Estimation in Threshold Autoregressive Models with a Stationary and a Unit Root Regime

Jiti Gao, Dag Tjøstheim†[1] and Jiying Yin**

The University of Adelaide, Australia and The University of Bergen, Norway†*

Abstract

This paper treats estimation in a class of new nonlinear threshold autoregressive models with both a stationary and a unit root regime. Existing literature on nonstationary threshold models have basically focused on models where the nonstationarity can be removed by differencing and/or where the threshold variable is stationary. This is not the case for the process we consider, and nonstandard estimation problems are the result.

This paper proposes a parameter estimation method for such nonlinear threshold autoregressive models using the theory of null recurrent Markov chains. Under certain assumptions, we show that the ordinary least squares (OLS) estimators of the parameters involved are asymptotically consistent. Furthermore, it can be shown that the OLS estimator of the coefficient parameter involved in the stationary regime can still be asymptotically normal while the OLS estimator of the coefficient parameter involved in the nonstationary regime has a nonstandard asymptotic distribution. In the limit, the rate of convergence in the stationary regime is asymptotically proportional to $n^{-\frac{1}{4}}$, whereas it is n^{-1} in the nonstationary regime. The proposed theory and estimation method are illustrated by both simulated and real data examples.

[1] *Correspondence*: Dag Tjøstheim, Department of Mathematics, University of Bergen, Bergen 5008, Norway. Email: dag.tjostheim@math.uib.no.

1 Introduction

Ordinary unit root models have just one regime, whereas ordinary threshold models have several regimes, but are stationary. In this paper, we study a threshold model that has unit–root behavior on one regime and acts as a stationary process in another regime. More specifically, we consider a parametric threshold auto–regressive (TAR) model of the form

$$y_t = \alpha_1 y_{t-1} I[y_{t-1} \in C_\tau] + \alpha_2 y_{t-1} I[y_{t-1} \in D_\tau] + e_t, \ 1 \le t \le n, \tag{1.1}$$

where C_τ is a subset of $R^1 = (-\infty, \infty)$ indexed by $\tau > 0$, $D_\tau = C_\tau^c = R^1 - C_\tau$ is the complement of C_τ, τ is essentially assumed to be known in the asymptotic analysis in this paper, $-\infty < \alpha_1, \ \alpha_2 < \infty$ are assumed to be unknown parameters, but will be estimated under the assumption that $\alpha_2 = 1$, the distribution of $\{e_t\}$ is absolutely continuous with respect to Lebesgue measure with $p_e(\cdot)$ being the density function satisfying $\inf_{x \in C} p_e(x) > 0$ for all compact sets C, $\{e_t\}$ is assumed to be a sequence of independent and identically distributed (i.i.d.) random errors with $E[e_1] = 0$, $0 < \sigma^2 = E[e_1^2] < \infty$ and $E[e_1^4] < \infty$, $\{e_t\}$ and $\{y_s\}$ are assumed to be mutually independent for all $s < t$, and n is the sample size of the time series. Let $y_0 = 0$ throughout this paper. Even though (1.1) is the simplest possible of the type of models we are discussing, it requires nonstandard techniques using the theory of null recurrent Markov chains.

The vast majority of threshold models used have been stationary models, i.e., models for which $|\alpha_1| < 1$ and $|\alpha_2| < 1$ in the first order case. Such models were introduced by Tong and Lim (1980). See also Tong (1983, 1990). Among the more recent contributions, Chan (1990, 1993) consider both estimation and testing problems for the case where $\{y_t\}$ of (1.1) is stationary, Pham, Chan and Tong (1991) consider a nonlinear unit–root problem and establish strong consistency results for the ordinary least squares (OLS) estimators of α_1 and α_2 for the case where (α_1, α_2) lies on the boundary, Hansen (1996) rigorously establishes an asymptotic theory for the likelihood ratio test for a threshold, Chan and Tsay (1998) discuss a related continuous–time TAR model, and Hansen (2000) proposes a new approach to estimating stationary TAR models. More recently, Liu, Ling and Shao (2009) extend the discussion of Pham, Chan and Tong (1991) by establishing an asymptotic distribution of the OLS estimator of α_2 for the case where $C_\tau = (-\infty, \tau]$ and either $\alpha_2 = 1$ and $\alpha_1 < 1$ or $\alpha_2 > 1$ and $\alpha_1 \leq 1$ holds.

Lately, there have been extensions to the nonstationary case, see in particular Caner and Hansen (2001), thus having a class of models that allow for both nonlinearity and nonstationarity, and where these properties can be (Caner and Hansen 2001) separately tested for. The nonstationarity of these models under the null hypothesis has been of a rather restricted form, thus typically regarding both $y_t - y_{t-1}$ and the threshold variable to be stationary. In the first order case, this leads to a somewhat degenerate model that under the null hypothesis has $H_0: \ \alpha_1 = \alpha_2 = 1$ in

$$y_t - y_{t-1} = (\alpha_1 - 1)\ y_{t-1} I[z_t \in C_\tau] + (\alpha_2 - 1)\ y_{t-1} I[z_t \in D_\tau] + e_t, \tag{1.2}$$

where $\{z_t\}$ is a sequence of stationary threshold variables, $C_\tau = (-\infty, \tau]$ and $D_\tau = (\tau, \infty)$. The parameters α_1 and α_2 can then be estimated under H_0, which leads to a pure random walk model for (1.2) but more general difference type models for the higher order case are treated in Caner and Hansen (2001). The authors also point out that there are several nonstationary alternatives when H_0 does not hold. One of the alternatives to H_0 is as follows:

$$H_1: \ |\alpha_1| \neq 1 \quad \text{and} \quad \alpha_2 = 1, \tag{1.3}$$

which does not imply $y_t - y_{t-1}$ is stationary under H_1.

We allow for more general forms of nonstationarity in which we do not require $y_t - y_{t-1}$ to be stationary, nor do we require the threshold variable to be stationary. To the best of our knowledge, estimation in this situation has not been treated before in the literature. In the present paper, for simplicity, we only treat the first order case, but the theory can be extended to higher order and vector models, making it possible to introduce threshold cointegration models in this context. It is also possible to allow nonlinear behavior in the regime C_τ. This is done by replacing the linear function $\alpha_1 y$ by a nonparametric function, also implicitly including an intercept in the model.

Although our focus in this paper is to estimate both α_1 and α_2 and then study asymptotic properties of the proposed estimates in Section 2.1 when τ is assumed to be known, we propose an estimation procedure for the τ parameter in Section

2.2 when τ is unknown. Since the case of both $|\alpha_1| < 1$ and $|\alpha_2| < 1$ and the case of $\alpha_1 = \alpha_2 = 1$ have already been discussed in the literature (Chen 1993; Hansen 2000 for example), we are interested in proposing an estimation method to deal with model (1.1) where C_τ is either a compact subset of R^1 or a set of type $(-\infty, \tau]$ or $[\tau, \infty)$, D_τ is the complement of C_τ, $|\alpha_1| < 1$ or $|\alpha_1| > 1$ and $\alpha_2 = 1$. Model (1.1) may be used to detect and then estimate structural change from one regime to another. Note that τ can be a vector of unknown parameters. In the case where $C_\tau = [\tau_1, \tau_2]$ with $-\infty < \tau_1 < \tau_2 < \infty$, the choice of τ is $\tau = (\tau_1, \tau_2)$.

It is shown in Section 2 below that the OLS estimator of α_1 is asymptotically consistent with a rate of convergence which in the limit is proportional to $n^{-\frac{1}{4}}$ where we can even let $|\alpha_1| > 1$ when C_τ is compact. By contrast, the OLS estimator of α_2 is asymptotically consistent with the super n–rate of convergence. In a related paper by Liu, Ling and Shao (2009), the authors have established similar results for $\widehat{\alpha}_2$, but have not established any asymptotic theory for $\widehat{\alpha}_1$.

The organization of this paper is as follows. Section 2 establishes asymptotic distributions of the OLS estimators of α_1 and α_2 and contains an estimation procedure for the threshold parameter τ. Section 3 discusses an extension of model (1.1) to a semiparametric threshold auto–regressive (SEMI–TAR) model. Examples of implementation are given in Section 4. The paper concludes in Section 5. We will use the theory of β–null recurrent Markov chains in this paper and some general results about these processes are given in Appendix A. Mathematical proofs of some lemmas are given in Appendix B.

2 Estimation in parametric threshold autoregressive models

We propose an ordinary least squares (OLS) estimation method for the unknown parameters α_1 and α_2 in Section 2.1. Discussion about estimation of the τ parameter is given in Section 2.2.

2.1 OLS estimation method and asymptotic theory

Consider model (1.1). It is obvious that α_1 and α_2 can be estimated by the ordinary least squares estimators

$$\widehat{\alpha}_1 = \widehat{\alpha}_1(\tau) = \frac{\sum_{t=1}^n y_t\ y_{t-1} I[y_{t-1} \in C_\tau]}{\sum_{t=1}^n y_{t-1}^2 I[y_{t-1} \in C_\tau]} \quad \text{and} \tag{2.1}$$

$$\widehat{\alpha}_2 = \widehat{\alpha}_2(\tau) = \frac{\sum_{t=1}^n y_t\ y_{t-1} I[y_{t-1} \in D_\tau]}{\sum_{t=1}^n y_{t-1}^2 I[y_{t-1} \in D_\tau]}. \tag{2.2}$$

This implies that

$$\widehat{\alpha}_1 - \alpha_1 = \frac{\sum_{t=1}^n e_t\ y_{t-1} I[y_{t-1} \in C_\tau]}{\sum_{t=1}^n y_{t-1}^2 I[y_{t-1} \in C_\tau]} \quad \text{and} \tag{2.3}$$

$$\widehat{\alpha}_2 - 1 = \frac{\sum_{t=1}^n e_t\ y_{t-1} I[y_{t-1} \in D_\tau]}{\sum_{t=1}^n y_{t-1}^2 I[y_{t-1} \in D_\tau]}. \tag{2.4}$$

In order to establish an asymptotic distribution for each of the estimators, we first need to state some auxiliary results. Observe that model (1.1) can be written as

$$y_t - y_{t-1} = (\alpha_1 - 1) y_{t-1} I[y_{t-1} \in C_\tau] + e_t \equiv u_t + e_t, \tag{2.5}$$

where $u_t = (\alpha_1 - 1) y_{t-1} I[y_{t-1} \in C_\tau]$.

Before our further discussion, we need to introduce Lemma 2.1 below. As it is a special case of Lemma 3.1 below, we need only to prove Lemma 3.1 in Appendix B.

Lemma 2.1 *Let $\{y_t\}$ be generated by model (1.1). Then $\{y_t\}$ is a β–null recurrent Markov chain with $\beta = \frac{1}{2}$.*

A β–null recurrent Markov chain possesses an invariant measure π_s and there is a variable $T(n)$ keeping track of the number of regenerations at time n. Note that the definitions of $\pi_s(\cdot)$ and $T(n)$ are given in detail in Appendix A below. If C_τ is compact, $T(n)$ may be taken to be proportional to the number of visits to C_τ, as is seen from the remark at the end of this subsection. Let $\mu_i = \int_{-\infty}^{\infty} y^i I[y \in C_\tau] \pi_s(dy)$ for $i = 1, 2$. Then Lemma A.1(i) implies that the following limits hold almost surely,

$$m_u \equiv \lim_{n\to\infty} \frac{1}{T(n)} \sum_{t=1}^n u_t = \lim_{n\to\infty} \frac{(\alpha_1 - 1)}{T(n)} \sum_{t=1}^n y_{t-1} I[y_{t-1} \in C_\tau] = (\alpha_1 - 1)\mu_1. \tag{2.6}$$

It follows from Lemma A.2 in Appendix A and then Lemma 2.1 that as $n \to \infty$

$$\frac{\sqrt{T([nr])}}{\sigma_u} \left(\frac{1}{T([nr])} \sum_{t=1}^{[nr]} u_t - m_u \right) \to_D B[M_\beta(r)] \tag{2.7}$$

uniformly in $0 \leq r \leq 1$, where the symbol " $\rightarrow_D$ " means weak convergence in cadlag space (see, for example, the appendix of KT 2001), $\sigma_u^2 = \mu_2 - \mu_1^2$, and $M_\beta(t)$ is the Mittag-Leffler process as defined in KT (2001, p 388). Finally, $[nr]$ is the largest integer part of nr.

Let $\eta_t = u_t + e_t$. Using (2.6) and (2.7), it then follows from the continuous mapping theorem (Corollary 2 of Billingsley 1968, p. 31) and Lemma A.2 that as $n \to \infty$

$$
\begin{aligned}
Q_n(r) \quad &\equiv \quad \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} \eta_t = \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} u_t + \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} e_t \\
&= \quad \frac{\sigma_u \sqrt{T([nr])}}{\sqrt{n}} \frac{\sqrt{T([nr])}}{\sigma_u} \left(\frac{1}{T([nr])} \sum_{t=1}^{[nr]} u_t - m_u \right) + \frac{T([nr])}{\sqrt{n}} m_u \\
&+ \quad \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} e_t = \frac{\sigma_u \sqrt{T([nr])}}{\sqrt{n}} B\left[M_\beta(r)\right] + \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} e_t + \frac{T([nr])}{\sqrt{n}} m_u \\
&+ \quad o_P(1) = \frac{\sigma}{\sqrt{n}\sigma} \sum_{t=1}^{[nr]} e_t + \frac{T([nr])}{\sqrt{n}} m_u + o_P(1) \\
&\rightarrow_D \quad \sigma B(r) + M_{\frac{1}{2}}(r)\ m_u \equiv Q(r) \qquad (2.8)
\end{aligned}
$$

uniformly in $0 < r \leq 1$, where Lemma A.4 in Appendix A has also been used.

This conclusion is summarized in Lemma 2.2.

Lemma 2.2 *Let $\{y_t\}$ be generated by model (1.1). Then as $n \to \infty$*

$$
Q_n(r) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} u_t + \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} e_t \rightarrow_D \sigma B(r) + M_{\frac{1}{2}}(r)\ m_u \equiv Q(r). \qquad (2.9)
$$

Note that when $\mu_1 = 0$ and thus $m_u = 0$, the contribution of $\{u_t\}$ to $\{y_t\}$ is asymptotically negligible. In this case, $\{y_t\}$ behaves like a random walk process.

We state the following lemma; its proof is given in Appendix B.

Lemma 2.3 *Assume that model (1.1) holds. Then as $n \to \infty$*

$$
\begin{aligned}
\frac{1}{T(n)} \sum_{t=1}^{n} y_{t-1}^2 I[y_{t-1} \in C_\tau] \quad &\rightarrow_P \quad \int_{-\infty}^{\infty} y^2 I[y \in C_\tau] \pi_s(dy), && (2.10) \\
\frac{1}{n^2} \sum_{t=1}^{n} y_{t-1}^2 I[y_{t-1} \in D_\tau] \quad &\xrightarrow{d} \quad \int_0^1 Q^2(r)\ dr, && (2.11) \\
\frac{1}{\sqrt{T(n)}} \sum_{t=1}^{n} y_{t-1} e_t I[y_{t-1} \in C_\tau] \quad &\xrightarrow{d} \quad N(0, \sigma_1^2), && (2.12) \\
\frac{1}{n} \sum_{t=1}^{n} y_{t-1} e_t I[y_{t-1} \in D_\tau] \quad &\xrightarrow{d} \quad \frac{1}{2} \left(Q^2(1) - \sigma^2 \right), && (2.13)
\end{aligned}
$$

where the symbol " $\xrightarrow{d}$ " denotes convergence in distribution, $\sigma_1^2 = \sigma^2 \int_{-\infty}^{\infty} y^2 I[y \in C_\tau]\pi_s(dy)$ and $Q(r) = \sigma B(r) + (\alpha_1 - 1)\mu_1\, M_{\frac{1}{2}}(r)$.

We now state the main results of this section.

THEOREM 2.1 *Assume that model (1.1) holds. Then as $n \to \infty$*

$$\sqrt{T(n)}\,(\widehat{\alpha}_1 - \alpha_1) \quad \xrightarrow{d} \quad N\left(0, \sigma^4\, \sigma_1^{-2}\right), \tag{2.14}$$

$$n\,(\widehat{\alpha}_2 - 1) \quad \xrightarrow{d} \quad \frac{(Q^2(1) - \sigma^2)}{2\int_0^1 Q^2(r)dr}. \tag{2.15}$$

Note that $Q(r) = \sigma B(r)$ when $\mu_1 = 0$. This implies that the asymptotic theory for $\widehat{\alpha}_2$ is the same as that for the unit–root case when $\mu_2 = \int_{-\infty}^{\infty} y I[y \in C_\tau]\pi_s(dy) = 0$, i.e., $\{y_t\}$ has some symmetrical structure in the stationary regime. In this symmetrical case, the asymptotic distribution in (2.15) corresponds to the main result in Theorem 2.1 of Liu, Ling and Shao (2009).

PROOF: Recall that

$$\widehat{\alpha}_1 - \alpha_1 \quad = \quad \frac{\sum_{t=1}^n e_t\; y_{t-1} I[y_{t-1} \in C_\tau]}{\sum_{t=1}^n y_{t-1}^2 I[y_{t-1} \in C_\tau]} \quad \text{and} \tag{2.16}$$

$$\widehat{\alpha}_2 - 1 \quad = \quad \frac{\sum_{t=1}^n e_t\; y_{t-1} I[y_{t-1} \in D_\tau]}{\sum_{t=1}^n y_{t-1}^2 I[y_{t-1} \in D_\tau]}. \tag{2.17}$$

The proof of Theorem 2.1 then follows immediately from Lemma 2.3 and the continuous mapping theorem.

REMARK 2.1. Theorem 2.1 shows that the rate of convergence of $\widehat{\alpha}_1$ to α_1 is proportional to $\sqrt{T(n)}$ while the rate of convergence of $\widehat{\alpha}_2$ to 1 is proportional to n. According to Lemmas 2.1 and 3.4 and Theorem 3.2 of Karlsen and Tjøstheim (2001), $T(n)$ behaves asymptotically as the Mittag–Leffler variable $M_{\frac{1}{2}}(\cdot)$ and in the limit can be associated with the deterministic convergence rate of $n^{-\frac{1}{2}}$. Our results can be translated to local–time terminology and are threshold autoregressive counterparts of the results in Phillips (1987), and Park and Phillips (2001). The results of those papers were for the nonlinear and nonstationary regression case and it is not clear whether the local time techniques used there can be extended to an autoregressive situation. Note that $T(n)$ may be replaced by $\frac{T_C(n)}{\pi_s(1_C)}$ (Lemma 3.6 of Karlsen and Tjøstheim 2001), where $T_C(n)$ is the number of visits to a small set C, which may be taken to be a subset of C_τ or C_τ itself if it is compact.

2.2 Discussion about estimation of the τ parameter

In both theory and practice, estimation of the τ parameter is of interest and importance.

Let $\widehat{e}_t(\tau) = y_t - \widehat{\alpha}_1 y_{t-1} I[y_{t-1} \in C_\tau] - \widehat{\alpha}_2 y_{t-1} I[y_{t-1} \in D_\tau]$ and then define the estimated variance by

$$\widehat{\sigma}^2(\tau) = \frac{1}{n}\sum_{t=1}^{n} \widehat{e}_t^2(\tau). \tag{2.18}$$

The τ parameter can then be estimated by

$$\widehat{\tau} = \arg \min_{\text{over all } \tau} \widehat{\sigma}^2(\tau). \tag{2.19}$$

In both the stationary and nonstationary unit–root cases, asymptotic properties of $\widehat{\tau}$ have been discussed (see, for example, Pham, Chan and Tong 1991; Chan 1993; Hansen 2000). In the paper by Chan (1993), the author shows that the rate of convergence of $\widehat{\tau}$ to τ can be as fast as the super–rate of n. Our simulation study in Section 4 however suggests that the rate of convergence of $\widehat{\tau}$ to τ may be related to $T(n)$, as will also be pointed out in the discussion below.

Studying asymptotic properties for $\widehat{\tau}$ in detail for the model we are considering requires a separate investigation, since even in the stationary case the theory is quite complex (see, for example, Chan 1993). In the present paper, we will only indicate how Chan's proof of consistency can be extended and comment on the rate that can be expected.

Chan (1993) restricts himself to the case of a single threshold τ, so that there is a stationary regime to the left of τ and another stationary regime to the right of τ. In our discussion we will use the same simplification but with one of the regimes being a unit root regime. Moreover, since we only look at the first order case, we take the threshold variable y_{t-d} to be y_{t-1}.

Chan makes use of ergodicity in his proof, which we do not have in our case, but his proof of consistency can nevertheless be adapted to our situation by noting that

$$L(\theta) = \sum_{t=1}^{n} \left(y_t - E_\theta(y_t|\mathcal{F}_{t-1})\right)^2$$

can be decomposed using the existence of the regeneration mechanism for a null recurrent process, such that (see (A.2) and (A.3) of Appendix A)

$$L(\theta) = U_0 + \sum_{k=1}^{T(n)} U_k + U_{(n)}. \tag{2.20}$$

Here, θ is the parameter composed of the AR coefficients and the threshold with θ belonging to a parameter space Θ. Moreover, $E_\theta(\cdot|\mathcal{F}_{t-1})$ is the conditional expectation with respect to the σ–field $\mathcal{F}_{t-1}$ generated by $\{y_s, 1 \le s \le t-1\}$, and

$$U_k = U_k(g,\theta) = \sum_{t=\tau_{k-1}+1}^{\tau_k} g(y_t, y_{t+1}, \theta)$$

with $g(y_{t-1}, y_t, \theta) = (y_t - E_\theta(y_t|\mathcal{F}_{t-1}))^2$, and where the τ_k's are regeneration times. Finally, U_0 and $U_{(n)}$ in (2.20) are a starting term and an ending term that can be neglected as $n \to \infty$. The sequence $\{U_k\}$ consists of random variables that are identically distributed and are 1–dependent. It essentially takes the place of the ergodic process $\{y_t\}$ in Chan's proof and of course $\{U_k\}$ trivially fulfills the stationarity and ergodicity requirement of his Theorem 1, where strong consistency is proved.

The decomposition of (3.2) of Chan (1993) can now be written as

$$\sum_{t=\tau_{k-1}+1}^{\tau_k} (y_t - E_\theta(y_t|\mathcal{F}_{t-1}))^2 = R_{1k}(\theta) + R_{2k}(\theta) + R_{3k}(\theta) + R_{4k}(\theta),$$

with

$$\begin{aligned}
R_{1k}(\theta) &= \sum_{t=\tau_{k-1}+1}^{\tau_k} (y_t - \beta_1 y_{t-1})^2 I(y_{t-1} \le z, y_{t-1} \le \tau), \\
R_{2k}(\theta) &= \sum_{t=\tau_{k-1}+1}^{\tau_k} (y_t - \beta_1 y_{t-1})^2 I(y_{t-1} \le z, y_{t-1} > \tau), \\
R_{3k}(\theta) &= \sum_{t=\tau_{k-1}+1}^{\tau_k} (y_t - \beta_2 y_{t-1})^2 I(y_{t-1} > z, y_{t-1} \le \tau), \\
R_{4k}(\theta) &= \sum_{t=\tau_{k-1}+1}^{\tau_k} (y_t - \beta_2 y_{t-1})^2 I(y_{t-1} > z, y_{t-1} > \tau),
\end{aligned}$$

where β_1, β_2 and z are neighboring points of the true values α_1, α_2 and τ in the parameter space Θ. Next, $R_{ik}(\theta)$, $i = 1, \cdots, 4$ can themselves be decomposed analogously to (3.3) in Chan (1993) by adding and substracting the true AR parameters α_1 and α_2. With these changes, the proof of Lemma 1 in Chan (1993) can be carried through. Moreover, in the proof of his claim 1, σ_θ^2 can be replaced by $E\left[\sum_{t=\tau_{k-1}+1}^{\tau_k} (y_t - E_{\theta_0}(y_t|\mathcal{F}_{t-1}))^2\right]$ with $\theta_0 = (\alpha_1, \alpha_2)$, and one has to introduce a truncation variable to ensure the existence of $E[R_{ik}(\theta)]$. The rest of the proof of consistency can be carried out along the lines of Chan (1993).

Chan obtains a rate for $\widehat{\tau} - \tau$ of order n^{-1}. By the decomposition in (2.20) we effectively introduce a 1–dependent process where $T(n)$ can be taken as the number of observations. We could therefore possibly expect a rate for $\widehat{\tau} - \tau$ of order $T^{-1}(n)$ which can be associated with $n^{-\frac{1}{2}}$. This is in agreement with the finite sample results for Case A of Example 4.3 below, which is an example where a threshold of this type was investigated by simulation. The finite sample rate for the other examples, where two thresholds are involved, is slower. It should be kept in mind, though, that the association of $T(n)$ with $\sqrt{n}$ is itself an asymptotic result (see, for example, Lemma 3.4 and Theorem 3.2 of Karlsen and Tjøstheim 2001). For a finite n, $T(n)$ will certainly depend on the set C_τ. Rigorous conditions and results about the rate and indeed about the asymptotic distribution would require an extension of Chan's Propositions 1 and 2 as well as Theorem 2. This is far from trivial and would require a separate paper.

3 Estimation in semiparametric threshold autoregressive models

This section considers a semiparametric threshold auto–regressive (SEMI–TAR) model of the form

$$\begin{aligned} y_t &= g(y_{t-1})I[y_{t-1} \in C_\tau] + \alpha y_{t-1} I[y_{t-1} \in D_\tau] + e_t \\ &= \begin{cases} g(y_{t-1}) + e_t & \text{if } y_{t-1} \in C_\tau, \\ \alpha\, y_{t-1} + e_t & \text{if } y_{t-1} \in D_\tau, \end{cases} \end{aligned} \tag{3.1}$$

where C_τ and D_τ are as defined in (1.1), $g(x)$ is an unknown and bounded function when $x \in C_\tau$, $\alpha = 1$, and $\{e_t\}$ is the same as assumed in (1.1). Let $y_0 = 0$. Model (3.1) may be used to detect and then estimate structural change from a nonlinear 'stationary' regime to a linear 'nonstationary' regime.

While the special case of $\alpha = 1$ of model (3.1) has been mentioned in Karlsen *et al* (2007) as an example of a null recurrent process, the asymptotic estimation theory for model (3.1) has not been studied in the literature. Existing results for the stationary nonlinear time series models (Tong 1990; Fan and Yao 2003; Gao 2007) are also not directly applicable to study such SEMI–TAR models. Our interest is to study asymptotic behavior of both a nonparametric estimator of $g(\cdot)$ and an OLS estimator of α.

In order to establish consistent estimates for $g(\cdot)$ and α, we need to introduce the following assumption.

ASSUMPTION 3.1 (i) The invariant measure π_s of $\{y_t\}$ has a locally continuous density $p_s(y)$ that is locally strictly positive; that is, $p_s(y) > 0$ for all $y \in R^1$.

(ii) Let $g(y)$ be twice differentiable and the second derivative $g''(y)$ be continuous at all $y \in R^1$.

(iii) Let $K(\cdot)$ be a symmetric probability kernel function with compact support $C(K)$. The bandwidth parameter h satisfies $\lim_{n\to\infty} h = 0$, $\lim_{n\to\infty} nh = \infty$ and $\limsup_{n\to\infty} n^{1+\delta_0} h^6 < \infty$ for some $0 < \delta_0 < \frac{1}{2}$.

(iv) In case C_τ is not compact, i.e. $C_\tau = (-\infty, \tau]$ or $C_\tau = [\tau, \infty)$, $|g(y)| \le c_g |y|$ with $0 < c_g < 1$ as $|y| \to \infty$.

Conditions in Assumption 3.1(i)(ii)(iii) are quite mild conditions (see, for example, Assumptions $B_0 - B_3$ of Karlsen and Tjøstheim 2001). Condition 3.1(iv) is to secure stationary type behavior on C_τ.

We need the following lemma; its proof is given in Appendix A below.

LEMMA 3.1 *Let $\{y_t\}$ be generated by model (3.1). If Assumption 3.1(i)(ii)(iv) holds, then $\{y_t\}$ is a β–null recurrent Markov chain with $\beta = \frac{1}{2}$.*

Similarly to (2.5), we have

$$y_t - y_{t-1} = \left(g(y_{t-1}) - y_{t-1}\right) I[y_{t-1} \in C_\tau] + e_t \equiv v_t + e_t. \tag{3.2}$$

Let $\mu_g = \int_{-\infty}^{\infty} g(y) I[y \in C_\tau] \pi_s(dy)$. Then Lemma A.1(i) below implies that the following limits hold almost surely,

$$g_v \equiv \lim_{n\to\infty} \frac{1}{T(n)} \sum_{t=1}^{n} v_t = \lim_{n\to\infty} \frac{1}{T(n)} \sum_{t=1}^{n} \left(g(y_{t-1}) - y_{t-1}\right) I[y_{t-1} \in C_\tau] = \mu_g - \mu_1, \tag{3.3}$$

where μ_1 is as defined in (2.6).

We state the following lemma; its proof is similar to equations (2.7)–(2.9).

LEMMA 3.2 *Let $\{y_t\}$ be generated by model (3.1). If Assumption 3.1(i)(ii)(iv) holds, then as $n \to \infty$*

$$P_n(r) \equiv \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} e_t + \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} v_t \to_D \sigma B(r) + M_{\frac{1}{2}}(r)\; g_v \equiv P(r). \tag{3.4}$$

Let $K(\cdot)$ be a probability kernel function and h be a bandwidth parameter satisfying Assumption 3.1(ii) above. It is obvious that $g(y)$ and α can be estimated by

$$\begin{aligned}
\widehat{g}(y) &= \widehat{g}(y,\tau) = \frac{\sum_{t=1}^n K\left(\frac{y-y_{t-1}}{h}\right) y_t I[y \in C_\tau]}{\sum_{t=1}^n K\left(\frac{y-y_{t-1}}{h}\right) I[y \in C_\tau]} \quad \text{and} \\
\widehat{\alpha} &= \widehat{\alpha}(\tau) = \frac{\sum_{t=1}^n y_t\ y_{t-1} I[y \in D_\tau]}{\sum_{t=1}^n y_{t-1}^2 I[y \in D_\tau]},
\end{aligned} \tag{3.5}$$

which imply that

$$\begin{aligned}
\widehat{g}(y) - g(y) &= \frac{\sum_{t=1}^n K\left(\frac{y-y_{t-1}}{h}\right) (g(y_{t-1}) - g(y))\, I[y \in C_\tau]}{\sum_{t=1}^n K\left(\frac{y-y_{t-1}}{h}\right) I[y \in C_\tau]} \\
&+ \frac{\sum_{t=1}^n K\left(\frac{y-y_{t-1}}{h}\right) e_t\ I[y \in C_\tau]}{\sum_{t=1}^n K\left(\frac{y-y_{t-1}}{h}\right) I[y \in C_\tau]}, \\
\widehat{\alpha} - 1 &= \frac{\sum_{t=1}^n e_t\ y_{t-1} I[y \in D_\tau]}{\sum_{t=1}^n y_{t-1}^2 I[y \in D_\tau]}.
\end{aligned} \tag{3.6}$$

We now state the main results of this section.

THEOREM 3.1 *Assume that both model (3.1) and Assumption 3.1 hold. Then as $n \to \infty$*

$$\sqrt{\sum_{t=1}^n K\left(\frac{y-y_{t-1}}{h}\right) I[y \in C_\tau]}\ \left(\widehat{g}(y) - g(y)\right) \ \stackrel{d}{\longrightarrow}\ N(0, \sigma_2^2), \tag{3.7}$$

$$n\left(\widehat{\alpha} - 1\right) \ \stackrel{d}{\longrightarrow}\ \frac{(P^2(1) - \sigma^2)}{2\int_0^1 P^2(r) dr}, \tag{3.8}$$

where $\sigma_2^2 = \sigma^2 \int K^2(u) du$ and $P(r) = \sigma B(r) + M_{\frac{1}{2}}(r)\ g_u$. Note that $P(r) = \sigma B(r)$ when $g_u = 0$.

PROOF: Because of Lemma 3.1, the proof of (3.8) is the same as that of (2.15). Let $W_{nt}(y) = \frac{K\left(\frac{y-y_{t-1}}{h}\right) I[y_{t-1} \in C_\tau]}{\sum_{t=1}^n K\left(\frac{y-y_{t-1}}{h}\right) I[y_{t-1} \in C_\tau]}$. In order to prove (3.7), in view of (3.6), it suffices to show that as $n \to \infty$

$$\sqrt{\sum_{t=1}^n K\left(\frac{y-y_{t-1}}{h}\right) I[y_{t-1} \in C_\tau]}\ \sum_{t=1}^n W_{nt}(y) \left(g(y_{t-1}) - g(y)\right) \to_P 0, \tag{3.9}$$

$$\sqrt{\sum_{t=1}^n K\left(\frac{y-y_{t-1}}{h}\right) I[y_{t-1} \in C_\tau]}\ \sum_{t=1}^n W_{nt}(y)\ e_t \stackrel{d}{\longrightarrow} N(0, \sigma^2). \tag{3.10}$$

Note that by Taylor expansions, Lemma A.1(i) and Assumption 3.1

$$
\begin{aligned}
&\frac{1}{T(n)h}\sum_{t=1}^{n} K\left(\frac{y-y_{t-1}}{h}\right) I[y_{t-1}\in C_\tau] = I[y\in C_\tau]\ p_s(y)(1+o_P(1)),\\
&\frac{1}{T(n)h}\sum_{t=1}^{n} K\left(\frac{y-y_{t-1}}{h}\right) I[y_{t-1}\in C_\tau]\,(g(y_{t-1})-g(y))\\
&= \frac{g'(y)h}{T(n)h}\sum_{t=1}^{n} K\left(\frac{y-y_{t-1}}{h}\right) I[y_{t-1}\in C_\tau]\frac{(y_{t-1}-y)}{h}\\
&+ \frac{g''(u)h^2}{2T(n)h}\sum_{t=1}^{n} K\left(\frac{y-y_{t-1}}{h}\right) I[y_{t-1}\in C_\tau]\frac{(y_{t-1}-y)^2}{h^2}\\
&= g'(y)\ h\int vK(v)dv\ I[y\in C_\tau]\ p_s(y) + o_P(h)\\
&+ \frac{h^2 g''(y)}{2}\int v^2K(v)dv\ I[y\in C_\tau]\ p_s(y) + o_P(h^2)\\
&= o_P(h), \qquad (3.11)
\end{aligned}
$$

where u is between y and y_{t-1}, and $p_s(y)$ is the density function of the invariant measure π_s of $\{y_t\}$.

This, along with Lemma A.1(ii) and using Assumption 3.1(iii), implies (3.9).

Let $a_{nt}(y) = \frac{K\left(\frac{y-y_{t-1}}{h}\right) I[y_{t-1}\in C_\tau]}{\sqrt{\sum_{t=1}^{n} K^2\left(\frac{y-y_{t-1}}{h}\right) I[y_{t-1}\in C_\tau]}}$. We can re–write (3.10) as

$$
\begin{aligned}
&\sqrt{\sum_{t=1}^{n} K\left(\frac{y-y_{t-1}}{h}\right) I[y_{t-1}\in C_\tau]}\ \sum_{t=1}^{n} W_{nt}(y)e_t\\
&= \sqrt{\frac{\sum_{t=1}^{n} K^2\left(\frac{y-y_{t-1}}{h}\right) I[y_{t-1}\in C_\tau]}{\sum_{t=1}^{n} K\left(\frac{y-y_{t-1}}{h}\right) I[y_{t-1}\in C_\tau]}}\ \sum_{t=1}^{n} a_{nt}(y)e_t. \qquad (3.12)
\end{aligned}
$$

Thus, in order to show (3.10), it suffices to show that as $n\to\infty$

$$
\sum_{t=1}^{n} a_{nt}(y)e_t \xrightarrow{d} N(0,\sigma^2). \qquad (3.13)
$$

To prove (3.13), by a conventional martingale central limit theorem (see, for example, Corollary 3.1 of Hall and Heyde 1980), it suffices to verify that as $n\to\infty$

$$
\begin{aligned}
\sum_{t=1}^{n} a_{nt}^2(y)E[e_t^2] &\to_P \sigma^2, \qquad (3.14)\\
\sum_{t=1}^{n} a_{nt}^4(y)E[e_t^4] &\to_P 0. \qquad (3.15)
\end{aligned}
$$

The proof of (3.14) follows automatically from $\sum_{t=1}^n a_{nt}^2(y) \equiv 1$, while the proof of (3.15) follows from

$$\frac{\sum_{t=1}^n K^4\left(\frac{y-y_{t-1}}{h}\right)}{\left(\sum_{t=1}^n K^2\left(\frac{y-y_{t-1}}{h}\right)\right)^2} = \frac{1}{T(n)} \frac{\frac{1}{T(n)}\sum_{t=1}^n K^4\left(\frac{y-y_{t-1}}{h}\right)}{\left(\frac{1}{T(n)}\sum_{t=1}^n K^2\left(\frac{y-y_{t-1}}{h}\right)\right)^2} = O_P\left(\frac{1}{T(n)}\right) = o_P(1) \tag{3.16}$$

using Lemma A.1(i). This therefore completes the proof of Theorem 3.1.

Remark 3.1. Compared with Theorem 2.1, Theorem 3.1 shows that while the parameter estimator $\widehat{\alpha}$ has the same asymptotic distribution as that of $\widehat{\alpha}_2$, the nonparametric estimator $\widehat{g}(\cdot)$ as expected has a rate of convergence slower than its parametric counterpart $\widehat{\alpha}_1$. In addition, Theorem 3.1 shows that the rate of convergence of $\widehat{g}(\cdot)$ is also slower than that of the corresponding nonparametric kernel estimator for the stationary case, as shown in Karlsen and Tjøstheim (2001), Karlsen, MyKelbust and Tjøstheim (2007), and Wang and Phillips (2009).

Other closely related papers in the field of nonparametric and semiparametric regression estimation involving nonstationary time series include Chen, Gao and Li (2008), Cai, Li and Park (2009), Gao *et al* (2009a, 2009b), and Wang and Phillips (2010).

4 Examples of implementation

This section gives several examples to evaluate the finite–sample performance of the proposed estimation method in several different cases. There are four simulation examples and one real date example.

Consider a general threshold autoregressive (TAR) model of the form

$$y_t = \alpha_1 y_{t-1} I[y_{t-1} \in C_\tau] + \alpha_2 y_{t-1} I[y_{t-1} \in C_\tau^c] + e_t, \ 1 \le t \le n, \tag{4.1}$$

where $\tau = (\tau_1, \tau_2)$, $C_\tau = [\tau_1, \tau_2]$ for $-\infty < \tau_1 < \tau_2 < \infty$ with both τ_1 and τ_2 being the threshold parameters, $C_\tau^c = (-\infty, \tau_1) \cup (\tau_2, \infty)$, and $\{e_t\}$ is assumed to be a sequence of independent and normally distributed random errors with $E[e_1] = 0$ and $\sigma^2 = E[e_1^2] = 1$. That is, $e_t \sim N(0,1)$. Let $y_0 = 0$.

The unknown parameters α_1, α_2 and τ are estimated by the ordinary least squares estimators:

$$\begin{aligned}
\widetilde{\alpha}_1 &= \widehat{\alpha}_1(\widehat{\tau}) = \frac{\sum_{t=1}^n y_t y_{t-1} I[y_{t-1} \in C_{\widehat{\tau}}]}{\sum_{t=1}^n y_{t-1}^2 I[y_{t-1} \in C_{\widehat{\tau}}]}, && (4.2)\\
\widetilde{\alpha}_2 &= \widehat{\alpha}_2(\widehat{\tau}) = \frac{\sum_{t=1}^n y_t\ y_{t-1} I[y_{t-1} \in C_{\widehat{\tau}}^c]}{\sum_{t=1}^n y_{t-1}^2 I[y_{t-1} \in C_{\widehat{\tau}}^c]}, && (4.3)\\
\widehat{\tau} &= \arg \min_{\text{over all } \tau} \widehat{\sigma}^2(\tau), && (4.4)
\end{aligned}$$

where $\widehat{\sigma}^2(\tau) = \frac{1}{n}\sum_{t=1}^n \left(y_t - \widehat{\alpha}_1(\tau) y_{t-1} I[y_{t-1} \in C_\tau] - \widehat{\alpha}_2(\tau) y_{t-1} I[y_{t-1} \in C_\tau^c]\right)^2$. Let $\widehat{\tau} = (\widehat{\tau}_1, \widehat{\tau}_2)$ for the asymmetrical case.

Example 4.1 consider a symmetrical case of the form $C_\tau = [-\tau, \tau]$. An asymmetrical bounded case where $C_\tau = [\tau_1, \tau_2]$ is discussed in Example 4.2 below. Example 4.3 examines the unbounded case where $C_\tau = (-\infty, \tau]$. Throughout Examples 4.1–4.3 below, we consider both the cases of $|\alpha_1| < 1$ and $|\alpha_1| > 1$.

- Consider the case of $n = 1000,\ 2000,\ 5000$ and 10000. Let $N = 1000$ be the number of replications and $\widetilde{\alpha}_i(j)$ and $\widehat{\tau}(j)$ be the respective value of $\widetilde{\alpha}_i$ and $\widehat{\tau}$ at the j–th replication.

- Calculate the standard deviations of the form

$$\text{std}(\widetilde{\alpha}_i) = \sqrt{\frac{1}{N-1}\sum_{j=1}^N \left(\widetilde{\alpha}_i(j) - \overline{\widetilde{\alpha}}_i\right)^2} \text{ and } \text{std}(\widehat{\tau}) = \sqrt{\frac{1}{N-1}\sum_{j=1}^N \left(\widehat{\tau}(j) - \overline{\widehat{\tau}}\right)^2} \quad (4.5)$$

  for $i = 1, 2$ and Cases A and B separately under $N = 1000$, where $\overline{\widetilde{\alpha}}_i = \frac{1}{N}\sum_{j=1}^N \widetilde{\alpha}_i(j)$ and $\overline{\widehat{\tau}} = \frac{1}{N}\sum_{j=1}^N \widehat{\tau}(j)$.

Example 4.1 Consider a symmetrical (bounded) threshold autoregressive (TAR) model of the form

$$y_t = \alpha_1 y_{t-1} I[|y_{t-1}| \le \tau] + \alpha_2 y_{t-1} I[|y_{t-1}| > \tau] + e_t,\ 1 \le t \le n. \quad (4.6)$$

This example then considers the following cases.

- Case A: $\alpha_1 = \frac{1}{2}$, $\alpha_2 = 1$ and $\tau = 2.5$; and

- Case B: $\alpha_1 = \frac{3}{2}$, $\alpha_2 = 1$ and $\tau = 2$.

The simulated results for Example 4.1 are given in Table 4.1 below.

Table 4.1 Simulation Results for Case A and Case B

Case A	std($\widetilde{\alpha}_1$)	std($\widetilde{\alpha}_2$)	std($\widehat{\tau}$)	Case B	std($\widetilde{\alpha}_1$)	std($\widetilde{\alpha}_2$)	std($\widehat{\tau}$)
$n = 1000$	0.1032	0.0144	0.1204	$n = 1000$	0.2703	0.0026	0.2040
$n = 2000$	0.0890	0.0059	0.0958	$n = 2000$	0.2270	0.0013	0.1761
$n = 5000$	0.0756	0.0015	0.0825	$n = 5000$	0.1817	0.0006	0.1439
$n = 10000$	0.0645	0.0007	0.0695	$n = 10000$	0.1558	0.0003	0.1238

Table 4.1 supports the rate results of Theorem 2.1. Case B has larger standard errors for $\widehat{\alpha}_1$ and smaller standard errors for $\widehat{\alpha}_2$, because the explosive behavior on C_τ in this case leads to more frequent stays in the random walk regime.

Example 4.2 Consider an asymmetrical (bounded) threshold autoregressive (TAR) model of the form

$$y_t = \alpha_1 y_{t-1} I[y_{t-1} \in C_\tau] + \alpha_2 y_{t-1} I[y_{t-1} \in C_\tau^c] + e_t, \ 1 \le t \le n. \tag{4.7}$$

We are then interested in the following cases:

- Case A: $\alpha_1 = \frac{1}{2}$, $\alpha_2 = 1$, $\tau_1 = -3$ and $\tau_2 = 2.5$; and
- Case B: $\alpha_1 = \frac{3}{2}$, $\alpha_2 = 1$, $\tau_1 = -1.5$ and $\tau_2 = 1$.

The simulated results for Example 4.2 are given in Table 4.2 below.

Table 4.2 Simulation Results for Cases A and B

Case A	std($\widetilde{\alpha}_1$)	std($\widetilde{\alpha}_2$)	std($\widehat{\tau}_1$)	std($\widehat{\tau}_2$)
$n = 1000$	0.0694	0.0208	0.2506	0.1396
$n = 2000$	0.0503	0.0074	0.2029	0.1186
$n = 5000$	0.0362	0.0024	0.1634	0.0754
$n = 10000$	0.0359	0.0008	0.1401	0.0659
Case B	std($\widetilde{\alpha}_1$)	std($\widetilde{\alpha}_2$)	std($\widehat{\tau}_1$)	std($\widehat{\tau}_2$)
$n = 1000$	0.7606	0.0028	0.2825	0.3146
$n = 2000$	0.7438	0.0015	0.2501	0.2937
$n = 5000$	0.6596	0.0006	0.2155	0.2799
$n = 10000$	0.6168	0.0003	0.1938	0.2535

Similarly to Table 4.1, Table 4.2 also demonstrates that the proposed estimation method still works well numerically even when two truncation parameters are involved in the model.

Example 4.3 Consider a threshold autoregressive (TAR) model with unbounded C_τ of the form

$$y_t = \alpha_1 y_{t-1} I[y_{t-1} \leq \tau] + \alpha_2 y_{t-1} I[y_{t-1} > \tau] + e_t, \ 1 \leq t \leq n. \tag{4.8}$$

This example also considers the following cases:

- Case A: $\alpha_1 = \frac{1}{2}$, $\alpha_2 = 1$, $\tau = 3$; and
- Case B: $\alpha_1 = \frac{3}{2}$, $\alpha_2 = 1$, $\tau = 3$.

The simulated results for Example 4.3 are given in Table 4.3 below.

Table 4.3 Simulation Results for Case A and Case B

Case A	std($\widetilde{\alpha}_1$)	std($\widetilde{\alpha}_2$)	std($\widehat{\tau}$)
$n = 1000$	0.0413	0.1177	0.1595
$n = 2000$	0.0373	0.0475	0.1133
$n = 5000$	0.0192	0.0155	0.0677
$n = 10000$	0.0169	0.0052	0.0556
Case B	std($\widetilde{\alpha}_1$)	std($\widetilde{\alpha}_2$)	std($\widehat{\tau}$)
$n = 1000$	0.2798	0.0034	0.1830
$n = 2000$	0.1712	0.0012	0.1530
$n = 5000$	0.1551	0.0005	0.1453
$n = 10000$	0.1340	0.0002	0.1247

Table 4.3 again supports the rate results of Theorem 2.1. Note that Case B is not covered by Assumption 3.1(iv), but it works well because the process "explodes" from $(-\infty, \tau]$ into the random walk part $[\tau, \infty)$.

In the following example, we consider a semiparametric threshold autoregressive model and then study the finite sample performance of the proposed semiparametric estimation method.

Example 4.4 Consider a semiparametric threshold auto–regressive (SEMI–TAR) model of the form

$$y_t = g(y_{t-1}) I[|y_{t-1}| \le \tau] + \alpha y_{t-1} I[|y_{t-1}| > \tau] + e_t, \tag{4.9}$$

where $\tau = 2.5$ and $e_t \sim N(0,1)$. Let $y_0 = 0$.

Let $K(x) = \frac{1}{2} I_{[-1,1]}(x)$. We then estimate $g(y)$ and α by

$$\widehat{g}(y, \widehat{\tau}) = \frac{\sum_{t=1}^n K\left(\frac{y - y_{t-1}}{\widehat{h}_{\mathrm{cv}}}\right) y_t I[|y_{t-1}| \le \widehat{\tau}]}{\sum_{t=1}^n K\left(\frac{y - y_{t-1}}{\widehat{h}_{\mathrm{cv}}}\right) I[|y_{t-1}| \le \widehat{\tau}]}, \tag{4.10}$$

$$\widetilde{\alpha} = \widehat{\alpha}(\widehat{\tau}) = \frac{\sum_{t=1}^n y_t \, y_{t-1} I[|y_{t-1}| > \widehat{\tau}]}{\sum_{t=1}^n y_{t-1}^2 I[|y_{t-1}| > \widehat{\tau}]}, \tag{4.11}$$

$$\widehat{\tau} = \arg \min_{\text{over all } \tau} \widehat{\sigma}^2(\tau), \tag{4.12}$$

where $\widehat{\sigma}^2(\tau) = \frac{1}{n} \sum_{t=1}^n \left(y_t - \widehat{g}(y_{t-1}, \tau) I[|y_{t-1}| \le \tau] - \widehat{\alpha}(\tau) y_{t-1} I[|y_{t-1}| > \tau]\right)^2$, and $\widehat{h}_{\mathrm{cv}}$ is chosen such that

$$\widehat{h}_{\mathrm{cv}} = \arg \min_{h \in H_n} \frac{1}{n} \sum_{t=1}^n \left(y_t \; I[|y_{t-1}| \le \widehat{\tau}] - \widehat{g}_{-t}(y_{t-1}; h) \; I[|y_{t-1}| \le \widehat{\tau}]\right)^2, \tag{4.13}$$

with $\widehat{g}_{-t}(y_{t-1}; h) = \frac{\sum_{s=1, \ne t}^n K\left(\frac{y_{t-1} - y_{s-1}}{h}\right) y_s \; I[|y_{s-1}| \le \widehat{\tau}]}{\sum_{s=1, \ne t}^n K\left(\frac{y_{t-1} - y_{s-1}}{h}\right) I[|y_{s-1}| \le \widehat{\tau}]}$ and $H_n = \left[n^{-1}, n^{-(1-\delta_0)}\right]$, in which $0 < \delta_0 < 1$ is chosen such that $\widehat{h}_{\mathrm{cv}}$ is achievable and unique in each individual case.

We are interested in the following cases:

- Case A: $g(y) = \frac{1}{1+y^2}$, $\alpha = 1$ and $\tau = 2.5$; and
- Case B: $g(y) = y^2$, $\alpha = 1$ and $\tau = 2.5$.

Consider the cases of $n = 250$, 600 and 1000. Let $\widehat{g}_j(y)$ be the estimated function of $\widehat{g}(y)$ at the j–th replication and $y_t(j)$ be the generated value of y_t at the j–th replication.

- Calculate the standard deviations of the form

$$\mathrm{std}(\widetilde{\alpha}) = \sqrt{\frac{1}{N-1} \sum_{j=1}^N \left(\widetilde{\alpha}(j) - \overline{\widetilde{\alpha}}\right)^2}$$

for Cases A and B separately under $N = 1000$, where $\overline{\widetilde{\alpha}} = \frac{1}{N} \sum_{j=1}^N \widetilde{\alpha}(j)$.

- For the case of $n = 250,\ 600$ and 1000, $N = 1000$ and Cases A and B, calculate the average of the standard deviations of the form

$$\mathrm{std}(\widehat{g}) = \sqrt{\frac{1}{N}\frac{1}{n-1}\sum_{j=1}^{N}\sum_{t=1}^{n}\left(\widehat{g}_j(y_{t-1}(j)) - \bar{\widehat{g}}_t\right)^2},$$

where $\bar{\widehat{g}}_t = \frac{1}{N}\sum_{j=1}^{N}\widehat{g}_j(y_{t-1}(j))$.

Table 4.4 Simulation Results for Case A and Case B

Case A	std($\widetilde{\alpha}$)	std($\widehat{g}$)	std($\widehat{\tau}$)	Case B	std($\widetilde{\alpha}$)	std($\widehat{g}$)	std($\widehat{\tau}$)
$n = 250$	0.0058	0.2304	0.6428	$n = 250$	0.1120	0.3305	0.2625
$n = 600$	0.0027	0.1944	0.6361	$n = 600$	0.0371	0.3255	0.1392
$n = 1000$	0.0018	0.1488	0.6005	$n = 1000$	0.0130	0.3106	0.1389

Table 4.4 also shows that the rate of $\widetilde{\alpha}$ to α is much faster than that of $\widehat{g}$ to g as shown in Theorem 3.1. Unlike Examples 4.1–4.3, the simulation study in Example 4.4 is more computationally intensive. This is because of the involvement of the nonparametric kernel estimation procedure and the cross–validation (CV) bandwidth selection method. Due to this, Table 4.4 provides only the simulation study results for the sample sizes of up to $n = 1000$. Meanwhile, we have only used the CV selection method in practice. Theoretical discussion about such an issue requires further study and is therefore left for future research.

Example 4.5 Finally, as a real data illustration, we now look at the 2–year (x_{1t}) and 30–year (x_{2t}) Australian government bonds, representing short–term and long–term series in the term structure of interest rates. The time frame of the study is January 1957 to March 2009, with 627 observations for each of x_{it}.

Similarly to Tsay (1998), we also employ the 3–month moving–average "spread" of logged interest rate as y_t, where

$$y_1 = s_1,\ y_2 = \frac{s_1 + s_2}{2} \ \text{ and } \ y_t = \frac{s_t + s_{t-1} + s_{t-2}}{3} \tag{4.14}$$

for $t \geq 3$, in which $s_t = z_{2t} - z_{1t}$ with $z_{it} = \ln(x_{it}) - \ln(x_{i,t-1})$. The plot of y_t is given in Figure 2 below.

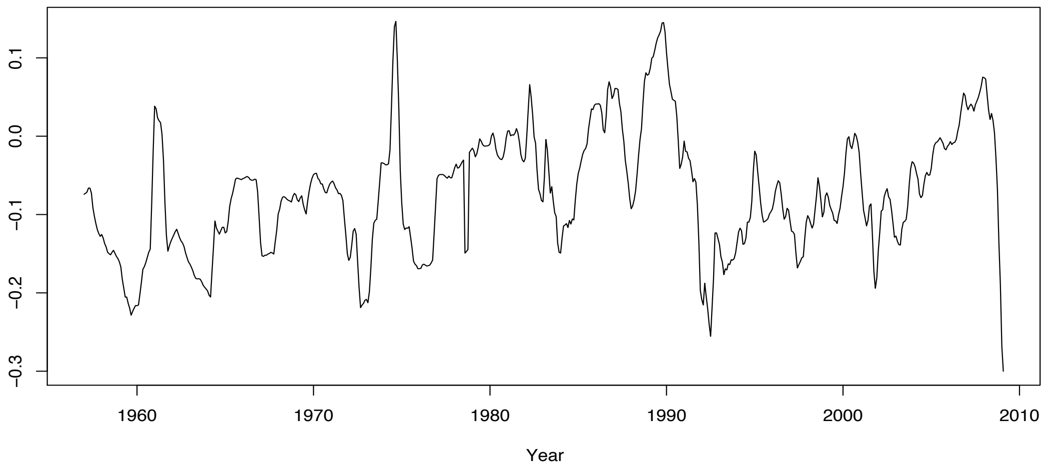

Figure 1: Plot of the time series y_t

Our estimation method also suggests a threshold model of the form

$$y_t = 0.8615\ y_{t-1}I(y_{t-1} \in C_\tau) + 0.9925\ y_{t-1}I\left(y_{t-1} \in C_\tau^c\right) + e_t, \tag{4.15}$$

where $C_\tau = [-0.128, -0.072]$ and $\widehat{\sigma}^2 = 3.1155 \times 10^{-4}$.

Model (4.15) therefore implies

$$y_t - y_{t-1} = 0.1385\ y_{t-1}I(y_{t-1} \in C_\tau) - 0.0075\ y_{t-1}I\left(y_{t-1} \in C_\tau^c\right) + e_t. \tag{4.16}$$

Model (4.16) shows that $\{y_t\}$ is nonstationary but does not necessary follow a random walk process, since the value of 0.1385 is significantly different from 0.

The finding in model (4.16) provides support from an empirical application point of view that there is some need to study a nonstationary threshold model of the form (1.1).

5 Conclusions and discussion

This paper has considered two classes of threshold autoregressive models with possible nonstationarity. The first one is a class of parametric threshold auto–regressive (TAR) models with possible nonstationarity. The slope parameters have been consistently estimated. The second class is a new class of semiparametric threshold auto–regressive (SEMI–TAR) models. We have estimated both the unknown slope parameter and unknown function using a semiparametrically consistent method.

One issue that has not been addressed is how to establish an asymptotic theory for $\widehat{\tau}$, a consistent estimate of τ, in this kind of nonlinear and nonstationary situation. While it is anticipated that an asymptotically normal estimator of τ may be established (similar to Theorem 2 of Chan 1993), detailed assumptions and rigorous proofs may involve both new tools and more technicalities and therefore are left for future research.

Another issue is possible extensions of the current discussion for the first–order univariate case to higher–order and vector models. If the latter is possible, one could introduce a class of threshold cointegration models with nonstationarity. Further discussion is also left for future research.

6 Acknowledgments

This work was started when the first author was visiting The University of Bergen in August/September in 2007. The authors would like to acknowledge constructive suggestions and comments from seminar and conference participants at Monash University, Singapore Management University, University of Technology Sydney, and the University of Hong Kong. Special thanks go to Shiqing Ling for making their working paper available to us.

The work of the first author was supported financially by The University of Bergen and the Australian Research Council Discovery Grants under Grant Numbers: DP0558602 and DP1096374. The second author would also like to acknowledge support from the Danish Research Council Grant Number: 2114-04-0001. The work of the third author was also supported financially by the Australian Research Council Discovery Grant under Grant Number: DP1096374.

Jiti Gao, School of Economics, The University of Adelaide, Adelaide SA 5005, Australia. Email: jiti.gao@adelaide.edu.au.

Dag Tjøstheim, Department of Mathematics, The University of Bergen, Bergen 5008, Norway. Email: dagt@mi.uib.no.

Jiying Yin, School of Economics, The University of Adelaide, Adelaide SA 5005, Australia. Email: jiying.yin@adelaide.edu.au.

7 Appendix A

In order to make this paper self–contained, we introduce some general results about β–null recurrent Markov chains in this appendix.

Let $\{y_t\}$ be a null recurrent Markov chain. We have n observations of this process and consider the sum

$$S_n(g) = \sum_{j=0}^{n} g(y_j, \ldots, _{j+r-1}) \tag{A.1}$$

for some function $g(\cdot)$. We let π be an invariant measure for $\{y_t\}$ and

$$S_n(g) = U_0 + \sum_{k=1}^{T(n)} U_k + U_{(n)} \tag{A.2}$$

be the decomposition of $S_n(g)$ as in (3.23) of Karlsen and Tjøstheim (2001) (hereafter KT). Moreover, $T(n)$ is the number of regenerations in the time interval $[0, n]$. We also need a notation for the moments w.r.t. the invariant measure π. Note from the decomposition (A.2) that

$$U_k = U_k(g) = \sum_{\tau_{k-1}+1}^{\tau_k} g(y_j, \ldots, y_{j+r-1}), \quad k = 1, 2, \ldots, \tag{A.3}$$

where the τ_k-s are regeneration times. The $U_k's$ are identically distributed and are $(r-1)$–dependent. (If $r = 1$ they are independent). If they exist, we denote the expectation and variance of these terms by $\mu(g) = E(U_k(g))$ and $\sigma^2(g) = \text{var}(U_k(g))$. Note that for $r = 1$, $\mu(g) = \int g(x)\pi_s(dx)$ and similarly for $\sigma^2(g)$. For $r > 1$

$$\mu(g) = \pi_s(g) \doteq \int \pi_s(dx_1)P(x_1, dx_2) \cdots P(x_{r-1}, dx_r)g(x_1, \ldots, x_r), \tag{A.4}$$

where s refers to the small function used in the minimization condition (see (3.4) of KT 2001) and $P(\cdot, \cdot)$ is the transition probability of the chain.

Finally, as in equation (4.4) of KT we introduce the notation

$$\bar{\sigma}^2 = \bar{\sigma}^2(g) = \sum_{k=-(r-1)}^{(r-1)} \text{cov}(U_{1+|k|}(g), U_1(g)). \tag{A.5}$$

We are now ready to formulate the lemmas:

Assumption A.1. Assume that the minorization condition ((3.4) of KT) is fulfilled and that $\{y_t\}$ is β–null recurrent as defined in Definition 3.2 and in Theorem 3.1 of KT. We let $u(n) = n^{\beta}L_s(n)$ where $0 < \beta < 1$ and the slowly varying function $L_s(n)$ is as in the tail condition (3.16) of KT.

Lemma A.1 Let Assumption A.1 hold. (i) *Let* $||g|| \in L^1_r(\pi_s)$ *and also the process have an arbitrary initial distribution* λ*. Then as* $n \to \infty$

$$\frac{S_n(g)}{T(n)} \to \pi_s(g) \quad \text{almost surely } (a.s.). \tag{A.6}$$

(ii) *Then for* n *large enough, the inequality* $n^{\frac{1}{2}-\epsilon_0} \leq T(n) \leq n^{\frac{1}{2}+\epsilon_0}$ *holds with probability one for some* $0 < \epsilon_0 < \frac{1}{4}$.

Proof. The proof of (i) follows from that of Lemma 3.2 of KT (2001) while the proof of (ii) follows from Lemma 3.4 of KT (2001).

Lemma A.2 *Let Assumption A.1 hold. If (i)* $\mu(|g|) < \infty$ *and (ii) there exists an* $m > 1$ *so that* $E\,|U(g) - \mu(g)|^{2m} \leq d_m$ *for some* $d_m > 0$*, then*

$$(\Delta_n, T_n) \to_{D^2} (B(M_\beta), M_\beta), \quad \textit{with } B \textit{ and } M_\beta \textit{ independent.} \tag{A.7}$$

where the symbol " $\to_{D^2}$ *" means weak convergence in cadlag space (see, for example, the appendix of KT 2001),* $T_n = \left\{\frac{T([nt])}{u(n)} : t \geq 0\right\}$, $\Delta_n(t) = u^{-1/2}(n)\bar{\sigma}^{-1}(g)\{S_{[nt]}(g) - \mu(g)T([nt])\}$, $[nt]$ *is the integer function and* $M_\beta(t)$ *is the Mittag-Leffler process as defined in KT on page 388.*

Proof. The proof is essentially the same as the proof of Theorem 4.1 in KT but much simpler. As in that proof one introduces the scaled variables

$$W_k(g) = \bar{\sigma}^{-1}(U_k(g) - \mu(g)). \tag{A.8}$$

(note that the existence of $\bar{\sigma}^2$ follows from condition ii), the definition of $\bar{\sigma}^2$ and the Schwartz inequality.) From condition ii) it also follows that there exists an $m > 1$ such that $E(W^{2m}) < d'_m$ for some constant d'_m from which

$$n^{-m} \sum_{k=1}^{[nt]} E(W_k^{2m}(g) \leq d'_m t n^{-m-1} = o(1). \tag{A.9}$$

It follows from standard limit theorems that

$$Q_n(t) \doteq n^{-1/2} \sum_{k=1}^{[nt]} W_k(g) \to_D B(t). \tag{A.10}$$

Tightness is then proved exactly as in KT (note that there is a misprint in the last formula on page 393 of KT: W_{2k-1} should be W_{2k-i}). It follows that the convergence can be strengthened to convergence in D^2. We can neglect the edge terms

$$\delta_{g,n}(t) \equiv u^{-1/2}(n)\bar{\sigma}^{-1}(g)\{U_0(t) + U_{(n)}(t)\}. \tag{A.11}$$

using the technique of part 2 of the proof of KT. The final part of the proof of KT only deals with the process T_n induced by the number of regenerations $T(n)$, and this is completely independent of the bandwidth considerations introduced in KT. The lemma follows.

The limit distribution in Lemma A.2 is non-Gaussian. However, as in Theorem 4.2 of KT (2001), a Gaussian distribution can be obtained by a stochastic normalization. We let $T_C(n)$ denote the number of visits of X_t to a small set C in the time period $[0, n]$. We have that $T_C(n)/T(n)$ converges with probability 1 to $\pi_s(C)$. We now have the following lemma.

Lemma A.3 *Assume that the conditions of Lemma A.2 hold and let C be a small set. Then*

$$T_C^{1/2}(n)\pi_s^{1/2}(C)\bar{\sigma}^{-1}(g)\{T_C^{-1}(n)S_n(g) - \pi_s^{-1}(C)\mu(g)\} \xrightarrow{d} N(0,1). \qquad \text{(A.12)}$$

In Lemma A.2 the process $B(M_\beta(t))$ enters. Concerning the existence of moments we have the following lemma.

Lemma A.4 *Let k be a positive integer. Then, $E\left[B(M_\beta(t))^{2k+1}\right] = 0$, and*

$$E\left[B(M_\beta(t))^{2k}\right] = (2k-1)(2k-3)\cdots 1 \cdot t^{\beta k}/(\Gamma(1+\beta))^k. \qquad \text{(A.13)}$$

Proof. We use double expectation and the independence of the processes B and M_β to obtain

$$\begin{aligned} E\left[B(M_\beta(t))^{2k}\right] &= E[E(B(M_\beta(t))^{2k}|M_\beta(t))] = E[(2k-1)(2k-3)\cdot 1 M_\beta(t)^k] \\ &= (2k-1)(2k-3)\cdots 1 \cdot t^{\beta k}/(\Gamma(1+\beta))^k, \end{aligned}$$

so that all moments exist.

8 Appendix B

This appendix provides the detailed proofs of Lemmas 2.1 and 2.3 given in Section 2 and Lemma 3.1 given in Section 3.

Proof of Lemma 2.1: Since the proof follows from that of Lemma 3.1 for the case of $g(y) = \alpha_1 y$, we omit the detail here.

Proof of Lemma 2.3: Recall from Lemma 2.2 that as $n \to \infty$

$$Q_n(r) = \frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]} u_t + \frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]} e_t \to_D \sigma B(r) + M_{\frac{1}{2}}(r)\ m_u \equiv Q(r) \qquad \text{(B.1)}$$

uniformly in $0 < r \leq 1$.

We then start by proving (2.10) and (2.11). It follows from Lemma A.1(i) that as $n \to \infty$

$$\frac{1}{T(n)} \sum_{t=1}^{[nr]} y_{t-1}^{i} I[y_{t-1} \in C_\tau] \to_P \mu_i \quad \text{for } i = 1, 2, \tag{B.2}$$

which implies the proof of (2.10).

Let $b_{nt} = \frac{y_{t-1} I[y_{t-1} \in C_\tau]}{\sqrt{\sum_{s=1}^{n} y_{t-1}^2 I[y_{t-1} \in C_\tau]}}$. In order to prove (2.12), it suffices to show that as $n \to \infty$

$$\sum_{t=1}^{n} \left(\frac{y_{t-1} I[y_{t-1} \in C_\tau]}{\sqrt{\sum_{s=1}^{n} y_{t-1}^2 I[y_{t-1} \in C_\tau]}} \right) e_t \xrightarrow{d} N\left(0, \sigma^2\right), \tag{B.3}$$

which follows from a conventional central limit theorem (see, for example, Corollary 3.1 of Hall and Heyde 1980) for the case where $\{e_t\}$ is a sequence of i.i.d. random errors by verifying

$$\sum_{t=1}^{n} b_{nt}^2 E[e_t^2] \quad \to_P \quad \sigma^2 \quad \text{and} \quad \sum_{t=1}^{n} b_{nt}^4(y) E[e_t^4] \to_P 0. \tag{B.4}$$

The first part follows automatically from $\sum_{t=1}^{n} b_{nt}^2 = 1$ while the second part of (B.4) follows from

$$\begin{aligned} \frac{\sum_{t=1}^{n} y_{t-1}^4 I[y_{t-1} \in C_\tau]}{\left(\sum_{t=1}^{n} y_{t-1}^2 I[y_{t-1} \in C_\tau]\right)^2} &= \frac{1}{T(n)} \frac{\frac{1}{T(n)} \sum_{t=1}^{n} y_{t-1}^4 I[y_{t-1} \in C_\tau]}{\left(\frac{1}{T(n)} \sum_{t=1}^{n} y_{t-1}^2 I[y_{t-1} \in C_\tau]\right)^2} \\ &= O_P\left(\frac{1}{T(n)}\right) = o_P(1) \end{aligned} \tag{B.5}$$

using Lemma 2.1 and then Lemma A.1(i).

We then prove equations (2.11) and (2.13). Recall that $y_t = \sum_{s=1}^{t} u_s + \sum_{s=1}^{t} e_s$. We now start to prove (2.11). Let $X_n(r) = \frac{1}{\sqrt{n}} \sum_{s=1}^{[nr]} (u_s + e_s)$. By the same arguments as in the proof of Theorem 3.1(a) of Phillips (1987), we have as $n \to \infty$

$$\begin{aligned} \frac{1}{n^2} \sum_{t=1}^{n} y_{t-1}^2 I[|y_{t-1}| \in D_\tau] &= \frac{1}{n^2} \sum_{t=1}^{n} y_{t-1}^2 - \frac{1}{n^2} \sum_{t=1}^{n} y_{t-1}^2 I[y_{t-1} \in C_\tau] \\ &= \frac{1}{n^2} \sum_{t=1}^{n} y_{t-1}^2 + o_P(1) \qquad \text{(B.6)} \\ &= \int_0^1 X_n^2(r) dr + o_P(1) \qquad \text{(B.7)} \\ &\xrightarrow{d} \int_0^1 Q^2(u) du, \end{aligned}$$

where (B.6) follows from the fact that $\frac{1}{T(n)} \sum_{t=1}^{n} y_{t-1}^2 I[y_{t-1} \leq \tau] \to_P \mu_2$ by Lemmas 2.1 and A.1(i), and Lemma 2.2 has been used in (B.7). The proof of (2.11) is now completed.

Recall $\eta_t = u_t + e_t$. We finally prove (2.13). Note that

$$\begin{aligned}
\frac{1}{n}\sum_{t=1}^{n} y_{t-1}e_t I[y_{t-1}\in D_\tau] &= \frac{1}{n}\sum_{t=1}^{n} y_{t-1}e_t - \frac{1}{n}\sum_{t=1}^{n} y_{t-1}e_t I[y_{t-1}\in C_\tau] \\
&= \frac{1}{n}\sum_{t=1}^{n} y_{t-1}e_t + o_P(1) \qquad \text{(B.8)} \\
&= \frac{1}{n}\sum_{t=1}^{n} y_{t-1}\eta_t - \frac{1}{n}\sum_{t=1}^{n} y_{t-1}u_t + o_P(1) \\
&= \frac{1}{n}\sum_{t=1}^{n} y_{t-1}\eta_t - \frac{(\alpha_1 - 1)}{n}\sum_{t=1}^{n} y_{t-1}^2 I[y_{t-1}\in C_\tau] + o_P(1) \\
&= \frac{1}{n}\sum_{t=1}^{n} y_{t-1}\eta_t + o_P(1) \qquad \text{(B.9)} \\
&= \frac{1}{n}\sum_{t=2}^{n}\left(\sum_{s=1}^{t-1}\eta_s\right)\eta_t + o_P(1) \\
&= \frac{1}{2n}\sum_{t=1}^{n}\left(\sum_{s=1}^{n}\eta_s\right)\eta_t - \frac{1}{2n}\sum_{t=1}^{n}\eta_t^2 + o_P(1) \\
&\xrightarrow{d} \frac{1}{2}\left(Q^2(1) - \sigma^2\right), \qquad \text{(B.10)}
\end{aligned}$$

where $Q(r) = \sigma B(r) + M_{\frac{1}{2}}(r)\ m_u$, equation (2.12) has been used in (B.8), Lemma A.1(i) has been used in (B.9) and Lemma 2.2 has been used in (B.10).

Therefore, the proof of Lemma 2.3 is completed.

Proof of Lemma 3.1: We shall use Theorem 3.1 of Karlsen and Tjøstheim (2001) (KT) to show that $\{y_t\}$ of (B.11) below is β–null recurrent with $\beta = \frac{1}{2}$ as in the random walk case. Recall the structure of model

$$y_t = g(y_{t-1})\ I[y_{t-1}\in C_\tau] + \alpha y_{t-1}\ I[y_{t-1}\in D_\tau] + e_t, \qquad \text{(B.11)}$$

where C_τ is either a compact subset of R^1 or $C_\tau = (-\infty, \tau]$ or $C_\tau = [\tau, \infty)$ and D_τ is the complement of C_τ.

Then the process $\{y_t\}$ is null recurrent (see Appendix B2 of Meyn and Tweedie 1994). Note that the proof in that book is easily modified to the situation of model (3.1) and a bounded $g(\cdot)$, see the remark at the bottom of page 303). This implies that there exists an invariant measure π and that the process recurs with probability 1, but with infinite expected recurrence time. The next step is to establish that the minorization condition (3.4) of KT holds. We first look at the case where C_τ is compact. Then the construction of Example 3.1 of KT can be used. The minorization condition then follows directly from Example 3.1 of KT with $f(x)$ of that paper given by $f(x) = g(x)I(x\in C) + x(1 - I(x\in C))$ with $C = C_\tau$ since it is assumed that the distribution of e_t is absolutely continuous with respect to Lebesgue measure. The fact that the minorization condition holds means that the split chain can be used, and as in KT, $\{S_\alpha\}$ is used to denote the recurrence times. They are iid and because of null recurrence $P(S_\alpha > n)$ must be asymptotically larger than $L_s(n)/n^{1+\varepsilon}$, where $L_s(n)$ is slowly varying and $\varepsilon > 0$.

We are free to choose any small set K_0 as a set of regeneration in (B.11). We choose K_0 as C_τ if C_τ is compact. This is because compact sets are small if the distribution of $\{e_t\}$ is absolutely continuous with respect to Lebesgue measure. There are two ways in which $\{y_t\}$ may regenerate:

1. The process $\{y_t\}$ does not leave the set C_τ before it regenerates. Let A_n be the event that y_t stays in C_τ in at least $n+i$ steps and regenerates at step $n+i$ for $i \geq 1$. The time S' to regeneration satisfies

$$P(S' > n) \quad = \quad P(A_n) \leq \sum_{i=n+1}^{\infty} \rho^i \leq M\rho^{n+1} = o(n^{-\gamma})$$

for any $0 < \gamma < 1$, where $0 < M < \infty$ is an absolute constant. Here $\rho = \rho_1\rho_2$ where $\rho_1 = \sup_{x \in C_\tau} P(x, C_\tau)$, where $P(\cdot,\cdot)$ is the transition probability of the chain. Note that $0 < \rho_1 < 1$. Similarly, $\rho_2 = 1 - a$, where a is defined in Example 3.1 of KT and $0 < a < 1$. From this, comparing to $O(n^{-\gamma})$, it is seen that these recurrence times do not contribute to the tail bahaviour of S_α.

2. The process $\{y_t\}$ does leave the set C_τ before it regenerates. Outside the set C_τ, $\{y_t\}$ behaves as a random walk, and therefore according to the paper by Kallianpur and Robbins (1954) and the fact that what goes on inside the set C_τ can be neglected compared to a probability of order $O(n^{-1/2})$, if S'' is such a recurrence time, $P(S'' > n) = O(n^{-1/2})$. This means that the tail behaviour of S_α is controlled by the tail behaviour of S'' and that $\{y_t\}$ is β–null recurrent with $\beta = \frac{1}{2}$.

Next we look at the case where $C_\tau = (-\infty, \tau]$ or $[\tau, \infty)$. Without loss of generality, we may assume $C_\tau = (-\infty, \tau]$ In this case we let the set of regeneration be the set $K_0 = [\tau', \tau]$ where τ' can be taken to be any real number smaller than τ. From Assumption 3.1(iv), we may assume that $\{y_t\}$ behaves as a stationary process to the left of τ' and like a random walk to the right of τ.

Again it follows from Appendix B2 of Meyn and Tweedie (1994) that $\{y_t\}$ is null recurrent. (In fact Meyn and Tweedie has $g(\cdot)$ linear). By the same reasoning as above, option 2 then splits into two cases: 2a) where $\{y_t\}$ leaves K_0 going to the stationary part of $\{y_t\}$ and then does not enter the random walk part before it regenerates. The associated recurrence time S''' has tail behaviour controlled by $P(S''' > n) = O(L_s(n)/n^{1+\varepsilon})$. The possibility 2b) is the case where the random walk part is visited before it regenerates, but here $P(S'' > n) = O(n^{-1/2})$, as time spent in the stationary part and in the set C_τ can be neglected as far as tail behaviour is concerned. This implies again that $\{y_t\}$ is β–null recurrent with $\beta = \frac{1}{2}$.

Remark: The process $\{y_t\}$ may even be explosive on the left–hand side, if it explodes in the direction of K_0 and the random walk regime. This is illustrated by the simulated example in Example 4.3.

REFERENCES

Billingsley, P. (1968). *Convergence of Probability Measures.* John Wiley, New York.

Cai, Z., Li, Q. and Park, J. (2009) Functional–coefficient models for nonstationary time series data. *Journal of Econometrics* **148** 101–113.

Caner, M. and Hansen, B. E. (2001). Threshold autoregression with a unit root. *Econometrica* **69** 1555–1596.

Chan, K. S. (1991). Testing for threshold autoregression. *Annals of Statistics* **18** 1886–1894.

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

Chan, K. S. and Tsay, R. S. (1998). Limiting properties of the least squares estimator of a continuous threshold autoregressive model. *Biometrika* **85** 413–426.

Chen, J., Gao, J. and Li, D. (2008). Semiparametric regression estimation in null recurrent time series. Working paper at http://www.adelaide.edu.au/directory/jiti.gao.

Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods.* Springer, New York.

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

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

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

Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and Its Application.* Academic Press, New York.

Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* **64** 413–430.

Hansen, B. E. (2000). Sample splitting and threshold estimation. *Econometrica* **68** 575–603.

Kallianpur, G. and Robbins, H. (1954). The sequence of sums of independent random variables. *Duke Mathematical Journal* **21** 285–307.

Karlsen, H. A. and Tjøstheim, D. (2001). Nonparametric estimation in null recurrent time series. *Annals of Statistics* **29** 372-416.

Karlsen, H. A., Mykelbust, T. and Tjøstheim, D. (2007). Nonparametric estimation in a nonlinear cointegration type model. *Annals of Statistics* **35** 252–299.

Liu, W., Ling, S. and Shao, Q. (2009). On nonstationary threshold autoregressive models. Working paper available from http://www.adelaide.edu.au/directory/jiti.gao.

Meyn, S. P. and Tweedie, R. L. (1994). *Markov Chains and Nonnegative Operators*. Cambridge University Press.

Park, J. and Phillips, P. C. B. (2001). Nonlinear regressions with integrated time series. *Econometrica* **69** 117-162.

Pham, D. T., Chan, K. S. and Tong, H. (1991). Strong consistency of the least squares estimator for a non–ergodic threshold autoregressive model. *Statistica Sinica* **1** 361–369.

Phillips, P. C. B. (1987). Time series regression with a unit root. *Econometrica* **55** 277–301.

Phillips, P. C. B. and Park, J. (1998). Nonstationary density estimation and kernel autoregression. Cowles Foundation Discussion Paper, No. 1181, Yale University.

Tsay, R. S. (1998). Testing and modeling multivariate threshold models. *Journal of the American Statistical Association* **93** 1188–1202.

Tong, H. (1983). Threshold Models in nonlinear Time Series Analysis. *Lecture Notes in Statistics* **21**. Springer-Verlag, New York

Tong, H. (1990). *Nonlinear Time Series. A Dynamical System Approach.* Oxford University Press, New York.

Tong, H. and Lim, K. S. (1980). Threshold autoregression, limit cycles and cyclical data (with Discussion). *Journal of the Royal Statistical Society Series B* **42** 245–292.

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. (2010). Structural nonparametric cointegrating regression. Cowles Foundation Discussion Paper No. 1657. Forthcoming in *Econometrica*.