Autocontour-based Evaluation of Multivariate Predictive Densities*

Gloria González-Rivera
Department of Economics
University of California, Riverside
Riverside, CA 92521
E-mail: gloria.gonzalez@ucr.edu

Emre Yoldas
Department of Economics
Bentley University
Waltham, MA 02452
E-mail: eyoldas@bentley.edu

April 2011

Abstract

We contribute to the rather thin literature on multivariate density forecasts by introducing a new framework for out-of-sample evaluation of multivariate density forecast models that builds upon the concept of "autocontour" proposed by González-Rivera et al. (2011). This approach uniquely combines formal testing with graphical devices. We work with the one-step-ahead quantile residuals, which under the null hypothesis of a correct density model must be i.i.d. (univariate and multivariate) normal. Their corresponding autocontours are mathematically very tractable and the tests based on them enjoy standard asymptotic properties. We show that parameter uncertainty is asymptotically irrelevant under certain conditions and, in general, a parametric bootstrap provides outstanding finite sample properties. We provide simulation evidence on finite sample performance of the tests and compare their performance with an alternative testing procedure. We also illustrate this methodology by evaluating bivariate density forecasts of the returns on U.S. value and growth portfolios.

Keywords: Probability Contour Plot, Probability Integral Transformation, Parameter Uncertainty, Forecasting Schemes.

*We would like to thank the participants at the International Symposium on Forecasting, San Diego 2010, and the ASA Joint Statistical Meetings, Vancouver 2010, for useful comments. Special thanks to Zeynep Senyuz for very helpful discussions. We are also grateful to the associate editor and two anonymous referees, whose constructive comments have greatly improved this manuscript. The usual caveat applies. This manuscript has been written under the auspices of the SAS/IIF Forecasting Research Award. Gloria González-Rivera acknowledges the financial support of the UC-Riverside University Scholar Award and Academic Senate grants.

1 Introduction

Evaluating density forecasts has been a very active field of research in recent years as both academics and practitioners emphasize the broader information content of a density forecast relative to a point forecast (see Tay and Wallis (2000) for a recent survey). Accurate density forecasts facilitate decision making by policy makers and business managers alike. Prime examples include the fields of financial risk management, e.g. Diebold et al. (1998), and Berkowitz (2001), and monetary policy, e.g. Bache et al. (2010).

The pioneering work of Diebold et al. (1998) proposed using the probability integral transformation (PIT) due to Rosenblatt (1952) to assess adequacy of predictive density models. The PITs are defined as $u_t = \int_{-\infty}^{y_t} f(v|\mathcal{F}_{t-1})dv$ where $f(y_t|\mathcal{F}_{t-1})$ is the conditional density of the process $\{y_t\}$. Under correct model specification, the PITs should be i.i.d. $U[0,1]$. Since then numerous articles have proposed extensions and alternative testing approaches to assess density forecasts by evaluating the statistical properties, uniformity and independence, of the $\{u_t\}$. Bai (2003) introduces a conditional Kolmogorov test to assess the properties of the $\{u_t\}$, but his test does not have power against violations of independence as noted by Corradi and Swanson (2006a), among others. Hong and Li (2005) propose a nonparametric-kernel-based test that has power against violations of both independence and density functional form and Hong et al. (2007) (HLZ hereafter) extend it to the out-of-sample framework. Corradi and Swanson (2006a,b) extend this literature in interesting directions by developing conditional Kolmogorov tests that allow dynamic misspecification with respect to the information set under the null hypothesis, or allow comparison of a number of possibly misspecified conditional density models. However, this literature, superbly reviewed by Corradi and Swanson (2006c), has only focused on univariate models. There has been only few studies that proposed methods to deal explicitly with multivariate predictive densities. [1] Diebold et al. (1999) generalized Diebold et al. (1998) approach to the multivariate case by decomposing the joint predictive distribution into its marginals and conditionals, whose respective PITs, as in the univariate case, should be i.i.d. $U[0,1]$. These properties are assessed by inspection of histograms and autocorrelograms of the PITs. More recently, Bai and Chen (2008) adopted the martingale transformation approach of Bai (2003) to the multivariate case, which requires the use of single-indexed empirical processes to make the computation of the test feasible. Kalliovirta (2008), extending the work of Berkowitz (2001) to the multivariate case, developed a battery of test statistics based on a further transformation of PITs to normality, i.e. $z_{it} = \Phi^{-1}(u_{it})$ for $i = 1, 2, \ldots, n$. The resulting processes z_{it} are also called *quantile residuals* of the assumed predictive model. Both Bai and Chen (2008) and Kalliovirta (2008) focus on in-sample dynamic specification testing.

In this paper we offer a new framework for out-of-sample evaluation of density forecasts in a multivariate context. We build up on the "autocontour" approach of González-Rivera

[1] There is a growing literature on testing the goodness of fit of various copula functions in the multivariate context, see Berg (2009) for a recent survey and power comparison.

et al. (2011) and González-Rivera and Yoldas (2010) applied to the second transformation of the PITs to normality, i.e. quantile residuals, as in Berkowitz (2001) and Kalliovirta (2008).

The autocontour approach is based on the generalized residuals of a location-scale time series model, i.e. $\varepsilon_t = (y_t - \mu_{t|t-1})/\sigma_{t|t-1}$ where $\mu_{t|t-1}$ is the conditional mean and $\sigma_{t|t-1}$ is the conditional standard deviation of the process. Under correct specification, the generalized residual should be i.i.d. with density $f(\varepsilon_t)$. The autocontour is the n-dimensional probability contour of the multivariate density of the process $\{\varepsilon_t\}$ under i.i.d-ness, e.g. for $n = 2$ and lag i the bivariate density is given by $f(\varepsilon_t, \varepsilon_{t-i}) = f(\varepsilon_t)f(\varepsilon_{t-i})$. Fixing the probability contained within a given autocontour is the basis of a testing procedure for model specification. We generalize this methodology to evaluate out-of-sample predictive densities based on quantile residuals. There are at least two advantages of working with quantile residuals. First, it allows for a broad range of specifications beyond the location-scale model, even though this is the most common in the econometrics literature. Our proposed methodology applies to density forecasts from non-linear models with conditional densities dependent on higher moments other than conditional mean and variance, such as Hansen (1994)'s autoregressive conditional density model; regime-switching models with state-dependent heteroskedasticity and Student-t innovations as in Perez-Quiros and Timmermann (2000); univariate and multivariate stochastic processes as in Liesenfeld and Richard (2003); and univariate and multivariate diffusion processes as in Hong and Li (2005). Generally speaking, for any specification, if the PITs can be retrieved, the quantile residuals will be easily calculated, and that will allow the implementation of autocontour-based testing. Secondly, the shape of autocontour may be difficult to obtain when the density functional form of the generalized residual becomes more complex. In this case, we need to implement numerical procedures to obtain the right probability mass of each autocontour, see González-Rivera et al. (2011). With the quantile residual we only deal with Gaussian autocontours that are analytically tractable and graphically very easy to implement. Our paper contributes to the limited literature on multivariate predictive density evaluation by proposing a computationally simple approach that uniquely combines formal testing with graphical illustration, making the visualization aspect one of the great advantages of this methodology. The proposed tests target the joint hypothesis of independence and normality of the quantile residual vector. The shape of the autocontours is the key to detect violations in both directions. The statistical properties of the tests developed in González-Rivera et al. (2011) translate easily into the out-of-sample context, so that standard asymptotic distributions hold. In some instances parameter uncertainty is asymptotically irrelevant, but in those where parameter estimation may play a role, we show that a parametric bootstrap procedure delivers very good finite sample properties of the tests. We illustrate our methodology with an empirical application on daily returns to value and growth equity portfolios. We evaluate the bivariate density forecasts of these two portfolios from 2006 to 2009. While in-sample, a DCC model with bivariate Student-t seems to be adequate,

out-of-sample, our tests rejects this density because it is unable to accommodate the high volatility events of 2007 and 2008.

The rest of the paper is organized as follows. In Section 2, we present the testing framework and discuss the role of parameter estimation in the distributions of the proposed tests. In Section 3, we provide a detailed assessment of the finite sample performance of the tests, and a comparison with the non-parametric tests of HLZ. In Section 4, we offer an empirical application on evaluating predictive densities for value and growth portfolios, and in Section 5 we conclude.

2 Testing Methodology

2.1 Quantile Residuals

Let $\mathbf{y}_t = (y_{1t}, \ldots, y_{nt})'$ denote the vector of interest with conditional density function $f(\mathbf{y}_t|\mathcal{F}_{t-1})$ where $\mathcal{F}_{t-1}$ is the information set available at time $t-1$, i.e. $\mathcal{F}_{t-1} = \sigma\{\mathbf{y}_t, \mathbf{y}_{t-1}, ...\}$.[2] Consider a parametric density forecast model for $\mathbf{y}_t$, say $g(\mathbf{y}_t|\mathbf{x}_{t-1}\ , \boldsymbol{\theta})$ where $\mathbf{x}_{t-1}$ is an $\mathcal{F}_{t-1}$ measurable vector and $\boldsymbol{\theta}$ is a vector of parameters such that $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^k$. Under correct density model specification we have $g(\mathbf{y}_t|\mathbf{x}_{t-1}\ , \boldsymbol{\theta}_0) = f(\mathbf{y}_t|\mathcal{F}_{t-1})$ *a.s.* for some unknown true parameter vector $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$. The predictive density function of $\mathbf{y}_t$ can be decomposed as follows

$$g(\mathbf{y}_t|\mathbf{x}_{t-1}, \boldsymbol{\theta}) = \prod_{j=1}^{n} g_j(y_{jt}|\mathbf{x}_{t-1}, \mathcal{A}_{j-1}, \boldsymbol{\theta}), \tag{1}$$

where $\mathcal{A}_{j-1} = \sigma\{Y_{1t}, \ldots, Y_{j-1,t}\}$.[3] Then, the PITs are given by $u_{jt} \equiv G_j(y_{jt}|\mathbf{x}_{t-1}, \boldsymbol{\theta}) = \int_{-\infty}^{y_{jt}} g_j(u|\mathbf{x}_{t-1}, \mathcal{A}_{j-1}, \boldsymbol{\theta})du$ and the vector of quantile residuals take the following form

$$\mathbf{z}_t(\boldsymbol{\theta}) = \begin{bmatrix} \Phi^{-1}(G_1(y_{1t}|\mathbf{x}_{t-1}, \boldsymbol{\theta})) \\ \Phi^{-1}(G_2(y_{2t}|\mathbf{x}_{t-1}, \boldsymbol{\theta})) \\ \vdots \\ \Phi^{-1}(G_n(y_{nt}|\mathbf{x}_{t-1}, \boldsymbol{\theta})) \end{bmatrix}, \tag{2}$$

where $\Phi^{-1}(.)$ denotes the inverse cumulative distribution function of the standard normal distribution. Under mild regularity conditions and correct specification of the conditional density, each $\Phi^{-1}(u_{jt})$ is i.i.d. N(0,1), so that the vector $\mathbf{z}_t(\boldsymbol{\theta}_0)$ will be i.i.d. N$(\mathbf{0}, \mathbf{I}_n)$.

Furthermore, Kalliovirta (2008) proposes a transformation of the vector of quantile residuals that yields a univariate stochastic process, which is also i.i.d. standard normal. Specifically, by generalizing the transformation proposed in Clements and Smith (2000) and Clements and

[2] For simplicity we do not consider predictor variables in the information set, but the extension is straightforward.

[3] In general, the ordering does not have to be from 1 to n, i.e. the joint density can be decomposed in $n!$ ways. The ordering does not have an impact on the asymptotic distribution of the tests but it may affect power in finite samples. See Hong and Li (2005) for a discussion of this point.

Smith (2002), she shows that the univariate process $v_t(\boldsymbol{\theta}_0) = w_t(\boldsymbol{\theta}_0) \sum_{j=0}^{n-1} \frac{(-1)^j}{j!} [\ln(w_t(\boldsymbol{\theta}_0))]^j$, where $w_t(\boldsymbol{\theta}_0) = \Pi_{j=1}^{n} G_j(y_{jt}|\mathbf{x}_{t-1}, \boldsymbol{\theta}_0)$, is i.i.d. uniformly distributed. Then, the transformation to normality $q_t(\boldsymbol{\theta}_0) = \Phi^{-1}(v_t(\boldsymbol{\theta}_0))$ delivers a quantile residual that is i.i.d. $\mathrm{N}(0,1)$.

2.2 Test Statistics

The i.i.d. normality of the quantile residuals will hold only when the assumed conditional density forecast model coincides with the true conditional density of $\mathbf{y}_t$, i.e. $g(\mathbf{y}_t|\mathbf{x}_{t-1}, \boldsymbol{\theta}_0) = f(\mathbf{y}_t|\mathcal{F}_{t-1})$ *a.s.* Hence, the adequacy of any density forecast model can be evaluated by checking the i.i.d. normality of the quantile residuals. For now we will assume that $\boldsymbol{\theta}_0$ is known. We will relax this assumption when we discuss the impact of parameter uncertainty on the distribution of our test statistics.

We are interested in testing the following null hypotheses on

(i) the transformed vector of quantile residuals

$$H_0 : q_t(\boldsymbol{\theta}_0) \sim \text{i.i.d. } \mathrm{N}(0,1). \tag{3}$$

and

(ii) the vector of the quantile residuals

$$H_0 : \mathbf{z}_t(\boldsymbol{\theta}_0) \sim \text{i.i.d. } \mathrm{N}(\mathbf{0}, \mathbf{I}_n), \tag{4}$$

In both cases H_1 is simply negation of the null. We develop test statistics that are designed to test the joint hypothesis of independence and density functional form on $\mathbf{z}_t$ and q_t.[4]

Let us now focus on the univariate aggregated quantile residual process and consider the joint distribution of q_t and q_{t-k} for $k \leq K < \infty$. Due to independence and normality implied by the null hypothesis, their joint pdf is given by $\phi(q_t, q_{t-k}) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}\left(q_t^2 + q_{t-k}^2\right)\right)$. For this process we define the autocontour, $ACR_q^{\alpha,k}$, as the set of points in the hyperplane (q_t, q_{t-k}) that results from horizontally slicing the joint density function at a fixed value, say ϕ_α, to guarantee that the resulting set contains $100\alpha\%$ of observations. This is effectively the probability contour plot of $\phi(q_t, q_{t-k})$ with probability mass equal to α. The formal definition of $ACR_q^{\alpha,k}$ is

$$ACR_q^{\alpha,k} := \left\{ B\left(q_t, q_{t-k}\right) \subset \mathbb{R}^2 \left| \int_{-\sqrt{a_\alpha}}^{\sqrt{a_\alpha}} \int_{-h(q_t)}^{h(q_t)} \frac{1}{2\pi} \exp\left(-\frac{1}{2}\left(q_t^2 + q_{t-k}^2\right)\right) dq_t dq_{t-k} \leq \alpha \right. \right\}, \tag{5}$$

where $B(.,.)$ is a set in $\mathbb{R}^2$, $a_\alpha = -2\ln(2\pi\phi_\alpha)$, and $h(q_t) = \sqrt{a_\alpha - q_t^2}$. $ACR_q^{\alpha,k}$ will have $100\alpha\%$ coverage only when both assumptions under the null, correct dynamic specification and density functional form, are satisfied. A graphical illustration for different coverage levels is given in Figure 1.

[4] For ease of exposition we will suppress the parameter vector argument in the quantile residuals until we deal with the parameter uncertainty problem.

Let T and P denote the number of observations in the full sample and the prediction sample respectively. We define an indicator series with respect to the $ACR_q^{\alpha,k}$ autocontour as follows

$$I_{q,t}^{\alpha,k} = \mathbb{I}\left((q_t, q_{t-k}) \notin ACR_q^{\alpha,k}\right) \quad t = R+k+1, \ldots, T, \tag{6}$$

where $R = T - P$, and $\mathbb{I}(.)$ denotes the usual indicator function. For the normal autocontour, we construct this indicator series as follows

$$I_{q,t}^{\alpha,k} = \mathbb{I}\left(q_t^2 + q_{t-k}^2 > a_\alpha\right), \quad t = R+k+1, \ldots, T. \tag{7}$$

Given the quantile residuals, we only need to obtain the a_α value to make this definition operational. Since $q_t^2 + q_{t-k}^2$ is chi-squared distributed with two degrees of freedom, it follows that $a_\alpha = -2\ln(1-\alpha)$. Let $p_\alpha = 1-\alpha$, then it is straightforward to show that $E[I_{q,t}^{\alpha,k}] = 1-\alpha$ and $Var(I_{q,t}^{\alpha,k}) = \alpha(1-\alpha)$. Furthermore, $I_{q,t}^{\alpha,k}$ is autocorrelated with autocovariance function

$$Cov\left(I_{q,t}^{\alpha,k}, I_{q,t-s}^{\alpha,k}\right) = \mathbb{I}(s=k)\left[P\left(I_{q,t}^{\alpha,k} = 1, I_{q,t-s}^{\alpha,k} = 1\right) - p_\alpha^2\right].$$

Hence, $I_{q,t}^{\alpha,k}$ exhibits a dependence structure similar to a restricted MA(k) process. By exploiting the statistical properties of this indicator process under the null, we will evaluate the adequacy of the one-step-ahead density forecast.

Two tests are provided. By fixing the probability α and the lag k, we can construct a t-test based on the sample values of p_α. Furthermore, by jointly analyzing several autocontour coverage levels, say $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m)'$ we can construct a chi-squared test based on the corresponding sample values of the vector $\mathbf{p}_{\boldsymbol{\alpha}} = (p_{\alpha_1}, \ldots, p_{\alpha_m})'$. Both statistics amount to test for the independence and density functional form of the aggregated quantile residual q_t. [5]

A t-test to evaluate one-step-ahead density forecasts *Define* $\hat{p}_q^{\alpha,k} = \frac{1}{P-k}\sum_{t=R+k+1}^{T} I_{q,t}^{\alpha,k}$. *Under the null hypothesis given in Equation* (3)*, we have*

$$t_q^{\alpha,k} = \frac{\sqrt{P-k}(\hat{p}_q^{\alpha,k} - p_\alpha)}{\sigma_q^{\alpha,k}} \xrightarrow{d} \mathrm{N}(0,1),$$

where $\sigma_q^{\alpha,k} = \sqrt{p_\alpha(1-p_\alpha) + 2Cov\left(I_{q,t}^{\alpha,k}, I_{q,t-k}^{\alpha,k}\right)}$.

We can examine the lag structure of $t_q^{\alpha,k}$ for $k = 1, \ldots, K$ and collect those t-statistics in a graph, which we call *autocontourgram*, (see Section 4 for various empirical examples). In certain applications, such as financial risk management, a specific coverage level may be of particular interest, which makes $t_q^{\alpha,k}$ very useful. In other instances, it may be desirable to

[5] The mathematical proofs for both tests are straightforward extensions of those provided in González-Rivera et al. (2011)

construct a test statistic that aggregates information from multiple autocontours and covers the entire density instead of specific regions. This is provided by the following test statistic.

A chi-squared test to evaluate one-step-ahead density forecasts Let us consider a set of coverage levels, say $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m)'$, and the vectors $\mathbf{p}_{\boldsymbol{\alpha}} = (p_{\alpha_1}, \ldots, p_{\alpha_m})'$, $\widehat{\mathbf{p}}_q^{\boldsymbol{\alpha},k} = (\hat{p}_q^{\alpha_1,k}, \ldots, \hat{p}_q^{\alpha_m,k})'$ where $\hat{p}_q^{\alpha_i,k} = \frac{1}{P-k}\sum_{t=R+k+1}^{T} I_{q,t}^{\alpha_i,k}$. Under the null hypothesis given in Equation (3), we have

$$J_q^{\boldsymbol{\alpha},k} = (P-k)(\widehat{\mathbf{p}}_q^{\boldsymbol{\alpha},k} - \mathbf{p}_{\boldsymbol{\alpha}})'\boldsymbol{\Xi}^{-1}(\widehat{\mathbf{p}}_q^{\boldsymbol{\alpha},k} - \mathbf{p}_{\boldsymbol{\alpha}}) \xrightarrow{d} \chi^2(m),$$

$$\xi_{ij} = \min(p_{\alpha_i}, p_{\alpha_j}) - p_{\alpha_i}p_{\alpha_j} + Cov(I_{q,t}^{\alpha_i,k}, I_{q,t-k}^{\alpha_j,k}) + Cov(I_{q,t}^{\alpha_j,k}, I_{q,t-k}^{\alpha_i,k}).$$

Now consider the vector of quantile residuals, $\mathbf{z}_t$ and let $\mathbf{r}_t = (\mathbf{z}_t', \mathbf{z}_{t-k}')'$. Then, $\mathbf{r}_t$ is i.i.d. $\mathrm{N}(\mathbf{0}, \mathbf{I}_{2n})$. In this case, the autocontour with $100\alpha\%$ coverage will be a $2n$-dimensional sphere. The formal definition is given by

$$ACR_{\mathbf{z}}^{\alpha,k} := \left\{ B(\mathbf{r}_t) \subset \mathbb{R}^{2n} \left| \int_{-h_1}^{h_1} \cdots \int_{-h_{2n}}^{h_{2n}} \frac{1}{(2\pi)^n} \exp\left(-\sum_{i=1}^{2n} r_{it}^2\right) dr_{1t} \ldots dr_{2n,t} \leq \alpha \right. \right\}, \quad (8)$$

where $h_1 = \sqrt{d_\lambda}$, $h_i = \sqrt{d_\lambda - \sum_{j=1}^{i-1} r_{jt}^2}$ for $i = 2, \ldots, 2n$, $\lambda \leq \alpha$, and $d_\alpha = \inf\{d : \aleph(d) \leq \alpha\}$ where $\aleph(.)$ is the cdf of a chi-squared random variable with $2n$ degrees of freedom. Given the dimension of the vector of quantile residuals and the coverage level, d_α can be easily computed with numerical methods.

As in the univariate case, we proceed to construct an indicator series with respect to this autocontour as follows

$$I_{\mathbf{z},t}^{\alpha,k} = \mathbb{I}\left(\mathbf{r}_t'\mathbf{r}_t > d_\alpha\right), \quad t = R+k+1, \ldots, T. \quad (9)$$

This indicator process has the same statistical properties as those of the indicator for the univariate case and, consequently the t and chi-squared test statistics will be constructed exactly in the same way as describe above. We will denote these test statistics as $t_{\mathbf{z}}^{\alpha,k}$ and $J_{\mathbf{z}}^{\boldsymbol{\alpha},k}$. Finally, we can follow the same strategy for the individual elements of the vector of quantile residuals to obtain a more refined picture of the properties of the density forecasts. We will denote the corresponding test statistics as $t_{z_i}^{\alpha,k}$ and $J_{z_i}^{\boldsymbol{\alpha},k}$.

These tests will have power to detect potential shortcomings of a density forecast, those coming either from misspecified dynamics, or from an incorrect density functional form, or from both. The fundamental reason is that the tests deal explicitly with both implications of the null hypothesis through the shape of the autocontours. Discrepancies between the theoretical autocontour under the null and the actual autocontour are the key to understand the power of the tests. For example, assume that the postulated density forecast model belongs to the location scale family and dynamics are correctly captured, but the assumed

density form is incorrect, e.g. the true density is leptokurtic while multivariate normality is assumed. The neglected leptokurtosis in the underlying process will be reflected in the quantile residuals. In that case $\mathbf{z}_t$ (and q_t) will be still i.i.d. but not normally distributed. The actual autocontours will deviate from the spheres (circles) implied by normality. The discrepancy between the two autocontours will cause a difference in the actual versus assumed coverage levels, which will cause the tests to reject the null model. Now suppose that the postulated model correctly captures the density functional form, but the dynamics are not fully modeled so that there is remaining linear dependence in quantile residuals. This will translate into actual elliptical autocontours as opposed to circles implied by the null and, again, the null will be rejected. Furthermore, whenever there is neglected linear dependence, both $t_q^{\alpha,k}$ and $J_q^{\boldsymbol{\alpha},k}$ ($t_{\mathbf{z}}^{\alpha,k}$ and $J_{\mathbf{z}}^{\boldsymbol{\alpha},k}$) statistics will exhibit a fast decaying pattern with respect to the lag displacement as linear dependence will die off rather quickly. On the other hand, they will display persistence when dynamic misspecification is of nonlinear type, e.g. neglected ARCH effects in financial data.

Though *a priori* it would be possible to choose α optimally for some defined objective function, e.g. the power function of the tests, from an empirical perspective, it is the application of interest that should guide the choice of α. For instance, if the researcher is interested in Value-at-Risk (VaR) calculations, the modeling of the tails of the conditional distribution is the most relevant, and consequently the α's of choice should be those corresponding to the extreme quantiles, i.e. 90, 95, 99%. In the financial duration literature, if the interest lies on the modeling of short durations (highly liquid assets), the most relevant α's will those corresponding to the most central quantiles, i.e. 1, 5, 10%. As a starting point in the implementation of our methodology, we recommend canvassing the full density with a chi-squared test, and upon rejection of the null, to examine individual autocontours to assess where the rejection comes from. As for the choice of k, this is analogous to examining the Q-statistics in a classical autocorrelogram where Q-statistics are reported for a large set of displacements. Likewise, with our proposed t-tests and chi-squared tests we will report their values for a large set of displacements (see simulations and empirical sections) and assess whether or not the dynamics of the model are well-specified.

Up to now we have assumed that the parameters of the density model are known but in practice the parameters will be estimated. Ignoring parameter uncertainty can result in substantial size distortions of the tests, as shown in a recent paper by Chen (2010) in the context of moment based tests for univariate density forecast models. In an out-of-sample context, the relevance of parameter uncertainty depends on the forecasting scheme (fixed, rolling, or recursive) as well as on the size of the prediction sample relative to the estimation sample. Here we provide a theoretical analysis in case of the fixed scheme, as in HLZ, but similar results can be obtained under recursive and rolling schemes.

Taking a mean value expansion of $\hat{p}_q^{\alpha,k}(\boldsymbol{\theta})$ around $\boldsymbol{\theta}_0$, we can obtain the following equation

$$\sqrt{P}\left(\hat{p}_q^{\alpha,k}(\widehat{\boldsymbol{\theta}}) - p_\alpha\right) = \sqrt{P}\left(\hat{p}_q^{\alpha,k}(\boldsymbol{\theta}_0) - p_\alpha\right) + \sqrt{\frac{P}{R}}\sqrt{R}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)' \lim_{P\to\infty} E\left[\left.\frac{\partial \hat{p}_q^{\alpha,k}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right] + o_p(1), \tag{10}$$

where $\hat{p}_q^{\alpha,k}(\widehat{\boldsymbol{\theta}}) = \frac{1}{P-k}\sum_{t=R+k+1}^{T} \mathbb{I}(q_t^2(\widehat{\boldsymbol{\theta}}) + q_{t-k}^2(\widehat{\boldsymbol{\theta}}) > a_\alpha)$ and $\widehat{\boldsymbol{\theta}}$ is the estimator obtained from the first R observations.[6] The parameter estimators are assumed to be $\sqrt{R}$-consistent, i.e. $(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = O_P(R^{-1/2})$. In general, this condition will be easily satisfied by m-estimators, such as the quasi maximum-likelihood (QML) estimator. If $R \to \infty$, $P \to \infty$, and $P/R \to 0$ as $T \to \infty$, the second term on the right hand side of Equation (10) will be asymptotically negligible provided that the gradient term is bounded. Therefore, as long as the ratio of the prediction sample to the estimation sample tends to zero as the total sample size grows indefinitely, our test statistics can be applied to quantile residuals based on estimated parameters without any adjustments. In situations where the condition $P/R \to 0$ is violated, we can bootstrap the tests to approximate their asymptotic distribution. Moreover, when $P/R \to 0$ condition is satisfied, we can obtain improvements in finite sample performance through bootstrap as the test statistics are asymptotically distribution free under this condition, e.g. Horowitz (2001). In our context, a parametric bootstrap is particularly relevant as the null model completely specifies the conditional distribution of the data. Specifically, we generate B samples of size T from $g(\mathbf{y}_t|\mathbf{x}_{t-1}, \widehat{\boldsymbol{\theta}})$. Let $\widehat{\boldsymbol{\theta}}_b$ denote the estimator under the fixed scheme from the bth bootstrap sample, then the quantile residuals are calculated from $g(\mathbf{y}_t|\mathbf{x}_{t-1}, \widehat{\boldsymbol{\theta}}_b)$. The resulting quantile residuals are used to calculate the test statistics, $t_q^{\alpha,k}(b)$ for $b = 1, \ldots, B$. Then, the bootstrap approximation to the p-value is given by

$$\tilde{p}(t_q^{\alpha,k}) = \frac{1}{B}\sum_{b=1}^{B} \mathbb{I}\left(|t_q^{\alpha,k}(b)| > |t_q^{\alpha,k}|\right). \tag{11}$$

Under suitable regularity conditions, asymptotic expansions can be used to show that the parametric bootstrap converges to the true distribution of the test statistic at a rate of $\sqrt{P}$ even when $P/R \approx 0$ condition is violated. When $P/R \approx 0$ holds, the parametric bootstrap converges to the true distribution of the test statistic faster than $\sqrt{P}$. This approach provides remarkable results in finite samples as shown by the following Monte Carlos simulation results.

3 Finite Sample Performance

In this section we examine the size and power of the tests for several bivariate data generating processes paying special attention to the size of the prediction sample relative to the estimation sample. We also offer a comparison with the tests of HLZ applied to the aggregated vector of quantile residuals. These authors entertain the same joint hypothesis of iid-ness and correct

[6] The analysis here exclusively focuses on the t-statistic for the aggregated quantile residual and an individual autocontour. The same line of reasoning applies to the vector of quantile residuals and the chi-squared statistic.

density functional form as our autocontour tests, so that the comparison will be fair and informative. In addition, since their tests are non-parametric, the comparison will offer an assessment of the merits of parametric autocontour-based tests versus non-parametric kernel-based tests.

HLZ tests are based on the PITs, which must be i.i.d. U[0,1] under correct specification of the density forecast. Their tests compare the joint density of the pair $\{u_t, u_{t-k}\}$ with the product of two independent U[0,1] densities, which is equal to unity under the null. They propose two tests: (i) for a given displacement k, $\hat{Q}(k) \to N(0,1)$ where the test statistic is the properly centered and scaled version of the non-parametric kernel based estimator of the joint density of $\{u_t, u_{t-k}\}$ using a boundry-modified kernel; and (ii) $\hat{W}(K) = K^{-1/2} \sum_{k=1}^{K} \hat{Q}(k) \to N(0,1)$. The $\hat{Q}(k)$ test is similar to our autocontour tests $t_q^{\alpha,k}$ and $J_q^{\boldsymbol{\alpha},k}$, for which the displacement k is also fixed.

3.1 Size

We simulate data from two VAR(1) systems under bivariate normality and bivariate Student-t:

$$\begin{aligned} \text{Size 1}: &\quad \mathbf{y}_t = \mathbf{A}\mathbf{y}_{t-1} + \boldsymbol{\Sigma}^{1/2}\boldsymbol{\epsilon}_t, \\ \text{Size 2}: &\quad \mathbf{y}_t = \mathbf{A}\mathbf{y}_{t-1} + \boldsymbol{\Sigma}^{1/2}\boldsymbol{\varepsilon}_t, \end{aligned} \tag{12}$$

where

$$\mathbf{A} = \begin{bmatrix} 0.15 & 0.05 \\ 0.15 & 0.45 \end{bmatrix},$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.3 \\ 0.3 & 1 \end{bmatrix},$$

$\boldsymbol{\epsilon}_t$ is an i.i.d. standard Normal vector, and $\boldsymbol{\varepsilon}_t$ i.i.d. Student-t vector with degrees of freedom equal to 5 with identity covariance matrix. We estimate $\mathbf{A}$ and $\boldsymbol{\Sigma}$ under the fixed forecasting scheme with Least Squares and apply our tests to the quantile residuals described above. The number of Monte Carlo replications is 1000. We set $T = 5000$ and consider $P \in \{250, 500, 1000, 2000\}$. The nominal size level is 5%. The set of autocontour coverage levels is given by $\boldsymbol{\alpha} = \{0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.99\}$. Finally, we set the number of bootrstrap replications equal to 500, i.e. $B = 500$.

The results for t-statistics are presented in Tables 1 and 2, and for the chi-squared statistics in Table 3. The t-statistics in general have very good size properties for both tests, t_q and $t_{\mathbf{z}}$, and for both assumed bivariate densities. For the smallest value of the evaluation sample size considered ($P = 250$), the test statistics are slightly undersized in some cases, such as $t_q^{0.99,1}$ under normal distribution. This is expected because there may not be enough variation in the data at the extreme coverage levels to obtain a reliable estimate of the violation percentage for relatively small values of P. As P increases, this tendency disappears as expected. Another critical observation is that the parametric bootstrap scheme delivers

excellent results in situations where asymptotic irrelevance of parameter estimation is hard to justify. This can be directly seen from the last rows of Tables 1 - 2 where $P/R = 0.67$.

The results for the chi-squared test statistics in Table 3 are similar to those of the t-statistics. They enjoy empirical sizes fluctuating around the nominal size of 5% even in those cases where the P/R ratio is high.[7]

We also apply the HLZ tests to the aggregated vector of quantile residuals, q_t. The critical values are calculated by implementing the distribution-free simulation procedure proposed by the authors, which is supposed to correct for finite sample bias. The results are presented in Table 4. When the prediction sample is relatively small the size of both tests, $\hat{Q}(k)$ and $\hat{W}(K)$ for bivariate Normal and bivariate Student-t, is acceptable. When the prediction sample is large, we observe severe size distortions, e.g. when $P = 2000$ and for the bivariate Student-t case, the size of the test $\hat{W}(5)$ is 13.5%, more than twice the nominal size of 5%. As P increases, the calculated critical values approach asymptotic critical values since the HLZ simulation procedure depends on sample size only. On the other hand, the effect of parameter estimation become more pronounced as P/R grows. These two effects combined result in the observed size distortions for large values of P in our simulations.

3.2 Power

In order to assess the power of the tests, we choose the VAR(1) specification with bivariate normality (described above as "Size 1") as the model under the null hypothesis, and we consider three alternative DGPs that deviate from the null in particular ways. First, we consider a model with linear dynamics as in the null model but with a non-normal density. Specifically we generate data from the multivariate Student-t distribution with 5 degrees of freedom. This model corresponds to that described above as "Size 2" and we will name it "Power 1". With this model, we will assess deviations from density functional form in the density forecast. The second DGP introduces non-linear dynamics in the conditional mean vector:

$$\text{Power 2}: \quad \mathbf{y}_t = \mathbb{I}(y_{1,t-1} < 0)\mathbf{A_1}\mathbf{y}_{t-1} + \mathbb{I}(y_{1,t-1} \geq 0)\mathbf{A_2}\mathbf{y}_{t-1} + \mathbf{\Sigma}^{1/2}\boldsymbol{\epsilon}_t, \tag{13}$$

where,

$$\mathbf{A_1} = \begin{bmatrix} 0.7 & 0 \\ 0.3 & 0.7 \end{bmatrix},$$

$\mathbf{A_2} = -\mathbf{A_1}$, and $\boldsymbol{\epsilon}_t$ is an i.i.d. standard Normal vector. Since we maintain bivariate normality of the vector of innovations, we would like to assess power in the direction of misspecified dynamics in the density forecast. Finally, the third DGP will combine non-normality with non-linear dynamics in higher moments than the mean. We consider a model with time-varying variances and correlations as in the Dynamic Conditional Correlation (DCC) model

[7] When we completely rely on asymptotic irrelevance arguments and do not bootstrap the distributions of the tests we observe some non-negligible size distortions especially when P/R is relatively large. These results are available upon request.

of Engle (2002). This third DGP is given by

$$\text{Power 3}: \quad \mathbf{y}_t = \mathbf{A}\mathbf{y}_{t-1} + \boldsymbol{\epsilon}_t, \tag{14}$$

where $\boldsymbol{\epsilon}_t \sim \mathrm{N}(\mathbf{0}, \mathbf{H}_t)$, $\mathbf{H}_t = \mathbf{D}_t\mathbf{R}_t\mathbf{D}_t$, $\mathbf{D}_t = \mathrm{diag}\{h_{11,t}, \ldots, h_{nn,t}\}$, and each element of $\mathbf{D}_t$ is modeled as a standard GARCH(1,1) process:

$$h_{ii,t} = \omega_i + \alpha_i \epsilon_{i,t-1}^2 + \beta_i h_{ii,t-1}, \quad i = 1, \ldots, n.$$

The model is complete by defining dynamics of the time-varying correlation matrix, $\mathbf{R}_t$. Let $\mathbf{e}_t = \mathbf{D}_t^{-1}\boldsymbol{\epsilon}_t$, then $R_{ij,t} = \gamma_{ij,t}/(\sqrt{\gamma_{ii,t}}\sqrt{\gamma_{jj,t}})$ where

$$\boldsymbol{\Gamma}_t = (1 - \alpha - \beta)\overline{\boldsymbol{\Gamma}} + \alpha \mathbf{e}_{t-1}\mathbf{e}_{t-1}' + \beta \boldsymbol{\Gamma}_{t-1},$$

and $\overline{\boldsymbol{\Gamma}} = E\left[\mathbf{e}_t\mathbf{e}_t'\right]$. We set $\alpha_i = 0.15$, $\beta_i = 0.8$, $\omega_i = 1 - \alpha_i - \beta_i$, $\alpha = 0.15$, and $\beta = 0.8$.

Table 5 summarizes rejection rates for $t_{\mathbf{z}}$ statistics at the 5% nominal size level. When the data is generated from "Power 1" the tests are extremely powerful in detecting deviations from normality. We observe rejection rates larger then 90% even with small prediction samples $P = 250$. The relatively high rejection rates at small and large coverage levels are due to neglected leptokurtosis. When the data is generated from "Power 2", the rejection rates are not as high as those from "Power 1", and it seems that we need larger samples than 250 to obtain rejection rates above 70%, but overall the results are quite satisfactory. The largest rejection rates when neglected nonlinearity is at stake happen for autocontours with probability mass $0.1 \leq \alpha \leq 0.4$ and $\alpha = 0.99$. This means that small to intermediate autocontours and the autocontour associated with the tail of the distribution are more sensitive to this form of deviation from the null. For $\alpha \in \{0.8, 0.9\}$ the rejection rates are very low implying that the null and the alternative are close to each other for those particular coverage levels. Finally, when we consider time varying variance and correlations under the alternative, we observe that the rejection rates tend to be larger than those for "Power 2" but smaller than those for "Power 1". For a given prediction sample, the power seems to be more uniform across autocontours than in "Power 1" and "Power 2". In Table 6 we present the rejection rates for the test with aggregated quantile residuals t_q. The conclusions are similar to those for Table 5, however the statistics $t_{\mathbf{z}}$ are more powerful than the t_q tests across the three DGPs considered.

Table 7 summarizes the results for the chi-squared tests. Similar to the case of t-statistics, the chi-squared statistics based on the vector of quantile residuals are more powerful than those based on the aggregated process. For DGPs "Power 1" and "Power 3", the rejection rates are similar with respect to the lag order. This is due to misspecification of density functional form (Power 1) and to neglected variance/correlation dynamics (Power 3) that create leptokurtic behavior in the quantile residuals at all lags. In DGP 'Power 2', the sharp

drop in the rejection rates of both $J_{\mathbf{z}}$ and J_q from $k = 1$ to $k \geq 2$ is particularly noteworthy as it distinguishes the case of misspecified mean dynamics from other sources of misspecification.

For comparison purposes, we also report the power simulation results for the HLZ tests in Table 8, where the tests statistics are calculated for the aggregated quantile residual process, q_t. Note that these power simulation results are not size adjusted. Under "Power 1" DGP, the HLZ tests have similar performance to the autocontour tests based on q_t, but they are less powerful compared to our autoconotur tests based on $\mathbf{z}_t$. Similar observations apply under the third DGP, "Power 3". For the case of "Power 2", the HLZ tests perform better than both autocontour tests, especially in small samples. Given the size distortions of the HLZ tests, especially under non-normal DGPs, the results indicate that the autocontour tests outperfom HLZ tests for detecting deviations from specified density and dependence through moments higher than conditional mean, while they perform comparably or slightly worse when it comes to violations of the null through dependence in conditional mean.

4 Empirical Illustration

In this section we apply our methodology to the daily returns on value and growth portfolios. Value and size are the most common styles in equity investments. For example, Morningstar provides an equity style box as a nine-cell grid that is used to identify the investment styles of domestic equity funds with respect to value and size. Style portfolios became subject of extensive academic research especially after the seminal work of Fama and French (1993). Even though they have been analyzed for portfolio allocation decisions at the monthly frequency, e.g. Guidolin and Timmermann (2008) and Patton (2004), to our knowledge no existing study has investigated the bivariate density forecast model for value and growth portfolios at the daily frequency.

We use the Fama-French data set available from the online Data Library of Kenneth R. French. Stocks are sorted into small and big categories with respect to their market capitalization. They are also sorted with respect to the ratio of market value to book value into three categories: value, neutral, and growth. Fama and French then consider the intersection of these categories to form a six cell grid.[8] We construct our value (growth) portfolio as the average of small value (growth) and big value (growth) portfolios. Formally, we have

$$\begin{aligned} r_t^{Value} &= \frac{1}{2}(\text{Small Capitalization Value} + \text{Big Capitalization Value}), \\ r_t^{Growth} &= \frac{1}{2}(\text{Small Capitalization Growth} + \text{Big Capitalization Growth}). \end{aligned} \tag{15}$$

Then, the vector of interest is given by $\mathbf{r}_t = (r_t^{Value}, r_t^{Growth})'$. Our daily sample runs from January 2, 1990 to October 30, 2009, providing a total of 5001 observations and we hold back

[8] For further details regarding the construction of portfolios and calculation of returns please refer to Kenneth French's web site.

the last 1000 observations for out-of-sample evaluation. Summary statistics of returns (%) can be found in Table 9. The mean daily return is close to zero for both portfolios in the estimation and prediction samples. We observe that the standard deviation, the range, and the kurtosis of both portfolios are substantially larger in the prediction sample than in the estimation sample. We should keep in mind that the prediction sample includes the turbulent financial events of 2008 and 2009.

We set the conditional mean equal to zero for the vector of returns, which is common practice when modeling conditional distributions of daily/weekly returns, e.g. J.P.Morgan (1996), and Capiello et al. (2006). This is mainly because the first moment is difficult to model at daily and higher frequencies due to the presence of noise. Furthermore, variation in the first moments is an order of magnitude smaller than the variation in the second moments for high frequency returns (see Andersen et al. (2010) for a detailed illustration of this point). We consider the DCC model of Engle (2002) under bivariate Normal and bivariate Student-t distributions to model the fat tails and the time varying second moments of the data. The model is given by

$$\mathbf{r}_t = \mathbf{H}_t^{1/2} \boldsymbol{\varepsilon}_t, \tag{16}$$

where $\mathbf{H}_t = \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t$, and $\mathbf{D}_t = \text{diag}\{h_{11,t}, \ldots, h_{22,t}\}$. In this setup, $\mathbf{R}_t$ is the time-varying correlation matrix. Based on model selection criteria and specification tests we model the individual variances as a threshold GARCH process as in Glosten et al. (1993):

$$h_{ii,t} = \omega_i + \alpha_i r_{i,t-1}^2 + \beta_i h_{ii,t-1} + \mathbb{I}(r_{i,t-1} < 0)\delta_i r_{i,t-1}^2, \quad i = 1, 2.$$

This specification captures the well known negative correlation between realized returns and volatility (the leverage effect/ the volatility feedback), e.g. Bollerslev et al. (2006). Let $\mathbf{e}_t = \mathbf{D}_t^{-1}\mathbf{r}_t$. Then, the dynamics of the correlation matrix is given by $R_{ij,t} = \gamma_{ij,t}/(\sqrt{\gamma_{ii,t}}\sqrt{\gamma_{jj,t}})$ where

$$\boldsymbol{\Gamma}_t = (1 - \alpha - \beta)\overline{\boldsymbol{\Gamma}} + \alpha \mathbf{e}_{t-1}\mathbf{e}_{t-1}' + \beta \boldsymbol{\Gamma}_{t-1},$$

and $\overline{\boldsymbol{\Gamma}} = E\left[\mathbf{e}_t \mathbf{e}_t'\right]$.[9] The density forecast model is complete with the specification of the distribution of $\boldsymbol{\varepsilon}_t$. We consider two popular densities to this end: multivariate normal and multivariate Student-t distributions. For the Student-t distribution, we still estimate the model under normal likelihood with a QML interpretation. The degrees of freedom for Student-t distribution is then estimated with the method of moments based on the conventional standardized residuals of the DCC model for each series, e.g. Bontemps and Meddahi (2005). The average of the estimates is taken as the common degrees of freedom for the density forecast model e.g. Pesaran et al. (2010). Based on this procedure we estimate the joint degrees of

[9] We also considered the asymmetric DCC model proposed in Capiello et al. (2006) and found that there is a weak but statistically significant leverage effect in correlation dynamics. However, this model created convergence problems in bootstrap replications.

freedom parameter as 10. Finally, we set the prediction sample $P = 1000$ and consider a fixed scheme.

The results for the chi-squared statistics are presented in Figures 2 (under normality) and 3 (under Student-t). All p-values are calculated using the parametric bootstrap scheme outlined in the previous section. From Panel-a of Figure 2 we observe that $J_{\mathbf{z}}^{\alpha,k}$ takes large values, with an average of approximately 38, and it is significant for all lags at 5% level. The monotonic behavior of the test statistic with respect to the lag order indicates that the DCC model does a satisfactory job of capturing the dynamics of the data. Similar observations apply to $J_q^{\alpha,k}$ statistic, presented in Panel-b of Figure 2. Under Student-t, Figure 3, the values of the statistics $J_{\mathbf{z}}^{\alpha,k}$ and $J_q^{\alpha,k}$ are smaller than those under normality indicating that the Student-t density forecast is a better fit than the Normal density forecast. However, $J_{\mathbf{z}}^{\alpha,k}$ points towards a mild rejection while $J_q^{\alpha,k}$ indicates a clear rejection of the Student-t density forecast at the conventional 5% significance level.

To understand where the rejection comes from, we examine the t-statistics $t_q^{\alpha,k}$ for different α coverage levels. In Figures 4 and 5 we report the values of the t-statistics for normal and Student-t for $\alpha \in \{0.1, 0.95\}$. For a normal density forecast, the tests fail to reject for the central autocontour ($\alpha = 0.1$) while it clearly rejects for the tail autocontour ($\alpha = 0.95$). For a Student-t density forecast, the rejection comes from both autocontours, though the rejection is much stronger for the 95% autocontour. Overall, the DCC model provides a good specification of the dynamics of the data and the bivariate Student-t density forecast is an improvement over the normal density forecast, but it is not entirely satisfactory as there are significant outlier returns in both portfolios, value and growth, coming from the high volatile periods in late 2008 and early 2009. The improvement provided by the Student-t distribution is also evident from Figure 6, which provides (q_t, q_{t-1}) scatter plots with normal autocontours superimposed on the quantile residuals. Figure 7 shows the HLZ $\hat{Q}$ test calculated for the aggregated quantile residual process under normal and Student-t densities. The results are in line with those of the autocontour tests based on q_t. The main difference is that our tests reject conditional normal DCC model more strongly while the HLZ tests indicate a stronger rejection of the conditional Student-t DCC model. This is likely due to the aforementioned size distortions of the HLZ tests, which are more pronounced under non-normal densities.

Though not directly comparable, our results are in contrast with those of Bai and Chen (2008). They provide in-sample evaluation of a bivariate system of monthly returns on IBM Stock and S&P 500 index and fail to reject a bivariate GARCH model coupled with Student-t distribution. On the other hand, we agree with Pesaran and Pesaran (2010) who conducted an out-of-sample evaluation of an equally-weighted portfolio of 17 assets. They apply the Kolmogorov-Smirnov test to the PITs of the Student-t DCC model and do not reject the null but they reject the Student-t DCC model with respect to VaR violations. They argue that tests focusing on the tail of the distribution prove to be more powerful. Our results are

consistent with this conclusion, which illustrates the usefulness of our methodology in terms of the flexibility it allows to focus on the entire distribution and/or specific regions.

5 Concluding Remarks

Noting that the literature on multivariate predictive densities is rather thin, we have aimed to develop a new framework for the out-of-sample evaluation of multivariate density forecasts building up on the concept of "autocontour" introduced in González-Rivera et al. (2011). The main advantage of our method is that the autocontours for a multivariate normal density are mathematically tractable regardless of the complexity of the dynamics of the model and the functional form of the assumed multivariate density. Once we obtain the quantile residuals of the model, through a second transformation of PITs to normality, we implement a battery of tests with standard asymptotic distributions and superior finite sample properties. In an out-of-sample context, the uncertainty created by parameter estimation depends on the size of the prediction sample relative to the estimation sample, which may be controlled easily by the researcher. Nonetheless we have shown that in all instances, whether parameter uncertainty is relevant or irrelevant, there are advantages to implementing a parametric bootstrap to correct mild size distortions in the tests. We provide Monte-Carlo evidence pertaining to finite-sample performance of our tests and compare them with those of Hong et al. (2007). We illustrate our approach by evaluating the bivariate density forecast of value and growth portfolio returns and concluded that a bivariate Student-t DCC density forecast is not fully satisfactory to model the events of 2008 and 2009. The rejection is not due to the dynamics provided by DCC, which seem to be adequate, but rather to the functional form of the bivariate density that seems to require even fatter tails in the prediction sample.

References

Andersen, T., T. Bollerslev, and F. Diebold (2010): "Parametric and Nonparametric Volatility Measurement," in *Handbook of Financial Econometrics*, ed. by Y. Aït-Sahalia and L. Hansen, Amsterdam: North-Holland, 67–138.

Bache, I., A. S. Jore, J. Mitchell, and S. Vahey (2010): "Combining VAR and DSGE Forecast Densities," *Journal of Economics Dynamics and Control*, forthcoming.

Bai, J. (2003): "Testing Parametric Conditional Distributions of Dynamic Models," *Review of Economics and Statistics*, 85(3), 532–549.

Bai, J. and Z. Chen (2008): "Testing Multivariate Distributions in GARCH Models,," *Journal of Econometrics*, 143, 19–36.

Berg, D. (2009): "Copula Goodness-of-fit Testing: An Overview and Power Comparison," *The European Journal of Finance*, 15(7-8), 675–701.

Berkowitz, J. (2001): "Testing Density Forecasts with Applications to Risk Management," *Journal of Business and Economic Statistics*, 19(4), 465–474.

Bollerslev, T., J. Litvinova, and G. Tauchen (2006): "Leverage and Volatility Feedback Effects in High-Frequency Data," *Journal of Financial Econometrics*, 4, 353–384.

Bontemps, C. and N. Meddahi (2005): "Testing Normality: A GMM Approach," *Journal of Econometrics*, 124, 149–186.

Capiello, L., R. Engle, and K. Shepard (2006): "Asymmetric Dynamics in the Correlations of Global Equity and Bond Returns," *Journal of Financial Econometrics*, 4, 537–572.

Chen, Y.-T. (2010): "Moment Tests for Density Forecast Evaluation in the Presence of Parameter Estimation Uncertainty," *Journal of Forecasting*, forthcoming.

Clements, M. and J. Smith (2000): "Evaluating the Forecast Densities of Linear and Nonlinear Models: Applications to Output Growth and Unemployment," *Journal of Forecasting*, 19, 255–276.

——— (2002): "Evaluating Multivariate Forecast Densities: A Comparison of Two Approaches," *International Journal of Forecasting*, 18, 397–407.

Corradi, V. and N. Swanson (2006a): "Bootstrap Conditional Distribution Tests In the Presence of Dynamic Misspecification," *Journal of Econometrics*, 133, 779–806.

——— (2006b): "Predictive Density and Conditional Confidence Interval Accuracy Tests," *Journal of Econometrics*, 135, 187–228.

——— (2006c): "Predictive Density Evaluation," in *Handbook of Forecasting*, ed. by G. Elliot, C. W. J. Granger, and A. Timmerman, Elsevier Science, 197–284.

Diebold, F. X., T. Gunther, and A. Tay (1998): "Evaluating Density Forecasts with Applications to Financial Risk Management," *International Economic Review*, 39, 863–883.

Diebold, F. X., J. Hahn, and A. S. Tay (1999): "Multivariate Density Forecast Evaluation and Calibration in Financial Risk Management: High-Frequency Returns on Foreign Exchange," *The Review of Economics and Statistics*, 81(4), 661–673.

Engle, R. (2002): "Dynamic Conditional Correlation-A simple Class of Multivariate GARCH Models," *Journal of Business and Economic Statistics*, 20, 339–350.

Fama, E. and K. French (1993): "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, 33, 3–56.

Glosten, L. R., R. Jaganathan, and D. E. Runkle (1993): "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks," *The Journal of Finance*, 48(5), 1779–1801.

González-Rivera, G., Z. Senyuz, and E. Yoldas (2011): "Autocontours: Dynamic Specification Testing," *Journal of Business and Economic Statistics*, 29, 186–200.

González-Rivera, G. and E. Yoldas (2010): "Multivariate Autocontours for Specification Testing in Multivariate GARCH Models," in *Volatility and Time Series Econometrics: Essays in Honor of Robert Engle*, ed. by T. Bollerslev, J. R. Russell, and M. W. Watson, Oxford University Press, 213–230.

Guidolin, M. and A. Timmermann (2008): "Size and Value Anomalies under Regime Shifts," *Journal of Financial Econometrics*, 6, 1–48.

Hansen, B. E. (1994): "Autoregressive Conditional Density Estimation," *International Economic Review*, 35, 705–730.

Hong, Y. and H. Li (2005): "Nonparametric Specification Testing for Continuous-Time Models with Applications to Term Structure of Interest Rates," *The Review of Financial Studies*, 18, 37–84.

Hong, Y., H. Li, and F. Zhao (2007): "Can the Random Walk Model be Beaten in out-of-sample Density Forecasts? Evidence from Intraday Foreign Exchange Rates," *Journal of Econometrics*, 141, 736–776.

Horowitz, J. L. (2001): "The Bootstrap in Econometrics," in *Handbook of Econometrics*, ed. by J. Heckman and E. Leamer, Elsevier Science.

J.P.Morgan (1996): *RiskMetrics-Technical Documents*, New York, 4 ed.

Kalliovirta, L. (2008): "Quantile Residuals for Multivariate Models," *Helsinki Center of Economic Research Discussion Papers*, No. 247.

Liesenfeld, R. and J.-F. Richard (2003): "Univariate and Multivariate Stochastic Volatility Models: Estimation and Diagnostics," *Journal of Empirical Finance*, 10, 505–531.

Patton, A. (2004): "On the Out-of-Sample Importance of Skewness and Asymmetric Dependence for Asset Allocation," *Journal of Financial Econometrics*, 2, 130–168.

Perez-Quiros, G. and A. Timmermann (2000): "Business Cycle Asymmetries in Stock Returns: Evidence from Higher Order Moments and Conditional Densities," *Journal of Econometrics*, 103, 259–306.

Pesaran, M. H. and B. Pesaran (2010): "Conditional Volatility and Correlations of Weekly Returns and the VaR Analysis of 2008 Stock Market Crash," *CESIFO WORKING PAPER NO. 3023*.

Pesaran, M. H., C. Schleicher, and P. Zaffaroni (2010): "Model Averaging in Risk Management with an Application to Futures Markets," *The Journal of Empirical Finance*, forthcoming.

Rosenblatt, M. (1952): "Remarks on a multivariate transformation," *The Annals of Mathematical Statistics*, 23(3), 470–472.

Tay, A. S. and K. F. Wallis (2000): "Density Forecasting: A Survey," *Journal of Forecasting*, 19, 235–254.

A Tables

Table 1: Size of t-statistics based on the Vector of Quantile Residuals

P	$t_{\mathbf{z}}^{0.01,1}$	$t_{\mathbf{z}}^{0.05,1}$	$t_{\mathbf{z}}^{0.1,1}$	$t_{\mathbf{z}}^{0.2,1}$	$t_{\mathbf{z}}^{0.3,1}$	$t_{\mathbf{z}}^{0.4,1}$	$t_{\mathbf{z}}^{0.5,1}$	$t_{\mathbf{z}}^{0.6,1}$	$t_{\mathbf{z}}^{0.7,1}$	$t_{\mathbf{z}}^{0.8,1}$	$t_{\mathbf{z}}^{0.9,1}$	$t_{\mathbf{z}}^{0.95,1}$	$t_{\mathbf{z}}^{0.99,1}$
Size 1													
250	4.1	4.3	5.5	5.2	5.9	5.2	4.6	5.9	6.2	5.7	5.4	5.1	4.0
500	4.8	5.2	5.3	4.5	5.2	6.2	4.9	5.7	5.2	5.7	5.5	5.6	5.0
1000	5.3	6.7	6.1	5.7	5.2	6.1	5.5	4.7	5.0	6.1	4.5	3.9	4.0
2000	4.4	4.8	5.2	4.1	4.7	4.4	5.0	5.5	5.1	4.5	4.7	3.5	4.3
Size 2													
250	4.4	5.0	5.1	5.7	5.5	5.9	5.7	6.2	5.2	5.2	5.1	5.2	3.1
500	4.6	4.5	4.8	6.0	5.3	4.3	4.3	4.7	4.5	5.9	5.2	6.4	5.3
1000	3.7	4.7	4.5	4.7	4.2	5.1	4.7	5.6	5.6	5.1	6.1	6.5	5.6
2000	4.1	5.6	5.7	5.6	4.3	5.3	4.9	5.3	5.7	4.7	5.4	4.0	4.8

Notes: This table reports simulated size of t-statistics based on the vector of quantile residuals. A parametric bootstrap is used to approximate the distribution of the test statistics. $T = 5000$, the number of Monte-Carlo replications is 1000, the number of bootstrap replications is 500, and the nominal size level is 5%. Details of the DGPs are given in the main text.

Table 2: Size of t-statistics based on the Aggregated Quantile Residual Process

P	$t_q^{0.01,1}$	$t_q^{0.05,1}$	$t_q^{0.1,1}$	$t_q^{0.2,1}$	$t_q^{0.3,1}$	$t_q^{0.4,1}$	$t_q^{0.5,1}$	$t_q^{0.6,1}$	$t_q^{0.7,1}$	$t_q^{0.8,1}$	$t_q^{0.9,1}$	$t_q^{0.95,1}$	$t_q^{0.99,1}$
Size 1													
250	3.9	3.9	4.5	4.1	5.6	6.3	5.5	5.5	5.3	6.0	5.4	4.7	2.6
500	4.2	4.3	5.5	3.9	4.4	5.6	5.0	5.3	6.4	5.0	3.5	3.3	4.1
1000	4.8	5.0	5.5	5.5	6.9	4.7	5.4	6.1	5.8	4.6	4.4	5.1	3.8
2000	5.3	4.6	5.4	5.5	6.3	5.7	5.9	6.2	6.9	4.9	5.0	5.9	3.7
Size 2													
250	4.8	5.0	4.4	5.5	5.2	5.7	4.7	4.3	5.0	4.5	5.1	3.7	3.2
500	3.8	4.6	4.8	5.1	5.2	5.1	5.5	5.6	5.5	5.9	5.7	3.6	4.9
1000	5.4	4.8	4.7	5.5	5.0	4.7	4.5	5.9	5.2	5.6	6.1	5.4	5.2
2000	5.7	4.3	5.2	4.9	4.5	4.6	4.8	4.7	5.0	5.3	5.1	5.3	4.1

Notes: This table reports simulated size of t-statistics based on the aggregated quantile residual process. A parametric bootstrap is used to approximate the distribution of the test statistics. $T = 5000$, the number of Monte-Carlo replications is 1000, the number of bootstrap replications is 500, and the nominal size level is 5%. Details of the DGPs are given in the main text.

Table 3: Size of chi-squared Statistics

P	$J_{\mathbf{z}}^{\boldsymbol{\alpha},1}$	$J_{\mathbf{z}}^{\boldsymbol{\alpha},2}$	$J_{\mathbf{z}}^{\boldsymbol{\alpha},3}$	$J_{\mathbf{z}}^{\boldsymbol{\alpha},4}$	$J_{\mathbf{z}}^{\boldsymbol{\alpha},5}$	$J_{q}^{\boldsymbol{\alpha},1}$	$J_{q}^{\boldsymbol{\alpha},2}$	$J_{q}^{\boldsymbol{\alpha},3}$	$J_{q}^{\boldsymbol{\alpha},4}$	$J_{q}^{\boldsymbol{\alpha},5}$
Size 1										
250	4.8	4.6	5.0	4.9	5.3	5.1	5.1	4.6	5.1	4.9
500	5.6	4.1	5.6	3.9	5.1	4.7	4.4	4.3	4.2	4.6
1000	5.6	4.9	5.7	4.9	5.4	4.8	4.0	4.6	4.5	3.8
2000	4.5	4.4	5.0	4.2	5.2	5.9	5.1	4.9	4.4	4.8
Size 2										
250	6.1	4.8	4.6	5.2	5.5	4.3	4.4	6.1	5.5	4.8
500	5.6	5.7	3.9	5.2	5.5	5.4	5.2	4.9	5.4	4.1
1000	5.0	5.1	5.0	6.1	4.3	5.4	5.6	5.5	4.6	5.0
2000	4.2	4.9	4.3	4.6	5.1	3.8	5.0	5.9	5.5	5.7

Notes: This table reports simulated size of chi-squared statistics based on a parametric bootstrap approximation to the finite sample distribution of the test statistics. $T = 5000$, the number of Monte-Carlo replications is 1000, the number of bootstrap replications is 500, the nominal size level is 5%, and $\boldsymbol{\alpha} = \{0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.99\}$. Details of the DGPs are given in the main text.

Table 4: Size of HLZ Statistics

P	$\hat{Q}(1)$	$\hat{Q}(2)$	$\hat{Q}(3)$	$\hat{Q}(4)$	$\hat{Q}(5)$	$\hat{W}(1)$	$\hat{W}(2)$	$\hat{W}(3)$	$\hat{W}(4)$	$\hat{W}(5)$
Size 1										
250	6.2	3.8	6.6	4.3	5.5	6.2	5.4	5.1	5.8	5.0
500	6.3	7.8	6.9	5.9	5.9	6.3	6.9	7.6	7.4	6.5
1000	5.8	5.7	6.5	4.7	5.1	5.8	6.2	7.4	6.4	6.6
2000	8.7	8.7	7.7	6.3	7.4	8.7	9.8	8.6	7.8	9.0
Size 2										
250	5.5	4.0	6.4	4.8	6.4	5.5	4.8	5.1	5.9	5.5
500	6.5	5.9	7.0	5.3	6.4	6.5	6.6	6.5	6.6	5.8
1000	6.0	5.8	8.2	6.2	6.3	6.0	6.3	7.7	7.5	8.1
2000	11.6	11.3	7.6	10.1	9.7	11.6	11.8	11.6	12.1	13.5

Notes: This table reports simulated size of the non-parametric kernel estimator based test statistics of HLZ applied to the aggregated quantile residual process, q_t. Finite sample critical values are calculated by Monte-Carlo simulation as described in HLZ Section 3.1. $T = 5000$, the number of Monte-Carlo replications is 1000, and the nominal size level is 5%. Details of the DGPs are given in the main text.

Table 5: Power of $t_{\mathbf{z}}$-statistics

P	$t_{\mathbf{z}}^{0.01,1}$	$t_{\mathbf{z}}^{0.05,1}$	$t_{\mathbf{z}}^{0.1,1}$	$t_{\mathbf{z}}^{0.2,1}$	$t_{\mathbf{z}}^{0.3,1}$	$t_{\mathbf{z}}^{0.4,1}$	$t_{\mathbf{z}}^{0.5,1}$	$t_{\mathbf{z}}^{0.6,1}$	$t_{\mathbf{z}}^{0.7,1}$	$t_{\mathbf{z}}^{0.8,1}$	$t_{\mathbf{z}}^{0.9,1}$	$t_{\mathbf{z}}^{0.95,1}$	$t_{\mathbf{z}}^{0.99,1}$
Power 1													
250	52.1	81.0	89.9	94.5	93.5	87.7	77.2	60.4	34.0	9.6	7.3	24.4	73.9
500	74.1	97.4	99.6	99.8	99.5	98.7	94.9	85.0	56.9	15.8	11.9	47.3	90.9
1000	91.4	99.9	100	100	100	100	99.7	97.5	78.3	24.5	18.9	75.7	99.5
Power 2													
250	21.4	31.8	42.1	51.7	48.6	44.2	35.6	25.6	13.6	7.0	10.5	25.7	58.6
500	29.8	55.2	70.2	79.9	77.5	69.3	55.6	39.9	19.7	6.7	16.9	43.2	75.4
1000	46.8	83.6	92.4	95.7	95.2	90.8	81.4	60.5	28.6	7.5	27.6	68.2	95.4
Power 3													
250	38.7	55.5	63.7	66.2	64.5	61.1	57.4	53.2	48.6	43.3	41.1	41.5	48.6
500	51.3	75.0	81.0	81.1	78.6	73.2	68.1	60.9	54.5	47.5	44.5	49.4	62.1
1000	69.3	89.9	92.7	92.9	91.1	87.2	81.0	70.7	58.2	46.0	43.1	59.3	82.3

Notes: This table reports simulated power of $t_{\mathbf{z}}$-statistics for the nominal size level of 5%. $T = 5000$, the number of Monte-Carlo replications is 1000, and the number of bootstrap replications is 500. Details of the DGPs are given in the main text.

Table 6: Power of t_q-statistics

P	$t_q^{0.01,1}$	$t_q^{0.05,1}$	$t_q^{0.1,1}$	$t_q^{0.2,1}$	$t_q^{0.3,1}$	$t_q^{0.4,1}$	$t_q^{0.5,1}$	$t_q^{0.6,1}$	$t_q^{0.7,1}$	$t_q^{0.8,1}$	$t_q^{0.9,1}$	$t_q^{0.95,1}$	$t_q^{0.99,1}$
Power 1													
250	14.7	24.8	33.6	47.9	53.9	53.5	52.4	44.0	29.2	14.5	4.6	8.4	43.8
500	17.9	40.2	59.0	77.4	81.6	82.9	78.7	70.1	49.9	23.3	5.7	16.5	58.3
1000	26.8	67.5	86.5	95.5	97.9	97.1	96.0	90.0	73.3	36.6	5.6	27.7	86.0
Power 2													
250	8.0	9.4	15.3	21.5	24.5	28.4	27.2	22.2	16.9	10.4	5.8	10.3	36.3
500	10.5	14.9	23.7	35.3	43.4	46.8	45.0	38.5	28.0	14.2	5.4	15.6	46.5
1000	11.9	28.7	42.9	61.0	68.3	73.4	71.3	61.5	45.3	18.8	5.4	27.0	74.4
Power 3													
250	13.9	22.6	32.1	43.4	47.4	49.6	49.7	49.2	45.9	42.7	37.0	33.3	33.1
500	17.7	34.1	46.3	55.8	60.5	60.4	58.5	57.1	50.7	44.1	38.7	38.4	46.1
1000	22.2	50.4	64.2	72.2	74.4	73.9	71.9	65.8	57.4	48.0	40.0	42.8	64.4

Notes: This table reports simulated power of t_q-statistics for the nominal size level of 5%. $T = 5000$, the number of Monte-Carlo replications is 1000, and the number of bootstrap replications is 500. Details of the DGPs are given in the main text.

Table 7: Power of chi-squared Statistics

P	$J_{\mathbf{z}}^{\boldsymbol{\alpha},1}$	$J_{\mathbf{z}}^{\boldsymbol{\alpha},2}$	$J_{\mathbf{z}}^{\boldsymbol{\alpha},3}$	$J_{\mathbf{z}}^{\boldsymbol{\alpha},4}$	$J_{\mathbf{z}}^{\boldsymbol{\alpha},5}$	$J_{q}^{\boldsymbol{\alpha},1}$	$J_{q}^{\boldsymbol{\alpha},2}$	$J_{q}^{\boldsymbol{\alpha},3}$	$J_{q}^{\boldsymbol{\alpha},4}$	$J_{q}^{\boldsymbol{\alpha},5}$
Power 1										
250	88.9	88.4	90.1	87.3	86.8	42.1	42.1	40.8	41.5	40.7
500	99.9	100	99.9	100	100	79.4	80.3	80.6	80.6	79.8
1000	100	100	100	100	100	99.1	99.1	99.2	98.9	99.2
Power 2										
250	47.5	25.1	24.1	24.1	24.3	20.6	11.0	11.4	10.0	12.0
500	86.6	50.4	49.4	50.3	51.2	49.3	17.0	19.0	19.8	22.4
1000	99.7	89.2	86.5	87.2	86.3	84.3	37.9	40.5	43.8	45.4
Power 3										
250	72.9	72.5	70.9	73.5	69.7	46.5	45.2	45.2	46.8	45.7
500	94.7	93.6	94.8	94.6	93.5	71.3	72.7	69.7	69.1	69.3
1000	99.7	100	99.7	99.8	99.9	93.1	92.4	91.9	91.5	92.0

Notes: This table reports simulated power of chi-squared statistics for the nominal size level of 5%. $T = 5000$, the number of Monte-Carlo replications is 1000, the number of bootstrap replications is 500, and $\boldsymbol{\alpha} = \{0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.99\}$. Details of the DGPs are given in the main text.

Table 8: Power of HLZ Statistics

P	$\hat{Q}(1)$	$\hat{Q}(2)$	$\hat{Q}(3)$	$\hat{Q}(4)$	$\hat{Q}(5)$	$\hat{W}(1)$	$\hat{W}(2)$	$\hat{W}(3)$	$\hat{W}(4)$	$\hat{W}(5)$
Power 1										
250	42.5	37.2	40.6	39.0	40.8	42.5	44.6	44.8	49.3	46.7
500	76.5	74.2	77.9	76.0	75.8	76.5	82.3	83.8	85.2	84.9
1000	97.9	96.7	98.3	96.2	96.7	97.9	98.1	98.8	98.8	99.1
Power 2										
250	97.4	63.0	67.8	60.3	62.5	97.4	94.2	91.3	90.6	87.4
500	100	96.4	97.2	95.7	95.8	100	100	100	100	100
1000	100	100	100	100	100	100	100	100	100	100
Power 3										
250	53.8	48.3	51.6	47.1	48.6	53.8	55.4	55.5	58.0	56.2
500	70.5	69.8	68.5	64.4	67.2	70.5	75.3	75.6	76.1	75.2
1000	90.7	87.6	86.4	85.2	83.2	90.7	91.5	92.8	92.6	93.3

Notes: This table reports simulated power of the non-parametric kernel estimator based test statistics of HLZ applied to the aggregated quantile residual process, q_t. Finite sample critical values are calculated by Monte-Carlo simulation as described in HLZ Section 3.1. $T = 5000$, the number of Monte-Carlo replications is 1000, and the nominal size level is 5%. Details of the DGPs are given in the main text.

Table 9: Descriptive Statistics

	Full sample		Estimation sample		Evaluation sample	
	Growth	Value	Growth	Value	Growth	Value
Mean	0.017	0.033	0.023	0.040	-0.005	0.001
Std Dev	1.205	1.140	1.089	0.786	1.584	2.007
Skewness	-0.164	-0.444	-0.235	-0.469	-0.032	-0.254
Kurtosis	9.168	18.057	6.590	7.055	9.476	8.761
Min	-8.888	-11.439	-7.430	-6.004	-8.888	-11.439
Max	10.757	9.177	5.919	4.283	10.757	9.177

Notes: The full sample runs from January 2, 1990 to October 30, 2009, with a total of 5001 observations. The evaluation sample is from November 10, 2005 to October 30, 2009 with a total of 1000 observations.

B Figures

Figure 1: Illustration of Autocontours

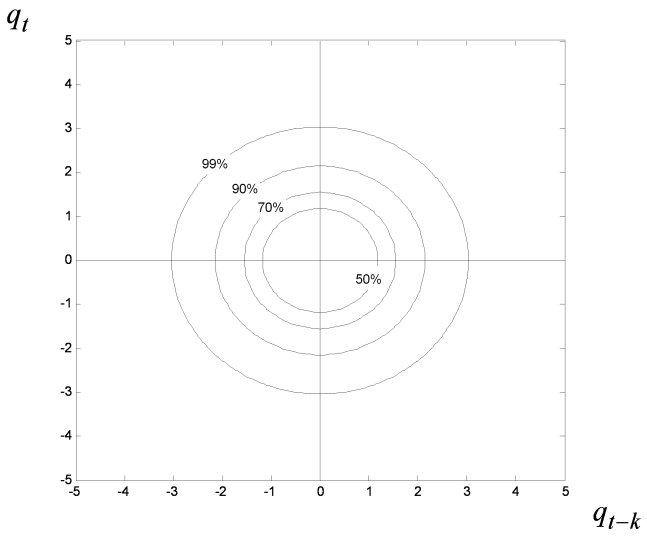
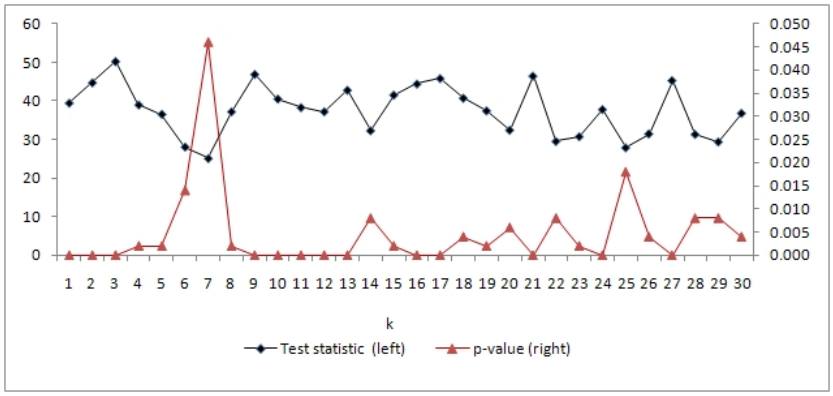


Figure 2: Chi-squared Statistics under Normal Distribution

Panel-a: $J_{\mathbf{z}}^{\boldsymbol{\alpha},k}$



Panel-b: $J_{q}^{\boldsymbol{\alpha},k}$

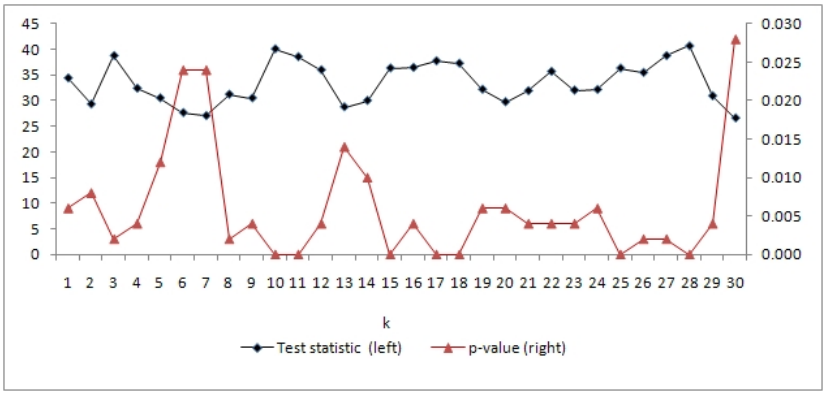
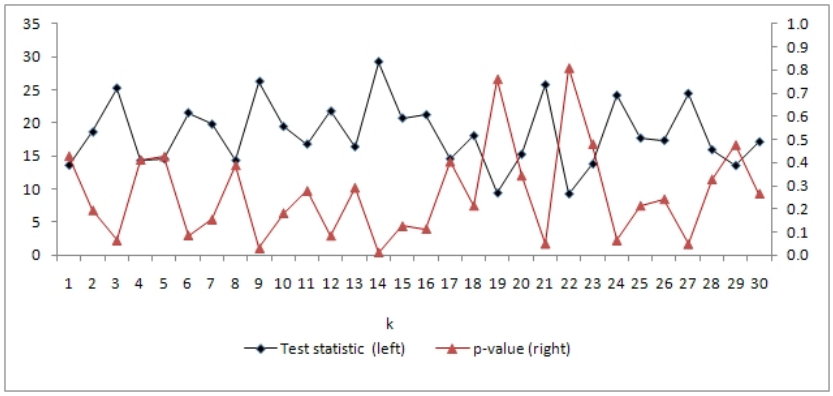


Figure 3: Chi-squared Statistics under Student-t Distribution

Panel-a: $J_{\mathbf{z}}^{\boldsymbol{\alpha},k}$



Panel-b: $J_{q}^{\boldsymbol{\alpha},k}$

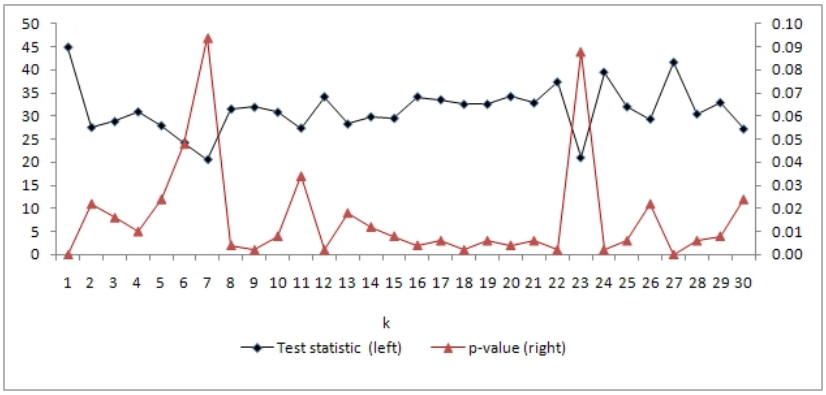
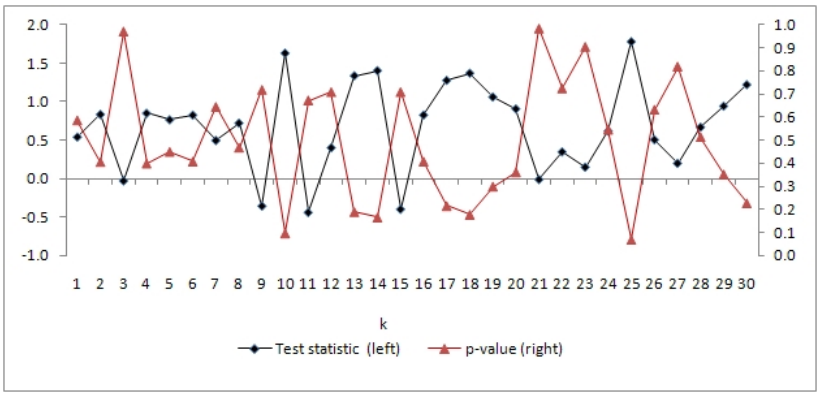


Figure 4: $t_q^{\alpha,k}$-Statistics under Normal Distribution

Panel-a: $\alpha = 0.1$



Panel-b: $\alpha = 0.95$

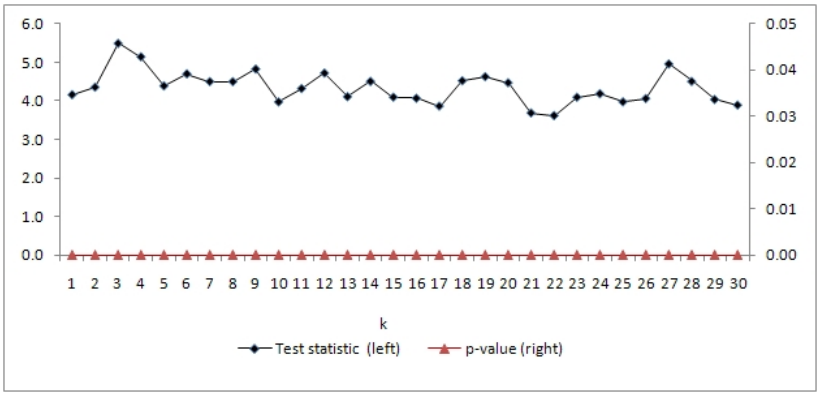
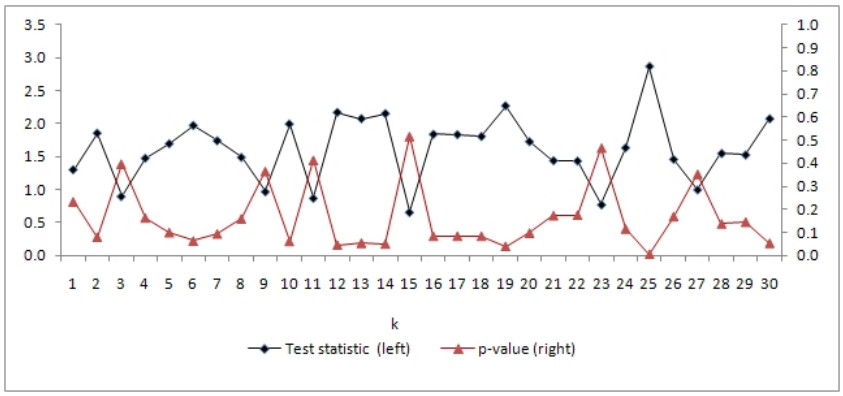


Figure 5: $t_q^{\alpha,k}$-Statistics under Student-t Distribution

Panel-a: $\alpha = 0.1$



Panel-b: $\alpha = 0.95$

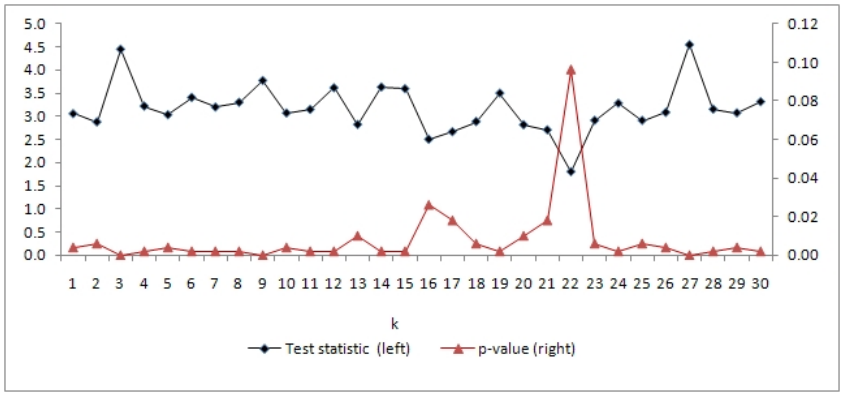
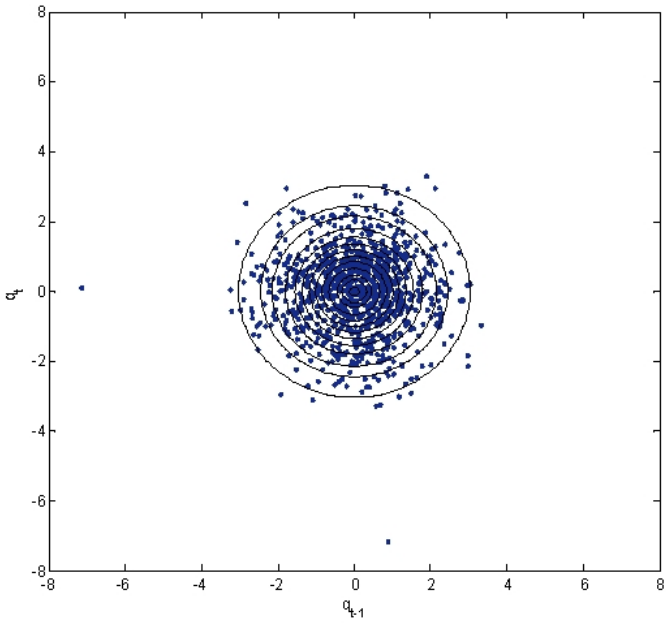


Figure 6: Data and Autocontours for the Aggregated Quantile Residual Process

Panel-a: Normal DCC



Panel-b: Student-t DCC

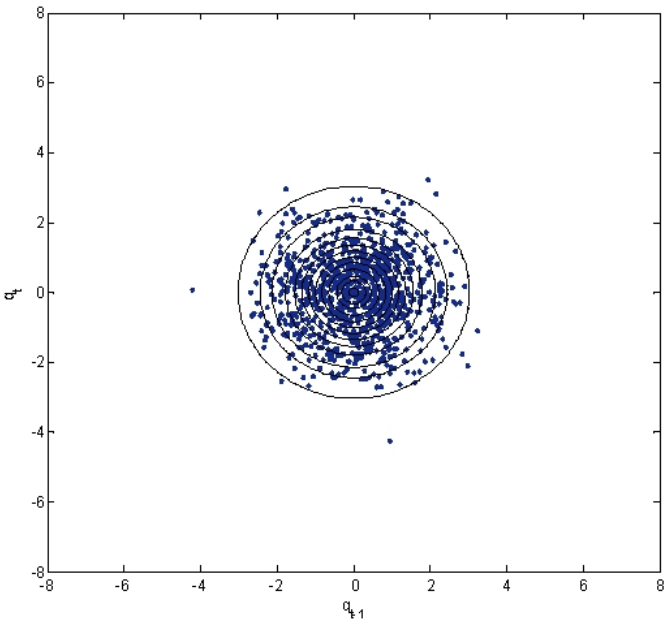
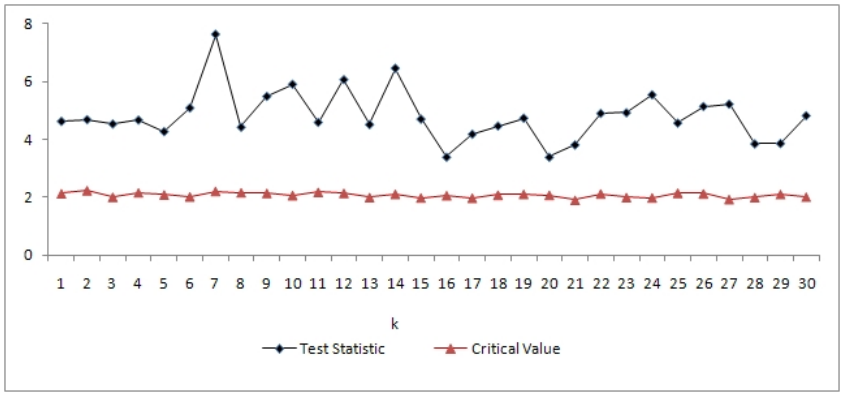


Figure 7: HLZ $\hat{Q}(k)$ Statistics

Panel-a: Normal



Panel-b: Student-t