A Simple Automatic Portmanteau Test for Conditional Goodness-of-Fit in Dynamic Models

Zaichao Du

Southwestern University of Finance and Economics, China

Juan Carlos Escanciano

Indiana University, Bloomington, USA

April 4, 2012

Abstract

In this paper, we propose a data-driven Portmanteu test for conditional goodness-of-fit in dynamic models. Our method uses the well-known fact that under the correct specification of the conditional distribution the generalized "errors" obtained after the conditional probability integral transformation are *iid* $U[0,1]$. The proposed test is a modified Box-Pierce statistic applied to the generalized errors, with a data-driven choice for the number of autocorrelations used. The test explicitly takes into account of the parameter estimation effect, and as a result it has a convenient standard chi-squared limit distribution. Hence, the main distinctive feature of our approach is its simplicity. The basic methodology is extended to conditional models for the tail, conditional hazard models and diffusion models. It is shown that, unlike existing approaches, our approach is applicable to a wide class of models, including ARMA-GARCH models with time varying higher order moments, such as Hansen's (1994) skewed t model. A simulation study shows that our test has a satisfactory size and power performance. Finally, an empirical application to the Nikkei Index data highlights the merits of the proposed test over competing alternatives.

Keywords and Phrases: Autocorrelation; Box-Pierce; Goodness-of-Fit; GARCH; Parameter estimation uncertainty; Skewed t distribution.

JEL Classifications: C12, C58, C52.

1. INTRODUCTION

The correct specification of a conditional distribution is important for a number of reasons. From the point of view of model estimation, misspecification of the conditional distribution leads to estimates that are at least inefficient (see Bollerslev, 1986), and often inconsistent. In particular, Engle and Gonzalez-Rivera (1991) showed that inefficiency of the Gaussian quasi-maximum likelihood estimator can be substantial when the true distribution is skewed. The correct specification of a conditional distribution is also crucial for density forecast, which is increasingly more important and commonplace, for example in financial risk management, portfolio choice and asset pricing. In other areas of time series analysis, conditional quantiles rather than densities are used. For instance, the most popular risk management tool is Value-at-Risk (VaR), which summarizes the downside risk of a portfolio in a single statistic. VaR is defined as the maximum loss on a portfolio which can be expected over a given horizon with a certain degree of confidence. There are several methods of estimating the VaR of a portfolio, but the most common is to assume a parametric conditional distribution function for portfolio returns and to estimate VaR as the appropriate quantile of this distribution (see Christoffersen 2009). Clearly the estimated VaR of a portfolio will be sensitive to the correct specification of the conditional distribution. Another general example is counterfactual policy analysis, which often relies on the correct specification of the conditional distribution, see e.g. the application to wage distribution in DiNardo, Fortin and Lemieux (1996) or the counterfactual analysis on the patent enforcement in the market for anti-bacterial drugs in India by Chaudhuri, Goldberg and Jia (2006). Conditional goodness-of-fit tests are then motivated by an interest in avoiding wrong conclusions that may arise in numerous applications relying on correct specification of a conditional model.

The problem of testing for the goodness-of-fit of an *unconditional* parametric distribution is a classical problem in statistics, see the seminal contributions by Pearson (1900), Kolmogorov (1933) and Smirnov (1939). These classical contributions were tailored to unconditional distributions of independent and identically distributed (*iid*) observations. Andrews (1997) extended the Kolmogorov test to a conditional distribution of a response variable given a covariate, but his test is still confined to *iid* observations. For other tests in this framework, see e.g. Heckman (1984), Zheng (2000) and Delgado and Stute (2008), among many others.

In contrast, this paper studies the problem of testing conditional distributions of *dynamic* models. More precisely, let Y_t be a real-valued dependent variable and Ω_{t-1} be the conditioning set that contains lagged values of Y_t as well as current and lagged values of other exogenous variables, say X_t. That is, $\Omega_{t-1} = \{X_t, X_{t-1}, ...; Y_{t-1}, Y_{t-2}, ...\}$. We aim to test that the conditional distribution of Y_t given Ω_{t-1} has a certain parametric form, say $G_t(\cdot, \Omega_{t-1}, \theta_0)$, where θ_0 is some unknown finite-dimensional parameter in a compact set $\Theta \subset \mathbb{R}^p$.

There are three main features that make dynamic conditional goodness-of-fit a challenging testing problem. First, as θ_0 is unknown, one has to base a test on some consistent estimator of θ_0, but then, the null limit distribution of tests will generally lose the "nuisance parameter-free" property, see Durbin (1973). This problem has been extensively studied in the literature and several solutions are already available. For instance, Bai (2003) applies a Khmaladze transformation (cf. Khmaladze, 1981) to get an asymptotically distribution-free test, while Hong and Li (2005) use a kernel smoothing method. Corradi and Swanson (2003), Li and Tkacz (2006) and Hidalgo and Zaffaroni (2007) use bootstrap to overcome the problem of the data-dependent asymptotic distributions. Recently, Koul and Ling (2006) propose a nuisance-parameter (but not distribution-) free test for some ARMA-GARCH models.

Second, the analysis is complicated by the fact that the whole history Ω_{t-1} is not observed. For Markov models this limitation does not pose a problem, as when e.g. $G_t(\cdot, \Omega_{t-1}, \theta_0) = G(\cdot, Y_{t-1}, \theta_0)$. Unfortunately, many popular models are non-Markovian and a truncated and estimated version of Ω_{t-1} is needed. Here, we denote generically this estimator by $\hat{\Omega}_{t-1}$, which contains $\{X_s, Y_{s-1}\}_{s=0}^{t}$ and possibly some other initial values. The observed sample is then $\{Y_t, \hat{\Omega}_{t-1}\}_{t=0}^{n}$, with $n \geq 1$ denoting the sample size.

Third, the growing dimension of the conditioning variables leads to a high-dimensional problem. To reduce the dimension the vast majority of existing tests deal with dynamic models driven by *iid* innovations or impose Markov restrictions; see e.g. Inglot and Stawiarski (2005), Li and Tkacz (2006), Koul and Ling (2006), Horvath and Zitikis (2006) or Neumann and Paparoditis (2008). There is, however, a vast empirical evidence documenting a lack of fit of models driven by *iid* variables with financial data. See e.g. Hansen

(1994), Harvey and Siddique (1999), Premaratne and Bera (2001), Jondeau and Rockinger (2003), Leon, Rubio, and Serna (2005) and Brooks et al. (2005), among many others. As mentioned earlier, it is also widely recognized that appropriate dynamic models for financial applications, e.g. GARCH models, are non-Markovian.

Our approach uses the well-known fact that for a continuous and correctly specified conditional distribution G_t, the sequence of generalized "errors"

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

comprises a sample of *iid* $U[0,1]$ variables. Many researchers have used this property to develop goodness-of-fit tests. Diebold et al. (1998) used this idea in the context of evaluating density forecast, but they did not provide a formal test, nor did they take into account of the parameter estimation effect; see also Berkowitz (2001). Bai (2003) also uses the probability integral transform, but his test is only tailored for whether u_t is $U[0,1]$ distributed. Hong and Li (2005) use the generalized errors $\{u_t\}$ to develop a joint test for pairwise independence and uniformity using kernel estimators. However, the resulting test depends on the choice of a bandwidth, for which there is no existing theory available.[1]

Our proposed test is a simple Box-Pierce-type test, modified to take into account of the parameter estimation effect, and fully automatic in the sense that the number of autocorrelations used is chosen from the data. The proposed test has several appealing properties. First, the test has a convenient chi-squared limit distribution and hence it is extremely easy to implement compared with other existing methods based on martingale transforms (see e.g. Bai, 2003) or bootstrap. Second, it can be applied to a wide class of models, including but not limited to ARMA-GARCH, autoregressive conditional duration and Hansen's (1994) skewed t with time-varying parameters.[2] In this respect our method allows for a richer set of dynamics than the vast majority of existing methods, and in particular, it allows for models with time varying conditional skewness and kurtosis like Hansen's skewed t distribution.

An ubiquitous feature of all existing tests is their inconsistency when general dynamics is allowed for. This is justified by the complexity of the testing problem. Our proposed

[1] To the best of our knowledge, it is not known whether the only avaible results in Gao and Gijbels (2008) can be extended to the present setting.

[2] Hansen's model has been applied by many authors, such as Jondeau and Rockinger (2003) and Patton (2004), to name a few.

test is no exception and is also inconsistent, but it is considerably simpler to implement than existing tests. Nevertheless, several simulations suggest that our test has satisfactory empirical power against a wide variety of fixed alternatives commonly used in applications, and that it compares favorably with one of the best competing methods proposed in the literature, namely Koul and Ling's (2006) test.

Another contribution of this paper is that we develop a goodness-of-fit test for the conditional *tail* distribution. In risk management people are often interested in the tail instead of the whole distribution. Examples include VaR, expected shortfall (the expected loss beyond the VaR) and stress testing. To the best of our knowledge, our proposed test is the first test specially designed for the conditional tail distribution.

The rest of the paper is organized as follows. In Section 2 we introduce our general automatic Portmanteau test. Section 3 shows some applications of our general method. Section 4 reports the results of some Monte Carlo simulations to study the finite-sample performance of our proposed tests. In Section 5 we apply our tests to the Nikkei Index data. In Section 6 we conclude. An Appendix contains the mathematical proofs of our results.

2. AN AUTOMATIC PORTMANTEAU TEST

In the sequel, we simplify the notations as follows: $u_t \equiv u_t(\theta_0)$ and

$$\widehat{u}_t = G_t(Y_t, \hat{\Omega}_{t-1}, \widehat{\theta}_n),$$

where $\widehat{\theta}_n$ is a $\sqrt{n}$-consistent estimator for θ_0, e.g the conditional maximum likelihood estimator. Let $\|\cdot\|$ denote the Euclidean norm, and let C be a generic constant that may change from expression to expression.

We aim to test

$$H_0 : Y_t|\Omega_{t-1} =^d G_t(\cdot, \Omega_{t-1}, \theta_0) \text{ almost surely (a.s.) for some } \theta_0 \in \Theta \subset \mathbb{R}^p, \tag{1}$$

where $=^d$ stands for equality in distribution. By a well-known property, dated back at least to Rosenblatt (1952), the null H_0 implies that

$$u_t = u_t(\theta_0) = G_t(Y_t, \Omega_{t-1}, \theta_0) \text{ is } iid \ U[0,1] \text{ distributed,}^3 \tag{2}$$

[3] Note that (2) is a necessary but not sufficient condition for (1). H_0 is actually characterized as $u_t|\Omega_{t-1} =^d U[0,1]$, which implies not only that u_t is *iid* but that u_t is independent of Ω_{t-1}.

which further implies that

$$cov(\varphi(u_t), \varphi(u_{t-j})) = 0, \ \forall j \geq 1, \tag{3}$$

where we focus on φ being a cadlag (right continuous with left limits) function on $[0, 1]$ that is of bounded variation or nondecreasing. This choice of φ already has many interesting applications, see the next section for details.

In this section, we construct an automatic Portmanteau test based on (3). Define the lag-j autocovariance and autocorrelation by

$$\begin{aligned} \gamma_j &= Cov(\varphi(u_t), \varphi(u_{t-j})) = E[(\varphi(u_t) - c_\varphi)(\varphi(u_{t-j}) - c_\varphi)] \text{ and} \\ \rho_j &= \frac{\gamma_j}{\gamma_0}, \end{aligned}$$

respectively, where $c_\varphi = E[\varphi(u_t)]$. The sample counterpart of γ_j based on a sample $\{u_t\}_{t=1}^n$ is

$$\gamma_{nj} = \frac{1}{n-j} \sum_{t=1+j}^{n} (\varphi(u_t) - c_\varphi)(\varphi(u_{t-j}) - c_\varphi).$$

However, in our present context $\{u_t\}_{t=1}^n$ is unobservable, as θ_0 is unknown and Ω_{t-1} is not completely observed. Then we substitute $\widehat{u}_t$ for u_t in γ_{nj} and obtain

$$\widehat{\gamma}_{nj} = \frac{1}{n-j} \sum_{t=1+j}^{n} (\varphi(\widehat{u}_t) - c_\varphi)(\varphi(\widehat{u}_{t-j}) - c_\varphi), \qquad j \geq 0,$$

and $\widehat{\rho}_{nj} = \widehat{\gamma}_{nj}/\widehat{\gamma}_{n0}$. The asymptotic distribution of $\widehat{\rho}_{nj}$ would generally be different to that of ρ_{nj} due to the parameter estimation effect. As a result classical Box-Pierce tests applied to $\widehat{\rho}_{nj}$ no longer have a chi-squared distribution. The next theorem addresses this problem under the following conditions:

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

Assumption A2: There is a $\sqrt{n}$-consistent estimator $\widehat{\theta}_n$ of the parameter θ_0, where θ_0 is in the interior point of Θ. Moreover, $\widehat{\theta}_n$ satisfies the following asymptotic (Bahadur) expansion under H_0,

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} l(Y_t, \Omega_{t-1}, \theta_0) + o_p(1),$$

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

Assumption A3: The effect of information truncation satisfies

$$\sup_{\theta \in \Theta_0} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left| G_t(Y_t, \widehat{\Omega}_{t-1}, \theta) - G_t(Y_t, \Omega_{t-1}, \theta) \right| = o_P(1),$$

where Θ_0 is an arbitrary neighborhood of $\theta_0 \in \Theta$, and $\widehat{\Omega}_{t-1}$ is the observed information set at time $t-1$.

Assumption A4: $F_t(\theta, x) \equiv \Pr[u_t(\theta) \leq x | \Omega_{t-1}] = G_t\left(G_t^{-1}(x, \Omega_{t-1}, \theta), \Omega_{t-1}, \theta_0\right)$ is continuously differentiable in θ and x a.s.. Moreover, $\left\| \frac{\partial^2 F_t(\theta_0, x)}{\partial x \partial \theta} \right\| = O_P(1)$,

$$E\left[\sup_{\theta \in \Theta_0, 0 \leq x \leq 1} \left\| \frac{\partial F_t(\theta, x)}{\partial x} \right\| \right] < C, \text{ and } \int_0^1 \sup_{\theta \in \Theta_0} \left\| \frac{\partial F_t(\theta_0, x)}{\partial \theta} \right\| dx = O_P(1).$$

Assumption A1 is made here for easy exposition. Our results are also valid for some non-stationary and non-ergodic sequences, see e.g. Escanciano (2007). Assumption A2 is satisfied by most commonly used estimators, such as the (quasi-)maximum likelihood estimator and the generalized method of moments estimator, see e.g. Bose (1998) and Wu (2007). Assumption A3 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). This assumption is not needed when the process is Markovian. Assumption A4 is required for the asymptotic tightness[4] of certain empirical processes, the uniform law of large numbers, integration by parts and the interchange of derivative and integral. With these assumptions in place we establish the null limit distribution of $\widehat{\rho}_n^{(m)} := (\widehat{\rho}_{n1}, \widehat{\rho}_{n2} ... \widehat{\rho}_{nm})'$ in the next Theorem.

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

$$\sqrt{n}\widehat{\rho}_n^{(m)} \longrightarrow^d N(0, \Sigma)$$

with the ij–th element of Σ given by

$$\Sigma_{ij} = \delta_{ij} + V_\varphi^{-2} R_i' S_j + V_\varphi^{-2} R_j' S_i + V_\varphi^{-2} R_i' E[l_t l_t'] R_j, \tag{4}$$

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

where $V_{\varphi} := var(\varphi(u_t))$, $R_{j,\alpha} = \left.\frac{\partial E[\varphi(u_t(\theta))(\varphi(u_{t-j})-c_{\varphi})]}{\partial \theta}\right|_{\theta_0}$, $S_j = E[(\varphi(u_t) - c_{\varphi})((\varphi(u_{t-j}) - c_{\varphi}))l_t]$, $l_t = l(Y_t, \Omega_{t-1}, \theta_0)$ *and* δ_{ij} *is the Kronecker delta function, which takes value 1 if* $i = j$, *and 0 otherwise.*

Notice that the Box-Pierce test statistic $n\widehat{\rho}_n^{(m)\prime}\widehat{\rho}_n^{(m)}$ is no longer chi-squared distributed when Σ is not the identity matrix. To get an asymptotically distribution-free test, we need take into account the parameter estimation effect. To that end, we consider a modified Box-Pierce test at lag m as follows

$$M_n(m) := n\widehat{\rho}_n^{(m)\prime}\widehat{\Sigma}^{-1}\widehat{\rho}_n^{(m)}, \tag{5}$$

where $\widehat{\Sigma}$ is a consistent estimator for Σ, i.e. $\widehat{\Sigma}_{ij} = \delta_{ij} + V_{\varphi}^{-2}\widehat{R}_i^{\prime}\widehat{S}_j + V_{\varphi}^{-2}\widehat{R}_j^{\prime}\widehat{S}_i + V_{\varphi}^{-2}\widehat{R}_i^{\prime}W_n\widehat{R}_j$, $W_n = n^{-1}\sum_{t=1}^{n}\widehat{l}_t\widehat{l}_t^{\prime}$, $\widehat{l}_t = l(Y_t, \Omega_{t-1}, \widehat{\theta}_n)$,

$$\widehat{R}_j = (n-j)^{-1} \left. \frac{\partial\left(\sum_{t=j+1}^{n}\varphi(u_t(\theta))(\varphi(\widehat{u}_{t-j}) - c_{\varphi})\right)}{\partial\theta}\right|_{\widehat{\theta}_n},$$

and $\widehat{S}_j = (n-j)^{-1}\sum_{t=j+1}^{n}(\varphi(\widehat{u}_t) - c_{\varphi})(\varphi(\widehat{u}_{t-j}) - c_{\varphi})\widehat{l}_t$. The next corollary gives the null limit distribution of $M_n(m)$ under the following additional assumptions:

Assumption A5: $l(Y_t, \Omega_{t-1}, \theta)$ is continuous in θ a.s., satisfying

$$E\left[\sup_{\theta\in\Theta} \| \, l(Y_t, \Omega_{t-1}, \theta) \, \|^2\right] < C.$$

Assumption A6: Σ is non-singular.

Corollary 1 *Under Assumptions A1-A6 and* H_0, $M_n(m) \longrightarrow^d \chi_m^2$.

There are at least two practical limitations of Portmanteau tests $M_n(m)$. First, inference with $M_n(m)$ can be sensitive to the selected number of autocorrelations m. To overcome this limitation, we propose a fully automatic Portmanteau test where m is not fixed but selected automatically from the data. We follow the suggestion in Inglot and Ledwina (2006a, b) and use a combination of Akaike (1974) and Schwarz (1978) criteria, see below for details.

Second, the test statistic $M_n(m)$ is unstable sometimes because $\widehat{\Sigma}$ is close to singular for some parameter values and moderate values of m, as we observed with some simulations.

These limitations motivate our data-driven test statistic given by

$$Q_n^* = Q_n(m^*), \tag{6}$$

where

$$Q_n(m) := n\widehat{\Sigma}_{11}^{-1}\widehat{\rho}_n^{(m)\prime}\widehat{\rho}_n^{(m)},$$

$$m^* = \min\{m : 1 \leq m \leq p; L_n(m) \geq L_n(h), h = 1, 2, ...p\},$$

$$L_n(m) = Q_n(m) - \pi_n(m, q),$$

p is an arbitrarily large but fixed upper bound,

$$\pi_n(m, q) = \begin{cases} m \log n, & \text{if} \quad \max_{1\leq j\leq p} \sqrt{n}\left|\widehat{\rho}_{nj}\right| \leq \sqrt{q \log n}, \\ 2m, & \text{if} \quad \max_{1\leq j\leq p} \sqrt{n}\left|\widehat{\rho}_{nj}\right| > \sqrt{q \log n}. \end{cases}$$

and q is some fixed positive number.

As explained in Inglot and Ledwina (2006a, b), the motivation of this selection rule for m is to combine the advantages of the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). On the one hand, tests constructed using the BIC criterion are able to properly control the type I error and are more powerful when the serial correlation is present in the first order autocorrelations. On the other hand, tests based on the AIC cannot properly control the type I error, but they are more powerful when the serial correlation is present in high order autocorrelations. This selection rule for m allows the data to choose the preferable criterion according to the data characteristics.

Extensive simulations in the literature suggest that the choice of $q = 2.4$ works well in finite samples; see Inglot and Ledwina (2006a, b) and Escanciano and Lobato (2009). Note that a small value for q would lead to the use of the AIC criterion, while a large q would lead to the choice of the BIC criterion. Moderate values, such as 2.4, provide a "switching effect" in which one combines the advantages of the two selection rules.

There are precedents of our data-driven test in the literature. Kallenberg and Ledwina (1999) suggested a data-driven test for independence of raw data, and Escanciano and

Lobato (2009) considered a data-driven Portmanteau test for serial correlation of raw data. In this paper, our data-driven test is applied to estimated generalized residuals rather than raw data, which turns out to have an important distinction. The parameter estimation effects in our setting destroy the covariance structure and the orthogonality properties of the components of the smooth tests in Kallenberg and Ledwina (1999) and Escanciano and Lobato (2009), which highly complicates the application of their results.

Our next theorem proves the asymptotic null distribution of the data-driven Portmanteau test.

Theorem 2 *Under Assumptions A1-A5 and* H_0, $Q_n^* \longrightarrow^d \chi_1^2$.

Remark 1: Under the null H_0 the generalized errors $\{u_t\}$ are uncorrelated, thus $Q_n(m)$ is "small" and due to the monotonicity of the penalization in m, m^* takes the minimum possible value with probability tending to one, which is $m^* = 1$. This explains the χ_1^2 limit distribution. Also because the data-driven method tends to choose $m^* = 1$ with probability 1, one only needs to weight $Q_n(m)$ by $\widehat{\Sigma}_{11}^{-1}$ for the data-driven statistic Q_n^* to have a standard limit distribution.

Remark 2: Although in this paper we do not study formally the power performance of our test, a similar analysis to that carried out in Escanciano and Lobato (2009) shows that our test is consistent against alternatives for which $\rho_k \neq 0$ for some $k \leq p$.

3. APPLICATIONS

In this section we show some applications of our general test developed above. By choosing φ's properly, we get a simple yet powerful goodness-of-fit test for the conditional distribution, as well as a test for the correct specification of the conditional *tail* distribution. We further show some other applications of the proposed method.

3.1. Test for Conditional Distribution

To test the null hypothesis (1), a natural and simple choice of φ is the identity function I, i.e.

$$\varphi(x) = I(x) = x.$$

Theorem 1, as well as the modified Box-Pierce test (5) and the data-driven test (6), can be applied directly to this setup with $c_\varphi = E(u_t) = 0.5$ and $V_\varphi = var(u_t) = 1/144$.

The simulations and real data applications below show that this simple choice of φ has satisfactory empirical power against a wide variety of fixed alternatives. Our general method can be applied to other choices of φ as well, for example $\varphi(x) = x^2$, which can be useful in detecting nonlinear dependence in $\{u_t\}$.

3.2. Test for Conditional Tail Distribution

In risk management people are often interested in the tail instead of the whole distribution. One example is Value-at-Risk (VaR), which is a widely used risk measure in commercial banks and other financial institutions, and it is defined as the maximum loss on a portfolio that can be expected over a given horizon with a certain degree of confidence. VaR is commonly modeled as the appropriate left tail quantile of the portfolio return distribution, see e.g. Christoffersen (2009), and hence depends crucially on the correct specification of the conditional tail distribution. Other examples include expected shortfall (the expected loss beyond the VaR) and stress testing, where people pay special attention to the conditional tail distribution.

We haven't seen in the literature any goodness-of-fit test specially designed for the conditional tail distribution, which is important as lack-of-fit of the whole distribution does not necessarily imply lack-of-fit of the tail distribution, see Section 5 for example. We propose such a test in this subsection simply by applying the general method in Section 2 to a special φ. Specifically, now we want to test the null hypothesis

$$H_{0r} : Y_t|\Omega_{t-1} =^d G_t(y, \Omega_{t-1}, \theta_0) \text{ for all } y \le c(\Omega_{t-1}, \theta_0),$$

for some $\theta_0 \in \Theta \subset \mathbb{R}^p$ and some threshold c, possibly depending on Ω_{t-1} and θ_0. A prominent example is $c(\Omega_{t-1}, \theta_0) = G_t^{-1}(\alpha, \Omega_{t-1}, \theta_0)$, the conditional α-th quantile or VaR model, e.g. at $\alpha = 0.01$. Henceforth, we only consider this example for simplicity in the exposition.

Notice that

$$\begin{aligned} G_t(Y_t \wedge G_t^{-1}(\alpha, \Omega_{t-1}, \theta_0), \Omega_{t-1}, \theta_0) &= u_t 1(u_t \le \alpha) + \alpha 1(u_t > \alpha) \\ &\equiv \varphi(u_t), \end{aligned}$$

where $a \wedge b = \min(a, b)$,

$$\varphi(x) = x1(x \leq \alpha) + \alpha 1(x > \alpha). \tag{7}$$

One can show that under H_{0r}

$$\Pr\left(\varphi(u_t) \leq u \mid \Omega_{t-1}\right) = u1(0 \leq u < \alpha) + 1(u = \alpha). \tag{8}$$

Hence, under H_{0r}, $\{\varphi(u_t)\}$ are *iid*, with marginal cdf given by (8). Therefore, Theorem 1, as well as the modified Box-Pierce test (5) and the data-driven test (6), can be applied with $c_\varphi = E[\varphi(u_t)] = \alpha(1 - \alpha/2)$ and $V_\varphi = var[\varphi(u_t)] = \alpha^3(1/3 - \alpha/4)$.

3.3. Other Extensions

In this subsection we illustrate the wide applicability of our proposed method by providing some other possible extensions. We show how our method can be used for testing the correct specification of conditional hazard models and continuous time models.

Conditional Hazard Models. We can test that the cumulative hazard funtion of Y_t given Ω_{t-1} is $\Lambda(\cdot, \Omega_{t-1}, \theta_0)$ by using the fact that under the correct specification

$$u_t = u_t(\theta_0) = \Lambda(Y_t, \Omega_{t-1}, \theta_0) \text{ is } iid \ \exp(1) \text{ distributed.}$$

One can simply let $\varphi = I$, and then apply Theorem 1, the modified Box-Pierce test (5) and the data-driven test (6) directly with $c_\varphi = E(u_t) = 1$ and $V_\varphi = var(u_t) = 1$.

Conditional Mean in Diffusion Models. Many continuous time models used in economics and finance have the form

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

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 (1965), Dubins and Schwarz (1965)) theorem implies that for any fixed positive number $\triangle$

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

are *iid* normal. Once again, one can apply our general method in Section 2 and hence derive a specification test for continuous time models.[5]

4. MONTE CARLO STUDY

To assess the finite sample performance of our proposed tests, especially for the choices of φ's in Subections 3.1 and 3.2, we carry out some Monte Carlo studies.

4.1. Conditional Distribution

As described in Subsection 3.1, one can apply our method to test the correct specification of a conditional distribution with $\varphi = I$. In the simulations, we test whether $Y_t|\Omega_{t-1}$ follows GARCH-type models with innovations being (1) $N(0,1)$ and (2) Hansen's (1994) skewed t, denoted by $GT(\cdot|\eta_t, \lambda_t)$ with η_t and λ_t depending on Ω_{t-1}.[6] All the simulations are done through the Quarry High Performance Cluster at Indiana University.

We first test whether $Y_t|\Omega_{t-1}$ follows an AR(1)-GARCH(1,1) model with normal innovations. For comparison purpuses, we use the same setup as Koul and Ling (2006, KL hereinafter)

$$Y_t = a_0 Y_{t-1} + \sigma_t \varepsilon_t, \quad \sigma_t^2 = \omega_0 + \alpha_0 \sigma_{t-1}^2 \varepsilon_{t-1}^2 + \beta_0 \sigma_{t-1}^2,$$

where the true parameters are $\theta_0 := (a_0, \omega_0, \alpha_0, \beta_0) = (0.5, 0.025, 0.25, 0.5)$. In the experiments, we generate 1000 realizations of $\{Y_t\}_{t=1}^n$, with $n = 400$ and 1000,[7] and the corresponding upper bounds $p = 15$ and 30, respectively. (We tried $p = 15, 20, 30$ and 40, and

[5] In this case, one can show that under some mild regulatory conditions on $\mu(\cdot, \cdot)$ and U_t, see e.g. Chang, Nguyen and Park (2009) for details, the estimation effect from estimating the quardratic variation is uniformly negligible.

[6] The density function of Hansen's skewed t distribution with fixed η and λ is given by $gt(z|\eta, \lambda) = \left[\begin{smallmatrix} bc\left(1+\frac{1}{\eta-2}(\frac{bz+a}{1-\lambda})^2\right)^{-(\eta+1)/2} & z<-a/b \\ bc\left(1+\frac{1}{\eta-2}(\frac{bz+a}{1+\lambda})^2\right)^{-(\eta+1)/2} & z\geq -a/b, \end{smallmatrix}\right.$ where $2 < \eta < \infty$, $-1 < \lambda < 1$, $a = 4\lambda c\left(\frac{\eta-2}{\eta-1}\right)$, $b^2 = 1 + 3\lambda^2 - a^2$, and $c = \Gamma((\eta+1)/2)/(\sqrt{\pi(\eta-2)}\Gamma(\eta/2))$.

[7] Large sample sizes such as $n = 400$ or 1000 are common in financial applications.

they all give us similar size and power performance.) Then we estimate the parameters by the maximum likelihood (MLE) method, and obtain $\widehat{u}_t = \Phi(\widehat{\varepsilon}_t)$, where $\widehat{\varepsilon}_t = (Y_t - \widehat{a}_n Y_{t-1})/\widehat{\sigma}_t$, Φ is the cumulative distribution function (cdf) of a standard normal; and $\widehat{\sigma}_t$ is as σ_t with estimated parameters replacing the true parameters.

Once we have $\widehat{u}_t$, we can easily calculate $\widehat{\rho}_{nj}$. To get our test statistics $M_n(m)$ and Q_n^*, we then need $\widehat{\Sigma}$, which reduces to $\widehat{R}_i$ and $\widehat{l}_t$. Some algebra shows that here $\partial u_t/\partial\theta = -\phi(\varepsilon_t)((Y_{t-1}/\sigma_t, 0, 0, 0)' + \varepsilon_t/\sigma_t \cdot \partial\sigma_t/\partial\theta)$, where ϕ is the density function of a standard normal and

$$\frac{\partial\sigma_t(\theta)}{\partial\theta} = \frac{0.5}{\sigma_t(\theta)}\sum_{j=1}^{t}\beta^{j-1}\left(-2\alpha(Y_{t-j} - aY_{t-j-1})Y_{t-j-1}, \frac{\beta^{1-j}}{t(1-\beta)}, (Y_{t-j} - aY_{t-j-1})^2, \sigma_{t-j}^2(\theta)\right)'.$$

Therefore

$$\begin{aligned}\widehat{R}_i &= \frac{1}{n}\sum_{t=1}^{n}\frac{\partial u_t(\widehat{\theta}_n)}{\partial\theta}(\widehat{u}_{t-i} - 0.5)\\ &= -\frac{1}{n}\sum_{t=1}^{n}(\widehat{u}_{t-i} - 0.5)\phi(\widehat{\varepsilon}_t)\left((\frac{Y_{t-1}}{\widehat{\sigma}_t}, 0, 0, 0)' + \frac{\widehat{\varepsilon}_t}{\widehat{\sigma}_t}\frac{\partial\sigma_t(\widehat{\theta}_n)}{\partial\theta}\right).\end{aligned}$$

Similarly,

$$\widehat{l}_t = l(Y_t, \Omega_{t-1}, \widehat{\theta}_n) = \widehat{H}^{-1}\left((1 - \widehat{\varepsilon}_t^2)\frac{1}{\widehat{\sigma}_t}\frac{\partial\sigma_t(\widehat{\theta}_n)}{\partial\theta} - \widehat{\varepsilon}_t(\frac{Y_{t-1}}{\widehat{\sigma}_t}, 0, 0, 0)'\right),$$

where $\widehat{H}$ is the Hessian matrix of the MLE objective function. $\widehat{S}_j$ is simply $n^{-1}\sum_{t=1}^{n}(\widehat{u}_t - 0.5)(\widehat{u}_{t-j} - 0.5)\widehat{l}_t$. With these components, one can calculate $\widehat{\Sigma}$ and hence $M_n(m)$ and Q_n^*.

We compare our tests with KL, which is one of the best available tests in this setup and is computed as

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

with $\widehat{\mathcal{I}}_n := 1/n\sum_{t=1}^{n}W_t(\widehat{\theta}_n)W_t(\widehat{\theta}_n)'$, $K_n(x, \widehat{\theta}_n) := 1/\sqrt{n}\sum_{t=1}^{n}W_t(\widehat{\theta}_n)[I(\widehat{\varepsilon}_t \le x) - \Phi(x)]$ and $W_t(\widehat{\theta}_n) = (\partial\widehat{\sigma}_t/\partial\omega, \partial\widehat{\sigma}_t/\partial\alpha, \partial\widehat{\sigma}_t/\partial\beta)/\widehat{\sigma}_t$.

We consider the following data generating processes:

H_0: AR(1)-GARCH(1,1) model (*AR1-GH*):

$$\begin{aligned}Y_t &= 0.5Y_{t-1} + v_t, \quad v_t = \sigma_t\varepsilon_t,\\ \sigma_t^2 &= 0.025 + 0.25v_{t-1}^2 + 0.5\sigma_{t-1}^2,\end{aligned} \tag{9}$$

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*): $Y_t = 0.2Y_{t-1} + 0.7v_{t-1} + v_t$.

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: Non-Linear Moving Average model (*NLMA*): $Y_t = 0.8\varepsilon_{t-1}^2 + \varepsilon_t$.

A_6: *ARFIMA*(0,d,0) model: $(1-B)^{0.3}Y_t = \varepsilon_t,\ \ BY_t = Y_{t-1}$.

A_7: Exponential Autoregressive model (*EXP*): $Y_t = 0.5Y_{t-1}\exp(-0.5Y_{t-1}^2) + \varepsilon_t$.

A_8: AR(1)-GARCH(1,1) model with standardized Student's t_5 innovations (*AR1-GH-T*).

A_9: 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*).

In models A_1 and A_2, v_t is defined as in (9), 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 (2010); whereas A_3, A_6 and A_7 are used in Hong (2000).

Table 1 gives the empirical sizes and size-corrected powers of the tests at 5% level. Our data-driven test Q_n^* has excellent size performance, while $M_n(9)$ has some size distortion because $\widehat{\Sigma}$ is close to singular sometimes. For A_1, A_4, A_6 and A_7 our tests have much better power than KL. For A_2 and A_5 both Q_n^* and KL have very high power. KL outperforms our test for A_3. For A_8 one can actually show that our test has no power, as under this alternative the resulting generalized errors are still *iid*. For A_9 the power of our test hinges on whether the mean of u_t equals 0.5. KL is specially designed for alternatives like A_8 and A_9,[8] and therefore it has better power than ours for these two alternatives. In summary, except A_8 and A_9, our test Q_n^* has satisfactory size and power performance.

[8] KL focuses on testing the correct specification of the innovation (ε_t) distribution, while assuming that the conditional mean and variance are correctly specified.

Table 1. Size and Power of M_n, Q_n^* ($\varphi = I$) and KL at 5%

	$n=400$				$n=1000$			
	$M_n(1)$	$M_n(9)$	Q_n^*	KL	$M_n(1)$	$M_n(9)$	Q_n^*	KL
	Size							
H_0 :*AR1-GH*	.055	.080	.057	.063	.047	.069	.049	.053
	Power, Size-Corrected							
A_1 :*AR2-GH*	.998	.952	.999	.061	1.00	1.00	1.00	.082
A_2 :*ARMA-GH*	.975	.874	.993	1.00	1.00	.831	1.00	1.00
A_3 :*TAR*	.243	.282	.293	.996	.446	.647	.542	1.00
A_4 :*BIL*	.288	.284	.316	.072	.676	.651	.695	.111
A_5 :*NLMA*	.423	.990	.967	1.00	.699	1.00	.999	1.00
A_6 :*ARFIMA*	.696	.543	.754	.479	.943	.721	.968	.636
A_7 :*EXP*	.196	.166	.201	.046	.516	.368	.515	.065
A_8 :*AR1-GH-T*	.042	.053	.050	.172	.069	.053	.073	.585
A_9 :*AR1-GH-KT*	.066	.119	.083	.865	.079	.137	.090	1.00

Next we apply our test to Hansen (1994) skewed t model. We test $H_0 : Y_t|\Omega_{t-1}$ follows a AR(1)-GARCH(1,1) model with $GT(\cdot|\eta_t, \lambda_t)$ innovations (*AR1-GH-H*), i.e.

$$\begin{aligned} Y_t &= a_0 Y_{t-1} + v_t, \quad v_t = \sigma_t \varepsilon_t, \\ \sigma_t^2 &= \omega_0 + \alpha_0 \sigma_{t-1}^2 \varepsilon_{t-1}^2 + \beta_0 \sigma_{t-1}^2, \\ \varepsilon_t &\sim GT(\cdot|\eta_t, \lambda_t), \quad \eta_t = g_{(2.1,30)}(\widetilde{\eta}_t), \quad \lambda_t = g_{(-0.9,0.9)}(\widetilde{\lambda}_t), \\ \widetilde{\eta}_t &= a_1 + b_1 Y_{t-1} + c_1 Y_{t-1}^2, \quad \widetilde{\lambda}_t = a_2 + b_2 Y_{t-1} + c_2 Y_{t-1}^2, \end{aligned} \tag{10}$$

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 $(a_2, b_2, c_2) = (-0.1, -0.15, -0.1)$, and generate 1000 realizations of $\{Y_t\}_{t=1}^n$, with $n =$ 200 and 400, and the associated upper bounds for Q_n^* are $p = 15$. Then we estimate the parameters by MLE. After obtaining the generalized residuals $\widehat{u}_t$, we calculate the test statistics $M_n(m)$ and Q_n^*.[9] Notice KL does not apply to this model, where the distribution of innovations $\{\varepsilon_t\}$ is also time-varying. For the power we consider alternatives $A_1 - A_7$.

Table 2 reports the empirical sizes and size-corrected powers of the tests at 5% level. We observe that the sizes of our tests are fairly close to the nominal size, and the power performance is satisfactory.

Table 2. Size and Power of M_n and Q_n^* ($\varphi = I$) at 5%

	$n = 200$			$n = 400$		
	$M_n(1)$	$M_n(9)$	Q_n^*	$M_n(1)$	$M_n(9)$	Q_n^*
	Size					
H_0 : *AR1-GH-H*	.032	.079	.051	.034	.046	.041
	Power, Size-Corrected					
A_1 : *AR2-GH*	.831	.645	.855	.985	.961	.989
A_2 : *ARMA-GH*	.833	.943	.967	.986	.999	.996
A_3 : *TAR*	.094	.094	.108	.094	.167	.131
A_4 : *BIL*	.123	.111	.122	.201	.236	.207
A_5 : *NLMA*	.147	.567	.300	.197	.857	.415
A_6 : *ARFIMA*	.221	.328	.380	.492	.724	.597
A_7 : *EXP*	.068	.058	.075	.128	.130	.131

4.2. Conditional Tail Distribution

Now we evaluate the finite sample performance of our proposed tests for the correct specification of the conditional tail distribution, i.e. we test

$$H_{0r} : Y_t|\Omega_{t-1} =^d G_t(y, \Omega_{t-1}, \theta_0) \text{ for all } y \leq G_t^{-1}(\alpha, \Omega_{t-1}, \theta_0),$$

for some $\theta_0 \in \Theta \subset \mathbb{R}^p$. As described in Subsection 3.2, one can apply our general method for this purpose with φ given in (7).

[9] Some algebra shows that in this case

$$\frac{\partial u_t(\theta)}{\partial \theta} = \begin{pmatrix} gt(\varepsilon_t|\eta_t, \lambda_t)(-\frac{Y_{t-1}}{\sigma_t} - \varepsilon_t \frac{\dot{\sigma}_t}{\sigma_t}) \\ \frac{\partial GT(\varepsilon_t|\eta_t,\lambda_t)}{\partial \eta_t} \dot{\eta}_t \\ \frac{\partial GT(\varepsilon_t|\eta_t,\lambda_t)}{\partial \lambda_t} \dot{\lambda}_t \end{pmatrix}$$

where $\dot{\eta}_t = (\frac{\partial \eta_t}{\partial a_1}, \frac{\partial \eta_t}{\partial b_1}, \frac{\partial \eta_t}{\partial c_1})'$ and $\dot{\lambda}_t = (\frac{\partial \lambda_t}{\partial a_2}, \frac{\partial \lambda_t}{\partial b_2}, \frac{\partial \lambda_t}{\partial c_2})'$.

We consider the following null data generating process for G_t:

H_{0r}: AR(1)-GARCH(1,1) model with standardized Student's innovations (*AR1-GH-T*):

$$\begin{aligned} Y_t &= 0.05Y_{t-1} + v_t, \quad v_t = \sigma_t \varepsilon_t, \\ \sigma_t^2 &= 0.05 + 0.1v_{t-1}^2 + 0.85\sigma_{t-1}^2. \end{aligned} \tag{11}$$

The parameter values we choose represent typical parameter values in empirical applications. For the power of the tests, we consider the following popular alternative data generating processes for G_t:

A_{1r}: AR(2)-GARCH(1,1) model (*AR2-GH*): $Y_t = 0.05Y_{t-1} + 0.3Y_{t-2} + v_t$, with v_t given in (11).

A_{2r}: AR(1)-ARCH(2) model (*AR1-AH2*): $Y_t = 0.05Y_{t-1} + v_t$, $v_t = \sigma_t \varepsilon_t$, $\sigma_t^2 = 0.1 + 0.1v_{t-1}^2 + 0.8v_{t-2}^2$.

A_{3r}: AR(1)-EGARCH(1,1) model (*AR1-EG*): $Y_t = 0.05Y_{t-1} + v_t$, $v_t = h_t \varepsilon_t$, $lnh_t^2 = 0.01 + 0.9lnh_{t-1}^2 + 0.3(|\varepsilon_{t-1}| - \sqrt{2/\pi}) - 0.8\varepsilon_{t-1}$.

A_{4r}: GARCH in Mean model (*GH-M*): $Y_t = 2.5\sigma_t^2 + v_t$, $v_t = \sigma_t \varepsilon_t$, $\sigma_t^2 = 0.01 + 0.29v_{t-1}^2 + 0.7\sigma_{t-1}^2$..

A_{5r}: Bilinear model (*BIL*): $Y_t = 0.05Y_{t-1} + 0.7\varepsilon_{t-1}Y_{t-2} + \varepsilon_t$.

A_{6r}: Non-Linear Moving Average model (*NLMA*): $Y_t = 0.8\varepsilon_{t-1}^2 + \varepsilon_t$.

A_{7r}: AR(1)-GARCH(1,1) model with mixed normal innovations (*AR1-GH-MN*): Y_t is as in (11), with $\varepsilon_t \sim [0.5 \cdot N(-3,1) + 0.5N(-3,1)]/\sqrt{10}$.

$\{\varepsilon_t\}$ is *iid* with a standardized Student's t_5 distribution in all the models except A_{7r}. Models $A_{3r}-$ A_{6r} are studied in Escanciano and Olmo (2010), whereas A_{7r} is used in Koul and Ling (2006).

Here we report the results for testing the lower 5% conditional tail distribution. Table 3 gives the empirical sizes and size-corrected powers of the tests at 5% level. We observe some size distortion for $M_n(9)$ as $\widehat{\Sigma}$ is close to singular sometimes. The sizes of our data-driven test Q_n^* are fairly close to the nominal level. Meanwhile Q_n^* has satisfactory power performance against a wide range of alternatives.

Table 3. Size and Power of M_n and Q_n^* (φ given in (7)) at 5%

	$n = 1000$			$n = 2500$		
	$M_n(1)$	$M_n(9)$	Q_n^*	$M_n(1)$	$M_n(9)$	Q_n^*
	Size					
H_{0r} :*AR1-GH-T*	.061	.166	.076	.051	.114	.062
	Power, Size-Corrected					
A_{1r} :*AR2-GH*	.089	.134	.212	.121	.294	.390
A_{2r} :*AR1-AH2*	.041	.149	.276	.110	.471	.584
A_{3r} :*AR1-EG*	.360	.187	.295	.580	.378	.586
A_{4r} :*GH-M*	.103	.147	.356	.214	.377	.486
A_{5r} :*BIL*	.051	.094	.167	.100	.244	.323
A_{6r} :*NLMA*	.152	.271	.545	.337	.791	.747
A_{7r} :*AR1-GH-MN*	.940	.953	.949	.949	.988	.950

5. EMPIRICAL APPLICATION

In this section, we apply our tests to the daily Nikkei Index data, one of the major stock indices in the world. Our data are obtained from *Freelunch.com* over the period 01/2004-12/2006, with a total of 705 observations.

We first test the hypothesis H_0 : the log-return, Y_t, follows an AR(1)-GARCH(1,1) process with Student's t innovations t_v, where the degree of freedom parameter v is discrete and unknown, i.e.

$$H_0 : Y_t = a_0 Y_{t-1} + \sigma_t \varepsilon_t,\ \sigma_t^2 = \omega_0 + \alpha_0 \sigma_{t-1}^2 \varepsilon_{t-1}^2 + \beta_0 \sigma_{t-1}^2,\ \varepsilon_t \sim t_v.$$

Our estimates for $(a_0, \omega_0, \alpha_0, \beta_0)$ using Gaussian MLE are given in Table 4, together with the estimate for v based on the method of moments. The values in the parentheses are the corresponding asymptotic standard deviations of the estimated parameters. We then calculate our data-driven test statistic Q_n^* with $\varphi = I$ and the upper bound $p = 15, 20, 30, 40$ and 50. All these different values of p give us the same value for Q_n^*. For comparison, we also calculate Koul and Ling (2006) test KL.

Table 4. Parameter estimates and p-values of Q_n^* ($\varphi = I$) and KL

a_0	ω_0	α_0	β_0	v	p-value of Q_n^*	p-value of KL
.022	.0034	.065	.92	10	.019	0.28
(.039)	(.0021)	(.017)	(.019)			

Our test Q_n^* rejects H_0 at 5% level, while KL fails to reject even at 10% level. Notice that KL is only designed for testing $\varepsilon_t \sim t_v$, while assuming the mean and variance are correctly specified. It is not clear how KL behaves if the mean or variance is incorrectly specified. Moreover, like Bai (2003), KL is not able to detect whether ε_t is *iid* or not. Our test Q_n^* overcomes the above mentioned problems of KL, and as a result it is able to detect this alternative.

For a robustness check, we apply the tests proposed in Escanciano (2006) and Escanciano (2010) to see whether the conditional mean and variance terms are correctly specified. By the former test we fail to reject that the condition mean term is correctly specified even at 10% level, while by the latter test we reject the AR(1)-GARCH(1,1) specification at 5% level. So the specification of the GARCH(1,1) part might be problematic, and hence the performance of KL is unclear here. On the other hand, the results support our test Q_n^*.

In risk management, people are often more interested in the conditional tail distribution. Next we apply our test statistic Q_n^* with φ given in (7) for the null hypothesis

$$H_{0r} : Y_t|\Omega_{t-1} =^d G_t(y, \Omega_{t-1}, \theta_0) \text{ for all } y \leq G_t^{-1}(\alpha, \Omega_{t-1}, \theta_0),$$

where G_t is the conditional distribution of an AR(1)-GARCH(1,1) process with Student's t innovations. We try $\alpha = 10\%$, 5% and 1%, and the p-values of Q_n^* are given in Table 5.

Table 5. p-values of Q_n^* (φ given in (7)) for tail distribution

	lower 10%	lower 5%	lower 1%
p-value of Q_n^*	0.440	0.089	0.834

This application illustrates clearly that the lack-of-fit of the whole distribution does not necessarily imply lack-of-fit of the tail distribution. Specifically, although an AR(1)-GARCH(1,1) process with Student's t innovations may be a bad fit for the whole distribution of Y_t, the p-values in Table 5 suggest that it may not be a bad fit for the tail distribution, especially for the lower 10% and 1% tails. Therefore, if one is only interested in modeling the tail part, for example Value-at-Risk or expected shortfall, the specification of AR(1)-GARCH(1,1) processes with Student's t innovations may not be a bad choice.

6. CONCLUSION

In this paper, we propose an automatic Portmanteu test for conditional goodness-of-fit. Our method is based on the fact that under the correct specification of the conditional distribution the generalized errors obtained after the probability integral transformation are *iid* $U[0, 1]$. The proposed test is a modified Box-Pierce statistic applied to the generalized errors, with a data-driven choice for the number of autocorrelations used. It is simple to implement, and applicable to a wide class of models.

Finally, we discuss some suggestions for further research. The correct specification of the conditional distribution implies the generalized errors are independent, but the current test is only for linear independence. One way to detect nonlinear dependence is to apply our test to e.g. the squared generalized residuals. Extensions in this direction follow straightforwardly. Moreover, one can extend our test to the multivariate case and construct conditional goodness-of-fit tests for multivariate models. See e.g. Chitturi (1974), Hosking (1980, 1981) and Escanciano et al. (2010) for multivariate extensions of Box-Pierce test.

7. APPENDIX

To prove Theorem 1, we need the following lemmas. Define the processes

$$
\begin{aligned}
R_{nj}(x,y) &= \frac{1}{n-j}\sum_{t=1+j}^{n}\{1(u_t \le x)-x\}\{1(u_{t-j}\le y)-y\},\\
\widehat{R}_{nj}(x,y) &= \frac{1}{n-j}\sum_{t=1+j}^{n}\{1(\widehat{u}_t \le x)-x\}\{1(\widehat{u}_{t-j}\le y)-y\}.
\end{aligned}
$$

LEMMA A1: *Under Assumptions A1-A4 and H_0, we have*

$$\sup_{0\le x\le 1, 0\le y\le 1}\left|\sqrt{n-j}[\widehat{R}_{nj}(x,y)-R_{nj}(x,y)]-\sqrt{n}(\widehat{\theta}_n-\theta_0)'E_j(x,y)\right| = o_p(1),$$

where

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

Lemma A1 is a special case of Theorem 1 in Du (2010).

LEMMA A2: *Denote by $I[0,1]$ the collection of all cadlag (right continuous with left limits) functions on $[0,1]$ that are of bounded variation or nondecreasing. Let $R(x,y)$ be a function defined on $[0,1]^2$ such that $R(\cdot,y)\in I[0,1]$ for $0\le y\le 1$, $R(x,\cdot)\in I[0,1]$ for $0\le x\le 1$ and $R=0$ on the boundaries, and denote by $\ell([0,1]^2)$ the set of all such functions. Then the mapping*

$$R \to \int_0^1\int_0^1 \varphi(x)\varphi(y)R(dx,dy)$$

is continuous in R for any $\varphi\in I[0,1]$.

PROOF OF LEMMA A2:

By the Integration by Parts Theorem (Theorem 11, Shiryaev 1996, pp. 206) and the definition of R, we have

$$\int_0^1\int_0^1 \varphi(x)\varphi(y)R(dx,dy) = \int_0^1\int_0^1 R(x,y)\varphi(dx)\varphi(dy).$$

Notice that

$$
\begin{aligned}
&\left|\int_0^1\int_0^1 R_1(x,y)\varphi(dx)\varphi(dy) - \int_0^1\int_0^1 R_2(x,y)\varphi(dx)\varphi(dy)\right|\\
\le\ & \sup|R_1(x,y)-R_2(x,y)|\int_0^1\int_0^1 |\varphi(dx)\varphi(dy)|,
\end{aligned}
$$

for any $R_1, R_2 \in \ell([0,1]^2)$, and $\int |\varphi(dx)| < \infty$ as $\varphi \in I[0,1]$, and the proof is complete.

■

With the above two lemmas in place, we are ready to prove Theorem 1.

Proof of Theorem 1:

We first consider the case of no information truncation. This occurs if G_t depends only on a finite number of lagged Y_t and X_t.

Notice that

$$\begin{aligned}
\sqrt{n-j}\widehat{\gamma}_{nj} &= \sqrt{n-j}\int_0^1\int_0^1 \varphi(x)\varphi(y)\widehat{R}_{nj}(dx,dy),, \\
\sqrt{n-j}\gamma_{nj} &= \sqrt{n-j}\int_0^1\int_0^1 \varphi(x)\varphi(y)R_{nj}(dx,dy).
\end{aligned}$$

By Lemma A1 and A2, we have

$$\sqrt{n-j}\widehat{\gamma}_{nj} = \sqrt{n-j}\gamma_{nj} + \sqrt{n}(\widehat{\theta}_n - \theta_0)'\int_0^1\int_0^1 \varphi(x)\varphi(y)E_j(dx,dy) + o_p(1),$$

where

$$\begin{aligned}
\int_0^1\int_0^1 \varphi(x)\varphi(y)E_j(dx,dy) &= \int_0^1\int_0^1 E\left\{\varphi(x)\frac{\partial^2 F_t(\theta_0,x)}{\partial x\partial\theta}\varphi(y)\frac{\partial[I(u_{t-j}\le y)-y]}{\partial y}dxdy\right\} \\
&= \int_0^1 E\left\{\varphi(x)\frac{\partial^2 F_t(\theta_0,x)}{\partial x\partial\theta}(\varphi(u_{t-j})-c_\varphi)dx\right\} \\
&= E\left\{\int_0^1 \varphi(x)\frac{\partial^2 F_t(\theta_0,x)}{\partial x\partial\theta}dx\ (\varphi(u_{t-j})-c_\varphi)\right\} \\
&= E\left\{\left[\varphi(x)\frac{\partial F_t(\theta_0,x)}{\partial\theta}|_0^1 - \int_0^1 \frac{\partial F_t(\theta_0,x)}{\partial\theta}d\varphi(x)\right]\ (\varphi(u_{t-j})-c_\varphi)\right\} \\
&= -E\left\{\int_0^1 \frac{\partial F_t(\theta_0,x)}{\partial\theta}d\varphi(x)\ (\varphi(u_{t-j})-c_\varphi)\right\}
\end{aligned}$$

The above integration by parts follows from Theorem 11 of Shiryaev (1996, pp. 206) and Assumption A4. We can also interchange the derivative and integral under Assupmtion A4 and get

$$
\begin{aligned}
\int_0^1 \frac{\partial F_t(\theta_0, x)}{\partial \theta} d\varphi(x) &= \left.\frac{\partial}{\partial \theta}\right|_{\theta_0} \int_0^1 G_t\left(G_t^{-1}(x, \Omega_{t-1}, \theta), \Omega_{t-1}, \theta_0\right) d\varphi(x) \\
&= \left.\frac{\partial}{\partial \theta}\right|_{\theta_0} \int_{-\infty}^{\infty} G_t\left(y, \Omega_{t-1}, \theta_0\right) d\varphi(G_t\left(y, \Omega_{t-1}, \theta\right)) \\
&= \left.\frac{\partial}{\partial \theta}\right|_{\theta_0} \left(\varphi(1) - \int_{-\infty}^{\infty} \varphi(G_t\left(y, \Omega_{t-1}, \theta\right)) dG_t\left(y, \Omega_{t-1}, \theta_0\right)\right) \\
&= \left.\frac{\partial}{\partial \theta}\right|_{\theta_0} \left(-E[\varphi(u_t(\theta))|\Omega_{t-1}]\right) \\
&= -\frac{\partial E[\varphi(u_t(\theta_0))|\Omega_{t-1}]}{\partial \theta},
\end{aligned}
$$

where we used $u_t(\theta) = G_t(Y_t, \Omega_{t-1}, \theta)$.

Therefore,

$$
\begin{aligned}
\int_0^1 \int_0^1 \varphi(x)\varphi(y) E_j(dx, dy) &= E\left[\frac{\partial E[\varphi(u_t(\theta_0))|\Omega_{t-1}]}{\partial \theta} \left(\varphi(u_{t-j}) - c_\varphi\right)\right] \\
&= \left.\frac{\partial E\left[\varphi(u_t(\theta))\ \left(\varphi(u_{t-j}) - c_\varphi\right)\right]}{\partial \theta}\right|_{\theta_0} \\
&= : R_j,
\end{aligned}
$$

where the interchange of the derivative and integral follows from Assupmtion A4. We proved that

$$
\begin{aligned}
\sqrt{n-j}(\widehat{\gamma}_{nj} - \gamma_{nj}) &= \sqrt{n}(\widehat{\theta}_n - \theta_0)' R_j + o_p(1). \\
\sqrt{n-j}(\widehat{\rho}_{nj} - \rho_{nj}) &= \frac{\sqrt{n-j}(\widehat{\gamma}_{nj} - \gamma_{nj})}{var(\varphi(u_t))} + o_p(1) \\
&= \frac{1}{var(\varphi(u_t))} R_j' \sqrt{n}(\widehat{\theta}_n - \theta_0) + o_p(1)
\end{aligned}
$$

Notice that $\sqrt{n}(\rho_{n1}, \rho_{n2} ... \rho_{nm})' \longrightarrow^d N(0, I_m)$, which implies Theorem 1 together with Assumption A2**.**

Next we consider the case of information truncation. Define $\widetilde{u}_t = G_t(Y_t, \Omega_{t-1}, \widehat{\theta}_n)$ and $\widetilde{\gamma}_{nj} = 1/(n-j) \sum_{t=1+j}^{n} (\varphi(\widetilde{u}_t) - c_\varphi)(\varphi(\widetilde{u}_{t-j}) - c_\varphi)$, and then we have

$$
\sqrt{n-j}(\widehat{\gamma}_{nj} - \gamma_{nj}) = \sqrt{n-j}(\widehat{\gamma}_{nj} - \widetilde{\gamma}_{nj}) + \sqrt{n-j}(\widetilde{\gamma}_{nj} - \gamma_{nj}).
$$

The first term on the right hand side is the extra term due to information truncation, which is $o_p(1)$ by Assumption A3. Then notice that the arguments above for $\sqrt{n-j}(\widehat{\gamma}_{nj} - \gamma_{nj})$ without information truncation can be applied directly to $\sqrt{n-j}(\widetilde{\gamma}_{nj} - \gamma_{nj})$, which completes the proof of Theorem 1.

$\blacksquare$

Proof of Theorem 2:

Define

$$m_{BIC} = \min\{m : 1 \leq m \leq p; L_{BIC}(m) \geq L_{BIC}(h), h = 1, 2..., p\},$$

where

$$L_{BIC}(m) = Q_n(m) - m \log n.$$

We need to prove that, under Assumptions A1-A5 and H_0,

$$\lim_{n\to\infty} P(m^* = m_{BIC}) = 1, \tag{12}$$

and that

$$\lim_{n\to\infty} P(m_{BIC} = 1) = 1. \tag{13}$$

We start by proving (12). Define the event

$$A_n(q) = \left\{ \max_{1\leq j\leq p} \sqrt{n} \left|\widehat{\rho}_{nj}\right| > \sqrt{q \log n} \right\}.$$

From Theorem 1 it follows that under H_0

$$\max_{1\leq j\leq p} \sqrt{n} \left|\widehat{\rho}_{nj}\right| = O_P(1).$$

Hence, $P(A_n(q)) = o(1)$, which implies (12).

To prove that (13) also holds, notice that

$$P(m_{BIC} = 1) = 1 - \sum_{j=2}^{p} P(m_{BIC} = j) \geq 1 - \sum_{j=2}^{p} P(L_{BIC}(j) \geq L_{BIC}(1)).$$

Now, for $1 < j \leq p$,

$$P(L_{BIC}(j) \geq L_{BIC}(1)) \leq P(Q_n(j) \geq (j-1)\log(n)) = o(1).$$

Therefore, (13) holds, and Theorem 2 follows from an application of the standard CLT for strictly stationary mds of Billingsley (1961).

$\blacksquare$

REFERENCES

[1] Akaike, H. (1974), "A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control*, 19, 716–723.

[2] Andrews, D.W.K. (1997), "A conditional Kolmogorov test," *Econometrica*, 65, 1097-1128.

[3] Bai, J. (2003), "Testing Parametric Conditional Distributions of Dynamic Models", *Review of Economics and Statistics*, Aug 2003, 85(3): 531-549.

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

[5] Billingsley, P. (1961), "The Lindeberg-Levy Theorem for Martingales," *Proceedings of the American Mathematical Society*, 12, 788-792.

[6] Bollerslev, T. (1986), "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, 31, 307–27.

[7] Bose, A. (1998), "Bahadur Representation of M Estimates", *Annals of Statistics*, 26(2), 771-777.

[8] 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.

[9] Brooks, C., Burke, S.P., Heravi, S. and Persand, G. (2005) "Autoregressive Conditional Kurtosis", *Journal of Financial Econometrics* 3, 399-421.

[10] Chang, Y., Nguyen, C. M. and Park, J. Y. (2009), "Testing Asset Pricing Models in Continuous Time", working paper.

[11] Chaudhuri, S., Goldberg, P. and Jia, P. (2006), "Estimating the Effects of Global Patent Protection in Pharmaceuticals: A Case Study of Quinolones in India", *American Economic Review*, Vol. 96 No. 5, 1477-1513.

[12] Chitturi, R.V. (1974), "Distribution of Residual Autocorrelations in Multiple Autoregressive Schemes," *Journal of American Statistical Association*, 69, 928-934.

[13] Christoffersen, P. F. (2009), "Value-at-Risk Models", *Handbook of Financial Time Series*, 753-766.

[14] Corradi, V. and Swanson, N. (2006), "Bootstrap conditional distribution tests in the presence of dynamic misspecification", *Journal of Econometrics* 133, 779 – 806

[15] Dambis, K. (1965), "On the Decomposition of Continuous Submartingales," *Theory of Probability and its Applications*, 10, 401.

[16] DiNardo, J., Fortin, N. M. and Lemieux, T. (1996), "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach," *Econometrica*, vol. 64(5), pages 1001-44

[17] Delgado, M.A. and Stute, W. (2008), "Distribution-free Specification Tests of Conditional Models", *Journal of Econometrics* v.143, p.37-55.

[18] Du, Z. (2010), "Nonparametric Bootstrap Tests for the Independence of Generalized Errors", working paper.

[19] Dubins, L. and Schwarz G. (1965), "On Continuous Martingales," *Proceedings of the National Academy of Sciences*, 53(5), 913-916.

[20] Durbin, J. (1973), "Weak convergence of the sample distribution function when parameters are estimated", *Annals of Statistics*, 1, 274-290.

[21] Engle, R.F. and Bollerslev, T. (1986), "Modelling the Persistence of Conditional Variances," *Econometric Reviews*, 5, 1-50.

[22] Engle, R. and Gonzalez-Rivera, G. (1991), "Semiparametric ARCH Models," *Journal of Business and Economic Statistics*, 9, 345–60.

[23] Escanciano, J.C. (2006), "Goodness-of-fit Tests for Linear and Nonlinear Time Series Models", *Journal of the American Statistical Association*, vol. 101, no. 474, pp. 531 - 541.

[24] Escanciano, J.C. (2007) "Weak Convergence of Non-Stationary Multivariate Marked Processes with Applications to Martingale Testing", *Journal of Multivariate Analysis*, 98, 1321-1336.

[25] Escanciano, J.C. (2010), "Asymptotic Distribution-Free Diagnostic Tests for Heteroskedastic Time Series Models," *Econometric Theory*, 26, 744-773.

[26] Escanciano, J.C. and Lobato, I.N. (2009), "An Automatic Portmanteau Test for Serial Correlation", *Journal of Econometrics*, 151, 2, 140-149.

[27] Escanciano, J.C., Lobato, I.N and Zhu, L. (2010), "Automatic Diagnostic Checking for Vector Autoregressions", working paper.

[28] Escanciano, J.C. and Olmo, J. (2010), "Backtesting Parametric Value-at-Risk with Estimation Risk", *Journal of Business and Economic Statistics*, 28, 36-51.

[29] Escanciano, J.C. and Velasco, C. (2010), "Specification Tests of Parametric Dynamic Conditional Quantiles", *Journal of Econometrics*, 159, 209-221.

[30] Gao, J. and Gijbels, I. (2008), "Bandwidth Selection in Nonparametric Kernel Testing", *Journal of the American Statistical Association*, 103, 1584-1594.

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

[32] Harvey, C.R. and Siddique, A. (1999) "Autoregressive Conditional Skewness", *Journal of Financial and Quantitative Analysis*, 34, 465–487.

[33] Hayashi, F. (2000), *Econometrics*, Princeton University Press.

[34] Hidalgo, J. and Zaffaroni, P. (2007), "A goodness-of-fit test for ARCH(∞) models" *Journal of Econometrics* 141, 835 – 875

[35] Heckman, J. (1984), "The χ^2 Goodness of Fit Statistic for Models with Parameters Estimated from Microdata", *Econometrica*, Vol. 52, No. 6, pp. 1543-1547

[36] Hong, Y. (2000), "Generalized Spectral Test for Serial Dependece", *Journal of the Royal Statistical Society Series B*, 62, part3, pp. 557-574.

[37] 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, 37-84.

[38] Horvath, L. and Zitikis, R. (2006), "Testing goodness of fit based on densities of GARCH innovations," *Econometric Theory* 22 (2006), pp. 457–482

[39] Hosking, J.R.M. (1980), "The Multivariate Portmanteau Statistic," *Journal of the American Statistical Association*, 75, 602-608.

[40] Hosking, J.R.M. (1981), "Equivalent Forms of Multivariate Portmanteau Tests," *Journal of the Royal Statistical Society, Series B*, 43, 261-262.

[41] Inglot, T. and Ledwina, T. (2006a), "Towards Data Driven Selection of a Penalty Function for Data Driven Neyman Tests," *Linear Algebra and its Applications*, 417, 124-133.

[42] Inglot, T. and Ledwina, T. (2006b), "Data Driven Score Tests of Fit for Semiparametric Homoscedastic Linear Regression Model," Preprint 665, *Institute of Mathematics Polish Academy of Sciences.*

[43] Inglot, T. and Stawiarski, B. (2005), "Data-driven score test of fit for conditional distribution in the GARCH(1,1) model," *Probability and Mathematical Statistics*, 25, pp. 331–362

[44] Jennrich, R.I. (1969), "Asymptotic properties of nonlinear least square estimators." *Annals of Mathematical Statistics*, 40, 633-643.

[45] 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.

[46] Kallenberg, W. C.M. and Ledwina, T. (1999), "Data driven rank tests for independence," *Journal of the American Statistical Association*, 94 (445). pp. 285-301.

[47] Khmaladze, E.V. (1981), "Martingale approach in the theory of Goodness-of-fit tests," *Theory of Probability and its Applications*, 26, 240-257.

[48] Kolmogorov, A. N. (1933), "Sulla Determinazione Empirica Di Una Legge Di Distribuzione," *Giornalle Dell'lnstituto Ital. Degli Attuari*, 4, 83-91.

[49] Koul, H. and Ling, S. (2006), "Fitting an error distribution in some heteroscedastic time series models", *Annals of Statistics* 34, 994-1012.

[50] Leon, A., Rubio, G. and Serna, G. (2005), "Autoregressive Conditional Volatility, Skewness and Kurtosis", *The Quarterly Review of Economics and Finance* 45, Issues 4-5, 599-618

[51] Li, F. and Tkacz, G. (2006) "A Consistent Bootstrap Test for Conditional Density Functions with Time-Series Data," *Journal of Econometrics* 133(2), 863-886.

[52] Ljung, G.M and Box, G.E.P. (1978), "A Measure of Lack of Fit in Time Series Models", *Biometrika* 65, 297-303.

[53] Neumann, M.H. and Paparoditis, E. (2008), "Goodness-of-fit tests for Markovian time series models: Central limit theory and bootstrap approximations," *Bernoulli* 14, No. 1, 14-46.

[54] Park, J.Y. and Vasudev R. (2006), "Testing for martingales in continuous time," Working paper, Dept. of Economics, Rice University.

[55] 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.

[56] Pearson, K,. (1900), "On the Criterion That a Given System of Deviations from the Probable in the Case of a Correlated System of Variables Is Such That It Can Be Reasonably Supposed to Have Arisen from Random Sampling," *The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science*, 50, 157-175.

[57] Premaratne, G. and Bera, A.K. (2001), "Modeling Asymmetry and Excess Kurtosis in Stock Return Data", Working Paper 00-123, Department of Economics, University of Illinois, Champaign.

[58] Rosenblatt, M. (1952), "Remarks on a Multivariate Transformation," *Annals of Mathematical Statistics*, 23, 470-472.

[59] Schwarz, G. (1978), "Estimating the dimension of a model," *Annals of Statistics*, 6, 461-464.

[60] Shiryaev, A.N. (1996), *Probability*, Springer, New York.

[61] Smirnov, N,. (1939), "On the Estimation of the Discrepancy between Empirical Curves of Distribution for Two Independent Samples," *Bulletin Mathématique de l'Université de Moscou*, 2, fasc. 2.

[62] van der Vaart, A. W. and Wellner, J.A.(1996), *Weak Convergence and Empirical Processes*, Springer, New York.

[63] Wu, W.B. (2007), "M-Estimation of Linear Models with Dependent Errors", *Annals of Statistics*, Vol. 35, No. 2, 495–521.

[64] Zheng, J.X. (2000), "A consistent test of conditional parametric distribution," *Econometric Theory*, Vol 16, 667-691.