Inference for Non-stationary Time Series Auto Regression

By Zhou Zhou [1]

University of Toronto

January 21, 2013

Abstract

The paper considers simultaneous inference for a class of non-stationary autoregressive models where the model parameters are allowed to vary smoothly over time. Simultaneous confidence tubes with asymptotically correct coverage probabilities are constructed to assess the overall patterns and magnitudes of the parameter functions over time. Simulation studies are conducted and a real time series dataset is analyzed to demonstrate the usefulness of the proposed methodology.

1 Introduction

Stochastic processes with time-varying probabilistic behaviors are common in many current scientific endeavors. For this reason there has been a great recent interest in non-stationary time series analysis. Particularly the class of locally stationary linear processes has attracted enormous attention in the statistical literature. See for instance Priestley (1988), Dahlhaus (1997), Dahlhaus (2000), Fryzlewicz and Nason (2006), Kitagawa and Gersch (1985) and Ombao et. al (2005) among others for various time and spectral domain approaches. We also refer to the recent excellent review of Dahlhaus (2011) for more discussions and references.

In this article we consider statistical inference for the following auto-regressive linear time series model with time-varying coefficients (tv-AR(p) model):

$$X_{n,i} = \beta_0(t_i) + \sum_{k=1}^{p} \beta_k(t_i) X_{n,i-k} + \varepsilon_i, \quad i = -\theta n, \ldots, n, \tag{1}$$

[1] Corresponding author. Department of Statistics, 100 St. George St. Toronto, ON M5S 3G3 Canada. *E-mail: zhou@utstat.toronto.edu*

Key words and phrases. Simultaneous confidence tubes, time-varying AR models, Gaussian approximation, locally stationary approximation.

with $X_{n,1}, X_{n,2}, \cdots, X_{n,n}$ observed and $\theta > 0$, where $t_i = i/n$ and ε_i's are uncorrelated innovations with $\mathbb{E}\varepsilon_i^2 < \infty$. In (1) we allow the recursion to start from $i = -\theta n$ to insure that the observed time series is sufficiently mixed. We shall omit the subscript n in $X_{n,i}$ in the sequel if no confusions will arise. By allowing the coefficient functions $\beta_i(t)$'s to change smoothly over time, one gains the capacity and flexibility to investigate the magnitude and pattern of the time-varying auto-regressive relationship, which is proved to be valuable in many real applications. Note that if $\beta_i(t)$'s are not varying with time, then (1) is reduced to the classic AR(p) model.

For the statistical inference of model (1), simultaneous confidence regions (SCR) for the coefficient functions are crucial. Similar to the rolls of confidence intervals in parametric models, SCRs enjoy the nice frequentist interpretation that the true coefficient functions fall in them with a prescribed probability. Clearly there are many important applications of SCRs. For instance, one can test whether certain coefficient function $\beta_i(t)$ is really time varying by checking whether a constant function can be embedded into the SCR of $\beta_i(t)$. Nevertheless, the construction of SCRs for parameter functions of locally stationary time series is a highly nontrivial task and to our knowledge there have been few results in the literature on this until this point. The purpose of the paper is to construct SCRs for $\beta_i(t)$'s in model (1) which achieve the correct coverage probability asymptotically. Consequently many formal statistical inference for model (1) can be carried out with the aid of the latter SCRs.

The construction of SCRs for model (1) relies heavily on Gaussian approximation results for dependent data. Recently, Wu and Zhou (2011) and Zhou and Wu (2010) developed Gaussian approximation results for a class of locally stationary time series. However, their results cannot be directly used here since $\{X_i\}$ in (1) does not have the simple locally stationary representation in the latter papers. In this paper, we establish a locally stationary approximation of X_i which states that X_i can be approximated by a simple locally stationary time series in the sense of Zhou and Wu (2010) with asymptotically negligible errors. As a consequence Gaussian approximation for X_i in model (1) can be established.

Most of the previous results on tv-AR models focused on independent errors. Deep locally stationary approximation and locally asymptotically normal results for such models were established in Dahlhaus (1997) and Dahlhaus (2011), among others. However, the independence assumption for the errors is restrictive in many applications. One important example is in finance where many return series are shown to be driven by conditionally heteroscedastic errors. See for instance the return series of the value-weighted index analyzed in Section 5. In this paper, we assume that the error series belong to a general class of nonlinear locally stationary time series in the sense of Zhou and Wu (2010). In particular, tv-AR models driven by time-varying ARCH or GARCH errors are allowed.

The rest of the paper is organized as follows. In Section 2 we establish locally stationary approximation for tv-AR models driven by non-stationary and nonlinear errors. SCRs with asymptotically correct coverage probabilities are constructed in Section 3. In Section 4 we shall conduct a simulation study to investigate the finite sample accuracy of the coverage probabilities of the SCRs. The return series of the value-weighted index was analyzed in Section 5. Finally the theoretical results are proved in Section 6.

2 Locally stationary approximations of tv-AR models.

As we discussed in the introduction, in many applications it is appropriate to model the error series $\{\varepsilon_i\}$ as non-stationary dependent sequence. In this paper, we shall assume that the error process is a zero mean nonlinear locally stationary sequence with the following representation (Zhou and Wu (2010))

$$\varepsilon_i = H(t_i, \mathcal{F}_i), \text{ where } \mathcal{F}_i = (\cdots, \eta_0, \eta_1, \cdots, \eta_i) \tag{2}$$

and η_i's are i.i.d. random variables. In (2), $H(t, \cdot)$ can be viewed as a time dependent nonlinear filter with innovations $\{\eta_i\}$. Since H is a function of t, the generated time series is non-stationary. Furthermore, if H is a smooth function of t in some appropriate sense, then we have local stationarity. To insure identifiability of the AR(p) model, we require that $H(t, \mathcal{F}_i)$ is a white noise system in the sense that

$$\mathrm{Cov}(H(t, \mathcal{F}_i), H(s, \mathcal{F}_j)) = 0 \text{ for all } t, s \in [-\theta, 1] \text{ and } i \neq j.$$

Model (2) allows the error sequence $\{\varepsilon_i\}$ to have very general non-stationary dependence structures. An important example of such system is the time varying GARCH(q, r) model

(Wu and Zhou (2011))

$$H(t, \mathcal{F}_i) = \eta_i V_i^{1/2}(t), \quad V_i(t) = c(t) + \sum_{j=1}^{q} \alpha_j(t) H^2(t, \mathcal{F}_{i-j}) + \sum_{k=1}^{r} \beta_k(t) V_{i-k}(t),$$

where η_i's are i.i.d. random variables with zero mean and finite variance. It is easy to see that the time-varying GARCH(q, r) model is a zero mean white noise system.

Associated with model (2), we define the following dependence measures to quantify the degree of the system's temporal dependence (Zhou and Wu (2010)):

Definition 1. *(Physical dependence measures). Assume for all $t \in [-\theta, 1]$, $\|H(t, \mathcal{F}_i)\|_q < \infty$, $q > 0$, where $\|\cdot\|_q = \{\mathbb{E}[|\cdot|^q]\}^{1/q}$ denotes the $\mathcal{L}^q$ norm of a random variable or vector. Let $(\eta_i')_{i\in\mathbb{Z}}$ be an i.i.d. copy of $(\eta_i)_{i\in\mathbb{Z}}$. For $j \geq 0$, define the physical dependence measure*

$$\delta(j, q) = \sup_{t\in[-\theta,1]} \|H(t, \mathcal{F}_i) - H(t, \mathcal{F}_{i,j})\|_q,$$

where $\mathcal{F}_{i,j} = (\mathcal{F}_{i-j-1}, \eta_{i-j}', \eta_{i-j+1}, \cdots, \eta_{i-1}, \eta_i)$.

The measures $\delta(j, q)$ quantify the changes in the filter's output when input of the system j steps ahead is changed to an i.i.d. copy. If the change is small, then we have short-range dependence and otherwise long-range dependence is demonstrated. We refer to Zhou and Wu (2009) for a more detailed discussion and calculation of such dependence measures.

Define $X_i^* = G(i/n, \mathcal{F}_i)$, where

$$G(t, \mathcal{F}_i) = \beta_0(t) + \sum_{j=1}^{p} \beta_j(t) G(t, \mathcal{F}_{i-j}) + H(t, \mathcal{F}_i) \tag{3}$$

Note that $\{X_i^*\}_{i=1}^n$ is a locally stationary time series in the sense of (2). More specifically, under the invertibility condition (C1) below, $G(t, \mathcal{F}_i)$ can be written in the following MA(∞) representation:

$$G(t, \mathcal{F}_i) = (1 - \sum_{j=1}^{p} \beta_j(t) B^j)^{-1} [\beta_0(t) + H(t, \mathcal{F}_i)],$$

where B denotes the back shift operator; i.e. $B[M(t, \mathcal{F}_i)] = M(t, \mathcal{F}_{i-1})$ for any function M. The following proposition is a key result in this paper which states that $\{X_i\}_{i=1}^n$ in (1) can be well approximated by $\{X_i^*\}_{i=1}^n$ under mild conditions. The result facilitates in-depth theoretical and methodological investigations of tv-AR models.

Proposition 1. *Assume that (C1): the roots of* $P(t,B) = 1 - \sum_{j=1}^{p} \beta_j(t) B^j$ *are uniformly bounded away from the unit circle on* $[-\theta, 1]$*; (C2): For all* t *and* i*,* $\|H(t, \mathcal{F}_i)\|_q < \infty$ *for some* $q \geq 2$ *and there exists a finite constant* C*, such that* $\|H(t, \mathcal{F}_0) - H(s, \mathcal{F}_0)\|_q \leq C|t-s|$ *for all* t, s*; (C3):* $\delta(k, q) = O(\chi^k)$ *for some* $0 \leq \chi < 1$ *and (C4):* $\beta_i(t)$*'s are* $\mathcal{C}^3$ *on* $[-\theta, 1]$*,* $i = 0, 1, \cdots, p$*. Then we have*

$$\max_{1 \leq i \leq n} \|X_i - X_i^*\|_q = O\Big(\frac{\log^2 n}{n}\Big). \tag{4}$$

The approximation rate $\log^2 n/n$ in (4) is optimal within a multiplicative logarithmic factor. We now discuss the regularity conditions (C1)-(C4). Condition (C1) is a classic assumption to avoid unit root or non-causal type of behavior in the sequence. Condition (C2) requires the innovation sequence to be locally stationary in the sense that H should be a smooth function of t. Condition (C3) requires that the innovations $\{\varepsilon_i\}$ are short range dependent with exponentially decaying physical dependence measures. Condition (C3) can be checked easily for various non-stationary linear and nonlinear models using the techniques in Zhou and Wu (2009). Finally, Condition (C4) requires the coefficient functions to be sufficiently smooth.

Remark 1. *The extra* $\log^2 n$ *term on the righthand side of* (4) *is due to the temporal dependence in the innovations* $\{\varepsilon_i\}$*. Indeed, if the series* $\{\varepsilon_i\}$ *is an independent sequence, then based on the proof of Proposition 1 in Section 6, it is easy to see that we can get rid of the extra* $\log^2 n$ *term in* (4) *and obtain the* $O(1/n)$ *optimal rate.*

Remark 2. *By the MA(∞) representation of* $G(t, \mathcal{F}_i)$ *and conditions (C1)-(C4), it is easy to show that* $G(t, \mathcal{F}_i)$ *also satisfies the stochastic Lipschitz condition in (C2). In other words,* $\|G(t, \mathcal{F}_0) - G(s, \mathcal{F}_0)\|_q \leq C|t-s|$ *for all* $t, s \in [0, 1]$.

3 SCRs for tv-AR models

In this section we shall construct SCRs for the coefficient functions $\beta_i(t)$. Theoretical results as well as practical implementations are discussed. First, we shall introduce local linear kernel regression which serves as the nonparametric methodology for the estimating of the model parameter functions.

3.1 Local linear estimation of parameter functions

Let $\boldsymbol{\beta}(t) = (\beta_0(t), \cdots, \beta_p(t))^\top$. In this paper, we will estimate the parameter function $\boldsymbol{\beta}(\cdot)$ via the local linear kernel method (Fan and Gijbels (1996))

$$(\hat{\boldsymbol{\beta}}(t), \hat{\boldsymbol{\beta}}'(t)) = \underset{\eta_0,\eta_1 \in \mathbb{R}^{p+1}}{\operatorname{argmin}} \sum_{i=p+1}^{n} (X_i - \bar{\mathbf{X}}_{i-1}^\top \eta_0 - \bar{\mathbf{X}}_{i-1}^\top \eta_1 (t_i - t))^2 K_{b_n}(t_i - t), \tag{5}$$

where $\bar{\mathbf{X}}_i = (1, X_i, X_{i-1}, \cdots, X_{i-p+1})^\top$, K is a kernel function, $b_n > 0$ is the bandwidth, and $K_c(\cdot) = K(\cdot/c)$, $c > 0$. Throughout paper we shall always assume that the kernel K is a symmetric density function with support $[-1, 1]$ and $K \in \mathcal{C}^1[-1, 1]$. Furthermore, define

$$\mu_i = \int_{\mathbb{R}} x^i K(x) dx \text{ and } \phi_i = \int_{\mathbb{R}} x^i K^2(x) dx, \quad i = 0, 1, \ldots.$$

Write

$$\mathbf{S}_{n,i}(t) = (nb_n)^{-1} \sum_{j=p+1}^{n} \bar{\mathbf{X}}_{j-1} \bar{\mathbf{X}}_{j-1}^\top [(t_j - t)/b_n]^i K_{b_n}(t_j - t),$$

for $i = 0, 1, \ldots$, here we let $0^0 = 1$, and

$$\mathbf{R}_{n,i}(t) = (nb_n)^{-1} \sum_{j=p+1}^{n} \bar{\mathbf{X}}_{j-1} X_j [(t_j - t)/b_n]^i K_{b_n}(t_j - t).$$

Let $\hat{\boldsymbol{\mu}}(t) = ([\hat{\boldsymbol{\beta}}(t)]^\top, b_n[\hat{\boldsymbol{\beta}}'(t)]^\top)^\top$. Then elementary calculations show that $\hat{\boldsymbol{\mu}}(t)$ has the following closed form representation

$$\hat{\boldsymbol{\mu}}(t) = \begin{pmatrix} \mathbf{S}_{n,0}(t) & \mathbf{S}_{n,1}^\top(t) \\ \mathbf{S}_{n,1}(t) & \mathbf{S}_{n,2}(t) \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{R}_{n,0}(t) \\ \mathbf{R}_{n,1}(t) \end{pmatrix} := \mathbf{S}_n^{-1}(t) \mathbf{R}_n(t). \tag{6}$$

3.2 Simultaneous confidence tubes

In practice, one may be interested in assessing the overall pattern of the regression functions $\boldsymbol{\beta}_{\mathbf{C}}(\cdot) := \mathbf{C}^\top \boldsymbol{\beta}(\cdot)$, where $\mathbf{C}$ is a fixed $(p+1) \times s$ full rank matrix with $s \leq p+1$. For instance, one may be interested in testing whether certain parameter function $\beta_k(\cdot)$ is different from zero, or whether it is constant or periodic over time. In this section we will approach the latter problems by constructing simultaneous confidence tubes (SCT) of $\boldsymbol{\beta}_{\mathbf{C}}(\cdot)$. Similar to the roles of confidence intervals in parametric inference, the SCT provides valuable information on the magnitude and pattern of the autoregressive relationship over time. The following theorem investigates the asymptotic behavior of the maximum deviation of estimated coefficient functions which facilitates the construction of SCT of $\boldsymbol{\beta}_{\mathbf{C}}(\cdot)$.

Theorem 1. *Assume that conditions (C1)-(C4) in Proposition 1 hold with* $q = 8$. *Further assume that* $\log^3 n/(n^{2/5}b_n) + nb_n^7 \log n \to 0$ *and that* $\Lambda(t)$ *is Lipschitz continuous and is positive definite on* $[0,1]$, *where*

$$\Lambda(t) = \sum_{i=-\infty}^{\infty} Cov(H(t,\mathcal{F}_0)\mathbf{G}(t,\mathcal{F}_0), H(t,\mathcal{F}_i)\mathbf{G}(t,\mathcal{F}_i)) \tag{7}$$

and $\mathbf{G}(t,\mathcal{F}_i) = (1, G(t,\mathcal{F}_{i-1}), G(t,\mathcal{F}_{i-2},) \cdots, G(t,\mathcal{F}_{i-p}))^\top$. *Then as* $n \to \infty$, *we have*

$$\mathbb{P}\left\{\frac{\sqrt{nb_n}}{\sqrt{\phi_0}} \sup_{t\in\mathcal{T}} |\Sigma_{\mathbf{C}}^{-1}(t)[\hat{\boldsymbol{\beta}}_{\mathbf{C}}(t) - \boldsymbol{\beta}_{\mathbf{C}}(t) - \frac{\mu_2 b_n^2 \boldsymbol{\beta}''_{\mathbf{C}}(t)}{2}]| - B_K(m^*) \le \frac{x}{\sqrt{2\log m^*}}\right\} \to e^{-2e^{-x}}, \tag{8}$$

where $\Sigma^2_{\mathbf{C}}(t) = \mathbf{C}^\top M^{-1}(t)\Lambda(t)M^{-1}(t)\mathbf{C}$, $\hat{\boldsymbol{\beta}}_{\mathbf{C}}(t) = \mathbf{C}^\top\hat{\boldsymbol{\beta}}(t)$, $\mathcal{T} = [b_n, 1-b_n]$, $m^* = 1/b_n$ *and*

$$B_K(m^*) = (2\log m^*)^{1/2} + \frac{\log C_K + (s/2 - 1/2)\log\log m^* - \log 2}{(2\log m^*)^{1/2}} \tag{9}$$

with $M(t) = \mathbb{E}[\mathbf{G}(t,\mathcal{F}_0)\mathbf{G}^\top(t,\mathcal{F}_0)]$ *and* $C_K = [\int_{-1}^1 |K'(u)|^2 du/(\phi_0\pi)]^{1/2}/\Gamma(s/2)$.

Based on (8), a $100(1-\alpha)\%$ SCT of $\boldsymbol{\beta}_{\mathbf{C}}(t)$ can be constructed as

$$\hat{\boldsymbol{\beta}}_{\mathbf{C}}(t) - \mu_2 b_n^2 \hat{\boldsymbol{\beta}}''_{\mathbf{C}}(t)/2 + \frac{\sqrt{\phi_0}}{\sqrt{nb_n}}\Big(B_K(m^*) + \frac{u_\alpha}{\sqrt{2\log m^*}}\Big)\hat{\Sigma}_{\mathbf{C}}(t)\mathcal{B}_s, \tag{10}$$

where $u_\alpha = -\log\log[(1-\alpha)^{-1/2}]$ and $\mathcal{B}_s = \{\mathbf{u} \in \mathbb{R}^s : |\mathbf{u}| \le 1\}$ is the unit ball.

3.3 Implementation

The SCT in (10) is of limited use in practice since it converges at $1/\sqrt{\log n}$ rate which is too slow to be useful. To alleviate the latter problem, we observe that $\hat{\boldsymbol{\beta}}_{\mathbf{C}}(t)$ can be well approximated by a Gaussian process uniformly on $[b_n, 1-b_n]$; see (11) below. Hence one could generate $\mathcal{B}$ (say 2000) i.i.d. copies of the latter Gaussian process and use the sampling distribution of those copies to construct the SCT. Furthermore, we notice that the asymptotic behavior of $\hat{\boldsymbol{\beta}}_{\mathbf{C}}(t)$ critically depends on the partial sum process of $\{\bar{\mathbf{X}}_{j-1}\varepsilon_j\}_{j=p+1}^n$. Recall that $\bar{\mathbf{X}}_i = (1, X_i, X_{i-1}, \cdots, X_{i-p+1})^\top$. Therefore we propose the following proposition which establishes a Gaussian approximation result for the latter partial sum process. The result is crucial for practical implementation of the SCTs.

Proposition 2. *Assume that conditions (C1)-(C4) in Proposition 1 hold. Then on a possibly richer probability space, there exist i.i.d. p-dimensional standard Gaussian random vectors $V_1, \cdots, V_n$ such that*

$$\max_{p+1\le i\le n} \Big| \sum_{j=p+1}^{i} \bar{\mathbf{X}}_{j-1}\varepsilon_j - \sum_{j=p+1}^{i} \Lambda(t_j)V_j \Big| = O_{\mathbb{P}}(n^{1/q'} \log^{3/2} n),$$

where $q' = \min\{q, 4\}$.

Based on Proposition 2, it is straightforward to derive that

$$\sup_{t\in\mathcal{T}} |\hat{\boldsymbol{\beta}}_{\mathbf{C}}(t) - \boldsymbol{\beta}_{\mathbf{C}}(t) - \frac{\mu_2 b_n^2 \boldsymbol{\beta}''_{\mathbf{C}}(t)}{2} - \boldsymbol{\Xi}(t)| = O_{\mathbb{P}}\Big(\frac{n^{-\nu}}{\sqrt{nb_n}\log^{1/2} n}\Big), \tag{11}$$

for some $\nu > 0$, where

$$\boldsymbol{\Xi}(t) = \Sigma_{\mathbf{C}}(t)\mu_{b_n}(t) \text{ with } \mu_{b_n}(t) = \sum_{i=1}^{n} V_i K_{b_n}(t_i - t)/(nb_n).$$

Hence the convergence is at algebraic rate. Now one can generate large i.i.d. copies of $\boldsymbol{\Xi}(t)$ and use their sampling maximum deviation to construct the SCT of $\hat{\boldsymbol{\beta}}_{\mathbf{C}}(t)$. The following are the detailed procedures:

(i) Find an appropriate bandwidth $\hat{b}_n$ and estimate $\hat{\Sigma}_{\mathbf{C}}(t) = [\mathbf{C}^\top \hat{M}^{-1}(t)\hat{\Lambda}(t)\hat{M}^{-1}(t)\mathbf{C}]^{1/2}$ using the methods described below.

(ii) Generate i.i.d. Gaussian vectors $V_1, V_2, \ldots, \sim N(0, \mathrm{I}_s)$ and calculate $\sup_{b_n\le t\le 1-b_n} |\mu_{b_n}(t)|$.

(iii) Repeat step (ii) for 2000 (say) times and obtain the estimated $(1-\alpha)$th quantile $\hat{q}_{1-\alpha}$ of $\sup_{b_n\le t\le 1-b_n} |\mu_{b_n}(t)|$.

(iv) Construct the $(1-\alpha)$th SCT of $\boldsymbol{\beta}_{\mathbf{C}}(t)$ as $\hat{\boldsymbol{\beta}}_{\mathbf{C}}(t) + \hat{\Sigma}_{\mathbf{C}}(t)\hat{q}_{1-\alpha}\mathcal{B}_s$.

To obtain $\hat{\Sigma}_{\mathbf{C}}(t)$, we suggest estimating $M(t)$ and $\Lambda(t)$ using the methods proposed in Zhou and Wu (2010). Specifically, $\hat{M}(t) = \mathbf{S}_{n,0}(t)$ and

$$\hat{\Lambda}(t) = \sum_{i=1}^{n} \omega(t,i)\Delta_i, \text{ where } \omega(t,i) = \frac{K_{\tau_n}(t_i - t)}{\sum_{k=1}^{n} K_{\tau_n}(t_k - t)},$$

$\Delta_i := (\sum_{j=-m}^{m} \mathbf{L}_{i+j})(\sum_{j=-m}^{m} \mathbf{L}_{i+j}^\top)/(2m+1)$ with $\mathbf{L}_i = \bar{\mathbf{X}}_{i-1}\hat{\varepsilon}_i$. Here m and τ_n are smoothing block size and bandwidth, respectively. The bandwidth b_n can be selected by the

Generalized Cross Validation (GCV) method discussed in Zhou and Wu (2010). The tuning parameters τ_n and m can be chosen by the minimum volatility method in Section 4.4 of Zhou and Wu (2010).

4 Simulation studies

There are two major purposes for the simulation studies conducted in this section. Firstly, we want to study the accuracy of the locally stationary approximation of model (1). According to the theory in Proposition 1, the errors in the latter approximation are asymptotically negligible. In this section we shall further study the influence of the approximation error on coverage probabilities of the SCT for finite samples. Secondly, we will study the accuracy of simulation based method in constructing SCT for finite samples. Specifically, consider the following time-varying AR(1) model

$$X_i = \zeta t_i^2 X_{i-1} + \varepsilon_i, \tag{12}$$

where $\{\varepsilon_i\}$'s are i.i.d. t distributed with 9 degrees of freedom and $\zeta \in (0,1)$ is a constant. Note that as ζ approaches 1, the dependence of the series is getting stronger and the locally stationary approximation to model (12)

$$X_i^* = G(t_i, \mathcal{F}_i) \text{ with } G(t, \mathcal{F}_i) = [1 - \zeta t^2 B]^{-1} \varepsilon_i \tag{13}$$

is expected to get less accurate. For models (12) and (13), simulated coverage probabilities for the SCT of $\beta_1(\cdot)$ with various choices of ζ and b_n are reported in Table 1 below. Sample sizes $n = 100$ and 400 with 1000 replicates.

We observe from Table 1 that the coverage probabilities for models (12) and (13) are almost identical, indicating that the locally stationary approximation for the tv-AR models are sufficiently accurate for finite sample inference. On the other hand, we observe that the simulation based method in Section 3.3 is reasonably accurate for the inference of $\beta_1(t)$, especially when b_n is in an appropriate range. Furthermore, we observe that coverage probabilities get worse when the dependence of the series become stronger and the accuracy improves as sample sizes get larger, which is consistent with one's intuition.

	Model (12)						Model (13)					
	90%			95%			90%			95%		
δ	0.5	0.7	0.9	0.5	0.7	0.9	0.5	0.7	0.9	0.5	0.7	0.9
$n = 100$												
$b = .1$	.248	.398	.647	.163	.272	.527	.244	.394	.67	.161	.278	.554
$b = .15$	.152	.276	.355	.084	.181	.245	.151	.27	.366	.083	.189	.259
$b = .2$	.115	.174	.206	.06	.106	.153	.115	.174	.213	.06	.109	.157
$b = .25$	.092	.151	.134	.048	.082	.085	.092	.15	.135	.048	.085	.087
$b = .3$	.087	.111	.104	.051	.058	.064	.086	.11	.105	.051	.059	.064
$n = 400$												
$b = .1$	.129	.172	.25	.073	.091	.167	.129	.172	.254	.073	.091	.167
$b = .15$	.106	.116	.153	.062	.063	.09	.106	.116	.149	.062	.063	.091
$b = .2$	.074	.095	.115	.046	.052	.054	.074	.095	.116	.046	.052	.055
$b = .25$	.071	.084	.104	.037	.046	.062	.071	.082	.105	.037	.046	.062
$b = .3$	.073	.072	.081	.037	.036	.043	.073	.072	.081	.037	.036	.043

Table 1. One minus simulated coverage probabilities of the SCT of $\beta_1(\cdot)$ at 90% and 95% nominal levels.

5 A real data example

In this section we shall analyze the monthly returns of the value-weighted index from the Center for Research in Security Prices (CRSP), University of Chicago. The period was Jan. 1926 to Dec. 1997 with a total of 864 observations. Figure 1 below displays the data. The series was used in Tsay (2005) to illustrate stationary AR and ARCH models. For a period as long as 72 years, one may suspect that the series is non-stationary and the AR relationship may change over time. On the other hand, there is a strong ARCH effect in the series indicating that it is of importance to model the errors of the AR model as dependent nonlinear series. In particular, previous tv-AR models with independent innovations do not seem to be appropriate for this data. In Tsay (2005), a stationary AR(3) model was fitted to the data. Here we shall fit the series with a tv-AR(3) model with dependent innovations. Particular interests are put in checking whether the auto regressive relationship are significant and whether the latter relationship is varying over time. Based on selection methods in Zhou and Wu (2010), the bandwidth was fixed at 0.2. 95% simultaneous confidence bands for the intercept and the auto-regressive coefficient functions are displayed in Figure 2 below.

We observe from Figure 2 that the lag 2 regression coefficient functions $\beta_2(t)$ is not significantly different from 0 and all other $\beta_i(t)$'s are significant, which is consistent with the findings in Tsay (2005). However, an important observation from Figure 2 is that $\beta_i(t)$'s $i = 1, 3$ are significantly time varying for this period of time. The time-varying association is most significant for $\beta_3(t)$, which exhibits a significant increasing trend for the first half of the period. The latter trends cannot be disclosed from stationary linear time series analysis. Furthermore, our finding indicates that forecasting the time series using the whole sequence as was performed in Tsay (2005) may lead to suboptimal results because of the change of the auto-regressive relationship over time. From Figure 2, it seems more appropriate to use the second half of the time series for the forecasting since the AR(3) relationship appears to be stable in the latter period.

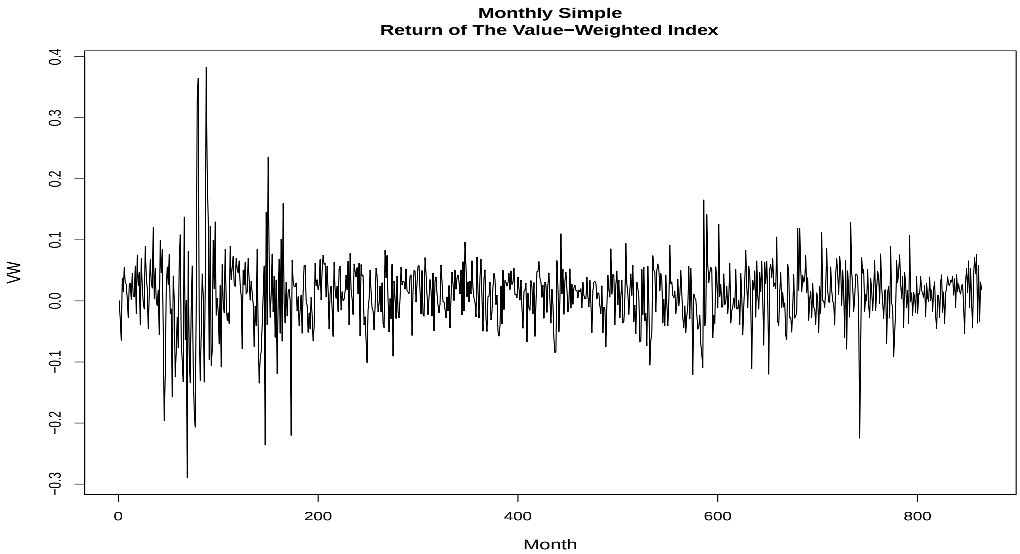

Figure 1: Time series plot of the monthly return of the value-weighted index from Jan. 1926 to Dec. 1997.

Acknowledgements. I am grateful to the anonymous referee for the many helpful com-

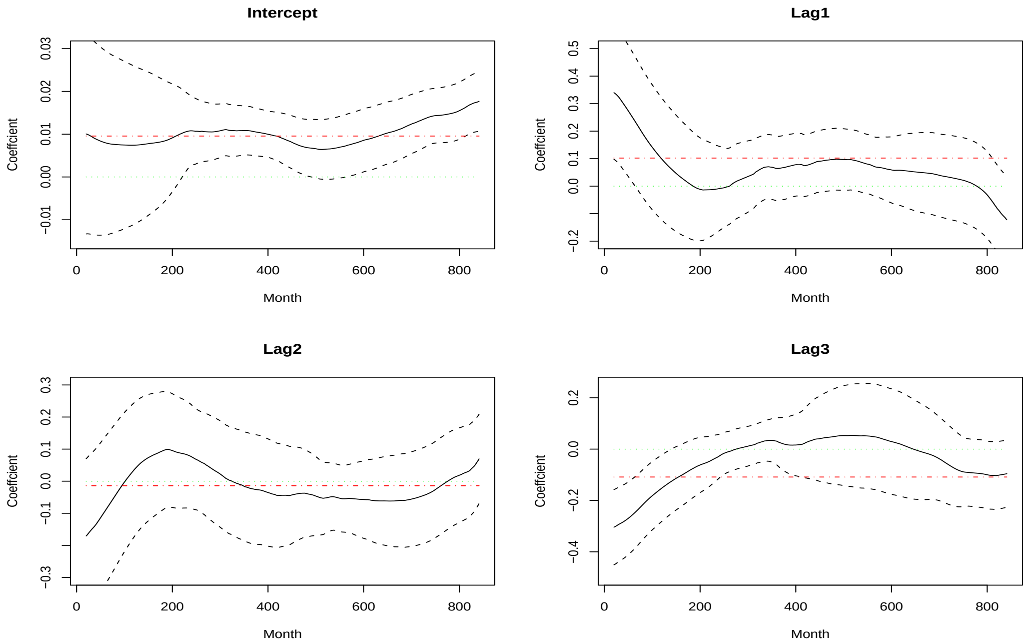

Figure 2: 95% simultaneous confidence bands for the AR coefficient functions $\beta_i(t)$ $i = 0, 1, 2, 3$ of the value-weighted index. The dotted horizontal lines are $\beta_i(t) = 0$, $i = 0, \cdots, 3$. The dotdash lines are the fitted constant trends $\beta_i(t) = \hat{c}_i$, $i = 0, \cdots, 3$.

ments which lead to a substantial improvement of the paper. The research was supported in part by NSERC of Canada.

6 Proofs

Proof of Proposition 1. For simplicity we shall only prove the case $\theta = 1$. The other cases follow by essentially the same arguments. Define $\mathbf{X}_i = (X_i, X_{i-1}, \cdots, X_{i-p+1})$. Then by the auto-regressive recursion (1), we have

$$\mathbf{X}_i = \boldsymbol{\beta}_0(t_i) + A(t_i)\mathbf{X}_{i-1} + \pi_i, \tag{14}$$

where $\pi_i = (\varepsilon_i, 0, 0, \cdots, 0)^\top$, $\boldsymbol{\beta}_0(t) = (\beta_0(t), 0, \cdots, 0)^\top$, $A_{1\cdot}(t) = (\beta_1(t), \beta_2(t), \cdots, \beta_p(t))$ and $A_{i\cdot}(t) = \mathbf{1}_{i-1}^\top$ for $2 \le i \le p$. Here $\mathbf{1}_i$ denotes the length-p vector with the ith entry being 1 and the other entries being 0. Note that the recursion in (14) is of order 1 and is easier to handle than the order p recursion in (1). It is easy to obtain that the recursion in (14) leads to

$$\mathbf{X}_i = \sum_{j=0}^{i+n} B_n(i,j)[\pi_{i-j} + \boldsymbol{\beta}_0(t_{i-j})] + \prod_{j=0}^{i+n} A(t_{i-j})\mathbf{X}_{-n-1},$$

where $B_n(i,j) = \prod_{k=0}^{j} A(t_{i-k})$, $j \geq 1$ and $B_n(i,0) := 1$. Define $B_n^*(i,j) = [A(t_i)]^j$, $j \geq 1$ and $B_n^*(i,0) = 1$. Based on condition (C1), we obtain that, for every i,j, $\max\{|B_n(i,j)|, |B_n^*(i,j)|\} = O(\rho^j)$ for some $\rho \in [0,1)$. Furthermore, an elementary induction argument with the aid of condition (C4) leads to

$$|B_n(i,j) - B_n^*(i,j)| \leq C(j+1)^2\rho^j/n \tag{15}$$

for every i,j. Now by Condition (C3) and Lemma 6 in Zhou (2011), we have

$$\max_i \|\mathbf{X}_i - \mathbf{X}_i^{**}\|_q = O(\sqrt{\sum_{j=0}^{\infty} |B_n(i,j) - B_n^*(i,j)|^2} = O\Big(\frac{1}{n}\Big),$$

where $\mathbf{X}_i^{**} = \sum_{j=0}^{i+n} B_n^*(i,j)[\pi_{i-j} + \boldsymbol{\beta}_0(t_{i-j})] + B_n^*(i,i+n)\mathbf{X}_{-n-1}$. Now define

$$\mathbf{X}_i^{***} = \sum_{j=0}^{i+n} B_n^*(i,j)[\pi_{i,i-j}^* + \boldsymbol{\beta}_0(t_i)] + B_n^*(i,i+n)\mathbf{X}_{-n-1},$$

where $\pi_{i,i-j}^* = H(t_i, \mathcal{F}_{i-j})$. Using the conditions (C2) and (C3), it is easy to show that

$$\max_{i,j,k,r} \|H(t_k,\mathcal{F}_i) - H(t_j,\mathcal{F}_i) - [H(t_k,\mathcal{F}_{i,r}) - H(t_j,\mathcal{F}_{i,r})]\|_q \leq C\min\{\frac{|j-k|}{n}, \delta(r,q)\}.$$

Hence by Lemma 6 in Zhou (2011), we obtain

$$\max_i \|\mathbf{X}_i^{**} - \mathbf{X}_i^{***}\|_q = O\Big(\sqrt{\sum_{j=0}^{\infty} \Big[\min\{\frac{j}{n}, \delta(j,q)\}\Big]^2}\Big) = O\Big(\frac{\log^2 n}{n}\Big).$$

Finally, let us define $\mathbf{X}_i^{****} = \sum_{j=0}^{\infty} B_n^*(i,j)[\pi_{i,i-j}^* + \boldsymbol{\beta}_0(t_i)]$. Then elementary arguments lead to

$$\max_i \|\mathbf{X}_i^{***} - \mathbf{X}_i^{****}\|_q = O\Big(\frac{1}{n}\Big).$$

Note that X_i^* is the first element in $\mathbf{X}_i^{****}$. Hence Proposition 1 follows. $\diamondsuit$

Lemma 1. *Let* $\bar{\mathbf{X}}_i^* = (1, X_i^*, X_{i-1}^*, \cdots, X_{i-p+1}^*)$,

$$\mathbf{S}_{n,i}^*(t) = (nb_n)^{-1} \sum_{j=p+1}^{n} \bar{\mathbf{X}}_{j-1}^* [\bar{\mathbf{X}}_{j-1}^*]^\top [(t_j - t)/b_n]^i K_{b_n}(t_j - t),$$

for $i = 0, 1, \ldots$, *here we let* $0^0 = 1$, *and*

$$\mathbf{R}_{n,i}^*(t) = (nb_n)^{-1} \sum_{j=p+1}^{n} \bar{\mathbf{X}}_{j-1}^* X_j^* [(t_j - t)/b_n]^i K_{b_n}(t_j - t).$$

Define $\mathbf{S}_n^*(t)$ *and* $\mathbf{R}_n^*(t)$ *similarly as in* (6). *Define* $\hat{\boldsymbol{\mu}}^*(t) = (\mathbf{S}_n^*(t))^{-1}\mathbf{R}_n^*(t)$. *Assume that conditions (C1)-(C4) in Proposition 1 hold with some* $q \geq 4$ *and* $b_n \to 0$ *with* $nb_n^2 \to \infty$. *Then we have*

$$\Big\| \sup_{t\in[0,1]} |\hat{\boldsymbol{\mu}}(t) - \hat{\boldsymbol{\mu}}^*(t)| \Big\|_{q/4} = O\Big(\frac{\log^2 n}{nb_n}\Big). \tag{16}$$

Proof. Note that

$$\begin{aligned} \hat{\boldsymbol{\mu}}(t) - \hat{\boldsymbol{\mu}}^*(t) &= [(\mathbf{S}_n(t))^{-1} - (\mathbf{S}_n^*(t))^{-1}][\mathbf{R}_n(t) - \mathbf{R}_n^*(t)] + [(\mathbf{S}_n(t))^{-1} - (\mathbf{S}_n^*(t))^{-1}]\mathbf{R}_n^*(t) \\ &+ (\mathbf{S}_n^*(t))^{-1}[\mathbf{R}_n(t) - \mathbf{R}_n^*(t)]. \end{aligned} \tag{17}$$

By the proof of Proposition 7 in Zhou and Wu (2009), it is easy to show that

$$\| \sup_{t\in[0,1]} (\mathbf{S}_n^*(t))^{-1} \|_{q/2} = O(1) \text{ and } \| \sup_{t\in[0,1]} \mathbf{R}_n^*(t) \|_{q/2} = O(1). \tag{18}$$

Additionally, by Proposition 1 and elementary algebra, we have

$$\| \sup_{t\in[0,1]} |(\mathbf{S}_n(t))^{-1} - (\mathbf{S}_n^*(t))^{-1}| \|_{q/2} + \| \sup_{t\in[0,1]} |\mathbf{R}_n(t) - \mathbf{R}_n^*(t)| \|_{q/2} = O\Big(\frac{\log^2 n}{nb_n}\Big). \tag{19}$$

Therefore the lemma follows from (17) to (19) and the Hölder's inequality. ♢

Proof of Theorem 1. According to Lemma 1, it suffices to show that (8) holds with $\hat{\mu}(t)$ therein replaced by $\hat{\mu}^*(t)$. On the other hand, by conditions (C1) and (C3), it is elementary to show that the physical dependence measures for $\{\bar{\mathbf{X}}_j^*\}$ and $\{\bar{\mathbf{X}}_{j-1}^* \varepsilon_j\}$ are exponentially decreasing and $M(t)$ is invertible for all $t \in [0, 1]$. Hence Theorem 1 follows from Remark 2 and Theorem 3 in Zhou and Wu (2010). Details are omitted. ♢

Proof of Proposition 2. Define $\mathbf{X}_i^* = (X_i^*, X_{i-1}^*, \cdots, X_{i-p+1}^*)$. Note that

$$\begin{aligned}\max_{p+1\le i\le n}\Big|\sum_{j=p+1}^{i}\mathbf{X}_{j-1}X_j - \sum_{j=p+1}^{i}\mathbf{X}_{j-1}^*X_j^*\Big| &\le \max_{p+1\le i\le n}\sum_{j=p+1}^{i}|\mathbf{X}_{j-1}X_j - \mathbf{X}_{j-1}^*X_j^*| \\ &= \sum_{j=p+1}^{n}|\mathbf{X}_{j-1}X_j - \mathbf{X}_{j-1}^*X_j^*|.\end{aligned}$$

On the other hand, by Cauchy's inequality and Proposition 1,

$$\begin{aligned}\Big\|\sum_{j=p+1}^{n}|\mathbf{X}_{j-1}X_j - \mathbf{X}_{j-1}^*X_j^*|\Big\|_{q/2} &\le \sum_{j=p+1}^{n}\|\mathbf{X}_{j-1}X_j - \mathbf{X}_{j-1}^*X_j^*\|_{q/2} \\ &\le \sum_{j=p+1}^{n}[\|\mathbf{X}_{j-1} - \mathbf{X}_{j-1}^*\|_q\|X_j\|_q + \|\mathbf{X}_{j-1}^*\|_q\|X_j - X_j^*\|_q] \\ &= O(\log^2 n).\end{aligned}$$

Hence

$$\max_{p+1\le i\le n}\Big|\sum_{j=p+1}^{i}\mathbf{X}_{j-1}X_j - \sum_{j=p+1}^{i}\mathbf{X}_{j-1}^*X_j^*\Big| = O_{\mathbb{P}}(\log^2 n).$$

Similarly $\max_{p+1\le i\le n}|\sum_{j=p+1}^{i}\mathbf{X}_{j-1}\varepsilon_j - \sum_{j=p+1}^{i}\mathbf{X}_{j-1}^*\varepsilon_j| = O_{\mathbb{P}}(\log^2 n)$. Note that for the locally stationary time series $\{\bar{\mathbf{X}}_{j-1}^*\varepsilon_j\}$, its physical dependence measures decay exponentially fast to zero. Hence Proposition 2 follows from Corollaries 1 and 2 in Wu and Zhou (2011). $\diamondsuit$

REFERENCES

Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Ann. Statist.* **25** 1–37.

Dahlhaus, R. (2000). A likelihood approximation for locally stationary processes. *Ann. Statist.* **28** 1762–1794.

Dahlhaus, R. (2011). Locally stationary processes - a review. *Handbook of Statistics* **30** to appear.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*, Chapman and Hall.

Fryzlewicz, P. and Nason, G. P. (2006). Haar-Fisz estimation of evolutionary wavelet spectra. *J. R. Statist. Soc. B* **68** 611–634.

Kitagawa, G. and Gersch, W. (1985). A smoothness priors time-varying AR coefficient modeling of the nonstationary covariance time series. *IEEE Trans. Automat. Control.* **30** 48–56.

Ombao, H.C., von Sachs, R. and Guo, W. (2005). The SLEX analysis of multivariate non-stationary time series. *J. Amer. Statist. Assoc.* **100** 519–531.

Priestley, M.B. (1988). *Nonlinear and Nonstationary Time Series Analysis.* Academic Press, London.

Tsay, R. S. (2005). *Analysis of Financial Time Series, second edition.* Wiley, Hoboken, New Jersey.

Wu, W. B. and Zhou, Z. (2011). Gaussian approximations for non-stationary multiple time series. *Statistica Sinica* **21** 1397–1413.

Zhou, Z. (2011). Heteroscedasticity and autocorrelation robust structural change detection. Preprint.

Zhou, Z. and Wu, W. B. (2009). Local linear quantile estimation for non-stationary time series. *Annals of Statistics* **37** 2696–2729.

Zhou, Z. and Wu, W. B. (2010). Simultaneous inference of linear models with time varying coefficients. *J. R. Statist. Soc. B* **72** 513–531.