International Journal of Forecasting 32 (2016) 293–302

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Score-driven exponentially weighted moving averages and Value-at-Risk forecasting

André Lucas[a], Xin Zhang[b,*]

[a] VU University Amsterdam and Tinbergen Institute, The Netherlands

[b] Sveriges Riksbank, Sweden

ARTICLE INFO

Keywords:
Dynamic volatilities
Dynamic higher-order moments
Integrated generalized autoregressive score models
Exponentially Weighted Moving Average (EWMA)
Value-at-Risk (VaR)

ABSTRACT

We present a simple methodology for modeling the time variation in volatilities and other higher-order moments using a recursive updating scheme that is similar to the familiar RiskMetrics™ approach. The parameters are updated using the score of the forecasting distribution, which allows the parameter dynamics to adapt automatically to any non-normal data features, and increases the robustness of the subsequent estimates. The new approach nests several of the earlier extensions to the exponentially weighted moving average (EWMA) scheme. In addition, it can be extended easily to higher dimensions and alternative forecasting distributions. The method is applied to Value-at-Risk forecasting with (skewed) Student's t distributions and a time-varying degrees of freedom and/or skewness parameter. We show that the new method is as good as or better than earlier methods for forecasting the volatility of individual stock returns and exchange rate returns.

1. Introduction

The time variation in second and higher-order moments is an important phenomenon for assessing (tail) risk, constructing hedge strategies, and pricing assets. Exponentially weighted moving average (EWMA) methods have proved to be useful tools for capturing such time variation in a parsimonious and effective way. Here, we develop a new empirical methodology that extends and improves upon the standard EWMA approach. Our framework uses the higher-moment properties of the forecasting distribution to drive the dynamics of volatilities and other time-varying parameters. This ensures that the new method is robust to outliers if a non-normal forecasting distribution is used, as is typically the case when forecasting financial asset returns. The new method is easy to implement and remains similar in spirit to the highly familiar EWMA approach of RiskMetrics™.

The score-driven EWMA (SD-EWMA) model that we propose builds on a new observation-driven methodology, namely the generalized autoregressive score (GAS) dynamics; see Creal, Koopman, and Lucas (2011, 2013) and Harvey (2013). In particular, we consider an integrated version of the score-driven dynamics. The analogy is simple: just as the standard EWMA approach is a special case of the IGARCH(1,1) model of Bollerslev (1986) and Engle (1982), the proposed SD-EWMA approach is a special case of the IGAS(1,1) model of Creal et al. (2013). Its key feature is the fact that the time-varying parameter dynamics are driven by the score of the forecasting distribution. Empirical evidence of the usefulness of score-driven dynamics is provided by Creal, Schwaab, Koopman, and Lucas (2014), Harvey and Luati (2014), and Lucas, Schwaab, and Zhang (2014), for example, while Blasques, Koopman, and Lucas (2015) demonstrate the information-theoretic optimality properties of score-driven updates.

* Corresponding author.
E-mail addresses: a.lucas@vu.nl (A. Lucas), xin.zhang@riksbank.se (X. Zhang).

http://dx.doi.org/10.1016/j.ijforecast.2015.09.003

The intuition for using the score is straightforward. As an example, consider forecasting a time-varying variance of a fat-tailed distribution. If one uses the standard EWMA approach, a large absolute return has a major impact on the next period's estimated variance, due to the use of squared returns in the variance updating equation. Given the integrated nature of the EWMA dynamics, this impact affects a large number of the subsequent volatility estimates. If one accounts for the fat-tailedness of the return distribution by using a score-driven propagation mechanism for the variances, the impact of incidental tail observations is reduced substantially. This mitigation or robustifying mechanism is particularly important in our current context with integrated (infinite memory) dynamics.

Our methodology is computationally simple and remains similar in spirit to the standard EWMA approach. We also show that the SD-EWMA approach encompasses other proposals in the literature for the modeling of time-varying parameters, such as the normal-based standard EWMA, the robust EWMA of Guermat and Harris (2002) based on the Laplace distribution, and the skewed EWMA of Gerlach, Lu, and Huang (2013) based on the asymmetric Laplace distribution. Given that we are interested in modeling the time variation in financial risk measures, we explicitly develop an SD-EWMA model based on the fat-tailed skewed Student's t distribution; see for example Poon and Granger (2003) for stylized facts about financial returns. However, it is clear that the modeler can easily substitute his/her own favorite forecasting distribution instead, such as the normal inverse Gaussian (NIG) or the generalized hyperbolic (GH) distribution. We illustrate this by also making the skewness and degrees of freedom parameter of a skewed Student's t forecasting distribution time-varying.

We apply our approach to forecasting of the Value-at-Risk (VaR) for individual stock returns and foreign exchange rate returns. It turns out that the (skewed) Student's t based SD-EWMA schemes work best for most of the series considered. All of the SD-EWMA methods uniformly improve on the normal-based EWMA method. We show that both the shape of the conditional distribution and the score-driven updates can be helpful for improving the value-at-risk forecasting performance.

Compared to previous methods, such as those of Jensen and Lunde (2001) and Wilhelmsson (2009), the SD-EWMA approach has the distinct advantage that it provides a unifying framework that embeds previous proposals from the literature, such as those of Gerlach et al. (2013) and Guermat and Harris (2002). In addition, the generality of the SD-EWMA approach also allows for a straightforward generalization to higher dimensions, estimating score-driven versions of both volatilities, covariances and correlations, and other higher-order moments.

The remainder of the paper is set up as follows. In Section 2, we introduce the basic methodology and convey the main idea, using the Student's t distribution as a leading example. Next, we extend the framework to the forecasting of distributions with time-varying skewness and/or kurtosis. In Section 3, we briefly review the tests used in our forecasting experiment for assessing the performances of quantile forecasts. In Section 4, we provide our empirical application to Value-at-Risk forecasting. Section 5 concludes.

2. Score-driven exponentially weighted moving averages

2.1. Standard Gaussian EWMA approach

Consider a time series $y_t \in \mathbb{R}$ observed over the sample period $t = 1, \ldots, T$. In our setting, y_t typically holds financial returns, such as stock returns or foreign exchange rate returns. We assume that y_t has a time-varying conditional distribution $\boldsymbol{p}(y_t|\mathcal{F}_{t-1}; f_t, \theta)$, where $\mathcal{F}_{t-1}$ is the information set available at time $t-1$, f_t is a vector of time-varying parameters, and θ is a vector of static parameters. For example, $\mathcal{F}_{t-1}$ may include lags of y_t and of exogenous variables, and f_t may include time-varying means and/or volatilities, while θ may hold the remaining parameters that characterize the distribution, such as skewness and excess kurtosis parameters.

The standard RiskMetrics™ approach sets $f_t = \sigma_t^2$ and uses the exponentially weighted moving average (EWMA) scheme

$$\sigma_{t+1}^2 = \lambda\sigma_t^2 + (1-\lambda)y_t^2, \qquad 0 < \lambda < 1. \tag{1}$$

The EWMA scheme in Eq. (1) corresponds to a zero-intercept IGARCH model,

$$\begin{aligned}\sigma_{t+1}^2 &= \omega + a y_t^2 + b\sigma_t^2 \\ &= \omega + a(y_t^2 - \sigma_t^2) + (a+b)\sigma_t^2,\end{aligned} \tag{2}$$

with $\omega = 0$, $b = \lambda$, and $a = 1 - b$, such that $a + b = 1$. Thus, the volatility is a weighted sum of past squared observations. In particular, the term $(y_t^2 - \sigma_t^2)$ is directly proportional to the score of the normal distribution with respect to σ_t^2. If the observations y_t are conditionally fat-tailed, using squared observations in Eq. (2) may not be optimal, as large realizations of y_t may occur regularly even if the variance has not changed substantially. If not properly accounted for, such large realizations may bias the estimates of the true underlying volatility. Due to the long memory of the integrated GARCH model in Eq. (2), the bias may persist for a long time and affect a large number of the subsequent volatility estimates.

2.2. Score-driven EWMA

To account for the shape of the conditional forecasting distribution in our construction of an EWMA scheme, we use the generalized autoregressive score (GAS) framework of Creal et al. (2011, 2013); see also Harvey (2013). Blasques et al. (2015) show that updating the time-varying parameters based on the score of the forecasting distribution always improves the local Kullback–Leibler divergence between the model and the true, unknown data generating process. The GAS(1,1) dynamics for the time-varying parameter f_t are given by

$$\begin{aligned}&f_{t+1} = \omega + As_t + Bf_t, \qquad s_t = \mathcal{S}_t \cdot \partial\ell_t/\partial f_t, \\ &\ell_t = \ln \boldsymbol{p}(y_t|\mathcal{F}_{t-1}; f_t, \theta),\end{aligned} \tag{3}$$

where $\mathcal{S}_t = \mathcal{S}(f_t, \mathcal{F}_{t-1}; \theta)$ is an $\mathcal{F}_{t-1}$-measurable scaling function. Note that the scaled score s_t is a function of y_t, f_t, and $\mathcal{F}_{t-1}$. Thus, the time-varying parameter f_t, as specified in Eq. (3), is observation-driven, according to the classification of Cox (1981). Dynamics more complicated than those specified in Eq. (3) can also be added to the specification; see for example Janus, Koopman, and Lucas (2014) for fractionally integrated dynamics, Creal et al. (2013) for higher-order dynamics, and Harvey and Luati (2014) for both higher-order dynamics and structural time series dynamics. For our current purposes, however, the GAS(1,1) dynamics suffice. For the scaling matrix $\mathcal{S}_t$, we propose the inverse diagonal of the Fisher conditional information matrix,

$$\mathcal{S}_t = \mathrm{diag}(\mathcal{I}_{t|t-1})^{-1} = \mathrm{diag}\left(\mathrm{E}_{t-1}\left[\,(\ell_t/\partial f_t)\,(\ell_t/\partial f_t)'\,\right]\right)^{-1}.$$

This form of scaling accounts for the local curvature of each of the score elements and embeds the standard GARCH dynamics as a special case; see Creal et al. (2013) for more details. In contrast to Creal et al. (2013), we use only the diagonal (rather than the full) information matrix for scaling. The advantage of this is that each parameter feeds directly only on its own score, rather than on a mix of scores for different parameters. This may be an advantage in the current EWMA setting, where the parameter dynamics are typically considered parameter by parameter. We also found that a diagonal scaling matrix increases the stability of the EWMA procedure, particularly if we consider the time-varying volatility, skewness, and degrees of freedom parameters jointly, for instance in the case of our skewed Student's t distribution.

Scaling by the inverse (diagonal) information matrix enables us to construct a score-driven EWMA (SD-EWMA) scheme by building on the analogy of the EWMA scheme in Eq. (1) and the IGARCH specification in Eq. (2). In particular, similarly to Eq. (2), our SD-EWMA uses the integrated GAS dynamics

$$f_{t+1} = A s_t + f_t, \tag{4}$$

also labeled a Newton score step by Blasques et al. (2015). This corresponds to an integrated GAS specification if we set $\omega = 0$ and $B = 1$ in Eq. (3). For example, if $\boldsymbol{p}(y_t|\mathcal{F}_{t-1}; f_t, \theta)$ is the Gaussian distribution with zero mean and variance $f_t = \sigma_t^2$, Creal et al. (2013) show that Eq. (4) reduces precisely to the standard EWMA scheme in Eq. (1) if we set $A = 1 - \lambda$.

However, there is no particular need to restrict oneself to the normal distribution. As it is well established that financial returns are typically fat-tailed, it makes much more sense to use an SD-EWMA scheme that is based upon a fat-tailed distribution. In this paper we follow Creal et al. (2011, 2013) and Harvey (2013) and use the Student's t (and later also the skewed Student's t) distribution with ν degrees of freedom,

$$\boldsymbol{p}(y_t|\mathcal{F}_{t-1}; f_t, \theta) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{(\nu-2)\pi\sigma_t^2}}\left(1 + \frac{y_t^2}{(\nu-2)\sigma_t^2}\right)^{-\frac{\nu+1}{2}}, \tag{5}$$

with $f_t = \sigma_t^2$ and $\theta = \nu > 2$. The corresponding SD-EWMA scheme is given by

$$\begin{aligned}\sigma_{t+1}^2 &= \sigma_t^2 + A\cdot(1+3\nu^{-1})\cdot\left(\frac{\nu+1}{\nu-2+y_t^2/\sigma_t^2}\cdot y_t^2 - \sigma_t^2\right)\\ &= (1-\lambda)\sigma_t^2 + \lambda\cdot\frac{\nu+1}{\nu-2+y_t^2/\sigma_t^2}\cdot y_t^2,\end{aligned} \tag{6}$$

with $\lambda = A\cdot(1+3\nu^{-1})$. One can either fix ν at a predetermined value such as 5, for robustness purposes, or estimate it using an initial estimation sample.

As was discussed by Creal et al. (2013) and Harvey (2013), the weight factor in front of y_t^2 in Eq. (6) has a robustifying effect on the volatility dynamics. If y_t lies in the tails of the conditional distribution at time t, the volatility is increased, but not by the full y_t^2. Part of the effect is attributed to the fat-tailedness of the Student's t distribution, as can be seen from the division by $(\nu - 2 + y_t^2/\sigma_t^2)$. As the SD-EWMA scheme has the same integrated dynamics as the original EWMA scheme, a more robust estimate of the volatility at time t has a persistent effect on the subsequent volatility estimates as well.

Though the SD-EWMA approach can adapt itself to any parametric distribution, there is a trade-off to be considered. If the conditional distribution depends on more parameters rather than only the time-varying parameter f_t, e.g., the degrees of freedom parameter ν, these parameters need to be estimated before the SD-EWMA scheme can be operationalized. An attractive feature of the EWMA approach for volatility filtering and forecasting is precisely the fact that no off-line estimation is needed. One way to achieve this is to estimate the auxiliary parameters on an estimation sample and to update them only infrequently. This approach works well for the Student's t SD-EWMA scheme and performs better than a number of competing schemes for a range of foreign exchange rates and stock returns; see the application in Section 4. For other distributions, however, more care may be needed.

2.3. The Skewed Student's t distribution with time varying higher-order moments

We note the flexibility of the SD-EWMA approach for accounting for other dynamic parameters beyond the volatility context. For example, the model can be extended easily to handle both volatilities and covariances, or volatilities and correlations, using the recursions of Creal et al. (2011) and the integrated GAS(1,1) specification in Eq. (4). In addition, the approach can also be generalized further to handle the time variation in higher-order moments, such as skewness and kurtosis, by putting the appropriate parameters into f_t rather than θ. An example that we use in our subsequent empirical analysis is a new SD-EWMA model with a time-varying degrees of freedom parameter. For this, consider the likelihood in Eq. (5) and set $f_t' = (f_{1,t}, f_{2,t})$ with $\sigma_t^2 = f_{1,t}$ and $\nu_t = 2 + \exp(f_{2,t})$. Using inverse Fisher information scaling, we obtain the following recursion for ν_t:

$$\begin{aligned}f_{2,t+1} = f_{2,t} - A_\nu\,\frac{2}{\nu_t-2}\Bigg[&\gamma''\left(\frac{\nu_t+1}{2}\right) - \gamma''\left(\frac{\nu_t}{2}\right)\\ &+ \frac{2(\nu_t+4)(\nu_t-3)}{(\nu_t+1)(\nu_t+3)(\nu_t-2)^2}\Bigg]^{-1}\\ &\times\Bigg[\gamma'\left(\frac{\nu_t+1}{2}\right) - \gamma'\left(\frac{\nu_t}{2}\right) - \frac{1}{\nu_t-2}\\ &- \ln\left(1+\frac{y_t^2}{(\nu_t-2)\,\sigma_t^2}\right)\\ &+ \frac{\nu_t+1}{\nu_t-2}\cdot\frac{y_t^2}{(\nu_t-2)\sigma_t^2+y_t^2}\Bigg],\end{aligned} \tag{7}$$

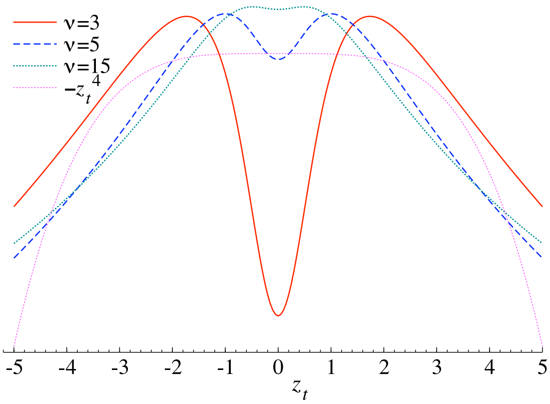

Fig. 1. *News impact curves for the time-varying degrees of freedom recursion in Eq.* (7). Scaled and re-centered news impact curves (Eq. (7)) as a function of $z_t = y_t^2/((\nu_t - 2)\sigma_t^2)$ for different values of ν_t. The (re-scaled and re-centered) curve of fourth order powers $-z_t^4$ is also shown as a benchmark.

where $A_\nu > 0$ is a scalar tuning parameter that is similar to the parameter A used for the volatility dynamics in Eq. (6), and $\gamma'(\cdot)$ and $\gamma''(\cdot)$ are the first and second order derivatives of $\gamma(\cdot) = \ln\Gamma(\cdot)$. The derivation of this result follows by using the results of Gómez, Torres, and Bolfarine (2007), for example, accounting for the fact that we model the variance of the Student's t distribution, rather than the scale parameter; see the online appendix (Appendix A) for further details. The reparameterization $\nu_t = 2 + \exp(f_{2,t})$ automatically ensures that the degrees of freedom parameter ν_t is always larger than 2, such that the variance of the Student's t distribution always exists. The score-based recursions account for this reparameterization automatically via the chain rule used in the score calculations.

Though the shape of the recursion for ν_t in (7) may look complicated at first sight, it is actually easy to implement. Interestingly, it does not use fourth order moments directly, as one might expect for the dynamics of a tail-shape parameter. Instead, it uses only a logarithmic moment, combined with the explicit information embedded in the tail shape of the Student's t distribution. One advantage of using the recursion in Eq. (7) is that it typically results in a much more stable path for the degrees of freedom parameter. In contrast, the fourth order moments of the data are notoriously unstable. Eq. (7) circumvents this instability problem by including squared data and the gamma functions and their derivatives. We provide some typical shapes for the news impact curves related to Eq. (7) for several values of ν_t in Fig. 1. The curves are re-centered and re-scaled so as to be comparable within a single figure. We also plot a fourth order polynomial $-z_t^4$ as a benchmark.

Fig. 1 shows that large values of $|z_t|$ result in a downward adjustment of ν_{t+1} for all curves considered. This is intuitive, as large values of $|z_t|$ can be associated with tails being fat. The decline in Eq. (7) for large values of z_t is comparable for different values of ν_t. Interestingly, the sensitivity of the GAS-based news impact curves for ν_{t+1} is much lower than that of the fourth order polynomial curve $-z_t^4$. This provides the SD-EWMA recursion for ν_t with its robustness feature. Also note that for fatter tailed distributions such as $\nu_t = 3$, values of z_t near zero also result in smaller values of ν_{t+1}. This is a consequence of the fact that, for the Student's t distribution, fat tails go hand in hand with leptokurtosis, i.e., 'peaked-ness' at the center of the distribution. The less leptokurtic the distribution, the smaller the downward effect of observations near zero compared to near, say, −1 or −2. The informativeness of observations in the center compared to tail observations only really becomes clear if the distribution is already fat-tailed, i.e., if ν_t is low. For higher values of ν_t, downward signals for ν_{t+1} must come predominantly from tail observations.

We note that the smoothing parameter A_ν for the ν_t recursion is typically smaller than that of the volatility recursion. Starting values for the estimation of A_ν for empirical data in the range of 0.001 work quite well. The low values of A_ν underline the stable path dynamics for ν_t described by Eq. (7). We show in Section 4 that allowing for a time-varying degrees of freedom parameter helps to improve the accuracy of tail probability estimates further for fat-tailed data.

Finally, the SD-EWMA also allows us to combine time-varying skewness and kurtosis, if so desired. One possible way forward is to use the skewed Student's t distribution with the associated score and information matrix expressions as derived by Gómez et al. (2007), for example, and discussed in a score-driven setting by Harvey (2013). The density of the skewed Student's t distribution is given by

$$\boldsymbol{p}(y_t|\mathcal{F}_{t-1}; f_t, \theta) = \frac{\Gamma\left(\frac{\nu_t+1}{2}\right)}{\Gamma\left(\frac{\nu_t}{2}\right)\sqrt{(\nu_t-2)\pi\bar{\sigma}_t^2}} \times \left(1 + \frac{y_t^2}{(1-\epsilon\cdot\operatorname{sign}(y_t-\bar{\mu}_t))(\nu_t-2)\bar{\sigma}_t^2}\right)^{-\frac{\nu_t+1}{2}}, \quad (8)$$

where $-1 < \epsilon_t < 1$ is the skewness parameter, and $\bar{\mu}_t$ and $\bar{\sigma}_t$ are the location and scale parameters, respectively. We can use the expressions for the mean μ_t and variance σ_t^2 of y_t as given by Gómez et al. (2007) to model the mean and time-varying variance rather than the location $\bar{m}_t$ and time-varying scale $\bar{\sigma}_t$. The precise equations are presented in the online appendix (see Appendix A) to this paper. The skewed Student's t model also allows us to illustrate the flexibility of the SD-EWMA approach to parameterize the model in such a way as to ensure proper parameter values for all values of f_t. For example, to ensure positive σ_t^2, $-1 < \epsilon_t < 1$, and $2 < \nu_t < 100$, we could choose $\sigma_t^2 = \exp(f_{1,t})$, $\epsilon_t = \tanh(f_{2,t})$, and $\nu_t = 51 + 49\tanh(f_{3,t})$. This reparameterization requires only slightly more involved expressions for the score, but leaves the rest of the SD-EWMA procedure untouched. For further details, see the online appendix (Appendix A).

2.4. *Extensions: other forecasting distributions*

Interestingly, the SD-EWMA approach also encompasses previous adaptations of the EWMA scheme proposed in the literature. For example, Guermat and Harris (2002) introduce a robust-EWMA scheme

$$\sigma_{t+1} = \lambda\sigma_t + (1-\lambda)\sqrt{2}|y_t|, \quad (9)$$

which is driven by absolute rather than squared observations. The authors relate their model to the GARCH-type models of Schwert (1990) and Taylor (1986), but Eq. (9) can also be seen as a special case of the SD-EWMA scheme in Eq. (4). To see this, consider the Laplace density

$$\boldsymbol{p}(y_t|\mathcal{F}_{t-1}; f_t, \theta) = \frac{1}{\sqrt{2}\sigma_t} \exp\left(-\sqrt{2}\,|y_t|/\sigma_t\right). \tag{10}$$

As for the standard EWMA, we set $f_t = \sigma_t^2$. The IGAS(1,1) for the Laplace distribution is

$$\begin{aligned} f_{t+1} &= \omega + 2A \cdot \sqrt{2}|y_t|\sigma_t + (B - 2A)f_t \\ \Leftrightarrow \sigma_{t+1}^2 &= \lambda\sigma_t^2 + \sigma_t \cdot (1-\lambda)\sqrt{2}|y_t|, \end{aligned} \tag{11}$$

if we set $\omega = 0$, $A = (1-\lambda)/2$, and $B = 1$. Except for the multiplication by σ_t, which is due to the parameterization $f_t = \sigma_t^2$ rather than $f_t = \sigma_t$, Eq. (11) is the same as Eq. (9). The robust-EWMA or Laplace based SD-EWMA model produces a modest increase in volatility for large values of $|y_t|$ compared to the standard EWMA (Eq. (1)). The derivation above reveals that the scheme can be considered as a score-driven approach based on the heavy-tailed Laplace distribution rather than the fat-tailed Student's t distribution in Eq. (6).

The SD-EWMA scheme introduced in Section 2.2 is very flexible. We can use it to accommodate the forecaster's favorite conditional distribution $\boldsymbol{p}(y_t|\mathcal{F}_{t-1}; f_t, \theta)$. As long as the conditional density has a parametric[1] form, we can compute the score and construct the SD-EWMA scheme. For example, Gerlach et al. (2013) introduces an EWMA scheme based on the asymmetric Laplace distribution

$$\begin{aligned} \boldsymbol{p}(y_t|\mathcal{F}_{t-1}; f_t, \theta) = \frac{k_t}{\sigma_t} \exp\Bigg(-\bigg(&\frac{1}{1-p_t}\mathbf{1}[y_t > 0] \\ &+ \frac{1}{p_t}\mathbf{1}[y_t < 0]\bigg)\frac{k_t\,|y_t|}{\sigma_t}\Bigg), \end{aligned} \tag{12}$$

with $f_t = (\sigma_t, p_t)$, and $k_t = (p_t^2 + (1-p_t)^2)^{1/2}$. Gerlach et al. (2013) introduce EWMA-type time variation in both σ_t and p_t, specified by the recursions

$$\begin{aligned} \sigma_{t+1} &= \lambda\sigma_t + (1-\lambda) \\ &\quad \times \left(\frac{k_t}{1-p_t}\mathbf{1}[y_t > 0] + \frac{k_t}{p_t}\mathbf{1}[y_t < 0]\right)|y_t|, \end{aligned} \tag{13}$$

$$\begin{aligned} u_{t+1} &= \beta_u u_t + (1-\beta_u)|y_t|\mathbf{1}[y_t > 0], \\ v_{t+1} &= \beta_v v_t + (1-\beta_v)|y_t|\mathbf{1}[y_t < 0], \\ p_{t+1} &= \left(1 + \sqrt{u_{t+1}/v_{t+1}}\right)^{-1}. \end{aligned} \tag{14}$$

We can also derive the IGAS(1,1) dynamics for σ_t^2 using $f_t = \sigma_t^2$ from Eq. (12) directly and obtain

$$\begin{aligned} \sigma_{t+1}^2 &= \lambda\sigma_t^2 + \sigma_t \cdot (1-\lambda) \\ &\quad \times \left(\frac{k_t}{1-p_t}\mathbf{1}[y_t > 0] + \frac{k_t}{p_t}\mathbf{1}[y_t < 0]\right)|y_t|, \end{aligned} \tag{15}$$

with $\lambda = 1 - 2A$. Again, we notice from Eq. (15) that the original robust and asymmetric EWMA scheme of Gerlach et al. (2013) can be interpreted as an SD-EWMA update if we set $f_t = \sigma_t$ rather than $f_t = \sigma_t^2$ as in the original EWMA.

[1] See Blasques, Ji, and Lucas (forthcoming) for an extension to a non-parametric density setting.

3. Value-at-Risk and backtesting

We now evaluate the performance of the SD-EWMA scheme for forecasting Value-at-Risk (VaR). We define the VaR $= -Y_\alpha$ at confidence level $(1-\alpha)$ as

$$Y_\alpha = \sup\left\{Y^* \,\middle|\, \mathrm{P}[Y < Y^*] \le \alpha\right\}.$$

The value of Y_α is highly dependent on the distributional assumptions for Y; see Chen and Lu (2010) for a recent survey. There is a trade-off between the fat-tailedness of the distribution of Y and the transition dynamics of the volatility updating mechanism. In the Student's t based SD-EWMA framework, the volatility updates are less responsive to extreme realized returns than in the standard Gaussian EWMA scheme. This makes the computed VaR less responsive to abrupt volatility changes. In contrast, if there are incidental tail observations, the Student's t based SD-EWMA scheme provides a much better and more robust estimate of the volatility at time t. Moreover, the fat-tailedness of the conditional Student's t distribution pushes the VaR levels farther out into the tails than for the Gaussian distribution with a fixed confidence level $(1-\alpha)$. The trade-off between all of these forces results in the relative performances of the different forecasting methods, which can be investigated only empirically across different confidence levels $(1-\alpha)$ and different datasets.

To assess the performances of alternative (SD)-EWMA methods, we consider a number of standard tests for the quality of tail probability forecasts: the unconditional coverage test, the independence test, the conditional coverage test, and the tail shape test of Berkowitz (2001). All of these tests are likelihood ratio (LR) based tests. A good VaR model should be consistent, in that the fraction of VaR violations, i.e., events $\{y_t < -VaR_t\}$, should equal α in large samples. Define the violation indicator

$$I_t = \mathbf{1}\{y_t < -VaR_t\},$$

and the number of violations $N = \sum_{t=1}^{T} I_t$ in T time periods. Following Christoffersen (1998), good VaR models produce serially independent I_ts. Our backtesting methods are all related to good coverage, serial independence, or both.

Kupiec (1995) tests the unconditional coverage (UC) of the VaR model using

$$LR_u = 2(\ln L_N - \ln L_\alpha) \sim \chi^2(1), \quad T \to \infty; \tag{16}$$

where $L_N = (1-N/T)^{T-N}(N/T)^N$ and $L_\alpha = (1-\alpha)^{T-N}\alpha^N$. Christoffersen (1998) proposes the independence (IN) test for the VaR violation indicator I_t. The transition matrix of the corresponding first-order Markov Chain is

$$\Pi = \begin{pmatrix} \pi_{00} & \pi_{01} \\ \pi_{10} & \pi_{11} \end{pmatrix},$$

$$\pi_{ij} = P(I_t = j | I_{t-1} = i) = T_{ij}/(T_{i0} + T_{i1}),$$

with T_{ij} recording the times of transition from states i to j, where $i, j \in \{0, 1\}$. The LR test for independence is

$$LR_{in} = 2(\ln L_\pi - \ln L_\alpha) \sim \chi^2(1), \quad T \to \infty, \tag{17}$$

where $L_\pi = \pi_{00}^{T_{00}}\pi_{01}^{T_{01}}\pi_{10}^{T_{10}}\pi_{11}^{T_{11}}$ and $L_\alpha = (1-\alpha)^{T_{01}+T_{11}}\alpha^{T_{00}+T_{10}}$. The simultaneous test for unconditional coverage and independence, namely the correct conditional coverage (CC) test, is

$$LR_c = LR_u + LR_{in} \sim \chi^2(2), \quad T \to \infty. \tag{18}$$

In practice, risk managers are concerned not only with the number of VaR failures, but also with the accuracy of the model for the tail shape beyond the VaR. This is relevant for assessing the potential magnitude of losses in the tail, and relates to the general shift in the industry and in regulation from VaR to expected loss (or conditional VaR) computations. To test for the general tail shape, we adopt the test proposed by Berkowitz (2001). The test operates on an inverse standard normal transformation of the probability integral transforms of the data, i.e.,

$$z_t = \Phi^{-1}\left(\hat{F}_t(y_t)\right), \tag{19}$$

where $\hat{F}_t(\cdot)$ denotes the estimated cumulative distribution function that is applicable at time t using the postulated VaR model, such as the Laplace, asymmetric Laplace, or (skewed) Student's t distribution, and $\Phi^{-1}(\cdot)$ denotes the inverse standard normal distribution function. The variable of interest is constructed by truncating the variable z_t at the threshold $\Phi^{-1}(\alpha) = -\text{VaR}$, such that $z_t = -\text{VaR}$ if $z_t \geq -\text{VaR}$. Estimating the mean and variance for a censored normal random variable can be achieved by maximizing the likelihood function

$$L(\mu,\sigma^2) = \sum_{z_t<-\text{VaR}} \left(-\frac{1}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}(z_t-\mu)^2\right) + \sum_{z_t\geq -\text{VaR}} \ln\left(1-\Phi\left(\frac{-\text{VaR}-\mu}{\sigma}\right)\right). \tag{20}$$

The Berkowitz (2001) test uses the maximum likelihood estimates to compute a likelihood ratio (LR) test for the null hypothesis $\mu = 0$ and $\sigma^2 = 1$. The corresponding LR test is

$$LR = -2(L(0,1) - L(\hat{\mu},\hat{\sigma}^2)),$$

which is asymptotically $\chi^2(2)$ distributed.

4. Empirical results

4.1. Data and descriptive statistics

In this section, we compare the performances of different SD-EWMA schemes. Note that, for the normal distribution, the SD-EWMA scheme coincides with the standard EWMA for volatility modeling. As was explained in Section 2, the SD-EWMA updating schemes in Eqs. (11) and (15), based on the Laplace and asymmetric Laplace distributions, respectively, are very close to the robust EWMA scheme (Eq. (9)) of Guermat and Harris (2002), and the skewed EWMA scheme (Eq. (13)) of Gerlach et al. (2013), respectively. For the dynamic asymmetric Laplace, we use the same dynamics for p_t in Eq. (14) as were used by Gerlach et al. (2013). As Gerlach et al. (2013) show that the GARCH and GJR-GARCH based on normal or Student's t distributions do not outperform the skewed EWMA models, we do not include them in our current study. We also benchmark our results against a standard EWMA scheme for the variance, while using a Student's t distribution to compute the relevant VaR and associated statistics.

We use 12 daily financial time series over the period January 5, 1999, to February 6, 2015. The dataset contains six exchange rate log returns and six equity log returns, with slightly over 4,000 observations per series. The exchange rates are always vis-à-vis the US Dollar and are taken from the database of the Federal Reserve Bank of St. Louis (FRED). We consider the Australian Dollar, Canadian Dollar, Euro, British Pound, Japanese Yen, and Swedish Kroner, denoted as AUD, CAD, EUR, GBP, JPY, and SEK, respectively. The stocks considered represent different industries and are all listed on the New York Stock Exchange: Alcoa Inc., Boeing Co., General Electric, IBM, Coca-Cola and AT&T, denoted as AA, BA, GE, IBM, KO, and T. Stock data are taken from Datastream.

From the descriptive statistics in Table 1, it is obvious that all series exhibit non-normal features such as non-zero skewness and excess kurtosis, particularly over the more recent sample period. Thus, we expect the Laplace-based SD-EWMA and Student's t SD-EWMA schemes to provide particular advantages relative to the standard EWMA scheme. We use the same distributional assumptions to set up the SD-EWMA recursions and to compute the VaR.

We split the sample into two subsamples. We use the sample from January 5, 1999, to December 29, 2006 (in-sample), to start off the estimation of the static parameters. In particular, for all models we use the estimation sample to estimate the optimal smoothing parameter A. We also estimate any remaining static parameters that may be needed, such as the degrees of freedom parameter ν for the Student's t distribution, or the skewness parameters p and ϵ for the asymmetric Laplace and skewed Student's t distributions, respectively. For the asymmetric Laplace or skewed Student's t with time-varying skewness, we estimate additional separate smoothing parameters for p_t, ϵ_t, and/or ν_t. In all cases, the estimated parameters are kept fixed over the entire forecasting period. This results in a computationally fast procedure. As parameters are unlikely to be kept fixed for the entire out-of-sample period of more than eight years in practice, we also carry out an analysis where all tuning parameters are updated recursively on a daily basis over the entire forecasting sample; see the discussion by Ardia and Hoogerheide (2014) for the potential benefits of such an approach.

4.2. Full results for the Euro–Dollar rate

For the Euro–Dollar exchange rate, we report the full results for all tests in Table 2. As usual, the normal-based standard EWMA scheme performs badly as we get deeper into the tails ($\alpha = 1\%, 0.5\%$). If we consider the hit rates (HR), we see that the normal and Student's t based approaches typically result in more VaR violations than the nominal level, whereas the Laplace-based models have fewer VaR violations. Considering the conditional

Table 1
Summary statistics.

Data	In-sample: 1999–2006				Out-of-sample: 2007–2015			
	Mean	SD	SK	EKS	Mean	SD	SK	EKS
	Exchange rate returns							
GBP	0.008	0.51	−0.02	0.57	−0.012	0.64	−0.40	6.94
AUD	0.012	0.68	−0.50	1.98	0.000	0.97	−0.71	12.83
JPY	0.003	0.63	−0.23	2.04	0.000	0.69	−0.26	5.07
CAD	−0.013	0.45	0.00	0.66	0.004	0.67	−0.08	5.98
SEK	−0.008	0.65	0.08	0.65	0.010	0.88	−0.20	4.14
EUR	0.006	0.61	0.02	0.73	−0.008	0.65	0.19	3.41
	Equity returns							
AA	0.032	2.34	0.22	2.63	−0.023	3.03	−0.34	6.95
BA	0.056	2.07	−0.38	5.76	0.035	1.90	−0.02	4.21
GE	0.014	1.86	0.05	4.12	−0.006	2.14	−0.06	9.87
IBM	0.006	2.08	−0.09	8.08	0.031	1.46	−0.07	5.49
KO	−0.009	1.61	−0.06	4.92	0.038	1.21	0.08	6.87
T	−0.004	2.04	−0.09	3.19	0.020	1.49	0.80	14.99

The descriptive statistics present the centered moments of the financial time series considered. The sample period is from January 5, 1999, to February 6, 2015. We split the sample into an in-sample estimation period and an out-of-sample forecasting period. The sample mean is multiplied by 100. A standard deviation (SD) of 1.28 denotes 1.28% per day. SK and EKS denote the skewness and excess kurtosis, respectively.

Table 2
Full SD-EWMA results for the Euro–Dollar exchange rate.

	No parameter updating					With parameter updating				
	CC	UC	IN	HR	BE	CC	UC	IN	HR	BE
	$\alpha = 0.5\%$									
N	18.0	17.3	0.7	1.28	29.9	19.9	19.2	0.7	1.33	29.8
T(ν)	3.1	2.9	0.3	0.79	0.1	2.2	2.0	0.2	0.74	0.2
T(ν_t)	0.5	0.3	0.1	0.59	2.7	0.9	0.7	0.2	0.64	0.5
ST(ε, ν_t)	4.1	3.8	0.3	0.84	0.2	0.9	0.7	0.2	0.64	1.4
ST(ε_t, ν)	0.9	0.7	0.2	0.64	0.0	0.1	0.0	0.1	0.49	0.9
ST(ε_t, ν_t)	0.9	0.7	0.2	0.64	0.0	0.5	0.3	0.1	0.59	0.8
T(ν)-RM	3.1	2.9	0.3	0.79	0.4	3.1	2.9	0.3	0.79	0.2
L(0.5)	7.0	7.0	0.0	0.15	19.8	7.0	7.0	0.0	0.15	20.0
L(p)	7.0	7.0	0.0	0.15	16.7	9.9	9.9	0.0	0.10	24.2
L(p_t)	7.0	7.0	0.0	0.15	19.6	4.9	4.9	0.0	0.20	16.6
	$\alpha = 1\%$									
N	12.5	11.1	1.4	1.82	29.4	12.5	11.1	1.4	1.82	29.3
T(ν)	5.0	4.1	0.9	1.48	1.0	5.0	4.1	0.9	1.48	1.0
T(ν_t)	2.7	2.0	0.7	1.33	3.4	2.7	2.0	0.7	1.33	0.6
ST(ε, ν_t)	5.0	4.1	0.9	1.48	1.3	4.1	3.3	0.8	1.43	3.4
ST(ε_t, ν)	2.1	1.5	0.7	1.28	0.1	2.7	2.0	0.7	1.33	1.9
ST(ε_t, ν_t)	2.1	1.5	0.7	1.28	0.1	2.7	2.0	0.7	1.33	1.9
T(ν)-RM	4.1	3.3	0.8	1.43	1.0	4.1	3.3	0.8	1.43	0.8
L(0.5)	6.6	6.5	0.1	0.49	24.2	5.3	5.2	0.1	0.54	22.3
L(p)	9.9	9.8	0.1	0.39	30.6	6.6	6.5	0.1	0.49	23.9
L(p_t)	3.2	3.1	0.2	0.64	19.3	5.3	5.2	0.1	0.54	19.5
	$\alpha = 5\%$									
N	5.0	4.0	1.0	6.00	29.0	6.1	4.4	1.7	6.05	28.9
T(ν)	13.8	9.3	4.5	6.54	5.8	9.0	7.2	1.8	6.35	4.6
T(ν_t)	10.4	7.7	2.7	6.39	6.0	9.3	7.7	1.7	6.39	5.0
ST(ε, ν_t)	13.5	8.8	4.7	6.49	5.5	6.4	4.8	1.6	6.10	5.6
ST(ε_t, ν)	11.8	5.3	6.5	6.15	2.8	5.0	4.0	1.0	6.00	3.9
ST(ε_t, ν_t)	11.8	5.3	6.5	6.15	2.8	5.0	4.0	1.0	6.00	4.0
T(ν)-RM	14.4	7.7	6.7	6.39	4.6	10.0	7.2	2.8	6.35	4.3
L(0.5)	1.6	0.0	1.6	5.02	26.2	1.3	0.1	1.2	5.16	29.0
L(p)	1.3	0.1	1.2	5.16	31.9	2.2	0.7	1.5	5.41	29.6
L(p_t)	1.3	0.1	1.2	5.16	23.1	0.7	0.0	0.7	5.02	21.5
Critical values	9.2	6.6	6.6	–	9.2	9.2	6.6	6.6	–	9.2

The test statistics correspond to the unconditional coverage (UC) test of Kupiec (1995), the independence (IN) and conditional coverage (CC) tests of Christoffersen (1998), and the Berkowitz (2001) test (BE). We use confidence levels for the VaR of $(1-\alpha) = 0.995/0.99/0.95$. Critical values ($\chi^2_{cv}$) at a 1% significance level are also displayed, as are the hit rates (HR) N/T of N VaR violations out of T observations, multiplied by 100. Static parameters are estimated over the period from Jan. 5, 1999, to Dec. 29, 2006, and held fixed over the forecast evaluation period from Jan. 3, 2007, to Feb. 6, 2015. The SD-EWMA schemes use the normal distribution (N), Laplace distribution (L) with skewness parameter 0.5, p, or p_t, the Student's t (T) and skewed Student's t (ST) distributions with degrees of freedom parameter ν or ν_t, and skewness parameter ϵ or ϵ_t. We provide the results for models with and without updated parameters in two different panels.

and unconditional coverage tests (CC, UC), the under-rejection for the Laplace is significant in several cases, whereas the over-rejection for the Student's t setting is never significant.

If we proceed by considering the tail shape beyond the VaR level using the Berkowitz test, we see that the Student's t based models perform better than both the normal and Laplace based approaches. We also note that a simple benchmark of standard Gaussian EWMA dynamics with a Student's t distribution for the VaR calculations also performs quite well (T(ν)-RM). For the Euro–Dollar rate, its behavior is quite close to that of the other models at VaR confidence levels of 99% and 99.5%. Less far out into the tails of the distribution, the performance of this method drops somewhat relative to those of the other Student's t based methods. However, if we consider the case where the tuning parameters are updated recursively, we see that the performance of T(ν)-RM starts to lag more substantially behind those of the skewed Student's t methods with time-varying parameters, particularly in terms of conditional coverages (CC).

To obtain an impression of the shape of the time-varying parameters, we plot σ_t^2, ϵ_t, and ν_t for the skewed Student's t model in Fig. 2. We can clearly see the increased volatility around the time of the financial crisis, as well as the higher volatility level during the European sovereign debt crisis (2010–2013). The skewness parameters indicate a positive skewness at the start of the sample. During the remainder of the sample period, the exchange rate returns are consistently negatively skewed, and particularly so around the time of the financial and European sovereign debt crises. The degrees of freedom parameter ranges from low values of around three near the end of the sample, to values of 15 in the periods of the great moderation, the financial crisis, and the European sovereign debt crisis.

We conclude that the skewed Student's t models with SD-EWMA dynamics for ϵ_t, ν_t, or both have the best overall performances in terms of coverage (CC, UC, IN) and tail shape beyond the VaR (BE), especially if we update the tuning parameters regularly based on the available data, as is commonly done in practice.

4.3. Full results: all series

To investigate the robustness of the results, we extend our analysis to other exchange rates as well as to individual stock returns. To save space, we present the results graphically for all series, three different confidence levels, and three tests: the conditional coverage test, the Berkowitz test, and the hit rate ($\hat{\alpha}/\alpha - 1$). As the setting with updated tuning parameters is most relevant from a practical point of view, we only present those results.

The results are shown in Fig. 3. Each column of three panels presents the results for the three different tests for a given VaR confidence level, $1 - \alpha = 0.995$, 0.99, or 0.95. The results for the exchange rate series are indicated by circles, and those for the stock returns by inverted triangles.

Looking at the top row of graphs, we confirm the results from Table 2 concerning the hit rates of the different methods. The normal and Student's t based methods typically result in somewhat more VaR violations than the nominal level. The Laplace-based approaches, on the other hand, result in substantially lower numbers of VaR violations. The further we go out into the tails, the worse the normal-based approach performs in terms of the hit rate. We also see that, across all series, the overall performance of the skewed Student's t based approaches in terms of hit rates is better than that of a standard RiskMetrics plus Student's t distribution approach (T(ν)-RM). This is particularly true for VaR confidence levels of 95% and 99.5%.

The above results are confirmed when we look at the second row of graphs, which indicate the significance of deviations from the nominal coverage, combined with possible violations of the independence assumption. Graphically, it is clear that, across different time series, the skewed Student's t based approaches perform best. The differences between using a skewed Student's t distribution with ϵ_t, ν_t, or both being time-varying appear to be much smaller.

If we consider the behaviors of different approaches for capturing the tail shape beyond the VaR, the bottom row of graphs in Fig. 3 shows that the Laplace distribution is clearly too thin-tailed to be able to describe the tail behavior of exchange rates and stock returns adequately. Note that the bottom row of graphs does not show the results for the normal distribution, as the Berkowitz test results for the normal are so high that they would distort the picture completely for the other models. The graphs also reveal that, for all VaR confidence levels, the polynomial tail shape of the (skewed or symmetric) Student's t distribution typically captures the stochastic behavior of extreme returns quite well. Note that, across all series, the skewed Student's t SD-EWMA results with time-varying ϵ_t and/or ν_t appear to be less susceptible to extreme outcomes for the tests than the other Student's t based approaches. Overall, the SD-EWMA approach with the time-varying skewed Student's t appears to have the best and most robust performance in our current volatility forecasting context.

5. Conclusion

We have developed a range of simple EWMA refinements that build on the recent literature on score-driven dynamics for time-varying parameters in non-normal models. In this paper we have shown that the standard EWMA and the robust Laplace based EWMA can be seen as special cases of the new score-driven EWMA (SD-EWMA) approach. In particular, as financial return series may typically be fat-tailed rather than heavy-tailed (such as Laplace), we developed a score-driven EWMA scheme based on the symmetric and skewed Student's t distributions. As the score-driven approach is not limited to time variation in volatilities only, we also developed a new SD-EWMA scheme for the simultaneous time series dynamics of the volatility, the degrees of freedom, and possibly the skewness parameter in a (skewed) Student's t distribution. The new schemes exhibit interesting robustness features for the time-varying parameter dynamics that make them

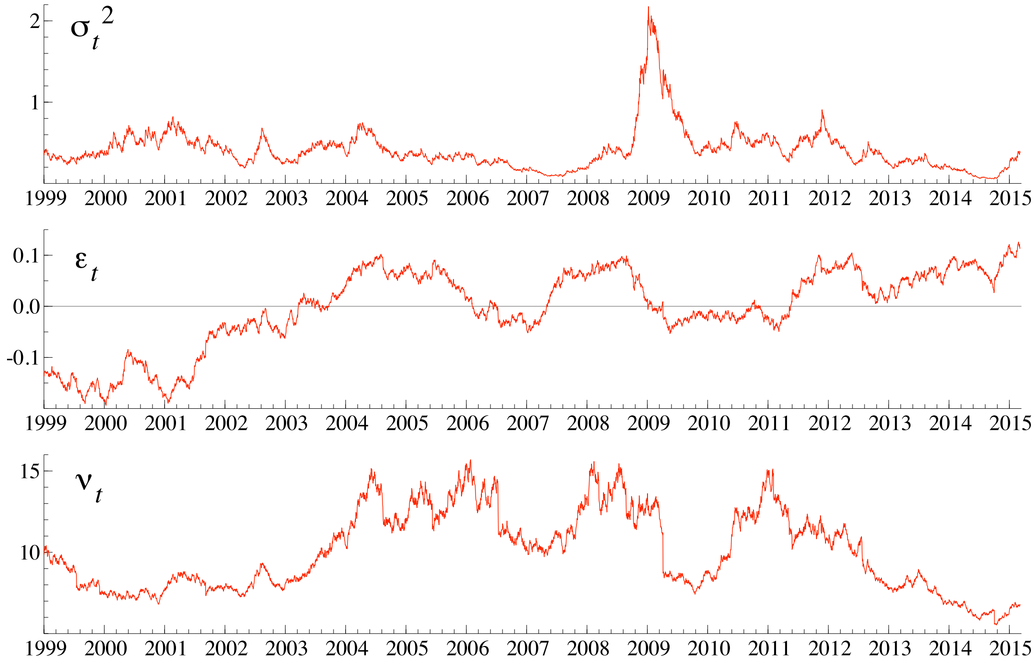

Fig. 2. Time-varying variance (σ_t^2), skewness (ϵ_t), and degrees of freedom (ν_t) for the Skewed Student's t model for the Euro–Dollar rate.

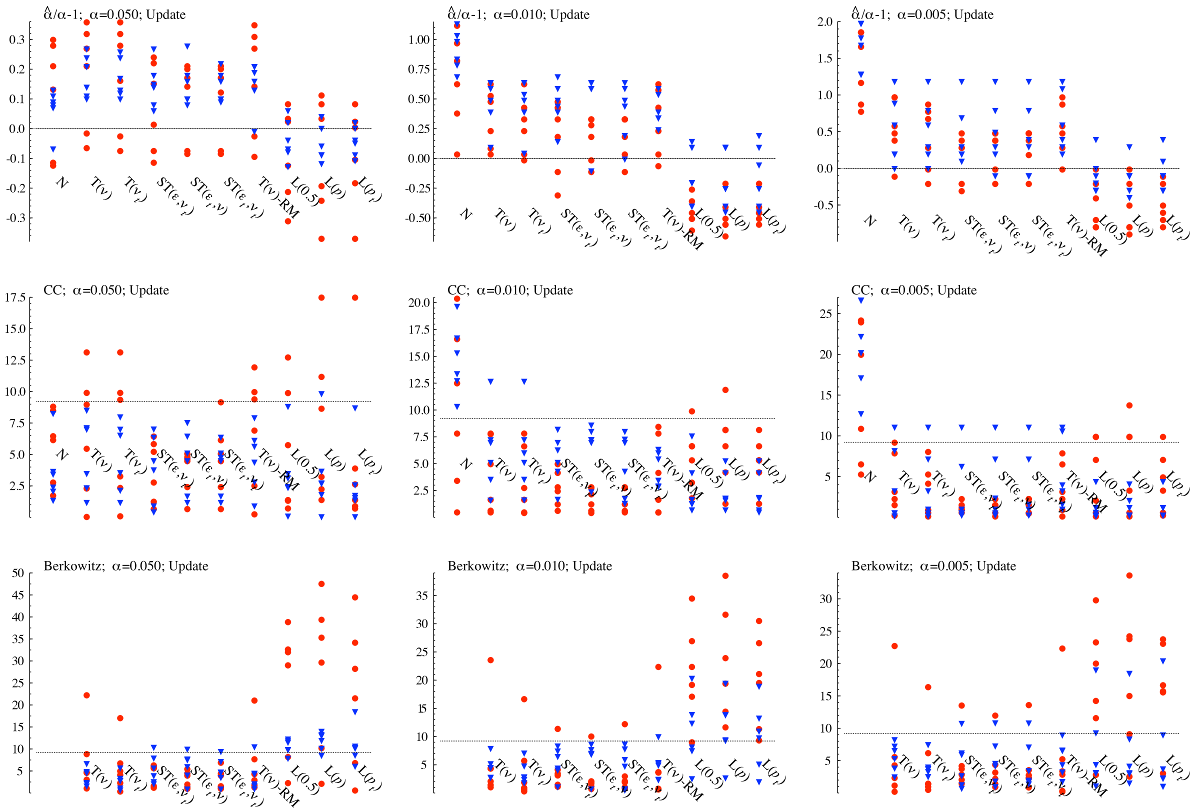

Fig. 3. *VaR performance.* Each panel contains the test results for 10 modeling methods using recursively estimated tuning parameters (see Table 2 for descriptions of the methods). Circles and inverted triangles indicate the test results for the six exchange rates and six stock return rates, respectively. The Berkowitz tests for the normal (N) are uniformly large, and are therefore left out.

particularly suitable in a context with non-Gaussian distributed observations.

We applied the new methods to the forecasting of Value-at-Risk (VaR) for exchange rate and stock return data. We found that the skewed Student's t based SD-EWMA model with time-varying volatility, degrees of freedom and/or skewness parameters had the best performance overall for different series and different VaR confidence levels. Thus, the new score-driven EWMA approach provides a unified and flexible tool for risk forecasting.

The score-driven EWMA approach can easily be adapted further to accommodate the researcher's preferred choice of forecasting distribution. For example, the ideas could be generalized further to semi-parametric approaches, such as the Gram-Charlier expansion of Gabrielsen, Zagaglia, Kirchner, and Liu (2012). Also note that the SD-EWMA can be adapted to handle multivariate observations; see for example Creal et al. (2011) and Lucas et al. (2014). Both of these possible extensions open up interesting avenues for further research.

Acknowledgments

André Lucas thanks the Dutch National Science Foundation (NWO; grant VICI453-09-005) for financial support. We thank conference participants at the 25th $(EC)^2$ conference and the CEF 2015 conference for helpful comments. The views expressed in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Executive Board of Sveriges Riksbank.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.ijforecast.2015.09.003.

References

Ardia, D., & Hoogerheide, L. F. (2014). Garch models for daily stock returns: Impact of estimation frequency on value-at-risk and expected shortfall forecasts. *Economics Letters*, *123*(2), 187–190.
Berkowitz, J. (2001). Testing density forecasts, with applications to risk management. *Journal of Business & Economic Statistics*, *19*(4), 465–474.
Blasques, F., Ji, J., & Lucas, A. (2015). Semiparametric score driven volatility models. *Computational Statistics & Data Analysis*, forthcoming.
Blasques, F., Koopman, S. J., & Lucas, A. (2015). Information-theoretic optimality of observation-driven time series models for continuous responses. *Biometrika*, *102*(2), 325–343.
Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327.
Chen, Y., & Lu, J. (2010). Value at risk estimation. In *Handbook of computational finance* (pp. 307–333). Springer.
Christoffersen, P. F. (1998). Evaluating interval forecasts. *International economic review*, 841–862.
Cox, D. R. (1981). Statistical analysis of time series: some recent developments. *Scandinavian Journal of Statistics*, *8*, 93–115.
Creal, D., Koopman, S. J., & Lucas, A. (2011). A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. *Journal of Economic and Business Statistics*, *29*(4), 552–563.
Creal, D., Koopman, S. J., & Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, *28*(5), 777–795.
Creal, D., Schwaab, B., Koopman, S. J., & Lucas, A. (2014). Observation driven mixed-measurement dynamic factor models with an application to credit risk. *Review of Economics and Statistics*, *96*(5), 898–915.
Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflations. *Econometrica*, *50*, 987–1008.
Gabrielsen, A., Zagaglia, P., Kirchner, A., & Liu, Z. (2012). Forecasting value-at-risk with time-varying variance, skewness and kurtosis in an exponential weighted moving average framework. arXiv preprint arXiv:1206.1380.
Gerlach, R., Lu, Z., & Huang, H. (2013). Exponentially smoothing the skewed laplace distribution for value-at-risk forecasting. *Journal of Forecasting*, *32*(6), 534–550.
Gómez, H. W., Torres, F. J., & Bolfarine, H. (2007). Large-sample inference for the epsilon-skew-*t* distribution. *Communications in Statistics. Theory and Methods*, *36*(1), 73–81.
Guermat, C., & Harris, R. D. (2002). Robust conditional variance estimation and value-at-risk. *Journal of Risk*, *4*, 25–42.
Harvey, A. C. (2013). *Dynamic Models for Volatility and Heavy Tails*. Cambridge University Press.
Harvey, A. C., & Luati, A. (2014). Filtering with heavy tails. *Journal of the American Statistical Association*, *109*, 1112–1122.
Janus, P., Koopman, S. J., & Lucas, A. (2014). Long memory dynamics for multivariate dependence under heavy tails. *Journal of Empirical Finance*, *29*, 187–206.
Jensen, M. B., & Lunde, A. (2001). The NIG-S&ARCH model: a fat-tailed, stochastic, and autoregressive conditional heteroskedastic volatility model. *The Econometrics Journal*, *4*(2), 319–342.
Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives*, *3*(2).
Lucas, A., Schwaab, B., & Zhang, X. (2014). Conditional euro area sovereign default risk. *Journal of Business and Economic Statistics*, *32*(2), 271–284.
Poon, S.-H., & Granger, C. W. (2003). Forecasting volatility in financial markets: A review. *Journal of Economic Literature*, *41*(2), 478–539.
Schwert, G. W. (1990). Stock volatility and the crash of '87. *Review of Financial Studies*, *3*(1), 77–102.
Taylor, S. J. (1986). *Modelling financial time series* (1st ed.). Wiley.
Wilhelmsson, A. (2009). Value at risk with time varying variance, skewness and kurtosis: the NIG-ACD model. *The Econometrics Journal*, *12*(1), 82–104.

André Lucas is professor of finance at VU University Amsterdam. He received his Ph.D. in Econometrics from Erasmus University Rotterdam and has published on financial and time series econometrics as well as risk management in journals such as the *Journal of Business and Economic Statistics*, *Journal of Econometrics*, and *Review of Economics and Statistics*. Together with Creal and Koopman, he propagates the use of generalized autoregressive score dynamics for time varying parameter models. He received a five year prestigious VICI research grant for this project from the Dutch national research council (NWO).

Xin Zhang received his Ph.D. degree from VU University Amsterdam and Tinbergen Institute. He also holds an M.Phil. in Econometrics and Finance from Tinbergen Institute. He was an external consultant for ECB in 2011. In fall 2012, he joined Sveriges Riksbank as an Economist in the research division. His research areas include time series econometrics, financial economics and credit risk. Xin's work has been published in the *Journal of Business and Economic Statistics*.