Asymptotic Expression for the Mean Squared Prediction Error of Self-Exciting Threshold Autoregressive Models

Chien-Ming Chi∗1, Ching-Kang Ing†2, and Howell Tong‡3

1Department of Economics, National Taiwan University
2Department of Statistics, National Tsing Hua University
3University of Electronic Science & Technology, China
3London School of Economics, United Kingdom

April 26, 2018

Abstract

Time series variables in economics and finance tend to be regime-dependent. For effective modelling, a non-linear time series model such as self-exciting threshold autoregressive (SETAR) model, among many others, is often applied to capture the underlying relation between variables. To most applicants, one of the major concern when modelling is which estimators to use, or equivalently provided SETAR(\(p\)) models (i.e. SETAR models of order \(p\)) are considered, how many lagged terms to include. This question poses serious challenge to researchers as most model evaluation/selection methods require model linearity and other types of analysis-friendly model specifications. In the absence of such model specifications, researches on SETAR model analysis, especially from a prediction error point of view, are often impeded much. In this paper, we overcome this difficulty and provide an answer to part of the above question by rigorously deriving and analyzing the asymptotic expression for mean squared prediction error (AMSPE) of the SETAR(\(p\)) model and extend the results to \(h\)-step prediction error. Specifically, we obtained a surprisingly intriguing closed-form expression for SETAR(\(p\)) \(h\)-step predictor under mild conditions that implies prediction error increases by \(k\sigma^2\) when \(k\) variables are overfitted; knowledge of such expression can then be used to develop an information criterion to consistently select ideal model. Simulations are given to illustrate our main results.

1 Introduction

For years, regime switching behaviour of economic variables such as output growth, interests rate, stock returns has been observed and reported by a substantial empirical

∗xbb66kw@gmail.com
†cking@stat.nthu.edu.tw
‡howell.tong@gmail.com
studies. To gain the benefit for modelling, either in the sense of better performance on a various measures of fits or drawing a more reliable inference, econometricians propose nonlinear time series models. Among many others, self-exciting threshold autoregressive (SETAR) models (Tong and Lim (1980)), Markov switching models (Hamilton (1989)), bilinear models (Granger and Andersen (1978)) and neural network models are arguably the most widely used nonlinear time series tools for such purposes. By design, SETAR model are especially easy to understand and interpret as it is just extending the simple autoregressive (AR) model to allow for nonlinear behavior. For this reason, numerous research works have applied SETAR to study, just to name a few, electricity price forecasting (Weron (2014)), financial time series (Tsay (2002)). The empirical usefulness of SETAR model attracts a lot theoretical interests. These researches range widely from modeling to techniques for testing the existence of threshold and selecting proper lags for SETAR model under certain scenarios (Hansen (1996, 2000), Tsay (1998), Gonzalo and Pitarakis (2002), Kapetanios (2001)).

On the other hand, mean squared prediction error, due to its straightforward interpretation, is frequently employed to construct proper measurement to evaluate fitted models from a various perspectives: inference about predictive ability (West (1996), Clark and West (2000), Greenway-McGrevy (2013)); properties of predictors (Kunitomo and Yamamoto (1985), Ing (2003)); model evaluation and model selection (Ing and Wei (2002), Hue et al. (2018)). Among these works, Kunitomo and Yamamoto (1985), Ing and Wei (2002) study the AMSPE when the fitted model is not correctly specified; realizing AMSPE remains the same with or without fitted model correctly specified, Hue et al. (2018) gives "misspecified resistance information criterion" that is able to perform consistent model selection without assuming true model. As pointed out by these works, what makes the perspective of prediction very valuable, besides its intuition, are the exclusive distinct features that mostly shared with evaluation methodologies constructed based on prediction viewpoint (see, in particular, Hue et al. (2018)).

However, the aforementioned analysis might not be easy to proceed in the context of SETAR process without fundamental analytical knowledge. As SETAR model has attracted so much attention from both practitioners and theorists, we consider the study of AMSPE of SETAR model a crucial task that deserved to be addressed. In this paper, we take the step toward this end and derive a closed-form expression for the asymptotic mean squared prediction error of a modified estimation procedure of the SETAR ($p$), $h$-step predictor with single threshold. To the authors' knowledge, these expressions (see Theorem 1 to 3) have never been mentioned or expected before our work. Moreover, these expressions are surprisingly instructive. To exemplify, the main result for one step prediction of SETAR ($p$) has a extraordinary implication that when we over-fit a SETAR ($p$) model by an extra $k$ lags, its asymptotical mean squared predition error simply increases by $2k\sigma^2$. 


In the language of model selection, the prediction error of the SETAR model is still linear in the number of overfitted lags as that in a linear time series model; and this can definitely benefit the authors’ analysis while conducting the model selection criterions like MRIC for SETAR modeling. Given the somewhat intricate asymptotic analysis results of SETAR in Chan(1992)(especially the result for threshold estimator), it is really to our amaze to see such simplicity in over-fitting prevention mechanism of SETAR models. Unlike the closed form expression in the main results, which are user-friendly for both practical and theoretical purpose, the rigorous derivation of which is by all mean a nasty task. Most works, including studies of a set of non-traditional moment bounds and fundamental results for SETAR process are relegated to the supplementary file. The rest of the paper is organized as follows. In Section 2, we introduce our modified estimation procedure and limiting distribution of normalized threshold estimator. Main results and simulation results are presented in section 3. Some of the most relevant propositions and moment bounds on normalized estimators are listed in section 4 with proof relegated to appendix and supplementary file. Assumptions for main results and comments and notes on which are deferred to appendix.

2 Model Setting and the Modified Estimates

Let \((\Omega, \mathcal{F}, P)\) denote the probability space. In this paper, we consider SETAR model with order \(p\) and only one threshold; the general model setting is similar to that in Chan(1992): for \(t > p\),

\[
x_t = \begin{cases} 
a_{10} + a_{11}x_{t-1} + \cdots + a_{1p}x_{t-p} + e_t = a_{10} + \tilde{a}_{1} \tilde{x}_{t-1} + e_t, & x_{t-d} \leq r, \\
a_{20} + a_{21}x_{t-1} + \cdots + a_{2p}x_{t-p} + e_t = a_{20} + \tilde{a}_{2} \tilde{x}_{t-1} + e_t, & x_{t-d} > r,
\end{cases}
\]

where \((a_{10}, \tilde{a}_{1})'\) and \((a_{20}, \tilde{a}_{2})'\), lie in \(\mathbb{R}^{p+1}\); \(e_t, t = 1, \ldots\) denote the innovative random variables which are independent of previous information, \(\mathcal{F}_{t-1}\). More assumptions on the SETAR process can be found in Appendix. The process \(\tilde{x}_t\) admits a stationary distribution whose properties have been thoroughly studies in Tong and Chan(1985), Chan(1988), and we shall denote the distribution’s density function as \(\tilde{\pi}\). The distributions of \(e_t\) and the marginal stationary distribution are denoted by \(f_e\) and \(\pi\), respectively. We assume the intercepts are zeros and drop \(a_{i0}\)'s for the ease of expression; all results in this work apply to the case where \(a_{i0} \neq 0, i = 1, 2\). A convention of notation usage we adapt in this paper is that when \(x_{i-d} \leq r\), then we say ”\(x_i\) is in state 1” and vice versa; we tend to express information accordingly(e.g. coefficients vector \(\tilde{a}_i, i = 1\) in state 1),
2.1 Conditional Ordinary Least Square Estimator

Given the data generated from SETAR\((p)\) model with sample size \(n\), presumably the most intuitive way to estimate the model parameters is by defining

\[
L(\tilde{\beta}_1, \tilde{\beta}_2, \beta_3, \beta_4) = \begin{cases} 
\infty, & \text{on } E, \\
\sum_{x_{i+1-\beta_4} \leq \beta_3} (\tilde{x}_i \tilde{\beta}_1 - x_{i+1})^2 + \sum_{x_{i+1-\beta_4} > \beta_3} (\tilde{x}_i \tilde{\beta}_2 - x_{i+1})^2, & \text{o.w.}
\end{cases}
\]

where

\[E = \begin{cases} 
\text{Either } \sum_{x_{i+1-\beta_4} \leq \beta_3} \tilde{x}_i \tilde{x}_i' \text{ or } \sum_{x_{i+1-\beta_4} > \beta_3} \tilde{x}_i \tilde{x}_i' \text{ is singular}
\end{cases},
\]

and summation is subject to \(p \leq i \leq n-1\); and calculating the conditional ordinary least square estimators (Chan(1992)) by

\[
(\hat{a}_{1n}, \hat{a}_{2n}, \hat{c}_n, \hat{d}_n) = \arg \min_{(\tilde{\beta}_1, \tilde{\beta}_2, \beta_3, \beta_4) \in T} L(\tilde{\beta}_1, \tilde{\beta}_2, \beta_3, \beta_4),
\]

where in \(T\), coefficients optimization is not subject to any restriction; threshold candidates of the optimization are all sample points; and thresholding lag is an integer smaller than the time series order \(p\). To be precise,

\[T = \{ (\tilde{\beta}_1, \tilde{\beta}_2, \beta_3, \beta_4) : \tilde{\beta}_i \in \mathbb{R}^p, i = 1, 2; \beta_3 \in \{x_1, \ldots, x_n\}; \beta_4 \in \{1, \ldots, p\}\}.
\]

We drop off the subscript \(n\) of the estimators if no notation confusion is possible.

2.2 Modified Estimation

To reduce unnecessary complexity of the derivation and facilitate understanding the contribution of this paper, we introduce a modified estimation scheme. The modification mainly focuses on

i) the finiteness of threshold parameter, and
ii) the boundness of the inverse sample covariance matrices.

For the finiteness, simply putting a restriction, for example, \([-C, C]\) for some \(C > 0\), on \(\beta_3\) in the optimization seems to be a straightforward idea of getting a restricted threshold parameter space; this however induces a tricky and non-trivial situation where the optimization is over all infinite values or an empty set: there is no guarantee samples will fall onto the real line nicely. We tackle this problem by defining a rare event, \(\mathcal{RE}_n\), such that on \(\mathcal{RE}_n\), for some \(1 \leq d \leq p\),

\[
\text{At least } p \ x_{i-d}'s, i = p + 1, \ldots, n, \text{ fall inside a compact set, } \Theta' \subset \mathbb{R}^1.
\]

A mathematical definition of \(\mathcal{RE}_n\) as well as \(\Theta'\) can be found in Appendix. The modified version of \(T\) is

\[
T' = \{ (\tilde{\beta}_1, \tilde{\beta}_2, \beta_3, \beta_4) : \tilde{\beta}_i \in \mathbb{R}^p, i = 1, 2; \beta_3 \in \{x_1, \ldots, x_n\} \cap \Theta'; \beta_4 \in \{1, \ldots, p\}\}.
\]
On the other hand, the boundness is embedded into another subset $U_n$:

$$U_n^c = \bigcup_{j=1,2} \left( B_j(\hat{r}, \hat{d}) \cap \left\{ (\hat{\beta}_1, \hat{\beta}_2, \hat{r}, \hat{d}) = \arg \min_{\hat{\beta}} L(\beta_1, \beta_2, r, d) \right\} \right) \cap \mathcal{RE}_n^c,$$

where $B_1(r, d) = \left\{ \left\| \sum_{x_{i+1-d} \leq r} \hat{x}_i \hat{x}_i' \right\|^2 \geq s n \right\}$ for some $s > 0$, and $B_2(r, d)$ is defined in an analogous way. The modified estimators are

$$\left\{ \begin{array}{l}
\arg \min_{(\hat{\beta}_1, \hat{\beta}_2, \beta_3, \beta_4) \in \mathcal{T}_L} L(\hat{\beta}_1, \hat{\beta}_2, \beta_3, \beta_4), \text{ on } U_n \cap \mathcal{RE}_n^c,
(\hat{a}_1, \hat{a}_2, \hat{r}, \hat{d}) = (0, 0, 0, 1), \text{ o.w.}
\end{array} \right.$$ 

Our modification is mild. Contrary to the linear time series model, where the negative moment bounds of sample covariance impose a serious challenge on the analytical issues, here, we simply set a lower bound, $sn$; we prove that $P(U_n^c)$ decreases to 0 in a speed faster than an arbitrarily high polynomial order of $n^{-1}$. This seemingly evasive approach to the problem of inverse moment bounds – from a well-established negative moment bound in the linear case to a mere probabilistic lower bound for minimal eigenvalue – actually reflects one nature of a SETAR process: $\# \{ i : x_{i+1-d} > r, p < i \leq n \}$ and its state 1 analog might be less than $p$, a situation where one of the minimal eigenvalue of the true covariance matrices is zero, with a non-zero probability for all $n$. In light of this observation, our analytical expediency is almost standing at the edge of what can be done, and indeed helps avoiding unnecessary notation burden.

### 2.3 Limiting Distribution of Threshold Estimator

In order to describe the limiting distribution of $n(\hat{r}_n^c - r)$, we define two compound Possion processes: Let $\mathcal{P}_k(z)$, $k = 1, 2$, be Possion processes with rate $\pi(r)$, $\mathcal{P}_k(0) = 0$ a.s., and the independent jumps for $\mathcal{P}_k$ are

$$\begin{align*}
\kappa_1 &\triangleq \left[ (\hat{a}_1 - \tilde{a}_2) \hat{x}_i^{\text{lim}} \right]^2 + 2\hat{x}_i^{\text{lim}} (\hat{a}_1 - \tilde{a}_2) \epsilon_1, \\
\kappa_2 &\triangleq \left[ (\hat{a}_2 - \tilde{a}_1) \hat{x}_i^{\text{lim}} \right]^2 + 2\hat{x}_i^{\text{lim}} (\hat{a}_2 - \tilde{a}_1) \epsilon_1,
\end{align*}$$

respectively, where $\hat{x}_i^{\text{lim}}$ is defined as

$$\hat{x}_i^{\text{lim}}(x_{T_m}, \ldots, x_{T_m+2-d}, r, x_{T_m-d}, \ldots, x_{T_m-p+1}),$$

where $T_m = \inf\{ i : x_{i+i-d} \in B_{am}(r), i > p \}$, for any $\{a_m\} \downarrow 0$. Since these random jumps have positive mean, there’s an unique random interval $[M_-, M_+]$ such that $\{ z : \mathcal{P}_1(-z)1_{z \leq 0} + \mathcal{P}_2(z)1_{z > 0} \}$ reaches its minimization on $[M_-, M_+]$ and $M_-, M_+$ are consecutive jump points. The limiting distribution $r_\infty \triangleq M_-$. We prove in the supplementary file that $E|r_\infty| < \infty$. 

5
2.4 *h*-step Estimator

Even with only one threshold considered in (1), the analysis and calculation of *h*-step \((d \geq h)\) predictor requires knowledge of the true/false vector \((1_{x_n+1-d>r}, \ldots, 1_{x_n+1-d>r})^t\); we thus define and use the function \(TH := \{1, 2\}^h \rightarrow \mathcal{F}_n\) for referring the corresponding \(2^h\) events of the zero/one vector (plus one). In particular, on \(TH(\{s_1, \ldots, s_h\})\),

\[
\begin{cases}
  x_{n+i-d} > r, & \text{if } s_i = 2, \\
  x_{n+i-d} \leq r, & \text{if } s_i = 1,
\end{cases}
\]

Then by (1), on \(TH(s), s = \{s_1, \ldots, s_h\}\),

\[
x_{n+h} = \hat{x}_n A_{s_1} \times \cdots \times A_{s_h} + \sum_{j=0}^{h-1} b_j(s)e_{n+h-j},
\]

(3)

where the functions \(b_j := \{1, 2\}^h \rightarrow \mathbb{R}^1, j = 0, \ldots, h - 1\), are defined by (3) and (1) with

\[
A_i = \left( \tilde{a}_i \begin{bmatrix} I_{p-1} \\ 0 \end{bmatrix} \right), i = 1, 2.
\]

With the sample analog of \(TH(s)\) and \(A(.)\), where \(r, \tilde{a}_j, d\) have been replaced with \(\hat{r}_n, \hat{a}_{jn}, \hat{d}_n\) (subscript \(n\) is often ignored), respectively, the \(h\)-step predictor is given by, on \(TH(s)\),

\[
\hat{x}_{n+h} = \begin{cases}
  \hat{x}_n^\prime \hat{A}_i \times \cdots \times \hat{a}_{s_h}, & \text{if } h \leq \hat{d}_n, \\
  0, & \text{if } h > \hat{d}_n,
\end{cases}
\]

where

\[
\hat{A}_i = \left( \hat{a}_i \begin{bmatrix} I_{p-1} \\ 0 \end{bmatrix} \right), i = 1, 2.
\]

In particular, one-step predictor is

\[
\hat{a}_{n+1} = \begin{cases}
  \hat{a}_n^\prime \hat{x}_n, x_{1+n-d} \leq \hat{r}_n, \\
  \hat{a}_n^\prime \hat{x}_n, x_{1+n-d} > \hat{r}_n.
\end{cases}
\]

the tilde is dropped off in the scalar case, the case with order 1. The multi-steps predictor depends on events \(TH(s), s \subset \{1, 2\}^h\); the probabilities of \(TH(s)\)’s therefore appears in the expected prediction error in contrast to the linear case in Ing(2003). By \(TH(s') \cap TH(s'') = \emptyset\) if \(s' \neq s''\),

\[
E \left[ (\hat{x}_{n+h}^o - x_{n+h})^2 | h \leq \hat{d}_n \right] = \sum_s E \left[ (\hat{x}_{n+h}^o - \hat{x}_n A(TH(s))^2 1_{TH(s)} | h \leq \hat{d}_n \right] \sum_{j=0}^{h-1} P(TH(s)) b_j(s) \sigma^2,
\]

(4)

where \(A(TH(s)) = A_{s_1} \times \cdots \times a_{s_h}\) and the summation is over all possible \(s \subset \{1, 2\}^h\).
3 Main results

3.1 Main Result for SETAR(1)

Assumptions and discussion of them are deferred to the appendix.

Theorem 1. Assume assumptions 1 to 5,

\[
\lim_{n \to \infty} n \left[ E \left( x_{n+1} - \hat{x}_{n+1}^o \right)^2 - \sigma^2 \right] = (a_1 - a_2)^2 r^2 \pi(r) E|r_\infty| + 2\sigma^2.
\]

From the RHS of the equation, the prediction error is decomposed into two terms, \((a_1 - a_2)^2 r^2 \pi(r) E|r_\infty|\) and \(2\sigma^2\). The first term resulted from wrong prediction of the state of the \(x_{n+1}\) has quite a novel structure. From the distribution of \(r_\infty\), the rate \(\pi(r)\) cancels the multiplication of \(\pi(r)\). This means \(\pi(r) E|r_\infty|\) is equivalent to \(E|r_\infty^c|\) for \(r_\infty^c\) distributed exactly the same as \(r_\infty\) but with rate 1. Based on this fact, we do not need to consider the point value of stationary density function. Also notice that as \((a_1 - a_2)^2 r^2\) decreases, the mean of independent jumps of the compound Poisson processes (see section 2.3) decreases and hence \(\pi(r) E|r_\infty|\) increases. In general they do not cancel each other, but we see the whole first term of prediction error remains stable in our simulation results. On the other hand, \(2\sigma^2\) represents the regular prediction error attributed to variance in coefficients estimation. This terms, which we also refer to as the AR prediction error if compared to Ing(2003), is considered as the penalty of model complexity and is increasing linearly as the number of lags added in the model; for this, also see the expression for SETAR\((p)\).

Proof of Theorem 1. Since

\[
(n \left[ E \left( x_{n+1} - \hat{x}_{n+1}^o \right)^2 - \sigma^2 \right]) =nE \left[ (a_1 x_n - \hat{a}_1^o x_n)^2 1_{(x_n \leq \hat{r}^o_n, r)} \right] \\
+nE \left[ (a_2 x_n - \hat{a}_2^o x_n)^2 1_{(x_n > \hat{r}^o_n, r)} \right] \\
+nE \left[ (a_1 x_n - \hat{a}_2^o x_n)^2 1_{(x_n \leq r, x > \hat{r}^o_n)} \right] \\
+nE \left[ (a_2 x_n - \hat{a}_1^o x_n)^2 1_{(x_n > r, x \leq \hat{r}^o_n)} \right] \quad (5)
\]

We refer to the event \(x_n\) falls into the gap between \(\hat{r}^o_n\) and \(r\) as a rare event; and we shall deal with the rare event part first,

\[
nE \left[ (a_1 x_n - \hat{a}_2^o x_n)^2 1_{(x_n \leq r, x > \hat{r}^o_n)} \right] =nE \left[ (a_2 - \hat{a}_2^o)^2 x_n^2 1_{(x_n \leq r, x > \hat{r}^o_n)} \right] \\
+2nE \left[ (a_2 - \hat{a}_2^o) (a_1 - a_2) x_n^2 1_{(x_n \leq r, x > \hat{r}^o_n)} \right] \\
+nE \left[ (a_2 - a_1)^2 x_n^2 1_{(x_n \leq r, x > \hat{r}^o_n)} \right] \equiv (I) + (II) + (III).
\]

By Proposition 1, 2, and \(E|r_\infty| < \infty(see (163))\), we have

\[
\limsup_{n \to \infty} nP(x_n \in B_{\hat{r}^o_n - r}) \leq C, \quad (6)
\]
where \( B(x) = [y+x, y] \cup [y, y+x] \), for \( x, y \in \mathbb{R}^1 \). By Cauchy-Schwartz inequality, Theorem 4, (6), we can show (I), (II) = \( o(1) \). For (III), we define
\[
R_n = nE \left[ (a_2 - a_1)^2 r^2 1_{(x_n \leq r; x > \hat{r}_n^o)} \right] \quad n(a_2 - a_1)^2 r^2 P(x_n \leq r; x > \hat{r}_n^o);
\]
then by Cauchy-Schwartz inequality, Theorem 4, (6) and the moment assumption,
\[
|nE \left[ (a_2 - a_1)^2 x_n^2 1_{(x_n \leq r; x > \hat{r}_n^o)} \right] - R_n| \\
\leq nE \left[ (a_2 - a_1)^2 (x_n - r)(x_n + r) 1_{(x_n \leq r; x > \hat{r}_n^o)} \right] \\
\leq (a_2 - a_1)^2 E^{1/2} r^2 (\hat{r}_n^o - r)^2 E^{1/4} (x_n + r)^4 P^{1/4} (x_n \leq r; x > \hat{r}_n^o) = o(1).
\]
Combining with the analogous result for \( \{ nE \left[ (a_2 x_n - \hat{a}_1^o x_n)^2 1_{(x_n \leq r; x > \hat{r}_n^o)} \right] \} \) and Proposition 1, 2, we have
\[
\lim_{n \to \infty} nE \left[ (a_2 x_n - \hat{a}_1^o x_n)^2 1_{(x_n \leq r; x > \hat{r}_n^o)} \right] + nE \left[ (a_1 x_n - \hat{a}_2^o x_n)^2 1_{(x_n \leq r; x > \hat{r}_n^o)} \right] = \lim_{n \to \infty} (a_2 - a_1)^2 r^2 nP(x_n \in B_{\hat{r}_n^o - r}) = (a_2 - a_1)^2 r^2 \pi_r E|r_\infty|.
\]
From Theorem 5 we have
\[
nE \left[ (a_2 x_n - \hat{a}_1^o x_n)^2 1_{(x_n \leq \hat{r}_n^o, r)} \right] + nE \left[ (a_2 x_n - \hat{a}_2^o x_n)^2 1_{(x_n > \hat{r}_n^o, r)} \right] = 2\sigma^2 + o(1).
\]
By (7), (8) we have finished the proof.

3.2 Extension to SETAR(p)

**Theorem 2.** Assume assumptions 1 to 5,
\[
\lim_{n \to \infty} n \left[ E(\hat{x}_{n+1}^o - x_{n+1})^2 - \sigma^2 \right] = E \left( \left[ (\hat{\phi}_2 - \hat{\phi}_1) \hat{x}_n \right] \right) | x_{n+1-d} = r \pi(r) E|r_\infty| + 2p\sigma^2.
\]
The prediction error is decomposed into the two terms in a similar way to that in SETAR(1). From the prediction point of view, once we over-fit a SETAR(\( p + k \)) model for the data, we end up with simply extra \( 2k\sigma^2 \) prediction error in an asymptotic sense as hoped. Notice how this finding coincides with those in Kunitomo and Yamamoto(1985) and Ing(2003) in a rather surprising way.

**Proof of Theorem 2.** Under the assumptions of Theorem 2, we show for each \( d \), all large \( n \),
\[
P(\hat{d} \neq d) = O(n^{-k})
\]
in supplementary file, where we see that (9) and standard inequalities can be applied to ‘replace’ \( \hat{d} \) with \( d \) in the integrations with threshold order estimator involved; accordingly, we assume \( d \) is known in the order \( p \) extensions. Not surprisingly, let us begin with a decomposition:
\[
n \left[ E(\hat{x}_{n+1}^o - x_{n+1})^2 - \sigma^2 \right] = n \left[ E(\hat{\phi}_2 \hat{x}_n - \hat{\phi}_1 \hat{x}_n)^2 1_{\hat{r}_n^o < x_{n+1-d} \leq r} \right]
\]
To deal with \((I) (II)\), we need moment bounds for normalized \(\hat{\alpha}_a\) and \(\hat{r}_o\) and a control over the probability of rare event. The order \(p\) analogs of Theorem 4 and (6) are stated as follows. For any \(s > 0\), there is some \(C > 0\) such that for all large \(n\), \(k = 1, 2,\)

\[
E \left\| n^{1/2} (\hat{\alpha}_k - \hat{\alpha}_k^o) \right\|^s < C, \quad (10)
\]

\[
E \left| n (r - \hat{r}_o) \right|^s < C, \quad (11)
\]

and

\[
\lim \sup_{n \to \infty} n P (x_{1+n-d} \in B_{\hat{r}_o-r}(r)) < \infty. \quad (12)
\]

(10), (11), (12) can be shown in a way very similar to that in SETAR(1) and hence the proofs are omitted; for more details, see the supplementary file for SETAR(\(p\)). By the moment bounds, standard inequalities, and (12), we have

\[
(I) = n \left[ E (\hat{\alpha}_2' \hat{x}_n - \hat{\alpha}_1' \hat{x}_n)^2 1_{r_{\hat{r}_o} < x_{1+n-d} \leq r} \right] + o(1). \quad (13)
\]

\[
(II) = n \left[ E (\hat{\alpha}_1' \hat{x}_n - \hat{\alpha}_2' \hat{x}_n)^2 1_{r_{\hat{r}_o} < x_{1+n-d} \leq r} \right] + o(1). \quad (14)
\]

By (13), (14), order \(p\) version of the propositions in Section 4.2,

\[
(I) + (II) = E \left\{ \left[ (\hat{\alpha}_2 - \hat{\alpha}_1)' \hat{x}_n \right]^2 | x_{n+1-d} = r \right\} \pi(r) E |r_\infty| + o(1).
\]

Extension of Theorem 5 to order \(p\) is straightforward so for whose proof we refer the reader to its SETAR(1) analog(i.e. Theorem 5); we note that by the previously stated theorem, propositions, and arguments similar to those in the proof of Theorem 5, we have

\[
(III) + (IV) = 2p\sigma^2 + o(1).
\]

The proof of Theorem 2 is finished. \(\square\)

**Remark 1.** If the intercepts \(a_0\)’s are not zeros but unknown constants required to be estimated, then the RHS naturally becomes

\[
E \left\{ \left[ (a_{20}, \hat{a}_2') - (a_{10}, \hat{a}_1') \right] (1, \hat{x}_n')^2 | x_{n+1-d} = r \right\} \pi(r) E |r_\infty| + 2(p + 1)\sigma^2.
\]
3.3 Extention to \(h\)-step Prediction Error

**Theorem 3.** Assume assumptions 1 to 5 and \(h \leq d\),

\[
\lim_{n \to \infty} n \left[ E(\tilde{x}_{n+h}^a - x_{n+h})^2 - \sum_s P(TH(s)) \sum_{j=0}^{h-1} b_j(s) \sigma^2 \right]
\]

\[
= \sum_s \sum_{k=1,2} tr(R(TH(s)) E_s k R_k^{-1} E_s k') \sigma^2
\]

\[
+ \sum_s \sum_{j=1,\ldots,h} E(\tilde{x}_n^a(A(TH(s)) - A(TH(s'))) 1_{TH(s')})^2 | x_{n+j-d} = r \),
\]

where, in the second term on the RHS, \(s^j = \{s_1^j, \ldots, s_h^j\}\), \(s_j^j = \begin{cases} s_i, \text{ if } i \neq j, \\ 3 - s_i, \text{ o.w.} \end{cases}\), \(\gamma_{3s} = \begin{cases} r_\infty > 0, s_j = 2, \\ r_\infty \leq 0, s_j = 1. \end{cases}\), \(R_2 = E(\tilde{x}_n^a 1_{x_{n+1-d}^r})\),

\[
E_{s,j} = \begin{cases} A_{s_1} \times \cdots \times A_{s_{j-1}} e_i^1 A_{s_j+1} \times \cdots \times \hat{a}_{s_h}, 1 < j < h, \\ I_p e_i^1 A_{s_2} \times \cdots \times \hat{a}_{s_h}, j = 1, \\ A_{s_1} \times \cdots \times A_{s_{h-1}}, j = h, \end{cases}
\]

and \(E^k_s = \sum_{s_j = k; 1 \leq j \leq h} E_{s,j}, k = 1, 2\).

Taking \(p \geq d \geq h = 2\) for example, we briefly illustrate the possibly most confusing terms \(E^k_s\) and \(A(TH(s)), A(TH(s'))\). In the first term on the RHS, the summation is summing over a grid of \(s = \{1, 2\}^2\) and \(k = 1, 2\). By definition we have \(E^1_{\{1,1\}} = E_{\{1,1\}, 1} + E_{\{1,1\}, 2}\), where \(E_{\{1,1\}, 1} = I_p e_i^1 \hat{a}_1\), \(E_{\{1,1\}, 2} = A_1\), and \(E^2_{\{1,1\}} = 0 \times I_p\); note that \(E^1_{\{1,1\}}, E^2_{\{1,2\}}(s_1 = s_2 = k)\) have the same structure (looking, to be heuristic) as that in (10) of Ing(2003), whereas the other \(E^k_s\)’s have a rather new structure (\(\hat{a}_1, \hat{a}_2\) both appear in one term) in SETAR process \(h\)-step prediction error. On the other hand, the second term on the RHS is summing over a grid of \(s = \{1, 2\}^2\) and \(j = 1, \ldots, h\). Take \(s = \{1, 2\}, j = 2\) for example we have \(A(TH(s)) = A_1 \hat{a}_2\) and \(A(TH(s')) = A_1 \hat{a}_1\). \(\cdot\) for the other combination of \(s, j\) can be constructed in the same way.

Two conlusions can be drawn from the comparison of Theorem 3 to the analog in Ing(2003) and Theorem 2: still the \(h\)-step prediction error consists of one regular and one rare event; the regular part of the right hand side, the term with tracing, now contains coefficient matrix, \(E^k_s\)’s, and appears to have a similar "looking" to that in Ing(2003). As Ing(2003) shows, when the predictor rolling forward more than one step, coefficients are involved in the regular event part of prediction error as well as the rare event part. Nevertheless, what makes the formula more delicate and notationally awkward than that.
in one step prediction error is the involvement of both the thresholding mechanism in (3) that introducing $TH(.)$’s we have to separately consider, and the fact that the occurrence of rare events now depends on not only $x_{n+1-d}$ but $(x_{n+h-d}, \ldots, x_{n+1-d})$, leading to extra terms to sum over.

Proof of Theorem 3. As explained in the proof of Theorem 2, we assume $d$ is a known parameter. With the sample analog of $TH(s)$ and $A(.)$ and (4), we have

$$n \left[ E(\hat{x}_{n+h}^0 - x_{n+h})^2 - \sum_s P(TH(s)) \sum_{j=0}^{h-1} b_j^2(s) \sigma^2 \right]$$

$$= n \sum_s E \left[ \left( \hat{x}_n(A(TH(s)) - A(TH(s))^1_{TH(s) \cap TH(s)} \right)^2 + n \sum_s \sum_{j=1, \ldots, h} E \left[ \left( \hat{x}_n(A(TH(s)) - A(TH(s)^j)) \right)^1_{TH(s) \cap TH(s)} \right)^2 + g_{1n} \equiv (I) + (II) + g_{1n},$$

where $s_i^j = \{s_i, \text{ if } i \neq j, 3 - s_i, \text{ o.w.}\}$ and $g_{1n}$, defined by (15), is the sum of integrations over very rare events, each of which is of occurrence of at least two mismatches between the paired values $(1_{x_{n+i-d} > r}, 1_{x_{n+i-d} > \hat{r}_n}), i = 1, \ldots, h$.

To deal with $(II)$ and $g_{1n}$, we apply (10), (11) and standards inequalities:

$$(II) + g_{1n} = n \sum_s \sum_{j=1, \ldots, h} E \left[ \left( \hat{x}_n(A(TH(s)) - A(TH(s)^j)) \right)^1_{TH(s) \cap TH(s)} \right)^2 + o(1).$$

By rearranging the above equation, we have $\lim_{n \to \infty} (II) + g_{1n}$ equivalent to

$$\sum_s \sum_{j=1, \ldots, h} \lim_{n \to \infty} E \left[ \left( \hat{x}_n(A(TH(s)) - A(TH(s)^j)) \right)^1_{TH(s) \cap TH(s)} \right)^2 | x_{n+j-d} \in \gamma_{1ns} \times n \times P(x_{n+j-d} \in \gamma_{1ns}),$$

where $\gamma_{1ns} = \{[r, \hat{r}_n], \text{ if } s_j = 2, [\hat{r}_n, r], \text{ if } s_j = 1\}$.

Porposition 1, 2 have order $p$ analog, whose proofs, with $d$ known, are essentially the same as the proofs for Porposition 1, 2; as a result of such the propositions are stated in the supplementary file without proof. By the Proposition 1 analog, we have for all $s \subset \{1, 2\}^h, j = 1, \ldots, h$,

$$n|P(x_{n+j-d} \in \gamma_{1ns}) - P(x_{1}^{in} \in \gamma_{1ns})| = o(1),$$

where $x_{1}^{in}$ is independent of original process and with identical distribution of $x_1$; as for the analog of Porposition 2, we have

$$\lim_{n \to \infty} nP(x_{1}^{in} \in \gamma_{1ns}) = \begin{cases} E(|r_\infty|; r_\infty > 0) \pi(r), s_j = 2, \\
E(|r_\infty|; r_\infty \leq 0) \pi(r), s_j = 1 \end{cases}.$$
To prove (21), note that if we let $\gamma_{3s} = \left\{ \begin{array}{ll} \gamma_{3s} > 0, & s_j = 2, \\ \gamma_{3s} \leq 0, & s_j = 1. \end{array} \right.$

For (I), we define $\tilde{A}_{i,j}, \tilde{a}_{i,j}$ by $\tilde{A}_{i,j} = \begin{cases} A_{s_i} - \bar{A}_{s_i}, & i = j, \\ A_{s_i}, & i \neq j, \end{cases}$ and $\tilde{a}_{i,j} = \begin{cases} \tilde{a}_{s_i} - \hat{a}_{s_i}, & i = j, \\ \tilde{a}_{s_i}, & i \neq j, \end{cases}$ respectively. Similar to Ing(2003), we have (I) equivalent to

$$n \sum_{s} E \left( \sum_{j=1}^{h} \tilde{x}_n \tilde{A}_{1,j} \times \cdots \times \tilde{a}_{h,j} 1_{TH(s)} \right)^2 + n \sum_{s} E(g_{2ns} + g_{3ns}),$$

where $g_{2ns}$ and $g_{3ns}$ are bounded by summations of finite terms; specifically:

$$g_{2ns} \leq C \sum_{l_1+l_2=3}^{2h} \left\| \tilde{\hat{a}}_1 - \tilde{a}_1 \right\| l_1^2 \left\| \tilde{\hat{a}}_2 - \tilde{a}_2 \right\| l_2 \left\| \tilde{x}_n \right\|,$$

$$g_{3ns} \leq C \sum_{l_1+l_2=2}^{2h} \left\| \tilde{\hat{a}}_1 - \tilde{a}_1 \right\| l_1^2 \left\| \tilde{\hat{a}}_2 - \tilde{a}_2 \right\| l_2 \left\| \tilde{x}_n \right\| 1_{TH(s) \cap TH^c(s)}.$$

To deal with the quadratic terms in (20), we have the following somewhat intriguing result for the expected product of normalized $\hat{a}_i$, $i = 1, 2$, coefficients estimators with $r$ known:

$$\lim_{n \to \infty} nE \left( \hat{a}_1 - \hat{a}_1 \right)' \left( \hat{a}_2 - \hat{a}_2 \right) = 0.$$  \hspace{1cm} (21)

To prove (21), note that if we let $Q_n = n^{-1/2} \sum_{i=1}^{n} \tilde{x}_i e_i + 1_{(x_{i+1-d} \leq r)}$, $W_n = n^{-1/2} \sum_{i=1}^{n} \tilde{x}_i e_i + 1_{(x_{i-1+d} > r)}$, then $E(Q_n W_n) = 0$; the same techniques in the proof of Theorem 5, especially the usage of (12), (10) and (11), lead us to (21).

The result of (21) together with almost identical steps in Ing(2003) and some algebraic manipulations leads to

$$nE \left( \sum_{j=1}^{h} \tilde{x}_n \tilde{A}_{1,j} \times \cdots \times \tilde{a}_{h,j} 1_{TH(s)} \right)^2 = \sum_{k=1,2} tr(R(TH(s)) E_s R_k^{-1} E_{s_k}^k) \sigma^2 + o(1),$$

for any fixing $s \in \{1, 2\}^h$, where $R(TH(s)) = E(\tilde{x}_n \tilde{x}_n 1_{TH(s)}); R_1 = E(\tilde{x}_n \tilde{x}_n 1_{x_{n+1-d} \leq r}); R_2 = E(\tilde{x}_n \tilde{x}_n 1_{x_{n-1+d} > r}); E_{s,j} = \begin{cases} A_{s_1} \times \cdots \times A_{s_{j-1}} e_1 A_{s_{j+1}} \times \cdots \times \tilde{a}_{s_h}, & 1 < j < h, \\ I_p e_1 A_{s_2} \times \cdots \times \tilde{a}_{s_h}, & j = 1, \\ A_{s_1} \times \cdots \times A_{s_{h-1}}, & j = h. \end{cases}$
and $E^k_s = \sum_{s_j=k;1 \leq j \leq h} E_{s,j}, k = 1, 2; \ e_1$ is the unit length vector with first element 1.

(20), (22), and (12), (10) and (11) as well as standard inequalities imply

$$\lim_{n \to \infty} (I) = \sum_s \sum_{k=1,2} tr(R(TH(s))E^k_s R_k^{-1} E^{k'}_s) \sigma^2.$$  \hspace{1cm} (23)

Combining (15), (19), (23), we have finished the proof.

3.4 Simulation Results

In this section, we do the Monte Carlo simulation based on Theorem 1 to 3 with $\sigma^2 = 1$. For illustration purpose, we discuss only the procedure for computing the limit of the conditional expectation and $\pi(r)E|_{r=\infty}$ in Theorem 2; $h$-step simulation can be done in the same way. First we note that $\pi(r)$ is the rate embedded in the two compound Poisson processes; the mutiplication of a $\pi(r)$ on $E|_{r=\infty}$ cancels the the rate in $r_\infty$. Therefore we directly consider the compound Poisson processes with rate 1. This is an amazing normalization which dramatically eases the computation loading and increases the accuracy of the outcome. To construct random variables distributed as $\tilde{x}_j^{lim}$, we set a small band, $a_m = 0.01$, and the resampling follows exactly the procedure described in (2): the vector $\tilde{x}_j$ is collected if $j = \inf\{i : x_{i+1-d} \in B_{am}(r), i \geq p\}$; after the collection, the whole process restarts again until we obtain sample of a certain size. With a minor midification, the same steps can be used for computing the limit of the conditional expectation. In general we can never be sure whether the cumulative summations in the Poisson process have reached their minimal, so we do some experiments; 250 jumps was eventually found to be a reasonable number such that the minimum occurred at the point beyound which shall be considered impossible.

The Monte Corlo results are reported as MSPErare*, which representes the rare event part in the decomposition of the AMSPE, and MSPE*, where $2p$ is included. $a_m = 0.001, 0.0001$ have also been examined for these results, no major difference (more than an amount up to 3 decimal) was found. MSPErare is the sample counter part of the rare event part in AMSPE; MSPEar will be the rest of that; MSPEboth is the sum of the first two terms. In the $h$-step prediction cases we report the aggregate results as separately reporting the rare and regular parts becomes too tedious.

Some comments on the simulation results are given as follows. In general as sample size increases, the convergence of the prediction error to its theoretical limit is evident. To better appreciate the result in Theorem 1, in the case of one step prediction we separately present the simulation results of the mismatch(MSPErare) and coefficients estimation(MSPEar; $2p$) prediction errors. This, however, is not so ready to be done in the case of mult-step prediction as we can see from the example given in Theorem 3; the summation is summing over a grid of $s = \{0, 1\}^h$ and $j = 1, \ldots, h$ rather than a single term. Overall the prediction error of $h$-step predictions are lower than 1-step predictions.
<table>
<thead>
<tr>
<th>n=300</th>
<th>500</th>
<th>700</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSPErare</td>
<td>3.771</td>
<td>3.084</td>
<td>3.819</td>
</tr>
<tr>
<td>MSPEar</td>
<td>4.107</td>
<td>4.242</td>
<td>3.967</td>
</tr>
<tr>
<td>MSPEboth</td>
<td>7.878</td>
<td>7.326</td>
<td>7.786</td>
</tr>
<tr>
<td>MSPErare*</td>
<td>3.39</td>
<td>3.39</td>
<td>3.39</td>
</tr>
<tr>
<td>MSPE*</td>
<td>7.39</td>
<td>7.39</td>
<td>7.39</td>
</tr>
</tbody>
</table>

\( (r, d) = (-1, 1), \tilde{a}_1 = (-0.5, -0.2), \tilde{a}_2 = (0.2, 0.15), \mathcal{N}(0, 1) \)

<table>
<thead>
<tr>
<th>n=300</th>
<th>500</th>
<th>700</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSPErare</td>
<td>6.804</td>
<td>5.354</td>
<td>4.395</td>
</tr>
<tr>
<td>MSPEar</td>
<td>6.186</td>
<td>6.107</td>
<td>5.869</td>
</tr>
<tr>
<td>MSPEboth</td>
<td>12.99</td>
<td>11.641</td>
<td>10.264</td>
</tr>
<tr>
<td>MSPErare*</td>
<td>2.987</td>
<td>2.987</td>
<td>2.987</td>
</tr>
<tr>
<td>MSPE*</td>
<td>8.987</td>
<td>8.987</td>
<td>8.987</td>
</tr>
</tbody>
</table>

\( (r, d) = (-1, 1), \tilde{a}_1 = (0.2, -0.5, -0.02), \tilde{a}_2 = (0.2, 0.15, 0.15), \mathcal{N}(0, 1) \)

<table>
<thead>
<tr>
<th>n=300</th>
<th>500</th>
<th>700</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSPErare</td>
<td>3.775</td>
<td>3.283</td>
<td>3.509</td>
</tr>
<tr>
<td>MSPEar</td>
<td>1.941</td>
<td>2.075</td>
<td>2.046</td>
</tr>
<tr>
<td>MSPEboth</td>
<td>5.716</td>
<td>5.358</td>
<td>5.555</td>
</tr>
<tr>
<td>MSPErare*</td>
<td>3.066</td>
<td>3.066</td>
<td>3.066</td>
</tr>
<tr>
<td>MSPE*</td>
<td>5.066</td>
<td>5.066</td>
<td>5.066</td>
</tr>
</tbody>
</table>

\( (r, d) = (-1, 1), \tilde{a}_1 = (0.4), \tilde{a}_2 = (-0.2), \mathcal{N}(0, 1) \)

<table>
<thead>
<tr>
<th>n=300</th>
<th>500</th>
<th>700</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSPErare</td>
<td>4.387</td>
<td>4.176</td>
<td>4.612</td>
</tr>
<tr>
<td>MSPEar</td>
<td>2.085</td>
<td>2.06</td>
<td>2.024</td>
</tr>
<tr>
<td>MSPEboth</td>
<td>6.472</td>
<td>6.236</td>
<td>6.636</td>
</tr>
<tr>
<td>MSPErare*</td>
<td>3.641</td>
<td>3.641</td>
<td>3.641</td>
</tr>
<tr>
<td>MSPE*</td>
<td>5.641</td>
<td>5.641</td>
<td>5.641</td>
</tr>
</tbody>
</table>

\( (r, d) = (-1, 1), \tilde{a}_1 = (0.3), \tilde{a}_2 = (0.3), \mathcal{N}(0, 1) \)

<table>
<thead>
<tr>
<th>n=300</th>
<th>500</th>
<th>700</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSPErare</td>
<td>10.107</td>
<td>5.34</td>
<td>5.935</td>
</tr>
<tr>
<td>MSPEar</td>
<td>6.817</td>
<td>8.023</td>
<td>6.428</td>
</tr>
<tr>
<td>MSPEboth</td>
<td>16.924</td>
<td>13.363</td>
<td>12.363</td>
</tr>
<tr>
<td>MSPErare*</td>
<td>3.815</td>
<td>3.815</td>
<td>3.815</td>
</tr>
</tbody>
</table>

\( (r, d) = (-1, 1), \tilde{a}_1 = (-0.3), \tilde{a}_2 = (0.3), \mathcal{N}(0, 1) \)
<table>
<thead>
<tr>
<th></th>
<th>n=300</th>
<th>500</th>
<th>700</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSPErare</td>
<td>3.415</td>
<td>2.531</td>
<td>3.702</td>
<td>2.962</td>
</tr>
<tr>
<td>MSPEar</td>
<td>4.247</td>
<td>4.307</td>
<td>3.934</td>
<td>4.174</td>
</tr>
<tr>
<td>MSPEboth</td>
<td>7.662</td>
<td>6.838</td>
<td>7.636</td>
<td>7.136</td>
</tr>
<tr>
<td>MSPErare*</td>
<td>3.591</td>
<td>3.591</td>
<td>3.591</td>
<td>3.591</td>
</tr>
<tr>
<td>MSPE*</td>
<td>7.591</td>
<td>7.591</td>
<td>7.591</td>
<td>7.591</td>
</tr>
</tbody>
</table>

\[(r,d) = (-1.5, 2), \hat{a}_1 = (-0.4, -0.2), \hat{a}_2 = (-0.7, 0.2), \text{ t distribution; d.f. = 80}\]

<table>
<thead>
<tr>
<th></th>
<th>n=300</th>
<th>500</th>
<th>700</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSPErare</td>
<td>5.55</td>
<td>3.599</td>
<td>2.595</td>
<td>3.857</td>
</tr>
<tr>
<td>MSPEboth</td>
<td>12.397</td>
<td>10.044</td>
<td>8.861</td>
<td>10.005</td>
</tr>
<tr>
<td>MSPErare*</td>
<td>3.937</td>
<td>3.397</td>
<td>3.397</td>
<td>3.397</td>
</tr>
</tbody>
</table>

\[(r,d) = (0.7, 2), \hat{a}_1 = (0.2, -0.5, 0.2), \hat{a}_2 = (-0.1, 0.1, -0.2), \text{ t distribution; d.f. = 80}\]

<table>
<thead>
<tr>
<th></th>
<th>n=300</th>
<th>500</th>
<th>700</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSPEboth</td>
<td>4.933</td>
<td>4.686</td>
<td>4.82</td>
<td>5.235</td>
</tr>
<tr>
<td>MSPE*</td>
<td>4.774</td>
<td>4.774</td>
<td>4.774</td>
<td>4.774</td>
</tr>
</tbody>
</table>

\[(r,d,h) = (-1.2, 2), \hat{a}_1 = (-0.3, -0.2), \hat{a}_2 = (0.3, 0.2), \mathcal{N}(0,1)\]

<table>
<thead>
<tr>
<th></th>
<th>n=300</th>
<th>500</th>
<th>700</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSPEboth</td>
<td>4.934</td>
<td>4.602</td>
<td>4.715</td>
<td>4.225</td>
</tr>
<tr>
<td>MSPE*</td>
<td>4.149</td>
<td>4.149</td>
<td>4.149</td>
<td>4.149</td>
</tr>
</tbody>
</table>

\[(r,d,h) = (-1.3, 3), \hat{a}_1 = (-0.3, -0.2, 0.12), \hat{a}_2 = (0.3, 0.2, -0.3), \mathcal{N}(0,1)\]

<table>
<thead>
<tr>
<th></th>
<th>n=300</th>
<th>500</th>
<th>700</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSPEboth</td>
<td>4.733</td>
<td>4.536</td>
<td>4.552</td>
<td>4.171</td>
</tr>
<tr>
<td>MSPE*</td>
<td>4.183</td>
<td>4.183</td>
<td>4.183</td>
<td>4.183</td>
</tr>
</tbody>
</table>

\[(r,d,h) = (-1.3, 3), \hat{a}_1 = (-0.3, -0.2, 0.12), \hat{a}_2 = (0.3, 0.2, -0.3), \text{ t distribution; d.f. = 80}\]

<table>
<thead>
<tr>
<th></th>
<th>n=300</th>
<th>500</th>
<th>700</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSPEboth</td>
<td>5.985</td>
<td>5.678</td>
<td>5.396</td>
<td>5.299</td>
</tr>
<tr>
<td>MSPE*</td>
<td>4.836</td>
<td>4.836</td>
<td>4.836</td>
<td>4.836</td>
</tr>
</tbody>
</table>

\[(r,d,h) = (1, 3, 3), \hat{a}_1 = (-0.5, 0.1, -0.07), \hat{a}_2 = (0.1, 0.5, -0.2), \text{ t distribution; d.f. = 80}\]

15
due to our model coefficients specification (see assumption 1). Finally, assuming moment assumptions such as assumption 5 facilitates our mathematical derivation, yet we see the equally nice performance in the cases even with \( t \) distribution of degree of freedom 80; this suggests our moment assumption might be too strong and our results are expected to apply to a more general case.

4 Statements Required for SETAR Process Analysis

The analysis of SETAR process is special in the sense that the analysis of SETAR(1) captures most the spirit of the central ideas. As the presentation of SETAR(\( p \)) tends to use more notation which might deteriorate the readability of the statements, we present our results under order 1 in this section. The other results as well as more details about the difference between the analysis of order 1 and \( p \) can be found in the supplementary file.

4.1 Moment Bounds for the Normalized Estimators

**Theorem 4.** Given \( q > 0 \). Assume assumption 1, 2, 5. Then we have some \( C > 0 \) such that for all large \( n, k = 1, 2 \),
\[
E \left| n^{1/2}(a_k - \hat{a}_k^o) \right|^q < C. \tag{24}
\]

Given \( q > 0 \). Assume assumption 1, 2, 3, 5. Then there exists some \( C > 0 \) such that for all large \( n, \)
\[
E \left| n(r - \hat{r}_n^o) \right|^q < C, \tag{25}
\]

4.2 Asymptotic Probabilities Relating to the State Prediction Error

To deal with \( nE \left[ (a_1x_n - \hat{a}_2^o x_n)^2 1_{(x_n \leq r; x > \hat{r}_n^o)} \right] + nE \left[ (a_2x_n - \hat{a}_1^o x_n)^2 1_{(x_n > r; x \leq \hat{r}_n^o)} \right] \), we need not only the moment bounds in Section 4.1 but also a rigorous analysis on the wrong state of \( x_{n+1} \) prediction, where the values comparison of \( x_n \) and \( r \) disagrees to that of \( x_n \) and \( \hat{r}_n^o \). The main idea of such an analysis is explained as follows. We first use the moment bounds in Theorem 4 and standard inequalities to extract \( nP(x_n \in B_{\hat{r}_n^o - r}(r)) \) from the integration. Proposition 1 states that \( x_n \) will be asymptotically independent of \( \hat{r}_n^o \). Proposition 2 states an intriguing result with intuition if we observe the probability of \( \{x_n \in B_{\hat{r}_n^o - r}(r)\} \) counts only when \( x_n \) falls around a ball center at \( r \) which is decreasing to singleton in probability due to \( (\hat{r}_n^o - r) \to 0 \); by this observation, the appearance of \( \pi(r) \) is understandable; \( E|x_\omega| \) is not unfamiliar if we notice \( r_\omega \) is the limiting distribution and that \( x_\omega \) is an independent random variable. We defer most of the proof to the online supplementary file.
**Proposition 1.** Assume assumption 1, 2, 3, 5. Let \( x_o \) denote an independent random variable with distribution as that of \( x_1 \)’s.

\[
|nP(x_n \in B_{r_n - r}(r)) - nP(x_o \in B_{r_n - r}(r))| = o(1).
\]

**Proposition 2.** Assume assumption 1 to 5,

\[
|nP(x_o \in B_{r_n - r}(r)) - \pi_r E|r_{\infty}| = o(1).
\]

### 4.3 The Variance of Coefficients Estimation in the Prediction Error

**Theorem 5.** Assume assumption 1, 2, 3, 5. Then

\[
nE \left[ (a_1 x_n - \hat{a}_0^n x_n)^2 1_{(x_n \leq r_n \land r)} \right] + nE \left[ (a_2 x_n - \hat{a}_2^n x_n)^2 1_{(x_n > r_n \lor r)} \right] = 2\sigma^2 + o(1).
\]

This result is rather typical, and the proof can be found in the online supplementary file.

### 5 Appendix

**Assumptions and Notation**

**Assumptions**

**Assumption 1.** Let \( e_t \) in model (1) be identically distributed independent random variables with \( E(e_t) = 0, E|e_t|^2 = \sigma^2 \). We also assume that the p.d.f. of \( e_t, f_e \), is uniformly continuous and bounded away from zero for any given compact set in \( \mathbb{R}^1 \) and sup \( f_e \leq M \). In addition, we require the coefficients to satisfy \( \max \left\{ \sum_{j=1}^{p} |a_{ij}|, \sum_{j=1}^{p} |a_{2j}| \right\} < 1 \) and there exist \( \tilde{z}_p \in \mathbb{R}^p \) with \( z_{p+1-d} = r \) such that \( ((a_{10}, \tilde{a}_1') - (a_{20}, \tilde{a}_2'))(1, \tilde{z}_p)' \neq 0 \). Finally, we assume the intercepts \( a_{00} \)'s are zeros.

**Assumption 2.** \( f_e \) is Holder continuous, i.e., there exist \( \theta_1, \theta_2 > 0 \) such that \( |f_e(x) - f_e(y)| \leq \theta_1 |x - y|^{\theta_2} \), for all \( x, y \in \mathbb{R}^1 \).

**Assumption 3.** There exist \( \delta_1, \delta_2, M_0 > 0 \) such that for all \( M \geq M_0, |y| \leq M \),

\[
\begin{align*}
P(y - M^{-\delta_1} \leq e \leq y) &\leq M^{-\delta_2} P(y \leq e), \\
P(y \leq e \leq y + M^{-\delta_1}) &\leq M^{-\delta_2} P(e \leq y).
\end{align*}
\]

**Assumption 4.** There exist \( \delta_3, \delta_4, M_0 > 0 \) such that for \( M \geq M_0, |z| < M \),

\[
\begin{align*}
\sup_{s \in B_{M^{-\delta_3}}(0)} f_e(z + s) &\leq \inf_{s \in B_{M^{-\delta_4}}(0)} f_e(z + s)(1 + M^{-\delta_3}), \\
\sup_{s \in B_{M^{-\delta_3}}(0)} f_e(z + s)(1 - M^{-\delta_3}) &\leq \inf_{s \in B_{M^{-\delta_4}}(0)} f_e(z + s).
\end{align*}
\]

**Assumption 5.** \( E \exp \left( w|e_t| \right) < \infty \) for some \( w > 0 \).
Tong and Chan (1985), Chan (1989) show that under assumption 1, the SETAR process in (1) has geometrically ergodic stationary distribution (see (43)) and we denote its joint stationary p.d.f. as \( \pi \) and marginal stationary distribution \( \pi \) (for SETAR(1), \( \pi \) and \( \pi \) coincide). Tweedie (1983), Theorem 1 then prove that under the same assumption and given \( E|e_1|^r < \infty, E\exp(|e_1|s) < \infty \) for some \( r, s > 0 \), we have \( E|x_1|^r < \infty, E\exp(|x_1|s) < \infty \), respectively. Assumption 1, 2 also give us an uniformly continuous and bounded away from zero \( \pi \) on any given compact set. In fact, assumption 2 is assumed for the sole purpose of having an uniformly continuous \( \pi \). Define function \( T : \mathbb{R}^p \to \mathbb{R}^1 \) such that

\[
T(\tilde{z}_p) = \begin{cases} 
\tilde{a}_1' \tilde{z}_p & \text{if } z_{p+1-d} \leq r, \\
\tilde{a}_2' \tilde{z}_p & \text{if o.w.}
\end{cases}
\]

By the definition of a stationary density function and its marginal density function, we have for \( z \in \mathbb{R}^1 \),

\[
\pi(z) = \int_{\tilde{y} \in \mathbb{R}^p} f_{\tilde{e}}(z - T(\tilde{y}))\tilde{\pi}(\tilde{y})d\tilde{y}.
\]

We see that by this invariat equation, \( \pi \) is bounded away from zero on any given comapct set if \( f_{\tilde{e}} \) is. For continuity, see Lemma 3, whose proof will be given in supplementary file when required notation is introduced. To the author’s knowledge, in order to reach the conclusion of Lemma 3, assumption 2 might not be needed; for this, see Chan (1992)’s comment on Condition 2 therein.

**Lemma 3.** Assume assumption 1, 2. \( \pi \) is uniformly continuous and bounded away from zero on any given compact set of \( \mathbb{R}^1 \).

• Assumption 3 and 4 are technical requirements. The usual smooth condition for p.d.f. is Holder continuity (Ing and Wei (2003)). Unlike Holder continuity, assumptions 4 control the smoothness of \( f_{\tilde{e}}(z) \) over the region \( |z| \leq M \). To have the condition stated in assumption 4, it suffices to have one smooth condition (Holder continuity suffices). Notice that this is not the reason we assume assumption 2; see the first comment) and one boundness against zero condition for \( |z| \leq M \). Less intuitively, assumption 3 requires the tailed probability density to be fat enough so we can bound the probability. The primary usage of this assumption is to extend an event such that a closed form expression of the probability of the extended event can be obtained without much probability loss.

Non-standard as these assumptions might be, they can be easily verified. Let \( e \sim N(0, 1) \). Let any \( \delta_1, \delta_2 \) such that \( 0 < \delta_2 < \delta_1 - 1 \) and \( \delta_2' \) such that \( \delta_2 < \delta_2' \leq \delta_1 - 1 \).
be given. For all sufficiently large $M$,

\[
\sup_{|y| \leq M} \frac{P(y \leq e \leq y + M^{-\delta_1})}{P(y \leq e)} \leq \sup_{0 < y \leq M} \frac{M^{-\delta_1} \exp(-y^2)}{M^{-\delta_1 + \delta_2} \exp(-y + M^{-\delta_1 + \delta_2})^2} \leq \frac{M^{-\delta_1} \exp(-M^2)}{M^{-\delta_1 + \delta_2} \exp(-(M + M^{-\delta_1 + \delta_2})^2)} \leq M^{-\delta_2};
\]

on the other hand, given any $\delta_3 - 1 > \delta_4 > 0$, we have for all sufficiently large $M$,

\[
\sup_{|z| \leq M : x, y \in B_{M^{-\delta_3}}(0)} \frac{f_e(z + x)}{f_e(z + y)} = \sup_{|z| \leq M : x, y \in B_{M^{-\delta_3}}(0)} \frac{\exp(-(z + x)^2)}{\exp(-(z + y)^2)} \leq \exp(-4M^{1-\delta_3}) < 1 + M^{-\delta_4}.
\]

The other side of inequalities can be shown in the same way.

We reparametrize assumption 3 and 4 so that based on the new set of parameters, the trade-off between smoothness and required moment bound on the innovative terms can be addressed. Let $n$ be the sample size. Assumption 3 implies there exists $\alpha > 0, \frac{1}{2} > s_2 > s_1 > 0$, such that for $|y| \leq n^{s_1}$, all large $n$,

\[
P(y - n^{-s_2} \leq e \leq y) \leq n^{-\alpha} P(y \leq e); \quad P(y \leq e \leq y + n^{-s_2}) \leq n^{-\alpha} P(e \leq y);
\]

and assumption 4 implies there exist $\frac{1}{2} > \nu > 0, 1 > \eta, \gamma > 0$ such that for $n^{-\eta}, |z| < n^\nu$, all large $n$,

\[
\sup_{s \in B_{n^{-\eta}}(0)} f_e(z + s) \leq \inf_{s \in B_{n^{-\eta}}(0)} f_e(z + s)(1 + n^{-\gamma}), \quad \sup_{s \in B_{n^{-\eta}}(0)} f_e(z + s)(1 - n^{-\gamma}) \leq \inf_{s \in B_{n^{-\eta}}(0)} f_e(z + s).
\]

In addition, we introduce Lemma 4, which is a direct result of assumption 1 and 4. The techniques used in the proof of the following lemma are constantly used throughout the paper (proofs in the supplementary file, especially).

**Lemma 4.** Assume assumption 1 and 4 and $E|e_1|^s < \infty$ for sufficient large $s$, then there exist $\beta < \gamma, 1 > b' \geq \eta$ such that for all large $n$, all $b \geq b'$,

\[
\sup_{s \in B_{n^{-b}}(0)} \pi(r + s) \leq (1 + n^{-\beta}) \inf_{s \in B_{n^{-b}}(0)} \pi(r + s), \quad \inf_{s \in B_{n^{-b}}(0)} \pi(r + s) \geq (1 - n^{-\beta}) \sup_{s \in B_{n^{-b}}(0)} \pi(r + s).
\]

**Proof of Lemma 4.** By the SETAR model’s coefficients assumption we have $|T(\tilde{z})| \leq \sqrt{p} \|\tilde{z}\|$; therefore, given arbitrarily small $\varepsilon > 0$,

\[
\sup_{\tilde{z} \in [-n^{\nu - \varepsilon}, n^{\nu - \varepsilon}]} |r + s - T(\tilde{z})| \leq n^\nu
\]

for all large $n$. The moment bound on $x_1$, continuity of $\pi(\cdot)$, and $\pi(r) > 0$ (see Lemma 3), and assumption 4 show for any $\gamma > \beta' > \beta$, $pP(|x_1| \geq n^{\nu - \varepsilon}) \leq (n^{-\beta} - n^{-\beta'}) \inf_{s \in B_{n^{-b'}}(0)} \pi(r + s)$.
s) for all large $n$. Combining this result, the fact that $\pi$ is stationary distribution’s p.d.f., assumption 4, and some algebraic manipulation we have

\[
(1+n^{-\beta}) \inf_{s \in B_{n-\beta'}(0)} \pi(r+s) \geq \int_{z_0 \in [-n^{-\epsilon}, n^{-\epsilon}]} (1+n^{-\beta'}) \inf_{s \in B_{n-\beta'}(0)} f_c(r+s - T(z_0)) \tilde{\pi}(z_0) dz_0 + pP(|x_1| \geq n^{\nu-\epsilon})
\]

\[
\geq \sup_{s \in B_{n-\beta'}(0)} \pi(r+s) \geq \sup_{s \in B_{n-\beta'}(0)} \pi(r+s)
\]

The other side of the inequality can be argued in the same way. ■

To discuss the trade-off between moment bound condition and smooth conditions, we define some more notation. Throughout the paper, let $0 < \alpha < \bar{\alpha}$. Define the sets

\[
\mathcal{DE}_{\infty}^1 = \left\{ x : \left( \frac{1}{50} \land \frac{\alpha}{12} \land \frac{1}{2} - \frac{s_2}{25} \right) > x > 0 \right\};
\]

\[
\mathcal{DE}_{\infty}^2 = \left\{ x : b + \beta > 1, \left( \frac{1}{50} \land \frac{\alpha}{12} \land \frac{1}{2} - \frac{s_2}{25} \land \frac{1 - \max\{b, \eta\}}{6} \right) > x > 0 \right\}.
\]

For SETAR($p$), a non-empty $\mathcal{DE}_{\infty}^2$ is a sufficient condition for required smooth condition of our main results; assumption 3 and 4 in turn are sufficient for a non-empty $\mathcal{DE}_{\infty}^1$, whereas assumption 3 implies a non-empty $\mathcal{DE}_{\infty}^1$. In general the more restricted $\mathcal{DE}_{\infty}^1, \mathcal{DE}_{\infty}^2$ are, the higher $q$ such that $E|e_1|^q < \infty$ we need. Since we assume assumption 5, assumption on moment bound of polynomial order is meaningless; therefore assumptions 3 to 5 are sufficient for required moment and smoothness conditions of our main results. However, most of our results, especially those for SETAR(1), do not need assumption 5; such assumption is mainly assumed when only brutal force arguments are available. Still assumption 5 is assumed throughout the paper to keep statements of propositions, lemmas, and theorems from becoming overly tedious.

**Notation**

In this section we define and explain a few more required notation and convention. $\bar{M} = \sup_{x \in \mathbb{R}^d} f_c(x)$, which will be a bounded number due to assumption 1. Let $F_{n(r_n-r)}(x)$, $F_{\infty}(x)$, $A$ denote the empirical distribution, the limiting distributions and $\max\{|a_1|, |a_2|\}$ respectively. For a vector-valued $x$, $\|x\|$ means the Euclidean distant from zero; for matrix $A$, $\|A\|^2 = \sup_{\|x\|} x' A' A x$. We also adapt the following three conventions:

1. We constantly use the subset arg min in many different contexts; we write $j = \arg\min i \ldots$ and $j$ will be the only element in the subset whenever with probability zero the number of the elements contained in the subset is more or less than one.
Moreover, comparison of a random subset $A$ with a scalar $x, y \in \mathbb{R}$, \{x $\geq A \geq y$\}, means the event \{x $\geq \sup A \cap \{\inf A \geq y\}$\.

2. Given a random variable $x$ and event $A, B$, $E(x; A; B)$ means taking expectation of $x$ on the subset $A, B$, and the same rule applies to $P(.)$.

3. Without potential confusion, the deterministic constants, $c, C, \alpha, \delta, \ldots$ will be stated without specific derivation. We note that a rigour derivation of these constants is possible and we do so for the sake of the simple context.

An analytically useful version of threshold estimator is

$$\tilde{r}_n = \begin{cases} \arg\min_{\beta_3, (\tilde{\beta}_1, \tilde{\beta}_2, \beta_3, \beta_4) \in \mathcal{T}^o} L(\tilde{\beta}_1, \tilde{\beta}_2, \beta_3, \beta_4), & \text{on } \mathcal{RE}^c_n, \\ 0, & \text{o.w.} \end{cases}$$

which only differs from $\hat{r}_n$ on $\mathcal{RE}_n$ (i.e. \{\tilde{r}_n = 0\} = $\mathcal{RE}_n$); the total number of $x_j, j = 1, \ldots, n$ falling inside $\theta \subset \mathbb{R}$' is denoted as $q^\theta = |\{k : x_k \in \theta, 1 \leq k \leq n\}|$. Then we define the subset, $\Pi^r$, which includes all integers lesser than the number of samples falling on the right hand side of $r$, and greater than the negative number of those on the left:

$$q^\theta_+ = q^{\theta, \infty}, q^\theta_- = q^{\theta, \infty}$$

$$\Pi^r = \{i : q^\theta_+ \leq i \leq 0, 0 < i \leq q^\theta_+, i \in \mathbb{Z}, \text{ the integer sets }\};$$

we also define the random indexes which focus on the relative position on the real line of each sample to $r$, rather than the time series index:

$$S^r_k = \begin{cases} i, x_i = x_{(k+q^\theta_-)}, k \in \Pi^r, \\ \infty, & \text{o.w.}, \end{cases}$$

where $x_{(i)}$ is the $i$-th ordered statistics of $x, i = 1, \ldots, n$.

We can define $x_\infty$ as anything, since throughout this paper we will not have any direct manipulation on $\{S^r_i = \infty\}$, and without potential confusion, we will drop off the superscript of $S^r_i$ to keep the context simple. Three subset of real line relating to threshold parameter are $\Theta, \Theta'$, and $\Theta''$, whose relative cardinalities are in the same order. The true parameter $r$ is assumed to be in $\Theta$; our estimation scheme fits the threshold only to samples lie on $\Theta'$; and we establish our analysis on $\Theta''$, so we can be sure both the threshold estimator and the true value are contained; the reader is encouraged to think of the three predetermined spaces as $[-M_1, M_1], [-M_2, M_2], [-M_3, M_3]$ with arbitrary $M_i$ such that $M_3 - 1 > M_2 \geq M_1$ greater than a given large-enough constant (100, for example, in our simulation cases the estimator never lies outside this range).

In the introduction, we have defined the compound Poisson processes in order to state our main result. Here, instead of focusing on the minimum position of the processes, we
also study the jump which minimizes the cumulative sum of the previous jumps in the processes. We define the jump of the processes with explicit indexes \( i \):

\[
\kappa_{1,i} = \{r^2(a_1 - a_2)^2 + re_{1,i}(a_1 - a_2)\}, \quad \kappa_{2,i} = \{r^2(a_2 - a_1)^2 + re_{2,i}(a_2 - a_1)\},
\]

\( i = 1, \ldots, \) and \( e_{k,i}\sim\mathcal{E}_1 \), for all \( i, k = 1, 2 \); the positions of the jump points, the discrete points of the two Poisson process with rate \( \pi(r) \), are denoted by \( p_{1,i}, p_{2,i}, i = 1, \ldots \) respectively.

\[
\begin{align*}
  j_1 &= \arg\min_{j \geq 1} \sum_{i=1}^{j} \kappa_{1,i}, \\
  j_2 &= \arg\min_{j \geq 1} \sum_{i=1}^{j} \kappa_{2,i}, \\
  j_3 &= \begin{cases} 
    -j_1, & \sum_{i=1}^{j_1} \kappa_{1,i} < \sum_{i=1}^{j_2} \kappa_{2,i}, \\
    j_2, & \sum_{i=1}^{j_2} \kappa_{2,i} < \sum_{i=1}^{j_1} \kappa_{1,i}.
  \end{cases}
\end{align*}
\]

(28)

By construction, for \( i > 0 \),

\[
P(r_\infty = p_{2,i}) = P(j_3 = i),
\]

and for \( i \leq 0 \),

\[
P(r_\infty = p_{1,-i+1}) = P(j_3 = i).
\]

5.1 Propositions and Theorems Required for the Proof of the Main Results

Statements in this section are important propositions for other theorems; their proof are deferred to supplementary file.

5.1.1 Consistency and the Convergent Rate of the Threshold Estimator \( \hat{r}_n \)

As we are analyzing the SETAR process, the limiting property of \( \hat{r}_n \) is the essential knowledge which we can learn from the study of \( \hat{r}_n \) instead, provided the probability of
$U_n^c$ is properly controlled. In fact, we will see in (29) that to control the probability upper bound of $U_n^c$, we need Proposition 5.

**Proposition 5.** Let $\hat{\theta} > \Delta > 0$ be given; assume assumptions 1, 2, and $E|e_1|^q < \infty$ for $q$ sufficiently large. There exist $c > 0, C > 0$ such that for all large $n$,

$$P(|\tilde{r}_n - r| \geq cn^{-(1-6\Delta)}) \leq C(\log n)n^{-\Delta}\left(\frac{c}{\epsilon}\right)^{+1}.$$ 

One application of Proposition 5 is the construction of the moment bound for $|n(\tilde{r} - r)|^q$,

$$E|n(\tilde{r}_n - r)|^q = \sum_{1 - n \leq i \leq n} E\left[|n(x_{S_i}' - r)|^q; x_{S_i}' = \tilde{r}_n \neq 0\right] + (n|r|)^qP(\tilde{r}_n = 0),$$

where we split the summation into cases $cn^{6\Delta} \leq i \leq n$, $1 - n \leq i \leq 1 - cn^{6\Delta}$ and the other; we should leave the other case alone for now. Intuitively, as the sample size increase, $x_{S_i}'$s will keep a certain distance away from $r$ with high probability, and hence Proposition 5 says these sample points are unlikely to be threshold estimator; the probability of this cases decrease in an order fast enough so we can use a sup-bound argument to get the moment bounds. A formal argument for the previous intuition can be found in (34) in the supplementray file, where corollary 11 and 6, lemma 12 are used for the argument that $x_{S_i}', cn^{6\Delta} \leq i \leq n$, $1 - n \leq i \leq 1 - cn^{6\Delta}$ would not get too close to $r$. We will constantly use this technique and also note that Proposition 7 is stated in a sample point fashion rather than the location fashion as in Proposition 5.

**Corollary 6.** Assume the assumptions for Proposition 5. Given $c_1, c_2, \varepsilon > 0$ and let

$$\mathcal{A}_1 = \{\text{More than } c_1n \text{ } x_i \text{'s } \in (\infty, r - 1]\cap B_\varepsilon(0)^c\},$$

$$\mathcal{A}_2 = \{\text{More than } c_2n \text{ } x_i \text{'s } \in [r + 1, \infty)\cap B_\varepsilon(0)^c\}.$$ 

There exists $s, c, \delta > 0, \Delta > 0, \Delta' > 0$ such that for all large $n$,

$$P(U_n^c) \leq P(|\tilde{r}_n - r| \leq n^{-(1-6\Delta)}; \mathcal{A}_1^c \cup \mathcal{A}_2^c) + P(|\tilde{r}_n - r| \geq n^{-(1-6\Delta)}) \leq C(\log n)n^{-\Delta}\left(\frac{c}{\epsilon}\right)^{+1};$$

$$P(\mathcal{R}E_n) \leq \exp\left(-\Delta'n\right).$$

**Proof of Corollary 6.** By Proposition 5, Corollary 11 and proper choice of the constants $c_1, c_2, \varepsilon > 0$, we have Corollary 6. On the other hand, (30) is just a version of Corollary 11; we state it here so we can have a compact record of both (29) and (30).

5.1.2 Asymptotic Distribution of $\tilde{r}_n$ Nearing $r$

**Proposition 7.** Assume assumptions 1, 2, 3, 5; given $\Delta \in \mathbb{D}C^1_\infty$. There exists $C, c > 0$ such that for all $1 - cn^{6\Delta} \leq j \leq cn^{6\Delta}$, $\beta \subset \mathbb{R}^1 \equiv [-\infty, \infty]$, all large $n$,

$$|P(\tilde{r}_n = x_{S_j}, n(x_{S_j} - r) \in \beta, j \in \Pi) - P(j_3 = j)P(n(x_{S_j} - r) \in \beta, j \in \Pi)|$$
\[ \leq C(\log n)^2 n^{-\alpha(1-18\Delta)}, \]

where \( S_j = S'_j, \Pi = \Pi' \).

Three comments are given. First, notice that \( \{1 - cn^{6\Delta}, cn^{6\Delta}\} \subset \Pi^c \) is a trivial set; take \( \beta = \mathbb{R}^1 \) we see \( P(\tilde{r}_n = x_{S_j}, j \in \Pi) \approx P(j = j) \), and we restrict the integration to the event \( \{j \in \Pi\} \) to maintain a well-defined \( S_j \). Second, \( P(j = j) \) is geometrically decayed in \( j \). Lastly, in Proposition 23 we use Proposition 7 to establish the convergent rate of distribution \( n(\tilde{r}_n - r) \) to its limiting distribution \( r_\infty \); to that end, we need to control \( n(x_{S_j} - r) \in \beta \) on \( \tilde{r}_n = x_{S_j} \).

### 5.2 Proof of Theorem 4

We use the notation \( n\Delta = cn^{6\Delta}, n_{s_1} = n\Delta n^{-\Delta} \) in the proof of Theorem 4.

**Proof of (24).** Let \( \Delta > 0 \) be small enough; \( E|e|^{s} < \infty \), for sufficiently large \( s \) (the requirement of Proposition 5). Given the threshold estimator \( \tilde{r}_n \), define the subsets \( \tilde{J}_1 = \{i : x_i \leq \tilde{r}_n, i = 1, \ldots, n-1\} \), \( \tilde{J}_2 = \{1, \ldots, n-1\}\setminus\tilde{J}_1 \); and we also have these true threshold analog of \( \tilde{J}_i \); \( \tilde{J}_{i,0} = \{i : x_i \leq r, i = 1, \ldots, n-1\} \), \( \tilde{J}_{2,0} = \{1, \ldots, n-1\}\setminus\tilde{J}_{1,0} \). Based on the \( \tilde{J}_k \) and the definition of \( \tilde{a}_k \), we have

\[
a_1 - \tilde{a}_1 = \left( \sum_{i \in \tilde{J}_1} x_i^2 \right)^{-1} \cdot \left( -\sum_{i \in \tilde{J}_1} x_i e_i + \sum_{i \in \tilde{J}_1 \setminus \tilde{J}_{1,0}} x_i^2 (a_1 - a_2) \right), \text{ on } \{\tilde{r}_n \neq 0\} \cap U_n,
\]

the same definition applies to \( i = 2 \). We need (31) to (34), whose proof will be mostly deferred to the succeeding paragraph, to finish the proof. Assume \( E|e|^{s} < \infty \) for sufficiently large \( s \), then \( q > 0 \) (for the cases \( 1 > q > 0 \) we use Jensen’s inequality),

\[
E \left| \frac{\sum_{i \in (\tilde{J}_{1,0})} x_i^2}{n^{1/2}} \right|^q \leq C; \tag{31}
\]

\[
E \left| \frac{\sum_{i \in \tilde{J}_1} x_i e_i}{n^{1/2}} \right|^q \leq C; \tag{32}
\]

\[
E \left( \left| n^{-1} \sum_{i \in \tilde{J}_1} x_i^2 \right|^{-q} 1_{U_n} \right) \leq C. \tag{33}
\]

By Proposition 5, Lemma 12, there exist \( c_1 > c'_1 > 0, c, \delta > 0 \) such that for \( n \geq j > n\Delta, 1-n \leq j < 1-n\Delta, \) all large \( n \),

\[
P(\tilde{r}_n = x_{S_j}, j \in \Pi') \leq P(|\tilde{r}_n - r| \geq c_1 n^{-(1-6\Delta)} + P(q^{(r-c'_1 n^{-(1-6\Delta)})-r} + c_1 n^{-(1-6\Delta)} \geq c_1 n^{6\Delta}) \leq C(\log n) n^{-\delta} (\alpha + 1).
\]

\[
(34)
\]
By standard inequalities, which includes Cauchy-Swartz, Jensen’s, Minkowski’s inequalities, the definition of $\hat{a}_q^n$ as well as that of $\tilde{J}_1$ and $U_n$, and (31) to (34), for all large $n$, $q \geq 1$ (for Minkowski’s inequality),

$$E(n^{1/2} |a_1 - \hat{a}_1^n|)^q$$

$$\leq E \left( \left( \sum_{i \in \tilde{J}_1} x_i^2 \right)^{1/2} \right)^{-1} \left( \sum_{i \in \tilde{J}_1} x_i \varepsilon_{i+1} + n^{-1/2} \sum_{i \in \tilde{J}_1 \setminus J_{1,0}} x_i^2 (a_1 - a_2) \right)^q 1_{U_n} + C$$

$$\leq E^{1/2} \left( \left( \frac{\sum_{i \in \tilde{J}_1} x_i^2}{n} \right)^{-q/2} 1_{U_n} \right)^q \left[ E^{2q} \left( \sum_{i \in \tilde{J}_1} x_i \varepsilon_{i+1} \right)^{q/2} + E^{2q} \left( \sum_{i \in \tilde{J}_1 \setminus J_{1,0}} x_i^2 (a_1 - a_2) \right)^{q/2} \right]^q + C$$

$$< C.$$

The case $0 < q < 1$ can be proven by Jensen’s inequality. \[\square\]

It remains to prove (31) to (33).

**Proof of (31).** Given $\Delta > 0$ small enough, $E|e_1|^s < \infty$ for $s$ large enough. Then for $q \geq 1$, all large $n$,

$$E \left| \sum_{i \in (\tilde{J}_1 \setminus J_{1,0}) \cup (J_{1,0} \setminus \tilde{J}_1)} x_i^2 \right|^q/n^{1/2}$$

$$\leq E \left( \sum_{1-n \leq \kappa \leq n} \left| \frac{Q_{k,1}}{n^{1/2}} \right|^q ; x_{S_{\kappa}} = \tilde{r}_n \neq 0 \right) + E \left( \sum_{i=1}^n x_i^2 ; \tilde{r}_n = 0 \right)$$

$$\leq E \left( \sum_{1-n \Delta \leq \kappa \leq n \Delta} \left| \frac{Q_{k,1}}{n^{1/2}} \right|^q ; x_{S_{\kappa}} = \tilde{r}_n \neq 0 \right) + E \left( \sum_{n \Delta < \kappa \leq n \Delta} \left| \frac{Q_{k,1}}{n^{1/2}} \right|^q ; x_{S_{\kappa}} = \tilde{r}_n \neq 0 \right)$$

$$+ E \left( \sum_{i=1}^n x_i^2 ; \tilde{r}_n = 0 \right) \leq C,$$

where

$$Q_{k,1} = \begin{cases} \sum_{i=1}^k x_i^2, & k > 0, \\ \sum_{i=0}^{|k|-1} x_i^2, & k \leq 0, \end{cases}$$

by standard inequalities, moment bound assumption, (77), (34), a version of Lemma 8. \[\square\]

To prove (32), we first need (35): For each $q \geq 1$, sufficiently large $s$ such that $E|e_1|^s < \infty$,

$$E \left| \sum_{i \in (\tilde{J}_1 \setminus J_{1,0}) \cup (J_{1,0} \setminus \tilde{J}_1)} x_i \varepsilon_{i+1} \right|^q \leq C. \quad (35)$$
Proof of (35). By standard inequalities, (34), Lemma 16, a version of Lemma 8, and sufficiently large $s$, $\Delta$ small enough, we have for all large $n$,
\[
E \left| \sum_{i \in (J_1 \setminus J_{1,0}) \cup (J_{1,0} \setminus J_1)} x_i \varepsilon_i x_i \varepsilon_{i+1} \right|^q \leq E \left( \sum_{1 - n_{\Delta} \leq k \leq n_{\Delta}} \left| \frac{Q_{k,2}}{n^{1/2}} \right|^q; \ v_{x_1} \neq 0 \right)
+ E \left( \sum_{n \leq j < n_{\Delta}, \ 1 - n \leq j < 1 - n_{\Delta}} (S^*_n)^q; \ v_{x_1} = 0 \neq 0 \right) + E ((S^*_n)^q; \ v_{x_1} = 0) \leq C,
\]
where
\[
Q_{k,2} = \begin{cases} \frac{\sum_{i=1}^k x_i \varepsilon_i x_i \varepsilon_{i+1}}{n^{1/2}}, k > 0, \\ \frac{\sum_{i=0}^{k-1} x_i \varepsilon_i x_i \varepsilon_{i+1}}{n^{1/2}}, k \leq 0, \end{cases}
\]
and
\[
S^*_n = \frac{\sum_{i=1}^{n-1} x_i \varepsilon_i}{n^{1/2}}.
\]
\[\blacksquare\]

Proof of (32). Given $q \geq 1$, $E|e_1|^s < \infty$ for $s$ large enough, by (35) and Lemma 20,
\[
E \left| \sum_{i \in J_1} x_i \varepsilon_{i+1} \right|^q \leq \left[ E^{-q} \left| \sum_{i \in J_{1,0}} x_i \varepsilon_{i+1} \right|^q + E^{-q} \left| \sum_{i \in (J_1 \setminus J_{1,0}) \cup (J_{1,0} \setminus J_1)} x_i \varepsilon_{i+1} \right|^q \right]^q \leq C.
\]
\[\blacksquare\]

Proof of 25. Let $q \geq 1$(the case $1 > q > 0$ we use Jensen’s inequality), $\Delta \in \mathcal{D} \mathcal{E}_1^\infty \cap \mathcal{M}_q \neq \emptyset$, where $\mathcal{M}_q = \left\{ x : 0 < x < \frac{\alpha}{12(1+q)} \right\}$(note that $q$ is the required target moment for $\hat{r}_n$; as $q$ grows, $\mathcal{M}_q$ become more restricted; the restriction $\mathcal{D} \mathcal{E}_1^\infty$ is for the need of Proposition 7; $\mathcal{M}_q$ is for (40)). We prove
\[
E|n(\hat{r}_n - r)|^q \leq C.
\]
By this, for all large $n$,
\[
E|n(\hat{r}_n - r)|^q \leq P(U_{n}) \times (n|r)|^q + E|n(\hat{r}_n - r)|^q \leq C.
\]
Let $E^c_{p_1,c_0} = \{ q^{\beta} \geq (1 - p_1)n \} \cap \{ q^{\beta} \geq c_0 n \}$. By Corollary 11, there exist $c_0, p_1, c, \delta > 0$ such that
\[
P(E^c_{p_1,c_0}) \leq \exp(-cn^\delta).
\]
For $E^c_{p_1,c_0}$, we have random indexes $\{ s'_1, \ldots, s'_{c_0} \}$ which are analogous to the random indexes $\{ s_1, \ldots, s_{(1-\delta)p_1} \}$ in Lemma 9. Notice that since $E^c_{p_1,c_0} \subset E_{p_1,c_0}$, $\{ s'_1, \ldots, s'_{c_0} \} \subset \{ s_1, \ldots, s_{(1-\delta)p_1} \}$. Then by a version of Lemma 9, there exists $m > 0$ such that for all $2 \leq k \leq c_0 n$, for any $X_{i=1}^{k-1} \beta_i \subset [r, M]^{k-1}, \beta_k \subset [r, M],
\[
P \left( \bigcap_{i=1}^{k-1} \beta_i = X_{i=1}^{k-1} \beta_i, E^c_{p_1,c_0} \right) \geq m^{[\beta_k]}.
\]
(36)
Define $u_1 = \arg \min_{x_{s_1}, \ldots, x_{s_{c_0}}} x_{s_i}$; $u_i = \arg \min_{x_{sj} \geq x_{ui-1}, 1 \leq j \leq c_0} x_{s_i}$; $c_0 \geq 1$. Let $Y_i \sim \text{Uniform}(r, r + m^{-1})$, $i = 1, \ldots, c_0 n$. Claim:

$$P \left( x_{u_j} \leq z \bigg| E'_{\rho_1, c_0} \right) \equiv F_{x_{u_j}|E'_{\rho_1, c_0}}(z) \geq F_{Y_j}(z), j > 0, \forall z \in \mathbb{R}.$$  

Proof of the claim.

$$F_{x_{u_j}|E'_{\rho_1, c_0}}(z) = P( \text{At least } j \text{ } x_{s_i}'s \in [r, r + z]) E'_{\rho_1, c_0} = \sum_{s=j}^{c_0 n} P( \text{Exact } s \text{ } x_{s_i}'s \in [r, r + z]) E'_{\rho_1, c_0} \geq \sum_{s=j}^{c_0 n} P( \text{Exact } s \text{ } Y_i's \in [r, r + z]) = F_{Y_j}(z).$$

The inequality follow from (36) and Lemma 10. \hfill ■

By the claim, for $1 \leq j \leq c_0 n$, $q \geq 1$,

$$E|Y_{(j)} - r|^q \geq E \left[ |x_{u_j} - r|^q \bigg| E'_{\rho_1, c_0} \right].$$  

(37)

Let $u_i, s_i,$ $i \leq 0$ denote the elements on the left side, then we also have for $0 \geq j \geq -c_0 n + 1$, $q \geq 1$,

$$E|Y_{(j+1)} - r|^q \geq E \left[ |x_{u_j} - r|^q \bigg| E'_{\rho_1, c_0} \right].$$  

(38)

Because $m(Y_{(j)} - r) \sim \text{Beta}(j, c_0 n - j + 1)$ for any $c_0 n \geq j > 0$, we have

$$E|Y_{(j)} - r|^q = m^{-q} \times \prod_{h=0}^{q} \left( \frac{j + h}{c_0 n + 1 + h} \right).$$  

(39)

By standard inequalities, (37) to (39), $P(E'_{\rho_1, c_0} > \frac{1}{2}$; Proposition 7, and the moment assumption $E|c_1|^s < \infty$ for sufficiently large $s$, the choice of $\Delta$, $P(j_3 = i)$ is geometrically decayed in $i$ and $|x + y|^\frac{1}{2} \leq |x|^\frac{1}{2} + |y|^\frac{1}{2}$, we have

$$\sum_{1-n_\Delta \leq j \leq n_\Delta} E \left[ n(x_{s_i'} - r)^q; 0 \neq \tilde{r}_n = x_{s_i'} \bigg| E'_{\rho_1, c_0} \right] \leq C \sum_{1-n_\Delta \leq j \leq n_\Delta} E^{1/2} (n(Y_{(j)} - r))^{2q} P^{1/2}(0 \neq \tilde{r}_n = x_{s_i'})$$

$$\leq C \sum_{1 \leq j \leq n_\Delta} \left\{ \left( \frac{n}{m} \right)^q \prod_{h=0}^{h=2q-1} \left( \frac{j + h}{c_0 n + 1 + h} \right) \right\}^{1/2} \times \left[ P(j_3 = j) + (\log n)^2 n^{-\left(\alpha \wedge (1-18\Delta)\right)} \right]^{1/2}$$

$$\leq C \sum_{1 \leq j \leq n_\Delta} \left\{ \left( \frac{n}{m} \right)^q \prod_{h=0}^{h=2q-1} \left( \frac{j + h}{c_0 n + 1 + h} \right) \right\}^{1/2} \times \sum_{1 \leq j \leq n_\Delta} \left[ (\log n)n^{-\frac{1}{2}(\alpha \wedge (1-18\Delta))} \right] \leq C,$$

(40)
where $C$ absorbs summation over the negatively indexed terms in the second inequality.

By (40), (34), standard inequalities and the moment assumption, for all large $n$,

$$
E\left[|n\tilde{r}_n - r|^q\right] = \sum_{1-n\leq i\leq n} E\left[|n(x_{S_i} - r)|^q; x_{S_i} = \tilde{r}_n \neq 0\right] + (n|r|^q) P(\tilde{r}_n = 0)
$$

$$
\leq \sum_{1-n\Delta\leq i\leq n\Delta} E\left[|n(x_{S_i} - r)|^q; x_{S_i} = \tilde{r}_n \neq 0\right] E_{\rho, c_0}^c + n \times \frac{\sum_{i=1}^n |n(x_i - r)|^q; E_{\rho, c_0}^c P(E_{\rho, c_0}^c)}{P(E_{\rho, c_0}^c)}
$$

$$
+ \sum_{\substack{n\geq j\geq n\Delta, \\
1-n\leq j<1-n\Delta}} E\left[|n(x_{S_i} - r)|^q; x_{S_i} = \tilde{r}_n \neq 0\right] + (n|r|^q) P(\tilde{r}_n = 0) \leq C.
$$

(41)

\[\blacksquare\]

References


B. E. Hansen. Inference When a nuisance parameter Is not identified under the null hypothesis *Econometrica*, 1996.


Supplementary file to SETAR(1) – This is the title

Notation for Proofs

The Model

We consider the self-exciting threshold autoregressive process (without intercepts). Namely, for $t > p$,

$$x_t = \begin{cases} 
a_{11}x_{t-1} + \cdots + a_{1p}x_{t-p} + e_t = \hat{a}_1'\tilde{x}_{t-1} + e_t, & x_{t-d} \leq r, \\
a_{21}x_{t-1} + \cdots + a_{2p}x_{t-p} + e_t = \hat{a}_2'\tilde{x}_{t-1} + e_t, & x_{t-d} > r, 
\end{cases} \tag{42}$$

where $e_t$’s satisfy assumption 1, $r \in \Theta \subset \Theta' \subset \Theta''$; $\Theta''$ is a compact set in $\mathbb{R}^1$ and $\Theta'$ is our threshold estimator searching region; some requirement of the choice of $\Theta''$ and $\Theta'$ can be found in Appendix. Chan and Tong (1985) shows that under the assumptions 1, 2 (see Appendix), $\hat{x}_t$ has a (limiting) stationary probability density function, $\tilde{\pi}$, which is bounded away from zero and infinity on any given compact set in $\mathbb{R}^p$; the univariate marginal stationary density function which is denoted as $\pi$. We assume $(x_1, \ldots, x_p) \overset{D}{\sim} \tilde{\pi}$, and define $\mathcal{F}_i = \sigma(x_1, \ldots, x_p, e_{p+1}, \ldots, e_i)$, the $\sigma$-field generated by $(x_1, \ldots, x_p, \ldots, e_i)$. Let $p(\tilde{z}, y), \tilde{z} \in \mathbb{R}^p, y \in \mathbb{R}^1$, denotes the transitional probability density function, i.e.

$$p(\tilde{z}, y) = \begin{cases} f_e(y - \hat{a}_1'\tilde{z}), & z_{p+1-d} \leq r, \\
- f_e(y - \hat{a}_2'\tilde{z}), & z_{p+1-d} > r, \end{cases}$$

where $f_e$ denote the pdf of $e_1$. By the definition of conditional expectation, for any $B \subset \mathbb{R}^1, i > p$,

$$P(x_i \in B | \mathcal{F}_{i-1}) = \int_B p(\tilde{x}_{i-1}, y) dy \ a.s.$$ 

We note that the joint probability density of $(x_1, \ldots, x_n)$ is $\tilde{\pi}(\tilde{z}_p) \times \cdots \times p(\tilde{z}_{n-1}, z_n)$, $(z_1, \ldots, z_n) \in \mathbb{R}^n$, so for $A \subset \mathbb{R}^n$,

$$P((x_1, \ldots, x_n) \in A) = \int_A \tilde{\pi}(\tilde{z}_p) \times \cdots \times p(\tilde{z}_{n-1}, z_n) dz_1 \cdots dz_n.$$ 

For $\tilde{z}_p \in \mathbb{R}^p, z_k \in \mathbb{R}^1$, define

$$p^k(\tilde{z}_p, z_{p+k}) = \int_{\mathbb{R}^{k-1}} p(\tilde{z}_p, z_{p+1}) \times \cdots \times p(\tilde{z}_{k+p-1}, z_{k+p}) dz_{p+1} \cdots dz_{k+p-1}$$

And we can have for any $B \subset \mathbb{R}^1$,

$$P(x_{s+k} \in B | \mathcal{F}_s) = \int_B p^k(\tilde{x}_s, z_k) dz_k, \ a.s.$$ 

For $\tilde{z}_p, \tilde{z}_{p+k} \in \mathbb{R}^p, k > p$, define

$$\hat{p}^k(\tilde{z}_p, \tilde{z}_{p+k}) = \int_{\mathbb{R}^{k-p}} p(\tilde{z}_p, z_{p+1}) \times \cdots \times p(\tilde{z}_{k+p-1}, z_{k+p}) dz_{p+1} \cdots dz_k$$
And we can have for any \( B \subset \mathbb{R}^p \),
\[
P(\tilde{x}_{s+k} \in B \mid \mathcal{F}_n) = \int_B \tilde{p}^k(\tilde{x}_s, \tilde{z}_k)dz_{k-p+1} \ldots dz_k, \text{ a.s.}
\]

With these notation, we proof Lemma 3.

**Proof of Lemma 3.** Boundness against zero can be readily seen from (27). The \((n+1)\)-th marginal density function given initial state \( \tilde{z}_0 \in \mathbb{R}^p \),
\[
\pi_{n+1}(\tilde{z}_0, z) = \int_{\tilde{z}_0 \in \mathbb{R}^p} f_e(z - T(\tilde{y}))\tilde{p}^n(\tilde{z}_0, \tilde{y})d\tilde{y}.
\]

Since \( f_e \) is Holder continuous, we have for any \( \delta > 0 \), all \( n \), \( |x - y| < \delta^{n+1} \), \( \tilde{z}_0 \in \mathbb{R}^p \),
\[
|\pi_{n+1}(\tilde{z}_0, x) - \pi_{n+1}(\tilde{z}_0, y)| \leq \int_{\tilde{z}_0 \in \mathbb{R}^p} |f_e(x - T(\tilde{y})) - f_e(y - T(\tilde{y}))|\tilde{p}^n(\tilde{z}_0, \tilde{y})d\tilde{y} \leq \theta_1 \delta^{n+1}.
\]

By (43) and the boundness of \( f_e \), there exists \( C > 0 \) such that for all \( n \) and \( x \in \mathbb{R}^1 \),
\[
|\pi_{n+1}(\tilde{z}_0, x) - \pi(x)| \leq C(1 + \|\tilde{z}_0\|)\rho^{n+1}.
\]

The last two equations show \( \pi \) is uniformly continuous on any given compact set. \( \blacksquare \)

From Chan and Tong(1985) and Chan(1988), we have a total variation norm convergence rate on the stationary Markov process: there exist \( 0 < \rho < 1, C > 0 \) such that for all \( p < k \in \mathbb{N}, \tilde{z}_p \in \mathbb{R}^p \),
\[
\int_{\mathbb{R}^p} \left| \pi(\tilde{y}) - \tilde{p}^k(\tilde{z}_p, \tilde{y}) \right|d\tilde{y} \leq C(1 + \|\tilde{z}_p\|)\rho^k; \tag{43}
\]

**Modified Estimation**

In the following we define the notation required for conducting an analysis on the modified estimators. Let \( \theta \in \Theta' \); the \( i \)-th times \( x_j \)'s falling inside \( \theta \): \( T_1^\theta = \inf\{k : x_k \in \theta, k \geq 1\} \), \( T_i^\theta = \inf\{k : x_k \in \theta, k > T_{i-1}^\theta\}, i \geq 2 \); the \( c \)-log non-repeating occurring counter part of \( T_i^\theta \): \( T'_1 = T_1^\theta \), \( T'_i = \inf\{k : x_k \in c \log n + T_{i-1}^\theta\}, i \geq 2 \); the total number of \( x_j, j = 1, \ldots, n \) falling inside \( \theta \): \( q^\theta = |\{k : x_k \in \theta, 1 \leq k \leq n\}| \); the analog of \( q^\theta \): \( q'^\theta = \sup\{k : T_k^\theta \leq n\} \); some related definitions: \( \inf\{\emptyset\} = \infty \), \( \sup\{\emptyset\} = 0 \), \( T_0^\theta = T_0'^\theta = 0 \). Then we define the subset, \( \Pi^r \), which includes all integers lesser than the number of samples falling on the right hand side of \( r \), and greater than the negative number of those on the left:
\[
q_+^r = -q_{-\infty}^r + 1, q_+^r = q_+^{r, \infty} \]
\[
\Pi^r = \{i : q_+^r \leq i \leq 0, 0 < i \leq q_+^r, i \in \mathbb{Z}, \text{ the integer sets}\};
\]
we also define the random indexes which focus on the relative position on the real line of each sample to \( r \), rather than the time series index:

\[
S^r_k = \begin{cases} 
i, x_i = x_{(k+q(-\infty,r])}, k \in \Pi^r, \\
\infty, \text{ o.w.,}
\end{cases}
\]

where \( x_{(i)} \) is the \( i \)-th ordered statistics of \( x_i, i = 1, \ldots, n \).

We can define \( x_\infty \) as anything, since throughout this paper we will not have any direct manipulation on \( \{S^r_i = \infty\} \), and without potential confusion, we will drop off the superscript of \( S^r_i \) to keep the context simple. We categorize \( x_i, p - d + 1 \leq i \leq n - d \) into two groups(two sets of random indexes): one is \( J_{1,j,d} = \{S^r_k : k \leq j, k \in \Pi^r\}/\{n, \ldots, n - d + 1, p - d, \ldots, 2, 1\} \) and the other is \( J_{2,j,d} = \{p - d + 1, \ldots, n - d\} \setminus J_{1,j,d} \). Note that \( J_{1,0,d} \cup J_{2,0,d} = J_{1,j,d} \cup J_{2,j,d} \) for all \( j \), and that indexed by such, the subset is regarded as the threshold candidates given the thresholding lag \( d \), i.e., given \( d \), only to which the threshold value is fitted can we have non-singular covariance matrices on both states(also regarded as the lag-\( d \) subsample).

The loss when the threshold is fitted to each of \( x_{S^r_j} \), \( 1 - n \leq j \leq n \) is

\[
Z_{j,d} = \begin{cases} 
\sum_{1+i-d \in J_{1,j,d}} (x_i+1 - \hat{x}_i+1)^2 + \sum_{1+i-d \in J_{2,j,d}} (x_i+1 - \hat{x}_i+1)^2, & \text{if } |J_{k,j,d}| \geq p, k = 1, 2, \\
\infty, \text{ o.w.,}
\end{cases}
\]

where on \( \{|J_{k,j,d}| \geq p, k = 1, 2\} \),

\[
\hat{x}_{i+1} = \begin{cases} 
\hat{a}_{J_{1,j,d}} \bar{x}_i, 1 + i - d \in J_{1,j,d}, \\
\hat{a}_{J_{2,j,d}} \bar{x}_i, 1 + i - d \in J_{2,j,d}
\end{cases}
\]

\[
\hat{a}_{J_{k,j,d}} = \left( \sum_{1+i-d \in J_{k,j,d}} \bar{x}_i \bar{x}_i' \right)^{-1} \left( \sum_{1+i-d \in J_{k,j,d}} \bar{x}_i x_{i+1} \right).
\]

Note that \( \hat{x}_{i+1} \) and \( \hat{a}_{J_{k,j,d}} \) are not required to be defined on \( \bigcup_{k=1,2} \{|J_{k,j,d}| < p\} \) in \( Z_{j,d} \). In essence, \( Z_{j,d} \) is \( L \), the loss function appears in the Section 2, subscripted by \( j \) and \( d \).
With subscript $d$, we are considering the lag-$d$ subsample, $x_{i-d}, i = p+1, \ldots, n$, threshold candidates. Given the thresholding lag $d$, the subscript $j > 0 (j \leq 0)$ indicates that we are measuring the fitted loss with threshold value fitted to the $j$-th largest($(j - 1)$-th smallest) among the lag-$d$ subsample on the right(left) hand side of the origin, $r$. Being explicit in $j$ and $d$ allows us to have a detailed expression of the loss, which is the only advantage of using $Z_{j,d}$ instead of $L$: for $|J_{k,j,d}| \geq p, k = 1, 2, j \geq 1$:

$$Z_{j,d} = \sum_{i=p}^{n-1} c_{i+1}^2 + \sum_{1+i-d \in J_{1,0,d}} \left( (\hat{a}_1 - \hat{a}_{J_{1,j,d}})'^2 \hat{x}_i \right) + 2 \sum_{1+i-d \in J_{1,0,d}} \left( \hat{a}_1 - \hat{a}_{J_{1,j,d}} \right) \hat{x}_i e_{i+1} + \sum_{1+i-d \in J_{2,0,d}} \left( (\hat{a}_2 - \hat{a}_{J_{2,j,d}})'^2 \hat{x}_i \right) + 2 \sum_{1+i-d \in J_{2,0,d}} \left( \hat{a}_2 - \hat{a}_{J_{2,j,d}} \right) \hat{x}_i e_{i+1} + \sum_{i=1}^{j} \left( (\hat{a}_2 - \hat{a}_{J_{1,j,d}}) \hat{x}_{S_{1}+d-1} \right) + 2 \sum_{i=1}^{j} \left( \hat{a}_2 - \hat{a}_{J_{1,j,d}} \right) \hat{x}_{S_{1}+d-1} e_{S_{1}+d},$$

and $Z_{j,d}, j \leq 0; |J_{k,j,d}| \geq p, k = 1, 2$, can be expressed in a similar way. This expression in turn allows us to, for example when $(j, d)$ is far away from the true values, mathematically establish the impossibility of $(j, d)$ being the optima. As the modified estimation has a confined threshold parameter space, we introduce $\bar{P}_{j,d}$, the analog of $Z_{j,d}$ but with bounded threshold parameter space:

$$\bar{P}_{j,d} = \begin{cases} Z_{j,d}, & \text{on } \{x_{S_j} \in \Theta', S_j \in J_{1,0,d} \cup J_{2,0,d}\}, \\ \infty, & \text{o.w.}. \end{cases}$$

Note that with probability one, $\arg \min_{1-n \leq i \leq n; 1 \leq d \leq p} \bar{P}_{i,d}$ contains only one set of $(j, d)$, and we shall write an equality, $=$, instead of $\in$ to save the notation usage.

From Section 2, we see the notation remains to be formally defined is the rare events, $\mathcal{RE}_n$ and $U_n$. A formal definition of $\mathcal{RE}_n$ is

$$\mathcal{RE}_n = \cap_{1 \leq s \leq p} \{ \# \{ i : i \in J_{1,0,s} \cup J_{2,0,s}, x_i \in \Theta', i \leq n \} < p \},$$

whereas that of $U_n$ is

$$U_n^c = \cup_{k=1,2} \left\{ \left\| \sum_{1+i-d \in J_{k,j,d}} \hat{x}_i \hat{x}_i' \right\|^{-1} \geq sn \right\} \cap \left\{ (j, d)' = \arg \min_{1-n \leq i \leq n; 1 \leq d \leq p} \bar{P}_{i,d} \right\} \cap \mathcal{RE}_n^c.$$
and \( \hat{\alpha}_k^o, k = 1, 2 \), are the corresponding least square estimators and \( \hat{0} \) otherwise.

In addition to \( \hat{r}_n^o \), we also introduce a version of threshold estimator, which is numerically really close to \( \hat{r}_n^o \) and can be analytically useful:

\[
\tilde{r}_n = \begin{cases} 
  x_{S^*_j}, & (\hat{j}, \hat{d})' = \arg\min_{-n+1 \leq i \leq n; 1 \leq d \leq p} \tilde{P}_{i,d}, \text{ on } \mathcal{R} \mathcal{E}_n^c, \\
  0, & \text{o.w.,}
\end{cases}
\]

### Preliminary Results

#### Outline introduction for Lemma 8 to 12

Our approach to deriving the main result relies on a list of preliminary works where only loose assumptions and stationary process are required; nevertheless, standard assumptions assumed in this paper remain assumed to avoid unnecessary complexity in the context. Although the usage of each lemmas is somewhat multi-folded, we manage to illustrate the basic need of the lemmas in a nutshell by one moment bound example. Let \( q_n, z_n \) denote some number depending on the sample size \( n; q \geq 1 \).

To bound \( E |\sum_{i=1}^n x_i^2|^q \), we make moment assumption and normalizing the summation by sample size \( n \). The challenge here is how we can maintain this property when we have \( 1_{q_n \in \Pi} \sum_{i=1}^{q_n} x_i^2, r > 0 \) in the integrand. The solution provided here is straightforward: we separately bound it on two collectively exhaustive events. One is that no more than \( q_n \) \( x_i \)’s fall inside a subset \([r, z_n] \); \( 1_{q_n \in \Pi} \sum_{i=1}^{q_n} x_i^2 \) is obviously bounded by \( q_n z_n^2 \) on this event.

The other is the complementary event, which is supposed to be a rare event and we can thus bound it by a sup-bound due to \( \sum_{i=1}^n x_i^2 \geq 1_{q_n \in \Pi} \sum_{i=1}^{q_n} x_i^2 \). Corollary 11, Lemma 12 mathematically quantify \( q_n, z_n \) and the probability of the rare event. For \( x_i \)’s to be an iid process, calculate and bound the exact probability of the event that more than \( q_n \) \( x_i \)’s fall inside \([r, z_n] \) is relatively easier than the same task for a stationary process; Lemma 9, 10 link the stationary process argument up to an iid process argument. One application of lemma 9, 10 can be seen in the proof of Lemma 12.

### Lemma 8 to Lemma 12

As the compact set \( \Theta'' \) is large enough, it is naturally to see most of the data \( x_i \)’s generated from the SETAR process will fall inside it. Lemma 8 describes this idea mathematically. In Lemma 9, we need \( c \) to be smaller than \( \frac{1}{3} \).

**Lemma 8.** For any small \( c > 0 \), there exist \( \alpha_1 > \alpha_2 > 1, p \) is an integer, such that

\[
\alpha_1 p m^*_p < c,
\]

\[
P(q^\Theta'' \leq (1 - \alpha_1 p m^*_p) n) \leq \frac{\alpha_2}{1 - \alpha_2},
\]

where \( \Theta'' = [-A^{-p}, A^{-p}] \), and \( m^* \equiv m^*_p = 2 \max\{E|e_1|, 1\} \times A^p (1 + p)^2 (1 - A)^{-1} \).
Proof of Lemma 8. Define $\rho_1 = \alpha_1 p m^*, x_0 = 0$. Let $t_1 = \inf \{ k : |x_{k-1}| \leq A^{-p}, |x_k| \geq A^{-p}, 1 \leq k \leq n \}$, $t_i = \inf \{ k : |x_k| \geq A^{-p}, t_{i-1} + p \leq k \leq n \}$, $i > 1$.

Claim: For $1 \leq l < j - p$, $k > 0$, $j + (k - 1)p \leq n$,

$$
P(t_1 = j, t_2 = j + p, \ldots, t_k = j + (k - 1)p) \leq m^k,
$$

$$
P(t_i = j, t_{i+1} = j + p, \ldots, t_{i+k} = j + (k - 1)p | t_{i-1} = l) \leq m^k, \forall i > 1.
$$

(44)

Proof of (44). First, without loss of generality, $i > 1$,

$$
P(t_i = j, t_{i-1} = l) \leq P(t_{i-1} = l, |x_{j-1}| \leq A^{-p}, |x_j| \geq A^{-p})
$$

$$
\leq P(t_{i-1} = l, |e_j| + A^{-p}A \geq A^{-p})
$$

$$
= P(t_{i-1} = l) \times P(|e_j| + A^{-p}A \geq A^{-p}),
$$

because $l < j - p$, $|x_j| \leq A|x_{j-1}| + |e_j|$ and $e_j$ is independent of $x_{i-1}, \forall i \geq 0$. By the above inequality,

$$
P(t_i = j | t_{i-1} = l) \leq P(|e_j| + A^{-p}A \geq A^{-p})
$$

$$
\leq E|e_j|A^{-p}(1 - A)^{-1} \leq m^*.
$$

Second, $l < j - p$. By

$$
|x_{j+p(k-1)}| \leq |e_{j+p(k-1)}| + A|x_{j+p(k-1)-1}|,
$$

we have

$$
\{ |x_{j+p(k-1)}| \geq A^{-p} \} \subset \{ |e_{j+p(k-1)}| \lor A|e_{j+p(k-1)-1}| \lor \cdots \lor A^{p-1}|e_{j+p(k-2)+1}| \geq A^{-p} \}
$$

$$
\cup \{ A_p|e_{j+p(k-2)}| \lor \cdots \lor A^{2p-1}|e_{j+p(k-3)+1}| \geq A^{-p} \}
$$

$$
\cdots
$$

$$
\cup \{ A^{(k-1)p}|x_j| \geq A^{-p} \}
$$

$$
eq A_{k-1,k-1} \cup \cdots \cup A_{k-1,0}.
$$

So we have

$$
P(t_{i+k-1} = j + p(k - 1), \ldots, t_i = j, t_{i-1} = l)
$$

$$
\leq P(\{ |x_{j+p(k-1)}| \geq A^{-p}, \ldots, |x_j| \geq A^{-p}, |x_{j-1}| \leq A^{-p}, t_{i-1} = l \})
$$

$$
\leq P(\cap_{u=0}^{k-1}(A_{u,0} \cup \cdots \cup A_{u,0}) \cap \{ |x_{j-1}| \leq A^{-p} \} \cap \{ t_{i-1} = l \}).
$$

Write

$$
\cap_{u=0}^{k-1}(A_{u,0} \cup \cdots \cup A_{u,0}) = \cup_j (A_{k-1,jk-1} \cap A_{k-2,jk-2} \cap \cdots \cap A_{0,0}),
$$

for proper $j(\cdot)$ indexes. Then we find that the RHS has unioned a lot trivial sets. For example,

$$
(A_{3,2} \cap A_{2,2} \cap A_{1,1} \cap A_{1,0}) \subset (A_{3,2} \cap A_{1,1} \cap A_{0,0}),
$$

35
\[ \bigcup_{k=1,2:s=0,1} (A_{3,1} \cap A_{2,k} \cap A_{1,s} \cap A_{0,0}) \subset (A_{3,1} \cap A_{0,0}), \]

and this observation applies to general case. In fact, we only need to consider events with form

\[
A_{k-1,s_1} \cap A_{s_1-1,s_2} \cap \cdots \cap A_{s_{q-1}-1,0}, s_1 > 0, \quad A_{k-1,0}, \text{ o.w.}
\]

with

\[
0 \leq s_i \leq s_{i-1} - 1 \leq k - 2, i \geq 2, \quad (\text{rule})
\]

There are totally \(2^{k-1}\) subsets if we count the events by indexes following the rule. \((S_{k-1} = \sum_{i=0}^{k-2} S_i, k > 2, S_0 = S_1 = 1)\). By the observation, we have

\[
\bigcup_{s_i,t_j} (A_{k-1,s_1} \cap A_{t_2,s_2} \cap \cdots \cap A_{t_q,0}) = \bigcap_{u=0}^{k-1} \bigcup_{u,u} (A_{u,u} \cup \cdots \cup A_{u,u}).
\]

Using this argument and the independence of \(A_{ij}\)'s, we have

\[
P(\bigcap_{u=0}^{k-1} (A_{u,u} \cup \cdots \cup A_{u,u}) \cap \{|x_j| \leq A^{-p} \cap \{t_{i-1} = l\})
\leq \sum_{\text{all possible } s_i,t_i} P((A_{k-1,s_1} \cap A_{t_2,s_2} \cap \cdots \cap A_{t_q,0}) \cap \{|x_j| \leq A^{-p} \cap \{t_{i-1} = l\})
\leq 2^{k-1} \frac{m^* k}{2} \leq m^* k.
\]

By (44), \(i_{j-1} < j_0 - p\),

\[
P(t_1 = i_1, \ldots, t_{j-1} = i_{j-1}, j = j_0 + (k - 1)p)
\leq P(t_1 = i_1, \ldots, t_{j-1} = i_{j-1}) \times m^* k
\]

Again, repeating the above argument we have, for all \(\{i_k\}_{k=1}^j ; j,i,k \leq n\),

\[
P(t_k = i_k, \forall k \leq j, t_{j+1} > n) \leq m^{*ij}.
\]

\[
P(q^{ \theta''} \leq (1 - \rho_1)n) \leq \sum_{n \geq j \geq \frac{p \rho_1}{2}} P(t_k = i_k, \forall k \leq j, t_{j+1} > n)
\leq \sum_{n \geq j \geq \frac{p \rho_1}{2}} C_j^m m^{*ij},
\]

by (45). Define

\[
P_j = C_j^m m^{*ij},
\]

then for \(n \geq j \geq \frac{\alpha_2 \rho_1 n}{2 \alpha_1 p} = \alpha_2 m^* n\),

\[
\frac{P_{j+1}}{P_j} = m^* \frac{j!(n - j)!}{(j + 1)!(n - j - 1)!} = m^* \frac{n - j}{j + 1} \leq m^* \frac{n}{\alpha_2 m^* n} = \alpha_2^{-1}.
\]
So
\[ \sum_{n \geq \frac{2}{\alpha_1} \frac{\rho_1^n}{p}} P_j \leq \sum_{i=0}^{(1-\frac{\alpha_2 \rho_1}{p})n} \alpha_2^{-i}, \]
due to \( P_j \leq 1, \forall j \geq 0 \). And hence
\[ (46) \leq \sum_{n \geq \frac{2}{\alpha_1} \frac{\rho_1^n}{p}} P_j \leq \sum_{i=0}^{(1-\frac{\alpha_2}{p})n} (\alpha_2^{-1})^{i+(\frac{\alpha_1 - \alpha_2}{\alpha_1} \frac{\rho_1}{p})n} \leq \frac{\alpha_2^{-(\alpha_1 - \alpha_2)n^*n}}{1 - \alpha_2^{-1}}, \]
by the definition of \( \rho_1 \).

**Notation for Lemma 9**

Let \( E_{\rho_1} = \{ q^{\Theta''} \geq (1-\rho_1)n \} \). On \( E_{\rho_1} \), we construct random indexes \( s_1, \ldots, s_{(1-3\rho_1)n} \) such that \( s_1 = \inf\{ k : (x_{k-1}, x_k, x_{k+1}) \in \Theta'' \times \Theta'' \times \Theta'', 1 \leq k \leq n, x_0, x_n+1 \equiv 0 \in \Theta'' \} \),
\[ s_i = \inf\{ k : (x_{k-1}, x_k, x_{k+1}) \in \Theta'' \times \Theta'' \times \Theta'', n \geq k > s_{i-1}, x_n+1 \equiv 0 \in \Theta'' \}. \]
We can pick \( \rho_1 < \frac{1}{3} \) in Lemma 8 so \( E_{\rho_1} \subset \{ s_{(1-3\rho_1)n} < \infty \} \).

By our assumption on \( f_e \),
\[ \infty > M_1 = \sup_{(x,y) \in \Theta''^2} p(x,y) \geq \inf_{(x,y) \in \Theta''^2} p(x,y) = m_1 > 0. \]

**Lemma 9.** For any \( \beta_1 \ldots \beta_k, \beta_i \subset \Theta'', \ i = 1, \ldots, k, \ k = 2, \ldots, (1-3\rho_1)n, \)
\[ P(x_{s_k} \in \beta_k | (x_{s_1}, \ldots, x_{s_{(k-1)}}) \in \beta_1 \times \cdots \times \beta_{k-1} \equiv \beta_{-k}, E_{\rho_1}) \geq \frac{m}{|A|}, \]
where \( \frac{|A|}{m} \) is the Lebesgue measure of subset \( A \), \( m = (\frac{m_1}{M_1})^2 |\Theta''|^{-1}. \)

**Proof of Lemma 9.** Let \( i = 1, \ldots, n, \)
\[ y_i = \begin{cases} 
1, x_i \in \Theta'', \\
0, \text{ o.w.}
\end{cases} \]
we see
\[ E_{\rho_1} = \left\{ \sum_{i=1}^{n} y_i \geq (1-\rho_1)n \right\} = \bigcup_{i=(1-\rho_1)n}^{n} \left\{ \sum_{j=1}^{n} y_j = i \right\} \]
From the above expression for \( E_{\rho_1} \), we observe there are \( N_{\rho_1,n} = \sum_{i=(1-\rho_1)n}^{n} C_i^n \) disjoint and exhaustive events in \( E_{\rho_1} \), and we denote these events by \( E_i, i = 1, \ldots, N_{\rho_1,n} \). Then we define \( \tilde{E}_i \subset \mathbb{R}^n \) such that \( E_i = \{ (x_1, \ldots, x_n) \in \tilde{E}_i \}. \) Note that on each \( E_j \), there are known constants \( b_i \)'s such that \( s_i = b_i, i = 1, \ldots, (1-3\rho_1)n \). We write
\[ P(x_{s_k} \in \beta_k | (x_{s_1}, \ldots, x_{s_{k-1}}) \in \beta_{-k}, E_{\rho_1}) \]
\[ = \frac{\sum_{1 \leq i \leq N_{\rho_1,n}} P(x_{s_k} \in \beta_k, (x_{s_1}, \ldots, x_{s_{k-1}}) \in \beta_{-k}, E_i)}{P((x_{s_1}, \ldots, x_{s_{k-1}}) \in \beta_{-k}, E_{\rho_1})}. \]
For each $E_i$ and constants $b_j$’s with $s_i = b_j, i = 1, \ldots, (1 - 3\rho_1)n$ on $E_i$, any $A_{b_k} \subset \mathbb{R}^{b_k-1}, A_{b_k} \subset \mathbb{R}^{n-b_k}$,

\[
\left(\frac{M_1}{m_1}\right)^2 P(x_{b_k} \in \beta_k, \cup_{i} \{(x_1, \ldots, x_{b_k-1}) \in A_{1,b_k}^* \} \cup \{(x_{b_k+1}, \ldots, x_n) \in A_{2,b_k}^* \}) , E_i
\]

\[
= \left(\frac{M_1}{m_1}\right)^2 \int_{\cup_{i}} \left( A_{1,b_k}^* \times \beta_k \times A_{2,b_k}^* \right) \cap E_i \pi(z_1) \ldots p(z_{b_k-2}, z_{b_k-1}) p(z_{b_k-1}, z_{b_k}) p(z_{b_k}, z_{b_k+1}) \ldots dz_1 \ldots dz_n
\]

\[
\geq \left(\frac{M_1}{m_1}\right)^2 \int_{\cup_{i}} \left( A_{1,b_k}^* \times \beta_k \times A_{2,b_k}^* \right) \cap E_i \inf_{i=1,2,3} \pi(z_1) \ldots p(z_{b_k-2}, z_{b_k-1}) p(z_{2}, z_{1}) p(z_{3}, z_{b_k+1}) \ldots dz_1 \ldots dz_n
\]

\[
= \left(\frac{M_1}{m_1}\right)^2 \int_{\cup_{i}} \left( A_{1,b_k}^* \times \beta_k \times A_{2,b_k}^* \right) \cap E_i \inf_{i=1,2,3} \frac{\beta_k}{|\Theta''|} \int_{\cup_{i}} \left( A_{1,b_k}^* \times \beta_k \times A_{2,b_k}^* \right) \cap E_i \pi(z_1) \ldots p(z_{b_k-2}, z_{b_k-1}) p(z_{2}, z_{1}) p(z_{3}, z_{b_k+1}) \ldots dz_1 \ldots dz_n
\]

\[
\geq \left(\frac{M_1}{m_1}\right)^2 \int_{\cup_{i}} \left( A_{1,b_k}^* \times \beta_k \times A_{2,b_k}^* \right) \cap E_i \sup_{i=1,2,3} \pi(z_1) \ldots p(z_{b_k-2}, z_{b_k-1}) p(z_{2}, z_{1}) p(z_{3}, z_{b_k+1}) \ldots dz_1 \ldots dz_n
\]

\[
= \left(\frac{M_1}{m_1}\right)^2 \int_{\cup_{i}} \left( A_{1,b_k}^* \times \beta_k \times A_{2,b_k}^* \right) \cap E_i \frac{\beta_k}{|\Theta''|} \int_{\cup_{i}} \left( A_{1,b_k}^* \times \beta_k \times A_{2,b_k}^* \right) \cap E_i \pi(z_1) \ldots p(z_{b_k-2}, z_{b_k-1}) p(z_{2}, z_{1}) p(z_{3}, z_{b_k+1}) \ldots dz_1 \ldots dz_n
\]

\[
= \frac{\beta_k}{|\Theta''|} \int_{\cup_{i}} \left( A_{1,b_k}^* \times \beta_k \times A_{2,b_k}^* \right) \cap E_i \pi(z_1) \ldots p(z_{b_k-2}, z_{b_k-1}) p(z_{2}, z_{1}) p(z_{3}, z_{b_k+1}) \ldots dz_1 \ldots dz_n
\]

\[
= \frac{\beta_k}{|\Theta''|} P(x_{b_k} \in \Theta''), \cup_{i} \{(x_1, \ldots, x_{b_k-1}) \in A_{1,b_k}^* \} \cup \{(x_{b_k+1}, \ldots, x_n) \in A_{2,b_k}^* \} , E_i
\]

\[
= \frac{\beta_k}{|\Theta''|} P(\cup_{i} \{(x_1, \ldots, x_{b_k-1}) \in A_{1,b_k}^* \} \cup \{(x_{b_k+1}, \ldots, x_n) \in A_{2,b_k}^* \} , E_i).
\]

(48)

By (48), for $i \leq N_{\rho_1,n}$,

\[
P(x_{s_k} \in \beta_k, (x_{s_1}, \ldots, x_{s_{k-1}}) \in \beta_{k-1}, E_i) \geq \left(\frac{m_1}{M_1}\right)^2 \frac{\beta_k}{|\Theta''|} P((x_{s_1}, \ldots, x_{s_{k-1}}) \in \beta_{k-1}, E_i)
\]

\[
= \frac{m \beta_k}{m_1} P((x_{s_1}, \ldots, x_{s_{k-1}}) \in \beta_{k-1}, E_i).
\]

So

\[
(47) \geq \sum_{1 \leq i \leq N_{\rho_1,n}} \frac{m \beta_k}{m_1} P((x_{s_1}, \ldots, x_{s_{k-1}}) \in \beta_{k-1}, E_i)
\]

\[
= m |\beta_k|.
\]

Lemma 10. Let $y_i \sim^j \text{Uniform}(\inf \Theta'', \inf \Theta'' + m^{-1}), i = 1, \ldots, n$. If for any rectangle $\beta \subset \Theta''_{s-1}, 2 \leq s \leq n, \beta \subset \Theta''$, any event $E$ that is independent of $y_i$’s,

\[
P \left( x_s \in \beta \mid (x_1, \ldots, x_{s-1}) \in \beta, E \right) \geq m |\beta|,
\]

(49)

then for any $\theta_1 \subset \Theta''$, $0 \leq j \leq n$,

\[
P \left( \text{At least } j \text{ } x_i's, i = 1, \ldots, n, \in \theta_1 \mid E \right) \geq P \left( \text{At least } j y_i's, i = 1, \ldots, n, \in \theta_1 \right).
\]

38
Proof of Lemma 10. For the case $j = 0$, the $RHS = LHS = 1$. In the below, we assume $n \geq j > 0$. Let $\theta_0 = \theta_0^j$. We define $q^j_i$ such that
\begin{align*}
q^j_i &= \begin{cases} 
    x_i, & i = 1, \ldots, n. \\
    y_i, & i \geq n - j + 1.
\end{cases} 
\end{align*}

By (49) and the independence of $y_i$, we have, for any $\{k_i\}_{i=1; i\neq n-j}^n$, $k_i \in \{0, 1\}$, $k_{n-j} = 1$, $0 \leq j \leq n - 1$,
\begin{align*}
P(q^j_i \in \theta_{k_i}, \forall i | E) &= P(x_i \in \theta_{k_i}, i \leq n - j; y_s \in \theta_{k_s}, n \geq s > n - j | E) \\
&= P(x_{n-j} \in \theta_{k_{n-j}} | x_i \in \theta_{k_i}, i \leq n - j - 1, E) \times P(x_i \in \theta_{k_i}, i \leq n - j - 1 | E) \\
&\quad \times P(y_s \in \theta_{k_s}, n \geq s > n - j) \\
&\geq P(y_{n-j} \in \theta_{k_{n-j}}) \times P(x_i \in \theta_{k_i}, i \leq n - j - 1 | E) \\
&\quad \times P(y_s \in \theta_{k_s}, n \geq s > n - j) \\
&= P(q^{j+1}_i \in \theta_{k_i}, \forall i | E).
\end{align*}
(50)

For any $\{k_i\}_{i=1; i\neq n-j}^n$, since $q^j_i, q^{j+1}_i, i = 1, \ldots, n$ only differ by $q^j_{n-j}, q^{j+1}_{n-j}$,
\begin{align*}
\sum_{u=0,1} P(q^j_i \in \theta_{k_i}, i \neq n-j, q^j_{n-j} \in \theta_u | E) = \sum_{u=0,1} P(q^{j+1}_i \in \theta_{k_i}, i \neq n-j, q^{j+1}_{n-j} \in \theta_u | E).
\end{align*}
(51)

By an observation on the pattern of all combination, (50), (51), for $0 \leq k \leq n - 1$,
\begin{align*}
\sum_{s=j}^n P(\text{Exact } s q^k_i's \in \theta_1 | E) &= P(\text{Exact } j q^k_i's \in \theta_1 \text{ with } q^{k}_{n-k} \in \theta_1 | E) \\
&\quad + \sum_{u=0,1} \sum_{s=j}^{n-1} P(\text{Exact } q^k, \ldots, q^k_{n-k-1}, q^k_{n-k+1}, \ldots, q^n \in \theta_1, q^{k}_{n-k} \in \theta_u | E) \\
&\geq P(\text{Exact } j q^{k+1}_i's \in \theta_1 \text{ with } q^{k+1}_{n-k} \in \theta_1 | E) \\
&\quad + \sum_{u=0,1} \sum_{s=j}^{n-1} P(\text{Exact } q^{k+1}, \ldots, q^{k+1}_{n-k-1}, q^{k+1}_{n-k+1}, \ldots, q^{k+1}_n \in \theta_1, q^{k+1}_{n-k} \in \theta_u | E) \\
&= \sum_{s=j}^n P(\text{Exact } s q^{k+1}_i's \in \theta_1 | E).
\end{align*}

Reapting $k$, the proof for $j > 0$ is finished. \hfill \blacksquare

We now introduce Corollary 11, which would be a refinement of Lemma 8. The major difference between these two statements is the subset $\theta \subset \Theta''$ we are considering. The need of this refinement will be clear in Proposition 5. Note that the proof of Corollary 11 is based on that of Lemma 8.
Corollary 11. For any $\theta \subset \Theta''$ with $|\theta \cap \Theta'| \geq c_0, 1 > c_0 > 0$. There exists $c_1, \delta > 0$ such that for all large $n$, 
\[
P(q^\theta < \rho_2 n) \leq \frac{\alpha_2}{1 - \alpha_1} \leq \exp(-c_1 n^\delta),
\]
where $\rho_2 = \frac{1}{2} \times c_0 \times (1 - 3\rho_1) \times m$, $\rho_1, \alpha_1, \alpha_2$ are defined in Lemma 8, $m$ are given in Lemma 9.

Proof of Corollary 11. $x_{s_1}, \ldots, x_{s(1-3\rho_1)n}$ are defined in Lemma 9, and $y_1, \ldots, y_{(1-3\rho_1)n}$ are defined in Lemma 10. By Lemma 10, for any $0 \leq j_0 \leq (1 - 3\rho_1)n$, 
\[
\sum_{j \leq j_0} P(\text{Exact } j'\text{s in } \theta \mid E_{\rho_1}) \leq \sum_{j \leq j_0} P(\text{Exact } j'\text{s in } \theta \mid E_{\rho_1})
\leq \sum_{j \leq j_0} C_j^{(1-3\rho_1)n}(c_0 m)^j(1 - c_0 m)^{(1-3\rho_1)n - j}
\equiv \sum_{j \leq j_0} P_j
\]
For any $j \leq \frac{1}{2}(1 - 3\rho_1)c_0 mn$,
\[
P_{j-1} = \frac{j}{(1 - 3\rho_1)n - j + 1} \left(\frac{1 - c_0 m}{c_0 m}\right) \leq \frac{c_0 m^{\frac{1}{2}}(1 - 3\rho_1)}{(1 - 3\rho_1)(1 - c_0 m^{\frac{1}{2}})} \frac{(1 - c_0 m)}{c_0 m} \leq \frac{1}{2},
\]
in the second inequality, we take $j = \frac{1}{2}(1 - 3\rho_1)c_0 mn$. So for all $j_0 \leq \frac{1}{2}(1 - 3\rho_1)c_0 mn$,
\[
\sum_{j \leq j_0} P_j \leq \frac{\left(\frac{1}{2}\right)^{\frac{1}{2}(1 - 3\rho_1)c_0 mn - j_0}}{1 - \frac{1}{2}}
\]
We have, for $j_0 = \frac{1}{2} \times \frac{1}{2}(1 - 3\rho_1)c_0 mn = \rho_2n$,
\[
P(q^\theta < \rho_2 n \mid E_{\rho_1}) \leq \sum_{j \leq \rho_2 n} P(\text{Exact } j'\text{s in } \theta \mid E_{\rho_1}) \leq \left(\frac{1}{2}\right)^{\frac{1}{2} \times \frac{1}{2}(1 - 3\rho_1)c_0 mn} \frac{(1 - c_0 m)}{c_0 m} \leq \left(\frac{1}{2}\right)^{\rho_2 n - 1}.
\]
Then we have
\[
P(q^\theta < \rho_2 n) \leq P(q^\theta < \rho_2 n \mid E_{\rho_1}) + P(E_{\rho_1}^c),
\]
and by Lemma 8, we have finished the proof.

While Corollary 11 cares about the $x_i$’s falling inside a constant compact set, we need to consider the situation of a shrinking subset as well.

Lemma 12. Given $1 > \Delta > 0$. There exist $c_1, c_2, \delta > 0$ whose values depend on $\Delta$, such that for $0 < b < 1 - \Delta$, $z \in \Theta', \theta_{Z,b} \equiv [z, z + c_2 n^{-b}] \subset \Theta''$, all large $n$,
\[
P(q^\theta_{z,b} \leq \rho_3 n^{1-b}) \leq \left(\frac{1}{2}\right)^{\rho_3 n^{1-b} - 1} + \frac{\alpha_2^{-\frac{-(\alpha_1 - \alpha_2)m^* n}{1 - \alpha_2}}}{\alpha_2^{-\frac{-(\alpha_1 - \alpha_2)m^*}{1 - \alpha_2}}} \leq \exp(-c_1 n^\delta),
\]
where \( \rho_3 = \frac{1}{4} m(1-3\rho_1) \), \( \rho_1, \alpha_1, \alpha_2, m^* \) are defined in Lemma 8, \( m \) are given in Lemma 9. Moreover, there exist \( c_3, c_4, \delta' > 0 \) whose values depend on \( \Delta \), such that for all \( 0 < b < 1 - \Delta \), \( z \in \Theta', \theta_{z,b} \equiv [z, z + c_4 n^{-b}] \subset \Theta'' \), all large \( n \),

\[
P(q^{z,b} \geq 4Mn^{1-b}) \leq \exp(-c_3 n^{\delta'}).
\]

**Proof of Lemma 12.** We prove the first assertion and without loss of generality we assume \( c_2 = 1 \). The other part can be done by the same argument.

Let \( y_i, x_{s,i}, i = 1, \ldots, (1 - 3\rho_1)n \) be defined in Lemma 10 and Lemma 9, respectively. Then

\[
\sup_{z \in \Theta'} P \left( q^{z,b} \leq \rho_3 n^{1-b} \right)
\leq \sup_{z \in \Theta'} \left[ P \left( q^{z,b} \leq \rho_3 n^{1-b} \mid E_{\rho_1} \right) + P \left( E_{\rho_1}^c \right) \right]
\leq \sup_{z \in \Theta'} \sum_{j=0}^{\rho_3 n^{1-b}} P( \text{Exact } j x_{s,i} \in \theta_{z,b} \mid E_{\rho_1}) + \frac{\alpha_2^-(\alpha_1 - \alpha_2) m^* n}{1 - \alpha_2},
\]

(52)

the second inequality is due to that conditional on \( E_{\rho_1} \), the event on the LHS is a subset of the event on the RHS. The third inequality came from the definition of \( y_i \)'s, \( \Theta' \subset \Theta'' \), and Lemma 10.

Let

\[
g(j) = P( \text{Exact } j y_i \in \theta_{z,b}) = C_j^{(1-3\rho_1)n}(mn^{-b})^j (1 - mn^{-b})^{(1-3\rho_1)n-j}.
\]

For all \( 0 < j \leq \frac{1}{2} m(1 - 3\rho_1)n^{1-b} \equiv 2\rho_3 n^{1-b} \),

\[
\frac{g(j - 1)}{g(j)} = \left( \frac{j}{(1 - 3\rho_1)n - j + 1} \right) \left( \frac{1 - mn^{-b}}{mn^{-b}} \right) \leq \frac{1}{2},
\]

\[
\sum_{j \leq \rho_3 n^{1-b}} g(j) \leq \frac{1}{1 - \frac{1}{2}} \times \left( \frac{1}{2} \right)^{(2\rho_3 - \rho_1)n^{1-b}} = \left( \frac{1}{2} \right)^{\frac{1}{2} m(1 - 3\rho_1)n^{1-b} - 1}.
\]

Hence for all \( 0 < b < 1 - \Delta \), there are constants \( c, \delta > 0 \) such that

\[
(52) \leq \left( \frac{1}{2} \right)^{\frac{1}{2} m(1 - 3\rho_1)n^{1-b} - 1} + \frac{\alpha_2^-(\alpha_1 - \alpha_2) m^* n}{1 - \alpha_2} \leq \exp(-cn^{\delta}).
\]

\[\blacksquare\]

**Outline introduction for Lemma 13, 14**

Let \( e_i, \bar{e}_i, e_{i,k}, k = 1, 2, i = 1, \ldots \) denote independent and identically distributed random variables. Lemma 13, 14 majorly serve to answer one simple question: Given \( z \in \Theta' \),
can we find out the asymptotical distribution of \((e_{S_i^+}, 1 \leq i \leq c_0)\), which is supposed to be that of \((\bar{e}_i, 1 \leq i \leq c_0)\)? The discussion of such a distribution is quite subtle than it appears to be. At first glance we notice the definition of \(e_{S_i^+}\) is not well-defined: Given any \(z > 0\), it’s always possible that there is no sample fall on the right side of \(z\); \(S_i^+\)’s value on such an event will be problematic. We circumvent this problem by considering a small interval \([z, z_n]\) for some \(z_n > z\) that depends on sample size \(n\) and the sample points falling on which. Then the distribution of interest is \((e_{S_i^+}, 1 \leq i \leq q^{(z, z_n)})\), where we can find a \(z_n\) large enough such that \(q^{(z, z_n)} \geq c_0\) with probability almost 1. To see the difficulty for making the connection between \((e_{S_i^+}, 1 \leq i \leq q^{(z, z_n)})\) and \((\bar{e}_i, 1 \leq i \leq q^{(z, z_n)})\), we compare these distributions with \((e_{T_i^{(z, z_n)}}, 1 \leq i \leq q^{(z, z_n)})\). By Theorem 4.1.3 in Durrett, or Lemma 20, the distribution of \((e_{T_i^{(z, z_n)}}, 1 \leq i \leq c_0)\) is iid to that of \((\bar{e}_i, 1 \leq i \leq c_0)\). However, \((e_{T_i^{(z, z_n)}}, 1 \leq i \leq q^{(z, z_n)})\) possesses no easy analytical property at all; the involvement of \(S_i^+\)’s, functioning as a rearrangement on \(e_{T_i^{(z, z_n)}+1}\)’s, furtherly make it impossible for establishing this connection through any traditional way. Nevertheless, we propose a two steps method to deal with this problem; Lemma 13 states that if we have \(z_n\) small enough, then the probability of the event that two \(x_i\)’s falling inside \([z, z_n]\) within any \(c \log n\) consecutive trials can be well-controlled. Lemma 14 consider the special stopping times \(T_i^{(z, z_n)}\)’s such that for any \(x_i\)’s falling inside \([z, z_n]\), the stopping time would ’wait’ for another \(c \log n\) trials; these stopping times are designed in a way so by (43), we can treat each pair of \((x_{T_i^{(z, z_n)}}, e_{T_i^{(z, z_n)}+1})\) as asymptotically iid random vectors with distribution \((x_{T_i^{(z, z_n)}}, e_{T_i^{(z, z_n)}+1})\). Note that by the Theorem 4.1.3 in Durrett, \(e_{T_i^{(z, z_n)}+1} \perp x_{T_i^{(z, z_n)}}\); hence, the rearrangement on \(x_{T_i^{(z, z_n)}}\)’s can no longer affect \(e_{T_i^{(z, z_n)}+1}\)’s. Since Lemma 14 is more involved than this simple case, the language devloped there is slightly more complicated.

**Lemma 13, 14**

Given \(z \in \Theta^1, \Delta > 0, 0 < b < 1 - \Delta, 3b > 2, \theta = [z, z + n^{-b}]\),

**Lemma 13.** There exist \(C > 0\) such that for all large \(n\),

\[
P\left(T_1^\theta \neq T_i^\theta, \text{ for some } 1 < i \leq q^\theta \right) \leq C(\log n)^2 n^{2-3b}.
\]

**Proof of Lemma 13.** Let \(y_i \sim_{i.i.d} \text{Uniform}(z, z + M^{-1})\), \(i = 1, \ldots, n\). We define the \(y_i\)’s stopping times, which are analogous to those of \(x_i\)’s,

\[
y_T^\theta = \inf\{k : y_k \in \theta, k > 0\},
\]

for \(i > 1,\)

\[
y_T^\theta = \inf\{k : y_k \in \theta, y_{T_{i-1}}^\theta < k < \infty\},
\]

\[
y_T^\theta = \inf\{k : y_k \in \theta, c \log n + y_{T_{i-1}}^\theta < k < \infty\},
\]

42
\[ yq^\theta = \sup\{k : yT_k^\theta \leq n, k \geq 1\}. \]

\(S_y\) counts the number of indexes \(i\)'s distants that are shorter than \(c \log n\), and \(y_i\)'s fall inside \(\theta\),

\[ S_y = |\{k + 1 : yT_{k+1}^\theta - yT_k^\theta \leq c \log n, 2 \leq k + 1 \leq yq^\theta\}|. \]

Claim:

\[ P(T_i^\theta \neq T_i'^\theta, \text{ for some } 1 < i \leq q^\theta) \leq P(yT_i^\theta \neq yT_i'^\theta, \text{ for some } 1 < i \leq yq^\theta) + C(\log n)^2 n^{2 - 3b}. \]

Proof of the claim. Define

\[ q_i^0 = y_i, i = 1, \ldots, n, \]

\[ q_i^1 = x_i, i = 1, \ldots, n, \]

\[ q_i^j = \begin{cases} x_i, 1 \leq i \leq j, \\ y_i, j + 1 \leq i \leq n, \end{cases} \quad j = 1, \ldots, n - 1. \]

\[ A_i^k = \{q_i^k \in \theta, q_{i+j}^k \in \theta \text{ for some } 1 \leq j \leq c \log n\}, \]

\[ B_i^k = \{q_i^k \in \theta, q_{i+j}^k \notin \theta \text{ for all } 1 \leq j \leq c \log n\}, \]

\[ C_i^k = \{q_i^k \notin \theta\}. \]

\[ \{T_i^\theta = T_i'^\theta \text{ for some } 1 < i \leq q^\theta\} = \{A_i^n \cup_{i=2}^{n-1} (A_i^n \cap_{s=1}^{i-1} (B_i^s \cup C_i^s))\}. \]

Proof. To prove the above equation, we can view the LHS event as \{ For some \(i, j\) with \(0 < j - i \leq c \log n, j \leq n, x_i, x_j \in \theta\}. And then we construct \(J\) as the first time these \(x_i, x_j \in \theta\), i.e.

\[ J = \inf\{k : x_k \in \theta, x_i \in \theta \text{ for some } k < i \leq (k + c \log n) \wedge n\}. \]

We have

\[ \{T_i^\theta = T_i'^\theta \text{ for some } 1 < i \leq q^\theta\} = \bigcup_{i=1}^{n-1}\{J = i\}. \]

Notice that \(\{J = 1\} = A_i^n\) and \(\{J = i\} = A_i^n \cap_{s=1}^{i-1} (B_i^s \cup C_i^s), n > i \geq 2.\)

We claim (53) to (56),

\[ P(A_i^k) \leq P(A_i^{k-1}) \text{ for } 1 \leq k \leq n \quad (53) \]

For \(2 \leq i \leq k, 2 \leq k \leq n\)

\[ P(A_i^k \cap_{s=1}^{i-1} (B_i^s \cup C_i^k)) \leq P(A_i^{k-1} \cap_{s=1}^{i-1} (B_i^{k-1} \cup C_i^{k-1})) \quad (54) \]

For \(k < i \leq (k + c \log n) \wedge n, 1 \leq k \leq n,\)

\[ P(\cap_{s=1}^{i-c \log n} (B_i^s \cup C_i^k) \cap_{s=i-c \log n+1}^{i-1} C_s^k) \leq P(\cap_{s=1}^{i-c \log n} (B_i^s \cup C_s^{k-1}) \cap_{s=i-c \log n+1}^{i-1} C_s^{k-1}) + \frac{M(2c \log n + 1)}{n^b}. \quad (55) \]
For $k + c \log n < i \leq n, 1 \leq k \leq n - c \log n$,

$$P(\cap_{s=1}^{i-c\log n} (B_s^k \cup C_s^k) \cap_{s=1}^{i-1} C_s^k)$$

$$\leq P(\cap_{s=1}^{i-c\log n} (B_s^{k-1} \cup C_s^{k-1}) \cap_{s=1}^{i-1} C_s^{k-1}) + \frac{M2c\log n}{n^b}. \quad (56)$$

In the cases of (55), (56), $A_i^k = A_i^{k-1}$ are independent of

$$\cap_{s=1}^{i-c\log n} (B_s^k \cup C_s^k) \cap_{s=1}^{i-1} C_s^k,$$

$$\cap_{s=1}^{i-c\log n} (B_s^{k-1} \cup C_s^{k-1}) \cap_{s=1}^{i-1} C_s^{k-1},$$

respectively. By these independent relations, (55), (56), we have for $k < i \leq (k + c \log n) \land n, 1 \leq k \leq n$, or $k + c \log n < i \leq n, 1 \leq k \leq n - c \log n$,

$$P(\cap_{s=1}^{i-c\log n} (B_s^k \cup C_s^k))$$

$$= P(\cap_{s=1}^{i-1} (B_s^k \cup C_s^k) \cap_{i-c\log n+1}^i C_s^k)$$

$$= P(\cap_{s=1}^{i-1} (B_s^k \cup C_s^k)) \cap_{i-c\log n+1}^i C_s^k$$

$$\leq P(A_i^{k-1}) \left\{ P(\cap_{s=1}^{i-1} (B_s^{k-1} \cup C_s^{k-1}) \cap_{i-c\log n+1}^i C_s^{k-1}) + \frac{M(2c\log n + 1)}{n^b} \right\}$$

$$\leq P(A_i^{k-1} \cap_{s=1}^{i-1} (B_s^{k-1} \cup C_s^{k-1}) \cap_{s=i-c\log n+1}^{i-1} C_s^{k-1}) + \frac{M(2c\log n + 1)}{n^b} \times \frac{M^2c\log n}{n^{2b}}$$

$$= P(A_i^{k-1} \cap_{s=1}^{i-1} (B_s^{k-1} \cup C_s^{k-1})) + \frac{M(2c\log n + 1)}{n^b} \times \frac{M^2c\log n}{n^{2b}}. \quad (57)$$

By repeating using (53), (54), (57) $n - 1$ times, we have

$$P(T_i^q = T_i^q \text{ for some } 1 < i \leq q^q) = P(\{A_1^q, A_1^{i-1} (A_i^{n} \cup C_s^m)\})$$

$$\leq P(A_1^q \cap_{i=1}^{q^q} (A_1^i \cap_{s=1}^{i-1} (B_s^k \cup C_s^m))) + C(\log n)^2 n^{2-3b}$$

$$= P(T_i^q = yT_i^q \text{ for some } 1 < i \leq q^q) + C(\log n)^2 n^{2-3b}$$

$$= P(S_y > 0) + C(\log n)^2 n^{2-3b}.$$

For the claims (53) to (56), (53) is obvious. In the following we prove (54) to (55).

**Proof of (54).** We notice that since $i \leq k, \cap_{s=1}^{i-1} (B_s^k \cup C_s^k) = \cap_{s=1}^{i-1} (B_s^{k-1} \cup C_s^{k-1})$. By the construction of $A_i^k, A_i^{k-1}$, the proof is finished. \[\blacksquare\]

**Proof of (55).** For $k < i \leq (k + c \log n) \land n, 1 \leq k \leq n$,

$$P(\cap_{s=1}^{i-c\log n} (B_s^k \cup C_s^k) \cap_{s=i-c\log n+1}^{i-1} C_s^k) \leq P(\cap_{s=1}^{i-c\log n} (B_s^k \cup C_s^k) \cap_{s=i-c\log n+1}^{i-1} C_s^k) \quad (58)$$

Define

$$E_k \equiv \cap_{s=1}^{i-c\log n} (B_s^k \cup C_s^k) \cap_{s=i-c\log n+1}^{i-1} C_s^k,$$
so the RHS of (58) is equivalent to

$$P(E_k \cap \bigcap_{s=-c \log n; s \neq 0} (C_k^k)^c) + P\left(\bigcup_{s=-c \log n; s \neq 0} (E_k \cap (C_k^k)^c)\right).$$

(59)

In the first event above, $q^k_{k+s}, |s| \leq c \log n, s \neq 0$ will not fall inside $\theta$, and $q^k_s$ is ”a free random variable”. This is also true for $q^k_{k+s}$ with $E_k$ being replaced with $E_{k-1}$. Besides, $y_s'$s are independent of $x_s'$s, so

$$P(E_k \cap \bigcap_{s=-c \log n; s \neq 0} (E_k \cap (C_k^k)^c)) = P(E_k \cap \bigcap_{s=-c \log n; s \neq 0} (C_k^k)^c).$$

(60)

Also

$$P\left(\bigcup_{s=-c \log n; s \neq 0} (E_k \cap (C_k^k)^c)\right) \leq P\left(\bigcup_{s=-c \log n; s \neq 0} (E_{k-1} \cap (C_k^k)^c)\right) + \frac{2\bar{M}c \log n}{n^b}. \quad (61)$$

$$P\left(\bigcap_{s=1}^{i-1} (B_s^{k-1} \cup C_s^{k-1}) \bigcap_{s=i}^{i-1} (C_s^{k-1}) \bigcap_{s=i-1}^{i-1} (C_s^{k-1}) \bigcap_{s=i}^{i-1} (C_s^{k-1}) + \frac{\bar{M}}{n^b}. \quad (62)$$

By (58) to (62), we have finished the proof. 

Proof of (56). For $k + c \log n < i \leq n, 1 \leq k \leq n - c \log n$, almost the same argument as in the proof of (55) applies to this case as well. Define

$$E_k \equiv \bigcap_{s=1}^{i-1} (B_s^{k} \cup C_s^{k}) \bigcap_{s=i}^{i-1} (C_s^{k}).$$

(56)

$$P\left(\bigcap_{s=1}^{i-1} (B_s^{k} \cup C_s^{k}) \bigcap_{s=i}^{i-1} (C_s^{k})\right) = P(E_k \cap \bigcap_{s=-c \log n; s \neq 0} (C_k^k)^c).$$

(63)

$$P(E_k \cap \bigcap_{s=-c \log n; s \neq 0} (E_k \cap (C_k^k)^c)) \leq P\left(\bigcup_{s=-c \log n; s \neq 0} (E_{k-1} \cap (C_k^k)^c)\right) + \frac{2\bar{M}c \log n}{n^b}. \quad (65)$$

By (63) to (65), we have finished the proof. 

$P(S_y > 0)$ is smaller than, for $c_2 > c_1 > 0$,

$$\sum_{j=2}^{n-1} \sum_{i=1}^{j-1} P(S_y = i; y_q^{\theta} = j) \leq \left\{ \sum_{j \geq 1}^{c_2 n^{1-b}} \sum_{i=1}^{j-1} P(S_y = i; y_q^{\theta} = j) \right\} + P(y_q^{\theta} \geq c_2 n^{1-b} \text{ or } y_q^{\theta} \leq c_1 n^{1-b}). \quad (66)$$

To bound (66), we note that there is $c_3, N > 0$ such that for all $n > N$, all $0 \leq j \leq c_2 n^{1-b}$,

$$\frac{n - j c \log n}{n - 2 j c \log n} = 1 + \frac{j c \log n}{n - 2 j c \log n} \leq 1 + c_3 n^{-b}. \quad (67)$$

45
There exists some \( c_4 > 0 \) such that for all large \( n \),
\[
\prod_{j=0}^{c_2 n^{1-b}} \left( \frac{n - j c \log n}{n - 2 j c \log n} \right) \leq \left( 1 + c_3 n^{-b} \right) c_2 n^{1-b} \leq \exp \left( 2 c_4 n^{1-2b} \right) \cdot \exp \left( \ln(...) \right)
\]

By the above calculation, for all \( j \leq c_2 n^{1-b} \), there is \( N > 0 \) such that for all \( n > N \),
\[
\frac{P(S_y = 1; yq = j)}{P(S_y = 0; yq = j)} \text{ is smaller than } \frac{2 j c \log n}{n \times (n - j c \log n) \times \cdots \times (n - j c \log n)} \leq 2 c_3 n^{-b} \exp \left( 2 c_4 n^{1-2b} \right), \quad (67)
\]
where \((n^{-b} \tilde{M}) (1 - n^{-b} \tilde{M})^{n-j(j)!-1}\) are canlcelled on both sides of the fraction. Note that the reason we divide the numerator by \( P(S_y = 0; yq = j) \) is to cancel out the the permutation term, which is somehow difficult to approximate. The fact that
\[
1 \geq P(S_y = 0; yq = j) \geq P(S_y = 1; yq = j)
\]
make the denominator a good helper for this purpose.

Given \( \Delta > 0 \), \( j \leq c_2 n^{1-b} \), there is \( N > 0 \) such that for all \( n > N \),
\[
\sum_{i=1}^{j-1} \frac{P(S_y = i; yq^\theta = j)}{P(S_y = 0; yq^\theta = j)} \leq \sum_{i=1}^{j-1} \frac{P(S_y = 1; yq^\theta = j)}{P(S_y = 0; yq^\theta = j)} \left( \frac{2 j c \log n}{n - j c \log n} \right)^{i-1} \leq c_5 \frac{P(S_y = 1; yq^\theta = j)}{P(S_y = 0; yq^\theta = j)} \quad \text{ (68)}
\]

Then by proper choice of the constant coefficients, Lemma 12, (67), (68), \( 1 < 2b \), we have \( \delta, c_2, c_4, c_5, C > 0 \) such that for all large \( n \),
\[
(66) \leq C (\log n)^2 n^{2-3b} + c_2 c_5 n^{1-b} n^{-b} \times \exp (2 c_4 n^{1-2b}) + \exp (-c n^4) \leq C (\log n)^2 n^{2-3b}.
\]

**Notation for Lemma 14**

Let \( \theta_1 = [r - n^{-b}, r], \theta_2 = [r, r + n^{-b}], \theta = [r - n^{-b}, r + n^{-b}] \),
\[
\bar{T}_i^{\theta_1} = \inf \{ T_j^\theta : x_{T_j^\theta} \in \theta_k, j \geq 1 \}, k = 1, 2, \quad \bar{T}_i^{\theta_2} = \inf \{ T_j^\theta : x_{T_j^\theta} \in \theta_k, T_j^\theta > \bar{T}_i^{\theta_1} \}, i > 1, k = 1, 2, \\
\bar{q}^{\theta_1} = \sup \{ i : \bar{T}_i^{\theta_1} \leq n, i \geq 1 \}, k = 1, 2,
\]
For the \( x_i, i = 1, \ldots, n \) with indexes being captured by \( T_i^{\theta_k} \), we set another random indexes in order to represent these \( x_i \)'s in an ordered fashion from \( r \) to upward, \( s_{i,2} \)'s , or \( s_{i,1} \)'s downward,
\[
\begin{align*}
  s_{1,1} &= \arg \max_{\bar{T}_k^{\theta_1} : x_{\bar{T}_k^{\theta_1}} < r, 0 \leq \bar{q}^{\theta_1} } x_{\bar{T}_k^{\theta_1}}, \quad s_{1,1} &= \arg \max_{\bar{T}_k^{\theta_1} : x_{\bar{T}_k^{\theta_1}} < x_{s_{i-1,1}} \leq \bar{q}^{\theta_1} } x_{\bar{T}_k^{\theta_1}} \geq i > 2,
\end{align*}
\]

46
Lemma 14. Define

$$E = \{ \times_{i=1}^{\infty} B_1^i, \times_{i=1}^{\infty} B_2^i \mid \times_{i=1}^{\infty} B_1^i \in \times_{i=1}^{\infty} R^{2i}, t = 1, 2 \}$$

If $E|e_1|^t < C$, then for each $n$,

$$\sup_{\{ \times_{i=1}^{\infty} A_1^i, \times_{i=1}^{\infty} A_2^i \}} \left| P \left( \bigcap_{k=1}^{\infty} \left\{ (x_{s_{i,k}}, e_{s_{i,k}+1}), 1 \leq i \leq q^{'q_{th}} \right\} \right) \cap W \right)$$

$$- P \left( \bigcap_{k=1}^{\infty} \left\{ (x_{s_{i,k}}, \bar{e}_{i,k}), 1 \leq i \leq q^{'q_{th}} \right\} \right) \cap W \right) \leq C(n^t \rho^\alpha \log n + n^{-2q+1}),$$

Proof of Lemma 14. Since we have continuous random variables, we consider only open subset $A_q$ in $E$. If we notice $q^{'q} = q^{'q_1} + q^{'q_2}$ and the permutation and position of order
statistics can be fully captured by a Borel sets, then for any \( \times_{i=1}^{\infty}A_i \in \times_{i=1}^{\infty}\mathbb{R}^{2i} \), \( A_i \)'s are open subsets, such that

\[
\cap_{k=1}^{\infty} \left\{ (x_{s_i,k}, e_{s_i,k+1}), 1 \leq i \leq \bar{q}^\theta_k \right\} \cap \mathcal{W} = \left\{ ((x_{T_i^\theta}, e_{T_i^\theta+1}), 1 \leq i \leq q^\theta) \in A_{q^\theta} \right\} \cap \mathcal{W},
\]

and

\[
P(\cap_{k=1}^{\infty} \left\{ (x_{s_i,k}, \bar{e}_{i,k}), 1 \leq i \leq \bar{q}^\theta_k \right\} \cap \mathcal{W})
= P(\{(x_{T_i^\theta}, \bar{e}_i), 1 \leq i \leq q^\theta \in A_{q^\theta} \} \cap \mathcal{W})).
\]

We control the probability in the second statement because the spaces of \( \bar{e}_i \) and \( \bar{e}_{i,k} \) are distinguished. To prove Lemma 14, it suffices to prove, for \( A_i \)'s are open subset,

\[
sup_{\times_{i=1}^{\infty}A_i \in \times_{i=1}^{\infty}\mathbb{R}^{2i}} \left| P(\{(x_{T_i^\theta}, e_{T_i^\theta+1}), 1 \leq i \leq q^\theta \in A_{q^\theta} \} \cap \mathcal{W}) - P(\{(x_{T_i^\theta}, \bar{e}_i), 1 \leq i \leq q^\theta \in A_{q^\theta} \} \cap \mathcal{W}) \right| \leq C(n^4 \rho c \log n + n^{-q}).
\]

Before we prove (69), we first state, with proof provided below the remark, the decreasing in \( n \) upper bound of the difference

\[
|P((x_1, \ldots, x_n) \in \gamma) - P((x_1, \ldots, x_{i+\log n}, x_{i+\log n+1}, \ldots, x_n) \in \gamma)| \leq (1 + n^2)M_0 \rho c \log n,
\]

where \( x_1, x_2 \)'s are two independent processes with distributions that are identical to those of \( x_i \)'s, and

\[
\gamma, i = 1, 2 \text{ are open sets with }
\gamma_1 \subset \mathbb{R}^{i-1} \times [-n^2, n^2], \gamma_2 \subset \mathbb{R}^{n-i-\log n}
\]

The closeness of these probabilities says after another \( c \log n \) (from \( x_{i+\log n} \), to \( x_{i+\log n} \)) free trials in the SETAR process, the whole process will asymptotically return to a new, independently identically distributed process. Note that due to the definition of \( T_i^\theta \), up to some modification, this is almost what we need for (69). In the following we deal with the subset \( \gamma \).

Let \( K > 0 \). By the idea Concentration in the following, we can bring (43) into our argument. We say \( \cup_{k=1}^{K} A_k \) is concentrating on \( i \)-th coordinate if

\[
A_k = [a_1^k, b_1^k] \times \cdots \times A_{k,i} \times \cdots \times [a_n^k, b_n^k],
\]

where \( A_{k,i} \subset \mathbb{R}^1 \) is a union of disjoint half-open rectangles, and for all \( 1 \leq t, s \leq K, t \neq s \), there is \( 1 \leq j \leq n, j \neq i \) such that

\[
[a_j^t, b_j^t] \cap [a_j^s, b_j^s] = \emptyset.
\]

48
This says when \( i \)-th coordinate is taken out from \( A_k \), the rest rectangles in \( A_k \)'s are disjoint. If \( \bigcup_{k=1}^{K} A_k \) is not concentrating on \( i \)-th coordinate but each \( A_k \) is the union of rectangles, i.e., \( A_k = \bigcup_{i=1}^{c_k} [a_{ij}^k, b_{ij}^k] \subset \mathbb{R}^n \), we can find \( c_K > 0 \) depending on \( K \) and \( B_k \subset \mathbb{R}^n, k = 1, \ldots, c_K \), such that

\[
\bigcup_{k=1}^{K} A_k = \bigcup_{k=1}^{c_K} B_k,
\]

and \( \bigcup_{k=1}^{c_K} B_k \) is concentrating on \( i \)-th coordinate.

Since \( (x_1, e_2, \ldots, e_k) \mapsto (x_1, x_2, \ldots, x_k) \) is an \( 1 \rightarrow 1 \) onto mapping, there is a \( \gamma^*_2 \in \mathbb{R}^{n-c \log n-i} \) such that \( \{(x_1, \ldots, x_n) \in \gamma \} = \{(x_1, \ldots, x_i+c \log n+1, e_{i+1} \log n+2, \ldots, e_n) \in \gamma_1 \times \mathbb{R}^{c \log n} \times \gamma^*_2 \equiv \gamma^* \} \). The transformation from \( x_i \)'s to \( e_i \)'s can help us tackling down the dependence of \( x_i \)'s happening on the \((i+c \log n+1)\)-th coordinate.

Given any such \( \gamma \), we have

\[
P((x_1^1, \ldots, x_n^1) \in \gamma) - P((x_1^1, \ldots, x_i+c \log n+1, x_{i+1}^2, x_i, \ldots, x_n^2) \in \gamma) = \left\{ \begin{array}{l}
\int_{\gamma} \pi(z_1) \ldots p(z_{n-1}, z_n) dz_1 \ldots dz_n \\
\quad - \int_{\gamma} \pi(z_1) \ldots p(z_i+c \log n-1, z_i+c \log n) \pi(z_i+c \log n+1) \ldots p(z_{n-1}, z_n) dz_1 \ldots dz_n \end{array} \right. = \int_{\gamma_1 \times \gamma_2} \pi(z_1) \ldots p(z_i-1, z_i) \left(p^{c \log n+1}(z_i, z_i+c \log n+1) - \pi(z_i+c \log n+1) \right) \ldots \\
\times p(z_{n-1}, z_n) dz_1 \ldots dz_i dz_{i+c \log n+1} \ldots dz_n = \int_{\gamma_1 \times \gamma_2} \pi(z_1) \ldots p(z_i-1, z_i) \left(p^{c \log n+1}(z_i, z_i+c \log n+1) - \pi(z_i+c \log n+1) \right) \ldots \\
\times \prod_{j=i+c \log n+2}^{n} f_c(z_j) dz_1 \ldots dz_i dz_{i+c \log n+1} \ldots dz_n.
\]

Since \( \gamma_1 \times \gamma^*_2 \) is an open subset, we can express it as \( \bigcup_{k \geq 1} \gamma^*_1 \times \gamma^*_3 \times \gamma^*_5 \), in which \( \gamma^*_1 \times \gamma^*_3 \times \gamma^*_5 \) are union of disjoint half-open rectangles such that \( \gamma^*_k \subset \mathbb{R}^{i-1} \times [-n^2, n^2], \gamma^*_3 \subset \mathbb{R}^1, \gamma^*_5 \subset \mathbb{R}^{n-i-c \log n-1} \). Due to (43), \( \gamma^*_k \subset \mathbb{R}^{i-1} \times [-n^2, n^2] \), continuity of probability measure, the fact that given \( K > 0, \bigcup_{k=1}^{c_k} \gamma^*_1 \times \gamma^*_3 \times \gamma^*_5 \) can be expressed as \( \bigcup_{k=1}^{c_K} \eta^*_1 \times \eta^*_3 \times \eta^*_5 \), which is concentrating in \( i \)-th coordinate if we view \( \gamma^*_1 \times \gamma^*_3 \times \gamma^*_5 \) as \( A_k \) and \( \eta^*_1 \times \eta^*_3 \times \eta^*_5 \) as \( B_k \), and independence of \( e_i \),

\[
(70) = \lim_{K \rightarrow \infty} \sum_{K \geq j \geq 1} \int_{\gamma^*_1 \times \gamma^*_3 \times \gamma^*_5} \pi(z_1) \ldots p^{c \log n+1}(z_i, z_i+c \log n+1) \ldots \\
\times \prod_{j=i+c \log n+2}^{n} f_c(z_j) dz_1 \ldots dz_i dz_{i+c \log n+1} \ldots dz_n = \lim_{K \rightarrow \infty} \sum_{c_k \geq j \geq 1} \int_{\eta^*_1 \times \eta^*_3 \times \eta^*_5} \pi(z_1) \ldots p^{c \log n+1}(z_i, z_i+c \log n+1) \ldots \\
\times \prod_{j=i+c \log n+2}^{n} f_c(z_j) dz_1 \ldots dz_i dz_{i+c \log n+1} \ldots dz_n
\]
\[
\leq \lim_{K \to \infty} \sum_{c, h \geq 1} \int_{y_i^1, y_i^2} \cdots \int_{y_i^n} p(z_{i-1}, z_i) \left\{ (1 + \|[-n^2, n^2]| |) C \rho^{c \log n} \right\} \\
\times \prod_{j=i+c \log n+2}^{n} f_c(z_j) dz_1 \cdots dz_i dz_{i+c \log n+2} \cdots dz_n \\
\leq (1 + \|[-n^2, n^2]| |) C \rho^{c \log n}
\]

Simply using the rectangles on \(x_i\) still cause dependence in \(p(z_{i+c \log n+1}, z_{i+c \log n+2})\). Then we use this upper bound to prove (69). Let \(\{e_i^k\}, k = 1, \ldots, n - c \log n\) be i.i.d with \(e_1^1 \equiv e_1, \{y_i^k\}_{i=1}^{n}, k = 1, \ldots, n - c \log n\) such that

\[
y_i^1 = x_1, \\
y_i^k = \begin{cases} a_1 y_{i-1}^k + e_i^k, y_{i-1}^k \leq r, & i > 1, \\
a_2 y_{i-1}^k + e_i^k, y_{i-1}^k > r, & i > 1. 
\end{cases}
\]

and for the first \(k\) times \(y_i^k \in B_{n-k}(r)\),

\[
y_i^{k+c \log n+1} \overset{d}{=} x_1, y_i^{k+c \log n+1} \perp y_i^{k-j}, j \geq 0,
\]

and \(y_i^{k+c \log n+s}, s \geq 1\) follows the same structure. Note that \(e_i^{k+c \log n+1}\) does not appear in the process.

Define the stopping times, random indexes and the corresponding event \(W\) on \(\{y_i^k\}\) as \(T_j^{\theta, k}, q_j^{\theta, k}, W_j^k\); respectively. To simplify the context, we drop off the superscript \(k\).

Given \(x_{i}^{\infty} A_i \in \times_{i=1}^{\infty} [-n^2, n^2]^{2i}\), \(\Gamma = \{w_1, \ldots, w_m\} \subset \{1, \ldots, n\}\), define \(A_i^{\Gamma}, A_i^{\Gamma, k}, 1 \leq i, 1 \leq k \leq n - c \log n\) such that

\[
\{(x_1, \ldots, x_n) \in A_m^{\Gamma} \} = \{T_i^{\theta} = w_i, i = 1, \ldots, m; q_i^{\theta} = m; (x_{T_i^{\theta}}, e_{T_i^{\theta}+1}^{\theta}), 1 \leq i \leq m\} \in A_m
\]

So is \(A_i^{\Gamma, k}\). Hence, for each \(m\),

\[
\{q_i^{\theta} = m, (x_{T_i^{\theta}}, e_{T_i^{\theta}+1}^{\theta}), 1 \leq i \leq m\} \in A_m \}
\]

so is \(A_i^{\Gamma, k}\). \(y_i^k\)'s do not always have control over \(e_i^{k+c \log n+1}\)'s since \(y_i^{k+c \log n+1}\)'s might be an independent random variable, for large \(n\), but \(y_i^k\)'s always have control over \(e_i^{k+c \log n+1}\)'s.

Thus, for any \(\Gamma, 1 \leq k \leq n - c \log n, 1 \leq i\), we have \(A_i^{\Gamma}, A_i^{\Gamma, k}\) such that

\[
A_i^{\Gamma} = A_i^{\Gamma, k}
\]
For $k = 1$, $\times_{i=1}^{\infty}A_i \in \times_{i=1}^{\infty}\mathbb{R}^{2^i}$,
\[
|P(((x_{t_{i}^j\theta}, e_{T_{i}^j\theta+1}), 1 \leq i \leq q^{\theta}) \in A_{q^{\theta}}) - P(((y_{t_{i}^j\theta}, e_{T_{i}^j\theta+1}), 1 \leq i \leq q^{\theta}) \in A_{q^{\theta}})|
\leq \sum_{m=1}^{n-c\log n} \sum_{i} \int_{A_m^r} Edz_1 \ldots dz_n
\leq \sum_{m=1}^{n-c\log n} \sum_{i} \int_{A_m^r} Edz_1 \ldots dz_n
= \sum_{i=1}^{n-c\log n} \int_{A_m^r} Edz_1 \ldots dz_n
\]
where the integrand $E$ in the above equations, which has a rather long expression so we put it in below, is
\[
E \equiv \pi(z_1) \times \ldots \times \left| p^{c\log n+1}(z_{w_1}, z_{w_1+c\log n+1}) - \pi(z_{w_1+c\log n+1}) \right| \times \cdots \times p(z_{n-1}, z_n)
\]
For distinguished $\Gamma$, $A_m^\Gamma$ are disjoint subsets, so we have the last equality. Apply the argument in (70) and notice that $P(W) \leq \exp(-cnA)$, $\times_{i}\Gamma = \{w_1, \ldots, w_m; w_1=i\};\{m \leq n; A_m^\Gamma; i = 1, \ldots, n - c\log n\}$ satisfies the requirement for $\gamma$ due to the construction of $T_{i, j}^{\theta}$, and the inner summation is smaller than $(1 + n^2)\rho^{c\log n}C$, so
\[
|P(((x_{t_{i}^j\theta}, e_{T_{i}^j\theta+1}, 1 \leq i \leq q^{\theta}) \in A_{q^{\theta}}, W) - P(((y_{t_{i}^j\theta}, e_{T_{i}^j\theta+1}, 1 \leq i \leq q^{\theta}) \in A_{q^{\theta}}, W)|
\leq n \times (1 + n^2)\rho^{c\log n}C
\]
To obtain the analogous results for $y_{i,k}^1$'s and $y_{i,k}^1$'s, we can fix $w_k$ instead of $w_1$. Repeating this argument at most $n - c\log n$ times, we have
\[
|P(((x_{t_{i}^j\theta}, e_{T_{i}^j\theta+1}), 1 \leq i \leq q^{\theta}) \in A_{q^{\theta}}, W) - P(((y_{t_{i}^j\theta}, e_{T_{i}^j\theta+1}), 1 \leq i \leq q^{\theta}) \in A_{q^{\theta}}, W)|
\leq n^2 \times (1 + n^2)\rho^{c\log n}C,
\]
(71)
Because $e_{T_{i}^j\theta+1}$ is independent of the stopping times when the indexes stay inside the sample size $n$,
\[
P(((y_{t_{i}^j\theta}, e_{T_{i}^j\theta+1}), 1 \leq i \leq q^{\theta}) \in A_{q^{\theta}}, W) = P(((y_{t_{i}^j\theta}, e_{T_{i}^j\theta+1}), 1 \leq i \leq q^{\theta}) \in A_{q^{\theta}}, W),
\]
(71) is equivalent to
\[
|P(((x_{t_{i}^j\theta}, e_{T_{i}^j\theta+1}), 1 \leq i \leq q^{\theta}) \in A_{q^{\theta}}, W) - P(((y_{t_{i}^j\theta}, e_{T_{i}^j\theta+1}), 1 \leq i \leq q^{\theta}) \in A_{q^{\theta}}, W)|
\leq n^2 \times (1 + n^2)\rho^{c\log n}C,
\]
(72)
By the independence of $\bar{e}_i$ and (71),
\[ |P\left(((x_{T_i^g}, \bar{e}_i), 1 \leq i \leq q^g) \in A_{q^g}; \mathcal{W}\right) - P\left(((y_{T_i^g}^{n-c\log n}, \bar{e}_i), 1 \leq i \leq q^g) \in A_{q^g}; \mathcal{W}\right) | \leq \sup_{x \in \times_{i=1}^\infty A_i \times \times_{i=1}^\infty \mathbb{R}^t} \left| P\left(((x_{T_i^g}), 1 \leq i \leq q^g) \in A_{q^g}; \mathcal{W}\right) - P\left(((y_{T_i^g}^{n-c\log n}), 1 \leq i \leq q^g) \in A_{q^g}; \mathcal{W}\right) \right| \leq n^2 \times (1 + n^2)\rho^{c\log n} C, \] (73)

The tail event can be bounded by $\theta \subset \Theta'$ for all large $n$. (The stopping time $x_{T_i^g} \notin \beta$ if $\beta \notin \Theta'$) and
\[ P\left(\sup_{i \leq n} |e_i| \geq n^2\right) \leq n^{-2q+1}E|e_1|^q \] (74)

By (71) to (74), we have (69), and hence finished the proof. 

\textbf{Corollary 15.} There exists $C > 0$ such that for any $n \geq g \geq t > 0$, $\beta_{1,g}, \beta_{2,g} \subset \mathbb{R}^g, \beta \subset \mathbb{R}$, we have
\[ \left| P\left\{n(x_{s,t} - r) \in \beta\right\} \cap \mathbb{R}^i \times \beta_{1,g} \times \mathbb{R}^{l-g}, i \geq g, \right| \]
\[ \left. - P\left\{n(x_{s,t} - r) \in \beta\right\} \cap \mathbb{R}^i \times \beta_{2,g} \times \mathbb{R}^{l-g}, i \geq g, \right| \leq C(n^4 \rho^{c\log n} + n^{-2q+1}), \]
\[ -g \leq t \leq 0 \text{ can be argued in the same way.} \]

\textbf{Proof of Corollary 15.} Given $n$, define
\[ A^1_t = \begin{cases} \mathbb{R}^i \times \beta_{1,g} \times \mathbb{R}^{l-g}, i \geq g, \\ \mathbb{R}^{2i}, i < g \end{cases} \]
\[ A^2_t = \begin{cases} \mathbb{R}^{l-1} \times \left(\frac{1}{n} \beta + r\right) \times \mathbb{R}^{l-t} \times \beta_{2,g} \times \mathbb{R}^{l-g}, i \geq g, \\ \mathbb{R}^{2i}, i < g \end{cases} \]

Notice that
\[ P\left(\cap_{k=1}^\infty \{(x_{s,k}, 1 \leq i \leq q^k, e_{s,k+1}, 1 \leq i \leq q^k) \in A_{q^k}^k, g \leq \min\{q^k, q^k\}\} \cap \mathcal{W} \cap \left\{g \leq \min\{q^k, q^k\}\right\} \right) \]
\[ = P\left\{n(x_{s,t} - r) \in \beta\right\} \cap \mathbb{R}^i \times \beta_{1,g} \times \mathbb{R}^{l-g}, i \geq g, \]
\[ P\left(\cap_{k=1}^\infty \{(x_{s,k}, 1 \leq i \leq q^k, e_{s,k}, 1 \leq i \leq q^k) \in A_{q^k}^k, g \leq \min\{q^k, q^k\}\} \cap \mathcal{W} \cap \left\{g \leq \min\{q^k, q^k\}\right\} \right) \]
\[ = P\left\{n(x_{s,t} - r) \in \beta\right\} \cap \mathbb{R}^i \times \beta_{2,g} \times \mathbb{R}^{l-g}, i \geq g, \]

Then by
\[ P\left\{g \leq \min\{q^k, q^k\}\right\} \leq \exp(-cn^\Delta) \]
and rearranging $((x_{s,k}, e_{s,k+1}), 1 \leq i \leq q^k)$ to $((x_{s,k+1}, 1 \leq i \leq q^k, e_{s,k+1}, 1 \leq i \leq q^k)$, we can apply Lemma 14 to finish the proof. 

52
Section M, Moment Bounds

In this section, we introduce some moment bounds for our own theoretical usage. Let 
\( z \in \Theta' \). \( S_i^z \)'s are defined in the introduction.

Let \( x_i^o = x_i 1_{F^c} + b_i 1_{F} \), \( \{ b_i \}_{i=1}^n \) are any deterministic real bounded sequence, and 
\( e_i^o = e_i 1_{F^c} \). We also define \( S_i^o \) for \( x_i^o \) in the same way we did for \( S_i \). Again, in order to 
save the notations problem, \( x_{S_i}^o \equiv x_{S_i}^o \), and the superscripts on \( S_i^z \) are dropped off.

The usage of \( \bar{F} \) is to such that the inverse moment of \( x_i^o \)'s is bounded. Although the 
choice of \( \bar{F} \) is arbitrary, we will always have \( \bar{F} \) being a rare event. For example, let 
\( \theta_1^i = [z, \infty], \theta_2^i = [-\infty, -\varepsilon] \), \( \bar{F} = \cup_{i=1,2} \{ \theta_i^q < \rho_2 n \} \).

By Corollary 8, Corollary 11, there exists \( c > 0 \), some \( \delta > 0 \) such that

\[
\sup_{z \in \Theta'} P(\bar{F}) \leq \exp(-cn^\delta).
\]

In the following we establish moment bound that is essential to analyze the SETAR process.

**Lemma 16.** 8 Given \( 1 > \Delta > 0 \). Let \( q \geq 1 \), \( E|e_1|^4 < \infty \). There exist constants \( C, c_1 \), 
whose values depend on \( \Delta, q \), such that for \( 0 < b < 1 - \Delta, z \in \Theta' \), \(-c_1n^{1-b} + 1 \leq k \leq 
\)
\( c_1n^{1-b}, \) all large \( n \),

\[
E \left| \sum_{i=1}^k x_{S_i}^o e_{S_i+1}^o \right|^q \leq Cn^{(1-b)^\frac{1}{2}}, k > 0,
\]

\[
E \left| \sum_{i=0}^{|k|} x_{S_i-1}^o e_{S_i-1+1}^o \right|^q \leq Cn^{(1-b)^\frac{1}{2}}, k \leq 0.
\]

**Proof of Lemma 16.** We assume \( k > 0 \). The moment bound for \( k \leq 0 \) can be proven in the same way.

\[
E|x_{S_i}^o e_{S_i+1}^o|^q \leq E^{1/2}|x_{S_i}^o|^2q E^{1/2}|e_{S_i+1}^o|^{2q}.
\] (75)

By Lemma 18(see below), there exists \( C > 0 \) such that for all \( 0 < i \leq c_1n^{1-b}, 1 - \Delta > 
\)
\( b > 0, z \in \Theta' \),

\[
E^{1/2}|e_{S_i+1}^o|^{2q} \leq n^{(1-b)^\frac{1}{2}} C.
\] (76)

There exist constants \( C > 0, c_1 > 0 \) such that for all \( c_1n^{1-b} \geq i > 0, 1 - \Delta > b > 0 \), 
\( z \in \Theta' \) so \( B_1(z) \subset \Theta'' \),

\[
E|x_{S_i}^o|^{2q} = E|x_{S_i}^o 1_{F^c} + b_i 1_{F}|^{2q} \leq \left\{ E^{1/2q}|x_{S_i}^o 1_{F^c}|^{2q} + C \right\}^{2q}
\]

\[
\leq \left\{ E^{1/2q} \left( \sup |\Theta'| + 1 \right) + \left( \sum_{i=1}^n |x_i| 1_{\{q^{a_1(v)} \leq c_1n \}} \right)^{2q} + C \right\}^{2q}
\]

\[
\leq C,
\] (77)

53
by triangular and Cauchy-Schwartz inequality, Lemma 12, and $E|x_1|^{4q} < \infty$. By (75) to (77),

$$E|x_{S_i}^{o}e_{S_i+1}^{o}|^q \leq n^{(1-b)\frac{q}{2}}C.$$ 

Hence for all $0 < k \leq c_1n^{1-b}$, $0 < b < 1 - \Delta$, $z \in \Theta'$,

$$E \left| n^{-(1-b)} \sum_{i=1}^{k} (x_{S_i}^{o}e_{S_i+1}^{o}) \right|^q \leq \left( n^{-(1-b)} \sum_{i=1}^{k} E^{1/q}|x_{S_i}^{o}e_{S_i+1}^{o}|^q \right)^q \leq Cn^{(1-b)\frac{q}{2}}.$$ 

**Corollary 17.** By picking $b = 1 - 2\Delta$ in Lemma 16, we see there exists $C$, whose value depends on $\Delta$, $q$, such that for $q \geq 1$, all $|k| \leq n^\Delta$, all large $n$,

$$E^{1/q} \left| n^{-3\Delta} \sum_{i=1}^{k} x_{S_i}^{o}e_{S_i+1}^{o} \right|^q \leq C, k > 0,$$

$$E^{1/q} \left| n^{-3\Delta} \sum_{i=0}^{k} x_{S_i}^{o}e_{S_i+1}^{o} \right|^q \leq C, k \leq 0.$$ 

By this, we have for $q \geq 1$, all $0 < b \leq 1 - 6\Delta$, $z \in \Theta'$, $|k| \leq n^\Delta$, all large $n$,

$$E^{1/q} \left| n^{-(1-b)\frac{q}{2}} \sum_{i=1}^{k} x_{S_i}^{o}e_{S_i+1}^{o} \right|^q \leq C, k > 0,$$

$$E^{1/q} \left| n^{-(1-b)\frac{q}{2}} \sum_{i=0}^{k} x_{S_i}^{o}e_{S_i+1}^{o} \right|^q \leq C, k \leq 0.$$ 

**Lemma 18.** Assume $E|e_1|^{2q} < \infty$. Given $\Delta > 0$. There exists $C, c_1$, whose value depend on $\Delta$, such that for $0 < b < 1 - \Delta$, $z \in \Theta'$, all large $n$,

$$E \left( \sup_{-c_1n^{1-b}+1 \leq i \leq c_1n^{1-b}} |e_{S_i}^{o}|^q \right) \leq Cn^{1-b}.$$ 

**Proof of Lemma 18.** We prove the case $i > 0$. Let $\theta_{z,b} = [z, z+n^{-b}]$. By Lemma 12, there exist $c_1, c_2, c$, whose value depend on $\Delta$, such that for $0 < b < 1 - \Delta$, $A_{2,z,b} = \{c_1n^{1-b} \leq q_{1}^{\theta_{z,b}} \leq c_2n^{1-b}\}e$, all large $n$,

$$\sup_{0 < b < 1 - \Delta z \in \Theta'} P(A_{2,z,b}) \leq \exp \left(-cn^\Delta \right).$$

Let $A_{1,z,b} = A_{2,z,b}$. Since $e_{i}^{o} = 0$ on $\bar{F}$,

$$\sup_{0 < i \leq c_1n^{1-b}} |e_{S_i}^{o}|^q \leq \sup_{0 < i \leq c_2n^{1-b}} |e_{T_i}^{o}\theta_{z,b}+1|^q \text{ on } A_{1,z,b},$$

and

$$\sup_{0 < i \leq c_1n^{1-b}} |e_{S_i}^{o}|^q \leq \sum_{i=1}^{n} |e_{i}|^q \text{ on } A_{2,z,b},$$

54
Combining these and $E|c_i|^{2q} < \infty$,

$$E \left( \sup_{0 < i \leq c_1 n^{1-b}} |e_{S_i+1}^0| q \right) \leq E \left( \sup_{0 < i \leq c_2 n^{1-b}} |e_{T_i}^{\theta_{z,b}}|^{q} 1_{A_{1,z,b}} + \sum_i^n |e_i|^{q} 1_{A_{2,z,b}} \right) \leq C n^{1-b},$$

by triangular and Cauchy-Schwartz inequality.

Define $z \in \Theta'$,

$$T_1^{\theta} = \inf \{ k : x_k^o \in \theta, k > 0 \}, T_{i+1}^{\theta} = \inf \{ k : x_k^o \in \theta, k > T_i^{\theta} \},$$

$$q^{\theta} = \sup \{ k : T_k^{\theta} \leq n \}.$$  

$$S_i^{\sup \theta_{z,b}} = \arg \min \left\{ x_k^o : n \geq k, x_k^o > \sup \theta_{z,b} \right\},$$

$$S_i^{\inf \theta_{z,b}} = \arg \min \left\{ x_k^o : n \geq k, x_k^o < \inf \theta_{z,b} \right\}.$$  

We drop the scripts and denote them as $T_i^\theta, S_i^{\sup \theta}, q^\theta$, respectively.

![Figure 4: Visualized random indexes. θ in the graphic can be θ_{z,b}.](image)

**Lemma 19.** Assume $E|c_i|^{2q} < \infty$. Given $0 < \Delta < 1$. There exists $C$, whose value depends on $\Delta$, $q$, such that for $0 < b \leq 1 - 6 \Delta$, $z \in \Theta'$, $\theta_{z,b} = [z, z + n^{-b}]$, $|k| \leq n^\Delta$, all large $n$,

$$E \left[ n^{-(1-b)\frac{1}{2}} \left( \sum_{i=1}^{q_{z,b}} x_{T_i}^o \theta_{z,b} e_{T_i}^o + \sum_{i=1}^{q_{z,b}} x_{S_i}^{\sup \theta_{z,b}} e_{S_i}^{\sup \theta_{z,b}} \right) \right]^{q} < C, k > 0,$$

$$E \left[ n^{-(1-b)\frac{1}{2}} \left( \sum_{i=1}^{q_{z,b}} x_{T_i}^o \theta_{z,b} e_{T_i}^o + \sum_{i=0}^{k} x_{S_i}^{\inf \theta_{z,b}} e_{S_i}^{\inf \theta_{z,b}} \right) \right]^{q} < C, k \leq 0.$$

**Proof of Lemma 19.** We prove the case $k > 0$. By Lemma 12, standard inequalities, Lemma 20, that fact that $x_{T_i}^o \theta_{z,b}$’s are bounded, there exist $C, c_2$, whose values depend on
\[ \Delta, \text{ such that } 0 < b < 1 - \Delta, \, z \in \Theta', \, \theta_{z,b} = [z, z + n^{-b}], \text{ all large } n, \]

\[ E^{1/q} \left| n^{-(1-b)^{1/2}} \sum_{i=1}^{q^{\theta_{z,b}}} X_i^{o \theta_{z,b}} e_i e_{T_i^{\theta_{z,b}+1}} \right|^q \]

\[ \leq E^{1/q} \left( \sup_{0 < j \leq c_2 n^{1-b}} \left| \frac{\sum_{i=1}^{j} X_i^{o \theta_{z,b}} e_i e_{T_i^{\theta_{z,b}+1}}}{n^{(1-b)^{1/2}}} \right|^q \right) \quad (78) \]

\[ + E^{1/2q} \left( \sup_{0 < j \leq n} \left| \sum_{i=1}^{j} X_i^{o \theta_{z,b}} e_i e_{T_i^{\theta_{z,b}+1}} \right| \right)^{2q} P^{1/2q} (q^{\theta_{z,b}} \geq c_2 n^{1-b}) \]

\[ \leq C, \]

For 0 < k \leq n^\Delta, z \in \Theta', 0 < b \leq 1 - 6\Delta, all large n, there exists C < \infty such that

\[ E \left| n^{-(1-b)^{1/2}} \sum_{i=1}^{q^{\theta_{z,b}}} X_i^{o \theta_{z,b}} e_i e_{T_i^{\theta_{z,b}+1}} \right|^q \]

\[ \leq \left( E^{1/q} \left| n^{-(1-b)^{1/2}} \sum_{i=1}^{q^{\theta_{z,b}}} X_i^{o \theta_{z,b}} e_i e_{T_i^{\theta_{z,b}+1}} \right|^q \right)^{1/q} + E^{1/2q} \left( \sup_{0 < j \leq n} \left| \sum_{i=1}^{j} X_i^{o \theta_{z,b}} e_i e_{T_i^{\theta_{z,b}+1}} \right| \right)^{2q} P^{1/2q} (q^{\theta_{z,b}} \geq c_2 n^{1-b}) \]

\[ \leq C. \]

by \( e_i^o = 0 \) on \( F \), (78) and Corollary 17. It is possible for sup \( \theta_{z,b} \neq \Theta' \), but this is not a problem if we restate \( z \in \Theta' \) to \( z \in \cap_{x \in \Theta'} B_x(n^{-b}). \]

**Lemma 20.** Given \( \theta \subset \mathbb{R}^1, j > 0 \), we have

\[ E \left( \sup_{k \leq j} \frac{\sum_{i=1}^{k} |X_i X_{T_i}|}{j^{1/2}} \right)^q \leq C E \sum_{i=1}^{j} X_i^2 j^{-1} \]

**Proof.** Denote \( T_i^q \) as \( T_i \). By theorem 4.1.3 in Durret with \( e_i = X_i \), we have

\( e_{T_i+1} \) is independent of \( \mathcal{F}_{T_i} = \sigma(e_1, \ldots, e_{T_i}) \), and \( e_{T_i+1} \sim \mathcal{D} e_1 \)

Hence \( \{e_{T_i+1}, \mathcal{F}_{T_i}\} \) is a sequence of martingale difference such that for some \( \alpha \geq 2 \) and \( c > 0 \),

\[ \sup_i E(|e_{T_i+1}|^\alpha |\mathcal{F}_{T_{i-1}})^{\frac{1}{\alpha} \leq E|e_1|^{\alpha} \leq C. \]

Lemma 2 in Wei(1987) can be applied in this case. \[ \]

**Proof of Statements in Appendix**

**Proof of Proposition 5.** Without loss of generality, we assume \( c = 1 \). For the subsets \( \Omega_{i,j,k}, \chi_{i,j}, 0 < j \leq C_{i,M}, i = 1, 2, k \geq 0 \) being such that

\[ \{ |\tilde{r}_n - r| \geq n^{-(1-6\Delta)} \} \subset \cup_{i,j,k} \{ \tilde{r}_n \in \Omega_{i,j,k} \} \cup_{i,j} \{ \tilde{r}_n \in \chi_{i,j} \} \cup \{ \tilde{r}_n = 0 \}, \]

56
we have
\[
P(\lvert \hat{r}_n - r \rvert \geq n^{-(1-6\Delta)}) \leq \sum_{i=1,2} \sum_{0<j \leq C_i,M;\ 0<k \leq \log n} P(\hat{r} \in \Omega_{i,j,k}) + \sum_{i=1,2} \sum_{0<j \leq C_i,M-1} P(\chi_{i,j}) + P(\hat{r} = 0).
\]
(79)

The following choice of subsets satisfy our needs; for all \(j > 0\), let \(\zeta_{i,j}, \Omega_{i,j,k}, z_{i,j}, C_{i,M}, i = 1,2, k \geq 0\) be such that
\[
\zeta_{2,j} = r + n^{-(1-6\Delta)} + n^{-1}(j-1) = r + n^{-z_{2,j}}, \quad \zeta_{1,j} = r - n^{-(1-6\Delta)} - n^{-1}(j-1) = r - n^{-z_{1,j}},
\]
\[
C_{2,M} = \arg \max_j \zeta_{2,j} \leq M, \quad C_{1,M} = \arg \max_j \zeta_{1,j} \geq -M,
\]
\[
\Omega_{2,j,k} = \begin{cases} 
[x_{S^j_{k}}^{\zeta_{2,j}}, x_{S^j_{k+1}}^{\zeta_{2,j}}) \cap \{0\}^c, & \text{if } k \in \Pi_{\zeta_{2,j}}, k > 0, \\
\emptyset, & \text{o.w.}
\end{cases}
\]
\[
\Omega_{1,j,k} = \begin{cases} 
[x_{S^j_{-k}}^{\zeta_{1,j}}, x_{S^j_{-k+1}}^{\zeta_{1,j}}) \cap \{0\}^c, & \text{if } -k \in \Pi_{\zeta_{1,j}}, k \geq 0, \\
\emptyset, & \text{o.w.}
\end{cases}
\]
\[
\chi_{i,j} = \{q^{[\zeta_{i,j-1}, \zeta_{i,j+1}]} \geq \log n\},
\]

From Corollary 11, there exists \(c, \delta > 0\) such that
\[
P(\hat{r}_n = 0) \leq P(\mathcal{RE}_n) \leq \exp(-cn^\delta).
\]
(80)

We are being very careful with the definition of \(\Omega_{s,j,k}\)’s. We claim the following results

Figure 5: Notation is slightly simplified. We flag \(\zeta\)’s on the horizontal line and bound the probability of \(\{\tilde{r}_n \in [x_{S^j}^{\zeta_{i}}, x_{S^j_{-k}}^{\zeta_{i-1}}]\}\) for each \(\zeta\), and then we argue the event "too many \(x_i\)’s falling inside the \(\zeta_i\)’s interval is also impossible" using the definition of \(\chi_{i,j}\).

with proof deferred to the succeeding subsection.
There exists $C > 0$ such that for $i = 1, 2, j$ such that $1 - 6\Delta \geq z_{i,j} > \frac{3}{2}\Delta$, $\log n \geq k \geq 0$, we have for all large $n$,

$$P(\tilde{r}_n \in \Omega_{i,j,k}) \leq C n^{-\Delta \left( \frac{n}{\pi} \right)}.$$  \hspace{1cm} (81)

There exists $C > 0$ such that for $i = 1, 2, j$ such that $2\Delta \geq z_{i,j} \geq z_{i,C_i,M}$, $\log n \geq k \geq 0$, we have for all large $n$,

$$P(\tilde{r}_n \in \Omega_{i,j,k}) \leq C n^{-\Delta \left( \frac{n}{\pi} \right)}.$$  \hspace{1cm} (82)

There exists $C > 0$ such that for $i = 1, 2, 1 \leq j \leq C_i,M - 1$, all large $n$,

$$P(\chi_{i,j}) \leq C n^{-\Delta \left( \frac{n}{\pi} \right)}.$$  \hspace{1cm} (83)

We separately state (81), (82) due to a technical reason. By (80) to (83), there is $c, \delta > 0$ such that for all large $n$,

$$(79) \leq C \log n \times n \times n^{-\Delta \left( \frac{n}{\pi} \right)} + C \exp \left( -cn^\delta \right) \leq C (\log n) n^{-\Delta \left( \frac{n}{\pi} \right)} + 1 \quad \blacksquare$$

In the following we shall prove (83), (81), and (82).

**Proof of (83).** For any $i = 1, 2, 1 \leq j \leq C_i,M$, all large $n$, $\tilde{M} = \sup_x f_\varepsilon(x)$,

$$P(\chi_{i,j}) \leq \sum_{j=\log n}^n C^n_j \left( \frac{\tilde{M}}{n} \right)^j (1 - \frac{\tilde{M}}{n})^{n-j}$$

$$\leq n \times \exp \left( (-n + \log n) \log \left( 1 - \frac{\log n}{n} \right) + \log n + \tilde{M} \log n - \log n \log \log n \right)$$

$$\leq \exp \left( C \log n - n \log \log n \right)$$

$$\leq C n^{-\Delta \left( \frac{n}{\pi} \right)}^{-1},$$

by Lemma 10, Stirling’s formula, $\log (1 - x) \leq -2x$, $0 < x < 1/2$. So

$$\sum_{i,j} P(\chi_{i,j}) \leq C n^{-\Delta \left( \frac{n}{\pi} \right)} \quad \blacksquare$$

**Proof of (81).** For $\zeta \in \mathbb{R}^1$, define

$$\theta_1^\zeta = [-\infty, \zeta], \theta_2^\zeta = [\zeta, \infty], \theta_3 = [r, \infty], \theta_4 = [-\infty, r], \theta_5^\zeta = [r, \zeta] \cup [\zeta, r].$$

We also define an union of at most $n$ events for a small constant $\tau > 0$,

$$E_k = \bigcup_{i=1,2} \bigcup_{0 < j \leq C_i,M} \left\{ q^{\theta_k \cap B(x,0)} \leq \rho_2 n \right\}, k = 1, 2,$$

$$E_k = \bigcup_{i=1,2} \bigcup_{0 < j \leq C_i,M} \left\{ q^{\theta_k \cap B(x,0)} \leq \rho_2 n \right\}, k = 4, 5,$$
\[
E_3 = \bigcup_{i=1,2} \bigcup_{0<j\leq C_{i,M}} \{ q^i_{\delta,i,j} \cap B^c_{\tau}(0) \leq \rho_3 n^{1-\delta_{i,j}} \},
\]
\[
\tilde{F} = \bigcup_{k=1}^n E_k,
\]

\(\rho_2, \rho_3 > 0\) are given in Collorary 11, Lemma 12. By these lemmas, there are \(c, \delta > 0\) such that for all large \(n\),
\[
P(\tilde{F}) \leq \exp(-cn^\delta).
\]

We consider only the case \(z_{i,j}, i = 2.\) Let \(x_i, e_i^2\) be defined in Lemma 16 with \(\tilde{F}\) and for a given \(n, \{b_i\}_{i=1}^n\) such that \(b_i = b_j, i \neq j,\) and each portioon of the number of \(b_i\)'s falling inside the subsets \([-\infty, C_{1,M}], [C_{2,M}, \infty], [r, r + n^{-(1-6\Delta)}], [r, r + n^{-(1-6\Delta)}]\) is not less than \(1/5.\)

Let \(\zeta = r + n^{-z}, 1 - 6\Delta \geq z > 0.\) The following we introduce another penalty analog that calculates the fitted loss with threshold fitted to one point \(z\) at real line instead of a sample point; define \(P^{o}_{z,k}:\)

\[
P^{o}_{z,k} = \sum_{i=1}^{q^i_{\delta}} x^{o^2}_{i \tau^{o}_{i}}(a_1 - \hat{a}^o_{1,z,k})^2 - 2 \sum_{i=1}^{q^i_{\delta}} x^{o^2}_{i \tau^{o}_{i}} e^{o}_{i \tau^{o}_{i}+1}(a_1 - \hat{a}^o_{1,z,k})
\]

\[
+ \sum_{i=1}^{q^i_{\delta}} x^{o^2}_{i \tau^{o}_{i}+} - \sum_{i=1}^{k} x^{o^2}_{i \tau^{o}_{i}}(a_2 - \hat{a}^o_{2,z,k})^2 + 2 \sum_{i=1}^{q^i_{\delta}} x^{o^2}_{i \tau^{o}_{i}+} e^{o}_{i \tau^{o}_{i}+1} - \sum_{i=1}^{k} x^{o^2}_{i \tau^{o}_{i}} e^{o}_{i \tau^{o}_{i}+1}(a_2 - \hat{a}^o_{2,z,k})
\]

\[
+ \sum_{i=1}^{q^i_{\delta}} x^{o^2}_{i \tau^{o}_{i}+} - \sum_{i=1}^{k} x^{o^2}_{i \tau^{o}_{i}}(a_2 - a_1)^2 + \sum_{i=1}^{q^i_{\delta}} x^{o^2}_{i \tau^{o}_{i}+} + \sum_{i=1}^{k} x^{o^2}_{i \tau^{o}_{i}}(a_2 - a_1)(a_1 - \hat{a}^o_{1,z,k})
\]

\[
+ \sum_{i=1}^{q^i_{\delta}} x^{o^2}_{i \tau^{o}_{i}+} - \sum_{i=1}^{k} x^{o^2}_{i \tau^{o}_{i}}(a_1 - \hat{a}^o_{1,z,k})^2
\]

\[
- 2 \sum_{i=1}^{q^i_{\delta}} x^{o^2}_{i \tau^{o}_{i}+} e^{o}_{i \tau^{o}_{i}+1}(a_2 - a_1) + 2 \sum_{i=1}^{q^i_{\delta}} x^{o^2}_{i \tau^{o}_{i}+} e^{o}_{i \tau^{o}_{i}+1} + \sum_{i=1}^{k} x^{o^2}_{i \tau^{o}_{i}} e^{o}_{i \tau^{o}_{i}+1}(a_1 - \hat{a}^o_{1,z,k})
\]

\[
+ \sum_{i=1}^{n} e^{o^2}_{i+1}
\]

\[
= (I) + (II) + (III) + (IV) + (V) + (VI) + (VII) + (VIII) + (X) + \sum_{i=1}^{n} e^{o^2}_{i+1},
\]

(84)
The idea of the proof is to show that $P$ where

sense. To do this, we need to identify at least one term in the decomposition of

$\Delta z$ that is absolutely large. The reason we separate state (81) and (82) is because as

MomentBounds

to $(V)$.

Lemma 21.

$\sup_{z \in \Theta', k=1, 2} E \left| \frac{n}{n} \sum_{i=1}^{n} x_{T_{i}^{k}} \right|^{q} \leq C$. 

60
Proof of Lemma 21. We only need to consider the cases where \( \theta_{z_1}^z, \theta_{z_0}^z, z_0 = \sup \Theta' > 0, z_1 = \inf \Theta' < 0 \), since \( x^2 \) is decreasing as we have \( x \) closer to zero. We consider the case \( \theta = \theta_{z_0}^z \).

Let \( c > 1 \). By Corollary 11, there exist \( \rho_2 > 0, C > 0 \), such that for all \( s \) with \( \rho_2 c^s > 1 \), all large \( n \),

\[
P(T_{\rho_2 c^s}^n > c^s n) = P(T_{\rho_2 c^s}^n > c^s n) \leq P(T_{\rho_2 c^s}^n > c^s n) \leq \exp (-C(c^s n)^{\delta}),
\]

if we take \( c^s n \) as \( n \) in the Corollary. Also we know if \( q > 1 \), \( E|x_1|^{2q} < \infty \), by triangular inequality,

\[
E^{1/q} \left| \sum_{i=1}^{c^s n} x_i^2 \right|^q \leq \sum_{i=1}^{c^s n} E^{1/q} x_i^{2q} \leq C c^s n.
\]

Given \( s_0 > 0 \),

\[
E \left| \sum_{i=1}^{n} \frac{x_i^2}{n} \right|^q \leq E \left| \sum_{i=1}^{c^s_0 n} \frac{x_i^2}{n} \right|^q + \sum_{i=1}^{\infty} \left( \sum_{j=1}^{c^s_0 n+i} x_j^2 \right)^q
\]

\[
\leq E \left| \sum_{i=1}^{c^s_0 n} \frac{x_i^2}{n} \right|^q + \sum_{i=1}^{\infty} \left( \sum_{j=1}^{c^s_0 n+i} x_j^2 \right)^q
\]

\[
\leq E \left| \sum_{i=1}^{c^s_0 n} \frac{x_i^2}{n} \right|^q + \sum_{i=1}^{\infty} \left( \sum_{j=1}^{c^s_0 n+i} x_j^2 \right)^q
\]

Let \( q > 1 \), and define the partial sum of the second term on the RHS

\[
U_{s,s+k} = E^{1/q} \left| \sum_{i=s}^{s+k} \left( \sum_{j=1}^{c^s_0+i} x_j^2 \right)^q \right|.
\]

then we see, by triangular inequality and (91), (92), there exist some \( s_0, n_{s_0}, C > 0, \delta > 0, c > 0 \) such that for all \( k > s > s_0 > 1, n > n_{s_0}, \)

\[
U_{s,s+k} = E^{1/q} \left| \sum_{j=0}^{k} \left( \sum_{i=1}^{c^s_0+i+j} x_i^2 \right)^q \right|
\]

\[
\leq E^{1/q} \left| \sum_{j=0}^{k} \left( \sum_{i=1}^{c^s_0+i+j} x_i^2 \right)^{2q} \right| \leq \sum_{j=0}^{k} \left( \exp \left( -c(c^s+j)^{\delta} \right) \right)^{\frac{1}{2q}} \times C n c^{s_0+1+s+j}
\]

\[
\leq \sum_{j=0}^{k} C \exp \left( -(c^s+j)^{\delta} \right) \times n c^{s+j}
\]

\[
\leq s^{-2}.
\]

So

\[
\limsup_{k \to \infty} U_{s,k} < \infty.
\]
By triangular inequality, Cauchy-Swartz inequality, $E|x_1|^{4q} < \infty$, (95), and monotone convergence theorem, we have (93) $< \infty$.

\textbf{Proof of (86).} By triangular inequality, the definition of $x_i^o, e_i^o$, Remark of Lemma 16, Lemma 20, Lemma 21, and moment assumption, there exists $C > 0$ such that for all $0 < k \leq \log n$, $w = 1, 2, 4, 5$,

$$
E \left| n^{-1/2} \left( \sum_{i=1}^{q^o_w} x_{T_i^o w}^o e_{T_i^o w+1}^o + \sum_{i=1}^k x_{S_i^o}^o e_{S_i^o+1}^o \right) \right|^q
\leq \left[ E^{1/2} \left( \frac{1}{n^{1/2}} \sum_{i=1}^{q^o_w} x_{T_i^o w}^o e_{T_i^o w+1}^o \right)^q + E^{1/2} \left( \frac{1}{n^{1/2}} \sum_{i=1}^k x_{S_i^o}^o e_{S_i^o+1}^o \right)^q \right]^q
\leq \left[ E^{1/2} \left( \frac{1}{n^{1/2}} \sup_{1 \leq j \leq n} \sum_{i=1}^j |x_{T_i^o w}^o e_{T_i^o w+1}^o| \right)^q + C \right]^q
\leq \left[ E^{1/2} \left( \frac{1}{n} \sum_{i=1}^n |x_{T_i^o w}^2| \right)^{q/2} + C \right]^q
\leq C.
$$

\textbf{Proof of (87).} This is a direct result of Lemma 19. Notice that we have required conditions all satisfied.

\textbf{Proof of (88).} The negative moment is bounded by the definition of $x_i^o, e_i^o$. For the positive moment, by triangular inequality, Lemma 21 and moment assumption, there exists $C > 0$ such that for all $w = 1, 2, 4, 5$, $|k| \leq c \log n$,

$$
E \left| n^{-1} \left( \sum_{i=1}^{q^o_w} x_{T_i^o w}^{o^2} + \sum_{i=1}^k x_{S_i^o}^{o^2} \right) \right|^q
\leq \left[ E^{1/4} \left( \frac{1}{n^{1/4}} \sum_{i=1}^n x_{T_i^o w}^{o^2} \right)^q + E^{1/4} \left( \frac{1}{n^{1/4}} \sum_{i=1}^n x_{S_i^o}^{o^2} \right)^q + C \right]^q
\leq \left[ E^{1/4} \left( \frac{1}{n^{1/4}} \sum_{i=1}^n x_{T_i^o w}^{o^2} \right)^q + E^{1/4} \left( \frac{1}{n^{1/4}} \sum_{i=1}^n x_{S_i^o}^{o^2} \right)^q + C \right]^q
\leq C.
$$

\textbf{Proof of (89).} By (77), the definition of $x_i^o, e_i^o, x_{T_i^o w}^{o^2} \in \Theta'$, standard inequalities, moment assumption, and Lemma 12,

$$
E \left| n^{-(1-z)} \left( \sum_{i=1}^{q^o_w} x_{T_i^o w}^{o^2} + \sum_{i=1}^k x_{S_i^o}^{o^2} \right) \right|^q
\leq \left[ E^{1/4} \left| n^{-(1-z)} \left( \sum_{i=1}^{q^o_w} x_{T_i^o w}^{o^2} \left( 1_{q^o_w \leq c_1 n^{1-z}} + 1_{q^o_w \geq c_2 n^{1-z}} \right) \right) \right|^q + C \right]^q
$$

62
By these moment bounds, we claim that (90) can be established simply by the definition of \( x_i^0, e_i^0 \).

**Proof of (90).** (90) can be established simply by the definition of \( x_i^0, e_i^0 \).

Assume \( E |e_1|^{16q} < \infty \). Given \( \frac{5}{6} > \Delta > 0 \), there exists \( C > 0 \) such that for \( 1 - 6\Delta \geq z \geq \frac{3}{2}\Delta, 0 < k \leq \log n, \) all large \( n \),

\[
P \left( \max \{|(I)|, |(II)|, |(III)|, |(IV)|, |(V)|, |(VI)|, |(VII)|, |(VIII)|, |(X)| \right) = (REST) \geq n^{1-z-2\Delta} \leq Cn^{-\Delta q}
\]

(96)

**Proof of the claim (96).** We show the results for (VIII) and (V), (VII); the others can be argued in a similar way. For (V), by (90), we have

\[
P \left( \left| \sum_{i=1}^{n} x_i^{o2} T_i^{\theta_i} + \sum_{i=1}^{k} x_i^{o2} S_i \right|^{-1} (a_2 - a_1)^2 \geq n^{-(1-z)+\Delta} \right) \leq Cn^{-\Delta q},
\]

(97)

Note that this is the desired result for (V). The same logic applies to all cases. For another example, (VIII): by (89),

\[
P \left( -2 \left| \sum_{i=1}^{n} x_i^{o} T_i^{\theta_i} c_i^{o} + \sum_{i=1}^{k} x_i^{o} S_i \right| (a_2 - a_1) \geq n^{1-z-2\Delta} \right) \leq Cn^{-\Delta q}
\]

(98)

One last example is much more serious; for (VII),

\[
P \left( \left| \sum_{i=1}^{n} x_i^{o2} T_i^{\theta_i} + \sum_{i=1}^{k} x_i^{o2} S_i \right| (a_1 - \hat{a}_{1,z,k}^o)^2 \geq n^{1-z-2\Delta} \right) \leq Cn^{-\Delta q} E \left( \sum_{i=1}^{n} x_i^{o2} T_i^{\theta_i} + \sum_{i=1}^{k} x_i^{o2} S_i \right) \left( a_1 - \hat{a}_{1,z,k}^o \right)^2 \]

\[
\leq Cn^{-\Delta q} E \left( \sum_{i=1}^{n} x_i^{o2} T_i^{\theta_i} + \sum_{i=1}^{k} x_i^{o2} S_i \right) \left( a_1 - \hat{a}_{1,z,k}^o \right)^2 \frac{n^{1-z}}{n^{2-3\Delta}} \]

\[
= Cn^{-\Delta q} E \left( \sum_{i=1}^{n} x_i^{o2} T_i^{\theta_i} + \sum_{i=1}^{k} x_i^{o2} S_i \right) \left( a_1 - \hat{a}_{1,z,k}^o \right)^2 \frac{n^{1-z}}{n^{2-3\Delta}} \]

63
\[ Cn^{-\Delta q} \left( \sum_{i=1}^{q^o} x_{i o} + \sum_{i=1}^{k} x_{i o}^o \right) \leq E_{\frac{1}{n^3}} \left( \sum_{i=1}^{q^o} x_{i o}^o + \sum_{i=1}^{k} x_{i o}^o \right)^2 \times E_{\frac{1}{n^2}} \left( \sum_{i=1}^{q^o} x_{i o}^o + \sum_{i=1}^{k} x_{i o}^o \right)^{-2} \times n^2 \]

\[ \leq Cn^{-\Delta q} \left( \sum_{i=1}^{q^o} x_{i o}^o + \sum_{i=1}^{k} x_{i o}^o \right)^2 \times n^{2-3\Delta} \]

\[ \leq Cn^{-\Delta q} \left( \sum_{i=1}^{q^o} x_{i o}^o + \sum_{i=1}^{k} x_{i o}^o \right)^2 \times n^{2-3\Delta} \]

\[ \equiv Cn^{-\Delta q} \left( (a) + (b) + (c) \right)^{8q} \leq Cn^{-\Delta q}, \]

by (F4), (F3), triangular inequality, provided \(2 - 3\Delta \geq \frac{1}{2}, \frac{3}{2} - 3\Delta \geq 1 - z, 2 - 3\Delta \geq 2(1 - z).\)

\(E|e_1|^{16q} < \infty\) is for (b). Note that the above conditions imply \(\Delta \leq \frac{5}{6}, \frac{3}{2} \Delta \leq z\). Combining with the \(1 - 6\Delta > b\) constraint in Lemma 19, we have finished the proof.

By (96), for all \(1 - 6\Delta \geq z \geq \frac{3}{2}\Delta,

\[ P \left( P_{\Delta} - \sum_{i=1}^{n-1} e_{i+1} \leq \left\{ n^{1-z-2\Delta} (n^\Delta - 1) \right\} \right) \leq P \left( \{ V \} \leq n^{1-z-\Delta} \cup \{ (REST) \geq n^{1-z-2\Delta} \} \right) \]

\[ \leq Cn^{-\Delta q}. \]

(99)

For the correct estimation, we note that if \(E|e_1|^{4q} < \infty\), for any \(\delta > 0,

\[ P \left( \sum_{i=1}^{n-1} e_{i+1} \geq n^\delta \right) \leq n^{-\delta q}. \]

(100)
Since \( \inf_{\frac{3}{2} \Delta < z \leq 1 - 6 \Delta} n^{1-z-2\Delta} (n^{\Delta} - 1) \geq C n^{5\Delta} \) for some \( C > 0 \), all large \( n \), we can pick \( \Delta < \delta < 5 \Delta \). Then by (99), (100), for all large \( n \), \( \frac{3}{2} \Delta < z \leq 1 - 6 \Delta, \)
\[
P(P_{z,k}^o - Z_0^o < 0) \leq C n^{-\Delta q}, \tag{101}
\]
for some \( C > 0 \). By the definition
\[
\Omega_{z,k} = \left[ x_{S_k^2}^{z}, x_{S_{k+1}^2}^{z} \right] \cap \{0\}^c, k > 0.
\]
We have, for all large \( n \), \( 1 - 6 \Delta \geq z > \frac{3}{2} \Delta \), \( 0 < k \leq \log n \),
\[
P(\hat{r}_n \in \Omega_{z,k}) \leq P(P_{z,k} \leq Z_0) \leq P(\hat{r}_n \leq Z_0^o) + P(\hat{r}_n) \leq C n^{-\Delta q} + P(\hat{r}_n),
\]
by (101). This finish the proof of (81).

**Proof of (82).** Still we consider \( z_{2,j} \)'s. In (84),
\[
(V) + (VI) + (VII) \geq 0. \tag{102}
\]
For \( 2 \Delta \geq z \geq z_{2,C_{2,M}} \), the biases \( |a_i - \hat{a}_{i,z,k}^o|, i = 1, 2 \), will be nearing constant instead of a quantity decreasing in \( n \). By (85),
\[
(I) = \sum_{i=1}^{q_{\theta_4}} x_{T_i^4}^{o} x_{T_i^4}^{\theta_4} \left( \sum_{i=1}^{q_{\theta_2}} x_{S_i^2}^{o} + \sum_{i=1}^{k} x_{S_i^2}^{o} \right)^{-1}
\times \left( \sum_{i=1}^{q_{\theta_1}} x_{T_i^1}^{o} e_{T_i^1}^{o} + \sum_{i=1}^{k} x_{S_i^2}^{o} e_{S_i^2}^{o} \right) + (a_1 - a_2) \left( \sum_{i=1}^{q_{\theta_3}} x_{T_i^3}^{o} + \sum_{i=1}^{k} x_{S_i^2}^{o} \right)^2
\equiv (I).a + (I).b + (I).c,
\]
where \( a, b, \) and \( c \) result from the expansion of the square.

**MomentBounds**

There exist an uniform upper bound \( C > 0 \) such that for \( 0 < k \leq \log n \), \( 2 \Delta \geq z \geq z_{2,C_{2,M}}, \) \( \zeta = r + n^{-z} \),
\[
E \left| n^{-\frac{1}{2}} \left( \sum_{i=1}^{q_{\theta_3}} x_{T_i^3}^{o} e_{T_i^3}^{o} + \sum_{i=1}^{k} x_{S_i^2}^{o} e_{S_i^2}^{o} \right) \right|^q \leq C; \tag{103}
\]
\[
E \left| n^{-1} \left( \sum_{i=1}^{q_{\theta_3}} x_{T_i^3}^{o} + \sum_{i=1}^{k} x_{S_i^2}^{o} \right) \right|^q \leq C; \tag{104}
\]
\[
E \left| n^{-(1-2\Delta)} \left( \sum_{i=1}^{q_{\theta_3}} x_{T_i^3}^{o} + \sum_{i=1}^{k} x_{S_i^2}^{o} \right) \right|^{-q} \leq C. \tag{105}
\]
The last one holds because of the definition of \( x_i^c \); the rest of the moment bounds, stated with the proofs omitted, as well as (86) and (88) hold almost trivially in this case. By the moment bounds, we claim

There exist \( C > 0 \) such that for all \( 0 < k \leq \log n, 2\Delta \geq z \geq z_{i,C_{i,M}} \),

\[
P\left( \max\{|(I).a|, |(I).b|, |(I).c|, |(IV)|, |(VIII)|, |(X)| \} \geq n^{1-\Delta} \right) \leq Cn^{-\Delta q},
\]

\[
P((I).c \leq n^{1-5\Delta}) \leq Cn^{-\Delta q}.
\]

**Proof of the claim.** We prove the case for \((I).c\). By (88), (105), Lemma 21, there exist a positive constant such that for all \( 0 < k \leq \log n, 2\Delta \geq z \geq z_{i,C_{i,M}} \),

\[
P \left( \sum_{i=1}^{q^{\emptyset}_j} x_i^{\emptyset_2} \sum_{i=1}^{k} x_i^{\emptyset_2} \sum_{i=1}^{\sum_{i=1}^{k} x_i^{\emptyset_2} S_i^j} C^{-2} \left( \sum_{i=1}^{k} x_i^{\emptyset_2} \sum_{i=1}^{\sum_{i=1}^{k} x_i^{\emptyset_2} S_i^j} \right)^{-2} \geq n^{-(1-4\Delta)+\Delta} \right) \leq n^{-\Delta q}.
\]

In the other cases, we have added some slackness to the order of normalizers to make the results tight and aligned.

By (102), (106) and the same argument in the proof of (83), we have \( C > 0 \) such that for all \( 0 < k \leq \log n, 2\Delta \geq z \geq z_{i,C_{i,M}} \),

\[
P(\bar{r}_n \in \Omega_{z,k}) \leq P(P_{z,k} \leq \Xi_0) \leq P(P_{z,k}^0 \leq \Xi_0^0) \leq P(\bar{F}) \leq Cn^{-\Delta q} + P(\bar{F}).
\]

**Statement and the Proof of Corollary 22**

**Corollary 22.** Assume the assumptions for Proposition 5. Let \( c > 0 \) and

\[A_1 = \{ \arg \min_{n \geq 2} \bar{P}_1 < 1 - cn^{6\Delta} \}, A_2 = \{ \arg \min_{n \geq 2} \bar{P}_1 > cn^{6\Delta} \}.\]

Then there exist \( c_1 > 0 \) such that for all large \( n \),

\[
P \left( A_1 \cup A_2, \{1 - c_1 n^{6\Delta}, c_1 n^{6\Delta} \} \in \Pi' \right) \leq C(\log n)n^{-\Delta(q^a)+1}.
\]

**Proof of Corollary 22.** By Proposition 5 and Lemma 12, there exists some \( c_1 > c'_1 > 0 \) small enough such that for all large \( n \),

\[
P \left( A_1 \cup A_2, \{-c_1 n^{6\Delta} + 1, c_1 n^{6\Delta} \} \in \Pi \right)
\]
where $S$ is large enough such that all large $n$ threshold $r > \beta$. By Corollary 22, we have for all $1 - nq_2 \leq j \leq nq_2$, all large $n$,

\[
|P(\tilde{r}_n = x_{S_j}, n(x_{S_j} - r) \in \beta, j \in \Pi) - P(\tilde{r}_n = x_{S_j}, n(x_{S_j} - r) \in \beta, \{nq_2, 1 - nq_2\} \subset \Pi)| \leq P(\tilde{r}_n = x_{S_j}, n(x_{S_j} - r) \in \beta, j \in \Pi, \{nq_2, 1 - nq_2\} \not\subset \Pi) \leq \exp(-cn^\delta).
\]

For all $1 - nq_2 \leq j \leq nq_2$, $\{nq_2, 1 - nq_2\} \subset \Pi$ guarantees a well-defined $S_j$; hence

\[
P(\tilde{r}_n = x_{S_j}, n(x_{S_j} - r) \in \beta, \{nq_2, 1 - nq_2\} \subset \Pi) \leq P \left( j = \arg \min_{1 - nq_2 \leq j \leq nq_2} \tilde{P}_i - \tilde{P}_0, n(x_{S_j} - r) \in \beta, \{nq_2, 1 - nq_2\} \subset \Pi \right).
\]

By Corollary 22, we have for all $1 - nq_2 \leq j \leq nq_2$, all large $n$,

\[
|P \left( j = \arg \min_{1 - nq_2 \leq j \leq nq_2} \tilde{P}_i - \tilde{P}_0, n(x_{S_j} - r) \in \beta, \{nq_2, 1 - nq_2\} \subset \Pi \right) - P \left( j = \arg \min_{1 - nq_2 \leq j \leq nq_2} \tilde{P}_i - \tilde{P}_0, n(x_{S_j} - r) \in \beta, \{nq_2, 1 - nq_2\} \subset \Pi \right) | = O(n^{-2}),
\]
where a sufficiently large $q$ is needed. We use $x_i^o$ in the proof of Proposition 5 to deal with the inverse moment, but here we can adopt a causal rare event: let $\theta_1 = [\infty, -\varepsilon] \cap \Theta', \theta_2 = [r, \infty] \cap \Theta', F = \{q^{(1)} \leq \rho_2 n\} \cup \{q^{(2)} \leq \rho_2 n\}$. By Corollary 11, there exist $c, \rho_2, \delta_1, \varepsilon > 0$ such that for all large $n$,

$$P(F) \leq \exp(-cn^\delta).$$  \hfill (110)

Now we construct our $x_i^o$'s on the newly adopted rare event $F$,

$$x_i^o = x_i 1_{F^c} + (|r| + 1 - \frac{1}{t^2})(-1_F)^i, \ e_i^n = c_i 1_{F^c}.$$

Define the $x_i^o$ analogy of $Z_j^o, \bar{P}_j$, and denote them as $Z_j^o, P_j^o$, and we drop off the superscript on the random indexes $S_i^o$'s, $J_i^o$'s; as we did in the proof of Proposition 5, we decompose $P_j - P_0^o$ into terms, whose specific definitions are deferred to the succeeding subsection. For $1 - n_q \leq j \leq n_q$,

$$P_j - P_0^o = Z_j^o - Z_0^o = (I)_j + \cdots + (VI)_j + S_j^o + (VII)_j + \cdots + (XI)_j, a.s., \hfill (111)$$

where

$$S_j^o = \begin{cases} \sum_i j \{r^2(a_2 - a_1)^2 + 2e_{S, a_2 + 1}r(a_2 - a_1)\}, j > 0, \\ \sum_i j \{r^2(a_1 - a_2)^2 + 2e_{S, a_1 + 1}r(a_1 - a_2)\}, j \leq 0, \end{cases}, k = 1, 2. \hfill (112)$$

Since $P(x_i^o \neq x_i, 1 \leq i \leq n) \leq P(F)$,

$$\left| P \left( j = \arg \min \left[ \bar{P}_j - \bar{P}_0, n(x_{S_j} - r) \in \beta, \{n_q, 1 - n_q\} \in \Pi \right] \right) - P \left( j = \arg \min \left[ P_i^o - P_0^o, n(x_{S_j} - r) \in \beta, \{n_q, 1 - n_q\} \in \Pi \right] \right) \right| \leq P(F). \hfill (113)$$

**Step1 : Bounding the irrelavent terms**

For $1 - n_q \leq j \leq n_q$, all large $n$,

$$P \left( \max_{1 - n_q \leq j \leq n_q} \max \{|(I)|_j, \ldots, |(XI)|_j\} \geq n^{-\frac{1}{2} - 25\Delta} \right) \leq n_{q_1}. \hfill (114)$$

Proof of (114) is deferred to next part of the proof. Define

$$(\ast) = \max_{-n_q + 1 \leq j \leq n_q} \max \{|(I)|_j, \ldots, |(XI)|_j\},$$

$$\Gamma_{n,j} = \begin{cases} (I)_j + \cdots + (XI)_j, (\ast) < n^{-\frac{1}{2} - 25\Delta}, \\ 0, (\ast) \geq n^{-\frac{1}{2} - 25\Delta}. \end{cases}$$
For $1 - n_{q_2} \leq j \leq n_{q_2}$, all large $n$,

$$
|P(j = \arg \min_{1-n_{q_2} \leq i \leq n_{q_2}} P_i^o - P_0^o, n(x_{S_j} - r) \in \beta, \{n_{q_2}, 1- n_{q_2}\} \in \Pi) - P \left( j = \arg \min_{1-n_{q_2} \leq i \leq n_{q_2}} S_i + \Gamma_{n,i}, n(x_{S_j} - r) \in \beta, \{n_{q_2}, 1- n_{q_2}\} \in \Pi \right) | \\
\leq P \left( \max_{1-n_{q_2} \leq i \leq n_{q_2}} \max\{|(I)_j|, \ldots, |(XIJ)_j|\} \geq n^{-\left(\frac{1}{2} - 25\Delta\right)} \right) \leq n_{q_1},
$$

(115)

by (114). For the $x_i$ analogy, $S_j$, we have for all $1 - n_{q_2} \leq j \leq n_{q_2}$, all large $n$,

$$
|P(j = \arg \min_{1-n_{q_2} \leq i \leq n_{q_2}} S_i^o + \Gamma_{n,i}, n(x_{S_j} - r) \in \beta, \{n_{q_2}, 1- n_{q_2}\} \in \Pi) - P(j = \arg \min_{1-n_{q_2} \leq i \leq n_{q_2}} S_i + \Gamma_{n,i}, n(x_{S_j} - r) \in \beta, \{n_{q_2}, 1- n_{q_2}\} \in \Pi) | \\
\leq P(F).
$$

(116)

**Step 2:** Construction of larger and smaller subsets

For $1 - n_{q_2} \leq j \leq n_{q_2}$, define

$$
\mathcal{L}_j = \left\{ \hat{z} : \tilde{z} = (\tilde{z}_{1- n_{q_2}}, \ldots, \tilde{z}_{n_{q_2}}) \in \mathbb{R}^{2n_{q_2}}, j = \arg \min_{1-n_{q_2} \leq i \leq n_{q_2}} \tilde{z}_i \right\},
$$

where $\tilde{z}_i$ is the $i + n_{q_2}$-th coordinate of $\hat{z}$. For $1 - n_{q_2} \leq j \leq n_{q_2}$,

$$
P\left( j = \arg \min_{1-n_{q_2} \leq i \leq n_{q_2}} S_i + \Gamma_{n,i}, n(x_{S_j} - r) \in \beta, \{n_{q_2}, 1- n_{q_2}\} \in \Pi \right) = P\left( (S_i + \Gamma_{n,i}, 1- n_{q_2} \leq i \leq n_{q_2}) \in \mathcal{L}_j, n(x_{S_j} - r) \in \beta, \{n_{q_2}, 1- n_{q_2}\} \in \Pi \right).
$$

We construct the augmented and diminished subset of $\mathcal{L}_j$, denoting them as $\hat{\mathcal{L}}_j$, $\hat{\mathcal{L}}_j$, respectively; for $1 - n_{q_2} \leq j \leq n_{q_2}$, let $v^j \in \mathbb{R}^{2n_{q_2}}$ be of unit length, $v^j_j = 1$; for $1 - n_{q_2} \leq j \leq n_{q_2}$,

$$
\hat{\mathcal{L}}_j = \left\{ \hat{z} + hv^j : \hat{z} \in \mathcal{L}_j, 0 \leq h \leq 2n^{-\left(\frac{1}{2} - 25\Delta\right)} \right\},
$$

$$
\hat{\mathcal{L}}_j = \left\{ \hat{z} - hv^j : \hat{z} \in \mathcal{L}_j, 0 \leq h \leq 2n^{-\left(\frac{1}{2} - 25\Delta\right)} \right\}.
$$

For all $1 - n_{q_2} \leq j \leq n_{q_2}$,

$$
P\left( (S_i, 1 - n_{q_2} \leq i \leq n_{q_2}) \in \mathcal{L}_j, n(x_{S_j} - r) \in \beta, \{n_{q_2}, 1- n_{q_2}\} \in \Pi \right) \\
\leq P\left( (S_i + \Gamma_{n,i}, 1 - n_{q_2} \leq i \leq n_{q_2}) \in \mathcal{L}_j, n(x_{S_j} - r) \in \beta, \{n_{q_2}, 1- n_{q_2}\} \in \Pi \right) \\
\leq P\left( (S_i, 1 - n_{q_2} \leq i \leq n_{q_2}) \in \hat{\mathcal{L}}_j, n(x_{S_j} - r) \in \beta, \{n_{q_2}, 1- n_{q_2}\} \in \Pi \right).
$$

(117)

In the following we discuss the upper bound. The lower bound shall be argued in the same way.
Step 3: The asymptotical independence and identical distribution

In this step, notation in Lemma 13, 14 are used, and we consider \( \theta_1 = [r - c_0 n^{-(1 - 6\Delta)}], \theta_2 = [r, r + c_0 n^{-(1 - 6\Delta)}] \) for some \( c_0 > 0 \) with value depending on the other parameters. Define

\[
S'_j = \begin{cases} 
\sum_{i=1}^{j} \{ r^2(a_2 - a_1)^2 + 2(a_2 - a_1)re_{s_{i-2}+1} \}, j > 0, \\
\sum_{i=1}^{[j]+1} \{ r^2(a_1 - a_2)^2 + 2(a_1 - a_2)re_{s_{i-1}+1} \}, j \leq 0,
\end{cases}
\]

By Lemma 12, there exists \( c_1, c_0, c, \delta > 0 \) such that for all \( 1 - n q_2 \leq j \leq n q_2 \),

\[
\left| P((S_i, 1 - n q_2 \leq i \leq n q_2) \in \tilde{L}_j, n(x_{S_j} - r) \in \beta, \{n q_2, 1 - n q_2 \} \in \Pi) \\
- P((S_i, 1 - n q_2 \leq i \leq n q_2) \in \tilde{L}_j, n(x_{S_j} - r) \in \beta, n q_2 \leq \min\{q^\theta_1, q^\theta_2\}) \right| \\
\leq |P ((S_1 \setminus S_2) \cup (S_2 \setminus S_1))| \leq \exp (-cn^\delta),
\]

where \( S_1 = \{ \{n q_2, 1 - n q_2 \} \in \Pi \}, S_2 = \{ n q_2 \leq \min\{q^\theta_1, q^\theta_2\} \}. \) By Lemma 13 and the fact that if

\[
T^\theta_i = T^\theta_i^k \text{ for all } 1 \leq i \leq \max\{q^\theta_1, q^\theta_2\}, k = 1, 2,
\]

then

\[
\begin{align*}
q^\theta_i &= q^\theta_i^k, k = 1, 2, \\
s_{j,2} &= S_j, n q_2 \geq j > 0, \\
s_{j+[j]+1} &= S_{-j}, 1 - n q_2 \leq j \leq 0,
\end{align*}
\]

we have for all \( 0 < j \leq n q_2, k = 2, \)

\[
\left| P((S_i, -n q_2 + 1 \leq i \leq n q_2) \in \tilde{L}_j, n(x_{S_j} - r) \in \beta, n q_2 \leq \min\{q^\theta_1, q^\theta_2\}) \\
- P((S'_i, -n q_2 + 1 \leq i \leq n q_2) \in \tilde{L}_j, n(x_{s_{j,k}} - r) \in \beta, n q_2 \leq \min\{q^\theta_1, q^\theta_2\}) \right| \\
\leq C(\log n)^2 n^{18\Delta - 1};
\]

and \( k = 1, 1 - n q_2 \leq j \leq 0 \) can be argued in the same way. By Corollary 15, for all \( 0 < j \leq n q_2, k = 2, \)

\[
\left| P((S'_i, 1 - n q_2 \leq i \leq n q_2) \in \tilde{L}_j, n(x_{s_{j,k}} - r) \in \beta, n q_2 \leq \min\{q^\theta_1, q^\theta_2\}) \\
- P((S^\kappa_i, 1 - n q_2 \leq i \leq n q_2) \in \tilde{L}_j, n(x_{s_{j,k}} - r) \in \beta, n q_2 \leq \min\{q^\theta_1, q^\theta_2\}) \right| \\
\leq C(n^4 \rho \log n + n^{-2q+1}),
\]

where

\[
S^\kappa_j = \begin{cases} 
\sum_{i=1}^{j} \kappa_{2,i}, j > 0, \\
\sum_{i=1}^{[j]+1} \kappa_{1,i}, j \leq 0;
\end{cases}
\]
and $0 \geq j \geq 1 - n_{q_2}$ can be argued in the same way. By Lemma 13 again, there exists $C > 0$ such that for all $n_{q_2} \geq j > 0$, $k = 2$, all large $n$,

$$
\left| P(n(x_{s_j} - r) \in \beta, n_{q_2} \leq \min\{q^2, q^3\}) - P(n(x_{s_j} - r) \in \beta, n_{q_2} \leq \min\{q^2, q^3\}) \right| 
\leq C(\log n)^2 n^{18\Delta - 1},
$$

(121)

and $1 - n_{q_2} \leq j \leq 0$, $k = 1$ can be argued in the same way. By Lemma 12, there exist $c, \delta > 0$ such that for all $1 - n_{q_2} \leq j \leq n_{q_2}$, all large $n$,

$$
\left| P(n(x_{s_j} - r) \in \beta, n_{q_2} \leq \min\{q^2, q^3\}) - P(n(x_{s_j} - r) \in \beta, j \in \Pi) \right| \leq C \exp(-cn^\delta).
$$

(122)

By (118) to (122), we have for all $1 - n_{q_2} \leq j \leq n_{q_2}$,

$$
\left| P((S_i, 1 - n_{q_2} \leq i \leq n_{q_2}) \in \tilde{L}_j, n(x_{s_j} - r) \in \beta, \{n_{q_2}, 1 - n_{q_2}\} \in \Pi) 
- P((S_i, 1 - n_{q_2} \leq i \leq n_{q_2}) \in \tilde{L}_j P(n(x_{s_j} - r) \in \beta, j \in \Pi) \right|
\leq C \exp(-cn^\delta) + C(\log n)^2 n^{18\Delta - 1} + C(n^4 \rho \log n + n^{-2q + 1}).
$$

(123)

**Step 4: Recovering**

Given $\tilde{z} \in \mathbb{R}^{2n_{q_2}}$, let

$$
S_{z_j}^z = \begin{cases}
\sum_{i=1}^j \{r^2(a_2 - a_1)^2 + 2(a_2 - a_1)r \tilde{z}_i\}, j > 0, \\
\sum_{i=0}^{j-1} \{r^2(a_1 - a_2)^2 + 2(a_1 - a_2)r \tilde{z}_{i-1}\}, j \leq 0.
\end{cases}
$$

We consider the case $j > 0$, $a_2 - a_1 > 0$ and establish the proof based on this assumption; the other cases can be shown in the same fashion with different set of notation. Let

$$
L_j' = \{ \tilde{z} : (S_{z_j}^z, 1 - n_{q_2} \leq i \leq n_{q_2}) \in L_j \},
$$

$$
\tilde{L}_j' = \{ \tilde{z} : (S_{z_j}^z, 1 - n_{q_2} \leq i \leq n_{q_2}) \in \tilde{L}_j \}.
$$

Let $\tilde{z} = \tilde{z} + \tilde{v}_j h [2r(a_2 - a_1)]^{-1} - \tilde{v}_{j+1} h [2r(a_2 - a_1)]^{-1}$ and we see $(S_{z_j}^z, 1 - n_{q_2} \leq i \leq n_{q_2}) = (S_{\tilde{z}_j}^z, 1 - n_{q_2} \leq i \leq n_{q_2}) + \tilde{v}_j h$. By this observation,

$$
\tilde{L}_j' = \{ (z_{1-n_{q_2}}, \ldots, z_j + h, z_{j+1} - h, \ldots, z_{n_{q_2}}) : \tilde{z} \in L_j', 0 \leq h \leq [2r(a_2 - a_1)]^{-1} n^{-\left(\frac{3}{2} - 25\Delta\right)} \}.
$$

Define

$$
\tilde{L}_{j,1} = \{ (z_{1-n_{q_2}}, \ldots, z_j + h_1, z_{j+1} - h_2, \ldots, z_{n_{q_2}}) : \tilde{z} \in L_j', 0 \leq h_i \leq [2r(a_2 - a_1)]^{-1} n^{-\left(\frac{3}{2} - 25\Delta\right)}, i = 1, 2 \},
$$

$$
\tilde{L}_{j,0} = \{ (z_{1-n_{q_2}}, \ldots, z_j + h, z_{j+1}, \ldots, z_{n_{q_2}}) : \tilde{z} \in L_j', 0 \leq h \leq [2r(a_2 - a_1)]^{-1} n^{-\left(\frac{3}{2} - 25\Delta\right)} \}.
$$
Here is another observation: due to \( j > 0, a_2 > a_1 \), we have if \( y \in \mathcal{L}_{j,1}' \), then \( y + y_{j+1} s \in \mathcal{L}_{j,1}' \), for all \( s \in \mathbb{R}_+ \), the same property holds for \( y_{j} \) with opposite sign. By this observation and assumption 3, there exists \( C > 0 \) such that for all large \( n \),

\[
P((S_i^\ast, -n_{q2} + 1 \leq i \leq n_{q2}) \in \mathcal{L}_j) = \int_{z \in \mathcal{L}_{j,0}'} \prod_{z_{q2+1} \leq i \leq n_{q2}} f_e(z_i) dz_i
\]

\[
\leq \int_{z \in \mathcal{L}_{j,1}'} \prod_{z_{q2+1} \leq i \leq n_{q2}} f_e(z_i) dz_i
\]

\[
= \int_{z \in \mathcal{L}_{j,1}'} f_e(z_{j+1}) dz_{j+1} \prod_{z_{q2+1} \leq i \neq j+1} f_e(z_i) dz_i
\]

\[
\leq (1 + n^{-\tilde{\alpha}}) \int_{z \in \mathcal{L}_{j,0}'} f_e(z_j) dz_j \prod_{z_{q2+1} \leq i \neq j} f_e(z_i) dz_i + Cn^{-2}
\]

\[
\leq (1 + n^{-\tilde{\alpha}})^2 \int_{z \in \mathcal{L}_{j,0}'} f_e(z_i) dz_i + Cn^{-2}
\]

\[
\leq (1 + n^{-\tilde{\alpha}})^2 P((S_i^\ast, -n_{q2} + 1 \leq i \leq n_{q2}) \in \mathcal{L}_j) + Cn^{-2}.
\]

For all large \( n \), there exist \( \tilde{\alpha} > \alpha > 0 \) such that

\[
0 \leq P((S_i^\ast, -n_{q2} + 1 \leq i \leq n_{q2}) \in \mathcal{L}_j) - P((S_i^\ast, -n_{q2} + 1 \leq i \leq n_{q2}) \in \mathcal{L}_j) \leq 2Cn^{-\alpha} + Cn^{-2}.
\]

(124)

For the second and third inequalities, we truncate \( j \)-th coordinate by \( n^{s_1} \), and then apply the Chebyshev’s inequality to bound the tailed probability(\( Cn^{-2} \); since we have the moment assumption, \( n^{-k} \) for any \( k > 0 \) is possible). Finally, there is \( 1 > \omega > 0 \) such that

\[
\left| P((S_i^\ast, -n_{q2} + 1 \leq i \leq n_{q2}) \in \mathcal{L}_j) - P(j = j_3) \right| \leq C\omega^{n_{q2}} = o(n^{-1}).
\]

(125)

By (108), (109), (113), (115) to (125), we finish the proof. \( \square \)

The following is the notation used in the main proof; it’s just another decomposition trick.

**Expression for** \( P_j^0 - P_0^0 \)

The following expression is based on \( j > 0, j \geq 0 \) can be dealt in the same fashion.

\[
P_j^0 - P_0^0 = \mathcal{Z}_j^0 - \mathcal{Z}_0^0 =
\]

\[
\sum_{i \in J_{j,0}} x_i^{q2} \left[ (a_1 - \hat{a}_{j,1})^2 - (a_1 - \hat{a}_{j,1,0})^2 \right] + 2 \sum_{i \in J_{j,1,0}} x_i^{q2} \left[ (a_1 - \hat{a}_{j,1}) - (a_1 - \hat{a}_{j,1,0}) \right]
\]

\[
+ \sum_{i \in J_{j,2,0}} x_i^{q2} \left[ (a_2 - \hat{a}_{j,2})^2 - (a_2 - \hat{a}_{j,2,0})^2 \right] + 2 \sum_{i \in J_{j,2,0}} x_i^{q2} \left[ (a_2 - \hat{a}_{j,2}) - (a_2 - \hat{a}_{j,2,0}) \right]
\]

\[
- \sum_{i=1}^j x_i^{q2} \left( a_2 - \hat{a}_{j,2,0} \right)^2 - 2 \sum_{i=1}^j x_i^{q2} \left( a_2 - \hat{a}_{j,2,0} \right)
\]

72
Claim 1

Proof of (114). There exists $C > 0$ such that for $1 - n_{q2} \leq j \leq n_{q2}$, all large $n$,

Claim 1

\[
P \left( \max_{1 - n_{q2} \leq j \leq n_{q2}} \{ |(I)_j|, \ldots, |(VI)_j|, |(IX)_j|, \ldots, |(XI)_j| \} \geq n^{-(1/2 + 24\Delta + \Delta)} \right) \leq C n^{-\Delta \frac{q}{2}}. \tag{127} \]

Claim 2

\[
P \left( |(VII)_j| \geq n^{12\Delta - 1 + \Delta} \right) \leq C n^{-\Delta \frac{q}{2}}, \tag{128} \]

\[
P \left( |(VIII)_j| \geq n^{18\Delta - 1 + \Delta} \right) \leq C n^{-\Delta \frac{q}{2}}. \tag{129} \]

Proof of Claim 1. Assume $E|e_2|^{4q} < \infty$, and we prove the case $n_{q2} \geq j > 0$. The below we list the required moment bounds. There exists $C > 0$ such that for all $0 < j \leq n_{q2}$, $k = 1, 2$, $0 < b < 1 - 6\Delta$, $q \geq 1$,

\[
E \left| \sum_{i=1}^{j} x_{S_i}^o e_{S_i+1}^o \right|^q < C; \tag{129} \]

Expression for $a_1 - \hat{a}_{J_{1,j}}, j > 0$

\[
a_1 - \hat{a}_{J_{1,j}} = \left( \sum_{i \in J_{1,j}} x_i^2 \right)^{-1} \left( \sum_{i \in J_{1,0}} x_i^o e_i^o + (a_1 - a_2) \sum_{i=1}^{j} x_{S_i}^o e_{S_i+1}^o + \sum_{i=1}^{j} x_{S_i}^o e_{S_i+1}^o \right),
\]

\[
a_2 - \hat{a}_{J_{2,j}} = \left( \sum_{i \in J_{2,j}} x_i^2 \right)^{-1} \left( \sum_{i \in J_{2,0}} x_i^o e_i^o + \sum_{i=1}^{j} x_{S_i}^o e_{S_i+1}^o \right),
\]

\[
(a_1 - \hat{a}_{J_{1,j}})^2 - (a_1 - \hat{a}_{J_{1,0}})^2 = (\hat{a}_{J_{1,0}} - \hat{a}_{J_{1,j}}) \times (2a_1 - \hat{a}_{J_{1,j}} - \hat{a}_{J_{1,0}})
\]

\[
= (a_1 - \hat{a}_{J_{1,j}} + a_1 - \hat{a}_{J_{1,0}}) \times \left( \sum_{J_{1,0}} x_i^2 \right)^{-1} \times \left( \sum_{J_{1,0}} x_i^2 \right)^{-1}
\]

\[
\times \left( \sum_{i=1}^{j} x_{S_i}^o \right) \left( \sum_{i \in J_{1,0}} x_i^o e_i^o \right) + \left( \sum_{i \in J_{1,0}} x_i^o \right) \left( a_1 - a_2 \right) \sum_{i=1}^{j} x_{S_i}^o + \sum_{i=1}^{j} x_{S_i}^o e_{S_i+1}^o \right). \tag{126} \]

The Remaining Proof for Proposition 7

Proof of (114). There exists $C > 0$ such that for $1 - n_{q2} \leq j \leq n_{q2}$, all large $n$,
$E \left| n^{-1/2} \sum_{i \in J_{k,j}} x_i^o e_i^o \right|^q < C$; \hspace{1cm} (130)

$E \left| n^{-1} \sum_{i \in J_{k,j}} x_i e_i^o \right|^{+q} < C$; \hspace{1cm} (131)

$E \left| n^{-(1-b)} \sum_{i=1}^{j} x_i^o \right|^q < C$. \hspace{1cm} (132)

**Proof of 129.** Since $q \geq 1$, we can reduce Lemma 16 to for all $0 < k \leq n_{q^2}$,

$$C \geq E \left| n^{-(1-b)} n^{-(1-b) \frac{1}{2}} \sum_{i=1}^{k} x_i^o e_{S_{i+1}}^o \right|^q \geq E \left| n^{-\frac{12\Delta}{q}} \sum_{i=1}^{k} x_i^o e_{S_{i+1}}^o \right|^q.$$

We sacrifice some power on $n$ in order to have a clear expression. \hfill ■

**Proof of 132.** By triangular inequality, $E|x_1|^{4q} < \infty$, (77), there exist $C > 0$ such that for all $0 < j \leq n_{q^2}$,

$$E \left| n^{-(1-b)} \sum_{i=1}^{j} x_i^o \right|^q \leq \left\{ n^{-(1-b)} \sum_{i=1}^{j} E_{\frac{q}{4}} |x_i^o|^{2q} \right\}^q \leq C.$$

\hfill ■

**Proof of 130, 131.** By the definition of $x_i^o$’s and the same method as in Proposition 5. \hfill ■

By these moment bounds and routine arguments, we can finish the proof. \hfill ■

**Proof of Claim 2.** By triangular and Cauchy-Swartz inequality and Lemma 12, there exists $C > 0$ such that for all $0 < j \leq n_{q^2}$,

$$E_{\frac{q}{4}} \left| x_j^o - r \right|^q \leq E_{\frac{q}{4}} \left| x_j^o \mathbf{1}_{q^{r, r+n-(1-\Delta)} \leq n_{q^2}} \right|^q + n^{-(1-6\Delta)} P_{\frac{q}{4}} \left( q^{r, r+n-(1-6\Delta)} > n_{q^2} \right)$$

$$\leq E_{\frac{q}{4}} \left| \sum_{i=1}^{n} x_i^o \right|^q \leq \frac{1}{q^2} \left( \frac{1}{q^2} \right) \leq n^{-(1-6\Delta)}$$

Assume $E|e_2|^{4q} < \infty$. There exists $C > 0$ such that for all $0 < j \leq n_{q^2}$,

$$E \left| \sum_{i=1}^{j} e_{S_{i+1}}^o (x_{S_i}^o - r) \right|^q \leq \left\{ \sum_{i=1}^{j} E_{\frac{q}{4}} \left| e_{S_{i+1}}^o (x_{S_i}^o - r) \right|^q \right\}^q$$

$$\leq \left\{ \sum_{i=1}^{j} E_{\frac{q}{4}} \left| e_{S_{i+1}}^o \right|^q E_{\frac{q}{4}} \left| x_{S_i}^o - r \right|^{2q} \right\}^q$$

$$\leq \left\{ n_{q^2} (C n^6 \Delta)^{\frac{1}{q^2}} n^{-(1-6\Delta)} \right\}^q \leq C n^{(18\Delta - 1)q},$$

74
by triangular (not Jensen) and Cauchy-Swartz inequality, Lemma 18, the previous result.

\[ E \left( \sum_{i=1}^{\frac{j}{2}} (x_{S_i}^0 - r)^2 \right)^q \leq \left\{ \sum_{i=1}^{\frac{j}{2}} E^{\frac{1}{q}} |x_{S_i}^0 - r|^{2q} E^{\frac{1}{q}} |x_{S_i}^0 + r|^{2q} \right\}^q \leq Cn^{(12\Delta - 1)q}. \]

by almost the same argument as above and Lemma 18. And note that these two moment bounds and Chybechev’s inequality finish the proof.

By these two claims we have (114).

\[ \text{Proof of Proposition 1} \]

\[ \text{Proof of Proposition 1.} \] Let \( \Delta \in \mathcal{D} \mathcal{E}^1_{\infty} \) and \( E|e|^q < \infty \) for sufficiently large \( q \); define the threshold estimator based on \( x_1, \ldots, x_{n-\log n} \) as \( \tilde{r}_n \) which is analogous to \( \hat{r}_n \). By the Corollary 6, for all large \( n \),

\[ P(U_n^c) \leq C(\log n)n^{-\Delta\left(\frac{1}{n}\right)+1}. \]

Hence, it suffices to prove

\[ |nP(x_n \in B_{\tilde{r}_n - r}(r)) - nP(x_n \in B_{\hat{r}_n - r}(r))| = o(1). \] (133)

\[ |nP(x_n \in B_{\hat{r}_n - r}(r)) - nP(x_o \in B_{\hat{r}_n - r}(r))| = o(1). \] (134)

\[ |nP(x_o \in B_{\hat{r}_n - r}(r)) - nP(x_o \in B_{\tilde{r}_n - r}(r))| = o(1). \] (135)

Intuitively, (133) is true because the difference for the events \( x_n \in B_{\tilde{r}_n - r}(r) \) and \( x_n \in B_{\hat{r}_n - r}(r) \) appears when \( x_{n-\log n+1}, \ldots, x_{n-1} \) fall inside a small neighbor of \( r \). This account for a decreasing order down almost to \( \frac{1}{n} \), and the other part of the decreasing order is due to \( x_n \) itself fall inside a small neighbor of \( r \). (134) is mainly due to the \( \log n \)-asymptotically independent result, and (135) is similar to (133). The proof of (133) to (135) can be found below.

\[ \text{Proof of (133).} \] Let \( \hat{r}_n' \) be a version of \( \tilde{r}_n \) with only the subsample of size \( n - c \log n \) used for estimation. By Lemma 12, there exists \( c_1', c_1 \) in \( n q_2, c, \delta > 0 \) such that for

\[ A = \{ x_n \in B_{\tilde{r}_n - r}(r) \}, \]

\[ B = \{ x_n \in B_{\hat{r}_n - r}(r) \}, \]

\[ D_1 = \{ \text{More than } n q_2 \text{ } x_i', i = 1, \ldots, n, \text{ fall inside } B_{c_1'-n^{-(1-\delta)}}(r) \}, \]

and all large \( n \),

\[ P(D_1) \leq \exp(-cn^\delta). \]

(133) is lesser than or equivalent to

\[ nP(\mathcal{A} \cap \mathcal{B}^c) \cup (\mathcal{A}^c \cap \mathcal{B}) \leq \left\{ n \left\{ P(\hat{r}_n - r) \geq c_1'n^{-(1-6\Delta)} \right\} + P(\hat{r}_n - r) \geq c_1'n^{-(1-6\Delta)} \right\} + P(D_1) \]

\[ + P((\mathcal{A} \cap \mathcal{B}^c) \cap (\mathcal{A}^c \cap \mathcal{B})_s; x_n \in B_{c_1'-n^{-(1-6\Delta)}}(r); \left| \hat{r}_n - r \right|, \left| \tilde{r}_n - r \right| \leq c_1'n^{-(1-6\Delta)}; D_1^c) \}; \] (136)
this is because if \( x_n \notin B_{c_1 n^{-1-6\Delta}}(r) \) and \(|\tilde{r}_n - r|, |\tilde{r}'_n - r| \leq c_1' n^{-(1-6\Delta)}\), then \( B^c \cap A^c \) is true. Let

\[
C_1 = \{ n_{q_2} \geq \arg \min_{-n+1 \leq i \leq n} \hat{P}_i \geq -n_{q_2} + 1 \},
\]

\[
C_2 = \{ n_{q_2} \geq \arg \min_{-n+c\log n+1 \leq i \leq n-c\log n} \hat{P}'_i \geq -n_{q_2} + 1 \},
\]

where \( \hat{P}'_i \)'s are defined the same as \( \hat{P}_i \)'s but consist of only \( x_1, \ldots, x_{n-c\log n} \). On \( \{ |\tilde{r}_n - r|, |\tilde{r}'_n - r| \leq c_1' n^{-(1-6\Delta)} \} \cap D'_1 \), for \( n \) large enough such that \( c_1' n^{-(1-6\Delta)} < r \), the third term on the RHS of (136) equals to

\[
P(x_n \in B_{c_1' n^{-(1-6\Delta)}}(r); (A \cap B^c) \cup (A^c \cap B); |\tilde{r}_n - r|, |\tilde{r}'_n - r| \leq c_1' n^{-(1-6\Delta)}; C_1; C_2; D'_1)
\]

\[
\leq P(x_n \in B_{c_1' n^{-(1-6\Delta)}}(r); \tilde{r}_n \neq \tilde{r}'_n; |\tilde{r}_n - r|, |\tilde{r}'_n - r| \leq c_1' n^{-(1-6\Delta)}; C_1; C_2).
\]

(137)

Note that we do not make the probability integrate over the event \( R\mathcal{E}_n \) to save the text simplicity; consequently on the rare event, confusion arises, and we shall circumvent this problem by directly redefining \( \arg \min \) to count out such rare events. In the following derivation the reader keeps in mind that any value comparison with a quantity of \( \arg \min \) over all infinite numbers involved results in an event of zero probability; \( P(R\mathcal{E}_n) \) is added and absorbed by \( o(1) \) in the end of the analysis to finish the proof. Define

\[
C_3 = \{ \arg \min_{-n+1 \leq i \leq n} \hat{P}_i \neq \arg \min_{-n+c\log n+1 \leq i \leq n-c\log n} \hat{P}'_i \},
\]

\[
C_4 = \{ \arg \min_{-n_{q_2}+1 \leq i \leq n_{q_2}} \hat{P}_i \neq \arg \min_{-n_{q_2}+1 \leq i \leq n_{q_2}} \hat{P}'_i \}
= \{ \arg \min_{-n_{q_2}+1 \leq i \leq n_{q_2}} S_i + R_i \neq \arg \min_{-n_{q_2}+1 \leq i \leq n_{q_2}} S'_i + R'_i \},
\]

\[
C_5 = \{ \arg \min_{-n_{q_2}+1 \leq i \leq n_{q_2}} S'_i + R'_i \neq \arg \min_{-n_{q_2}+1 \leq i \leq n_{q_2}} S'_i + R'_i \},
\]

where \( S_j \)'s are defined in the same way in Proposition 7; the definition of \( S'_j \) follows that of \( S_j \) but that the subsample insted of full sample is used for construction; moreover, \( R_j = \hat{P}_j - S_j, R'_j = \hat{P}'_j - S'_j \). We have already used the notation \( S'_j \), but it shall result in no confusion. The null analog are constructed in the same way:

\[
C_6 = \{ \arg \min_{-n_{q_2}+1 \leq i \leq n_{q_2}} S'_i + R'_i \neq \arg \min_{-n_{q_2}+1 \leq i \leq n_{q_2}} S'_i + R'_i \}
\]

\[
\Gamma = \{ \sup_{-n_{q_2}+1 \leq i \leq n_{q_2}} |R'_i - R'_0|, \sup_{-n_{q_2}+1 \leq i \leq n_{q_2}} |R'_0 - R'_i| \leq n^{-(\frac{1}{2} - 25\Delta)} \},
\]

\[
D_2 = \{ \text{Less than } n_{q_2} \text{ and absorbed by } o(1) \},
\]

where \( c_2, c, \delta > 0 \) are such that for all large \( n \), \( P(D_2) \leq \exp(-cn^\delta) \). Let \( \Pi' \) be the analog of \( \Pi \) based on only \( x_i, 1 \leq i \leq n - c\log n \). The RHS of (137) is lesser than or equivalent
to
\[
P(x_n \in B_{c_1^n(1-6\Delta)}(r); \mathcal{C}_3; |\tilde{r}_n - r|, |\tilde{r}_n' - r| \leq n^{-1-6\Delta}; \mathcal{C}_1; \mathcal{C}_2) \\
+ P(x_n \in B_{c_1^n(1-6\Delta)}(r); \tilde{r}_n' \neq \tilde{r}_n'; \mathcal{C}_3; |\tilde{r}_n - r|, |\tilde{r}_n' - r| \leq n^{-1-6\Delta}; \mathcal{C}_1; \mathcal{C}_2) \\
\leq P(x_n \in B_{c_1^n(1-6\Delta)}(r); \mathcal{C}_4) + \sum_{i=1}^{c \log n - 1} P(x_n, x_{n-i} \in B_{c_1^n(1-6\Delta)}(r)) \\
\leq P(x_n \in B_{c_1^n(1-6\Delta)}(r); \mathcal{C}_4; \mathcal{D}_2) + P(\mathcal{D}_2) + \sum_{i=1}^{c \log n - 1} P(x_n, x_{n-i} \in B_{c_1^n(1-6\Delta)}(r)) \\
\leq P(x_n \in B_{c_1^n(1-6\Delta)}(r); \{n_{q_2}, -n_{q_2} + 1\} \in \Pi') + P(\mathcal{D}_2) \\
+ 2 \sum_{i=1}^{c \log n - 1} P(x_n, x_{n-i} \in B_{c_1^n(1-6\Delta)}(r)) + P(\{n_{q_2}, -n_{q_2} + 1\} \not\in \Pi') \\
\leq \sum_{j=-n_{q_2} + 1}^{n_{q_2}} P(x_n \in B_{c_1'n(1-6\Delta)}(r); (S'_i, -n_{q_2} + 1 \leq i \leq n_{q_2}) \in \hat{L}_j \setminus L_j; \{n_{q_2}, -n_{q_2} + 1\} \in \Pi') \\
+ 2P(\mathcal{C}_5 \Delta \mathcal{C}_5') + P(\Gamma_0) + P(\mathcal{D}_2) + 2 \sum_{i=1}^{c \log n} P(x_n, x_i \in B_{n^{-1-6\Delta}}(r)) + P(\{n_{q_2}, -n_{q_2} + 1\} \not\in \Pi') \\
= \sum_{j=-n_{q_2} + 1}^{n_{q_2}} P(x_n \in B_{c_1'n(1-6\Delta)}(r); (S'_i, -n_{q_2} + 1 \leq i \leq n_{q_2}) \in \hat{L}_j \setminus L_j; \{n_{q_2}, -n_{q_2} + 1\} \in \Pi') + o(n^{-1}); \\
(138)
\]

for the first inequality, we note that on \(\{\tilde{r}_n' \neq \tilde{r}_n'; \mathcal{C}_3; |\tilde{r}_n - r|, |\tilde{r}_n' - r| \leq c_1'n^{-1-6\Delta}\}\), we have \(x_{n-j} \in B_{c_1'n(1-6\Delta)}(r)\) for some \(1 \leq j < c \log n\), and \(\mathcal{C}_1 \cap \mathcal{C}_2 \cap \mathcal{C}_3 \subset \mathcal{C}_4\). The Third inequality follows from the fact that on \(\mathcal{D}_5\), if \(S_i \neq S'_i\), for some \(-n_{q_2} + 1 \leq i \leq n_{q_2}\), then some \(x_i, n \geq i > n - c \log n\) are falling inside \(B_{c_2n^{-1}(1-\Delta)}(r)\). The last inequality is similar to step 1 in Proposition 7, and we skip the detail of the step replacing \(x_i\) with \(x_i^0\), and then replacing it back to \(x_i\). For \(-n_{q_2} + 1 \leq j \leq n_{q_2}\), let
\[
\mathcal{C}_{i,j} = \{(S'_i, -n_{q_2} + 1 \leq i \leq n_{q_2}) \in \hat{L}_j \setminus L_j; \{n_{q_2}, -n_{q_2} + 1\} \in \Pi'\}
\]
For any \(-n_{q_2} + 1 \leq j \leq n_{q_2}\), we have this trick to isolate \(x_n\):
\[
|P(x_n \in B_{c_1'n(1-6\Delta)}(r); \mathcal{C}_{i,j}) - P(x_o \in B_{c_1'n(1-6\Delta)}(r); \mathcal{C}_{i,j})| \\
\leq E \left[ P(x_n \in B_{c_1^n(1-6\Delta)}(r); \mathcal{C}_{i,j} \setminus \mathcal{F}_{n \log n}) - P(x_o \in B_{c_1^n(1-6\Delta)}(r); \mathcal{C}_{i,j} \setminus \mathcal{F}_{n \log n}) \right] \\
\leq C\rho^{c \log n}(1 + n^2) + Cn^{-2q+1},
\]
these two terms come from evaluating the expectation on \(\{|x_n^{c \log n}| \leq n^2\}\) using (43) and on \(\{\sup_{1 \leq i \leq n} |x_i| \geq n^2\}\), \(P(.) \leq 1\). Finally, by the step 3 in Proposition 7, there
exists \( C > 0 \) such that for all large \( n \),
\[
\sum_{j=-n_{q2}+1}^{n_{q2}} P(x_o \in B_{c_j' n^{-(1-6\delta)}}(r); C_{6,j}) \leq \sum_{j=-n_{q2}+1}^{n_{q2}} P(x_o \in B_{c_j' n^{-(1-6\delta)}}(r)) \times (A_1 - A_2) \\
\leq Cn_{q2} \bar{M} c_j' n^{-(1-6\delta)} n^{-\alpha},
\]
(140)

where
\[
A_1 = P((\mathcal{S}_i', -n_{q2} + 1 \leq i \leq n_{q2}) \in \mathcal{L}_j; \{n_{q2}, -n_{q2} + 1\} \subset \Pi'), \\
A_2 = P((\mathcal{S}_i', -n_{q2} + 1 \leq i \leq n_{q2}) \in \mathcal{L}_j; \{n_{q2}, -n_{q2} + 1\} \subset \Pi').
\]

By (136) to (140), we have (133).

Note that (134) can be proved by the method in (139); (135) can be proved in a way similar to (133).

**Statement and Proof of Proposition 23**

**Proposition 23.** Assume assumption 1, 2, 3, 4, 5 and let \( 0 < \alpha < \bar{\alpha}, \Delta \in \mathcal{D} \mathcal{E}^1_{\infty} \). There exists \( \tilde{q} > 0, C > 0 \) such that for all large \( n \),
\[
\sup_{x \in \mathbb{R}^1} \left| F_n(r_{n-r}(x) - F_{r_{\infty}}(x) \right| \leq C n^{-\tilde{q}}.
\]

**Proof of Proposition 23.** Due to Corollary 6, it suffices to prove, for all large \( n \),
\[
\sup_{x \in \mathbb{R}^1} \left| F_n(r_{n-r}(x) - F_{r_{\infty}}(x) \right| \leq C n^{-\tilde{q}}.
\]

We have
\[
\sup_{x \in \mathbb{R}^1} \left| F_n(r_{n-r}(x) - F_{r_{\infty}}(x) \right| \leq \sup_{\beta \equiv [a,b] \subset \mathbb{R}^1} \left| P(n(\tilde{r}_n - r) \in \beta) - P(r_{\infty} \in \beta) \right|,
\]

where \( \mathbb{R}^1 \) denotes extended real line. Note that
\[
P(n(\tilde{r}_n - r) \in \beta) = \sum_{j=-n+1}^{n} P(\tilde{r}_n = x_{S_j^*}; n(x_{S_j^*} - r) \in \beta, j \in \Pi) + P(\tilde{r}_n = 0; -nr \in \beta),
\]
\[
P(r_{\infty} \in \beta) = \sum_{i=1}^{0} P(r_{\infty} = p_{2,i}; p_{2,i} \in \beta) + \sum_{i=-\infty}^{0} P(r_{\infty} = p_{1,-i+1}; p_{1,-i+1} \in \beta),
\]

where \( p_{k,i} \)'s are defined in notation section; \( n_{q2} = c_1 n^{6\Delta} \). Pick \( c_1' \) small enough such that given \( c_1 \), there exist some \( c, \delta > 0 \) such that for all large \( n \),
\[
P(q^{r_{n+c_1' n^{-(1-6\delta)}}} \geq n_{q2}) \leq \exp (-cn^{\delta})
\]

Without loss of generality, we consider for any \( \beta = [a,b] \subset \mathbb{R}^1, b > a \geq r, \) any \( \nu_1 > 0, \)
\[
|P(n(\tilde{r}_n - r) \in \beta) - P(r_{\infty} \in \beta)\right| \leq \sum_{i=1}^{\nu_1 \log n} |P(\tilde{r}_n = x_{S_j^*}; n(x_{S_j^*} - r) \in \beta; j \in \Pi) - P(r_{\infty} = p_{2,i}; p_{2,i} \in \beta)|
\]

78
\[ + \sum_{i=-v_1 \log n+1}^{0} |P(\hat{r}_n = x_{S_i^r}; n(x_{S_i^r} - r) \in \beta; j \in \Pi) - P(r_\infty = p_{1,-i+1}; p_{1,-i+1} \in \beta)| \]

\[ + P(\hat{r}_n = 0) \]

\[ + \sum_{i=v_1 \log n+1}^{n} \sum_{i=1}^{n_{q2}} |P(\hat{r}_n = x_{S_i^r}; n(x_{S_i^r} - r) \in \beta; j \in \Pi) - P(r_\infty = p_{2,i}; n(x_{S_i^r} - r) \in \beta)| \]

\[ + P(r_\infty = p_{2,i})|P(n(x_{S_i^r} - r) \in \beta) - P(p_{2,i} \in \beta)| \]

\[ + \sum_{i=-n_{q2}+1}^{-1} \sum_{i=-v_1 \log n}^{n} |P(\hat{r}_n = x_{S_i^r}; n(x_{S_i^r} - r) \in \beta; j \in \Pi) - P(r_\infty = p_{1,-i+1}; n(x_{S_i^r} - r) \in \beta; j \in \Pi)| \]

\[ + \sum_{i=-n_{q2}+1}^{-1} \sum_{i=-v_1 \log n}^{n} P(r_\infty = p_{1,-i+1})|P(n(x_{S_i^r} - r) \in \beta; j \in \Pi) - P(p_{1,-i+1} \in \beta)| \]

\[ + \sum_{i=n_{q2}+1}^{n} P(|\hat{r}_n - r| \geq c_1' n^{-1-6\Delta}) + nP(q^{r,r+c_1' n^{-1-6\Delta}} \geq n_{q2}) \]

\[ + \sum_{i=-n_{q2}+1}^{-1} P(|\hat{r}_n - r| \geq c_1' n^{-1-6\Delta}) + nP(q^{r-r-c_1' n^{-1-6\Delta}, r} \geq n_{q2}) \]

\[ + \sum_{i=-\infty}^{\infty} P(j_3 = i) + \sum_{i=-\infty}^{n_{q2}+1} P(j_3 = i) \]

By Proposition 5, 7, definition of \( c_1' \), some basic probability results, and Corollary 11 for no \( x_i \)'s in \( \Theta' \), there exists \( \bar{q} > 0 \) such that for all large \( n \),

\[ |P(n(\hat{r}_n - r) \in \beta) - P(r_\infty \in \beta)| \leq \sum_{i=1}^{v_1 \log n} |P(\hat{r}_n = x_{S_i^r}; n(x_{S_i^r} - r) \in \beta; j \in \Pi) - P(r_\infty = p_{2,i}; p_{2,i} \in \beta)| \]

\[ + \sum_{i=-v_1 \log n}^{0} |P(\hat{r}_n = x_{S_i^r}; n(x_{S_i^r} - r) \in \beta; j \in \Pi) - P(r_\infty = p_{1,-i+1}; p_{1,-i+1} \in \beta)| + n^{-q}, \]

(141)

In the following we deal with the first summation term in (141), the second summation term can be handled by the same way. By Proposition 7, for all large \( n \),

\[ \sum_{i=1}^{v_1 \log n} |P(\hat{r}_n = x_{S_i^r}; n(x_{S_i^r} - r) \in \beta; j \in \Pi) - P(r_\infty = p_{2,i}; p_{2,i} \in \beta)| \]

\[ \leq \sum_{i=1}^{v_1 \log n} |P(\hat{r}_n = x_{S_i^r}; n(x_{S_i^r} - r) \in \beta; j \in \Pi) - P(j_3 = i)P(n(x_{S_i^r} - r) \in \beta)| \]

(142)

\[ + \sum_{i=1}^{v_1 \log n} |P(j_3 = i)P(n(x_{S_i^r} - r) \in \beta; j \in \Pi) - P(j_3 = i)P(p_{2,i} \in \beta)| \]

\[ \leq n^{-q} + \sum_{i=1}^{v_1 \log n} |P(n(x_{S_i^r} - r) \in \beta; j \in \Pi) - P(p_{2,i} \in \beta)|, \]
For the second term, given any \( \theta_n \subset \mathbb{R}^1 \),

\[
\sum_{i=1}^{v_1 \log n} |P(n(x_{S_i}^c - r) \in \beta; j \in \Pi) - P(p_{2,i} \in \beta)|
\leq \sum_{i=1}^{v_1 \log n} |P(n(x_{S_i}^c - r) \in \beta \cap \theta_n; j \in \Pi) - P(p_{2,i} \in \beta \cap \theta_n)|
\quad (143)
\]

\[
+ \sum_{i=1}^{v_1 \log n} |P(n(x_{S_i}^c - r) \in \beta \cap \theta_n^c; j \in \Pi) - P(p_{2,i} \in \beta \cap \theta_n^c)|.
\]

Now we control the second term in (143). The following two statements, (144) and (145), extends the results of polynomial order in Lemma 12 to logarithmic order. Define \( \theta_n \equiv [0, c_1 \log n] \).

\[
P(q_{v_1 \log n} \leq v_1 \log n) \leq P( \text{Less than } v_1 \log n x_i \text{'s, } i = 1, \ldots, (1-3\rho_1)n \in \theta_n \big| E_{\rho_1}^c) + P(E_{\rho_1}^c)
\leq \sum_{j=0}^{v_1 \log n} C_j (1-3\rho_1)n (1-c_1 m \log n \theta_n^{(1-3\rho_1)n-j})^j (c_1 m \log n \theta_n - j)^j + P(E_{\rho_1}^c)
\equiv \left( \sum_{j=0}^{v_1 \log n} W_j^1 \right) + P(E_{\rho_1}^c) \leq \left( \frac{1}{2} \right)^{v_1 \log n - 1} + P(E_{\rho_1}^c);
\quad (144)
\]

\[
P(p_{2,v_1 \log n} \leq c_1 \log n) \leq \sum_{j=0}^{v_1 \log n} \exp(-c_1 \log n \pi(r)) \frac{(c_1 \log n \pi(r))^j}{j!}
\equiv \left( \sum_{j=0}^{v_1 \log n} W_j^2 \right) \leq \left( \frac{1}{2} \right)^{v_1 \log n - 1};
\quad (145)
\]

the last steps in (144), (145) follow because given \( v_1 \), we pick \( c_1 \) large enough such that for all \( 1 \leq j \leq 2v_1 \log n \), all large \( n \),

\[
\frac{W_{j-1}^1}{W_j^1} = \frac{j}{n - j + 1} \left( \frac{1 - c_1 m \log n \theta_n}{c_1 m \log n \theta_n} \right) \leq \frac{1}{2},
\]

\[
\frac{W_{j-1}^2}{W_j^2} = \frac{j}{c_1 \log n \pi(r)} \leq \frac{1}{2}.
\]

By (144), (145), we have for \( 1 \leq j \leq v_1 \log n \),

\[
\sum_{i=1}^{v_1 \log n} |P(n(x_{S_i}^c - r) \in \beta \cap \theta_n^c; j \in \Pi) - P(p_{2,i} \in \beta \cap \theta_n^c)| \leq C v_1 \log n \left[ P(E_{\rho_1}^c) + \left( \frac{1}{2} \right)^{v_1 \log n - 1} \right].
\quad (146)
\]

Before we turn to the first term in (143), we define \( q^\theta_p \) and \( v_2 \): for \( \theta \subset \mathbb{R}^1, i > 0 \), define \( q^\theta_p \) such that

\[
q^\theta_p = | \{ i : p_{2,i} \in \theta; i > 0 \} |;
\]
given \( v_2 \geq v_1 \) large enough such that for all large \( n \),
\[
P(q_{\frac{[r,b]\cap\theta_n}{n}} > v_2 \log n) + P(q_{\frac{[r,a]\cap\theta_n}{n}} > v_2 \log n) \leq n^{-\tilde{q}}; \tag{147}
\]
by an arguments similar to (144) and (145), the existence of \( v_2 \) is guaranteed. To deal with the first term in (143), we note that for \( 1 \leq j \leq v_1 \log n \leq v_2 \log n \),
\[
P(n(x_{S_j}^t - r) \in \beta \cap \theta_n; j \in \Pi) - \sum_{s=j}^{v_2 \log n} \sum_{t=0}^{j-1} P(q_{\frac{[r,b]\cap\theta_n}{n}} = s - t; q_{\frac{[r,a]\cap\theta_n}{n}} = t) \leq \sum_{s=v_2 \log n+1}^{n} P(q_{\frac{[r,b]\cap\theta_n}{n}} = s). \tag{148}
\]
\[
P(p_{2j} \in \beta \cap \theta_n) - \sum_{s=j}^{v_2 \log n} \sum_{t=0}^{j-1} P(q_{\frac{[r,b]\cap\theta_n}{n}} = s - t; q_{\frac{[r,a]\cap\theta_n}{n}} = t) \leq \sum_{s=v_2 \log n+1}^{n} P(\bar{q}_p = s). \tag{149}
\]
Claim:

There exists some \( h > 0 \) such that for \( 0 \leq t < s \leq v_2 \log n \), all large \( n \),
\[
|P(q_{\frac{[r,b]\cap\theta_n}{n}} = s - t; q_{\frac{[r,a]\cap\theta_n}{n}} = t) - P(q_{\frac{[r,b]\cap\theta_n}{n}} = s - t; q_{\frac{[r,a]\cap\theta_n}{n}} = t)| \leq n^{-h}
\]
By this claim, choosing \( \tilde{q} = h - \varepsilon \) for some small \( \varepsilon > 0 \), (147) to (149), for all large \( n \),
\[
\sum_{i=1}^{v_1 \log n} |P(n(x_{S_i}^t - r) \in \beta \cap \theta_n; j \in \Pi) - P(p_{2i} \in \beta \cap \theta_n)| \leq n^{-\tilde{q}} \tag{150}
\]
By (141) to (150), we have finished the proof. \( \blacksquare \)

Proof of the Claim. We prove for \( 0 \leq t < s \leq v_2 \log n \), all large \( n \),
\[
P(q_{\frac{[r,b]\cap\theta_n}{n}} = s - t; q_{\frac{[r,a]\cap\theta_n}{n}} = t) \leq P(q_{\frac{[r,b]\cap\theta_n}{n}} = s - t; q_{\frac{[r,a]\cap\theta_n}{n}} = t) + n^{-h},
\]
the other side can be proven by the same method. Let
\[
\begin{align*}
\theta_0^1 &= \theta_0^2 = \left( \frac{[r,b] \cap \theta_n}{n} \right)^c \cap [-n^v, n^v], \\
\theta_1^1 &= \left( \frac{[r,b] \cap \theta_n}{n} \right) \cap [-n^v, n^v] = \frac{[r,b] \cap \theta_n}{n}, \text{ for all large } n, \\
\theta_2^1 &= \left( \frac{[r,a] \cap \theta_n}{n} \right) \cap [-n^v, n^v] = \frac{[r,a] \cap \theta_n}{n}, \text{ for all large } n, \\
\theta_0^2 &= \left( \frac{[a,b] \cap \theta_n}{n} \right) \cap [-n^v, n^v] = \frac{[a,b] \cap \theta_n}{n}, \text{ for all large } n,
\end{align*}
\]
where \( v \) is defined in assumption 4. Given \( 1 \leq k \leq v_2 \log n \),
\[
\Omega_k^1 = \cup_{t_i,i=1,\ldots,n} G_{1,k} \theta_0^1 \times \cdots \theta_0^1,
\]
\( G_{1,k} \) includes all possible elements such that if \( \{t_i, i = 1, \ldots, n\} \in G_{1,k} \), then \( t_i \)’s satisfies
\( i \) \( t_i \in \{0, 1\} \) for all \( i \), \( ii \) \( 1 \leq \sum_{i=1}^{n} t_i = k \leq v_2 \log n \),
\( \Omega_k^1 \) is defined as \( \Omega_k^1 \) but with \( \theta_0^1 \) being replaced by \( \left( \frac{[r,b]\cap\theta_n}{n}\right)^c \).
Given $1 \leq k \leq v_2 \log n$, fixing a subset $\{u_1, \ldots, u_k\}$ such that $u_i \in \{2, 3\}$ for all $1 \leq i \leq k$ and $k \geq t = |\{i : u_i = 2\}| \geq 0$. Let

$$\Omega_k^2 = \cup_{\{t_i, i = 1, \ldots, n\} \in G_{2k}} \theta_{t_1}^2 \times \cdots \times \theta_{t_n}^2,$$

$G_{2,k}$ includes all possible elements such that if $\{t_i, i = 1, \ldots, n\} \in G_{2,k}$, $t_i$’s satisfies

(iii) $t_i \in \{0, 2, 3\}$, for all $i$,

(iv) $|\{i : t_i \neq 0\}| = k \leq v_2 \log n$,

and given $\{t_1, \ldots, t_n\}$ satisfying (iii), (iv), let $\{a_i\}_{i=1}^k$ be distinct acendeing indexes such that $t_{a_i} \neq 0, i = 1, \ldots, k$,

(v) $t_{a_i} = u_i, i = 1, \ldots, k$.

$\Omega_k^2$ is defined in the same way as $\Omega_k^2$ but with $\theta_0^2$ being replaced by $(\frac{r_k}{n})^c$. Note that both $\Omega_k^2, \Omega_k^2$ depend on $\{u_1, \ldots, u_s\}$, and $\Omega_0^2 = \Omega_0^1$. Given $k > 0$ and a fixed subset $\{u_1, \ldots, u_k\}$, then for each $\{t_1, \ldots, t_n\} \in G_{2,k}$, there is a corresponding $\{t'_1, \ldots, t'_n\} \in G_{1,k}$ such that

$$n - k = |\{i : t'_i = t_i = 0\}|.$$

Claim: Given any $1 \leq k \leq v_2 \log n$, a fixing subset $\{u_1, \ldots, u_k\}$ with $k \geq t = |\{i : u_i = 2\}|$ in $\Omega_k^2$,

$$P((x_1, \ldots, x_n) \in \Omega_k^2) = P((x_1, \ldots, x_n) \in \Omega_k^1) \times (1 + n^{-\gamma}c)^{2v_2 \log n} \times \frac{|\theta_2^2| |\theta_3^2|^{k-t}}{|\theta_2^2 \cup \theta_3^2|^k} + C \times \exp(-n^{wv}) \times n,$$

$$P((x_1, \ldots, x_n) \in \Omega_k^2) \geq P((x_1, \ldots, x_n) \in \Omega_k^1) \times (1 - n^{-\gamma}c)^{2v_2 \log n} \times \frac{|\theta_2^2| |\theta_3^2|^{k-t}}{|\theta_2^2 \cup \theta_3^2|^k} - C \times \exp(-n^{wv}) \times n$$

Proof of the Claim. We consider each $\{t_1, \ldots, t_n\} \in G_{2,k}$ and the corresponding $\{t'_1, \ldots, t'_n\} \in G_{1,k}$, then by the assumption 4 and the usage of a similar argument in Lemma 9, we have

$$P((x_1, \ldots, x_n) \in \theta_{t_1}^2 \times \cdots \times \theta_{t_n}^2) \leq P((x_1, \ldots, x_n) \in \theta_{t'_1}^2 \times \cdots \times \theta_{t'_n}^2) \times (1 + n^{-\gamma}c)^{2v_2 \log n} \times \frac{|\theta_2^2| |\theta_3^2|^{k-t}}{|\theta_2^2 \cup \theta_3^2|^k}$$

Then we sum over all possible $t_i$’s and $t'_i$’s,

$$P((x_1, \ldots, x_n) \in \Omega_k^2) \leq P((x_1, \ldots, x_n) \in \Omega_k^1) \times (1 + n^{-\gamma}c)^{2v_2 \log n} \times \frac{|\theta_2^2| |\theta_3^2|^{k-t}}{|\theta_2^2 \cup \theta_3^2|^k}, \quad (151)$$

To control the difference between $\Omega_k^1$ and $\Omega_k^2$, we note that by the assumption 5, for $k \leq v_2 \log n$,

$$P((x_1, \ldots, x_n) \in \Omega_k^1 \setminus \Omega_k^2) \leq C_{v_2 \log n} \times \exp(-n^{wv}) \times n \times C,$$

$$P((x_1, \ldots, x_n) \in \Omega_k^2 \setminus \Omega_k^1) \leq C_{v_2 \log n} \times \exp(-n^{wv}) \times n \times C, \quad (152)$$

82
since $|G_{i,k}| \leq C_{v_2 \log n}^n, i = 1, 2, k \leq v_2 \log n$, and there are at most $n x_i$'s falling outside $[-n^v, n^v]$.

By (151), (152), we have finished this part of the proof. The same argument can be applied to the inequality on the other side.

By the above claim, we have for $0 \leq t < s \leq v_2 \log n$,

$$P(q_{\frac{n_a}{n}}^n = s - t; q_{\frac{0,n}{n}}^n = t) = \sum_{t<s} P((x_1, \ldots, x_n) \in \Omega_s^1) \leq C_t^s \times P((x_1, \ldots, x_n) \in \Omega_s^1) \times (1 + n^{-\gamma}c)^{2v_2 \log n} \times \frac{[\theta_2^t]^s - [\theta_3^t]^s}{[\theta_3^2 \cup \theta_2^2]^s} + C \times C_t^s \times C_{v_2 \log n}^n \times \exp(-n^{uv}) \times n,$$  

(153)

where the summation, given $t, s$, is over all combination of $\{u_1, \ldots, u_s\}$ with $u_i \in \{2, 3\}$, $t = \{|i : u_i = 2\}|$. Now we turn to the RHS of (153). Since

$$P((x_1, \ldots, x_n) \in \Omega_s^1) = P(q_{\frac{n_a}{n}}^n = s),$$

by the Poisson Approximation lemma, we have for $1 \leq s \leq v_2 \log n$,

$$(153) \leq C_t^s \times P(q_{\frac{n_a}{n}}^n = s) \times (1 + n^{-\gamma}c)^{2v_2 \log n} \times \frac{[\theta_2^t]^s - [\theta_3^t]^s}{[\theta_3^2 \cup \theta_2^2]^s} + C \times C_t^s \times (1 + n^{-\gamma}c)^{2v_2 \log n} \times \frac{[\theta_2^t]^s - [\theta_3^t]^s}{[\theta_3^2 \cup \theta_2^2]^s} \times \frac{\log n^4}{n}$$  

(154)

$$+ C_t^s \times C_{v_2 \log n}^n \times \exp(-n^{uv}) \times n \times C.$$  

Simple calculation show

$$\sum_{j=0}^{s} C_j^s \left( \frac{[\theta_2^t]^j}{[\theta_3^2 \cup \theta_2^2]} \right) \left( \frac{[\theta_3^t]^j}{[\theta_3^2 \cup \theta_2^2]} \right)^{s-j} = 1.$$  

(155)

For $0 \leq t < s \leq v_1 \log n$,

$$C_t^s \leq 2^{v_1 \log n + 1} \leq 2n^{v_1 \log 2}$$  

(156)

For $0 \leq t < s \leq v_1 \log n$,

$$C_t^s \times P(q_{\frac{n_a}{n}}^n = s) \times \frac{[\theta_2^t]^s - [\theta_3^t]^s}{[\theta_3^2 \cup \theta_2^2]^s} = P(q_{\frac{n_a}{n}}^n = t) \times P(q_{\frac{n_a}{n}}^n = s - t).$$  

(157)

For all large $n$,

$$(1 + n^{-\gamma}c)^{2v_1 \log n} \leq \exp(2v_1 \log n \times \ln(1 + n^{-\gamma}c)) \leq (1 + c_2 n^{-\gamma} \log n).$$  

(158)

By (154) to (158), for all large $n$,

$$P(q_{\frac{n_a}{n}}^n = t) \times P(q_{\frac{n_a}{n}}^n = s - t) + C(n^{-\gamma} \log n + o(n^{-1})).$$

The existence of $h > 0$ is now obvious.
Statement and Proof of the Poisson Approximation Lemma

The classical result for Poisson approximation is constructed in the independent and identically distributed process; Chen(1973) extends the result to $m$-dependent process, where the random variables are independent if the indexes of such differ by more than $m$. Using the techniques developed in Chen(1973), Lemma 24 furtherly extends the result to the process possessing the property (43). For $\theta \subset \mathbb{R}^1, i > 0$, let $q^\theta_p$ be such that 

$$q^\theta_p = |\{i : p_{2,i} \in \theta; i > 0\}|.
$$

Lemma 24 (Poisson Approximation Lemma). Given $[a,b] \subset \mathbb{R}^1$, there exists $C > 0$ such that for $s \geq 0$, $\theta_n \equiv [0,c_1 \log n]$, all large $n$,

$$|P(q^{[a,b] \cap \theta_n} = s) - P(q_p^{[a,b] \cap \theta_n} = s)| \leq C\left(\frac{\log n}{n}\right)^3.
$$

Proof of Lemma (24), Poisson Approximation Lemma. By the assumption 4 and 5, for $1 \leq i,j \leq n$, there exists some constant $C > 0$ such that 

$$P(x_i \in [a,b] \cap \theta_n)P(x_j \in [a,b] \cap \theta_n) \leq C\left(\frac{\log n}{n}\right)^2,
$$

$$E(1_{x_i \in [a,b] \cap \theta_n} \times 1_{x_j \in [a,b] \cap \theta_n}) \leq C\left(\frac{\log n}{n}\right)^2.
$$

Hence, for some $C > 0, c_3 > 0$,

$$\sum_{|i-j| \leq c_3 \log n} P(x_i \in [a,b] \cap \theta_n)P(x_j \in [a,b] \cap \theta_n) \leq C\left(\frac{\log n}{n}\right)^3. \quad (159)
$$

$$\sum_{0 < |i-j| \leq c_3 \log n} E(1_{x_i \in [a,b] \cap \theta_n} \times 1_{x_j \in [a,b] \cap \theta_n}) \leq C\left(\frac{\log n}{n}\right)^3. \quad (160)
$$

The choice of the constant $c_3$ comes from our next proposition, whose proof consist of regular calculation, mainly based on (43), so we skip it off; let $\sigma(x)$ denotes the sigma field generated by $x$:

Proposition 25. Assume $E|x|^q < \infty$. There exists a constant $C > 0$ such that for $n \geq k \geq (c_3 \log n)^2$, any $\Gamma \in \sigma(x_j, \ldots, x_n)$, $n \geq j \geq 1$,

$$|P(\Gamma | \sigma(x_1, \ldots, x_{j-k})) - P(\Gamma)| \leq Cn^{-q}.
$$

In this lemma, we adopt the definition $x_i \equiv 0$ if $i < 1$.

By (159), (160), Proposition 25 with $q > 2$, (4.12) in Chen (1973), we have for $s \geq 0$,

$$|P(q^{[a,b] \cap \theta_n} = s) - P(q_p^{[a,b] \cap \theta_n} = s)| \leq C(n^{-1} + \left(\frac{\log n}{n}\right)^3). \quad (161)$$
Proof of Proposition 2

Proof of Proposition 2. Let $\Delta \in \mathcal{D}^2$ be small enough (see (165)). Due to Corollary 22, it suffices to prove

$$nP(x_o \in B_{r_o}(r)) - \pi_r E|r_\infty| = o(1).$$

(162)

With assumption 1 and 5, we have

$$E|r_\infty|^k < \infty, \text{ for each } k > 0.$$

(163)

(Sketch of the) Proof of (163). The distribution of $r_\infty$ is decided by two Poisson processes with rate $\pi(r)$ and independent jumps with positive mean. First we note that considering only one side of the Poisson process, the $j$-th jumps occurs at a distant distributed as Gamma($j, \pi(r)$), whose moment, given $k$, increases in a $j$ order as $j$ grows. To prove (163), we show the probability that the $j$-th partial sum of jumps reaches the minimum of all partial sum of jumps is decreasing in $j$ exponentially. Specifically, we make two claims. The first: Assume assumption 5, $E |\tilde{z}_1|^{2+\delta_1} < \infty$ for any $\delta_1 > 0$, $S_n = \sum_{i=1}^{n} x_i$, there exist $M > 0$ such that $P(S_{n_0} > 0) > \frac{1}{2}$ for all $n_0 \geq M^2c_0^{-1} \log 2$.

By this claim, we see the probability for partial sum $S_{i+jn_0}$ to reach the minimum is indeed decreasing exponentially. As $n_0$ is finite, by separately constructing the moment bounds for the subsequence $i+jn_0, i=1, \ldots, n_0$ and some more tedious algebraic manipulation, we can finish the proof. We now prove the claim

(Sketch) Proof of the first claim. Due to the SETAR process defined in (1) and assumption 5, we have for each $k$, there exists $C > 0$ such that for all $n$,

$$\int \tilde{p}^n(\tilde{z}_0, \tilde{z}_1) \|\tilde{z}_1\|^k d\tilde{z}_1 < C.$$

(164)

Note the stationary distribution admits $k$-th moment bound under assumption 5. Combining these moment bounds with (43) there exists some $C > 0$ such that for all $n$, $\tilde{z}_0 \in \mathbb{R}^p, M > 0$,

$$\int \|\tilde{p}^n(\tilde{z}_0, \tilde{z}_1) - \tilde{p}(\tilde{z}_1)\| \|\tilde{z}_1\|^k d\tilde{z}_1 < C(M^{-k} + M^k(1 + \|\tilde{z}_0\|)^{-n}).$$

Then by the techniques used in (192) and Cauchy-Swartz inequality and assumption 5 for necessary truncation, we can finish the proof. 

Proof of the second claim. Define $x_1^M$ as the truncated $x_1$ such that $x_1^M = M$ if $x_1 > M$, $x_1^M = -M$ if $x_1 < -M$, $x_1^M = x_1$ otherwise; and that $Ex_1^M \geq \frac{1}{2} E x_1$.

Hoeffding’s inequality says $P(|S_n^M - E(S_n^M)| > \frac{1}{2} E(S_n^M)) \leq 2 \exp(-\frac{n c_0}{2 M^2})$; and note that the truncated $S_n^M$ is likely to coincide with the orginal $S_{n_0}$: $P(S_{n_0}^M \neq S_{n_0}) \leq$
\[ \sum_{1 \leq i \leq n_0} P(|x_1| \geq M) \leq n_0 M^{-2-\delta_i} C. \] By these arguments, we see if \( n_0 \geq M^2 c_0^{-1} \log 2 \) and \( M \) is large enough, the claim is proven (on the event \( \{|S_{n_0}^M - E(S_{n_0}^M)| \leq \frac{1}{2} E(S_{n_0}^M)\}, S_{n_0}^M > 0 \)).

Let \( \mathcal{F}_x \) denote the sigma field generated by \( x \). We begin with \( P(x_0 \in B_{\bar{r}_n}(r)) \),

\[
\begin{align*}
|P(x_0 \in B_{\bar{r}_n-r}(r)) - P(n(x_0 - r) \in B_{r_\infty}(0))| &= |P(n(x_0 - r) \in B_{n(\bar{r}_n-r)}(0)) - P(n(x_0 - r) \in B_{r_\infty}(0))| \\
&\leq \left| E \left\{ E \left[ 1_{n(x_0-r) \in B_{n(\bar{r}_n-r)}(0)} - 1_{n(x_0-r) \in B_{r_\infty}(0)} \right] |\mathcal{F}_x \right\} \right| \\
&\leq E \left\{ E \left[ 1_{n(x_0-r) \in B_{n(\bar{r}_n-r)}(0)} - 1_{n(x_0-r) \in B_{r_\infty}(0)} ; n|\bar{r}_n - r|; |r_\infty| \leq n^6 \Delta \right] |\mathcal{F}_x \right\} \\
&\quad + 2P(|\bar{r}_n - r| \geq n^{-1-6\Delta}) + 2P(|r_\infty| \geq n^6 \Delta) \\
&\equiv (I) + (II) + (III).
\end{align*}
\]

By Proposition 23, for all large \( n \),

\[
(I) = \int_r^\infty \int_{n(y-r)}^{n(y-r)+n^6 \Delta} dF_{n(\bar{r}_n-r)}(z) - \int_{n(y-r)}^{n^6 \Delta} dF_{r_\infty}(z) \left| \pi(y)dy \right| \\
+ \int_{-\infty}^r \int_{-n^6 \Delta}^{n(y-r)} dF_{n(\bar{r}_n-r)}(z) - \int_{-n^6 \Delta}^{n(y-r)} dF_{r_\infty}(z) \left| \pi(y)dy \right| \\
= \int_r^{r+n^{-6 \Delta}} \int_{n(y-r)}^{n^6 \Delta} dF_{n(\bar{r}_n-r)}(z) - dF_{r_\infty}(z) \left| \pi(y)dy \right| \\
+ \int_r^r \int_{r-n^{-6 \Delta}}^{n(y-r)} dF_{n(\bar{r}_n-r)}(z) - dF_{r_\infty}(z) \left| \pi(y)dy \right|
\]

\[ \leq C n^{-\bar{q} - (1-6\Delta)}, \]

Since \( \bar{q} > 0 \), we can pick \( \Delta \) such that \( C n^{-\bar{q} - (1-6\Delta)} = o(n^{-1}) \). By the above inequality, moments on \( |r_\infty| \),

\[
|P(x_0 \in B_{\bar{r}_n-r}(r)) - P(n(x_0 - r) \in B_{r_\infty}(0))| \leq C n^{-\bar{q} - (1-6\Delta)} + o(n^{-1}). \tag{165}
\]

Let \( Y \sim D \) Uniform \( (r - \frac{1}{2r}, r + \frac{1}{2r}) \). By the Lemma 4, \( \Delta \in \mathcal{D} \mathcal{E}^2_{\infty} \),

\[
\begin{align*}
&|P(n(x_0 - r) \in B_{r_\infty}(0)) - P(n(Y - r) \in B_{r_\infty}(0))| \\
&\leq \left| E \left\{ E \left[ 1_{n(x_0-r) \in B_{r_\infty}(0)} - 1_{n(Y-r) \in B_{r_\infty}(0)} ; |r_\infty| \leq n^6 \Delta \right] |\mathcal{F}_{r_\infty} \right\} \right| + 2P(|r_\infty| \geq n^6 \Delta) \\
&\leq \int_{y \in \mathbb{R}_+} \int_{r-n^{-6 \Delta}}^{r+n^{-6 \Delta}} n^{-\beta} \pi(r)dzdF_{r_\infty}(y) + 2P(|r_\infty| \geq n^6 \Delta) \tag{166}
\end{align*}
\]

\[ \leq C n^{-\beta - b} + 2P(|r_\infty| \geq n^6 \Delta). \]
By the moment bound on \(r_\infty\), for all large \(n\),

\[
|nP(n(Y - r) \in B_{r_\infty}(0)) - \pi(r)E\{|r_\infty|; |r_\infty| \leq n^{6\Delta}\}| \\
= n \left[ \int_{\mathbb{R}}^{\infty} \int_{-\frac{n^{6\Delta}}{2\pi(r)}}^{\frac{n^{6\Delta}}{2\pi(r)}} \pi(r) \frac{1}{n} dF_{r_\infty}(z) + \int_{-\infty}^{0} \int_{-\frac{n^{6\Delta}}{2\pi(r)}}^{\frac{n^{6\Delta}}{2\pi(r)}} \pi(r) \frac{1}{n} dF_{r_\infty}(z) \right] \\
+ \int_{-\infty}^{0} \int_{-\frac{n^{6\Delta}}{2\pi(r)}}^{\frac{n^{6\Delta}}{2\pi(r)}} \pi(r) \frac{1}{n} dF_{r_\infty}(z) \\
= n \left[ \int_{\mathbb{R}}^{\infty} \frac{1}{2} dF_{r_\infty}(z) + \int_{-\infty}^{0} \frac{1}{2} dF_{r_\infty}(z) \right] + n \left[ \pi(r) \int_{-\infty}^{-\frac{n^{6\Delta}}{2\pi(r)}} dF_{r_\infty}(z) + \pi(r) \int_{\frac{n^{6\Delta}}{2\pi(r)}}^{0} dF_{r_\infty}(z) \right] \\
\leq 2CnP(|r_\infty| \geq \frac{n}{2\pi(r)}) + 2Cn^2P(|r_\infty| \geq n^{6\Delta}) \\
= o(n^{-1}).
\] (167)

By Cauchy-Swartz inequality, the moment bound on \(r_\infty\), for all large \(n\),

\[
\pi(r)E\{|r_\infty|; |r_\infty| \leq n^{6\Delta}\} - \pi(r)E|r_\infty| \\
\leq \pi(r)E\{|r_\infty|; |r_\infty| \geq n^{6\Delta}\} \\
\leq \pi(r)E^{1/2}(r_\infty^2)P^{1/2}(r_\infty \geq n^{6\Delta}) \\
= o(n^{-1}).
\] (168)

By (165) to (168), we have (162). \(\blacksquare\)

**Proof of Theorem 5**

**Proof of Theorem 5.** In the following derivation, we consider

\[
nE \left[ (a_1x_n - \hat{a}_1^o x_n)^2 1_{(x_n \leq \tilde{r}_n \wedge r)} \right] = \sigma^2 + o(1),
\] (169)

and the other part can be done by the same argument. Since \(\tilde{r}_n^o\) differs from \(\tilde{r}_n\) only on \(U_n^c \cap R_n^c\), by Corollary 6, the moment assumptions, and standard inequalities we have

\[
nE \left[ (a_1x_n - \hat{a}_1^o x_n)^2 1_{(x_n \leq \tilde{r}_n \wedge r)} \right] = nE \left[ (a_1x_n - \hat{a}_1^o x_n)^2 1_{(x_n \leq \tilde{r}_n \wedge r)} \right] + o(1)
\] (170)

Now define \(\hat{a}_1^o\) as the coefficient estimator bases on the subsample, \(x_1, \ldots, x_{n-c}\log n\); by standard inequalities,

\[
nE \left[ (a_1 - \hat{a}_1^o)^2 x_n^2 1_{(x_n \leq \tilde{r}_n \wedge r)} \right] - nE \left[ (a_1 - \hat{a}_1^o)^2 x_n^2 1_{(x_n \leq \tilde{r}_n \wedge r)} \right] \\
= E \left[ (\hat{a}_1^o - \hat{a}_1^o)^2 (2a_1 - \hat{a}_1^o - \hat{a}_1^o) x_n^2 1_{x_n \leq \tilde{r}_n \wedge r} \right] \\
\leq E^{1/2} \left[ (\hat{a}_1^o - \hat{a}_1^o)^2 n 1_{U_n} \right] E^{1/2} \left[ (2a_1 - \hat{a}_1^o - \hat{a}_1^o)^2 x_n^4 1_{x_n \leq \tilde{r}_n \wedge r} \right] \\
\leq E^{1/2} \left[ (\hat{a}_1^o - \hat{a}_1^o)^2 n 1_{U_n} \right] \left\{ E^{1/4} \left[ (a_1 - \hat{a}_1^o)^4 x_n^2 1_{U_n} \right] + E^{1/4} \left[ (a_1 - \hat{a}_1^o)^4 x_n^2 1_{U_n} \right] \right\} E^{1/4} (x_n^4 1_{x_n \leq \tilde{r}_n \wedge r}).
\] (171)
By the techniques in step 1 of Proposition 7, Theorem 4, the definition of $U_n$, and that the same moment argument can be directly applied to the model with sample size $n - c \log n$, we have

$$E^{1/2} [ (\hat{a}_1 - \tilde{a}_1)^2 n 1_{U_n} ] = o(1).$$

By standard inequalities, Theorem 4, (172), and the moment assumptions,

$$n = o(1).$$

By standard inequalities, Theorem 4,

$$nE \left[ (a_1 - \tilde{a}_1) \frac{2}{x_n^2 1_{(x_n \leq \tilde{r}_n, r)} 1_{U_n}} \right] - nE \left[ (a_1 - \tilde{a}_1) \frac{2}{x_n^2 1_{(x_n \leq r)} 1_{U_n}} \right]
\leq E^{1/2} \left[ \left( a_1 - \tilde{a}_1 \right)^4 1_{r \geq \sum \Delta \gamma} \right] \times E^{1/2} \left[ \left( x_n \left( 1_{x_n \leq \tilde{r}_n, r} - 1_{x_n \leq r} \right) \right) 1_{U_n} \right]
\leq E^{1/2} \left[ \left( a_1 - \tilde{a}_1 \right)^4 1_{r \geq \sum \Delta \gamma} \right] \times E^{1/2} \left[ \left( x_n \left( 1_{x_n \leq \tilde{r}_n, r} - 1_{x_n \leq \tilde{r}_n, r} \right) \right) 1_{U_n} \right]
= o(1),$$

where the resulting $o(1)$ is due to an application of (43) and the moment assumptions. By an argument similar to (171), we can get rid off the prime estimator,

$$\left| nE \left[ (a_1 - \tilde{a}_1)^2 1_{U_n} \right] - nE \left[ (a_1 - \tilde{a}_1)^2 1_{U_n} \right] \right| = o(1)$$

Let $R_1 = E(x_i^2 1_{(x_i \leq r)})$. The autocovariance function of $x_i$, which might not be easy to state in a closed-form, is surely absolutely summation due to its AR structure. Finally we deal with $nE \left[ (a_1 - \tilde{a}_1)^2 1_{U_n} \right],$

$$nE \left[ (a_1 - \tilde{a}_1)^2 1_{U_n} \right] = nE \left[ \left( \sum_{i \in J_1} x_i^2 \right)^{-2} \left( a_1 \sum_{i \in J_1} x_i^2 - \sum_{i \in J_1} x_i x_{i+1} \right) 1_{U_n} \right] + o(1) \equiv E [(I)(II)] + o(1),$$

where

$$\left( I \right) = \left( \sum_{i \in J_1} x_i^2 + \sum_{i \in J_1 \setminus J_1 \cup J_1 \setminus J_1} x_i^2 \right)^{-2} 1_{U_n},$$
\[(II) = n^{-1} \left( (a_1 - a_2) \sum_{i \in J \setminus J_1} x_i^2 - \sum_{i \in J_1} x_i e_{i+1} - \sum_{i \in J_1 \setminus J_1} x_i e_{i+1} + \sum_{i \in J \setminus J_1} x_i e_{i+1} \right)^2.\]

To deal with \((I)\), we note that

\[
E^{1/2} \left| (I) - R_1^{-2} \right|^2 = E^{1/2} \left| \left( \sum_{i \in J_1} x_i^2 + \sum_{i \in J \setminus J_1 \setminus J_1} x_i^2 \right) - \sum_{i \in J_1} x_i e_{i+1} + \sum_{i \in J \setminus J_1} x_i e_{i+1} \right| - n^{-1} E \left( - \sum_{i \in J_1} x_i e_{i+1} \right)^2 = o(1), \tag{179}\]

by standard inequalities, techniques in Theorem 4. For \((II)\), by a similiar argument we have

\[
(II) - n^{-1} E \left( - \sum_{i \in J_1} x_i e_{i+1} \right)^2 = o(1), \tag{179}\]

and also that

\[
n^{-1} R_1^{-2} E \left( - \sum_{i \in J_1} x_i e_{i+1} \right)^2 = R_1^{-1} \sigma^2 + o(1) \tag{180}\]

By (170) to (180), standard inequalities, and the definition of \(x_o\) and \(R_1\), we have (169).

\textbf{Supplementray File to SETAR(p)}

\textbf{Introduction}

For a linear autoregressive time series model with regular conditions, which might exhibit a variety of forms depending on the context, hold, to improve the result of AMSPE from
order 1 to order $p$ might be difficult due to the heavy techniques involved in the establishment of the negative moment bounds for sample covariates matrix. This, however, is not the case for the SETAR model. From the decomposition of $n\left[ E\left( x_{n+1} - \hat{x}_{n+1}^{o} \right)^{2} - \sigma^{2} \right]$, we see that given the perfect delay and threshold estimator, so $d$ and $S_{r}^{0}$ are both known, we are actually dealing with the AMSPE of two autoregressive model for each regression function in the SETAR model. The result of such AMSPE will be $2p\sigma^{2}$, provided

$$E\left( \left\| \frac{\sum_{i+1-d \in J_{k,0}} \tilde{x}_{i} \tilde{x}_{i}'}{n} \right\|^{-q} \right) < C, k = 1, 2,$$

hold. The nature of SETAR process is that the event \{No $x_{i}$ falling into one side of $r$\} is always with positive probability, and this fact force us to make a refinement, which introduce the event $U_{n}$, on our estimators since one of $J_{k,0}$, $k = 1, 2$, might be empty set. Note that for the modified estimators, $\hat{a}_{k}^{o}$, negative moment bounds on the sample covariates matrix are trivial.

For this reason, in this online supplementary file, we provide the sketch proof with details only for the features of SETAR(p) when the order $p$ is non-trivially involved.

**Preliminary Results**

**Lemma 26.** For any small $\frac{1}{3} \geq c > 0$, there exist $\alpha_{1} > \alpha_{2} > 1$, $s$ is an integer, such that

$$\alpha_{1}sm_{s}^{*} < c,$$

$$P(q^{\Theta''} \leq (1 - \alpha_{1}sm_{s}^{*})n) \leq \frac{\alpha_{2}^{-(\alpha_{1} - \alpha_{2})m_{s}^{*}n}}{1 - \alpha_{2}^{2}},$$

where $\Theta'' = [-A^{-s}, A^{-s}]$, and $m_{s}^{*} \equiv m_{s}^{*} = 2 \max\{E|e_{1}|, 1\} \times A^{s}(1 + s)^{2}(1 - A)^{-1}$, $x_{0} \equiv 0$.

**Proof of Lemma 26.** Define $\rho_{1} = \alpha_{1}sm_{s}^{*}$. Let $t_{1} = \inf\{k : |x_{k-1}| \leq A^{-s}, |x_{k}| \geq A^{-s}, 1 \leq k \leq n\}$, $t_{i} = \inf\{k : |x_{k}| \geq A^{-s}, t_{i-1} + s \leq k \leq n\}$, $i > 1$.

Claim: For $1 \leq l < j - s, k > 0$, $j + (k - 1)s \leq n$,

$$P(t_{1} = j, t_{2} = j + s, \ldots, t_{k} = j + (k - 1)s) \leq m^{sk},$$

$$P(t_{i} = j, t_{i+1} = j + s, \ldots, t_{i+k-1} = j + (k - 1)s|t_{i-1} = l) \leq m^{sk}, i > 1.$$

Note that by the multivariate form of the SETAR(p) model,

$$\tilde{x}_{n} = \begin{cases} A_{1}\tilde{x}_{n-1} + \tilde{e}_{n}, x_{n-d} \leq r \\ A_{2}\tilde{x}_{n-1} + \tilde{e}_{n}, x_{n-d} > r, \end{cases}$$

where

$$A_{i} = \left( \frac{\hat{a}'_{i}}{I_{p-1}^{0}} \right), \tilde{e}_{n} = (e_{n}, \tilde{0})'.$$
By this we have

\[ |x_n| \leq \|\tilde{x}_n\| \leq |e_n| + A \|\tilde{x}_{n-1}\|; \]

hence for all \( k > 0 \)

\[ |x_n| \leq |e_n| + \sum_{i=1}^{k} A^k|e_{n-i}| + A^{k+1} \|\tilde{x}_{n-k-1}\|. \]

Then by a argument almost the same with that in SETAR(1), we can prove (181); the rest of the proof is routine.

Let \( E_{p_1} = \left\{ s \left( \frac{1-\rho_0}{p} \right)_n < \infty \right\} \) and

\[ s_1 = \inf\{i : (x_{i-2p+1}, \ldots, x_{i+p}) \in \Theta''^3, n > i > 2p + 1\}, \]

\[ s_j = \inf\{i : (x_{i-2p+1}, \ldots, x_{i+p}) \in \Theta''^3, n > i > s_{j-1} + p\}, j > 1. \]

Notice that \( \{q^{\theta''} > (1 - \rho_1)n\} \subset E_{p_1}. \)

**Lemma 27.** For \( 1 \leq k \leq \left( \frac{1-\rho_0}{p} \right)n \equiv \rho_0n, \tilde{\beta} \subset \Theta''^p, \beta \subset \Theta''^{(k-1)p}, \)

\[ P \left( \tilde{x}_{s_k} \in \tilde{\beta}\left| (\tilde{x}_{s_i}, 1 \leq i < k - 1) \in \beta, E_{p_1} \right. \right) \geq |\tilde{\beta}| m^*, \]

where \( m^* = \frac{(mM^{-1})^{2p}}{|\Theta''|^p}. \)

**Proof of Lemma 27.** The proof is essentially the same with that in SETAR(1).

Let \( \theta_i \subset \Theta'', i = 1, 2, \) and

\[ E_{p_1}^2 = \{ \text{More than } c_2 n \ (x_{s_i-d+1}, x_{s_i-d+1})'s \text{ fall inside } \theta_1 \times \theta_2, 1 \leq i \leq \rho_0n\}, \]

\[ u_1 = \inf\{s_i : (x_{s_i-d+1}, x_{s_i-d+1}) \in \theta_1 \times \theta_2, 1 \leq i \leq \rho_0n\}, \]

\[ u_j = \inf\{s_i : (x_{s_i-d+1}, x_{s_i-d+1}) \in \theta_1 \times \theta_2, u_{j-1} \leq i \leq \rho_0n\}, j > 1, \]

\[ \tilde{\theta} \equiv A_1 \times \cdots \times A_p, A_{p-d+1} = \theta_1, A_{p-d+1} = \theta_2, A_j = \Theta''. \]

Notice that by a way similar to Lemma 26, we can show that there is some \( c_2, c, \delta > 0 \) such that for all large \( n, \)

\[ P(E_{p_1}^2) \leq \exp(-cn^\delta). \]

By almost the same techniques in the proof of Lemma 27, we have

**Corollary 28.** For any \( \theta_i \subset \Theta'', i = 1, 2 \) in the above definition of \( u_i \)'s and \( \tilde{\theta} \), there exists \( m^* > 0 \) such that for \( 1 \leq k \leq c_2 n, \tilde{\beta} \subset \tilde{\theta}, \beta_{(k-1)p} \subset \Theta''^{(k-1)p}, \)

\[ P \left( \tilde{x}_{u_k} \in \tilde{\beta}\left| (\tilde{x}_{u_i}, 1 \leq i < k - 1) \in \beta_{(k-1)p}, E_{p_1}^2 \right. \right) \geq m^* |\tilde{\beta}|. \]
Figure 6: Visualized random indexes for $T^\theta_k$'s and $\tilde{T}^\theta_k$'s. Note that $\tilde{T}^\theta_k$'s capture distinguished subset of the indexes than $T^\theta_k$'s. In this graphic, $s_{1,2} = T^\theta_2 + d - 1, s_{2,2} = T^\theta_1 + d - 1$.

Let $\theta_1 = [r - n^{-b}, r], \theta_2 = [r, r + n^{-b}], \theta = [r - n^{-b}, r + n^{-b}]$, and for $k = 1,2,$
$$\tilde{T}^\theta_k = \inf\{T^\theta_j : x^\theta_j \in \theta_k, j \geq p + 1\}, \tilde{T}^\theta_k = \inf\{T^\theta_j : x^\theta_j \in \theta_k, T^\theta_j > \tilde{T}^\theta_k\}, i > 1,$$
$$q^\theta_k = \sup\{i : \tilde{T}^\theta_k \leq n, i \geq 1\},$$

For the $x_{i}, i = 1, \ldots, n$ with indexes being captured by $\tilde{T}^\theta_k$, we set another random indexes in order to represent these $x_i$'s in an ordered fashion from $r$ to upward, $s_{1,2}$'s , or $s_{1,1}$'s downward,

$$s_{1,1} = \arg\max_{T^\theta_k + d - 1 : x^\theta_{k,1} < r, 0 < k \leq q^\theta_1} x^\theta_{k,1}, s_{i,1} = \arg\max_{T^\theta_k + d - 1 : x^\theta_{k,1} < x_{i-1,1}, 1 \leq i \leq q^\theta_1} x^\theta_{k,1}, q^\theta_2 \geq i > 2,$$

$$s_{1,2} = \arg\min_{T^\theta_k + d - 1 : x^\theta_{k,2} > r, 0 < k \leq q^\theta_2} x^\theta_{k,2}, s_{i,2} = \arg\min_{T^\theta_k + d - 1 : x^\theta_{k,2} > x_{i-1,1}, 1 \leq i \leq q^\theta_2} x^\theta_{k,1}, q^\theta_2 \geq i > 2,$$

and define
$$\arg\max_{t \in \emptyset} x_t \equiv 0, \arg\min_{t \in \emptyset} x_t \equiv \infty.$$

$x_0, x_{\infty}$ can be anything. Define $(r_1,i, \ldots, r_{d-1,i}, r_{d-1+i}, \ldots, r_{p-1,i})'$'s, which are independent and identically distributed processes, and $\tilde{x}_{s_{1,2}}$'s as follow

$$\tilde{x}_{s_{1,2}} = (r_1,i, \ldots, r_{d-1,i}, x_{s_{i,1}-d+1}, r_{d-1+i}, \ldots, r_{p-1,i})'$$

$$(r_1,i, \ldots, r_{d-1,i}, r_{d-1+i}, \ldots, r_{p-1,i})' \overset{D}{=} (x_{T'_1}, \ldots, x_{T'_{i-d+2}}, x_{T'_{i-d}}, \ldots, x_{T'_{i-p+1}})'$$

where
$$T'_1 = \{i : x_{1+i-d} \in [r, r + cn^{-1-6\Delta}], i > p\},$$

and $x_i^{in}$'s are independent and identically distributed process of $x_i$'s; $\tilde{x}_{s_{1,2}}$ is defined in a similar way. Note that the bar-hat denotes that we are using the iid process analog.
Let \( e_i, \bar{e}_i, \bar{e}_{i,k}, k = 1, 2; i = 1, \ldots \) denote independent and identically distributed random variables. For subset \( A_i \)'s, we introduce the product set

\[
x_k \prod_{i=1}^k (A_i) = A_1 \times \cdots \times A_k
\]

The next lemma, which provide us a control over the asymptotic distribution of the random vector \( (e_{s,i,k+1}, 1 \leq i \leq q^\theta_k) \), \( k = 1, 2 \), has a simple interpretation if we make the odd coordinates of \( A_k^i \in \mathbb{R}^{2i}, k = 1, 2, i = 1, \ldots \) real lines: the limiting distribution of the former is that of \( (\bar{e}_{i,k}, 1 \leq i \leq q^\theta_k) \), \( k = 1, 2 \). Notice that we use the same random index \( q^\theta_k \) as the length of the random vectors.

**Lemma 30.** Define

\[
E = \left\{ \prod_{i=1}^\infty B_{i,1}, \prod_{i=1}^\infty B_{i,2} \mid \prod_{i=1}^\infty B_{i,s} \in \prod_{i=1}^\infty \mathbb{R}^{(p+1)i}, s = 1, 2 \right\}.
\]

If \( E|\epsilon_1| < C \), then there exists a constant such that for each \( n \),

\[
\sup_{\prod_{i=1}^\infty A_{i,1}, \prod_{i=1}^\infty A_{i,2} \in E} \left| P\left( \cap_{k=1}^2 \left\{ \left( (\bar{x}_{s,i,k}, e_{s,i,k+1}), 1 \leq i \leq q^\theta_k \right) \in A_k^i \right\} \right) \right| - P\left( \cap_{k=1}^2 \left\{ \left( (\bar{x}_{s,i,k}, e_{i,k}), 1 \leq i \leq q^\theta_k \right) \in A_k^i \right\} \right) \leq C(n^4 \rho \log n + n^{-2q+1}),
\]

**Proof of Lemma 29.** The proof can be found in SETAR(1).

**Corollary 30.** There exists \( C, c > 0, 1 > \rho > 0 \) such that for any \( n \geq g \geq t > 0, \beta_{1,g}, \beta_{2,g} \subset \mathbb{R}^{(1+p)g}, \beta \subset \mathbb{R} \), all large \( n \), we have

\[
\left| P\left( \{ (n(x_{s,2+1-d} - r) \in \beta \} \cap_{k=1}^2 \left\{ \left( (\bar{x}_{s,i,k}, e_{s,i,k+1}), 1 \leq i \leq g \right) \in \beta_{k,g} \right\} \right) \right| - P\left( \{ (n(x_{s,2+1-d} - r) \in \beta \} \cap_{k=1}^2 \left\{ \left( (\bar{x}_{s,i,k}, e_{i,k}), 1 \leq i \leq g \right) \in \beta_{k,g} \right\} \right) \leq C(n^4 \rho \log n + n^{-2q+1}),
\]

\(-g \leq t \leq 0 \) can be argued in the same way.

**Proof of Corollary.** The proof can be found in SETAR(1).

**Features in SETAR(p)**

**Lemma 31.** Given \( d, d' \in \{1, \ldots, p\}, d' \neq d \), there exists \( c_1, c_2 > 0 \) such that for all large \( n \), all \( \theta_1, \theta_2 \in \Theta^p \) with \( |\theta| \geq c_0 > 0 \),

\[
P\left( \lambda_{\min} \left( \sum_{x_{1+i-d} \in \theta_1, x_{1+i-d'} \in \theta_2} \bar{x}_i \bar{x}_i' \right) \leq c_1 n \right) \leq \exp(-cn^\delta).
\]

93
Proof of Lemma 31. Given $p$-dimensional random vectors $\tilde{y}_i, i = 1, \ldots, c_3 n$ such that for all $\theta_i, i = 1, 2, \beta \in \Theta_\beta^{(i-1)p}$,
\[
P \left( \tilde{y}_i \in \tilde{\theta} \left| \tilde{y}_j, 1 \leq j \leq i - 1 \right. \right) \geq m^*|\tilde{\theta}|,
\]  (182)
where $\tilde{\theta} = A_1 \times \cdots \times A_p, A_{p-d+1} = \theta_1, A_{p-d+2} = \theta_2, A_j = \Theta'\beta$ otherwise. Then fixing a large enough $q_{c_0} > 1$, by (182) and the argument similar to that in the proof of Lemma 11, there exist $c_5, c_6, c_7, \delta_2 > 0$ such that for all $\|v\| = 1$, all large $n$,
\[
P \left( \text{Fewer than } c_5 n \text{ } \tilde{y}_i \text{'s fall inside } A_v \right) \leq \exp \left( -c_6 n^{\delta_2} \right),
\]
where
\[
A_v = \left\{ y : \inf_{z : |z'| \geq \frac{1}{q_{c_0}}} \sum_{i=1}^{c_3 n} |\tilde{y}_i y_i'| \geq c_7 > 0, y \in \tilde{\theta} \right\}.
\]
Note that when $c_0$ is fixed, a large enough $q_{c_0}$ and small enough $c_7$ ensure a non-empty $A_v$. Then we can pick $c_4 \leq c_5 c_7$ small enough such that
\[
P \left( \lambda_{\min} \sum_{i=1}^{c_3 n} \tilde{y}_i y_i' \right) \leq c_4 n \leq P \left( \text{Fewer than } c_5 n \text{ } \tilde{y}_i \text{'s falling inside } A_v \right) \leq \exp \left( -c_6 n^{\delta_2} \right),
\]  (183)
Since $q_{c_0}$ is a constant, there exists $N > 0, \nu_1, \ldots, \nu_N$ such that $\|\nu_i\| = 1$ and
\[
\partial B_1(0) \subset \bigcup_{j=1}^{N} \{ z : |z'| \nu_i | \geq \frac{1}{q_{c_0}}, \|z\| = 1 \}. \]  (184)
By (183), (184), we have
\[
P \left( \lambda_{\min} \sum_{i=1}^{c_3 n} \tilde{y}_i y_i' \right) \leq c_4 n \leq N \exp \left( -c n^{\delta} \right).
\]
Notice that from Corollary 28, we have $u_i, i = 1, \ldots, c_2 n$ such that $x_{u_i}$'s satisfy (182). 

Lemma 32. Assume $E|e_i|^{2q} < \infty$. There exists $C > 0$ such that for all $1 > \Delta > 0, z \in \Theta', 1 \leq d \leq p$, all large $n$,
\[
\sum_{x_{x_i} \leq \Theta^u} E \|\tilde{x}_i\|^q \leq C (\log n) n^\Delta,
\]
where $\theta_{z, \Delta} = [z, z + c_1 n^{-\Delta}]$.

proof of Lemma 32. Without loss of generality, we assume $c_1 = 1$. Let
\[
Q_1 = \inf \{ j : x_{j+1-d} \in \theta_{z, \Delta}, j > p \},
\]
\[
Q_i = \inf \{ j : x_{j+1-d} \in \theta_{z, \Delta}, j > Q_{i-1} + c_2 \log n \}, i > 1,
\]
\[
A = \{ \text{More than } c_4 n^{\Delta} x_i \text{'s falling inside } \theta_{z, \Delta} \}.
\]
$c > 0$ such that $P(A) < \exp(-c_3 n^\delta)$ for some $c_3, \delta > 0$. Note that the definition of $Q_i$’s is slightly different than $T_i$’s. Using (43) and arguments similar to those in the supplementary file of SETAR(1) and Lemma 10, we have for all large $n$,

$$\sum_{i=1}^{c_4 n^\Delta} E \left| x_{Q_i} - \hat{x}_o \right|^q \leq C (\log n)n^\Delta. \quad (185)$$

We divide the indexes $\{i : x_{i+1-d} \in \theta_{z,\Delta}, i > p\}$ into two subsets: those whose closest antecedent index is within $c \log n$ and the others. We bound the first kinds with brutal-force bounds; for the others we use $c \log n$ asymptotical independence property,

$$\sum_{x_{i+1-d} \in \theta_{z,\Delta}, i > p} E \left| x_i - \hat{x}_o \right|^q \leq \left[ \sum_{i=1}^{c_4 n^\Delta} \sum_{j=0}^{n} E \left| x_{Q_i+1-d+j} - \hat{x}_o \right|^q + \sum_{i=p}^{n} E \left( \left| x_i - \hat{x}_o \right|^q ; A \right) \right]$$

$$\leq C (\log n)n^\Delta,$$

by (185), standard inequalities and moment assumption.

**Moment bounds**

**Lemma 33.** Assume $E|\epsilon_i|^q < \infty$. Given $\Delta > 0$, there exists $C > 0$ such that for all large $n$, $z \in \Theta'$, $1 \leq d \leq p$,

$$\sup_{i \in \{j : x_{j-d+1} \in \theta_{z,\Delta} \equiv [z, z+n^{-(1-\Delta)}]\}} E|\epsilon_{i+1}|^q \leq C n^\Delta.$$

The following moment bounds will be constantly used in our argument. Their proofs are all routinely applying Lemma 32, 33 and standard inequalities.

**Lemma 34.** Assume $E|\epsilon_i|^{2q} < \infty$. Given $1 > \Delta > 0$, then there exists $C > 0$ such that for all $h \leq n^\Delta$, $z \in \Theta'$, all large $n$,

$$E \left| \frac{\sum_{1+i-d \in (J_{1,z,h,d} \cup J_{2,z,d})} x_i \epsilon_{i+1}}{(\log n)n^{3\Delta}} \right|^q \leq C,$$

$$E \left| \frac{\sum_{1+i-d \notin (J_{1,z,h,d} \cup J_{2,z,d})} x_i \epsilon_{i+1}}{(\log n)n^{3\Delta}} \right|^q \leq C,$$

where $J_{1,z,h,d} = J_{1,z,d} \cup \{S_j^z : 1 \leq j \leq h\}$, $J_{1,z,d} = \{i : x_i < z, n-d \geq i \geq p-d+1\}$; $J_{2,z,d}$, $J_{2,z,h,d}$ are defined in a similar way; $e_0 \equiv 0$ for this lemma.

**Lemma 35.** Assume $E|\epsilon_i|^{2q} < \infty$. Given $1 > \Delta > 0$, then for all $0 < b < 1-\Delta$, $z \in \Theta'$, $\theta_{z,b} \equiv [z, z+n^{-b}]$, all large $n$,

$$E \left| \frac{\sum_{x_{i+1-d} \in \theta_{z,b}} x_i \epsilon_{i+1}}{(\log n)n^{3(1-b)}} \right|^q \leq C.$$
Figure 7: Visualized random indexes subset $J_{1,z,d}, J_{2,z,d}, J_{1,z,h,d}$ with $z = 0.5, h = 3$. The definition of $J_{i,z,d}$'s excludes \{1, \ldots, p - d, n - d + 1, \ldots, n\}.

On the empirical distribution of $\tilde{r}_n$

**Proposition 36.** Given $\Delta > 0$ small enough. For each $k > 0$, assume $E|e_1|^q$ for $q$ large enough. Then there exist $C, c > 0$ such that for all large $n$,

$$P(|\tilde{r}_n - r| \geq cn^{-(1-6\Delta)}) = O(n^{-k}).$$

**Proof.** The major difference between this proof and its analog in SETAR(1) is that to show

$$(\hat{j}, \hat{d})' \in \arg\min_{-n+1 \leq i \leq n; 1 \leq d \leq p} \tilde{P}_{i,d},$$

we now need for all large $n$,

$$P(\hat{d} \neq d) = O(n^{-k}). \quad (186)$$

To prove (186), we argue in the same way as in Proposition 36 for SETAR(1): first we construct intervals with length $\frac{1}{n}$ spreading over the compact set $\Theta'$ with the amount no more than $(|\Theta'|+1)n$, and then we show that threshold estimator cannot fall on any such intervals in $\Theta'$. We will use $\tilde{x}_i^q = x_i e + q_i 1_F$, which provides us bounded inverse covariate moments with $P(F) = O(n^{-k})$ for all large $n$. Specifically, $F = \cup_s E_{1,s} \cup_s E_{2,s} \cup E_3$,

$E_{1,s} = \left\{ \lambda_{\min} \left( \sum_{x_{1+i-d} < \tilde{x}_i \tilde{x}'_s} \tilde{x}_i \tilde{x}'_s \right) \leq c_1 n \right\},$

$E_{2,s} = \left\{ \lambda_{\min} \left( \sum_{x_{1+i-d} \geq \tilde{x}_i \tilde{x}'_s} \tilde{x}_i \tilde{x}'_s \right) \leq c_1 n \right\},$

$E_3 = \left\{ \text{More than } c \log n \text{ } x_i's \text{ fall inside any interval} \right\},$
where \(a_i, b_i \in \Theta\) are the cutting points of the intervals. By Lemma 31 and proper choice for the constants and \(q_i\)'s, the argument for SETAR(1) can also be used for this proof; we should omit the detail of \(q_i\)'s to keep the context simple. It turns out that we can establish (186) by constructing lower bound with high probability of the fitted loss, \(\mathbb{Z}_{z,d}^o\), for all \(d \neq d, z \in \Theta\), where

\[
\mathbb{Z}_{z,d}^o = \sum_{i=p}^{n-1} c_{i+1}^o + \sum_{x_i = 1} \left( (\hat{a}_1 - \hat{a}_{1,z,d}) x_i^o \right)^2 + 2 \sum_{x_i = 1} (\hat{a}_1 - \hat{a}_{1,z,d}) x_i^o e_{i+1}^o \\
+ \sum_{x_i = 1} \left( (\hat{a}_2 - \hat{a}_{2,z,d}) x_i^o \right)^2 + 2 \sum_{x_i = 1} (\hat{a}_2 - \hat{a}_{2,z,d}) x_i^o e_{i+1}^o,
\]

where \(\hat{a}_{1,z,d} = \left( \sum_{x_i = 1} \hat{x}_i^o \right)^{-1} \left( \sum_{x_i = 1} x_i^o \hat{x}_i^o \right)\),

The definition of \(\hat{a}_{2,z,d}\) follows this way. Constructing such a lower bound is a simple task if we observe

\[
\hat{a}_1 - \hat{a}_{1,z,d'} = \left( \sum_{x_i = 1} \hat{x}_i^o \right)^{-1} \left( \sum_{x_i = 1} x_i^o e_{i+1}^o + \sum_{x_i = 1} \hat{x}_i^o (\hat{a}_1 - \hat{a}_2) \right),
\]

and by Lemma 31, standard inequalities and the definition of \(x_i^o, e_i^o\)'s, there exists some \(\frac{1}{2} > \varepsilon > 0\) such that for all large \(n\),

\[
P\left( \left\| \hat{a}_1 - \hat{a}_{1,z,d} \right\| < n^{-\varepsilon} \right) = O(n^{-k}).
\]

This result is intuitive since fitting a wrong delay causes huge fitted loss and hence the coefficients estimator based on \(d'\) can never get as close to \(a_i\)'s as the estimators based on \(d\). By this and the preliminary results and arguments similar to those in SETAR(1), we can establish the lower bound of \(\mathbb{Z}_{z,d}^o\). The rest of the proof is standard as in that of SETAR(1).

We note that when \(d \neq d\), any positive moment bound can be bounded by sup-bounds due to Proposition 36 and standard inequalities; and the negative moment bounds can be dealt with the event \(U_n\). Hence, the consideration while \(d \neq d\) will be trivial and we assume \(d\) is known from now on; also we drop the underscript \(d\) in \(J_{k,j,d}\) off. Let \(\hat{x}_{i,k}, \hat{x}_{lim}, \hat{e}_{i,k}, e_1, k = 1, 2\) denote independent and identically distributed processes, respectively. Define

\[
G_j^lim = \begin{cases} 
\sum_{i=1}^j \left[ \hat{x}_{i,2}^lim (\hat{a}_1 - \hat{a}_2) \right]^2 + 2 \sum_{i=1}^j \hat{x}_{i,2}^lim (\hat{a}_1 - \hat{a}_2) e_{i,2}, j > 0, \\
\sum_{i=1}^{j-1} \left[ \hat{x}_{i,1}^lim (\hat{a}_2 - \hat{a}_1) \right]^2 + 2 \sum_{i=1}^{j-1} \hat{x}_{i,1}^lim (\hat{a}_2 - \hat{a}_1) e_{i,1}, j \leq 0
\end{cases},
\]

\[j = \arg\min_i G_i^lim.
\]
Using (188) and arguments similar to those in SETAR(1), we have

\[ P(\tilde{r}_n = x_{S_i^j}, j \in \Pi, n(x_{S_i^j} - r) \in \beta) = P(j = j)P(n(x_{S_i^j} - r) \in \beta) + C(n^{18\Delta-1} \vee n^{-\alpha} \vee n^{-\gamma + 6\Delta}). \]

**Proof of Proposition 37.** Assume \( d \) is known; given any \( c > 0 \), by lemma 26, 11, and Proposition 36, we have \( C > 0 \) such that for all large \( n \), \( \Lambda \equiv \{ j : 1 - cn^{6\Delta} \leq j \leq cn^{6\Delta} \} \),

\[ \sum_{j \notin \Lambda} P(j = \arg \min_i \tilde{P}_i) \leq Cn^{-1}. \]  

(188)

Given \( 1 - cn^{6\Delta} \leq j \leq cn^{6\Delta} \), \( \beta \subset \mathbb{R}^1 \), let

\[ W \equiv \{ 1 - cn^{6\Delta}, cn^{6\Delta} \in \Pi\}, W_j \equiv \{ j \in \Pi\}, Q \equiv \{ n(x_{S_i^j} - r) \in \beta\}, W' \equiv \{ s_{cn^{6\Delta}, 1} \neq 0, s_{cn^{6\Delta}, 2} \neq \infty\}, \]

\[ Q' \equiv \begin{cases} 
\{ n(x_{S_i^j} - r) \in \beta\}, j > 0, \\
\{ n(x_{S_i^j} - r) \in \beta\}, j \leq 0. 
\end{cases} \]

Again we use \( x_i^o \)'s to deal with the inverse moment bounds; \( F \) is the accompanying 'exception' set such that \( P(F) \leq \exp(-cn^d) \) for some constant \( c, \delta > 0 \) according to Lemma 31 and a version of Lemma 8, with details omitted as this method has been used several times throughout the article. The arguments provided below are all uniform in \( j \in \Lambda \).

Using (188) and arguments similar to those in SETAR(1), we have

\[ P(\tilde{r}_n = x_{S_i^j}, W_j, Q) = P(\tilde{r}_n = x_{S_i^j}, W, Q) + o(n^{-1}) \]

\[ = P(j = \arg \min_{j \in \Pi} \tilde{P}_i, W, Q, \mathcal{E}^c_n) + o(n^{-1}) \]

\[ = P(j = \arg \min_{j \in \Pi} \tilde{P}_i - \tilde{P}_0, W, Q, \mathcal{E}^c_n) + o(n^{-1}) \]

\[ = P(j = \arg \min_{j \in \Pi} P_j^o - P_0^o, W, Q) + o(n^{-1}) \]

\[ = P(j = \arg \min_{j \in \Lambda} P_j^o - P_0^o, W, Q) + o(n^{-1}). \]  

(189)

To deal with (189), we decompose \( P_j^o - P_0^o \) for \( j \in \Lambda \) as we did for SETAR(1),

\[ P_j^o - P_0^o = [(I) + (II) + (III) + (IV)] + [(V) + (VI) + (VII) + (VIII) + (X) + (XI)] + [(XII) + (XIII) + (XX) + (XXI)] + S^o, \]

where, we denote \( J_{i,k}^o \) by \( J_{i,k}^o \), \( \tilde{a}_{i,k}^o \) as \( \tilde{a}_{i,k}^o \) to maintain clean texts,

\[ [(I) + (II) + (III) + (IV)] = \sum_{1+i-d \in J_{i,0} \cap J_{i,j}} \tilde{x}_i'(\tilde{a}_{i,0}^o - \tilde{a}_{i,1}^o) \tilde{x}_i' \left[ (\tilde{a}_1 - \tilde{a}_1^o) + (\tilde{a}_1 - \tilde{a}_1^o) \right] \]

\[ + \sum_{1+i-d \in J_{i,0} \cap J_{i,j}} \tilde{x}_i'(\tilde{a}_{i,2}^o - \tilde{a}_{i,2}^o) \tilde{x}_i' \left[ (\tilde{a}_2 - \tilde{a}_2^o) + (\tilde{a}_2 - \tilde{a}_2^o) \right] \]

\[ + 2 \sum_{1+i-d \in J_{i,0} \cap J_{i,j}} \tilde{x}_i'\left( \tilde{a}_{i,1}^o - \tilde{a}_{i,1}^o \right) e_i^{o+1} + 2 \sum_{1+i-d \in J_{i,0} \cap J_{i,j}} \tilde{x}_i'\left( \tilde{a}_{i,2}^o - \tilde{a}_{i,2}^o \right) e_i^{o+1}, \]
\[(V) + (VI) + (VII) + (VIII) + (X) + (XI) = \sum_{1+i-d \in J_1,0\cap J_2,j} \left[ \tilde{x}^o_i (\tilde{a}_1 - \tilde{a}_{2,j}) \right]^2 + \sum_{1+i-d \in J_2,0\cap J_1,j} \left[ \tilde{z}^o_i (\tilde{a}_1 - \tilde{a}_{1,j}) \right]^2 + \sum_{1+i-d \in J_1,0\cap J_2,j} \left[ \tilde{x}_{i}^o (\tilde{a}_1 - \tilde{a}_{2,j}) \right]^2 + \sum_{1+i-d \in J_2,0\cap J_1,j} \left[ \tilde{x}_{i}^o (\tilde{a}_1 - \tilde{a}_{2,j}) \right]^2 + 2 \sum_{J_1,0\cap J_2,j} \tilde{x}_{i}^o (\tilde{a}_1 - \tilde{a}_{2,j}) e_{i+1}^o + 2 \sum_{J_2,0\cap J_1,j} \tilde{x}_{i}^o (\tilde{a}_1 - \tilde{a}_{1,j}) e_{i+1}^o,\]

\[\sum_{J_2,0\cap J_1,j} \left[ (\tilde{x}_{i}^o - \tilde{z}_{i}^o) (\tilde{a}_1 - \tilde{a}_{1,j}) \right]^2 + 2 \sum_{J_1,0\cap J_2,j} (\tilde{x}_{i}^o - \tilde{z}_{i}^o) (\tilde{a}_1 - \tilde{a}_{2,j}) e_{i+1}^o + 2 \sum_{J_2,0\cap J_1,j} (\tilde{x}_{i}^o - \tilde{z}_{i}^o) (\tilde{a}_1 - \tilde{a}_{1,j}) e_{i+1}^o,\]

\[S_j^o = \sum_{J_1,0\cap J_2,j} \left[ \tilde{x}_{i}^o (\tilde{a}_1 - \tilde{a}_{2,j}) \right]^2 + 2 \sum_{J_1,0\cap J_2,j} \tilde{x}_{i}^o (\tilde{a}_1 - \tilde{a}_{2,j}) e_{i+1}^o + \sum_{J_2,0\cap J_1,j} \left[ \tilde{x}_{i}^o (\tilde{a}_2 - \tilde{a}_{1,j}) \right]^2 + 2 \sum_{J_2,0\cap J_1,j} \tilde{x}_{i}^o (\tilde{a}_2 - \tilde{a}_{1,j}) e_{i+1}^o,\]

The check-hat denotes the replacement of the \(d\)-th coordinate in the vectors with \(r\). For \(1 - cn_6^\Delta \leq j \leq cn_6^\Delta\), define

\[L_j = \left\{ \tilde{z} : \tilde{z} = (\tilde{z}_1 - cn_6^\Delta, \ldots, \tilde{z}_cn_6^\Delta) \in \mathbb{R}^{2cn_6^\Delta}, j = \arg\min_{1 - cn_6^\Delta \leq j \leq cn_6^\Delta} \tilde{z} \right\},\]

\[\hat{L}_j = \left\{ \tilde{z} + h\tilde{v}_j^d : \tilde{z} \in L_j, 0 \leq h \leq 2n^{-\left(\frac{1}{2} - 25\Delta\right)} \right\},\]

\[\tilde{L}_j = \left\{ \tilde{z} - h\tilde{v}_j^d : \tilde{z} \in L_j, 0 \leq h \leq 2n^{-\left(\frac{1}{2} - 25\Delta\right)} \right\}.\]

where \(\tilde{z}_i\) is the \(i + cn_6^\Delta\)-th coordinate of \(\tilde{z}\), \(\tilde{v}_j^d \in \mathbb{R}^{2cn_6^\Delta}\) be of unit length, \(\tilde{v}_j^d = 1\). Claim:

\[(189) \leq P((S_j^o, 1 - cn_6^\Delta \leq i \leq cn_6^\Delta) \in \hat{L}_j, W, Q) + o(n^{-1}). \quad (190)\]

**Proof of the claim. Moment Bounds**

By the preliminary lemmas, there exists \(C > 0\) such that for all \(0 < j \leq cn_6^\Delta\), all large \(n\),

\[E \left\| \frac{\sum_{i=1}^{j} \tilde{x}_{S_i+d-1} e_{S_i+d}}{(\log n) n^{12\Delta}} \right\|^q < C,\]

\[E \left\| \frac{\sum_{1+i-d \in J_k,j} \tilde{x}_{i}^o e_{i+1}^o}{n^{1/2}} \right\|^q < C, k = 1, 2,\]

\[E \left\| \frac{\sum_{1+i-d \in J_k,j} \tilde{x}_{i}^o \tilde{x}_{i}^o}{n} \right\|^{+q} < C, k = 1, 2,\]

\[E \left\| \frac{\sum_{i=1}^{j} \tilde{x}_{S_i+d-1} \tilde{x}_{i}^o}{(\log n) n^{6\Delta}} \right\|^q < C,\]

99
$1-cn^{6\Delta} \leq j \leq 0$ can be established in the same way. By these moment bounds and the same argument as that in Proposition 37 of SETAR(1), we have (190).

Define

$$S_j' = \left\{ \sum_{i=1}^{j} \left[ \tilde{x}_{s_i,2} (\tilde{\alpha}_1 - \tilde{\alpha}_2) \right]^2 + 2 \sum_{i=1}^{j} \tilde{x}_{s_i,1} (\tilde{\alpha}_2 - \tilde{\alpha}_1)e_{s_{i+1}}: j > 0, \right\}
$$

$$\sum_{i=1}^{j+1} \left[ \tilde{x}_{s_i,1} (\tilde{\alpha}_2 - \tilde{\alpha}_1) \right]^2 + 2 \sum_{i=1}^{j+1} \tilde{x}_{s_i,1} (\tilde{\alpha}_2 - \tilde{\alpha}_1)e_{s_{i+1}}: j \leq 0. \right\}$$

By $P(S_i' \neq S_j) = o(n^{-1})$ and lemma 10,

$$\left| (190) - P \left( (S'_i, 1-cn^{6\Delta} \leq i \leq cn^{6\Delta}) \in \tilde{L}_j, W', Q' \right) \right| = O(n^{-(1-18\Delta)}). \quad (191)$$

Define the identically distributed and independent processes such that

\[
\tilde{x}_{i,2}^\prime \tilde{x}_{i,1} \in (x_{T_i}', r, \ldots, x_{T_{i-p+1}}'), i = 1, \ldots,
\]

where

$$T_1 = \inf \{ i : x_{i+d} \in [r, r+cn^{-(1-6\Delta)}], i > p \},$$

and $\tilde{x}_{i,1}$'s are defined in the same way.

\[
\tilde{G}_j = \left\{ \sum_{i=1}^{j} \left[ \tilde{x}_{i,2} (\tilde{\alpha}_1 - \tilde{\alpha}_2) \right]^2 + 2 \sum_{i=1}^{j} \tilde{x}_{i,1} (\tilde{\alpha}_2 - \tilde{\alpha}_1)e_{i,2}, j > 0, \right\}
\]

$$\sum_{i=1}^{j+1} \left[ \tilde{x}_{i,1} (\tilde{\alpha}_2 - \tilde{\alpha}_1) \right]^2 + 2 \sum_{i=1}^{j+1} \tilde{x}_{i,1} (\tilde{\alpha}_2 - \tilde{\alpha}_1)e_{i,1}, j \leq 0. \right\}$$

By Corollary 30, Lemma 10, and that $P(W')$ is trivial,

$$\left| P \left( (S'_i, 1-cn^{6\Delta} \leq i \leq cn^{6\Delta}) \in \tilde{L}_j, W', Q' \right) - P \left( (G_i, 1-cn^{6\Delta} \leq i \leq cn^{6\Delta}) \in \tilde{L}_j \right) \right| = O(n^{-(1-18\Delta)}).$$

Now we connect $\tilde{x}_{1,1}$ to its limiting distribution $\tilde{x}^{lim}$. By (196), assumption 4, 5, lemma 4, 26, 27, we can show (see below) that there exists $C, \gamma^* > 0$ such that for all large $n$, any $\beta^{1,j} \subset \mathbb{R}^{p-d}, \beta^{2,j} \subset \mathbb{R}^{d-1},$ and $\beta = \cup_1^{d,j} \times \{ r \} \times \beta^{2,j} \subset \mathbb{R}^p$,

\[
\sup_{\beta^{1,j}, \beta^{2,j} \geq 0} \left| P(\tilde{x}_{1,1} \in \beta) - P(\tilde{x}^{lim} \in \beta) \right| \leq Cn^{-\gamma^*}. \quad (192)
\]

Note that the limiting process is indexed in $n$(absent script), and (192) confirms the existence of the limiting distribution of $\tilde{x}^{lim}$. By (192) and the frequently used in SETAR(1) technique that we replace each of $\tilde{x}_{cn^{6\Delta},2}, \ldots, \tilde{x}_{1,2}, \tilde{x}_{1,1}, \ldots, \tilde{x}_{cn^{6\Delta},1}$ with their $\tilde{x}_{j,k}^{lim}$ counter part one by one in $G_i, i \in \Lambda$, and sum over the probability differences resulted from each replacement, we have for some $C > 0$, all large $n$,

\[
\left| P((G_i, 1-cn^{6\Delta} \leq i \leq cn^{6\Delta}) \in \tilde{L}_j) - P((G_i^{lim}, 1-cn^{6\Delta} \leq i \leq cn^{6\Delta}) \in \tilde{L}_j) \right| \leq Cn^{-\gamma^*+6\Delta}. \quad (193)
\]

As we have done by assumption 3 for SETAR(1), the last step is to recover the probability for the case $L_j$ for all large $n$. To fit in the argument in that for SETAR(1), where $\tilde{e}_{i,k}$'s
are the only randomness in the cumulative summation, we can conditional on \( x_{i,k}^{\text{lim}} \)'s in \( G_i^{\text{lim}} \)'s; apply the argument similar to that in SETAR(1); do the truncation on \( |\tilde{e}_{i,k}| \geq n^\alpha \) provided the moment bound assumption and assumption 3 holds. Hence

\[
P((G_i^{\text{lim}}, 1 - cn^{6\Delta} \leq i \leq cn^{6\Delta}) \in \tilde{L}_j) \leq P((G_i^{\text{lim}}, 1 - cn^{6\Delta} \leq i \leq cn^{6\Delta}) \in L_j) + Cn^{-\alpha}.
\]

(194)

\[
P((G_i^{\text{lim}}, 1 - cn^{6\Delta} \leq i \leq cn^{6\Delta}) \in L_j) \leq P(j = \arg \min_i G_i^{\text{lim}}) + o(n^{-1}),
\]

(195)

By (189) to (191), (192) to (195), we have finished the proof for \( \tilde{L} \) part. The \( L \) part can be proven in the same fashion, and henceforth we have finished the proof.

Proof of (192). Assume assumption 4 (abuse of notation occurs since \( \beta \) denotes a set in some situation and a scalar in other times) Without loss of generality, let \( c = 1 \) (in \( T_1 \)). Given \( \Delta > 0 \) with \( v_4 = 1 - 6\Delta > \max \{ b, \eta \} > 0 \), where \( \eta \) appers in assumption 4 and \( b \) in Lemma 4. For better expression, in this proof we denote \( x_{1,1} \) as \( x_n \); the process is now explicitly subscripted by \( n \). Define \( \beta_n = \cup_j \beta^{1,j} \times [r, r + n^{-v_4}] \times \beta^{2,j} \subset \mathbb{R}^p \). For \( m \geq n \), there exists some \( v_6 > 0 \), whose value depends on other parameters, such that

\[
P\left( x_m \in \beta \right) = \sum_{j \geq 0} P\left( x_i \notin B_{m^{-v_4}}(r), i \leq j, x_{p+j} \in \beta_m \right)
\]

\[
= \sum_{j \geq 0} P\left( x_i \notin B_{m^{-v_4}}(r), i \leq j, x_{p+j} \in \beta_m \right) + o(m^{-1})
\]

\[
\geq \left\{ (1 + m^{-\gamma})^{1+p} \left[ 1 + \left( \frac{m^{-v_4}}{m^{-\eta} - m^{-v_4}} \right) \right] \right\}^{-e \log m}
\]

\[
\times \sum_{j \geq 0} P\left( x_i \notin B_{m^{-v_4}}(r), i \leq j - c \log m, x_{p+j} \in \beta_m \right) + o(m^{-1})
\]

\[
= P(x_p \in \beta_m) \sum_{j \geq 0} P\left( x_i \notin B_{m^{-v_4}}(r), i \leq j \right) + o(m^{-v_6})
\]

(196)

The second equality is due to Lemma 8, 9, and that

\[
\sum_{j > m^2} P\left( x_i \notin B_{m^{-v_4}}(r), i \leq j, x_{p+j} \in \beta_m \right) \leq P\left( x_i \notin B_{m^{-v_4}}(r), i \leq m^2 \right).
\]

The first inequality comes from moment assumption on \( x_1 \), assumption 4, and techniques similar to those used in Lemma 9. For a sufficient large \( c \), the third equality can be verified by (43) and the same set of techniques constantly used throughout this paper.

We now relate \( P(x_p \in \beta_m) \sum_{j \geq 0} m^{-e \log m} P\left( x_i \notin B_{m^{-v_4}}(r), i \leq j \right) \) to their \( n \)-subscripted analogs. Define \( Q_n = \inf \{ i : x_i \in B_{n^{-v_4}}(r), i > 0 \} \). A standard probability result shows

\[
\sum_{i \geq 1} P\left( Q_n = i \right) = 1.
\]

(197)
By some algebraic manipulation, techniques used previously, and (197), there exists $v_8 > 0$ such that for all large $m$,

$$1 + m^{-v_8} \geq P(x_1 \in B_{m-v_4}(r)) \sum_{j \geq 0} P(x_i \notin B_{m-v_4}(r), i \leq j) \geq 1 - m^{-v_8}. \quad (198)$$

By Lemma 4, for all $m \geq n$ large enough,

$$\left( \frac{m}{n} \right)^{-v_4} P(x_1 \in B_{n-v_4}(r))(1+n^{-\beta}) \geq P(x_1 \in B_{m-v_4}(r)) \geq \left( \frac{m}{n} \right)^{-v_4} P(x_1 \in B_{n-v_4}(r))(1-n^{-\beta}). \quad (199)$$

Combining (198), (199), there exists $v_9 > 0$, whose value depends on other parameters, such that for all large $n, m \geq n$,

$$\sum_{j \geq 0} P(x_i \notin B_{m-v_4}(r), i \leq j) \geq \sum_{j \geq 0} P(x_i \notin B_{n-v_4}(r), i \leq j) \geq 1 - m^{-v_8}. \quad (200)$$

By assumption 4, moment assumption on $x_1$, and the constantly used techniques,

$$P(\tilde{x}_p \in \beta_m) \geq \frac{1}{(1+n^{-\gamma})^{1+p}} \left( \frac{m}{n} \right)^{-v_4} P(\tilde{x}_p \in \beta_n) \quad (201)$$

By (196), (200) and (201) and more algebraic manipulation, there exists $v_{10} > 0$, whose value depends on other parameters, such that

$$P(\tilde{x}_m \in \beta) \geq P(\tilde{x}_n \in \beta) + o(n^{-v_{10}}). \quad (202)$$

The other side of the inequality can be proven in the same way. We have proven the Cauchy sequence converges to a limit; now let $\gamma^*$ be $v_{10}$ to finish the proof. ■

**Moment bound on the threshold and coefficients estimators**

**Theorem 6.** Assume assumption 1 to 5. Then for any $s > 0$, we have some $C > 0$ such that for all large $n, k = 1, 2$,

$$E\left\| n^{1/2}(\tilde{a}_k - \hat{a}_k) \right\|_s^s < C, \quad (203)$$

and

$$E|n(r - \hat{r}_n)|^s < C. \quad (204)$$

**Proof of Theorem 6.** Combining the moment bound results, Proposition 36, $d$ is known and that

$$\hat{a}_1 - \tilde{a}_1 = \left( \sum_{1+i-d \in J_1} \tilde{x}_i \tilde{x}_i' + \sum_{1+i-d \in J_1 \setminus J_1} \tilde{x}_i \tilde{x}_i' \right)^{-1}$$

102
We can finish the proof by the above result, Cauchy-Swartz inequality and a standard H"{o}lder inequality to prove
\[ nE \left[ \left( \hat{a}_1' \tilde{x}_n - \hat{a}_2' \tilde{x}_n \right)^2 1_{x_{1+n-d} \in B_{\tilde{r}_n}(r)} \right] - nE \left[ \left( \hat{a}_1' \tilde{z}_p - \hat{a}_2' \tilde{z}_p \right)^2 1_{x_{1+p-d} \in B_{\tilde{r}_n}(r)} \right] = o(1). \]

Proof of Proposition 38. Define the threshold estimator based on \( x_1, \ldots, x_{n-c\log n} \) as \( \tilde{r}_n \), which is analogous to \( \tilde{r}_n \). By the Proposition 36, for all large \( n \),
\[ P(U_n^c) = o(n^{-1}). \]

Since
\[ E \left[ \left( \hat{a}_1' \tilde{x}_n - \hat{a}_2' \tilde{x}_n \right)^2 1_{x_{1+n-d} \in B_{\tilde{r}_n}(r)} \right] - E \left[ \left( \hat{a}_1' \tilde{z}_p - \hat{a}_2' \tilde{z}_p \right)^2 1_{x_{1+p-d} \in B_{\tilde{r}_n}(r)} \right] = \]
\[ E \left[ \left( \hat{a}_1' \tilde{x}_n - \hat{a}_2' \tilde{x}_n \right)^2 1_{x_{1+n-d} \in B_{\tilde{r}_n}(r)} \right] - E \left[ \left( \hat{a}_1' \tilde{x}_n - \hat{a}_2' \tilde{x}_n \right)^2 1_{x_{1+p-d} \in B_{\tilde{r}_n}(r)} \right] = \]
\[ E \left[ \left( \hat{a}_1' \tilde{x}_n - \hat{a}_2' \tilde{x}_n \right)^2 1_{x_{1+n-d} \in B_{\tilde{r}_n}(r)} \right] - E \left[ \left( \hat{a}_1' \tilde{x}_n - \hat{a}_2' \tilde{x}_n \right)^2 1_{x_{1+p-d} \in B_{\tilde{r}_n}(r)} \right]. \]

We can finish the proof by the above result, Cauchy-Swartz inequality and a standard asymptotical independence argument for \( \tilde{x}_n \) and \( \tilde{r}_n \), and
\[ |nP(x_{1+n-d} \in B_{\tilde{r}_n}(r)) - nP(x_{1+p-d} \in B_{\tilde{r}_n}(r))| = o(1), \]
\[ |nP(x_{1+p-d} \in B_{\tilde{r}_n}(r)) - nP(x_{1+n-d} \in B_{\tilde{r}_n}(r))| = o(1). \]

Hence, it still suffices to prove
\[ |nP(x_{1+n-d} \in B_{\tilde{r}_n}(r)) - nP(x_{1+n-d} \in B_{\tilde{r}_n}(r))| = o(1). \quad (205) \]
\[ |nP(x_{1+p-d} \in B_{\tilde{r}_n}(r)) - nP(x_{1+p-d} \in B_{\tilde{r}_n}(r))| = o(1). \quad (206) \]
\[ |nP(x_{1+p-d} \in B_{\tilde{r}_n}(r)) - nP(x_{1+n-d} \in B_{\tilde{r}_n}(r))| = o(1). \quad (207) \]

SETAR(p) is barely involved in these arguments, and we have the results follow the arguments in SETAR(1).

\[ 103 \]
Proposition 39. Assume assumption 1 to 5,

\[ \left| nP(x_{1+p-d}^n \in B_{r_n^2-r}(r)) - \pi_r E |r_\infty| \right| = o(1). \]

**Proof.** See the proof for SETAR(1). 

Proposition 40. Assume assumption 1 to 5. Then

\[ \lim_{n \to \infty} E \left[ \left( \tilde{a}_2 \tilde{x}_p^m - \tilde{a}_1 \tilde{x}_p^m \right)^2 \right| x_{p-d+1}^m \in B_{r_n^2-r}(r) \] = \[ \frac{nE \left[ \left( \tilde{a}_2 \tilde{x}_p^m - \tilde{a}_1 \tilde{x}_p^m \right)^2 1_{x_{p-d+1}^m \in B_{r_n^2-r}(r)} \right]}{nP \left( x_{p-d+1}^m \in B_{r_n^2-r}(r) \right)} . \]

where \( \tilde{x}_p^m = (x_1^m, \ldots, x_p^m) \) is an independent random vector distributed as \( \tilde{x}_p \).

**Proof.** Note that

\[ E \left[ \left( \tilde{a}_2 \tilde{x}_p^m - \tilde{a}_1 \tilde{x}_p^m \right)^2 \right| x_{p-d+1}^m \in B_{r_n^2-r}(r) \] = \[ \frac{nE \left[ \left( \tilde{a}_2 \tilde{x}_p^m - \tilde{a}_1 \tilde{x}_p^m \right)^2 1_{x_{p-d+1}^m \in B_{r_n^2-r}(r)} \right]}{nP \left( x_{p-d+1}^m \in B_{r_n^2-r}(r) \right)} . \]

We argue the existence of limit of the numerator. By the now standard techniques used throughout this paper,

\[ nE \left[ \left( \tilde{a}_2 \tilde{x}_p^m - \tilde{a}_1 \tilde{x}_p^m \right)^2 1_{x_{p-d+1}^m \in B_{r_n^2-r}(r)} \right] - nE \left[ \left( \tilde{a}_2 \tilde{x}_p^m - \tilde{a}_1 \tilde{x}_p^m \right)^2 1_{n(x_{p-d+1}^m \in B_{r_n^2-r}(r))} \right] = o(1). \] (208)

To deal with the RHS of (208), note that by the moment bound condition on \( r_\infty \) (see (163)), we have for all \( m \geq n \), all large \( n \),

\[ E \left[ \left( \tilde{a}_2 \tilde{x}_p^m - \tilde{a}_1 \tilde{x}_p^m \right)^2 1_{n(x_{p-d+1}^m \in B_{r_n^2-r}(r))} \right] \]

\[ = \int_0^n \int_{\mathbb{R}^{2p-d \times [r, r + \frac{p}{n}] \times \mathbb{R}^{d-1}}} \left[ \tilde{z}_p (\tilde{a}_1 - \tilde{a}_2) \right]^2 f_c(z_p - T(\tilde{z}_{p-1})) \times \cdots \times f_c(z_1 - T(\tilde{z}_0)) \tilde{\pi}(\tilde{z}_0) d\tilde{z}_p d\tilde{z}_0 dF_{r_\infty}(y) \]

\[ + \int_0^n \int_{\mathbb{R}^{2p-d \times [r, r + \frac{p}{n}] \times \mathbb{R}^{d-1}}} \left[ \tilde{z}_p (\tilde{a}_1 - \tilde{a}_2) \right]^2 f_c(z_p - T(\tilde{z}_{p-1})) \times \cdots \times f_c(z_1 - T(\tilde{z}_0)) \tilde{\pi}(\tilde{z}_0) d\tilde{z}_p d\tilde{z}_0 dF_{r_\infty}(y) \]

\[ + o(n^{-1}). \] (209)

In (209) we expand the integration to consider more lag variables \( (\tilde{z}_p^, \tilde{z}_0^) = (z_p, \ldots, z_{p+1}) \)

and make one truncation on the value of \( r_\infty \); the notation \( T(.) \) stands for the non-linear autoregressive function of SETAR process. By techniques similar to those used in Lemma 9, assumption 4, moment bounds for \( x_1^m \), \( f_c(.) \), we have

\[ (209) \geq (1 + n^{-\gamma})^{-d} \left( \frac{m}{n} \right) E \left[ \left( \tilde{a}_2 \tilde{x}_p^m - \tilde{a}_1 \tilde{x}_p^m \right)^2 1_{n(x_{p-d+1}^m \in B_{r_n^2-r}(r))} \right] + o(n^{-1}). \] (210)

The other side of inequality can be proven in the same way. By (209) and (210), we see

\[ nE \left[ \left( \tilde{a}_2 \tilde{x}_p^m - \tilde{a}_1 \tilde{x}_p^m \right)^2 1_{n(x_{p-d+1}^m \in B_{r_n^2-r}(r))} \right] \]

is a Cauchy sequence and hence converges to a limit. Then by (208) we see the numerator converges to a limit.

The limit of denominator has already been handled (see Proposition 39).