行政院國家科學委員會專題研究計畫 成果報告

新的一致性條件動差檢定
研究成果報告(精簡版)

計 畫 類 別 ：個別型
計 畫 編 號 ：NSC 96-2415-H-004-008-
執 行 期 間 ：96 年 08 月 01 日至 97 年 10 月 31 日
執 行 單 位 ：國立政治大學經濟學系

計 畫 主 持 人 ：林馨怡

計畫參與人員： 碩士班研究生-兼任助理人員：周東慶

處 理 方 式 ：本計畫可公開查詢

中 華 民 國 97 年 12 月 11 日

1 Introduction

Many economic and econometric models are represented by conditional moment restrictions, for example, the rational expectation model, the market disequilibrium model, the conditional probability model, the discrete choice model and the nonlinear simultaneous equations model. The validity of those model specifications is evaluated by testing the associated moment conditions. The resulting test is the conditional moment test or M-test which has been developed by Newey (1985), Tauchen (1985), and White (1987). However, these conditional moment tests may not be consistent because they check only necessary conditions of the conditional moment restrictions and there exist alternatives that cannot be detected by these testing procedures. Therefore, our research focuses on constructing consistent conditional moment tests.

There is an abundance of literature on constructing conditional moment tests. The nonparametric tests employ a nonparametric estimator to a parametric conditional moment function that equals zero under the null. See, for example, Eubank and Spiegelman (1990), Lewbel (1995), Hong and White (1995), Fan and Li (1996), Li and Wang (1998), Chen and Fan (1999), Zheng (1998a, 1998b, 2000), Horowitz and Spokoiny (2001), Delgado and González Manteiga (2001), Li, Hsiao, and Zinn (2003) and Tripathi and Kitamura (2003), among others. However, the test statistics of nonparametric tests are subjective in choosing smoothing parameters and could be computationally costly. Another approach of the conditional moment test is based on infinitely many unconditional moment functions with uncountably many weighted functions indexed by continuous nuisance parameters (Stichcombe and White, 1998). Therefore, a consistent conditional moment test can be obtained by checking these orthogonality functions. Bierens (1982, 1984, 1990), de Jong (1996), Bierens and Ploberger (1997), and Bierens and Ginther (2001) choose the exponential function while Stute (1997), Stute, Thies and Zhu (1998), Koul and Stute (1999), and Stute and Zhu (2002) use the indicator function as the weight functions. The former tests are henceforth called the Bierens test while the latter tests are the the Stute test.

It is noted that both the Bierens and the Stute tests are not asymptotically pivotal in general; that is their limiting distributions depend on model characteristics and critical values cannot be tabulated. For example, the auxiliary nuisance parameters in the exponential weight function of the Bierens test lead to the limiting distributions depending on the data generating process. Although Bierens and Ploberger (1997) derived case-independent upper bounds of the critical values to solve this problem, their test may be too conservative in practice. The Stute test is also not asymptotically pivotal because of the estimation effects (Durbin, 1973) and it is case dependent. Stute, González Manteiga and Presedo Quindïmil

(1998), Whang (2000, 2001, 2004), and Dominguez and Lobato (2006) try to avoid the problem by using the bootstrap to approximate the limiting distributions. Stute, Thies and Zhu (1998), Koul and Stute (1999) and Stute and Zhu (2002) employe the martingale transformation of Khmaladze (1981) to obtain asymptotically distribution-free test statistics. However, due to the difficulty of having a proper definition of a multitime parameter martingales, these tests, except for Song (2007), cannot be carried out with multivariate regressors. Note that in Song (2007), the nonparametric estimation of the conditional moment function is required but a high dimensional nonparametric estimation is very complicated to compute. Kuan and Lin (2008) propose a test which centers the estimation effect out by an average of empirical processes and thus obtained an asymptotically pivotal test. Their test can be applied to multivariate regressor cases and the limiting distribution is a sup-norm of a multi-parameter Kiefer type process.

This paper proposes a consistent conditional moment test by checking an infinite set of unconditional moment conditions with indicator weight functions. The test statistic is based on the subsampling marked empirical process with sample size b instead of the whole size n such that $b < n$. The subsampling, investigated by Politis and Romano (1994) and Politis, Romano and Wolf (1999) is a method of estimating the distribution of an estimator or test statistic by drawing subsamples from the original data. Existing studies in the literature include Andrews and Guggenberger (2005), Chernozhukov and Fernándes-Val (2005), Guggenberger and Wolf (2004), Hong and Scaillet (2006), Linton, Massoumi and Whang (2005) and Whang (2004). The subsampling method has not been used to construct the test statistics, which is done in this article. Other advantages of this paper are: (i) Instead of computing the sample average of the conditional moment function with the whole sample, the test statistic is obtained by the subsampling marked empirical process. The estimation effect disappears when the relative sample size of subsampling to that of the whole sample is zero asymptotically. Therefore, the proposed test does not suffer from the Durbin problem and is asymptotically pivotal. (ii) The present paper differs from the Stute test in that they use a martingale transformation technique which is applied to the model with a univariate regressor, and the test subsamples the marked empirical processes and can be applied to models with multivariate regressors. (iii) Unlike the tests using nonparametric smoothing methods, the proposed test therein avoids the user-chosen smoothing methodology. (iv) The test statistic can be computed using any $\sqrt{n}$-consistent estimator and different estimation methods can be applied. (v) The proposed test does not use the bootstrap, the martingale transformation, or the nonparametric method, resulting in significant simplifications in computing the test statistics. One disadvantage of our test is that it is powerful against local alternatives with rates $b^{-1/2}$, but the proposed test is incapable of detecting local alternatives at rate $n^{-1/2}$. Our test shares the same disadvantage of most nonparametric tests. The Monte Carlo simulations show that the proposed test has good finite sample performances and the test is robust with respect to different values of b.

This paper is arranged as follows. Section 2 presents the conditional moment restriction and the proposed test. Section 3 shows the consistency of the test and the asymptotic behavior under different local alternatives. Section 4 shows the Monte carlo simulation results and follows the conclusions in section 5. All proofs are given in the Appendix.

2 A New Test

2.1 Conditional Moment Restrictions

Consider general conditional moment restrictions

$$\mathbb{E}[m(y, \boldsymbol{x}, \boldsymbol{\beta}_o)\big|\boldsymbol{x}] = 0, \tag{1}$$

where $\mathbb{E}[\cdot|\boldsymbol{x}]$ denotes the expectation conditional on the information set of $\boldsymbol{x}$ and $m(\cdot)$ is a function on data, and $\{y, \boldsymbol{x}\}$ is a sequence of random variables with $\boldsymbol{x} = (x_1, \cdots, x_k)'$ and parameters $\boldsymbol{\beta} \in \boldsymbol{B}$ with $\boldsymbol{B} \in \boldsymbol{R}^k$. The conditional moment restrictions can be obtained from existing models such as the parametric nonlinear regression models where $m(y, \boldsymbol{x}, \boldsymbol{\beta}_o)$ is the difference between y and $g(\boldsymbol{x}', \boldsymbol{\beta})$, with $g(\cdot)$ a nonlinear function. To test the condition moment restrictions, the null and alternative hypotheses are as follows. The null hypothesis is the conditional moment function being equal to zero:

$$H_0 : \mathbb{P}\big\{\mathbb{E}(m(y, \boldsymbol{x}, \boldsymbol{\beta}_o)\big|\boldsymbol{x}) = 0\big\} = 1, \text{ for some } \boldsymbol{\beta}_o \in \boldsymbol{B},$$

against the alternative hypothesis is, for all $\boldsymbol{\beta} \in \boldsymbol{B}$, $\mathbb{E}(m(y, \boldsymbol{x}, \boldsymbol{\beta})\big|\boldsymbol{x}) \neq 0$ with a positive probability:

$$H_1 : \mathbb{P}\big\{\mathbb{E}(m(y, \boldsymbol{x}, \boldsymbol{\beta})\big|\boldsymbol{x}) = 0\big\} < 1, \text{ for all } \boldsymbol{\beta} \in \boldsymbol{B},$$

with $\boldsymbol{B} \in \boldsymbol{R}^k$ a compact set.

As has been shown by Stinchcombe and White (1998), the conditional moment condition (1) equals infinitely many unconditional moment functions

$$\mathbb{E}[m(y, \boldsymbol{x}, \boldsymbol{\beta}_o)\omega(\boldsymbol{x}, \boldsymbol{\xi})] = 0, \forall \boldsymbol{\xi} \in \boldsymbol{R}^k, \tag{2}$$

where $\omega(\cdot)$ is an infinite set of weights indexed by continuous parameters $\boldsymbol{\xi}$ and $\omega(\cdot)$ may be any analytic function that is not polynomial. Therefore, testing (2) constructs a consistent conditional moment test. For example, Bierens (1982, 1984, 1990), de Jong (1996) and Bierens and Ploberger (1997) and Bierens and Ginther (2001) use the exponential weighted function $\omega(\boldsymbol{x}, \boldsymbol{\xi}) = \exp(\boldsymbol{x}'\boldsymbol{\xi})$ for their integrated conditional moment test. Stute (1997), Stute, Thies and Zhu (1998), Koul and Stute (1999) and Stute and Zhu (2002) take the indicator function

$$\omega(\boldsymbol{x}, \boldsymbol{\xi}) = \mathbb{1}_{\{\boldsymbol{x} \leq \boldsymbol{\xi}\}} := \mathbb{1}_{\{x_1 \leq \xi_1\}} \cdots \mathbb{1}_{\{x_k \leq \xi_k\}},$$

where $\mathbb{1}_A$ denotes the indicator function of even A. The present paper implements the indicator function and the conditional moment restrictions (1) can be rewritten by the infinitely many unconditional moment functions as follows:

$$\mathbb{E}[m(y, \boldsymbol{x}, \boldsymbol{\beta}_o) \mathbb{1}_{\{\boldsymbol{x} \leq \boldsymbol{\xi}\}}] = 0, \forall \boldsymbol{\xi} = (\xi_1, \cdots, \xi_k)' \in \boldsymbol{R}^k. \tag{3}$$

The above moment functions allow for multivariate regressors.

2.2 Test Statistics

The specification test considered in this paper examines infinitely many unconditional moment functions (3) that are equivalent to the conditional moment restriction (1) and therefore it is a consistent conditional moment test. To test the moment function $\mathbb{E}[m(y, \boldsymbol{x}, \boldsymbol{\beta}_o) \mathbb{1}_{\{\boldsymbol{x} \leq \boldsymbol{\xi}\}}]$ being equal to zero, it is natural to consider the normalized sample average of the moment function:

$$M_n(\boldsymbol{\xi}; \boldsymbol{\beta}_o) := \frac{1}{\sqrt{n}} \sum_{t=1}^{n} m\big(y_i, \boldsymbol{x}_i, \boldsymbol{\beta}_o\big) \mathbb{1}_{\{\boldsymbol{x}_i \leq \boldsymbol{\xi}\}},$$

with $\{y_i, \boldsymbol{x}_i\}_{i=1}^n$ a sequence of random variable, and $\mathbb{1}_{\{\boldsymbol{x}_i \leq \boldsymbol{\xi}\}} = 1_{\{x_{i1} \leq \xi_1\}} \cdots 1_{\{x_{ik} \leq \xi_k\}}$. The function $m\big(y_i, \boldsymbol{x}_i, \boldsymbol{\beta}\big) \mathbb{1}_{\{\boldsymbol{x}_i \leq \boldsymbol{\xi}\}}$ is considered the marked empirical process with marks given by the moment function m. The function M_n is the average of the market empirical processes with sample size n. Our objective is to test that $M_n(\boldsymbol{\xi}; \boldsymbol{\beta}_o)$ is close to zero or not. If $M_n(\boldsymbol{x}_i; \boldsymbol{\beta}_o)$ is close to zero, then we do not reject the null hypothesis; otherwise, we reject the null hypothesis and conclude that the conditional moment restriction does not hold.

Since the true parameter $\boldsymbol{\beta}_o$ is unknown, we replace $\boldsymbol{\beta}_o$ by its consistent estimator, $\hat{\boldsymbol{\beta}}_n$, and the sample average of the marked empirical processes becomes

$$M_n(\boldsymbol{\xi}; \hat{\boldsymbol{\beta}}_n) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} m(y_i, \boldsymbol{x}_i, \hat{\boldsymbol{\beta}}_n) \mathbb{1}_{\{\boldsymbol{x}_i \leq \boldsymbol{\xi}\}}.$$

Now, rewrite the process M_n:

$$M_n(\boldsymbol{\xi}; \hat{\boldsymbol{\beta}}_n) = M_n(\boldsymbol{\xi}; \boldsymbol{\beta}_o) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big(m(y_i, \boldsymbol{x}_i, \hat{\boldsymbol{\beta}}_n) - m(y_i, \boldsymbol{x}_i, \boldsymbol{\beta}_o)\big) \mathbb{1}_{\{\boldsymbol{x}_i \leq \boldsymbol{\xi}\}}.$$

If $m(y_i, \boldsymbol{x}_i, \boldsymbol{\beta})$ is once differentiable with first derivative $\nabla_{\boldsymbol{\beta}} m(y_i, \boldsymbol{x}_i, \boldsymbol{\beta}_o)$, then

$$\begin{aligned} M_n(\boldsymbol{\xi}; \hat{\boldsymbol{\beta}}_n) &= M_n(\boldsymbol{\xi}; \boldsymbol{\beta}_o) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla_{\boldsymbol{\beta}} m(y_i, \boldsymbol{x}_i, \boldsymbol{\beta}_o)(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_o) \mathbb{1}_{\{\boldsymbol{x}_i \leq \boldsymbol{\xi}\}} + o_p(1) \\ &= M_n(\boldsymbol{\xi}; \boldsymbol{\beta}_o) + \sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_o) \frac{1}{n} \sum_{i=1}^{n} \nabla_{\boldsymbol{\beta}} m(y_i, \boldsymbol{x}_i, \boldsymbol{\beta}_o) \mathbb{1}_{\{\boldsymbol{x}_i \leq \boldsymbol{\xi}\}} + o_p(1). \end{aligned}$$

It is seen that $M_n(\boldsymbol{\xi}; \hat{\boldsymbol{\beta}}_n)$ and $M_n(\boldsymbol{\xi}; \boldsymbol{\beta}_o)$ are not asymptotically equivalent due to the presence of the second term on the right hand side of the second equality. This term is the estimation effect discussed in Durbin (1973). The second term depends on a model characteristic that makes the test based on $M_n(\boldsymbol{\xi}; \hat{\boldsymbol{\beta}}_n)$ not asymptotically pivotal; this is the well know Durbin's problem. To eliminate the estimation effect, Stute, Thies and Zhu (1998), Koul and Stute (1999), and Stute and Zhu (2002) use the martingale transformation of Khmaladze (1981). However, due to the difficulty of having a proper definition of a multitime parameter martingales, these tests are used for univariate $\boldsymbol{x}$; see also Bai (2003). Song (2007) extends the research to multiparameter processes but because a nonparametric estimation of the conditional moment restriction is required, his test is complicated to compute.

This paper considers a subsampling version of the M_n process. Instead of using the whole sample of data with sample size n to compute the process, we use a subsample of data with sample size b to compute the sample average and construct the following process, for $b < n$:

$$M_b(\boldsymbol{\xi}; \hat{\boldsymbol{\beta}}_n) := \frac{1}{\sqrt{b}} \sum_{i=1}^{b} m\big(y_i, \boldsymbol{x}_i, \hat{\boldsymbol{\beta}}_n\big) \mathbb{1}_{\{\boldsymbol{x}_i \leq \boldsymbol{\xi}\}},$$

where $\hat{\boldsymbol{\beta}}_n$ can be any $\sqrt{n}$-estimator associated with the model of interest by the whole sample. M_b permits the following expansion:

$$\begin{aligned} M_b(\boldsymbol{\xi}; \hat{\boldsymbol{\beta}}_n) &= M_b(\boldsymbol{\xi}; \boldsymbol{\beta}_o) + \frac{1}{\sqrt{b}} \sum_{i=1}^{b} \nabla_{\boldsymbol{\beta}} m(y_i, \boldsymbol{x}_i, \boldsymbol{\beta}_o)(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_o) \mathbb{1}_{\{\boldsymbol{x}_i \leq \boldsymbol{\xi}\}} + o_p(1) \\ &= M_b(\boldsymbol{\xi}; \boldsymbol{\beta}_o) + \sqrt{\frac{b}{n}} \sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_o) \left[\frac{1}{b} \sum_{i=1}^{b} \nabla_{\boldsymbol{\beta}} m(y_i, \boldsymbol{x}_i, \boldsymbol{\beta}_o) \mathbb{1}_{\{\boldsymbol{x}_i \leq \boldsymbol{\xi}\}} \right] + o_p(1). \end{aligned}$$

It is interesting to see that if $b \to \infty$, $n \to \infty$ and $b/n \to 0$, and there are some regularity conditions, then the second term on the right-hand-side of the second equality of the above equality converges to zero. Thus, $M_b(\boldsymbol{\xi}; \hat{\boldsymbol{\beta}}_n)$ and $M_b(\boldsymbol{\xi}; \boldsymbol{\beta}_o)$ are asymptotically equivalent. Subsampling the marked empirical process annihilates the estimation effect. Kuan and Lin (2008) considers a partial sum of the data and eliminate the estimation effect by centering a sequentially marked empirical process. Let $D(\boldsymbol{R}^k)$ be the space of the cadlag function on $\boldsymbol{R}^k$ endowed with the Skorohod topology. Here, M_b is in $D(\boldsymbol{R}^k)$. In what follows, let

$\Rightarrow$ denote the convergence in distribution, and $\xrightarrow{p}$ denote the convergence in probability. The following assumptions are sufficient for the weak convergence of the subsampling marked empirical processes.

[A1] $\{y_i, \boldsymbol{x}_i\}_{i=1}^n$ is ergodic and strictly stationary where $\boldsymbol{x}_i$ has the continuous distribution function F and the density function is f.

[A2] (i) $\mathbb{E}[m(y_i, \boldsymbol{x}_i, \boldsymbol{\beta})^2|\boldsymbol{x}_i] < \infty$,
(ii) $\mathbb{E}m(y_i, \boldsymbol{x}_i, \boldsymbol{\beta})^4 = \kappa < \infty$,
(iii) $\mathbb{E}[m(y_i, \boldsymbol{x}_i, \boldsymbol{\beta})^4||\boldsymbol{x}_i||^{1+\eta}] < \infty$, for some $\eta > 0$.

[A3] The conditional density $f_{\boldsymbol{x}_i|\mathcal{F}_{i-1}}$ is bounded and continuous, where $\mathcal{F}_{i-1}$ is the σ algebra generated by $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_{i-1}$.

[A4] $m(\cdot)$ is once continuously differentiable in a neighborhood $\boldsymbol{\beta}_o$ and satisfies

$$\mathbb{E}\left[\sup_{\boldsymbol{\beta}\in\boldsymbol{B}_o} |\nabla_{\boldsymbol{\beta}} m(y_i, \boldsymbol{x}_i, \boldsymbol{\beta})|\right] < \infty,$$

where $\boldsymbol{B}_o$ denotes a neighborhood of $\boldsymbol{\beta}_o$.

[A5] $\hat{\boldsymbol{\beta}}_n$ is a $\sqrt{n}$-consistent estimator; that is $\sqrt{n}\big(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_o\big) = O_p(1)$.

Assumption [A1] permits data with weak dependence. Assumptions in [A2] restrict the dependence of the moment function. Given [A2] (i), the conditional variance function $\sigma^2(\boldsymbol{x}_i)$ of $m(y_i, \boldsymbol{x}_i, \boldsymbol{\beta})$ is defined with

$$\sigma^2(\boldsymbol{u}) := \mathrm{var}\big[m(y_i, \boldsymbol{x}_i, \boldsymbol{\beta})\big|\boldsymbol{x}_i = \boldsymbol{u}].$$

For $\boldsymbol{\xi} = (\xi_1, \cdots, \xi_k)'$ and $\boldsymbol{u} = (u_1, \cdots, u_k)'$, define

$$V(\boldsymbol{\xi}) := \mathbb{E}\left[\sigma^2(\boldsymbol{x}_i)\mathbb{1}_{\{\boldsymbol{x}_i\le\boldsymbol{\xi}\}}\right] = \int_{-\infty}^{\boldsymbol{\xi}} \sigma^2(\boldsymbol{u}) F_{\boldsymbol{x}}(d\boldsymbol{u}),$$

with $\int_{-\infty}^{\boldsymbol{\xi}} := \int_{-\infty}^{\xi_1} \cdots \int_{-\infty}^{\xi_k}$. Following Koul and Stute (1999), assumptions [A2](ii) and (iii) together with [A3] are required to obtain the uniform tightness in the space $D[-\infty, \infty]$. Assumption [A4] is a standard smoothness assumption. [A4] can be relaxed to non-smooth moment function when considering the stochastic equicontiunity of m. Assumption [A5] is weak and could be applied to most existing estimation method. In the following, we obtain the weak convergence of M_b.

Theorem 2.1. *Under H_0 and assumptions* [A1]-[A5]*, if $b \to \infty$, $n \to \infty$ and $b/n \to 0$, then one has:*

$$M_b(\boldsymbol{\xi}; \hat{\boldsymbol{\beta}}) \Rightarrow B\big(V(\boldsymbol{\xi})\big),$$

where $B(\cdot)$ denotes a Brownain sheet process.

The limiting distribution of M_b is a centered Gaussian process which is a multi-parameter Brownian motion process on $[0,1]^k$ with covariance function

$$V(\boldsymbol{\xi}_1 \wedge \boldsymbol{\xi}_2) = \int_{-\infty}^{\boldsymbol{\xi}_1 \wedge \boldsymbol{\xi}_2} \sigma^2(\boldsymbol{u}) F(d\boldsymbol{u}),$$

where $\int_{-\infty}^{\boldsymbol{\xi}_1 \wedge \boldsymbol{\xi}_2} = \int_{-\infty}^{\xi_{11} \wedge \xi_{21}} \cdots \int_{-\infty}^{\xi_{1k} \wedge \xi_{2k}}$. In particular, when $\boldsymbol{x}_i$ is a univariate, the process B is the standard Brownian motion. The limit of $M_b(\boldsymbol{\xi}; \boldsymbol{\beta}_o)$ and that of $M_b(\boldsymbol{\xi}; \hat{\boldsymbol{\beta}}_n)$ are the same and the Durbin problem disappears because of the convergence rate of b to infinity is slower that that of n. In addition, it is seen that $V(\boldsymbol{\xi})$ plays an important in the test. Since $V(\boldsymbol{\xi})$ still depends on the distribution of $\boldsymbol{x}_i$ and σ^2, the process $M_b(\boldsymbol{\xi}; \hat{\boldsymbol{\beta}}_n)$ is not asymptotically distribution free. When $\sigma^2(\boldsymbol{x}_i) = \sigma_0^2$ (the conditional homoskedasticity case), which is a constant, we obtain $V(\boldsymbol{\xi}) = \sigma_0^2 F(\boldsymbol{\xi})$. We follow Koul and Stute (1999) to consider an estimator of $V(\boldsymbol{\xi})$:

$$\hat{V}_n(\boldsymbol{\xi}) = \frac{1}{n} \sum_{i=1}^{n} m^2(y_i, \boldsymbol{x}_i, \boldsymbol{\beta}) \mathbb{1}_{\{\boldsymbol{x}_i \leq \boldsymbol{\xi}\}}, \ \boldsymbol{\xi} \in \boldsymbol{R}^k.$$

Tests for H_0 can be based on an appropriately scaling of M_b. Consider a consistent estimator $\hat{\sigma}_b^2$ for σ_0^2 in homoskedasticity case. We have the "scale invariant"version of subsampling marked empirical processes:

$$\tilde{M}_b(\boldsymbol{\xi}; \hat{\boldsymbol{\beta}}_n) := \frac{1}{\sqrt{b}} \hat{\sigma}_b^{-1} \sum_{i=1}^{b} m\big(y_i, \boldsymbol{x}_i, \hat{\boldsymbol{\beta}}_n\big) \mathbb{1}_{\{\boldsymbol{x}_i \leq \boldsymbol{\xi}\}}.$$

Theorem 2.2. *Under H_0 and assumptions* [A1]-[A5]*, if $b \to \infty$, $n \to \infty$, $b/n \to 0$ and $\hat{\sigma}_b^2 \to \sigma_0^2$, then one has*

$$\tilde{M}_b(\boldsymbol{\xi}; \hat{\boldsymbol{\beta}}_n) \Rightarrow B\big(F(\boldsymbol{\xi})\big),$$

with $B(\cdot)$ a standard Brownian sheet.

The computational counterpart of the scaled invariant version of $\tilde{M}_b(\boldsymbol{\xi}; \hat{\boldsymbol{\beta}})$ is considered as follows:

$$\tilde{M}_b(\boldsymbol{x}_j; \hat{\boldsymbol{\beta}}_n) := \frac{1}{\sqrt{b}} \hat{\sigma}_b^{-1} \sum_{i=1}^{b} m\big(y_i, \boldsymbol{x}_i, \hat{\boldsymbol{\beta}}_n\big) \mathbb{1}_{\{\boldsymbol{x}_i \leq \boldsymbol{x}_j\}}, j = 1, \cdots, n,$$

where each realization $\boldsymbol{x}_j$ is used as a $\boldsymbol{\xi}$ in the indicator function. Consider two goodness-of-fit statistics, the Kolmogorov-Smirnov and Cramer-von Mises test statistics:

$$KS_n = \sup_{\boldsymbol{x}_j \in \boldsymbol{R}^k} \big| \tilde{M}_b(\boldsymbol{x}_j; \hat{\boldsymbol{\beta}}_n) \big|,$$

$$CM_n = \frac{1}{n}\sum_{j=1}^{n} \tilde{M}_b(\boldsymbol{x}_j; \hat{\boldsymbol{\beta}}_n)^2.$$

By Theorem 2.2 and the continuous mapping theorem, for large n, one has, with $\boldsymbol{\omega} \in [0,1]^k$

$$KS_n \Rightarrow \sup_{\boldsymbol{\xi}\in \boldsymbol{R}^k} \big|B\big(F(\boldsymbol{\xi})\big)\big| = \sup_{0\le \boldsymbol{\omega}\le 1} \big|B(\boldsymbol{\omega})\big|,$$

and

$$CM_n = \int_{-\infty}^{\infty} \tilde{M}_b(\boldsymbol{\xi}; \hat{\boldsymbol{\beta}}_n)^2 F(d\boldsymbol{\xi}) \Rightarrow \int_{-\infty}^{\infty} B\big(F(\boldsymbol{\xi})\big)^2 F(d\boldsymbol{\xi}) = \int_0^1 B(\boldsymbol{\omega})^2 d\boldsymbol{\omega}.$$

The critical values of the test statistics KS_n and CM_n can be found in existing literature; see the book of Shorack and Wellner (1986). It is interesting to note that the proposed test is asymptotically pivotal and the limiting distribution of the proposed test does not depend on a data generating process. Therefore, we have the following corollary.

Corollary 2.3. *Under all the assumptions in Theorem* 2.2.

$$KS_n \Rightarrow \sup_{0\le \boldsymbol{\omega}\le 1} \big|B(\boldsymbol{\omega})\big|,$$

$$CM_n \Rightarrow \int_0^1 B(\boldsymbol{\omega})^2 d\boldsymbol{\omega},$$

with $B(\cdot)$ *the standard Brownian sheet.*

3 Power of the Tests

To investigate the power performance of the proposed test, two types of alternatives are considered. One is the general type of alternatives:

$$H_1 : \mathbb{E}[m(y, \boldsymbol{x}, \boldsymbol{\beta}_o)|\boldsymbol{x}] = \boldsymbol{\mu}(\boldsymbol{x}) \neq \boldsymbol{0},\ \forall\, \boldsymbol{\xi} = (\xi_1, \cdots, \xi_k) \in \boldsymbol{R}^k,$$

and the other is the local alternatives:

$$H_1^L : \mathbb{E}\big[m(y, \boldsymbol{x}, \boldsymbol{\beta}_o)\big|\boldsymbol{x}\big] = \frac{\boldsymbol{\delta}(\boldsymbol{x})}{\sqrt{b}},$$

with $\boldsymbol{\delta}(\boldsymbol{x}) \neq 0$. We then have the following theorem.

Theorem 3.1. *Assume assumptions* [A1]-[A5] *hold. Assume also* $b \to \infty$, $n \to \infty$ *and* $b/n \to 0$. *Therefore:*

(i) *Under the fixed alternative* H_1*:*

$$\tilde{M}_b(\boldsymbol{\xi}; \hat{\boldsymbol{\beta}}_n) \to \infty.$$

(ii) *Under the local alternatives* H_1^L*:*

$$\tilde{M}_b(\boldsymbol{\xi};\hat{\boldsymbol{\beta}}_n) \Rightarrow B(F(\boldsymbol{\xi})) + \sigma_0^{-1}\mathbb{E}[\delta(\boldsymbol{x}_i)\mathbb{1}_{\{\boldsymbol{x}_i\leq\boldsymbol{\xi}\}}].$$

By Theorem 3.1 and the continuous mapping theorem, we additionally have the following corollary.

Corollary 3.2. *Assume assumptions* [A1]-[A5] *hold. Assume also* $b\to\infty$*,* $n\to\infty$ *and* $b/n\to 0$*. Therefore:*

(i) *Under the fixed alternative* H_1*:*

$$KS_n \to \infty,$$

$$CM_n \to \infty.$$

(ii) *Under the local alternatives* H_1^L*:*

$$KS_n \Rightarrow \sup_{\boldsymbol{\xi}\in\boldsymbol{R}^k} \left|B(F(\boldsymbol{\xi})) + \sigma_0^{-1}\mathbb{E}[\delta(\boldsymbol{x}_i)\mathbb{1}_{\{\boldsymbol{x}_i\leq\boldsymbol{\xi}\}}]\right|$$

$$CM_n \Rightarrow \int_0^1 \left(B(F(\boldsymbol{\xi})) + \sigma_0^{-1}\mathbb{E}[\delta(\boldsymbol{x}_i)\mathbb{1}_{\{\boldsymbol{x}_i\leq\boldsymbol{\xi}\}}]\right)^2 d\boldsymbol{\omega}.$$

The first part of Corollary 3.2 gives the consistency of the proposed test and the second part implies that the proposed test has nontrivial powers against local alternatives at rate $b^{-1/2}$. However, there may exist local alternatives at rate $n^{-1/2}$ as follows.

$$H_2^L : \mathbb{E}\left[m(y,\boldsymbol{x},\boldsymbol{\beta}_o)\middle|\boldsymbol{x}\right] = \frac{\boldsymbol{\delta}(\boldsymbol{x})}{\sqrt{n}}.$$

We have Theorem 3.3.

Theorem 3.3. *Assume assumptions* [A1]-[A5] *hold. Assume also* $b\to\infty$*,* $n\to\infty$ *and* $b/n\to 0$*. Under the local alternatives* H_2^L*:*

$$\tilde{M}_b(\boldsymbol{\xi};\hat{\boldsymbol{\beta}}_n) \Rightarrow B(F(\boldsymbol{\xi})).$$

Under the local alternatives H_2^L, the limiting distribution of $\tilde{M}_b(\boldsymbol{\xi};\hat{\boldsymbol{\beta}}_n)$ is the same as that under the null hypothesis; see Theorem 2.1. The proposed test is incapable of detecting local alternatives at rate $n^{-1/2}$. Therefore, our test shares the same disadvantage of most nonparametric tests; see the discussion in Whang (2000) therein. The proposed test has no local power against this type of alternatives. Thus our test cannot be used in this scenario.

4 Monte Carlo Simulations

This section reports some simulation results to examine the finite sample performance of the test statistic KS_n. We consider the following null data generating processes (DGPs).

(A) $y_i = x_{i1} + e_i$,

(B) $y_i = x_{i1} + 5 + e_i$,

(C) $y_i = x_{i1} + \exp(z_i) + e_i$,

(D) $y_i = x_{i2} + x_{i3} + e_i$,

(E) $y_i = x_{i2} + x_{i3} + 5 + e_i$,

(F) $y_i = x_{i2} + x_{i3} + \exp(z_i) + e_i$.

Here x_{i1}, x_{i2}, x_{i3} and z_i are independent and identically distributed (i.i.d.) $N(0,1)$ distribution and e_i is i.i.d. $N(0, \sigma_0^2)$ with $\sigma_0^2 = 1, 2, 3, 4$. The test statistic KS_n for one regressor is:

$$KS_1 = \max_j \left| \frac{1}{\sqrt{b}} \hat{\sigma}_1^{-1} \sum_{i=1}^{b} (y_i - x_{i1}'\hat{\beta}_1) \mathbb{1}_{\{x_{i1} \le x_j\}} \right|,$$

for DGPs (A), (B) and (C) and a statistic for two regressors:

$$KS_2 = \max_j \left| \frac{1}{\sqrt{b}} \hat{\sigma}_2^{-1} \sum_{i=1}^{b} (y_i - x_{i2}'\hat{\beta}_2 - x_{i3}'\hat{\beta}_3) \mathbb{1}_{\{x_{i2} \le x_{j1}\}} \mathbb{1}_{\{x_{i3} \le x_{j2}\}} \right|,$$

for DGPs (D), (E) and (F) where $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$ are least square estimates, $\hat{\sigma}_1^2 = b^{-1} \sum_{i=1}^{b} (y_i - x_{i1}'\hat{\beta}_1)^2$ and $\hat{\sigma}_2^2 = b^{-1} \sum_{i=1}^{b} (y_i - x_{i2}'\hat{\beta}_2 - x_{i3}'\hat{\beta}_3)^2$. In each simulation experiment, the number of replications is 2000. The significance level is 0.05. We choose different values of b in this simulation. The choice of b is considered for the formula $b = n^p$ with $p = 0.5, 0.55, \cdots, 0.95$.

Table 1, given $\sigma_0^2 = 1$, reports the rejection frequencies of the tests for different values of n and p. For DGPs (A) and (D), the rejection values are finite sample sizes of the test. In the column of DGP (A), all values are close to the significance level 0.05 except at the values of $p = 0.5$. However, in the column of DGP (D), the proposed test is under-sized for large p. We can see that when the number of regressors of the regression increases, or if b increases, then the sizes of the test are lower. For DGPs (B), (C), (E) and (F), the rejection rates are the finite sample powers of the proposed test. In columns of DGPs (B) and (E) that have fixed alternatives, the finite sample powers are 1 showing that the test has good power performances in different values of n and b(or$p$). In addition, the values on columns of

DGPs (C) and (F) present that the test performs well when the alternatives are a random variable. For DGP (C), there are good power performances of the test for large values of p. When n increases, the powers of the test are closer to 1. For DGP (F), powers are lower for $n = 100$, and as n and $b(\text{or}p)$ increase, the powers increase. To sum up, the test has correct sizes for one regressor and is slightly under-sized for two regressors in the regression. When there are fixed alternatives, the power performances are very good. The powers of the test increase along with both n and b. Table 2 reports the rejection frequencies of the test for six DGPs with different σ_0^2 and p. The sample size is 500. The finite sample performances in Table 2 are similar to those in Table 1. Moreover, we find that when the variety of error term increases, the powers of the test decrease.

5 Conclusions

This paper proposes a consistent conditional moment test based on infinitely many unconditional moment conditions. The test statistic is a subsampling marked empirical processes and the Durbin problem is eliminated as the convergence rate of the subsampling size is slower than that of the whole sample size. We thus obtain an asymptotically pivotal test. The proposed test is consistent against a general type of alternatives and is powerful against local alternatives at rates $b^{-1/2}$. However, our test is not powerful against $n^{-1/2}$ local alternatives. In addition, the test performs well in finite sample simulations and the power performances are good with most values of b.

Table 1: Rejection frequencies of the conditional moment tests

		KS_1			KS_2		
n	p	(A)	(B)	(C)	(D)	(E)	(F)
100	0.50	0.076	1.000	0.742	0.057	1.000	0.596
	0.55	0.067	1.000	0.817	0.043	1.000	0.661
	0.60	0.057	1.000	0.877	0.037	1.000	0.775
	0.65	0.048	1.000	0.947	0.038	1.000	0.847
	0.70	0.044	1.000	0.977	0.036	1.000	0.919
	0.75	0.047	1.000	0.992	0.031	1.000	0.968
	0.80	0.049	1.000	0.999	0.024	1.000	0.985
	0.85	0.042	1.000	1.000	0.021	1.000	0.995
	0.90	0.046	1.000	0.999	0.025	1.000	0.999
	0.95	0.041	1.000	1.000	0.018	1.000	0.999
200	0.50	0.063	1.000	0.863	0.046	1.000	0.783
	0.55	0.045	1.000	0.937	0.038	1.000	0.867
	0.60	0.051	1.000	0.978	0.040	1.000	0.950
	0.65	0.053	1.000	0.992	0.031	1.000	0.981
	0.70	0.039	1.000	0.997	0.035	1.000	0.990
	0.75	0.054	1.000	1.000	0.023	1.000	0.998
	0.80	0.048	1.000	1.000	0.028	1.000	1.000
	0.85	0.043	1.000	1.000	0.023	1.000	1.000
	0.90	0.042	1.000	1.000	0.025	1.000	1.000
	0.95	0.049	1.000	1.000	0.022	1.000	1.000
500	0.50	0.068	1.000	0.964	0.057	1.000	0.950
	0.55	0.042	1.000	0.989	0.030	1.000	0.982
	0.60	0.047	1.000	0.998	0.035	1.000	0.994
	0.65	0.044	1.000	0.999	0.036	1.000	0.998
	0.70	0.045	1.000	1.000	0.040	1.000	0.999
	0.75	0.040	1.000	1.000	0.030	1.000	1.000
	0.80	0.055	1.000	1.000	0.028	1.000	1.000
	0.85	0.040	1.000	1.000	0.029	1.000	1.000
	0.90	0.036	1.000	1.000	0.019	1.000	1.000
	0.95	0.035	1.000	1.000	0.028	1.000	1.000

Note: The significant level is 0.05. $b = n^p$. The values in the 3rd and 6th columns are the finite sample sizes and the values in the 4th, 5th, 7th and 8th columns are the finite sample powers of the proposed test.

Table 2: Rejection frequencies of the conditional moment tests

		KS_1			KS_2		
σ_0^2	p	(A)	(B)	(C)	(D)	(E)	(F)
2	0.50	0.058	1.000	0.906	0.037	1.000	0.862
	0.55	0.053	1.000	0.972	0.041	1.000	0.955
	0.60	0.043	1.000	0.995	0.036	1.000	0.984
	0.65	0.044	1.000	0.999	0.037	1.000	0.998
	0.70	0.049	1.000	1.000	0.039	1.000	0.999
	0.75	0.044	1.000	1.000	0.029	1.000	1.000
	0.80	0.052	1.000	1.000	0.030	1.000	1.000
	0.85	0.040	1.000	1.000	0.033	1.000	1.000
	0.90	0.040	1.000	1.000	0.022	1.000	1.000
	0.95	0.036	1.000	1.000	0.023	1.000	1.000
3	0.50	0.060	1.000	0.843	0.050	1.000	0.797
	0.55	0.050	1.000	0.942	0.039	1.000	0.915
	0.60	0.054	1.000	0.987	0.037	1.000	0.972
	0.65	0.051	1.000	0.999	0.039	1.000	0.994
	0.70	0.048	1.000	1.000	0.034	1.000	0.999
	0.75	0.037	1.000	1.000	0.033	1.000	1.000
	0.80	0.046	1.000	1.000	0.038	1.000	1.000
	0.85	0.052	1.000	1.000	0.030	1.000	1.000
	0.90	0.040	1.000	1.000	0.023	1.000	1.000
	0.95	0.041	1.000	1.000	0.027	1.000	1.000
4	0.50	0.048	1.000	0.776	0.042	1.000	0.703
	0.55	0.053	1.000	0.897	0.042	1.000	0.845
	0.60	0.044	1.000	0.969	0.037	1.000	0.949
	0.65	0.042	1.000	0.992	0.031	1.000	0.988
	0.70	0.048	1.000	0.999	0.033	1.000	0.999
	0.75	0.046	1.000	1.000	0.035	1.000	1.000
	0.80	0.053	1.000	1.000	0.038	1.000	1.000
	0.85	0.047	1.000	1.000	0.027	1.000	1.000
	0.90	0.043	1.000	1.000	0.024	1.000	1.000
	0.95	0.042	1.000	1.000	0.029	1.000	1.000

Note: The significant level is 0.05. $b = n^p$. The values in the 3rd and 6th columns are the finite sample sizes and the values in the 4th, 5th, 7th and 8th columns are the finite sample powers of the proposed test.

Appendix

Proof of Theorem 2.1. By assumption [A4], the subsampling marked empirical process M_b permits the Taylor expansion:

$$\frac{1}{\sqrt{b}}\sum_{i=1}^{b} m\big(y_i,\boldsymbol{x}_i,\hat{\boldsymbol{\beta}}_n\big)\,\mathbb{1}_{\{\boldsymbol{x}_i\le\boldsymbol{\xi}\}}$$
$$=\frac{1}{\sqrt{b}}\sum_{i=1}^{b} m\big(y_i,\boldsymbol{x}_i,\boldsymbol{\beta}_o\big)\,\mathbb{1}_{\{\boldsymbol{x}_i\le\boldsymbol{\xi}\}}+\frac{1}{\sqrt{b}}\sum_{i=1}^{b}\nabla_{\boldsymbol{\beta}} m\big(y_i,\boldsymbol{x}_i,\boldsymbol{\beta}_o\big)(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_o)\,\mathbb{1}_{\{\boldsymbol{x}_i\le\boldsymbol{\xi}\}}+o_p(1).$$

Because $b/n\to 0$ and assumption [A5],

$$\sqrt{b}(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_o)=\sqrt{\frac{b}{n}}\sqrt{n}(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_o)\xrightarrow{p}0.$$

In addition, by assumptions [A1] and [A4], Hölder's inequality and ergodic theorem, we have the following law of large numbers of ergodic and stationary sequence:

$$\frac{1}{b}\sum_{i=1}^{b}\nabla_{\boldsymbol{\beta}} m\big(y_i,\boldsymbol{x}_i,\boldsymbol{\beta}_o\big)\,\mathbb{1}_{\{\boldsymbol{x}_i\le\boldsymbol{\xi}\}}\xrightarrow{p}\mathbb{E}\left[\nabla_{\boldsymbol{\beta}} m\big(y_i,\boldsymbol{x}_i,\boldsymbol{\beta}_o\big)\,\mathbb{1}_{\{\boldsymbol{x}_i\le\boldsymbol{\xi}\}}\right],$$

we then obtain

$$\frac{1}{\sqrt{b}}\sum_{i=1}^{b}\nabla_{\boldsymbol{\beta}} m\big(y_i,\boldsymbol{x}_i,\boldsymbol{\beta}_o\big)(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_o)\,\mathbb{1}_{\{\boldsymbol{x}_i\le\boldsymbol{\xi}\}}=\left[\frac{1}{b}\sum_{i=1}^{b}\nabla_{\boldsymbol{\beta}} m\big(y_i,\boldsymbol{x}_i,\boldsymbol{\beta}_o\big)\,\mathbb{1}_{\{\boldsymbol{x}_i\le\boldsymbol{\xi}\}}\right]\sqrt{b}(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_o)\xrightarrow{p}0.$$

Therefore,

$$\frac{1}{\sqrt{b}}\sum_{i=1}^{b} m\big(y_i,\boldsymbol{x}_i,\hat{\boldsymbol{\beta}}\big)\,\mathbb{1}_{\{\boldsymbol{x}_i\le\boldsymbol{\xi}\}}=\frac{1}{\sqrt{b}}\sum_{i=1}^{b} m\big(y_i,\boldsymbol{x}_i,\boldsymbol{\beta}_o\big)\,\mathbb{1}_{\{\boldsymbol{x}_i\le\boldsymbol{\xi}\}}+o_p(1).$$

$M_b(\boldsymbol{\xi};\hat{\boldsymbol{\beta}}_n)$ and $M_b(\boldsymbol{\xi};\boldsymbol{\beta}_o)$ are asymptotically equivalent. Estimating parameter $\boldsymbol{\beta}$ does not affect the limiting distribution of the statistic and the Durbin problem does not appear.

The process M_b belongs to the Shorohod space $D(\boldsymbol{R}^k)$ and the weak convergence of $M_b(\boldsymbol{\xi};\boldsymbol{\beta}_o)$ in the space $D(\boldsymbol{R}^k)$ to a continuous limit is implies by the tightness of M_b and the finite dimensional convergence of $M_b(\boldsymbol{\xi};\boldsymbol{\beta}_o)$. In the following, we first follow Bickel and Wichura (1971), Koul and Stute (1999) and Domínguez and Lobato (2004) to show the tightness of M_b and then the weak convergence of $M_b(\boldsymbol{\xi};\boldsymbol{\beta}_o)$. Define $I_1=(s^1,t^1]=\times_{j=1}^{k}(s_j^1,t_j^1]$, and $I_2=(s^2,t^2]=\times_{j=1}^{k}(s_j^2,t_j^2]$ be two subsets in $\boldsymbol{R}^k$. Then I_1 and I_2 are neighbor subsets if and only if for some $j^*\in\{1,2,\cdots,k\}$, $(s_{j^*}^1,t_{j^*}^1]\neq(s_{j^*}^2,t_{j^*}^2]$, $\times_{j\neq j^*}^{k}(s_j^1,t_j^1]=\times_{j\neq j^*}^{k}(s_j^2,t_j^2]$ and $t_{j^*}^1=s_{j^*}^2$; that is they are next to each other and share the j^*th face. Then the process M_b indexed by a parameter in $\boldsymbol{R}^k$ has an associated process indexed by the intervals as follows: for $h = 1, 2$,

$$
\begin{aligned}
M_b(I_h; \boldsymbol{\beta}) &:= \frac{1}{\sqrt{b}} \sum_{i=1}^{b} m(y_i, \boldsymbol{x}_i; \boldsymbol{\beta}) \mathbb{1}_{\{\boldsymbol{x}_i \in I_h\}} \\
&= \sum_{e_1=0}^{1} \cdots \sum_{e_k=0}^{1} (-1)^{k - \sum_{j=1,\cdots,k} e_j} M_b(s_1^h + e_1(t_1^h - s_1^h), \cdots, s_k^h + e_k(t_k^h - s_k^h); \boldsymbol{\beta}),
\end{aligned}
$$

which is the increment of M_b around I_h. Let $m(y_i, \boldsymbol{x}_i; \boldsymbol{\beta}) = m_i$. Following Bickel and Wichura (1971, Theorem 3 and example II), if

$$
\mathbb{E}\left(M_b(I_1; \boldsymbol{\beta})^2, M_b(I_2; \boldsymbol{\beta})^2\right) = \frac{1}{b^2} \mathbb{E}\left(\left[\sum_{i=1}^{b} m_i \mathbb{1}_{\{\boldsymbol{x}_i \in I_1\}} \right]^2 \left[\sum_{i=1}^{b} m_i \mathbb{1}_{\{\boldsymbol{x}_i \in I_2\}} \right]^2 \right).
$$

is bounded, then for any $\lambda > 0$ and $\gamma > 1$,

$$
\mathbb{P}(M_b \geq \lambda) \leq \lambda^{-4} \mu(I_1 \cup I_2)^{\gamma},
$$

with some measure μ. The above result asserts that the process M_b is tight.

Let $\mathcal{F}_i$ denote the natural filtration. Under H_0 and assumption [A1], $\{m_i \mathbb{1}_{\{\boldsymbol{x}_i \leq \boldsymbol{\tau}\}}, \mathcal{F}_{i-1}\}$ is a strictly stationary and ergodic martingale difference sequence. When a subindex appears once in the summation, the corresponding term is zero by the law of iterated expectation and the martingale difference property. Moreover, since I_1 and I_2 are disjoint sets, when a subindex appears more than twice, the corresponding term is zero. Therefore, similar to Koul and Stute (1999) and Domínguez and Lobato (2004),

$$
\begin{aligned}
&\mathbb{E}\left(M_b(I_1; \boldsymbol{\beta})^2, M_b(I_2; \boldsymbol{\beta})^2\right) \\
&= \frac{1}{b^2} \mathbb{E}\left[\sum_{i=1}^{b} m_i^2 \mathbb{1}_{\{\boldsymbol{x}_i \in I_1\}} \left(\sum_{j=1}^{i-1} m_j \mathbb{1}_{\{\boldsymbol{x}_j \in I_2\}} \right)^2 \right] + \frac{1}{b^2} \mathbb{E}\left[\sum_{i=1}^{b} m_i^2 \mathbb{1}_{\{\boldsymbol{x}_i \in I_2\}} \left(\sum_{j=1}^{i-1} m_j \mathbb{1}_{\{\boldsymbol{x}_j \in I_1\}} \right)^2 \right].
\end{aligned}
$$

The first and the second terms in the above equation are similar and the only difference is the indexing set I_h; we then focus on the first term. By assumption [A2](i),

$$
\begin{aligned}
&\frac{1}{b^2} \sum_{i=1}^{b} \mathbb{E}\left[m_i^2 \mathbb{1}_{\{\boldsymbol{x}_i \in I_1\}} \left(\sum_{j=1}^{i-1} m_j \mathbb{1}_{\{\boldsymbol{x}_j \in I_2\}} \right)^2 \right] \\
&= \frac{1}{b^2} \sum_{i=1}^{b} \mathbb{E}\left[\sigma^2(\boldsymbol{x}_i, \mathcal{F}_{i-1}) \mathbb{1}_{\{\boldsymbol{x}_i \in I_1\}} \left(\sum_{j=1}^{i-1} m_j \mathbb{1}_{\{\boldsymbol{x}_j \in I_2\}} \right)^2 \right] \\
&= \frac{1}{b^2} \sum_{i=1}^{b} \mathbb{E}\left[\int_{I_1} \sigma^2(\boldsymbol{u}, \mathcal{F}_{i-1}) f_{\boldsymbol{x}_i | \mathcal{F}_{i-1}}(\boldsymbol{u}) d\boldsymbol{u} \left(\sum_{j=1}^{i-1} m_j \mathbb{1}_{\{\boldsymbol{x}_j \in I_2\}} \right)^2 \right].
\end{aligned}
$$

By Fubini's Theorem, the above equation equals

$$\frac{1}{b^2}\sum_{i=1}^{b}\int_{I_1}\mathbb{E}\left[\sigma^2(\boldsymbol{u},\mathcal{F}_{i-1})f_{\boldsymbol{x}_i|\mathcal{F}_{i-1}}(\boldsymbol{u})\left(\sum_{j=1}^{i-1}m_j\,\mathbb{1}_{\{\boldsymbol{x}_j\in I_2\}}\right)^2\right]d\boldsymbol{u}.$$

Using Cauchy-Schwarz's inequality, we have

$$\begin{aligned}&\frac{1}{b^2}\sum_{i=1}^{b}\int_{I_1}\mathbb{E}\left[\sigma^2(\boldsymbol{u},\mathcal{F}_{i-1})f_{\boldsymbol{x}_i|\mathcal{F}_{i-1}}(\boldsymbol{u})\left(\sum_{j=1}^{i-1}m_j\,\mathbb{1}_{\{\boldsymbol{x}_j\in I_2\}}\right)^2\right]d\boldsymbol{u}\\&\le\frac{1}{b^2}\sum_{i=1}^{b}\int_{I_1}\left[\left\{\mathbb{E}\left[\sigma^2(\boldsymbol{u},\mathcal{F}_{i-1})f_{\boldsymbol{x}_i|\mathcal{F}_{i-1}}(\boldsymbol{u})\right]^2\right\}^{1/2}\left\{\mathbb{E}\left(\sum_{j=1}^{i-1}m_j\,\mathbb{1}_{\{\boldsymbol{x}_j\in I_2\}}\right)^4\right\}^{1/2}\right]d\boldsymbol{u}.\end{aligned}$$

By Burkholder's inequality and the moment inequality yield, with some constant C,

$$\mathbb{E}\left(\sum_{j=1}^{i-1}m_j\,\mathbb{1}_{\{\boldsymbol{x}_i\in I_2\}}\right)^4\le C\,\mathbb{E}\left(\sum_{j=1}^{i-1}m_j^2\,\mathbb{1}^2_{\{\boldsymbol{x}_i\in I_2\}}\right)^2\le C(i-1)^2\,\mathbb{E}(m_1^4\,\mathbb{1}_{\{\boldsymbol{x}_1\in I_2\}}).$$

It follows that

$$\begin{aligned}&\frac{1}{b^2}\sum_{i=1}^{b}\mathbb{E}\left[m_i^2\,\mathbb{1}_{\{\boldsymbol{x}_i\in I_1\}}\left(\sum_{j=1}^{i-1}m_j\,\mathbb{1}_{\{\boldsymbol{x}_j\in I_2\}}\right)^2\right]\\&\le\frac{1}{b^2}\sum_{i=1}^{b}\int_{I_1}\left[\left\{\mathbb{E}\left[\sigma^2(\boldsymbol{u},\mathcal{F}_{i-1})f_{\boldsymbol{x}_i|\mathcal{F}_{i-1}}(\boldsymbol{u})\right]^2\right\}^{1/2}\left\{C(i-1)^2\,\mathbb{E}(m_1^4\,\mathbb{1}_{\{\boldsymbol{x}_1\in I_2\}})\right\}^{1/2}\right]d\boldsymbol{u}\\&=\frac{1}{b^2}\left[C\,\mathbb{E}(m_1^4\,\mathbb{1}_{\{\boldsymbol{x}_1\in I_2\}})\right]^{1/2}\sum_{i=1}^{b}(i-1)\int_{I_1}\left\{\mathbb{E}\left[\sigma^2(\boldsymbol{u},\mathcal{F}_{i-1})f_{\boldsymbol{x}_i|\mathcal{F}_{i-1}}(\boldsymbol{u})\right]^2\right\}^{1/2}d\boldsymbol{u}.\end{aligned}$$

$\mathbb{E}(m_1^4\,\mathbb{1}_{\{\boldsymbol{x}_1\in I_2\}})\le\mathbb{E}(m_1^4)$ which is bounded by assumption [A2] (ii). In addition, from Koul and Stute (1999), $\int_{I_1}\{\mathbb{E}[\sigma^2(\boldsymbol{u})f_{\boldsymbol{x}_i|\mathcal{F}_{i-1}}(\boldsymbol{u})]^2\}^{1/2}d\boldsymbol{u}$ is bounded by assumptions [A2](iii) and [A3]; see more detail discuss therein. Therefore, under H_0 and assumptions [A1]–[A3], the process M_b is tight. Note that our assumption [A2] (ii) and (iii) are similar to the assumption (A)(a) in Koul and Stute (1999). In Domínguez and Lobato (2004), they use stricter conditions (see, [A7] and [A8]) and Hölder's inequality to obtain the boundness of $\mathbb{E}(m_1^4\,\mathbb{1}_{\{\boldsymbol{x}_1\in I_1\}})$.

Under assumptions [A1] and [A2] (i), and by a central limit theorem for the ergodic stationary martingale difference sequence, we have for any $\boldsymbol{\tau}\in\boldsymbol{R}^k$,

$$M_b(\boldsymbol{\xi};\boldsymbol{\beta}_o)\Rightarrow N(0,V(\boldsymbol{\xi})).$$

For $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2 \in \boldsymbol{R}^k$,

$$\begin{aligned}
&\operatorname{Cov}\big(M_b(\boldsymbol{\xi}_1;\boldsymbol{\beta}_o), M_b(\boldsymbol{\xi}_2;\boldsymbol{\beta}_o)\big)\\
&\quad= \frac{1}{b}\sum_{i=1}^{b} \mathbb{E}\big[m(y_i,\boldsymbol{x}_i;\boldsymbol{\beta}_o)^2\,\mathbb{1}_{\{\boldsymbol{x}_i\le\boldsymbol{\xi}_1\}}\,\mathbb{1}_{\{\boldsymbol{x}_i\le\boldsymbol{\xi}_2\}}\big]\\
&\quad\xrightarrow{p} \int_{-\infty}^{\boldsymbol{\xi}_1\wedge\boldsymbol{\xi}_2} \sigma^2(\boldsymbol{u})\,F(\mathrm{d}\boldsymbol{u})\\
&\quad= V(\boldsymbol{\xi}_1\wedge\boldsymbol{\xi}_2),
\end{aligned}$$

where the first equality holds by the property of martingale difference sequence. Since $V(\boldsymbol{\xi})$ is nondecreasing and nonnegative, M_b admits a asymptotically distributed as $B(V(\boldsymbol{\xi}))$, where $B(\cdot)$ is a standard Brownian sheet. □

Proof of Theorem 2.2. In this proof, we show that using a consistent estimator $\hat{\sigma}_b^2$ to replace σ_0^2 does not affect the asymptotics of the scale invariant subsampling marked empirical process. Rewrite the process $\tilde{M}_b(\boldsymbol{\xi};\hat{\boldsymbol{\beta}}_n)$:

$$\begin{aligned}
&\frac{1}{\sqrt{b}}\hat{\sigma}_b^{-1}\sum_{i=1}^{b} m\big(y_i,\boldsymbol{x}_i,\hat{\boldsymbol{\beta}}_n\big)\,\mathbb{1}_{\{\boldsymbol{x}_i\le\boldsymbol{\xi}\}}\\
&= \big(\hat{\sigma}_b^{-1}-\sigma_0^{-1}\big)\frac{1}{\sqrt{b}}\sum_{i=1}^{b} m\big(y_i,\boldsymbol{x}_i,\hat{\boldsymbol{\beta}}_n\big)\,\mathbb{1}_{\{\boldsymbol{x}_i\le\boldsymbol{\xi}\}} + \frac{1}{\sqrt{b}}\sigma_0^{-1}\sum_{i=1}^{b} m\big(y_i,\boldsymbol{x}_i,\hat{\boldsymbol{\beta}}_n\big)\,\mathbb{1}_{\{\boldsymbol{x}_i\le\boldsymbol{\xi}\}}.
\end{aligned}$$

Since $\hat{\sigma}_b^{-1}-\sigma_0^{-1}=o_p(1)$, and by Theorem 2.1,

$$b^{-1/2}\sum_{i=1}^{b} m\big(y_i,\boldsymbol{x}_i,\hat{\boldsymbol{\beta}}_n\big)\,\mathbb{1}_{\{\boldsymbol{x}_i\le\boldsymbol{\xi}\}} = O_p(1),$$

then

$$\tilde{M}_b(\boldsymbol{\xi};\hat{\boldsymbol{\beta}}_n) = \frac{1}{\sqrt{b}}\sigma_0^{-1}\sum_{i=1}^{b} m\big(y_i,\boldsymbol{x}_i,\hat{\boldsymbol{\beta}}_n\big)\,\mathbb{1}_{\{\boldsymbol{x}_i\le\boldsymbol{\xi}\}} + o_p(1).$$

Denote

$$\tilde{M}_b^o(\boldsymbol{\xi},\boldsymbol{\beta}) := \frac{1}{\sqrt{b}}\sigma_0^{-1}\sum_{i=1}^{b} m\big(y_i,\boldsymbol{x}_i,\boldsymbol{\beta}\big)\,\mathbb{1}_{\{\boldsymbol{x}_i\le\boldsymbol{\xi}\}}.$$

The processes $\tilde{M}_b(\boldsymbol{\xi};\hat{\boldsymbol{\beta}}_n)$ and $\tilde{M}_b^o(\boldsymbol{\xi},\hat{\boldsymbol{\beta}}_n)$ have the same limiting distribution. In addition, similar to the proof of Theorem 2.1, replacing $\boldsymbol{\beta}_o$ by $\hat{\boldsymbol{\beta}}_n$ in $\tilde{M}_b^o$ does not affected the asymptotics of $\tilde{M}_b^o$. It follows that

$$\tilde{M}_b(\boldsymbol{\xi};\hat{\boldsymbol{\beta}}_n) \approx \tilde{M}_b^o(\boldsymbol{\xi};\hat{\boldsymbol{\beta}}_n) \approx \tilde{M}_b^o(\boldsymbol{\xi};\boldsymbol{\beta}_o), \tag{4}$$

and it suffices to focus on the limiting behavior of $\tilde{M}_b^o(\boldsymbol{\xi};\boldsymbol{\beta}_o)$. The tightness of M_b can be obtained in Theorem 2.1 as σ_o^2 is continuous. Since $\{m(y_i,\boldsymbol{x}_i,\boldsymbol{\beta})\,\mathbb{1}_{\{\boldsymbol{x}_i\le\boldsymbol{\xi}\}},\mathcal{F}_{i-1}\}$ is a martingale difference sequence, we then use the central limit theorem for ergodic and stationary martingale difference sequence to obtain the limiting distribution, which is a Gaussian process with zero mean and for $\boldsymbol{\xi}_1,\boldsymbol{\xi}_2\in\boldsymbol{R}^k$,

$$
\begin{aligned}
&\mathrm{Cov}\Big(\tilde{M}_b^o(\boldsymbol{\xi}_1,\boldsymbol{\beta}_o),\tilde{M}_b^o(\boldsymbol{\xi}_2,\boldsymbol{\beta}_o)\Big)\\
&\quad=\frac{1}{b}\sigma_0^{-2}\sum_{i=1}^{b}\mathbb{E}\big[m(y_i,\boldsymbol{x}_i;\boldsymbol{\beta}_o)^2\,\mathbb{1}_{\{\boldsymbol{x}_i\le\boldsymbol{\xi}_1\}}\,\mathbb{1}_{\{\boldsymbol{x}_i\le\boldsymbol{\xi}_2\}}\big]\\
&\quad\xrightarrow{p}\int_{-\infty}^{\boldsymbol{\xi}_1\wedge\boldsymbol{\xi}_2}F(\mathrm{d}\boldsymbol{u})\\
&\quad=F(\boldsymbol{\xi}_1\wedge\boldsymbol{\xi}_2).
\end{aligned}
$$

Hence, $\tilde{M}_b(\boldsymbol{\xi};\hat{\boldsymbol{\beta}}_n)\Rightarrow B(F(\boldsymbol{\xi}))$, with B a Brownian sheet. □

Proof of Theorem 3.1. $\tilde{M}_b(\boldsymbol{\xi};\hat{\boldsymbol{\beta}}_n)$ and $\tilde{M}_b^o(\boldsymbol{\xi};\boldsymbol{\beta}_o)$ are asymptotically equivalent from (4). It suffices to discuss the limit of $\tilde{M}_b^o(\boldsymbol{\xi};\boldsymbol{\beta}_o)$ under two different types of alternatives.

For part (i), rewrite $\tilde{M}_b^o(\boldsymbol{\xi};\boldsymbol{\beta}_o)$:

$$
\begin{aligned}
&\frac{1}{\sqrt{b}}\sigma_0^{-1}\sum_{i=1}^{b}m(y_i,\boldsymbol{x}_i,\boldsymbol{\beta}_o)\,\mathbb{1}_{\{\boldsymbol{x}_i\le\boldsymbol{\xi}\}}\\
&=\frac{1}{\sqrt{b}}\sigma_0^{-1}\sum_{i=1}^{b}\big[m(y_i,\boldsymbol{x}_i,\boldsymbol{\beta}_o)-\boldsymbol{\mu}(\boldsymbol{x}_i)\big]\,\mathbb{1}_{\{\boldsymbol{x}_i\le\boldsymbol{\xi}\}}+\frac{1}{\sqrt{b}}\sigma_0^{-1}\sum_{i=1}^{b}\boldsymbol{\mu}(\boldsymbol{x}_i)\,\mathbb{1}_{\{\boldsymbol{x}_i\le\boldsymbol{\xi}\}}.
\end{aligned}
$$

Under H_1 and assumptions [A1]–[A5], by the previous proofs, the first part of the above equation converges to $B(F(\boldsymbol{\xi}))$. In addition, if $\mathbb{E}|\boldsymbol{\mu}(\boldsymbol{x}_i)\,\mathbb{1}_{\{\boldsymbol{x}_i\le\boldsymbol{\xi}\}}|<\infty$, the probability limit of $b^{-1/2}\sigma_0^{-1}\sum_{i=1}^{b}\boldsymbol{\mu}(\boldsymbol{x}_i)\,\mathbb{1}_{\{\boldsymbol{x}_i\le\boldsymbol{\xi}\}}$ will be

$$
\frac{1}{\sqrt{b}}\sigma_0^{-1}\sum_{i=1}^{b}\mathbb{E}[\boldsymbol{\mu}(\boldsymbol{x}_i)\,\mathbb{1}_{\{\boldsymbol{x}_i\le\boldsymbol{\xi}\}}]=\sqrt{b}\sigma_0^{-1}\mathbb{E}[\boldsymbol{\mu}(\boldsymbol{x}_i)\,\mathbb{1}_{\{\boldsymbol{x}_i\le\boldsymbol{\xi}\}}].
$$

As $b\to\infty$, $\tilde{M}_b^o(\boldsymbol{\xi};\boldsymbol{\beta}_o)\to\infty$. Thus

$$
\tilde{M}_b(\boldsymbol{\xi};\hat{\boldsymbol{\beta}}_n)\to\infty.
$$

For part (ii), rewrite $\tilde{M}_b^o(\boldsymbol{\xi};\boldsymbol{\beta}_o)$:

$$
\begin{aligned}
&\frac{1}{\sqrt{b}}\sigma_0^{-1}\sum_{i=1}^{b}m(y_i,\boldsymbol{x}_i,\boldsymbol{\beta}_o)\,\mathbb{1}_{\{\boldsymbol{x}_i\le\boldsymbol{\xi}\}}\\
&=\frac{1}{\sqrt{b}}\sigma_0^{-1}\sum_{i=1}^{b}\left[m(y_i,\boldsymbol{x}_i,\boldsymbol{\beta}_o)-\frac{\boldsymbol{\delta}(\boldsymbol{x}_i)}{\sqrt{b}}\right]\mathbb{1}_{\{\boldsymbol{x}_i\le\boldsymbol{\xi}\}}+\frac{1}{b}\sigma_0^{-1}\sum_{i=1}^{b}\boldsymbol{\delta}(\boldsymbol{x}_i)\,\mathbb{1}_{\{\boldsymbol{x}_i\le\boldsymbol{\xi}\}}.
\end{aligned}
$$

Under H_1^L and assumptions [A1]–[A5], by the previous proofs, the first part of the above equation converges to $B(F(\boldsymbol{\xi}))$. If $\mathbb{E}|\boldsymbol{\delta}(\boldsymbol{x}_i)\,\mathbb{1}_{\{\boldsymbol{x}_i\leq\boldsymbol{\xi}\}}| < \infty$, the probability limit of $b^{-1}\sigma_0^{-1}\sum_{i=1}^{b}\boldsymbol{\delta}(\boldsymbol{x}_i)\,\mathbb{1}_{\{\boldsymbol{x}_i\leq\boldsymbol{\xi}\}}$ will be $\sigma_0^{-1}\mathbb{E}[\boldsymbol{\delta}(\boldsymbol{x}_i)\,\mathbb{1}_{\{\boldsymbol{x}_i\leq\boldsymbol{\xi}\}}]$. Therefore, under H_1^L, $\tilde{M}_b(\boldsymbol{\xi};\hat{\boldsymbol{\beta}}_n)$ converges to a Brownian sheet process plus a non-zero constant term $\sigma_0^{-1}\mathbb{E}[\boldsymbol{\delta}(\boldsymbol{x}_i)\,\mathbb{1}_{\{\boldsymbol{x}_i\leq\boldsymbol{\xi}\}}]$. □

Proof of Theorem 3.3. Similar to the proof of Theorem 3.1, rewrite $\tilde{M}_b^o(\boldsymbol{\xi};\boldsymbol{\beta}_o)$:

$$\begin{aligned}
&\frac{1}{\sqrt{b}}\sigma_0^{-1}\sum_{i=1}^{b} m\big(y_i,\boldsymbol{x}_i,\boldsymbol{\beta}_o\big)\,\mathbb{1}_{\{\boldsymbol{x}_i\leq\boldsymbol{\xi}\}}\\
&=\frac{1}{\sqrt{b}}\sigma_0^{-1}\sum_{i=1}^{b}\left[m\big(y_i,\boldsymbol{x}_i,\boldsymbol{\beta}_o\big)-\frac{\boldsymbol{\delta}(\boldsymbol{x}_i)}{\sqrt{n}}\right]\mathbb{1}_{\{\boldsymbol{x}_i\leq\boldsymbol{\xi}\}}+\frac{1}{\sqrt{b}\sqrt{n}}\sigma_0^{-1}\sum_{i=1}^{b}\boldsymbol{\delta}(\boldsymbol{x}_i)\,\mathbb{1}_{\{\boldsymbol{x}_i\leq\boldsymbol{\xi}\}}.
\end{aligned}$$

The probability limit of the second term on the right-hand-side of the above equation will be

$$\frac{1}{\sqrt{b}\sqrt{n}}\sigma_0^{-1}\sum_{i=1}^{b}\boldsymbol{\delta}(\boldsymbol{x}_i)\,\mathbb{1}_{\{\boldsymbol{x}_i\leq\boldsymbol{\xi}\}}=\frac{\sqrt{b}}{\sqrt{n}}\left[\sigma_0^{-1}\frac{1}{b}\sum_{i=1}^{b}\boldsymbol{\delta}(\boldsymbol{x}_i)\,\mathbb{1}_{\{\boldsymbol{x}_i\leq\boldsymbol{\xi}\}}\right]\xrightarrow{p}0,$$

with $b/n\to 0$ and $b^{-1}\sum_{i=1}^{b}\boldsymbol{\delta}(\boldsymbol{x}_i)\,\mathbb{1}_{\{\boldsymbol{x}_i\leq\boldsymbol{\xi}\}}\xrightarrow{p}\mathbb{E}[\boldsymbol{\delta}(\boldsymbol{x}_i)\,\mathbb{1}_{\{\boldsymbol{x}_i\leq\boldsymbol{\xi}\}}]$. Therefore, $\tilde{M}_b(\boldsymbol{\xi};\hat{\boldsymbol{\beta}}_n)$ converges to a Brownian sheet process under both H_0 and H_2^L. □

References

Andrews, D. and P. Guggenberger (2005). Hybrid and size-corrected subsample methods, unpublished manuscript, Cowels Foundation, Yale University.

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

Bickel, P. and M. Wichura (1971). Convergence criteria for multiparameter stochastic processes and some applications, *Annals of Mathematical Statistics*, **42**, 1656–1670.

Bierens, H. (1982). Consistent model specification tests, *Journal of Econometrics*, **20**, 105–134.

——— (1984). Model specification testing of time series regressions, *Journal of Econometrics*, **26**, 323–353.

——— (1990). A consistent conditional moment test of functional form, *Econometrica*, **58**, 1443–1458.

Bierens, H. and D. Ginther (2001). Integrated conditional moment testing of quantile regression models, *Empirical Economics*, **26**, 307–324.

Bierens, H. and W. Ploberger (1997). Asymptotic theory of integrated conditional moment tests, *Econometrica*, **65**, 1129–1151.

Chen, X. and Y. Fan (1999). Consistent hypothesis testing in semiparametric and nonparametric models for econometric time series, *Journal of Econometrics*, **91**, 373–401.

Chernozhukov, V. and I. Fernández-Val (2005). Subsampling inference on quantile regression process, *Sankha*, **67**, 253–276.

de Jong, R. (1996). On the Bierens test under data dependence, *Journal of Econometrics*, **72**, 1–32

Delgado, M. and W. González-Manteiga (2001). Significance testing in nonparametric regression based on the bootstrap, *Annals of Statistics*, **29**, 1469–1507.

Domínguez, M and I. Lobato (2006). Consistent estimation of models defined by conditional moment restrictions, *Econometrica*, **72**, 1601–1615.

Durbin, J. (1973). Weak convergence of the sample distribution function when parameters are estimated, *Annals of Statistics*, **1**, 279–290.

Eubank, R. and C. Spiegelman (1990). Testing the goodness-of-fit of a linear model via nonparametric regression techniques, *Journal of the American Statistical Association*, **85**, 387–392.

Fan, Y. and Q. Li (1996). Consistent model specification tests: omitted variables and semiparametric functional forms, *Econometrica*, **64**, 865–890.

Guggenberger, P. and M. Wolf (2004). Subsampling tests of parameter hypotheses and overifentifying restrictions with possible failure of identification, working paper, Department of Economics, UCLA.

Hong, H. and O. Scaillet (2006). A fast subsampling method for nonlinear dynamics models, *Journal of Econometrics*, **133**, 557–578.

Hong, Y. and H. White (1995). Consistent specification testing via nonparametric series regression, *Econometrica*, **63**, 1133–1159.

Horowitz, J. and V. Sponoiny (2001). An adaptive, rate-optimal test of a parametric mean=regression model against a nonparametric alternative, *Econometrica*, **69**, 599–631.

Khmaladze, E. (1981). Martingale approach in the theory of goodness-of-fit tests, *Theory of*

Probability and its applications, **XXVI**, 240–257.

Koul, H. and E. Stute (1999). Nonparametric model checks for time series, *Annals of Statistics*, **27**, 204–236.

Kuan, C.-M. and H.-y. Lin (2008). A consistent and asymptotical pivotal test for conditional moment restrictions, working paper.

Lewbel, A. (1995). Cinsistent nonparametric hypothesis tests with an application to Slutsky symmetry, *Journal of Econometrics*, **67**, 379–401.

Li, Q. and S. Wang (1998). A simple consistent bootstrap test for a parametric regression function, *Journal of Econometrics*, **87**, 145–165.

Li, W, C. Hsiao, and J. Zinn (2003). Consistent specification tests for semiparametric/nonparametric models based on series estimation methods, *Journal of Econometrics*, **112**, 295–325.

Linton, O., Maasoumi, E. and Y.-J. Whang (2005). Consistent testing for stochastic dominance under general sampling schemes, *Review of Economic Studies*, *72*, 735–765.

Newey, W. (1985). Maximum likelihood specification testing and conditional moment tests, *Econometrica*, **53**, 1047–1070.

Politis, D. N. and J. P. Romano (1994). Large sample confidence regions based on subsamples under minimal assumptions, *Annals of Statistics*, **22**, 2031–2050.

Politis, D. N., J. P. Romano, and M. Wolf (1999). *Subsampling*, New York: Springer.

Shorack, G. and J. Wellner (1986). *Empirical Processes with Applications to Statistics*, New York: John Wiley & Sons.

Song. K. (2007). Testing semiparametric conditional moment restrictions using conditional martingale transforms, working paper.

Stichcombe, M. and H. White (1998). Consistent specification testing with nuisance parameters present only under the alternative, *Econometric theory*, **14**, 295–325.

Stute, W. (1997). Nonparametric model checks for regression, *Annals of Statistics*, **25**, 613–641.

Stute, W., W. González Manteiga and M. Presedo Quindïmil (1998). Bootstrap approximations in model check for regression, *Journal of the American Statistical Association*, **93**, 141–149.

Stute, W., S. Thies, and L. Zhu (1998). Model checks for regression: an innovation process approach, *Annals of Statistics*, **26**, 1916–1934.

Stute, W. and L. Zhu (2002). Model checks for generalized linear models, *Scandinavian Journal of Statistics*, **29**, 535–545.

Tauchen, G. (1985). Diagnostic testing and evaluation of maximum likelihood models, *Journal of Econometrics*, **30**, 415–443.

Tripathi, G. and Y. Kitamura (2003). Testing conditional moment restrictions, *Annals of Statistics*, **31**, 2059–2095.

Whang, Y. (2000). Consistent bootstrap tests of parametric regression functions, *Journal of Econometrics*, **98**, 27–46.

——— (2001). Consistent specification testing for conditional moment restrictions, *Economics Letters*, **71**, 299–306.

——— (2004). Consistent specification testing for quantile regression models, working paper.

White, H. (1987). Specification testing in dynamic models. In T. Bewley (ed.), *Advances in Econometrics–Fifth World Congress*, **1**, New York: Cambridge University Press.

Zheng, J. (1998a). A consistent nonparametric test of parametric regression models under conditional quantile restrictions, *Econometric Theory*, **14**, 123–138.

——— (1998b). Consistent specification testing for conditional symmetry, *Econometric Theory*, **14**, 139–149.

——— (2000). A consistent test of conditional parametric distributions, *Econometric Theory*, **16**, 667–691.