JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **25**: 197–231 (2010)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/jae.1158

REALISING THE FUTURE: FORECASTING WITH HIGH-FREQUENCY-BASED VOLATILITY (HEAVY) MODELS

NEIL SHEPHARD[a,b]* AND KEVIN SHEPPARD[b,a]

[a] *Oxford-Man Institute, University of Oxford, UK*
[b] *Department of Economics, University of Oxford, UK*

SUMMARY

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 realised 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 allows us to estimate the entire predictive distribution of returns. We also provide an analysis of missing data in the context of these models. Copyright © 2010 John Wiley & Sons, Ltd.

Received 10 July 2009; Revised 8 October 2009

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 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 'Oxford-Man Institute's realised library' of historical volatility statistics, constructed using high-frequency data. 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 interest to us. The first is realised variance, which was systematically studied by Andersen *et al*. (2001a) 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 *et al*. (2008). Alternatives to the realised kernel include the multiscale estimators of Zhang *et al*. (2005) and Zhang (2006) and the pre-averaging estimator of Jacod *et al*. (2009).[1]

* Correspondence to: Professor Neil Shephard, Oxford-Man Institute, Eagle House, Walton Well Road, Oxford OX2 6ED, UK. E-mail: neil.shephard@economics.ox.ac.uk

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

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 co-authors, 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 *et al*. (2001a,b, 2003, 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 *et al*. (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, HEAVY models of squared returns have some similarities with the component GARCH model of Engle and Lee (1999). However, HEAVY models are much easier 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 realised 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 as follows. 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 'Oxford-Man Institute's realised 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 discusses possible extensions. Section 6 draws some conclusions.

2. HEAVY MODELS

2.1. Assumed Data Structure

Our analysis will be based on daily financial returns:

$$r_1, r_2, \ldots, r_T$$

and a corresponding sequence of daily realised measures:

$$\mathrm{RM}_1, \mathrm{RM}_2, \ldots, \mathrm{RM}_T$$

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

DOI: 10.1002/jae

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 *et al.* (2009) and Barndorff-Nielsen and Shephard (2007).

The simplest realised measure is realised variance:

$$\mathrm{RM}_t = \sum_{0 \le t_{j-1,t} < t_{j,t} \le 1} x_{j,t}^2, \quad x_{j,t} = X_{t+t_{j,t}} - X_{t+t_{j-1,t}} \tag{1}$$

where $t_{j,t}$ are the normalised times of trades or quotes (or a subset of them) on the tth 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 tth day. It was formalised econometrically by Andersen *et al.* (2001a) and Barndorff-Nielsen and Shephard (2002). In practice, market microstructure noise plays an important part and the above authors use 1- to 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 maximum 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: pre-averaging (Jacod *et al.*, 2009), multiscale (Zhang, 2006; Zhang *et al.*, 2005) and realised kernel (Barndorff-Nielsen *et al.*, 2008).[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 that there is no adjustment for mean and the sums are not scaled by their sample size):

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

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

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

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 *et al.* (2009a), 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.[3]

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

[3] We could also have included jump robust measures, which typically lead to an increase in predictive power. See, for example, Andersen *et al.* (2007) and Barndorff-Nielsen and Shephard (2006). This has virtues but then we would also need to forecast these terms in making multistep-ahead forecasts. See the work of Engle and Gallo (2006) in this context.

DOI: 10.1002/jae

2.2. Definitions

We will write a sequence of daily returns as $r_1, r_2, \ldots, r_T$, while we will use $\mathcal{F}^{\mathrm{LF}}_{t-1}$ 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

$$\mathrm{var}(r_t|\mathcal{F}^{LF}_{t-1}) = \sigma_t^2 = \omega_G + \alpha_G r_{t-1}^2 + \beta_G \sigma_{t-1}^2$$

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

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

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

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

$$\left\{ \begin{array}{l} \mathrm{var}(r_t|\mathcal{F}^{\mathrm{HF}}_{t-1}) \\ \mathrm{E}(\mathrm{RM}_t|\mathcal{F}^{\mathrm{HF}}_{t-1}) \end{array} \right\}, \quad t = 2, 3, \ldots, T$$

where $\mathcal{F}^{\mathrm{HF}}_{t-1}$ 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

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

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

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.[4] 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.

Note that (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 stationary HEAVY-r model and GARCH models as

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

[4] Of course, the most basic realised measure is the squared daily return, so in some sense the GARCH model is a HEAVY-r model. This point was made to us by Frank Diebold. From this point of view one might think that a HEAVY-r model is a 'turbo-charged' GARCH.

Another interpretation of HEAVY models is that one could unravel $\mathrm{var}(r_t|\mathcal{F}^{\mathrm{HF}}_{t-1})$ in terms of many lags of RM, which relates it directly back in some sense to the forecasting models considered by Andersen, Bollerslev and Diebold in various papers in which they focused on forecasting RM using lags of RM.

DOI: 10.1002/jae

In applied work we will typically estimate β to be around 0.6–0.7 and ω to be small. Thus 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 it has much more memory and thus it averages more data points.

Note that, unlike GARCH models, the HEAVY-r model has no feedback and so the properties of the realised measures determine the properties of $\text{var}(r_t|\mathcal{F}_{t-1}^{\text{HF}})$.

The predictive model for the times series of realised measures is not novel. The work of Andersen *et al*. (2001a,b, 2003, 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), based on 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. Lu (2005) looked at extensions of GARCH models allowing the inclusion of lagged realised variance. He provides extensive empirical analysis of these GARCHX models.

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). 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 I, 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. Further, their out-of-sample comparison was based only on 217 observations, which makes their analysis somewhat noisy. Perhaps these two features 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 $h_t = \omega + \alpha \text{RM}_{t-1} + \beta h_{t-1}$ 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 perform 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.

DOI: 10.1002/jae

Table I. A description of the 'OMI's realised 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 27 March 2009

Asset	Med dur	Start date	T	Asset	Med dur	Start date	T
Dow Jones Industrials	2	2-1-1996	3278	MSCI Australia	60	2-12-1999	2323
Nasdaq 100	15	2-1-1996	3279	MSCI Belgium	60	1-7-1999	2442
S&P 400 Midcap	15	2-1-1996	3275	MSCI Brazil	60	4-10-2002	1587
S&P 500	15	2-1-1996	3284	MSCI Canada	60	12-2-2001	2013
Russell 3000	15	2-1-1996	3279	MSCI Switzerland	60	9-6-1999	2434
Russell 1000	15	2-1-1996	3279	MSCI Germany	60	1-7-1999	2448
Russell 2000	15	2-1-1996	3281	MSCI Spain	60	1-7-1999	2423
CAC 40	30	2-1-1996	3322	MSCI France	60	1-7-1999	2455
FTSE 100	15	20-10-1997	2862	MSCI UK	60	8-6-1999	2451
German DAX	15	2-1-1996	3317	MSCI Italy	60	1-7-1999	2437
Italian MIBTEL	60	3-7-2000	2194	MSCI Japan	15	2-12-1999	2240
Milan MIB 30	60	2-1-1996	3310	MSCI South Korea	60	3-12-1999	2263
Nikkei 250	60	5-1-1996	3177	MSCI Mexico	60	4-10-2002	1612
Spanish IBEX	5	2-1-1996	3288	MSCI Netherlands	60	1-7-1999	2454
S&P TSE	15	31-12-1998	2546	MSCI World	60	11-2-2001	2101
British pound	2	3-1-1999	2584				
Euro	1	3-1-1999	2600				
Swiss franc	3	3-1-1999	2579				
Japanese yen	2	3-1-1999	2599				

Bollerslev *et al*. (2009) model multiple volatility indicators and daily returns, where the return model has a conditional variance which is contemporaneous realised variance.

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

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

This could be thought of as a GARCHX type model, but that name suggests it is the squared returns which drives the model, whereas 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 on the rare occasions that γ_X is estimated to be statistically significant. 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

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

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 with respect to $\mathcal{F}_{t-1}^{\mathrm{HF}}$.[5]

[5] 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.

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

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

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

The memory in the HEAVY model is governed by

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

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 interest but allows some theoretical insights.

Assume that $\alpha_R, \beta_R, \beta \in [0, 1)$ and $\alpha_R + \beta_R < 1$. Define $u_t = r_t^2 - h_t$, $u_{Rt} = \mathrm{RM}_t - \mu_t$, which under the model are martingale difference sequences with respect to $\mathcal{F}_{t-1}^{\mathrm{HF}}$. We can write out the process for the r_t^2 from a HEAVY model as

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

where L is the lag operator. Therefore

$$(1 - \beta L)r_t^2 = \omega + \alpha \mathrm{RM}_{t-1} + (1 - \beta L)u_t$$

Likewise:

$$\{1 - (\alpha_R + \beta_R)L\}\mathrm{RM}_t = \omega_R + (1 - \beta_R L)u_{Rt}, \quad u_{Rt} = \mathrm{RM}_t - \mu_t$$

Combining delivers the result

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

where

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

If we assume that

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

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

$$
\begin{pmatrix} r_t^2 \\ \mathrm{RM}_t \end{pmatrix} = \begin{pmatrix} \omega \\ \omega_R \end{pmatrix} + \begin{pmatrix} \beta & \alpha \\ 0 & \alpha_R+\beta_R \end{pmatrix} \begin{pmatrix} r_{t-1}^2 \\ \mathrm{RM}_{t-1} \end{pmatrix} + \begin{pmatrix} (1-\beta L)u_t \\ (1-\beta_R L)u_{Rt} \end{pmatrix}
$$

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

$$
\Delta r_t^2 = \omega + \alpha(\mathrm{RM}_{t-1} - \gamma r_{t-1}^2) + (1-\beta L)u_t, \quad \text{where} \quad \gamma = \frac{1-\beta}{\alpha} \tag{7}
$$

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}).

The weak GARCH(2,2) representation has some similarities with the component model of Engle and Lee (1999, equations (2.4) and (2.5)), which models

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

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.

2.3.3 Momentum

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, although the realised measure still drives volatility. The component model of Engle and Lee (1999) is also a weak GARCH(1,2) if $\rho_C=0$. The sophisticated model of Engle and Gallo (2006) is capable of generating momentum effects, of course.

DOI: 10.1002/jae

If $\beta_R = \beta$ then

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

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

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

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

2.3.4 Integrated HEAVY Models

The marginal process (8) can be rewritten 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 one, 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

$$\text{var}\ (r_t|\mathcal{F}_{t-1}^{\text{HF}}) = h_t = \omega + \alpha \text{RM}_{t-1} + \beta h_{t-1}, \quad \omega, \alpha \geq 0, \quad \beta \in [0,1) \tag{9}$$
$$\text{E}\ (\text{RM}_t|\mathcal{F}_{t-1}^{\text{HF}}) = \mu_t = \alpha_R \text{RM}_{t-1} + (1-\alpha_R)\mu_{t-1}, \quad \alpha_R \in [0,1) \tag{10}$$

We call this the 'integrated HEAVY model'. We will see later that this very simple model can generate reliable multiperiod forecasts.

2.3.5 Iterative Multistep-Ahead Forecasts

Multistep-ahead forecasts of volatility are very important for asset allocation or risk assessment since these tasks are 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.

DOI: 10.1002/jae

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

$$\begin{pmatrix} \mathrm{var}(r_{t+s}|\mathcal{F}_{t-1}^{\mathrm{HF}}) \\ \mathrm{E}(\mathrm{RM}_{t+s}|\mathcal{F}_{t-1}^{\mathrm{HF}}) \end{pmatrix} = \begin{pmatrix} h_{t+s|t-1} \\ \mu_{t+s|t-1} \end{pmatrix} = (I + B + \ldots + B^s)w + B^{s+1}\begin{pmatrix} h_{t-1} \\ \mu_{t-1} \end{pmatrix} \tag{11}$$

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

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

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

$$\mathrm{var}(r_t + r_{t+1} + \ldots + r_{t+s}|\mathcal{F}_{t-1}^{\mathrm{HF}}) = \sum_{j=0}^{s} \mathrm{var}(r_{t+j}|\mathcal{F}_{t-1}^{\mathrm{HF}})$$

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

2.3.6 Targeting 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

$$\begin{aligned} \mu_t &= \omega_R + \alpha_R \mathrm{RM}_{t-1} + \beta_R \mu_{t-1}, \quad \alpha_R, \beta_R \geq 0, \quad \alpha_R + \beta_R < 1, \\ &= \mu_R(1 - \alpha_R - \beta_R) + \alpha_R \mathrm{RM}_{t-1} + \beta_R \mu_{t-1} \end{aligned} \tag{12}$$

so that $\mathrm{E}(\mathrm{RM}_t) = \mu_R$. For the HEAVY-r equation it is less clear since the realised 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

$$\begin{aligned} h_t &= \omega_r + \alpha \mathrm{RM}_{t-1} + \beta h_{t-1} \\ &= \mu(1 - \alpha\kappa - \beta) + \alpha \mathrm{RM}_{t-1} + \beta h_{t-1}, \quad \kappa = \frac{\mu_R}{\mu} \leq 1 \end{aligned} \tag{13}$$

Taken together we call (13) and (12) the 'targeting parameterisation' for the HEAVY model.
This parameterisation of the HEAVY model has the virtue that it is possible to use the estimators[6]

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

of μ_R, μ and κ. Thus this reparameterisation is the HEAVY extension of variance targeting introduced by Engle and Mezrich (1996). When these estimators are plugged into the quasi-likelihood functions it makes optimisation easier, as the dimension is smaller, but it does alter the resulting asymptotic standard errors. This is discussed in the next subsection.

[6] 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.

DOI: 10.1002/jae

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:

$$\mathrm{var}(r_t|\mathcal{F}_{t-1}^{\mathrm{HF}}) = h_t = \omega + \alpha \mathrm{RM}_{t-1} + \beta h_{t-1},$$
$$\mathrm{E}(\mathrm{RM}_t|\mathcal{F}_{t-1}^{\mathrm{HF}}) = \mu_t = \omega_R + \alpha_R \mathrm{RM}_{t-1} + \beta_R \mu_{t-1}$$

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 *et al.*, 2007).

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

$$\log Q_1(\omega, \psi) = \sum_{t=2}^{T} l_t^r, \quad \text{where} \quad l_t^r = -\frac{1}{2}(\log h_t + r_t^2/h_t), \quad \psi = (\alpha, \beta)' \tag{14}$$

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

$$\log Q_2(\omega_R, \psi_R) = \sum_{t=2}^{T} l_t^{\mathrm{RM}} \quad \text{where} \quad l_t^{\mathrm{RM}} = -\frac{1}{2}(\log \mu_t + \mathrm{RM}_t/\mu_t), \quad \psi_R = (\alpha_R, \beta_R)' \tag{15}$$

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

In inference we will regard the parameters as having no link between the HEAVY-r and HEAVY-RM models, i.e. (ω, ψ) and (ω_R, ψ_R) are variation free (e.g. Engle *et al.*, 1983), which we will see in the next subsection is important for inference. It then follows that equation-by-equation optimisation is all that is necessary to maximise the quasi-likelihood. This is convenient as existing GARCH type code can simply be used in this context. We will write $\theta = (\omega, \psi', \omega_R, \psi_R')'$ and the resulting maximum of the quasi-likelihoods as $\widehat{\theta}$.

The alternative targeting parameterisation has

$$h_t = \mu(1 - \alpha\kappa - \beta) + \alpha \mathrm{RM}_{t-1} + \beta h_{t-1}, \quad \kappa = \frac{\mu_R}{\mu} \le 1,$$
$$\mu_t = \mu_R(1 - \alpha_R - \beta_R) + \alpha_R \mathrm{RM}_{t-1} + \beta_R \mu_{t-1}, \quad \alpha_R + \beta_R < 1$$

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

$$\widehat{\mu} = \frac{1}{T} \sum_{t=1}^{T} r_t^2 \quad \text{and} \quad \widehat{\mu}_R = \frac{1}{T} \sum_{t=1}^{T} \mathrm{RM}_t$$

and then we compute

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

DOI: 10.1002/jae

This reduces the dimension of the optimisations by one each time; this 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 (14) and (15). We stack the scores so that

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

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

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

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

$$\sqrt{T}(\widehat{\theta} - \theta^*) \xrightarrow{d} N(0, \mathcal{J}^{-1}\mathcal{I}J^{-1'})$$

where the Hessian is

$$\mathcal{J} = p \lim_{T \longrightarrow \infty} \widehat{\mathcal{J}}_T, \quad \text{where} \quad \widehat{\mathcal{J}}_T = -\frac{1}{T} \begin{pmatrix} \sum_{t=2}^{T} \frac{\partial^2 l_t^r}{\partial \lambda \partial \lambda'} & 0 \\ 0 & \sum_{t=2}^{T} \frac{\partial^2 l_t^{\mathrm{RM}}}{\partial \lambda_R \partial \lambda_R'} \end{pmatrix} \tag{16}$$

and

$$\mathcal{I} = p \lim_{T \longrightarrow \infty} \widehat{\mathcal{I}}_T, \quad \text{where} \quad \widehat{\mathcal{I}}_T = \frac{1}{T} \sum_{t=2}^{T} m_t(\widehat{\theta}) m_t(\widehat{\theta})' \tag{17}$$

The block diagonality of (16) is due to the variation-free property of the parameters, while it is not necessary to use an HAC estimator in (17) 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 discussed extensively in Cipollini *et al*. (2007).

The most important implication of the block diagonality of the Hessian (16) 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 targeting parameterisation then the moment conditions change to

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

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:

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

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

2.4.3 Non-nested Tests

One natural way to assess the forecasting power of the HEAVY model is to compare it to that generated by the GARCH model. This can be assessed at distinct horizons by comparing the performance using the QLIK loss function:

$$\text{loss}(r_{t+s}^2, \widetilde{\sigma}_{t+s|t-1}^2) = \frac{r_{t+s}^2}{\widetilde{\sigma}_{t+s|t-1}^2} - \log\left(\frac{r_{t+s}^2}{\widetilde{\sigma}_{t+s|t-1}^2}\right) - 1, \quad s = 0, 1, \ldots, S \tag{18}$$

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

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

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

The temporal average $(s+1)$-step-ahead relative loss between a HEAVY and GARCH model will be

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

where

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

DOI: 10.1002/jae

Here $h_{t+s|t-1}$ is the forecast from the HEAVY model, $\sigma^2_{t+s|t}$ 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 $L_s = \mathrm{E}(L_{t,s})$, $s = 0, 1, \ldots, S$, 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 $s = 1, 2, \ldots, S$. 'Weakly forecast-dominates' means that $L_s \leq 0$ for all $s = 1, 2, \ldots, S$ 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.[7]

The above scheme can be implemented 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

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

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

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 *et al*. (2006) and Ghysels *et al*. (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 *et al*. (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:

$$\log Q_{1,s}(\omega_s, \psi_s) = \sum_{t=2}^{T} l^r_{t,s}, \quad \text{where} \quad l^r_{t,s} = -\frac{1}{2}\left(\log h_{t+s|t-1} + \frac{r^2_{t+s}}{h_{t+s|t-1}}\right), \psi_s = (\alpha_s, \beta_s)',$$

$$\log Q_{2,s}(\omega_{R,s}, \psi_{R,s}) = \sum_{t=2}^{T} l^{\mathrm{RM}}_{t,s} \quad \text{where} \quad l^{\mathrm{RM}}_{t,s} = -\frac{1}{2}\left(\log \mu_{t+s|t-1} + \frac{\mathrm{RM}_{t+s}}{\mu_{t+s|t-1}}\right), \psi_{R,s} = (\alpha_{R,s}, \beta_{R,s})'$$

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

[7] In the context of forecasting this is related to Diebold and Mariano (1995). 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.

DOI: 10.1002/jae

horizon tuning for the HEAVY-RM forecasts.[8] The same exercise can be carried out for a GARCH model.

2.4.5 Bootstrapping

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

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

or

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

A simple way of carrying this out is via a model-based bootstrap. We use the representation $r_t = \zeta_t h_t^{1/2}$, $\mathrm{RM}_t = \eta_t \mu_t$, $\mathrm{E}(\zeta_t^2|\mathcal{F}_{t-1}^{\mathrm{HF}}) = 1$, $\mathrm{E}(\eta_t|\mathcal{F}_{t-1}^{\mathrm{HF}}) = 1$ and then assume that $(\zeta_t, \eta_t)' \overset{\text{i.i.d.}}{\sim} F_{\zeta,\eta}$. Typically these bivariate variables will be contemporaneously correlated. For equities we would expect a sharp negative correlation reflecting statistical leverage. 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[9]

$$\widehat{\zeta}_t = r_t/h_t^{1/2}, \quad \widehat{\eta}_t = (\mathrm{RM}_t/\mu_t)^{1/2}, \quad t = 2, 3, \ldots, T \tag{21}$$

and computing the empirical distribution function $\widehat{F}_{\zeta,\eta}$. Then we can sample with replacement pairs from this population,[10] which can then be used to drive a simulated joint path of the pair $(r_t, \mathrm{RM}_t)'$, $(r_{t+1}, \mathrm{RM}_{t+1})'$, ..., $(r_{t+s}, \mathrm{RM}_{t+s})'$. 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.

3. OMI'S REALISED LIBRARY 0.1

3.1. A List of Assets and Data Cleaning

This paper uses the database 'Oxford-Man Institute's realised library' version 0.1, which has been produced by Heber *et al.* (2009).[11]

The version 0.1 of the library currently starts on 2 January 1996 and finishes on 27 March 2009. Some of the series are available throughout this period, but quite a number start after 1996, as detailed in Table I. 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. Table I gives the basic features of

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

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

[10] There may be 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. However, we have not explored that here.

[11] Available at http://realized.oxford-man.ox.ac.uk

DOI: 10.1002/jae

the data used to compute the library, indicating the frequency of the base data used in the calculations.

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 are regarded as being of unacceptably low quality 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. As a result, for example, 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.[12] Realised kernels are computed in tick time using every available data point, after cleaning. Data cleaning is discussed in our data appendix at the end of this paper.

3.2. Summary Statistics for the Library

Table II 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. They show much higher standard deviations for squared returns than for their realised measure cousins. The ACF figures are the serial correlation coefficients at one lag. This shows the modest degree of serial correlation of squared returns and much higher numbers of the realised variances and realised kernels. These are the expected results.

[12] 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).

DOI: 10.1002/jae

Table II. Calculations use 100 times differences of the log price (i.e. roughly percent 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. SD is the daily standard deviation of percent daily returns or realised measure. The same data are used to compute the ACFs (serial correlations) at 1 lag

Asset	r_t^2			Realized variance			Realized kernel		
	Avol	SD	ACF$_1$	Avol	SD	SCF$_1$	Avol	SD	ACF$_1$
Dow Jones Industrials	19.4	4.81	0.125	15.2	1.94	0.663	15.0	1.95	0.655
Nasdaq 100	28.1	8.35	0.180	17.8	2.22	0.664	18.7	2.52	0.646
S&P 400 Midcap	21.7	5.68	0.260	13.5	1.90	0.800	13.7	1.96	0.799
S&P 500	20.8	5.46	0.209	15.5	2.09	0.699	15.9	2.14	0.701
Russell 3000	20.3	5.32	0.127	14.3	1.86	0.694	14.5	1.90	0.697
Russell 1000	20.4	5.38	0.125	14.7	1.91	0.692	14.9	1.94	0.695
Russell 2000	23.3	6.02	0.313	13.2	1.85	0.715	13.4	1.96	0.720
CAC 40	23.7	5.95	0.236	18.1	2.18	0.662	18.3	2.21	0.669
FTSE 100	20.7	4.66	0.229	15.2	1.62	0.645	15.6	1.74	0.620
German DAX	25.1	6.57	0.163	21.1	3.10	0.659	21.3	3.22	0.626
Italian MIBTEL	20.1	5.07	0.218	13.1	1.34	0.665	13.7	1.52	0.662
Milan MIB 30	23.2	5.69	0.214	16.5	1.84	0.624	17.0	1.99	0.615
Nikkei 250	24.9	6.96	0.241	16.0	1.37	0.691	16.5	1.48	0.668
Spanish IBEX	23.7	6.57	0.295	16.7	1.76	0.639	16.5	1.73	0.655
S&P TSE	20.9	5.54	0.292	14.1	1.82	0.785	14.3	1.89	0.774
MSCI Australia	16.4	3.05	0.229	8.8	0.53	0.763	9.1	0.57	0.749
MSCI Belgium	23.4	10.5	0.159	16.4	1.66	0.718	16.1	1.84	0.684
MSCI Brazil	43.7	24.3	0.155	28.5	6.30	0.796	29.6	7.21	0.749
MSCI Canada	19.5	5.05	0.320	12.6	1.67	0.819	13.1	1.88	0.761
MSCI Switzerland	20.6	5.25	0.330	14.5	1.44	0.727	14.5	1.56	0.700
MSCI Germany	25.7	6.94	0.163	21.1	3.10	0.677	20.8	2.99	0.692
MSCI Spain	24.0	6.08	0.225	17.5	1.84	0.690	17.6	1.92	0.676
MSCI France	23.9	6.29	0.238	18.2	2.23	0.682	18.4	2.32	0.669
MSCI UK	20.0	4.95	0.233	15.6	1.84	0.615	15.7	1.89	0.649
MSCI Italy	21.4	5.35	0.247	16.0	1.82	0.672	16.2	1.93	0.670
MSCI Japan	23.7	6.40	0.273	14.2	1.27	0.746	14.4	1.26	0.755
MSCI South Korea	32.0	9.63	0.131	21.6	2.61	0.700	21.9	2.80	0.682
MSCI Mexico	29.6	11.8	0.144	16.3	2.59	0.675	17.5	2.87	0.678
MSCI Netherlands	23.9	6.14	0.281	17.7	2.09	0.733	17.9	2.25	0.716
MSCI World	17.7	4.22	0.250	13.1	1.44	0.766	13.6	1.68	0.691
British pound	9.2	0.75	0.215	9.8	0.51	0.876	9.4	0.51	0.879
Euro	10.4	0.79	0.103	11.1	0.45	0.668	10.5	0.45	0.658
Swiss franc	11.0	0.91	0.133	11.6	0.39	0.690	10.8	0.38	0.650
Japanese yen	10.9	1.32	0.134	11.6	0.64	0.698	11.2	0.63	0.696

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 targeting GARCH:

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

DOI: 10.1002/jae

and the non-targeting GARCHX models. The HEAVY models are set up in their targeting parameterisation:

$$\mu_t = \mu_R(1 - \alpha_R - \beta_R) + \alpha_R \mathrm{RM}_{t-1} + \beta_R \mu_{t-1}, \quad \alpha_R + \beta_R < 1,$$
$$h_t = \mu(1 - \alpha\kappa - \beta) + \alpha \mathrm{RM}_{t-1} + \beta h_{t-1}, \quad \kappa = \frac{\mu_R}{\mu} \leq 1, \quad \alpha + \beta < 1$$

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 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 for imposing a unit root on the HEAVY-RM model, in which case we model

$$\begin{aligned} \mu_t &= \alpha_R \mathrm{RM}_{t-1} + (1 - \alpha_R)\mu_{t-1}, \quad \alpha_R < 1 \\ h_t &= \omega + \alpha RM_{t-1} + \beta h_{t-1}, \quad \alpha + \beta < 1 \end{aligned} \tag{22}$$

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

The results are presented in some detail in Table III 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 0.35–0.45 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, as can be seen from Table IV.

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 two 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 indices.

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 typically there are small falls in the estimates.

Table IV 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 V shows the HEAVY's model's average in sample iterated multistep-ahead QLIK loss compared to the GARCH model, using the methodology discussed above ('Iterative Multistep-Ahead Forecasts'). Here the parameters are estimated using the quasi-likelihood, which means

DOI: 10.1002/jae

Table III. 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 estimated using tracking parameterisation. Integrated models are IGARCH and Int-HEAVY-RM

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_G	α_R
Dow Jones Industrials	0.407	0.737	0.407	0.737	0.000	0.082	0.912	0.411	0.567	0.062	0.336
Nasdaq 100	0.730	0.658	0.439	0.744	0.051	0.081	0.916	0.428	0.567	0.063	0.349
S&P 400 Midcap	0.848	0.641	0.270	0.794	0.083	0.100	0.886	0.392	0.603	0.073	0.333
S&P 500	0.378	0.773	0.378	0.773	0.000	0.076	0.918	0.417	0.564	0.054	0.340
Russell 3000	0.448	0.747	0.448	0.747	0.000	0.081	0.911	0.403	0.574	0.059	0.313
Russell 1000	0.397	0.768	0.397	0.768	0.000	0.078	0.916	0.402	0.577	0.057	0.315
Russell 2000	0.949	0.678	0.244	0.812	0.102	0.106	0.885	0.387	0.622	0.077	0.322
CAC 40	0.526	0.674	0.526	0.674	0.000	0.081	0.917	0.417	0.573	0.067	0.350
FTSE 100	0.613	0.656	0.613	0.656	0.000	0.105	0.892	0.441	0.556	0.085	0.369
German DAX	0.447	0.673	0.447	0.673	0.000	0.093	0.903	0.457	0.536	0.075	0.376
Italian MIBTEL	0.806	0.630	0.806	0.630	0.000	0.107	0.889	0.512	0.486	0.080	0.436
Milan MIB 30	0.496	0.748	0.342	0.779	0.047	0.102	0.895	0.484	0.518	0.075	0.417
Nikkei 250	0.508	0.772	0.508	0.772	0.000	0.079	0.905	0.346	0.641	0.065	0.295
Spanish IBEX	0.640	0.669	0.481	0.713	0.035	0.113	0.885	0.393	0.603	0.084	0.343
S&P TSE	0.643	0.692	0.637	0.693	0.002	0.067	0.930	0.362	0.635	0.054	0.324
Index's median	0.526	0.678	0.447	0.744	0.000	0.082	0.905	0.411	0.573	0.067	0.340
MSCI Australia	0.214	0.645	0.976	0.668	0.043	0.098	0.894	0.324	0.670	0.069	0.292
MSCI Belgium	0.769	0.568	0.374	0.692	0.093	0.143	0.854	0.399	0.608	0.105	0.359
MSCI Brazil	0.662	0.652	0.661	0.653	0.001	0.096	0.876	0.433	0.536	0.071	0.375
MSCI Canada	0.515	0.765	0.485	0.769	0.009	0.074	0.914	0.364	0.630	0.060	0.329
MSCI Switzerland	0.699	0.638	0.699	0.638	0.000	0.131	0.860	0.474	0.508	0.093	0.425
MSCI Germany	0.568	0.592	0.568	0.592	0.000	0.107	0.885	0.461	0.529	0.083	0.388
MSCI Spain	0.589	0.659	0.589	0.659	0.000	0.090	0.907	0.417	0.579	0.067	0.365
MSCI France	0.596	0.628	0.596	0.628	0.000	0.090	0.908	0.453	0.543	0.074	0.386
MSCI UK	0.582	0.616	0.582	0.616	0.000	0.110	0.886	0.456	0.543	0.086	0.393
MSCI Italy	0.583	0.659	0.583	0.659	0.000	0.100	0.896	0.537	0.462	0.075	0.467
MSCI Japan	0.741	0.720	0.741	0.720	0.000	0.088	0.902	0.459	0.533	0.075	0.387
MSCI South Korea	0.765	0.661	0.765	0.661	0.000	0.071	0.928	0.432	0.564	0.059	0.392
MSCI Mexico	0.872	0.711	0.723	0.725	0.032	0.095	0.885	0.364	0.624	0.068	0.328
MSCI Netherlands	0.538	0.678	0.538	0.678	0.000	0.105	0.889	0.453	0.541	0.084	0.396
MSCI World	0.339	0.798	0.339	0.798	0.000	0.084	0.910	0.377	0.610	0.068	0.340
MSCI's median	0.596	0.659	0.589	0.661	0.000	0.096	0.894	0.433	0.543	0.074	0.386
British pound	0.162	0.810	0.162	0.810	0.000	0.042	0.950	0.283	0.699	0.035	0.264
Euro	0.055	0.936	0.034	0.947	0.013	0.030	0.969	0.247	0.746	0.028	0.223
Swiss franc	0.046	0.948	0.045	0.947	0.002	0.027	0.971	0.239	0.748	0.024	0.220
Japanese yen	0.173	0.772	0.173	0.772	0.000	0.048	0.934	0.398	0.552	0.035	0.341
Currency's median	0.109	0.873	0.104	0.879	0.001	0.036	0.959	0.265	0.722	0.031	0.244

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 targeting parameterisation. The right-hand side shows the corresponding result for the 'integrated HEAVY' model, which is discussed in (22). Recall that negative t-statistics indicate a statistically significant preference for HEAVY models. The final column examines the log-likelihood loss from excluding the smoothing parameter from the HEAVY-RM model ($\beta = 0$). In all cases the decrease in log-likelihood is substantial, indicating that averaging over the most recent 4 or 5 days is highly desirable.

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

Asset	Compare to extended HEAVY-r		Impose unit root		No momentum
	HEAVY−r	GARCH	GARCH	HEAVY-RM	$\beta = 0$
Dow Jones Industrials	0.0	−199.5	−48.4	−19.5	−56.6
Nasdaq 100	−15.9	−108.5	−31.1	−14.4	−72.9
S&P 400 Midcap	−64.6	−61.8	−61.4	−11.0	−88.8
S&P 500	0.0	−211.1	−50.6	−17.9	−67.2
Russell 3000	0.0	−187.3	−49.8	−21.1	−61.3
Russell 1000	0.0	−186.3	−45.3	−20.0	−61.9
Russell 2000	−163.2	−64.9	−57.4	−13.3	−131.1
CAC 40	0.0	−149.1	−30.8	−14.5	−67.3
FTSE 100	0.0	−125.5	−32.4	−12.3	−55.2
German DAX	0.0	−153.4	−47.0	−16.0	−63.7
Italian MIBTEL	0.0	−141.2	−40.5	−9.9	−38.1
Milan MIB 30	−16.5	−100.7	−48.3	−13.0	−75.6
Nikkei 250	0.0	−116.5	−64.5	−9.9	−84.6
Spanish IBEX	−9.3	−113.9	−59.0	−12.1	−78.4
S&P TSE	−0.0	−120.8	−17.3	−5.6	−72.3
Index's median	0.0	−125.5	−48.3	−13.3	−67.3
MSCI Australia	−6.6	−96.6	−31.2	−3.9	−55.8
MSCI Belgium	−22.7	−66.2	−60.2	−4.1	−56.9
MSCI Brazil	0.0	−60.2	−35.5	−7.1	−23.6
MSCI Canada	−0.4	−75.0	−22.9	−4.4	−56.7
MSCI Switzerland	0.0	−153.4	−65.8	−9.1	−32.7
MSCI Germany	0.0	−136.9	−45.0	−10.7	−44.5
MSCI Spain	0.0	−106.7	−31.5	−7.5	−44.5
MSCI France	0.0	−158.3	−27.7	−9.4	−47.1
MSCI UK	0.0	−134.3	−37.1	−9.3	−44.5
MSCI Italy	0.0	−154.7	−38.3	−8.7	−35.4
MSCI Japan	0.0	−111.8	−33.7	−6.2	−28.0
MSCI South Korea	0.0	−118.6	−15.1	−4.1	−43.5
MSCI Mexico	−3.4	−61.2	−36.5	−3.5	−43.1
MSCI Netherlands	0.0	−117.8	−40.8	−7.6	−46.8
MSCI World	0.0	−92.9	−25.6	−6.3	−104.0
MSCI's median	0.0	−111.8	−35.5	−7.1	−44.5
British pound	0.0	−50.4	−16.0	−1.8	−28.3
Euro	−2.7	−18.5	−6.0	−1.6	−44.6
Swiss franc	−0.1	−33.0	−5.9	−1.7	−40.5
Japanese yen	0.0	−67.4	−38.6	−8.4	−26.1
Currency's median	−0.0	−41.7	−11.0	−1.8	−34.4

The results are striking. They shows that in sample and pointwise the standard HEAVY model forecast dominates the GARCH model, but that the out-performance gets weaker as the forecast horizon increases. The integrated HEAVY model performs slightly more poorly than the unconstrained HEAVY model.

This picture is remarkably stable across assets with two counter-examples: the mid-cap series Russell 2000 and the S&P 400 Midcap. These have lower quasi-likelihoods and this under-performance continues when applied at multistep-ahead periods.

DOI: 10.1002/jae

Table V. In-sample likelihood ratio tests for losses generated by HEAVY and GARCH models. Negative values favour HEAVY models. Both models are estimated using the quasi-likelihood, i.e. tuned to one-step-ahead predictions

Asset	t-statistic for non-nested LR tests for iterative forecasts											
	Horizon $h = s + 1$: HEAVY model						Horizon $h = s + 1$: Int HEAVY model					
	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

4.2. Direct Forecasting

The above estimation strategy fixes the parameters at the QMLE values and uses these to iterate through the multistep-ahead forecast formula to produce multistep-ahead forecasts and corresponding estimated losses. We call this indirect estimation. We now move on to a second approach, which allows different parameters to be used at different forecast horizon, maximising the multistep-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

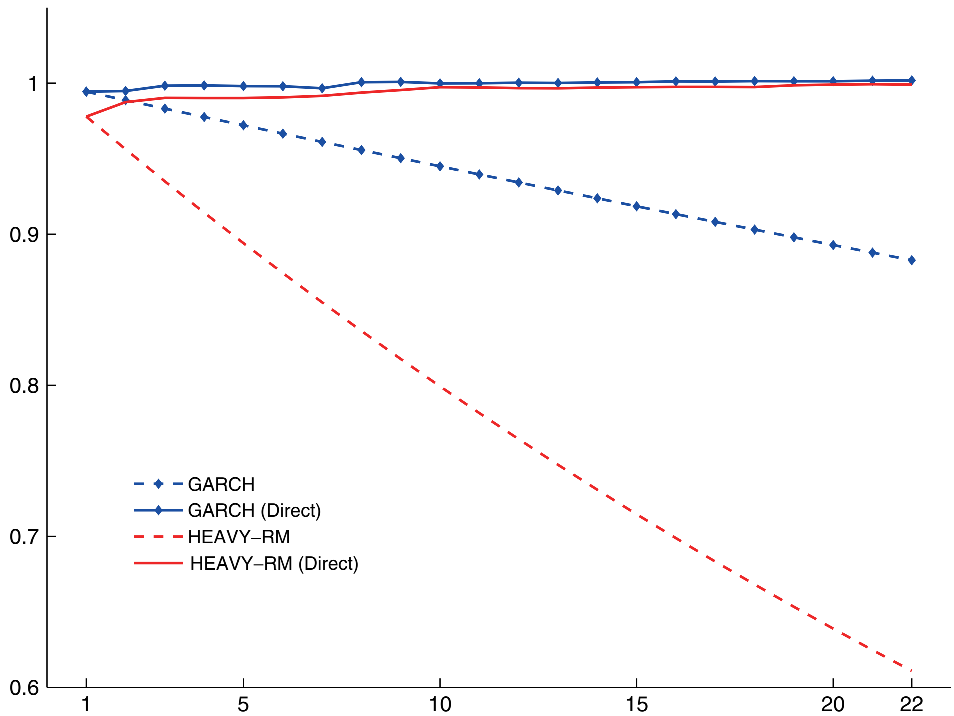

Figure 1. Direct and indirect method for Dow Jones Industrial case. Estimates of $(\alpha_R + \beta_R)^{s+1}$ and $(\alpha_G + \beta_G)^{s+1}$ drawn against forecast horizon $s + 1$. The figure shows the impact of strong mean reversion on the HEAVY-RM model when it is indirectly estimated and the weaker mean reversion in the direct case. This figure is available in color online at www.interscience.wiley.com/journal/jae

memory in the HEAVY-RM and GARCH models:

$$(\alpha_R + \beta_R)^{s+1}, \quad \text{and} \quad (\alpha_G + \beta_G)^{s+1} \tag{23}$$

plotted against s when we use the quasi-likelihood, which is tuned to perform well at one step. We see that, although the estimated values of these parameters are not very different, at long lags the difference becomes magnified. By the time we are 1 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; the model wants around 90% of the weight to come from the recent data and only 10% to come from the unconditional mean.

Figure 1 also shows the profile of (23) now for the directly estimated parameters, tuning each estimator to the appropriate forecast horizon. When we do this the persistence of 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, the rise in β_R was sharper, 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 VI 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 out-performance of HEAVY models appears for nearly all assets and forecast horizons. The results are given in Table VI.

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

Asset	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		
	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

DOI: 10.1002/jae

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

Asset	Pointwise comparison						Cumulative comparison					
	Direct HEAVY vs. Direct GARCH			Int. HEAVY vs. GARCH			Direct HEAVY vs. Direct GARCH			Int. HEAVY vs. GARCH		
	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

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 (1008 observations) and parameters were updated daily. Forecasts were then produced for one through 22 steps ahead. Table VII shows 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 optimised by fitting the realised measure (HEAVY-RM) or squared return (GARCH) models at the forecasting horizon. All HEAVY models used

DOI: 10.1002/jae

the same HEAVY-r model, which was optimised for the one-step horizon. The second compares the performance of the Integrated HEAVY-RM specification with a standard GARCH, where both sets of parameters were optimised for one-step prediction. The standard HEAVY model based on one-step tuning is not included since the memory parameter chosen was often implausibly small. Neither the directly estimated HEAVY model nor the Integrated HEAVY suffers 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 favour 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 realised 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 one-step is identical to the pointwise (and so replaced by the five-step horizon). HEAVY models perform well at all horizons, with statistically significant out-performance 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 4 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 but relatively modest movements in the corresponding HEAVY parameter α_R. Percentage changes are important as the time variation in the conditional variance is scaled by α parameters.

The bottom of Figure 2 shows the rolling estimate of the persistence parameters for the GARCH model $\alpha_G + \beta_G$ and the HEAVY-RM model $\alpha_R + \beta_R$. The latter shows consistently less memory than the former but, interestingly, the two sequences of parameter estimates are moving around in lock step. Figure 2 shows results for the α parameter. 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:

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

In this section we evaluate the performance using the quasi-likelihood criteria.

Figure 3 shows these innovations for the Dow Jones Index example, which is fairly typical of results we have seen for other series. At 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 occurred at the start of 1996 and at the start of 2007. There are no remarkable moves during the credit crunch.

DOI: 10.1002/jae

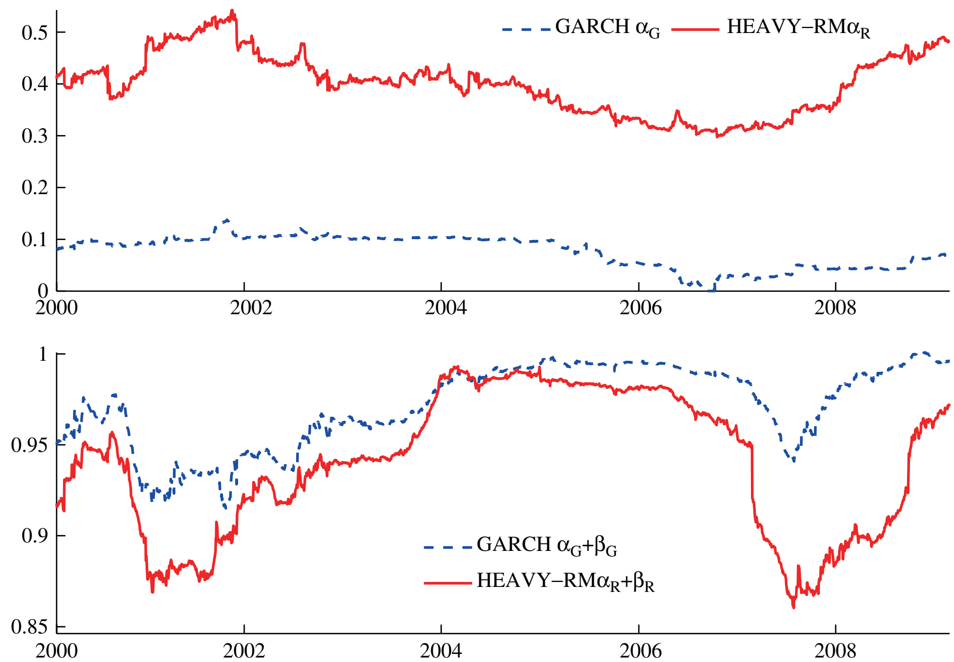

Figure 2. Recursive parameter estimates using a quasi-likelihood for GARCH and HEAVY model for the Dow Jones Industrial example. This figure is available in color online at www.interscience.wiley.com/journal/jae

At the top right-hand side of Figure 3 there is a time series plot of $\widehat{\eta}_t$, which has large moves at the same time as the large moves in $\widehat{\zeta}_t$. This is confirmed at the bottom left-hand side of the figure, which cross-plots $\widehat{\zeta}_t$ and $\widehat{\eta}_t$, suggesting some dependence in the bottom right-hand quadrant. The bottom right side 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.

Summary statistics for the innovations for all the series are given in Table VIII. We have chosen 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 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. The final column reports the first-order autocorrelation of $\widehat{\eta}_t^2$, which was small but generally positive. This may indicate that a more complex specification could be justified for the HEAVY-RM model, which is a topic of ongoing research.

Another features of the table which is 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.

DOI: 10.1002/jae

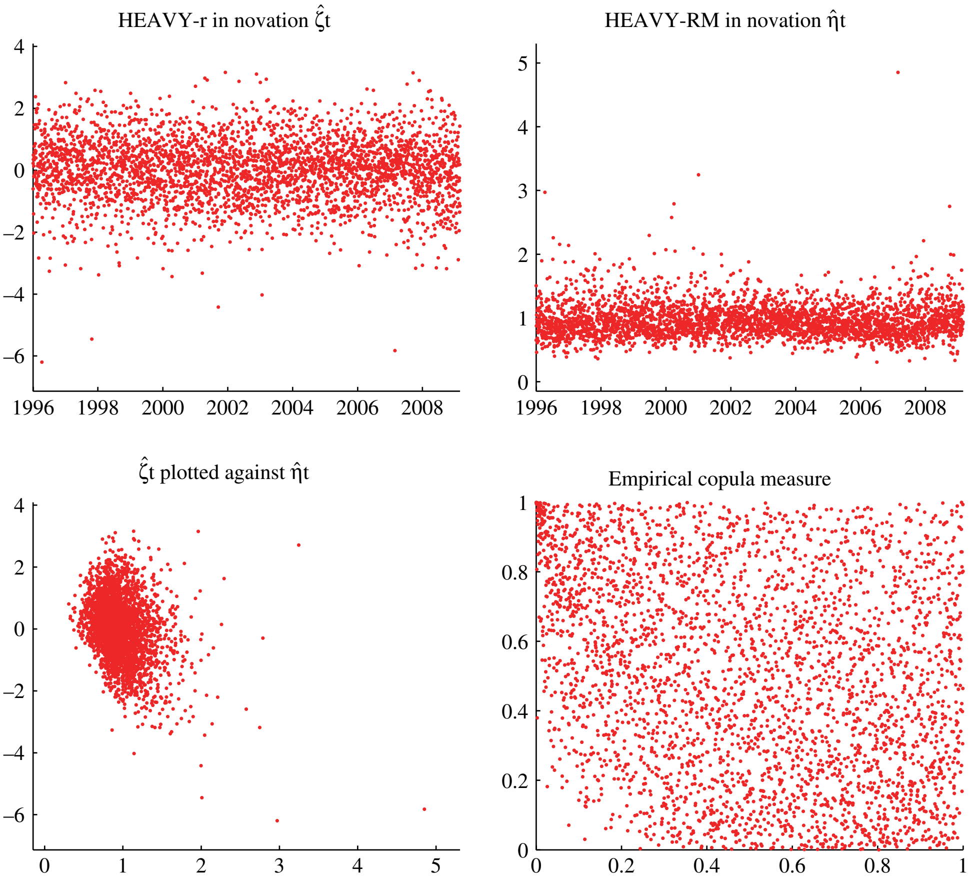

Figure 3. Innovations $\widehat{\zeta}_t$ and $\widehat{\eta}_t$ from the HEAVY model fitted to the DJI. Top left: the HEAVY-r model innovations $\widehat{\zeta}_t$, which should be roughly martingale difference sequences with unit variance. Top right is $\widehat{\eta}_t$, which should have unit conditional means and be uncorrelated. Bottom left is a cross-plot of $\widehat{\zeta}_t$ and $\widehat{\eta}_t$, while bottom right is the equivalent version mapped into copula spaces using the marginal empirical distribution functions to calculate the empirical copula measure. This figure is available in color online at www.interscience.wiley.com/journal/jae

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 dashed line, over the next month. The corresponding results for the GARCH model are also shown using a solid line with added symbols, with the multistep-ahead forecasts being shown using a dotted line.

DOI: 10.1002/jae

Table VIII. 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. The first five columns are estimated moments of their marginal distributions. r denotes the correlation, r_s is the Spearman rank correlation coefficient and ρ is the first-order autocorrelation

Asset	$\min(\widehat{\zeta}_t)$	$\max(\widehat{\zeta}_t)$	$E(\widehat{\zeta}_t^3)$	$Sd(\widehat{\zeta}_t^2)$	$Sd(\widehat{\eta}_t)$	$r(\widehat{\zeta}_t, \widehat{\eta}_t)$	$r_s(\widehat{\zeta}_t, \widehat{\eta}_t)$	$\rho(\eta_t^2)$
Dow Jones Industrials	−6.19	3.15	−0.336	1.82	0.270	−0.313	−0.280	0.008
Nasdaq 100	−5.90	4.25	−0.149	1.66	0.264	−0.321	−0.323	0.043
S&P 400 Midcap	−7.47	3.51	−0.366	1.87	0.257	−0.351	−0.334	0.036
S&P 500	−6.86	3.61	−0.396	1.90	0.270	−0.331	−0.312	0.027
Russell 3000	−7.01	3.97	−0.338	1.88	0.276	−0.339	−0.335	0.019
Russell 1000	−7.40	3.91	−0.346	1.92	0.275	−0.337	−0.329	0.021
Russell 2000	−7.08	3.48	−0.385	1.83	0.285	−0.289	−0.264	0.030
CAC 40	−4.37	3.64	−0.212	1.53	0.262	−0.350	−0.323	0.017
FTSE 100	−4.29	3.90	−0.307	1.52	0.263	−0.330	−0.312	0.012
German DAX	−5.30	3.79	−0.216	1.59	0.259	−0.396	−0.367	0.026
Italian MIBTEL	−4.90	3.40	−0.464	1.63	0.255	−0.430	−0.421	0.010
Milan MIB 30	−5.38	4.87	−0.076	1.79	0.263	−0.334	−0.339	0.012
Nikkei 250	−5.78	3.95	−0.343	1.77	0.260	−0.198	−0.162	0.043
Spanish IBEX	−6.95	5.19	−0.237	1.92	0.262	−0.328	−0.297	0.035
S&P TSE	−5.82	3.48	−0.225	1.63	0.254	−0.282	−0.286	0.045
MSCI Australia	−6.19	3.63	−0.318	1.75	0.238	−0.244	−0.207	0.040
MSCI Belgium	−5.78	3.04	−0.391	1.77	0.239	−0.310	−0.267	0.032
MSCI Brazil	−5.08	3.61	−0.194	1.59	0.258	−0.327	−0.311	0.041
MSCI Canada	−4.52	3.47	−0.232	1.58	0.247	−0.309	−0.298	0.064
MSCI Switzerland	−5.98	3.43	−0.453	1.84	0.231	−0.396	−0.346	0.012
MSCI Germany	−4.94	3.21	−0.333	1.57	0.246	−0.390	−0.370	0.042
MSCI Spain	−5.48	3.63	−0.211	1.60	0.243	−0.312	−0.297	0.033
MSCI France	−4.55	3.06	−0.249	1.48	0.250	−0.345	−0.335	0.027
MSCI UK	−4.71	3.17	−0.381	1.60	0.251	−0.347	−0.328	0.006
MSCI Italy	−4.44	3.17	−0.392	1.56	0.241	−0.396	−0.385	0.014
MSCI Japan	−5.95	3.41	−0.351	1.69	0.235	−0.274	−0.212	0.031
MSCI South Korea	−5.64	3.37	−0.239	1.71	0.222	−0.233	−0.229	0.001
MSCI Mexico	−5.19	3.75	−0.107	1.74	0.241	−0.262	−0.222	0.071
MSCI Netherlands	−5.00	3.23	−0.296	1.55	0.242	−0.368	−0.352	0.040
MSCI World	−5.36	4.34	−0.197	1.62	0.259	−0.227	−0.225	0.061
British pound	−3.58	3.76	−0.061	1.51	0.170	−0.050	−0.030	0.065
Euro	−4.20	3.48	0.060	1.54	0.196	0.014	0.017	0.053
Swiss franc	−4.49	3.91	−0.182	1.57	0.184	−0.101	−0.080	0.064
Japanese yen	−4.65	3.71	−0.322	1.80	0.222	−0.193	−0.128	0.028

The figure shows the GARCH model always slowly mean reverting back to its long-term average. It also shows from the start of September a sequence of upward moves in 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. Another feature that is 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.

DOI: 10.1002/jae

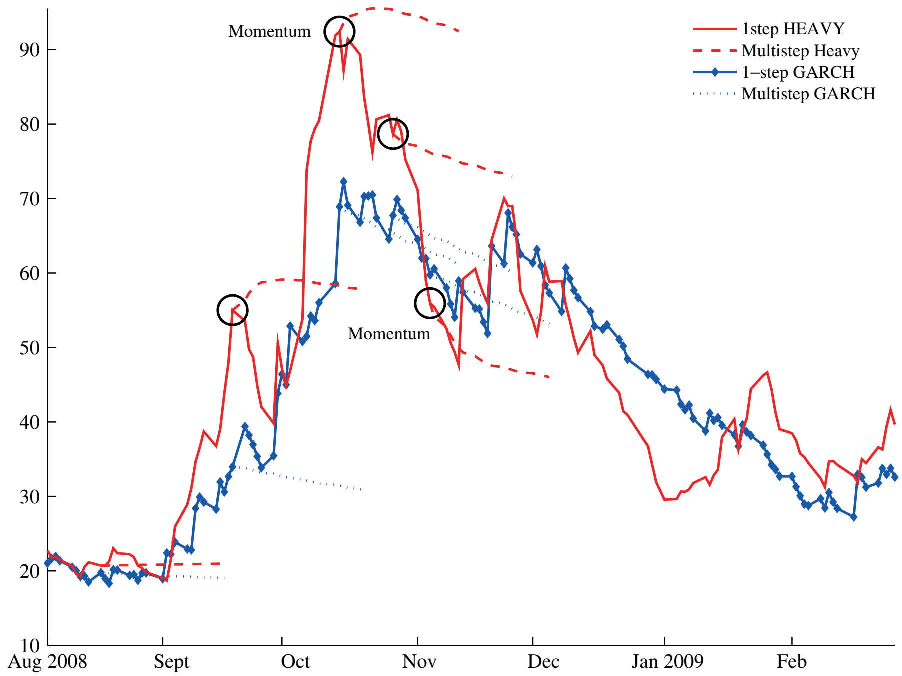

Figure 4. Volatility hedgehog plot for annualised volatility for the MSCI Canada series. The hedgehog plots are given for both HEAVY and GARCH models. Areas of momentum are indicated by ellipses. This figure is available in color online at www.interscience.wiley.com/journal/jae

Overall the main impressions are 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 as though 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. The result is some interesting arcs which appear in the volatility hedgehog plot given in Figure 5. The evidence in Table III 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. 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^*).

DOI: 10.1002/jae

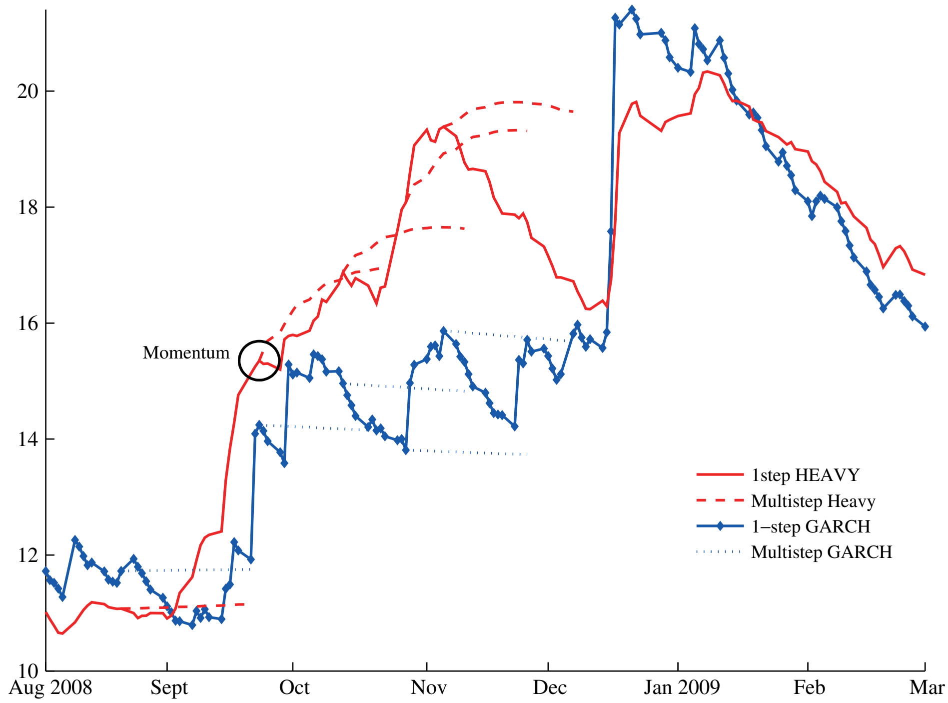

Figure 5. Extreme case of momentum. Volatility hedgehog plot for annualised volatility for the Swiss franc against the US dollar. The hedgehog plots are given for both HEAVY and GARCH models. This figure is available in color online at www.interscience.wiley.com/journal/jae

Realised semivariances (sums of squared negative returns) were introduced by Barndorff-Nielsen *et al*. (2009b) and further emphasised in empirical work by Patton and Sheppard (2009b). Now our model becomes

$$\begin{aligned}
\mathrm{var}(r_t|\mathcal{F}^{\mathrm{HF}}_{t-1}) &= h_t = \omega + \alpha \mathrm{RM}_{t-1} + \alpha^* \mathrm{RM}^*_{t-1} + \beta h_{t-1}, \quad \alpha^* \geq 0,\\
\mathrm{E}(\mathrm{RM}_t|\mathcal{F}^{\mathrm{HF}}_{t-1}) &= \mu_t = \omega_R + \alpha_R \mathrm{RM}_{t-1} + \beta_R \mu_{t-1},\\
\mathrm{E}(\mathrm{RM}^*_t|\mathcal{F}^{\mathrm{HF}}_{t-1}) &= \mu^*_t = \omega^*_R + \alpha^*_R \mathrm{RM}^*_{t-1} + \beta^*_R \mu^*_{t-1}, \quad \alpha^*_R, \beta^*_R \geq 0, \quad \alpha^*_R + \beta^*_R < 1
\end{aligned}$$

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 too, 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

$$\begin{aligned}
\mathrm{var}(r_t|\mathcal{F}^{\mathrm{HF}}_{t-1}) &= h_t = \omega + \alpha \mathrm{RM}_{t-1} + \alpha^* i_{t-1} \mathrm{RM}_{t-1} + \beta h_{t-1}, \quad \alpha^* \geq 0 \qquad (24)\\
\mathrm{E}(\mathrm{RM}_t|\mathcal{F}^{\mathrm{HF}}_{t-1}) &= \mu_t = \omega_R + \alpha_R \mathrm{RM}_{t-1} + \alpha i_{t-1} \mathrm{RM}_{t-1} + \beta_R \mu_{t-1}, \quad \alpha^*_R \geq 0
\end{aligned}$$

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

One approach to this is to assume that

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

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

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

Typically we would assume that $\mathrm{E}(i_{t+h}|\mathcal{F}_{t-1}^{\mathrm{HF}}) = \mathrm{E}(i_{t+h})$, 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 simulation through (24). This method of dealing with statistical leverage has the virtue 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.

5.2. 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

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

This is a nonparametric location-scale mixture.[13] Now $\varepsilon_t = \eta_t^{-1/2}[\zeta_t - \beta\{\eta_t - \mathrm{E}(\eta_t)\}]$ and so we can 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 $\mathrm{cov}(\zeta_t, \eta_t) = \beta$.

5.3. 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. The HEAVY model then becomes

$$\mathrm{var}(r_t|\mathcal{F}_{t-1}^{\mathrm{HF}}) = h_t = \omega + \alpha\mathrm{RM}_{t-1} + \beta h_{t-1} + \gamma r_{t-1}^2, \quad \beta + \gamma < 1$$
$$\mathrm{E}(\mathrm{RM}_t|\mathcal{F}_{t-1}^{\mathrm{HF}}) = \mu_t = \omega_R + \alpha_R\mathrm{RM}_{t-1} + \beta_R h_{t-1}, \quad \alpha_R + \beta_R < 1$$

It is then 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.

[13] If the parametric assumption that F_η was a generalised inverse Gaussian distribution and F_ε was Gaussian, then the resulting distribution for ζ_t would be the well-known generalised hyperbolic distribution.

DOI: 10.1002/jae

The derivation of this result is as follows:

$$r_t^2 = h_t + u_t, \quad h_t = \omega + \alpha \mathrm{RM}_{t-1} + \beta h_{t-1} + \gamma r_{t-1}, \quad \text{so}$$
$$\{1 - (\beta + \gamma)L\} r_t^2 = \omega + \alpha \mathrm{RM}_{t-1} + (1 - \beta L) u_t$$

where L is the lag operator. Likewise:

$$\{1 - (\alpha_R + \beta_R)L\}\mathrm{RM}_t = \omega_R + (1 - \beta_R L) v_t, \quad v_t = \mathrm{RM}_t - \mu_t$$

Combining delivers the result. In particular:

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

Thus

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

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 that 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.

ACKNOWLEDGEMENTS

We thank Tim Bollerslev, Frank Diebold, Rob Engle, Nathaniel Frank, Giampiero Gallo, Eric Ghysels, Andrew Patton, Natalia Sizova, Jun Yu and two referees for various helpful suggestions. We are responsible for any errors.

REFERENCES

Andersen TG, Bollerslev T, Diebold FX, Labys P. 2001a. The distribution of exchange rate volatility. *Journal of the American Statistical Association* **96**: 42–55. Correction published 2003, vol. 98, p. 501.

Andersen TG, Bollerslev T, Diebold FX, Ebens H. 2001b. The distribution of realized stock return volatility. *Journal of Financial Economics* **61**: 43–76.

Andersen TG, Bollerslev T, Diebold FX, Labys P. 2003. Modeling and forecasting realized volatility. *Econometrica* **71**: 579–625.

Andersen TG, Bollerslev T, Meddahi N. 2006. *Market microstructure noise and realized volatility forecasting*. Working paper, Department of Economics, Duke University. Forthcoming, Journal of Econometrics.

Andersen TG, Bollerslev T, Diebold FX. 2007. Roughing it up: including jump components in the measurement, modeling and forecasting of return volatility. *Review of Economics and Statistics* **89**: 707–720.

DOI: 10.1002/jae

Andersen TG, Bollerslev T, Diebold FX. 2009. Parametric and nonparametric measurement of volatility. In *Handbook of Financial Econometrics*, Aït-Sahalia Y, Hansen LP (eds). North-Holland: Amsterdam 67–137.
Andrews DWK. 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59**: 817–858.
Bandi FM, Russell JR. 2006. Separating microstructure noise from volatility. *Journal of Financial Economics* **79**: 655–692.
Bandi FM, Russell JR. 2008. Microstructure noise, realized variance, and optimal sampling. *Review of Economic Studies* **75**: 339–369.
Barndorff-Nielsen OE, Shephard N. 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 OE, Shephard N. 2006. Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics* **4**: 1–30.
Barndorff-Nielsen OE, Shephard N. 2007. Variation, jumps and high frequency data in financial econometrics. In *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress*, Blundell R, Persson T, Newey WK (eds). *Econometric Society Monographs*, Cambridge University Press: Cambridge, UK; 328–372.
Barndorff-Nielsen OE, Hansen PR, Lunde A, Shephard N. 2008. Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica* **76**: 1481–1536.
Barndorff-Nielsen OE, Hansen PR, Lunde A, Shephard N. 2009a. Realised kernels in practice: trades and quotes. *Econometrics Journal* **12**: C1–C32.
Barndorff-Nielsen OE, Kinnebrouck S, Shephard N. 2009b. Measuring downside risk: realised semivariance. In *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle*, Bollerslev T, Russell J, Watson M (eds). Oxford University Press: Oxford; 117–136.
Bollerslev T. 1986. Generalised autoregressive conditional heteroskedasticity. *Journal of Econometrics* **51**: 307–327.
Bollerslev T, Wooldridge JM. 1992. Quasi maximum likelihood estimation and inference in dynamic models with time varying covariances. *Econometric Reviews* **11**: 143–172.
Bollerslev T, Kretschmer U, Pigorsch C, Tauchen G. 2009. A discrete-time model for daily S&P 500 returns and realized variations: jumps and leverage effects. *Journal of Econometrics* **150**: 151–166.
Brownlees CT, Gallo GM. 2006. Financial econometrics at ultra-high frequency: data handling concerns. *Computational Statistics and Data Analysis* **51**: 2232–2245.
Brownlees CT, Gallo GM. 2009. Comparison of volatility measures: a risk management perspective. *Journal of Financial Econometrics* (forthcoming).
Christensen K, Podolskij M. 2007. Asymptotic theory for range-based estimation of integrated volatility of a continuous semi-martingale. *Journal of Econometrics* **141**: 323–349.
Cipollini F, Engle RF, Gallo G. 2007. *A model for multivariate non-negative valued processes in financial econometrics*. Working paper, Stern School of Business, New York University.
Clements MP, Hendry DF. 1999. *Forecasting Non-stationary Economic Time Series: The Zeuthen Lectures on Economic Forecasting*. MIT Press: Cambridge, MA.
Corradi V, Distaso W. 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 DR. 1961a. Prediction by exponentially weighted moving averages and related methods. *Journal of the Royal Statistical Society, Series B* **23**: 414–422.
Cox DR. 1961b. Tests of separate families of hypotheses. *Proceedings of the Berkeley Symposium* **4**: 105–123.
Diebold FX, Mariano RS. 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* **13**: 253–263.
Drost FC, Nijman TE. 1993. Temporal aggregation of GARCH processes. *Econometrica* **61**: 909–927.
Engle RF. 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom inflation. *Econometrica* **50**: 987–1007.
Engle RF. 2002. New frontiers for ARCH models. *Journal of Applied Econometrics* **17**: 425–446.
Engle RF, Gallo JP. 2006. A multiple indicator model for volatility using intra daily data. *Journal of Econometrics* **131**: 3–27.

Engle RF, Lee GGJ. 1999. A permanent and transitory component model of stock return volatility. In *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive W. J. Granger*, Engle RF, White H (eds). Oxford University Press: Oxford; 475–497.
Engle RF, Mezrich J. 1996. GARCH for groups. *Risk*: **9**(8): 36–40.
Engle RF, Hendry DF, Richard JF. 1983. Exogeneity. *Econometrica* **51**: 277–304.
Falkenberry TN. 2002. *High frequency data filtering*. Technical report, Tick Data.
Fan J, Wang Y. 2007. Multi-scale jump and volatility analysis for high-frequency financial data. *Journal of the American Statistical Association* **102**: 1349–1362.
Ghysels E, Rubia A, Valkanov R. 2009. *Multi-period forecasts of volatility: direct, iterated and mixed-data approaches*. Working paper, Department of Economics, UNC at Chapel Hill.
Giacomini R, White H. 2006. Tests of conditional predictive ability. *Econometrica* **74**: 1545–1578.
Giot P, Laurent S. 2004. Modelling daily value-at-risk using realized volatility and ARCH type models. *Journal of Empirical Finance* **11**: 379–398.
Golub GH, Van Loan CF. 1989. *Matrix Computations*. Johns Hopkins University Press: Baltimore, MD.
Hansen PR, Lunde A. 2006. Realized variance and market microstructure noise (with discussion). *Journal of Business and Economic Statistics* **24**: 127–218.
Heber G, Lunde A, Shephard N, Sheppard KK. 2009. *OMI's realised library*, Version 0.1. Oxford-Man Institute: University of Oxford.
Hendry DF. 1995. *Dynamic Econometrics*. Oxford University Press: Oxford.
Jacod J, Li Y, Mykland PA, Podolskij M, Vetter M. 2009. Microstructure noise in the continuous case: the pre-averaging approach. *Stochastic Processes and Their Applications* **119**: 2249–2276.
Lu Y. 2005. *Modeling and forecasting daily stock return volatility with intra-day price fluctuations information*. PhD thesis, Department of Economics, University of Pennsylvania.
Maheu JM, McCurdy TH. 2009. *Do high-frequency measures of volatility improve forecasts of return distributions?* Working paper, Department of Economics, Toronto University.
Marcellino M, Stock JH, Watson MW. 2006. A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics* **135**: 499–526.
Patton AJ. 2009. Volatility forecast evaluation and comparison using imperfect volatility proxies. *Journal of Econometrics* (forthcoming).
Patton AJ, Sheppard KK. 2009a. Evaluating volatility forecasts. In *Handbook of Financial Time Series*, Andersen TG, Davis RA, Kreiss JP, Mikosch T (eds). Springer: Berlin; 801–838.
Patton AJ, Sheppard KK. 2009b. *Good volatility, bad volatility: signed jumps and persistence of volatility*. Working paper, Oxford-Man Institute, University of Oxford.
Rivers D, Vuong QH. 2002. Model selection for nonlinear dynamic models. *Econometrics Journal* **5**: 1–39.
Vuong QH. 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, Mykland PA, Aït-Sahalia Y. 2005. A tale of two time scales: determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association* **100**: 1394–1411.

DOI: 10.1002/jae

APPENDIX: DATA CLEANING

The Realised Library is based on underlying high-frequency data, which we obtain through Reuters. We are not in a position to make available these base data, or its cleaned version, for commercial reasons, as Reuters owns the copyright to it. Although the raw data are of high quality, they do need to be cleaned so they are 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 mis-recordings of prices or large amounts of turbulence are encountered at the start of a trading day then they may sometimes give false signals. Barndorff-Nielsen *et al*. (2009a) 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 have more variation in structure than those dealt with in Barndorff-Nielsen *et al*. (2009a) 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 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 timestamp outside the interval when the exchange is open.

Quote data for the exchange rates are very plentiful and have 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 than 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 median centred but excluding the observation under consideration of 50 observations (i.e. 25 observations before and 25 after).

In addition, we have made various manual edits in the library when the results were 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.

DOI: 10.1002/jae