Duisenberg school of finance - Tinbergen Institute Discussion Paper

TI 14-092/IV/DSF77

Score Driven Exponentially Weighted Moving Average and Value-at-Risk Forecasting

André Lucas[1]
Xin Zhang[2]

[1] *Faculty of Economics and Business Administration, VU University Amsterdam, and Tinbergen Institute, the Netherlands;*
[2] *Sveriges Riksbank, Sweden.*

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

More TI discussion papers can be downloaded at http://www.tinbergen.nl

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Duisenberg school of finance is a collaboration of the Dutch financial sector and universities, with the ambition to support innovative research and offer top quality academic education in core areas of finance.

DSF research papers can be downloaded at: http://www.dsf.nl/

Duisenberg school of finance
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 8579

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*

Abstract

A simple methodology is presented for modeling time variation in volatilities and other higher-order moments using a recursive updating scheme similar to the familiar RiskMetrics™ approach. We update parameters using the score of the forecasting distribution. This allows the parameter dynamics to adapt automatically to any non-normal data features and robustifies the subsequent estimates. The new approach nests several of the earlier extensions to the exponentially weighted moving average (EWMA) scheme. In addition, it can easily be extended 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 competitive to or better than earlier methods in forecasting volatility of individual stock returns and exchange rate returns.

Keywords: dynamic volatilities, dynamic higher-order moments, integrated generalized autoregressive score models, Exponentially Weighted Moving Average (EWMA), Value-at-Risk (VaR).

JEL: C51, C52, C53, G15.

---

1. Introduction

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 to capture

☆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.

Email addresses: `a.lucas@vu.nl` (André Lucas), `xin.zhang@riksbank.se` (Xin Zhang)

September 9, 2015

such time variation in a parsimonious and effective way. Here, we develop a new empirical methodology to extend and improve upon the standard EWMA approach. Our framework exploits the higher-moment properties of the forecasting distribution to drive the dynamics of volatilities and other time-varying parameters. By doing so, the new method is robust to outliers if a non-normal forecasting distribution is used, which is typically the case when forecasting financial asset returns. The new method is easy to implement and remains close in spirit to the highly familiar EWMA approach of RiskMetrics$^{\text{TM}}$.

The score-driven EWMA (SD-EWMA) model we propose builds on a new observation driven methodology, namely the generalized autoregressive score (GAS) dynamics; see Creal et al. [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 Engle [1982] and Bollerslev [1986], the proposed SD-EWMA approach is a special cases of the IGAS(1,1) model of Creal et al. [2013]. Its key feature is that the time-varying parameter dynamics are driven by the score of the forecasting distribution. Empirical evidence for the usefulness of score driven dynamics is provided in for example Creal et al. [2014], Lucas et al. [2014], and Harvey and Luati [2014], while Blasques et al. [2015] demonstrate the information-theoretic optimality properties of score-driven updates.

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 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 subsequent volatility estimates. If one accounts for the fat-tailedness of the return distribution using a score-driven propagation mechanism for the variances, the impact of incidental tail observations is substantially mitigated. This mitigation or robustifying mechanism is particularly important in our current context with integrated (infinite memory) dynamics.

Our methodology is computationally simple and remains close in spirit to the standard EWMA approach. We also show that the SD-EWMA approach encompasses other proposals from the literature to model 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 et al. [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. It is clear, however, 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 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 better for most of the series considered. All SD-EWMA methods improve uniformly on the normal based EWMA method. We show that both the shape of the conditional distribution and the score-driven updates can be helpful to improve the value-at-risk forecasting performance.

Compared to previous methods, such as 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 Guermat and Harris [2002] and Gerlach et al. [2013]. 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 intuition using the Student's t distribution as a leading example. Next, we extend the framework to forecasting distributions with time varying skewness and/or kurtosis. In Section 3, we briefly review the tests used in our forecasting experiment to assess the performance 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 characterizing the distribution, such as skewness and excess kurtosis parameters.

The standard RiskMetrics$^{\text{TM}}$ 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 (1) corresponds to a zero-intercept IGARCH model,

$$\sigma_{t+1}^2 = \omega + \alpha y_t^2 + \beta \sigma_t^2 = \omega + \alpha (y_t^2 - \sigma_t^2) + (\alpha + \beta) \sigma_t^2, \tag{2}$$

with $\omega = 0$, $\beta = \lambda$, and $\alpha = 1 - \beta$, such that $\alpha + \beta = 1$. The volatility is thus 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 (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 (2), the bias may persist for a long time and affect a large number of subsequent volatility estimates.

2.2. *Score Driven EWMA*

To account for the shape of the conditional forecasting distribution in constructing 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 by the score of the forecasting distribution always locally improves the 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

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

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 $\mathcal{F}_{t-1}$. The time-varying parameter f_t as specified in (3) is thus observation driven in the classification of Cox [1981]. More complicated dynamics than the ones specified in equation (3) can be added to the specification; see for example Janus et al. [2011] for fractionally integrated dynamics, Creal et al. [2013] for higher-order dynamics, and Harvey and Luati [2014] for higher order dynamics as well as 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 only feeds directly 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 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 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 equation (1) and the IGARCH specification in (2). In particular, similar to (2) our SD-EWMA uses the integrated GAS dynamics

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

also labeled a Newton score step in Blasques et al. [2015]. This corresponds to an integrated GAS specification by setting $\omega = 0$ and $B = 1$ in equation (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 (4) reduces precisely to the standard EWMA scheme in (1) if we set $A = 1 - \lambda$.

There is, however, 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 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,

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

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

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

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 discussed in Creal et al. [2013] and Harvey [2013], the weight factor in front of y_t^2 in equation (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 subsequent volatility estimates as well.

Though the SD-EWMA approach adapts itself to any parametric distribution, there is a trade-off to be considered. If the conditional distribution depends on more parameters than the time-varying parameter f_t only, 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 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. For the Student's t SD-EWMA scheme this approach works well and better than a number of competing schemes for a range of foreign exchange rate 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 to account for other dynamic parameters beyond the volatility context. For example, the model can easily be extended to handle both volatilities and covariances, or volatilities and correlations, using the recursions in Creal et al. [2011] and the integrated GAS(1,1) specification in (4). In addition, the approach can be further generalized to handle 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 equation (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,

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

where $A_\nu > 0$ is a scalar tuning parameter similar to the parameter A used for the volatility dynamics in (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 results in for example Gómez et al. [2007], accounting for the fact that we model the variance of the Student's t distribution, rather than the scale parameter; see the online appendix 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 automatically account for this reparameterization 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 directly use fourth order moments as one may have expected for the dynamics of a tail-shape parameter. Rather, it only uses a logarithmic moment, combined with the explicit information embedded in the tail shape of the Student's t distribution. An advantage of using the recursion in (7) is that it typically results in a much more stable path of the degrees of freedom parameter. Fourth order moments of the data, by contrast, are notoriously unstable. The composition of squared data and the gamma functions and their derivatives in (7) circumvent this problem of instability. We provide some typical shapes of the news impact curves related to equation (7) for several values of ν_t in Figure 1. The curves are re-centered and re-scaled to be comparable within one figure. We also plot a fourth order polynomial $-z_t^4$ as a benchmark.

Figure 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 (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 z_t near zero also result in smaller values of ν_{t+1}. This is a consequence of the fact that fat-tails for the Student's t distribution 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

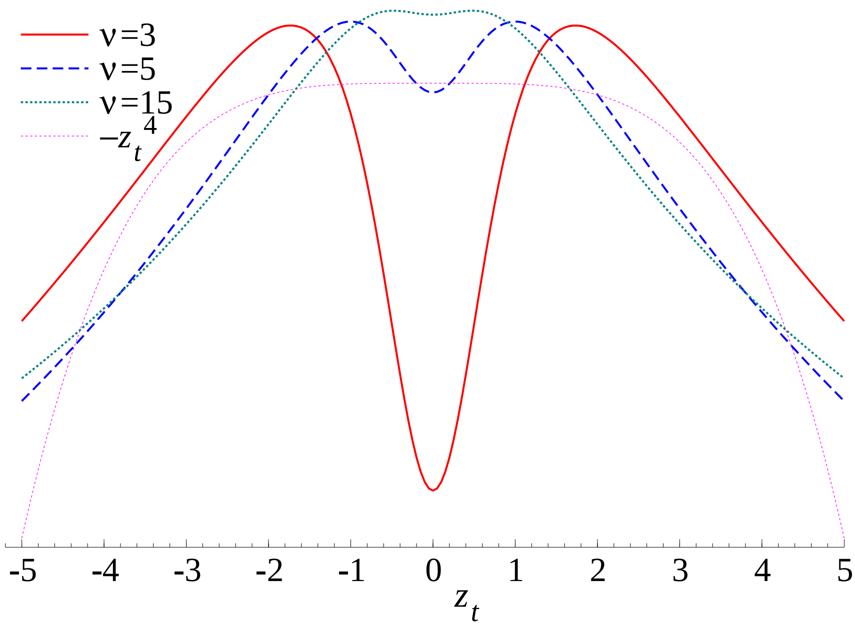

Figure 1: News impact curves for the time-varying degrees of freedom recursion (7)

Scaled and recentered news impact curves (7) as a function of $z_t = y_t^2/((\nu_t - 2)\sigma_t^2)$ for different values of ν_t. The (rescaled and recentered) curve of fourth order powers $-z_t^4$ is also shown as a benchmark.

low. For higher values of ν_t, downward signals for ν_{t+1} predominantly must come 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 (7). We show in Section 4 that allowing for a time-varying degrees of freedom parameter helps to further improve the accuracy of tail probability estimates for fat-tailed data.

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

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

where $-1 < \epsilon_t < 1$ is the skewness parameter, and $\bar{\mu}_t$ and $\bar{\sigma}_t$ are the location and scale parameter, respectively. We can use the expressions for the mean μ_t and variance σ_t^2 of y_t as given in 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 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 can for instance 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 only causes slightly more involved expressions for the score, but leaves the rest of the SD-EWMA procedure untouched. Further details can be found in the online appendix.

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

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

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

Laplace density

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

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

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

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 $f_t = \sigma_t$, (11) is the same as (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 (1). The derivation above reveals that the scheme can be motivated as a score-driven approach based on the heavy-tailed Laplace distribution rather than the fat-tailed Student's t distribution in (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 $p(y_t|\mathcal{F}_{t-1}; f_t, \theta)$. As long as the conditional density has a parametric[1]form, we can compute the score and construct the SD-EWMA scheme. The scheme also works for asymmetric distributions. For example, Gerlach et al. [2013] introduces an EWMA scheme based on the asymmetric Laplace distribution

$$p(y_t|\mathcal{F}_{t-1}; f_t, \theta) = \frac{k_t}{\sigma_t} \exp\left(-\left(\frac{1}{1-p_t}\mathbf{1}[y_t > 0] + \frac{1}{p_t}\mathbf{1}[y_t < 0]\right)\frac{k_t\,|y_t|}{\sigma_t}\right), \tag{12}$$

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

[1]See Blasques et al. [2015] for an extension to a non-parametric density setting.

type time variation in both σ_t and p_t, specified by the recursions

$$\begin{aligned} \sigma_{t+1} &= \lambda\sigma_t + (1-\lambda)\left(\frac{k_t}{1-p_t}\mathbf{1}[y_t > 0] + \frac{k_t}{p_t}\mathbf{1}[y_t < 0]\right)|y_t|, \quad (13) \\ u_{t+1} &= \beta_u u_t + (1-\beta_u)|y_t|\mathbf{1}[y_t > 0], \\ v_{t+1} &= \beta_v v_t + (1-\beta_v)|y_t|\mathbf{1}[y_t < 0], \\ p_{t+1} &= \left(1 + \sqrt{u_{t+1}/v_{t+1}}\right)^{-1}. \quad (14) \end{aligned}$$

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

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

with $\lambda = 1 - 2\,A$. Again we notice from (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.

3. Value-at-Risk and backtesting

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

$$Y_a = \sup\left\{Y^* \;\middle|\; \mathrm{P}[Y < Y^*] \leq a\right\}.$$

The value of Y_a hinges tightly together with the distributional assumptions for Y; see Chen and Lu [2012] 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 compared to the standard Gaussian EWMA scheme. This makes the computed VaR less responsive to abrupt volatility changes. By contrast, if there are incidental tail observations, the Student's t based SD-EWMA scheme provides a much better and 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 compared to the Gaussian distribution for a fixed confidence level $(1 - a)$. The trade-off between all these forces results in the relative performance of the different methods for forecasting, which can only be investigated empirically across different confidence levels $(1 - a)$ and different datasets.

To assess the performance 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 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 a in large samples. Define the violation indicator

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

and the number of violation $N = \sum_{t=1}^{T} I_t$ out of 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 indicators I_t. The transition matrix of the corresponding first-order Markov Chain is

$$\Pi = \begin{pmatrix} \pi_{00} & \pi_{01} \\ \pi_{10} & \pi_{11} \end{pmatrix}, \qquad \pi_{ij} = P(I_t = j \mid I_{t-1} = i) = T_{ij}/(T_{i0} + T_{i1}),$$

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

$$LR_{in} = 2(\ln L_A - \ln L_0) \sim \chi^2(1), \quad T \to \infty, \tag{17}$$

where $L_A = \pi_{00}^{T_{00}} \pi_{01}^{T_{01}} \pi_{10}^{T_{10}} \pi_{11}^{T_{11}}$ 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 not only concerned 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.,

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

where $\hat{F}_t(\cdot)$ denotes the estimated cumulative distribution function 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 $\Phi^{-1}(a) = -\text{VaR}$, 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

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

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 performance of different SD-EWMA schemes. Note that for the normal distribution, the SD-EWMA scheme coincides with the standard EWMA for volatility modeling. As explained in Section 2, the SD-EWMA updating schemes (11) and (15) based on the Laplace and asymmetric Laplace distribution, respectively, are very close to the robust EWMA scheme (9) of Guermat and Harris [2002], and the skewed EWMA scheme(13) of Gerlach et al. [2013], respectively. For the dynamic asymmetric Laplace, we use the same dynamics for p_t in (14) as used in Gerlach et al. [2013]. As Gerlach et al. [2013] show that the GARCH and GJR-GARCH based on a normal or Student's t distribution 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 6 exchange rate log returns and 6 equity log returns with

Table 1: Summary Statistics

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

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

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 St. Louiss (FRED). We consider the Australian Dollar, the Canadian Dollar, the 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 at 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. We thus expect the Laplace based SD-EWMA and Student's t SD-EWMA schemes to provide particular advantages compared 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 estimate the optimal smoothing parameter A using the estimation sample. We also estimate any remaining static parameters 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 distribution, 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 in practice parameters are unlikely to be kept fixed for the entire out-of-sample period of more than 8 years, we also carry an analysis where all tuning parameters are recursively updated on a daily basis over the entire forecasting sample; see the discussion in 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 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 compared to the nominal level, whereas the Laplace based models have fewer VaR violations. Considering the conditional 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.

Table 2: Full SD-EWMA Results for the Euro-Dollar Exchange Rate

The test statistics correspond to the unconditional coverage (UC) test of Kupiec [1995], the independence (ID) and Conditional Coverage (CC) test of Christoffersen [1998], and the Berkowitz [2001] test (BE). We use a confidence level for the VaR equal to $1 - a = 0.995/0.99/0.95$. Critical values (χ^2_{cv}) at a 1% significance level are also displayed, as are the Hit Rate (HR) N/T of N VaR violations out of T observations, multiplied with 100. Static parameters are estimated over Jan 5, 1999 to Dec 29, 2006, and held fixed over the forecast evaluation period 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) and skewed Student's t (ST) distribution with degrees of freedom parameter ν, ν_t, and skewness parameter ϵ or ϵ_t. We separate the results for models with and without updated parameters in two different panels.

	No parameter updating					With parameter updating				
	CC	UC	IN	HR	BE	CC	UC	IN	HR	BE
					$a = 0.5\%$					
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
					$a = 1\%$					
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
					$a = 5\%$					
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

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 1% and 0.5%. Less far out into the tails of the distribution, the performance of this method drops somewhat compared to that of the other Student's t based methods. If, however, 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 compared to that of the skewed Student's t methods with time-varying parameters, particularly in terms of conditional coverage (CC).

To get an impression about 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 Figure 2. We 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 indicates positive skewness at the start of the sample. During the remainder of the sample period, the exchange rate returns are repeatedly 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 around 3 near the end of the sample, to values of 15 in the period 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 either ϵ_t, ν_t, or both, have the best overall performance in terms of coverage (CC, UC, IN) and tail shape beyond the VaR (BE), especially if we regularly update the tuning parameters based on the available data, as is commonly done in practice.

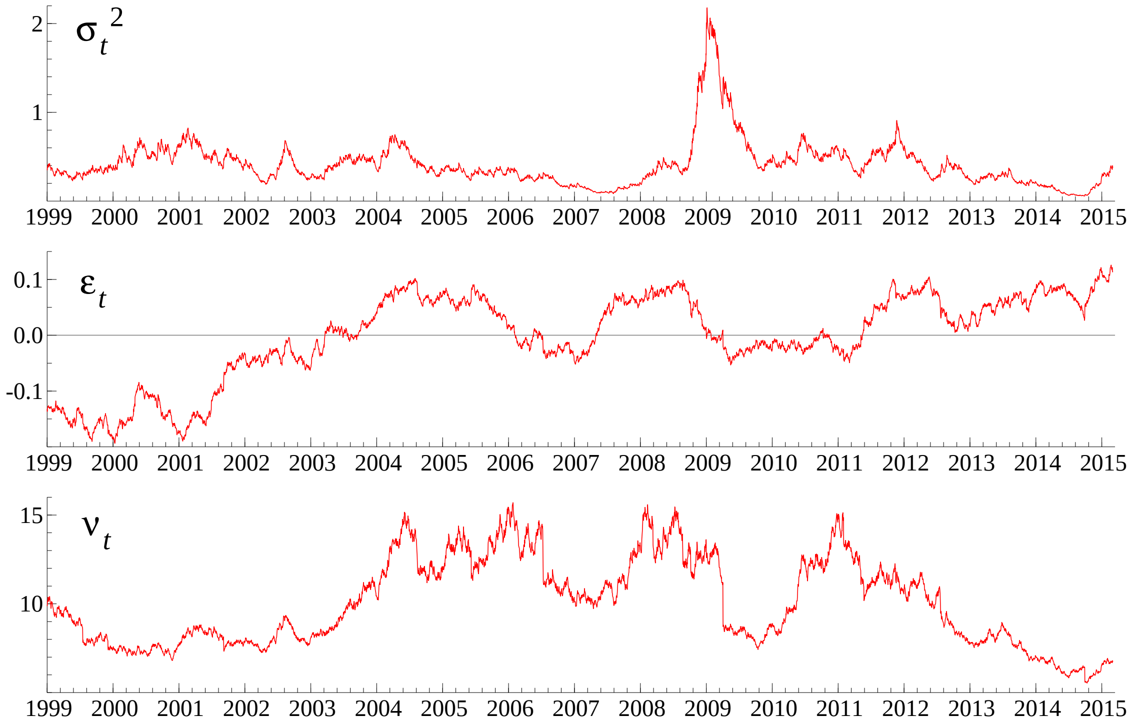

Figure 2: Time-varying variance (σ_t^2), skewness (ϵ_t), and degrees of freedom (ν_t) for the Skewed Student's t model for the Euro-Dollar rate

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 for 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.

The results are shown in Figure 3. Each column of three panels presents the results for the three different tests for a given VaR confidence level. The columns contain the results for the three different VaR confidence levels, $\alpha = 0.005, 0.01, 0.05$. 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

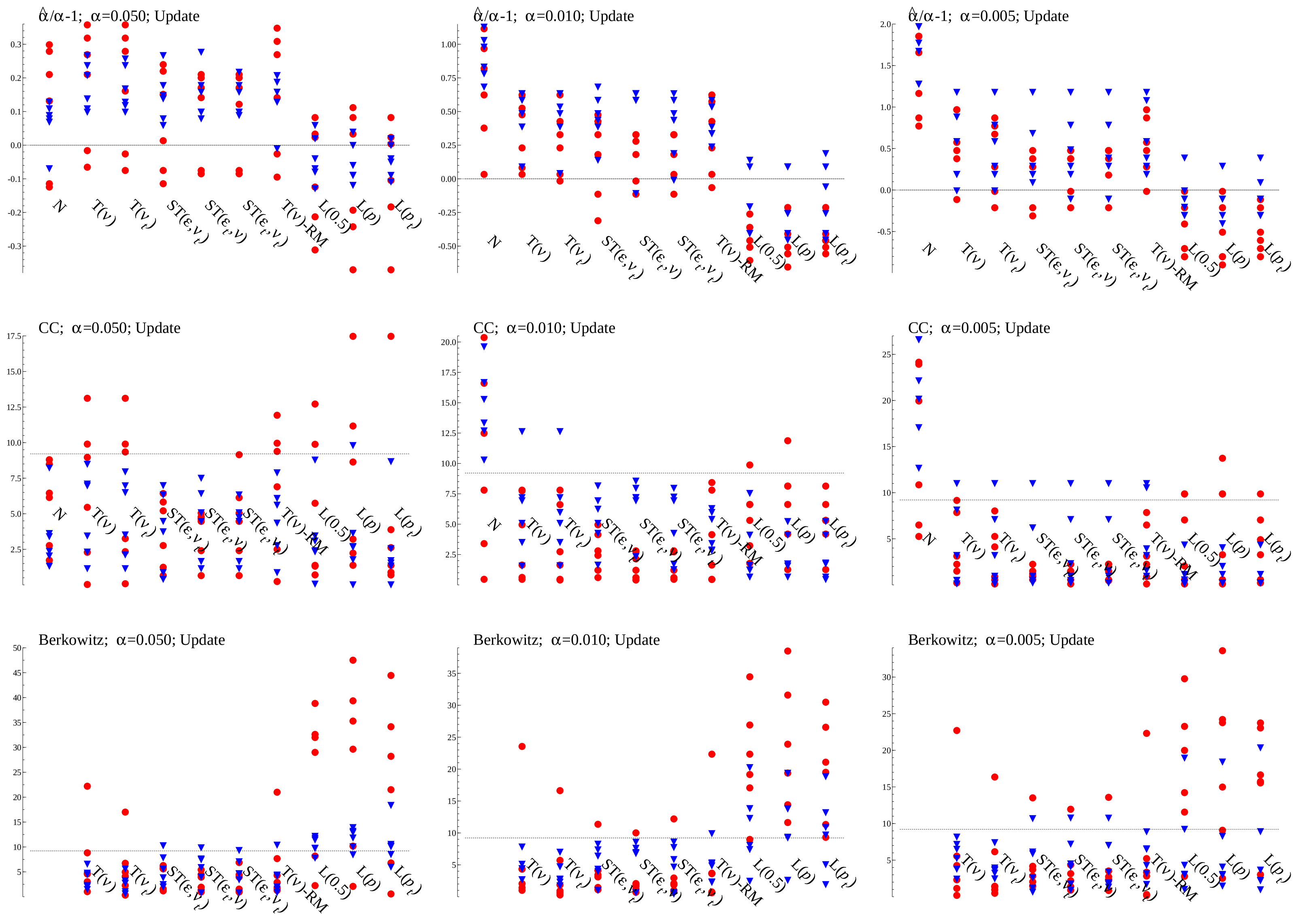

Figure 3: VaR Performance

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 6 exchange rates and 6 stock return rates, respectively. Berkowitz tests for the normal (N) are uniformly large and therefore left out.

hit rates of the different methods. The normal and Student's t based method typically result in somewhat more VaR violations compared to the nominal level. The Laplace based approaches, by contrast, result in a substantially lower number of VaR violations. The further we go out into the tails, the worse the normal based approach works in terms of 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 looking 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 either ϵ_t time-varying, ν_t, or both, appear to be much smaller.

If we consider the behavior of different approaches in capturing the tail shape beyond the VaR, the bottom row of graphs in Figure 3 shows that the Laplace distribution is clearly too thin-tailed to adequately describe the tail behavior of exchange rate and stock returns. Note that the bottom row of graphs does not show the results for the normal distribution. The Berkowitz test results for the normal are so high that they would completely distort the picture 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 less susceptible to extreme outcomes for the tests than the other Student's t based approaches. Overall, the SD-EWMA approach on 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 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. We showed that the standard EWMA and the robust Laplace based EWMA can all 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 distribution. 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 particularly suited in a context with non-Gaussian distributed observations.

We applied the new methods to forecast 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 parameter had the best overall performance for different series and different VaR confidence levels. The new score-driven EWMA approach thus 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 et al. [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 an interesting avenue for further research.

References

Ardia, D. and L. F. Hoogerheide (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., J. Ji, and A. Lucas (2015). Semiparametric score driven volatility models. *Computational Statistics & Data Analysis*, forthcoming.

Blasques, F., S. J. Koopman, and A. Lucas (2015). Information-theoretic optimality of observation-driven time series models for continuous responses. *Biometrika*, forthcoming.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*(3), 307–327.

Chen, Y. and J. Lu (2012). 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., S. J. Koopman, and A. Lucas (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., S. J. Koopman, and A. Lucas (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics 28*(5), 777–795.

Creal, D., B. Schwaab, S. J. Koopman, and A. Lucas (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., P. Zagaglia, A. Kirchner, and Z. Liu (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., Z. Lu, and H. Huang (2013). Exponentially smoothing the skewed laplace distribution for value-at-risk forecasting. *Journal of Forecasting 32*(6), 534–550.

Gómez, H. W., F. J. Torres, and H. Bolfarine (2007). Large-sample inference for the epsilon-skew-t distribution. *Communications in Statistics — Theory and Methods 36*(1), 73–81.

Guermat, C. and R. D. Harris (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. and A. Luati (2014). Filtering with heavy tails. *Journal of the American Statistical Association 109*, 1112–1122.

Janus, P., S. J. Koopman, and A. Lucas (2011). Long memory dynamics for multivariate dependence under heavy tails. *Tinbergen Institute Discussion Paper 11-175/2/DSF28*.

Jensen, M. B. and A. Lunde (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., B. Schwaab, and X. Zhang (2014). Conditional euro area sovereign default risk. *Journal of Business and Economic Statistics 32*(2), 271–284.

Poon, S.-H. and C. W. Granger (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*. Number 1st. 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.

Online supplementary appendix

A.1. Symmetric Student's t with time-varying ν_t

In this appendix we show the direct derivations needed for the symmetric Student's *t* SD-EWMA scheme. Alternatively, we could use the results from Gómez et al. [2007] concerning expressions for the score and information matrix to arrive at the same result. In that case, however, one should make sure to account for the fact that we model the variance rather than the scale parameter of the Student's *t* distribution. As the information matrix is non-diagonal between the scale parameter and the degrees of freedom parameter, and the variance is a function of both the scale parameter and the degrees of freedom parameter, this affects the precise form of the appropriate derivatives.

Define $\gamma(x) = \ln \Gamma(x)$, with first and second order derivatives $\gamma'(x)$ and $\gamma''(x)$, respectively. Given the density of the Student's *t* distribution with variance σ_t^2,

$$\ell_t(\sigma_t^2, \nu_t) = \gamma\left(\frac{\nu_t + 1}{2}\right) - \gamma\left(\frac{\nu_t}{2}\right) - \frac{1}{2}\ln\left((\nu_t - 2)\pi\sigma_t^2\right) - \frac{1}{2}(\nu_t + 1)\ln\left(1 + \frac{y_t^2}{(\nu_t - 2)\sigma_t^2}\right), \tag{A1}$$

we obtain

$$\begin{aligned}\frac{\partial \ell_t(\sigma_t^2, \nu_t)}{\partial \nu_t} &= \frac{1}{2}\gamma'\left(\frac{\nu_t + 1}{2}\right) - \frac{1}{2}\gamma'\left(\frac{\nu_t}{2}\right) - \frac{1}{2}\frac{1}{\nu_t - 2} - \frac{1}{2}\ln\left(1 + \frac{y_t^2}{(\nu_t - 2)\sigma_t^2}\right) \\ &\quad + \frac{1}{2}(\nu_t + 1)\frac{\frac{y_t^2}{(\nu_t-2)^2\sigma_t^2}}{1 + \frac{y_t^2}{(\nu_t-2)\sigma_t^2}},\end{aligned} \tag{A2}$$

with $\mathrm{E}[\partial \ell_t(\sigma_t^2, \nu_t)/\partial \nu_t] = 0$. Taking further derivatives, we obtain

$$\begin{aligned}\frac{\partial^2 \ell_t(\sigma_t^2, \nu_t)}{(\partial \nu_t)^2} &= \frac{1}{4}\gamma''\left(\frac{\nu_t + 1}{2}\right) - \frac{1}{4}\gamma''\left(\frac{\nu_t}{2}\right) + \frac{1}{2}\frac{1}{(\nu_t - 2)^2} + \frac{\frac{y_t^2}{(\nu_t-2)^2\sigma_t^2}}{1 + \frac{y_t^2}{(\nu_t-2)\sigma_t^2}} \\ &\quad + \frac{1}{2}(\nu_t + 1)\frac{-2\left(1 + \frac{y_t^2}{(\nu_t-2)\sigma_t^2}\right)\frac{y_t^2}{(\nu_t-2)^3\sigma_t^2} + \frac{y_t^2}{(\nu_t-2)^2\sigma_t^2}\left(\frac{y_t^2}{(\nu_t-2)^2\sigma_t^2}\right)}{\left(1 + \frac{y_t^2}{(\nu_t-2)\sigma_t^2}\right)^2} \\ &= \frac{1}{4}\gamma''\left(\frac{\nu_t + 1}{2}\right) - \frac{1}{4}\gamma''\left(\frac{\nu_t}{2}\right) + \frac{1}{2}\frac{1}{(\nu_t - 2)^2} + \frac{\frac{y_t^2}{(\nu_t-2)^2\sigma_t^2}}{1 + \frac{y_t^2}{(\nu_t-2)\sigma_t^2}} \\ &\quad - \frac{1}{2}(\nu_t + 1)\frac{2\frac{y_t^2}{(\nu_t-2)^3\sigma_t^2} + \frac{y_t^4}{(\nu_t-2)^4\sigma_t^4}}{\left(1 + \frac{y_t^2}{(\nu_t-2)\sigma_t^2}\right)^2},\end{aligned}$$

Using the transformation of variables $\nu_t^{1/2}(\nu_t-2)^{-1/2}y_t/\sigma_t \to y_t$, we have that for some $a,b>0$

$$
\begin{aligned}
q(a,b,\nu_t) &= \mathrm{E}\left[\frac{\left(\frac{y_t^2}{(\nu_t-2)\sigma_t^2}\right)^a}{\left(1+\frac{y_t^2}{(\nu_t-2)\sigma_t^2}\right)^b}\right] \\
&= \int \frac{\left(\frac{y^2}{(\nu_t-2)\sigma_t^2}\right)^a}{\left(1+\frac{y^2}{(\nu_t-2)\sigma_t^2}\right)^b} \frac{\Gamma\left(\frac{\nu_t+1}{2}\right)}{\Gamma\left(\frac{\nu_t}{2}\right)\sqrt{(\nu_t-2)\sigma_t^2\pi}} \frac{1}{\left(1+\frac{y^2}{(\nu_t-2)\sigma_t^2}\right)^{(\nu_t+1)/2}}\mathrm{d}y \\
&= \int \frac{\Gamma\left(\frac{\nu_t+1}{2}\right)}{\Gamma\left(\frac{\nu_t}{2}\right)\sqrt{(\nu_t-2)\sigma_t^2\pi}} \frac{\left(\frac{y^2}{(\nu_t-2)\sigma_t^2}\right)^a}{\left(1+\frac{y^2}{(\nu_t-2)\sigma_t^2}\right)^{(\nu_t+2b+1)/2}}\mathrm{d}y \\
&= \int \frac{\Gamma\left(\frac{\nu_t+1}{2}\right)}{\Gamma\left(\frac{\nu_t}{2}\right)\sqrt{\nu_t\pi}} \frac{(y^2/\nu_t)^a}{\left(1+\frac{y^2}{\nu_t}\right)^{(\nu_t+2b+1)/2}}\mathrm{d}y \\
&= \frac{\Gamma\left(\frac{\nu_t+1}{2}\right)\Gamma\left(\frac{\nu_t+2b}{2}\right)(\nu_t+2b)^a}{\Gamma\left(\frac{\nu_t+2b+1}{2}\right)\Gamma\left(\frac{\nu_t}{2}\right)\nu_t^a}\mathrm{E}[z_1^{2a}z_2^{-a}],
\end{aligned}
$$

with $z_1 \sim \mathrm{N}(0,\nu_t/(\nu_t+2b))$, $z_2 \sim \chi^2(\nu_t+2b)$, $z_3 = (\nu_t+2b)^{1/2}z_1/\nu_t^{1/2} \sim \mathrm{N}(0,1)$, z_1 and z_2 are independent. Note that if $z_4 \sim \chi^2(\nu_t)$, then

$$
\mathrm{E}[z_4^a] = \int \frac{z^a}{\Gamma\left(\frac{\nu_t}{2}\right)2^{\nu_t/2}} z^{(\nu_t/2)-1}e^{-z/2}\mathrm{d}z = \frac{\Gamma\left(\frac{\nu_t}{2}+a\right)2^a}{\Gamma\left(\frac{\nu_t}{2}\right)},
$$

such that

$$
\begin{aligned}
q(a,b,\nu_t) &= \frac{\Gamma\left(\frac{\nu_t+1}{2}\right)\Gamma\left(\frac{\nu_t+2b}{2}\right)}{\Gamma\left(\frac{\nu_t+2b+1}{2}\right)\Gamma\left(\frac{\nu_t}{2}\right)}\mathrm{E}[z_3^{2a}z_2^{-a}] \\
&= \frac{\Gamma\left(\frac{\nu_t+1}{2}\right)\Gamma\left(\frac{\nu_t+2b}{2}\right)\Gamma\left(\frac{1}{2}+a\right)\Gamma\left(\frac{\nu_t+2b}{2}-a\right)}{\Gamma\left(\frac{\nu_t+2b+1}{2}\right)\Gamma\left(\frac{\nu_t}{2}\right)\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{\nu_t+2b}{2}\right)} \\
&= \frac{\Gamma\left(\frac{\nu_t+1}{2}\right)\Gamma\left(\frac{1}{2}+a\right)\Gamma\left(\frac{\nu_t+2b}{2}-a\right)}{\Gamma\left(\frac{\nu_t+2b+1}{2}\right)\Gamma\left(\frac{\nu_t}{2}\right)\Gamma\left(\frac{1}{2}\right)} = \frac{\Gamma\left(\frac{\nu_t+1}{2}\right)}{\Gamma\left(\frac{\nu_t+1}{2}+b\right)} \frac{\Gamma\left(\frac{\nu_t}{2}+b-a\right)}{\Gamma\left(\frac{\nu_t}{2}\right)} \frac{\Gamma\left(\frac{1}{2}+a\right)}{\Gamma\left(\frac{1}{2}\right)}.
\end{aligned}
$$

We use the expression for $q(a,b,\nu_t)$ to rewrite

$$\begin{aligned}
\mathrm{E}\left[\frac{\partial^2 \ell_t(\sigma_t^2,\nu_t)}{(\partial \nu_t)^2}\right] &= \frac{1}{4}\gamma''\left(\frac{\nu_t+1}{2}\right) - \frac{1}{4}\gamma''\left(\frac{\nu_t}{2}\right) + \frac{1}{2}\,\frac{1}{(\nu_t-2)^2} + \frac{q(1,1,\nu_t)}{\nu_t-2} \\
&\qquad -\frac{1}{2}(\nu_t+1)\left(\frac{2\,q(1,2,\nu_t)+q(2,2,\nu_t)}{(\nu_t-2)^2}\right) \\
&= \frac{1}{4}\gamma''\left(\frac{\nu_t+1}{2}\right) - \frac{1}{4}\gamma''\left(\frac{\nu_t}{2}\right) + \frac{1}{2}\,\frac{1}{(\nu_t-2)^2} + \frac{1}{(\nu_t+1)(\nu_t-2)} \\
&\qquad -\frac{1}{2}(\nu_t+1)\left(\frac{\frac{2\nu_t}{(\nu_t+1)(\nu_t+3)}+\frac{3}{(\nu_t+1)(\nu_t+3)}}{(\nu_t-2)^2}\right) \\
&= \frac{1}{4}\gamma''\left(\frac{\nu_t+1}{2}\right) - \frac{1}{4}\gamma''\left(\frac{\nu_t}{2}\right) + \frac{1}{2}\,\frac{1}{(\nu_t-2)^2(\nu_t+1)(\nu_t+3)} \times \\
&\qquad ((\nu_t+1)(\nu_t+3)+2(\nu_t+3)(\nu_t-2)-(\nu_t+1)(2\nu_t+3)) \\
&= \frac{1}{4}\gamma''\left(\frac{\nu_t+1}{2}\right) - \frac{1}{4}\gamma''\left(\frac{\nu_t}{2}\right) + \frac{1}{2}\,\frac{1}{(\nu_t-2)^2(\nu_t+1)(\nu_t+3)} \times \\
&\qquad \left(\nu_t^2+4\nu_t+3+2\nu_t^2+2\nu_t-12-2\nu_t^2-5\nu_t-3\right) \\
&= \frac{1}{4}\gamma''\left(\frac{\nu_t+1}{2}\right) - \frac{1}{4}\gamma''\left(\frac{\nu_t}{2}\right) + \frac{1}{2}\,\frac{\nu_t^2+\nu_t-12}{(\nu_t-2)^2(\nu_t+1)(\nu_t+3)} \\
&= \frac{1}{4}\gamma''\left(\frac{\nu_t+1}{2}\right) - \frac{1}{4}\gamma''\left(\frac{\nu_t}{2}\right) + \frac{1}{2}\,\frac{(\nu_t+4)(\nu_t-3)}{(\nu_t-2)^2(\nu_t+1)(\nu_t+3)}
\end{aligned}$$

Note that if we use the parameterization $\nu(f_t)$ with first and second derivatives $\dot{\nu}_t = \dot{\nu}(f_t) = \partial\nu(f_t)/\partial f_t$ and $\ddot{\nu}_t = \partial\dot{\nu}(f_t)/\partial f_t$, respectively, we have

$$\begin{aligned}
\mathrm{E}\left[\frac{\partial\ \ell_t(\sigma_t^2,\nu_t)}{\partial f_t}\right] &= \dot{\nu}_t\,\mathrm{E}\left[\frac{\partial \ell_t(\sigma_t^2,\nu_t)}{\partial \nu_t}\right] = 0, \\
\mathrm{E}\left[\frac{\partial^2 \ell_t(\sigma_t^2,\nu_t)}{(\partial f_t)^2}\right] &= \ddot{\nu}_t\,\mathrm{E}\left[\frac{\partial \ell_t(\sigma_t^2,\nu_t)}{\partial \nu_t}\right] + (\dot{\nu}_t)^2\,\mathrm{E}\left[\frac{\partial^2 \ell_t(\sigma_t^2,\nu_t)}{(\partial \nu_t)^2}\right] = (\dot{\nu}_t)^2\,\mathrm{E}\left[\frac{\partial^2 \ell_t(\sigma_t^2,\nu_t)}{(\partial \nu_t)^2}\right].
\end{aligned}$$

With inverse Fisher information scaling and thus using minus the expected hessian, we obtain the steps

$$-(\dot{\nu}_t)^{-1}\left(\mathrm{E}\left[\frac{\partial^2 \ell_t(\sigma_t^2,\nu_t)}{(\partial \nu_t)^2}\right]\right)^{-1}\frac{\partial \ell_t(\sigma_t^2,\nu_t)}{\partial \nu_t}.$$

A.2. Skewed Student's t with time-varying ϵ_t and ν_t

We make the following definitions. Let $\bar{\mu}_t$ and $\bar{\sigma}_t$ denote the location and scale parameter of the skewed Student's t distribution with skewness parameter ϵ_t and degrees of freedom parameter ν_t. Let μ_t and σ_t denote the mean and standard deviation of the Student's t distribution, assuming $\nu_t > 2$. Following Gómez et al. [2007], we have

$$
\begin{aligned}
c(\nu_t) &= \frac{\Gamma\left(\frac{\nu_t+1}{2}\right)}{\sqrt{\nu_t \pi}\Gamma(\nu_t/2)}, \\
\mu_t &= \bar{\mu}_t - \frac{4c(\nu_t)\epsilon_t\bar{\sigma}_t\nu_t}{\nu_t - 1}, \\
\sigma_t^2 &= \bar{\sigma}_t^2 \left(\frac{\nu_t(1+3\epsilon_t^2)}{\nu_t - 2} - \frac{16c(\nu_t)^2\epsilon_t^2\nu_t^2}{(\nu_t-1)^2}\right), \\
\mathcal{I}_t &= \frac{1}{\nu_t+3}\begin{pmatrix} \frac{\nu_t+1}{\bar{\sigma}_t^2(1-\epsilon_t^2)} & 0 & \frac{-4c(\nu_t)(\nu_t+1)}{\bar{\sigma}_t(1-\epsilon_t^2)} & 0 \\ 0 & \frac{\nu_t}{2\bar{\sigma}_t^4} & 0 & \frac{-1}{\bar{\sigma}_t^2(\nu_t+1)} \\ \frac{-4c(\nu_t)(\nu_t+1)}{\bar{\sigma}_t(1-\epsilon_t^2)} & 0 & \frac{3(\nu_t+1)}{1-\epsilon_t^2} & 0 \\ 0 & \frac{-1}{\bar{\sigma}_t^2(\nu_t+1)} & 0 & -h(\nu_t)(\nu_t+3) - \frac{\nu_t+5}{2\nu_t(\nu_t+1)} \end{pmatrix}, \\
h(\nu_t) &= 0.25 \cdot (\psi'((\nu_t+1)/2) - \psi'(\nu_t/2)), \\
c'(\nu_t) &= 0.5c(\nu_t) \cdot (\psi((\nu_t+1)/2) - \psi(\nu_t/2) - \nu_t^{-1}),
\end{aligned}
$$

where ψ is the digamma function, and ψ' the trigamma function. Define the transformations from the main text,

$$
\begin{aligned}
\sigma_t^2 &= \exp(f_{1,t}), \\
\epsilon_t &= \tanh(f_{2,t}), \\
\nu_t &= 51 + 49\tanh(f_{3,t}),
\end{aligned}
$$

with $f_t = (f_{1,t}, f_{2,t}, f_{3,t})'$. We have

$$H_{1,t} = \frac{\partial\left(\bar{\mu}, \bar{\sigma}_t^2, \epsilon_t, \nu_t\right)}{\partial\left(\bar{\mu}, \bar{\sigma}_t^2, \epsilon_t, \nu_t\right)'} = \left(\frac{\partial\left(\bar{\mu}, \bar{\sigma}_t^2, \epsilon_t, \nu_t\right)}{\partial\left(\bar{\mu}, \bar{\sigma}_t^2, \epsilon_t, \nu_t\right)'}\right)^{-1}$$

$$= \begin{pmatrix} 1 & \frac{-2c(\nu_t)\epsilon_t\nu_t}{\bar{\sigma}_t(\nu_t-1)} & \frac{-4c(\nu_t)\bar{\sigma}_t\nu_t}{\nu_t-1} & \frac{4c(\nu_t)\epsilon_t\bar{\sigma}_t}{(\nu_t-1)^2} - \frac{4c'(\nu_t)\epsilon_t\bar{\sigma}_t\nu_t}{\nu_t-1} \\ 0 & \frac{\nu_t(1+3\epsilon_t^2)}{\nu_t-2} - \frac{16c(\nu_t)^2\epsilon_t^2\nu_t^2}{(\nu_t-1)^2} & \bar{\sigma}_t^2\left(\frac{6\epsilon_t\nu_t}{\nu_t-2} - \frac{32c(\nu_t)^2\epsilon_t\nu_t^2}{(\nu_t-1)^2}\right) & \bar{\sigma}_t^2\left(\frac{-2(1+3\epsilon_t^2)}{(\nu_t-2)^2} - \frac{32c'(\nu_t)c(\nu_t)\epsilon_t^2\nu_t^2}{(\nu_t-1)^2} + \frac{32c(\nu_t)^2\epsilon_t^2\nu_t}{(\nu_t-1)^3}\right) \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}^{-1},$$

$$H_{2,t} = \frac{\partial\left(\mu, \sigma_t^2, \epsilon_t, \nu_t\right)}{\partial f_t'} = \begin{pmatrix} 0 & 0 & 0 \\ \sigma_t^2 & 0 & 0 \\ 0 & 1-\epsilon_t^2 & 0 \\ 0 & 0 & 49-(\nu_t-51)^2/49 \end{pmatrix},$$

and

$$\nabla_t = \frac{\partial \log p(y_t|f_t)}{\partial(\bar{\mu}_t, \bar{\sigma}_t, \epsilon_t, \nu_t)'} = \begin{pmatrix} \frac{(\nu_t+1)(y_t-\bar{\mu}_t)/(\bar{\sigma}_t^2\nu_t\tilde{\epsilon}(y_t))}{1+Y_t^2} \\ \frac{-1}{2\bar{\sigma}_t^2} + \frac{\nu_t+1}{2}\frac{Y_t^2/\bar{\sigma}_t^2}{1+Y_t^2} \\ (\nu_t+1)(1-\operatorname{sign}(y_t-\bar{\mu}_t)\cdot\epsilon_t)\operatorname{sign}(y_t-\bar{\mu}_t)\frac{Y_t^2/\tilde{\epsilon}(y_t)}{1+Y_t^2} \\ \frac{c'(\nu_t)}{c(\nu_t)} - \frac{1}{2}\log(1+Y_t^2) + \frac{(\nu_t^{-1}+1)Y_t^2}{2(1+Y_t^2)} \end{pmatrix},$$

with $\tilde{\epsilon}(y_t) = (1-\operatorname{sign}(y_t-\bar{\mu}_t)\cdot\epsilon_t)^2$, and $Y_t^2 = (y_t-\bar{\mu}_t)^2/(\bar{\sigma}_t^2\nu_t\tilde{\epsilon}(y_t))$.

The score with respect to f_t is now given by

$$\frac{\partial \log p(y_t|f_t)}{\partial f_t} = H_{2,t}'H_{1,t}'\nabla_t.$$

We scale each of these elements by the inverse diagonal elements of

$$H_{2,t}'H_{1,t}'\mathcal{I}_tH_{1,t}H_{2,t}$$

to obtain three univariate recursions.

If, for example, only σ_t^2 and ϵ_t follow an SD-EWMA scheme, while ν_t is constant, define the selection matrix S such that

$$Sf_t = (f_{1,t}, f_{2,t})'.$$

The two univariate recursions are then driven by

$$\frac{\partial \log p(y_t|f_t)}{\partial (f_{1,t}, f_{2,t})'} = S H_{2,t}' H_{1,t}' \nabla_t,$$

scaled by the inverse diagonal elements of

$$S H_{2,t}' H_{1,t}' \mathcal{I}_t H_{1,t} H_{2,t} S'.$$

Similar formulas hold for other combinations of score-driven and fixed parameters.