An Introduction to Nonstationary Time Series Analysis

Ting Zhang¹
tingz@bu.edu

Department of Mathematics and Statistics
Boston University

August 15, 2016

¹I am grateful for the support of NSF Grant DMS-1461796. Some of the figures used in this presentation are from Wikipedia or other websites.
Welcome to Boston

ボストンへようこそ

Ting Zhang (BU) Nonstationary Time Series
Introduction

- A time series is a sequence of measurements collected over time, and examples include
  - electroencephalogram (EEG);
  - stock price;
  - temperature series;
  - and many others.

The first human EEG recording obtained by Hans Berger in 1924. Upper: EEG. Lower: timing signal.
An important feature of time series is the temporal dependence.

- Observations collected at different time points depend on each other.
- The common assumption of independence no longer holds.

**Example:** Let $X_1, \ldots, X_n$ be random variables, sharing common marginal distribution $N(\mu, \sigma^2)$.

- If independent, then

$$\bar{X}_n = \frac{X_1 + \cdots + X_n}{n} \sim N\left(\mu, \frac{\sigma^2}{n}\right),$$

and a $(1 - \alpha)$-th confidence interval for $\mu$ is given by

$$\left[\bar{X}_n \pm q_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}\right].$$

- Under dependence, the above constructed confidence interval may no longer preserve the desired nominal size, as the distribution of $\bar{X}_n$ can be different.
Example (Continued): We shall here consider an illustrative dependence case. Let $\epsilon_0, \ldots, \epsilon_n$ be independent standard normal random variables, and set $X_i = \mu + \sigma (\epsilon_i + \epsilon_{i-1}) / \sqrt{2}$, then

- $X_i \sim N(\mu, \sigma^2)$ has the same marginal distribution, but not independent. In fact, $\text{cor}(X_i, X_{i-1}) = 0.5$.
- In this case, it can be shown that (exercise?)

$$
\bar{X}_n = \frac{X_1 + \cdots + X_n}{n} \sim N\left\{ \mu, \frac{\sigma^2 (2n - 1)}{n^2} \right\},
$$

and a $(1 - \alpha)$-th (asymptotic) confidence interval for $\mu$ is given by

$$
\left[ \bar{X}_n \pm q_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \times \sqrt{2} \right].
$$

Dependence makes a difference!

- Will the “magic” $\sqrt{2}$ adjustment work for other dependent data? Generally not!
To incorporate the dependence, parametric models have been widely used, and popular ones are:

- **Autoregressive (AR) models:**
  \[ X_i = a_1 X_{i-1} + \cdots + a_p X_{i-p} + \epsilon_i; \]

- **Moving average (MA) models:**
  \[ X_i = \epsilon_i + a_1 \epsilon_{i-1} + \cdots + a_q \epsilon_{i-q}; \]

- **Threshold AR models:**
  \[ X_i = \rho |X_{i-1}| + \epsilon_i; \]

- and many others.

When using parametric models, the dependence structure is fully described except for a few unknown parameters, which makes the inference procedure easier and likelihood-based methods are often used.

**Model misspecification issues!**
Stationarity: Another Common Assumption

- A process is said to be stationary when its joint probability distribution does not change when shifted in time.
  - Implying weak stationarity where
    \[
    E(X_i) = E(X_0), \quad i = 1, \ldots, n;
    \]
    \[
    \text{cov}(X_i, X_{i+k}) = \text{cov}(X_0, X_k), \quad i = 1, \ldots, n.
    \]
  - Under (weak) stationarity, it makes sense to estimate the mean as a single parameter. The same holds for the marginal variance and autocorrelation.

- However, stationarity is a strong assumption and can be violated in practice.

The first human EEG recording obtained by Hans Berger in 1924.

Upper: EEG. Lower: timing signal.
Introduction

A Sample Problem

Other Interesting Problems

Ting Zhang (BU) Nonstationary Time Series

S&P 500

[Graph of S&P 500 stock prices from 1950 to 2013]
Nonstationary time series analysis has been a challenging but active area of research.

When the assumption of stationarity fails, parameters of interest may no longer be a constant. In this case, they are naturally modeled as functions of time, which are infinite dimensional objects.

Questions of interests:

- How to estimate those functions?
- Can parametric models be used to describe the time-varying pattern?
- How to make statistical inference, including hypothesis testing and constructing simultaneous confidence bands?
- Is it possible to provide a rigorous theoretical justification for those methods?
- Can we use the developed results to better address some real life problems?
Testing Parametric Assumptions on Trends of Nonstationary Time Series

A Sample Problem
Motivating Examples (Tropical Cyclone Data)

Figure: Satellite-derived lifetime-maximum wind speeds of tropical cyclones during 1981–2006

- In atmospheric science (*constancy or not?):
  - Global warming has an impact on tropical cyclones (Emanuel, 1991, Holland, 1997 and Bengtsson et al., 2007);

- Is the mean really a constant?
Figure: Annual central England temperature series from 1659 to 2009

- **In climate science** (*various models proposed*):
  - Linear (Jones and Hulme, 1997);
  - Quadratic (Benner, 1999, *Int. J. Climatol.*);
  - Local polynomial (Harvey and Mills, 2003).

- Which model should we use?
Suppose we observe:

\[ y_i = \mu(i/n) + e_i, \quad i = 1, \ldots, n, \]

- \( \mu(t), \ t \in [0, 1], \) is an unknown smooth trend function;
- \((e_i)\) is the error process (dependent and nonstationary).

Motivated by the CET data, we want to test:

\[ H_0 : \mu(t) = f(\theta, t). \]

Examples: \( f(\theta, t) = \theta_0 \) or \( f(\theta, t) = \theta_0 + \theta_1 t. \)

A natural strategy:

Nonparametric: \( \hat{\mu}_n(t) \) v.s Parametric: \( f(\hat{\theta}_n, t). \)

We form the \( \mathcal{L}^2 \)-distance

\[ \Delta = \int_0^1 \{\hat{\mu}_n(t) - f(\hat{\theta}_n, t)\}^2 dt, \]

and reject \( H_0 \) if \( \Delta \) is large.

Ref: Fan and Gijbels (1996)
Introduction

A Sample Problem

Other Interesting Problems

Introduction

A Sample Problem

Other Interesting Problems

“Locally fits a linear line”: $\hat{\mu}_n(t) = \sum_{i=1}^{n} y_i w_i(t)$.

- Bias: $O(b_n^2)$;
- Variance: $O(1/\sqrt{nb_n})$.

- The window size $b_n$ satisfies $b_n \to 0$ and $nb_n \to \infty$;

Ref: Fan and Gijbels (1996)
Recall the test statistic
\[ \Delta = \int_0^1 \{ \hat{\mu}_n(t) - f(\hat{\theta}_n, t) \}^2 dt. \]

In order to develop a rigorous statistical test, we need an asymptotic theory on the closely related integrated squared error:
\[ \text{ISE} = \int_0^1 \{ \hat{\mu}_n(t) - \mu(t) \}^2 dt. \]

Reason: the error of \( \hat{\mu}_n(t) \) dominates that of \( f(\hat{\theta}_n, t) \).

In the literature:
- Independent errors:
- Stationary linear error processes:
  - González-Manteiga and Vilar Fernández (1995);
  - Biedermann and Dette (2000).
- Nonstationary nonlinear error processes: ??? Need a good framework.
Time-Varying Causal Representation

The Framework
We assume that the error process has the causal representation:

\[ e_i = H(i/n; \mathcal{F}_i), \quad \mathcal{F}_i = (\ldots, \epsilon_{i-1}, \epsilon_i), \]

for some measurable function \( H : [0, 1] \times \mathbb{R}^\infty \to \mathbb{R}. \)

“A time-varying function of past innovations”

Examples:

- Stationary causal processes: \( e_i = H(\mathcal{F}_i) = H(\ldots, \epsilon_{i-1}, \epsilon_i); \)

  “Include popular linear and nonlinear time series as special cases”

- Nonstationary linear processes: \( e_i = \epsilon_i + a_1(i/n)\epsilon_{i-1} + \cdots. \)

  “Nonstationary generalization by allowing time-varying parameters”

Easy to work with, and makes developing asymptotic theory of complicated statistics possible.

Under this framework, an asymptotic theory was developed in Zhang and Wu (2011, *Biometrika*) for the test statistic, and the cut-off value can then be obtained.
Tropical Cyclone Data

In the literature:
- Linear trends for quantiles (Elsner et al., 2008, Nature);
- $\mathcal{L}^\infty$-based test: accept the mean constancy at the 5% level (Zhou, 2010, Ann. Statist.).

Our procedure:
- Reject the mean constancy at the 5% level.
  “$\mathcal{L}^2$-based tests can be more powerful than $\mathcal{L}^\infty$-based ones”

Figure: Satellite-derived lifetime-maximum wind speeds of tropical cyclones during 1981–2006
Central England Temperature Data

Figure: Annual central England temperature series from 1659 to 2009 with dashed curve representing the global cubic trend

- **In climate science:**
  - **Linear** trend (Jones and Hulme, 1997);
  - **Quadratic** trend (Benner, 1999, *Int. J. Climatol.*);
  - **Local polynomial** trend (Harvey and Mills, 2003).

- **Our procedure:**
  - **Quadratic** trend (*Reject* with *p*-value 0.00);
  - **Cubic** trend (*Accept* with *p*-value 0.47).

*Ref: Jones and Bradley (1992)*
Other Interesting Problems
Introduction

A Sample Problem

Other Interesting Problems

Regression Setting

Figure: Daily hospital admissions (top) and measurements of pollutants in Hong Kong between January 1, 1994 and December 31, 1995.

**Multivariate Setting**

*Figure:* Air temperature measurements at 15 measurement facilities in the Southern Great Plains region of the United States from 10/06/2005 to 10/30/2005.

*Ref:* Degras et al. (2012, IEEE), Zhang (2013, JASA) and Zhang (2016, Sinica)
Potential Jump Setting

Figure: Monthly global temperature anomalies in Celsius from 01/1850 to 12/2012. The period between the two dashed lines corresponds to 11/1944–12/1946.

Thank You!

NAME: TING ZHANG

ADDRESS: DEPARTMENT OF MATHEMATICS AND STATISTICS
          BOSTON UNIVERSITY
          BOSTON, MA 02215, U.S.A.

E-MAIL: tingz@bu.edu