Nonparametric Tests of Moment Condition Stability

by Ted Juhl and Zhijie Xiao

2013

Ted Juhl and Zhijie Xiao. (2013). Nonparametric Tests of Moment Condition Stability. *Econometric Theory* 29(1):91-114

Published version: http://dx.doi.org/10.1017/S0266466612000151

Econometric Theory, **29**, 2013, 90–114.
doi:10.1017/S0266466612000151

NONPARAMETRIC TESTS OF MOMENT CONDITION STABILITY

TED JUHL
University of Kansas

ZHIJIE XIAO
Boston College

This paper considers testing for moment condition instability for a wide variety of models that arise in econometric applications. We propose a nonparametric test based on smoothing the moment conditions over time. The resulting test takes the form of a U-statistic and has a limiting normal distribution. The proposed test statistic is not affected by changes in the distribution of the data, so long as certain simple regularity conditions hold. We examine the performance of the test through a small Monte Carlo experiment.

1. INTRODUCTION

Many econometric models are estimated by imposing some population moment conditions on the sample. An important assumption in these models is that the moment condition holds throughout the entire sample. Given that this assumption is important to further analysis of the model, a number of tests for moment condition stability have been proposed. In particular, Andrews and Fair (1988) proposed Wald, likelihood ratio-type, and Lagrange multiplier-type tests, and Ghysels and Hall (1990) proposed a predictive test for the case where the structural break date is known. These techniques are also extended to the case with unknown breakpoint by Andrews (1993), Andrews and Ploberger (1994), Sowell (1996), Ghysels, Guay, and Hall (1997), and Hall and Sen (1999a): By calculating a chosen statistic for each possible breakpoint, a sequence of statistics can be obtained and a test for structural change with unknown breakpoint can be constructed based on a function of the sequence. The literature on testing for structural breaks is vast; in addition to the above-mentioned work, an incomplete list includes Mankiw and Shapiro (1986), Ploberger and Kramer (1996), Kuan (1998), Vogelsang (1997, 1998), Bai and Perron (1998), Hall and Sen (1999b), Hansen (2000), and Maynard and Shimotsu (2009). Some tests that involve nonparametric estimation include Hidalgo (1995) and Inoue (2001).

In this paper we consider the inference problem of testing for moment (or conditional moment) instability. The proposed tests are nonparametric. In

We thank Peter Phillips, Jeff Wooldridge, and three referees for very helpful comments on early versions of this paper. Address correspondence to Zhijie Xiao, Department of Economics, Boston College, Chestnut Hill, MA 02467, U.S.A.; e-mail: xiaoz@bc.edu.

particular, we construct test statistics based on smoothing the moment conditions over time. The approach follows the idea of Robinson (1989), who develops a kernel-based nonparametric approach to estimating time-varying parameters in a linear regression model. We propose a test of the stability of the population moment condition based on this idea. Our tests take a U-statistic form, and we show that the statistics have an asymptotic normal distribution.

The contribution of this paper is to provide tests of moment stability for a wide variety of conditional models that arise in econometric applications. The proposed test is not affected by changes in the distribution of the data, so long as the moment conditions tested are not violated and certain regularity conditions hold. A leading example covered by our model is the inference problem of structural change of unknown timing in regression models. In this example, our primary interest is usually on the constancy of regression parameters, and not particularly concerned with the distribution of the regressors. Most existing tests (see, inter alia, Andrews, 1993; Andrews and Ploberger, 1994) are based on the assumption that the regressors are stationary and thus cannot discriminate between instability of parameters and structural change in the distribution of the regressors. Hansen (2000) proposes a "fixed regressor bootstrap" test for parameter constancy in regression models that allows for structural change in the marginal distribution of the regressors. When specialized to the regression models, the current paper provides an alternative test for this inference problem[1].

The rest of the paper is organized as follows: In Section 2 we introduce the model and propose a test for conditional moment instability. The asymptotic theory is provided in Section 3. A generalization to dependent moments is studied in Section 4. We present the results from a small Monte Carlo study in Section 5, and Section 6 concludes. The proofs are given in the Appendix.

2. THE MODEL

Given observed data $\{z_t, t = 1, \dots, T\}$, we consider a parametric model with population moment conditions:

$$\mathbb{E}(m(z_t, \theta_0)) = 0, \tag{2.1}$$

where θ is a $p \times 1$ vector of parameters and m is an $L \times 1$ vector where $L \geq p$. We estimate the model using, say, the generalized method of moments (GMM). Thus, the parameters, θ, are estimated under the assumption that the population moment condition is not time varying, and hence zero for all time periods. This assumption is crucial to further analysis of the model. We are interested in whether or not the moment conditions (2.1) indeed hold over time. In this paper, we propose a nonparametric test to check the moment stability.

Many important inference problems can be expressed in the form of moment instability. For example, the well-known problem of structural change in regression models can be rewritten into a form covered by our model. Consider the linear regression model

$$y_t = x_t^\top \theta + u_t,$$

and suppose that we are interested in whether or not θ is constant. To reformulate it into an inference problem of moment stability, let $z_t = (y_t, x_t^\top)^\top$ and

$$m(z_t, \theta) = \left(y_t - x_t^\top \theta \right) x_t.$$

Then, under the null hypothesis that θ is constant, the moment condition (2.1) holds for all t. Otherwise, for any vector θ of constants, $\mathbb{E}(m(z_t, \theta)) \neq 0$ for some nonnegligible fraction of the sample. Most studies in the literature consider the case where the distribution of x_t is assumed to be stationary. Furthermore, we may allow for changes in distribution of x_t but focus our interest on the constancy of parameters θ. In this case our model reduces to a conditional model similar to that of Hansen (2000).

The test that we propose to check moment instability is a variant of nonparametric conditional moment tests as developed in Fan and Li (1999), Hjellvik, Yao, and Tjøstheim (1998), Lavergne and Vuong (2000), Li and Wang (1998), and Zheng (1996), among others.

For convenience of our analysis, we introduce some notation. We use the notation $m_t = m(z_t, \theta)$, and denote the density of m_t by $f_t(m_t)$, where the subscript t in f_t indicates that the density may be time varying. In addition, we use the notation $\mu(t)$ for $\mathbb{E}(m_t)$ to indicate that $\mathbb{E}(m_t)$ may be time varying (under the alternative of moment instability). Thus,

$$\mu(t) = \mathbb{E}(m_t) = \int_{-\infty}^{\infty} m_t f_t(m_t) dm_t.$$

The null hypothesis of interest here is one of moment stability:

$$H_0 : \text{There exists a } \theta \text{ such that} \quad \mu(t) = 0 \quad \text{almost everywhere (a.e.)},$$

and the alternative hypothesis is: $H_1 : \mu(t) \neq 0$ for some nonnegligible fraction of the sample. Suppose that

$$\frac{1}{T} \sum_{t=1}^{T} m_t^\top \mu(t) \xrightarrow{p} \lambda.$$

Notice that under the null hypothesis, we have $\lambda = 0$. Under regularity conditions limiting the dependence of m_t and bounding the variances, it is easy to show that

$$\frac{1}{T} \sum_{t=1}^{T} m_t^\top \mu(t) = \frac{1}{T} \sum_{t=1}^{T} \mu(t)^\top \mu(t) + o_p(1). \tag{2.2}$$

Then $\lambda = 0$ implies that $\mu(t) = 0$ almost everywhere. If $\lambda \neq 0$, then $\mu(t) \neq 0$ for some nonnegligible fraction of the sample.

We want to base a statistic on the quantity λ that requires an estimate of $\mu(t)$. Following Robinson (1989), we can view the moment conditions as depending on the scaled time index t/T. That is, we denote $\mu(t) = \mu(\frac{t}{T})$. This representation allows for more points to accumulate near a given t as T grows to infinity. In this way, we can consistently estimate $\mu(\frac{t}{T})$ using a leave-one-out version of the kernel-based Priestley–Chao (1972) estimator

$$\widehat{\mu}(t) = \frac{1}{Th} \sum_{s \neq t}^{T} K\left(\frac{t-s}{Th}\right) \hat{m}_s, \tag{2.3}$$

where h is a bandwidth parameter, $\hat{m}_t = m(z_t, \hat{\theta}_T)$ is an estimate of m_t using a preliminary estimate of θ, and $K(\cdot)$ is a kernel function. This estimator essentially takes averages of m_s points with s close to t. A smaller bandwidth uses the points where s is closer to t and a larger bandwidth uses more points. The proposed test statistics are based on the above kernel-smoothed estimator. We propose a test for conditional moment stability and consider its asymptotic distribution in the next section, and study the inference problem of testing for moment condition instability in general dynamic models in Section 4.

3. A NONPARAMETRIC TEST OF CONDITIONAL MOMENT STABILITY

We first consider the case that m_t is a martingale difference sequence ($\mathbb{E}(m_t | \mathcal{F}_{-\infty}^{t-1}) = 0$), and test for conditional moment stability. Many econometric models impose some conditional moment assumptions on the sample. For example, rational expectations-based economic models often imply that the conditional expectation of some variable, conditional on past information, is zero. For this reason, testing conditional moment stability is in and of itself interesting and important, and worthy of discussion separately. We propose a test for conditional moment stability in this section and extend our analysis to general moment stability inference in the next section.

The proposed test is motivated based on our discussion in the previous section, in particular, by plugging the nonparametric estimator (2.3) into equation (2.2). Thus, the proposed test statistic is based on the following quantity:

$$\hat{\lambda}_T = \frac{1}{T^2 h} \sum_{t=1}^{T} \sum_{s \neq t}^{T} K\left(\frac{t-s}{Th}\right) \hat{m}_t^\top \hat{m}_s.$$

To facilitate asymptotic analysis of the proposed test statistic under the null, we introduce the following assumptions.

Assumption 1. Let $\mathcal{F}_s^t = \sigma(m_s, \ldots, m_t)$, $\mathbb{E}(m_t | \mathcal{F}_{-\infty}^{t-1}) = 0$ and $\mathbb{E}(m_t m_t^\top) = \Sigma(\frac{t}{T}) < \infty$. In addition, $\Sigma(v)$ is a continuous function on $v \in (0, 1)$.

Assumption 2. The process $\{(z_t)\}$ is absolutely regular. That is,

$$\beta(\tau) = \sup_s \mathbb{E}\left\{ \sup_{A \in \mathcal{G}_{s+\tau}^{\infty}} |P(A|\mathcal{G}_{-\infty}^{s}) - P(A)| \right\} \to 0, \quad \text{as } \tau \to \infty,$$

where $\mathcal{G}_s^t$ is the σ-field generated by $\{(z_j) : j = s, \ldots, t\}$. For a constant $\delta > 0$, $\sum_{\tau=1}^{\infty} \tau^2 \beta(\tau)^{\delta/(1+\delta)} < \infty$.

Assumption 3. $\nabla m(z_t, \cdot)$ and $\nabla^2 m(z_t, \cdot)$ represent the first and second derivatives of $m(z_t, \cdot)$. They are bounded by a function $M_g(z_t)$ that has finite second moments.

Assumption 4.

$$\max\{M_1, M_2, M_3, M_4, M_5\} < \infty$$

where M_i, $i = 1, \ldots, 5$ are defined in the Appendix.

Assumption 5. $\sqrt{T}(\hat{\theta}_T - \theta_0) = O_p(1)$.

Assumption 6. $K(u)$ is a density function such that $K(0) \geq K(u)$ for all u, and $\int K(u)^2 du < \infty$.

Assumption 7. $h \to 0$, $Th^2 \to \infty$.

Assumption 1 states that m_t is a martingale difference sequence. We allow that the variance of the moment conditions may be time-varying, but continuous. For this purpose, following the literature of nonparametric estimation, we restandardize the time index so that they are on a compact set [0, 1], and assume that the variance function is continuous on [0,1]. Assumption 2 limits the dependence found in the data, z_t, which is used to form the moment conditions. Absolute regularity, or beta mixing, is a condition that is stronger than α-mixing but weaker than ϕ-mixing. See Fan and Li (1997) for a list of well-known processes that satisfy absolute regularity, and Fan and Li (1999) and Li and Wooldridge (2002) for studies of nonparametric estimation with absolutely regular processes. Notice that as stated, absolute regularity does not require stationarity. That is, the mixing coefficients are the sup over s of all events that are τ units of time apart. This means that the expectation may vary over different values of s, and we take the largest as the β coefficient. A simple example is an independent process with some heteroskedasticity. This is a very important distinction since many tests of structural change are not robust to, say, a simple mean or variance change in the regressors. Hansen (2000) provides a thorough analysis of tests for structural change when there is a change in the marginal distribution of regressors. Assumption 3 is standard in nonparametric conditional moment tests such as Fan and Li (1999). In Assumption 4 we have several moment conditions. Notice that although we do not require stationarity, our moment conditions in Assumption 4 rule out unit root processes or trending regressors. In addition, this assumption requires slightly more than eighth moments for the data. Assumption 5 is a high-level assumption stating that the parameters of the model are estimated at rate $T^{1/2}$. The limiting distribution of $\hat{\theta}_T$ does not affect the distribution of our test since we are using nonparametric estimation of the residuals. The rate of convergence of the nonparametric estimation procedure is slower than $T^{1/2}$, so that the final limiting distribution is not affected by the distribution of $\hat{\theta}_T$. Assumptions 6 and 7 are standard assumptions in the kernel regression literature.

THEOREM 3.1. *Suppose that Assumptions 1–7 hold. Then*

$$\frac{Th^{1/2}\hat{\lambda}_T}{\hat{\sigma}} \xrightarrow{d} N(0,1),$$

where

$$\hat{\sigma}^2 = \frac{2}{T^2 h}\sum_{t=1}^{T}\sum_{s\neq t}^{T} K\left(\frac{t-s}{Th}\right)^2 \left(\hat{m}_t^\top \hat{m}_s\right)^2$$

is a consistent estimate of $\sigma^2 = 2\int_0^1 K(u)^2 du \int_0^1 \text{trace}\,(\Sigma(v)\Sigma(v))\,dv$.

The proof of the theorem uses the U-statistic structure of the statistic. As in Hall (1984) and Hjellvik et al. (1998), we show that a martingale central limit theorem applies. However, in our case the arguments of the kernel function are deterministic, and results for strictly stationary series do not apply. Moreover, we allow for the distribution of the data to vary over time.

The limiting distribution of the standardized statistic is standard normal. This limiting result is robust to changes in the mean or variance of the data provided that the moment condition tested does not change ($\mathbb{E}(m_t) = 0$) and that the moment conditions in Assumption 4 still hold. One implication is that when the tested moment conditions fail to hold over the entire sample, we can attribute a rejection of the test to parameter instability. We explore this important special case in a Monte Carlo experiment in Section 5.

The proposed testing statistic has a similar structure to the U-statistic tests proposed by Zheng (1996) and Li (1999). Although we do not explore local power in this paper, we conjecture that following a similar analysis, our test will also have power against local alternatives that are $O_p(T^{-1/2}h^{-1/4})$.

An important issue in the calculation of our tests is the choice of bandwidth parameter. The conditions of the theorem only provide a range of possible bandwidth choices. In this paper we follow Zheng (1996) and Li (1999) in that we explore several choices of bandwidth to examine the sensitivity of size and power for the tests. An alternative strategy was suggested by Horowitz and Spokoiny (2001). They propose calculating the supremum of nonparametric tests over a range of possible bandwidth parameters as a test statistic. This approach may be possible with our test. However, the distribution would change, and the proofs would become much more complicated. We leave this idea for future research.

4. TESTING MOMENT STABILITY WITH DEPENDENCE IN THE MOMENTS

An important assumption that we imposed in the previous section is that m_t is a martingale difference sequence. This type of test is of interest in practice because many econometric models are estimated by imposing some conditional moment assumptions. In this section we relax the assumption that m_t is a martingale difference sequence and extend our analysis to inference problems of moment stability for general dynamic models where m_t are allowed to be serially correlated. To accommodate this extension, we need to modify our assumptions and the test statistic.

In designing a test for conditional moment stability, we use a leave-one-out estimator (2.3). In order to extend our test for moment stability and to the case that m_t is serially correlated, we propose the following leave-k-out estimator,

$$\widetilde{\mu}(t) = \frac{1}{Th} \sum_{s \notin B(t)} K\left(\frac{t-s}{Th}\right) \hat{m}_s,$$

where $B(t) = [t - Th_2, t + Th_2]$ is a growing neighborhood around (t) and h_2 is a bandwidth parameter that satisfies the assumption that $h_2/h \to 0$, as $T \to \infty$. To this end, the proposed test statistic[2] is

$$\widetilde{\lambda}_T = \frac{1}{T^2 h} \sum_{t=1}^{T} \sum_{s \notin B(t)} K\left(\frac{t-s}{Th}\right) \hat{m}_t^\top \hat{m}_s.$$

We modify Assumptions 1 and 7 as follows.

Assumption 1′. $\mathbb{E}(m_t) = 0$ and $\mathbb{E}(m_t m_t^\top) = \Sigma(\frac{t}{T}) < \infty$. In addition, $\Sigma(v)$ is a continuous function on $v \in (0,1)$. The process $\{(m_t)\}$ is absolutely regular. That is,

$$\beta(\tau) = \sup_s \mathbb{E}\left\{ \sup_{A \in \mathcal{G}_{s+\tau}^{\infty}} \left| P\left(A | \mathcal{G}_{-\infty}^{s}\right) - P(A) \right| \right\} \to 0, \quad \text{as } \tau \to \infty,$$

where $\mathcal{G}_s^t$ is the σ-field generated by $\{(m_j) : j = s, \ldots, t\}$. For a constant $\delta > 0$, $\sum_{\tau=1}^{\infty} \tau^2 \beta(\tau)^{\delta/(1+\delta)} < \infty$.

Assumption 7′. As $T \to \infty$, $h \to 0$, $h_2 \to 0$, $h_2/h \to 0$, $Th^2 \to \infty$, and $T^4 h_2^5 \to \infty$.

THEOREM 4.1. *Suppose that Assumptions 1′, 2–6, and 7′ hold, then*

$$\frac{T h^{1/2} \tilde{\lambda}_T}{\hat{\sigma}} \xrightarrow{d} N(0,1),$$

where

$$\hat{\sigma}^2 = \frac{2}{T^2 h} \sum_{t=1}^{T} \sum_{s \notin B(t)} K\left(\frac{t-s}{Th}\right)^2 \left(\hat{m}_t^\top \hat{m}_s\right)^2$$

is a consistent estimate of $\sigma^2 = 2\int_0^1 K(u)^2 du \int_0^1 \mathrm{trace}\,(\Sigma(v)\Sigma(v))\, dv$.

Even though β mixing processes include many standard cases in econometrics (such as autoregressive moving average (ARMA) models with some restrictions on the density), it may be possible to extend our results to more general cases. Chen and Fan (1999) suggest that many of their results for U-statistics can be extended to near-epoch-dependent processes.[3] We leave this direction for future research.

5. MONTE CARLO

We examine the finite sample performance of our proposed statistic using a small set of Monte Carlo experiments. Our test is applied to the moment condition $\mathbb{E}(x_t \epsilon_t) = 0$, and we wish to see how our test compares to other tests in detecting a structural change in regression parameters. In the first set of experiments, we examine the size of our test using the model

$$y_t = \alpha_0 + \alpha_1 x_t + \epsilon_t.$$

Initially, we consider four specifications of the model. In case 1, ϵ_t is homoskedastic and follows a standard normal distribution and x_t follows a t-distribution with five degrees of freedom. This case is denoted *iid* in our tables. Following Hansen (2000), we also treat several cases involving heteroskesticity in ϵ_t and changes in the distribution of x_t. In case 2, $\epsilon_t \sim N(0, 1 + 0.25x_t^2)$, and we denote this as *het* in the tables. For case 3, ϵ_t again follows the distribution in case 2 but now there is a mean shift in x_t so that

$$x_t = \begin{cases} \eta_t & \text{for } t < 0.5T \\ \eta_t + 5 & \text{for } t \geq 0.5T, \end{cases}$$

where η_t follows a t-distribution with five degrees of freedom. We label this case by *hetmb* for heteroskedasticity mean break. Case 4 uses a similar specification for ϵ_t and η_t, but now there is a change in the variance of x_t so that

$$x_t = \begin{cases} \eta_t & \text{for } t < 0.5T \\ 5\eta_t & \text{for } t \geq 0.5T, \end{cases}$$

and this case is denoted *hetvb*. For case 5, we allow for x_t to have an autoregressive structure. In particular,

$$x_t = \omega_t + \rho x_{t-1} + \epsilon_t,$$

where $\rho = 0.7$ and

$$\omega_t = \begin{cases} 0 & \text{for } t < 0.5T \\ 5 & \text{for } t \geq 0.5T. \end{cases}$$

This case corresponds to a break in an autoregressive process, and we label the results *arbreak* in the table. It is well documented that an autoregressive process with a break is often mistaken for a unit root process (see Perron, 1989). We consider this case to see what happens to all of the tests under this type of change in the marginal distribution of x_t.

Without loss of generality, we set the values of $\alpha_0 = 1$ and $\alpha_1 = 1$. We compare our nonparametric moment stability test, denoted NPMS, to some other standard tests. Our test requires the choice of a bandwidth. To determine the sensitivity of our test, we include three bandwidth choices, $h = 0.125 \times T^{-1/5}$, $h = 0.25 \times T^{-1/5}$, and $h = 0.5 \times T^{-1/5}$, all of which are larger than the bandwidth parameters used in the first set of experiments. The three bandwidths generate tests indexed as NPMS1, NPMS2, and NPMS3. We also consider other tests. First, we include the SupF test analyzed by Andrews (1993) and others. Essentially, this test is calculated by dividing the sample into two parts and testing that α_0 and α_1 are identical in each part of the sample. The data are broken into two parts beginning with 15% of the data until 85% of the data, and the maximum of all the F tests becomes the statistic. In addition, we conduct a Wald test in the same manner but we use a heteroskedasticity-consistent covariance matrix when calculating the variances of the estimated α parameters. Finally, following Hansen (2000), we include a fixed regressor bootstrap version of the SupF test, which we denote HetBoot in our tables. The number of bootstrap replications is set to 1,000. For all experiments, the number of replications for each test is 2,000. The results for the size experiment appear in Table 1. In the independent and identically distributed (i.i.d.) case, we note that with the exception of the SupWald test, all of the tests have good size or are conservative. As the bandwidth gets larger, the NPMS test becomes more conservative. This in and of itself is not a problem so long as power does not suffer. We will explore this possibility later in the paper. In general, the tests improve as the sample size increases. In the heteroskedasticity case, the results are slightly different. The fixed regressor bootstrap is slightly oversized, and the size gradually improves as the sample size increases. The SupF test is oversized, and the effect gets worse as the sample size increases. The SupWald test is initially oversized and has worse performance than the SupF test. However, as the sample size increases, the SupWald has better size as the improved estimation of the heteroskedasticity-consistent covariance matrix helps reduce the initial size distortion. The nonparametric moment stability tests are conservative for all of the bandwidth choices in this first experiment. Moreover, the test becomes more conservative as the bandwidth size increases. When there is heteroskedasticity and a change in the mean of the regressors, the results are magnified. The SupF and SupWald tests are severely oversized. These results match the findings given

TABLE 1. Empirical size

	T	SupF	SupWald	HetBoot	NPMS1	NPMS2	NPMS3
iid	50	0.046	0.313	0.049	0.031	0.024	0.015
	100	0.050	0.172	0.057	0.031	0.024	0.020
	200	0.045	0.102	0.057	0.044	0.036	0.023
	800	0.049	0.068	0.060	0.035	0.032	0.022
het	50	0.148	0.384	0.085	0.036	0.028	0.017
	100	0.178	0.222	0.076	0.037	0.031	0.019
	200	0.232	0.143	0.074	0.041	0.037	0.024
	800	0.309	0.074	0.049	0.040	0.035	0.026
hetmb	50	0.208	0.435	0.078	0.032	0.028	0.012
	100	0.228	0.254	0.071	0.041	0.028	0.016
	200	0.225	0.161	0.061	0.033	0.029	0.015
	800	0.275	0.096	0.055	0.036	0.032	0.021
hetvb	50	0.578	0.995	0.118	0.027	0.020	0.010
	100	0.714	0.999	0.096	0.026	0.031	0.014
	200	0.775	1.000	0.067	0.030	0.023	0.010
	800	0.915	0.999	0.052	0.031	0.029	0.022
arbreak	50	0.073	0.512	0.024	0.031	0.027	0.010
	100	0.062	0.307	0.037	0.034	0.029	0.012
	200	0.069	0.214	0.039	0.040	0.033	0.014
	800	0.075	0.121	0.044	0.049	0.038	0.026

in Hansen (2000) for a similar experiment design. The nonparametric moment stability tests are still conservative as expected, since the assumptions of our theorem allow for such changes in the distribution of x_t. The *arbreak* specification seems to affect the SupWald test the most, however; as the sample size increases, the effect is once again mitigated by the improved estimation of the covariance matrix.

These results of the experiment, particularly those when there is a change in the distribution of the regressors, reveal an advantage in the new nonparametric tests. Notice that the SupWald and SupF tests are extremely oversized in the presence of a change in distribution of regressors, while the proposed nonparametric tests are not. Hence, when the SupWald and SupF tests are used, a change in the distribution of x_t will be incorrectly interpreted as a change in the model through the parameter α. This interpretation is false since the response of y_t to changes in x_t has not changed. Therefore, our nonparametric tests are useful in differentiating between changes in α and changes in the mean or variance of the regressors. The fixed regressor bootstrap of Hansen (2000) provides some protection against this phenomenon as well.

The model we consider in the second experiment is the same as the first with two exceptions. First, the regressors are governed by the autoregressive process

$$x_t = 0.5x_{t-1} + \eta_t.$$

Second, the error process is given by

$$\epsilon_t = \rho\epsilon_{t-1} + u_t,$$

where u_t are i.i.d. and $N(0,1)$. As shown in Theorem 4.1, it is possible to modify the nonparametric moment stability test so that it is valid given a certain amount of dependence in the moments. The modification of the leave-k-out estimation requires a second bandwidth. We set this as $h_2 = h \times T^{-1/5}$ so that the conditions of Assumption 7′ are satisfied.

It is well known that as the dependence increases, the optimal bandwidth should increase for nonparametric estimation. In light of this fact, we consider three bandwidths with $h = T^{-1/5}$, $h = 1.5 \times T^{-1/4}$, and $h = 2 \times T^{-1/5}$. The tests associated with these bandwidths are denoted NPMS4, NPMS5, and NPMS6. Altman (1990) shows that the optimal bandwidth in nonparametric regression with dependent data should be proportional to

$$\left(\gamma_0 + 2\sum_{i=1}^{\infty}\gamma_i\right)^{1/5} T^{-1/5},$$

where γ_i is the ith covariance. We approximate this quantity in a similar fashion to Andrews (1991) using a plug-in method based on a first-order autoregressive process. That is, we assume

$$m_t = \phi_m m_{t-1} + \eta_t,$$

where the variance of η_t is given by σ_η^2. This assumption is only used for the bandwidth selection. Then the optimal bandwidth is proportional to

$$\left(\frac{\sigma_\eta}{(1-\phi_m)^2}\right)^{1/5} T^{-1/5}.$$

We denote the test associated with this data-dependent bandwidth as NPMSD.

The SupF test and the fixed regressor bootstrap of Hansen (2000) are not applicable here since ϵ_t is not a martingale difference sequence. However, we include the SupF test for comparison purposes.[4] The SupWald test can be modified to make it asymptotically valid for such an experiment. This requires the use of a heteroskedaticity autocorrelation consistent (HAC) covariance matrix for the regression parameters. We choose the HAC given in Andrews (1991), which uses a quadratic spectral kernel with the corresponding data-dependent bandwidth.[5] The results appear in Table 2.

From the results, we see that size generally increases as ρ increases. For the SupWald test, performance is very bad for small sample sizes but improves rapidly as sample size increases. The size for SupF worsens as the sample size increases. The leave-k-out version of the nonparametric moment stability tests is conservative for small values of ρ but has a larger size distortion as ρ increases. The

TABLE 2. Size for dependent errors

ρ	T	SupF	SupWald	NPMS4	NPMS5	NPMS6	NPMSD
0.30	50	0.203	0.571	0.043	0.050	0.050	0.008
0.30	100	0.200	0.394	0.057	0.054	0.054	0.006
0.30	200	0.215	0.262	0.053	0.057	0.051	0.008
0.30	400	0.228	0.174	0.057	0.049	0.045	0.012
0.30	800	0.262	0.125	0.063	0.044	0.041	0.014
0.50	50	0.382	0.703	0.056	0.071	0.080	0.014
0.50	100	0.468	0.553	0.080	0.082	0.078	0.013
0.50	200	0.467	0.347	0.081	0.082	0.103	0.013
0.50	400	0.501	0.236	0.102	0.084	0.089	0.021
0.50	800	0.519	0.132	0.092	0.087	0.087	0.027
0.70	50	0.662	0.832	0.069	0.108	0.123	0.023
0.70	100	0.729	0.681	0.108	0.132	0.138	0.024
0.70	200	0.780	0.506	0.133	0.137	0.149	0.028
0.70	400	0.795	0.327	0.171	0.148	0.162	0.032
0.70	800	0.829	0.217	0.147	0.132	0.138	0.054

data-dependent bandwidth version of the test appears to have the best performance. Of the tests considered here, it is least likely to be oversized.

To explore power properties, we let the slope parameter vary so that

$$\alpha_1 = \begin{cases} 0 & \text{for } t = 1, \ldots, 0.5T \\ \gamma & \text{for } t = 0.5T + 1, \ldots, T. \end{cases}$$

The error and regressor processes are as given in the first experiment, where the error is i.i.d., and we set $T = 100$. The size-adjusted power curves appear in Figure 1. We notice that size-adjusted power for the nonparametric moment stability tests improves as the bandwidth gets larger and that for the largest bandwidth, the test is competitive in terms of power. However, this may be purely an artifact of the size adjustment. That is, the nonparametric moment stability test with the largest bandwidth is the most undersized, so that the correction gives the largest power increase to this test. To explore this possibility, we show the raw power for our tests in Figure 2. Notice that when γ is zero, we plot the size and the NPMS3 test has the smallest size. However, even when raw power is examined, the power is greatest for the most undersized test (NPMS3). Therefore, we conclude the high power of this test is not solely an artifact of the size adjustment. The larger power for the more conservative test (using the larger bandwidth) is unexpected, but this is a small sample phenomenon. Evidently, for the larger bandwidth, the numerator of our test statistic is more sensitive to changes in α than the denominator in small samples.

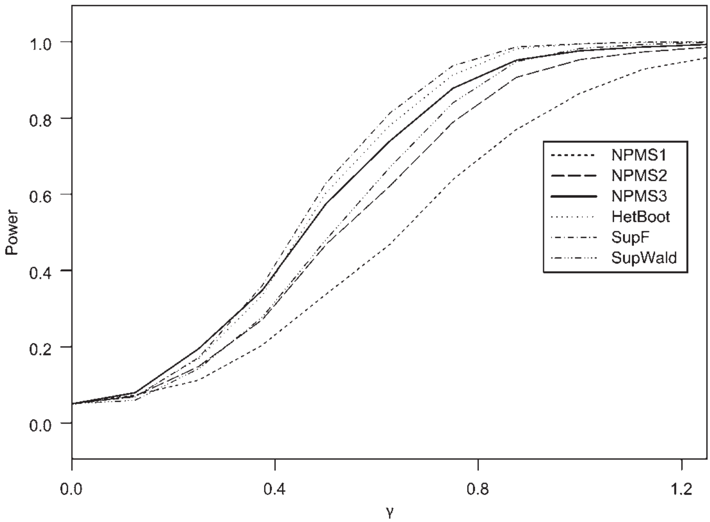

FIGURE 1. Size-Adjusted power

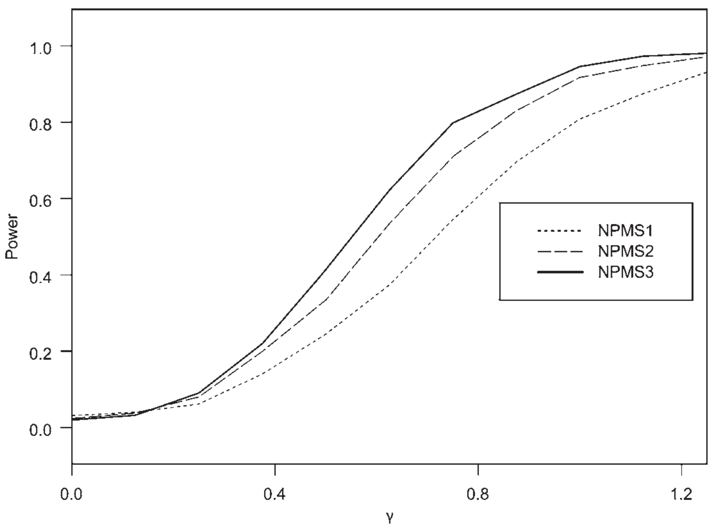

FIGURE 2. Unadjusted power

6. CONCLUSION

The tests proposed in this paper have several desirable properties. First, the limiting distribution is standard normal even if there is a change in mean or variance in the data. This eliminates the need for a bootstrapping procedure to obtain first-order asymptotics. The first version of the test (which is appropriate when the moments are martingale differences) is conservative yet has power against a variety of structural changes in parameters. In our Monte Carlo experiment, we found that the bandwidth choice that makes the test the most conservative (under the null) does not necessarily have the lowest size unadjusted power. The second version of the test is applicable when there is dependence in the moments. We find that this test has better size properties than the competing tests, but all of the tests have size distortions when there is more dependence.

NOTES

1. Our model does not consider unit root or trending regressors.

2. As a referee pointed out, there is a connection between our proposed test and long-run variance estimation of the process m_t. Suppose that m_t is a scalar process. Then denote the long-run variance estimator of m_t as $\hat{\Omega} = \hat{\gamma}_0 + 2\sum_{j=1}^{T-1} K\left(\frac{j}{hT}\right)\hat{\gamma}_j$, where $\hat{\gamma}_j$ is the usual jth autocovariance. Then $\hat{\lambda}_T = \frac{1}{hT}\hat{\Omega}$. Now define $\hat{\Omega}^* = \hat{\gamma}_0 + 2\sum_{j=1}^{T-1} K\left(\frac{j}{hT}\right)1(j \le k)\hat{\gamma}_j$, where k is from the leave-k-out estimator. Then $\tilde{\lambda}_T = \frac{1}{hT}\left(\hat{\Omega} - \hat{\Omega}^*\right)$. Both $\hat{\Omega}$ and $\hat{\Omega}^*$ are estimators of the spectral density at zero, and hence can be made asymptotically normal given the rates of k and h.

3. See Wooldridge (1994) for a discussion of near-epoch-dependent processes in econometrics.

4. Simulations not shown here indicate that the performance of the fixed regressor bootstrap and SupF are similar in this design.

5. We use the AR(1) approximation of ϵ_t to select the data-dependent bandwidth used in the long-run variance estimator. This bandwidth diverges at rate $T^{1/5}$.

REFERENCES

Altman, N.S. (1990) Kernel smoothing of data with correlated errors. *Journal of the American Statistical Association* 85, 749–759.

Andrews, D.W.K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.

Andrews, D.W.K. (1993) Tests for parameter instability and structural change with unknown change point. *Econometrica* 61, 821–856.

Andrews, D.W.K. & R.C. Fair (1988) Inference in nonlinear econometric models with structural change. *Review of Economic Studies* 55, 615–640.

Andrews, D.W.K. & W. Ploberger (1994) Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* 62, 1383–1414.

Bai, J. & P. Perron (1998) Estimating and testing linear models with multiple structrual changes. *Econometrica* 66, 47–78.

Chen, X. & Y. Fan (1999) Consistent hypothesis testing in semiparmetric and nonparametric models for econometric time series. *Journal of Econometrics* 91, 373–402.

Fan, Y. & Q. Li (1997) A consistent nonparametric test for linearity of AR(p) models. *Economics Letters* 55, 53–59.

Fan, Y. & Q. Li (1999) Central limit theorem for degenerate U-statistics of absolutely regular processes with applications to model specification testing. *Journal of Nonparametric Statistics* 10, 245–271.

Ghysels, E., A. Guay, & A. Hall (1998) Predictive test for structural change with unknown breakpoint. *Journal of Econometrics* 82, 209–233.

Ghysels, E. & A. Hall (1990) A test for structural stability change of Euler conditions parameters estimated via the generalized method of moments estimator. *International Economic Review* 31, 355–364.

Hall, A. & A. Sen (1999a) Structural stability testing in models estimated via the generalized method of moments estimation. *Journal of Business and Economic Statistics* 17, 335–348.

Hall, A. & A. Sen (1999b) Two further aspects of some new tests for structural stability. *Structural Change and Economic Dynamics* 10, 431–443.

Hall, P. (1984) Central limit theorem for integrated square error of multivariate nonparametric density estimators. *Journal of Multivariate Analysis* 14, 1–16.

Hansen, B.E. (2000) Testing for structural change in conditional models. *Journal of Econometrics* 97, 93–115.

Hidalgo, J. (1995) A nonparametric conditional moment test for structural stability. *Econometric Theory* 11, 671–698.

Hjellvik, V., Q. Yao, & D. Tjøstheim (1998) Linearity testing using local polynomial approximation. *Journal of Statistical Planning and Inference* 68, 295–321.

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

Inoue, A. (2001) Testing for distributional change in time series. *Econometric Theory* 17, 156–187.

Kuan, C.M. (1998) Tests for changes in models with polynomial trends. *Journal of Econometrics* 84, 75–92.

Lavergne, P. & Q. Vuong (2000) Nonparametric significance testing. *Econometric Theory* 16, 576–601.

Li, Q. (1999) Consistent model specification tests for time series econometric models. *Journal of Econometrics* 92, 101–148.

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

Li, Q. & J.M. Wooldridge (2002) Semiparametric estimation of partially linear models for dependent data with generated regressors. *Econometric Theory* 18, 625–645.

Mankiw, N. & M. Shapiro (1986) Do we reject too often, small sample properties of tests of rational expectations models. *Economic Letters* 20, 139–145.

Maynard, A. & K. Shimotsu (2009) Covariance based orthogonality tests for regressors with unknown persistence. *Econometric Theory* 25, 63–116.

Perron, P. (1989) The great crash, the oil price shock and the unit root hypothesis. *Econometrica* 57, 1361–1401.

Ploberger, W. & W. Kramer (1996) A trend resistant test for structural change based on OLS residuals. *Journal of Econometrics* 70, 175–186.

Priestley, M.B. & M.T. Chao (1972) Non-parametric function fitting. *Journal of the Royal Statistical Society, Series B* 34, 386–92.

Robinson, P.M. (1989) Nonparametric estimation of time-varying parameters. In P. Hackl (ed.) *Statistical Analysis and Forecasting of Economic Structural Change.* Springer Verlag.

Sowell, F. (1996) Optimal tests for parameter instability in the generalized method of moments framework. *Econometrica* 64, 1085–1107.

Vogelsang, T.J. (1997) Wald-type tests for detecting breaks in the trend function of a dynamic time series. *Econometric Theory* 13, 818–849.

Vogelsang, T.J. (1998) Trend function hypothesis testing in the presence of serial correlation. *Econometrica* 66, 123–148.

Volkonskii, V.A. & Y.A. Rozanov (1961) Some limit theorems for random functions II. *Theory of Probability and Its Applications* 6, 186–198.

Wooldridge, J.M. (1994) Estimation and inference for dependent processes. In R. Engle & D. McFadden (eds.), *Handbook of Econometrics*, vol. 4. North Holland.

Yoshihara, K. (1976) Limiting behavior of U-statistics for stationary, absolutely regular processes. *Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete* 35, 237–252.

Zheng, J.X. (1996) A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics* 75, 263–290.

APPENDIX A. Proofs

We first define the following quantities:

$$M_1 = \max\{M_{11}, M_{12}\},$$

$$M_{11} = \sup_{s,s',t} \int \left| m_t^\top m_s m_t^\top m_{s'} \right|^{1+\delta} dF(z_t, z_s, z_{s'}),$$

$$M_{12} = \sup_{s,s',t} \int \left| m_t^\top m_s m_t^\top m_{s'} \right|^{1+\delta} dF(z_t, z_{s'}) dF(z_s),$$

$$M_2 = \max\{M_{21}, M_{22}\},$$

$$M_{21} = \sup_{s,t,t',r} \int \left| m_t^\top m_s m_{t'}^\top m_r \right|^{2(1+\delta)} dF(z_t, z_s, z_{t'}, z_r),$$

$$M_{22} = \sup_{s,t,t',r} \int \left| m_t^\top m_s m_{t'}^\top m_r \right|^{2(1+\delta)} dF(z_t, z_s) dF(z_{t'}, z_r),$$

$$M_3 = \max\{M_{31}, M_{32}\},$$

$$M_{31} = \sup_{r,r',s,s',t,t'} \int \left| m_t^\top m_s m_t^\top m_{s'} m_{t'}^\top m_r m_{t'}^\top m_{r'} \right|^{(1+\delta)} dF(z_t, z_s, z_{s'}, z_{t'}, z_r, z_{r'}),$$

$$M_{32} = \sup_{r,r',s,s',t,t'} \int \left| m_t^\top m_s m_t^\top m_{s'} m_{t'}^\top m_r m_{t'}^\top m_{r'} \right|^{(1+\delta)} dF(z_t, z_s, z_{s'}, z_{t'}, z_r) dF(z_{r'}),$$

$$M_4 = \sup_{t,s,t',s'} \mathbb{E} \left| m_t^\top m_s m_t^\top m_{s'} m_{t'}^\top m_s m_{t'}^\top m_{s'} \right|,$$

$$M_5 = \max\{M_{51}, M_{52}\},$$

$$M_{51} = \sup_{s,s',t,t'} \int \left| \nabla m_t^\top \nabla m_{t'} m_s m_{s'} \right|^{(1+\delta)} dF(z_s, z_{s'}, z_t, z_{t'}),$$

$$M_{52} = \sup_{s,s',t,t'} \int \left| \nabla m_t^\top \nabla m_{t'} m_s m_{s'} \right|^{(1+\delta)} dF(z_s, z_t, z_{t'}) dF(z_{s'}).$$

LEMMA A.1. *Given Assumptions 1–4, and 6–7,*

$$\frac{U_T}{s_T} \xrightarrow{d} N(0,1),$$

where

$$U_T = \frac{1}{Th^{1/2}} \sum_{t=2}^{T} \sum_{s=1}^{t-1} K_{ts} m_t^\top m_s$$

and $s_T^2 = \mathbb{E}(U_T^2)$.

Proof. Let

$$V_t = \frac{1}{Th^{1/2}} \sum_{s=1}^{t-1} K_{ts} m_t^\top m_s,$$

which is a martingale difference sequence. We need to show

$$s_T^{-2}\sum_{t=2}^{T}V_t^2 \xrightarrow{p} 1 \tag{A.1}$$

and

$$s_T^{-4}\sum_{t=2}^{T}\mathbb{E}\left(V_t^4\right) \to 0. \tag{A.2}$$

First,

$$s_T^2 = \frac{1}{T^2h}\sum_{t=2}^{T}\sum_{s=1}^{t-1}K_{ts}^2\mathbb{E}\left(m_t^\top m_s\right)^2 + \frac{1}{T^2h}\sum_{t=2}^{T}\sum^{t-1}\sum_{s\neq s'}^{t-1}K_{ts}K_{ts'}\mathbb{E}\left(m_t^\top m_s m_t^\top m_{s'}\right)$$
$$= B_1 + B_2.$$

We can write B_2 as

$$\frac{2}{T^2h}\sum_{t=3}^{T}\sum_{s'=2}^{t-1}\sum_{s=1}^{s'-1}K_{ts}K_{ts'}\mathbb{E}\left(m_t^\top m_s m_t^\top m_{s'}\right)$$

since we have $s < s' < t$. Suppose that $s' - s \geq t - s'$. Now by Lemma (A.2),

$$\left|\int m_t^\top m_s m_t^\top m_{s'} dF(m_t, m_s, m_{s'}) - \int m_t^\top m_s m_t^\top m_{s'} dF(m_t, m_{s'}) dF(m_s)\right|$$
$$\leq M_1^{1/(1+\delta)}\beta(s'-s)^{\delta/(1+\delta)},$$

Then we have

$$|B_2| \leq \frac{2}{T^2h}\sum_{s=1}^{T-2}\sum_{s'=s+1}^{T-1}\sum_{t=s'+1}^{s'+(s'-s)}K_{ts}K_{ts'}M_1^{1/(1+\delta)}\beta(s'-s)^{\delta/(1+\delta)}$$
$$\leq \frac{2}{T^2h}\sum_{s=1}^{T-2}\sum_{s'=s+1}^{T-1}\sum_{t=s'+1}^{s'+(s'-s)}K(0)^2M_1^{1/(1+\delta)}\beta(s'-s)^{\delta(1+\delta)},$$

since $K(0) > K_{ts}$. Then

$$|B_2| \leq \frac{2}{T^2h}\sum_{s=1}^{T-2}\sum_{s'=s+1}^{T-1}(s'-s)K(0)^2M_1^{1/(1+\delta)}\beta(s'-s)^{\delta/(1+\delta)}$$
$$\leq \frac{2}{T^2h}\sum_{s=1}^{T-2}\sum_{\tau=1}^{T-1}K(0)^2M_1^{1/(1+\delta)}\tau\beta(\tau)^{\delta/(1+\delta)}$$
$$= O\left((Th)^{-1}\right),$$

since $\sum_{\tau}^{\infty}\tau\beta(\tau)^{\delta/(1+\delta)} \leq \infty$. The case where $s' - s \leq t - s'$ is similar.

Now

$$\mathbb{E}\left(\sum_{t=2}^{T}V_t^2 - s_T^2\right)^2 = \mathbb{E}\left(\sum_{t=2}^{T}V_t^2\right)^2 - s_T^4,$$

so we consider

$$\begin{aligned}\mathbb{E}\left(\sum_{t=2}^{T} V_t^2\right)^2 &= \sum_{t=2}^{T} \mathrm{E}\left(V_t^4\right) + 2\sum\sum_{2\le t<t'\le T} \mathbb{E}\left(V_t^2 V_{t'}^2\right) \\ &= \frac{1}{T^4h^2}\sum_{t=1}^{T}\sum_{s=1}^{t-1}\sum_{s'=1}^{t-1}\sum_{r=1}^{t-1}\sum_{r'=1}^{t-1} K_{ts}K_{ts'}K_{tr}K_{tr'}\mathbb{E}\left(m_t^\top m_s m_t^\top m_{s'} m_t^\top m_r m_t^\top m_{r'}\right) \\ &\quad + \frac{2}{T^4h^2}\sum_{t'=2}^{T}\sum_{t=2}^{t'-1}\sum_{s=1}^{t-1}\sum_{s'=1}^{t-1}\sum_{r=1}^{t'-1}\sum_{r'=1}^{t'-1} K_{ts}K_{ts'}K_{t'r}K_{t'r'} \\ &\quad \times \mathbb{E}\left(m_t^\top m_s m_t^\top m_{s'} m_{t'}^\top m_r m_{t'}^\top m_{r'}\right) \\ &= C_1 + C_2.\end{aligned}$$

C_2 can be decomposed as

$$\begin{aligned}C_2 &= \frac{2}{T^4h^2}\sum_{t'=2}^{T}\sum_{t=2}^{t'-1}\sum_{s=1}^{t-1}\sum_{r=1}^{t'-1} K_{ts}^2K_{t'r}^2\mathbb{E}\left(m_t^\top m_s\right)^2\left(m_{t'}^\top m_r\right)^2 \\ &\quad + \frac{4}{T^4h^2}\sum_{t'=2}^{T}\sum_{t=2}^{t'-1}\sum_{s=1}^{t-1}\sum_{s'=1}^{t-1} K_{ts}K_{ts'}K_{t's}K_{t's'}\mathbb{E}\left(m_t^\top m_s m_t^\top m_{s'} m_{t'}^\top m_s m_{t'}^\top m_{s'}\right) \\ &\quad + \frac{2}{T^4h^2}\sum_{t'=2}^{T}\sum_{t=2}^{t'-1}\sum^{t-1}\sum^{t'-1}\sum_{s\neq r\neq r'}^{t'-1} K_{ts}^2K_{t'r}K_{t'r'}\mathbb{E}\left(\left(m_t^\top m_s\right)^2 m_{t'}^\top m_r m_{t'}^\top m_{r'}\right) \\ &\quad + \frac{2}{T^4h^2}\sum_{t'=2}^{T}\sum_{t=2}^{t'-1}\sum^{t-1}\sum^{t-1}\sum_{s\neq s'\neq r}^{t'-1} K_{ts}K_{ts'}K_{t'r}^2\mathbb{E}\left(m_t^\top m_s m_t^\top m_{s'}\left(m_{t'}^\top m_r\right)^2\right) \\ &\quad + \frac{8}{T^4h^2}\sum_{t'=2}^{T}\sum_{t=2}^{t'-1}\sum^{t-1}\sum^{t-1}\sum_{s\neq s'\neq r}^{t'-1} K_{ts}K_{ts'}K_{t's}K_{t'r}\mathbb{E}\left(m_t^\top m_s m_t^\top m_{s'} m_{t'}^\top m_s m_{t'}^\top m_r\right) \\ &\quad + \frac{2}{T^4h^2}\sum_{t'=2}^{T}\sum_{t=2}^{t'-1}\sum^{t-1}\sum^{t-1}\sum^{t'-1}\sum_{s\neq s'\neq r\neq r'}^{t'-1} K_{ts}K_{ts'}K_{t'r}K_{t'r'} \\ &\quad \times \mathbb{E}\left(m_t^\top m_s m_t^\top m_{s'} m_{t'}^\top m_r m_{t'}^\top m_{r'}\right) \\ &= C_{21} + C_{22} + C_{23} + C_{24} + C_{25}.\end{aligned}$$

We can write

$$\begin{aligned}s_T^4 &= \frac{1}{T^4h^2}\left(\sum_{t=2}^{T}\sum_{s=1}^{t-1} K_{ts}^2\mathbb{E}\left(m_t^\top m_s\right)^2\right)\left(\sum_{t'=2}^{T}\sum_{s'=1}^{t'-1} K_{t's'}^2\mathbb{E}\left(m_{t'}^\top m_{s'}\right)^2\right) \\ &= \frac{2}{T^4h^2}\sum_{t'=2}^{T}\sum_{t=2}^{t'-1}\sum_{s=1}^{t-1}\sum_{r=1}^{t'-1} K_{ts}^2K_{t'r}^2\mathbb{E}\left(m_t^\top m_s\right)^2\mathbb{E}\left(m_{t'}^\top m_r\right)^2.\end{aligned}$$

Then

$$C_{21}-s_T^4=\frac{2}{T^4h^2}\sum_{t'=2}^{T}\sum_{t=2}^{t'-1}\sum_{s=1}^{t-1}\sum_{r=1}^{t'-1}K_{ts}^2K_{t'r}^2\left(\mathbb{E}\left(m_t^\top m_s\right)^2\left(m_{t'}^\top m_r\right)^2-\mathbb{E}\left(m_t^\top m_s\right)^2\right.$$
$$\left.\times\,\mathbb{E}\left(m_{t'}^\top m_r\right)^2\right).$$

Suppose that $s<t<r<t'$. By Lemma A.2,

$$\left|\mathbb{E}\left(m_t^\top m_s\right)^2\left(m_{t'}^\top m_r\right)^2-\mathbb{E}\left(m_t^\top m_s\right)^2\mathbb{E}\left(m_{t'}^\top m_r\right)^2\right|\le M_2^{1/(1+\delta)}\beta(r-t)^{\delta/(1+\delta)}.$$

Then

$$\begin{aligned}\left|C_{21}-s_T^4\right|&\le\frac{2}{T^4h^2}\sum_{t'=2}^{T}\sum_{t=2}^{t'-1}\sum_{s=1}^{t-1}K_{ts}^2\sum_{r=1}^{t'-1}K_{t'r}^2M_2^{1/(1+\delta)}\beta(r-t)^{\delta/(1+\delta)}\\&\sim\frac{2}{T^4h^2}\sum_{t'=2}^{T}\sum_{t=2}^{t'-1}\sum_{s=1}^{t-1}K_{ts}^2\\&=O\left(\frac{1}{Th}\right),\end{aligned}$$

since

$$\sum_{t=2}^{t'-1}\sum_{s=1}^{t-1}K_{ts}^2\sim\int_0^T\int_0^t K\left(\frac{t-s}{Th}\right)ds\,dt=T^2h\int_0^1\int_0^{v/h}K(u)du\,dv=O\left(T^2h\right).$$

The cases $s<r<t<t'$ and $r<s<t<t'$ are similar.

Now we show that $C_{25}\to 0$. Without loss of generality, suppose that $s<s'$ and $r<r'$ so that $s<s'<t$ and $r<r'<t'$. Denote $s<s'<r<r'<t<t'$ as case 1 and consider the following subcases.

Case 1a: $s'-s\ge\max\{r-s',r'-r,t-r',t'-t\}$
Case 1b: $r-s'\ge\max\{s'-s,r'-r,t-r',t'-t\}$
Case 1c: $r'-r\ge\max\{s'-s,r-s',t-r',t'-t\}$
Case 1d: $t-r'\ge\max\{s'-s,r-s',r'-r,t'-t\}$
Case 1e: $t'-t\ge\max\{s'-s,r-s',r'-r,t-r'\}$.

For case 1a, we have

$$\begin{aligned}|C_{25}|&=\frac{1}{T^4h^2}\sum_{s=1}^{T=5}\sum_{s'=s+1}^{T-4}\sum_{r=s'+1}^{T-3}\sum_{r'=r+1}^{T-2}\sum_{t=r'+1}^{T-1}\sum_{t'=t+1}^{T}K_{ts}K_{ts'}K_{t'r}K_{t'r'}\\&\quad\times\left|\mathbb{E}\left(m_t^\top m_s m_t^\top m_{s'}m_{t'}^\top m_r m_{t'}^\top m_{r'}\right)\right|\\&\le\frac{1}{T^4h^2}\sum_{s=1}^{T=5}\sum_{s'=s+1}^{T-4}\sum_{r=s'+1}^{T-3}\sum_{r'=r+1}^{T-2}\sum_{t=r'+1}^{T-1}\sum_{t'=t+1}^{T}K_{ts}K_{ts'}K_{t'r}K_{t'r'}M_{T3}^{1/(1+\delta)}\\&\quad\times\beta\left(s'-s\right)^{\delta/(1+\delta)}\end{aligned}$$

$$\leq \frac{1}{T^4h^2}\sum_{s=1}^{T=5}\sum_{s'=s+1}^{T-4}\left(s'-s\right)^2 M_3^{1/(1+\delta)}\beta\left(s'-s\right)^{\delta/(1+\delta)}\sum_{r'=r+1}^{T-2}\sum_{t'=t+1}^{T}K_{t'r'}K(0)^3$$

$$= O\left(\frac{1}{Th}\right).$$

For case 1b,

$$|C_{25}| = \frac{1}{T^4h^2}\sum_{s=1}^{T=5}\sum_{s'=s+1}^{T-4}\sum_{r=s'+1}^{T-3}\sum_{r'=r+1}^{T-2}\sum_{t=r'+1}^{T-1}\sum_{t'=t+1}^{T}K_{ts}K_{ts'}K_{t'r}K_{t'r'}$$
$$\times \left|\mathbb{E}\left(m_t^\top m_s m_t^\top m_{s'} m_{t'}^\top m_r m_{t'}^\top m_{r'}\right)\right|$$
$$\leq \frac{1}{T^4h^2}\sum_{s=1}^{T=5}\sum_{s'=s+1}^{T-4}\sum_{r=s'+1}^{T-3}\sum_{r'=r+1}^{T-2}\sum_{t=r'+1}^{T-1}\sum_{t'=t+1}^{T}K_{ts}K_{ts'}K_{t'r}K_{t'r'}M_3^{1/(1+\delta)}$$
$$\times \beta\left(r-s'\right)^{\delta/(1+\delta)}$$
$$\leq \frac{1}{T^4h^2}\sum_{s=1}^{T=5}\sum_{s'=s+1}^{T-4}\sum_{r=s'+1}^{T-3}\left(r-s'\right)^2 M_3^{1/(1+\delta)}\beta\left(r-s'\right)^{\delta/(1+\delta)}\sum_{t=1}^{T}K_{ts'}K(0)^3$$
$$= O\left(\frac{1}{Th}\right).$$

For case 1c,

$$|C_{25}| = \frac{1}{T^4h^2}\sum_{t'=6}^{T}\sum_{t=5}^{t'-1}\sum_{r'=4}^{t-1}\sum_{r=3}^{r'-1}\sum_{s'=2}^{r-1}\sum_{s=1}^{s'-1}K_{ts}K_{ts'}K_{t'r}K_{t'r'}$$
$$\times \left|\mathbb{E}\left(m_t^\top m_s m_t^\top m_{s'} m_{t'}^\top m_r m_{t'}^\top m_{r'}\right)\right|$$
$$\leq \frac{1}{T^4h^2}\sum_{t'=6}^{T}\sum_{t=5}^{t'-1}\sum_{r'=4}^{t-1}\sum_{r=3}^{r'-1}\sum_{s'=2}^{r-1}\sum_{s=1}^{s'-1}K_{ts}K_{ts'}K_{t'r}K_{t'r'}M_3^{1/(1+\delta)}\beta\left(r'-r\right)^{\delta/(1+\delta)}$$
$$\leq \frac{1}{T^4h^2}\sum_{t'=6}^{T}\sum_{t=5}^{t'-1}\sum_{r'=4}^{t-1}\sum_{r=3}^{r'-1}K(0)^3K_{t'r}M_3^{1/(1+\delta)}\left(r'-r\right)^2\beta\left(r'-r\right)^{\delta/(1+\delta)}$$
$$= O\left(\frac{1}{Th}\right).$$

For case 1d,

$$|C_{25}| = \frac{1}{T^4h^2}\sum_{t'=6}^{T}\sum_{t=5}^{t'-1}\sum_{r'=4}^{t-1}\sum_{r=3}^{r'-1}\sum_{s'=2}^{r-1}\sum_{s=1}^{s'-1}K_{ts}K_{ts'}K_{t'r}K_{t'r'}$$
$$\times \left|\mathbb{E}\left(m_t^\top m_s m_t^\top m_{s'} m_{t'}^\top m_r m_{t'}^\top m_{r'}\right)\right|$$
$$\leq \frac{1}{T^4h^2}\sum_{t'=6}^{T}\sum_{t=5}^{t'-1}\sum_{r'=4}^{t-1}\sum_{r=3}^{r'-1}\sum_{s'=2}^{r-1}\sum_{s=1}^{s'-1}K_{ts}K_{ts'}K_{t'r}K_{t'r'}M_3^{1/1+\delta}\beta\left(t-r'\right)^{\delta/(1+\delta)}$$

$$\leq \frac{1}{T^4h^2}\sum_{t'=6}^{T}\sum_{t=5}^{t'-1}\sum_{r'=4}^{t-1} M_3^{1/(1+\delta)}\left(t-r'\right)^2\beta\left(t-r'\right)^{\delta/(1+\delta)}\sum_{r=1}^{T} K_{t'r}K(0)^3$$
$$= O\left(\frac{1}{Th}\right).$$

For case 1e, we consider the second-order statistic of $\{s'-s, r-s', r'-r, t-r'\}$ and repeat the arguments for cases 1a–1d. Hence, we have shown that $C_{25}\to 0$.

The term C_1 can be decomposed as

$$C_1 = \frac{3}{T^4h^2}\sum_{t=1}^{T}\sum_{s=1}^{t-1}\sum_{s'=1}^{t-1} K_{ts}^2K_{ts'}^2\mathbb{E}\left(m_t^\top m_s\right)^2\left(m_t^\top m_{s'}\right)^2$$
$$+\frac{6}{T^4h^2}\sum_{t=1}^{T}\sum_{s=1}^{t-1}\sum^{t-1}\sum_{r\neq r'}^{t-1} K_{ts}^2K_{tr}K_{tr'}\mathbb{E}\left(\left(m_t^\top m_s\right)^2 m_t^\top m_r m_t^\top m_{r'}\right)$$
$$+\frac{1}{T^4h^2}\sum_{t=1}^{T}\sum^{t-1}\sum^{t-1}\sum^{t-1}\sum_{s\neq s'\neq r\neq r'=1}^{t-1} K_{ts}K_{ts'}K_{tr}K_{tr'}\mathbb{E}\left(m_t^\top m_s m_t^\top m_{s'} m_t^\top m_r m_t^\top m_{r'}\right)$$
$$= C_{11}+C_{12}+C_{13}.$$

Following the method used for C_2, we have $C_{11}=O(T^{-1})$, $C_{12}=O(T^{-1})$, and $C_{13}=O((Th)^{-2})$. In a similar manner, C_{23} and C_{24} are at most $O((Th)^{-1})$. Finally,

$$|C_{22}| \leq \frac{4}{T^4h^2}\sum_{t'=2}^{T}\sum_{t=2}^{t'-1}\sum^{t-1}\sum_{s\neq s'}^{t-1} K_{ts}K_{ts'}K_{t's}K_{t's'}\mathbb{E}\left|m_t^\top m_s m_t^\top m_{s'} m_{t'}^\top m_s m_{t'}^\top m_{s'}\right|$$
$$\leq \frac{4}{T^4h^2}\sum_{t'=2}^{T}\sum_{t=2}^{t'-1}\sum^{t-1}\sum_{s\neq s'}^{t-1} K_{ts}K_{ts'}K_{t's}K_{t's'}M_4$$
$$= O\left(h^2\right).$$

Combining these results, we have $\mathbb{E}\left(\sum_{t=2}^{T}V_t^2 - s_T^2\right)^2 = O(h^2)+O((Th)^{-1}) = o(1)$ so that (A.1) holds. Since $s_T^2 = O(1)$ and $C_1 = O(T^{-1})$, (A.2) holds. ■

Proof of Theorem 3.1.

$$\hat{U}_T = \frac{1}{Th^{1/2}}\sum_{t=2}^{T}\sum_{s=1}^{t-1} K_{ts}\hat{m}_t^\top\hat{m}_s$$
$$= \frac{1}{Th^{1/2}}\sum_{t=2}^{T}\sum_{s=1}^{t-1} K_{ts}\left(m_t+(\hat{m}_t-m_t)\right)^\top\left(m_s+(\hat{m}_s-m_s)\right)$$
$$= \frac{1}{Th^{1/2}}\sum_{t=2}^{T}\sum_{s=1}^{t-1} K_{ts}m_t^\top m_s + \frac{1}{Th^{1/2}}\sum_{t=2}^{T}\sum_{s=1}^{t-1} K_{ts}m_t^\top(\hat{m}_s-m_s)$$
$$+\frac{1}{Th^{1/2}}\sum_{t=2}^{T}\sum_{s=1}^{t-1} K_{ts}(\hat{m}_t-m_t)^\top m_s + \frac{1}{Th^{1/2}}\sum_{t=2}^{T}\sum_{s=1}^{t-1} K_{ts}(\hat{m}_t-m_t)^\top(\hat{m}_s-m_s)$$
$$= F_1+F_2+F_3+F_4.$$

Consider F_3. Suppose that $L = 1$ so that m_t is 1×1. Then we can write

$$\hat{m}_t = m_t + \nabla m_t(\theta_0)\left(\hat{\theta}_T - \theta_0\right) + \left(\hat{\theta}_T - \theta_0\right)^\top \nabla^2 m_t(\bar{\theta})\left(\hat{\theta}_T - \theta_0\right),$$

where $\bar{\theta}$ lies between $\hat{\theta}_T$ and θ_0. Then F_3 can be written as

$$\begin{aligned}
F_3 &= \frac{1}{Th^{1/2}} \sum_{t=2}^{T} \sum_{s=1}^{t-1} K_{ts} \nabla m_t(\theta_0) m_s \left(\hat{\theta}_T - \theta_0\right) \\
&\quad + \frac{1}{Th^{1/2}} \sum_{t=2}^{T} \sum_{s=1}^{t-1} K_{ts} \left(\hat{\theta}_T - \theta_0\right)^\top \nabla^2 m_t(\bar{\theta}) m_s \left(\hat{\theta}_T - \theta_0\right) \\
&= F_{31}\left(\hat{\theta}_T - \theta_0\right) + \left(\hat{\theta}_T - \theta_0\right)^\top F_{32}\left(\hat{\theta}_T - \theta_0\right).
\end{aligned}$$

Now

$$\begin{aligned}
\mathbb{E}\left(F_{31}^\top F_{31}\right) &= \frac{1}{T^2 h} \sum_{t=2}^{T} \sum_{s=1}^{t-1} \sum_{t'=2}^{T} \sum_{s'=1}^{t'-1} K_{ts} K_{t's'} \mathbb{E}\left(\nabla m_t^\top(\theta_0) \nabla m_{t'}(\theta_0) m_s m_{s'}\right) \\
&= \frac{1}{T^2 h} \sum_{t=2}^{T} \sum_{s=1}^{t-1} K_{ts}^2 \mathbb{E}\left(\nabla m_t^\top(\theta_0) \nabla m_t(\theta_0) m_s^2\right) \\
&\quad + \frac{2}{T^2 h} \sum_{t=2}^{T} \sum_{s=2}^{t-1} \sum_{s'=1}^{s-1} K_{ts} K_{ts'} \mathbb{E}\left(\nabla m_t^\top(\theta_0) \nabla m_t(\theta_0) m_s m_{s'}\right) \\
&\quad + \frac{2}{T^2 h} \sum_{t=3}^{T} \sum_{t'=2}^{t-1} \sum_{s=1}^{t'-1} K_{ts} K_{t's} \mathbb{E}\left(\nabla m_t^\top(\theta_0) \nabla m_{t'}(\theta_0) m_s^2\right) \\
&\quad + \frac{1}{T^2 h} \sum_{t=2}^{T} \sum_{t' \neq t}^{T} \sum_{s=1}^{t-1} \sum_{s' \neq s}^{t'-1} K_{ts} K_{t's'} \mathbb{E}\left(\nabla m_t^\top(\theta_0) \nabla m_{t'}(\theta_0) m_s m_{s'}\right) \\
&= F_{311} + F_{312} + F_{313} + F_{314}.
\end{aligned}$$

For the terms F_{314}, suppose that $s < t < s' < t'$ and that $s' - t \geq \max\{t - s, t' - s'\}$. Then

$$\begin{aligned}
F_{314} &\leq \frac{1}{T^2 h} \sum_{t'=4}^{T} \sum_{s'=3}^{t'-1} \sum_{t=2}^{s'-1} \sum_{s=1}^{t-1} K_{ts} K_{t's'} \beta\left(s' - t\right)^{\delta/(1+\delta)} M_5 \\
&\leq \frac{1}{T^2 h} \sum_{t'=4}^{T} \sum_{s'=3}^{t'-1} \sum_{t=2}^{s'-1} K(0)^2 \left(s' - t\right) \beta\left(s' - t\right)^{\delta/(1+\delta)} M_5 \\
&= O\left(h^{-1}\right),
\end{aligned}$$

the other cases being similar so that F_{314} is at most $O(1)$. By Cauchy Schwarz, F_{311} is $O(1)$, and F_{312} and F_{313} are $O(Th)$. These results imply that F_{31} is $O_p(T^{1/2}h^{1/2}) + O_p(h^{-1/2})$. Now

$$\begin{aligned}
\mathbb{E}|F_{32}| &\leq \frac{1}{Th^{1/2}} \sum_{t=2}^{T} \sum_{s=1}^{t-1} K_{ts} \mathbb{E}|M_g(z_t)| \\
&= O\left(Th^{1/2}\right),
\end{aligned}$$

so that $F_3 = O_p(h^{1/2}) + O_p((Th)^{-1/2}) = o_p(1)$. The proof is identical for F_2.

For F_4, we follow Fan and Li (1999) and note that

$$\left\|\hat{m}_t - m_t\right\| \le dM_g(z_t)\left\|\hat{\theta}_T - \theta_0\right\|$$

for some generic constant d so that

$$|F_4| \le \frac{1}{Th^{1/2}} \sum_{t=2}^{T} \sum_{s=1}^{t-1} K_{ts} d^2 M_g(z_t) M_g(z_s) \left\|\hat{\theta}_T - \theta_0\right\|^2$$
$$= O_p\left(h^{1/2}\right) = o_p(1).$$

Note that F_1 is normally distributed by Lemma A.1 and it is easy to show that the variance is consistently estimated using $\hat{\sigma}^2$. ■

Proof of Theorem 4.1. The proof follows the steps of the proof of Theorem (3.1) in that $\hat{m}_t$ can be replaced by m_t. However, we have to now deal with the dependence of the moment conditions. We write

$$\tilde{U}_T = \frac{1}{Th^{1/2}} \sum_{t=2}^{T} \sum_{s=1}^{t-Th_2} K_{ts} m_t^\top m_s.$$

The summands are no longer a martingale difference since $\mathbb{E}(m_t^\top m_s | \mathcal{F}_{-\infty}^{t-1}) \neq 0$ where $\mathcal{F}_{-\infty}^{t-1} = \sigma(m_{t-1}, m_{t-2}, \ldots)$. We construct martingale differences in the following way. Let

$$V_t = \frac{1}{Th} \sum_{s=1}^{t-1} K_{ts} m_t^\top m_s$$
$$= \frac{1}{Th} \sum_{s=1}^{t-1} V_{ts}$$
$$U_T^* = \sum_{t=2}^{T} \left(V_t - \mathbb{E}\left(V_t | \mathcal{F}_{-\infty}^{t-1}\right)\right)$$
$$= \sum_{t=2}^{T} V_t^*$$
$$= \sum_{t=2}^{T} \sum_{s=1}^{t-1} V_{ts}^*,$$

so that the summands U_T^* are martingale differences since $\mathbb{E}\left(V_t^* | \mathcal{F}_{-\infty}^{t-1}\right) = 0$ by construction. Hence, we can apply Lemma A.1 to U_T^*. Moreover, we have

$$U_T^* = \tilde{U}_T + \frac{1}{Th^{1/2}} \sum_{t=2}^{T} \sum_{s=t-Th_2+1}^{t-1} V_{ts}^*$$
$$- \frac{1}{Th^{1/2}} \sum_{t=2}^{T} \sum_{s=1}^{t-Th_2} K_{ts} \mathbb{E}\left(m_t^\top m_s \middle| \mathcal{F}_{-\infty}^{t-1}\right)$$
$$= \tilde{U}_T + U_{1T}^* - U_{2T}^*.$$

First, we have

$$\mathbb{E}\left[\left(U_{1T}^{*}\right)^{2}\right]=\frac{1}{T^{2}h}\mathbb{E}\left[\sum_{t=2}^{T}\sum_{s=t-Th_{2}+1}^{t-1}\sum_{t'=2}^{T}\sum_{s'=t'-Th_{2}+1}^{t'-1}V_{ts}^{*}V_{t's'}^{*}\right].$$

Just as in the proof of Lemma A.1, the dominant term is of the form

$$\begin{aligned}G_{1}&=\frac{1}{T^{2}h}\sum_{t=2}^{T}\sum_{s=t-Th_{2}+1}^{t-1}\mathbb{E}\left[\left(V_{ts}^{*}\right)^{2}\right]\\&=\frac{1}{T^{2}h}\sum_{t=2}^{T}\sum_{s=t-Th_{2}+1}^{t-1}K_{ts}^{2}\mathbb{E}\left[m_{t}^{\top}m_{s}-\mathbb{E}\left(m_{t}^{\top}m_{s}\,\middle|\,\mathcal{F}_{-\infty}^{t-1}\right)\right]^{2}\\&=\frac{1}{T^{2}h}O\left(T^{2}h_{2}\right),\end{aligned}$$

which goes to zero since $h_2/h\to 0$, so that $U_{1T}^{*}\xrightarrow{p}0$.

Now for U_{2T}^{*}, we have

$$\begin{aligned}\mathbb{E}\left|U_{2T}^{*}\right|&\le\frac{1}{Th^{1/2}}\sum_{t=2}^{T}\sum_{s=1}^{t-Th_{2}}\mathbb{E}\left|K_{ts}\mathbb{E}\left(m_{t}^{\top}m_{s}\,\middle|\,\mathcal{F}_{-\infty}^{t-1}\right)\right|\\&=\frac{1}{Th^{1/2}}\sum_{t=2}^{T}\sum_{s=1}^{t-Th_{2}}K_{ts}\mathbb{E}\left|\mathbb{E}\left(m_{t}^{\top}m_{s}\,\middle|\,\mathcal{F}_{-\infty}^{t-1}\right)-\mathbb{E}\left(m_{t}^{\top}m_{s}\right)+\mathbb{E}\left(m_{t}^{\top}m_{s}\right)\right|\\&\le\frac{1}{Th^{1/2}}\sum_{t=2}^{T}\sum_{s=1}^{t-Th_{2}}K_{ts}\left[\mathbb{E}\left|\mathbb{E}\left(m_{t}^{\top}m_{s}\,\middle|\,\mathcal{F}_{-\infty}^{t-1}\right)-\mathbb{E}\left(m_{t}^{\top}m_{s}\right)\right|+\left|\mathbb{E}\left(m_{t}^{\top}m_{s}\right)\right|\right]\\&\le\frac{1}{Th^{1/2}}\sum_{t=2}^{T}\sum_{s=1}^{t-Th_{2}}K_{ts}M\beta(t-s)^{\delta/(1+\delta)}\\&=\frac{1}{h^{1/2}}O\left(\beta(Th_{2})^{\delta/(1+\delta)}\right),\end{aligned}$$

where the last inequality comes from Lemma A.0 of Fan and Li (1999) and an application of Lemma 1 of Yoshihara (1976). By Assumption 1′, $\sum_{\tau=1}^{\infty}\tau^{2}\beta(\tau)^{\delta/(1+\delta)}<\infty$, which implies that $\beta(Th_2)^{\delta/(1+\delta)}=O\left[(Th_2)^{-2-\epsilon}\right]$ for some $\epsilon>0$. Hence $\mathbb{E}|U_{2T}|=O(h^{-1/2}(Th_2)^{-2-\epsilon})$, which is $o(1)$ since $T^4h_2^5\to\infty$ and $h_2/h\to 0$. ■

LEMMA A.2. *Let $x_{t_1},x_{t_2},\ldots,x_{t_k}$ (with $t_1<t_2<\cdots<t_k$) be absolutely regular random vectors with mixing coefficients β. Let $h(x_{t_1},x_{t_2},\ldots,x_{t_k})$ be a Borel measurable function, and let there be a $\delta>0$ such that*

$$P=\max\{P_1,P_2\}<\infty,$$

where

$$P_{1}=\sup_{t_{1},t_{2},\ldots,t_{k}}\int\left|h(x_{t_{1}},x_{t_{2}},\ldots,x_{t_{k}})\right|^{1+\delta}dF(x_{t_{1}},x_{t_{2}},\ldots,x_{t_{k}}),$$

$$P_{2}=\sup_{t_{1},t_{2},\ldots,t_{k}}\int\left|h(x_{t_{1}},x_{t_{2}},\ldots,x_{t_{k}})\right|^{1+\delta}dF(x_{t_{1}},\ldots,x_{t_{j}})dF(x_{t_{j+1}},\ldots,x_{t_{k}}).$$

Then

$$\left|\int h(x_{t_1},x_{t_2},\ldots,x_{t_k})dF(x_{t_1},x_{t_2},\ldots,x_{t_k}) - h(x_{t_1},x_{t_2},\ldots,x_{t_k})dF(x_{t_1},\ldots,x_{t_j})dF(x_{t_{j+1}},\ldots,x_{t_k})\right| \leq 4P^{1/(1+\delta)}\beta_\tau^{\delta/(1+\delta)}$$

for all $\tau = t_{j+1} - t_j$.

Proof. The proof is similar to Lemma 1 in Yoshihara (1976) with the exception that we use the definition of absolute regularity given in Assumption 2, which applies to observations with heterogeneous distributions. See Lemma 4.1 in Volkonskii and Rozanov (1961). ■