Smooth Tests for Correct Specification of Conditional Predictive Densities

Juan Lin∗ Ximing Wu†

April 9, 2015

Abstract

We develop two specification tests of predictive densities based on that the generalized residuals of correctly specified predictive density models are i.i.d. uniform. The simultaneous test compares the joint density of generalized residuals with product of uniform densities; the sequential test examines the hypotheses of serial independence and uniformity sequentially based on the copula representation of a joint density. We propose data-driven smooth tests to construct the test statistics. We derive the asymptotic null distributions of the tests, which are nuisance parameter free, and establish their consistency. Monte Carlo simulations demonstrate excellent finite sample performance of the tests. We apply the proposed tests to evaluate some commonly used models of stock returns.

JEL Classification Codes: C12; C52; C53

Keywords: specification testing; density forecast; smooth test; sequential test; copula

∗Department of Finance, School of Economics & Wang Yanan Institute for Studies in Economics, Xiamen University, China. Email: linjuan.bnu@gmail.com.
†Corresponding author. Department of Agricultural Economics, Texas A&M University, College Station, TX 77843; Email: xwu@tamu.edu.
1 Introduction

Density forecast is of fundamental importance for decision making under uncertainty, for which good point estimates might not be adequate. Accurate density forecasts of key macroeconomic and financial variables, such as inflation, unemployment rate, stock returns and exchange rate, facilitate informed decision making of policy makers and financial managers, particularly when a forecaster’s loss function is asymmetric and the underlying process is non-Gaussian. Given the importance of density forecast, great caution should be exercised in judging the quality of density forecast models.

In a seminal paper, Diebold et al. (1998) introduced the method of dynamic Probability Integral Transform (PIT) to evaluate out-of-sample density forecasts. The transformed data are often called the generalized residuals of a forecast model. They showed that if a forecast model is correctly specified, the generalized residuals are \( i.i.d. \) uniformly distributed on \([0, 1]\). The serial independence signifies correct dynamic structure while uniformity characterizes correct specification of the stationary distribution.

For the purpose of evaluating density forecasts, Diebold et al. (1998) proposed some intuitive graphical methods to assess separately the serial independence and uniformity of the generalized residuals. Violation to either property indicates misspecification of predictive densities. Subsequently, a number of authors proposed methods to test formally the specification of predictive densities, including Berkowitz (2001), Bai (2003), Chen and Fan (2004), Hong and Li (2005), Corradi and Swanson (2006a), Hong et al. (2007), and Chen (2011), among others. For general overviews of the literature on the specification testing and evaluation of predictive densities, see Corradi and Swanson (2006b, 2012) and references therein.

Denote by \( \{Z_t\}_{t=1}^{N} \) the generalized residuals associated with certain density forecast model. Throughout the paper, we assume that \( \{Z_t\} \) is a stationary Markov process of order \( j \) with a marginal distribution \( G_0 \). The properties of \( \{Z_t\} \) can be captured by the joint distribution of \( Z_t \) and \( Z_{t-j} \), denoted by \( P_0 \). According to Sklar’s (1959) Theorem, there exists a copula function \( C_0 : [0, 1]^2 \rightarrow [0, 1] \) such that

\[
P_0(Z_t, Z_{t-j}) = C_0(G_0(Z_t), G_0(Z_{t-j})),
\]

where \( C_0 \) completely characterizes the dependence structure between \( Z_t \) and \( Z_{t-j} \). In this study, we propose two tests for the \( i.i.d. \) uniformity of the generalized residuals based on (1). The simultaneous test, based on the joint distribution, compares \( P_0 \) with the product of two uniform distributions. The sequential test, based on the copula representation of a joint distribution, examines sequentially whether \( C_0 \) is the independent copula and whether \( G_0 \) is
the uniform distribution. The first stage of the sequential procedure tests the hypothesis of independent copula and is robust against misspecification of marginal distributions. Rejection of the independence hypothesis effectively terminates the test. Otherwise, a subsequent uniformity test is conducted. Proceeding in this particular order ensures that the independence test is not affected by violation to uniformity and the uniformity test (if needed) is not affected by violation to independence. The asymptotic independence between the two stages facilitates proper control of the overall type I error of the sequential testing procedure.

We employ Neyman’s smooth test to construct the test statistics. Inspired by Ledwina (1994) and Kallenberg and Ledwina (1999), we propose data driven methods to select a suitable set of ‘directions’ and focus the tests in those directions. In addition to their ease of implementation, the proposed tests are marked by the following features: (i) The tests are omnibus as they adapt to the unknown underlying distributions and are consistent against essentially all alternatives. (ii) Adjustment is taken to account for the influence of nuisance parameters. Under the null hypothesis the limiting distributions of the tests are nuisance parameter free and can be easily tabulated. (iii) As is demonstrated by Monte Carlo simulations, the tests exhibit excellent finite sample size and power performance against a variety of alternatives and under different forecasting schemes. (iv) The two components of the sequential tests can be used as stand-alone tests for correct specification of dynamic structure and stationary distributions of forecast models. Their excellent finite sample performances are confirmed by Monte Carlo simulations as well.

The remainder of the paper proceeds as follows. In section 2, we briefly review the relevant literature on the dynamic probability integral transformation and Neyman’s smooth tests. We present in Section 3 the sequential tests for correct density forecasts via the copula approach and in Section 4 the simultaneous tests. In Section 5 we report Monte Carlo simulation results, followed by an application of the proposed tests to evaluate a variety of forecast models of stock returns. The last section concludes. Technical assumptions and proofs of theorems are relegated to the appendices.

2 Background

2.1 Dynamic Probability Integral Transformation

Given a time series \( \{Y_i\}_{i=1}^N \), we are interested in the one-step-ahead forecast of its conditional density \( f_0(\cdot|\Omega_{t-1}) \), where \( \Omega_{t-1} \) represents the information set available at time \( t-1 \). We split the sample of \( N \) observations into an in-sample period of size \( R \) for model estimation and an out-of-sample period of size \( n = N - R \) for forecast performance evaluation. Throughout the
paper, we assume that $R$ and $n$ both increase with $N$ and $\lim_{N \to \infty} n/R = \tau$, for some fixed number $0 \leq \tau < \infty$. Denote by $F_t(\cdot|\Omega_{t-1}, \theta)$ and $f_t(\cdot|\Omega_{t-1}, \theta)$ some conditional distribution and density functions of $Y_t$ given $\Omega_{t-1}$, where $\theta \in \Theta \subset \mathbb{R}^q$. The dynamic probability integral transform (PIT) of the data $\{Y_t\}_{t=R+1}^N$, with respect to the density forecast model $f_t(\cdot|\Omega_{t-1}, \theta)$, is defined as follows:

$$Z_t(\theta) = F_t(Y_t|\Omega_{t-1}, \theta) = \int_{-\infty}^{Y_t} f_t(v|\Omega_{t-1}, \theta) dv, \quad t = R + 1, \ldots, N. \quad (2)$$

The transformed data, $Z_t(\theta)$, are often called the generalized residuals of a forecast model.

Suppose that the forecast model is correctly specified in the sense that there exists some $\theta_0$ such that $f_t(\cdot|\Omega_{t-1}, \theta_0)$ coincides with the true but unknown conditional density function $f_0(\cdot|\Omega_{t-1})$. Under this condition, Diebold et al. (1998) showed that the generalized residuals $\{Z_t(\theta_0)\}$ are i.i.d. uniform on $[0, 1]$. Therefore, the test on a generic conditional density function $f_t(\cdot|\Omega_{t-1}, \theta)$ is equivalent to a test on the joint hypotheses that $\{Z_t(\theta)\}$ are i.i.d. uniform. Hereafter we shall write $Z_t(\theta)$ as $Z_t$ for simplicity whenever there is no ambiguity.

Diebold et al. (1998) used some intuitive graphical methods to separately examine the serial independence and uniformity of the generalized residuals. Subsequently, many authors have adopted the approach of PIT and proposed formal specification testing of predictive densities. Berkowitz (2001) further transformed the generalized residuals to $\Phi^{-1}(Z_t)$, where $\Phi^{-1}(\cdot)$ is the inverse of the standard normal distribution function, and proposed tests of serial independence under the assumption of linear autoregressive dependence structure. Bai (2003) and Corradi and Swanson (2006a) proposed Kolmogorov type tests. Hong and Li (2005) and Hong et al. (2007) constructed nonparametric tests by comparing kernel estimate of the joint density of $(Z_t, Z_{t-j})$ with the product of two uniform densities. Chen and Fan (2004) suggested copula based tests of serial independence against alternative parametric copulas. Park and Zhang (2010) and Chen (2011) considered moment based tests. Corradi and Swanson (2006b, 2012) provided excellent overviews of the fast-growing literature on specification testing and evaluation of predictive densities.

2.2 Neyman’s Smooth Test

Omnibus tests are desirable in goodness-of-fit tests, wherein alternative hypotheses are often vague. Some classic omnibus tests, such as the Kolmogorov-Smirnov test or Cramér-von Mises test, are known to be consistent but can only detect a few deviations from a null hypothesis for moderate sample sizes. In fact, these omnibus tests behave very much like a parametric test for a certain alternative that often corresponds to smooth departures from
the null distribution. In this study we adopt Neyman’s smooth test, which enjoys attractive theoretical and finite sample properties and can be tailored to adapt to unknown underlying distributions. For a general review of smooth tests, see Rayner and Best (1990). Here we briefly review the smooth test and some extensions. For simplicity, suppose for now that \( \{Z_i\}_{i=1}^n \) are i.i.d. random samples from a distribution \( G_0 \) defined on the unit interval. To test the uniformity hypothesis, Neyman (1937) considered a family of smooth alternative distributions given by
\[
g(z) = \exp \left( \sum_{i=1}^k b_i \psi_i(z) + b_0 \right), \quad z \in [0, 1], \tag{3}
\]
where \( b_0 \) is a normalization constant such that \( g \) integrates to unity and \( \psi_i \)'s are shifted Legendre polynomials given by
\[
\psi_i(z) = \frac{\sqrt{2i + 1}}{i!} \frac{d^i}{dz^i} (z^2 - z)^i, \quad i = 1, \ldots, k, \tag{4}
\]
which are orthonormal with respect to the standard uniform distribution.

Under the assumption that \( G_0 \) is a member of (3), testing uniformity amounts to testing the hypothesis \( B \equiv (b_1, \ldots, b_k)' = 0 \), to which the likelihood ratio test can be readily applied. Alternatively, one can construct a score test, which is asymptotically locally optimal and also computationally easy. Define \( \hat{\psi}_i = n^{-1} \sum_{t=1}^n \psi_i(Z_t) \) and \( \hat{\psi}_{(k)} = (\hat{\psi}_1, \ldots, \hat{\psi}_k)' \). Neyman’s smooth test for uniformity is constructed as
\[
N_k = n\hat{\psi}_{(k)}' \hat{\psi}_{(k)}. \tag{5}
\]
Under uniformity, \( N_k \) converges in distribution to the \( \chi^2 \) distribution with \( k \) degrees of freedom as \( n \to \infty \).

One limitation of the smooth test is the lack of clear guidance on how to select the number of terms \( k \) in (3), which may affect its power. A customary practice is to use a small \( k \), usually less than 4. Some studies that consider data driven methods in the selection of \( k \) have emerged in the past two decades. These methods employ a two step procedure: A certain model selection rule is applied to choose a suitable distribution within the family of (3), followed by the smooth test based on the chosen distribution. This approach was pioneered by Ledwina (1994), who applied the Bayesian Information Criterion (BIC) to construct a data driven Neyman’s test. Under uniformity, the number of terms converges in probability to one as \( n \to \infty \); consequently the asymptotic distribution of the test statistic
(5) is approximated by the $\chi^2$ distribution with one degree of freedom. She showed that, via simulations, the selected number of terms converges to one quickly even under rather small sample sizes. On the other hand, the $\chi^2$ approximation to the finite sample distribution of the test statistic is not adequate, due to its data driven nature. Ledwina (1994) suggested a simple simulation method to calculate the critical values of her data driven smooth tests. Extensive numerical experiments by Ledwina (1994) and Kallenberg and Ledwina (1995) demonstrated outstanding performance of this test. Kallenberg and Ledwina (1995) established the consistency of the data driven smooth test against essentially all fixed alternatives and Claeskens and Hjort (2004) studied this test under local alternatives. For the applications of smooth tests in econometrics, see for example Bera and Ghosh (2002), Bera et al. (2013), Lin and Wu (2015) and references therein.

Subsequent studies have extended Ledwina’s test in several directions. Inglot et al. (1997) and Kallenberg and Ledwina (1997) considered composite hypotheses. Kallenberg (2002) and Inglot and Ledwina (2006) suggested that many alternative criteria can be employed to construct consistent data driven smooth tests. For instance, they showed that both the AIC and BIC penalties lead to consistent data-driven smooth tests. Kallenberg and Ledwina (1999) and Kallenberg (2009) proposed smooth tests for bivariate distributions. Let $\psi_{i_1i_2}(v_1, v_2) = \psi_{i_1}(v_1)\psi_{i_2}(v_2)$ and $\Psi$ be a non-empty subset of $\{\psi_{i_1i_2} : 1 \leq i_1, i_2 \leq M\}$. Denote the cardinality of $\Psi$ by $k = |\Psi|$ and write $\Psi = \{\Psi_1, \ldots, \Psi_k\}$. They considered a bivariate version of (3) for the joint distribution of the empirical distributions of two random variables,

$$g(z_1, z_2) = \exp \left( \sum_{i=1}^{k} b_i \Psi_i(z_1, z_2) + b_0 \right), \quad (z_1, z_2) \in [0, 1]^2,$$

where $b_0$ is a normalization constant. The test of independence within the exponential family (6) is equivalent to a test of the hypothesis: $B \equiv (b_1, \ldots, b_k)' = 0$. A smooth test of independence can then be constructed analogously to the test of univariate uniformity.

Intuitively, the data driven smooth test selects a set of suitable ‘directions’ $\Psi$ and focuses the test in those directions. It adapts to the underlying unknown distribution via data driven selection of $\Psi$. Thus the specification of $\Psi$ is crucial to the consistency of the test. Ledwina (1994) used a truncation rule to select the number of terms in her test. This simple rule, albeit natural for univariate distributions, can be cumbersome in a multivariate framework: It tends to select a larger-than-desired set of terms due to the curse of dimensionality. In their study of bivariate distributions, Kallenberg and Ledwina (1999) suggested two specifications. The first method considers only ‘diagonal’ entries such as $\psi_{11}, \psi_{22}$ and so on, to which the BIC truncation rule is applied. The second method is more flexible...
and allows ‘non-diagonal’ entries. In particular, let \( \{Z_{1,t}, Z_{2,t}\}_{t=1}^n \) be an i.i.d. sample and \( \hat{\Psi}_i = n^{-1} \sum_{t=1}^n \Psi_i(Z_{1,t}, Z_{2,t}), i = 1, \ldots, K \). The set \( \Psi \) is arranged such that \( \Psi_1 = \psi_{11} \) and the remaining entries are arranged in the descending order according to the absolute value of \( \hat{\Psi}_i \). The BIC truncation rule is then applied to this ordered set. In a similar spirit, Kallenberg (2009) used a threshold rule to select significant elements of \( \hat{\Psi} \) for copula specifications.

3 Sequential Tests of Correct Density Forecasts

In this section we develop a sequential procedure for evaluating density forecasts, taking advantage of the copula representation of a joint distribution. A copula is a multivariate probability distribution with uniform marginals. Copulas provide a natural way to separately model the marginal behavior of \( \{Z_t\} \) and its serial dependence structure. This separation permits us to test the serial independence and uniformity of \( \{Z_t\} \) sequentially. Rejection of serial independence effectively terminates the procedure; otherwise, a subsequent test on uniformity is conducted. Below we shall explain why we proceed in this particular order and how to obtain the desired overall type I error of the sequential test. To ease composition, we start with the test on uniformity of the marginal distribution.

3.1 Test of Uniformity

Given a density forecast model \( f_t(v|\Omega_{t-1}, \theta) \), define the generalized residuals

\[
\hat{Z}_t \equiv Z_t(\hat{\theta}_t) = \int_{-\infty}^{Y_t} f_t(v|\Omega_{t-1}, \hat{\theta}_t)dv, \quad t = R + 1, \ldots, N,
\]

where \( \hat{\theta}_t \) is the maximum likelihood estimate (MLE) of \( \theta \) based on the data up to time \( t - 1 \). We allow for three different estimation schemes for density forecast. The fixed scheme estimates the parameters of the density function \( \theta \) once using observations from time 1 to \( R \). The recursive scheme uses all past observations from time 1 to \( t - 1 \) to estimate \( \theta \). The rolling scheme fixes a rolling sample of fixed size \( R \) and drops distant observations as recent ones enter the estimation, using observations from time \( t - R \) to \( t - 1 \).

Testing correct stationary distribution is equivalent to testing the hypothesis that \( \hat{Z}_t \) is uniformly distributed. Let \( \Psi_U \) be a non-empty subset of basis functions \( \{\psi_1, \ldots, \psi_M\} \) for some integer \( M > 1 \) and denote its cardinality by \( K_U \equiv |\Psi_U| \). Write \( \Psi_U = (\Psi_{U,1}, \ldots, \Psi_{U,K_U})' \).
We consider a smooth alternative distribution given by
\[
g(z) = \exp \left( \sum_{i=1}^{K_U} b_{U,i}(z) + b_{U,0} \right), \quad z \in [0,1],
\] (8)
where \(b_{U,0}\) is a normalization constant. Let \(B_U = (b_{U,1}, \ldots, b_{U,K_U})'\). Under the assumption that \(\hat{Z}_t\) is distributed according to (8), testing for uniformity is equivalent to testing the following hypothesis:
\[
H_{0U}: B_U = 0.
\]
Correspondingly, one can construct a smooth test based on the sample moments \(\hat{\Psi}_U = (\hat{\Psi}_{U,1}, \ldots, \hat{\Psi}_{U,K_U})'\), where \(\hat{\Psi}_{U,i} = n^{-1} \sum_{t=R+1}^{N} \Psi_{U,i}(\hat{Z}_t)\) for \(i = 1, \ldots, K_U\).

Compared with test (5) derived under a simple hypothesis, the present test is complicated by the presence of nuisance parameters \(\hat{\theta}_t\) and depends on the estimation scheme used in the density prediction. Proper adjustments are required to account for their influences. Let \(s_t = \frac{\partial}{\partial \theta} \ln f_t(Y_t|\Omega_{t-1}, \theta)\) be the gradient function of the predictive density. Define
\[
\begin{align*}
s_{0t} &= s_t|_{\theta=\theta_0}, \quad Z_{0t} = Z_t(\theta_0), \\
A &= E[s_{0t}s_{0t}'], \quad D = E[\Psi_U(Z_{0t})s_{0t}'].
\end{align*}
\]
Also define
\[
\eta = \begin{cases} 
\tau, & \text{fixed}, \\
0, & \text{recursive}, \\
-\frac{\tau^2}{3}, & \text{rolling (}\tau \leq 1), \\
-1 + \frac{2}{3\tau}, & \text{rolling (}\tau > 1).
\end{cases}
\]

Applying the results of West and McCracken (1998) to our test statistic and following the arguments of Chen (2011), we establish the following asymptotic properties of \(\hat{\Psi}_U:\)

**Theorem 1.** Suppose that conditions C1-C3 given in the Appendix hold. Under \(H_{0U}\), \(\hat{\Psi}_U \xrightarrow{p} 0\) and \(\sqrt{n}\hat{\Psi}_U \xrightarrow{d} N(0, V_U)\) as \(n \to \infty\), where \(V_U = I_U + \eta D A^{-1} D'\) with \(I_U\) being a \(K_U\)-dimensional identity matrix.

**Remark 1.** When the parameter \(\theta\) in the forecast density model \(f_t(\cdot|\Omega_{t-1}, \theta)\) is known, \(V_U\) is reduced to \(I_U\), the variance obtained under the simple hypothesis. The adjustment term \(D A^{-1} D'\) for nuisance parameters is multiplied by a factor \(\eta\) to account for the estimation scheme of the density forecast model. If instead the entire sample is used in the estimation (as in full sample estimation rather than prediction), \(\eta = 1\).
Remark 2. Bai (2003) used Khmaladze’s martingale transformation to construct a Kolmogorov type test that is nuisance parameter free asymptotically. Hong et al. (2007) constructed specification tests based on kernel estimates of the densities of generalized residuals. Because the nuisance parameters converge at rate $n^{-1/2}$ while the test statistics converge at nonparametric rates, the effects of nuisance parameter estimation is asymptotically negligible. The convenience of not having to directly account for parameter estimation error is gained at the prices of bandwidth selection for kernel densities and slower convergence rates. Corradi and Swanson (2006a) proposed an alternative Kolmogorov type test that converges at a parametric rate and does not require bandwidth selection. Its disadvantage is that the limiting distribution is not nuisance parameter free and one needs to rely on bootstrap techniques to obtain valid critical values.

Remark 3. West and McCracken (1998) required the moment functions $\psi_i$’s to be continuously differentiable. McCracken (2000) extended the results of West and McCracken (1998) to allow non-differentiable moment functions. However, $E[\psi_i(Z_t(\theta))]$ is still required to be continuously differentiable with respect to $\theta$. We note that the basis functions considered in this study, such as the Legendre polynomials and cosine series, satisfy regularity conditions given in West and McCracken (1998). Furthermore, condition C2 in Appendix A ensures the differentiability of $E[\psi_i(Z_t(\theta))]$ with respect to $\theta$.

Next define

$$\hat{\eta} = \eta|_{\tau=n/R}, \quad \hat{s}_t = \frac{\partial}{\partial \theta} \ln f_t(Y_t|\Omega_{t-1}, \theta)|_{\theta=\hat{\theta}},$$

$$\hat{A} = n^{-1} \sum_{t=R+1}^{N} \hat{s}_t^t \hat{s}^{t'}_t, \quad \hat{D} = n^{-1} \sum_{t=R+1}^{N} \Psi_U(\hat{Z}_t) \hat{s}^{t'}_t. \tag{10}$$

We can estimate $V_U$ consistently using its sample counterpart:

$$\hat{V}_U = I_U + \hat{\eta} \hat{D} \hat{A}^{-1} \hat{D}'. \tag{11}$$

We then construct a smooth test of uniformity as follows.

Theorem 2. Suppose that the conditions of Theorem 1 hold. Let the test of stationary distribution of model (2) be given by

$$\hat{N}_U = n \hat{\Psi}_U' \hat{V}_U^{-1} \hat{\Psi}_U. \tag{12}$$

Under the null hypothesis of correct stationary distribution, $\hat{N}_U \xrightarrow{d} \chi^2_{K_U}$ as $n \to \infty$. 

8
A critical component of data driven smooth tests is the selection of suitable basis functions from \( \{ \psi_1, \ldots, \psi_M \} \), i.e., the configuration of \( \Psi_U \) in the present test. In the spirit of Kallenberg and Ledwina (1999), we rearrange the candidate set \( \{ \psi_1, \ldots, \psi_M \} \) to \( \Psi_u = \{ \psi_{u,1}, \ldots, \psi_{u,M} \} \) such that \( \psi_{u,1} = \psi_1 \) and the rest of the set corresponds to the elements of \( \{ \psi_2, \ldots, \psi_M \} \) arranged in the descending order according to \( |\hat{\psi}_i| \equiv |n^{-1} \sum_{t=R+1}^N \psi_i(\hat{Z}_t)| \) for \( i = 2, \ldots, M \). A truncation rule according to some information criterion is then applied to the empirically ordered candidate set \( \Psi_u \) to select \( \Psi_U \), based on which the test statistic is constructed.

Intuitively large value of \( |\hat{\psi}_i| \) indicates probable deviation from the hypothesized uniform distribution in the ‘direction’ characterized by \( \psi_i \). Thus the data driven selection of \( \Psi_U \) gives priorities to those potentially important directions in the testing procedure. On the other hand, we set \( \psi_{u,1} = \psi_1 \) for the following reasons: (i) low frequency deviations are generally more likely, especially when the unknown underlying distributions are smooth; \( \psi_1 \), corresponding to a shift in the mean, appears to be a natural point of departure; (ii) low order moments are more robust to possible outliers; (iii) under uniformity, the number of selected basis functions converges to one as \( n \to \infty \). Fixing \( \psi_{u,1} = \psi_1 \) ensures that under the null the data driven \( \Psi_U \) converges in probability to one fixed element \( \psi_1 \), rather than a random element of \( \{ \psi_1, \ldots, \psi_M \} \) corresponding to the maximum of their sample analogs. Subsequently, the asymptotic distribution of \( \hat{N}_U \) is approximately \( \chi^2_1 \), which simplifies the theoretical analysis.

Given the ordered candidate set \( \Psi_u \), we proceed to use an information criterion to select \( \Psi_U \). Denote the subset of \( \Psi_u \) with its first \( k \) elements by \( \Psi_{u,(k)} = \{ \psi_{u,1}, \ldots, \psi_{u,k} \} \), \( k = 1, \ldots, M \) and the corresponding \( V_{u,(k)} \) and \( N_{u,(k)} \), as given in (11) and (12), are similarly defined; their sample analogs are denoted by \( \hat{V}_{u,(k)} \), \( \hat{N}_{u,(k)} \) and \( \hat{\Psi}_{u,(k)} \) respectively. For each \( k \), let \( \hat{\Psi}_{u,(k)}^* = \nu_{u,(k)}' \hat{\Psi}_{u,(k)} \), where \( \nu_{u,(k)} \) is given by \( \hat{V}_{u,(k)}^{-1} \nu_{u,(k)} = \nu_{u,(k)}' \hat{V}_{u,(k)}^{-1} \nu_{u,(k)}' \). Following Inglot and Ledwina (2006), we use the following criterion to select a suitable \( \Psi_U \), whose cardinality is denoted by \( K_U \):

\[
K_U = \min \{ k : \hat{N}_{u,(k)} - \Gamma(k, n) \geq \hat{N}_{u,(s)} - \Gamma(s, n), 1 \leq k, s \leq M \}.
\]

The complexity penalty \( \Gamma(s, n) \) is defined as

\[
\Gamma(s, n) = \begin{cases} 
  s \log n & \text{if } \max_{1 \leq k \leq M} |\sqrt{n} \hat{\Psi}_{u,k}^*| \leq 2.4 \log n, \\
  2s & \text{if } \max_{1 \leq k \leq M} |\sqrt{n} \hat{\Psi}_{u,k}^*| > 2.4 \log n,
\end{cases}
\]

where \( \hat{\Psi}_{u,k}^* \) is the \( k \)th element of \( \hat{\Psi}_{u,(M)}^* \) for \( k = 1, \ldots, M \). Note that the penalty factor
is ‘adaptive’ in the sense that either the AIC or BIC is adopted in a data driven manner, depending on the empirical evidence pertinent to the magnitude of deviation from uniformity. In the spirit of the general score information criterion proposed by Aerts et al. (2000), we use in the selection procedure \( \hat{N}_{u,(k)} \) as given by (12), rather than its unweighted counterpart \( n \hat{\Psi}'_{u,(k)} \hat{\Psi}_{u,(k)} \) as in Inglot and Ledwina (2006). Our numerical experiments suggest that the criterion based on \( \hat{N}_{u,(k)} \) provides better size performance.

Next we present the asymptotic properties of the proposed test \( \hat{N}_U \) based on a set of basis functions \( \Psi_U \) selected according to the procedure described above. The first part of the theorem below provides the asymptotic distribution of the test statistic under the null hypothesis and the second part establishes the consistency of the test.

**Theorem 3.** Let \( K_U \) be selected according to (13). Suppose that conditions C1-C8 given in the Appendix hold. (a) Suppose that \( Z_t(\theta) \) given by (2) follows a uniform distribution. Then \( \Pr(K_U = 1) \xrightarrow{p} 1 \) and \( \hat{N}_U \xrightarrow{d} \chi^2_1 \) as \( n \to \infty \). (b) Suppose instead that \( Z_t(\theta) \) is distributed according to a distribution \( P \) such that \( E_P[|\hat{\psi}_i(Z_t(\theta))|] \neq 0 \) for some \( i \) in \( 1, \ldots, M \). Let \( \theta^*_0 \) be associated with \( P \) and \( \hat{\theta} \) such that \( ||\hat{\theta} - \theta^*_0|| \xrightarrow{P} 0 \) as \( n \to \infty \) under \( P \). Assume that conditions C2 and C3 are satisfied with \( \theta_0 \) replaced by \( \theta^*_0 \). Then \( \hat{N}_U \xrightarrow{d} \infty \) as \( n \to \infty \).

**Remark 4.** It is necessary that \( M \) increases with sample size such that the smooth test is consistent. Kallenberg and Ledwina (1995) showed that the alternative considered in Theorem 3(b) is general enough to encompass essentially all deviations from uniformity. Letting \( M \) increase with sample size ensures that deviations from uniformity will be detected asymptotically with probability one by the proposed data driven configuration of \( \Psi_U \). In practice, a moderately large \( M \) suffices; extensive simulations by Ledwina (1994) and Kallenberg and Ledwina (1995) indicate that \( M = 10 \) gives satisfactory results.

Since the proposed test is constructed in a data driven fashion, the number of functions selected is random even under the null hypothesis. Consequently, the \( \chi^2_1 \) distribution does not provide an adequate approximation to the distribution of the test statistic for moderate sample sizes. Nonetheless, Theorem 3 establishes that \( \hat{N}_U \) is asymptotically distribution free under uniformity, implying that its distribution can be approximated via simple simulations based on \( i.i.d. \) uniform samples. Below we present a description of the simulation procedure.

- **Step U1:** Generate an \( i.i.d. \) random sample \( \{Z_t\}_{t=1}^n \) from the standard uniform distribution.
- **Step U2:** Construct the candidate set such that \( \Psi_{u,1} = \psi_1 \) and \( \{\Psi_{u,i}\}_{i=2}^M \) correspond to \( \{\hat{\psi}_i\}_{i=2}^M \) arranged in the descending order according to \( |\hat{\psi}_i| \).
• Step U3: Choose a set of basis functions $\Psi_U$ according to the selection rule given in (13); Compute the corresponding test statistic $\hat{N}_U$.

• Step U4: Repeat steps U1 to U3 $L$ times; denote the results by $\{\hat{N}_U^{(l)}\}_{l=1}^L$.

• Step U5: Use the $q$th percentile of $\{\hat{N}_U^{(l)}\}_{l=1}^L$ to approximate the $q$th percentile critical value of $\hat{N}_U$.

Our numerical results reported in Section 5 suggest that the simulated critical values provide good size performance for sample size as small as $n = 250$.

### 3.2 Robust Test of Serial Independence

The copula function of random variables completely captures their dependence structure. Thus testing for serial independence between $Z_t$ and $Z_{t-j}$ can be based on their copula distribution. In particular, testing their independence is equivalent to testing the hypothesis that their copula density is constant at unity.

The copula distribution, $C_0(G_0(Z_t), G_0(Z_{t-j}))$, takes as input the marginal distribution $G_0$ of $Z_t$ and $Z_{t-j}$. Given $\{\hat{Z}_t\}$, we can estimate $G_0$ using parametric or nonparametric methods. Lacking a priori information on the marginal distributions, we choose to estimate $G_0$ nonparametrically using their rescaled empirical distribution:

$$G_n(v) = \frac{1}{n+1} \sum_{t=R+1}^N I(\hat{Z}_t \leq v),$$

where $I(\cdot)$ is the indicator function. Unlike parametric estimates, the empirical distribution is free of possible misspecification errors. More importantly, Chen and Fan (2006) established that subsequent copula tests based on the empirical distributions of the marginals are asymptotically invariant to the estimation of $\theta$ in the first stage. Consequently, under the null hypothesis of independence, no adjustments for the influence of the nuisance parameter $\hat{\theta}$ are required in the construction of smooth tests based on $G_n(\hat{Z}_t)$’s. In addition to this simplified test construction, Chen and Fan (2004) showed that one key advantage of this approach is that a so-constructed test on independent copula is robust to misspecification of the marginal distribution of $Z_t$.

Let $\Psi_C = \{\Psi_{C,1}, \ldots, \Psi_{C,K_C}\}$ be a non-empty subset of $\{\psi_{i_1i_2} : 1 \leq i_1, i_2 \leq M\}$, where $\psi_{i_1i_2}(u_1, u_2) = \psi_{i_1}(u_1)\psi_{i_2}(u_2)$ and $K_C = |\Psi_C|$. Define $U_t = G_0(Z_t), t = R+1, \ldots, N$, which is distributed according to the standard uniform distribution. In the spirit of Neyman’s
smooth test for uniformity, we consider the following alternative bivariate density

\[ c(u_1, u_2) = \exp \left\{ \sum_{i=1}^{K_C} b_{C,i} \Psi_{C,i}(u_1, u_2) + b_{C,0} \right\}, \quad (u_1, u_2) \in [0, 1]^2, \]  

(14)

where \( b_{C,0} \) is a normalization constant. Let \( B_C = (b_{C,1}, \ldots, b_{C,K_C})' \). Under the assumption that the underlying copula density \( c_0 \) is a member of (14), testing for independence is equivalent to testing the following hypothesis:

\[ H_{0C} : B_C = 0. \]  

(15)

Define \( \hat{U}_t = G_n(\hat{Z}_t) \) and \( \hat{\Psi}_{C,i}(j) = (n-j)^{-1} \sum_{t=R+j+1}^N \Psi_{C,i}(\hat{U}_t, \hat{U}_{t-j}) \) for \( i = 1, \ldots, K_C \). A test on the hypothesis (15) is readily constructed as

\[ \hat{Q}_C(j) = (n-j) \hat{\Psi}_C(j) \hat{\Psi}_C(j), \]  

(16)

where \( \hat{\Psi}_C(j) = (\hat{\Psi}_{C,1}(j), \ldots, \hat{\Psi}_{C,K_C}(j))' \). This test is particularly appealing as it is free of nuisance parameters.

**Theorem 4.** Suppose that conditions C1-C3 given in the Appendix hold. Under the null hypothesis of independence, \( \hat{Q}_C(j) \overset{d}{\to} \chi^2_{K_C} \) as \( n \to \infty \).

The copula test of independence (16) depends crucially on the configuration of the basis functions included in \( \Psi_C \) to capture potential deviations from independence.\(^1\) Similar to the construction of uniformity test, we start with a candidate set \( \Psi_c \) such that \( \Psi_{c,1} = \psi_{11} \) and the rest of its elements correspond to \( \psi_{i_1 i_2} \) ordered in the descending order according to \( |\hat{\psi}_{i_1 i_2}| = |1/(n-j) \sum_{t=R+j+1}^N \psi_{i_1 i_2}(\hat{U}_t, \hat{U}_{t-j})| \).\(^2\) Denote the cardinality of \( \Psi_c \) by \( |\Psi_c| \) and let \( \Psi_{c,(k)} = \{\Psi_{c,1}, \ldots, \Psi_{c,k}\}, k = 1, \ldots, |\Psi_c| \). Further define \( \hat{\Psi}_{c,(k)}(j) = (\hat{\Psi}_{c,1}(j), \ldots, \hat{\Psi}_{c,k}(j))' \) and \( \hat{Q}_{c,(k)}(j) = (n-j) \hat{\psi}_{c,(k)}(j) \hat{\psi}_{c,(k)}(j) \). A truncation rule is then used to select a suitable subset \( \Psi_C \), with cardinality \( K_C(j) \), according to the following information criterion:

\[ K_C(j) = \min\{k : \hat{Q}_{c,(k)}(j) - \Gamma(k, n-j) \geq \hat{Q}_{c,(s)}(j) - \Gamma(s, n-j), 1 \leq k, s \leq |\Psi_c|\}, \]  

(17)

---

\(^1\)Kallenberg and Ledwina (1999) considered two configurations: the ‘diagonal’ test includes only terms of the form \( \psi_{ii}, i = 1, 2, \ldots \), while the ‘mixed’ test allows both diagonal and off-diagonal entries. In this study we focus on the latter case, which is more general.

\(^2\)When \( \Psi_c \) contains only \( \psi_{11} \), the corresponding test statistic is proportional to the square of Spearman’s rank correlation coefficient.
where the complexity penalty $\Gamma(k, n - j)$ is given by

$$
\Gamma(k, n - j) = \begin{cases} 
  k \log(n - j) & \text{if } \max_{1 \leq i \leq |\Psi|} |\sqrt{n - j} \hat{\Psi}_{c,i}(j)| \leq \sqrt{2.4 \log(n - j)}, \\
  2k & \text{if } \max_{1 \leq i \leq |\Psi|} |\sqrt{n - j} \hat{\Psi}_{c,i}(j)| > \sqrt{2.4 \log(n - j)}. 
\end{cases} 
$$

(18)

The following theorem characterizes the asymptotic behavior $\hat{Q}_C(j)$ under the null hypothesis and its consistency.

**Theorem 5.** Let $K_C(j)$ be selected according to (17). Suppose that conditions C1-C3 given in the Appendix hold and $M \to \infty$ as $n \to \infty$ with $M = o(\log n/\log \log n)$. (a) Suppose that $C_0(\cdot, \cdot)$ is the independent copula. Then $\Pr(K_C(j) = 1) \xrightarrow{p} 1$ and $\hat{Q}_C(j) \xrightarrow{d} \chi^2_1$ as $n \to \infty$. (b) Suppose instead $(U_t, U_{t-j})$ is distributed according to a distribution $P$ such that $E_P[\psi_{i_1i_2}(U_t, U_{t-j})] \neq 0$ for some $i_1, i_2$ in $1, \ldots, M$. Then $\hat{Q}_C(j) \to \infty$ as $n \to \infty$.

**Remark 5.** The copula-based test of Chen and Fan (2004) tests the hypothesis of independent copula against some commonly used parametric copulas. In contrast, our tests consider nonparametric alternatives in the family of (14). As suggested by the theorem above, data driven specification of the nonparametric alternative ensures that our tests have power against essentially all alternatives.

The test $\hat{Q}_C(j)$ is designed to detect serial dependence between the generalized residuals $j$ periods apart. In practice, it is desirable to jointly test the independence hypothesis at a number of lags. Following Hong et al. (2007), we also consider the following portmanteau test statistic:

$$
\hat{W}_C(J) = \sum_{j=1}^{J} \hat{Q}_C(j),
$$

(19)

where $J \geq 1$ is the longest prediction horizon of interest.

The asymptotic properties of $\hat{W}_C(J)$ follows readily from Theorem 5.

**Theorem 6.** (a) Suppose that the conditions for Theorem 5(a) hold for $j = 1, \ldots, J$. Then $\hat{W}_C(J) \xrightarrow{d} \chi^2_J$ as $n \to \infty$. (b) Suppose that the conditions for Theorem 5(b) hold for at least one $j$ in $j = 1, \ldots, J$. Then $\hat{W}_C(J) \to \infty$ as $n \to \infty$.

Similar to the inference procedure for the uniformity test, we use simulations to approximate the distributions of $\hat{Q}_C(j)$ and $\hat{W}_C(J)$ for $j = 1, \ldots, J$. The simulations are rather straightforward and computationally simple for both test statistics, which are asymptotically distribution free. Below we present a description of the simulation procedure.
• Step C1: Generate an i.i.d. random sample \( \{Z_t\}_{t=1}^n \) from the standard uniform distribution; calculate its empirical distributions denoted by \( \{\hat{U}_t\}_{t=1}^n \).

• Step C2: For \( j = 1, \ldots, J \), select a set of basis functions \( \Psi_C \) according to (17); calculate the test statistic \( \hat{Q}_C(j) \).

• Step C3: Calculate the portmanteau test statistic \( \hat{W}_C(J) = \sum_{j=1}^J \hat{Q}_C(j) \).

• Step C4: Repeat steps C1 to C3 \( L \) times to obtain \( \{\hat{Q}_C(j)^{(l)}\}_{l=1}^L \) for \( j = 1, \ldots, J \) and \( \{\hat{W}_C(J)^{(l)}\}_{l=1}^L \).

• Step C5: Use the \( q \)th percentile of \( \{\hat{Q}_C(j)^{(l)}\}_{l=1}^L \) as the \( q \)th percent critical value of \( \hat{Q}_C(j) \) for \( j = 1, \ldots, J \); use the \( q \)th percentile of \( \{\hat{W}_C(J)^{(l)}\}_{l=1}^L \) as the \( q \)th percent critical value of \( \hat{W}_C(J) \).

3.3 Construction of Sequential Test and Inference

We have presented two separate smooth tests for the stationary distribution and serial independence under the copula representation. Here we proceed to construct a sequential test for the hypothesis of correct density forecast, which is equivalent to the i.i.d. uniformity of the generalized residuals. Under the null hypothesis, the sequential test is invariant to the ordering of the two sub-tests. However, this invariance may be compromised if either serial independence or uniformity does not hold. The test for uniformity is constructed under the assumption of serial independence. Its limiting distribution and critical values, derived under the assumption of correct dynamic specification, are generally not valid in the presence of dynamic misspecification.\(^3\) Using the uniformity test as the first step of a sequential test suffers size distortion if the independence condition is violated. Fortunately, the robust test of independent copula is asymptotically invariant to possible deviations from uniformity of the marginal distributions. Therefore, we choose to use the independence test in the first stage of the sequential test. The test is terminated if the independence hypothesis is rejected; a subsequent test on uniformity is conducted only when the independence hypothesis is not rejected. Consequently, the uniformity test, if needed, is not compromised by possible violation to independence.

The sequential nature of the proposed test complicates its inference: Ignoring the two-stage nature of the design can severely inflate the type I error. Suppose that the significance levels for the first and second stage of a sequential test is set at \( \alpha_1 \) and \( \alpha_2 \) respectively.

\(^3\)Corradi and Swanson (2006a) proposed a test of uniformity that is robust to dynamic misspecification.
Denote by $p_{2|1}$ the probability of rejecting the second stage hypothesis, conditional on not rejecting the first stage hypothesis. The overall type I error of the two-stage test, denoted by $\alpha$, is given by

$$\alpha = \alpha_1 + p_{2|1}(1 - \alpha_1). \quad (20)$$

If the tests from the first and second stage are independent, (20) is simplified to

$$\alpha = \alpha_1 + \alpha_2(1 - \alpha_1). \quad (21)$$

Next we show that the proposed tests on serial independence and uniformity are asymptotically independent under the null hypothesis.

**Theorem 7.** Under the null hypothesis of correct specification of density forecast, the test statistics $\hat{Q}_C(j)$ and $\hat{N}_U$ are asymptotically independent for $j = 1, \ldots, J$; similarly, $\hat{W}_C(J)$ and $\hat{N}_U$ are asymptotically independent.

The asymptotic independence suggested by this theorem facilitates the control of type I error $\alpha$ via proper choice of $\alpha_1$ and $\alpha_2$ based on (21). According to Qiu and Sheng (2008), in the absence of a priori guidance on the significance levels of the two stages, a natural choice is to set $\alpha_1 = \alpha_2$ for a given $\alpha$. We then have

$$\alpha_1 = \alpha_2 = 1 - \sqrt{1 - \alpha}.$$

For instance, $\alpha = 5\%$ implies that $\alpha_1 = \alpha_2 \approx 2.53\%$. After $\alpha_1$ and $\alpha_2$ have been determined, the overall $p$-values can be defined as,

$$p\text{-value} = \begin{cases} p_1, & \text{if} \ p_1 \leq \alpha_1, \\ \alpha_1 + p_2(1 - \alpha_1), & \text{otherwise}, \end{cases} \quad (22)$$

where $p_1$ and $p_2$ are the first and second stage $p$-values. A two-stage test using the $p$-value (22) rejects the overall null hypothesis when one of the following occurs: (i) the first stage null hypothesis is rejected (i.e. $p_1 \leq \alpha_1$); (ii) the first stage null hypothesis is not rejected and the second stage null hypothesis is rejected (i.e. $p_1 > \alpha_1$ and $p_2 \leq \alpha_2$).

4 Simultaneous Test for Correct Density Forecasts

In accordance with the literature (e.g., Hong and Li (2005) and Hong et al. (2007)), we also develop a smooth test to simultaneously test the i.i.d. uniformity of the generalized residuals.
In particular, we examine departure of the joint density function of \((Z_t, Z_{t-j})\) from unity. The null hypothesis is given by

\[ H_0 : p_0(Z_t, Z_{t-j}) = 1, \]  

where \(p_0\) is the joint density function of \((Z_t, Z_{t-j})\).

Let \(\Psi_S\) be a non-empty subset of \(\{\psi_i : 1 \leq i \leq M\} \cup \{\psi_{i_1,i_2} : 1 \leq i_1, i_2 \leq M\}\). Define \(K_S = |\Psi_S|\) and write \(\Psi_S = \{\Psi_{S,1}, \ldots, \Psi_{S,K_S}\}\). Similar to the two tests presented in the previous section, we consider a family of general exponential distribution as the alternative to \(H_0\):

\[ p(z_1, z_2) = \exp \left\{ \sum_{i=1}^{K_S} b_{S,i} \Psi_{S,i}(z_1, z_2) + b_{S,0} \right\}, \quad (z_1, z_2) \in [0,1]^2, \]  

(24)

where \(b_{S,0}\) is a normalization constant. For \(i = 1, \ldots, K_S\), if \(\Psi_{S,i} \in \{\psi_j\}_{j=1}^M\), we set \(\Psi_{S,i}(z_1, z_2) = \Psi_{S,i}(z_1)\) without loss of generality for \(Z_t\) and \(Z_{t-j}\) follow the same marginal distribution.

Let \(B_S = (b_{S,1}, \ldots, b_{S,K_S})'\). Under the assumption that the joint density of \((Z_t, Z_{t-j})\) belongs to (24), testing the hypothesis \(H_0\) is equivalent to testing the following hypothesis:

\[ H_{0S} : B_S = 0. \]

**Remark 6.** A key difference between the simultaneous test and the copula independence test in the previous section is that the copula function takes \(G_0(Z_t)\) and \(G_0(Z_{t-j})\), the CDF's of the marginals, as arguments. Therefore it does not require elements \(\{\psi_j\}_{j=1}^M\), basis functions for the marginal distributions, in the modeling of alternative copula densities. The test on the joint density of \(Z_t\) and \(Z_{t-j}\), on the other hand, makes no assumption about the marginal distributions. Consequently, it includes \(\{\psi_j\}_{j=1}^M\) in its candidate set such that the test is able to detect deviations of the marginal distribution from the uniform distribution.

Naturally, for the purpose of detecting possible deviations of the marginal distributions of \(\hat{Z}_t\) from uniformity, we shall use \(\hat{Z}_t\)'s rather than their empirical distributions to construct the simultaneous test. As a result, proper adjustment to account for the influence of nuisance parameters is warranted. Let \(\hat{\Psi}_S(j) = (\hat{\Psi}_{S,1}(j), \ldots, \hat{\Psi}_{S,K_S}(j))'\) with \(\hat{\Psi}_{S,i}(j) = (n - j)^{-1} \sum_{t=R+j+1}^N \Psi_{S,i}(\hat{Z}_t, \hat{Z}_{t-j})\) for \(i = 1, \ldots, K_S\). Also define \(A = E[s_{0t}s_{0t}]\) and \(B(j) = E[\Psi_S(Z_{0t},Z_{0,t-j})s_{0t}]\) where \(s_{0t}\) and \(Z_{0t}\) are given in (9). The asymptotic properties of \(\hat{\Psi}_S(j)\) are given in the following theorem.

**Theorem 8.** Suppose that conditions C1-C3 given in the Appendix hold. Under \(H_0\), \(\hat{\Psi}_S(j) \overset{p}{\to} \)
and given by of correct density forecast model, Theorem 9. Its asymptotic properties under the null hypothesis are as follows:

Next define \( \hat{A}(j) = (n-j)^{-1} \sum_{t=R+j+1}^{N} \hat{s}_t \hat{s}_t' \) and \( \hat{B}(j) = (n-j)^{-1} \sum_{t=R+j+1}^{N} \Psi_s(\hat{Z}_t, \hat{Z}_{t-j}) \hat{s}_t' \), where \( \hat{s}_t \) is given in (10). We can estimate \( \Sigma_s(j) \) by its sample counterpart:

\[
\hat{\Sigma}_s(j) = I_S + \hat{\eta} \hat{B}(j) \hat{A}(j)^{-1} \hat{B}(j)',
\]

where \( \hat{\eta} \) is given in (10). A simultaneous smooth test on the i.i.d. uniformity of \( \hat{Z}_t \)'s is then given by

\[
\hat{Q}_s(j) = (n-j) \hat{\Psi}_s(j)' \hat{\Sigma}_s(j)^{-1} \hat{\Psi}_s(j).
\]

Its asymptotic properties under the null hypothesis are as follows:

**Theorem 9.** Suppose that the conditions given in Theorem 8 hold. Under the null hypothesis of correct density forecast model, \( \hat{Q}_s(j) \xrightarrow{d} \chi^2_{K_S} \) as \( n \to \infty \).

Similar to the two previous tests, we use data driven method to select a suitable \( \Psi_s(j) \) to capture possible deviations from the null distribution. We start with a candidate set \( \Psi_s \) such that \( \Psi_{s,1} = \psi_{11} \) and the rest of its elements correspond to basis functions in \( \{ \psi_i : 1 \leq i \leq M \} \cup \{ \psi_{i_1i_2} : 1 \leq i_1, i_2 \leq M, (i_1, i_2) \neq (1,1) \} \) arranged in the descending order according to the absolute value of their sample averages. Denote by \( |\Psi_s| \) the cardinality of \( \Psi_s \). Let \( \Psi_s, (k) = \{ \Psi_{s,1}, \ldots, \Psi_{s,k} \}, k = 1, \ldots, |\Psi_s| \) and the corresponding \( \Sigma_s, (k) \) and \( Q_s, (k) \), as given in (25) and (26), are similarly defined; their sample analogs are denoted by \( \hat{\Psi}_s, (k) \), \( \hat{\Sigma}_s, (k) \) and \( \hat{Q}_s, (k) \) respectively. For each \( k \), let \( \hat{\Psi}_s, (k)' = \hat{\nu}_s, (k)' \hat{\Psi}_s, (k) \), where \( \hat{\nu}_s, (k) \) is given by \( \hat{\Sigma}_s, (k)^{-1} = \hat{\nu}_s, (k)' \hat{\Sigma}_s, (k) \). We then use the following information criterion to select \( \Psi_s(j) \), whose cardinality is denoted by \( K_S(j) \):

\[
K_S(j) = \min \{ k : \hat{Q}_s, (k) - \Gamma (k, n) \geq \hat{Q}_s, (h, n), 1 \leq k, h \leq |\Psi_s| \},
\]

where the complexity penalty \( \Gamma (k, n) \) is the same as (18) with \( \hat{\Psi}_s, (j) \) taking the place of \( \hat{\Psi}_c, (j) \).

The asymptotic properties of the data driven test \( \hat{Q}_s(j) \) and its consistency can be established similarly as those of the copula test \( \hat{Q}_C(j) \).

**Theorem 10.** Let \( K_S(j) \) be selected according to (27). Suppose Conditions C1-C8 given in the Appendix hold. (a) Under the null hypothesis \( H_{DS} \), \( \Pr (K_S(j) = 1) \xrightarrow{P} 1 \) and \( \hat{Q}_s(j) \xrightarrow{d} \chi^2_1 \) as \( n \to \infty \). (b) Suppose that \( (Z_t, Z_{t-j}) \) is distributed according to a distribution \( P \) such that
\[ E_P[\psi_i(Z_t)] \neq 0 \text{ for some } i \in 1, \ldots, M, \text{ or } E_P[\psi_{i_1i_2}(Z_t, Z_{t-j})] \neq 0 \text{ for some } i_1, i_2 \in 1, \ldots, M. \]

Let \( \theta^*_0 \) be associated with \( P \) and \( \tilde{\theta} \) such that \( ||\tilde{\theta} - \theta^*_0|| \to 0 \) as \( n \to \infty \) under \( P \). Assume that Conditions C2 and C3 are satisfied with \( \theta_0 \) replaced by \( \theta^*_0 \). Then \( \hat{Q}_S(j) \to \infty \) as \( n \to \infty \).

We can also construct a portmanteau test that examines possible deviations from serial independence up to \( J \) lags:

\[ \hat{W}_S(J) = \sum_{j=1}^J \hat{Q}_S(j). \quad (28) \]

The following theorem describes the asymptotic properties of \( \hat{W}_S(J) \).

**Theorem 11.** (a) Suppose that the conditions for Theorem 10(a) hold for \( j = 1, \ldots, J \). Then \( \hat{W}_S(J) \overset{d}{\to} \chi^2_J \) as \( n \to \infty \). (b) Suppose that the conditions for Theorem 10(b) hold for at least one \( j \) in \( j = 1, \ldots, J \). Then \( \hat{W}_S(J) \to \infty \) as \( n \to \infty \).

We conclude this section with a step-by-step description of a simulation procedure, similar to that for the copula test, for the inference of the simultaneous test.

- **Step S1:** Generate an i.i.d. random sample \( \{Z_t\}_{t=1}^n \) from the standard uniform distribution.
- **Step S2:** For \( j = 1, \ldots, J \), select a set of basis functions \( \Psi_S \) according to (27); calculate the test statistic \( \hat{Q}_S(j) \).
- **Step S3:** Calculate the portmanteau test statistic \( \hat{W}_S(J) = \sum_{j=1}^J \hat{Q}_S(j) \).
- **Step S4:** Repeat steps S1-S3 \( L \) times to obtain \( \{\hat{Q}_S(j)^{(l)}\}_{l=1}^L \) for \( j = 1, \ldots, J \) and \( \{\hat{W}_S(J)^{(l)}\}_{l=1}^L \).
- **Step S5:** Use the \( q \)th percentile of \( \{\hat{Q}_S(j)^{(l)}\}_{l=1}^L \) as the \( q \)th percent critical value of \( \hat{Q}_S(j) \) for \( j = 1, \ldots, J \); use the \( q \)th percentile of \( \{\hat{W}_S(J)^{(l)}\}_{l=1}^L \) as the \( q \)th percent critical value of \( \hat{W}_S(J) \).

**5 Monte Carlo Simulations**

In this section, we examine the finite sample properties of the proposed sequential test (hereafter, “SQT”) and the simultaneous test (hereafter, “SMT”). We generate random samples of length \( N = R + n \) and split the samples into \( R \) in-sample observations for estimation and \( n \) out-of-sample observations for density forecast evaluation. We consider
three out-of-sample sizes: \( n = 250, 500 \) and \( 1,000 \); for each \( n \), we consider three estimation-evaluation ratios: \( R/n = 1, 2, \) and \( 3 \). We repeat each experiment \( 3,000 \) times. We set the confidence level at \( \alpha = 5\% \) and for each \( n \), calculate the critical values based on Monte-Carlo simulations with \( 10,000 \) repetitions under the null hypothesis of \( i.i.d. \) uniformity of \( Z_t \). For computational convenience, we use the fixed scheme in our estimation as in several other studies (see, e.g., Diebold et al. (1998), Chen (2011) and Hong et al. (2007), among others).

The construction of a data driven smooth test starts with a candidate set of basis functions. As suggested by Ledwina (1994) and Kallenberg and Ledwina (1995), the type of orthogonal basis functions makes little difference; e.g., the shifted Legendre polynomials defined on \([0, 1]\) and the cosine series, given by \( \sqrt{2} \cos(i\pi z), i = 1, 2, \ldots \), provide largely identical results. Moreover, the test statistics are not sensitive to \( M \), the number of basis functions included in the candidate set. We use the Legendre polynomials in our simulations. Following Kallenberg and Ledwina (1999), we set \( M = 2 \) and \( K_C \leq 2 \) in the first stage copula test of serial independence of the sequential test. In the second stage test on uniformity (if needed), we set \( M = 10 \), following Ledwina (1994). In the simultaneous test, we set \( M = 4 \) and \( i_1 + i_2 \leq 4 \). For each test, we then apply its corresponding information criterion prescribed in the previous sections to select a suitable set of basis functions, based on which the test statistic is calculated. In the copula test of independence and simultaneous test of independence, we consider the single-lag tests \( \hat{Q}_C(j) \) and \( \hat{Q}_S(j) \) for \( j = 1, 5 \) and \( 10 \), and the portmanteau tests \( \hat{W}_C(J) \) and \( \hat{W}_S(J) \) for \( J = 5, 10 \) and \( 20 \). All tests provide satisfactory results. Since the single-lag tests with \( j = 5 \) or \( 10 \) are generally dominated by those with \( j = 1 \) and the portmanteau tests, we choose not to report them to save space.

To facilitate comparison with the existent literature, we follow the experiment design of Hong et al. (2007) in our simulations. To assess the size of the proposed tests, we consider the following two commonly used models:

1. Random-Walk-Normal model (RW-N):

\[
Y_t = 2.77 \varepsilon_t, \quad \varepsilon_t \sim i.i.d. \ N(0,1)
\]  

(29)

2. GARCH(1,1)-Normal Model(GARCH-N):

\[
\begin{align*}
Y_t &= \sqrt{h_t} \varepsilon_t, \quad \varepsilon_t \sim i.i.d. \ N(0,1) \\
h_t &= 0.76 + 0.14Y_{t-1}^2 + 0.77h_{t-1}.
\end{align*}
\]  

(30)

Table 1 reports the size performances of the SQT and SMT tests. For comparison, we also report the results of the nonparametric omnibus tests by Hong et al. (2007) (hereafter
Both our tests and HLZ tests are asymptotically distribution free and use simulated critical values. The sizes for the two smooth tests are generally close to the 5% theoretical value and do not seem to vary across the $R/n$ ratios. There is little difference between the SQT and SMT tests across experiments, sample sizes and $R/n$ ratios. In contrast, the sizes of HLZ tests vary considerably with the $R/n$ ratio. With $R/n = 1$, their sizes average around 8% and 10% for the RW-N and GARCH-N models respectively. This oversize problem improves with the $R/n$ ratio but seems to persist as sample size increases.

Next we investigate the power of the SQT and SMT tests under three types of departure from the null hypothesis: (i) misspecification present only in the stationary distribution of predictive density, (ii) misspecification present only in the dependence structure of predictive density, (iii) misspecifications in both aspects of predictive density. Following Hong et al. (2007), we focus on testing the correctness of RW-N model against the following three particular DGP’s:

- **DGP1**: Random-Walk-$T$ model (RW-$T$

\[ Y_t = 2.78 \varepsilon_t, \quad \varepsilon_t \sim \text{i.i.d.} \sqrt{\frac{\nu - 2}{\nu}} t(\nu), \tag{31} \]

where $\nu = 3.39$.

- **DGP2**: GARCH-$N$ model defined in (30).

- **DGP3**: Regime-Switching-$T$ model (RS-$T$

\[ Y_t = \sigma(s_t)\varepsilon_t, \quad \varepsilon_t \sim \text{m.d.s.} \sqrt{\frac{\nu(s_t) - 2}{\nu(s_t)}} t(\nu(s_t)), \]

where $s_t = 1$ or 2, and $\text{m.d.s.}$ stands for martingale difference sequence. The transition probability between the two regimes is defined as

\[ P(s_t = l | s_{t-1} = l) = \frac{1}{1 + \exp(-c_l)}, l = 1, 2 \]

where $(\sigma(1), \sigma(2), \nu(1), \nu(2)) = (1.81, 3.67, 6.92, 3.88)$ and $(c_1, c_2) = (3.12, 2.76)$.

The empirical powers of the two smooth tests, together with those of HLZ tests, are reported in Table 2. The smooth tests generally outperform the HLZ tests, often by substantial margins. Under the RW-$T$ alternative, the powers of the smooth tests are larger than 80% while those of the HLZ tests are below 40% when $n = 250$. Although all tests
improve with sample sizes, the performance gaps persist: When \( n = 1,000 \), the powers of all smooth tests reach 100%, while those of the HLZ tests average around 65%. Similarly, the smooth tests dominate the HLZ tests under the GARCH-N alternative across sample sizes and \( R/n \) ratios. When \( n = 250 \), the smooth tests register an average power around 45%, in contrast to an average of 35% for the HLZ tests. Moreover, the performance gaps seem to increase with sample size: When \( n = 1000 \), the corresponding power comparison is roughly 95% v.s. 60%.

Under the RS-T alternative, all three tests are largely comparable. However we caution that the comparison under the third alternative might not be as informative as the first two since the powers of all three tests are close to 100%; alternatives more contiguous to the null hypothesis are likely to be more discriminant among the competing tests. Given the substantial oversize of the HLZ tests, we also conjecture that the seemingly comparable performance between the smooth tests and the HLZ tests might show some difference if we adjust the powers by their corresponding empirical sizes.

Lastly we note that the sequential tests provide an additional benefit in that it diagnoses the source of misspecification.\(^4\) In addition, the two components of the sequential test are of interest themselves as they can serve as stand-alone tests for serial independence and uniformity of the generalized residuals. For completeness, here we briefly discuss simulation results of these two tests, which are reported in Appendix B. Table B.1 reports the results of the robust tests of independent copula, focusing on the hypothesis of RW-N model. The first three columns reflect the empirical sizes, which are centered about the nominal 5% significance level. The middle three columns show the substantial powers of the proposed tests against the alternative of GARCH-N model. The last three columns report the results against the RW-T model, which are correctly centered about the 5% level despite the misspecification of the stationary distribution under the null hypothesis of RW-N model. This third experiment confirms our theoretical analysis that the rank-based copula test of serial independence is robust against misspecification in the stationary distributions. Table B.2 reports the finite sample performance of the uniformity tests. We again focus on the hypothesis of RW-N model and consider three data generating processes: RW-N, RW-T and RS-T. The first three columns show that the uniformity tests are correctly sized and the rest of the tables demonstrate their excellent powers against two alternative models, averaging more than 85% even when the sample size is as small as 250.

\(^4\)The simultaneous tests, to a lesser degree, can also provide similar information. The composition of the selected basis functions provides pointers to possible source of misspecification.
Table 1: Simulation results: size performance

<table>
<thead>
<tr>
<th>DGP</th>
<th>Test</th>
<th>n=250</th>
<th>n=500</th>
<th>n=1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>RW-N</td>
<td>R/n=1</td>
<td>SQT</td>
<td>4.4</td>
<td>5.0</td>
</tr>
<tr>
<td></td>
<td>SMT</td>
<td>4.3</td>
<td>4.4</td>
<td>4.5</td>
</tr>
<tr>
<td></td>
<td>HLZ</td>
<td>7.3</td>
<td>7.5</td>
<td>8.0</td>
</tr>
<tr>
<td>R/n=2</td>
<td>SQT</td>
<td>4.5</td>
<td>4.6</td>
<td>5.2</td>
</tr>
<tr>
<td></td>
<td>SMT</td>
<td>5.0</td>
<td>5.0</td>
<td>4.6</td>
</tr>
<tr>
<td></td>
<td>HLZ</td>
<td>5.6</td>
<td>6.3</td>
<td>6.3</td>
</tr>
<tr>
<td>R/n=3</td>
<td>SQT</td>
<td>4.1</td>
<td>4.6</td>
<td>4.5</td>
</tr>
<tr>
<td></td>
<td>SMT</td>
<td>4.8</td>
<td>4.4</td>
<td>4.5</td>
</tr>
<tr>
<td></td>
<td>HLZ</td>
<td>4.8</td>
<td>5.5</td>
<td>5.8</td>
</tr>
<tr>
<td>GARCH-N</td>
<td>R/n=1</td>
<td>SQT</td>
<td>6.5</td>
<td>7.4</td>
</tr>
<tr>
<td></td>
<td>SMT</td>
<td>6.7</td>
<td>7.5</td>
<td>7.4</td>
</tr>
<tr>
<td></td>
<td>HLZ</td>
<td>10.4</td>
<td>12.4</td>
<td>12.1</td>
</tr>
<tr>
<td>R/n=2</td>
<td>SQT</td>
<td>6.2</td>
<td>6.4</td>
<td>6.1</td>
</tr>
<tr>
<td></td>
<td>SMT</td>
<td>6.4</td>
<td>6.4</td>
<td>6.5</td>
</tr>
<tr>
<td></td>
<td>HLZ</td>
<td>6.6</td>
<td>7.8</td>
<td>7.9</td>
</tr>
<tr>
<td>R/n=3</td>
<td>SQT</td>
<td>5.2</td>
<td>5.8</td>
<td>5.4</td>
</tr>
<tr>
<td></td>
<td>SMT</td>
<td>4.8</td>
<td>5.5</td>
<td>5.5</td>
</tr>
<tr>
<td></td>
<td>HLZ</td>
<td>5.8</td>
<td>6.1</td>
<td>6.5</td>
</tr>
</tbody>
</table>

Note: This table reports the size performances of the sequential tests (SQT) using $\hat{Q}_C(1)$ or $\hat{W}_C(J)$, $J = 5, 10, 20$ as the first stage test statistic and those of the simultaneous tests (SMT). For comparison, the results of Hong et al. (2007)’s nonparametric tests (HLZ), rounded to the first decimal place, are also reported. The nominal size is 0.05. $R$ denotes the number of in-sample observations, $n$ denotes the number of out-of-sample observations. Results are based on 3,000 replications.
Table 2: Simulation results: power performance

<table>
<thead>
<tr>
<th>DGP</th>
<th>Test</th>
<th>n=250</th>
<th>n=500</th>
<th>n=1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>RW-T</td>
<td>R/n=1</td>
<td>SQT</td>
<td>80.0</td>
<td>80.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SMT</td>
<td>77.2</td>
<td>82.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HLZ</td>
<td>36.8</td>
<td>39.8</td>
</tr>
<tr>
<td>RW-T</td>
<td>R/n=2</td>
<td>SQT</td>
<td>86.1</td>
<td>85.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SMT</td>
<td>82.8</td>
<td>87.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HLZ</td>
<td>33.4</td>
<td>34.5</td>
</tr>
<tr>
<td>RW-T</td>
<td>R/n=3</td>
<td>SQT</td>
<td>88.6</td>
<td>88.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SMT</td>
<td>86.0</td>
<td>90.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HLZ</td>
<td>30.6</td>
<td>33.0</td>
</tr>
<tr>
<td>GARCH-N</td>
<td>R/n=1</td>
<td>SQT</td>
<td>40.0</td>
<td>47.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SMT</td>
<td>45.8</td>
<td>43.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HLZ</td>
<td>36.8</td>
<td>30.4</td>
</tr>
<tr>
<td>GARCH-N</td>
<td>R/n=2</td>
<td>SQT</td>
<td>42.7</td>
<td>48.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SMT</td>
<td>47.5</td>
<td>47.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HLZ</td>
<td>33.4</td>
<td>34.5</td>
</tr>
<tr>
<td>GARCH-N</td>
<td>R/n=3</td>
<td>SQT</td>
<td>42.3</td>
<td>49.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SMT</td>
<td>47.5</td>
<td>47.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HLZ</td>
<td>30.6</td>
<td>33.0</td>
</tr>
<tr>
<td>RS-T</td>
<td>R/n=1</td>
<td>SQT</td>
<td>80.4</td>
<td>81.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SMT</td>
<td>78.3</td>
<td>81.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HLZ</td>
<td>72.7</td>
<td>78.8</td>
</tr>
<tr>
<td>RS-T</td>
<td>R/n=2</td>
<td>SQT</td>
<td>85.9</td>
<td>86.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SMT</td>
<td>83.4</td>
<td>87.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HLZ</td>
<td>80.6</td>
<td>86.2</td>
</tr>
<tr>
<td>RS-T</td>
<td>R/n=3</td>
<td>SQT</td>
<td>88.0</td>
<td>88.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SMT</td>
<td>84.2</td>
<td>87.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HLZ</td>
<td>82.1</td>
<td>87.9</td>
</tr>
</tbody>
</table>

Note: This table reports the power performances of the sequential tests (SQT) using $Q_C(1)$ or $W_C(J)$, $J = 5, 10, 20$ as the first stage test statistic and those of the simultaneous tests (SMT). For comparison, the sizes of Hong et al. (2007)'s nonparametric tests (HLZ), rounded to the first decimal place, are also reported. The results of $W(20)$ for HLZ are left blank because they are not reported in HLZ. The nominal size is 0.05. $R$ denotes the number of in-sample observations, $n$ denotes the number of out-of-sample observations. Results are based on 3,000 replications.
6 Empirical Application

In this section, we apply the proposed smooth tests to evaluate various forecast models of stock returns. In particular, we study the daily value-weighted S&P500 returns, with dividends, from July 3, 1962 to December 29, 1995. These data have been analyzed by Diebold et al. (1998) and Chen (2011), among others. Following Diebold et al. (1998), we divide the sample roughly into two halves: Observations from July 3 1962 through December 29, 1978 (with a total of 4133 observations) are used for estimation, while those from January 2, 1979 through December 29, 1995 (with a total of 4298 observations) are used for evaluation.

Diebold et al. (1998) considered three models: i.i.d. normal, MA(1)-GARCH(1,1)-N, and MA(1)-GARCH(1,1)-T. Based on an intuitive graphical approach, they found that the GARCH models significantly outperform the i.i.d. normal model. Their results lend support to the MA(1)-GARCH(1,1)-T model. Although simple, their graphical approach does not account for the influence of nuisance parameters. Chen (2011) proposed a family of moment-based tests on the uniformity and serial independence of generalized residuals, explicitly taking into account the parameter estimation uncertainty. In addition to the three models studied in Diebold et al. (1998), he also considered the MA(1)-EGARCH(1,1)-N and MA(1)-EGARCH(1,1)-T models. His results indicate that for the out-of-sample moment test, the GARCH-T model outperforms the GARCH-N model in the uniformity test; however, both the GARCH and EGARCH models fail to correctly predict the dynamics of the return series in the forecast period.

We revisit this empirical study and consider the following six models: RW-N, RW-T, GARCH(1,1)-N, GARCH(1,1)-T, EGARCH(1,1)-N, and EGARCH(1,1)-T. In accordance with Diebold et al. (1998) and Chen (2011), we use the fixed estimation scheme and adopt an MA(1) specification in the conditional mean for all GARCH-type models. After calculating the generalized residuals according to density forecast models, we test their i.i.d. uniformity using both the sequential and simultaneous tests. To save space, we restrict our discussion to portmanteau tests with \( J = 5, 10 \) and 20. The test results, together with their corresponding simulated critical values, are reported in Table 3. It transpires that the hypothesis of correct density forecast is rejected for all models. Examination of the test results provide the following insights.

- The test statistics of both the sequential and simultaneous tests are substantially larger than their corresponding critical values across all models and specifications of test statistics, rejecting decisively the i.i.d. uniformity of the generalized residuals.

- The first stage of the sequential tests decisively reject the hypothesis of serial indepen-
dence, indicating that none of the models in consideration can adequately describe the
dynamics of the daily S&P500 returns. Similar results are reported by Chen (2011).

- Comparison among the models suggests that GARCH-type models generally outper-
form random walk models, underscoring the importance of modeling volatility cluster-
ing. Allowing asymmetric behavior in volatility through the EGARCH model further
improves the specification of the dynamic structure as is suggested by the robust tests
of serial dependence. On the other hand, the simultaneous tests seem to favor GARCH
models over EGARCH models.

- The rejection of serial independence in the first stage of the sequential test effectively
terminates the test. Nonetheless, we reported the results of stand-alone uniformity tests
as they may provide useful information regarding the goodness-of-fit of the stationary
distributions. The results decisively reject all models with a normal innovation. In
contrast, models with a $t$ innovation are either marginally rejected and not rejected.
This finding is consistent with the consensus that the distributions of stock returns are
fat-tailed.

- Both the sequential and simultaneous tests seem to favor the GARCH-$T$ model among
all models under consideration, which is consistent with Diebold et al. (1998). We plot
the histograms of $\hat{Z}_t$ and the correlograms of $(\hat{Z}_t - \bar{Z})^i$ with $\bar{Z} = n^{-1} \sum_{t=R+1}^{N} \hat{Z}_t$ and
$i = 1, 2, 3, 4$ for the MA(1)-GARCH(1,1)-$T$ model in Figure 1 and 2. Consistent with
the test on uniformity, the histogram of $\hat{Z}_t$ is close to uniform. On the other hand, the
sample autocorrelations of $(\hat{Z}_t - \bar{Z})$ and $(\hat{Z}_t - \bar{Z})^3$ are significantly different from zero
at lag one, indicating that the GARCH(1,1) model fails to adequately characterize the
dynamic structure of stock returns.
Figure 1: Histogram of $\hat{Z}_t$ from the MA(1)-GARCH(1,1)-T model

Figure 2: Correlograms of the powers of $\hat{Z}_t$ from the MA(1)-GARCH(1,1)-T model
Table 3: Test results for estimated density forecast models

<table>
<thead>
<tr>
<th></th>
<th>Critical Value</th>
<th>RW-N</th>
<th>RW-T</th>
<th>GARCH-N</th>
<th>GARCH-T</th>
<th>EGARCH-N</th>
<th>EGARCH-T</th>
</tr>
</thead>
<tbody>
<tr>
<td>Robust test of serial independence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{W}_C(5) )</td>
<td>16.2</td>
<td>294.4</td>
<td>240.5</td>
<td>147.3</td>
<td>160.8</td>
<td>75.5</td>
<td>70.8</td>
</tr>
<tr>
<td>( \hat{W}_C(10) )</td>
<td>24.9</td>
<td>558.3</td>
<td>448.0</td>
<td>150.6</td>
<td>164.1</td>
<td>79.1</td>
<td>74.4</td>
</tr>
<tr>
<td>( \hat{W}_C(20) )</td>
<td>40.7</td>
<td>1158.5</td>
<td>928.9</td>
<td>160.4</td>
<td>174.0</td>
<td>90.4</td>
<td>86.9</td>
</tr>
<tr>
<td>Test of uniformity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{N}_U )</td>
<td>5.8</td>
<td>376.5</td>
<td>21.4</td>
<td>83.8</td>
<td>1.2</td>
<td>92.7</td>
<td>12.7</td>
</tr>
<tr>
<td>Simultaneous tests of ( i.i.d ) Uniformity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{W}_S(5) )</td>
<td>16.9</td>
<td>2184.9</td>
<td>425.7</td>
<td>447.1</td>
<td>166.6</td>
<td>470.6</td>
<td>144.1</td>
</tr>
<tr>
<td>( \hat{W}_S(10) )</td>
<td>26.3</td>
<td>4299.8</td>
<td>812.3</td>
<td>732.7</td>
<td>169.7</td>
<td>826.9</td>
<td>206.8</td>
</tr>
<tr>
<td>( \hat{W}_S(20) )</td>
<td>43.5</td>
<td>8616.9</td>
<td>1659.0</td>
<td>1280.9</td>
<td>179.3</td>
<td>1545.0</td>
<td>338.0</td>
</tr>
</tbody>
</table>

Note: The parameters of the models are estimated from the estimation sample (from July 3, 1962 through December 29, 1978 with a total of 4133 observations). The generalized residuals are obtained using the evaluation sample (from January 2, 1979 through December 29, 1995 with a total of 4298 observations.) In the sequential tests, the hypothesis of independent copula is rejected for all models, effectively terminating the tests. The stand-alone uniformity tests are reported for completeness as they provide useful information about the modeling of marginal distributions.

7 Concluding remarks

We have proposed two data driven smooth tests for the specification of predictive density models. The tests are shown to have nuisance free asymptotic distributions under the null hypothesis of correct specification of predictive density, and have powers against essentially all alternatives. The two components of the sequential test can be used as stand-alone tests for the serial independence and uniformity of generalized residuals associated with forecast models. Monte Carlo simulations demonstrate the excellent performances of these tests.

We have focused on testing correct density forecast models in the present study. All models of stock returns considered in the previous section are rejected by our tests. Although most tests favor the MA-GARCH-T model, it is not clear if this model is significantly better than some of its competitors. For this purpose, formal model selection procedure is needed. In fact, another equally important subject of the predictive density literature is how to select a best model from a set of competing models that might all be misspecified. We conjecture that the methods proposed in the present study can be extended to formal model comparison and model selection. We leave these topics for future study.
References


Appendix A

Some conditions are required to establish the asymptotic properties of the proposed selection rule and test statistics. These conditions should hold in an open neighborhood of the “true” value of $\theta$, denoted by $\theta_0$. This region will then be called $\Theta_0$. Most of the conditions are also required by Inglot et al. (1997) and Hong et al. (2007). We use $C$ to denote a generic bounded constant and $\partial \partial_{\theta} Z_t(\theta_0)$ to denote $\partial \partial_{\theta} Z_t(\theta)|_{\theta=\theta_0}$.

C1. The time series $\{Y_t\}_{t=1}^N$ is generated from an unknown conditional probability density function $f_{\theta}(\cdot|\Omega_{t-1})$, where $\Omega_{t-1}$ is the information set available at time $t-1$.

C2. Let $\Theta$ be a finite-dimensional parameter space. (i) For each $\theta \in \Theta$, $f_t(v|\Omega_{t-1}, \theta)$ is a conditional density model for $\{Y_t\}_{t=1}^N$, and is measurable function of $(v, \Omega_{t-1})$; (ii) with probability one, $f_t(v|\Omega_{t-1}, \theta)$ is twice-continuously differentiable with respect to $\theta$ in $\Theta_0$, with $\lim_{n \to \infty} n^{-1} \sum_{t=R+1}^N E \sup_{\theta \in \Theta_0} |\frac{\partial}{\partial \theta} Z_t(\theta_0)|^2 \leq C$ for some constant $v > 1$ and $\lim_{n \to \infty} n^{-1} \sum_{t=R+1}^N E \sup_{\theta \in \Theta_0} |\frac{\partial^2}{\partial \theta \partial \theta'} Z_t(\theta_0)|^2 \leq C$, where $Z_t(\theta)$ is defined in (2).

C3. (i) $G_{t-1}(z) \equiv E[\frac{\partial}{\partial \theta} Z_t(\theta_0)|Z_t(\theta_0) = z, \Omega_{t-1}]$ is a measurable function of $(z, \Omega_{t-1})$; (ii) with probability one, $G_{t-1}(z)$ is continuously differentiable with respect to $z$, and $\lim_{n \to \infty} n^{-1} \sum_{t=R+1}^N E|G'_{t-1}(Z_t(\theta_0))|^2 \leq C$. 

30
C4. \( \hat{\theta}_t \) is the MLE estimator of \( \theta \) based on the data up to time \( t - 1 \). There exist positive constants \( \kappa_1, \kappa_2, \rho_1 \) and \( n_1 \) such that the estimator \( \hat{\theta}_t \) of \( \theta \in \Theta_0 \) satisfies

\[
P( \sup_{R+1 \leq t \leq N} \| \sqrt{n}(\hat{\theta}_t - \theta) \| \geq r ) \leq \kappa_1 \exp( -\kappa_2 r^2 )
\]

for all \( r = \rho \sqrt{\log n} \) with \( 0 < \rho \leq \rho_1 \) and \( n \geq n_1 \).

C5. There exist positive constants \( m_1, m_2, \kappa_3 \) and \( \kappa_4 \) such that

\[
\sup_{x \in [0,1]} |\psi_i'(x)| \leq \kappa_3 m_1, \quad \sup_{x \in [0,1]} |\psi_i''(x)| \leq \kappa_4 m_2,
\]

for each \( i = 1, 2, \ldots, M \).

C6. \( \{MV_M\}^2n^{-1} \log n \to 0 \) as \( n \to \infty \), where \( V_k = \max_{1 \leq i \leq k} \sup_{x \in [0,1]} |\psi_i(x)| \).

C7. \( M = o(\{n/\log n\}^{(2m)-1}) \) as \( n \to \infty \) where \( m = \max(m_1, m_2) \).

C8. \( M = o(n^c) \) as \( n \to \infty \) for some \( c < \kappa_2 b^{-2} \) if \( \rho_1 b \geq 1 \), and with \( c = \kappa_2 \rho_1^2 \), otherwise, where \( \kappa_2 \) and \( \rho_1 \) are given by C4 and

\[
b = \left\{ \sum_{j=1}^q \var\frac{\partial}{\partial \theta_j} \log f_t(v|\Omega_{t-1}, \theta) \right\}^{1/2}.
\]

Condition C1 describes the data generating process of \( \{Y_t\}_{t=1}^N \). We allow the conditional density function \( f_0(t|\Omega_{t-1}) \) to be time-varying. Conditions C2 and C3 are regularity conditions on the conditional density function \( f_t(v|\Omega_{t-1}, \theta) \). Condition C4 characterizes the convergence rate of the estimator \( \hat{\theta}_t \). By Lemma A1 of West and McCracken (1998), for 0 \( \leq a < 0.5 \), \( \sup |n^a(\hat{\theta}_t - \theta)| \overset{p}{\to} 0 \). It’s noted that Condition C4 satisfies this requirement. Condition C5 concerns the orthonormal system \( \psi_i, i = 1, 2, \ldots \). If \( \{\psi_i\} \) is the orthonormal Legendre polynomials on \([0,1]\), C5 is satisfied with \( m_1 = 5/2 \) and \( m_2 = 9/2 \); If \( \{\psi_i\} \) is the cosine system, it is satisfied with \( m_1 = 1 \) and \( m_2 = 2 \).

Conditions C6-C8 concern \( M \), which governs the dimension of the candidate set, for the test of stationary distribution and for the joint tests. If \( \{\psi_i\} \) is the orthonormal Legendre polynomials on \([0,1]\), \( V_k = (2k+1)^{1/2} \) and hence C6 reduces to \( M^3n^{-1} \log n \to 0 \) as \( n \to \infty \). If \( \{\psi_i\} \) is the cosine system, \( V_k = \sqrt{2} \) and hence C6 reduces to \( M^2n^{-1} \log n \to 0 \) as \( n \to \infty \).

Proofs of Theorem 1 and 2
We introduce some additional notation: \( V_0 = \sum_{k=-\infty}^{\infty} E[\Psi_U(Z_{0,t})\Psi_U(Z_{0,t-k})] \), \( \Xi = E[\frac{\partial}{\partial \theta} s_t]|_{\theta=\theta_0} \), 
\( C = \frac{\partial}{\partial \theta} E[\Psi_U(Z_t)]|_{\theta=\theta_0} \), \( D_0 = \sum_{k=-\infty}^{\infty} E[\Psi_U(Z_0)s'_{0,t-k}] \) and \( A_0 = \sum_{k=-\infty}^{\infty} E[s_{0,t} s'_{0,t-k}] \). 
Recall that \( \hat{\Psi}_U \) is a \( K_U \)-dimensional vector of sample moments. By applying Lemma 4.1 and Lemma 4.2 of West and McCracken (1998), we can obtain the asymptotic normality of \( \hat{\Psi}_U \), 
\[ \sqrt{n} \hat{\Psi}_U \xrightarrow{d} N(0, \Omega), \]
where 
\[ \Omega = V_0 - \eta_1(D_0 \Xi^{-1} C' + C \Xi^{-1} D_0') + \eta_2 C \Xi^{-1} A_0 \Xi^{-1} C', \]
in which \( \eta_1 \) and \( \eta_2 \) depend on the estimation scheme,
\[
\eta_1 = \begin{cases} 
0, & \text{fixed}, \\
1 - \frac{1}{\tau} \ln(1 + \tau), & \text{recursive}, \\
\frac{\tau}{\tau^2}, & \text{rolling (} \tau \leq 1 \text{)}, \\
1 - \frac{1}{\tau^2}, & \text{rolling (} \tau > 1 \text{)}.
\end{cases}
\]
\[
\eta_2 = \begin{cases} 
\tau, & \text{fixed}, \\
2(1 - \frac{1}{\tau} \ln(1 + \tau)), & \text{recursive}, \\
\tau - \frac{\tau^2}{3}, & \text{rolling (} \tau \leq 1 \text{)}, \\
1 - \frac{1}{\tau^2}, & \text{rolling (} \tau > 1 \text{)}.
\end{cases}
\]
Under the null hypothesis of i.i.d. and uniformity of \( Z_t \), by following the argument of Chen (2011), \( \Omega \) can be simplified to \( V_U := I_{K_U} + \eta DA^{-1}D \), where \( D, A \) and \( \eta \) are defined in Theorem 1. Furthermore, we can show that \( \hat{\Psi}_U \xrightarrow{p} \Psi_U \) and \( \hat{V}_U \xrightarrow{p} V_U \). It follows for \( \hat{N}_U = n \hat{\Psi}'_U \hat{V}^{-1}_U \hat{\Psi}_U \), \( \hat{N}_U \xrightarrow{d} \chi^2_{K_U} \).

**Proof of Theorem 3**

Define the simplified BIC and AIC as follows,
\[
SK_U = \min\{k: \hat{N}_{u,(k)} - k \log n \geq \hat{N}_{u,(s)} - s \log n, 1 \leq k, s \leq M\},
\]
\[
AK_U = \min\{k: \hat{N}_{u,(k)} - 2k \geq \hat{N}_{u,(s)} - 2s, 1 \leq k, s \leq M\},
\]
We first prove Theorem 3(a). By Appendix A of Inglot and Ledwina (2006), we can show that when \( \textbf{C6} \) holds, we have \( \Pr(K_U = SK_U) \rightarrow 1 \) as \( n \rightarrow \infty \). Therefore, to prove that \( \Pr(K_U = 1) \rightarrow 1 \), it suffices to show that \( \Pr(SK_U = 1) \rightarrow 1 \).

Let \( \Psi^*_u(k) \) be defined in the same way as \( \hat{\Psi}^*_u(k) \) with \( Z_0t \) defined in (9) replacing \( \hat{Z}_t \). Also define \( a_k = (k-1)n^{-1} \log n \). By Theorem 2.1 of Inglot et al. (1997), we can show that under \( H_{0U} \),
\[ \lim_{n \rightarrow \infty} \sum_{k=2}^{M} \Pr(||\hat{\Psi}^*_u(k) - \Psi^*_u(k)|| \geq (1 - \varepsilon)a_k^{1/2}) = 0. \] (A.1)

Define \( \hat{N}_{u,1} = n \hat{\Psi}^2_{u,1} \hat{V}^{-1}_{u,1} \), where \( \hat{V}_{u,1} \) is the variance of \( \hat{\Psi}_{u,1} \). Because \( SK_U = k \), where \( k \neq 1 \) implies that dimension \( k \) is selected over dimension 1, we get by definition of \( SK_U \),
using (A.1) and $\hat{N}_{u,1} \geq 0$,
\[
\Pr(SK_U \geq 2) = \sum_{k=2}^{M} \Pr(SK_U = k) \\
\leq \sum_{k=2}^{M} \Pr(\hat{N}_{u,(k)} \geq na_k) \\
\leq \sum_{k=2}^{M} \Pr(||\hat{\Psi}^*_u,(k)|| > a_k^{1/2}) \\
\leq \sum_{k=2}^{M} \Pr(||\hat{\Psi}^*_u,(k)|| \geq \varepsilon a_k^{1/2}) + \Pr(||\hat{\Psi}^*_u,(k) - \Psi^*_u,(k)|| \geq (1 - \varepsilon)a_k^{1/2}) \\
= \sum_{k=2}^{M} \Pr(||\Psi^*_u,(k)|| \geq \varepsilon a_k^{1/2}) + o(1)
\]
Next we apply equation (2) of Prohorov (1973) with, in the notation of that paper, $\rho = \varepsilon \{na_k\}^{1/2}$, $m = k$, $\lambda = 1$, $a = \varepsilon a_k^{1/2} k^{1/2} V_k$, yielding
\[
\sum_{k=2}^{M} \Pr(||\Psi^*_u,(k)|| \geq \varepsilon a_k^{1/2}) \leq \kappa_5 \sum_{k=2}^{M} \exp\{-\frac{4}{9} \varepsilon^2 (k - 1) \log n \} \tag{A.2}
\]
for some positive constant $\kappa_5$. The right hand side of (A.2) tends to 0 as $n \to \infty$. This completes the proof that $\Pr(SK_U \geq 2) = 0$.

Further we have,
\[
\Pr(\hat{N}_U \leq x) = \Pr(\hat{N}_1 \leq x) - \Pr(\hat{N}_1 \leq x, K_U \geq 2) + \Pr(\hat{N}_U \leq x, K_U \geq 2)
\]
Because $\hat{N}_1 \xrightarrow{d} \chi^2_1$ under $H_{0U}$ and $\Pr(K_U \geq 2) = 0$, it follows immediately that $\hat{N}_U \xrightarrow{d} \chi^2_1$.

We now prove Theorem 3(b). Under the alternative distribution $P$, it holds that
\[
\lim_{n \to \infty} \Pr \left( \max_{1 \leq k \leq M} |\sqrt{n} \hat{\Psi}^*_u,k| > \sqrt{2.4 \log n} \right) = 1.
\]
It implies that $\Pr(K_U = AK_U) \to 1$ as $n \to \infty$. Therefore, to prove the consistency of $K_U$, it suffices to prove the consistency of $AK_U$.

Denote by $\hat{N}_i$ the test statistic constructed from $\psi_i$, where $\hat{N}_i = n \hat{\psi}_i^2 \hat{V}_i^{-1}$, in which $\hat{V}_i$ is the variance of $\hat{\psi}_i$. Since $E_P[\psi_i(Z_i(\theta))] \neq 0$, we have $n^{-1} \hat{N}_i \to \{E_P[\psi_i(Z_i(\theta))V_i^{-1/2}]\}^2 \geq 0$ under the alternative distribution $P$. 

33
Since $AK_U \geq 1$, we have
\[ n^{-1} \hat{N}_U \geq n^{-1}[\hat{N}_U - 2(AK_U - 1)]. \quad \text{(A.3)} \]

Because $\Psi_U$ is the optimal subset chosen by the rule $AK_U$, we get by definition of $AK_U$,
\[ n^{-1}(\hat{N}_U - 2AK_U) \geq n^{-1}\hat{N}_i - 2n^{-1}. \quad \text{(A.4)} \]

Therefore, by combining (A.3) and (A.4), we can show that $n^{-1}\hat{N}_U \geq n^{-1}\hat{N}_i \geq 0$. The consistency of $AK_U$ easily follows.

**Proofs of Theorem 4**
This theorem is a direct result of Chen and Fan (2006). According to Proposition 3.2 of their paper, when the marginal distributions are estimated by the empirical CDFs, the subsequent copula tests are asymptotically invariant to the estimation of the parameters $\theta$. By applying this result and following Lemma A.1, Lemma A.2 and Lemma A.3 of Kallenberg and Ledwina (1999), we can show that $\hat{\Psi}_{C,i}(j) = (n - j)^{-1} \sum_{t=R+j}^{N} \psi_{C,i}(U_t, U_{t-j}) + o_p(1), i = 1, ..., K_C$. It follows readily, as in the tests for simple hypotheses, that $(n - j)^{-1/2} \sum_{t=R+j}^{N} \psi_{C,i}(U_t, U_{t-j}) \xrightarrow{d} N(0, I_K)$, where $I_K$ is an identity matrix of size $K_C$. It follows immediately that $\hat{Q}_C(j) \xrightarrow{d} \chi^2_{K_C}$ as $n \to \infty$.

**Proofs of Theorem 5**
Theorem 5(a) and (b) are similar to Theorem 1b and Theorem 2b in Kallenberg and Ledwina (1999). For brevity, the proofs are not produced here. We note that Kallenberg and Ledwina (1999) established their results for rank-based independence tests for simple hypothesis in the absence of nuisance parameters. In contrast, our tests are based on the generalized residuals of estimated conditional densities. Nonetheless, Theorem 4 indicates that copula tests based on empirical distributions of the marginal are asymptotically invariant to the estimation of nuisance parameters. Consequently, Theorem 1b and Theorem 2b in Kallenberg and Ledwina (1999) apply to our present directly tests.

**Proof of Theorem 6**
The proof of Theorem 6 is similar to that of Theorem 11, which is more general, and therefore omitted.

**Proofs of Theorem 7**
We first consider the asymptotic independence between $\hat{Q}_C(j)$ and $\hat{N}_U$ for any given $j$. Recalling that $\hat{\Psi}_C(j) = (\hat{\psi}_{C,1}(j), ..., \hat{\psi}_{C,K_C}(j))^\prime$, we have $\sqrt{n - j}\hat{\Psi}_C(j) \xrightarrow{d} N(0, I_{K_C})$. Recalling that $\hat{\Psi}_U = (\hat{\psi}_{U,1}, ..., \hat{\psi}_{U,K_U})^\prime$ and $\hat{\Psi}_U = \hat{\nu}_U^\prime \hat{\psi}_U$, we have $\sqrt{n}\hat{\psi}_U \xrightarrow{d} N(0, I_{K_U})$. 

34
Then, we have,
\[
\left[ \frac{\sqrt{n} - j \hat{\Psi}_C(j)}{\sqrt{n} \hat{\Psi}_U^*} \right] \overset{d}{\to} N\left( 0, \begin{bmatrix} I_{KC} & \Sigma_{12} \\ \Sigma_{21} & I_{K_U} \end{bmatrix} \right)
\]

Now we consider \( \Sigma_{12} \), which is the variance-covariance of \( \sqrt{n} - j \hat{\Psi}_C(j) \) and \( \sqrt{n} \hat{\Psi}_U^* \), where \( \Sigma_{12} = \sqrt{n(n-j)}E[\hat{\Psi}_C(j)\hat{\Psi}_U^*] \). Due to the orthonormal properties of the Legendre polynomials and the \( i.i.d. \) uniformity properties of \( \hat{Z}_t \), it’s easy to show that \( \Sigma_{12} = 0_{KC \times K_U} \), where \( 0_{KC \times K_U} \) is a zero matrix with \( K_C \times K_U \) dimension. Therefore, \( \hat{Q}_C(j) \) and \( \hat{N}_U \) are independent. It follows immediately that \( \hat{W}_C(J) \) and \( \hat{N}_U \) are independent.

**Proofs of Theorem 8 and 9**

The proofs of Theorem 8 and 9 are similar to those of Theorem 1 and 2, and thus omitted.

**Proofs of Theorem 10**

The proof of Theorem 10(a) and (b) are similar to that of Theorem 3 (a) and (b), respectively, and thus omitted.

**Proofs of Theorem 11**

Recalling that \( \hat{W}_S(J) = \sum_{j=1}^J \hat{Q}_S(j) \), where \( \hat{Q}_S(j) = (n-j)\hat{\Psi}_S^*(j)\hat{\Psi}_S^*(j) \) for a given \( j \). Following the result of Theorem 10, we have that under \( H_{0S} \), \( \Pr(K_S(j) = 1) \overset{p}{\to} 1 \) and \( \sqrt{n-j}\hat{\Psi}_S^*(j) \overset{p}{\to} \sqrt{n-j}\hat{\Psi}_{S,1}^*(j) \).

Define
\[
\hat{\Psi}_{S,1,(j)}^* \equiv (\sqrt{n-1}\hat{\Psi}_{S,1}^*(1), \sqrt{n-2}\hat{\Psi}_{S,1}^*(2), ..., \sqrt{n-J}\hat{\Psi}_{S,1}^*(J))'.
\]
We have, \( (\sqrt{n-1}\hat{\Psi}_{S,1}^*(1)', \sqrt{n-2}\hat{\Psi}_{S,1}^*(2)', ..., \sqrt{n-J}\hat{\Psi}_{S,1}^*(J)')' \overset{p}{\to} \hat{\Psi}_{S,1,(J)}^* \). Since \( \sqrt{n-j}\hat{\Psi}_{S,1}^*(j) \overset{d}{\to} N(0,1) \) and \( \hat{\Psi}_{S,1}^*(j) \) and \( \hat{\Psi}_{S,1}^*(i), i \neq j, i,j = 1, ..., J \) are independent, we have \( \hat{\Psi}_{S,1,(J)}^* \overset{d}{\to} N(0, I_J) \). According to Lemma 17.1 in Van der Vaart (2000), we have,
\[
\hat{W}_S(J) \overset{p}{\to} \hat{\Psi}_{S,1,(J)}^*\hat{\Psi}_{S,1,(J)} \overset{d}{\to} \chi_J^2.
\]

Therefore, \( \hat{W}_S(J) \) converges to the chi-square distribution with \( J \) degrees of freedom.

The consistence of \( \hat{W}_S(J) \) follows immediately from the consistency of \( \hat{Q}_S(j), j = 1, ..., J \).
Appendix B

Table B.1: Simulation results of robust tests of serial independence

<table>
<thead>
<tr>
<th>DGP</th>
<th>RW-N</th>
<th></th>
<th>GARCH-N</th>
<th></th>
<th>RW-T</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$R/n=1$</td>
<td>$R/n=2$</td>
<td>$R/n=3$</td>
<td></td>
<td>$R/n=1$</td>
<td>$R/n=2$</td>
<td>$R/n=3$</td>
</tr>
<tr>
<td>$n=250$</td>
<td>$Q_C(1)$</td>
<td>5.5</td>
<td>4.7</td>
<td>4.8</td>
<td>36.6</td>
<td>34.7</td>
<td>36.9</td>
<td>5.4</td>
</tr>
<tr>
<td></td>
<td>$W_C(5)$</td>
<td>5.6</td>
<td>5.7</td>
<td>5.2</td>
<td>43.1</td>
<td>42.2</td>
<td>43.3</td>
<td>4.8</td>
</tr>
<tr>
<td></td>
<td>$W_C(10)$</td>
<td>5.4</td>
<td>5.9</td>
<td>4.9</td>
<td>40.3</td>
<td>40.4</td>
<td>40.6</td>
<td>4.7</td>
</tr>
<tr>
<td></td>
<td>$W_C(20)$</td>
<td>5.1</td>
<td>5.3</td>
<td>4.4</td>
<td>35.3</td>
<td>35.1</td>
<td>35.7</td>
<td>4.7</td>
</tr>
<tr>
<td>$n=500$</td>
<td>$Q_C(1)$</td>
<td>5.3</td>
<td>5.6</td>
<td>5.3</td>
<td>62.5</td>
<td>63.3</td>
<td>62.6</td>
<td>4.7</td>
</tr>
<tr>
<td></td>
<td>$W_C(5)$</td>
<td>5.4</td>
<td>4.0</td>
<td>4.6</td>
<td>72.3</td>
<td>72.7</td>
<td>73.0</td>
<td>4.9</td>
</tr>
<tr>
<td></td>
<td>$W_C(10)$</td>
<td>4.7</td>
<td>4.8</td>
<td>4.1</td>
<td>70.1</td>
<td>69.9</td>
<td>71.1</td>
<td>4.8</td>
</tr>
<tr>
<td></td>
<td>$W_C(20)$</td>
<td>5.1</td>
<td>5.0</td>
<td>4.8</td>
<td>63.0</td>
<td>64.4</td>
<td>65.1</td>
<td>5.1</td>
</tr>
<tr>
<td>$n=1000$</td>
<td>$Q_C(1)$</td>
<td>5.1</td>
<td>4.8</td>
<td>5.3</td>
<td>89.7</td>
<td>90.2</td>
<td>89.4</td>
<td>5.4</td>
</tr>
<tr>
<td></td>
<td>$W_C(5)$</td>
<td>5.7</td>
<td>5.7</td>
<td>4.7</td>
<td>95.6</td>
<td>96.3</td>
<td>95.7</td>
<td>5.7</td>
</tr>
<tr>
<td></td>
<td>$W_C(10)$</td>
<td>5.8</td>
<td>4.7</td>
<td>5.0</td>
<td>95.6</td>
<td>95.8</td>
<td>95.2</td>
<td>5.1</td>
</tr>
<tr>
<td></td>
<td>$W_C(20)$</td>
<td>5.6</td>
<td>4.2</td>
<td>4.1</td>
<td>93.5</td>
<td>93.4</td>
<td>92.3</td>
<td>5.3</td>
</tr>
</tbody>
</table>

Note: This table reports the empirical sizes and powers of the robust tests $\hat{Q}_C(1)$ and $\hat{W}_C(J)$, $J = 5, 10, 20$. The null hypothesis is that the data are generated from RW-N. The nominal size is 0.05. $R$ denotes the number of in-sample observations, $n$ denotes the number of out-of-sample observations. Results are based on 3000 replications.

Table B.2: Simulation results of uniformity test

<table>
<thead>
<tr>
<th>DGP</th>
<th>RW-N</th>
<th></th>
<th>RW-T</th>
<th></th>
<th>RS-T</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$R/n=1$</td>
<td>$R/n=2$</td>
<td>$R/n=3$</td>
<td></td>
<td>$R/n=1$</td>
<td>$R/n=2$</td>
<td>$R/n=3$</td>
</tr>
<tr>
<td>$n=250$</td>
<td>4.2</td>
<td>4.3</td>
<td>4.4</td>
<td>83.5</td>
<td>88.1</td>
<td>91.0</td>
<td>82.4</td>
<td>87.5</td>
</tr>
<tr>
<td>$n=500$</td>
<td>4.8</td>
<td>5.2</td>
<td>4.8</td>
<td>99.0</td>
<td>99.5</td>
<td>99.9</td>
<td>98.4</td>
<td>99.0</td>
</tr>
<tr>
<td>$n=1000$</td>
<td>4.6</td>
<td>5.2</td>
<td>5.7</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

Note: This table reports the empirical sizes and powers of the uniformity tests $\hat{N}_V$. The nominal size is 0.05. $R$ denotes the number of in-sample observations, $n$ denotes the number of out-of-sample observations. Results are based on 3000 replications.