



City Research Online

City St George's, University of London

Citation: Wongsa-art, P., Gao, J. & Kim, N. H. (2014). Semiparametric methods in nonlinear time series analysis: a selective review. *Journal of Nonparametric Statistics*, 26(1), pp. 141-169. doi: 10.1080/10485252.2013.840724

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/33941/>

Link to published version: <https://doi.org/10.1080/10485252.2013.840724>

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to [City Research Online policy](#).

Semiparametric Methods in Nonlinear Time Series Analysis: A Selective Review

PATRICK SAART^{†12}, JITI GAO[‡] AND NAM HYUN KIM[†]
University of Canterbury[†] and Monash University[‡]

Abstract

Time series analysis is a tremendous research area in statistics and econometrics. In a previous review, the author was able break down up to fifteen key areas of research interest in time series analysis. Nonetheless, the aim of the review in this current paper is not to cover a wide range of somewhat unrelated topics on the subject, but the key strategy of the review in this paper is to begin with a core, the “curse of dimensionality” in nonparametric time series analysis, and explore further in a metaphorical domino-effect fashion into other closely related areas in semiparametric methods in nonlinear time series analysis.

JEL Classification: C12, C14, C22

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

1. Introduction

In time series regression, nonparametric methods have been quite popular both for prediction and for 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 or equal to three. When d is greater than three, the conditional mean can still be estimated using the NW estimator and asymptotic theory can be constructed. However, due to a well-known problem often referred to in the literature as the curse of dimensionality, this may not be recommended in practice unless the number of data points is extremely large. There are multiple phenomena in the literature which are referred to as the curse of dimensionality in various domains, e.g. numerical analysis, sampling, combinatorics, etc. For the sake of clarity, let us give a simple example of the curse of dimensionality in nonparametric regression.

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

$$U = g(V) + \text{noise with a mean of zero.}$$

¹Patrick 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>

²The authors thank the three anonymous referees for their comments and suggestions.

29 The data $(\mathbf{V}, \mathbf{U}) \doteq \{(V_t, U_t)\}_{t=1}^n$ are assumed to be drawn identically and indepen-
 30 dently distributed (i.i.d.) from a distribution over a joint input-output space $\mathcal{V} \times \mathcal{U}$.
 31 The input space \mathcal{V} is usually assumed to be a subset of \mathbb{R}^d , i.e. V is a vector of d
 32 features. The output space \mathcal{U} is assumed to be a subset of \mathbb{R}^d and is a random vec-
 33 tor satisfying $E[U|V = v] = g(v)$. An objective of the nonparametric regression is to
 34 approximate g with a nonparametric regressor, say, g_n . Under some smoothing condi-
 35 tion of g ; e.g. Lipschitz, a number of nonparametric estimators can be shown to satisfy
 36 $E_{\mathbf{V}, \mathbf{U}} \|g_n - g\|^2 \leq O(n^{-2/(2+d)})$. For instance, this is the rate for a kernel estimator. Such
 37 a rate implies that we need a sample size n exponential in d in order to approximate g .
 38 Hence, when d is high, as is often the case in modern applications, $n > 2^d$ is impractical.
 39 Furthermore, to get some intuition into the reason for such a rate, consider that non-
 40 parametric approaches, such as the NW estimator, operate by approximating the target
 41 function locally (on its domain \mathcal{V}) by simpler functions. There are necessarily some local
 42 errors and these errors aggregate globally. To approximate the entire function well, we
 43 need to do well in most local areas. Suppose, for instance, that the target function is well
 44 approximated by constants in regions with a radius of at most $0 < r < 1$. In how many
 45 ways can we divide up the domain \mathcal{V} into smaller regions with a radius of at most r ? If
 46 \mathcal{V} is d -dimensional then the smallest such partition is of size $O(r^{-d})$. We will need data
 47 points to fall into each such region if we hope to do well locally everywhere. In the other
 48 words, we will need a data set that is exponential in d in size. ■

49 Over recent years, a number of review papers on nonparametric and semiparametric
 50 methods have become available in the literature. Below, let us introduce a few of those
 51 that are the most relevant to the materials presented in the current paper. Firstly, there
 52 are two books by Härdle et al. (2000) and Gao (2007) that introduced some nonparametric
 53 and semiparametric nonlinear time series models as well as establishing various new results
 54 to enrich the literature. Meanwhile, a review by Fan (2005) focuses on nonparametric
 55 techniques used for estimating stochastic diffusion models, especially the drift and the
 56 diffusion functions, based on either discretely or continuously observed data. The paper
 57 begins with a brief review of some useful stochastic models for modeling stock prices and
 58 bond yields, which includes the Cox Ingersoll Ross model by Cox et al. (1985), Vasicek
 59 model by Vasicek (1977) and the Chan Karolyi Longstaff Sanders model of Chan et al.
 60 (2012). Furthermore, the paper reviews in the paper techniques for estimating state price
 61 densities and transition densities, and their applications in asset pricing and testing for
 62 parametric diffusion models. Some important references in that review are Aït-Sahalia
 63 (1996), Aït-Sahalia (2002), Aït-Sahalia and Lo (2002), and Fan et al. (1996).

64 Secondly, the review by Härdle et al. (2007) provides a fairly broad survey of many

65 nonparametric analysis techniques for time series. Specifically, the review discusses non-
66 parametric methods for estimating the spectral density, the conditional mean, higher order
67 conditional moments and conditional densities. Density estimation with correlated data,
68 bootstrap methods for time series and nonparametric trend analysis were also reviewed.

69 Finally, the two review papers by Gao (2012) and Sun and Li (2012) summarize some
70 recent theoretical developments in nonparametric and semiparametric techniques as ap-
71 plied to nonstationary or near-nonstationary variables. The first paper introduces a class
72 of semi-linear time series models that incorporate both nonstationarity and endogeneity.
73 The author also introduces and then discusses a class of the so-called “nearly integrated”
74 time series models. The second paper begins with a review on various concepts of the in-
75 tegrated series of order zero and of order one, and cointegration for a linear model as they
76 are available in the literature. It then discusses some popular nonlinear parametric models
77 beginning with those for stationary data, such as the self-exciting threshold autoregres-
78 sive models (Tong and Lim (1980), Chan (1993)) and the smooth transition autoregressive
79 models (Chan and Tong (1986), and Van Dijk et al. (2002)), then some nonlinear error
80 correction models and nonlinear cointegrating models (Teräsvirta et al. (2011), Dufrénot
81 and Mignon (2002)). Thirdly, it discusses nonparametric models with nonstationary data
82 in a similar fashion to the above parametric cases (nonparametric autoregressive then
83 nonparametric cointegrating models). The review concentrates on existing works on the
84 consistency of nonparametric estimators, with some key studies being Wang and Phillips
85 (2009a), Wang and Phillips (2009b), Karlsen and Tjøstheim (2001), Karlsen et al. (2007),
86 Karlsen et al. (2010). Finally, it presents a discussion on semiparametric models with
87 nonstationary data. The focus of the review is on semiparametric varying coefficient
88 cointegrating models and on semiparametric binary choice models.

89 The current paper complements these existing reviews by filling in some gaps, which
90 are currently left unexplored. First, it discusses various issues (e.g. identification con-
91 ditions, estimation procedures and asymptotic properties) involving semiparametric time
92 series models, which are described as tools for circumventing the curse of dimensional-
93 ity. Establishing ways to circumventing the curse of dimensionality is traditionally an
94 important objective for a large number of studies in nonparametric statistics. There are
95 essentially two approaches discussed in the literature. The first is largely concerned with
96 dimension reduction; some well-known examples of studies that fall into this category are
97 Li (1991), Cook (1998) and Xia et al. (2002). The review in the current paper focuses
98 on studies in the second category, namely function approximation using semiparametric
99 specifications. The paper first introduces three of the most well-known and successfully
100 applied semiparametric time series regression specifications in the literature, namely the

101 partially linear (PL), additive and the single-index models. It then reviews various speci-
102 fication tests for the semiparametric models in detail, including tests for linear regression.
103 Since the focus is on time series, the current paper also presents a thorough discussion on
104 re-specification of the above semiparametric models to form semiparametric autoregres-
105 sive models and their specifications test. It is argued that although these semiparametric
106 models are non-nested, they share some important similarities, especially their intolerance
107 to the endogeneity of the error term in order to obtain consistent estimates of the mod-
108 els. Addressing such a problem, in practice, involves directly estimating semiparametric
109 models with generated regressors. The current paper presents (in Section 5) a review
110 of a recent method of addressing the endogeneity problem in semiparametric time series
111 models and other semiparametric models with generated regressors, which share similar
112 characteristics. One model, in particular, has a direct application to financial econo-
113 metrics and explores a similar research area to those reviewed in Fan (2005). Finally,
114 the current paper reviews semiparametric models with nonstationary data. However, the
115 focus of this review is quite different to that of Sun and Li (2012). Since the issues of
116 semiparametric models with nonstationary data is such a large area of research, which in
117 itself warrants a separated review, the current review focuses mainly on (i) semiparamet-
118 ric models that have been established to help detect and estimate trend and seasonality,
119 and (ii) semiparametric models involved both endogeneity and nonstationarity.

120 In summary, the logic of this paper can be described metaphorically as a domino-effect
121 as follows. The first point of impact is on nonlinear time series analysis. The second,
122 third and the fourth dominoes to fall are the curse of dimensionality, the semiparametric
123 time series models and their specification testings, respectively. The fifth is the required
124 conditions shared by these popular semiparametric models, namely the exogeneity of
125 the error term and stationarity of the time series. The stationarity condition can then
126 be linked to the respecification of the semiparametric time series models to construct
127 nonlinear autoregressive time series models. Furthermore, addressing the breakdown in
128 the exogeneity condition leads to the emergence of semiparametric models with generated
129 regressors, while addressing the breakdown in the stationarity leads to semiparametric
130 models of nonstationary data.

131 The remainder of this paper is structured as follows. Section 2 discusses semipara-
132 metric models for time series, while Section 3 considers some specification tests for these
133 models. Section 4 discusses nonlinear autoregressive models and their specification test-
134 ing. Section 5 reviews the endogeneity problem in semiparametric time series models
135 and models with generated regressors. Section 6 discusses semiparametric models with
136 nonstationary data. Section 7 concludes and presents a discussion on future research.

137 **2. Semiparametric Models for Time Series**

138 This section discusses various issues involving the estimation and identification of three
 139 of the most well known and successfully applied semiparametric time series regression
 140 models in the literature, namely the PL, additive and the single-index models. Below, let
 141 us begin with the semiparametric PL time series models.

142 *Partially Linear Semiparametric Model for Time Series*

143 Since their introduction to economic literature in the 1980s by Engle et al. (1986), the PL
 144 model has attracted much attention among econometricians and applied statisticians; see
 145 Heckman (1986), Robinson (1988), Fan et al. (1995), Härdle et al. (2000) and Gao (2007)
 146 for example. In some empirical studies, the PL model is able to help avoid the impact
 147 of the curse of dimensionality by allowing *a priori* information concerning the possible
 148 linearity of some of the components to be included in the model. More specifically, the
 149 PL models look at approximating the conditional mean function $m(X_t) = m(U_t, V_t) =$
 150 $E[Y_t|U_t, V_t]$ by a semiparametric function of the form:

$$m_1(U_t, V_t) = \mu + U_t^\tau \beta + g(V_t) \tag{2.1}$$

151 such that $E[Y_t - m_1(U_t, V_t)]^2$ is minimized over a class of semiparametric functions of
 152 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
 153 is an unknown parameter, $\beta = (\beta_1, \dots, \beta_q)^\tau$ is a vector of unknown parameters, $g(\cdot)$ is
 154 an unknown function over \mathbb{R}^p , and $U_t = (U_{t1}, \dots, U_{tq})^\tau$ and $V_t = (V_{t1}, \dots, V_{tp})^\tau$ may be
 155 vectors of time series variables. Such a minimization problem is equivalent to minimizing
 156 $E[Y_t - \mu - U_t^\tau \beta - g(V_t)]^2 = E[E\{(Y_t - \mu - U_t^\tau \beta - g(V_t))^2|V_t\}]$ over some (μ, β, g) . This
 157 implies 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:

$$\beta = \Sigma^{-1} E[(U_t - E[U_t|V_t])(Y_t - E[Y_t|V_t])], \tag{2.2}$$

158 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
 159 also shows that $m_1(U_t, V_t)$ is identifiable under the assumption of $E[g(V_t)] = 0$. Some
 160 important motivations for using the functional form in (2.1) for both independent and
 161 time series data analysis can be found in Härdle et al. (2000). Based on an i.i.d. random
 162 sample, it has been shown that the parameter vector β in various versions of (2.1) can be
 163 consistently estimated at \sqrt{n} -rate, see Heckman (1986), Robinson (1988) and Fan et al.
 164 (1995), for example. For dependent processes, traditionally such a result is established
 165 under a set of somewhat more stringent conditions, e.g. the independence between $\{U_t\}$
 166 and $\{V_t\}$, as in Truong and Stone (1994). On the other hand, Fan and Li (1999b) extend

167 the \sqrt{n} -consistency and asymptotic normality results of Robinson (1988) and Fan et al.
 168 (1995) for independent observations to a strictly stationary, absolutely regular β -mixing
 169 processes under a similar set of conditions. However, these results are not applicable to a
 170 weaker condition of strong mixing processes and the case where $p > 3$.

171 Although the PL specification can reduce the dimensionality of nonparametric time
 172 series regression significantly in some cases, it is also true that the PL time series model in
 173 (2.1) may still suffer from the curse of dimensionality when $g(\cdot)$ is not necessarily additive
 174 and $p \geq 3$. A method of addressing such an issue in the literature is to establish an
 175 effective model selection procedure to ensure that both the linear and the nonparametric
 176 components of the model are of the smallest possible dimension. Gao and Tong (2004),
 177 for example, propose using a semiparametric leave n_v out cross-validation function for the
 178 choice of both the parametric and nonparametric regressors, where $n_v > 1$ is a positive
 179 integer satisfying $n_v \rightarrow \infty$ as the number observations expands to infinity. Although the
 180 details of the test can be found in the paper (see also Gao (2007)), let us note an important
 181 advantage of such a method which is the fact that it provides a general model selection
 182 procedure in determining asymptotically whether both the linear time series component
 183 and the nonparametric time series component are of the smallest possible dimension.
 184 Hence, it can help to reduce the impact of the curse of dimensionality arising from using
 185 nonparametric techniques to estimate $g(\cdot)$ in (2.1).

186 *Additive Semiparametric Model for Time Series*

187 When $g(\cdot)$ 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

$$m_1(U_t, V_t) = \mu + U_t^\tau \beta + \sum_{i=1}^p g_i(V_{ti}), \quad (2.3)$$

188 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 (2.3),
 189 where $g_i(\cdot)$ for $1 \leq i \leq p$ are all unknown one dimensional functions over \mathbb{R}^1 . The main
 190 ideas of the discussion on the semiparametric additive model above can be taken from
 191 Gao et al. (2006), who established an estimation procedure for semiparametric spatial
 192 regression. The semiparametric kernel estimation approach, as discussed in Gao et al.
 193 (2006), involves a few important steps. The first step is to estimate μ and $g(\cdot)$ by assuming
 194 that β is known. Observe that under such an assumption, we have:

$$g(x) = g(x, \beta) = E[Y_t - \mu - U_t^\tau \beta | V_t = x] = E[(Y_t - E[Y_t]) - (U_t - E[U_t])^\tau \beta | V_t = x], \quad (2.4)$$

195 using the fact that $\mu = E[Y_t] - E[U_t^\tau \beta]$, which can be estimated by the standard local linear
 196 estimation. (See, e.g. Fan and Gijbels (1996)) The second step is to apply the marginal

197 integration technique of Linton and Nielsen (1995) to obtain g_1, \dots, g_p of (2.3) based on
 198 $g(V_t) = g(V_{t1}, \dots, V_{tp}) = \sum_{l=1}^p g_l(V_{tl})$. Since $E[g_l(V_{tl})] = 0$ for $l = 1, \dots, p$, we have, for a
 199 fixed value of k , $g_k(x_k) = E[g(V_{t1}, \dots, x_k, \dots, V_{tp})]$. Therefore, this method of estimating
 200 $g(\cdot)$ is based on an additive marginal integration projection on the set of additive functions,
 201 where the projection is taken with the product measure of V_{tl} , for $l = 1, \dots, p$, unlike in the
 202 backfitting case of Nielsen and Linton (1998), and Mammen et al. (1999). Although the
 203 marginal integration technique is inferior to backfitting in asymptotic efficiency for purely
 204 additive models, it seems well suited to the framework of PL estimation; see also Fan et al.
 205 (1998), and Fan and Li (2003) for details. The third and final step involves the estimation
 206 of β using the weighted least squares estimator $\hat{\beta}$ of β derived in (2.2). The estimation
 207 procedure is completed by reintroducing $\hat{\beta}$ into the previous steps. For the independent
 208 data case, orthogonal series estimation has been used as an alternative to some other
 209 nonparametric estimation method, such as the kernel method (see Eubank (1999), for
 210 example). By approximating each $g_i(\cdot)$ using an orthogonal series $\sum_{j=1}^{n_i} f_{ij}(\cdot)\theta_{ij}$ with
 211 $\{f_{ij}(\cdot)\}$ being a sequence of orthogonal functions and $\{n_i\}$ being a sequence of positive
 212 integers, we have an approximate model of the form:

$$Y_t = \mu + U_t^\tau \beta + \sum_{i=1}^p \sum_{j=1}^{n_i} f_{ij}(V_{ti})\theta_{ij} + e_t. \quad (2.5)$$

213 Model (2.5) covers some natural parametric time series models. For example, when $U_{tl} =$
 214 U_{t-l} and $V_{ti} = Y_{t-i}$, model (2.5) becomes a parametric nonlinear additive time series
 215 model:

$$Y_t = \mu + \sum_{l=1}^q U_{t-l}\beta_l + \sum_{i=1}^p \sum_{j=1}^{n_i} f_{ij}(Y_{t-i})\theta_{ij} + e_t \quad (2.6)$$

216 The least squares estimators of (β, θ, μ) can be derived using (2.5):

$$\hat{\beta} = \hat{\beta}(n) = (\hat{U}^\tau \hat{U})^+ \hat{U}^\tau \hat{Y}, \quad \hat{\theta} = (F^\tau F)^+ F^\tau (\tilde{Y} - \tilde{U} \hat{\beta}), \quad \text{and} \quad \hat{\mu} = \bar{Y} - \bar{U}^\tau \hat{\beta}, \quad (2.7)$$

217 where

$$\begin{aligned} \theta &= (\theta_1^\tau, \dots, \theta_p^\tau)^\tau, \quad \theta_i = (\theta_{i1}, \dots, \theta_{in_i})^\tau, \\ F &= (F_1, F_2, \dots, F_p), \quad F_i = F_{in_i} = (F_i(V_{1i}), \dots, F_i(V_{ni}))^\tau, \\ \bar{U} &= \frac{1}{n} \sum_{t=1}^n U_t, \quad \tilde{U} = (U_1 - \bar{U}, \dots, U_n - \bar{U})^\tau, \quad \bar{Y} = \frac{1}{n} \sum_{t=1}^n Y_t, \quad \tilde{Y} = (Y_1 - \bar{Y}, \dots, Y_n - \bar{Y})^\tau, \\ P &= F(F^\tau F)^+ F^\tau, \quad \hat{U} = (I - P)\tilde{U}, \quad \hat{Y} = (I - P)\tilde{Y}, \quad n = (n_1, \dots, n_p)^\tau \end{aligned}$$

221 and where A^+ denotes the Moore-Penrose inverse of A . A detailed discussion about the
 222 orthogonal series estimation method is available in Chapter 2 of Gao (2007).

223 *(Extended Generalized) Single-Index Semiparametric Model*

224 Alternatively, we may approximate the conditional mean function, $m(U_t, V_t)$, using a
 225 semiparametric function of the form:

$$m_2(U_t, V_t) = U_t^\tau \theta + \psi(V_t^\tau \eta), \quad (2.8)$$

226 where θ and η are unknown vector parameters and $\psi(\cdot)$ is an unknown function. A more
 227 general model, which has recently become available in the literature, is the semiparametric
 228 single-index model:

$$m_3(X_t) = X_t^\tau \theta + \psi(X_t^\tau \eta), \quad (2.9)$$

229 where $\{X_t^\tau\}$ is a stationary and α -mixing sequence with a mixing coefficient $\alpha(k) = O(c^k)$
 230 for some large enough $0 < c < 1$. Xia et al. (1999) refer to the functional form in (2.9) as
 231 the extended generalized partially linear single-index (EG-PLSI) model.

232 **Assumption 2.1.** *Suppose that $X_t^\tau \theta + \psi(X_t^\tau \eta)$ can be written as $X_t^\tau (\theta + c\eta) + \psi(X_t^\tau \eta) -$
 233 $cX_t^\tau \theta$ such that all the roots of the function $x^d - (\theta_1 + c\eta_1)x^{d-1} - \dots - (\theta_d + c\eta_d)$ are inside
 234 the unit circle. Moreover, suppose that $\lim_{|u| \rightarrow \infty} |\psi(u)/u| = 0$.*

235 In this case, the geometrical ergodicity of $\{Y_t\}$ is ensured under the conditions stated
 236 in Assumption 2.1. (See Theorem 3 of Xia et al. (1999) for details.) Furthermore, in
 237 order to ensure the estimatability of the model, it must also be the case that θ and η are
 238 perpendicular to each other with $\|\eta\| = 1$ and the first nonzero element must be positive.
 239 Now, let us define the following:

$$S(\theta, \eta) = E[Y_t - \varphi_\eta(X_t^\tau \eta) - \{X_t - \Gamma_\eta(X_t^\tau \eta)\}^\tau \theta]^2, \quad (2.10)$$

240 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 $\mathcal{W}(\eta) = E[\{X -$
 241 $\Gamma_\eta(X_t^\tau \eta)\} \{X - \Gamma_\eta(X_t^\tau \eta)\}^\tau]$ and $\mathcal{V}(\eta) = E[\{X - \Gamma_\eta(X_t^\tau \eta)\} \{Y_t - \varphi_\eta(X_t^\tau \eta)\}]$. Xia et al. (1999)
 242 show that the minimum point of $S(\theta, \eta)$ with $\theta \perp \eta$ is unique at η and $\theta = \{\mathcal{W}(\eta)\}^+ \mathcal{V}(\eta)$,
 243 where $\{\mathcal{W}(\eta)\}^+$ is the Moore–Penrose inverse.

244 To estimate the model, Xia et al. (1999) introduce an estimation procedure, which is
 245 a semiparametric extension of the one introduced in Härdle et al. (1993) for a nonpara-
 246 metric single-index model. The procedure consists of four important steps as follows: (i)
 247 Compute the estimate $\hat{\theta}_\eta$ of θ given η and the delete-one estimators of φ_η and Γ_η . Let Θ
 248 denote all the unit vectors in \mathbb{R}^p . (ii) Estimate $\eta \in \Theta$ and the bandwidth, h , using those
 249 values $\hat{\eta}$ and \hat{h} that minimize $\hat{S}(\eta, h)$ (an estimate of $S(\theta, \eta)$), where θ is replaced by $\hat{\theta}_\eta$,
 250 and φ_η and $\Gamma_\eta(\cdot)$ are replaced by their nonparametric estimators; (iii) Re-estimate θ as in

251 the first step, but with η being replaced by $\hat{\eta}$; (iv) Estimate $\psi(\cdot)$ using the nonparametric
 252 kernel estimates and the fact that $\psi(x) = \varphi(x) - \theta\Gamma(x)$.

253 In order to illustrate the statistical validity of such an estimation procedure, the
 254 following asymptotic results are established:

255 (i) \sqrt{n} -consistency

$$\tilde{n} \left(\hat{\theta} - \theta \right) \rightarrow N(0, \mathbb{C}^+) \quad \text{and} \quad \tilde{n} (\hat{\eta} - \eta) \rightarrow N(0, \mathbb{D}^+) \quad (2.11)$$

256 in distribution, where \tilde{n} is the number of elements in $\mathcal{A} \subset \mathbb{R}$, i.e. the union of a number
 257 of open convex sets such that $f(x) > M$ for some constant $M > 0$, and \mathbb{C}^+ and \mathbb{D}^+ are
 258 some positive, finite constants; see also the corollary in page 836 of Xia et al. (1999);

259 (ii) Uniform convergence, where, almost surely:

$$\sup_{v \in \{x^\tau \eta : x \in \mathcal{A}\}} \left| \hat{\psi}_{\hat{\eta}}(v) - \psi(v) \right| = O \left\{ (n^{-4/5} \log n)^{1/2} \right\}. \quad (2.12)$$

260 See also Theorem 5 of Xia et al. (1999).

261 These results warrant a few remarks. The asymptotic normality is a direct extension
 262 of the one presented in Härdle et al. (1993), but under α -mixing and a larger parameter
 263 cone Ω_n such that $\Omega_n = \{\eta : \|\eta - \eta\| \leq Mn^{-\delta}\}$ for some constant M , where $\frac{3}{10} < \delta < \frac{1}{2}$.
 264 The proof of such results is made possible using a decomposition of \hat{S} into various parts
 265 as follows.

266 Xia et al. (1999) derive the decomposition of $\hat{S}(\eta, h)$ into a few important terms (see
 267 Theorem 4 of the paper). While one of these is shown to be $o(1)$, the remaining are:

$$\tilde{S}(\eta) = \sum_{X_t \in \mathcal{A}} \{y_t - X_t^\tau \theta_\eta - \psi(X_t^\tau \eta)\}^2 \quad \text{and} \quad T(h) = \sum_{X_t \in \mathcal{A}} \{\hat{\psi}_\eta(X_t^\tau \eta) - \psi(X_t^\tau \eta)\}^2, \quad (2.13)$$

268 where $\{y_t\}$ is a stationary and α -mixing process. Such a result suggests that estimating
 269 the EG-PLSI model can also be done in iterative steps, such as: (i) estimating h given an
 270 initial estimator of η , e.g. $\check{\eta}$; (ii) update $\check{\eta}$ using \check{h} from the previous step; (iii) repeat the
 271 first two steps. In this setting, since it is clear that Step (i) is simply an estimation of a
 272 PL model for time series, such a result highlights a close connection between the PL and
 273 the EG-PLSI models.

274 Finally, some basic modifications to the formulation of the models bring about various
 275 special cases, which are well-known in the literature. For instance, if $\theta = 0$, (2.9) reduces
 276 to:

$$m_4(X_t) = \psi(X_t^\tau \eta), \quad (2.14)$$

277 which is the single index model discussed in Härdle et al. (1993). Furthermore, by par-
 278 titioning $X_t = (U_t^\tau, V_t^\tau)^\tau$ and by taking $\theta = (\beta^\tau, 0, \dots, 0)^\tau$ and $\eta = (0, \dots, 0, \alpha^\tau)^\tau$, the

279 EG-PLSI model becomes the generalized partially linear single-index (G-PLSI) model
 280 introduced by Carroll et al. (1997) of the form

$$m_5(X_t) = U_t^T \theta + \psi(V_t^T \alpha), \quad (2.15)$$

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

282 In the next section, let us review some useful procedures for testing the semiparametric
 283 specifications of these time series models.

284 3. Some Specification Tests for Semiparametric Models

285 In this section, we focus first on tests for a semiparametric (either PL or single-index)
 286 form against a nonparametric form. We then introduce the corresponding PL model to
 287 the EG-PLSI model in order to, finally, discuss testing a linear regression model against
 288 a semiparametric model.

289 *Specification Tests for Semiparametric vs. Nonparametric Form*

290 Most nonlinear time series specification tests, discussed in the literature, concentrate
 291 mainly on testing either:

$$\text{Nonparametric model: } y_t = m(X_t) + e_t \text{ or Single-index model: } y_t = \psi(X_t^T \eta) + e_t, \quad (3.1)$$

292 where $\{X_t\}$ is a sequence of strictly stationary time series variables. As we will discuss
 293 below, these tests can be conveniently adopted to hypothesis testing in the semiparametric
 294 time series models discussed above.

295 Based on the nonparametric model in (3.1), Gao and Gijbels (2008) discuss a non-
 296 parametric testing procedure to test hypotheses of the form:

$$H_{01} : m(x) = m_{\theta_0}(x) \text{ versus } H_{11} : m(x) = m_{\theta_1}(x) + C_n \Delta_n(x) \text{ for all } x \in \mathbb{R}^d,$$

297 where both θ_0 and $\theta_1 \in \Theta$ are unknown parameters, Θ is a parameter space of \mathbb{R}^d , C_n
 298 is a sequence of real numbers and $\Delta_n(x)$ is a sequence of nonparametrically unknown
 299 functions over \mathbb{R}^d , such that model (3.1) becomes a semiparametric time series model of
 300 the form:

$$y_t = m_{\theta_0}(X_t) + e_t \quad (3.2)$$

301 under H_{01} . Gao and Gijbels (2008) assume that $\{X_t\}$ is strictly stationary and α -mixing,
 302 with the mixing coefficient defined by

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

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

305 Prior to Gao and Gijbels (2008), Härdle and Mammen (1993) suggest that one way of
 306 establishing the nonparametric kernel test statistic for such hypothesis is to do so based
 307 on the L_2 -distance function:

$$M_{1n}(h) = nh^{\frac{d}{2}} \int \{\widehat{m}_h(x) - \widetilde{m}_{\widehat{\theta}}(x)\}^2 w(x) dx, \quad (3.3)$$

308 where $w(x)$ is some non-negative weight function, $\widehat{m}_h(x)$ is the nonparametric kernel
 309 estimator of $m(\cdot)$ defined by:

$$\widehat{m}_h(x) = \frac{\sum_{t=1}^n K_h(x - X_t) y_t}{\sum_{t=1}^n K_h(x - X_t)} \quad (3.4)$$

310 and $\widetilde{m}_{\widehat{\theta}}(x)$ is its parametric counterpart:

$$\widetilde{m}_{\widehat{\theta}}(x) = \frac{\sum_{t=1}^n K_h(x - X_t) m_{\widehat{\theta}}(X_t)}{\sum_{t=1}^n K_h(x - X_t)}, \quad (3.5)$$

311 where $\widehat{\theta}$ is a \sqrt{n} -consistent estimator of θ_0 . Recently, a number of studies derived the
 312 nonparametric test statistics based on a modified version of the L_2 -distance function in
 313 (3.3). An example is the work by Horowitz and Spokoiny (2001), who used a discrete
 314 approximation to $M_{1n}(h)$ of the form

$$M_{2n}(h) = \sum_{t=1}^n (\widehat{m}_h(X_t) - \widetilde{m}_{\widehat{\theta}}(X_t))^2, \quad (3.6)$$

315 where $\{X_t\}$ is only a sequence of fixed designs. They also considered a multiscale nor-
 316 malized version of the form:

$$M_{2n} = \max_{h \in H_n} \frac{M_{2n}(h) - \widehat{M}_n(h)}{\widehat{V}_n(h)}, \quad (3.7)$$

317 where H_n is a set of suitable bandwidth:

$$\widehat{M}_n(h) = \sum_{t=1}^n \left(\sum_{s=1}^n W_h(X_s, X_t) \right) \widehat{\sigma}_n^2(X_t)$$

318 and:

$$\widehat{V}_n^2(h) = 2 \sum_{s=1}^n \sum_{t=1}^n \left(\sum_{\ell=1}^n W_h(X_\ell, X_t) \right)^2 \widehat{\sigma}_n^2(X_s) \widehat{\sigma}_n^2(X_t),$$

319 where $W_h(\cdot, X_t) = \frac{K_h(\cdot - X_t)}{\sum_{u=1}^n K_h(\cdot - S_u)}$ and $\hat{\sigma}_n^2(X_s)$ is a consistent estimator of the variance
 320 function $\sigma_n^2(X_t) = E[e_t^2]$. They then show that M_{2n} is asymptotically consistent with an
 321 optimal rate of convergence for hypothesis testing.

322 An alternative approach employed by Gao and Gijbels (2008) is to consider a different
 323 type of distance function for the nonparametric kernel test statistic. In order to discuss
 324 this method, let us first rewrite the nonparametric model into a notational version so
 325 that, under the H_0 , we have:

$$Y = m_{\theta_0}(X) + e, \quad (3.8)$$

326 where X is assumed to be random, θ_0 is the true value of θ under H_0 and $E[e|X] = 0$. In
 327 this case, the distance function employed can be written as follows:

$$E[eE(e|X)\pi(X)] = E[(E^2(e|X))\pi(X)], \quad (3.9)$$

328 where $\pi(\cdot)$ is the marginal density function of X . In order to establish the asymptotic
 329 distribution of their test statistic, Gao and Gijbels (2008) suggested studying asymptotic
 330 distribution and proposing an Edgeworth expansion for the quadratic form of the following
 331 type:

$$R_n(h) = \sum_{s=1}^n \sum_{t=1}^n e_s \phi_n(X_s, X_t) e_t, \quad (3.10)$$

332 where $\phi_n(\cdot, \cdot)$ may depend on n , the bandwidth h and the kernel function K . This is
 333 because, to derive the test statistic, they are able to use a normalized kernel-based sample
 334 analogue of (3.9) of the form

$$L_{1n}(h) = \frac{h^{\frac{d}{2}}}{n} \sum_{s=1}^n \sum_{t=1}^n \hat{e}_s K_h(X_t - X_s) \hat{e}_t, \quad (3.11)$$

335 where $\hat{e}_t = y_t - m_{\hat{\theta}}(X_t)$, which turns out to be simply the leading term of the quadratic
 336 form in (3.10). In order to proceed, let us now define the following:

$$\hat{L}_{1n}(h) = \frac{L_{1n}(h) - E[L_{1n}(h)]}{\sqrt{\text{var}[L_{1n}(h)]}}. \quad (3.12)$$

337 For each given h , we may also define a stochastically normalized version of the form

$$\bar{L}_{1n}(h) = \frac{\sum_{s=1}^n \sum_{t=1, \neq s}^n \hat{e}_s K_h(X_s - X_t) \hat{e}_t}{\sqrt{2 \sum_{s=1}^n \sum_{t=1}^n \hat{e}_s^2 K_h(X_s - X_t) \hat{e}_t^2}}. \quad (3.13)$$

338 Furthermore, it has been shown in Gao (2007), and Gao and Gijbels (2008) that we have:

$$\bar{L}_{1n}(h) = L_{1n}(h) + o_P(1) \quad (3.14)$$

339 for each given h . Hence, we may use the distribution of $\bar{L}_n(h)$ to approximate that of
 340 $\hat{L}_n(h)$. Since the main objective of the research in Gao and Gijbels (2008) is to propose a
 341 suitable selection criterion for the choice of h (such that while the size function is appro-
 342 priately controlled, the power function is maximized at this h), they also give Edgeworth
 343 expansions of both the size and power functions of the test. Nonetheless, instead of dis-
 344 cussing these in detail here, we suggest that interested readers should consult Section 3
 345 of Gao and Gijbels (2008).

346 In this review, let us proceed with hypothesis testing of the semiparametric time
 347 series specifications. We will begin with the corresponding hypothesis testing for the PL
 348 regression:

$$H_{02} : m(x) = u^\tau \beta + g(v) \quad \text{versus} \quad H_{12} : m(x) = u^\tau \beta + g(v) + C_n \Delta_n(x) \quad \text{for all } x \in \mathbb{R}^d,$$

349 where C_n and $\Delta_n(\cdot)$ are as defined previously, and u and v are subvectors of $x = (u^\tau, v^\tau)^\tau$.
 350 In this case, the test statistic can be written as:

$$L_{2n}(h) = \sum_{s=1}^n \sum_{t=1}^n \hat{y}_s K \left(\frac{X_s - X_t}{h} \right) \hat{y}_t, \quad (3.15)$$

351 where $\hat{y}_s = y_s - U_s^\tau \hat{\beta} - \hat{g}(V_s)$, $\hat{\beta} = (\tilde{U}^\tau \tilde{U})^{-1} \tilde{U}^\tau \tilde{y}$, $\hat{g}(V_s) = \sum_{t=1}^n w_{2st} (y_t - U_t^\tau \hat{\beta})$, $\tilde{U} =$
 352 $(I - W_2)U$, $U = (U_1, \dots, U_n)^\tau$, $\tilde{y} = (I - W_2)Y$ and $W_2 = \{w_{2st}\}$ is a $n \times n$ matrix such
 353 that $w_{2st} = \frac{K_2\left(\frac{V_s - V_t}{h}\right)}{\sum_{u=1}^n K_2\left(\frac{V_s - V_u}{h}\right)}$ with $K_2(\cdot)$ being a kernel function. Some existing results for
 354 a similar test statistic to $L_{2n}(h)$ as defined in (3.15) can be found in, for example, Fan
 355 and Li (1996) and Fan and Li (1997) (see also the detailed review of Fan and Li (1997)
 356 below).

357 Fan and Li (1996) consider a consistent test for a PL model where $\{U_t^\tau, V_t^\tau\}_{t=1}^n$ is
 358 a set of n i.i.d. observations on $\{U^\tau, V^\tau\}^\tau$ with U being $p \times 1$ and V being the $q \times 1$
 359 regressors. Nonetheless, there are two useful results in the literature that may enable
 360 an extension of Fan and Li (1996) procedure to hypothesis testing in time series data,
 361 namely the Central Limit Theorem (CLT) established in Fan and Li (1999a) and the
 362 \sqrt{n} -consistent estimation of partially linear time series models in Fan and Li (1999b).
 363 Together, these two results can be used in the generalization of the consistent test of Fan
 364 and Li (1996) for testing a PL model versus a nonparametric regression model in the time
 365 series framework. This work is done by Li (1999). An important issue that should be
 366 noted is the dependence structure assumed. Fan and Li (1999a), Fan and Li (1999b), and
 367 Li (1999) considered absolutely regular (β -mixing) processes (though it is well known that
 368 such absolute regularity is stronger than strong mixing). This is because their method
 369 of mathematical proof relied on an inequality for β -mixing processes due to Yoshihara

370 (1976), which was not available for α -mixing. However, there are other recent works that
 371 studied PL models of α -mixing processes, such as Gao and Yee (2000), and Härdle et al.
 372 (2000).

373 With regard to the semiparametric single-index model, as a natural extension to the
 374 above tests, we may consider testing

$$H_{03} : m(x) = u^\tau \beta + \psi(x^\tau \eta) \text{ versus } H_{13} : m(x) = u^\tau \beta + \psi(x^\tau \eta) + C_n \Delta_n(x) \text{ for all } x \in R^d,$$

375 where both θ and η are vectors of unknown parameters, and $\psi(\cdot)$ is an unknown function.
 376 In this case, the test statistic can be written, similar to (3.15), as:

$$L_{3n}(h) = \sum_{s=1}^n \sum_{t=1}^n \tilde{y}_s K \left(\frac{(X_s - X_t)^\tau \hat{\eta}}{h} \right) \tilde{y}_t, \quad (3.16)$$

377 where $\hat{\theta}$, $\hat{\eta}$ and $\hat{\psi}(\cdot)$ are consistent estimators discussed previously, and we have:

$$\tilde{y}_t = \left(y_t - U_t^\tau \hat{\beta} - \hat{\psi}(X_t^\tau \hat{\eta}) \right) \hat{f}_3(X_t^\tau \hat{\eta}),$$

378 in which $\hat{f}_3(X_t^\tau \hat{\eta}) = \frac{1}{h} \sum_{t=1}^n K \left(\frac{(X_s - X_t)^\tau \hat{\eta}}{h} \right)$.

379 *Specification Test for Linear Regression vs. a Semiparametric Form*

380 For the case where $\{X_t\}$ is a vector time series regressor and $g(\cdot)$ is an unknown function
 381 defined on \mathbb{R}^p (where $1 \leq p \leq 3$), an attempt is made in the work of Gao (2012) to extend
 382 the semiparametric PL models in (2.1) to the semi-linear (SL) model of the form

$$m_6(X_t) = \mu + X_t^\tau \beta + g(X_t), \quad (3.17)$$

383 which is a direct counterpart of the EG-PLSI model in (2.9), where the SL model has
 384 different motivations and applications from the conventional semiparametric time series
 385 model presented in (2.1) as follows: (i) In (3.17), the linear component in many cases
 386 plays the leading role, while the nonparametric component behaves like a type of unknown
 387 departure from such classic linear model. In order to establish the empirical support for
 388 such a condition, Gao (2012) uses the SL model to investigate time series properties of
 389 quarterly consumer price index numbers of 11 classes of commodities for eight Australian
 390 capital cities between 1994 and 2008. The author has found that linearity remains the
 391 leading component of the trending component of the consumer price index data. (ii) The
 392 SL model can be motivated as a model to address some endogenous problems involved
 393 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
 394 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

395 an i.i.d. error. Unfortunately, in the process of estimating β and $g(\cdot)$, existing methods
 396 are not directly applicable, especially given the fact that $\Sigma = (E[U_t - E[U_t|U_t]](E[U_t -$
 397 $E[U_t|U_t]]^\tau)) = 0$. To this end, Gao (2012) studies the estimation of the SL model and
 398 its asymptotic properties in two different contexts, namely (i) where $\{X_t\}$ is a vector of
 399 stationary time series regressors; (ii) where $\{X_t\}$ is stochastically nonstationary.

400 In the following, we focus first on the case of stationary time series regressors, while
 401 the case of nonstationary regressors will be considered later. In this case, essential as-
 402 sumptions are the identifiability and the smallness conditions of $g(\cdot)$.

403 **Assumption 3.1.** (*Assumption 2.1(i) of Gao (2012)*) Let $g(\cdot)$ be an integrable function

$$\int \|x\|^i |g(x)|^i dF(x) \leq \infty$$

404 for $i = 1, 2$ and $\int xg(x)dF(x) = 0$, where $F(x)$ is the cumulative distribution function of
 405 $\{X_t\}$ and $\|\cdot\|$ denotes the conventional Euclidean norm.

406 Under such conditions, the parameter β is identifiable and chosen such that $E[Y_t -$
 407 $X_t^\tau \beta]^2$ is minimized over β , which implies $\beta = (E[X_1 X_1^\tau])^{-1} E[X_1 Y_1]$, provided that the
 408 inverse matrix exists. Such a definition of β suggests that $\int xg(x)dF(x) = 0$, so β can be
 409 estimated by the ordinary least squares estimator of the form:

$$\hat{\beta} = \left(\sum_{t=1}^n X_t X_t^\tau \right)^{-1} \left(\sum_{t=1}^n X_t Y_t \right) \quad \text{such that} \quad \hat{g}(x) = \sum_{t=1}^n w_{nt}(x) \left(Y_t - X_t^\tau \hat{\beta} \right), \quad (3.18)$$

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

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

$$\hat{L}_{1n} = \frac{\sqrt{n}}{\hat{\sigma}_1} \left(\frac{1}{n} \sum_{t=1}^n \left(Y_t - X_t^\tau \hat{\beta} \right)^2 - \hat{\sigma}_0^2 \right) \xrightarrow{D} N(0, 1), \quad (3.19)$$

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

419 Finally, let us note that the EG-PLSI model discussed in the previous section can
 420 always be used for the case where $p \geq 4$. For the sake of convenience and clarity, we will
 421 leave the discussion on the case of a nonstationary time series to a later section.

422 In the next section, let us shift our attention to a number of nonlinear autoregressive
423 models that can be derived based on the semiparametric models defined above and their
424 specification testing in practice.

425 4. Nonlinear Autoregressive Models and Their Specification Testing

426 Since the focus of the current review is on time series, it is important that we also
427 discuss the re-specification of the semiparametric models in the previous section to form
428 semiparametric autoregressive models and their specification testing.

429 If the observations are allowed to be taken over time, then the above mentioned semi-
430 parametric models give rise to a number of well-known nonlinear autoregressive models
431 discussed in the literature as follows:

432 (i) A similar partitioning of X_t to that of (2.1) such that $U_t = (Y_{t-c_1}, Y_{t-c_2}, \dots, Y_{t-c_p})^\tau$
433 and $V_t = (Y_{t-d_1}, Y_{t-d_2}, \dots, Y_{t-d_p})^\tau$, where $c_i \neq d_j$ for all $1 \leq i \leq p$ and $1 \leq j \leq q$, giving
434 rises to the autoregressive semiparametric PL additive model discussed in Gao and Yee
435 (2000). Gao and Yee (2000) found that the PL regression is more appropriate than a
436 completely nonparametric autoregression for the Canadian lynx data, which comprises of
437 the annual record of the number of lynx trapped in the MacKenzie River district in the
438 Canadian Northwest Territories from 1821 to 1934.

439 (ii) The autoregressive single-index model discussed in Xia et al. (1999) is obtained
440 simply by letting $X_t = (Y_{t-1}, Y_{t-2}, \dots, Y_{t-p})^\tau$ in (2.9). Using the projection pursuit
441 method to investigate the autoregressive process of sunspot numbers in a year, Xia et al.
442 (1999) found some strong empirical evidence in support of such a model. Furthermore, a
443 specification test of linearity can be developed based on the fact that statistical insignif-
444 icance of the nonlinear component signals the superiority of a linear model. To see this,
445 let us write an autoregressive EG-PLSI model in the form:

$$y_t = \beta^\tau X_t + \phi(\eta^\tau X_t) + \varepsilon_t, \quad (4.1)$$

446 where $X_t = (y_{t-1}, y_{t-2}, y_{t-3})^\tau$. Such a specification can be tested against a linear regressive
447 model through testing $H_0 : \phi(u) \equiv 0$. Xia et al. (1999) suggested that the testing procedure
448 can be developed based on the method discussed in Xia (1998). In their empirical analysis
449 of the shape-invariant Engel curves in Australia, Kim et al. (2013) follow this suggestion
450 and construct the Bonferroni-type variability bands in order to determine the statistical
451 significance of, for example, $\phi(\cdot)$.

452 (iii) Another useful alternative is to establish an autoregressive SL model. In this case,
453 the process $\{Y_t\}$ is stochastically stationary and α -mixing under the following conditions:

454 **Assumption 4.1.** (*Assumption 4.1 of Gao (2012)*) (i) $\beta = (\beta_1, \dots, \beta_p)^\tau$ satisfy $Y^p -$
 455 $\beta_1 Y^{p-1} - \dots - \beta_{p-1} Y - \beta_p \neq 0$ for any $|Y| \geq 1$; (ii) $g(X)$ is bounded on any bounded
 456 Borel measurable set and satisfy $g(X) = o(\|X\|)$ as $\|X\| \rightarrow \infty$, where $\|\cdot\|$ denotes the
 457 conventional Euclidean norm.

458 In this case, the test statistic described in (3.19) above can be used to test a linear autore-
 459 gressive model against a semiparametric alternative. Clearly, this is the corresponding
 460 test to that in Xia et al. (1999) above.

461 In order to provide a brief background and introduction into issues surrounding the
 462 specification testing of the autoregressive semiparametric models, let us first consider the
 463 following general autoregressive model of a finite order p :

$$Y_t = g(Y_{t-1}, \dots, Y_{t-p}) + \epsilon_t, \quad (4.2)$$

464 where the autoregressive function g is unknown and $\{\epsilon_t\}$ is a sequence of martingale
 465 differences. The process $\{Y_t\}$ is absolutely regular with a coefficient $\phi_\tau = O(\rho^\tau)$, where ρ
 466 is a constant $0 < \rho < 1$. One of the first natural steps in the analysis of time series is to
 467 decide whether to use a nonlinear model. For convenience, we let $X_t = (Y_{t-1}, \dots, Y_{t-p})^\tau$
 468 so that we observe X_1, \dots, X_{n+1} . To this end, Fan and Li (1997) establish a consistent
 469 nonparametric test for the linearity of AR(p) models. In terms of X_t , the hypotheses can
 470 be written as:

$$H_{03} : P(g(X_t) = \alpha^\tau X_t) = 1 \text{ and } H_{13} : P(g(X_t) = \alpha^\tau X_t) < 1 \quad (4.3)$$

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

$$I_n = \frac{1}{n(n-1)h^p} \sum_{s \neq t} \hat{\epsilon}_t \hat{\epsilon}_s K_{st}, \quad (4.4)$$

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

$$I_{n1} = \frac{1}{n(n-1)h^p} \sum_{s \neq t} \epsilon_t \epsilon_s K_{st}. \quad (4.5)$$

479 To this end, Fan and Li (1997) derive the asymptotic normality of $nh^{p/2}I_{n1}$ by invoking on
 480 the CLT for degenerate U-statistics of absolutely regular processes of Khashimov (1993).
 481 In addition, Fan and Li (1999a), focus on one of the conditions in Khashimov (1993)
 482 which requires the error term ϵ_t to bounded and to provide a new CLT that can be used
 483 to relax such a boundedness. (Note that in the model specification testing introduced in
 484 Fan and Li (1999a) the error term is defined instead as $Y_t - g(X_t, \gamma)$ to reflect the null
 485 hypothesis which involves a specific parametric family.)

486 In the previous sections, we have noted results in the literature which suggested a
 487 close connection between the semiparametric models reviewed above. Another important
 488 feature that these models share is their intolerance of the endogeneity of the error term
 489 in order to obtain consistent estimates of the models. In the following section, we review
 490 the literature on the endogeneity problem in semiparametric models and a few methods
 491 of dealing with it. It will soon be clear that these can be directly related to studies of
 492 model estimation with generated regressors.

493 5. Endogeneity and Semiparametric Models with Generated Regressors

494 As noted previously, consistent estimation of the above mentioned PL and EG-PLSI
 495 models for time series relies on the exogeneity of the error term with respect to both the
 496 parametric and nonparametric regressors. The breakdown of such a condition is famously
 497 known in the literature as the endogeneity problem (see Blundell and Powell (2003), for
 498 example). Let e_t (for $t = 1, 2, \dots, n$) form a sequence of i.i.d. random errors with a mean
 499 of zero and a finite variance of σ^2 , so that the PL model for time series can be written as:

$$y_t = \mu + U_t^\tau \beta + g(V_t) + e_t. \quad (5.1)$$

500 An important assumption, which is required to ensure the consistent estimation of β
 501 and $g(\cdot)$, is the exogeneity of the error term with respect to both the parametric and
 502 nonparametric regressors, mathematically described as $E[e|U = u] = 0$ and $E[e|V = v] =$
 503 0. Such an exogeneity condition is also needed for the EG-PLSI model:

$$y_t = X_t^\tau \theta + \psi(X_t^\tau \eta) + e_t, \quad (5.2)$$

504 where, in this case, it is necessary that $E[e|X^\tau \eta = v] = 0$. Kim and Saart (2013), and
 505 Kim et al. (2013) discuss in detail a set of simulation exercises to illustrate the seriousness
 506 of the impacts of endogeneity problem in semiparametric regression models.

507 While Kim and Saart (2013) attempted to address the endogeneity problem in the
 508 PL model, Kim et al. (2013) did so for the EG-PLSI model. In principle, the methods

509 considered in Kim and Saart (2013) closely followed the logic of Robinson's (1988) two-step
510 estimation procedure mentioned previously, i.e. first obtaining consistent estimators of
511 the unknown parameters and then using them in order to identify an unknown structural
512 function. If the parametric regressors are exogenous, then the least-squares estimators of
513 the parametric parameters are consistent. Otherwise, if parametric endogeneity is present,
514 then the parametric instrumental variable (PIV) estimation can be used. The consistency
515 of the parametric estimators is important not only in its own right but also for identifying
516 an unknown nonlinear function, $g(\cdot)$.

517 The presence of nonparametric endogeneity can induce further complication in the
518 identification of the unknown function. There are two alternative methods in the literature
519 which may be helpful in identifying the unknown function in such a case, namely the
520 nonparametric instrumental variable (NpIV) estimation and the control function (CF)
521 approach. Newey and Powell (2003), Hall and Horowitz (2005), and Darolles et al. (2011)
522 developed the NpIV estimation for a pure nonparametric model, while Ai and Chen (2003)
523 did so for semiparametric models, which included the PL model as a special case. One
524 of the difficulties with using NpIV estimation resides in the well-known ill-posed inverse
525 problem; see O'Sullivan (1986), for example. To overcome such an obstacle, Ai and
526 Chen (2003) based their estimation on a complex sieve estimation under some regularity
527 conditions on the inversion matrix and a constraint on the space of the reduced relation
528 to keep it compact. On the other hand, Newey et al. (1999) and Pinkse (2000) considered
529 the CF approach in a pure nonparametric model, while Blundell and Powell (2004) did
530 so for a special case of a single index model, i.e. a case where only the discrete dependent
531 variable was considered. With regard to the nonparametric estimation employed, Newey
532 et al. (1999) and Pinkse (2000) relied on series approximation, while Su and Ullah (2008)
533 used the local polynomial estimation of Fan and Gijbels (1996). Blundell and Powell
534 (2004), on the other hand, relied on the local constant kernel estimation method.

535 Kim and Saart (2013) addressed nonparametric endogeneity in the estimation and
536 inference of the PL model in a simple but widely-used framework of nonparametric si-
537 multaneous equations, specifically a nonparametric triangular model. Although the full
538 details can be found in the paper, let us discuss this briefly here. They considered the
539 following model:

$$y = x'\beta + g(v) + \epsilon, \quad (5.3)$$

540 where x may be either exogenous or endogenous, while v is endogenous. In addition, the
541 following nonparametric reduced-form equation exists:

$$v = m_v(z) + \eta, \quad (5.4)$$

542 where z is a vector of the instrumental variables such that $E(\eta|z) = 0$ and $E(\epsilon|z, \eta) =$
 543 $E(\epsilon|\eta) \neq 0$. In order to identify and to estimate the structural function $g(\cdot)$, they take
 544 the CF approach, as in Newey et al. (1999), namely

$$E(y|v, \eta) = E(x|v, \eta)' \beta + g(v) + \iota(\eta), \quad (5.5)$$

545 where the endogeneity (i.e. $E(\epsilon|\eta) = \iota(\eta) \neq 0$) is controlled by introducing an additional
 546 unknown function. This structure enabled Kim and Saart (2013) to write the model as
 547 a simple nonparametric additive structure and, therefore, to employ the local constant
 548 kernel estimation and the marginal integration technique of Linton and Nielsen (1995),
 549 and Tjøstheim and Austad (1996) to identify the unknown function. As discussed in Kim
 550 et al. (2013), this procedure can also be used to address an endogeneity problem in the
 551 EG-PLSI model for the time series of Xia et al. (1999).

552 Nonetheless, this estimation procedure involves a generated regressor in the sense
 553 that an estimate of η must be used in estimating the conditional expectation in (5.5).
 554 In fact, there are many nonparametric and semiparametric models in econometrics that
 555 contain generated regressors. For example, Lewbel and Linton (2007) dealt with non-
 556 parametrically generated regressors when considering homothetically separable functions.
 557 Moreover, Newey et al. (1999) and Su and Ullah (2008) studied the nonparametric estima-
 558 tion of triangular simultaneous equation models. Li and Wooldridge (2002) considered the
 559 semiparametric estimation of PL models for dependent data with generated regressors. In
 560 a sense, Li and Wooldridge's (2002) model can be seen as a special case of the regression
 561 model in (5.5). Let $\mathcal{W}_t = \{Y_t, U_t^\tau, S_t, Z_t^\tau\}$ be a stationary and absolutely regular process,
 562 i.e., as $\tau \rightarrow \infty$:

$$\beta_\tau = \sup_{s \in \mathcal{N}} E \left[\sup_{A \in \mathcal{M}_{s+\tau}^\infty} \{|P(A|\mathcal{M}_{-\infty}^s(\mathcal{W})) - P(A)|\} \right] \rightarrow 0, \quad (5.6)$$

563 where $\mathcal{M}_s^t(\mathcal{W})$ denotes $\sigma(\mathcal{W}_s, \dots, \mathcal{W}_t)$, the sigma algebra generated by $(\mathcal{W}_s, \dots, \mathcal{W}_t)$, for
 564 $s \leq t$. Li and Wooldridge's (2002) model can be written as:

$$Y_t = U_t^\tau \beta + g(\eta_t) + \varepsilon_t \quad (5.7)$$

$$\eta_t = S_t - Z_t^\tau \alpha \quad (5.8)$$

565 such that $E(\varepsilon_t|U_t, Z_t, \eta_t) = 0$ and $E(\eta_t|Z_t) = 0$, where U_t is $p \times 1$, Z_t is $q \times 1$, Y_t and S_t are
 566 scalars, β and α are the vectors of unknown parameters, and $g(\cdot)$ is an unknown smooth
 567 function. The model can be modified so that nested within it are a nonlinear regression
 568 model and a Tobit-3 model. These modifications have been found to be very useful in
 569 practice. Bachmeier (2002), for example, applies a modified version of Li and Wooldridge's

570 (2002) model, which is written in the form of a semiparametric error correction model, to
 571 investigate nonlinearity in the term structure; see also Galego and Pereira (2010) for an
 572 application of the model to labour economics.

573 Overall, the model's estimation procedure is similar to that introduced in Robinson
 574 (1988), which we discussed earlier. The only exception in this case is the fact that the
 575 parametric estimation of η , as defined in (5.8), is now required in the first step. Hence,
 576 the mathematical proof of the \sqrt{n} -consistency of the unknown parameters must rely
 577 on an assumption that a \sqrt{n} -consistent estimator of α exists. Compared to those in
 578 Robinson (1988), Li and Wooldridge (2002) have to impose slightly stronger moment and
 579 smoothness conditions on the regression, density and kernel functions. This is mainly
 580 because they have to use Taylor expansions in their proof to deal with the regressor η_t ,
 581 which was initially generated parametrically.

582 A similar generated regressor problem was also encountered by Saart et al. (2013)
 583 in order to develop their so-called semiparametric autoregressive conditional duration
 584 (SEMI-ACD) model. This is with an exception to the fact that, in their study, the
 585 unobservable regressor is computed semiparametrically based on an iterative estimation
 586 algorithm instead of using a linear regression as stated in (5.8). Saart et al. (2013) first
 587 derived the uniform consistency of the estimation algorithm, then used the Taylor expan-
 588 sions (together with the uniform convergence rates for kernel estimation with dependent
 589 data derived in Hansen (2008)) in the proof to deal with the generated regressor. Below,
 590 let us discuss the SEMI-ACD model in more detail. Let Y_t denotes financial duration, i.e.
 591 the waiting time between two consecutive financial events, associated with the t -th event.
 592 Engle and Russell (1998) develop the ACD model by assuming that

$$Y_t = \psi_t \varepsilon_t, \tag{5.9}$$

593 where $\{\varepsilon_t\}$ is an i.i.d. innovation series with non-negative support density $p(\varepsilon; \phi)$, in which
 594 ϕ is a vector of parameters and:

$$\psi_t \equiv \sum_{j=1}^p \alpha_j Y_{t-j} + \sum_{k=1}^q \beta_k \psi_{t-k}, \tag{5.10}$$

595 where $\{\psi_t\}$ denotes the process of conditional expectation, which summarizes the dynam-
 596 ics of the duration process. Suppose that the processes $\{Y_t\}$ and $\{\psi_t\}$ are both strictly
 597 stationary and α -mixing with the mixing coefficients $\alpha_x(n)$ and $\alpha_\psi(n)$ satisfying $\alpha_x(n) \leq$
 598 $C_x q_x^n$ and $\alpha_\psi(n) \leq C_\psi q_\psi^n$, respectively, where $0 < C_x, C_\psi < \infty$ and $0 < q_x, q_\psi < 1$.

599 The ACD model in (5.9) is considered by many to be too restrictive to take care of the
 600 dynamics of the duration process in practice. Furthermore, estimating the model requires

601 the imposition of a distributional assumption on ε_t , a requirement that is not popular in
 602 the literature; see Pacurar (2008) for an excellent review of the ACD literature. Saart
 603 et al. (2013) attempt to minimize impacts of such issues by introducing the SEMI-ACD
 604 model such that

$$\psi_t \equiv \sum_{j=1}^p \gamma_j Y_{t-j} + \sum_{k=1}^q \mathbf{g}_k(\psi_{t-k}), \quad (5.11)$$

605 where γ_j is an unknown parameter and $\mathbf{g}_k(\cdot)$ is an unknown function on the real line.
 606 Even though the above mentioned distributional assumption is not required to estimate
 607 these semiparametric models, a latency problem arises because the conditional duration
 608 (ψ) is not observable in practice.

609 To estimate the model, the authors rely on an iterative estimation algorithm. For a
 610 special case of the model where $p = q = 1$, i.e. the so-called SEMI-ACD(1,1) model, the
 611 algorithm can be summarized as follows: *Step 1:* Choose the starting values for the vector
 612 of the n conditional durations. Index these values with a zero. Let $\{\widehat{\psi}_{t,0}; 1 \leq t \leq n\}$ satisfy
 613 $\widehat{\psi}_{t,0} = \psi_{t,0}$. Set $m = 1$. *Step 2:* Compute $\widehat{\gamma}_m$ and $\widehat{g}_{h,m}$, by regressing $\{Y_t; 2 \leq t \leq n\}$
 614 against $\{Y_{t-1}; 2 \leq t \leq n\}$ and the estimates of ψ computed in the previous step, i.e.
 615 $\{\widehat{\psi}_{t-1,m-1}; 2 \leq t \leq n\}$. *Step 3:* Compute $\{\widehat{\psi}_{t,m}; 2 \leq t \leq n\}$. Furthermore, use the average
 616 of $\{\widehat{\psi}_{t,m}; 2 \leq t \leq n\}$ as a proxy for $\widehat{\psi}_{1,m}$, which cannot be computed recursively. *Step*
 617 *4:* For $1 \leq m < m^*$, where $m^* = O(\log(n))$ is the (pre-specified) maximum number of
 618 iterations, increment m and return to Step 2. At $m = m^*$, perform the final estimation
 619 to obtain the final estimates of γ and g .

620 Saart et al. (2013) studied the asymptotic properties of such a procedure for the
 621 SEMI-ACD(1,1) model by first deriving the consistency of the estimation algorithm, i.e.

$$\left\| \widehat{\Psi}_m - \Psi \right\|_{1e} \leq \Delta_{1n}(\widehat{\psi}) C_m(G) + G^m \Delta_{2n}(\psi), \quad (5.12)$$

622 where $0 < G < 1$, $\widehat{\Psi}_m = (\widehat{\psi}_{m+1,m}, \dots, \widehat{\psi}_{n,m})^\tau$ and $\Psi = (\psi_{m+1}, \dots, \psi_n)^\tau$. Although the
 623 details are shown in Theorem 3.1 of the paper, let us simply note that while the first
 624 term on the right side of (5.12) converges to zero uniformly over all the possible values of
 625 the bandwidth, h , at the same rate as the mean squared error of a usual PL time series
 626 model, the limit of the second term is zero as $m \rightarrow 0$. Hence, m^* in this case is selected so
 627 that the second term is bounded in probability by the first. Note that the proof of (5.12)
 628 requires the following contraction property on g .

629 **Assumption 5.1.** (*Assumption 3.1 of Saart et al. (2013)*) Suppose that the function g
 630 on the real line satisfies the following Lipschitz type condition:

$$|g(x + \delta) - g(x)| \leq \varphi(x)|\delta| \quad (5.13)$$

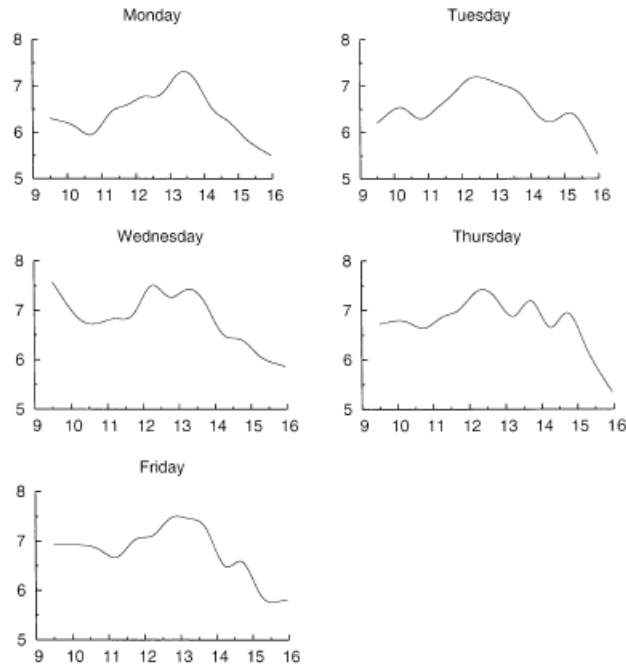
631 for each given $x \in S_\omega$, where S_ω is a compact support. Furthermore, $\varphi(\cdot)$ is a nonnegative
 632 measurable function such that with probability one and for some $0 < G < 1$:

$$\max_{t \geq 1} E [\varphi^2(\psi_t) | (\psi_{t-1}, \dots, \psi_1)] \leq G^2; \max_{t \geq 1} E [\varphi^2(\psi_{t,m}) | (\psi_{t-1,m-1}, \dots, \psi_{1,1})] \leq G^2.$$

633 The asymptotic normality of the least squares estimator of the unknown parameter is
 634 then proved using the Taylor expansions together with the convergence rates for kernel
 635 estimation derived in Hansen (2008). Saart et al. (2013) also showed that an extension
 636 of the SEMI-ACD(1,1) model to, for example, a SEMI-ACD(p,q) model, where $q \leq 3$,
 637 is also possible without requiring an additional assumption. This is in the sense that the
 638 \sqrt{n} -asymptotic normality of the relevant estimators still holds, provided that $q \leq 3$. Such
 639 a claim is supported by the results found in Robinson (1988) and Fan and Li (1999b).

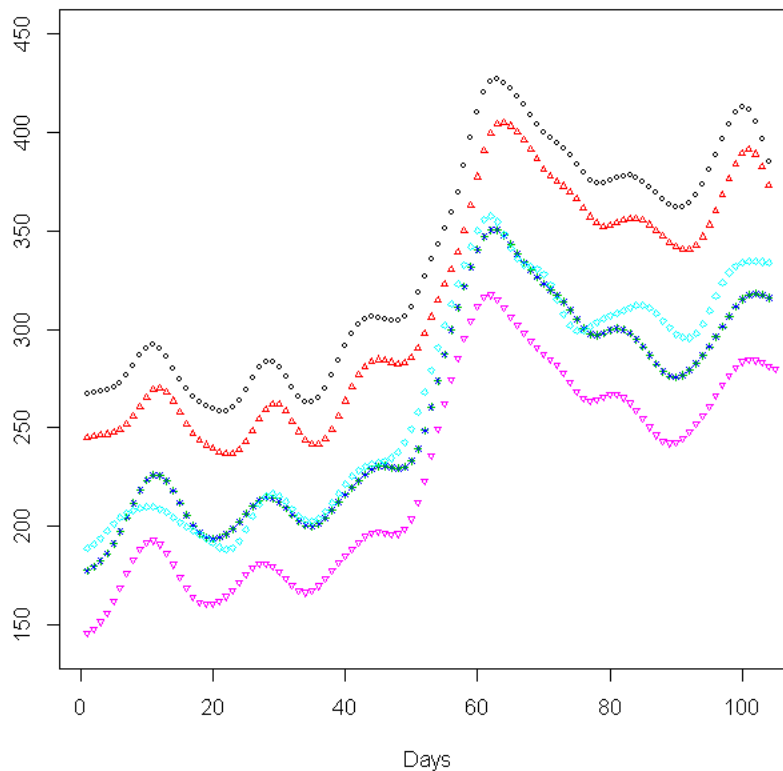
640 Let us finish this section with discussion of some important issues on the use of the
 641 SEMI-ACD model in practice. Firstly, the construction of the ACD models is done under
 642 the assumption that the duration process is stationary. However, it is well known that
 643 intraday financial data often involve some strongly diurnal patterns. Hence, the first
 644 step toward the econometric analysis of financial duration is always to perform a diurnal
 645 adjustment.

646 **Figure 5.1.** *Diurnal Patterns of Financial Trading for Monday to Friday*



647 Nonetheless, the remaining problems are other trading patterns, for example day of
 648 the week effects, that have not been taken care of. To illustrate the problem, let us
 649 consider Figure 5.1, which is a reproduction of Figure 7 of Meitz and Teräsvirta (2006).
 650 The figure presents the nonparametric kernel estimate of diurnal components of the price
 651 duration series for IBM for each of the trading days. The data are high-frequency data
 652 for IBM shares between July 2002 and December 2002. The results show a similar but
 653 not exactly identical inverted U-shaped pattern in the moving average of durations over
 654 the days. An idea, which is a work in progress, is to use the fact that the EG-PLSI model
 655 in (2.9) allows the nonparametric shape-invariant analysis (Härdle and Marron (1990))
 656 and to jointly model the regular components of the duration process without unpooling
 657 the data. We have applied the idea to the total number of daily hospital admissions of
 658 circulatory and respiratory patients in Hong Kong between 1994 to 1996, i.e. the data set
 659 originally used in Xia et al. (2002). Figure 5.2 below presents the estimated time trend
 660 taking day-effects into account (\circ Monday, ∇ Tuesday, \times Wednesday, \diamond Thursday and \triangle
 661 Friday). (Details of this work are available from the authors upon request.)

662 **Figure 5.2.** *Trend and Day-Effects for Hong Kong Patients*



663 Finally, the SEMI-ACD model can potentially be used in various empirical studies of
 664 financial market microstructure. Saart and Gao (2012), for example, applied the SEMI-
 665 ACD model to model the intertemporal dynamics of the price change duration process
 666 in stock exchange markets. The probability distribution of the resulting estimates of the
 667 so-called standardized durations were then hypothetically tested in order to obtain some
 668 information about that of the true duration processes. Although the details can be found
 669 in the paper, it is noted here that the outcomes of the above mentioned test are different
 670 when it is implemented based on the SEMI-ACD model compared to when it is based
 671 on the parametric ACD model. Furthermore, a work in progress is being conducted in
 672 the use of the resulting standardized duration from the SEMI-ACD model to study the
 673 exogeneity of trade arrivals in the financial market (details of this work are available from
 674 the authors upon request).

675 **6. Semiparametric Models with Nonstationary Data**

676 Firstly, in this section we will review a number of semiparametric models, which have
 677 been established to help detect and estimate trend and seasonality. Furthermore, since
 678 we have discussed the endogeneity problem in semiparametric models in detail in the
 679 previous section, it will also be of particular interest to also review the estimation of
 680 semiparametric models that involves both endogeneity and nonstationarity.

681 *Semiparametric Detection and Estimation of Trend and Seasonality*

682 Many important macroeconomic and financial data, such as income, unemployment and
 683 retail sale, are found to exhibit deterministic/stochastic trends. The closest semipara-
 684 metric model to the PL model that explicitly allows for a trend detection is the PL time
 685 series error model introduced by Gao and Hawthorne (2006), of the form:

$$Y_t = U_t^\tau \beta + g\left(\frac{t}{n}\right) + \varepsilon_t, \quad t = 1, 2, \dots, n, \quad (6.1)$$

686 where $\{Y_t\}$ is a response variable (e.g. the mean temperature series), $U_t = (U_{t1}, \dots, U_{tq})^\tau$
 687 is a vector of q -explanatory variables (e.g. the southern oscillation index), t is the time
 688 in years, β is a vector of unknown coefficients for the explanatory variables, $g(\cdot)$ is an
 689 unknown smooth function of time representing the trend and $\{\varepsilon_t\}$ represents a sequence
 690 of stationary time series errors with $E[e_t] = 0$ and $0 < \text{var}[e_t] = \sigma^2 < \infty$. In order to
 691 estimate the model, Gao and Hawthorne (2006) introduce an estimation procedure, which
 692 is closely similar to that of the above mentioned PL time series model: (i) compute an
 693 estimate of $g(\cdot)$ for a given β , i.e. similar to the second term of (3.18); (ii) compute the

694 least-squares estimate of β ; (iii) compute the new estimate of $g(\cdot)$ based on that of β
 695 estimated in the previous step. Gao and Hawthorne (2006) also consider an alternative
 696 case where $\{\varepsilon_t\}$ is allowed to be $I(1)$. This is to say that $\{\varepsilon_t\}$ itself may be nonstationary,
 697 but its differences $\delta_t = \varepsilon_t - \varepsilon_{t-1}$ are assumed to be stationary. In this case, we need only
 698 to consider the first differenced version of (6.1) of the form

$$V_t = W_t^\tau \beta + m\left(\frac{t}{n}\right) + \delta_t, \quad t = 1, 2, \dots, n, \quad (6.2)$$

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

700 These models enable us to study an important issue in practice, which is to deter-
 701 mine whether a linear trend is able to approximate the behavior of the series in question
 702 adequately. Using the model in (6.1), such a problem can be written as the hypotheses

$$H_0 : g\left(\frac{t}{n}\right) = \alpha_0 + \gamma_0 t \quad \text{versus} \quad H_1 : g\left(\frac{t}{n}\right) \neq \alpha + \gamma t \quad (6.3)$$

703 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 .
 704 Hence, this issue is coherent with the general interest in statistics and econometrics, which
 705 involves testing the hypotheses of a parametric form against a nonparametric alternative.
 706 Inspired by Horowitz and Spokoiny (2001), Gao and Hawthorne (2006) propose a novel
 707 test for linearity in the trend function $g(\cdot)$ under such semiparametric settings such as
 708 (6.1) and (6.2). For each given value of bandwidth h , to test H_0 , Gao and Hawthorne
 709 (2006) propose using the following:

$$L_{4n}(h) = \frac{\sum_{t=1}^n \sum_{s=1, \neq t}^n K\left(\frac{s-t}{nh}\right) \tilde{\varepsilon}_s \tilde{\varepsilon}_t}{\tilde{S}_n}, \quad (6.4)$$

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

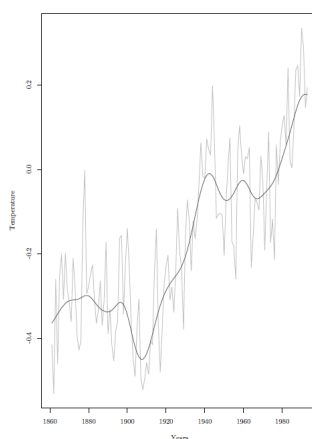
712 By applying the above method, Gao and Hawthorne (2006) shows that the trend esti-
 713 mate of the global temperature series for 1867 to 1993 appears to be distinctly nonlinear.
 714 Figure 5.3 below is a reproduction Figure 4 of Gao and Hawthorne (2006), which shows
 715 the global temperature series for 1867 to 1993 and the estimated trend. Furthermore,
 716 Gao and Hawthorne (2006) also consider the possible nonstationarity of the residuals in
 717 the models by applying the first differenced version of the model defined in (6.2). The hy-
 718 pothesis testing described in (6.3) and (6.4) is then employed. They report that a similar
 719 conclusion – rejecting the linearity in the trend – can be drawn by using either the level
 720 or differenced version of the data. Note that in order to perform a nonparametric kernel
 721 testing such as this, bandwidth selection can be crucial and may significantly affect the

722 outcome of the test. A novel idea about the testing procedure in Gao and Hawthorne
 723 (2006) is the use of a maximized version of the test such that:

$$L^* = \max_{h \in H_n} L_{4n}(h). \quad (6.5)$$

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

727 **Figure 5.3.** *Global temperature series for 1867 to 1993 (light line) and the estimated trend (solid curve)*



728 More recently, there is a new semiparametric PL time series model has been developed
 729 by Chen et al. (2011). Although this model does not assist us with the dimension reduction
 730 problem, Chen et al. (2011) procedure provides a convenient estimation of the following
 731 extended version of the PL time series model:

$$Y_t = \beta(U_t, \theta_1) + g(U_t) + \varepsilon_t, \quad (6.6)$$

732 where $\beta(\cdot, \theta_1)$ is the known link function indexed by an unknown parameter vector $\theta_1 \in$
 733 $\Theta \subset \mathbb{R}^p$ ($p \geq 1$). An important point to note about the model in (6.6) is the fact that
 734 $\{U_t\}$ is allowed to be generated by

$$U_t = H\left(\frac{t}{n}\right) + u_t, \quad (6.7)$$

735 where $H(t)$ is unknown functions defined on \mathbb{R}^d and $\{u_t\}$ is a sequence of i.i.d. random
 736 errors. In other words, it allows for the existence of deterministic trends in the regressors.
 737 Chen et al. (2011) studied a case where nonstationarity was allowed and was driven by
 738 a deterministic trending component. Regarding the model's estimation procedure, Chen

739 et al. (2011) provided two alternative methods, namely the nonlinear least squares (see
 740 Gao (1995) and Gao (2012) for example) and the semiparametric weighted least squares
 741 estimations (see Härdle et al. (2000) for example). Among these methods, the former
 742 first estimates θ_1 ; such an estimate is then used in order to compute that of $g(\cdot)$, while
 743 the latter operates in just the reverse order. More important issues, however, are the
 744 identifiability and estimatability of the model. The following conditions are needed in
 745 Chen et al. (2011) in order to ensure that θ_1 in (6.6) is identifiable and estimable.

746 **Assumption 6.1.** (*Assumption A2 of Chen et al. (2011)*) (i) $\beta(U_t, \theta)$ is twice differen-
 747 tiable with respect to θ , and both $g(\cdot)$ and $H(\cdot)$ are continuous. (ii) Denoting the partial
 748 derivative of $\beta(U_t, \theta)$ with respect to θ by $\dot{\beta}(U_t, \theta)$, then

$$749 \quad \Gamma(\theta) := \int_0^1 \left\{ \int g(v) \dot{\beta}(v, \theta) p_u(v - H(r)) dv \right\} dr = 0$$

750 for all $\theta \in \Theta$ and $\int_0^1 \left\{ \int [\beta(v, \theta_1) - \beta(v, \theta)] \dot{\beta}(v, \theta) p_u(v - H(r)) dv \right\} dr \neq 0$ uniformly in
 751 $\theta \in \Theta(\delta) = \{\theta : \|\theta - \theta_1\| \leq \delta\}$ for any $\delta > 0$.

752 In addition, there is an alternative model that is closely related to (6.6), which is dis-
 753 cussed in Gao (2012). Unfortunately, due to the unavailability of the asymptotic results,
 754 the study focuses only on the case where $p = 1$. Gao's (2012) model can be obtained sim-
 755 ply by replacing the parametric component with $x_t \beta$ and the nonparametric component
 756 with $g(x_t)$, whereby the regressor is defined as in the assumption below.

757 **Assumption 6.2.** (*Assumption 3.2(i) of Gao (2012)*) Let $x_t = x_{t-1} + u_t$ with $x_0 = 0$
 758 and $u_t = \sum_{i=0}^{\infty} \psi_i \eta_{t-i}$, where $\{\eta_t\}$ is a sequence of independent and identically distributed
 759 random errors, and $\{\psi_i : i \geq 0\}$ is a sequence of real numbers such that $\sum_{i=0}^{\infty} i^2 |\psi_i| < \infty$
 760 and $\sum_{i=0}^{\infty} \psi_i \neq 0$.

761 The required smallness conditions on $g(\cdot)$ are provided for two cases: stationary and
 762 nonstationary regressors. While the conditions of the stationary regressors are discussed
 763 in details in our discussion of the SL model for time series, those required for the nonsta-
 764 tionary regressor case are the following:

765 **Assumption 6.3.** (*Assumption 3.1 of Gao (2012)*) (i) Let $g(\cdot)$ be a real function on
 766 $\mathbb{R}^1 = (-\infty, \infty)$ such that $\int |x|^i |g(x)|^i dx < \infty$ for $i = 1, 2$, and $\int xg(x)dx \neq 0$; (ii) In
 767 addition, let $g(\cdot)$ satisfy $\int \left| \int e^{ixy} yg(y) dy \right| dx < \infty$ when $\int xg(x)dx = 0$.

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

770 smallness condition in practice is presented in Example 5.3 of Gao (2012). The author
 771 considers the logarithm of British pound/American dollar real exchange rate defined by:

$$y_t = \log(e_t) + \log(p_t^{UK}) - \log(p_t^{US}), \quad (6.8)$$

772 where $\{e_t\}$ is the monthly average of the nominal exchange rate, and $\{p_t^j\}$ denotes the
 773 consumer price index of country j . He finds that $\{y_t\}$ approximately follows a threshold
 774 model of the form

$$y_t = y_{t-1} - 1.1249y_{t-1}I[|y_{t-1}| \leq 0.0134] + e_t. \quad (6.9)$$

775 This result suggests that, although $\{y_t\}$ does not necessarily follow an integrated time
 776 series model, e.g. $y_t = y_{t-1} + e_t$, it behaves like a nearly integrated time series, since
 777 the nonlinear component is a small departure function (see also the discussion on the
 778 semiparametric threshold models in Section 7).

779 In the literature, there is a number of mathematical approaches which have been
 780 established as tools for deriving an asymptotic theory for the nonparametric estimation
 781 of univariate models of nonstationary data. Below, let us mention a couple (see also the
 782 review in Sun and Li (2012) for details). Firstly, we have the Markov splitting technique
 783 used in; for example, Karlsen and Tjøstheim (2001) and Karlsen et al. (2007) that is
 784 used to model univariate time series with a null recurrent structure. Secondly, we have
 785 the local time methods developed by Phillips (2009) and Wang and Phillips (2011) used
 786 to derive an asymptotic theory for the nonparametric estimation of univariate models
 787 with an integrated time series. In Gao's (2012) model, since $\{x_t\}$ is nonstationary, the
 788 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
 789 leading to

$$\hat{\beta} = \left(\sum_{t=1}^n x_t^2 \right)^{-1} \left(\sum_{t=1}^n x_t y_t \right), \quad (6.10)$$

790 which is closely related to the results of (3.18). Although the details are discussed in
 791 the paper, we note here that in order to establish an asymptotic distribution for $\hat{\beta}$, it is
 792 necessary that, as $n \rightarrow \infty$, we have:

$$\frac{1}{n} \sum_{t=1}^n x_t g(x_t) \rightarrow_P 0.$$

793 Regarding the case of a nonstationary regressor, $\int xg(x)dx$ may or may not be zero.
 794 The asymptotic distribution of the estimators, namely the above ordinary least squares
 795 estimator of the unknown parameter β and the nonparametric estimator of $g(\cdot)$, are based
 796 very much on Theorem 2.1 of the studies by Wang and Phillips (2009a), and Wang and
 797 Phillips (2011).

799 In the case of independent and stationary time series data, semiparametric methods have
 800 been shown to be particularly useful in modelling economic data in a way that retains
 801 generality where it is most needed while reducing dimensionality problems. Gao and
 802 Phillips (2013) sought to pursue these advantages in a wider context that allows for
 803 nonstationarities and endogeneities within a vector semiparametric regression model. In
 804 their study, the time series $\{(Y_t, X_t, V_t) : 1 \leq t \leq n\}$ were assumed to be modeled in a
 805 system of multivariate nonstationary time series models of the form:

$$Y_t = AX_t + g(V_t) + e_t; \quad (6.11)$$

$$X_t = H(V_t) + U_t \quad t = 1, 2, \dots, n;$$

$$E[e_t|V_t] = E[e_t] = 0; \text{ and} \quad (6.12)$$

$$E[U_t|V_t] = 0, \quad (6.13)$$

806 where n is the sample size, A is a $p \times d$ -matrix of unknown parameters, $Y_t = (y_{t1}, \dots, y_{tp})^\tau$,
 807 $X_t = (x_{t1}, \dots, x_{td})^\tau$ and V_t is a sequence of univariate integrated time series regressors,
 808 $g(\cdot) = (g_1(\cdot), \dots, g_p(\cdot))^\tau$ and $H(\cdot) = (h_1(\cdot), \dots, h_d(\cdot))^\tau$ are all unknown functions, and
 809 both e_t and U_t are vectors of stationary time series. Note that $\{X_t\}$ can be stationary only
 810 when $\{X_t\}$ and $\{V_t\}$ are independent. The identification condition $E[e_t|V_t] = E[e_t] = 0$
 811 in (6.12) eliminates endogeneity between e_t and V_t while retaining endogeneity between e_t
 812 and X_t and potential nonstationarity in both X_t and V_t . In this setting, such a condition
 813 corresponds to the condition $E[e_t|V_t, U_t] = E[e_t|U_t]$ that is assumed in Newey et al. (1999),
 814 for example. The rationale behind (6.12) is the fact that

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

815 when U_t is independent of V_t and $E[e_t] = 0$. These conditions are less restrictive than
 816 the exogeneity condition between e_t and (X_t, V_t) that is common in the literature for the
 817 stationary case. In the study by Gao and Phillips (2013), the model is treated as a vector
 818 semiparametric structural model, and considers the case where X_t and V_t may be vectors
 819 of nonstationary regressors and X_t may be endogenous. The main contribution of the
 820 study resides in the derivation of a semiparametric instrumental variable least squares
 821 estimate of A to deal with endogeneity in X_t and a nonparametric estimator for the
 822 function $g(\cdot)$. Let us assume that there exists a vector of stationary variables η_t for which
 823 we have:

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

824 The derivation of the semiparametric instrumental variable least squares estimate of A
825 can now be done based on the following expanded version of the system (6.11):

$$Y_t = AX_t + g(V_t) + e_t \quad t = 1, 2, \dots, n \quad (6.14)$$

$$X_t = H(V_t) + U_t$$

$$Q_t = J(V_t) + \eta_t$$

$$E[e_t|V_t] = E[e_t] = 0, \quad E[U_t|V_t] = 0 \quad \text{and} \quad E[\eta_t|V_t] = 0, \quad (6.15)$$

826 where $Q_t = (q_{t1}, \dots, q_{td})^\tau$ is a vector of possible instrumental variables for X_t generated
827 by a reduced form equation involving V_t , and $I(\cdot) = (J_1(\cdot), \dots, J_d(\cdot))^\tau$ is a vector of
828 unknown functions. The limiting theory in this kind of nonstationary semiparametric
829 model depends on the probabilistic structure of the regressors and errors e_t , U_t , η_t and V_t ,
830 as well as the functional forms of $g(\cdot)$, $H(\cdot)$ and $J(\cdot)$. Gao and Phillips (2013) provide a
831 list of the conditions required including, their detailed explanation in Appendix A of the
832 paper.

833 7. Conclusions and Discussion

834 We have seen in the literature that theoretical and empirical research in time series analy-
835 sis may be conducted on a large number of topics. Among these, we personally believe that
836 perhaps nonlinear time series models are the most studied over recent years. In order to
837 take the nonlinearity in time series regression in to account, nonparametric methods have
838 been very popular both for predicting and characterizing nonlinear dependence. However,
839 their developments has been significantly dampened by the so-called curse of dimension-
840 ality. Firstly, we reviewed a number of semiparametric time series models offered in the
841 literature as the methods for combating the curse of dimensionality and their specifica-
842 tion testings. In order to proceed along a linked sequence of materials, we identified two
843 links between these semiparametric models, namely exogeneity and stationarity condi-
844 tions. Addressing the breakdown in the former led to the emergence of semiparametric
845 models with generated regressors, while addressing the breakdown in stationarity led to
846 semiparametric models of nonstationary data. We presented a detailed review of recent
847 developments of these models. Nonetheless, since time series models for nonstationary
848 data provide a large field of research, the review in this paper focused on semiparametric
849 models established to help detect and estimate trend and seasonality, and semiparametric
850 models that involve both endogeneity and nonstationarity.

851 In many places throughout the previous sections, we have provided our views on future
852 research. In the following, let us discuss some additional open questions about this area of

853 research. In our view, the problems of endogeneity and nonstationarity are both impor-
854 tant issues that future research in the area of semiparametric time series should be based
855 on. As reviewed previously, Kim and Saart (2013) and Kim et al. (2013) successfully ad-
856 dressed the endogeneity problem in the PL and the EG-PLSI models. However, a number
857 of questions are left unanswered, especially the importance of weak/strong instruments
858 and the characteristics of the control function on the performance of the CF approach.
859 Furthermore, a more detailed comparison between the CF and the NpIV approaches than
860 what was done in Kim et al. (2013) is required, especially on the characteristics of the
861 endogeneity for which these methods would be more advantageous. An advantage of the
862 CF approach is its ability to disentangle the structural nonparametric relationship and the
863 effect of endogeneity. Hence, a simple test of exogeneity can be developed by testing the
864 statistical significance of the above mentioned effect of endogeneity. The first attempt by
865 Kim et al. (2013) was to use bias-corrected confidence bands in nonparametric regression
866 of Xia (1998). However, we believe that a more formal test can be developed based on
867 this idea.

868 More recently, there has also been an attempt by Gao et al. (2013) to detect and to
869 estimate a structural change from a nonlinear stationary regime to a linear nonstationary
870 regime using a semiparametric threshold autoregressive model, which can be conveniently
871 expressed as

$$\begin{aligned}
Y_t &= g(Y_{t-1})I[Y_{t-1} \in C_\tau] + \alpha Y_{t-1}I[Y_{t-1} \in D_\tau] + \varepsilon_t \\
&= \begin{cases} g(Y_{t-1}) + \varepsilon_t & \text{if } Y_{t-1} \in C_\tau \\ \alpha Y_{t-1} + \varepsilon_t & \text{if } Y_{t-1} \in D_\tau, \end{cases} \tag{7.1}
\end{aligned}$$

872 where C_τ is either a compact subset of R^1 or a set of the type $(-\infty, \tau]$ or $[\tau, \infty)$, D_τ
873 is the complement of C_τ , $g(x)$ is an unknown and bounded function when $x \in C_\tau$ and
874 $\alpha = 1$. Lemma 3.1 of the paper shows a special case of the model where $\alpha = 1$ is a
875 β -null recurrent Markov Chain process; see also a detailed discussion on a null recurrent
876 process in Karlsen et al. (2007). The existing asymptotic results for the stationary non-
877 linear time series models (for instance in Fan and Yao (2003), and Gao (2007)) are not
878 directly applicable. While Gao et al. (2013) studied the asymptotic behavior of both a
879 nonparametric estimator of $g(\cdot)$ and the least square estimator of α , their mathematical
880 proof relied heavily on a number of general results of the β -null recurrent Markov chains
881 discussed in Karlsen and Tjøstheim (2001).

882 As an alternative, we could establish a new threshold autoregressive model such that
883 the response variable Y depends on the vector of stochastic explanatory variables or

884 stochastic covariates $X = (X_1, \dots, X_p)^T$, where ($p \geq 2$) as follows:

$$\begin{aligned}
 Y &= \beta_0^T X + \phi(\theta_0^T X) I[\theta_0^T X \in C_\tau] + \varepsilon \\
 &= \begin{cases} \beta_0^T X + \phi(\theta_0^T X) + \varepsilon, & \text{if } \theta_0^T X \in C_\tau \\ \beta_0^T X + \varepsilon, & \text{otherwise,} \end{cases} \quad (7.2)
 \end{aligned}$$

885 where the conditions stated in Assumption 2.1 hold and C_τ is either a compact subset of \mathbb{R}^1
 886 or a set of the type $(-\infty, \tau]$ or $[\infty, \tau)$. We refer to the model as the partially-linear single-
 887 index threshold autoregressive (PLSi-TAR) model. Let us state a few remarks regarding
 888 the PLSi-TAR model: (i) Compared to the SEMI-TAR model of Gao et al. (2013), the
 889 PLSi-TAR model offers alternative types of flexibility, which can be quite useful when
 890 attempting to perform dimension reduction in modeling time series data. (ii) The model
 891 can be used to detect the structural change from a nonlinear stationary regime to a linear
 892 stationary regime. However, by relaxing some of the conditions in Assumption 2.1 the
 893 model can also be used to detect the structural change from a nonlinear stationary regime
 894 to a linear nonstationary regime. As another alternative, we may introduce an extended
 895 PLSi-TAR model of the form:

$$\begin{aligned}
 Y_t &= g(X_t, \theta_1) + \phi(\theta_0^T X_t) I[\theta_0^T X_t \in C_\tau] + \varepsilon_t \\
 &= \begin{cases} g(X_t, \theta_1) + \phi(\theta_0^T X_t) + \varepsilon_t & \text{if } \theta_0^T X_t \in C_\tau, \\ g(X_t, \theta_1) + \varepsilon_t & \text{otherwise,} \end{cases} \quad (7.3)
 \end{aligned}$$

896 where $g(\cdot, \theta_1)$ is a known link function indexed by an unknown parameter vector $\theta_1 \in$
 897 \mathbb{R}^p ($p \geq 1$).

898 Finally, let us give a remark on another important research direction, which focuses
 899 instead on improving parametric time series modeling. This line of development may be
 900 worth further exploration in parallel with those methods discussed previously. Clearly, an
 901 important benefit of the above semiparametric models resides in the additional flexibility
 902 that they provide given a constraint in the form of the curse of dimensionality. However,
 903 if we take a different point of view, for example, that all models are wrong, but some are
 904 useful (Box (1976)), then the usual arguments in favor of non/semi parametric models are
 905 substantially weakened (especially in time series analysis). As suggested by an anonymous
 906 referee, in this case, it is perhaps particularly relevant to explore ways to fit mis-specified
 907 parametric models more earnestly. An example of the studies in this area is that of Xia
 908 and Tong (2011).

References

- Ai, C., Chen, X., 2003. Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica* 71 (6), 1795–1843.
- Aït-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.
- Blundell, R., Powell, J., 2003. Endogeneity in nonparametric and semiparametric regression models. *Econometric Society Monographs* 36, 312–357.
- Blundell, R., Powell, J., 2004. Endogeneity in semiparametric binary response models. *Review of Economics Studies* 71, 655–679.
- Box, G. E., 1976. Science and statistics. *Journal of the American Statistical Association* 71 (356), 791–799.
- 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.
- Chan, K. S., Tong, H., 1986. On estimating thresholds in autoregressive models. *Journal of time series analysis* 7 (3), 179–190.
- 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.
- Darolles, S., Fan, Y., Florens, J.-P., Renault, E., 2011. Nonparametric instrumental regression. *Econometrica* 79 (5), 1541–1565.

- Dufrénot, G., Mignon, V., 2002. *Recent Developments in Nonlinear Cointegration with Applications to Macroeconomics and Finance*. Springer.
- 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 relation between weather and electricity sales. *Journal of the American Statistical Association* 81 (394), 310–320.
- Eubank, R., 1999. *Nonparametric Regression and Spline Smoothing*. Vol. 157. CRC Press: Boca Raton.
- 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 Modeling and its Applications*. Chapman & Hall, London.
- 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: New York.
- 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* 64 (4), 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 semi-parametric 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., 2007. *Nonlinear Time Series: Semiparametric and Nonparametric Methods*. Chapman & Hall, London.
- Gao, J., 2012. *Handbook on Non- and Semi-Parametric Econometrics and Statistics*. Oxford University Press, Oxford, Ch. Identification, estimation and specification in a class of semi-linear time series models.
- 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., Lu, Z., Tjøstheim, D., 2006. Estimation in semiparametric spatial regression. *The Annals of Statistics* 34 (3), 1395–1435.
- Gao, J., Phillips, P. C., 2013. Semiparametric estimation in triangular system equations with nonstationarity. *Journal of Econometrics*.
- Gao, J., Tjøstheim, D., Yin, J., 2013. Estimation in threshold autoregressive models with nonstationarity. *Journal of Econometrics* 172, 1 – 13.
- 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.
- Hall, P., Horowitz, J., 2005. Non-parametric method for inference in the presence of instrumental variables. *The Annals of Statistics* 33, 2904–2929.
- 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., Mammen, E., 1993. Comparing nonparametric versus parametric regression fits. *The Annals of Statistics* 21 (4), 1926–1947.

- Härdle, W., Marron, J., 1990. Semiparametric comparison of regression curves. *The Annals of Statistics* 18 (1), 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. L., Spokoiny, V. G., 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., Tjøstheim, 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.
- Kim, N., Saart, P., 2013. Estimation in semiparametric partially linear models with parametric and/or nonparametric endogeneity. SSRN Working Paper Series. Available at SSRN: <http://ssrn.com/abstract=2213071> (accessed 18 June 2013).
- Kim, N., Saart, P., Gao, J., 2013. Semi-parametric analysis of shape-invariant Engel curves with control function approach. SSRN Working Paper Series. Available at SSRN: <http://ssrn.com/abstract=2213067> (accessed 18 June 2013).
- 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., 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.
- Meitz, M., Teräsvirta, T., 2006. Evaluating models of autoregressive conditional duration. *Journal of Business & Economic Statistics* 24 (1).
- Newey, W., Powell, J., 2003. Instrumental variable estimation of nonparametric models. *Econometrica* 71 (5), 1565–1578.
- 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.
- O’Sullivan, F., 1986. A statistical perspective on ill-posed inverse problems. *Statistical Science* 1 (4), 502–518.
- Pacurar, M., 2008. Autoregressive conditional duration models in finance: a survey of the theoretical and empirical literature. *Journal of Economic Surveys* 22, 711–751.
- Phillips, P., 2009. Local limit theory and spurious nonparametric regression. *Econometric Theory* 25 (06), 1466–1497.
- Pinkse, J., 2000. Nonparametric two-step regression estimation when regressors and error are dependent. *The Canadian Journal of Statistics* 28 (2), 289–300.
- Robinson, P., 1988. Root-N-consistent semiparametric regression. *Econometrica* 56, 931–954.
- Saart, P. W., Gao, J., 2012. An alternative nonparametric specification test in nonlinear autoregressive conditional duration model. Working Paper Series. Available at SSRN: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2130454 (accessed on 18 June 2013).
- Saart, P. W., Gao, J., Allen, D., 2013. Semiparametric autoregressive conditional duration model: theory and practice. *Econometric Reviews* (in press) 33.
- 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 non-stationary time series. Working Paper.
- Teräsvirta, T., Tjøstheim, D., Granger, C. W., 2011. *Modelling nonlinear economic time series*. Oxford University Press, New York.
- Tjøstheim, D., Auestad, R., 1996. Nonparametric identification of nonlinear time series: projection. *Journal of the American Statistical Association* 89, 1398–1409.

- Tong, H., Lim, K., 1980. Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society. Series B (Methodological)* 42 (3), 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 models: a 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.
- 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., 1998. Bias-corrected confidence bands in nonparametric regression. *Journal of the Royal Statistical Society, Series B* 60, 797–811.
- Xia, Y., Tong, H., 2011. Feature matching in time series modeling. *Statistical science* 26 (1), 21–46.
- 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.
- Yoshihara, K.-I., 1976. Limiting behavior of U-statistics for stationary, absolutely regular processes. *Probability Theory and Related Fields* 35 (3), 237–252.