A Sequential Test for the Specification of Predictive Densities

Juan Lin* Ximing Wu†

February 14, 2017

Abstract

We develop a specification test of predictive densities based on that the generalized residuals of correctly specified predictive density models are *i.i.d.* uniform. The proposed *sequential* test examines the hypotheses of serial independence and uniformity in two stages, wherein the first stage test of serial independence is robust to violation of uniformity. The approach of data driven smooth test is employed to construct the test statistics. The asymptotic independence between the two stages facilitates proper control of the overall type I error of the sequential test. We derive the asymptotic null distribution of the test, which is nuisance parameter free, and establish its consistency. Monte Carlo simulations demonstrate excellent finite sample performance of the test. We apply this test to evaluate some commonly used models of stock returns.

JEL Classification Codes: C12; C52; C53

Keywords: predictive density; sequential testing; portmanteau test; data driven smooth 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, USA; Email: xwu@tamu.edu.

We thank Dennis Kristensen (co-editor) and an anonymous referee for many suggestions that greatly improved this article. We are also grateful to Xiaohong Chen, Yi-Ting Chen, Zaichao Du, Wilbert Kallenberg, Teresa Ledwina and Anthony Tay for their valuable inputs. Lin acknowledges support from the National Natural Science Foundation of China (grant 71501164).

1 Introduction

Density forecast is of fundamental importance for decision making under uncertainty, wherein 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 transformation to evaluate out-of-sample density forecasts. The transformed data are often called the generalized residuals of a forecast model. Given a time series $\{Y_t\}$, denote by $\{Z_t\}$ the generalized residuals associated with some density forecast model, which is defined in the next section. They showed that if a forecast model is correctly specified, $\{Z_t\}$ is *i.i.d.* uniformly distributed on $[0, 1]$. The serial independence signifies correct dynamic structure while uniformity characterizes correct specification of the unconditional distribution. Subsequently, many formal tests have been developed based on this approach, extending their original method to accommodate issues such as the influence of nuisance parameters, dynamic misspecification errors, multiple-step-ahead forecasts, etc. For general overviews of this literature see Corradi and Swanson (2006c, 2012) and references therein.

Suppose that $\{Z_t\}$ is a strictly stationary process with an invariant marginal distribution G_0. Let P_0 be the joint distribution of Z_{t-j} and Z_t, where j is a positive integer. According to Sklar's (1959) Theorem, there exists a copula function $C_0 : [0, 1]^2 \to [0, 1]$ such that

$$P_0(Z_{t-j}, Z_t) = C_0(G_0(Z_{t-j}), G_0(Z_t)), \quad (1)$$

where C_0 completely characterizes the dependence structure between Z_{t-j} and Z_t. Most of the existing work evaluate density forecasts by simultaneously testing the serial independence and uniformity hypotheses based on the LHS of (1), comparing the density of P_0 with the product of two standard uniform densities. However, if the joint null hypothesis is rejected, the *simultaneous* test "generally provides no guidance as to why "(Diebold et al., 1998). Is the rejection attributable to a violation of the uniformity of the unconditional distribution of $\{Z_t\}$, a violation of the serial independence of $\{Z_t\}$, or both?

To address this issue, we propose a sequential test for the *i.i.d.* uniformity of the generalized residuals. This *sequential* test, based on the copula representation of a joint distribution as in the RHS of (1), examines firstly whether C_0 is the independent copula and then whether

G_0 is the uniform distribution. Rejection of the independence hypothesis effectively terminates the test; otherwise, a subsequent uniformity test for the unconditional distribution is conducted. We establish the large sample properties of the proposed tests, which are distribution free asymptotically. An appealing feature of our sequential test is that the first stage test of independent copula is constructed to be robust to misspecification of marginal distributions. Therefore it remains valid even if $\{Z_t\}$ is not uniformly distributed (because of misspecified marginal distributions). On the other hand, since the first stage independence test is consistent with asymptotic power equal to 1, the uniformity test (if necessary) is not affected by violation of independence asymptotically. Generally the overall type I error of sequential tests can be difficult to control. Nonetheless, we establish that the two stages of the proposed sequential test are asymptotically independent and suggest a simple method to properly control the overall type I error of our test.

Compared with *simultaneous* tests, the proposed *sequential* test enjoys certain advantages. The first stage robust test for serial independence can be easily constructed since parameter estimation uncertainty does not affect its limiting distribution. Upon the rejection of serial independence, the testing procedure terminates. The testing task is simplified in such cases. More importantly, it facilitates the diagnosis of the source of misspecification. Tay et al. (2000) show that serial dependence in $\{Z_t\}$ may signal poorly captured dynamics, whereas non-uniformity may indicate improper distributional assumptions, or poorly captured dynamics, or both. If the serial independence test is not rejected and the uniformity test is rejected, we may improve deficient density forecast by calibration, see Diebold et al. (1999). On the other hand, if the serial independence of the generalized residuals is rejected, we can further apply uniformity tests that are robust to violations of independence to examine the specification of the marginal distributions.[1] In addition, our serial independence test is based on copulas, which are invariant to strictly monotone transformations of random variables. Therefore, the rich information captured by the copula of $\{Z_t\}$ can be used to diagnose the dynamic structure of the forecast model, in case the serial independence is rejected (see e.g., Chen et al. (2004)).

Below we first briefly review the relevant literature on the specification test of predictive densities and Neyman's smooth tests. We present in Section 3 the sequential test of correct density forecasts and its theoretical properties. In Section 4, we use a series of Monte Carlo simulations to demonstrate the excellent finite sample performance of the proposed tests

[1]Several uniformity tests for dependent data have been proposed in the literature. For example, Munk et al. (2011) extended the data-driven smooth test by Ledwina (1994) to time series, adjusting for the estimation of cumulative autocovariance. Corradi and Swanson (2006a) and Rossi and Sekhposyan (2015) proposed tests for uniformity robust to violations of independence that allow for dynamic misspecification under the null.

under a variety of circumstances. We then apply these tests to evaluate a host of commonly used models of stock market returns in Section 5. The last section concludes. Technical assumptions and proofs of theorems are relegated to Appendix.

2 Background

2.1 Specification Test of Predictive Densities

For simplicity, consider the one-step-ahead forecast of the conditional density $f_{0t}(\cdot|\Omega_{t-1})$ of Y_t,[2] where Ω_{t-1} represents the information set available at time $t-1$. We split a sample of N observations $\{Y_t\}_{t=1}^N$ into an in-sample subset of size R for model estimation and an out-of-sample subset of size $n = N - R$ for forecast performance evaluation. 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 Ω_{t-1}, where $\theta \in \Theta \subset \mathbb{R}^q$. The dynamic Probability Integral Transformation (PIT) of the data $\{Y_t\}_{t=R+1}^N$, with respect to the density forecast $f_t(\cdot|\Omega_{t-1}, \theta)$, is defined as

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

The transformed data, $Z_t(\theta)$, are often called the generalized residuals of a forecast model. Suppose that the density forecast model is correctly specified in the sense that there exists some θ_0 such that $f_{0t}(y|\Omega_{t-1}) = f_t(y|\Omega_{t-1}, \theta_0)$ almost surely (a.s.) and for all ts. Under this condition, Diebold et al. (1998) showed that the generalized residuals $\{Z_t(\theta_0)\}$ should be *i.i.d.* uniform on $[0, 1]$. Therefore, the test of a generic conditional density function $f_t(\cdot|\Omega_{t-1}, \theta)$ is equivalent to a test of the joint hypothesis

$$H_0 : \{Z_t(\theta_0)\} \text{ is a sequence of } i.i.d. \text{ uniform random variables for some } \theta_0 \in \Theta \subset \mathbb{R}^q. \tag{3}$$

The alternative hypothesis is the negation of the null (3). Hereafter we shall write $Z_t(\theta)$ as Z_t for simplicity whenever there is no ambiguity. Since the generalized residuals are constructed using out-of-sample predictions, tests based on generalized residuals are out-of-sample tests.

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 to develop formal specification tests of predictive densities. Diebold et al. (1999) extend the method to bi-variate data. Berkowitz (2001)

[2] Following Diebold et al. (1998), our approach may be extended to handle h-step-ahead density forecasts by partitioning the generalized residuals into groups that are h-periods apart and using Bonferroni bounds.

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. Chen et al. (2004) suggested copula based tests of serial independence of the generalized residuals against alternative parametric copulas. These tests do not consider the effect of parameter estimation in their test statistics. By explicitly accounting for the impact of parameter estimation uncertainty, Bai (2003) proposed Kolmogorov type tests. Hong et al. (2007) constructed nonparametric tests by comparing kernel estimate of the joint density of (Z_{t-j}, Z_t) with the product of two uniform densities. Park and Zhang (2010) proposed data-driven smooth tests, which simultaneously test the uniformity and independence. Chen (2011) considered a family of moment based tests. Recently, Corradi and Swanson (2006a,c,b) proposed Kolmogorov type tests that allow for dynamic misspecification. Rossi and Sekhposyan (2015) proposed a new test wherein parameter estimation error is preserved under the null hypothesis.

2.2 Neyman's Smooth Test

Omnibus tests are desirable in goodness-of-fit testing, wherein the 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 only have good powers to detect a few deviations from the null hypothesis under moderate sample sizes; see e.g, Fan (1996). 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 (see Rayner and Best (1990) for a general review). Here we briefly review the smooth test. For simplicity, suppose for now that $\{Z_t\}_{t=1}^n$ is an *i.i.d.* sample from a distribution G_0 defined on the unit interval. To test the uniformity hypothesis, Neyman (1937) considered an alternative family of smooth distributions given by

$$g(z) = \exp\left(\sum_{i=1}^{k} b_i \psi_i(z) + b_0\right), z \in [0,1], \tag{4}$$

where b_0 is a normalization constant such that g integrates to unity and ψ_i's are shifted Legendre polynomials, given by

$$\psi_i(z) = \frac{\sqrt{2i+1}}{i!} \frac{d^i}{dz^i}(z^2 - z)^i, i = 1, \ldots, k, \tag{5}$$

which are orthonormal with respect to the standard uniform distribution. Consequently, $E[\psi_i(z)] = 0$ for all i's and $E[\psi_i(z)\psi_j(z)] = 0,\ i \neq j$ if z follows the standard uniform distribution.

Under the assumption that G_0 is a member of (4), 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 $\hat{\psi}_{(k)} = (\hat{\psi}_1, ..., \hat{\psi}_k)'$. Neyman's smooth test for uniformity is constructed as

$$N_k = n\hat{\psi}'_{(k)}\hat{\psi}_{(k)}. \tag{6}$$

Under uniformity, N_k converges in distribution to the χ^2 distribution with k degrees of freedom as $n \to \infty$. The performance of smooth test depends on the choice of k. Ledwina (1994) proposed a data-driven approach to select a proper k. Various aspects of this adaptive smooth test are studied by Kallenberg and Ledwina (1995, 1997), Inglot et al. (1997) and Claeskens and Hjort (2004). For applications of smooth tests in econometrics, see for example Bera and Ghosh (2002), Bera et al. (2013), Lin and Wu (2015) and references therein.

3 Sequential Test of Correct Density Forecasts

In this study we propose 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 standard uniform margins. Copulas provide a natural way to separately examine 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 start with the copula-based test of serial independence, followed by the uniformity test of the univariate marginal distributions. We shall then explain the rationale of this sequential test, why we place the independence test in the first stage, and lastly how to obtain desired overall type I error of the test.

3.1 Robust Test of Serial Independence

Copula completely characterizes the dependence structure among random variables. Thus testing for serial independence between Z_{t-j} and Z_t can be based on their copula function. In particular, testing their independence is equivalent to testing the hypothesis that their copula density is constant at unity.[3]

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

$$\hat{Z}_t \equiv Z_t(\hat{\theta}_t) = \int_{-\infty}^{Y_t} f_t(v|\Omega_{t-1},\hat{\theta}_t)dv, \quad t = R+1,...,N, \tag{7}$$

where $\hat{\theta}_t$ is the maximum likelihood estimate (MLE) of θ given by

$$\hat{\theta}_t \equiv \hat{\theta}_{[t_0:t_1]} = \arg\max_{\theta} \sum_{t=t_0}^{t_1} \ln f_t(y_t|\Omega_{t-1},\theta), 1 \leq t_0 < t_1 \leq N.$$

We allow for three different estimation schemes for density forecast. Under the fixed, recursive and rolling schemes, $\hat{\theta}_{[t_0:t_1]} = \hat{\theta}_{[1:R]}$, $\hat{\theta}_{[1:t-1]}$ and $\hat{\theta}_{[t-R:t-1]}$, respectively.

Let $Z_{0t} = Z_t(\theta_0)$. The arguments of the copula distribution function $C_0(G_0(Z_{0,t-j}), G_0(Z_{0t}))$, as defined in (1), are the marginal distribution G_0 of $Z_{0,t-j}$ and Z_{0t}. 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:

$$\hat{G}_n(z) = \frac{1}{n+1}\sum_{t=R+1}^{N} I(Z_t(\hat{\theta}_t) \leq z), \quad z \in [0,1], \tag{8}$$

where $I(\cdot)$ is the indicator function and we divide the summation by $n+1$ rather than the usual n to avoid possibly unbounded copula densities at the boundary. Unlike parametric estimates, the empirical distribution is free of possible misspecification errors.

To fix the idea, we first present the smooth test of serial independence for a given lag j, $1 \leq j \leq n-1$. To ease exposition, whenever there is no confusion, we suppress j in the notation. Denote by $\psi_i : [0,1] \to \mathbb{R}, i = 1,2,\ldots$, a series of basis functions orthonormal with respect to the standard uniform distribution. Define $\psi_{i_1 i_2}(u_1,u_2) = \psi_{i_1}(u_1)\psi_{i_2}(u_2)$.[4] Let Ψ_C be a non-empty subset of $\psi_C = \{\psi_{i_1 i_2} : 1 \leq i_1, i_2 \leq M\}$, where M is a given upper bound.[5] Let $K_C = |\Psi_C|$ be the cardinality of Ψ_C and we write $\Psi_C = \{\Psi_{C,1},\ldots,\Psi_{C,K_C}\}$.

[3] Although the independence test developed here is similar to that of Kallenberg and Ledwina (1999), the purposes of these two tests are rather different: Our test is designed to detect temporal dependence of a univariate time series, while theirs is for contemporaneous dependence between two univariate variables.

[4] 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, which is more general.

[5] Note that Escanciano and Lobato (2009) and Escanciano et al. (2013) also considered a given upper bound in the selection criterion.

Define $U_t = G_0(Z_{0t})$. We consider the following alternative bivariate density of (U_{t-j}, U_t),

$$c(u_1, u_2) = \exp\left\{\sum_{i=1}^{K_C} b_{C,i}\Psi_{C,i}(u_1, u_2) + b_{C,0}\right\}, (u_1, u_2) \in [0, 1]^2. \tag{9}$$

Let c_0 be the density function of C_0 and $B_C = (b_{C,1}, \ldots, b_{C,K_C})'$. Note that if $B_C = 0$, $c(u_1, u_2) = \exp(0) = 1$, yielding the independent copula. Thus under the assumption that the underlying copula density c_0 is a member of (9), testing for independent copula is equivalent to testing the following hypothesis:

$$H_{0C} : B_C = 0. \tag{10}$$

Define $\hat{U}_t = \hat{G}_n(\hat{Z}_t)$ and $\hat{\Psi}_{C,i}(j) = (n-j)^{-1}\sum_{t=R+j+1}^{N}\Psi_{C,i}(\hat{U}_{t-j}, \hat{U}_t)$ for $i = 1, ..., K_C$. A test on the hypothesis (10) is readily constructed as

$$\hat{Q}_C(j) = (n-j)\hat{\Psi}'_C(j)\hat{\Psi}_C(j), \tag{11}$$

where $\hat{\Psi}_C(j) = (\hat{\Psi}_{C,1}(j), \ldots, \hat{\Psi}_{C,K_C}(j))'$. This test is a rank-based test constructed using the empirical CDF's of the generalized residuals. It is particularly appealing as parameter estimation uncertainty in $\hat{\theta}_t$ does not affect the asymptotic distribution of $\hat{Q}_C(j)$. This result is established by the following theorem.

Theorem 1. *Suppose that **C1-C5** and **C7** given in Appendix hold. Under the null hypothesis of independence,* $\hat{Q}_C(j) \xrightarrow{d} \chi^2_{K_C}$ *as* $n \to \infty$.

A critical component of data driven smooth tests is the selection of suitable basis functions Ψ_C from the candidate set ψ_C. In the spirit of Kallenberg and Ledwina (1999), we rearrange the candidate set to Ψ_c such that $\Psi_{c,1} = \psi_{11}$ and the rest of its elements correspond to $\{\psi_{i_1 i_2} : 1 \leq i_1, i_2 \leq M, (i_1, i_2) \neq (1, 1)\}$ arranged in the descending order according to $|\hat{\psi}_{i_1 i_2}| = |(n-j)^{-1}\sum_{t=R+j+1}^{N}\psi_{i_1 i_2}(\hat{U}_{t-j}, \hat{U}_t)|$.[6] Fixing $\Psi_{c,1} = \psi_{11}$ ensures that under the null the data driven Ψ_C converges in probability to one fixed element, ψ_{11}, rather than a random element of $\{\psi_{i_1 i_2} : 1 \leq i_1, i_2 \leq M\}$ associated with the maximum of their sample analogs. Consequently, the asymptotic distribution of $\hat{Q}_C(j)$ is approximately χ^2_1, which simplifies the theoretical analysis.

Given the ordered candidate set Ψ_c, we proceed to use an information criterion to select Ψ_C. Denote the cardinality of Ψ_c by $|\Psi_c|$ and let $\Psi_{c,(k)} = \{\Psi_{c,1}, ..., \Psi_{c,k}\}, k = 1, \ldots, |\Psi_c|$.

[6] When Ψ_c contains only ψ_{11}, the corresponding test statistic is proportional to the square of Spearman's rank correlation coefficient.

Further define $\hat{\Psi}_{c,(k)}(j) = (\hat{\Psi}_{c,1}(j), ..., \hat{\Psi}_{c,k}(j))'$ and $\hat{Q}_{c,(k)}(j) = (n-j)\hat{\Psi}'_{c,(k)}(j)\hat{\Psi}_{c,(k)}(j)$. Following Inglot and Ledwina (2006), we use the following criterion to select a suitable Ψ_C, whose cardinality is denoted by $K_C(j)$:

$$K_C(j) = \min\{k : \hat{Q}_{c,(k)}(j) - \Gamma_1(k, n-j, \zeta) \geq \hat{Q}_{c,(s)}(j) - \Gamma_1(s, n-j, \zeta), 1 \leq k, s \leq |\Psi_c|\}. \quad (12)$$

The penalty $\Gamma_1(s, n, \zeta)$ of this information criterion is given by

$$\Gamma_1(s, n, \zeta) = \begin{cases} s \log n, & \text{if } \max_{1 \leq k \leq |\Psi_c|} |\sqrt{n}\hat{\Psi}_{c,k}(j)| \leq \sqrt{\zeta \log n}; \\ 2s, & \text{if } \max_{1 \leq k \leq |\Psi_c|} |\sqrt{n}\hat{\Psi}_{c,k}(j)| > \sqrt{\zeta \log n}, \end{cases} \quad (13)$$

where $\hat{\Psi}_{c,k}(j)$ is the kth element of $\hat{\Psi}_{c,(k)}(j)$ and $\zeta = 2.4$. Note that this penalty 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 independence.[7]

Next we present the asymptotic properties of the proposed test $\hat{Q}_C(j)$ based on a set of basis functions Ψ_C 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 its consistency.

Theorem 2. *Let $K_C(j)$ be selected according to* (12). *Suppose that the conditions of Theorem 1 hold. (a) Suppose that $C_0(\cdot, \cdot)$ is the independent copula. Then* $\lim_{n\to\infty} \Pr(K_C(j) = 1) = 1$ *and $\hat{Q}_C(j) \xrightarrow{d} \chi^2_1$ as $n \to \infty$. (b) Let $\mathscr{P}$ be an alternative and G_0 be the marginal distribution of Z_t under $\mathscr{P}$. Suppose that $E_{\mathscr{P}}[\psi_{i_1 i_2}(G_0(Z_{t-j}), G_0(Z_t))] \neq 0$ for some i_1, i_2 in $1, \ldots, M$. Then $\hat{Q}_C(j) \to \infty$ as $n \to \infty$.*

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

$$\hat{W}_Q(p) = \sum_{j=1}^{p} \hat{Q}_C(j), \quad (14)$$

where p is the longest prediction horizon of interest. One limitation of this test is that the selection of p can be arbitrary. In order to address this limitation, Escanciano and Lobato (2009) proposed an adaptive portmanteau Box-Pierce test for serial correlation, which selects

[7] Given its asymptotic equivalence to the AIC, the method of cross-validation can also be used for moment selection. We opt for AIC/BIC in this study mainly because of their ease of implementation.

the unknown order of autocorrelation in a data-driven manner. This strategy has been employed to test the correct specification of a vector autoregression model by Escanciano et al. (2013). In a similar spirit, we further consider an adaptive portmanteau test and select the optimal number of lags p according to the following criterion

$$\tilde{p} = \min\{k : \hat{W}_Q(k) - \Gamma_2(k,n,\zeta) \geq \hat{W}_Q(s) - \Gamma_2(s,n,\zeta), 1 \leq k, s \leq p\}, \tag{15}$$

where the complexity penalty $\Gamma_2(k,n,\zeta)$ is the same as (13) with $\max_{1\leq j\leq p}\max_{1\leq i\leq K(j)}|\sqrt{n}\hat{\Psi}_{C,i}(j)|$ taking the place of $\max_{1\leq k\leq|\Psi_c|}|\sqrt{n}\hat{\Psi}_{c,k}(j)|$.

The asymptotic properties of $\hat{W}_Q(\tilde{p})$ follow readily from Theorem 2.

Theorem 3. *Let $\tilde{p}$ be selected according to* (15). *(a) Suppose that the conditions for Theorem 2(a) hold for $j = 1,\ldots,p$. Then $\lim_{n\to\infty}\Pr(\tilde{p}=1) = 1$ and $\hat{W}_Q(\tilde{p}) \xrightarrow{d} \chi^2_1$ as $n\to\infty$. (b) Suppose instead that the conditions for Theorem 2(b) hold for at least one j in $j=1,\ldots,p$. Then $\hat{W}_Q(\tilde{p})\to\infty$ as $n\to\infty$.*

We conclude this section by noting that when the serial independence of the generalized residuals is rejected, it is often of interest to explore whether the serial dependence comes primarily through the conditional mean or higher conditional moments (see Diebold et al., 1998). In order to address this issue, Hong et al. (2007) proposed separate out-of-sample inference procedures that can detect serial dependence of $\{Y_t\}$ in terms of the level, volatility, skewness, kurtosis, and leverage effect, etc. Similarly, we also propose a simple separate test for this purpose.

Let $\mu_k = E[(U-1/2)^k]$ and $\sigma_k^2 = var[(U-1/2)^k]$, where U is a standard uniform random variable. Denote

$$\hat{\Phi}_{k,l}(j) = \frac{1}{n-j}\sum_{t=R+j+1}^{N}\left[\frac{(\hat{U}_{t-j}-1/2)^k-\mu_k}{\sigma_k}\right]\left[\frac{(\hat{U}_t-1/2)^l-\mu_l}{\sigma_l}\right]$$

We further consider the following test

$$\hat{R}_{k,l}(j) = (n-j)\hat{\Phi}_{k,l}(j)^2. \tag{16}$$

Although similar to the independence test of Diebold et al. (1998), the current test uses $\hat{U}_t = \hat{G}_n(\hat{Z}_t)$, rather than $\hat{Z}_t$, to construct the test statistic (16). This assures that the asymptotic distribution of $\hat{R}_{k,l}(j)$ is not affected by the estimation of θ_0, which is needed to obtain $\hat{Z}_t$ (see also Chen, 2011).

Following Hong et al. (2007), we consider $(k,l) = (1,1),(2,2),(3,3),(4,4),(1,2),(2,1)$ in (16) in order to detect possible autocorrelations in level, volatility, skewness, kurtosis, ARCH-in-mean and leverage effects of $\{Y_t\}$, respectively. For example, we can use $\hat{R}_{2,2}(j)$ to test autocorrelation in the volatility of $\{Y_t\}$. Similarly to $\hat{W}_Q(\tilde{p})$, we can further construct a data driven portmanteau test $\hat{W}_R^{(k,l)}(\tilde{p})$ based on $\hat{R}_{k,l}(j)$. Using Theorems 2 and 3, it is straightforward to show that under the null hypothesis, $\hat{R}_{k,l}(j) \xrightarrow{d} \chi_1^2$ and $\hat{W}_R^{(k,l)}(\tilde{p}) \xrightarrow{d} \chi_1^2$ as $n \to \infty$. The critical value of $\hat{R}_{k,l}(j)$ can be tabulated and that of $\hat{W}_R^{(k,l)}(\tilde{p})$ can be calculated following the simulation approach proposed for $\hat{W}_Q(\tilde{p})$, which is given below. The consistency of these tests can also be established in a similar manner.

3.2 Test of Uniformity

Testing correct unconditional distribution of the forecast model is equivalent to testing the hypothesis that the generalized residuals $\{\hat{Z}_t\}$ are uniformly distributed. Since Z_{t-j} and Z_t share the same marginal distribution $G_0(\cdot)$ under (1), testing the uniformity of $\{\hat{Z}_t\}$ is equivalent to testing the uniformity of $\{\hat{Z}_{t-j}\}$. However, conducting the test on $\{\hat{Z}_{t-j}\}$ is advantageous as it allows us to construct a test in the second stage that is independent to the first stage test in the presence of estimated parameter uncertainty. Therefore, in this section, we shall conduct the uniformity test on $\{\hat{Z}_{t-j}\}$.

Let $\Psi_U = \{\Psi_{U,1}, \ldots, \Psi_{U,K_U}\}$ be a non-empty subset of a candidate set $\psi_U = \{\psi_1, \ldots, \psi_M\}$, where $K_U \equiv |\Psi_U|$. We consider a smooth alternative distribution given by

$$g(z) = \exp\left(\sum_{i=1}^{K_U} b_{U,i}\Psi_{U,i}(z) + b_{U,0}\right), z \in [0,1], \tag{17}$$

where $b_{U,0}$ is a normalization constant. Let $B_U = (b_{U,1}, \ldots, b_{U,K_U})'$. Clearly, $B_U = 0$ yields $g(z) = 1$, coinciding with the uniform density. Under the assumption that $\{\hat{Z}_{t-j}\}$ are distributed according to (17), 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 $\hat{\Psi}_U(j) = (\hat{\Psi}_{U,1}(j), ..., \hat{\Psi}_{U,K_U}(j))'$, where $\hat{\Psi}_{U,i}(j) = (n-j)^{-1}\sum_{t=R+j+1}^{N} \Psi_{U,i}(\hat{Z}_{t-j})$ for $i = 1, \ldots, K_U$.

Compared with test (6) 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 forecast. 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 of the predictive density. Define

$$\begin{aligned} s_{0,t-j} &= s_{t-j}|_{\theta=\theta_0}, \quad Z_{0,t-j} = Z_{t-j}(\theta_0), \\ A &= E[s_{0,t-j}s'_{0,t-j}], \quad D = E[\Psi_U(Z_{0,t-j})s'_{0,t-j}]. \end{aligned} \tag{18}$$

We assume that $R, n \to \infty$ as $N \to \infty$ and $\lim_{N\to\infty} n/R = \tau$, for some fixed number $0 \leq \tau < \infty$. Also define

$$\eta = \begin{cases} \tau, & \text{fixed,} \\ 0, & \text{recursive,} \\ -\frac{\tau^2}{3}, & \text{rolling } (\tau \leq 1), \\ -1 + \frac{2}{3\tau}, & \text{rolling } (\tau > 1). \end{cases} \tag{19}$$

Below we show that the asymptotic variance of $\hat{\Psi}_U(j)$ is given by

$$V_U = I_{K_U} + \eta D A^{-1} D',$$

where I_{K_U} is a K_U-dimensional identity matrix. Next define

$$\begin{aligned} \hat{\eta} &= \eta|_{\tau=n/R}, \quad \hat{s}_t = \frac{\partial}{\partial\theta} \ln f_t(Y_t|\Omega_{t-1}, \theta)|_{\theta=\hat{\theta}_t} \\ \hat{A} &= \frac{1}{n-j} \sum_{t=R+j+1}^{N} \hat{s}_{t-j}\hat{s}'_{t-j}, \quad \hat{D} = \frac{1}{n-j} \sum_{t=R+j+1}^{N} \Psi_U(\hat{Z}_{t-j})\hat{s}'_{t-j}. \end{aligned} \tag{20}$$

We can estimate V_U consistently using its sample counterpart:

$$\hat{V}_U = I_{K_U} + \hat{\eta}\hat{D}\hat{A}^{-1}\hat{D}'. \tag{21}$$

We then construct a smooth test of uniformity as follows

$$\hat{N}_U(j) = (n-j)\hat{\Psi}_U(j)'\hat{V}_U^{-1}\hat{\Psi}_U(j). \tag{22}$$

Note here the dependence on j is made explicit in the notations to emphasize that $\hat{N}_U(j)$ is constructed based on $\{\hat{Z}_{t-j}\}$.

Applying the results of West and McCracken (1998) to this test and following the arguments of Chen (2011), we establish the following results.

Theorem 4. *Suppose that **C1**-**C6** given in Appendix hold. Under H_{0U} as $n \to \infty$, (a) $\hat{\Psi}_U(j) \xrightarrow{p} 0$ and $\sqrt{n-j}\hat{\Psi}_U(j) \xrightarrow{d} N(0, V_U)$; (b) the test statistic $\hat{N}_U(j) \xrightarrow{d} \chi^2_{K_U}$.*

Remark 1. *When the parameter* θ *in the forecast density model* $f_t(\cdot|\Omega_{t-1}, \theta)$ *is known,* V_U *is reduced to* I_{K_U}*, the variance obtained under the simple hypothesis. The adjustment* $DA^{-1}D'$ *for nuisance parameters is multiplied by a factor* η *that reflects the estimation scheme of the density forecast.*

Remark 2. *West and McCracken (1998) required the moment functions* ψ_i*'s to be continuously differentiable. McCracken (2000) extended the results of West and McCracken (1998) to allow for non-differentiable moment functions, but their expectations are still required to be continuously differentiable with respect to* θ*. We note that the basis functions considered in this study, such as the Legendre polynomials and cosine series, satisfy the regularity conditions given in West and McCracken (1998).*

Like the copula test of independence presented above, the test of uniformity (22) depends crucially on the configuration of Ψ_U to capture potential deviations from uniformity. In the spirit of Kallenberg and Ledwina (1999), we start with a candidate set Ψ_u such that $\Psi_{u,1} = \psi_1$ and the rest of the set correspond to the elements of $\{\psi_2, ..., \psi_M\}$ arranged in the descending order according to their corresponding entries in the vector $\hat{V}_u^{-1/2}|\sqrt{n-j}\hat{\Psi}_u|$, where $\hat{V}_u$ is the estimated covariance matrix of $\sqrt{n-j}\hat{\Psi}_u$.[8]

Given the ordered candidate set Ψ_u, we proceed to use an information criterion to select Ψ_U. Denote the subset of Ψ_u with its first k elements by $\Psi_{u,(k)} = \{\Psi_{u,1}, ..., \Psi_{u,k}\}$, $k = 1, ..., M$ and the corresponding $V_{u,(k)}$ and $N_{u,(k)}(j)$, as given in (21) and (22), are similarly defined; their sample analogs are denoted by $\hat{\Psi}_{u,(k)}(j), \hat{V}_{u,(k)}$ and $\hat{N}_{u,(k)}(j)$ respectively. For each k, let $\hat{\Psi}^*_{u,(k)}(j) = \hat{V}_{u,(k)}^{-1/2}\hat{\Psi}_{u,(k)}(j)$. Following Inglot and Ledwina (2006), we use the following criterion to select a suitable Ψ_U, whose cardinality is denoted by $K_U(j)$:

$$K_U(j) = \min\{k : \hat{N}_{u,(k)}(j) - \Gamma(k, n-j, \zeta) \geq \hat{N}_{u,(s)}(j) - \Gamma(s, n-j, \zeta), 1 \leq k, s \leq M\}, \quad (23)$$

where the penalty $\Gamma(k, n, \zeta)$ is the same as (13) with $\max_{1\leq k\leq M} |\sqrt{n}\hat{\Psi}^*_{u,k}(j)|$ taking the place of $\max_{1\leq k\leq|\Psi_c|} |\sqrt{n}\hat{\Psi}_{c,k}(j)|$.

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

Theorem 5. *Let* $K_U(j)$ *be selected according to* (23)*. Suppose that* ***C1-C6*** *given in Appendix hold. (a) Suppose that* $Z_{0,t-j}$ *given by* (18) *follows a uniform distribution. Then*

[8]Note that standardization of the moments is necessary for our tests as the elements of $\hat{\Psi}_u$ are generally correlated and differ in variance. In contrast, the test in Kallenberg and Ledwina (1999) utilizes rank-based sample moments that are free of nuisance parameters and asymptotically orthonormal under the null hypothesis. Therefore, their procedure does not require standardization of the moments.

$\lim_{n\to\infty} \Pr(K_U(j) = 1) = 1$ *and* $\hat{N}_U(j) \xrightarrow{d} \chi^2_1$ *as* $n \to \infty$*. (b) Suppose instead that* Z_{t-j} *is distributed according to an alternative distribution* $\mathscr{P}$ *such that* $E_{\mathscr{P}}[\psi_S(Z_{t-j})] \neq 0$ *for some* $S \in \{1, 2, \ldots, M\}$*. Then* $\hat{N}_U(j) \to \infty$ *as* $n \to \infty$*.*

3.3 Construction of Sequential Test and Inference

We have presented two separate smooth tests for the serial independence and uniformity of the generalized residuals $\{Z_t\}$. 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 $\{Z_t\}$.

As is indicated in Introduction, our sequential test facilitates the diagnostics of misspecification in density forecast. Under the null hypothesis of *i.i.d.* uniformity, a sequential test is valid regardless whether the independence test or the uniformity test comes in first. 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. In the presence of dynamic misspecification, serial independence of $\{Z_t\}$ is violated and the uniformity test suffers size distortion. In contrast, the robust test of independent copula is asymptotically invariant to possible deviations from uniformity due to misspecified marginal distributions. Therefore, we choose to place 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. This arrangement assures that the uniformity test (if necessary) is not compromised by possible violation of serial independence.

The sequential nature of the proposed test complicates its inference: ignoring the two-stage nature of the design can sometimes inflate the type I error. Suppose that the significance levels for the first and second stage of a sequential test are set at α_1 and α_2 respectively. 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 α, is given by

$$\alpha = \alpha_1 + p_{2|1}(1 - \alpha_1). \tag{24}$$

If the tests from the first and second stage are independent, (24) is simplified to

$$\alpha = \alpha_1 + \alpha_2(1 - \alpha_1). \tag{25}$$

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

Theorem 6. *Under the null hypothesis of correct specification of density forecast, the test statistics $\hat{Q}_C(j)$ and $\hat{N}_U(j)$ are asymptotically independent for $j = 1, \ldots, p$; similarly, $\hat{W}_Q(p)$ ($\hat{W}_Q(\tilde{p})$) and $\hat{N}_U(p)$ are asymptotically independent.*

The asymptotic independence suggested by this theorem facilitates the control of type I error α via proper choice of α_1 and α_2 based on (25). Qiu and Sheng (2008) suggest that, in the absence of a priori guidance on the significance levels of the two stages, a natural choice is to set $\alpha_1 = \alpha_2$. It follows that

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

For instance, setting $\alpha = 5\%$ yields $\alpha_1 = \alpha_2 \approx 2.53\%$. After α_1 and α_2 have been determined, the overall p-value is given by

$$p\text{-value } = \begin{cases} p_1, & \text{if } p_1 \leq \alpha_1; \\ \alpha_1 + p_2(1 - \alpha_1), & \text{otherwise,} \end{cases} \tag{26}$$

where p_1 and p_2 are p-values of the first and second stage. A two-stage test using the p-value (26) rejects the overall null hypothesis when either 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 $p_2 \leq \alpha_2$).

We conclude this section with a procedure to calculate the critical values for our sequential tests. Since the proposed independence test and uniformity test are constructed in a data driven fashion, the number of functions selected is random even under the null hypothesis. Consequently, the χ_1^2 distribution does not provide an adequate approximation to the distribution of the test statistics under moderate sample sizes. To deal with this problem, we propose a simple simulation procedure to obtain the critical values. Since both tests derived in the previous sections are asymptotically distribution free, their limiting distributions can be approximated via simple simulations by drawing repeatedly from the standard uniform distribution. The procedures are described below.

Serial Independence Test

- For $l = 1 : L$
  - Generate an *i.i.d.* random sample $\{Z_{l,t}\}_{t=1}^n$ from the standard uniform distribution.
  - Calculate the empirical distribution of $\{Z_{l,t}\}_{t=1}^n$, denoted by $\{\hat{U}_{l,t}\}_{t=1}^n$.

- For $j = 1, \ldots, p$, select a set of basis functions Ψ_C according to (12); calculate $\hat{Q}_C^{(l)}(j)$ according to (11).
- Select the optimal number of lags $\tilde{p}$ according to (15); calculate the adaptive portmanteau test $\hat{W}_Q^{(l)}(\tilde{p}) = \sum_{j=1}^{\tilde{p}} \hat{Q}_C^{(l)}(j)$.

- Use the $(1-\alpha_1)^{th}$ percentile of $\{\hat{Q}_C^{(l)}(j)\}_{l=1}^{L}$ as the $(1-\alpha_1)^{th}$ percent critical value of $\hat{Q}_C(j)$ for $j = 1, \ldots, p$; use the $(1-\alpha_1)^{th}$ percentile of $\{\hat{W}_Q^{(l)}(\tilde{p})\}_{l=1}^{L}$ as the $(1-\alpha_1)^{th}$ percent critical value of $\hat{W}_Q(\tilde{p})$.

Uniformity Test (if necessary)

- For $l = 1 : L$
  - Generate an *i.i.d.* random sample $\{Z_{l,t}\}_{t=1}^{n}$ from the standard uniform distribution.
  - Select a set of basis functions Ψ_U according to the selection rule given in (23); Compute the corresponding test statistic $\hat{N}_U^{(l)}(j)$.

- Use the $(1-\alpha_2)^{th}$ percentile of $\{\hat{N}_U^{(l)}(j)\}_{l=1}^{L}$ to approximate the $(1-\alpha_2)^{th}$ percent critical value of $\hat{N}_U(j)$.

Note that the approximated critical values are obtained via simple simulations based on the uniform distribution. These procedures do not require sampling from the data; nor do they entail any estimation based on the data. Numerical experiments reported in Section 4 suggest that the simulated critical values provide good size performance under small sample sizes.

4 Monte Carlo Simulations

In this section, we use numerical simulations to examine the finite sample performance of the proposed sequential test. To facilitate comparison with the existent literature, we closely follow the experiment design of Hong et al. (2007) in our simulations. In particular, 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 four estimation-evaluation ratios: $R/n = 1, 2$ and 3.[9] We repeat each experiment $3,000$ times.

[9]As suggested by a referee, we also examine the case of $R/n = 1/2$. No comparable results are available from Hong et al. (2007) nor Park and Zhang (2010). To save space, the results are reported in the Appendix.

We set the confidence level at $\alpha = 5\%$ and for each n, calculate the critical values using the simulation procedures described above, with $10,000$ repetitions under the null hypothesis of $i.i.d.$ uniformity of $\{Z_t\}$.

Following Hong et al. (2007), we use the following two commonly used models to assess the size of the proposed tests:

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

$$Y_t = 2.77\varepsilon_t, \quad \varepsilon_t \sim \text{i.i.d. N(0,1)} \tag{27}$$

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

$$Y_t = \sqrt{h_t}\varepsilon_t, \quad \varepsilon_t \sim \text{i.i.d. N(0,1)}, \quad h_t = 0.76 + 0.14Y_{t-1}^2 + 0.77h_{t-1} \tag{28}$$

We also focus on testing the correctness of RW-N model against the following alternative DGP's:

- DGP1: Random-Walk-T model (RW-T)[10]

$$Y_t = 2.78\varepsilon_t, \quad \varepsilon_t \sim \text{i.i.d.} \sqrt{\frac{\nu - 2}{\nu}}t(\nu), \quad \nu = 3.39. \tag{29}$$

- DGP2: GARCH-N model defined in (28).

- DGP3: Regime-Switching-T model (RS-T)

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

  where $s_t = 1$ or 2, and $m.d.s.$ 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 $(c_1, c_2) = (3.12, 2.76)$.

[10]We also experimented with the RW-T DGP with the degrees of freedom being 4,5 or 6. The power of our tests is good in general. To save the space, these results are not reported; they are available from the authors upon request.

The construction of a data driven smooth test starts with a candidate set of basis functions. As is shown 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 the size of the candidate set ψ_U or ψ_C as defined in Section 3. We use the shifted Legendre polynomials in our simulations. Following Kallenberg and Ledwina (1999), we set $M = 2$ and $K_C(j) \leq 2$ in the first stage copula test of serial independence. In the second stage test on uniformity (if necessary), we set $M = 10$, following Ledwina (1994). In either 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, we consider the single-lag test $\hat{Q}_C(j)$ with $j = 1, 5$ and 10, the portmanteau tests $\hat{W}_Q(p)$ with $p = 5, 10$ and 20 and the automatic portmanteau test $\hat{W}_Q(\tilde{p})$. In the uniformity test, we use $\hat{N}_U(20)$, which is constructed based on $\hat{Z}_{t-20}$. 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.

Table 1 reports the empirical sizes of the sequential test (hereafter "SQT"). For comparison, we also report the results of the nonparametric omnibus test by Hong et al. (2007) (hereafter "HLZ") and of the simultaneous data-driven smooth test by Park and Zhang (2010) (hereafter "PZ"). Hong et al. (2007) constructed nonparametric tests that jointly test the uniformity and serial independence of $\{Z_t\}$ by comparing kernel estimator of the joint density of (Z_{t-j}, Z_t) with the product of two uniform densities. Because the nuisance parameters converge at a root-n rate while the test statistics converge at nonparametric rates, the effects of nuisance parameter estimation are 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. Park and Zhang (2010) adopted the data driven smooth test to evaluate the accuracy of the conditional density function. Unlike the proposed test, theirs is a simultaneous test of the uniformity and independence that relies on bootstrap critical values. This test is found to have low power against misspecification of the marginal distributions, which will be discussed in more details below.

All three tests use the fixed estimation scheme in their estimation. Both the SQT and PZ tests use the approach of data driven smooth test; their sizes are generally close to the 5% theoretical value and do not seem to vary across the R/n ratios. The sizes of HLZ test vary noticeably 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.

The empirical powers of the SQT test, together with those of the HLZ and PZ tests, are reported in Table 2. The SQT test generally outperforms the HLZ and PZ tests. In particular, under the RW-T DGP, the SQT test dominates the other two tests by substantial margins; under the GARCH-N DGP, the SQT and PZ tests are comparable and dominate the HLZ test; under the RS-T DGP, the SQT and HLZ tests are comparable and dominate the PZ test.

Some remarks are in order. (i) The PZ test essentially focuses on testing the copula density, implicitly assuming that the marginal distributions are uniform. Consequently, it has good power against misspecification in the serial dependence (GARCH-N vs RW-N in our simulations). On the other hand, it has weak power against misspecification in the marginal distributions: In the presence of misspecified marginal distributions, the PZ test is dominated by the other two tests, especially so when the misspecification only occurs in the marginal distributions (RW-T vs RW-N). (ii) The HLZ test, based on the joint density of $(\hat{Z}_{t-j}, \hat{Z}_t)$, is seen to provide good powers against the RS-T alternative, which deviates from the null in both the marginal distributions and serial dependence. On the other hand, it has relatively low power when the violation only occurs in one aspect, as in the RW-T or GARCH-N case. (iii) The proposed SQT tests $\hat{Q}_C(1), \hat{W}_Q(5)$, $\hat{W}_Q(10)$, $\hat{W}_Q(20)$ and $\hat{W}_Q(\tilde{p})$ provide largely similar performance, with slight variations in a few cases. Since the adaptive portmanteau test $\hat{W}_Q(\tilde{p})$ performs well across all alternatives, it is recommended unless there are strong reasons to focus on a specific lag or time horizon in the testing. We shall therefore focus on this test in the following discussion of independence test and the empirical investigations.

As is discussed above, the first stage robust test can serve as stand-alone test for serial independence of the generalized residuals. Table 3 reports simulation results of the robust test $\hat{W}_Q(\tilde{p})$ 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 unconditional distribution under the null hypothesis of RW-N model. This experiment confirms our theoretical analysis that the rank-based copula test of serial independence is robust to misspecification in the unconditional distribution.

Table 1: Simulation results: empirical sizes (fixed estimation scheme)

DGP		Test	n=250					n=500					n=1000				
			$Q(1)$	$W(5)$	$W(10)$	$W(20)$	$W(\tilde{p})$	$Q(1)$	$W(5)$	$W(10)$	$W(20)$	$W(\tilde{p})$	$Q(1)$	$W(5)$	$W(10)$	$W(20)$	$W(\tilde{p})$
RW-N	R/n=1	SQT	4.9	4.9	4.7	5.2	4.9	5.7	6.1	5.1	4.9	5.5	5.5	5.0	5.4	5.0	5.1
		HLZ	7.3	7.5	8.0	7.9		7.5	7.9	8.2	7.8		7.5	7.9	8.2	7.8	
		PZ		5.2	5.2	5.7			5.1	5.1	4.8			5.5	5.6	5.7	
	R/n=2	SQT	4.9	5.7	5.1	5.0	5.2	4.7	5.2	5.1	5.3	4.8	4.6	4.9	4.9	5.1	4.7
		HLZ	5.6	6.3	6.3	6.2		5.9	6.6	6.2	6.4		5.9	6.6	6.2	6.4	
		PZ		4.9	5.2	4.7			4.2	4.8	4.8			4.9	4.5	5.0	
	R/n=3	SQT	4.8	4.3	4.3	4.5	4.4	4.9	5.1	4.7	4.8	5.0	6.0	5.3	5.9	5.6	6.2
		HLZ	4.8	5.5	5.8	5.9		5.3	5.8	5.8	5.9		5.7	5.8	5.6	5.9	
		PZ		5.4	4.7	5.2			6.1	5.9	5.8			4.6	4.2	4.5	
GARCH-N	R/n=1	SQT	7.6	7.8	7.2	7.0	7.5	7.3	7.2	7.1	6.9	7.6	6.6	6.3	6.1	5.8	7.2
		HLZ	10.4	12.4	12.1	12		10.7	10.9	10.7	10.7		8.2	9.8	10.1	9.9	
		PZ		6.2	6.0	6.0			5.3	5.8	6.2			4.6	5.7	6.3	
	R/n=2	SQT	6.5	6.7	6.7	6.4	6.4	6.8	5.9	6.6	6.5	5.5	5.7	5.5	6.1	5.3	5.4
		HLZ	6.6	7.8	7.9	8		7.0	7.2	7.8	7.7		6.2	6.8	6.9	6.8	
		PZ		6.3	6.4	5.4			6.8	6.4	5.9			4.6	4.3	4.6	
	R/n=3	SQT	5.3	5.0	5.0	4.5	5.0	5.4	4.8	5.4	5.1	5.1	5.9	5.4	5.7	5.5	5.5
		HLZ	5.8	6.1	6.5	6.7		6.4	7.0	7.2	6.8		6.0	6.3	6.2	5.7	
		PZ		5.6	4.9	5.5			4.8	5.2	5.1			4.6	4.3	4.4	

Note: This table reports the empirical sizes of the sequential tests(SQT) under the fixed estimation scheme with $\hat{Q}_C(1)$, $\hat{W}_Q(p)$, $p = 5, 10, 20$ or $\hat{W}_Q(\tilde{p})$ in the first stage. The sizes of Hong et al. (2007)'s nonparametric tests (HLZ) and Park and Zhang (2010)'s simultaneous smooth tests (PZ), rounded to the first decimal place, are also reported. The nominal size is 0.05. R/n denotes the estimation-evaluation ratio. Results are based on $3,000$ replications.

Table 2: Simulation results: empirical powers (fixed estimation scheme)

DGP		Test	n=250					n=500					n=1000				
			$Q(1)$	$W(5)$	$W(10)$	$W(20)$	$W(\tilde{p})$	$Q(1)$	$W(5)$	$W(10)$	$W(20)$	$W(\tilde{p})$	$Q(1)$	$W(5)$	$W(10)$	$W(20)$	$W(\tilde{p})$
RW-T	R/n=1	SQT	82.7	82.6	82.8	82.8	82.7	98.6	98.6	98.5	98.5	98.5	100	100	100	100	100
		HLZ	36.8	39.8	38.8	37.9		49.3	50.4	47.1	43.2		65	68.2	61.9		
		PZ		24.2	24.3	25.1			40.3	48.1	56			60	67.2	74.2	
	R/n=2	SQT	86.6	86.6	86.6	86.5	86.8	99.5	99.5	99.5	99.5	99.6	100	100	100	100	100
		HLZ	33.4	34.5	33.3	32.2		45.1	47.3	43.5	38.9		65.2	66.4	60		
		PZ		28.3	32.7	36			40	46	54.8			61.6	70.9	77	
	R/n=3	SQT	87.4	87.6	87.4	87.5	87.4	99.5	99.5	99.5	99.5	99.5	100	100	100	100	100
		HLZ	30.6	33	31.2	30.1		41.6	44.6	41.6	36.8		64.1	65.6	58.2		
		PZ		26.3	34	37.8			40.1	47.6	54.1			61.6	71.7	80.1	
GARCH-N	R/n=1	SQT	41.5	48.1	45.8	42.5	44.2	69	76.8	74.9	69.8	73.1	91.9	96.3	96.1	94.1	95.8
		HLZ	36.8	30.4	39.8	38.8		37.9	36.9	35	50.4		47.1	52.5	43.4		
		PZ		46.3	45	39.9			84.4	81.8	74.5			98.3	97.9	96.4	
	R/n=2	SQT	42.3	47.6	46.3	43.1	44.9	70.4	77.3	75.7	70	74.4	92.7	96.8	96.5	94.7	96.5
		HLZ	33.4	34.5	33.3	32.2		45.1	47.3	43.5	38.9		65.2	66.4	60		
		PZ		55.6	54.6	48.5			85	81.2	75			98.2	97.7	96.1	
	R/n=3	SQT	43.6	49.4	48	43.9	46	69.5	77	75.9	70.2	74	92.7	96.5	96.6	94.9	96.4
		HLZ	30.6	33	31.2	30.1		41.6	44.6	41.6	36.8		64.1	65.6	58.2		
		PZ		54.7	50.9	46.4			81.1	78.2	71.3			98.2	97.8	95.8	
RS-T	R/n=1	SQT	80.5	80.7	80.6	80.2	80.4	98.4	98.5	98.6	98.5	98.4	100	100	100	100	100
		HLZ	72.67	78.77	79.4	79.83		96.07	97.83	98.17	98.1		99.97	100	100		
		PZ		60.3	55.3	49.4			87.7	89.4	89			98.8	99.3	98.9	
	R/n=2	SQT	85.6	85.9	85.9	85.7	85.7	99.3	99.3	99.3	99.3	99.3	100	100	100	100	100
		HLZ	80.63	86.17	86.83	87.1		98.23	99.13	99.27	99.33		100	100	100		
		PZ		66.8	70.2	68.3			87.8	89.2	89.5			99.1	99.5	99.1	
	R/n=3	SQT	87.5	87.7	87.6	87.5	87.6	99.3	99.2	99.2	99.2	99.2	100	100	100	100	100
		HLZ	82.13	87.87	88.43	88.93		98.9	99.33	99.4	99.43		100	100	100		
		PZ		65.9	68.2	66.6			88.4	89.6	89.3			98.9	99.3	99.4	

Note: This table reports the empirical powers of the sequential tests (SQT) with $\hat{Q}_C(1)$, $\hat{W}_Q(p)$, $p = 5, 10, 20$ or $\hat{W}_Q(\tilde{p})$ in the first stage. The powers of Hong et al. (2007)'s nonparametric tests (HLZ) and Park and Zhang (2010)'s simultaneous smooth tests (PZ), 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/n denotes the estimation-evaluation ratio. Results are based on $3,000$ replications.

Table 3: Simulation results of robust tests for serial independence (fixed estimation scheme)

DGP	RW-N			GARCH-N			RW-T		
	$R/n=1$	$R/n=2$	$R/n=3$	$R/n=1$	$R/n=2$	$R/n=3$	$R/n=1$	$R/n=2$	$R/n=3$
n=250	4.7	4.9	5.5	39.8	39.7	40.1	4.9	5	5
n=500	5.1	5.4	5.3	73.1	73.5	72.6	4.4	5.3	5.3
n=1000	4.8	5.1	5.9	95.8	96	95.6	5.2	5.8	5.0

Note: This table reports the empirical sizes and powers of the data driven portmanteau test statistic $\hat{W}_Q(\tilde{p})$ for serial independence. The null hypothesis is that the data are generated from the RW-N model. The nominal size is 0.05. R/n denotes the estimation-evaluation ratio. Results are based on $3,000$ replications.

5 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. Diebold et al. (1998) proposed some intuitive graphical methods to assess separately the serial independence and uniformity of the generalized residuals. Although simple, their graphical approach does not account for the influence of nuisance parameters. Chen (2011) proposes a generalized PIT-based moment test to unify the existing uniformity tests and serial independence tests. In particular, he considered tests based on some pre-determined moment functions and accounted for the parameter estimation uncertainty using the West-McCracken method for out-of-sample tests. 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 4,133 observations) are used for estimation, while those from January 2, 1979 through December 29, 1995 (with a total of 4,298 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. 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. In addition to the three models studied in Diebold et al. (1998), Chen (2011) also considered the MA(1)-EGARCH(1,1)-N and MA(1)-EGARCH(1,1)-T models. His results suggest that 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 eight models: RW-N, RW-T, GARCH(1,1)-N, GARCH(1,1)-T, EGARCH(1,1)-N, EGARCH(1,1)-T, RiskMetrics-N

(RM-N) and RiskMetrics-T (RM-T). This particular set of competitors is considered in this empirical investigation because they facilitate comparison with Diebold et al. (1999) and Chen (2011) and represent the most commonly used models for stock returns. Note that the last two RiskMetrics models were also considered in Hong et al. (2007). 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 and RiskMetrics models.

After calculating the generalized residuals according to density forecast models, we test their *i.i.d.* uniformity using the proposed sequential tests. To save space, we focus on the sequential test with the automatic portmanteau test $\hat{W}_Q(\tilde{p})$ in the first stage. Recall that the significance levels of both stages are set at 2.53% to obtain a 5% overall significance level of the sequential test. The simulated critical value of $\hat{W}_Q(\tilde{p})$ at the 2.53% significance level is 9.12. The test results are reported in Table 4. In order to provide useful information about the specification of marginal distributions, the results for the stand-alone uniformity test $\hat{N}_U(20)$ are also reported.[11] The simulated critical value of $\hat{N}_U(20)$ at the 5% significance level is 5.64. It transpires that the hypothesis of correct density forecast is rejected for all models.

Table 4: Test results for estimated density forecast models

	RW-N	RW-T	GARCH-N	GARCH-T	EGARCH-N	EGARCH-T	RM-N	RM-T
$\hat{W}_Q(\tilde{p})$	838.5	838.5	186.5	189.5	179.9	194.5	182.5	211
$\hat{N}_U(20)$	370.8	241.6	79.9	11.6	70.8	32.5	116.7	47

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.) The hypothesis of independent copula is rejected for all models, effectively terminating the tests.

Examination of the test results provides the following insights.

- The first stage of the robust test decisively rejects the hypothesis of serial independence, indicating that none of the models in consideration can adequately describe the dynamics of the daily S&P500 returns. Similar findings are reported in Chen (2011).
- Comparison among the models suggests that the GARCH-type and RiskMetrics models generally outperform random walk models, underscoring the importance of accounting for volatility clustering. Allowing asymmetric behavior in volatility through the

[11] Note that here we conduct the uniformity test under the null hypotheses of uniformity and independence of $\{Z_t\}$. Therefore, the test results can not be interpreted solely as the deviation from the uniformity of $\{Z_t\}$. Nonetheless, it may provide useful insights.

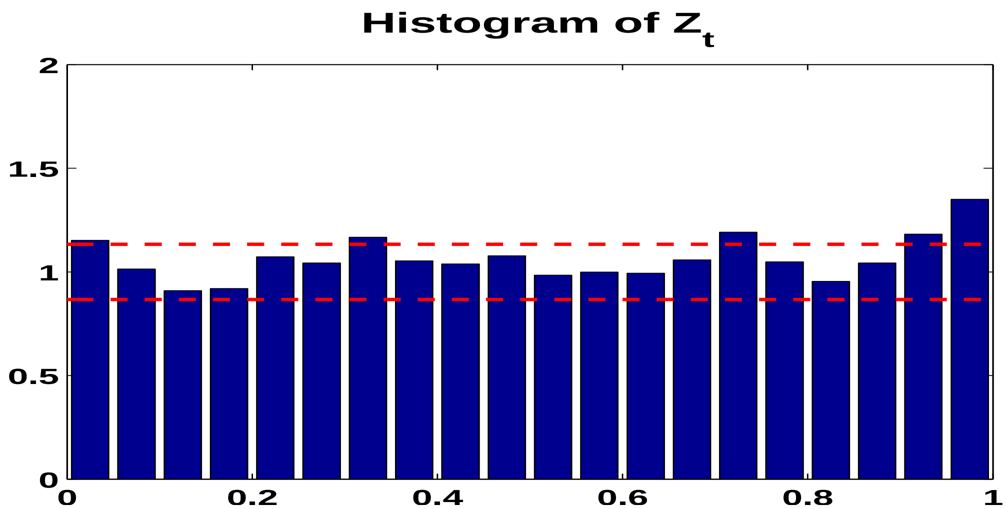

Figure 1: Histogram of $\hat{Z}_t$ from the MA(1)-GARCH(1,1)-T model

EGARCH model or applying the exponential smoothing technique through the RiskMetrics model does not seem to improve the specification of the dynamic structure.

- The rejection of serial independence in the first stage of the sequential test effectively terminates the test. Nonetheless, we report the results of stand-alone uniformity test as they may provide useful information regarding the goodness-of-fit of the unconditional distributions. All models with a normal innovation are rejected decisively, while those with a t innovation are marginally rejected. This finding is consistent with the general consensus that the distributions of stock returns are fat-tailed.

- Both the robust test and stand-alone uniformity test seem to favor the GARCH-T model among all models under consideration, which agrees with the findings of 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 Figures 1 and 2. Consistent with the test on uniformity, the histogram of $\hat{Z}_t$ is nearly 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.

Lastly, we conduct the out-of-sample separate inference tests, as described at the end of Section 3.1, for $\{\hat{Z}_t\}$ derived from MA(1)-GARCH(1,1)-T model. At the 5% significance level, the critical values of $\hat{R}_{k,l}(1)$ at lag one and $\hat{W}_R^{(k,l)}(\tilde{p})$ are 3.84 and 3.97, respectively. The test statistics $\hat{R}_{k,k}(1)$ with $k = 1, 2, 3, 4$ are 153.9, 3.2, 59.2 and 5.3, respectively and the corresponding data driven portmanteau test statistics $\hat{W}_R^{(k,k)}(\tilde{p})$ are 170.9, 18.6, 70.6 and 21.1,

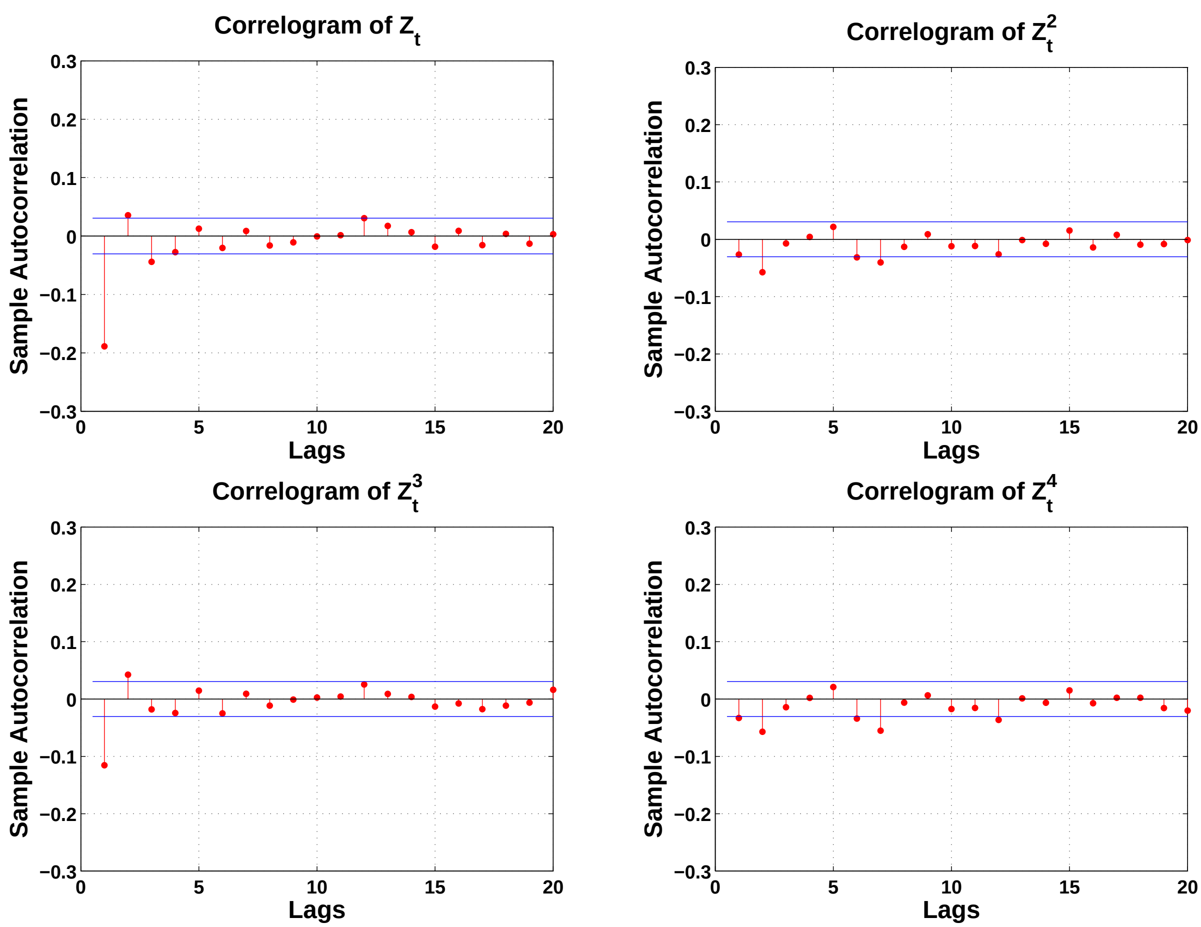

Figure 2: Correlograms of the powers of $\hat{Z}_t$ from the MA(1)-GARCH(1,1)-T model

respectively. These results are consistent with the correlograms reported in Figure 2. Those figures show that at lag one, there are strong autocorrelations in level and skewness, minor autocorrelation in kurtosis and no autocorrelation in variance of $\{\hat{Z}_t\}$. Jointly considering the first 20 lags, there exists strong autocorrelations in the level and skewness of $\{\hat{Z}_t\}$ and small autocorrelations in variance and kurtosis of $\{\hat{Z}_t\}$.

6 Concluding remarks

We have proposed a sequential test for the specification of predictive density models. The proposed test is shown to have a nuisance-parameter-free asymptotic distribution under the null hypothesis of correct specification of predictive density. One attractive feature of the test is that it facilitates the diagnosis of the potential sources of misspecification by separating the independence test and uniformity test of the generalized residuals. Monte Carlo simulations demonstrate excellent performances of the test.

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 an MA-GARCH-T model is preferred, 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

Bai, J. (2003), "Testing parametric conditional distributions of dynamic models," *Review of Economics and Statistics*, 85, 531–549.

Bera, A. K. and Ghosh, A. (2002), "Neymans smooth test and its applications in econometrics," in *Handbook of Applied Econometrics and Statistical Inference*, eds. Ullah, A., Wan, A., and Chaturvedi, A., pp. 177–230.

Bera, A. K., Ghosh, A., and Xiao, Z. (2013), "A smooth test for the equality of distributions," *Econometric Theory*, 29, 419–446.

Berkowitz, J. (2001), "Testing density forecasts, with applications to risk management," *Journal of Business & Economic Statistics*, 19, 465–474.

Chen, X., Fan, Y., and Patton, A. (2004), "Simple tests for models of dependence between multiple financial time series, with applications to US equity returns and exchange rates," *London Economics Financial Markets Group Working Paper*.

Chen, Y.-T. (2011), "Moment tests for density forecast evaluation in the presence of parameter estimation uncertainty," *Journal of Forecasting*, 30, 409–450.

Claeskens, G. and Hjort, N. L. (2004), "Goodness of Fit via Non-parametric Likelihood Ratios," *Scandinavian Journal of Statistics*, 31, 487–513.

Corradi, V. and Swanson, N. R. (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 Economic Forecasting*, eds. Granger, C. W., Elliot, G., and Timmerman, A., Elsevier, pp. 197–284.

— (2012), "A Survey of Recent Advances in Forecast Accuracy Comparison Testing, with an Extension to Stochastic Dominance," in *Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis: Essays in Honor of Halbert L. White Jr*, eds. Chen, X. C. and Swanson, N. R., Springer, pp. 121–144.

de Wet, T. and Randles, R. (1987), "On the effects of substituting parameter estimators in limiting χ^2 U and V statistics," *Annals of Statistics*, 15, 398–412.

Diebold, F. X., Gunther, T. A., and Tay, A. S. (1998), "Evaluating Density Forecasts with Applications to Financial Risk Management," *International Economic Review*, 39, 863–83.

Diebold, F. X., Hahn, J., and Tay, A. S. (1999), "Multivariate density forecast evaluation and calibration in financial risk management: high-frequency returns on foreign exchange," *Review of Economics and Statistics*, 81, 661–673.

Escanciano, J. C. and Lobato, I. N. (2009), "An automatic portmanteau test for serial correlation," *Journal of Econometrics*, 151, 140–149.

Escanciano, J. C., Lobato, I. N., and Zhu, L. (2013), "Automatic specification testing for vector autoregressions and multivariate nonlinear time series models," *Journal of Business & Economic Statistics*, 31, 426–437.

Escanciano, J. C. and Olmo, J. (2010), "Backtesting parametric value-at-risk with estimation risk," *Journal of Business & Economic Statistics*, 28, 36–51.

Fan, J. (1996), "Test of significance based on wavelet thresholding and Neyman's truncation," *Journal of the American Statistical Association*, 91, 674–688.

Genest, C., Ghoudi, K., and Rivest, L. (1995), "A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions," *Biometrika*, 82(3), 543–552.

Hong, Y., Li, H., and Zhao, F. (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.

Inglot, T., Kallenberg, W. C., and Ledwina, T. (1997), "Data driven smooth tests for composite hypotheses," *The Annals of Statistics*, 25, 1222–1250.

Inglot, T. and Ledwina, T. (2006), "Towards data driven selection of a penalty function for data driven Neyman tests," *Linear algebra and its applications*, 417, 124–133.

Jennrich, R. I. (1969), "Asymptotic properties of non-linear least squares estimators," *The Annals of Mathematical Statistics*, 40, 633–643.

Kallenberg, W. C. and Ledwina, T. (1995), "Consistency and Monte Carlo Simulation of a Data Driven Version of Smooth Goodness-of-Fit Tests," *The Annals of Statistics*, 23, 1594–1608.

— (1997), "Data-driven smooth tests when the hypothesis is composite," *Journal of the American Statistical Association*, 92, 1094–1104.

— (1999), "Data-driven rank tests for independence," *Journal of the American Statistical Association*, 94, 285–301.

Ledwina, T. (1994), "Data driven version of the Neyman smooth test of fit," *Journal of American Statistical Association*, 89, 1000–1005.

Lin, J. and Wu, X. (2015), "Smooth Tests of Copula Specifications," *Journal of Business and Economic Statistics*, 33, 128–143.

McCracken, M. W. (2000), "Robust out-of-sample inference," *Journal of Econometrics*, 99, 195–223.

Munk, A., Stockis, J.-P., Valeinis, J., and Giese, G. (2011), "Neyman smooth goodness-of-fit tests for the marginal distribution of dependent data," *Annals of the Institute of Statistical Mathematics*, 63, 939–959.

Park, S. Y. and Zhang, Y. (2010), "Density Forecast Evaluation Using Data-Driven Smooth Test," *unpublished manuscript*.

Qiu, P. and Sheng, J. (2008), "A two-stage procedure for comparing hazard rate functions," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70, 191–208.

Randles, R. (1984), "On tests applied to residuals," *Journal of American Statistical Association*, 79, 349–354.

Rayner, J. and Best, D. (1990), "Smooth tests of goodness of fit: an overview," *International Statistical Review*, 58, 9–17.

Rossi, B. and Sekhposyan, T. (2015), "Alternative tests for correct specification of conditional predictive densities," *Available at* *https://www.aeaweb.org/aea/2016conference/program/retrieve.php?pdfid=1017*.

Sklar, A. (1959), "Fonctions de rpartition n dimensions et leurs marges," *Publications de l'Institut de Statistique de l'Universit de Paris*, 8, 229–231.

Tay, A. S., Wallis, K. F., et al. (2000), "Density forecasting: a survey," *Journal of forecasting*, 19, 235–254.

West, K. D. and McCracken, M. W. (1998), "Regression-Based Tests of Predictive Ability," *International Economic Review*, 39, 817–840.

Appendix A. Proof of Theorems

We first introduce some notations. For a vector $a = (a_1, ..., a_k)'$ and $r > 0$, define $|a|_r = (\sum_{i=1}^{k} |a_i|^r)^{1/r}$, $|a| = |a|_2$ and $|a|_\infty = \max_{1\le i\le k} |a_i|$. For a matrix $A = (a_{ij})_{1\le i\le m, 1\le j\le n}$, let $\|A\|$ denote the Euclidean norm and $\|A\|_\infty = \max_{1\le i\le m, 1\le j\le n} |a_{ij}|$. For convenience, define $n_j = n - j$.

Some conditions are required to establish the asymptotic properties of the proposed selection rule and test statistics. These conditions should hold in an arbitrary neighbourhood of $\theta_0 \in \Theta$. This region will then be called Θ_0. We use C to denote a generic constant, which might vary from one place to another.

C1 $\{Y_t\}_{t=1}^N$ is generated from an unknown conditional probability density function $f_{0t}(v|\Omega_{t-1})$, where Ω_{t-1} is the information set available at time $t-1$.

C2 The generalized residuals $\{Z_t\}_{t=R+1}^N$ of the density forecast model $f_{0t}(v|\Omega_{t-1})$ is a sample of strictly stationary and ergodic process.

C3 Let Θ 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 a measurable function of (v, Ω_{t-1}); (ii) $f_t(v|\Omega_{t-1}, \theta)$ is twice-continuously differentiable with respect to θ in Θ_0 with probability one.

C4 The expected moment function $E[\psi_i(G_0(Z_t(\theta)))]$ is differentiable with respect to θ.

C5 $\hat{\theta}_t$ is a $\sqrt{t}$-consistent estimator for θ_0. Moreover, $\hat{\theta}_t$ satisfies $\hat{\theta}_t - \theta_0 = \Xi(t)^{-1}S(t)$, where (a) $\Xi(t) \to \Xi$, a matrix of rank q; (b) Depending on the forecasting scheme, $S(t)$ is a $q \times 1$ vector such that $S(t) = (t-1)^{-1}\sum_{k=1}^{t-1} s_{0k}$ (recursive), $S(t) = R^{-1}\sum_{k=t-R}^{t-1} s_{0k}$ (rolling) or $S(t) = R^{-1}\sum_{k=1}^{R} s_{0k}$ (fixed), where s_{0t} is defined in (18); (c) $E[s_{0k}] = 0$ a.s..

C6 $V_U = I_{K_U} + \eta D A^{-1} D'$ is finite and positive definite, where D, A and η are defined in (18) and (19).

C7 Denote $G_t(\theta, z) = \Pr(Z_t(\theta) \le z|\Omega_{t-1})$. $G_t(\theta, z)$ is continuously differentiable in θ and z $(a.s.)$. Moreover,

$$E\left[\sup_{\theta\in\Theta_0, z\in R}\left|\frac{\partial G_t(\theta, z)}{\partial z}\right|\right] < C, \quad E\left[\sup_{\theta\in\Theta_0, z\in R}\left|\frac{\partial G_t(\theta, z)}{\partial \theta}\right|\right] < C,$$

Condition **C1** describes the data generating process of $\{Y_t\}_{t=1}^N$. We allow the conditional density function $f_{0t}(\cdot|\Omega_{t-1})$ to be time-varying. Condition **C2** imposes regularity conditions on the generalized residuals $\{Z_t\}$. We allow but do not require $\{Z_t\}$ to be a stationary Markov process. We need this assumption to prove the convergence results. Condition **C3** prescribes regularity conditions on the conditional density function $f_t(v|\Omega_{t-1},\theta)$. Condition **C4** is discussed in details in remark 2. Condition **C5** is satisfied by most estimators used in the literature, including the maximum likelihood estimators and standard GMM estimators. Under this assumption, we have $\max_{R+1\leq t\leq N}\sqrt{t}(\hat{\theta}_t-\theta_0)=O_p(1)$. Condition **C7** is required for the asymptotic equicontinuity of certain empirical processes and the uniform law of large numbers.

Below we shall start with the proof of Theorems 4 and 5 for the univariate uniformity test. Some results in their proofs are used to prove Theorems 1-3 on copula tests of serial independence.

Proof of Theorem 4

We introduce some additional notation: $\Upsilon=\frac{\partial}{\partial\theta'}E[\Psi_U(Z_{t-j})]|_{\theta=\theta_0}$, $\Xi=E[\frac{\partial}{\partial\theta'}s_{t-j}]|_{\theta=\theta_0}$, $V_0=\sum_{k=-\infty}^{\infty}E[\Psi_U(Z_{0,t-j})\Psi_U(Z_{0,t-j-k})']$, $D_0=\sum_{k=-\infty}^{\infty}E[\Psi_U(Z_{0,t-j})s'_{0,t-j-k}]$ and $A_0=\sum_{k=-\infty}^{\infty}E[s_{0,t-j}s'_{0,t-j-k}]$. Note that the gradient function $s_{0,t-j}$ is defined in (18) and the matrix Ξ is defined in **C5**. 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), under Condition **C2** and **C5**, we can obtain the asymptotic normality of $\hat{\Psi}_U(j)$,

$$\sqrt{n_j}\hat{\Psi}_U(j)\xrightarrow{d}N(0,\Omega),$$

where

$$\Omega=V_0-\eta_1(D_0\Xi^{-1}\Upsilon'+\Upsilon\Xi^{-1}D_0')+\eta_2\Upsilon\Xi^{-1}A_0\Xi^{-1}\Upsilon',$$

in which η_1 and η_2 depend on the estimation scheme as follows

$$\eta_1=\begin{cases}0, & \text{fixed},\\ 1-\frac{1}{\tau}\ln(1+\tau), & \text{recursive},\\ \frac{\tau}{2}, & \text{rolling } (\tau\leq 1),\\ 1-\frac{1}{2\tau}, & \text{rolling } (\tau>1)\end{cases}\qquad \eta_2=\begin{cases}\tau, & \text{fixed},\\ 2(1-\frac{1}{\tau}\ln(1+\tau)), & \text{recursive},\\ \tau-\frac{\tau^2}{3}, & \text{rolling } (\tau\leq 1),\\ 1-\frac{1}{3\tau}, & \text{rolling } (\tau>1)\end{cases}\tag{A.1}$$

Under the null hypothesis of *i.i.d.* and uniformity of Z_t, following the proofs of Chen (2011) on page 423, by the law of iterated expectations and the martingale-difference conditions: $E[\Psi_U(Z_{0,t-j})|\Omega_{t-j-1}] = 0$ and $E[s_{0,t-j}|\Omega_{t-j-1}] = 0$ for all $(t-j)$'s, we can show that $\forall k \neq 0$, $E[\Psi_U(Z_{0,t-j})\Psi_U(Z_{0,t-j-k})'] = 0$, $E[\Psi_U(Z_{0,t-j})s'_{0,t-j-k}] = 0$ and $E[s_{0,t-j}s'_{0,t-j-k}] = 0$. Accordingly, we can obtain the simplified results $V_0 = I_{K_U}$, $D_0 = D$ and $A_0 = A$, and hence $\Omega = I_{K_U} - \eta_1(D\Xi^{-1}\Upsilon' + \Upsilon\Xi^{-1}D') + \eta_2\Upsilon\Xi^{-1}A\Xi^{-1}\Upsilon'$, where D and A are defined in (18). Furthermore, we can use the generalized information matrix equality to write $\Xi + A = 0$ and $\Upsilon + D = 0$ under H_0. Therefore, we can further simplify Ω to $V_U := I_{K_U} + \eta DA^{-1}D$, where η are defined in (19). Furthermore, due to the consistency of n/R for τ and the MLE $\hat{\theta}_t$ for θ_0 under H_0, we can show that $\hat{\Psi}_U \xrightarrow{p} \Psi_U$ and $\hat{V}_U \xrightarrow{p} V_U$ by the uniform law of large number theorem (ULLN) of Jennrich (1969, Theorem 2). Under **C6**, it follows that $\hat{N}_U(j) \xrightarrow{d} \chi^2_{K_U}$.

Proof of Theorem 5 Define the simplified BIC as follows

$$SK_U = \min\{k : \hat{N}_{u,(k)}(j) - k\log n_j \geq \hat{N}_{u,(s)}(j) - s\log n_j, 1 \leq k, s \leq M\},$$

where M is a given number that is sufficiently large.

In order to prove Theorem 5(a), we need to establish that, under the null hypothesis,

$$\lim_{n\to\infty} \Pr(K_U(j) = SK_U) = 1, \tag{A.2}$$

and

$$\lim_{n\to\infty} \Pr(SK_U = 1) = 1. \tag{A.3}$$

We start by proving (A.2). Recall that $\hat{\Psi}^*_{u,(k)}(j) = \hat{V}^{-1/2}_{u,(k)}\hat{\Psi}_{u,(k)}(j)$, which is a $k \times 1$ vector. Define the event

$$A_n(\zeta) = \left\{\sqrt{n_j}|\hat{\Psi}^*_{u,(k)}(j)|_\infty > \sqrt{\zeta \log n_j}\right\}.$$

By the definition of $\Gamma(k, n, \zeta)$ defined in (23), in order to prove (A.2), it suffices to prove $P(A_n(\zeta)) = o(1)$. Define $\tilde{\Psi}^*_{u,(k)}(j) = V^{-1/2}_{u,(k)}[n_j^{-1}\sum_{t=R+j+1}^N \Psi_{u,(k)}(Z_{0,t-j})]$. We can rewrite $\hat{\Psi}^*_{u,(k)}(j)$ as,

$$\sqrt{n_j}\hat{\Psi}^*_{u,(k)}(j) = \sqrt{n_j}\hat{V}^{-1/2}_{u,(k)}V^{1/2}_{u,(k)}\tilde{\Psi}^*_{u,(k)}(j) + R_{n(k)},$$

where

$$R_{n(k)} = \sqrt{n_j}\hat{V}^{-1/2}_{u,(k)}\left\{\hat{\Psi}_{u,(k)}(j) - \frac{1}{n_j}\sum_{t=R+j+1}^{N}\Psi_{u,(k)}(Z_{0,t-j})\right\}. \tag{A.4}$$

Since $|\hat{\theta}_t - \theta_0| = O_p(t^{-1/2})$ by Condition **C5**, we have $\hat{\Psi}_{u,(k)}(j) = n_j^{-1}\sum_{t=R+1}^N \Psi_{u,(k)}(Z_{0,t-j}) + O_p(n^{-1/2})$. By (A.4), we obtain $|R_{n(k)}|_\infty = O_p(1)$. Therefore, for any $\epsilon > 0$, we have $P(|R_{n(k)}|_\infty > 2^{-1}\sqrt{\zeta\log n_j}) \leq \epsilon/2$ for a sufficiently large n.

Next define the event

$$A_J = \left\{ \|\hat{V}_{u,(k)}^{-1/2} V_{u,(k)}^{1/2}\| \le J \right\},$$

where J is a positive constant to be defined later. Under **C2**, by the consistency of $\hat{\theta}_t$ and the ULLN, we have $\|\hat{V}_{u,(k)} - V_{u,(k)}\| = o_p(1)$. Under Condition **C6**, it's straightforward to show that there exists a positive constant J such that $P(A_J^c) < \epsilon/2$ for all $\epsilon > 0$, where A_J^c is the complementary set of A_J.

Notice that

$$\begin{aligned}
P(A_n(\zeta)) &\le P(A_n(\zeta) \cap A_J) + P(A_J^c) \\
&\le P\left(\sqrt{n_j} |\hat{V}_{u,(k)}^{-1/2} V_{u,(k)}^{1/2} \tilde{\Psi}_{u,(k)}^*(j)|_\infty > 2^{-1}\sqrt{\zeta \log n_j}, A_J \right) \\
&\quad + P\left(|R_{n(k)}|_\infty > 2^{-1}\sqrt{\zeta \log n_j} \right) + \epsilon/2 \\
&\le P\left(\sqrt{n_j} \|\hat{V}_{u,(k)}^{-1/2} V_{u,(k)}^{1/2}\| |\tilde{\Psi}_{u,(k)}^*(j)| > 2^{-1}\sqrt{\zeta \log n_j}, A_J \right) + \epsilon \\
&\le P\left(\sqrt{n_j} |\tilde{\Psi}_{u,(k)}^*(j)| > (2J)^{-1}\sqrt{\zeta \log n_j} \right) + \epsilon.
\end{aligned}$$

Since $\sqrt{n_j}|\tilde{\Psi}_{u,(k)}^*(j)| = O_p(1)$, it follows that

$$P\left(\sqrt{n_j} |\tilde{\Psi}_{u,(k)}^*(j)| > (2J)^{-1}\sqrt{\zeta \log n_j} \right) \to 0$$

as $n \to \infty$. We then have $P(A_n(\zeta)) \le \epsilon$ for an arbitrary ϵ, from which (A.2) follows.

We shall next prove (A.3). Note that $\Pr(SK_U = 1) = 1 - \sum_{k=2}^M \Pr(SK_U = k)$. Because $SK_U = k$ implies that dimension k "beats "dimension 1 and $\hat{N}_{u,(1)}(j) \ge 0$, we obtain by the definition of SK_U the following

$$\Pr(SK_U = k) \le \Pr(\hat{N}_{u,(k)}(j) - k \log n_j \ge \hat{N}_{u,(1)}(j) - \log n_j) = \Pr(\hat{N}_{u,(k)}(j) \ge (k-1)\log n_j).$$

It follows that

$$\Pr(\hat{N}_{u,(k)}(j) \ge (k-1)\log n_j) \le Q_{1n} + Q_{2n},$$

where $Q_{1n} = \Pr(N_{u,(k)}(j) \ge (k-1)\log n_j/2)$ and $Q_{2n} = \Pr(n_j|\hat{\Psi}_{u,(k)}^*(j) - \tilde{\Psi}_{u,(k)}^*(j)|^2 \ge (k-1)\log n_j/2)$.

Since under the null, $N_{u,(k)}(j)$ converges to a non-degenerate (χ_k^2-distributed) random variable for any k, it's straightforward to show that $Q_{1n} \to 0$ as $n \to \infty$. Because

$\sqrt{n_j}|\hat{\Psi}^*_{u,(k)}(j) - \tilde{\Psi}^*_{u,(k)}(j)| = O_p(1)$, $Q_{2n} \to 0$ as $n \to \infty$. It follows immediately that $\lim_{n\to\infty} \Pr(SK_U = 1) = 1$.

Further note that

$$\Pr(\hat{N}_U(j) \le x) = \Pr(\hat{N}_{u,(1)}(j) \le x) - \Pr(\hat{N}_{u,(1)}(j) \le x, SK_U \ge 2) + \Pr(\hat{N}_U(j) \le x, SK_U \ge 2).$$

Because under the null $\hat{N}_{u,(1)}(j) \xrightarrow{d} \chi^2_1$ and $\Pr(SK_U \ge 2) \xrightarrow{p} 0$ as $n \to \infty$, it follows immediately that $\hat{N}_U(j) \xrightarrow{d} \chi^2_1$.

We now proceed to Theorem 5(b). Define the simplified AIC as follows,

$$AK_U = \min\{k : \hat{N}_{u,(k)}(j) - 2k \ge \hat{N}_{u,(s)}(j) - 2s, 1 \le k, s \le M\}.$$

In order to prove Theorem 5(b), we need to establish that under the alternative distribution $\mathscr{P}$,

$$\lim_{n\to\infty} \Pr(K_U(j) = AK_U) = 1, \tag{A.5}$$

and

$$\Pr(AK_U \ge S) \to 1. \tag{A.6}$$

We denote by $A_n(\zeta)^c$ the complementary set of $A_n(\zeta)$. Under the alternative distribution $\mathscr{P}$, there exists an $S \le M$ such that $E_{\mathscr{P}}[\psi_S(Z_{t-j})] \ne 0$. Denote by S' the corresponding index of ψ_S in the ordered set Ψ_u, where $S' \le M$. It follows that $E_{\mathscr{P}}[\tilde{\Psi}^*_{u,(S')}(j)] \ne 0$. We then have

$$P(A_n(\zeta)^c) \le P\left(\sqrt{n_j}|\hat{\Psi}^*_{u,(S')}(j)| < \sqrt{\zeta \log n_j}\right) \to 0,$$

since $\hat{\Psi}^*_{u,(S')}(j) = \tilde{\Psi}^*_{u,(S')}(j) + o_p(1)$ by the ULLN. Hence, (A.5) holds.

Now, according to the ULLN, for $k = 1, ..., S' - 1$, $|\hat{\Psi}^*_{u,(k)}(j)|^2 \to 0$ and $|\hat{\Psi}^*_{u,(S')}(j)|^2 \to |\tilde{\Psi}^*_{u,(S')}(j)|^2 > 0$. We obtain

$$\begin{aligned} P(AK_U = k) &\le P(\hat{N}_{u,(k)}(j) - 2k \ge \hat{N}_{u,(S')}(j) - 2S') \\ &\le P(n_j|\hat{\Psi}^*_{u,(k)}(j)|^2 \ge 2(k - S') + n_j|\hat{\Psi}^*_{u,(S')}(j)|^2) \to 0. \end{aligned}$$

Therefore, (A.6) also holds.

For a generic constant $C > 0$, we have

$$\Pr(\hat{N}_U(j) \le C) = \Pr(\hat{N}_U(j) \le C, AK_U \ge S') + o(1) \le P(n_j|\hat{\Psi}^*_{u,(S')}|^2 \le C) + o(1) = o(1),$$

where the last equality follows the ULLN. Then, $\hat{N}_U(j) \to \infty$ as $n \to \infty$.

Proof of Theorem 1

Recall that $Z_{0t} = Z_t(\theta_0)$, $\hat{Z}_t = Z_t(\hat{\theta}_t)$, $U_t = G_0(Z_{0t})$, $\hat{G}_n(z) = (n+1)^{-1} \sum_{t=R+1}^{N} I(Z_t(\hat{\theta}_t) \leq z)$ and $G_t(\theta, z) = E[I(Z_t(\theta) \leq z)|\Omega_{t-1}]$. Let $G_n(z) = (n+1)^{-1} \sum_{t=R+1}^{N} I(Z_t(\theta_0) \leq z)$. We first develop some preliminary results that will be used in the proof of Theorem 1.

Lemma A.1. *Under* ***C2***, ***C5*** *and* ***C7***, *we have, uniformly over* $z \in R$

$$\hat{G}_n(z) - G_0(z) = G_n(z) - G_0(z) + h(\theta_0, z) \left[\frac{1}{n} \sum_{t=R+1}^{N} (\hat{\theta}_t - \theta_0) \right] + o_p(n^{-1/2}),$$

where

$$h(\theta_0, z) = E\left[\frac{\partial G_t(\theta_0, z)}{\partial \theta} \right]$$

Proof of Lemma A.1: Define the process

$$\begin{aligned} K_n(c, z) = \frac{1}{\sqrt{n}} \sum_{t=R+1}^{N} \Bigg(& I\left(Z_t\left(\theta_0 + \frac{c}{\sqrt{N}} \right) \leq z \right) \\ & - E\left[I\left(Z_t\left(\theta_0 + \frac{c}{\sqrt{N}} \right) \leq z \right) \Bigg| \Omega_{t-1} \right] \Bigg), \end{aligned}$$

indexed by $(c, z) \in C_K \times \mathbb{R}$, where $C_K = \{c \in \mathbb{R}^q : |c| \leq C\}$, and $C > 0$ is an arbitrary but fixed constant. By (A.1) of Escanciano and Olmo (2010), we can show $\sup_{z \in R} |K_n(\hat{c}, z) - K_n(0, z)| = o_p(1)$ for any $\hat{c} = O_p(1)$. Set $\hat{c} = \max_{R+1 \leq t \leq N} \sqrt{t}(\hat{\theta}_t - \theta_0)$. It follows that,

$$\begin{aligned} \sqrt{n} \sup_z |\hat{G}_n(z) - G_n(z)| &= \sup_z \left| \frac{1}{\sqrt{n}} \sum_{t=R+1}^{N} \{I(\hat{Z}_t \leq z) - I(Z_{0t} \leq z)\} \right| \\ &= \sup_z \left| \frac{1}{\sqrt{n}} \sum_{t=R+1}^{N} \{G_t(\hat{\theta}_t, z) - G_t(\theta_0, z)\} \right| + o_p(1) \qquad \text{(A.7)} \end{aligned}$$

Next by the mean value theorem and interchanging expectation and differentiation, we have

$$\begin{aligned} D_n &:= \sup_z \left| \frac{1}{\sqrt{n}} \sum_{t=R+1}^{N} \left\{ G_t(\hat{\theta}_t, z) - E[G_t(\hat{\theta}_t, z)] - G_t(\theta_0, z) + E[G_t(\theta_0, z)] \right\} \right| \\ &= \sup_z \left| \frac{1}{\sqrt{n}} \sum_{t=R+1}^{N} \left(\frac{\partial G_t(\tilde{\theta}_t, z)}{\partial \theta} - E\left[\frac{\partial G_t(\tilde{\theta}_t, z)}{\partial \theta} \right] \right) (\hat{\theta}_t - \theta_0) \right|, \end{aligned}$$

where $\tilde{\theta}_t$ is between $\hat{\theta}_t$ and θ_0. Note that under **C5** and **C7**, by the uniform law of large numbers (ULLN) of Jennrich (1969, Theorem 2), we can show that $D_n = o_p(1)$. Hence, by (A.7), we have,

$$\sqrt{n}\sup_z |\hat{G}_n(z) - G_n(z)| = \sup_z \left| E\left[\frac{\partial G_t(\theta_0, z)}{\partial \theta}\right] \frac{1}{\sqrt{n}} \sum_{t=R+1}^{N} (\hat{\theta}_t - \theta_0) \right| + o_p(1)$$

This completes the proof of Lemma A.1.

By the mean-value theorem, we expand $\hat{\Psi}_C(j)$ to obtain,

$$\begin{aligned}\sqrt{n_j}\hat{\Psi}_C(j) &= \frac{1}{\sqrt{n_j}} \sum_{t=R+j+1}^{N} \Big\{ \Psi_C(U_{t-j}, U_t) + \Psi_C^{(1)}(\tilde{U}_{t-j}, \tilde{U}_t)[\hat{G}_n(\hat{Z}_{t-j}) - G_0(Z_{0,t-j})], \\ &\quad + \Psi_C^{(2)}(\tilde{U}_{t-j}, \tilde{U}_t)[\hat{G}_n(\hat{Z}_t) - G_0(Z_{0,t})] \Big\} \\ &= A_{1n} + A_{2n} + A_{3n}, \end{aligned} \tag{A.8}$$

where $\tilde{U}_t = \tilde{G}_n(Z_t(\tilde{\theta}))$ is some random value between $\hat{G}_n(\hat{Z}_t)$ and $G_0(Z_{0t})$, $\Psi_C^{(l)}(u_1, u_2) = \partial\Psi_C(u_1, u_2)/\partial u_l$, $l = 1, 2$, and

$$A_{1n} = \frac{1}{\sqrt{n_j}} \sum_{t=R+j+1}^{N} \Psi_C(U_{t-j}, U_t), \tag{A.9}$$

$$\begin{aligned}A_{2n} &= \frac{1}{\sqrt{n_j}} \sum_{t=R+j+1}^{N} \Big\{ \Psi_C^{(1)}(U_{t-j}, U_t)[\hat{G}_n(\hat{Z}_{t-j}) - G_0(Z_{0,t-j})] \Big\} \\ &\quad + \frac{1}{\sqrt{n_j}} \sum_{t=R+j+1}^{N} \Big\{ \Psi_C^{(2)}(U_{t-j}, U_t)[\hat{G}_n(\hat{Z}_t) - G_0(Z_{0t})] \Big\} = B_{1n} + B_{2n} \end{aligned} \tag{A.10}$$

$$\begin{aligned}A_{3n} &= \frac{1}{\sqrt{n_j}} \sum_{t=R+j+1}^{N} \Big\{ [\Psi_C^{(1)}(\tilde{U}_{t-j}, \tilde{U}_t) - \Psi_C^{(1)}(U_{t-j}, U_t)][\hat{G}_n(\hat{Z}_{t-j}) - G_0(Z_{0,t-j})] \\ &\quad + [\Psi_C^{(2)}(\tilde{U}_{t-j}, \tilde{U}_t) - \Psi_C^{(2)}(U_{t-j}, U_t)][\hat{G}_n(\hat{Z}_t) - G_0(Z_{0t})] \Big\}, \end{aligned} \tag{A.11}$$

where B_{1n} and B_{2n} are implicitly defined.

In order to show that $\sqrt{n_j}\hat{\Psi}_C(j) = n_j^{-1/2} \sum_{t=R+j+1}^{N} \Psi_C(U_{t-j}, U_t) + o_p(1)$, it suffices to show that $A_{2n} = o_p(1)$ and $A_{3n} = o_p(1)$.

We first consider A_{2n}. By applying Lemma A.1 to B_{1n} defined in (A.10), we have

$$\begin{aligned} B_{1n} &= \frac{1}{\sqrt{n_j}} \sum_{t=R+j+1}^{N} \Psi_C^{(1)}(U_{t-j}, U_t)[G_n(Z_{0,t-j}) - G_0(Z_{0,t-j})] \\ &\quad + \frac{1}{n_j} \sum_{t=R+j+1}^{N} \Psi_C^{(1)}(U_{t-j}, U_t) h(\theta_0, Z_{0,t-j}) \left[\frac{1}{\sqrt{n_j}} \sum_{s=R+j+1}^{N} (\hat{\theta}_{s-j} - \theta_0) \right] + o_p(1) \\ &= C_{1n} + C_{2n} + o_p(1), \end{aligned} \tag{A.12}$$

where C_{1n} and C_{2n} are implicitly defined.

Under H_{0C}, following the proof of proposition 2.1 of Genest et al. (1995), we can show that

$$C_{1n} = \int_0^1 \int_0^1 \Psi_C^{(1)}(u_1, u_2)[I(U_{t-j} \leq u_1) - u_1] du_1 du_2 + o_p(1). \tag{A.13}$$

Recall that Ψ_C is the tensor product of basis functions, which are orthonormal with respect to uniform distribution. We can rewrite

$$\Psi_C^{(1)}(u_1, u_2) = \Psi_1'(u_1) \circ \Psi_2(u_2) \tag{A.14}$$

where Ψ_1 and Ψ_2 are vectors of basis functions, $\Psi_1'(u) = \partial \Psi_1(u)/\partial u$ and $\circ$ denotes the Hadamard product. We then have

$$C_{1n} = \left[\int_0^1 \Psi_1'(u_1)[I(U_{t-j} \leq u_1) - u_1] du_1 \right] \circ \left[\int_0^1 \Psi_2(u_2) du_2 \right] + o_p(1) = 0 + o_p(1), \tag{A.15}$$

where the last equality holds due to the orthogonality of Ψ_2 with respect to the standard uniform distribution.

Under **C5** and **C7**, by the law of large numbers, we have,

$$C_{2n} = E_Z[\Psi_C^{(1)}(U_{t-j}, U_t) h(\theta_0, Z_{0,t-j})] \left[\frac{1}{\sqrt{n_j}} \sum_{s=R+j+1}^{N} (\hat{\theta}_{s-j} - \theta_0) \right] + o_p(1) \tag{A.16}$$

By the same proof as that for (A.15), we have $E_Z[\Psi_C^{(1)}(U_{t-j}, U_t) h(\theta_0, Z_{0,t-j})] = 0$ since $E[\Psi_2(U)] = 0$, where U is uniformly distributed on $[0,1]$. Thus, $C_{2n} = o_p(1)$. We note here the estimation effect of θ is ignorable because the U_t's are defined based on the ranks of the generalized residuals and therefore are exactly uniformly distributed. For general treatments of rank-based tests with nuisance parameters, see Randles (1984) and de Wet and Randles

(1987). By (A.12), it follows that $B_{1n} = o_p(1)$. Similarly, we can show $B_{2n} = o_p(1)$. Then by (A.10), we have $A_{2n} = o_p(1)$.

Now turn to A_{3n}. By (A.7), it's straightforward to show that $A_{3n} = o_p(1)$.

Therefore, we have $\sqrt{n_j}\hat{\Psi}_C(j) = n_j^{-1/2}\sum_{t=R+j+1}^{N} \Psi_C(U_{t-j}, U_t) + o_p(1)$. Under H_{0C}, it's easy to show that $n_j^{-1/2}\sum_{t=R+j+1}^{N} \Psi_C(U_{t-j}, U_t) \xrightarrow{d} N(0, I_{K_C})$. It follows that $\sqrt{n_j}\hat{\Psi}_C(j) \xrightarrow{d} N(0, I_{K_C})$. This completes the proof of Theorem 1.

Proof of Theorems 2 and 3

The proof of Theorems 2 and 3 are similar to that of Theorem 5 and therefore omitted.

Proof of Theorem 6

We first establish the asymptotic independence between $\hat{Q}_C(j)$ and $\hat{N}_U(j)$ for a given j. Recall that $\hat{\Psi}_C(j) = (\hat{\Psi}_{C,1}(j), ..., \hat{\Psi}_{C,K_C}(j))'$ and $\hat{\Psi}_U(j) = (\hat{\Psi}_{U,1}(j), ..., \hat{\Psi}_{U,K_U}(j))'$. We shall show that $(\sqrt{n_j}\hat{\Psi}_C(j)', \sqrt{n_j}\hat{\Psi}_U(j)')'$ has a multivariate normal distribution and they are uncorrelated. Let

$$X(N) = \sum_{t=R+j+1}^{N} \Psi_C(U_{t-j}, U_t)$$
$$Y(N) = \sum_{t=R+j+1}^{N} \left\{ \Psi_U(Z_{0,t-j}) + E\left[\frac{\partial \Psi_U(Z_{0,t-j})}{\partial \theta}\right] \Xi^{-1} S(t-j) \right\}$$

Following reasoning analogous to that for Theorem 1 and for Theorem 4, under **C5**, we can show that $\sqrt{n_j}\hat{\Psi}_C(j) = n_j^{-1/2}X(N) + o_p(1)$, $\sqrt{n_j}\hat{\Psi}_U(j) = n_j^{-1/2}Y(N) + o_p(1)$ and

$$\begin{bmatrix} n_j^{-1/2}X(N) \\ n_j^{-1/2}Y(N) \end{bmatrix} \xrightarrow{d} N\left(\begin{bmatrix} 0_{K_C\times 1} \\ 0_{K_U\times 1} \end{bmatrix}, \begin{bmatrix} I_{K_C} & \Sigma_{12} \\ \Sigma_{21} & V_U \end{bmatrix} \right).$$

where $0_{k\times l}$ is a k-by-l matrix of zeros, I_k is the identity matrix of size k, I_{K_C}, V_U are covariance matrices of $n_j^{-1/2}X(N)$ and $n_j^{-1/2}Y(N)$, whereas $\Sigma_{12} = \Sigma_{21}'$ is the cross-covariance matrix of $n_j^{-1/2}X(N)$ and $n_j^{-1/2}Y(N)$.

Now we compute Σ_{12}. By applying the results from Lemma 4.1 and Lemma 4.2 of West and McCracken (1998), following the proof of Theorem 4, we can show,

$$\begin{aligned} \Sigma_{12} &= \lim_{n\to\infty} Cov(n_j^{-1/2}X(N), n_j^{-1/2}Y(N)) \\ &= E\{\Psi_C(U_{t-j}, U_t)\Psi_U(Z_{0,t-j})'\} - \eta_1 \Pi_0 \Xi^{-1} \Upsilon', \end{aligned}$$

where η_1 is defined in (A.1) and

$$\Pi_0 = \sum_{k=-\infty}^{\infty} E\left\{\Psi_C(U_{t-j}, U_t)s'_{0,t-j-k}\right\}$$

By using the same notation as (A.14), we rewrite $\Psi_C(U_{t-j}, U_t) = \Psi_1(U_{t-j}) \circ \Psi_2(U_t)$. Therefore, under the null hypothesis that $\{Z_t(\theta_0)\}$ is *i.i.d.* uniform, by applying the law of iterative expectation and using the fact $E[\Psi_2(U_t)] = 0$, we can show that

$$E\{\Psi_C(U_{t-j}, U_t)\Psi_U(Z_{0,t-j})'\} = 0. \tag{A.17}$$

Define $\Pi = E\{\Psi_C(U_{t-j}, U_t)s'_{0,t-j}\}$. Following the same argument as the proof in Theorem 4, we can show that $\Pi_0 = \Pi$. Note that $s_{0,t-j}$ is a function of Ω_{t-j}. Following the same arguments used to derive (A.17), we can show that $\Pi = 0$. Combining with (A.17), we have shown that $\Sigma_{21} = 0_{K_C \times K_U}$. Similarly, we can show that $\Sigma_{21} = 0_{K_U \times K_C}$. So, $(n_j^{-1/2}X(N)', n_j^{-1/2}Y(N)')'$ has a multivariate normal distribution and they are uncorrelated. It follows that $(\sqrt{n_j}\hat{\Psi}_C(j)', \sqrt{n_j}\hat{\Psi}_U(j)')'$ converges in distribution to a multivariate normal distribution and they are uncorrelated. Therefore, $\sqrt{n_j}\hat{\Psi}_C(j)$ and $\sqrt{n_j}\hat{\Psi}_U(j)$ are asymptotically independent of each other. So are $\hat{Q}_C(j)$ and $\hat{N}_U(j)$.

Since this result holds for an arbitrary j, the asymptotic independence between the (automatic) portmanteau test $\hat{W}_Q(p)$ ($\hat{W}_Q(\tilde{p})$) and $\hat{N}_U(p)$ follows immediately. The proof is now finished.

Appendix B. Additional Simulation Results

In this appendix, we report results of some additional simulations suggested by a referee.

First we consider the case of $R/n = 1/2$. The results under the fixed estimation scheme are reported in Table B.1. We set the out-of-sample evaluation period $n = 100$, 250, 500 and 1000. The overall size and power performances are satisfactory and improve with sample size.

Table B.1: Empirical sizes and powers for $R/n = 1/2$ (fixed estimation scheme)

DGP NULL		RW-N RW-N	GARCH-N GARCH-N	RW-T RW-N	GARCH-N RW-N	RS-T RW-N
$n = 100$	Q(1)	4.1	8.5	34.2	16.1	31.1
	W(5)	3.5	7.5	34.1	19.8	32.2
	W(10)	4.3	6.5	34	18.4	31.8
	W(20)	4.2	6.1	34.2	17.2	31.3
	$W_Q(\tilde{p})$	4.1	6	34	16.2	30.5
$n = 250$	Q(1)	4	8.5	75.9	40.2	72.9
	W(5)	4.5	8.4	76.2	46.9	73.5
	W(10)	4.6	8	75.9	45.4	73.4
	W(20)	4.5	7.6	76.1	41.3	73.1
	$W_Q(\tilde{p})$	4.7	7.6	75.9	43.6	73
$n = 500$	Q(1)	4.8	8.1	96.2	66.2	95.8
	W(5)	4.7	8.1	96.2	75.9	95.9
	W(10)	4.8	9	96.1	74	95.8
	W(20)	4.8	7.6	96.1	67.4	95.6
	$W_Q(\tilde{p})$	4.4	8.7	96.2	73.8	95.8
$n = 1000$	Q(1)	5	8.9	99.97	92	99.97
	W(5)	4.5	9.3	99.97	96.5	99.97
	W(10)	5	9	99.97	96.1	99.97
	W(20)	4.3	7.9	99.97	93.6	99.97
	$W_Q(\tilde{p})$	4.3	8.7	99.97	95.5	99.97

Note: This table reports the empirical sizes and powers of the sequential tests (SQT) with $\hat{Q}_C(1)$, $\hat{W}_Q(p)$, $p = 5, 10, 20$ or $\hat{W}_Q(\tilde{p})$ in the first stage. The nominal size is 0.05. R/n denotes the estimation-evaluation ratio. Results are based on $3,000$ replications.

Second, we examine the empirical sizes of SQT under the rolling scheme. As shown in Table B.2, the tests have better overall size under the rolling scheme compared with those under the fixed scheme.

Table B.2: Empirical Sizes of the sequential tests under the rolling estimation scheme

DGP		n=250					n=500					n=1000				
		$Q(1)$	$W(5)$	$W(10)$	$W(20)$	$W(\tilde{p})$	$Q(1)$	$W(5)$	$W(10)$	$W(20)$	$W(\tilde{p})$	$Q(1)$	$W(5)$	$W(10)$	$W(20)$	$W(\tilde{p})$
RW-N	R/n=1	5.5	5.8	5.4	5.2	5.6	5.5	5.3	5.5	5.9	5.8	5.7	5.9	5.8	5.7	5.4
	R/n=2	5.3	5.4	5.2	5.4	5.3	4.9	5.1	4.8	5.3	5.2	5.2	4.8	5	5.1	4.7
	R/n=3	5.5	5.5	5.6	6	5.9	5.2	5.2	5.8	5.9	5.6	4.6	4.9	4.8	4.8	5.5
GARCH-N	R/n=1	7	7.7	7.9	8.1	7.6	5.4	5.3	5.6	5.6	5.7	4.9	5.2	5.5	5.7	5.3
	R/n=2	5.1	4.9	4.9	5.2	5.1	5.2	4.8	5.6	5.7	5.4	4.6	5	4.9	5.4	4.8
	R/n=3	5.7	5.2	5.1	6.1	5.3	5.5	5.5	5.6	5.8	5.8	5.2	5.2	5.6	5.8	4.7

Note: This table reports the empirical sizes of the sequential tests(SQT) under the rolling scheme with $\hat{Q}_C(1)$, $\hat{W}_Q(p)$, $p = 5, 10, 20$ or $\hat{W}_Q(\tilde{p})$ in the first stage. The nominal size is 0.05. R/n denotes the estimation-evaluation ratio. Results are based on $3,000$ replications.