



MONASH University

Australia

Department of Econometrics and Business Statistics

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

**Identification, Estimation and Specification
In a Class of Semiparametric Time Series Models**

Jiti Gao

March 2012

Working Paper 06/12

Identification, Estimation and Specification in a Class of Semi–Linear Time Series Models

BY JITI GAO^{†1}

[†]*Monash University*

Abstract

In this paper, we consider some identification, estimation and specification problems in a class of semi–linear time series models. Existing studies for the stationary time series case have been reviewed and discussed. We also establish some new results for the integrated time series case. In the meantime, we propose a new estimation method and establish a new theory for a class of semi–linear nonstationary autoregressive models. In addition, we discuss certain directions for further research.

Keywords: Asymptotic theory, departure function, kernel method, nonlinearity, nonstationarity, semiparametric model, stationarity, time series

JEL Classifications: C13, C14, C22.

¹Jiti Gao is from Department of Econometrics and Business Statistics, Monash University, Caulfield East, Melbourne, VIC 3145, Australia. Email: jiti.gao@monash.edu. [Http://www.jitigao.com](http://www.jitigao.com).

1. Introduction

Consider a class of semi-linear (semiparametric) time series models of the form

$$y_t = x_t^\top \beta + g(x_t) + e_t, \quad (1.1)$$

where $\{x_t\}$ is a vector of time series regressors, β is a vector of unknown parameters, $g(\cdot)$ is an unknown function defined on R^d , $\{e_t\}$ is a sequence of martingale differences, and n is the number of observations. This paper mainly focuses on the case of $1 \leq d \leq 3$. As discussed in Section 2.2 below, for the case of $d \geq 4$, one may replace $g(x_t)$ by a semiparametric single-index form $g(x_t^\top \beta)$.

Various semiparametric regression models have been proposed and discussed extensively in recent years. Primary interest focuses on general nonparametric and semiparametric time series models under stationarity assumption. Recent studies include Tong (1990), Fan and Gijbels (1996), Härdle, Liang and Gao (2000), Fan and Yao (2003), Gao (2007), Li and Racine (2007), and Teräsvirta, Tjøstheim and Granger (2010) as well as the references therein. Meanwhile, model estimation and selection as well as model specification problems have been discussed for one specific class of semiparametric regression models of the form

$$y_t = x_t^\top \beta + \psi(v_t) + e_t, \quad (1.2)$$

where $\{v_t\}$ is a vector of time series regressors and $\psi(\cdot)$ is an unknown function. As discussed in the literature (see, for example, Robinson 1988; Chapter 6 of Härdle, Liang and Gao 2000; Gao 2007; Li and Racine 2007), a number of estimation and specification problems have already been studied for the case where both x_t and v_t are stationary and the covariance matrix $\Sigma = E[(x_t - E[x_t|v_t])(x_t - E[x_t|v_t])^\top]$ is positive definite. In recent years, attempts have also been made to address some estimation and specification testing problems for model (1.2) for the case where x_t and v_t may be stochastically nonstationary (see, for example, Juhl and Xiao 2005; Chen, Gao and Li 2012; Gao and Phillips 2011).

The focus of our discussion in this paper is on model (1.1). Model (1.1) has different types of motivations and applications from the conventional semiparametric time series model of the form (1.2). In model (1.1), the linear component in many cases plays the leading role while the nonparametric component behaves like a type of unknown departure from the classic linear model. Since such departure is usually unknown, it is not unreasonable to treat $g(\cdot)$ as a nonparametrically unknown function. In comparison with the literature by Glad (1998), Martins-Filho, Mishra and Ullah (2008), and Fan, Wu and Feng (2009), the unknown nonparametric component $g(\cdot)$ is being treated as a ‘small’ departure function in our discussion for both estimation and specification testing. In the case of model specification testing, we treat model (1.1) as an alternative when there is not enough evidence to suggest accepting a parametric true model of the form $y_t = x_t^\top \beta + e_t$. In addition, model (1.1) will also be motivated as a model to address some endogenous problems

involved in a class of linear models of the form $y_t = x_t^\tau \beta + \varepsilon_t$, where $\{\varepsilon_t\}$ is a sequence of errors with $E[\varepsilon_t] = 0$ but $E[\varepsilon_t|x_t] \neq 0$. In the process of estimating both β and $g(\cdot)$ consistently, existing methods, as discussed in the literature by Robinson (1988), Härdle, Liang and Gao (2000), Gao (2007), and Li and Racine (2007) for example, are not valid and directly applicable because $\Sigma = E[(x_t - E[x_t|x_t])(x_t - E[x_t|x_t])^\tau] = 0$. The main contribution of this paper is summarized as follows. We discuss some recent developments for the stationary time series case of model (1.1) in Section 2 below. Sections 3 and 4 establish some new theory for model (1.1) for the integrated time series case and a nonstationary autoregressive time series case, respectively.

The organization of this paper is summarised as follows. Section 2 discusses model (1.1) for the case where $\{x_t\}$ is a vector of stationary time series regressors. Section 2 also proposes an alternative model to model (1.1) for the case where $d \geq 4$. The case where $\{x_t\}$ is a vector of nonstationary time series regressors is discussed in Section 3. Section 4 considers an autoregressive case of $d = 1$ and $x_t = y_{t-1}$ and then establishes some new theory. Section 5 gives some examples to show why the proposed models are relevant and how to implement the proposed theory and estimation method in practice. This paper concludes with some remarks in Section 6.

2. Stationary models

Note in the rest of this paper that we refer to $g(\cdot)$ as a ‘small’ function if $g(\cdot)$ satisfies either Assumption 2.1(i), or, Assumption 2.3(i), or, Assumption 3.1, or Assumption 4.2(ii) below. Note also that the symbol “ $\implies_{\mathcal{D}}$ ” denotes weak convergence, the symbol “ $\rightarrow_{\mathcal{D}}$ ” denotes convergence in distribution, and “ \rightarrow_P ” denotes convergence in probability.

In this section, we give some review about the development of model (1.1) for the case where $\{x_t\}$ is a vector of stationary time series regressors. Some identification and estimation issues are then reviewed and discussed. Section 2.1 discusses the case of $1 \leq d \leq 3$, while Section 2.2 considers using a single-index model to deal with the case of $d \geq 4$.

2.1 Case of $1 \leq d \leq 3$

While the literature may mainly focus on model (1.2), model (1.1) itself has its own motivations and applications. As a matter of fact, there is also a long history about the study of model (1.1). Owen (1991) considers model (1.1) for the case where $\{x_t\}$ is a vector of independent regressors and then treats $g(\cdot)$ as a misspecification error before an empirical likelihood estimation method is proposed. Gao (1992) systematically discusses model (1.1) for the case where $\{x_t\}$ is a vector of independent regressors and then considers both model estimation and specification issues. Before we start our discussion, we introduce an identifiability condition of the form in Assumption 2.1.

Assumption 2.1. (i) Let $g(\cdot)$ be an integrable function such that $\int \|x\|^i |g(x)|^i dF(x) < \infty$ for $i = 1, 2$ and $\int xg(x)dF(x) = 0$, where $F(x)$ is the cumulative distribution function of $\{x_t\}$ and $\|\cdot\|$ denotes the conventional Euclidean norm.

(ii) For any vector γ , $\min_{\gamma} E [g(x_1) - x_1^{\tau} \gamma]^2 > 0$.

Note that Assumption 2.1(i) implies both the identifiability and the ‘smallness’ conditions on $g(\cdot)$. Assumption 2.1(ii) is imposed to exclude any cases where $g(x)$ is a linear function of x . Under Assumption 2.1, parameter β is identifiable and chosen such that

$$E [y_t - x_t^{\tau} \beta]^2 \text{ is minimised over } \beta, \quad (2.1)$$

which implies $\beta = (E [x_1 x_1^{\tau}])^{-1} E [x_1 y_1]$ provided that the inverse matrix does exist. Note that the definition of $\beta = (E [x_1 x_1^{\tau}])^{-1} E [x_1 y_1]$ implies $\int x g(x) dF(x) = 0$, and vice versa. As a consequence, β may be estimated by the ordinary least squares estimator of the form

$$\hat{\beta} = \left(\sum_{t=1}^n x_t x_t^{\tau} \right)^{-1} \left(\sum_{t=1}^n x_t y_t \right). \quad (2.2)$$

Gao (1992) then establishes an asymptotic theory for $\hat{\beta}$ and a nonparametric estimator of $g(\cdot)$ of the form

$$\hat{g}(x) = \sum_{t=1}^n w_{nt}(x) (y_t - x_t^{\tau} \hat{\beta}), \quad (2.3)$$

where $w_{nt}(x)$ is a probability weight function and is commonly chosen as $w_{nt}(x) = \frac{K(\frac{x_t - x}{h})}{\sum_{s=1}^n K(\frac{x_s - x}{h})}$, in which $K(\cdot)$ and h are the probability kernel function and the bandwidth parameter, respectively.

As a result of such an estimation procedure, one may be able to determine whether $g(\cdot)$ is small enough to be negligible. A further testing procedure may be used to test whether the null hypothesis $H_0 : g(\cdot) = 0$ may be accepted. Gao (1995) proposes a simple test and then shows that under H_0 ,

$$\hat{L}_{1n} = \frac{\sqrt{n}}{\hat{\sigma}_1} \left(\frac{1}{n} \sum_{t=1}^n (y_t - x_t^{\tau} \hat{\beta})^2 - \hat{\sigma}_0^2 \right) \rightarrow_D N(0, 1), \quad (2.4)$$

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.

In recent years, model (1.1) has been commonly used as a semiparametric alternative to a simple parametric linear model when there is not enough evidence to suggest accepting the simple linear model. In such cases, interest is mainly on establishing an asymptotic distribution of the test statistic under the null hypothesis. Alternative models are mainly used in small sample simulation studies when evaluating the power performance of the proposed test. There are some exceptions that further interest is in estimating the $g(\cdot)$ function involved before establishing a closed-form expression of the power function and then studying its large-sample and small-sample properties (see, for example, Gao 2007; Gao and Gijbels 2008). Even in such cases, estimation of $g(\cdot)$ becomes a secondary issue. Therefore, there has been no primary need to rigorously deal with such an estimation issue under suitable identifiability conditions similar to Assumption 2.1.

To state some general results for $\hat{\beta}$ and $\hat{g}(\cdot)$, we introduce the following conditions.

Assumption 2.2. (i) Let (x_t, e_t) be a vector of stationary and α -mixing time series with mixing coefficient $\alpha(k)$ satisfying $\sum_{k=1}^{\infty} \alpha^{\frac{\delta}{2+\delta}}(k) < \infty$ for some $\delta > 0$, where $\delta > 0$ is chosen such that $E \left[|x_1 \varepsilon_1|^{2+\delta} \right] < \infty$, in which $\varepsilon_t = e_t + g(x_t)$.

(ii) Let $E[e_1|x_1] = 0$ and $E[e_1^2|x_1] = \sigma_e^2 < \infty$ almost surely. Let also $\Sigma_{11} = E[x_1 x_1^T]$ be a positive definite matrix.

(iii) Let $p(x)$ be the marginal density of x_1 . The first derivative of $p(x)$ is continuous in x .

(iv) The probability kernel function $K(\cdot)$ is a continuous and symmetric function with compact support.

(v) The bandwidth h satisfies $\lim_{n \rightarrow \infty} h = 0$, $\lim_{n \rightarrow \infty} nh^d = \infty$ and $\limsup_{n \rightarrow \infty} nh^{d+4} < \infty$.

Assumption 2.2 is a set of conditions similar to what has been used in the literature (such as, Gao 2007; Li and Racine 2007; Gao and Gijbels 2008). As a consequence, its suitability may be verified similarly.

We now state the following proposition.

Proposition 2.1 (i) *Let Assumptions 2.1 and 2.2 hold. Then as $n \rightarrow \infty$*

$$\sqrt{n} \left(\hat{\beta} - \beta \right) \rightarrow_D N \left(0, \sigma_{1\varepsilon}^2 \Sigma_{11}^{-2} \right), \quad (2.5)$$

where $\sigma_{1\varepsilon}^2 = E[x_1 x_1^T \varepsilon_1^2] + 2 \sum_{t=2}^{\infty} E[\varepsilon_1 \varepsilon_t x_1 x_t^T]$.

(ii) *If, in addition, the first two derivatives of $g(x)$ are continuous, we have as $n \rightarrow \infty$*

$$\sqrt{nh^d} \left(\hat{g}(x) - g(x) - c_n \right) \rightarrow_D N \left(0, \sigma_g^2(x) \right) \quad (2.6)$$

at such x that $p(x) > 0$, where $c_n = \frac{h^2(1+o(1))}{2} \left(g''(x) + \frac{2g'(x)p'(x)}{p(x)} \right) \int u^\tau u K(u) du$ and $\sigma_g^2(x) = \frac{\int K^2(u) du}{p(x)}$, in which $p(x)$ is the marginal density of x_1 .

The proof of Proposition 2.1 is relatively straightforward using existing results for central limit theorems for partial sums of stationary and α -mixing time series (see, for example, Fan and Yao 2003). Obviously, one may use a local-linear kernel weight function to replace $w_{nt}(x)$ in order to correct the bias term involved in c_n . Since such details are not essential to the primary interest of the discussion of this kind of problem, we there omit such details here.

Furthermore, in a recent paper by Chen, Gao and Li (2011), the authors consider an extended case of model (2.1) of the form

$$y_t = f(x_t^T \beta) + g(x_t) + e_t \quad \text{with} \quad x_t = \lambda_t + u_t, \quad (2.7)$$

where $f(\cdot)$ is parametrically known, $\{\lambda_t\}$ is an unknown deterministic function of t , and $\{u_t\}$ is a sequence of independent errors. In addition, $g(\cdot)$ is allowed to be a sequence of functions of the form $g_n(\cdot)$ in order to directly link model (2.7) with a sequence of local alternative functions under an alternative hypothesis as has been widely discussed in the literature (see, for example,

Gao 2007; Gao and Gijbels 2008). By the way, the finite-sample results presented in Chen, Gao and Li (2011) further confirm that the pair $(\hat{\beta}, \hat{g}(\cdot))$ has better performance than a semiparametric weighted least squares (SWLS) estimation method proposed for model (1.2), since the so-called “SWLS” estimation method, as pointed out before, is not theoretically sound for model (1.1). Obviously, there are certain limitations with the paper by Chen, Gao and Li (2011) and further discussion may be needed to fully take issues related to endogeneity and stationarity into account.

As also briefly mentioned in the introduction, model (1.1) may be motivated as a model to address a kind of ‘weak’ endogenous problem. Consider a simple linear model of the form

$$y_t = x_t^\tau \beta + \varepsilon_t \quad \text{with} \quad E[\varepsilon_t | x_t] \neq 0, \quad (2.8)$$

where $\{\varepsilon_t\}$ is a sequence of stationary errors.

Let $g(x) = E[\varepsilon_t | x_t = x]$. Since $\{\varepsilon_t\}$ is unobservable, it may not be unreasonable to assume that the functional form of $g(\cdot)$ is unknown. Meanwhile, empirical evidence broadly supports either full linearity or semi-linearity. It is therefore that one may assume that $g(\cdot)$ is kind of ‘small’ function satisfying Assumption 2.1. Let $e_t = \varepsilon_t - E[\varepsilon_t | x_t]$. In this case, model (2.8) can be rewritten as model (1.1) with $E[e_t | x_t] = 0$. In this case, $g(x_t)$ may be used as an ‘instrumental variable’ to address a ‘weak’ endogeneity problem involved in model (2.8). As a consequence, β can be consistently estimated by $\hat{\beta}$ under Assumption 2.1 and the so-called “instrumental variable” $g(x_t)$ may be asymptotically ‘found’ by $\hat{g}(x_t)$.

2.2 Case of $d \geq 4$

As discussed in the literature (see, for example, Chapter 7 of Fan and Gijbels 1996; Chapter 2 of Gao 2007), one may need to encounter the so-called “problem of the curse of dimensionality” when estimating high dimensional (with the dimensionality $d \geq 4$) functions. We therefore propose using a semiparametric single-index model of the form

$$y_t = x_t^\tau \beta + g(x_t^\tau \beta) + e_t \quad (2.9)$$

as an alternative to model (1.1). To be able to identify and estimate model (2.9), Assumption 2.1 will need to be modified as follows.

Assumption 2.3. (i) Let $g(\cdot)$ be an integrable function such that $\int \|x\|^i |g(x^\tau \beta_0)|^i dF(x) < \infty$ for $i = 1, 2$ and $\int x g(x^\tau \beta_0) dF(x) = 0$, where β_0 is the true value of β and $F(x)$ is the cumulative distribution function of $\{x_t\}$.

(ii) For any vector γ , $\min_\gamma E[g(x_1^\tau \beta_0) - x_1^\tau \gamma]^2 > 0$.

Under Assumption 2.2, β is identifiable and estimable by $\hat{\beta}$. The conclusions of Proposition 2.1 still remain valid except the fact that $\hat{g}(\cdot)$ is now modified as

$$\hat{g}(u) = \frac{\sum_{t=1}^n K\left(\frac{x_t^\tau \hat{\beta} - u}{h}\right) y_t}{\sum_{s=1}^n K\left(\frac{x_s^\tau \hat{\beta} - u}{h}\right)}. \quad (2.10)$$

We think that model (2.9) is probably the most feasible and easily implementable alternative to model (1.1), although there are some other alternatives. One of them is a semiparametric single-index model of the form

$$y_t = x_t^\tau \beta + g(x_t^\tau \gamma) + e_t, \quad (2.11)$$

where γ is another vector of unknown parameters. As discussed in Xia, Tong and Li (1999), model (2.11) is a better alternative to model (1.2) than to model (1.1). Another of them is a semiparametric additive model of the form

$$y_t = x_t^\tau \beta + \sum_{j=1}^d g_j(x_{tj}) + e_t, \quad (2.12)$$

where each $g_j(\cdot)$ is an unknown and univariate function. In this case, Assumption 2.1 may be replaced by Assumption 2.4 below.

Assumption 2.4. (i) Let each $g_j(\cdot)$ satisfy $\max_{1 \leq j \leq d} \int \|x\|^i |g_j(x_j)|^i dF(x) < \infty$ for $i = 1, 2$ and $\sum_{j=1}^d \int x g_j(x_j) dF(x) = 0$, where each x_j is the j -th component of $x = (x_1, \dots, x_j, \dots, x_d)^\tau$ and $F(x)$ is the cumulative distribution function of $\{x_t\}$.

(ii) For any vector γ , $\min_\gamma E \left[\sum_{j=1}^d g_j(x_{tj}) - x_t^\tau \gamma \right]^2 > 0$, where each x_{tj} is the j -th component of $x_t = (x_{t1}, \dots, x_{tj}, \dots, x_{td})^\tau$.

Under Assumption 2.4, β is still identifiable and estimable by $\hat{\beta}$. The estimation of $\{g_j(\cdot)\}$ however involves an additive estimation method, such as the marginal integration method discussed in Chapter 2 of Gao (2007). Under Assumptions 2.2 and 2.4 as well as some additional conditions, asymptotic properties may be established for the resulting estimators of $g_j(\cdot)$ in a similar way to Section 2.3 of Gao (2007).

We have so far discussed some issues for the case where $\{x_t\}$ is stationary. In order to establish an asymptotic theory in each individual case, various conditions may be imposed on the probabilistic structure $\{e_t\}$. Both our own experience and the literature show that it is relatively straightforward to establish an asymptotic theory for $\hat{\beta}$ and $\hat{g}(\cdot)$ under either the case where $\{e_t\}$ satisfies some martingale assumptions or the case where $\{e_t\}$ is a linear process. In Section 3 below, we provide some necessary conditions before we establish a new asymptotic theory for the case where $\{x_t\}$ is a sequence of nonstationary regressors.

3. Nonstationary models

This section focuses on the case where $\{x_t\}$ is stochastically nonstationary. Since the paper by Chen, Gao and Li (2011) already discusses the case where nonstationarity is driven by a deterministic trending component, this section focuses on the case where the nonstationarity of $\{x_t\}$ is driven by a stochastic trending component. Due to the limitation of existing theory, we only discuss the case of $d = 1$ in the nonstationary case.

Before our discussion, we introduce some necessary conditions.

Assumption 3.1. (i) Let $g(\cdot)$ be a real function on $R^1 = (-\infty, \infty)$ such that $\int |x|^i |g(x)|^i dx < \infty$ for $i = 1, 2$ and $\int xg(x)dx \neq 0$.

(ii) In addition, let $g(\cdot)$ satisfy $\int | \int e^{ixy} yg(y)dy | dx < \infty$ when $\int xg(x)dx = 0$.

In comparison with Assumption 2.1, there is no need to impose a condition similar to Assumption 2.1(ii), since Assumption 3.1 itself already excludes the case where $g(x)$ is a simple linear function of x .

In addition to Assumption 3.1, we will need the following conditions.

Assumption 3.2. (i) Let $x_t = x_{t-1} + u_t$ with $x_0 = 0$ and $u_t = \sum_{i=0}^{\infty} \psi_i \eta_{t-i}$, where $\{\eta_t\}$ is a sequence of independent and identically distributed random errors with $E[\eta_1] = 0$, $0 < E[\eta_1^2] = \sigma_\eta^2 < \infty$ and $E[|\eta_1|^{4+\delta}] < \infty$ for some $\delta > 0$, in which $\{\psi_i : i \geq 0\}$ is a sequence of real numbers such that $\sum_{i=0}^{\infty} i^2 |\psi_i| < \infty$ and $\sum_{i=0}^{\infty} \psi_i \neq 0$. Let $\varphi(\cdot)$ be the characteristic function of η_1 satisfying $|r|\varphi(r) \rightarrow 0$ as $r \rightarrow \infty$.

(ii) Suppose that $\{(e_t, \mathcal{F}_t) : t \geq 1\}$ is a sequence of martingale differences satisfying $E[e_t^2 | \mathcal{F}_{t-1}] = \sigma_e^2 > 0$, a.s., and $E[e_t^4 | \mathcal{F}_{t-1}] < \infty$ a.s. for all $t \geq 1$. Let $\{x_t\}$ be adapted to \mathcal{F}_{t-1} for $t = 1, 2, \dots, n$.

(iii) Let $E_n(r) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} e_t$ and $U_n(r) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} u_t$. There is a vector Brownian motion (E, U) such that $(E_n(r), U_n(r)) \Rightarrow_{\mathcal{D}} (E(r), U(r))$ on $D[0, 1]^2$ as $n \rightarrow \infty$, where $\Rightarrow_{\mathcal{D}}$ stands for the weak convergence.

(iv) The probability kernel function $K(\cdot)$ is a bounded and symmetric function. In addition, there is a real function $\Delta(x, y)$ such that, when h is small enough, $|g(x + hy) - g(x)| \leq h \Delta(x, y)$ for all y and $\int K(y)\Delta(x, y)dy < \infty$ for each given x .

(v) The bandwidth h satisfies $h \rightarrow 0$, $nh^2 \rightarrow \infty$ and $nh^6 \rightarrow 0$ as $n \rightarrow \infty$.

Similar sets of conditions have been used in Gao and Phillips (2011), Li *et al* (2011), and Chen, Gao and Li (2012). The verification and suitability of Assumption 3.2 may be given in a similar way to Remark A.1 of Appendix A of Li *et al* (2011).

Since $\{x_t\}$ is nonstationary, we replace equation (2.1) by a sample version of the form

$$\frac{1}{n} \sum_{t=1}^n [y_t - x_t \beta]^2 \text{ is minimised over } \beta, \quad (3.1)$$

which implies $\hat{\beta} = (\sum_{t=1}^n x_t^2)^{-1} (\sum_{t=1}^n x_t y_t)$ as has been given in equation (2.2). A simple expression implies

$$n(\hat{\beta} - \beta) = \left(\frac{1}{n^2} \sum_{t=1}^n x_t^2 \right)^{-1} \left(\frac{1}{n} \sum_{t=1}^n x_t e_t \right) + \left(\frac{1}{n^2} \sum_{t=1}^n x_t^2 \right)^{-1} \left(\frac{1}{n} \sum_{t=1}^n x_t g(x_t) \right). \quad (3.2)$$

Straightforward derivations imply as $n \rightarrow \infty$

$$\frac{1}{n^2} \sum_{t=1}^n x_t^2 = \frac{1}{n} \sum_{t=1}^n x_{tn}^2 \Rightarrow_{\mathcal{D}} \int_0^1 U^2(r) dr, \quad (3.3)$$

$$\frac{1}{n} \sum_{t=1}^n x_t e_t = \frac{1}{\sqrt{n}} \sum_{t=1}^n x_{tn} e_t \Longrightarrow_{\mathcal{D}} \int_0^1 U(r) dE(r), \quad (3.4)$$

where $x_{tn} = \frac{x_t}{\sqrt{n}}$.

In view of equations (3.2)–(3.4), in order to establish an asymptotic distribution for $\widehat{\beta}$, it is expected to show that as $n \rightarrow \infty$

$$\frac{1}{n} \sum_{t=1}^n x_t g(x_t) \rightarrow_P 0. \quad (3.5)$$

To be able to show (3.5), we need to consider the case of $\int xg(x)dx = 0$ and the case of $\int xg(x)dx \neq 0$ separately. In the case of $\int xg(x)dx \neq 0$, existing results (such as, Theorem 2.1 of Wang and Phillips 2009) imply as $n \rightarrow \infty$

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n x_t g(x_t) = \frac{d_n}{n} \sum_{t=1}^n (d_n x_{tn}) g(d_n x_{tn}) \rightarrow_D L_U(1, 0) \cdot \int_{-\infty}^{\infty} z g(z) dz, \quad (3.6)$$

where $d_n = \sqrt{n}$ and $L_U(1, 0)$ is the local-time process associated with $U(r)$. This then implies as $n \rightarrow \infty$

$$\frac{1}{n} \sum_{t=1}^n x_t g(x_t) = \frac{1}{\sqrt{n}} \cdot \frac{1}{\sqrt{n}} \sum_{t=1}^n x_t g(x_t) \rightarrow_P 0. \quad (3.7)$$

In the case of $\int xg(x)dx = 0$, existing results (such as, Theorem 2.1 of Wang and Phillips 2011) also imply as $n \rightarrow \infty$

$$\sqrt{\frac{1}{\sqrt{n}}} \sum_{t=1}^n x_t g(x_t) = \sqrt{\frac{d_n}{n}} \sum_{t=1}^n (d_n x_{tn}) g(d_n x_{tn}) \rightarrow_D \sqrt{L_U(1, 0)} \cdot N(0, 1) \cdot \sqrt{\int_{-\infty}^{\infty} z^2 g^2(z) dz}, \quad (3.8)$$

where $N(0, 1)$ is a standard normal random variable independent of $L_U(1, 0)$. This shows that equation (3.5) is also valid for the case of $\int xg(x) = 0$.

We therefore summarise the above discussion into the following proposition.

Proposition 3.1 (i) *Let Assumptions 3.1 and 3.2(i)–(iii) hold. Then as $n \rightarrow \infty$*

$$n \left(\widehat{\beta} - \beta \right) \Longrightarrow_{\mathcal{D}} \left(\int_0^1 U^2(r) dr \right)^{-1} \int_0^1 U(r) dE(r). \quad (3.9)$$

(ii) *If, in addition, Assumption 3.2(iv)(v) holds, then as $n \rightarrow \infty$*

$$\sqrt{\sum_{t=1}^n K \left(\frac{x_t - x}{h} \right)} \left(\widehat{g}(x) - g(x) \right) \rightarrow_D N \left(0, \sigma_{1g}^2 \right), \quad (3.10)$$

where $\sigma_{1g}^2 = \sigma_e^2 \int K^2(u) du$.

The proof of (3.9) follows from equations (3.2)–(3.8). To show (3.10), one may be seen that

$$\widehat{g}(x) - g(x) = \sum_{t=1}^n w_{nt}(x) e_t + \sum_{t=1}^n w_{nt}(x) (g(x_t) - g(x)) + \sum_{t=1}^n w_{nt}(x) x_t \left(\beta - \widehat{\beta} \right). \quad (3.11)$$

The first two terms may be dealt with in the same way as in existing studies (such as the proof of Theorem 3.1 of Wang and Phillips 2009). To deal with the third term, one may have the following derivations:

$$\sum_{t=1}^n w_{nt}(x)x_t = h \cdot \frac{\sum_{t=1}^n K\left(\frac{x_t-x}{h}\right)\left(\frac{x_t-x}{h}\right)}{\sum_{t=1}^n K\left(\frac{x_t-x}{h}\right)} + x = o_P(1) \quad (3.12)$$

by the fact that $\int uK(u)du = 0$ and an application of Theorem 2.1 of Wang and Phillips (2011). Equations (3.11) and (3.12), along with (3.9), complete the proof of (3.10).

Meanwhile, as in the stationary case, model (1.1) can also be considered as an alternative model to a simple linear model of the form $y_t = x_t^\tau \beta + e_t$ in the nonstationary case. A nonparametric test of the form

$$\widehat{L}_{2n} = \frac{\sum_{t=1}^n \sum_{s=1, \neq t}^n \widehat{e}_s K\left(\frac{x_t-x_s}{h}\right) \widehat{e}_t}{\sqrt{2 \sum_{t=1}^n \sum_{s=1, \neq t}^n \widehat{e}_s^2 K^2\left(\frac{x_t-x_s}{h}\right) \widehat{e}_t^2}} \quad (3.13)$$

has been proposed to test $H_0 : P(g(x_t) = 0) = 1$ and studied in recent years (see, for example, Gao *et al* 2009a; Li *et al* 2011; Wang and Phillips 2012), where $\widehat{e}_t = y_t - x_t^\tau \widehat{\beta}$, in which $\widehat{\beta}$ is the ordinary least squares estimator based on model (1.1) under H_0 . Obviously, Assumption 2.1 is no longer needed for this kind of testing problem.

This section has so far considered the case where $\{x_t\}$ is an integrated time series. In Section 4 below, we consider an autoregressive version of model (1.1) and then discuss stationary and nonstationary cases separately.

4. Nonlinear autoregressive models

Consider an autoregressive version of model (1.1) of the form

$$y_t = x_t^\tau \beta + g(x_t) + e_t, \quad (4.1)$$

where $x_t = (y_{t-1}, \dots, y_{t-d})^\tau$, and the others are the same as before.

As has been discussed in the literature (see, for example, Tong 1990; Masry and Tjøstheim 1995; Chapter 6 of Härdle, Liang and Gao 2000), $\{y_t\}$ can be stochastically stationary and α -mixing when β satisfies Assumption 4.1(i) below and $g(\cdot)$ satisfies Assumption 4.1(ii) below.

Assumption 4.1. (i) Let $\beta = (\beta_1, \dots, \beta_d)^\tau$ satisfy $y^d - \beta_1 y^{d-1} - \dots - \beta_{d-1} y - \beta_d \neq 0$ for any $|y| \geq 1$.

(ii) Let $g(x)$ be bounded on any bounded Borel measurable set and satisfy $g(x) = o(\|x\|)$ as $\|x\| \rightarrow \infty$.

Under Assumption 4.1, $\{y_t\}$ is stationary and α -mixing. This, along with Assumption 2.1, implies that the estimation of β and $g(\cdot)$ may be done in the same way as in Section 2.1 for the case of $1 \leq d \leq 3$ and in Section 2.2 for the case of $d \geq 4$. Therefore, discussion of model (4.1) is relatively straightforward.

In the rest of this section, we then focus on the case where $\{y_t\}$ is nonstationary and discuss about how to estimate β and $g(\cdot)$ consistently. To present the main ideas of our discussion, we focus on the case of $d = 1$ to imply a semiparametric autoregressive model of the form

$$y_t = \beta y_{t-1} + g(y_{t-1}) + e_t. \quad (4.2)$$

While model (4.2) might look too simple, as discussed below, the study of the nonstationarity of $\{y_t\}$ may not be so easy at all. This is mainly because the nonstationarity may be driven by either the case of $\beta = 1$, or, the case where the functional form of $g(\cdot)$ may be too ‘explosive’, or a mixture of both. Our interest of this section is to focus on the case where $g(\cdot)$ is a ‘small’ departure function and the true value of β is $\beta = 1$. In a recent paper by Gao, Tjøstheim and Yin (2012), the authors discuss a threshold version of model (4.2), in which $g(\cdot)$ is being treated as a conventional unknown function (not necessarily a ‘small’ function) defined on a compact subset.

In a similar fashion to (3.1), we estimate β by minimising

$$\frac{1}{n} \sum_{t=1}^n [y_t - y_{t-1}\beta]^2 \quad \text{over } \beta, \quad (4.3)$$

which implies $\hat{\beta} = (\sum_{t=1}^n y_{t-1}^2)^{-1} (\sum_{t=1}^n y_{t-1} y_t)$. The unknown departure function $g(\cdot)$ can then be estimated by

$$\hat{g}(y) = \sum_{t=1}^n w_{nt}(y) (y_t - \hat{\beta} y_{t-1}) \quad \text{with } w_{nt}(y) = \frac{K\left(\frac{y_{t-1}-y}{h}\right)}{\sum_{s=1}^n K\left(\frac{y_{s-1}-y}{h}\right)}. \quad (4.4)$$

When $\beta = 1$, we have

$$\hat{\beta} - 1 = \left(\sum_{t=1}^n y_{t-1}^2 \right)^{-1} \left(\sum_{t=1}^n y_{t-1} e_t \right) + \left(\sum_{t=1}^n y_{t-1}^2 \right)^{-1} \left(\sum_{t=1}^n y_{t-1} g(y_{t-1}) \right). \quad (4.5)$$

In order to establish an asymptotic distribution for $\hat{\beta}$, we will need to understand the probabilistic structure of $\{y_t\}$. Obviously, $\{y_t\}$ is not integrated unless $g(\cdot) \equiv 0$. Thus, existing theory for the integrated time series case is not applicable here. We therefore impose some specific conditions on $g(\cdot)$ and $\{e_t\}$ to ensure that certain probabilistic structure can be deduced for $\{y_t\}$.

Assumption 4.2. (i) Let $\{e_t\}$ be a sequence of independent and identically distributed continuous random errors with $E[e_1] = 0$ and $0 < E[e_1^2] = \sigma_e^2 < \infty$. Let $\{e_t\}$ and $\{y_s\}$ be independent for all $s < t$. In addition, the probability density, $p(x)$, of e_1 satisfies $\inf_{x \in C_p} p(x) > 0$ for all compact sets C_p .

(ii) Let $g(y)$ be twice differentiable and the second derivative of $g(y)$ be continuous in $y \in R^1 = (-\infty, \infty)$. In addition, $\int |g(y)|^i \pi_s(dy) < \infty$ for $i = 1, 2$, where $\pi_s(\cdot)$ is the invariant measure of $\{y_t\}$.

(iii) Furthermore, $\int |yg(y)|^i \pi_s(dy) < \infty$ for $i = 1, 2$.

Assumption 4.2(i) is needed to show that $\{y_t\}$ can be a λ -null recurrent Markov chain with $\lambda = \frac{1}{2}$. Assumption 4.2(ii) is required to restrict that the functional form of $g(\cdot)$ is not too ‘explosive’ in a similar fashion to Assumption 4.1(ii). If the functional form of $g(\cdot)$ is too ‘explosive’ in this case, the nonstationarity of $\{y_t\}$ may be too strong to be controllable. Assumption 4.2(iii) imposes additional integrability conditions on $yg(y)$ in a way similar to Assumption 3.1(i) for the integrated case. Note that there is no need to require $\int yg(y)\pi_s(dy) = 0$ and also we need not consider the case of $\int yg(y)\pi_s(dy) = 0$ specifically as in Section 3.

In order to establish an asymptotic theory for $(\hat{\beta}, \hat{g}(\cdot))$, we need to introduce the following proposition.

Proposition 4.1. *Let Assumption 4.2(i)(ii) hold. Then $\{y_t\}$ is a λ -null recurrent Markov chain with $\lambda = \frac{1}{2}$.*

The proof of Proposition 4.1 follows similarly from that of Lemma 3.1 of Gao, Tjøstheim and Yin (2012). More details about null recurrent Markov chains are available in Karlsen and Tjøstheim (2001) and Appendix A of Gao, Tjøstheim and Yin (2012). Proposition 4.1 shows that $\{y_t\}$ is a nonstationary Markov chain, although it cannot be an integrated time series when $g(\cdot) \neq 0$. As a consequence, one may establish the following asymptotic theory in Proposition 4.2 below.

Proposition 4.2. (i) *Let Assumption 4.2 hold. Then as $n \rightarrow \infty$*

$$n(\hat{\beta} - 1) \xrightarrow{\mathcal{D}} \frac{(Q^2(1) - \sigma_e^2)}{2 \int_0^1 Q^2(r) dr}, \quad (4.6)$$

where $Q(r) = \sigma_e B(r) + M_{\frac{1}{2}}(r)\mu_g$, in which $B(r)$ is the conventional Brownian motion, $M_{\frac{1}{2}}(t)$ is the Mittag-Leffler process as defined in Karlsen and Tjøstheim (2001, p 388) and $\mu_g = \int g(y)\pi_s(dy)$.

(ii) *If, in addition, Assumption 3.2(iv)(v) holds, then as $n \rightarrow \infty$*

$$\sqrt{\sum_{t=1}^n K\left(\frac{y_{t-1} - y}{h}\right)} (\hat{g}(y) - g(y)) \rightarrow_D N(0, \sigma_{2g}^2), \quad (4.7)$$

where $\sigma_{2g}^2 = \sigma_e^2 \int K^2(u) du$.

The proof of Proposition 4.2 is given in Appendix A below. Note that Proposition 4.2 shows that the super rate- n of convergence is still achievable for $\hat{\beta}$ even when $\{y_t\}$ is not an integrated time series. In addition, $Q(r) = \sigma_e B(r)$ when $\mu_g = 0$. In other words, $\hat{\beta}$ retains the same asymptotic behaviour as if $\{y_t\}$ were integrated when the ‘small’ departure function $g(\cdot)$ satisfies $\int g(y)\pi_s(dy) = 0$. Meanwhile, the asymptotic theory of $\hat{g}(\cdot)$ remains the same as in the integrated case (see, for example, Proposition 3.1(ii)).

Remark 4.1. (i) While Assumptions 4.1 and 4.2 are assumed respectively for the stationary and nonstationary cases, there are some common features in both assumptions. To present the main ideas in this discussion, we focus on the case of $d = 1$ in Assumption 4.1(i). When $|\beta| < 1$,

Assumption 4.1(ii) basically requires that the rate of $g(y)$ decaying to infinity is slower than that of $|y| \rightarrow \infty$ in order to ensure that $\{y_t\}$ is stochastically stationary. In the case of $\beta = 1$, in addition to the ‘smallness’ condition in Assumption 4.2(iii), Assumption 4.2(ii) also imposes certain conditions on the rate of divergence of $g(\cdot)$ to deduce that $\{y_t\}$ is a nonstationary Markov chain, although, in the case of $g(\cdot) \neq 0$, $\{y_t\}$ is not an integrated time series. This is mainly because it may be impossible to study such nonlinear autoregressive models when $g(\cdot)$ behaves too ‘explosive’.

(ii) $\{y_t\}$ could be generated recursively by a nonlinear autoregressive time series of the form $y_t = y_{t-1} + g(y_{t-1}) + e_t$ if $\beta = 1$ and $g(\cdot)$ were known. In the paper by Granger, Inoue and Morin (1997), the authors propose some parametric specifications for $g(\cdot)$ and treat $g(\cdot)$ as a stochastic trending component. The authors then suggest estimating $g(\cdot)$ nonparametrically before checking whether $g(\cdot)$ is negligible. Gao *et al* (2009b) further consider this model and propose a nonparametric unit-root test for testing $H_0 : g(\cdot) = 0$. As pointed out above, what we have been interested is to deal with the case where $g(\cdot)$ is not negligible, but is a ‘small’ departure function. Proposition 4.2 implies that model (4.2) may generate a class of “nearly integrated” time series models when $\beta = 1$ and $\int g(y)\pi_s(dy) = 0$. This may motivate us to further develop some general theory for the so-called class of “nearly integrated” time series models.

5. Extensions and examples of implementation

Since many practical problems (see, for example, Examples 5.1 and 5.2 below) may require the inclusion of a general polynomial function as the main mean function of y_t , model (1.1) may need to be extended to accommodate a general class of parametric functions. In this case, model (1.1) can be written as

$$y_t = f(x_t, \beta) + g(x_t) + e_t, \quad (5.1)$$

where $f(x, \beta)$ is a parametrically known function indexed by a vector of unknown parameters β . In the stationary case, equation (2.1) now becomes

$$E [y_t - f(x_t, \beta)]^2 \quad \text{is minimised over } \beta. \quad (5.2)$$

In the integrated time series case, equation (3.1) can be replaced by minimising

$$\frac{1}{n} \sum_{t=1}^n [y_t - f(x_t, \beta)]^2 \quad \text{over } \beta, \quad (5.3)$$

which is similar to the discussion used in Park and Phillips (2001). Obviously, various other identifiability conditions imposed in Sections 2 and 3 can be modified straightforwardly. Thus, further discussion is omitted here.

In the autoregressive time series case, model (5.1) becomes

$$y_t = f(y_{t-1}, \beta) + g(y_{t-1}) + e_t, \quad (5.4)$$

and equation (4.3) is now replaced by minimising

$$\frac{1}{n} \sum_{t=1}^n [y_t - f(y_{t-1}, \beta)]^2 \quad \text{over } \beta. \quad (5.5)$$

In the threshold case where $f(y, \beta) = \beta y I[y \in C_\tau]$ with C_τ being a compact set indexed by parameter τ , Gao, Tjøstheim and Yin (2012) show that $\{y_t\}$ is a $\frac{1}{2}$ -null recurrent Markov chain under Assumption 4.2(i)(ii). In general, further discussion about model (5.4) is needed and therefore left for future research.

Examples 5.1–5.3 below show why the proposed models and estimation methods are relevant and how the proposed estimation methods may be implemented in practice.

Example 5.1. This data set consists of quarterly consumer price index (CPI) numbers of 11 classes of commodities for 8 Australian capital cities spanning from 1994 to 2008 (available from the Australian Bureau of Statistics at www.abs.gov.au). Figure 5.1 below gives the scatter plots of the log food CPI and the log all-group CPI.

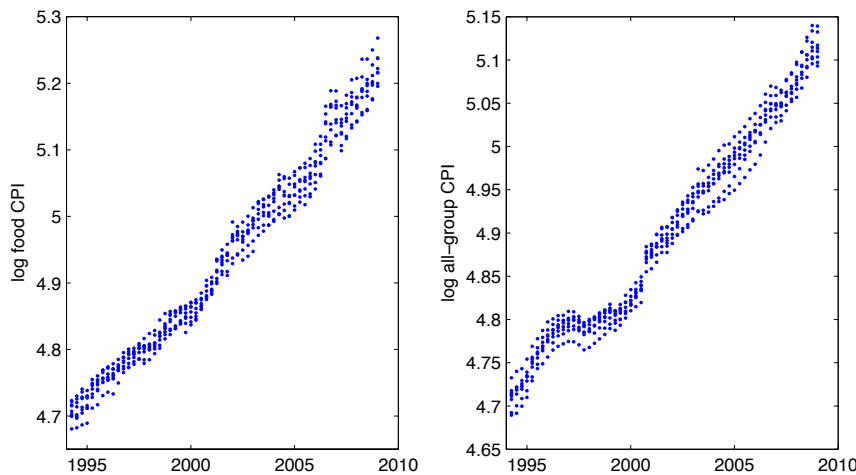


Figure 5.1. Scatter plots of the log food CPI and the log all-group CPI.

Figure 5.1 shows that either a simple linear trending function or a second-order polynomial form may be sufficient to capture the main trending behaviour for each of the CPI data sets. Similarly, many other data sets available in climatology, economics and finance also show that linearity remains the leading component of the trending component of the data under study. Figure 5.2 clearly shows that it is not unreasonable to assume a simple linear trending function for a disposable income data set (a quarter data set from the first quarter of 1960 to the last quarter of 2009 available from the Bureau of Economic Analysis at <http://www.bea.gov>).

The following example is the same as Example 5.2 of Li *et al* (2011). We use this example to show that in some empirical models, a second-order polynomial model is more accurate than a simple linear model.

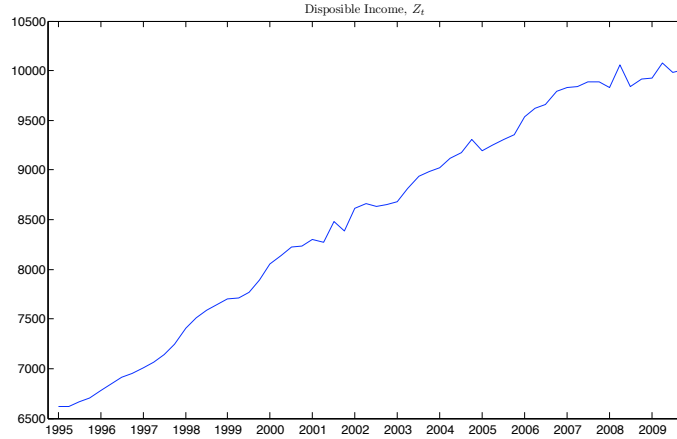


Figure 5.2. Plot of the disposable income data

Example 5.2. In this example, we consider the 2-year (x_{1t}) and 30-year (x_{2t}) Australian government bonds, which represent short-term and long-term series in the term structure of interest rates. Our aim is to analyze the relationship between the long-term data $\{x_{2t}\}$ and short-term data $\{x_{1t}\}$. We first apply the transformed versions defined by $y_t = \log(x_{2t})$ and $x_t = \log(x_{1t})$. The time frame of the study is during January 1971 to December 2000, with 360 observations for each of $\{y_t\}$ and $\{x_t\}$.

Consider the null hypothesis defined by

$$H_0 : y_t = \alpha_0 + \beta_0 x_t + \gamma_0 x_t^2 + e_t, \quad (5.6)$$

where $\{e_t\}$ is an unobserved error process.

In case there is any departure from the simple linear model, we propose using a nonparametric kernel estimate of the form

$$\hat{g}(x) = \sum_{t=1}^n w_{nt}(x) \left(y_t - \hat{\alpha}_0 - \hat{\beta}_0 x_t - \hat{\gamma}_0 x_t^2 \right), \quad (5.7)$$

where $\hat{\alpha}_0 = -0.2338$, $\hat{\beta}_0 = 1.4446$, $\hat{\gamma}_0 = -0.1374$, and $\{w_{nt}(x)\}$ is as defined in (3.2), in which $K(x) = \frac{3}{4}(1 - x^2)I\{|x| \leq 1\}$ and an optimal bandwidth \hat{h}_{optimal} is chosen by a cross-validation method.

Figure 5.3 shows that the relationship between y_t and x_t may be approximately modelled by a second-order polynomial function of the form $y = -0.2338 + 1.4446 x - 0.1374 x^2$.

The following example is the same as Example 4.5 of Gao, Tjøstheim and Yin (2012). We use it here to show that a parametric version of model (4.2) is a valid alternative to a conventional integrated time series model in this case.

Example 5.3. We look at the logarithm of the British pound/American dollar real exchange rate, y_t , defined as $\log(e_t) + \log(p_t^{UK}) - \log(p_t^{USA})$, where $\{e_t\}$ is the monthly average of the

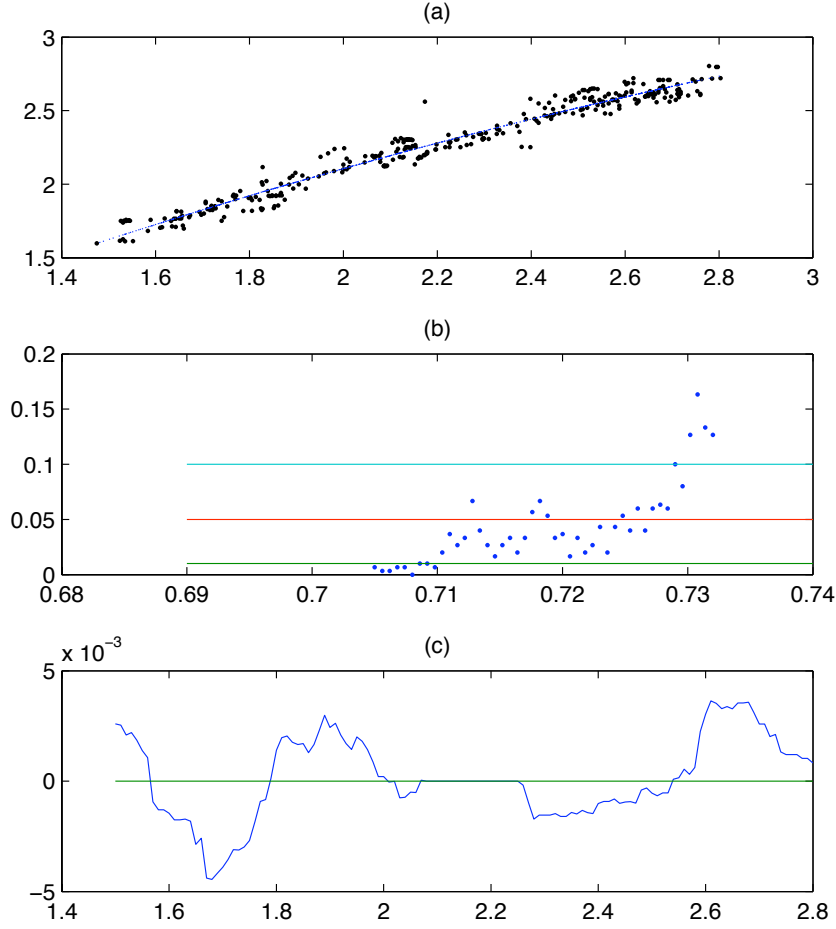


Figure 5.3: (a) provides the scatter chart of (y_t, x_t) and a nonparametric kernel regression plot $\hat{y} = \hat{m}(x)$; (b) gives p -values of the test for different bandwidths; and (c) is the plot of $\hat{g}(x)$ and its values are between -5×10^{-3} and 5×10^{-3} .

nominal exchange rate, and $\{p_t^i\}$ denotes the consumption price index of country i . These CPI data come from website: <http://www.rateinflation.com/> and the exchange rate data are available at <http://www.federalreserve.gov/>, spanning from January 1988 to February 2011, with sample size $n = 278$.

Our estimation method suggests that $\{y_t\}$ approximately follows a threshold model of the form

$$y_t = y_{t-1} - 1.1249 y_{t-1} I[|y_{t-1}| \leq 0.0134] + e_t. \quad (5.8)$$

Note that model (5.7) indicates that while $\{y_t\}$ does not necessarily follow an integrated time series model of the form $y_t = y_{t-1} + e_t$, $\{y_t\}$ behaves like a “nearly integrated” time series, because the nonlinear component $g(y) = -1.1249 y I[|y| \leq 0.0134]$ is a ‘small’ departure function with an upper bound of 0.0150.

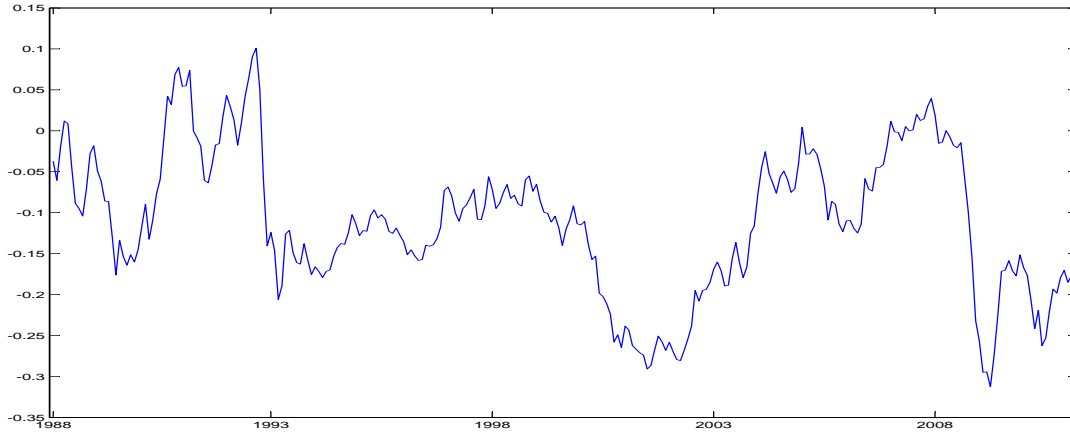


Figure 5.4. $y_t = \log(e_t) + \log(p_t^{UK}) - \log(p_t^{USA})$.

6. Conclusions and discussion

This paper has discussed a class of “nearly linear” models in Sections 2–4. Section 2 has summarised the history of model (1.1) and then explained why model (1.1) is important and has different theory to what has been commonly studied for model (1.2). Sections 3 and 4 have further explored such models to the nonstationary cases with the co-integrating case being discussed in Section 3 and the autoregressive case being discussed in Section 4. As shown in Sections 3 and 4, respectively, while the conventional “local-time” approach is applicable to establish the asymptotic theory in Proposition 3.1, one may need to develop the so-called “Markov chain” approach for the establishment of the asymptotic theory in Proposition 4.2.

As discussed in Remark 4.1, model (4.2) introduces a class of non-integrated but “nearly integrated” autoregressive time series models. Such a class of nonstationary models, along with a class of nonstationary threshold models proposed in Gao, Tjøstheim and Yin (2012), may provide existing literature with two new classes of nonlinear nonstationary models as alternatives to the class of integrated time series models already commonly and popularly studied in the literature. It is hoped that such models proposed in (4.2) and Gao, Tjøstheim and Yin (2012) along with the technologies developed could motivate us to develop some general classes of nonlinear and nonstationary autoregressive time series models.

7. Acknowledgments

The author acknowledges benefits from discussions and collaborations with his co-authors in the last few years, particularly with Jia Chen, Maxwell King, Degui Li, Peter Phillips and Dag Tjøstheim. The author also acknowledges useful comments and suggestions by Hua Liang. Thanks go to the Australian Research Council for its financial support under the Discovery Grants Scheme under Grant Number: DP1096374.

Appendix A

In order to help the reader of this paper, we introduce some necessary notation and useful lemmas for the proof of Proposition 4.2.

Let $\{y_t\}$ be a null-recurrent Markov chain. It is well-known that for a Markov chain on a countable state space which has a point of recurrence, a sequence split by the regeneration times becomes independent and identically distributed (i.i.d.) by the Markov property (see, for example, Chung 1967). For a general Markov process that does not have an obvious point of recurrence, as in Nummelin (1984), the Harris recurrence allows one to construct a split chain that decomposes the partial sum of the Markov process $\{y_t\}$ into blocks of i.i.d. parts and the negligible remaining parts.

Let z_t only take the values 0 and 1, and $\{(y_t, z_t), t \geq 0\}$ be the split chain. Define

$$\tau_k = \begin{cases} \inf\{t \geq 0 : z_t = 1\}, & k = 0, \\ \inf\{t > \tau_{k-1} : z_t = 1\}, & k \geq 1, \end{cases} \quad (\text{A.1})$$

and denote the total number of regenerations in the time interval $[0, n]$ by $T(n)$, that is,

$$T(n) = \begin{cases} \max\{k : \tau_k \leq n\}, & \text{if } \tau_0 \leq n, \\ 0, & \text{otherwise.} \end{cases} \quad (\text{A.2})$$

Note that $T(n)$ plays a central role in the proof of Proposition 4.2 below. While $T(n)$ is not observable, it may be replaced by $\frac{T_C(n)}{\pi_s(\mathbf{I}_C)}$ (see, for example, Lemma 3.6 of Karlsen and Tjøstheim 2001), where $T_C(n) = \sum_{t=1}^n I[y_t \in C]$, C is a compact set and \mathbf{I}_C is the conventional indicator function. In addition, Lemma 3.2 of Karlsen and Tjøstheim (2001) and Theorem 2.1 of Wang and Phillips (2009) imply that $T(n)$ is asymptotically equivalent to $\sqrt{n}L_B(1, 0)$, where $L_B(1, 0) = \lim_{\delta \rightarrow 0} \frac{1}{2\delta} \int_0^1 I[|B(s)| < \delta] ds$ is the local-time process of the Brownian motion $B(r)$.

We are now ready to establish some useful lemmas before the proof of Proposition 4.2. The proofs of Lemmas A.1 and A.2 below follow similarly from those of Lemmas 2.2 and 2.3 of Gao, Tjøstheim and Yin (2012), respectively.

Lemma A.1 *Let Assumption 4.2(i)(ii) hold. Then as $n \rightarrow \infty$*

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} e_t + \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} g(y_{t-1}) \rightarrow_{\mathcal{D}} \sigma_e B(r) + M_{\frac{1}{2}}(r) \mu_g \equiv Q(r), \quad (\text{A.3})$$

where $M_{\frac{1}{2}}(r)$, μ_g and $Q(r)$ are the same as defined in Proposition 4.2.

Lemma A.2 *Let Assumption 4.2 hold. Then as $n \rightarrow \infty$*

$$\frac{1}{T(n)} \sum_{t=1}^n y_{t-1} g(y_{t-1}) \rightarrow_P \int_{-\infty}^{\infty} yg(y) \pi_s(dy), \quad (\text{A.4})$$

$$\frac{1}{n^2} \sum_{t=1}^n y_{t-1}^2 \rightarrow_{\mathcal{D}} \int_0^1 Q^2(r) dr, \quad (\text{A.5})$$

$$\frac{1}{n} \sum_{t=1}^n y_{t-1} e_t \rightarrow_{\mathcal{D}} \frac{1}{2} (Q^2(1) - \sigma_e^2). \quad (\text{A.6})$$

Proof of Proposition 4.2. The proof of the first part of Proposition 4.2 follows from Lemma A.2 and

$$\begin{aligned}
 n(\widehat{\beta} - 1) &= \left(\frac{1}{n^2} \sum_{t=1}^n y_{t-1}^2 \right)^{-1} \left(\frac{1}{n} \sum_{t=1}^n y_{t-1} e_t \right) \\
 &+ \left(\frac{1}{n^2} \sum_{t=1}^n y_{t-1}^2 \right)^{-1} \left(\frac{T(n)}{n} \frac{1}{T(n)} \sum_{t=1}^n y_{t-1} g(y_{t-1}) \right). \tag{A.7}
 \end{aligned}$$

The proof of the second part of Proposition 4.2 follows similarly from that of Proposition 3.1(ii).

References

- Chen, J., Gao, J. and Li, D. (2011). Estimation in semiparametric time series models (invited paper). *Statistics and Its Interface* **4**, 243–252.
- Chen, J., Gao, J. and Li, D. (2012). Estimation in semiparametric regression with nonstationarity. Forthcoming in *Bernoulli* **18**. Available at <http://www.bernoulli-society.org/index.php/publications/bernoulli-journal/bernoulli-journal-papers>.
- Chung, K. L. (1967). *Markov Chains with Stationary Transition Probabilities*. Springer–Verlag, 2nd Edition.
- Fan, J. and Gijbels, I. (1996). *Local Polynomial Modeling and Its Applications*. Chapman & Hall, London.
- Fan, J., Wu, Y. and Feng, Y. (2009). Local quasi–likelihood with a parametric guide. *Annals of Statistics* **37**, 4153–4183.
- Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: parametric and nonparametric methods*. Springer, New York.
- Gao, J. (1992) *Large Sample Theory in Semiparametric Regression*. Doctoral Thesis at the Graduate School of the University of Science and Technology of China, Hefei, China.
- Gao, J. (1995). Parametric test in a partially linear model. *Acta Mathematica Scientia* (English Edition) **15**, 1–10.
- Gao, J. (2007). *Nonlinear Time Series: semi– and non–parametric methods*. Chapman & Hall, London.
- Gao, J. and Gijbels, I. (2008). Bandwidth selection in nonparametric kernel testing. *Journal of the American Statistical Association* **484**, 1584–1594.
- Gao, J., King, M. L., Lu, D. and Tjøstheim, D. (2009a). Nonparametric specification testing for nonlinear time series with nonstationarity. *Econometric Theory* **25**, 1869–1892.
- Gao, J., King, M. L., Lu, D. and Tjøstheim, D. (2009b). Specification testing in nonlinear and nonstationary time series autoregression. *Annals of Statistics* **37**, 3893–3928.
- Gao, J. and Phillips, P. C. B. (2011). Semiparametric estimation in multivariate nonstationary time series models. Working paper available at <http://www.buseco.monash.edu.au/ebs/pubs/wpapers/2011/wp17-11.pdf>.

- Gao, J., Tjøstheim, D. and Yin, J. (2012). Estimation in threshold autoregressive models with a stationary and a unit root regime. Available at <http://www.buseco.monash.edu.au/ebs/pubs/wpapers/2011/wp21-11.pdf>. Forthcoming in *Journal of Econometrics* **170**.
- Glad, I. (1998). Parametrically guided nonparametric regression. *Scandinavian Journal of Statistics* **25**, 649–668.
- Granger, C. W. J., Inoue, T. and Morin, N. (1997). Nonlinear stochastic trends. *Journal of Econometrics* **81**, 65–92.
- Härdle, W., Liang, H. and Gao, J. (2000). *Partially Linear Models*. Springer Series in Economics and Statistics. Physica–Verlag, New York.
- Juhl, T. and Xiao, Z. (2005). Partially linear models with unit roots. *Econometric Theory* **21**, 877–906.
- Karlsen, H. A. and Tjøstheim, D. (2001). Nonparametric estimation in null recurrent time series. *Annals of Statistics* **29**, 372–416.
- Li, D., Gao, J., Chen, J. and Lin, Z. (2011). Nonparametric estimation and specification testing in nonlinear and nonstationary time series models. Working paper available at www.jitigao.com.
- Li, Q. and Racine, J. (2007). *Nonparametric Econometrics: theory and practice*. Princeton University Press, Princeton.
- Martins-Filho, C., Mishra, S. and Ullah, A. (2008). A class of improved parametrically guided nonparametric regression estimators. *Econometric Reviews* **27**, 542–573.
- Masry, E. and Tjøstheim, D. (1995). Nonparametric estimation and identification of nonlinear ARCH time series. *Econometric Theory* **11**, 258–289.
- Nummelin, E. (1984). *General Irreducible Markov Chains and Non-negative Operators*. Cambridge University Press.
- Owen, A. (1991). Empirical likelihood for linear models. *Annals of Statistics* **19**, 1725–1747.
- Park, J. Y. and Phillips, P. C. B. (2001). Nonlinear regressions with integrated time series. *Econometrica* **69**, 117–162.
- Robinson, P. M. (1988). Root–N–consistent semiparametric regression. *Econometrica* **56**, 931–964.
- Teräsvirta, T., Tjøstheim, D., Granger, C. W. J. (2010). *Modelling Nonlinear Economic Time Series*. Oxford University Press, Oxford.
- Tong, H. (1990). *Nonlinear Time Series: a Dynamical System Approach*. Oxford University Press, Oxford.
- Wang, Q. Y. and Phillips, P. C. B. (2009). Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory* **25**, 710–738.
- Wang, Q. Y. and Phillips, P. C. B. (2011). Asymptotic theory for zero energy functionals with nonparametric regression application. *Econometric Theory* **27**, 235–259.
- Wang, Q. Y. and Phillips, P. C. B. (2012). Specification testing for nonlinear cointegrating regression. Forthcoming in *Annals of Statistics* **40**.
- Xia, Y., Tong, H. and Li, W. K. (1999). On extended partially linear single–index models. *Biometrika* **86**, 831–842.