UNIVERSITY OF
CAMBRIDGE

Cambridge Working Papers in Economics

Modeling the Interactions between Volatility and Returns

Andrew Harvey and Rutger-Jan Lange

CWPE 1518

Modeling the Interactions between Volatility and Returns

Andrew Harvey and Rutger-Jan Lange
Faculty of Economics, Cambridge University

July 1, 2015

Abstract

Volatility of a stock may incur a risk premium, leading to a positive correlation between volatility and returns. On the other hand the leverage effect, whereby negative returns increase volatility, acts in the opposite direction. We propose a reformulation and extension of the ARCH in Mean model, in which the logarithm of scale is driven by the score of the conditional distribution. This EGARCH-M model is shown to be theoretically tractable as well as practically useful. By employing a two component extension we are able to distinguish between the long and short run effects of returns on volatility. The EGARCH formulation allows more flexibility in the asymmetry of the response (leverage) and this enables us to find that the short-term response is, in some cases, close to being anti-asymmetric. The long and short run volatility components are shown to have very different effects on returns, with the long-run component yielding the risk premium. A model in which the returns have a skewed t distribution is shown to fit well to daily and weekly data on some of the major stock market indices.

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

1 Introduction

Volatility of a stock may incur a risk premium. A popular textbook model for introducing a time-varying risk premium into a return is Autoregressive

Conditional Heteroscedasticity in Mean, or simply ARCH-M; see, for example, Taylor (2005, p 205, 252-4) and Mills and Markellos (2008, p 287-93). Recent papers on the topic include Baillie and Morana (2009), Lundblad (2007), Guo and Neely (2008) and Christiansen et al (2010, 2012); Koopman and Uspensky (2002) approach the problem in an unobserved components framework.

Current ARCH-M models lack a comprehensive asymptotic theory for the maximum likelihood (ML) estimator. Here we introduce an exponential formulation, with the dynamics driven by the score of the conditional distribution, and show that it is theoretically tractable as well as practically useful. Models constructed using the conditional score were introduced into the literature by Creal, Koopman and Lucas (2011, 2013), where they are called Generalized Autoregressive Score (GAS) models and Harvey (2013), where they are called Dynamic Conditional Score (DCS) models. The asymptotic distribution theory for the ML estimator for the proposed class of DCS EGARCH-M models can be obtained by extending the results in Harvey (2013) and Blasques et al (2014). This contrasts with standard formulations of ARCH-M models, where the development of a corresponding asymptotic theory has proved to be difficult. Other theoretical results, such as the existence of moments for a conditional Student t distribution, may also be derived. As regards the practical value of DCS models, there is already a good deal of evidence (in the references cited) to show that they tend to outperform standard models, one of the principal reasons being the way in which the dynamic equations deal with observations which, in a Gaussian world, would be regarded as outliers.

A model with two components of volatility has a number of attractions, one of which is to account for the long memory behavior often seen in the autocorrelations of absolute values of returns or their squares; see Alizadeh et al (2002). 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. Short-run volatility can lead to a 'news effect', as described by Chou (1988) and Schwert (1989), that makes investors nervous of risk and so predicts a negative correlation between volatility and return. This negative relationship contrasts with the positive relationship between long-run volatility and return predicted by Merton's (1973) intertemporal capital asset pricing model (ICAPM); see French et al (1987). Failure to model both aspects of volatility has led to inconclusive results regarding the sign of the risk premium. For example, the risk premium is negative and significant according to Campbell (1987) and 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). Indeed Lundblad (2007) argues that, when existing GARCH-M models are used, even a hundred years of data may not be sufficient to draw clear inferences. Here we show that a carefully specified two-component model can resolve these apparent contradictions because it enables the researcher to investigate the possibility that when long-run volatility goes up it tends to be followed by an increasing level of returns, whereas an increase in short-run volatility leads to a fall.

Returns may have an asymmetric effect on volatility. For example, considerations of leverage suggests that negative returns are associated with increased volatility; see Bekeart and Wu (2000) for an excellent discussion of the finance issues involved. Indeed the term leverage is often loosely used to indicate any kind of asymmetry in the response of volatility to returns. Figure 1, which highlights some of the relationships between volatility and returns, shows why disentangling the various effects can be so hard. Again a two component model plays an important role in that it allows us to determine whether asymmetric effects are different in the short-run and the long-run. For example it may be that an asymmetric response is confined to the short-run volatility component, as found by Engle and Lee (1999) and others. Indeed short-run volatility may even decrease after a good day, because of the calming effect this has on the market[1]. Standard GARCH models are unable to identify this possibility because they are constrained to model leverage with an indicator variable rather than the sign variable which can be adopted in EGARCH.

Section 2 sets out the DCS formulation of the basic EGARCH-M model with a conditional t-distribution and establishes the conditions for the existence of moments, explores the patterns of autocorrelation functions and discusses predictions. The asymptotic distribution theory is outlined and this is followed by sub-sections describing the extension the leverage effects, two components and skewed distributions. In Section 3 various models are fitted to weekly and daily data on NASDAQ and NIKKEI returns. Section 4 sketches out extensions to multivariate time series and the inclusion of the risk-free rate as an explanatory variable, as in Scruggs (1998). In the latter

[1] Chen and Ghysels (2008) have recently identified the presence of such effects in diurnal data.

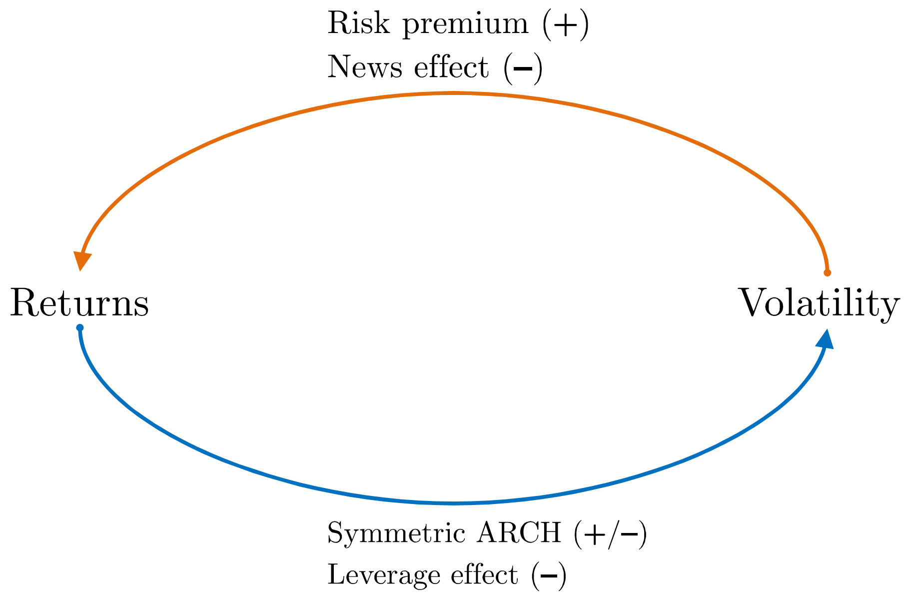

Figure 1: Interactions between returns and volatility.

case, we present results for fitting models to weekly excess returns on S&P 500 over a 60-year period. Section 5 concludes by summarizing the extent to which our econometric modeling has succeeded in disentangling the various interactions between returns and volatility.

2 Model Formulation

In the ARCH-M model, as introduced by Engle et al (1987), the model for a series of returns subject to a time-varying risk premium is

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

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, $\sigma^2_{t|t-1}$, is assumed to be generated by a dynamic equation that is dependent on squared observations, so in the GARCH(1,1) case

$$\sigma^2_{t+1|t} = \gamma + \beta\sigma^2_{t|t-1} + \delta y_t^2, \tag{2}$$

where γ, β and δ are parameters. If the mean, μ, can be dropped, the ARCH-M effects are estimated more precisely; see Lanne and Saikkonen (2006).

Engle et al (1987), and most subsequent studies, find the standard deviation, that is $m = 1$, gives a better fit than variance (although the ICAPM theory suggests the latter). This is fortunate because it turns out that the asymptotic theory for ML estimators of DCS EGARCH-M models is most easily developed with the standard deviation or, more generally, the scale. Thus the DCS EGARCH-M is set up as

$$y_t = \mu + \alpha \exp(\lambda_{t|t-1}) + \varepsilon_t \exp(\lambda_{t|t-1}), \quad t = 1, ..., T, \tag{3}$$

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 of the logarithm of the probability density function at time t. The stationary first-order dynamic model for $\lambda_{t|t-1}$ is

$$\lambda_{t+1|t} = \omega(1 - \phi) + \phi\lambda_{t|t-1} + \kappa u_t, \quad |\phi| < 1, \tag{4}$$

where u_t is the conditional score and ϕ, κ and ω are parameters, with ω denoting the unconditional mean, and $\lambda_{1|0} = \omega$. The score with respect to $\lambda_{t|t-1}$ for a conditional t-distribution with ν degrees of freedom and scale $\exp(\lambda_{t|t-1})$ is

$$u_t = (\nu + 1)b_t - 1 + \alpha(1 - b_t)[(\nu + 1)/\nu]\varepsilon_t, \tag{5}$$

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

$$b_t = \frac{\varepsilon_t^2/\nu}{1 + \varepsilon_t^2/\nu}, \qquad 0 \leq b_t \leq 1, \quad 0 < \nu < \infty, \tag{6}$$

is distributed as $beta(1/2, \nu/2)$ at the true parameter values. In the absence of the ARCH-M effect, 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. Letting $\nu \to \infty$ yields the normal distribution, that is ε_t is $NID(0, 1)$, in which case the scale is the standard deviation and the score is simply

$$u_t = \varepsilon_t^2 - 1 + \alpha\varepsilon_t, \quad t = 1, ..., T. \tag{7}$$

The last term in (5) 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.

Remark 1 *Note that the ARCH-M score term in (7), that is ε_t, is not a leverage term, despite the resemblance to the leverage term in the classic EGARCH model of Nelson (1991).*

2.1 Moments, autocorrelations and predictions

One of the important features of the Beta-t-EGARCH model, is that when $\lambda_{t|t-1}$ is stationary, the unconditional moments of the observations exist whenever the corresponding conditional moment exists[2]. This is not generally true for GARCH models where issues surrounding the existence of unconditional moments can be quite complex. Furthermore exact analytic expressions for moments and autocorrelations of absolute values of the observations raised to any non-negative power may be derived; see Harvey (2013, ch 4). For the Beta-t-EGARCH-M model the inclusion of the ARCH-M score term, $\alpha(1-b_t)[(\nu+1)/\nu]\varepsilon_t$, in (5) makes it difficult to derive corresponding analytic expressions. However, the boundedness of the score, and hence $\exp(m\lambda_{t|t-1})$, for the t-distribution means that the unconditional $m-th$ moment still exists whenever the corresponding conditional $m-th$ moment does.

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

$$\rho(\tau)=\frac{E(\exp(\lambda_{t|t-1}+\lambda_{t-\tau|t-\tau-1}))-[E(\exp\lambda_{t|t-1})]^2}{(1+\sigma_\varepsilon^2/\alpha^2)E(\exp 2\lambda_{t|t-1})-[E(\exp\lambda_{t|t-1})]^2},\qquad \tau=1,2,3,..., \tag{8}$$

when the distribution of the ε_t's is symmetric; see Appendix A. These autocorrelations do not depend on ω. Clearly $\rho(\tau)=0$ for $\alpha=0$. When $\alpha\neq 0$, $\rho(\tau)$ has the same sign[3] as the corresponding autocorrelation of the volatility,

[2] In Nelson's classic EGARCH model the dynamics are driven by absolute values of returns; see, for example, Christensen et al (2010). If the conditional distribution is Student's t, the returns have no moments of any order.

[3] In the GARCH-M model, the autocorrelations of the $y_t's$ must all be positive because of the positivity constraints on the GARCH parameters; see Hong(1991).

$\exp \lambda_{t|t-1}$, with a pattern derived from that of the $\lambda'_{t|t-1}s$. As already noted, analytic expressions can be found for terms like $E(\exp \lambda_{t|t-1})$ if the ARCH-M score term is dropped from the dynamic equation.

Expressions for conditional moments of future observations can be readily obtained when the ARCH-M score term is not included. But the full predictive distribution is often what is needed and simulating it is not difficult because, even with the inclusion of the ARCH-M score term, it depends only on independent Student-t variates.

2.2 Large sample distribution of the maximum likelihood estimator

The asymptotic distribution of the ML estimator for the Beta-t-EGARCH model is relatively straightforward to obtain; see Harvey (2013, ch 4). This tractability is retained when the ARCH-M term is included as in (5). In contrast, no asymptotic theory is available for the standard ARCH-M model based on the GARCH equation, (2). To see why the DCS model is more straightforward, we simplify by focussing on the Gaussian model with known μ. In this case, the information matrix (for a single observation) for the dynamic parameters, $\boldsymbol{\psi} = (\kappa,\, \phi,\, \omega)'$, and the ARCH-M parameter, α, can be approximated by

$$\mathbf{I}\begin{bmatrix} \boldsymbol{\psi} \\ \alpha \end{bmatrix} \simeq \begin{bmatrix} (2+\alpha^2)\mathbf{D} & \alpha\mathbf{d} \\ \alpha\mathbf{d}' & 1 \end{bmatrix}, \tag{9}$$

where $\mathbf{D}$ is defined in Harvey (2013, p 37) and $\mathbf{d} = (0, 0, (1-\phi)/(1-\phi+2\kappa+\alpha^2))$. The exact information matrix is given in Appendix B, but it turns out that the additional terms are negligible in practice. The model was simulated 10,000 times with typical values of the parameters and ML estimation carried out. Table 1 compares the resulting RMSEs with those obtained by inverting the information matrix in (9) above. (Note that this matrix does not depend on ω). As can be seen, the simulated RMSEs are very close to the asymptotic standard errors (ASEs) for $T = 10,000$. They are similarly close for $T = 1000$, except for ϕ. However, this may not be too surprising because the value of $\phi = 0.98$ is quite close to unity, and the estimates of ϕ were constrained to be less than one. When the exact ASEs are computed using the formulae in the appendix the only difference is that for α they are (to the four decimal places computed) exactly equal to the RMSEs.

	True parameter	$T=10,000$			$T=1000$		
		Mean	RMSE	ASE	Mean	RMSE	ASE
κ	0.04	0.0399	0.0024	0.0023	0.0397	0.0077	0.0074
ϕ	0.98	0.9794	0.0030	0.0028	0.9734	0.0155	0.0090
ω	0.10	0.0984	0.0350	0.0354	0.0882	0.1123	0.1119
α	0.05	0.0499	0.0099	0.0100	0.0498	0.0314	0.0316

Table 1: Monte Carlo results for Gaussian EGARCH-M model

When the conditional distribution is Student's t, the proof of consistency and asymptotic normality is an extension of the one outlined in Harvey (2013, ch 2), which is based on the main theorem in Jensen and Rahbek (2004); the boundedness of the score plays a key role.

Remark 2 *Using the variance as the risk premium variable for the Gaussian model, that is $m = 2$ in (1), gives an ARCH-M score term of $\alpha\varepsilon_t \exp(\lambda_{t|t-1})$. More generally, for any $m > 0$, $u_t = \varepsilon_t^2 - 1 + \alpha m \varepsilon_t \exp(\lambda_{t|t-1}(m-1))$. The presence of $\exp(\lambda_{t|t-1}(m-1))$ in the score for $m \neq 1$ leads to a less tractable asymptotic theory (because the scores are no longer IID). This remains true when the ARCH-M term is $\ln \sigma_{t|t-1}$, as suggested by Engle et al (1987) for some series.*

2.3 Asymmetric impact curves (leverage)

The leverage effect in finance suggests that volatility rises when the asset price falls. This asymmetry can be captured by an EGARCH model by modifying the dynamic equation in (4) to

$$\lambda_{t+1|t} = \omega\,(1-\phi) + \phi\,\lambda_{t|t-1} + \kappa\,u_t + \kappa^* u_t^*, \tag{10}$$

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 expected to be positive; see Harvey (2013, p109). When the distribution of ε_t is symmetric, u_t^* has zero mean and is orthogonal to u_t, in that $E(u_t u_t^*) = 0$.

The rise in volatility following a fall in the asset price need not necessarily be due to leverage as such. For example the label 'news impact curve' is often used instead of leverage, reflecting the idea that a sharp fall in asset price may induce more uncertainty and hence higher variability. Figure 2 shows the response, that is $\kappa\,u_t + \kappa^* u_t^*$, for the Beta-t-EGARCH model with

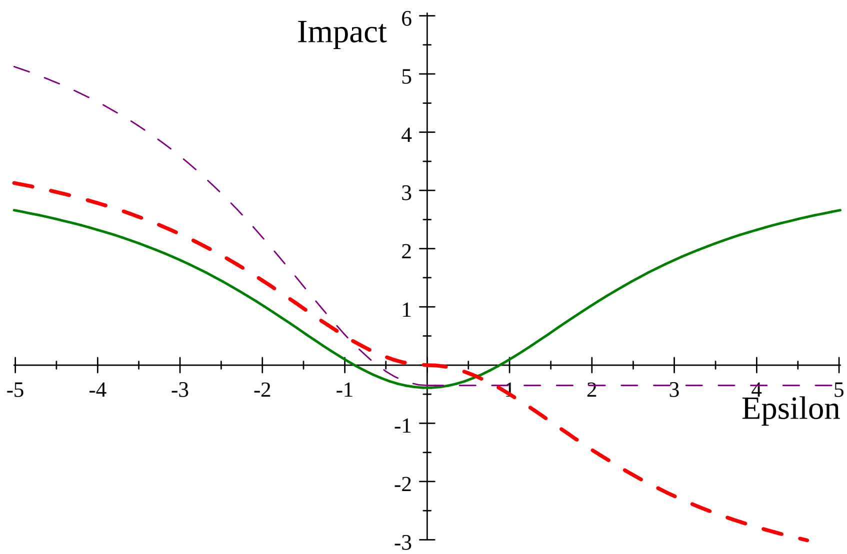

Figure 2: Impact curves for t_9. Solid is symmetric with $\kappa^* = 0$, thin dash is $\kappa = \kappa^*$ and thick dash is anti-symmetric with $\kappa = 0$.

a conditional t_9 distribution. The curves, which are plotted against standardized observations, ε_t, range from the symmetric, in which $\kappa^* = 0$, to the anti-symmetric in which $\kappa = 0$. (The non-zero values of κ and κ^* are different for each curve so as to keep them separate.)

The standard way of incorporating leverage effects into GARCH models is by including a variable in which the squared observations are multiplied by an indicator, $I(y_t < 0)$, taking the value one for $y_t < 0$ and zero otherwise. The sign cannot be used because negative values could give a negative conditional variance. Glosten, Jagannathan and Runkle (1993, p 1788), who introduced this way of modeling leverage, acknowledge that being unable to allow for the possibility of a negative effect on volatility is a weakness and they note that the problem does not arise with EGARCH. Scruggs (1998, p 582) uses EGARCH precisely for this reason. However, the drawback with classic EGARCH is that $\lambda_{t+1|t}$ responds to $|\varepsilon_t|$ and ε_t, rather than to the scores, as in (10). The linear response means that $\lambda_{t|t-1}$ is sensitive to outliers and y_t has no unconditional moments for a conditional t-distribution.

The information matrix for a Beta-t-EGARCH model with dynamics as in (10) is as in Harvey (2013, pp. 121-4) and it may be extended to incorporate

ARCH-M effects. Identifiability requires only that either κ or κ^* be non-zero. Thus Wald and LR tests of the null hypothesis that *either* κ *or* κ^* is zero can be carried out.

2.4 Two components

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

$$\begin{aligned} \lambda_{t|t-1} &= \omega + \lambda_{1,t|t-1} + \lambda_{2,t|t-1}, & t = 1, ..., T, \\ \lambda_{i,t+1|t} &= \phi_i \, \lambda_{i,t|t-1} + \kappa_i \, u_t + \kappa_i^* \, u_t^*, & i = 1, 2, \end{aligned} \tag{11}$$

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 $\kappa_1 \neq 0$ or $\kappa_1^* \neq 0$ and $\kappa_2 \neq 0$ or $\kappa_2^* \neq 0$.

The two component model allows the leverage effect to impact differently in the long run and short run; see also Engle and Lee (1999). It also makes it possible to separate out the effects of long-run and short-run movements in volatility on the mean. Thus equation (3) 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 risk premium is then captured by the long-run component. When volatility is at its equilibrium level, that is $\lambda_{1,t|t-1}$ and $\lambda_{2,t|t-1}$, are equal to zero, the risk premium is $\mu' + \alpha_1 \exp(\omega)$. (If ϕ_1 is set to one and $\omega = 0$, $\lambda_{1,t+1|t}$ is the long-run forecast of volatility and $\alpha_1 \exp(\lambda_{1,t+1|t})$ is the corresponding forecast of the risk premium.)

2.5 Skew t distribution

Skewness can be introduced into a distribution by means of the Fernandez and Steel (1998) method. Harvey and Sucarrat (2014) use this method to give the skew Student t distribution

$$f(y) = \frac{2}{\gamma + \gamma^{-1}} \, \frac{1}{\varphi \, \nu^{1/2}} \frac{1}{B(\nu/2, 1/2)} \left(1 + \frac{1}{\nu} \frac{(y-\mu)^2}{\gamma^{2 \, \mathrm{sgn}(y-\mu)} \varphi^2} \right)^{-(\frac{\nu}{2} + \frac{1}{2})}, \quad -\infty < y < \infty,$$

where $0 < \gamma < \infty$ and $\gamma = 1$ gives symmetry. The score in the Skew-t-EGARCH model is

$$u_t = (\nu + 1)\frac{(y-\mu)^2 e^{-2\lambda_{t|t-1}}/\nu}{(y-\mu)^2 e^{-2\lambda_{t|t-1}}/\nu + \gamma^{2\,\mathrm{sgn}(y_t-\mu)}} - 1. \tag{13}$$

The introduction of skewness means that the conditional expectation of ε_t is no longer zero[4]. As a result, the ARCH-M effect is confounded with the conditional expectation of $\varepsilon_t \exp(\lambda_{t|t-1})$. The mean of a skewed distribution is $\mu_\varepsilon = (\gamma - \gamma^{-1})E(|\varepsilon|)$, where $E(|\varepsilon|)$ refers to the original (non-skewed) variable, and so the conditional expectation of y_t in (3) is $\mu + (\alpha + \mu_\varepsilon)\exp(\lambda_{t|t-1})$. Thus the ARCH-M effect is not given by the estimate of α in (3) but by $\alpha^\dagger = \alpha + \mu_\varepsilon$. When γ is less than one, which is typically the case, μ_ε is negative and so the ARCH-M effect is reduced. A more convenient model formulation is obtained by subtracting μ_ε from the disturbance term so that

$$y_t = \mu + \alpha^\dagger \exp\lambda_{t|t-1} + (\varepsilon_t - \mu_\varepsilon)\exp\lambda_{t|t-1}, \quad t = 1, ..., T,$$

where $\alpha^\dagger = \alpha + \mu_\varepsilon$. For the skewed Student-t distribution

$$\mu_\varepsilon = \left(\gamma - \frac{1}{\gamma}\right)\frac{\Gamma(\nu/2 - 1/2)}{\Gamma(1/2)\Gamma(\nu/2)}\nu^{1/2}.$$

For the two-component model, a modification of (12) yields

$$y_t = \mu' + \alpha_1 \exp(\omega + \lambda_{1,t|t-1}) + \alpha_2[\exp(\lambda_{2,t|t-1}) - 1] + (\varepsilon_t - \mu_\varepsilon)\exp(\lambda_{t|t-1}), \tag{14}$$

where α_1 plays the same role as $\alpha^\dagger$ in the earlier expression.

3 Results

The benefits of using the DCS EGARCH-M are best illustrated with weekly data. The results for daily and monthly observations, while slightly less clear-cut, are consistent with the weekly findings.

The return in period t is defined as the continuously compounded percentage $100 * [\log(I_t) - \log(I_{t-1})]$, where I_t is the adjusted closing price of

[4]A similar problem was noted in Harvey and Sucarrat (2014), where a correction for skewness was made in order for the returns to be martingale differences. There the model was re-formulated. Here it is re-arranged.

the index. The data are drawn from Yahoo Finance (yahoo.finance.com). 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), Poon and Taylor (1992) and Koopman and Uspensky (2002).

The excess return, y_t, is defined as the return minus the risk-free return, that is $y_t = 100 * [\log(I_t) - \log(I_{t-1})] - r_{f,t}$. The risk-free return is proxied by the secondary market for 3-month Treasury bills[5]. The risk-free return for weekly data is defined as the continuously compounded interest rate per week.

Estimation of all models was carried out by ML, with standard errors estimated numerically. Residual serial correlation in location and scale was assessed by constructing Box-Ljung statistics, $Q(20)$, from the first 20 residual autocorrelations. Serial correlation in location was measured by autocorrelations in the residuals and the scores with respect to location, while serial correlation in scale was measured by autocorrelations in the squared residuals and the scores with respect to scale; see Harvey (2013, ch 2). With a conditional t-distribution the score residuals are less affected by outliers than the raw residuals. Outliers usually weaken the serial correlation, though it is sometimes found that autocorrelations from raw residuals are bigger when there are two consecutive outliers. Either way test statistics from the raw residuals can be misleading about underlying residual autocorrelation and they are not quoted after Table 2. As they stand, the (score-based) Q-statistics are only a rough guide to residual serial correlation as they are not, in general, asymptotically χ^2_{20}. Further work on diagnostics for this kind of situation in DCS models is currently under way.

3.1 Weekly data

Table 2 shows the results of fitting first-order Beta-t-EGARCH and GARCH-t to weekly NASDAQ excess returns from 8 Feb 1972 to 3 Nov 2014 (2,282 observations). We note, for future reference, that the mean of $r_{f,t}$ is 0.09 and that the average excess return, y_t, is 0.07. On the other hand, the median is 0.24. Leverage effects are significant for both models and the Beta-t-EGARCH model gives a slightly better fit than the GARCH$-t$ model, that

[5] This data is available from Table H.15 of the Federal Reserve (http://www.federalreserve.gov/releases/h15/data.htm), from 4 January 1954 onwards. The table cites yearly interest rates so these had to be adjusted for the relevant frequency.

is (2) with the additional term $\delta^* y_t^2 I(y_t < 0)$; the parameter ω in the tables is obtained by setting $\gamma = \omega(\beta + \delta + \delta^*/2)$. The first-order Beta-t-EGARCH model with leverage, (10) outperforms the GARCH model in terms of fit. However, there is strong evidence of serial correlation in location and in neither model is the ARCH-M coefficient, α, statistically significant (though it is four times as big in the EGARCH model).

The first row of Table 3 shows the ML estimates from fitting a Beta–t-EGARCH two-component model with leverage in both components. The long-term news impact curve is close to being 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. Thus $\exp(\lambda_2)$ can be viewed as a measure of both good and bad sentiment: it is high after bad weeks, and low after good weeks. The EGARCH formulation allows the calming effect of positive returns to be identified. The asymmetric function of the score used to capture leverage in equation (10) is uncorrelated with the score itself. This orthogonality helps to distinguish the two components of volatility when one is symmetric and the other is anti-symmetric. Figure 3 shows the net result: the initial effect of a positive shock is a lowering of total volatility. Figure 4 shows the dynamic responses of $\exp \lambda_1$ and $\exp \lambda_2$ to a large positive shock to standardized returns. The individual responses are in different directions and the response of their sum, like that of λ_2, is initially negative. After about four weeks, the more persistent λ_1 begins to dominate in the total ARCH-M variable, $\exp(\lambda_1 + \lambda_2)$.

The next three rows show what happens when there is only one ARCH-M term, that is either λ_1 or λ_2, or their aggregate. The important point to note is that α_1 is positive, whereas α_2 is negative. The estimate of α, obtained from the sum of λ_1 and λ_2, is negative, so it cannot be estimating the risk premium. This model also displays strong autocorrelation in the residuals and location scores, that is ε_t and $u_{\mu,t}$; there is no evidence for such residual autocorrelation when a short-run ARCH-M term enters separately.

The model in the fifth row, where both short-run and long-run ARCH-M terms are included in their own right, resolves most of the difficulties. The long-run ARCH-M coefficient, α_1, is positive and the short-run coefficient, α_2, is negative. Autocorrelation is no longer present in any form. Since $\alpha_2 < 0$, good sentiment, as reflected by $\exp(\lambda_2) < 1$, tends to lead to further gains, whereas bad sentiment, $\exp(\lambda_2) > 1$, tends to induce further losses. Because $\phi_2 \simeq 0.75$, these effects are short term, but the 'momentum' may persist for several weeks.

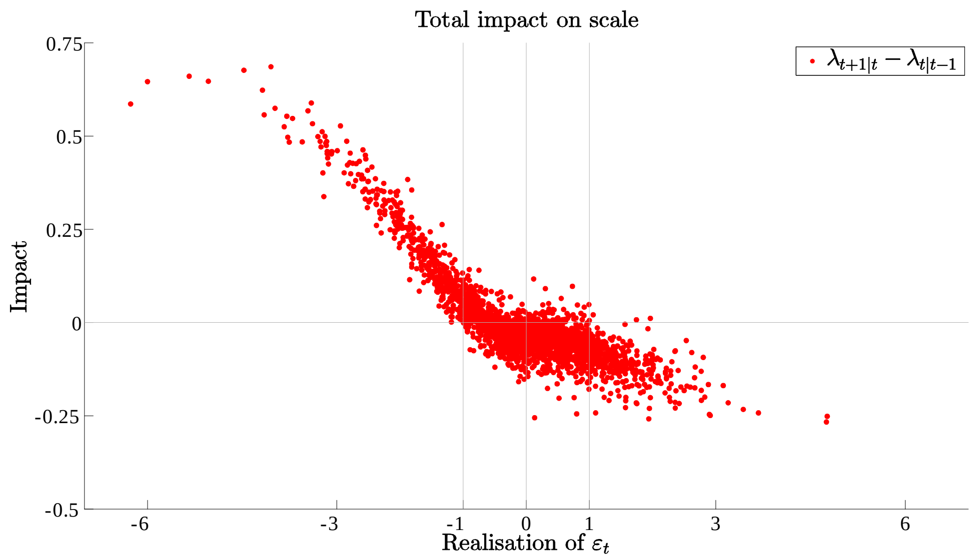

Figure 3: Initial impact on total volatility of each NASDAQ observation plotted against the standardized return.

The estimates indicate that $\mu \simeq -\alpha_2$. The sixth line shows the results with the constraint $\mu = -\alpha_2$ imposed. This constrained model can be conveniently estimated by re-writing the equation for returns as in (12) with the new mean, μ', set equal to zero. The estimated value of α_1 is similar to what it was before, but it is now statistically significant. There is no indication of serial correlation and, according to both AIC and BIC, the fit is better than for other specifications in the table. Figure 5 shows the long and short run components of volatility. The plot of the long-run component, $\alpha_1 \exp(\omega + \lambda_{1,t|t-1})$, contains the constant term and it is this quantity which measures the time-varying risk premium.

The fact that κ_1^* and κ_2 are small and statistically insignificant in most models suggests setting them to zero. When this is done, the remaining parameter estimates are virtually unchanged and the resulting BIC is often better. The entries below the line in Table 3 show the results for the models in lines 1, 5 and 6.

Remark 3 *For purposes of comparison, a two-component GARCH model as in Engle and Lee was estimated. For each of the estimated models in Table 3, the fit is worse than for the corresponding Beta-t-EGARCH model above.*

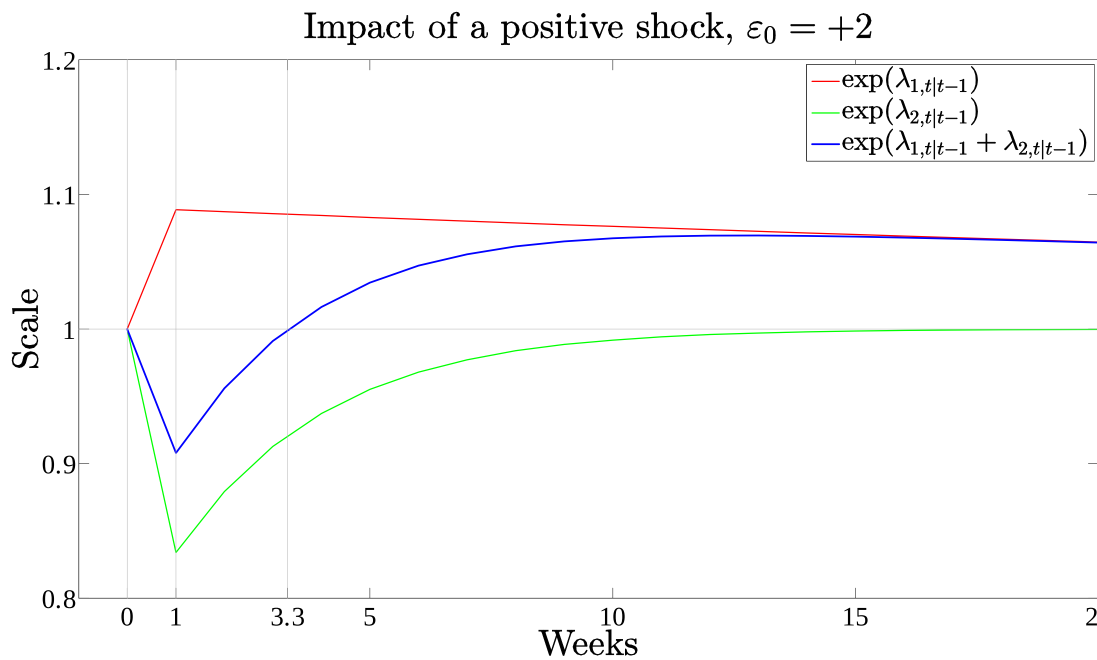

Figure 4: Response of long-run and short-run volatility to a large positive value for standardized returns.

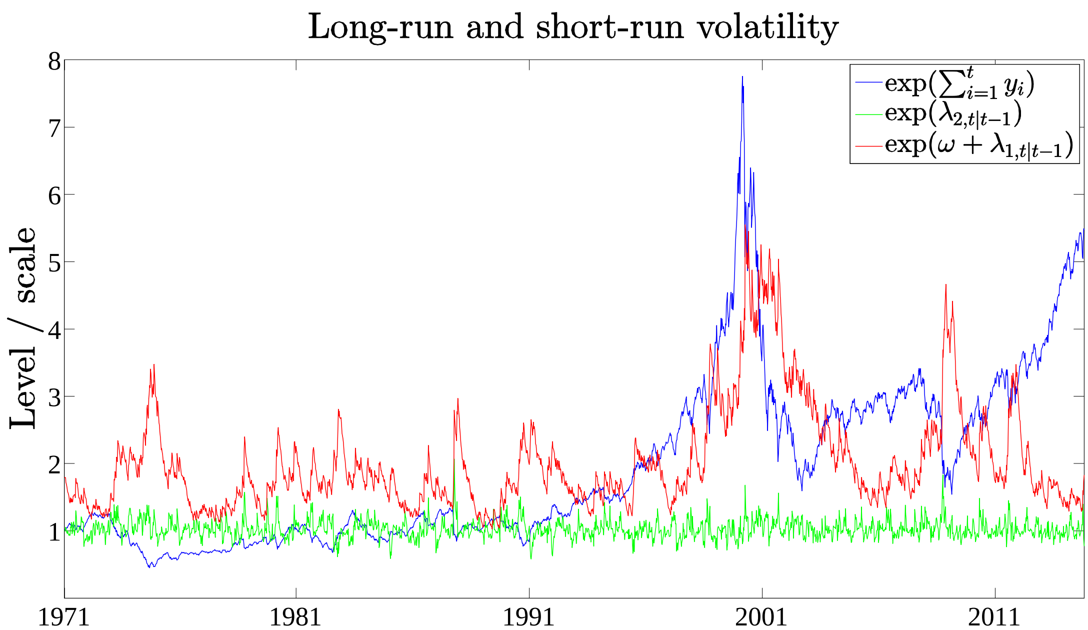

Figure 5: Long and short run volatilities for NASDAQ, together with index constructed from excess returns.

It is reassuring, however, that the findings are qualitatively similar. As in the Beta-t-EGARCH case, we find that the long-term news impact function is entirely symmetric (here, the constraint $\kappa_1^* \geq 0$ *is always binding). Again, the short-term news impact function is driven almost entirely by leverage (*κ_2 *is never significant, and the constraint* $\kappa_2 \geq 0$ *is often binding). An important difference, however, is that the short-term news impact function in the GARCH model is driven by the indicator function, rather than by the sign as in the EGARCH formulation. This is necessary for the variance to remain positive. However, since the resulting news impact function is zero for positive returns, GARCH models cannot find the calming effect of positive returns. Indeed, even in the GARCH model,* κ_2 *is estimated to be negative if the restriction* $\kappa_2, \kappa_2^* \geq 0$ *is not enforced; thereby mimicking the calming effect we find in the EGARCH formulation. The restriction must generally be enforced, however, if the variance is to remain positive. The interpretation of* $\exp(\lambda_2)$ *as sentiment is still possible, except it now refers to 'bad sentiment' only, as it is not affected by positive returns (at least not directly). As in the Beta-t-EGARCH model, we find strong evidence for volatility feedback in that* α_2 *is significantly negative, such that bad sentiment tends to lead to further losses (at least in the short term). Again,* α_1 *is estimated to be positive and significant.*

The fit of all the models in Table 3 is improved if the t-distribution is skewed. In all cases the estimate of the skewing parameter, γ, was close to 0.80 and the evidence against symmetry, that is $\gamma = 1$, is overwhelming. Figure 6 shows the histogram of residuals from fitting a (two-component) skew-t model, together with the corresponding density function and the one obtained when a symmetric $t-$distribution is used. Table 4 shows the results for the pure volatility model and the ARCH-M models reported in the last two rows, that is those where the long and short-run volatilities enter separately. Again the preferred specification has $\mu = -\alpha_2$, that is (14) with $\mu' = 0$. In this case, the weekly equilibrium risk premium estimate is $E(y_t) \simeq \alpha_1 \exp(\omega) = 0.051 \times \exp(0.676) = 0.10$. By continuous compounding, the average yearly risk premium is $\exp(0.10/100 \times 52) = 1.054$, that is 5.4% per year. This figure is bigger than the raw mean but much smaller than the median. The reason is that the raw mean is affected by one or two extreme negative values, corresponding to crashes, whereas in the fitted model the effect of these outliers is reduced by the (soft) trimming implied by the location score. Without skewness the weekly equilibrium risk premium,

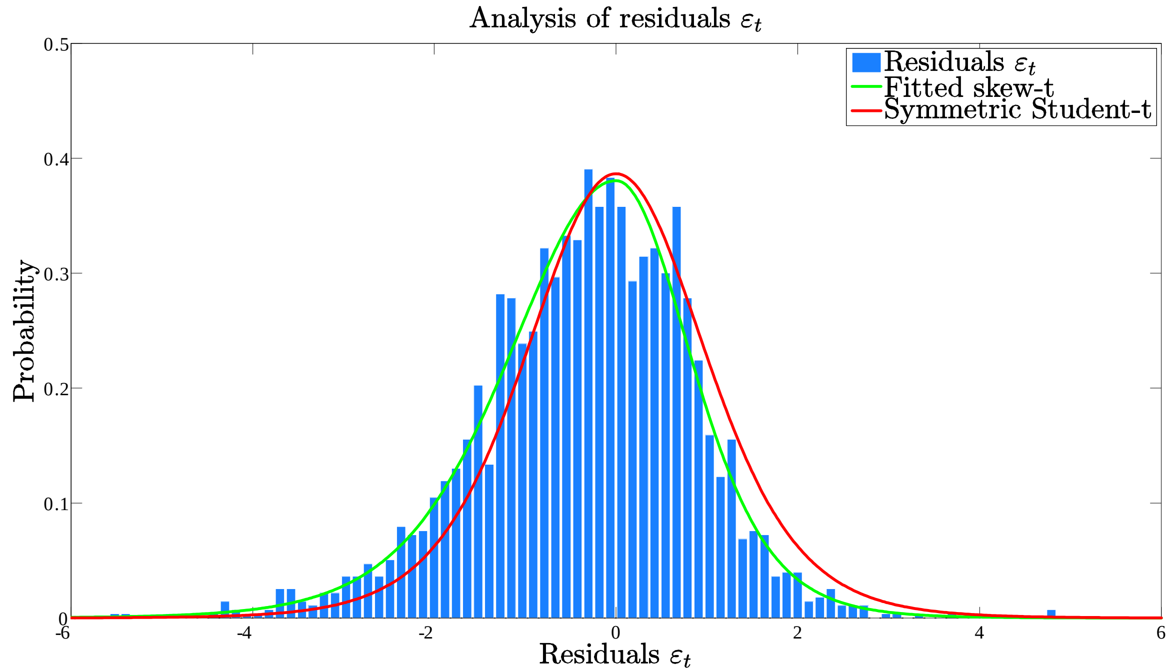

Figure 6: Histogram of standardized residuals from skewed t models together with density functions of fitted skewed and symmetric t-distributions.

as given by the last line of Table 3, is 0.245 which corresponds to a much higher annual risk premium of 13.6%.

We now examine a different market to see if the above findings apply more generally. Tables 5 and 6 show results for NIKKEI weekly returns from 4 Jan 1984 to 20 Oct 2014 (1,595 observations). Here the average excess return, y_t, is 0.027 and the median is 0.243. As with NASDAQ, the long-term news impact curve is essentially symmetric, whereas the short-run curve is close to being anti-symmetric. The short-term ARCH-M coefficient, α_2, is significantly negative and $\mu \approx -\alpha_2$. Residual autocorrelation is less of an issue than for NASDAQ, but again the short-term ARCH-M term is instrumental in reducing serial correlation in the location scores. The long-run ARCH-M coefficient, α_1, is positive and significant at conventional levels when μ is set equal to $-\alpha_2$. It is smaller than the corresponding NASDAQ coefficient, but the simple average of Nikkei returns is much smaller than the NASDAQ average.

The skewness parameter in Table 6, like the degrees of freedom, is very similar to the corresponding parameter for NASDAQ. Here, it is assumed that the long-term (short-term) news-impact curves are fully symmetric (asymmetric). The estimate of α_1, which requires $\mu = -\alpha_2$ to have the right sign, is now rather small and in fact the fit of the model is virtually identical to the model with only μ and α_2. The weekly equilibrium risk premium estimate is 0.057, that is $0.025 \times \exp(0.832)$, which implies an average return of 3% per year. This number actually represents average gross returns, rather than average excess return, because for the NIKKEI no risk-free rate was subtracted.

3.2 Daily data

Table 7 reports results obtained by fitting the preferred two-component EGARCH and EGARCH-M models to daily NASDAQ observations from the start of 1990. The results tell much the same story as weekly returns. In particular, the long-run news impact curve appears to be symmetric whereas the short-run curve is anti-symmetric. The model with a skew-t distribution and the constraint $\mu = -\alpha_2$ fits best and the estimates of α_1 and α_2 are plausible. Thus high short-term volatility may be associated with a fall in returns rather than an increase because of the risk premium. As regards ARCH effects, there is evidence of a calming effect of positive returns on short-run volatility. Note that the estimates of the autoregressive parameters are consistent with those in Table 4. For the weekly data ϕ_1 is further from unity and ϕ_2 is much smaller.

4 Extensions

4.1 Multivariate EGARCH-M models

A multivariate extension of the ARCH-M model was investigated by Bollerslev, Engle and Wooldridge (1988). However, the specification in Bekeart and Wu(2000) is much closer to the generalization proposed below.

The multivariate Beta-t-EGARCH model for an $N \times 1$ vector of returns, $\mathbf{y}_t$, has mean $\boldsymbol{\mu}$ and a conditional scale matrix specified as $\mathbf{\Omega}_{t|t-1} = \mathbf{D}_{t|t-1}\mathbf{R}_{t|t-1}\mathbf{D}_{t|t-1}$, where $\mathbf{D}_{t|t-1}$ is a diagonal matrix containing the time-varying scales for each series and $\mathbf{R}_{t|t-1}$ is a positive definite matrix with diagonal elements equal to unity; see Creal *et al* (2011, p. 557) and Harvey (2013, chapter 7). In a Gaussian model $\mathbf{R}_{t|t-1}$ is the correlation matrix. The logarithms of the diagonal elements of $\mathbf{D}_{t|t-1}$ are given by $\boldsymbol{\omega}+\boldsymbol{\lambda}_{t|t-1}$, where $\boldsymbol{\omega}$ is an $N \times 1$ vector of constants and the dynamic equations are

$$\boldsymbol{\lambda}_{t+1|t}= \boldsymbol{\Phi}\boldsymbol{\lambda}_{t|t-1}+\mathbf{K}\mathbf{u}_t, \qquad t=1,...,T, \tag{15}$$

where $\mathbf{u}_t$ is the score with respect to $\boldsymbol{\lambda}_{t|t-1}$ and $\boldsymbol{\Phi}$ and $\mathbf{K}$ are matrices of parameters. When the conditional scores are standardized by pre-multiplying by the inverse of the information matrix, the interpretation is more straightforward. For example, restricting $\boldsymbol{\Phi}$ and $\mathbf{K}$ to be diagonal leads to simple univariate filters. Leverage effects may be introduced into (15) by adding the term $\mathbf{K}^*\mathbf{u}_t^*$, where the $i-th$ element in the $N \times 1$ vector $\mathbf{u}_t^*$ is $sgn(-\varepsilon_{it})u_{it}$, $i=1,..,N$.

The multivariate EGARCH in mean model has

$$\mathbf{y}_t= \boldsymbol{\mu}+\mathbf{D}_{t|t-1}\boldsymbol{\alpha}+\mathbf{D}_{t|t-1}\boldsymbol{\varepsilon}_t, \quad t=1,...,T, \tag{16}$$

where $\boldsymbol{\alpha}$ is an $N \times 1$ vector. The extension to two components in $\boldsymbol{\lambda}_{t|t-1}$, and hence two ARCH-M components, is straightforward.

4.2 Risk-free rate

There is ample evidence in the literature of a negative correlation between excess stock returns and the risk-free rate. A variety of explanations have been proposed for what, at first sight, seems to be an anomaly. Early papers include Nelson (1976) and Fama and Schwert (1977). More recent contributions include those by Scruggs (1998) and Bjornland and Leitemo (2009).

The basic DCS EGARCH-M model incorporating the current risk-free rate, $r_{f,t}$, is as follows:

$$y_t=\mu+\alpha\exp(\lambda_{t|t-1})+\delta r_{f,t}+\varepsilon_t\exp(\lambda_{t|t-1}), \quad t=1,...,T, \tag{17}$$

with

$$\lambda_{t+1|t}=\omega(1-\phi)+\phi\lambda_{t|t-1}+\kappa u_t+\kappa^* u_t^*+\zeta r_{f,t+1}. \tag{18}$$

The model may be extended to include two components and to have a skew distribution. The results from estimating such a model for S&P are shown in Table 9. The S&P series was chosen because it is a long one, consisting of 3,175 weekly observations running from 4 Jan 1954 to 3 Nov 2014. The risk-free rate only appears in the equation for y_t, that is (17), because it was found to be small and statistically insignificant in the equation for $\lambda_{t+1|t}$. The only difference between the models in Table 9 and those reported for NASDAQ in Table 4 comes from the presence of δ. The other coefficients are not dissimilar.

The weekly risk premium is now estimated as $\mu + \alpha_1 \exp(\omega) + \delta\, \bar{r}_{f,t}$, where $\bar{r}_{f,t}$ is the sample average. The first model in Table 9 implies that the average excess return per week is $E(y_t) = \mu + \delta\, \bar{r}_{f,t} = 0.197 - 1.62 \times 0.09 = 0.051\%$ per week. This implies an annual excess return of 2.7% which is close to the historic average of around 2.8% per year. For the third model in Table 9, the weekly risk premium is $0.015 + 0.156 \exp(0.493) - 1.82 \times 0.09 = 0.075\%$, implying an annual average risk premium of 4.0%. The reason this figure is higher than the historic excess return is that the effects of extreme values receive less weight[6].

Scruggs (1998) proposes a two equation model in which the roles of y_t and $r_{f,t}$ are interchanged in the second equation. Such a model can be handled within the framework outlined in the previous sub-section.

5 Conclusions

The way in which the EGARCH-M is formulated resolves a number of contradictions and puzzles in the literature on the relationship between volatility and the risk premium. The ease with which a two component dynamic 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 actually reduce short-term volatility, an effect that cannot be found when leverage is introduced into GARCH models. Thus while returns have a symmetric effect on volatility in the long-run, they have something approaching an anti-symmetric effect in the short-run. As regards the risk premium, our results for both weekly and daily data allow us to reject both a constant and a rapidly varying risk premium in favour of a risk premium that is associated with the slowly varying component of volatility. Whereas long-term volatility is associated with a higher return, the opposite appears to be the case with short-term volatility, presumably because increased uncertainty drives away

[6] On the other hand, it is somewhat lower than the results in earlier papers. A possible explanation is that our sample includes both the dot-com and financial crises.

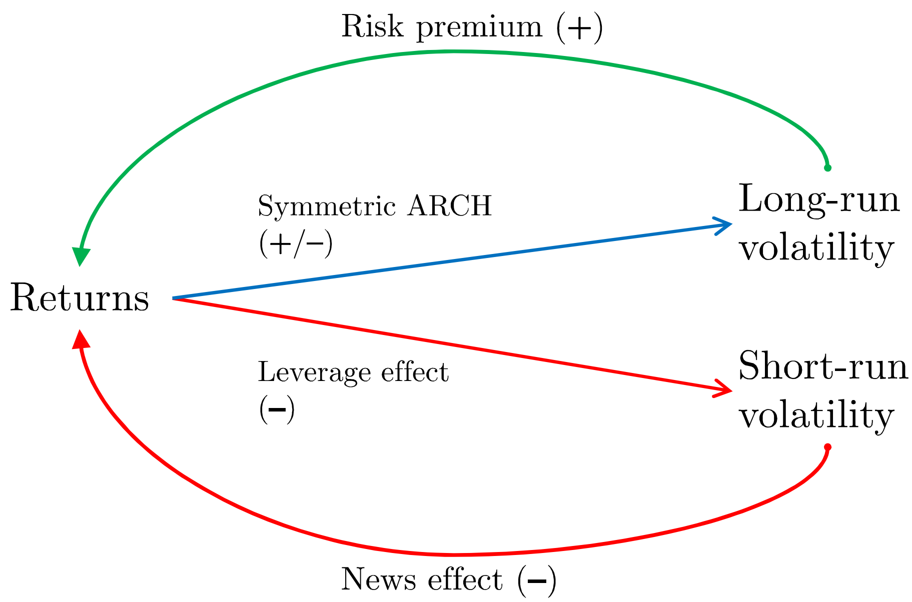

Figure 7: Interactions between returns and volatility for a two component model

nervous investors and less uncertainty has a calming effect[7]. If volatility is modeled with only one component, these effects tend to cancel each other out which is why many previous studies have been inconclusive. Figure 7 shows how the various puzzles of Figure 1 are resolved.

From the technical point of view, using of the score of the conditional distribution to drive the dynamics enables statistical properties of the model to be derived. The fact that the score is bounded for a $t-$distribution simplifies the asymptotic theory, as well as having the practical advantage of curbing

[7] A second explanation is possible, where the causality is reversed. Rather than downward moves causing future volatility, it is possible that (the expectation of) future volatility causes downward moves in the present. That is, if risk-averse investors anticipate an increase (decrease) in volatility, they may adjust their exposure ex ante by selling (buying), thus driving prices down (up). For example, risk-averse investors may reduce their exposure before a large news event or, conversely, increase their exposure if they foresee calmer waters. If the investors' expectations regarding changes in volatility are on average correct, then negative (positive) returns will tend to be followed by higher (lower) volatility. According to this interpretation, the antisymmetric nature of short-term volatility is therefore a by-product of risk-averse investors correctly anticipating changes in volatility.

the impact of extreme observations on the measure of volatility. Allowing the $t-$distribution to be skewed is straightforward and it seems to be produce the best fitting models with the most plausible estimates of the risk premium.

There are several avenues for future research. One is to estimate multivariate models. Another is to exploit the relatively large sample sizes for financial time series to estimate a richer class of conditional distributions. For example, the upper and lower tails may have different degrees of freedom, as in Zhu and Galbraith (2010) and the generalized t distribution may be used, as in the ARCH-M model of Theodossiou and Savva (2015). Some preliminary results indicate the benefit of adopting such distributions, but the matter is not pursued here in order not to detract from the main thrust of the paper.

Acknowledgements

The work was carried out when Rutger-Jan Lange was a Post-Doctoral Research Associate on the project Dynamic Models for Volatility and Heavy Tails. We are grateful to the Keynes Fund for financial support. Earlier versions of this paper were presented at the conference Recent Developments in Financial Econometrics and Empirical Finance at the University of Essex, at a Cambridge Finance-Tinbergen Institute meeting in Amsterdam and at the CFE conference in Pisa. We are grateful to Peter Boswijk, Andre Lucas, Siem-Jan Koopman, Mark Salmon and Robert Taylor for helpful comments.

APPENDIX

A Autocorrelation function for EGARCH-M

To derive (8) first note that the unconditional moments about μ are given by $E(\alpha+\varepsilon_t)^m\, E(\exp m\lambda_{t|t-1})$, $m = 1, 2, ...$ Thus the mean is $\mu+\alpha E(\exp\lambda_{t|t-1}) = \mu + \mu_\lambda$, whereas the variance is

$$Var(y_t) = (\sigma_\varepsilon^2+\alpha^2)E(\exp 2\lambda_{t|t-1})-\mu_\lambda^2 = (\sigma_\varepsilon^2+\alpha^2)E(\exp 2\lambda_{t|t-1})-\alpha^2[E(\exp\lambda_{t|t-1})]^2.$$

The autocovariance of y_t at $lag\ \tau$ is

$$\begin{aligned}\gamma(\tau) &= E(y_t-\mu-\mu_\lambda)(y_{t-\tau}-\mu-\mu_\lambda)] \\ &= \alpha^2 E(\exp(\lambda_{t|t-1}+\lambda_{t-\tau|t-\tau-1}))+\alpha E(\varepsilon_{t-\tau}\exp(\lambda_{t|t-1}+\lambda_{t-\tau|t-\tau-1}))-\mu_\lambda^2,\end{aligned}$$

because the two terms containing ε_t have zero expectation (by the law of iterated expectations). The term $E(\varepsilon_{t-\tau}\exp(\lambda_{t|t-1}+\lambda_{t-\tau|t-\tau-1}))$ is also zero because, although $\lambda_{t|t-1}$ depends on $\varepsilon_{t-\tau}$, we can write

$$\begin{aligned}E(\varepsilon_{t-\tau}\exp(\lambda_{t|t-1}+\lambda_{t-\tau|t-\tau-1})) &= \Pr(\varepsilon_{t-\tau}>0)E(|\varepsilon_{t-\tau}|\exp(\lambda_{t|t-1}+\lambda_{t-\tau|t-\tau-1})) \\ -\Pr(\varepsilon_{t-\tau} &< 0)E(|\varepsilon_{t-\tau}|\exp(\lambda_{t|t-1}+\lambda_{t-\tau|t-\tau-1}))\end{aligned}$$

and this is zero when there are no leverage effects and ε_t is symmetric. Thus

$$\gamma(\tau)=\alpha^2(E(\exp(\lambda_{t|t-1}+\lambda_{t-\tau|t-\tau-1}))-[E(\exp\lambda_{t|t-1})]^2),\quad \tau=1,2,3,...$$

and (8) follows.

B Information matrix

The information matrix for the dynamic parameters in $\boldsymbol{\psi}=(\kappa\ \phi\ \omega)'$ is obtained by noting that

$$\frac{d\ln f_t(y_t\mid Y_{t-1};\boldsymbol{\psi})}{d\boldsymbol{\psi}}=\frac{\partial\ln f_t(y_t\mid Y_{t-1};\lambda_{t|t-1})}{\partial\lambda_{t|t-1}}\frac{d\lambda_{t|t-1}}{d\boldsymbol{\psi}}$$

so

$$E\left[\frac{d\ln f_t(y_t\mid Y_{t-1};\boldsymbol{\psi})}{d\boldsymbol{\psi}}\frac{d\ln f_t(y_t\mid Y_{t-1};\boldsymbol{\psi})}{d\boldsymbol{\psi}'}\right]=I_{\lambda\lambda}\mathbf{D}(\boldsymbol{\psi}) \tag{19}$$

as in Harvey (2013, p 37).

When there are additional parameters, such as α and ν, these need to be included in the full information matrix. Denote these parameters by the vector $\boldsymbol{\xi}$. The score for $\boldsymbol{\xi}$ breaks down into two parts, the first is conditional on $\lambda_{t|t-1}$ and the second is conditional on past observations, Y_{t-1}. Thus[8]

$$\frac{d\ln f(y_t\mid Y_{t-1};\boldsymbol{\xi})}{d\boldsymbol{\xi}}=\frac{\partial\ln f(y_t\mid\lambda_{t|t-1};\boldsymbol{\xi})}{\partial\boldsymbol{\xi}}+\frac{\partial\ln f(y_t\mid Y_{t-1};\lambda_{t|t-1})}{\partial\lambda_{t|t-1}}\frac{d\lambda_{t|t-1}}{d\boldsymbol{\xi}} \tag{20}$$

Conditioning on $\lambda_{t|t-1}$ automatically implies conditioning on Y_{t-1}. The converse is not true. Although Y_{t-1} is fixed, $\lambda_{t|t-1}$ may change with $\boldsymbol{\xi}$. This is the

[8] In our notation, the chain rule reads $\frac{d\,f(x,g(x))}{dx}=\frac{\partial f(x,g)}{\partial x}+\frac{\partial f(x,g)}{\partial g}\frac{dg(x)}{dx}$; i.e. the total derivative d takes into account all dependencies, while the partial derivative ∂ treats all input parameters that are not differentiated as constants.

case with α because it appears in the score, u_t. The simplified information matrix given in (9) ignores the contribution of this second term in (20).

The full information matrix for $\boldsymbol{\xi}$ is

$$E\left[\frac{d\ln f_t(y_t \mid Y_{t-1};\boldsymbol{\psi},\boldsymbol{\xi})}{d\boldsymbol{\xi}}\frac{d\ln f_t(y_t \mid Y_{t-1};\boldsymbol{\psi},\boldsymbol{\xi})}{d\boldsymbol{\xi}'}\right]$$
$$= \mathbf{I}_{\xi\xi} + I_{\lambda\lambda}E\left[\frac{d\lambda_{t|t-1}}{d\boldsymbol{\xi}}\frac{d\lambda_{t|t-1}}{d\boldsymbol{\xi}'}\right] + \mathbf{I}_{\xi\lambda}E\left[\frac{d\lambda_{t|t-1}}{d\boldsymbol{\xi}'}\right] + E\left[\frac{d\lambda_{t|t-1}}{d\boldsymbol{\xi}}\right]\mathbf{I}_{\lambda\xi},$$

where $\mathbf{I}_{\lambda\xi}$ and $\mathbf{I}_{\xi\xi}$, like $I_{\lambda\lambda}$, are as in the static information matrix. When the expectations of derivatives and cross-products of terms of u_t do not depend on the time-varying parameter,

$$E\left[\frac{d\ln f_t(y_t \mid Y_{t-1};\boldsymbol{\psi},\boldsymbol{\xi})}{d\boldsymbol{\xi}}\frac{d\ln f_t(y_t \mid Y_{t-1};\boldsymbol{\psi},\boldsymbol{\xi})}{d\boldsymbol{\xi}'}\right]$$
$$= \mathbf{I}_{\xi\xi} + I_{\lambda\lambda}\frac{\kappa^2}{1-b}\left[E\left(\frac{\partial u_t}{\partial\boldsymbol{\xi}}\frac{\partial u_t}{\partial\boldsymbol{\xi}'}\right) + \frac{1}{1-a}\left(2\phi\mathbf{I}_{\xi\lambda}\mathbf{I}_{\lambda\xi} - \kappa\mathbf{I}_{\xi\lambda}E\frac{\partial u_t}{\partial\lambda}\frac{\partial u_t}{\partial\boldsymbol{\xi}'} - \kappa\left[\mathbf{I}_{\xi\lambda}E\frac{\partial u_t}{\partial\lambda}\frac{\partial u_t}{\partial\boldsymbol{\xi}'}\right]'\right)\right]$$
$$-\frac{2\kappa}{1-a}\mathbf{I}_{\xi\lambda}\mathbf{I}_{\lambda\xi},$$

where

$$\begin{aligned} a &= \phi + \kappa E\left(\frac{\partial u_t}{\partial\lambda}\right), \qquad c = \kappa E\left(u_t\frac{\partial u_t}{\partial\lambda}\right) \qquad (21) \\ \textit{and} \qquad b &= \phi^2 + 2\phi\kappa E\left(\frac{\partial u_t}{\partial\lambda}\right) + \kappa^2 E\left(\frac{\partial u_t}{\partial\lambda}\right)^2. \end{aligned}$$

The derivation follows along the lines of that for (19) in Harvey (2013, ch 2). Details are available on request.

The block involving $\boldsymbol{\psi}$ and $\boldsymbol{\xi}$ is

$$E\left[\frac{d\ln f_t(y_t \mid Y_{t-1};\boldsymbol{\psi},\boldsymbol{\xi})}{\partial\boldsymbol{\xi}}\frac{d\ln f_t(y_t \mid Y_{t-1};\boldsymbol{\psi},\boldsymbol{\xi})}{\partial\boldsymbol{\psi}'}\right]$$
$$= \mathbf{I}_{\boldsymbol{\xi}\lambda}E\frac{d\lambda_{t|t-1}}{d\boldsymbol{\psi}'} + I_{\lambda\lambda}E\left[\frac{d\lambda_{t|t-1}}{d\boldsymbol{\xi}}\frac{\partial\lambda_{t|t-1}}{\partial\boldsymbol{\psi}'}\right]$$

and the individual terms for κ, ϕ and ω are

$$I_{\lambda\lambda}\frac{\kappa}{1-b}\left[E\left(u_t\frac{\partial u_t}{\partial\boldsymbol{\xi}}\right) - \frac{c}{1-a}\mathbf{I}_{\boldsymbol{\xi}\boldsymbol{\lambda}}\right], \qquad I_{\lambda\lambda}\frac{1}{1-b}\frac{\kappa^2}{1-a\phi}\left(E\left(u_t\frac{\partial u_t}{\partial\boldsymbol{\xi}}\right) - \frac{c}{1-a}\mathbf{I}_{\xi\lambda}\right)$$

and

$$\frac{1-\phi}{1-a}\mathbf{I}_{\xi\lambda}+I_{\lambda\lambda}\frac{\kappa}{1-b}\frac{1-\phi}{1-a}\left[\kappa E\left(\frac{\partial u_t}{\partial\lambda}\frac{\partial u_t}{\partial\boldsymbol{\xi}}\right)-(1+\phi)\mathbf{I}_{\boldsymbol{\xi\lambda}}\right]$$

respectively.

In the Gaussian EGARCH-M model $\partial u_t/\partial\alpha=-\left(y_t\exp(-\lambda_{t|t-1})-\alpha\right)-\alpha=-(\varepsilon_t+\alpha)$. Thus $E(\partial u_t/\partial\alpha)=-I_{\lambda\alpha}=-\alpha$ and $E[(\partial u_t/\partial\alpha)^2]=1+\alpha^2$. Also $\partial u_t/\partial\lambda=-(2\varepsilon_t+\alpha)(\varepsilon_t+\alpha)$, so $E(\partial u_t/\partial\lambda)=-(2+\alpha^2)$ and $E[(\partial u_t/\partial\lambda)^2]=\alpha^4+13\alpha^2+12$. The expressions in (21) are therefore $a=\phi-\kappa(2+\alpha^2)$, $b=\phi^2-2\phi\kappa(2+\alpha^2)+\kappa^2(12+13\alpha^2+\alpha^4)$ and $c=-\kappa(4+3\alpha^2)$. Finally $E[(\partial u_t/\partial\lambda)(\partial u_t/\partial\alpha)]=\alpha^3+5\alpha$ and $E[u_t(\partial u_t/\partial\alpha)]=-\alpha$. As a result, the information quantity for α in the bottom right-hand corner of (9) changes from unity to

$$1+\frac{\kappa^2(2+\alpha^2)}{1-b}\left(1+\alpha^2+\frac{2\alpha^2}{1-a}\left(\phi-\kappa(\alpha^2+5)\right)\right)-\frac{2\alpha^2\kappa}{1-a}.$$

As regards off-diagonal terms, the following expressions need to be added to $\alpha\mathbf{d}$ in the last column of (9) and to $\alpha\mathbf{d}'$ in the bottom row:

$$\frac{-\alpha(2+\alpha^2)\kappa}{1-b}\left(1+\frac{c}{1-a}\right),\qquad -\alpha\frac{(2+\alpha^2)}{1-b}\frac{\kappa^2}{1-a\phi}\left(1+\frac{c}{1-a}\right)$$

and

$$\frac{-\kappa(2+\alpha^2)}{1-b}\frac{(1-\phi)\alpha}{1-a}(1+\phi-\kappa(\alpha^2+5)).$$

References

Alizadeh, S., Brandt, M. W and F. X. Diebold (2002). Range-based estimation of stochastic volatility models, *Journal of Finance* 57, 1047-1092.

Baillie, R. T. and C. Morana (2009). Modelling long memory and structural breaks in conditional variances: An adaptive FIGARCH approach. *Journal of Economic Dynamics and Control*, 33, 1577–1592.

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

Bjørnland, H. C., and K. Leitemo. (2009). Identifying the interdependence between US monetary policy and the stock market. *Journal of Monetary Economics*, 56, 275-282.

Blasques, F., Koopman, S. J. and A. Lucas (2014). Maximum likelihood estimation for generalized autoregressive score models. Tinbergen Institute Discussion Paper, TI 2014-029/III, Amsterdam.

Bollerslev, T., Engle, R. F. and J. M. Wooldridge (1988). A Capital Asset Pricing Model with Time-Varying Covariances. *Journal of Political Economy*, 96, 116-131.

Campbell, J.Y. (1987). Stock returns and the term structure. *Journal of Financial Economics* 18, 373–399.

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

Chen, X. and E. Ghysels (2011). News—good or bad—and its impact on volatility predictions over multiple horizons. *Review of Financial Studies*, 24.1, 46-81.

Chou, R.Y. (1988), Volatility persistence and stock valuations: Some empirical evidence using GARCH. *Journal of Applied Econometrics*, 3.4, 279-294.

Creal, D., Koopman, S. J. and A. Lucas (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, S. J. and A. Lucas (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics,* **28**, 777-95.

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

Christensen, B.J., Dahl, C. M. and E. M. Iglesias (2012). Semiparametric inference in a GARCH-in-mean model. *Journal of Econometrics* 167, 458–472

Engle, R. F. and G. G. J. Lee (1999). A long-run and short-run component model of stock return volatility. In Engle, R. F. and H. White (eds.). *Cointegration, causality, and forecasting: a festschrift in honour of Clive W.J. Granger*. Oxford: Oxford University Press.

Engle, R.F., Lilien, D. M. and R. P. Robins (1987) Estimating Time Varying Risk Premia in the Term Structure: The Arch-M Model. *Econometrica*, 55, 391-407

Fama, E. F. and G. W. Schwert (1977) Asset returns and inflation, *Journal of Financial Economics* 5, 115-146.

Fernandez, C. and Steel, M.F.J. (1998). On Bayesian Modeling of Fat Tails and Skewness. *Journal of the American Statistical Association* 99, 359-371.

French, K. R., Schwert, G.W. and R. F. Stambaugh (1987), Expected Stock Returns and Volatility, *Journal of Financial Economics*, 19, 3-29.

Guo, H. and Neely, C.J. (2008). Investigating the intertemporal risk-return relation in international stock markets with the component GARCH model. *Economic letters*, 99, 371-374

Glosten, L. R., Jagannathan, R. and D. E. Runkle (1993). On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *Journal of Finance*, 48, 1779–801.

Harvey, A.C. (2013). Dynamic Models for Volatility and Heavy Tails: with applications to financial and economic time series. Econometric Society Monograph, Cambridge University Press.

Harvey, A. C. and G. Sucarrat (2014). EGARCH Models with Fat Tails, Skewness and Leverage. *Computational Statistics and Data Analysis*, 26, 320-338

Hong, E P. (1991). The autocorrelation structure for the GARCH-M process. *Economics Letters* 37, 129-132.

Koopman, S. J. and E. H. Uspensky (2002). The stochastic volatility in mean model: empirical evidence from international stock markets. *Journal of Applied Econometrics,* 17, 667–689.

Lanne, M and P. Saikkonen (2006). Why is it so difficult to uncover the risk–return tradeoff in stock returns? *Economics Letters* 92, 118–125.

Lundblad, C. (2007). The risk return tradeoff in the long run: 1836–2003. *Journal of Financial Economics*, 85, 1, 123–150.

Merton, R.C. (1973). An intertemporal captial asset pricing model. *Econometrica* 41, 867-887

Mills, T. and R.N. Markellos (2008). *The Econometric Modelling of Financial Time Series,* 3rd ed. Cambridge University Press.

Nelson, C. R. (1976) Inflation and rates of return on common stocks. Journal of Finance 31, 471-483.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* 59, 347-370.

Poon, S. and S.J. Taylor (1992). Stock returns and volatility: an empirical study of the UK stock market. *Journal of Banking and Finance,* 16, 37–59.

Schwert, G.W. (1989). Why does stock market volatility change over time? *The Journal of Finance* 44, 1115-1153.

Scruggs, J.T. (1998). Resolving the puzzling intertemporal relation between market risk premium and conditional market variance: a two-factor approach. *Journal of Finance,* 53, 575-603.

Taylor, S. (2005). *Asset Price Dynamics, Volatility, and Prediction.* Princeton University Press.

Theodossiou, P. and C.S. Savva. (2015), *Skewness and the relation between risk and return*, forthcoming in *Management Science*

Turner, C.M., R. Startz, and C.R. Nelson (1989). A Markov model of heteroskedasticity, risk, and learning in the stock market. *Journal of Financial Economics*, 25, 3–22.

Zhu, D. and J.W. Galbraith (2010). A generalized asymmetric Student-t distribution with application to financial econometrics. *Journal of Econometrics* 157, 297–305.

Table 2: NASDAQ weekly excess returns (2,282 observations from 8 Feb 1972 to 3 Nov 2014)

Model	**Volatility**				**Mean**		**Shape**	**Fit**			**Q(20)**			
	κ	κ^*	φ	ω	μ	α	ν	LogL	AIC	BIC	ε_t	$u_{\mu,t}$	ε_t^2	$u_{\lambda,t}$
Beta-t-EGARCH	0.056	0.021	0.970	0.594	0.208		7.20	-5135.2	4.5059	4.5210	51.6	49.1	25.3	20.9
	0.008	0.005	0.006	0.070	0.043		0.93				.00	.00	.19	.40
	0.056	0.020	0.970	0.595	0.137	0.041	7.20	-5135.1	4.5067	4.5243	53.2	50.5	26.2	21.4
	0.008	0.005	0.006	0.069	0.152	0.084	0.93				.00	.00	.16	.38
	δ	δ^*	β	ω	μ	α	ν	LogL	AIC	BIC	ε_t	$u_{\mu,t}$	ε_t^2	$u_{\lambda,t}$
GARCH	0.075	0.119	0.815	0.297	0.211		8.24	-5140.4	4.5104	4.5255	52.8		27.8	
	0.019	0.031	0.025	0.066	0.043		1.22				.01		.58	
	0.076	0.117	0.815	0.296	0.192	0.010	8.23	-5140.4	4.5113	4.5288	53.3		27.9	
	0.020	0.033	0.025	0.065	0.156	0.073	1.22				.01		.58	

Table 3: NASDAQ weekly excess returns

Volatility							**Mean**				**Shape**	**Fit**			**Q(20)**	
κ_1	κ_1^*	φ_1	κ_2	κ_2^*	φ_2	ω	μ	α_1	α_2	α	ν	LogL	AIC	BIC	$u_{\mu,t}$	$u_{\lambda,t}$
0.034	0.004	0.989	0.024	0.049	0.724	0.565	0.231				7.60	-5118.3	4.4937	4.5163	53.1	21.0
0.006	0.004	0.003	0.013	0.009	0.066	0.104	0.042				1.02				.00	.40
0.036	0.002	0.988	0.022	0.051	0.726	0.592	-0.012	0.140			7.63	-5117.4	4.4937	4.5189	53.0	20.7
0.007	0.004	0.002	0.013	0.009	0.063	0.095	0.187	0.105			1.04				.00	.41
0.043	0.007	0.985	-0.009	0.061	0.711	0.580	2.872		-2.644		7.65	-5093.5	4.4729	4.4980	16.6	20.4
0.006	0.005	0.003	0.009	0.010	0.048	0.089	0.553		0.548		1.02				.68	.44
0.033	0.005	0.990	0.022	0.054	0.722	0.551	0.445			-0.122	7.52	-5117.3	4.4937	4.5188	46.5	21.1
0.006	0.004	0.002	0.013	0.010	0.060	0.105	0.160			0.089	1.01				.00	.39
0.042	0.006	0.984	-0.006	0.061	0.709	0.585	2.611	0.147	-2.643		7.73	-5092.8	4.4731	4.5007	16.6	20.1
0.006	0.005	0.003	0.009	0.010	0.048	0.085	0.607	0.122	0.562		1.04				.68	.45
0.043	0.006	0.984	-0.006	0.061	0.709	0.584	$\mu \equiv -\alpha_2$	0.130	-2.640		7.72	-5092.8	4.4722	4.4973	16.6	20.2
0.006	0.005	0.003	0.009	0.010	0.048	0.085		0.033	0.549		1.03				.68	.45
0.045		0.984		0.053	0.743	0.610	0.234				7.33	-5120.2	4.4936	4.5112	53.5	21.8
0.007		0.005		0.009	0.058	0.088	0.042				0.94				.00	.35
0.043		0.985		0.066	0.727	0.601	2.220	0.176	-2.297		7.68	-5093.1	4.4716	4.4943	16.1	18.8
0.006		0.005		0.009	0.045	0.089	0.449	0.122	0.408		1.00				.71	.53
0.043		0.985		0.066	0.727	0.609	$\mu \equiv -\alpha_2$	0.133	-2.290		7.68	-5093.8	4.4714	4.4915	16.1	18.8
0.006		0.005		0.009	0.045	0.088		0.033	0.405		1.00				.71	.54

Table 4: NASDAQ weekly excess returns with skew t-distribution

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	LogL	AIC	BIC	$u_{\mu,t}$	$u_{\lambda,t}$
0.036	0.010	0.988	0.032	0.044	0.731	0.654	0.154			8.23	0.80	-5095.0	4.4741	4.4993	53.7	25.5
0.007	0.004	0.004	0.012	0.009	0.075	0.101	0.044			1.17	0.03				.00	.18
0.043	0.015	0.982	0.012	0.059	0.700	0.684	2.567	0.153	-2.768	8.27	0.81	-5069.8	4.4539	4.4840	22.5	24.7
0.007	0.005	0.004	0.015	0.010	0.052	0.086	0.614	0.176	0.512	1.22	0.03				.31	.21
0.044	0.015	0.984	0.009	0.059	0.698	0.676	$\mu \equiv -\alpha_2$	0.051	-2.717	8.18	0.81	-5070.2	4.4533	4.4810	22.5	24.5
0.006	0.005	0.003	0.010	0.010	0.053	0.093		0.036	0.463	1.15	0.03				.32	.22

Table 5: NIKKEI weekly returns (1,595 observations from 4 Jan 1984 to 20 Oct 2014)

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	LogL	AIC	BIC	$u_{\mu,t}$	$u_{\lambda,t}$
0.047	0.006	0.969	0.002	0.049	0.729	0.791	0.174			8.20	-3800.6	4.7770	4.8073	24.4	11.8
0.016	0.008	0.014	0.023	0.012	0.089	0.067	0.059			1.27				.23	.92
0.062	0.009	0.958	-0.029	0.048	0.766	0.796	2.142	-0.120	-1.713	8.40	-3795.3	4.7727	4.8098	16.6	14.5
0.017	0.011	0.015	0.021	0.014	0.081	0.065	0.818	0.138	0.730	1.29				.68	.81
0.049	0.008	0.962	-0.016	0.054	0.780	0.795	$\mu \equiv -\alpha_2$	0.071	-1.709	8.41	-3796.2	4.7726	4.8063	17.9	14.6
0.013	0.010	0.010	0.015	0.011	0.069	0.065		0.035	0.774	1.33				.60	.80
0.047		0.969		0.054	0.763	0.804	0.176			8.27	-3800.9	4.7748	4.7984	24.2	11.5
0.008		0.010		0.010	0.061	0.068	0.059			1.29				.23	.93
0.046		0.971		0.055	0.810	0.800	1.490	-0.032	-1.266	8.36	-3796.7	4.7721	4.8024	23.1	14.8
0.008		0.010		0.009	0.053	0.070	0.535	0.145	0.465	1.30				.28	.79
0.046		0.970		0.055	0.817	0.804	$\mu \equiv -\alpha_2$	0.070	-1.312	8.37	-3796.9	4.7711	4.7980	17.2	14.9
0.008		0.010		0.009	0.051	0.068		0.035	0.463	1.30				.64	.79

Table 6: NIKKEI weekly returns with skew t-distribution

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	LogL	AIC	BIC	$u_{\mu,t}$	$u_{\lambda,t}$
0.053		0.961		0.055	0.794	0.822	0.111			8.61	0.85	-3791.5	4.7642	4.7912	20.0	11.5
0.008		0.010		0.009	0.051	0.063	0.061			1.43	0.03				.46	.93
0.052		0.963		0.061	0.826	0.831	1.296	-0.036	-1.156	8.70	0.85	-3787.1	4.7613	4.7950	13.9	14.8
0.008		0.009		0.009	0.043	0.066	0.810	0.172	0.557	1.45	0.03				.83	.79
0.052		0.962		0.061	0.829	0.832	$\mu \equiv -\alpha_2$	0.025	-1.141	8.70	0.85	-3787.2	4.7602	4.7905	14.0	14.9
0.008		0.009		0.009	0.042	0.064		0.036	0.391	1.45	0.03				.83	.79

Table 7: NASDAQ daily excess returns (6,048 observations from 2 Jan 1991 - 31 Dec 2014)

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	LogL	AIC	BIC	$u_{\mu,t}$	$u_{\lambda,t}$
0.026	0.005	0.997	0.007	0.043	0.853	-0.270	0.068			9.36		-9721.7	3.1148	3.1245	65.4	39.3
0.004	0.002	0.001	0.005	0.005	0.031	0.144	0.012			0.95					.00	.01
0.029	0.007	0.996	-0.001	0.047	0.815	-0.266	0.375	-0.003	-0.306	9.21		-9718.1	3.1143	3.1262	58.6	39.2
0.004	0.003	0.001	0.007	0.006	0.057	0.134	0.133	0.043	0.124	0.91					.00	.01
0.028	0.006	0.996	0.001	0.046	0.833	-0.222	$\mu \equiv -\alpha_2$	0.071	-0.298	9.37		-9719.7	3.1145	3.1253	60.5	40.1
0.004	0.003	0.001	0.007	0.006	0.047	0.117		0.016	0.120	0.94					.00	.01
0.026	0.007	0.997	0.016	0.042	0.866	-0.087	0.046			10.23	0.86	-9682.2	3.1025	3.1133	60.6	33.1
0.003	0.002	0.001	0.005	0.004	0.025	0.132	0.012			1.12	0.02				.00	.03
0.028	0.009	0.997	0.012	0.044	0.859	-0.064	0.264	-0.010	-0.218	10.05	0.86	-9679.8	3.1024	3.1153	56.0	32.1
0.004	0.003	0.001	0.006	0.005	0.034	0.143	0.109	0.041	0.101	1.08	0.02				.00	.04
0.027	0.008	0.996	0.013	0.044	0.863	-0.036	$\mu \equiv -\alpha_2$	0.037	-0.232	10.16	0.85	-9680.5	3.1023	3.1141	56.7	32.2
0.003	0.003	0.001	0.006	0.005	0.032	0.125		0.016	0.106	1.10	0.02				.00	.04
0.035		0.995		0.043	0.902	-0.065	0.049			9.99	0.87	-9690.8	3.1046	3.1132	59.0	31.4
0.003		0.001		0.004	0.015	0.102	0.012			1.09	0.02				.00	.05
0.034		0.995		0.045	0.908	-0.076	0.229	-0.004	-0.184	9.86	0.87	-9688.1	3.1044	3.1152	55.0	31.2
0.003		0.001		0.004	0.018	0.110	0.084	0.031	0.088	1.04	0.02				.00	.05
0.033		0.995		0.045	0.908	-0.072	$\mu \equiv -\alpha_2$	0.041	-0.185	9.94	0.87	-9688.8	3.1043	3.1140	55.3	31.6
0.003		0.001		0.004	0.017	0.104		0.017	0.079	1.07	0.02				.00	.05

Table 8: NIKKEI daily returns (7,601 observations from 5 Jan 1984 to 18 Nov 2014)

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	LogL	AIC	BIC	$u_{\mu,t}$	$u_{\lambda,t}$
0.047	0.012	0.987	-0.007	0.043	0.735	-0.028	0.045			7.76		-12205.4	3.2139	3.2221	17.9	23.7
0.004	0.003	0.002	0.008	0.005	0.049	0.060	0.011			0.61					.60	.26
0.049	0.014	0.987	-0.01	0.043	0.741	-0.025	0.321	-0.062	-0.227	7.48		-12202.1	3.2135	3.2236	18.4	22.3
0.004	0.003	0.002	0.007	0.006	0.047	0.063	0.131	0.036	0.128	0.56					.56	.33
0.047	0.014	0.985	-0.009	0.042	0.751	-0.011	$\mu \equiv -\alpha_2$	0.034	-0.285	7.86		-12205.7	3.2142	3.2234	19.9	21.7
0.004	0.003	0.002	0.007	0.006	0.046	0.056		0.014	0.127	0.63					.46	.36
0.047	0.014	0.986	-0.003	0.041	0.743	0.026	0.034			7.94	0.94	-12198.2	3.2123	3.2214	18.8	22.9
0.005	0.003	0.002	0.008	0.005	0.049	0.060	0.012			0.64	0.02				.53	.29
0.049	0.016	0.987	-0.006	0.041	0.750	0.045	0.325	-0.071	-0.236	7.96	0.94	-12194.1	3.2117	3.2227	19.4	21.1
0.004	0.003	0.002	0.007	0.005	0.046	0.066	0.139	0.037	0.136	0.65	0.02				.50	.39
0.048	0.016	0.985	-0.005	0.041	0.762	0.053	$\mu \equiv -\alpha_2$	0.017	-0.297	8.02	0.94	-12197.4	3.2123	3.2223	20.6	21.1
0.004	0.003	0.002	0.007	0.005	0.046	0.059		0.014	0.130	0.65	0.02				.42	.39
0.047		0.988		0.046	0.873	0.019	0.038			7.92	0.95	-12207.7	3.2142	3.2215	17.8	25.6
0.003		0.002		0.005	0.025	0.062	0.012			0.64	0.02				.60	.18
0.048		0.988		0.047	0.885	0.023	0.257	-0.071	-0.164	7.93	0.95	-12204.0	3.2138	3.2229	17.9	24.5
0.003		0.002		0.005	0.024	0.064	0.088	0.036	0.081	0.64	0.02				.59	.22
0.047		0.987		0.046	0.891	0.029	$\mu \equiv -\alpha_2$	0.021	-0.186	7.98	0.94	-12207.2	3.2144	3.2226	18.3	23.8
0.003		0.002		0.005	0.023	0.061		0.015	0.080	0.65	0.02				.57	.25

Table 9: SP500 weekly excess returns (3,175 observations from 4 Jan 1954 - 3 Nov 2014) with correction for the risk-free rate

Volatility							**Mean**				**Shape**			**Fit**			**Q(20)**	
κ_1	κ_1^*	φ_1	κ_2	κ_2^*	φ_2	ω	μ	α_1	α_2	α	ν	γ	δ	LogL	AIC	BIC	$u_{\mu,t}$	$u_{\lambda,t}$
0.033	0.009	0.989	0.026	0.058	0.798	0.503	0.197				10.36	0.82	-1.62	-6384.7	4.0288	4.0498	26.8	15.9
0.005	0.004	0.002	0.009	0.008	0.047	0.088	0.053				1.56	0.02	0.57				.14	.72
0.032	0.009	0.987	0.026	0.058	0.799	0.493		0.147			10.61	0.82	-1.85	-6383.1	4.0278	4.0488	27.4	16.7
0.005	0.004	0.002	0.009	0.008	0.047	0.075		0.039			1.62	0.02	0.63				.13	.67
0.032	0.009	0.987	0.026	0.058	0.800	0.493	-0.015	0.156			10.62	0.82	-1.84	-6383.1	4.0284	4.0513	27.3	16.7
0.006	0.005	0.004	0.010	0.008	0.053	0.074	0.085	0.063			1.59	0.02	0.63				.13	.67
0.033	0.010	0.989	0.026	0.059	0.798	0.506	0.282		-0.091		10.33	0.82	-1.59	-6384.6	4.0294	4.0523	27.0	16.2
0.005	0.005	0.002	0.010	0.008	0.048	0.089	0.156		0.162		1.55	0.02	0.60				.14	.71
0.032	0.008	0.988	0.026	0.057	0.796	0.490	0.075			0.100	10.56	0.82	-1.82	-6383.6	4.0288	4.0517	27.4	16.1
0.005	0.004	0.002	0.009	0.008	0.046	0.079	0.126			0.071	1.62	0.02	0.74				.13	.71
0.032	0.009	0.987	0.026	0.058	0.798	0.496	0.082	0.145	-0.086		10.58	0.82	-1.82	-6383.0	4.0290	4.0538	27.6	16.9
0.005	0.005	0.003	0.009	0.008	0.049	0.076	0.246	0.110	0.235		1.62	0.02	0.63				.12	.66
0.031	0.011	0.988	0.026	0.057	0.814	0.561	$\mu \equiv -\alpha_2$	0.132	-0.252		10.54	0.82	-1.59	-6383.6	4.0287	4.0516	28.1	16.2
0.005	0.005	0.002	0.009	0.008	0.049	0.096		0.042	0.192		1.60	0.02	0.65				.11	.71