ESTIMATION OF A THRESHOLD AUTOREGRESSIVE MODEL UNDER MISSPECIFICATION

MYUNG HWAN SEO

Abstract. This paper obtains an asymptotic distribution for the least squares estimator of the self-exciting threshold autoregressive model, which was introduced by Tong (1983), under the assumption that the model is an approximation to a more complicated system. Under some moderate assumptions on the true data generating process, it is shown that the least squares estimator is mere cube root n-consistent to a pseudo true value, where n is the sample size, and the limit distribution is characterized by the minimizer of a non-zero-mean Gaussian process. This is in sharp contrast to the standard super-consistency of threshold estimates. Some univariate economic time-series data are examined to demonstrate the slower convergence. We also show that the smoothed least squares estimater can improve upon the rate arbitrarily close to square root n under some model smoothness assumption and yields the asymptotic normality.

1. Introduction

The self-exciting threshold autoregressive (SETAR) model, which was introduced and popularized by Tong (1983, 1990), can generate many important features that a useful nonlinear time series should produce, with parsimony. It is commonly estimated by least squares (LS) principle. Suppose that a sample $\{y_t\}_{t=0}^n$ is fit by the SETAR model of order 1

$$y_t = (\mu_1 + \alpha_1 y_{t-1})\, 1\{y_{t-1} \le \gamma\} + (\mu_2 + \alpha_2 y_{t-1})\, 1\{y_{t-1} > \gamma\} + e_t. \tag{1}$$

Let $\beta = (\beta_1', \beta_2')'$, where $\beta_i = (\mu_i, \alpha_i)'$ for $i = 1, 2$, and $\theta = (\beta', \gamma)' \in \Theta = B \times \Gamma$, which are compact. We will also use $\delta = \beta_2 - \beta_1$ to denote the change in the paramter values. Define

$$\pi_\gamma(y) = (1\{y \le \gamma\}, 1\{y > \gamma\})' \otimes (1, y)'.$$

Then, the LS estimate is a minimizer of

$$\mathbb{S}_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} \left(y_t - \beta' \pi_\gamma(y_{t-1})\right)^2.$$

For a fixed γ, the LS estimate of β is

$$\hat{\beta}(\gamma) = \left(\sum_{t=1}^{n} \pi_\gamma(y_{t-1})\, \pi_\gamma(y_{t-1})'\right)^{-1} \left(\sum_{t=1}^{n} \pi_\gamma(y_{t-1})\, y_t\right).$$

Date: March 2011; **Preliminary and Comments Welcome**.

Key words and phrases. Threshold, Autoregression, Misspecification, Pseudo True Values, Least Squares, Smoothed Least Squares, Cube Root Asymptotics.

This paper benefits from disucssion with Howell Tong and Bruce Hansen. The author thanks participants in Recent Advances in Time Series Analysis Workshop and seminars in LSE and SNU..

Let $\hat{\mathbb{S}}_n(\gamma) = \mathbb{S}_n\left(\hat{\beta}(\gamma), \gamma\right)$, then

$$\hat{\gamma} = \underset{\gamma \in \Gamma}{\text{argmin}}\ \hat{\mathbb{S}}_n(\gamma)$$

and

$$\hat{\beta} = \beta(\hat{\gamma}).$$

In finite sample, $\hat{\gamma}$ is not unique but an interval thus defined as the minimum in the interval.

The sampling distribution of $\hat{\theta}$ has been approximated by various asymptotic distributions. Chan (1993) established the n-consistency of $\hat{\gamma}$, asymptotic distribution of $\hat{\gamma}$, and the asymptotic normality of $\hat{\beta}$ assuming that (1) is the true data generating process of $\{y_t\}$ and the regression function has a jump at the threshold value. Gonzalo and Wolf (2005) derived the asymptotic normality of $\hat{\theta}$ when there is no jump, while Chan and Tsay (1998) showed the asymptotic normality of the restricted least squares of θ that imposes the continuity of the regression function. On the other hand, Hansen (2000) obtained another asymptotic distribution for $\hat{\gamma}$ under a diminishing threshold assumption, where δ vanishes as $n \to \infty$. It is viewed as a technical device to obtain an asymptotic distribution, which enables an asymptotic inference without resorting to a resampling method, but maybe not as a true data generating process. Common in Chan's and Hansen's results is the asymptotic independence between $\hat{\beta}$ and $\hat{\gamma}$, which is convenient but may not be a good approximation to the true sampling distribution.

This paper explores the asymptotic property of $\hat{\theta}$ in the spirit of Huber (1967) and White (1982). The former showed that under general conditions the maximum likelihood estimator for a class of smooth likelihood functions, not necessarily gaussian, converges to a well-defined limit, i.e. the pseudo true value, even if the likelihood function employed in the estimation is not true. The latter established the asymptotic normality of the quasi maximum likelihood estimator extending Huber's result. As in the discussion of White (1982), we interpret the limit of $\hat{\theta}$ as the projection coefficients in the projection of the true unknown function on the space of the piecewise linear functions in terms of mean squared error. We show that under certain regularity conditions the estimator $\hat{\theta}$ converges to the pseudo true value at the rate of $n^{1/3}$ and the asymptotic distribution is characterized by the minimizer of a gaussian process. This extends the cube-root asymptotics of Kim and Pollard (1990) for a dynamic model and the asymptotic distribution can be consistently approximated by subsampling, see .e.g. Politis, Romano, and Wolf (1999).

We illustrate the practical relevance of our result in Section 3 using many economic and financial time series data, which have been examined in the SETAR framework. In particular, the convergence rates of $\hat{\gamma}$ are estimated by a method proposed by Politis et al. (1999). Among four series examined, only one series has an estimate closer to n and the other three have estimates closer to $n^{1/3}$. This highlights the practical relevance of the possible misspecification.

Section 4 analyses the smoothed least squares estimator proposed by Seo and Linton (2007), which has shown the asymptotic normality of the estimator under correct specification. We reestablish the asymptotic normality under misspecification but the convergence rate turns out to be slower. We provide a range of admissible rates for the smoothing parameters, where the estimator has a proper asymptotic normality. Section 5 concludes. Proofs of main theorems are relegated to Appendix.

2. Asymptotic Distribution Under Misspecification

Assumption 1. *Assume that*

(a) $\{y_t\}$ *is strictly stationary, ergodic, and has a density function* $p(\cdot)$ *that is continuous and positive everywhere in* $\mathbb{R}$.

(b) *Let* $S(\theta) = \mathrm{E}\left(y_t - \beta' \pi_\gamma(y_{t-1})\right)^2 < \infty$ *for all* $\theta \in \Theta$. *Then, there exists* θ_0 *that minimizes* $S(\theta)$ *uniquely.*

Assumption 1 is a minimal set of high-level assumptions that yields the consistency of $\hat{\theta}$ to the population projection coefficients θ_0. That is, β_0 is interpreted as the projection coefficients in each subsample as in the linear projection and γ provides best splitting of the sample in terms of mean squared error. In the likelihood setup, White (1982) formalizes the information theoretic interpretation of the pseudo true value, which was implicit in Berk (1966, 1970). The global identification condition in Assumption 1 (b) might be stated more specifically when more structure is imposed on the dynamics of the process y_t. We provide an examples of y_t that satisfies Assumption 1.

Lemma 1. *Under Assumption 1,* $\hat{\theta} \xrightarrow{p} \theta_0$.

Now, we impose more structure in the process $\{y_t\}$ to obtain an asymptotic distribution for the LS estimates.

Assumption 2. *Assume that*

(a) *The process* $\{y_t\}$ *is a Markov process such that*

$$y_t = f(y_{t-1}) + \varepsilon_t, \tag{2}$$

where f *is continuously differentiable with bounded derivatives and* $\{\varepsilon_t\}$ *is a sequence of independent and identically distributed random variables such that* $\mathrm{E}(\varepsilon_t) = 0$, $\mathrm{E}\left(\varepsilon_t^2\right) = \sigma^2$, $\mathrm{E}\left|\varepsilon_t\right|^{2+\zeta} < \infty$ *for some* $\zeta > 0$.

(b) $\{y_t\}$ *is* ρ*-mixing with* ρ*-mixing coefficients satisfying* $\sum_{m=1}^{\infty} \rho_m^{1/2} < \infty$ *and* $\mathrm{E}\left|y_t\right|^{4+\zeta}$ *for some* $\zeta > 0$.

(c) $S(\theta)$ *is twice continuously differentiable at* θ_0 *with a positive definite second derivative matrix.*

Remarks

1. The existence of a stationary solution for the system (2) is well-known, see e.g. Proposition 6 of Wu (2007) , which mainly requires some Lipschitz continuity of f. See also Meyn and Tweedie (1993) for mixing properties of Markov chains.

2. The existence of a unique β that minimizes $S(\theta)$ for a given γ is obvious but that of a unique γ may depend on f. However, the existence of such an f is straightforward. For instance, consider

$$y_t = \alpha y_{t-1} + \delta y_{t-1}\left(1 + e^{-\rho(y_{t-1}-\gamma)}\right)^{-1} + \varepsilon_t.$$

Figure 1 plots the profiled $S(\gamma)$ with $\alpha = -0.3, \delta = 0.8, \rho = 50$, and $\gamma = 1$, which shows $S(\gamma)$ is globally minimized at $\gamma = 1$, whereas Figure 2 depicts the fitted regression function compared to the true regression function.

Figure 1 and 2 about here

3. Condition (c) is useful to obtain an asymptotic distribution. In particular, it requires that

$$-\delta_0' \begin{pmatrix} 1 \\ \gamma_0 \end{pmatrix} \left(f'(\gamma_0) - (\alpha_{10} + \alpha_{20})/2\right) < 0, \tag{3}$$

which in turn implies that the threshold regression function is discontinuous. Note that the term $(1,\gamma_0)\,\delta_0$ measures the size of jump in the pseudo regression function. It is essential to the subsequent asymptotic analysis.

The modulus of continuity of $\mathbb{S}_n$ at $\theta=\theta_0$ is smaller under the current setup than under the correctly specified case. This results in much slower convergence rate as shown in the subsequent lemma.

Lemma 2. *Under Assumption 1 and 2,* $\hat{\theta}=\theta_0+O_p\left(n^{-1/3}\right)$.

The following theorem presents the asymptotic distribution of $\hat{\theta}$.

Theorem 3. *Let B_1 and B_2 be two independent standard Brownian motions and define*

$$\begin{aligned} G(\theta) &= \sqrt{4\sigma^2\omega^2 p_0}\,\left(B_1(\gamma)\,1\{\gamma\leq 0\}+B_2(\gamma)\,1\{\gamma>0\}\right) \\ V(\theta) &= p_0\omega\left[\phi\gamma^2-2\gamma\left(\beta_1 1\{\gamma\leq 0\}+\beta_2 1\{\gamma>0\}\right)'\begin{pmatrix}1\\ \gamma_0\end{pmatrix}\right]+\beta' M\beta, \end{aligned}$$

where $M=\mathrm{E}\left(\pi_{\gamma_0}\pi_{\gamma_0}'\right)$, $p_0=p(\gamma_0)$, $\omega=-(1,\gamma_0)\,\delta_0$, *and* $\phi=f'(\gamma_0)-(\alpha_{10}+\alpha_{20})/2$. *Then, under Assumption 1 and 2,*

$$n^{1/3}\left(\hat{\theta}-\theta_0\right)\xrightarrow{d}\underset{\theta}{\operatorname{argmin}}\,G(\theta)+V(\theta).$$

Proof. See Appendix. □

Remarks

1.Since $\omega\phi<0$ due to Assumption 2 (c), that is, (3), the limit process is uniquely minimized with respect to γ almost surely, see e.g. Kim and Pollard (1990). With regard to β, the process is quadratic and the minimizer has a closed-form. Thus, the limit process can be represented as only a function of γ after concentration. Also note that M is a block diagonal matrix whose diagonal elements are $\mathrm{E}\,(1,y_{t-1})\,(1,y_{t-1})'\,1\{y_{t-1}\leq\gamma_0\}$ and $\mathrm{E}\,(1,y_{t-1})\,(1,y_{t-1})'\,1\{y_{t-1}>\gamma_0\}$.

2. The asymptotic distribution is not pivotal and there seems to be no obvious way to studentize the estimator to construct an asymptotically pivotal statistic. However, the asymptotic distribution can be simulated by estimating the finite-dimensional unknown quantities. Alternatively, the subsampling can estimate the asymptotic distribution consistently, see e.g. Politis, Romano, and Wolf (1999). It can aslo estimate the asymptotic distribution obtained in Chan (1993) under the correctly specified threshold autoregression. Thus, the subsampling is an inference tool, which is valid under both correctly- and mis-specified models. In this case, we may estimate the convergence rate as well using the subsampling.

3. Examples

In this application, we revisit four popular time series data, which have been extensively studied and often fit by the SETAR model. Then, we estimate the convergence rate of the threshold estimate $\hat{\gamma}$ to see how the data match with the different asymptotic theories. Since the change of convergence rate is arguably the most striking feature under misspecification, this exercise would provide a good insight on how plausible the misspecification is with the real data.

First of the series is the annual sunspot means as reported in Appendix 3 of Tong (1990), for the period of 1700-1988, and transformed by the square root transformation, $y_t=2\left(\sqrt{1+y_t^*}-1\right)$, where y_t^* is the original series, following Ghaddar and Tong (1981). This series has been analyzed by many. A common specification is the SETAR model with lag order 11 and the threshold variable being y_{t-2}, see e.g. Tong and Lim (1980) and Hansen (1999). We exercise this specification.

US monthly Industrial Production Series is the second. The data set is taken from Hansen (1999), which spans from 1960.01 to 1998.09 and is transformed by taking annualized growth rate, $y_t = 100 * \left(\ln y_t^* - \ln y_{t-12}^*\right)$. As in Hansen, the SETAR model is fit with lag order 16 while threshold variable being y_{t-6}.

US GNP series is another series intensively examined by e.g. Potter (1995). Our data are for the period 1947 Q2-2007 Q4 and we set lag order at 5 and y_{t-2} as the threshold variable following Potter. Finally, we consider quarterly US civilian unemployment rate series from 1948 to 1993. It was analyzed by Chan and Tsay (1998) among others. In particular, we take growth of quarterly averages to fit the SETAR model with lag order 2 and y_{t-2} as threshold variable.

Plots of the four series are given in Figure 3-6.

Figure 3 - 6 about here

The convergence rate $n^{-\alpha}$ of $\hat{\gamma} - \gamma_0$ is estimated by the method proposed by Politis, Romano, and Wolf (1999 Section 8.2). Given distinct block sizes $b_1, ..., b_J$, we estimate the sampling distribution of the un-scaled statistic $\left(\hat{\theta}_b - \hat{\theta}_n\right)$. Since $\hat{\theta}_b - \hat{\theta}_n = O_p\left(b^{-\alpha}\right)$, they propose to estimate α by the linear regression of quantiles of $\log\left|\hat{\theta}_b - \hat{\theta}_n\right|$ on the block sizes b_js. Trying to distinguish the continuous SETAR model from the discontinuous one, Gonzalo and Wolf (2005) performed simulation study to guide the choice of tuning parameters, which we follow. To be more specific, let $J = 4$ and $b_j = n^{a_j}$, where $a_j = 0.8\left(1 + \ln\left(j+1\right)/5\right)/\ln 100$, and $j = 1, ..., J$. Further, define a vector q of 4 evenly spaced values between 0.7 and 0.99. Then, we consider two different regressions whose dependent variables are, respecitvely, the averages of the logarithms of q quantiles of $\left|\hat{\theta}_{b_j} - \hat{\theta}_n\right|$ for each j and those of interval lengths of q and $1-q$ quantiles of $\left(\hat{\theta}_b - \hat{\theta}_n\right)$. The regressors are the constant and b_j in both regressions.

The estimated coefficients of b_j of all the regressions are reported in Table 1. As our scenario concerns two possible values of $\alpha = 1/3$ and 1, it appears reasonable to set $\hat{\alpha} = 1/3$ if the estimated coefficient is below 2/3, which is the half point between 1/3 and 1, and $\hat{\alpha} = 1$ otherwise. Then, there is one case of $\hat{\alpha}$ being 1 and are three cases of $\hat{\alpha}$ being 1/3. Despite the fact that the estimation of the convergence rate is difficult, this result comes as a surprise and is indicative of that the misspecification appears prominent.

	Sunspots	Industrial Prod.	US GNP	Unemployment
interval	0.105367	0.848217	0.109178	0.131173
absolute value	0.288774	0.844709	0.565566	0.090619

Table 1: Estimated Convergence Rates of the threshold parameter estimates. The reported values are estimated α, where $\hat{\gamma} - \gamma_0 = O_p\left(n^{-\alpha}\right)$.

4. Smoothed Least Squares

In the smoothed least squares (SLS) estimation, proposed by Seo and Linton (2007), the indicator function in $\mathbb{S}_n$ is replaced by a smooth bounded function $\mathcal{K}$ satisfying that

$$\lim_{s \to -\infty} \mathcal{K}\left(s\right) = 0, \lim_{s \to +\infty} \mathcal{K}\left(s\right) = 1.$$

Two versions are proposed. Specifically, let

$$\chi_{\gamma}(y) = \left(\left(1 - \mathcal{K}_{\gamma}(y), \mathcal{K}_{\gamma}(y)\right) \otimes \begin{pmatrix} 1 \\ y \end{pmatrix} \right),$$

where $\mathcal{K}_{\gamma}(y) = \mathcal{K}\left(h_n^{-1}(y-\gamma)\right)$ and $h_n \to 0$ as $n \to \infty$, and define

$$\mathbb{S}_n^{+}(\theta) = \frac{1}{n}\sum_{t=1}^{n}\left(y_t - \beta' \chi_{\gamma}(y_{t-1})\right)^2.$$

Here the dependence of χ on n is suppressed to ease notation. Then the first estimator is defined as

$$\hat{\theta}^{+} = \underset{\theta}{\operatorname{argmin}}\, \mathbb{S}_n^{+}(\theta).$$

The other version is based on the observation that

$$\mathbb{S}_n(\theta) = \frac{1}{n}\sum_{t=1}^{n}\left(y_t - x_t'\beta_1\right)^2 + \frac{1}{n}\sum_{t=1}^{n}\left\{\left(x_t'\delta\right)^2 - 2x_t'\delta\left(y_t - x_t'\beta_1\right)\right\} 1\left\{y_{t-1} > \gamma\right\},$$

where $x_t = (1, y_{t-1})$, since the square of the indicator is the indicator itself. Let

$$\mathbb{S}_n^{*}(\theta) = \frac{1}{n}\sum_{t=1}^{n}\left(y_t - x_t'\beta_1\right)^2 + \frac{1}{n}\sum_{t=1}^{n}\left\{\left(x_t'\delta\right)^2 - 2x_t'\delta\left(y_t - x_t'\beta_1\right)\right\} \mathcal{K}_{\gamma}(y_{t-1}),$$

and the second SLS estimator is given by

$$\hat{\theta}^{*} = \underset{\theta}{\operatorname{argmin}}\, \mathbb{S}_n^{*}(\theta).$$

This also allows for concentration as in $\hat{\theta}^{+}$, that is, for a given γ,

$$\begin{bmatrix} \hat{\beta}_1^{*}(\gamma) \\ \hat{\delta}_n^{*}(\gamma) \end{bmatrix} = \left(\sum_{t=1}^{n} \begin{bmatrix} 1 & \mathcal{K}_{\gamma}(y_{t-1}) \\ \mathcal{K}_{\gamma}(y_{t-1}) & \mathcal{K}_{\gamma}(y_{t-1}) \end{bmatrix} \otimes x_t x_t' \right)^{-1} \sum_{t=1}^{n} \begin{bmatrix} x_t y_t \\ \mathcal{K}_{\gamma}(y_{t-1})\, x_t y_t \end{bmatrix}.$$

The following conditions are imposed on $\mathcal{K}$. The integral $\int$ is taken over the real line $\mathbb{R}$ unless specified otherwise and the first and second derivatives of $\mathcal{K}$ are denoted by $\mathcal{K}'$ and $\mathcal{K}''$, respectively.

Assumption 3. *Assume that*

(a) *$\mathcal{K}$ is twice differentiable everywhere, $\mathcal{K}'$ is symmetric around zero, $|\mathcal{K}'(\cdot)|$ and $|\mathcal{K}''(\cdot)|$ are uniformly bounded, and: $\int |\mathcal{K}'(v)|^4 dv < \infty$, $\int |\mathcal{K}''(v)|^2 dv < \infty$, $\int \left|v^2 \mathcal{K}''(v)\, dv\right| < \infty$.*

(b) *For some integer $\ell \geq 2$, each integer i $(1 \leq i \leq \ell)$, $\int \left|v^i \mathcal{K}'(v)\, dv\right| < \infty$, and*

$$\int s^{i-1}\mathcal{K}'(s)\, ds = 0, \ \textit{and} \int s^{\ell}\mathcal{K}'(s)\, ds \neq 0,$$

and $\mathcal{K}(x) - \mathcal{K}(0) \gtrless 0$ if $x \gtrless 0$.

(c) *For each integer i $(0 \leq i \leq \ell)$, and $\eta > 0$, and any sequence $\{h_n\}$ converging to 0,*

$$\lim_{n\to\infty} h_n^{i-\ell} \int_{|h_n s| > \eta} \left|s^i \mathcal{K}'(s)\right| ds = 0, \ \textit{and} \ \lim_{n\to\infty} h_n^{-1} \int_{|h_n s| > \eta} \left|\mathcal{K}''(s)\right| ds = 0.$$

(d) *For some $\mu \in (0, 1]$, a positive constant C, and all $x, y \in \mathbb{R}$,*

$$\left|\mathcal{K}''(x) - \mathcal{K}''(y)\right| \leq C\left|x - y\right|^{\mu}.$$

The two estimators have different asymptotic distributions. Let

$$\mathbb{T}_n^*(\theta) = \frac{\partial}{\partial\theta}\mathbb{S}_n^*(\theta), \text{ and } \mathbb{Q}_n^*(\theta) = \frac{\partial^2}{\partial\theta\partial\theta'}\mathbb{S}_n^*(\theta).$$

Similarly, define $\mathbb{T}_n^+(\theta)$ and $\mathbb{Q}_n^+(\theta)$. Whereas $\mathbb{Q}_n^*(\theta)$ and $\mathbb{Q}_n^+(\theta)$ converge to the same limit in a certain shrinking neighborhood of θ_0, the properly scaled $\mathbb{T}_n^*$ and $\mathbb{T}_n^+$ have different weak limits. Furthermore, the admissible ranges of the bandwidth h_n are not the same. It turns out that $\mathbb{T}_n^+$ demands $h_n = O\left(n^{-1/3}\right)$ to achieve the asymptotic normality. This implies that the convergence rate for $\hat{\theta}^+$ is at best $n^{-1/3}$ and that the uniform convergence of $\mathbb{Q}_n^+(\theta)$ is not guaranteed. While here we only present the asymptotic normality for $\hat{\theta}^*$, Lemma 5 and 7 in Appendix show the consistency and rate result for $\hat{\theta}^+$ as well as $\hat{\theta}^*$.

Let us introduce a symmetric matrix

$$Q = \begin{pmatrix} M & \cdot \\ \left(1, \frac{1}{2}\right) \otimes (1, \gamma_0)\,\omega p_0 & \omega\phi p_0 \end{pmatrix},$$

and a constant

$$V = \sigma^2\omega^2 p_0 \int |\mathcal{K}'(s)|^2 ds.$$

Then,

Theorem 4. *Suppose that Assumption 1, 2, and 3 hold with some $\ell \geq 2$ and that $p(y)$ and $f(y)$ are ℓ times continuously differentiable at γ_0 with bounded derivatives. Then, if $h_n = o\left(n^{-1/(2\ell+1)}\right)$ and $h_n^{-1} = o\left(n^{1/3}\right)$,*

$$\sqrt{nh_n}\left(\hat{\theta}^* - \theta_0\right) \xrightarrow{d} Q^{-1}\iota\mathcal{N}(0, V),$$

where $\iota = (0, .., 0, 1)'$.

Proof. See Appendix. □

Remarks

1. The convergence rate of the estimator is $\sqrt{nh_n}$, which is faster than $n^{1/3}$ under the given conditions on the bandwidth h_n, and it can get arbitrarily close to $n^{1/2}$ provided the underlying model satisfies proper smoothness condition. This is similar to Horowitz (1992), which proposed the smoothed maximum score estimator. However, the limit distribution of $\hat{\theta}^*$ is a degenerate normal, which is caused by the slower convergence of $\hat{\gamma}$ than $n^{1/2}$.

2. The asymptotic variance of $\hat{\theta}^*$ contains unknowns Q and V, which involve the unknown density of the true regression error. However, Q can be consistently estimated by $\mathbb{Q}_n^*\left(\hat{\theta}^*\right)$, the explicit formula of which is given in Appendix, and V by

$$\hat{V} = \frac{h_n}{n}\sum_{t=1}^{n}\left|\left\{\left(x_t'\hat{\delta}^*\right)^2 - 2x_t'\hat{\delta}^*\left(y_t - x_t'\hat{\beta}_1^*\right)\right\}\mathcal{K}'\left(\frac{y_{t-1} - \hat{\gamma}^*}{h_n}\right)\frac{1}{h_n}\right|^2.$$

3. The admissible range of rates for h_n and the range given in Seo and Linton (2007) are mutually exclusive. However, carefull reading of the proof reveals that they imposed the restriction that $h_n = o\left(n^{-1/3}\right)$ to ensure the asymptotic independence between $\hat{\beta}$ and $\hat{\gamma}$. This can be relaxed to $h_n = o\left(n^{-1/4}\right)$ at the expense of the asymptotic independence. This is important because it enables us to make a robust inference. That is, the asymptotic inference based on the t-statistic is valid whether or not the model is correctly specified, provided that $h_n = Kn^{-\eta}$ for some $1/4 < \eta < 1/3$ and $0 < K < \infty$.

5. Conclusion

We examined the asymptotic property of the least squares estimator of the SETAR model under a weaker condition than the standard conditional moment condition. In particular, the convergence rate of the estimator is much slower than the super-consistent rate obtained in Chan (1993). Our result may represent a less favorable case, where it is not easy to identify the threshold value. Some of the examples indeed illustrate such a case. It may also invalidate some convenient inferential procedures, which makes use of the super-consistent rate, such as the Oracle property and the sequential estimation and testing procedure for multiple regime threshold models suggested by Hansen (1999).

We now have at least three different asymptotic distributions, that is, from Chan (1993), Hansen (2000), and this paper. It is rather subjective which one to employ, whereas it would be a meaningful future research to find a more objective criterion. A prior information may well be useful. Our asymptotic distribution would be useful in case where the threshold value cannot be easily distinguishable and it affects estimation of the regression coefficients. It is natural to explore a robust inference procedure so that we do not have to specifiy each case beforehand. The smoothed estimation in Section 4 can be an option.

References

Andrews, D. W. K. (1987): "Consistency in nonlinear econometric models: a generic uniform law of large numbers," *Econometrica*, 55(6), 1465–1471.

Berk, R. H. (1966): "Limiting behavior of posterior distributions when the model is incorrect," *Annals of Mathematical Statistics*, 37, 745–746.

——— (1970): "Consistency a posteriori," *Annals of Mathematical Statistics*, 41, 894–906.

Chan, K. S. (1993): "Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model," *The Annals of Statistics*, 21, 520–533.

Chan, K. S., and R. S. Tsay (1998): "Limiting properties of the least squares estimator of a continuous threshold autoregressive model," *Biometrika*, 85(2), 413–426.

Davidson, J. (1994): *Stochastic limit theory.* Oxford University Press.

Ghaddar, D. K., and H. Tong (1981): "Data transformation and self-exciting threshold autoregression," *Journal of the Royal Statistical Society. Series C*, 30, 238–248.

Gonzalo, J., and M. Wolf (2005): "Subsampling Inference in Threshold Autoregressive Models," *Journal of Econometrics*, 127(201-224), 209–233.

Hansen, B. (1999a): "Testing for linearity," *Journal of Economic Surveys*, 13, 551–576.

Hansen, B. E. (1999b): "Threshold effects in non-dynamic panels: estimation, testing, and inference," *Journal of Econometrics*, 93(2), 345–368.

——— (2000): "Sample splitting and threshold estimation," *Econometrica*, 68, 575–603.

Horowitz, J. L. (1992): "A smoothed maximum score estimator for the binary response model," *Econometrica*, 60(3), 505–531.

Huber, P. J. (1967): "The behavior of maximum likelihood estimates under nonstandard conditions," in *Proc. Fifth Berkeley Sympos. Math. Statist. and Probability (Berkeley, Calif., 1965/66), Vol. I: Statistics*, pp. 221–233. Univ. California Press, Berkeley, Calif.

Kim, J., and D. Pollard (1990): "Cube root asymptotics," *The Annals of Statistics*, 18(1), 191–219.

Meyn, S. P., and R. L. Tweedie (1993): *Markov chains and stochastic stability*, Communications and Control Engineering Series. Springer-Verlag London Ltd., London.

Peligrad, M. (1982): "Invariance Principles for Mixing Sequences of Random Variables," *The Annals of Probability*, 10(4), 968–981.

Politis, D. N., J. P. Romano, and M. Wolf (1999): *Subsampling*, Springer Series in Statistics. Springer-Verlag, New York.

Potter, S. M. (1995): "A Nonlinear Approach to US GNP," *Journal of Applied Econometrics*, 10(2), 109–25.

Seo, M., and O. Linton (2007): "A Smoothed Least Squares Estimator For The Threshold Regression," *Journal of Econometrics*, 141, 704–735.

Tong, H. (1983): *Threshold models in nonlinear time series ananlysis*, Lecture Notes in Statistics. Springer, Berlin.

TONG, H. (1990): *Nonlinear time series*, vol. 6 of *Oxford Statistical Science Series*. The Clarendon Press Oxford University Press, New York.
TONG, H., AND K. S. LIM (1980): "Threshold Autoregression, Limit Cycles and Cyclical Data," *Journal of the Royal Statistical Society. Series B (Methodological)*, 42(3), 245–292.
VAN DER VAART, A. W., AND J. A. WELLNER (1996): *Weak Convergence and Empirical Process*. Springer, New York.
WHITE, H. (1982): "Maximum likelihood estimation of misspecified models," *Econometrica*, 50(1), 1–25.
WU, W. B. (2007): "*M*-estimation of linear models with dependent errors," *The Annals of Statistics*, 35(2), 495–521.
——— (2008): "empirical processes of stationary sequences," *Statistica Sinica*, 18, 313–333.

Appendix A. Proof of Theorems

Proof of Lemma 1. The condition (a) in Assumption 1 is sufficient to apply a ULLN to $\mathbb{S}_n$ on any compact subset K, e.g., Davidson (1994). The condition (b) in Assumption 1 and the continuity of $S(\theta)$, due to Assumption 1 (a), are sufficient for the consistency, see e.g. van der Vaart and Wellner (1996) (Corollary 3.2.3). □

Proof of Lemma 2. The conditions for Theorem 3.2.5 in van der Vaart and Wellner (1996) are verified for $-\mathbb{S}_n$ and $-S$ as the theorem is for the maximization problem. The first condition on $-S$ is trivially satisfied from the differentiability condition in Assumption 2 (c). Then, it remains to show that

$$\mathrm{E}\sup_{|\theta-\theta_0|<\zeta}\left|(\mathbb{S}_n-S)(\theta)-(\mathbb{S}_n-S)(\theta_0)\right|\lesssim\frac{\sqrt{\zeta}}{\sqrt{n}}, \tag{4}$$

for some $\zeta>0$. Let $\mathbb{P}_n=n^{-1}\sum_{t=1}^{n}$, the empirical measure, and $\mathbb{G}_n=\sqrt{n}\,(\mathbb{P}_n-\mathrm{E})$, the empirical process. And unless specified otherwise, $\pi_\gamma=\pi_\gamma(y_{t-1})$. Also define

$$\begin{aligned}
A_n &= \left[(\beta+\beta_0)'\,\mathbb{G}_n\left(\pi_{\gamma_0}\pi_{\gamma_0}'\right)-2\mathbb{G}_n y_t\pi_{\gamma_0}'\right](\beta-\beta_0)\\
B_n &= \mathbb{G}_n\left(\beta'\left(\pi_\gamma-\pi_{\gamma_0}\right)\right)^2-2\mathbb{G}_n\left[f-\beta'\pi_{\gamma_0}\right]\left(\pi_\gamma-\pi_{\gamma_0}\right)'\beta\\
C_n &= -2\mathbb{G}_n\varepsilon_t\left(\pi_\gamma-\pi_{\gamma_0}\right)'\beta,
\end{aligned}$$

and verify the inequality in (4) for each term, as $\sqrt{n}\left[(\mathbb{S}_n-S)(\theta)-(\mathbb{S}_n-S)(\theta_0)\right]$ is the sum of the three.

For A_n, it is sufficient to show that variances of $\mathbb{G}_n\left(\pi_{\gamma_0}\pi_{\gamma_0}'\right)$ and $\mathbb{G}_n y_t\pi_{\gamma_0}'$ are bounded, which is trivial and omitted. To check the condition for B_n, assume $\gamma>\gamma_0$. The other case can be verified in the same manner and omitted. Then,

$$\pi_\gamma(y)-\pi_{\gamma_0}(y)=(1,y,-1,-y)\,1\{\gamma_0<y\le\gamma\}.$$

As f and $\beta'\pi_\gamma$ are bounded on a compact set, we consider an empirical process indexed by a class of functions,

$$\{h_\gamma(y)=g(y)\,1\{\gamma_0<y\le\gamma\}:|\gamma_0-\gamma|<\zeta\},$$

for a bounded function g. We follow the proof of theorem 2 of Wu (2008) to get the modulus of continuity of such an empirical process at γ_0. While the theorem itself is for empirical distribution functions, the multiplication by a bounded function g does not change the result. In particular, the theorem implies that for a generic constant C

$$\mathrm{E}\left[\sup_{|\gamma_0-\gamma|<\zeta}|\mathbb{G}_n h_\gamma|\right]\le C\zeta^{q/2-1}\left(F(\gamma_0+\zeta)-F(\gamma_0)\right),$$

where in our case $q = 1$ and $F(\gamma) = \mathrm{E}(g(y)1\{\gamma_0 < y \leq \gamma\})$. As

$$F(\gamma_0 + \zeta) - F(\gamma_0) \leq C\zeta,$$

it is shown that B_n satisfies (4).

Finally, we apply Lemma A.3 of Hansen (2000) for C_n noting that ε_t is a martingale difference sequence. In particular, apply the lemma with $\eta = \sqrt{n}2^{2j-2}r_n^{-2}$ and $\zeta = 2^j r_n^{-1}$, where $r_n = n^{1/3}$. Strictly speaking, the proof of Theorem 3.2.5 shows that the condition (4) is a sufficient condition to ensure the inequality given in Lemma A.3. This completes the proof. □

Proof of Theorem 3. As $S(\theta)$ is differentiable and minimized at θ_0, the first order condition of the minimization implies that

$$2f(\gamma_0) = (1, \gamma_0, 1, \gamma_0)\beta_0, \tag{5}$$

and

$$\mathrm{E}\left[\pi_{\gamma_0}(y_{t-1})e_t\right] = 0, \tag{6}$$

where $e_t = \left(y_t - \beta_0'\pi_{\gamma_0}(y_{t-1})\right)$.

Due to Lemma 2, we may apply reparametrization $b = r_n(\beta - \beta_0)$, $g = r_n(\gamma_n - \gamma_0)$, where $r_n = n^{1/3}$ and $|g|, |b| \leq K < \infty$, throughout the proof. Partition of $b = (b_1', b_2')'$ according to that of $\beta = \left(\beta_1', \beta_2'\right)'$. Then,

$$r_n^2(\mathbb{S}_n(\theta) - \mathbb{S}_n(\theta_0)) = A_n + B_n + C_n,$$

where

$$\begin{aligned}
A_n &= b'\mathbb{P}_n\pi_{\gamma_0}\pi_{\gamma_0}'b + 2r_n\mathbb{P}_n b'\pi_{\gamma_0}\left(\pi_\gamma - \pi_{\gamma_0}\right)'\beta, \\
B_n &= b'\mathbb{P}_n\left(\pi_\gamma - \pi_{\gamma_0}\right)\left(\pi_\gamma - \pi_{\gamma_0}\right)'b - 2\frac{r_n}{\sqrt{n}}\mathbb{G}_n e_t\pi_\gamma'(y_{t-1})b, \\
C_n &= \frac{r_n^2}{n}\sum_{t=1}^n\left(\beta_0'\left(\pi_\gamma - \pi_{\gamma_0}\right) - 2e_t\right)\left(\pi_\gamma - \pi_{\gamma_0}\right)'\beta_0.
\end{aligned}$$

We apply a generic uniform law of large numbers to A_n, see e.g., Andrews (1987, Corollary 1). Note that $\pi_{\gamma_0}\left(\pi_\gamma - \pi_{\gamma_0}\right)'$ equals $(1, y, 0, 0)'\left(\pi_\gamma - \pi_{\gamma_0}\right)'$ if $\gamma < \gamma_0$ and $(0, 0, 1, y)\left(\pi_\gamma - \pi_{\gamma_0}\right)'$ otherwise. And, for $k = 0$ or 1 and $g \geq 0$,

$$r_n\mathrm{E}\left\{y_{t-1}^k 1\left(\gamma_0 < y_{t-1} \leq \gamma_0 + gr_n^{-1}\right)\right\} \to \gamma_0^k\, p(\gamma_0)\, g,$$

using the change-of-variables. Then, a simple algebra yields that

$$A_n \xrightarrow{p} b'Mb + 2p_0\delta_0'\begin{pmatrix} 1 & \gamma_0 \\ \gamma_0 & \gamma_0^2 \end{pmatrix}(b_1 1\{g \leq 0\} + b_2 1\{g > 0\})g,$$

uniformly in b and g. Next, the first term in B_n is analysed in the same manner. And, it follows from (6) and Lemma 3.4 of Peligrad (1982) that there is a $K < \infty$ such that

$$\operatorname{var}\mathbb{G}_n e_t\pi_{\gamma_0}(y_{t-1}) \leq K\ \operatorname{var}\left(e_t\pi_{\gamma_0}(y_{t-1})\right) < \infty,$$

where the last inequality is due to Assumption 2 (b). Then,

$$\sup_b|B_n| = O_p\left(r_n^{-2} + \frac{r_n}{\sqrt{n}}\right) = o_p(1).$$

Now consider C_n. Recall that $e_t = \varepsilon_t + f(y_{t-1}) - \beta_0'\pi_{\gamma_0}(y_{t-1})$ and note that if y_{t-1} lies between γ and γ_0,

$$1\{y_{t-1} \leq \gamma\} + 1\{y_{t-1} \leq \gamma_0\} = 1\{y_{t-1} > \gamma\} + 1\{y_{t-1} > \gamma_0\} = 1,$$

to yield

$$C_n = -\frac{2r_n}{\sqrt{n}}\mathbb{G}_n\varepsilon_t\left(\pi_{\gamma_0+g/r_n}-\pi_{\gamma_0}\right)'\beta_0 + r_n^2\mathbb{P}_n\xi\left(y_{t-1}\right)\left(\pi_{\gamma_0+g/r_n}-\pi_{\gamma_0}\right)'\beta_0,$$

where $\xi\left(y\right) = \left(1,y,1,y\right)\beta_0 - 2f\left(y\right)$. We show below that the empirical process in C_n, the first term, converges weakly to a zero-mean Gaussian process and the second converges in probability to a deterministic function uniformly in g. However, the tightness of the empirical process follows from Lemma A.3 of Hansen (2000) (as proceeding in Lemma A.11 therein). To characterize the covariance kernel, note that

$$\left(\pi_{\gamma_0+g_1/r_n}-\pi_{\gamma_0}\right)\left(\pi_{\gamma_0+g_2/r_n}-\pi_{\gamma_0}\right)' = 0,$$

if g_1 and g_2 take opposite signs. Thus let g_1 and $g_{2.}$ be positive. Then, as $r_n^4/n = r_n$, the covariance kernel of the limit process is given by

$$\lim_{n\to\infty} r_n 4\mathrm{E}\varepsilon_t^2\int_{\gamma_0}^{\gamma_0+(g_1\wedge g_2)/r_n}\left(\left(1,y\right)\left(\beta_{10}-\beta_{20}\right)\right)^2 p\left(y\right)dy = 4\sigma^2\omega^2 p_0\left(g_1\wedge g_2\right),$$

using change-of-variables. Proceed similary for negative values of g_1 and g_2. Next, note that $\xi\left(\gamma_0\right) = 0$ due to (5) and ξ is differentiable. Using the mean value expansion $\xi\left(y_{t-1}\right) = \xi'\left(\vec{y}_{t-1}\right)\left(y_{t-1}-\gamma_0\right)$ and change-of-variables, it is easy to see that

$$r_n^2\mathrm{E}\xi\left(y_{t-1}\right)\left(\pi_{\gamma_0+g/r_n}-\pi_{\gamma_0}\right)'\beta_0 \to \frac{1}{2}\xi'\left(\gamma_0\right)\left(\left(1,\gamma_0\right)\left(\beta_{10}-\beta_{20}\right)\right)p_0 g^2,$$

where $\xi'\left(\gamma_0\right) = \left(0,1,0,1\right)\beta_0 - 2f'\left(\gamma_0\right) = \alpha_{10}+\alpha_{20}-2f'\left(\gamma_0\right)$. Then, a ULLN yields that

$$\frac{r_n^2}{n}\sum_{t=1}^{n}\xi\left(y_{t-1}\right)\left(\pi_{\gamma_0+g/r_n}-\pi_{\gamma_0}\right)'\beta_0 \xrightarrow{p} \frac{1}{2}\xi'\left(\gamma_0\right)\left(\left(1,\gamma_0\right)\left(\beta_{10}-\beta_{20}\right)\right)p_0 g^2,$$

uniformly in g. This concludes the proof. □

The proof of Theorem 4 takes several steps.

Lemma 5. *$\hat{\theta}^+$ and $\hat{\theta}^*$ are consistent.*

Proof. Proving Theorem 1 of Seo and Linton (2007), they showed that $\mathbb{S}_n - \mathbb{S}_n^+ = o_p\left(1\right)$ uniformly over Θ. This and Lemma 1 yield the consistency of $\hat{\theta}^+$ and similar argument applies for that of $\hat{\theta}^*$. □

Lemma 6.

$$\mathrm{E}\mathbb{T}_n^+\left(\theta_0\right) = O\left(h_n\right) \text{ and } nh_n \operatorname{var}\mathbb{T}_n^+\left(\theta_0\right) \to \mathbb{V}^+ \tag{7}$$

$$\mathrm{E}\mathbb{T}_n^*\left(\theta_0\right) = O\left(h_n^\ell\right) \text{ and } nh_n \operatorname{var}\mathbb{T}_n^*\left(\theta_0\right) \to \mathbb{V}^*, \tag{8}$$

where the last diagonal elements of $\mathbb{V}^+$ *and* $\mathbb{V}^*$ *are given by, respectively,* $4\sigma^2\omega^2 p_0\int\left|\mathcal{K}'\right|^2 + \omega^4 p_0\int\left|\mathcal{K}'\right|^2 + 4\omega^2 p_0\int\left(1\left(y>0\right)-\mathcal{K}\left(y\right)\right)^2\left|\mathcal{K}'\left(y\right)\right|^2 dy$ *and* $4\sigma^2\omega^2 p_0\int\left|\mathcal{K}'\right|^2$, *and the other elements in them are zero.*

Proof. Let $\kappa_t = \kappa_t\left(\gamma_0\right)$, where $\kappa_t\left(\gamma\right) = \partial\mathcal{K}\left(\frac{y_{t-1}-\gamma}{h_n}\right)/\partial\gamma = \frac{-1}{h_n}\mathcal{K}'\left(\frac{y_{t-1}-\gamma}{h_n}\right)$. Then,

$$\mathbb{T}_n^+\left(\theta\right) = \frac{2}{n}\sum_{t=1}^{n}\begin{pmatrix} -\left(y_t-\chi_\gamma\left(y_{t-1}\right)'\beta\right)\chi_\gamma\left(y_{t-1}\right) \\ \left(y_t-\chi_\gamma\left(y_{t-1}\right)'\beta\right)x_t'\delta\kappa_t\left(\gamma\right)\end{pmatrix}. \tag{9}$$

We first show (7) for the second element in (9). The algebra is similar for the first element and thus omitted. Replace $\left(y_t - \chi_\gamma\left(y_{t-1}\right)'\beta\right)$ with

$$\varepsilon_t + \left(f\left(y_{t-1}\right) - \pi_\gamma\left(y_{t-1}\right)'\beta\right) + \left(\pi_\gamma - \chi_\gamma\right)\left(y_{t-1}\right)'\beta,$$

and first consider

$$A_n = \frac{1}{n}\sum_{t=1}^{n} \varepsilon_t x_t' \delta_0 \kappa_t.$$

As ε_t is iid sequence, $\mathrm{E}A_n = 0$ and applying the change of variables

$$\begin{aligned} nh_n \mathrm{E}A_n^2 &= \sigma^2 \int \left(\left(1, y\right)\delta_0\right)^2 \mathcal{K}'\left(\frac{y-\gamma_0}{h_n}\right)^2 p\left(y\right)\frac{dy}{h_n} \\ &= \sigma^2 \int \left(\left(1, h_n y + \gamma_0\right)\delta_0\right)^2 \mathcal{K}'\left(y\right)^2 p\left(h_n y + \gamma_0\right) dy \\ &\to \sigma^2\omega^2 p_0 \int \left|\mathcal{K}'\right|^2. \end{aligned}$$

Next, consider

$$B_n = \frac{1}{n}\sum_{t=1}^{n}\left(f\left(y_{t-1}\right) - \pi_{\gamma_0}\left(y_{t-1}\right)'\beta_0\right) x_t'\delta_0\kappa_t.$$

Then, applying the change of variables

$$\mathrm{E}B_n = \int \left(f\left(h_n y + \gamma_0\right) - \pi_{\gamma_0}\left(h_n y + \gamma_0\right)'\beta_0\right)\left(1, h_n y + \gamma_0\right)\delta_0 \mathcal{K}'\left(y\right) p\left(h_n y + \gamma_0\right) dy.$$

Since $\mathcal{K}'$ is symmetric, $\int \mathcal{K}' = 1$, and $\int |y|\, \mathcal{K}'\left(y\right) dy = 0$, we have

$$\begin{aligned} &\int \pi_{\gamma_0}\left(h_n y + \gamma_0\right)'\beta_0\left(1, \gamma_0\right)\delta_0\mathcal{K}'\left(y\right) p\left(h_n y + \gamma_0\right) dy \\ =\ & \omega\left(1, \gamma_0\right)\left(\beta_{10}\int_{-\infty}^{0}\mathcal{K}'\left(y\right) p\left(h_n y + \gamma_0\right) dy + \beta_{20}\int_{0}^{\infty}\mathcal{K}'\left(y\right) p\left(h_n y + \gamma_0\right) dy\right) \\ =\ & \omega\left(1, \gamma_0\right)\left(\frac{1}{2}\left(\beta_{10} + \beta_{20}\right) p_0 + O\left(h_n^2\right)\right), \end{aligned} \tag{10}$$

where the last equality can be derived by expanding $p\left(\cdot\right)$ at γ_0, and similarly,

$$\int \pi_{\gamma_0}\left(h_n y + \gamma_0\right)'\beta_0\left(0, h_n y\right)\delta_0\mathcal{K}'\left(y\right) p\left(h_n y + \gamma_0\right) dy = O\left(h_n^2\right). \tag{11}$$

In the same manner, it can be deduced that

$$\int f\left(h_n y + \gamma_0\right)\left(1, h_n y + \gamma_0\right)\mathcal{K}'\left(y\right) p\left(h_n y + \gamma_0\right) dy = \omega f\left(\gamma_0\right) p_0 + O\left(h_n^2\right). \tag{12}$$

However, the first order condition (5) implies that

$$\mathrm{E}B_n = (12) - (10) - (11) = O\left(h_n^2\right).$$

For $nh_n \operatorname{var}\left(B_n\right)$, we refer to Lemma 2 of Seo and Linton (2007), which shows the negligibility of the sum of autocovariance terms across t, that is, writing $B_n = \frac{1}{n}\sum_{t=1}^{n} B_{nt}$,

$$nh_n \operatorname{var}\left(B_n\right) = h_n\left(\mathrm{E}B_{nt}^2 - \left(\mathrm{E}B_{nt}\right)^2\right) = h_n\mathrm{E}B_{nt}^2 - O\left(h_n^5\right),$$

and the same algebra as above yields

$$h_n\mathrm{E}B_{nt}^2 \to \frac{1}{4}\omega^4 p_0 \int \left|\mathcal{K}'\right|^2.$$

Then, consider

$$\begin{aligned} C_n &= \frac{1}{n}\sum_{t=1}^{n} \beta_0' \left(\pi_{\gamma_0} - \chi_{\gamma_0}\right)(y_{t-1})\, x_t'\delta_0\kappa_t \\ &= \frac{1}{n}\sum_{t=1}^{n} \left(1\left(y_{t-1} > \gamma_0\right) - \mathcal{K}_{\gamma_0}\left(y_{t-1}\right)\right)\kappa_t\ \left(x_t'\delta_0\right)^2, \end{aligned}$$

since $\left(\pi_\gamma - \chi_\gamma\right)(y) = \left(1\left(y > \gamma\right) - \mathcal{K}\left(\left(y-\gamma\right)/h_n\right)\right)\left((-1,1)\otimes(1,y)\right)'$. We may note the fact that $\int \left(1\left(y>0\right) - \mathcal{K}\left(y\right)\right)\mathcal{K}'\left(y\right)dy = 0$ to deduce

$$\begin{aligned} \mathrm{E}C_n &= \int \left(1\left(y>0\right) - \mathcal{K}\left(y\right)\right)\mathcal{K}'\left(y\right)\left(\left(1, h_n y + \gamma_0\right)\delta_0\right)^2 p\left(h_n y + \gamma_0\right)dy. \\ &= h_n\omega\left[\omega p'\left(\gamma_0\right) + 2\delta_{20}p_0\right]\int \left(1\left(y>0\right) - \mathcal{K}\left(y\right)\right)\mathcal{K}'\left(y\right)y dy + o\left(h_n\right). \end{aligned}$$

It can be shown that $\int \left(1\left(y>0\right) - \mathcal{K}\left(y\right)\right)\mathcal{K}'\left(y\right)ydy > 0$ by integral by parts and by noting that $\left(1\left(y>0\right) - \mathcal{K}\left(y\right)\right)y$ is symmetric. Proceeding similarly for $\mathrm{var}\left(B_n\right)$, define C_{nt} and deduce $nh_n \,\mathrm{var}\left(C_n\right) = h_n \mathrm{E}C_{nt}^2 + o\left(1\right)$, where

$$h_n \mathrm{E}C_{nt}^2 \to \omega^2 p_0 \int \left(1\left(y>0\right) - \mathcal{K}\left(y\right)\right)^2 \left|\mathcal{K}'\left(y\right)\right|^2 dy.$$

Finally, $nh_n \,\mathrm{cov}\left(A_n, B_n\right) = nh_n\,\mathrm{cov}\left(A_n, C_n\right) = 0$ and

$$nh_n\,\mathrm{cov}\left(B_n, C_n\right) = h_n \mathrm{E}\left(B_{nt}C_{nt}\right) = O\left(h_n\right).$$

Turning to

$$\mathbb{T}_n^*\left(\theta\right) = \frac{2}{n}\sum_{t=1}^{n}\begin{pmatrix} -\left(y_t - x_t'\beta_1 - x_t'\delta\mathcal{K}_\gamma\left(y_{t-1}\right)\right)x_t \\ -\left\{y_t - x_t\beta_2\right\}x_t\mathcal{K}_\gamma\left(y_{t-1}\right) \\ x_t'\delta\left(\varepsilon_t + f\left(y_{t-1}\right) - x_t'\left(\beta_1+\beta_2\right)/2\right)\kappa_t \end{pmatrix}, \tag{13}$$

we examine the last element in (13) at $\theta = \theta_0$, which is $2\left(A_n + D_n\right)$, where

$$D_n = \frac{1}{n}\sum_{t=1}^{n}\left(f\left(y_{t-1}\right) - x_t'\left(\beta_{10}+\beta_{20}\right)/2\right)x_t'\delta_0\kappa_t.$$

Applying the change of variables, expansions, (5), and Assumption 3 (a), we may deduce

$$\begin{aligned} \mathrm{E}D_n &= \int \left(\left(f'\left(\gamma_0\right) - \frac{\alpha_{10}+\alpha_{20}}{2}\right)h_n y + \frac{f''\left(\tilde{\gamma}\right)}{2}h_n^2 y^2\right)\left(1, h_n y + \gamma_0\right)\delta_0 \\ &\quad \times\mathcal{K}'\left(y\right)p\left(h_n y + \gamma_0\right)dy, \\ &= O\left(h_n^{\ell}\right). \end{aligned}$$

And, defining D_{nt} as in B_{nt} and proceeding as in the calculation of $\mathrm{E}D_n$,

$$\begin{aligned} \mathrm{E}D_{nt}^2 &= \int \left(\left(f'\left(\gamma_0\right) - \frac{\beta_{12}+\beta_{22}}{2}\right)h_n y + \frac{f''\left(\tilde{\gamma}\right)}{2}h_n^2 y^2\right)^2 \left(\left(1, h_n y + \gamma_0\right)\delta_0\right)^2 \\ &\quad \times\mathcal{K}'\left(y\right)^2 p\left(h_n y + \gamma_0\right)dy \\ &\to 0. \end{aligned}$$

□

Lemma 7. $\sqrt{nh_n}\left(\hat{\theta}^* - \theta_0\right) = O_p\left(1\right)$, *and* $\sqrt{nh_n}\left(\hat{\theta}^+ - \theta_0\right) = O_p\left(1\right)$.

Proof. We apply Theorem 3.4.1 in van der Vaart and Wellner (1996). Let $S_n^*(\theta) = \mathrm{E}\mathbb{S}_n^*(\theta)$. The first condition, which concerns $-S_n^*$, is obvious from its differentiability. Conditions are verified for $-\mathbb{S}_n^*$ and $-S_n^*$ as their theorem is for the maximization problem. The second condition is

$$\mathrm{E}\sup_{|\theta-\theta_0|<\zeta}\left|\left(\mathbb{S}_n^*-S_n^*\right)(\theta)-\left(\mathbb{S}_n^*-S_n^*\right)(\theta_0)\right| \lesssim \frac{\zeta}{\sqrt{nh_n}}. \tag{14}$$

Some algebra yields that

$$\begin{aligned}
&\sqrt{nh_n}\left[\left(\mathbb{S}_n^*-S_n^*\right)(\theta)-\left(\mathbb{S}_n^*-S_n^*\right)(\theta_0)\right]\\
= & -\sqrt{h_n}\mathbb{G}_n\left(2y_t - x_t'(\beta_1+\beta_{10}) - 2x_t'\delta\mathcal{K}_{\gamma_0}(y_{t-1})\right)x_t'(\beta_1-\beta_{10})\\
& -\sqrt{h_n}\mathbb{G}_n\left(2y_t - 2x_t'\beta_{10} - x_t'(\delta+\delta_0)\right)\mathcal{K}_{\gamma_0}(y_{t-1})\,x_t'(\delta-\delta_0)\\
& -\sqrt{h_n}\mathbb{G}_n\left\{2y_t - x_t'(\beta_1+\beta_2)\right\}\beta'\left(\chi_\gamma-\chi_{\gamma_0}\right)(y_{t-1}).
\end{aligned}$$

Since $\chi_\gamma(y)-\chi_{\gamma_0}(y) = -(1,y,-1,-y)\,\mathcal{K}'\left(\frac{y-\tilde{\gamma}}{h_n}\right)\frac{1}{h_n}(\gamma-\gamma_0)$, it is sufficient to show that the terms preceding $(\beta_1-\beta_{10})$, $(\delta-\delta_0)$, and $(\gamma-\gamma_0)$ have bounded variances. However, this can be shown by direct calculation as in Lemma 6.

For $\hat{\theta}^+$, we may write

$$\begin{aligned}
&\sqrt{nh_n}\left[\left(\mathbb{S}_n^+-S_n^+\right)(\theta)-\left(\mathbb{S}_n^+-S_n^+\right)(\theta_0)\right]\\
= & \left[(\beta+\beta_0)'\sqrt{h_n}\mathbb{G}_n\chi_{\gamma_0}\chi_{\gamma_0}' - 2\sqrt{h_n}\mathbb{G}_n\left(y_t\chi_{\gamma_0}(y_{t-1})'\right)\right](\beta-\beta_0)\\
& -\sqrt{h_n}\mathbb{G}_n\left[2y_t-\beta'\left(\chi_\gamma+\chi_{\gamma_0}\right)\right]\left(\chi_\gamma-\chi_{\gamma_0}\right)'\beta,
\end{aligned}$$

and proceed similalry as above. □

Proof of Theorem 4. In view of Lemma 3 of Seo and Linton (2007), Lemma 6 is sufficient to claim that $\sqrt{nh_n}\mathbb{T}_n^*(\theta_0)$ is asymptotically normal and it remains to derive the convergence of the second derivative matrix $\mathbb{Q}_n^*$, which is given explicitly by

$$\frac{2}{n}\sum_{t=1}^n\left[\begin{array}{cc}\left[\begin{array}{cc}1 & \mathcal{K}_\gamma(y_{t-1})\\ \cdot & \mathcal{K}_\gamma(y_{t-1})\end{array}\right]\otimes x_tx_t' & \begin{array}{c}x_tx_t'\delta\kappa_t(\gamma)\\ (x_tx_t'\beta_2-y_tx_t)\,\kappa_t(\gamma)\end{array}\\ \cdot & x_t'\delta\left(y_t - x_t'\frac{\beta_1+\beta_2}{2}\right)\mathcal{K}''\left(\frac{y_{t-1}-\gamma}{h_n}\right)\frac{1}{h_n^2}\end{array}\right]. \tag{15}$$

Given Lemma 7, it is sufficient to show that

$$\mathbb{Q}_n^*(\theta)\xrightarrow{p}2Q,$$

uniformly in θ such that $|\theta-\theta_0|\le C(nh_n)^{-1/2}$ for any $C<\infty$. However, Lemma 5 and 6 of Seo and Linton (2007) derived the uniform convergence of $\mathbb{Q}_n^*$ in a $o(h_n)$ neighborhood of θ_0 and $(nh_n)^{-1/2}$ is smaller order than h_n under the condition of the theorem. The pointwise convergence of each term follows from direct calculation similarly as above. We only illustrate the convergence of the last diagonal element,

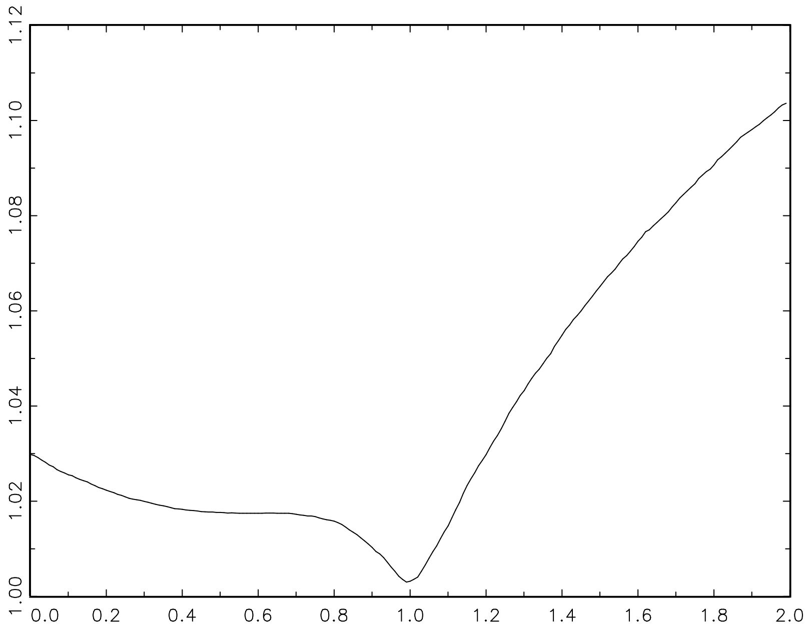

FIGURE 1. Plot of $S(\gamma)$

for which we apply the change of variables and reall (5) to deduce

$$\begin{aligned}
& \mathrm{E}\left[x_t'\delta\left(\varepsilon_t + f\left(x_t\right) - x_t'\left(\beta_1+\beta_2\right)/2\right)\mathcal{K}''\left(\frac{y_{t-1}-\gamma}{h_n}\right)\frac{-1}{h_n^2}\right] \\
= & -\int\left(1, h_n y+\gamma_0\right)\delta_0\left(\left(f'\left(\gamma_0\right)-\frac{\alpha_1+\alpha_2}{2}\right)y+\frac{f''\left(\tilde{\gamma}\right)}{2}h_n y^2\right)\mathcal{K}''\left(y\right)p\left(h_n y+\gamma_0\right)dy \\
\to & \ \omega\left(f'\left(\gamma_0\right)-(0,1)\frac{\left(\beta_1+\beta_2\right)}{2}\right)p_0,
\end{aligned}$$

as $\int y\mathcal{K}''(y)\,dy = -1$. This completes the proof. □

LONDON SCHOOL OF ECONOMICS, HOUGHTON STREET, WC2A 2AE, LONDON, UK
URL: http://personal.lse.ac.uk/seo
E-mail address: m.seo@lse.ac.uk

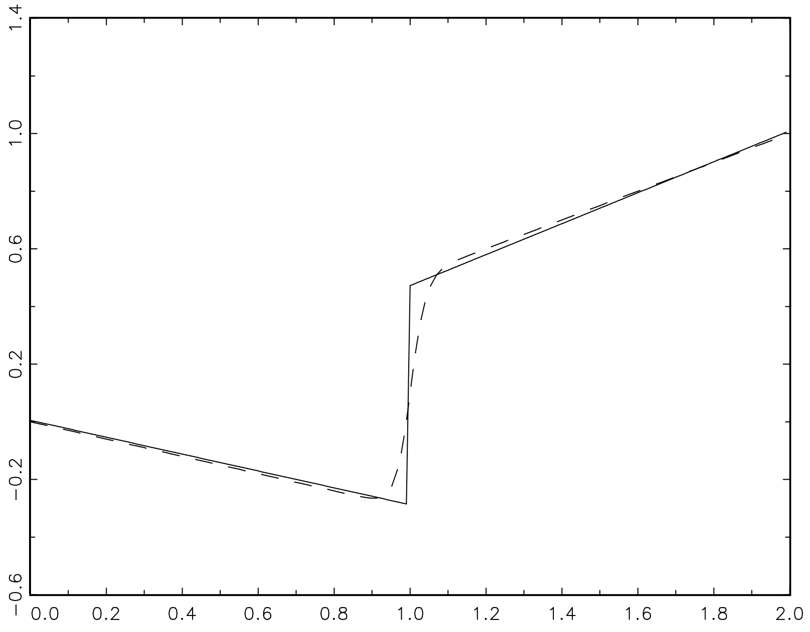

FIGURE 2. Fitted (solid line) and True (dashed line) Regression Functions.

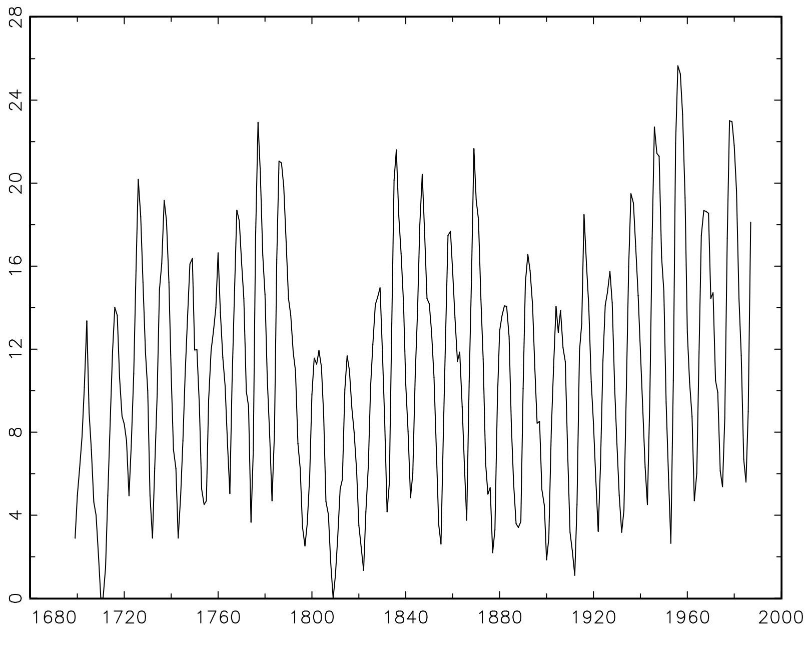

FIGURE 3. Annual Sunspot Means, 1700-1988

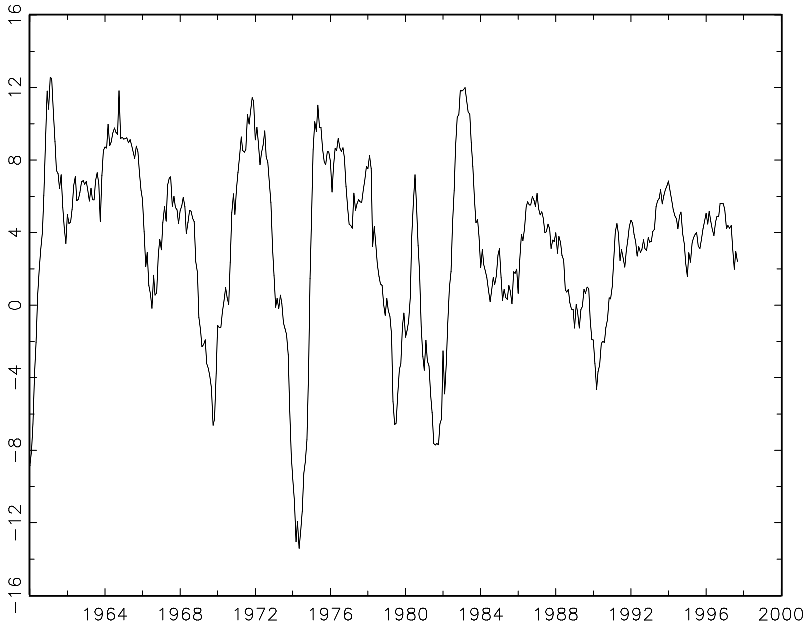

FIGURE 4. U.S. Monthly Industrial Production, Annual Growth, 1960-1998

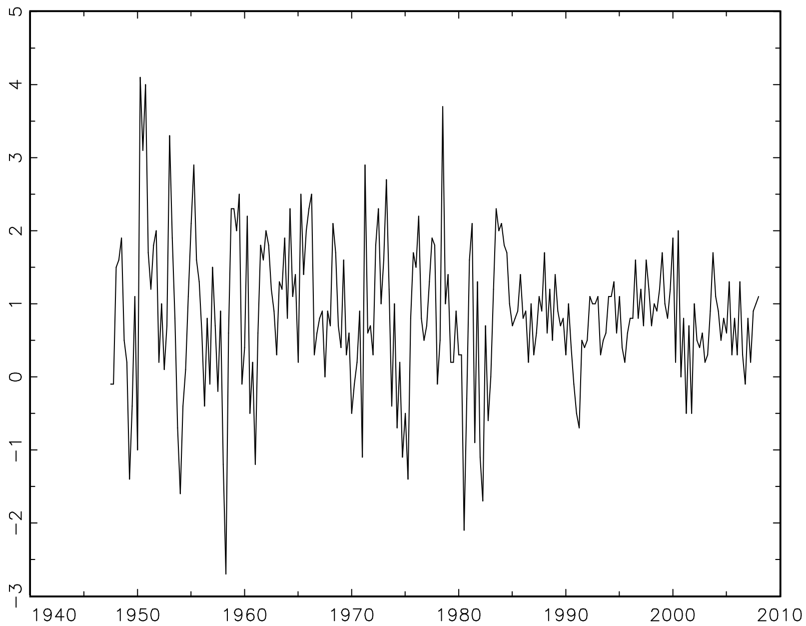

FIGURE 5. U.S. GNP Quarterly Growth, 1947-2007

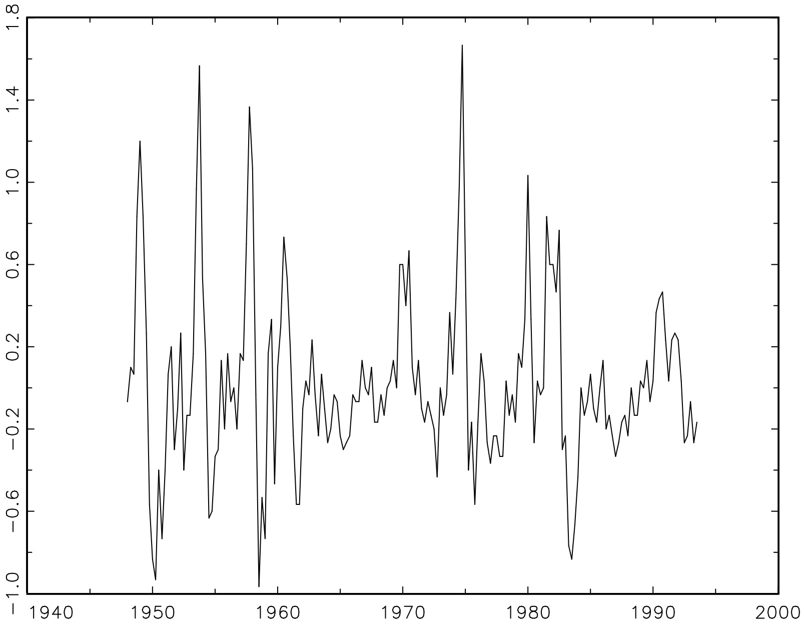

FIGURE 6. U.S. Unemployment Rate, Quaterly Growth, 1948-1993