

A Test for Density Forecast Comparison with Applications to Risk Management*

Yong Bao[†]
UC Riverside

Tae-Hwy Lee[‡]
UC Riverside

Burak Saltoğlu[§]
Marmara University

June 2004

ABSTRACT

In this paper we propose a testing procedure for comparing the predictive abilities of possibly misspecified density forecast models. We use the minimum Kullback-Leibler Information Criterion (KLIC) divergence measure to define the distance between the candidate density forecast model and the true model. We use the fact that the inverse-normal transform of the probability integral transforms (PIT) should be IID standard normal as discussed in Berkowitz (2001). To compare the performance of density forecast models in the tails, we use the censored likelihood functions to compute the tail minimum KLIC. The reality check test of White (2000) is then constructed using our distance measure as a loss function. To highlight the merits of our approach, we use the daily S&P500 and NASDAQ return series to conduct an empirical density forecast comparison exercise. A large set of distributions, including some recently proposed flexible distributions to accommodate higher moments, and the ARCH-family volatility specifications are studied. Our empirical findings lend further support of fat-tailedness and skewness of return distributions. In addition, the choice of conditional distribution specification appears to be a much more dominant factor in determining the quality of density forecasts than the choice of volatility specification.

Key Words: density forecast comparison, KLIC, predictive ability, reality check.

JEL Classification: C3, C5, G0.

*We would like to thank Jin-Seo Cho, Alexei V. Egorov, Lutz Kilian, Norm Swanson, Hal White, Ximing Wu, as well as seminar participants at the 2003 Midwest Econometrics Group Meeting and the 2004 Winter Econometric Society Meeting, for helpful discussions and comments. We also thank Canlin Li for providing the CRSP data. All remaining errors are our own.

[†]Department of Economics, University of California, Riverside, CA 92521, U.S.A. Tel: +1 (909)7887998. Fax: +1 (909)7875685. Email: yongbao@mail.ucr.edu.

[‡]Corresponding author. Department of Economics, University of California, Riverside, CA 92521, U.S.A. Tel: +1 (909) 8271509. Fax: +1 (909)7875685. Email: tae.lee@ucr.edu.

[§]Department of Economics, Marmara University, Istanbul, 81040, Turkey. Tel: +90 (216)3368487. Fax: +90 (216)3464356. Email: saltoglu@marmara.edu.tr.

1 Introduction

Forecasting density has been at the core of finance and economics research agenda. For instance, most of the classical finance theories such as asset pricing, portfolio selection, and option valuation aim to model the surrounding uncertainty via a parametric distribution function. Extracting information about market participants' expectations from option prices can be considered another form of density forecasting exercise (Soderling and Svensson, 1997; Jackwerth and Rubinstein, 1996). Besides, there have been increasing interest in density forecasts of inflation, unemployment and output (Clements and Smith, 2000; Clements, 2002). Some popular risk measurement tools such as the Value-at-Risk (VaR) and expected shortfalls (ES), which aim to model the tails of portfolio return distributions, can also be evaluated from the perspective of density forecasting.

Consequently, in time series econometrics, there has been extensive literature on evaluating density forecast models: Diebold *et al.* (1998), Diebold *et al.* (1999), Clements and Smith (2000), Berkowitz (2001), Hong (2002), among others. In particular, Diebold *et al.* (1998), Granger (1999a, 1999b), Granger and Pesaran (2000a, 2000b), Pesaran and Skouras (2001), and Patton and Timmermann (2003) discussed the issue of forecast optimality.¹

While the research on evaluating each density forecast model has been very versatile since the seminal paper of Diebold *et al.* (1998), there has been much less effort in comparing alternative density forecast models. Considering the recent empirical evidence on volatility clustering and asymmetry and fat-tailedness in financial return series, we believe that a formal test of relative optimality of a given model compared with alternative distribution and volatility specifications in the context of density forecasts will contribute to the existing literature.² Deciding on which distribution and/or volatility specification to use for a particular asset is a common task even for finance practitioners and risk professionals. For instance, in spite of the massive literature on volatility forecasting, a clear consensus on which model to use has not yet been reached. As argued in Poon and Granger (2003), most of the (volatility) forecasting studies do not produce very conclusive results because only a subset of alternative models are compared, with a potential bias towards the method developed by the authors.³ It is further claimed that lack of a uniform forecast evaluation technique makes volatility forecasting a difficult task. Being able to choose the most suitable volatility *and* distribution specifications is a more demanding task. However, in this paper we demonstrate that this gap can be filled

¹Tay and Wallis (2000) surveyed the recent studies on density forecasts; Bauwens *et al.* (2003) applied density forecast to the ACD models and Freichs and Löffler (2003) to credit portfolio risk evaluation within the same framework.

²For more details, see Engle (1982), Harvey and Siddique (1999), Mittnik *et al.* (1999), Lambert and Laurent (2001), Jondeau and Rockinger (2001, 2003), and Mittnik and Paoletta (2003).

³They further stated (p. 507), "However, it seems clear that one form of study that is included is conducted just to support a viewpoint that a particular method is useful. It might not have been submitted for publication if the required result had not been reached. This is one of the obvious weaknesses of a comparison such as this; the papers being prepared for different reasons, use different data sets, many kinds of assets, various intervals between readings, and a variety of evaluation techniques."

by a rigorous density forecast comparison methodology.

The main aim of this paper is to propose a test for comparing various density forecast models, all of which can be possibly misspecified. The proposed test for density forecast comparison enables us to assess which volatility and/or distribution are statistically more appropriate to mimic the time series behavior of a return series. Since every particular density forecast model may be misspecified, comparison of a set of candidate models should be based on the “distances” of these models to the true, unknown model. Our starting point is to utilize Kullback and Leibler’s (1951) information criterion (KLIC) and then relate it to the likelihood ratio (LR) statistics of Berkowitz (2001) for density forecast evaluation. In particular, we define the distance of a postulated density forecast model to the true model as the minimum KLIC divergence measure between the two distributions, as suggested by Vuong (1989). This distance can be estimated by the sample likelihood ratio, which is equivalent to the sample likelihood ratio of the density of the inverse normal transform of the probability integral transforms (PIT) of the process with respect to the models’s density forecast to a standard normal density. This equivalence enables us to use the LR statistic developed in Berkowitz (2001) to estimate conveniently the minimum KLIC. One immediate extension of the proposed test is to compare the predictive abilities of alternative density forecast models in the tails within the context of VaR and ES.⁴ For this purpose, a tail minimum KLIC discrepancy measure based on the censored likelihoods is used. Another innovation of this paper is to use our distance measure as a forecast loss function in the framework of White’s (2000) reality check.

As an application, in the empirical section we demonstrate how the proposed testing methodology can be used to assess density forecasts in the context of risk management. We show that, by using the proposed test it is possible to differentiate the relative importance of volatility and distribution specifications for a return series. To this end, we conduct a density forecast comparison of 117 ($= 13 \times 9$) models constructed from thirteen different distributions and nine different volatility models. Our empirical findings based on the daily S&P500 and NASDAQ return series confirm the recent evidence on financial return asymmetry and persistence of volatility. The Hansen- t distribution developed by Hansen (1994) appears to be the one of the most successful distribution for the tails. In addition, the empirical results strongly indicate that the choice of conditional distribution specification is a much more dominant factor in determining the quality of density forecasts than the choice of volatility specification. The distribution ranking given a volatility model is much more consistent and show less variability than the volatility ranking given a distribution model. This finding, may constitute an explanation why the volatility forecasting studies in the massive literature are so inconclusive in ranking various volatility models. Meanwhile, a model that provides superior density

⁴The proposed test can also be useful for comparing density forecasts of portfolio returns and evaluating alternative credit portfolio risk models.

forecasts for the whole return distribution may not necessarily meet the needs of risk managers who care much more about the tails. These findings may explain why many different VaR models based on different distribution and volatility specifications do produce different predictive performance, as examined in Bao *et al.* (2003). The results obtained in the paper may shed some light on how both the distribution and volatility specifications can be treated for a better forecasting practice.

This paper is organized as follows. In Section 2, we develop our distance measure based on the minimum KLIC divergence measure between a candidate model and the true model. The reality check test using this distance measure as a loss function to compare a set of competing models with a benchmark model is also discussed. In Section 3 we present the design of the empirical experiment we shall conduct in this paper. In Section 4, empirical results obtained by using the daily S&P 500 and NASDAQ return series are presented. Finally, Section 5 concludes. More detailed description of the distribution and volatility models described in Section 3 and used in Section 4 is given in the appendix.

2 Comparing Density Forecast Models

Consider a financial return series $\{y_t\}_{t=1}^T$, generated by the probability law $\Pr(y_t \leq y | \mathcal{F}_{t-1})$ conditional on the information set \mathcal{F}_{t-1} (σ -field) at time $t-1$, possibly including some exogenous variables. Suppose that y_t has a true, unknown conditional density function $\varphi_t(y_t) \equiv \varphi_t(y_t | \mathcal{F}_{t-1})$. Let $\psi_t(y_t; \boldsymbol{\theta}) \equiv \psi_t(y_t | \mathcal{F}_{t-1}; \boldsymbol{\theta})$ be a one-step-ahead conditional density forecast model, where $\boldsymbol{\theta} \in \Theta$ is a finite-dimensional vector of parameters in a compact parameter space Θ . If $\psi_t(\cdot; \boldsymbol{\theta}_0) = \varphi_t(\cdot)$ for some $\boldsymbol{\theta}_0 \in \Theta$, then the one-step-ahead density forecast is correctly specified and hence optimal, as it dominates all other density forecasts for any loss function (e.g., Diebold *et al.*, 1998; Granger and Pesaran, 2000a, 2000b).

In practice, it is rarely the case that we can find an optimal model. As it is very likely that “the true distribution is in fact too complicated to be represented by a simple mathematical function” (Sawa, 1978), all the models proposed by different researchers can be possibly misspecified and we regard each model as an approximation to the truth. Our task is then to investigate which model can approximate the true model most closely. We have to first define a metric to measure the distance of a given model to the truth and then compare different models in terms to this distance.

2.1 The Distance Measure

The adequacy of a postulated distribution may be appropriately measured by the Kullback-Leibler (1951) Information Criterion (KLIC) divergence measure between two conditional densities, $I(\varphi : \psi, \boldsymbol{\theta}) = \mathbb{E}[\ln \varphi_t(y_t) - \ln \psi_t(y_t; \boldsymbol{\theta})]$, where the expectation is with respect to the true distribution. Following Vuong (1989), we define

the distance between a model and the true density as the minimum of the KLIC

$$I(\varphi : \psi, \boldsymbol{\theta}^*) = \mathbb{E}[\ln \varphi_t(y_t) - \ln \psi_t(y_t; \boldsymbol{\theta}^*)], \quad (1)$$

and $\boldsymbol{\theta}^*$ is the pseudo-true value of $\boldsymbol{\theta}$, the parameter value that gives the minimum $I(\varphi : \psi, \boldsymbol{\theta})$ for all $\boldsymbol{\theta} \in \Theta$ (e.g., Sawa, 1978; White, 1982).⁵ The smaller this distance, the closer the model $\psi(\cdot)$ to the true density.⁶ Hence we use this distance measure to compare a battery of competing models.

However, $I(\varphi : \psi, \boldsymbol{\theta}^*)$ is generally unknown, since we can not observe $\varphi(\cdot)$ and hence the expectation. Under some regularity conditions, it can nevertheless be shown that $\mathbb{E}[\ln \varphi_t(y_t) - \ln \psi_t(y_t; \boldsymbol{\theta}^*)]$ can be consistently estimated by

$$\hat{I}(\varphi : \psi, \boldsymbol{\theta}^*) = \frac{1}{T} \sum_{t=1}^T [\ln \varphi_t(y_t) - \ln \psi_t(y_t; \boldsymbol{\theta}^*)], \quad (2)$$

where $\boldsymbol{\theta}^*$ can be consistently estimated by $\hat{\boldsymbol{\theta}}_T$ that maximizes $\frac{1}{T} \sum_{t=1}^T \ln \psi_t(y_t; \boldsymbol{\theta})$

But we still do not know $\varphi_t(\cdot)$. For this, we utilize a inverse normal transform of the probability integral transform (PIT) of the actual realizations of the process with respect to the models's density forecast. The equivalence between $\ln[\varphi_t(y_t)/\psi_t(y_t; \hat{\boldsymbol{\theta}}_T)]$ and the log likelihood ratio of the transformed PITs enables us to consistently estimate $I(\varphi : \psi, \boldsymbol{\theta}^*)$ and hence to compare possibly misspecified models in terms of their distance to the true model.

The PIT of the realization of the process with respect to the models's density forecast is defined as

$$u_t = \int_{-\infty}^{y_t} \hat{\psi}_t(y) dy, \quad (3)$$

where $\hat{\psi}_t(y) = \psi_t(y; \hat{\boldsymbol{\theta}}_T)$. It is well known that if $\hat{\psi}_t(\cdot)$ coincides with the true density $\varphi_t(\cdot)$, then the sequence $\{u_t\}_{t=1}^T$ is IID and uniform on the interval $[0, 1]$ ($U[0, 1]$ henceforth). This provides a powerful approach to evaluating the quality of a density forecast model. If the transformed series $\{u_t\}$ is not IID $U[0, 1]$, then $\psi_t(\cdot; \boldsymbol{\theta})$ is not an optimal density forecast model.⁷ Therefore, a natural test of optimality of a density forecast model is to test the IID $U[0, 1]$ properties of $\{u_t\}$, see, e.g., Diebold *et al.* (1998), Diebold *et al.* (1999b), Clements and Smith (2000), Hong and Li (2002), Noceti *et al.* (2003), among others. Our task, however, is not to evaluate a single model, but to compare a battery of competing models. Our purpose of utilizing the PITs is to exploit the following equivalence between $\ln[\varphi_t(y_t)/\hat{\psi}_t(y_t)]$ and the log likelihood ratio of the transformed PITs and hence to construct the distance measure.⁸ The inverse normal transform of the PIT

⁵We use the word "distance" loosely because the KLIC *divergance* measure does not possess some basic properties of a distance measure. By "distance", we mean the KLIC divergance measure evaluated at the pseudo-true parameter value.

⁶This motivates Vuong (1989) to design a model selection test for *two* competing models, say, $\psi_1(\cdot)$ and $\psi_2(\cdot)$, by comparing $\mathbb{E}[\psi_1(y; \boldsymbol{\theta}_1^*)]$ and $\mathbb{E}[\psi_2(y; \boldsymbol{\theta}_2^*)]$, while Giacomini (2002) considered a weighted out-of-sample test.

⁷Diebold *et al.* (1999) showed that when the innovation distribution belongs to a location-scale family and the conditional location and scale are adequately specified, $\{u_t\}$ will be IID, but not $U[0, 1]$ if the innovation distribution is misspecified. If the innovation distribution does not belong to the location-scale family, both serial dependence and non-uniformity of $\{u_t\}$ may be caused by either poorly modelled location/scale or misspecified innovation distribution

⁸In fact, any other continuous transformation (the inverse of a continuous CDF) will give us an equivalence relationship. We use the normal inverse transform in particular for the sake of convenience.

is

$$x_t = \Phi^{-1}(u_t), \quad (4)$$

where $\Phi(\cdot)$ is the CDF of the standard normal. If the sequence $\{u_t\}_{t=1}^T$ is IID $U[0, 1]$, then $\{x_t\}_{t=1}^T$ is IID standard normal $N(0, 1)$ (IID $N(0, 1)$ henceforth).⁹ More importantly, Berkowitz (2001) (Proposition 2, p. 467) showed that

$$\ln \left[\varphi_t(y_t) / \hat{\psi}_t(y_t) \right] = \ln [p_t(x_t) / \phi(x_t)], \quad (5)$$

where $p_t(\cdot)$ is the density of x_t and $\phi(\cdot)$ is the standard normal density. Therefore, the distance of a density forecast model to the unknown true model can be equivalently estimated by the departure of $\{x_t\}_{t=1}^T$ from IID $N(0, 1)$,

$$\tilde{I}(\varphi : \psi, \hat{\boldsymbol{\theta}}_T) = \frac{1}{T} \sum_{t=1}^T [\ln p_t(x_t) - \ln \phi(x_t)]. \quad (6)$$

At a first glance, it may look like a loop: we do not know $\varphi_t(\cdot)$ and we make use of the transformed PITs, but we do not know $p_t(\cdot)$ either. Ideally, the unknown density $p_t(\cdot)$ should be able to accommodate heterogeneity, dependency, and nonnormality, possibly existing in the transformed PITs due to some misspecification of the density forecast model. We transform the departure of $\psi_t(\cdot; \boldsymbol{\theta})$ from $\varphi_t(\cdot)$ to the departure of $p_t(\cdot)$ from IID $N(0, 1)$. From a methodological point of view, however, we regard this transformation as nontrivial. The true distribution about $\{y_t\}_{t=1}^T$ is in fact maybe “too complicated” and we do not even know what it should look like, so departure of a model from this unknown object will also be too complicated. The truth about the transformed PITs $\{x_t\}_{t=1}^T$ is nevertheless quite simpler: it should behave as IID $N(0, 1)$ if the density forecast model hits the true density. In our point of view, measuring departure of the unknown $p_t(\cdot)$ from IID $N(0, 1)$ is more straightforward than measuring departure of the postulated $\psi_t(\cdot; \boldsymbol{\theta})$ from something unknown in the sense that we can at least specify a flexible $p_t(\cdot)$ to include IID $N(0, 1)$ as a special case, but we when we specify $\psi_t(\cdot; \boldsymbol{\theta})$ there is no guarantee that the postulated $\psi_t(\cdot; \boldsymbol{\theta})$ will accommodate the complicated $\varphi_t(\cdot)$, which is unknown at all *a priori*.

Saying so, we do not intend to claim that specifying departure from IID $N(0, 1)$ is an easy task. We want $p_t(\cdot)$ to be as flexible as possible to reflect the true distribution of $\{x_t\}_{t=1}^T$ and at the same time it can be IID $N(0, 1)$ if the density model coincides with the true model. We follow Berkowitz (2001) by specifying $\{x_t\}_{t=1}^T$ as an AR(L) process

$$x_t = \boldsymbol{\rho}' X_{t-1} + \sigma \eta_t, \quad (7)$$

where $X_{t-1} = (1, x_{t-1}, \dots, x_{t-L})'$, $\boldsymbol{\rho}$ is an $(L+1) \times 1$ vector of parameters, and η_t is IID distributed. In Berkowitz (2001), η_t is further assumed to be normally distributed. Apparently, if we specify $p_t(\cdot)$ as such

⁹This provides an equivalent approach to evaluating the quality of a density forecast model, see, e.g., Berkowitz (2001). Of course, testing IID $N(0, 1)$ can also be regarded as a specification test, see, e.g., Jarque and Bera (1980), Lee (1984), Hall (1990), Bai (2003), Bontemps and Meddahi (2003), Duan (2003), and van der Klaauw and Koning (2003).

(IID and normal), then our comparison based on the distance measure (6) will suffer the same criticism of the LR test of Berkowitz, as pointed out by Clements and Smith (2000) and Noceti *et al.* (2003): a maintained assumption on $\{x_t\}_{t=1}^T$ is normality and in some circumstances “a test of the normality assumption itself may prove more powerful” (Noceti *et al.*, 2003, p. 448). An obvious remedy is to specify a flexible alternative distribution for η_t , say, $p_t(\eta_t; \boldsymbol{\vartheta}_\eta)$, where $\boldsymbol{\vartheta}_\eta$ is a vector of distribution parameters such that when $\boldsymbol{\vartheta}_\eta = \boldsymbol{\vartheta}_\eta^*$, $p_t(\cdot; \boldsymbol{\vartheta}_\eta^*)$ is IID $N(0, 1)$. A test for IID $N(0, 1)$ of $\{x_t\}_{t=1}^T$ *per se* can be constructed by testing elements of the parameter vector $\boldsymbol{\vartheta} = (\boldsymbol{\rho}', \sigma, \boldsymbol{\vartheta}_\eta')'$, say, $\boldsymbol{\rho} = \mathbf{0}$, $\sigma = 1$, and $\boldsymbol{\vartheta}_\eta = \boldsymbol{\vartheta}_\eta^*$. Jarque and Bera (1980) specified $p_t(\cdot; \boldsymbol{\vartheta}_\eta)$ to be the Pearson distribution; Hall (1990) assumed the seminonparametric (SNP) density of Gallant and Nychka (1987) for η_t ; Kiefer and Salmon (1983) and Smith (1989) assumed η_t to admit the Gram-Charlier/Edgeworth-Sargan density. All these works constructed the tests using the LM principle. In fact, as the LM test requires estimation under the null only, the test statistics derived as such can be simplified substantially. However, the LM statistics, taking the form of a chi-square, is not a sample average, which makes the reality check (see next subsection) unoperational. The LR statistics, which is also the estimated distance measure of a density forecast model to the true model, on the other hand, takes the form of a sample average, which makes the reality check to compare competing models readily operational. To construct the LR statistics, we require the estimation of $\boldsymbol{\vartheta}$, in particular, $\boldsymbol{\vartheta}_\eta$. For the three flexible densities mentioned above (Pearson, SNP, and Edgeworth/Gram-Charlier), there is documented evidence in estimating the Pearson density (in particular, the type IV density, see Johnson *et al.*, 1994), and the SNP density is much easier to estimate compared with the Gram-Charlier/Edgeworth-Sargan density, when the nonnegativeness constraint on the density function is imposed (also see Section 3). For this, we assume that $\{x_t\}_{t=1}^T$ follows the AR process (7) with η_t IID distributed with the SNP density function of order K

$$p(\eta_t; \boldsymbol{\vartheta}_\eta) = \frac{\left(\sum_{k=0}^K r_k \eta_t^k\right)^2 \phi(\eta_t)}{\int_{-\infty}^{+\infty} \left(\sum_{k=0}^K r_k u^k\right)^2 \phi(u) du}, \quad (8)$$

where $r_0 = 1$. Now $\boldsymbol{\vartheta}_\eta = (r_1, \dots, r_K)$. Setting $r_k = 0$, $k = 1, \dots, K$, $p(\eta_t) = \phi(\eta_t)$.

Given (7) and (8), the density of x_t is

$$p(x_t; \boldsymbol{\vartheta}) = \frac{p[(x_t - \boldsymbol{\rho}' X_{t-1})/\sigma; \boldsymbol{\vartheta}_\eta]}{\sigma},$$

which degenerates into IID $N(0, 1)$ by setting $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}^* = (\mathbf{0}', 1, \mathbf{0}')'$. Then the estimated minimum KLIC divergence measure is

$$\tilde{I}(\varphi : \psi; \hat{\boldsymbol{\theta}}_T) = \frac{1}{T} \sum_{t=1}^T \left[\ln \frac{p[(x_t - \boldsymbol{\rho}' X_{t-1})/\sigma; \boldsymbol{\vartheta}_\eta]}{\sigma} - \ln \phi(x_t) \right].$$

The LR test statistics of the adequacy of the density forecast model $\psi(\cdot; \boldsymbol{\theta})$ in Berkowitz (2001) is simply the above formula with $p(\cdot) = \phi(\cdot)$. The parameter vector $\boldsymbol{\vartheta}$ can be estimated by maximizing the sample

likelihood function $\frac{1}{T} \sum_{t=1}^T \ln p(x_t; \boldsymbol{\vartheta})$. Then $\tilde{I}(\varphi : \psi, \hat{\boldsymbol{\theta}}_T)$ can be estimated by

$$\begin{aligned} \tilde{I}(\varphi : \psi, \hat{\boldsymbol{\theta}}_T) &= \frac{1}{T} \sum_{t=1}^T \left[\ln p(x_t; \hat{\boldsymbol{\vartheta}}_T) - \ln \phi(x_t) \right] \\ &= \frac{1}{T} \sum_{t=1}^T \left[\ln \frac{p \left[(x_t - \hat{\boldsymbol{\rho}}_T' X_{t-1}) / \hat{\sigma}_T; \hat{\boldsymbol{\vartheta}}_{\eta T} \right]}{\hat{\sigma}_T} - \ln \phi(x_t) \right], \end{aligned} \quad (9)$$

where $\hat{\boldsymbol{\vartheta}}_T = (\hat{\boldsymbol{\rho}}_T', \hat{\sigma}_T, \hat{\boldsymbol{\vartheta}}_{\eta T}')'$ is the MLE.

We note that closely related works to compare density forecast models are Corradi and Swanson (2003a, 2003b). Their approach is to compare the cumulative distribution function of a density forecast model to the empirical distribution (EDF) of the series in question. The distance of a density forecast model to the true model is measured by the mean square error of the CDF and the EDF, integrated out over the domain of the series. Our approach and theirs share the same flavor in the sense that we quantify the distance of a density forecast model with reference to the true distribution. They estimate the true cumulative distribution by the empirical distribution, but in our case the (transformed) true density is always IID $N(0, 1)$. The equivalence relationship (5) enables us to map the misspecification of a model to the deviation of $\{x_t\}_{t=1}^T$ from IID $N(0, 1)$. The relationship (5) also tells us that the LR statistic based on the transformed PITs is actually an estimate of the KLIC divergence measure between the model and the truth. In addition, we shall show soon that our distance measure can also be used to investigate the accuracy of the density forecast model over a specific region. Therefore, we answer both questions posed in Corradi and Swanson (2003a) (the applicability of the PIT approach to the context of density forecast comparison) and Corradi and Swanson (2003b) (the applicability of the KLIC discrepancy measure over some specific regions). Nevertheless, we agree with them on that neither distance measure is more “natural” than the other. From a Bayesian point of view, a model selection procedure based on the KLIC based distance measure, however, will give a model that with the highest posterior probability (Fernandez-Villaverde and Rubio-Ramirez, 2002). In practice, our approach should be simpler since it is based on a simple LR statistics, while their statistics are the Kolmogorov-Smirnov type statistics. An empirical comparison of our approach with theirs may be the subject of a further study.

As stated before, we may also want to check the performance of a density forecast model in certain regions of the distribution. It is quite possible that a suboptimal forecast model does better than another in predicting a certain region of the distribution (e.g., the left 5% tail), but worse in predicting another aspect of the distribution (e.g., the left 1% tail). Risk managers care more about the extreme values in the lower tail (large losses) than about the values in other regions of the distribution (small losses or gains). Therefore, a density forecast model that accurately predicts tail events, though not necessarily all the events, is of more interest to risk managers. To compare the performance of various competing density forecast models in the

tails, we can easily modify our distance measure tailored for the tail parts only. For the ease of exposition, we focus on the lower tails only. We define the censored random variable

$$x_t^\tau = \begin{cases} \Phi^{-1}(\alpha) \equiv \tau & \text{if } x_t \geq \tau \\ x_t & \text{if } x_t < \tau. \end{cases} \quad (10)$$

For example, $\tau = -1.645$ for $\alpha = 0.05$, the 5% tail. As before, we consider an AR model (7) with η_t distributed as in (8). Let $\mathbf{1}(\cdot)$ denote an indicator function that takes 1 if its argument is true and 0 otherwise. Then the censored random variable x_t^τ has the distribution function

$$p^\tau(x_t^\tau; \boldsymbol{\vartheta}) = \left[1 - P\left(\frac{\tau - \boldsymbol{\rho}'X_{t-1}}{\sigma}; \boldsymbol{\vartheta}_\eta\right) \right]^{\mathbf{1}(x_t \geq \tau)} \left[\frac{p[(x_t - \boldsymbol{\rho}'X_{t-1})/\sigma]}{\sigma} \right]^{\mathbf{1}(x_t < \tau)}, \quad (11)$$

in which $P(\cdot; \boldsymbol{\vartheta}_\eta)$ is the CDF of the SNP density function (8). The CDF function P of the SNP density can be easily calculated by utilizing the following: for integer $m \geq 1$,

$$\int_{-\infty}^v u^m \phi(u) du = \begin{cases} (-1)^m 2^{(m-1)/2} \Gamma\left(\frac{m+1}{2}, \frac{v^2}{2}\right) / \sqrt{2\pi} & \text{if } v \leq 0 \\ 2^{(m-1)/2} \left[(1 + (-1)^m) \Gamma\left(\frac{m+1}{2}\right) - \Gamma\left(\frac{m+1}{2}, \frac{v^2}{2}\right) \right] / \sqrt{2\pi} & \text{if } v > 0, \end{cases}$$

where $\Gamma(\cdot)$ is the complete gamma function and $\Gamma(\cdot, \cdot)$ is the upper incomplete gamma function (see Appendix A). Given $p^\tau(x_t^\tau; \boldsymbol{\vartheta})$, the tail minimum KLIC divergence measure can be estimated analogously

$$\hat{I}^\tau(\varphi : \psi, \hat{\boldsymbol{\theta}}_T) = \frac{1}{T} \sum_{t=1}^T [\ln p^\tau(x_t^\tau; \boldsymbol{\vartheta}) - \ln \phi^\tau(x_t^\tau)],$$

where $\phi^\tau(x_t^\tau) = [1 - \Phi(\tau)]^{\mathbf{1}(x_t \geq \tau)} [\phi(x_t)]^{\mathbf{1}(x_t < \tau)}$. Similarly, $\hat{I}^\tau(\varphi : \psi, \hat{\boldsymbol{\theta}}_T)$ can be estimated by

$$\bar{I}^\tau(\varphi : \psi, \hat{\boldsymbol{\theta}}_T) = \frac{1}{T} \sum_{t=1}^T [\ln p^\tau(x_t^\tau; \hat{\boldsymbol{\vartheta}}_T) - \ln \phi^\tau(x_t^\tau)], \quad (13)$$

where $\hat{\boldsymbol{\vartheta}}_T$ maximizes $\frac{1}{T} \sum_{t=1}^T \ln p^\tau(x_t^\tau; \boldsymbol{\vartheta})$.

2.2 Out-of-Sample Density Forecast Comparison

Suppose there are $l + 1$ models ($k = 0, 1, \dots, l$) in a set of the competing models, possibly misspecified. To establish the notation with model index k , let the density forecast model k ($k = 0, 1, \dots, l$) be denoted by $\psi_{k,t}(y; \boldsymbol{\theta}_k)$. There are different schemes (recursive, fixed, and rolling) to estimate $\boldsymbol{\theta}_k$. As our purpose is to compare the out-of-sample predictive abilities among competing density forecast models, we consider two subsamples $\{y_t\}_{t=1}^R$ and $\{y_t\}_{t=R+1}^T$: we use the first sample to estimate the unknown parameter vector $\boldsymbol{\theta}_k$ and the second subsample to check if the transformed PITs are IID $N(0, 1)$. That is, we first construct

$$u_{k,t} = \int_{-\infty}^{y_t} \psi_{k,t}(y; \hat{\boldsymbol{\theta}}_{k,t-1}) dy, \quad t = R + 1, \dots, T, \quad (14)$$

where $\hat{\boldsymbol{\theta}}_{k,t-1}$ can be based on the whole subsample $\{y_{t-1}, \dots, y_1\}$, or a fixed sample $\{y_R, \dots, y_1\}$, or a rolling sample $\{y_{t-1}, \dots, y_{t-R}\}$ of size R .¹⁰ Suppose we use the rolling sample scheme, then we check the IID $N(0, 1)$

¹⁰See West and McCracken (1998, p. 819) for more discussion on these three forecasting schemes. While using the whole sample has the advantage of using more observations, we will use the rolling sample in Section 4 because it may be more robust to a possible parameter variation in the presence of potential structural breaks.

properties of the transformed out-of-sample PITs $\{x_{k,t}\}_{t=R+1}^T$ by defining an estimate of the *out-of-sample* minimum KLIC discrepancy as

$$\bar{I}_k(\varphi : \psi_k, \hat{\boldsymbol{\theta}}_{k,R}) = \frac{1}{n} \sum_{t=R+1}^T \left[\ln p(x_{k,t}; \hat{\boldsymbol{\theta}}_{k,n}) - \ln \phi(x_{k,t}) \right], \quad (15)$$

where the R in $\hat{\boldsymbol{\theta}}_{k,R}$ is to denote the sample size used to estimate $\boldsymbol{\theta}_k$ and since we are using a rolling sample scheme, it is understood that $\hat{\boldsymbol{\theta}}_{k,R}$ is time-varying.

Model comparison between a single model (model k) and the benchmark model (model 0) can be conveniently formulated as hypothesis testing of some suitable moment conditions. Consider

$$d_{k,t} = d_k(\psi_0, \psi_k) = [\ln \varphi_t(y_t) - \ln \psi_0(y_t; \boldsymbol{\theta}_0^*)] - [\ln \varphi_t(y_t) - \ln \psi_k(y_t; \boldsymbol{\theta}_k^*)] \quad (16)$$

where $1 \leq k \leq l$. Note that $\mathbb{E}(d_{k,t}) = I_0(\varphi : \psi, \boldsymbol{\theta}_0^*) - I_k(\varphi : \psi, \boldsymbol{\theta}_k^*)$ is the difference in the minimum KLIC of model 0 and model k . When we compare model k with a benchmark (model 0), the null hypothesis is that model k is no better than the benchmark:

$$\mathbb{H}_1 : \mathbb{E}(d_{k,t}) \leq 0 \text{ for each } k = 1, \dots, l. \quad (17)$$

We can use the out-of-sample predictive ability tests of Diebold and Mariano (1995) and West (1996) by referring to the standard asymptotic result for $\sqrt{n}\bar{d}_{k,n} = n^{-1/2} \sum_{t=R+1}^T \hat{d}_{k,t} = \sqrt{n}[\bar{I}_0(\varphi : \psi_0, \hat{\boldsymbol{\theta}}_{0,R}) - \bar{I}_k(\varphi : \psi_k, \hat{\boldsymbol{\theta}}_{k,R})]$ to test \mathbb{H}_1 , where

$$\hat{d}_{k,t} = \left[\ln p(x_{0,t}; \hat{\boldsymbol{\theta}}_{0,n}) - \ln \phi(x_{0,t}) \right] - \left[\ln p(x_{k,t}; \hat{\boldsymbol{\theta}}_{k,n}) - \ln \phi(x_{k,t}) \right].$$

When we compare multiple l models against a benchmark jointly, the null hypothesis of interest is that the best model is no better than the benchmark:

$$\mathbb{H}_2 : \max_{1 \leq k \leq l} \mathbb{E}(d_k) \leq 0. \quad (18)$$

To formulate the test statistic, consider an $l \times 1$ vector of moments, $\mathbb{E}(\mathbf{d})$, where \mathbf{d} is an $l \times 1$ vector of loss differential, with elements d_k , $1 \leq k \leq l$. These moments have incorporated the information of the l models and the benchmark model. Define the out-of-sample $l \times 1$ moment vector $\bar{\mathbf{d}}_n \equiv n^{-1} \sum_{t=R+1}^T \hat{\mathbf{d}}_t$, where $\hat{\mathbf{d}}_t$ has the k th element $\hat{d}_{k,t}$. By a suitable central limit theorem, we have $\sqrt{n}[\bar{\mathbf{d}}_n - \mathbb{E}(\mathbf{d}_t)] \rightarrow N(0, \Omega)$ in distribution as $n \equiv n(T) \rightarrow \infty$ when $T \rightarrow \infty$. In general, the covariance matrix Ω is rather complicated because it depends on a component due to parameter estimation uncertainty (West, 1996). However, when $\mathbb{E}(\partial \mathbf{d} / \partial \theta) = 0$ or $n/R \rightarrow 0$ as $T \rightarrow \infty$, West (1996, Theorem 4.1) showed that, under proper regularity conditions, Ω does not depend on parameter estimation uncertainty as is the case of Diebold and Mariano (1995). White (2000) proposed the following test statistic for \mathbb{H}_2 :

$$\bar{V}_n \equiv \max_{1 \leq k \leq l} \sqrt{n}[\bar{d}_{k,n} - \mathbb{E}(d_k)]. \quad (19)$$

To obtain the p -value for \bar{V}_n , White (2000) suggested and justified using the stationary bootstrap of Politis and Romano (1994). This bootstrap p -value for testing \mathbb{H}_2 is called the “reality check p -value” for data snooping. As discussed in Hansen (2001), White’s reality check p -value may be considered as an upper bound of the true p -value. Hansen (2001) considered a modified reality check test which removes the “bad” models from the comparison and thereby improves the size and the power of the test. In Section 4 we will report both p -values of White (2000) and Hansen (2001), denoted as *White* and *Hansen* respectively.

The reality check to compare the performance of density forecast models in the tails can be implemented analogously as outlined above.

2.3 Remarks on Parameter Estimation Error

Let us remark on the parameter estimation uncertainty in $\psi_t(y_t; \hat{\theta}_R)$ ($t = R + 1, \dots, T$). If the parameter θ^* is known so that it needs not be estimated and if the density forecast model $\psi_t(y_t; \theta^*)$ coincides with the true density $\varphi_t(y_t)$, then $\{u_t\}$ is IID $U[0, 1]$ and $\{x_t\}$ is IID $N(0, 1)$. However, if the parameter $\hat{\theta}_R$ is estimated, it is possible that $\{u_t\}$ departs from IID $U[0, 1]$ and $\{x_t\}$ departs from IID $N(0, 1)$ even when the density forecast model $\psi_t(y_t; \theta^*)$ coincides with the true density $\varphi_t(y_t)$. We generate the PIT $\{u_t\}$ from (3), that is $u_t = \int_{-\infty}^{y_t} \psi_t(y; \hat{\theta}_R) dy$, instead from $\int_{-\infty}^{y_t} \psi_t(y; \theta^*) dy$ because θ^* is unknown. Hence, our reality check is based on $\hat{I}(\varphi : \psi, \hat{\theta}_R) \equiv \frac{1}{n} \sum_{t=R+1}^T [\ln \varphi_t(y_t) - \ln \psi_t(y_t; \hat{\theta}_R)]$ rather than $\hat{I}(\varphi : \psi, \theta^*) = \frac{1}{n} \sum_{t=R+1}^T [\ln \varphi_t(y_t) - \ln \psi_t(y_t; \theta^*)]$ in (2). The KLIC loss $\hat{I}(\varphi : \psi, \hat{\theta}_R)$ may be decompose as

$$\hat{I}(\varphi : \psi, \hat{\theta}_R) = \hat{I}(\varphi : \psi, \theta^*) + \frac{1}{n} \sum_{t=R+1}^T [\ln \psi_t(y_t; \theta^*) - \ln \psi_t(y_t; \hat{\theta}_R)], \quad (20)$$

where the first term is the loss due to model-specification error and the second term is the loss due to the parameter estimation error. It is important to note that in the comparison of the density forecast models we treat the parameter estimation error as a loss. A density forecast model that suffers less from the parameter estimation error (i.e., a more parsimonious model) will thus be preferred if it is correctly specified. This is in line with the model selection practice for which the bigger models are penalized. In this paper, a density forecast model will be said to be superior to another if the KLIC loss $\hat{I}(\varphi : \psi, \hat{\theta}_R)$ is smaller, and thus when it comes to comparing models for forecasting, both the model-specification-error and the parameter-estimation-error should be taken into a loss function. (To implement the reality check, the loss $\hat{I}(\varphi : \psi, \hat{\theta}_R)$ is estimated by $\bar{I}(\varphi : \psi, \hat{\theta}_R)$ from the AR(L)-SNP(K) model for $\{x_t\}$.)

Early work on the effects of the parameter estimation uncertainty on portfolio selection by Bawa *et al* (1979) has shown that the predictive distribution of an asset return that is obtained by integrating the conditional distribution over the parameter space is different from the predictive distribution that is obtained when the parameters are treated as known. The difference, that may be regarded as an additional

risk due to the parameter uncertainty, has been termed as “estimation risk”. More recently, Kandel and Stambaugh (1996), Barberis (2000), and Xia (2001), among others, explore the effect of the estimation risk on the economic importance in predicting stock returns, in selecting portfolio, in hedging, in long-horizon investment decision, and in market timing. We thus consider the estimation risk as a part of the loss function in our density forecast comparison as shown in (20).

Also, in financial risk management, both the regulators and financial risk professionals are very much interested in forecasting the risk of a given portfolio. In the risk management practice, financial institutions are faced with the model-risk which is the result of choosing a specific model for (say) interest rates or stock prices. The uncertainty arising from this potential misspecification is known as model-risk. However, since the institution is very much after the *total* risk measure due to the forecast errors, a distinction between “model risk” and “estimation risk” is not a concern. As a result, the evaluation tool in this context is backtesting the actual realized values and the forecasts from various alternative models. Throughout this process, it is not important to distinguish how much of the forecast error is sourcing from the model error or from the estimation error. For instance, the Bank for International Settlements (BIS) assesses the performance of a given institution’s risk system on the basis of its risk forecasting ability. In this assessment, what counts in the forecast comparison of various alternative models is the total forecast error regardless of the source of error.

2.4 Evaluating the Adequacy of Each Model

Not only we may use the KLIC measure to compare the different density forecast models but also we may use it to evaluate the adequacy of a density forecast model. Noting that the KLIC loss $\bar{I}_k(\varphi : \psi_k, \hat{\boldsymbol{\theta}}_{k,R})$ is related to the likelihood ratio (LR), it is interesting to use the LR statistic to assess the adequacy of each single model. To evaluate the adequacy of the density forecast, the LR statistic is to test the null hypothesis that $\{x_t\}$ follows IID $N(0, 1)$, that is $\boldsymbol{\rho} = \mathbf{0}, \sigma = 1$, and $\boldsymbol{\vartheta}_\eta = (r_1, \dots, r_K) = \mathbf{0}$. The LR statistic is equal to $2 \times n \times \bar{I}_k(\varphi : \psi_k, \hat{\boldsymbol{\theta}}_{k,R})$, which has the approximate asymptotic chi-squared distribution with the degrees of freedom of $(L + 1) + 1 + K$. With $L = 3$ and $K = 8$ (that we use in Section 4), its approximate asymptotic critical values are 19.81 (at 10% level), 22.36 (at 5% level), and 27.69 (at 1% level). In Section 4 we report the KLIC loss $\bar{I}_k(\varphi : \psi_k, \hat{\boldsymbol{\theta}}_{k,R})$ in the tables instead of the LR statistics. Thus the “critical” sample loss values for $\bar{I}_k(\varphi : \psi_k, \hat{\boldsymbol{\theta}}_{k,R})$ are these critical values divided by $2n$, which are 0.0058 (at 10% level), 0.0066 (at 5% level), and 0.0081 (at 1% level) for $n = 1700$. If the out-of-sample averaged loss of Model k , $\bar{I}_k(\varphi : \psi_k, \hat{\boldsymbol{\theta}}_{k,R})$, reported in the tables is greater than these values (say, 0.0066), then the model can be rejected as an adequate density forecast model at the pre-specified significance level.¹¹ This LR statistic is a

¹¹For an AR(3) model of x_t , we in fact lose the first three observations, so the critical value for the sample loss should be $c/2(n - 3)$. But there is almost no difference between $c/2n$ and $c/2(n - 3)$ for a large n .

generalized version of the LR statistic of Berkowitz (2001), who used $K = 0$ in the SNP density (i.e., normal density).

3 Application

In this section, we apply the reality check test under the KLIC-based loss function to investigate the adequacy of various density forecast models for the daily S&P500 and NASDAQ return series. For a typology of various models of $\{y_t\}_{t=1}^T$, let it follow the stochastic process

$$y_t = \mu_t + \varepsilon_t \equiv \mu_t + z_t \sigma_t, \quad (21)$$

where $\mu_t = \mathbb{E}(y_t | \mathcal{F}_{t-1})$, $\sigma_t^2 = \mathbb{E}(\varepsilon_t^2 | \mathcal{F}_{t-1})$, and $z_t \equiv \varepsilon_t / \sigma_t$. If the dependence structure of $\{y_t\}$ can be fully described by the first two conditional moments, $\{z_t\}$ may be independently and identically distributed (IID). We simply assume that μ_t follows an MA(1) process without a constant term (see, e.g., Anderson *et al.*, 2002; Hong and Lee, 2003). As can be seen, a density forecast model based on (21) can be decomposed into two parts: specification of σ_t^2 and specification of the conditional distribution of $\{z_t\}_{t=1}^T$. The symmetric and skewed parametric distributions to be described below can be used in very general situations. They formulate the basis of a likelihood function, based on which we can implement estimation, hypothesis testing, and so on. Also, inversion of the CDF function can give the quantile of the standardized residuals, and hence VaR of the return series. For the purpose of VaR alone, however, we may not necessarily assume a parametric form of the distribution to get the quantile. Most commonly used in the literature is a simple nonparametric model, namely, the historical distribution.

3.1 Distribution Specifications

We can specify the distribution of the standardized residuals $\{z_t\}_{t=1}^T$ by a conditional density function $f_t(z) = f_t(z | \mathcal{F}_{t-1}) = f_t(z | \mathcal{F}_{t-1}; \boldsymbol{\theta}_d)$ with a vector of distribution parameters $\boldsymbol{\theta}_d$ and a corresponding cumulative distribution function (CDF) $F_t(z) = F_t(z | \mathcal{F}_{t-1}) = F_t(z | \mathcal{F}_{t-1}; \boldsymbol{\theta}_d)$. Note that $\boldsymbol{\theta}_d$ may be time varying and a time subscript can be added, $\boldsymbol{\theta}_{d,t}$. Further, as common in the literature, $\boldsymbol{\theta}_{d,t}$ is specified as $\boldsymbol{\theta}(\mathcal{F}_{t-1}; \boldsymbol{\theta}_d)$ where the “true” parameter vector $\boldsymbol{\theta}_d$ is time-invariant. For time-varying distribution parameters, it is the case that z_t is not IID. For simplicity, we do not consider such complications in this paper and z_t is assumed to be IID. Hence we can drop the time subscript and write $f_t(z) = f(z)$. Given $f(\cdot)$ and conditional volatility $\sigma_t^2 = \sigma_t^2(\boldsymbol{\theta}_v)$ with a vector of volatility parameters $\boldsymbol{\theta}_v$, the conditional distribution of ε_t is $\psi_t(\varepsilon_t | \mathcal{F}_{t-1}) = \psi_t(\varepsilon_t | \mathcal{F}_{t-1}; \boldsymbol{\theta}) = f(z_t) / \sigma_t$, on which we build the sample log likelihood function

$$\mathcal{L}(\boldsymbol{\theta}) = \sum_{t=1}^T \ln[\psi_t(\varepsilon_t | \mathcal{F}_{t-1})] = \sum_{t=1}^T \ln \left[\frac{1}{\sigma_t} f(z_t) \right], \quad (22)$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}'_d, \boldsymbol{\theta}'_v)'$. Throughout, we use γ_1 and γ_2 , if they exist, to denote the skewness and *excess* kurtosis parameters specified by a distribution, respectively.

We have to confess that the candidate distributions included in this paper are far from being complete. Distributions that have been used in the literature but we do not use in this paper include the exponentially generalized beta distribution of type II (McDonald and Xu, 1995; Wang *et al.*, 2001), the generalized t (McDonald and Newey, 1988) and its skewed generalization (Theodossiou, 1998), the semiparametric density (González-Rivera and Engle, 1993), nonparametric density (Bao *et al.*, 2003), the local density (Gourieroux and Jasiak, 2001), the generalized hyperbolic density (Eberlin and Keller, 1995; Prause, 1999), and the stable Paretian distribution (Panorska *et al.*, 1995; Mittnik *et al.*, 1999), and others. In general, the cost of using these more general or complicated models is computational time and problem of sensitivity of the parameters to starting values and outliers in the data. Further, for some distributions, the second and higher moments do not even exist. Nonexistence of moments, especially the second moment, can cause some fundamental problem in pricing theory. For example, even though it is nested within the generalized t distribution, whose moment generating function does not exist, the generalized error distribution (GED) discussed below possesses the moment generating function, which renders straightforward derivation of moments that are needed to price derivative assets. Also, as discussed in Theodossiou (2000), the estimated distribution parameters of the GED are not that sensitive to outliers in the data compared with the generalized t distribution. A second example is the stable Paretian distribution, whose variance is infinite.

3.1.1 Symmetric Parametric Distributions

For symmetric distributions, $\gamma_1 = 0$, if the third moment exists. The standard Gaussian normal density has long been used in the literature primarily due to its simplicity and its well understood properties. The density function $\phi(z)$, the CDF $\Phi(z)$, together with its inverse $\Phi^{-1}(\alpha)$, have been well studied. Further, there is no distribution parameter to estimate. However, the assumption of normality is quite dubious in that $\phi(z)$ has a very restrictive shape with $\gamma_1 = 0$ and $\gamma_2 = 0$, which may not be consistent with the actual data. Departure from normality of financial data was first documented in Mandelbrot (1963). In particular, most empirical findings suggest that the standardized residuals remain leptokurtic. The GARCH-family models may explain some, but not all, of the leptokurtosis in the unconditional distributions. Hence we have many other candidate distributions in the following to accommodate excess kurtosis (as well as skewness for some distributions). Even though wrongly assuming normality can give consistency of estimated volatility parameters, efficiency can be potential problems in small samples. However, when proceeding along this line we should be cautious: wrongly assuming normality still yields consistency, but wrongly assuming a nonnormal distribution renders inconsistency of the parameters, let alone efficiency. The problem is that

we never know the data generating process (DGP). Hence we may regard all the distributions as well as volatilities as possibly misspecified and we are to compare these possibly misspecified models using an appropriate measure of the distance to the unknown DGP.

A second symmetric distribution widely used is the Student t distribution, which has fatter tails compared with the standard normal distribution. Since the kurtosis of the Student t distribution is determined by its degrees of freedom parameter ν , we estimate it simultaneously as a free parameter to capture the excess kurtosis that can not be explained by the GARCH-family models. A standardized Student t distribution is symmetric and $\gamma_1 = 0$ provided that $\nu > 3$. For $\nu > 4$, it gives $\gamma_2 = 6/(\nu - 4)$.

Nelson (1991) used the generalized error distribution (GED) of Subbotin (1923) to model the distribution of stock market returns. The GED is still symmetric, but is quite flexible in the tails through a parameter ν : when $\nu = 2$, it degenerates into a standard normal; when $\nu < 2$, it has fatter tails than the normal (for $\nu = 1$ it degenerates into a double exponential distribution); when $\nu > 2$, it has thinner tails than the normal (for $\nu = \infty$, it becomes a uniform distribution). The excess kurtosis parameter γ_2 is $\Gamma(1/\nu)\Gamma(5/\nu)/[\Gamma(3/\nu)]^2$. The double-exponential distribution (also known as the Laplace distribution), a special case of the GED, was used in Mittnik and Rachev (1993) to model unconditional distributions of asset returns. Granger and Ding (1995) and González-Rivera (1997) used it in conjunction with the GARCH model. It gives $\gamma_2 = 3$. The Double Weibull distribution (to be denoted as DW henceforth) with a tail parameter a , which generalizes the Laplace distribution, but differs from the GED, was also proposed in Mittnik and Rachev (1993) and can model $\gamma_2 = \Gamma(1 + 4/a)/[\Gamma(1 + 2/a)]^2$.

3.1.2 Skewed Parametric Distributions

While the GED and double Weibull distribution allow for considerable flexibility for γ_2 , they can not model skewness observed in many financial series. For example, Smikowitz and Beedles (1980) and Singleton and Wingender (1986) found evidence of positive skewness of individual stock returns; Badrinath and Chatterjee (1988) and Alles and Kling (1994) found the presence of skewness in indices of security and bond markets; Chen *et al.* (2000) observed that small firm's stocks tend to possess positive skewness in contrast to the observed negative skewness for large firm's stocks. Existence of skewness can be explained by the leverage effect (Black, 1976), volatility feedback (Pindyck, 1984), stochastic bubble (Blanchard and Watson, 1982), and heterogeneity among investors (Hong and Stein, 1999; Chen *et al.*, 2000). Neglecting skewness can lead to biased inference about risk and Harvey and Siddique (2000) incorporated measures of conditional (co)skewness to a asset pricing model. In this paper, we include eight flexible parametric distributions which *can* model both skewness and excess kurtosis.

The skewed t was first proposed by Fernández and Steel (1998) as a four-parameter skewed Student distri-

bution where the four parameters specifying the location, dispersion, skewness and kurtosis have meaningful interpretations. It is based on a mixture of two truncated symmetric Student distributions. The standardized version with only two parameters given in Lambert and Laurent (2001) was used with an APARCH model to model the NASDAQ return series and was found to be adequate in describing the data compared to the normal and Student distribution. γ_1 and γ_2 , functions of the two parameters ξ and ν , for a skewed t distribution are given in the appendix.

Closely related is Hansen’s (1994) generalized t with time-varying parameters. A skewness parameter λ_t directly controls the extent of asymmetry. When $\lambda_t = 0$, the distribution degenerates to the standard Student t distribution and when $\lambda_t > 0$ (< 0), the distribution is skewed to the right (left). A time invariant form of Hansen (1994) is used in Bond (2001), where the two parameters η and λ are estimated directly. Compared with the skewed t of Lambert and Laurent (2001), Hansen (1994) does not relate η and λ directly to the kurtosis and skewness, though this is done recently in Jondeau and Rockinger (2003). The allowed γ_1 and γ_2 are quite complicated, as shown in Jondeau and Rockinger (2003). The skewness or kurtosis may even cease to exist for some ranges of domain of (η, λ) . For a given kurtosis, the allowed skewness is quite restricted. If we impose existence of the density function given the first four moments, the excess kurtosis is bounded below by 0, which means that Hansen t does not allow tails to be thinner than the normal distribution. Instead of modelling the distribution parameters as function of \mathcal{F}_{t-1} , Harvey and Siddique (1999) modeled the skewness (only) directly as $\gamma_1 = \gamma_1(\mathcal{F}_{t-1})$, which implicitly determines the non-central t parameters used for the conditional distribution of the standardized series. It is claimed in Harvey and Siddique (1999) that Hansen’s (1994) t is an alternative parameterization of the non-central t distribution. For simplicity, we use a time-invariant version of Hansen (1994) in this paper.

Theodossiou (2000) gave the skewed version of the GED with two parameters which control skewness and kurtosis respectively. The skewed GED was used to investigate the empirical distribution of returns of several financial assets at different frequencies and is found to provide a good fit.

The inverse hyperbolic sine (IHS) transformation with two parameters is used in Hansen *et al.* (2000) and Choi (2001) to model asymmetric and fat-tailed distributions. It is based on Johnson’s (1949) hyperbolic sine transformation of a standard normal distribution. Besides being “one of the most flexible distribution among all skewed and leptokurtic distributions” (Choi, 2000, p.8), calculation of the inverse CDF (quantile) is extremely simple compared with other skewed parametric distributions. He used the IHS model on several stock market indices and found that it can capture tail behavior better than the extreme-value-theory-based models.

Another distribution which can exhibit skewness and excess kurtosis is the mixture of two normal distributions (denoted as MN henceforth). See Venkataraman (1997), Bee (2000), and Wang (2001) for its

application. Of course, in principle, the components to be mixed can be more than two and need not necessarily be normal. For example, Neumann (1998) used the mixture of lognormal distributions to model option pricing. For simplicity, in this paper we focus on a mixture of two normals. After standardization, it has only three parameters.

The Double Gamma distribution (DG, henceforth) is another candidate distribution, which was first proposed by Knight *et al.* (1995) as a method of capturing asymmetry in the distribution of financial returns. Bond (2002) applied it to model the downside risk in exchange rate data in a conditional volatility framework. It has three parameters.

The Gram-Charlier / Edgeworth-Sargan density (to be simply denoted as Sargan henceforth), has also been used in the literature to model skewness and excess kurtosis. See Mauleón and Perote (1995), Perote and Del Brío (1999), Jondeau and Rockinger (2001), and Straja (2001). It takes the form

$$f_n(z) = \sum_{i=0}^n c_i H_i(z) \phi(z),$$

where n is the order of the approximating density, $H_i(\cdot)$ is the i th-order Hermite polynomial, defined as $(-1)^i (\partial\phi/\partial z^i)/\phi(z)$. Usually we set $n = 4$ or 6 , which give the Edgeworth type-A and type-B density, respectively, as follows

$$\begin{aligned} f_4(z) &= \left[1 + \frac{\gamma_1}{6} H_3(z) + \frac{\gamma_2}{24} H_4(z) \right] \phi(z), \\ f_6(z) &= \left[1 + \frac{\gamma_1}{6} H_3(z) + \frac{\gamma_2}{24} H_4(z) + \frac{\gamma_1^2}{72} H_6(z) \right] \phi(z). \end{aligned}$$

As shown by Barton and Dennis (1952), to guarantee the positiveness of the approximate density, the range of skewness and excess kurtosis parameters is smaller for type-B density than type-A density. Following Jondeau and Rockinger (2001), we use $f_4(z)$ and regard γ_1 and γ_2 as parameters to be estimated by the maximum likelihood method. In practice, we adopt the algorithm developed by Jondeau and Rockinger (2001) to implement the positiveness constraint. Briefly, the envelope of the domain of admissible skewness and excess kurtosis, $\mathcal{D}(\gamma_1, \gamma_2)$, which solves simultaneous $f_4(z) = 0$ and $f_4'(z) = 0$, is characterized by

$$\gamma_1(z) = -24 \frac{H_3(z)}{d(z)}, \quad \gamma_2(z) = 72 \frac{H_2(z)}{d(z)},$$

where $d(z) = 4H_3^2(z) - 3H_2^2(z)H_4(z)$. Therefore, we have a one-to-one correspondence between γ_2 and $|\gamma_1|$ and the range of γ_2 is $[0, 4]$ while the range of γ_1 is symmetric around a given γ_2 . The maximal $|\gamma_1|$ is 1.0493 when $\gamma_2 = 2.4508$. To begin with, we partition $[0, 4]$ into $N - 1$ equally spaced intervals with grid points γ_{2i} , $i = 1, \dots, N$. For each γ_{2i} , we solve for $|\gamma_{1i}|$. For a given γ_2 , the upper and lower bounds, γ_1^U and γ_1^L , of γ_1 can be found by linear interpolation. That is, find i such that $\gamma_{2i} < \gamma_2 < \gamma_{2,i+1}$ and the bounds of γ_1 are

$$\gamma_1^U = a_i + b_i \gamma_2, \quad \gamma_1^L = -\gamma_1^U,$$

$$a_i = \frac{|\gamma_{1i}| \gamma_{2,i+1} - \gamma_{2i} |\gamma_{1,i+1}|}{\gamma_{2,i+1} - \gamma_{2i}}, \quad b_i = \frac{|\gamma_{1,i+1}| - |\gamma_{1i}|}{\gamma_{2,i+1} - \gamma_{2i}}.$$

We can use the logistic transformation to impose the constraints on γ_1 and γ_2 so that we have a unconstrained maximizing exercise.

3.1.3 Historical Distribution

For the calculation of VaR, a quite simple way is to use is the historical distribution (HS, henceforth). One criticism of this methodology is that it ignores information updating. However, it is so easy to implement that it is still widely used in practice for various purposes. This method uses the empirical quantiles of the historical distribution of past observations to estimate $F^{-1}(\alpha)$. In other words, for a sample of size n , we simply assume that $f(z_t) = 1/n$ and $F(z) = \frac{1}{n} \sum_{t=1}^n 1(z_t \leq z)$.

3.2 Volatility Specifications

The conditional variance σ_t^2 can be specified with various volatility models. We can specify it nonparametrically (Bühlmann and McNeil, 2001), parametrically (the GARCH family), or through some stochastic volatility model (Taylor, 1986, 1994; Ghysels *et al.*, 1996). For comparison of these different volatility specifications, see Poon and Granger (2003) and references therein. In this paper, we focus on the GARCH family.

Engel (1982) first introduced the ARCH model, where σ_t^2 is a function of past squared errors. Bollerslev (1986) developed the GARCH model, where σ_t^2 depends upon its own lagged values as well as the squared errors. The exponentially weighted moving average (EWMA) of Riskmetrics is a special case of the nonstationary version of GARCH, the so-called integrated GARCH (IGARCH), where all the volatility parameters (excluding the constant term) sum to 1. The GARCH model captures a salient feature of financial series: volatility persistence. However, while the GARCH model features an exponential decay in the autocorrelation of conditional variance, the IGARCH model assumes an infinite horizon of decay. Empirical findings suggest that a shock in volatility seems to have long, but not infinite, memory. This gives rise to the fractionally integrated GARCH model (FIGARCH) of Bollerslev and Mikkelsen (1996), where a fractional parameter d controls the rate of hyperbolic decay in the autocorrelation of conditional variance. A generalization of FIGARCH is the hyperbolic GARCH (HYGARCH) of Davidson (2002). The component GARCH (CGARCH) model of Ding and Granger (1996) and Engle and Lee (1999) is also capable of capturing slow decay in the second moments. An issue of interest is to investigate and compare the adequacy of the two different types of volatility models that can both produce the very slow decay in the autocorrelation of conditional variance.

The above mentioned GARCH-family models are symmetric in the sense that positive and negative shocks have the same effects on the conditional variance. Engle and Ng (1993) proposed using the news impact curve

to evaluate asymmetric GARCH models, in which the feedback mechanisms of positive and negative shocks (“good” and “bad” news in their terms) are modelled differently. Asymmetric GARCH models include the exponential GARCH (EGARCH) model of Nelson (1991), the GJR-GARCH of Glosten *et al.* (1993), the threshold GARCH (TGARCH) of Zakoian (1994), the smooth transition GARCH (STGARCH) of González-Rivera (1998), the asymmetric power ARCH (APARCH) of Ding *et al.* (1993), and others. Note that the APARCH model nests ARCH if $\delta = 2$, $\varphi_i = 0$ and $\beta_j = 0$, nests GARCH if $\delta = 2$ and $\varphi_i = 0$, nests GJR if $\delta = 2$ and nests TGARCH if $\delta = 1$ (see the appendix). If we combine the idea of fractional integration with asymmetry GARCH models, we can easily have the FIEGARCH, FIAPARCH, and so on.

In practice, we have to impose some conditions on the volatility parameters to guarantee the positiveness of the conditional variance, except for the EGARCH model. For some special cases, further restrictions can be imposed to achieve stationarity. Given the possibility of hundreds of various GARCH-family models (e.g., Hansen and Lunde, 2002), we select only the following models in this paper: EWMA, GARCH, GJR, APARCH, EGARCH, STGARCH, FIGARCH, FIAPARCH, CGARCH. For all the fractionally integrated models, we use its generalized version HYGARCH, say. That is, we estimate an additional parameter κ (see the appendix). Also, we fix p , lag of the conditional variance, and q , lag of the squared errors, to be both 1.

4 Empirical Results

In this section, we use two data sets to compare the 117 density forecast models described in Section 3 (thirteen different distribution models and nine different volatility models). The two data sets are the daily S&P 500 and NASDAQ return series, retrieved from *finance.yahoo.com* and CRSP.¹² They are from January 3, 1990 to June 30, 2003 ($T = 3403$). We split the sample into two parts (roughly into two halves): one for in-sample estimation of size $R = 1703$ and another for out-of-sample density forecast of size $n = 1700$. We use a rolling-sample scheme. That is, the first density forecast is based on observations 1 through R (January 3, 1990 to September 24, 1996), the second density forecast is based on observations 2 through $R + 1$ (January 4, 1990 to September 25, 1996), and so on.

Table 1 and Table 2 give the empirical results of density forecast comparison under the KLIC loss function for S&P 500 and NASDAQ. A model in each cell, corresponding to a particular density specification (row) in conjunction with a particular volatility specification (column), is regarded as a benchmark model and it is compared with the remaining 116 models. We set the number of bootstraps for the reality check to be 1000 and the mean block length 4, which corresponds to $q = 0.25$ in White (2000). The averaged out-of-sample KLIC loss $\bar{I}_k(\varphi : \psi_k, \hat{\theta}_{k,R})$ (the first number in each cell) as well as the reality check p -value of White (2000)

¹²There were a few missing observations in the NASDAQ series from *finance.yahoo.com*, which were checked against the CRSP data provided by Canlin Li. Other than these few observations, the Nasdaq series from the two sources were consistent. Regarding the week following September 11, 2001, we treat it as a long holiday.

(the second number in each cell) and the modified reality check p -value of Hansen (2001) (the third number in each cell) are presented.

With the AR(L)-SNP(K) model specified in (7) and (8), we use the Akaike and Schwarz information criteria (AIC, SIC) for the selection of L and K . In general, for most of the 117 models under comparison, large values of the SNP order (e.g., $K = 8$) seem to be better than smaller values (e.g., $K = 3$) in terms of AIC and SIC, indicating the prevalent departures of $\{x_t\}$ from the normality. On the other hand, the different AR orders $L = 1, 2$, or 3 make less difference in AIC or SIC, indicating the less prevalent departures of $\{x_t\}$ from the IID-ness across the various density forecast models. For instance, the reality check test based on AR(3)-SNP(8) is more powerful than the test based on AR(3)-SNP(3), while AR(3)-SNP(8) and AR(1)-SNP(8) give somewhat similar results. This may be consistent with our empirical findings that the choice of the distributional models is much more important than the choice of the volatility models. Once a right distribution has been chosen, a simple GARCH model may be as good as more complicated volatility models. Therefore, in reporting our results for all 117 models, we choose $L = 3$ for the AR model in (7) and $K = 8$ for the SNP model in (8).

The estimated out-of-sample KLIC (15) and its censored versions as defined from (10) with four different values of τ , each corresponding to $\alpha = 1.00, 0.10, 0.05$, and 0.01 , are reported in Panels A, B, C, and D for each of Tables 1 and 2. Note that in these tables a smaller value of the first number of each cell indicates a lower sample loss and hence a better density forecast model from a pair of volatility and distribution. The larger reality check p -values (the second or the third numbers of each cell), the better density the forecast model corresponding to the cell, as we fail to reject the null hypothesis that the other remaining 116 competing models is no better than this model (the benchmark model). (In general, a low sample loss should indicate a high reality check p -value, but this correlation is not perfect since the testing not only depends on the point value of the loss differential, but also its variance.)

4.1 Results on S&P500

In Table 1 the reality check results for the S&P500 data with the 100%, 10%, 5%, and 1% tails presented through Panels A-D.

The best specification in Panel A for $\alpha = 1.00$ is obtained by the model of DG+GJR where the loss is 0.0003 and the p -values of White (2000) and Hansen (2001) are 1.000 and 0.977, respectively. As discussed in Section 2.4, the sample loss values $\bar{I}_k(\varphi : \psi_k, \hat{\theta}_{k,R})$ greater than 0.0058 (at 10% level), 0.0066 (at 5% level), and 0.0081 (at 1% level) would indicate inadequacy of the density forecast. For Normal, Student- t , DW, MN, DG, and Sargan, there are several volatility models that give adequate density forecast models. All the other distributions (HS, GED, Laplace, Skewed- t , Hansen- t , Skewed GED, IHS) do not provide any

adequate density forecast models in combination with any of the volatility models. Laplace, Skewed- t , and Hansen- t are the weakest distribution specifications for the whole distribution.

The results for the 10% tail of the S&P500 returns presented in Panel B show that Hansen- t with APARCH, EGARCH, HYAPARCH, GJR, and HYGARCH are clearly among the best models, with large reality check p -values. All the remaining models are clearly dominated by these models as indicated by the small reality check p -values of those models. Skewed- t is the second best distributional model. However, no model may be considered as adequate because even the best model has $\bar{I}_k(\varphi : \psi_k, \hat{\theta}_{k,R}) = 0.0120$ that is larger than 0.0081.

In Panel C, the results on S&P500 for the 5% tail, Hansen- t with HYGARCH attains the smallest loss, followed by Hansen- t with GJR. These two models has the sample-averaged loss smaller than 0.0058, indicating their adequacy as a density forecast model. All the other alternative models are not adequate, with very low reality check p -values and the large loss values.

For the 1% tail (Panel D), Hansen- t generates again the best density forecast model in combination with most volatility models considered. Sargan may be the next to Hansen- t . Perhaps surprisingly, many more distributions (for the 1% tail than for 5% and 10% tails) turn out to be adequate as the loss values are small and the reality check p -values are large. For HS, Laplace, Skewed- t , Hansen- t , MN, DG, and Sargan, at least 8 out of 9 volatility models produce the adequate density forecast models (with the loss values smaller than 0.0081). The worst distribution model for the 1% tail is Normal, which does not produce any adequate density forecast model with any combination of the nine volatility models. We also note that the distributional model exhibits much more robust performance across the different combinations with the volatility models than the other way around. That is, a good volatility model for some distributional model can often become a very bad choice with other distributional model. Once a good distributional model has been chosen, the choice of the volatility model may not be important. Hence the distribution choice is much more important than the volatility choice.

4.2 Results on NASDAQ

In Table 2, the results for the 100%, 10%, 5%, and 1% tails of the NASDAQ daily returns are presented in Panels A-D.

In Panel A (for the 100% tail or for the whole distribution), the best density forecast model is DG+GJR where the loss is 0.0002 with $White = 0.991$ and $Hansen = 0.856$. For DG and GED, there are a couple of volatility models that give adequate density forecast models. All the other distributions do not provide any adequate density forecast models in combination with any of the volatility models. Hansen- t is the weakest distribution specifications for the whole distribution.

For the 10% tail, Hansen- t with GARCH specification produces the best performance. None of the other distribution and volatility specifications produce adequate density forecasts.

When we turn to the 5% tail, as in the case of S&P500, the best pairs of distribution and volatility specification are Hansen- t with GARCH, GJR, APARCH, HYGARCH, HYAPARCH, EGARCH. Hansen- t is clearly the best distributional model with robust performance with many different volatility specifications. The weakest volatility specification appears to be EWMA. All the other distribution specifications produced rather weak predictive power for this tail. Weak performance of Normal distribution is also evident.

Finally, when we move to the 1% tail, a clear advantage of Hansen- t is distribution is still observed. In eight of nine volatility models, Hansen- t produced the best predictive power. This distribution is followed by MN, DG, Skewed- t , Sargan, and Laplace. Weakness of Normal and HS distributions is also evident for NASDAQ return tails. Similar observations to those for S&P500 may also be found in NASDAQ – once a good distributional model has been chosen, the choice of the simple volatility model such as GARCH may be as good as any other complicated volatility models.

4.3 Summary

The main findings for S&P500 and NASDAQ are more or less the same. In terms of the tail density forecast, a larger number of density forecast models turn out to be adequate for the 1% tail density forecast, than for the 5% tail density forecast or for the 10% tail density forecast. For the whole return distributions of S&P500 and NASDAQ, DG and Normal produce the best density forecast models and Hansen- t is the worst. For tails however, the opposite has been observed. Hansen- t generally produces the best tail density forecast models together with most volatility specifications considered, while Normal is the worst.

Once a good distribution is chosen, the choice of the volatility model may be secondary particularly for the tail density forecasting. For the tail density forecast, if Hansen- t is chosen, a more complex volatility model than the standard GARCH model may not be necessary. This is more strongly observed for a more extreme tail (i.e., more so for the 1% tail than for the 5% tail, even more than the 10% tail). For the whole return distribution, however, the choice of the volatility may be important as only a few volatility models are shown to be adequate for the best choice of the distributional model. Therefore, perhaps to the contrary to some common perception, these results show that it is more difficult to come up with a good density forecast model for the entire return distribution than for the tails.

We summarize our findings now.

1. In terms of the density forecasts, the choice of conditional distribution is more important than the choice of conditional volatility models. Given a fixed volatility model, the variation in the sample loss

values as well as the reality check p -values across various distribution models are much bigger compared with the variation across volatility models given a fixed distribution model.

2. For tails and both in S&P500 and NASDAQ, Hansen- t appears to be the best distribution for the 10%, 5%, and 1% tails.
3. For modeling the whole return distribution, on average, Double Gamma distribution appears to be the best choice among other alternatives. Unlike its success on the tails, performance of Hansen- t appears to be the weakest for modeling the whole distribution.
4. The consistent success of Hansen- t distribution for tails, regardless of the volatility choice, indicates the existence of skewness and fat-tailedness in both S&P500 and NASDAQ return series. The results are similar for these two data sets. However, Hansen- t was very weak for modeling the whole distribution.
5. There are several lessons for the practitioners and academics. First, the commonly used Normal distribution is not capable of generating adequate tail density forecast models. Second, along the lines of Poon and Granger's (2003) findings, volatility forecasting can be made better if the underlying conditional distribution is chosen rigorously. Third, without explicitly taking into account of fat tails and conditional skewness, modeling tails and risk forecasting can not be conducted in an adequate accuracy. Relative success of Hansen- t and other skewed distributions on modeling the return distributions of S&P500 and NASDAQ is a clear indication of this finding. In addition, very poor performance of Normal and HS regardless of the choice of volatility constitutes another reason why special care is required to model tails for risk forecasting purposes.

5 Conclusions

In this paper we propose a new method to compare various possibly misspecified density forecast models. We use the minimum Kullback-Leibler Information Criterion (KLIC) divergence measure between a candidate model and the true model. We use the flexible SNP density to construct the LR statistics, generalizing the LR statistics of Berkowitz (2001) for density forecast evaluation without assuming normality. To compare the performance of density forecast models in the tails, we also use a censored LR statistic to estimate the tail minimum KLIC.

Our empirical findings based on the daily S&P500 and NASDAQ return series confirm the recent evidence on fat-tailedness and asymmetry in financial return distributions. Hansen- t , a distribution that captures these two properties, appears to produce the best density forecast in tails. Another important finding of this paper is that successful density forecast depends much more heavily on the choice of distributional model than the

choice of volatility model. This finding may be related to why the vast of volatility forecasting studies in the literature may be inconclusive in ranking the volatility models.

A model that provides superior density forecasts for the whole distribution does not necessarily meet the needs of risk managers who care much more about the tails. Related to this issue, we have found, modeling the whole return distribution and the tails of the return distribution can not be easily accomplished with the same density forecast models. Besides, the weak performance of Normal and Historical distributions questions their wide use in financial modeling, including the VaR measures used by practitioners.

The method developed in this paper can be used to compare alternative credit risk models and to compare density forecasts of portfolio of assets, which we leave for future study.

References

- Alles, L. A. and J. L. Kling. 1994. Regularities in the variation of skewness in asset returns. *Journal of Financial Research* 17: 427-438.
- Anderson, T. G., L. Benzoni, and J. Lund. 2002. An empirical investigation of continuous-time equity return models. *Journal of Finance* 57: 1239-1284.
- Badrinath, S. G. and S. Chatterjee. 1988. On measuring skewness and elongation in common stock return distributions: the case of the market index. *Journal of Business* 61: 451-472.
- Bai, J. 2003. Testing parametric conditional distributions of dynamic models. *Review of Economics and Statistics* 85: 531-549.
- Bao, Y., T.-H. Lee, and B. Saltoglu. 2003. Evaluating predictive performance of value-at-risk models in emerging markets: a reality check. Working paper, University of California, Riverside.
- Barberis, N. 2000. Investing for the long run when returns are predictable. *Journal of Finance* 55: 225-264.
- Barton, D. E. and K. E. R. Dennis. 1952. The conditions under which Gram-Charlier and Edgeworth curves are positive definite and unimodal. *Biometrika* 29: 425-427.
- Bauwens L., P. Giot, J. Grammig, and D. Veredas. 2003. A comparison of financial duration models via density forecasts. CORE Discussion Papers 20, forthcoming, *International Journal of Forecasting*.
- Bawa, V. S., S. J. Brown, and R. W. Klein. 1979. *Estimation Risk and Optimal Portfolio Choice*. North-Holland: New York.
- Bee, M. 2001. Mixture models for VaR and stressing testing. Technical Report ALEA 12, Informatica e Studi Aziendali, University of Trento.
- Berkowitz, J. 2001. Testing density forecasts with applications to risk management. *Journal of Business and Economic Statistics* 19: 465-474.
- Black, F. 1976. Studies of stock price volatility changes. *Proceedings of the American Statistical Association (Business and Statistics Section)*, 177-181.
- Blanchard, O. J. and M. W. Watson. 1982. Bubbles, rational expectations, and financial markets, 295-316. In *Crises in Economic and Financial Structure*, P. Wachtel (ed.), Lexington, MA: D.C. Heath and Company.
- Bollerslev, T. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31: 307-327.
- Bollerslev, T. and H. O. Mikkelsen. 1996. Modelling and pricing long memory in stock market volatility. *Journal of Econometrics* 73: 151-184.
- Bond, S. A. 2001. A review of conditional density functions in autoregressive conditional heteroscedasticity models, 21-46. In *Return Distributions in Finance*, J. Knight and S. E. Satchell (eds.), Oxford: Butterworth and Heinemann.
- Bond, S. A. 2002. Asymmetry and downside risk in foreign exchange markets. Working paper, University of Cambridge.
- Bontemps, C. and N. Meddahi. 2003. Testing normality: A GMM approach. Working paper, Université de Montréal.
- Bühlmann, P. and A. J. McNeil. 2001. An algorithm for nonparametric GARCH modelling. Working paper, Federal Institute of Technology, Zurich, Germany.
- Chen, J., H. Hong, and J. C. Stein. 2000. Forecasting crashes: trading volume, past returns and conditional skewness in stock prices. NBER Working Paper 7687.
- Choi, P. 2001. Estimation of value at risk using Johnson S_U -normal distribution. Working paper, Texas A&M University.
- Christoffersen, P. F. 1998. Evaluating interval forecasts. *International Economic Review* 39: 841-864.
- Christoffersen, P. and F. X. Diebold. 1996. Further results on forecasting and model selection under asymmetric loss. *Journal of Applied Econometrics* 11: 561-572.

- Christoffersen, P. and F. X. Diebold. 1997. Optimal prediction under asymmetric loss. *Econometric Theory* 13: 808-817.
- Clements, M. P. and J. Smith. 2000. Evaluating the forecast densities of linear and non-linear models: applications to output growth and unemployment. *Journal of Forecasting* 19: 255-276.
- Clements, M. P. 2002. An evaluation of the survey of professional forecasters probability distributions of expected inflation and output growth. Working paper, University of Warwick.
- Corradi, V. and N. Swanson. 2003a. The block bootstrap for recursive m-estimations with application to predictive evaluations. Working paper, University of Exeter and Rutgers University.
- Corradi, V. and N. Swanson. 2003b. A test for comparing multiple misspecified conditional distributions. Working paper, University of Exeter and Rutgers University.
- Davidson, J. 2002. Moment and memory properties of linear conditional heteroscedasticity models. Working paper.
- Diebold, F. X., T. A. Gunther, and A. S. Tay. 1998. Evaluating density forecasts with applications to financial risk management. *International Economic Review* 39: 863-883.
- Diebold, F. X., J. Hahn, and A. S. Tay. 1999. Multivariate density forecast evaluation and calibration in financial risk management: high-frequency returns of foreign exchange. *Review of Economics and Statistics* 81: 661-673.
- Diebold, F. X. and R. Mariano. 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* 13: 253-265.
- Ding, Z. and C. W. J. Granger. 1996. Modeling volatility persistence of speculative returns: a new approach. *Journal of Econometrics* 73: 185-215.
- Ding, Z., C. W. J. Granger, and R. F. Engle. 1993. A long memory property of stock market returns and a new model. *Journal of Empirical Finance* 1: 83-106.
- Duan, J. C. 2003. A specification test for time series models by a normality transformation. Working paper, University of Toronto.
- Engle, R. F. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation", *Econometrica* 50: 987-1008.
- Engle, R. F. and G. G. L. Lee. 1999. A long-run and short-run component model of stock return volatility. In *Cointegration, Causality, and Forecasting*, R. F. Engle, and H. White (eds.), Oxford University Press.
- Engle, R. and V. K. Ng. 1993. Measuring and testing the impact of news on volatility. *Journal of Finance* 48: 1749-1778.
- Eberlin, E. and U. Keller. 1995. Hyperbolic distributions in finance. *Bernoulli* 1: 281-299.
- Fernandez-Villaverde, J. and J. F. Rubio-Ramirez. 2002. Comparing dynamic equilibrium models to data. Working paper, University of Pennsylvania.
- Freichs, H. and G. Löffler. 2003. Evaluating credit risk models using loss density forecasts. Working Paper, University of Frankfurt.
- Fernández, C. and M. Steel. 1998. On Bayesian modelling of fat tails and skewness. *Journal of the American Statistical Association* 93: 359-371.
- Gallant, A. R. and D. W. Nychka. 1987. Semi-nonparametric maximum likelihood estimation. *Econometrica* 55: 363-390.
- Ghysels, E., A. C. Harvey, and E. Renault. 1996. Stochastic volatility. In *Handbook of Statistics*, Vol. 14, C. R. Rao and G. S. Maddala (eds.); Amsterdam: North-Holland.
- Giacomini. R. 2002. Comparing density forecasts via weighted likelihood ratio tests: Asymptotic and bootstrap methods. Working paper, University of California, San Diego.
- Glosten, L., R. Jagannathan, and D. Runkle. 1993. Relationship between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* 48: 1779-1801.

- González-Rivera, G. 1997. A Note on adaptation in GARCH models. *Econometric Reviews*, 16: 55-68.
- González-Rivera, G. 1998. Smooth-transition GARCH models. *Studies in Nonlinear Dynamics and Econometrics* 3: 61-78.
- González-Rivera, G. and R. F. Engle. 1991. Semiparametric ARCH models. *Journal of Business and Economic Statistics* 9:345-360.
- Gourieroux, C. and J. Jasiak. 2001. Local likelihood density estimation and value at risk. Working paper. CREST and CEPREMAP and York University.
- Granger, C. W. J. 1999a. Outline of forecast theory using generalized cost functions. *Spanish Economic Review* 1: 161-173.
- Granger, C. W. J. 1999b. *Empirical Modeling in Economics*. Cambridge: Cambridge University Press.
- Granger, C. W. J. and Ding Z. 1995. Some properties of absolute returns, an alternative measure of risk. *Annales d'Economie et de Statistique* 40: 67-91.
- Granger, C. W. J. and M. H. Pesaran. 2000a. A decision theoretic approach to forecasting evaluation. In *Statistics and Finance: An Interface*, W. S. Chan, W. K. Li, and Howell Tong (eds.), London: Imperial College Press.
- Granger, C. W. J. and M. H. Pesaran. 2000b. Economic and statistical measures of forecast accuracy. *Journal of Forecasting* 19: 537-560.
- Hall, A. 1990. Lagrange multiplier tests for normality against semiparametric alternatives. *Journal of Business and Economic Statistics* 8: 417-426.
- Hansen, B. E. 1994. Autoregressive conditional density estimation. *International Economic Review* 35: 705-730.
- Hansen, C. B., J. B. McDoanld, and P. Theodossiou. 2000. Some flexible parametric models for skewed and leptokurtic data. Working paper, Brigham Young University and Rutgers University.
- Hansen, P. R. 2001. An unbiased and powerful test for superior predictive ability. Working paper, Brown University.
- Hansen, P. R. and A. Lunde. 2002. A forecast comparison of volatility models: does anything beat a GARCH(1,1)? Working paper, Brown University and Aarhus School of Business.
- Harvey, C. and A. Siddique. 1999. Autoregressive conditional skewness. *Journal of Financial and Quantitative Analysis* 34: 465-488.
- Harvey, C. R. and A. Siddique. 2000. Conditional skewness in asset pricing tests. *Journal of Finance* 55: 1263-1295.
- Hong, H. and J. L. Stein. 1999. Differences of opinion, rational arbitrage and market crashes. NBER Working Paper 7376.
- Hong, Y. 2002. Evaluation of out-of-sample probability density forecasts with applications to stock prices. Working paper, Cornell University.
- Hong, Y. and T.-H. Lee. 2003. Diagnostic checking for adequacy of nonlinear time series models. *Econometric Theory* 19: 1065-1121.
- Hong, Y. and H. Li. 2002. Nonparametric specification testing for continuous-time models with application to spot interest rates. Working paper, Cornell University.
- Jackwerth, J. C. and M. Rubinstein. 1996. Recovering probability distributions from option prices. *Journal of Finance* 51: 1611-1631.
- Jarque, C. M. and A. K. Bera. 1980. Efficient tests for normality, heteroscedasticity, and serial dependence of regression models. *Economics Letters* 6: 255-259.
- Johnson, N. L. 1949. Systems of frequency curves generated by method of translation. *Biometrika* 36: 149-176.
- Johnson, N. L., S. Kotz, and N. Balakrishnan. 1994. *Continuous univariate distributions*. New York: Wiley.

- Jondeau, E. and M. Rockinger. 2001. Gram-Charlier densities. *Journal of Economic Dynamics and Control* 25: 1457-1483.
- Jondeau, E. and M. Rockinger. 2003. Conditional volatility, skewness, and kurtosis: existence, persistence, and comovements. *Journal of Economic Dynamics and Control* 27: 1699-1737.
- Kandel, S. and R. F. Stambaugh. 1996. On the predictability of stock returns: An asset-allocation perspective. *Journal of Finance* 51: 385-424.
- Kiefer, N. M. and M. Salmon. Testing normality in econometric models. *Economics Letters* 11: 123-127.
- Knight, J. L., S. E. Satchell, and K. C. Tran. 1995. Statistical modelling of asymmetric risk in asset returns. *Applied Mathematical Finance* 2: 155-172.
- Kullback, L. and R. A. Leibler. 1951. On information and sufficiency. *Annals of Mathematical Statistics* 22: 79-86.
- Lambert, P. and S. Laurent. 2001. Modelling financial time series using GARCH-type models with a skewed Student distribution for the innovations. Working paper, Université Catholique de Louvain and Université de Liège, Belgium.
- Lee, L.-F. 1984. Maximum likelihood estimation and a specification test for non-normal distributional assumptions for the accelerated failure time models. *Journal of Econometrics* 24: 159-179.
- Mandelbrot, B. B. 1963. The variation of certain speculative prices. *Journal of Business* 36: 394-419.
- Mauleón, I. and J. Perote. 1995. Testing densities with financial data: an empirical comparison of the Edgeworth-Sargan density to Student's t . *European Journal of Finance* 6: 225-239.
- McDonald, J. B. and W. Newey. 1988. Partially adaptive estimation of regression models via the generalized t Distribution. *Econometric Theory* 4: 428-457.
- McDonald, J. B. and Y. J. Xu. 1995. A generalization of the beta distribution with applications. *Journal of Econometrics* 66: 133-152.
- Mittnik, S. and T. Rachev. 1993. Modeling asset returns with alternative stable models. *Econometric Reviews* 12: 261-330.
- Mittnik, S., S. T. Rachev, T. Doganoglu, and D. Chenyao. 1999. Maximum likelihood estimation of stable paretian models. *Mathematical and Financial Modelling* 29: 275-293.
- Mittnik, S. and M. Paoletta. 2003. Prediction of financial downside risk with heavy tailed conditional distributions. CFS Working Paper, 2003-04.
- Nelson, D. B. 1991. Conditional heteroscedasticity in asset returns: a new approach. *Econometrica* 59: 347-370.
- Neumann, M. 1998. Option pricing under the mixture of distributions hypothesis. Working paper, Institut für Entscheidungstheorie und Unternehmensforschung, Universität Karlsruhe.
- Noceti, P., J. Smith, and S. Hodges. 2003. An evaluation of tests of distributional forecasts. *Journal of Forecasting* 22: 447-455.
- Panorska, A. K., S. Mittnik, and S. T. Rachev. 1995. Stable GARCH models for financial time series. *Applied Mathematics Letters* 8: 33-37.
- Patton, A. and A. Timmermann. 2003. Properties of optimal forecasts. Working paper, University of California, San Diego.
- Perote, J. and E. Del Brío. 1999. Edgeworth-Sargan forecasting for high volatility periods. Working paper, Universidad de Salamanca.
- Pesaran, M. H. and S. Skouras. 2001. Decision-based methods for forecast evaluation. In *Companion to Economic Forecasting*, M. P. Clements and D. F. Hendry (eds.), Malden: Basil Blackwell.
- Pindyck, R. S. 1984. Risk, inflation, and the stock market. *American Economic Review* 74: 334-351.
- Politis, D. N. and J. P. Romano. 1994. The stationary bootstrap. *Journal of American Statistical Association* 89: 1303-1313.

- Poon, S. and C. Granger. 2003. Forecasting volatility in financial markets: A review. *Journal of Economic Literature* 41: 478-539.
- Prause, K. 1999. The generalized hyperbolic model: estimation, financial derivatives, and risk measures. Doctoral thesis, University of Freiburg.
- Sawa, T. 1978. Information criteria for discriminating among alternative regression models. *Econometrica* 46: 1273-1291.
- Singleton, J. C., and J. Wingender. 1986. Skewness persistence in common stock returns. *Journal of Financial and Quantitative Analysis* 21: 335-341.
- Smikowitz, M. A. and W. L. Beedles. 1980. Asymmetric stable distributed security returns. *Journal of the American Statistical Association* 75: 306-312.
- Smith, R. 1989. On the use of distributional misspecification checks in limited dependent variable models. *Economic Journal Supplement* 99: 178-192.
- Soderling, P. and L. Svensson. 1997. New techniques to extract market expectations from financial instruments. *Journal of Monetary Economics* 40: 383-429.
- Straja, S. R. 2001. Application of the Gram-Charlier approximation for option valuation. Montgomery Investment Technology Inc., Radnor, PA.
- Subbotin, M. T. H. 1923. On the law of frequency of error. *Mathematicheskii Sbornik* 31: 296-301.
- Tay, A. S. and K. F. Wallis. 2000. Density forecasting: a survey. *Journal of Forecasting* 19: 235-254.
- Taylor, S. J. 1986. *Modelling Financial Time Series*. New York: Wiley.
- Taylor, S. J. 1994. Modelling stochastic volatility. *Mathematical Finance* 4: 183-204.
- Theodossiou, P. 1998. Financial data and skewed generalized t distribution. *Management Science* 44: 1650-1661.
- Theodossiou, P. 2000. Distribution of financial asset prices, the skewed generalized error distribution, and the pricing of options. Working paper, School of Business, Rutgers University.
- van der Klaauw, B. and R. H. Koning. 2003. Testing normality assumption in the sample selection model with application to travel demand. *Journal of Business and Economic Statistics* 21: 31-42.
- Venkataraman, S. 1997. Value at risk for a mixture of normal distributions: the use of quasi-Bayesian estimation techniques. *Economic Perspectives*, March, 2-13, Federal Reserve Bank of Chicago.
- Vuong, Q. 1989. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* 57: 307-333.
- Wang, J. 2001. Generating daily changes in market variables using a multivariate mixture of normal distributions, 283-289. In *Proceedings of the 2001 Winter Simulation Conference*, B. A. Peters, J. S. Smith, D. J. Medeiros, and M. W. Rohrer (eds.), Arlington, Virginia.
- Wang, K.-L., C. Fawson, C. Barrett and J. B. McDonald. 2001. A flexible parametric GARCH model with an application to exchange rates. *Journal of Applied Econometrics* 16: 521-536.
- West, K. 1996. Asymptotic inference about prediction ability. *Econometrica* 64: 1067-1084.
- West, K. D. and M. W. McCracken. 1998. Regression-based tests of predictive ability. *International Economic Review* 39: 817-840.
- White, H. 1982. Maximum likelihood estimation of misspecified models. *Econometrica* 50: 1-25.
- White, H. 1994. *Estimation, Inference, and Specification Analysis*. Cambridge: Cambridge University Press.
- White, H. 2000. A reality check for data snooping. *Econometrica* 68: 1097-1028.
- Xia, Y. 2001. Learning about predictability: The effects of parameter uncertainty on dynamic asset allocation. *Journal of Finance* 56: 205-246.
- Zakoian, J. 1994. Threshold heteroskedastic models. *Journal of Economic Dynamics and Control* 18: 931-955.

APPENDIX A: Parametric Distributions

In the following, $f(z; \theta_d)$ with parameter(s) θ_d , if any, is the *standardized* density function of the standardized residuals $\{z_t\}$. We use $\phi(z)$ (with corresponding CDF $\Phi(z)$)

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)$$

and $g(\cdot | \nu)$ (with corresponding CDF $G(z | \nu)$)

$$g(z | \nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \frac{1}{\sqrt{\pi\nu}} \left(1 + \frac{z^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

to denote the distribution function of a standard normal and a Student t distribution with degrees of freedom ν , respectively. Also, $\gamma(a, x) = \int_0^x e^{-t} t^{a-1} dt$ is the lower incomplete gamma function, $\Gamma(a, x) = \int_x^{+\infty} e^{-t} t^{a-1} dt$ is the upper incomplete gamma function, $\Gamma(a) = \gamma(a, x) + \Gamma(a, x)$ is the complete gamma function, and $\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) dt$ is the error function.

Symmetric Distribution

1. Gaussian Normal

$$f(z) = \phi(z), \quad F(z) = \Phi(z), \quad F^{-1}(\alpha) = \Phi^{-1}(\alpha), \quad \mathbb{E}|z| = \sqrt{2/\pi}, \quad \gamma_2 = 0.$$

2. Student t

$$\begin{aligned} f(z; \nu) &= \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \frac{1}{\sqrt{\pi(\nu-2)}} \left(1 + \frac{z^2}{\nu-2}\right)^{-\frac{\nu+1}{2}}, \quad \nu > 2, \\ F(z; \nu) &= G\left(z \sqrt{\frac{\nu}{\nu-2}} \mid \nu\right), \\ F^{-1}(\alpha) &= \sqrt{\frac{\nu-2}{\nu}} G^{-1}(\alpha \mid \nu), \\ \mathbb{E}|z| &= \frac{4\Gamma(\frac{1+\nu}{2}) \sqrt{\nu-2}}{1 + \sqrt{\pi}\Gamma(\frac{\nu}{2})(\nu-1)}, \\ \gamma_2 &= \frac{6}{\nu-4}, \quad \nu > 4. \end{aligned}$$

3. GED

$$\begin{aligned} f(z; \nu) &= \frac{\nu \exp(-0.5|z/\lambda|^\nu)}{\lambda 2^{(1+1/\nu)} \Gamma(1/\nu)}, \quad \nu > 0, \quad \lambda = \sqrt{2^{(-2/\nu)} \Gamma(1/\nu) / \Gamma(3/\nu)}, \\ F(z; \nu) &= \begin{cases} \frac{\Gamma(1/\nu, \frac{1}{2}(-\frac{z}{\lambda})^\nu)}{2\Gamma(1/\nu)} & \text{if } z \leq 0 \\ \frac{1}{2} + \frac{\gamma(1/\nu, \frac{1}{2}(\frac{z}{\lambda})^\nu)}{2\Gamma(1/\nu)} & \text{if } z > 0 \end{cases}, \\ F^{-1}(\alpha) &= \begin{cases} -2\lambda\Gamma^{-1}(1/\nu, 2\alpha\Gamma(1/\nu)) & \text{if } \alpha \leq 0.5 \\ 2\lambda\gamma^{-1}(1/\nu, 2(\alpha - \frac{1}{2})\Gamma(1/\nu)) & \text{if } \alpha > 0.5 \end{cases}, \\ \mathbb{E}|z| &= \lambda 2^{1/\nu} \frac{\Gamma(2/\nu)}{\Gamma(1/\nu)}, \\ \gamma_2 &= \frac{\Gamma(1/\nu) \Gamma(5/\nu)}{[\Gamma(3/\nu)]^2}. \end{aligned}$$

4. Double Exponential / Laplace

$$\begin{aligned}
f(z) &= \frac{1}{\sqrt{2}} \exp(-\sqrt{2}|z|), \\
F(z) &= \begin{cases} \frac{1}{2} \exp(\sqrt{2}z) & \text{if } z \leq 0 \\ 1 - \frac{1}{2} \exp(-\sqrt{2}z) & \text{if } z > 0 \end{cases}, \\
F^{-1}(\alpha) &= \begin{cases} \frac{1}{\sqrt{2}} \ln(2\alpha) & \text{if } \alpha \leq \frac{1}{2} \\ -\frac{1}{\sqrt{2}} \ln[2(1-\alpha)] & \text{if } \alpha > \frac{1}{2} \end{cases}, \\
\mathbb{E}|z| &= \frac{\Gamma(2)}{\sqrt{2}}, \\
\gamma_2 &= 3.
\end{aligned}$$

5. Double Weibull

$$\begin{aligned}
f(z; a) &= \frac{a}{2\sigma} \left| \frac{z}{\sigma} \right|^{a-1} \exp\left(-\left| \frac{z}{\sigma} \right|^a\right), \quad a > 0, \quad \sigma = \sqrt{\frac{1}{\Gamma\left(\frac{a+2}{a}\right)}} \\
F(z; a) &= \begin{cases} \frac{1}{2} \exp\left(-\left(-\frac{z}{\sigma}\right)^a\right) & \text{if } z \leq 0 \\ 1 - \frac{1}{2} \exp\left(-\left(\frac{z}{\sigma}\right)^a\right) & \text{if } z > 0 \end{cases}, \\
F^{-1}(\alpha) &= \begin{cases} -\sigma [-\ln(2\alpha)]^{\frac{1}{a}} & \text{if } \alpha \leq \frac{1}{2} \\ \sigma (-\ln[2(1-\alpha)])^{\frac{1}{a}} & \text{if } \alpha > \frac{1}{2} \end{cases}, \\
\mathbb{E}|z| &= \frac{\sigma \Gamma\left(\frac{1}{a}\right)}{a}, \\
\gamma_2 &= \frac{\Gamma\left(\frac{4+a}{a}\right)}{[\Gamma\left(\frac{2+a}{a}\right)]^2} - 3.
\end{aligned}$$

Asymmetric Distributions

6. Skewed t

$$\begin{aligned}
f(z; \xi, \nu) &= \begin{cases} \frac{2s}{\xi + \xi^{-1}} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\sqrt{\pi(\nu-2)}} \left(1 + \frac{\xi^2(sz+m)^2}{\nu-2}\right)^{-\frac{\nu+1}{2}} & \text{if } z \leq -\frac{m}{s} \\ \frac{2s}{\xi + \xi^{-1}} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\sqrt{\pi(\nu-2)}} \left(1 + \frac{\xi^{-2}(sz+m)^2}{\nu-2}\right)^{-\frac{\nu+1}{2}} & \text{if } z > -\frac{m}{s} \end{cases}, \\
F(z; \xi, \nu) &= \begin{cases} \frac{2}{\xi^2+1} G\left[\xi(sz+m) \sqrt{\frac{\nu}{\nu-2}} \mid \nu\right] & \text{if } z \leq -\frac{m}{s} \\ 1 - \frac{2}{\xi^{-2}+1} G\left[-\xi^{-1}(sz+m) \sqrt{\frac{\nu}{\nu-2}} \mid \nu\right] & \text{if } z > -\frac{m}{s} \end{cases}, \\
F^{-1}(\alpha) &= \begin{cases} \frac{\frac{1}{\xi} G^{-1}\left[\frac{\alpha}{s}(1+\xi^2)\right] \nu - m}{s} & \text{if } \alpha \leq \frac{1}{1+\xi^2} \\ \frac{-\xi G^{-1}\left[\frac{1-\alpha}{s}(1+\xi^{-2})\right] \nu - m}{s} & \text{if } \alpha > \frac{1}{1+\xi^2} \end{cases}, \\
\mathbb{E}|z| &= \frac{4\xi^2 \Gamma\left(\frac{1+\nu}{2}\right) \sqrt{\nu-2}}{\xi + \frac{1}{\xi} \sqrt{\pi} \Gamma\left(\frac{\nu}{2}\right) (\nu-1)}, \\
\gamma_1 &= \frac{(\nu-2)^{3/2} (\xi^2-1) (\xi^4+1) \Gamma[(\nu-3)/2]}{\sqrt{\pi} \xi^3 s^3 \Gamma(\nu/2)} - \frac{m(m^2+3s)}{s^3}, \\
\gamma_2 &= \frac{3(\xi^5 + \xi^{-5})(\nu-2)}{(\xi + \xi^{-1})(\nu-4)s^4} + \frac{3m^2(m^2+2s)}{s^4} \\
&\quad - \frac{4m(\nu-2)^{3/2} (\xi^2-1) (\xi^4+1) \Gamma[(\nu-3)/2]}{\sqrt{\pi} \xi^3 s^4 \Gamma(\nu/2)},
\end{aligned}$$

where $\xi > 0$, $\nu > 2$, $I = \mathbf{1}(z \geq -m/s)$, $m = \Gamma[(\nu - 1)/2] \sqrt{(\nu - 2)/\pi} (\xi - 1/\xi) / \Gamma(\nu/2)$, $s = \sqrt{(\xi^2 + 1/\xi^2 - 1) - m^2}$, γ_1 exists if $\nu > 3$ and γ_2 exists if $\nu > 4$.

7. Hansen *t*

$$\begin{aligned}
f(z; \eta, \lambda) &= \begin{cases} bc \left(1 + \frac{1}{\eta-2} \left(\frac{bz+a}{1-\lambda}\right)^2\right)^{-(\eta+1)/2} & \text{if } z \leq -\frac{a}{b} \\ bc \left(1 + \frac{1}{\eta-2} \left(\frac{bz+a}{1+\lambda}\right)^2\right)^{-(\eta+1)/2} & \text{if } z > -\frac{a}{b} \end{cases}, \\
F(z; \eta, \lambda) &= \begin{cases} (1-\lambda) G\left(\frac{bz+a}{1-\lambda} \frac{\sqrt{\eta}}{\sqrt{\eta-2}} \mid \eta\right) & \text{if } z \leq -\frac{a}{b} \\ (1+\lambda) \left[G\left(\frac{bz+a}{1+\lambda} \frac{\sqrt{\eta}}{\sqrt{\eta-2}} \mid \eta\right)\right] - \lambda & \text{if } z > -\frac{a}{b} \end{cases}, \\
F^{-1}(\alpha) &= \begin{cases} \frac{1}{b} \left[(1-\lambda) \sqrt{\frac{\eta-2}{\eta}} G^{-1}\left(\frac{\alpha}{1-\lambda} \mid \eta\right) - a\right] & \text{if } \alpha \leq \frac{1-\lambda}{2} \\ \frac{1}{b} \left[(1+\lambda) \sqrt{\frac{\eta-2}{\eta}} G^{-1}\left(\frac{\alpha+\lambda}{1+\lambda} \mid \eta\right) - a\right] & \text{if } \alpha > \frac{1-\lambda}{2} \end{cases}, \\
\mathbb{E}|z| &= \begin{cases} \left[2G\left(\frac{a}{1-\lambda} \sqrt{\frac{\eta}{\eta-2}} \mid \eta\right)\right] \frac{a(1-\lambda)}{b} + \frac{2c[a^2+(1-\lambda)^2(\eta-2)]^{-\frac{\eta-1}{2}} [(1-\lambda)^2(\eta-2)]^{\frac{\eta+1}{2}}}{b(\eta-1)} & \text{if } \lambda \leq 0 \\ \left[2G\left(\frac{a}{1+\lambda} \sqrt{\frac{\eta}{\eta-2}} \mid \eta\right) - 1\right] \frac{a(1+\lambda)}{b} + \frac{2c[a^2+(1+\lambda)^2(\eta-2)]^{-\frac{\eta-1}{2}} [(1+\lambda)^2(\eta-2)]^{\frac{\eta+1}{2}}}{b(\eta-1)} & \text{if } \lambda > 0 \end{cases}, \\
\gamma_1 &= \frac{2a^3 - 3a(1 + 3\lambda^2)}{b^3} + \frac{16c\lambda(\eta-2)^2(1+\lambda^2)}{(\eta-1)(\eta-3)b^3}, \quad \eta > 3, \\
\gamma_2 &= \frac{6a^2(1 + 3\lambda^2) - 3a^4}{b^4} + \frac{3(\eta-2)(1 + 10\lambda^2 + 5\lambda^4)}{(\eta-4)b^4} - \frac{64ac\lambda(\eta-2)^2(1+\lambda^2)}{(\eta-1)(\eta-3)b^4} - 3,
\end{aligned}$$

where $\eta > 2$, $1 > \lambda > -1$, $a = 4\lambda c(\eta - 2) / (\eta - 1)$, $b^2 = 1 + 3\lambda^2 - a^2$, $c = \Gamma((\eta + 1)/2) / [\sqrt{\pi(\eta - 2)} \Gamma(\eta/2)]$, and γ_2 exists if $\eta > 4$.

8. Skewed GED

$$\begin{aligned}
f(z; k, \lambda) &= \begin{cases} c \exp\left(-w_1 |z - \delta|^k\right) & \text{if } z \leq \delta \\ c \exp\left(-w_2 |z - \delta|^k\right) & \text{if } z > \delta \end{cases}, \\
F(z; k, \lambda) &= \begin{cases} \frac{(1-\lambda)\Gamma(1/k, w_1(\delta-z)^k)}{2\Gamma(1/k)} & \text{if } z \leq \delta \\ \frac{1-\lambda}{2} + \frac{(1+\lambda)\Gamma(1/k, w_2(z-\delta)^k)}{2\Gamma(1/k)} & \text{if } z > \delta \end{cases}, \\
F^{-1}(\alpha) &= \begin{cases} \delta - \frac{1-\lambda}{\phi} \left[\Gamma^{-1}\left(\frac{1}{k}, \frac{2\alpha\Gamma(1/k)}{1-\lambda}\right)\right]^{1/k} & \text{if } \alpha \leq \frac{1-\lambda}{2} \\ \delta + \frac{1+\lambda}{\phi} \left[\Gamma^{-1}\left(\frac{1}{k}, \frac{2(\alpha-\frac{1-\lambda}{2})\Gamma(1/k)}{1+\lambda}\right)\right]^{1/k} & \text{if } \alpha > \frac{1-\lambda}{2} \end{cases}, \\
\mathbb{E}|z| &= \begin{cases} c(-A_1 - B_1 + C_1) & \text{if } \delta \leq 0 \\ c(-A_2 + B_2 + C_2) & \text{if } \delta > 0 \end{cases}, \\
\gamma_1 &= \frac{4\lambda(1 + \lambda^2)\Gamma(4/k)}{\phi^3\Gamma(1/k)} + 3\delta + \delta^3, \\
\gamma_2 &= \frac{(1 + 10\lambda^2 + 5\lambda^4)\Gamma(5/k)}{\phi^4\Gamma(1/k)} + \frac{16\lambda\delta(1 + \lambda^2)\Gamma(4/k)}{\phi^3\Gamma(1/k)} + 6\delta^2 + 3\delta^4 - 3,
\end{aligned}$$

where $k > 0$, $1 > \lambda > -1$,

$$\begin{aligned}
c &= \frac{k\phi}{2\Gamma(1/k)}, \quad \phi = s \sqrt{\frac{\Gamma(3/k)}{\Gamma(1/k)}}, \quad \delta = -2\lambda A s^{-1} \\
s &= \sqrt{1 + 3\lambda^2 - 4A^2\lambda^2}, \quad A = \frac{\Gamma(2/k)}{\sqrt{\Gamma(1/k)\Gamma(3/k)}},
\end{aligned}$$

$$\begin{aligned}
w_1 &= \frac{\phi^k}{(1-\lambda)^k}, \quad w_2 = \frac{\phi^k}{(1+\lambda)^k}, \\
A_1 &= \frac{w_1^{-2/k} \left[-\Gamma\left(\frac{2}{k}\right) + k\delta w_1^{1/k} \Gamma\left(1 + \frac{1}{k}\right) \right]}{k}, \\
B_1 &= \frac{w_2^{-2/k} \left[\Gamma\left(\frac{2}{k}\right) - \Gamma\left(\frac{2}{k}, (-\delta)^k w_2\right) + \delta w_2^{1/k} \left(k\Gamma\left(1 + \frac{1}{k}\right) - \Gamma\left(\frac{1}{k}, (-\delta)^k w_2\right) \right) \right]}{k}, \\
C_1 &= \frac{w_2^{-2/k} \left[\Gamma\left(\frac{2}{k}\right) + k\delta w_2^{1/k} \Gamma\left(1 + \frac{1}{k}\right) \right]}{k} - B_1, \\
A_2 &= \frac{w_1^{-2/k} \left[\Gamma\left(\frac{2}{k}\right) - k\delta w_1^{1/k} \Gamma\left(1 + \frac{1}{k}\right) \right]}{k} - B_2, \\
B_2 &= \frac{w_1^{-2/k} \left[-\Gamma\left(\frac{2}{k}\right) + \Gamma\left(\frac{2}{k}, \delta^k w_1\right) + \delta w_1^{1/k} \left(k\Gamma\left(1 + \frac{1}{k}\right) - \Gamma\left(\frac{1}{k}, \delta^k w_1\right) \right) \right]}{k}, \\
C_2 &= \frac{w_2^{-2/k} \left[\Gamma\left(\frac{2}{k}\right) + k\delta w_2^{1/k} \Gamma\left(1 + \frac{1}{k}\right) \right]}{k}.
\end{aligned}$$

9. Inverse Hyperbolic Sine

$$\begin{aligned}
f(z; \lambda, \delta) &= \frac{s}{\sqrt{2\pi} \left[(zs + \mu)^2 + 1 \right] \delta^2} \exp\left(-\frac{[\sinh^{-1}(zs + \mu) - \lambda]^2}{2\delta^2}\right), \\
F(z; \lambda, \delta) &= \Phi\left(\frac{\sinh^{-1}(zs + \mu) - \lambda}{\delta}\right), \\
F^{-1}(\alpha) &= \frac{\sinh[\lambda + \delta\Phi^{-1}(\alpha)] - \mu}{s}, \\
\mathbb{E}|z| &= \frac{1}{2} \exp\left(\frac{\delta^2}{2}\right) \cosh(\lambda) \left[\operatorname{erf}\left(\frac{\delta^2 - \lambda}{\sqrt{2}\delta}\right) + \operatorname{erf}\left(\frac{\delta^2 + \lambda}{\sqrt{2}\delta}\right) \right], \\
\gamma_1 &= \frac{\omega^{1/2} (\omega - 1)^2 [\omega(\omega + 2) \sinh(3\lambda) + 3 \sinh(\lambda)]}{4s^3}, \\
\gamma_2 &= \frac{(\omega - 1)^2 [\omega^2 (\omega^4 + 2\omega^3 + 3\omega^2 - 3) \cosh(4\lambda) + 4\omega^2 (\omega + 2) \cosh(2\lambda) + 3(2\omega + 1)]}{8s^4} \\
&\quad - 3,
\end{aligned}$$

where $\delta > 0$, $\mu = \omega^{1/2} \sinh(\lambda)$, $s = \sqrt{(\omega - 1) [\omega \cosh(2\lambda) + 1]}/2$, $\omega = \exp(\delta^2)$. Note that

$$\sinh(x) = \frac{e^x - e^{-x}}{2}, \quad \sinh^{-1}(x) = \ln\left(x + \sqrt{x^2 + 1}\right), \quad \cosh(x) = \frac{e^x + e^{-x}}{2}.$$

10. Mixture of Normals

$$\begin{aligned}
f(z; p_1, \mu_1, \sigma_1) &= p_1 \frac{1}{\sqrt{2\pi}\sigma_1^2} \exp\left[-\frac{(z - \mu_1)^2}{2\sigma_1^2}\right] + p_2 \frac{1}{\sqrt{2\pi}\sigma_2^2} \exp\left[-\frac{(z - \mu_2)^2}{2\sigma_2^2}\right], \\
F(z; p_1, \mu_1, \sigma_1) &= p_1 \Phi\left(\frac{z - \mu_1}{\sigma_1}\right) + p_2 \Phi\left(\frac{z - \mu_2}{\sigma_2}\right), \\
\mathbb{E}|z| &= \sqrt{2}p_1\sigma_1 \exp\left(-\frac{\mu_1^2}{2\sigma_1^2}\right) / \sqrt{\pi} + \sqrt{2}p_2\sigma_2 \exp\left(-\frac{\mu_2^2}{2\sigma_2^2}\right) / \sqrt{\pi} \\
&\quad - 2p_1\mu_1 \Phi\left(\frac{-\mu_1}{\sigma_1}\right) - 2p_2\mu_2 \Phi\left(\frac{-\mu_2}{\sigma_2}\right),
\end{aligned}$$

$$\begin{aligned}\gamma_1 &= p_1 (3\mu_1\sigma_1 + \mu_1^3) + p_2 (3\mu_2\sigma_2 + \mu_2^3), \\ \gamma_2 &= p_1 (6\mu_1^2\sigma_1^2 + \mu_1^4 + 3\sigma_1^4) + p_2 (6\mu_2^2\sigma_2^2 + \mu_2^4 + 3\sigma_2^4) - 3,\end{aligned}$$

where p_2 , μ_2 , and σ_2 are determined through the constraints $p_1 + p_2 = 1$, $p_1 \geq 0$, $p_2 \geq 0$, $\sum_{i=1}^2 p_i \mu_i = 0$, and $\sum_{i=1}^2 p_i (\mu_i^2 + \sigma_i^2) = 1$. As for the inverse CDF function $F^{-1}(\alpha)$, there is no closed-form solution, but numerical inversion is always possible since $F(z; p_1, \mu_1, \sigma_1)$ is monotonic in z .

11. Double Gamma

$$\begin{aligned}f(z; \alpha_1, \alpha_2, p) &= \begin{cases} \frac{(1-p)\lambda_1^{\alpha_1}}{\Gamma(\alpha_1)} |z|^{(\alpha_1-1)} \exp(-\lambda_1 |z|) & \text{if } z \leq 0 \\ \frac{p\lambda_2^{\alpha_2}}{\Gamma(\alpha_2)} z^{(\alpha_2-1)} \exp(-\lambda_2 z) & \text{if } z > 0 \end{cases}, \\ F(z; \alpha_1, \alpha_2, p) &= \begin{cases} \frac{(1-p)}{\Gamma(\alpha_1)} \Gamma(\alpha_1, -\lambda_1 z) & \text{if } z \leq 0 \\ 1 - p + \frac{p}{\Gamma(\alpha_2)} \Gamma(\alpha_2, \lambda_2 z) & \text{if } z > 0 \end{cases}, \\ F^{-1}(\alpha) &= \begin{cases} -\Gamma^{-1}\left(\alpha_1, \frac{\alpha\Gamma(\alpha_1)}{1-p}\right) / \lambda_1 & \text{if } \alpha \leq 1 - p \\ \Gamma^{-1}\left(\alpha_2, \frac{(\alpha+p-1)\Gamma(\alpha_2)}{p}\right) / \lambda_2 & \text{if } \alpha > 1 - p \end{cases}, \\ \mathbb{E}|z| &= \frac{p\Gamma(1+\alpha_2)}{\lambda_2\Gamma(\alpha_2)} + \frac{(1-p)\Gamma(1+\alpha_1)}{\lambda_1\Gamma(\alpha_1)}, \\ \gamma_1 &= \frac{p\Gamma(3+\alpha_2)}{\lambda_2^3\Gamma(\alpha_2)} - \frac{(1-p)\Gamma(3+\alpha_1)}{\lambda_1^3\Gamma(\alpha_1)}, \\ \gamma_2 &= \frac{p\Gamma(4+\alpha_2)}{\lambda_2^4\Gamma(\alpha_2)} - \frac{(1-p)\Gamma(4+\alpha_1)}{\lambda_1^4\Gamma(\alpha_1)} - 3,\end{aligned}$$

where $1 > p > 0$, $\alpha_1 > 0$, $\alpha_2 > 0$,

$$\lambda_1 = \frac{p\alpha_1\lambda_2}{(1-p)\alpha_2}, \quad \lambda_2 = \sqrt{\alpha_2(1-p) \left[\frac{(\alpha_1+1)\alpha_2(1-p)}{p\alpha_1} + (\alpha_2+1) \right]}.$$

12. Gram-Charlier / Edgeworth-Sargan Density

$$\begin{aligned}f_4(z; \gamma_1, \gamma_2) &= \left[1 + \frac{\gamma_1}{6} H_3(z) + \frac{\gamma_2}{24} H_4(z) \right] \phi(z), \\ F(z; \gamma_1, \gamma_2) &= \Phi(z) - \phi(z) \left[\frac{\gamma_1(z^2-1)}{6} + \frac{\gamma_2(z^3-3z)}{24} \right], \\ \mathbb{E}|z| &= \sqrt{\frac{2}{\pi}} - \frac{\gamma_2}{12} \sqrt{\frac{1}{2\pi}},\end{aligned}$$

where $4 \geq \gamma_2 \geq 0$, $\gamma_1^U \geq \gamma_1 \geq \gamma_1^L$, $H_3(z) = z^3 - 3z$, $H_4(z) = z^4 - 6z^2 + 3$. As for the inverse CDF function $F^{-1}(\alpha)$, there is no closed-form solution, and numerical inversion can in general give multiple solutions since $F(z; p_1, \mu_1, \sigma_1)$ is not guaranteed to be monotonic in z . However, as long as the positiveness constraint on $f_4(z; \gamma_1, \gamma_2)$ is imposed and the grids over the domain of admissible γ_1 and γ_2 are fine enough, $F(z; p_1, \mu_1, \sigma_1)$ should be almost monotonic in z and hence numerical inversion is possible.

APPENDIX B: Volatility Specification

In the following, in some cases we use $\alpha(L) = \sum_{i=1}^p \alpha_i L^i$, $\beta(L) = \sum_{j=1}^q \beta_j L^j$, where L is the backshift operator. Similarly, $\phi(L)$ is a polynomial in L .

1. EWMA

$$\sigma_t^2 = \lambda \sigma_{t-1}^2 + (1 - \lambda)(y_{t-1} - \hat{\mu}_t)^2,$$

where $\lambda = 0.94$ and $\hat{\mu}_t = \frac{1}{t-1} \sum_{j=1}^{t-1} y_{t-j}$.

2. GARCH(p, q)

$$\begin{aligned} \sigma_t^2 &= \omega + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2, \\ \omega &> 0, \alpha_i \geq 0, \beta_j \geq 0, \sum_{i=1}^q \beta_i + \sum_{j=1}^q \alpha_j < 1. \end{aligned}$$

3. GJR(p, q)

$$\begin{aligned} \sigma_t^2 &= \omega + \sum_{i=1}^q (\alpha_i \varepsilon_{t-i}^2 + \varphi_i S_{t-i} \varepsilon_{t-i}^2) + \sum_{j=1}^p \beta_j \sigma_{t-j}^2, \\ S_{t-i} &= \begin{cases} 1 & \text{if } \varepsilon_{t-i} < 0 \\ 0 & \text{if } \varepsilon_{t-i} \geq 0 \end{cases}, \\ \omega &> 0, \beta_j > 0, \alpha_i + \varphi_i > 0. \end{aligned}$$

4. APARCH(p, q)

$$\begin{aligned} \sigma_t^\delta &= \omega + \sum_{i=1}^q \alpha_i (|\varepsilon_{t-i}| - \varphi_i \varepsilon_{t-i})^\delta + \sum_{j=1}^p \beta_j \sigma_{t-j}^\delta, \\ \omega &> 0, \beta_i > 0, \alpha_j > 0, 1 > \varphi_i > -1, \delta > 0. \end{aligned}$$

5. EGARCH(p, q)

$$\ln \sigma_t^2 = \omega + [1 - \beta(L)]^{-1} [1 + \alpha(L)] [\varphi_1 z_{t-1} + \varphi_2 (|z_{t-1}| - E|z_{t-1}|)].$$

6. STGARCH(p, q)

$$\begin{aligned} \sigma_t^2 &= \omega + \sum_{i=1}^p \beta_i \sigma_{t-i}^2 + \sum_{j=1}^q [\alpha_{1j} \varepsilon_{t-j}^2 + \alpha_{2j} \varepsilon_{t-j}^2 F(\varepsilon_{t-d})], \\ F(\varepsilon_{t-d}) &= \frac{1}{1 + \exp(\varphi \varepsilon_{t-d})} - \frac{1}{2}, \\ \omega &> 0, \beta_i > 0, \alpha_{1j} > \frac{1}{2} \alpha_{2j} > 0, \varphi > 0. \end{aligned}$$

7. FIGARCH(p, d, q) / HYGARCH(p, d, q)

Bollerslev and Mikkelsen (1996) introduced a FIGARCH model as follows (by setting $\kappa = 1$, where κ is an additional parameter for its generalization HYGARCH)

$$\begin{aligned} \sigma_t^2 &= \omega [1 - \beta(L)]^{-1} + \left(1 - [1 - \beta(L)]^{-1} \phi(L) \left\{ 1 + \kappa \left[(1 - L)^d - 1 \right] \right\} \right) \varepsilon_t^2 \\ &= \omega [1 - \beta(L)]^{-1} + [1 - \beta(L)]^{-1} \lambda(L) \varepsilon_t^2. \end{aligned}$$

If we rewrite $\phi(L) = 1 - \phi^*(L)$ (note that the definition of $\phi(L) = (1 - \alpha(L) - \beta(L))(1 - L)^{-1}$ implies that it has a constant term), we have

$$\begin{aligned}\lambda(L) &= 1 - \beta(L) - (1 - \phi^*(L)) \left\{ 1 + \kappa \left[(1 - L)^d - 1 \right] \right\} \\ &= 1 - \beta(L) - 1 - \kappa \left[(1 - L)^d - 1 \right] + \phi^*(L) + \phi^*(L) \kappa \left[(1 - L)^d - 1 \right] \\ &= -\beta(L) + \phi^*(L) - \kappa \left[(1 - L)^d - 1 \right] + \phi^*(L) \kappa \left[(1 - L)^d - 1 \right],\end{aligned}$$

where

$$(1 - L)^d - 1 = \sum_{k=1}^{\infty} \delta_k L^k, \quad \delta_k = \begin{cases} -d, & k = 1 \\ \frac{(k-1-d)}{k} \delta_{k-1}, & k \geq 2 \end{cases}.$$

Hence we have

$$\begin{aligned}\lambda(L) &= -\beta(L) + \phi^*(L) - \kappa \sum_{k=1}^{\infty} \delta_k L^k + \kappa \phi^*(L) \sum_{k=1}^{\infty} \delta_k L^k \\ &= -\beta(L) + \phi^*(L) - \kappa \sum_{k=1}^{\infty} \delta_k L^k + \kappa \sum_{k=1}^{\infty} \delta_k^* L^k,\end{aligned}$$

where

$$\delta_k^* = \begin{cases} 0 & k = 1 \\ \sum_{j=1}^{k-1} \phi_j^* \delta_{k-j} & k = 2, \dots, q+1 \\ \sum_{j=1}^q \phi_j^* \delta_{k-j} & k > q+1 \end{cases}.$$

Therefore, we can define a HYGARCH(p, d, q) model as

$$\begin{aligned}\sigma_t^2 &= \omega + \beta(L) \sigma_t^2 + \left[\phi^*(L) - \beta(L) + \sum_{k=1}^{\infty} \pi_k L^k \right] \varepsilon_t^2, \\ \pi_k &= \kappa (\delta_k^* - \delta_k), \quad \beta(L) = \sum_{i=1}^p \beta_i L^i, \quad \phi^*(L) = \sum_{j=1}^q \phi_j^* L^j.\end{aligned}$$

The parameters are $\theta_v = (\omega, \beta, \phi^{*'}, \kappa, d)'$. For a HYGARCH(1, d , 1) model, the following constraints are sufficient for positiveness of σ_t^2 : $\phi^* - \beta + \kappa d \geq 0$, $\phi^* \leq 0$, $\beta \geq 0$, $\omega \geq 0$.

8. FIAPARCH(p, d, q) / HYAPARCH(p, d, q)

$$\sigma_t^\delta = \omega + \left(1 - [1 - \beta(L)]^{-1} \phi(L) \left\{ 1 + \kappa \left[(1 - L)^d - 1 \right] \right\} \right) (|\varepsilon_t| - \varphi \varepsilon_t)^\delta,$$

which can be rewritten as

$$\sigma_t^\delta = \omega^* + \beta(L) \sigma_t^\delta + \left[\phi^*(L) - \beta(L) + \sum_{k=1}^{\infty} \pi_k L^k \right] (|\varepsilon_t| - \varphi \varepsilon_t)^\delta,$$

where $\phi^*(L)$, $\beta(L)$, and π_k are defined in FIGARCH/HYGARCH. The parameters are $\theta_v = (\omega^*, \beta, \phi^{*'}, \varphi, \kappa, d, \delta)'$. For a HYAPARCH(1, d , 1) model, the following constraints are sufficient for positiveness of σ_t^2 : $1 > \varphi > -1$, $\delta > 0$, $\phi^* - \beta + \kappa d \geq 0$, $\phi^* \leq 0$, $\beta \geq 0$, $\omega^* \geq 0$.

9. CGARCH(1,1)

$$\begin{aligned}\sigma_t^2 &= \Omega_t + \beta (\sigma_{t-1}^2 - \Omega_{t-1}) + \alpha (\varepsilon_{t-1}^2 - \Omega_{t-1}) \\ \Omega_t &= \omega + \rho \Omega_{t-1} + \phi (\varepsilon_{t-1}^2 - \Omega_{t-1}), \\ 1 &> \rho > \beta + \alpha > 0, \quad \beta > \phi > 0, \quad \alpha > 0, \quad \omega > 0,\end{aligned}$$

where Ω_t is the long-run volatility.

Table 1: Reality Check Results for S&P500

Panel A: Whole Distribution

	EWMA	GARCH	GJR	APARCH	EGARCH	STGARCH	HYGARCH	HYAPARCH	CGARCH
Historical	0.0436	0.0183	0.0207	0.0213	0.0187	0.0261	0.0116	0.0148	0.0195
	0.001	0.254	0.171	0.150	0.237	0.203	0.721	0.496	0.211
	0.001	0.216	0.153	0.134	0.205	0.185	0.621	0.436	0.178
Normal	0.0256	0.0058	0.0258	0.0023	0.0084	0.0096	0.0291	0.0220	0.0252
	0.165	0.942	0.174	0.991	0.862	0.811	0.154	0.227	0.175
	0.136	0.796	0.151	0.894	0.747	0.694	0.128	0.195	0.148
Student t	0.0039	0.0107	0.0131	0.0108	0.0122	0.0082	0.0111	0.0105	0.0091
	0.983	0.777	0.640	0.763	0.698	0.870	0.746	0.779	0.840
	0.805	0.646	0.553	0.641	0.595	0.739	0.621	0.654	0.695
GED	0.0132	0.0147	0.0164	0.0143	0.0158	0.0182	0.0144	0.0140	0.0133
	0.621	0.478	0.363	0.556	0.428	0.247	0.512	0.577	0.602
	0.509	0.383	0.294	0.465	0.348	0.189	0.412	0.480	0.493
Laplace	0.0408	0.0363	0.0446	0.0409	0.0424	0.0426	0.0333	0.0405	0.0359
	0.000	0.004	0.000	0.001	0.000	0.000	0.014	0.001	0.004
	0.000	0.004	0.000	0.001	0.000	0.000	0.013	0.001	0.004
Double Weibull	0.0132	0.0061	0.0062	0.0065	0.0166	0.0191	0.0143	0.0035	0.0135
	0.617	0.941	0.941	0.927	0.378	0.222	0.538	0.981	0.585
	0.515	0.795	0.796	0.786	0.313	0.178	0.447	0.788	0.483
Skewed t	0.0488	0.0628	0.0384	0.0429	0.0178	0.0450	0.0630	0.0437	0.0715
	0.000	0.000	0.007	0.000	0.328	0.000	0.000	0.000	0.000
	0.000	0.000	0.006	0.000	0.301	0.000	0.000	0.000	0.000
Hansen t	0.0867	0.0602	0.1421	0.1461	0.1367	0.0600	0.1464	0.1433	0.0728
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Skewed GED	0.0108	0.0120	0.0144	0.0127	0.0142	0.0155	0.0119	0.0126	0.0109
	0.789	0.707	0.499	0.663	0.545	0.428	0.713	0.678	0.794
	0.656	0.596	0.415	0.557	0.459	0.350	0.604	0.558	0.663
IHS	0.0072	0.0093	0.0113	0.0102	0.0114	0.0128	0.0091	0.0099	0.0080
	0.921	0.839	0.753	0.797	0.751	0.657	0.849	0.809	0.896
	0.770	0.695	0.643	0.673	0.641	0.561	0.703	0.681	0.738
Mixture	0.0055	0.0138	0.0057	0.0107	0.0124	0.0151	0.0105	0.0104	0.0136
	0.965	0.528	0.955	0.783	0.633	0.488	0.756	0.805	0.524
	0.795	0.423	0.828	0.629	0.521	0.412	0.628	0.658	0.427
Double Gamma	0.0174	0.0138	0.0003	0.0120	0.0175	0.0054	0.0119	0.0105	0.0029
	0.380	0.586	1.000	0.698	0.367	0.944	0.703	0.777	0.990
	0.289	0.492	0.977	0.573	0.310	0.771	0.594	0.647	0.842
Sargan	0.0042	0.0286	0.0040	0.0028	0.0267	0.0309	0.0266	0.0219	0.0209
	0.987	0.134	0.988	0.995	0.156	0.111	0.169	0.231	0.251
	0.811	0.113	0.864	0.869	0.137	0.097	0.149	0.200	0.220

- Notes:
1. For each combination of distribution and volatility, the first number is the sample averaged loss, the second and the third numbers are the reality check p -values of White's (2000) test and Hansen's (2001) test, respectively. Each model from a combination of distribution and volatility is regarded as a benchmark model and is compared with the remaining 116 models from different combinations.
 2. We retrieve the S&P500 returns series from *finance.yahoo.com*. The sample observations are from January 3, 1990 to June 30, 2003 ($T = 3303$), the in-sample observations are from January 3, 1990 to September 24, 1996 ($R = 1703$), and the out-of-sample observations are from September 25, 1996 to June 30, 2003 ($n = 1700$).
 3. We use an MA(1) model, without a constant, for the mean, under rolling scheme. The loss is based on an AR(3)-SNP(8) model for the transformed PITs $\{x_t\}$.

Table 1: Reality Check Results for S&P500

Panel B: 10% Tail

	EWMA	GARCH	GJR	APARCH	EGARCH	STGARCH	HYGARCH	HYAPARCH	CGARCH
Historical	0.0397	0.0467	0.0471	0.0452	0.0485	0.1246	0.0475	0.0446	0.0482
	0.109	0.044	0.039	0.051	0.034	0.000	0.041	0.056	0.040
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Normal	0.0425	0.0425	0.0456	0.0315	0.0451	0.0445	0.0927	0.0436	0.0397
	0.092	0.079	0.059	0.309	0.069	0.066	0.001	0.078	0.118
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Student <i>t</i>	0.0535	0.0554	0.0549	0.0533	0.0539	0.0580	0.0553	0.0533	0.0540
	0.016	0.013	0.012	0.020	0.018	0.010	0.014	0.019	0.014
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
GED	0.0516	0.0532	0.0514	0.0504	0.0524	0.0560	0.0531	0.0505	0.0523
	0.020	0.017	0.028	0.029	0.021	0.012	0.018	0.029	0.017
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Laplace	0.0610	0.0580	0.0567	0.0544	0.0562	0.0602	0.0548	0.0542	0.0566
	0.007	0.008	0.011	0.015	0.012	0.008	0.013	0.015	0.011
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Double Weibull	0.0540	0.0535	0.0531	0.0534	0.0542	0.0574	0.0554	0.0519	0.0536
	0.015	0.017	0.016	0.019	0.015	0.011	0.012	0.021	0.015
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Skewed <i>t</i>	0.0259	0.0229	0.0259	0.0261	0.0269	0.0257	0.0229	0.0266	0.0209
	0.448	0.573	0.451	0.439	0.414	0.469	0.572	0.426	0.666
	0.000	0.006	0.000	0.000	0.000	0.000	0.006	0.000	0.009
Hansen <i>t</i>	0.0851	0.0278	0.0129	0.0120	0.0122	0.0282	0.0142	0.0123	0.0234
	0.000	0.399	0.992	1.000	0.993	0.381	0.956	0.995	0.559
	0.000	0.001	0.522	0.959	0.745	0.001	0.207	0.707	0.010
Skewed GED	0.0502	0.0492	0.0502	0.0492	0.0500	0.0528	0.0499	0.0490	0.0503
	0.023	0.028	0.030	0.035	0.030	0.016	0.025	0.036	0.021
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
IHS	0.0529	0.0539	0.0528	0.0513	0.0536	0.0566	0.0539	0.0510	0.0523
	0.018	0.014	0.016	0.023	0.016	0.011	0.015	0.025	0.016
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Mixture	0.0509	0.0575	0.0542	0.0529	0.0530	0.0592	0.0553	0.0532	0.0554
	0.017	0.010	0.013	0.020	0.019	0.009	0.013	0.019	0.011
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Double Gamma	0.0558	0.0540	0.0531	0.0527	0.0543	0.0576	0.0538	0.0506	0.0534
	0.011	0.012	0.018	0.017	0.013	0.009	0.011	0.023	0.013
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Sargan	0.1660	0.2037	0.1917	0.0552	0.2350	0.1848	0.0257	0.0434	0.1006
	0.000	0.000	0.000	0.053	0.000	0.000	0.469	0.078	0.000
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

- Notes:
1. For each combination of distribution and volatility, the first number is the sample averaged loss, the second and the third numbers are the reality check p -values of White's (2000) test and Hansen's (2001) test, respectively. Each model from a combination of distribution and volatility is regarded as a benchmark model and is compared with the remaining 116 models from different combinations.
 2. We retrieve the S&P500 returns series from *finance.yahoo.com*. The sample observations are from January 3, 1990 to June 30, 2003 ($T = 3303$), the in-sample observations are from January 3, 1990 to September 24, 1996 ($R = 1703$), and the out-of-sample observations are from September 25, 1996 to June 30, 2003 ($n = 1700$).
 3. We use an MA(1) model, without a constant, for the mean, under rolling scheme. The loss is based on an AR(3)-SNP(8) model for the transformed PITs $\{x_t\}$.

Table 1: Reality Check Results for S&P500

Panel C: 5% Tail

	EWMA	GARCH	GJR	APARCH	EGARCH	STGARCH	HYGARCH	HYAPARCH	CGARCH
Historical	0.0112	0.0171	0.0141	0.0163	0.0158	0.1441	0.0155	0.0147	0.0169
	0.978	0.912	0.949	0.930	0.927	0.010	0.933	0.944	0.919
	0.065	0.006	0.015	0.006	0.010	0.000	0.019	0.014	0.007
Normal	0.0429	0.0414	0.0344	0.0311	0.0407	0.0401	0.0064	0.0303	0.0473
	0.526	0.543	0.670	0.722	0.573	0.566	1.000	0.740	0.465
	0.006	0.005	0.006	0.006	0.006	0.004	0.345	0.007	0.003
Student <i>t</i>	0.0260	0.0255	0.0219	0.0227	0.0247	0.9193	0.0254	0.0228	0.0939
	0.806	0.807	0.854	0.829	0.819	0.000	0.808	0.848	0.090
	0.000	0.000	0.007	0.003	0.001	0.000	0.000	0.003	0.000
GED	0.9085	0.0270	0.0311	0.9079	0.0227	0.0267	0.0333	0.0264	0.0223
	0.000	0.799	0.720	0.000	0.856	0.807	0.686	0.806	0.857
	0.000	0.002	0.001	0.000	0.001	0.002	0.002	0.002	0.001
Laplace	0.0448	0.0193	0.0183	0.0204	0.0206	0.5739	0.0739	0.0168	0.0177
	0.513	0.899	0.906	0.882	0.876	0.000	0.189	0.923	0.912
	0.000	0.002	0.006	0.002	0.002	0.000	0.000	0.011	0.004
Double Weibull	0.0256	0.0228	0.0237	0.8970	0.9065	0.6465	0.0261	0.0231	0.0216
	0.817	0.852	0.844	0.000	0.000	0.000	0.817	0.854	0.868
	0.001	0.000	0.002	0.000	0.000	0.000	0.001	0.001	0.001
Skewed <i>t</i>	0.0164	0.0149	0.0333	0.0127	0.0126	0.0103	0.0093	0.0134	0.0089
	0.931	0.947	0.694	0.963	0.962	0.990	0.991	0.947	0.992
	0.045	0.048	0.002	0.029	0.043	0.041	0.106	0.051	0.107
Hansen <i>t</i>	0.0349	0.8413	0.0055	0.6470	0.6395	0.0127	0.0038	0.6484	0.0087
	0.665	0.000	1.000	0.000	0.000	0.970	1.000	0.000	0.992
	0.000	0.000	0.504	0.000	0.000	0.090	0.702	0.000	0.233
Skewed GED	0.0240	0.0180	0.0212	0.9226	0.0204	0.0231	0.0217	0.0214	0.0202
	0.841	0.899	0.876	0.000	0.886	0.849	0.867	0.868	0.888
	0.001	0.019	0.002	0.000	0.002	0.001	0.001	0.009	0.003
IHS	0.0203	0.0207	0.0201	0.0219	0.0198	0.0238	0.0209	0.0204	0.0215
	0.885	0.875	0.885	0.861	0.892	0.834	0.873	0.884	0.863
	0.003	0.003	0.001	0.000	0.002	0.001	0.003	0.001	0.000
Mixture	0.0351	0.0284	0.0256	0.0249	0.0267	0.0314	0.0268	0.0253	0.0278
	0.657	0.779	0.816	0.827	0.803	0.719	0.811	0.826	0.791
	0.000	0.005	0.002	0.004	0.001	0.005	0.002	0.004	0.007
Double Gamma	0.0231	0.0223	0.0216	0.0194	0.0176	0.0238	0.8250	0.0170	0.0174
	0.846	0.863	0.871	0.896	0.911	0.836	0.000	0.916	0.915
	0.002	0.001	0.001	0.002	0.005	0.001	0.000	0.011	0.001
Sargan	0.7745	0.0387	0.0215	0.2310	0.0257	0.0243	0.0239	0.0275	0.0234
	0.000	0.611	0.873	0.000	0.811	0.834	0.841	0.779	0.851
	0.000	0.000	0.001	0.000	0.004	0.004	0.004	0.003	0.003

- Notes:
1. For each combination of distribution and volatility, the first number is the sample averaged loss, the second and the third numbers are the reality check p -values of White's (2000) test and Hansen's (2001) test, respectively. Each model from a combination of distribution and volatility is regarded as a benchmark model and is compared with the remaining 116 models from different combinations.
 2. We retrieve the S&P500 returns series from *finance.yahoo.com*. The sample observations are from January 3, 1990 to June 30, 2003 ($T = 3303$), the in-sample observations are from January 3, 1990 to September 24, 1996 ($R = 1703$), and the out-of-sample observations are from September 25, 1996 to June 30, 2003 ($n = 1700$).
 3. We use an MA(1) model, without a constant, for the mean, under rolling scheme. The loss is based on an AR(3)-SNP(8) model for the transformed PITs $\{x_t\}$.

Table 1: Reality Check Results for S&P500

Panel D: 1% Tail

	EWMA	GARCH	GJR	APARCH	EGARCH	STGARCH	HYGARCH	HYAPARCH	CGARCH
Historical	0.0021	0.0050	0.0049	0.0039	0.0044	0.0155	0.0077	0.0039	0.0051
	0.935	0.609	0.630	0.781	0.697	0.289	0.435	0.780	0.604
	0.693	0.159	0.207	0.424	0.296	0.014	0.066	0.423	0.153
Normal	0.0102	0.0122	0.0103	0.0102	0.0108	0.0121	0.0120	0.0101	0.0106
	0.362	0.325	0.358	0.358	0.348	0.324	0.329	0.363	0.360
	0.015	0.004	0.007	0.006	0.004	0.006	0.004	0.007	0.006
Student t	0.0079	0.0071	0.0095	0.0099	0.0096	0.0098	0.0075	0.0099	0.0071
	0.418	0.443	0.374	0.365	0.371	0.367	0.424	0.365	0.442
	0.038	0.048	0.015	0.007	0.013	0.012	0.032	0.007	0.048
GED	0.0071	0.0104	0.0092	0.0095	0.0093	0.0094	0.0074	0.0103	0.0070
	0.451	0.361	0.379	0.370	0.378	0.376	0.430	0.358	0.448
	0.043	0.005	0.010	0.009	0.013	0.015	0.031	0.008	0.041
Laplace	0.0005	0.0037	0.0073	0.0065	0.0058	0.0032	0.0066	0.0058	0.0066
	0.998	0.795	0.440	0.517	0.548	0.860	0.494	0.547	0.484
	0.926	0.391	0.048	0.104	0.115	0.498	0.117	0.112	0.102
Double Weibull	0.0072	0.0073	0.0093	0.0091	0.0093	0.0079	0.0067	0.0094	0.0066
	0.438	0.446	0.376	0.382	0.380	0.417	0.462	0.378	0.478
	0.046	0.062	0.010	0.012	0.010	0.033	0.056	0.010	0.051
Skewed t	0.0080	0.0047	0.0033	0.0034	0.0034	0.9100	0.0049	0.0038	0.0042
	0.456	0.636	0.828	0.820	0.827	0.000	0.616	0.770	0.719
	0.100	0.220	0.487	0.481	0.490	0.000	0.210	0.394	0.350
Hansen t	0.0087	0.0013	0.0000	0.0000	0.0000	0.0013	0.0000	0.0000	0.0013
	0.401	0.956	1.000	1.000	1.000	0.956	1.000	1.000	0.956
	0.041	0.763	0.943	0.943	0.943	0.767	0.943	0.943	0.768
Skewed GED	0.0069	0.0065	0.0098	0.0103	0.0084	1.0374	0.0066	0.0083	0.0038
	0.451	0.484	0.365	0.363	0.404	0.000	0.480	0.410	0.715
	0.042	0.054	0.009	0.016	0.020	0.000	0.052	0.017	0.369
IHS	0.0074	0.0067	0.0085	0.0080	0.0083	0.0069	0.0066	0.0081	0.0075
	0.432	0.472	0.399	0.417	0.414	0.458	0.473	0.417	0.443
	0.046	0.062	0.016	0.026	0.022	0.055	0.062	0.026	0.074
Mixture	0.0066	0.0059	0.0069	0.0069	0.0073	0.0075	0.0066	0.0065	0.0066
	0.481	0.522	0.467	0.468	0.440	0.428	0.497	0.493	0.484
	0.109	0.091	0.058	0.058	0.034	0.037	0.121	0.069	0.095
Double Gamma	0.0070	0.0061	0.0078	0.0069	0.0074	0.0062	0.0062	0.0044	0.0057
	0.448	0.499	0.421	0.452	0.442	0.497	0.509	0.713	0.536
	0.055	0.066	0.032	0.040	0.042	0.072	0.091	0.307	0.101
Sargan	0.0000	0.0000	0.0014	0.0187	0.0015	0.0000	0.0118	0.0101	0.0049
	1.000	1.000	0.951	0.217	0.935	1.000	0.334	0.363	0.654
	0.943	0.943	0.767	0.000	0.765	0.943	0.005	0.007	0.338

- Notes:
1. For each combination of distribution and volatility, the first number is the sample averaged loss, the second and the third numbers are the reality check p -values of White's (2000) test and Hansen's (2001) test, respectively. Each model from a combination of distribution and volatility is regarded as a benchmark model and is compared with the remaining 116 models from different combinations.
 2. We retrieve the S&P500 returns series from *finance.yahoo.com*. The sample observations are from January 3, 1990 to June 30, 2003 ($T = 3303$), the in-sample observations are from January 3, 1990 to September 24, 1996 ($R = 1703$), and the out-of-sample observations are from September 25, 1996 to June 30, 2003 ($n = 1700$).
 3. We use an MA(1) model, without a constant, for the mean, under rolling scheme. The loss is based on an AR(3)-SNP(8) model for the transformed PITs $\{x_t\}$.

Table 2: Reality Check Results for NASDAQ

Panel A: Whole Distribution

	EWMA	GARCH	GJR	APARCH	EGARCH	STGARCH	HYGARCH	HYAPARCH	CGARCH
Historical	0.0282	0.0603	0.0144	0.0487	0.0207	0.0607	0.0471	0.0487	0.0402
	0.340	0.116	0.744	0.138	0.521	0.117	0.156	0.142	0.206
	0.086	0.025	0.410	0.015	0.062	0.027	0.024	0.015	0.040
Normal	0.0068	0.0275	0.0319	0.0306	0.0302	0.0300	0.0274	0.0305	0.0220
	0.991	0.386	0.314	0.341	0.348	0.334	0.394	0.343	0.516
	0.624	0.140	0.109	0.130	0.132	0.109	0.157	0.132	0.167
Student t	0.0167	0.0104	0.0161	0.0014	0.0227	0.0222	0.0118	0.0147	0.0070
	0.711	0.947	0.735	1.000	0.491	0.492	0.861	0.787	0.990
	0.198	0.428	0.203	0.965	0.066	0.046	0.350	0.256	0.639
GED	0.0241	0.0004	0.0278	0.0264	0.0023	0.0260	0.0230	0.0261	0.0210
	0.451	1.000	0.366	0.408	1.000	0.405	0.484	0.414	0.538
	0.180	0.894	0.096	0.133	0.836	0.113	0.052	0.137	0.072
Laplace	0.0624	0.0107	0.0676	0.0417	0.0602	0.0126	0.0540	0.0594	0.0569
	0.041	0.942	0.035	0.129	0.045	0.891	0.070	0.047	0.059
	0.000	0.396	0.000	0.008	0.000	0.417	0.000	0.000	0.000
Double Weibull	0.0245	0.0310	0.0358	0.0115	0.0252	0.0257	0.0234	0.0246	0.0193
	0.435	0.279	0.200	0.853	0.429	0.412	0.463	0.449	0.605
	0.161	0.081	0.047	0.513	0.132	0.119	0.042	0.044	0.124
Skewed t	0.0326	0.0272	0.0222	0.0201	0.0201	0.0116	0.0277	0.0204	0.0308
	0.273	0.387	0.504	0.557	0.560	0.875	0.370	0.553	0.307
	0.062	0.149	0.104	0.163	0.140	0.507	0.137	0.152	0.097
Hansen t	0.1064	0.4548	0.2357	0.8554	0.1468	0.0436	0.1433	0.1444	0.0531
	0.004	0.000	0.000	0.000	0.000	0.133	0.000	0.000	0.078
	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.001
Skewed GED	0.0156	0.0146	0.0215	0.0126	0.0219	0.0213	0.0197	0.0219	0.0171
	0.742	0.822	0.568	0.891	0.508	0.526	0.570	0.506	0.672
	0.272	0.332	0.138	0.417	0.056	0.049	0.079	0.063	0.162
IHS	0.0141	0.0159	0.0132	0.0128	0.0127	0.0180	0.0167	0.0137	0.0069
	0.802	0.782	0.848	0.859	0.861	0.700	0.743	0.837	0.992
	0.333	0.268	0.383	0.409	0.415	0.233	0.278	0.361	0.658
Mixture	0.0175	0.0247	0.0251	0.0241	0.0228	0.0262	0.0204	0.0220	0.0216
	0.689	0.420	0.415	0.457	0.489	0.372	0.560	0.537	0.517
	0.160	0.022	0.025	0.029	0.045	0.011	0.056	0.124	0.052
Double Gamma	0.0267	0.0285	0.0002	0.0328	0.0069	0.0017	0.0108	0.0124	0.0295
	0.375	0.331	0.991	0.249	0.982	1.000	0.932	0.862	0.306
	0.097	0.078	0.856	0.041	0.709	0.937	0.522	0.433	0.072
Sargan	0.0152	0.0240	0.0181	0.0191	0.0329	0.0387	0.0070	0.0326	0.0334
	0.778	0.466	0.677	0.642	0.220	0.186	0.992	0.228	0.300
	0.282	0.091	0.237	0.189	0.020	0.023	0.595	0.022	0.116

- Notes:
1. For each combination of distribution and volatility, the first number is the sample averaged loss, the second and the third numbers are the reality check p -values of White's (2000) test and Hansen's (2001) test, respectively. Each model from a combination of distribution and volatility is regarded as a benchmark model and is compared with the remaining 116 models from different combinations.
 2. We retrieve the NASDAQ returns series from *finance.yahoo.com*. The sample observations are from January 3, 1990 to June 30, 2003 ($T = 3303$), the in-sample observations are from January 3, 1990 to September 24, 1996 ($R = 1703$), and the out-of-sample observations are from September 25, 1996 to June 30, 2003 ($n = 1700$).
 3. We use an MA(1) model, without a constant, for the mean, under rolling scheme. The loss is based on an AR(3)-SNP(8) model for the transformed PITs $\{x_t\}$.

Table 2: Reality Check Results for NASDAQ

Panel B: 10% Tail

	EWMA	GARCH	GJR	APARCH	EGARCH	STGARCH	HYGARCH	HYAPARCH	CGARCH
Historical	0.0697	0.1271	0.1245	0.1426	0.0806	0.1122	0.1128	0.1596	0.1181
	0.026	0.000	0.001	0.000	0.006	0.001	0.001	0.000	0.001
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Normal	0.0433	0.0468	0.0500	0.0495	0.0501	0.0479	0.0462	0.0495	0.0463
	0.203	0.152	0.122	0.125	0.118	0.142	0.158	0.125	0.159
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Student t	0.0508	0.0541	0.0601	0.0578	0.0580	0.0566	0.0529	0.0570	0.0540
	0.110	0.086	0.049	0.057	0.060	0.068	0.092	0.063	0.093
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
GED	0.0545	0.0527	0.0564	0.0559	0.0556	0.0550	0.0518	0.0554	0.0530
	0.079	0.097	0.070	0.073	0.078	0.080	0.100	0.077	0.100
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Laplace	0.0635	0.0633	0.0658	0.0608	0.0595	0.0667	0.0553	0.0602	0.0608
	0.041	0.042	0.033	0.049	0.057	0.033	0.079	0.051	0.053
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Double Weibull	0.0563	0.0424	0.0443	0.0457	0.0599	0.0559	0.0524	0.0579	0.0551
	0.069	0.234	0.199	0.180	0.050	0.074	0.099	0.061	0.086
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Skewed t	0.0293	0.0298	0.0290	0.0293	0.0315	0.0322	0.0289	0.0288	0.8013
	0.504	0.482	0.511	0.496	0.433	0.415	0.506	0.517	0.000
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Hansen t	0.0877	0.0065	0.0116	0.9604	0.0133	0.7039	0.0124	0.0122	0.7005
	0.002	1.000	0.973	0.000	0.940	0.000	0.946	0.961	0.000
	0.000	0.969	0.193	0.000	0.082	0.000	0.139	0.123	0.000
Skewed GED	0.0540	0.0436	0.0427	0.0416	0.0528	0.0507	0.0493	0.0529	0.0488
	0.088	0.212	0.221	0.239	0.102	0.113	0.128	0.103	0.129
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
IHS	0.0567	0.0470	0.0466	0.0461	0.0473	0.0449	0.0460	0.0451	0.0528
	0.065	0.174	0.181	0.184	0.175	0.199	0.184	0.202	0.099
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Mixture	0.0513	0.0537	0.0578	0.0563	0.0540	0.0565	0.0520	0.0434	0.0551
	0.110	0.092	0.065	0.072	0.094	0.072	0.101	0.214	0.088
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Double Gamma	0.0547	0.0553	0.0558	0.0553	0.0512	0.0454	0.0482	0.0523	0.0545
	0.079	0.080	0.075	0.076	0.124	0.195	0.158	0.103	0.090
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Sargan	0.0570	0.0438	0.0432	0.0426	0.0603	0.0818	0.2441	0.0613	0.1278
	0.064	0.215	0.224	0.227	0.054	0.013	0.000	0.051	0.001
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

- Notes:
1. For each combination of distribution and volatility, the first number is the sample averaged loss, the second and the third numbers are the reality check p -values of White's (2000) test and Hansen's (2001) test, respectively. Each model from a combination of distribution and volatility is regarded as a benchmark model and is compared with the remaining 116 models from different combinations.
 2. We retrieve the NASDAQ returns series from *finance.yahoo.com*. The sample observations are from January 3, 1990 to June 30, 2003 ($T = 3303$), the in-sample observations are from January 3, 1990 to September 24, 1996 ($R = 1703$), and the out-of-sample observations are from September 25, 1996 to June 30, 2003 ($n = 1700$).
 3. We use an MA(1) model, without a constant, for the mean, under rolling scheme. The loss is based on an AR(3)-SNP(8) model for the transformed PITs $\{x_t\}$.

Table 2: Reality Check Results for NASDAQ

Panel C: 5% Tail

	EWMA	GARCH	GJR	APARCH	EGARCH	STGARCH	HYGARCH	HYAPARCH	CGARCH
Historical	0.0272	0.2049	0.1902	0.1478	0.0452	0.1637	0.1674	0.2049	0.1706
	0.692	0.000	0.000	0.000	0.282	0.000	0.000	0.000	0.000
	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Normal	0.0268	0.0320	0.0308	0.0306	0.0313	0.0325	0.0321	0.0297	0.0294
	0.701	0.588	0.609	0.623	0.605	0.576	0.588	0.639	0.648
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Student <i>t</i>	0.0300	0.0355	0.0355	0.0339	0.0331	0.0368	0.0355	0.0336	0.0331
	0.629	0.499	0.507	0.544	0.561	0.462	0.500	0.546	0.560
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
GED	0.0310	0.0327	0.0324	0.0320	0.0325	0.0347	0.0325	0.0319	0.0301
	0.608	0.567	0.576	0.589	0.579	0.517	0.566	0.592	0.631
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Laplace	0.0307	0.0316	0.0304	0.0271	0.0280	0.0329	0.0295	0.0263	0.0294
	0.613	0.591	0.629	0.697	0.679	0.562	0.637	0.722	0.651
	0.000	0.000	0.000	0.001	0.001	0.000	0.000	0.001	0.000
Double Weibull	0.0322	0.0237	0.7445	0.0259	0.0325	0.0350	0.0341	0.0313	0.0325
	0.586	0.748	0.000	0.709	0.577	0.510	0.529	0.612	0.579
	0.000	0.002	0.000	0.001	0.000	0.000	0.000	0.000	0.000
Skewed <i>t</i>	0.0132	0.0093	0.0115	0.0118	0.0127	0.0108	0.0086	0.0185	0.0086
	0.940	0.984	0.964	0.964	0.952	0.968	0.985	0.876	0.986
	0.005	0.073	0.018	0.016	0.007	0.025	0.111	0.005	0.096
Hansen <i>t</i>	0.0447	0.0025	0.0026	0.0029	0.0048	0.0130	0.0032	0.0045	0.0103
	0.315	1.000	1.000	1.000	0.996	0.936	1.000	0.997	0.965
	0.000	0.987	0.766	0.702	0.543	0.034	0.706	0.549	0.071
Skewed GED	0.0336	0.0253	0.0267	0.0269	0.0264	0.9544	0.9437	0.0262	0.0266
	0.555	0.725	0.692	0.691	0.716	0.000	0.000	0.715	0.709
	0.000	0.001	0.001	0.001	0.001	0.000	0.000	0.000	0.000
IHS	0.0360	0.0269	0.7443	0.0299	0.0295	0.0258	0.0264	0.0291	0.0280
	0.503	0.695	0.000	0.628	0.637	0.714	0.703	0.643	0.685
	0.000	0.001	0.000	0.000	0.000	0.001	0.001	0.000	0.000
Mixture	0.0217	0.0299	0.0264	0.0264	0.0272	0.0314	0.0291	0.0281	0.0298
	0.814	0.631	0.713	0.715	0.694	0.591	0.644	0.664	0.636
	0.008	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.000
Double Gamma	0.0249	0.0267	0.0287	0.0272	0.0291	0.0228	0.0247	0.8431	0.0268
	0.745	0.705	0.663	0.697	0.647	0.763	0.727	0.000	0.705
	0.001	0.000	0.000	0.000	0.000	0.003	0.001	0.000	0.001
Sargan	0.0354	0.0264	0.7235	0.7235	0.0282	0.1827	0.0262	0.0281	0.1621
	0.520	0.704	0.000	0.000	0.675	0.000	0.725	0.675	0.000
	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000

- Notes:
1. For each combination of distribution and volatility, the first number is the sample averaged loss, the second and the third numbers are the reality check *p*-values of White's (2000) test and Hansen's (2001) test, respectively. Each model from a combination of distribution and volatility is regarded as a benchmark model and is compared with the remaining 116 models from different combinations.
 2. We retrieve the NASDAQ returns series from *finance.yahoo.com*. The sample observations are from January 3, 1990 to June 30, 2003 ($T = 3303$), the in-sample observations are from January 3, 1990 to September 24, 1996 ($R = 1703$), and the out-of-sample observations are from September 25, 1996 to June 30, 2003 ($n = 1700$).
 3. We use an MA(1) model, without a constant, for the mean, under rolling scheme. The loss is based on an AR(3)-SNP(8) model for the transformed PITs $\{x_t\}$.

Table 2: Reality Check Results for NASDAQ

Panel D: 1% Tail

	EWMA	GARCH	GJR	APARCH	EGARCH	STGARCH	HYGARCH	HYAPARCH	CGARCH
Historical	0.0035	0.0814	0.0072	0.0171	0.0058	0.0896	0.0194	0.0154	0.1036
	0.972	0.005	0.886	0.611	0.928	0.004	0.557	0.664	0.006
	0.752	0.005	0.442	0.113	0.462	0.004	0.081	0.156	0.006
Normal	0.0075	0.0096	0.0114	0.0117	0.0121	0.0160	0.0095	0.0116	0.0083
	0.858	0.802	0.753	0.738	0.729	0.618	0.803	0.744	0.830
	0.339	0.235	0.137	0.137	0.120	0.123	0.241	0.137	0.280
Student <i>t</i>	0.0068	0.0073	0.0092	0.0101	0.0103	0.0072	0.0073	0.0098	0.0055
	0.885	0.865	0.824	0.787	0.783	0.864	0.865	0.790	0.931
	0.356	0.311	0.252	0.207	0.203	0.310	0.309	0.224	0.468
GED	0.0091	0.0071	0.0102	0.0105	0.0103	0.0079	0.0074	0.0099	0.0052
	0.788	0.869	0.794	0.772	0.783	0.850	0.868	0.789	0.935
	0.283	0.312	0.208	0.191	0.206	0.321	0.345	0.223	0.509
Laplace	0.0056	0.0061	0.0067	0.0063	0.0051	0.0072	0.0081	0.0070	0.0046
	0.942	0.916	0.906	0.906	0.947	0.882	0.895	0.894	0.959
	0.463	0.402	0.350	0.382	0.502	0.311	0.307	0.340	0.570
Double Weibull	0.0164	0.0045	0.0051	0.0056	0.0088	0.0062	0.0062	0.0083	0.0169
	0.559	0.933	0.908	0.895	0.823	0.909	0.908	0.836	0.662
	0.054	0.596	0.574	0.523	0.264	0.378	0.376	0.287	0.154
Skewed <i>t</i>	0.0028	0.0029	0.0057	0.0034	0.0035	0.0029	0.0028	0.0040	0.0057
	0.988	0.988	0.914	0.970	0.970	0.988	0.988	0.958	0.952
	0.837	0.818	0.469	0.717	0.716	0.813	0.826	0.636	0.490
Hansen <i>t</i>	0.0075	0.0000	0.0000	0.0000	0.0000	0.0008	0.0000	0.0000	0.0017
	0.837	1.000	1.000	1.000	1.000	0.999	1.000	1.000	0.983
	0.381	0.955	0.955	0.955	0.955	0.982	0.955	0.955	0.877
Skewed GED	0.0116	0.0068	0.0079	0.0078	0.0058	0.0867	0.0058	0.0067	0.1147
	0.698	0.855	0.820	0.821	0.928	0.008	0.935	0.890	0.008
	0.153	0.427	0.348	0.351	0.445	0.008	0.444	0.352	0.008
IHS	0.0115	0.0078	0.0079	0.0087	0.0104	0.0069	0.0078	0.0082	0.0047
	0.708	0.830	0.820	0.796	0.745	0.853	0.827	0.811	0.958
	0.160	0.349	0.349	0.308	0.223	0.412	0.355	0.333	0.561
Mixture	0.0035	0.0056	0.0063	0.0063	0.0060	0.0057	0.0057	0.0069	0.0696
	0.975	0.937	0.911	0.912	0.927	0.936	0.935	0.848	0.018
	0.712	0.452	0.380	0.381	0.426	0.453	0.453	0.426	0.018
Double Gamma	0.0056	0.0051	0.0057	0.0059	0.0068	0.0364	0.0049	0.0052	0.0053
	0.946	0.953	0.928	0.919	0.855	0.299	0.925	0.911	0.954
	0.473	0.521	0.448	0.453	0.438	0.123	0.556	0.575	0.502
Sargan	0.0105	0.0074	0.0077	0.0076	0.0008	0.0000	0.4629	0.0135	0.0000
	0.748	0.844	0.818	0.818	0.994	1.000	0.000	0.692	1.000
	0.203	0.389	0.371	0.371	0.935	0.955	0.000	0.261	0.955

- Notes:
1. For each combination of distribution and volatility, the first number is the sample averaged loss, the second and the third numbers are the reality check p -values of White's (2000) test and Hansen's (2001) test, respectively. Each model from a combination of distribution and volatility is regarded as a benchmark model and is compared with the remaining 116 models from different combinations.
 2. We retrieve the NASDAQ returns series from *finance.yahoo.com*. The sample observations are from January 3, 1990 to June 30, 2003 ($T = 3303$), the in-sample observations are from January 3, 1990 to September 24, 1996 ($R = 1703$), and the out-of-sample observations are from September 25, 1996 to June 30, 2003 ($n = 1700$).
 3. We use an MA(1) model, without a constant, for the mean, under rolling scheme. The loss is based on an AR(3)-SNP(8) model for the transformed PITs $\{x_t\}$.