Specification Tests for Nonlinear Time Series Models

Igor Kheifets*
Department of Economics
Universidad Carlos III de Madrid

February 20, 2008

Abstract

This paper proposes a new parametric model adequacy test for possibly nonlinear time series models such as generalized autoregressive conditional heteroskedasticity (GARCH) and autoregressive conditional duration (ACD). We consider the correct specification of parametric conditional distributions, not only some particular conditional characteristics. Using the true parametric conditional distribution under the null hypothesis we transform data to uniform iid series. The uniformity and serial independence of these transformed series is then examined simultaneously by comparing joint empirical distribution functions with the product of its theoretical uniform marginals by means of Cramer-von Mises or Kolmogorov-Smirnov statistics. We study consistency and asymptotic properties of such tests taking into account parameter estimation effect. Since asymptotic distribution is case dependent, critical values can not be tabulated and we use a bootstrap approximation. The test analyzed in this paper can be extended to in two ways: higher order joint distributions and more lags could be considered. The performance of the test is compared with classical specification checks when estimating ACD models.

Keywords: Goodness of fit, diagnostic test, parametric conditional distribution, serial independence, parameter estimation effect, GARCH models, ACD models, bootstrap.

JEL classification: C12, C22, C52.

*Address: Departamento de Economía, Universidad Carlos III de Madrid, Calle Madrid 126, 28903 Getafe (Madrid), Spain. Email: igor.kheifets @ uc3m.es. I thank Carlos Velasco for his comments.

1 Introduction

Testing the specification of nonlinear time series models is important in macroeconomics and finance to make relevant analysis. Moreover, it is often not enough to check only conditional moments, while specification of the conditional distribution function (df) is necessary. For instance, knowing the true distribution is important to apply efficient maximum likelihood (ML) methods to many models including generalized autoregressive conditional heteroskedasticity (GARCH) models introduced by Engle (1982) and Bollerslev (1986). In autocorrelation conditional duration (ACD) model Engle and Russel (1998) proposed to model conditional expected duration of frequently arriving transaction data on stock exchange. The arrival times are treated as random variables together with volume, price, bid ask spread information. To measure and forecast the intensity of transactions arrivals which is parameterized as a function of time between past events and characteristics of past transactions and to estimate hazard function, the conditional distribution is required. To measure and manage the market risk of the asset portfolios trading organizations such as investment banks or securities firms use Value at Risk (VaR). By definition, VaR is a quantile of the *conditional distribution* of returns on the portfolio given the information set.

Many specification tests are based on an old idea of *integral transform*, dating at least to Segal (1938) and Rosenblatt (1952), that the series can be transformed using the truth conditional distribution to independent identically (iid) uniform distributed on [0,1] random variables (rvs). The challenge with this approach is that we do not know the truth parameters, while the integral transform using estimated parameters does not necessary provide iid uniform rvs. Asymptotic properties and critical values of the test with estimated parameters based on integral transform become invalid and *parameter estimation effect* has to be addressed.

Estimation effect problem is not specific to integral transform, it was considered by Durbin (1973) in the context of goodness-of-fit Kolmogorov test. Newey (1995) and Tauchen (1985) use ML method to deal with the problem of estimation uncertainty. The extension of Kolmogorov test to goodness-of-fit for parametric conditional distribution function, is done by Andrews (1997) in an iid (static) context. The estimation effect does change the asymptotic distribution of the statistics and makes it *data dependent*. Andrews (1997) proves that parametric bootstrap provides correct critical values in this case using linear expansion of the estimation effect, which arise naturally under the ML method. The idea of orthogonal projecting the test statistics against the estimation effect due to Wooldridge (1990), were used in parametric moment tests Bontemps and Meddahi (2006). The continuous version of the projection, often called Khmaladze (1981) transformation, was employed in test of Koul and Stute (1999) to specify the conditional mean, Bai and Ng (2001), Bai (2003), Delgado and Stute (2008) to specify the conditional distribution. These projection tests are not model invariant since they require to compute conditional mean and variance derivatives, and also projections may cause the loss in power. An alternative approach, proposed in Chen (2006), is to use information about estimation error from lower moment conditions in the test based on other (higher) moment conditions. His test is model invariant, robust to square root consistent estimators of conditional mean and variance parameters.

There are different tests with the integral transform that are able to handle appropriately the estimation effect. Hong (1999, 2000) and Hong and Li (2005) rely on nonparametric kernel estimators and hence have slow convergence rate. Thomson (2008) use Generalized Method of Moments (GMM), his tests are not consistent and may not detect some alternatives but achieve better power in large samples against a number relevant alternatives. Bai (2003) and Corradi and Swanson (2001) use empirical function approach which checks distributional assumption but do not detect directly departures from independence. Gonzalez-Rivera, Senyuz and Yoldas (2007) use a novel approach of "autocontours" but without integral transform nor empirical function.

Independence tests can be based on the difference between joint distribution (density) and the product of the marginals. Nonparametric tests on distributions are defined through empirical distribution functions are proposed in Hoeffding (1948) and Blum et al. (1961) while Rosenblatt (1975) used smooth kernel estimates of the densities. The extension of Blum et al. (1961) to a time series framework is provided in Skaug and Tjøstheim (1993a). The test is consistent against lag one dependence. The test has the same limiting distribution as Blum et al. (1961) if we have continuous marginal distributions. It can be applied to time series with infinite variance, which is often the case in high frequency financial data. This test is generalized to the higher dimensions in Delgado (1996). He tests that the joint distribution of p random variables is equal to the product of marginals and derives asymptotic distribution under the null and fixed alternatives, which appeared to be different from Blum et al. (1961). Ghoudi, Kulperger and Remillard (2001) point out that to approximate critical values by permutation method used by Delgado (1996) is a very sensitive procedure since any pseudo-random generator has some kind of dependence. Hong (1998) generalizes the test of Skaug and Tjøstheim (1993a) to larger but finite number of lags, which increases with sample size and recent lags receive larger weights. He obtains standard normal limiting distribution, developing a dependent degenerate V-statistics projection theory. Simulations show that the test has good power against both short and long memory linear processes, while low power appears for ARCH mode. Delgado and Mora (2000) consider testing independence of *unobservable errors* in the linear regression model. Surprisingly, in the test of independence via Blum et al. (1961) statistics the additional random term which appears from joint distribution with estimated parameters cancels out with random term which appears from product of marginal distributions with estimated parameters. It would be interesting to extend this technique to nonlinear regression models.

The goal of this paper is to propose a new specification test for dynamic models, using integral transform but taking into account parameter estimation effect, and using empirical function approach but capture both uniformity and iid. The drawback is that once model is rejected we do not know the source of the misspecification as in any omnibus test. We introduce an empirical process which incorporates the difference between the empirical joint distribution and the product of the uniform marginals and then apply Cramer-von Mises or Kolmogorov-Smirnov measures. Under standard conditions we prove consistency and show asymptotic properties of such tests taking into account parameter estimation effect. Since asymptotic distribution is case dependent, critical values can not be tabulated and we prove that the bootstrap distribution approximation is valid. Our framework allows to consider higher order joint distributions as well as to incorporate many lags. In this view, Bai (2003) empirical process can be considered as a zero-lag case. We also derive algebraic formula to compute our statistics to avoid numeric integration.

The rest of the paper is organized as follows. Section 2 introduces specification test statistics. Asymptotic properties are shown in Sections 3 and bootstrap justification provided in Section 4. Examples of the GARCH and ACD models are considered in Section 5. Section 6 concludes.

2 New test statistics

In this secction we introduce our statistics. Suppose that a sequence of observations $Y_1, Y_2, \ldots, Y_n$ is given. Let $\Omega_t = (Y_{t-1}, Y_{t-2}, \ldots)$ be the information set at time t (not including Y_t). We consider family of conditional df's $F_t(y|\Omega_t, \theta)$, parameterized by $\theta \in \Theta$, where $\Theta \subseteq R^L$. We allow nonstationarity. Moreover, apart that condition (information) set Ω_t changes with the time, we permit change in functional form of df using subscript t in F_t. Our null hypothesis of correct specification is

H_0 : The conditional distribution of Y_t conditional on Ω_t is in the parametric family $F_t(y|\Omega_t, \theta)$ for some $\theta_0 \in \Theta$, where $\Theta \subseteq R^L$ is the parameter space.

We use the fact that under the null $U_t = F_t(Y_t|\Omega_t, \theta_0)$ are uniform on [0,1] and i.i.d. random variables (rv's), so that $\Pr(U_t \le r_1, U_{t-1} \le r_2) = r_1 r_2$ or $\mathrm{E}\, I(U_t \le r_1) I(U_{t-1} \le r_2) = r_1 r_2$. This motivates us to consider the following empirical process, for $r = (r_1, r_2) \in [0,1]^2$

$$V_{2n}(r) = \frac{1}{\sqrt{n-1}} \sum_{t=2}^{n} \left[I(U_t \le r_1) I(U_{t-1} \le r_2) - r_1 r_2 \right]. \tag{1}$$

If we do not know θ_0 either $\{Y_t, t \le 0\}$, we approximate U_t with $\hat{U}_t = F_t(Y_t|\tilde{\Omega}_t, \hat{\theta})$ where $\hat{\theta}$ is an estimator of θ_0 and truncated information is $\tilde{\Omega}_t = (Y_{t-1}, Y_{t-2}, \ldots, Y_1)$ and write

$$\hat{V}_{2n}(r) = \frac{1}{\sqrt{n-1}} \sum_{t=2}^{n} \left[I(\hat{U}_t \le r_1) I(\hat{U}_{t-1} \le r_2) - r_1 r_2 \right]. \tag{2}$$

The process $\hat{V}_{2n}(r)$ measure the distance to the null hypothesis for each r, so we need to choose the metrics in $[0,1]^2$ to aggregate for all r. For any continuous functional $\Gamma(\cdot)$ from $\ell^\infty([0,1]^2)$, the set of uniformly bounded real functions on $[0,1]^2$, to R,

$$D_{2n} = \Gamma(\hat{V}_{2n}(r)).$$

For example if we take CvM or KS metrics in $[0,1]^2$, we get the following statistics

$$D_{2n}^{CvM} = \int_{[0,1]^2} \hat{V}_{2n}(r)^2 dr \text{ or } D_{2n}^{KS} = \max_{[0,1]^2} \left| \hat{V}_{2n}(r) \right|.$$

Recall that (2) checks only pairwise dependence. To check p-wise independence in a similar way to Delgado (1996) we can write

$$\hat{V}_{pn}(r) = \frac{1}{\sqrt{n-(p+1)}} \sum_{t=p+1}^{n} \left[\prod_{j=1}^{p} I(\hat{U}_{t-j} \le r_j) - r_1 r_2 \ldots r_p \right] \tag{3}$$

and use the test statistics

$$D_{pn}^{CvM} = \int_{[0,1]^p} \hat{V}_{pn}(r)^2 dr \text{ or } D_{pn}^{KS} = \max_{[0,1]^p} \left| \hat{V}_{pn}(r) \right|. \tag{4}$$

Then one may also wish to aggregate the statistics for different p. One further possibility is to test j-lag pairwise independence. Define process

$$\hat{V}_{2n,j}(r) = \frac{1}{\sqrt{n-j}} \sum_{t=j+1}^{n} \left[I(\hat{U}_t \le r_1) I(\hat{U}_{t\text{-}j} \le r_2) - r_1 r_2 \right]. \tag{5}$$

with test statistics, for example

$$D_{2n,j}^{CvM} = \int_{[0,1]^2} \hat{V}_{2n,j}(r)^2 dr \text{ or } D_{2n,j}^{KS} = \max_{[0,1]^2} \left| \hat{V}_{2n,j}(r) \right|. \tag{6}$$

We can aggregate across p or j summing possibly with different weights $k(\cdot)$, we get generalized statistics

$$ADP_n = \sum_{p=1}^{n-1} k(p) D_{pn}, \text{ or } ADJ_n = \sum_{j=1}^{n-1} k(j) D_{2n,j}. \tag{7}$$

For $p = 1$ the process $\hat{V}_{1n}$ and the statistic D_{1n}^{KS} were studied by Bai (2003). He focuses on conditional distributions of dynamic models (the same null hypothesis) for a sequence of observations (Y_1, X_1), $(Y_2, X_2), ..., (Y_n, X_n)$ with the information set $\Omega_t = (X_t, X_{t-1}, \ldots; Y_{t-1}, Y_{t-2}, \ldots)$ and truncated version is $\hat{\Omega}_t = (X_t, X_{t-1}, \ldots, X_1, 0, 0, \ldots; Y_{t-1}, Y_{t-2}, \ldots, Y_1, 0, 0, \ldots)$.

To check nonparametrically serial independence of a sequence $\{Z_t, t = 1, ..., n\}$ Skaug and Tjøstheim (1993a) used the following process for $z = (z_1, z_2) \in \mathbb{R}^2$

$$S_{2n}(z) = \sqrt{n} \left[\frac{1}{n} \sum_{t=2}^{n} I(Z_t \le z_1) I(Z_{t-1} \le z_2) - \frac{1}{n} \sum_{t=1}^{n} I(Z_t \le z_1) \frac{1}{n} \sum_{t=2}^{n} I(Z_{t-1} \le z_2) \right], \tag{8}$$

which was proposed by Hoeffding (1948) and Blum et al. (1961) in a different non time series setup. We will see that our process V_{2n} has a different asymptotic distribution than S_{2n}.

Delgado and Mora (2000) showed that the distribution of S_{2n} does not change if estimated errors (residuals) from linear regression are used. It would be interesting to study whether their result hold for nonlinear models. In our case, asymptotic distribution of $\hat{V}_{2n}$ is different from V_{2n}, and moreover, is not distribution free. Therefore we use bootstrap approximation to estimate critical values.

3 Asymptotic properties

In this section we derive asymptotic properties of our statistics. We start with the simple case when we know parameters, then study how changes the asymptotic distribution if we estimate parameters. We provide analysis under the null, under the local and fixed alternatives. We will need standard assumptions on conditional dfs and the form of parametric family of dfs and on the estimator.

Assumption 1 *The conditional df $F_t(y|\Omega_t, \theta)$ are continuously differentiable with respect to θ, and continuous and strictly increasing in y.*

The following proposition provides result about integral transform.

Proposition 1 *Suppose Assumption 1 holds. Then under H_0 random variables $U_t = F_t(Y_t|\Omega_t, \theta_0)$ are iid uniform.*

We first describe the asymptotic behavior of the process $V_{2n}(r)$ under H_0.

Proposition 2 *Suppose Assumption 1 holds. Then under H_0*

$$V_{2n}(r) \xrightarrow{d} V_{2\infty}(r),$$

where $V_{2\infty}(r)$ is asymptotically zero mean Gaussian process with Covariance

$$\mathrm{Cov}_{V_{2\infty}}(r, s) = (r_1 \wedge s_1)(r_2 \wedge s_2) + (r_1 \wedge s_2)r_2 s_1 + (r_2 \wedge s_1)r_1 s_2 - 3r_1 r_2 s_1 s_2. \quad (9)$$

The Covariance differs from

$$\lim_{n\to\infty} \mathrm{Cov}_{S_n}(r, s) = ((r_1 \wedge s_1) - r_1 s_1)((r_2 \wedge s_2) - r_2 s_2),$$

therefore our new test has different asymptotic distribution from Skaug and Tjøstheim(1993a) and Delgado and Mora (2000).

In practice estimating the parameter will affect asymptotic distribution. To take this into account, we use Taylor expansion to approximate $\hat{V}_{2n}(r)$ with $V_{2n}(r)$. Let $\|\cdot\|$ denote Euclidian norm for matrices, i.e. $\|A\| = \sqrt{\mathrm{tr}(A'A)}$ and $B(a, \rho)$ is an open ball in R^L with the center in the point a and the radius ρ. Define

$$\begin{aligned} h_t(r, u, v) \;=\; & \dot{F}_{t\theta}(F_t^{-1}(r_1|\Omega_t, u)|\Omega_t, v) F_{t-1}(F_{t-1}^{-1}(r_2|\Omega_{t-1}, u)|\Omega_{t-1}, v) \\ & + \dot{F}_{t-1\theta}(F_{t-1}^{-1}(r_2|\Omega_{t-1}, u)|\Omega_{t-1}, v) F_t(F_t^{-1}(r_1|\Omega_t, u)|\Omega_t, v). \end{aligned}$$

Assumption 2 *There exists a uniformly continuous (vector) function $h(r)$ from $[0, 1]^2$ to R^L such that for every $M > 0$*

$$\sup_{u,v \in B(\theta_0, Mn^{-1/2})} \left\| \frac{1}{n} \sum_{t=2}^{n} h_t(r, u, v) - h(r) \right\| = o_p(1)$$

uniformly in r.

Note: uniform continuity on $[0,1]^2$ implies uniform boundness.

To identify the limit of $\hat{V}_{2n}(r)$ we need to study limiting distribution of $\sqrt{n}(\hat{\theta}-\theta_0)$. We assume the following

Assumption 3 *When the sample is generated by the null df $F_t(y|\Omega_t,\theta_0)$, the estimator $\hat{\theta}$ admits a linear expansion*

$$\sqrt{n}(\hat{\theta}-\theta_0)=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\psi(\Omega_t)l(U_t)+o_p(1), \tag{10}$$

with $E_{F_t}\left(l(U_t)|\Omega_t\right)=0$ and

$$\frac{1}{n}\sum_{t=1}^{n}\psi(\Omega_t)\psi(\Omega_t)'l^2(U_t)\overset{p_{F_t}}{\to}\Psi.$$

This assumption is satisfied for ML estimator. We will show it in the following section. Define

$$C_n(r,s,\theta)=E\begin{pmatrix} V_{2n}(r) \\ \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\psi(\Omega_t)l(U_t)\end{pmatrix}\begin{pmatrix} V_{2n}(s) \\ \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\psi(\Omega_t)l(U_t)\end{pmatrix}'$$

and let $(V_{2\infty}(r),\psi_\infty')'$ be a zero mean Gaussian process with covariance function $C(r,s,\theta_0)=\lim_{n\to\infty}C_n(r,s,\theta_0)$. Dependence on θ on right hand side (rhs) comes through U_t, since they are obtained with integral transform.

Suppose the conditional distribution function $H_t(y|\Omega_t)$ is not in the parametric family $F_t(y|\Omega_t,\theta)$, i.e. for each $\theta\in\Theta$ there exists $y\in\mathbb{R}$ and one Ω_{t_0} that occurs with positive probability and $H_{t_0}(y|\Omega_{t_0})\neq F_{t_0}(y|\Omega_{t_0},\theta)$. For any $n_0\in\{0,1,2,...,\}$ and $n\geq n_0$ define conditional on Ω_t conditional df

$$G_{nt}(y|\Omega_t,\theta)=\left(1-\frac{\sqrt{n_0}}{\sqrt{n}}\right)F_t(y|\Omega_t,\theta)+\frac{\sqrt{n_0}}{\sqrt{n}}H_t(y|\Omega_t).$$

Now we define fixed and local alternatives.

H_1: Conditional df of Y_t is not in the parametric family $F_t(y|\Omega_t,\theta)$.

H_{1n}: Conditional df of Y_t is equal to $G_{nt}(y|\Omega_t,\theta_0)$ with $n_0\neq 0$.

Rvs Y_t with conditional df $G_{nt}(y|\Omega_t,\theta_0)$ allows us to study all three cases: H_0 if $n_0=0$, H_{1n} if $n=n_0,n_0+1,n_0+2,...$ and $n_0\neq 0$ and H_1 if we fix $n=n_0$. To derive asymptotical distribution under alternatives we need assumption on $H_t(y|\Omega_t)$ similar to Assumptions 1.

Assumption 4 *The conditional df $H_t(y|\Omega_t)$ are continuous and strictly increasing in y.*

Under Assumptions 1 and 4 we also have that the conditional df $G_{nt}(y|\Omega_t, \theta)$ are continuously differentiable with respect to θ, and continuous and strictly increasing in y.

Under alternative we will require that estimator converges in probability.

Assumption 5 $\hat{\theta} \xrightarrow{p} \theta_1$ *for some* $\theta_1 \in \Theta$.

In the next proposition we provide the result on asymptotic distribution of our statistics under the null, under the local and fixed alternatives.

Proposition 3 *a) Suppose Assumptions 1, 2, 3 hold. Then under H_0*

$$\Gamma(\hat{V}_{2n}) \xrightarrow{d} \Gamma(\hat{V}_{2\infty}(r)),$$

where

$$\hat{V}_{2\infty}(r) = V_{2\infty}(r) - h(r)'\psi_\infty.$$

b) Suppose Assumptions 1, 2, 3, 4 hold. Then under H_{1n}

$$\Gamma(\hat{V})_{2n} \xrightarrow{d} \Gamma(\hat{V}_{2\infty}(r) + \sqrt{n_0}k(r) - \sqrt{n_0}h(r)'m),$$

where

$$\begin{aligned} k(r) \quad &= \quad \text{plim}\, \frac{1}{n} \sum_{t=2}^{n} \left\{ \left[H_t(F_t^{-1}(r_1|\Omega_t, \theta_0)|\Omega_t) - r_1 \right] r_2 \right. \\ &\qquad \left. + \left[H_{t-1}(F_{t-1}^{-1}(r_2|\Omega_{t-1}, \theta_0)|\Omega_{t-1}) - r_2 \right] r_1 \right\}. \end{aligned}$$

and

$$m = \text{plim}\, \frac{1}{n} \sum_{t=1}^{n} \psi(\Omega_t) l(U_t). \tag{11}$$

c) Suppose Assumptions 1, 2, 5 hold. Then for all sequences of random variables $\{c_n : n \geq 1\}$ with $c_n = O_p(1)$, under H_1, we have

$$\lim_{n\to\infty} P(\Gamma(\hat{V}_{2n}) > c_n) = 1.$$

Under G_{nt}, rv's $U_t = F_t(Y_t|\Omega_t, \theta_0)$ are not any more iid, instead $U_t^* = G_{nt}(Y_t|\Omega_t, \theta_0)$ are uniform iid. Due to this fact we have the term $k(r)$ in asymptotic distribution. Under alternative we may have also (10) not centered, since $E_{G_{nt}}\left(l(U_t)|\Omega_t\right) = \frac{\sqrt{n_0}}{\sqrt{n}} E_{H_t}\left(l(U_t)|\Omega_t\right)$, therefore m may be nonzero, which stands for information from estimation. This term does not appear in Bai (2003) method, since he projects out the estimation. Due to Assumption 3, we may also take plim in (11) with respect to probability measure generated by $H_t(y|\Omega_t)$.

4 Parametric bootstrap approximation

There are different bootstrap and sampling techniques to approximate asymptotic distribution, see for example Shao and Dongsheng (1995), Politis, Romano and Wolf (1999).. Since under H_0 we know parametric conditional distribution, we apply parametric bootstrap to mimic the H_0 distribution. We introduce the algorithm now.

1. Estimate model with initial data Y_t, $t = 1, 2, ..., n$, get parameter estimator $\hat{\theta}$, get test statistic $\Gamma(\hat{V})$.
2. Simulate Y_t^* with $F_t(\cdot|\Omega_t^*, \hat{\theta})$ recursively for $t = 1, 2, ..., n$, where $\Omega_t^* = (Y_{t-1}^*, Y_{t-2}^*, ...)$.
3. Estimate model with simulated data Y_t^*, get θ^*, get bootstrapped statistics $\Gamma(\hat{V}^*)$.
4. Repeat 2-3 B times, compute the percentiles of the empirical distribution of the B boostrapped statistics.
5. Reject H_0 if $\Gamma(\hat{V})$ is greater than the $(1-\alpha)$th percentile.

We will prove that $\Gamma(\hat{V}_{2n}^*)$ has the same limiting distribution as $\Gamma(\hat{V}_{2n})$. We say that the sample is distributed under $\{\theta_n : n \geq 1\}$ when there is a triangular array of rv's $\{Y_{nt} : n \geq 1, t \leq n\}$ with (n,t) element generated by $F_t(\cdot|\Omega_{nt}, \theta_n)$, where $\Omega_{nt} = (Y_{nt-1}, Y_{nt-2}, ...)$.

Assumption 6 *For all nonrandom sequences $\{\theta_n : n \geq 1\}$ for which $\theta_n \to \theta_0$, we have*

$$\sqrt{n}(\hat{\theta} - \theta_n) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \psi(\Omega_{nt}) l(U_{nt}) + o_p(1),$$

under $\{\theta_n : n \geq 1\}$, where $E\left(l(U_{nt}, \theta_0)|\Omega_{nt}\right) = 0$ and

$$\frac{1}{n} \sum_{t=1}^{n} \psi(\Omega_{nt}) \psi(\Omega_{nt})' l^2(U_{nt}) \xrightarrow{p} \Psi.$$

Note, functions $\psi(\cdot)$ and $l(\cdot)$ are the same as in Assumption 3. We ask that estimators of close to θ_0 points have *the same* linear representation as the estimator of θ_0 itself. Assumption 3 itself will guarantee only expansion with $\psi^{\theta_n}(\cdot)$ and $l^{\theta_n}(\cdot)$ depending on θ_n.

Proposition 4 *Suppose Assumptions 1, 2, 6 hold. Then for any nonrandom sequence* $\{\theta_n : n \geq 1\}$ *for which* $\theta_n \to \theta_0$*, under* $\{\theta_n : n \geq 1\}$

$$\Gamma(\hat{V}_{2n}(r)) \xrightarrow{d} \Gamma(\hat{V}_{2\infty}(r)). \tag{12}$$

Denote $c_{\alpha n}(\theta_n)$ the level α critical value obtained from $\Gamma(\hat{V}_{2n})$ distribution under $\{\theta_n : n \geq 1\}$ and $c_\alpha(\theta_0)$ the level α critical value obtained by the distribution of $\Gamma(\hat{V}_{2\infty})$, i.e.

$$P_{\theta_n}(\Gamma(\hat{V}_{2n}) > c_{\alpha n}(\theta_n)) = \alpha \tag{13}$$

and

$$P(\Gamma(\hat{V}_{2\infty}) > c_\alpha(\theta_0)) = \alpha$$

and $c_{\alpha nB}(\theta_n)$ bootstrap approximate for $c_{\alpha n}(\theta_n)$.

Proposition 5 *(a) Suppose Assumptions 1, 2, 6 hold and* $B \to \infty$ *as* $n \to \infty$*. Then*

$$\begin{aligned} &c_{\alpha nB}(\hat{\theta}) \xrightarrow{p} c_\alpha(\theta_0) \\ &P_{\theta_0}(\Gamma(\hat{V}_{2n}) > c_{\alpha nB}(\hat{\theta})) \to \alpha. \end{aligned}$$

(b) Suppose Assumptions 1 and Assumptions 2, 6 hold for any $\theta_0 \in \Theta$*, and* $B \to \infty$ *as* $n \to \infty$*. Then*

$$\sup_{\theta_0 \in \Theta} \lim_{n \to \infty} P_{\theta_0}(\Gamma(\hat{V}_{2n}) > c_{\alpha nB}(\hat{\theta})) = \alpha.$$

(c) Suppose Assumptions 1, 5 and Assumptions 2, 6 hold for any $\theta_0 \in \Theta$*, and* $B \to \infty$ *as* $n \to \infty$*. Then*

$$c_{\alpha nB}(\hat{\theta}) \xrightarrow{p} c_\alpha(\theta_1).$$

Bootstrapped critical values are $O_p(1)$ both under the null and under the alternative, therefore the test based on bootstrapped critical values is consistent according to Proposition 3.

5 Applications

The autoregressive conditional heteroskedastic (ARCH) model was first introduced by Engle (1982) and has been very popular nonlinear financial time series model. Here we define *generalized* ARCH (GARCH) proposed by Bollerslev (1986). GARCH(p,q) process has the conditional variance function

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{q} \alpha_j (Y_{t-j} - C)^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2,$$

where $p > 0$ and $\beta_j \geq 0, 1 \leq j \leq p$ and

$$Y_t = C + \varepsilon_t \sigma_t,$$

where

$$\varepsilon_t \sim \text{ i.i.d. with cdf } F_\varepsilon(\varepsilon) \text{ or } Y_t|\Omega_t \sim \text{ i.i.d. with cdf } F_{Y_t|\Omega_t}(y) = F_\varepsilon(y/\sigma_t).$$

Therefore the transform is just $U_t = F_t(Y_t|\Omega_t) = F_{Y_t|\Omega_t}(Y_t) = F_\varepsilon(Y_t/\sigma_t) = F_\varepsilon(\varepsilon_t)$. Here we have in mind $F_\varepsilon(\varepsilon)$ being normal or t-distribution. Assumption **??** is obviously satisfied.

If we have data $Y_1, Y_2, ..., Y_n$, quasi-maximum likelihood estimators (QMLE) we can obtain by maximizing log-likelihood function

$$\begin{aligned} L_n(Y_1, Y_2, ..., Y_n; \theta) &= L_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} l_t(\theta) \\ &= -\frac{1}{2n} \sum_{t=1}^{n} \ln \sigma_t^2 - \frac{1}{2n} \sum_{t=1}^{n} \frac{(Y_t - C)^2}{\sigma_t^2}. \end{aligned}$$

We consider QMLE for GARCH(1,1) as studied in Lee and Hansen (1994).

Assumption 7 *(o) The parameter space $\Theta \subseteq R^4$ is a compact and convex. For any vector $(C, \alpha_0, \alpha, \beta) \in \Theta$, $-\frac{\tilde{C}}{2} \leq C \leq \frac{\tilde{C}}{2}$, $\delta \leq \alpha_0 \leq \tilde{\alpha}_0$, $\delta \leq \alpha \leq 1 - \delta$, and $\delta \leq \beta \leq 1 - \delta$, for some constant δ, where $\tilde{C}$, $\tilde{\alpha}_0$, and δ are given a priori.*

(i) ε_t is strictly stationary and ergodic.

(ii) ε_t^2 is non degenerate

(iii) For some $\delta > 0$, there exists an $S_b < \infty$ such that $E\left(\varepsilon_t^{2+\delta}|\Omega_t\right) \leq S_b < \infty$ a.s.

(iv) $\sup_t E\left[\ln\left(\alpha\varepsilon_t^2 + \beta\right)|\Omega_t\right] < 0$ *a.s.*

Assumption 8 *(i) $E\left(\varepsilon_t^4|\Omega_t\right) \leq S_b < \infty$ a.s.*

(ii) The true parameter vector θ_0 is in the interior of Θ.

The first assumption is enough to show consistency of QMLE and coupled with the second provides asymptotic normality and in particular sufficient for Assumption 3. It can be shown that $plim\ L_{n\theta\theta}^{-1}(\theta_0)$ exists. Using Taylor expansion of $L_n(\theta)$ around θ_0 we have

$$\begin{aligned}\sqrt{n}(\theta-\theta_0) &= \sqrt{n}L_{n\theta\theta}^{-1}(\theta_0)L_{n\theta}(\theta)+o_p(1)=L_{n\theta\theta}^{-1}(\theta_0)\frac{1}{\sqrt{n}}\sum_{t=1}^{n}l_{t\theta}(\theta)+o_p(1)\\ &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\frac{(Y_t-C)^2}{\sigma_t^2}-1\right)\frac{(\sigma_t^2)_\theta}{\sigma_t^2}+o_p(1).\end{aligned}$$

Now choose $\psi(\Omega_t)=\frac{(\sigma_t^2)_\theta}{\sigma_t^2}$ and $l(U_t)=\frac{(F_t^{-1}(U_t|\Omega_t)-C)^2}{\sigma_t^2}-1$ to get expansion of the estimator as in Assumption 3.

Following Engle and Russell (1998), the class of autoregressive conditional duration (ACD) models is defined as follows. Denote by $Y_i=x_i-x_{i-1}$ the duration between two consecutive events, where x_i is the time at which the i-th event (trade, quote, price change etc.) occurs. Let F_{i-1} be the information set consisting of all information up to and including time x_{i-1}. Let ψ_i be the expectation of the i-th duration

$$E(Y_i|Y_{i-1},\ldots,Y_1)=\psi_i(Y_{i-1},\ldots,Y_1;\theta)\equiv\psi_i.$$

We assume

$$Y_i=\psi_i\varepsilon_i,$$

where

$$\{\varepsilon_i\}\sim\ \text{i.i.d. with density } f(\varepsilon;\phi).$$

The specification of ACD model requires to know the expected duration and the distribution of ε_i. One such specification is the exponential distribution.

One of possible specifications of conditional intensity is

$$\psi_i=\omega+\sum_{j=0}^{m}\alpha_jY_{i-j}+\sum_{j=0}^{q}\beta_j\psi_{i-j},$$

which is called ACD(m,q). This can be formulated in terms of ARMA(m,q):

$$Y_i=\omega+\sum_{j=0}^{\max(m,q)}(\alpha_j+\beta_j)Y_{i-j}-\sum_{j=0}^{q}\beta_j\eta_{i-j}+\eta_i,$$

where $\eta_i\equiv Y_i-\psi_i$ is Martingale Difference Sequence (MDS) and highly non-Gaussian.

Also there is a close connection with GARCH models. For example some results for GARCH can be rewritten in terms of ACD. For instance see corollary on page 1135 of Engle and Russell (1998) about QMLE properties.

The ACD model falls in a large field of *marked point process* and *durational analysis* (see Lancaster (1979) for durational analysis in econometrics), but since our main target is the specification testing we will not go there.

Engle and Russel (1998) apply their model to financial markets, to check if the frequency of transactions carry information about state of the market and if there is clustering of transactions. To check the adequacy of the model they use Portmanteau tests for independence. We are going to repeat these checks via our new test statistics and compare the performance for IBM transaction data.

6 Conclusion

We have proposed a new test for checking goodness-of-fit of conditional distribution in nonlinear time series models. Specification of the conditional distribution (but not only conditional moments) is important in many macroeconomics and financial applications, in particular in VAR analysis. Under standard conditions, although allowing nonstanarity, we obtain asymptotic properties of our statistics. Due due to parameter estimation effect, the asymptotic distribution depends on data. We show that parametric bootstrap provides right distribution. The test can be applied to wide range of financial models including popular GARCH and ACD models.

APPENDIX

7 Proofs

Proof of Proposition 1.

First we show that U_t are uniform:

$$\begin{aligned} P(U_t \le r|\Omega_t) &= P(F_t(Y_t|\Omega_t) \le r|\Omega_t) = P(Y_t \le F_t^{-1}(r|\Omega_t)|\Omega_t) \\ &= F_t(F_t^{-1}(r|\Omega_t)|\Omega_t) = r. \end{aligned}$$

Unconditionally,

$$P(U_t \le r) = E\left[E\left(I(U_t \le r)|\Omega_t\right)\right] = EP(U_t \le r|\Omega_t) = r.$$

Now we check pairwise lag-1 conditional independence, general independence can be shown in a similar manner. The joint distribution of U_t and U_{t-1} condition on Ω_{t-1} will be

$$\begin{aligned} P(U_t \le r_1, U_{t-1} \le r_2|\Omega_{t-1}) &= E(I(U_t \le r_1, U_{t-1} \le r_2)|\Omega_{t-1}) \\ &= E\left[I(U_t \le r_1)I(U_{t-1} \le r_2)|\Omega_{t-1}\right] \\ &= E\left[E(I(U_t \le r_1)I(U_{t-1} \le r_2)|\Omega_t)|\Omega_{t-1}\right] \\ &= E\left[E(I(U_t \le r_1)|\Omega_t)I(U_{t-1} \le r_2)|\Omega_{t-1}\right] \\ &= E\left[P(U_t \le r_1|\Omega_t)I(U_{t-1} \le r_2)|\Omega_{t-1}\right] \\ &= E\left[r_1 I(U_{t-1} \le r_2)|\Omega_{t-1}\right] \\ &= r_1 E\left[I(U_{t-1} \le r_2)|\Omega_{t-1}\right] \\ &= r_1 r_2 = P(U_t \le r_1)P(U_{t-1} \le r_2) \end{aligned}$$

Proof of Proposition 2.

We use functional CLT of Pollard (1984, Theorem 10.12). We need to check equicontinuity and convergence of fidis. Equicontinuity follows from Theorems 1-3 of Andrews(1994). Indeed, each component of V_{2n} is a function of type II (see p. 2270 of Andrews (1994)), therefore for V_{2n} by Theorems 2-3 Assumption A holds with envelope 1, so Assumption B also holds and we can apply Theorem 1 to get equicontinuity.

We check that the process has zero mean and covariance converging to (9).

The fidis converge by CLT for 1-dependent data (see for example ..).

$$
\begin{aligned}
EV_{2n}(r) &= \\
&= \frac{1}{\sqrt{n-1}} \sum_{t=2}^{n} \left[P(U_t \le r_1, U_{t-1} \le r_2) - r_1 r_2\right] \\
&= \frac{1}{\sqrt{n-1}} \sum_{t=2}^{n} \left[P(U_t \le r_1) P(U_{t-1} \le r_2) - r_1 r_2\right] \\
&= \frac{1}{\sqrt{n-1}} \sum_{t=2}^{n} \left[r_1 r_2 - r_1 r_2\right] = 0.
\end{aligned}
$$

In the third equality we used independence of U_t and in the fourth that they are uniformly distributed under then null.

Now we derive the variance. Again, using that U_t are iid uniform we have

$$
\begin{aligned}
& E\left[I(U_t \le r_1, U_{t-1} \le r_2) - r_1 r_2\right]\left[I(U_{t'} \le s_1, U_{t'-1} \le s_2) - s_1 s_2\right] \\
= \; & EI(U_t \le r_1, U_{t-1} \le r_2, U_{t'} \le s_1, U_{t'-1} \le s_2) + r_1 r_2 s_1 s_2 \\
& -E\left[I(U_t \le r_1, U_{t-1} \le r_2) s_1 s_2\right] - E\left[I(U_{t'} \le s_1, U_{t'-1} \le s_2) r_1 r_2\right] \\
= \; & P(U_t \le r_1, U_{t-1} \le r_2, U_{t'} \le s_1, U_{t'-1} \le s_2) + r_1 r_2 s_1 s_2 \\
& -P(U_t \le r_1, U_{t-1} \le r_2) s_1 s_2 - P(U_{t'} \le s_1, U_{t'-1} \le s_2) r_1 r_2 \\
= \; & P(U_t \le r_1, U_{t-1} \le r_2, U_{t'} \le s_1, U_{t'-1} \le s_2) + r_1 r_2 s_1 s_2 \\
& -P(U_t \le r_1) P(U_{t-1} \le r_2) s_1 s_2 - P(U_{t'} \le s_1) P(U_{t'-1} \le s_2) r_1 r_2 \\
= \; & P(U_t \le r_1, U_{t-1} \le r_2, U_{t'} \le s_1, U_{t'-1} \le s_2) + r_1 r_2 s_1 s_2 \\
& -r_1 r_2 s_1 s_2 - s_1 s_2 r_1 r_2 \\
= \; & P(U_t \le r_1, U_{t-1} \le r_2, U_{t'} \le s_1, U_{t'-1} \le s_2) - r_1 r_2 s_1 s_2 \\
= \; & \begin{cases}
P(U_t \le r_1) P(U_{t-1} \le r_2) P(U_{t'} \le s_1) P(U_{t'-1} \le s_2), & \text{if } |t - t'| > 1 \\
P(U_t \le r_1 \wedge s_1) P(U_{t-1} \le r_2 \wedge s_1), & \text{if } t' = t \\
P(U_t \le r_1 \wedge s_2) P(U_{t-1} \le r_2) P(U_{t'} \le s_1), & \text{if } t' = t + 1 \\
P(U_{t-1} \le r_2 \wedge s_1) P(U_t \le r_1) P(U_{t'-1} \le s_2), & \text{if } t' = t - 1.
\end{cases} \\
& -r_1 r_2 s_1 s_2 \\
= \; & \begin{cases}
r_1 r_2 s_1 s_2 - r_1 r_2 s_1 s_2 = 0, & \text{if } |t - t'| > 1 \\
(r_1 \wedge s_1)(r_2 \wedge s_2) - r_1 r_2 s_1 s_2 =: V_1, & \text{if } t' = t \\
(r_1 \wedge s_2) r_2 s_1 - r_1 r_2 s_1 s_2 =: V_2, & \text{if } t' = t + 1 \\
(r_2 \wedge s_1) r_1 s_2 - r_1 r_2 s_1 s_2 =: V_3, & \text{if } t' = t - 1.
\end{cases}
\end{aligned}
$$

Then covariance of the process will be

$$
\begin{aligned}
& EV_{2n}(r)V_{2n}(s) \\
= \ & \frac{1}{(n-1)} \sum_{t=2}^{n} \sum_{t'=2}^{n} E\left[I(U_t \le r_1, U_{t-1} \le r_2) - r_1 r_2\right]\left[I(U_{t'} \le s_1, U_{t'-1} \le s_2) - s_1 s_2\right] \\
= \ & \frac{1}{(n-1)} \left[\sum_{\substack{t=2 \\ t'=t}}^{n} V_1 + \sum_{\substack{t=2 \\ t'=t+1}}^{n-1} V_2 + \sum_{\substack{t=3 \\ t'=t-1}}^{n} V_3 \right] \\
= \ & \frac{1}{n-1}\left[(n-1)V_1 + (n-2)V_2 + (n-2)V_3\right] \stackrel{n\to\infty}{\to} V_1 + V_2 + V_3.
\end{aligned}
$$

Lemma 1 *Under Assumptions 1, 2, 3 (6), 4, the following asymptotic representation holds under G_{nt} (under $\{\theta_n : n \ge 1\}$)*

$$
\hat{V}_{2n}(r) = V_{2n}(r) + \sqrt{n_0}k(r) + h(r)'\sqrt{n}(\hat{\theta} - \theta_n) + o_p(1)
$$

uniform in r.

NOTE: this lemma holds for data generated by
1) $F_t(\cdot|\Omega_t, \theta_0)$, by putting $n_0 = 0$ and $\theta_n = \theta_0$, under Assumptions 1, 2, 3;
2) $F_t(\cdot|\Omega_t, \theta_n)$ with $\theta_n - \theta_0 = o(1)$, by putting $n_0 = 0$, under Assumptions 1, 2, 6;
3) $G_{nt}(\cdot|\Omega_t, \theta_0)$, by putting $\theta_n = \theta_0$, under Assumptions 1, 2, 3, 4.

Proof of Lemma 1.

For $i \in \{1, 2\}$ denote $\xi_t(r_i, a, b) = F_t(F_t^{-1}(r_i|a)|b)$, $\zeta_t(r_i, a, b) = H_t(F_t^{-1}(r_i|a)|b)$ and $e_t(r_i, a, b) = \frac{\sqrt{n_0}}{\sqrt{n}}\left[\zeta_t(r_i, a, b) - \xi_t(r_i, a, b)\right]$. For particular arguments $\hat{\theta}$ and θ_n, we will use even shorter notation: $\xi_{t,i} = F_t(F_t^{-1}(r_i|\hat{\theta})|\theta_n)$, $\zeta_{t,i} = H_t(F_t^{-1}(r_i|\hat{\theta})|\theta_n)$ and $e_{t,i} = \frac{\sqrt{n_0}}{\sqrt{n}}\left[\zeta_t(r_i, \hat{\theta}, \theta_n) - \xi_t(r_i, \hat{\theta}, \theta_n)\right]$.

$$
\begin{aligned}
\hat{U}_t \le r_i \quad & \Leftrightarrow \quad F_t(Y_t|\hat{\theta}) \le r_i \\
& \Leftrightarrow \quad F_t(G_{nt}^{-1}(U_t^*|\theta_n)|\hat{\theta}) \le r_i \\
& \Leftrightarrow \quad U_t^* \le G_{nt}(F_t^{-1}(r_i|\hat{\theta})|\theta_n) \\
& \Leftrightarrow \quad U_t^* \le \xi_{t,i} + e_{ti}.
\end{aligned}
$$

Then

$$
\begin{aligned}
\hat{V}_{2n}(r) \quad & = \quad \frac{1}{\sqrt{n-1}} \sum_{t=2}^{n} \left[I(\hat{U}_t \le r_1, \hat{U}_{t-1} \le r_2) - r_1 r_2 \right] \\
& = \quad \frac{1}{\sqrt{n-1}} \sum_{t=2}^{n} \left[I(U_t^* \le \xi_{t,1} + e_{t,1}, U_{t-1}^* \le \xi_{t-1,2} + e_{t-1,2}) - r_1 r_2 \right] \\
& = \quad V_{2n}(r) + d_n(r) + R_n(r, \hat{\theta}),
\end{aligned}
$$

where

$$
\begin{aligned}
d_n(r) &= -\frac{1}{\sqrt{n-1}}\sum_{t=2}^{n}\left[r_1r_2-(\xi_{t,1}+e_{t,1})(\xi_{t-1,2}+e_{t-1,2})\right] \\
&= -\frac{1}{\sqrt{n-1}}\sum_{t=2}^{n}\Big[\xi_t(r_1,\hat\theta,\hat\theta)\xi_{t-1}(r_2,\hat\theta,\hat\theta)-\xi_t(r_1,\hat\theta,\theta_n)\xi_{t-1}(r_2,\hat\theta,\theta_n) \\
&\quad -\xi_{t-1,2}e_{t,1}-\xi_{t,1}e_{t-1,2}-O_p\left(\frac{1}{n}\right)\Big] \\
&= -\frac{1}{\sqrt{n-1}}\sum_{t=2}^{n}\frac{\partial}{\partial\theta}\left[\xi_t(r_1,\hat\theta,\theta)\xi_{t-1}(r_2,\hat\theta,\theta)\right]\Big|_{\theta=\theta^*}(\hat\theta-\theta_n)+\sqrt{n_0}k(r)+o_p\left(1\right) \\
&= -\frac{1}{n-1}\sum_{t=2}^{n}\left[\left.\frac{\partial\xi_t(r_1,\hat\theta,\theta)}{\partial\theta}\right|_{\theta=\theta^*}\xi_{t-1}(r_2,\hat\theta,\theta^*)+\xi_t(r_1,\hat\theta,\theta^*)\left.\frac{\partial\xi_{t-1}(r_2,\hat\theta,\theta)}{\partial\theta}\right|_{\theta=\theta^*}\right] \\
&\quad \times\sqrt{n-1}(\hat\theta-\theta_n)+\sqrt{n_0}k(r)+o_p\left(\frac{1}{\sqrt{n}}\right) \\
&= -\frac{1}{n-1}\sum_{t=2}^{n}h_t(r,\hat\theta,\theta^*(r))\sqrt{n-1}(\hat\theta-\theta_n)+\sqrt{n_0}k(r)+o_p\left(\frac{1}{\sqrt{n}}\right) \\
&= -h(r)'\sqrt{n}(\hat\theta-\theta_n)+\sqrt{n_0}k(r)+o_p(1),
\end{aligned}
$$

uniformly in r by mean value theorem for some θ^* between $\hat\theta$ and θ_0 (or Taylor expansion around $\hat\theta$). We have used

$$
e_{t,1}e_{t-1,2}=\frac{n_0}{n}\left[\zeta_{t,1}-\xi_{t,1}\right]\left[\zeta_{t-1,2}-\xi_{t-1,2}\right]=O_p\left(\frac{1}{n}\right)
$$

uniformly in r and t and

$$
\begin{aligned}
&\frac{1}{\sqrt{n-1}}\sum_{t=2}^{n}\xi_{t-1,2}e_{t,1}+\xi_{t,1}e_{t-1,2} \\
&=\frac{\sqrt{n_0}}{\sqrt{(n-1)n}}\sum_{t=2}^{n}\left(\xi_{t-1,2}\left[\zeta_{t,1}-\xi_{t,1}\right]+\xi_{t,1}\left[\zeta_{t-1,2}-\xi_{t-1,2}\right]\right) \\
&=\frac{\sqrt{n_0}}{\sqrt{(n-1)n}}\sum_{t=2}^{n}\left(\left[r_2+o_p(1)\right]\left[\zeta_{t,1}-r_1+o_p(1)\right]+\left[r_1+o_p(1)\right]\left[\zeta_{t-1,2}-r_2+o_p(1)\right]\right) \\
&=\frac{\sqrt{n_0}}{n}\sum_{t=2}^{n}\left(r_2\left[\zeta_{t,1}-r_1\right]+r_1\left[\zeta_{t-1,2}-r_2\right]\right)+o_p(1) \\
&=\sqrt{n_0}k(r)+o_p(1)
\end{aligned}
$$

since $\hat\theta$ is a consistent estimator of θ_n and by Assumption 1

$$
\xi_{t,i}-r_i=F_t(F_t^{-1}(r_i|\hat\theta)|\theta_n)-r_i=F_t(F_t^{-1}(r_i|\hat\theta)|\theta_n)-F_t(F_t^{-1}(r_i|\hat\theta)|\hat\theta)=o_p(1)
$$

uniformly in r and t. The rest

$$\begin{aligned} R_n(r,\hat{\theta}) &= \frac{1}{\sqrt{n-1}}\sum_{t=2}^{n}[I(U_t^* \le \xi_{t,1}+e_{t,1}, U_{t-1}^* \le \xi_{t-1,2}+e_{t-1,2}) \\ &\quad -(\xi_{t,1}+e_{t,1})(\xi_{t-1,2}+e_{t-1,2}) \\ &\quad -I(U_t^* \le r_1, U_{t-1}^* \le r_2)+r_1r_2] \end{aligned}$$

is $o_p(1)$ uniformly in r.

Lemma 2 *Under assumptions 1, 2, 3 (6) under G_{nt} (under $\{\theta_n : n \ge 1\}$) we have*

$$\begin{pmatrix} V_{2n}(r) \\ \sqrt{n}(\hat{\theta}-\theta_n) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} V_{2\infty}(r) \\ \psi_\infty + m \end{pmatrix}.$$

Proof of Lemma 2. Repeating the argument under the null, under $\{\theta_n : n \ge 1\}$ we have $V_{2n}(r) \xrightarrow{d} V_{2\infty}(r)$ and $\sqrt{n}(\hat{\theta}-\theta_0)$ is asymptoticaly $N(m,\Psi)$ by CLT for MDS. Denote

$$v_t = \begin{pmatrix} I(U_{nt} \le r_1)I(U_{nt-1} \le r_2) - r_1r_2 \\ \psi(\Omega_{nt})l(U_{nt}) \end{pmatrix}$$

To prove vector convergence we use functional CLT of Pollard (1984, Theorem 10.12)). We need to check equicontinuity and convergence of fidis. Equicontinuity follows from the fact that equicontiuty of the vector is equivaletn to equicontinuity of its components, equicontinuity of the first component provided in Proposition **2**, the second components is equiconinuous automaticaly since it does not depend on the parameter.

To check convergence of fidis we apply CLT to v_t, where the first component of the vector is the normalized sum of 1-dependent processes, while the second is the sum of MDS. One way to prove CLT for the sum of 1-dependent rv's is to split them in groups skipping 1 element between to form the sum of iid rv's. Fix $p > 1$and denote $m = \lfloor \frac{n}{p} \rfloor$. Consider the following decomposition

$$\begin{aligned} \begin{pmatrix} V_{2n}(r) \\ \sqrt{n}(\hat{\theta}-\theta_n) \end{pmatrix} &= \frac{1}{\sqrt{n-1}}\sum_{t=2}^{n} v_t + o_p(1) \\ &= \frac{1}{\sqrt{n-1}}\sum_{t=2}^{p} v_t + v_{p+1} + \sum_{t=p+2}^{2p} v_t + v_{2p+1} + \ldots \\ &\quad + \sum_{t=(m-1)p+2}^{t=mp} v_t + C_{n,p} + o_p(1) \\ &= A_{n,p} + B_{n,p} + C_{n,p} + o_p(1) \end{aligned}$$

where $A_{n,p} = \frac{1}{\sqrt{n-1}} \sum_{k=1}^{m} A_{n,p,k}$ denotes the sum of blocks $A_{n,p,k} = \sum_{t=(k-1)p+2}^{kp} v_t$ of lenght $p-1$, $B_{n,p} = \frac{1}{\sqrt{n-1}} \sum_{i=1}^{(m-1)} v_{ip+1}$ and $C_{n,p}$ denotes the sum of remaining terms, no more than $p+1$ terms. We will show now that $A_{n,p}$ converges to the right limit, $B_{n,p}$ and $C_{n,p}$ are $o_p(1)$. $A_{n,p,k}$ is a martingale difference array (MDA) with respect to the new filtration $\Omega'_k = \Omega_{(k-1)p+1}$. Indeed, for the first component $A^{(1)}_{n,p,k}$ of $A_{n,p,k}$ we have

$$\begin{aligned} E\left(A^{(1)}_{n,p,k}|\Omega'_l\right) &= \sum_{t=(k-1)p+2}^{kp} E\left(I(U_{nt} \leq r_1)I(U_{nt-1} \leq r_2) - r_1 r_2|\Omega_{(l-1)p+1}\right) = \\ & \begin{cases} 0, \text{ for } k \geq l \\ A^{(1)}_{n,p,k}, \text{ for } k < l \end{cases}. \end{aligned}$$

For the second component $A^{(2)}_{n,p,k}$ of $A_{n,p,k}$ we have $E\left(A^{(2)}_{n,p,k}|\Omega'_l\right) = A^{(2)}_{n,p,k}$ for $k < l$ and for $k \geq l$

$$\begin{aligned} E\left(A^{(2)}_{n,p,k}|\Omega'_l\right) &= \sum_{t=(k-1)p+2}^{kp} E\left(\psi(\Omega_{nt})l(U_{nt})|\Omega_{(l-1)p+1}\right) = \\ &= \sum_{t=(k-1)p+2}^{kp} E\left(E\left(\psi(\Omega_{nt})l(U_{nt})|\Omega_t\right)|\Omega_{(l-1)p+1}\right) \\ &= \sum_{t=(k-1)p+2}^{kp} E\left(\psi(\Omega_{nt})E\left(l(U_{nt})|\Omega_t\right)|\Omega_{(l-1)p+1}\right) = 0. \end{aligned}$$

There for we can apply CLT for MDA to obtain the following

$$A_{n,p} = \frac{1}{\sqrt{n-1}} \sum_{k=1}^{m} A_{n,p,k} = \sqrt{\frac{m}{n-1}} \frac{1}{\sqrt{m}} \sum_{k=1}^{m} A_{n,p,k} \xrightarrow{d} N(0, V_p)$$

where variance

$$\begin{aligned} V_p &= \frac{1}{p} \operatorname*{plim}_{n\to\infty} \frac{1}{m} \sum_{k=1}^{m} E A_{n,p,k} A'_{n,p,k} \\ &= \frac{1}{p} \operatorname*{plim}_{n\to\infty} \frac{1}{m} \sum_{k=1}^{m} E\left(\sum_{t=(k-1)p+2}^{kp} v_t\right)\left(\sum_{t=(k-1)p+2}^{kp} v_t\right)' \\ &= \operatorname*{plim}_{n\to\infty} \frac{1}{m} \sum_{k=1}^{m} \left(\frac{1}{p} \sum_{t=(k-1)p+2}^{kp} E v_t v'_t + \frac{2}{p} \sum_{t=(k-1)p+2}^{kp-1} E v_t v'_{t+1}\right) \end{aligned}$$

(if stationary)for say $k = 1$

$$\begin{aligned}
V_p &= \frac{1}{p} E A_{n,p,k} A'_{n,p,k} \\
&= \frac{1}{p} E \left(\sum_{t=(k-1)p+2}^{kp} v_t \right) \left(\sum_{t=(k-1)p+2}^{kp} v_t \right)' \\
&= \left(\frac{1}{p} \sum_{t=(k-1)p+2}^{kp} E v_t v'_t + \frac{2}{p} \sum_{t=(k-1)p+2}^{kp-1} E v_t v'_{t+1} \right) \\
&= \left(\frac{p-1}{p} E v_t v'_t + \frac{(p-2)}{p} \left(E v_t v'_{t+1} + E v_{t+1} v'_t \right) \right) \\
&\overset{p\to\infty}{\to} E v_t v'_t + E v_t v'_{t+1} + E v_{t+1} v'_t
\end{aligned}$$

Next we apply Bernstein lemma (Billingsley 1999, Theorem 3.2) for $Y_{n,p} = A_{n,p}$, $Z_{n,p} = B_{n,p} + C_{n,p}$

$$\begin{aligned}
P\left(|Z_{n,p}| > \varepsilon\right) &\leq \frac{E Z'_{n,p} Z_{n,p}}{\varepsilon^2} = \frac{E\left(B_{n,p} + C_{n,p}\right)'\left(B_{n,p} + C_{n,p}\right)}{\varepsilon^2} \\
&\leq \frac{2\left[E B'_{n,p} B_{n,p} + E C'_{n,p} C_{n,p}\right]}{\varepsilon^2} \\
&\leq \frac{2}{(n-1)\varepsilon^2} \left[\sum_{i=1}^{(m-1)} E v'_{ip+1} v_{ip+1} + 2^{p+1}(p+1) \max_{t=mp,..,n} E v'_t v_t \right].
\end{aligned}$$

Hence

$$0 \leq \limsup_{p\to\infty} \limsup_{n\to\infty} P\left(|Z_{n,p}| > \varepsilon\right) \leq \limsup_{p\to\infty} \frac{2}{p\varepsilon^2} C = 0.$$

Next, for fixed p

$$Y_{n,p} = A_{n,p} \xrightarrow{d} N(0, V_p)$$

and for $p \to \infty$

$$Y_{n,p} = A_{n,p} \overset{p\to\infty}{\to} N(0, E v_t v'_t + E v_t v'_{t+1} + E v_{t+1} v'_t)$$

and we get the result

$$A_{n,p} + B_{n,p} + C_{n,p} \xrightarrow{d} N(0, E v_t v'_t + E v_t v'_{t+1} + E v_{t+1} v'_t).$$

Now we derive the covariance between two components. Under $\{\theta_n : n \geq 1\}$

$$
\begin{aligned}
& Cov(V_{2n}(r), \frac{1}{\sqrt{n}}\sum_{t'=2}^{n}\psi(\Omega_{nt'})l(U_{nt'})) \\
= \; & \frac{1}{n}\sum_{t=2}^{n}\sum_{t'=2}^{n}E\left[(I(U_{nt}\le r_1, U_{nt-1}\le r_2) - r_1 r_2)\,\psi(\Omega_{nt'})l(U_{nt'})\right] \\
= \; & \frac{1}{n}\sum_{t=2}^{n}\sum_{\substack{t'=2 \\ t'>t}}^{n}EE\left[I\left(U_{nt}\le r_1, U_{nt-1}\le r_2\right)\psi(\Omega_{nt'})l(U_{nt'})|\Omega_{nt'}\right] \\
& +\frac{1}{n}\sum_{t=2}^{n}\sum_{\substack{t'=2 \\ t'<t-1}}^{n}EE\left[I\left(U_{nt}\le r_1, U_{nt-1}\le r_2\right)\psi(\Omega_{nt'})l(U_{nt'})|\Omega_{nt}\right] \\
& +\frac{1}{n}\sum_{\substack{t=2 \\ t'=t}}^{n}EE\left[I\left(U_{nt}\le r_1, U_{nt-1}\le r_2\right)\psi(\Omega_{nt'})l(U_{nt'})|\Omega_{nt'}\right] \\
& +\frac{1}{n}\sum_{\substack{t=3 \\ t'=t-1}}^{n}EE\left[I\left(U_{nt}\le r_1, U_{nt-1}\le r_2\right)\psi(\Omega_{nt'})l(U_{nt'})|\Omega_{nt}\right] \\
= \; & \frac{1}{n}\sum_{t=2}^{n}\sum_{\substack{t'=2 \\ t'>t}}^{n}E\left[I\left(U_{nt}\le r_1, U_{nt-1}\le r_2\right)\psi(\Omega_{nt'})E\left(l(U_{nt'})|\Omega_{nt'}\right)\right] \\
& +\frac{1}{n}\sum_{t=2}^{n}\sum_{\substack{t'=2 \\ t'<t-1}}^{n}E\left[\psi(\Omega_{nt'})l(U_{nt'})E\left(I\left(U_{nt}\le r_1, U_{nt-1}\le r_2\right)|\Omega_{nt-1}\right)\right] \\
& +\frac{1}{n}\sum_{t=2}^{n}E\left[I(U_{nt-1}\le r_2)\psi(\Omega_{nt})E\left(I(U_{nt}\le r_1)l(U_{nt})|\Omega_{nt}\right)\right] \\
& +\frac{1}{n}\sum_{t=3}^{n}E\left[I(U_{nt-1}\le r_2)D_0\psi(\Omega_{nt-1})l(U_{nt-1})E\left(I(U_{nt}\le r_1)|\Omega_{nt}\right)\right] \\
= \; & 0 \\
& +r_1 r_2\sum_{t=2}^{n}\sum_{\substack{t'=2 \\ t'<t-1}}^{n}E\left[D_0\psi(\Omega_{nt'})l(U_{nt'})\right] \\
& +\frac{1}{n}\sum_{t=2}^{n}E\left[I(U_{nt-1}\le r_2)D_0\psi(\Omega_{nt})E\left(I(U_{nt}\le r_1)l(U_{nt})|\Omega_{nt}\right)\right] \\
& +\frac{1}{n}\sum_{t=3}^{n}E\left[I(U_{nt-1}\le r_2)D_0\psi(\Omega_{nt-1})l(U_{nt-1})E\left(I(U_{nt}\le r_1)|\Omega_{nt}\right)\right] \\
= \; & \frac{1}{n}\sum_{t=2}^{n}E\left[I(U_{nt-1}\le r_2)D_0\psi(\Omega_{nt})E\left(I(U_{nt}\le r_1)l(U_{nt})|\Omega_{nt}\right)\right] \\
& +\frac{1}{n}r_1\sum_{t=3}^{n}E\left[I(U_{nt-1}\le r_2)D_0\psi(\Omega_{nt-1})l(U_{nt-1})\right]
\end{aligned}
$$

or

$$
\begin{aligned}
&= \frac{1}{n}\sum_{t=2}^{n} E\left[\left(I(U_{nt}\le r_1)I(U_{nt-1}\le r_2)+I(U_{nt+1}\le r_1)I(U_{nt}\le r_2)\right)D_0\psi(\Omega_{nt})l(U_{nt})\right] \\
&= \frac{1}{n}\sum_{t=2}^{n} E\left[\left(I(U_{nt}\le r_1)I(U_{nt-1}\le r_2)+r_1 I(U_{nt}\le r_2)\right)\psi(\Omega_{nt})l(U_{nt})\right]
\end{aligned}
$$

Now, using result from Lemma 2 we have the following.

Proof of Proposition 3 a) and b). Apply functional CMT, see Pollard (1984, Theorem IV.12, p.70).

Proof of Proposition 3 c). Apply functional CLT and CMT, see Pollard (1984, Theorem IV.12, p.70).

Proof of Proposition 4. Apply functional CMT, see Pollard (1984, Theorem IV.12, p.70).

Proof of Proposition 5 See discussion on pages 1107-1109 of Andrews (1997).

8 Computation of norms of V_{kn} process

In this section we derive algebraic formula to compute our statistics to avoid numeric integration. Since $V_{kn}(r)$ is piece-wise linear, the computation of norms of this process can be significantly simplified. We divide domain $[0,1]^k$ into parts where indicators are constant, compute statistics there, and then aggregate information.

8.1 Cramer-von-Mises norm of V_n process

Assume $U_t \in (0,1)$, $t = 1,...,n$ are all different. Let $U_{(t)}$ be ordered series U_t for $t = 1,...,n$ and $U_{(0)} = 0$ and $U_{(n+1)} = 1$. We divide the unit interval $[0,1]$ on $n+1$ subintervals with endpoints at $U_{(t)}$ and use additivity of the integrals

$$\int_{[0,1]} V_{1n}(r)^2 dr = \sum_{k=0}^{n}\int_{U_{(k)}}^{U_{(k+1)}} V_{1n}(r)^2 dr.$$

On each subinterval we use piece-wise linearity of the statistics

$$V_{1n}(r) = \sqrt{n}\left[\frac{1}{n}\sum_{t=1}^{n} I(U_t \le r) - r\right] = \sqrt{n}\left(\frac{k}{n} - r\right) \text{ for } r \in [U_{(k)}, U_{(k+1)}).$$

Therefore

$$\int_{U_{(k)}}^{U_{(k+1)}} V_{1n}(r)^2 dr = \int_{U_{(k)}}^{U_{(k+1)}} n\left[\frac{1}{n}\sum_{t=1}^{n} I(U_t \le r) - r\right]^2 dr = n\int_{U_{(k)}}^{U_{(k+1)}} \left(\frac{k}{n} - r\right)^2 dr$$

$$= n\int_{U_{(k)}}^{U_{(k+1)}} \left(\frac{k^2}{n^2} - 2\frac{k}{n}r + r^2\right) dr = n\left(\frac{k^2}{n^2}r - \frac{k}{n}r^2 + \frac{1}{3}r^3\right)\Bigg|_{U_{(k)}}^{U_{(k+1)}}.$$

8.2 Cramer-von-Mises norm of V_{2n} process

Assume $U_t^{(i)} \in (0,1)$, $t = 1,\ldots,n$ are different for each $i = 1,2$. Let $U_{(t)}^{(i)}$ be ordered series $U_t^{(i)}$ and $U_{(0)}^{(i)} = 0$ and $U_{(n+1)}^{(i)} = 1$. We divide the unit square $[0,1]\times[0,1]$ on $(n+1)\times(n+1)$ squares $Q_{lm} = [U_{(l)}^{(1)}, U_{(l+1)}^{(1)}) \times [U_{(m)}^{(2)}, U_{(m+1)}^{(2)})$, $l,m = 0,\ldots,n$ and use additivity of the integrals

$$\int_{[0,1]^2} V_{2n}(r)^2 dr = \sum_{l=0}^{n}\sum_{m=0}^{n}\int_{Q_{lm}} V_{2n}(r)^2 dr.$$

On each square we use linearity of the statistics

$$V_{2n}(r) = \sqrt{n}\left[\frac{1}{n}\sum_{t=1}^{n} I(U_t^{(1)} \le r_1) I(U_t^{(2)} \le r_2) - r_1 r_2\right] = \sqrt{n}\,(a_{lm} - r_1 r_2) \text{ for } r \in Q_{lm},$$

where

$$a_{lm} = \frac{1}{n}\sum_{t=1}^{n} I(U_t^{(1)} \le U_{(l)}^{(1)}) I(U_t^{(2)} \le U_{(m)}^{(2)}).$$

Therefore

$$\int_{Q_{lm}} V_{2n}(r)^2 dr = \int_{[U_{(m)}^{(2)}, U_{(m+1)}^{(2)})}\int_{[U_{(l)}^{(1)}, U_{(l+1)}^{(1)})} n\,(a_{lm} - r_1 r_2)^2 \, dr_1 dr_2$$

$$= n\int_{[U_{(m)}^{(2)}, U_{(m+1)}^{(2)})}\int_{[U_{(l)}^{(1)}, U_{(l+1)}^{(1)})} (a_{lm}^2 - 2a_{lm} r_1 r_2 + r_1^2 r_2^2)\, dr_1 dr_2$$

$$= n a_{lm}^2 \left(U_{(l+1)}^{(1)} U_{(m+1)}^{(2)} - U_{(l+1)}^{(1)} U_{(m)}^{(2)} - U_{(l)}^{(1)} U_{(m+1)}^{(2)} + U_{(l)}^{(1)} U_{(m)}^{(2)}\right)$$

$$-n\frac{1}{2}a_{lm}\left({U_{(l+1)}^{(1)}}^2 {U_{(m+1)}^{(2)}}^2 - {U_{(l+1)}^{(1)}}^2 {U_{(m)}^{(2)}}^2 - {U_{(l)}^{(1)}}^2 {U_{(m+1)}^{(2)}}^2 + {U_{(l)}^{(1)}}^2 {U_{(m)}^{(2)}}^2\right)$$

$$+n\frac{1}{9}\left({U_{(l+1)}^{(1)}}^3 {U_{(m+1)}^{(2)}}^3 - {U_{(l+1)}^{(1)}}^3 {U_{(m)}^{(2)}}^3 - {U_{(l)}^{(1)}}^3 {U_{(m+1)}^{(2)}}^3 + {U_{(l)}^{(1)}}^3 {U_{(m)}^{(2)}}^3\right)$$

Computation of Cramer-von-Mises norm of V_{kn}, $k > 2$, can be done in a similar way.

8.3 Kolmogorov-Smirnov norm of V_{kn} process

For generalized process

$$V_{kn}(r) = \sqrt{n}\left[\frac{1}{n}\sum_{t=1}^{n}\prod_{j=1}^{k} I(U_t^{(j)} \le r_j) - r_1 r_2 \dots r_k\right].$$

define Kolmogorov-Smirnov norm as

$$\sup_{r\in[0,1]^k} |V_{kn}(r)|.$$

Assume $U_t^{(i)} \in (0,1),\ t = 1,...,n$ are different for each $i = 1,...,k$. Let $U_{(t)}^{(i)}$ be ordered series $U_t^{(i)}$ and $U_{(0)}^{(i)} = 0$ and $U_{(n+1)}^{(i)} = 1$. We divide the unit cube $[0,1]^k$ on $(n+1)^k$ cubes $Q_l = [U_{(l_1)}^{(1)}, U_{(l_1+1)}^{(1)}) \times \dots \times [U_{(l_k)}^{(k)}, U_{(l_k+1)}^{(k)})$ where $l = (l_1,\dots,l_k), l_i = 0,\dots,n$, then

$$\sup_{r\in[0,1]^k} |V_{kn}(r)| = \max_{l} \sup_{r\in Q_l} |V_{kn}(r)|.$$

We do not need to consider "border" points (r, s.t. some $r_i = 1$) since the value of statistics at these points can be approximated by values of statistics at inside points, i.e

$$\forall r \notin [0,1)^k \qquad \exists r^{<m>} \in [0,1)^k \qquad V_{kn}(r^{<m>}) \to V_{kn}(r).$$

On each cube the statistics is linear, therefore extremum arguments are on the corners

$$\arg\sup_{r\in Q_l} |V_{kn}(r)| = \arg\sup_{r\in Q_l} \sqrt{n}\left|\frac{1}{n}\sum_{t=1}^{n}\prod_{j=1}^{k} I(U_t^{(j)} \le r_j) - r_1 r_2 \dots r_k\right|$$

$$= \arg\sup_{r\in Q_l} \sqrt{n}\,|a_l - r_1 r_2 \dots r_k| = \arg\max_{r=c_l,c_{l+1}} \sqrt{n}\,|a_l - r_1 r_2 \dots r_k|$$

where $c_l = (U_{(l_1)}^{(1)},\dots,U_{(l_k)}^{(k)})$ and $c_{l+1} = (U_{(l_1+1)}^{(1)},\dots,U_{(l_k+1)}^{(k)})$ is the bottom-left ("smallest") and top-right ("biggest") corner points of the square Q_l and

$$a_l = \frac{1}{n}\sum_{t=1}^{n}\prod_{i=1}^{n} I(U_t^{(i)} \le U_{(l)}^{(i)}) = \frac{1}{n}\sum_{t=1}^{n}\prod_{i=1}^{n} I(U_t^{(i)} < U_{(l+1)}^{(i)}).$$

Thus

$$\sup_{r\in Q_l} |V_{kn}(r)| = \max\left(\sqrt{n}\left|a_l - U_{(l_1)}^{(1)} U_{(l_2)}^{(2)} \dots U_{(l_k)}^{(k)}\right|, \sqrt{n}\left|a_l - U_{(l_1+1)}^{(1)} U_{(l_2+1)}^{(2)} \dots U_{(l_k+1)}^{(k)}\right|\right).$$

If we define the strong analog of V_{kn} as

$$\tilde{V}_{kn}(r) = \sqrt{n} \left[\frac{1}{n} \sum_{t=1}^{n} \prod_{j=1}^{k} I(U_t^{(j)} < r_j) - r_1 r_2 \dots r_k \right],$$

then the last equation can be rewritten as

$$\sup_{r \in Q_l} |V_{kn}(r)| = \max \left(|V_{kn}(c_l)|, |\tilde{V}_{kn}(c_{l+1})| \right).$$

References

[1] Andrews, D.W.K. (1994) Empirical process methods in econometrice. Handbook of Econometrics vol. IV

[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, The Review of Economics and Statistics, 85(3): 531-549.

[4] Blum, J. R., Kiefer, J. and Rosenblatt, M. (1961) Distribution free tests of independence based on sample distribution function. Ann. Nath. Statist., 32, 485-98.

[5] Bollerslev, T. (1986) "Generalized autoregressive conditional heteroscedasticity," Journal of Econometrics, 31, 307-327.

[6] Chen Yi-Ting, (2006) Moment tests for standardized error distributions: a simple robust approach.

[7] Delgado, M. (1996) Testing serial independence using the sample distribution function, Journal of Time Series Analysis 17, 271-285.

[8] Delgado, M. and Mora (2000) A nonparametric test for serial independence of regression errors, Biometrika, 87, 1, pp 228-234.

[9] Delgado, M. and Stute W. (2008)

[10] Durbin, J. (1973): "Weak convergence of sample distribution functions when parameters are estimated," Annals of Statistics, 1, 279-290.

[11] Engle, R.F., (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50, 987-1007.

[12] Engle, R.F., and Russell (1998). Autoregressive conditional duration: a new model for irregulary spaced data. Econometrica 66, 5, 1127-1162.

[13] Ghoudi K., Kulperger R. and Remillard B. (2001) a nonparametric test of serial independence for time series residuals. Journal of Multivariate Analysis 79, 191-218.

[14] Hoeffding W. (1948). A nonparametric test of independence. Ann. Math. Statist. 26, 189-211.

[15] Hong, Y., (1998). Testing for pairwise serial independence via the empirical distribution function. Journal Royal Statistical Society 60, part 2, 429-453.

[16] Hong, Y., and Lee, Y. J., (2003). Diagnostic checking for adequacy of noninear time series model. Econometric theory, 19, 2003, 1065-1121 72, 499-541.

[17] Hong, Y., and Li, Y. J., (2005). Generalized spectral tests for conditional mean models in time series with conditional heteroskedasticity of unknown form. Review of Economic Studies 72, 499-541.

[18] Khmaladze, E.V. (1981): "Martingale approach in the theory of goodness-of-tests," Theory of probability and its applications, XXVI, 240-257.

[19] Koul, H.L. and W. Stute (1999): "Nonparametric model checks for time series," Annals of Statistics, Vol 27, 204-236.

[20] Lancaster (1979) Econometric methods for the duration of unemployment, Econometrica, 47, 939-956.

[21] Lee and Hansen (1994) Asymptotic theory for the GARCH(1,1) QMLE, Econometric theory 10, 29-52.

[22] Politis, Romano and Wolf (1999). Subsampling. New York: Springer-Verlag.

[23] Pollard D. (1984) Convergence of stochastic processes. New York: Springer-Verlag.

[24] Pollard D. (1990) Empirical processes: Theory and application, CBMS Conferences Series in Probability and Statistics, Vol. 2. Hayward, CA: Institute of Mathematical Statistics.

[25] Rosenblatt, M. (1952) Remarks on a Multivariate Transformation. Annals of Mathematical Statistics 23(3):47072.

[26] Rosenblatt, M. (1975) A quadratic measure of deviation of two-dimensional density estimates and a test of independence, Ann. Stat, 3, 1-14.

[27] Segal, I. E. (1938). Fudicial Distribution of Several Parameters with Application to a Normal System. Cambridge Philosophical Society Proceedings 33:4147.

[28] Shao and Dongsheng (1995) The Jackknife and bootstrap. New York: Springer-Verlag.

[29] Skaug, H. J. and Tjøstheim, D. (1993a) Nonparametric test of serial independence based on the empirical distribution function, Biometrika 80, 591-602.

[30] Stute, W., (1997). Nonparametic model checks for regression. Annals of Statistics 25, 613-641.