ISSN 1440-771X

MONASH University

Australia

Department of Econometrics and Business Statistics

http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/

Semiparametric Methods in Nonlinear Time Series Analysis: A Selective Review

Patrick Saart and Jiti Gao

November 2012

Working Paper 21/12

Semiparametric Methods in Nonlinear Time Series Analysis: A Selective Review

Patrick Saart[†1] and Jiti Gao[‡]

University of Canterbury[†] and *Monash University*[‡]

Abstract

Time series analysis is a tremendous research area in statistics and econometrics. As remarked in a review by Howell Tong in 2001, *for about 100 years up to 2001 Biometrika* (alone) *published over 400 papers on the subject.* [Tong (2001)] Furthermore, in the review, Howell Tong is able break down up to fifteen key areas of research interest in time series analysis. Nonetheless, unlike that of Howell Tong, the aim of the review in this paper is not to cover a wide range of topics on the subject, but is to concentrate on a small, but extremely essential, point he made on the *semiparametric methods in nonlinear time series analysis* and to explore into various aspects of this research area in much more detail. It is also an objective of this review to provide some discussion on a future research where appropriate.

JEL Classification: C12, C14, C22

Keywords: Autoregressive time series; nonparametric model; nonstationary process; partially linear structure, semiparametric method

[1]Patrick Saart, Department of Mathematics and Statistics, University of Canterbury, New Zealand, Email: patrick.wsaart@canterbury.ac.nz, Website: https://sites.google.com/site/patrickwsaart/home

1. Introduction

In time series regression, nonparametric methods have been very popular both for prediction and characterizing nonlinear dependence. Let $\{Y_t\}$ and $\{X_t\}$ be the one–dimensional and d-dimensional time series data, respectively. For a vector of time series data $\{Y_t, X_t\}$, the conditional mean function $E[Y_t|X_t = x]$ of Y_t on $X_t = x$ may be estimated nonparametrically by the Nadaraya–Watson (NW) estimator when the dimensionality d is less than three. When d is greater than three, the conditional mean can still be estimated using the NW estimator, and an asymptotic theory can be constructed. In practice, however, because of the so–called ***curse of dimensionality***, this may not be recommended unless the number of data points is extremely large.

Nonetheless, there are multiple phenomena referred to by the name "curse of dimensionality" in domains such as numerical analysis, sampling, combinatorics, machine learning, data mining and databases. For the sake of clarity, below let us give a simple example of the curse of dimensionality problem in nonparametric regression.

Example 1. *Let there be a set of data points (U, V), where*

$$U = g(V) + \quad \textit{noise (of mean zero)}. \tag{1}$$

The data $(\mathbf{V}, \mathbf{U}) \doteq \{(V_i, U_i)\}_{i=1}^n$ are assumed to be drawn i.i.d from a distribution over a joint input–output space $\mathcal{V} \times \mathcal{U}$. The input space $\mathcal{V}$ is usually assumed to be a subset of $\mathbb{R}^d$, i.e. V is a vector of d features. The output space $\mathcal{U}$ is assumed to be a subset of $\mathbb{R}^{d'}$, and is a random vector satisfying $E[U|V = v] = g(v)$. An objective of the nonparametric regression is to approximate g with a nonparametric regressor, say, g_n. Under some smoothing condition, e.g. Lipschitz, of g, a number of nonparametric estimators can be shown to satisfy

$$E_{\mathbf{V},\mathbf{U}}||g_n - g||^2 \leq O\left(n^{-2/(2+d)}\right). \tag{2}$$

For instance, this is the rate for a kernel estimator. Such a rate implies that we need a sample size n exponential in d in order to approximate g. Hence, when d is high as it is often the case in modern application, $n > 2^d$ is impractical. ■

In order to get an intuition as to the reason for such a rate, consider that nonparametric approaches such as the aforementioned operate by approximating the target function locally (on its domain $\mathcal{V}$) by simpler functions. There is necessarily some local errors and these errors aggregate globally. Thus to approximate the entire function well, we need to do well in most local areas. Suppose for instance that the target function is well approximated by constants in regions of radius at most $0 < r < 1$. In how many ways can we divide up the domain $\mathcal{V}$ into smaller regions of radius at most r? If $\mathcal{V}$ is d-dimensional then the smallest such partition is of size $O(r^{-d})$. We will need data points to fall into each such region if we hope to do well locally everywhere. In other words, we will need a data size exponential in d.

Over the recent years, a number of survey papers on nonparametric and semiparametric statistics have become available in the literature. Below let us introduce only a few of those that are the most relevant to the materials presented in the current paper. These are the works of Fan (2005), Härdle et al. (2007), and Sun and Li (2012). The main themes of these papers are ***highlighted***.

Fan (2005) presents a selective review of ***nonparametric methods in financial econometrics***. The main focus of the paper is on nonparametric techniques used for estimating stochastic diffusion models, especially the drift and the diffusion functions based on either discretely or continuously observed data. The paper begins with a brief review of some useful stochastic models for modeling stock prices and bond yields. These may include, for example, the CIR model by Cox et al. (1985), Vasicek model by Vasicek (1977) and the CKLS model of Chan et al. (2012). Furthermore, it is reviewed in the paper techniques for estimating state price densities and transition densities and their applications in asset pricing and testing for parametric diffusion models. Some important references are, for example, Ait-Sahalia (1996), Aït-Sahalia (2002), Aït-Sahalia and Lo (2002) and Fan et al. (1996).

Härdle et al. (2007) provide a fairly broad survey of many ***nonparametric analysis techniques for time series***. Specifically, they review nonparametric methods for estimating the spectral density, the conditional mean, higher order conditional moments or conditional densities. Moreover density estimation with correlated data, bootstrap methods for time series and nonparametric trend analysis are described.

Sun and Li (2012) focus on some recent theoretical developments in ***nonparametric and semiparametric techniques as applied to nonstationary or near–nonstationary variables***. Firstly, they review both the various concepts of I(0), I(1) and cointegrating for a linear model as they are available in the literature and some existing works on extending these concepts to nonlinear framework. Secondly, they discuss some popular nonlinear parametric models beginning with those for stationary data, such as the self–exciting threshold autoregressive (SETAR) (e.g. Tong and Lim (1980); Chan (1993)) models and smooth transition autoregressive (STAR) (e.g. Van Dijk et al. (2002)) models, then some nonlinear error correction (NEC) models and nonlinear cointegrating models (e.g.Terasvirta et al. (2011); Dufrénot and Mignon (2002)). Thirdly, the authors review nonparametric models with nonstationary data in a similar fashion as to the above parametric cases, i.e. nonparametric autoregressive followed by nonparametric cointegrating models. The review concentrates on existing works on the consistency of nonparametric estimators with some key studies such as Wang and Phillips (2009a); Wang and Phillips (2009b); Karlsen and Tjostheim (2001); Karlsen et al. (2007); Karlsen et al. (2010). Finally, it is the discussion on semiparametric models and model specification tests with stationary data; see also Gao et al. (2009a); Gao et al. (2009b); Gao and King (2012); Gao et al. (2012b) for some recent studies related to hypothesis testing of nonstationarity which are not yet available in Sun and Li (2012).

In each of the above three paragraphs, the highlighted phrase gives the main theme of the review paper being introduced. The current paper complements these existing reviews by filling in some gaps, which are currently left unexplored. We believe that the most appropriate description for this paper is ***semiparametric time series models as tools to circumventing curse of dimensionality and other usefulness recently explored in the literature***. Finding tools to circumventing the curse of dimensionality is traditional a major objective for a large number of studies in nonparametric statistics. There are essentially two approaches: the first is largely concerned with dimension reduction while second with function approximation. Some well–known examples of a study that falls into the first category are Li (1991); Cook (1998); Xia et al. (2002). Regarding the review in the current paper, the above–given description clearly indicates its focus on studies in the latter category that is function approximation using semiparametric specifications.

In this paper, we review in Section 2 various aspects of three of the most well known and successfully applied semiparametric time series models in the literature, namely the *partially linear*, *additive* and the *single–index* models. Secondly, we comprehensively review in Section 3 their recent extensions and other usefulness which are discussed in the literature. Our discussion in this section will focus on the involvement of semiparametric time series methods in the following issues: (i) hypothesis testing, (ii) detection and estimation of trend and seasonality, (iii) multivariate nonstationary time series models, (iv) semiparametric models with generated regressors and a semiparametric conditional duration model, and (v) nonlinear autoregressive models. Finally, Section 4 gives some concluding remarks.

2. Semiparametric Time Series Models

Let us begin with the most well–known model in its class, i.e. the semiparametric partially linear time series models.

Semiparametric Partially Linear Time Series Models

Since their introduction to the economic literature in the 1980s by Engle et al. (1986), the partially linear models have attracted much attention among econometricians and applied statisticians; see Heckman (1986), Robinson (1988) and Fan et al. (1995) for example. An advantage of using the semiparametric approach is the fact that a priori information concerning possible linearity of some of the components can now be included in the model. More specifically, we will look at approximating the conditional mean function $m(X_t) = m(U_t, V_t) = E[Y_t|U_t, V_t]$ by a semiparametric function of the form

$$m_1(U_t, V_t) = \mu + U_t^\tau \beta + g(V_t) \tag{3}$$

such that $E[Y_t - m_1(U_t, V_t)]^2$ is minimized over a class of semiparametric functions of the form $m_1(U_t, V_t)$ subject to $E[g(V_t)] = 0$ for the identifiability of $m_1(U_t, V_t)$, where μ is an unknown parameter, $\beta = (\beta_1, \ldots, \beta_q)^\tau$ is a vector of unknown parameters, $g(\cdot)$ is an unknown function over $\mathbb{R}^p$, both $U_t = (U_{t1}, \ldots, U_{tq})^\tau$ and $V_t = (V_{t1}, \ldots, V_{tp})^\tau$ may be vectors of time series variables. Such a minimization problem is equivalent to minimizing $E[Y_t - \mu - U_t^\tau \beta - g(V_t)]^2 = E\left[E\left\{(Y_t - \mu - U_t^\tau \beta - g(V_t))^2|V_t\right\}\right]$ over some (μ, β, g). This imples that $g(V_t) = E[(Y_t - \mu - U_t^\tau \beta)|V_t]$ and $\mu = E[Y_t - U_t^\tau \beta]$ with β being given by

$$\beta = \Sigma^{-1} E[(U_t - E[U_t|V_t])(Y_t - E[Y_t|V_t])] \tag{4}$$

provided that the inverse $\Sigma^{-1} = (E[U_t - E[U_t|V_t])(E[U_t - E[U_t|V_t])^\tau])^{-1}$ exists. This also shows that $m_1(U_t, V_t)$ is identifiable under the assumption of $E[g(V_t)] = 0$. Some important motivations for using the form (3) for independent data analysis can be found in Härdle et al. (2000). Based on an independently and identically distributed (i.i.d.) random sample, it has been shown that the parameter vector β in various versions of (3) can be consistently estimated at $\sqrt{n}$-rate, see Heckman (1986), Robinson (1988) and Fan et al. (1995) for example. For dependent processes, traditionally such a result is established under a set of somewhat more stringent conditions, e.g. the independence between $\{U_t\}$ and $\{V_t\}$ in Truong and Stone (1994). On the contrary, Fan and Li (1999b) extend the $\sqrt{n}$-consistency and asymptotic normality results for independent observations to weakly dependent, i.e. absolutely regular (β-mixing), processes under conditions similar to those employed in Robinson (1988) and Fan et al. (1995). These results are not applicable to a weaker condition of strong mixing processes and a case by which $p > 3$; however. ■

As with the independent data case, estimating $g(\cdot)$ in model (3) may still suffer from the curse of dimensionality when $g(\cdot)$ is not necessarily additive and $p \geq 3$. One way of addressing such an issue is to establish an effective model selection method to ensure that both the linear and the nonparametric components of the model are of the smallest possible dimension as done in Gao and Tong (2004) using a *cross–validation* type of selection procedure. Otherwise, the literature introduces alternative dimension–reduction models, namely the *semiparametric additive* and *semiparametric single–index* models. Below let us provide a detailed review for each of these models in turn.

Semiparametric Additive Model

The main ideas of proposing the semiparametric additive model can be taken from Gao et al. (2006), who have established an estimation procedure for semiparametric spatial regression. When $g(\cdot)$ itself is additive, i.e. $g(x) = \sum_{i=1}^{p} g_i(x_i)$, the form of $m_1(U_t, V_t)$ can be written as

$$m_1(U_t, V_t) = \mu + U_t^\tau \beta + \sum_{i=1}^{p} g_i(V_{ti}) \quad (5)$$

subject to $E[g_i(V_{ti})] = 0$ for all $1 \leq i \leq p$ for the identifiability of $m_1(U_t, V_t)$ in (5), where $g_i(\cdot)$ for $1 \leq i \leq p$ are all unknown one dimensional functions over $\mathbb{R}^1$. The semiparametric kernel estimation approach as discussed in the literature involves three important steps.

The *first step* is to estimate μ and $g(\cdot)$ by assuming β to be known. Observe that we have under such an assumption

$$g(x) = g(x, \beta) = E[Y_t - \mu - U_t^\tau \beta)|V_t = x] = E[(Y_t - E[Y_t] - (U_t - E[U_t])^\tau \beta)|V_t = x] \quad (6)$$

using the fact that $\mu = E[Y_t] - E[U_t^\tau \beta]$, which can be estimated by standard local linear estimation of Fan and Gijbels (1996). The *second step* is to apply the marginal integration technique of Linton and Nielsen (1995) to obtain $g_1, \ldots, g_p$ of (5) based on $g(V_t) = g(V_{t1}, \ldots, V_{tp}) = \sum_{t=1}^p g_l(V_{tl})$. Since $E[g_l(V_{tl})] = 0$ for $l = 1, \ldots, p$, we have for k fixed $g_k(x_k) = E[g(V_{t1}, \ldots, x_k, \ldots, V_{tp})]$. This method of estimating $g(\cdot)$ is therefore based on an additive marginal integration projection on the set of additive functions where, unlike the backfitting case; see Nielsen and Linton (1998); Mammen et al. (1999) for examples, the projection is taken with the product measure of V_{tl} for $l = 1, \ldots p$. Although the marginal integration technique is inferior to backfitting in asymptotic efficiency for purely additive models, it seems well suited to the framework of partially linear estimation; see also Fan et al. (1998); Fan and Li (2003) for details. The *third step* involves the estimation of β using the weighted least squares estimator $\widehat{\beta}$ of β derived in (4). The estimation procedure is completed by reintroducing $\widehat{\beta}$ into the above first and the second of steps. For the independent data case, the orthogonal series estimation method has been used as an alternative to some other nonparametric estimation methods such as the kernel method; see Eubank (1999) for example. By approximating each $g_i(\cdot)$ by an orthogonal series $\sum_{j=1}^{ni} f_{ij}(\cdot)\theta_{ij}$ with $\{f_{ij}(\cdot)\}$ being a sequence of orthogonal functions and $\{n_i\}$ being a sequence of positive integers, we have an approximate model of the form

$$Y_t = \mu + U_t^\tau \beta + \sum_{i=1}^{p} \sum_{j=1}^{n_i} f_{ij}(V_{ti})\theta_{ij} + e_t \tag{7}$$

which covers some natural parametric time series models. For example, when $U_{tl} = U_{t-l}$ and $V_{ti} = Y_{t-i}$, model (7) becomes a parametric nonlinear additive time series model

$$Y_t = \mu + \sum_{l=1}^{q} U_{t-l}\beta_l + \sum_{i=1}^{p} \sum_{j=1}^{n_i} f_{ij}(Y_{t-i})\theta_{ij} + e_t = \mu + \sum_{l=1}^{q} U_{t-l}\beta_l + \sum_{i=1}^{p} F_i(V_{ti})\theta_i(n) + e_t, \tag{8}$$

where $\theta_i = (\theta_{i1}, \ldots, \theta_{in_i})^\tau$ and $F_i = F_{in_i} = (F_i(V_{1i}), \ldots, F_i(V_{Ti}))^\tau$. The least squares estimators of (β, θ, μ) can then be defined using the approximate model (8). However, it follows from (8) that the prediction equation depends both on the series functions $\{f_{ij} : 1 \leq j \leq n_i, 1 \leq i \leq p\}$ and more importantly on n, the vector of truncation parameters. To this end, Gao (1998) applied a generalized cross–validation criterion to choose smoothing truncation parameters for the time series case; see also Wahba (1990) for an excellent survey. ■

(Extended Generalized) Semiparametric Single–Index Model

As an alternative to (5), we assume that $m(x) = E[Y_t|X_t = x] = m_2(X_t)$ is given by the semiparametric single–index

$$m_2(X_t) = X_t^\tau \theta_0 + \psi(X_t^\tau \eta_0), \tag{9}$$

where θ_0 and η_0 are unknown vector parameters and $\psi(\cdot)$ is an unknown function. Suppose that $X_t^\tau \theta_0 + \psi(X_t^\tau \eta_0)$ can be written as $X_t^\tau(\theta_0 + c\eta_0) + \psi(X_t^\tau \eta_0) - cX_t^\tau \theta_0$, the geometrical ergodicity of $\{Y_t\}$ is ensured under the following conditions.

Assumption 1. *(i) All the roots of function $x^d - (\theta_{01} + c\eta_{01})x^{d-1} - \ldots - (\theta_{0d} + c\eta_{0d})$ are inside the unit circle; (ii) $\lim_{|u|\to\infty} |\psi(u)/u| \to 0$; see also the conditions stated in Theorem 3 of Xia et al. (1999).*

Xia et al. (1999) name the functional form in (9) the *extended generalized partially linear single–index (EG–PLSI)*. In order to ensure the estimability of model, it must also be the case that θ_0 and η_0 are perpendicular to each other with $||\eta_0|| = 1$ and its first nonzero element positive. Now, define $S(\theta, \eta) = E[Y_t - \varphi_\eta(X_t^\tau \eta) - \{X_t - \Gamma_\eta(X_t^\tau \eta)\}^\tau \theta]^2$, where $\varphi_\eta(u) = E[Y_t|X_t^\tau \eta = u]$ and $\Gamma_\eta(u) = E[X_t|X_t^\tau \eta = u]$, and let $\mathcal{W}(\eta) = E[\{X - \Gamma_\eta(X_t^\tau \eta)\}\{X - \Gamma_\eta(X_t^\tau \eta)\}^\tau]$; $\mathcal{V}(\eta) = E[\{X - \Gamma_\eta(X_t^\tau \eta)\}\{Y_t - \varphi_\eta(X_t^\tau \eta)\}$. Xia et al. (1999) show that the minimum point of $S(\theta, \eta)$ with $\theta \perp \eta$ is unique at η_0 and $\theta_0 = \{\mathcal{W}(\eta_0)\}^+ \mathcal{V}(\eta_0)$, where $\{\mathcal{W}(\eta_0)\}^+$ is the Moore–Penrose inverse; see also a Theorem 2 of Xia et al. (1999).

The suggested estimation procedure of the model, which is a semiparametric extension of that introduced in Härdle et al. (1993) for a nonparametric single –index model, involves *four* important steps which can be summarized as the followings:

- The *first step* is to compute the estimate $\widehat{\theta}_\eta$ of θ given η and based on the "delete–one" estimators of φ_η and Γ_η. Let Θ denote all the unit vectors in $\mathbb{R}^p$.

- The *second step* is to estimate $\eta \in \Theta$ and the bandwidh, h, by those values $\widehat{\eta}$ and $\widehat{h}$ which minimize $\widehat{S}(\widehat{\theta}_\eta, \eta, h)$, i.e. an estimate of $S(\theta, \eta)$.

- The *third step* is to re–estimate θ as in the first step, but with η replaced by $\widehat{\eta}$.

- The *forth step* is to estimate $\psi(\cdot)$ using the nonparametric kernel estimates and the fact that $\psi(x) = \varphi(x) - \theta\Gamma(x)$.

In order to illustrate the statistical validity of such a estimation procedure, Xia et al. (1999) prove the following asymptotic results.

(i) $\sqrt{n}$-consistency:

$$\widetilde{n}\left(\widehat{\theta}-\theta_0\right) \to N(0,\mathbb{C}^+) \quad \text{and} \quad \widetilde{n}\left(\widehat{\eta}-\eta_0\right) \to N(0,\mathbb{D}^+) \tag{10}$$

in distribution, where $\widetilde{n}$ is the number of elements in $\mathcal{A} \subset \mathbb{R}$, i.e. the union of a number of open convex sets such that $f(x) > M$ for some constant $M > 0$, and $\mathbb{C}^+$ and $\mathbb{D}^+$ are some positive, finite constants; see also Corollary of Xia et al. (1999).

(ii) Uniform convergence:

$$\sup_{v \in \{\theta^T x : x \in \mathcal{A}\}} \left|\widehat{\psi}_{\widehat{\eta}}(v) - \psi(v)\right| = O\left\{(n^{-4/5}\log n)^{1/2}\right\} \tag{11}$$

almost surely; see also Theorem 5 of Xia et al. (1999).

The above results warrant a few remarks. The above asymptotic normality is a direct extension of the one presented in Härdle et al. (1993), but under α-mixing and a larger parameter cone Ω_n such that $\Omega_n = \{\eta : ||\eta - \eta_0|| \leq M_0 n^{-\delta}\}$ for some constant M_0, where $\frac{3}{10} < \delta < \frac{1}{2}$. Furthermore, the proof of such results are made possible using a decomposition of $\widehat{S}$ into various parts. While two of which are asymptotically trivial, the remaining two can be used to investigate the properties of the estimator of h_0 assuming that η_0 is known and of the estimator of η assuming h_0 is known, respectively. Xia et al. (1999) present such asymptotic results based on the Nadaraya–Watson method, but it should be noted also that such results can also be derived for local linear smoothing of Fan and Gijbels (1996).

Finally, some basic modifications to the formulation of the models bring about various special cases, which are well–known in the literature. For instance, the EG–PLSI model in (9) includes many existing models as special cases. If $\theta_0 = 0$, it reduces to

$$m_2(X_t) = \psi(X_t^\tau \eta_0), \tag{12}$$

which is the *single index* model discussed in Härdle et al. (1993). Furthermore, by partitioning $X_t = (U_t^\tau, V_t^\tau)^\tau$ and taking $\theta = (\beta^\tau, 0, \ldots, 0)^\tau$ and $\eta = (0, \ldots, 0, \alpha^\tau)^\tau$, the EG–PLSI becomes the *generalized partially linear single–index (G–PLSI)* introduced by Carroll et al. (1997) of the form

$$m_2(X_t) = U_t^\tau \theta + \psi(V_t^\tau \alpha), \tag{13}$$

which is a special case of the multiple–index model of Ichimura and Lee (1991). ■

Clearly, the *single-index*, *additive* and the *partially linear* models are non–nested. There are single–index models that are neither additive nor partially linear, additive models that are neither single–index nor partially linear, and partially linear models that are neither single–index nor additive. Chapters 7 to 9, of a recent book by Li and Racine (2011) provide a very comprehensive treatment of these models and their asymptotic properties. On the other hand, two other books by Pagan and Ullah (1999) and Horowitz (2009) provide a more basic treatment, but with an extensive discussion about various applications of such models in economics.

Finally, a recently–developed model, which is closely related to the above discussed semiparametric time series specifications is the so–called Semi–Linear model introduced by Gao (2012). For the sake of completion, let us also provide a detailed review of the model.

Semi–Linear (SL) Time Series Model

For the case by which $\{X_t\}$ is a vector of stationary time series regressors and $g(\cdot)$ is an unknown function defined on $\mathbb{R}^p$ (where $1 \leq p \leq 3$), an attempt is made in the work of Gao (2012) to extend the semiparametric partially linear models in (3) to the case

$$m_1(X_t) = \mu + X_t^\tau \beta + g(X_t), \tag{14}$$

which is a direct counterpart of the EG–PLSI model in (9). The SL model has different types of motivations and applications from the conventional semiparametric time series model presented in (3). In (14), the linear component in many cases plays the leading role while the nonparametric component behaves like a type of unknown departure from such classic linear model. Furthermore, the SL model can be motivated as a model to address some endogenous problems involved in a class of linear models of the form

$$Y_t = X_t^\tau \beta + \varepsilon_t,$$

where $\{\varepsilon_t\}$ is a sequence of errors with $E[\varepsilon] = 0$ but $E[\varepsilon_t|X_t] \neq 0$, i.e. it might be the case that $\varepsilon_t = g(X_t) + e_t$ where e_t is an i.i.d. error. Unfortunately, in the process of estimating β and $g(\cdot)$, existing methods are not directly applicable especially given the fact that $\Sigma = (E[U_t - E[U_t|U_t])(E[U_t - E[U_t|U_t])^\tau]) = 0$.

To this end, Gao (2012) studies the estimation of the SL model and its asymptotic properties in two different contexts, namely (i) $\{X_t\}$ is a *vector of stationary time series regressors*; (ii) $\{X_t\}$ is *stochastically nonstationary.* An essential assumption, which must be imposed, is to ensure the identifiability and the *"smallness"* conditions on $g(\cdot)$.

Assumption 2. *(Assumption 2.1(i) of Gao (2012)) Let $g(\cdot)$ be an integrable function $\int ||x||^i |g(x)|^i dF(x) \leq \infty$ for $i = 1, 2$ and $\int xg(x)dF(x) = 0$, where $F(x)$ is the cumulative distribution function of $\{X_t\}$ and $||\cdot||$ denotes the conventional Euclidean norm; see also Assumption 2.1 of the paper.*

Under such a condition, the parameter β is identifiable and chosen such that $E[Y_t - X_t^\tau \beta]^2$ is minimized over β, which implies $\beta = (E[X_1 X_1^\tau])^{-1} E[X_1 Y_1]$ provided that the inverse matrix does exist. Such a definition of β suggests that $\int xg(x)dF(x) = 0$ so that β can be estimated by the ordinary least squares estimator of the form

$$\widehat{\beta} = \left(\sum_{t=1}^{n} X_t X_t^\tau \right)^{-1} \left(\sum_{t=1}^{n} X_t Y_t \right) \quad \text{such that} \quad \widehat{g}(x) = \sum_{t=1}^{n} w_{nt}(x) \left(Y_t - X_t^\tau \widehat{\beta} \right), \tag{15}$$

where $w_{nt}(x)$ is a probability (kernel) weight function. Gao (2012) then establishes the asymptotic normality of such estimators. Nonetheless, the full proof of such results is not shown in the paper since it is relatively straightforward using existing results for central limit theorems for partial sums of stationary and α-mixing time series; see Fan and Yao (2003) for example.

To this end an existing hypothesis testing procedure, which can be used to determine whether $g(\cdot)$ is small enough to be negligible, is that developed by Gao (1995). The null hypothesis in this case is $H_0 : g(\cdot) = 0$, while the asymptotic distribution of the test statistic is derived as

$$\widehat{L}_{1n} = \frac{\sqrt{n}}{\widehat{\sigma}_1} \left(\frac{1}{n} \sum_{t=1}^{n} \left(Y_t - X_t^\tau \widehat{\beta} \right)^2 - \widehat{\sigma}_0^2 \right) \xrightarrow{D} N(,1), \tag{16}$$

where $\widehat{\sigma}_1^2$ and $\widehat{\sigma}_0^2$ are consistent estimators of $\sigma_1^2 = E[e_1^4] - \sigma_0^4$ and $\sigma_0^2 = E[e_1^2]$, respectively.

Finally, let us note that the EG–PLSI model discussed in the previous section can alway be used for the case by which $p \geq 4$. For the sake of convenience and clarity, we will leave the discussion on the case of a nonstationary time series to a later section. ■

3. Recent Explorations and Other Usefulnesses

During the 1980s and the early 1990s, a large number of studies on the semiparametric models focused on finding the best estimation procedures and establishing their asymptotic behavior for the case of independent data; for example Engle et al. (1986), Heckman (1986), Robinson (1988) and Fan et al. (1995). Then in the 1990s, many studies attempted to move toward time series data; for example Truong and Stone (1994) and Fan and Li (1999b). More recently; however, enormous attention has been paid on finding new extensions, additional flexibility and other usages of the above–discussed semiparametric models. In the remaining of this paper, we provides a comprehensive review of these recent development by categorizing them into a number of key areas as the followings.

Hypothesis Testings

In order to provide a brief background and introduction into the topic, let us consider firstly the following general autoregressive model of a finite order p

$$Y_t = g(Y_{t-1}, \ldots, Y_{t-p}) + \epsilon_t, \tag{17}$$

where the autoregressive function g is unknown and $\{\epsilon_t\}$ is a sequence of martingale differences. *"Among the first natural step in analysis a time series is to decide whether to use a nonlinear model or not."* (Tjøstheim (1994)) For convenience, we let $X_t = (Y_{t-1}, \ldots, Y_{t-p})^\tau$ so that we observe $X_1, \ldots, X_{n+1}$. To this end, Fan and Li (1997) establish a consistent nonparametric test for linearity of AR(p) models such that, in terms of X_t, the hypotheses can be written as

$$H_0 : P(g(X_t) = \alpha_0^\tau X_t) = 1 \text{ and } H_1 : P(g(X_t) = \alpha_0^\tau X_t) < 1 \tag{18}$$

for some $\alpha_0 \in (-1, 1)^p$ and for all $\alpha_0 \in (-1, 1)^p$, respectively. If the null hypothesis holds, then the ordinary least squares (OLS) estimator $\widehat{\alpha}$, for example, provides a consistent estimator of α_0. Furthermore, by letting $\widehat{\epsilon}_t = Y_t - \widehat{\alpha}^\tau X_t$, the test statistic of Fan and Li (1997) is based on the kernel estimate of the sample analogue $E[\epsilon_t E(\epsilon_t | X_t) f(X_t)]$, i.e.

$$I_n = \frac{1}{n(n-1)h^p} \sum \sum_{s \neq t} \widehat{\epsilon}_t \widehat{\epsilon}_s K_{st}, \tag{19}$$

where $h \equiv h_n \to 0$ is a sequence of smoothing parameters, $K_{st} = K((X_s - X_t)/h)$, $K(\cdot)$ is a kernel function satisfying certain conditions and $\sum \sum_{s \neq t} = \sum_{s=1}^n \sum_{t \neq s, t=1}^n$. Under the null hypothesis, it is the case that $\widehat{\epsilon}_t = \epsilon_t - (\widehat{\alpha} - \alpha_0)^\tau X_t$ so that the asymptotic distribution of I_n is determined by that of $nh^{p/2} I_{n1}$, where

$$I_{n1} = \frac{1}{n(n-1)h^p} \sum \sum_{s \neq t} \epsilon_t \epsilon_s K_{st}. \tag{20}$$

To this end, Fan and Li (1997) derive the asymptotic normality of $nh^{p/2} I_{n1}$ by invoking on the CLT for degenerate U–statistics of absolutely regular processes of Khashimov (1993). In addition, Fan and Li (1999a) focus on one of the condition in Khashimov (1993) which requires that the error term ϵ_t be bounded and provide a new CLT that can be used to relax such a boundedness. (Note that in the model specification testing introduced in Fan and Li (1999a) the error term is defined instead as $Y_t - g(X_t, \gamma_0)$ to reflect the null hypothesis which involve a specific parametric family.) Furthermore, the usefulness of such a result can be extended beyond the simple test for regression functional form as discussed in Fan and Li (1997) and Fan and Li (1999a). The CLT established in Fan and Li (1999a) can be used together with the $\sqrt{n}$-consistent estimation of partially linear time series models in Fan and Li (1999b) to generalize the consistent test of Fan and Li (1996) for testing a partially linear model versus a nonparametric regression model to time series framework. This work is done by Li (1999). In principle, the test statistic considered in Li (1999) has the form as that of Fan and Li (1996). However, the derivation of the asymptotic distribution of the test statistic is much more complex due mainly to the involvement of a nonparametric or a semiparametric regression under the null hypothesis, whose rate is slower than the parametric rate of $\sqrt{n}$. ■

Detection and Estimation of Trend and Seasonality

Many important macroeconomic and financial data, such as income, unemployment and retail sale, are found to exhibit deterministic/stochastic trend. A recent semiparametric model that explicitly allows for a trend detection is the *partially linear time series error model* introduced by Gao and Hawthorne (2006) of the form

$$Y_t = U_t^\tau \beta + g\left(\frac{t}{n}\right) + \varepsilon_t, \ t = 1, 2, \ldots, n, \tag{21}$$

where $\{Y_t\}$ is a response variable (e.g. the mean temperature series of interest), $U_t = (U_{t1}, \ldots, U_{tq})^\tau$ is a vector of q-explanatory variables (e.g. the southern oscillation index), t is time in years, β is a vector of unknown coefficients for the explanatory variables, $g(\cdot)$ is an unknown smooth function of time representing the trend and $\{\varepsilon_t\}$ represents a sequence of stationary time series errors with $E[e_t] = 0$ and $0 < \mathrm{var}[e_t] = \sigma^2 < \infty$. In order to estimate the model, Gao and Hawthorne (2006) introduce an estimation procedure, which is closely similar to that of the above partially linear time series model, i.e. (i) Compute an estimate of $g(\cdot)$ for a given β, i.e. similarly to the second term of (15) above; (ii) Compute the least square estimate of β using $Y_t - U_t^\tau \beta = g\left(\frac{t}{n}\right) + \varepsilon_t$; (iii) Compute the new estimate of $g(\cdot)$ based on that of β estimated in the previous step. Gao and Hawthorne (2006) also consider an alternative case by which $\{\varepsilon_t\}$ is allowed to be $I(1)$. That is to say that $\{\varepsilon_t\}$ itself may be nonstationary, but its differences $\delta_t = \varepsilon_t - \varepsilon_{t-1}$ are assumed to be stationary. In this case, we need only to consider the first differenced version of (21) of the form

$$V_t = W_t^\tau \beta + m\left(\frac{t}{n}\right) + \delta_t,\ t = 1, 2, \ldots, n, \tag{22}$$

where $V_t = Y_t - Y_{t-1}$, $W_t = U_t - U_{t-1}$, and $m\left(\frac{t}{n}\right) = g\left(\frac{t}{n}\right) - g\left(\frac{t-1}{n}\right)$.

Such models enable a study of an important question in practice that is whether a linear trend is able to adequately approximates the behaviour of the series in question. Using the model in (21), such a question can be written as the hypotheses

$$H_0 : g\left(\frac{t}{n}\right) = \alpha_0 + \gamma_0 t \quad \text{versus} \quad H_1 : g\left(\frac{t}{n}\right) \neq \alpha_0 + \gamma_0 t \tag{23}$$

for some $\theta_0 = (\alpha_0, \gamma_0) \in \Theta$ and all $\theta = (\alpha, \gamma) \in \Theta$, where Θ is a parameter space in R^2. This issue is coherent with the general interest in statistics and econometrics which involves nonparametric specification testing for linearity.

To this end, Horowitz and Spokoiny (2001) establish an adaptive test for the nonparametric mean function of a fixed–design regression model versus a parametric counterpart. Inspired by Horowitz and Spokoiny (2001), Gao and Hawthorne (2006) propose a novel test for linearity in the trend function $g(\cdot)$ under such semiparametric settings as in (21) and (22). For each given value of bandwidth h, to test $H_)$ Gao and Hawthorne (2006) propose using

$$L_n(h) = \frac{\sum_{t=1}^n \sum_{s=1,\neq t}^n K\left(\frac{s-t}{nh}\right) \widetilde{\varepsilon}_s \widetilde{\varepsilon}_t}{\widetilde{S}_n}, \tag{24}$$

where $\widetilde{S}_n^2 = 2\sum_{t=1}^n \sum_{s=1}^n K^2\left(\frac{s-t}{nh}\right) \widetilde{\varepsilon}_s^2 \widetilde{\varepsilon}_t^2$, $\widetilde{\varepsilon}_t = Y_t - U_t^\tau \widetilde{\beta} - f(t, \widetilde{\theta})$ in which $f(t, \widetilde{\theta})$ is the least-squares estimate of $f(t, \theta_0)$.

In order to perform a nonparametric kernel testing such as this, bandwidth selection can be crucial and may significantly affect the outcome of the test. A novel idea about a testing procedure in Gao and Hawthorne (2006) is the use of a maximized version of the test such that

$$L^* = \max_{h \in H_n} L_n(h). \tag{25}$$

The main theoretical results of the paper show the consistency such test statistics; see also Gao and King (2004); Gao and Gijbels (2008); Saart and Gao (2012) for related works on nonparametric kernel testing and bandwidth selection.

More recently, there is a development of a new semiparametric partially linear time series model by Chen et al. (2011). Although such a model does not assist us with the dimensionality reduction problem, the Chen et al. (2011) procedure enables a convenient estimation of the following extended version of the partially linear time series model

$$Y_t = \beta(U_t, \theta_1) + g(U_t) + \varepsilon_t, \tag{26}$$

where $\beta(\cdot, \theta_1)$ is a known link function indexed by an unknown parameter vector $\theta_1 \in \Theta \subset \mathbb{R}^p$ $(p \geq 1)$. An important point to note about the model in (26) is the fact that $\{U_t\}$ is allowed to be generated by

$$U_t = H\left(\frac{t}{n}\right) + u_t, \tag{27}$$

where $H(t)$ is unknown functions defined on $\mathbb{R}^d$ and $\{u_t\}$ is a sequence of i.i.d. random errors. In the other words, it allows for the existence of deterministic trend in the regressors. Hence, Chen et al. (2011) study a case by which nonstationarity is allowed and is driven by a deterministic trending component. Regarding the model's estimation procedure, Chen et al. (2011) provide two alternative methods, namely the nonlinear LS (see Gao (1995) and Gao (2012) for example) and the semiparametric weighted LS estimations (see Härdle et al. (2000) for example). Among these methods, the former firstly estimate θ_1, then use such an estimate is used in order to compute that of $g(\cdot)$, while the latter is just the opposite. A more important issue; however, is the identifiability and estimability of the model. The following conditions are needed in Chen et al. (2011) in order to ensure that θ_1 in (26) is identifiable and estimable.

Assumption 3. *(Assumption A2 of Chen et al. (2011)) (i) $\beta(U_t, \theta)$ is twice differentiable with respect to θ, and both $g(\cdot)$ and $H(\cdot)$ are continuous. (ii) Denoting the partial derivative of $\beta(U_t, \theta)$ with respect to θ by $\dot{\beta}(U_t, \theta)$, then*

$$\Gamma(\theta) := \int_0^1 \left\{ \int g(v) \dot{\beta}(v, \theta) p_u(v - H(r)) dv \right\} dr = 0 \tag{28}$$

for all $\theta \in \Theta$ *and*

$$\int_0^1 \left\{ \int [\beta(v,\theta_1) - \beta(v,\theta)]\dot{\beta}(v,\theta)p_u(v - H(r))dv \right\} dr \neq 0 \tag{29}$$

uniformly in $\theta \in \Theta(\delta) = \{\theta : ||\theta - \theta_1|| \leq \delta\}$ *for any* $\delta > 0$. ■

Nonstationary Time Series Models

Before discussing the recent development of semiparametric nonstationary time series models, it is useful to note the following mathematical approaches which have been established as tools to deriving an asymptotic theory for nonparametric estimation of univariate models of nonstationary data.

- Firstly, it is the *"Markov splitting technique"* used in; for example, Karlsen and Tjostheim (2001) and Karlsen et al. (2007) in order to model univariate time series with a null–recurrent structure.

- Secondly, it is the *"local–time methods"* developed by Phillips (2009) and Wang and Phillips (2011) used in order to derive an asymptotic theory for nonparametric estimation of univariate models with integrated time series.

In addition, there is an alternative model that is closely related to (26), which is discussed in Gao (2012). Unfortunately, due to the unavailability of the asymptotic results, the study focuses only on the case by which $p = 1$. The Gao's (2012) model can be obtained simply by replacing the parametric component with $x_t\beta$ and the nonparametric component with $g(x_t)$ whereby the regressor is defined as in the assumption below.

Assumption 4. *(Assumption 3.2(i) of Gao (2012)) Let* $x_t = x_{t-1} + u_t$ *with* $x_0 = 0$ *and* $u_t = \sum_{i=0}^{\infty} \psi_i \eta_{t-i}$, *where* $\{\eta_t\}$ *is a sequence of independent and identically distributed random errors, and* $\{\psi_i : i \geq 0\}$ *is a sequence of real numbers such that* $\sum_{i=0}^{\infty} i^2 |\psi_i| < \infty$ *and* $\sum_{i=0}^{\infty} \psi_i \neq 0$.

The required smallness conditions on $g(\cdot)$ are provided for two cases; namely stationary and nonstationary regressors. While those of the stationary regressors are discussed in details in our discussion of the semi–linear time series model, those of the nonstationary regressor case are the following:

Assumption 5. *(Assumption 3.1 of Gao (2012)) (i) Let $g(\cdot)$ be a real function on $\mathbb{R}^1 = (-\infty, \infty)$ such that $\int |x|^i |g(x)|^i dx < \infty$ for $i = 1, 2$, and $\int xg(x)dx \neq 0$; (ii) In addition, let $g(\cdot)$ satisfy $\int \left| \int e^{ixy} yg(y)dy \right| dx < \infty$ when $\int xg(x)dx = 0$.*

An important point to note about these assumptions is the fact that both exclude the case where $g(x)$ is a simple linear function of x.

Furthermore, since $\{x_t\}$ is nonstationary, the parameter β is identifiable and chosen such that $\frac{1}{n}\sum_{t=1}^{n}[y_t - x_t\beta]^2$ is minimized over β leading to

$$\widehat{\beta} = \left(\sum_{t=1}^{n} x_t^2\right)^{-1} \left(\sum_{t=1}^{n} x_t y_t\right), \tag{30}$$

which is closely related to that of (15). Although the detail is discussed in the paper, here we note that in order to establish an asymptotic distribution for $\widehat{\beta}$ it is expected to show that as $n \to \infty$ $\frac{1}{n}\sum_{t=1}^{n} x_t g(x_t) \to_P 0$. Regarding the case of a nonstationary regressor, $\int xg(x)dx$ "may" or "may not" be zero. The asymptotic distribution of the estimators, namely the ordinary least squares estimator of the unknown parameter β and the nonparametric estimator of $g(\cdot)$, in these cases are derived based very much on Theorem 2.1 of the above–mentioned studies by Wang and Phillips (2009a), and Wang and Phillips (2011); see also the discussion on an asymptotic study for nonstationarity in a semiparametric autoregressive model below.
■

Semiparametric Estimation in Multivariate Nonstationary Time Series Models

In the case of independent and stationary time series data, semiparametric methods have been shown to be particularly useful in modelling economic data in a way that retains generality where it is most needed while reducing dimensionality problems. Gao and Phillips (2012) seeks to pursue these advantages in a wider context that allows for nonstationarities and endogeneities within a vector semiparametric regression model. In their study, the time series $\{Y_t, X_t, V_t) : 1 \leq t \leq n\}$ are assumed to be modeled in a system of multivariate nonstationary time series models of the form

$$\begin{aligned} Y_t &= AX_t + g(V_t) + e_t && (31) \\ X_t &= H(V_t) + U_t \quad t = 1, 2, \ldots, n, && \\ E[e_t|V_t] &= E[e_t] = 0 \text{ and} && (32) \\ E[U_t|V_t] &= 0, && (33) \end{aligned}$$

where n is the sample size, A is a $p \times d$-matrix of unknown parameters, $Y_t = (y_{t1}, \ldots, y_{tp})^\tau$, $X_t = (x_{t1}, \ldots, x_{td})^\tau$ and V_t *is a sequence of univariate integrated time series regressors*, $g(\cdot) = (g_1(\cdot), \ldots, g_p(\cdot))^\tau$ and $H(\cdot) = (h_1(\cdot), \ldots, h_d(\cdot))^\tau$ are all unknown functions and both e_t and U_t are vectors of stationary time series. *Note that* $\{X_t\}$ *cab be stationary only when* $\{X_t\}$ *and* $\{V_t\}$ *are independent.* The identification condition

$$E[e_t|V_t] = E[e_t] = 0$$

in (32) eliminates endogeneity between e_t and V_t while retaining *endogeneity between* e_t *and* X_t *and potential nonstationarity in both* X_t *and* V_t. In this setting, such a condition corresponds to the condition $E[e_t|V_t, U_t] = E[e_t|U_t]$ that is assumed in Newey et al. (1999) and Su and Ullah (2008) for example. The rational behind (32) is the fact that

$$E[e_t|V_t] = E(E[e_t|U_t, V_t]|V_t) = E(E[e_t, |U_t]|V_t) = E(E[e_t|U_t]) = E[e_t]$$

when U_t is independent of V_t and $E[e_t] = 0$. These conditions are less restrictive than the exogeneity condition between e_t and (X_t, V_t) that is common in the literature for the stationary case. The model in (31) corresponds to similar structures that have been used in the independent case; see Su and Ullah (2008) for example. In the study by Gao and Phillips (2012), the model is treated as *a vector semiparametric structural model* and considers the case where X_t and V_t may be vectors of nonstationary regressors and X_t may be endogenous. The main contribution of the study resides in the derivation of a semiparametric instrumental variable least squares (SIV) estimate of A; and its asymptotic properties, to deal with endogeneity in X_t and a nonparametric estimator for the function $g(\cdot)$. Let us assume that there exists a vector of stationary variables η_t for which

$$E[U_t \eta_t^\tau] \neq 0 \text{ and } E[e_t|\eta_t] = 0.$$

Then the derivation of the SIV estimate of A is done based on the following expanded version of the system (31)

$$\begin{aligned} Y_t &= AX_t + g(V_t) + e_t \quad t = 1, 2, \ldots, n && (34) \\ X_t &= H(V_t) + U_t \\ Q_t &= J(V_t) + \eta_t \\ E[e_t|V_t] &= E[e_t] = 0, \;\; E[U_t|V_t] = 0 \text{ and } E[\eta_t|V_t] = 0, && (35) \end{aligned}$$

where $Q_t = (q_{t1}, \ldots, q_{td})^\tau$ is a vector of possible instrumental variables for X_t generated by a reduced form equation involving V_t, and $I(\cdot) = (J_1(\cdot), \ldots, J_d(\cdot))^\tau$ is a vector of unknown functions. The limit theory in this kind of nonstationary semiparametric model depend on the probabilistic structure of the regressors and errors e_t, U_t, η_t and V_t as well as the functional forms of $g(\cdot)$, $H(\cdot)$ and $J(\cdot)$. Gao and Phillips (2012) provide a list of conditions required including their detailed explanation in Append A of the paper. ■

Semiparametric Models with Generated Regressors

In econometrics, estimation problems with "generated regressors" (or "generated covariates") are rather common. Some particular contexts that we may be facing with such problems are estimation problems with endogeneity, simultaneous equation systems, link function testing with unknown indices, selection bias problems and unobserved variables like missing or data mathching problems; see also Sperlich (2009) for details. Pagan (1984) and Oxley and McAleer (1993) study the generated–regressor problem in the parametric context. Furthermore, Rodríguez-Póo et al. (2005), and Lewbel and Linton (2007) respectively analyze semiparametric simultaneous equation systems and deal with nonparametrically generated regressors when considering homothetically separable functions. However, establishing asymptotic properties of semiparametric models with generated regressors is not a straightforward task and to the best of our knowledge not a significant number of studies in statistic and econometric literature have attempted to do so. An example is a recent work by Mammen et al. (2011), who study a general class of semiparametric optimization estimators of a vector–valued parameter. The main contribution of the paper is a new stochastic expansion that characterizes the influence of generated covariates in the model's nonparametric component on the asymptotic properties of the final estimator. The usefulness of such a new expansion resides in the asymptotic penalty that must be incurred when other methods, for example the Taylor expansion, are employed; see the review of a work by Li and Wooldridge (2002) below for more details.

In addition, there are many examples of time series models that contain generated regressors. For instance the so–called rational expectation model; i.e. a model by which a surprise term represented by the error of a parametric model but not the exogenous variable itself, has an impact on the response variable. Another obvious example is the error correction model (ECM) of Engle and Granger (1987). To this end, Li and Wooldridge (2002) introduce an alternative formulation of the semiparametric partially linear model. By letting $\mathcal{W}_t = \{Y_t, U_t^\tau, S_t, Z_t^\tau\}$ be a stationary and absolutely regular process; see Section 2 of the paper for detials, the Li and Wooldridge's (2002) model can be written as

$$Y_t = U_t^\tau \beta + g(\eta_t) + \varepsilon_t \tag{36}$$

with $\eta_t = S_t - Z_t^\tau \alpha$ such that $E(\varepsilon_t|U_t, Z_t, \eta_t) = 0$ and $E(\eta_t|U_t, Z_t) = 0$, where U_t is $p \times 1$, Z_t is $q \times 1$, Y_t and S_t are scalars, β and α are vectors of unknown parameters, and $g(\cdot)$ is an unknown smooth function. The model can be modified so that nested within it are a nonlinear regression and a Tobit–3 models. Such model and modifications are found to be very useful in practice; see for example Galego and Pereira (2010), and Bachmeier (2002) for some interesting applications.

Overall, the model's estimation procedure is closely similar to that introduced in Robinson (1988), which we discussed earlier. The only exception in this case is the fact that the parametric estimation of η is required in the first step. Hence, the mathematical proof of the $\sqrt{n}$-consistency of the unknown parameters must rely on an assumption that there exists a $\sqrt{n}$-consistent estimator of α. Compared to those in Robinson (1988), Li and Wooldridge (2002) must impose slightly stronger moment and smoothness conditions on the regression, density, and the kernel functions. This is mainly due to the fact that they must use Taylor expansions in their proof to deal with the regressor η_t, which is initially parametrically–generated.

A similar problem is also encountered by Saart et al. (2012) in order to develop their so–called semiparametric autoregressive conditional duration (SEMI–ACD) model. In their study, the unobservable regressor is computed semiparametrically based on an iterative estimation algorithm. Saart et al. (2012) first derive the uniform consistency of the estimation algorithm, then use the Taylor expansions together with the uniform convergence rates for kernel estimation with dependent data derived in Hansen (2008) in the proof to deal with the generated regressor.

Let Y_t denotes the financial duration in question, i.e. the waiting time between two consecutive financial events, associated with the t-th event. Engle and Russell (1998) develop the ACD model by assuming that

$$Y_t = \psi_t \varepsilon_t, \tag{37}$$

where $\{Y_t\}$ is the stationary process of financial durations, $\{\varepsilon_i\}$ is an independent and identically distributed (i.i.d.) innovation series with non–negative support density $p(\varepsilon;\phi)$, in which ϕ is a vector of parameters and

$$\psi_t \equiv \sum_{j=1}^{p} \alpha_j Y_{t-j} + \sum_{k=1}^{q} \beta_k \psi_{t-k} \tag{38}$$

with $\{\psi_t\}$ denotes the process of conditional expectation which summarizes the dynamics of the duration process. The ACD model in (37) is considered by many to be too restrictive to take care of the dynamics of the duration process in practice. Furthermore, to estimate the model requires an imposition of a distributional assumption on ε_t, a requirement that is not popular in the literature; see Pacurar (2008) for an excellent review of the ACD literature. Saart et al. (2012) attempt to minimize impacts of such issues by introducing SEMI–ACD model such that

$$\psi_t \equiv \sum_{j=1}^{p} \gamma_j Y_{t-j} + \sum_{k=1}^{q} \mathsf{g}_k\left(\psi_{t-k}\right), \tag{39}$$

where γ_j are unknown parameters and $\mathsf{g}_k(\cdot)$ are unknown functions on the real line. Even though the above–mentioned distributional assumption is not required to estimate such a semiparametric model, there exists a latency problem which arises due to the fact that the conditional duration (ψ) is not observable in practice. To estimate the model the authors rely on an iterative estimation algorithm, which is established using a similar idea to the one introduced previously in Bühlmann and McNeil (2002).

For a special case of the model by which $p = q = 1$, i.e. the so–called SEMI–ACD(1,1), the algorithm can be summarized as the following:

Step 1: Choose the starting values for the vector of the n conditional durations. Index these values by a zero. Let $\{\widehat{\psi}_{t,0}; 1 \le t \le n\}$ satisfy $\widehat{\psi}_{t,0} = \psi_{t,0}$. Set $m = 1$.

Step 2: Compute $\widehat{\gamma}_m$ and $\widehat{g}_{h,m}$, by regressing $\{Y_t; 2 \le t \le n\}$ against $\{Y_{t-1}; 2 \le t \le n\}$ and the estimates of ψ computed in the previous step, i.e. $\{\widehat{\psi}_{t-1,m-1}; 2 \le t \le n\}$.

Step 3: Compute $\{\widehat{\psi}_{t,m}; 2 \le t \le n\}$. Furthermore, use the average of $\{\widehat{\psi}_{t,m}; 2 \le t \le n\}$ as a proxy for $\widehat{\psi}_{1,m}$, which cannot be computed recursively.

Step 4: For $1 \le m < m^*$, where $m^* = O\left(\log(n)\right)$ is the (pre–specified) maximum number of iterations, increment m and return to Step 2. At $m = m^*$, perform the final estimation to obtain the final estimates of γ and g.

Saart et al. (2012) study the asymptotic properties of such a procedure for the SEMI–ACD(1,1) model by first deriving the consistency of the estimation algorithm, i.e.

$$\left|\left|\widehat{\Psi}_m - \Psi\right|\right|_{1e} \leq \Delta_{1n}(\widehat{\psi})\ C_m(G) + G^m\ \Delta_{2n}(\psi), \tag{40}$$

where $0 < G < 1$, $\widehat{\Psi}_m = (\widehat{\psi}_{m+1,m}, \ldots, \widehat{\psi}_{n,m})^\tau$ and $\Psi = (\psi_{m+1}, \ldots, \psi_n)^\tau$. Although details are shown in Theorem 3.1 of the paper, here let us simply note that while the first term on the right side of (40) converges to zero uniformly over all the possible values of the bandwidth, h, at the same rate as the mean squared error of a usual partially linear time series model, the limit of the second term is zero as $m \to 0$. Hence, m^* in this case is selected so that the second term is bounded in probability by the first. Note that the proof of (40) requires the following contraction property on g.

Assumption 6. *(Assumption 3.1 of Saart et al. (2012)) Suppose that function g on the real line satisfies the following Lipschitz type condition:*

$$|g(x+\delta) - g(x)| \leq \varphi(x)|\delta| \tag{41}$$

for each given $x \in S_\omega$, where S_ω is a compact support. Furthermore, $\varphi(\cdot)$ is a nonnegative measurable function such that with probability one,

$$\max_{i\geq 1} E\left[\varphi^2(\psi_i)|(\psi_{i-1}, \cdots, \psi_1)\right] \leq G^2;\ \max_{i\geq 1} E\left[\varphi^2(\psi_{i,m})|(\psi_{i-1,m-1}, \cdots, \psi_{1,1})\right] \leq G^2 \tag{42}$$

for some $0 < G < 1$.

The asymptotic normality of the LS estimator of the unknown parameter is then proved using the Taylor expansions together with the convergence rates for kernel estimation derived in Hansen (2008). Saart et al. (2012) also shows that an extension of the SEMI–ACD(1,1) model to, for example, a SEMI–ACD(p,q) model, where $q \leq 3$, is also possible without requiring an additional assumption. This is in the sense that the $\sqrt{n}$-asymptotic normality of the concerned estimators still holds provided that $q \leq 3$. Such a claim is supported by the results found in Robinson (1988) and Fan and Li (1999b).

An important issue about the construction of both of the parametric ACD and SEMI–ACD models mentioned above is the fact that it is done under the stationarity assumption of the duration process. However, it is well known that intraday financial data often involve strongly some diurnal patterns. Hence, the first step toward the econometric analysis of financial duration is always to perform a diurnal adjustment. The remaining problems; nonetheless, are other trading patterns; for example the so–called day–of–the–week effects, that are not taken care off. An idea, which is a work in progress, is to use the fact that the extended generalized single–index model in (9) allows a shape–invariant modeling and structural analysis; see Härdle and Marron (1990), and joinly model the regular components and the dynamics of the duration process.

Finally, a close similarity of the ACD to the famous *generalized autoregressive conditional heteroskedasticity (GARCH)* model suggests that the methods introduced in Saart et al. (2012) can also be used in an establishment of a semiparametric GARCH (SEMI–GARCH) model; see also Bühlmann and McNeil (2002).

Nonlinear Autoregressive Models

If the observations are allowed to be taken over time, then the above mentioned semiparametric models may give rise to a few well–known nonlinear autoregressive models, which have been discussed in the literature. Firstly, a similar partitioning of X_t to that of (13) such that $U_t = (Y_{t-c_1}, Y_{t-c_2}, \ldots, Y_{t-c_p})^\tau$ and $V_t = (Y_{t-d_1}, Y_{t-d_2}, \ldots, Y_{t-d_p})^\tau$, where $c_i \neq d_j$ for all $1 \leq i \leq p$ and $1 \leq j \leq q$, give rises to the *autoregressive semiparametric partially linear additive* model discussed in Gao and Yee (2000), who have found that the partially linear regression is more appropriate than a completely nonparametric autoregression for some sets of data. Secondly, the *autoregressive single–index* model discussed in Xia et al. (1999) is obtained simply by letting $X_t = (Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p})^\tau$ in (9). Xia et al. (1999) have found, while using the projection pursuit method to investigate the autoregressive relationship between Y_t and X_t, where Y_t = sunspot number in year, some strong empirical evidence in support of such a model. Furthermore, since the statistical insignificance of the nonlinear component signals superiority of a linear model, the model can be tested against linearity; a testing procedure for such a hypothesis is being developed in the meantime and an up to date draft of this work is available from the authors upon request. In addition, a semiparametric autoregressive partially linear counterpart of the Xia et al. (1999) model, i.e. an autoregressive version of (14), is studied by Gao (2012). In this case, the process $\{Y_t\}$ is stochastically stationary and α-mixing under the following conditions:

Assumption 7. *(Assumption 4.1 of Gao (2012)) (i)* $\beta = (\beta_1, \ldots, \beta_p)^\tau$ *satisfy* $Y^p - \beta_1 Y^{p-1} - \ldots - \beta_{p-1} Y - \beta_p \neq 0$ *for any* $|Y| \geq 1$*; (ii)* $g(X)$ *is bounded on any bounded Borel measurable* *set and satisfy $g(X) = o(||X||)$ as $||X|| \to \infty$, where $||\cdot||$ denotes the conventional Euclidean norm.*

Such conditions together with Assumption 2 above suggest that the estimation of β and $g(\cdot)$ can be done in the same way as that of the semi–linear time series models.

On the contrary, an analysis of the nonstationarity of $\{Y_t\}$ may be done only for the case by which $d = 1$ since some essential results for such integrated time series are currently available only for nonparametric estimation of univariate models; see Phillips (2009) and Wang and Phillips (2011) for example. Furthermore, such analysis can be quite complicated given that nonstationarity may be driven by either $\beta = 1$ or $g(\cdot)$ being too explosive or a mixture of both. To perform such an analysis, Gao (2012) imposes various conditions as listed in his Assumption 4.2 in order to ensure that $\{Y_t\}$ is a λ-null recurrent Markov chain with $\lambda = \frac{1}{2}$ so that the asymptotic properties in this case can be derived based on the results in Karlsen et al. (2007).

More recently, there has also been an attempt by Gao et al. (2012a) to detect and to estimate a structural change from a nonlinear stationary regime to a linear nonstationary regime using a *semiparametric threshold autoregressive* model, which can be conveniently expressed as

$$\begin{aligned} Y_t &= g(Y_{t-1}) I[Y_{t-1} \in C_\tau] + \alpha Y_{t-1} I[Y_{t-1} \in D_\tau] + \varepsilon_t \\ &= \begin{cases} g(Y_{t-1}) + \varepsilon_t & \text{if } Y_{t-1} \in C_\tau \\ \alpha Y_{t-1} + \varepsilon_t & \text{if } Y_{t-1} \in D_\tau, \end{cases} \end{aligned} \tag{43}$$

where C_τ is either a compact subset of R^1 or a set of type $(-\infty, \tau]$ or $[\tau, \infty)$, D_τ is the complement of C_τ, $g(x)$ is an unknown and bounded function when $x \in C_\tau$ and $\alpha = 1$. It is shown in Lemma 3.1 of the paper that such a special case of the model through which $\alpha = 1$ is a β-null recurrent Markov Chain process; see also a detailed discussion on a null recurrent process in Karlsen et al. (2007). As the results, the existing asymptotic results for the stationary nonlinear time series models; for instance in Fan and Yao (2003); Gao (2007), are not directly applicable. While Gao et al. (2012a) study the asymptotic behaviour of both a nonparametric estimator of $g(\cdot)$ and the least square estimator of α, their mathematical proof relies heavily on a number of general results of the β-null recurrent Markov chains discussed in Karlsen and Tjostheim (2001). ■

4. Conclusions

We have seen in literature that theoretical and empirical research in time series analysis may be conducted within a large number of topics. Among these, we personally believe that perhaps nonlinear time series is one of the most studied over the recent years. In order to take into account the nonlinearity in time series regression, nonparametric methods have been very popular both for prediction and characterising nonlinear dependence. However, its usefulness has been significantly dampen due to the so–called curse of dimensionality. Semiparametric time series methods may offer a valuable assistance in dealing with such a problem. However, these models are nonnested and each possesses different sets of strengths and weaknesses. In this paper, we have provided a selective review on various aspects of the three of the most well–known semiparametric time series models in the literature. Furthermore, we have comprehensively discussed their recent extensions and other usefulnesses which are discussed in a more recent literature. We have also focused on a number of important issues, such as, hypothesis testings, nonlinear autoregressive models, semiparametric models with generated regressors, semiparametric estimation in multivariate nonstationary time series models.

5. References

Ait-Sahalia, Y., 1996. Testing continuous-time models of the spot interest rate. Review of Financial Studies 9 (2), 385–426.

Aït-Sahalia, Y., 2002. Transition densities for interest rate and other nonlinear diffusions. The Journal of Finance 54 (4), 1361–1395.

Aït-Sahalia, Y., Lo, A., 2002. Nonparametric estimation of state-price densities implicit in financial asset prices. The Journal of Finance 53 (2), 499–547.

Bachmeier, L., 2002. Is the term structure nonlinear? a semiparametric investigation. Applied Economics Letters 9 (3), 151–153.

Bühlmann, P., McNeil, A., 2002. An algorithm for nonparametric GARCH modelling. Computational Statistics & Data Analysis 40, 665–683.

Carroll, R., Fan, J., Wand, M., 1997. Generalized partially linear single-index models. Journal of the American Statistical Association 92 (438), 477–489.

Chan, K., 1993. Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. The Annals of Statistics 21 (1), 520–533.

Chan, K., Karolyi, G., Longstaff, F., Sanders, A., 2012. An empirical comparison of alternative models of the short-term interest rate. The Journal of Finance 47 (3), 1209–1227.

Chen, J., Gao, J., Li, D., 2011. Estimation in semiparametric time series regression. Statistics and Its Interface 4 (4), 243–252.

Cook, R., 1998. Regression Graphics. Wiley New York:.

Cox, J., Ingersoll, E., Ross, S. A., 1985. A theory of the term structure of interest rates1. Econometrica 53 (2), 385–407.

Dufrénot, G., Mignon, V., 2002. Recent developments in nonlinear cointegration with applications to macroeconomics and finance. Springer.

Engle, R., Granger, C., 1987. Co-integration and error correction: representation, estimation, and testing. Econometrica: journal of the Econometric Society, 251–276.

Engle, R., Russell, J., 1998. Autoregressive conditional duration: a new model for irregularly spaced transaction data. Econometrica 66, 1127–1162.

Engle, R. F., Granger, C., Rice, J., Weiss, A., 1986. Semiparametric estimates of the relationship between weather and electricity sales. Journal of the American Statistical Association 81 (394).

Eubank, R., 1999. Nonparametric regression and spline smoothing. Vol. 157. CRC Press.

Fan, J., 2005. A selective overview of nonparametric methods in financial econometrics. Statistical Science 20 (4), 317–337.

Fan, J., Gijbels, I., 1996. Local polynomial modelling and its applications. Chapman & Hall/CRC.

Fan, J., Härdle, W., Mammen, E., 1998. Direct estimation of low-dimensional components in additive models. The Annals of Statistics 26 (3), 943–971.

Fan, J., Yao, Q., 2003. Nonlinear time series: Nonparametric and parametric methods. Springer Verlag.

Fan, J., Yao, Q., Tong, H., 1996. Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. Biometrika 83 (1), 189–206.

Fan, Y., Li, Q., 1996. Consistent model specification tests: omitted variables and semiparametric functional forms. Econometrica: Journal of the Econometric Society, 865–890.

Fan, Y., Li, Q., 1997. A consistent nonparametric test for linearity of ar (p) models. Economics Letters 55 (1), 53–59.

Fan, Y., Li, Q., 1999a. Central limit theorem for degenerate u-statistics of absolutely regular processes with applications to model specification testing. Journal of Nonparametric Statistics 10 (3), 245–271.

Fan, Y., Li, Q., 1999b. Root-n-consistent estimation of partially linear time series models. Journal of Nonparametric Statistics 11 (1-3), 251–269.

Fan, Y., Li, Q., 2003. A kernel-based method for estimating additive partially linear models. Statistica Sinica 13 (3), 739–762.

Fan, Y., Li, Q., Stengos, T., 1995. Root-n-consistent semiparametric regression with conditional heteroscedastic disturbances. Journal of Quantitative Economics 11, 229–140.

Galego, A., Pereira, J., 2010. Evidence on gender wage discrimination in portugal: Parametric and semiparametric approaches. Review of Income and Wealth 56 (4), 651–666.

Gao, J., 1995. Parametric test in a partially linear model. Acta Mathematica Scientia (English Edition) 15, 1–10.

Gao, J., 1998. Semiparametric regression smoothing of non-linear time series. Scandinavian Journal of Statistics 25 (3), 521–539.

Gao, J., 2007. Nonlinear Time Series: Semiparametric and Nonparametric Methods. Chapman & Hall/CRC, London.

Gao, J., 2012. Identification, estimation and specification in a class of semi–linear time series models. Monash Econometrics and Business Statistics Working Papers. Available at `http://ideas.repec.org/f/pga362.html`.

Gao, J., Gijbels, I., 2008. Bandwidth selection in nonparametric kernel testing. Journal of the American Statistical Association 103 (484), 1584–1594.

Gao, J., Hawthorne, K., 2006. Semiparametric estimation and testing of the trend of temperature series. Econometrics Journal 9, 332–355.

Gao, J., King, M. L., 2004. Adaptive testing in continuous-time diffusion models. Econometric Theory 20 (05), 844–882.

Gao, J., King, M. L., 2012. An improved nonparametric unit–root test. Monash Econometrics and Business Statistics Working Papers 16/12. Available at `http://ideas.repec.org/f/pga362.html`.

Gao, J., King, M. L., Lu, Z., Tjøstheim, D., 2009a. Nonparametric specification testing for nonlinear time series with nonstationarity. Econometric Theory 25 (06), 1869–1892.

Gao, J., King, M. L., Lu, Z., Tjøstheim, D., 2009b. Specification testing in nonlinear and nonstationary time series autoregression. The Annals of Statistics 37 (6B), 3893–3928.

Gao, J., Lu, Z., Tjøstheim, D., 2006. Estimation in semiparametric spatial regression. The Annals of Statistics 34 (3), 1395–1435.

Gao, J., Phillips, P. C. B., 2012. Semiparametric estimation in multivariate nonstationry time series models. Monash Econometrics and Business Statistics Working Papers. Available at http://ideas.repec.org/f/pga362.html.

Gao, J., Tjøstheim, D., Yin, J., 2012a. Estimation in autoregressive threshold models with nonstationarity. Journal of Econometrics (in press).

Gao, J., Tjøstheim, D., Yin, J., 2012b. Model specification between parametric and nonparametric cointegration. Monash Econometrics and Business Statistics Working Papers 18/12. Available at http://ideas.repec.org/f/pga362.html.

Gao, J., Tong, H., 2004. Semiparametric non-linear time series model selection. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 66 (2), 321–336.

Gao, J., Yee, T., 2000. Adaptive estimation in partially linear autoregressive models. Canadian Journal of Statistics 28, 571–586.

Hansen, B., 2008. Uniform convergence rates for kernel estimation with dependent data. Econometric Theory 24, 726–748.

Härdle, W., Hall, P., Ichimura, H., 1993. Optimal smoothing in single-index models. The annals of Statistics 21 (1), 157–178.

Härdle, W., Liang, H., Gao, J., 2000. Partially Linear Models. Springer-Verlag, Berlin.

Härdle, W., Lütkepohl, H., Chen, R., 2007. A review of nonparametric time series analysis. International Statistical Review 65 (1), 49–72.

Härdle, W., Marron, J., 1990. Semiparametric comparison of regression curves. The Annals of Statistics, 63–89.

Heckman, N., 1986. Spline smoothing in a partly linear model. Journal of the Royal Statistical Society, Series B 48, 244–248.

Horowitz, J., 2009. Semiparametric and nonparametric methods in econometrics. Vol. 692. Springer.

Horowitz, J., Spokoiny, V., 2001. An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. Econometrica 69 (3), 599–631.

Ichimura, H., Lee, L., 1991. Semiparametric least squares estimation of multiple index models: single equation estimation. In: Nonparametric and semiparametric methods in econometrics and statistics: Proceedings of the Fifth International Symposium in Economic Theory and Econometrics. Cambridge. pp. 3–49.

Karlsen, H., Myklebust, T., Tjøstheim, D., 2007. Nonparametric estimation in a nonlinear cointegration type model. The Annals of Statistics 35 (1), 252–299.

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

Karlsen, H. A., Myklebust, T., Tjøstheim, D., 2010. Nonparametric regression estimation in a null recurrent time series. Journal of Statistical Planning and Inference 140 (12), 3619–3626.

Khashimov, S., 1993. Limiting behavior of generalized u-statistics of weakly dependent stationary processes. Theory of Probability & Its Applications 37 (1), 148–150.

Lewbel, A., Linton, O., 2007. Nonparametric matching and efficient estimators of homothetically separable functions. Econometrica 75 (4), 1209–1227.

Li, K., 1991. Sliced inverse regression for dimension reduction. Journal of the American Statistical Association 86 (414), 316–327.

Li, Q., 1999. Consistent model specification tests for time series econometric models. Journal of Econometrics 92 (1), 101–147.

Li, Q., Racine, J., 2011. Nonparametric econometrics: Theory and practice. Princeton University Press.

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

Linton, O., Nielsen, J., 1995. A kernel method of estimating structural nonparametric regression based on marginal integration. Biometrika 82, 93–100.

Mammen, E., Linton, O., Nielsen, J., 1999. The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. The Annals of Statistics 27 (5), 1443–1490.

Mammen, E., Rothe, C., Schienle, M., 2011. Semiparametric estimation with generated covariates.

Newey, W., Powell, J., Vella, F., 1999. Nonparametric estimation of triangular simultaneous equations models. Econometrica 67 (3), 565–603.

Nielsen, J., Linton, O., 1998. An optimization interpretation of integration and back-fitting estimators for separable nonparametric models. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 60 (1), 217–222.

Oxley, L., McAleer, M., 1993. Econometric issues in macroeconomic models with generated regressors. Journal of Economic Surveys 7 (1), 1–40.

Pacurar, M., 2008. Autoregressive conditional duration models in finance: a survey of the theoretical and empirical literature. Journal of Economic Surveys 22, 711–751.

Pagan, A., 1984. Econometric issues in the analysis of regressions with generated regressors. International Economic Review 25 (1), 221–247.

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

Phillips, P., 2009. Local limit theory and spurious nonparametric regression. Econometric Theory 25 (06), 1466–1497.

Robinson, P., 1988. Root-N-consistent semiparametric regression. Econometrica 56, 931–954.

Rodríguez-Póo, J., Sperlich, S., Fernández, A., 2005. Semiparametric three-step estimation methods for simultaneous equation systems. Journal of Applied Econometrics 20 (6), 699–721.

Saart, P. W., Gao, J., 2012. An alternative nonparametric specification test in nonlinear autoregressive conditional duration model. Unpublished Working Paper: Available at SSRN: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2130454.

Saart, P. W., Gao, J., Allen, D., 2012. Semiparametric autoregressive conditional duration model: Theory and practice. Unpublished Working Paper: Available at SSRN: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2109200.

Sperlich, S., 2009. A note on non-parametric estimation with predicted variables. The Econometrics Journal 12 (2), 382–395.

Su, L., Ullah, A., 2008. Local polynomial estimation of nonparametric simultaneous equations models. Journal of Econometrics 144 (1), 193–218.

Sun, Y., Li, Q., 2012. Nonparametric and semiparametric estimation and hypothesis testing with nonstationary time series. Working Paper.

Terasvirta, T., Tjostheim, D., Granger, C., 2011. Modelling nonlinear economic time series. OUP Catalogue.

Tjøstheim, D., 1994. Nonlinear time series: a selective review. Scandinavian Journal of Statistics 21, 97–130.

Tong, H., 2001. A personal journey through time series in biometrika. Biometrika 88 (1), 195–218.

Tong, H., Lim, K., 1980. Threshold autoregression, limit cycles and cyclical data. Journal of the Royal Statistical Society. Series B (Methodological), 245–292.

Truong, Y., Stone, C., 1994. Semiparametric time series regression. Journal of Time Series Analysis 15 (4), 405–428.

Van Dijk, D., Teräsvirta, T., Franses, P., 2002. Smooth transition autoregressive modelsa survey of recent developments. Econometric Reviews 21 (1), 1–47.

Vasicek, O., 1977. An equilibrium characterization of the term structure. Journal of Financial Economics 5 (2), 177–188.

Wahba, G., 1990. Spline models for observational data. Vol. 59. Society for Industrial Mathematics.

Wang, Q., Phillips, P., 2009a. Asymptotic theory for local time density estimation and nonparametric cointegrating regression. Econometric Theory 25 (3), 710.

Wang, Q., Phillips, P., 2009b. Structural nonparametric cointegrating regression. Econometrica 77 (6), 1901–1948.

Wang, Q., Phillips, P., 2011. Asymptotic theory for zero energy functionals with nonparametric regression applications. Econometric Theory 27 (2), 235.

Xia, Y., Tong, H., Li, W., 1999. On extended partially linear single-index models. Biometrika 86 (4), 831–842.

Xia, Y., Tong, H., Li, W., Zhu, L., 2002. An adaptive estimation of dimension reduction space. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64 (3), 363–410.