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 two component ARCH in Mean model to separate the two effects; such a model also picks up the long memory features of the data. An exponential formulation, with the dynamics driven by the score of the conditional distribution, is shown to be theoretically tractable as well as practically useful. In particular it enables us to write down the asymptotic distribution of the maximum likelihood estimator, something that has not proved possible for standard formulations of ARCH in Mean. Our EGARCH-M model in which the returns have a conditional skewed generalized-t distribution is shown to give a good fit to daily S&P500 excess returns over a 60-year period (4 Jan 1954 - 30 Dec 2013).

Keywords: Dynamic conditional score (DCS) model; generalized t-distribution; leverage; returns; risk premium; two component model; volatility.

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 (p 205, p 252-4), Mills and Markellos (2008, 3rd ed, p. 287-93) and Martin et al (2013). Recent papers on the topic include Baillie and Morana (2009), Lundblad

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 theory for the proposed class of DCS EGARCH-M models can be obtained by extending the results in Harvey (2013) and Blasques et al. (2014). 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. When coupled with their theoretical advantages, this makes them extremely attractive.

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. We develop this model further by investigating the viability of a long-run, or persistent, component that derives from a specification which, in a linear state space model, yields smoothed estimates that are essentially a cubic spline. Hence it plays a similar role to the splines used by Engle and Rangel (2008) to capture long-term movements in volatility. It also provides an alternative to the trigonometric terms used by Baillie and Morana (2009), Christiansen et al. (2010) and others. Our stochastic specification can be incorporated into the filter, and hence updated with each new observation, and it can be modified so as to make it stationary.

A two-component model enables the researcher to distinguish between the effects of short and long-run volatility. This distinction is similar to that implied by the ‘news effect’ described by Chou (1988) and Schwert (1989), which predicts a negative correlation between unexpected (i.e. short-run) volatility and return because it makes investors nervous of risk, as opposed to that emanating from Merton’s (1973) intertemporal capital asset pricing model (ICAPM), which predicts a positive correlation between expected (i.e. long-run) volatility and return; see French et al. (1987). Failure to model both types 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.

Leverage plays an important role in volatility models. It acts in the opposite direction to the risk premium effect, in that negative returns are associated with increased volatility; see Bekeart and Wu (2000) for an excellent discussion of the finance issues involved. Figure 1 highlights some of the relationships between volatility and returns. Again a two component model plays an important role in that it allows us to whether the effect of leverage is different in the short-run and the long-run. For example it may be that leverage mainly affects the volatility short-run 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. 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. It is interesting to note that Chen and Ghysels (2008) have recently identified the presence of such effects in diurnal data.

The next section sets out the DCS formulation of the EGARCH-M model and section 3 gives the large sample distribution. Section 4 establishes the conditions for the existence of moments, explores the patterns of autocorrelation functions.

Figure 1: Some interactions between volatility and returns.
and discusses predictions. Section 5 introduces the skew generalized Student’s $t$-distribution for returns. Section 6 fits the new DCS EGARCH models to S&P500 excess returns over a 60-year period (4 Jan 1954 - 30 Dec 2013) and compares them with standard GARCH-M models. Section 7 indicates how the DCS EGARCH modeling framework can be generalized to multivariate time series. Section 8 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

$$y_t = \mu + \alpha \sigma_{t-1}^m + \varepsilon_t \sigma_{t-1}, \quad t = 1, \ldots, T,$$

where $\sigma_{t-1}$ is the conditional standard deviation, $\varepsilon_t$ is a serially independent standard normal variable, that is $\varepsilon_t \sim NID(0,1)$, $\mu$ and $\alpha$ are parameters and $m$ is typically set to one or two (or it can be estimated). The conditional variance, $\sigma_{t-1}^2$, is assumed to be generated by a dynamic equation that is dependent on squared observations, so in the GARCH(1,1) case

$$\sigma_{t+1|t}^2 = \omega(1 - \phi) + \phi \sigma_{t|t}^2 + \kappa(y_t^2 - \sigma_{t|t}^2),$$

where $\omega, \phi$ and $\kappa$ are parameters.

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 will be set up as

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

where $\exp(\lambda_{t-1})$ is the scale, with the dynamic equation for $\lambda_{t-1}$ 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 $t$. When $\varepsilon_t$ is $NID(0,1)$, in which case the scale is the standard deviation, the score with respect to $\lambda_{t-1}$ is

$$u_t = (y_t - \mu)e^{-\lambda_{t-1}}((y_t - \mu)e^{-\lambda_{t-1}} - \alpha) - 1, \quad t = 1, \ldots, T,$$

which, on substituting $(y_t - \mu)e^{-\lambda_{t-1}} = \alpha + \varepsilon_t$, becomes

$$u_t = \varepsilon_t^2 - 1 + \alpha \varepsilon_t, \quad t = 1, \ldots, T.$$
At the true parameter values, the scores are IID\(\{0, \sigma_u^2\}\), where \(\sigma_u^2 = E[(\varepsilon_t^2 - 1 + \alpha \varepsilon_t)^2] = 2 + \alpha^2\) because \(\varepsilon_t^2 \sim \chi^2_1\) and so its variance is two.

In a pure DCS-EGARCH volatility model, that is \(\alpha = 0\), the dynamics are driven by the squares of the \(\varepsilon_t's\) when the conditional distribution is Gaussian. In the DCS-EGARCH-M model, the value of \(\varepsilon_t\) itself appears, reflecting the fact that it, too, is informative about the movements in \(\lambda_{t+1|t}\). This term does not appear in the GARCH-M model, (2), where the variable corresponding to \(u_t\) is \(y_t^2 - \sigma_{t-1}^2 = \sigma_{t-1}^2(\varepsilon_t^2 - 1)\). Having said that, it is important to note that the inclusion in \(u_t\) of \(\alpha \varepsilon_t\), which we will hereafter call the ARCH-M score term, is not crucial to the model because \(\alpha\) is typically very small: dropping it makes very little practical difference and the theoretical properties of the model as a whole become much simpler.

Remark 1 Note that \(\alpha \varepsilon_t\) is not a leverage term, despite the superficial resemblance to the leverage term in the classic EGARCH model of Nelson (1991). Indeed it will normally act in the other direction because, as a rule, \(\alpha\) will be positive. In a Gaussian DCS EGARCH model leverage is captured by the variable \(\varepsilon_t^2 \text{sgn}(-\varepsilon_t)\); see the discussion in sub-section 2.2.

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 \(\alpha \varepsilon_t \exp(\lambda_{t-1})\). More generally, for any \(m > 0\), \(u_t = \varepsilon_t^2 - 1 + \alpha \varepsilon_t \exp(\lambda_{t-1}(m - 1))\). The presence of \(\lambda_{t-1}\) in the score for \(m \neq 1\) leads to a less tractable asymptotic theory. The use of \(\ln \sigma_{t-1}\), as suggested by Engle et al (1987) for some series, means that \(\alpha \varepsilon_t \exp(\lambda_{t-1})\) is replaced by \(\alpha \lambda_{t-1}\) and the ARCH-M score term again depends on \(\lambda_{t-1}\).

2.1 Student-t distribution

The Student-t distribution usually improves the fit to returns. The score for a conditional t-distribution with \(\nu\) degrees of freedom and scale \(\exp(\lambda_{t-1})\) is

\[ u_t = (\nu + 1)b_t - 1 + \alpha(1 - b_t)([\nu + 1]/\nu)\varepsilon_t, \]  

(6)

where

\[ b_t = \frac{\varepsilon_t^2/\nu}{1 + \varepsilon_t^2/\nu}, \quad 0 \leq b_t \leq 1, \quad 0 < \nu < \infty, \]  

(7)

is distributed as \(\text{beta}(1/2, \nu/2)\) at the true parameter values. In the absence of the ARCH-M effect, this model is known\(^1\) as Beta-t-EGARCH.

\(^1\)The (potential) instability in conventional EGARCH-M models noted by St. Pierre (1998) seems not to be an issue with Beta-t-EGARCH-M, presumably because the score downweights outliers.
A Student-t distribution, or a generalized Student-t distribution as in section 5, is assumed when models are estimated. However, the formulae associated with the Gaussian distribution (obtained by $\nu \to \infty$) are often much simpler and are sometimes employed to illustrate theoretical points.

### 2.2 Dynamics and leverage

The basic dynamic model for $\lambda_{t|t-1}$ is the stationary first-order process\(^2\)

\[
\lambda_{t+1|t} = \omega (1 - \phi) + \phi \lambda_{t|t-1} + \kappa u_t, \quad |\phi| < 1, \tag{8}
\]

where $\omega$, $\phi$ and $\kappa$ are parameters, with $\omega$ denoting the unconditional mean, and $\lambda_{1|0} = \omega$.

The leverage effect in finance suggests that volatility rises when the asset price falls. This asymmetry can be captured by a DCS EGARCH model by modifying the dynamic equation in (8) to

\[
\lambda_{t+1|t} = \omega (1 - \phi) + \phi \lambda_{t|t-1} + \kappa u_t + \kappa^* \text{sgn}(-\varepsilon_t)(u_t + 1), \tag{9}
\]

where $\kappa^*$ 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). The rise in volatility following a fall in the asset price need not necessarily be due to leverage as such\(^3\). For example a sharp fall in asset price may induce more uncertainty and hence higher variability.

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; see Taylor (2005, p. 220-1). The sign cannot be used because negative values could give a negative conditional variance.

### 2.3 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

\[
\begin{align*}
\lambda_{t|t-1} &= \omega + \lambda_{1.1|t-1} + \lambda_{2.1|t-1}, \\
\lambda_{i.t+1|t} &= \phi_i \lambda_{i.t|t-1} + \kappa_i u_{i.t} + \kappa^*_i \text{sgn}(-\varepsilon_t)(u_{i.t} + 1), \quad i = 1, 2, 
\end{align*} \tag{10}
\]

\(^2\)More general dynamic models can be constructed and explanatory variables may be added; see Harvey (2013).

\(^3\)Bekaert and Wu (2002) write ‘To many, leverage effects have become synonymous with asymmetric volatility, the asymmetric nature of the volatility response to return shocks.’
where $\phi_1 > \phi_2$ if $\lambda_{1,t-1}$ denotes the long-run component. The scores, $u_{i,t}$, will usually be the same but may be different when the ARCH-M effect is not driven by total volatility.

An attraction of a two component model is that it allows the leverage effect to impact differently in the long run and short run. For example, Engle and Lee (1999) and Harvey (2013, p 134-5) find that leverage is confined to the short-term component. If this is the case, the evolution of the long-run component is even less susceptible to the influence of strongly negative returns and so may be more suitable for capturing the ARCH-M effect. If $\phi_1$ is set to one (and $\omega = 0$), $\lambda_{1,t+1|t}$ is the long-run forecast of volatility and $\alpha \exp(\lambda_{1,t+1|t})$ is the corresponding forecast of the risk premium.

The possibility of including a short-term ARCH-M effect may also be investigated. Equation (3) may be amended to become

$$y_t = \mu + \alpha_1 \exp(\omega + \lambda_{1,t-1}) + \alpha_2 \exp(\lambda_{2,t-1} - 1) + \varepsilon_t \exp(\lambda_{t-1}), \quad t = 1, ..., T,$$

so that when volatility is at its equilibrium level, the risk premium is $\mu + \alpha_1 \exp(\omega)$.

### 2.4 Damped spline filter

Engle and Rangel (2008) propose the use of cubic splines for capturing slowly changing movements in volatility. Their approach is to fit the splines nonparametrically with knots at selected points throughout the sample. However, the smoothed estimates of an integrated random walk trend in a linear unobserved components model are known to be closely related to a cubic spline; see Harvey (2013, pp 77-9). The same idea may be adapted for modeling the long-run component of volatility in a DCS EGARCH model. The filter is

$$\lambda_{1,t+1|t} = \lambda_{1,t|t-1} + \lambda_{1,t|t-1}^\beta + \kappa_1 u_t$$

where the starting values, $\lambda_{1,1|0}$ and $\lambda_{1,1|0}^\beta$, are treated as fixed and estimated along with the other parameters. A persistent stationary component may be constructed by introducing a damping factor, $\phi$, which is close to one, into the equations. Thus

$$\lambda_{1,t+1|t} = \phi_1 \lambda_{1,t|t-1} + \lambda_{1,t|t-1}^\beta + \kappa_1 u_t, \quad |\phi_1| < 1, \quad (12)$$

$$\lambda_{1,t+1|t}^\beta = \phi_1 \lambda_{1,t|t-1}^\beta + \frac{\kappa_1^2}{2 - \kappa_1} u_t.$$

We will refer to this (constrained) second-order model as the damped spline filter.
Remark 2 The use of the score to drive the dynamics poses one further issue. As noted in Remark 1, the ARCH-M score term, $\alpha \varepsilon_t$ in the Gaussian model, will normally act in the opposite direction from that of the leverage term. One solution is to drop the ARCH-M score term from the dynamic equation. However, if the ARCH-M effect depends only on $\lambda_{1,t-1}$, the DCS model has a short-term dynamic equation\(^4\) that does not contain an ARCH-M term in the score.

Remark 3 Identifiability requires $\phi_1 \neq \phi_2$, which we have implicitly assumed by setting $\phi_1 > \phi_2$, together with $\kappa_1 \neq 0$ and $\kappa_2 \neq 0$.

3 Large sample distribution of the maximum likelihood estimator

When the conditional distribution of returns is Student’s t and $\mu$ is known to be zero, an analytic expression for the information matrix can be derived. The reason this is possible is that the information matrix for the static model, that is $\lambda_{t|t-1} = \lambda$, is independent of $\lambda$, as are $E(\partial u_t/\partial \lambda)$ and $E(u_t \partial u_t/\partial \lambda)$.

Proposition 4 Suppose that the distribution of $y_t$ conditional on past observations in (3) is Student’s t with scale $\exp \lambda_{t|t-1}$ and mean $\alpha \exp \lambda_{t|t-1}$, where $\lambda_{t|t-1}$ is generated by the stationary first-order dynamic process (8). The degrees of freedom, $\nu$, is assumed to be known. Let $\psi = (\kappa, \phi, \omega)'$ and define $a$, $b$ and $c$ as

\[
a = \phi + \kappa E\left(\frac{\partial u_t}{\partial \lambda}\right) \tag{13}
b = \phi^2 + 2\phi \kappa E\left(\frac{\partial u_t}{\partial \lambda}\right) + \kappa^2 E\left(\frac{\partial u_t}{\partial \lambda}\right)^2 \geq 0 \tag{13}
c = \kappa E\left(u_t \frac{\partial u_t}{\partial \lambda}\right),
\]

where the model is formulated in such a way that the unconditional and conditional expectations in (13) are the same. Assuming that $b < 1$ and $\kappa \neq 0$, then $(\tilde{\psi}', \alpha)'$, the ML estimator of $(\psi', \alpha)'$, is consistent and the limiting distribution of $\sqrt{T}(\tilde{\psi}' - \psi', \tilde{\alpha} - \alpha)'$ is multivariate normal with mean vector zero and covariance matrix given by $\text{Var}(\tilde{\psi} - \psi, \tilde{\alpha} - \alpha) = I^{-1}(\psi, \alpha)$, where the information matrix is

\[
I \begin{bmatrix} \psi & \alpha \\ \alpha & 1 \end{bmatrix} = \frac{\nu + 1}{\nu + 3} \begin{bmatrix} (2(\nu/(\nu + 1)) + \alpha^2)D(\psi) & \alpha d(\psi) \\ \alpha d'(\psi) & 1 \end{bmatrix} \tag{14}
\]

with $d(\psi) = (0, 0, (1 - \phi)/(1 - a))$ and $D(\psi)$ as in Appendix A.

\(^4\)The score for the long-term component is then $u_t = \varepsilon^2 + \alpha \varepsilon_t \exp(-\lambda_{2,t-1}) - 1$ in the Gaussian model. More generally the ARCH-M score term is multiplied by $\exp(-\lambda_{2,t-1})$. 

8
Remark 5 When \( \nu \) is estimated, (14) is extended accordingly; see Harvey (2013, p 116). Note that \( E(\partial \ln f_t/\partial \alpha \cdot \partial \ln f_t/\partial \nu) = 0 \).

The form of (14) follows from Corollary 9 in Harvey (2013, p 47). The expectations of terms involving \( u_t \) and its derivatives can be found from properties of the beta distribution. For example,

\[
\sigma_u^2 = E[((\nu + 1)b_t - 1)^2] + 2\alpha[((\nu + 1)/\nu)E[(1 - b_t)((\nu + 1)b_t - 1)]\varepsilon_t] \\
+ \alpha^2 [(\nu + 1)/\nu]^2 E[(((1 - b_t)\varepsilon_t)^2] \\
= \frac{2\nu}{\nu + 3} + 0 + \alpha^2 \frac{(\nu + 1)^2}{\nu^2} \frac{\nu^2}{(\nu + 3)(\nu + 1)} = \frac{2\nu}{\nu + 3} + \alpha^2 \frac{\nu + 1}{\nu + 3}
\]

The expression for \( \sigma_u^2 \) gives \( a \) in (13) because it is the same as \( E(\partial u_t/\partial \lambda) \). It is straightforward but tedious to derive \( b \) and \( c \). However, for a Gaussian conditional distribution, it is easy to see that

\[
E(u_t^2) = E(2\varepsilon_t^2 + 3\alpha\varepsilon_t + \alpha^2)^2 = 12 + 13\alpha^2 + \alpha^4 \quad \text{and} \quad E(u_t u_t') = -4 - 3\alpha^2,
\]

and so

\[
a = \phi - \kappa(2 + \alpha^2), \quad c = -\kappa(4 + 3\alpha^2), \quad b = \phi^2 - 2\phi\kappa(2 + \alpha^2) + \kappa^2(12 + 13\alpha^2 + \alpha^4) \geq 0.
\]  

Because \( \alpha \) is typically very small, the inclusion of terms involving \( \alpha^2 \) and \( \alpha^4 \) in the above formulae makes very little practical difference. If the ARCH-M score term is dropped, or simply ignored in the evaluation of the information matrix, the expressions for the elements of the \( D(\psi) \) matrix for the \( t \)-distribution revert to those in the Beta-t-EGARCH model. Leverage effects can then be included as set out in Harvey (2013, pp 121-4).

A small Monte Carlo experiment was run in order to investigate the finite sample properties of the ML estimator. A Gaussian model was simulated and ML estimation carried out for 10,000 iterations with typical values of the parameters; but note that, according to (14) the asymptotic distribution does not depend on \( \omega \). As can be seen from Table 1, 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 \( \phi \). However, this may not be too surprising because the value of \( \phi = 0.98 \) is quite close to unity, and the estimates of \( \phi \) were constrained to be less than one

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean ( T = 10,000 )</th>
<th>RMSE ( T = 10,000 )</th>
<th>ASE ( T = 10,000 )</th>
<th>Mean ( T = 1000 )</th>
<th>RMSE ( T = 1000 )</th>
<th>ASE ( T = 1000 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \kappa )</td>
<td>0.04</td>
<td>0.0399</td>
<td>0.0024</td>
<td>0.0023</td>
<td>0.0397</td>
<td>0.0077</td>
</tr>
<tr>
<td>( \phi )</td>
<td>0.98</td>
<td>0.9794</td>
<td>0.0030</td>
<td>0.0028</td>
<td>0.9734</td>
<td>0.0155</td>
</tr>
<tr>
<td>( \omega )</td>
<td>0.10</td>
<td>0.0984</td>
<td>0.0350</td>
<td>0.0354</td>
<td>0.0882</td>
<td>0.1123</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>0.05</td>
<td>0.0499</td>
<td>0.0099</td>
<td>0.0100</td>
<td>0.0498</td>
<td>0.0314</td>
</tr>
</tbody>
</table>
Table 1: Monte Carlo results for Gaussian EGARCH-M model

When there is leverage in Beta-t-EGARCH, the information matrix is as in Harvey (2013, pp. 121-4). The main point to note is that identifiability requires only that either $\kappa$ or $\kappa^*$ be non-zero. (When there is no leverage term, $D(\psi)$ is singular if $\kappa = 0$ and the model is underidentified.) Thus Wald and LR tests of the null hypothesis that either $\kappa$ or $\kappa^*$ is zero can be carried out and this continues to be the case when there is an ARCH-M term.

**Remark 6** When $\mu$ is included in the model, the score with respect to $\mu$ is $\varepsilon_t \exp(\lambda_{t-1})$. The static information matrix for the Gaussian model is then

$$I = \begin{bmatrix} \lambda \\ \alpha \\ \mu \end{bmatrix} = \begin{bmatrix} 2 + \alpha^2 & \alpha e^{-\lambda} & 0 \\ \alpha & 1 & 0 \\ e^{-\lambda} & 0 & e^{-2\lambda} \end{bmatrix},$$

(16)

and a decomposition like the one in (14) is not possible when $\lambda$ is stochastic because of the presence of $e^{-\lambda_{t-1}}$ in the off-diagonal elements of the information matrix of the $t$-th observation conditional on the observations available at time $t - 1$.

**Remark 7** Because the above information matrix in (14) does not depend on $\alpha$ when $\alpha = 0$, a Lagrange multiplier test of the null hypothesis that $\alpha = 0$ against the alternative $\alpha \neq 0$ is very easy. To be specific, the statistic, which is asymptotically distributed as $\chi^2_1$ under the null hypothesis of a Beta-t-EGARCH model, is

$$LM = \frac{(\nu + 1)(\nu + 3)}{2\nu^2} \sum_{t=1}^{T} \left( \frac{y_{t} - \tilde{\mu}}{\exp(\lambda_{t-1})} \right)^2.$$

4 Moments, autocorrelations and predictions

One of the important features of the Beta-t-EGARCH model, is that when $\lambda_{t-1}$ is stationary, the unconditional moment of the observations exists whenever the corresponding conditional moment exists. 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). This section investigates the extent to which it is possible to generalize the Beta-t-EGARCH results to a model which includes an ARCH-M term.

---

5 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.
4.1 ACFs of powers of absolute values

For the DCS EGARCH-M model the addition of the ARCH-M score term, \( \alpha(1 - b_t)(\nu + 1)/\nu \varepsilon_t \), makes it difficult to derive corresponding analytic expressions, but the boundedness of the score, and hence \( \exp(m \lambda_{t-1}) \), for the t-distribution means that the unconditional \( m \) th moment exists whenever the corresponding conditional \( m \) th moment does.

Because \( \alpha \) is typically small, the moments and autocorrelations of absolute values are unlikely to be very different from those of the Beta-t-EGARCH model. The (unconditional) expectation of volatility in the stationary Beta-t-EGARCH model is

\[
E(\exp \lambda_{t-1}) = e^\omega \prod_{j=1}^{\infty} e^{-\psi_j} \beta_\nu(\psi_j), \tag{17}
\]

where \( \psi_j, j = 1, 2, \ldots \) is the coefficient of \( u_{t-j} \) in the moving average representation of \( \lambda_{t-1} \) and \( \beta_\nu(a) \) is Kummer’s (confluent hypergeometric) function. The correction factor \( \prod_{j=1}^{\infty} e^{-\psi_j} \beta_\nu(\psi_j) \) is greater than or equal to one because \( E(\exp \lambda_{t-1}) \geq e^\omega \) by Jensen’s inequality.

4.2 ACF of returns

The unconditional moments about \( \mu \) are given by \( E(\alpha + \varepsilon_t)^m E(\exp m \lambda_{t-1}), m = 1, 2, \ldots \). Thus the mean is \( \mu + \alpha E(\exp \lambda_{t-1}) = \mu + \mu_\lambda \), whereas the variance is

\[
\text{Var}(y_t) = (\sigma_2^2 + \alpha^2)E(\exp 2 \lambda_{t-1}) - \mu_\lambda^2 = (\sigma_2^2 + \alpha^2)E(\exp 2 \lambda_{t-1}) - \alpha^2[E(\exp \lambda_{t-1})]^2
\]

The autocovariance of \( y_t \) at lag \( \tau \), is

\[
\gamma(\tau) = E(y_t - \mu - \mu_\lambda)(y_{t-\tau} - \mu - \mu_\lambda)
= \alpha^2 E(\exp(\lambda_{t-1} + \lambda_{t-\tau-\tau-1}))
+ \alpha E(\varepsilon_{t-\tau} \exp(\lambda_{t-1} + \lambda_{t-\tau-\tau-1})) - \mu_\lambda^2, \tag{18}
\]

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

\[
E(\varepsilon_{t-\tau} \exp(\lambda_{t-1} + \lambda_{t-\tau-\tau-1})) = \Pr(\varepsilon_{t-\tau} > 0)E(|\varepsilon_{t-\tau}| \exp(\lambda_{t-1} + \lambda_{t-\tau-\tau-1}))
- \Pr(\varepsilon_{t-\tau} < 0)E(|\varepsilon_{t-\tau}| \exp(\lambda_{t-1} + \lambda_{t-\tau-\tau-1})) \tag{19}
\]

and this is zero when there are no leverage effects and \( \varepsilon_t \) is symmetric. Thus

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

11
and

\[ \rho(\tau) = \frac{E(\exp(\lambda_{t+1} + \lambda_{t+\tau-1})) - [E(\exp \lambda_{t+1})]^2}{(\sigma^2/\alpha^2)E(\exp 2\lambda_{t+1}) + E(\exp 2\lambda_{t+1}) - [E(\exp \lambda_{t+1})]^2}, \quad \tau \geq 1. \]

These autocorrelations do not depend on \( \omega \). Clearly the autocorrelations are zero for \( \alpha = 0 \). For non-zero \( \alpha \), \( \rho(\tau) \) has the same sign as the corresponding autocorrelation of the volatility, \( \exp \lambda_{t+1} \), with a pattern derived from that of the \( \lambda'_{t+1} \)’s. Note that in the GARCH-M model, the autocorrelations of the \( y' \)’s must all be positive because of the positivity constraints on the GARCH parameters; see Hong(1991). (Of course in practice the autocorrelations in the first-order model, (8), are almost invariably positive.)

When the dynamic equation contains a leverage term, the second term in (19) will be greater than the first\(^6\) for \( \kappa^* > 0 \) and so the second term in (18) will be negative for positive \( \alpha \), so weakening the autocorrelations. In fact they may no longer be positive.

### 4.3 Predictions

Expressions for conditional moments of future observations can be worked out for the pure EGARCH models. To do so here is more difficult because of the additional term in \( u_t \). But the full predictive distribution is often what is needed and simulating it is not difficult because it depends only on independent Student-t variates.

Of course if the ARCH-M score term is dropped from the dynamic equation, predictions of moments can be made as in a Beta-t-EGARCH model, but with the additional feature that the conditional mean is no longer constant. The MMSE predictor of the observation \( \ell \) steps ahead is

\[
E_T(y_{T+\ell}) = \alpha E_T \left( e^{\lambda_{T+\ell} + \lambda_{T+\ell-1}} \right) = \alpha e^{\lambda_{T+\ell}} \prod_{j=1}^{\ell-1} e^{-\psi_j} \beta_{\rho}(\psi_j), \quad \ell = 2, 3, \ldots \quad (20)
\]

where \( \lambda_{T+\ell} \) is the MMSE predictor of \( \lambda_{T+\ell} + \lambda_{T+\ell-1} \); see expression (4.30) in Harvey (2013, p 111).

### 5 Skew generalized Student-t distribution

In the Gamma-GED-EGARCH model, the conditional distribution of \( y_t \) is a general error distribution (GED), with time-varying scale parameter \( \exp(\lambda_{t+1}) \) and

\(^a\)A similar term appears in the ACF of the Beta–t-EGARCH model and can be evaluated as in Harvey (2013, p 239-41).
\( \lambda_{t-1} \) evolving as a linear function of the conditional score variable

\[
u_t = \frac{\nu}{2} \left( \frac{|y_t|}{\exp(\lambda_{t-1})} \right)^\nu - 1, \quad t = 1, \ldots, T, \quad (21)
\]

where \( \nu \) is a shape parameter that is two for the normal and one for the Laplace, or double exponential, distribution. When \( \nu < 2 \), the response is less sensitive to outliers than it is for a normal distribution, but it is far less robust than for a Beta-t-EGARCH model with small degrees of freedom; it is clear from (21) that \( u_t \) is not bounded. At the true parameter values, the \( u_t' \)s are independently distributed as gamma variates with shape parameters 2 and \( 1/\nu \); see Harvey (2013, section 4.4).

The generalized Student-t distribution contains the GED and Student-t distributions as limiting cases. It was introduced by McDonald and Newey (1987), who proposed its use for static regression models, and it was subsequently employed by Theodossiou (1998) for financial data. The generalized Student-t distribution can be used to construct a general DCS EGARCH model, which we will call Beta-Gen-t-EGARCH. A generalized t variable, \( y \), is such that \( |y| \) is distributed as a generalized beta of the second kind (GB2) with two of the shape parameters constrained so that the mode is at the origin. The PDF is

\[
f(y) = \frac{\nu}{2 \varphi^{1/\nu} B(\nu, 1/\nu)} \left( 1 + \frac{1}{\nu} \frac{|y - \mu|^\nu}{\varphi^\nu} \right)^{-\left(\frac{\nu}{2} + \frac{1}{\nu}\right)} \quad \infty < y < \infty, \quad (22)
\]

where \( \mu \) is a location parameter, \( \varphi \) is a scale parameter, whereas \( \nu \) and \( \nu \) are positive shape parameters\(^7\). Setting \( \nu = 2 \) gives \( t \), whereas \( GED(\nu) \) is obtained when \( \nu \to \infty \). The tail index is \( \nu \nu / 2 \); hence it is \( \nu \) for Student’s \( t \), but \( \nu / 2 \) for generalized Pareto, a distribution given by setting \( \nu = 1 \). The generalized Student-t introduces more flexibility into the shape of the distribution: for a given tail index and standard deviation, the distribution is more peaked for lower \( \nu \) whereas, when \( \nu \) is fixed, the peak becomes higher as the tail index decreases. The moments about the mean for \( m = 2, 4, \ldots \) are\(^8\)

\[
E[(y - \mu)^m] = \frac{\Gamma\left(\frac{1}{\nu} + \frac{m}{\nu}\right) \Gamma\left(\frac{\nu}{2} - \frac{m}{\nu}\right)}{\Gamma\left(\frac{1}{\nu}\right) \Gamma\left(\frac{\nu}{2}\right)} \nu^{\frac{m}{\nu}} \varphi^m, \quad -1 < m < \nu \nu / 2.
\]

The Beta-Gen-t-EGARCH model has \( \varphi = \exp(\lambda_{t-1}) \) and the score is

\[
u_t = \frac{\partial \ln f_t}{\partial \lambda_{t-1}} = \left( \frac{\nu \nu}{2} + 1 \right) b_t - 1, \quad (23)
\]

\(^7\)Note that \( \partial f(y)/\partial y = 0 \) at \( y = \mu \) for \( \nu > 1 \). (For \( \nu < 1 \), \( \partial f(y)/\partial y = \infty \).)

\(^8\)For example \( m = 2 \) with \( t \) gives \( Var(y) = \varphi^2 \nu / (\nu - 2) \).
where

\[ b_t = \frac{(\|y_t - \mu\| e^{-\lambda_{t-1}})^\nu / \nu}{(\|y_t - \mu\| e^{-\lambda_{t-1}})^\nu / \nu + 1} \]

is distributed as beta(1/\nu, \nu/2) at the true parameter values. For \( \nu = 2 \) we obtain the score as in (6). Letting \( \nu \to \infty \) gives the score as in (21). The score is bounded for \( \nu < \infty \). A bounded score means that the existence of moments is not affected by volatility, although they may become hugely inflated. Exact expressions for the unconditional moments can be derived in the same way as for Beta-t-EGARCH; see subsection 4.1. When \( \nu = \infty \), the existence of unconditional moments requires constraints on the parameters. Thus for the first-order model, (8), the \( m \)-th order moment only exists for \( m < 1/\nu \). When the model is estimated by ML, an exact expression for the information matrix for a single component first-order model can be constructed in a similar way as was done for the GB2 model in Harvey (2013, sub-section 5.3).

Skewness can be introduced into the distribution by means of the Fernandez and Steel (1998) method, as used by Harvey and Sucarrat (2014) for the Student-t distribution. The PDF in (22) then becomes

\[ f(y) = \frac{2}{\gamma + \frac{1}{\gamma}} \frac{\nu}{\varphi \nu^{1/\nu}} \frac{1}{B\left(\frac{\nu}{2}, \frac{1}{\nu}\right)} \left(1 + \frac{1}{\nu} \gamma^{\nu \text{sgn}(y-\mu) \varphi^\nu}\right)^{-\left(\frac{\nu}{2} + \frac{1}{\nu}\right)}, \quad \infty < y < \infty, \quad (24) \]

where \( 0 < \gamma < \infty \) and \( \gamma = 1 \) gives symmetry. This distribution is similar to, but not quite the same as, the skew generalized t distribution of McDonald et al (2009), which is based on the skew-t distribution of Hansen (1994). The score in the EGARCH model now becomes

\[ u_t = \left(\frac{\nu \nu}{2} + 1\right) \frac{(\|y_t - \mu\| e^{-\lambda_{t-1}})^\nu / \nu}{(\|y_t - \mu\| e^{-\lambda_{t-1}})^\nu / \nu + \gamma^\nu \text{sgn}(y_t - \mu) - 1}. \quad (25) \]

The introduction of skewness implies that \( \mu \) is still the mode but no longer the mean. This raises the new issue that the conditional expectation of \( \varepsilon_t \) is no longer zero. As a result, the ARCH-M effect is confounded with the conditional expectation \( \varepsilon_t \exp(\lambda_{t-1}) \). The mean of a skewed distribution is \( (\gamma + \gamma^{-1})E(|\varepsilon|) \), where \( E(|\varepsilon|) \) refers to the original (non-skewed) variable; see e.g. Harvey and Sucarrat (2014). For the skewed generalized Student-t distribution, we have

\[ \mu_\varepsilon = \left(\gamma - 1/\gamma\right) E(|\varepsilon|) = \left(\gamma - 1/\gamma\right) \frac{\Gamma(2/\nu)\Gamma(\nu/2 - 1/\nu)}{\Gamma(1/\nu)\Gamma(\nu/2)} \nu^{1/\nu} \]

\[ 9 \text{There may therefore be a case for approximating Gamma-GED-EGARCH by a Beta-Gen-t-EGARCH in which } \nu \text{ is large but bounded.} \]

\[ 10 \text{A similar problem was noted in Harvey and Sucarrat (2014), where a correction for skewness was made in order for the returns to be martingale differences. There the model was re-formulated. Here it is re-arranged.} \]
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\footnote{This assumes that the total volatility, rather than just the long-run component, enters into the ARCH-M effect.} is not given by the estimate of $\alpha$ in (3) but by $\alpha^\dagger = \alpha + \mu_\varepsilon$. When $\gamma$ is less than one, which is typically the case, $\mu_\varepsilon$ is negative and hence the ARCH-M effect is reduced. A more convenient model formulation is obtained by subtracting $\mu_\varepsilon$ from the disturbance term so that

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

where $\alpha^\dagger = \alpha + \mu_\varepsilon$. For the two-component model, a modification of (11) yields

$$y_t = \mu + \alpha_1 \exp(\omega + \lambda_{1,t,t-1}) + \alpha_2 [\exp(\lambda_{2,t,t-1}) - 1] + (\varepsilon_t - \mu_\varepsilon) \exp(\lambda_{1,t,t-1}), \quad t = 1, \ldots, T,$$

(26)

where $\alpha_1$ plays the same role as $\alpha^\dagger$ in the previous expression. If the mean, $\mu$, can be dropped, the ARCH-M effects are estimated more precisely; see Lanne and Saikkonen (2006).

### 6 A location-scale model for the S&P500

This section investigates the relationship between the conditional mean and conditional variance of S&P500 excess returns, but without imposing a particular relationship. Instead, we allow the filtered location and the filtered scale to vary independently. In particular, we assume the daily excess return $y_t$ is given by

$$y_t = \rho y_{t-1} + \mu_{t|t-1} + (\varepsilon_t - \mu_\varepsilon) \exp(\lambda_{t|t-1}),$$

(27)

where $\rho$ captures possible auto-correlation in the returns, $\mu_{t|t-1}$ is the filtered location, and $\exp(\lambda_{t|t-1})$ is the filtered scale. The residuals $\varepsilon_t$ are i.i.d., but may have non-zero mean due to skewness. The mean of the residuals is denoted by $\mu_\varepsilon$, such that the scale is multiplied by an i.i.d. random variable with zero mean. This section makes no a priori assumption as to how location and scale should move together, but examines a possible intertemporal relationship a posteriori.

Typically, in (27), the filtered location is two orders of magnitudes smaller than the filtered scale. Indeed, the ambiguous results in the ARCH-M literature are caused mainly by the fact that it is hard to filter location accurately. During 1994-2014, daily excess returns had an unconditional mean (standard deviation) of 0.016% (1.22%). The signal to noise ratio regarding the location, therefore, is very low. Further, the location filter must be robust to large values of $|y_t|$, which occur frequently as financial returns have fat tails.
This paper assumes that the residuals $\varepsilon_t$ are distributed as a (possibly skewed) generalized Student-t. The generalized Student-t distribution was introduced by McDonald and Newey (1987) and employed by Theodossiou (1998) for financial data. The p.d.f. for $\varepsilon_t$ with unit scale is

$$P(\varepsilon_t \in dx) = p(x) \, dx = \frac{v}{2^{\nu/2}} \frac{1}{B(\nu/2, 1/2)} \left( 1 + \frac{|x|^\nu}{\nu} \right)^{-(1/2 + \frac{1}{\nu})} \, dx,$$

for $x \in (-\infty, \infty)$, and where $v$ and $\nu$ are positive shape parameters. For $\nu = 2$, the Student-t distribution with $\nu$ degrees of freedom is obtained. For $\nu \to \infty$, the GED with shape parameter $\nu$ is obtained. The tail index is $\nu\nu/2$. The mean and mode are both at zero. Skewness can be introduced by method in Fernandez and Steel (1998), such that the p.d.f. becomes

$$p(x) = \frac{2}{\gamma + \frac{1}{\gamma}} \frac{v}{2^{\nu/2}} \frac{1}{B(\nu/2, 1/2)} \left( 1 + \frac{1}{\nu} \frac{|x|^\nu}{\nu \gamma |\text{sgn}(x)|} \right)^{-(1/2 + \frac{1}{\nu})},$$

(28)

for positive $\gamma$. The choice $\gamma = 1$ gives symmetry, whereas $\gamma < 1$ implies negative skew. If $\gamma \neq 1$, then $E(\varepsilon_t) \neq 0$. In Harvey and Sucarrat (2014) it is shown that $E(\varepsilon_t)$ is

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

such that $E(\varepsilon_t)$ depends on three shape parameters. If $\gamma = 1$, then $\mu_\varepsilon = 0$. In (27), the scale is multiplied by the zero-mean i.i.d. random variable $\varepsilon_t - \mu_\varepsilon$. From the p.d.f. of $\varepsilon_t$, it follows that the p.d.f. of $y_t$ equals

$$f(y_t) = \exp(-\lambda_{t|t-1}) \, p\left( \frac{y_t - \mu_{t|t-1} - \rho y_{t-1}}{\exp(\lambda_{t|t-1})} + \mu_\varepsilon \right) = \exp(-\lambda_{t|t-1}) \, p\left( \varepsilon_t \right),$$

(29)

where $p$ is given by (28) and $\varepsilon_t = (y_t - \rho y_{t-1} - \mu_{t|t-1}) \exp(-\lambda_{t|t-1}) + \mu_\varepsilon$ is simply the measurement equation (27) re-written. In dynamic conditional score (DCS) models, the dynamics of a time-varying parameter is driven by an innovation that is proportional to the conditional score. The score $u_{t}$ is defined as the derivative of the log-likelihood with respect to the time-varying parameter, i.e.:

$$u_{\lambda t} := \frac{\partial \ln f(y_t)}{\partial \lambda_{t|t-1}} = \left( \frac{\nu}{2} + 1 \right) \left( 1 - \frac{\mu_\varepsilon}{\varepsilon_t} \right) \frac{|\varepsilon_t|^\nu}{|\varepsilon_t|^\nu + \nu \gamma |\text{sgn}(\varepsilon_t)|} - 1,$$

$$u_{\mu t} := \frac{\partial \ln f(y_t)}{\partial \mu_{t|t-1}} = \left( \frac{\nu}{2} + 1 \right) \left( 1 - \frac{\mu_\varepsilon}{\varepsilon_t} \right) \frac{|\varepsilon_t|^\nu}{|\varepsilon_t|^\nu + \nu \gamma |\text{sgn}(\varepsilon_t)|} \exp(\lambda_{t|t-1}),$$

(30)

where $f$ is the predictive density (29). The scores can be written as functions of $y_t$ or of $\varepsilon_t$; the latter leads to more compact expressions. For Gaussian residuals
\((\gamma = 1, \mu = 0, \nu = 2 \text{ and } \nu \to \infty)\), the scores are
\[
\begin{align*}
  u_{\lambda,t} &= \lim_{\nu \to \infty} \frac{\nu + 1}{\nu \varepsilon_t^2 + \nu} \varepsilon_t^2 - 1 = \varepsilon_t^2 - 1, \\
  u_{\mu,t} &= \lim_{\nu \to \infty} \frac{\nu + 1}{\nu \varepsilon_t^2 + \nu} \varepsilon_t = \frac{\varepsilon_t}{\exp(\lambda_{t|t-1})}.
\end{align*}
\]

The dynamic equation for the location and the scale involve the scores and are as follows:
\[
\begin{align*}
  \lambda_{t+1|t} &= \omega_{\lambda} (1 - \varphi_{\lambda}) + \varphi_{\lambda} \lambda_{t|t-1} + \kappa_{\lambda} u_{\lambda,t} + \kappa^{*} \text{sgn}(-\varepsilon_t) (u_{\lambda,t} + 1), \\
  \mu_{t+1|t} &= \omega_{\mu} (1 - \varphi_{\mu}) + \varphi_{\mu} \mu_{t|t-1} + \kappa_{\mu} u_{\mu,t},
\end{align*}
\]
where, in both equations, \(\omega\) denotes the long-term average, \(\varphi < 1\) denotes the persistence, and \(\kappa\) and \(\kappa^{*}\) govern the sensitivity with respect to the score. In volatility modeling, it is well-known that positive returns and negative returns may have different impact on the volatility. The asymmetric response of volatility to returns has become known as the ‘leverage effect’. The leverage parameter \(\kappa^{*} > 0\) implies that a fall in the asset price increases uncertainty by a larger amount than a corresponding rise in the asset price would have done. The leverage parameter \(\kappa^{*}\) is multiplied by \((u_{\lambda,t} + 1)\), such that \(\lambda_{t+1|t}\) is a continuous function of \(y_t\) (compare with (30)).

7 ARCH-M in SP500

In this section we fit a variety of EGARCH-M models to the S&P500 stock market index. The return on day \(t\) is defined as the continuously compounded percentage \(100 \times [\log(I_t) - \log(I_{t-1})]\), where \(I_t\) is the adjusted closing price of the index.

The data are drawn from Yahoo Finance (yahoo.finance.com), spanning a 60-year period from 4 January 1954 to 31 December 2013; a total of 15,104 observations. No adjustment was made for dividends, as the consensus (e.g. French et al. 1987, Poon & Taylor 1992 and Koopman & Uspensky 2002) seems to be that they have little or no effect on the estimates.

The risk-free return is proxied by the secondary market for 3-month Treasury bills\(^{12}\). The risk-free return, \(r_{f,t}\), is defined as the continuously compounded interest rate per business day.

\(^{12}\)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. To calculate the (continuously compounded) interest rate per business day, we assume 250 business days per year. Thus the daily interest rate in percentages is \(r_{f,t} = 100/250 \times \log(1+\text{yearly interest rate in percentages}/100)\).
The daily excess return, $y_t$, is defined as the daily S&P500 return minus the daily risk-free return, i.e. $y_t = 100 \times [\log(I_t) - \log(I_{t-1})] - r_{f,t}$. Our analysis focuses on the time series $y_t$.

Four basic models were estimated (see Appendix 2 for details). In all cases the conditional distribution was assumed to be skew generalized Student-t. In some cases we set $\gamma = 1$ and $\upsilon = 2$, such that the conditional distribution is simply Student-t.

**Model 1** has a single component and is specified as in (3) with dynamic equation (9) driven by a score as in (25). All other models have two components, as in (10). They differ in the way in which volatility affects returns. In **Model 2**, the EGARCH-M effect is driven by the overall volatility, that is $\omega$ and the two components, with identifiability enforced by the restriction $\phi_1 > \phi_2$. In **Model 3**, it is driven instead by the sum of $\omega$ and the long-run component, $\lambda_{1,t|t-1}$. The fact that only one component is associated with the EGARCH-M effect is enough to ensure identifiability; in fact unrestricted estimation always yields a model in which the EGARCH-M effect is associated with the more persistent component. **Model 4** permits two separate EGARCH-M effects; one associated with the long-run and one with the short-run component of volatility.

In **Models 2b, 3b and 4b**, the dynamic equation for the first component is adjusted to include a damped cubic spline as in equation (12). In **Models 1c, 2c, 3c and 4c**, the risk premium parameter $\alpha$ is set to zero in the score and appears only in the predictive density. **Models 2bc, 3bc and 4bc** both include a spline and have the risk premium parameter set to zero in the score.

Tables 2 and 3 in Appendix 3 show the results. For all models reported in Table 2, $\varepsilon_t$ is Student-t distributed. For all models in Table 3, $\varepsilon_t$ is skew generalized t distributed.

Model 1 in Table 2 shows that leverage is necessary to find a positive ARCH-M effect. When leverage is not explicitly included in the model, $\alpha$ takes the role of leverage and is estimated to be negative. This can be explained by the appearance of $\alpha$ in the score, where it acts like a leverage term if $\alpha$ is negative. To separate the effect of leverage from the effect of the risk premium, we may set $\alpha$ to zero in the score (Model 1c). Then the estimate of $\alpha$ is positive even when the leverage term as such is excluded from the model. Leverage is crucial for a good fit, and Model 1c outperforms the GARCH model when it is included.

CAPM suggests that excess returns are possible only at the cost of extra risk. This suggests that $\mu$ must be set to 0, thus eliminating a ‘free lunch’. We explicitly test for this by incorporating both $\mu$ and $\alpha$ in Model 1. Model 1 with $\alpha$ (and no $\mu$) and Model 1 with $\mu$ (and no $\alpha$) perform similarly in terms of fit. This suggests that the ARCH-M effect is fairly constant or moving slowly over time, thus motivating a two-component model.
Model 2 is a marked improvement in log-likelihood over Model 1; although the EGARCH-M effect is still associated with the overall volatility, the volatility itself is modelled using two components. However, the estimate of $\alpha$ remains insignificant unless $\mu$ is set to 0.

In Models 2 and 3 we find $\kappa^*_2 > \kappa_2$, implying that short-run volatility decreases after a good day (see also Figure 3 in Appendix 2). Herein lies the added value of our EGARCH formulation, which allows us to expressly identify the calming effect of a good day\textsuperscript{13}. In contrast, the very setup of the GARCH model precludes such a finding; its dynamic equation uses the indicator rather than the sign. At most, a GARCH model can find that volatility does not go up because the coefficient of the indicator, $\kappa_2$, is restricted to be non-negative to prevent variance becoming negative. If the constraint on $\kappa_2$ is not imposed, it is estimated to be negative, mimicking the calming effect we find with our EGARCH model.

Model 3 clearly outperforms the two-component GARCH model in terms of fit, but the estimates for $\alpha$ are insignificant unless $\mu$ is dropped.

In Table 3, the estimates for $\alpha_1$ are positive and significant, irrespective of which model is used: Models 3 and 4 both find a slowly varying risk premium. Specifically, Model 3 rejects a constant risk premium ($\mu$ is insignificant), while Model 4 rejects a quickly varying risk premium ($\alpha_2$ is insignificant). In fact, the best BIC is obtained by setting both $\mu$ and $\alpha_2$ to zero, as in Model 3c (without $\mu$).

To corroborate this finding, we also consider a model that simultaneously incorporates all three parameters $\mu$, $\alpha_1$ and $\alpha_2$. All three are now significant: specifically, short-term volatility drives prices down (i.e. $\alpha_2 < 0$) and long-term volatility drives prices up ($\alpha_1 > 0$). The former can be tentatively ascribed to the ‘news effect’ as described in Chou (1988) and Schwert (1989), where unanticipated (i.e. short-term) volatility drives prices down as risk-averse investors flee. The latter comes into play once the higher volatility has been incorporated into the (lower) price, after which the now anticipated (i.e. long-term) volatility is expected to drive prices up. Investors willing to bear the risk can then reap the reward of a risk premium. This positive correlation between risk and return may partly hinge on the preceding negative one, in that prices must come down before they are expected to go up again. Given the subtle interplay between the three parameters, Model 4 should be interpreted with caution. Still, it makes intuitive sense in that both the news effect and the risk premium point in the expected direction and, moreover, Model 4 with all three parameters has the best AIC of all models. Both the news effect and the risk premium have been discussed at length in the literature; to date, however, it has not yet been found that both could be true.

\textsuperscript{13}Of course, this advantage of EGARCH applies only to the short-run component; the long-run component of volatility is affected by big swings in either direction.
separated only in time.

A comparison of Table 3 (with the skew generalized t distribution) and Table 2 (without it) suggests that the skew generalized t distribution is crucial to establishing the existence of a risk premium. In Table 3, the parameter estimates for $\gamma$ and $\upsilon$ always differ significantly from 1 and 2, respectively. The need for negative skewness in models for stock market returns is well known; the need for a generalized Student-t is less so. The latter is especially valuable in combination with skewness, however, because skewness pushes probability into the tail. In doing so it lowers the central peak, but the flexibility of the generalized Student-t is able to counteract this undesirable tendency.

The introduction of negative skew implies a negative mean for the residuals $\varepsilon_t$. Without $\mu$ or $\alpha$, therefore, the stock market would go down on average. The negative mean of $\varepsilon_t$ allows both $\mu$ and $\alpha$ to take higher absolute values. This makes it easier to distinguish between them, and $\alpha_1$ is always favoured over $\mu$ in this setting.

What do our findings imply about annual excess returns? In Model 3c without $\mu$ and a generalized t distribution, the expected daily excess return is

$$E \left[ \varepsilon_t \exp(\lambda_{lt-1}) + \alpha_1 \exp(\omega + \lambda_{1,lt-1}) \right] = 0.0104\%.$$  

The first term contributes to the expectation, since $\varepsilon_t$ is negatively skewed and has negative expectation. As daily returns are continuously compounded, the expected annual excess return is

$$100 \left[ \exp \left( 250 \times \frac{0.0104}{100} \right) - 1 \right] = 2.63\%$$

where we have assumed 250 business days. Historically, excess returns on the S&P have been $\sim 2.8\% \text{14}$.

Skewness resolves difficulties noted in Guo and Neely (2008).??

This section has shown that the risk premium is positive, significant and slowly varying when estimated with a two-component DCS EGARCH model, (26) and (10), in which leverage enters differently into long and short run components. We also find that skewness must be complemented by the generalized t distribution for the best results, and that with the EGARCH model we can, and do, find a calming effect of positive returns on short-run volatility. Finally there is evidence that high

\text{14}One dollar, invested in the S&P500 on 1 Jan 1954, would have grown to $74.5 over the course of the next 60 years, representing an annual return of $\sim 7.5\%$. The same dollar invested in 3-month U.S. Treasury bills would have accumulated only to a fifth of that, i.e. to $15.4$, equivalent to an annual return of $\sim 4.7\%$. The average annual excess return over the period 1954-2013 has thus been $\sim 2.8\%$
short-term volatility may cause returns to go down. These results demonstrate the versatility and viability of our EGARCH-M formulation. However, they are obtained from a very long sample for just one stock index, albeit a very important one. From the practitioners point of view the next question to ask is whether the substantive findings carry over to shorter time periods and other indices.

8 Conclusions and extensions

The fact that the asymptotic theory can be established for a basic version of the DCS EGARCH-M model provides re-assurance on the viability of the form adopted, that is modeling the logarithm of the scale by a linear function of the conditional scores of past observations and letting the risk premium effect depend on the scale (or standard deviation). More general models constructed along these lines are likely to be equally effective.

The simulation evidence shows that the asymptotic theory gives a good approximation in moderate size samples and the applications indicate that our DCS model outperforms other specifications in terms of goodness of fit.

Leverage effects play an important role in volatility models and the use of two components, coupled with the EGARCH formulation, reveals some interesting, and perhaps surprising, aspects of leverage: positive returns seem to actively reduce short-term volatility. This suggests that short-term volatility can be thought of as ‘bad sentiment’, which drives prices down, increases after negative shocks and is tempered by positive ones. Long-term volatility, on the other hand, increases with large swings in either direction. In other words, negative shocks have an upward effect on both short- and long-term volatility, whereas positive shocks reduce short-term but not long-term volatility.

Our results allow us to reject both a constant and a rapidly varying risk premium; instead, we have demonstrated that the risk premium is associated with the slowly varying component of volatility. The inclusion of skewness seems to be key to this result, and is ideally complemented with the generalized Student-t distribution.
9 Appendix 1: Asymptotics

When all other parameters are known, the information matrix for $\psi = (\phi, \kappa, \omega)'$ for a single observation in a DCS model is given by $I(\psi) = I.D(\psi)$, where $I$ is the information quantity for a single observation and

$$
D(\psi) = D \begin{pmatrix}
\kappa \\
\phi \\
\omega
\end{pmatrix}
= \frac{1}{1-b} \begin{bmatrix}
A & D & E \\
D & B & F \\
E & F & C
\end{bmatrix}, \quad b < 1, \quad (31)
$$

with

$$
A = \sigma_u^2, \quad B = \frac{\kappa^2 \sigma_u^2 (1 + a\phi)}{(1 - \phi^2)(1 - a\phi)}, \quad C = \frac{(1 - \phi)^2(1 + a)}{1 - a},
$$

$$
D = \frac{a\kappa \sigma_u^2}{1 - a\phi}, \quad E = c(1 - \phi)/(1 - a) \quad \text{and} \quad F = \frac{ack(1 - \phi)}{(1 - a)(1 - a\phi)}.
$$

The extension to situations in which there are fixed parameters in addition to those in the dynamic equation for $\lambda_{t|t-1}$ is discussed in Harvey (2013, p 47).
10 Appendix 2: Estimated models

In all models, $y_t$ is the daily excess return. **One-component GJR GARCH-M:**

$$
\varepsilon_t \sim \text{Student-t}(0, \sqrt{\frac{\nu - 2}{\nu}}, \nu), \text{ such that } E\varepsilon_t^2 = 1,
$$

$$
y_t = \mu + \alpha \lambda_{t|t-1} + \varepsilon_t \lambda_{t|t-1},
$$

$$
\lambda_{t+1|t}^2 = \omega + \varphi \lambda_{t|t-1}^2 + (\kappa + \kappa^* I(\varepsilon_t < 0)) \lambda_{t|t-1}^2 \varepsilon_t^2.
$$

**Two-component GJR GARCH-M:**

$$
\varepsilon_t \sim \text{Student-t}(0, \sqrt{\frac{\nu - 2}{\nu}}, \nu), \text{ such that } E\varepsilon_t^2 = 1,
$$

$$
y_t = \mu + \alpha \lambda_{1,t|t-1} + \varepsilon_t \lambda_{t|t-1},
$$

$$
\lambda_{t+1|t}^2 = \lambda_{1,t+1|t}^2 + \lambda_{2,t+1|t}^2,
$$

$$
\lambda_{1,t+1|t}^2 = \omega + \varphi_1 \lambda_{1,t|t-1}^2 + (\kappa_1 + \kappa_1^* I(\varepsilon_t < 0)) \lambda_{t|t-1}^2 \varepsilon_t^2,
$$

$$
\lambda_{2,t+1|t}^2 = \varphi_2 \lambda_{2,t|t-1}^2 + (\kappa_2 + \kappa_2^* I(\varepsilon_t < 0)) \lambda_{t|t-1}^2 \varepsilon_t^2.
$$

**Model 1A:**

$$
\varepsilon_t \sim \text{skew Gen-Student-t}(\nu, \upsilon, \gamma),
$$

$$
y_t = \rho y_{t-1} + \mu + (\varepsilon_t - \mu_t) \exp(\lambda_t),
$$

$$
\lambda_{t+1|t} = \omega (1 - \varphi) + \varphi \lambda_{t|t-1} + \kappa u_t + \kappa^* \text{sgn}(\varepsilon_t) (u_t + 1),
$$

$$
u_t = \left(1 - \frac{\mu_t}{\varepsilon_t^2}ight) \left(\frac{\nu \upsilon}{2} + 1\right) \frac{|\varepsilon_t|^\nu}{|\varepsilon_t|^\nu + \nu \gamma^u \text{sgn}(\varepsilon_t)} - 1.
$$

**Model 1B:**

$$
\varepsilon_t \sim \text{skew Gen-Student-t}(\nu, \upsilon, \gamma),
$$

$$
y_t = \rho y_{t-1} + \mu_{t|t-1} + (\varepsilon_t - \mu_t) \exp(\lambda_t),
$$

$$
\lambda_{t+1|t} = \omega (1 - \varphi) + \varphi \lambda_{t|t-1} + \kappa u_t + \kappa^* \text{sgn}(-\varepsilon_t) (u_t + 1),
$$

$$
\mu_{t+1|t} = \omega_{\mu} (1 - \varphi_{\mu}) + \varphi_{\mu} \mu_{t|t-1} + \kappa_{\mu} u_{\mu,t},
$$

$$
u_{\lambda,t} = \left(\frac{\nu \upsilon}{2} + 1\right) \left(1 - \frac{\mu_t}{\varepsilon_t^2}\right) \frac{|\varepsilon_t|^\nu}{|\varepsilon_t|^\nu + \nu \gamma^u \text{sgn}(\varepsilon_t)} - 1,
$$

$$
u_{\mu,t} = \left(\frac{\nu \upsilon}{2} + 1\right) \frac{|\varepsilon_t|^\nu}{|\varepsilon_t|^\nu + \nu \gamma^u \text{sgn}(\varepsilon_t)} \exp(\lambda_{t|t-1}).
Model 1C:

\[ \varepsilon_t \sim \text{skew Gen-Student-t}(\nu, v, \gamma), \]
\[ y_t = \rho y_{t-1} + \mu + \alpha_k e^{\lambda_{t-k|t-k-1} + (\varepsilon_t - \mu_e) \exp(\lambda_{t|t-1})}, \]
\[ \lambda_{t+1|t} = \omega (1 - \varphi) + \varphi \lambda_{t|t-1} + \kappa u_t + \kappa^* \text{sgn}(-\varepsilon_t) (u_t + 1), \]
\[ u_t = \left( \frac{\nu v}{2} + 1 \right) \left( 1 - \frac{\mu_e}{\varepsilon_t} \right) \frac{|\varepsilon_t|^v}{|\varepsilon_t|^v + \nu \gamma^v \text{sgn}(\varepsilon_t) - 1}, \]

where \( \lambda_{t-k|t-k-1} = \lambda_{\max(t-k,1)|\max(t-k-1,0)} \), such that \( \lambda_{t-k|t-k-1} \) is defined for all \( t \) and \( k \), and can be applied to the same data set as other models.

Model 2A:

\[ \varepsilon_t \sim \text{skew Gen-Student-t}(\nu, v, \gamma), \]
\[ y_t = \rho y_{t-1} + \mu + (\varepsilon_t - \mu_e) \exp(\lambda_{t|t-1}), \]
\[ \lambda_{t|t-1} = \omega + \lambda_{1,t|t-1} + \lambda_{2,t|t-1}, \]
\[ \lambda_{i,t+1|t} = \varphi_i \lambda_{i,t|t-1} + \kappa_i u_t + \kappa^*_i \text{sgn}(\varepsilon_t) (u_t + 1), \quad i = 1, 2 \]
\[ u_t = \left( \frac{\nu v}{2} + 1 \right) \left( 1 - \frac{\mu_e}{\varepsilon_t} \right) \frac{|\varepsilon_t|^v}{|\varepsilon_t|^v + \nu \gamma^v \text{sgn}(\varepsilon_t) - 1}. \]

Model 2B:

\[ \varepsilon_t \sim \text{skew Gen-Student-t}(\nu, v, \gamma), \]
\[ y_t = \rho y_{t-1} + \mu_{t|t-1} + (\varepsilon_t - \mu_e) \exp(\lambda_{t|t-1}), \]
\[ \lambda_{t|t-1} = \omega + \lambda_{1,t|t-1} + \lambda_{2,t|t-1}, \]
\[ \lambda_{i,t+1|t} = \varphi_i \lambda_{i,t|t-1} + \kappa_{i,t} u_{\lambda,t} + \kappa^*_{i,t} \text{sgn}(\varepsilon_t) (u_{\mu,t} + 1), \quad i = 1, 2 \]
\[ \mu_{t+1|t} = \omega_{\mu} (1 - \varphi_{\mu}) + \varphi_{\mu} \mu_{t|t-1} + \kappa_{\mu} u_{\mu,t}, \]
\[ u_{\lambda,t} = \left( \frac{\nu v}{2} + 1 \right) \left( 1 - \frac{\mu_e}{\varepsilon_t} \right) \frac{|\varepsilon_t|^v}{|\varepsilon_t|^v + \nu \gamma^v \text{sgn}(\varepsilon_t) - 1}, \]
\[ u_{\mu,t} = \left( \frac{\nu v}{2} + 1 \right) \frac{|\varepsilon_t|^v}{|\varepsilon_t|^v + \nu \gamma^v \text{sgn}(\varepsilon_t) \exp(\lambda_{t|t-1}). \]
Model 2C:

\[ \varepsilon_t \sim \text{skew Gen-Student-t}(\nu, \upsilon, \gamma), \]

\[ y_t = \rho y_{t-1} + \mu + \alpha_1 \exp(\omega + \lambda_{1,t|t-1}) + \alpha_2 \exp(\lambda_{2,t|t-1}) + (\varepsilon_t - \mu_e) \exp(\lambda_{t|t-1}), \]

\[ \lambda_{t|t-1} = \omega + \lambda_{1,t|t-1} + \lambda_{2,t|t-1}, \]

\[ \lambda_{i,t+1|t} = \varphi_i \lambda_{i,t|t-1} + \kappa_i u_{i,t} + \kappa^*_i \text{sgn}(\varepsilon_t) (u_{i,t} + 1), \quad i = 1, 2 \]

\[ u_{1,t} = \left( 1 - \frac{\mu_e}{\varepsilon_t} + \frac{\alpha_1 \exp(-\varepsilon_{2,t|t-1})}{\varepsilon_t} \right) \left( \frac{\nu \upsilon}{2} + 1 \right) \frac{|\varepsilon_t|^\upsilon}{|\varepsilon_t|^\upsilon + \nu \gamma^u \text{sgn}(\varepsilon_t)} - 1, \]

\[ u_{2,t} = \left( 1 - \frac{\mu_e}{\varepsilon_t} + \frac{\alpha_2 \exp(-\omega - \lambda_{1,t|t-1})}{\varepsilon_t} \right) \left( \frac{\nu \upsilon}{2} + 1 \right) \frac{|\varepsilon_t|^\upsilon}{|\varepsilon_t|^\upsilon + \nu \gamma^u \text{sgn}(\varepsilon_t)} - 1. \]

When \( \gamma \) and \( \upsilon \) are not estimated, they are set to 1 and 2, respectively. In models with a star, e.g. Model 2A*, the dynamic equation for the first component is adjusted to include a damped cubic spline \( \lambda^\beta \) as follows:

\[ \lambda_{1,t+1|t} = \varphi_1 \lambda_{1,t|t-1} + \lambda_{1,t|t-1}^\beta + \kappa_1 u_{1,t} + \kappa_1^* \text{sgn}(-\varepsilon_t) (u_{1,t} + 1), \]

\[ \lambda_{1,t+1|t}^\beta = \varphi_1 \lambda_{1,t|t-1}^\beta + \frac{\kappa_2}{2 - \kappa_1} u_{1,t}, \]

\[ 25 \]
## 11 Appendix 3: Parameter estimates

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>15,104 obs for e</td>
<td>k_2 lev_2 phi_2</td>
<td>k_1 lev_1 phi_1</td>
<td>om.</td>
<td>mu</td>
<td>a_1 a_2</td>
<td>nu upsil gam.</td>
<td>LogL</td>
<td>AIC</td>
</tr>
<tr>
<td>GARCH</td>
<td>t</td>
<td>0.023 0.093 0.921</td>
<td>0.008</td>
<td>0.022</td>
<td>0.009</td>
<td>7.55</td>
<td>-17866.2</td>
<td>2.3667</td>
</tr>
<tr>
<td>Model 1</td>
<td>without leverage</td>
<td>t</td>
<td>0.043</td>
<td>0.990</td>
<td>-0.388</td>
<td>0.111</td>
<td>-0.128</td>
<td>7.28</td>
</tr>
<tr>
<td>Model 1c</td>
<td>without leverage</td>
<td>t</td>
<td>0.042</td>
<td>0.991</td>
<td>-0.378</td>
<td>0.015</td>
<td>0.041</td>
<td>7.38</td>
</tr>
<tr>
<td>Model 1c</td>
<td>without alpha</td>
<td>t</td>
<td>0.036</td>
<td>0.023</td>
<td>0.989</td>
<td>-0.456</td>
<td>0.025</td>
<td>0.001</td>
</tr>
<tr>
<td>Model 1c</td>
<td>without mu</td>
<td>t</td>
<td>0.036</td>
<td>0.023</td>
<td>0.989</td>
<td>-0.456</td>
<td>0.004</td>
<td>0.007</td>
</tr>
<tr>
<td>Model 2c</td>
<td>t</td>
<td>0.007 0.035 0.903</td>
<td>0.024</td>
<td>0.007</td>
<td>0.997</td>
<td>-0.498</td>
<td>0.040</td>
<td>-0.024</td>
</tr>
<tr>
<td>GARCH</td>
<td>t</td>
<td>0.000 0.146 0.849</td>
<td>0.011</td>
<td>0.000</td>
<td>0.975</td>
<td>0.002</td>
<td>0.015</td>
<td>0.023</td>
</tr>
<tr>
<td>GARCH</td>
<td>unconstrained</td>
<td>t</td>
<td>-0.019 0.167 0.877</td>
<td>0.022</td>
<td>-0.022</td>
<td>0.974</td>
<td>0.002</td>
<td>0.010</td>
</tr>
<tr>
<td>Model 3</td>
<td>t</td>
<td>0.007 0.034 0.903</td>
<td>0.023</td>
<td>0.006</td>
<td>0.997</td>
<td>-0.515</td>
<td>0.036</td>
<td>-0.016</td>
</tr>
<tr>
<td>Model 3b</td>
<td>t</td>
<td>0.013 0.033 0.910</td>
<td>0.019</td>
<td>0.007</td>
<td>0.992</td>
<td>-0.463</td>
<td>0.034</td>
<td>-0.013</td>
</tr>
<tr>
<td>Model 3c</td>
<td>t</td>
<td>0.007 0.034 0.904</td>
<td>0.023</td>
<td>0.006</td>
<td>0.996</td>
<td>-0.497</td>
<td>0.028</td>
<td>-0.002</td>
</tr>
<tr>
<td>Model 3bc</td>
<td>t</td>
<td>0.013 0.033 0.911</td>
<td>0.019</td>
<td>0.007</td>
<td>0.992</td>
<td>-0.456</td>
<td>0.027</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Table 2: Parameter estimates for one and two component GARCH-M models, and corresponding DCS EGARCH-M models. All models have \( \varepsilon_t \) distributed as Student-t.
<table>
<thead>
<tr>
<th>Model</th>
<th>Distribution</th>
<th>Fit</th>
<th>ARCH at 20 lags</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 3</td>
<td>Gen-t</td>
<td>0.010 0.034 0.917</td>
<td>1954-2013 Distr. Vol.</td>
</tr>
<tr>
<td>Model 3b</td>
<td>Gen-t</td>
<td>0.017 0.032 0.925</td>
<td>0.011 0.033 0.915</td>
</tr>
<tr>
<td>Model 3c</td>
<td>Gen-t</td>
<td>0.004 0.003 0.030</td>
<td>0.003 0.003 0.010</td>
</tr>
<tr>
<td>Model 3bc</td>
<td>Gen-t</td>
<td>0.017 0.031 0.924</td>
<td>0.010 0.034 0.917</td>
</tr>
<tr>
<td>Model 4</td>
<td>Gen-t</td>
<td>0.011 0.034 0.920</td>
<td>0.012 0.034 0.922</td>
</tr>
<tr>
<td>Model 4b</td>
<td>Gen-t</td>
<td>0.015 0.027 0.900</td>
<td>0.006 0.032 0.899</td>
</tr>
</tbody>
</table>

**Table 3:** Parameter estimates for DCS EGARCH-M models with $\varepsilon_t$ skew generalized t-distributed.
<table>
<thead>
<tr>
<th>1954-2013</th>
<th>Short-run</th>
<th>Long-run</th>
<th>Vol.</th>
<th>ARCH-M</th>
<th>Generalised-T</th>
<th>Fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>n=15,104</td>
<td>( \kappa_2 ) 0.010 0.034 0.033</td>
<td>( \kappa_3 ) 0.023 0.027 0.009</td>
<td>( \phi_2 ) 0.023 0.007 0.997</td>
<td>( \phi_1 ) 0.023 0.007 0.997</td>
<td>( \omega ) -0.491 0.032 0.888</td>
<td>( \mu \alpha_1 \alpha_2 ) 0.010 1.869 0.939</td>
</tr>
<tr>
<td>Model 3</td>
<td>0.004 0.003 0.003</td>
<td>0.002 0.022 0.000</td>
<td>0.002 0.022 0.000</td>
<td>0.002 0.022 0.000</td>
<td>0.088 0.017 0.035</td>
<td>1.423 0.065 0.01</td>
</tr>
<tr>
<td>Model 3b</td>
<td>0.017 0.032 0.025</td>
<td>0.018 0.007 0.993</td>
<td>0.018 0.007 0.993</td>
<td>0.018 0.007 0.993</td>
<td>-0.485 0.027 0.099</td>
<td>10.75 1.840 0.939</td>
</tr>
<tr>
<td>Model 3c</td>
<td>0.011 0.033 0.015</td>
<td>0.023 0.006 0.997</td>
<td>0.023 0.006 0.997</td>
<td>0.023 0.006 0.997</td>
<td>-0.610 0.025 0.107</td>
<td>10.62 1.849 0.937</td>
</tr>
<tr>
<td>Model 3bc</td>
<td>0.017 0.031 0.024</td>
<td>0.018 0.007 0.993</td>
<td>0.018 0.007 0.993</td>
<td>0.018 0.007 0.993</td>
<td>-0.548 0.019 0.118</td>
<td>10.85 1.840 0.936</td>
</tr>
<tr>
<td>Model 3</td>
<td>0.012 0.034 0.022</td>
<td>0.022 0.007 0.996</td>
<td>0.022 0.007 0.996</td>
<td>0.022 0.007 0.996</td>
<td>-0.420 0.145 0.021</td>
<td>10.57 1.848 0.939</td>
</tr>
<tr>
<td>Model 3b</td>
<td>0.018 0.031 0.030</td>
<td>0.017 0.007 0.993</td>
<td>0.017 0.007 0.993</td>
<td>0.017 0.007 0.993</td>
<td>-0.456 0.148 0.021</td>
<td>11.00 1.832 0.938</td>
</tr>
<tr>
<td>Model 3c</td>
<td>0.011 0.033 0.018</td>
<td>0.022 0.006 0.996</td>
<td>0.022 0.006 0.996</td>
<td>0.022 0.006 0.996</td>
<td>-0.590 0.150 0.021</td>
<td>10.72 1.847 0.936</td>
</tr>
<tr>
<td>Model 3bc</td>
<td>0.017 0.031 0.027</td>
<td>0.018 0.006 0.993</td>
<td>0.018 0.006 0.993</td>
<td>0.018 0.006 0.993</td>
<td>-0.543 0.151 0.021</td>
<td>10.93 1.838 0.936</td>
</tr>
<tr>
<td>Model 4</td>
<td>0.011 0.034 0.020</td>
<td>0.022 0.007 0.997</td>
<td>0.022 0.007 0.997</td>
<td>0.022 0.007 0.997</td>
<td>-0.439 0.126 0.011</td>
<td>10.54 1.849 0.938</td>
</tr>
<tr>
<td>Model 4b</td>
<td>0.018 0.032 0.029</td>
<td>0.018 0.007 0.993</td>
<td>0.018 0.007 0.993</td>
<td>0.018 0.007 0.993</td>
<td>-0.458 0.142 0.003</td>
<td>11.00 1.832 0.938</td>
</tr>
<tr>
<td>Model 4c</td>
<td>0.011 0.033 0.018</td>
<td>0.022 0.006 0.997</td>
<td>0.022 0.006 0.997</td>
<td>0.022 0.006 0.997</td>
<td>-0.593 0.143 0.004</td>
<td>10.71 1.847 0.936</td>
</tr>
<tr>
<td>Model 4bc</td>
<td>0.017 0.031 0.027</td>
<td>0.018 0.006 0.993</td>
<td>0.018 0.006 0.993</td>
<td>0.018 0.006 0.993</td>
<td>-0.543 0.154 0.001</td>
<td>10.93 1.838 0.936</td>
</tr>
<tr>
<td>Model 4</td>
<td>0.006 0.032 0.099</td>
<td>0.026 0.011 0.997</td>
<td>0.026 0.011 0.997</td>
<td>0.026 0.011 0.997</td>
<td>-0.589 0.208 0.064</td>
<td>10.70 1.841 0.947</td>
</tr>
<tr>
<td>Model 4b</td>
<td>0.015 0.027 0.000</td>
<td>0.022 0.013 0.991</td>
<td>0.022 0.013 0.991</td>
<td>0.022 0.013 0.991</td>
<td>-0.532 0.211 0.098</td>
<td>11.06 1.826 0.943</td>
</tr>
</tbody>
</table>
Table 2: Estimates for Model 1A

<table>
<thead>
<tr>
<th>Volatility</th>
<th>Mean</th>
<th>Shape</th>
<th>Fit</th>
<th>ARCH at 100 lags</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full sample</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[1954, 2014]</td>
<td>0.038</td>
<td>0.027</td>
<td>0.988</td>
<td>0.392</td>
</tr>
<tr>
<td>Subsample I</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[1954, 1974]</td>
<td>0.050</td>
<td>0.039</td>
<td>0.969</td>
<td>0.626</td>
</tr>
<tr>
<td>Subsample II</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[1974, 1994]</td>
<td>0.026</td>
<td>0.011</td>
<td>0.991</td>
<td>-0.301</td>
</tr>
<tr>
<td>Subsample III</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[1994, 2014]</td>
<td>0.036</td>
<td>0.040</td>
<td>0.985</td>
<td>-0.273</td>
</tr>
</tbody>
</table>

Table 3: Estimates for Model 1B

<table>
<thead>
<tr>
<th>Volatility</th>
<th>Mean</th>
<th>Shape</th>
<th>Fit</th>
<th>ARCH at 100 lags</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full sample</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[1954, 2014]</td>
<td>0.036</td>
<td>0.029</td>
<td>0.989</td>
<td>-0.380</td>
</tr>
<tr>
<td>Subsample I</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[1954, 1974]</td>
<td>0.045</td>
<td>0.040</td>
<td>0.976</td>
<td>-0.610</td>
</tr>
<tr>
<td>Subsample II</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[1974, 1994]</td>
<td>0.026</td>
<td>0.010</td>
<td>0.991</td>
<td>-0.313</td>
</tr>
<tr>
<td>Subsample III</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[1994, 2014]</td>
<td>0.033</td>
<td>0.041</td>
<td>0.987</td>
<td>-0.276</td>
</tr>
</tbody>
</table>
Table 4: Estimates for Model 1C

<table>
<thead>
<tr>
<th>Volatility</th>
<th>Mean</th>
<th>Shape</th>
<th>Fit</th>
<th>ARCH at 100 lags</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\kappa$</td>
<td>$\kappa^*$</td>
<td>$\varphi$</td>
<td>$\omega$</td>
</tr>
<tr>
<td>Full sample</td>
<td>0.038</td>
<td>0.028</td>
<td>0.987</td>
<td>-0.398</td>
</tr>
<tr>
<td>[1954, 2014]</td>
<td>0.002</td>
<td>0.002</td>
<td>0.001</td>
<td>0.050</td>
</tr>
<tr>
<td>Subsample I</td>
<td>0.049</td>
<td>0.040</td>
<td>0.969</td>
<td>-0.624</td>
</tr>
<tr>
<td>[1954, 1974]</td>
<td>0.005</td>
<td>0.004</td>
<td>0.005</td>
<td>0.060</td>
</tr>
<tr>
<td>Subsample II</td>
<td>0.026</td>
<td>0.011</td>
<td>0.991</td>
<td>-0.306</td>
</tr>
<tr>
<td>[1974, 1994]</td>
<td>0.003</td>
<td>0.002</td>
<td>0.002</td>
<td>0.077</td>
</tr>
<tr>
<td>Subsample III</td>
<td>0.035</td>
<td>0.040</td>
<td>0.984</td>
<td>-0.296</td>
</tr>
<tr>
<td>[1994, 2014]</td>
<td>0.004</td>
<td>0.003</td>
<td>0.002</td>
<td>0.091</td>
</tr>
</tbody>
</table>

Table 5: Estimates for Model 1C with $\mu$ dropped

<table>
<thead>
<tr>
<th>Volatility</th>
<th>Mean</th>
<th>Shape</th>
<th>Fit</th>
<th>ARCH at 100 lags</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\kappa$</td>
<td>$\kappa^*$</td>
<td>$\varphi$</td>
<td>$\omega$</td>
</tr>
<tr>
<td>Full sample</td>
<td>0.038</td>
<td>0.027</td>
<td>0.988</td>
<td>-0.420</td>
</tr>
<tr>
<td>[1954, 2014]</td>
<td>0.002</td>
<td>0.002</td>
<td>0.001</td>
<td>0.051</td>
</tr>
<tr>
<td>Subsample I</td>
<td>0.050</td>
<td>0.039</td>
<td>0.969</td>
<td>-0.645</td>
</tr>
<tr>
<td>[1954, 1974]</td>
<td>0.005</td>
<td>0.004</td>
<td>0.005</td>
<td>0.059</td>
</tr>
<tr>
<td>Subsample II</td>
<td>0.026</td>
<td>0.010</td>
<td>0.991</td>
<td>-0.306</td>
</tr>
<tr>
<td>[1974, 1994]</td>
<td>0.003</td>
<td>0.002</td>
<td>0.002</td>
<td>0.078</td>
</tr>
<tr>
<td>Subsample III</td>
<td>0.035</td>
<td>0.040</td>
<td>0.984</td>
<td>-0.294</td>
</tr>
<tr>
<td>[1994, 2014]</td>
<td>0.003</td>
<td>0.003</td>
<td>0.002</td>
<td>0.081</td>
</tr>
</tbody>
</table>
Table 6: Estimates for Model 2A

<table>
<thead>
<tr>
<th></th>
<th>Volatility</th>
<th>Mean</th>
<th>Shape</th>
<th>Fit</th>
<th>ARCH at 100 lags</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\kappa_2$ $\kappa_2^*$ $\varphi_2$</td>
<td>$\kappa_1$ $\kappa_1^*$ $\varphi_1$</td>
<td>$\omega$</td>
<td>$\rho$ $\mu$</td>
<td>$\nu$ $\nu$ $\gamma$</td>
</tr>
<tr>
<td><strong>Full sample</strong></td>
<td>0.013 0.037 0.909</td>
<td>0.022 0.007 0.997</td>
<td>-0.417</td>
<td>0.090 0.012</td>
<td>9.53 1.908 0.948</td>
</tr>
<tr>
<td>[1954, 2014)</td>
<td>0.004 0.003 0.014</td>
<td>0.002 0.002 0.001</td>
<td>0.079</td>
<td>0.008 0.006</td>
<td>1.33 0.066 0.011</td>
</tr>
<tr>
<td><strong>Subsample I</strong></td>
<td>0.032 0.047 0.891</td>
<td>0.017 0.009 0.998</td>
<td>-0.565</td>
<td>0.217 0.011</td>
<td>5.61 2.200 0.940</td>
</tr>
<tr>
<td>[1954, 1974)</td>
<td>0.007 0.005 0.024</td>
<td>0.003 0.003 0.001</td>
<td>0.152</td>
<td>0.014 0.007</td>
<td>1.13 0.151 0.019</td>
</tr>
<tr>
<td><strong>Subsample II</strong></td>
<td>-0.009 0.020 0.825</td>
<td>0.028 0.005 0.992</td>
<td>-0.298</td>
<td>0.088 -0.004</td>
<td>8.52 1.955 0.999</td>
</tr>
<tr>
<td>[1974, 1994)</td>
<td>0.009 0.005 0.095</td>
<td>0.004 0.003 0.002</td>
<td>0.079</td>
<td>0.014 0.011</td>
<td>2.33 0.138 0.020</td>
</tr>
<tr>
<td><strong>Subsample III</strong></td>
<td>-0.049 0.049 0.578</td>
<td>0.040 0.032 0.986</td>
<td>-0.261</td>
<td>-0.032 0.025</td>
<td>15.81 1.704 0.895</td>
</tr>
<tr>
<td>[1994, 2014)</td>
<td>0.011 0.007 0.089</td>
<td>0.004 0.004 0.003</td>
<td>0.100</td>
<td>0.013 0.014</td>
<td>5.39 0.102 0.017</td>
</tr>
</tbody>
</table>
This section contains some graphs for different models.

Figure 1: The bottom graphs shows the value of one dollar invested in the S&P500 versus the value of a dollar invested in 3-month Treasury bills. By continuously compounding the excess returns $y_t$, the top graph is obtained. Its value is equal to the ratio of the S&P500 over the T-bills from the bottom graph.
Estimated impact curves $\kappa_i u_t + \kappa_i^* sgn(-\epsilon_t)(u_t + 1)$

Figure 2: Model 1A, full sample. Estimated news impact curve.
Figure 3: Model 1A, Subsample I. Estimated news impact curve, showing a high degree of asymmetry.
Estimated impact curves $\kappa_i u_t + \kappa^*_i \text{sgn}(-\epsilon_t)(u_t + 1)$

Figure 4: Model 1A, Subsample II. The estimated news impact curve is quite symmetric.
Figure 5: Model 1A, Subsample III. Estimated news impact curve, showing that positive returns have a calming effect on the market.
Figure 6: Model 1B, full sample. The filtered location increases when estimated $\varepsilon_t$ is positive, and decreases when it is negative. It does not do so linearly, however. This graph shows that unexpected excess returns of roughly 3% or more affect the location less than linearly. In effect, very large returns attributed to the heavy tail of $\varepsilon_t$ and are thus down-weighted as far as the updating of the location is concerned.
Figure 7: Model 1B, full sample. Estimated location $\mu_{t|t-1}$, multiplied by 250, tracks roughly the historic 250-day return.
Figure 8: Model 1B, full sample. Estimated location $\mu_{t|t-1}$ and daily excess returns. Clearly, the filtered location is very low in comparison with the noise. Typically, the filtered location is two orders of magnitude lower than the filtered scale.
Contemporaneous relation between location and scale

\[ y = (0.044) + (-0.064) \cdot x + (0.014) \cdot x^2 \]

\[ R^2 = 20\% \]

Figure 8: Model 1B, full sample. Contemporaneous relation between filtered location and filtered scale. Due to the leverage effect, the filtered scale is likely to be high after a series of negative returns, i.e. when the location is low. This leads to a negative contemporaneous correlation between location and scale.
Figure 8: Model 1B, full sample. To investigate the intertemporal relationship between location and scale, we plot the difference in the location over a 100-day horizon. Clearly, the relationship is now positive and $R^2 = 11\%$ is reasonably large.
Estimated risk-return relation over different horizons

Figure 8: Model 1B, full sample. Estimated risk-return relationship for different horizons. For 10-40 days, the risk-return relationship is relatively flat (and explanatory power is low, see next graph). For 80-180 days, the relationship is fairly strong and explanatory power is relatively high (see next graph).
Figure 8: Model 1B, full sample. Explanatory power for the risk-return relationship over different horizons. The $R^2$ increases over time, showing that the risk premium needs about 120 days to materialize, on average.
Figure 8: Model 1C, full sample. Estimates of $\alpha_k$ as a function of $k$. Clearly, the risk premium is significantly positive for $k$ roughly between 60 and 160.
Figure 8: Model 2A, full sample. The cross correlation between \( \exp(\lambda_{1,t|t-1}) \) and \( \epsilon_{t+k} \) shows that high long-term volatility is predictive of positive residuals up to 20 days.
Estimated impact curves $\kappa_i u_t + \kappa_i^* \mathrm{sgn}(\epsilon_t)(u_t + 1)$

![Graph showing impact of $\epsilon_t$ on both components of volatility. The dots correspond to actual realizations of $\epsilon_t$. It is seen that positive returns reduce short-term volatility.](image)

**Figure 3:** This graph shows the impact of $\epsilon_t$ on both components of volatility. The dots correspond to actual realizations of $\epsilon_t$. It is seen that positive returns reduce short-term volatility.
Figure 4: This graph shows the filtered estimate of the scale. The scale multiplies $\epsilon_t$, which is distributed as a skew generalized Student-t. For parameter estimates in Model 3c, a scale of around 4 implies a 95% confidence interval of around $[-9\%, 8\%]$, which is a huge range for daily returns.
Figure 5: The estimated scale is separated as a product of two terms: \( \exp(\omega + \lambda_1, t|t-1) \) and \( \exp(\lambda_2, t|t-1) \). The first is highly persistent, while second mean-reverts relatively quickly to its equilibrium level of 1.
Figure 6: Behavior of the stock market and estimated scale around Black Monday (19 Oct 1987).
S&P500 around Black Monday (19 Oct 1987)

Figure 7: Behavior of estimated long-run and short-run volatility around Black Monday (19 Oct 1987).
Figure 8: Behavior of long-run and short-run volatility around the collapse of Lehman Brothers (15 Sep 2008). Due to a combination of large negative and large positive shocks, most uncertainty is incorporated into the long-run component of volatility.
During the Booming Nineties, the longest period on record of continuous economic expansion in the US, long-term volatility is low and flat.
Figure 10: Rolling annual realized return, and predicted return based on risk premium.
Figure 11: Implied residuals and estimated distribution of the $\epsilon_t$ in Model 3c. This distribution is compared with a (non-skew, non-generalized) Student-t with $\nu = 7.8$, as typically implied by other models.
13 Acknowledgements

Rutger-Jan Lange is 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 in June 2014 and at the Cambridge Finance-Tinbergen Institute meeting in Amsterdam in May 2014. We are grateful to Peter Boswijk, Andre Lucas, Siem-Jan Koopman, Mark Salmon and Robert Taylor for helpful comments.

14 References


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


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