CAEPR Working Paper
#023-2009

Nonparametric Bootstrap Tests for Independence of Generalized Errors

Zaichao Du

Indiana University
December 6, 2009

This paper can be downloaded without charge from the Social Science Research Network electronic library at: <http://ssrn.com/abstract=1524850>.

The Center for Applied Economics and Policy Research resides in the Department of Economics at Indiana University Bloomington. CAEPR can be found on the Internet at: <http://www.indiana.edu/~caepr>. CAEPR can be reached via email at caepr@indiana.edu or via phone at 812-855-4050.

©2008 by NAME. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Nonparametric Bootstrap Tests for Independence of Generalized Errors

ZAICHAO DU*

Indiana University

December 6, 2009

Abstract

In this paper, we develop a general method of testing for independence when unobservable generalized errors are involved. Our method can be applied to testing for serial independence of generalized errors, and testing for independence between the generalized errors and observable covariates. The former can serve as a unified approach to testing adequacy of time series models, as model adequacy often implies that the generalized errors obtained after a suitable transformation are independent and identically distributed. The latter is a key identification assumption in many nonlinear economic models. Our tests are based on a classical sample dependence measure, the Hoeffding-Blum-Kiefer-Rosenblatt-type empirical process applied to generalized residuals. We establish a uniform expansion of the process, thereby deriving an explicit expression for the parameter estimation effect, which causes our tests not to be nuisance parameter-free. To circumvent this problem, we propose a multiplier-type bootstrap to approximate the limit distribution. Our bootstrap procedure is computationally very simple as it does not require a reestimation of the parameters in each bootstrap replication. In a simulation study, we apply our method to test the adequacy of ARMA-GARCH and Hansen (1994) skewed t models, and document a good finite sample performance of our test. Finally, an empirical application to some daily exchange rate data highlights the merits of our approach.

Keywords and Phrases: Independence; Generalized errors; Goodness-of-Fit; Identification; Empirical process; Parameter estimation uncertainty; Bootstrap; GARCH; Skewed t distribution.

*I am greatly indebted to my advisor Prof. Juan Carlos Escanciano for his numerous guidance and help. I also thank Prof. David Jacho-Chavez, Dr. Konstantin Tyurin, as well as seminar participants at Indiana University and 2009 Midwest Econometric Group Meeting for helpful comments. All the errors are my own.

1. INTRODUCTION

Testing for independence continues to be a major interest of time series analysis. Not only do many statistical inferences, including many asymptotic results, rely on the independence assumption, it is also a key assumption for identification in many economic models. Therefore, it is important to develop specification tests for the independence assumption. So far there have been plenty of researches on testing independence, but most of them have focused on raw data but not on unobservable generalized errors, which hampers the application of these existing procedures to many important situations in econometrics (see the examples below). In this paper, we develop a general method of testing for independence when generalized errors are involved, thereby considerably broadening the scope of possible applications of the existing methods.

Before we present our motivational examples, let us introduce some general notations. Suppose that we observe data $\{Y_t, \Omega_{t-1}\}_{t=1}^n$, where Y_t is a real-valued dependent variable and Ω_{t-1} may contain lagged values of Y_t as well as current and lagged values of other exogenous variables, say X_t . Furthermore, assume that θ_0 is some unknown parameter in a compact set $\Theta \subset \mathbb{R}^p$.

Example 1. SERIAL INDEPENDENCE OF ERRORS IN LOCATION-SCALE MODELS

Our leading motivational example is the conditional mean and variance (or location-scale) model,

$$Y_t = \mu(\Omega_{t-1}, \theta_0) + \sigma(\Omega_{t-1}, \theta_0)u_t,$$

where $\mu(\Omega_{t-1}, \theta_0) = E[Y_t \mid \Omega_{t-1}]$ and $\sigma^2(\Omega_{t-1}, \theta_0) = \text{Var}[Y_t \mid \Omega_{t-1}]$ almost surely (a.s.). This type of models, with the popular ARMA-GARCH model as a special case, are widely used in financial applications, for example, in modeling stock returns, term structure of interest rates, etc. In these models, one usually makes the *iid* assumption for the standardized errors

$$u_t = u_t(\theta_0) = \frac{Y_t - \mu(\Omega_{t-1}, \theta_0)}{\sigma(\Omega_{t-1}, \theta_0)}.$$

The *iid* assumption on u_t is a crucial assumption for the validity of inferences about the parameter θ_0 and for specification tests of the parametric mean and variance. In particular, the *iid* assumption often implies that the Value at Risk (*VaR*) model at level α can be expressed as

$$\text{VaR}_\alpha = \mu(\Omega_{t-1}, \theta_0) + \sigma(\Omega_{t-1}, \theta_0)F_u^{-1}(\alpha),$$

where $F_u^{-1}(\alpha)$ is the quantile function of u_t . *VaR* models play an important role in the assessment of market risk at commercial banks and other financial institutions.

Example 2. CONDITIONAL GOODNESS-OF-FIT

A (Conditional) Goodness-of-Fit (GOF) test problem arises when one assumes that the conditional distribution of $Y_t|\Omega_{t-1}$ has a certain parametric form $H(\cdot, \Omega_{t-1}, \theta_0)$. One way to construct a test is to use the well-known fact that for a continuous and correctly specified H , the generalized error

$$u_t = u_t(\theta_0) \equiv H(Y_t, \Omega_{t-1}, \theta_0)$$

has an *iid* $U[0, 1]$ distribution. Bai (2003) uses this idea to test the conditional distributions of dynamic models, although his test is tailored only for the $U[0, 1]$ distributed part, not for the *iid* part. Hong and Li (2005) also use this idea to construct a specification test for the term structure of interest rates. Along these lines, our proposed test for *iid* of generalized errors can be used as an alternative to existing conditional Goodness-of-Fit tests, see Section 5 and 6 for further details.

Example 3. SPECIFICATION TEST FOR CONTINUOUS TIME MODELS

Many economic and financial models can be specified in continuous time as

$$dY_t = \mu(X_t, \theta_0)dt + dU_t,$$

where (Y_t) and (X_t) are stochastic processes, $(\mathcal{F}_t)$ is a filtration to which both (Y_t) and (X_t) are adapted; $\mu(X_t, \theta_0)dt = E(dY_t|\mathcal{F}_t)$ and (U_t) is a martingale with respect to the filtration $(\mathcal{F}_t)$. If we apply a time change (T_t) ,

$$T_t = \inf_{s>0} \{\langle U \rangle_s > t\},$$

where $(\langle U \rangle_t)$ is the quadratic variation of (U_t) , then the DDS (Dambis, Dubins-Schwarz) theorem implies that

$$u_t(\theta_0) = \Delta^{-1/2} \left(Y_{T_{t\Delta}} - Y_{T_{(t-1)\Delta}} - \int_{T_{(t-1)\Delta}}^{T_{t\Delta}} \mu(X_t, \theta_0)dt \right), \quad t = 1, \dots, n,$$

are *iid* normal. Park and Vasudev (2006) uses this idea to develop a test for martingale in continuous time. In a similar way, our test for *iid* of generalized errors can serve as a specification test for continuous time models.

Example 4. IDENTIFICATION IN NONLINEAR MODELS

Many nonlinear economic models have the form

$$u_t(\theta_0) = H(Y_t, Z_t, \theta_0),$$

where $Z_t \subset \Omega_{t-1}$ is a k -dimensional vector, and H is a known function up to the parameters θ_0 . In these models, one often requires $u_t(\theta_0)$ to be independent of Z_t for identification, see e.g. Brown

(1983) and Brown and Wegkamp (2002). Brown et al. (2007) develop a test for independence between $u_t(\theta_0)$ and Z_t in a cross-section setup. Our results generalize theirs to time series models.

Motivated by the above examples, in this paper we develop a general method of testing for independence when unobservable generalized errors are involved. Our method can be applied to testing the null hypothesis

$$H_0 : \{u_t(\theta_0)\}_{t \in \mathbb{Z}} \text{ is a sequence of iid r.v, for some } \theta_0 \in \Theta \subset \mathbb{R}^p, \quad (1)$$

as well as

$$H'_0 : u_t(\theta_0) \text{ is independent of } Z_t \text{ for some } \theta_0 \in \Theta \subset \mathbb{R}^p, \quad (1')$$

where $u_t(\theta_0) \equiv H(Y_t, \Omega_{t-1}, \theta_0)$ is a generalized error, given by a known measurable transformation H and an unknown parameter θ_0 in $\Theta \subset \mathbb{R}^p$, and $Z_t \subset \Omega_{t-1}$ is a k -dimensional vector. The alternative hypotheses are the negation of the null (1) and (1'), respectively.

Notice that the null hypotheses (1) and (1') are composite, since we do not know θ_0 . Given a $\sqrt{n}$ -consistent estimator of θ_0 , say $\hat{\theta}_n$, we can construct generalized residuals $\hat{u}_t = u_t(\hat{\theta}_n)$, and base our test on $\hat{u}_t$, but then, the null limit distribution of the test will generally lose the “nuisance parameter-free” property, which is a tenet of the empirical process theory when parameters are estimated, see Durbin (1973) for example. To circumvent this problem, three main approaches have been proposed. One is using martingale transformations (due to Khmaladze 1981) to get an asymptotically distribution-free test, see e.g. Bai (2003) in a conditional GOF setup. Another approach is using bootstrap to overcome the problem of the data-dependent asymptotic distributions, see Corradi and Swanson (2003) and Hidalgo and Zaffaroni (2007) in conditional GOF setups, and Brown et al. (2007) in a cross-section setup analogous to Example 4. The third approach comprises nonparametric smoothing methods, which involve bandwidth selection, see Hong and Lee (2003) in a setup similar to Example 1. In this paper we propose to use a multiplier-type bootstrap¹ to solve the aforementioned problem in a more general setup than the existing approaches.

Our tests are constructed upon a classical sample dependence measure, the Hoeffding-Blum-Kiefer-Rosenblatt-type empirical process applied to residuals $\hat{u}_t$. We establish a uniform expansion of the process, thereby deriving an explicit expression for the parameter estimation effect, which causes our tests not to be nuisance parameter-free. To circumvent this problem, we propose a multiplier-type bootstrap to approximate the limit distribution, and our bootstrap procedure is computationally very simple as it does not require a reestimation of the parameters in each bootstrap replication.

Our proposed test for (1) has a similar form to Box-Ljung-Pierce test (Box and Pierce (1970) and Ljung and Box (1978), BLP hereinafter) except that our test is based on the sample dependence

¹We call it a multiplier-type bootstrap since the theoretical basis for this is a multiplier central limit theorem, see van der Vaart and Wellner (1996, p.176)

measure mentioned above instead of the sample autocorrelations. As a result, our test is powerful against any type of pairwise dependence. Furthermore, our test has a wide variety of applications. When applied to testing for serial independence of errors in location-scale models, our test can serve as an alternative to Hong and Lee (2003), as it does not require selection of bandwidth parameters. When used for conditional goodness-of-fit purpose as explained in Example 2, our test can be viewed as a complement to Bai (2003) as we test both for the *iid* and $U[0, 1]$ distribution assumptions of the generalized errors, and as an alternative to Hong and Li (2005) as it does not involve any bandwidth selection. Moreover, our test can be used to test for Student- t distribution with unknown degrees of freedom, as well as Hansen's (1994) skewed t distribution.

Our test for (1') is a generalization of Brown et al. (2007) to time series models. Brown et al. (2007) consider the test problem (1') in a cross-section setup. Their test is constructed on a similar dependence measure to ours, and they use a version of Efron's (1979) bootstrap to deal with the parameter estimation effect. Here we extend their test to time series models, and our bootstrap procedure is computationally simpler than theirs.

In a simulation study, we apply our method to test the conditional goodness-of-fit of ARMA-GARCH and Hansen (1994) skewed t models, and we find that our test has empirical sizes close to the nominal levels, and reasonable power against a wide variety of alternatives to the null models. Finally, an empirical application to some daily exchange rate data highlights the merits of our approach.

The rest of the paper is organized as follows. In Section 2 we establish a uniform expansion of the Hoeffding-Blum-Kiefer-Rosenblatt-type empirical process applied to residuals and show some of its applications. In Section 3 we introduce our test statistics and derive their limit distributions. In Section 4 we propose a multiplier-type bootstrap to approximate the data-dependent asymptotic distributions. In Section 5 we do some Monte Carlo simulations to study the performance of our proposed tests. In Section 6 we apply our tests to real exchange rate data. In Section 7 we conclude and suggest some directions for future research. The proofs are gathered in the Appendix.

In the sequel, we simplify the notations as follows: $F_t(\theta, x) \equiv \Pr[u_t(\theta) \leq x | \Omega_{t-1}]$, $u_t \equiv u_t(\theta_0)$, $\hat{u}_t \equiv u_t(\hat{\theta}_n)$, where $\hat{\theta}_n$ is a $\sqrt{n}$ -consistent estimator for θ_0 , and $\hat{F}_n(x) = 1/n \sum_{t=1}^n I(\hat{u}_t \leq x)$. Furthermore, let $F(x)$ and $F_j(x, y) := \Pr(u_t \leq x, u_{t-j} \leq y)$ denote the marginal and joint distribution functions of (u_t, u_{t-j}) , respectively; let f be the density function of u_t . Finally, Let $I(A)$ denote the indicator function for a set A , and let $\|\cdot\|$ be the Euclidean norm.

2. AN ASYMPTOTIC UNIFORM EXPANSION

Our tests are based on a classical sample dependence measure, the Hoeffding-Blum-Kiefer-Rosenblatt-type empirical process applied to residuals. In this section we establish a uniform expansion of the process, thereby deriving an explicit expression for the parameter estimation effect.

2.1 Serial Dependence Measure

Our test for the null hypothesis (1) is based on the following dependence measure

$$\gamma_j(x, y) = \text{Cov}[I(u_t \leq x), I(u_{t-j} \leq y)] = F_j(x, y) - F(x)F(y),$$

which was proposed by Hoeffding (1948). Unlike correlation that only measures linear dependence, $\gamma_j(x, y)$ captures all types of pairwise dependencies. The sample counterpart of $\gamma_j(x, y)$ based on a sample $\{u_t\}_{t=1}^n$ is

$$\gamma_{nj}(x, y) = \frac{1}{n-j} \sum_{t=1+j}^n I(u_t \leq x) I(u_{t-j} \leq y) - \frac{1}{(n-j)^2} \left\{ \sum_{t=1+j}^n I(u_t \leq x) \right\} \left\{ \sum_{t=1+j}^n I(u_{t-j} \leq y) \right\}$$

Notice that under the null hypothesis (1),

$$\gamma_j(x, y) = 0 \quad \forall j \geq 1, \forall (x, y) \in \mathbb{R}^2,$$

hence we can test (1) based on the distance between γ_{nj} and 0. Hoeffding (1948) used this idea to test the independence between two *iid* random variables. Along the same line, Hong (2000) proposed a generalized spectral test for serial independence of observable stationary time series. However, in our present context $\{u_t\}_{t=1}^n$ is unobservable as θ_0 is unknown. Then we substitute $\hat{u}_t$ for u_t in γ_{nj} and obtain

$$\hat{\gamma}_{nj}(x, y) = \frac{1}{n-j} \sum_{t=1+j}^n I(\hat{u}_t \leq x) I(\hat{u}_{t-j} \leq y) - \frac{1}{(n-j)^2} \left\{ \sum_{t=1+j}^n I(\hat{u}_t \leq x) \right\} \left\{ \sum_{t=1+j}^n I(\hat{u}_{t-j} \leq y) \right\}, \quad (2)$$

but the asymptotic distribution of $\hat{\gamma}_{nj}(x, y)$ would generally be different to that of $\gamma_{nj}(x, y)$. The next theorem gives an explicit expression for the difference under the following conditions:

Assumption A1: $\{Y_t, \Omega_{t-1}\}_{t=1}^n$ is strictly stationary and ergodic.

Assumption A2: $F_t(\theta, x)$ is continuously differentiable in θ and x (a.s.), $E\left[\sup_{\theta \in \Theta, x \in \mathbb{R}} \left\| \frac{\partial F_t(\theta, x)}{\partial \theta} \right\| \right] < C$

and $E\left[\sup_{\theta \in \Theta, x \in \mathbb{R}} \left\| \frac{\partial F_t(\theta, x)}{\partial x} \right\| \right] < C$.

Assumption A3: There is a $\sqrt{n}$ -consistent estimator $\hat{\theta}_n$ of the pseudo-parameter θ_* , where θ_* is a unique interior point of Θ . Under H_0 , $\theta_* = \theta_0$.

Assumption A4: $n^{-1} \sum_{t=1}^n \left| H(Y_t, \hat{\Omega}_{t-1}, \hat{\theta}_n) - H(Y_t, \Omega_{t-1}, \theta_0) \right| = o_P(1)$, where $\hat{\Omega}_{t-1}$ is the observed information set at time $t - 1$.

Assumption A1 is made here for easy exposition. Our results are also valid for some non-stationary and non-ergodic sequences, see Escanciano (2007) for details. Assumption A2 is required for the asymptotic tightness² of the empirical processes and the uniform law of large numbers. Assumption A3 is satisfied under mild conditions. Assumption A4 is on the effect of information truncation due to the unavailability of the infinite history of observations, and it easily holds for many time series models, including stationary and invertible ARMA processes, GARCH processes etc., see e.g. the discussions in Bai (2003) and Hong and Lee (2003). With these assumptions in place we are in position to establish the first important result of the paper.

Theorem 1 *Under Assumptions A1-A4 and H_0 , we have*

$$\sup_{(x,y) \in \mathbb{R}^2} \left| \sqrt{n-j} [\hat{\gamma}_{nj}(x,y) - \gamma_{nj}(x,y)] - \sqrt{n} (\hat{\theta}_n - \theta_0)' E_j(x,y) \right| = o_P(1)$$

where

$$E_j(x,y) := E \left\{ \frac{\partial F_t(\theta_0, x)}{\partial \theta} [I(u_{t-j} \leq y) - F(y)] \right\}. \quad (3)$$

Due to the presence of this additional term $\sqrt{n}(\hat{\theta}_n - \theta_0)' E_j(x,y)$, what we call *estimation effect*, $\hat{\gamma}_{nj}(x,y)$ would not have the same asymptotic distribution as $\gamma_{nj}(x,y)$ in general. Therefore, unlike tests based on $\gamma_{nj}(x,y)$, tests based on $\hat{\gamma}_{nj}(x,y)$ are no longer asymptotically distribution-free. However, there are some special cases where the estimation effect vanishes, see Du and Escanciano (2008) for details.

Next we derive expressions for the estimation effects in our motivational examples by applying Theorem 1.

Example 1. (Continued) Testing for Serial Independence of Errors in Location-Scale

Models: $Y_t = \mu(\Omega_{t-1}, \theta_0) + \sigma(\Omega_{t-1}, \theta_0)u_t$.

For these models, we want to test

$$H_0 : u_t = u_t(\theta_0) = \frac{Y_t - \mu(\Omega_{t-1}, \theta_0)}{\sigma(\Omega_{t-1}, \theta_0)} \text{ are iid for some } \theta_0 \in \Theta \subset \mathbb{R}^p.$$

²For the definition of asymptotic tightness, see van der Vaart and Wellner (1996).

Some algebra shows that in this setup

$$\frac{\partial F_t(\theta_0, x)}{\partial \theta} = f(x) \frac{\dot{\mu}_t + x \dot{\sigma}_t}{\sigma_t},$$

where $\sigma_t = \sigma(\Omega_{t-1}, \theta_0)$, $\dot{\mu}_t = \frac{\partial \mu(\Omega_{t-1}, \theta_0)}{\partial \theta}$ and $\dot{\sigma}_t = \frac{\partial \sigma(\Omega_{t-1}, \theta_0)}{\partial \theta}$. Plug the above display into (3), we get

$$E_j(x, y) = f(x) \cdot E \left\{ \frac{\dot{\mu}_t + x \dot{\sigma}_t}{\sigma_t} [I(u_{t-j} \leq y) - F(y)] \right\} \quad (4)$$

Example 2. (Continued) Conditional Goodness-of-Fit.

One way to test that $Y_t | \Omega_{t-1}$ follows a continuous conditional distribution $H(\cdot, \Omega_{t-1}, \theta_0)$ is to test the following hypothesis

$$H_0 : u_t = u_t(\theta_0) = H(Y_t, \Omega_{t-1}, \theta_0) \text{ is iid } U[0, 1] \text{ distributed.}$$

To test for u_t being both *iid* and $U[0, 1]$ distributed, we accordingly adjust the sample dependence measure to

$$\gamma_{nj}(x, y) = \frac{1}{n-j} \sum_{t=1+j}^n (I(u_t \leq x) - x) I(u_{t-j} \leq y), \quad (5)$$

whose feasible counterpart is

$$\hat{\gamma}_{nj}(x, y) = \frac{1}{n-j} \sum_{t=1+j}^n (I(\hat{u}_t \leq x) - x) I(\hat{u}_{t-j} \leq y). \quad (6)$$

Theorem 1 now holds with

$$E_j(x, y) = E \left\{ \frac{\partial F_t(\theta_0, x)}{\partial \theta} I(u_{t-j} \leq y) \right\}, \quad (7)$$

where

$$\frac{\partial F_t(\theta_0, x)}{\partial \theta} = h(H^{-1}(x, \Omega_{t-1}, \theta_0), \Omega_{t-1}, \theta_0) \frac{\partial H^{-1}(x, \Omega_{t-1}, \theta_0)}{\partial \theta}, \quad (8)$$

and $h(\cdot, \Omega_{t-1}, \theta_0)$ is the density function corresponding to $H(\cdot, \Omega_{t-1}, \theta_0)$. Here we require that H^{-1} exists and is differentiable with respect to θ .

2.2 Dependence Measure between Errors and Covariates

The dependence measure that we use for testing (1') is as the one in the previous subsection, with Z_t replacing u_{t-j} . To be specific, our sample dependence measure in this case is

$$\gamma_n(x, z) = \frac{1}{n} \sum_{t=1}^n I(u_t \leq x) I(Z_t \leq z) - \frac{1}{n^2} \left\{ \sum_{t=1}^n I(u_t \leq x) \right\} \left\{ \sum_{t=1}^n I(Z_t \leq z) \right\}.$$

Substitute the unobservable u_t with $\hat{u}_t$, and we get

$$\hat{\gamma}_n(x, z) = \frac{1}{n} \sum_{t=1}^n I(\hat{u}_t \leq x) I(Z_t \leq z) - \frac{1}{n^2} \left\{ \sum_{t=1}^n I(\hat{u}_t \leq x) \right\} \left\{ \sum_{t=1}^n I(Z_t \leq z) \right\}. \quad (9)$$

As in Theorem 1, we establish a uniform expansion of the above process, thereby deriving an explicit expression for the parameter estimation effect.

Theorem 2 *Under Assumptions A1-A4 and the null hypothesis (1'), we have*

$$\sup_{(x,z) \in \mathbb{R} \times \mathbb{R}^k} \left| \sqrt{n}[\hat{\gamma}_n(x, z) - \gamma_n(x, z)] - \sqrt{n}(\hat{\theta}_n - \theta_0)' E(x, z) \right| = o_p(1)$$

where

$$\begin{aligned} E(x, z) &: = E \left\{ \frac{\partial F_t(\theta_0, x)}{\partial \theta} [I(Z_t \leq z) - F_Z(z)] \right\}, \\ \frac{\partial F_t(\theta_0, x)}{\partial \theta} &= f(x) \frac{\partial H(H^{-1}(x, Z_t, \theta), Z_t, \theta_0)}{\partial \theta} \Big|_{\theta=\theta_0}, \end{aligned} \quad (10)$$

and F_Z is the CDF of Z_t . Here we require that H^{-1} exists and is differentiable with respect to θ .

3. TEST STATISTICS

3.1 Test for Serial Independence of Generalized Errors

Our proposed test for the null hypothesis (1) is based on the sample dependence measure $\hat{\gamma}_{nj}$ defined in (2). Specifically, the test statistic is given as follows

$$Q(m) := \sum_{j=1}^m (n-j) \hat{\gamma}_n^2(j) w_j, \quad (11)$$

where

$$\hat{\gamma}_n^2(j) = \int_{\mathbb{R}^2} \hat{\gamma}_{nj}^2(x, y) d\hat{F}_n(x) d\hat{F}_n(y) = n^{-2} \sum_{r=1}^n \sum_{s=1}^n \hat{\gamma}_{nj}^2(\hat{u}_r, \hat{u}_s);$$

m is the number of lags considered, and w_j is some weighting scheme.

Here we can actually consider all the possible lags by letting $m = n - 1$ and $w_j = 1/(j\pi)^2$, and then our test is powerful against any type of pairwise dependence. The resulting test can be interpreted as a generalized spectral test, see Hong (2000) for details. For the choice of $w_j = 1$ and m fixed, our test $Q(m)$ has a similar form to the Box-Ljung-Pierce test, except that the former is based on the sample dependence measure $\hat{\gamma}_n^2(j)$ instead of the sample autocorrelations of $\hat{u}_t$. The above two choices for w_j are what we use in our simulations. Generally, how to choose the optimal w_j to maximize the power of the test against certain alternatives is an issue we are going to address in future research.

The next theorem establishes the null limit distribution of (11), for which we need the following additional assumption

Assumption A5: $\hat{\theta}_n$ satisfies the asymptotic expansion,

$$\sqrt{n}(\hat{\theta}_n - \theta_*) = \frac{1}{\sqrt{n}} \sum_{t=1}^n l(Y_t, \Omega_{t-1}, \theta_*) + o_p(1),$$

where l is such that $E[l(Y_t, \Omega_{t-1}, \theta_*)] = 0$ and $E[l(Y_t, \Omega_{t-1}, \theta_*)l'(Y_t, \Omega_{t-1}, \theta_*)]$ exists. Under H_0 , $\theta_* = \theta_0$.

Assumption A5 is satisfied by most $\sqrt{n}$ -consistent estimators, such as the quasi-maximum likelihood and the generalized method of moments estimators.

Theorem 3 Under Assumptions A1-A5 and H_0 ,

$$Q(m) \xrightarrow{d} Q_\infty(m) = \int_{[0,1] \times \mathbb{R}^2} (S_\infty^m(\lambda, x, y))^2 d\lambda dF(x) dF(y),$$

where $S_\infty^m(\cdot)$ is a Gaussian process centered at 0 with covariance structure

$$E[S_\infty^m(\lambda_1, x_1, y_1)S_\infty^m(\lambda_2, x_2, y_2)] = 2 \sum_{j=1}^m \sum_{k=1}^m E(U_{t,j}(x_1, y_1)U_{t,k}(x_2, y_2)) \sqrt{w_j w_k} \sin j\pi\lambda_1 \sin k\pi\lambda_2,$$

where $U_{t,j}(x, y) = (I(u_t \leq x) - F(x))(I(u_{t-j} \leq y) - F(y)) + E_j(x, y)'l(Y_t, \Omega_{t-1}, \theta_0)$, and $E_j(x, y)$ is given in (3).

If the estimation effect vanishes, i.e. $E_j(x, y) = 0$, then $S_\infty^m(\cdot)$ will be a two-parameter Brownian bridge, and $Q_\infty(m)$ will be distribution-free. Now with the E_j term, the limit distribution of $Q(m)$ depends on the unknown data-generating process and therefore cannot be tabulated in general. We propose a multiplier-type bootstrap to overcome this problem in Section 4.

Example 2. (Continued) Conditional Goodness-of-Fit Test.

Our proposed test $Q(m)$ can be applied directly to Example 1, and in this subsection we apply a slightly modified version of it to Example 2, conditional goodness-of-fit test. The test includes testing *iid* normal as a special case, therefore it can be used for testing the correct specification of continuous time models as explained in Example 3.

Here we want to test that $Y_t|\Omega_{t-1}$ follows a continuous conditional distribution $H(\cdot, \Omega_{t-1}, \theta_0)$ by testing the hypothesis

$$H_0 : u_t = u_t(\theta_0) = H(Y_t, \Omega_{t-1}, \theta_0) \text{ is iid } U[0, 1] \text{ distributed.}$$

Bai's (2003) test uses this same idea, but his test is tailored only for the $U[0, 1]$ distributed part, not for the *iid* part. Our proposed test here complements Bai (2003) as we test both for the *iid* and $U[0, 1]$ distribution of u_t .

Our test statistic is still $Q(m)$ defined in (11), now with $\hat{\gamma}_{nj}(\cdot, \cdot)$ given by (6) as we also test for u_t being $U[0, 1]$ distributed. Apply Theorem 3 to this setup, and we get the same limit distribution $Q_\infty(m)$, except now $U_{t,j}(x, y) = (I(u_t \leq x) - x)I(u_{t-j} \leq y) + E_j(x, y)'l(Y_t, \Omega_{t-1}, \theta_0)$ with $E_j(x, y)$ given in (7).

3.2 Test for Independence between Errors and Covariates

Our test statistic for (1') is given by

$$Q' := n\hat{\gamma}_n^2, \quad (12)$$

with

$$\hat{\gamma}_n^2 = \int_{\mathbb{R} \times \mathbb{R}^k} \hat{\gamma}_n^2(x, z) d\hat{F}_n(x) dF_n(z) = n^{-2} \sum_{r=1}^n \sum_{s=1}^n \hat{\gamma}_n^2(\hat{u}_r, Z_s),$$

where $\hat{\gamma}_n(x, z)$ is defined in (9), and $F_n(z) = 1/n \sum_{t=1}^n I(Z_t \leq z)$. The next Theorem gives the null limit distribution of (12).

Theorem 4 *Under Assumptions A1-A5 and H'_0 ,*

$$Q' \xrightarrow{d} Q'_\infty = \int_{\mathbb{R} \times \mathbb{R}^k} S_\infty^2(x, z) dF(x) dF_Z(z),$$

where $S_\infty(\cdot)$ is a Gaussian process centered at 0 with covariance structure

$$E[S_\infty(x_1, z_1) S_\infty(x_2, z_2)] = E(U_t(x_1, z_1) U_t(x_2, z_2)),$$

$U_t(x, z) = (I(u_t \leq x) - F(x))(I(Z_t \leq z) - F_Z(z)) + E(x, z)' l(Y_t, \Omega_{t-1}, \theta_0)$, and $E(x, z)$ is given in (10).

Again, the limit distribution Q'_∞ is generally not distribution-free because of the $E(x, z)$ term introduced by parameter estimation effect.

4. A MULTIPLIER-TYPE BOOTSTRAP

In this section, we propose a multiplier-type bootstrap to overcome the problem of the data-dependent asymptotic distributions of our test statistics. For easy exposition, we demonstrate our bootstrap procedure using Example 2. The procedure can be easily adapted to other cases.

Before we state our bootstrap procedure, we introduce some further notations. First recall that for Example 2, γ_{nj} , $\hat{\gamma}_{nj}$ and E_j are given by (5), (6) and (7), respectively. Next, let

$$\begin{aligned} l_{1jt}(x, y) &= E_j(x, y)' l(Y_t, \Omega_{t-1}, \theta_0), \\ l_{0jt}(x, y) &= (I(u_t \leq x) - x) I(u_{t-j} \leq y), \end{aligned}$$

and then $\sqrt{n-j} \hat{\gamma}_{nj}(x, y) = 1/\sqrt{n-j} \sum_{t=1+j}^n l_{0jt}(x, y)$. By Theorem 1 and Assumption A5 we have

$$\sqrt{n-j} \hat{\gamma}_{nj}(x, y) = \frac{1}{\sqrt{n-j}} \sum_{t=1+j}^n l_{0jt}(x, y) + \frac{1}{\sqrt{n}} \sum_{t=1}^n l_{1jt}(x, y) + o_p(1).$$

Furthermore, let $\widehat{l}_{0jt}(x, y)$ be defined the same way as $l_{0jt}(x, y)$ with $\widehat{u}_t$ in place of u_t , and $\widehat{l}_{1jt}(x, y)$ be the feasible counterpart of $l_{1jt}(x, y)$, i.e. $\widehat{l}_{1jt}(x, y) = \widehat{E}_j(x, y)'l(Y_t, \Omega_{t-1}, \widehat{\theta}_n)$ with

$$\widehat{E}_j(x, y) = \frac{1}{n} \sum_{t=1}^n \left\{ \frac{\partial \widehat{F}_t(\theta_0, x)}{\partial \theta} I(\widehat{u}_{t-j} \leq y) \right\},$$

where $\widehat{\frac{\partial F_t(\theta_0, x)}{\partial \theta}}$ is as (8) with $\widehat{\theta}_n$ replacing θ_0 .

We are now ready to state our bootstrap procedure. We approximate the asymptotic null distribution of $Q(m)$ by that of

$$Q^*(m) = \sum_{j=1}^m (n-j) \widehat{\gamma}_n^{*2}(j) w_j, \quad (13)$$

where

$$\widehat{\gamma}_n^{*2}(j) = n^{-2} \sum_{r=1}^n \sum_{s=1}^n \widehat{\gamma}_{nj}^{*2}(\widehat{u}_r, \widehat{u}_s),$$

and $\widehat{\gamma}_{nj}^*(x, y)$ is obtained from the following two simple steps:

Step 1. Generate a sequence of *iid*, mean 0, unit variance and bounded support random variables $\{\zeta_t\}$ that are independent of $\{Y_t, \Omega_{t-1}\}_{t=1}^n$. One example of $\{\zeta_t\}$ is a sequence of *iid* Bernoulli variables with $P(\zeta_t = .5(1 - \sqrt{5})) = b$, $P(\zeta_t = .5(1 + \sqrt{5})) = 1 - b$ and $b = (1 + \sqrt{5})/2\sqrt{5}$.

Step 2. Get $\widehat{\gamma}_{nj}^*(x, y) = \frac{1}{n-j} \sum_{t=1+j}^n \widehat{l}_{0jt}(x, y) \zeta_t + \frac{1}{\sqrt{n-j}\sqrt{n}} \sum_{t=1}^n \widehat{l}_{1jt}(x, y) \zeta_t$.

The above bootstrap is very easy to implement, as one only needs to generate a sequence of $\{\zeta_t\}$ and does not need to reestimate $\widehat{\theta}_n^*$ in each bootstrap replication. The next theorem theoretically justifies the bootstrap approximation, for which we need the following additional assumption

Assumption A6: $l(Y_t, \Omega_{t-1}, \theta)$ is differentiable in θ (a.s.) in an open neighborhood $\mathcal{N}(\theta_*)$ of θ_* , and $E\left[\sup_{\theta \in \mathcal{N}(\theta_*)} \left\| \frac{\partial l(Y_t, \Omega_{t-1}, \theta)}{\partial \theta} \right\| \right] < C$.

Theorem 5 Under Assumptions A1-A6

$$\sup_{x \in \mathbb{R}} |\Pr(Q^*(m) \leq x | \{Y_t, \Omega_{t-1}\}_{t=1}^n) - \Pr(Q_\infty(m) \leq x)| = o_p(1),$$

where $Q_\infty(m)$ is as in Theorem 3 with θ_* in place of θ_0 .

The bootstrap procedure introduced above can be readily adapted to other examples, as long as we have a consistent estimator for the estimation effect term $E_j(x, y)$ in Theorem 1, or $E(x, z)$ in Theorem 2.

5. MONTE CARLO STUDY

One application of our proposed methodology is conditional goodness-of-fit test, as illustrated in Section 3, Example 2. We are going to evaluate its finite sample performance in this section. To be specific, we test whether $Y_t|\Omega_{t-1}$ follows an AR(1)-GARCH(1,1) process with (1) $N(0, 1)$ innovations, (2) Student's t innovations with unknown degrees of freedom v , denoted by t_v , and (3) Hansen's (1994) skewed t innovations, denoted by $KT(\cdot|\eta_t, \lambda_t)$. All the simulations are performed using the Quarry High Performance Cluster at Indiana University. We use the software R for models (1) and (2), and Matlab for model (3).

5.1 AR-GARCH Models with Gaussian or Student's t Errors

We first consider the following AR(1)-GARCH(1,1) (*AR1-GH*) data generating process (DGP) under the null hypothesis

$$\begin{aligned} Y_t &= a_0 Y_{t-1} + v_t, & v_t &= \sigma_t \varepsilon_t, \\ \sigma_t^2 &= \omega_0 + \alpha_0 v_{t-1}^2 + \beta_0 \sigma_{t-1}^2, \end{aligned} \tag{14}$$

where the true parameters are $(a_0, \omega_0, \alpha_0, \beta_0) = (0.5, 0.025, 0.25, 0.5)$, and ε_t is *iid* $N(0, 1)$. This is the model considered by Koul and Ling (2006, *KL* hereinafter). In the experiments, we generate 1000 realizations of $\{Y_t\}_{t=1}^n$, with $n = 100$ and 200 . Then we estimate the parameters by the maximum likelihood method, and obtain the generalized residuals $\hat{u}_t = \Phi(\hat{\varepsilon}_t)$, where Φ denotes the CDF of standard normal. Our test statistic is $Q(m)$ with $w_j = 1/(j\pi)^2$ and $\hat{\gamma}_{nj}(\cdot, \cdot)$ given in (6),³ and its limit distribution is approximated by the bootstrap procedure described in Section 4.

We compare our tests with *KL*, which is probably the best test available in this setup. It is computed as

$$KL = \sup_{x \in \mathbb{R}} [K_n(x, \hat{\theta}_n)' \hat{\mathcal{I}}_n^{-1} K_n(x, \hat{\theta}_n)],$$

with $\hat{\mathcal{I}}_n := 1/n \sum_{t=1}^n W_t(\hat{\theta}_n) W_t(\hat{\theta}_n)'$, $K_n(x, \hat{\theta}_n) := 1/\sqrt{n} \sum_{t=1}^n W_t(\hat{\theta}_n) [I(\hat{\varepsilon}_t \leq x) - \Phi(x)]$ and $W_t(\hat{\theta}_n) = (\frac{\partial \hat{\sigma}_t}{\partial \omega}, \frac{\partial \hat{\sigma}_t}{\partial \alpha}, \frac{\partial \hat{\sigma}_t}{\partial \beta}) / \hat{\sigma}_t$. In this case, the limit distribution of *KL* is nuisance-parameter-free, although not asymptotically distribution-free, and hence can be approximated by direct simulation.

For the power of the test, we consider the following alternatives :

A_1 : AR(2)-GARCH(1,1) model (*AR2-GH*): $Y_t = 0.5Y_{t-1} + 0.3Y_{t-2} + v_t$.

A_2 : ARMA(1,1)-GARCH(1,1) model (*ARMA-GH*): $Y_t = 0.2Y_{t-1} + 0.7v_{t-1} + v_t$.

³We also tried the choice of $w_j = 1$. The resulting tests have slightly different sizes and powers, and we did not report them here to save space.

A_3 : TAR model: $Y_t = -0.5Y_{t-1} + \varepsilon_t$ if $Y_{t-1} \leq 1$ and $Y_t = 0.4Y_{t-1} + \varepsilon_t$ if $Y_{t-1} > 1$.

A_4 : Bilinear model (*BIL*): $Y_t = 0.5Y_{t-1} + 0.7\varepsilon_{t-1}Y_{t-2} + \varepsilon_t$.

A_5 : AR(1)-GARCH(1,1) model with standardized Student's t_5 innovations (*AR1-GH-T*).

A_6 : AR(1)-GARCH(1,1) model with Hansen's (1994) skewed t innovations, whose degree of freedom parameter $\eta = 5$, and skewness parameter $\lambda = -0.5$ (*AR1-GH-KT*).

A_7 : AR(1)-GARCH(1,1) model with mixed normal innovations $0.7N(0, 0.352^2) + 0.3N(0, 2.295^2)$ (*AR1-GH-MN*).

In models A_1 and A_2 , v_t is defined as in (14), and $\{\varepsilon_t\}$ is *iid* $N(0, 1)$ in all the models. Models A_2 , A_4 and A_5 are studied in Escanciano and Velasco (2009), and A_3 is considered in Hong (2000).

Table 1 gives the empirical sizes and powers of the tests. For our test $Q(m)$, we tried many different values of m , and found that the size and power performances are really similar when $m > 15$. Here we only report the results for $m = 1$ and 9. From the table we can see that, our tests and KL have similar sizes that are fairly close to the nominal levels. For the power, our tests have better performance than KL in general, and the only exceptions are A_3 and A_7 , where KL has slightly better power than ours. Also $Q(9)$ has better power performance than $Q(1)$ in general.

Table 1. Size and Power of $Q(1)$, $Q(9)$ and KL at 10% and 5%, $n = 200$

	$\alpha = 10\%$			$\alpha = 5\%$		
	$Q(1)$	$Q(9)$	KL	$Q(1)$	$Q(9)$	KL
Size						
$H_0 : AR1-GH$	.099	.094	.103	.049	.055	.058
Power						
$A_1 : AR2-GH$	.424	.433	.123	.308	.273	.063
$A_2 : ARMA-GH$	.799	.814	.570	.703	.719	.438
$A_3 : TAR$	.777	.781	.962	.649	.685	.935
$A_4 : BIL$	.177	.155	.121	.104	.091	.068
$A_5 : AR1-GH-T$	.220	.235	.160	.096	.105	.097
$A_6 : AR1-GH-KT$	.701	.752	.625	.524	.581	.511
$A_7 : AR1-GH-MN$	.933	.933	1.00	.930	.932	1.00

Next, we test whether $Y_t|\Omega_{t-1}$ follows an AR(1)-GARCH(1,1) process with Student's t innovations t_v (*AR1-GH-T*), where the degree of freedom parameter v is discrete and unknown. In this case, the estimation effect only stems from the uncertainty of estimating the location and scale parameters. Some unreported simulations confirmed this. See also Escanciano and Olmo (2007) for similar

findings. Therefore, we can treat v as known, although it is estimated. As a result of this, KL is also applicable here.

Our null DGP here is same as (14) except that $\{\varepsilon_t\}$ is *iid* t_5 now. We estimate a_0, ω_0, α_0 and β_0 using QMLE, then estimate v using the 4th moment of ε_t , i.e. $E(\varepsilon_t^4) = 3(v-2)/(v-4)$, and round v to the nearest integer. We also tried estimating v by the maximum likelihood method, and got similar results.

For the power of the tests, we consider alternatives $A'_1, A'_2, A_3, A_4, A_6$ and A_7 , where A'_1 and A'_2 are same as A_1 and A_2 , respectively, except that $\{\varepsilon_t\}$ is *iid* t_5 now instead of *iid* $N(0, 1)$.

Table 2 gives the empirical sizes and powers of the tests. We can see that, our tests have similar size performance as KL . For the power, our tests have better performance than KL in general except for A_3, A_7 and 5% level under A_6 . Besides, $Q(9)$ has better power performance than $Q(1)$ except at 5% level under A_1 . Also worth of mentioning is the DGP in A_7 has the same first 4 moments as the DGP under the null (*AR1-GH-T*), but our tests, as well as KL , still have excellent power against such alternatives.

Table 2. Size and Power of $Q(1)$, $Q(9)$ and KL at 10% and 5%, $n = 200$

	$\alpha = 10\%$			$\alpha = 5\%$		
	$Q(1)$	$Q(9)$	KL	$Q(1)$	$Q(9)$	KL
Size						
$H_0 : AR1-GH-T$	.109	.113	.094	.046	.045	.048
Power						
$A'_1 : AR2-GH-T$	.446	.460	.133	.309	.293	.066
$A'_2 : ARMA-GH-T$	.872	.883	.796	.806	.827	.662
$A_3 : TAR$	.833	.835	.981	.785	.792	.965
$A_4 : BIL$	.327	.340	.231	.202	.220	.148
$A_6 : AR1-GH-KT$	.701	.752	.633	.453	.499	.512
$A_7 : AR1-GH-MN$	.926	.930	.996	.923	.924	.985

5.2 AR-GARCH Models with Skewed t Errors

Hansen (1994) proposes a GARCH model with innovations following a skewed t distribution $KT(\cdot|\eta_t, \lambda_t)$. By allowing η_t and λ_t to depend on the past information Ω_{t-1} , he introduces a way to model the conditional skewness and kurtosis. Since then, Hansen's skewed t distribution has been applied by many authors, such as Jondeau and Rockinger (2003) and Patton (2004), to name a few. However, we have not seen any formal goodness-of-fit test for this distribution in the literature. In this subsection we apply our method to test the adequacy of Hansen's (1994) model, and evaluate

our test's finite sample performance.

Now we consider the following AR(1)-GARCH(1,1) process with Hansen's skewed t innovations (*AR1-GH-KT*)

$$\begin{aligned}
Y_t &= a_0 Y_{t-1} + v_t, & v_t &= \sigma_t \varepsilon_t, \\
\sigma_t^2 &= \omega_0 + \alpha_0 v_{t-1}^2 + \beta_0 \sigma_{t-1}^2, \\
\varepsilon_t &\sim KT(\cdot | \eta_t, \lambda_t), & \eta_t &= g_{(2.1, 30)}(\tilde{\eta}_t), & \lambda_t &= g_{(-0.9, 0.9)}(\tilde{\lambda}_t), \\
\tilde{\eta}_t &= a_1 + b_1 v_{t-1} + c_1 v_{t-1}^2, & \tilde{\lambda}_t &= a_2 + b_2 v_{t-1} + c_2 v_{t-1}^2,
\end{aligned} \tag{15}$$

where g is the logistic function, i.e. $g_{(L,U)}(x) = L + (U - L)/(1 + \exp(-x))$.

For the size, we set $(a_0, \omega_0, \alpha_0, \beta_0) = (0.5, 0.025, 0.25, 0.5)$, $(a_1, b_1, c_1) = (-1, -0.5, -0.1)$ and $(a_2, b_2, c_2) = (-0.1, -0.15, -0.10)$, and generate 1000 realizations of $\{Y_t\}_{t=1}^n$, with $n = 200$ and 400. Then we estimate the parameters by MLE. After obtaining the generalized residuals $\hat{u}_t$, we calculate $\hat{\gamma}_{nj}(\hat{u}_t, \hat{u}_s)$ given in (6) and the test statistic $Q(m)$, whose limit distribution is approximated by the bootstrap procedure described in Section 3. We consider alternatives A_1'' , A_3 , A_4 and A_7 for the power, where A_1'' is same as A_1 except that $\{\varepsilon_t\}$ are $KT(\cdot | \eta_t, \lambda_t)$ innovations now instead of *iid* $N(0, 1)$.

Table 3 reports the empirical sizes and powers for $n = 200$. We observe that the sizes of our tests are fairly close to the α 's, and the powers are reasonably high.

Table 3. Size and Power of $Q(1)$ and $Q(9)$ at 10% and 5%, $n = 200$

	$\alpha = 10\%$		$\alpha = 5\%$	
	$Q(1)$	$Q(9)$	$Q(1)$	$Q(9)$
Size				
$H_0 : AR1-GH-KT$	.113	.115	.066	.061
Power				
$A_1'' : AR2-GH-KT$	.326	.304	.212	.182
$A_3 : TAR$	.648	.675	.501	.530
$A_4 : BIL$	.148	.144	.091	.086
$A_7 : AR1-GH-MN$	.126	.137	.066	.068

6. EMPIRICAL APPLICATION

In this section, we apply our test to the daily exchange rates data. We use AU/US, EU/US, YEN/US, UK/US for the amount of Australian Dollar, Euro, Japanese Yen, and British Pound necessary to purchase one US dollar. Our data are obtained from *Freelunch.com* over the period 01/2004- 10/2007, with a total of 965 observations.

For each of the four series we test the hypothesis H_0 : the log-return, Y_t , follows an AR(1)-GARCH(1,1) process with $N(0, 1)$ innovations, and the results are summarized in Table 4.

Table 4. Empirical Results for Exchange Rates Fitted AR(1)-GARCH(1,1)-N(0,1) Models

Currencies	p-value of $Q(1)$	p-value of $Q(9)$	p-value of KL
AU/US	.008	.002	.342
EU/US	.006	.004	.956
YEN/US	.01	.004	.068
UK/US	..016	.002	.996

Our tests reject H_0 at 5% level for all four series, while KL rejects none of them. Notice that under H_0 , $u_t = u_t(\theta_0) = \Phi((Y_t - a_0 Y_{t-1})/\sigma_t)$, with σ_t defined in (14), should be *iid* and $U[0, 1]$ distributed. We make QQ-plots for the generalized residuals $\hat{u}_t = u_t(\hat{\theta}_n)$, and see whether they are close to a $U[0, 1]$ distribution. The plots are given in Fig. 1. For all four series, $\hat{u}_t$ appears to be fairly close to a $U[0, 1]$ distribution, although the plots should be taken with caution as they do not take into account of the parameter estimation effect. Nonetheless, the plots may help explain why KL fails to reject H_0 , as KL is only designed for testing the $U[0, 1]$ part.

Fig. 1. QQ-plot of $\hat{u}_t$ against $U[0, 1]$ quantiles, where $\hat{u}_t = H(Y_t, \Omega_{t-1}, \hat{\theta}_n)$, with $H(\cdot, \Omega_{t-1}, \theta_0)$ the conditional distribution of an AR(1)-GARCH(1,1)-N(0,1) process, Y_t the log-return of AU/US, EU/US, YEN/US and UK/US exchange rate, respectively, and $\hat{\theta}_n$ the MLE of θ_0 . The dotted lines are the 95% confidence bands.

Next we examine whether u_t is *iid*. For that we plot the autocorrelations of $\hat{u}_t$ in Fig. 2. We observe that $\hat{u}_t$ is highly autocorrelated in all four series. Although this does not necessarily indicate high autocorrelations of u_t as the plots do not take into account of the parameter estimation effect, the plots provide one possible explanation for our tests' small p-values: u_t is not *iid*.

This application highlights the merits of using our method for conditional goodness-of-fit purpose. Our tests here complement *KL* and Bai (2003) as we test both for the *iid* and $U[0, 1]$ distribution of u_t .

Fig. 2. Autocorrelations of $\hat{u}_t$, where $\hat{u}_t = H(Y_t, \Omega_{t-1}, \hat{\theta}_n)$, with $H(\cdot, \Omega_{t-1}, \theta_0)$ the conditional distribution of an AR(1)-GARCH(1,1)-N(0,1) process, Y_t the log-return of AU/US, EU/US, YEN/US and UK/US exchange rate, respectively, and $\hat{\theta}_n$ the MLE of θ_0 . The dotted lines are the 95% confidence bands.

7. CONCLUSION

In this paper, we develop a general method of testing for independence when unobservable generalized errors are involved. Our method can be applied to testing for serial independence of generalized errors, and testing for independence between the generalized errors and observable covariates. Our tests are functional of the classical Hoeffding-Blum-Kiefer-Rosenblatt-type empirical process applied to residuals. We establish a uniform expansion of the process, and thereby deriving an explicit expression for the parameter estimation effect in a general setup. Due to the nonstandard limit distribution of the tests, we use a multiplier-type bootstrap to approximate their null limit distributions.

Finally, we point out some possible directions for future work. One remaining issue in the current paper is how to choose the optimal w_j in (11) to maximize the power of the test against certain alternatives. We may address this problem along the lines of Delgado and Velasco (2009), who study a similar problem for Box-Pierce type tests. Moreover, as we already have the expression for the parameter estimation effect, we may consider applying some transformations to get asymptotically distribution-free test. For example, in the setup of Section 3.2, we may use a Khmaladze transformation to annihilate the parameter estimation effect.

8. APPENDIX

PROOF OF THEOREM 1:

We can write

$$\begin{aligned} \sqrt{n-j}\{\hat{\gamma}_{nj}(x, y) - \gamma_{nj}(x, y)\} &= \frac{1}{\sqrt{n-j}} \sum_{t=1+j}^n \{I(\hat{u}_t \leq x)I(\hat{u}_{t-j} \leq y) - I(u_t \leq x)I(u_{t-j} \leq y)\} - \\ &\quad \frac{1}{(n-j)^{3/2}} \left\{ \left[\sum_{t=1+j}^n I(\hat{u}_t \leq x) \right] \left[\sum_{t=1+j}^n I(\hat{u}_{t-j} \leq y) \right] - \left[\sum_{t=1+j}^n I(u_t \leq x) \right] \left[\sum_{t=1+j}^n I(u_{t-j} \leq y) \right] \right\} \\ &:= A_{1n}(x, y) - A_{2n}(x, y) \end{aligned}$$

To handle A_{1n} , define the process

$$K_n(c, x, y) = \frac{1}{\sqrt{n-j}} \sum_{t=1+j}^n \{I(u_t(\theta_0 + \frac{c}{\sqrt{n}}) \leq x) - E[I(u_t(\theta_0 + \frac{c}{\sqrt{n}}) \leq x) | \Omega_{t-1}]\} I(u_{t-j}(\theta_0 + \frac{c}{\sqrt{n}}) \leq y),$$

indexed by $(c, x, y) \in C_K \times \mathbb{R}^2$, where $C_K = \{c \in \mathbb{R}^p : |c| \leq K\}$, and $K > 0$ is an arbitrary but fixed constant.

LEMMA A1: Under Assumptions A1-A2, $K_n(c, x, y)$ is asymptotically tight with respect to $(c, x, y) \in C_K \times \mathbb{R}^2$.

PROOF OF LEMMA A1: Let d_K be the metric on C_K induced by the Euclidian norm, and define a metric d_R on $\mathbb{R}$ by $d_R(x_1, x_2) = |F(x_1) - F(x_2)|$. Then a metric d on $C_K \times \mathbb{R}^2$ can be defined as $d((c_1, x_1, y_1), (c_2, x_2, y_2)) = \max(d_K(c_1, c_2), d_R(x_1, x_2), d_R(y_1, y_2))$. We can see that $C_K \times \mathbb{R}^2$ is compact with respect to d , and we then get the result by following the steps of Lemma A1 in Escanciano and Olmo (2007).

LEMMA A2: Under Assumptions A1-A2, we have $\sup_{(x, y) \in \mathbb{R}^2} |K_n(\hat{c}, x, y) - K_n(0, x, y)| = o_p(1)$, for any $\hat{c} = O_p(1)$.

PROOF OF LEMMA A2: Let $\varepsilon, \varsigma > 0$ be given. We wish to show that

$$P \left(\sup_{(x, y) \in \mathbb{R}^2} |K_n(\hat{c}, x, y) - K_n(0, x, y)| > \varsigma \right) < \varepsilon \quad \text{as } n \rightarrow \infty.$$

Since $\hat{c} = O_p(1)$, there exists $K > 0$ such that $P(|\hat{c}| > K) < \varepsilon/2$.

Some algebra shows that, for any fixed $(c, x, y) \in C_K \times \mathbb{R}^2$, $E[|K_n(c, x, y) - K_n(0, x, y)|^2] = o(1)$, which implies, by Chebyshev's Inequality,

$$|K_n(c, x, y) - K_n(0, x, y)| = o_p(1) \quad \forall (c, x, y) \in C_K \times \mathbb{R}^2.$$

The above display, the compactness of $C_K \times \mathbb{R}^2$ with respect to d and the asymptotic tightness of $K_n(c, x, y)$ imply

$$P\left(\sup_{(c, x, y) \in C_K \times \mathbb{R}^2} |K_n(c, x, y) - K_n(0, x, y)| > \varsigma\right) \leq \frac{\varepsilon}{2},$$

for n sufficiently large. Therefore, for n sufficiently large

$$\begin{aligned} P\left(\sup_{(x, y) \in \mathbb{R}^2} |K_n(\hat{c}, x, y) - K_n(0, x, y)| > \varsigma\right) &\leq P(|\hat{c}| > K) + P\left(\sup_{(c, x, y) \in C_K \times \mathbb{R}^2} |K_n(c, x, y) - K_n(0, x, y)| > \varsigma\right) \\ &< \varepsilon. \end{aligned}$$

Apply Lemma A2 to $\hat{c} = \sqrt{n}(\hat{\theta}_n - \theta_0) = O_p(1)$, and we get

$$\begin{aligned} \sup_{(x, y) \in \mathbb{R}^2} \frac{1}{\sqrt{n-j}} \left| \sum_{t=1+j}^n \{I(\hat{u}_t \leq x) - E[I(\hat{u}_t \leq x)|\Omega_{t-1}]\} I(\hat{u}_{t-j} \leq y) - \right. \\ \left. \sum_{t=1+j}^n \{I(u_t \leq x) - E[I(u_t \leq x)|\Omega_{t-1}]\} I(u_{t-j} \leq y) \right| = o_p(1). \end{aligned}$$

Hence,

$$\begin{aligned} A_{1n}(x, y) &= \frac{1}{\sqrt{n-j}} \sum_{t=1+j}^n \{I(\hat{u}_t \leq x)I(\hat{u}_{t-j} \leq y) - I(u_t \leq x)I(u_{t-j} \leq y)\} \\ &= \frac{1}{\sqrt{n-j}} \sum_{t=1+j}^n \{E[I(\hat{u}_t \leq x)|\Omega_{t-1}] - E[I(u_t \leq x)|\Omega_{t-1}]\} I(\hat{u}_{t-j} \leq y) + \\ &\quad \frac{1}{\sqrt{n-j}} \sum_{t=1+j}^n E[I(u_t \leq x)|\Omega_{t-1}] \{I(\hat{u}_{t-j} \leq y) - I(u_{t-j} \leq y)\} + o_p(1) \\ &\equiv B_{1n}(x, y) + B_{2n}(x, y) + o_p(1). \end{aligned}$$

Notice that $F_t(\theta, x) = E[I(u_t(\theta) \leq x) | \Omega_{t-1}]$, then

$$\begin{aligned}
B_{1n}(x, y) &= \frac{1}{\sqrt{n-j}} \sum_{t=1+j}^n \{F_t(\hat{\theta}_n, x) - F_t(\theta_0, x)\} I(\hat{u}_{t-j} \leq y) \\
&= \frac{1}{\sqrt{n-j}} \sum_{t=1+j}^n (\hat{\theta}_n - \theta_0)' \frac{\partial F_t(\tilde{\theta}_n, x)}{\partial \theta} I(\hat{u}_{t-j} \leq y) \\
&= \sqrt{n}(\hat{\theta}_n - \theta_0)' \frac{1}{\sqrt{n}} \frac{1}{\sqrt{n-j}} \sum_{t=1+j}^n \frac{\partial F_t(\tilde{\theta}_n, x)}{\partial \theta} I(\hat{u}_{t-j} \leq y) \\
&= \sqrt{n}(\hat{\theta}_n - \theta_0)' E\left[\frac{\partial F_t(\theta_0, x)}{\partial \theta} I(u_{t-j} \leq y)\right] + o_p(1) \quad (\text{By ULLN}).
\end{aligned}$$

$$B_{2n}(x, y) = \frac{1}{\sqrt{n-j}} \sum_{t=1+j}^n F_t(\theta_0, x) \{I(\hat{u}_{t-j} \leq y) - I(u_{t-j} \leq y)\}.$$

Using the same argument as we use for A_{1n} , we get

$$\begin{aligned}
B_{2n}(x, y) &= \frac{1}{\sqrt{n-j}} \sum_{t=1+j}^n F_t(\theta_0, x) \{F_{t-j}(\hat{\theta}_n, y) - F_{t-j}(\theta_0, y)\} + o_p(1) \\
&= \sqrt{n}(\hat{\theta}_n - \theta_0)' F(x) \frac{1}{\sqrt{n}} \frac{1}{\sqrt{n-j}} \sum_{t=1+j}^n \frac{\partial F_{t-j}(\tilde{\theta}_n, y)}{\partial \theta} + o_p(1) \\
&\quad (\text{need } u_t \text{ independent of } \Omega_{t-1}, \text{ which is often implied by } H_0) \\
&= \sqrt{n}(\hat{\theta}_n - \theta_0)' F(x) E\left[\frac{\partial F_t(\theta_0, y)}{\partial \theta}\right] + o_p(1) \quad (\text{By ULLN}).
\end{aligned}$$

Now, turn to A_{2n}

$$\begin{aligned}
A_{2n}(x, y) &= \frac{1}{(n-j)^{3/2}} \left\{ \left[\sum_{t=1+j}^n I(\hat{u}_t \leq x) \right] \left[\sum_{t=1+j}^n I(\hat{u}_{t-j} \leq y) \right] - \left[\sum_{t=1+j}^n I(u_t \leq x) \right] \left[\sum_{t=1+j}^n I(u_{t-j} \leq y) \right] \right\} \\
&= \frac{1}{(n-j)^{3/2}} \left\{ \left[\sum_{t=1+j}^n I(\hat{u}_t \leq x) \right] \left[\sum_{t=1+j}^n I(\hat{u}_{t-j} \leq y) \right] - \left[\sum_{t=1+j}^n I(u_t \leq x) \right] \left[\sum_{t=1+j}^n I(\hat{u}_{t-j} \leq y) \right] \right\} + \\
&\quad \frac{1}{(n-j)^{3/2}} \left\{ \left[\sum_{t=1+j}^n I(u_t \leq x) \right] \left[\sum_{t=1+j}^n I(\hat{u}_{t-j} \leq y) \right] - \left[\sum_{t=1+j}^n I(u_t \leq x) \right] \left[\sum_{t=1+j}^n I(u_{t-j} \leq y) \right] \right\} \\
&\equiv C_{1n}(x, y) + C_{2n}(x, y) \\
C_{1n}(x, y) &= \frac{1}{(n-j)^{3/2}} \left\{ \left[\sum_{t=1+j}^n I(\hat{u}_t \leq x) \right] \left[\sum_{t=1+j}^n I(\hat{u}_{t-j} \leq y) \right] - \left[\sum_{t=1+j}^n I(u_t \leq x) \right] \left[\sum_{t=1+j}^n I(\hat{u}_{t-j} \leq y) \right] \right\} \\
&= \frac{1}{n-j} \left\{ \sum_{t=1+j}^n I(\hat{u}_{t-j} \leq y) \right\} \frac{1}{\sqrt{n-j}} \left\{ \sum_{t=1+j}^n [I(\hat{u}_t \leq x) - I(u_t \leq x)] \right\}.
\end{aligned}$$

Using the same argument as we use for B_{2n} , we get

$$\begin{aligned}
C_{1n}(x, y) &= [F(y) + o_p(1)] \left\{ \frac{1}{\sqrt{n-j}} \sum_{t=1+j}^n [F_t(\hat{\theta}_n, x) - F_t(\theta_0, x)] + o_p(1) \right\} \\
&= \sqrt{n}(\hat{\theta}_n - \theta_0)' F(y) \frac{1}{\sqrt{n}} \frac{1}{\sqrt{n-j}} \sum_{t=1+j}^n \frac{\partial F_t(\tilde{\theta}_n, x)}{\partial \theta} + o_p(1) \\
&= \sqrt{n}(\hat{\theta}_n - \theta_0)' F(y) E\left[\frac{\partial F_t(\theta_0, x)}{\partial \theta}\right] + o_p(1) \quad (\text{By ULLN}).
\end{aligned}$$

Similarly,

$$\begin{aligned}
C_{2n}(x, y) &= \frac{1}{(n-j)^{3/2}} \left\{ \left[\sum_{t=1+j}^n I(u_t \leq x) \right] \left[\sum_{t=1+j}^n I(\hat{u}_{t-j} \leq y) \right] - \left[\sum_{t=1+j}^n I(u_t \leq x) \right] \left[\sum_{t=1+j}^n I(u_{t-j} \leq y) \right] \right\} \\
&= \frac{1}{n-j} \left\{ \sum_{t=1+j}^n I(u_t \leq x) \right\} \frac{1}{\sqrt{n-j}} \left\{ \sum_{t=1+j}^n [I(\hat{u}_{t-j} \leq y) - I(u_{t-j} \leq y)] \right\} \\
&= \sqrt{n}(\hat{\theta}_n - \theta_0)' F(x) E\left[\frac{\partial F_t(\theta_0, y)}{\partial \theta}\right] + o_p(1)
\end{aligned}$$

Therefore,

$$\begin{aligned}
\sup_{(x,y) \in \mathbb{R}^2} \left| \sqrt{n-j} [\hat{\gamma}_{nj}(x, y) - \gamma_{nj}(x, y)] \right| &= \sup_{(x,y) \in \mathbb{R}^2} |A_{1n}(x, y) - A_{2n}(x, y)| \\
&= \sup_{(x,y) \in \mathbb{R}^2} |B_{1n}(x, y) + B_{2n}(x, y) - C_{1n}(x, y) - C_{2n}(x, y) + o_p(1)| \\
&= \sqrt{n}(\hat{\theta}_n - \theta_0)' E \left\{ \frac{\partial F_t(\theta_0, x)}{\partial \theta} [I(u_{t-j} \leq y) - F(y)] \right\} + o_p(1) \\
&\equiv \sqrt{n}(\hat{\theta}_n - \theta_0)' E_j(x, y) + o_p(1),
\end{aligned}$$

where $E_j(x, y) := E \left\{ \frac{\partial F_t(\theta_0, x)}{\partial \theta} [I(u_{t-j} \leq y) - F(y)] \right\}$.

■

PROOF OF THEOREM 2:

$$\begin{aligned}
\sqrt{n}\{\hat{\gamma}_n(x, z) - \gamma_n(x, z)\} &= \frac{1}{\sqrt{n}} \sum_{t=1}^n [I(\hat{u}_t \leq x) - I(u_t \leq x)] I(Z_t \leq z) - \\
&\quad F_n(z) \cdot \frac{1}{\sqrt{n}} \sum_{t=1}^n [I(\hat{u}_t \leq x) - I(u_t \leq x)] \\
&\equiv A_{1n}(x, z) - A_{2n}(x, z)
\end{aligned}$$

Define the process

$$K_n(c, x, z) = \frac{1}{\sqrt{n}} \sum_{t=1}^n \{I(u_t(\theta_0 + \frac{c}{\sqrt{n}}) \leq x) - E[I(u_t(\theta_0 + \frac{c}{\sqrt{n}}) \leq x) | \Omega_{t-1}]\} I(Z_t \leq z),$$

indexed by $(c, x, z) \in C_K \times \mathbb{R} \times \mathbb{R}^k$, where $C_K = \{c \in \mathbb{R}^p : |c| \leq K\}$, and $K > 0$ is an arbitrary but fixed constant. The rest of the proof then follows from that of Theorem 1.

■

PROOF OF THEOREM 3:

Notice that $Q(m) = \int_{[0,1] \times \mathbb{R}^2} (S_n^m(\lambda, x, y))^2 d\lambda d\widehat{F}_n(x) d\widehat{F}_n(y)$, where

$$S_n^m(\lambda, x, y) = \sum_{j=1}^m (n-j)^{\frac{1}{2}} \widehat{\gamma}_{nj}(x, y) \sqrt{2w_j} \sin j\pi\lambda.$$

By the Continuous Mapping Theorem, $Q(m) \xrightarrow{d} Q_\infty(m)$ if $S_n^m(\cdot)$ converges weakly to $S_\infty^m(\cdot)$, which follows from Theorem 1.

■

PROOF OF THEOREM 4:

The result follows directly from Theorem 2 and the Continuous Mapping Theorem.

■

PROOF OF THEOREM 5:

Let $l_{0jt}^*(x, y)$ and $l_{1jt}^*(x, y)$ be defined in the same way as $l_{0jt}(x, y)$ and $l_{1jt}(x, y)$, respectively, with θ_* in place of θ_0 . Define

$$\begin{aligned} \gamma_{nj}^0(x, y) &= \frac{1}{n-j} \sum_{t=1+j}^n l_{0jt}^*(x, y) + \frac{1}{\sqrt{n-j}\sqrt{n}} \sum_{t=1}^n l_{1jt}^*(x, y), \\ \gamma_{nj}^{0*}(x, y) &= \frac{1}{n-j} \sum_{t=1+j}^n l_{0jt}^*(x, y) \zeta_t + \frac{1}{\sqrt{n-j}\sqrt{n}} \sum_{t=1}^n l_{1jt}^*(x, y) \zeta_t. \end{aligned}$$

By the multiplier central limit theorem, see van der Vaart and Wellner (1996, p.176), $\sqrt{n-j}\gamma_{nj}^{0*}(\cdot, \cdot)$ converges weakly to the same limit process as $\sqrt{n-j}\gamma_{nj}^0(\cdot, \cdot)$.

Under Assumptions A1-A6, it is immediate that

$$\sup_{(x,y) \in \mathbb{R}^2} \left| \sqrt{n-j}[\widehat{\gamma}_{nj}^*(x, y) - \gamma_{nj}^{0*}(x, y)] \right| = o_p(1).$$

Therefore, $\sqrt{n-j}\widehat{\gamma}_{nj}^*(\cdot, \cdot)$ converges weakly to the same limit process as $\sqrt{n-j}\gamma_{nj}^0(\cdot, \cdot)$. The conclusion then follows from the Continuous Mapping Theorem.

■

REFERENCES

- [1] BAI, J. (2003), "Testing Parametric Conditional Distributions of Dynamic Models", *Review of Economics and Statistics*, Aug 2003, 85(3): 531-549
- [2] BOX, G. AND PIERCE, D. (1970), "Distribution of residual Autocorrelations in Autorregressive Integrated Moving Average Time Series Models", *Journal of American Statistical Association*, 65, 1509-1527.
- [3] BROWN, B.W. (1983), "The Identification Problem in Systems Nonlinear in the Variables," *Econometrica*, 51, 175-196.
- [4] BROWN, D., DEB, R. AND WEGKAMP, M.H. (2007), "Tests of independence in separable econometric models: Theory and Application," Cowles Foundation Discussion Paper No. 1395RR.
- [5] BROWN, D. AND WEGKAMP, M. (2002), "Weighted Minimum Mean-Square Distance from Independence Estimation," *Econometrica*, 70(5), 2035-2051.
- [6] CORRADI, V. AND SWANSON, N. (2006), "Bootstrap conditional distribution tests in the presence of dynamic misspecification", *Journal of Econometrics* 133, 779 – 806
- [7] DELGADO, M.A., AND VELASCO, C. (2009), "A New Class of Distribution-Free Tests for Time Series Models Specification," Universidad Carlos III de Madrid working paper.
- [8] DU, Z. AND ESCANCIANO, J.C. (2008), "A nonparametric test for serial independence of errors", working paper.
- [9] DURBIN, J. (1973), "Weak convergence of the sample distribution function when parameters are estimated", *Annals of Statistics*, 1, 274-290.
- [10] EFRON, B. (1979), "Bootstrap methods: Another look at jackknife". *Annals of Statistics*, 7:1-26.
- [11] ESCANCIANO, J.C. (2007) "Weak Convergence of Non-Stationary Multivariate Marked Processes with Applications to Martingale Testing", *Journal of Multivariate Analysis*, 98, 1321-1336.
- [12] ESCANCIANO, J.C. AND OLMO, J. (2007), "Estimation Risk Effects on Backtesting for Parametric Value-at-Risk Models". CAEPR Working Paper No. 2007-005.
- [13] ESCANCIANO, J.C. AND VELASCO, C. (2009), "Specification Tests of Parametric Dynamic Conditional Quantiles", working paper.
- [14] HANSEN, B.E. (1994) "Autoregressive Conditional Density Estimation", *International Economic Review*, 35, 705–730.

- [15] HIDALGO, J. AND ZAFFARONI, P. (2007): "A goodness-of-fit test for ARCH(∞) models" *Journal of Econometrics* 141, 835 – 875
- [16] HOEFFDING, W. (1948), "A nonparametric test of independence", *Annals of Mathematical Statistics* 26, 189-211.
- [17] HONG, Y. (2000), "Generalized Spectral Test for Serial Dependence", *Journal of the Royal Statistical Society Series B*, 62, part3.
- [18] HONG, Y. AND LEE T.H. (2003), "Diagnostic Checking for Adequacy of Linear and Nonlinear Time Series Models", *Econometric Theory*, 19, 2003, 1065–1121.
- [19] HONG, Y. AND LI, H. (2005), "Nonparametric Specification Testing for Continuous-Time Models with Applications to Term Structure of Interest Rates", *The Review of Financial Studies*, 18, 1; 2005.
- [20] JONDEAU, E. AND ROCKINGER, M. (2003), "Conditional volatility, skewness, and kurtosis: existence, persistence, and comovements," *Journal of Economic Dynamics and Control*, vol. 27(10), 1699-1737.
- [21] KHMALADZE, E.V. (1981): "Martingale approach in the theory of Goodness-of-fit tests," *Theory of Probability and its Applications*, 26, 240-257.
- [22] KOUL, H. AND LING, S. (2006), "Fitting an error distribution in some heteroscedastic time series models", *Annals of Statistics* 34, 994-1012.
- [23] LJUNG, G.M AND BOX, G.E.P. (1978), "A Measure of Lack of Fit in Time Series Models", *Biometrika* 65, 297-303.
- [24] PARK, J.Y. AND VASUDEV R. (2006), "Testing for martingales in continuous time," Working paper, Dept. of Economics, Rice University.
- [25] PATTON, A. J. (2004), "On the Out-of-Sample Importance of Skewness and Asymmetric Dependence for Asset Allocation", *Journal of Financial Econometrics* Vol. 2, No. 1, pp. 130-168
- [26] VAN DER VAART, A. W. AND WELLNER, J.A.(1996), *Weak Convergence and Empirical Processes*. Springer, New York.