12-2009

Some New Problems in Changepoint Analysis

Jonathan Woody
Clemson University, jwoody@clemson.edu

Some New Problems in Changepoint Analysis

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Mathematical Science

---

by
Jonathan R. Woody
December 2009

Accepted by:
Dr. Robert B. Lund, Committee Chair
Dr. Peter C. Kiessler
Dr. Colin M. Gallagher
Dr. James R. Brannan

Abstract

Climatological studies have often neglected changepoint effects when modeling various physical phenomena. Here, changepoints are plausible whenever a station location moves or its instruments are changed. There is frequently meta-data to perform sound statistical inferences that account for changepoint information. This dissertation focuses on two such problems in changepoint analysis.

The first problem we investigate involves assessing trends in daily snow depth series. Here, we introduce a stochastic storage model. The model allows for seasonal features, which permits the analysis of daily data. Changepoint times are shown to greatly influence estimated trends in one snow depth series and are accounted for in this analysis. The model is fitted by numerically minimizing a sum of squares of daily prediction errors. Standard errors for the model parameters, useful in making trend inferences, are presented. The methods are illustrated in the analysis of a century of daily snow depth observations from Napoleon, North Dakota. The results here show that snow depths are significantly declining at Napoleon, with spring ablation occurring earlier, and that changepoint features are very influential in deriving realistic trend estimates.

The second problem considers the asymptotic statistical properties of parameters in a general linear model experiencing infinitely many level shifts occurring at known times. It is felt that this setting is more realistic than standard infill asymptotics, where the number of data points between all changepoint times converges to infinity. Here, least squares estimators for m trend parameters are derived, and consistency is proven in the case of a short-memory time series error process.

Dedication

This dissertation is dedicated to, in accordance with King Diamond's perverse religious philosophies, my family and friends.

Table of Contents

List of Tables

List of Figures

Chapter 1

Changepoint Modeling

1.1 Changepoints in Climatology

A changepoint (also called breakpoint) is a time of discontinuity in the structure of a time series of data $\{X_1, X_2, \ldots\}$. A series could experience a change in first moment, variance, or even in distribution. In general, a changepoint is any change in the marginal distribution of the sequence of data. Statisticians now recognize the importance of changepoints in inferential aspects. Many changepoints, particularly in climate settings, manifest themselves as level shifts in the mean of a stochastic process. This aspect will be studied further here.

Although changepoint studies can be quite technical, the general idea is to segment data into several homogeneous regimes before any statistical inferences on the data are conducted. Consideration of changepoints can enhance many statistical estimation techniques, simultaneously increasing a model's adherence to that assumed for the process and reducing estimation variance. As such, changepoint analysis is a crucial component in the study of many natural processes.

Changepoints have become a popular avenue of theoretical investigation in the second half of the twentieth century. Mathematical studies have taken multiple lines of interest. For example, changepoint detection reached into an ever broadening array of data driven studies. Page (1954, 1955) introduced undocumented changepoint probems. Csörgő and Horváth (1998) provide a technical asymptotic analysis of changepoint detection methods. Brodsky and Darkhovsky (1993) provide a consolidated text on nonparametric changepoint methods.

Real world examples of processes with heterogeneous data structures are ubiquitous in the physical sciences. For example, the ARCH and GARCH models are popular in financial time series modeling. Kokoszka and Leipus (2000) study changepoint estimation techniques in ARCH models. Such techniques can be used to detect changepoints in the variance of stock-return data, which could be useful to the financial engineer when constructing exotic financial devices to capitalize on risk associated with market fluctuations.

In climate settings, one must deal with both documented and undocumented changepoints. Work on undocumented changepoint methods in applied to climate studies continue to grow (Lund and Reeves (2002), Menne Williams (2005)). Additionally, one often has numerous documented changepoints for a given set of data. For example, temperature stations in the US average about 6 documented changepoints per century, including changes in station location and instrumentation (Mitchell (1953)).

Trend analysis methods that account for changepoints have already been applied to temperature data (Lund et al. (2001)). Figure 1.1 shows a century of average yearly temperatures at Tuscaloosa, AL. There are three documented changepoints in the meta-data record. The first changepoint occurred in 1939 when the thermometer was changed. The second occurred in 1957 when the station was moved, and finally, the third in 1987 when the station location and thermometer were changed simultaneously.

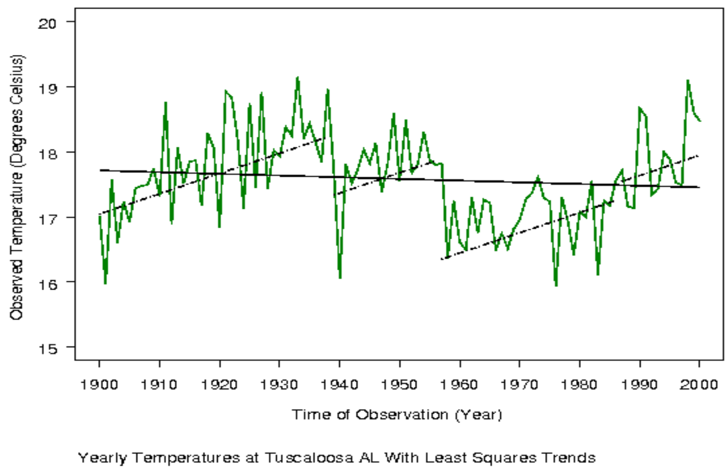

Figure 1.1: Annual Temperatures at Tuscaloosa, AL

A study of the 1957-2000 sub-series judged the 1987 changepoint statistically insignificant via standard t-tests. In Figure 1.1, the solid line represents a "naive" trend estimate computed via ordinary least squares that ignores the three changepoint times. The term naive refers to the assumption of homogeneous data. The dashed lines depict a fitted least square model with the assumption of four distinct homogeneous, but mean shifted, regimes. Here, the trend slope of the four segments is required to be the same.

The naive approach yields a trend slope estimate of $-0.302 \pm 0.776\,^{\circ}$C/Century. The data fitted with three changepoints yields a trend estimate of $3.517 \pm 0.419\,^{\circ}$C/Century. We note that the two trend estimates take a different sign, and are radically different under any reasonable statistical assumptions. The standard errors quoted represent one standard deviation and account for autocorrelation in the series. One sees the importance of changepoints.

1.2 A Storage Model for the Assessment of Snow Depth Trends

It is important to understand snow cover trends. Environmental effects of changes in the snow cover are numerous. Since seasonal snow melt plays a crucial role in water resources of which wildlife and agricultural interests are sensitive, a detailed study of snow cover processes is prudent. Since snow cover is a function of both temperature and precipitation, changes in global climate can induce large fluctuations in snow cover processes.

In the next chapter, we study snow cover trends. Increasing global temperatures do not necessarily translate into decreasing snow cover. Indeed, slightly warming subfreezing air allows it to hold more moisture, which could increase the mean amount of snow deposited per snowfall as well as altering the number or snowfall events induced by changes in atmospheric patterns. Also, snowfall morphology depends on temperature as shown in Figure 1.2 (Libbrecht (2005)). This figure illustrates that the characteristics of snow flakes change as a function of temperature and atmospheric H_2O supersaturation. With such vastly different snow flake structure, the overall water density per unit volume will significantly change. Phrased another way, the snow water equivalent of 10 inches of new snow at 10°F is much less than the snow water equivalent of 10 inches of new snow at 32°F.

Historical climate data in the United States includes measurements of maximum and minimum daily temperatures, precipitation, snow depth, and daily snowfall. Snow depth records are the richest source of long-term snow cover data suitable

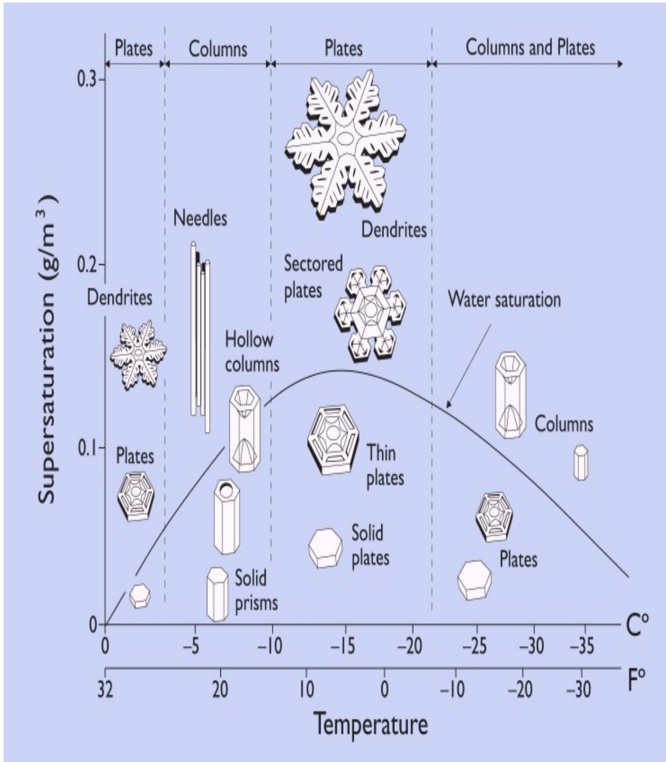

Figure 1.2: Snowfall Morphology Picture.

for statistical study. The National Weather Service Cooperative Observer Program (COOP) is a collection of over 11,000 volunteers that forms the U.S. Cooperative Observer Network (USCON). Unfortunately, the length and quality of each series varies significantly in the USCON. Nonetheless, such a richness in spatial data should be useful in joint space-time analyses. Additionally, the United States Historical Climatology Network (USHCN) is a subset of (USCON), as depicted in Figure 1.3, and consists of approximately 1,200 locations judged to be of higher quality in the contiguous United States.

Also, there are issues with data measurements being taken under different metrics. For example, the Napoleon, ND data, which will be the focus of the next chapter, was originally taken in inches and later converted to centimeters. This conversion produced a plethora of measurements of exactly 2.54 centimeters. This is

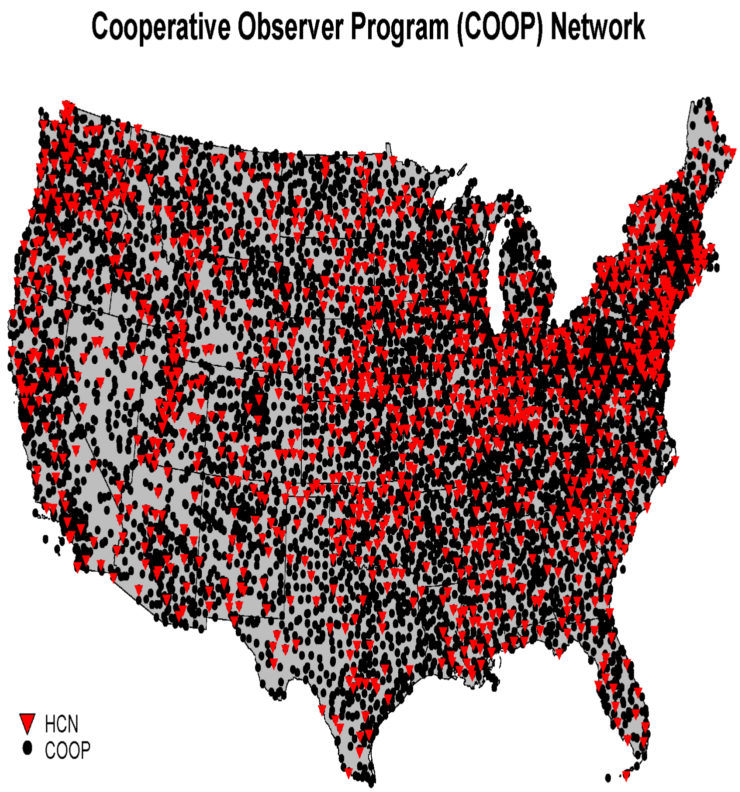

Figure 1.3: Continental United States COOP Weather Stations.

essentially a rounding issue as snow depths in the US are observed to the nearest inch. Additionally, observers and observer's standards of measurement are not uniform across North America. For example, daily snow depth records are often taken by volunteer (cooperative) observers (Robinson and Hughes (1991)). Volunteer observers usually only record snow depths once per day as opposed to measurements being taken every 6 hours by National Weather Service stations.

Changes in observer, time of observation, and location of observation can degrade trend estimation. Changes in time of observation are important since snow undergoes compactification throughout the day; therefore, afternoon and evening measurements are often less than that in the morning for a given day when there is no new additional snowfall. New observers can introduce different decision making processes when dealing with drifts or very low snow cover. For example, observer judgment is key in determining snow cover when the overall snow cover is not 100% of ground. The snow depth is considered to be a trace when 50% or less of the ground is covered with snow, but the observer must make this decision. Therefore, lower snow depths are particularly sensitive to changes in observer. Finally, different locations will have different drift patterns, soil makeups, and vegetation; changing any of these factors may induce a changepoint in snow depth record. In short, snow data are thoroughly maligned with changepoint effects. Trend detection with snow depth data has yet to receive attention within the context of changepoints. In the next chapter, we introduce a stochastic storage model with periodic dynamics that is capable of modeling daily snow depths and assessing their trends. The model allows for multiple known changepoint times.

In the course of this investigation, shot noise, diffusions, and other stochastic models were considered but were not adopted for various physical and statistical reasons. For example, shot noise models did not seem to accurately portray snow depth changes.

1.3 A General Linear Model with Infinitely Many Level Shifts

Chapter 3 investigates fitting a general linear model (GLM) to accommodate infinitely many mean level shifts. For a time series $\{X_t\}$ sampled from a stochastic process, the governing regression is

$$X_t = \mu + \theta_1 f_1(t) + \cdots + \theta_m f_m(t) + \epsilon_t, \qquad 1 \le t \le n, \tag{3.1}$$

where $\theta_1, \ldots, \theta_m$ are unknown parameters, $f_1(t), \ldots, f_m(t)$ are regression factors, μ is the intercept, and ϵ_t is the error at time t. We assume that $\{\epsilon_t\}$ is a stationary zero-mean time series.

To allow the intercept μ to shift infinitely often at the known times $\tau_1 \ldots, \tau_k$, we examine the regression equation

$$X_t = \theta_1 f_1(t) + \cdots + \theta_m f_m(t) + \delta_t + \epsilon_t. \tag{3.2}$$

Here, $\{\delta_t\}$ represents the level shift factor. For a fixed sample size n and $1 \le t \le n$, the level shift at time t is

$$\delta_t = \begin{cases} \Delta_1, & 1 \le t < \tau_1 \\ \Delta_2, & \tau_1 \le t < \tau_2 \\ \vdots & \vdots \\ \Delta_k, & \tau_{k-1} \le t \le \tau_n \end{cases},$$

where $k = k(n)$ is the number of changepoints up to time n and $\tau_0, \tau_1, ..., \tau_{k-1}$ are the ordered level shift times, with $\tau_0 = 1$ being a convention. We consider the times of all level shift changes, $\{\tau_i\}_{i=0}^{\infty}$, to be known. Hence, $\tau_0, ..., \tau_{k-1}$ partition $\{1, \ldots, n\}$ into $k(n)$ regimes.

In Chapter 3, we identify and prove consistency of the least squares estimators of $\theta_1, \ldots, \theta_m$. This is a difficult task under general time series errors $\{\epsilon_t\}$.

References

Brodsky, B.E., and Darkhovsky, B.S. (1993), *Nonparametric Methods in Change-point Problems*, Kluwer Academic Publishers, AA Dordrecht, The Netherlands.

Csörgő, M, and Horváth, L. (1998), *Limit Theorems in Change-point Analysis*, John Wiley and Sons Ltd., New York City.

Kokoszka, P., and Leipus, R. (2000), Change-point estimation in ARCH models, *Bernoulli*, 60, 513-539.

Menne, M.J., and Williams, C.N. Jr. (2005), Detection of undocumented changepoints using multiple test statistics and composite reference series hydrology, *J. Climate*, 18, 4271-4286.

Mitchell, J.M., Jr. (1953), On the causes of instrumentally observed secular temperature trends, *Journal of Meteorology*, 10, 244-261.

Page, E.S. (1954), Continuous inspection schemes, *Biometrika*, 41, 100-105.

Page, E.S. (1955), A test for a change in a parameter occurring at an unknown point, *Biometrika*,42, 523-527.

Libbrecht, K.G. (2005), The physics of snow crystals, *Reports on Progress in Physics*, 68, 855-895.

Lund, R., Seymour, and L. Kafadar, K. (2001), Temperature trends in the United States, *Environmetrics*, 12, 673-690.

Lund, R., and J. Reeves (2002), Detection of undocumented changepoints: a revision of the two-phase regression model, *J. Climate*, 15, 2547-2554 .

Robinson, D.A., and M.G. Hughes (1991), Snow cover variability in the Northern and Central Great Plains, *Great Plains Research*, 1, 93-113.

Chapter 2

A Storage Model for the Assessment of Snow Depth Trends

2.1 Introduction

This chapter introduces a statistical method to assess long-term trends in snow depth time series. Snow is an important geophysical and environmental quantity that is very sensitive to climate change since its magnitude depends on both temperature and precipitation (Kukla (1979), Barry (1990)). Global climate models indicate that snow cover changes will considerably impact the cryospheric portion of the water budget in a greenhouse enhanced climate (Barnett et al. (2005)). Studies suggest that the duration of snow cover may decrease by 40% in the Canadian Prairies and by 70% in the Great Plains (Boer et al. (1992), Brown et al. (1995)). Negative trends in snow have been observed across many areas in the Western United States (Mote et al. (2005), Hamlet et al. (2005)). Such changes can dramatically impact local hydrology. An earlier Spring ablation in some regions of North America has been linked to earlier maximum stream flow dates (Barnett et al. (2005), Burn (1994)).

Reduced snow cover has been linked to a lengthened growing season at high latitudes (Myeni et al. (1997)), while changes in snow depths have been tied to soil temperature changes across the Great Plains (Schmidt et al. (2001), Grundstein et al. (2005)).

Ground based snow depths, which are commonly measured and often have long periods of record, are frequently used in hydrological and climatological studies as surrogates for snow mass and as measures of thermal insulation of the ground. Furthermore, the long period of record available at many stations makes these data very useful for climate change assessments.

While issues of data quality (Robinson (1993)) and data scaling Blöschl (1999)) have previously been addressed, snow data have not been critically evaluated with regard to breakpoints. A breakpoint occurs whenever there is a change in the location of the station, a change in the observer, or a change in the station instrumentation. Breakpoints profoundly affect geophysical data such as temperatures (Lu and Lund (2007)), but no study has quantitatively examined breakpoint effects on snow data. Because of the blowing and drifting nature of snow, accounting for station location changes is critical: an observation taken in a shaded area near a fence can be much deeper than one taken in the open. Fortunately, our data comes replete with meta-data, a history of the conditions under which the data was observed.

This paper attempts to accurately quantify trends in daily snow depth series in a statistical manner. We focus on one station, illuminating the features in the record that are influential in trend assessment. The model introduced is capable of describing daily observations. Modeling daily data is a challenge for several reasons. First, day-to-day snow depths are highly correlated in time and inference methods that assume independent and identically distributed data will give exaggerated levels of statistical confidence (Fuller (1996) is a comprehensive reference on such issues). Second, daily snow depths have a seasonal structure, with larger depths being more common in mid-winter through early Spring. Third, snow depths cannot be negative, but have a positive probability of being zero. In fact, snow depths are zero during the summer at all but the most alpine or Arctic stations. Accurate inferences should take into account this "zero modified support set issue". Marsh (1999) is a good general snow dynamic reference.

The model we adopt is rooted in storage modeling. The only other author to pursue storage models in describing snow depths is Perona et al. (2007), who employs a model with two phases: increasing snow depths (the accumulation phase) and decreasing snow depths (the ablation phase). The model in Perona et al. (2007) does not have a trend component nor does it allow for melting during the accumulation phase or accumulation during the ablation phase.

The rest of this paper proceeds as follows. Section 2.2 introduces the data set that we study, a century of daily snow depth observations from Napoleon, North Dakota. Section 2.3 describes a storage model for the snow depth process; such a model allows for the features noted above. Section 2.4 shows how to estimate the parameters in the model and Section 2.5 applies the methods to make inferences about snow depth changes at Napoleon. Section 2.6 concludes with comments.

2.2 The Napoleon Data

Figure 2.1 displays 37595 daily snow depth observations taken at Napoleon, North Dakota from 01 January 1901 — 31 December 2003. Napoleon is located at latitude $46°30'17''$N, longitude $99°46'1''$W, resides at 1959 feet above sea level, has a mean annual temperature of 40.45 degrees F, receives 17.9 inches of liquid precipitation a year, and 36.4 inches of snow per year. This data has a well documented station history and extends back over 100 years. Only about 1.5% of the daily observations are missing, mostly at "seemingly random times". Robinson (1988) judged the Napoleon record to be of a uniquely high quality after performing intensive quality checks of stations across the United States. In Figure 2.1, one can see that peak snow depths vary considerably from year to year. The late 1970s are exceptionally snowy. A simple linear estimate of the annual snowfalls at Napoleon from 1931-2003 shows a positive increase of 0.633 cm per year. This increase in new snow may not translate to increasing snow depths for a variety of reasons, including increasing temperatures which may lead to snowpack compaction and ablation.

To gain a feel for an annual snow depth cycle, Figure 2.2 shows the observations for the 1977-1978 winter season. For this snow season, depths peaked in early March, followed by Spring ablation in early April. Snow from October and November storms is evident, but rapidly melted off. Snow cover was continual from early December until Spring ablation in early April. These depths do not appear to compactify between snowfall events. It is unclear to us whether this is a consequence of the observering techniques, the climatological characteristics of the site, or the fact that the original depths were rounded to the nearest inch (or some combination thereof). As our model accommodates data of any type, this issue is not overly important in what follows.

Table 2.1 below shows the meta-data record for the Napoleon station. Over the 103 years of record, there are some 18 changes in station location, observer, and observation times. Any of these changes may induce a level shift in the series. A casual inspection of Figure 2.1 and Table 2.1 reveals concerns. In particular, the three largest snow depths occurred during the 1976-1985 regime when the recording station is listed as Warren Wentz's Mother's house. Whether or not this era was truly this snowy is confounded with breakpoint and station location issues. Because of drifting tendencies, snow depths can significantly vary when the measuring location is moved only a few feet. While reference station comparisons could partly resolve such issues,

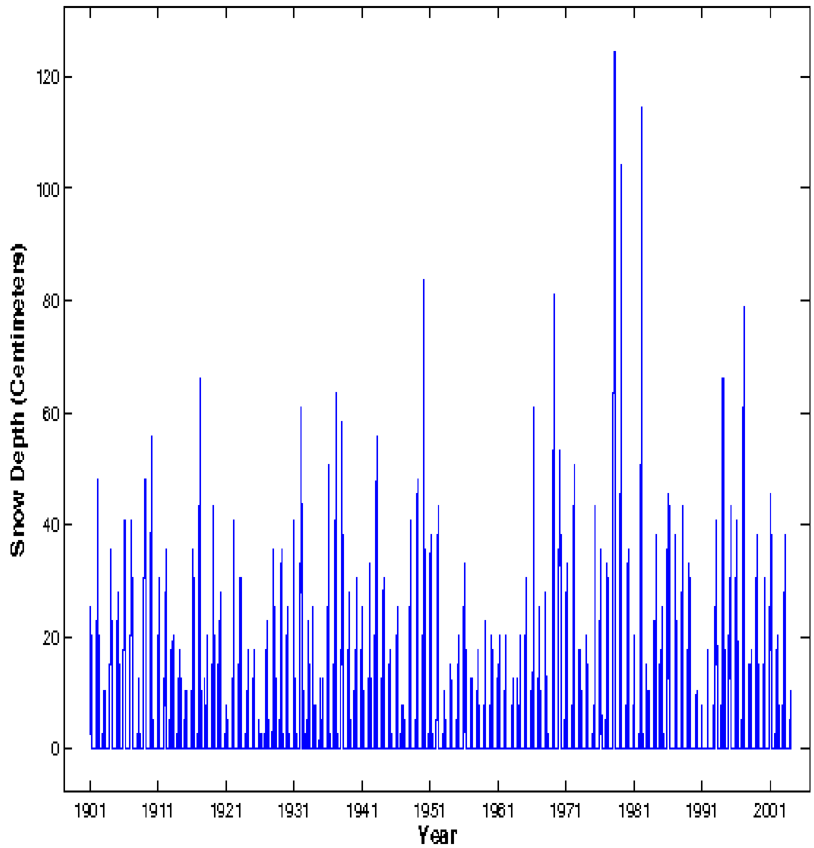

Figure 2.1: Daily Snow Depths at Napoleon, ND from 01 January 1901 — 06 December 2003.

no good reference station exists that spans the duration of the record. The point here is that single station methods will need to take into account breakpoint information in examining long-term trends in the snow depth record. With the problem and data now elaborated upon, we introduce a model capable of assessing trends.

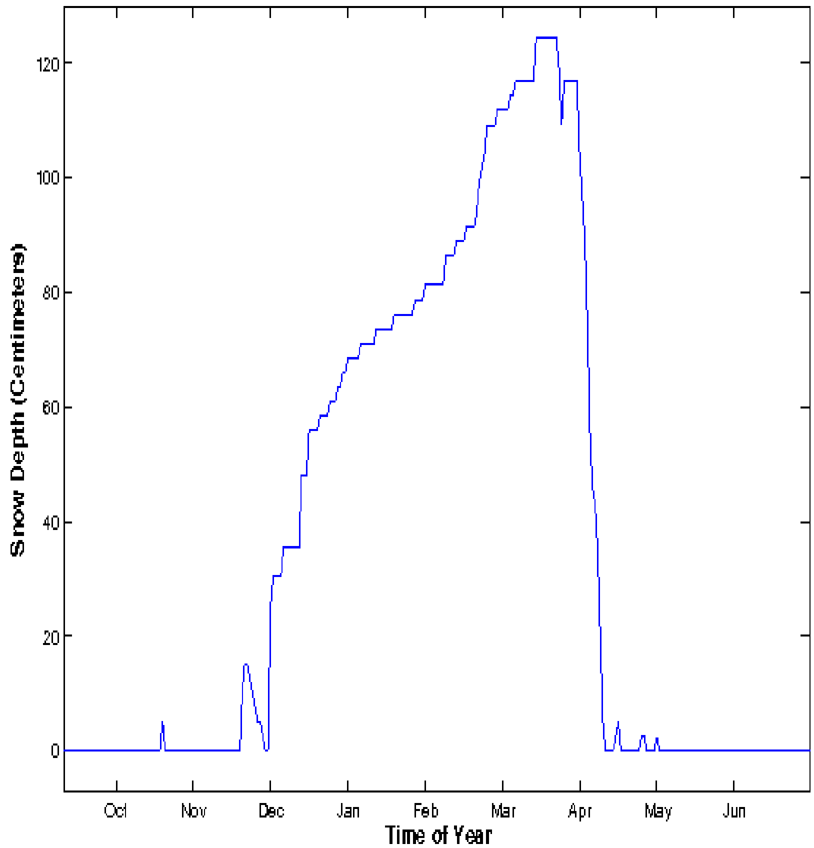

Figure 2.2: Daily Napoleon Snow Depths During the Winter of 1977-78.

2.3 A Discrete Time Periodic Storage Model

This section introduces our model of the snow depth process. Let X_t denote the snow depth at time t. Because the amount of snow on the ground is a stored quantity, we adopt a storage type model. Our model is based on the storage balance equation

Table 2.1: Napoleon Meta-data

Date	#	Time	Change	Observer	Note
8/19/1939		18:00	Site was already established	CJ Hoof	Station began 4/1/1889 1.3 mi SE of PO
6/20/1946		7:45	New observer, no move	Gladys Peterson	Station 1.5 mi SE of post office
7/8/1948			Station moved to another part of farm	Gladys Peterson	New thermometer support
11/11/1949		18:30	Moved 70 feet SE to improve exposure	Gladys Peterson	1.3 mi SE of post office
3/17/1954	1	18:00	Moved to 1.5 mi NW of Napoleon	Ted Frank	0.3 mi N of post office
4/18/1956	1A		New observer, no move	Alvin Schuchard	
2/19/1957	1B		New observer, no move	Warren Wentz	
5/8/1957	2	18:00	New observer, moved	Gladys Peterson	At ice cream store 3 blocks east of PO, 3.5 blocks SE of old location
7/1/1958	2A		Recording rain gauge removed		
8/28/1958	3	18:00	Equipment moved 0.6 mi W to observer's house	Gladys Peterson	Moved to 0.3 mi W of PO
9/30/1965	4	18:00	No move, update form	Gladys Peterson	
9/10/1968	4A		New observer, no move	Warren Wentz	
8/18/1969	5	7:00	Moved to Soo Depot, 0.5 mi E	Warren Wentz	Moved to more convenient location, station at 0.1 mi NE of PO
12/1/1973	5A		New observer, no move	Terry Wentz	No move
6/14/1976	6	7:00	Moved to mother's house 0.4 mi NE	Warren Wentz	Station at 0.5 mi NE
	6A		Address correction		
7/11/1985	6B	7:00	Moved across street to observer's house	Warren Wentz	MMTS installed
12/23/1987	7	7:00	No move, update form		
10/20/1992	8	8:00	Moved 0.1 mi SW to new residence	Bruce Wentz	Son of previous observer, now 0.4 mi NE of PO

$$X_t = \max\{X_{t-1} + Z_t, 0\}, \tag{3.1}$$

where Z_t is a random variable quantifying statistical changes in the pack occurring from time $t-1$ to t. View Z_t as the net change of snow (new minus meltoff) should the snow pack at time $t-1$ be so deep as to preclude ablation by day t. Because the data are observed daily (and not continuously) and snow depths cannot be negative, the maximum in (3.1) serves to prevent the pack content from becoming negative. We assume that $\{Z_t\}$ is white noise, independent of $\{X_t\}$, with periodic dynamics: $E[Z_t] = m_t$ and $\mathrm{Var}(Z_t) = w_t^2$. Periodic dynamics allow depth increases to be more likely in winter seasons. For convenience, we assume that Z_t is normally distributed. As the inferential objective here lies with trends, which are changes in the first moment, the normal assumption is not overly crucial. Should one be interested in extremes of the snow depths, then a marginal distribution with heavier tails could be used for $\{Z_t\}$.

For seasonal dynamics, we assume that $\{m_t\}$ and $\{w_t\}$ are periodic in time with period $T = 365$ days. Using a first-order Fourier expansion to describe the seasonal mean component, we write

$$m_t = P_t \left\{ A + B \cos\left(\frac{2\pi(t-\rho)}{T}\right) + \delta_t + \alpha t \right\}, \tag{3.2}$$

where ρ denotes the expected time of maximal daily increase of the pack and A and B are the mean and amplitude, respectively, of the sinusoidal expansion. The quantity α is the slope in a linear trend component and is the focal point of our future inferences. In (3.2), P_t is a deterministic indicator that is unity at times of the year when snowfall is possible and zero otherwise. This quantity stabilizes the ensuing numerical optimizations; specifically, it keeps the model optimization step and search routine from examining candidate models where large snow depths are possible during the summer. The quantity δ_t accounts for the breakpoints in the series; we assume the step form

$$\delta_t = \begin{cases} \Delta_1, & 1 \le t < \eta_1 \\ \Delta_2, & \eta_1 \le t < \eta_2 \\ \vdots & \vdots \\ \Delta_k, & \eta_{k-1} \le t \le N \end{cases},$$

where $\eta_1 < \eta_2 < \ldots < \eta_{k-1} \le N$ denote the ordered breakpoint times in the meta-data. Perusing the meta-data in Table 2.1, we consider 18 breakpoint times with $\eta_1 = 8/19/1939$, $\eta_2 = 6/20/1946, \ldots, \eta_{18} = 10/20/1992$. While undocumented breakpoint times (changepoints) may exist, changepoint methods are beyond the scope of this paper. The form of δ_t is a step function which depends on the regime at which the data was recorded. We take $\Delta_1 = 0$ to keep all model parameters identifiable (else, take $A = 0$ as a baseline). The interpretation is that regime ℓ is more snowy than regime $\ell - 1$ (assuming a zero trend) when $\Delta_\ell > \Delta_{\ell-1}$.

The model is easily modified to permit a seasonal trend component. For this, (3.2) is changed to

$$m_t = P_t \left\{ A + B \cos\left(\frac{2\pi(t-\rho)}{T} \right) + \delta_t + \left[C + D \cos\left(\frac{2\pi(t-\xi)}{T} \right) \right] t \right\}. \tag{3.3}$$

In (3.3), C is the average trend and D and ξ are the amplitude and phase parameters of the seasonal trend deviations about C.

Given these dynamics, $\{X_t\}$ is a discrete-time Markov chain on the state space $[0, \infty)$ with periodic transition probabilities. In continuous time, one can regard the process as a periodic diffusion with a boundary at state zero (diffusions are discussed in Cox and Miller, (1965)). The data can be viewed as observations of such a continuous time process taken on a discrete time lattice. Such notions are not overly important — we will simply need to be able to compute predictions of the next snow depth measurement from past observations. We comment that the same storage model may prove useful in describing streamflow series with periodic features that run dry. Here, precipitation is regarded as input into the store and stream discharge is regarded as the output.

It is tempting to try to develop more elaborate models for the snow depth process, for instance one in which Z_t is decomposed into a daily new snow accumulation minus a daily meltoff. However, as we will need to fit the model by conditional moment methods (this is discussed in the next section), parameter identifiability issues in such a decomposition would arise. Because of this, we do not pursue models that have separate components for new snow accumulations and meltoff.

2.4 Model Estimation

This section discusses parameter estimation in the model. While likelihood estimators generally have the most favorable sampling properties, the likelihood function of the model is essentially intractable (this is known from queueing theory, (Basawa and Rao (1980))) and alternative approaches need to be considered. We will adopt a conditional moment approach based on quasilikelihood and estimating equations. For clarity of exposition, we assume there is no missing data and ignore leap year effects.

Let $\boldsymbol{\theta} = (A, B, \rho, \alpha, \Delta_2, \ldots, \Delta_k)'$ be a vector containing all model parameters (add C, D, and ξ to this vector should a seasonal trend be considered). A simple sum of squares based on one-step-ahead prediction errors is $\sum_{t=1}^{N}(X_t - \hat{X}_t)^2$, where $\hat{X}_t = E[X_t|X_{t-1}, \ldots, X_1]$. Here, $E[\cdot]$ denotes expectation and $E[X|Y]$ indicates the conditional expectation of X given Y. Since $\{X_t\}$ is a periodic Markov chain, $\hat{X}_t = E[X_t|X_{t-1}]$. Because the snow depth process has periodic characteristics, it is preferable to use weighted least squares techniques in lieu of ordinary least squares. This entails scaling the prediction errors during each season by an estimate of its variability. In particular, the sum of squares function that we will minimize is

$$S(\boldsymbol{\theta}) = \sum_{t=1}^{N} \frac{(X_t - \hat{X}_t)^2}{\sigma_t^2} = \sum_{n=0}^{d-1} \sum_{\nu=1}^{T} \frac{(X_{nT+\nu} - \hat{X}_{nT+\nu})^2}{\sigma_\nu^2}, \tag{4.1}$$

where $\sigma_t^2 = E[(X_t - \hat{X}_t)^2 | X_{t-1}]$ and $d = N/T = 103$ is the number of years of data.

It is not important to be precise with the seasonal weights σ_ν^2. In fact, optimizing a version of $S(\boldsymbol{\theta})$ without any seasonal weights gives asymptotically consistent estimators of all non-breakpoint parameters; however, these estimators will not have a minimal variance ((Fuller (1996), Chapter 9) is a good reference discussing such issues). Here, we will use the simple sample variance weight

$$\sigma_\nu^2 = \frac{1}{d-1} \sum_{n=0}^{d-1} (X_{nT+\nu} - X_{nT+\nu-1})^2. \tag{4.2}$$

This is because the conditional variance $\mathrm{Var}(X_{nT+\nu} - \hat{X}_{nT+\nu} | X_{nT+\nu-1})$ is the optimal choice of σ_ν^2 and away from the state space boundary at zero (bare ground) and

$$X_{nT+\nu} - \hat{X}_{nT+\nu} \approx X_{nT+\nu} - X_{nT+\nu-1} - Z_{nT+\nu}.$$

Phrased another way, $w_\nu \approx \sigma_\nu$.

Figure 2.3 plots such empirical values of σ_ν^2 over all seasons ν. Also shown is a curve that smooths the empirical values; the smoothed versions were developed by fitting sinusoids to the Spring and Fall components and will be used in subsequent computations. Separate sinusoids were used before and after January 1. This is because snow depth changes are the most variable during early Spring when a full winter's pack has accumulated; depth changes during the height of winter appear to be less variable at Napoleon (likely due to consistent subfreezing midwinter temperatures). To aid numerical stability, we do not allow a weight to be less than unity. If a weight is very small, this season will contribute heavily to the sum of squares.

An explicit form is needed for the conditional mean in (3.1). Elementary Calculations yield

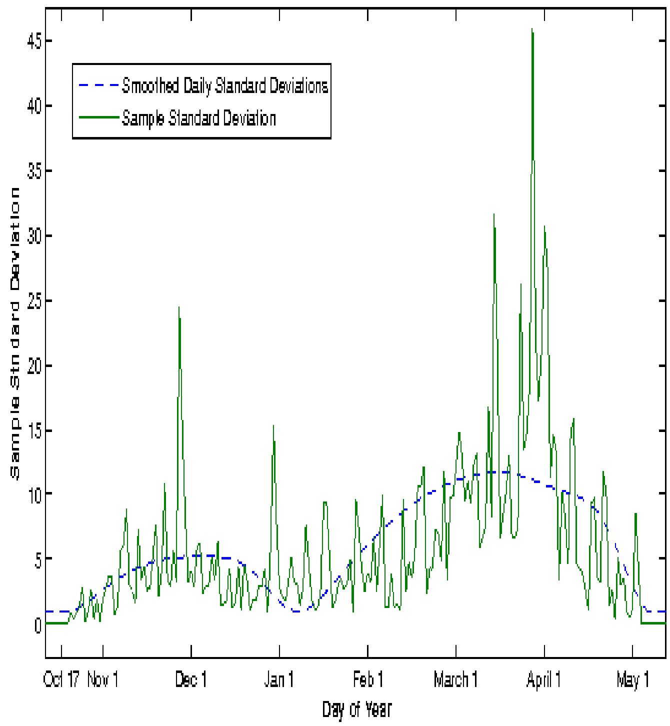

Figure 2.3: Estimated Values of σ_ν^2.

$$E[X_{t+1}|X_t] = \{X_t + m_{t+1}\}\left\{1 - \Phi\left(\frac{-X_t - m_{t+1}}{w_{t+1}}\right)\right\} + w_{t+1}\phi\left(\frac{X_t + m_{t+1}}{w_{t+1}}\right), \quad (4.3)$$

where $\Phi(x) = Pr[Z \leq x]$ is the cumulative distribution function of the standard normal random variable and $\phi(x) = \Phi'(x)$ is the standard normal density function.

Estimates of the parameters in $\boldsymbol{\theta}$ are found by numerically minimizing $S(\boldsymbol{\theta})$. Because of the seasonal weights and other nonlinearities of the objective function in the model parameters, this is necessarily a numerical task. Standard MATLAB packages have capably handled our problem. We use $w_\nu = \sigma_\nu$ in (4.3). One can consider strategies where w_ν is jointly optimized with m_ν, but this does not appear to be needed here.

Inferences can be made by invoking asymptotic normality principles for estimators derived from sum of squares criterion (see Klimko and Nelson, (1978)). In particular, under appropriate sampling schemes, it can be shown that the estimator $\boldsymbol{\theta}$ that minimizes $S(\boldsymbol{\theta})$ is consistent and asymptotically normal in that the distributional convergence

$$\hat{\boldsymbol{\theta}} \xrightarrow{\mathcal{D}} N(\boldsymbol{\theta}, F/d),$$

as $d \to \infty$ is achieved. Here, F is a positive definite covariance (information) matrix and N represents a normal distribution. Following the classical arguments of Klimko and Nelson (1978), F/d can be approximated by the inverse of the second derivative matrix of $S(\boldsymbol{\theta})$ evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$:

$$F/d \approx \left[\frac{\partial^2 S(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right]^{-1} \Bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}. \tag{4.4}$$

This relation allows standard errors for the model parameter estimates to be obtained; these standard errors are simply the square roots of the diagonal components of F/d. Such standard errors enable us to tune the model fit, eliminating any breakpoint mean shift parameters that do not induce significant changes. Eliminating insignificant model parameters allows one to improve accuracy margins of the trend estimator.

Two caveats need to be added to the above stated asymptotic normality. First, a proof of asymptotic normality needs all snow processes to continue infinitely far into the future. When $\alpha < 0$, there will be a last time where it snows and snow processes eventually cease. Of course, α will be small in most practical situations and the last time where it snows may occur so far into the future as to render the sampling distribution of $\hat{\alpha}$ close to Gaussian. Second, the sampling scheme needed to easily prove asymptotic normality assumes that the number of data points sampled in each regime (in between all breakpoint times) converges to infinity as $d \to \infty$. Such infill asymptotics seem unrealistic here; in fact, it seems more plausible that the station location will move infinitely often over an infinite time horizon. An implication of the latter sampling scheme is that the mean shift sizes $\Delta_2, \ldots, \Delta_k$ cannot be consistently estimated. This said, it is expected, akin to Lu and Lund (2007), that asymptotic normality of $\hat{\alpha}$ could still be proven as long as the station does not move too often. The mathematics to such arguments is intense and will not investigated here. The next section presents a simulation showing that the sampling distribution of the trend estimator is very close to Gaussian for our situation (approximately a century of data with five significant breakpoints). We also point out that asymptotic normality is a limiting property that is routinely applied to finite samples of reasonable size — it should provide reasonable guidance here.

2.5 Results

This section fits our storage model to the Napoleon data. We make two comments before proceeding. First, leap year effects will be ignored; in fact, the Napoleon data record lists 365 days for each and every year in the study (we are not sure how leap year effects were accounted for). Regardless, leap year effects should not change results appreciably. Second, a snow season is taken to run from October 17 to May 1 — so $P(t)$ is zero for times t sampled between May 2 and October 16. No appreciable snow was observed outside of these days. Third, missing data must be dealt with.

This issue is minimal with the Napoleon data as only about 1.5% of the record is missing during the snow season. If the datum point is missing at time t, we simply omit the terms from the summation in (4.1) at times t and $t+1$ (omitting the time $t+1$ observation is necessary since $\hat{X}_{t+1}$ also depends on X_t). Since most of the missing observations are not isolated, but rather occur in longer strings of weeks or months (often in summer months when $P(t)$ is zero), the total percentage of snow-season times missing from (4.1) is less than 3%. Figure 2.4 shows how the model fits the snow depths during the winter season of 1975-76. This graphic displays the snow depth values against their predictions; 95% confidence bands are included and are derived from (3.1) and normal $\{Z_t\}$. Indeed, the predictions appear to be tracking the data reasonably well. However, the model does not resolve individual days in that the estimated snowpack is too small on days when it snows heavily. The estimates "catch back up" with the snow depths on the first day thereafter on which heavy snow does not fall.

When all breakpoint times are ignored, the fitted trend estimate and one standard error is $\hat{\alpha} = 0.2250 \pm 0.0466$ cm per century. Because breakpoints induce mean uncertainties, their presence (or lack thereof) greatly impacts the estimated trend (Lu and Lund (2007) encounter the same issue in estimating temperature trends).

Our next goal is to investigate the effects of breakpoints on the trend estimate. Of the 18 breakpoint times listed in Table 2.1, two occur during the summer of 1958. Because snow depths are zero during summer, both of these breakpoints cannot be "identified" with our data and we proceed with 17 total breakpoints, one during the summer of 1958. Such a duplication issue occurs only once. With all 17 breakpoints in the model, the trend estimate and one standard error become $\hat{\alpha} = -0.5004 \pm 0.1927$ cm per century. Observe that the trend is now significantly negative.

Many of the breakpoint times may not be accompanied with significant mean

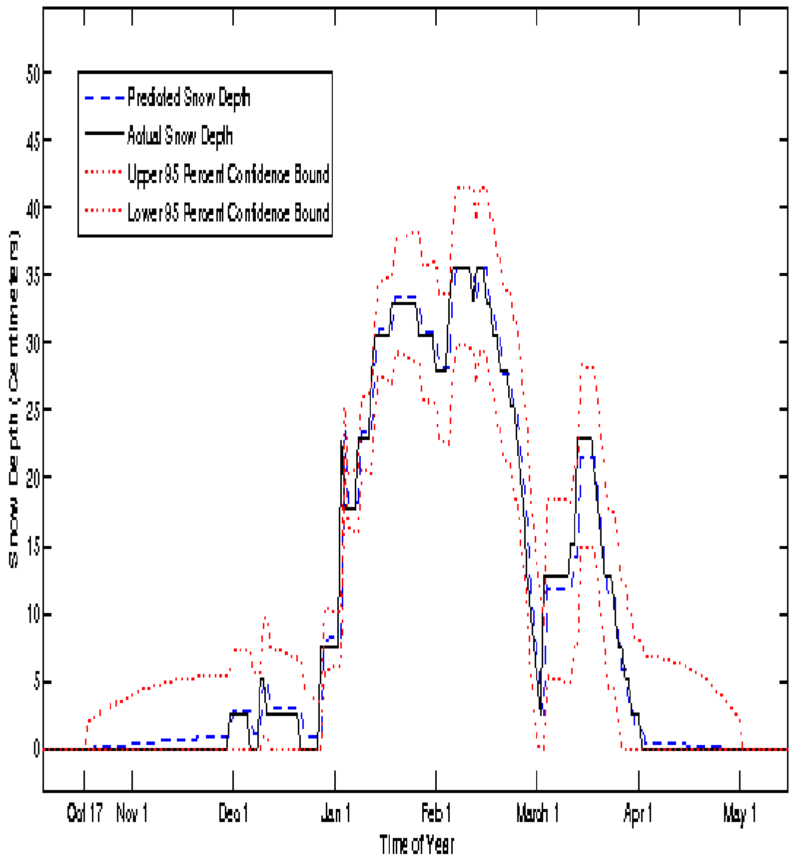

Figure 2.4: Snow Depths with One-Day-Ahead Predictions.

shifts; hence, our next task is to eliminate all insignificant breakpoint times in the metadata and recompute the trend estimate with only the significant breakpoints included. Such a procedure enables the trend and other model parameters to be more accurately estimated. The 17 remaining breakpoint times were subjected to a backwards regression elimination procedure at level 95% (see Anderson et al., (1994)). The magnitude of the level shift from regime i to regime $i+1$ is $\Delta_{i+1} - \Delta_i$. The standard error for its estimated difference is hence

$$\mathrm{Var}(\hat{\Delta}_{i+1} - \hat{\Delta}_i) = \mathrm{Var}(\hat{\Delta}_{i+1}) + \mathrm{Var}(\hat{\Delta}_i) - 2\mathrm{Cov}(\hat{\Delta}_{i+1}, \hat{\Delta}_i), \tag{5.1}$$

which is easily estimated from components of the second derivative matrix F/d. We then compute the Z-score

$$z_i = \frac{\hat{\Delta}_{i+1} - \hat{\Delta}_i}{\mathrm{Var}(\hat{\Delta}_{i+1} - \hat{\Delta}_i)^{1/2}}$$

for each breakpoint time i. At each iteration of the backwards regression, the breakpoint with the smallest $|z_i|$ is eliminated as long as $|z_i|$ is smaller than the 95th standard normal percentile (two-sided) of 1.96. If all $|z_i|$ exceed 1.96, the elimination procedure is stopped. After each eliminated breakpoint, the model is refitted and the parameter estimates and standard errors are recomputed. In the end, five significant breakpoint times are retained. We do not see any significant patterns in the type of retained breakpoint (two are observer changes, two are station relocations, etc). The trend estimate becomes -0.4748 ± 0.1803 cm per century. Notice that the standard error has decreased slightly from that for 17 breakpoints. The results of the backward elimination procedure are summarized in Table 2.2. This table shows estimates of the model parameters when 1) all breakpoints are ignored, 2) all seventeen of the breakpoint times are included, and 3) when the five significant mean shifts are accounted for. The times of the breakpoints were extracted from Table 2.1 and are noted in Table 2.2.

At the 95% confidence level, the trend estimate with all 17 breakpoints is concluded to be negative. In fact, a two-sided p-value of 0.0086 is obtained for the hypothesis test that snow depths are not changing. Hence, controlling for breakpoints,

Table 2.2: Summary of Model Parameter Estimates

Parameters	17 Breakpoints	5 Breakpoints	No Breakpoints
A	-2.5982 (0.0755)	-2.5951 (0.0742)	-2.7224 (0.0678)
B	2.9144 (0.0707)	2.9062 (0.0700)	2.9002 (0.0700)
ρ	3.8374 (0.6811)	3.7964 (0.6823)	3.8986 (0.6810)
α	-0.5004 (0.1927)	-0.4748 (0.1803)	0.2250 (0.0466)
Δ_1(8/19/1939)	0.2558 (0.0728)	0.2263 (0.0629)	-
Δ_2(6/20/1946)	0.1269 (0.1114)	-	-
Δ_3(7/8/1948)	0.4639 (0.1421)	-	-
Δ_4(11/11/1949)	0.2419 (0.0911)	-	-
Δ_5(3/17/1954)	0.1924 (0.1196)	-	-
Δ_6(4/18/1956)	-0.0360 (0.1949)	0.0074 (0.0960)	-
Δ_7(2/19/1957)	0.3707 (0.5387)	-	-
Δ_8(5/8/1957)	-0.2248 (0.1905)	-	-
Δ_9(7/1/1958)	-	-	-
Δ_{10}(8/28/1958)	0.0626 (0.0990)	-	-
Δ_{11}(9/30/1965)	-0.0381 (0.1327)	-	-
Δ_{12}(9/10/1968)	0.5633 (0.2352)	0.5626 (0.1118)	-
Δ_{13}(8/18/1969)	0.4975 (0.1225)	-	-
Δ_{14}(12/1/1973)	0.7380 (0.1559)	-	-
Δ_{15}(6/14/1976)	0.5899 (0.1285)	-	-
Δ_{16}(7/11/1985)	0.2021 (0.1685)	0.2835 (0.1442)	-
Δ_{17}(12/23/1987)	0.3923 (0.1685)	-	-
Δ_{18}(10/20/1992)	0.6332 (0.1565)	0.6123 (0.1473)	-

snow depths appear to be decreasing at Napoleon. The non-trend parameters all test as being significantly non-zero; hence, the model cannot be reduced further. Figure 2.6 graphically portrays the structure of the fitted models by plotting estimates of m_t against time t for no breakpoints, all seventeen breakpoints, and only the five significant breakpoints. It is instructive to compare the bottom two graphics in this panel as they depict which regimes have been eliminated. For instance, mean shifts for the very short regime in the late 1960s were deemed insignificant from the model fits and this regime was assimilated into a larger regime. The fitted model assigned the very snowy period during the late 1970s as part of a longer regime that had a significant upwards mean shift. Since inferences were based on asymptotic normality,

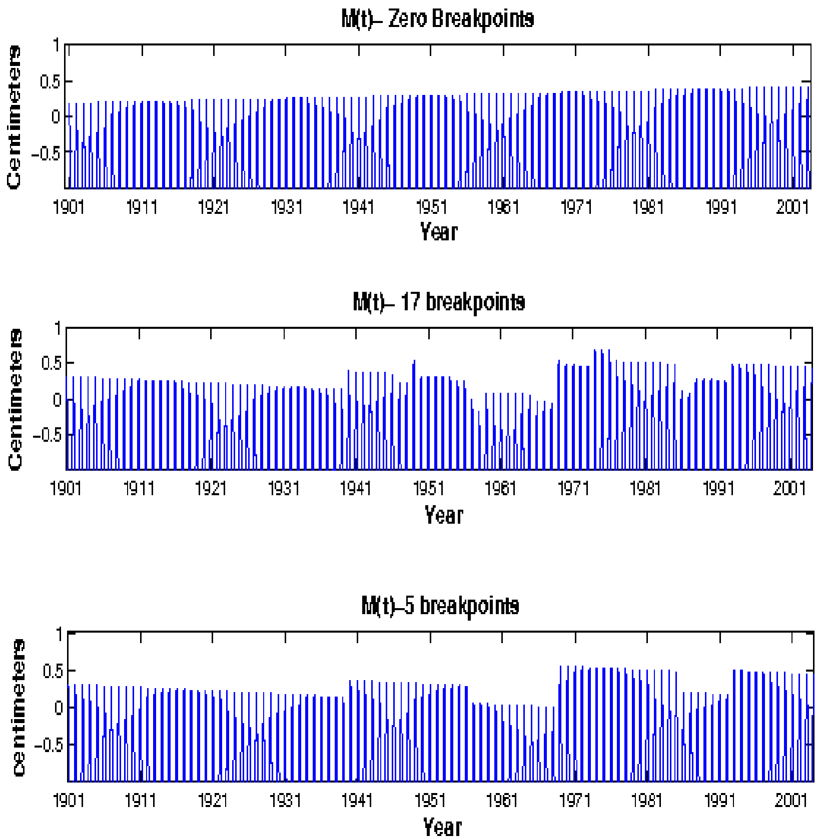

Figure 2.5: Structural Form of m_t.

a simulation was run to check asymptotic normality of the sampling distributions. For this, 1000 replicates of the fitted time series were generated. In each simulation run, 103 years of the daily snow depth process were generated. Each simulation run has the five breakpoint times shown in the third column of Table 2.2. The magnitude of the mean shift at each breakpoint time and simulation run was randomly generated from a uniform distribution taking values over $[-1, 1]$. The other parameters in (3.2) were taken as those fitted to the Napoleon series, except that the trend α was taken as zero in all replications. Figure 2.6 shows a kernel density smoothed histogram of the 1000 estimates of α (one from each simulation run). Observe that these estimates center around the true trend of zero and that their frequency distribution is unimodal and approximately normal. Hence, Gaussianity appears quite plausible for our sample sizes.

Next, the model with the periodic trend in (3.3) was fitted allowing for the five significant breakpoints identified above; an estimate and one standard error of the seasonal trend parameter D is $\hat{D} = -0.6121 \pm 0.2400$ cm per century, suggesting with a p-value of 0.0108 that trends are non-seasonal. The other parameter estimates in the model are $\hat{A} = -2.8687 \pm 0.1345$, $\hat{B} = 3.2190 \pm 0.1673$, $\hat{\rho} = 3.5016 \pm 1.0762$, $\hat{C} = 0.0329 \pm 0.2844$, and $\hat{\xi} = -0.0761 \pm 9.6588$. The five mean shift estimates do not change appreciably from those listed in Table 2.2.

To check on the seasonal fit, Figure 2.6 displays linear trend estimates for the depth observations for each day of year; for example, the January 1 trend is simply the trend slope, scaled to units of cm per century, of the 103 observations taken on January 1. These daily trend estimates, computed via Equation (3) in Lu and Lund (2007), account for the five significant breakpoint times. The cosine wave fitted for the seasonal trend is superimposed on the graphic in Figure 2.7 and matches the rough structure of the daily trend estimates. Specifically, mid to late Winter snow depths are decreasing most rapidly (ablation appears to be occurring earlier) and Spring snow depths are showing a slight increase. It is also noted that a single cosine wave does not seem to describe the seasonal trends well. Higher order Fourier fits, wavelet based expansions, or hinge-type structures such as those in Livezey et al. (2007) could be explored.

Before concluding, it is instructive to compare our methods to a naive trend analysis with seasonal snow tallies. Specifically, yearly snow totals were computed by adding the snow depth observations over all days during each snow season. As some seasons have too many missing data points to be considered reliable, seasons where more than 10% of the observations are missing were discarded in this comparison (i.e., are deemed as missing). Also, the 1901 Spring and 2003 Fall records were discarded since a complete record is not available for these winter seasons. In years where 10% or less of the snow season observations are missing, the seasonal total is made by summing all non-missing snow depths. These seasonal totals were then divided by the number of non-missing days during each snow season (which is 195 if no data are missing), giving an average daily snow depth for each non-missing season. A simple linear regression was fitted to the non-missing yearly average daily depths, yielding a slope of 0.426 cm per century when all breakpoints are ignored and -1.548 cm per century when the five significant breakpoint are taken into account. Observe that the sign of these estimates agree with the ones fitted in our above computations.

Some model validation diagnostics were performed. First, parameters for the seasonal model were estimated from the first 93 years of data only. This model does not contain a seasonal trend and only includes the five breakpoints deemed significant in the above analyses. A simulation was then conducted to estimate the time-varying mean (the unconditional mean) of this fitted model. This estimated mean is plotted against the last ten years of snow depths in Figure 2.8. Except for 1997, the fitted model seems to describe this 10 years of data well. Notice that the fitted model has a slightly negative trend estimate. Because a breakpoint occurred in October of 1992, the mean shift used for this breakpoint was taken as that estimated from all 103 years of data. Second, a set of residuals was computed for the daily model with a periodic trend and five breakpoints. These residuals, which are simply the

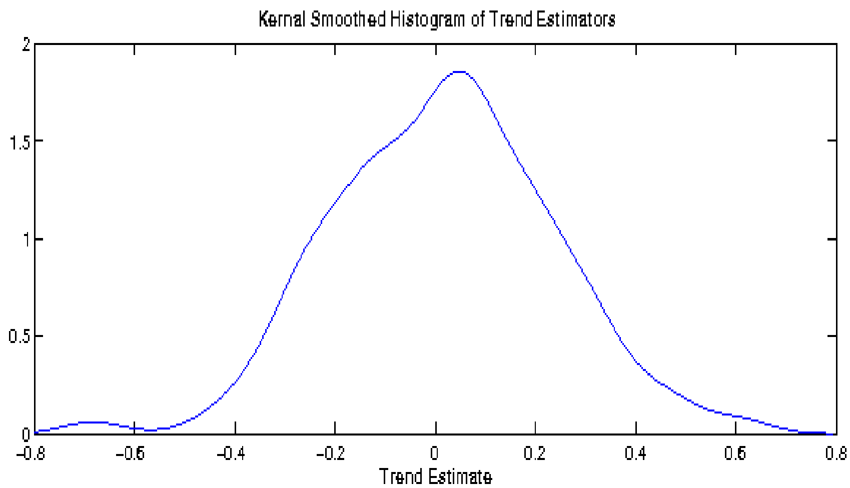

Figure 2.6: Kernel Smoothed Histogram of Estimated Trends Showing Approximate Normality.

difference between the observations and the one-day-ahead predictions scaled by a daily standard deviation, are analyzed in Figure 2.9. The residuals appear to be symmetric about a zero mean (as seen in the center plot). The standard normal quantile plot in the top graphic shows that the residuals have some non-Gaussian features, but Gaussianity is not a requirement of residuals from a time series analysis.

No one-day-ahead residual exceeds 10 in magnitude, nor as the bottom plot shows, does there appear to be significant lag one autocorrelation. This said, the figure suggests that the model could be improved. Elaborating, large snowstorms frequently induce large positive residuals, especially when they occur at times when snow is typically not on the ground. Years where missing data is prevalent during mid-winter (such as 1991) have very small residuals. Residuals during years with minimal snow tend to be small and negative. While incorporating a shot-type component into the model (i.e., a component that allows for rapid inputs of large amounts of snow) might remedy some of these aspects, this would be a difficult extension of this work.

2.6 Comments

As in temperature trend analyses, breakpoints appear to be the most critical aspect to account for in estimating trends. With the Napoleon data, the sign of the trend estimate changes when breakpoints are ignored. In temperature studies, the effects of breakpoints are frequently illuminated by making reference series comparisons (Menne and Williams Jr., (2005)). Such a tactic is not possible here as reference series are not readily available — it is questionable if any suitable reference station exists for Napoleon over its entire record.

As snow depths can vary considerably over short geographical distances, reference station comparisons for snow data may also be more untrustworthy than those for temperature data. Also, the trend at any one station should be loosely interpreted. Ideally, trend estimates at many stations would be computed and spatially aggregated to make firm conclusions about changes in a geographic area. Such an endeavor requires a spatial analysis of trends from many stations, each of which could be computed as in this article. Modeling improvements might be possible if an accom-

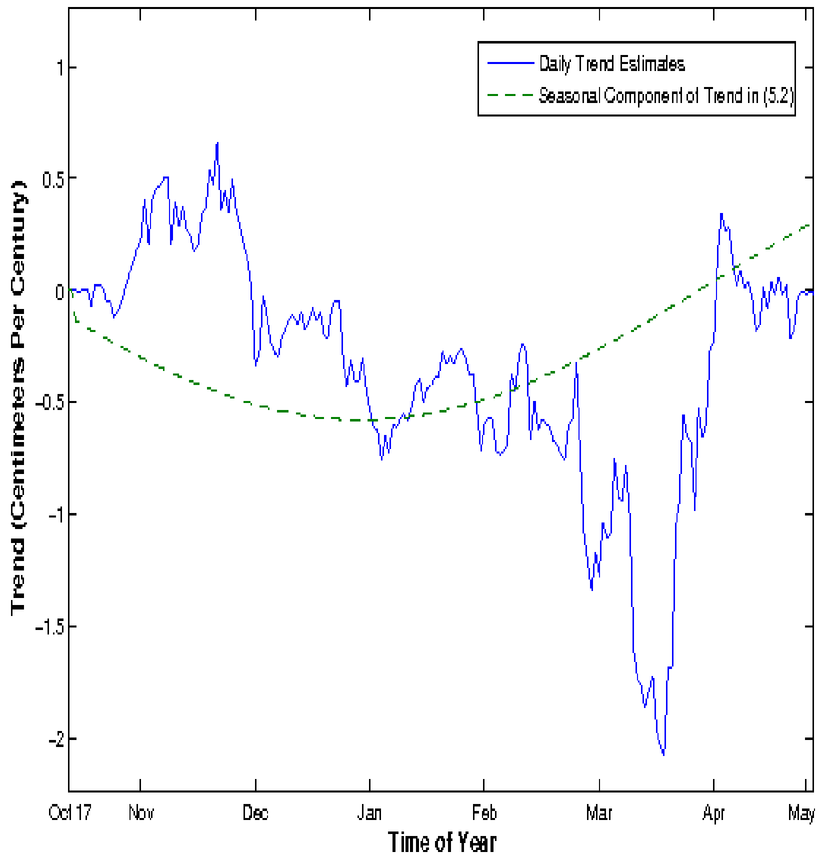

Figure 2.7: Daily Trend Estimates with Superimposed Fourier Fit.

panying temperature record (or other covariates) were available. Finally, true snow depth changes are likely to be non-linear in time; of course, linear trends, regardless of the true trend structure, can always be interpreted as an average rate of change over the record.

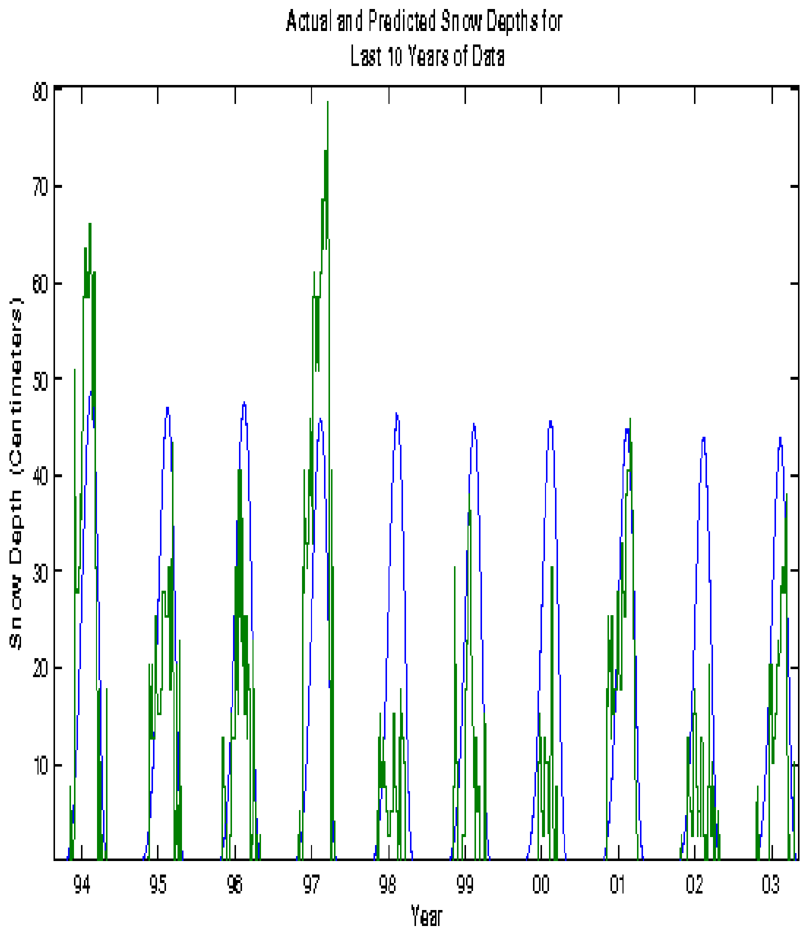

Figure 2.8: Predictions of Last 10 Years of Data Against Observed Values.

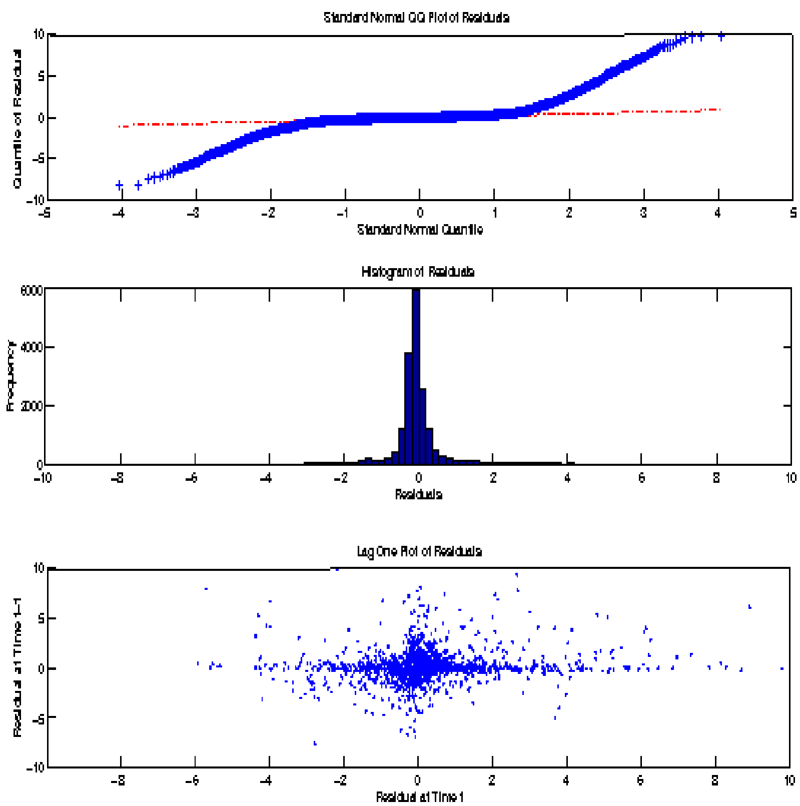

Figure 2.9: Diagnostic Plots of Model Residuals.

References

Anderson, D.R., D.J. Sweeney, and T.A. Williams (1994), *Introduction to Statistics. Concepts and Applications*, Third Edition, West Publishing Company, St. Paul, MN.

Barnett, T.P., J.C. Adam, and P. Lettenmaier (2005), Potential impacts of a warming climate on water availability in snow-dominated regions, *Nature*, 438, 303-309.

Barry, R.G. (1990), Evidence of recent changes in global snow and ice cover, *GeoJournal*, 20, 121-127.

Basawa, I.V., and B.L.S. Prakasa Rao (1980), *Statistical Inference for Stochastic Processes*, Academic Press, New York City.

Blöschl, G. (1999), Scaling issues in snow hydrology, *Hydrological Processes*, 13, 2149-2175.

Boer, G.J., N. McFarlane, and M. Lazare (1992), Greenhouse gas-induced climate change simulated with the CCC second-generation general circulation model, *Journal of Climate*, 5, 1045-1077.

Brown, R.D., M. Hughes, and D. Robinson (1995), Characterizing the long term variability of snow cover extent over the interior of North America, *Annals of Glaciology*, 21, 45-50.

Burn, D.H. (1994), Hydrologic effects of climatic change in West-Central Canada, *Journal of Hydrology*, 160, 53-70.

Cox, D.A., and H.D. Miller (1965), *The Theory of Stochastic Processes*, Chapman and Hall, London.

Fuller, W.A. (1996), *Introduction to Statistical Time Series*, Second Edition, John Wiley and Sons, New York.

Grundstein, A., P. Todhunter, and T.L. Mote (2005), Snowpack control over the thermal offset of air and soil temperatures in eastern North Dakota, *Geophysical Research Letters*, 32, L08503, doi: 10.1029/2005GL022532.

Hamlet, A.F., P.W. Mote, M.P. Clark, and D.P. Lettenmaier (2005), Effects of temperature and precipitation variability on snowpack trends in the Western United States, *Journal of Climate*, 18, 4545-4561.

Klimko, L.A, and P.I. Nelson (1978), On conditional least squares estimation for stochastic processes, *Annals of Statistics*, 6, 629-642.

Kukla, G.J. (1979), Climatic role of snow covers. In: *Sea Level, Ice and Climatic Change*, IAHS Publication 131, 79-107.

Lu, Q., and R.B. Lund (2007), Simple linear regression with multiple changepoints, *Canadian Journal of Statistics*, 37, 447-458.

Marsh, P. (1999), Snowcover formation and melt: recent advances and future prospects, *Hydrological Processes*, 13, 2117-2134.

Mote, P.W., A.F. Hamlet, M.P. Clark, and D.P. Lettenmaier (2005), Declining mountain snowpack in Western North America, *Bulletin of the American Meteorological Society*, 86, 39-49.

Menne, M.J., and C.J. Williams Jr. (2005), Detection of undocumented changepoints using multiple test statistics and composite reference series, *Journal of Climate*, 18, 4271-4286.

Myeni., R.B., C.D. Keeling, C.J. Tucker, G. Asrar, and R.R. Nemani (1997), Increased plant growth in the northern high latitudes from 1981 to 1991, *Nature*, 386, 698-702.

Perona, P.A., A. Porporato, and L. Ridolfi (2007), A stochastic process for the inter-annual snow storage and melting dynamics, *Journal of Geophysical Research, Atmo-*

spheres, 112 D08107, doi:10.1029/2006JD007798.

Robinson, D.A. (1993), Historical daily climatic data for the United States, In: *Proceedings of the Eighth Conference on Applied Climatology*, Anaheim, CA, 264-269.

Schmidt, W.L., W.D. Gosnold, and J.W. Enz (2001), A decade of air-ground temperature exchange from Fargo, ND, *Global and Planetary Change*, 29, 311-325.

Chapter 3

General Linear Models with Infinitely Many Level Shifts

3.1 Introduction

Consider a general linear regression model (GLM) for a time series $\{X_t\}$ sampled from a stochastic process. We write the model in the form

$$X_t = \mu + \theta_1 f_1(t) + \cdots + \theta_m f_m(t) + \epsilon_t, \qquad 1 \leq t \leq n, \tag{1.1}$$

where $\theta_1, \ldots, \theta_m$ are unknown parameters, $f_1(t), \ldots, f_m(t)$ are regression factors, μ is the intercept, and ϵ_t is the error at time t. Thorough treatments of the GLM are provided in Graybill (1976), Bickel and Doksum (2001), etc. The case of correlated $\{\epsilon_t\}$ is considered in detail by Brockwell and Davis (1991) and Fuller (1996). Applications of the GLM are ubiquitous in modern scientific studies.

Often, level shifts that occur in the underlying stochastic process are unaccounted for in the modeling procedure. For example, it is known that climatological data often contain level shifts due to changing location of a measuring station, its instrumentation, or even the observer taking the data (see Lu and Lund (2007), Solow (1987), Easterling and Peterson (1994), Menne and Williams Jr. (2008)). Climatological meta-data often list changes that can induce level shifts in a process.

The problem that we tackle here involves inference of $\boldsymbol{\Theta} = (\theta_1, \dots, \theta_m)$ when the process is experiencing probabilistically regular level shifts. Specifically, we consider the case where the number of level shifts converges to infinity as $n \to \infty$. As United States temperature series experience six level shifts per century on average (see Mitchell (1953)), it is felt that this scenario is more realistic than classical infill asymptotics where the number of data points taken between all changepoints converges to infinity as the sample size converges to infinity.

3.2 The Model

A version of the model in (1.1) that allows for infinitely many level shifts in the process is

$$X_t = \theta_1 f_1(t) + \cdots + \theta_m f_m(t) + \delta_t + \epsilon_t. \tag{2.1}$$

Here, $\{\delta_t\}$ represents the level shift factor. For a fixed sample size n and $1 \leq t \leq n$, the level shift at time t is

$$\delta_t = \begin{cases} \Delta_1, & 1 \leq t < \tau_1 \\ \Delta_2, & \tau_1 \leq t < \tau_2 \\ \vdots & \vdots \\ \Delta_k, & \tau_{k-1} \leq t \leq \tau_n \end{cases},$$

where $k = k(n)$ is the number of changepoints up to time n and $\tau_0, \tau_1, ..., \tau_{k-1}$ are the ordered level shift times, with $\tau_0 = 1$ being a convention. We consider the times of all level shift changes to be known. Hence, $\tau_0, ..., \tau_{k-1}$ partition $\{1, \ldots, n\}$ into $k(n)$ regimes. We use r to index the r^{th} regime of the data, $Y_r = \{\tau_{r-1}, \ldots, \tau_r - 1\}$ as the set of times when the series is in regime r, and $l_r = \tau_r - \tau_{r-1}$ to be the length of the r^{th} regime. The model above becomes

$$X_t = \theta_1 f_1(t) + \cdots + \theta_m f_m(t) + \delta_{r(t)} + \epsilon_t, \tag{2.2}$$

where $r(t)$ is the regime the series was experiencing at time t. The errors $\{\epsilon_t\}$ are assumed to be a zero-mean stationary process with lag h autocovariance $\gamma(h) = \operatorname{cov}(\epsilon_t, \epsilon_{t+h})$. The m deterministic factors $f_1(t), \ldots, f_m(t)$ are functions in time, known for $t = 1, 2, \ldots, n$, with additional structure to be imposed below. Since we allow $k(n) \to \infty$ as $n \to \infty$, the model accommodates an infinite number of level shifts in the limit.

Our task is to estimate $\theta_1, \ldots, \theta_m$ consistently, and if possible, construct estimates of these parameters that are asymptotically normal. One must realize that the mean shift parameters $\Delta_1, \Delta_2, \ldots$ cannot be estimated consistently as the only information for Δ_r arises from the data from regime r, which may be a finite set.

3.3 Derivation of the Estimators

Our first task is to derive estimators of $\boldsymbol{\Theta} = (\theta_1, \ldots, \theta_m)^T$. We proceed by deriving the ordinary least squares estimator of $\boldsymbol{\Theta}$. Let

$$S(\boldsymbol{\Theta}) \;=\; \sum_{t=1}^{n} (X_t - E[X_t])^2 \tag{3.1}$$

$$= \sum_{r=1}^{k(n)} \sum_{t \in Y_r} (X_t - (\theta_1 f_1(t) + \cdots + \theta_m f_m(t) + \Delta_r))^2 \tag{3.2}$$

be the sum of squares when the regression parameters are $\theta_1, \ldots, \theta_m$.

Taking the partial derivative of (3.2) with respect to Δ_r and setting the result to zero gives

$$-2 \sum_{t \in Y_r} (X_t - (\theta_1 f_1(t) + \cdots + \theta_m f_m(t) + \Delta_r)) = 0.$$

Hence,

$$l_r \Delta_r = \sum_{t \in Y_r} \left(X_t - \left(\sum_{u=1}^{m} \theta_m f_u(t) \right) \right).$$

This, in turn, gives

$$\bar{X}_r = \left(\sum_{t \in Y_r} \sum_{u=1}^{m} \theta_u \bar{f}_{u,r} \right) + \Delta_r, \tag{3.3}$$

where

$$\bar{f}_{u,r} = \frac{1}{l_r} \sum_{t \in Y_r} f_u(t) \tag{3.4}$$

is the average of the u^{th} factor over regime r.

Next, taking partials of (3.2) with respect to $\theta_1, \ldots, \theta_m$ and making use of (3.3), we obtain

$$\begin{aligned}
\left(\frac{-1}{2} \right) \frac{\partial S(\Theta)}{\partial \theta_v} &= \sum_{r=1}^{k(n)} \sum_{t \in Y_r} \left(X_t - \left[\left(\sum_{u=1}^{m} \theta_u f_u(t) \right) + \Delta_r \right] \right) f_v(t) \\
&= \sum_{r=1}^{k(n)} \sum_{t \in Y_r} \left(X_t - \left[\left(\sum_{u=1}^{m} \theta_u \bar{f}_{u,r} \right) + \Delta_r \right] \right) f_v(t) \\
&+ \sum_{r=1}^{k(n)} \sum_{t \in Y_r} \left(\sum_{u=1}^{m} \theta_u \left[f_u(t) - \bar{f}_{u,r} \right] \right) f_v(t)
\end{aligned}$$

$$= \sum_{r=1}^{k(n)} \sum_{t \in Y_r} \left(X_t - \bar{X}_r\right) f_v(t)$$

$$- \sum_{r=1}^{k(n)} \sum_{t \in Y_r} \left(\sum_{u=1}^{m} \theta_u \left[f_u(t) - \bar{f}_{u,r}\right]\right) f_v(t).$$

Equating these derivatives to zero gives

$$\begin{aligned}
\sum_{r=1}^{k(n)} \sum_{t \in Y_r} \theta_v \left(f_v(t) - \bar{f}_{v,r}\right) f_v(t) &= \sum_{r=1}^{k(n)} \sum_{t \in Y_r} \left(X_t - \bar{X}_r\right) f_v(t) \\
&- \sum_{r=1}^{k(n)} \sum_{t \in Y_r} \left(f_v(t) \sum_{u=1;u\neq v}^{m} \theta_u f_u(t)\right) \\
&+ \sum_{r=1}^{k(n)} \sum_{t \in Y_r} \left(f_v(t) \sum_{u=1;u\neq v}^{m} \theta_u \bar{f}_{u,r}\right).
\end{aligned}$$

Rewriting this gives, for $1 \leq v \leq m$,

$$\begin{aligned}
\theta_v &= \frac{\sum_{r=1}^{k(n)} \sum_{t \in Y_r} \left(X_t - \bar{X}_r\right)\left(f_v(t) - \bar{f}_{v,r}\right)}{\sum_{r=1}^{k(n)} \sum_{t \in Y_r} \left(f_v(t) - \bar{f}_{v,r}\right)^2} \\
&- \frac{\sum_{r=1}^{k(n)} \sum_{t \in Y_r} \left(\sum_{u=1;u\neq v}^{m} \theta_u \left[f_u(t) - \bar{f}_{u,r}\right]\right)\left(f_v(t) - \bar{f}_{v,r}\right)}{\sum_{r=1}^{k(n)} \sum_{t \in Y_r} \left(f_v(t) - \bar{f}_{v,r}\right)^2} \\
&= \frac{\sum_{r=1}^{k(n)} \sum_{t \in Y_r} \left(X_t - \bar{X}_r\right)\left(f_v(t) - \bar{f}_{v,r}\right)}{\sum_{r=1}^{k(n)} \sum_{t \in Y_r} \left(f_v(t) - \bar{f}_{v,r}\right)^2} \\
&- \sum_{u=1;u\neq v}^{m} \theta_u \left\{\frac{\sum_{r=1}^{k(n)} \sum_{t \in Y_r} \left(f_u(t) - \bar{f}_{u,r}\right)\left(f_v(t) - \bar{f}_{v,r}\right)}{\sum_{r=1}^{k(n)} \sum_{t \in Y_r} \left(f_v(t) - \bar{f}_{v,r}\right)^2}\right\}.
\end{aligned}$$

Writing these in a linear system gives

$$\theta_v + \sum_{u=1;u\neq v}^{m} \theta_u \left\{ \frac{\sum_{r=1}^{k(n)} \sum_{t\in Y_r} \left(f_u(t) - \bar{f}_{u,r}\right)\left(f_v(t) - \bar{f}_{v,r}\right)}{\sum_{r=1}^{k(n)} \sum_{t\in Y_r} \left(f_v(t) - \bar{f}_{v,r}\right)^2} \right\} =$$

$$\frac{\sum_{r=1}^{k(n)} \sum_{t\in Y_r} \left(X_t - \bar{X}_r\right)\left(f_v(t) - \bar{f}_{v,r}\right)}{\sum_{r=1}^{k(n)} \sum_{t\in Y_r} \left(f_v(t) - \bar{f}_{v,r}\right)^2}.$$

This gives the $m \times m$ linear system

$$\mathbf{B}\hat{\mathbf{\Theta}} = \mathbf{Z}, \tag{3.5}$$

where the $(i,j)^{th}$ element of $\mathbf{B}$ is

$$B_{i,j} = \frac{\sum_{r=1}^{k(n)} \sum_{t\in Y_r} \left(f_i(t) - \bar{f}_{i,r}\right)\left(f_j(t) - \bar{f}_{j,r}\right)}{\sum_{r=1}^{k(n)} \sum_{t\in Y_r} \left(f_i(t) - \bar{f}_{i,r}\right)^2} \tag{3.6}$$

for $1 \leq i,j \leq m$ and

$$Z_i = \frac{\sum_{r=1}^{k(n)} \sum_{t\in Y_r} \left(X_t - \bar{X}_r\right)\left(f_i(t) - \bar{f}_{i,r}\right)}{\sum_{r=1}^{k(n)} \sum_{t\in Y_r} \left(f_i(t) - \bar{f}_{i,r}\right)^2} \tag{3.7}$$

for $1 \leq i \leq m$. Here, $\mathbf{Z} = (Z_1, \ldots, Z_m)^T$

Since the notation for the above system is cumbersome, we introduce some shorthand notation. Define

$$\eta_{i,t} = f_i(t) - \bar{f}_{i,r(t)}, \quad 1 \leq t \leq n, \quad 1 \leq i \leq m. \tag{3.8}$$

For convenience, we often suppress the t in the ensuing analysis. Observe that $\boldsymbol{\eta}_i = \{\eta_i\}_{t=1}^{n}$ is an $n \times 1$ vector. Define the inner product

$$< \boldsymbol{\eta}_i, \boldsymbol{\eta}_j >= \sum_{r=1}^{k(n)} \sum_{t \in Y_r} \left(f_i(t) - \bar{f}_{i,r} \right) \left(f_j(t) - \bar{f}_{j,r} \right), \quad 1 \leq t \leq n, \quad 1 \leq i \leq m. \tag{3.9}$$

Defining $\mathbf{X} = (X_1, \ldots, X_n)^T$, the components in (3.5) are seen to be

$$\mathbf{B} = \begin{pmatrix} \frac{<\boldsymbol{\eta}_1, \boldsymbol{\eta}_1>}{<\boldsymbol{\eta}_1, \boldsymbol{\eta}_1>} & \ldots & \frac{<\boldsymbol{\eta}_1, \boldsymbol{\eta}_m>}{<\boldsymbol{\eta}_1, \boldsymbol{\eta}_1>} \\ \vdots & \ddots & \vdots \\ \frac{<\boldsymbol{\eta}_m, \boldsymbol{\eta}_1>}{<\boldsymbol{\eta}_m, \boldsymbol{\eta}_m>} & \ldots & \frac{<\boldsymbol{\eta}_m, \boldsymbol{\eta}_m>}{<\boldsymbol{\eta}_m, \boldsymbol{\eta}_m>} \end{pmatrix} \tag{3.10}$$

$$\mathbf{Z} = \begin{pmatrix} \frac{<\mathbf{X}, \boldsymbol{\eta}_1>}{<\boldsymbol{\eta}_1, \boldsymbol{\eta}_1>} \\ \vdots \\ \frac{<\mathbf{X}, \boldsymbol{\eta}_m>}{<\boldsymbol{\eta}_m, \boldsymbol{\eta}_m>} \end{pmatrix}. \tag{3.11}$$

Hence, we have finally identified the least squares parameter estimators as

$$\hat{\boldsymbol{\Theta}} = \begin{pmatrix} \hat{\theta}_1 \\ \vdots \\ \hat{\theta}_m \end{pmatrix} = \mathbf{B}^{-1}\mathbf{Z}. \tag{3.12}$$

3.4 Consistency of the Estimators

This section investigates the consistency of $\hat{\boldsymbol{\Theta}}$. If the number of level shifts is finite, one can construct a consistent estimate of the regression parameters $\theta_1, \ldots, \theta_m$ from data taken from the last regime only. Therefore, we focus on the case where the number of level shifts, $k(n)$, goes to infinity as $n \to \infty$, while noting the techniques below also work when there are finite or even no level shifts.

Taking variances in (3.12) gives $\text{Var}\left(\hat{\boldsymbol{\Theta}}\right) = \mathbf{B}^{-1}\text{Var}\left(\mathbf{Z}\right)\mathbf{B}^{-1}$. However, the computations can be simplified if we transform the problem as follows. Let

$$\mathbf{E} = \begin{pmatrix} \langle \boldsymbol{\eta}_1, \boldsymbol{\eta}_1 \rangle & 0 & \dots & 0 \\ 0 & \langle \boldsymbol{\eta}_2, \boldsymbol{\eta}_2 \rangle & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & \langle \boldsymbol{\eta}_m, \boldsymbol{\eta}_m \rangle \end{pmatrix} \tag{4.1}$$

and left multiply both sides of (3.5) by $\mathbf{E}$ to get

$$\mathbf{G}\hat{\boldsymbol{\Theta}} = \mathbf{E}\mathbf{Z}, \tag{4.2}$$

with

$$\mathbf{G} = \begin{pmatrix} \langle \boldsymbol{\eta}_1, \boldsymbol{\eta}_1 \rangle & \langle \boldsymbol{\eta}_1, \boldsymbol{\eta}_2 \rangle & \dots & \langle \boldsymbol{\eta}_1, \boldsymbol{\eta}_m \rangle \\ \langle \boldsymbol{\eta}_2, \boldsymbol{\eta}_1 \rangle & \langle \boldsymbol{\eta}_2, \boldsymbol{\eta}_2 \rangle & \dots & \langle \boldsymbol{\eta}_2, \boldsymbol{\eta}_m \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \boldsymbol{\eta}_m, \boldsymbol{\eta}_1 \rangle & \langle \boldsymbol{\eta}_m, \boldsymbol{\eta}_2 \rangle & \dots & \langle \boldsymbol{\eta}_m, \boldsymbol{\eta}_m \rangle \end{pmatrix}. \tag{4.3}$$

The matrix $\mathbf{G}$ is of a special form and is known as the Gram matrix. The determinant of the Gram matrix is called the Gramian. The Gramian of linearly independent vectors is positive, and zero for linearly dependent vectors. Thus, the factors of (2.2) should be chosen so that $\{\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_m\}$ are a set of m linearly independent $n \times 1$ vectors; otherwise, (4.2) cannot be solved uniquely and multiple solutions to the least squares equations exist (See Gantmacher (1977)).

Observing that $\mathbf{G}$ is symmetric, we see that

$$\text{Var}\left(\hat{\boldsymbol{\Theta}}\right) = \mathbf{G}^{-1}\text{Var}\left(\mathbf{E}\mathbf{Z}\right)\mathbf{G}^{-1}. \tag{4.4}$$

Let us turn our attention to $\text{Var}\left(\mathbf{E}\mathbf{Z}\right)$. For the error sequence $\{\epsilon_t\}_{t=1}^{n}$, define the $n \times 1$ error vector $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)^T$. Then

$$\text{Var}(\mathbf{EZ}) = \begin{pmatrix} \text{Cov}(\langle \boldsymbol{\epsilon}, \boldsymbol{\eta}_1 \rangle, \langle \boldsymbol{\epsilon}, \boldsymbol{\eta}_1 \rangle) & \ldots & \text{Cov}(\langle \boldsymbol{\epsilon}, \boldsymbol{\eta}_1 \rangle, \langle \boldsymbol{\epsilon}, \boldsymbol{\eta}_m \rangle) \\ \vdots & \ddots & \vdots \\ \text{Cov}(\langle \boldsymbol{\epsilon}, \boldsymbol{\eta}_m \rangle, \langle \boldsymbol{\epsilon}, \boldsymbol{\eta}_m \rangle) & \ldots & \text{Cov}(\langle \boldsymbol{\epsilon}, \boldsymbol{\eta}_m \rangle, \langle \boldsymbol{\epsilon}, \boldsymbol{\eta}_m \rangle) \end{pmatrix}. \tag{4.5}$$

Since $\{\epsilon_t\}_{t=1}^{\infty}$ is a zero mean process,

$$\begin{aligned} \text{Cov}\Big(< \boldsymbol{\epsilon}, \boldsymbol{\eta}_i >, < \boldsymbol{\epsilon}, \boldsymbol{\eta}_j >\Big) &= \text{E}\Big[< \boldsymbol{\epsilon}, \boldsymbol{\eta}_i >< \boldsymbol{\epsilon}, \boldsymbol{\eta}_j >\Big] \\ &= \gamma(0) \sum_{t=1}^{n} \eta_{i,t}\eta_{j,t} \\ &+ \sum_{h=1}^{n-1} \gamma(h) \left(\sum_{t=1}^{n-h} \eta_{i,t}\eta_{j,t+h} + \eta_{i,t+h}\eta_{j,t} \right). \end{aligned}$$

Before proceeding, we define an "angle" matrix $\mathbf{C}$ via

$$\mathbf{C} = \mathbf{E}^{-\frac{1}{2}} \mathbf{G} \mathbf{E}^{-\frac{1}{2}}. \tag{4.6}$$

We call $\mathbf{C}$ an angle matrix due to its representation as

$$\mathbf{C} = \begin{pmatrix} \frac{<\boldsymbol{\eta}_1, \boldsymbol{\eta}_1>}{\|\boldsymbol{\eta}_1\| \|\boldsymbol{\eta}_1\|} & \frac{<\boldsymbol{\eta}_1, \boldsymbol{\eta}_2>}{\|\boldsymbol{\eta}_1\| \|\boldsymbol{\eta}_2\|} & \cdots & \frac{<\boldsymbol{\eta}_1, \boldsymbol{\eta}_m>}{\|\boldsymbol{\eta}_1\| \|\boldsymbol{\eta}_m\|} \\ \frac{<\boldsymbol{\eta}_2, \boldsymbol{\eta}_1>}{\|\boldsymbol{\eta}_2\| \|\boldsymbol{\eta}_1\|} & \frac{<\boldsymbol{\eta}_2, \boldsymbol{\eta}_2>}{\|\boldsymbol{\eta}_2\| \|\boldsymbol{\eta}_2\|} & \cdots & \frac{<\boldsymbol{\eta}_2, \boldsymbol{\eta}_m>}{\|\boldsymbol{\eta}_2\| \|\boldsymbol{\eta}_m\|} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{<\boldsymbol{\eta}_m, \boldsymbol{\eta}_1>}{\|\boldsymbol{\eta}_m\| \|\boldsymbol{\eta}_1\|} & \frac{<\boldsymbol{\eta}_m, \boldsymbol{\eta}_2>}{\|\boldsymbol{\eta}_m\| \|\boldsymbol{\eta}_2\|} & \cdots & \frac{<\boldsymbol{\eta}_m, \boldsymbol{\eta}_m>}{\|\boldsymbol{\eta}_m\| \|\boldsymbol{\eta}_i\|} \end{pmatrix}. \tag{4.7}$$

Now, we introduce some lemmas helpful in proving our main results. Until this point, we have used notation that has suppressed the dependence of matrices on

n. As we now investigate properties related to convergence, we do not suppress n. For example, $\mathbf{C}(n)$ signifies that $\mathbf{C}$ depends on n. Thus $\mathbf{C}(n)$, the n^{th} term in the sequence $\{\mathbf{C}(i)\}_{i=1}^{\infty}$, is to be distinguished from $\mathbf{C}$.

LEMMA 1: *Let $\lambda_i(n)$ for $1 \leq i \leq m$ be the eigenvalues of $\mathbf{G}(n)$ and assume that $< \boldsymbol{\eta}_i(n), \boldsymbol{\eta}_i(n) > \to \infty$ as $n \to \infty$ for each i satisfying $1 \leq i \leq m$. Then* $\min_{\{1 \leq i \leq m\}}\{\lambda_i(n) : 1 \leq i \leq m\} \to \infty$ *as $n \to \infty$.*

Proof:

To avoid trite work, we assume that n is large enough so that $\sum_{t=1}^{n} \eta_{i,t}(n)^2$ is strictly positive for each i in $\{1, 2, \ldots, m\}$ (under any non-degenerate experimental design, $\sum_{t=1}^{n} \eta_{i,t}(n)^2 = 0$ only when each and every time index in $\{1, \ldots, n\}$ is a changepoint time). Since $\sum_{t=1}^{n} \eta_{i,t}^2(n)$ is nondecreasing in n for each i, it makes sense to assume this positivity for each and every large n uniformly in i. Let $0 \leq \lambda_1(n) \leq \ldots \leq \lambda_m(n)$ be the eigenvalues of $\mathbf{G}(n)$; the assumed linear independence of $\boldsymbol{\eta}_i$ justifies positivity of each eigenvalue. Let $\mathbf{U}_i(n)$ be unit length eigenvectors (with respect to the Euclidean) norm satisfying

$$\mathbf{G}(n)\mathbf{U}_i(n) = \lambda_i(n)\mathbf{U}_i(n)$$

for $1 \leq i \leq m$. Observe that $\mathbf{U}_i^T(n)\mathbf{U}_i(n) = 1$ for all $1 \leq i \leq m$ and $n \geq 1$ and define $\mathbf{V}(n) = \mathbf{E}(n)^{\frac{1}{2}}\mathbf{U}_1(n)$. Then

$$\mathbf{V}^T(n)\mathbf{C}(n)\mathbf{V}(n) \quad = \quad \mathbf{U}_1^T(n)\mathbf{E}(n)^{\frac{1}{2}}\mathbf{E}^{-\frac{1}{2}}(n)\mathbf{G}(n)\mathbf{E}^{-\frac{1}{2}}(n)\mathbf{E}^{\frac{1}{2}}(n)\mathbf{U}_1(n)$$

$$= \mathbf{U}_1^T(n)\mathbf{G}(n)\mathbf{U}_1(n)$$

$$= \lambda_1(n).$$

Noting that $\mathbf{C}_{i,i}(n) = 1$ for all $1 \leq i \leq m$ and $n \geq 1$, we have

$$\begin{aligned}
\mathbf{V}^T(n)\mathbf{C}(n)\mathbf{V}(n) &= \mathbf{U}_1^T(n)\mathbf{E}^{\frac{1}{2}}(n)\mathbf{C}(n)\mathbf{E}^{\frac{1}{2}}(n)\mathbf{U}_1(n) \\
&= \sum_{i=1}^{m} \mathbf{E}_{i,i}(n)\mathbf{U}_{i,1}^2(n)\mathbf{C}_{i,i}(n) \\
&\geq \left(\min_{\{1\leq i\leq m\}}\{\mathbf{E}_{i,i}(n)\}\right)\left(\sum_{i=1}^{m} \mathbf{U}_{i,1}^2(n)\right) \\
&= \min_{\{1\leq i\leq m\}}\{\mathbf{E}_{i,i}(n)\}.
\end{aligned}$$

Combining the above two relations gives

$$\lambda_i(n) \geq \min_{\{1\leq i\leq m\}}\{\mathbf{E}_{i,i}(n)\}.$$

To finish the proof, simply use the assumption that $<\boldsymbol{\eta}_i(n), \boldsymbol{\eta}_i(n)>= \mathbf{E}_{i,i}(n) \rightarrow \infty$ as $n \rightarrow \infty$. Thus, $\lambda_i(n) \rightarrow \infty$ for each i with $1 \leq i \leq m$. #

LEMMA 2: *If* $\min_{\{1\leq i\leq m\}}\{\lambda_i(n) : 1 \leq i \leq m\} \to \infty$ *as* $n \to \infty$, *then* $\left\|\boldsymbol{G}^{-1}(n)\right\| \to 0$, *where* $\|\cdot\|$ *denotes the Euclidean norm defined by* $\|\mathbf{A}\| = \max_{\|\mathbf{x}\|=1} \|\mathbf{A}\mathbf{x}\|$.

Proof:

Observe that $\mathbf{G}(n)$ is symmetric. As in the last result, we assume that it is also positive definite for each large n. Then for such n,

$$\mathbf{G}(n) = \mathbf{U}(n)\mathbf{D}(n)\mathbf{U}^T(n),$$

with $\mathbf{D}(n)$ being a diagonal matrix whose elements are the eigenvalues of $\mathbf{G}(n)$. Thus,

$$\begin{aligned}
\left\|\mathbf{G}^{-1}(n)\right\| &= \left\|\left(\mathbf{U}(n)\mathbf{D}(n)\mathbf{U}^T(n)\right)^{-1}\right\| \\
&= \left\|\mathbf{U}(n)\mathbf{D}^{-1}(n)\mathbf{U}^T(n)\right\| \\
&\leq \|\mathbf{U}(n)\| \left\|\mathbf{D}^{-1}(n)\right\| \left\|\mathbf{U}^T(n)\right\| \\
&= \left\|\mathbf{D}^{-1}(n)\right\| \\
&= \left|\frac{1}{\min_{1\leq i\leq m}\{\lambda_i(n)\}}\right|,
\end{aligned}$$

where we have used that $\mathbf{D}(n)$ is diagonal. Thus, if $\min_{1\leq i\leq m}\{\lambda_i(n)\} \to \infty$ as $n \to \infty$, we get the result. #

We can now prove the consistency of $\hat{\mathbf{\Theta}}(n)$ in the case where $\{\epsilon_t\}$ is independent and identically distributed (IID) and $k(n) \to \infty$ as $n \to \infty$.

THEOREM 1: *Assume that $\{\epsilon_t\}$ is a zero mean sequence of independent and identically distributed random variables with finite second moment σ^2. Assume that $< \boldsymbol{\eta}_i(n), \boldsymbol{\eta}_i(n) > \to \infty$ as $n \to \infty$ for each i with $1 \leq i \leq m$. Then $\hat{\mathbf{\Theta}}(n)$ is consistent as $n \to \infty$.*

Proof:

By (4.5) and (4.6), $\mathrm{Var}\left(\mathbf{E}(n)\mathbf{Z}(n)\right) = \sigma^2\mathbf{G}(n)$ when $\{\epsilon_t\}$ is zero mean and IID. Thus,

$$\begin{aligned}\left\|\mathrm{Var}\left(\hat{\mathbf{\Theta}}(n)\right)\right\| &= \left\|\mathbf{G}^{-1}(n)\sigma^2\mathbf{G}(n)\mathbf{G}^{-1}(n)\right\| \\ &= \sigma^2\left\|\mathbf{G}^{-1}(n)\right\|.\end{aligned}$$

But by Lemmas 1 and 2, $\left\|\mathbf{G}^{-1}(n)\right\| \to 0$. #

THEOREM 2 *Assume that $\{\epsilon_t\}$ is stationary and has short memory in that*

$$\sum_{h=0}^{\infty} |\gamma(h)| < \infty.$$

Assume that $\mathbf{C}_{i,j}(n) \to \mathbf{C}_{i,j}$ as $n \to \infty$, with $\mathbf{C}$ being a symmetric positive definite

$m \times m$ matrix. Then $\hat{\mathbf{\Theta}}(n)$ is consistent as $n \to \infty$.

Proof:

Let $\mathbf{\Sigma}(n) = \mathrm{Var}\left(\mathbf{E}(n)\mathbf{Z}(n)\right) = \mathbf{E}(n)\mathrm{Var}\left(\mathbf{Z}(n)\right)\mathbf{E}(n)$. Using (4.2), we obtain

$$\left\|\mathrm{Var}\left(\hat{\mathbf{\Theta}}(n)\right)\right\| = \left\|\mathbf{G}^{-1}(n)\mathbf{\Sigma}(n)\mathbf{G}^{-1}(n)\right\|. \tag{4.8}$$

Using (4.6) gives

$$\begin{aligned}
\left\|\mathrm{Var}\left(\hat{\mathbf{\Theta}}(n)\right)\right\| &= \left\|\left(\mathbf{E}^{\frac{1}{2}}(n)\mathbf{C}(n)\mathbf{E}^{\frac{1}{2}}(n)\right)^{-1}\mathbf{\Sigma}(n)\left(\mathbf{E}^{\frac{1}{2}}(n)\mathbf{C}(n)\mathbf{E}^{\frac{1}{2}}(n)\right)^{-1}\right\| \\
&\leq \left\|\mathbf{E}^{-\frac{1}{2}}(n)\mathbf{C}^{-1}(n)\right\|^2 \left\|\mathbf{E}^{-\frac{1}{2}}(n)\mathbf{\Sigma}(n)\mathbf{E}^{-\frac{1}{2}}(n)\right\| \\
&\leq \left\|\mathbf{E}^{-\frac{1}{2}}(n)\right\|^2 \left\|\mathbf{C}^{-1}(n)\right\|^2 \left\|\mathbf{E}^{-\frac{1}{2}}(n)\mathbf{\Sigma}(n)\mathbf{E}^{-\frac{1}{2}}(n)\right\|.
\end{aligned}$$

We now we construct a bound for the elements of $\mathbf{\Sigma}(n)$. By definition,

$$\begin{aligned}
\mathbf{\Sigma}_{i,j}(n) &= \gamma(0)\sum_{t=1}^{n}\eta_{i,t}(n)\eta_{j,t}(n) \\
&+ \sum_{h=1}^{n-1}\gamma(h)\sum_{t=1}^{n-h}\left(\eta_{i,t}(n)\eta_{j,t+h}(n) + \eta_{i,t+h}(n)\eta_{j,t}(n)\right).
\end{aligned}$$

For $1 \le i, j \le m$, and $n \ge 1$,

$$\begin{aligned}
|\mathbf{\Sigma}_{i,j}(n)| &\le \gamma(0)\left(\sum_{t=1}^{n}\eta_{i,t}^2(n)\right)^{\frac{1}{2}}\left(\sum_{t=1}^{n}\eta_{j,t}^2(n)\right)^{\frac{1}{2}} \\
&+ \sum_{h=1}^{n-1}|\gamma(h)|\left(\sum_{t=1}^{n}\eta_{i,t+h}^2(n)\right)^{\frac{1}{2}}\left(\sum_{t=1}^{n}\eta_{j,t}^2(n)\right)^{\frac{1}{2}} \\
&+ \sum_{h=1}^{n-1}|\gamma(h)|\left(\sum_{t=1}^{n}\eta_{i,t}^2(n)\right)^{\frac{1}{2}}\left(\sum_{t=1}^{n}\eta_{j,t+h}^2(n)\right)^{\frac{1}{2}} \\
&\le \gamma(0)\,\|\boldsymbol{\eta_i}(n)\|\,\|\boldsymbol{\eta_j}(n)\| + 2\sum_{h=1}^{n-1}|\gamma(h)|\,\|\boldsymbol{\eta_i}(n)\|\,\|\boldsymbol{\eta_j}(n)\| \\
&\le \|\boldsymbol{\eta_i}(n)\|\,\|\boldsymbol{\eta_j}(n)\|\left(\gamma(0) + 2\sum_{h=1}^{n-1}|\gamma(h)|\right).
\end{aligned}$$

Now apply the assumed short memory of $\{\epsilon_t\}$ to infer that there exist $W_{i,j}$ with $0 \le W_{i,j} < \infty$ satisfying

$$|\mathbf{\Sigma}_{i,j}(n)| \le \|\boldsymbol{\eta_i}(n)\|\,\|\boldsymbol{\eta_j}(n)\|\; W_{i,j}. \tag{4.9}$$

Thus, for all $n \ge 1$,

$$\left\|\mathbf{E}^{-\frac{1}{2}}(n)\mathbf{\Sigma}(n)\mathbf{E}^{-\frac{1}{2}}(n)\right\| \le \max_{1\le i,j\le m} W_{i,j} \tag{4.10}$$

and $\left\|\mathbf{E}^{-\frac{1}{2}}(n)\mathbf{\Sigma}(n)\mathbf{E}^{-\frac{1}{2}}(n)\right\|$ is bounded. By assumption, $\mathbf{C}(n) \to \mathbf{C}$ as $n \to \infty$ with $\mathbf{C}$ being invertible, thus, $\mathbf{C}^{-1}(n) \to \mathbf{C}^{-1}$ as $n \to \infty$ and $\left\|\mathbf{C}^{-1}(n)\right\| \to \left\|\mathbf{C}^{-1}\right\|$ as $n \to \infty$. Thus $\|\mathbf{C}^{-1}(n)\|$ is bounded. Finally, we note that $\left\|\mathbf{E}^{-\frac{1}{2}}(n)\right\| \to 0$ as $n \to \infty$. Combining these results, we see that

$$\left\|\mathbf{E}^{-\frac{1}{2}}(n)\right\|^2\left\|\mathbf{C}^{-1}(n)\right\|^2\left\|\mathbf{E}^{-\frac{1}{2}}(n)\mathbf{\Sigma}(n)\mathbf{E}^{-\frac{1}{2}}(n)\right\| \to 0 \tag{4.11}$$

as $n \to \infty$. Hence $\text{Var}\left(\hat{\mathbf{\Theta}}(n)\right) \to 0$ as $n \to \infty$. #

One may ask when there exists limits for the $\mathbf{C}_{i,j}(n)$'s. This may be shown in simple linear regression settings with a renewal structure imposed upon the change-point times. Specifically suppose that

$$f_1(t) = t \tag{4.12}$$

and that the τ_i's are the points in a renewal sequence that are independent of the errors $\{\epsilon_t\}$. Then

$$\frac{1}{n}\sum_{t=1}^{n} \eta_{i,t}(n) \longrightarrow \text{E}\left[A_\infty\right], \text{ a.s.} \tag{4.13}$$

where A_∞ is the limiting expected age of the renewal process in its stationary setting. Here, one can show that

$$\frac{1}{n}\sum_{t=1}^{n} \eta_{i,j}^2(n) \tag{4.14}$$

also converges to a non-zero limit almost surely.

We are currently investigating the existence of limits for $\mathbf{C}_{i,j}(n)$ in the case of a polynomial regression.

Also, we have not established joint asymptotic normality of $\hat{\theta}_1, \ldots, \hat{\theta}_m$. This is currently being investigated under the assumptions of Theorem 2 and a polynomial design. Such results are currently being investigated.

References

Bickel, P.J, and K.A. Doksum (2001), *Mathematical Statistics: Basic Ideas and Selected Topics Volume One*, Second Edition, Prentice Hall, Upper Saddle River, New Jersey.

Gantmacher, F.R. (1977), *Matrix Theory Volume One*, Third Edition, Chelsea Publishing Company, United States of America.

Graybill, F.J. (1976), *Theory and Application of the Linear Model*, Duxbury, Pacific Grove, CA.

Lu, Q., and R.B. Lund (2007), Simple linear regression with multiple level shifts, *Canadian Journal of Statistics*, **37**, 447-458.

Mitchell, J.M., Jr., (1953), On the causes of instrumentally observed secular temperature trends, *Journal of Meteorology*, **10**, 244-261.

Menne, M.J., and Williams, C.N. Jr. (2005), Detection of undocumented changepoints using multiple test statistics and composite reference series hydrology, *Journal of Climate*, 18, 4271-4286.

Peterson, T.C., and D.R. Easterling (1994), Creation of homogeneous composite climatological reference series, *International Journal of Climatology*, **14**: 671-9.

Solow, A.R. (1987), Testing for climate change: An application of the two-phase regression model, *Journal of Climate and Applied Meteorology*, **26**: 1401-5.