model which was optimized for the 1-step horizon. The second compares the performance of the Integrated HEAVY-RM specification with a standard GARCH, where both sets of parameters were optimized for 1-step prediction. The standard HEAVY model with based on 1-step tuning is not included since the memory parameter chosen was often implausibly small. Neither the directly estimated HEAVY model nor the Integrated HEAVY suffer from this issue.

The left panel contains pointwise comparisons which assess the forecasting performance at a specific horizon where performance is assessed using Giacomini and White (2006) tests which evaluate the loss of both the innovation and the parameter estimation uncertainty. These results strongly favor the HEAVY models in both cases, especially at shorter horizons. The results for the S&P 400 Midcap index and the Russell 2000 further highlight the strength of the HEAVY model – despite decidedly worse performance in full-sample comparisons, HEAVY models outperform GARCH models in out-of-sample evaluation. This difference is likely due to the higher signal-to-noise ratio of realized measures.

The right panel contains cumulative comparisons for the two sets of models. Cumulative loss measures the performance on the total variation over the forecast horizon, and so the 1-step is identical to the pointwise (and so replaced by the 5-step horizon). HEAVY models perform well at all horizons, with statistically significant outperformance in most series while never being outperformed by GARCH-based forecasts.

4.4 Parameter stability

Figure 2 shows time series plots of the estimated HEAVY and GARCH parameters estimated using the quasi-likelihood based on a moving window of four years of data, recording the estimates at the time of the last data point in the sample. The top of the plot shows very dramatic percentage changes in the GARCH α_G parameter while relatively modest movements in the corresponding HEAVY parameter α_R.

The bottom of Figure 2 shows the rolling estimate of the persistence parameters for the GARCH

	Pointwise Comparison						Cumulative Comparison					
	Direct HEAVY vs. Direct GARCH			Int. HEAVY vs. GARCH			Direct HEAVY vs. Direct GARCH			Int. HEAVY vs. GARCH		
Asset	1	10	22	1	10	22	5	10	22	5	10	22
Dow Jones Industrials	-5.94	-2.74	0.39	-5.81	-3.04	-0.60	-5.19	-4.87	-2.83	-4.83	-4.75	-3.03
Nasdaq 100	-5.43	-1.00	-2.67	-5.28	-3.55	-2.55	-4.51	-3.50	-2.94	-4.33	-4.64	-3.45
S&P 400 Midcap	-2.87	-0.81	-2.50	-2.98	-0.07	-1.25	-2.01	-1.89	-1.29	-1.90	-2.47	-2.50
S&P 500	-6.55	-1.96	0.24	-6.57	-3.14	-0.34	-5.40	-4.55	-2.03	-5.12	-4.79	-2.67
Russell 3000	-6.00	-1.87	-0.88	-5.89	-3.48	-1.17	-5.29	-4.37	-2.64	-5.22	-5.23	-3.49
Russell 1000	-6.01	-1.82	-0.66	-5.90	-3.41	-0.91	-5.37	-4.36	-2.53	-5.24	-5.22	-3.24
Russell 2000	-0.97	0.20	-0.80	-1.07	0.24	-0.73	0.17	0.48	-0.42	-0.43	-0.11	-0.30
CAC 40	-4.82	-0.20	-2.08	-4.76	-1.06	-1.46	-4.31	-1.76	-1.72	-4.02	-2.72	-2.07
FTSE 100	-5.45	-2.02	-2.84	-5.57	-1.72	-2.13	-3.78	-2.90	-2.85	-3.85	-2.86	-2.44
German DAX	-3.96	-2.49	-3.57	-4.12	-2.11	-1.32	-3.84	-3.33	-4.02	-3.89	-3.25	-2.55
Italian MIBTEL	-2.87	-0.81	-2.50	-2.98	-0.07	-1.25	-1.88	-1.61	-2.61	-1.51	-0.71	-0.86
Milan MIB 30	-4.18	-1.13	-3.33	-4.28	-1.19	-1.41	-3.36	-2.94	-3.69	-3.27	-2.59	-2.32
Nikkei 250	-3.35	-0.64	-0.03	-3.36	-0.68	0.93	-3.74	-3.35	-0.37	-3.41	-2.76	-0.90
Spanish IBEX	-3.13	-0.52	-2.96	-3.19	-0.87	-1.28	-2.88	-2.10	-2.06	-2.54	-1.68	-1.33
S&P TSE	-3.29	-1.78	-0.46	-3.25	-1.03	0.63	-3.07	-2.36	-1.72	-2.91	-1.97	-0.53
MSCI Australia	-2.60	-2.15	-1.65	-2.61	-1.48	-1.25	-2.01	-2.10	-2.91	-1.95	-1.79	-1.49
MSCI Belgium	-3.28	-3.69	-3.29	-3.26	-2.79	-3.54	-3.16	-3.67	-4.79	-2.52	-2.63	-3.45
MSCI Brazil	-2.21	-1.52	0.54	-2.27	-1.65	-0.62	-1.58	-1.61	-0.97	-1.46	-1.57	-0.92
MSCI Canada	-3.41	-1.98	-1.49	-3.34	-1.04	0.17	-3.01	-2.30	-1.88	-2.81	-1.82	-0.85
MSCI Switzerland	-5.15	-2.22	-2.65	-5.13	-1.90	-2.91	-4.47	-3.34	-3.83	-4.64	-3.25	-3.81
MSCI Germany	-3.15	-3.67	-1.93	-3.18	-1.93	-1.64	-3.26	-3.46	-3.95	-2.83	-2.35	-1.97
MSCI Spain	-2.82	-1.39	-2.88	-2.84	-0.96	-1.28	-2.89	-2.38	-2.30	-2.50	-1.68	-1.23
MSCI France	-4.38	-2.06	-2.81	-4.39	-1.31	-2.01	-4.64	-2.91	-3.67	-3.99	-2.64	-2.12
MSCI UK	-4.30	-1.09	-3.13	-4.32	-1.56	-3.04	-3.79	-2.74	-2.29	-3.25	-2.60	-2.52
MSCI Italy	-4.08	-2.64	-2.88	-4.08	-1.40	-2.19	-3.37	-3.78	-4.53	-3.02	-2.38	-2.44
MSCI Japan	-2.73	-0.18	-0.25	-2.62	0.15	0.60	-2.72	-1.79	-0.58	-2.43	-1.44	-0.58
MSCI South Korea	-4.08	0.14	1.12	-4.10	-1.68	0.18	-2.65	-1.62	0.34	-3.06	-2.74	-1.38
MSCI Mexico	-2.23	-1.34	-0.63	-2.24	-1.28	-0.92	-1.53	-1.47	-0.89	-1.47	-1.43	-1.09
MSCI Netherlands	-4.58	-3.35	-3.08	-4.55	-2.36	-1.62	-4.21	-4.21	-3.54	-4.09	-3.28	-2.49
MSCI World	-3.30	-0.08	0.20	-3.59	-0.93	-1.15	-2.07	-1.22	-0.99	-2.41	-1.73	-1.29
British Pound	-2.53	-1.53	-1.60	-2.59	-0.94	-0.97	-2.24	-1.90	-1.82	-2.16	-1.65	-1.24
Euro	-1.05	-0.03	0.70	-1.03	-0.69	-0.56	-0.65	-0.10	0.08	-0.85	-0.69	-0.58
Swiss Franc	-2.09	-0.82	-2.33	-2.10	-1.58	-2.32	-2.12	-1.51	-1.94	-2.24	-2.03	-2.22
Japanese Yen	-2.22	-1.12	-1.34	-2.22	-1.44	-0.54	-1.82	-1.76	-1.64	-1.75	-1.85	-1.23

Table 7: Out-of-sample t-stat based LR tests comparing losses generated by the HEAVY and GARCH models. Negative values favour the HEAVY model. The left columns of each panel compared HEAVY and GARCH models using horizon tuned parameters and the right columns compare Integrated HEAVY against a standard GARCH model using 1-step-ahead tuned parameters.

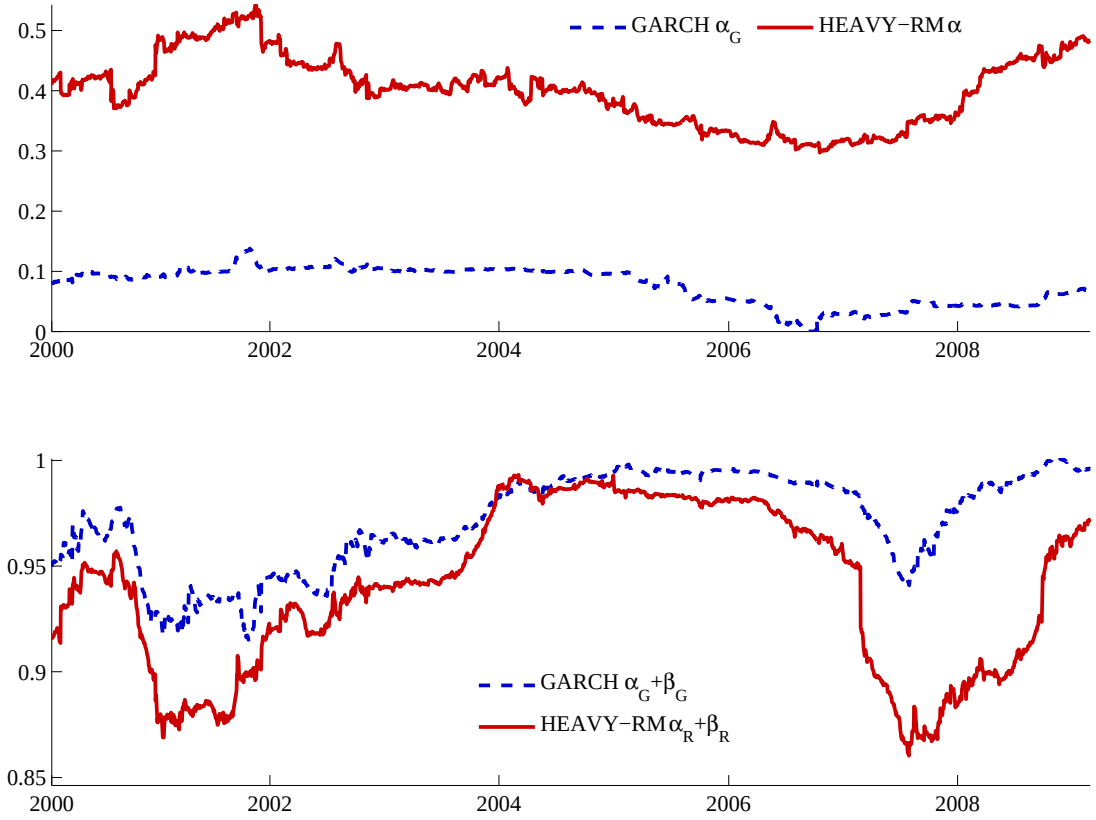

Figure 2: Recurisve parameter estimates using a quasi-likelihood for GARCH and HEAVY model.

model $\alpha_G + \beta_G$ and the HEAVY-RM model $\alpha_G + \beta_G$. The latter shows consistently less memory than the former, but interestingly the two sequences of parameter estimates are moving around in lock step.

The top of Figure 2 shows a rolling estimate of the HEAVY-r's α parameter which controls the immediate impact on the predicted conditional variance of the lagged realised measure. It is a volatile picture, but the percentage moves are actually quite modest.

4.5 Properties of the innovations

One way of thinking about the performance of the model is by computing the one-step ahead innovations from the model

$$\widehat{\zeta}_t = r_t/h_t^{1/2}, \quad \widehat{\eta}_t = (RM_t/\mu_t)^{1/2}, \quad t = 2, 3, ..., T.$$

In this section we will do this based on the model fitted using the quasi-likelihood criteria.

Figure 3 shows these innovations for the Dow Jones Index example, which is pretty typical of results we have seen for other series. In the top left hand side of the Figure we have a time series plot of $\widehat{\zeta}_t$. It does not show much volatility clustering, but there are some quite large negative

innovations, with a couple of days reporting falls which are larger than -5. These happened at the start of 1996 and at the start of 2007. Notice there are no remarkable moves during the credit crunch.

In the top right hand side of the Figure 3 there is a time series plot of $\widehat{\eta}_t$, which has large moves in at the same time as the large moves in $\widehat{\zeta}_t$. This is confirmed in the bottom left hand side of the Figure, which cross plots $\widehat{\zeta}_t$ and $\widehat{\eta}_t$, which suggests some dependence in the bottom right hand quadrant. The bottom right shows the empirical copula for $\widehat{\zeta}_t$ and $\widehat{\eta}_t$, from which it is hard to see much dependence, although there is little mass in the bottom left hand quadrant and a cluster of points in the bottom right.

Asset	$\min(\widehat{\zeta}_t)$	$\max(\widehat{\zeta}_t)$	$E(\widehat{\zeta}_t^3)$	$Sd(\widehat{\zeta}_t^2)$	$Sd(\widehat{\eta}_t)$	$r(\widehat{\zeta}_t,\widehat{\eta}_t)$	$r_s(\widehat{\zeta}_t,\widehat{\eta}_t)$	$r_s(\widehat{\zeta}_t^2,\widehat{\eta}_t)$
Dow Jones Industrials	-6.19	3.15	-.336	1.82	.270	-.313	-.280	.165
Nasdaq 100	-5.90	4.25	-.149	1.66	.264	-.321	-.323	.091
S&P 400 Midcap	-7.47	3.51	-.366	1.87	.257	-.351	-.334	.130
S&P 500	-6.86	3.61	-.396	1.90	.270	-.331	-.312	.147
Russell 3000	-7.01	3.97	-.338	1.88	.276	-.339	-.335	.137
Russell 1000	-7.40	3.91	-.346	1.92	.275	-.337	-.329	.140
Russell 2000	-7.08	3.48	-.385	1.83	.285	-.289	-.264	.193
CAC 40	-4.37	3.64	-.212	1.53	.262	-.350	-.323	.214
FTSE 100	-4.29	3.90	-.307	1.52	.263	-.330	-.312	.195
German DAX	-5.30	3.79	-.216	1.59	.259	-.396	-.367	.218
Italian MIBTEL	-4.90	3.40	-.464	1.63	.255	-.430	-.421	.158
Milan MIB 30	-5.38	4.87	-.076	1.79	.263	-.334	-.339	.156
Nikkei 250	-5.78	3.95	-.343	1.77	.260	-.198	-.162	.211
Spanish IBEX	-6.95	5.19	-.237	1.92	.262	-.328	-.297	.206
S&P TSE	-5.82	3.48	-.225	1.63	.254	-.282	-.286	.182
MSCI Australia	-6.19	3.63	-.318	1.75	.238	-.244	-.207	.111
MSCI Belgium	-5.78	3.04	-.391	1.77	.239	-.310	-.267	.208
MSCI Brazil	-5.08	3.61	-.194	1.59	.258	-.327	-.311	.132
MSCI Canada	-4.52	3.47	-.232	1.58	.247	-.309	-.298	.177
MSCI Switzerland	-5.98	3.43	-.453	1.84	.231	-.396	-.346	.209
MSCI Germany	-4.94	3.21	-.333	1.57	.246	-.390	-.370	.223
MSCI Spain	-5.48	3.63	-.211	1.60	.243	-.312	-.297	.210
MSCI France	-4.55	3.06	-.249	1.48	.250	-.345	-.335	.211
MSCI UK	-4.71	3.17	-.381	1.60	.251	-.347	-.328	.190
MSCI Italy	-4.44	3.17	-.392	1.56	.241	-.396	-.385	.177
MSCI Japan	-5.95	3.41	-.351	1.69	.235	-.274	-.212	.161
MSCI South Korea	-5.64	3.37	-.239	1.71	.222	-.233	-.229	-.001
MSCI Mexico	-5.19	3.75	-.107	1.74	.241	-.262	-.222	.239
MSCI Netherlands	-5.00	3.23	-.296	1.55	.242	-.368	-.352	.216
MSCI World	-5.36	4.34	-.197	1.62	.259	-.227	-.225	.313
British Pound	-3.58	3.76	-.061	1.51	.170	-.050	-.030	.344
Euro	-4.20	3.48	.060	1.54	.196	.014	.017	.335
Swiss Franc	-4.49	3.91	-.182	1.57	.184	-.101	-.080	.344
Japanese Yen	-4.65	3.71	-.322	1.80	.222	-.193	-.128	.357

Table 8: Descriptive statistics of the estimated innovations $\widehat{\zeta}_t$ and $\widehat{\eta}_t$ from the fitted HEAVY model. Their empirical variance and mean were, respectively, very close to one and so are not reported here. First five columns are estimated moments of their marginal distributions. r denotes the correlation, r_s is the Spearman rank correlation coefficient.

Summary statistics for the innovations for all the series are given in Table 8. We have chosen

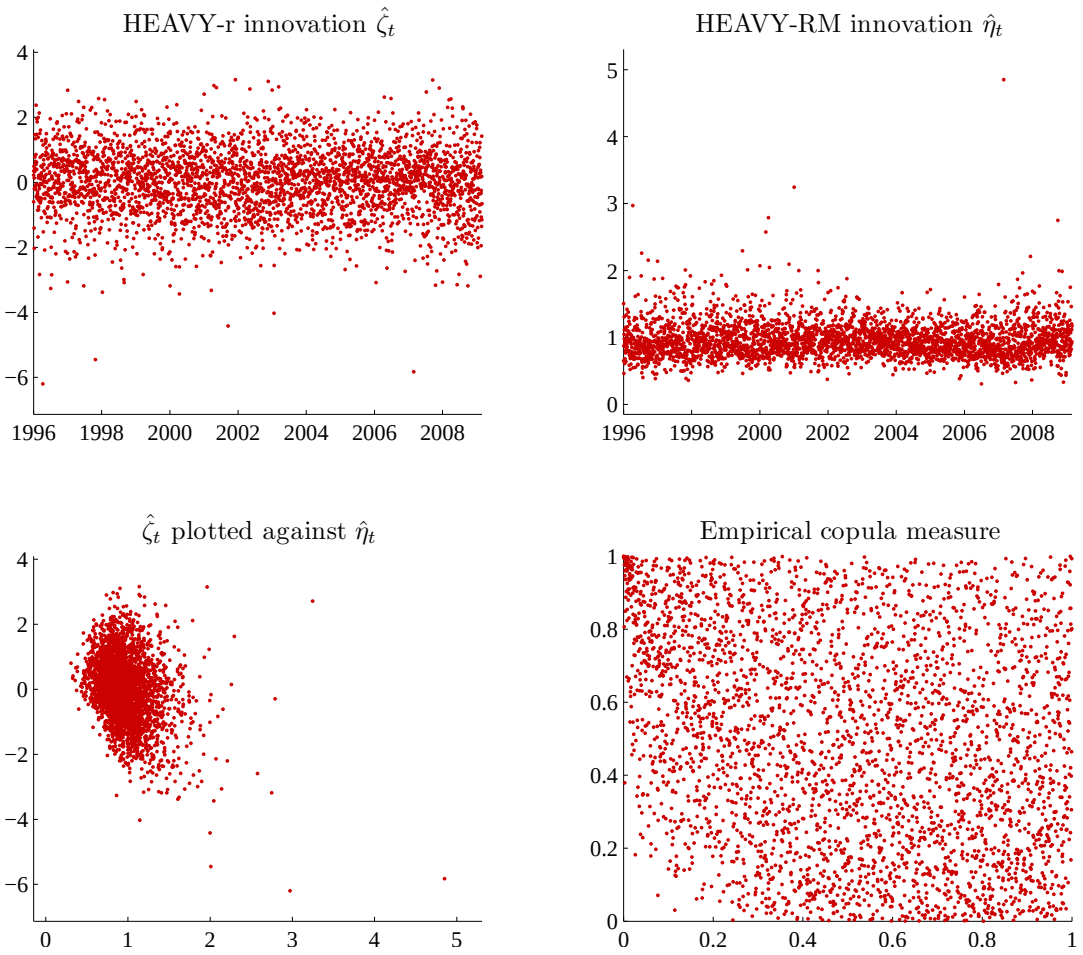

Figure 3: Innovations $\widehat{\zeta}_t$ and $\widehat{\eta}_t$ from the HEAVY model fitted to the DJI. Top left: the HEAVY-r model innovations $\widehat{\zeta}_t$, which should be roughly martingale difference sequences with unit variance. The top right is $\widehat{\eta}_t$ with should have unit conditional means and be uncorrelated. Botton left is a cross plot of $\widehat{\zeta}_t$ and $\widehat{\eta}_t$, while bottom right is the equivilent version mapped into copula spaces using the marginal empirical distributions functions to calculate the empirical copula measure.

not to report the estimated $\mathrm{E}(\widehat{\zeta}_t^2)$ and $\mathrm{E}(\widehat{\eta}_t)$ as these are for all series extremely close to one. Here r denotes the estimated correlation coefficient and r_s denotes the Spearman's rank coefficient. We will first focus on the first row, the Dow Jones series. The raw correlation shows a large amount of negative correlation between the $\widehat{\zeta}_t$ and $\widehat{\eta}_t$ for all the equity series. This negative dependence is a measure of statistical leverage — that is falls in equity prices are associated with rises in volatility. For exchange rates the correlation is roughly zero. The Spearman's rank correlations show the same pattern, while $\widehat{\zeta}_t^2, \widehat{\eta}_t$ are mildly positively correlated based on ranks.

The other features of the Table which are interesting is that there is strong evidence that $\widehat{\zeta}_t$ has a negative skew and that the standard deviation of $\widehat{\zeta}_t^2$ is not far from two. The latter suggests that the marginal distribution of $\widehat{\zeta}_t$ is not very thick tailed. These results are common across different series except for the exchange rates which are closer to symmetry, except for the Yen.

4.6 Volatility hedgehog plots

It is challenging to plot sequences of multistep ahead volatility forecasts. We carry this out using what we call "volatility hedgehog plots" and illustrate it through the credit crunch of late 2008. An example of this is Figure 4, which is calculated for the MSCI Canada series. It plots the time series of one step ahead forecasts from the HEAVY-r model h_t; these are joined together using a thick solid red line. For a selected number of days (if all days are plotted then it is hard to see the details) we also draw off the one step ahead forecast the corresponding multistep ahead forecast drawn using a red dotted line over the next month. The corresponding results for the GARCH model are also shown using a thick blue line with added symbols, with the multistep ahead forecasts being shown using a dashed line.

The Figure shows the GARCH model always slowly mean reverting back to its long term average. In this picture it also shows from the start of September a sequence of upward moves in the volatility, caused by the slow adjustment of the GARCH model.

The HEAVY model has a rather different profile. This is most clearly seen by the highest volatility point, where the multistep ahead forecast shows momentum. This is highlighted by displaying an ellipse. The model expected volatility to increase even further than we had already seen in the data. The other features which are interesting is that the HEAVY model has, in the first half of the data sample, much higher levels of volatility. After the end of October volatility falls, with the HEAVY model indicating very fast falls suggesting a lull in volatility during November 2008, before it kicks back up in December before falling to around 45% for the remaining 3 months of the data. GARCH models do not see this lull, instead from half way through October until the end of December the GARCH model shows historically very high levels of volatility with a slow decline.

Overall the main impressions we get from this graph is the slow and steady adjustments of the GARCH model and the more rapid movements implied by the HEAVY model. There is some evidence that GARCH was behind the curve during the peak of the financial crisis, while HEAVY models rapidly adjust. Likewise it looks like GARCH's volatility was too high during late December and early January as the model could not allow the conditional variance to fall rapidly enough. The momentum effects of the HEAVY model are not very large in these figures but they do have an impact. Basically local trends are followed through before mean reversion overcomes them.

More dramatic momentum effects can be seen from the Swiss Franc case, which is the most extreme example of momentum we have seen in our empirical work. For the HEAVY model β is much higher than is typical for equities, being around 0.95. This means the momentum feature has considerable memory. The result is some interesting arcs which appear in the volatility hedgehog

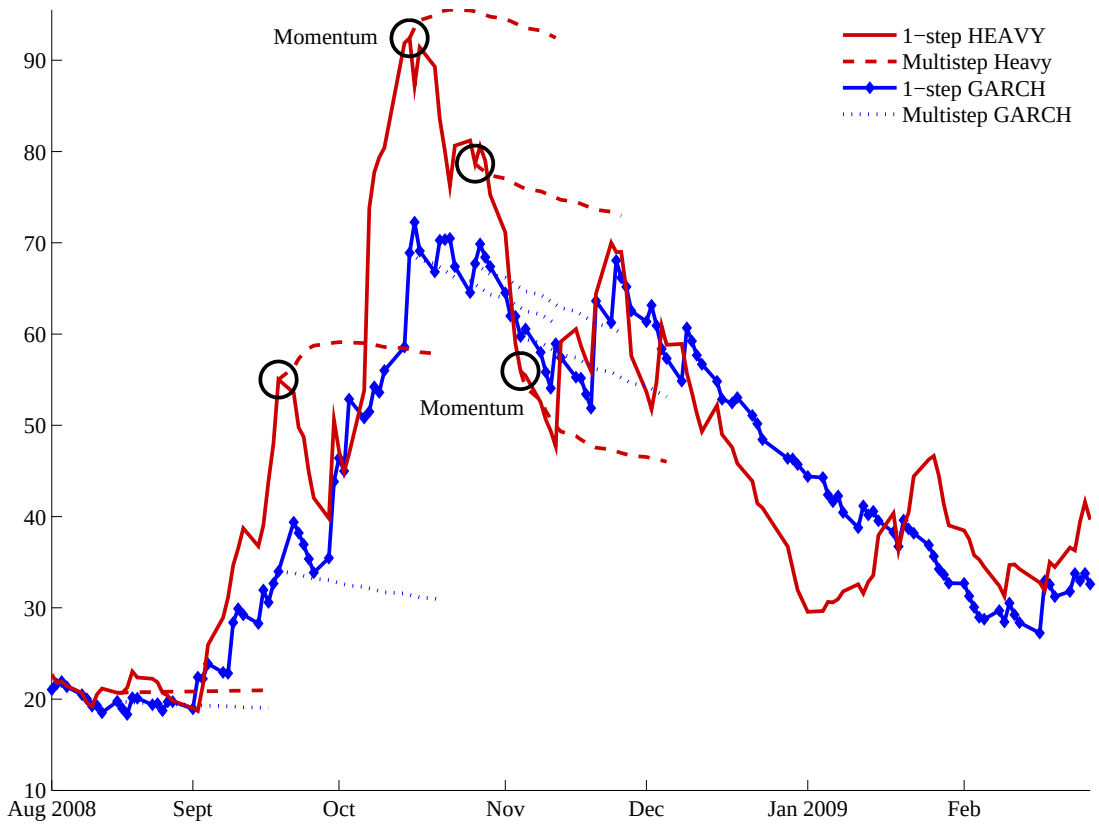

Figure 4: Volatility hedgehog plot for annualised volatility for the MSCI Canada series. The hedgehog plots are given for both HEAVY and GARCH models. Areas of momentum are indicated by elipses.

plot given in Figure 5. The evidence in Table 3 is that the HEAVY model is a better fit than for GARCH models but the difference is very modest for exchange rates in the library while for other assets it is quite substantial.

5 Extensions

5.1 A semiparametric model for $F_{\zeta,\eta}$

The joint distribution of the innovations $F_{\zeta,\eta}$ can be approximated by the joint empirical distribution function, which can be used inside a bootstrap procedure.

We could impose a model on the joint distribution via the following simple structure. Let $\eta_t \sim F_\eta$ and

$$\zeta_t|\eta_t \stackrel{L}{=} \beta\left\{\eta_t - \mathrm{E}(\eta_t)\right\} + \eta_t^{1/2}\varepsilon_t, \quad \varepsilon_t \sim F_\varepsilon, \quad \eta_t \perp\!\!\!\perp \varepsilon_t.$$

This is a nonparametric location scale mixture[11]. Now $\varepsilon_t = \eta_t^{-1/2}\left[\zeta_t - \beta\left\{\eta_t - \mathrm{E}(\eta_t)\right\}\right]$ and so we

[11] If the parametric assumption that F_η was a generalised inverse Gaussian distribution and F_ε was Gaussian, then

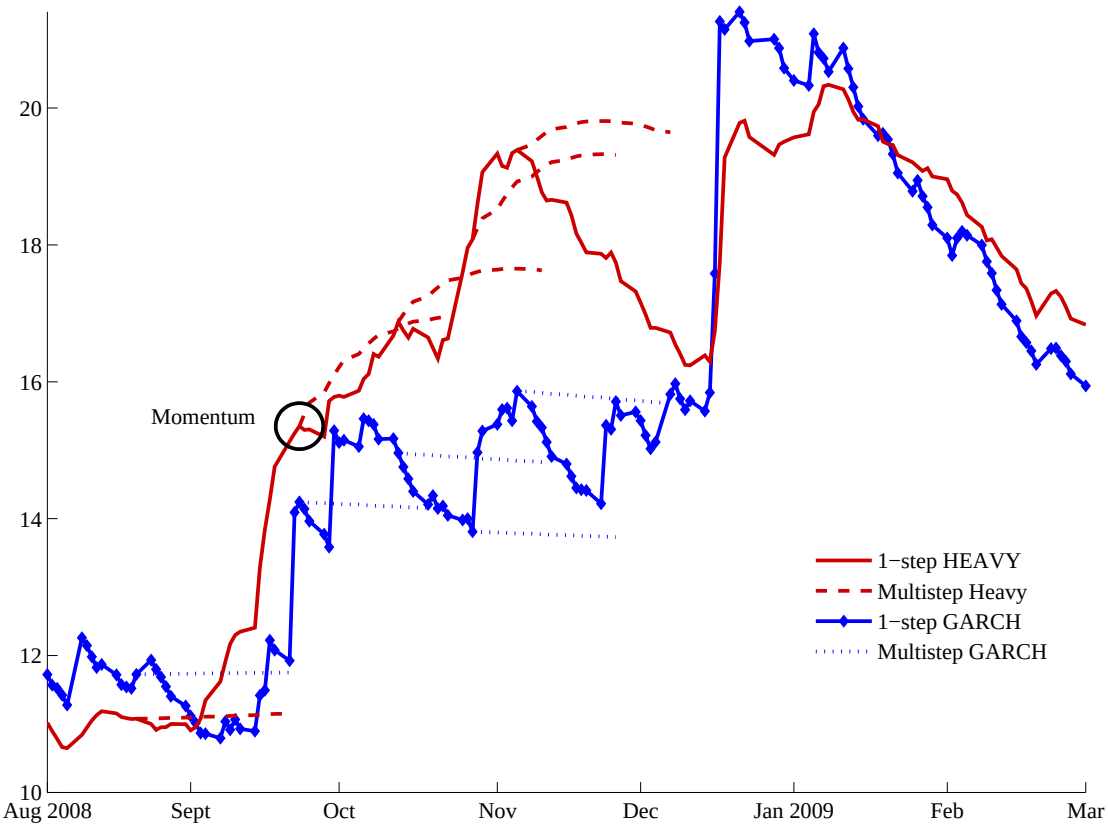

Figure 5: Extreme case of momentum. Volatility hedgehog plot for annualised volatility for the Swiss Franc against the US Dollar. The hedgehog plots are given for both HEAVY and GARCH models.

can be estimate the distribution functions F_η and F_ε by their univariate empirical distribution functions, having estimated β by using the fact that under this model $\text{Cov}(\zeta_t, \eta_t) = \beta$.

5.2 Extending HEAVY-r

In some cases where the realised measure is inadequate it may be better to extend the HEAVY-r model to allow a GARCHX structure. Then the HEAVY model becomes

$$\begin{aligned} \text{Var}(r_t|\mathcal{F}_{t-1}^{HF}) &= h_t = \omega + \alpha RM_{t-1} + \beta h_{t-1} + \gamma r_{t-1}^2, \quad \beta + \gamma < 1 \\ \text{E}(RM_t|\mathcal{F}_{t-1}^{HF}) &= \mu_t = \omega_R + \alpha_R RM_{t-1} + \beta_R h_{t-1}, \quad \alpha_R + \beta_R < 1. \end{aligned}$$

Then it is straightforward to see that r_t^2 has an ARMA(2,2) representation with autoregressive roots $\alpha_R + \beta_R$ and $\beta + \gamma$. The moving average roots are not changed by having $\gamma > 0$. Thus this extension has more momentum than the standard HEAVY model.

the resulting distribution for ζ_t would be the well known generalised hyperbolic distribution.

The derivation of this result is as follows.

$$\begin{aligned} r_t^2 &= h_t + u_t, \quad h_t = \omega + \alpha RM_{t-1} + \beta h_{t-1} + \gamma r_{t-1}, \quad \text{so} \\ \{1 - (\beta+\gamma)\, L\}\, r_t^2 &= \omega + \alpha RM_{t-1} + (1-\beta L)\, u_t, \end{aligned}$$

where L is the lag operator. Likewise

$$\{1 - (\alpha_R + \beta_R)\, L\}\, RM_t = \omega_R + (1 - \beta_R L)\, v_t, \quad v_t = RM_t - \mu_t.$$

Combining delivers the result. In particular

$$\{1 - (\beta+\gamma)\, L\}\, r_t^2 = \omega + \alpha \frac{\{\, \omega_R + (1-\beta_R L)\, v_{t-1}\}}{\{1 - (\alpha_R + \beta_R)\, L\}} + (1 - \beta L)\, u_t.$$

So

$$\begin{aligned} \{1 - (\alpha_R + \beta_R)\, L\}\, \{1 - (\beta+\gamma)\, L\}\, r_t^2 = \{1 - (\alpha_R + \beta_R)\}\, \omega + \alpha\, \{\, \omega_R + (1-\beta_R L)\, v_{t-1}\} \\ + \{1 - (\alpha_R + \beta_R)\, L\}\, (1 - \beta L)\, u_t. \end{aligned}$$

5.3 Missing data

Although much of financial data is of high quality, there are often gaps of various types due to public holidays or datafeed breakdowns or concerns over opening and closing auctions in equity markets. In the multivariate case missing data is very important due to asynchronous high frequency data reflecting trading at different times and due to differential market openings and closing around the globe when using low frequency daily data. Early analysis of some of these issues includes Scholes and Williams (1977) and Lo and MacKinlay (1990). Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008b) look at estimating multivariate quadratic variation in the presence of non-synchronous data. Burns, Engle, and Mezrich (1998) wrote about some aspects of asynchronous data and market closures in the context of multivariate GARCH models. They fit a vector first order moving average to returns and extract estimated prices which they call "synchronised." Our approach is distinct from this.

It is attractive to have a principled way of dealing with missing data. One approach is through a data augmentation exercise, which is pursued in detail in Fiorentini, Sentana, and Shephard (2004). However, that approach is certainly cumbersome.

The approach we use in this paper follows Shephard and Sheppard (2009), adapted to the HEAVY model. There we replace the function of data in our models by the relevant conditional expectations. In the GARCH case this is

$$\begin{aligned} \sigma_t^2 &= \omega_G + \alpha_G \mathrm{E}(r_{t-1}^2 | \mathcal{F}_{t-1}^{LF}) + \beta_G \sigma_{t-1}^2 \\ &= \omega_G + \alpha_G \left\{ \mathrm{E}(r_{t-1}^2 | \mathcal{F}_{t-1}^{LF}) - \mathrm{E}(r_{t-1}^2 | \mathcal{F}_{t-2}^{LF}) \right\} + (\alpha_G + \beta_G)\, \sigma_{t-1}^2. \end{aligned}$$

This is attractive as

$$\mathrm{E}(r_{t-1}^2|\mathcal{F}_{t-1}^{LF}) - \mathrm{E}(r_{t-1}^2|\mathcal{F}_{t-2}^{LF})$$

is a martingale difference sequence with respect to $\mathcal{F}_{t-2}^{LF}$. This structure leaves us the generic task of building a model for $\mathrm{E}(r_{t-1}^2|\mathcal{F}_{t-1}^{LF})$.

In the HEAVY case we need to write

$$\begin{aligned} h_t &= \omega + \alpha \mathrm{E}\left(RM_{t-1}|\mathcal{F}_{t-1}^{HF}\right) + \beta h_{t-1}, \\ \mu_t &= \omega_R + \alpha_R \mathrm{E}\left(RM_{t-1}|\mathcal{F}_{t-1}^{HF}\right) + \beta_R \mu_{t-1}. \end{aligned}$$

If the data is "missing at random" then one would replace

$$\mathrm{E}(r_{t-1}^2|\mathcal{F}_{t-1}^{LF}) = \sigma_{t-1}^2, \quad \mathrm{E}\left(RM_{t-1}|\mathcal{F}_{t-1}^{HF}\right) = \mu_{t-1}.$$

If the data is available then

$$\mathrm{E}(r_{t-1}^2|\mathcal{F}_{t-1}^{LF}) = r_{t-1}^2, \quad \mathrm{E}\left(RM_{t-1}|\mathcal{F}_{t-1}^{HF}\right) = RM_{t-1}.$$

This principled way of dealing with missing data is helpful in the multivariate context, where information in other series may allow us to model this conditional expectation sensibly. This is the problem which Burns, Engle, and Mezrich (1998) attempted to address.

Example 1 *Suppose we have a point in time when the realised measure is missing but we do have squared returns. Then we need to compute*

$$\mathrm{E}\left(RM_{t-1}|\mathcal{F}_{t-1}^{HF}\right) = \mathrm{E}\left(RM_{t-1}|\mathcal{F}_{t-2}^{HF}, r_{t-1}^2\right).$$

The simplest version of this would be to model

$$\mathrm{E}\left(RM_{t-1}|\mathcal{F}_{t-2}^{HF}, r_{t-1}^2\right) = \alpha_M r_{t-1}^2 + \beta_M \mu_{t-1}, \quad \alpha_M, \beta_M \in [0,1].$$

This would imply

$$\begin{aligned} h_t &= \omega + \beta h_{t-1} + \alpha \left[1_{M_{t-1}} \left\{\alpha_M r_{t-1}^2 + \beta_M \mu_{t-1}\right\} + \left(1 - 1_{M_{t-1}}\right) RM_{t-1}\right] \\ &= \begin{cases} \omega + \beta h_{t-1} + \alpha RM_{t-1}, & \textit{not missing} \\ \omega + \beta h_{t-1} + \alpha \mu_{t-1} + \alpha \left\{\alpha_M r_{t-1}^2 - (1-\beta_M)\mu_{t-1}\right\}, & \textit{missing,} \end{cases} \end{aligned}$$

while

$$\begin{aligned} \mu_t &= \omega_R + \beta_R \mu_{t-1} + \alpha_R \left[1_{M_{t-1}} \left\{\alpha_M r_{t-1}^2 + \beta_M \mu_{t-1}\right\} + \left(1 - 1_{M_{t-1}}\right) RM_{t-1}\right] \\ &= \begin{cases} \omega_R + \beta_R \mu_{t-1} + \alpha_R RM_{t-1}, & \textit{not missing} \\ \omega_R + \beta_R \mu_{t-1} + \alpha_R \mu_{t-1} + \alpha_R \left\{\alpha_M r_{t-1}^2 - (1-\beta_M)\mu_{t-1}\right\}, & \textit{missing.} \end{cases} \end{aligned}$$

In practice we would expect α_M to be quite small and β_M to be close to one. It is somewhat tempting to estimate α_M, β_M using an additional quasi-likelihood (like (16)), available for all the data in the sample except when RM_t is missing, or one can apply the above more directly. It is likely to former does well if there is very little missing data and the latter is better for forecasting if there is a quite a lot (using the theory of misspecified models).

6 Conclusions

In this paper we have given a self-contained and sustained analysis of a particular model of conditional volatility based on high frequency data. HEAVY models are relatively easy to estimate and have both momentum and mean reversion. We show these models are more robust to level breaks in the volatility than conventional GARCH models, adjusting to the new level much faster. Further, as well as showing mean reversion, HEAVY models exhibit momentum, a feature which is missing from traditional models.

References

Andersen, T. G., T. Bollerslev, and F. X. Diebold (2007). Roughing it up: Including jump components in the measurement, modeling and forecasting of return volatility. *Review of Economics and Statistics 89*, 707–720.

Andersen, T. G., T. Bollerslev, and F. X. Diebold (2009). Parametric and nonparametric measurement of volatility. In Y. Aït-Sahalia and L. P. Hansen (Eds.), *Handbook of Financial Econometrics*. Amsterdam: North Holland. Forthcoming.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens (2001). The distribution of realized stock return volatility. *Journal of Financial Economics 61*, 43–76.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001). The distribution of exchange rate volatility. *Journal of the American Statistical Association 96*, 42–55. Correction published in 2003, volume 98, page 501.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2003). Modeling and forecasting realized volatility. *Econometrica 71*, 579–625.

Andersen, T. G., T. Bollerslev, and N. Meddahi (2006). Market microstructure noise and realized volatility forecasting. Unpublished paper: Department of Economics, Duke University.

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica 59*, 817–858.

Bandi, F. M. and J. R. Russell (2006). Seperating microstructure noise from volatility. *Journal of Financial Economics 79*, 655–692.

Bandi, F. M. and J. R. Russell (2008). Microstructure noise, realized variance, and optimal sampling. *Review of Economic Studies 75*, 339–369.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2008a). Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica 76*, 1481–1536.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2008b). Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. Unpublished paper: Oxford-Man Institute, University of Oxford.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2009). Realised kernels in practice: trades and quotes. *Econometrics Journal*. Forthcoming.

Barndorff-Nielsen, O. E., S. Kinnebrouck, and N. Shephard (2009). Measuring downside risk: realised semivariance. In T. Bollerslev, J. Russell, and M. Watson (Eds.), *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle*. Oxford University Press. Forthcoming.

Barndorff-Nielsen, O. E. and N. Shephard (2002). Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B 64*, 253–280.

Barndorff-Nielsen, O. E. and N. Shephard (2007). Variation, jumps and high frequency data in financial econometrics. In R. Blundell, T. Persson, and W. K. Newey (Eds.), *Advances in Economics and Econometrics. Theory and Applications, Ninth World Congress*, Econometric Society Monographs, pp. 328–372. Cambridge University Press.

Bollerslev, T. (1986). Generalised autoregressive conditional heteroskedasticity. *Journal of Econometrics 51*, 307–327.

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

Brownlees, C. T. and G. M. Gallo (2006). Financial econometrics at ultra-high frequency: data handling concerns. *Computational Statistics and Data Analysis 51*, 2232–2245.

Brownlees, C. T. and G. M. Gallo (2009). Comparison of volatility measures: a risk management perspective. *Journal of Financial Econometrics*. Forthcoming.

Burns, P., R. F. Engle, and J. Mezrich (1998). Correlations and volatilities of asynchronous data. *Journal of Derivatives*, 7–18.

Christensen, K. and M. Podolskij (2007). Asymptotic theory for range-based estimation of integrated volatility of a continuous semi-martingale. *Journal of Econometrics 141*, 323–349.

Christensen, P. F., K. Jacobs, and Y. Wang (2008). Option valuation with long-run and short-run volatility components. *Journal of Financial Economics 90*, 272–297.

Cipollini, F., R. F. Engle, and G. Gallo (2007). A model for multivariate non-negative valued processes in financial econometrics. Unpublished paper: Stern School of Business, New York University.

Clements, M. P. and D. F. Hendry (1999). *Forecasting Non-stationary Economic Time Series: The Zeuthen Lectures on Economic Forecasting*. Cambridge, Mass.: MIT Press.

Corradi, V. and W. Distaso (2006). Semiparametric comparison of stochastic volatility models using realized measures. *Review of Economic Studies 73*, 635–667.

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

Cox, D. R. (1961a). Prediction by exponentially weighted moving averages and related methods. *Journal of the Royal Statistical Society, Series B 23*, 414–422.

Cox, D. R. (1961b). Tests of seperate families of hypotheses. *Proceedings of the Berkeley Symposium 4*, 105–123.

Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics 13*, 253–263.

Drost, F. C. and T. E. Nijman (1993). Temporal aggregation of GARCH processes. *Econometrica 61*, 909–927.

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom inflation. *Econometrica 50*, 987–1007.

Engle, R. F. (2002). New frontiers for ARCH models. *Journal of Applied Econometrics 17*, 425–446.

Engle, R. F. and J. P. Gallo (2006). A multiple indicator model for volatility using intra daily data. *Journal of Econometrics 131*, 3–27.

Engle, R. F., D. F. Hendry, and J. F. Richard (1983). Exogeneity. *Econometrica 51*, 277–304.

Engle, R. F. and G. G. J. Lee (1999). A permanent and transitory component model of stock return volatility. In R. F. Engle and H. White (Eds.), *Cointegration, Causality, and Forecasting. A Festschrift in Honour of Clive W.J. Granger*, pp. 475–497. Oxford: Oxford University Press.

Engle, R. F. and J. Mezrich (1996). GARCH for groups. *Risk*, 36–40.

Falkenberry, T. N. (2002). High frequency data filtering. Technical report, Tick Data.

Fan, J. and Y. Wang (2007). Multi-scale jump and volatility analysis for high-frequency financial data. *Journal of the American Statistical Association 102*, 1349–1362.

Fiorentini, G., E. Sentana, and N. Shephard (2004). Likelihood-based estimation of latent generalised ARCH structures. *Econometrica 72*, 1481–1517.

Ghysels, E., A. Rubia, and R. Valkanov (2009). Multi-period forecasts of volatility: Direct, iterated and mixed-data approaches. Unpublished paper: Department of Economics, UNC at Chapel Hill.

Giacomini, R. and H. White (2006). Tests of conditional predictive ability. *Econometrica 74*, 1545–1578.

Giot, P. and S. Laurent (2004). Modelling daily value-at-risk using realized volatility and ARCH type models. *Journal of Empirical Finance 11*, 379–398.

Golub, G. H. and C. F. Van Loan (1989). *Matrix Computations*. Baltimore: John Hopkins University Press.

Hansen, P. R. and A. Lunde (2006). Realized variance and market microstructure noise (with discussion). *Journal of Business and Economic Statistics 24*, 127–218.

Heber, G., A. Lunde, N. Shephard, and K. K. Sheppard (2009). OMI's realised measure library. Version 0.1, Oxford-Man Institute, University of Oxford.

Hendry, D. F. (1995). *Dynamic Econometrics*. Oxford: Oxford University Press.

Jacod, J., Y. Li, P. A. Mykland, M. Podolskij, and M. Vetter (2009). Microstructure noise in the continuous case: the pre-averaging approach. *Stochastic Processes and Their Applications*. Forthcoming.

Lo, A. and C. MacKinlay (1990). The econometric analysis of nonsynchronous trading. *Journal of Econometrics 45*, 181–211.

Maheu, J. M. and T. H. McCurdy (2009). Do high-frequency measures of volatility improve forecasts of return distributions? Unpublished paper: Department of Economics, Toronto University.

Marcellino, M., J. H. Stock, and M. W. Watson (2006). A comparison of direct and iterated multistep ar methods for forecasting macroeconomic time series. *Journal of Econometrics 135*, 499–526.

Patton, A. J. (2009). Volatility forecast evaluation and comparison using imperfect volatility proxies. *Journal of Econometrics*. Forthcoming.

Patton, A. J. and K. K. Sheppard (2009a). Evaluating volatility forecasts. In T. G. Andersen, R. A. Davis, J. P. Kreiss, and T. Mikosch (Eds.), *Handbook of Financial Time Series*, pp. 801–838. Springer Verlag.

Patton, A. J. and K. K. Sheppard (2009b). Good volatility, bad volatility: signed jumps and persistence of volatility. Unpublished paper: Oxford-Man Institute, University of Oxford.

Rivers, D. and Q. H. Vuong (2002). Model selection for nonlinear dynamic models. *The Econometrics Journal 5*, 1–39.

Scholes, M. and J. Williams (1977). Estimating betas from nonsynchronous trading. *Journal of Financial Economics 5*, 309–27.

Shephard, N. and K. K. Sheppard (2009). Missing value based ARCH type model. Unpublished paper: Oxford-Man Institute, University of Oxford.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica 57*, 307–333.

Zhang, L. (2006). Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach. *Bernoulli 12*, 1019–1043.

Zhang, L., P. A. Mykland, and Y. Aït-Sahalia (2005). A tale of two time scales: determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association 100*, 1394–1411.