# Bad Environments, Good Environments: A Non-Gaussian Asymmetric Volatility Model[*]

Geert Bekaert

Columbia University and the National Bureau of Economic Research

Eric Engstrom

Board of Governors of the Federal Reserve System[†]

July 22, 2013

**Abstract**

We propose an extension of standard asymmetric volatility models in the generalized autoregressive conditional heteroskedasticity (GARCH) class that admits conditional non-Gaussianities in a tractable fashion. Our "bad environment-good environment" (BEGE) model utilizes two gamma-distributed shocks and generates a conditional shock distribution with time-varying heteroskedasticity, skewness, and kurtosis. The BEGE model features nontrivial news impact curves and closed-form solutions for higher-order moments. In an empirical application to stock returns, the BEGE model outperforms standard asymmetric GARCH and regime-switching models along several dimensions.

# 1    Introduction

Since the seminal work of Engle (1982) and Bollerslev (1986) on volatility clustering, thousands of articles have applied models in the generalized autoregressive conditional heteroskedasticity (GARCH) class to capture volatility clustering in economic and financial time series data. In the basic GARCH (1,1) model, today's conditional variance is a deterministic linear function of the past conditional variance and contemporaneous squared shocks to the process describing the data. Nelson (1991) and Glosten, Jagannathan, and Runkle (1993, GJR henceforth), motivated by empirical work on stock return data, provide important extensions, accommodating asymmetric responses of conditional volatility to negative versus positive shocks. Engle and Ng (1993) compare the response of conditional variance to shocks ("news impact curves") implied by various econometric models and find evidence that the GJR model fits stock return data the best.

The original models in the GARCH class assumed Gaussian innovations, but nonetheless imply non-Gaussian unconditional distributions. Unfortunately, time-varying volatility models with Gaussian innovations generally do not generate sufficient unconditional non-Gaussianity to match some financial asset return data (see, e.g., Poon and Granger, 2003). Additional evidence of conditional non-Gaussianity has come from two corners. First, empirical work by Evans and Wachtel (1993), Hamilton and Susmel (1994), Kim and White (2004), and many others has documented conditional nonlinearities in economic data. Second, in finance, a voluminous literature on the joint properties of option prices and stock returns (see, e.g., Broadie, Chernov, and Johannes (2009)) has also suggested the need for models with time-varying nonlinearities. In principle, one can estimate GARCH models consistently using quasi maximum likelihood (see Lumsdaine, 1996; Lee and Hansen, 1994), not worrying about modeling the non-Gaussianity in the shocks. However, fitting the actual non-Gaussianities in the data can lead to more efficient estimates and is critical if the model is to be used in actual applications (for example, option pricing or risk management) that require an estimate of the conditional distribution. Of course, it is not difficult to introduce

non-Gaussian shocks into a GARCH model, and several authors have done so (see, e.g., Bollerslev (1987) and Hsieh (1989), who used the t-distribution, and Mittnik, Paolella and Rachev (2002), who used shocks with a distribution in the stable Paretian class). However, extant models in this vein generally cannot fit time-varying non-Gaussianities that are evident in the data.

We present an extension of models in the GARCH class that accommodates conditional non-Gaussianity in a tractable fashion, offering simple closed-form expressions for conditional moments. Our "bad environment–good environment" (BEGE) model utilizes two gamma-distributed shocks that together imply a conditional shock distribution with time-varying heteroskedasticity, skewness, and kurtosis. This is accomplished by allowing the shape parameters of the two distributions to vary through time. Hence, our model features nontrivial news impact curves for higher-order moments. We apply the model to stock returns, showing that the model outperforms extant alternatives using a variety of specification tests. Of course, nonlinear models exist outside the GARCH class that may also fit the data quite well. Regime-switching models, in particular, have shown promise in many applications. We therefore also estimate regime-switching models on our stock returns data sample and show that the BEGE model also significantly outperforms models in this class.

The remainder of the article is organized as follows. In section 2, we present the BEGE model, describe how it nests the standard GJR–GARCH model as a special case, and present an alternative regime-switching model. In section 3, we describe the estimation methodology and the specification tests that we conduct. In section 4, we confront several models from the above classes, including the BEGE model, to monthly U.S. stock return data from 1929 through 2010.

# 2 The BEGE–GARCH Model

Before introducing the BEGE model, we begin with a review of the seminal GJR asymmetric GARCH model.

## 2.1 Traditional GJR–GARCH

Consider a time series $r_t$ with constant mean $\mu$. The GJR model assumes that the series follows

$$
\begin{aligned}
r_{t+1} &= \mu + u_{t+1}, \\
u_{t+1} &\sim N\left(0, h_t\right), \\
\text{and } h_t &= h_0 + \rho_h h_{t-1} + \phi^+ u_t^2 I_{u_t \geq 0} + \phi^- u_t^2 \left(1 - I_{u_t \geq 0}\right).
\end{aligned}
\tag{1}
$$

That is, the innovation to returns, $u_{t+1}$, has time-varying conditional variance, $var_t\left(r_{t+1}\right) = h_t$, which is assumed to be a linear function of its own lagged value and squared innovations to returns. One key feature of this model that enables it to better fit many economic time series is the differential response of the conditional variance of shocks following positive versus negative innovations. In stock return and economic activity data, it is typically found that $\phi^- > \phi^+$, so that negative shocks result in more of an increase in variance than do positive shocks.

## 2.2 BEGE GJR–GARCH

The BEGE model that we propose relaxes the assumption of Gaussianity by assuming that the $u_{t+1}$ innovation consists of two components. We assume that $\omega_{p,t+1}$, a good environment shock, and $\omega_{n,t+1}$, a bad environment shock, are drawn from "demeaned" (or "centered") gamma distributions that have a mean equal to zero.[1] The overall innovation is a linear combination of the two component shocks, which are assumed to be conditionally

---

[1]The centered gamma distribution with shape parameter $k$ and scale parameter $\theta$, which we denote $\widetilde{\Gamma}(k, \theta)$, has probability density function, $\phi(x) = \frac{1}{\Gamma(k)\theta^k}(x + k\theta)^{k-1}\exp\left(-\frac{1}{\theta}(x + k\theta)\right)$ for $x > -k\theta$, and with $\Gamma(\cdot)$ representing the gamma function.

independent. The gamma distributions are assumed to have constant scale parameters, but we let their shape parameters vary through time. More precisely, the BEGE framework assumes:

$$
\begin{aligned}
u_{t+1} &= \sigma_p \omega_{p,t+1} - \sigma_n \omega_{n,t+1}, \text{ where} \\
\omega_{p,t+1} &\sim \widetilde{\Gamma}\left(p_t, 1\right), \text{ and} \\
\omega_{n,t+1} &\sim \widetilde{\Gamma}\left(n_t, 1\right), \text{ and}
\end{aligned}
\tag{2}
$$

where $\widetilde{\Gamma}\left(k, \theta\right)$ denotes a centered gamma distribution with shape and scale parameters, $k$ and $\theta$, respectively. Thus, $p_t$ $(n_t)$ is the shape parameter for the good (bad) environment shock. Figure 1 provides a visual representation of the flexibility of the BEGE distribution. Plotted are the 1st and 99th percentiles of two sequences of hypothetical distributions. The blue stars illustrate a series of BEGE distributions for which $p_t$ is fixed at 1.5, but $n_t$ varies from 0.1 to 3.0, which are the values across the horizontal axis. The lower line of blue asterisks shows the 1st percentiles of these distributions, while the upper line of blue stars shows the 99th percentiles. Clearly, increases in $n_t$ have an outsized effect on the lower tail, particularly at low values of $n_t$. The upper tail is relatively insensitive to changes in $n_t$. The green plus symbols show results from the complementary exercise: holding $n_t$ fixed at 1.5 and varying $p_t$ from 0.1 through 3.0. Clearly $p_t$ impacts the upper tail of the distribution much more than it impacts the lower tail. These results highlight the potential benefits of the BEGE distribution. As we will demonstrate, financial data provide evidence that some shocks primarily affect the lower tail of the distribution of returns, but leave the upper tail relatively unchanged (see section 4). This is exactly the kind of effect that BEGE is designed to accommodate.

We model the time variation in the shape parameters in a manner that is analogous to that for $h_t$ in the GJR specification:

$$
\begin{aligned}
p_t &= p_0 + \rho_p p_{t-1} + \phi_p^+ u_t^2 I_{u_t \geq 0} + \phi_p^- u_t^2 \left(1 - I_{u_t \geq 0}\right) \\
\text{and } n_t &= n_0 + \rho_n n_{t-1} + \phi_n^+ u_t^2 I_{u_t \geq 0} + \phi_n^- u_t^2 \left(1 - I_{u_t \geq 0}\right).
\end{aligned}
\tag{3}
$$

The overall conditional variance of $u_{t+1}$ follows trivially from the moment-generating function of the centered gamma distribution[2]

$$h_t \equiv var_t\left(r_{t+1}\right) = \sigma_p^2\ p_t + \sigma_n^2\ n_t, \tag{4}$$

where, with some abuse of notation, $h_t$ now represents the conditional variance under the BEGE model. Higher-order moments also follow in a straightforward manner from the moment-generating function of the gamma distribution. For instance, conditional (unscaled) skewness and excess kurtosis are given by

$$
\begin{aligned}
s_t &\equiv& skw_t\left(r_{t+1}\right) = 2\left(\sigma_p^3\ p_t - \sigma_n^3\ n_t\right) \\
\text{and } k_t &\equiv& kur_t\left(r_{t+1}\right) = 6\left(\sigma_p^4\ p_t + \sigma_n^4\ n_t\right).
\end{aligned}
\tag{5}
$$

The expression for skewess shows that larger values for $p_t$ generate more positive skewness, while larger values of $n_t$ generate more negative skewness. Moments of order higher than four are equally trivial to compute using the moment generating function and are also affine in $p_t$ and $n_t$. It is this affine structure that makes the model both parsimonious and tractable. The model thus allows for positive or negative skewness, and the sign of skewness may vary through time. Excess kurtosis is always positive, but its magnitude varies as well. Note that the key innovation of the model is to let the shape, not the scale, parameters vary through time. Because the shape parameters determine the shape of the distribution, we parsimoniously generate time-variation in all higher order moments simultaneously.

Asymmetric volatility under the BEGE specification can be generated by either the "good volatility" ($p_t$) component or the "bad volatility" ($n_t$) component, or both:

$$
\frac{\partial h_{t+1}}{\partial u_t^2} = \begin{cases} \sigma_p^2\phi_p^+ + \sigma_n^2\phi_n^+ & \text{if } u_t \geq 0 \\ \sigma_p^2\phi_p^- + \sigma_n^2\phi_n^- & \text{otherwise} \end{cases}
\tag{6}
$$

Similar expressions are readily calculated for conditional skewness and kurtosis under

---

[2]The moment-generating function for a random variable, $x$, with the demeaned gamma distribution with shape parameter, $k$, and scale parameter, $\theta$, is given by

$$mgf_x\left(s\right) \equiv E\left[\exp\left(sx\right)\right] = \exp\left(-k\left(\ln\left(1 - \theta s\right) + \theta s\right)\right).$$

Successive differentiation of $mgf_x\left(s\right)$ with respect to $s$, and evaluation at $s = 0$, yields, for the first few moments: $E\left[x\right] = 0, E\left[x^2\right] = \theta^2 k, E\left[x^3\right] = 2\theta^3 k$, and $E\left[x^4\right] - E\left[x^2\right]^2 = 6\theta^4 k$

the BEGE model and

$$\frac{\partial s_{t+1}}{\partial u_t^2} = \begin{cases} 2\left(\sigma_p^3 \phi_p^+ - \sigma_n^3 \phi_n^+\right) & \text{if } u_t \geq 0 \\ 2\left(\sigma_p^3 \phi_p^- - \sigma_n^3 \phi_n^-\right) & \text{otherwise} \end{cases} \tag{7}$$

and

$$\frac{\partial k_{t+1}}{\partial u_t^2} = \begin{cases} 6\left(\sigma_p^4 \phi_p^+ + \sigma_n^4 \phi_n^+\right) & \text{if } u_t \geq 0 \\ 6\left(\sigma_p^4 \phi_p^- + \sigma_n^4 \phi_n^-\right) & \text{otherwise.} \end{cases} \tag{8}$$

Of course, under the traditional Gaussian GJR–BEGE model, conditional skewness and excess kurtosis are fixed at zero. The BEGE model thus allows for richer dynamics for the conditional distribution of the data process, with tractable expressions for conditional moments.

An intuitive feature of the model arises from the fact that for a gamma-distributed random variable, as the shape parameter goes to infinity, the distribution converges to a Gaussian distribution. Therefore, the BEGE model can get arbitrarily close to the traditional GARCH model, even in terms of the conditional Gaussianity of the shocks. More concretely, suppose that the two gamma shocks in the BEGE model are symmetric in their autoregressive behavior and in their responses to the innovation, $u_{t+1}$. That is, suppose, $\rho_h = \rho_p = \rho_n$, $\phi_h^+ = \phi_p^+ = \phi_n^+$, and $\phi_h^- = \phi_p^- = \phi_n^-$. Substituting, we find

$$
\begin{aligned}
h_t &= \left(\sigma_p^2 p_0 + \sigma_n^2 n_0\right) + \rho_h \left(\sigma_p^2 p_{t-1} + \sigma_n^2 n_{t-1}\right) \\
&\quad + \phi_h^+ \left(\sigma_p^2 + \sigma_n^2\right) u_t^2 I_{u_t \geq 0} + \phi_n^- \left(\sigma_p^2 + \sigma_n^2\right) u_t^2 \left(1 - I_{u_t \geq 0}\right) \\
&= h_0 + \rho_h h_{t-1} + \widetilde{\phi}_h^+ u_t^2 I_{u_t \geq 0} + \widetilde{\phi}_h^- u_t^2 \left(1 - I_{u_t \geq 0}\right),
\end{aligned} \tag{9}
$$

with the notations $\widetilde{\phi}_h^+$ and $\widetilde{\phi}_h^+$ implicitly defined. Inspection confirms that this volatility process is isomorphic to that of traditional GJR–GARCH. Moreover, if the constants $p_0$ and $n_0$ are allowed to become arbitrarily large, the gamma distributions will approach their Gaussian limits, and the BEGE–GJR–GARCH process collapses to the traditional Gaussian GJR–GARCH specification.

## 2.3  Regime-switching (RS) models

An alternative approach for generating conditional nonlinearities is the regime-switching model introduced by Hamilton (1989) to model GDP growth dynamics. In this model, an unobserved Markov variable causes the process to switch among two or more regimes, generating a conditional mixture of normals model. As Timmermann (2000) stresses, such a model can generate time-variation in moments up to order five. In the specific two-regime model on which we focus, the process is assumed to follow

$$r_{t+1} = \mu + \mu_{12} J_{12,t+1} + \mu_{21} J_{21,t+1} + \sigma_{s_{t+1}} e_{t+1}, \tag{10}$$

where $s_t$ is a hidden Markov variable. Specifically, $s_t$ can take on the value of 1 or 2. The transition probabilities are defined as $p_{ij} = prob\,(s_{t+1} = j | s_t = i)$, and are assumed to be constant. The innovation, $e_t$, is a standard normal random variable. The $J$ variables are dummy variables specified as

$$J_{12,t+1} = \begin{cases} 1 & \text{if } s_t = 1 \text{ and } s_{t+1} = 2 \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

and similarly for $J_{21,t+1}$. Hence, they determine the mean return conditional on a transition between regimes. These "jump" terms are inspired by Mayfield (2004) and are specifically included for our stock return application. The conditional mean specification allows, for instance, that in the high-variance regime, the conditional mean is potentially higher than in the low-variance regime, because an eventual jump to the low-variance regime is expected, and the return associated with this transition is positive. The reverse could be true for the low-variance regime.

In this model, the conditional distribution of the shock is a mixture of normals with moments that depend on the current regime. For example, the first three uncentered moments of the distribution conditional on being in regime $s_t = 1$ are given by

$$
\begin{aligned}
E_{s_t=1}\left(r_{t+1}\right) &= p_{11}\left(\mu\right) + p_{12}\left(\mu + \mu_{12}\right), \\
E_{s_t=1}\left(r_{t+1}^2\right) &= p_{11}\left(\mu^2 + \sigma_1^2\right) + p_{12}\left(\left(\mu + \mu_{12}\right)^2 + \sigma_2^2\right),
\end{aligned}
$$

$$E_{s_t=1}\left(r_{t+1}^3\right) \quad = \quad p_{11}\left(\mu^3 + 3\mu\sigma_1^2\right) + p_{12}\left(\left(\mu + \mu_{12}\right)^3 + 3\left(\mu + \mu_{12}\right)\sigma_2^2\right), \tag{12}$$

and analytic expressions are also available for higher-order moments, centered moments, and moments conditional on $s_t = 2$. While the mixture-of-normal distributions have a fair amount of flexibility to match moments, the distributions can be multi-modal, a feature for which evidence in most financial data is not readily apparent. A complete description of the RS model specifications that we estimate, as well as our estimation procedure, is provided in Appendix B.

## 3    Estimation and Test Statistics

This section briefly describes the estimation techniques for the models and then introduces the specification tests that we use to assess model performance.

### 3.1    Estimation

We estimate all models using maximum likelihood (ML) and report Huber (1967)–White (1982) standard errors. Alternative estimation methods are, of course, possible. In particular, given that the models have closed-form expressions for conditional moments, a moments-based estimator could also be used.

While conditional ML estimation procedures for Gaussian GARCH and regime-switching models are well established, evaluation of the BEGE likelihood function is slightly more involved. The BEGE distribution is simply a four parameter distribution, and an analytic, if complex, expression is available for the evaluation of its density. This analytic expression for the BEGE density is derived in Appendix A. However, direct evaluation of this expression is often computationally costly, and we use numerical integration to evaluate the density in most of our calculations. The procedure for numerical integration is straightforward: Random variables with the BEGE density take the form $u = \omega_p - \omega_p$ (suppressing time subscripts) where $u$ is the BEGE-distributed variable, and $\omega_p$ and $\omega_p$ are demeaned gamma

distributions. The BEGE density, $f_{BEGE}(u)$, can be represented

$$f_{BEGE}(u) = \int_{\omega_p} f_{BEGE}(u|\omega_p)\, df_{\omega_p}$$

$$= \int_{\omega_p} f_{\omega_n}(\omega_p - u)\, df_{\omega_p}, \tag{13}$$

where $f_{\omega_p}$ and $f_{\omega_n}$ are the densities of $\omega_p$ and $\omega_n$, respectively. Numerical integration is straightforward. In practice, we find that numerical evaluation of the BEGE density is faster and more stable when we employ an alternative representation for the BEGE distribution function:

$$F_{BEGE}(u) = 1 - \int_{\omega_p} F_{\omega_n}(\omega_p - u)\, df_{\omega_p}, \tag{14}$$

where $F_{BEGE}(\cdot)$ denotes the cumulative distribution function of BEGE. That is, we first evaluate the integral above numerically and then use a finite difference approximation of $F_{BEGE}$ to arrive at the BEGE density.[3]

## 3.2  Specification tests

While the ML estimation yields the likelihood value for all models, the standard likelihood ratio test can only be used for the nested models. To assess the relative performance of the models, we report Akaike information criterion (AIC) and Bayesian information criterion (BIC) values for all models. To further parse the performance of the various models with respect to nonlinearities, we employ a battery of additional tests.

### 3.2.1  Likelihood ratio tests for non-nested models

We consider likelihood ratio tests of Vuong (1989), Rivers and Vuong (2002), and Calvet and Fisher (2004). Vuong (1989), develops the test statistic:

$$\sum_{t=1}^{T} ln\left(\frac{f(r_t|R_{t-1}, \hat{\theta}_T)}{g(r_t|R_{t-1}, \hat{\theta}_T)}\right) \equiv \sum_{t=1}^{T} \hat{a}_t, \tag{15}$$

---

[3]Matlab routines that evaluate the BEGE density and distribution functions are available from the authors upon request.

where $R_t = [r_t, r_{t-1}, ..., r_0]$, $f$ and $g$ are probability densities for the models being compared, $\hat{\theta}_T$ is a vector comprised of the estimated parameters for the models, and $\hat{a}_t$ is implicitly defined. The statistic follows $N(0, T\sigma^2)$ under the null hypothesis that the models describe the data equally well. In the basic case of i.i.d. $r_t$ analyzed in Vuong (1989), $\sigma^2$ is just the variance of $a_t$. In the case of non-i.i.d. observations, Calvet and Fisher (2004) argue that the distribution of the test statistic stays the same with $\sigma^2$ now being the heteroskedasticity and autocorrelation- (HAC-) adjusted variance of $a_t$, for example the Newey-West (1987) estimator.

### 3.2.2 Unconditional moments

It is useful to investigate to what extent the various models are able to generate the unconditional moments observed in the data. Because closed-form solutions for unconditional moments are generally not available for the models that we examine, we use a Monte Carlo methodology to execute these tests. In each Monte Carlo sample, a sequence of observations (of the same length as the historical time series) is generated by randomly drawing error terms and using the estimated parameters for each model. Next, the values of variance, skewness, and kurtosis are computed for the generated time series. In the case of the regime-switching models, we first draw the sequence of regimes randomly given the estimated initial distribution of the regimes and the transition probability matrix. Then, conditioning on the regimes, the returns are drawn from the regime distributions. Repeating the procedure 10,000 times yields the null distributions of variance, skewness, and kurtosis under each model.

### 3.2.3 Conditional distribution: quantile shifts

We also examine several conditional quantile tests to determine which models best match the conditional distribution of returns. In particular, we condition on the return in the previous period having been positive or negative. We consider two cases. In the first

case, *positive* and *negative* simply refer to $r_{t-1}$ being greater or (weakly) less than zero, respectively. In the second case *positive* and *negative* are defined as returns that exceed (fall short) of the unconditional mean of the series plus (minus) one standard deviation. Our sample is sufficiently large to measure these conditional quantiles in the data with reasonable accuracy, and we focus on the quantiles corresponding to the 5th, 10th, 50th, 90th, and 95th percentiles. Specifically, we measure the quantiles based on the entire sample, the quantiles for a restricted sample in which the previous month's return is *negative*, and finally for a restricted sample in which the previous month's return is *positive*. We refer to the differences between unconditional quantiles and those for the restricted samples as quantile shifts. To quantitatively investigate how the various estimated models match the observed quantile shifts, we again use the simulation methodology described above. The simulation procedure yields 10,000 random samples of the same length as our data sample, and for each simulated sample we can compute the quantile shifts under the null of the various models. Finally, we calculate the probability of observing the historical quantile shift under each model.

### 3.2.4 Conditional distribution: Engle–Manganelli "hit" test

These tests were developed by Engle and Manganelli (2004) (EM henceforth) to test whether estimates of conditional quantiles under a given model are consistent with the data. EM define the variable $hit_t^{pr}$ as

$$hit_{t+1}^{pr} = I_{r_{t+1} < \widehat{q_t(pr)}} - pr, \tag{16}$$

where $\widehat{q_t(pr)}$ is the model-implied estimate of the conditional $pr$ quantile (e.g. the 1st percentile of the distribution). EM exploit that under correct model specification,

$$E\left[hit_{t+1}^{pr} z_t\right] = 0 \tag{17}$$

for any time $t$ measurable vector of instruments $z_t$, with dimensionality $m$. For example, if $z_t = 1$, then this test assesses, loosely speaking, whether $r_{t+1}$ falls below the $pr^{th}$ conditional

quantile in $pr$ percent of observations, consistent with proper specification. The test statistic,

$$\frac{G_T' \widehat{V}_T^{-1} G_T}{p(1-p)}, \tag{18}$$

where $G_T = \sum_{t=1}^{T} \left( hit_{t+1}^{pr} z_t \right)$ and $\widehat{V}_T = E\left[ \left( hit_{t+1}^{pr} z_t \right) \left( hit_{t+1}^{pr} z_t \right)' \right]$, converges to a $\chi^2$ distribution with $m$ degrees of freedom under certain conditions.

# 4 Empirical application: Monthly U.S. Stock returns, 1925–2010

The data we use are monthly log U.S. stock returns including dividends from 1925–2010 from the Center of Research in Securities Prices (CRSP). We first describe the parameter estimates of various models, then present the results of several specification tests, and end with a discussion of news impact curves.

## 4.1 Model estimation results

We estimate three traditional GARCH models that have been previously proposed:

1. the standard Gaussian GARCH (1,1) model, labeled "GARCH" in the table

2. the asymmetric GJR model, labeled "GJR–GARCH," with Gaussian innovations

3. the asymmetric GJR model assuming a Student's t-distribution for the shock, labeled "TDIST–GJR–GARCH"

We estimate several nested versions of the BEGE–GJR–GARCH model:

1. the full-fledged BEGE-GJR model, described in section 2, "Full BEGE–GJR"

2. a restricted version that imposes that all $p_t$ and $n_t$ coefficients are identical ($p_0 = n_0$, $\rho_p = \rho_n$, $\sigma_p = \sigma_n$, $\phi_p^+ = \phi_n^+$, $\phi_p^- = \phi_n^-$). Naturally, these restrictions lead to $p_t = n_t$

for all $t$. Relative to a GARCH(1,1) model, this model introduces conditional non-Gaussianity, but without admitting any non-zero conditional skewness. We estimate symmetric-GARCH and GJR versions of this model, labeled "Symmetric BEGE" and "Symmetric BEGE–GJR" respectively.

3. a restricted version with identical scale parameters ($\sigma_p = \sigma_n$) but unrestricted processes for the shape parameters, $p_t$ and $n_t$, labeled "BEGE–GJR different shapes"

4. a restricted version with only identical shape parameters ($p_0 = n_0$, $\rho_p = \rho_n$, $\phi_p^+ = \phi_n^+$, $\phi_p^- = \phi_n^-$) but without imposing equality of $\sigma_p$ and $\sigma_n$, labeled "BEGE–GJR different scales"

Within the RS class, we consider three different models:

1. a two-regime model with the special jump dynamics described in section 2, "2-regime with jump."

2. a standard two-regime model with constant mean across regimes, "2-regime"

3. a standard three regime model, "3-regime."

We find that some of the models that we estimate, particularly those with the highest numbers of parameters, are difficult to identify using data for returns alone. As a robustness check, we also estimate the full BEGE model using time series for realized variance in addition to returns. Realized monthly variances are computed for each period as the sum of intra-period squared returns. We assume the following model for the realized variance, $rvar_t$, which is defined as the sum of squared realized innovations in a given interval:

$$rvar_t = E_{t-1}rvar_t + \sigma_v \varepsilon_t, \tag{19}$$

where $\varepsilon_t \sim N(0,1)$ is a Gaussian error term and $E_{t-1}rvar_t$ is the model-dependent conditional variance. Under the BEGE model,

$$E_t rvar_{t+1} = \sigma_p^2 p_t + \sigma_n^2 n_t. \tag{20}$$

13

The total log likelihood for these estimations is the sum of the log likelihood for returns and the log likelihood for realized variance  For the Full BEGE–GJR "2-series" model, the Huber–White standard errors are much closer to those based solely on the Hessian, consistent with better identification.

Table 1 shows likelihood values for a variety of different models and their respective AIC and BIC criteria.  The models are ranked according to their BIC criterion.  As the table clearly indicates, the full BEGE models dominate in terms of AIC and BIC criteria, performing not only better than the standard Gaussian GARCH models and the GJR–GARCH model with an underlying t-distribution, but also better than the regime switching models. Among the regime switching models, the two-regime model with jumps performs best, and we restrict attention to that RS model henceforth.  Within the class of BEGE models, the model with different scales, but otherwise identical $p_t$ and $n_t$ parameters, performs best in terms of the BIC criterion (it is a very parsimonious model), but not in terms of the AIC criterion, where the full BEGE model performs best.  The traditional GARCH and GJR–GARCH models perform the worst.  We also investigate likelihood ratio tests among models within in the same class.  Note that full symmetry is rejected for both the GARCH and the BEGE models.  Within the class of the BEGE models, likelihood ratio tests reject all simpler models at the 1 percent level compared to the full BEGE–GJR–GARCH specification.  In the remainder of the paper, we focus our attention on the best performing models from each class:  Gaussian GJR–GARCH, henceforth referred to as "GJR," the two full BEGE–GJR–GARCH (henceforth referred to simply as BEGE and BEGE, 2-series), and the two-regime RS model including jumps.

Table 2 reports the parameter estimates for the two GARCH models.   The second column reports results from the BEGE model.  The third column reports the results from the BEGE, 2-series estimation.

Below every parameter estimate are two sets of standard errors; the first line is based on the inverse on the Hessian and the second uses the usual White (1982) standard errors.

14

It is well-known that in well-specified models, these standard errors should be close to one another.

Several well-known features of the data emerge upon inspection of the parameter values in Table 2. First, under the GJR specification, the conditional variance has a relatively high degree of persistence, with $\rho_h$ estimated at 0.85. Moreover, $h_t$ responds positively to squared innovations whether the innovations are positive or negative, as can be seen by the positive estimates for $\phi_h^+$ and $\phi_h^-$, but the response to negative shocks is about twice as large as that to positive shocks. The time series for raw returns and for $h_t$ are plotted in Figure 2. The large response of volatility to negative shocks is evident, for instance, following the 1987 crash.

Relative to this baseline, the parameter estimates from the BEGE model significantly refine our description of return dynamics. First, $\rho_p$ is estimated at about 0.91 while $\rho_n$ is estimated at 0.78, indicating that the good-environment volatility variable is significantly more persistent than the bad-environment variable. Although these estimates are not statistically distinct (under the inverse Hessian-based estimate of the parameter covariance matrix) for the 1-series estimates of the model, in the 2-series estimates, the standard errors for these parameters are significantly smaller, and $\rho_p$ and $\rho_n$ are statistically distinct. In terms of responses of volatility to positive versus negative shocks, the BEGE model suggests more intricate return dynamics. The parameter $\phi_p^+$ is substantially larger than $\phi_p^-$, indicating that good volatility responds to positive shocks more than it does to negative shocks. In contrast, $\phi_n^+$ is estimated to be negative (slightly positive) under the 1-series (2-series) estimation, while $\phi_n^-$ is strongly positive and much larger in magnitude. This indicates that bad volatility, or the negative tail of the return distribution, substantially increases following negative shocks. This, of course, is a feature of the data that has substantial risk-management implications but which standard Gaussian models cannot hope to match. Figure 3 shows the time series patterns of $p_t$ and $n_t$ from the BEGE model. Using the 1987 crash as an example again, that negative shock sharply increases the bad volatility regime

variable, $n_t$, but it hardly affects $p_t$ at all. This result implies that the negative tail of the return distribution widened following the crash, but the upper tail was less affected.

Table 3 reports on the parameter estimates from the RS models. We identify the regimes by defining them to be increasing in the innovation variances. As is typically found, the innovation volatilities are very different across regimes. Under the two-regime specification including jumps, the first regime registers a 3.7 percent shock volatility, but the second regime has a 10.7 percent shock volatility. Also typical is the finding that the low-volatility regime is more persistent than the high-volatility regime (see also Ang and Bekaert, 2002). In the models including jumps, note that a transition from the low-volatility to the high-volatility regime is associated with a negative return of 10 percent, whereas a transition from the high variance to the low variance regime entails a positive return of 5 percent. The jump terms imply that the conditional mean in the high-variance regime is higher than in the low-variance regime. In fact, using the estimated transition probabilities, the mean in the high-variance regime is 1.8 percent, but in the low variance regime it is just 0.9 percent. These differences can be contrasted with the overall unconditional mean of 1.10 percent as reported for the two-regime model without jumps. Figure 4 plots ex-ante and smoothed estimates of the probability of being in the high-volatility regime, which are calculated in the usual manner (see Appendix B). High-volatility regimes include the Great Depression, the pre-war period, the first oil shock, the October 1987 crash, the period following September 11, 2001, the 1998 Russia and LTCM crises, and the recent global financial crisis. The relatively low persistence of the high-volatility regime is readily apparent.

## 4.2 Specification test results

In Table 4, Vuong–Fisher–Calvet likelihood ratio tests for non-nested models are reported. In the table, positive (negative) entries indicate that the model listed in the first column dominates (underperforms) the model listed in the first row. In every case, the BEGE models dominate the competing GARCH and RS models. For the simple Vuong

tests, rejections are at the 1 percent level, and the 1-series BEGE model rejects the bivariate BEGE-GARCH model at the 5 percent level.[4] For the Calvet–Fisher test, the 1-series BEGE model rejects the RS model only at the 5 percent level.

Table 5 tests how well the various models are able to match the unconditional moments of returns observed in the data. The GJR model performs especially poorly, significantly undershooting the magnitude of unconditional skewness in the data, which is negative, and also undershooting kurtosis. No rejections are found for the RS or 1-series BEGE model, but the 2-series BEGE model is narrowly rejected for both unconditional volatility and kurtosis.

Tests regarding the conditional distribution of returns are presented in Tables 6 and 7. Specifically, the tests examine how well the models can replicate shifts in the conditional distribution of returns that occur following positive and negative return shocks. Tests of the changes in the lower tail of the distribution coincide with value-at-risk measures, a popular risk-management tool. In Table 6, we test how well the models fit the change in the distribution of returns following negative and positive return realizations. In the upper portion of the table, the column labeled "sample" reports the estimated change in various quantiles (down the rows) following negative return realizations. The lower portion of the table reports on quantile shifts following positive realizations. For instance, the 5th percentile of the distribution falls by about 2.2 percentage points following a negative return, while the 95th percentile rises by about 1.3 percent. Following positive realizations, the 5th percentile typically rises by about 2.1 percentage points, while the 95th percentile falls by about 0.5 percentage points. The top panel of Figure 5 graphically depicts these quantile shifts. The blue squares plot the unconditional distribution of returns. The green triangles plot the distribution of returns following a positive realization in the previous period, and the red triangles plot the distribution following a negative realization. Clearly, the lower tail of the return distribution is more sensitive to recent return realizations (positive and negative) than are the upper tails of the distribution. Also, negative (positive) shocks lead

---

[4]Of course, this test only considers the return equation, ignoring any difference in the performance of the two models in terms of matching the realized variance series.

to a wider (narrower) probability distribution for the next period.

Returning to Table 6, the columns to the right show how well the various models match these historical patterns. To implement the test, we draw 10,000 random samples under each model using the parameters reported in Tables 2 and 3. Each sample has length equal to that of our data sample, 1,020 observations. For each random sample, we calculate quantile shifts exactly as we do for the data. Finally, we report what fraction of observed quantile shifts in the random samples are lower than those observed in the sample. If this fraction is very small or very large, we conclude that the model is inconsistent with the sample data for that quantile shift. That is, we can observe rejections at either tail of the distribution. We denote rejections at the 1, 5, and 10 percent levels using one, two, or three asterisks.

The second column from the left column reports results for a trivial model in which the conditional distribution at each point is simply equal to the unconditional distribution observed for the sample. This model is strongly rejected using tests at the tails of the distribution. This result indicates that the observed quantile shifts in the sample are very unlikely to be observed if the true underlying conditional distribution is constant. The remaining columns show results for our four key models. All of the models suffer some rejections for quantile shifts in the lower portion of the distribution. The GJR model fares the poorest, with strong rejections for the 5th and 10th quantiles. The RS and BEGE models perform somewhat better, although all the models fail to capture the increase in the lower quantiles following a positive return realization. However, the p-values are smaller for the BEGE (1-series) model.

Table 7 repeats this exercise, but examining quantile shifts following larger (in magnitude) return realizations. Specifically, we now examine return realizations one (unconditional) standard deviation above and below the unconditional mean. In the data, strong quantile shifts are evident following the larger negative shocks, with the lower 5th, 10th, and 25th percentiles falling 6, 5, and 4 percent, respectively. In contrast, there is less evidence of large quantile shifts following positive realizations, as shown in the lower portion of the

table. These quantile shifts are illustrated in the lower panel of Figure 5. It is little surprise that an unconditional model fits the positive quantile shifts but fails to fit the quantile shifts after a negative realization for almost every quantile. The fours models that we examine again feature a number of rejections. The GJR model again fares the worst missing the quantile shifts following a negative shock for the four lowest quantiles (5, 10, 25 and 50), although none at the 1 percent significance level. Perhaps surprisingly, this model also fails to fit the 5 and 95 percent quantile shifts following a positive shock. The same is true for the regime switching model, but that model fits the shifts following a negative shock for the lowest percentiles relatively well. The RS model can also not match the negative shock quantile shifts for the 25th and 50th percentiles of the distribution. The BEGE model (1 series) again performs the best. Nevertheless, three rejections occur at the 5 percent level: quantile shifts at the 25th and 50th percentiles following a negative shock and the quantile shift following a positive shock for the 95th percentile. While the fit is thus not perfect, of the few rejections we record for the BEGE model, none is at the 1 percent level.

To further examine which model provides the most accurate description of the conditional distribution of returns, we turn to the tests of EM. In doing so, we will focus on the lower portion of the distribution, specifically the 1st and 5th percentiles, which have implications as value-at-risk metrics. To make this determination, we begin by plotting selected quantiles for our models of focus. Figure 6 plots the conditional quantiles for the GJR model, the two-regime RS model with jumps, and the one-series BEGE models (the two-series version looks very similar). The GJR model distributions are Gaussian and thus symmetric about zero. Quantiles from the GJR model exhibit quite a bit more persistence in the extreme quantiles than do those from the RS model. Some non-Gaussian features of the BEGE distribution are readily evident. For instance, the lower quantiles of the distribution have larger magnitudes than do the corresponding upper quantiles. This is equivalent to negative conditional (quantile) skewness.

Armed with the conditional quantiles implied by each model, we proceed to implement

the EM tests. For each quantile and model tested, we begin by defining the sequence of hits, $hit_{t+1}^{pr}$, as described in section 3. We select a small set of instruments for the test. Specifically, we choose

$$z_t = [1, hit_t^{pr}, r_t] \tag{21}$$

so that we are testing that the mean rate of exceedences of the quantile in question is accurate (e.g., the 1st quantile should be exceeded in 99 percent of observations), as well as orthogonality of $hit_{t+1}^{pr}$ to $hit_t^{pr}$ and $r_t$. The latter two instruments are intuitive, as one would surely prefer a model for which hits are not autocorrelated and also for which hits are not forecastable by lagged returns. We test for orthogonality of the instruments individually. To do so, we calculate the statistic,

$$\frac{G_T' \widehat{V}_T^{-1} G_T}{p(1-p)} \xrightarrow{d} \chi_1^2 \tag{22}$$

where $G_T = \sum_{t=1}^{T} \left( hit_{t+1}^{pr} z_t \right)$ and $\widehat{V}_T = E\left[ \left( hit_{t+1}^{pr} z_t \right) \left( hit_{t+1}^{pr} z_t \right)' \right]$. We compare this statistic to critical values of the standard $\chi_1^2$ distribution. In doing so, we ignore that our test is conducted on an in-sample basis, which, as EM point out, alters the sampling distribution of the test statistic. Our tests are thus informal. We use a measure of the covariance matrix, $\widehat{V}_T$, that is constant across models so that results across models are more comparable.[5]

The top panel of Table 8 shows results for the 1st quantile of the return distribution. The GJR model fails every test, including that based on $z_t = 1$. That is, we can reject that the GJR model-implied 1st percentile is exceeded 99 percent of the time. We also reject that GJR hit errors are orthogonal to lagged values of $hit_t^p$ or lagged returns. The BEGE and RS models perform somewhat better. We do not reject those models for $z = 1$, and we reject for the other instruments only at more stringent levels. In particular, the BEGE (1 series) model performs the best, producing only rejections at the 10 percent level. Results for the 5th percentile, shown in the lower panel of Table 8, are broadly similar. The GJR model performs miserably. In this case, the RS model is also rejected for each of the

---

[5]In the results reported, we used an estimate of $\widehat{V}_T$ that is based on the BEGE 1-series models. For robustness, we tried using $\widehat{V}_T$ estimates from all of the models, which yielded similar results.

instruments. In contrast, the BEGE models are rejected only for $z_t = r_t$. The joint tests of orthogonality to the instruments, provided rejections at the 1 percent level for all of the models for both the 1st and 5th percentiles. In sum, the EM tests appear to be challenging for all of the models. However, at least against $z_t = 1$ and $z_t = hit_t$ individually, the BEGE models perform fairly well, and quite a bit better than the competing models that we tested.

## 4.3  Impact curves

In Figure 7, we report conditional moment impact curves for the GARCH models as inspired by Engle and Ng (1993). That is, the curve describes the relationship between $h_t$ and the past shock, $u_{t-1}$, holding constant (at unconditional means) all information at time $(t-2)$. The analytic expressions describing the impact of a squared shock represent the derivatives of the conditional variance function with respect to the squared shock and were presented in section 2. For all the panels shown, shocks are represented on the horizontal axes, ranging from minus to positive 20 percentage points, representing the range of return shocks present in the data. On the vertical axes are the responses of various conditional moments to the shocks under the model listed. For instance, the upper-left panel shows the response of conditional variance under the GJR model.

As expected, negative shocks are associated with a larger increase in conditional variance than are positive shocks of the same magnitude. The effect is even more pronounced under the BEGE model, shown to the right, which suggests that conditional variance is little affected by positive shocks. The second row of panels plot the responses of conditional scaled skewness to return innovations. For the GJR, the effect is identically zero, an artifact of the assumed conditional Gaussianity. The BEGE model, in contrast, suggests an increase in (the generally negative) conditional scaled skewness of returns, which is much stronger, following a positive shock. Only at very large positive shocks does the skewness become positive. The third row shows the responses of conditional scaled kurtosis. Again for the standard GJR model, these are zero, by definition, whereas the BEGE model suggests

that conditional scaled kurtosis is generally decreasing the larger the shock is in magnitude, regardless of sign.

It is also instructive to examine the responses of unscaled skewness and kurtosis, to help discern the effects on the third and fourth moments from effects on volatility. These results are shown in the bottom two panels of the figure. For unscaled skewness, the BEGE model generates sharp drops for negative shocks (as the negatively skewed component of the BEGE distribution becomes more important) but increases in skewness for positive shocks, although these are less steep. Therefore, the reason that scaled skewness actually increases with negative shocks is that volatility (cubed) goes up by even more than the third moment decreases when negative shocks occur. For unscaled kurtosis, we obtain a flat pattern for positive shocks, and a rather sharp increase for negative shocks. Since actual kurtosis falls with both positive and negative shocks, it must be that volatility effects dominate. All in all, the BEGE model suggests a rich pattern of news impact curves for higher-order moments, which conditional Gaussian models cannot match. The quantile test results in the previous section show that these patterns are necessary to help explain conditional quantile shifts in the data. We suspect that such patterns may also be important for explaining option price dynamics.

# 5   Conclusion

We have introduced an extension of standard asymmetric volatility models in the GARCH class that admits conditional non-Gaussianities in a tractable fashion. Our bad environment–good environment (BEGE) model features two gamma-distributed shocks that imply a conditional shock distribution with time-varying heteroskedasticity, skewness and kurtosis. Our model features nontrivial news impact curves for higher-order moments. In an empirical application to monthly U.S. stock returns, the model outperforms standard asymmetric GARCH and regime-switching models along several dimensions.

In this application, we have embedded the BEGE structure in a GARCH framework, which provides for easy estimation since the factors driving conditional volatility and the conditional distribution of returns are essentially observable conditional on the model parameters and the sequence of returns. We believe a number of interesting applications, for example, to risk management, are therefore possible and very tractable. Useful applications in macroeconomics are conceivable as well. While in financial returns the BEGE framework helps fit asymmetries on the downside, for inflation data, the ability of the model to generate positive conditional skewness could help model inflation scares – periods in which very high inflation becomes more probable.

Of course, the BEGE model remains very simple, and may miss some important features of economic data. For instance, there is evidence suggesting that the volatility and possibly higher-order moments as well depend on factors that are imperfectly correlated with realized returns. The option pricing literature typically relies on stochastic volatility models rather than GARCH-type models, partially because of tractability, but also because having volatility shocks that are not perfectly correlated with returns helps models fit option prices. It is feasible to create a version of the BEGE framework where the BEGE factors have independent shocks. An additional advantage of the BEGE framework in this regard is tractability. Some important features of options prices, such as risk-neutral moments, have closed form solutions in a BEGE framework with independent latent factors.

# Appendices

## A    Evaluating the BEGE density

Random variables with the BEGE density take the form $x = \omega_p - \omega_n$, where $x$ is the BEGE-distributed variable, and $\omega_p$ and $\omega_n$ are demeaned gamma distributions with parameter vectors (shape and scale) of $\left(k_{\omega_p}, \theta_{\omega_p}\right)$ and $(k_{\omega_n}, \theta_{\omega_n})$, respectively. We seek an expression for the density of $x$, $f_{BEGE}(x)$. To begin, using Bayes's rule,

$$
f_{BEGE}(x) = \int_{\omega_p} f_x(x|\omega_p) f(\omega_p) d\omega_p
$$

$$
= \int_{\omega_p} f_{\omega_n}(\omega_p - x) f(\omega_p) d\omega_p
$$

Now, let us specialize to the demeaned gamma distribution for $\omega_p$ and $\omega_n$:

$$
f_{\omega_p}(\omega_p) = \frac{\left(\omega_p - \underline{\omega_p}\right)^{k_{\omega_p}-1} \exp\left(-\left(\omega_p - \underline{\omega_p}\right)/\theta_{\omega_p}\right)}{\Gamma\left(k_{\omega_p}\right)\theta_{\omega_p}^{k_{\omega_p}}} \text{ for } \omega_p > \underline{\omega_p}
$$

$$
f_{\omega_n}(\omega_n) = \frac{\left(\omega_n - \underline{\omega_n}\right)^{k_{\omega_n}-1} \exp\left(-\left(\omega_n - \underline{\omega_n}\right)/\theta_{\omega_n}\right)}{\Gamma\left(k_{\omega_n}\right)\theta_{\omega_n}^{k_{\omega_n}}} \text{ for } \omega_n > \underline{\omega_n}
$$

where $\underline{\omega_p} = -k_{\omega_p}\theta_{\omega_p}$ and $\underline{\omega_n} = -k_{\omega_n}\theta_{\omega_n}$. The upper limit of integration in the expression for $f_{BEGE}(x)$ is infinity. The lower limit for $\omega_p$ must satisfy both $\omega_p > \underline{\omega_p}$ and $(\omega_p - x) > \underline{\omega_n}$ or $\omega_p > x + \underline{\omega_n}$. Define $\overline{\omega_p}$ as $max\left(\underline{\omega_p}, x + \underline{\omega_n}\right)$, then,

$$
f_{BEGE}(x) = \int_{\omega_p=\overline{\omega_p}}^{\infty} f_{\omega_p}(\omega_p) f_{\omega_n}(\omega_p - x) d\omega_p
$$

$$
= A_1 A_2 A_3 \int_{\omega_p=\overline{\omega_p}}^{\infty} \left(\omega_p - \underline{\omega_p}\right)^{k_{\omega_p}-1} \left(\omega_p - x - \underline{\omega_n}\right)^{k_{\omega_n}-1} \exp\left(-\omega_p\widetilde{\theta}\right) d\omega_p
$$

where $A_1 = \frac{1}{\Gamma\left(k_{\omega_p}\right)\theta_{\omega_p}^{k_{\omega_p}}} \frac{1}{\Gamma(k_{\omega_n})\theta_{\omega_n}^{k_{\omega_n}}}$, $A_2 = \exp\left(\underline{\omega_p}/\theta_{\omega_p} + \underline{\omega_n}/\theta_{\omega_n}\right)$, $A_3 = \exp\left(x/\theta_{\omega_n}\right)$ and $\widetilde{\theta} = \left(1/\theta_{\omega_p} + 1/\theta_{\omega_n}\right)$. There are known solutions for integrals of the form

$$
W_{k,m}(z) = \frac{\exp\left(-z/2\right) z^k}{\Gamma\left(\frac{1}{2} - k + m\right)} \int_{t=0}^{\infty} t^{(-k-1/2+m)} \left(1 + \frac{t}{z}\right)^{(k-1/2+m)} \exp\left(-t\right) dt
$$

where $W_{k,m}(z)$ is the Whittaker W function. To use this result, we use a change of variables, defining, $\widetilde{\omega_p} = \omega_p \widetilde{\theta} - \overline{\omega_p} \widetilde{\theta}$. Then, $\omega_p = \frac{1}{\widetilde{\theta}} \widetilde{\omega_p} + \overline{\omega_p}$. Substituting,

$$f_{BEGE}(x) = A_1 A_2 A_3 \int_{\widetilde{\omega_p}=0}^{\infty} \left( \frac{1}{\widetilde{\theta}} \widetilde{\omega_p} + \overline{\omega_p} - \underline{\omega_p} \right)^{k_{\omega_p}-1} \left( \frac{1}{\widetilde{\theta}} \widetilde{\omega_p} + \overline{\omega_p} - x - \underline{\omega_n} \right)^{k_{\omega_n}-1}$$

$$\exp\left( -\widetilde{\omega_p} - \overline{\omega_p} \widetilde{\theta} \right) \frac{1}{\widetilde{\theta}} d\widetilde{\omega_p}$$

This integral simplifies for the specific cases at hand. First, if $\overline{\omega_p} = \underline{\omega_p}$. Then the integral becomes

$$f_{BEGE}(x) = A_1 A_2 A_3 A_4 \cdot$$

$$\int_{\widetilde{\omega_p}=0}^{\infty} \left( \frac{1}{\widetilde{\theta}} \widetilde{\omega_p} \right)^{k_{\omega_p}-1} \left( \frac{1}{\widetilde{\theta}} \widetilde{\omega_p} + \underline{\omega_p} - x - \underline{\omega_n} \right)^{k_{\omega_n}-1} \exp\left( -\widetilde{\omega_p} \right) \frac{1}{\widetilde{\theta}} d\widetilde{\omega_p}$$

$$= A_1 A_2 A_3 A_4 A_5 A_6 \cdot$$

$$\int_{\widetilde{\omega_p}=0}^{\infty} \widetilde{\omega_p}^{k_{\omega_p}-1} \left( \frac{\widetilde{\omega_p}}{\widetilde{\theta}\left( \underline{\omega_p} - x - \underline{\omega_n} \right)} + 1 \right)^{k_{\omega_n}-1} \exp\left( -\widetilde{\omega_p} \right) d\widetilde{\omega_p}$$

where $A_4 = \exp\left( -\underline{\omega_p} \widetilde{\theta} \right)$, $A_5 = \left( \frac{1}{\widetilde{\theta}} \right)^{k_{\omega_p}}$, $A_6 = \left( \underline{\omega_p} - x - \underline{\omega_n} \right)^{k_{\omega_n}-1}$. The integral term is now isomorphic to that in the expression for $W_{k,m}(z)$ above. Substitution and algebra yields the final expression,

$$f_{BEGE}(x) = A_1 A_2 A_3 A_4 A_5 A_6 A_7 A_8 W_{k,m}(z)$$

where $A_7 = \Gamma\left( \frac{1}{2} - k + m \right)$, $A_8(z) = \exp(z/2) z^{-k}$, $z = \left( \underline{\omega_p} - x - \underline{\omega_n} \right) \widetilde{\theta}$, $m = \frac{1}{2}\left( k_{\omega_n} + k_{\omega_p} - 1 \right)$, and $k = \frac{1}{2}\left( k_{\omega_n} - k_{\omega_p} \right)$.

In the second case, $\overline{\omega_p} = x + \underline{\omega_n}$, and similar calculations lead to

$$f_{BEGE}(x) = A_1 A_2 A_3 A_4' A_5' A_6' A_7' A_8' W_{k',m'}(z')$$

where $A_4' = \exp\left( -\left( x + \underline{\omega_n} \right) \widetilde{\theta} \right)$, $A_5' = \left( \frac{1}{\widetilde{\theta}} \right)^{k_{\omega_n}}$, $A_6' = \left( x + \underline{\omega_n} - \underline{\omega_p} \right)^{k_{\omega_p}-1}$, $A_7' = \Gamma\left( \frac{1}{2} - k' + m' \right)$, $A_8' = \exp(z'/2) z'^{-k}$, $z' = -z$., $m' = m$, $k' = -k$.

# B Regime-switching model specification and estimation

We estimate three regime-switching models: benchmark models with two and three regimes as well as a jump model. The log-likelihood function for this model is:

$$L(\{y_1, y_2, ..., y_T\}; \theta) = \sum_{t=1}^{T} \log f(y_t | Y_{t-1}; \theta),$$

where $Y_t$ is the history of observations up to time $t$ and $f$ is the probability density function. To evaluate the likelihood, note that:

$$f(y_t | Y_{t-1}; \Theta) = \sum_s p(s_t | Y_{t-1}) f(y_t | Y_{t-1}, s),$$

where $p(s_t | Y_{t-1})$ is the probability of the regime $s$ at time $t$ conditioned on the observations up to time $t$ and can be computed as:

$$
\begin{aligned}
p(s_t | Y_{t-1}) &= \sum_{s'_{t-1}} P\left(s_t | s'_{t-1}\right) p\left(s'_{t-1} | Y_{t-1}\right) \\
&= \sum_{s'_{t-1}} P\left(s_t | s'_{t-1}\right) \frac{p\left(s'_{t-1} | Y_{t-2}\right) f\left(y_{t-1} | Y_{t-2}, s'_{t-1}\right)}{\sum_{s''_{t-1}} p\left(s''_{t-1} | Y_{t-2}\right) f\left(y_{t-1} | Y_{t-2}, s''_{t-1}\right)}
\end{aligned}
$$

Each observation is assumed to follow:

$$r_t = \mu + \sigma(s_t) e_t,$$

where $e_t$ is i.i.d. standard normal, so $y_t = r_t$. We consider the models with 2 and 3 regimes. The parameters to estimate are the mean return ($\mu$), the standard deviations of the regime distributions ($\sigma_i$), and the transition probability matrix ($P(s_{t+1} = i | s_t = j)$). The prior distribution over regimes $p(s_0)$ is set equal to the unconditional probabilities.

Formally, the estimation is done by numerically maximizing the likelihood function. In order to avoid local maxima, we use different initial parameters for the optimization algorithm. We also check the stability of the final solution by randomly deviating from the estimates, and verifying that the routine returns to the provisional maxima.

In the model allowing for two regimes and jumps, each observation is assumed to follow

$$r_t = \mu_0 + \mu_{12} J_{12,t} + \mu_{21} J_{21,t} + \sigma(s_t) e_t,$$

where $e_t$ is again i.i.d. standard normal and $J$ is a dummy variable specified as

$$
J_{12,t} = \begin{cases} 1 & \text{if } s_{t-1} = 1 \text{ and } s_t = 2, \\ 0 & \text{otherwise,} \end{cases}
$$

and

$$J_{21,t} = \begin{cases} 1 & \text{if } s_{t-1} = 2 \text{ and } s_t = 1, \\ 0 & \text{otherwise} \end{cases}$$

For the RS models including jumps, the likelihood function is more complex than in the benchmark case:

$$f\left(y_t | Y_{t-1}, s\right) \sim \begin{cases} N\left(\mu_0, \sigma_i^2\right) & \text{if } s_{t-1} = s_t = i \\ N\left(\mu_0 + \mu_{ji}, \sigma_i^2\right) & \text{if } s_{t-1} = j \neq i = s_t \end{cases}$$

This model can be recast as a regime switching model with 4 states, after which the usual likelihood construction can proceed.

# References

Ang, A., and G. Bekaert, 2002, "International Asset Allocation with Regime Shifts." Review of Financial Studies, 15, 4, pp. 1137-1187.

Bollerslev, T., 1986, "Generalized Autoregressive Conditional Heteroskedasticity," Journal of Econometrics 31, pp. 307–28.

Bollerslev, T., 1987, "A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return," The Review of Economics and Statistics, MIT Press, vol. 69(3), pp 542–47.

Broadie, M., Chernov, M., and M. Johannes, 2009, "Understanding Index Option Returns," Review of Financial Studies, Society for Financial Studies, vol. 22(11), pp 4493–4529.

Calvet, L. and A. Fisher, 2004, "How to Forecast Long-Run Volatility: Regime-Switching and Estimation of Multifractal Processes," Journal of Financial Econometrics, 2, pp.49–83.

Engle, R. F., 1982, "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," Econometrica, Econometric Society, vol. 50(4), pp 987–1007.

Engle R.F. and S. Manganelli, 2004, "CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles," Journal of Business & Economic Statistics, American Statistical Association, vol. 22, pp 367–381.

Engle, R. F. and V. Ng, 1993, "Measuring and Testing the Impact of News on Volatility," Journal of Finance, American Finance Association, vol. 48(5), pp 1749–78.

Evans, M. and P. Wachtel, 1993, "Inflation Regimes and the Sources of Inflation Uncertainty," Journal of Money, Credit and Banking, 25, 3, pp. 475–511.

French, M. and D. Sichel, 1993, "Cyclical Patterns in the Variance of Economic Activity," Journal of Economic and Business Statistics, 11, 1, pp. 113–119.

Glosten, L. R, Jagannathan, R. and D. Runkle, 1993, "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks," Journal of Finance, American Finance Association, vol. 48(5), pp 1779–1801.

Hamilton, J., 1989, "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," Econometrica, Econometric Society, vol. 57(2), pp 357–84.

Hamilton, J. and R. Susmel, 1994, "Autoregressive conditional heteroskedasticity and changes in regime," Journal of Econometrics, Elsevier, vol. 64(1-2), pp 307–333.

Hansen, L., 1982, "Large Sample Properties of Generalized Method of Moments Estimators," Econometrica , Vol. 50, No. 4, pp. 1029–1054.

Hsieh, D., 1989, "Testing for Nonlinear Dependence in Daily Foreign Exchange Rates," Journal of Business, Vol. 62, 3, pp.339–368.

Huber, P., 1967, "The Behavior of Maximum Likelihood Estimates Under Nonstandard Conditions," Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, pp. 221-233.

Kim, C.J., 1994, "Dynamic linear models with Markov-switching," Journal of Econometrics, Elsevier, vol. 60(1-2), pp 1–22.

Kim, T.H.. and H. White, 2004, "On more robust estimation of skewness and kurtosis,". Finance Research Letters 1, pp.56-73.

Lumsdaine, R.L., 1996, "Consistency and Asymptotic Normality of the quasi-Maximum Likelihood Estimator in IGARCH(1,1) and Covariance Stationary GARCH(1,1) Models," Econometrica, Vol. 64(3), pp. 575-596.

Mayfield, E.S., 2004, "Estimating the Market Risk Premium," Journal of Financial Economics 73, no. 3, pp. 465–496.

Mittnik, S., Paotella, M.S., and S.T. Rachev, 2002, "Stationarity of Stable Power-GARCH Processes," Journal of Econometrics, Vol. 106(1), pp. 97-107.

Nelson, D. B., 1991, "Conditional heteroskedasticity in asset returns: A new approach", Econometrica 59, pp. 347–370.

Newey, W. and K. West, 1987,. "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," Econometrica, Econometric Society, Vol. 55(3), pp. 703–08.

Poon, S.H. and C. Granger, 2003, "Forecasting Volatility in Financial Markets: A Review," Journal of Economic Literature, American Economic Association, vol. 41(2), pp. 478–539.

Rivers, D. and Q. Vuong, 2002, "Model selection tests for nonlinear dynamic models," The Econometrics Journal, Vol. 5, 1, pp. 1–39.

Timmermann, A., 2000, "Moments of Markov switching models," Journal of Econometrics, Elsevier, vol. 96(1), pp. 75–111.

Vuong, Q.H., 1989, "Likelihood Ratio Tests for Model Selection and Non-nested Hypotheses," Econometrica, Vol. 57, 2, pp. 307–33.

White, H., 1982, "Maximum Likelihood Estimation of Misspecified Models," Econometrica, 50, pp. 1–26.

Table 1: Model performance comparison

| Model | # parameters | Log-likelihood | AIC | BIC |
|---|---|---|---|---|
| BEGE, different scales | 7 | 1716.54 | -3419.08 | -3384.59 |
| Full BEGE–GJR | 11 | 1724.26 | -3426.53 | -3372.32 |
| TDIST–GJR–GARCH | 6 | 1703.60 | -3395.20 | -3365.63 |
| BEGE–GJR, different shapes | 10 | 1716.67 | -3413.34 | -3364.07 |
| Symmetric BEGE–GJR | 6 | 1695.50 | -3379.00 | -3349.43 |
| Symmetric BEGE | 5 | 1690.27 | -3370.55 | -3345.91 |
| Full BEGE–GJR, 2-series | 11 | 1709.75 | -3397.50 | -3343.30 |
| 2-regime w/jump | 8 | 1692.74 | -3369.48 | -3330.06 |
| 2-regime | 6 | 1682.93 | -3353.87 | -3324.30 |
| 3-regime | 12 | 1698.69 | -3373.37 | -3314.24 |
| GARCH | 4 | 1668.72 | -3329.44 | -3309.73 |
| GJR–GARCH | 5 | 1671.77 | -3330.54 | -3308.90 |

Table 2: Parameter estimates for GARCH models

| Model | GJR | BEGE | BEGE, 2-series |
|---|---|---|---|
| $\mu$ | 0.0089 | 0.0100 | 0.0115 |
| | (0.0013) | (0.0013) | (0.0010) |
| | (0.0013) | (0.0015) | (0.0010) |
| $p_0/h_0$ | 0.00008 | 0.0890 | 0.1728 |
| | (0.00003) | (0.0917) | (0.0152) |
| | (0.00005) | (0.0070) | (0.0140) |
| $n_0$ | | 0.2204 | 0.4112 |
| | | (0.0839) | (0.0410) |
| | | (0.0111) | (0.0490) |
| $\rho_h/\rho_p$ | 0.8516 | 0.9099 | 0.8774 |
| | (0.0195) | (0.0855) | (0.0083) |
| | (0.0313) | (0.0009) | (0.0037) |
| $\rho_n$ | | 0.7822 | 0.6879 |
| | | (0.1200) | (0.0260) |
| | | (0.0086) | (0.0436) |
| $\sigma_p$ | | 0.0072 | 0.0101 |
| | | (0.0008) | (0.0007) |
| | | (0.0008) | (0.0004) |
| $\sigma_n$ | | 0.02823 | 0.0224 |
| | | (0.0092) | (0.0017) |
| | | (0.0087) | (0.0021) |
| $\phi_h^+/\phi_p^+$ | 0.0800 | 0.0964 | 0.1156 |
| | (0.0283) | (0.0440) | (0.0081) |
| | (0.0404) | (0.0426) | (0.0054) |
| $\phi_h^-/\phi_p^-$ | 0.1563 | 0.0128 | -0.0116 |
| | (0.0283) | (0.0125) | (0.0033) |
| | (0.0431) | (0.0139) | (0.0029) |
| $\phi_n^+$ | | -0.0789 | 0.0939 |
| | | (0.0400) | (0.0264) |
| | | (0.0524) | (0.0385) |
| $\phi_n^-$ | | 0.3548 | 0.4234 |
| | | (0.0473) | (0.0490) |
| | | (0.0901) | (0.0612) |
| $\sigma_v$ | | | 0.0040 |
| | | | (0.0001) |
| | | | (0.0001) |

For the 2-series case, the monthly realized variances are computed as the sum of squared daily realized logarithmic returns. Asymptotic standard errors are in parentheses. The first standard errors are inverse Hessian standard errors. The second standard errors estimates used are of the Huber–White sandwich type.

Table 3: Parameter estimates for regime-switching models

| Model | 2-regime | 3-regime | 2-regime w/jump |
|---|---|---|---|
| $\mu$ | 0.01102 | 0.01114 | 0.01128 |
| | (0.00135) | (0.00132) | (0.00133) |
| $\mu_{lh}$ | - | - | -0.1048 |
| | | | (0.0217) |
| $\mu_{hl}$ | - | - | 0.05302 |
| | | | (0.01038) |
| $\sigma_{low}$ | 0.03749 | 0.03033 | 0.03707 |
| | (0.00120) | (0.00511) | (0.00115) |
| $\sigma_{middle}$ | - | 0.04744 | - |
| | | (0.00908) | |
| $\sigma_{high}$ | 0.1098 | 0.1237 | 0.1069 |
| | (0.0104) | (0.0190) | (0.0102) |
| $P(s_{t+1}=l|s_t=l)$ | 0.9834 | 0.9678 | 0.9806 |
| | (0.0060) | (0.0519) | (0.0059) |
| $P(s_{t+1}=m|s_t=l)$ | - | 0.03220 | - |
| | | (0.01695) | |
| $P(s_{t+1}=m|s_t=m)$ | - | 0.9551 | - |
| | | (0.0325) | |
| $P(s_{t+1}=h|s_t=m)$ | - | 0.01846 | - |
| | | (0.02556) | |
| $P(s_{t+1}=m|s_t=h)$ | - | 0.06121 | - |
| | | (0.06070) | |
| $P(s_{t+1}=h|s_t=h)$ | 0.9019 | 0.9300 | 0.8785 |
| | (0.0362) | (0.0440) | (0.0362) |

The data are logged monthly dividend-adjusted stock returns from December 1925 to December 2010. The asymptotic Huber-White standard errors are in parentheses.

Table 4: Likelihood ratio tests for non-nested models

| | Vuong (1989) $t$-test | | | |
|---|---|---|---|---|
| Model 1/Model 2 | GJR-GARCH | 2 regime w/ jump | BEGE | BEGE, 2 series |
| BEGE, 1 series | 3.0710*** | 3.1426*** | - | 2.5016** |
| BEGE, 2 series | 2.5242** | 1.9776** | -2.5016** | - |
| GJR | - | -1.0577 | -3.0710*** | -2.5242** |

| | Fisher-Calvet (2004) $t$-test | | | |
|---|---|---|---|---|
| Model 1/Model 2 | GJR | 2 regime w/jump | BEGE | BEGE, 2 series |
| BEGE, 1 series | 2.9411*** | 2.4281** | - | 2.7407*** |
| BEGE, 2 series | 2.2915** | 1.5006 | -2.7407*** | - |
| GJR | - | -0.8999 | -2.9411*** | -2.2915** |

The data are logged monthly dividend-adjusted stock returns from December 1925 to December 2010. A positive number means that model 1 (listed in the first column) is better than model 2 (listed in the first row), and a negative number means that model 1 is worse than model 2. For the BEGE models estimated from both returns and variance time series, the likelihood is computed only for the time series of returns. In the Calvet–Fisher tests, the HAC-adjusted variance estimator is the Newey–West (1987) estimator with 12 lags. The asterisks, *,**, and *** correspond to the statistical significance at 10, 5, and 1 percent levels, respectively.

Table 5: Unconditional moment test

|  | Sample est. | CDF value from 10,000 samples Monte Carlo | | | |
|---|---|---|---|---|---|
|  |  | GJR | 2 regime w/ jump | BEGE | BEGE, 2 series |
| Stdev | 0.0545 | 0.7104 | 0.5422 | 0.8424 | 0.9554* |
| Skew | -0.5742 | 0.0086** | 0.2903 | 0.6994 | 0.5389 |
| Excess Kurt | 6.6134 | 0.9812** | 0.8811 | 0.9481 | 0.9704* |

The data are logged monthly dividend-adjusted stock returns from December 1925 to December 2010. In each Monte Carlo sample the time series of 1,020 observations (the same length as the historical time series) is generated by randomly drawing error terms and using the estimated parameters for each model. Next, the values of variance, skewness, and kurtosis are computed for that generated time series. The reported cumulative distribution function (CDF) values are probabilities that the value less than or equal to the historical value is observed under the simulated distribution. The asterisks, *,**, and *** correspond to the statistical significance at 10, 5, and 1 percent levels, respectively.

Table 6: Quantile shift test 1

| | Sample | CDF value from 10,000 samples Monte Carlo | | | | |
|---|---|---|---|---|---|---|
| | | Unc | GJR | 2 regime w/jump | BEGE | BEGE,2 series |
| $Q_5^n - Q_5^u$ | -0.0220 | 0.0060** | 0.0065** | 0.0527 | 0.0857 | 0.0492* |
| $Q_{10}^n - Q_{10}^u$ | -0.0200 | 0.0000*** | 0.0004*** | 0.0015*** | 0.0223** | 0.0105** |
| $Q_{25}^n - Q_{25}^u$ | -0.0070 | 0.0022*** | 0.0199** | 0.0070** | 0.1097 | 0.0833 |
| $Q_{50}^n - Q_{50}^u$ | -0.0029 | 0.0703 | 0.0950 | 0.0700 | 0.0595 | 0.0614 |
| $Q_{75}^n - Q_{75}^u$ | 0.0025 | 0.8731 | 0.6907 | 0.7110 | 0.3627 | 0.4206 |
| $Q_{90}^n - Q_{90}^u$ | 0.0059 | 0.9889** | 0.8282 | 0.7935 | 0.5467 | 0.6326 |
| $Q_{95}^n - Q_{95}^u$ | 0.0130 | 0.9916** | 0.9557* | 0.8729 | 0.8609 | 0.9083 |
| $Q_5^p - Q_5^u$ | 0.0205 | 0.9995*** | 0.9975*** | 0.9969*** | 0.9923** | 0.9981*** |
| $Q_{10}^p - Q_{10}^u$ | 0.0128 | 1.0000*** | 0.9984*** | 0.9999*** | 0.9920** | 0.9963*** |
| $Q_{25}^p - Q_{25}^u$ | 0.0022 | 0.9174 | 0.7754 | 0.8897 | 0.7182 | 0.774 |
| $Q_{50}^p - Q_{50}^u$ | 0.0020 | 0.9497 | 0.9003 | 0.9402 | 0.9479 | 0.9706* |
| $Q_{75}^p - Q_{75}^u$ | -0.0007 | 0.2934 | 0.5258 | 0.4846 | 0.7861 | 0.7287 |
| $Q_{90}^p - Q_{90}^u$ | -0.0045 | 0.0057** | 0.1558 | 0.1124 | 0.2417 | 0.1522 |
| $Q_{95}^p - Q_{95}^u$ | -0.0048 | 0.0033*** | 0.2571 | 0.3178 | 0.4024 | 0.2945 |

The data are logged monthly dividend-adjusted stock returns from December 1925 to December 2010. $Q_i^n$ is the $i^{th}$ percentile of the $r_t$ distribution conditioning on $r_{t-1} < 0$. $Q_i^u$ is the $i-th$ percentile of the unconditional $r_t$ distribution. $Q_i^p$ is the $i^{th}$ percentile of the $r_t$ distribution conditioning on $r_{t-1} \geq 0$. In each Monte Carlo sample the time series of 1,020 observations (the same length as the historical time series) was generated by randomly drawing the errors and using the optimal parameters for each model. Next, the values of $Q_i^n - Q_i^u$ and $Q_i^p - Q_i^u$ were computed for that generated time series. Repeating the procedure 10,000 times yielded the null distributions of $Q_i^n - Q_i^u$ and $Q_i^p - Q_i^u$ under each model. The reported cumulative distribution function (CDF) values are probabilities that the value less or equal to the historical value is observed under the simulated distribution. $Unc$ refers to the case where the time series was formed by randomly sampling (with replacement) historical monthly returns. The asterisks, *,**, and *** correspond to statistical significance at 10, 5, and 1 percent levels, respectively.

Table 7: Quantile shift test 2

| | Sample | CDF value from 10000 samples Monte Carlo | | | | |
|---|---|---|---|---|---|---|
| | | Unc | GJR | 2 regime w/jump | BEGE | BEGE, 2 series |
| $Q_5^n - Q_5^u$ | -0.0611 | 0.0027*** | 0.0484* | 0.2617 | 0.0949 | 0.0453* |
| $Q_{10}^n - Q_{10}^u$ | -0.0488 | 0.0002*** | 0.0370* | 0.1468 | 0.0687 | 0.0271* |
| $Q_{25}^n - Q_{25}^u$ | -0.0359 | 0.0000*** | 0.0114** | 0.0057** | 0.0141** | 0.0035*** |
| $Q_{50}^n - Q_{50}^u$ | -0.0106 | 0.0106** | 0.0389* | 0.0153** | 0.0209** | 0.0140** |
| $Q_{75}^n - Q_{75}^u$ | 0.0072 | 0.9450 | 0.4461 | 0.2839 | 0.2203 | 0.2607 |
| $Q_{90}^n - Q_{90}^u$ | 0.0104 | 0.9681* | 0.2956 | 0.1048 | 0.1368 | 0.1951 |
| $Q_{95}^n - Q_{95}^u$ | 0.0221 | 0.9588* | 0.5605 | 0.1517 | 0.4581 | 0.5728 |
| $Q_5^p - Q_5^u$ | 0.0137 | 0.7802 | 0.9965*** | 0.9936** | 0.9395 | 0.9800** |
| $Q_{10}^p - Q_{10}^u$ | 0.0017 | 0.5709 | 0.9306 | 0.9233 | 0.7765 | 0.8762 |
| $Q_{25}^p - Q_{25}^u$ | 0.0020 | 0.6395 | 0.8973 | 0.8241 | 0.8441 | 0.8900 |
| $Q_{50}^p - Q_{50}^u$ | 0.0043 | 0.8356 | 0.7985 | 0.7002 | 0.8805 | 0.8549 |
| $Q_{75}^p - Q_{75}^u$ | 0.0010 | 0.5866 | 0.2456 | 0.1870 | 0.3275 | 0.2274 |
| $Q_{90}^p - Q_{90}^u$ | -0.0034 | 0.3004 | 0.0508 | 0.0586 | 0.0580 | 0.0232** |
| $Q_{95}^p - Q_{95}^u$ | -0.0059 | 0.2405 | 0.0303* | 0.0142** | 0.0406* | 0.0187** |

The data are logged monthly dividend-adjusted stock returns from December 1925 to December 2010. $Q_i^n$ is the $i^{th}$ percentile of the $r_t$ distribution conditioning on $r_{t-1} < \mu - \sigma$, where $\mu$ and $\sigma$ are unconditional mean and standard deviation of the returns time series. $Q_i^u$ is the $i^{th}$ percentile of the unconditional $r_t$ distribution. $Q_i^p$ is the $i^{th}$ percentile of the $r_t$ distribution conditioning on $r_{t-1} > \mu + \sigma$. In each Monte Carlo sample the time series of 1,020 observations (the same length as the historical time series) was generated by randomly drawing the errors and using the optimal parameters for each model. Next, the values of $Q_i^n - Q_i^u$ and $Q_i^p - Q_i^u$ were computed for that generated time series. Repeating the procedure 10,000 times yielded the null distributions of $Q_i^n - Q_i^u$ and $Q_i^p - Q_i^u$ under each model. The reported cumulative distribution function (CDF) values are probabilities that the value less or equal to the historical value is observed under the simulated distribution. $Unc$ distribution refers to the case where the time series was formed by randomly sampling (with replacement) historical monthly returns. The asterisks, *,**, and *** correspond to statistical significance at 10, 5, and 1 percent levels, respectively.

Table 8: Quantile hit test

| | 1% Hit ratio test | | |
|---|---|---|---|
| Model/Instrument | 1 | $hit_{t-1}$ | $r_{t-1}$ |
| GJR-GARCH | 8.8859*** | 150.6957*** | 5.6194** |
| 2-regime w/jump | 0.0734 | 4.0332** | 6.3024** |
| BEGE | 0.0046 | 2.7828* | 2.8697* |
| BEGE, 2 series | 0.3718 | 5.5149** | 4.7248** |
| | 5% Hit ratio test | | |
| Model/Instrument | 1 | $hit_{t-1}$ | $r_{t-1}$ |
| GJR | 3.2132* | 4.6906** | 7.8811*** |
| 2-regime w/jump | 3.7711* | 7.6508*** | 7.6541*** |
| BEGE | 1.4281 | 0.6922 | 4.9023** |
| BEGE, 2 series | 1.4281 | 0.0553 | 6.0696** |

Each cell reports EM hit test statistic as described in the text. The asterisks, *,**,and *** correspond to statistical significance at 10%, 5%, and 1% levels, respectively using the chi-squared distribution.

Figure 1: BEGE distribution tail properties
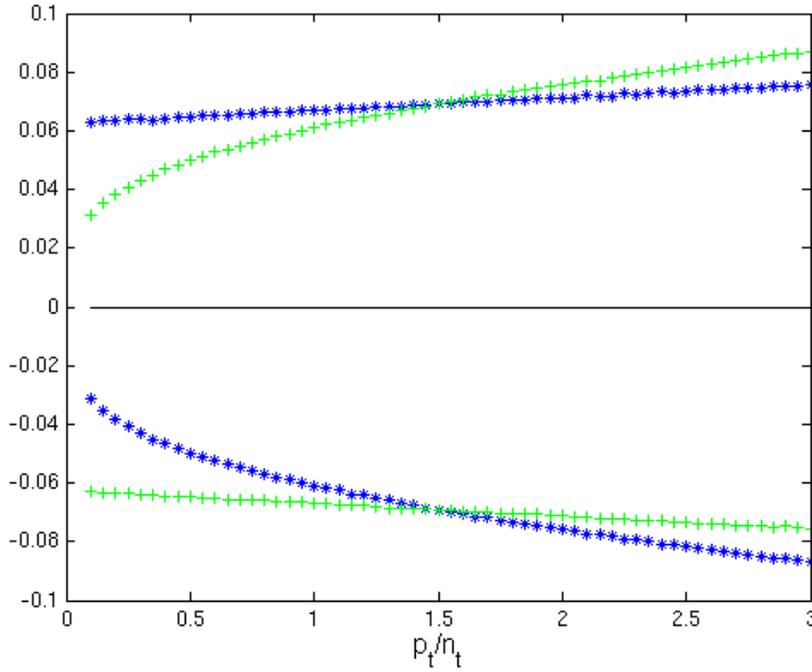


This plot shows the 99th percentiles and 1st percentiles for two sequences of BEGE distributions, which take the form.

$$
\begin{aligned}
u_{t+1} &= \omega_{p,t+1} - \omega_{n,t+1} \\
\omega_{p,t+1} &\sim \widetilde{\Gamma}\left(p_t, \sigma_p\right) \\
\omega_{n,t+1} &\sim \widetilde{\Gamma}\left(n_t, \sigma_n\right)
\end{aligned}
$$

where $\widetilde{\Gamma}$ denotes the centered gamma distribution. Throughout, we maintain that $\sigma_n = \sigma_p = 0.015$. The lines of blue asterisks show the quantiles for distributions in which $p_t$ is fixed at 1.5, but $n_t$ varies from 0.1 through 3.0. Conversely, the lines of green plus symbols show the quantiles for distributions in which $p_t$ varies from 0.1 through 3.0 while $n_t$ is held fixed at 1.5.
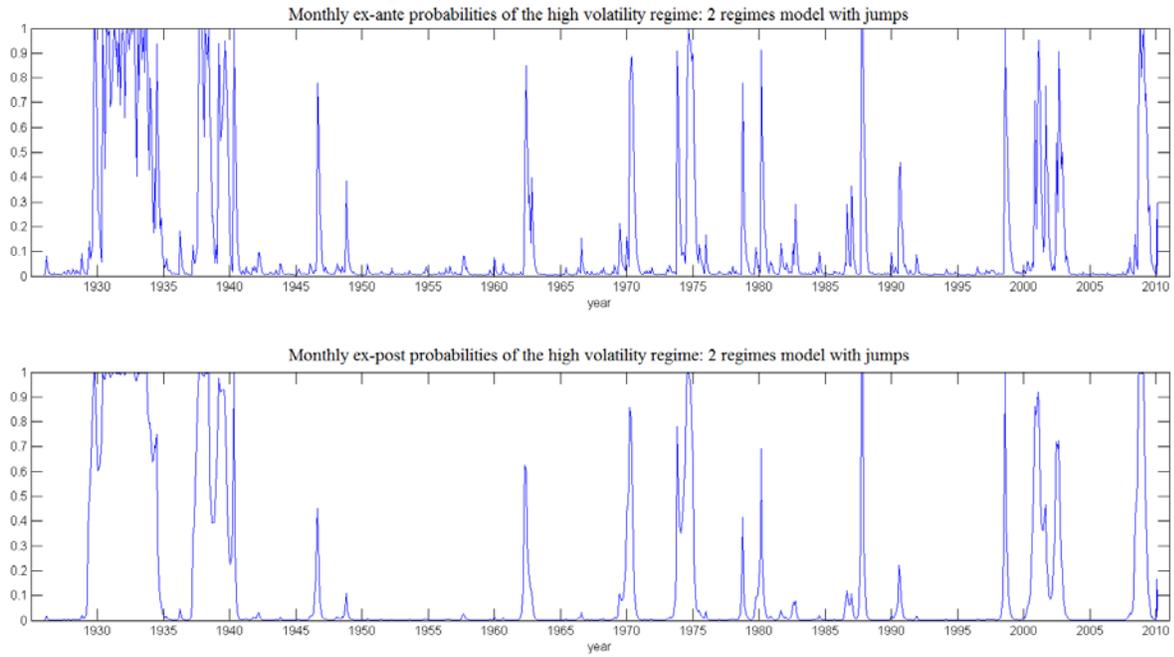
Figure 2: Results from GJR-GARCH estimation
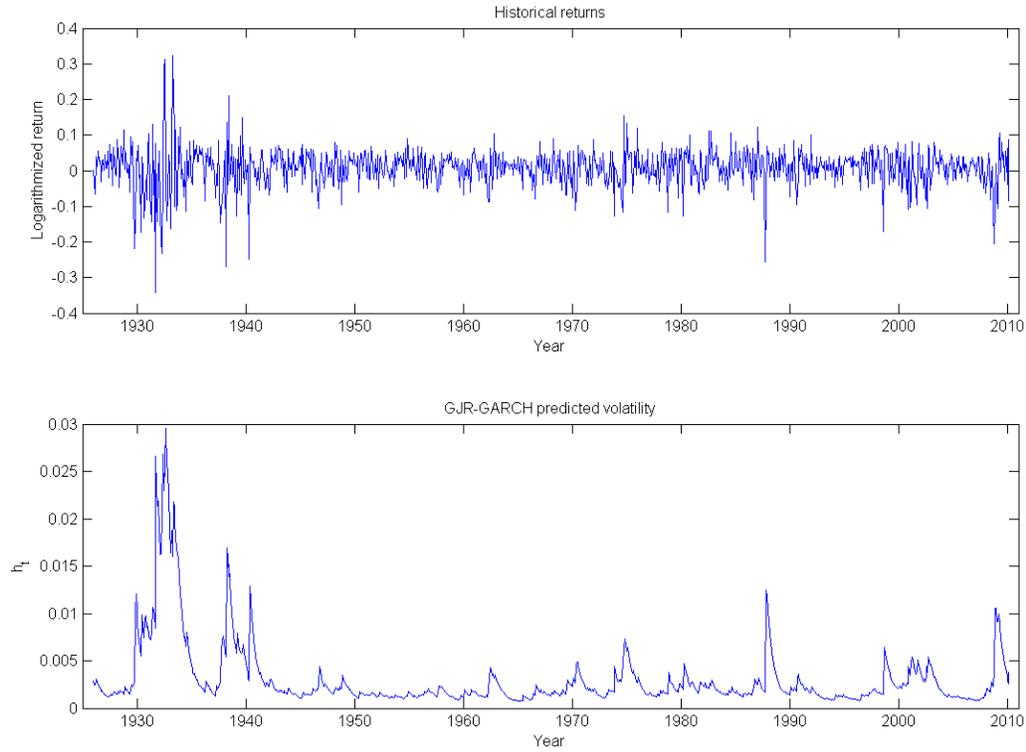


The data are logged monthly dividend-adjusted stock returns from December 1925 to December 2010. The top panel shows the raw return series. The bottom panel shows estimates of $h_t$ from the GJR-GARCH model, using the parameter estimates reported in Table 2.
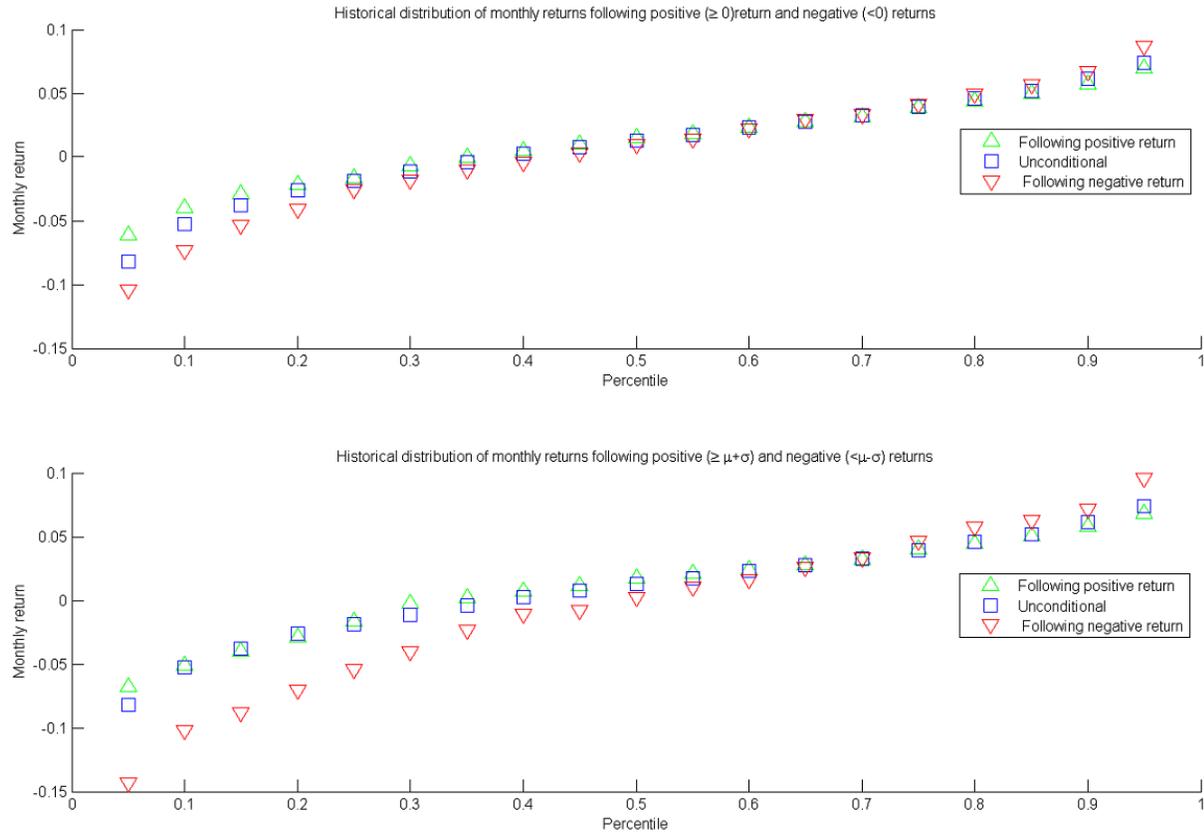
Figure 3: Results from BEGE estimation
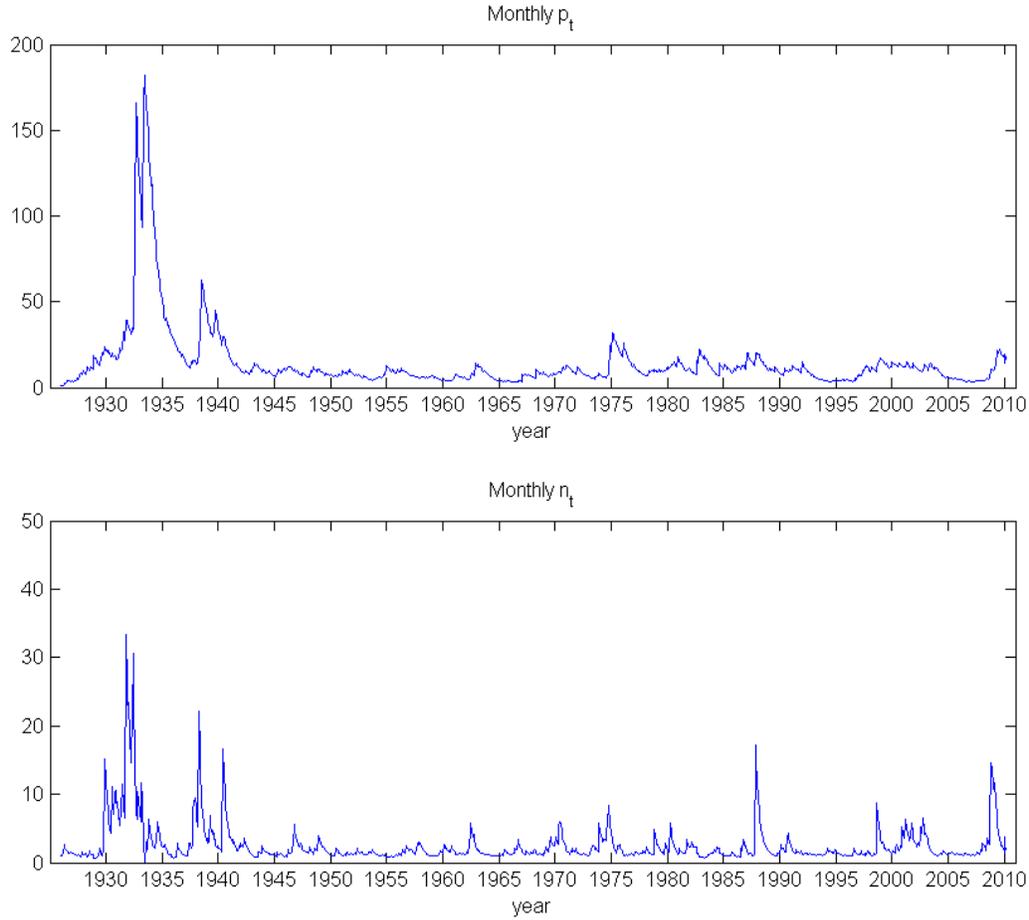


The top panel shows estimates of $p_t$ from the BEGE GJR (1-series) model, using the parameter estimates reported in table 2. The bottom panel shows estimates of $n_t$.

Figure 4: Results from regime–switching model estimation



The top panel shows the ex-ante probabilities of being in regime 2 for the 2-regime RS model with jumps using the parameter estimates reported on table 3. The bottom panel shows smoothed probabilities.

Figure 5: Quantile shift results



Historical distribution of monthly returns following positive (≥ 0)return and negative (<0) returns

Historical distribution of monthly returns following positive (≥ μ+σ) and negative (<μ-σ) returns

The data are logged monthly dividend-adjusted stock returns from December 1925 to December 2010. The top panel shows the values of unconditional quantiles (squares), quantiles for a restricted sample in which the return in the previous period was positive (up triangles) and quantiles for a restricted sample in which the return in the previous period was negative (down triangles). The bottom panel reports results further restricting the "positive" ("negative") subsample to those for which the previous return is one (unconditional) standard deviation above (below) the unconditional mean return.

Figure 6: Conditional quantile estimates
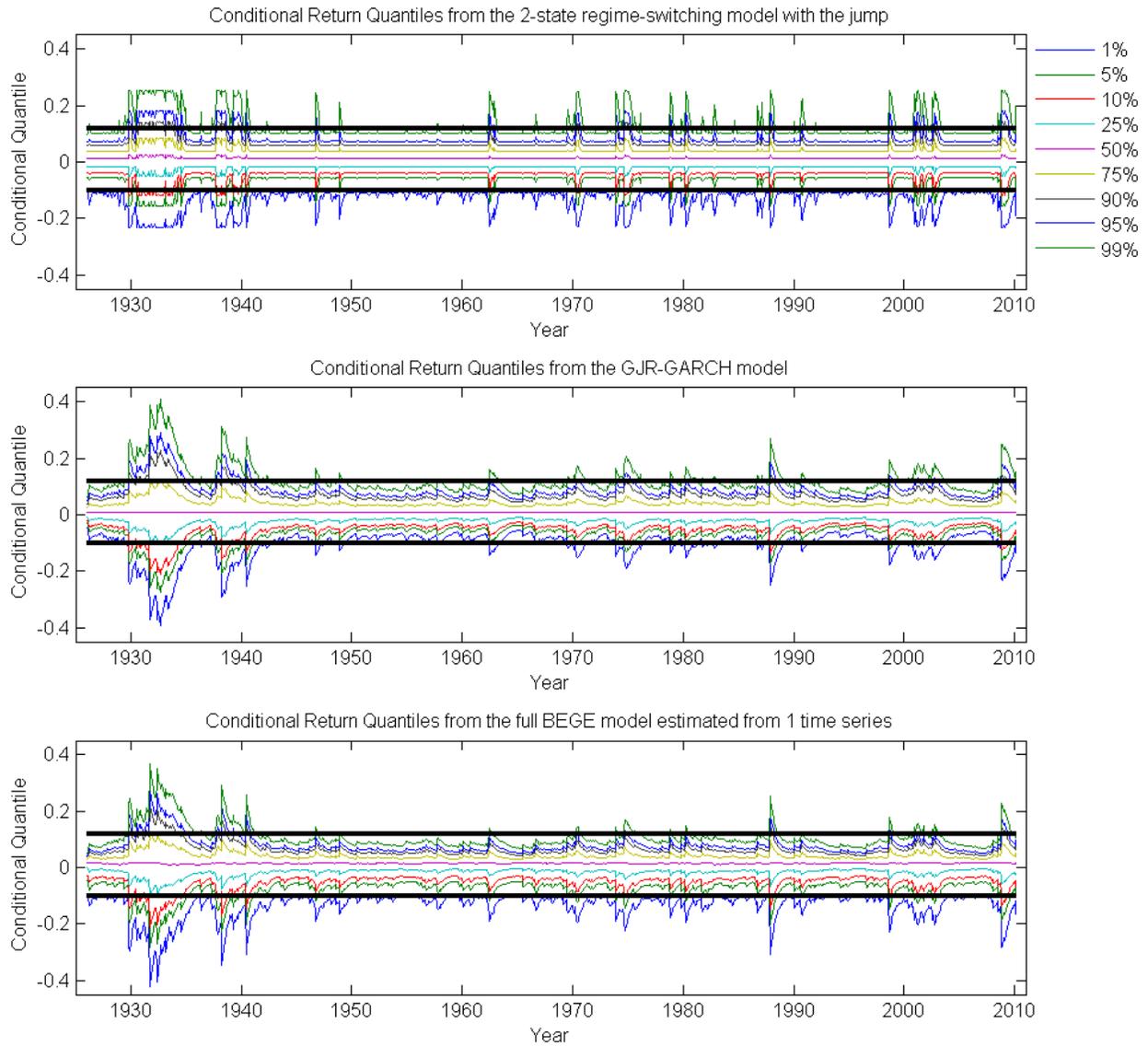


The figure reports estimates of conditional quantiles under the models specified in the panel headings.

Figure 7: Conditional moment impact curves
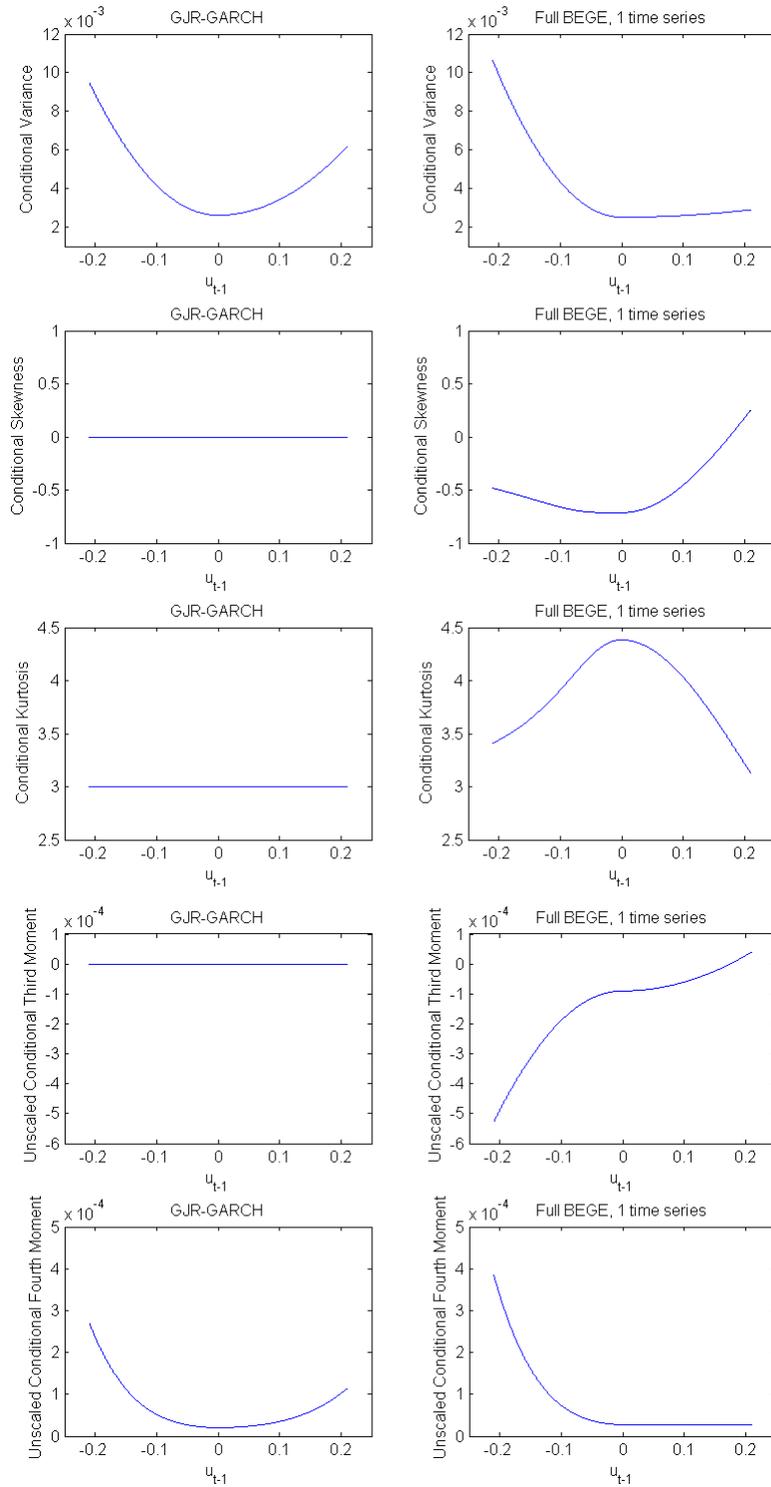


The news impact curves for the GJR and BEGE models. $p_{t-1}$ and $n_{t-1}$ are assumed to be equal to the average values of $p_t$ and $n_t$ over the observed time series.