6-2002

How should we interpret evidence of time varying conditional skewness?

Gamini PREMARATNE
National University of Singapore

Anthony S. TAY
Singapore Management University, anthonytay@smu.edu.sg

How Should We Interpret Evidence of Time Varying Conditional Skewness?

Gamini Premaratne and Anthony S. Tay*

Department of Economics,
National University of Singapore
1 Arts Link, Singapore 117570
Republic of Singapore

June 2002

Abstract:
Several recent articles report evidence of predictability in the skewness of equity returns, raising hopes that predictability in third moments will be useful for forecasting the probability of tail events. The evidence is unfortunately difficult to interpret, partly because they were obtained mainly from parametric models of time-varying conditional skewness, and because little is known about the behavior of such models, for instance, when there are outliers. We investigate a non-parametric approach to testing for predictability in skewness. Specifically, we explore the size and power of a Runs tests, and compare this approach with other tests. A re-examination of daily market returns reveals mild evidence of predictability in skewness. Incorporating this conditional heteroskewness into standard volatility models hardly improves out-of-sample forecasts of tail probabilities.

* Corresponding author: Tay, at email: anthonytay@nus.edu.sg, Tel: (65) 6874-3991, Fax: (65) 6775-2646.
Financial support from the National University of Singapore (NUS) Research Grant: R-122-000-052-112 "Conditional Skewness in Stock Returns" and assistance from the NUS Supercomputing and Visualization Unit is gratefully acknowledged.

1. Introduction

There is strong evidence that the standardized residuals from conditionally heteroskedastic models fitted to stock returns are asymmetrically distributed (Bai and Ng, 2001; Bera and Premaratne, 2001). Does the degree and extent of the asymmetry vary over time, and is this variation predictable? If there is predictability in the shape of the conditional distribution of stock returns, the implications for asset and derivative pricing and risk management are immediate. Under general assumptions about utility functions, the shape of the distribution of asset returns will be priced (Kraus and Litzenberger, 1976; Harvey and Siddique, 2000). Even if the first two moments are sufficient for asset pricing, variation in the shape of the distribution may affect the estimation of the conditional mean and conditional variance of an asset return, just as (non-varying) asymmetry affects the estimation of the conditional mean and variance (Newey and Stiegerwald, 1997; Bera and Premaratne, 2001). Likewise, the shape of the distribution of the underlying asset will in general affect option prices. Risk management practices often focus on the left tail of the distribution of future changes in the value of a portfolio (Duffie and Pan, 1997), and the probability of large negative changes in the value of a portfolio, for any given level of mean and volatility, may depend on the shape of the distribution.

The question of predictability in the shape of a variable's conditional distribution is usually framed in terms of predictability in conditional skewness, or 'conditional heteroskewness', with attention focusing primarily on predictability using the variable's past history. Recent investigations have uncovered some evidence of such predictability in stock returns. Specifying the conditional distribution of the standardized residuals of a GARCH-M model as a non-central t-distribution, with skewness depending on the conditional skewness in the previous period, Harvey and Siddique (1999) found evidence of autoregressive behavior in the conditional skewness of daily US, German and Japanese stock index returns. Chen, Hong and Stein (2001) used cross-sectional regressions of skewness in the daily stock returns of individual firms, measured over six month periods, and found that periods of high return and unusually high turnover tend to be followed by periods of negative skewness. Building on the Autoregressive Conditional Density (ARCD) model of Hansen (1994), Jondeau and Rockinger (2000) found evidence of predictability in the third moments of the daily returns of several major stock markets, but not at the weekly frequency (see also Rockinger and Jondeau, 2001). Using the ARCD framework, Hashmi and Tay (2001) found predictability in the skewness of weekly returns of a World and an Asia-Pacific stock index. Perez-Quiros and Timmermann (2001) found time-variation in the skewness of size-sorted portfolios US stocks. In an application to a daily French stock index return, El Babsiri and Zakoian (2001) found that a model with conditional heteroskewness, heterokurtosis, and leverage effects in volatility improves upon models without these effects.

The evidence of predictability in the skewness of stock returns is, however, difficult to interpret, particularly its implications for risk management. The majority of studies on this issue proceed by fitting a model that allows for predictability in skewness, and testing if the parameters that embody conditional heteroskewness are statistically significant (e.g., Harvey and Siddique, 1999; Rockinger and Jondeau, 2000; Hashmi and Tay, 2001). However, little is known about the behavior of models with time-varying conditional skewness. In particular, these models may not be robust to outliers. On the other hand, the models may not be able to pick up predictability in extreme realizations, even if predictability exists, as extreme realizations occur infrequently (Rockinger and Jondeau, 2002, pg 139). Another problem with parametric tests for predictability in skewness is the need to specify a law of motion for conditional skewness. Estimation of the parametric model is difficult, and usually also requires a distributional assumption, and either an incorrect specification of skewness dynamics or an incorrect distributional assumption may lead to faulty inferences.

We explore a non-parametric approach, involving the use of a Runs test, to testing for predictability in skewness. The following section explains why the Runs test is well suited for detecting this particular departure from randomness. In section 3, we use Monte Carlo experiments to evaluate the size and power properties of the Runs test under various situations. We also compare the Runs test with other tests for predictability in skewness. We report in Section 4 the results of an application of the tests to three daily stock index returns: the S&P 500, FTSE100 and Nikkei 225. We find mild evidence of predictability in skewness in two of the three series. We then investigate whether this predictability represents a systematic component in the shape of the distribution that may assist in forecasting tail events. We do so by generating out-of-sample one-step-ahead forecasts of tail probabilities using two sets of models, one with and the other without conditional heteroskewness, and compare the models' ability to forecast tail events.

2. A Runs Test for Predictability in Conditional Skewness

2.1 The Runs Test

The Runs test is a test designed to detect non-randomness in a sequence of observations of a binary variable, such as a sequence of 1's and 0's. A 'run' is defined as a string of consecutive 1's or 0's, and a run may consist of just one observation. The idea of the runs test is that non-randomness in a string of 1's and 0's often manifests itself as either too few runs, or too many runs. In a sequence of observations of a univariate continuous random variable, non-randomness can be detected using a Runs test, by assigning '1' to realizations that fall within some category (e.g., $y \geq 0$), and '0' otherwise. The distribution of the number of runs under the null of randomness has long been derived (see for instance Mood, 1940). Although the exact distribution is known, we will use the asymptotic result of Wald and Wolfowitz (1940): in a sequence of length n with n_1 1's and n_0 0's, the number of runs is asymptotically normal with mean

$$\mu = 1 + \frac{2n_0 n_1}{n}$$

and standard deviation

$$\sigma = \sqrt{\frac{2n_0 n_1 (2n_0 n_1 - n)}{n^2 (n-1)}}.$$

The Runs test has been used in a wide range of applications in numerous fields of research. We will not elaborate upon applications of the Runs test, except to note that in finance it has been used to test stock returns for departures from independence, e.g., Fama (1965), and in the econometric forecasting literature, the Runs test has been used to evaluate interval forecasts, e.g., Christoffersen and Diebold (2000).

Take a sequence of 1's and 0's $\{I_t\}_{t=1}^{n}$ to be indicators of the sign of the corresponding realizations $\{y_t\}_{t=1}^{n}$ of a random variable, say, stock returns. Any form of predictability in y_t that results in predictability in the probability of observing positive and negative values would result in non-randomness in $\{I_t\}_{t=1}^{n}$. For example, in an AR(1) process with strong positive autocorrelation and a symmetric conditional distribution there will be long periods of time during which the probability of observing one sign is greater than the probability of the opposite sign. This "clustering of probabilities" results in long runs of 1's and 0's in $\{I_t\}_{t=1}^{n}$, and consequently, of too few runs. This is the rationale behind the usual interpretation of a rejection of the Runs test as evidence of predictability in the mean (linear or otherwise). This same idea is explored in detail and used in Christoffersen and Diebold (2002) to study the relationship between market timing (i.e., predictability of the sign of returns) and volatility dynamics.

On the other hand, any form of dependence that leaves the conditional probability of a positive realization unchanged will result in randomness in I_t, since the conditional probability at any time t of observing a positive value would be the same as the marginal probability of a positive realization. If y_t is a process with zero conditional mean, I_t is random even if the process is conditionally heteroskedastic. If in addition the conditional distribution is symmetric, I_t will also be random under conditional heterokurtosis, since neither predictability in the second moments, nor in any of the higher even moments, would result in variation in the conditional probability of observing positive values of y_t. It is reasonable, therefore, to interpret a rejection of the Runs test, when applied to the signs of a zero-conditional mean process, as evidence of predictability in the variation in the degree of asymmetry present in the conditional distribution of y_t. Finally, the Runs test will be robust to the presence of outliers, since it focuses on the sign of the observations and not their size. We therefore propose the following non-parametric approach to testing for predictability in conditional skewness: first model the conditional mean, perhaps using a non-parametric regression technique, and generate the indicator sequence I_t of the signs of the residuals. Next, carry out the Runs test on I_t. This procedure can be repeated for different ways of modeling the conditional mean, to provide a more robust view regarding predictability in conditional skewness.

We focus on the Runs test, but of course, the Runs test can be replaced by other tests for forecastability of the signs. Probit models relating I_t to past residuals, for instance, might prove useful. There are also other ways to test for conditional heteroskewness that use the residuals themselves, rather than their signs. For instance, we could evaluate the statistical significance of a regression of $\hat{\varepsilon}_t^3$ on $\hat{\varepsilon}_{t-1}$, where $\hat{\varepsilon}$ represents the residuals after estimating and removing the conditional mean of the process. We can informally view this test as having a foundation based on an Information Matrix (IM) test; Bera and Lee (1993) show that the IM test statistic for a linear regression model with possibly autocorrelated Gaussian errors can be decomposed into several components, one of which tests for 'heteroclicticity' in a similar fashion. The size and power properties of the Runs test for predictability in skewness will be investigated using Monte Carlo experiments, but first we briefly discuss a parametric model with conditional heteroskewness. The properties of this approach will also be investigated and compared with the Runs test.

2.2 A Parametric Model for Predictability in Skewness

Another way to test for predictability in conditional skewness is to fit a model that allows for such a phenomenon, and test if the relevant parameters are statistically significant. An early time series model that allows for predictability in conditional skewness is the Autoregressive Conditional Density (ARCD) model of Hansen (1994). Building on the standard setup of a conditionally heteroskedastic model, Hansen (1994) specifies:

$$r_t = \mu(r_{t-1}, \theta) + \varepsilon_t,\ \varepsilon_t = \sigma_t z_t, \tag{2.1}$$

$$\sigma_t^2 = \sigma^2(x_t, \theta) \tag{2.2}$$

$$z_t \sim g\left(z_t \middle| \lambda_t, \eta_t\right) \tag{2.3}$$

where $g\left(z_t \mid \lambda_t, \eta_t\right)$ is a skewed t-distribution standardized to have zero mean and unit variance:

$$g\left(z_t \mid \lambda_t, \eta_t\right)=\begin{cases} b_t\, c_t\left(1+\dfrac{1}{\eta_t-2}\left(\dfrac{b_t\, z_t+a_t}{1-\lambda_t}\right)^2\right)^{-\frac{\eta_t+1}{2}} & \text{when } z_t<-a_t / b_t \\ b_t\, c_t\left(1+\dfrac{1}{\eta_t-2}\left(\dfrac{b_t\, z_t+a_t}{1+\lambda_t}\right)^2\right)^{-\frac{\eta_t+1}{2}} & \text{when } z_t \geq -a_t / b_t \end{cases} \tag{2.4}$$

with $a_t=4 \lambda_t\, c_t\left(\dfrac{\eta_t-2}{\eta_t-1}\right)$, $b_t^2=1+3 \lambda_t^2-a_t^2$, and $c_t=\dfrac{\Gamma\left(\dfrac{\eta_t+1}{2}\right)}{\sqrt{\pi\left(\eta_t-2\right)}\, \Gamma\left(\dfrac{\eta_t}{2}\right)}$. The "parameters" λ_t and η_t determine the shape of the distribution. If $\lambda_t=0$ the distribution reduces to the standardized t distribution with degrees of freedom η_t, otherwise the distribution is skewed to the left when $\lambda_t<0$ and skewed to the right when $\lambda_t>0$. The parameter λ_t is restricted to lie within the interval (–1.0, 1.0), and a logistic transformation of λ_t is usually applied to achieve this effect. Predictability in the parameters λ_t and η_t is achieved by allowing them to be driven by past shocks, much in the same way that past shocks drive variation in the conditional variance. As the skewness and kurtosis are functions of λ_t and η_t, this framework generates predictability in the third and fourth conditional moments. The properties of this model and applications using it can be found in Hansen (1994), Jondeau and Rockinger (2000) and Hashmi and Tay (2002). In principle, the skewed t-distribution can be substituted by any non-symmetric distribution, including the non-central t-distribution (Harvey and Siddique, 1999), Pearson Type IV and Log-Gamma (Brannas and Nordman, 2001) and predictability in skewness (and kurtosis) induced by specifying the relevant parameters to vary with past information. In an interesting and significant departure from this framework, Rockinger and Jondeau (2002) model time-varying skewness and kurtosis, and estimate entropy densities to match the skewness and kurtosis at each point in time.

Once these models are fitted to the data, it is a simple matter to test for the presence of predictability in conditional skewness (or higher moments) by testing if the relevant parameters are significantly different from zero. For instance, if the Hansen (1994) framework is applied, and the skewness equation is specified as

$$\lambda_t = \beta_0 + \beta_1 z_{t-1} \tag{2.5}$$

so that skewness in the conditional distribution is driven by past standardized residuals, then the hypothesis of no predictability of conditional skewness is simply the hypothesis $H_0 : \beta_1 = 0$, which can be tested using a Wald test or a Likelihood Ratio test.

3. Monte Carlo Analysis

In this section, we report the results of Monte Carlo experiments to evaluate the size and power of the Runs Test. We begin with size, in relation to alternatives with outliers, and with dependence in the second and fourth moments. We use 1000 replications in each experiment, and the tests are evaluated over two sample sizes of 500 and 2000 observations, roughly corresponding to two years of daily data or ten years of weekly data, and 8 years of daily data, the intention being to study the behavior of the tests for small and for reasonably large sample sizes. In all cases, our DGPs will be zero-conditional mean processes, so the Runs test is applied directly to the pseudo-data, without any need to model a conditional mean process.[1]

3.1 Size

To check for robustness of the Runs test to the presence of random outliers, we apply the test to simulated data with outliers. Pseudo-data with outliers are generated in several ways:

a. $y_t \overset{iid}{\sim} t(5)$,

b. $y_t \overset{iid}{\sim} g(\lambda = -0.5, \eta = 5)$ where $g(\lambda, \eta)$ is as in (2.4),

[1] All computations in this section were carried out in MATLAB. Maximum Likelihood Estimation of the parametric model in the next subsection uses the MATLAB numerical optimization routine *fminunc,* and robust standard errors are used.

c. $y_t \overset{iid}{\sim} t(5)$ with one outlier (–10 standard deviations) per 500 observations, randomly located, and

d. $y_t \overset{iid}{\sim} t(5)$ with two outliers (–10 standard deviations) per 500 observations, randomly located.

The first DGP specifies an iid process with fat-tails. The second DGP adds leftward skewness so that the probability of negative outliers is increased, and the likelihood of positive outliers reduced. The last two DGPs, like the first one, specifies an iid t(5) process, but adds one and two very large outliers per 500 observations respectively. The locations of these outliers were determined by drawing uniformly distributed numbers over the sample range. The results of this experiment are displayed in Table 1. The results are not surprising; in all four cases, the Runs test is perfectly sized.

Conditionally heteroskedastic and conditionally heterokurtic (i.e., predictable conditional kurtosis) alternatives are generated using the ARCD model given in (2.1) - (2.4). In all cases, we set $r_t = \varepsilon_t$, $\varepsilon_t = \sigma_t z_t$, so we have zero-conditional mean. For conditionally heteroskedastic alternatives, we set $\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \max\left(0, \varepsilon_{t-1}\right)^2 + \alpha_3 \sigma_{t-1}^2$, $\lambda_t = \lambda$ and $\eta_t = \eta = 5$. The form of the variance equation is that of Glosten, Jagannathan and Runkle (1993) which allows for positive and negative shocks to affect future volatility differently; setting $\alpha_2 = 0$ gives the usual GARCH specification. $\lambda = 0$ gives us the usual conditionally student-t GARCH process. We also include a DGP with $\lambda = -0.5$. For conditionally heterokurtic alternatives, we set $\sigma_t^2 = \sigma^2 = 1$, $\lambda_t = 0$ and $\eta_t = f(\eta_t')$, $\eta_t' = \gamma_0 + \gamma_1 z_{t-1} + \gamma_2 \eta_{t-1}'$ where $f(.)$ is a logistic transformation to restrict η_t to the interval (3,50). The empirical size of the Runs test for these alternative at various parameter values are given in Table 2; other parameter values were also used, with similar results. Again, the empirical size of the Runs test is close to that of the nominal size of the test, highlighting the robustness of the Runs test to dependence in even moments.

3.2 Power

To ascertain the power of the Runs tests, a second set of experiments is run using DGPs with predictability in conditional skewness. For this purpose, we again simulate data from the Hansen ARCD model as specified in equations (2.1) to (2.4), setting the mean and variance equations to be constant as in the conditionally heterokurtic case. Again, we set $\eta_t = \eta = 5$, but specify a law of motion for λ_t. Specifically, we specify $\lambda_t' = \beta_0 + \beta_1 z_{t-1}$, setting β_0 and β_1 to various values, and use the logistic transformation $\lambda_t = -0.99 + \dfrac{1.98}{1+\exp(\lambda_t')}$ to restrict the asymmetry parameter to the unit interval. The values of β_0 and β_1 were chosen to control the range and variability of λ_t, noting that as $\lambda_t' = \beta_0 + \beta_1 z_{t-1}$ behaves as a random variable with mean β_0 and variance β_1^2 (since z_{t-1} is a zero mean unit variance process). The values chosen are combinations of $\beta_0 \in \{-0.6, 0,\ 0.6\}$, corresponding approximately to an average value of –0.3, 0 and 0.3 respectively for λ_t (which in turn corresponds to an average value of the skewness coefficient of about –1.0, 0, and 1.0 respectively), and $\beta_1 \in \{-0.5,\ 0.5\}$, so there are altogether six different DGPs. In the case $\beta_0 = 0$, the values of β_1 implies a range of λ_t of approximately -0.5 to 0.5. These ranges can be inferred from the relations between λ and η and the skewness and kurtosis coefficients given in Jondeau and Rockinger (2000). Figure 1 displays a simulated series from the ARCD for one of these sets of parameter values.

The results of this experiment are shown in Table 3, and demonstrates the power of the Runs test to detect predictability in conditional skewness. Two observations stand out, the first is that power appears to be influenced by the value of β_0. Specifically, the test is most powerful when $\beta_0 = 0$, and power can fall by over twenty points when β_0 is set at 0.6 or –0.6. This corresponds with the analysis of David (1947) who demonstrated that the Runs test becomes less powerful as the relative frequency of 1's and 0's becomes uneven. As we pointed out earlier, in our DGPs the value of β_0 is the average value of λ_t, and therefore corresponds with the unconditional skewness of the data, which in turn influences the relative probability of observing 1's and 0's. The results in Table 3 also emphasizes the need for a fairly large sample size in order to achieve reasonable power.

3.3 A Comparison with Other Tests

We repeat the experiments using the two other approaches discussed in section 2.2, namely a parametric test, and the test of the significance of the regression of y_t^3 on y_{t-1}; we denote the latter test by ρ_{31}. The parametric test is carried out by first estimating the model described in equations (2.1) - (2.4), assuming $\lambda_t = \beta_0 + \beta_1 z_{t-1}$ with λ_t again restricted to the $(-1,1)$ interval using a logistic regression. Other skewed distributions can be used in place of (2.4) although we expect the results to be similar for the other specifications. We then test for the presence of predictability of conditional skewness by testing the statistical significance of β_1. Starting values in our numerical optimization procedures are always set with the presumption that there is no predictability in skewness, and no unconditional skewness, i.e., we set the initial values of both β_0 and β_1 to zero. To focus on predictability in conditional skewness, we set $\eta_t = \eta$ to be constant, and allow only for λ_t to vary over time. The mean and variance equations are set at constants $\mu(x_t) = \mu$ and $\sigma_t^2(x_t) = \sigma^2$.

The results are displayed in Table 4a. To improve the readability of the table we left out the results for nominal size 0.01, and only report the results of a small subset of specifications. The results in panel (B) indicate that outliers cause the parametric test to be oversized, but the effect is mild. The numbers in panel (A) show that both tests have power to detect predictability in skewness; the ρ_{31} test having similar power to the Runs test. The parametric test is much more powerful than both the Runs test and the ρ_{31} test . It should be borne in mind, however, that the experiments are heavily stacked in favor of the parametric test, since we have assumed in our estimation model both the correct distribution and the correct specification for the asymmetry equation. Some power is naturally sacrificed when moving from a parametric test to a non-parametric test, in return for not having to make any distributional and specification assumptions.

The cost of misspecification is highlighted by the results in panel (C) and (D) of Table 4a. Here, the parametric test and the ρ_{31} test are applied to conditionally heteroskedastic and conditionally heterokurtic DGPs. Note that for the parametric test we have assumed constant values for the variance and degrees of freedom parameters. When leverage effect is present, the parametric test is severely oversized. This is consistent with the Harvey and Siddique (1999) finding that incorporating conditional heteroskewness sometimes reduces leverage effects in conditional variance. The ρ_{31} test is also oversized when there is conditionally heteroskedasticity. The parametric test performs better when the misspecification is removed. In Table 4b, the parametric test is re-evaluated for conditionally heteroskedastic DGPs but without the misspecification in the estimation model. The empirical size of the test is much closer to the nominal size.

4. Application

We now test for predictability in the conditional skewness of stock market returns, and check if this predictability can assist in the prediction of tail events. We study daily returns on three stock market index returns, namely the FTSE100, the Nikkei 225 and the S&P 500, over the period Jan, 1984 to Dec 2001, giving a sample size of 4,695 observations. The data is obtained from Datastream.

4.1 Predictability in Conditional Skewness

The Runs test for conditional heteroskewness requires the series being tested to have zero-conditional mean. While our three series all show little correlation, we take care to account for possible non-linear dependence in the data. We do so in a number of ways, using parametric and non-parametric methods, and apply the Runs test, the parametric test and the ρ_{31} test to the residuals after the removal of the conditional mean. We also apply the tests to the returns series itself. The original returns series is labeled as series (a) in what follows. In series (b) and (c) we model the conditional mean using an ARMA – EGARCH-in-mean specification and an ARMA(2,2) specification respectively. For series (d) a nonlinear specification was used for the conditional mean: the returns were regressed on two lags of returns, their squares, and an interaction term between the two lagged returns. For series (e), the conditional mean is estimated non-parametrically, using a multivariate version of the local linear regression smoother as described and studied in Fan (1992), see also Fan and Gijbels (1996). Denoting the observations by $\left\{\left(y_t, x_{1t}, \ldots, x_{kt}\right)\right\}_{i=1}^{T} = \left\{\left(y_t, \mathbf{X}_t\right)\right\}_{i=1}^{T}$, the local linear approximation of the conditional mean of y at $\mathbf{x}$ is obtained as $\hat{\delta}_0$ in the weighted regression

$$\min \sum_{i=1}^{T} \left\{ y_i - \delta_0 - \sum_{j=1}^{k} \delta_j \left(x_{ij} - x_j \right) \right\}^2 K_B \left(\mathbf{X}_i - x \right) \tag{4.1}$$

where $K_B(.)$ is a multivariate kernel density $K_B(\mathbf{u}) = \frac{1}{|B|} K\left(B^{-1}\mathbf{u}\right)$ with bandwidth matrix B. We use the same predictors as in series (d), and use a multivariate gaussian kernel with a diagonal covariance matrix. The bandwidth is taken as $B = h\mathbf{I}$, with the constant h selected by cross-validation. The multivariate local polynomial regression is closely related to the *LOESS* procedure of Cleveland and Devlin (1988) and Cleveland, Grosse and Shyu (1992). Fan (1992) demonstrates that the performance of the local polynomial regression technique is superior to that of other non-parametric estimators when the regressors are highly clustered, such as in our application. The local polynomial regressor also does not require boundary modifications to achieve higher rates of convergence at the boundary of the support of $f(\mathbf{X})$.

The parametric test uses the ARCD model with the Hansen skewed-t distribution in (2.4) using various specifications for the variance as well as the asymmetry equation. Only the results for the best fitting model is reported. In all cases, the variance equation uses the Glosten et al (1993) specification. The asymmetry equation varies between

$$\lambda_t = \beta_0 + \beta_1 z_{t-1} \tag{4.2}$$

$$\lambda_t = \beta_0 + \beta_1 z_{t-1} + \beta_3 \lambda_{t-1} \tag{4.3}$$

$$\lambda_t = \beta_0 + \beta_1 z_{t-1} + \beta_2 z_{t-1}^2 + \beta_3 \lambda_{t-1} \tag{4.4}$$

The specifications in (4.3) and (4.4) can be described as allowing for "autoregressive conditional skewness". The lagged λ term is included only if the coefficient of either z_{t-1} or z_{t-1}^2 is significant. The parametric test for predictability in conditional skewness is a Wald test of the null hypothesis $H_0 : \beta_j = 0 \; \forall \; j \neq 0$.

The results are shown in columns (i), (ii) and (iii) of Table 5. The Runs test shows mild evidence of predictable conditional skewness in the FTSE100 returns series: for all five versions of the FTSE100 returns series, the Runs test rejects at 10% the null hypothesis that there is no predictability in skewness, a conclusion that the ρ_{31} test agrees with in all but one case. For the NK225 series, the Runs test does not detect any predictability, and again the ρ_{31} test is in general agreement. The results for the S&P500 series are more difficult to interpret. The Runs test gives conflicting results. The ρ_{31} test however, strongly rejects the null. Quite strikingly, the Wald test for every series and in every case indicates strongly the predictability of skewness. Our experiments have demonstrated that misspecifications can result in greater likelihood of a false rejection of the null in the parametric and ρ_{31} tests, and our results may be a reflection of possible specification problems in our parametric models. The presence of outliers such as the October 1987 crash may have also affected the results, although as we have seen, the effects of outliers on the parametric test is small, and outliers should not effect the ρ_{31} test. It seems more likely, given the sample skewness coefficients in column (iv) of Table 5, that in the case of the S&P500 series the Runs test may lack power; in our experiments, the power for the Runs test was found to be strongly influenced by the degree of skewness in the unconditional distribution, and the S&P500 returns series are strongly negatively skewed. Our conclusion is that there appears to be predictability in conditional skewness in the FTSE100 returns series, and very likely in the S&P500 returns series, but we find no evidence in the Nikkei 225 returns series. Certainly, the occurrence of predictability in third moments does not appear to be anywhere as strong or as widespread as predictability in variance.

4.2 Forecasting Tail Events

We next ask if the predictability in skewness is nonetheless useful in forecasting tail probabilities. We answer this question by comparing out-of-sample forecasts from two sets of models, one with and the other without conditional heteroskewness. We split the sample into two parts, a 'estimation sample' from Jan, 1984 to Dec, 1995 (3130 observations), and a 'forecast sample' from Jan, 1996 to Dec, 2001 (1567 observations). We re-estimate the GARCH and GARCH-ARCD models to all fifteen series over the estimation sample. Again, we use the Glosten, Jaganathan and Runkle (1993) specification for the variance equation, allowing for different reactions of volatility to negative and positive shocks. Using the best fitting estimated models, one-step ahead probability forecasts of the event $\Pr\left(y_t < -2.5\right)$ are generated over the forecast sample period. We also generate the series of indicator variables defined as

$$I_t = \begin{cases} 1 & \text{if } y_t < -2.5 \\ 0 & \text{otherwise} \end{cases} \tag{4.5}$$

As there appears to be only mild conditional heteroskewness in our data, we might also expect the tail probability forecasts to be similar across our models. This is indeed the case. Figure 2(a) plots the tail probability forecasts for the S&P500 returns series, without removing the conditional mean. While there is substantial variation in the probability forecasts, the two series are difficult to distinguish visually; their correlation is 0.957! Figure 2(b) plots the difference between the GARCH and the ARCD forecasts. The differences are mostly small, although the differences are larger in volatile periods. Nevertheless, it does appear that conditional heteroskewness plays a much smaller role in forecasting tail probabilities than conditional heteroskedasticity. Similar statements can be made about all fifteen series.

To compare the forecasts more formally, we use Probit regressions of the form

$$\Pr\left(I_t = 1\right) = \Phi\left(\delta_0 + \delta_1 \hat{p}_{it}\right),\ i = 1,2 \tag{4.6}$$

where $\hat{p}_{1t}$ and $\hat{p}_{2t}$ are the out-of-sample probability forecasts from the constant conditional skewness model ('GARCH') and the heteroskewness model ('GARCH-ARCD') respectively. The estimated coefficient $\hat{\delta}_1$ would give an indication of how well the probability forecasts match the actual occurrence of a realization of y_t below -2.5. Our method is similar in spirit to the Mincer-Zarnowitz (1969) regression, and is adapted for use with binary dependent variables. We compare the GARCH and GARCH-ARCD forecasts by comparing the value of the McFadden R^2:

$$R^2 = 1 - \frac{\log l\left(\hat{\delta}_0\right)}{\log l\left(\hat{\delta}_0 + \hat{\delta}_1 \hat{p}_{it}\right)}$$

for the two sets of probit regressions, and view a larger R^2 as evidence of improved probability forecasts.

The results of the comparison is shown in Table 6. Several features stand out. First, The R^2 values for the NK225 series are identical across the two models for all but one series, a result which corresponds with our findings from the previous section. For the FTSE100 and SP500 returns series, the R^2 values for the GARCH-ARCD forecasts are all larger than the corresponding values for the GARCH forecasts, which suggests that incorporating third moment dynamics into our forecasting models improve forecasts of tail probabilities[2]. The differences are all, however, very small, which suggests that the improvements are small. While this may be a reflection of the restrictions imposed in our forecast models (e.g., constant conditional kurtosis) and the specific distributional assumption made, it appears that the predictability in conditional skewness, based on the past returns, has little to add to forecasts of tail probabilities using standard volatility models

5. Concluding Remarks

We presented and studied a non-parametric approach to testing for predictability in conditional skewness. A comparative evaluation of the test with some alternative procedures demonstrated that the Runs test has good power while being robust to outliers, and other forms of dependence. In an application

[2] We note also that the coefficients on the probability forecasts in the probit regressions are all significant.

of the test to three stock market indexes we find that while there is some evidence of predictability in conditional skewness, this evidence is mild. An out-of-sample forecast exercise suggests that this predictability improves upon forecasts of tail probabilities from conventional volatility forecasts, but that the improvements are small and insignificant.

To conclude, we mention several caveats and highlight a few avenues for further research. First, our analysis has focused on daily returns, which may be less relevant to risk managers than returns over longer periods. The analysis of Christoffersen and Diebold (2000), however, suggests that even volatility models may have little to add to unconditional forecasts at lower frequencies, so it is unlikely that an analysis at lower frequencies would provide a different conclusion. Our use of probability forecasts deviates from the usual focus on quantile forecasts in risk management applications, but allows us a straightforward way of comparing our models. Alternatively, one could extend the Value-at-Risk comparison test developed in Christoffersen, Hahn and Inoue (2001) to include models with conditionally heteroskewness to evaluate quantile forecasts from our models, an activity we leave to future research.

Our study is also limited to forecastability using only the history of returns, so while our evidence suggests only weak predictability, this does not preclude that there may be other variables that would explain or forecast the shape of the distribution, and in turn provide more informative forecasts of tail events. Our conclusions are also limited to three stock indexes. A further analysis on a wider range of stocks and stock index returns, and other financial assets, would also be of interest.

Finally, we note that perhaps the most limiting aspect of our analysis is the strict (and intentional) focus on conditional heteroskewness. We look forward to continued investigations into the nature of higher order moment dynamics and its implications for asset and options pricing, risk-management, and for market timing, as noted by Christoffersen and Diebold (2002).

References:

Bai, J. and Ng, S. (2001), "A Consistent Test for Conditional Symmetry in Time Series Models", *Journal of Econometrics* 103, 235-258.

Bera, A. and Premaratne, G. (2001), "Modeling Asymmetry and Excess Kurtosis in Stock Return Data", manuscript, University of Illinois.

Bera, A. and Lee, S. (1993) "Information Matrix Test, Parameter Heterogeneity and ARCH: A Synthesis", *Review of Economic Studies* 60, 229-240.

Brannas, K. and Nordman, N. (2001), "Conditional Skewness Modelling for Stock Returns", manuscript, Umea University.

Chen, J., Hong, H. and Stein, J. C. (2001) "Forecasting crashes: trading volume, past returns and conditional skewness in stock prices", *Journal of Financial Economics* 61, 345-381.

Christoffersen, P. F. and Diebold, F.X. (2000), "How Relevant is Volatility Forecasting for Financial Risk Management?", *Review of Economics and Statistics* 82, 12-22.

Christoffersen, P. F. and Diebold, F. X. (2002) "Financial Asset Returns, Market Timing, and Volatility Dynamics", manuscript, McGill University and University of Pennsylvania.

Christoffersen, P.F., Hahn, J. and Inoue, A. (2001), "Testing and Comparing Value-at-Risk Measures", *Journal of Empirical Finance* 8, 325-342.

Cleveland, W. S. and Devlin, S. J. (1988), "Locally-weighted regression: An Approach to Regression Analysis by Local Fitting", *Journal of the American Statistical Association* 83, 597-610.

Cleveland, W. S., Grosse, E., and Shyu, W. M. (1992) "Local regression Models" in Chambers, J.M. and Hastie, T. J. (eds), *Statistical Models in S*, 309-376. Wadsworth & Brooks, California.

David, F. N. (1947) "A Power Function for Tests of Randomness in a Sequence of Alternatives", *Biometrika* 34, 3/4, 335-339.

Duffie, D. and Pan J. (1997) "An Overview of Value at Risk", *Journal of Derivatives* 4, 7-49.

El Babsiri, M. and Zakoian, J. M. (2001), "Contemporaneous Asymmetry in GARCH processes", *Journal of Econometrics* 101, 257-294.

Fama, E. (1965) "The Behavior of Stock Market Prices", *Journal of Business* 38. 34-105.

Fan, J. (1992) "Design-Adaptive Nonparametric Regression", *Journal of the American Statistical Association* 82, 998-1004.

Fan, J. and Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications*, Chapman & Hall, London.

Glosten, L. R., Jaganathan, R., and Runkle, D. E., 1993. "On the relation between the the expected value and the volatility of the nominal excess return on stocks" *Journal of Finance* 48, 1779-1802.

Hansen, B. E., 1994. “Autoregressive Conditional Density Estimation” *International Economic Review* 35, 705-730

Hashmi, A. R. and Tay, A. S. (2001), “Global and Regional Sources of Risk in Equity Markets: Evidence from Factor Models with Time-Varying Conditional Skewness”, Department of Economics Working Paper No. 0116, National University of Singapore.

Harvey, C. R. and Siddique, A. (1999), “Autoregressive Conditional Skewness”, *Journal of Financial and Quantitative Analysis* 34, 465-487.

Harvey C. R. and Siddique, A. (2000) “Conditional skewness in asset pricing tests”, *Journal of Finance* 55, 1263-1295.

Jondeau, E. and Rockinger, M. (2000) “Conditional volatility, skewness, and kurtosis: existence and persistence”, manuscript, HEC-School of Management.

Kraus, A. and Litzenberger, R. H. (1976) “Skewness Preference and the Valuation of Risk Assets”, *Journal of Finance* 31, 1085- 1100.

Mincer, J. and Zarnowitz, V. (1969), “The Evaluation of Economic Forecasts” in J. Mincer (ed.) *Economic Forecasts and Expectations*. NBER, New York.

Mood, A. M. (1940) “The Distribution Theory of Runs”, *Annals of Mathematical Statistics* 11, 4, 367-392.

Newey, W. K. and Steigerwald, D. G. (1997) “Asymptotic Bias for Quasi-Maximum-Likelihood Estimators in Conditional Heteroskedasticity Models”, *Econometrica*, **65**, 3, 587-599.

Perez-Quiros, G. and Timmermann, A. (2001) “Business cycle asymmetries in stock returns: evidence from higher order moments and conditional densities”. *Journal of Econometrics* 103, 259-306.

Premaratne, G. and Bera, A. (2001) “A Test for Asymmetry with Leptokurtic Financial Data”, Manuscript, University of Illinois.

Rockinger, M. and Jondeau, E. (2002) “Entropy Densities with an Application to Autoregressive Conditional Skewness and Kurtosis”, *Journal of Econometrics* 106, 119-142.

Wald, A. and Wolfowitz, J. (1940), “On a Test Whether Two Samples are from the Same Population”, *Annals of Mathematical Statistics* 11, 2, 147-162.

Table 1
Empirical Size of Runs Test (Outlier Alternatives) [a]

		Nominal Size		
DGP	Sample Size	0.01	0.05	0.10
iid t(5)	500	0.014	0.062	0.130
	2000	0.013	0.060	0.112
iid Hansen-t with $\lambda = -0.5, \eta = 5$	500	0.004	0.048	0.100
	2000	0.009	0.047	0.096
iid t(5) 1 outlier [b] per 500 obs	500	0.016	0.067	0.128
	2000	0.015	0.062	0.114
iid t(5) 2 outliers [b] per 500 obs	500	0.015	0.075	0.126
	2000	0.013	0.060	0.105

[a] Frequency of rejection of the null hypothesis that there is no predictability in conditional skewness, measured over 1000 replications.

[b] Outlier of –10 standard deviations

Table 2
Empirical Size of Runs Test (Conditionally Heteroskedastic and Conditionally Heterokurtic Alternatives) [a]

Sample Size	DGP	Nominal Size 0.01	0.05	0.10
(A) *Conditionally Heteroskedastic Alternatives*				
500	$\alpha_0 = 0.05$ $\alpha_1 = 0.15$ $\alpha_2 = 0.0$ $\alpha_3 = 0.8$	0.011	0.056	0.109
2000	$\lambda = 0.0$	0.014	0.060	0.115
500	$\alpha_0 = 0.05$ $\alpha_1 = 0.15$ $\alpha_2 = -0.08$ $\alpha_3 = 0.8$	0.014	0.059	0.106
2000	$\lambda = 0.0$	0.010	0.058	0.126
500	$\alpha_0 = 0.05$ $\alpha_1 = 0.15$ $\alpha_2 = 0.0$ $\alpha_3 = 0.8$	0.010	0.050	0.109
2000	$\lambda = -0.5$	0.010	0.049	0.110
(B) *Conditionally Heterokurtic Alternatives*				
500	$\lambda = 0.0$	0.015	0.051	0.115
2000	$\gamma_0 = -4.0$ $\gamma_1 = 0.4$	0.013	0.055	0.107
500	$\lambda = 0.0$	0.013	0.060	0.112
2000	$\gamma_0 = -4.0$ $\gamma_1 = 0.4$ $\gamma_2 = 0.5$	0.009	0.058	0.114

[a] Frequency of rejection of the null hypothesis that there is no predictability in conditional skewness, measured over 1000 replications. The DGP in panel (A) is

$$y_t = \varepsilon_t,\ \varepsilon_t = h_t^{1/2} z_t,\ z_t \sim Hansen - t(\lambda, 5),$$
$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}{}^2 + \alpha_2 \max(0, \varepsilon_{t-1})^2 + \alpha_3 h_{t-1}.$$

The DGP in panel (B) is

$$y_t = \varepsilon_t,\ \varepsilon_t = z_t,\ z_t \sim Hansen - t(0, \eta_t),$$
$$\eta_t = f(\eta_t'),\ \eta_t' = \gamma_0 + \gamma_1 z_{t-1} + \gamma_2 \eta_{t-1}'$$

where $f(.)$ is the logistic distribution restricting η_t to the interval (3, 50).

Table 3
Power of Runs Test [a]

		Nominal Size		
Sample Size	DGP	0.01	0.05	0.10
500	$\beta_0 = 0.6\ \beta_1 = 0.5$	0.091	0.210	0.316
2000		0.362	0.612	0.722
500	$\beta_0 = 0.0\ \beta_1 = 0.5$	0.142	0.353	0.457
2000		0.691	0.868	0.923
500	$\beta_0 = -0.6\ \beta_1 = 0.5$	0.083	0.215	0.313
2000		0.333	0.594	0.711
500	$\beta_0 = 0.6\ \beta_1 = -0.5$	0.114	0.252	0.366
2000		0.445	0.677	0.780
500	$\beta_0 = 0.0\ \beta_1 = -0.5$	0.147	0.343	0.459
2000		0.679	0.842	0.902
500	$\beta_0 = -0.6\ \beta_1 = -0.5$	0.097	0.240	0.350
2000		0.434	0.690	0.787

[a] Frequency of rejection of the null hypothesis that there is no predictability in conditional skewness, measured over 1000 replications. The DGP is

$$y_t = \varepsilon_t,\ \varepsilon_t = z_t,\ z_t \sim Hansen - t(\lambda_t, 5)\ ,$$
$$\lambda_t = f(\lambda_t'),\ \lambda_t' = \beta_0 + \beta_1 z_{t-1}$$

where $f(.)$ is the logistic distribution restricting λ_t to the interval (−1, 1).

Table 4a
Size and Power of Other Tests for Predictability in Conditional Skewness

		Wald		ρ_{31}	
Sample Size	DGP	0.05	0.10	0.05	0.10
(A) *Power*					
500	$\beta_0 = 0.0$ $\beta_1 = 0.5$	0.943	0.969	0.262	0.388
2000		1.000	1.000	0.662	0.773
500	$\beta_0 = -0.6$ $\beta_1 = -0.5$	0.919	0.953	0.192	0.296
2000		1.000	1.000	0.493	0.622
(B) *Size (Outlier Alternatives)*					
500	iid t(5)	0.074	0.139	0.042	0.089
2000		0.058	0.107	0.047	0.097
500	iid Hansen-t	0.102	0.162	0.049	0.079
2000	$\lambda = -0.5, \eta = 5$	0.068	0.123	0.044	0.085
(C) *Size (Conditionally Heteroskedastic Alternatives)*					
500	$\alpha_0 = 0.05$ $\alpha_1 = 0.15$	0.057	0.117	0.183	0.248
2000	$\alpha_2 = -0.08$ $\alpha_3 = 0.8$; $\lambda = 0.0$	0.040	0.112	0.289	0.351
500	$\alpha_0 = 0.05$ $\alpha_1 = 0.15$	0.128	0.198	0.277	0.351
2000	$\alpha_2 = 0.0$ $\alpha_3 = 0.8$; $\lambda = -0.5$	0.245	0.346	0.447	0.521
(D) *Size (Conditionally Heterokurtic Alternatives)*					
500	$\lambda = 0.0$	0.077	0.120	0.039	0.065
2000	$\gamma_0 = -4.0$ $\gamma_1 = 0.4$	0.052	0.107	0.039	0.073
500	$\lambda = -0.5$	0.150	0.229	0.046	0.066
2000	$\gamma_0 = -4.0$ $\gamma_1 = 0.4$	0.106	0.193	0.051	0.075

[a] Frequency of rejection of the null hypothesis that there is no predictability in conditional skewness, measured over 1000 replications. The DGPs are

Panel (A): $y_t = \varepsilon_t,\ \varepsilon_t = z_t,\ z_t \sim Hansen-t(\lambda_t, 5)$, $\lambda_t = f(\lambda_t')$, $\lambda_t' = \beta_0 + \beta_1 z_{t-1}$
where $f(.)$ is the logistic distribution restricting λ_t to lie in the interval (–1,1).

Panel (C): $y_t = \varepsilon_t,\ \varepsilon_t = h_t^{1/2} z_t,\ z_t \sim Hansen-t(\lambda, 5)$, $h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \max(0, \varepsilon_{t-1})^2 + \alpha_3 h_{t-1}$.

Panel (D): $y_t = \varepsilon_t,\ \varepsilon_t = z_t,\ z_t \sim Hansen-t(\lambda, \eta_t)$, $\eta_t = f(\eta_t')$, $\eta_t' = \gamma_0 + \gamma_1 z_{t-1} + \gamma_2 \eta_{t-1}'$
where $f(.)$ is the logistic distribution restricting η_t to lie in the interval (3, 50).

'Wald' refers to a test the null hypothesis $H_0 : \beta_1 = 0$ in the ARCD model $y_t = \varepsilon_t$, $\varepsilon_t = h_t^{1/2} z_t$, $z_t \sim Hansen-t(\lambda_t, 5)$, $\lambda_t = \beta_0 + \beta_1 z_{t-1}$. ρ_{31} refers to the significance of the regression of y_t^3 on y_{t-1}.

Table 4b
Size of Alternate Wald Test for Predictability in Conditional Skewness
(Conditionally Heteroskedastic Alternatives)

Sample Size	DGP	Wald (b) 0.05	Wald (b) 0.10
500	$\alpha_0 = 0.05$ $\alpha_1 = 0.15$	0.062	0.121
2000	$\alpha_2 = -0.08$ $\alpha_3 = 0.8$; $\lambda = 0.0$	0.058	0.114
500	$\alpha_0 = 0.05$ $\alpha_1 = 0.15$	0.097	0.158
2000	$\alpha_2 = 0.0$ $\alpha_3 = 0.8$; $\lambda = -0.5$	0.080	0.131

[a] Frequency of rejection of the null hypothesis that there is no predictability in conditional skewness, measured over 1000 replications. The DGP is

$$y_t = \varepsilon_t,\ \varepsilon_t = h_t^{1/2} z_t,\ z_t \sim Hansen - t(\lambda, 5),$$
$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \max(0, \varepsilon_{t-1})^2 + \alpha_3 h_{t-1}.$$

Wald (b) refers to the same test using the model

$$y_t = \varepsilon_t,\ \varepsilon_t = h_t^{1/2} z_t,\ z_t \sim Hansen - t(\lambda_t, \eta),$$
$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \max(0, \varepsilon_{t-1})^2 + \alpha_3 h_{t-1},$$
$$\lambda_t = \beta_0 + \beta_1 z_{t-1}$$

Table 5
Tests for Predictability in Conditional Skewness

Series	(i) Wald Test	(ii) Runs Test	(iii) ρ_{31}	(iv) skewness
FTSE100				
(a)	0.000	0.087	0.000	-0.7405
(b)	0.001	0.052	0.000	-0.6922
(c)	0.000	0.057	0.000	-0.7173
(d)	0.000	0.066	0.001	-0.5599
(e)	0.000	0.066	0.359	-0.4694
NK225				
(a)	0.000	0.903	0.069	0.0760
(b)	0.000	0.120	0.023	0.1782
(c)	0.001	0.293	0.174	0.0283
(d)	0.000	0.221	0.912	-0.1205
(e)	0.000	0.221	0.439	-0.1113
SP500				
(a)	0.000	0.143	0.000	-1.8752
(b)	0.000	0.057	0.000	-2.0261
(c)	0.000	0.169	0.000	-2.0389
(d)	0.000	0.249	0.000	-2.1203
(e)	0.000	0.003	0.000	-1.9663

[a] p-values of tests of the null hypothesis that there is no predictability in conditional skewness over the period Jan 1984 to Dec 2001. "Wald Test" refers to a Wald test of the null hypothesis that the coefficients of the asymmetry equation (T5.1) in the model

$$\begin{aligned} y_t &= \mu + \varepsilon_t,\ \ \varepsilon_t = h_t^{1/2} z_t,\ \ z_t \sim Hansen - t(\lambda_t, 5), \\ h_t &= \alpha_0 + \alpha_1 \varepsilon_{t-1}{}^2 + \alpha_2 \max(0, \varepsilon_{t-1})^2 + \alpha_3 h_{t-1} \\ \lambda_t &= g(z_{t-1,}\ z_{t-2}, \ldots) \end{aligned} \qquad \text{(T5.1)}$$

are zero. Several specifications for (T5.1) were used, including lags, squared lags, and lagged λ_t. The result from the best fitting model is reported. (a) refers to the returns series. In (b) to (e), the conditional mean is first modeled and subtracted from the returns series. The conditional mean is modeled as follows: (b) ARMA-GARCH-in-Mean, (c) ARMA(2,2), (d) regression of returns on two lags of returns and squared returns and an interaction term between returns at one and two lags, (e) non-parametrically (multivariate local polynomial smoothing) using the same predictors as in (d). ρ_{31} refers to the significance of the regression of y_t^3 on y_{t-1}. 'Skewness' is the sample skewness coefficient.

Table 6
Mincer-Zarnowitz Regressions for Tail Probabilities [a]

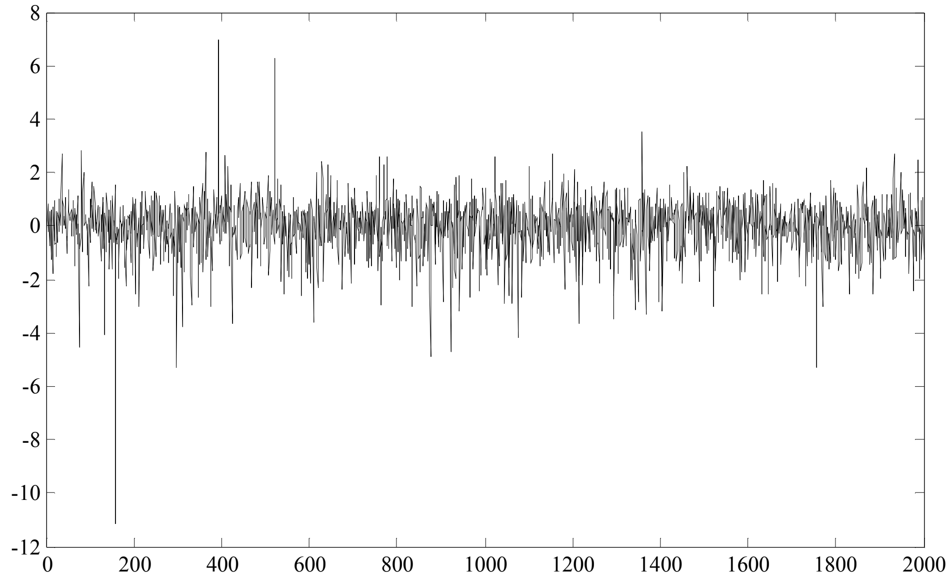
	GARCH	GARCH-ARCD
FTSE100		
(a)	0.098	0.106
(b)	0.298	0.396
(c)	0.277	0.334
(d)	0.099	0.110
(e)	0.096	0.109
NK225		
(a)	0.037	0.037
(b)	0.258	0.306
(c)	0.203	0.203
(d)	0.036	0.036
(e)	0.037	0.037
SP500		
(a)	0.030	0.050
(b)	0.133	0.246
(c)	0.133	0.197
(d)	0.024	0.049
(e)	0.023	0.059

[a] McFadden R^2 for estimated coefficients in the probit regression

$$\Pr(y_i < -2.5) = \Phi\left(\delta_0 + \delta_1 \hat{p}_{it}\right),\ i = 1,2,$$

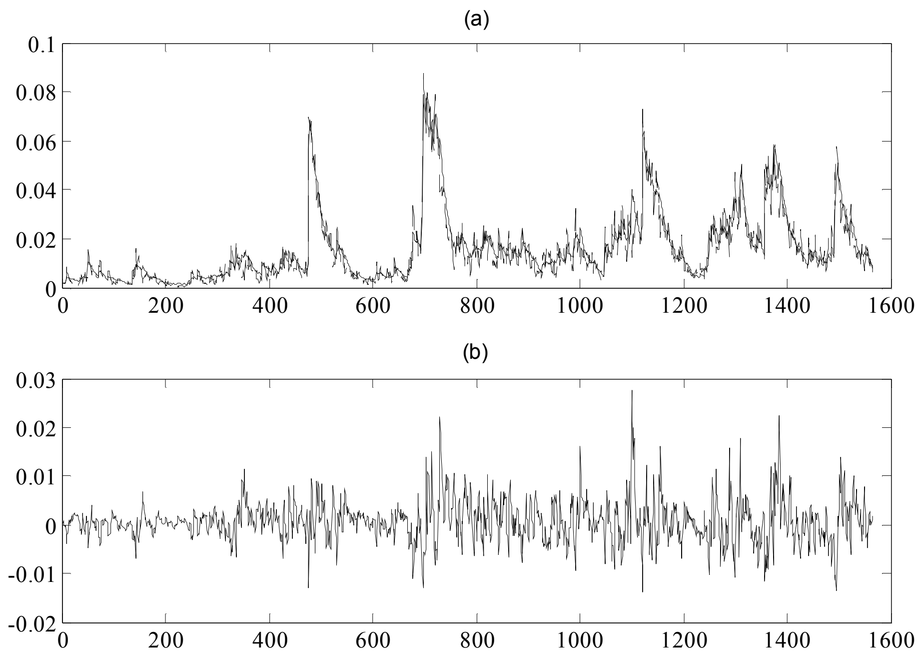
where $\hat{p}_{1t}$ and $\hat{p}_{2t}$ represent out-of-sample probability forecasts of the event $y_t < -2.5$ from an estimated GARCH model, and an estimated GARCH-ARCD model, respectively. (a) refers to the returns series. In (b) to (e), the conditional mean is first modeled and subtracted from the returns series. The conditional mean is modeled as follows: (b) ARMA-GARCH-in-Mean, (c) ARMA(2,2), (d) regression of returns on two lags of returns and squared returns and an interaction term between returns at one and two lags, (e) non-parametrically (multivariate local polynomial smoothing) using the same predictors as in (d).

Figure 1 Simulated Data Series with Predictability in Conditional Skewness[a]



[a] Data generated from a Hansen ARCD distribution as given in equations (2.1) to (2.4), with the asymmetry equation specified as $\lambda_t = -0.6 + 0.5 z_{t-1}$.

Figure 2
Tail Probability Forecasts



Panel (a) Forecasted probabilities $\Pr(y_t < -2.5)$ for SP500 returns. Solid line: GARCH model. Dashed line: GARCH-ARCD model. Panel (b): differences between the forecasted probabilities.