CIRJE-F-369

Scanning Multivariate Conditional Densities with Probability Integral Transforms

Isao Ishida
University of Tokyo

September 2005

Scanning Multivariate Conditional Densities with Probability Integral Transforms*

Isao Ishida
University of Tokyo

September 2005

Abstract

This paper introduces new ways to construct probability integral transforms of random vectors that complement the approach of Diebold, Hahn, and Tay (1999) for evaluating multivariate conditional density forecasts. Our approach enables us to "scan" multivariate densities in various different ways. A simple bivariate normal example is given that illustrates how "scanning" a multivariate density from particular angles leads to tests with no power or high power. An empirical example is also given that applies several different probability integral transforms to specification testing of Engle's (2002) dynamic conditional correlation model for multivariate financial returns time series with multivariate normal and t errors.

Key words: Density forecast evaluation, multivariate conditional density, probability integral transform, multivariate GARCH

1 Introduction

For economic decision makers, having a good forecast of the entire probability distribution (equivalently the entire probability density if it exists) of the relevant variables is helpful. However, until recently it was considered a very difficult task to assess the quality of such forecasts in non-i.i.d. time series settings since probability distributions are not observed even ex post. Given the importance of distributional forecasts, it is not surprising that the method of probability integral

*This paper is based on Chapter 3 of my Ph.D. dissertation, Ishida (2004). I would like to thank Jin Seo Cho, Jin-Chuan Duan, Graham Elliott, Rob Engle, Jim Hamilton, Bruce Lehmann, Andrew Patton, Allan Timmermann, and Halbert White for their comments and suggestions. Contact information: Graduate School of Public Policy and Faculty of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-Ku, Tokyo, 113-0033, Japan. Email: ishida@e.u-tokyo.ac.jp.

transforms (PITs) for conditional density forecast evaluation have quickly become a widely used statistical tool in empirical work since Diebold, Gunther, and Tay (1998) introduced it to the econometric mainstream; See, e.g., Bauwens et al. (2000), Andersen et al. (2003), and Bliss and Panigirtzoglou (2004) for financial applications. The econometrics of forecast density evaluation is still nascent, but has been an area of intense research in the past few years and the literature has already made impressive strides. Some of the theoretical advances include Diebold, Hahn, and Tay (1999), Berkowitz (2001), Hong, Li, and Zhao (2002), Hong and Li (2004), Giacomini (2002), Giacomini and White (2003), Bao, Lee, and Saltoğlu (2003), Corradi and Swanson (2004), Bai (2003), Duan (2003).

More often than not, economic decisions are multivariate ones requiring joint distributional forecasts for a set of random variables. Diebold, Hahn, and Tay (1999) (DHT) have proposed to apply the PIT approach for multivariate density forecast evaluation problems. They suggest a decomposition of each period's joint density into a marginal density for the first variable and a sequence of conditional densities. Intuitively speaking, their method amounts to viewing each period's multivariate density from one set of particular angles. However, when viewing a multidimensional object such as a multivariate density, some angles are revealing while others hide the true shape. This paper demonstrates with examples that it is not particularly advantageous to adhere to the DHT transforms and that there are alternative ways to construct probability integral transforms to "scan" multivariate densities from different angles that lead to more powerful tests of model specifications and forecast accuracy.

The rest of the paper is organized as follows. Section 2 briefly reviews the approach of DHT. Section 3 introduces our alternative approach with several examples. Section 4 discusses a potential application of our approach to path density forecast evaluation. Section 5 contains a real data application of our approach to testing the specification of a bivariate exchange rate example of Engle's (2002) dynamic conditional correlation model using Duan's (2003) testing procedure. Section 6 concludes.

2 A brief review of the dynamic probability integral approach

Suppose that we have a series of one-period-ahead multivariate conditional density forecasts $\left\{\widehat{f}_{Y_t|\mathcal{F}_{t-1}}\left(y \mid \mathcal{F}_{t-1}\right)\right\}$ of an N-dimensional time series $\{Y_t\}$ where $\mathcal{F}_{t-1} \supset \sigma\left\{Y_{t-1}, Y_{t-2}, \cdots\right\}$ is the information set available at $t-1$. In what follows, we suppress $\mathcal{F}_{t-1}$ out of our notations for conditional density functions and cumulative distribution functions, and simply write $\widehat{f}_{Y_t}(y)$ for $\widehat{f}_{Y_t|\mathcal{F}_{t-1}}\left(y \mid \mathcal{F}_{t-1}\right)$, $\widehat{f}_{Y_{2t}|Y_{1t}}\left(y_2 \mid Y_{1t}\right)$ for $\widehat{f}_{Y_{2t}|\sigma(Y_{1t})\vee\mathcal{F}_{t-1}}\left(y_2 \mid \sigma\left(Y_{1t}\right) \vee \mathcal{F}_{t-1}\right)$, and so forth. Diebold, Gunther, and Tay (1998) (hereafter DGT) have recently popularized the approach that relies on Rosenblatt's (1957) dynamic probability integral transforms (PITs)

$\left\{U_t := \int_{-\infty}^{Y_t} \widehat{f}_{Y_t}(y)\, dy\right\}$ for testing the accuracy of these forecasts when $N = 1$. Central to the approach is the fact that if $\widehat{f}_{Y_t}(y) = f_{Y_t}(y)$ for all y and all t where $f_{Y_t}(y)$ is the true density of Y_t conditional on $\mathcal{F}_{t-1}$ (which is assumed to exist) then $\{U_t\}$ is a sequence of i.i.d. $U(0,1)$ (uniform over the interval (0,1)) random variables (equalities involving conditional densities should be read $=_{a.s.}$). A host of formal tests such as the Kolmogorov-Smirnov test of uniformity and the Ljung-Box test of no serial correlation then become available. We may also visually inspect histograms of $\{U_t\}$ to detect possible forecast failures (a method recommended by DGT).

Extending the approach of DGT to multivariate time series settings, Diebold, Hahn, and Tay (1999) (hereafter DHT) proposed a particular set of PITs that transform an N-dimensional time series $\{Y_t = (Y_{1t}, \cdots, Y_{Nt})\}$ into another N-dimensional series $\left\{U_t^{DHT} = \left(U_{1t}^{DHT}, \cdots, U_{Nt}^{DHT}\right)\right\}$ using multivariate density forecasts $\left\{\widehat{f}_{Y_t}(y)\right\}$. Specifically,

$$U_{1t}^{DHT} \;:=\; \int_{-\infty}^{Y_{1t}} \widehat{f}_{Y_{1t}}(y)\, dy, \tag{1}$$

$$U_{it}^{DHT} \;:=\; \int_{-\infty}^{Y_{it}} \widehat{f}_{Y_{it}|Y_{1t},\cdots,Y_{i-1,t}}(y \mid Y_{1t}, \cdots, Y_{i-1,t})\, dy \quad \text{for } i \geq 2. \tag{2}$$

U_t^{DHT} for each t is a vector of N independent $U(0,1)$ random variables and $\left\{U_t^{DHT}\right\}$ is an i.i.d. series if the density forecast is correct for each t. We will be referring to the DHT transform[1] as $\left(U_{1t}^{DHT}, \cdots, U_{Nt}^{DHT}\right)$, while keeping in mind that there are $N!$ ways to construct U_t^{DHT} since U_t^{DHT} depends on the order in which the joint density is factored into a multiple of conditional densities.

In this paper, we demonstrate that many other PITs (in most cases, an infinity of them) for a multivariate time series producing a univariate i.i.d. uniform (0,1) time series under the null of correct density forecasts are available under certain conditions. Alternative PITs produced by our framework may serve as building blocks for designing tests of multivariate density forecast accuracy that are more powerful than the tests based on DHT's transformation.

3 An alternative approach to producing multivariate probability integral transforms

First, consider an N-dimensional time series $\{Y_t\}$ and its transform

$$\{W_t = (W_{1t}, \cdots, W_{Mt})\},$$

[1] Although we will be using the term the *DHT transform*, we note that this transform due to Rosenblatt (1952) had been considered by other authors in the statistics literature prior to Diebold, Hahn, and Tay (1999) in the context of goodness-of-fit tests for models of densities; See, e.g., Quesenberry (1986). The use of probability integral transforms for goodness-of-fit tests has a longer history; See, e.g., Pearson (1933).

an M-dimensional time series where $W_{mt} = g_{mt}(Y_t)$, $m = 1, \cdots, M$. We allow the *transform function* $g_{mt}(\cdot) : \mathbb{R}^N \to \mathbb{R}$ to be stochastic but require it to be known at time $t-1$, i.e., $g_{mt}(y) \in \mathcal{F}_{t-1}$ for all y. We propose to further transform W_{mt} to U_{mt}, $m = 1, \cdots, M$, as follows:

$$\begin{aligned} U_{mt} \; &:= \widehat{F}_{W_{mt}}(W_{mt}) \\ &:= \int_{\{y \in \mathbb{R}^N:\ g_{mt}(y) \leq W_{mt}\}} \widehat{f}_{Y_t}(y)\, dy. \end{aligned} \tag{3}$$

where $\widehat{f}_{Y_t|\mathcal{F}_{t-1}}(y)$ for each t is our forecast of the joint (N-dimensional) density $f_{Y_t|\mathcal{F}_{t-1}}(y)$ of Y_t conditional on $\mathcal{F}_{t-1}$ (assumed to exist).

Proposition: Consider any particular m. Suppose that the transform is such that, for all t, W_{mt} has a conditional density (conditional on $\mathcal{F}_{t-1}$). Then, $\{U_{mt}\}$ is a sequence of i.i.d. $U(0,1)$ random variables if $\widehat{f}_{Y_t}(y) = f_{Y_t}(y)$ for all y and all t.

Proof: If $\widehat{f}_{Y_t}(y) = f_{Y_t}(y)$ for all y and all t, then $\widehat{F}_{W_{mt}}(\cdot)$ is the true conditional c.d.f. of W_{mt} (conditional on $\mathcal{F}_{t-1}$) $\forall t$, and U_{mt}'s are identically $U(0,1)$. A proof essentially identical to the proof of the independence part of Lemma 1 in Bai (2003, p.537) may be applied here to show the independence of U_{mt}'s. $\square$

Remarks:

(1) Our method simply transforms the original random vector Y_t for each t into U_{mt}, by way of a probability integral transform of an intermediate random variable W_{mt} ensuring that $\{U_{mt}\}$ is an i.i.d. $U(0,1)$ series under correct forecasts. Figuratively speaking, using different "scanning" functions for transforming Y_t into W_{mt} is like viewing the density forecast $\widehat{f}_{Y_t}$ from different directions.

(2) The assumptions that the time index set for $\{Y_t\}$ consists of equally-spaced points in time and that the dimension N of Y_t is fixed through time are for notational convenience only. Just as with the original DHT transformation, the time index set may consist of irregularly-spaced points in time instead. We may also allow the dimension to change through time.

(3) The Jacobian of the transformation from Y_t to $U_t := (U_{1t}, \cdots, U_{Mt})$ is nonvanishing and U_t has an M-dimensional joint density if the Jacobian of the transformation from Y_t to $W_t := (W_{1t}, \cdots, W_{Mt})$ is nonvanishing. If we do not care whether the transformed random vector has a conditional joint density or not, we can make the dimension M of U_t arbitrarily large. It would be interesting to explore ways in which we may combine the information contained in more than one transformed series, perhaps even a continuum of them, but it is beyond the scope of this paper. The DHT approach produces N univariate i.i.d. series that are by construction cross-sectionally independent as well under the null of correct forecasts so that one can simply stack N series into $\left(U_{11}^{DHT}, \cdots, U_{TN}^{DHT}\right)'$, a univariate i.i.d. $U(0,1)$ series of length NT. As demonstrated by Hong and Li (2004) in the context of specification testing for interest rate models, however, tests based on a long, stacked series constructed this way may be less powerful than those based on a single series $\left(U_{n1}^{DHT}, \cdots, U_{nT}^{DNT}\right)$ for some n; See also Clements and Smith (2001).

(4) If we set $g_{1t}(y) = y_1$, then

$$\begin{aligned}
&\int_{\{y\in\mathbb{R}^N:\ g_{1t}(y)\le Y_{1t}\}} \widehat{f}_{Y_t}(y)\,dy \\
=\ &\int_{-\infty}^{Y_{1t}}\int_{\{(y_2,\cdots,y_N)\in\mathbb{R}^{N-1}\}} \widehat{f}_{(Y_{2t},\cdots,Y_{Nt})|Y_{1t}}(y_2,\cdots,y_N \mid Y_{1t})\,\widehat{f}_{Y_{1t}}(y_1)\,dy_1\cdots dy_N \\
=\ &\int_{-\infty}^{Y_{1t}} \widehat{f}_{Y_t}(y_1)\,dy_1 = U_{1t}^{DHT}.
\end{aligned}$$

Note also that our method does not require the user to explicitly decompose the joint density forecast into conditional densities.

(5) Applying the DHT transform method to our intermediate random vector W_t is another possibility:

$$U_{nt}^{W} := \int_{-\infty}^{W_{nt}} \widehat{f}_{W_{nt}|W_{1t},\cdots,W_{n-1,t}}(w \mid W_{1t},\cdots,W_{n-1,t})\,dw.$$

3.1 Example 1: Linear transform function

Write $y = (y_1, y_2, \cdots, y_N)$, and define, for $m = 1, \cdots, M \le N$, $g_{mt}(y) := a_m \cdot y := \sum_{i=1}^{N} a_{mi} y_i$ where $a_m := (a_{m1}, a_{m2}, \cdots, a_{mN}) \in \mathbb{R}^N$ and $a_m \neq 0$ and $W_{mt} := g_{mt}(Y_t)$

$$U_{mt} := \int_{\{a_m\cdot y\ \le\ a_m\cdot Y_t\}} \widehat{f}_{Y_t}(y)\ dy \tag{4}$$

Under correct density forecasts, $\{U_{mt}\}$ is an $U(0,1)$ i.i.d. sequence for each n. Furthermore, if $M = N$ and the square matrix $A := (a_{mi})$ is nonsingular, then $\{U_{1t}, \cdots, U_{Nt}\}_{t=1}^{T}$ is an i.i.d. series with some joint N-dimensional density. Figure 1 illustrates this transform for the case of $N = 2$ and $M = 1$. We note that, for the case of $N = 2$ with $a_{11} = 1$, $a_{12} = 0$, we obtain $U_{1t} = U_{1t}^{DHT}$.

Next, consider a simple special case in which the true conditional densities are i.i.d. bivariate normal:

$$Y_t \sim i.i.d.N(0, \Sigma_0) \text{ where } \Sigma := \begin{pmatrix} 1 & \rho_0 \\ \rho_0 & 1 \end{pmatrix}. \tag{5}$$

Suppose that the forecaster has the correct knowledge of this except for the value of the correlation coefficient ρ_0, and hence uses an estimate $\widehat{\rho}$ for ρ_0. For this case,

$$U_t^a(\widehat{\rho}) := \int_{\{y_1+ay_2\ \le\ Y_{1t}+aY_{2t}\}} \widehat{f}_{Y_t}(y)\,dy = \Phi\left(\frac{Y_{1t}+aY_{2t}}{\sqrt{1+a^2+2a\widehat{\rho}}}\right) \tag{6}$$

where Φ is the standard normal c.d.f. Since $\frac{Y_{1t}+aY_{2t}}{\sqrt{1+a^2+2a\widehat{\rho}}} \sim N\left(0, \frac{1+2a\rho_0+a^2}{1+2a\widehat{\rho}+a^2}\right)$, for $a \neq 0$, $U_t^a(\widehat{\rho}) \sim i.i.d.U(0,1)$ if and only if $\widehat{\rho} = \rho_0$. But $U_t^{a=0}(\widehat{\rho}) = \Phi(Y_{1t}) \sim$

$i.i.d.U\left(0,1\right)$ regardless the values of $\widehat{\rho}$ and ρ_0 under the maintained hypothesis of this example. So any test based on $\left\{U_t^{a=0}\left(\widehat{\rho}\right)\right\}$ completely lacks power. Since

$$U_{1t}^{DHT}\left(\widehat{\rho}\right) := \int_{-\infty}^{Y_{1t}} \widehat{f}_{Y_{1t}}\left(y\right)\ dy = \Phi\left(Y_t\right) = U_t^{a=0}\left(\widehat{\rho}\right), \tag{7}$$

tests based solely on $U_{1t}^{DHT}\left(\widehat{\rho}\right)$ have no power at all. We also have

$$\begin{aligned} U_{2t}^{DHT}\left(\widehat{\rho}\right) &= \int_{-\infty}^{Y_{2t}} \widehat{f}_{Y_{2t}|Y_{1t}}\left(y \mid Y_{1t}\right)\ dy \\ &= \int_{-\infty}^{Y_{2t}} \left(2\pi\right)^{-1/2}\left(1-\widehat{\rho}^2\right)^{-1/2} \exp\left(-\frac{\left(y-\widehat{\rho}Y_1\right)^2}{2\left(1-\widehat{\rho}^2\right)}\right) dy \\ &= \Phi\left(\frac{Y_{2t}-\widehat{\rho}Y_{1t}}{\sqrt{1-\widehat{\rho}^2}}\right) \sim i.i.d.U\left(0,1\right). \end{aligned} \tag{8}$$

So $U_{2t}^{DHT}\left(\widehat{\rho}\right) = U_t^{a=-\widehat{\rho}}\left(\widehat{\rho}\right)$.

Since $Var\left(Y_{1t}\right) = Var\left(Y_{2t}\right) = 1$ under the maintained hypothesis, the equi-probability contours of the forecast joint density $\widehat{f}_{Y_t}\left(y\right)$ are ellipses with major and minor axes having slopes of 1 and -1 respectively on the $\left(y_1, y_2\right)$ plane when $\widehat{\rho} > 0$ (flip the signs when $\widehat{\rho} < 0$) and we would expect tests based on $U_t^{a=1}\left(\widehat{\rho}\right)$ and $U_t^{a=-1}\left(\widehat{\rho}\right)$ to have power. To verify this by simulation, we generated 20,000 sample paths of i.i.d bivariate, unit-variance normal series with various values of ρ. For each of the simulated sample paths, we calculated the Kolmogorov-Smirnov (KS) statistics for the null hypotheses of $U_t^a\left(\widehat{\rho}\right) \sim i.i.d.U\left(0,1\right)$, for various values of a and $\widehat{\rho}$, all of which are equivalent to the null of $\widehat{\rho} = \rho$. Figure 3 exhibits the power function (at 5% size) of the Kolmogorov-Smirnov statistics for $a = -1.0, -.8, \cdots, 1.0$ and $\widehat{\rho} = -.99$ (the curve labeled "Contour" will be explained in the next subsection). As we have already shown analytically, the KS statistic based on $U_t^{a=0}\left(-.99\right)$ for $H_0 : \widehat{\rho} = \rho$ has no power (always rejects H_0 at 5%). When $a = 1$, the graph of $y_1 + ay_2 = Y_{1t} + aY_{2t}$ on the $\left(y_1, y_2\right)$ plane (for given values of Y_{1t} and Y_{2t}) is a line orthogonal to the major axis of the elliptical equi-probability contours of $\widehat{f}_{Y_t}\left(y\right)$. Therefore, it is no surprise that the KS test based on $U_t^{a=1}\left(-.99\right)$ appears to be uniformly most powerful among the ones considered here. On Figures 4-17, we see that the same conclusion holds as long as $\widehat{\rho} < 0$ and that the power properties of the KS tests with $U_t^{a=1}\left(\widehat{\rho}\right)$ and $U_t^{a=-1}\left(\widehat{\rho}\right)$ are reversed when $\widehat{\rho} > 0$. Note that $\widehat{\rho}$ is of course an observable quantity for the forecaster and that a is a quantity that he can choose. So, for this simple example, the forecaster should set $a = 1$ ($a = -1$) if $\widehat{\rho} < 0$ ($\widehat{\rho} > 0$). If $\{Y_t\}$ is a financial time series, it is likely that the conditional variance of each element is time-varying, and this conclusion may not hold.

As another example of linear transform, consider a multiperiod portfolio selection problem with rebalancing. A vector of portfolio weights is calculated each period based on the available information, and hence stochastic but known at the start of each period. So testing the i.i.d. $U(0,1)$ property of the PITs of the returns time series of the optimal portfolio is one relevant way to scan the multivariate density forecasts.

3.2 Example 2: Using the forecast densities for transformation

As an example of a series of stochastic function $\{g_t(\cdot)\}$ ($g_t(\cdot)$ known at time $t-1$), consider the conditional multivariate density forecasts themselves that are being evaluated. $W_t := g_t(y) := \widehat{f}_{Y_t}(y)$ may be used to produce

$$U_t^{Contour} := \int_{\left\{\widehat{f}_{Y_t}(y)\ \leq\ \widehat{f}_{Y_t}(Y_t)\right\}} \widehat{f}_{Y_t}(y)\,dy \tag{9}$$

See Figure 2. A simple example of this is the case in which $\widehat{f}_{Y_t}(\cdot)$ for each t is a conditionally multivariate normal density with conditional mean $\widehat{\mu}_t$ and conditional covariance matrix $\widehat{\Sigma}_t$:

$$\widehat{f}_{Y_t}(y) = (2\pi)^{-N/2}\left|\widehat{\Sigma}_t\right|^{-1/2}\exp\left((y-\widehat{\mu}_t)'\,\widehat{\Sigma}_t^{-1}(y-\widehat{\mu}_t)\right) \tag{10}$$

When $\widehat{f}_{Y_t}$ belongs to the class of elliptical distributions, we have

$$U_t^{Contour} = \int_{\left\{(y-\widehat{\mu}_t)'\widehat{\Sigma}_t^{-1}(y-\widehat{\mu}_t)\geq(Y_t-\widehat{\mu}_t)'\widehat{\Sigma}_t^{-1}(Y_t-\widehat{\mu}_t)\right\}} \widehat{f}_{Y_t}(y)\,dy. \tag{11}$$

When $\widehat{f}_{Y_t}$ is multivariate normal as in our case here, the calculation is straightforward since

$$U_t^{Contour} = 1 - F_{\chi^2_N}\left((Y_t-\widehat{\mu}_t)'\,\widehat{\Sigma}_t^{-1}(Y_t-\widehat{\mu}_t)\right) \tag{12}$$

where $F_{\chi^2_N}(\cdot)$ is the cumulative distribution function of a χ^2 r.v. with N degrees of freedom. When $\widehat{f}_{Y_t}$ is multivariate t of a particular kind, $U_t^{Contour}$ is similarly easy to calculate (See Section 5).

As a further simplification, consider the simple bivariate normal case of Example 1 in which the true DGP is (5). The curve labeled "Contour " on each of Figures 4-17 represents the power function (size 5%) for the KS statistic based on $U_t^{Contour}(\widehat{\rho}) := 1 - F_{\chi^2_2}\left(\frac{Y_1^2-2Y_1Y_2\rho+Y_2^2}{1-\widehat{\rho}^2}\right)$, obtained by simulation as in Example 1. We see that, although $U_t^{Contour}(\widehat{\rho})$ has high power when $|\widehat{\rho}|\approx 1$, the power deteriorates (relative to the most powerful linear transformation) as $|\widehat{\rho}|$ gets closer to zero. We note, however, that tests based on $U_t^{a=-1}(\widehat{\rho})$ of the previous subsection (with the true μ unknown to the forecaster) does not have any power against deviations (parallel to the minor axis of the equi-probability contours) of $\widehat{\mu}$ from the true μ, while $U_t^{Contour}$ is sensitive to this type of forecast errors as well.

4 Application to the evaluation of path density forecasts

We may have a series of multiperiod density forecasts (a "path density" forecast) for a time series. For example, in pricing path-dependent derivatives, we care about the probability distribution for the whole trajectory that the value of the underlying goes through before maturity. Of course, if a series of single-period-ahead forecasts $\left\{\widehat{f}_{Y_t}(y)\right\}$ are correct, then multiperiod-ahead, single-period density forecasts and multiperiod path density forecasts implied by $\left\{\widehat{f}_{Y_t}(y)\right\}$ are correct as well. However, parametric models generating density forecasts, for example, may have errors (due to misspecification and/or parameter estimation uncertainty) that manifest themselves differently depending on the forecast horizon. So just as we are sometimes interested in multiperiod-ahead point forecasts, we may be interested in directly evaluating path density forecasts (or multiperiod-ahead, single-period density forecasts).

Suppose that we forecast the next K-period path density. Consider a univariate time series $\{Y_t\}$. Since we can reshape it into an K-dimensional time series $\{Y_t^*\}_{t=1,2,\cdots}$ (a series of blocks of length N) where $Y_t^* = (Y_t, Y_{t+1}, \cdots, Y_{t+K-1})$, we may use our PIT framework in designing tests for "path density" forecast accuracy.

$$U_t := \int_{g_t(y) \leq g_t(Y_t^*)} \widehat{f}_{Y_t^*|\mathcal{F}_{t-1}}(y \mid \mathcal{F}_{t-1})\, dy \tag{13}$$

where as before $\mathcal{F}_{t-1} \supset \sigma(Y_{t-1}, Y_{t-2}, \cdots)$ is the information set available to the forecaster at time $t-1$. Assuming that g_t is known at time $t-1$ and that $\widehat{f}_{Y_t^*|\mathcal{F}_{t-1}}(y \mid \mathcal{F}_{t-1})$ implies a density for $W_t = g_t(Y_t^*)$, $\{U_t\}_{t=1,2,\cdots}$ will be an $MA(K-1)$ process with an identical marginal $U(0,1)$ distribution, and $\{U_t\}_{t\in\mathcal{T}}$ for $\mathcal{T} = \{k, k+K, \ k+2K, \cdots\}$ for any $k \in \mathbb{N}$ will be a sequence of i.i.d. $U(0,1)$ random variables, under the null of correct forecasts. If $\{Y_t\}$ is an N-dimensional multivariate time series instead of a univariate one, nothing basically changes in the above argument. In such a case, the forecast path density functions and $g_t(Y_t^*)$ are NK-variate functions. We may also consider Y_t^* consisting of a subset of $(Y_t, Y_{t+1}, \cdots, Y_{t+K-1})$. For example, if we set $Y_t^* := Y_{t+N-1}$, the problem becomes an N-period-ahead, single-period density forecast evaluation one. It is usually difficult to explicitly derive $\widehat{f}_{Y_t^*|\mathcal{F}_{t-1}}$ implied by a dynamic model defined in terms of a system of difference equations such as an ARMA or a GARCH model. In some cases, we may nevertheless be able to approximate PITs by simulating paths as in Rosernberg and Engle (2002).

4.1 Example 3: Volatility of multiperiod returns

Consider a financial return series $\{Y_t\}$. Suppose that we are interested in testing the accuracy of our forecasts for the conditional variances of K-period returns,

$Var_{t-1}\left(\sum_{s=0}^{K-1} Y_{t+s}\right)$ where, for a random variable X,

$$
\begin{aligned}
Var_{t-1}(X) &= E_{t-1}\left[(X - E_{t-1}[X])^2\right], \\
E_{t-1}[X] &:= E[X \mid \mathcal{F}_{t-1}]
\end{aligned}
$$

with $\mathcal{F}_{t-1}$ being the conditioning information set available at $t-1$. In comparing the predictive accuracy of the long-horizon (up to 250 trading days) predictions for the S&P 500 volatility implied by the standard GARCH(1,1) model with parameters frequently re-estimated with past data versus those based a simple moving average of squared past returns, Stărică (2003) applied the MSE criterion to the error, i.e., the difference between the K-period conditional variance and the so called "realized volatility" $\sum_{s=0}^{K-1} Y_{t+s}^2$ proxying for $Var_{t-1}\left(\sum_{s=0}^{K-1} Y_{t+s}\right)$ (see, e.g., Andersen and Bollerslev (1997) and Andersen et al. (2003)).

One problem with this approach is that time series models with finite (unconditional) variances, let alone those with finite fourth moments, may not be suitable for the daily returns on the S&P 500 or other equity indices, a possibility raised by the empirical evidence of the literature (including Stărică's (2003) own and Engle and Ishida (2002)), in which case the MSE criterion is inappropriate. When comparing the accuracy of path density forecasts using the PITs described in this section, we do not encounter this problem. For example, we may use $g_t(y_t, y_{t+1}, \cdots, y_{t+K-1}) := \sum_{s=1}^{K} y_{t+s-1}^2$ as the transform function, and obtain $U_t = \int_{-\infty}^{W_t} \widehat{f}_{W_t}(w)\, dw$ (again $\mathcal{F}_t$ is suppressed out of our notation). Interestingly, $W_t := \sum_{s=1}^{N} Y_{t+s}^2$ is the realized volatility and $\widehat{f}_{W_t}(w)$, $w \in \mathbb{R}^+$, is the implied density (conditional on $\mathcal{F}_t$) for the realized volatility. So checking the accuracy of the realized volatility density forecasts implied by a volatility model is one particular way of checking the accuracy of its path density forecasts.

5 Application to real data

In this section, we demonstrate the use of alternative PITs as well as the original DHT transforms in constructing statistics to test the specification of Engle's (2002) dynamic conditional correlation (DCC) model. The real data for our analysis are from a bivariate exchange rate series: the daily log difference series of the US dollar/Japanese yen and the US dollar/British pound exchange rates, October 12, 1983 - December 31, 2004, obtained from Datastream (4937 observations).

The DCC model relaxes the restrictive assumption of constant conditional correlation imposed in Bollerslev's (1990) constant conditional correlation (CCC) model, while not sacrificing the CCC's parsimony too much and keeping under control the curse of higher dimension that plagues some of the other multivariate GARCH models such as the VEC model and the BEKK model (see Bauwens, Laurent, and Rombouts (2003) for an extensive survey of multivariate GARCH models). The DCC model essentially specifies a univariate GARCH model for each of the conditional variance processes and a univariate GARCH-like model for the conditional correlation process of each pair of assets being studied. We consider a special case of the first-order DCC model that may be expressed formally as follows:

$$Y_t = \mu_t + \epsilon_t$$

where Y_t is an N-dimensional multivariate time series of our interest, and $\mu_t := E_{t-1}[Y_t]$. Although $N = 2$ in our case, we keep the notation N, anticipating a larger empirical study with more than two exchange rate series. The conditional variance of each component series of $\{Y_t\}$ is assumed to satisfy the standard GARCH(1,1) equation:

$$h_{it} := E_{t-1}\left[\epsilon_{it}^2\right] = (1-\alpha_i-\beta_i)\,\sigma_i^2 + \beta_i h_{it} + \alpha_i \epsilon_{it}^2, \quad i = 1, \cdots, N, \tag{14}$$

where ϵ_{it} is the i-th element of ϵ_t, $\sigma_i^2 := Var(\epsilon_{it})$, and α_i, β_i are nonnegative scalar parameters with the restriction $\alpha_i + \beta_i < 1$. Let $D_t := diag\left(h_{1t}^{1/2}, \cdots, h_{Nt}^{1/2}\right)$ and $\eta_t := D_t^{-1}\epsilon_t$. The DCC specifies the conditional covariance matrix of Y_t to satisfy:

$$H_t := Var_{t-1}(Y_t) = D_t E_{t-1}\left[\epsilon_t \epsilon_t'\right] D_t = D_t R_t D_t. \tag{15}$$

The conditional correlation matrix R_t is assumed to satisfy

$$R_t = (diag\ Q_t)^{-1/2}\, Q_t\, (diag\ Q_t)^{-1/2} \tag{16}$$

The matrix Q_t follows the dynamic

$$Q_t = (1-\alpha_c-\beta_c)\,\overline{Q} + \alpha_c \eta_{t-1}\eta_{t-1}' + \beta_c Q_{t-1} \tag{17}$$

where $\overline{Q} := E\left[\eta_{t-1}\eta_{t-1}'\right]$ and α_c, β_c are nonnegative scalar parameters with the restriction $\alpha_c + \beta_c \leq 1$.

To simplify estimation, we first assume that the conditional mean for each component series is constant through time and equal to the sample mean, and work with the demeaned series $\{\epsilon_t\}$. Also, we follow Engle (2002) and use the variance targeting procedure of Engle and Mezrich (1996), i.e., set the unconditional covariance matrix equal to the sample covariance matrix. Since our data series is only two dimensional, we estimate all of the remaining parameters in one step by maximum likelihood (conditional on the initial values for H_t and the unconditional covariance matrix being equal to the sample covariance matrix, and etc.) instead of using the two-step procedure proposed by Engle (2002). We estimate the DCC with two alternative specifications for the standardized error vector $Z_t := H_t^{-1/2}\epsilon_t$ where $H_t^{1/2}$ is an $N \times N$ lower triangular matrix such that $H_t^{1/2}\left(H_t^{1/2}\right)' = H_t$. One is $Z_t \sim i.i.d.N(0, I)$ as in the original DCC formulation of Engle (2002). For this specification, the unknown parameter vector θ to be estimated is $(\alpha_1, \beta_1, \alpha_2, \beta_2, \alpha_c, \beta_c)$. And the other is $Z_t \sim i.i.d.$standardized multivariate t with degrees of freedom parameter v. Motivated by the success of Student-t-based univariate GARCH models in improving the fit (see, e.g., Bollerslev (1987) and Bollerslev, Engle, and Nelson (1994)) over their normality-based counterparts by capturing excess conditional kurtosis, several authors have considered multivariate generalization of the t distribution for multivariate GARCH modeling. There are several different multivariate extensions of the t distribution (see Johnson and Kotz (1972, p.134)). We follow the common choice in the literature and use the joint density of a vector of N independent standard normal variables divided by a common denominator $\sqrt{\chi_v^2/v}$, where χ_v^2 is distributed χ^2 with degrees of freedom v, and multiplied by a standardizing coefficient $\sqrt{(v-2)/v}$. For this specification, $\theta = (\alpha_1, \beta_1, \alpha_2, \beta_2, \alpha_c, \beta_c, \phi)$, where $\phi := \nu^{-1}$ so that the multivariate normality is nested, and the log likelihood function is:

$$
\begin{aligned}
L(\theta; \{\epsilon_t\}) &= T \ln \Gamma\left(\frac{v+N}{2}\right) - T \ln \Gamma\left(\frac{v}{2}\right) - \frac{TN}{2} \ln \pi (v-2) \\
&\quad - \frac{1}{2} \sum_{t=1}^{T} \ln |R_t| - \sum_{t=1}^{T} \ln |D_t| - \frac{N+v}{2} \sum_{t=1}^{T} \ln \left(1 + \frac{\epsilon_t' D_t^{-1} R_t^{-1} D_t^{-1} \epsilon_t}{v-2}\right)
\end{aligned}
\tag{18}
$$

where $\Gamma(\cdot)$ is the gamma function. Table 1 shows the estimation results for our data of 4937 bivariate observations. $\widehat{\phi}$ for the DCC-t (multivariate-t-based DCC), is 0.1793, highly significantly above 0. And the likelihood ratio is also large enough to reject $H_0 : \phi = 0$ (i.e., DCC-n, the normality-based DCC) at conventional levels.

A next step is to conduct specification tests, for which we use several alternative PITs together with the testing procedure recently developed by Duan (2003). The best-known test for a distributional assumption is the Kolmogorov-Smirnov test. However, there are a few problems associated with using such tests as the classical Kolmogorov-Smirnov test for testing the i.i.d. $U(0,1)$ property of the probability integral transforms as a check on model specification. First, as stressed by Hong and Li (2004), the Kolmogorov-Smirnov test in this context tests the uniformity under the i.i.d. assumption, and does not test the uniformity and i.i.d. jointly. Secondly, if we use a model with estimated parameters for producing conditional density forecasts, the PITs will not be i.i.d. uniform even when the model is correctly specified. The effect of parameter estimation uncertainty on the distribution of test statistics in general remains even asymptotically. Bai's (2002) test, while removing the problem of parameter estimation uncertainty through an extension of Khmaladze's (1981) martingale transform technique, does not test the i.i.d. property. The Hong and Li (2004) test and the Duan (2003) test are two of the few PIT-based specification tests that jointly test the i.i.d. and $U(0,1)$ properties of the PITs and are asymptotically free of the effects of estimation uncertainty. Here, we rely on Duan's (2003) approach.

To review the Duan test briefly, let $\widehat{\theta}_T$ be a $\sqrt{T}$-consistent estimator of the true parameter θ_0 (a $k_\theta \times 1$ vector) where T is the sample size. The Duan test first transforms a PIT series $U_t\left(\widehat{\theta}_T\right)$, implied by the estimated model whose specification is being tested, into $\xi_t\left(\widehat{\theta}_T\right) := \Phi^{-1}\left(U_t\left(\widehat{\theta}_T\right)\right)$ using the inverse standard normal c.d.f. $\Phi^{-1}(\cdot)$ as in Berkowitz (2002). Duan (2003) originally considered testing univariate model specifications, but his testing procedure may be applied to our PITs as well as DHT's PITs arising from multivariate dynamic models. Further transform $\left\{\xi_t\left(\widehat{\theta}_T\right)\right\}_{t=1}^{T}$ into $\left\{q_{m,i}^{(p)}\left(\widehat{\theta}_T\right)\right\}_{i=1}^{[T/m]}$, $m=1,2,\cdots,n$, where

$$\begin{aligned} q_{m,i}^{(p)}(\theta) &:= \sum_{j=1}^{m} \xi_{(i-1)m+j}^{p}(\theta) \text{ for } p=1,2 \\ q_{m,i}^{(3)}(\theta) &:= m^{-1}\left(\sum_{j=1}^{m} \xi_{(i-1)m+j}(\theta)\right)^2 \\ q_{m,i}^{(4)}(\theta) &:= m^{-2}\left(\sum_{j=1}^{m} \xi_{(i-1)m+j}^{2}(\theta)-m\right)^2. \end{aligned}$$

Since $\{\xi_t(\theta_0)\}$ is a series of i.i.d. standard normal random variables, the CDF $R_m^{(p)}(x)$ for $q_{m,i}^{(p)}(\theta_0)$, $p=1,2,3,4$, are known (and invariant with respect to the value of θ_0). Define

$$X_{m,T}^{(p)}(\theta) := m^{-1/2}\left[T/m\right]^{-1}\sum_{i=1}^{[T/m]} R_m^{(p)}\left(q_{m,i}^{(p)}(\theta)\right)-\frac{1}{2}$$

and $X_T^{(p)}(\theta) := \left(X_{1,T}^{(p)}(\theta),\cdots,X_{n,T}^{(p)}(\theta)\right)$. Then, since $\left\{R_m^{(p)}\left(q_{m,i}^{(p)}(\theta_0)\right)-\frac{1}{2}\right\}$ is a series of i.i.d $U\left(-\frac{1}{2},\frac{1}{2}\right)$ random variables, $\sqrt{T}X_T^{(p)}(\theta_0)$ converges to a zero-mean multivariate normal random variable by the central limit theorem. The problem is that we do not know the value of θ_0 and that the effect of parameter estimation error on the distribution of $\sqrt{T}X_T^{(p)}\left(\widehat{\theta}_T\right)$ used in lieu of $\sqrt{T}X_T^{(p)}(\theta_0)$ does not disappear even asymptotically. Duan (2003) shows that it is possible under a set of mild regularity conditions to find $n\in\mathbb{N}$ and a sequence of $k\times n$ weighting matrices $\Xi\left(\widehat{\theta}_T\right)$ that asymptotically cancels out the components of $\sqrt{T}X_T^{(p)}\left(\widehat{\theta}_T\right)$ due to estimation error so that, as $T\to\infty$,

$$J_T^{(p)}\left(\widehat{\theta}_T\right) = T\left\|\Xi\left(\widehat{\theta}_T\right)X_T^{(p)}\left(\widehat{\theta}_T\right)\right\|^2 \xrightarrow{D} \chi^2(k) \tag{19}$$

where $k\geq 1$ is some integer and $\|\cdot\|$ is the Euclidean norm. For those $\left\{U_t^{DHT}\left(\widehat{\theta}_T\right)\right\}$ that stack N PIT series into a single univariate one, we need to use NT in place of T in (19). We set $k=2$ and follow Duan's (2003) procedure step by step[2] to construct $\Xi\left(\widehat{\theta}_T\right)$ for the test statistics $J_T^{(p)}\left(\widehat{\theta}_T\right)$ for $p=1,2,3,4$ using various

[2]Except that we use a long simulated sample (100,000 observations) in estimating $\lim_{T\to\infty}\partial Z_T^{(p)}(\theta_0)/\partial\theta$ that comprises a part of the weigthing matrix, whereas Duan (2003) for each of his empirical examples used a shorter simulated sample with size set equal to that of the corresponding real data set.

PIT series from our estimation of the bivariate DCC model. Our sample has size $T = 4937$.

One of the PITs we consider, defined only for the case $N = 2$, is:

$$U_t\left(\widehat{\theta}_T\right) := \int_{\{g_t(e)\leq g_t(\epsilon_t)\}} f_{\epsilon_t}\left(e;\widehat{\theta}_T\right) de \tag{20}$$

where $f_{\epsilon_t}\left(\epsilon;\widehat{\theta}_T\right)$, $e := (e_1, e_2) \in \mathbb{R}^2$, is the conditional density for ϵ_t implied by the estimated model,

$$g_t(e) := \begin{cases} e_1 - e_2 \text{ if } \widehat{\rho}_t \geq 0 \\ e_1 + e_2 \text{ if } \widehat{\rho}_t < 0 \end{cases}, \tag{21}$$

and $\widehat{\rho}_t$ is the estimated conditional correlation (the (2,1), or (1,2), element of R_t). This transform function is motivated by Example 1.1. Note, however, that unlike Example 1.1 in which the true conditional variance of each series is constant and known to be one, here we use conditional variances produced by the estimated DCC model, which is likely to be misspecified to some degree whether the distribution of the standardized error vector is modeled as multivariate normal or multivariate t.

For the DCC-n and the DCC-t, it is easy to calculate $U_t^{Contour}\left(\widehat{\theta}_T\right)$ introduced in 3.2:

$$U_t^{Contour}\left(\widehat{\theta}_T\right) := \int_{\left\{f_{\epsilon_t}\left(e;\widehat{\theta}_T\right)\leq f_{\epsilon_t}\left(\epsilon_t;\widehat{\theta}_T\right)\right\}} f_{\epsilon_t}\left(e;\widehat{\theta}_T\right) de \tag{22}$$

$$= \int_{\left\{e'D_t^{-1}\left(\widehat{\theta}_T\right)R_t^{-1}\left(\widehat{\theta}_T\right)D_t^{-1}\left(\widehat{\theta}_T\right)e\geq\epsilon_t'D_t^{-1}\left(\widehat{\theta}_T\right)R_t^{-1}\left(\widehat{\theta}_T\right)D_t^{-1}\left(\widehat{\theta}_T\right)\epsilon_t\right\}} f_{\epsilon_t}\left(e;\widehat{\theta}_T\right) de$$

$$= \int_{\left\{z'z\geq\widehat{Z}_t'\widehat{Z}_t\right\}} f_{Z_t}\left(z;\widehat{\theta}_T\right) dz$$

where $Z_t(\theta) := \epsilon_t' D_t^{-1}(\theta) R_t^{-1}(\theta) D_t^{-1}(\theta)\epsilon_t$, $D_t(\theta)$ and $R_t(\theta)$ are D_t and R_t given θ, and $\widehat{Z}_t := Z_t\left(\widehat{\theta}_T\right)$. For the DCC-$n$ model, $f_{Z_t}\left(z;\widehat{\theta}_T\right) := \phi_N(z)$ where ϕ_N is the multivariate standard normal p.d.f., and for the DCC-t model, $f_{Z_t}\left(z;\widehat{\theta}_T\right) := f_{t_{\widehat{v}}}(z)$ where f_{t_v} is the multivariate t p.d.f. (v degrees of freedom). So we have

$$U_t^{Contour}\left(\widehat{\theta}_T\right) = \begin{cases} \int_{\left\{z'z\geq\widehat{Z}_t'\widehat{Z}_t\right\}} \phi_N(z)\, dz = 1 - \Psi_{\chi^2_N}\left(\widehat{Z}_t'\widehat{Z}_t\right) \\ \text{for DCC-}n \\ \int_{\left\{\frac{\widehat{v}}{N(\widehat{v}-2)}z'z\geq\frac{\widehat{v}}{N(\widehat{v}-2)}\widehat{Z}_t'\widehat{Z}_t\right\}} f_{t_{\widehat{v}}}(z)\, dz = 1 - \Psi_{F_{N,\widehat{v}}}\left(\frac{\widehat{v}}{N(\widehat{v}-2)}\widehat{Z}_t'\widehat{Z}_t\right) \\ \text{for DCC-}t \end{cases} \tag{23}$$

where $\Psi_{\chi^2_N}$ is the χ^2 c.d.f. (N degrees of freedom) and $\Psi_{F_{N,v}}$ is the F c.d.f. (N numerator and v denominator degrees of freedom). $\Psi_{F_{N,\widehat{v}}}$ for the DCC-t case is obtained because, by the definition of our multivariate t density, $\frac{\widehat{v}}{N(\widehat{v}-2)}\widehat{Z}_t'\widehat{Z}_t$ has the distribution of a $\left(\chi^2_N/N\right)/\left(\chi^2_{\widehat{v}}/\widehat{v}\right)$ random variable if $\widehat{\theta}_T = \theta_0$.

The DHT transforms for the DCC-n and DCC-t are straightforward to calculate as well. Let $\widehat{\eta}_{it}$ denote the ith element of $\widehat{\eta}_t := \eta_t\left(\widehat{\theta}_T\right)$. We have

$$U_{1t}^{DHT}\left(\widehat{\theta}_T\right) = \begin{cases} \Phi\left(\widehat{\eta}_{1t}\right) & \text{for DCC-}n \\ \Phi_{t_{\widehat{v}}}\left(\sqrt{\frac{\widehat{v}-2}{\widehat{v}}}\widehat{\eta}_{1t}\right) & \text{for DCC-}t \end{cases} \tag{24}$$

where $\Phi_{t_{\widehat{v}}}$ is the c.d.f. of the (univariate) t-distribution with $\widehat{v}$ degrees of freedom, and

$$U_{2t}^{DHT}\left(\widehat{\theta}_T\right) = \begin{cases} \Phi\left(\frac{\widehat{\eta}_{2t}-\widehat{\rho}_{12,t}\widehat{\eta}_{1t}}{\sqrt{1-\widehat{\rho}_{12,t}^2}}\right) & \text{for DCC-}n \\ \Psi_{t_{\widehat{v}}}\left(\frac{\widehat{\eta}_{2t}-\widehat{\rho}_{12,t}\widehat{\eta}_{1t}}{\sqrt{\frac{\widehat{v}-2}{\widehat{v}}\left(1-\widehat{\rho}_{12,t}^2\right)}}\right) & \text{for DCC-}t \end{cases} \tag{25}$$

The expression for $U_{2t}^{DHT}\left(\widehat{\theta}_T\right)$ with the DCC-t model is due to the fact that if $\zeta := (\zeta_1, \cdots, \zeta_N) = A^{1/2}\delta$ where δ is an $N \times 1$ random vector with the multivariate t distribution with v degrees of freedom and A is some $N \times N$ matrix satisfying conditions for it to be a correlation matrix, then the conditional distribution of η_j given η_i is that of $\sqrt{1-\rho_{ij}^2}t_v + \rho_{ij}\eta_i$ where t_v a random variable with the t distribution with v degrees of freedom and ρ_{ij} is the (i,j) element of A (cf. Johnson and Kotz (1972, p.135)) .

Table 2 shows two sets of Duan's (2003) J test statistics $J_T^{(p)}\left(\widehat{\theta}_T\right)$, $p = 1, 2, 3, 4$, using the linear transform function (21) (the row "Linear") and the contour transform function (22) (the row "Contour") for each of the two specifications: the DCC-n and the DCC-t. The J statistics constructed of the linear-transform-function-based PITs, all but the one for $p = 1$, reject the null at 1% level[3] that the DCC-n is correctly specified. All of the J statistics using the contour-based PITs cleanly reject the DCC-n specification at 1% level. On the other hand, looking at the second column, none of the test statistics rejects the DCC-t at 5% level. On Table 3, the three groups of rows show the J statistics based on $\left\{U_{1t}^{DHT}\left(\widehat{\theta}_T\right)\right\}$ (DHT1), $\left\{U_{2t}^{DHT}\left(\widehat{\theta}_T\right)\right\}$ (DHT2), and the two series of PITs stacked (DHT3), with the first series being the \$/£ series and the second being the ¥/\$ series. Neither the DCC-n nor the DCC-t is rejected by any of these statistics at 5% level, which presumably is due to the low power of PITs. The statistics on the next three groups of rows are similarly calculated with DHT4 corresponding to DHT1, and etc., but with the ¥/\$ (resp. \$/£) series being the first (resp. second) asset. Since $U_{1t}^{DHT}\left(\widehat{\theta}_T\right)$ is based on the marginal distribution of the first series (still conditional on the past realizations of the series), the DHT4 panel reveals a possible inadequacy of the standard univariate GARCH(1,1) specification for the ¥/\$ conditional variance process.

[3] Of course the usual caveat regarding multiple testing applies here since we do not know the overall size.

Although some of the J statistics in the last three panels DHT4-DHT5 reject the DCC-n specification strongly, the rejection is not as clear-cut as in the case of the tests with the linear- and contour-PIT-based statistics. Overall, it is reassuring for the users of the DCC-t model to know that viewing of the conditional density in different ways does not reveal problems with the specification.

The DCC model is designed to deal with many asset returns, more than just two, at once. It would be interesting to see how high-dimensional DCC models would stand these specification tests.

6 Conclusion

In this paper, we have introduced a new approach for producing probability integral transforms that "scan" conditional probability density forecasts from different ways, and given several examples of specific "scanning" functions that transform the original multivariate time series into an i.i.d. series. An interesting example is to use the conditional density forecast itself as a scanning function for each period. Various PIT series that are produced by our method as well as DHT's serve as building blocks for constructing formal tests for forecast accuracy and specification adequacy. Multiple PIT series may be combined in clever ways to produce powerful omnibus tests, which we leave as a topic for future research.

References

[1] Andersen, Torben G., and Tim Bollerslev, 1988, Answering the critics: Yes, ARCH models do provide good volatility forecasts, *International Economic Review* **39**, 885-905.

[2] Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Paul Labys, 2003, Modeling and forecasting realized volatility, *Econometrica* **71**, 579-625.

[3] Bai, J., 2003, Testing parametric conditional distributions of dynamic models, *Review of Economics and Statistics* **85**, 531-549.

[4] Bao, Yong, Tae-Hwy Lee, and Burak Saltoğlu, 2003, A test for density forecast comparison with application to risk management, Working paper, University of California Riverside.

[5] Bauwens, Luc, Pierre Giot, Joachim Grammig, and David Veredas, 2000, A comparison of financial duration models via density forecasts, CORE Discussion Paper 2000-60.

[6] Bauwens, Luc, and Sébastien Laurent, 2002, A new class of multivariate skew densities, with application to GARCH models, CORE Discussion Paper 2002-20.

[7] Berkowitz, Jeremy, 2001, Testing density forecasts with applications to risk management, *Journal of Business and Economic Statistics* **19**, 465-474.

[8] Bollerslev, 1986, Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* **31**, 307-327.

[9] Bollerslev, Tim, 1987, A conditional heteroskedastic time series model for speculative prices and rate of return, *Review of Economics and Statistics* **69**, 542-547.

[10] Bollerslev, Tim, Robert F. Engle, and Daniel B. Nelson, 1994, ARCH models, in Daniel McFadden and Robert F. Engle, eds., *Handbook of Econometrics Vol.4*, Elsevier: Amsterdam.

[11] Clements, Michael P., and Jeremy Smith, 2002, Evaluating multivariate forecast densities: A comparison of two approaches, *International Journal of Forecasting* **18**, 397-407.

[12] Corradi, Valentina, and Norman R. Swanson, 2003, Predictive density accuracy tests, Working paper, Queen Mary, University of London and Rutgers University.

[13] Diebold, Francis X., T. Gunther, and Anthony Tay, 1998, Evaluating density forecasts, with applications to financial risk management, *International Economic Review* **39**, 863-883.

[14] Diebold, Francis X., Jinyong Hahn, and Anthony Tay, 1999, Multivariate density forecast evaluation and calibration in financial risk management: High-frequency returns on foreign exchange, *Review of Economics and Statistics* **81**, 661-673.

[15] Duan, Jin-Chuan, 2003, A specification test for time series models by a normality transformation, Working paper, University of Toronto.

[16] Engle, Robert F., 2002, Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models, *Journal of Business and Economic Statistics* **20**, 339-350.

[17] Engle, Robert F., and J. Mezrich, 1996, GARCH for groups, *Risk* **9**, 36-40.

[18] Engle, Robert F., and Isao Ishida, 2002, Modeling variance of variance: The Square-root, the affine, and the CEV GARCH models, Working paper, New York University.

[19] Giacomini, Rafaella, 2002, Comparing density forecasts via weighted likelihood ratio tests: Asymptotic and bootstrap methods, UCSD Discussion Paper, 2002-12.

[20] Giacomini, Rafaella, and Halbert White, 2003, Tests of conditional predictive ability, UCSD Discussion Paper 2003-09.

[21] Hong, Yongmiao, Haitao Li, and Feng Zhao, 2002, Can the random walk model be beaten in out-of-sample density forecasts: Evidence from intraday foreign exchange rates, Working paper, Cornell University.

[22] Hong, Yongmiao, and Haitao Li, 2004, Nonparametric specification testing for continuous-time models with applications to spot interest rates, forthcoming in *Review of Financial Studies*.

[23] Ishida, Isao, 2004, Essays on Financial Time Series, Unpublished Ph.D. dissertation, University of California, San Diego.

[24] Johnson, Norman L., and Samuel Kotz, 1972, Distributions in Statistics: Continuous Multivariate Distributions, John Wiley & Sons: New York.

[25] Joshua V. Rosenberg, and Robert F. Engle, 2002, Empirical pricing kernels, *Journal of Econometrics* **64**, 341-372.

[26] Pearson, Karl, 1933, On a method of determining whether a sample of size n supposed to have been drawn from a parent population having a known probability integral has probably been drawn at random, *Biometrika* **25**, 134-148.

[27] Quesenberry, Charles, 1986, Some transformation methods in goodness-of-fit, in Ralph B. D'Agostino and Michael A. Stephens, eds., *Goodness-of-Fit Techniques*, Marcel Dekker: New York.

[28] Rosenblatt, Murray, 1952, Remarks on a multivariate transformation, *Annals of Mathematical Statistics* **23**, 470-472.

[29] Stărică, Cătălin, 2003, Is GARCH(1,1) as good a model as the Nobel prize accolades would imply? Working paper, Chalmers University of Technology.

Table 1: Estimation Results for the DCC-n and the DCC-t Models

	DCC-n	S.E.	DCC-t	S.E.
α_1	0.0403	(0.0044)	0.0409	(0.0052)
β_1	0.9520	(0.0058)	0.9528	(0.0066)
α_2	0.0498	(0.0063)	0.0424	(0.0054)
β_2	0.9250	(0.0110)	0.9496	(0.0075)
α_c	0.0220	(0.0030)	0.0263	(0.0038)
β_c	0.9746	(0.0036)	0.9705	(0.0045)
ϕ			0.1793	(0.0075)
Log-likelihood	-8712.81		-8318.01	

Table 2: Duan's Test Statistics for the DCC Models

		DCC-n		DCC-t	
Test type	p	J statistics	p-value	J statistics	p-value
Linear	1	2.7370	0.2545	3.2797	0.1940
	2	90.4322	0.0000*	2.3826	0.3038
	3	58.5665	0.0000*	0.6238	0.7321
	4	15.8063	0.0004*	1.6553	0.4371
Contour	1	436.0785	0.0000*	5.2772	0.0715
	2	19.1397	0.0001*	2.1702	0.3379
	3	11.1605	0.0038*	4.7761	0.0918
	4	44.2296	0.0000*	3.1134	0.2108

* Significant at 1% level.

Table 3: Duan's test statistics for the DCC models

		DCC-n		DCC-t	
Test Type	p	J statistics	p-value	J statistics	p-value
DHT 1	1	0.5690	0.7524	0.7962	0.6716
	2	4.0137	0.1344	0.6234	0.7322
	3	0.9683	0.6162	2.6230	0.2694
	4	1.3097	0.5195	1.2928	0.5239
DHT 2	1	5.4570	0.0653	6.7065	0.0350
	2	0.0010	0.9995	3.9924	0.1359
	3	0.6527	0.7215	0.5081	0.7757
	4	1.2216	0.5429	2.8864	0.2362
DHT 3	1	4.3748	0.1122	5.6708	0.0587
	2	1.0323	0.5968	0.8347	0.6588
	3	4.7112	0.0948	1.7634	0.4141
	4	2.3681	0.3060	0.2251	0.8936
DHT 4	1	4.3697	0.1125	5.7812	0.0555
	2	8.6598	0.0132	0.1031	0.9498
	3	3.4449	0.1786	0.4198	0.8107
	4	21.3210	0.0000*	0.3859	0.8245
DHT 5	1	0.8930	0.6399	1.2188	0.5437
	2	1.4267	0.4900	0.8674	0.6481
	3	1.8381	0.3989	2.7417	0.2539
	4	1.5010	0.4721	1.2323	0.5400
DHT 6	1	4.1030	0.1285	5.7975	0.0551
	2	17.4569	0.0002*	0.4803	0.7865
	3	8.8117	0.0122	2.5069	0.2855
	4	11.4230	0.0033*	0.4070	0.8159

* Significant at 1% level.

Figure 1: Bivariate PIT with $g(y_1, y_2) = \omega_1 y_1 + \omega_2 y_2$

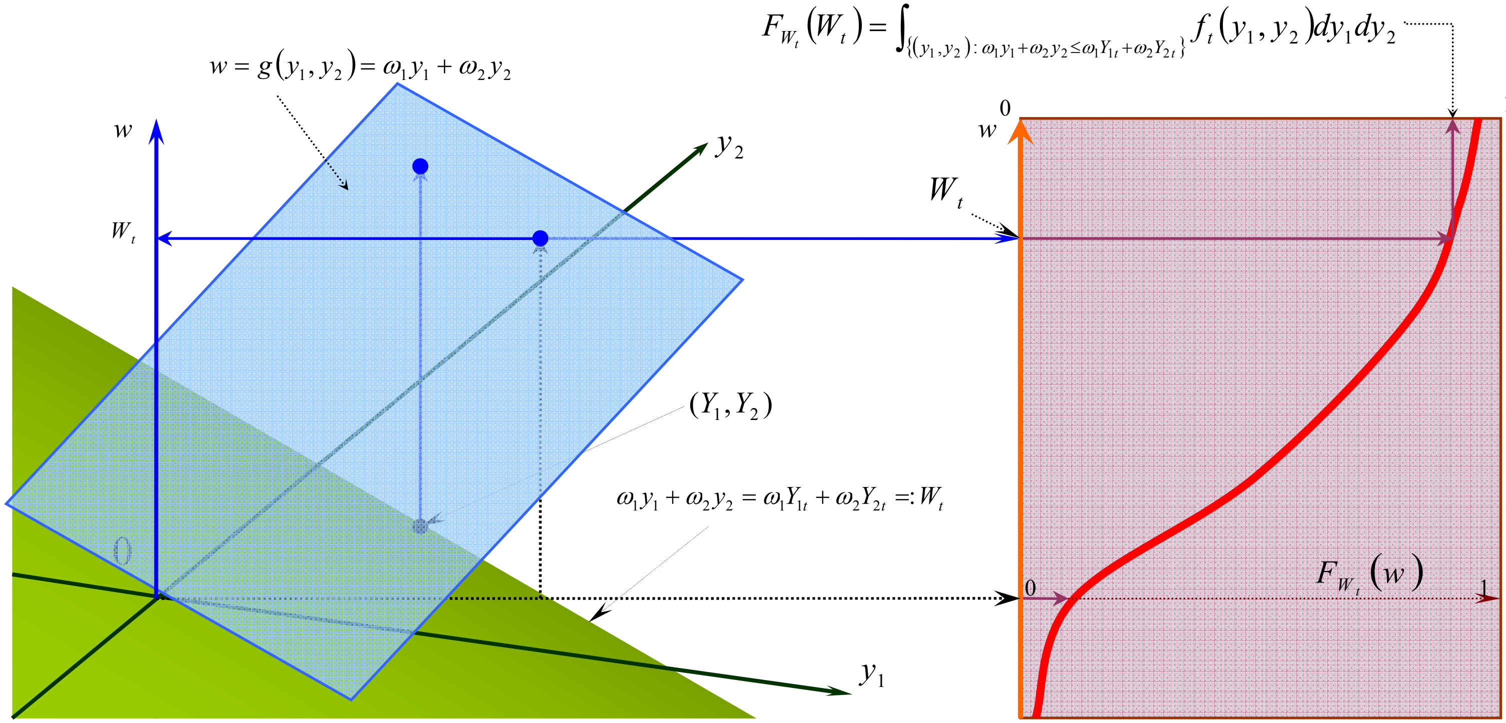

Figure 2 : Bivariate PIT with $g_t(y_1, y_2) = \hat{f}_t(y_1, y_2)$

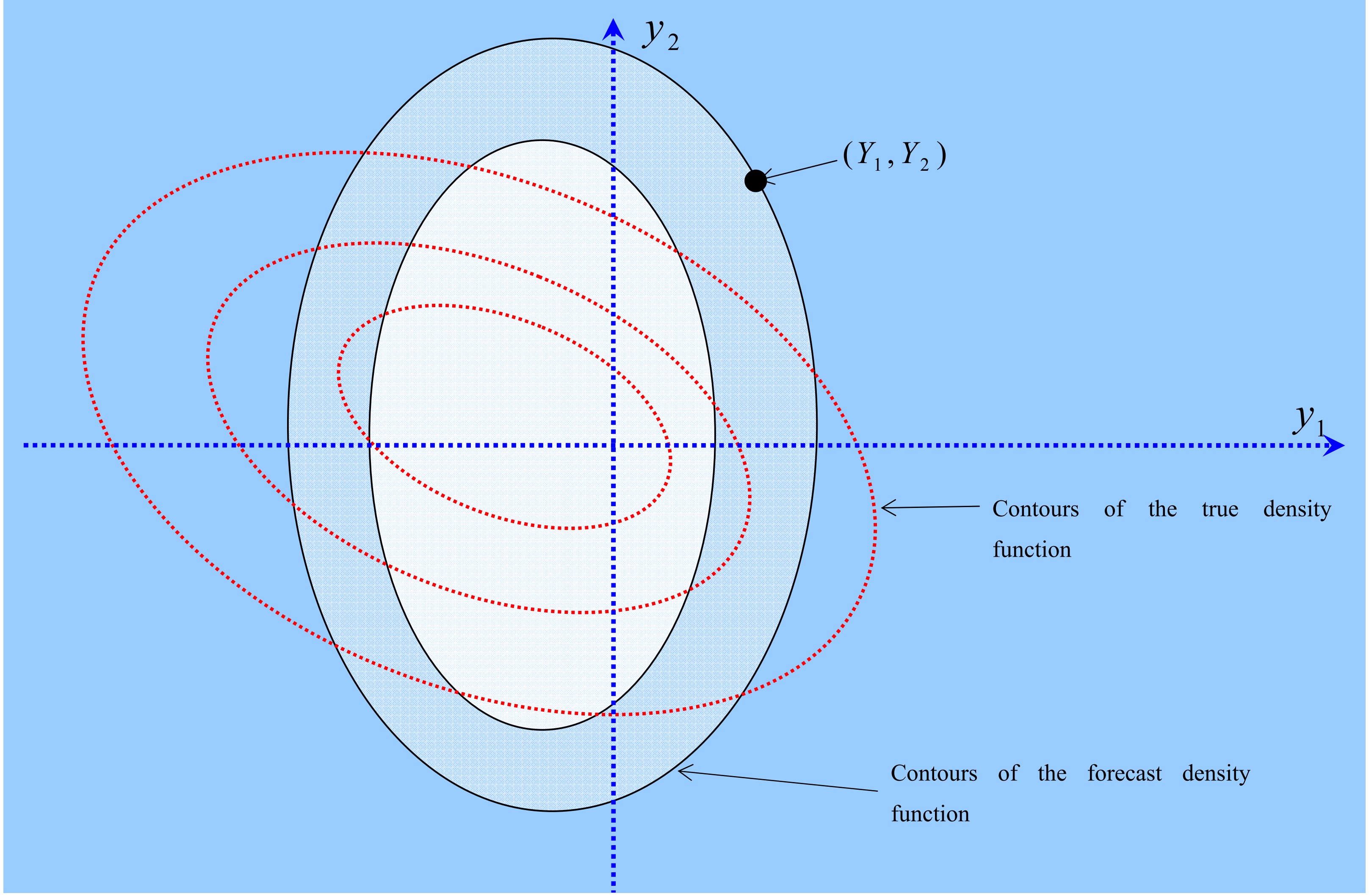

Figure 3 $\hat{\rho} = -0.99$

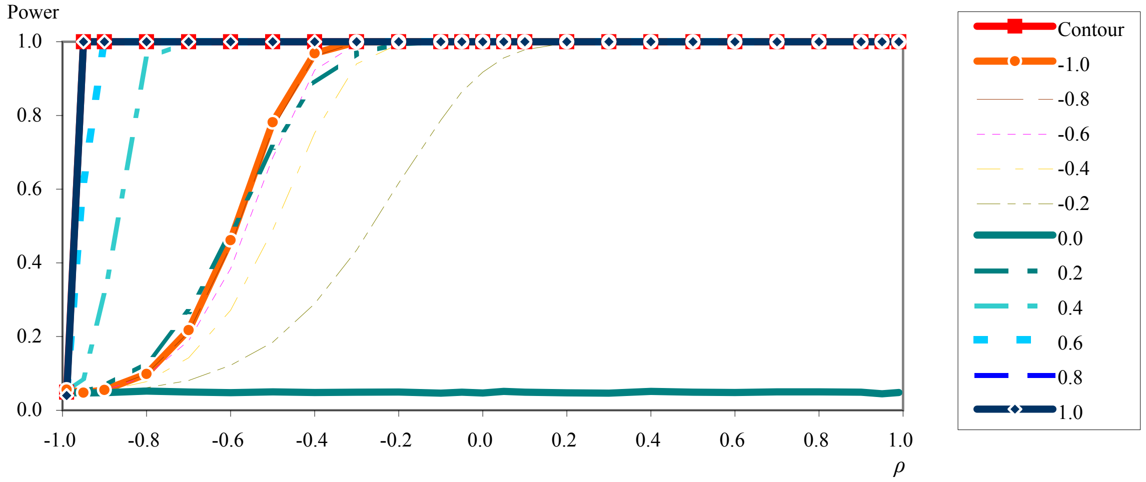

Figure 4 $\hat{\rho} = -0.95$

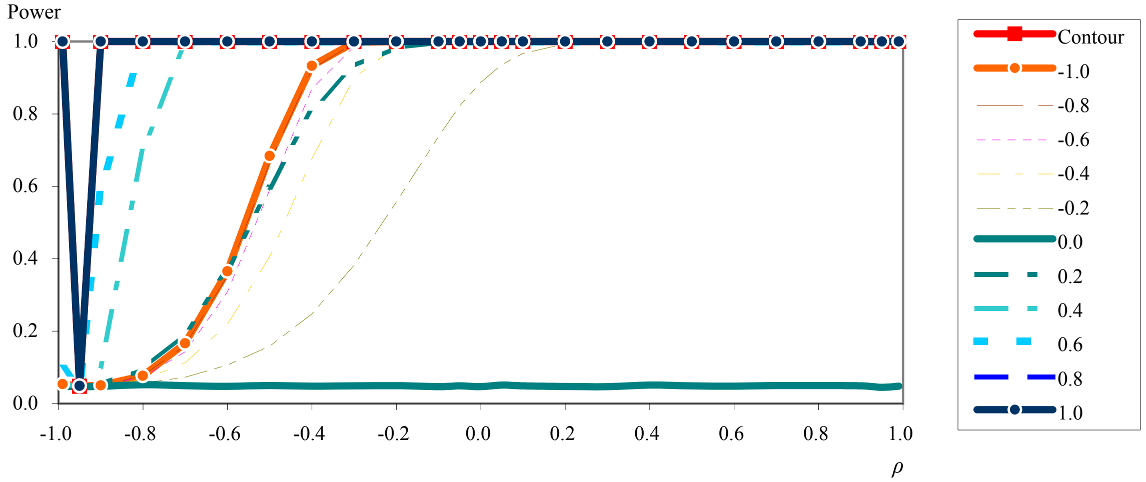

Figure 5 $\hat{\rho} = -0.90$

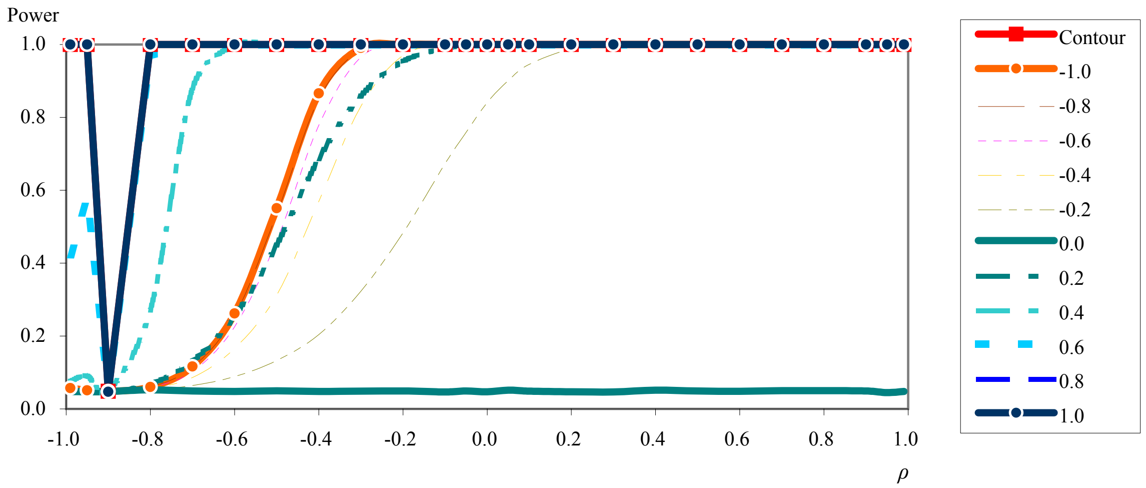

Figure 6 $\hat{\rho} = -0.80$

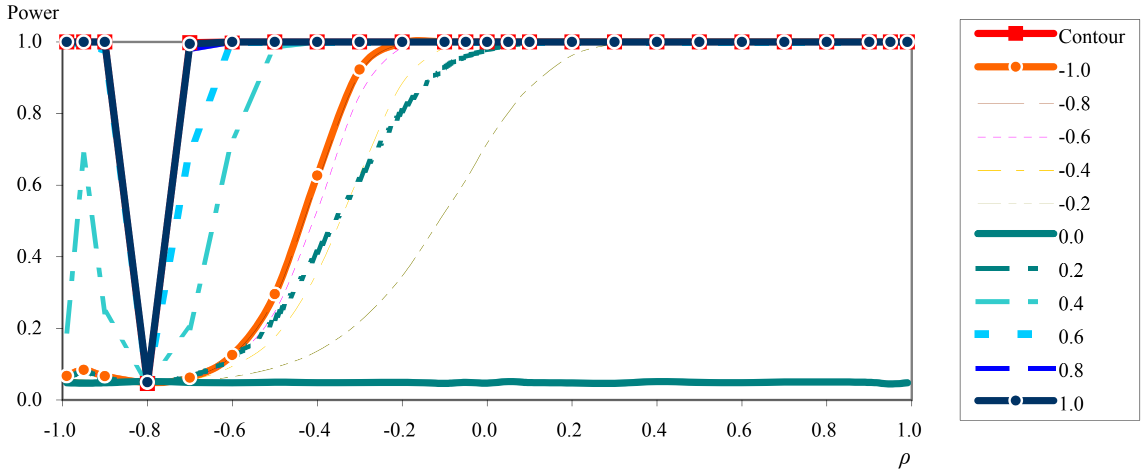

Figure 7 $\hat{\rho} = -0.60$

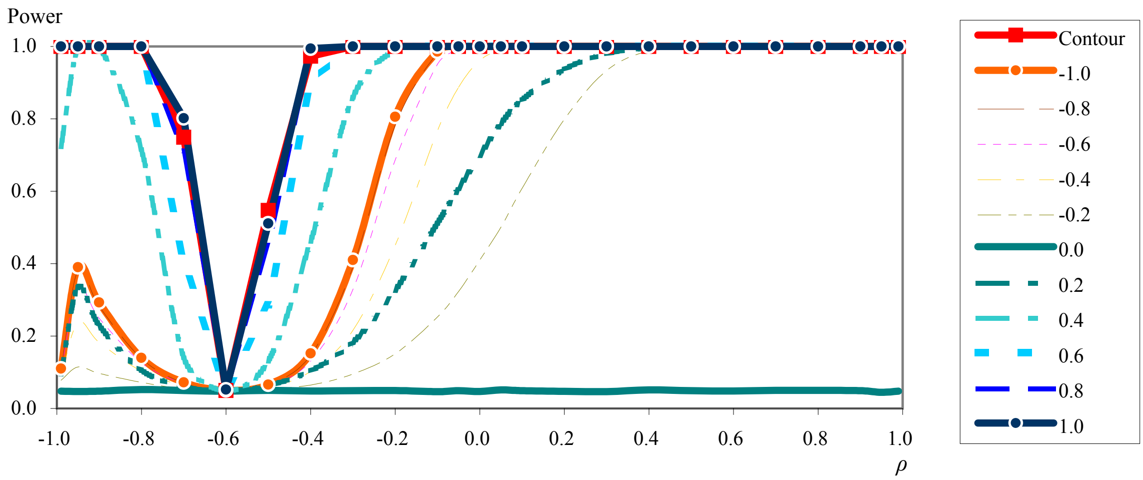

Figure 8 $\hat{\rho} = -0.40$

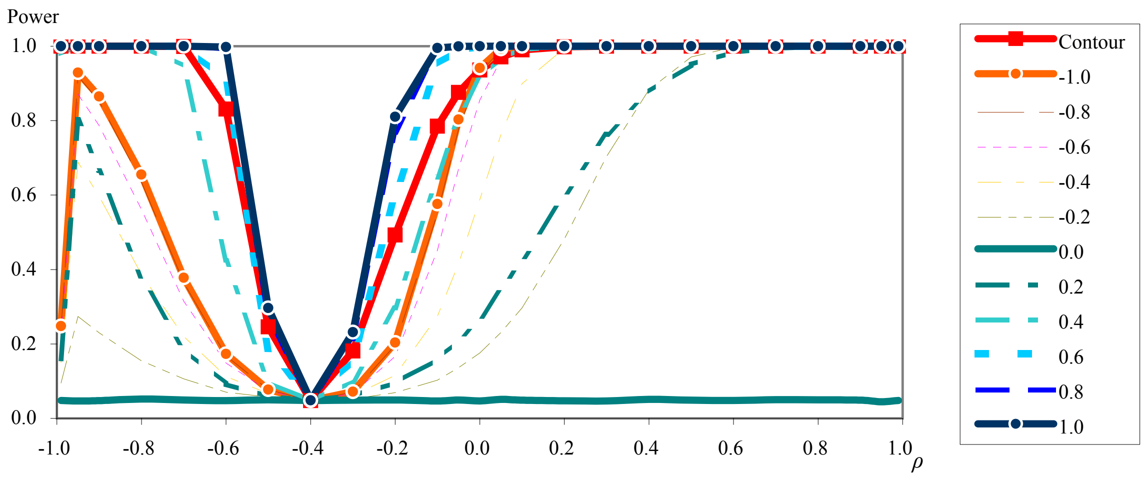

Figure 9 $\hat{\rho} = -0.20$

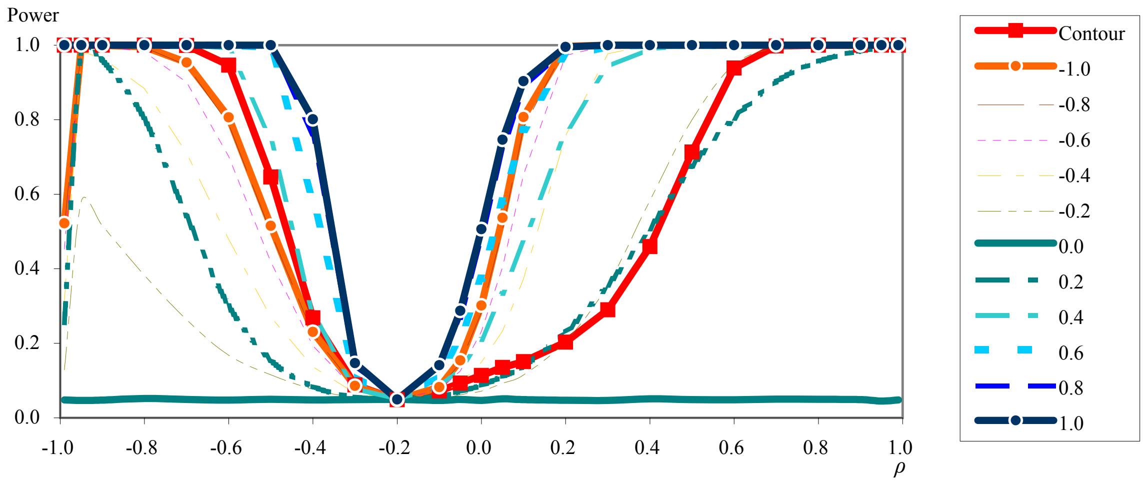

Figure 10 $\hat{\rho} = 0.00$

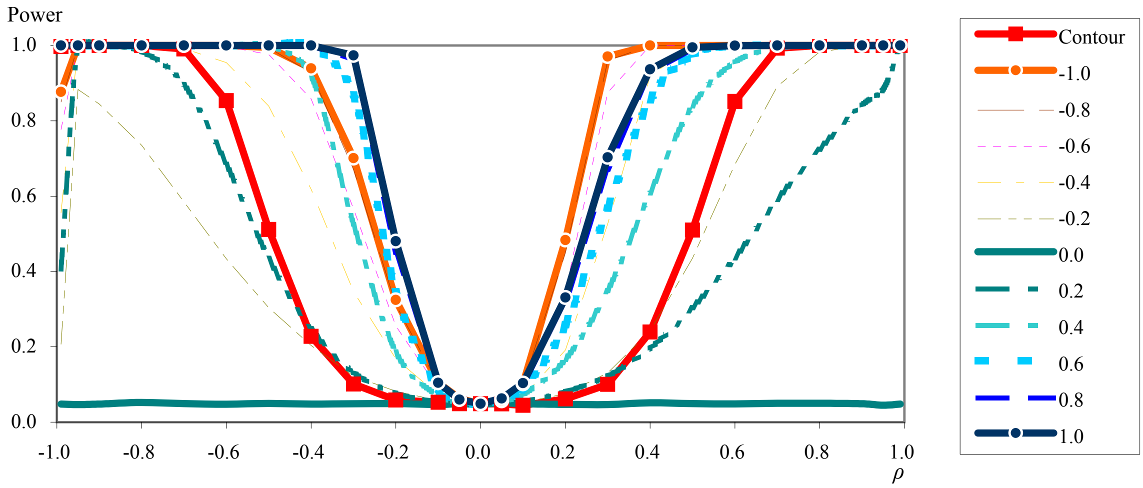

Figure 11 $\hat{\rho} = 0.20$

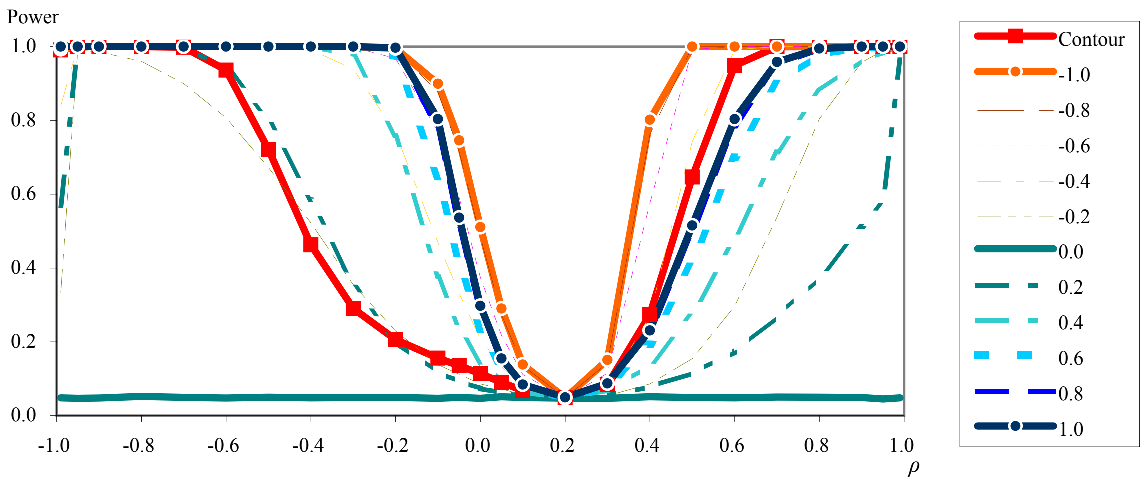

Figure 12

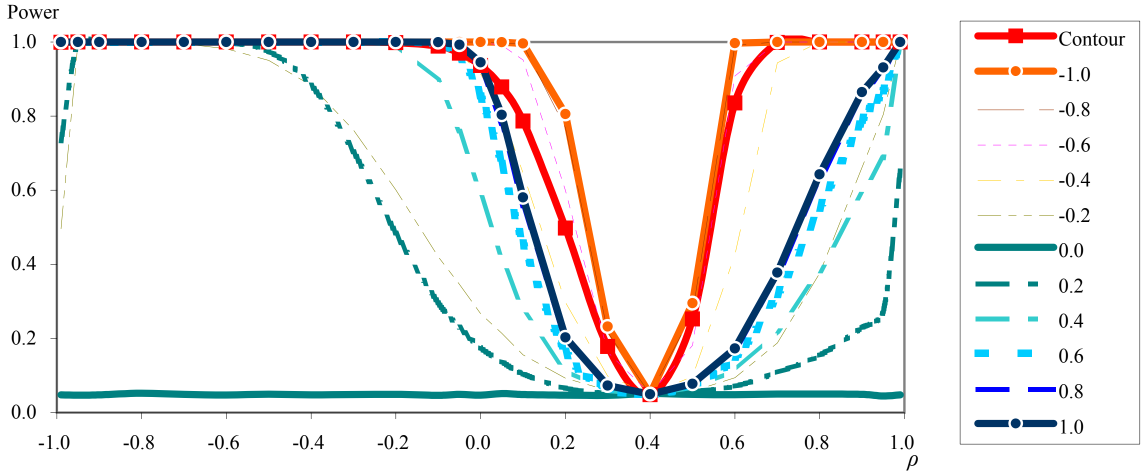

Figure 13

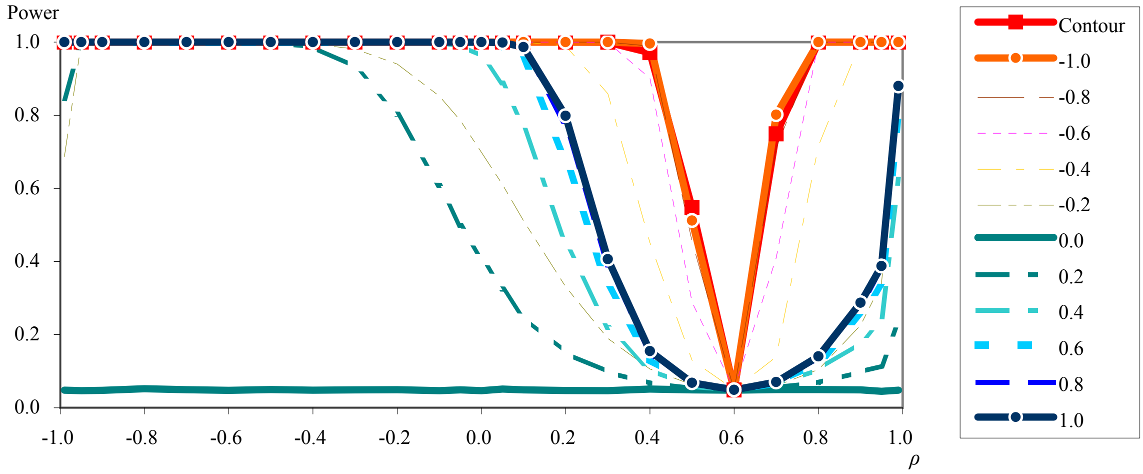

Figure 14 $\hat{\rho} = 0.80$

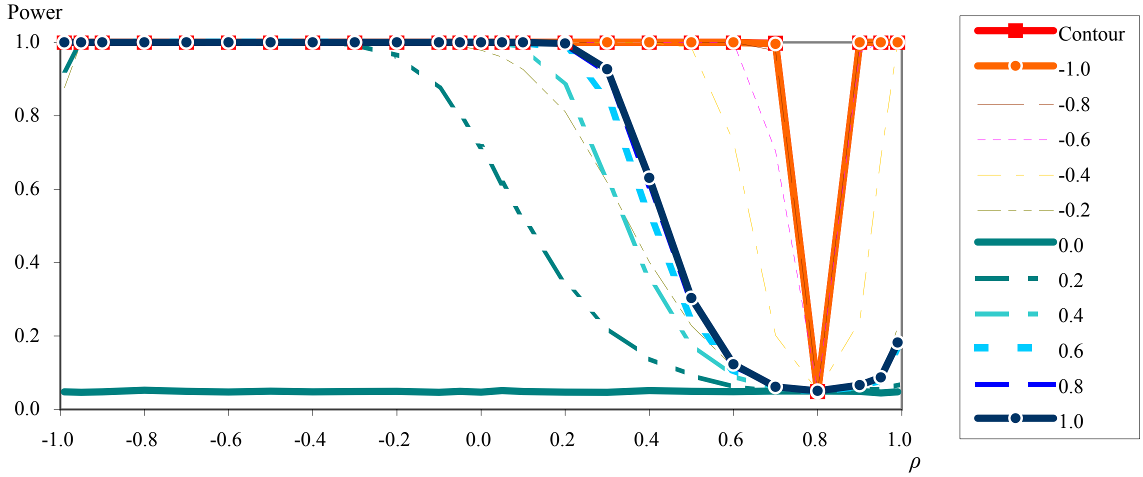

Figure 15 $\hat{\rho} = 0.90$

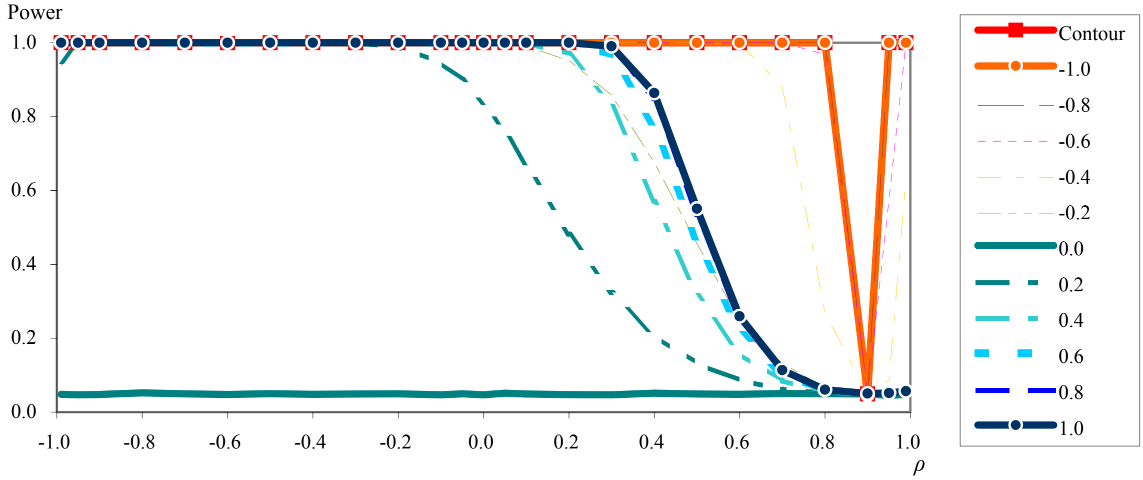

Figure 16 $\hat{\rho} = 0.95$

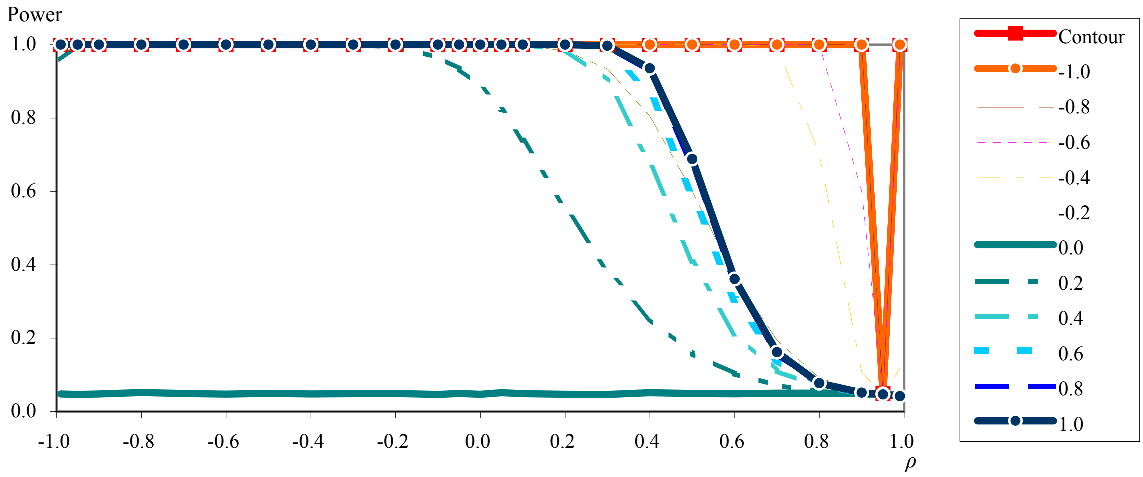

Figure 17 $\hat{\rho} = 0.99$