JOURNAL OF TIME SERIES ANALYSIS
J. Time Ser. Anal. **39**: 909–919 (2018)
Published online 29 July 2018 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1111/jtsa.12419

SPECIAL ISSUE

MODELING THE INTERACTIONS BETWEEN VOLATILITY AND RETURNS USING EGARCH-M

ANDREW HARVEY[a*] AND RUTGER-JAN LANGE[b]
[a] *Faculty of Economics, University of Cambridge, UK*
[b] *Econometric Institute, Erasmus University Rotterdam, Rotterdam, The Netherlands*

An EGARCH-M model, in which the logarithm of scale is driven by the score of the conditional distribution, is shown to be theoretically tractable as well as practically useful. A two-component extension makes it possible to distinguish between the short- and long-run effects of returns on volatility, and the resulting short- and long-run volatility components are then allowed to have different effects on returns, with the long-run component yielding the equity risk premium. The EGARCH formulation allows for more flexibility in the asymmetry of the volatility response (leverage) than standard GARCH models and suggests that, for weekly observations on two major stock market indices, the short-term response is close to being anti-symmetric.

Received 02 March 2018; Accepted 15 June 2018

Keywords: ARCH-in-mean; dynamic conditional score (DCS) model; equity risk premium; leverage; two-component model.

JEL. C22; G15.

MOS subject classification: 91B84; 62M10; 37M10; 91G70.

1. INTRODUCTION

Uncertainty in stock returns, as captured by volatility, contributes to the equity risk premium so causing a positive correlation between volatility and returns; see French *et al.* (1987). The standard textbook model for introducing a time-varying risk premium into returns is the autoregressive conditional heteroscedasticity in mean model, or simply the ARCH-M model; see, for example, Taylor (2005, pp. 205, 252–254).

An ARCH-M model with two components of volatility has a number of attractions, one of which is to account for the long-memory behaviour often seen in the autocorrelations of absolute values of returns or their squares; see, for example, Harvey (2013, p. 135). However, in contrast to a long-memory model, a two-component model enables the researcher to distinguish between the effects of short- and long-run volatility on returns. Short-run volatility can lead to a 'news effect', as described by Chou (1988) and Schwert (1989), that makes investors nervous of risk and predicts a negative correlation between volatility and return. This negative relationship contrasts with the positive relationship between long-run volatility and return predicted by the inter-temporal capital asset pricing model (ICAPM). Failure to model both aspects of volatility has led to inconclusive results regarding the sign of the correlation between volatility and the equity risk premium. For example, the risk premium is negative and significant according to Nelson (1991), positive but insignificant according to French *et al.* (1987) and Campbell and Hentschel (1992), and positive or negative (depending on the method) according Glosten *et al.* (1993) and Turner *et al.* (1989).

Returns may have an asymmetric effect on volatility, with negative returns typically associated with increased volatility; see, for example, Bekaert and Wu (2000). A two-component model is able to allow for different asymmetric responses in the short and long run. In particular, it may be that asymmetry is confined to the short-run

* Correspondence to: Andrew Harvey, Faculty of Economics, Austin Robinson Building, Sidgwick Avenue, Cambridge CB3 9DD, UK.
Email: andrew.harvey@econ.cam.ac.uk

component, as found by Engle and Lee (1999) and others. Indeed, short-run volatility may even decrease after a good day, because it calms the market. Standard GARCH models are unable to identify this effect.

Adrian and Rosenberg (2008) proposed a two-component exponential GARCH-M (EGARCH-M) model. They showed that it captures the asymmetric response of short- and long-term volatility to returns, and that the long-run component is positively correlated with the equity risk premium. Here we demonstrate that letting the dynamics be driven by the score of the conditional distribution yields a model that is theoretically tractable as well as practically useful. Models constructed using the conditional score were introduced into the literature by Creal *et al.* (2011, 2013), where they are called Generalized Autoregressive Score (GAS) models, and Harvey (2013), where they are called Dynamic Conditional Score (DCS) models. The classic EGARCH specification in Nelson (1991), which is the one used by Adrian and Rosenberg, is sensitive to outliers and it has the unfortunate theoretical property that unconditional moments do not exist when the conditional distribution is Student's t. The DCS model resolves these problems and, in doing so, yields a specification which is open to the development of a full asymptotic theory for the distribution of the maximum likelihood estimator. This contrasts with standard formulations of ARCH-M models, where the asymptotic theory appears to be intractable.

Section 2 sets out the DCS formulation of the basic EGARCH-M model with a conditional Student's t-distribution and outlines the associated statistical theory, including an easily implemented condition for invertibility. This is followed by subsections on the asymmetric response of volatility to returns (leverage), two components, skewness and splines. In Section 3 various models are fitted to weekly data on NASDAQ and NIKKEI returns. The risk-free rate of return is then introduced as an explanatory variable, as in Scruggs (1998), and the model is fitted to weekly excess returns on S&P500 over a 60-year period. Section 4 reports the results of a small forecasting study to provide reassurance on the predictive performance of DCS models. Section 5 concludes. Appendices A to C are provided online as Appendix S1 (supporting information).

2. MODEL FORMULATION

In the ARCH-M model, as introduced by Engle *et al.* (1987), returns, y_t, are subject to a time-varying equity risk premium as follows:

$$y_t = \mu + \alpha\sigma^m_{t|t-1} + \varepsilon_t\sigma_{t|t-1}, \quad t = 1, \ldots, T, \tag{1}$$

where $\sigma_{t|t-1}$ is the conditional standard deviation, ε_t is a serially independent standard normal variable, that is $\varepsilon_t \sim NID(0, 1)$, μ and α are parameters and m is typically set to one or two. The conditional variance, $\sigma^2_{t|t-1}$, depends on past squared observations, as in a GARCH model.

Engle *et al.* (1987), and most subsequent studies, find that the standard deviation, that is $m = 1$, gives the best fit and so the DCS EGARCH-M is set up as

$$y_t = \mu + \alpha \exp(\lambda_{t|t-1}) + \varepsilon_t \exp(\lambda_{t|t-1}), \quad t = 1, \ldots, T, \tag{2}$$

where $\exp(\lambda_{t|t-1})$ is the scale, with the dynamic equation for $\lambda_{t|t-1}$ driven by the score of the conditional distribution of y_t at time t, that is the first derivative with respect to $\lambda_{t|t-1}$ of the logarithm of the probability density function at time t. The stationary first-order dynamic model for $\lambda_{t|t-1}$ is

$$\lambda_{t+1|t} = \omega(1 - \phi) + \phi\lambda_{t|t-1} + \kappa u_t, \quad |\phi| < 1, \tag{3}$$

where u_t is the conditional score and ϕ, κ and ω are parameters, with ω denoting the unconditional mean of $\lambda_{t|t-1}$, and $\lambda_{1|0} = \omega$. The score with respect to $\lambda_{t|t-1}$ for a conditional Student's t-distribution with ν degrees of freedom and scale $\exp(\lambda_{t|t-1})$ is

$$u_t = (\nu + 1)b_t - 1 + \alpha(1 - b_t)[(\nu + 1)/\nu]\varepsilon_t, \tag{4}$$

DOI: 10.1111/jtsa.12419

where $\varepsilon_t = (y_t - \mu)e^{-\lambda_{t|t-1}} - \alpha$ and

$$b_t = \frac{\varepsilon_t^2/\nu}{1+\varepsilon_t^2/\nu}, \qquad 0 \le b_t \le 1, \quad 0 < \nu < \infty, \tag{5}$$

is distributed as $beta(1/2, \nu/2)$. In the absence of the ARCH-M effect in (2), this model is known as Beta-t-EGARCH. The fact that the score is bounded when ν is finite means that the impact of outliers is limited.

The last term in (4) depends on ε_t, the appearance of which reflects the fact that, like the first term, it is informative about the movements in $\lambda_{t+1|t}$. However, the inclusion of this 'ARCH-M score term' is not crucial to the model because α is typically very small: dropping it makes very little practical difference and the theoretical properties of the model as a whole are simpler.

2.1. Moments, Autocorrelations and Predictions

When $\lambda_{t|t-1}$ is stationary, the unconditional moments of the observations exist whenever the corresponding conditional moment exists. Furthermore, if the ARCH-M score term is dropped from (4), exact analytic expressions for moments and autocorrelations of $|y_t|^c$ may be derived for any nonnegative c; compare Harvey (2013, ch 4). Moments of future observations can also be found and the full predictive distribution readily simulated.

The presence of the ARCH-M component means that returns are serially correlated with the autocorrelation function given by

$$\rho(\tau) = \frac{E(\exp(\lambda_{t|t-1} + \lambda_{t-\tau|t-\tau-1})) - [E(\exp\lambda_{t|t-1})]^2}{(1+\sigma_\varepsilon^2/\alpha^2)E(\exp 2\lambda_{t|t-1}) - [E(\exp\lambda_{t|t-1})]^2}, \qquad \tau = 1, 2, 3, \ldots, \tag{6}$$

when the distribution of the ε_t's is symmetric; see Appendix A. These autocorrelations do not depend on ω. When $\alpha \neq 0$, $\rho(\tau)$ has the same sign as the corresponding autocorrelation of the volatility, $\exp\lambda_{t|t-1}$, with a pattern derived from that of the $\lambda_{t|t-1}$'s; compare the GARCH-M autocorrelations in Hong (1991).

2.2. Invertibility and Estimation

Following the discussion in Blasques *et al.* (2014, 2018), a sufficient condition for invertibility can be found by generalizing the result in Harvey and Lange (2017, p. 182). The proposition below is for the *Beta-t-EGARCH-M* model, (2) to (5), but with the ARCH-M score term dropped from (4).

Proposition 1. For the Beta-t-EGARCH-M model to be invertible, it is sufficient that

$$\left|\phi + \frac{\kappa\alpha^2(\nu+1)}{2\nu}\right| < 1 \quad \text{and} \quad \left|\phi - \frac{\kappa(\nu+1)}{2}\frac{\nu}{\nu + 2\alpha^2 - 2|\alpha|\sqrt{\alpha^2+\nu}}\right| < 1. \tag{7}$$

The proof is contained in Appendix B. Although condition (7) is almost certainly too strong, the admissible parameter space is sufficiently large to be useful for empirical work and an upper bound of thirty degrees of freedom[1] will usually guarantee invertibility. Classic EGARCH-(M), as used by Adrian and Rosenberg (2008), does not satisfy our sufficient condition for invertibility.

To derive the asymptotic distribution of the maximum likelihood (ML) estimator for Beta-t-EGARCH-M, we follow the approach in Harvey (2013).

[1] For example, for $\phi = 0.99$, $\kappa = 0.10$ and $\nu = 30$, the first condition implies $|\alpha| < 0.44$, while the second only needs $|\alpha| < 0.69$. Values of α found in empirical work depend on the frequency of observations, but rarely exceed 0.10.

DOI: 10.1111/jtsa.12419

Proposition 2. Let the sufficient condition for invertibility (7) hold. Suppose μ and ν are known. Let $\widetilde{\boldsymbol{\psi}}$ denote the ML estimator of the parameters $\boldsymbol{\psi} = (\kappa, \phi, \omega, \alpha)'$, while $\boldsymbol{\psi}_0$ denotes the true value. Then the limiting distribution of $\sqrt{T}(\widetilde{\boldsymbol{\psi}} - \boldsymbol{\psi}_0)$ is multivariate normal with mean zero and a covariance matrix given by the inverse of the information matrix $\mathbf{I}(\boldsymbol{\psi}_0)$, which is given in Appendix C but can be approximated by

$$\mathbf{I}\begin{bmatrix} \boldsymbol{\psi} \\ \alpha \end{bmatrix} \simeq \frac{\nu+1}{\nu+3}\begin{bmatrix} (2(\nu/(\nu+1)) + \alpha^2)\mathbf{D} & \alpha\mathbf{d} \\ \alpha\mathbf{d}' & 1 \end{bmatrix}, \tag{8}$$

where $\mathbf{D}$ is defined in Harvey (2013, p. 37) and $\mathbf{d} = (0, 0, (1-\phi)/(1-a))$.

The Monte Carlo experiments reported in Appendix C support the asymptotic theory and show that the approximation is a good one. Although the results are for a very basic model, they suggests that the standard asyptototic theory will apply to more general EGARCH-M specifications.

In the model of Adrian and Rosenberg (2008), the ARCH-M term is the logarithm of the scale, so $\alpha\lambda_{t|t-1}$ appears in place of $\alpha\exp(\lambda_{t|t-1})$ in (2). In this case an analytic expression for the information matrix cannot be obtained.

2.3. Asymmetric Impact Curves (Leverage)

Asymmetry in the impact of returns on volatility is captured in a Beta-t-EGARCH model by modifying the dynamic equation in (3) to

$$\lambda_{t+1|t} = \omega(1-\phi) + \phi\,\lambda_{t|t-1} + \kappa\,u_t + \kappa^* u_t^*, \tag{9}$$

where $u_t^* = sgn(-\varepsilon_t)(u_t + 1)$ and κ^* is a new parameter which, because the negative of the sign of the return is taken, is usually positive; see Harvey (2013, p. 109). When the distribution of ε_t is symmetric and u_t is defined without the ARCH-M term of (4), u_t^* has zero mean and is orthogonal to u_t, in that $E(u_t u_t^*) = 0$. The model allows for the possibility that volatility can go down when returns are positive, something that is not normally possible with a GARCH model; see Glosten *et al.* (1993, p. 1788). Identifiability requires only that either κ or κ^* is nonzero. The invertibility condition (7) may be extended to account for leverage (details are available on request) and the information matrix in Harvey (2013, pp. 121–124) generalized.

2.4. Two Components

Instead of capturing long memory by a fractionally integrated process, as in Christensen *et al.* (2010), two components may be used. Thus

$$\lambda_{t|t-1} = \omega + \lambda_{1,t|t-1} + \lambda_{2,t|t-1}, \qquad t = 1, \ldots, T, \tag{10}$$

$$\lambda_{i,t+1|t} = \phi_i\,\lambda_{i,t|t-1} + \kappa_i\,u_t + \kappa_i^*\,u_t^*, \qquad i = 1, 2, \tag{11}$$

where $\phi_1 > \phi_2$ if $\lambda_{1,t|t-1}$ denotes the long-run component. Identifiability requires $\phi_1 \neq \phi_2$, which is implicitly assumed by setting $\phi_1 > \phi_2$, together with the restrictions (i) $\kappa_1 \neq 0$ or $\kappa_1^* \neq 0$ and (ii) $\kappa_2 \neq 0$ or $\kappa_2^* \neq 0$. To investigate the impact of short- and long-term volatility on returns, the equation for y_t, that is (2), is replaced by

$$y_t = \mu' + \alpha_1\exp(\omega + \lambda_{1,t|t-1}) + \alpha_2[\exp(\lambda_{2,t|t-1}) - 1] + \varepsilon_t\exp(\lambda_{t|t-1}), \tag{12}$$

where $\mu' = \mu + \alpha_2$. The equity risk premium is then captured by the long-run component. When volatility is at its equilibrium level, the risk premium is $\mu' + \alpha_1\exp\omega$. Lanne and Saikkonen (2006) showed that if the mean in a GARCH-M model can be dropped, the ARCH-M effects are estimated more precisely. The same is likely to be true here if μ' can be set to zero.

  J. Time Ser. Anal. **39**: 909–919 (2018)
DOI: 10.1111/jtsa.12419

2.5. Skewness

Using the approach of Harvey and Sucarrat (2014) to introduce skewness into the t-distribution of the EGARCH-M model means that μ_ϵ, the expectation of ϵ_t in (12), is no longer zero. The model is therefore best estimated by subtracting μ_ϵ from the disturbance term, so that

$$y_t = \mu' + \alpha_1 \exp(\omega + \lambda_{1,t|t-1}) + \alpha_2[\exp(\lambda_{2,t|t-1}) - 1] + (\epsilon_t - \mu_\epsilon)\exp(\lambda_{t|t-1}). \tag{13}$$

2.6. Explanatory Variables and Splines

The long-term component of volatility in (10) may be modelled using exogenous variables rather than a filter. Hence $\lambda_{1,t|t-1} = \boldsymbol{\delta}'\mathbf{x}_t$, $t = 1, \ldots, T$, where $\boldsymbol{\delta}$ is a vector of parameters, $\mathbf{x}_t$ contains exogenous variables and the dynamic equation for $\lambda_{2,t|t-1}$ remains as in (11). Engle and Rangel (2008) propose the use of a quadratic spline, in which case the elements of $\mathbf{x}_t$ are chosen such that

$$\lambda_{1,t|t-1} = \delta_0\, t + \sum_{i=1}^{k} \delta_i \max\{t - t_{i-1}, 0\}^2, \qquad t = 1, \ldots, T, \tag{14}$$

where $\{t_0 = 0, t_1, \ldots, t_k = T\}$ denotes the partitioning of the horizon T into k equally spaced intervals and the "optimal" number of knots k is determined by an information criterion. The spline model is less convenient for forecasting, but a referee has suggested that it may be better for estimating the ARCH-M effect over a fixed period of time.

3. RESULTS

Excess returns y_t are defined as the log return minus the risk-free return $r_{f,t}$, that is $y_t = 100 \ln(I_t/I_{t-1}) - r_{f,t}$, where I_t is the closing price of the index and $r_{f,t}$ is proxied by the secondary market for 3-month US Treasury bills.[2] No adjustments were made for dividends, as the consensus seems to be that they have little or no effect on the estimates; for example, see French *et al.* (1987) and Poon and Taylor (1992). Estimation was carried out by ML and standard errors were computed numerically. Residual serial correlation in location and scale was computed using scores rather than raw residuals, since the latter can be sensitive to outliers. Residual serial correlation is measured by Box–Ljung test statistics constructed from the first 20 autocorrelations.[3]

3.1. NASDAQ

We first consider weekly NASDAQ excess returns from 8 February 1972 to 3 November 2014 (2282 observations). The first set of estimates in Table I are from fitting a two-component Beta-t-EGARCH model with leverage in both components. The long-term news impact curve is nearly symmetric because κ_1^* is small and insignificant. By contrast, in the short-term news impact curve κ_2^* is bigger than κ_2 so the response to large positive shocks is actually to lower volatility. A positive shock results in the two components moving in opposite directions with the initial net effect being a lowering of total volatility, because the short-term impact initially dominates.

The second set of estimates, which include both short- and long-run ARCH-M terms, produce no evidence of residual autocorrelation. The long-run ARCH-M coefficient, α_1, is positive but the short-run coefficient, α_2, is negative, suggesting that short-term volatility may induce a fall in returns, consistent with the 'news effect'.

[2] Index data are publicly available from many sources; we used Yahoo Finance. Risk-free rates for the US are available from Table H.15 of the Federal Reserve from 4 January 1954 onwards.

[3] These $Q(20)$ statistics are only a rough guide to residual serial correlation as they are not, in general, asymptotically χ^2_{20}.

DOI: 10.1111/jtsa.12419

Table I. Two-component Beta-*t*-EGARCH model estimates, with their standard errors (*in the next row*) for weekly NASDAQ excess returns from 8 February 1972 to 3 November 2014 (2282 observations)

Volatility							Mean			Shape	Fit			$Q(20)$	
κ_1	κ_1^*	ϕ_1	κ_2	κ_2^*	ϕ_2	ω	μ	α_1	α_2	ν	Log L	AIC	BIC	$u_{\mu,t}$	$u_{\lambda,t}$
.034	.004	.989	.024	.049	.724	.565	.231			7.60	−5118.3	4.4937	4.5163	53.1	21.0
.006	*.004*	*.003*	*.013*	*.009*	*.066*	*.104*	*.042*			*1.02*					
.042	.006	.984	−.006	.061	.709	.585	2.611	.147	−2.643	7.73	−5092.8	4.4731	4.5007	16.6	2.1
.006	*.005*	*.003*	*.009*	*.010*	*.048*	*.085*	*.607*	*.122*	*.562*	*1.04*					
.043		.985		.066	.727	.609	$\mu = -\alpha_2$	.133	−2.290	7.68	−5093.8	4.4714	4.4915	16.1	18.8
.006		*.005*		*.009*	*.045*	*.088*		*.033*	*.405*	*1.00*					

Table II. Two-component Beta-skew-*t*-EGARCH model estimated on NASDAQ returns as in Table I

Volatility							Mean			Shape		Fit			$Q(20)$	
κ_1	κ_1^*	ϕ_1	κ_2	κ_2^*	ϕ_2	ω	μ	α_1	α_2	ν	γ	Log L	AIC	BIC	$u_{\mu,t}$	$u_{\lambda,t}$
.043	.015	.982	.012	.059	.700	.684	2.567	.153	−2.768	8.27	.81	−5069.8	4.4539	4.4840	22.5	24.7
.007	*.005*	*.004*	*.015*	*.010*	*.052*	*.086*	*.614*	*.176*	*.512*	*1.22*	*.03*					
.044	.015	.984	.009	.059	.698	.676	$\mu = -\alpha_2$	.051	−2.717	8.18	.81	−5070.2	4.4533	4.4810	22.5	24.5
.006	*.005*	*.003*	*.010*	*.010*	*.053*	*.093*		*.036*	*.463*	*1.15*	*.03*					

The estimates indicate that $\mu \simeq -\alpha_2$. The constraint $\mu = -\alpha_2$ can be imposed by setting $\mu' = 0$ in (12). When this is done, the estimated value of α_1 is similar to before, but it is now statistically significant. The fact that κ_1^* and κ_2 are small and statistically insignificant suggests setting them to zero and, when this is done in conjunction with $\mu' = 0$, the third set of estimates is obtained. The resulting AIC and BIC are smaller than in any previous model.

Table II shows results for the EGARCH-M model with skewness. Setting $\kappa_1^* = \kappa_2 = 0$ makes very little difference to the fit. The skewness parameter, γ, which lies in the range $0 < \gamma < \infty$, is 0.81 and the evidence against symmetry, that is $\gamma = 1$, is overwhelming. Again the preferred specification has $\mu = -\alpha_2$, that is (13) with $\mu' = 0$. The expected weekly equilibrium risk premium, where $\lambda_{i,t|t-1}$ for $i = 1, 2$ are assumed to be at their equilibrium positions, is $\alpha_1 \exp(\omega) = 0.051 \exp(0.676) = 0.10$. Using continuous compounding, the average yearly risk premium is $\exp(0.10/100 \times 52) = 1.054$, that is 5.4% per year.

3.2. NIKKEI

This section examines NIKKEI weekly returns from 4 January 1984 to 20 October 2014 (1595 observations) to see if the above findings apply more generally. We examine gross returns rather than excess returns because no risk-free rate was available. As with NASDAQ, we find the long-term (short-term) news impact curve is essentially symmetric (anti-symmetric). These constraints are imposed in Table III, which shows the results for the skewed Student's *t*-distribution. The skewness parameter, like the degrees of freedom, is similar to the corresponding parameter for NASDAQ. The second set of estimates show that the short-term ARCH-M coefficient, α_2, is significantly negative and approximately equal to the estimate of $-\mu$. Residual autocorrelation is less of an issue than for NASDAQ, but again the short-term ARCH-M component is instrumental in reducing serial correlation in the location scores. The long-run ARCH-M coefficient, α_1, is positive when μ is set equal to $-\alpha_2$, but much smaller than the corresponding NASDAQ coefficient.[4] The weekly equilibrium risk premium estimate is $0.057 = 0.025 \exp(0.832)$, which implies an average return of 3% per year.

[4] This may be because the NIKKEI simple average is much smaller than the NASDAQ simple average.

  J. Time Ser. Anal. **39**: 909–919 (2018)
DOI: 10.1111/jtsa.12419

Table III. Two-component Beta-skew-t-EGARCH model estimated on weekly NIKKEI returns from 4 January 1984 to 20 October 2014 (1595 observations)

Volatility					Mean			Shape		Fit			$Q(20)$	
κ_1	ϕ_1	κ_2^*	ϕ_2	ω	μ	α_1	α_2	ν	γ	Log L	AIC	BIC	$u_{\mu,t}$	$u_{\lambda,t}$
.052	.963	.061	.826	.831	1.296	−.036	−1.156	8.70	.85	−3787.1	4.7613	4.7950	13.9	14.8
.008	*.009*	*.009*	*.043*	*.066*	*.810*	*.172*	*.557*	*1.45*	*.03*					
.052	.962	.061	.829	.832	$\mu = -\alpha_2$	.025	−1.141	8.70	.85	−3787.2	4.7602	4.7905	14.0	14.9
.008	*.009*	*.009*	*.042*	*.064*		*.036*	*.391*	*1.45*	*.03*					

Table IV. Two-component Beta-skew-t-EGARCH model with correction for risk-free rate estimated on weekly SP500 excess returns from 4 January 1954 to 3 November 2014 (3175 observations)

Volatility							Mean			Shape			Fit			$Q(20)$	
κ_1	κ_1^*	ϕ_1	κ_2	κ_2^*	ϕ_2	ω	μ	α_1	α_2	ν	γ	ξ	Log L	AIC	BIC	$u_{\mu,t}$	$u_{\lambda,t}$
.032	.009	.987	.026	.058	.798	.496	.082	.145	−.086	10.58	.82	−1.82	−6383.0	4.0290	4.0538	27.6	16.9
.005	*.005*	*.003*	*.009*	*.008*	*.049*	*.076*	*.246*	*.110*	*.235*	*1.62*	*.02*	*.63*					
.031	.011	.988	.026	.057	.814	.561	$\mu = -\alpha_2$	.132	−.252	10.54	.82	−1.59	−6383.6	4.0287	4.0516	28.1	16.2
.005	*.005*	*.002*	*.009*	*.008*	*.049*	*.096*		*.042*	*.192*	*1.60*	*.02*	*.65*					

Table V. Two-component Beta-skew-t-EGARCH model with correction for the risk-free rate, but the long-term component modelled as a quadratic spline with k knots

Knots	Volatility				Mean			Shape			Fit					
#	κ_2	κ_2^*	ϕ_2	ω	μ	α_1	α_2	ν	γ	ξ	Log L	*rank*	AIC	*rank*	BIC	*rank*
1	.038	.056	.957	.505	.090	−.033	.196	10.75	.82	−1.66	−6407.3	(15)	4.0437	(15)	4.0666	(1)
2	.038	.056	.956	.451	.012	.015	.199	10.86	.82	−1.73	−6405.1	(14)	4.0429	(14)	4.0677	(2)
3	.040	.056	.950	.210	.053	.007	.201	11.12	.82	−2.13	−6401.5	(13)	4.0413	(13)	4.0680	(4)
4	.041	.055	.949	.413	−.164	.172	.188	11.38	.82	−2.69	−6397.0	(12)	4.0390	(7)	4.0677	(2)
5	.042	.054	.943	.495	−.059	.100	.188	11.52	.82	−2.51	−6394.1	(11)	4.0378	(6)	4.0684	(5)
6	.044	.055	.935	.537	−.002	.050	.205	11.59	.82	−2.38	−6391.4	(7)	4.0368	(2)	4.0693	(6)
7	.044	.055	.938	.633	−.024	.070	.190	11.51	.82	−2.35	−6391.4	(7)	4.0374	(3)	4.0718	(7)
8	.045	.055	.933	.674	−.036	.062	.211	11.55	.82	−2.28	−6390.9	(6)	4.0377	(5)	4.0740	(8)
9	.044	.055	.938	.671	−.039	.080	.189	11.45	.82	−2.33	−6392.3	(10)	4.0392	(10)	4.0774	(9)
10	.044	.056	.934	.675	.015	.039	.196	11.46	.82	−2.26	−6391.8	(9)	4.0396	(11)	4.0797	(10)
11	.044	.055	.934	.579	−.008	.045	.210	11.62	.82	−2.25	−6389.9	(5)	4.0390	(7)	4.0810	(12)
12	.046	.055	.924	.574	−.014	.038	.229	12.15	.82	−2.28	−6385.1	(1)	4.0366	(1)	4.0805	(11)
13	.046	.053	.931	.596	−.051	.083	.209	11.98	.81	−2.49	−6385.6	(2)	4.0375	(4)	4.0833	(13)
14	.046	.054	.931	.565	−.020	.052	.227	11.96	.81	−2.45	−6386.9	(3)	4.0390	(7)	4.0867	(14)
15	.045	.055	.935	.495	−.027	.045	.229	11.73	.81	−2.25	−6388.7	(4)	4.0408	(12)	4.0904	(15)

3.3. S&P500 and the Risk-free Rate

This section investigates the risk-free return as an exogenous variable for explaining y_t, both with and without fitting a spline for the long-term volatility component. For this purpose we investigate the S&P500, which is a particularly long series of 3175 weekly observations running from 4 January 1954 to 3 November 2014.

As in Scruggs (1998) and Bjornland and Leitemo (2009), we include the risk-free rate directly as an explanatory variable in the equation for y_t, that is in (13). Table IV presents the estimates for our two-component model with skewness and leverage in both components. As expected, the estimates of ξ, the coefficient of $r_{f,t}$, are significant and negative, while the short-term (long-term) impact of returns on volatility is again close to being symmetric (anti-symmetric).

Next, we investigate whether it is beneficial to replace the filter for the long-term component by the spline formulation of Engle and Rangel (2008), as given in (14). Table V reports the results with up to fifteen knots. Using twelve knots achieves the best fit, but the resulting log likelihood falls short of the log likelihoods reported in

DOI: 10.1111/jtsa.12419

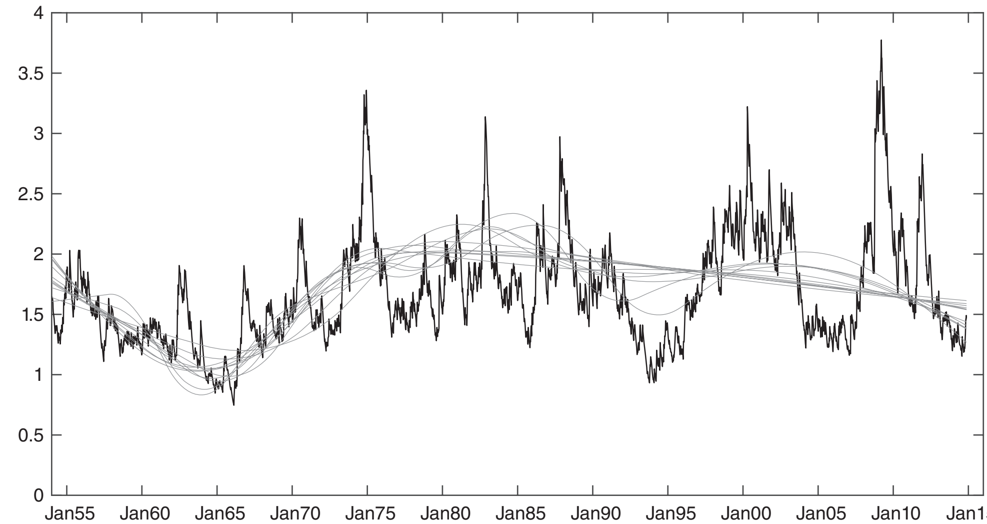

Figure 1. Comparison of long-run components from Beta-skew-t-EGARCH models fitted to S&P with the risk-free rate. Filtered scale, shown in black, and quadratic splines, with more than five knots, shown in grey

Table IV. The estimates of ϕ_2 and κ_2 in Table V are noticeably increased in comparison with Table IV, presumably to account for symmetric volatility dynamics not picked up by the spline. As a result, the short-term component becomes more persistent and its news impact curve becomes more symmetric. Hence it is not surprising that the estimates of α_1 in Table V are smaller than those in Table IV while those of α_2 are bigger. Although the other parameters are largely unaffected, it seems that using the spline makes it harder to disentangle the 'news effect' from the equity risk premium.

Figure 1 plots the long-term scale, that is $\exp(\omega + \lambda_{1,t|t-1})$, where $\lambda_{1,t|t-1}$ is from the filter, (10) or the spline, (14). The filter uses the first set of estimates in Table IV, while the splines use the estimates in Table V with at least five knots. The splines seem to match the filter reasonably well, but only if enough knots are allowed, which comes at a cost in terms of AIC or BIC. Even then, the splines, which are implemented by fixing an equally spaced partition in advance, have difficulty picking up the peak in volatility at the time of the European debt crisis, which starts in 2009. On the evidence of this application, therefore, we conclude that, from all points of view, modelling the long-term component of volatility by a filter is preferable to fitting a spline.

4. OUT-OF-SAMPLE VOLATILITY PREDICTION

Although our primary aim is to understand the interactions between volatility and returns, we can also use the model to make predictions.[5] We therefore consider weekly S&P500 log returns, with an in-sample period from 4 January 1954 to 3 January 2000 (2401 observations) and an out-of-sample period from 10 January 2000 to 10 November 2014 (774 observations). Any gains over competing models will be most apparent in volatility forecasts and so we ignore time variation in the mean, by setting $\alpha_1 = \alpha_2 = 0$, and the effect of the risk-free rate, that is $\xi = 0$. We also assume a symmetric conditional t−distribution by letting $\gamma = 1$. This leaves the model as in Table IV, but with four fewer parameters. For simplicity we estimate the parameters once, based on the in-sample period, and use these parameters to make one- and four-step-ahead predictions of the standard deviation of weekly returns in the out-of-sample period.

As the benchmark, we take the classic Engle and Lee (1999) two-component GARCH model, which, for comparability with our model, is modified so as to have a conditional Student's t-distribution and leverage in both

[5] This extension was suggested by a referee.

DOI: 10.1111/jtsa.12419

Table VI. Comparison of one- and four-step-ahead predictions of the standard deviation of weekly S&P500 returns using six loss functions as in Hansen and Lunde (2005). For each loss function and horizon the winner is shown in bold

	MSE_1	QLIKE	MAE_1	MSE_2	R^2LOG	MAE_2
One step ahead ($T = 774$)						
Our model	**0.51**	**6.67**	**0.49**	**44.6**	**0.35**	**2.76**
Engle and Lee (1999)	0.70	6.91	0.60	45.0	0.48	3.40
Four steps ahead ($T = 771$)						
Our model	**1.06**	**10.29**	**0.63**	97.8	**0.56**	**3.55**
Engle and Lee (1999)	1.11	11.96	0.65	**92.3**	0.59	3.74

components. The performance of both models is compared against a proxy for the 'true' standard deviation of weekly returns, which we take as the square root of the sum of five daily realized variances based on five-minute intraday returns obtained from the Oxford Man Institute library.[6] These data are available from 3 January 2000: hence the starting date of the out-of-sample period.

To evaluate the predictions, we take the six loss functions proposed by Hansen and Lunde (2005). These are:

$$\begin{aligned} \mathrm{MSE}_1 &= T^{-1} \textstyle\sum_{t=1}^{T} (\sigma_t - h_t)^2 & \mathrm{MSE}_2 &= T^{-1} \textstyle\sum_{t=1}^{T} (\sigma_t^2 - h_t^2)^2 \\ \mathrm{QLIKE} &= T^{-1} \textstyle\sum_{t=1}^{T} (\ln(h_t) + \sigma_t^2/h_t^2)^2 & \mathrm{R^2LOG} &= T^{-1} \textstyle\sum_{t=1}^{T} [\ln(\sigma_t^2/h_t^2)]^2 \\ \mathrm{MAE}_1 &= T^{-1} \textstyle\sum_{t=1}^{T} |\sigma_t - h_t| & \mathrm{MAE}_2 &= T^{-1} \textstyle\sum_{t=1}^{T} |\sigma_t^2 - h_t^2|, \end{aligned}$$

where σ_t is the square root of the realized variance and h_t denotes the predicted standard deviation. For the two-component Beta-t -EGARCH model, $h_{t|t-1} = \exp(\lambda_{t|t-1})\,\sigma_\varepsilon$, where $\sigma_\varepsilon = \sqrt{\nu/(\nu-2)}$.

The superior predictive ability of the Beta-t-EGARCH model is borne out by Table VI, which contains results for both forecasting horizons and six loss functions. For one-step-ahead predictions, the MSE_1 column shows that the root mean squared error of $\sqrt{0.51} \approx 0.71$ as produced by our model improves on the value of $\sqrt{0.70} \approx 0.84$ from the corresponding GARCH model. As the remaining five columns show, this outperformance is evident for all loss functions. It seems that whereas big swings in the market do not adversely affect the predictions of Beta-t -EGARCH, they appear to induce overprediction in the GARCH. This reflects the fact that GARCH models are sensitive to outliers and are unable to model the calming effect of positive returns. When the forecasting horizon increases to four steps ahead, the difference between the two models decreases, primarily because the short-run component is less influential. Nevertheless two-component Beta-t-EGARCH still outperforms the two-component GARCH model for five out of six loss functions.

5. CONCLUSION

The various contradictions and puzzles described in Section 1 are resolved by the DCS EGARCH-M model. The ease with which a dynamic two-component model for volatility can be estimated and interpreted plays a key role, as does the flexibility in the leverage term. In particular, it seems that positive returns can actually reduce short-term volatility. Thus whereas returns have a symmetric effect on volatility in the long run, the short-run response is sometimes close to being anti-symmetric. As regards the equity risk premium, our results for weekly data allow us to reject both a constant and a rapidly varying risk premium in favour of one that is associated with the long-run component of volatility. The effect of short-term volatility is to reduce returns, presumably because increased uncertainty drives away nervous investors, whereas a reduction in volatility has a calming effect. When volatility is modelled using a single component, these countervailing forces tend to cancel, which is perhaps why many previous studies have been inconclusive.

[6] We use the variable 'rv5' from https://realized.oxford-man.ox.ac.uk

DOI: 10.1111/jtsa.12419

The above conclusions are consistent with those of Adrian and Rosenberg (2008). However, our modifications to their EGARCH-M model are theoretically important and practically useful. By using the score of the conditional distribution to drive the dynamics, we are able to derive the statistical properties of the model and to develop an asymptotic theory for the maximum likelihood estimator when the conditional distribution is Student's t. A simple condition for invertibility is also given. The fact that the score is bounded has the advantage of curbing the impact of extreme observations.

Allowing the Student's t-distribution to be skewed is straightforward and seems to produce the best models, with the most plausible estimates of the risk premium. Further generalizations are possible. For example the generalized t-distribution may be used, as in the ARCH-M model of Theodossiou and Savva (2016), and this may be extended to allow for different degrees of freedom in the upper and lower tails, as in Harvey and Lange (2017). Adding an ARCH-M vector to a multivariate DCS volatility model of the kind proposed by Creal *et al.* (2013) is also possible, yielding a model which is much closer to the specification of Bekaert and Wu (2000) than that of Bollerslev *et al.* (1988).

ACKNOWLEDGEMENTS

When the original work was done, Rutger-Jan Lange was a Post-Doctoral Research Associate on the project Dynamic Models for Volatility and Heavy Tails at Cambridge University. We are grateful to the Keynes Fund for financial support. Earlier versions of this article were presented at the conference Recent Developments in Financial Econometrics and Empirical Finance at the University of Essex in June 2014 and at the Cambridge Finance-Tinbergen Institute meeting in Amsterdam in May 2014. We are grateful to Francisco Blasques, Peter Boswijk, Andre Lucas, Siem Jan Koopman, Mark Salmon, Robert Taylor and an anonymous referee for helpful comments.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

REFERENCES

Adrian T, Rosenberg J. 2008. Stock returns and volatility: pricing short-run and long-run components of market risk. *The Journal of Finance* **63**: 2997–3030.

Bekaert G, Wu G. 2000. Asymmetric volatility and risk in equity markets. *Review of Financial Studies* **13**: 1–42.

Bjørnland HC, Leitemo K. 2009. Identifying the interdependence between US monetary policy and the stock market. *Journal of Monetary Economics* **56**: 275–282.

Blasques F, Koopman SJ, Lucas A. 2014. *Maximum likelihood estimation for generalized autoregressive score models*. Amsterdam: Tinbergen Institute. Discussion Paper, TI 2014-029/III.

Blasques F, Gorgi P, Koopman SJ, Wintenberger O. 2018. Feasible invertibility conditions and maximum likelihood estimation for observation-driven models. *Electronic Journal of Statistics* **12**: 1019–1052.

Bollerslev T, Engle RF, Wooldridge JM. 1988. A capital asset pricing model with time-varying covariances. *Journal of Political Economy* **96**: 116–131.

Campbell JY, Hentschel L. 1992. No news is good news: an asymmetric model of changing volatility in stock returns. *Journal of Financial Economics* **31**: 281–318.

Chou RY. 1988. Volatility persistence and stock valuations: some empirical evidence using GARCH. *Journal of Applied Econometrics* **4**: 279–294.

Creal D, Koopman SJ, Lucas A. 2011. A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. *Journal of Business and Economic Statistics* **29**: 552–63.

Creal D, Koopman SJ, Lucas A. 2013. Generalized autoregressive score models with applications. *Journal of Applied Econometrics* **28**: 777–95.

Christensen BJ, Nielsen MO, Zhu J. 2010. Long memory in stock market volatility and the volatility-in-mean effect: the FIEGARCH-M Model. *Journal of Empirical Finance* **17**: 460–470.

Engle RF, Lee GGJ. 1999. *A long-run and short-run component model of stock return volatility* Engle RF, White H. (eds.) Oxford: Oxford University Press.

Engle RF, Rangel JG. 2008. The spline-GARCH model for low-frequency volatility and its global macroeconomic causes. *The Review of Financial Studies* **21**: 1187–1222.
Engle RF, Lilien DM, Robins RP. 1987. Estimating time-varying risk premia in the term structure: the ARCH-M model. *Econometrica* **55**: 391–407.
French KR, Schwert GW, Stambaugh RF. 1987. Expected stock returns and volatility. *Journal of Financial Economics* **19**: 3–29.
Glosten LR, Jagannathan R, Runkle DE. 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance* **48**: 1779–801.
Hansen PR, Lunde A. 2005. A forecast comparison of volatility models: does anything beat a GARCH (1, 1)? *Journal of Applied Econometrics* **20**: 873–889.
Harvey AC. 2013. *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series. Econometric Society Monograph*. Cambridge: Cambridge University Press.
Harvey AC, Sucarrat G. 2014. EGARCH models with fat tails, skewness and leverage. *Computational Statistics and Data Analysis* **26**: 320–338.
Harvey AC, Lange R-J. 2017. Volatility modelling with a generalized *t*-distribution. *Journal of Time Series Analysis* **38**: 175–90.
Hong EP. 1991. The autocorrelation structure for the GARCH-M process. *Economics Letters* **37**: 129–132.
Lanne M, Saikkonen P. 2006. Why is it so difficult to uncover the risk-return tradeoff in stock returns? *Economics Letters* **92**: 118–125.
Nelson DB. 1991. Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* **59**: 347–370.
Poon S, Taylor SJ. 1992. Stock returns and volatility: an empirical study of the UK stock market. *Journal of Banking and Finance* **16**: 37–59.
Schwert GW. 1989. Why does stock market volatility change over time? *The Journal of Finance* **44**: 1115–1153.
Scruggs JT. 1998. Resolving the puzzling intertemporal relation between market risk premium and conditional market variance: a two-factor approach. *The Journal of Finance* **53**: 575–603.
Taylor SJ. 2005. *Asset Price Dynamics, Volatility, and Prediction*: Princeton University Press.
Theodossiou P, Savva CS. 2016. Skewness and the relation between risk and return. *Management Science* **62**: 1598–1609.
Turner CM, Startz R, Nelson CR. 1989. A Markov model of heteroskedasticity, risk, and learning in the stock market. *Journal of Financial Economics* **25**: 3–22.

DOI: 10.1111/jtsa.12419