Comparing Density Forecasts in a Risk Management Context

Cees Diks *[1 2] and Hao Fang[1 2]

[1]Center for Nonlinear Dynamics in Economics and Finance (CeNDEF), University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands

[2]Tinbergen Institute, Gustav Mahlerplein 117, 1082 MS Amsterdam, The Netherlands

5th August 2016

Abstract

We compare multivariate and univariate approaches to assessing the accuracy of competing density forecasts of a portfolio return in the downside part of the support. We argue that the common practice to perform multivariate forecast comparisons can be problematic in the context of assessing portfolio risk, since better multivariate forecasts do not necessarily correspond to better aggregate portfolio return forecasts. This is illustrated by examples involving (skew) elliptical distributions and an application to daily returns of a number of US stock prices. Additionally, time-varying test statistics and Value-at-Risk forecasts provide empirical evidence for regime changes over the last decades.

Keywords: Density forecast evaluation; scoring rules; skew-elliptical distributions; portfolio risk assessment; Value-at-Risk forecasts

JEL codes: C12, C46, C52, C53, G17, G31

*Corresponding author: Center for Nonlinear Dynamics in Economics and Finance, Amsterdam School of Economics, Faculty of Economics and Business, University of Amsterdam, Roetersstraat 11, 1018 WB, Amsterdam, The Netherlands. E-mail: `C.G.H.Diks@uva.nl`, Phone: +31 20 525 5329.

1 Introduction

Since Markowitz (1952) laid the foundation of modern portfolio theory, theoretical research in portfolio theory has mainly centered around the first two predictive moments of portfolio returns, which are determined by the expected returns on the constituting assets and their conditional variance-covariance structure. See e.g. Merton (1992) for the economic theory behind portfolio selection and Elton and Gruber (1997) and Elton et al. (2009) for a detailed review of modern portfolio analysis.

However, as Granger (2003) pointed out, "*... it is most natural to consider the whole conditional distribution ... particularly in many dimensions.*" Ever-increasing computational power and memory now actually enable the estimation of complicated multivariate models; multivariate GARCH models with various distributional assumptions on the error terms are now commonly used to describe and forecast the temporal dependence of several assets and markets. A comprehensive survey of multivariate GARCH models is provided by Bauwens et al. (2006).

Along with the prosperity of multivariate modeling of financial returns, portfolio managers are now weaponed with various tools to estimate and forecast the distribution of asset returns, which forms the basis for risk measurement and position adjustment. In portfolio decision and risk assessment, density forecasts form the basis for risk measures such as Value-at-Risk (VaR) and Expected Shortfall (ES); see Alexander (2009) for recent applications. Risk management and portfolio selection involve the assessment of expected returns and risk of portfolios. In this paper we focus on the latter: assessing the (short-term) risk associated with a given portfolio. Dynamic portfolio selection strategies (trading strategies) and their associated (long-term) returns and risk are also important in practice, but go beyond the more modest aim of this paper, which is to discuss some caveats of comparing density forecasts for the returns of a given portfolio.

Nowadays many alternative density forecasts for financial returns are available, and a natural question both researchers and practitioners face for any given portfolio therefore is: Among a number of competing density forecasts for a portfolio, which most accurately describes the distribution of losses (the left tail of the return distribution)?

Traditionally multivariate normality has been taken for granted in portfolio theory research; see, e.g., Kandel and Stambaugh (1996), Guidolin and Timmermann (2007) for asset allocation under multivariate normal distributions. However, over the past two decades it has become clear that financial multivariate time series are typically non-normally distributed, exhibiting fat tails, skewness and nonlinear co-movements or asymmetric dependence structure; see, for instance, Harvey and Siddique (1999), Brooks et al. (2005) and Patton (2004). Patton (2004) and Boubaker and Sghaier (2013) studied portfolio optimization with dependence measured

by competing copula densities; in fact in both studies the Gumbel copula turned out to be preferred. Giot and Laurent (2003) showed that a model with skew t innovations outperforms those with symmetric distributions in VaR estimation; Huang et al. (2015) suggested that better portfolio performance is achieved with a time-varying copula, particularly the Clayton copula. Diks et al. (2010) and Diks et al. (2014) found that for daily exchange rate returns, the t copula is favored over its counterparts, while in the government bond market, a mixture of t and Clayton copulas performs best for the daily changes of yields on government bonds of the G7 countries.

Tests for the relative accuracy of two or more competing density forecasts can be based on loss functions measuring the distance between predicted density and the (unknown) true density. Possible loss functions include the integrated squared difference discussed by Sarno and Valente (2004) and the mean squared error (MSE) considered by Corradi and Swanson (2005) and Giacomini and White (2006). Amisano and Giacomini (2007) proposed a so-called weighted logarithmic scoring rule, with a weight function allowing predictive accuracy comparison in a particular region of the distribution, which is a desirable property for testing in tails. A score-based approach is mathematically convenient as it leads to simple pseudo-likelihood ratio type t-tests.

An important restriction we wish to impose on weighted scoring rules is properness (Matheson and Winkler, 1976). Intuitively, a scoring rule is proper if it never assigns a higher expected score to any density forecast than to the true conditional density. Gneiting and Raftery (2007) suggested that apart from the logarithmic score, the continuous ranked probability score ($crsp$) is proper. Gneiting and Ranjan (2011) pointed out that the weighted logarithmic scoring rule by Amisano and Giacomini (2007) is improper and introduced a proper weighted $csrp$ rule to compare density forecasts. Diks et al. (2011) also noted the problem of weighted likelihood tests, and came up with two proper related scoring rules: the conditional likelihood (cl) and the censored likelihood (csl) scoring rule. Like the log-likelihood scoring rule, these are based on the Kullback-Leibler Information Criterion (KLIC), a well-known and widely-used measure of divergence between two probability distributions; see, e.g., Vuong (1989); Giacomini and White (2006) and Diks et al. (2011). Motivated by desirable properties of weighted scoring rules, Pelenis (2014) also proposed two alternative scoring rules: the penalized weighted likelihood (pwl) scoring rule and the incremental weighted $crsp$ rule. Since the extension of the $csrp$ scoring rules to the multivariate setting is not straightforward, we restrict ourselves to the KLIC-based scoring rules, cl, csl and pwl, which can be readily applied to multivariate density forecasts.

When the dimension (number of assets) increases, the tests for predictive accuracy of density forecasts based on alternative multivariate models will become increasingly demanding compu-

tationally. Hence it then becomes attractive to calculate scores based on a univariate density forecasts for the portfolio return. This leads to the second question this paper wishes to investigate: Given a portfolio, should density forecast selection be based on comparing scores for (i) multivariate density forecasts of the joint asset returns distribution, (ii) univariate projections of these multivariate density forecasts onto the portfolio return, or (iii) direct univariate density forecasts for the portfolio return?

Common practice is to follow approach (i), that is, to construct a (usually model-based) multivariate density forecast first, and then use the implied univariate portfolio density forecast to assess portfolio risk. The problem of this routine is that the forecast that is better according to the multivariate evaluation needs not correspond to a better portfolio return forecast. Even worse, high-dimensional information could be misleading. Section 2.1 provides a simple example to illustrate this.

When comparing approach (i), the evaluation of multivariate density forecasts, with (ii), the evaluation of the corresponding univariate projections, we focus on density forecasts within (skew) elliptical families of distributions. These families of elliptical distributions are introduced here for mathematical convenience since they are closed under linear affine transformations. Typical elliptical families of distributions are the Normal and t-distributions. Besides these symmetric elliptical distributions, two more flexible families of distributions, the skew normal and skew t-distributions defined by Azzalini and Dalla Valle (1996) and Azzalini and Capitanio (2003), respectively, are considered as well.

The contribution of this paper is three-fold. First, we introduce skew elliptical distributions in financial returns modeling and provide comprehensive comparisons between candidate densities, both across dimensions and across families of distributions, using proper scoring rules. Second, we shed some light on the pitfalls of multivariate modeling in portfolio risk assessment. Third, the time-varying nature of which family of forecast densities performs best is demonstrated empirically and further confirmed by the superior performance of dynamically selected densities in VaR estimation.

The remainder of this paper is organized as follows. Section 2 describes the methodology, including the testing environment, the score rules, test statistics and the distributional assumptions. Section 3 provides Monte Carlo simulation experiments to study the empirical size and power of the tests. In Section 4 we apply the tests to returns on a portfolio and on its underlying stocks. Section 5 develops a simple technique to dynamically select the distribution for VaR forecasts. The final section summarizes and concludes.

2 Methodology

2.1 Density Comparison within Elliptical Families

We are concerned with the evaluation of the predictive accuracy of two competing multivariate density forecasts versus evaluating their corresponding univariate projections. From a portfolio perspective, the forecasted densities of a portfolio consisting of some assets could be evaluated in multivariate space, or in the univariate space of the portfolio return. The implied distribution of the portfolio return, if explicitly available, will be directly relevant to the assessment of the portfolio risk. Indeed, there are some circumstances in which the higher dimensional information is irrelevant, or even misleading. When we perform a statistical test on two competing density forecasts $\hat{f}$ and $\hat{g}$, it may happen that the multivariate scoring rule favors distribution $\hat{f}$ over $\hat{g}$, while the univariate scoring rule suggests the opposite. The following example illustrates this paradox.

Example 1. Suppose we intend to compare the accuracy of two competing density forecasts, where the true density follows a standard bivariate Normal distribution $\boldsymbol{Y} = (Y_1, Y_2) \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{I})$. Suppose the two forecasts are

$$\begin{aligned} \hat{f}: &\quad \boldsymbol{Y} \sim \boldsymbol{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1), \text{ with } \boldsymbol{\mu}_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix} \text{ and } \Sigma_1 = \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix} \\ \hat{g}: &\quad \boldsymbol{Y} \sim \boldsymbol{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2), \text{ with } \boldsymbol{\mu}_2 = \begin{pmatrix} 0.2 \\ 0.2 \end{pmatrix} \text{ and } \Sigma_2 = \begin{pmatrix} 1 & -0.2 \\ -0.2 & 1 \end{pmatrix}. \end{aligned}$$

Contour plots of the density of the DGP and two density forecasts are shown in Fig. 1(a). The dotted circles correspond to $\boldsymbol{N}(\boldsymbol{0}, \boldsymbol{I})$, the true distribution from the DGP, while the solid and dashed lines give the contours of 'forecasts' 1 and 2, respectively. Clearly forecast 2 is closer to the true distribution, and a reasonable test for the predictive ability of the bivariate densities should be in favor of $\hat{g}$ over $\hat{f}$. However, if we construct a portfolio defined as $Z = b_1 Y_1 + b_2 Y_2$, the order of the predictive ability of the forecasts may change. Consider the simple case where $b_1 = b_2 = 1$. Fig. 1(b) gives the univariate densities of the DGP (dotted line), forecast 1 (solid line) and forecast 2 (dashed line). In contrast with the previous conclusion, the density of forecast 1 clearly is more close to the DGP now.

[Figure 1 about here.]

Therefore, even though density forecast $\hat{g}$ is more close to the true distribution in the bivariate space, for the portfolio risk evaluation it makes more sense to make decisions based on the distribution of $\hat{f}$. Even without a formal test, this plausible example illustrates the dilemma

faced by academia as well as industry; when the true distribution is unknown and with a linear combination of some marginal elements in hand, what is the basis for portfolio analysis, the high-dimensional information or the univariate distribution after aggregation?

In general, the comparison of density forecast evaluation across different dimensions is complicated by the fact that not many distributions are analytically tractable upon aggregation by taking linear combinations. In this paper we therefore deliberately limit ourselves to distributions from the elliptical class (ED class) of distributions for both mathematical convenience and numerical accuracy. Elliptical distributions, which are closed under affine transformations and linear aggregation, provide a tool by which we can analytically trace closed-form expressions for the densities. Specifically, the property of the ED family which facilitates our analysis is the following.

Property 1. If the d-variate random variable $\mathbf{Y} \sim \mathrm{ED}(\mathbf{\Delta}, \mathbf{\Omega})$, given any $\ell \times d$ matrix A of rank $\ell \leq d$, the random vector $\mathbf{Z} = A\mathbf{Y} \sim \mathrm{ED}(A\mathbf{\Delta}, A\mathbf{\Omega}A')$.

Typical families of distributions in the elliptical class include Normal, t and Cauchy. We refer to Kelker (1970), Cambanis et al. (1981), Fang et al. (1990), Arellano-Valle and Bolfarine (1995) for developments in the multivariate ED distribution theory.

Here we apply the Normal and t-distributions as these two better appear to fit financial data than the Cauchy distribution. Property 1 is crucial to our study in the sense that a linear combination of elliptically distributed assets remains in the ED family. Hence compared to the corresponding multivariate density forecast, which is usually costly to estimate, the density forecast of the portfolio return is easier to handle. Some handy properties and analytical expressions for multivariate t-distributions are summarized in Appendix A.1 for ease of reference.

Naturally, the next question is whether the class of ED models is rich enough to describe the random variables involving risk management and asset allocation accurately. To answer that, we wish to consider more families of elliptical distributions than just the symmetric Normal and t-distributions. Section 2.4 introduces skew normal and skew t-distributions developed in a number of papers by Azzalini and co-authors (1996, 2003, 2005, 2013). The merit of using these two skew families of distributions is that they allow for additional asymmetry while remaining closed under affine transformations.

2.2 Testing Approach

Consider a stochastic vector $\{(\boldsymbol{Y}_t', \boldsymbol{X}_t)\}$, $t = 1, 2, ..., T$, where $\boldsymbol{Y}_t = (Y_{1,t}, ..., Y_{d,t})'$ represents the $d-$dimensional vector the density of which is of interest, and where $\boldsymbol{X}_t$ is a vector of exogenous or observable predictor variables. In the context of time series data it is natural to consider

the conditional distribution $F_{\boldsymbol{Y}_{t+1}}(\boldsymbol{y}|\mathscr{F}_t)$, where $\mathscr{F}_t$ denotes the information available at time t. For simplicity we restrict ourselves to one-step-ahead density forecasts; generalizations to multi-step-ahead forecasts are straightforward but would provide little extra insight.

Following Giacomini and White (2006), we compare *forecast methods* rather than forecast models. By forecast method, we mean the model on which the forecast is based, along with estimation methods, applied observation weights and all other choices one makes at the time of the prediction. The only restriction is that the density forecasts depend on a finite number m of most recent observations. The advantage of this framework is that it allows for treating parameter estimation uncertainty as an integral part of the density forecasts. A fixed rolling window of length m is used for estimation. As shown by Giacomini and White (2006), this limited memory scheme affords considerable analytical convenience for asymptotic theory of the test of equal predictive accuracy of two competing density forecasts $\hat{f}_t$ and $\hat{g}_t$.

A prevalent approach to comparing the relative performance of density forecasts is based on scoring rules, which are loss functions whose arguments are the density forecast and the actual outcome of the variable; see Diebold and Lopez (1996). In the current context, the scoring rule for one-step-ahead forecasts is of the form $S(\hat{f}_t; \boldsymbol{y}_{t+1})$, depending on the density forecast $\hat{f}_t$ of $\boldsymbol{Y}_{t+1}$ given $\mathscr{F}_t$ and the actually observed value $\boldsymbol{y}_{t+1}$, such that a 'better' prediction receives a higher score on average.

Given a scoring rule $S^*(.)$, two competing density forecasts $\hat{f}_t$ and $\hat{g}_t$ and the corresponding realizations of the d-dimensional variable $\boldsymbol{y}_{t+1}$ for $t = m, ..., T-1$, we may compare $\hat{f}_t$ and $\hat{g}_t$ based on their mean scores. The test we perform in this paper, following Giacomini and White (2006), is an unconditional predictive ability test (although the density forecasts are made conditional on m in-sample observations). Under the moving window scheme, $n = T - m$ out-of-sample observations are available, and a formal test for whether $\hat{f}_t$ and $\hat{g}_t$ are significantly different is performed. Defining the score differences as

$$d^*_{t+1} = S^*(\hat{f}_t; \boldsymbol{y}_{t+1}) - S^*(\hat{g}_t; \boldsymbol{y}_{t+1}),$$

the null hypothesis of equal predictive ability is given by

$$H_0: \ \mathbb{E}(d^*_{t+1}) = 0, \qquad \text{for} \quad t = m, ..., T-1,$$

which is tested against the alternative hypothesis $H_1 : \mathbb{E}(d^*_{t+1}) \neq 0$ (or > 0 or < 0 for a one-sided test). Let $\bar{d}^*_{m,n}$ denote the out-of-sample average score difference: $\bar{d}^*_{m,n} = n^{-1} \sum_{t=m}^{T-1} d^*_t$, a

Diebold and Mariano (1995) type statistic is given by

$$t_{m,n} = \frac{\bar{d}^*_{m,n}}{\sqrt{\hat{\sigma}^2_{m,n}/n}}, \tag{1}$$

where $\hat{\sigma}^2_{m,n}$ is a heteroskedasticity and autocorrelation-consistent(HAC) variance estimator for the long-run variance $\sigma^2_{m,n} = \mathrm{Var}(\sqrt{n}\bar{d}^*_{m,n})$ given by $\hat{\sigma}^2_{m,n} \equiv \hat{\gamma}_0^2 + 2\sum_{k=1}^{K-1} a_k\hat{\gamma}_k$, in which $\hat{\gamma}_k$ denotes the sample covariance of sequence $\{d^*_{t+1}\}$ at lag k and the Bartlett weight $a_k = 1-k/K$ with the truncated length $K = \lfloor n^{1/4} \rfloor$.

Under the assumptions of a fixed estimation window size m and mild mixing conditions on the sequence of score differences $\{d^*_{t+1}\}$, the test statistic $t_{m,n}$ is asymptotically standard normally distributed by Theorem 4 in Giacomini and White (2006). A test of (asymptotic) significance level α therefore rejects the null hypothesis of equal performance when $|t_{m,n}| > z_{\alpha/2}$ for two-sided test, where $\alpha/2$ is the $1-\alpha/2$ quantile of standard normal distribution $N(0,1)$. Note that the sign of the test statistic $t_{m,n}$ indicates which of the two density forecasts $\hat{f}_t$ and $\hat{g}_t$ performs better, as scores will be defined such that a higher average score is preferred.

2.3 Weighted Logarithmic Scoring Rules

In this paper we focus on three different logarithmic scoring rules as they can be applied regardless of the dimension of $\boldsymbol{Y}$. A typical logarithmic scoring rule is of the form $S(\hat{f}_t; \boldsymbol{y}_{t+1}) = \log \hat{f}_t(\boldsymbol{y}_{t+1})$, and testing for equality of the average scores of $\hat{f}_t$ and $\hat{g}_t$ leads to the pseudo-likelihood ratio test originally developed by Vuong (1989). It has been shown in many studies that the pseudo-likelihood ratio test is closely related to the Kullback-Leibler Information Criterion (KLIC), an information theoretical goodness-of-fit measure of divergence between two probability distributions; see Vuong (1989), Mitchell and Hall (2005) and Bao et al. (2004). KLIC-based scores therefore quantify the divergence between a candidate density forecast $\hat{f}_t$ and the true density p_t. For our purpose of comparing candidate density $\hat{f}_t$ and $\hat{g}_t$ on a particular region such as the left tail, a weighted scoring rule is used, with a weight function $w(\boldsymbol{y}_{t+1})$ emphasizing the region of interest.

It is a natural requirement that the true density p_t is rewarded with the highest average score, otherwise the associated test might suggest that some wrong density forecast is significantly better than the true predictive density. To avoid such situations, we focus on proper scoring rules, which imply that no density forecast $\hat{f}_t$ receives a higher score, on average, than the true (unknown) density p_t, that is,

$$\mathbb{E}_t(S(\hat{f}_t; \boldsymbol{y}_{t+1})) \leq \mathbb{E}_t(S(p_t; \boldsymbol{y}_{t+1})), \text{ for all } t.$$

As mentioned above, Matheson and Winkler (1976) and Gneiting and Raftery (2007) suggested many suitable proper scoring rules, but these can not be readily extended to a multivariate context. We therefore restrict ourselves to KLIC-based scoring rules, which are practically convenient and closely related to pseudo-likelihood tests. Particularly, we consider the following three scoring rules, the first two of which were suggested by Diks et al. (2011), and the last by Pelenis (2014).

1. **The conditional likelihood scoring rule** (cl)

$$S^{cl}(\hat{f}_t; \boldsymbol{y}_{t+1}) = w_t(\boldsymbol{y}_{t+1}) \log\left(\frac{\hat{f}_t(\boldsymbol{y}_{t+1})}{\int w_t(\boldsymbol{s})\hat{f}_t(\boldsymbol{s})d\boldsymbol{s}}\right); \tag{2}$$

2. **The censored likelihood scoring rule** (csl)

$$S^{csl}(\hat{f}_t; \boldsymbol{y}_{t+1}) = w_t(\boldsymbol{y}_{t+1}) \log(\hat{f}_t(\boldsymbol{y}_{t+1})) + (1 - w_t(\boldsymbol{y}_{t+1})) \log\left(1 - \int w_t(\boldsymbol{s})\hat{f}_t(\boldsymbol{s})d\boldsymbol{s}\right); \tag{3}$$

3. **The penalized weighted likelihood scoring rule** (pwl)

$$S^{pwl}(\hat{f}_t; \boldsymbol{y}_{t+1}) = w_t(\boldsymbol{y}_{t+1})\log(\hat{f}_t(\boldsymbol{y}_{t+1})) - \int w_t(\boldsymbol{s})\hat{f}_t(\boldsymbol{s})d\boldsymbol{s} + w(\boldsymbol{y}_{t+1}) + w_t(\boldsymbol{y}_t). \tag{4}$$

At this point, we make the following assumptions.

Assumption 1. The density forecasts $\hat{f}_t$ and $\hat{g}_t$ satisfy $\text{KLIC}(\hat{f}_t) < \infty$ and $\text{KLIC}(\hat{g}_t) < \infty$, where $\text{KLIC}(h_t) = \int p_t(y) \log(p_t(y)/h_t(y))dy$ is the Kullback-Leibler divergence between the density forecast h_t and the true conditional density p_t.

Assumption 2. The weight function $w_t(y)$ is such that (a) it is determined by the information available at time t, and hence a function of $\mathscr{F}_t$, (b) $0 \leq w_t(y) \leq 1$, and (c) $\int w_t(y)p_t(y)dy > 0$.

Under Assumptions 1 and 2, the cl scoring rule (2) and the csl scoring rule (3) are proper, as established by Lemma 1 in Diks et al. (2011). In a similar manner, Appendix A.2 shows that the pwl scoring rule (4) is proper in terms of Kullback-Leibler divergences between weighted density forecast and the true density.

To illustrate the usage of the above-mentioned weighted scoring rules, we revisit Example 1 focusing on the left tail of the distributions by adopting the threshold weight function $w(z) = I(z \leq r)$, where $I(\cdot)$ denotes the indicator function taking the value 1 if its argument is true, and 0 otherwise, and where $z = b_1y_1 + b_2y_2$. In order to make the rejection rate more comparable for different threshold values r, we let the sample size n be determined by the threshold value

as $n = c/P(Z \leq r)$, for some $c > 0$, so that the expected number of observations in the region of interest, is fixed at c.

Example 2. We generate $10,000$ independent observations from the bivariate vector $\boldsymbol{Y} \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{I})$ for $c = 40$ and compare the performance of the two forecast candidates $\hat{f}$ and $\hat{g}$ defined in Example 1. Again we take $b_1 = b_2 = 1$ as the portfolio weights. Fig. 2 shows the observed one-sided rejection rates of the *csl* scoring rule for $r \in [-2.5, 2.5]$, evaluated in both bivariate space and univariate space, denoted by the solid and dashed lines, respectively. The rejection rates for the *cl* and *pwl* scoring rule behave similarly, and are available from the authors upon request.

[Figure 2 about here.]

Several interesting conclusions can be drawn from Fig. 2. First, the test based on the bivariate *csl* scoring rule significantly rejects the null hypothesis of equal predictive accuracy and favors $\hat{g}$, while the test based on the univariate *csl* scoring rule suggests the other way around. Especially for the left tail, when $r \leq -2$, the power of this one-sided test against the alternative that $\hat{f}$ is better, is higher than 0.5. This result is quite robust, even for $c = 20$ observations in the region of interest (on average), the rejection rate for the univariate scoring rule against superior predictive ability of $\hat{f}$ tends to 0.4 in the left tail.

Second, when comparing Figs 1 and 2, the apparently conflicting conclusions from the Diebold-Mariano type tests are not surprising after all. When constructing the linear combination $Z = b_1 Y_1 + b_2 Y_2$, essentially we project the bivariate vector $\boldsymbol{Y}$ from the plane $\mathbb{R}^2$ onto the real line $\mathbb{R}$. During this projection, the relative distance between the density forecasts and the true distribution changes. Fig. 1 suggests that the univariate projection of $\hat{f}$ is more accurate than $\hat{g}$, even though $\hat{f}$ is in fact the worse multivariate forecast.

Third, in this simple example it is not hard to see that it would be a mistake for a risk manager to believe that $\hat{g}$ is more appropriate based on the bivariate scores. Since the projected distribution of $\hat{g}$ is thinner-tailed than that of the DGP in the left tail (see Fig. 1(b)), the estimated VaR and ES will be insufficient to reflect the real risk level and large losses might be incurred.

Examples 1 and 2 illustrate the fact that a better forecast in $\mathbb{R}^d$ space may not lead to a better risk assessment from a portfolio perspective. Note that if one forecast, say $\hat{f}$, is perfect in the sense that $\hat{f} \equiv p$, where p is the true conditional density implied by the DGP, we will not be in such a dilemma because the perfect forecast $\hat{f}$ will dominate $\hat{g}$ in both the univariate space and the bivariate space. However, the issue of evaluating competing densities with multivariate scoring rule concerns us because in reality p is never reached; $\hat{f}$ and $\hat{g}$ are

always based on a misspecified model and therefore always incorrect to some extent. Even when one of the models would be correctly specified, parameter estimation uncertainty would prevent its associated density forecast from being perfect.

Hence, we observe that decisions based on the multivariate scoring rule need not coincide with that based on the corresponding univariate scoring rule. Arguably, from a portfolio risk assessment perspective, good univariate density forecasts are needed for portfolio returns, and these may not necessarily be the projected multivariate optimal forecasts.

2.4 Skew Elliptical Distributions

As Bauwens and Laurent (2005) noted, one well-established stylized fact of financial returns is that they often exhibit fat tails and skewness, and a more suitable distribution than the multivariate normal is of primary importance in modeling and inference. A general class of multidimensional distributions which allow for heavy tails and skewness will be useful to modeling multivariate random variables. Applications from Bauwens and Laurent (2005) and Giot and Laurent (2003) show that the skew t-distribution, when fully taking into account skewness and fat tails, provides a better fit for the value-at risk of portfolios.

However, the density defined by Bauwens and Laurent (2005), which was derived using a non-linear transformation of a symmetric density function, is inspiring, yet complicated. Moreover, their method is difficult to apply in the present context, where the aim is to compare the predictive distributions in different spaces. As we are unable to show that their family of distributions is closed under affine transformations, the distribution of a linear combination of marginal variables remains unknown. Jones and Faddy (2003) also considered a skew generalization of the t-distribution which is tractable for all moments, but which is difficult to generalize to the multivariate case.

Alternatively, skewness can be modeled by perturbing a symmetric distribution to generate asymmetric densities. This method was introduced by Azzalini and Dalla Valle (1996) to construct a multivariate skew-normal distribution. Later, Azzalini and Capitanio (2003) generalized the approach to distributions of elliptical families. A coincident result of multivariate skew-elliptical distributions is given by Branco and Dey (2001), although a different method is used. According to Azzalini and Capitanio (2003), there are at least two avenues to construct a skew-distributed random variable, first, the *conditioning method*, as used by Branco and Dey (2001), and second, the *transformation method* (e.g. Azzalini and Dalla Valle, 1996). We will not cover the mathematical details in this paper, but instead refer to Azzalini (2005) and Azzalini (2013) for the intensive development of distribution theory in this direction.

Before introducing the density and some properties of skew t-distributions, we first introduce the simpler skew normal distributions. Following Azzalini and Capitanio (2003), given

a $d \times d$ variance-covariance matrix $\boldsymbol{\Sigma}$, define the square root of the main diagonal elements matrix $\boldsymbol{\omega} = (diag(\boldsymbol{\Sigma}))^{1/2}$ such that $\bar{\boldsymbol{\Sigma}} = \boldsymbol{\omega^{-1}\Sigma\omega^{-1}}$ is the associated correlation matrix. A d-dimensional random vector $\boldsymbol{Y}$ has a skew normal distribution, denoted by $\boldsymbol{Y} \sim \mathrm{SN}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\alpha})$ if it is continuous with density function of the type

$$f(\boldsymbol{y}) = 2\phi(\boldsymbol{y}, \boldsymbol{\mu}, \boldsymbol{\Sigma})\Phi\left(\boldsymbol{\alpha}^{\mathrm{T}}\omega^{-1}(\boldsymbol{y} - \boldsymbol{\mu})\right), \tag{5}$$

referring to $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ and $\boldsymbol{\alpha}$ as the location, scale and skewness parameters, respectively. Here $\phi(\boldsymbol{y}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the density function of a d-dimensional normal variate with mean $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$, and $\Phi\left(\boldsymbol{\alpha}^{\mathrm{T}}\omega^{-1}(\boldsymbol{y} - \boldsymbol{\mu})\right)$ represents the standard univariate normal distribution function. The construction approach of our skew normal distribution is closely linked to the elliptical family of distributions, as a consequence, the distribution in (5) shares various properties of the elliptical family. Among these, particularly two are of major interest to us, namely, the properties related to moments and affine transformations given in Property 2.

Property 2. (a) Moments: If $\boldsymbol{Y} \sim \mathrm{SN}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\alpha})$, $E(\boldsymbol{Y}) = \boldsymbol{\mu} + \omega\boldsymbol{\mu_S}$ and $\mathrm{Cov}(\boldsymbol{Y}) = \boldsymbol{\Sigma} - \omega\boldsymbol{\mu_S}\boldsymbol{\mu_S}^T\omega$, where $\boldsymbol{\mu_S} = b\delta$ is the expectation of standard skew normal variate (i.e. $\boldsymbol{S} \sim \mathrm{SN_d}(\boldsymbol{0}, \bar{\boldsymbol{\Sigma}}, \boldsymbol{\alpha})$) with a correlation matrix $\bar{\boldsymbol{\Sigma}}$ such that $\boldsymbol{\Sigma} = \omega\bar{\boldsymbol{\Sigma}}\omega$, with b and δ defined as

$$\begin{aligned} b &= \sqrt{2/\pi} \\ \delta &= (1 + \boldsymbol{\alpha}^T\bar{\boldsymbol{\Sigma}}\boldsymbol{\alpha})^{-1/2}\bar{\boldsymbol{\Sigma}}\boldsymbol{\alpha}. \end{aligned}$$

(b) Affine transformation: If $\boldsymbol{Y} \sim \mathrm{SN}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\alpha})$, given $\boldsymbol{Z} = c + A\boldsymbol{Y}$, with $\ell \times d$ matrix A of rank $\ell \leq d$ and $c \in \mathbb{R}^r$, we have $\boldsymbol{Z} \sim \mathrm{SN}_\ell(\boldsymbol{\mu_Z}, \boldsymbol{\Sigma_Z}, \boldsymbol{\alpha_Z})$ where

$$\begin{aligned} \boldsymbol{\mu_Z} &= c + A\boldsymbol{\mu}, \\ \boldsymbol{\Sigma_Z} &= A\boldsymbol{\Sigma}A', \\ \boldsymbol{\alpha_Z} &= (1 - \delta^T\omega A^T\boldsymbol{\Sigma_Z}^{-1}A\omega\delta)^{-1/2}\omega_{\mathbf{Z}}\boldsymbol{\Sigma_Z}^{-1}A\omega\delta. \end{aligned} \tag{6}$$

We adopt the skew normal and t-distribution defined by Azzalini and Capitanio (2003) based on two considerations, (i) the closed-form expression of the multidimensional density is available and straightforward for likelihood inference; (ii) the skew t-distribution is a special case of skew elliptical densities which possesses the properties of the distributions from the elliptical family, particularly the closure under affine transformations that allow us to project the multidimensional distribution to the real line. Next we provide the density of the skew t-distribution. We say that $\boldsymbol{Y}$ follows a skew t-distribution, denoted by $\boldsymbol{Y} \sim \mathrm{St}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}, \nu)$, if $\boldsymbol{Y}$ is a d-variate skew t-distribution with location and scale parameter $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ respectively, defined as above; ν is the number of degrees of freedom and $\boldsymbol{\alpha} \in \boldsymbol{R}^d$ is the shape or skewness parameter. When

$\boldsymbol{\alpha} = \mathbf{0}$, the $t_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ density is obtained again. The density function of $\boldsymbol{Y} \sim \mathrm{St}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}, \nu)$ is

$$f(\boldsymbol{y}) = 2t_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) T_1\left(\boldsymbol{\alpha}^{\mathrm{T}} \omega^{-1}(\boldsymbol{y} - \boldsymbol{\mu})(\frac{\nu + d}{Q_{\boldsymbol{y}} + \nu})^{1/2}; \nu + d\right), \tag{7}$$

where $t_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ is the density function of a $d-$dimensional t-variate with ν degrees of freedom, $T_1(y; \nu + d)$ represents the standard univariate t-distribution function with $\nu + d$ degrees of freedom, and ω, the diagonal matrix formed by standard deviations within $\boldsymbol{\Sigma}$, $Q_{\boldsymbol{y}}$ are defined as

$$\begin{aligned} \omega &= \mathrm{diag}(\boldsymbol{\Sigma})^{1/2}, \\ Q_{\boldsymbol{y}} &= (\boldsymbol{y} - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\boldsymbol{y} - \boldsymbol{\mu}). \end{aligned}$$

Fig. 3 gives a graphical illustration of the skew t-distribution for univariate density in the upper two panels and contour plots in the lower panels. The above two panels show how α and ν may change the density function of the skew t-distribution. When $\alpha = 0$, we are back in the symmetric situation, i.e. the density for the standard t-distribution; and $\mathrm{St}_1(y; \nu)$ will converge to a skew normal distribution as ν goes to infinity. Panel (c) shows that when the degrees of freedom parameter ν increases, the contours of the skew t-distribution converge to those of the skew normal distribution with the same skewness and correlation. In panel (d), we see how the correlation ρ and the skewness vector $\boldsymbol{\alpha}$ alter the appearance of the skew t contours.

[Figure 3 about here.]

Analogous to Property 2 for the skew normal distribution, we have corresponding properties for the skew t-distribution, summarized in Property 3.

Property 3. (a) Moments: If $\mathbf{Y} \sim \mathrm{St}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}, \nu)$, $\mathrm{E}(\mathbf{Y}) = \boldsymbol{\mu} + \omega \boldsymbol{\mu}_{\mathbf{S}}$ for $\nu > 1$ and $\mathrm{Cov}(\mathbf{Y}) = \frac{\nu}{\nu-2}\boldsymbol{\Sigma} - \omega \boldsymbol{\mu}_{\mathbf{S}} \boldsymbol{\mu}_{\mathbf{S}}^T \omega$ for $\nu > 2$, where $\boldsymbol{\mu}_{\mathbf{S}} = b_\nu \delta$ is the expectation of standard skew t-variate (i.e. $\boldsymbol{S} \sim \mathrm{St}_d(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{S}}, \boldsymbol{\alpha}, \nu)$ with the correlation matrix $\boldsymbol{\Sigma}_{\boldsymbol{S}}$, hence $\boldsymbol{\Sigma} = \omega \boldsymbol{\Sigma}_{\boldsymbol{S}} \omega$), with δ defined as in (6) and

$$b_\nu = \left(\frac{\nu}{\pi}\right)^{1/2} \frac{\Gamma\{\frac{1}{2}(\nu - 1)\}}{\Gamma(\frac{1}{2}\nu)}, \quad \text{for } \nu > 1. \tag{8}$$

(b) Affine transformation: If $\mathbf{Y} \sim \mathrm{St}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}, \nu)$, given $\mathbf{Z} = c + A\mathbf{Y}$, with $\ell \times d$ matrix D of rank $\ell \leq d$ and $c \in \mathbb{R}^r$, we have $\mathbf{Z} \sim \mathrm{St}_\ell(\boldsymbol{\mu}_{\mathbf{Z}}, \boldsymbol{\Sigma}_{\mathbf{Z}}, \boldsymbol{\alpha}_{\mathbf{Z}}, \nu)$, where $\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}$ are the same as in Eq. (6).

Note that when $d = 1$ in Eqs (5) and (7), we obtain the special cases of the univariate skew normal distribution $\mathrm{SN}(\mu, \Sigma, \alpha)$ and skew t-distribution $\mathrm{St}(\mu, \Sigma, \alpha, \nu)$, respectively, with $\delta = \alpha/\sqrt{1 + \alpha^2}$ and $\omega = \sqrt{\Sigma}$.

3 Monte Carlo Simulations

In this section we will examine some finite-sample properties of our test of equal predictive ability. In the first two simulation experiments we illustrate the fact that the skew t assumption is more flexible in the sense that it incorporates skewness and fat tails simultaneously, even when the two competing density forecasts are both misspecified. Since we are more interested in comparing density forecasts in tails, we adopt a simple weight function $w(Y) = I(Y \leq r)$. To make the rejection rates obtained for different values of r more comparable, we determine the sample size $n = c/P(Y \leq r)$ based on the threshold value in such a way that the expected number of observations in the region of interest is fixed at c across all selected values of r. Besides, we standardize all forecast distributions to the same mean (zero) and variance (one) for a fair comparison. Appendix A.3 presents the standardized skew Normal and t-distributions.

To be realistic, we limit ourselves to small values of c. Fig. 4 shows the observed rejection frequency at the 5% significance level for $c = 20$, based on 10,000 replications. The data are drawn from the standardized skew $t(5)$-distribution for the upper panels and the standardized skew normal in the lower panels. In Fig. 4, the null hypothesis is equal predictive accuracy of skew $t(5)$ with 5 degrees of freedom and skewness parameter $\alpha = -2$ and skew normal with the same α. Thus the left (right) column reports rejection rates against better prediction of skew t density (skew normal) as a function of the threshold value r.

[Figure 4 about here.]

In a similar manner, Fig. 5 presents rejection rates of equal predictive accuracy against better performance of the skew $t(5)$ distribution with $\alpha = -2$ (left column) and the $t(5)$ distribution (right column) based on 10,000 replications, when for the upper (lower) two panels data are generated from the skew $t(5)$ (standardized $t(5)$) distribution, respectively.

[Figure 5 about here.]

Two conclusions can be drawn from Figs. 4 and 5. First, in both experiments the three scoring rules all give satisfactory powers, and the power of *pwl* is close to that of *csl* all the time. They both outperform the *cl* rule when the threshold r takes very negative values, which is not strange as the *cl* rule does not take into account the probability of the region of interest, as noted by Diks et al. (2011). As $r \to \infty$, the three scoring rules behave even more similarly.

Second, it is quite straightforward that if the number of degrees of freedom ν increases, the power curve in Fig. 4 will decrease as $St(\nu)$ converges to SN in the limit; similarly, when $\alpha \to 0$, $St(5)$ will become indistinguishable from $t(5)$. This is observed in simulations (not included in this paper due to space considerations).

Next, we illustrate the idea of density comparison under a linear projection of a density from higher-dimensional space into one-dimensional space, in the framework of the skew elliptical distributions. Fig. 6 shows the rejection rates of test on equal predictive accuracy of $\mathrm{St}_4(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}, \nu)$ and skew normal $\mathrm{SN}_4(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\alpha})$ and their corresponding univariate projections with the equal-weight mapping vector $(1/4, 1/4, 1/4, 1/4)$ over the whole support, where $\nu = 5$, $\boldsymbol{\alpha} = (-2, 1, -1, -1)$. We assume a strong correlation $\rho_{ij} = 0.7$ with $i = 1, ..., 4$ and $j \neq i$. We standardize the distribution in such a way that the data are centered around the origin and each marginal distribution has unit variance. After the affine transformation (projection), the data are standardized again to have mean zero and variance one. The sample size n takes values 100, 200, 300, 500, 800, 1,000, 2,000, 3,000, 5,000 and 8,000.

[Figure 6 about here.]

The first and last panels in Fig. 6 clearly show that the null hypothesis of equal predictive ability is rejected. In both panels, it can be observed that the test based on the univariate density is less powerful than the test based on the four-dimensional density forecasts as the dashed power curve lies below the real line, especially for small sample sizes. However the power of the test based on the univariate density increases rapidly with the sample size. As seen in the two off-diagonal panels, neither test has spurious power.

Fig. 6 suggests that the test based on the univariate scoring rule after projection to the portfolio linear combination is as reliable as the test based on the multivariate distribution, although the former suffers from lack of power for small samples, which is caused by information loss during the affine transformation. The relative lack of power, however, is not necessarily a bad thing. It only means that the two competing densities after projection are not distinguishable. The high rejection rates seen from the multivariate scoring rule might be just 'spurious' power in the sense that the strongly favored density forecast in the high-dimensional space may not be a better forecast for the portfolio density. Recall that for Example 2 we obtained contradicting test results based on the univariate scoring rule and the bivariate scoring rule.

A final comment concerns the number of degrees of freedom parameter ν. Although not presented in detail in this paper, simulations show that when $\nu = 20$ the power of the test deteriorates; since the two competing densities become very similar, the power based on the univariate densities drops, and spurious power can be observed in the off-diagonal panels.

4 Empirical Application

In this section, we present the comparison of out-of-sample density forecasts for a daily portfolio consisting of three asset returns using the parametric univariate and multivariate model of

the GARCH family. The comparison is twofold: a comparison between using forecasted multivariate densities versus their univariate projections, and a comparison between the univariate projections and directly forecasted portfolio densities.

We consider daily data for three US stocks (Source: Yahoo Finance): the Alcoa Stock (AA), the McDonald's stock (MCD) and the Merck stock (MRK) over the period from April 2, 1984 until May 14, 2015. For these stock prices, daily log-returns defined as $y_t = \log(P_t/P_{t-1})$ are examined, with the adjusted closing price P_t at day t.[1] This yields a total of $7,845$ valid observations. A weighted portfolio with fixed weights allocated for 70% to AA, 20% to MCD and 10% to MRK is studied.[2]

Our work is similar to that of Giot and Laurent (2003) in the following respects: first, both papers look at the same portfolio of daily asset returns with fixed weights. Secondly, they also provide a comparison of the performance of the assumed distributions that the innovation terms follow; there normal, standardized t and skew t are considered, although the expression of the skew t-distribution differs from ours. Thirdly, a 'stability window' of 50 days is applied there for parameter updating, which looks similar to our moving window scheme at first glance. However, a moving window is used in the present paper to satisfy a condition for the asymptotic distribution of the test statistic to be normal. Besides, their estimation sample is augmented as time moves forward, in contrast to the fixed estimation window of observations used here. More importantly, our paper shifts focus from finding a better VaR measure to testing the more suitable distribution of the given portfolio return. Put differently, this paper confronts a comparison based on multivariate densities with that based on the corresponding univariate projections.

Instead of back-testing the in-sample density, the out-of-sample forecasts are of more practical interest. As Bao et al. (2007) found, the specification of the standardized innovations has a larger impact on density forecast accuracy than the volatility specification. Therefore, to illustrate the idea of comparing density forecasts under projections, we define four forecast methods which differ only in the prior assumption of the family of distributions from which the standardized innovations are drawn, namely Standard normal, t, skew normal and skew t. Parametric GARCH$(1,1)$ models, jointly with DCC$(1,1)$ updating of the correlations in the multivariate case, are used to estimate based on past returns and make one-step-ahead

[1] Strictly speaking the portfolio return is a linear combination of individual stock returns only for simple returns $(P_t - P_{t-1})/P_{t-1}$. However, for daily returns the differences between the two types of returns are small, and since the return distributions proposed in the literature commonly are defined for (joint) log-returns between $-\infty$ to ∞, it is more natural to work with log-returns in the applications.

[2] We also performed the same procedure on two other portfolios, namely with the equal allocation weights $(1/3, 1/3, 1/3)$ and the reverse weights on AA and MRK, $(0.1, 0.2, 0.7)$, than the that discussed in the main text. It turns out that the choice of portfolio weight may lead to slightly different evaluation results, which confirms our initial conjecture that the projection vector plays a role when we generate the univariate portfolio density from the combination of multivariate marginals.

forecasts. The log-returns series are treated as having conditional mean zero, and conditional variance-covariance matrix, following Engle (2002), given by

$$\boldsymbol{Y}_t|\mathscr{F}_{t-1} \sim \text{Dist.}(\boldsymbol{0}, \boldsymbol{H_t}).$$

The covariance matrix $\boldsymbol{H_t}$ can be decomposed as

$$\boldsymbol{H}_t = \boldsymbol{D}_t \boldsymbol{R}_t \boldsymbol{D}_t,$$

where $\boldsymbol{D}_t$ is the $d \times d$ diagonal matrix of time-dependent standard deviations from the univariate GARCH(1,1) models with $\sqrt{h_{i,t}}$ the $i-th$ diagonal element and $\boldsymbol{R}_t$ is the estimated time-varying correlation matrix. The evolution of individual conditional variances $h_{i,t}$, $i = 1, \ldots, d$, and the correlations is determined by

$$\begin{aligned} h_{i,t} &= \omega_i + \alpha_i \varepsilon_{i,t-1}^2 + \beta_i h_{i,t-1}, && (9) \\ \boldsymbol{Q}_t &= (1-a-b)\bar{\boldsymbol{Q}} + a\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}' + b\boldsymbol{Q}_{t-1}, && (10) \\ \boldsymbol{R}_t &= \boldsymbol{Q}_t^{*-1}\boldsymbol{Q}_t\boldsymbol{Q}_t^{*-1}, && (11) \end{aligned}$$

where $\boldsymbol{\varepsilon}_t = (\varepsilon_{1,t}, ..., \varepsilon_{d,t})'$ is the vector of the standardized residuals $\varepsilon_{i,t} = y_{i,t}/\sqrt{h_{i,t}}$ and where $\bar{\boldsymbol{Q}} = \mathbb{E}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t')$ represent the conditional correlation matrix. Last, $\boldsymbol{Q}^*$ is a diagonal matrix with the square root of the $i-th$ diagonal element of $\boldsymbol{Q}$ on its $i-th$ diagonal position. Once each individual GARCH$(1,1)$ model (9) has been estimated, $\boldsymbol{\varepsilon_t}$ is stored to estimate the dynamics of the correlation (10)-(11).

We use a moving-window scheme to predict the one-day-ahead density. The estimation window is set to $m = 1,000$, and the remaining observations are used for out-of-sample evaluation. The Diebold-Mariano test defined in (1) is applied, with scoring rules (2)-(4). As mentioned earlier, the left tail of the distribution of the portfolio return is of interest. Given the portfolio return series $\{r_t\}$ with the fixed individual weights, we use the truncated weight function $w_t(r_t) = I(r_t \leq \hat{r}_t^q)$. The time-varying threshold $\hat{r}_t^q$ is calculated as the empirical qth quantile of all return observations in the most recent estimation window, where $q = 0.01, 0.05, 0.1, 0.15$ are considered.

Table 1 gives the average score differences and the test result for the censored likelihood scoring rule (3). Similar results from the conditional likelihood rule (2) and penalized weighted likelihood rule (4) are shown in Table 2 and Table 3. The four sub-panels in Table 1 present results based on different choices for the threshold quantile q. The numbers in parentheses under the score differences are the values of the corresponding Diebold-Mariano test statistics, where those with an asterisk are significant at the 5% significance level. The test statistics

are calculated based on the average score differences $\bar{d}$ with the HAC estimator adjusted for serial dependence. As mentioned before, we want to compare the evaluation of the density forecasts in both high-dimensional space and in the projected univariate space. The label csl_{multi} denotes test results based on trivariate density evaluation, and csl_{proj} uses the projected parameter values for density evaluation in one-dimensional space, where the asset allocation weight $(0.7, 0.2, 0.1)$ plays the role of the projection vector. The test results for csl_{1d} in the last row is obtained using portfolio return series $\{r_t\}$ for estimation and prediction directly.

[Table 1 about here.]

[Table 2 about here.]

[Table 3 about here.]

Table 1 indicates that, firstly, the scores of the t and skew t forecasts overwhelm those of the standard Normal and skew Normal in most comparisons, which is not surprising, as stock returns are often characterized by fat tails. Secondly, the order of priority among the other two pairs is hard to assess. The signs of the score differences suggest that the standard Normal and t forecasts outperform the skew Normal and skew t forecasts. However, significant test statistics are rarely seen.

Thirdly, test results from csl_{proj} do not always lead to the same conclusion as csl_{multi}. For example, for $q = 0.05$, csl_{multi} suggests that the t forecast is significantly better than the skew t forecast, while only insignificant score difference are witnessed by csl_{proj}. The divergence is even larger for the comparison between standard Normal and skew Normal for $q = 0.05$, where csl_{multi} favors the skew Normal since it delivers negative score difference; on the contrary, csl_{proj} prefers the standard Normal. The contradictive results from different dimensional spaces illustrates the idea of Example 1 and 2 again; the better forecasts for the multivariate distribution may not yield better forecasts for the univariate projection. Therefore, from the portfolio risk assessment point of view, it seems better to consider with the univariate portfolio returns rather than the underlying multivariate returns.

Fourthly, when comparing the test results of csl_{proj} and csl_{uni}, only two differences are observed. From the pair comparison t-SN at $q = 1\%$ and t-St at $q = 15\%$, both scoring rules favor the t forecasts. However csl_{proj} delivers a sharp conclusion at the 5% level, while csl_{uni} shows less significant results. Apart from that, the results from csl_{proj} and csl_{uni} are highly in agreement. This is numerical support for univariate rather than multivariate modelling, avoiding excessive computations for multivariate density forecast evaluation, without losing crucial information.

Further investigation of the projected densities and univariate densities is carried out by direct pairwise comparison of the corresponding univariate density forecast accuracies. Table 4 reports

the average score difference and the associated test statistics for the univariate model-based densities against their rivaling projected densities, based on scoring rules (2)-(4). From there, we see very robust result, especially for the *csl* and *pwl* rules. The t and skew t forecasts strongly outperform those of the normal and skew normal, regardless of whether these forecasts are generated using univariate modeling or projected multivariate modeling. Besides, all diagonal statistics in each sub-panel are insignificant, which indicates again that the test results based on the projected densities are in line with the results from testing univariate densities.

[Table 4 about here.]

4.1 Dynamic selection of forecast distribution families

So far, the results provided in Tables 1-4 made use of the full sample, yielding the conclusion that the skew t and t forecasts are best over the entire sample. However, assuming a fixed family of distributions for stock returns across the entire historical sample is not very realistic, especially for a sample over the last thirty years, which witnessed a number of severe crises. A more practically relevant question is, what is the underlying distribution of the portfolio returns within shorter sub-periods? The remaining part of this section will seek to provide insights into the time-varying nature of the conditional densities by using the test statistic dynamically within sub-samples.

We apply an iterative procedure on the univariate daily return series, namely, the portfolio consisting of 70% investment in an aluminum producer AA, 20% in service industry for MCD and the last 10% in medicine manufacturer MRK. Each sub-sample consists of an estimation window of length $m = 1,000$, the same as before, and an out-of-sample evaluation window. The length of the evaluation window is also fixed at $1,000$, which corresponds to about four years. At each iteration, the univariate GARCH(1,1) model is estimated using the portfolio return data available in the sample, and a one-step-ahead forecast is calculated and stored for later evaluation. As the estimation window rolls forward until all data except the last in the evaluation sample are consumed, we calculate the test statistic. For example, in Figs 7-9 the first estimation window and the first evaluation window are denoted by a light blue rectangle (labeled I) and a cyan rectangle (labeled II), respectively. The iteration procedure takes steps of 25 days, which means that after every sub-sample of length $2,000$ the model parameters are estimated and the forecasts evaluated; this procedure repeats itself every 25 trading days. In total 234 sequential test statistics are calculated across time, and plotted.

[Figure 7 about here.]

Fig. 7 reports the dynamics of the test statistics for six pair-wise comparisons, based on scoring rules (2)-(4) at empirical quantile level $q = 0.05$ and Fig. 8 compares the dynamics

of the test statistics of censored likelihood rule for the same competing densities of different quantile levels. The area between the two black dashed lines represents a (point-wise) 5% forecast interval. Whenever the upper boundary is reached by the test statistic the former distribution from that pair comparison is preferred significantly, and vice versa. Even without significant results, since we have many sequential test statistics, the signs of these test statistics may be indicative of the relative predictive ability under the different distribution assumptions. For example, in the first sub-panel of Fig. 7, the standard Normal distribution is tested against the t-distribution; the statistics almost lie below zero all the time except for a short period in the 1990s, suggesting that the Normal distribution is more suitable.

The dynamical graphs enable us to see which of the two competing innovation distributions is more appropriate during different periods. Figs 7 and 8 show rather consistent results across different scoring rules and different quantiles qs. Some are in accord with our findings in Table 1. For example, t innovations all along outperform standard Normal and skew Normal innovations, and the superiority of the standard Normal or skew Normal densities varies mostly within the significance bands. Interesting dynamics can be observed for the other three pairs, where more violent patterns occur as a sign of regime switching from time to time. Another eye-catching phenomenon is the relatively stable regime of the statistics in the late 90s.

[Figure 8 about here.]

In the remaining part of this section we try to provide some insights into these phenomena. We take the S&P500 Index as a barometer of the stock market. The blue line in Fig. 9 corresponds to the S&P price index labeled on the left y axis while the right y axis denotes test statistic values. The red dash-dotted line is the *csl* statistic for equal predictive ability of t and skew t innovations, while the green dashed line corresponds to the test of skew Normal against skew t, both replicated from Fig. 7. The first period of interest runs from November 1997 to October 2001, the test statistics stay quite steady for all pair-comparisons, denoted by the green rectangle (labeled III) in Fig. 9.

[Figure 9 about here.]

To understand this, recall that we use an estimation window of length $1,000$ to generate one-step-ahead forecasts and another $1,000$ day data window to evaluate the two competing densities. Within this framework, the first statistic in the green region is based on a set of out-of-sample density forecasts roughly between December 1993 and October 1997. In other words, almost 8 years of out-of-sample forecasts are used to generate the dynamics inside the green area. Therefore, during the US stock market boom period of 1994–2001 before 9/11, all six pair comparisons of our test show stable, near constant, results, among which t innovations are suggested to perform better than their counterparts for the given portfolio.

Next, the pink rectangle (labeled IV) in Fig. 9 witnesses a striking decline of the *csl* statistics of the t against the skew t from 2.83 in October, 2002 to -4.78 in August, 2009, which starts right after the dot-com bubble burst and ends by the bear market rally during the recent financial crisis. The stock market experienced a wild fluctuation in the first decade of this century, reflected by frequent ups and downs of S&P500 Index. In this period, t innovations give way to the skew t assumption gradually, suggesting that the symmetric t-distribution is unable to model drastic negative shocks in a volatile market.

In this period we also observe an increase in the test statistics of the skew Normal against skew t before it drops to the negative significance boundary. A potential reason for the superiority of skew Normal innovations at the beginning could be the fact that the rolling estimation window does not incorporate enough negative return observations up to April 2000, when the market bubble did not burst yet. As the iterative procedure moves forward, more and more extreme shocks are included for estimation, giving rise to better model estimation and prediction using skew t innovations.

5 Value-at-Risk(VaR) Forecasts and Evaluation

This section illustrates how time-varying score averages can be used in an application of dynamic one-step-ahead VaR forecasting in the context of both the univariate portfolio modeling and the multivariate modeling, based on different distributional choices. With the findings from the last subsection that the relative performances of four density assumptions are time-varying, apart from generating the VaR forecasts from the four density candidates, we may use the scoring rules (2)-(4) to daily select the recently 'best' performing distribution to base the out-of-sample VaR forecast on.

Intuitively, for a fixed estimation window up to and including time t, we estimate a parametric GARCH model. With the observed returns we can compute the scores based on (2)-(4) and select the density with the highest score for one-day-ahead VaR calculation.

Under the rolling window scheme, the parametric GARCH(1,1) and GARCH(1,1)-DCC(1,1) model are estimated and updated for every trading day for the univariate modeling and multivariate modeling, respectively. In the next step, the one-step-ahead VaR forecasts are generated under different distributional assumptions. Apart from the four fixed densities, the dynamically selected density by both the univariate and trivariate scoring rules for three different modeling methods, i.e. univariate modeling, multivaiate modeling and the projection of the latter, are also used in forecasting the VaR. The estimation window is set equal to $m = 1,000$, yielding 6844 out-of-sample VaR forecasts.

Conceptually, the VaR is nothing but a time-varying conditional quantile, and our goal is to

provide the evolution of the quantile forecast over time. In this sense, our technique and the CAViaR model proposed by Engle and Manganelli (2004) are alike, though we seek to select the highest scoring density for quantile forecasting indirectly while a direct quantile regression is performed by Engle and Manganelli (2004). One may also think of this adaptive selection technique as an analogue of the regime switching model proposed by Pelletier (2006) and Garcia and Tsafack (2011) for time-varying dependence structure. Chollete et al. (2009) constructed a multivariate regime switching copula model for VaR forecasting. This section is similar to their model not only in dynamically updating conditional dependence for outperforming fixed distributions in VaR forecasts, but also in the two-step estimation methodology for multivariate modeling. However, our technique is more parsimonious than regime switching models since no latent variable or transition probabilities are involved. Another advantage of our technique is that more than two 'regimes' can be easily accommodated with scoring rules, whereas regime switching models typically assume two regimes, e.g. a symmetric and asymmetric regime for normal and extreme situations, respectively. It is also noteworthy that our methodology of density selection is in line with Opschoor et al. (2014), where a technique of combining densities based on scoring rules is applied to obtain VaR forecasts.

To test the accuracy of the VaR forecasts from different distributions, namely four fixed families of densities N, t, SN, St and the dynamically selected densities based on (2)-(4), we first compute the coverage rates for 1%, 5%, 10% and 15% VaR forecasts, e.g. the percentages of the observations such that $y_{t+1} \leq \mathrm{VaR}_{t+1}$, then consider three different types of tests. The first two tests only keep track on the statistical accuracy of the VaR values by checking the violation frequency of the VaR forecasts with the correct unconditional coverage (CUC) of Kupiec (1995) and the correct conditional coverage (CCC) of Christoffersen (1998).

In addition, we compare two competing VaR forecasting models in the style of Diebold and Mariano (1995) by using a loss function which measures the distance between the observed returns and forecasted VaR values. The linear asymmetric linear loss function, defined by Giacomini and Komunjer (2005), is given by

$$L^q(e_{t+1}) = (q - I(e_{t+1} < 0))e_{t+1}, \tag{12}$$

where $e_{t+1} = y_{t+1} - \mathrm{VaR}^q_{t+1}$ and $q \in \{0.01, 0.05, 0.1, 0.15\}$ is the quantile considered. Let $d_{AB} = L^q_A - L^q_B$, where L^q_A and L^q_B are the loss functions for distributions A and B, respectively. A negative value of d_{AB} indicates that the forecast based on A is superior to that on B.

A comparison is made for the univariate modeling of the portfolio return directly, the projected multivariate modeling with the allocation vector, and the multivariate modeling over the three individual stock returns. The results, reported in Table 5, summarize the coverage

rates, the p-values for the CUC test, the p-values for the CCC test and the t-statistic of the Diebold-Mariano test for the equal conditional predictive ability of two VaR models. In all three modeling frameworks, we choose the VaR based on the *csl* rule selected density as the benchmark, and compare the other six models against it. A negative statistic suggests that the VaR_{t+1}^{csl} is preferred and vice versa.

[Table 5 about here.]

Several conclusions can be drawn from Table 5. First, the empirical VaR violation rates using the dynamically selected density based on *pwl* and *csl* rules are closer to the nominal frequencies than the VaR forecasts based on the fixed families of distributions for the 5%, 10% and 15% quantiles, independent of the modeling method. However this advantage is not seen from the *cl* rule. For small quantiles the *pwl* and *csl* rule-based VaR forecasts deliver identical results, which is consistent with the finding before that the performance of the two scoring rules is very similar.

Second, the CUC and the CCC tests reject the null hypothesis of a correct coverage rate for the fixed density based VaR more than the dynamically selected density-based VaR, indicating again that it is more reasonable to assume that the underlying portfolio return distribution is time-varying rather than fixed. Among the four distributional assumptions, the t and skew t-distributions perform as good as, if not better than, the other two distributions in the sense that the null hypothesis is rejected less often, which is not surprising given the fact that t and skew t forecasts in general gain a higher score in the density forecast comparisons than the Normal and skew Normal forecasts.

Third, the last column for each modeling method reports the t-value for the Diebold-Mariano test. It is suggested that the *csl* rule based VaR forecast consistently outperforms the fixed distribution based VaR forecasts since the t-statistics are negative almost everywhere. The Normal density and the skew Normal density based VaR estimates are significantly dominated by the benchmark for univariate modeling, projected modeling and direct multivariate modeling for quantiles 5%, 10% and 15%. Regarding the comparison among different dynamic-selected densities, the conclusion is less sharp. The *pwl* rule based VaR forecasts tend to be superior to the *csl* rule based ones in the univariate modeling, but not in the projected or multivariate models.

Finally, we compare the VaR forecast results across the different modeling methods in Tables 6 and 7 by using the loss function (12) again. Table 6 reports results of the Diebold-Mariano test for univariate modeling and projected modeling against the multivariate modeling with four fixed densities and three dynamic-selected densities. A negative test statistic suggests that the VaR forecasts from multivariate modeling are better. Similarly, Table 7 presents the results for

tests on the equal VaR accuracy between the univariate modeling or projected modeling and projected modeling with different density assumptions.

[Table 6 about here.]

[Table 7 about here.]

From Table 6 it can be seen that multivariate modeling does not improve the VaR accuracy. Although the dynamic-selection of densities based on trivariate scoring rules (2)-(4) end up with negative t-statistics almost all the time, especially when compared with the VaR estimates generated by fixed univariate densities, the test of equal predictive accuracy of the trivariate scoring rule-based VaR and univariate scoring rules-based VaR is not conclusive. Besides, the diagonal statistics in each quantile sub-panel for univariate modeling seem to suggest that the VaR based on the parsimonious GARCH(1,1) model is at least as good as the forecasts from the involved DCC(1,1)-GARCH(1,1) model when comparing identical distribution families with different dimensions.

To summarize, with the foresight of the time-varying distribution for the given portfolio, we may come up with a simple selection method for the underlying distribution to be assumed for the VaR estimation based on scoring rules (2)-(4), evaluated for both univariate densities and multivariate densities. It turns out that this dynamically-selected distribution based VaR forecast outperforms the fixed distributions. Consistent results are observed. For example, the Normal density and the skew Normal density perform worse, and the multivariate modeling method does not deliver better VaR forecasts than the univariate modeling method.

6 Summary and Conclusions

To evaluate the risk of a given portfolio of assets, one can evaluate and select competing density forecasts for the relevant future random variables. The traditional way is to focus on forecasting the multivariate distribution of the composing assets first, and then assess the downside risk of the portfolio. However, as we showed in a simple example, information in high-dimensional space can be irrelevant or even misleading for portfolio risk measurement.

In this paper, we developed a testing framework for comparing the accuracy of competing density forecasts in a selected part of the support across different dimensions. Based on the Kullback-Leibler information criterion, three proper scoring rules including conditional likelihood, censored likelihood and penalized weighted likelihood were used for evaluating the predictive ability of (out-of-sample) densities. We focused on distributions from skew elliptical families since these distributions are closed under affine transformations. The merits are that we may explicitly trace the distribution function of the aggregate portfolio return. Monte

Carlo simulations indicate that the test in higher-dimensional space could deliver a similar or opposite conclusion as the univariate test, depending on the shape and location of the projected densities.

Our empirical application to daily return series of a fixed portfolio consisting of three S&P500 stocks indicates that the t-distribution outperforms the Normal, Skew t and Skew Normal distributions in the left tail of the support over a thirty years time horizon. The parametric GARCH(1,1) model, jointly with DCC updating for correlation in the multivariate setting, is used to successively generate one-step-ahead density forecasts. Given the fact that the dependence structure may vary with respect to sign and/or magnitude, we iteratively perform our test on portfolio return forecasts in a moving sub-sample to generate the dynamics of test statistics. The changes in the test statistics are visualized, from which it can be observed that the t-distribution overwhelms during the booming period, while the skew t-distribution is more appropriate in a declining market. The dynamical results show that the optimal family of distributions changes more than once in the sample period.

To further investigate the time-varying nature of the conditional distributions, a natural extension to the univariate and multivariate VaR estimation is proposed; a scoring rule-based technique to choose the distribution for VaR forecasts is applied, allowing dynamically selected forecast distributions for VaR forecasts. The empirical VaR forecasts based on this approach show improvements, confirming the time-varying nature of the underlying distribution.

The method discussed in this paper can also be used to compare multi-step ahead density forecasts, or to compare non-parametrically estimated densities. It may be possible to improve the forecasts by considering more general distributions, but this is left for future research. Besides, given the potential dynamics of the underlying distribution, more advanced techniques could be developed for density forecasting and VaR estimation. Lastly, dynamic portfolio optimization driven by time-varying density forecasts form another topic for further research.

References

Alexander C. 2009. *Market Risk Analysis, Value at Risk Models*, volume 4. John Wiley & Sons.

Amisano G, Giacomini R. 2007. Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business & Economic Statistics* **25**: 177–190. DOI: 10.1198/073500106000000332.

Arellano-Valle RB, Bolfarine H. 1995. On some characterizations of the t-distribution. *Statistics & Probability Letters* **25**: 79–85. ISSN 0167-7152. DOI: 10.1016/0167-7152(94)00208-p.

Azzalini A. 2005. The skew-normal distribution and related multivariate families. *Scandinavian Journal of Statistics* **32**: 159–188. DOI: 10.1111/j.1467-9469.2005.00426.x.

Azzalini A. 2013. *The skew-normal and related families*, volume 3. Cambridge University Press. DOI: 10.1017/cbo9781139248891.

Azzalini A, Capitanio A. 2003. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **65**: 367–389. DOI: 10.1111/1467-9868.00391.

Azzalini A, Dalla Valle A. 1996. The multivariate skew-normal distribution. *Biometrika* **83**: 715–726. DOI: 10.1017/cbo9781139248891.006.

Bao Y, Lee TH, Saltoğlu B. 2004. A test for density forecast comparison with applications to risk management. *Department of Economics, UC Riverside* .

Bao Y, Lee TH, Saltoğlu B. 2007. Comparing density forecast models. *Journal of Forecasting* **26**: 203–225. DOI: 10.1002/for.1023.

Bauwens L, Laurent S. 2005. A new class of multivariate skew densities, with application to generalized autoregressive conditional heteroscedasticity models. *Journal of Business & Economic Statistics* **23**: 346–354. DOI: 10.1198/073500104000000523.

Bauwens L, Laurent S, Rombouts JV. 2006. Multivariate garch models: a survey. *Journal of applied econometrics* **21**: 79–109. DOI: 10.2139/ssrn.411062.

Boubaker H, Sghaier N. 2013. Portfolio optimization in the presence of dependent financial returns with long memory: A copula based approach. *Journal of Banking & Finance* **37**: 361–377. ISSN 0378-4266. DOI: 10.1016/j.jbankfin.2012.09.006.

Branco MD, Dey DK. 2001. A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis* **79**: 99–113. DOI: 10.1006/jmva.2000.1960.

Brooks C, Burke SP, Heravi S, Persand G. 2005. Autoregressive conditional kurtosis. *Journal of Financial Econometrics* **3**: 399–421. DOI: 10.1093/jjfinec/nbi018.

Cambanis S, Huang S, Simons G. 1981. On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis* **11**: 368–385. ISSN 0047-259X. DOI: 10.1016/0047-259x(81)90082-8.

Casella G, Berger RL. 2002. *Statistical inference*, volume 2. Duxbury Pacific Grove, CA. DOI: 10.2307/1269442.

Chamberlain G. 1983. A characterization of the distributions that imply mean-variance utility functions. *Journal of Economic Theory* **29**: 185–201. ISSN 0022-0531. DOI: 10.1016/0022-0531(83)90129-1.

Chollete L, Heinen A, Valdesogo A. 2009. Modeling international financial returns with a multivariate regime-switching copula. *Journal of Financial Econometrics* : nbp014DOI: 10.2139/ssrn.1102632.

Christoffersen PF. 1998. Evaluating interval forecasts. *International Economic Review* **39**: 841–862. DOI: 10.2307/2527341.

Corradi V, Swanson NR. 2005. A test for comparing multiple misspecified conditional interval models. *Econometric Theory* **21**: 991–1016. DOI: 10.1017/s0266466605050498.

Diebold FX, Lopez JA. 1996. Forecast evaluation and combination. In *Handbook of Statistics*. DOI: 10.3386/t0192.

Diebold FX, Mariano RS. 1995. Comparing predictive accuracy. *Journal of Business & Economic Statistics* **13**: 253–263. DOI: 10.3386/t0169.

Diks C, Panchenko V, Sokolinskiy O, van Dijk D. 2014. Comparing the accuracy of multivariate density forecasts in selected regions of the copula support. *Journal of Economic Dynamics and Control* **48**: 79–94. ISSN 0165-1889. DOI: 10.1016/j.jedc.2014.08.021.

Diks C, Panchenko V, Van Dijk D. 2010. Out-of-sample comparison of copula specifications in multivariate density forecasts. *Journal of Economic Dynamics and Control* **34**: 1596–1609. DOI: 10.2139/ssrn.1294250.

Diks C, Panchenko V, van Dijk D. 2011. Likelihood-based scoring rules for comparing density forecasts in tails. *Journal of Econometrics* **163**: 215–230. ISSN 0304-4076. DOI: 10.1016/j.jeconom.2011.04.001.

Elton EJ, Gruber MJ. 1997. Modern portfolio theory, 1950 to date. *Journal of Banking & Finance* **21**: 1743–1759. ISSN 0378-4266. DOI: 10.1016/s0378-4266(97)00048-4.

Elton EJ, Gruber MJ, Brown SJ, Goetzmann WN. 2009. *Modern Portfolio Theory and Investment Analysis*. John Wiley & Sons.

Engle R. 2002. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics* **20**: 339–350. DOI: 10.1198/073500102288618487.

Engle RF, Manganelli S. 2004. Caviar: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics* **22**: 367–381. DOI: 10.1198/073500104000000370.

Fang KT, Samuel K, Kai WN. 1990. *Symmetric Multivariate and Related Distributions*. Champman and Hall. DOI: 10.1007/978-1-4899-2937-2.

Garcia R, Tsafack G. 2011. Dependence structure and extreme comovements in international equity and bond markets. *Journal of Banking & Finance* **35**: 1954–1970. DOI: 10.2139/ssrn.891172.

Giacomini R, Komunjer I. 2005. Evaluation and combination of conditional quantile forecasts. *Journal of Business & Economic Statistics* **23**: 416–431. DOI: 10.2139/ssrn.318728.

Giacomini R, White H. 2006. Tests of conditional predictive ability. *Econometrica* **74**: 1545–1578. DOI: 10.1111/j.1468-0262.2006.00718.x.

Giot P, Laurent S. 2003. Value-at-risk for long and short trading positions. *Journal of Applied Econometrics* **18**: 641–663. DOI: 10.1002/jae.710.

Gneiting T, Raftery AE. 2007. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association* **102**: 359–378. DOI: 10.1198/016214506000001437.

Gneiting T, Ranjan R. 2011. Comparing density forecasts using threshold-and quantile-weighted scoring rules. *Journal of Business & Economic Statistics* **29**: 411–422. DOI: 10.1198/jbes.2010.08110.

Granger CW. 2003. Time series concepts for conditional distributions. *Oxford Bulletin of Economics and Statistics* **65**: 689–701. DOI: 10.1046/j.0305-9049.2003.00094.x.

Guidolin M, Timmermann A. 2007. Asset allocation under multivariate regime switching. *Journal of Economic Dynamics and Control* **31**: 3503–3544. DOI: 10.2139/ssrn.940652.

Gupta AK, Varga T, Bodnar T. 2013. *Elliptically Contoured Models in Statistics and Portfolio Theory*. Springer. DOI: 10.1007/978-1-4614-8154-6.

Harvey CR, Siddique A. 1999. Autoregressive conditional skewness. *Journal of financial and quantitative analysis* **34**: 465–487. DOI: 10.2139/ssrn.61332.

Huang CW, Hsu CP, Chiou WJP. 2015. Can time-varying copulas improve the mean-variance portfolio? In *Handbook of Financial Econometrics and Statistics*. Springer New York, 233–251. DOI: 10.1007/978-1-4614-7750-1_8.

Jones M, Faddy M. 2003. A skew extension of the t-distribution, with applications. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **65**: 159–174. DOI: 10.1111/1467-9868.00378.

Kandel S, Stambaugh RF. 1996. On the predictability of stock returns: An asset-allocation perspective. *The Journal of Finance* **51**: 385–424. DOI: 10.3386/w4997.

Kelker D. 1970. Distribution theory of spherical distributions and a location-scale parameter generalization. *The Indian Journal of Statistics, Series A* **32**: 419–430.

Kotz S, Nadarajah S. 2004. *Multivariate t-distributions and their applications*. Cambridge University Press. DOI: 10.1017/cbo9780511550683.

Kupiec PH. 1995. Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives* **3**: 73–84. DOI: 10.3905/jod.1995.407942.

Landsman ZM, Valdez EA. 2003. Tail conditional expectations for elliptical distributions. *North American Actuarial Journal* **7**: 55–71. DOI: 10.1080/10920277.2003.10596118.

Markowitz H. 1952. Portfolio selection. *The Journal of Finance* **7**: 77–91. DOI: 10.2307/2975974.

Matheson JE, Winkler RL. 1976. Scoring rules for continuous probability distributions. *Management Science* **22**: 1087–1096. DOI: 10.1287/mnsc.22.10.1087.

Merton RC. 1992. *Continuous-Time Finance*. Macroeconomics and Finance Series. Wiley-Blackwell. ISBN 9780631185086.

Mitchell J, Hall SG. 2005. Evaluating, comparing and combining density forecasts using the KLIC with an application to the Bank of England and NIESR 'fan' charts of inflation. *Oxford bulletin of economics and statistics* **67**: 995–1033. DOI: 10.1111/j.1468-0084.2005.00149.x.

Opschoor A, Van Dijk DJ, Van der Wel M. 2014. Improving density forecasts and value-at-risk estimates by combining densities. Technical report. Tinbergen Institute Discussion Paper 14-090/III. DOI: 10.2139/ssrn.2469024.

Owen J, Rabinovitch R. 1983. On the class of elliptical distributions and their applications to the theory of portfolio choice. *The Journal of Finance* **38**: 745–752. ISSN 00221082. DOI: 10.2307/2328079.

Patton AJ. 2004. On the out-of-sample importance of skewness and asymmetric dependence for asset allocation. *Journal of Financial Econometrics* **2**: 130–168. DOI: 10.1093/jjfinec/nbh006.

Pelenis J. 2014. Weighted scoring rules for comparison of density forecasts on subsets of interest. Technical report, Institute for Advanced Studies, Vienna.

Pelletier D. 2006. Regime switching for dynamic correlations. *Journal of Econometrics* **131**: 445–473. DOI: 10.1016/j.jeconom.2005.01.013.

Sarno L, Valente G. 2004. Comparing the accuracy of density forecasts from competing models. *Journal of Forecasting* **23**: 541–557. DOI: 10.1002/for.930.

Shaw W, Lee K. 2008. Bivariate student distributions with variable marginal degrees of freedom and independence. *Journal of Multivariate Analysis* **99**: 1276–1287. ISSN 0047-259X. DOI: 10.1016/j.jmva.2007.08.006.

Vuong QH. 1989. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* **57**: 307–333. DOI: 10.2307/1912557.

A Appendix

A.1 Density Function for Multivariate t-distribution

There are many different candidates for the multivariate generalization of Student's t-distribution; Kotz and Nadarajah (2004) provide an extensive survey. The multivariate t-distribution is defined as an extension of the classical univariate t-distribution. For example, the probability density function of a d-dimensional t-distributed vector $\mathbf{Y} = \{Y_1, ..., Y_d\}$ is defined by Kotz and Nadarajah (2004) as

$$f(\mathbf{y}) = \frac{\Gamma(\frac{\nu+d}{2})|\boldsymbol{R}|^{-\frac{1}{2}}}{\Gamma^d(\frac{1}{2})\Gamma(\frac{\nu}{2})\nu^{\frac{d}{2}}} \left(1 + \frac{1}{\nu}(\boldsymbol{y} - \boldsymbol{\mu})^T \boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{\mu})\right)^{-\frac{\nu+d}{2}}, \tag{A.1}$$

with mean vector $\boldsymbol{\mu}$, correlation matrix $\boldsymbol{R}$ and degrees of freedom ν, provided $\boldsymbol{R}$ is positive definite. When $\boldsymbol{\mu} = \mathbf{0}$, we render the standard $d-$dimensional form

$$f(\mathbf{y}) = \frac{\Gamma(\frac{\nu+d}{2})|\boldsymbol{R}|^{-\frac{1}{2}}}{\Gamma^d(\frac{1}{2})\Gamma(\frac{\nu}{2})\nu^{\frac{d}{2}}} \left(1 + \frac{1}{\nu}\boldsymbol{y}^T \boldsymbol{R}^{-1}\boldsymbol{y}\right)^{-\frac{\nu+d}{2}}.$$

If $d = 1$, together with $\boldsymbol{R} = 1$, we have the univariate probability density function

$$f(y) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi\nu}\Gamma(\frac{\nu}{2})} \left(1 + \frac{y^2}{\nu}\right)^{-\frac{\nu+1}{2}}.$$

The generalization of the multivariate t-distribution in Eq. (A.1) indeed uses the elliptical methodology, by which we mean the distribution theory for elliptical distributions (ED) and for the distributions obtained from elliptical distributions by introducing a location vector and a scale matrix, i.e. $\boldsymbol{\mu}$ and $\boldsymbol{R}$ used above. Shaw and Lee (2008) pointed out a difficulty with the standard ED representation of the multivariate t-distribution, which is that the joint distribution in the form of Eq. (A.1), does not factorize into the product of the d marginal one-dimensional densities. However, we stick with the traditional ED representation for the multivariate t-distribution based on the desirable properties of ED family distributions. Indeed, since Chamberlain (1983) and Owen and Rabinovitch (1983), elliptical distributions are widely applied to characterize the distribution of portfolio returns. For example, Landsman and Valdez (2003) decomposed the portfolio risk into individual components under multivariate elliptical distribution setting. See Gupta et al. (2013) for a detailed discussion with the application of elliptical models in portfolio theory.

In this paper, we consider a generalized form of Eq. (A.1) for the multivariate t-distribution,

with location vector $\boldsymbol{\mu}$ and scale matrix $\boldsymbol{\Sigma}$ and degrees of freedom ν. The density is defined as

$$f(\mathbf{y}) = \frac{\Gamma(\frac{\nu+d}{2})|\boldsymbol{\Sigma}|^{-\frac{1}{2}}}{\Gamma^d(\frac{1}{2})\Gamma(\frac{\nu}{2})\nu^{\frac{d}{2}}}\left(1+\frac{1}{\nu}(\boldsymbol{y}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\right)^{-\frac{\nu+d}{2}}. \tag{A.2}$$

It is worth mentioning that in Eq. (A.2), $\boldsymbol{\Sigma}$ is not correlation matrix anymore, but a positive definite matrix which is proportional to the covariance matrix. Here we show two useful properties particularly for multivariate t-distribution defined in Eq. (A.2):

1. *Moments*

   If $\mathbf{Y} \sim t_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, then $E(\mathbf{y}) = \boldsymbol{\mu}$ and $Cov(\mathbf{y}) = \nu\boldsymbol{\Sigma}/(\nu-2)$, provided with $\nu > 2$.

2. *Affine Transformation*

   Given any $(\ell \times d)$ matrix A of rank $\ell \leq d$, the random vector $\mathbf{Z} = A\mathbf{Y} \sim t_d(A\boldsymbol{\mu}, A\boldsymbol{\Sigma}A', \nu)$.

Notice that the second property is a special case of Property 1 summarized in Section 2.1.

Next, we intend to show the density of 'standardized' t-distribution, i.e. the contour is centered around the origin and the covariance is an identity matrix, or for the univariate scenario, the variance equals one. Following Theorems 2.1.5 and 3.5.6 in Casella and Berger (2002), suppose $\mathbf{Y} \sim t_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ with the density function as Eq. (A.2), we may construct $\mathbf{Z} = \sqrt{\frac{\nu-2}{\nu}}\boldsymbol{\Sigma}^{-1/2}(\mathbf{Y}-\mu)$, which follows $t_d(\mathbf{0}, \frac{\nu-2}{\nu}\boldsymbol{I}, \nu)$, then the density function of $\mathbf{Z}$ and relationship between $\mathbf{Y}$ and $\mathbf{Z}$ are

$$\begin{aligned} f_{\mathbf{Z}}(\mathbf{z}) &= \frac{\Gamma(\frac{\nu+d}{2})}{\Gamma^d(\frac{1}{2})\Gamma(\frac{\nu}{2})(\nu-2)^{\frac{d}{2}}}\left(1+\frac{1}{\nu-2}\boldsymbol{z}^T\boldsymbol{z}\right)^{-\frac{\nu+d}{2}}, \\ f_{\mathbf{Y}}(\mathbf{y}) &= \sqrt{\frac{\nu-2}{\nu}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}} f_{\mathbf{Z}}(\sqrt{\frac{\nu-2}{\nu}}\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\mu)). \end{aligned} \tag{A.3}$$

The density of aggregated marginal variables from the multivariate distribution is also of interest. The linear combination of two arguments from a bivariate t-distributed random vector $\boldsymbol{Y} = (Y_1, Y_2)'$ is derived for simplicity. Assuming the weighting vector $A = (d_1, d_2)$, we could partition $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ into $\boldsymbol{\mu} = (\mu_{Y_1}, \mu_{Y_2})'$, $\boldsymbol{\Sigma} = (\sigma^2_{Y_1}, \sigma_{Y_1,Y_2}; \sigma_{Y_1,Y_2}, \sigma^2_{Y_2})$ and derive the portfolio distribution of $Z = A\boldsymbol{Y}$, with location parameter μ_Y and scale parameter σ^2_Y:

$$\begin{aligned} \mu_Z &= d_1\mu_{Y_1} + d_2\mu_{Y_2} \\ \sigma^2_Z &= d_1^2\sigma^2_{Y_1} + d_2^2\sigma^2_{Y_2} + 2d_1d_2\sigma_{Y_1,Y_2}. \end{aligned}$$

Note that σ^2_Z is not the variance of Z; the variance of Z is defined by $Var(Z) = \nu\sigma^2_Z/(\nu-2)$

as Property 1 states. The density function of $Z \sim t(\mu_Z, \sigma_Z^2, \nu)$ is defined as

$$f(z) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi\sigma_Z^2}\Gamma(\frac{\nu}{2})}\left(1 + \frac{(z-\mu_Z)^2}{\nu\sigma_Z^2}\right)^{-\frac{\nu+1}{2}}. \tag{A.4}$$

Finally, we provide the 'standardized' univariate t-distribution for a fair comparison with standard normal distribution $N(0,1)$. Given $Z \sim t(0, \frac{\nu-2}{\nu}, \nu)$, which is a t-distributed random variable with mean 0 and variance 1, we have $Z = \frac{Y-\mu_Y}{\sqrt{\frac{\nu}{\nu-2}\sigma_Y^2}}$ and

$$\begin{aligned} f_Z(z) &= \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{(\nu-2)\pi}\Gamma(\frac{\nu}{2})}\left(1 + \frac{z^2}{\nu-2}\right)^{-\frac{\nu+1}{2}}, \\ f_Y(y) &= \sqrt{\frac{\nu-2}{\nu\sigma_Y^2}} f_{\mathbf{Z}}\left(\frac{Y-\mu_Y}{\sqrt{\frac{\nu}{\nu-2}\sigma_Y^2}}\right). \end{aligned} \tag{A.5}$$

A.2 Proof of the Properness of PWL

Define $P_t = \int w_t(\boldsymbol{s})p_t(\boldsymbol{s})d\boldsymbol{s}$, and similarly $\hat{F}_t = \int w_t(\boldsymbol{s})\hat{f}_t(\boldsymbol{s})d\boldsymbol{s}$. Assumption 2(b), scaling $w_t(\boldsymbol{y})$ between 0 and 1, implies that P_t and $\hat{F}_t$ can be interpreted as the corresponding probability distribution functions for the true conditional density $p_t(\boldsymbol{s})$ and forecasted density $\hat{f}_t(\boldsymbol{s})$, respectively.

It is to be proven that $E_t(S(p_t; \boldsymbol{y}_{t+1}) - S(\hat{f}_t; \boldsymbol{y}_{t+1})) \geq 0$. Denote the conditional score difference for density forecast p_t and $\hat{f}_t$ as $d_{t+1}^{pwl}(p_t, \hat{f}_t) = S(p_t; \boldsymbol{y}_{t+1}) - S(\hat{f}_t; \boldsymbol{y}_{t+1})$, then

$$\begin{aligned} E_t\left(d_{t+1}^{pwl}(p_t, \hat{f}_t)\right) &= \int p_t(\boldsymbol{y})\left(w_t(\boldsymbol{y})\log(p_t(\boldsymbol{y})) - P_t - w_t(\boldsymbol{y})\log(\hat{f}_t(\boldsymbol{y})) + \hat{F}_t\right)d\boldsymbol{y} \\ &= \int w_t(\boldsymbol{y})p_t(\boldsymbol{y})\log\left(\frac{p_t(\boldsymbol{y})}{\hat{f}_t(\boldsymbol{y})}\right)d\boldsymbol{y} + \hat{F}_t - P_t \\ &= \int \left(w_t(\boldsymbol{y})p_t(\boldsymbol{y})\log\left(\frac{p_t(\boldsymbol{y})/P_t}{\hat{f}_t(\boldsymbol{y})/\hat{F}_t}\right) + w_t(\boldsymbol{y})p_t(\boldsymbol{y})\log\left(\frac{P_t}{\hat{F}_t}\right)\right)d\boldsymbol{y} + \hat{F}_t - P_t \\ &= P_t\int \frac{w_t(\boldsymbol{y})p_t(\boldsymbol{y})}{P_t}\log\left(\frac{w_t(\boldsymbol{y})p_t(\boldsymbol{y})/P_t}{w_t(\boldsymbol{y})\hat{f}_t(\boldsymbol{y})/\hat{F}_t}\right)d\boldsymbol{y} + P_t\log\left(\frac{P_t}{\hat{F}_t}\right) + \hat{F}_t - P_t \\ &= P_t\int \frac{w_t(\boldsymbol{y})p_t(\boldsymbol{y})}{P_t}\log\left(\frac{w_t(\boldsymbol{y})p_t(\boldsymbol{y})/P_t}{w_t(\boldsymbol{y})\hat{f}_t(\boldsymbol{y})/\hat{F}_t}\right)d\boldsymbol{y} + P_t\log(P_t) - P_t\log(\hat{F}_t) + \hat{F}_t - P_t \\ &= P_t K\left(\frac{w_t(\boldsymbol{y})p_t(\boldsymbol{y})}{P_t}, \frac{w_t(\boldsymbol{y})\hat{f}_t(\boldsymbol{y})}{\hat{F}_t}\right) + P_t\log(P_t) - P_t\log(\hat{F}_t) + \hat{F}_t - P_t, \end{aligned} \tag{A.6}$$

where $K(.,.)$ is the Kullback-Leibler divergence, which measures the difference between two probability distributions. According to Assumptions 1 and 2(c), $w_t(\boldsymbol{y})p_t(\boldsymbol{y})/P_t$ and $w_t(\boldsymbol{y})\hat{f}_t(\boldsymbol{y})/\hat{F}_t$ can be treated as pdfs, and Assumption 2(a) makes sure that the weighting function $w_t(\boldsymbol{y})$ is a given function of $\boldsymbol{y}$ conditional on $\mathscr{F}_t$. Thus the first term in the last line of (A.6), as a distance measurement of its arguments, is non-negative and finite. The remaining terms reach the minimum value zero if and only if $\hat{F}_t = P_t$, provided with the positive semidefinite Hessian matrix except for the trivial situation where $P_t = 0$.

A.3 Standardization of Skew Elliptical Distributions

In this subsection, we give the density expressions for the 'standardized' skew Normal and 'standardized' t-distributions. We consider this problem because the way we form skew Normal and t-distributions does not assign location and scale parameter $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ as the expectation and variance, respectively. When we assume the value of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, the expectation and covariance are pinned down according to Properties 2 and 3. To make the wto competing densities more comparable, we adjust the parameters such that the two competing densities are centerd around the origin and of unit variance, i.e. the 'standardized' covariance matrix coincides with the correlation matrix.

First consider the skew Normal distribution. Given $\boldsymbol{Y} \sim \mathrm{SN_n}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\alpha})$, and the pdf of the form in Eq. (5), the expectation and covariance of $\boldsymbol{Y}$ are defined by Property 2 as $E(\boldsymbol{Y}) = \boldsymbol{\mu} + \omega\boldsymbol{\mu_S}$ and $\mathrm{Cov}(\boldsymbol{Y}) = \boldsymbol{\Sigma} - \omega\boldsymbol{\mu_S}\boldsymbol{\mu_S}^T\omega$, with $\boldsymbol{\mu_S} = \sqrt{2/\pi}(1 + \boldsymbol{\alpha}^T\bar{\boldsymbol{\Sigma}}\boldsymbol{\alpha})^{-1/2}\bar{\boldsymbol{\Sigma}}\boldsymbol{\alpha}$. Note that $\boldsymbol{\Sigma} = \omega\bar{\boldsymbol{\Sigma}}\omega$, where ω is defined as $\omega = (\mathrm{diag}(\boldsymbol{\Sigma}))^{1/2}$. Thus to standardize $\boldsymbol{Y}$ we may express the covariance of $\boldsymbol{Y}$ as

$$\mathrm{Cov}(\boldsymbol{Y}) = \omega\bar{\boldsymbol{\Sigma}}\omega - \omega\boldsymbol{\mu_S}\boldsymbol{\mu_S}^T\omega \equiv \bar{\boldsymbol{\Sigma}}. \tag{A.7}$$

Eq. (A.7) gives the solution to ω. For n-dimensional variate, ω is a diagonal matrix of the form

$$\omega = \begin{pmatrix} \omega_1 & & \\ & \ddots & \\ & & \omega_n \end{pmatrix},$$

where ω_i is the standard deviation for the ith component. Using the Hadamard product instead

of the inner product, we may rewrite Eq. (A.7) as

$$\begin{aligned}\mathrm{Cov}(\boldsymbol{Y}) &= \bar{\boldsymbol{\Sigma}} \\ &= \omega(\bar{\boldsymbol{\Sigma}} - \boldsymbol{\mu_S}\boldsymbol{\mu_S}^T)\omega \\ &= \omega^2 \circ (\bar{\boldsymbol{\Sigma}} - \boldsymbol{\mu_S}\boldsymbol{\mu_S}^T).\end{aligned}$$

Given the standardized ω, $\boldsymbol{\mu} = -\omega\boldsymbol{\mu_S}$ and $\boldsymbol{\Sigma} = \omega\bar{\boldsymbol{\Sigma}}\omega$ will deliver the standardized skew Normal distribution.

For $\boldsymbol{Y} \sim \mathrm{St}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}, \nu)$, the only difference is now $\boldsymbol{\mu_S} = (\frac{\nu}{\pi})^{1/2}\frac{\Gamma\{\frac{1}{2}(\nu-1)\}}{\Gamma(\frac{1}{2}\nu)}(1+\boldsymbol{\alpha}^T\bar{\boldsymbol{\Sigma}}\boldsymbol{\alpha})^{-1/2}\bar{\boldsymbol{\Sigma}}\boldsymbol{\alpha}$, and Eq. (A.7) still holds. Similar as before, we could solve for the standardizing parameter $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

When we project a multivariate distribution onto the real line, we must do the standardization again for the univariate distribution. Following the previous notation, if $\boldsymbol{Y} \sim \mathrm{SN_n}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\alpha})$, $Z = A\boldsymbol{Y} \sim \mathrm{SN}(\mu_Z, \Sigma_Z, \alpha_Z, \nu)$ with $(1 \times n)$ matrix A and the scalar parameter μ_Z, Σ_Z and α_Z are defined same as Eq. (7) (see Propety 2(b)). Recall $\boldsymbol{Y}$ is already standardized centered around zero, hence $E(Z) = E(A\boldsymbol{Y}) = 0$ and we only need to consider re-scale variance to be unit, which requires

$$\mathrm{Var}(Z) = \Sigma'_Z - \omega'_Z\mu_s^2\omega'_Z \equiv 1, \tag{A.8}$$

where $\mu_s = \sqrt{2}\alpha_Z/\sqrt{\pi(1+\alpha_Z^2)}$ and $\omega'_Z = \sqrt{\Sigma'_Z}$. From Eq. (A.8), it is easy to show that $\Sigma'_Z = 1/(1-\mu_s^2)$. The derivation for standardizing the skew t-distribution is similar, apart from the definition of $\mu_s = \sqrt{\frac{\nu}{\pi}}\frac{\Gamma\{(\nu-1)/2\}}{\Gamma(\nu/2)}$ and the requirement for the variance becomes

$$\mathrm{Var}(Z) = \frac{\nu}{\pi}\Sigma'_Z - \omega'_Z\mu_s^2\omega'_Z \equiv 1.$$

Solving $\mathrm{Var}(Z) = 1$ gives the scalar parameter Σ'_Z.

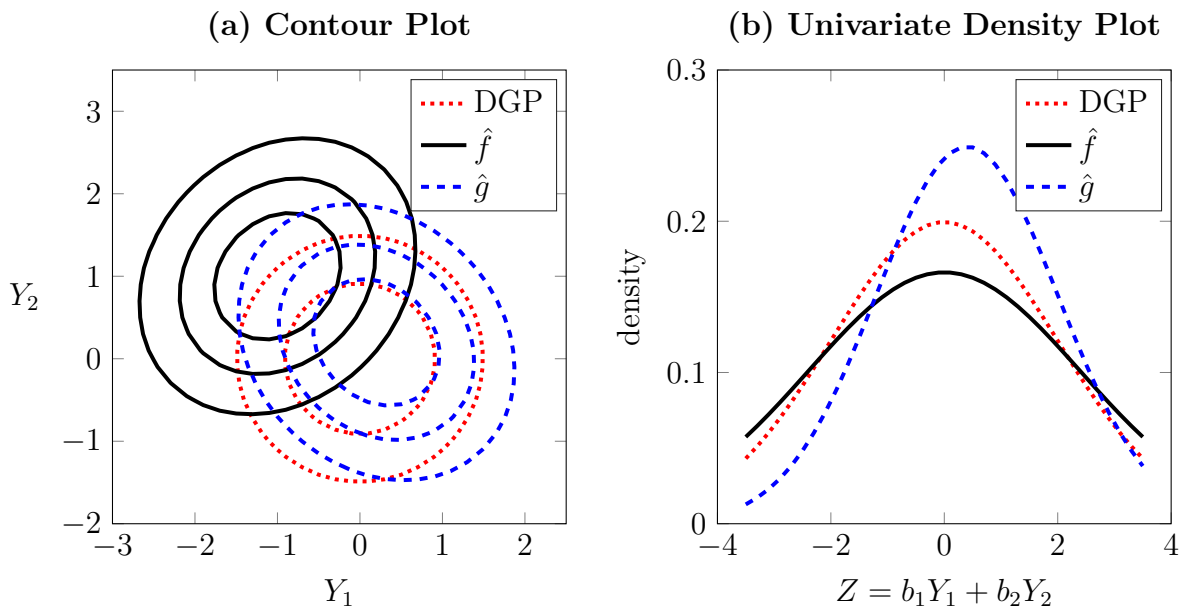

Figure 1: Density comparison in bivariate space (a) and univariate space (b).

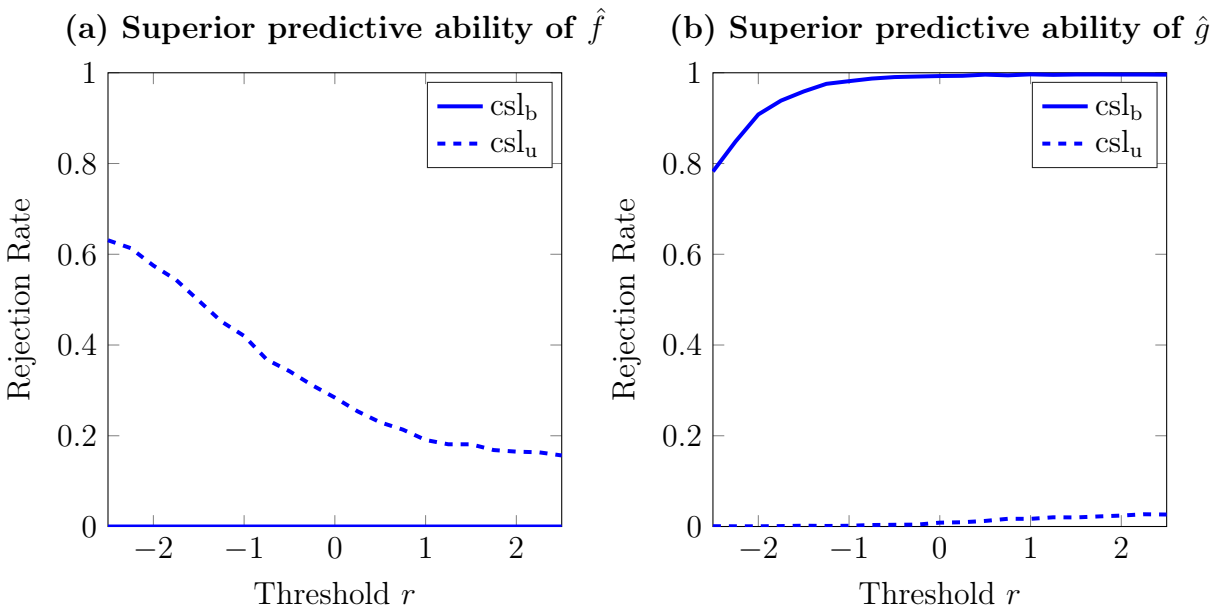

Figure 2: One-sided rejection rates (at nominal size 5%) of Diebold-Mariano type test statistic of equal predictive accuracy under the threshold weight function $w(z) = I(z \leq r)$ for $c = 40$ expected observations in the left tail, based on 10,000 replications. The DGP is i.i.d. standard bivariate normal. Panel (a) shows high rejection rates against superior predictive ability of $\hat{f}$ as a function of r, while rejection in panel (b) suggests $\hat{g}$ is more accurate.

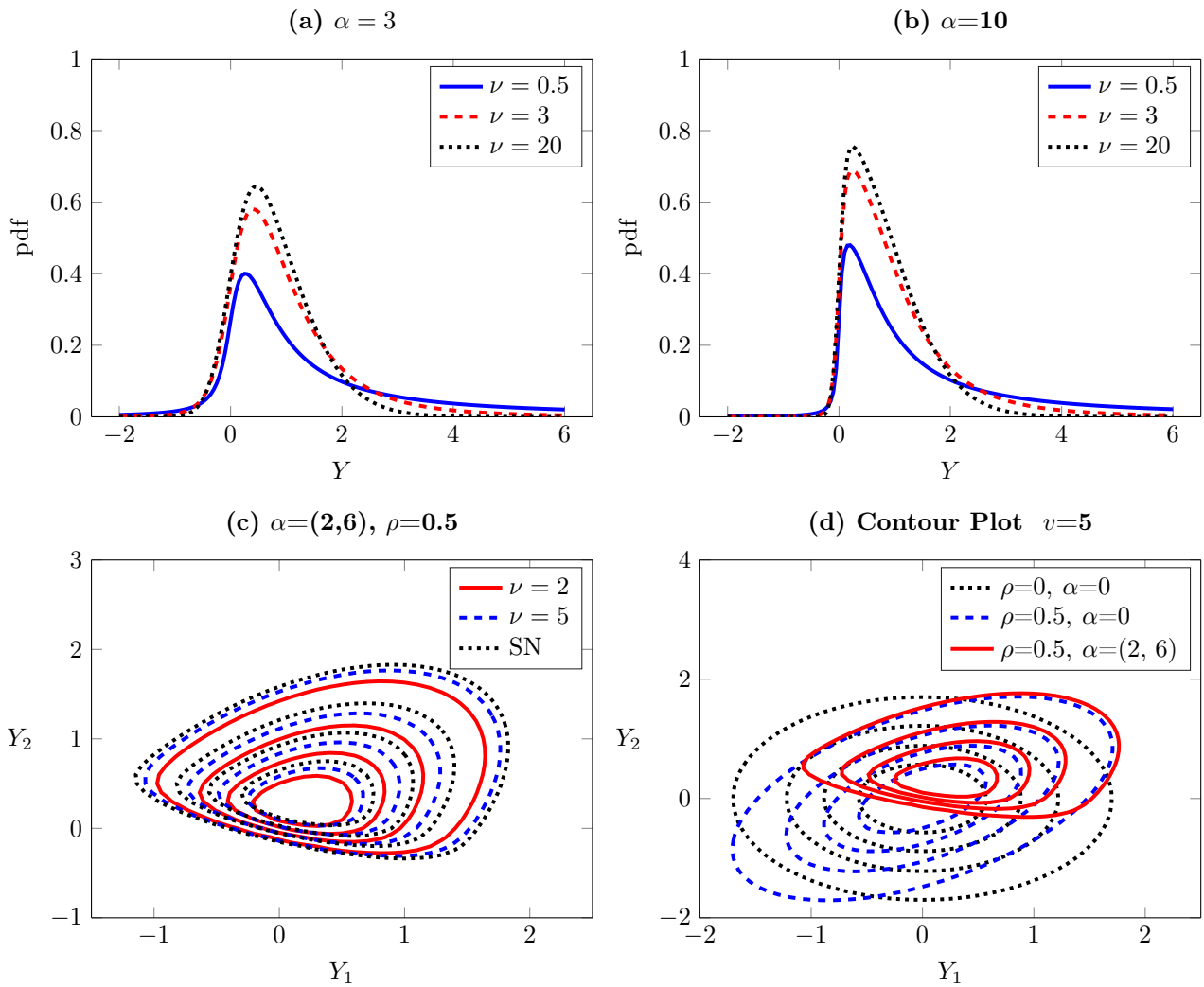

Figure 3: Graphical illustration of skew t-distribution. The top two panels presents the univariate skew t-density functions for $\alpha = 3$ in the left and $\alpha = 10$ in the right column, for different values of ν. The bottom panels show contour plots for the standardized bivariate skew t-distribution. In the lower-left panel, given skewness vector $\boldsymbol{\alpha} = (2, 6)$ and correlation $\rho = 0.5$, the contours of the skew t and skew normal distributions are shown; the lower-right panel illustrates changes in the contour plot when correlation and skewness are introduced.

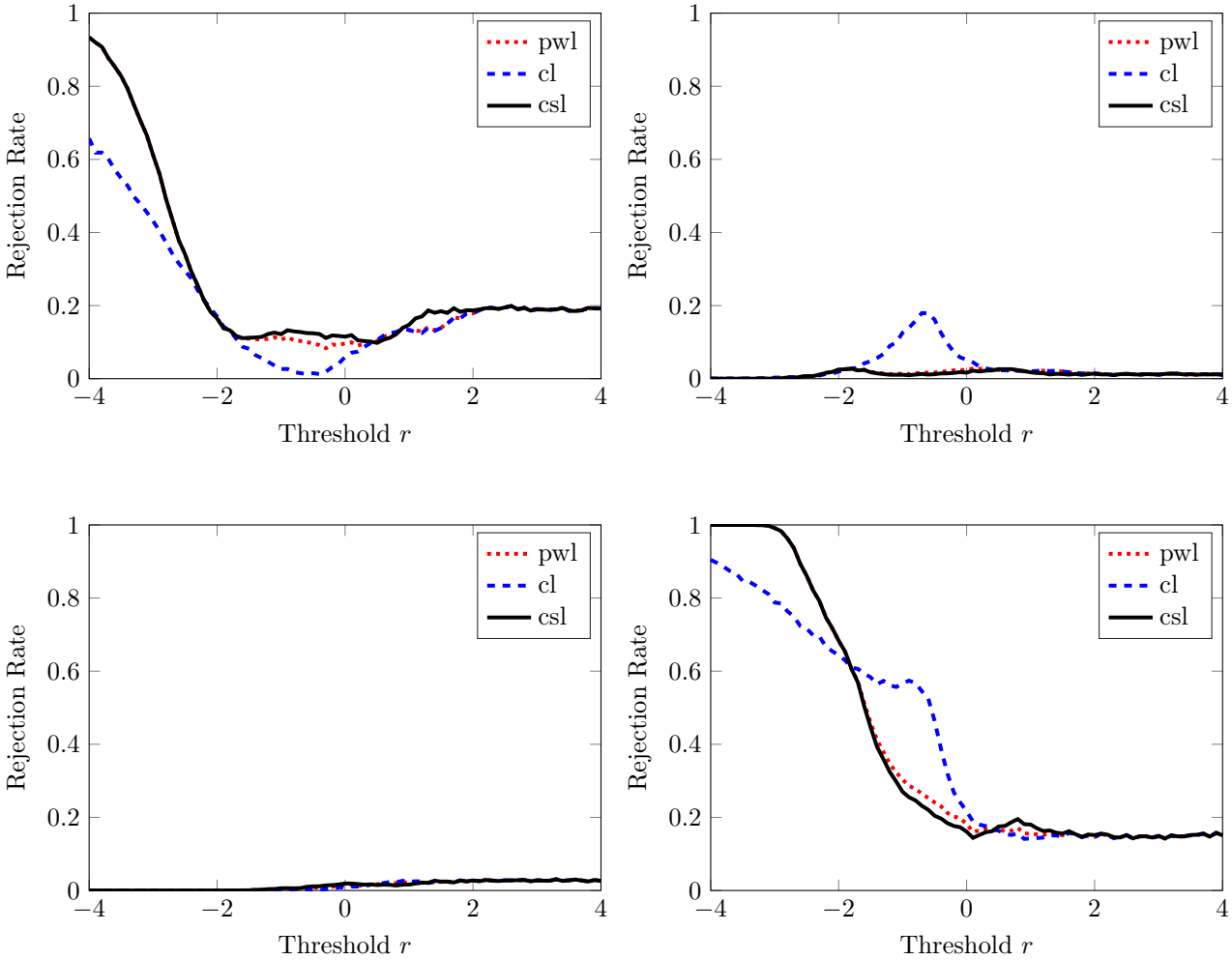

Figure 4: One-sided rejection rates (at nominal size 5%) of Diebold-Mariano type test statistic of equal predictive accuracy under the threshold weight function $w(y) = I(y \leq r)$ for $c = 20$ expected observations in the left tail, based on 10,000 replications. The DGP is i.i.d. skew $t(5)$ (skew normal) with $\alpha = -2$ for the upper (lower) two panels. The left (right) column shows rejection rates against superior predictive ability of skew $t(5)$ (skew normal) with $\alpha = -2$ as a function of the threshold parameter r.

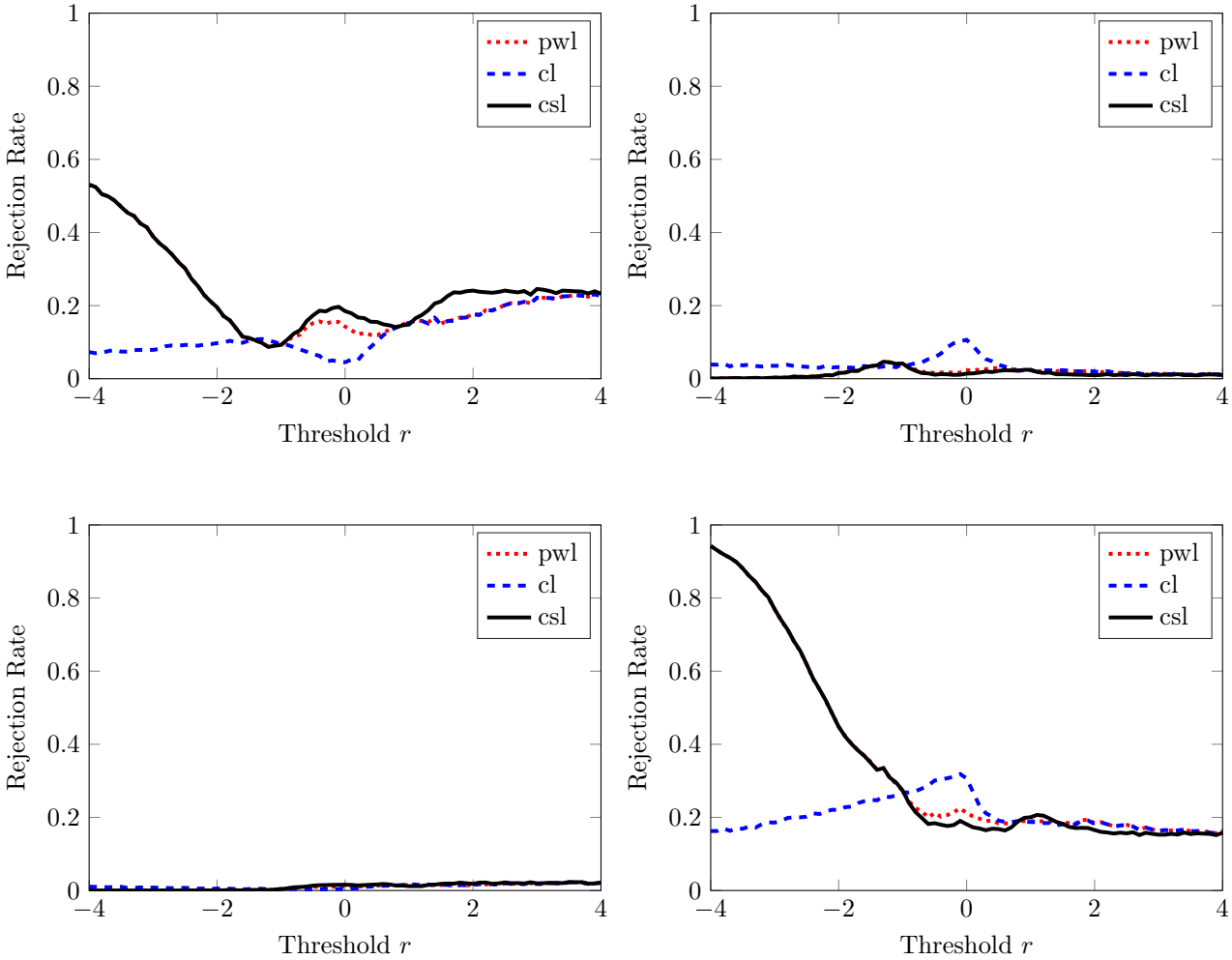

Figure 5: One-sided rejection rates (at nominal size 5%) of Diebold-Mariano type test statistic of equal predictive accuracy under the threshold weight function $w(z) = I(z \leq r)$ for $c = 20$ expected observations in the left tail, based on 10,000 replications. The DGP is i.i.d. skew $t(5)$ with $\alpha = -2$ ($t(5)$) in upper (lower) two panels. The left(right) column shows rejection rates against superior predictive ability of skew $t(5)$ with $\alpha = -2$ ($t(5)$) as a function of the threshold parameter r.

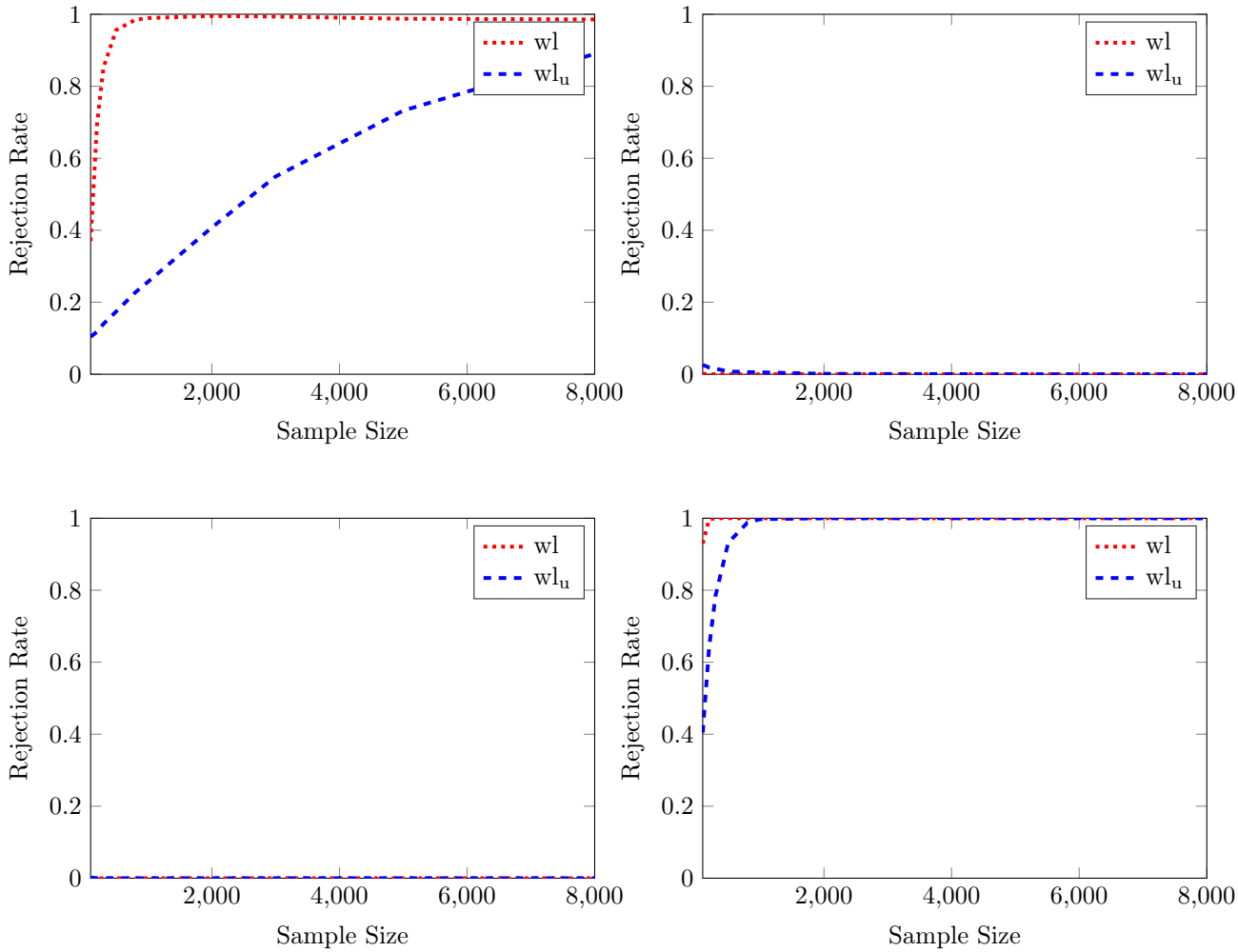

Figure 6: One-sided rejection rates (at nominal size 5%) of Diebold-Mariano type test statistic of equal predictive accuracy, based on 10,000 replications. The DGP is i.i.d $\mathrm{S}t_4(5)$ ($\mathrm{S}N_4$) for the upper (lower) two panels, with $\boldsymbol{\alpha} = (-2, 1, -1, -1)$. The left (right) column shows rejection rates against superior predictive ability of $\mathrm{S}t_4(5)$ ($\mathrm{S}N_4$) as a function of sample size. wl denotes the test of weighted likelihood scoring rule directly on four-dimensional density, while wl_u is the test based on the univariate density projected by vector $(1/4, 1/4, 1/4, 1/4)$.

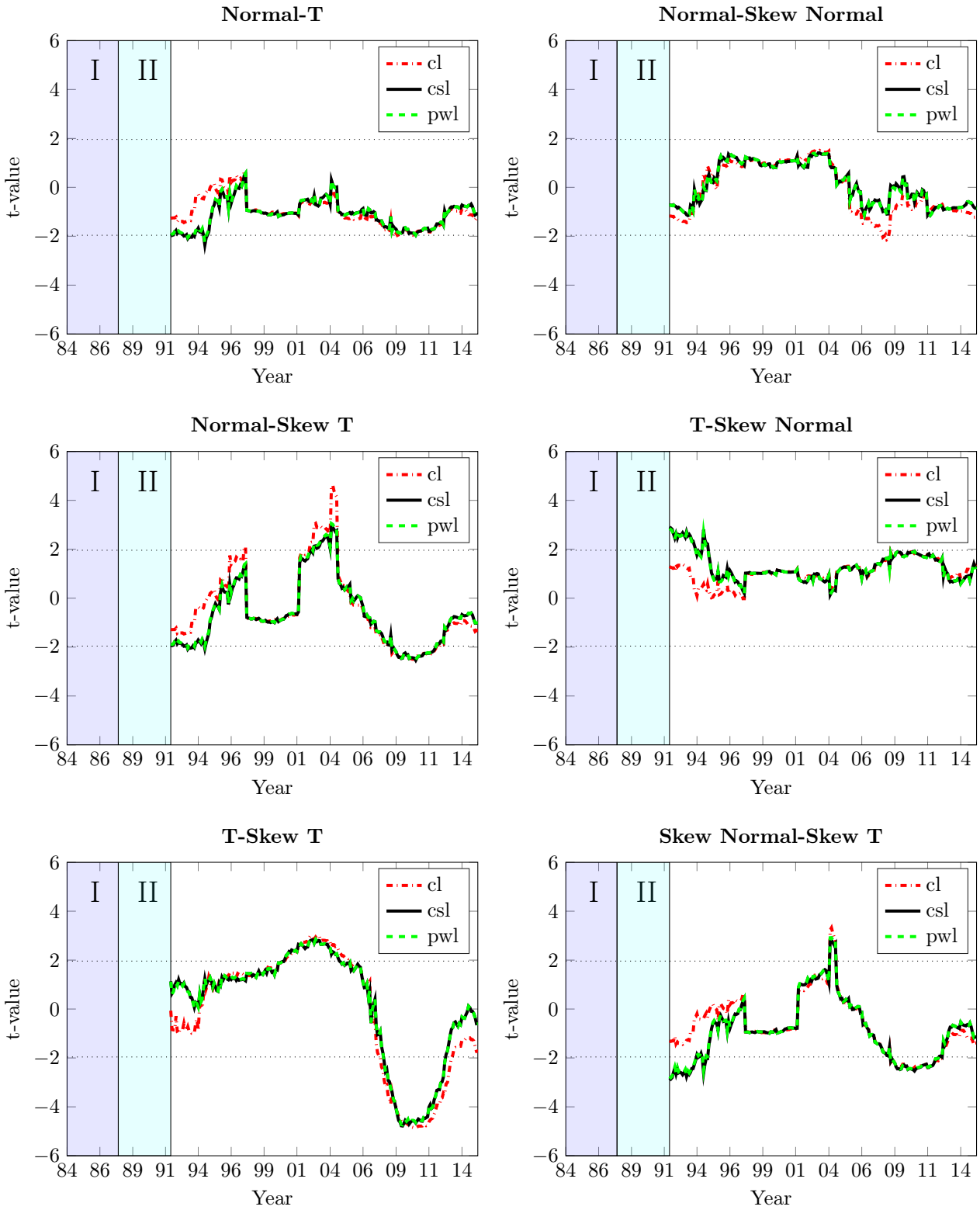

Figure 7: Dynamics of the two-sided Diebold-Mariano type test statistic (at nominal size 5%) of equal predictive accuracy, based on the *cl*, *csl* and *pwl* scoring rules. Empirical quantile $q = 0.05$ in the indicator weight function $w_t(r_t) = I(r_t \leq \hat{r}_t^q)$ is applied for selecting the left tail region. The light blue (I) and cyan (II) rectangles represent the first estimation and evaluation window.

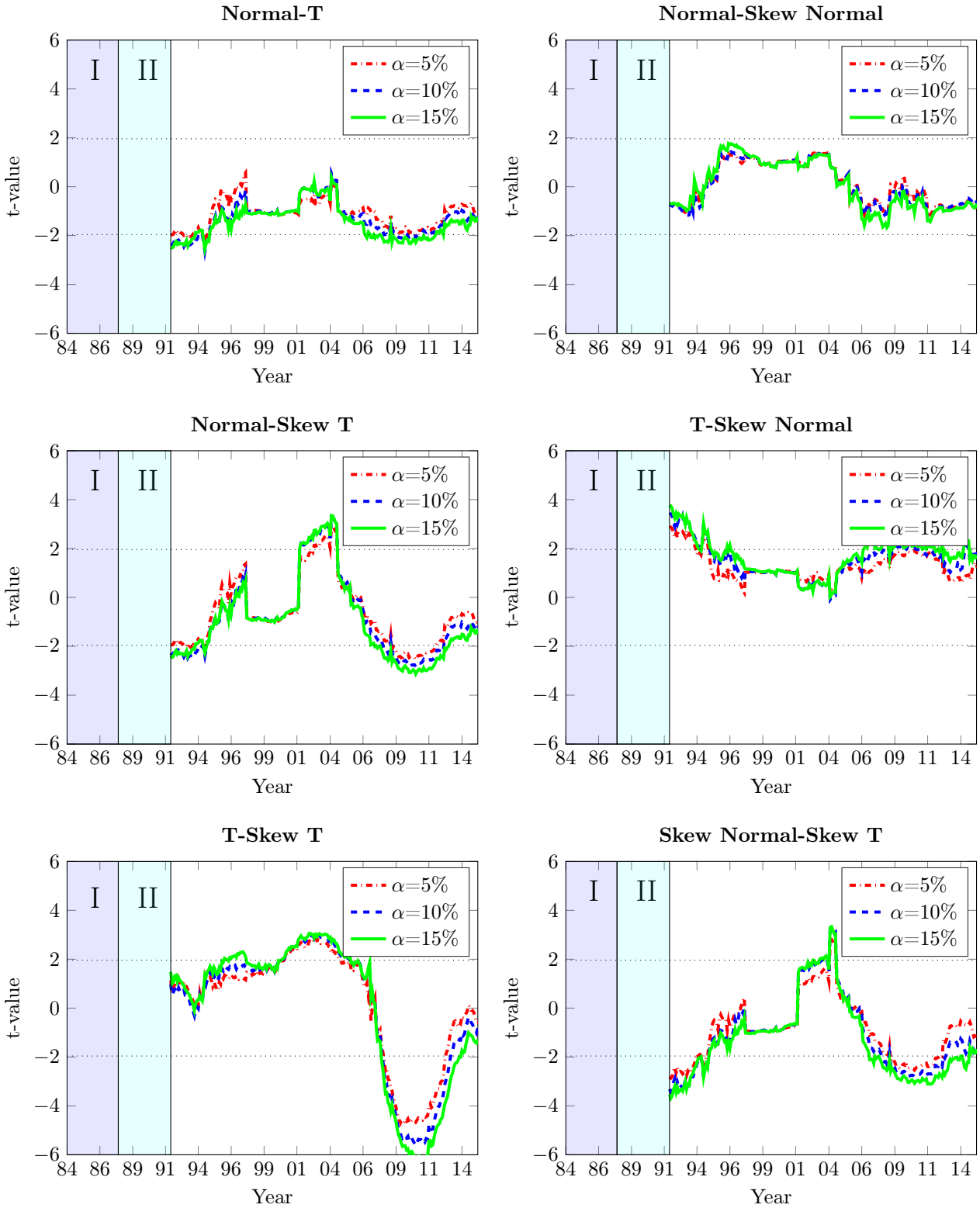

Figure 8: Dynamics of the two-sided Diebold-Mariano type test statistic (at nominal size 5%) of equal predictive accuracy, based on the *csl* scoring rule. Empirical quantile $q = 0.05, 0.10, 0.15$ in the indicator weight function $w_t(r_t) = I(r_t \leq \hat{r}_t^q)$ are applied for selecting the left tail region. The light blue (I) and cyan (II) rectangles represent the first estimation and evaluation window.

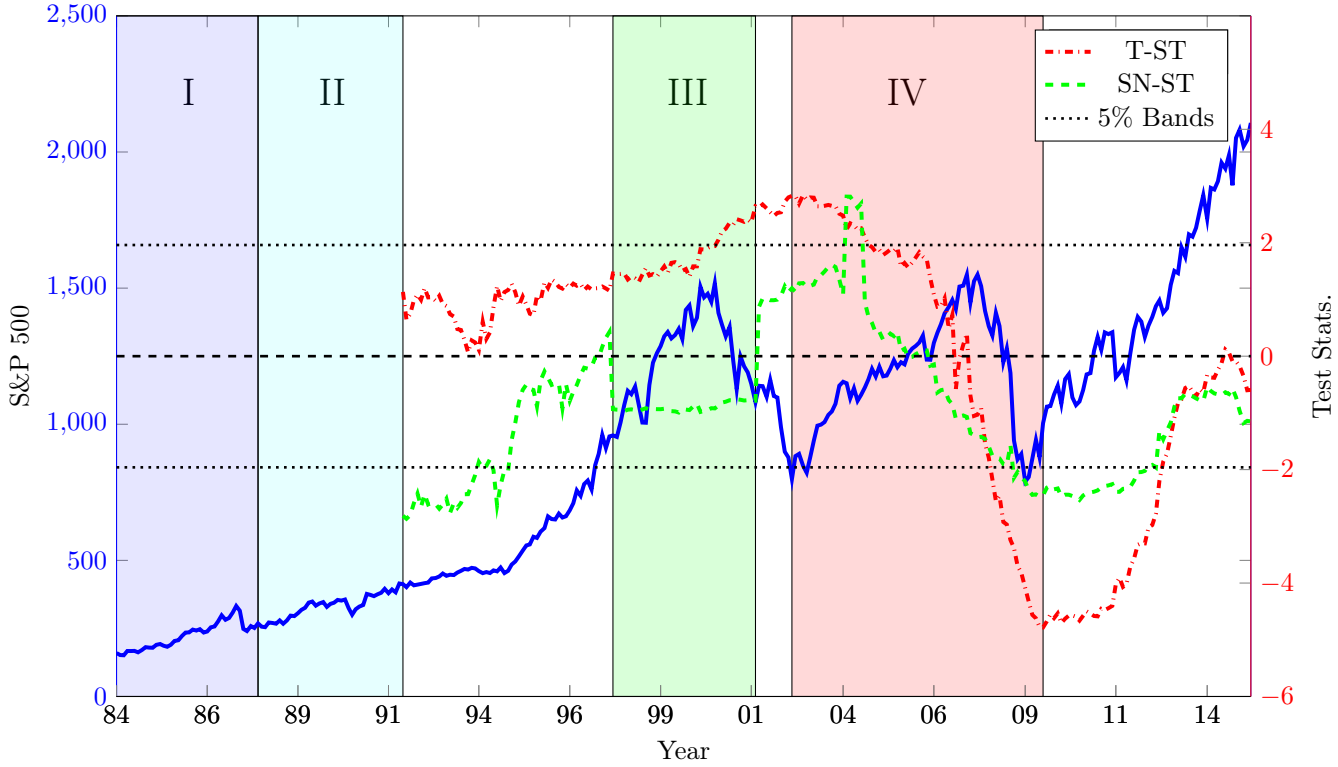

Figure 9: S&P500 Index and the dynamics of the two-sided Diebold-Mariano type test statistic (at nominal size 5%) of equal predictive accuracy for t against skew t, and skew Normal against skew t, based on the *csl* scoring rule. Empirical quantile $q = 0.05$ in the indicator weight function $w_t(r_t) = I(r_t \leq \hat{r}_t^q)$ is applied for selecting the left tail region. The light blue (I) and cyan (II) rectangles represent the first estimation and evaluation windows, respectively. The green (III) and pink (IV) regions indicate two periods discussed in the main text.

Table 1: Average score differences and test statistics based on the *csl* rule

scoring rule	N − t	N − SN	N − St	t − SN	t − St	SN − St
			$q = 0.01$			
csl_{multi}	-0.0104	0.0017	-0.0089	0.0122	0.0016	-0.0106
	(-2.7627*)	(0.7224)	(-2.6501*)	(3.3909*)	(1.2037)	(-3.2392*)
csl_{proj}	-0.0067	0.0001	-0.0061	0.0068	0.0005	-0.0063
	(-2.4209*)	(0.0736)	(-2.5974*)	(3.2254*)	(0.6275)	(-3.3745*)
csl_{uni}	-0.0057	0.0003	-0.0048	0.0060	0.0009	-0.0051
	(-2.1393*)	(0.2811)	(-2.0449*)	(1.8565)	(1.4043)	(-1.7973*)
			$q = 0.05$			
	N − t	N − SN	N − St	t − SN	t − St	SN − St
csl_{multi}	-0.0216	0.0069	-0.0174	0.0285	0.0042	-0.0243
	(-4.3507*)	(1.9909*)	(-3.7816*)	(5.5157*)	(2.5909*)	(-5.0617*)
csl_{proj}	-0.0089	0.0016	-0.0074	0.0105	0.0015	-0.0090
	(-3.0321*)	(0.6765)	(-2.7956*)	(4.1646*)	(1.5886)	(-3.7488*)
csl_{uni}	-0.0076	0.0003	-0.0062	0.0079	0.0014	-0.0065
	(-2.6652*)	(0.2676)	(-2.3725*)	(2.3920*)	(1.8705)	(-2.2157*)
			$q = 0.10$			
	N − t	N − SN	N − St	t − SN	t − St	SN − St
csl_{multi}	-0.0283	0.0099	-0.0223	0.0382	0.0060	-0.0322
	(-5.4127*)	(2.5551*)	(-4.5035*)	(7.0949*)	(3.3665*)	(-6.4189*)
csl_{proj}	-0.0103	0.0019	-0.0084	0.0122	0.0019	-0.0103
	(-3.4735*)	(0.7644)	(-3.0973*)	(4.5199*)	(1.9493)	(-3.9436*)
csl_{uni}	-0.0086	0.0003	-0.0073	0.0089	0.0013	-0.0076
	(-3.0036*)	(0.2705)	(-2.7583*)	(2.6935*)	(1.7593)	(-2.5736*)
			$q = 0.15$			
	N − t	N − SN	N − St	t − SN	t − St	SN − St
csl_{multi}	-0.1244	-0.0183	-0.1167	0.1061	0.0077	-0.0984
	(-1.4249)	(-0.5578)	(-1.3356)	(1.9351)	(4.1372*)	(-1.7933)
csl_{proj}	-0.0114	0.0020	-0.0090	0.0135	0.0025	-0.0110
	(-3.8191*)	(0.8007)	(-3.3008*)	(4.8685*)	(2.3632*)	(-4.1300*)
csl_{uni}	-0.0095	0.0002	-0.0082	0.0097	0.0013	-0.0085
	(-3.2908*)	(0.2128)	(-3.0791*)	(2.9310*)	(1.6951)	(-2.8520*)

Note: This table presents the average score difference $\bar{d}$ and the corresponding test statistic (the number in parentheses) for the censored likelihood scoring rule (3) for different empirical quantiles q in the indicator weight function $w_t(r_t) = I(r_t \leq \hat{r}_t^q)$, where for the four sub-panels, $q = 0.01, 0.05, 0.1, 0.15$ respectively. The superscript * denotes that the test statistic is significant at the 5% level of significance. There are six pair-comparisons since we have four distributional assumptions about the innovations. The test result in the first row of each sub-panel is based on multivariate regression model, and the corresponding projection onto one-dimensional space yields the outcome in the second row. The bottom row of each sub-panel gives test results based on the univariate regression model of portfolio return series.

Table 2: Average score differences and test statistics based on the cl rule

scoring rule	N − t	N − SN	N − St	t − SN	t − St	SN − St
			$q = 0.01$			
cl_{multi}	-0.0084	-0.0010	-0.0068	0.0074	0.0015	-0.0059
	(-2.5060*)	(-0.5256)	(-2.3475*)	(2.0735*)	(1.5836)	(-1.8065)
cl_{proj}	-0.0046	-0.0025	-0.0041	0.0021	0.0005	-0.0016
	(-2.0268*)	(-2.9047*)	(-2.2145*)	(1.2272)	(0.9889)	(-1.2484)
cl_{uni}	-0.0043	0.0006	-0.0032	0.0049	0.0010	-0.0039
	(-1.8019)	(0.8646)	(-1.5998)	(1.6195)	(1.9506)	(-1.4604)
			$q = 0.05$			
	N − t	N − SN	N − St	t − SN	t − St	SN − St
cl_{multi}	-0.0204	0.0021	-0.0174	0.0225	0.0031	-0.0194
	(-4.1319*)	(0.6889)	(-3.8511*)	(4.3397*)	(2.1550*)	(-4.1049*)
cl_{proj}	-0.0077	-0.0033	-0.0074	0.0044	0.0003	-0.0041
	(-2.7353*)	(-2.3576*)	(-3.0417*)	(2.2436*)	(0.4336)	(-2.5451*)
cl_{uni}	-0.0066	-0.0000	-0.0056	0.0066	0.0011	-0.0055
	(-2.4056*)	(-0.0403)	(-2.2393*)	(2.0364*)	(1.6494)	(-1.9237)
			$q = 0.10$			
	N − t	N − SN	N − St	t − SN	t − St	SN − St
cl_{multi}	-0.0263	0.0056	-0.0216	0.0319	0.0047	-0.0272
	(-4.9598*)	(1.5979)	(-4.3784*)	(5.8326*)	(2.8460*)	(-5.4410*)
cl_{proj}	-0.0083	-0.0024	-0.0077	0.0058	0.0006	-0.0052
	(-2.8226*)	(-1.5076)	(-2.9940*)	(2.8247*)	(0.7479)	(-2.9624*)
cl_{uni}	-0.0072	0.0001	-0.0060	0.0073	0.0012	-0.0061
	(-2.5176*)	(0.0944)	(-2.3030*)	(2.2104*)	(1.6319)	(-2.0855*)
			$q = 0.15$			
	N − t	N − SN	N − St	t − SN	t − St	SN − St
cl_{multi}	-0.1215	-0.0224	-0.1152	0.0991	0.0063	-0.0928
	(-1.3915)	(-0.6834)	(-1.3184)	(1.8069)	(3.5978*)	(-1.6911)
cl_{proj}	-0.0085	-0.0021	-0.0074	0.0065	0.0011	-0.0054
	(-2.8861*)	(-1.1609)	(-2.8563*)	(3.0549*)	(1.2834)	(-2.8686*)
cl_{uni}	-0.0075	0.0003	-0.0062	0.0078	0.0013	-0.0065
	(-2.6176*)	(0.2855)	(-2.3658*)	(2.3603*)	(1.6528)	(-2.2245*)

Note: This table presents the average score difference $\bar{d}$ and the corresponding test statistic (the number in parentheses) for the conditional likelihood scoring rule (2) for different empirical quantiles q in the indicator weight function $w_t(r_t) = I(r_t \leq \hat{r}_t^q)$, where for the four sub-panels, $q = 0.01, 0.05, 0.1, 0.15$, respectively. The superscript * denotes that the test statistic is significant at the 5% level of significance. There are six pair-comparisons since we have four distributional assumptions about the innovations. The test result in the first row of each sub-panel is based on multivariate regression model, and the corresponding projection onto one-dimensional space yields the outcome in the second row. The bottom row of each sub-panel gives testing result based on the univariate regression model of portfolio return series.

Table 3: Average score differences and test statistics based on the *pwl* rule

scoring rule	N − t	N − SN	N − St	t − SN	t − St	SN − St
			$q = 0.01$			
pwl_{multi}	-0.0104	0.0015	-0.0088	0.0119	0.0016	-0.0103
	(-2.7584*)	(0.6313)	(-2.6425*)	(3.3268*)	(1.2167)	(-3.1664*)
pwl_{proj}	-0.0066	-0.0001	-0.0061	0.0065	0.0005	-0.0060
	(-2.4172*)	(-0.0606)	(-2.5927*)	(3.1492*)	(0.6457)	(-3.2981*)
pwl_{uni}	-0.0057	0.0003	-0.0048	0.0060	0.0009	-0.0050
	(-2.1356*)	(0.2768)	(-2.0393*)	(1.8519)	(1.4161)	(-1.7906)
			$q = 0.05$			
	N − t	N − SN	N − St	t − SN	t − St	SN − St
pwl_{multi}	-0.0215	0.0060	-0.0173	0.0275	0.0042	-0.0233
	(-4.3302*)	(1.7873)	(-3.7774*)	(5.3561*)	(2.5726*)	(-4.9076*)
pwl_{proj}	-0.0088	0.0007	-0.0074	0.0095	0.0014	-0.0081
	(-3.0078*)	(0.3123)	(-2.8031*)	(3.9747*)	(1.5495)	(-3.5782*)
pwl_{uni}	-0.0074	0.0003	-0.0061	0.0077	0.0014	-0.0064
	(-2.6340*)	(0.2664)	(-2.3442*)	(2.3619*)	(1.8647)	(-2.1863)
			$q = 0.10$			
	N − t	N − SN	N − St	t − SN	t − St	SN − St
pwl_{multi}	-0.0281	0.0088	-0.0222	0.0369	0.0059	-0.0310
	(-5.3676*)	(2.3429*)	(-4.4909*)	(6.8909*)	(3.3378*)	(-6.2352*)
pwl_{proj}	-0.0101	0.0008	-0.0083	0.0108	0.0018	-0.0090
	(-3.4151*)	(0.3408)	(-3.0963*)	(4.3593*)	(1.8907)	(-3.8081*)
pwl_{uni}	-0.0084	0.0003	-0.0071	0.0087	0.0013	-0.0074
	(-2.9532*)	(0.2873)	(-2.7100*)	(2.6515*)	(1.7423)	(-2.5323)
			$q = 0.15$			
	N − t	N − SN	N − St	t − SN	t − St	SN − St
pwl_{multi}	-0.1240	-0.0198	-0.1165	0.1043	0.0075	-0.0967
	(-1.4199)	(-0.6012)	(-1.3332)	(1.9015)	(4.0886*)	(-1.7636)
pwl_{proj}	-0.0110	0.0006	-0.0087	0.0116	0.0023	-0.0094
	(-3.7005*)	(0.2699)	(-3.2605*)	(4.6087*)	(2.2821*)	(-3.8978*)
pwl_{uni}	-0.0092	0.0003	-0.0079	0.0095	0.0013	-0.0082
	(-3.1992*)	(0.2562)	(-2.9819*)	(2.8581*)	(1.6820)	(-2.7730*)

Note: This table presents the average score difference $\bar{d}$ and the corresponding test statistic (the number in parentheses) for the penalized weighted likelihood scoring rule (4) for different empirical quantiles q in the indicator weight function $w_t(r_t) = I(r_t \leq \hat{r}_t^q)$, where for the four sub-panels, $q = 0.01, 0.05, 0.1, 0.15$ respectively. The superscript * denotes that the test statistic is significant at the 5% level of significance. There are six pair-comparisons since we have four distributional assumptions about the innovations. The test result in the first row of each sub-panel is based on multivariate regression model, and the corresponding projection onto one-dimensional space yields the outcome in the second row. The bottom row of each sub-panel gives testing result based on the univariate regression model of portfolio return series.

Table 4: Average score differences and test statistics of univariate andp projected densities

	$q = 0.01$											
	cl				*csl*				*pwl*			
	N_{uni}	t_{uni}	SpN_{uni}	St_{uni}	N_{uni}	t_{uni}	SN_{uni}	St_{uni}	N_{uni}	t_{uni}	SN_{uni}	St_{uni}
N_{proj}	0.0004	0.0047	-0.0002	0.0036	0.0011	0.0068	0.0009	0.0059	0.0011	0.0068	0.0009	0.0059
	(1.0490)	(2.0338*)	(-0.2504)	(1.8524)	(1.4575)	(2.4573*)	(0.5940)	(2.3774*)	(1.4624)	(2.4515*)	(0.5913)	(2.3704*)
t_{proj}	-0.0042	0.0000	-0.0049	-0.0010	-0.0055	0.0002	-0.0058	-0.0007	-0.0055	0.0002	-0.0058	-0.0008
	(-1.7909)	(0.6433)	(-1.6117)	(-1.9412)	(-2.0637*)	(0.6276)	(-1.8009)	(-1.1593)	(-2.0631*)	(0.6250)	(-1.7985)	(-1.1803)
SN_{proj}	-0.0021	0.0021	-0.0028	0.0011	0.0013	0.0070	0.0010	0.0061	0.001	0.0067	0.0007	0.0058
	(-2.6583*)	(1.2468)	(-1.9341)	(0.7798)	(0.7819)	(3.3199*)	(0.4600)	(3.0842*)	(0.6455)	(3.2397*)	(0.3503)	(3.0010*)
St_{proj}	-0.0037	0.0005	-0.0044	-0.0005	-0.0050	0.0007	-0.0053	-0.0002	-0.005	0.0007	-0.0053	-0.0002
	(-1.9451)	(1.0273)	(-1.7113)	(-1.7364)	(-2.2662*)	(0.7653)	(-2.0177*)	(-0.4717)	(-2.2626*)	(0.7785)	(-2.0116*)	(-0.4665)
	$q = 0.05$											
	cl				*csl*				*pwl*			
	N_{uni}	t_{uni}	SN_{uni}	St_{uni}	N_{uni}	t_{uni}	SN_{uni}	St_{uni}	N_{uni}	t_{uni}	SN_{uni}	St_{uni}
N_{proj}	0.0013	0.0080	0.0014	0.0069	0.0016	0.0091	0.0013	0.0077	0.0016	0.0090	0.0013	0.0076
	(2.0439*)	(2.7957*)	(0.9998)	(2.6455*)	(1.5837)	(3.0650*)	(0.7666)	(2.7932*)	(1.6454)	(3.0402*)	(0.7847)	(2.7725*)
t_{proj}	-0.0064	0.0002	-0.0064	-0.0008	-0.0073	0.0002	-0.0076	-0.0012	-0.0072	0.0002	-0.0075	-0.0011
	(-2.327*)	(1.5013)	(-1.9720*)	(-1.3369)	(-2.5526*)	(0.5086)	(-2.3049*)	(-1.4447)	(-2.5281*)	(0.5352)	(-2.2799*)	(-1.4632)
SN_{proj}	-0.0020	0.0047	-0.0019	0.0036	0.0031	0.0107	0.0028	0.0093	0.0022	0.0097	0.0020	0.0083
	(-1.653)	(2.3509*)	(-1.1106)	(2.0119*)	(1.5294)	(4.2659*)	(1.1357)	(3.7357*)	(1.1924)	(4.0796*)	(0.8286)	(3.5451*)
St_{proj}	-0.0061	0.0005	-0.0061	-0.0005	-0.0059	0.0017	-0.0062	0.0003	-0.0058	0.0016	-0.0061	0.0003
	(-2.6386*)	(0.7302)	(-2.2631*)	(-1.3953)	(-2.3245*)	(1.5741)	(-2.2224*)	(0.4457)	(-2.3274*)	(1.5507)	(-2.2191*)	(0.3883)
	$q = 0.10$											
	cl				*csl*				*pwl*			
	N_{uni}	t_{uni}	SN_{uni}	St_{uni}	N_{uni}	t_{uni}	SN_{uni}	St_{uni}	N_{uni}	t_{uni}	SN_{uni}	St_{uni}
N_{proj}	0.0013	0.0085	0.0012	0.0073	0.0018	0.0104	0.0015	0.0091	0.0018	0.0102	0.0014	0.0089
	(1.6952)	(2.8561*)	(0.811)	(2.6528*)	(1.7325)	(3.4469*)	(0.8830)	(3.2168*)	(1.7796)	(3.3913*)	(0.8695)	(3.1639*)
t_{proj}	-0.0070	0.0002	-0.0071	-0.001	-0.0085	0.0001	-0.0088	-0.0012	-0.0083	0.0001	-0.0086	-0.0012
	(-2.4510*)	(0.7559)	(-2.159*)	(-1.4063)	(-2.9315*)	(0.2187)	(-2.6416*)	(-1.4655)	(-2.8907*)	(0.2336)	(-2.6071*)	(-1.4933)
SN_{proj}	-0.0011	0.0060	-0.0012	0.0048	0.0037	0.0123	0.0034	0.0110	0.0025	0.0109	0.0022	0.0096
	(-0.8199)	(2.9008*)	(-0.6517)	(2.5127*)	(1.6705)	(4.5486*)	(1.2904)	(4.0796*)	(1.2812)	(4.3941*)	(0.9111)	(3.9134*)
St_{proj}	-0.0064	0.0008	-0.0065	-0.0004	-0.0066	0.0020	-0.0069	0.0007	-0.0065	0.0019	-0.0068	0.0006
	(-2.6355*)	(0.8998)	(-2.3957*)	(-0.8504)	(-2.5669*)	(1.7664)	(-2.4507*)	(0.8733)	(-2.5782*)	(1.7298)	(-2.4619*)	(0.7986)
	$q = 0.15$											
	cl				*csl*				*pwl*			
	N_{uni}	t_{uni}	SN_{uni}	St_{uni}	N_{uni}	t_{uni}	SN_{uni}	St_{uni}	N_{uni}	t_{uni}	SN_{uni}	St_{uni}
N_{proj}	0.0014	0.0089	0.0011	0.0077	0.0019	0.0115	0.0017	0.0102	0.0019	0.0111	0.0016	0.0098
	(1.6794)	(2.9644*)	(0.7146)	(2.7276*)	(1.8609)	(3.7584*)	(1.0007)	(3.5597*)	(1.8995)	(3.6588*)	(0.9634)	(3.4557*)
t_{proj}	-0.0071	0.0004	-0.0074	-0.0009	-0.0095	0.0000	-0.0097	-0.0013	-0.0091	0.0001	-0.0094	-0.0012
	(-2.4882*)	(1.2235)	(-2.2557*)	(-1.1601)	(-3.2454*)	(0.0658)	(-2.9026*)	(-1.5031)	(-3.1485*)	(0.1786)	(-2.8246*)	(-1.4794)
SN_{proj}	-0.0006	0.0069	-0.0009	0.0056	0.0040	0.0135	0.0037	0.0122	0.0025	0.0117	0.0022	0.0104
	(-0.4083)	(3.1984*)	(-0.4617)	(2.7332*)	(1.7579)	(4.8582*)	(1.4079)	(4.4337*)	(1.2591)	(4.6236*)	(0.9157)	(4.1704*)
St_{proj}	-0.0060	0.0015	-0.0063	0.0002	-0.0070	0.0025	-0.0073	0.0012	-0.0068	0.0023	-0.0071	0.0011
	(-2.4592*)	(1.5465)	(-2.3366*)	(0.3821)	(-2.7286*)	(2.0729*)	(-2.5741*)	(1.3512)	(-2.7040*)	(2.0527*)	(-2.5633*)	(1.3136)

Note: This table presents the average score difference $\bar{d}$ and the corresponding test statistic (the number in parentheses) of univariate densities against projected densities for scoring rules (2)-(4) for different empirical quantiles q in the indicator weight function $w_t(r_t) = I(r_t \leq \hat{r}_t^q p)$, where for the four sub-panels, $q = 0.01, 0.05, 0.1, 0.15$, respectively. The superscript * denotes that the test statistic is significant at the 5% level of significance.

Table 5: Evaluation of One-day-ahead VaR Results

	Univariate Modeling				Projected Modeling				Multivariate Modeling			
	coverage	p_{CUC}	p_{CCC}	t stat.	coverage	p_{CUC}	p_{CCC}	t stat.	coverage	p_{CUC}	p_{CCC}	t stat.
$q = 0.01$												
N	0.0111	0.3668	0.3525	-0.2755	0.0120	0.1101	0.0648	-0.6611	0.0120	0.1101	0.0648	-0.6873
t	0.0080	0.0909	0.0651	-0.2976	0.0076	0.0370	0.0250	0.2960	0.0076	0.0370	0.0250	0.2651
SN	0.0111	0.3668	0.3525	-1.0227	0.0101	0.9458	0.1935	0.1938	0.0101	0.9458	0.1935	0.1693
St	0.0088	0.2949	0.2271	0.0544	0.0096	0.7656	0.5910	0.3001	0.0096	0.7656	0.5910	0.2734
cl	0.0082	0.1187	0.0865	-0.1587	0.0077	0.0508	0.0350	0.5071	0.0076	0.0370	0.0250	0.4752
pwl	0.0123	0.0679	0.0669	1.0558	0.0132	0.0125	0.0121	0.0000	0.0132	0.0125	0.0121	1.0006
csl	0.0123	0.0679	0.0669	-	0.0132	0.0125	0.0121	-	0.0132	0.0125	0.0121	-
$q = 0.05$												
N	0.0443	0.0267	**0.0015**	-3.2765*	0.0432	**0.0088**	**0.0001**	-2.2739*	0.0432	**0.0088**	**0.0001**	-2.3566*
t	0.0481	0.4613	**0.0024**	-0.3436	0.0469	0.2350	**0.0007**	-0.9715	0.0469	0.2350	**0.0007**	-0.7552
SN	0.0450	0.0538	**0.0011**	-4.8656*	0.0378	**0.0000**	**0.0000**	-3.1530*	0.0378	**0.0000**	**0.0000**	-2.7415*
St	0.0492	0.7725	**0.0015**	-0.7333	0.0473	0.3086	**0.0023**	-1.8461*	0.0473	0.3086	**0.0023**	-1.6027
cl	0.0476	0.3653	0.0115	0.3445	0.0465	0.1746	**0.0036**	-0.5367	0.0466	0.1933	**0.0021**	0.6663
pwl	0.0494	0.8155	0.0409	0.4504	0.0506	0.8334	0.1666	-1.2784	0.0506	0.8334	0.1666	2.2185*
csl	0.0494	0.8155	0.0409	-	0.0503	0.9205	0.1549	-	0.0506	0.8334	0.1666	-
$q = 0.10$												
N	0.0830	**0.0000**	**0.0000**	-3.4011*	0.0828	**0.0000**	**0.0000**	-2.7017*	0.0828	**0.0000**	**0.0000**	-3.1207*
t	0.0954	0.2027	**0.0002**	-0.9657	0.0935	0.0708	**0.0000**	-0.9880	0.0935	0.0708	**0.0000**	-1.0245
SN	0.0840	**0.0000**	**0.0000**	-3.9061*	0.0754	**0.0000**	**0.0000**	-3.7668*	0.0754	**0.0000**	**0.0000**	-3.5593*
St	0.0959	0.2495	**0.0001**	-1.2980	0.0915	0.0171	**0.0002**	-2.0164*	0.0915	0.0171	**0.0002**	-2.2490*
cl	0.0947	0.1393	**0.0009**	0.1409	0.0923	0.0327	**0.0001**	-1.1621	0.0922	0.0295	**0.0001**	-0.5729
pwl	0.0944	0.1187	**0.0007**	2.6402*	0.0970	0.4090	0.0130	-1.2545	0.0977	0.5336	**0.0094**	0.4234
csl	0.0942	0.1093	**0.0008**	-	0.0967	0.3644	0.0113	-	0.0973	0.4567	**0.0075**	-
$q = 0.15$												
N	0.1265	**0.0000**	**0.0000**	-3.9748*	0.1236	**0.0000**	**0.0000**	-2.3177*	0.1236	**0.0000**	**0.0000**	-3.0833*
t	0.1448	0.2259	**0.0004**	-0.5218	0.1417	0.0535	**0.0001**	-0.2348	0.1417	0.0535	**0.0001**	-0.7853
SN	0.1268	**0.0000**	**0.0000**	-3.9140*	0.1184	**0.0000**	**0.0000**	-4.0342*	0.1184	**0.0000**	**0.0000**	-4.4127*
St	0.1452	0.2677	**0.0002**	-0.6109	0.1359	**0.0009**	**0.0000**	-1.2131	0.1359	0.0009	**0.0000**	-1.8858
cl	0.1430	0.1048	**0.0023**	0.2316	0.1397	0.0158	**0.0006**	-0.3925	0.1401	0.0209	**0.0004**	-0.7453
pwl	0.1425	0.0786	**0.0037**	1.6639	0.1419	0.0579	**0.0062**	-0.6927	0.1438	0.1469	0.0105	-0.3217
csl	0.1423	0.0729	**0.0042**	-	0.1413	0.0419	**0.0045**	-	0.1433	0.1204	**0.0089**	-

Note: This table presents the accuracy of one-day-ahead VaR estimates of the daily portfolio returns, obtained by assuming the densities to be Normal, t, skew Normal, skew t and three dynamically-selected densities based on the scoring rules *cl*, *pwl* and *csl*. 1%, 5%, 10% and 15% VaR results for the univariate modeling, projected modeling and multivariate modeling are reported, respectively. In each sub-panel for different quantiles, the table provides the corresponding coverage rate of VaR forecasts, the p-values of the CUC test and the CCC test and the t-statistic of the Diebold-Mariano test for equal conditional predictive ability against better *csl* rule based forecasts, using the loss function (12). Bold numbers highlight the p-values smaller than 1% and asterisks suggest that the t-statistic is significant at the 5% level.

Table 6: Evaluation of One-day-ahead VaR Results against the Multivariate Modeling

	Univariate Modeling							Projected Modeling						
	N_{uni}	t_{uni}	SN_{uni}	St_{uni}	cl_{uni}	pwl_{uni}	csl_{uni}	N_{proj}	t_{proj}	SN_{proj}	St_{proj}	cl_{proj}	pwl_{proj}	csl_{proj}
$q = 0.01$														
N_{tri}	2.2895*	0.8689	1.0888	0.9421	0.8927	1.3193	1.3176	-	-	-	-	0.8095	0.6611	0.6611
t_{tri}	0.2070	0.6525	-0.3451	1.0999	0.9263	0.5011	0.4989	-0.7436	-	-	-	0.9959	-0.296	-0.296
SN_{tri}	0.1901	0.2784	-0.212	0.5304	0.4118	0.4046	0.403	-0.5601	0.0583	-	-	0.3136	-0.1938	-0.1938
St_{tri}	0.1195	0.2134	-0.3098	1.0133	0.4016	0.3514	0.3497	-0.6305	-0.0923	-0.0922	-	0.3312	-0.3001	-0.3001
cl_{tri}	-0.0001	-0.093	-0.4895	0.6265	0.2554	0.2365	0.2344	-0.8215	-1.1865	-0.3086	-0.3224	0.0746	-0.5054	-0.5054
pwl_{tri}	0.5228	0.4401	-0.1093	0.6518	0.543	1.4545	1.4508	-0.6919	0.2583	0.164	0.2677	0.4706	-1.9660*	-1.9660*
csl_{tri}	0.5314	0.4469	-0.1025	0.6579	0.5497	1.4705	1.4669	-0.6873	0.2651	0.1693	0.2734	0.4772	-1.9430	-1.9430
$q = 0.05$														
N_{tri}	1.7816	3.3304*	0.6813	2.9049*	3.5407*	3.2473*	3.2418*	-	-	-	-	2.7446*	2.1698*	2.2739*
t_{tri}	-1.5868	0.6655	-2.6899*	0.3798	1.0111	0.7088	0.7033	-3.4239*	-	-	-	0.9612	0.5206	0.9715
SN_{tri}	1.4517	2.1904*	0.9869	2.0710*	2.3391*	2.2024*	2.2011*	0.6337	1.9247	-	-	2.7091*	2.6029*	3.1530*
St_{tri}	-0.2992	1.5994	-1.3265	1.7401	1.8920	1.8754	1.8711	-1.5425	1.3757	-1.4900	-	1.8708	1.3551	1.8461
cl_{tri}	-2.1173*	-0.9614	-2.7551*	-1.0760	-0.7493	-0.8220	-0.8255	-2.7309*	-1.2889	-3.2211*	-2.1543*	-1.0972	-1.0519	-0.0312
pwl_{tri}	-1.6948	-0.5722	-2.3367*	-0.6939	-0.3719	-0.4799	-0.4833	-2.4006*	-0.8181	-2.795*	-1.6754	-0.2457	-1.7429	0.5812
csl_{tri}	-1.6346	-0.5053	-2.2778*	-0.6249	-0.3030	-0.4089	-0.4123	-2.3566*	-0.7552	-2.7415*	-1.6027	-0.1454	-1.4121	0.6963
$q = 0.10$														
N_{tri}	1.8060	3.9432*	1.4240	3.7074*	4.5249*	4.2425*	4.1842*	-	-	-	-	2.9101*	2.5969*	2.7017*
t_{tri}	-2.1278*	0.1907	-2.5185*	-0.0887	0.9195	0.8374	0.7457	-4.2863*	-	-	-	0.3346	0.4456	0.9880
SN_{tri}	2.0823*	2.5990*	1.9195	2.5016*	2.8053*	2.8106*	2.7911*	1.3153	2.5102*	-	-	3.0545*	3.1646*	3.7668*
St_{tri}	-0.6473	1.9927*	-1.0387	1.9127	2.4599*	2.4955*	2.4209*	-2.1263*	2.2066*	-2.0390*	-	1.6426	1.5676	2.0164*
cl_{tri}	-2.4529*	-0.7571	-2.6842*	-0.8524	-0.4000	-0.3721	-0.4273	-3.1434*	-0.8382	-3.4194*	-2.1044*	-1.5335	-0.5055	0.6291
pwl_{tri}	-2.5653*	-1.0401	-2.7873*	-1.1345	-0.739	-0.7249	-0.7762	-3.169*	-1.092	-3.6343*	-2.3122*	-1.5232	-1.5986	0.2136
csl_{tri}	-2.5244*	-0.9778	-2.7477*	-1.0748	-0.6705	-0.6573	-0.7090	-3.1207*	-1.0245	-3.5593*	-2.2490*	-1.3178	-1.3739	0.3077
$q = 0.15$														
N_{tri}	0.7701	3.7424*	1.1181	3.6622*	4.1199*	4.1368*	4.0518*	-	-	-	-	2.7414*	2.3245*	2.3177*
t_{tri}	-3.1221*	-0.2513	-3.0337*	-0.3444	0.2719	0.3449	0.1590	-4.1144*	-	-	-	-0.0332	-0.0918	0.2348
SN_{tri}	2.7607*	3.4080*	2.8414*	3.3558*	3.5677*	3.5832*	3.5564*	2.3607*	3.3922*	-	-	3.7945*	3.7639*	4.0342*
St_{tri}	-1.9181	1.7089	-1.7436	1.6390	2.0583*	2.0914*	1.9460	-2.5391*	2.1949*	-2.9890*	-	1.2490	1.0066	1.2131
cl_{tri}	-2.9825*	-0.5176	-2.8382*	-0.5399	-0.302	-0.2414	-0.3446	-3.1773*	-0.426	-4.0955*	-1.7219	-1.1973	-0.7569	-0.1040
pwl_{tri}	-2.9375*	-0.7781	-2.8088*	-0.7943	-0.6196	-0.5714	-0.6643	-3.0593*	-0.6958	-4.3720*	-1.8307	-1.3165	-1.6186	-0.6743
csl_{tri}	-2.9955*	-0.869	-2.8695*	-0.8834	-0.7142	-0.6745	-0.7639	-3.0833*	-0.7853	-4.4127*	-1.8858	-1.1758	-1.4791	-0.7337

Note: This table presents the t-statistic of Diebold-Mariano type for the equal conditional predictive ability of pairs of competing VaR forecasts, using the loss function (12). Both univariate modeling and projected modeling methods are evaluated against direct multivariate modeling for different quantiles q. Asterisks indicate that the t-statistic is significant at 5% level.

Table 7: Evaluation of One-day-ahead VaR Results against the Projected Modelling

	Univariate Modeling							Projected Modeling						
	N_{uni}	t_{uni}	SN_{uni}	St_{uni}	cl_{uni}	pwl_{uni}	csl_{uni}	N_{proj}	t_{proj}	SN_{proj}	St_{proj}	cl_{proj}	pwl_{proj}	csl_{proj}
$q = 0.01$														
N_{proj}	2.2895*	0.8689	1.0888	0.9421	0.8927	1.3193	1.3176	-	-	-	-	-	-	-
t_{proj}	0.2070	0.6525	-0.3451	1.0999	0.9263	0.5011	0.4989	-0.7436	-	-	-	-	-	-
SN_{proj}	0.1901	0.2784	-0.2120	0.5304	0.4118	0.4046	0.4030	-0.5601	0.0583	-	-	-	-	-
St_{proj}	0.1195	0.2134	-0.3098	1.0133	0.4016	0.3514	0.3497	-0.6305	-0.0923	-0.0922	-	-	-	-
cl_{proj}	-0.0036	-0.0989	-0.4826	0.6042	0.2390	0.2302	0.2281	-0.8095	-0.9959	-0.3136	-0.3312	-	-	-
pwl_{proj}	0.5679	0.4778	-0.0698	0.6857	0.5800	1.5379	1.5343	-0.6611	0.2960	0.1938	0.3001	0.5071	-	-
csl_{proj}	0.5679	0.4778	-0.0698	0.6857	0.5800	1.5379	1.5343	-0.6611	0.2960	0.1938	0.3001	0.5071	0.0000	-
$q = 0.05$														
	Univariate Modelling							Projected Modelling						
	N_{uni}	t_{uni}	SN_{uni}	St_{uni}	cl_{uni}	pwl_{uni}	csl_{uni}	N_{proj}	t_{proj}	SN_{proj}	St_{proj}	cl_{proj}	pwl_{proj}	csl_{proj}
N_{proj}	1.7816	3.3304*	0.6813*	2.9049*	3.5407*	3.2473*	3.2418*	-	-	-	-	-	-	-
t_{proj}	-1.5868	0.6655	-2.6899	0.3798	1.0111	0.7088	0.7033	-3.4239*	-	-	-	-	-	-
SN_{proj}	1.4517	2.1904*	0.9869	2.0710*	2.3391*	2.2024*	2.2011*	0.6337	1.9247	-	-	-	-	-
St_{proj}	-0.2992	1.5994	-1.3265	1.7401	1.8920	1.8754	1.8711	-1.5425	1.3757	-1.4900	-	-	-	-
cl_{proj}	-1.9566	-0.5528	-2.7194*	-0.6816	-0.2748	-0.3770	-0.3814	-2.7446*	-0.9612	-2.7091*	-1.8708	-	-	-
pwl_{proj}	-1.4095	-0.2565	-2.0600*	-0.3709	-0.0461	-0.1434	-0.147	-2.1698*	-0.5206	-2.6029*	-1.3551	0.2255	-	-
csl_{proj}	-1.7579	-0.7716	-2.3343*	-0.8953	-0.5983	-0.6989	-0.7025	-2.2739*	-0.9715	-3.1530*	-1.8461	-0.5367	-1.2784	-
$q = 0.10$														
	Univariate Modeling							Projected Modeling						
	N_{uni}	t_{uni}	SN_{uni}	St_{uni}	cl_{uni}	pwl_{uni}	csl_{uni}	N_{proj}	t_{proj}	SN_{proj}	St_{proj}	cl_{proj}	pwl_{proj}	csl_{proj}
N_{proj}	1.8060	3.9432*	1.4240	3.7074*	4.5249*	4.2425*	4.1842*	-	-	-	-	-	-	-
t_{proj}	-2.1278*	0.1907	-2.5185*	-0.0887	0.9195	0.8374	0.7457	-4.2863*	-	-	-	-	-	-
SN_{proj}	2.0823*	2.5990*	1.9195	2.5016*	2.8053*	2.8106*	2.7911*	1.3153	2.5102*	-	-	-	-	-
St_{proj}	-0.6473	1.9927*	-1.0387	1.9127	2.4599*	2.4955*	2.4209*	-2.1263*	2.2066*	-2.0390*	-	-	-	-
cl_{proj}	-2.0214*	-0.2417	-2.2791*	-0.3589	0.1795	0.1973	0.1382	-2.9101*	-0.3346	-3.0545*	-1.6426	-	-	-
pwl_{proj}	-1.8425	-0.3822	-2.0618*	-0.4796	-0.0588	-0.0400	-0.0886	-2.5969*	-0.4456	-3.1646	-1.5676	-0.3425	-	-
csl_{proj}	-2.2467*	-0.9485	-2.4274*	-1.0231	-0.6921	-0.7073	-0.7487	-2.7017*	-0.988	-3.7668*	-2.0164*	-1.1621	-1.2545	-
$q = 0.15$														
p	Univariate Modeling							Projected Modeling						
	N_{uni}	t_{uni}	SN_{uni}	St_{uni}	cl_{uni}	pwl_{uni}	csl_{uni}	N_{proj}	t_{proj}	SN_{proj}	St_{proj}	cl_{proj}	pwl_{proj}	csl_{proj}
N_{proj}	0.7701	3.7424	1.1181	3.6622*	4.1199*	4.1368*	4.0518*	-	-	-	-	-	-	-
t_{proj}	-3.1221*	-0.2513	-3.0337*	-0.3444	0.2719	0.3449	0.159	-4.1144*	-	-	-	-	-	-
SN_{proj}	2.7607*	3.4080*	2.8414*	3.3558*	3.5677*	3.5832*	3.5564*	2.3607*	3.3922*	-	-	-	-	-
St_{proj}	-1.9181	1.7089	-1.7436	1.6390	2.0583*	2.0914*	1.9460	-2.5391*	2.1949*	-2.9890*	-	-	-	-
cl_{proj}	-2.4699*	-0.0694	-2.3289*	-0.1042	0.1770	0.2256	0.1236	-2.7414*	0.0332	-3.7945*	-1.2490	-	-	-
pwl_{proj}	-2.0900*	0.0045	-1.9571	-0.0268	0.2197	0.2662	0.1774	-2.3245*	0.0918	-3.7639*	-1.0066	0.1102	-	-
csl_{proj}	-2.2038*	-0.3093	-2.0724*	-0.3303	-0.1465	-0.1090	-0.1871	-2.3177*	-0.2348	-4.0342*	-1.2131	-0.3925	-0.6927	-

Note: This table presents the t-statistic of Diebold-Mariano type for the equal conditional predictive ability of pairs of competing VaR forecasts, using the loss function (12). Both univariate modeling and projected modeling methods are evaluated against the projected modeling for different quantiles q. Asterisks suggest that the t-statistic is significant at 5% level.