Estimation and model specification testing in nonparametric and semiparametric econometric models

Gao, Jiti and King, Maxwell

The University of Adelaide, Monash University

March 2003

Estimation and Model Specification Testing in Nonparametric and Semiparametric Econometric Models

Jiti Gao[1] and Maxwell King

School of Mathematics and Statistics, The University of Western Australia, Crawley WA 6009, Australia; and Faculty of Business and Economics, Monash University, Clayton VIC 3800, Australia

Abstract. This paper considers two classes of semiparametric nonlinear regression models, in which nonlinear components are introduced to reflect the nonlinear fluctuation in the mean. A general estimation and testing procedure for nonparametric time series regression under the α–mixing condition is introduced. Several test statistics for testing nonparametric significance, linearity and additivity in nonparametric and semiparametric time series econometric models are then constructed. The proposed test statistics are shown to have asymptotic normal distributions under their respective null hypotheses. Moreover, the proposed testing procedures are illustrated by several simulated examples. In addition, one of the proposed testing procedures is applied to a continuous-time model and implemented through a set of the US Federal interest rate data. Our research suggests that it is unreasonable to assume the linearity in the drift for the given data as required by some existing studies.

JEL classification: Primary C52; Secondary C14

Keywords: Estimation; Model specification; Semiparametric Error Correction Model; Stochastic Process.

1. Introduction and motivation

The problem of estimating nonlinear econometric models has gained much attention in recent years. This is mainly due to the recent development in nonparametric and semiparametric econometrics. See Pagan and Ullah (1999) for a recent survey up to 1999. Due to the curse of dimensionality, however, nonparametric multivariate smoothing techniques are in practice not very useful when there are more than two or three predictor variables [see Chapter 7 of Fan and Gijbels (1996)]. In recent years, nonparametric and semiparametric approaches have been proposed to deal with the curse of dimensionality problem and some related problems as well. These include the construction of consistent model specification tests and additive nonparametric and semiparametric regression modelling. For the case of model specification tests, interest focuses on tests for a parametric model versus a nonparametric model, tests for a semiparametric (partially linear or single-index) model against a nonparametric model, and tests for the significance of a subset of regressors. For example, Härdle and Mammen (1993) have developed consistent tests for a parametric specification

[1] *Correspondence to*: Professor Jiti Gao, School of Mathematics and Statistics, The University of Western Australia, Crawley WA 6009, Australia. Tel: 618 9380 3354; Fax: 618 9380 1028; Email: jiti.gao@uwa.edu.au.

by employing the kernel regression estimation technique; Hong and White (1995), Gao and Liang (1997), Gao and Shi (1997), and Gao, Tong and Wolff (2002a, 2002b) have applied the method of series estimation to consistent testing for a parametric regression model; Shively, Kohn and Ansley (1994) consider testing for linearity in a semiparametric regression model based on the stochastic interpretation of spline smoothing given in Wahba (1978); Wooldridge (1992), Yatchew (1992), Gozalo (1993), Samarov (1993), Whang and Andrews (1993), Horowitz and Härdle (1994), Fan and Li (1996), Lavergne and Vuong (1996), Zheng (1996), Li (1999), and Lavergne and Vuong (2000) have developed consistent tests for a semiparametric model (partially linear or single-index) versus a nonparametric alternative for either the independent and identically distributed (i.i.d.) case or the time series case. Other related references include Robinson (1989), Eubank and Spiegelman (1990), Eubank and Hart (1992), Hjellvik and Tjøstheim (1995), Jayasuriva (1996), Kreiss, Neumann and Yao (1997), Hjellvik, Yao and Tjøstheim (1998), Li and Hsiao (1998), Li and Wang (1998), Härdle and Kneip (1999), and Härdle, Liang and Gao (2000). More recently, Aït-Sahalia, Bickel and Stoker (2001), Chen, Härdle and Li (2001), Fan and Huang (2001), Fan, Zhang and Zhang (2001), Gozalo and Linton (2001), Horowitz and Spokoiny (2001), and Lavergne (2001) further consider nonparametric and semiparametric test problems for either the i.i.d. case or the fixed design case.

For the case of additive nonparametric and semiparametric modelling, Fan, Härdle and Mammen (1998) have provided an efficient and direct way to deal with the dimensionality reduction problem. In practice, however, before applying the additive nonparametric regression technique to model real sets of data, a crucial problem is whether an additive nonparametric regression model is appropriate for a given set of data. In other words, we should test for nonparametric additivity before using an additive nonparametric regression to model a given set of data. When an additive nonparametric regression model is not appropriate for a given set of data, one needs to find alternative methods to solve the dimensionality reduction problem. As an alternative, one suggests using the additive partially linear regression to deal with the dimensionality reduction problem. In theory, one can assume that the process (Y_t, X_t) satisfies the following model

$$Y_t = E[Y_t|X_t] + e_t = m(X_t) + e_t = U_t^\tau \beta + g(V_t) + e_t, \tag{1.1}$$

where $X_t = (U_t^\tau, V_t^\tau)^\tau$, $m(X_t) = E[Y_t|X_t]$, and $e_t = Y_t - E[Y_t|X_t]$ is the error process and allowed to depend on X_t. In model (1.1), U_t and V_t are allowed to be two different time series. For example, U_t could be a vector of endogenous time series while V_t could be a vector of exogenous time series. In practice, a crucial problem is how to identify U_t and V_t before applying model (1.1) to model real sets of data. For some cases, the identification problem can be solved easily by using empirical studies. For example, when modelling electricity sales, it is natural to assume the impact of temperature on electricity consumption to be nonlinear, as both high and low temperatures lead to increased consumption, whereas a linear relationship may be assumed for other regressors. See Engle, Granger, Rice and Weiss (1986). Similarly, when modelling the dependence of earnings on qualification and labour market experience variables, our research [see Härdle, Liang and Gao (2000)] shows that the impact of qualification on earnings to be linear, while the dependence of earnings on labour market experience appears to be nonlinear. For many other cases, however, the identification problem should be solved theoretically before using model (1.1). More recently, Härdle, Liang and Gao (2000, §6.2) have extended the discussion of Chen and Chen (1991) for the i.i.d. case to the time series case, and therefore the identification problem for both the i.i.d. case and the time series case has been solved.

A selective review of the recent development of model (1.1) can be found in Härdle, Liang and Gao (2000).

We now consider a new class of partially linear models of the form

$$Y_t = X_t^\tau \beta + g(X_t) + e_t, \tag{1.2}$$

where $X_t = (X_{t1}, \cdots, X_{tp})^\tau$ is a vector of time series, $\beta = (\beta_1, \cdots, \beta_p)^\tau$ is a vector of unknown parameters, $g(\cdot)$ is an unknown function and can be viewed as a misspecification error, and e_t may be interpreted as a measurement error. In model (1.2), the error process e_t is allowed to depend on X_t. Obviously, model (1.2) cannot be viewed as a special form of model (1.1).

The main motivation for systematically studying model (1.2) is that the partially linear regression model (1.2) can play a significant role in modelling some nonlinear problems, although the linear regression normally fails to appropriately model nonlinear phenomena. We therefore suggest using the semiparametric partially linear regression (1.2) to model nonlinear phenomena, and then determine whether the nonlinearity is significant for a given data set (X_t, Y_t). In addition, some special cases of model (1.2) have already been considered by econometricians. In Section 2 below, one can see that some special forms of model (1.2) have already been used to model economic and financial data.

This paper then considers some estimation and model specification testing procedures for models (1.1) and (1.2), in particular, the model specification testing for the nonparametric component involved in models (1.1) and (1.2), as one needs to determine whether the nonlinear component is significant before applying either model (1.1) or (1.2) to fit a given set of data. For example, before using a stochastic differential equation to model a given financial data, one needs to determine whether the linearity in the drift is appropriate for the given financial data. This is particularly important as pointed out by some authors [see Aït-Sahalia (1996a); Ahn and Gao (1999)], the linearity of the drift imposed in the literature appears to be the main source of misspecification.

The rest of the paper is organised as follows. Section 2 presents some important examples. Section 3 discusses estimation and model specification testing procedures for models (1.1) and

(1.2). Examples of implementation and applications of the procedures to economic models and financial data are given in Section 4. Section 4 further considers some extensions and generalizations. Mathematical details are relegated to Appendices A–C.

2. Examples and models

Before proposing our estimation and model specification testing procedures, we present some interesting examples and models, which are either special forms or extended forms of models (1.1) and (1.2).

2.1. *Special and extended forms of model (1.1)*

Example 2.1 (Partially linear autoregressive models): Let $u_1, u_2, \ldots$ be an endogenous time series, $Y_t = u_t$, $U_t = (u_{t-1}, \ldots, u_{t-r})^\tau$, and $V_t = (v_{t1}, \ldots, v_{tq})^\tau$ be a vector of exogenous time series. Now model (1.1) is a partially linear autoregressive model of the form

$$u_t = \sum_{i=1}^{r} \beta_i u_{t-i} + g(v_{t1}, \ldots, v_{tq}) + e_t. \tag{2.1}$$

Example 2.2 (Partially nonlinear autoregressive models): Let $v_1, v_2, \ldots$ be an endogenous time series, $Y_t = v_t$, $V_t = (v_{t-1}, \ldots, v_{t-q})^\tau$, and $U_t = (u_{t1}, \ldots, u_{tr})^\tau$ be a vector of exogenous variables. Then model (1.1) is a partially nonlinear autoregressive model of the form

$$v_t = \sum_{i=1}^{r} \alpha_i u_{ti} + g(v_{t-1}, \ldots, v_{t-q}) + e_t. \tag{2.2}$$

Some estimation results for models (2.1) and (2.2) can be found from the literature. See for example, Robinson (1988), Teräsvirta, Tjøstheim and Granger (1994), Gao and Liang (1995), Gao (1998), Li and Hsiao (1998), Härdle, Liang and Gao (2000), and Gao, Tong and Wolff (2002a, 2002b).

In recent years, some other semiparametric regression models have also been discussed. We now review two related models, which are given in Examples 2.3–2.4.

Example 2.3: Consider a linear regression with a nonparametric error model of the form

$$Y_t = X_t^\tau \beta + u_t, \;\; u_t = g(u_{t-1}) + \epsilon_t, \tag{2.3}$$

where X_t and β are p-dimensional column vectors, X_t is stationary with finite second moments, Y_t and u_t are scalars, $g(\cdot)$ is an unknown function, possibly nonlinear, and is such that u_t is at least stationary with zero mean and finite variance i.i.d. innovations ϵ_t.

Model (2.3) was proposed by Hidalgo (1992) and then estimated by a kernel based procedure.

Example 2.4: Consider a nonparametric regression with an AR(1) error model of the form

$$Y_t = g(X_t) + u_t, \ u_t = \theta u_{t-1} + \epsilon_t, \tag{2.4}$$

where (X_t, Y_t) is a bivariate stationary time series, θ, satisfying $|\theta| < 1$, is an unknown parameter, $g(\cdot)$ is an unknown function, and ϵ_t is i.i.d. with zero mean and finite variance $0 < \sigma^2 < \infty$.

Truong and Stone (1994) proposed an estimation procedure for model (2.4).

In addition to the application of models (2.1)–(2.4) in economics and finance, one can extend models (2.1)–(2.4) to derive some very useful models such as semiparametric error correction models (SECMs). Before our derivation, we review a class of parametric error correction models (PECMs).

Parametric error correction models have been discussed extensively in the literature. See for example, Phillips and Loretan (1991), Mills (1993, §6.5–§6.6), and van Dijk and Franses (2000). We now consider one general parametric error correction model discussed in Phillips and Loretan (1991).

Example 2.5: Consider a parametric error correction model of the form

$$Y_t = X_t^\tau \beta + \sum_{s=1}^{q} \gamma_s^\tau \bigtriangledown X_{t-s} + \epsilon_t, \tag{2.5}$$

where $X_t = (X_{t1}, \ldots, X_{tp})^\tau$ is a vector of endogenous time series, $\beta = (\beta_1, \ldots, \beta_p)^\tau$ is a vector of unknown parameters, $\bigtriangledown X_t = X_t - X_{t-1}$, γ_s is another vector of unknown parameters, and ϵ_t is a white noise error term.

Model (2.5), as discussed in Mills (1993), can be used to model financial relations, such as the relationship between equity prices, dividends and gilt yields. In practical applications, however, whether the dependence of Y_t on $\bigtriangledown X_t$ is linear cannot be known for some data sets, in particular, the financial data. Therefore, one would suggest using some nonparametric models in practice and let the data speak for themselves.

For convenience, we first consider the case of $p = 1$ and introduce the following notation.

$$u_t = Y_t - X_t\beta, \ v_{ts} = \bigtriangledown X_{t-s}, \ \text{and } v_t = (v_{t1}, \ldots, v_{tq})^\tau.$$

Example 2.6 (Semiparametric error correction model): If the pair (u_t, v_t) satisfies model (2.1), then a semiparametric form of model (2.5) can be written as

$$Y_t = X_t\beta + u_t, \ u_t = g(v_{t1}, \ldots, v_{tq}) + \epsilon_t. \tag{2.6}$$

As can be seen from model (2.6), in order to determine whether leads of $\bigtriangledown X_t$ are being included in model (2.6), it suffices to test whether the null hypothesis $H_0: \ g = 0$ holds.

For model (2.6), the absence of the nonlinearity in $\bigtriangledown X_t$ implies that the error process $u_t = Y_t - X_t\beta$ is just white noise. Obviously, model (2.7) without nonlinearity is much simpler in theory and may be more applicable in practice. Thus, one would suggest determining whether the simpler model is appropriate before applying model (2.6) to actual data.

As in Phillips and Loretan (1991), X_t in (2.5) is a vector of endogenous time series, one needs to consider the following extension.

For the case where $X_t = (X_{t1}, \ldots, X_{tp})^\tau$ is a vector of endogenous time series, model (2.5) can be extended to the following additive semiparametric error correction model

$$Y_t = X_t^\tau \beta + \sum_{j=1}^{q} \sum_{i=1}^{p} g_{ij}(\bigtriangledown X_{t-j,i}) + \epsilon_t, \tag{2.7}$$

where $\{g_{ij} : 1 \leq i \leq p, 1 \leq j \leq q\}$ are unknown functions defined on R^1 and the others are as defined before.

As this paper mainly considers applications of the proposed model specification testing procedures to model (2.6), estimation and testing procedures associated with model (2.7) will not be detailed. We leave them for future research.

2.2. *Special and extended forms of model (1.2)*

Example 2.9: Model (1.2) has some special forms. This example considers the case where $p = 1$, both X_t and e_t are i.i.d., and e_t is independent of X_t with $E[e_t] = 0$ and $E[e_t^2] < \infty$. Consider a partially linear model of the form

$$Y_t = X_t\beta + g(X_t) + e_t. \tag{2.8}$$

For the discussion of model (2.8) in the i.i.d. case, see Eubank and Spiegelman (1990), Eubank and Hart (1992), Chen (1994), Shively, Kohn and Ansley (1994), and Jayasuriva (1996).

Example 2.10 (Partially linear ARCH models): For the case where $p = 1$, Y_t is a sequence of time series, $X_t = Y_{t-1}$, and e_t depends on Y_{t-1}, model (1.2) is a partially linear ARCH model of the form

$$Y_t = \beta Y_{t-1} + g(Y_{t-1}) + e_t, \tag{2.9}$$

where e_t is assumed to be stationary, both β and g are identifiable, and $\sigma^2(y) = E[e_t^2|Y_{t-1} = y]$ is a smooth function of y. Hjellvik and Tjøstheim (1995), and Hjellvik, Yao and Tjøstheim (1998) considered testing for linearity in model (2.9). Granger, Inoue and Morin (1997) have considered some estimation problems for the case of $\beta = 1$ in model (2.9).

Example 2.11 (Nonparametric stochastic differential equations): This example involves using model (1.2) to approximate a continuous-time process of the form

$$dr_t = \mu(r_t)dt + \sigma(r_t)dB_t, \tag{2.10}$$

where $\mu(\cdot)$ and $\sigma(\cdot) > 0$ are respectively the drift and volatility functions of the process, and B_t is standard Brownian motion. We now consider a discretized version of model (2.10) of the form

$$r_{t\Delta} - r_{(t-1)\Delta} = \mu(r_{(t-1)\Delta})\Delta + \sigma(r_{(t-1)\Delta})[B_{t\Delta} - B_{(t-1)\Delta}],\ t = 1, 2, \cdots, \tag{2.11}$$

where Δ is the time between successive observations. In practice, Δ is small but fixed, as most continuous-time models in finance are estimated with monthly, weekly, daily, or higher frequency observations.

Suppose that data are sampled at time $t\Delta$ for $t = 1, 2, \cdots, T$. Let

$$Y_t = (r_{t\Delta} - r_{(t-1)\Delta})/\Delta,\ X_t = r_{(t-1)\Delta} \text{ and } \mu(X_t) = \beta X_t + g(X_t),$$

where both β and $g(\cdot)$ are identifiable. Model (2.11) now can be written as

$$Y_t = X_t\beta + g(X_t) + \sigma(X_t)\epsilon_t, \tag{2.12}$$

where ϵ_t is a Gaussian random error with $E[\epsilon_t] = 0$ and $\text{var}[\epsilon_t] = \Delta^{-1}$. Obviously, model (2.12) is a special form of model (1.2).

In Section 4 below, we will use model (2.12) to fit a given set of financial data.

As mentioned earlier, when p in model (1.2) is more than two or three, model (1.2) itself is not very feasible in practice due to the curse of dimensionality. As an alternative, one can use either

$$Y_t = X_t^\tau\beta + g(X_t^s) + e_t \tag{2.13}$$

or

$$Y_t = X_t^\tau\beta + \sum_{j=1}^{p} g_j(X_{tj}) + e_t, \tag{2.14}$$

where X_t^s is a sub-vector of X_t and each g_j is an unknown function defined on R^1. As both β and g_j are required to be identifiable, some orthogonality conditions on g_j are needed. The null hypothesis $H_0:\ g(\cdot) = 0$ has not been considered yet. When each g_j is approximated by a series of orthogonal functions as used in Gao, Tong and Wolff (2002a, 2002b), a test statistic for testing $H_0:\ g_j = 0$ can be constructed and its asymptotic distribution can be established. In general, each g_j can be estimated by using the so-called marginal integration method [see Linton and Härdle (1996); Linton (1997, 2000); Sperlich, Tjøstheim and Yang

(2002)] and then a test statistic can be constructed. As the detail is lengthy and extremely technical, it will not be given in this paper.

3. Estimation and model specification testing procedures

This section first considers an estimation and model specification procedure for a general nonparametric regression model. Specific applications of the proposed procedure to some of the above models are discussed later.

3.1. *Estimation and testing in a general model*

Suppose that (X, Y) is a $p+1$-dimensional process with $X = (X_1, \ldots, X_p)^\tau \in R^p$ and $Y \in R^1$. Consider a general nonparametric regression model of the form

$$Y = E[Y|X] + e = m(X) + e, \tag{3.1}$$

where $m(x) = E[Y|X = x]$ is an unknown function, e is an error process with mean zero and allowed to depend on X.

We first consider a general testing problem of the form

$$H_0: \ m(x) = 0.$$

As the choice of a test statistic depends on not only the type of estimator used for $m(\cdot)$ but also the type of distance measure, we suggest using a distance measure of the form

$$\pi_0 = E\{YE[Y|X]f(X)\} = E\{[E(Y|X)]^2 f(X)\} \geq 0, \tag{3.2}$$

where $f(\cdot)$ is the density function of X. It follows that $\pi_0 \equiv 0$ holds if and only if H_0 is true.

This section then constructs a test statistic for testing H_0. In order to do so, one needs to estimate the unknown function $m(\cdot)$ first.

Let $\{(X_t, Y_t) : 1 \leq t \leq T\}$ be a set of observations, T be the number of observations, and W be a $T \times T$ matrix depending on $(X_1, \ldots, X_T)$ and T. Let $\widehat{m}(\cdot)$ denote the general nonparametric estimator of $m(\cdot)$. Assume that

$$\widehat{M} = (\widehat{m}(X_1), \ldots, \widehat{m}(X_T))^\tau = WY, \tag{3.3}$$

where $Y = (Y_1, \ldots, Y_T)^\tau$ and W depends mainly on the type of nonparametric estimator used.

Assume that there are two sequences $\{p_{st}\}$ and $\{d_{st}\}$ with $\min_{1 \leq s,t \leq T} d_{st} > 0$ such that the $s \times t$ element, w_{st}, of W can be represented by $w_{st} = \frac{p_{st}}{d_{st}}$. Now equations (3.1)–(3.3) suggest using the following test statistic

$$L_T = \frac{\sum_{t=1}^T \sum_{s \neq t} p_{st} Y_s Y_t}{\widehat{S}_T}, \tag{3.4}$$

where $\widehat{S}_T^2 = 2\sum_{t=1}^T \sum_{s=1}^T p_{st}^2 Y_s^2 Y_t^2$. Note that the explicit form of W depends on the type of nonparametric estimator used in (3.3).

Before establishing the asymptotic distribution of (3.4), we consider the following examples.

Example 3.1 [Nonparametric kernel method]: Let K be a kernel function on R^p and h be a bandwidth parameter depending on T with $h = h_T \to 0$ as $T \to \infty$. In this example, we consider estimating $m(\cdot)$ by either the PC (see Priestley and Chao 1972) kernel estimator

$$\widehat{m}(x) = \frac{1}{Th^p} \sum_{s=1}^T K_h(x - X_s) Y_s$$

or the Nadaraya–Watson (NW) kernel estimator [see equation (2.4) of Fan and Gijbels (1996)]

$$\widehat{m}(x) = \frac{\frac{1}{Th^p} \sum_{s=1}^T K_h(x - X_s) Y_s}{\frac{1}{Th^p} \sum_{s=1}^T K_h(x - X_s)}.$$

Thus, in (3.4) one can choose

$$p_{st} = \frac{1}{Th^p} K_h(X_s - X_t) \quad \text{and} \quad d_{st} \equiv 1$$

for the PC case, and

$$p_{st} = \frac{1}{Th^p} K_h(X_s - X_t) \quad \text{and} \quad d_{st} = \frac{1}{Th^p} \sum_{u=1}^T K_h(X_s - X_u) \tag{3.5}$$

for the NW case, where $K_h(\cdot) = K(\cdot/h)$.

Remark 3.1. Note that L_T is similar to that proposed in Li (1999). In Li (1999), the author considers testing the hypothesis $H_0 : E[e|X] = 0$ in model (3.1) and uses the Nadaraya–Watson kernel estimator of the form

$$\widehat{m}(x) = \frac{\sum_{t=1}^T K_h(x - X_t) Y_t}{\sum_{t=1}^T K_h(x - X_t)}$$

and then constructs test statistics based on $K_{st} = K_h(X_s - X_t)$. In order to avoid the random denominator problem, the author chooses a modified test statistic of the form

$$\widetilde{L}_T = \frac{\sum_{t=1}^T \sum_{s \neq t} K_{st} \hat{e}_t \hat{f}_t \hat{e}_s \hat{f}_s}{\widetilde{S}_T},$$

where $\widetilde{S}_T^2 = 2\sum_{t=1}^T \sum_{s=1}^T K_{st}^2 \hat{e}_t^2 \hat{f}_t^2 \hat{e}_s^2 \hat{f}_s^2$, $\hat{e}_t = Y_t - \widehat{m}(X_t)$, and $\hat{f}_t = \hat{f}(X_t) = \frac{1}{Th^p} \sum_{s=1}^T K_{ts}$.

The next example involves the nonparametric series estimation method.

Example 3.2 [Nonparametric series method]: Assume that there are a sequence of series functions $\{z_i(\cdot) : 1 \leq i \leq k\}$ and a vector of unknown parameters $\{\gamma_i : 1 \leq i \leq k\}$

such that $m(x)$ can be approximated by $\sum_{i=1}^{k} z_i(x)\gamma_i$. Let $Z(x) = (z_1(x), \ldots, z_k(x))^\tau$, $Z = (Z(X_1), \ldots, Z(X_T))^\tau$, $\gamma = (\gamma_1, \ldots, \gamma_k)^\tau$ and $W = Z(Z^\tau Z)^+ Z^\tau$, in which $k = k_T$ is an integer, $k = k_T \to \infty$ as $T \to \infty$, and $(\cdot)^+$ denotes the Moore-Penrose inverse.

It can be shown that the least squares estimator of γ is given by $\hat{\gamma} = (Z^\tau Z)^+ Z^\tau Y$. The nonparametric series estimator of $m(\cdot)$ is defined as $\widehat{m}(x) = Z(x)^\tau \hat{\gamma}$.

We now have

$$\widehat{M} = (\widehat{m}(X_1), \ldots, \widehat{m}(X_T))^\tau = Z(Z^\tau Z)^+ Z^\tau Y = WY.$$

Without loss of generality, assume $c_i^2 = E[z_i^2(X_s)] = 1$ for all $i \geq 1$. Define the diagonal matrix I =diag$(1, \cdots, 1)$. Let d_{st} be the $s \times t$ element of the matrix

$$D = \left(I + T(ZZ^\tau)^+ \left(Z(Z^\tau Z)^+ Z^\tau - T^{-1} ZZ^\tau\right)\right)^+.$$

For the series method, one chooses

$$p_{st} = \frac{1}{T} \sum_{i=1}^{k} z_i(X_s) z_i(X_t) \text{ and } \ w_{st} = \frac{p_{st}}{d_{st}} \tag{3.6}$$

in (3.4).

Remark 3.2. One can consider using the following test statistic directly

$$\overline{L}_T = \frac{\sum_{t=1}^{T} \sum_{s \neq t} w_{st} Y_s Y_t}{\overline{S}_T}, \tag{3.7}$$

where $\overline{S}_T^2 = 2 \sum_{t=1}^{T} \sum_{s=1}^{T} w_{st}^2 Y_s^2 Y_t^2$, in which w_{st} is the $s \times t$ element of $W = Z(Z^\tau Z)^+ Z^\tau$.

In theory, it can be shown that $\overline{L}_T$ of (3.7) is asymptotically equivalent to L_T of (3.4) with p_{st} defined by (3.6). In practice, however, one would prefer to use L_T, as it avoids the random denominator problem.

We conclude the examples by pointing out that one can consider the case where $m(\cdot)$ is approximated by the B–spline. We shall not detail this case, as our experience shows that for the dependent time series observations the B–spline approximation is very difficult to be implemented in practice.

We now establish the first result of this paper.

Theorem 3.1. Assume that Assumptions A.1, A.2 and A.4 listed in Assumption A hold. Then under H_0

$$L_T \to_D N(0, 1) \text{ as } T \to \infty.$$

Furthermore, under $H_1: \ m(\cdot) \neq 0$, we have $\lim_{T\to\infty} P(L_T \geq C_T) = 1$, where C_T is any positive, nonstochastic sequence with $C_T = o(Tq^{-1/2})$, in which $q = q_T \to \infty$ as $T \to \infty$ is as defined in Assumption A.2.

Remark 3.1. Theorem 3.1 establishes the asymptotic distribution of the proposed statistic L_T of (3.4). It extends Theorem 3.1 of Li (1999) from the β–mixing condition to the α–mixing case. In addition, the test statistic L_T doesn't depend on a particular nonparametric estimation method, although one needs to identify the form of $\{p_{st}\}$ when implementing the test statistic in practice.

For Examples 3.1 and 3.2, we have the following corresponding results.

Corollary 3.1. (i) *Let p_{st} in (3.4) be defined by (3.5). Assume that Assumptions A.1, A.4 and A.5 hold. Then the conclusions of Theorem 3.1 remain true.*

(ii) *Let p_{st} in (3.4) be defined by (3.6). Assume that Assumptions A.1, A.4 and A.6 hold. Then the conclusions of Theorem 3.1 remain true.*

The proofs of Theorem 3.1 and Corollary 3.1 are relegated to Appendix B.

3.2. *Testing for nonparametric significance and linearity*

As the test statistics for some models are similar to those for others, one will only consider testing for nonparametric significance and linearity for some special forms of models (1.1) and (1.2).

3.2.1. Testing for nonparametric significance

Before discussing model (1.1), we consider a general nonparametric regression model of the form

$$Y_t = m(X_t) + e_t = m(U_t, V_t) + e_t, \tag{3.8}$$

where $X_t = (U_t^\tau, V_t^\tau)^\tau$, and $U_t = (U_{t1}, \cdots, U_{td})^\tau$ ($d \leq p-1$) and $V_t = (V_{t1}, \cdots, V_{tc})^\tau$ ($c = p-d$) are allowed to be two different time series. For example, U_t could be a vector of endogenous time series while V_t could be a vector of exogenous time series. Due to the curse of dimensionality problem arising from using nonparametric regression modelling, before applying model (3.8) in practice one needs to consider whether Y_t depends only on the time series U_t. In other words, one needs to test whether the null hypothesis $H_0: \ E[Y_t|X_t] - E[Y_t|U_t] = 0$ holds.

Under Assumption A.3, one can estimate $m_1(U_t) = E[Y_t|U_t]$ by

$$\widehat{m}_1(U_t) = \sum_{s=1}^{T} w_{1ts} Y_s,$$

where $w_{1st} = \frac{p_{1st}}{c_{1st}}$ is as defined in Assumption A.3(ii).

Thus one can estimate $m_2(X_t) = m(X_t) - m_1(U_t)$ by

$$\widehat{m}_2(X_t) = \sum_{s=1}^{T} w_{ts} Y_s - \sum_{s=1}^{T} w_{1ts} Y_s. \tag{3.9}$$

This suggests using the following test statistic

$$L_{1T} = \frac{\sum_{t=1}^T \sum_{s \neq t} p_{st} \hat{Y}_s \hat{Y}_t}{\hat{S}_{1T}}, \tag{3.10}$$

where $\hat{Y}_t = [Y_t - \widehat{m}_1(U_t)] d_{1Tt}$, $d_{1Tt} = \frac{1}{T}\sum_{s=1}^T c_{1ts}$, and $\hat{S}_{1T}^2 = 2\sum_{t=1}^T \sum_{s=1}^T p_{st}^2 \hat{Y}_s^2 \hat{Y}_t^2$.

When model (3.8) is a nonparametric additive model of the form

$$Y_t = m_1(U_t) + m_2(V_t) + e_t, \tag{3.11}$$

one can use (3.10) to test the nonparametric hypothesis $H_0: \ m_2 = 0$.

Remark 3.3. The test statistic L_{1T} of (3.10) is similar to that of (6) of Li (1999). For the NW case, L_{1T} is actually identical to that of (6) of Li (1999). An empirical study of the test statistic for testing $H_0: \ m_2 = 0$ is given in Example 4.1.

For model (3.8), we now consider the null hypothesis $H_0: \ m_2(X_t) = 0$ versus the alternative $H_1: \ m_2(X_t) \neq 0$. For model (3.11), we consider the null hypothesis $H_0: m_2(V_t) = 0$ versus the alternative $H_1: \ m_2(V_t) \neq 0$.

Theorem 3.2. Let $X_t = (U_t^\tau, V_t^\tau)^\tau$. Assume that (X_t, Y_t) and e_t satisfy Assumption A.1. In addition, suppose that Assumptions A.3 and A.4 hold. Then the conclusions of Theorem 3.1 hold for L_{1T}.

As a special case of model (3.11), one can consider model (1.1) given by

$$Y_t = U_t^\tau \beta + g(V_t) + e_t.$$

For model (1.1), one can estimate the nonparametric component $g(\cdot)$ and then define the estimators of β and $g(\cdot)$ by [see Härdle, Liang and Gao (2000, §1.2)],

$$\hat{\beta}_1 = (\tilde{U}^\tau \tilde{U})^+ \tilde{U}^\tau \tilde{Y} \text{ and } \hat{g}(V_t) = \sum_{s=1}^T w_{2ts}(Y_s - U_s^\tau \hat{\beta}_1),$$

respectively, where $\tilde{U} = (I - W_2)U$, $U = (U_1, \ldots, U_T)^\tau$, $\tilde{Y} = (I - W_2)Y$, $W_2 = \{w_{2st}\}$ is a $T \times T$ matrix with w_{2st} as its $s \times t$ element, and the definition of $w_{2st} = \frac{p_{2st}}{c_{2st}}$ with $\min_{1 \leq s \leq T} c_{2st} > 0$ is similar to that of w_{1st}.

Similar to (3.10), one can construct the following test statistic

$$L_{2T} = \frac{\sum_{t=1}^T \sum_{s \neq t} p_{2st} \hat{Y}_s \hat{Y}_t}{\hat{S}_{2T}},$$

where $\hat{S}_{2T}^2 = 2\sum_{t=1}^T \sum_{s=1}^T p_{2st}^2 \hat{Y}_s^2 \hat{Y}_t^2$ and $\hat{Y}_t = Y_t - U_t^\tau \hat{\beta}_1$.

For model (1.1), we now have the following result for the null hypothesis $H_0: \ g = 0$ versus the alternative $H_1: \ g \neq 0$.

Theorem 3.3. Let $X_t = (U_t^\tau, V_t^\tau)^\tau$. Assume that (X_t, Y_t) and e_t satisfy Assumption A.1. In addition, suppose that Assumptions A.3 and A.4 hold. Then the conclusions of Theorem 3.1 hold for L_{2T}.

For the series method case, Theorem 3.3 is similar to Theorem 2.3 of Gao, Tong and Wolff (2002a). The proofs of Theorems 3.2 and 3.3 are relegated to Appendix B.

Corollary 3.2. As models (2.1), (2.2), (2.4) and (2.6) are special cases of model (1.1), the corresponding test statistics and their asymptotic distributions follow immediately.

Corollary 3.3. For model (2.3), one needs to construct the following test statistic

$$L_{3T} = \frac{\sum_{t=1}^T \sum_{s \neq t} \hat{p}_{st} \hat{Y}_s \hat{Y}_t}{\hat{S}_{3T}},$$

where $\hat{S}_{3T}^2 = 2 \sum_{t=1}^T \sum_{s=1}^T \hat{p}_{st}^2 \hat{Y}_s^2 \hat{Y}_t^2$, $\hat{Y}_t = Y_t - X_t^\tau \hat{\beta}_2$, $\hat{p}_{st} = p(\hat{Y}_{s-1}, \hat{Y}_{t-1})$, and $\hat{\beta}_2$ is the solution of

$$\min_\beta \sum_{t=1}^T \left[u_t - \sum_{s=1}^T p(u_{t-1}, u_{s-1}) \right]^2,$$

where $u_t = Y_t - X_t^\tau \beta$.

Under some additional conditions on the nonparametric estimation function p_{st}, the asymptotic normality of L_{3T} can be established, although the detail is extremely technical.

3.2.2. Testing for linearity in model (1.2)

Consider a generalized form of model (1.2) given by

$$Y_t = \alpha + X_t^\tau \beta + g(X_t) + e_t, \tag{3.12}$$

where α, β and $g(\cdot)$ are identifiable.

It follows that the least squares estimators of α and β can be defined as

$$\hat{\alpha} = \bar{Y} - \bar{X}^\tau \hat{\beta}_3,$$

$$\hat{\beta}_3 = \left(\sum_{t=1}^T (X_t - \bar{X})(X_t - \bar{X})^\tau \right)^+ \sum_{t=1}^T (X_t - \bar{X})(Y_t - \bar{Y}),$$

where $\bar{X} = \frac{1}{T} \sum_{t=1}^T X_t$ and $\bar{Y} = \frac{1}{T} \sum_{t=1}^T Y_t$.

One can now suggest the following test statistic for testing $H_0: \ g = 0$,

$$L_{4T} = \frac{\sum_{t=1}^T \sum_{s \neq t} p_{st} \hat{Y}_s \hat{Y}_t}{\hat{S}_{4T}},$$

where $\hat{S}_{4T}^2 = 2 \sum_{t=1}^T \sum_{s=1}^T p_{st}^2 \hat{Y}_s^2 \hat{Y}_t^2$, $\hat{Y}_t = Y_t - \hat{\alpha} - X_t^\tau \hat{\beta}_3$, and p_{st} is as defined in (3.4).

We now have the following result for the null hypothesis $H_0: \ g = 0$ versus the alternative $H_1: \ g \neq 0$.

Theorem 3.4. Assume that Assumptions A.1, A.2 and A.4 hold. In addition, suppose that α, β and $g(\cdot)$ are identifiable. Then the conclusions of Theorem 3.1 hold for L_{4T}.

The proof of Theorem 3.4 is relegated to Appendix B.

Corollary 3.4. As models (2.8), (2.9) and (2.12) are special cases of model (1.2), the corresponding test statistics and their asymptotic distributions for the models can be established immediately.

For model (2.10), we will use model (2.12) to approximate it and then apply model (2.10) to fit a set of financial data in Section 4.3. In the meantime, some alternative estimators for both the drift and the diffusion are also provided and compared in some detail.

4. Implementation and applications

This section illustrates the proposed estimation and testing procedure by three simulated examples and one real data analysis. We consider only small sample studies and applications for a nonparametric additive model and some special cases of model (1.2) due to the following reasons:

(i) Small sample studies for model (1.1) are similar to those for the additive model and model (1.2);
(ii) some special cases of model (1.1) have already been discussed [see Härdle, Liang and Gao (2000, §6.2)]; and
(iii) model (1.2) has econometric applications.

4.1. *Testing for nonparametric significance*

In this section, we illustrate the test statistic L_{1T} of (3.10) by a simulated example. Rejection rates of the test statistic L_{1T} are detailed in Example 4.1. Let $X \sim U(a, b)$ denote that X is uniformly distributed over $[a, b]$, and $e \sim N(\mu, \sigma^2)$ denote that e is normally distributed with mean μ and variance σ^2.

Example 4.1. Consider a nonparametric additive model of the form

$$Y_t = 0.3\cos(U_t) + \phi\sin(V_t) + e_t,\ t = 1, 2, \ldots, T, \tag{4.1}$$

$$U_t = -0.5U_{t-1} + \epsilon_t,\ V_t = 0.5V_{t-1} + \zeta_t,\ e_t = \sigma_0\eta_t\sqrt{0.25 + 0.5U_t^2},$$

where $0 \le \phi \le 1$ is a constant, $\{\epsilon_t : t \ge 1\}$, $\{\zeta_t : t \ge 1\}$ and $\{\eta_t : t \ge 1\}$ are mutually independent and identically distributed, $\{U_t : t \ge 1\}$ and $\{V_t : t \ge 1\}$ are independent, $\{\epsilon_t : t \ge 1\}$ are independent of U_0, $\{\zeta_t : t \ge 1\}$ are independent of V_0, $\epsilon_t \sim U(-0.5, 0.5)$, $\zeta_t \sim U(-0.5, 0.5)$, $U_0 \sim U(-1, 1)$, $V_0 \sim U(-1, 1)$, $\eta_t \sim N(0, 1)$, and $\sigma_0 > 0$ is to be specified. It is clear from (4.1) that Assumption A.1 holds.

This example then considers the small sample behaviour of the proposed test statistic

$$L_{1T} = \frac{\sum_{t=1}^{T} \sum_{s \neq t} p_{st} \hat{Y}_s \hat{Y}_t}{\hat{S}_{1T}},$$

where

$$\hat{Y}_t = [Y_t - \widehat{m}_1(U_t)]\hat{f}_t, \ \hat{f}_t = \sum_{s=1}^{T} p_{1ts}, \ \widehat{m}_1(U_t) = \sum_{s=1}^{T} w_{1ts} Y_s, \ \widehat{S}_{1T}^2 = 2\sum_{t=1}^{T}\sum_{s=1}^{T} p_{st}^2 \hat{Y}_s^2 \hat{Y}_t^2,$$

$$p_{st} = \frac{1}{Th} K_h(X_s - X_t), \ p_{1st} = \frac{1}{Th} K_h(U_s - U_t), \ w_{1st} = \frac{\frac{1}{Th} K_h(U_s - U_t)}{\sum_{t=1}^{T} \frac{1}{Th} K_h(U_s - U_t)},$$

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \text{ and } h^{-1} = 2T^{1/4}.$$

Obviously, Assumptions A.4 and A.5 hold. For Example 4.1, we use the asymptotic critical value $L_0 = 1.65$ at the 5% level. For model (4.1) we consider the cases where $T = 50$, 150 and 250. The simulation results were performed 1500 times and the rejection rates are tabulated in Table 4.1 below.

Table 4.1. Rejection Rates For Example 4.1

Sample	bandwidth	variance	Rejection rate of L_{1T}			
T	h	σ_0^2	$\phi = 0$	$\phi = 0.1$	$\phi = 0.25$	$\phi = 0.5$
50	0.1871	0.5	0.000	0.020	0.073	0.520
150	0.1426	0.5	0.000	0.273	0.960	1.000
250	0.1256	0.5	0.006	0.706	1.000	1.000
50	0.1871	1.0	0.000	0.000	0.026	0.300
150	0.1426	1.0	0.000	0.046	0.460	1.000
250	0.1256	1.0	0.006	0.106	0.926	1.000
50	0.1871	1.5	0.000	0.000	0.013	0.120
150	0.1426	1.5	0.000	0.006	0.226	0.993
250	0.1256	1.5	0.006	0.040	0.666	1.000

Remark 4.1. Table 4.1 shows that the rejection rates seem relatively sensitive to the values of T, ϕ, and σ_0. The power increased as ϕ increased while the power decreased as σ_0 increased for almost all cases. This shows that the rejection rates depend strongly on the values of σ_0^2 as well as ϕ. In addition, Table 4.1 shows that the overall rejection rate is high. For example, for the case where $\phi = 0.25$, $\sigma_0^2 = 0.5$ and $T = 250$, the rejection rate is already 100%. In the meantime, our small sample studies show that the test statistic is very sensitive in accepting the null hypothesis for the case of $\phi = 0$. For example, almost all acceptance rates for the case of $\phi = 0$ are 100%. We think that the reason why the test statistic is very sensitive in

terms of accepting or rejecting the hypothesis is probably because the test statistic suggested by Li (1999) overcomes the random denominator problem, which could slow the rejection rates. We also computed the rejection rates for a modified form of the test statistic L_{1T} with $\hat{Y}_t = [Y_t - \widehat{m}_1(U_t)]\hat{f}_t$ in L_{1T} replaced by $\tilde{Y}_t = Y_t - \widehat{m}_1(U_t)$. Our small sample studies show that the rejection rates of L_{1T} are always higher than those of the modified form. Theoretically, however, we haven't been able to show that the test statistic L_{1T} is more powerful than the modified form.

4.2. *Testing for linearity*

In this section, we illustrate the test statistic L_{4T} by a simulated example. Rejection rates of the test statistic L_{4T} are detailed in Example 4.2.

Example 4.2. Consider a state-space model of the form

$$\begin{aligned} Y_t &= 0.3X_t + \phi X_t^2 + e_t, \ t = 1, 2, \dots, T, \\ X_t &= 0.5X_{t-1} + \epsilon_t, \ e_t = \sigma_0 \eta_t \sqrt{0.5 + 0.25X_t^2}, \end{aligned} \tag{4.2}$$

where $0 \le \phi \le 1$ is a constant, both $\{\epsilon_t : t \ge 1\}$ and $\{\eta_t : t \ge 1\}$ are mutually independent and identically distributed, the $\{\epsilon_t : t \ge 1\}$ are independent of X_0, the $\{\eta_t : t \ge 1\}$ are independent of X_0, $\epsilon_t \sim U(-0.5, 0.5)$, $X_0 \sim U(-1, 1)$, $\eta_t \sim N(0, 1)$, and $\sigma_0 > 0$ is to be specified.

First, it is clear from (4.2) that Assumption A.1 holds. Second, in the calculation of L_{4T}, we choose the following quantities

$$p_{st} = \frac{1}{Th} K_h(X_s - X_t), \ K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \text{ and } h^{-1} = 4T^{1/5}.$$

Obviously, Assumptions A.4 and A.5 hold. For Example 4.2, we use the asymptotic critical value $L_0 = 1.65$ at the 5% level. For model (4.2) we consider the cases where $T = 50$, 150, 250 and 350. The simulation results were performed 1500 times and the rejection rates are tabulated in Table 4.2 below.

Table 4.2. Rejection Rates For Example 4.2

T	h	σ_0^2	$\phi = 0$	$\phi = 0.10$	$\phi = 0.25$	$\phi = 0.50$
50	0.114	0.15	0.026	0.026	0.106	0.386
150	0.091	0.15	0.026	0.080	0.433	0.993
250	0.083	0.15	0.033	0.140	0.700	1.000
350	0.077	0.15	0.046	0.226	0.880	1.000
50	0.114	0.10	0.026	0.033	0.153	0.573
150	0.091	0.10	0.026	0.100	0.606	1.000
250	0.083	0.10	0.033	0.206	0.880	1.000
350	0.077	0.10	0.046	0.273	0.973	1.000

Remark 4.2. Table 4.2 shows that the rejection rates seem relatively sensitive to the choice of T, ϕ, and σ_0. The power increased as ϕ increased while the power decreased as σ_0 increased for the case of $\phi \neq 0$. This shows that the rejection rates depend strongly on the choice of σ_0^2. For example, the rejection rate for model with $\phi = 0.25$, $\sigma_0^2 = 0.10$ and $T = 350$ is already 97.3%. Moreover, the rejection rates for the case where $\sigma_0^2 = 0.15$ and $\phi = 0.10$ or 0.25 are higher than those for the case where $\sigma_0^2 = 0.10$ and $\phi = 0.10$ or 0.25. For the case where $\phi = 0$, however, the rejection rates for the case of $\sigma_0^2 = 0.15$ are indistinguishable to those for the case of $\sigma_0^2 = 0.10$. Similarly, we computed the rejection rates for the case where the distribution of η_t is replaced by $U(-1, 1)$. Our simulation results show that the performance of L_{4T} under the normal error is better than that under the uniform error.

For examples 4.1 and 4.2, we also computed the rejection rates for the series based test statistics, and the simulation results are similar to those based on the kernel method.

4.3. *Implementation and application in financial models*

Recently, several researchers have used nonparametric techniques to estimate continuous-time diffusion processes that are observed at discrete intervals. For example, Aït-Sahalia (1996a) estimated the diffusion function (or volatility function) nonparametrically, given a linear specification for the drift function. Stanton (1997) constructed a family of approximations to the drift and diffusion of a diffusion process, and estimated the approximations nonparametrically. Fan and Yao (1998) considered using the local linear kernel method to estimate both the drift and the diffusion of a class of discrete time series models, and presented asymptotic properties as well as practical applications.

While estimating the diffusion function nonparametrically is quite reasonable, it is too restrictive to impose the linearity on the drift as there is evidence of substantial nonlinearity in the drift [see Aït-Sahalia (1996b) for example]. As pointed out by Ahn and Gao (1999), the linearity of the drift imposed in the literature appears to be the main source of misspecification. To avoid misspecification for the drift function, it would be better to consider a model specification problem before determining whether one should impose the linearity on the drift. Aït-Sahalia (1996b) already considered testing the parametric specification of diffusion processes. Pritsker (1998) conducted the finite sample simulation of one of Aït-Sahalia (1996b) nonparametric tests of continuous time models of the short-term riskless rate. See also Jiang and Knight (1997), and Chapman and Pearson (2000).

Consider model (2.10). It follows from Aït-Sahalia (1996a) and Stanton (1997) that

$$\mu(x) = \frac{1}{2\pi(x)} \frac{d}{dx}[\sigma^2(x)\pi(x)] \tag{4.3}$$

and

$$\sigma^2(x) = \frac{2}{\pi(x)} \int_0^x \mu(u)\pi(u)du, \tag{4.4}$$

where π is the stationary density of X_t.

Equation (4.3) allows us to estimate the drift function nonparametrically, given a nonparametric estimate of the stationary density, π, but only if we know the diffusion, σ. Conversely, equation (4.4) allows us to estimate the diffusion function nonparametrically, given a nonparametric estimate of the stationary density, π, but only if we know the drift, μ. Aït-Sahalia (1996a) assumed a linear drift

$$\mu(x) = \kappa[\theta - x] \tag{4.5}$$

in (4.4) and then estimated σ^2 nonparametrically. As argued by Stanton (1997) and some other authors, however, there is mounting evidence that condition (4.5) is not suitable.

In order to determine the linearity in the drift, we suggest testing the null hypothesis

$$H_0: \ \mu \text{ is linear } \ \text{versus } H_1: \ \mu \text{ is nonlinear.}$$

As can be seen from models (2.10) and (2.12), in order to test whether the drift function is linear, it suffices to test whether $H_0: \ g(\cdot) = 0$ holds in model (2.12).

Before using our test statistic L_{4T} in practice, we review some related estimation and testing methods. It follows from (4.3) and (4.4) that the estimator of $\sigma^2(\cdot)$ can be constructed based on the estimator of $\mu(\cdot)$, and vice versa. As mentioned earlier, Aït-Sahalia (1996a) uses (4.4) to estimate $\sigma(\cdot)$ based on the linear estimator of $\mu(\cdot)$. Jiang and Knight (1997) estimated $\sigma(\cdot)$ and then use (4.3) to estimate $\mu(\cdot)$. More recently, Chapman and Pearson (2000) conducted small sample studies for the estimators proposed in Aït-Sahalia (1996a) and Stanton (1997). Their conclusion is that there is no definitive answer to the question that the drift function of short-term interest rate data is nonlinear.

Thus we suggest using the test statistic L_{4T} for testing the linearity. Unlike the testing procedure proposed in Aït-Sahalia (1996b), we test for linearity in the drift rather than in both the drift and the diffusion. As pointed out by Aït-Sahalia (1996b), in order to test both the drift and the diffusion, it suffices to test whether the stationary density $\pi(\cdot)$ belongs to a specific family of density functions.

In this section, we then illustrate Theorem 3.4 using one simulated example and one real example. Rejection rates of the test statistic are detailed in Example 4.3.

Example 4.3. Consider the interest rate model proposed by Ahn and Gao (1999),

$$dr_t = \kappa(\theta - r_t)r_t dt + \sigma r_t^{1.5} dB_t, \ t = 1, 2, \cdots, \tag{4.6}$$

with parameter values $\kappa > 0$, $\theta > 0$ and $\sigma > 0$, where B_t is standard Brownian motion. Model (4.6) was proposed by Ahn and Gao (1999). The authors show that the necessary and sufficient conditions for stationarity of the process are $\kappa > 0$ and $\theta > 0$ [see Appendix

A of Ahn and Gao (1999)]. The authors also consider estimating the parameters κ, θ and σ [see Table 3 of Ahn and Gao (1999)].

Assume that the initial interest rate is $r_0 = 0.06$. In this example, we consider the discretized model (2.12)

$$Y_t = \beta X_t + g(X_t) + \sigma(X_t)\epsilon_t, \tag{4.7}$$

where $\epsilon_t \sim N(0, \Delta^{-1})$,

$$Y_t = \frac{r_{t\Delta} - r_{(t-1)\Delta}}{\Delta} \quad \text{and} \quad X_t = r_{(t-1)\Delta},$$

$$\mu(X_t) = \kappa(\theta - X_t)X_t = \beta X_t + g(X_t) \quad \text{and} \quad \sigma(X_t) = \sigma X_t^{1.5},$$

in which both $\beta = \kappa\theta$ and $g(X_t) = -\kappa X_t^2$ are identifiable. In the following small sample study, we consider three different choices of Δ: $\Delta = \frac{1}{250}$ (daily), $\Delta = \frac{5}{250}$ (weekly), and $\Delta = \frac{20}{250}$ (monthly).

In this example, one considers using the series approximation to $g(\cdot)$. The family of orthogonal series used here is

$$\{\cos(\pi v), \cdots, \cos(k\pi v)\}, \tag{4.8}$$

where $k = 4\left[T^{\frac{1}{5}}\right]$ and $v \in [-1, 1]$.

First, it is clear that Assumption A.1 holds. See for example, Lu (1998). Second, applying the property of trigonometric functions, we have

$$E[\cos(i\pi V_s)\cos(i\pi V_t)] = 0 \text{ and } E[\cos(i\pi V_t)\cos(j\pi V_t)] = 0$$

for all $i \neq j$ and $s \neq t$, where $V_t = 100(X_t - \overline{X})$ and $\overline{X} = \frac{1}{T}\sum_{t=1}^{T} X_t$. As the simulated values of r_t are generally small, we use V_t instead of X_t in the sample simulation. Therefore Assumption A.6 holds. Finally, as pointed out by Hong and White (1995), Eumunds and Moscatelli's (1977) results can be applied to show that non-periodic functions can still be approximated by the family of trigonometric series (4.8). Thus Assumption A.6 holds with $\mu = 1$. Moreover, the optimum convergence rate given in Assumption A.6(i) is obtained as in the periodic case.

Based on (3.6) and (4.6)–(4.8), we can now compute L_{4T}. For the calculation of rejection rates of the null hypothesis H_0, one needs to use the asymptotic critical value $L_0 = 1.65$ at the 5% level.

This example uses $\kappa = 3.5$, $\theta = 0.08$ and $\sigma = 1.28$ [see Table 3 of Ahn and Gao (1999)] for the detailed simulation. The simulation results below were performed 1500 times and the rejection rates are tabulated in Table 4.3 below.

Table 4.3. Rejection Rates For Example 4.3

T	k	$\Delta = \frac{1}{250}$	$\Delta = \frac{5}{250}$	$\Delta = \frac{20}{250}$
50	8	0.280	0.240	0.272
500	13	0.439	0.448	0.424
1050	16	0.700	0.600	0.409
1550	17	0.900	0.666	0.474

Remark 4.3: Table 4.3 shows that the rejection rates seem relatively sensitive to the choice of both k and Δ, although the choice of k is not so significant for the case of $\Delta = \frac{20}{250}$. For the case where $T = 1050$ or 1550, the rejection rates decrease as the values of Δ increase. This demonstrates that the rejection rates depend heavily on how the continuous process r_t is discretised. When $T = 1550$ and $\Delta = \frac{1}{250}$, the rejection rate is as high as 90%.

Example 4.4: This example considers using model (2.10) to fit the US Federal interest rate data, monthly from January 1963 through December 1998. Let r_t denote the interest rate data, $X_t = 100r_t$ and $Y_t = (X_{t+1} - X_t)/\Delta$ for $t = 1, 2, \cdots, 431$, where $\Delta = 20/250$.

This example considers using the test statistic of L_{4T} to determine whether nonlinearity in the drift is appropriate for the interest rate data. In the calculation of L_{4T}, we choose the following quantities

$$p_{st} = \frac{1}{Th} K_h(X_s - X_t), \;\; K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \text{ and } h^{-1} = 4 \cdot T^{1/5}.$$

In this example, we estimate $\sigma^2(\cdot)$ by

$$\hat{\sigma}^2(X_t) = \Delta \cdot \sum_{s=1}^{T} w_{ts} [Y_s - X_s \hat{\beta}_3 - \hat{g}(X_s)]^2,$$

where $w_{ts} = \frac{K_h(X_t - X_s)}{\sum_{u=1}^T K_h(X_t - X_u)}$ and $\hat{g}(X_t) = \sum_{s=1}^T w_{ts}(Y_s - X_s \hat{\beta}_3)$.

By computing L_{4T} for model (4.7), we obtain that

$$L_{4T} = 2.66 > 1.65,$$

which is the asymptotic critical value at the 5% level. This conclusion is the same as Gao (2000), who considered using the series based test statistic. As suggested by some other existing studies [see Aït-Sahalia (1996a); Stanton (1997)], our research suggests that it is unreasonable to assume the linearity in the drift. Moreover, as one can see from the plots, the drift function appears to be nonlinear while the diffusion looks neither linear, nor like a square root function, but appears closer to the $r^{1.5}$ given in Example 4.3. The model was suggested by Ahn and Gao (1999). Some other studies already show that parametric nonlinear models can also be used to fit the data. See for example, Aït-Sahalia (1999).

5. Conclusion

In this paper, we consider the general nonparametric time series regression model (3.1), estimate the mean by the nonparametric weight function (3.3), and then propose the model specification testing statistic (3.4) for testing the mean under the α–mixing condition. As an application of the model specification procedure, we consider testing for nonparametric significance in the nonparametric time series regression model (3.8). Testing for nonparametric additivity and linearity has also been discussed. The results for nonparametric time series regression models under the α–mixing condition complement some existing results under the β–mixing condition. See for example, Li (1999). In order to deal with the α–mixing condition, we establish some general results for moment inequalities [see Lemma C.2] and limit theorems [see Lemma B.1] for degenerate U–statistics of strongly dependent processes. Both Lemmas B.1 and C.2 are applicable to some other nonparametric estimation and testing of time series with the α–mixing condition. In addition, we consider testing for linearity in the partially linear regression model (1.2). Applications of the estimation and model specification procedure for model (1.2) to three simulated examples and one real data set are given in some detail.

The main drawback of the proposed model specification testing procedures is that the smoothing parameter q involved in the procedures is nonrandom and fixed. In the examples, we use some theoretically optimum values for the bandwidth parameter h and the truncation parameter k. In theory, we hope to show that the conclusions of Theorems 3.1–3.4 remain unchanged when q is replaced by a random data-driven $\hat{q}$. More recently, Gao and Tong (2001b) suggest that for the series case asymptotic normality of series based test statistics remains true when the truncation parameter k is replaced by a random data-driven $\hat{k}$. Theorem 3 of Lavergne (2001) states that it is also true for the kernel case. As the detailed discussion is extremely technical, we do not discuss the problem any further in this paper.

The results given in this paper can be extended in a number of directions. First, it is possible to consider testing for linearity for models (2.7) and (2.14). Second, the results of this paper for the short-range dependent time series case can be extended to the long-range dependent time series case, for which one needs to modify Lemmas B.1 and C.2 given below. Third, one probably can relax the strict stationarity and the mixing condition, as the recent work by Karlsen and Tjøstheim (2001) indicates that it may be possible to do such work without the stationarity and the mixing condition. This part is particularly important for the two reasons: (i) for the long-range dependent case one needs to avoid assuming both the long-range dependence and the mixing condition, as they contradict each other; and (ii) some important models are nonstationary. For example, when $\beta = 1$, model (2.9) is nonstationary. Some of the issues are left for possible future research.

A. Appendix A

This appendix lists the necessary assumptions for the establishment and the proof of the main results given in Section 3.

Assumption A.1. (i) Assume that the process (X_t, Y_t) is strictly stationary and α-mixing with the mixing coefficient $\alpha(t) \leq C_\alpha \alpha^t$ defined by

$$\alpha(t) = \sup\{|P(A \cap B) - P(A)P(B)| : A \in \Omega_1^s, B \in \Omega_{s+t}^\infty\}$$

for all $s, t \geq 1$, where $0 < C_\alpha < \infty$ and $0 < \alpha < 1$ are constants, and Ω_i^j denotes the σ-field generated by $\{(X_t, Y_t) : i \leq t \leq j\}$.

(ii) Assume that $e_t = Y_t - E[Y_t|X_t]$ satisfies for all $t \geq 1$

$$E[e_t|\Omega_{t-1}] = 0,$$

where $\Omega_t = \sigma\{(X_{s+1}, Y_s) : 1 \leq s \leq t\}$ is a sequence of σ-fields generated by $\{(X_{s+1}, Y_s) : 1 \leq s \leq t\}$.

(iii) In addition, assume

$$E[|e_t^{4+\xi}|] < \infty \text{ and } E\left[\left|e_{t_1}^{i_1} e_{t_2}^{i_2} \cdot e_{t_l}^{i_l}\right|^{1+\eta}\right] < \infty$$

for some small $\xi > 0$ and $\eta > 0$, where $2 \leq l \leq 4$ is an integer, $0 \leq i_j \leq 4$ and $\sum_{j=1}^{l} i_j \leq 8$.

Assumption A.2. (i) There are two measurable functions $\{p_{st}\}$ and $\{d_{st}\}$ with $\min_{1 \leq s,t \leq T} d_{st} > 0$ such that the $s \times t$ element, w_{st}, of W can be represented by $w_{st} = \frac{p_{st}}{d_{st}}$. Moreover, assume that $p_{st} = p(X_s, X_t)$ is a symmetric and continuous function of (X_s, X_t). There is a positive number q such that

$$\max_{1 \leq s,t \leq T} |p_{st}| \leq \frac{C_0 q}{T},$$

where $q = q_T$ satisfies $q_T \to \infty$ as $T \to \infty$.

(ii) Let $\sigma_{ij}^2 = \text{var}(e_s p_{st} e_t)$ and $\sigma_T^2 = \sum_{1 \leq s < t \leq T} \sigma_{ij}^2$. Assume that

$$\lim_{T \to \infty} \frac{q^4}{T \sigma_T^4} = 0 \text{ and } \lim_{T \to \infty} \frac{q^3}{T \sigma_T} = 0.$$

(iii) For $1 \leq i, j \leq T$, let $P(X_i)$ and $P(X_i, X_j)$ be the probability measures of X_i and (X_i, X_j), respectively. Define $\sigma^2(x) = E[e_t^2|X_t = x]$,

$$\psi(X_i, X_j) = \int \sigma^2(x) p(x, X_i) p(x, X_j) dP(x),$$

$$C_{1T} = \max_{1 \leq i < j < k \leq T} \left\{ E\left[|p_{ik} p_{jk}|^{(1+\delta)(1+\delta_1)}\right], \int\int\int |p_{ik} p_{jk}|^{(1+\delta)(1+\delta_1)} dP(X_i) dP(X_j, X_k) \right\},$$

$$C_{2T} = \max_{1 \leq i < j \leq T} \left\{ E|\psi(X_i, X_j)|^{2(1+\delta)}, \int\int |\psi(X_i, X_j)|^{2(1+\delta)} dP(X_i) dP(X_j) \right\},$$

$$C_{3T} = E|\psi(X_i, X_j)|^2,$$

where $0 < \delta < 1$ and $0 < \delta_1 < 1$ satisfy $\frac{1+\delta}{3-\delta} < \delta_1 < \frac{1-\delta}{1+\delta}$. Assume that as $T \to \infty$

$$\frac{T^2 C_{1T}^{\frac{1}{(1+\delta)(1+\delta_1)}}}{\sigma_T^2} \to 0, \quad \frac{T^2 C_{2T}^{\frac{1}{2(1+\delta)}}}{\sigma_T^2} \to 0, \quad \frac{T^2 C_{3T}^{\frac{1}{2}}}{\sigma_T^2} \to 0.$$

Assumption A.3. (i) Assumptions A.2(i)–A.2(iii) hold.

(ii) Assume that there are two measurable functions $\{p_{1st}\}$ and $\{c_{1st}\}$ with $\min_{1\le s,t\le T} c_{1st} > 0$ such that the $s \times t$ element, w_{1st}, of L_{1T} can be represented by $w_{1st} = \frac{p_{1st}}{c_{1st}}$. Moreover, assume that $p_{1st} = p_1(U_s, U_t)$ is a symmetric and continuous function of (U_s, U_t). Let $d_{1Ts} = \frac{1}{T}\sum_{t=1}^{T} c_{1st}$. Assume that there is a positive, continuous and bounded function $d_{1s} = d_1(U_s)$ such that as $T \to \infty$

$$d_{1Ts} - d_{1s} \to_p 0$$

uniformly in $s \ge 1$.

(iii) Let $\delta_{1t} = m_1(U_t) - \widehat{m}_1(U_t)$ and $\eta_{1t} = d_{1Tt} - d_{1t}$. In addition to Assumption A.3(ii), suppose that as $T \to \infty$

$$\frac{1}{\sigma_{1T}} \sum_{1\le s<t\le T} E\left[e_s^2 \eta_{1s}^2 |p_{st}|\right] \to 0,$$

$$\frac{1}{\sigma_{1T}} \sum_{1\le s<t\le T} E\left[e_s^2 \delta_{1s}^2 d_{1Ts}^2 |p_{st}|\right] \to 0,$$

and

$$\frac{1}{\sigma_{1T}^2} \sum_{1\le s<t\le T} E\left[\delta_{1s}^2 d_{1Ts}^2 \delta_{1t}^2 d_{1Tt}^2 p_{st}^2\right] \to 0,$$

where $\sigma_{1T}^2 = 2\sum_{t=1}^{T}\sum_{s=1}^{T} E\left[p_{st}^2 e_s^2 e_t^2 d_{1s}^2 d_{1t}^2\right]$.

Assumption A.4. (i) Let $\sigma^2(x) = E[e_t^2|X_t = x]$ and $\mu_4(x) = E[e_t^4|X_t = x]$. Assume that $\sigma^2(x)$ and $\mu_4(x)$ satisfy some Lipschitz conditions:

$$\left|\sigma^2(u+v) - \sigma^2(u)\right| \le D(u)||v|| \text{ and } |\mu_4(u+v) - \mu_4(u)| \le D(u)||v||$$

with $v \in S$ (a compact set of R^q) and $E\left[|D(X_t)|^{2+\zeta}\right] < \infty$ for some small $\zeta > 0$, where $||\cdot||$ denotes the Euclidean norm.

(ii) Assume that the first two derivatives of $m(\cdot)$ and $m_1(\cdot)$ exist and are bounded.

(iii) Let $f_{\tau_1,\tau_2,\cdots,\tau_l}(\cdot)$ be the joint probability density of $(X_{1+\tau_1}, \ldots, X_{1+\tau_l})$ $(1 \le l \le 4)$. Assume that $f_{\tau_1,\tau_2,\cdots,\tau_l}(\cdot)$ exists and satisfies the following Lipschitz condition:

$$|f_{\tau_1,\tau_2,\cdots,\tau_l}(x_1 + v_1, \cdots, x_l + v_l) - f_{\tau_1,\tau_2,\cdots,\tau_l}(x_1, \cdots, x_l)| \le D_{\tau_1,\cdots,\tau_l}(x_1, \cdots, x_l)||v||$$

for $v \in S$, where S is a compact subset and $D_{\tau_1,\cdots,\tau_l}(x_1, \cdots, x_l)$ is integrable and satisfies the following conditions

$$\int D_{\tau_1,\cdots,\tau_l}(x_1, \cdots, x_l)||x||^{2\theta} dx < M_1 < \infty,$$

$$\int D_{\tau_1,\cdots,\tau_l}(x_1, \cdots, x_l) f_{\tau_1,\tau_2,\cdots,\tau_l}(x_1, \cdots, x_l) dx < M_2 < \infty$$

for some $\theta > 1$ and constants $M_1 > 0$ and $M_2 > 0$.

Assumption A.5. (i) Assume that the univariate kernel function $k(\cdot)$ is bounded and symmetric with $\int k(u)du = 1$, $\int uk(u)du = 0$ and $\int u^2 k(u)du < \infty$. In addition, $k(x)$ is continuous on $R^1 = (-\infty, \infty)$. This paper considers using

$$K(x_1, \cdots, x_p) = \prod_{i=1}^{p} k(x_i).$$

(ii) The bandwidth parameter h satisfies that

$$\lim_{T\to\infty} Th^{\frac{5p}{2}} = \infty \text{ and } \limsup_{T\to\infty} Th^{5p} < \infty.$$

Before introducing the following assumption, we need to give some notation.

Let $m^{(\mu)}$ be the μ-order derivative of function $m(\cdot)$ and C_0 be a constant. Define

$$M_\mu = \left\{m : \left|m^{(\mu)}(s_1) - m^{(\mu)}(s_2)\right| \le C_0||s_1 - s_2||\right\},$$

where $s_1, s_2 \in S$ and S is a compact subset of R^p.

Assumption A.6. (i) For $m \in M_\mu$ and $\{z_j(\cdot) : j = 1, 2, \ldots\}$ given above, there exists a vector of unknown parameters $\gamma = (\gamma_1, \ldots, \gamma_k)^\tau$ such that for a constant C_0 $(0 \le C_0 < \infty)$ independent of T

$$k^{2(\mu+1)+p} E\left[\sum_{j=1}^{k} z_j(X_t)\gamma_j - m(X_t)\right]^2 \approx C_0$$

where $\mu + 1 > p$.

(ii) The truncation parameter k is chosen as $k = \left[h^{-1}\right]$ with h defined in Assumption A.5., where $[x] \le x$ denotes the largest integer part of x.

(iii) Z is of full column rank k. Each $z_i(x)$ is continuous with $\sup_{(x,i)} |z_i(x)| < \infty$.

(iv) Assume that $0 < c_i^2 = E[z_i^2(X_t)] < \infty$ exists and that

$$E[z_i(X_s)z_i(X_t)] = 0 \text{ and } E[z_i(X_t)z_j(X_t)] = 0$$

for all $i \ge 1$, $i \ne j$ and $s \ne t$.

Some detailed remarks on the assumptions are relegated to Appendix D.

B. Appendix B

This appendix lists a very general lemma for the proof of the main results given in Section 3. The lemma establishes central limit theorems for degenerate U–statistics of strongly dependent processes.

B.1. A technical lemma

Lemma B.1. Let ξ_t be a r-dimensional strictly stationary and strong mixing (α–mixing) stochastic process. Let $\theta(\cdot, \cdot)$ be a symmetric Borel function defined on $R^r \times R^r$. Assume that $E[\theta(\xi_s, \xi_t)] = 0$ for all $1 \le s, t \le T$ and further that for any fixed $x, y \in R^r$, $E[\theta(\xi_1, y)] = E[\theta(x, \xi_1)] = 0$. Let $\theta_{st} = \theta(\xi_s, \xi_t)$ and $\sigma_T^2 = \sum_{1\le s<t\le T} \mathrm{Var}[\theta_{st}]$. For some small constant $0 < \delta < 1$, let

$$\begin{aligned} M_{T1} &= \max_{1\le i<j<k\le T} \max\left\{E|\theta_{ik}\psi_{jk}|^{1+\delta}, \int |\theta_{ik}\theta_{jk}|^{1+\delta} dP(\xi_i)dP(\xi_j, \xi_k)\right\}, \\ M_{T21} &= \max_{1\le i<j<k\le T} \max\left\{E|\theta_{ik}\theta_{jk}|^{2(1+\delta)}, \int |\theta_{ik}\theta_{jk}|^{2(1+\delta)} dP(\xi_i)dP(\xi_j, \xi_k)\right\}, \end{aligned}$$

$$
\begin{aligned}
M_{T22} &= \max_{1\le i<j<k\le T} \max\left\{ \int |\theta_{ik}\theta_{jk}|^{2(1+\delta)} dP(\xi_i,\xi_j)dP(\xi_k), \int |\theta_{ik}\theta_{jk}|^{2(1+\delta)} dP(\xi_i)dP(\xi_j)dP(\xi_k) \right\}, \\
M_{T3} &= \max_{1\le i<j<k\le T} E|\theta_{ik}\theta_{jk}|^2, \quad M_{T4} = \max_{\substack{1<i,j,k\le 2T \\ i,j,k \ \textit{different}}} \left\{ \max_P \int |\theta_{1i}\theta_{jk}|^{2(1+\delta)} dP \right\},
\end{aligned}
$$

where the maximization over P in the equation for M_{T4} is taken over the four probability measures $P(\xi_1,\xi_i,\xi_j,\xi_k)$, $P(\xi_1)P(\xi_i,\xi_j,\xi_k)$, $P(\xi_1)P(\xi_{i_1})P(\xi_{i_2},\xi_{i_3})$, and $P(\xi_1)P(\xi_i)P(\xi_j)P(\xi_k)$, where (i_1,i_2,i_3) is the permutation of (i,j,k) in ascending order;

$$
\begin{aligned}
M_{T51} &= \max_{1\le i<j<k\le T} \max\left\{ E\left| \int \theta_{ik}\theta_{jk}\theta_{ik}\theta_{jk} dP(\xi_i) \right|^{2(1+\delta)} \right\}, \\
M_{T52} &= \max_{1\le i<j<k\le T} \max\left\{ \int \left| \int \theta_{ik}\theta_{jk}\theta_{ik}\theta_{jk} dP(\xi_i) \right|^{2(1+\delta)} dP(\xi_j)dP(\xi_k) \right\}, \\
M_{T6} &= \max_{1\le i<j<k\le T} E\left| \int \theta_{ik}\theta_{jk} dP(\xi_i)\right|^2, \quad M_{T7} = \max_{1\le i<j<T} E\left[|\theta_{ij}|^{1+\delta} \right].
\end{aligned}
$$

Assume that all the $M_T's$ are finite. Let

$$
\begin{aligned}
M_T &= \max\left\{ T^2 M_{T1}^{\frac{1}{1+\delta}}, T^2 M_{T51}^{\frac{1}{2(1+\delta)}}, T^2 M_{T52}^{\frac{1}{2(1+\delta)}}, T^2 M_{T6}^{\frac{1}{2}} \right\}, \\
N_T &= \max\left\{ T^{\frac{3}{2}} M_{T21}^{\frac{1}{2(1+\delta)}}, T^{\frac{3}{2}} M_{T22}^{\frac{1}{2(1+\delta)}}, T^{\frac{3}{2}} M_{T3}^{\frac{1}{2}}, T^{\frac{3}{2}} M_{T4}^{\frac{1}{2(1+\delta)}}, T^{\frac{3}{2}} M_{T7}^{\frac{1}{1+\delta}} \right\}.
\end{aligned}
$$

If $\lim_{T\to\infty} \frac{\max\{M_T,N_T\}}{\sigma_T^2} = 0$, then

$$
\frac{1}{\sigma_T} \sum_{1\le s<t\le T} \theta(\xi_s,\xi_t) \to_D N(0,1) \quad \text{as} \quad T\to\infty.
$$

Remark B.1. Lemma B.1 establishes central limit theorems for degenerate U–statistics of strongly dependent processes. The lemma extends and complements some existing results for the β–mixing case. See for example, Lemma 3.2 of Hjellvik, Yao and Tjøstheim (1998) and Theorem 2.1 of Fan and Li (1999). It should be pointed out that the conclusion of Lemma B.1 remains true when the usual martingale assumption that $E[\phi(\xi_i,\xi_j)|I_j] = 0$ for any $i<j$ is removed, where I_t is a sequence of σ–field generated by $\{\xi_s : 1\le s\le t\}$. Such a martingale assumption is used only for a direct application of an existing central limit theorem (CLT) for martingales.

Proof of Lemma B.1: For a given constant $0<\rho_0\le \frac{1}{4}$, choose $q=[T^{\rho_0}]>2$ as the largest integer part of T^{ρ_0}. Obviously, $\sum_{T=1}^{\infty} e^{-d_0 q_T} < \infty$ for any given $d_0>0$. Recall the notation of θ_{st} and define

$$
\phi_{st} = \theta_{st} - E\left[\theta_{st}|I_{t-q}\right] \quad \text{and} \quad \psi_{st} = E\left[\theta_{st}|I_{t-q}\right]. \tag{B.1}
$$

Observe that

$$
\begin{aligned}
L_T &= \sum_{t=2}^{T}\sum_{s=1}^{t-1} \theta_{st} = \sum_{t=q+1}^{T}\sum_{s=1}^{t-q} \phi_{st} + \sum_{t=q+1}^{T}\sum_{s=1}^{t-q} \psi_{st} \\
&+ \sum_{t=2}^{T}\sum_{s=t+1-q}^{t-1} \phi_{st} + \sum_{t=2}^{T}\sum_{s=t+1-q}^{t-1} \psi_{st} \equiv \sum_{j=1}^{4} L_{jT}.
\end{aligned} \tag{B.2}
$$

To establish the asymptotic distribution of L_T, it suffices to show that as $T \to \infty$

$$\frac{L_{1T}}{\sigma_T} \to N(0,1) \quad \text{and} \quad \frac{L_{jT}}{\sigma_T} \to_p 0 \quad \text{for} \quad j = 2,3,4. \tag{B.3}$$

Let $V_t = \sum_{s=1}^{t-q} \phi_{st}$. Then $E[V_t | I_{t-q}] = 0$. This implies that V_t is a martingale difference with respect to I_{t-q}. We now start proving the first part of (B.3). Applying a central limit theorem for martingale sequences (see Theorem 1 of Chapter VIII of Pollard 1984), in order to prove the first part of (B.3), it suffices to show that

$$\frac{1}{\sigma_T^2} \sum_{t=q+1}^{T} V_t^2 \to_p 1 \quad \text{and} \quad \frac{1}{\sigma_T^4} \sum_{t=q+1}^{T} E\left[V_t^4\right] \to 0. \tag{B.4}$$

To verify (B.4), we first need to calculate some useful quantities. Recall the definition of V_t and observe that

$$\begin{aligned} V_t^2 &= \sum_{s=1}^{t-q} \phi_{st}^2 + 2 \sum_{s_1=2}^{t-q} \sum_{s_2=1}^{s_1-1} \phi_{s_1 t}\phi_{s_2 t} \\ \sum_{t=q+1}^{T} E[V_t^2] &= \sum_{t=q+1}^{T} \sum_{s=1}^{t-q} E[\phi_{st}^2] + 2 \sum_{t=q+2}^{T} \sum_{s_1=2}^{t-q} \sum_{s_2=1}^{s_1-1} E\left[\phi_{s_1 t}\phi_{s_2 t}\right] \equiv \sigma_{1T}^2 + \Delta_{1T}. \end{aligned} \tag{B.5}$$

We now show that as $T \to \infty$

$$\sigma_{1T}^2 = \sigma_T^2 \,(1 + o(1)) \quad \text{and} \quad \Delta_{1T} = o\left(\sigma_T^2\right). \tag{B.6}$$

By Lemma C.1 (with $\eta_1 = \phi_{s_1 t}$, $\eta_2 = \phi_{s_1 t}$, $l = 2$, $p_i = 2(1+\delta)$ and $Q = \frac{1}{1+\delta}$),

$$E\left|\phi_{s_1 t}\phi_{s_2 t}\right| \le 10 M_{T1}^{\frac{1}{1+\delta}} \beta^{\frac{\delta}{1+\delta}}(s_1 - s_2).$$

Therefore,

$$\Delta_{1T} \le 10 T^2 M_{T1}^{\frac{1}{1+\delta}} \sum_{i=1}^{T} \alpha^{\frac{\delta}{1+\delta}}(i) \le C T^2 M_{T1}^{\frac{1}{1+\delta}} \tag{B.7}$$

using $\sum_{i=1}^{\infty} \alpha^{\frac{\delta}{1+\delta}}(i) < \infty$. This, together with the conditions of Lemma B.1, impliles that $\Delta_{1T} = o\left(\sigma_T^2\right)$ as $T \to \infty$.

We now start to verify the first part of (B.6). Let $\sigma_{st}^2 = E[\phi_{st}^2]$. Observe that

$$\begin{aligned} E\left(\sum_{t=q+1}^{T} V_t^2 - \sigma_{1T}^2\right)^2 &\le 2E\left\{\sum_{t=q+1}^{T} \sum_{s=1}^{t-q} \left[\phi_{st}^2 - \sigma_{st}^2\right]\right\}^2 \\ &+ 8E\left\{\sum_{t=q+2}^{T} \sum_{s_1=2}^{t-q} \sum_{s_2=1}^{s_1-1} \phi_{s_1 t}\phi_{s_2 t}\right\}^2 \equiv Q_{1T} + Q_{2T}. \end{aligned} \tag{B.8}$$

In the following, we first show that as $T \to \infty$

$$Q_{2T} = o\left(\sigma_T^4\right). \tag{B.9}$$

Using Lemma C.1 again, we can show that as $T \to \infty$

$$\begin{aligned} Q_{2T} &= 8E\left\{\sum_{t=q+2}^{T}\sum_{s_1=2}^{t-q}\sum_{s_2=1}^{s_1-1}\phi_{s_1 t}\phi_{s_2 t}\right\}^2 \\ &\leq 8\left(\sum_{t_1\neq t_2}\sum_{s_1\neq s_2}\sum_{r_1\neq r_2}\left|E\left[\phi_{s_1t_1}\phi_{s_2t_1}\phi_{r_1t_2}\phi_{r_2t_2}\right]\right|\right) \\ &\leq 8\max\left\{M_T^2, N_T^2\right\} = o\left(\sigma_T^4\right) \end{aligned}$$

under the conditions of Lemma B.1.

Let $C_\phi = \int \phi_{12}^2\phi_{34}^2 dP_1(\xi_1)dP_1(\xi_2)dP_1(\xi_3)dP_1(\xi_4)$, where $P_1(\xi_i)$ denotes the probability measure of ξ_i.

Using Lemma C.1 repeatedly, we have that for different i, j, k, l

$$\begin{aligned} \left|E\left[\phi_{ij}^2\phi_{kl}^2\right] - C_\phi\right| &\leq 10\left\{\alpha(\Delta(i,j,k,l))\right\}^{1-\frac{1}{1+\delta}} M_{T4}^{\frac{1}{1+\delta}} \\ &= 10M_{T4}^{\frac{1}{1+\delta}}\left\{\alpha(\Delta(i,j,k,l))\right\}^{\frac{\delta}{1+\delta}}, \end{aligned} \tag{B.10}$$

where $\Delta(i, j, k, l)$ is the minimum increment in the sequence which is the permutation of i, j, k, l in ascending order.

Similarly to (B.10), we can have for all different i, j, k, l

$$\left|\sigma_{ij}^2\sigma_{kl}^2 - C_\phi\right| \leq 10M_{T4}^{\frac{1}{1+\delta}}\left\{\alpha(\Delta(i,j,k,l))\right\}^{\frac{\delta}{1+\delta}}. \tag{B.11}$$

Therefore, using (B.10) and (B.11),

$$\begin{aligned} Q_{1T} &= 2E\left\{\sum_{t=q+2}^{T}\sum_{s=1}^{t-q}\left[\phi_{st}^2 - \sigma_{st}^2\right]\right\}^2 \\ &\leq 2\left(\sum_{t_1,t_2}\sum_{s_1,s_2}\left|E\left[\phi_{ij}^2\phi_{kl}^2\right] - \sigma_{ij}^2\sigma_{kl}^2\right|\right) \\ &\leq 2\left(\sum_{t_1,t_2}\sum_{s_1,s_2}\left|E\left[\phi_{ij}^2\phi_{kl}^2\right] - C_\phi\right| + \left|C_\phi - \sigma_{ij}^2\sigma_{kl}^2\right|\right) \\ &\leq \left\{O\left(T^3M_{T4}^{\frac{1}{1+\delta}}\right) + O\left(T^3M_{T3}\right)\right\} = o(\sigma_T^4). \end{aligned} \tag{B.12}$$

It now follows from (B.8)–(B.12) that for any $\epsilon > 0$

$$P\left\{\left|\frac{1}{\sigma_{1T}^2}\sum_{t=q+1}^{T}V_t^2 - 1\right| \geq \epsilon\right\} \leq \frac{1}{\sigma_T^4\epsilon^2}E\left[\sum_{t=q+1}^{T}V_t^2 - \sigma_{1T}^2\right]^2 \to 0. \tag{B.13}$$

Thus, the first part of (B.4) is proved.

Note that for $q+1 \le k \le T$,

$$\begin{aligned}
E[V_k^4] &= E\left\{\sum_{i=1}^{k-q}\phi_{it}^2 + 2\sum_{1\le i<j<k-q}\phi_{ik}\phi_{jk}\right\}^2 \\
&= E\left\{\sum_{i=1}^{k-q}\phi_{ik}^4 + 6\sum_{1\le i<j<k-q}\phi_{ik}^2\phi_{jk}^2 + 4\sum_{l=1}^{k-q}\sum_{1\le i<j<k-q}\phi_{lk}^2\phi_{ik}\phi_{jk}\right\} \\
&+ 4E\left\{\sum_{q\le i<j<k, 1\le s<t<k-q, (i,j)\neq(s,t)}\phi_{ik}\phi_{jk}\phi_{sk}\phi_{tk}\right\} \\
&= 4\sum_{l=1}^{k-q}\sum_{1\le i<j<k-q} E\left[\phi_{lk}^2\phi_{ik}\phi_{jk}\right] + 4\sum_{1\le i<j<k-q,\ 1\le s<t<k-q, (i,j)\neq(s,t)} E\left[\phi_{ik}\phi_{jk}\phi_{sk}\phi_{tk}\right] \\
&+ O\left(T^2 M_{T3}\right). \qquad \text{(B.14)}
\end{aligned}$$

It is easy to see that

$$\int |\phi_{ik}\phi_{jk}\phi_{sk}\phi_{tk}|^{1+\delta}\,dP \le \left\{\int |\phi_{ik}\phi_{jk}|^{2(1+\delta)}\,dP \int |\phi_{sk}\phi_{tk}|^{2(1+\delta)}\,dP\right\}^{1/2} \le M_{T4}.$$

Similarly to (B.10), we can have for any $(i,j)\neq(s,t)$,

$$|E[\phi_{ik}\phi_{jk}\phi_{sk}\phi_{tk}]| \le 10 M_{T4}^{\frac{1}{1+\delta}}\left\{\alpha(\Delta(i,j,s,t))\right\}^{\frac{\delta}{1+\delta}}, \qquad \text{(B.15)}$$

where $\Delta(\cdot)$ is as defined before.

Consequently,

$$\sum_{k=q+1}^{T} E[V_k^4] = O\left(T^3 M_{T4}^{\frac{1}{1+\delta}}\right) = o(\sigma_T^4). \qquad \text{(B.16)}$$

This finishes the proof of the first part of (B.4), and therefore the proof of (B.4).

Applying Lemma C.3 implies that as $T\to\infty$

$$E|L_{2T}| \le \sum_{t=q+1}^{T}\sum_{s=1}^{t-q} E\left|E\left[\theta_{st}|I_{t-q}\right]\right| \le C\left(TqM_{T7}^{\frac{1}{1+\delta}}\right) = o(\sigma_T^2) \qquad \text{(B.17)}$$

using the conditions of Lemma B.1.

The second part of (B.3) for L_{4T} follows from the conditions of Lemma B.1 and

$$\begin{aligned}
E|L_{4T}| &\le \sum_{t=2}^{T}\sum_{s=t+1-q}^{t-1} E\left(E\left[|\theta_{st}|\,|I_{t-q}\right]\right) \\
&= \sum_{t=2}^{T}\sum_{s=t+1-q}^{t-1} E\left[|\theta_{st}|\right] \le \left(TqM_{T7}^{\frac{1}{1+\delta}}\right) = o(\sigma_T^2) \qquad \text{(B.18)}
\end{aligned}$$

We finally prove the second part of (B.3) for L_{3T}. Similarly, using Lemma C.1, we can show that as $T\to\infty$

$$\left|\sum_{t=2}^{T}\sum_{s_1=t+1-q}^{t-1}\sum_{s_2\neq s_1, s_2=t+1-q}^{t-1} E\left[\phi_{s_1t}\phi_{s_2t}\right]\right| \le \sum_{t=2}^{T}\sum_{s_1=t+1-q}^{t-1}\sum_{s_2\neq s_1, s_2=t+1-q}^{t-1} E\left[|\phi_{s_1t}\phi_{s_2t}|\right]$$

$$\leq \quad o\left(T^2 q \; M_{T3}\right),$$

$$\left|\sum_{t_1=3}^{T}\sum_{t_2=t_1+1-q}^{t_1-1}\sum_{s_1=t_1+1-q}^{t_1-1}\sum_{s_2=t_2+1-q}^{t_2-1} E\left[\phi_{s_1 t_1}\phi_{s_2 t_2}\right]\right| \leq o\left(T^2 q^2 \; M_{T3}\right). \tag{B.19}$$

Using (B.19) implies that as $T \to \infty$

$$\begin{aligned} E\left[L_{3T}^2\right] &= \sum_{t=2}^{T}\sum_{s=t+1-q}^{t-1} E\left[\phi_{st}^2\right] + \sum_{t=2}^{T}\sum_{s_1=t+1-q}^{t-1}\sum_{s_2\neq s_1, s_2=t+1-q}^{t-1} E\left[\phi_{s_1 t}\phi_{s_2 t}\right] \\ &+ 2\sum_{t_1=q+2}^{T}\sum_{t_2=2}^{t_1-q}\sum_{s_1=t_1+1-q}^{t_1-1}\sum_{s_2=t_2+1-q}^{t_2-1} E\left[\phi_{s_1 t_1}\phi_{s_2 t_2}\right] \\ &+ 2\sum_{t_1=2}^{T}\sum_{t_2=t_1+1-q}^{t_1-1}\sum_{s_1=t_1+1-q}^{t_1-1}\sum_{s_2=t_2+1-q}^{t_2-1} E\left[\phi_{s_1 t_1}\phi_{s_2 t_2}\right] \\ &= \sum_{t=2}^{T}\sum_{s=t+1-q}^{t-1} E\left[\phi_{st}^2\right] + \sum_{t=2}^{T}\sum_{s_1=t+1-q}^{t-1}\sum_{s_2\neq s_1, s_2=t+1-q}^{t-1} E\left[\phi_{s_1 t}\phi_{s_2 t}\right] \\ &+ 2\sum_{t_1=3}^{T}\sum_{t_2=t_1+1-q}^{t_1-1}\sum_{s_1=t_1+1-q}^{t_1-1}\sum_{s_2=t_2+1-q}^{t_2-1} E\left[\phi_{s_1 t_1}\phi_{s_2 t_2}\right] \\ &= O\left(T^3 M_{T3}\right) + O\left(T^2 q^2 M_{T3}\right) = o\left(\sigma_T^4\right) \end{aligned} \tag{B.20}$$

noting that the third term of (B.20) is zero because of $E[\phi_{s_2 t_2}\phi_{s_1 t_1}|I_{t_1-q}] = 0$. This completes the proof of Lemma B.1.

B.2. Proof of Theorem 3.1

Let $\xi_t = (e_t, X_t^\tau)$, $\theta(\xi_s, \xi_t) = p_{st} e_s e_t$ and $\sigma_T^2 = \sum_{1\leq s<t\leq T} \mathrm{Var}(p_{st} e_s e_t)$, where $p_{st} = p(X_s, X_t)$ is a symmetric and continuous function as defined in Assumption A.3.

In order to apply Lemma B.1, one needs to justify the conditions of Lemma B.1 hold for $\theta(\xi_s, \xi_t) = p_{st} e_s e_t$. We now verify only the following condition listed in Lemma B.1,

$$\frac{\max\{M_T, N_T\}}{\sigma_T^2} \to 0 \quad \text{as } T \to \infty,$$

as the other conditions can be justified similarly.

For the M_T part, one justifies only

$$\frac{T^2 M_{T1}^{\frac{1}{1+\delta}}}{\sigma_T^2} \to 0 \quad \text{as } T \to \infty.$$

The others follow similarly.

It follows that for some $0 < \delta < 1$ and $1 \leq i < j < k \leq T$

$$E\left[|\theta_{ik}\theta_{jk}|^{1+\delta}\right] = E\left[|e_i e_j e_k^2 p_{ik} p_{jk}|^{1+\delta}\right]$$

$$\leq \left\{E\left[|e_i e_j e_k^2|^{2(1+\delta)(1+\delta_2)}\right]\right\}^{\frac{1}{2(1+\delta_2)}} \left\{E\left[|p_{ij} p_{ik}|^{(1+\delta)(1+\delta_1)}\right]\right\}^{\frac{1}{(1+\delta_1)}}.$$

Since $0 < \delta_1 < 1$ and $0 < \delta_2 < 1$ satisfy $\frac{1}{1+\delta_1} + \frac{1}{2(1+\delta_2)} = 1$ and $\frac{1+\delta}{3-\delta} < \delta_1 < \frac{1-\delta}{1+\delta}$, we have that

$$1 < \zeta = (1+\delta)(1+\delta_2) < 2 \quad \text{and} \quad 1 < \eta = (1+\delta)(1+\delta_1) < 2.$$

By the second part of Assumption A.1(iii) and Assumption A.2(iii), one can have

$$\frac{T^2 M_{T1}^{\frac{1}{1+\delta}}}{\sigma_T^2} \to 0 \quad \text{as } T \to \infty.$$

Similarly, one can verify the above is true for the second part of M_{T1}.

For the N_T part, one needs only to use Assumptions A.1(iii) and A.2(i)(ii) to show that

$$\frac{N_T}{\sigma_T^2} \to 0 \quad \text{as } T \to \infty.$$

We consider the first part of M_{T21} for example. For $1 \le i < j < k \le T$ and $0 < \delta < 1$

$$E\left[|\theta_{ik}\theta_{jk}|^{2(1+\delta)}\right] = E\left[|e_i e_j e_k^2 p_{ik} p_{jk}|^{2(1+\delta)}\right]$$

$$\le C\left(\frac{q}{T}\right)^{4(1+\delta)} E\left[|e_i e_j e_k^2|^{2(1+\delta)}\right]$$

using Assumption A.2(i).

Thus as $T \to \infty$

$$\frac{T^{\frac{3}{2}} M_{T21}^{\frac{1}{2(1+\delta)}}}{\sigma_T^2} \le C\frac{T^{\frac{3}{2}}}{\sigma_T^2}\left(\frac{q}{T}\right)^2 = C\frac{q^2}{\sqrt{T}\sigma_T^2} \to 0$$

using Assumptions A.1(iii) and A.2(ii).

Analogously, one can verify the other parts of N_T.

It follows that the conditions of Lemma B.1 hold. Thus as $T \to \infty$

$$\frac{\sum_{1\le s<t\le T} p_{st} e_s e_t}{\sigma_T} \to_D N(0,1). \tag{B.9}$$

Thus, in order to finish the proof of Theorem 3.1, it suffices to show that as $T \to \infty$

$$\frac{\hat{\sigma}_T^2}{\sigma_T^2} - 1 \to_p 0, \tag{B.10}$$

where $\hat{\sigma}_T^2 = \sum_{1\le s<t\le T} p_{st}^2 e_s^2 e_t^2$.

In view of the definition of $\hat{\sigma}_T^2$ and σ_T^2, in order to prove (B.10), it suffices to show that

$$\sigma_T^{-2} \sum_{1\le s<t\le T} \left\{p_{st}^2 e_s^2 e_t^2 - E\left[p_{st}^2 e_s^2 e_t^2\right]\right\} \to_p 0. \tag{B.11}$$

The proof of (B.11) is the same as that of the first part of (B.4). This finally finishes the proof of Theorem 3.1.

B.3. Proof of Theorems 3.2–3.4.

Let $w_t = Y_t - U_t^\tau \beta$, $u_t = Y_t - X_t^\tau \beta$,

$$\tilde{L}_{2T} = \frac{\sum_{t=1}^T \sum_{s \neq t} p_{2st} w_s w_t}{\tilde{\sigma}_{2T}} \text{ and } \tilde{L}_{4T} = \frac{\sum_{t=1}^T \sum_{s \neq t} p_{st} u_s u_t}{\tilde{\sigma}_{4T}},$$

where $\tilde{\sigma}_{2T}^2 = 2\sum_{t=1}^T \sum_{s=1}^T p_{2st}^2 w_s^2 w_t^2$ and $\tilde{\sigma}_{4T}^2 = 2\sum_{t=1}^T \sum_{s=1}^T p_{st}^2 u_s^2 u_t^2$.

In order to prove Theorems 3.3–3.4, one needs first to show that

$$L_{2T} = \tilde{L}_{2T} + o_p(1) \text{ and } L_{4T} = \tilde{L}_{4T} + o_p(1). \qquad (B.12)$$

The remainder of the proof follows from Theorem 3.1. As the proof of (B.12) is much simpler than that of Theorem 3.2, we give only the proof of Theorem 3.2 in some detail.

Let $\eta_{1t} = d_{1Tt} - d_{1t}$ and $\delta_{1t} = m_1(U_t) - \hat{m}_1(U_t)$. Note that

$$\hat{Y}_t = [e_t + \delta_{1t}] d_{1Tt} = e_t d_{1t} + e_t \eta_{1t} + \delta_{1t} d_{1Tt},$$

$$\hat{Y}_s \hat{Y}_t = e_s e_t d_{1s} d_{1t} + e_s e_t \eta_{1s} \eta_{1t} + \delta_{1s} d_{1Ts} \delta_{1t} d_{1Tt}$$

$$+ e_s e_t \eta_{1s} d_{1t} + e_s e_t d_{1s} \eta_{1t} + \delta_{1s} d_{1Ts} e_t d_{1t} + d_{1Ts} \delta_{1s} e_t \eta_{1t} + d_{1Tt} \delta_{1t} e_s \eta_{1s} + d_{1Tt} \delta_{1t} e_s \eta_{1s}$$

$$\equiv e_s e_t d_{1s} d_{1t} + e_s e_t \eta_{1s} \eta_{1t} + \delta_{1s} d_{1Ts} \delta_{1t} d_{1Tt} + r_{st},$$

where

$$r_{st} = e_s e_t \eta_{1s} d_{1t} + e_s e_t d_{1s} \eta_{1t} + \delta_{1s} d_{1Ts} e_t d_{1t} + d_{1Ts} \delta_{1s} e_t \eta_{1t} + d_{1Tt} \delta_{1t} e_s \eta_{1s} + d_{1Tt} \delta_{1t} e_s \eta_{1s}.$$

Now one can have the following decomposition

$$L_{1T} = \left[\widetilde{L}_{1T} + \Delta_{1T} \right] \frac{\hat{\sigma}_{1T}}{\hat{S}_{1T}},$$

where

$$\widetilde{L}_{1T} = \frac{1}{\hat{\sigma}_{1T}} \sum_{s \neq t} p_{st} e_s e_t d_{1s} d_{1t} \text{ and } \Delta_{1T} = \frac{1}{\hat{\sigma}_{1T}} \sum_{s \neq t} p_{st} \left[e_s e_t \eta_{1s} \eta_{1t} + \delta_{1s} d_{1Ts} \delta_{1t} d_{1Tt} + r_{st} \right]. \qquad (B.13)$$

In view of (B.13), in order to prove

$$L_{1T} = \widetilde{L}_{1T} + o_p(1),$$

it suffices to show that as $T \to \infty$

$$\Delta_{1T} \to_p 0 \text{ and } \frac{\hat{\sigma}_{1T}}{\hat{S}_{1T}} \to_p 1. \qquad (B.14)$$

Let $\epsilon_t = e_t d_{1t}$, $\hat{\sigma}_{1T} = 2\sum_{s=1}^T \sum_{t=1}^T p_{st}^2 \epsilon_s^2 \epsilon_t^2$ and $\sigma_{1T}^2 = 2\sum_{s=1}^T \sum_{t=1}^T E\left[p_{st}^2 \epsilon_s^2 \epsilon_t^2\right]$. In view of (B.13) and (B.14), it suffices to show that

$$\frac{\sum_{s \neq t} p_{st} e_s e_t \eta_{1s} \eta_{1t}}{\sigma_{1T}} \to_p 0, \; \frac{\sum_{s \neq t} p_{st} \delta_{1s} d_{1Ts} \delta_{1t} d_{1Tt}}{\sigma_{1T}} \to_p 0, \; \frac{\sum_{s \neq t} p_{st} r_{st}}{\sigma_{1T}} \to_p 0, \qquad (B.15)$$

$$\frac{\hat{S}_{1T}^2 - \hat{\sigma}_{1T}^2}{\sigma_{1T}^2} \to_p 0 \text{ and } \frac{\hat{\sigma}_{1T}^2}{\sigma_{1T}^2} \to_p 1. \qquad (B.16)$$

We then prove only the first two parts of (B.15) and (B.16). The proof of the third part of (B.15) is similar to that of the first part of (B.15).

Obviously,

$$E\left|\sum_{s\neq t} p_{st} e_s e_t \eta_{1s}\eta_{1t}\right| \leq \sum_{s\neq t} E\left|p_{st} e_s e_t \eta_{1s}\eta_{1t}\right|$$

$$\leq \frac{1}{2}\sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} E\left\{\left[e_s^2\eta_{1s}^2 + e_t^2\eta_{1t}^2\right]|p_{st}|\right\} = \sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} E\left\{e_s^2\eta_{1s}^2|p_{st}|\right\} = o(\sigma_{1T})$$

using Assumption A.3(iii). This implies that the first part of (B.15) holds.

Similarly, one can have

$$E\left|\sum_{s\neq t} p_{st}\delta_{1s}d_{1Ts}\delta_{1t}d_{1Tt}\right| \leq \sum_{s\neq t} E\left|p_{st}\delta_{1s}d_{1Ts}\delta_{1t}d_{1Tt}\right|$$

$$\leq \frac{1}{2}\sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} E\left\{\left[\delta_{1s}^2 d_{1Ts}^2 + \delta_{1t}^2 d_{1Tt}^2\right]|p_{st}|\right\} = \sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} E\left\{\delta_{1s}^2 d_{1Ts}^2|p_{st}|\right\} = o(\sigma_{1T})$$

using Assumption A.3(iii). This implies that the second part of (B.15) holds.

In view of the definition of $\hat{\sigma}_{1T}^2$ and σ_{1T}^2, in order to prove (B.16), it suffices to show that

$$\sigma_{1T}^{-2}\sum_{1\leq s<t\leq T}\left\{p_{st}^2\left[\hat{Y}_s^2\hat{Y}_t^2 - \epsilon_s^2\epsilon_t^2\right]\right\} \to_p 0 \qquad (B.17)$$

and

$$\sigma_{1T}^{-2}\sum_{1\leq s<t\leq T}\left\{p_{st}^2\epsilon_s^2\epsilon_t^2 - E\left[p_{st}^2\epsilon_s^2\epsilon_t^2\right]\right\} \to_p 0, \qquad (B.18)$$

where $\epsilon_t = e_t d_{1t}$.

The proof of (B.18) follows similarly from that of (B.11). In view of the definition of $\hat{Y}_t$, in order to prove (B.17), one proves only

$$J_{1T} = \sum_{1\leq s<t\leq T} p_{st}^2 e_s^2\eta_{1s}^2 e_t^2\eta_{1t}^2 = o_p(\sigma_{1T}^2) \qquad (B.19)$$

and

$$J_{2T} = \sum_{1\leq s<t\leq T} p_{st}^2\delta_{1s}^2 d_{1Ts}^2\delta_{1t}^2 d_{1Tt}^2 = o_p(\sigma_{1T}^2), \qquad (B.20)$$

as the other parts follow similarly.

The proof of (B.19) follows from

$$J_{1T} \leq \max_{1\leq t\leq T}\left|\frac{d_{1Tt}}{d_{1t}} - 1\right|^4 \sum_{1\leq s<t\leq T} p_{st}^2 e_s^2 d_{1s}^2 e_t^2 d_{1t}^2 = o_p(\sigma_{1T}^2)$$

using Assumption A.3(ii). The proof of (B.20) follows from

$$E[J_{2T}] = \sum_{1\leq s<t\leq T} E\left[p_{st}^2\delta_{1s}^2 d_{1Ts}^2\delta_{1t}^2 d_{1Tt}^2\right] = o(\sigma_{1T}^2)$$

using Assumption A.3(iii). The proof of

$$\widetilde{L}_{1T} \to N(0,1)$$

follows from that of Theorem 3.1. This finally finishes the proof of Theorem 3.2.

B.4. Proof of Corollaries 3.1–3.4.

The proof of Corollary 3.1 follows from that of Theorem 3.1 immediately, as either Assumptions A.4 and A.5 or Assumptions A.4 and A.6 implies Assumption A.3.

The proof of Corollary 3.2 follows from Theorem 3.2 immediately. For the proof of Corollary 3.3, one needs to follow the proof of (B.10) to show that

$$\tilde{L}_{2T} = L_{2T} + o_p(1),$$

where $\tilde{L}_{2T} = \frac{\sum_{t=1}^{T}\sum_{s\neq t} p(u_{s-1},u_{t-1})u_s u_t}{\tilde{S}_{2T}}$, $\tilde{S}_{2T}^2 = 2\sum_{t=1}^{T}\sum_{s=1}^{T} p(u_{s-1},u_{t-1})^2 u_s^2 u_t^2$, and $u_s = Y_s - X_s^\tau \beta$.

In the detailed proof, the continuity of the nonparametric function $p(u,v)$ is used.

The proof of Corollary 3.4 follows from Theorem 3.3 immediately. We now finish the proof of the main results given in Section 3.

Appendix C

The following two technical lemmas have already been used in the proof of Lemma B.1 and the proof of Theorem 3.1. The two lemmas are of general interest and can be used for other nonparametric estimation and testing problems associated with the α–mixing condition.

Lemma C.1. Suppose that M_m^n are the σ-fields generated by a stationary α-mixing process ξ_i with the mixing coefficient $\alpha(i)$. For some positive integers m let $\eta_i \in M_{s_i}^{t_i}$ where $s_1 < t_1 < s_2 < t_2 < \cdots < t_m$ and suppose $t_i - s_i > \tau$ for all i. Assume further that

$$||\eta_i||_{p_i}^{p_i} = E|\eta_i|^{p_i} < \infty,$$

for some $p_i > 1$ for which

$$Q = \sum_{i=1}^{l} \frac{1}{p_i} < 1.$$

Then

$$\left| E\left[\prod_{i=1}^{l} \eta_i\right] - \prod_{i=1}^{l} E[\eta_i]\right| \leq 10(l-1)\alpha(\tau)^{(1-Q)} \prod_{i=1}^{l} ||\eta_i||_{p_i}.$$

Proof. See Roussas and Ionnides (1987).

Lemma C.2. (i) *Let $\psi(\cdot,\cdot,\cdot)$ be a symmetric Borel function defined on $R^r \times R^r \times R^r$. Let the process ξ_i be defined as in Lemma B.1. Assume that for any fixed $x, y \in R^r$, $E[\psi(\xi_1, x, y)] = 0$. Then*

$$E\left\{\sum_{1\leq i<j<k\leq T} \psi(\xi_i,\xi_j,\xi_k)\right\}^2 \leq CT^3 M^{\frac{1}{1+\delta}},$$

where $0 < \delta < 1$ *is a small constant,* $C > 0$ *is a constant independent of* T *and the function* ψ*,* $M = \max\{M_1, M_2, M_3\}$*, and*

$$M_1 = \max_{1<i<j\le T} \max\left\{E|\psi(\xi_1,\xi_i,\xi_j)|^{2(1+\delta)}, \int |\psi(\xi_1,\xi_i,\xi_j)|^{2(1+\delta)} dP(\xi_1)dP(\xi_i,\xi_j)\right\},$$

$$M_2 = \max_{1<i<j\le T} \max\left\{\int |\psi(\xi_1,\xi_i,\xi_j)|^{2(1+\delta)} dP(\xi_j)dP(\xi_1,\xi_i)\right\},$$

$$M_3 = \max_{1<i<j\le T} \max\left\{\int |\psi(\xi_1,\xi_i,\xi_j)|^{2(1+\delta)} dP(\xi_1)dP(\xi_i)dP(\xi_j)\right\}. \qquad (C.1)$$

(ii) *Let* $\phi(\cdot,\cdot)$ *be a symmetric Borel function defined on* $R^r \times R^r$*. Let the process* ξ_i *be defined as in Lemma B.1. Assume that for any fixed* $x \in R^r$*,* $E[\phi(\xi_1, x)] = 0$*. Then*

$$E\left\{\sum_{1\le i<j\le T} \phi(\xi_i,\xi_j)\right\}^2 \le CT^2 M_4^{\frac{1}{1+\delta}},$$

where $\delta > 0$ *is a constant,* $C > 0$ *is a constant independent of* T *and the function* ϕ*, and*

$$M_4 = \max_{1<i<j\le T} \max\left\{E|\phi(\xi_1,\xi_i)|^{2(1+\delta)}, \int |\phi(\xi_1,\xi_i)|^{2(1+\delta)} dP(\xi_1)dP(\xi_i)\right\}. \qquad (C.2)$$

Remark C.1. Lemma C.2 is useful in itself for providing moment inequalities for strictly stationary and mixing processes.

Proof: As the proof of (ii) is similar to that of (i), one proves only (i). Let $i_1, \ldots, i_6$ be distinct integers and $1 \le i_j \le T$, let $1 \le k_1 < \cdots < k_6 \le T$ be the permutation of $i_1, \ldots, i_6$ in ascending order and let d_c be the c–th largest difference among $k_{j+1} - k_j$, $j = 1, \cdots, 5$. Let

$$H(k_1, \cdots, k_6) = \psi(\xi_{i_1}, \xi_{i_2}, \xi_{i_3})\psi(\xi_{i_4}, \xi_{i_5}, \xi_{i_6}).$$

By Lemma C.1 (with $\eta_1 = \psi(\xi_{i_1}, \xi_{i_2}, \xi_{i_3})$, $\eta_2 = \psi(\xi_{i_4}, \xi_{i_5}, \xi_{i_6})$, $l = 2$, $p_i = 2(1+\delta)$ and $Q = \frac{1}{1+\delta}$),

$$|E[H(k_1,\cdots,k_6)]| \le \begin{cases} 10M^{\frac{1}{1+\delta}}\alpha^{\frac{\delta}{1+\delta}}(k_6-k_5) & \text{if } k_6-k_5 = d_1 \\ 10M^{\frac{1}{1+\delta}}\alpha^{\frac{\delta}{1+\delta}}(k_2-k_1) & \text{if } k_2-k_1 = d_1. \end{cases}$$

Thus,

$$\sum_{\substack{1\le k_1<\cdots<k_6\le T \\ k_2-k_1=d_1}} |E[H(k_1,\cdots,k_6)]|$$

$$\le \sum_{k_1=1}^{T-5} \sum_{k_2=k_1+\max_{j\ge3}\{k_j-k_{j-1}\}} \sum_{k_3=k_2+1}^{T-3} \cdots \sum_{k_6=k_5+1}^{T} \left\{10M^{\frac{1}{1+\delta}}\alpha^{\frac{\delta}{1+\delta}}(k_2-k_1)\right\}$$

$$\le 10M^{\frac{1}{1+\delta}} \sum_{k_1=1}^{T-5} \sum_{k_2=k_1+1}^{T-4} (k_2-k_1)^2 \alpha^{\frac{\delta}{1+\delta}}(k_2-k_1)$$

$$\le 10TM^{\frac{1}{1+\delta}} \sum_{k=1}^{T} k^4 \alpha^{\frac{\delta}{1+\delta}}(k) \le CTM^{\frac{1}{1+\delta}}. \qquad (C.2)$$

Similarly,

$$\sum_{\substack{1 \le k_1 < \cdots < k_6 \le T \\ k_2 - k_1 = d_1}} |E[H(k_1, \cdots, k_6)]| \le CTM^{\frac{1}{1+\delta}}. \qquad (C.3)$$

Analogously, it can be shown in a similar way that

$$\sum_{\substack{1 \le k_1 < \cdots < k_6 \le T \\ k_6 - k_5 = d_2 \text{ or } k_2 - k_1 = d_2}} |E[H(k_1, \cdots, k_6)]| \le CT^2 M^{\frac{1}{1+\delta}}, \qquad (C.4)$$

$$\sum_{\substack{1 \le k_1 < \cdots < k_6 \le T \\ k_6 - k_5 = d_3 \text{ or } k_2 - k_1 = d_3}} |E[H(k_1, \cdots, k_6)]| \le CTM^{\frac{1}{1+\delta}}. \qquad (C.5)$$

On the other hand, if $\{k_6 - k_5, k_2 - k_1\} = \{d_4, d_5\}$, by using Lemma C.1 three times we have the inequality

$$|E[H(k_1, \cdots, k_6)]| \le 10M^{\frac{1}{1+\delta}} \sum_{i=1}^{3} \alpha^{\frac{\delta}{1+\delta}}(d_i).$$

Hence,

$$\sum_{\substack{1 \le k_1 < \cdots < k_6 \le T \\ \{k_6 - k_5, k_2 - k_1\} = \{d_4, d_5\}}} |E[H(k_1, \cdots, k_6)]|$$

$$\le \sum_{\substack{1 \le k_1 < \cdots < k_6 \le T \\ \max\{k_6 - k_5, k_2 - k_1\} \\ \le \min_{2 \le j \le 4}\{k_{j+1} - k_j\}}} \left\{10M^{\frac{1}{1+\delta}} \left[\alpha^{\frac{\delta}{1+\delta}}(k_3 - k_2) + \alpha^{\frac{\delta}{1+\delta}}(k_4 - k_3) + \alpha^{\frac{\delta}{1+\delta}}(k_5 - k_4)\right]\right\}$$

$$\le 30M^{\frac{1}{1+\delta}} \sum_{\substack{1 \le k_1 < \cdots < k_6 \le T \\ \max\{k_6 - k_5, k_2 - k_1\} \le d_3}} \alpha^{\frac{\delta}{1+\delta}}(d_3) \le 30CT^3 M^{\frac{1}{1+\delta}}. \qquad (C.6)$$

It follows from (C.2)–(C.6) that

$$\sum_{\substack{1 \le i, j, k, r, s, t \le T \\ i, j, k, r, s, t \text{ different}}} |E[\psi(\xi_i, \xi_j, \xi_k)\psi(\xi_r, \xi_s, \xi_t)]| \le CT^3 M^{\frac{1}{1+\delta}}. \qquad (C.7)$$

Similar to (C.7), one can show that

$$\sum_{\substack{1 \le i, j, k, r, s, t \le T \\ i, j, k, s, t \text{ different}}} |E[\psi(\xi_i, \xi_j, \xi_k)\psi(\xi_i, \xi_s, \xi_t)]| \le CT^3 M^{\frac{1}{1+\delta}}, \qquad (C.8)$$

$$\sum_{\substack{1 \le i,j,k,l \le T \\ i,j,k,l \text{ different}}} |E[\psi(\xi_i, \xi_j, \xi_k)\psi(\xi_i, \xi_j, \xi_l)]| \le CT^3 M^{\frac{1}{1+\delta}}. \qquad (C.9)$$

Finally, it is easy to see that

$$\sum_{1 \le i < j < k \le T} E[\psi(\xi_i, \xi_j, \xi_k)^2] \le T^3 \max_{1<i<j} E[\psi(\xi_1, \xi_i, \xi_j)^2]. \qquad (C.10)$$

The conclusion of Lemma C.2(i) follows immediately from (C.7)–(C.10).

Lemma C.3. Let $\phi(\cdot, \cdot)$ *be a symmetric Borel function defined on* $R^r \times R^r$. *Let the process* ξ_i *be defined as in Lemma C.2. Assume that for any fixed* $x, y \in R^r$, $E[\phi(x, \xi_1)] = E[\phi(\xi_1, y)] = 0$. *Then for* $1 \le i < j \le T$,

$$|E[\phi(\xi_i, \xi_j)|I_i]| \le C\alpha^{\frac{\delta}{1+\delta}}(j-i) \left(E\left[|\phi(\xi_i, \xi_j)|^{1+\delta}\right]\right)^{\frac{1}{1+\delta}},$$

where $0 < \delta < 1$ *is some constant such that* $\max_{1 \le i < j \le T} E\left[|\phi(\xi_i, \xi_j)|^{1+\delta}\right] < \infty$.

Proof: See Yoshihara (1989) or Roussas and Ionnides (1987).

Appendix D

This appendix gives the verification of Assumptions A.1–A.6 listed in Appendix A. It can be seen that the assumptions are justifiable for both the kernel method and the series case.

Remark D.1. (i) Assumption A.1(i) is quite common in the α-mixing case. Assumption A.1(ii) imposes some necessary conditions on the error process. Assumption A.1(iii) is adopted from Condition A1(iii) of Li (1999).

(ii) The justifications of Assumptions A.2 and A.3 are relegated to Remark D.2 and Remark D.3 respectively.

(iii) Assumption A.4 is similar to Conditions A1(iii) and A1(iv) of Li (1999).

(iv) Assumption A.5(i) is similar to Condition A2(i) of Li (1999). Assumption A.5(ii) is quite natural for the kernel method.

(v) Assumption A.6 is the corresponding version of Assumption A.5 for the series case. As the orthogonality conditions are assumed in Assumption A.5, we don't need to assume $\frac{k^{4(\mu+1)+p}}{T^{2p}} \to \infty$ as used in Theorem 3.1 of Hong and White (1995). More justifications for Assumption A.6 can be found in Gao and Tong (2001b).

Remark D.2. (i) Assumption A.2(i) is a general condition for the form of p_{st}. It holds automatically when the form of p_{st} is chosen as either the kernel based weight function or the series based weight function.

(ii) Assumption A.2(ii) imposes some conditions on σ_T^2 and q. This assumption can be justified for the kernel method with $q = h^{-p}$ and and the series method with $q = k^p$. The detailed justification follows from (iii) and (iv) below.

(iii) We now verify Assumption A.2(iii) in some detail. We first consider the kernel case.

Let $p_{st} = \frac{1}{Th^p} K\left(\frac{X_s - X_t}{h}\right)$ as in (3.5) and $\eta = (1+\delta)(1+\delta_1)$. It follows that for $1 \le i < j < k \le T$,

$$E[|p_{ik}p_{jk}|^\eta] = \frac{1}{(Th^p)^{2\eta}} \int\int\int \left|K\left(\frac{u-w}{h}\right)\right|^\eta \left|K\left(\frac{v-w}{h}\right)\right|^\eta f(u,v,w)dudvdw$$

$$= \frac{h^{2p}}{(Th^p)^{2\eta}} \int\int\int |K(x)K(y)|^\eta f(xh+z, yh+z, z)dxdydz \le C_1 \frac{h^{2p}}{(Th^p)^{2\eta}} \qquad (D.1)$$

under Assumptions A.4 and A.5(i).

Similarly, one can find

$$\int\int\int |p_{ik}p_{jk}|^\eta dP(X_i)dP(X_j, X_k) \le C_2 \frac{h^{2p}}{(Th^p)^{2\eta}} = \frac{1}{T^{2\eta}} h^{2p(1-\eta)}. \qquad (D.2)$$

Let $\sigma_s^2 = \sigma^2(X_s) = E[e_s^2|X_s]$. It follows similarly that

$$\sigma_{1T}^2 \equiv \sum_{1\le s<t\le T} E[\sigma_s^2\sigma_t^2 p_{st}^2] = \sum_{1\le s<t\le T} \frac{1}{(Th^p)^2} \int\int \sigma^2(u)\sigma^2(v)K^2\left(\frac{u-v}{h}\right) f(u,v)dudv$$

$$= \frac{T(T-1)}{(Th^p)^2} h^p \int\int \sigma^2(xh+y)\sigma^2(y)K^2(x)f(xh+y,y)dxdy = C_3 h^{-p}(1+o(1)) \qquad (D.3)$$

as $T \to \infty$.

Finally, it can be shown that as $T \to \infty$

$$\frac{\sigma_T^2 - \sigma_{1T}^2}{\sigma_{1T}^2} \to 0. \qquad (D.4)$$

Equations (D.1)–(D.4) imply as $T \to \infty$

$$\frac{T^2 M_{1T}^{\frac{1}{\eta}}}{\sigma_T^2} \le C_4 h^{\frac{(2-\eta)p}{\eta}} \to 0. \qquad (D.5)$$

Similarly, one can verify the others.

(iv) For the series case, without loss of generality one can choose $c_i^2 \equiv 1$ and

$$p_{st} = \frac{1}{T}\sum_{i=1}^k z_i(X_s)z_i(X_t).$$

As the justification of Assumption A.2(iii) depends on the choice of $\{z_i(\cdot)\}$, we now consider the case where $p = 1$ and $z_i(x) = \cos(ix)$. Similarly, one can verify the case where $z_i(x) = \sin(ix)$ and $z_j(x) = \cos(jx)$ for $i \ne j$.

Observe that

$$\sum_{i=1}^k \cos(ix)\cos(iy) = \frac{1}{2}\sum_{i=1}^k \cos((x+y)i) + \frac{1}{2}\sum_{i=1}^k \cos((x-y)i).$$

It follows that for any real number u

$$\sum_{i=1}^k \cos(iu)\sin(u/2) = \frac{1}{2}\sum_{i=1}^k \left\{\sin\left(iu + \frac{u}{2}\right) - \sin\left(iu - \frac{u}{2}\right)\right\}$$

$$= \frac{1}{2}\left\{\sin\left(ku + \frac{u}{2}\right) - \sin\left(\frac{u}{2}\right)\right\} = \cos\left(\frac{(k+1)u}{2}\right)\sin\left(\frac{k}{2}u\right).$$

For $u = \frac{X_s + X_t}{2}$ or $\frac{X_s - X_t}{2}$, let

$$I_k(u) = \frac{1}{2}\frac{\cos((k+1)u)\sin(ku)}{\sin(u)}.$$

Then

$$p_{st} = \frac{1}{T} I_k\left(\frac{X_s + X_t}{2}\right) + \frac{1}{T} I_k\left(\frac{X_s - X_t}{2}\right).$$

Similar to (D.1), one can have for the same η

$$\begin{aligned}
M_\eta &= E\left[|p_{ik}p_{jk}|^\eta\right] \\
&\le \frac{C_\eta}{T^{2\eta}}\int\int\int\left\{\left|I_k\left(\frac{u+w}{2}\right)I_k\left(\frac{v+w}{2}\right)\right|^\eta\right\} f(u,v,w)dudvdw \\
&+\frac{C_\eta}{T^{2\eta}}\int\int\int\left\{\left|I_k\left(\frac{u+w}{2}\right)I_k\left(\frac{v-w}{2}\right)\right|^\eta\right\} f(u,v,w)dudvdw \\
&+\frac{C_\eta}{T^{2\eta}}\int\int\int\left\{\left|I_k\left(\frac{u-w}{2}\right)I_k\left(\frac{v+w}{2}\right)\right|^\eta\right\} f(u,v,w)dudvdw \\
&+\frac{C_\eta}{T^{2\eta}}\int\int\int\left\{\left|I_k\left(\frac{u-w}{2}\right)I_k\left(\frac{v-w}{2}\right)\right|^\eta\right\} f(u,v,w)dudvdw \\
&\equiv M_{1\eta} + M_{2\eta} + M_{3\eta} + M_{4\eta}. \qquad (D.6)
\end{aligned}$$

In the following, we consider $M_{4\eta}$ only and the others follow similarly.

$$M_{4\eta} = \frac{C_\eta}{T^{2\eta}}\int\int\int\left|I_k\left(\frac{u-w}{2}\right)I_k\left(\frac{v-w}{2}\right)\right|^\eta f(u,v,w)dudvdw$$

$$= \frac{C_\eta}{2^{2\eta}T^{2\eta}}\int\int\int \frac{\left|\cos\left((k+1)\left(\frac{u-w}{2}\right)\right)\sin\left(k\left(\frac{u-w}{2}\right)\right)\right|^\eta}{\left|\sin\left(\frac{u-w}{2}\right)\right|^\eta}\frac{\left|\cos\left((k+1)\left(\frac{v-w}{2}\right)\right)\sin\left(k\left(\frac{v-w}{2}\right)\right)\right|^\eta}{\left|\sin\left(\frac{v-w}{2}\right)\right|^\eta}$$

$$\cdot f(u,v,w)dudvdw$$

$$= \frac{D_\eta}{T^{2\eta}}\int\int\int\frac{|\cos((k+1)x)\sin(kx)|^\eta}{|\sin(x)|^\eta}\frac{|\cos((k+1)y)\sin(ky)|^\eta}{|\sin(y)|^\eta} f(2x+z, 2y+z, z)dxdydz$$

$$= \frac{D_\eta}{(k+1)^2T^{2\eta}}\int\int\int\frac{\left|\cos(x_1)\sin\left(x_1 - \frac{x_1}{k+1}\right)\right|^\eta}{\left|\sin\left(\frac{x_1}{k+1}\right)\right|^\eta}\frac{\left|\cos(x_2)\sin\left(x_2 - \frac{x_2}{k+1}\right)\right|^\eta}{\left|\sin\left(\frac{x_2}{k+1}\right)\right|^\eta}$$

$$\cdot f\left(x_3 + \frac{2x_1}{k+1}, x_3 + \frac{2x_2}{k+1}, x_3\right)dx_1dx_2dx_3$$

$$= \frac{D_\eta(k+2)^{2\eta}}{(k+1)^2T^{2\eta}}\int\int\int\frac{\left|\cos(x_1)\sin\left(x_1 - \frac{x_1}{k+1}\right)\right|^\eta}{|x_1|^\eta\,|\psi_k(x_1)|^\eta}\frac{\left|\cos(x_2)\sin\left(x_2 - \frac{x_2}{k+1}\right)\right|^\eta}{|x_2|^\eta\,|\psi_k(x_2)|^\eta}$$

$$\cdot f\left(x_3 + \frac{2x_1}{k+1}, x_3 + \frac{2x_2}{k+1}, x_3\right)dx_1dx_2dx_3$$

$$\equiv D_\eta\frac{(k+1)^{2(\eta-1)}}{T^{2\eta}}C_\eta(k),$$

where $\psi_k(x) = \frac{\sin\left(\frac{x}{k+1}\right)}{\frac{x}{k+1}}$, $0 < D_\eta < \infty$ is a constant independent of k, and $C_\eta(k)$ is a function of k.

Hence, using properties of the trigonometric series and Assumption A.3, one can have as $T \to \infty$

$$M_{4\eta} = O\left(\frac{(k+1)^{2(\eta-1)}}{T^{2\eta}}\right). \tag{D.7}$$

Thus it can be seen that $M_{4\eta}$ has the same order as (D.1) when $p = 1$ and $k = h^{-1}$.

Analogously, one can find the corresponding versions of (D.2)–(D.5) for the series case. Therefore, it can be shown that Assumption A.2(iii) is justifiable for both the kernel case and the series case.

Before justifying Assumption A.3 for the kernel method, one needs to introduce the following assumption.

Assumption D. In addition to Assumption A.5, assume that

(i) the univariate kernel function $k(\cdot)$ is of bounded variation on $R^1 = (-\infty, \infty)$;

(ii) there is a second kernel function $l(\cdot)$ that satisfies the same conditions as $k(\cdot)$ does; and

(iii) there is a second bandwidth parameter h_1 such that for $0 < \eta < \frac{7p}{8}$

$$h = O\left(T^{-\eta}\right),\ h_1 \to 0 \text{ as } T \to \infty,\ \lim_{T\to\infty} \frac{h^p}{h_1^{2d}} = 0 \text{ and } \lim_{T\to\infty} T h^{p/2} h_1^4 = 0.$$

Assumption D(i) is required for the uniform convergence of d_{1Ts} assumed in Assumption A.3 (ii). Assumption D(ii)(iii) is taken from Condition (A2) of Li (1999).

Remark D.3. (i) This remark needs only to justify Assumption A.3(ii)(iii). Assumption A.3(ii) holds for both the kernel method and the series case. As can be seen from Remark D.2(iv), it suffices to verify the kernel case. For the kernel method, one can take

$$d_{1Tt} = \frac{1}{T h_1^d} \sum_{s\neq t} L_h(U_t - U_s) \text{ and } d_{1t} = f_1(U_t),$$

which is the density function of U_t, where $L_{h_1}(\cdot) = L(\cdot/h_1)$ and $L(u_1, \cdots, u_d) = \prod_{i=1}^d l(u_i)$. Under Assumptions A.5(i)(ii) and D(i), one can show that $\max_{t\geq 1} |d_{1Tt} - d_{1t}| = o_p(1)$. See Lemmas A.1 and A.3 of Härdle, Liang and Gao (2000) for example.

(ii) For both the kernel method and the series case, one can justify the following three equations

$$\frac{1}{\sigma_{1T}} \sum_{1\leq s<t\leq T} E\left[e_s^2 \eta_{1s}^2 |p_{st}|\right] \to 0,$$

$$\frac{1}{\sigma_{1T}} \sum_{1\leq s<t\leq T} E\left[e_s^2 \delta_{1s}^2 d_{1Ts}^2 |p_{st}|\right] \to 0,$$

and

$$\frac{1}{\sigma_{1T}^2} \sum_{1\leq s<t\leq T} E\left[\delta_{1s}^2 d_{1Ts}^2 \delta_{1t}^2 d_{1Tt}^2 p_{st}^2\right] \to 0.$$

For the kernel case, by Assumption D, Lemma C.3 of Li (1999) can be used for the justification. For this case, one can use

$$p_{st} = \frac{1}{Th^p} \sum_{s \neq t} K_h(X_s - X_t) \text{ and } \sum_{s=1, \neq t} p_{st} = \hat{f}(X_t)$$

as an estimate to $f(X_t)$ in the first two equations. For the kernel case, one can compute that $\sigma_{1T}^2 = Ch^{-p}(1 + o(1))$ using Assumption A.4.

For both the second equation and the third equation, one can substitute $\delta_{1t} d_{1Tt}$ by

$$\begin{aligned}
\delta_{1t} d_{1Tt} &= \left[m_1(U_t) - \sum_{s=1}^T w_{1ts} Y_s \right] d_{1Tt} = -\sum_{s=1}^T p_{1ts} e_s + \left[m_1(U_t) - \sum_{s=1}^T w_{1ts} m_1(U_s) \right] d_{1Tt} \\
&= -\sum_{s=1}^T p_{1ts} e_s + \left[m_1(U_t) - \sum_{s=1}^T w_{1ts} m_1(U_s) \right] d_{1t} + \left[m_1(U_t) - \sum_{s=1}^T w_{1ts} m_1(U_s) \right] \eta_{1t} \\
&\equiv -\sum_{s=1}^T p_{1ts} e_s + \eta_{2t} d_{1t} + \eta_{2t} \eta_{1t},
\end{aligned}$$

where $\eta_{1t} = d_{1Tt} - d_{1t}$ and $\eta_{2t} = m_1(U_t) - \sum_{s=1}^T w_{1ts} m_1(U_s)$.

For the first part, one can use Lemma C.1 again to estimate the order. Assumptions A.4(ii), A.5 and D can be used to estimate the order of the second and third parts. For the series case, by Assumption A.6 and Remark D.2(iv) one can verify Assumption A.3(iii).

Remark D.4. (i) As one can see, Assumption A.2(iii) is a necessary condition, but its justification depends heavily on the explicit form of $\{p_{st}\}$. If one replaces σ_T^2 by

$$\Sigma_T^2 = \sigma_T^2 + 2 \sum_{1 \le s < t < u \le T} \text{cov}(e_s p_{su} e_u, e_t p_{tu} e_u), \tag{D.8}$$

Assumption A.2(iii) is not required. Certainly, one needs to ensure that $\inf_T \Sigma_T^2 > 0$.

For this case, the test statistics proposed need to be modified. For example, L_T of (3.4) needs to be replaced by

$$\widetilde{L}_T = \frac{\sum_{t=1}^T \sum_{s \neq t} p_{st} Y_s Y_t}{\widetilde{S}_T},$$

where $\widetilde{S}_T^2 = 2 \sum_{t=1}^T \sum_{s=1}^T p_{st}^2 Y_s^2 Y_t^2 + 4 \sum_{t=1}^T \sum_{s=1}^T \sum_{u=1}^T p_{su} p_{tu} Y_s Y_t Y_u^2$.

As can be seen from the proof of (B.4) and (B.5), Assumption A.2(iii) is not required for this case.

For the form of (D.8) itself, one needs to point out that (D.8) is of general interest. For example, when e_t is a long-range dependent process, the second part of (D.8) cannot be estimated by using the α–mixing condition and Lemma C.1, as the long-range dependence and the α–mixing condition contradict each other. For the long-range dependent case, the second part of (D.8) therefore needs to be included.

(ii) For the short-range dependent case, in order to aviod using Assumption A.2(iii), one may further impose some conditions on the structure of e_t to ensure that the second term of (D.8) equals to zero. That is

$$E[e_s e_t p_{su} p_{tu} e_u^2] = 0 \text{ for any } s \neq t \neq u.$$

This holds for example when $e_t = h(X_t, Y_{t-1})\epsilon_t$, $h(\cdot, \cdot)$ is a measurable function, and ϵ_t satisfies

$$E[\epsilon_t|\Omega_{t-1}] = 0 \;\text{ and }\; E[\epsilon_s \epsilon_t \epsilon_u^2|\Omega_{v-1}] = 0 \text{ for any } s \neq t \neq u,$$

where $v = \max\{s, t, u\}$ and Ω_t is as defined in Assumption A.1. For the independence case, this holds automatically.

References

Ahn, D. H., Gao, B., 1999. A parametric nonlinear model of term structure dynamics. Review of Financial Studies 12, 721–762.

Aït-Sahalia, Y., 1996a. Nonparametric pricing of interest rate derivative securities. Econometrica 64, 527–560.

Aït-Sahalia, Y., 1996b. Testing continuous-time models of the spot interest rate. Review of Financial Studies 9, 385–426.

Aït-Sahalia, Y., 1999. Transition densities for interest rate and other nonlinear diffusions. Journal of Finance 54, 1361–1395.

Aït-Sahalia, Y., Bickel, P., Stoker, T., 2001. Goodness-of-fit tests for regression using kernel methods. Journal of Econometrics 105, 363–412.

Chapman, D., Pearson, N., 2000. Is the short rate drift actually nonlinear ? Journal of Finance 54, 355–388.

Chen, H., Chen, K, 1991. Selection of the splined variables and convergence rates in a partial spline model. The Canadian Journal of Statistics 19, 323–339.

Chen, S., Härdle, W., Li, M., 2001. An empirical likelihood goodness-of-fit test for time series. Private communication.

Engle, R. F., Granger, C. W. J., Rice, J. A., Weiss, A., 1986. Semiparametric estimates of the relation between weather and electricity sales. Journal of the American Statistical Association 81, 310–320.

Eubank, R. L., Hart, J. D., 1992. Testing goodness-of-fit in regression via order selection. The Annals of Statistics 20, 1412–1425.

Eubank, R. L., Spiegelman, C. H., 1990. Testing the goodness of fit of a linear model via nonparametric regression techniques. Journal of the American Statistical Association 85, 387–392.

Eumunds, D. E., Moscatelli, V. B., 1977. Fourier approximation and embeddings of Sobolev space. Dissertationae Mathematicae. Polish Scientific Publishers, Warsaw.

Fan, J., 1992. Design-adaptive nonparametric regression. Journal of the American Statistical Association 87, 998–1004.

Fan, J., Gijbels, I., 1996. Local Polynomial Modelling and Its Applications. Chapman and Hall, London.

Fan, J., Härdle, W., Mammen, E., 1998. Direct estimation of low dimensional components in additive models. The Annals of Statistics 26, 943–971.

Fan, J., Huang, L. S., 2001. Goodness-of-fit tests for parametric regression models. Journal of the American Statistical Association 453, 640–652.

Fan, J., Yao, Q., 1998. Efficient estimation of conditional variance functions in stochastic regression. Biometrika 85, 645–660.

Fan, J., Zhang, C. M., Zhang, J., 2001. Generalized likelihood ratio statistics and Wilks phenomenon. The Annals of Statistics 29, 153–193.

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

Fan, Y., Li, Q., 1998. Central limit theorem for degenerate U–statistics of absolutely regular processes with applications to model specification testing. Journal of Nonparametric Statistics 10, 245–271.

Gao, J., 1998. Semiparametric regression modelling of nonlinear time series. Scandinavian Journal of Statistics 25, 521–539.

Gao, J., 2000. A semiparametric approach to pricing interest rate derivative securities. Unpublished paper presented at Quantitative Methods in Finance & Bernoulli Society 2000 Conference, Sydney, 4–9 December, 2000. Available from www.maths.uwa.edu.au/˜jiti/uts.pdf.

Gao, J., Liang, H., 1995. Asymptotic normality of pseudo-LS estimator for partially linear autoregressive models. Statistics and Probability Letters 23, 27–34.

Gao, J., Liang, H., 1997. Statistical inference in single-index and partially nonlinear regression models. Annals of the Institute of Statistical Mathematics 49, 493–517.

Gao, J., Shi, P., 1997. M-type smoothing splines in nonparametric and semiparametric regression models. Statistica Sinica 7, 1155–1169.

Gao, J., Tong, H., Wolff, R., 2002a. Model specification tests in nonparametric stochastic regression models. Journal of Multivariate Analysis Vol. 83, 324–359.

Gao, J., Tong, H., Wolff, R., 2002b. Adaptive orthogonal series estimation in additive stochastic regression models. Statistica Sinica 12, 409–428.

Gozalo, P. L., 1993. A consistent model specification test for nonparametric estimation of regression function models. Econometric Theory 9, 451–477.

Gozalo, P. L., Linton, O. B., 2001. Testing additivity in generalized nonparametric regression models with estimated parameters. Journal of Econometrics 104, 1–48.

Granger, C. W. J., Inoue, T., Morin, N., 1997. Nonlinear stochastic trends. Journal of Econometrics 81, 65–92.

Härdle, W., Kneip, A., 1999. Testing a regression model when we have smooth alternatives in mind. Scandinavian Journal of Statistics 26, 221–238.

Härdle, W., Liang, H., Gao, J., 2000. Partially Linear Models. Springer Series: Contributions to Statistics. Physica-Verlag, New York.

Härdle, W., Mammen, E., 1993. Comparing nonparametric versus parametric regression fits. The Annals of Statistics 21, 1926–1947.

Hidalgo, F. J., 1992. Adaptive semiparametric estimation in the presence of autocorrelation of unknown form. Journal of Time Series Analysis 13, 47–78.

Hjellvik, V., Tjøstheim, D., 1995. Nonparametric tests of linearity for time series. Biometrika 82, 351–368.

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

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

Horowitz, J. L., Härdle, W., 1994. Testing a parametric model against a semiparametric alternative. Econometric Theory 10, 821–848.

Horowitz, J. L., Spokoiny, V. G., 2001. An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. Econometrica 69, 599–632.

Jayasuriya, B., 1996. Testing for polynomial regression using nonparametric regression techniques. Journal of the American Statistical Association 91, 1626–1631.

Jiang, G., Knight, J., 1997. A nonparametric approach to the estimation of diffusion processes with an application to a short-term interest rate model. Econometric Theory 13, 615–645.

Karlsen, H., Tjøstheim, D., 2001 Nonparametric estimation in null recurrent time series. The Annals of Statistics 29, 372–416.

Kreiss, J. P., Neumann, M. H., Yao, Q., 1997. Bootstrap tests for simple structures in nonparametric time series regression. Private communication.

Lavergne, P., 2001. An equality test across nonparametric regressions. Journal of Econometrics 103, 307–344.

Lavergne, P., Vuong, Q. H., 1996. Nonparametric selection of regressors: the nonnested case. Econometrica 64, 207–219.

Lavergne, P., Vuong, Q. H., 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, 101–147.

Li, Q., Hsiao, C., 1998. Testing serial correlation in semiparametric panel data models. Journal of Econometrics 87, 207–237.

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

Linton, O. B., 1997. Efficient estimation of additive nonparametric regression models. Biometrika 84, 469–473.

Linton, O. B., 2000. Efficient estimation of generalized additive nonparametric regression models. Econometric Theory 16, 502–523.

Linton, O, B., Härdle, W., 1996. Estimation of additive regression models with known links. Biometrika 83, 529–540.

Mills, T., 1993. The Econometric Modelling of Financial Time Series. Cambridge University Press, New York.

Pagan, A., Ullah, A., 1999. Nonparametric Econometrics. Cambridge University Press, New York.

Phillips, P., Loretan, M., 1991. Estimating long-run economic equilibria. Review of Economic Studies 58, 407–436.

Priestley, M., Chao, M., 1972. Nonparametric function fitting. Journal of the Royal Statistical Society Series B 34, 384–392.

Pritsker, M., 1998. Nonparametric density estimation and tests of continuous time interest rate models. Review of Financial Studies 11, 449–487.

Robinson, P., 1988. Root-N-consistent semiparametric regression. Econometrica 56, 931–964.

Robinson, P., 1989. Hypothesis testing in semiparametric and nonparametric models for econometric time series. Review of Economic Studies 56, 511–534.

Roussas, G., Ioannides, D., 1987. Moment inequalities for mixing sequences of random variables. Stochastic Analysis and Applications 5, 61–120.

Samarov, A. M., 1993. Exploring regression structure using nonparametric functional estimation. Journal of the American Statistical Association 423, 836–847.

Shively, T., Kohn, R., Ansley, C. F., 1994. Testing for linearity in a semiparametric regression model. Journal of Econometrics 64, 7–96.

Sperlich, S., Tjøstheim, D., Yang, L., 2002. Nonparametric estimation and testing of interaction in additive models. Econometric Theory 18, 197–251.

Stanton, R., 1997. A nonparametric model of term structure dynamics and the market price of interest rate risk. Journal of Finance 52, 1973–2002.

Teräsvirta, T., Tjøstheim, D., Granger, C., 1994. Aspects of modelling nonlinear time series, in R. F. Engle and D. L. McFadden (eds), Handbook of Econometrics 4, 2919–2957.

Truong, Y. K., Stone, C. J., 1994. Semiparametric time series regression. Journal Time Series Analysis 15, 405–428.

van Dijk, D., Franses, P., 2000. Nonlinear error-correction models for interest rates in the Netherlands. International Symposia in Economic Theory and Econometrics (eds. W. Barnett, D. Hendry, S. Hylleberg, T. Teräsvirta, D. Tjøstheim and A. Würtz) 11, 203–227.

Wahba, G., 1978. Improper priors, spline smoothing and the problem of guarding against model errors in regression. Journal of the Royal Statistical Society Series B 40, 364–372.

Whang, Y. J., Andrews, D. W. K., 1993. Tests of specification for parametric and semiparametric models. Journal of Econometrics 57, 277–318.

Wooldridge, J., 1992. A test for functional form against nonparametric alternatives. Econometric Theory 8, 452–475.

Yatchew, A. J., 1992. Nonparametric regression tests based on least squares. Econometric Theory 8, 435–451.

Yoshihara, K., 1989. Limiting behavior of generalized quadratic forms generated by absolutely regular sequences. In *Proc. Fourth Prague Conference on Asymptotic Statist.*, 539–547.

Zheng, J. X., 1996. A consistent test of functional form via nonparametric estimation techniques. Journal of Econometrics 75, 263–289.