5-2012

Asymptotics and confidence estimation in segmented regression models.

Rebekah Ann Robinson
University of Louisville

ASYMPTOTICS AND CONFIDENCE ESTIMATION IN SEGMENTED REGRESSION MODELS

By

Rebekah Ann Robinson
B.S., University of Evansville, 2006
M.A., University of Louisville, 2008

A Dissertation
Submitted to the Faculty of the
College of Arts and Sciences of the University of Louisville
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

Department of Mathematics
University of Louisville
Louisville, KY

May 2012

ASYMPTOTICS AND CONFIDENCE ESTIMATION IN SEGMENTED REGRESSION MODELS

Submitted by

Rebekah Ann Robinson

A Dissertation Approved on

April 6, 2012

(Date)

by the Following Reading and Examination Committee:

Ryan Gill, Dissertation Director

Kiseop Lee

Christine Rich

Prasanna Sahoo

Cristina Tone

ACKNOWLEDGEMENTS

I would like to thank my advisor, Professor Ryan Gill, for the endless amount of time he devoted to working with me during my time as a graduate student. He inspired me, very early on, to choose statistics as my area of study and I am so very grateful to have had my ideas and research guided by him. I could not have done this without his patience and unwavering support and encouragement.

I would also like to thank the members of my committee, Professors Kiseop Lee, Christine Rich, Prasanna Sahoo, and Cristina Tone, for taking the time to read and understand my dissertation and for providing encouragement and guidance along the way.

My thanks also goes to the mathematics faculty at the University of Louisville and the University of Evansville for supporting me through my years of education. I would especially like to thank Dr. Thomas Riedel for all of the insight and encouragement he has provided.

Finally, I would like to thank my family for their unending patience and love, and a lifetime of support. A very special thanks goes to my husband, Chris, for always believing in my abilities and my dreams. I am forever thankful to have him by my side.

ABSTRACT

ASYMPTOTICS AND CONFIDENCE ESTIMATION IN SEGMENTED REGRESSION MODELS

Rebekah Ann Robinson

May 11, 2012

Standard regularity assumptions for regression models are not satisfied in segmented regression models with an unknown change point, and consequently standard asymptotic results and inferential methods for confidence estimation are not applicable. This dissertation considers a clustered segmented regression model with a continuity constraint and considers estimators of the model parameters based on the likelihood principle. The strong consistency of the maximum likelihood estimators is established. To consider the asymptotic distribution, two cases must be considered. Case 1 occurs when the true change point occurs between two of the observation times, while Case 2 occurs when the true change point occurs at one of the observation times. In each case, the asymptotic distribution of relevant estimators is derived. These results are used to develop a new comprehensive algorithm for constructing a confidence interval for the change point parameter which works for both cases using all available data in determining the confidence bounds. This algorithm is compared to an existing method known as the removal algorithm. A slight modification to the comprehensive algorithm is also considered. Finally, these methods for obtaining confidence intervals are compared by simulation studies and applied to a real data set.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

This chapter provides an introduction to change point models, a literature review regarding the topic, and a detailed explanation of the intention of this dissertation.

1.1 Change point models

Whether approximating the point at which an economic policy decision affects the market value of a company or determining the time at which an athlete's body no longer maximizes metabolic burn of energy, change point analysis is invaluable in the modern world. Change point problems take a standard model and allow a parameter to change at an unknown time, namely, the *change point.* The classical change point model consists of data following a distribution with parameter θ_0 up until some time τ, whereafter the parameter changes to θ_1. The primary focus of such a model is the estimation of the location of the change point τ. It is also of interest to estimate the magnitude of the change, i.e., the pre-change parameter, θ_0, and the post-change parameter, θ_1. The model can be stated as follows. Observe $x_1, \ldots, x_\tau, x_{\tau+1}, \ldots, x_n$, where

$$x_j \sim \begin{cases} P(\cdot|\theta_0) & \text{for } j = 1, \ldots, \tau \\ P(\cdot|\theta_1) & \text{for } j = \tau + 1, \ldots, n. \end{cases}$$

The most primitive way of locating a change point is to simply view the data and visuall approximate the location of a change. Such a method of approximation is

imprecise and has little applicability. Change point analysis is highly applicable in various fields and thus, developing better methods to approximate these models has been receiving considerable attention since the 1950s. In the health field, change point problems are popular as tools to model disease trends and mortality rates. In economics, change point problems are useful as a means to investigate whether an economic variable, such as the stock market, borrowing and lending behavior, or a change in government policy, has experienced a structural change. In manufacturing, change point analysis can facilitate quality control measures: manufacturers desire rapid detection of deteriorations in the quality of the good they are producing. Finally, in the midst of current debates on global warming, change point problems are useful in modeling weather patterns.

Change point problems date back to Page (1954). His emphasis was primarily on applications to industry and he proposed various process inspection schemes to detect deteriorations in quality of continuous mass production. One such sequential inspection scheme that was mentioned in some detail was the use of control charts. If a sample point falls outside the pre-determined control limits (or a specified proportion falls outside of the warning lines) on such a chart, the process is deemed out of control. He considered *average run length* as a means to measure performance of an inspection scheme. Average run length is the expected number of items sampled before intervention of the process is taken. Lai (1995) added to the work of Page (1954) by unifying the theory of sequential changepoint detection. He considered both a *moving average* and a CUSUM inspection scheme in comparison with the average run length.

1.2 Segmented regression models

A segmented regression model is a standard linear regression model where the coefficient values change at an unknown time. In the context of regression, the estimation of the location of the change point and the pre-change and post-change regression coefficients are of interest. While this is the main goal, there are distinctions between the types of models that have been previously considered.

One very important distinction is whether the change is assumed to be continuous or abrupt. Specifically, suppose that there is no continuity constraint imposed on the model, i.e., there is an abrupt change. Such a model can be stated as having n data points $(u_1, y_1), \ldots, (u_n, y_n)$ observed where the values of the explanatory variables are $u_1 \leq \ldots \leq u_n$ and the independent responses are

$$y_i \sim \begin{cases} \mathcal{N}(\omega_{01} + \omega_{02} u_i, \sigma_0^2) & \text{for } i = 1, \ldots, k_0 \\ \mathcal{N}(\omega_{03} + \omega_{04} u_i, \sigma_0^2) & \text{for } i = k_0 + 1, \ldots, n. \end{cases}$$

In this case, the unknown change, $k_0 \in \{1, \ldots, n\}$, is not identifiable between observations. In other words, the change can only occur at an observation. The intercept for each phase is ω_{01} and ω_{03}. respectively, and the respective slope for each phase is ω_{02} and ω_{04}.

This can be compared to a segmented regression model with a continuity constraint. Observe $(u_1, y_1), \ldots, (u_n, y_n)$, where $u_1 \leq \ldots \leq u_n$ and

$$y_i \sim \mathcal{N}(\alpha_0 + \beta_0 u_i + \delta_0 (u_i - \tau_0)_+, \sigma_0^2), i = 1, \ldots, n.$$

Here, $x_+ = \max\{0, x\}$. The unknown changepoint parameter, $\tau_0 \in (u_1, u_n)$, can be identified either at or between observations. In this model, α_0 is the unknown intercept, β_0 is the unknown slope, and δ_0 is the unknown change in slope after τ_0.

There are also differences in the methods used to estimate the parameters, i.e., maximum likelihood, least squares, Bayesian, nonparametric, etc. The assumption about the error terms differs between models as well. Another important

distinction is in the number of change points being considered. The AMOC (at most one change) model has received considerable attention in the past, while the multiple change point model has recently become of more interest. It is important to note whether the number of changes is known and just the location is to be estimated or if it is of interest to also estimate the number of changes. These different features of the model lend themselves to various research done on the topic.

The abrupt change model was studied by Quandt (1958), Bai (1997), Bai and Perron (1998), and Chong (2001). Quandt (1958) considered a two-phase linear model with exactly one change. He assumed the regression lines on either side of the change did not necessarily meet. He proposed estimating the change point sequentially by dividing the data at each observation and estimating separate regression lines for each division. The change point estimate was the observation whose division maximized the likelihood function. Quandt was also the first to suggest using a hypothesis test to determine whether or not a change had even occurred. Bai (1997) and Bai and Perron (1998), on the other hand, considered multiple change point models with no continuity constraint. Bai (1997) proposed a procedure for sequentially (one-by-one) estimating the changes. This was done by first treating the model as if it had only one change and estimating it by minimizing the sum of squares function. Then, the sample was divided at the estimated change and a break was estimated in each of the subsamples. Bai also derived the asymptotic distribution of the change points using a "repartition" method. Bai and Perron (1998) discussed different ways to incorporate hypothesis testing into the idea of sequential detection of multiple change points. Chong (2001) applied hypothesis testing and sample-splitting to a model with an unknown number of changes.

Sylwester (1965) and Hudson (1966) were among the first to consider a broken line model with a continuity constraint. Sylwester (1965) dealt primarily with MLE for a two-phase linear regression model assuming equally spaced observations in the

interval $[0, 1]$, assuming the errors were independent normal random variables with mean 0 and constant variance. Sylwester's broken line model had constant slope before the change and zero slope after the change. He derived the MLE's for the parameters of the model and proposed a "pseudo-problem" to derive the asymptotic normality of the estimators. In the "pseudo-problem", a decreasing interval around the true change was deleted, which enabled the asymptotic normality to be derived for the "pseudo" estimates. By showing that the the MLE's and the pseudo-MLE's have the same asymptotic distribution, he was able to discuss the asymptotics of the MLE's for this particular model. Feder (1975) extended the idea of using a "pseudo-problem" to derive asymptotic normality to a more general model.

Hudson (1966) added to Quandt's (1958) model by giving an iterative method for estimating the intersection point of two regression lines using LSE. Hinkley (1971) then extended Hudson's idea to derive the asymptotic normality of the MLE's. Küchenhoff (1997) proposed estimation methods for generalized linear models with change points. Liu, Wu, and Zidek (1997) dealt with segmented multivariate regression and discussed adding a penalty term to the unknown number of changes in a multiple change model to prevent overfitting the model.

1.3 Clustered segmented regression model

The special case of a linear segmented regression model that will be considered in detail in this dissertation is a clustered segmented regression model. Consider the model with m observations at each of N unique observation points. Given a fixed N and $mN = n$, the random variables

$$y_{ij} \sim \mathcal{N}(\alpha_0 + \beta_0 u_i + \delta_0 (u_i - \tau_0)_+, \sigma_0^2)$$

are independent for $i = 1, \ldots, N$ and $j = 1, \ldots, m$. Assume that $-\infty < u_1 < \ldots <$

$u_N < \infty$. Let $\boldsymbol{\theta}_0 = (\alpha_0, \beta_0, \delta_0, \tau_0)^T$ be the vector of true values of the parameters, where τ_0 is the unknown changepoint parameter in (u_1, u_N), α_0 is the unknown intercept, β_0 is the unknown slope coefficient, and δ_0 is the change in slope after τ_0.

Gill, et. al (2009) considered a similar clustered changepoint model with an unknown change. The focus was on clustered logistic changepoint regression, the MLE's of the model's parameters, the consistency and asymptotics of the estimators, and the behavior of the bootstrap method for confidence estimation. The bounds of the confidence intervals were shown to be consistent if the true change was located between two observations but not if the true change was located at an observation. A removal algorithm was described to dealt with the latter case.

This dissertation is organized as follows. In Chapter 2, the MLE's are derived for three types of segmented regression models; a one change abrupt model, a one change model with a continuity constraint, and a one change clustered model with a continuity constraint. The rest of the dissertation deals only with the one change clustered model with a continuity constraint. Next, the consistency and asymptotic normality of the MLE's are discussed in Chapter 3. Chapter 4 describes two different algorithms used in confidence estimation of the change point. The removal algorithm from Gill, et. al (2009) is described along with the comprehensive algorithm that I have written. Strengths and weaknesses for the comprehensive algorithm are discussed and an alternative idea is suggested to fix the weaknesses of the comprehensive algorithm. Chapter 5 compares the methods discussed in Chapter 4. Simulation results and an example are also included in Chapter 5. Some of the major proofs and results in Chapter 4 are located in the Appendix.

CHAPTER 2
MAXIMUM LIKELIHOOD ESTIMATION

The general method of maximum likelihood estimation will be used to derive the estimates for the parameters in the models described in Chapter 1. The idea behind maximum likelihood estimation is to choose the unknown parameter values in a way that maximizes the probability of getting the sample values.

DEFINITION 2.1. *Let $x_1, \ldots, x_n$ be a random sample of observed values of a random variable $\boldsymbol{X}$ with unknown parameter $\boldsymbol{\theta}$. The **likelihood function** is given by $L(\boldsymbol{\theta}) = f(\boldsymbol{x}; \boldsymbol{\theta})$ where $f(\boldsymbol{x}; \boldsymbol{\theta})$ is the joint probability density of the sample.*

DEFINITION 2.2. *The value of $\boldsymbol{\theta}$ that maximizes $L(\boldsymbol{\theta})$ is referred to as the **maximum likelihood estimator** of $\boldsymbol{\theta}$ and is denoted by $\hat{\boldsymbol{\theta}}$. This can be stated as $\hat{\boldsymbol{\theta}} = \max_{\boldsymbol{\theta}} L(\boldsymbol{\theta})$.*

The likelihood function, $L(\boldsymbol{\theta})$, is a function of the parameter $\boldsymbol{\theta}$ for given sample values, $\boldsymbol{x}$, while the probability density (or distribution) function is a function of random variables $\boldsymbol{X}$ for a given value of the parameter. The method of maximum likelihood estimation chooses the value of the parameter that most likely produced the given observations $\boldsymbol{x}$.

2.1 Without continuity constraint

A segmented regression model with no continuity constraint can be defined by observing a discrete set of data points $(u_1, y_1), \ldots, (u_n, y_n)$ such that $u_1 \leq \ldots \leq u_n$ and $y_1, \ldots, y_n$ are independent random variables such that

$$y_i \sim \begin{cases} \mathcal{N}(\omega_{01} + \omega_{02}u_i, \sigma_0^2) & \text{for } i = 1, \ldots, k_0 \\ \mathcal{N}(\omega_{03} + \omega_{04}u_i, \sigma_0^2) & \text{for } i = k_0 + 1, \ldots, n \end{cases};$$

let $f(y_1 \ldots, y_n; k_0, \boldsymbol{\omega}_0, \sigma_0^2)$ denote the joint density of $y_1, \ldots, y_n$. Using maximum likelihood estimation (MLE) and least squares estimation (LSE), parameters $\boldsymbol{\omega}_0 = (\omega_{01}, \omega_{02}, \omega_{03}, \omega_{04})^T, k_0$, and σ_0^2 can be estimated. The likelihood function is

$$\begin{aligned} &L(k, \boldsymbol{\omega}, \sigma^2) = f(y_1, \ldots, y_n; k, \boldsymbol{\omega}, \sigma^2) \\ &= \prod_{i=1}^{n} f(y_i; k, \boldsymbol{\omega}, \sigma^2) \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left[\frac{-1}{2\sigma^2}\left(\sum_{i=1}^{k}(y_i - \omega_1 - \omega_2 u_i)^2 + \sum_{i=k+1}^{n}(y_i - \omega_3 - \omega_4 u_i)^2\right)\right]. \end{aligned}$$

The values of $k, \boldsymbol{\omega}$, and σ^2 that maximize the likelihood function $L(k, \boldsymbol{\omega}, \sigma^2)$ will also maximize the log-likelihood function

$$l(k, \boldsymbol{\omega}, \sigma^2) = -\tfrac{n}{2}\ln(2\pi\sigma^2) - \tfrac{1}{2\sigma^2}\left(\sum_{i=1}^{k}(y_i - \omega_1 - \omega_2 u_i)^2 + \sum_{i=k+1}^{n}(y_i - \omega_3 - \omega_4 u_i)^2\right).$$

Since the sum of squares function for this model is

$$Q(k, \boldsymbol{\omega}) = \sum_{i=1}^{k}(y_i - \omega_1 - \omega_2 u_i)^2 + \sum_{i=k+1}^{n}(y_i - \omega_3 - \omega_4 u_i)^2,$$

the log-likelihood function can be re-written as

$$l(k, \boldsymbol{\omega}, \sigma^2) = -\tfrac{n}{2}\ln(2\pi\sigma^2) - \tfrac{Q(k,\boldsymbol{\omega})}{2\sigma^2}.$$

Notice that maximizing $l(k, \boldsymbol{\omega}, \sigma^2)$ is equivalent to minimizing $Q(k, \boldsymbol{\omega})$, so the MLE's for k and $\boldsymbol{\omega}$ are the same as the LSE's for k and $\boldsymbol{\omega}$. Letting $\boldsymbol{y} = (y_1, \ldots, y_n)^T$ and letting $\boldsymbol{U}_k$ be the design matrix (for a fixed k) defined as

$$\boldsymbol{U}_k = \begin{bmatrix} 1 & u_1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & u_k & 0 & 0 \\ 0 & 0 & 1 & u_{k+1} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & u_n \end{bmatrix},$$

the sum of squares function can be re-written in matrix form as

$$\begin{aligned} Q(k,\boldsymbol{\omega}) &= \|\boldsymbol{y} - \boldsymbol{U}_k\boldsymbol{\omega}\|^2 \\ &= (\boldsymbol{y} - \boldsymbol{U}_k\boldsymbol{\omega})^T(\boldsymbol{y} - \boldsymbol{U}_k\boldsymbol{\omega}) \\ &= \boldsymbol{y}^T\boldsymbol{y} - 2\boldsymbol{\omega}^T\boldsymbol{U}_k^T\boldsymbol{y} + \boldsymbol{\omega}^T\boldsymbol{U}_k^T\boldsymbol{U}_k\boldsymbol{\omega}. \end{aligned}$$

To obtain the LSE of $\boldsymbol{\omega}$ (which is equivalent to the MLE of $\boldsymbol{\omega}$), fix k and minimize $Q(k,\boldsymbol{\omega})$ with respect to $\boldsymbol{\omega}$. Then, the partial derivative with respect to $\boldsymbol{\omega}$ is $\frac{\partial Q(k,\boldsymbol{\omega})}{\partial \boldsymbol{\omega}} = -2\boldsymbol{U}_k^T\boldsymbol{y} + 2\boldsymbol{U}_k^T\boldsymbol{U}_k\boldsymbol{\omega}$. Setting this equal to 0 and solving for $\boldsymbol{\omega}$, the MLE $\hat{\boldsymbol{\omega}}_k$ is obtained as follows.

$$\begin{aligned} -2\boldsymbol{U}_k^T\boldsymbol{y} + 2\boldsymbol{U}_k^T\boldsymbol{U}_k\hat{\boldsymbol{\omega}}_k &= \boldsymbol{0} \\ \boldsymbol{U}_k^T\boldsymbol{U}_k\hat{\boldsymbol{\omega}}_k &= \boldsymbol{U}_k^T\boldsymbol{y} \\ \hat{\boldsymbol{\omega}}_k &= (\boldsymbol{U}_k^T\boldsymbol{U}_k)^{-1}\boldsymbol{U}_k^T\boldsymbol{y}. \end{aligned}$$

The MLE for σ_0^2 can be found similarly using the log-likelihood function. Holding k fixed, and using $\hat{\boldsymbol{\omega}}_k$,

$$\begin{aligned} \boldsymbol{0} &\equiv \frac{\partial l(k,\hat{\boldsymbol{\omega}}_k,\sigma_k^2)}{\partial \sigma_k^2} = -\frac{n}{2\sigma_k^2} + \frac{Q(k,\hat{\boldsymbol{\omega}}_k)}{2(\sigma_k^2)^2} \\ &\Rightarrow \frac{n}{2\hat{\sigma}_k^2} = \frac{Q(k,\hat{\boldsymbol{\omega}}_k)}{2(\hat{\sigma}_k^2)^2} \\ &\Rightarrow \hat{\sigma}_k^2 = \frac{Q(k,\hat{\boldsymbol{\omega}}_k)}{n}. \end{aligned}$$

Both $\hat{\boldsymbol{\omega}}_k$ and $\hat{\sigma}_k^2$ are conditioned upon a fixed k. To estimate k, choose the value

of k that maximizes the profile likelihood function (or equivalently minimizes the profile sum of squares function).

$$\hat{k} \equiv \text{argmax}_k l(k, \hat{\boldsymbol{\omega}}_k, \hat{\sigma}_k^2) = \text{argmin}_k Q(k, \hat{\boldsymbol{\omega}}_k).$$

Thus, $\hat{\boldsymbol{\omega}} = \hat{\boldsymbol{\omega}}_{\hat{k}}$ and $\hat{\sigma}^2 = \hat{\sigma}_{\hat{k}}^2$.

2.2 With continuity constraint

Recall that a segmented regression model with a continuity constraint is defined as follows. Observe $(u_1, y_1), \ldots, (u_n, y_n)$ where $u_1 \leq \ldots \leq u_n$ and given that $x_+ = \max\{0, x\}$, the random variables

$$y_i \sim \mathcal{N}(\alpha_0 + \beta_0 u_i + \delta_0 (u_i - \tau_0)_+, \sigma_0^2)$$

are independent for $i = 1, \ldots, n$. Similar to before, the log-likelihood function is $l(\boldsymbol{\theta}, \sigma^2) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{Q(\boldsymbol{\theta})}{2}$, where $Q(\boldsymbol{\theta}) = \sum_{i=1}^{n}(y_i - \mu_i(\boldsymbol{\theta}))^2$ and

$$\mu_i(\boldsymbol{\theta}) = \alpha + \beta u_i + \delta(u_i - \tau)_+ = \begin{cases} \alpha + \beta u_i & \text{if } u_i \leq \tau \\ \alpha + \beta u_i + \delta(u_i - \tau) & \text{if } u_i > \tau \end{cases}.$$

To obtain estimates for $\boldsymbol{\theta}_0 = (\alpha_0, \beta_0, \delta_0, \tau_0)^T$ and σ_0^2, the log-likelihood function can be maximized, or equivalently, the sum of squares function, $Q(\boldsymbol{\theta})$, minimized. Since the true change in this model, τ_0, is identified either at an observation or between two observations, both cases must be considered when finding the global minimizer of $Q(\boldsymbol{\theta})$.

Suppose for every $k = 1, \ldots, n-1$, the restriction of $\tau \in (u_k, u_{k+1})$ is made and estimates corresponding to that interval are searched for. Then,

$$\begin{aligned} Q_k(\boldsymbol{\theta}) &= \sum_{i=1}^{k}(y_i - \alpha - \beta u_i)^2 + \sum_{i=k+1}^{n}(y_i - (\alpha - \delta\tau) - (\beta + \delta)u_i)^2 \\ &= \sum_{i=1}^{k}(y_i - \omega_1 - \omega_2 u_i)^2 + \sum_{i=k+1}^{n}(y_i - \omega_3 - \omega_4 u_i)^2 \end{aligned}$$

where $\omega_1 = \alpha, \omega_2 = \beta, \omega_3 = \alpha - \delta\tau$, and $\omega_4 = \beta + \delta$. Using this notation, define the kth *super log-likelihood* function as

$$l_k(\boldsymbol{\theta}, \sigma^2) = -\tfrac{n}{2}\ln(2\pi\sigma^2) - \tfrac{Q_k(\boldsymbol{\theta})}{2}$$

for any $k = 1, \ldots, n-1$. Now, letting $\boldsymbol{\omega} = (\omega_1, \omega_2, \omega_3, \omega_4)^T$, it is useful to consider the function

$$\psi(\boldsymbol{\omega}) = \begin{bmatrix} \alpha \\ \beta \\ \delta \\ \tau \end{bmatrix} = \begin{bmatrix} \omega_1 \\ \omega_2 \\ \omega_4 - \omega_2 \\ \frac{\omega_1 - \omega_3}{\omega_4 - \omega_2} \end{bmatrix}.$$

Letting

$$\boldsymbol{U}_k = \begin{bmatrix} 1 & u_1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & u_k & 0 & 0 \\ 0 & 0 & 1 & u_{k+1} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & u_n \end{bmatrix} \text{ and } \boldsymbol{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

gives $Q_k(\boldsymbol{\omega}) = \|\boldsymbol{y} - \boldsymbol{U}_k\boldsymbol{\omega}\|^2$. Then the LSE (MLE) for the restricted case when $\tau \in (u_k, u_{k+1})$ is $\hat{\boldsymbol{\omega}}_k = (\boldsymbol{U}_k^T\boldsymbol{U}_k)^{-1}\boldsymbol{U}_k^T\boldsymbol{y} = (\hat{\omega}_{1k}, \hat{\omega}_{2k}, \hat{\omega}_{3k}, \hat{\omega}_{4k})^T$. By the invariance of the MLE, $\psi(\hat{\boldsymbol{\omega}}_k) = \hat{\boldsymbol{\theta}}_k = (\hat{\alpha}_k, \hat{\beta}_k, \hat{\delta}_k, \hat{\tau}_k)^T$ is a local minimizer of Q as long as $\hat{\tau}_k \in (u_k, u_{k+1})$. So for every value of $k = 2, \ldots, n-1$, $\hat{\boldsymbol{\theta}}_k$ is a local (and thus possibly global) minimizer of Q if $\hat{\tau}_k \in (u_k, u_{k+1})$.

Note that if $\tau = u_i$ for any $i = 2, \ldots, n-1$, the function Q is continuous but not differentiable at τ. So if the change occurs at an observation point, the minimizer of $Q(\alpha, \beta, \delta, u_k)$ for $k = 1, \ldots, n-1$ is computed using the design matrix

$$\boldsymbol{U}_k^{\#} = \begin{bmatrix} 1 & u_1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & u_k & 0 \\ 1 & u_{k+1} & u_{k+1} - u_k \\ \vdots & \vdots & \vdots \\ 1 & u_n & u_n - u_k \end{bmatrix}$$

and $\boldsymbol{\theta}^{\#} = (\alpha, \beta, \delta)^T$. So $Q(\boldsymbol{\theta}^{\#}) = \|\boldsymbol{y} - \boldsymbol{U}_k^{\#}\boldsymbol{\theta}^{\#}\|^2$. Since u_k is held fixed for each value of $k = 1, \ldots, n-1$, Q can be differentiated with respect to α, β, and δ. Thus,

$$\hat{\boldsymbol{\theta}}_k^{\#} = \begin{bmatrix} \hat{\alpha}_k^{\#} \\ \hat{\beta}_k^{\#} \\ \hat{\delta}_k^{\#} \end{bmatrix} = (\boldsymbol{U}_k^{\#T}\boldsymbol{U}_k^{\#})^{-1}\boldsymbol{U}_k^{\#T}\boldsymbol{y}$$

is the vector of estimators to obtain $Q(\hat{\alpha}_k^{\#}, \hat{\beta}_k^{\#}, \hat{\delta}_k^{\#}, u_k)$ for $k = 2, \ldots, n-1$.

Finally, the global minimizer, $\hat{\boldsymbol{\theta}} = (\hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{\tau})^T$, of Q is obtained by evaluating Q at each of the local minimizers, $\hat{\boldsymbol{\theta}}_k$, as well as at each point $(\hat{\alpha}_k^{\#}, \hat{\beta}_k^{\#}, \hat{\delta}_k^{\#}, u_k)$ for $k = 2, \ldots, n-1$. Then the MLE for σ_0^2 is found as before; $\hat{\sigma}^2 = \frac{Q(\hat{\boldsymbol{\theta}})}{n}$.

2.3 Clustered model

Recall the clustered model as defined in Section 1.3. For fixed N and $mN = n$, suppose that

$$y_{ij} \sim \mathcal{N}(\alpha_0 + \beta_0 u_i + \delta_0(u_i - \tau_0)_+, \sigma_0^2), i = 1, \ldots, N, j = 1, \ldots, m.$$

The estimators of $\boldsymbol{\theta}_0 = (\alpha_0, \beta_0, \delta_0, \tau_0)^T$ and σ_0^2 can be found, as before, by maximizing the log-likelihood function

$$l(\boldsymbol{\theta}, \sigma^2) = -\tfrac{n}{2}\ln(2\pi\sigma^2) - \tfrac{Q(\boldsymbol{\theta})}{2}$$

where $Q(\boldsymbol{\theta})$ is the sum of squares function. Let $\bar{y}_i = \frac{1}{m}\sum_{i=1}^{m} y_{ij}$ and

$$\begin{aligned}
\mu_i(\boldsymbol{\theta}) &= \alpha + \beta u_i + \delta(u_i - \tau)_+ \\
&= \begin{cases} \alpha + \beta u_i & \text{if } u_i \leq \tau \\ (\alpha - \delta\tau) + (\beta + \delta)u_i & \text{if } u_i > \tau \end{cases} \\
&= \begin{cases} \omega_1 + \omega_2 u_i & \text{if } u_i \leq \tau \\ \omega_3 + \omega_4 u_i & \text{if } u_i > \tau \end{cases} \\
&= \mu_i(\boldsymbol{\omega}).
\end{aligned}$$

Then, the sum of squares function can be re-expressed as

$$Q(\boldsymbol{\omega}) = \sum_{i=1}^{N}\sum_{j=1}^{m}(y_{ij} - \mu_i(\boldsymbol{\omega}))^2 = \sum_{i=1}^{N}\sum_{j=1}^{m}(y_{ij} - \bar{y}_i)^2 + m\sum_{i=1}^{N}(\bar{y}_i - \mu_i(\boldsymbol{\omega}))^2$$

because $Q(\boldsymbol{\theta}) = Q(\boldsymbol{\omega})$. Since the second part of this equation is the only part that depends on the parameters, minimizing $Q(\boldsymbol{\omega})$ is equivalent to minimizing $\widetilde{Q}(\boldsymbol{\omega}) = m\sum_{i=1}^{N}(\bar{y}_i - \mu_i(\boldsymbol{\omega}))^2$.

Local minimizers of $\widetilde{Q}$ are found where $\tau \in (u_k, u_{k+1})$ for $k = 1, \ldots, N-1$. Define $\bar{\boldsymbol{y}} = (\bar{y}_1, \ldots, \bar{y}_N)^T$ and the design matrix,

$$\boldsymbol{Z}_k = \begin{bmatrix} 1 & u_1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & u_k & 0 & 0 \\ 0 & 0 & 1 & u_{k+1} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & u_N \end{bmatrix} \text{ for any } k = 1, \ldots, N.$$

Then $\widetilde{Q}_k(\boldsymbol{\omega}) = m\|\bar{\boldsymbol{y}} - \boldsymbol{Z}_k\boldsymbol{\omega}\|^2$ and the LSE for $\boldsymbol{\omega}_0$ is $\hat{\boldsymbol{\omega}}_k = (\boldsymbol{Z}_k^T\boldsymbol{Z}_k)^{-1}\boldsymbol{Z}_k^T\bar{\boldsymbol{y}}$. As before, by the invariance property, $\hat{\boldsymbol{\theta}}_k = \psi(\hat{\boldsymbol{\omega}}_k)$ is the MLE of $\boldsymbol{\theta}_0$. Finally, holding k fixed, the estimate for σ_0^2 is $\hat{\sigma}_k^2 = \frac{Q(\hat{\boldsymbol{\omega}}_k)}{n}$.

CHAPTER 3
ASYMPTOTIC RESULTS

Now that the MLE of $\boldsymbol{\theta}_0$ has been derived, it is of interest to consider how 'close' this estimator gets to the true parameter value as the sample size gets large. The following modes of convergence are used throughout this chapter

DEFINITION 3.1. ***Almost sure convergence*** *(convergence with probability one) A sequence of random variables* $\{\boldsymbol{X}_n\}$ ***converges almost surely*** *(or, equivalently, converges with probability one) to the random variable* $\boldsymbol{X}$ *if*

$$P(\lim_{n\to\infty} \boldsymbol{X}_n = \boldsymbol{X}) = 1.$$

This can be written as $\boldsymbol{X}_n \to_{a.s.} \boldsymbol{X}$ *as* $n \to \infty$.

DEFINITION 3.2. ***Convergence in distribution***
A sequence of random variables $\{\boldsymbol{X}_n\}$ *with cumulative distribution function* F_n ***converges in distribution*** *to a random variable* $\boldsymbol{X}$ *with cumulative distribution function* F *if*

$$\lim_{n\to\infty} F_n(x) = F(x)$$

at every value $x \in \mathbb{R}$ *where* F *is continuous. This can be written as* $\boldsymbol{X}_n \to_d \boldsymbol{X}$.

The results in the remainder of this chapter utilize the following versions of standard theorems from probability theory described, for example, in Bilodeau and Brenner (1999) and van der Vaart (1998) .

THEOREM 3.1. ***Strong Law of Large Numbers (SLLN)***

Let $\{\boldsymbol{X}_n\}$ *be a sequence of independent and identically distributed (i.i.d.) random variables with* $E(\boldsymbol{X}_i) = \boldsymbol{\mu}$ *and finite covariance matrix. As the sample size,* n*, increases without bound, the sample mean converges almost surely to the expected value. In other words,* $\bar{\boldsymbol{X}}_n \rightarrow_{a.s.} \boldsymbol{\mu}$ *as* $n \rightarrow \infty$*.*

THEOREM 3.2. ***Central Limit Theorem***

Let $\{\boldsymbol{X}_n\}$ *be a sequence of independent and identically distributed (i.i.d.) random variables with* $E(\boldsymbol{X}_i) = \boldsymbol{\mu}$ *and* $var(\boldsymbol{X}_i) = \boldsymbol{\Sigma}$*. As* $n \rightarrow \infty$*,* $n(\bar{\boldsymbol{X}}_n - \boldsymbol{\mu}) \rightarrow_d \boldsymbol{Z}$ *as* $n \rightarrow \infty$ *where* $\boldsymbol{Z} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma})$*.*

THEOREM 3.3. ***Continuous Mapping Theorem***

If g *is a function from* $\mathbb{R}^p$ *to* $\mathbb{R}^q$ *which is continuous at every point in a set* C *such that* $P(\boldsymbol{X} \in C) = 1$*, then the following results hold:*

1. $\boldsymbol{X}_n \rightarrow_d \boldsymbol{X} \Rightarrow g(\boldsymbol{X}_n) \rightarrow_d g(\boldsymbol{X})$
2. $\boldsymbol{X}_n \rightarrow_{a.s.} \boldsymbol{X} \Rightarrow g(\boldsymbol{X}_n) \rightarrow_{a.s.} g(\boldsymbol{X})$

3.1 Strong consistency of the MLE

Consistency is one common criteria used to evaluate the goodness of an estimator. Since consistency depends on sample size, it is considered a large sample property. As the sample size gets large, i.e. approaches infinity, the values of a strongly consistent estimator approach the true value of a parameter in the sense that events where this convergence does not occur have probability zero.

THEOREM 3.4. *For the clustered segmented regression model with the continuity constraint,* $\hat{\boldsymbol{\theta}} \rightarrow_{a.s} \boldsymbol{\theta}_0$ *as* $m \rightarrow \infty$*.*

It is important to first discuss the motivation behind the proof for this theorem. The assumption is made that the true change occurs in the kth interval, i.e.,

$\tau_0 \in [u_k, u_{k+1}]$. Based on this assumption, it can be shown that the MLE based on the kth interval, namely $\hat{\boldsymbol{\omega}}_{k,m}$, is strongly consistent as the sample size gets large. Note that the estimator is double indexed by k and m. The index k corresponds to the assumed interval used to find a local minimizer of $\tilde{Q}$. The estimator is also indexed by m to show its dependence on sample size.

The first part of this proof shows that if the estimator is from the correct interval, then it is a strongly consistent estimator of the true parameter, i.e., $\hat{\boldsymbol{\omega}}_{k,m} \rightarrow_{a.s} \boldsymbol{\omega}_0$ as $m \rightarrow \infty$. Next, it is established that among all of the local minimizers of Q, for any $l = 1, \ldots, N-1$, the estimator, $\hat{\boldsymbol{\omega}}_{l,m}$ that minimizes the sum of squares is the estimator based on the correct interval, i.e., $\hat{\boldsymbol{\omega}}_{k,m}$. In other words, $Q(\hat{\boldsymbol{\omega}}_{l,m}) > Q(\hat{\boldsymbol{\omega}}_{k,m})$ for any $l \neq k$. This means $\hat{\boldsymbol{\omega}}_k = \hat{\boldsymbol{\omega}}$.

Putting this together, the estimator that globally minimizes Q is the estimator based on the correctly specified interval and this estimator is strongly consistent. Finally, using the invariance property of the MLE, if $\hat{\boldsymbol{\omega}}$ is a strongly consistent estimator of $\boldsymbol{\omega}_0$, then $\hat{\boldsymbol{\theta}}$ is a strongly consistent estimator of $\boldsymbol{\theta}_0$.

Proof. Define $\boldsymbol{F}^{(l)}(\bar{\boldsymbol{y}}_m, \boldsymbol{\omega}) = \frac{\partial \tilde{Q}^{(l)}(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}} = -2m\boldsymbol{Z}_l^T(\bar{\boldsymbol{y}}_m - \boldsymbol{Z}_l\boldsymbol{\omega})$. Since $\frac{\partial^2 \tilde{Q}^{(l)}(\boldsymbol{\omega})}{\partial \boldsymbol{\omega} \partial \boldsymbol{\omega}^T}$ exists and is positive definite, $\frac{\partial \boldsymbol{F}^{(l)}}{\partial \boldsymbol{\omega}}$ exists. Because it is differentiable, $\boldsymbol{F}^{(l)}$ is a continuous function. Then, by continuity and the Strong Law of Large Numbers,

$$\begin{aligned} \lim_{m\to\infty} \boldsymbol{F}^{(k)}(\bar{\boldsymbol{y}}_m, \hat{\boldsymbol{\omega}}_{k,m}) &= \boldsymbol{F}^{(k)}(\lim_{m\to\infty} \bar{\boldsymbol{y}}_m, \lim_{m\to\infty} \hat{\boldsymbol{\omega}}_{k,m}) \\ &= \boldsymbol{F}^{(k)}(\boldsymbol{\mu}_0, \lim_{m\to\infty} \hat{\boldsymbol{\omega}}_{k,m}) \end{aligned}$$

with probability 1. By definition of LSE, $\lim_{m\to\infty} \boldsymbol{F}^{(k)}(\bar{\boldsymbol{y}}_m, \hat{\boldsymbol{\omega}}_{k,m}) = \lim_{m\to\infty} \boldsymbol{0} = \boldsymbol{0}$ with probability 1. Putting this together, $\boldsymbol{F}^{(k)}(\boldsymbol{\mu}_0, \lim_{m\to\infty} \hat{\boldsymbol{\omega}}_{k,m}) = \boldsymbol{0}$ with probability 1. As defined, $\boldsymbol{Z}_k\boldsymbol{\omega}_0 = \boldsymbol{\mu}_0$. Thus, $\boldsymbol{F}^{(k)}(\boldsymbol{\mu}_0, \boldsymbol{\omega}_0) = -2m\boldsymbol{Z}_k^T(\boldsymbol{\mu}_0 - \boldsymbol{Z}_k\boldsymbol{\omega}_0) = \boldsymbol{0}$. Now, $\boldsymbol{F}^{(k)}(\boldsymbol{\mu}_0, \lim_{m\to\infty} \hat{\boldsymbol{\omega}}_{k,m}) = \boldsymbol{F}^{(k)}(\boldsymbol{\mu}_0, \boldsymbol{\omega}_0)$ with probability 1. Thus, $\lim_{m\to\infty} \hat{\boldsymbol{\omega}}_{k,m} = \boldsymbol{\omega}_0$ with probability 1. Therefore, if $\tau_0 \in [u_k, u_{k+1}]$, then $\hat{\boldsymbol{\omega}}_{k.m} \rightarrow_{a.s} \boldsymbol{\omega}_0$ as $m \rightarrow \infty$.

Now assume without loss of generality, that $l < k$ and $\tau_0 \notin [u_l, u_{l+1}]$. Then,

$$Q^{(k)}(\hat{\boldsymbol{\omega}}_{k,m}) = \sum_{i=1}^{N}\sum_{j=1}^{m}(y_{ij} - \bar{y}_{i,m})^2 + m\sum_{i=1}^{N}(\bar{y}_{i,m} - \mu_i(\hat{\boldsymbol{\omega}}_{k,m}))^2 \text{ and}$$

$$Q^{(l)}(\hat{\boldsymbol{\omega}}_{l,m}) = \sum_{i=1}^{N}\sum_{j=1}^{m}(y_{ij} - \bar{y}_{i,m})^2 + m\sum_{i=1}^{N}(\bar{y}_{i,m} - \mu_i(\hat{\boldsymbol{\omega}}_{l,m}))^2.$$

Since both sums contain $\sum_{i=1}^{N}\sum_{j=1}^{m}(y_{ij} - \bar{y}_{i.m})^2$, it is only necessary to compare

$\widetilde{Q}^{(k)}(\hat{\boldsymbol{\omega}}_{k,m}) = m\sum_{i=1}^{N}(\bar{y}_{i,m} - \mu_i(\hat{\boldsymbol{\omega}}_{k,m}))^2$ and $\widetilde{Q}^{(l)}(\hat{\boldsymbol{\omega}}_{l,m}) = m\sum_{i=1}^{N}(\bar{y}_{i,m} - \mu_i(\hat{\boldsymbol{\omega}}_{l,m}))^2$.
It has already been shown that if $\tau_0 \in [u_k, u_{k+1}]$, then $\hat{\boldsymbol{\omega}}_{k,m} \rightarrow_{a.s.} \boldsymbol{\omega}_0$ as $m \rightarrow \infty$. By definition, $\mu_i(\boldsymbol{\omega}_j)$ is a continuous function because

$$\mu_i(\boldsymbol{\omega}_j) = \begin{cases} \omega_1 + \omega_2 u_i & \text{if } i \leq j \\ \omega_3 + \omega_4 u_i & \text{if } i > j \end{cases}.$$

Then, by the Continuous Mapping Theorem, $\mu_i(\hat{\boldsymbol{\omega}}_{k,m}) \rightarrow_{a.s} \mu_i(\boldsymbol{\omega}_0)$ as $m \rightarrow \infty$ where $\mu_i(\boldsymbol{\omega}_0) = \mu_{0i}$. Since $y_{i1}, \ldots, y_{im}$ are i.i.d with $E(y_{ij}) = \mu_{0i}$ for all $i = 1, \ldots, N$, the SLLN implies that $\bar{y}_{i,m} \rightarrow_{a.s} \mu_{0i}$ as $m \rightarrow \infty$ for all $i = 1, \ldots, N$. Thus, $\sum_{i=1}^{N}(\bar{y}_{i,m} - \mu_i(\hat{\boldsymbol{\omega}}_{k,m}))^2 \rightarrow_{a.s} 0$ as $m \rightarrow \infty$. This means that $\frac{1}{m}\widetilde{Q}^{(k)}(\hat{\boldsymbol{\omega}}_{k,m}) \rightarrow_{a.s} 0$ as $m \rightarrow \infty$.

Now, consider $\frac{1}{m}\widetilde{Q}^{(l)}(\hat{\boldsymbol{\omega}}_{l,m}) = \sum_{i=1}^{N}(\bar{y}_{i,m} - \mu(\hat{\boldsymbol{\omega}}_{l,m}))^2$. It can be shown that $\frac{1}{m}\tilde{Q}^{(l)}(\hat{\boldsymbol{\omega}}_{l,m})$ is bounded away from 0, i.e., that $\frac{1}{m}\tilde{Q}^{(l)}(\hat{\boldsymbol{\omega}}_{l,m}) \rightarrow_{a.s} c > 0$ as $m \rightarrow \infty$ for some constant c. Note that

$$\begin{aligned} \frac{1}{m}\tilde{Q}^{(l)}(\hat{\boldsymbol{\omega}}_{l,m}) &= \| \bar{\boldsymbol{y}}_m - \boldsymbol{Z}_l\hat{\boldsymbol{\omega}}_{l,m} \|^2 \\ &= \| \bar{\boldsymbol{y}}_m - \boldsymbol{Z}_l(\boldsymbol{Z}_l^T\boldsymbol{Z}_l)^{-1}\boldsymbol{Z}_l^T\bar{\boldsymbol{y}}_m \|^2 \\ &= \| (\boldsymbol{I} - \boldsymbol{Z}_l(\boldsymbol{Z}_l^T\boldsymbol{Z}_l)^{-1}\boldsymbol{Z}_l^T)\bar{\boldsymbol{y}}_m \|^2 \\ &= \| (\boldsymbol{I} - \boldsymbol{H}_l)\bar{\boldsymbol{y}}_m \|^2 \end{aligned}$$

where $\boldsymbol{H}_l = \boldsymbol{Z}_l(\boldsymbol{Z}_l^T\boldsymbol{Z}_l)^{-1}\boldsymbol{Z}_l^T$ is an $N \times N$ hat matrix of rank 4 which projects an N-dimensional vector onto the space spanned by the columns of $\boldsymbol{Z}_l$. Then, using the Continuous Mapping Theorem and the fact that $(\boldsymbol{I} - \boldsymbol{H}_l)$ is not random,

$$\tfrac{1}{m}\tilde{Q}^{(l)}(\hat{\boldsymbol{\omega}}_{l,m}) = \| (\boldsymbol{I} - \boldsymbol{H}_l)\bar{\boldsymbol{y}}_m \|^2 \to_{a.s} \| (\boldsymbol{I} - \boldsymbol{H}_l)\boldsymbol{\mu}_0 \|^2 \text{ as } m \to \infty.$$

It can be shown that $\boldsymbol{\mu}_0 \neq \boldsymbol{H}_l\boldsymbol{\mu}_0$ so that $(\boldsymbol{I} - \boldsymbol{H}_l)\boldsymbol{\mu}_0 \neq \boldsymbol{0}$. Since $\boldsymbol{H}_l\boldsymbol{\mu}_0$ is in the space spanned by the columns of $\boldsymbol{Z}_l$,

$$\boldsymbol{H}_l\boldsymbol{\mu}_0 = \boldsymbol{Z}_l\boldsymbol{e} = \begin{bmatrix} e_1 + e_2u_1 \\ \vdots \\ e_1 + e_2u_l \\ e_3 + e_4u_{l+1} \\ \vdots \\ e_3 + e_4u_{k-1} \\ e_3 + e_4u_k \\ e_3 + e_4u_{k+1} \\ \vdots \\ e_3 + e_4u_N \end{bmatrix}$$

for some $\boldsymbol{e} = [e_1, e_2, e_3, e_4]^T$. Note that

$$\frac{(e_3 + e_4u_k) - (e_3 + e_4u_{k-1})}{u_k - u_{k-1}} = e_4 = \frac{(e_3 + e_4u_{k+1}) - (e_3 + e_4u_k)}{u_{k+1} - u_k}.$$

But $\dfrac{\mu_{0k} - \mu_{0,k-1}}{u_k - u_{k-1}} = \beta_0 \neq \beta_0 + \delta_0 = \dfrac{\mu_{0,k+1} - \mu_{0k}}{u_{k+1} - u_k}$, so $\boldsymbol{\mu}_0$ is not in the space spanned by the columns of $\boldsymbol{Z}_l$. □

THEOREM 3.5. *$\hat{\sigma}^2_{k,m} \to_{a.s} \sigma^2_0$ as $m \to \infty$.*

Proof. The previously defined function $\mu_i(\boldsymbol{\omega})$ can be written as

$$\mu_i^{(l)}(\boldsymbol{\omega}) = \begin{cases} \omega_{1l} + \omega_{2l}u_i \text{ if } i \leq l \\ \omega_{3l} + \omega_{4l}u_i \text{ if } i > l \end{cases}.$$

For $i = 1, \ldots, N, \mu_i^{(l)}(\boldsymbol{\omega}) : \mathbb{R}^4 \to \mathbb{R}$ is a continuous function. So the Continuous Mapping Theorem gives

$$\phi_i^{(k)}(\hat{\boldsymbol{\omega}}_{k,m}) \to_{a.s} \phi_i^{(k)}(\boldsymbol{\omega}_0) = \begin{cases} \omega_{01} + \omega_{02}u_i \text{ if } i \leq k \\ \omega_{03} + \omega_{04}u_i \text{ if } i > k \end{cases} = \mu_{0i}$$

for $i = 1, \ldots, N$ as $m \to \infty$. It has been previously established that since $y_{i1}, \ldots, y_{im}$ are i.i.d with $E(y_{ij}) = \mu_{0i}$, the Strong Law of Large Numbers implies

$$\bar{y}_{i,m} \to_{a.s} \mu_{0i} \text{ as } m \to \infty$$

for $i = 1, \ldots, N$. It follows from $\text{var}(y_{ij}) = \sigma_0^2$ and $\text{var}(y_{ij}) = E(y_{ij}^2) - [E(y_{ij})]^2$ that

$$E(y_{ij}^2) = \sigma_0^2 + \mu_{0i}^2$$

for $i = 1, \ldots, N$. The Strong Law of Large Numbers applied to the i.i.d random variables $y_{i1}^2, \ldots, y_{im}^2$ implies that

$$\tfrac{1}{m} \sum_{j=1}^{m} y_{ij}^2 \to_{a.s} \sigma_0^2 + \mu_{0i}^2 \text{ as } m \to \infty$$

for $i = 1, \ldots, N$. Since $\hat{\sigma}_{k,m}^2 = \frac{Q^{(k)}(\hat{\boldsymbol{\omega}}_{k,m})}{n}$, consider the following:

$$Q^{(k)}(\hat{\boldsymbol{\omega}}_{k,m}) = \sum_{i=1}^{N} \sum_{j=1}^{m} (y_{ij} - \bar{y}_{i,m})^2 + m \sum_{i=1}^{N} (\bar{y}_{i,m} - \phi_i^{(k)}(\hat{\boldsymbol{\omega}}_{k,m}))^2$$

$$\begin{aligned}
&= \sum_{i=1}^{N} \sum_{j=1}^{m} (y_{ij}^2 - 2y_{ij}\bar{y}_{i,m} + \bar{y}_{i,m}^2)^2 + m \sum_{i=1}^{N} (\bar{y}_{i,m} - \phi_i^{(k)}(\hat{\boldsymbol{\omega}}_{k,m}))^2 \\
&= \sum_{i=1}^{N} \left[(y_{i1}^2 - 2y_{i1}\bar{y}_{i,m} + \bar{y}_{i,m}^2) + \ldots + (y_{im}^2 - 2y_{im}\bar{y}_{i,m} + \bar{y}_{i,m}^2) \right] \\
&\quad + m \sum_{i=1}^{N} (\bar{y}_{i,m} - \phi_i^{(k)}(\hat{\boldsymbol{\omega}}_{k,m}))^2 \\
&= \sum_{i=1}^{N} \left[(y_{i1}^2 + \ldots + y_{im}^2) - 2\bar{y}_{im}(y_{i1} + \ldots + y_{im}) + m\bar{y}_{im}^2 \right] \\
&\quad + m \sum_{i=1}^{N} (\bar{y}_{i,m} - \phi_i^{(k)}(\hat{\boldsymbol{\omega}}_{k,m}))^2 \\
&= \sum_{i=1}^{N} \left[\sum_{j=1}^{m} y_{ij}^2 - 2m\bar{y}_{im}^2 + m\bar{y}_{im}^2 \right] + m \sum_{i=1}^{N} (\bar{y}_{i,m} - \phi_i^{(k)}(\hat{\boldsymbol{\omega}}_{k,m}))^2 \\
&= \sum_{i=1}^{N} m \left(\frac{1}{m} \sum_{j=1}^{m} y_{ij}^2 - \bar{y}_{im}^2 \right) + m \sum_{i=1}^{N} (\bar{y}_{i,m} - \phi_i^{(k)}(\hat{\boldsymbol{\omega}}_{k,m}))^2 \\
&= m \left\{ \sum_{i=1}^{N} \left(\frac{1}{m} \sum_{j=1}^{m} y_{ij}^2 - \bar{y}_{im}^2 \right) + \sum_{i=1}^{N} (\bar{y}_{i,m} - \phi_i^{(k)}(\hat{\boldsymbol{\omega}}_{k,m}))^2 \right\}.
\end{aligned}$$

So, $\frac{1}{m}Q^{(k)}(\hat{\boldsymbol{\omega}}_{k,m}) = \sum_{i=1}^{N}\left(\frac{1}{m}\sum_{j=1}^{m} y_{ij}^2 - \bar{y}_{im}^2\right) + \sum_{i=1}^{N}(\bar{y}_{i,m} - \phi_i^{(k)}(\hat{\boldsymbol{\omega}}_{k,m}))^2$. Now, define a continuous function $g : \mathbb{R}^{3N} \to \mathbb{R}$ such that

$$g(x_1, \ldots, x_{3N}) = \sum_{i=1}^{N}(x_i - x_{N+i}^2) + \sum_{i=1}^{N}(x_{N+i} - x_{2N+i})^2.$$

Let $x_1, \ldots, x_N$ correspond to $\frac{1}{m}\sum_{j=1}^{m} y_{ij}^2$ for $i = 1, \ldots, N$, respectively, let $x_{N+1}, \ldots, x_{2N}$ correspond to $\bar{y}_{1,m} \ldots, \bar{y}_{N.m}$, respectively, and let $x_{2N+1}, \ldots, x_{3N}$ correspond to $\phi_1^{(k)}(\hat{\boldsymbol{\omega}}_{k.m})$, ..., $\phi_N^{(k)}(\hat{\boldsymbol{\omega}}_{k,m})$, respectively. Then $g(x_1, \ldots, x_{3N}) = \frac{1}{m}Q_k(\hat{\boldsymbol{\omega}}_{k.m})$. As was established, $\frac{1}{m}\sum_{j=1}^{m} y_{ij}^2 \to_{a.s} \sigma_0^2 + \mu_{0i}^2$, $\bar{y}_{i,m} \to_{a.s} \mu_{0i}$, and $\phi_i^{(k)}(\hat{\boldsymbol{\omega}}_{k,m}) \to_{a.s} \mu_{0i}$ as $m \to \infty$ for $i = 1, \ldots, N$. Since g is a continuous function, the Continuous Mapping Theorem can be applied. Thus, as $m \to \infty$,

$$\begin{aligned}
\hat{\sigma}_{k,m}^2 &= \frac{Q_k(\hat{\boldsymbol{\omega}}_{k,m})}{n} \\
&= \frac{Q_k(\hat{\boldsymbol{\omega}}_{k,m})}{Nm} \\
&= \frac{1}{N}\cdot g\left(\frac{1}{m}\sum_{j=1}^{m} y_{ij}^2, \ldots, \frac{1}{m}\sum_{j=1}^{m} y_{Nj}^2, \bar{y}_{1m}, \ldots, \bar{y}_{Nm}, \phi_1^{(k)}(\hat{\boldsymbol{\omega}}_{k,m}), \ldots, \phi_N^{(k)}(\hat{\boldsymbol{\omega}}_{k,m})\right) \\
\to_{a.s} \ & \frac{1}{N}\cdot g(\sigma_0^2 + \mu_{01}^2, \ldots, \sigma_0^2 + \mu_{0N}^2, \mu_{01}, \ldots, \mu_{0N}, \mu_{01}, \ldots, \mu_{0N}) \\
&= \frac{1}{N}\left(\sum_{i=1}^{N}(\sigma_0^2 + \mu_{0i}^2 - \mu_{0i}^2) + \sum_{i=1}^{N}(\mu_{0i}^2 - \mu_{0i}^2)\right) \\
&= \frac{1}{N}\sum_{i=1}^{N}\sigma_0^2 \\
&= \frac{1}{N}\cdot N \cdot \sigma_0^2 \\
&= \sigma_0^2
\end{aligned}$$

Therefore, $\hat{\sigma}_{k,m}^2 \to_{a.s} \sigma_0^2$ as $m \to \infty$. □

3.2 Asymptotic normality

Now that strong consistency has been established, it is useful to consider

whether the estimator of interest, $\hat{\tau}$, is asymptotically normal, i.e., if the distribution of $\hat{\tau}$ near τ_0 approaches a normal distribution as the sample size increases. This is a useful property for the estimator to have. Since the actual distribution of the estimator $\hat{\tau}$ is unknown, asymptotic normality suggests that as the sample size gets large, the distribution of the estimator can be approximated by the normal distribution. This allows the properties of the normal distribution to be used when working with $\hat{\tau}$ and large sample sizes.

As has been the case when dealing with a segmented regression model with a continuity constraint, the asymptotic behavior of the least squares estimator, $\hat{\tau}$, depends on the location of τ_0. For this reason, the asymptotic normality will be dealt with in two cases. The first case is where the true change occurs between two observations and the second case is where the true change occurs at an observation. For both cases, define

$$\psi(\boldsymbol{\omega}) = \begin{bmatrix} \alpha \\ \beta \\ \delta \\ \tau \end{bmatrix} = \begin{bmatrix} \omega_1 \\ \omega_2 \\ \omega_4 - \omega_2 \\ \frac{\omega_1 - \omega_3}{\omega_4 - \omega_2} \end{bmatrix}$$

and

$$\boldsymbol{V}_{a,b} = \begin{bmatrix} b-a & \sum_{i=a+1}^{b} u_i \\ \sum_{i=a+1}^{b} u_i & \sum_{i=a+1}^{b} u_i^2 \end{bmatrix}.$$

The following standard result from probability (see Basu, 1999) is also used for both cases.

Cramér-Wold Criterion: A random vector $\boldsymbol{X}_n \rightarrow_d \boldsymbol{X}$ iff $\boldsymbol{b}^T \boldsymbol{X}_n \rightarrow_d \boldsymbol{b}^T \boldsymbol{X}$ for all $\boldsymbol{b}$.

3.2.1 Case 1

THEOREM 3.6. *For the clustered segmented regression model with the continuity constraint, if* $\tau_0 \in (u_k, u_{k+1})$ *for some* k, *then*

$$\sqrt{m}(\hat{\tau} - \tau_0) \rightarrow_d \mathcal{N}(\mathbf{0}, \sigma_0^2\ \tilde{\boldsymbol{a}}^T\ \widetilde{\boldsymbol{\Sigma}}\ \tilde{\boldsymbol{a}})\ as\ m \rightarrow \infty$$

where

$$\widetilde{\boldsymbol{\Sigma}} = \begin{bmatrix} \boldsymbol{V}_{1,k}^{-1} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{V}_{k+1,N}^{-1} \end{bmatrix}$$

and

$$\tilde{\boldsymbol{a}} = \frac{\partial \psi_{k,4}(\boldsymbol{\omega}_0)}{\partial \boldsymbol{\omega}_0} = \begin{bmatrix} \frac{1}{\omega_{04}-\omega_{02}} & \frac{\omega_{01}-\omega_{03}}{(\omega_{02}-\omega_{04})^2} & \frac{-1}{\omega_{04}-\omega_{02}} & \frac{-(\omega_{01}-\omega_{03})}{(\omega_{02}-\omega_{04})^2} \end{bmatrix}^T.$$

Notation: For any distributions **A** and **B**, $\mathbf{A} \sim \mathbf{B}$ means that **A** and **B** have the same distribution.

Proof. As previously defined, $\boldsymbol{\mu}_l = \boldsymbol{Z}_l\boldsymbol{\omega}_l$ for any $l = 1, \ldots, N$. Then $\boldsymbol{\omega}_l = (\boldsymbol{Z}_l^T\boldsymbol{Z}_l)^{-1}\boldsymbol{Z}_l^T\boldsymbol{\mu}$. Define a function $\boldsymbol{g}_l(\boldsymbol{\mu}) = (\boldsymbol{Z}_l^T\boldsymbol{Z}_l)^{-1}\boldsymbol{Z}_l^T\boldsymbol{\mu}$. Then, $\boldsymbol{g}_k(\boldsymbol{\mu}_0) = \boldsymbol{g}(\boldsymbol{\mu}_0) = \boldsymbol{\omega}_0$ and $\boldsymbol{g}_k(\hat{\boldsymbol{\mu}}_{k,m}) = \boldsymbol{g}(\hat{\boldsymbol{\mu}}_k) = \hat{\boldsymbol{\omega}}_{k,m}$. Using this notation, $\sqrt{m}(\hat{\boldsymbol{\omega}}_{k,m} - \boldsymbol{\omega}_0) = \sqrt{m}(\boldsymbol{g}(\hat{\boldsymbol{\mu}}_{k,m}) - \boldsymbol{g}(\boldsymbol{\mu}_0))$.

Let $\boldsymbol{b} \in \mathbb{R}^4$ be any 4x1 vector. Since $\boldsymbol{g}$ is infinitely differentiable with respect to $\boldsymbol{\mu}$, it can be represented with a Taylor series expansion. Letting $\boldsymbol{f}(\boldsymbol{\mu}) = \boldsymbol{b}^T\boldsymbol{g}(\boldsymbol{\mu})$,

$$\sqrt{m}\cdot\boldsymbol{b}^T(\boldsymbol{g}(\hat{\boldsymbol{\mu}}_k)-\boldsymbol{g}(\boldsymbol{\mu}_0)) = \sqrt{m}\cdot[\nabla\boldsymbol{f}(\boldsymbol{\mu}_0)]^T(\hat{\boldsymbol{\mu}}_k-\boldsymbol{\mu}_0)+\frac{\sqrt{m}}{2}(\hat{\boldsymbol{\mu}}_k-\boldsymbol{\mu}_0)^T[\nabla^2\boldsymbol{f}(\tilde{\boldsymbol{\mu}})](\hat{\boldsymbol{\mu}}_k-\boldsymbol{\mu}_0).$$

where $\tilde{\boldsymbol{\mu}}$ lies somewhere on the line joining $\hat{\boldsymbol{\mu}}_k$ and $\boldsymbol{\mu}_0$ and

$$\begin{aligned} \nabla\boldsymbol{f}(\boldsymbol{\mu}) = \boldsymbol{b}^T\frac{\partial\boldsymbol{g}}{\partial\boldsymbol{\mu}} &= [b_1, b_2, b_3, b_4]\begin{bmatrix} \frac{\partial\boldsymbol{g}_1}{\partial\boldsymbol{\mu}} \\ \frac{\partial\boldsymbol{g}_2}{\partial\boldsymbol{\mu}} \\ \frac{\partial\boldsymbol{g}_3}{\partial\boldsymbol{\mu}} \\ \frac{\partial\boldsymbol{g}_4}{\partial\boldsymbol{\mu}} \end{bmatrix}. \\ &= b_1\frac{\partial\boldsymbol{g}}{\partial\boldsymbol{\mu}_1} + b_2\frac{\partial\boldsymbol{g}}{\partial\boldsymbol{\mu}_2} + b_3\frac{\partial\boldsymbol{g}}{\partial\boldsymbol{\mu}_3} + b_4\frac{\partial\boldsymbol{g}}{\partial\boldsymbol{\mu}_4} \\ &= \frac{\partial\boldsymbol{g}}{\partial\boldsymbol{\mu}^T}\boldsymbol{b}. \end{aligned}$$

Thus, $\left[\nabla \boldsymbol{f}(\boldsymbol{\mu}_0)\right]^T = \boldsymbol{b}^T \frac{\partial \boldsymbol{g}}{\partial \boldsymbol{\mu}_0}$ and $\left[\nabla^2 \boldsymbol{f}(\boldsymbol{\mu})\right] = \frac{\partial}{\partial \boldsymbol{\mu}}\left[(\nabla \boldsymbol{f}(\boldsymbol{\mu}))^T\right] = \frac{\partial}{\partial \boldsymbol{\mu}}\left[\boldsymbol{b}^T \frac{\partial \boldsymbol{g}}{\partial \boldsymbol{\mu}}\right]$.
Letting $\boldsymbol{G}(\boldsymbol{\mu}) = \frac{\partial \boldsymbol{g}}{\partial \boldsymbol{\mu}} = (\boldsymbol{Z}_k^T \boldsymbol{Z}_k)^{-1} \boldsymbol{Z}_k^T$ gives

$$\begin{aligned}
[\nabla^2 \boldsymbol{f}(\tilde{\boldsymbol{\mu}})] &= \frac{\partial}{\partial \tilde{\boldsymbol{\mu}}}[\boldsymbol{b}^T \boldsymbol{G}(\tilde{\boldsymbol{\mu}})] \\
&= \frac{\partial}{\partial \tilde{\boldsymbol{\mu}}} \boldsymbol{b}^T (\boldsymbol{Z}_k^T \boldsymbol{Z}_k)^{-1} \boldsymbol{Z}_k^T \\
&= \mathbf{0}.
\end{aligned}$$

By substitution back into the Taylor expansion,

$$\sqrt{m} \cdot \boldsymbol{b}^T(\boldsymbol{g}(\hat{\boldsymbol{\mu}}_k) - \boldsymbol{g}(\boldsymbol{\mu}_0)) = \sqrt{m} \cdot \boldsymbol{b}^T \boldsymbol{G}(\boldsymbol{\mu}_0)(\hat{\boldsymbol{\mu}}_k - \boldsymbol{\mu}_0)$$

where $\boldsymbol{b}^T \boldsymbol{G}(\boldsymbol{\mu}_0)$ is fixed and not random. The Central Limit Theorem implies $\sqrt{m}(\hat{\boldsymbol{\mu}}_k - \boldsymbol{\mu}_0) \to_d \mathcal{N}(\mathbf{0}, \sigma_0^2 \mathbf{I}_4)$. Then,

$$\begin{aligned}
\boldsymbol{b}^T \boldsymbol{G}(\boldsymbol{\mu}_0)[\sqrt{m}(\hat{\boldsymbol{\mu}}_k - \boldsymbol{\mu}_0)] &\to_d \mathcal{N}(\mathbf{0}, \boldsymbol{b}^T \boldsymbol{G}(\boldsymbol{\mu}_0) \sigma_0^2 (\boldsymbol{G}(\boldsymbol{\mu}_0))^T \boldsymbol{b}) \\
&\sim \mathcal{N}(\mathbf{0}, \sigma_0^2 \boldsymbol{b}^T \boldsymbol{G}\boldsymbol{G}^T \boldsymbol{b})
\end{aligned}$$

where

$$\begin{aligned}
\boldsymbol{G}\boldsymbol{G}^T &= (\boldsymbol{Z}_k^T \boldsymbol{Z}_k)^{-1} \boldsymbol{Z}_k^T \boldsymbol{Z}_k (\boldsymbol{Z}_k^T \boldsymbol{Z}_k)^{-1} \\
&= (\boldsymbol{Z}_k^T \boldsymbol{Z}_k)^{-1} \\
&= \begin{bmatrix} (\boldsymbol{V}_{1.k})^{-1} & \boldsymbol{O} \\ \boldsymbol{O} & (\boldsymbol{V}_{k+1.n})^{-1} \end{bmatrix} \\
&= \widetilde{\boldsymbol{\Sigma}}.
\end{aligned}$$

It follows that

$$\begin{aligned}
\boldsymbol{b}^T \sqrt{m}(\boldsymbol{g}(\hat{\boldsymbol{\mu}}_k) - \boldsymbol{g}(\boldsymbol{\mu}_0)) = \boldsymbol{b}^T \boldsymbol{G}(\boldsymbol{\mu}_0)[\sqrt{m}(\hat{\boldsymbol{\mu}}_k - \boldsymbol{\mu}_0)] &\to_d \mathcal{N}(\mathbf{0}, \boldsymbol{b}^T \sigma_0^2 (\boldsymbol{Z}_k^T \boldsymbol{Z}_k)^{-1} \boldsymbol{b}) \\
&\sim \mathcal{N}(\mathbf{0}, \boldsymbol{b}^T \sigma_0^2 \widetilde{\boldsymbol{\Sigma}} \boldsymbol{b}).
\end{aligned}$$

The Cramér-Wold device implies $\sqrt{m}(\boldsymbol{g}(\hat{\boldsymbol{\mu}}_k) - \boldsymbol{g}(\boldsymbol{\mu}_0)) \to_d \mathcal{N}(\mathbf{0}, \sigma_0^2 \widetilde{\boldsymbol{\Sigma}})$. Therefore, by definition, $\sqrt{m}(\hat{\boldsymbol{\omega}}_{k.m} - \boldsymbol{\omega}_0) \to_d \mathcal{N}(\mathbf{0}, \sigma_0^2 \widetilde{\boldsymbol{\Sigma}})$. Since $\tau = \psi_{k,4}(\boldsymbol{\omega}) = \frac{\omega_3 - \omega_1}{\omega_2 - \omega_4}$, the Delta Method implies

$$\sqrt{m}(\psi_{k,4}(\hat{\boldsymbol{\omega}}_k) - \psi_{k,4}(\boldsymbol{\omega}_0)) \to_d \mathcal{N}(\mathbf{0}, \sigma_0^2\ (\nabla\psi_{k,4}(\boldsymbol{\omega}_0)^T\ \widetilde{\boldsymbol{\Sigma}}\ (\nabla\psi_{k,4}(\boldsymbol{\omega}_0)))$$

which means

$$\sqrt{m}(\hat{\tau}_k - \tau_0) \to_d \mathcal{N}(\mathbf{0}, \sigma_0^2\ \check{\boldsymbol{a}}^T\ \widetilde{\boldsymbol{\Sigma}}\ \check{\boldsymbol{a}}).$$

□

3.2.2 Case 2

THEOREM 3.7. *For the clustered segmented regression model with the continuity constraint, if* $\tau_0 = u_k$ *for some* k, *then*

$$\sqrt{m}\left(\begin{bmatrix} \hat{\tau}_{k-1} \\ \hat{\tau}_k \end{bmatrix} - \begin{bmatrix} \tau_0 \\ \tau_0 \end{bmatrix} \right) \to_d \mathcal{N}(\mathbf{0}, \sigma_0^2\ (\boldsymbol{a}^\star)^T\ \boldsymbol{\Sigma}^\star\ \boldsymbol{a}^\star)$$

where

$$\boldsymbol{\Sigma}^\star = \begin{bmatrix} \boldsymbol{V}_{1,k-1}^{-1} & \boldsymbol{O} & \boldsymbol{V}_{1,k}^{-1} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{V}_{k,N}^{-1} & \boldsymbol{V}_{k,N}^{-1}\boldsymbol{V}_{k,k}\boldsymbol{V}_{1,k}^{-1} & \boldsymbol{V}_{k,N}^{-1} \\ \boldsymbol{V}_{1,k}^{-1} & \boldsymbol{V}_{1,k}^{-1}\boldsymbol{V}_{k,k}\boldsymbol{V}_{k,N}^{-1} & \boldsymbol{V}_{1,k}^{-1} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{V}_{k,N}^{-1} & \boldsymbol{O} & \boldsymbol{V}_{k+1,N}^{-1} \end{bmatrix}.$$

and $\boldsymbol{a}^\star = \frac{\partial\psi^\star(\boldsymbol{\omega}_0,\boldsymbol{\omega}_0)}{\partial(\boldsymbol{\omega}_{k-1}^T,\boldsymbol{\omega}_k^T)^T}$ *where* $\psi^\star(\boldsymbol{\omega}_\ell, \boldsymbol{\omega}_j) = \begin{bmatrix} \frac{\omega_{\ell 3}-\omega_{\ell 1}}{\omega_{\ell 2}-\omega_{\ell 4}} & \frac{\omega_{j3}-\omega_{j1}}{\omega_{j2}-\omega_{j4}} \end{bmatrix}$.

The proof for this theorem is similar to that of Theorem 3.6, except that the joint distribution of $\hat{\tau}_{k-1}$ and $\hat{\tau}_k$ must be dealt with for this case.

Proof. Define a function $\boldsymbol{g} : \mathbb{R}^N \to \mathbb{R}^8$ as

$$\boldsymbol{g}(\boldsymbol{\mu}) = \begin{bmatrix} (\boldsymbol{Z}_{k-1}^T\boldsymbol{Z}_{k-1})^{-1}\boldsymbol{Z}_{k-1}^T\boldsymbol{\mu} \\ (\boldsymbol{Z}_k^T\boldsymbol{Z}_k)^{-1}\boldsymbol{Z}_k^T\boldsymbol{\mu} \end{bmatrix}.$$

Then, defining $\boldsymbol{G}(\boldsymbol{\mu}) = \frac{\partial \boldsymbol{g}}{\partial \boldsymbol{\mu}^T}$ gives

$$\boldsymbol{G}(\boldsymbol{\mu}) = \begin{bmatrix} (\boldsymbol{Z}_{k-1}^T\boldsymbol{Z}_{k-1})^{-1}\boldsymbol{Z}_{k-1}^T \\ (\boldsymbol{Z}_k^T\boldsymbol{Z}_k)^{-1}\boldsymbol{Z}_k^T \end{bmatrix}.$$

Next, use a Taylor series expansion, the Central Limit Theorem, and the fact that $\frac{\partial(\boldsymbol{b}^T\boldsymbol{g})^2}{\partial\boldsymbol{\mu}\partial\boldsymbol{\mu}^T} = \mathbf{0}$ to write

$$\begin{aligned}
\sqrt{m}\cdot\boldsymbol{b}^T(\boldsymbol{g}(\hat{\boldsymbol{\mu}}_m)-\boldsymbol{g}(\boldsymbol{\mu}_0)) &= \sqrt{m}\boldsymbol{b}^T\boldsymbol{G}(\boldsymbol{\mu}_0)(\hat{\boldsymbol{\mu}}_m-\boldsymbol{\mu}_0)\\
&= \boldsymbol{b}^T\boldsymbol{G}(\boldsymbol{\mu}_0)[\sqrt{m}(\hat{\boldsymbol{\mu}}_m-\boldsymbol{\mu}_0)]\\
&\to_d \mathcal{N}(\mathbf{0},\boldsymbol{b}^T\boldsymbol{G}(\boldsymbol{\mu}_0)\sigma_0^2\boldsymbol{G}(\boldsymbol{\mu}_0)^T\boldsymbol{b})\\
&\sim \mathcal{N}(\mathbf{0},\sigma_0^2\boldsymbol{b}^T\boldsymbol{G}(\boldsymbol{\mu}_0)\boldsymbol{G}(\boldsymbol{\mu}_0)^T\boldsymbol{b})\\
&\sim \mathcal{N}(\mathbf{0},\sigma_0^2\boldsymbol{b}^T\boldsymbol{\Sigma}^\star\boldsymbol{b}).
\end{aligned}$$

The Cramér-Wold device yields

$$\sqrt{m}(\boldsymbol{g}(\hat{\boldsymbol{\mu}}_m)-\boldsymbol{g}(\boldsymbol{\mu}_0))\to_d\mathcal{N}(\mathbf{0},\sigma_0^2\boldsymbol{\Sigma}^\star),$$

which implies

$$\sqrt{m}\left(\begin{bmatrix}\hat{\boldsymbol{\omega}}_{k-1}\\ \hat{\boldsymbol{\omega}}_k\end{bmatrix}-\begin{bmatrix}\boldsymbol{\omega}_0\\ \boldsymbol{\omega}_0\end{bmatrix}\right)\to_d\mathcal{N}(\mathbf{0},\sigma_0^2\boldsymbol{\Sigma}^\star).$$

The above argument is verified by showing that $\boldsymbol{G}(\boldsymbol{\mu}_0)\boldsymbol{G}(\boldsymbol{\mu}_0)^T=\boldsymbol{\Sigma}^\star$.

$$\boldsymbol{G}(\boldsymbol{\mu}_0)\boldsymbol{G}(\boldsymbol{\mu}_0)^T=\begin{bmatrix}(\boldsymbol{Z}_{k-1}^T\boldsymbol{Z}_{k-1})^{-1}\boldsymbol{Z}_{k-1}^T\\ \\ (\boldsymbol{Z}_k^T\boldsymbol{Z}_k)^{-1}\boldsymbol{Z}_k^T\end{bmatrix}\begin{bmatrix}\boldsymbol{Z}_{k-1}(\boldsymbol{Z}_{k-1}^T\boldsymbol{Z}_{k-1})^{-1} & \boldsymbol{Z}_k(\boldsymbol{Z}_k^T\boldsymbol{Z}_k)^{-1}\end{bmatrix}$$

$$=\begin{bmatrix}(\boldsymbol{Z}_{k-1}^T\boldsymbol{Z}_{k-1})^{-1} & (\boldsymbol{Z}_{k-1}^T\boldsymbol{Z}_{k-1})^{-1}\boldsymbol{Z}_{k-1}^T\boldsymbol{Z}_k(\boldsymbol{Z}_k^T\boldsymbol{Z}_k)^{-1}\\ \\ (\boldsymbol{Z}_k^T\boldsymbol{Z}_k)^{-1}\boldsymbol{Z}_k^T\boldsymbol{Z}_{k-1}(\boldsymbol{Z}_{k-1}^T\boldsymbol{Z}_{k-1})^{-1} & (\boldsymbol{Z}_k^T\boldsymbol{Z}_k)^{-1}\end{bmatrix}.$$

As in Theorem 3.6, the first and fourth elements of this matrix can be written as follows:

$$(\boldsymbol{Z}_{k-1}^T\boldsymbol{Z}_{k-1})^{-1}=\begin{bmatrix}\boldsymbol{V}_{1,k-1}^{-1} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{V}_{k,N}^{-1}\end{bmatrix}\quad\text{and}\quad(\boldsymbol{Z}_k^T\boldsymbol{Z}_k)^{-1}=\begin{bmatrix}\boldsymbol{V}_{1,k}^{-1} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{V}_{k+1,N}^{-1}\end{bmatrix}.$$

Consider the second element $(\boldsymbol{Z}_{k-1}^T\boldsymbol{Z}_{k-1})^{-1}\boldsymbol{Z}_{k-1}^T\boldsymbol{Z}_k(\boldsymbol{Z}_k^T\boldsymbol{Z}_k)^{-1}$:

$$
\begin{aligned}
\boldsymbol{Z}_{k-1}^T\boldsymbol{Z}_k &= \begin{bmatrix} 1 & \dots & 1 & 0 & \dots & 0 \\ u_1 & \dots & u_{k-1} & 0 & \dots & 0 \\ 0 & \dots & 0 & 1 & \dots & 1 \\ 0 & \dots & 0 & u_k & \dots & u_N \end{bmatrix} \begin{bmatrix} 1 & u_1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & u_k & 0 & 0 \\ 0 & 0 & 1 & u_{k+1} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & u_N \end{bmatrix} \\
&= \begin{bmatrix} k-1 & \sum_{i=1}^{k-1} u_i & 0 & 0 \\ \sum_{i=1}^{k-1} u_i & \sum_{i=1}^{k-1} u_i^2 & 0 & 0 \\ 1 & u_k & N-k & \sum_{i=k+1}^{N} u_i \\ u_k & u_k^2 & \sum_{i=k+1}^{N} u_i & \sum_{i=k+1}^{N} u_i^2 \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{V}_{1,k-1} & \boldsymbol{O} \\ \boldsymbol{V}_{k,k} & \boldsymbol{V}_{k+1,N} \end{bmatrix}.
\end{aligned}
$$

Then,

$$
\begin{aligned}
&(\boldsymbol{Z}_{k-1}^T\boldsymbol{Z}_{k-1})^{-1}\boldsymbol{Z}_{k-1}^T\boldsymbol{Z}_k(\boldsymbol{Z}_k^T\boldsymbol{Z}_k)^{-1} = \\
&\quad \begin{bmatrix} \boldsymbol{V}_{1,k-1}^{-1} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{V}_{k,N}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{V}_{1,k-1} & \boldsymbol{O} \\ \boldsymbol{V}_{k,k} & \boldsymbol{V}_{k+1,N} \end{bmatrix} \begin{bmatrix} \boldsymbol{V}_{1,k}^{-1} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{V}_{k+1,N}^{-1} \end{bmatrix}
\end{aligned}
$$

$$= \begin{bmatrix} \mathbf{I} & \boldsymbol{O} \\ \boldsymbol{V}_{k,N}^{-1}\boldsymbol{V}_{k,k} & \boldsymbol{V}_{k,N}^{-1}\boldsymbol{V}_{k+1,N} \end{bmatrix} \begin{bmatrix} \boldsymbol{V}_{1,k}^{-1} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{V}_{k+1,N}^{-1} \end{bmatrix}$$

$$= \begin{bmatrix} \boldsymbol{V}_{1,k}^{-1} & \boldsymbol{O} \\ \boldsymbol{V}_{k,N}^{-1}\boldsymbol{V}_{k,k}\boldsymbol{V}_{1,k}^{-1} & \boldsymbol{V}_{k,N}^{-1} \end{bmatrix}.$$

Likewise, the third component $(\boldsymbol{Z}_k^T\boldsymbol{Z}_k)^{-1}\boldsymbol{Z}_k^T\boldsymbol{Z}_{k-1}(\boldsymbol{Z}_{k-1}^T\boldsymbol{Z}_{k-1})^{-1}$ can be simplified as follows;

$$\boldsymbol{Z}_k^T\boldsymbol{Z}_{k-1} = \begin{bmatrix} 1 & \dots & 1 & 0 & \dots & 0 \\ u_1 & \dots & u_k & 0 & \dots & 0 \\ 0 & \dots & 0 & 1 & \dots & 1 \\ 0 & \dots & 0 & u_{k+1} & \dots & u_N \end{bmatrix} \begin{bmatrix} 1 & u_1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & u_{k-1} & 0 & 0 \\ 0 & 0 & 1 & u_k \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & u_N \end{bmatrix}$$

$$= \begin{bmatrix} k-1 & \sum_{i=1}^{k-1} u_i & 1 & u_k \\ \sum_{i=1}^{k-1} u_i & \sum_{i=1}^{k-1} u_i^2 & u_k & u_k^2 \\ 0 & 0 & N-k & \sum_{i=k+1}^{N} u_i \\ 0 & 0 & \sum_{i=k+1}^{N} u_i & \sum_{i-k+1}^{N} u_i^2 \end{bmatrix}$$

$$= \begin{bmatrix} \boldsymbol{V}_{1,k-1} & \boldsymbol{V}_{k,k} \\ \boldsymbol{O} & \boldsymbol{V}_{k+1,N} \end{bmatrix}$$

and

$$(\boldsymbol{Z}_k^T\boldsymbol{Z}_k)^{-1}\boldsymbol{Z}_k^T\boldsymbol{Z}_{k-1}(\boldsymbol{Z}_{k-1}^T\boldsymbol{Z}_{k-1})^{-1}$$

$$= \begin{bmatrix} \boldsymbol{V}_{1.k}^{-1} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{V}_{k+1.N}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{V}_{1.k-1} & \boldsymbol{V}_{k.k} \\ \boldsymbol{O} & \boldsymbol{V}_{k+1.N} \end{bmatrix} \begin{bmatrix} \boldsymbol{V}_{1,k-1}^{-1} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{V}_{k,N}^{-1} \end{bmatrix}$$

$$= \begin{bmatrix} \boldsymbol{V}_{1,k}^{-1} & \boldsymbol{V}_{1,k}^{-1}\boldsymbol{V}_{k,k}\boldsymbol{V}_{k,N}^{-1} \\ \\ \boldsymbol{O} & \boldsymbol{V}_{k,N}^{-1} \end{bmatrix}.$$

Thus, it follows that

$$\boldsymbol{G}(\boldsymbol{\mu}_0)\boldsymbol{G}(\boldsymbol{\mu}_0)^T = \begin{bmatrix} \boldsymbol{V}_{1:k-1}^{-1} & \boldsymbol{O} & \boldsymbol{V}_{1:k}^{-1} & \boldsymbol{O} \\ \\ \boldsymbol{O} & \boldsymbol{V}_{k:N}^{-1} & \boldsymbol{V}_{k:N}^{-1}\boldsymbol{V}_{k:k}\boldsymbol{V}_{1:k}^{-1} & \boldsymbol{V}_{k:N}^{-1} \\ \\ \boldsymbol{V}_{1:k}^{-1} & \boldsymbol{V}_{1:k}^{-1}\boldsymbol{V}_{k:k}\boldsymbol{V}_{k:N}^{-1} & \boldsymbol{V}_{1:k}^{-1} & \boldsymbol{O} \\ \\ \boldsymbol{O} & \boldsymbol{V}_{k:N}^{-1} & \boldsymbol{O} & \boldsymbol{V}_{k+1:N}^{-1} \end{bmatrix} = \boldsymbol{\Sigma}^\star,$$

where $\boldsymbol{\Sigma}^\star$ is the covariance matrix for the joint distribution of $\hat{\boldsymbol{\omega}}_{k-1}$ and $\hat{\boldsymbol{\omega}}_k$.

Finally, since

$$\sqrt{m}\left(\begin{bmatrix} \hat{\boldsymbol{\omega}}_{k-1} \\ \hat{\boldsymbol{\omega}}_k \end{bmatrix} - \begin{bmatrix} \boldsymbol{\omega}_0 \\ \boldsymbol{\omega}_0 \end{bmatrix} \right) \rightarrow_d \mathcal{N}(\mathbf{0}, \sigma_0^2\boldsymbol{\Sigma}^\star),$$

the Delta Method implies

$$\sqrt{m}\left(\psi\left(\begin{bmatrix} \hat{\boldsymbol{\omega}}_{k-1} \\ \hat{\boldsymbol{\omega}}_k \end{bmatrix} \right) - \psi\left(\begin{bmatrix} \boldsymbol{\omega}_0 \\ \boldsymbol{\omega}_0 \end{bmatrix} \right) \right) \rightarrow_d$$

$$\mathcal{N}\left(\mathbf{0}, \sigma_0^2 \left(\nabla\psi\left(\begin{bmatrix} \boldsymbol{\omega}_0 \\ \boldsymbol{\omega}_0 \end{bmatrix} \right) \right)^T \boldsymbol{\Sigma}^\star \left(\nabla\psi\left(\begin{bmatrix} \boldsymbol{\omega}_0 \\ \boldsymbol{\omega}_0 \end{bmatrix} \right) \right) \right).$$

Thus,

$$\sqrt{m}\left(\begin{bmatrix} \hat{\tau}_{k-1} \\ \hat{\tau}_k \end{bmatrix} - \begin{bmatrix} \tau_0 \\ \tau_0 \end{bmatrix} \right) \rightarrow_d \mathcal{N}(\mathbf{0}, \sigma_0^2 \ (\boldsymbol{a}^\star)^T \ \boldsymbol{\Sigma}^\star \ \boldsymbol{a}^\star).$$

□

CHAPTER 4
CONFIDENCE ESTIMATION

In order to construct a confidence interval based on the asymptotic theory, we need to have an algorithm that works regardless of where τ_0 is located. In other words, our algorithm should work for both

Case 1: $\tau_0 \in (u_k, u_{k+1})$ for some $k = 2, \ldots, N-2$ and

Case 2: $\tau_0 = u_k$ for some $k = 2, \ldots, N-1$.

Before discussing the following algorithms, it must be shown that parameters can be replaced by consistent estimators in proofs and statements related to confidence estimation. In particular, since the true variance, σ_0^2, in the asymptotic theory described above is unknown, it is desirable to use the estimated variance, $\hat{\sigma}^2$, in the construction of confidence intervals for the estimated changepoint. The following lemma allows this, i.e., it shows that parameters can be replaced by consistent estimators in regards to confidence estimation.

Lemma 4.1. *Suppose T_n is a consistent and asymptotically normal estimator of T. In other words, suppose $T_n \to_p T$ as $n \to \infty$ and $\sqrt{n}(T_n - T) \to_d \mathcal{N}(0, U)$ as $n \to \infty$ for some constant U. If U_n is any consistent estimator of U, then*

$$\left(\frac{T_n - T}{\sqrt{\frac{U_n}{n}}}\right) \to_d \mathcal{N}(0.1).$$

Proof. Since T_n is consistent and asymptotically normal, the following hold:

$$T_n \to_p T \text{ as } n \to \infty \tag{4.1}$$

$$\sqrt{n}(T_n - T) \to_d \mathcal{N}(0, U) \text{ as } n \to \infty. \tag{4.2}$$

Statement (4.2) implies that $\frac{T_n - T}{\sqrt{\frac{U}{n}}} \rightarrow_d Z$ as $n \rightarrow \infty$ where Z is standard normal. Then, using Slutsky's Theorem and the Continuous Mapping Theorem,

$$\begin{aligned}
\frac{T_n - T}{\sqrt{\frac{U_n}{n}}} = \frac{\sqrt{n}(T_n - T)}{\sqrt{U_n}} &= \frac{\sqrt{U}}{\sqrt{U}} \cdot \frac{\sqrt{n}(T_n - T)}{\sqrt{U_n}} \\
&= \frac{\sqrt{U}}{\sqrt{U_n}} \cdot \frac{\sqrt{n}(T_n - T)}{\sqrt{U}} \\
&\rightarrow_d \frac{\sqrt{U}}{\sqrt{U}} \cdot Z \\
&= Z.
\end{aligned}$$

Thus, $\frac{T_n - T}{\sqrt{\frac{U_n}{n}}} \rightarrow_d \mathcal{N}(0,1)$ as $n \rightarrow \infty$. □

In regards to the clustered model, since $\hat{\sigma}^2 \rightarrow_{a.s.} \sigma_0^2$ as $m \rightarrow \infty$ by Theorem 3.5, it follows that $\hat{\sigma}^2 \rightarrow_p \sigma_0^2$ as $m \rightarrow \infty$, as convergence in probability is implied by almost sure convergence. Putting this together with Theorem 3.6 and Theorem 3.7, confidence intervals for this model can be constructed using $\hat{\sigma}^2$ in place of σ_0^2.

4.1 Removal algorithm

One way to deal with the two cases is to always force Case 1. So, regardless of whether τ_0 is located at an observation (Case 2) or between two observations (Case 1), the removal algorithm deletes the appropriate observation to guarantee that τ_0 is between two observations. Thus, for the purpose of confidence estimation, even if τ_0 is truly at an observation, the removal of that observation forces τ_0 between two observations. Then the calculations only require the use of asymptotic normality of Case 1, i.e., Theorem 3.6.

The following removal algorithm gives a method for constructing a $100(1 - \alpha)\%$ confidence interval for the estimator of the change point, $\hat{\tau}$. The advantage of the removal algorithm is that it works regardless of the location of the true change point, τ_0. In other words, it works for both Case 1 and Case 2. In particular, it

works for the case when $\tau_0 = u_k$ for some $k = 2, \ldots, N-1$ (Case 2) without dealing with the complicated joint distribution of $\hat{\tau}_{k-1}$ and $\hat{\tau}_k$. The disadvantage of the removal algorithm is that by throwing away an observation, an increasing amount of data is lost. This affects the efficiency of the algorithm.

Removal Algorithm

- Compute $\hat{\boldsymbol{\theta}}$. This is a consistent estimator of $\boldsymbol{\theta}_0$, so if m is sufficiently large, $\hat{\tau}$ will be close to τ_0.

- Let u_k by the observation point that is closest to $\hat{\tau}$. Then for sufficiently large m, u_k will be the observation point that is closest to τ_0.

- Remove the observations at u_k and re-compute the MLE of $\boldsymbol{\theta}_0$, namely, $\hat{\boldsymbol{\theta}}^- = (\hat{\alpha}^-, \hat{\beta}^-, \hat{\delta}^-, \hat{\tau}^-)^T$.

- Compute the $100(1-\alpha)\%$ confidence interval, $\mathcal{I}_{1-\alpha}$, based on Theorem 3.6:

$$\hat{\tau}^- \pm z_{\frac{\alpha}{2}} \sqrt{\frac{(\hat{\sigma}^-)^2 \, \tilde{\boldsymbol{a}}^T \, \widetilde{\Sigma}^- \, \tilde{\boldsymbol{a}}}{m}}$$

  where z_α is the $100(1-\alpha)$th percentile of a standard normal distribution and

$$\widetilde{\Sigma}^- = \begin{bmatrix} \boldsymbol{V}_{1.k-1}^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{V}_{k+1.N}^{-1} \end{bmatrix}.$$

4.2 Comprehensive algorithm

The advantage of this algorithm over the removal algorithm is that there is no loss of information and it works regardless of the location of the true change. The interval chosen for the estimator and confidence interval for each of the 9 scenarios stated below was based off intuition and the information that was gathered through simulation.

- Set the confidence level, $100(1-\alpha)\%$.
- Estimate $\hat{\tau}$ using LSE. Record an interval of adjacent observations times which contains $\hat{\tau}$.
- Using the location of $\hat{\tau}$, locate the next closest interval. Denote these two adjacent intervals as (u_j, u_{j+1}) for the left interval and (u_k, u_{k+1}) for the right interval. Note that $j+1=k$.
- Compute $\hat{\tau}_j$ and $\hat{\tau}_k$, the estimators based on splitting the data at the jth and kth intervals, respectively.
- Compute the $100(1-\alpha)\%$ confidence intervals based on both $\hat{\tau}_j$ and $\hat{\tau}_k$. Denote these as $C_{\alpha,j}$ and $C_{\alpha,k}$. There are 9 different scenarios that can occur based on the location of the two confidence intervals.
  - A1: $C_{\alpha,j} \subset (u_j, u_k)$ and $C_{\alpha,k} \subset (u_j, u_k)$
  - A2: $C_{\alpha,j} \subset (u_k, u_{k+1})$ and $C_{\alpha,k} \subset (u_k, u_{k+1})$
  - B1: $u_k \in C_{\alpha,j}$ and $u_k \in C_{\alpha,k}$
  - B2: $C_{\alpha,j} \subset (u_j, u_k)$ and $u_k \in C_{\alpha,k}$
  - B3: $u_k \in C_{\alpha,j}$ and $C_{\alpha,k} \subset (u_j, u_k)$
  - B4: $C_{\alpha,j} \subset (u_k, u_{k+1})$ and $u_k \in C_{\alpha,k}$

- B5: $u_k \in C_{\alpha,j}$ and $C_{\alpha,k} \subset (u_k, u_{k+1})$
- C1: $C_{\alpha,j} \subset (u_j, u_k)$ and $C_{\alpha,k} \subset (u_k, u_{k+1})$
- C2: $C_{\alpha,j} \subset (u_k, u_{k+1})$ and $C_{\alpha,k} \subset (u_j, u_k)$

- If scenario A1, B2, or B3 occurs, choose $\hat{\tau} = \hat{\tau}_j$ and $C = C_{\alpha,j}$.
- If scenario A2, B4, or B5 occurs, choose $\hat{\tau} = \hat{\tau}_k$ and $C = C_{\alpha,k}$.
- If scenario B1 or C1 occurs, choose the $\hat{\tau}$ and C based on which interval contains the MLE for τ_0. If $l_k(\hat{\boldsymbol{\theta}}_k, \hat{\sigma}^2) > l_j(\hat{\boldsymbol{\theta}}_j, \hat{\sigma}^2)$, then choose $\hat{\tau} = \hat{\tau}_k$ and $C = C_{\alpha,k}$. If $l_j(\hat{\boldsymbol{\theta}}_j, \hat{\sigma}^2) > l_k(\hat{\boldsymbol{\theta}}_k, \hat{\sigma}^2)$, then choose $\hat{\tau} = \hat{\tau}_j$ and $C = C_{\alpha,j}$.
- If scenario C2 occurs, then the MLE is neither $\hat{\tau}_j$ nor $\hat{\tau}_k$ because they fall outside of their respective intervals. For this decision, consider the super-log-likelihood functions, $l_j(\boldsymbol{\theta}_j, \sigma^2)$ and $l_k(\boldsymbol{\theta}_k, \sigma^2)$. If $l_j(\hat{\boldsymbol{\theta}}_j, \hat{\sigma}^2) > l_k(\hat{\boldsymbol{\theta}}_k, \hat{\sigma}^2)$, choose $\hat{\tau} = \hat{\tau}_j$ and $C = C_{\alpha,j}$. If $l_k(\hat{\boldsymbol{\theta}}_k, \hat{\sigma}^2) > l_j(\hat{\boldsymbol{\theta}}_j, \hat{\sigma}^2)$, choose $\hat{\tau} = \hat{\tau}_k$ and $C = C_{\alpha,k}$.
- $P(\tau_0 \in C) \to 1 - \alpha$ as $m \to \infty$.

The last part of this algorithm will be proved by cases, i.e., based on the location of τ_0. For Case 1, the position of τ_0 must be considered. The following paragraphs provide a general outline of the proof.

Suppose $\tau_0 \in (u_l, u_{l+1})$ for some $l = 2, \ldots, N-2$ with τ_0 falling closer to u_l than to u_{l+1}. Then, Scenario A2 occurs wp1. Choose $\hat{\tau} = \hat{\tau}_l$. By Theorem 3.6, $\sqrt{m}(\hat{\tau} - \tau_0) \to_d \mathcal{N}(0, \sigma_0^2 \tilde{a}^T \tilde{\Sigma} \tilde{a})$ as $m \to \infty$. Let $C = C_{\alpha,l} = \hat{\tau} \pm z_{\alpha/2} \sqrt{\frac{\sigma_0^2 \tilde{a}^T \tilde{\Sigma} \tilde{a}}{m}}$. Then, $P(\tau_0 \in C) \to 1 - \alpha$ as $m \to \infty$.

On the other hand, for Case 1, suppose $\tau_0 \in (u_l, u_{l+1})$ for some $l = 2, \ldots, N-2$ with τ_0 falling closer to u_{l+1} than to u_l. Then, Scenario A1 occurs wp1. Choose $\hat{\tau} = \hat{\tau}_l$. By Theorem 3.6, $\sqrt{m}(\hat{\tau} - \tau_0) \to_d \mathcal{N}(0, \sigma_0^2 \tilde{a}^T \tilde{\Sigma} \tilde{a})$ as $m \to \infty$. Let $C = C_{\alpha,l} = \hat{\tau} \pm z_{\alpha/2} \sqrt{\frac{\sigma_0^2 \tilde{a}^T \tilde{\Sigma} \tilde{a}}{m}}$. Then, $P(\tau_0 \in C) \to 1 - \alpha$ as $m \to \infty$.

For Case 2, suppose that $\tau_0 = u_l$ for some $l = 2, \ldots, N-1$. Then, as $m \to \infty$,

$$\begin{aligned}
P(\tau_0 \in C) &= P(u_k \in C) \\
&= P\Big((u_k \in C) \cap (\text{A1} \cup \ldots \cup \text{C2})\Big) \\
&= P((u_k \in C) \cap \text{A1}) + \ldots + P((u_k \in C) \cap \text{C2}) \\
&= P\Big((u_k \in C) \cap (B1 \cup B3 \cup B4)\Big) \\
&= P\Big((u_k \in C) \Big| (B1 \cup B3 \cup B4)\Big) \cdot P(B1 \cup B3 \cup B4) \\
&= 1 \cdot P(B1 \cup B3 \cup B4) \\
&= P(B1) + P(B3) + P(B4) \\
&\to 1 - \alpha.
\end{aligned}$$

Since this is just an outline to show why this algorithm works, there are several pieces that need to be proved in order to use this idea. However, notice that regardless of the location of τ_0, this algorithm provides a way to construct a confidence interval of the appropriate level using information regarding the 9 scenarios stated. It is useful to consider these 9 scenarios graphically in Figure 4.1. To form a graphical interpretation, it is helpful to rewrite the scenarios the following way. Let σ_j and σ_k denote the standard errors for the respective splits. Then, for sufficiently large m,

A1: $C_{\alpha,j} \subset (u_j, u_k)$ and $C_{\alpha,k} \subset (u_j, u_k) \Rightarrow$
$(u_j + z_{\frac{\alpha}{2}} \cdot \sigma_j) < \hat{\tau}_j < (u_k - z_{\frac{\alpha}{2}} \cdot \sigma_j)$ and $(u_j + z_{\frac{\alpha}{2}} \cdot \sigma_k) < \hat{\tau}_k < (u_k - z_{\frac{\alpha}{2}} \cdot \sigma_k)$

A2: $C_{\alpha,j} \subset (u_k, u_{k+1})$ and $C_{\alpha,k} \subset (u_k, u_{k+1}) \Rightarrow$
$(u_k + z_{\frac{\alpha}{2}} \cdot \sigma_j) < \hat{\tau}_j < (u_{k+1} - z_{\frac{\alpha}{2}} \cdot \sigma_j)$ and $(u_k + z_{\frac{\alpha}{2}} \cdot \sigma_k) < \hat{\tau}_k < (u_{k+1} - z_{\frac{\alpha}{2}} \cdot \sigma_k)$

B1: $u_k \in C_{\alpha,j}$ and $u_k \in C_{\alpha,k} \Rightarrow$

$(u_k - z_{\frac{\alpha}{2}} \cdot \sigma_j) < \hat{\tau}_j < (u_k + z_{\frac{\alpha}{2}} \cdot \sigma_j)$ and $(u_k - z_{\frac{\alpha}{2}} \cdot \sigma_k) < \hat{\tau}_k < (u_k - z_{\frac{\alpha}{2}} \cdot \sigma_k)$

B2: $C_{\alpha,j} \subset (u_j, u_k)$ and $u_k \in C_{\alpha.k} \Rightarrow$

$(u_j + z_{\frac{\alpha}{2}} \cdot \sigma_j) < \hat{\tau}_j < (u_k - z_{\frac{\alpha}{2}} \cdot \sigma_j)$ and $(u_k - z_{\frac{\alpha}{2}} \cdot \sigma_k) < \hat{\tau}_k < (u_k - z_{\frac{\alpha}{2}} \cdot \sigma_k)$

B3: $u_k \in C_{\alpha,j}$ and $C_{\alpha,k} \subset (u_j, u_k) \Rightarrow$

$(u_k - z_{\frac{\alpha}{2}} \cdot \sigma_j) < \hat{\tau}_j < (u_k + z_{\frac{\alpha}{2}} \cdot \sigma_j)$ and $(u_j + z_{\frac{\alpha}{2}} \cdot \sigma_k) < \hat{\tau}_k < (u_k - z_{\frac{\alpha}{2}} \cdot \sigma_k)$

B4: $C_{\alpha,j} \subset (u_k, u_{k+1})$ and $u_k \in C_{\alpha,k} \Rightarrow$

$(u_k + z_{\frac{\alpha}{2}} \cdot \sigma_j) < \hat{\tau}_j < (u_{k+1} - z_{\frac{\alpha}{2}} \cdot \sigma_j)$ and $(u_k - z_{\frac{\alpha}{2}} \cdot \sigma_k) < \hat{\tau}_k < (u_k + z_{\frac{\alpha}{2}} \cdot \sigma_k)$

B5: $u_k \in C_{\alpha,j}$ and $C_{\alpha,k} \subset (u_k, u_{k+1}) \Rightarrow$

$(u_k - z_{\frac{\alpha}{2}} \cdot \sigma_j) < \hat{\tau}_j < (u_k + z_{\frac{\alpha}{2}} \cdot \sigma_j)$ and $(u_k + z_{\frac{\alpha}{2}} \cdot \sigma_k) < \hat{\tau}_k < (u_{k+1} - z_{\frac{\alpha}{2}} \cdot \sigma_k)$

C1: $C_{\alpha,j} \subset (u_j, u_k)$ and $C_{\alpha,k} \subset (u_k, u_{k+1}) \Rightarrow$

$(u_j + z_{\frac{\alpha}{2}} \cdot \sigma_j) < \hat{\tau}_j < (u_k - z_{\frac{\alpha}{2}} \cdot \sigma_j)$ and $(u_k + z_{\frac{\alpha}{2}} \cdot \sigma_k) < \hat{\tau}_k < (u_{k+1} - z_{\frac{\alpha}{2}} \cdot \sigma_k)$

C2: $C_{\alpha,j} \subset (u_k, u_{k+1})$ and $C_{\alpha,k} \subset (u_j, u_k) \Rightarrow$

$(u_k + z_{\frac{\alpha}{2}} \cdot \sigma_j) < \hat{\tau}_j < (u_{k+1} - z_{\frac{\alpha}{2}} \cdot \sigma_j)$ and $(u_j + z_{\frac{\alpha}{2}} \cdot \sigma_k) < \hat{\tau}_k < (u_k - z_{\frac{\alpha}{2}} \cdot \sigma_k)$

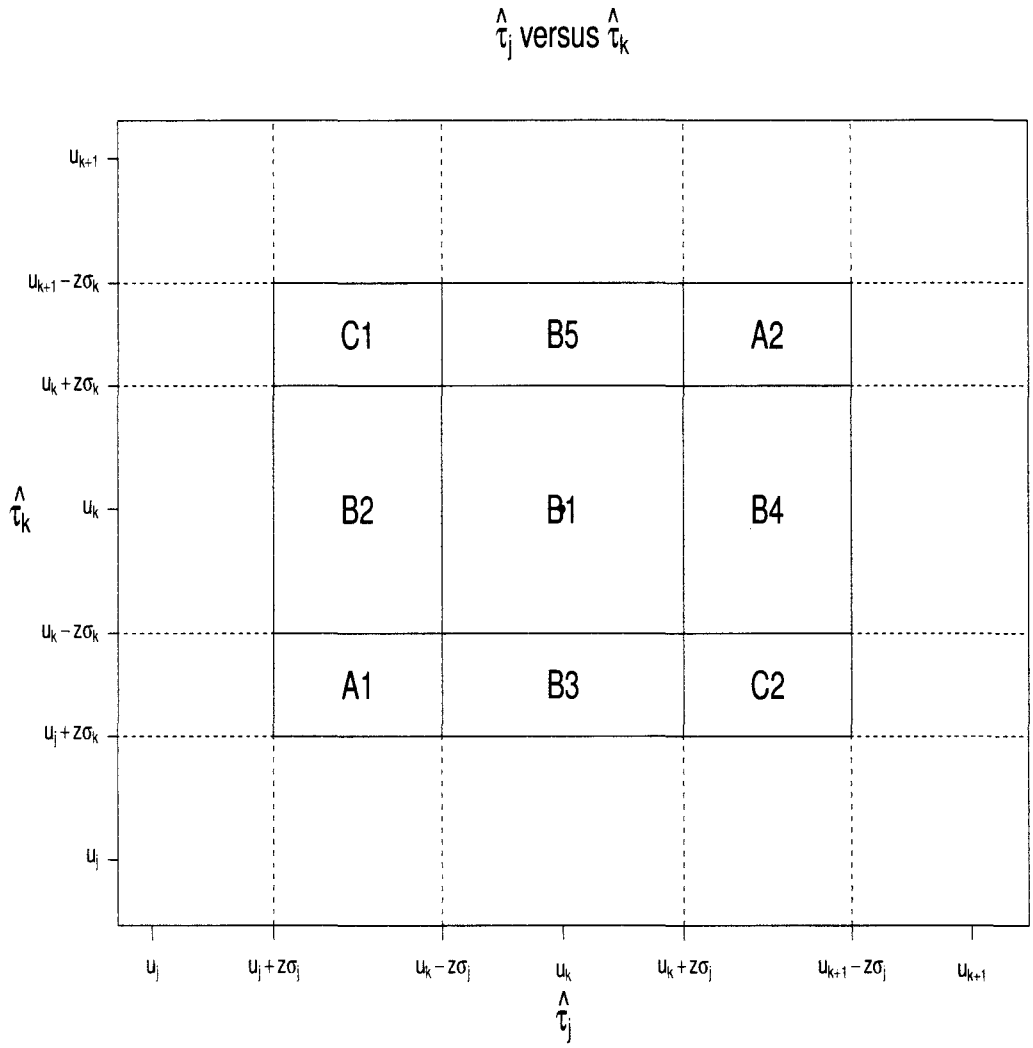


Figure 4.1: 9 Scenarios

First, Case 1, where τ_0 is located between two observations, will be considered.

Lemma 4.2. *If* $\tau_0 \in (u_l, u_{l+1})$ *for some* $l = 2, \ldots, N-2$ *with* τ_0 *closer to* u_l, *then, based on the comprehensive algorithm, scenario A2 occurs with probability 1 as* $m \to \infty$.

Proof. See Appendix. □

Lemma 4.3. *If $\tau_0 \in (u_l, u_{l+1})$ for some $l = 2, \ldots, N-2$ with τ_0 closer to u_{l+1}, then, based on the comprehensive algorithm, scenario A1 occurs with probability 1 as $m \to \infty$.*

Proof. See Appendix. □

Lemma 4.4. *If $\tau_0 \in (u_l, u_{l+1})$ for some $l = 2, \ldots, N-2$ with $\tau_0 = \frac{1}{2}(u_l + u_{l+1})$, then, based on the comprehensive algorithm, either scenario A1 or scenario A2 occurs with probability 1 as $m \to \infty$.*

Proof. See Appendix. □

THEOREM 4.1. *(Case 1): If $\tau_0 \in (u_l, u_{l+1})$ for some $l = 2, \ldots, N-2$, then*

$$P(\tau_0 \in C) \to 1 - \alpha \text{ as } m \to \infty.$$

Proof. By Lemma 4.2, Lemma 4.3, and Lemma 4.4, depending on the exact position for τ_0 in the interval (u_l, u_{l+1}), either scenario A1 or A2 occurs. Following the comprehensive algorithm for these two scenarios, choose $\hat{\tau} = \hat{\tau}_l$. By Theorem 3.6, $\sqrt{m}(\hat{\tau} - \tau_0) \to_d \mathcal{N}(0, \sigma_0^2 \tilde{a}^T \tilde{\Sigma} \tilde{a})$ as $m \to \infty$. Then the choice of confidence interval, $C = C_{\alpha,k} = \hat{\tau} \pm z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma_0^2 \tilde{a}^T \tilde{\Sigma} \tilde{a}}{m}}$ yields $P(\tau_0 \in C) \to 1 - \alpha$ as $m \to \infty$. □

THEOREM 4.2. *If $\tau_0 = u_l$ for some $l = 2, \ldots, N-1$, then $P(\tau_0 \in C) \to 1 - \alpha$ as $m \to \infty$.*

Proof. As $m \to \infty$,

$$
\begin{aligned}
P(\tau_0 \in C) &= P(u_l \in C) \\
&= P\Big(u_l \in C \cap (\text{A1} \cup \text{A2} \cup \text{B1} \cup \text{B2} \cup \text{B3} \cup \text{B4} \cup \text{B5} \cup \text{C1} \cup \text{C2})\Big) \\
&= P(u_l \in C \cap \text{A1}) + P(u_l \in C \cap \text{A2}) + P(u_l \in C \cap \text{B1}) + P(u_l \in C \cap \text{B2}) + \\
&\quad P(u_l \in C \cap \text{B3}) + P(u_l \in C \cap \text{B4}) + P(u_l \in C \cap \text{B5}) + P(u_l \in C \cap \text{C1}) + \\
&\quad P(u_l \in C \cap \text{C2}) \\
&= P\Big(u_l \in C \cap (B1 \cup B3 \cup B4)\Big) \\
&= P\Big(u_l \in C \Big| (B1 \cup B3 \cup B4)\Big) \cdot P(B1 \cup B3 \cup B4) \\
&= 1 \cdot P(B1 \cup B3 \cup B4) \\
&= P(B1) + P(B3) + P(B4) \\
&\to 1 - \alpha.
\end{aligned}
$$

The calculation for $P(B1) + P(B3) + P(B4)$ is based on the bivariate normal curve and is shown in detail in Appendix B.1. □

CHAPTER 5
SIMULATIONS AND APPLICATIONS

5.1 Hybrid algorithm

The comprehensive algorithm presented in Section 4.2 provides an efficient algorithm for obtaining a confidence interval for τ_0 which makes logical choices for what to do based on the scenario. One important aspect of this algorithm is that it avoids the need to know the distribution of the overall MLE $\hat{\tau}$. Depending on one's perspective, though, the comprehensive algorithm's procedure may not be ideal. Specifically, consider the decision for the confidence interval in scenario C2. From a logical perspective, it makes sense that u_k should not be in the confidence interval if it is not in $C_{\alpha,j}$ or $C_{\alpha,k}$. This is also convenient since it makes it easy to select the endpoints of $C_{\alpha,j}$ and $C_{\alpha,k}$ in such a way that the confidence level of the comprehensive algorithm is $100(1-\alpha)\%$.

On the other hand, consider this decision from a likelihood-based perspective. In scenario C2, $\hat{\tau}_j \in (u_k, u_{k+1})$ and $\hat{\tau}_k \in (u_j, u_k)$. In this case, the continuity of the likelihood function implies that $\hat{\tau} = u_k$. Thus, from a likelihood-based perspective, u_k should be in the confidence interval in scenario C2.

Consider a revision to the comprehensive algorithm in a manner so that u_k is in the confidence interval in scenarios B1, B3, B4, and C2 and it is not in the interval otherwise, and refer to such an algorithm as a hybrid algorithm. There are many ways that the endpoints for the confidence interval could be defined for scenario C2,

but choosing between the various ways goes beyond the scope of this discussion and the only important aspect of defining the interval for a hybrid algorithm is that u_k is in the interval. Proceeding as in Appendix 2.1, select z so that the probability of the shaded area in Figure 5.1 equals $1-\alpha$ when Case 2 is true (i.e., $\tau_0 = u_k$).

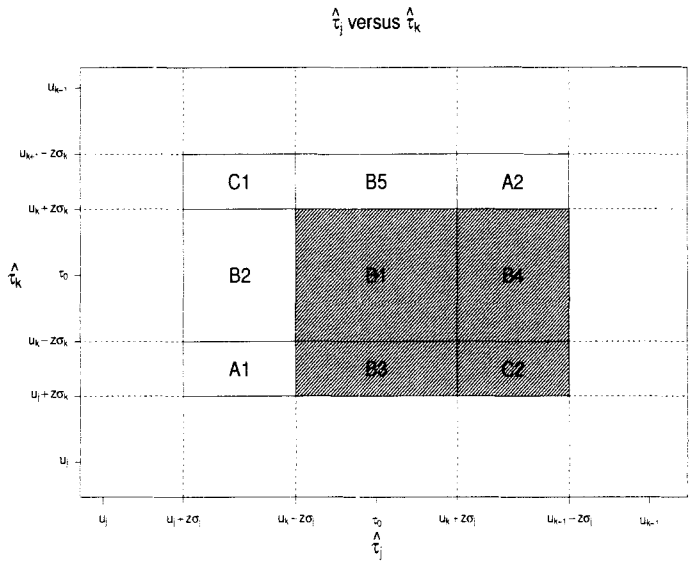


Figure 5.1: Shaded region for hybrid algorithm.

Specifically, the goal is to find z such that $\mathcal{I}_z = 1-\alpha$ where

$$\mathcal{I}_z = \int_{-\infty}^{u_k+z\sigma_k} \int_{u_k-z\sigma_j}^{\infty} \frac{1}{2\pi\sigma_j\sigma_k\sqrt{1-\rho^2}} \times e^{\frac{-1}{2(1-\rho^2)}\left(\frac{(\hat{\tau}_j-\tau_0)^2}{\sigma_j^2} - \frac{2\rho(\hat{\tau}_j-\tau_0)(\hat{\tau}_k-\tau_0)}{\sigma_j\sigma_k} + \frac{(\hat{\tau}_k-\tau_0)^2}{\sigma_k^2}\right)} \mathrm{d}\hat{\tau}_j \mathrm{d}\hat{\tau}_k.$$

It is the case that $z \in (z_\alpha, z_{\alpha/2})$ since $I_{z_\alpha} < 1-\alpha$, $I_{z_{\alpha/2}} > 1-\alpha$ and I_z is increasing in z. There are no standard functions available in statistical software packages to compute this, so a numerical method to obtain z is needed. In this case, the secant method described in Mathews (1987) provides a simple and computationally efficient method of determining z.

The secant method iteratively searches for the solution to the equation $\mathcal{I}_z = 1 - \alpha$. The method proceeds as follows.

1. Specify the tolerance ϵ (some small positive number).
2. Set $z_a = z_\alpha$ and $z_b = z_{\alpha/2}$.
3. Compute $z = z_a + \dfrac{(1-\alpha) - \mathcal{I}_{z_a}}{\mathcal{I}_{z_b} - \mathcal{I}_{z_a}}(z_b - z_a)$.
4. Compute $\mathcal{I}_z$.
5. If $\mathcal{I}_z < 1 - \alpha - \epsilon$, then set $z_a = z$ and go back to step 3.
6. If $\mathcal{I}_z > 1 - \alpha + \epsilon$, then set $z_b = z$ and go back to step 3.
7. If $|\mathcal{I}_z - (1-\alpha)| \leq \epsilon$, then the secant method has converged. Use z.

However, care must be taken to design the hybrid algorithm so that the asymptotic confidence level is correct regardless of the case. To create an algorithm that will also work in Case 1, consider the two-step procedure described below.

1. Determine the scenario based on $z = z_{\alpha/2}$.
2. If scenario B1, B3. B4, or C2 occurs, then recompute z based on the hybrid algorithm and use the resulting confidence interval.

This two step procedure works because, in case 1, none of the scenarios mentioned in step 2 will occur so that the same result is obtained as if the comprehensive algorithm had been used, while in case 2, the level is adjusted to include C2 in the region that would lead to u_k being in the confidence interval.

EXAMPLE 1. Consider the data set in Table 5.1 to illustrate the secant method and the hybrid algorithm.

u_i	1	2	3	4	5	6
$\bar{y}_i$	.9010	.8004	.6987	.6505	.6000	.5502
$m = 100000$	$\sum_{i=1}^{6}\sum_{j=1}^{m}(y_{ij} - \bar{y}_i)^2 = 96086.9604$					

Table 5.1: Data set for Example 1.

First, compute the MLE $\hat{\tau} = 3$ to determine $j = 2$ and $k = 3$. To determine the initial scenario, compute confidence intervals $C_{.05,2} = (2.894079, 3.127490)$ and $C_{.05,3} = (2.880648, 3.054646)$. Since $u_3 = 3$ is in both intervals, this is scenario B1 and thus according to the hybrid algorithm, z must be recomputed. Then the value $z = 1.959905$ is determined by the secant method with $\epsilon = 10^{-6}$ as shown in Table 5.2.

Iteration	z_a	z	z_b	$\mathcal{I}_{z_a}$	$\mathcal{I}_z$	$\mathcal{I}_{z_b}$
0	1.644854		1.959964	0.9000854		0.9500074
1	1.644854	1.959917	1.959964	0.9000854	0.9500019	0.9500074
2	1.644854	1.959905	1.959917	0.9000854	0.9500005	0.9500019

Table 5.2: Secant method for choosing z in the hybrid algorithm for Example 1.

Using this new value of z, compute $2(1 - \Phi^{-1}(z)) = 0.05000692$ and redetermine the scenario by computing the confidence intervals $C_{.05000692,2} = (2.894083, 3.127486)$ and $C_{.05000692,3} = (2.880650, 3.054644)$. Since $u_3 = 3$ is still in both intervals, this is still scenario B1 and a hybrid algorithm will conclude that 3 is in the confidence interval.

5.2 Comparison of confidence estimation methods

The three algorithms presented in this dissertation can be compared numerically by simulation using the statistical software R. The following tables give the results for running 250,000 simulations for varying values of m. Letting $N = 10$,

$\alpha_0 = 1$, $\beta_0 = -0.1$, $\delta_0 = 0.05$, and $\sigma_0 = 0.4$, the three algorithms can be compared for Case 1 by letting $\tau_0 = 5.5$. Table 5.3 displays the results for using the removal algorithm for Case 1. Table 5.4 displays the results for using the comprehensive algorithm for Case 1.

The second columns of Table 5.3 and Table 5.4 give the estimated probability that the confidence interval will contain the true value of the change point. All values in both tables are close to .95 and therefore support the asymptotic results obtained in Chapter 4 regarding the coverage probabilities.

Note that there is a difference in the third and fourth columns of these tables. The third column reports the average width of every confidence interval computed. The fourth column only reports the average width of those confidence intervals that actually contain the true change. These are the important widths to consider because it is of more importance to compare the average width of the "correct" confidence intervals between the different algorithms. However, regardless of which column is used to assess the performance of the algorithms, the comprehensive algorithm outperforms the removal algorithm. Specifically, the width of the 95% confidence intervals obtained by the comprehensive algorithm are roughly 76% of the width of the corresponding intervals for the removal algorithm.

The results for using the hybrid algorithm are displayed in Table 5.6. Table 5.5 and Table 5.7 show the number of times each of the 9 scenarios occur for each value of m.

m	$\hat{P}(\tau_0 \in C_{.95})$	avg width of $C_{.95}$	avg width of $C_{.95}$ given $\tau_0 \in C_{.95}$
10^4	0.943096	0.528222600	0.537994700
10^5	0.946724	0.167653200	0.166318700
10^6	0.949276	0.053121820	0.053209200
10^7	0.949032	0.016809240	0.016796630
10^8	0.950064	0.005316718	0.005316825

Table 5.3: Removal algorithm: Case 1

m	$\hat{P}(\tau_0 \in C_{.95})$	avg width of $C_{.95}$	avg width of $C_{.95}$ given $\tau_0 \in C_{.95}$
10^4	0.950204	0.403635700	0.419388000
10^5	0.949908	0.127400400	0.126725300
10^6	0.950468	0.040283140	0.040158530
10^7	0.950076	0.012738230	0.012730330
10^8	0.949984	0.004028191	0.004028782

Table 5.4: Comprehensive algorithm: Case 1

m	A1	A2	B1	B2	B3	B4	B5	C1	C2
10^4	46870	46811	1047	77757	41	43	77395	12	0
10^5	124692	125308	0	0	0	0	0	0	0
10^6	125044	124956	0	0	0	0	0	0	0
10^7	124739	125261	0	0	0	0	0	0	0
10^8	125087	124913	0	0	0	0	0	0	0

Table 5.5: Results for 9 scenarios - Comprehensive algorithm: Case 1

m	$\hat{P}(\tau_0 \in C_{.95})$	avg width of $C_{.95}$ given $\tau_0 \in C_{.95}$
10^4	0.950204	0.419388000
10^5	0.949908	0.126725300
10^6	0.950468	0.040158530
10^7	0.950076	0.012730330
10^8	0.949984	0.004028782

Table 5.6: Hybrid algorithm: Case 1

m	A1	A2	B1	B2	B3	B4	B5	C1	C2
10^4	46870	46811	1047	77757	41	43	77395	12	0
10^5	124692	125308	0	0	0	0	0	0	0
10^6	125044	124956	0	0	0	0	0	0	0
10^7	124739	125261	0	0	0	0	0	0	0
10^8	125087	124913	0	0	0	0	0	0	0

Table 5.7: Results for 9 scenarios - Hybrid algorithm: Case 1

Again, letting $N = 10$, $\alpha_0 = 1$, $\beta_0 = -0.1$, $\delta_0 = 0.05$, and $\sigma_0 = 0.4$, the three algorithms can be compared for Case 2 by letting $\tau_0 = 5$. Table 5.8 displays the results for using the removal algorithm for Case 2. Table 5.9 displays the results for using the comprehensive algorithm for Case 2. The results for using the hybrid algorithm are displayed in Table 5.11. Table 5.10 and Table 5.12 show the number of times each of the 9 scenarios occur for each value of m.

Again, the second columns of Table 5.8 and Table 5.9 give the estimated probability that the confidence interval will contain the true value of the change point and all values in both tables are close to .95. The third and fourth columns show that the comprehensive algorithm outperforms the removal algorithm. This time, the width of the 95% confidence intervals obtained by the comprehensive al-

gorithm are roughly 85% of the width of the corresponding intervals for the removal algorithm.

m	$\hat{P}(\tau_0 \in C_{.95})$	avg width of $C_{.95}$	avg width of $C_{.95}$ given $\tau_0 \in C_{.95}$
10^4	0.951260	0.50717800	0.50504350
10^5	0.950500	0.15994730	0.15971700
10^6	0.949924	0.05056696	0.05063142
10^7	0.949780	0.01599036	0.01599157
10^8	0.950076	0.005056557	0.00505168

Table 5.8: Removal algorithm: Case 2

m	$\hat{P}(\tau_0 \in C_{.95})$	avg width of $C_{.95}$	avg width of $C_{.95}$ given $\tau_0 \in C_{.95}$
10^4	0.944996	0.427455500	0.444053500
10^5	0.948264	0.135069600	0.129316500
10^6	0.949560	0.042733470	0.041056020
10^7	0.949652	0.013516740	0.014105450
10^8	0.950200	0.004274465	0.004456335

Table 5.9: Comprehensive algorithm: Case 2

m	A1	A2	B1	B2	B3	B4	B5	C1	C2
10^4	985	866	227363	6226	4721	4165	5672	1	1
10^5	969	889	227012	5567	5149	4905	5502	5	2
10^6	906	955	226737	5387	5340	5313	5355	4	3
10^7	905	937	226654	5322	5502	5257	5415	5	3
10^8	938	980	226952	5306	5254	5344	5217	4	5

Table 5.10: Results for 9 scenarios - Comprehensive algorithm: Case 2

m	$\hat{P}(\tau_0 \in C_{.95})$
10^4	0.945000
10^5	0.948272
10^6	0.949572
10^7	0.949664
10^8	0.950220

Table 5.11: Hybrid algorithm: Case 2

m	A1	A2	B1	B2	B3	B4	B5	C1	C2
10^4	985	866	227363	6226	4721	4165	5672	1	1
10^5	969	889	227012	5567	5149	4905	5502	5	2
10^6	906	955	226737	5387	5340	5313	5355	4	3
10^7	905	937	226654	5322	5502	5257	5415	5	3
10^8	938	980	226952	5306	5254	5344	5217	4	5

Table 5.12: Results for 9 scenarios - Hybrid algorithm: Case 2

5.3 Real data examples

Data that was obtained from the GSS (General Social Surveys) database, Smith, et al. (1972-2010), will be used to provide an example of the ideas presented in this dissertation. The independent variables that were used from this survey are the years 1972-2010. The response variable used from this survey was the highest grade of schooling, 0-20, for which the respondent's mother received credit. Starting in 1994, data was only obtained every other year. Three other years are unrepresented in this dataset; 1979, 1981, and 1992. There are 1156 responses for every represented year for a total of 32,368 responses. The scatterplot of the raw data is shown in Figure 5.2 and the scatterplot of the average education level for

each year is shown in Figure 5.3 below.

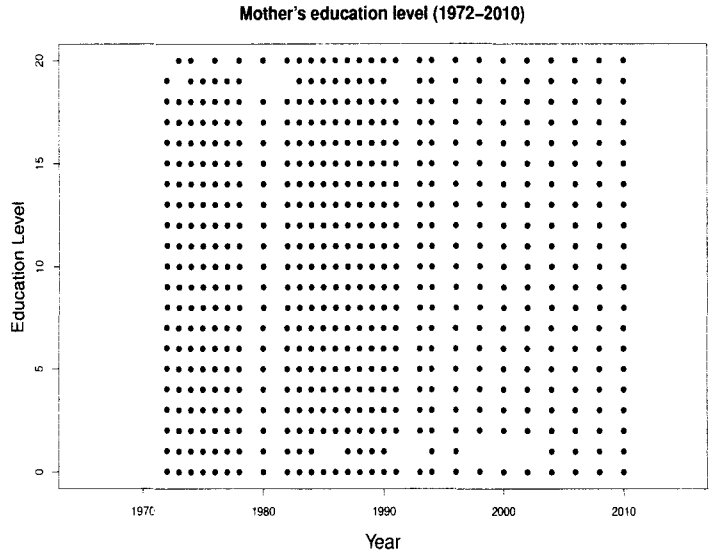


Figure 5.2: Scatterplot of mother's education data.

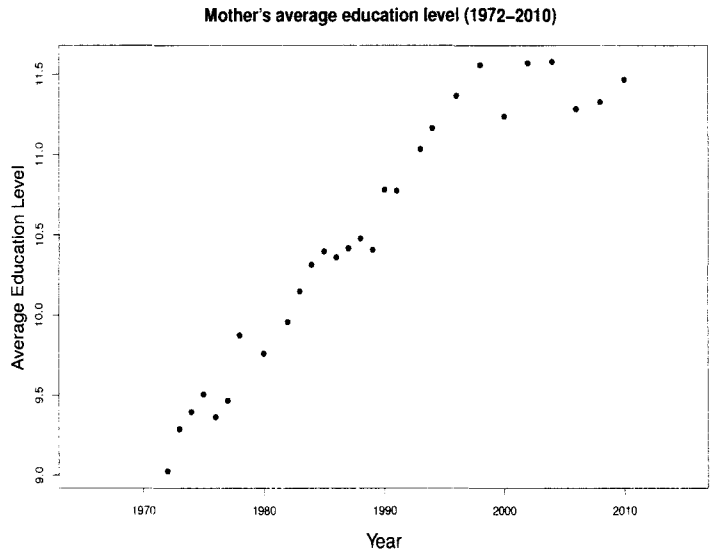


Figure 5.3: Scatterplot of average mother's education data.

When the data is assumed to have one change and fit to a clustered segmented regression model assuming a continuity constraint as described in Section 2.3, the resulting segmented regression lines are shown below in Figure 5.4.

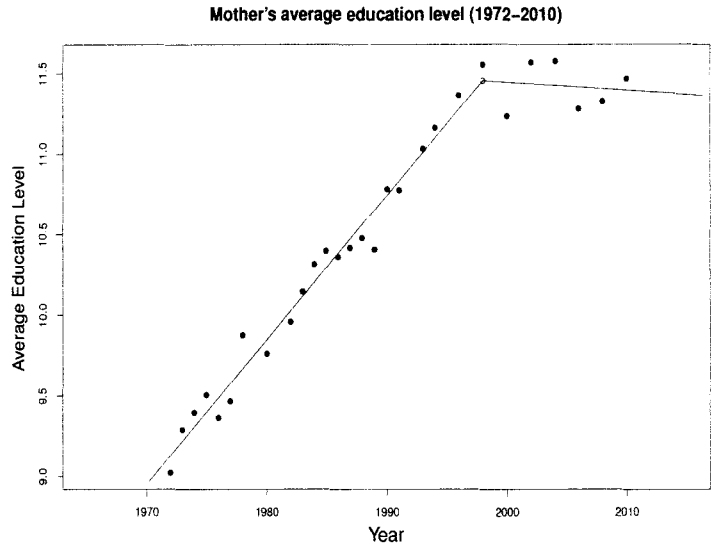


Figure 5.4: Segmented regression line

Using the removal algorithm on this data results in the removal of $u_{22} = 1998$. Then the re-computed estimator for the change point is $\hat{\tau} = 1997.405$ and the 95% confidence interval is $C = (1997.392, 1997.418)$. The length of C is 0.02609835.

If the comprehensive algorithm is used instead, the jth interval is found to be $(u_{21}, u_{22}) = (1996, 1998)$ and the kth interval is found to be $(u_{22}, u_{23}) = (1998, 2000)$. The jth and kth change point estimators are $\hat{\tau}_{21} = 1998.208$ and $\hat{\tau}_{22} = 1997.185$, respectively. The 95% confidence intervals associated with these estimators are $C_{.05,21} = (1998.199, 1998.217)$ and $C_{.05,22} = (1997.172, 1997.198)$, respectively. This is an example of scenario C2. Following the comprehensive algorithm, the decision on what estimator and confidence interval to use will be based off which super log-likelihood function is the greatest. Since $Q_j(\hat{\boldsymbol{\theta}}_{21}) = 0.3530153$ and $Q_k(\hat{\boldsymbol{\theta}}_{22}) = 0.3486604$, then $l_{21}(\hat{\boldsymbol{\theta}}_{21}, \hat{\sigma}^2) < l_{22}(\hat{\boldsymbol{\theta}}_{22}, \hat{\sigma}^2)$. This means that the estimator that is chosen using the comprehensive algorithm is $\hat{\tau} = \hat{\tau}_{22} = 1997.185$ and the confidence interval that will be used is $C = C_{.05,22} = (1997.172, 1997.198)$. Here, the length of C is 0.02609835.

REFERENCES

[1] Jushan Bai, *Estimating multiple breaks one at a time*, Econometric Theory **13** (1997), no. 3, 315–352. MR 1455175 (98h:62131)

[2] Jushan Bai and Pierre Perron, *Estimating and testing linear models with multiple structural changes*, Econometrica **66** (1998), no. 1, 47–78. MR 1616121 (98m:62184)

[3] A. K. Basu, *Measure theory and probability*, PHI Learning Pvt. Ltd., New Delhi, 1999.

[4] Martin Bilodeau and David Brenner, *Theory of multivariate statistics*, Springer Texts in Statistics, Springer-Verlag, New York, 1999. MR 1705291 (2000f:62002)

[5] Terence Tai-Leung Chong, *Estimating the locations and number of change points by the sample-splitting method*, Statist. Papers **42** (2001), no. 1, 53–79. MR 1821174 (2002k:62109)

[6] Paul I. Feder, *On asymptotic distribution theory in segmented regression problems—identified case*, Ann. Statist. **3** (1975), 49–83. MR 0378267 (51 #14436a)

[7] Ryan Gill, Grzegorz A. Rempala, and Michal Czajkowski, *Confidence estimation via the parametric bootstrap in logistic joinpoint regression*, J. Statist. Plann. Inference **139** (2009), no. 9, 3132–3141. MR 2535189 (2010j:62066)

[8] Gene H. Golub and Charles F. Van Loan, *Matrix computations*, third ed., Johns Hopkins Studies in the Mathematical Sciences, Johns Hopkins University Press, Baltimore, MD, 1996. MR 1417720 (97g:65006)

[9] David V. Hinkley, *Inference in two-phase regression*, J. Amer. Statist. Assoc. **66** (1971), no. 336, 736–743.

[10] Derek J. Hudson, *Fitting segmented curves whose join points have to be estimated*. J. Amer. Statist. Assoc. **61** (1966), 1097–1129. MR 0210243 (35 #1137)

[11] Helmut Küchenhoff, *An exact algorithm for estimating breakpoints in segmented linear models*, Comput. Statist. Data Anal **12** (1997), no. 2, 235–247.

[12] Tze Leung Lai. *Sequential changepoint detection in quality control and dynamical systems*, J. Roy. Statist. Soc. Ser. B **57** (1995), no. 4, 613–658, With discussion and a reply by the author. MR 1354072 (96j:62154)

[13] Jian Liu, Shiying Wu, and James V. Zidek, *On segmented multivariate regression*, Statistica Sinica **7** (1997), 497–525.

[14] John H. Mathews, *Numerical methods for computer science, engineering, and mathematics*, Prentice Hall Inc., Englewood Cliffs, NJ, 1987. MR 914373 (89j:65004)

[15] E. S. Page, *Continuous inspection schemes*, Biometrika **41** (1954), 100–115. MR 0088850 (19,589d)

[16] Richard E. Quandt, *The estimation of the parameters of a linear regression system obeying two separate regimes*, J. Amer. Statist. Assoc. **53** (1958), 873–880. MR 0100314 (20 #6747)

[17] Tom W Smith, Peter Marsden, Michael Hout, and Jibum Kim, *General social surveys, 1972-2010*, [machine-readable data file] /Principal Investigator, Tom

W. Smith; Co-Principal Investigator, Peter V. Marsden; Co-Principal Investigator, Michael Hout: Sponsered by National Science Foundation. –NORC ed.– Chicago: National Opinion Research Center [producer]; Storrs, CT: The Roper Center for Public Opinion Research, University of Connecticut [distributor], 2011, Accessed at http://sda.berkeley.edu/cgi-bin/hsda?harcsda+gss10nw on February 23, 2012 at 2:39pm.

[18] David L. Sylwester, *On maximum likelihood estimation for two-phase linear regression*, September 1965, Prepared under the auspices of public health service grant USPHS=5T1GM 25-07.

[19] A. W. van der Vaart, *Asymptotic statistics*, Cambridge Series in Statistical and Probabilistic Mathematics, vol. 3, Cambridge University Press, Cambridge, 1998. MR 1652247 (2000c:62003)

APPENDIX

A.1 Proof for Lemma 4.2

Assume $\tau_0 \in (u_l, u_{l+1})$ for some $l = 2, \ldots, N-2$ with τ_0 closer to u_l. To be consistent with the comprehensive algorithm, let $l = k$. Let $\hat{\tau} = \hat{\tau}_k$ be the estimator based on splitting at the kth observation and $\tilde{\tau} = \hat{\tau}_j$ be the estimator based on splitting at the jth observation.

First, consider splitting the data at the kth observation. This means the observations on the left side are $\{u_1, \ldots, u_k\}$ and the observations on the right side are $\{u_{k+1}, \ldots, u_N\}$. The estimates for the intercept and slope of the regression line fit through the data on the left are $\hat{\omega}_1$ and $\hat{\omega}_2$, respectively. Likewise, $\hat{\omega}_3$ and $\hat{\omega}_4$ are the estimates for the respective intercept and slope for the right side. The point where these two lines cross gives the estimate for the change point. Thus, the change point estimator based on this split is found as follows:

$$\begin{aligned} \hat{\omega}_1 + \hat{\omega}_2\hat{\tau} &= \hat{\omega}_3 + \hat{\omega}_4\hat{\tau} \\ (\hat{\omega}_2 - \hat{\omega}_4)\hat{\tau} &= \hat{\omega}_3 - \hat{\omega}_1 \\ \hat{\tau} &= \frac{\hat{\omega}_3 - \hat{\omega}_1}{\hat{\omega}_2 - \hat{\omega}_4} \end{aligned}$$

Strong consistency gives $\hat{\tau} \rightarrow_{a.s.} \tau_0$ as $m \rightarrow \infty$. So for sufficiently large m, it is the case that $\hat{\tau} \in (u_k, u_{k+1})$ wp1 and $C_{\alpha,k} \subset (u_k, u_{k+1})$ wp1.

If the data is split at the jth observation, the left side consists of $\{u_1, \ldots, u_j\}$ and the right side is $\{u_k, \ldots, u_N\}$. Since the left side is specified the same as it was when the data was split at the kth observation, the respective estimates $\tilde{\omega}_1$ and $\tilde{\omega}_2$

for the intercept and slope on the left are strongly consistent: that is, $\tilde{\omega}_1 \rightarrow_{a.s.} \omega_{01}$ and $\tilde{\omega}_2 \rightarrow_{a.s.} \omega_{02}$ as $m \rightarrow \infty$. However, since $\tau_0 \in (u_k, u_{k+1})$, the kth observation truly follows the trend of the left side so including the kth observation on the right makes the right side mis-specified. This changes the estimates for the intercept and slope, namely, $\tilde{\omega}_3$ and $\tilde{\omega}_4$. The change point estimator based on this split is found similar to before. Let

$$\tilde{\tau} = \frac{\tilde{\omega}_3 - \tilde{\omega}_1}{\tilde{\omega}_2 - \tilde{\omega}_4}$$

Here, strong consistency does not guarantee that $\tilde{\tau} \in (u_k, u_{k+1})$ or that $C_{\alpha,j} \subset (u_k, u_{k+1})$. This must be established in a different way. The following argument justifies why $u_k < \tilde{\tau} < \tau_0$ wp1 which implies that $C_{\alpha,j} \subset (u_k, u_{k+1})$ as $m \rightarrow \infty$.

In order to compare the right side parameter estimates for the different splits, define

$$\boldsymbol{U}_R = \begin{bmatrix} \boldsymbol{U}_1 \\ \boldsymbol{U}_2 \end{bmatrix} \text{ where } \boldsymbol{U}_1 = \begin{bmatrix} 1 & u_k \end{bmatrix} \text{ and } \boldsymbol{U}_2 = \begin{bmatrix} 1 & u_{k+1} \\ \vdots & \vdots \\ 1 & u_N \end{bmatrix}$$

$$\bar{\boldsymbol{y}}_R = \begin{bmatrix} \bar{\boldsymbol{y}}_1 \\ \bar{\boldsymbol{y}}_2 \end{bmatrix} \text{ where } \bar{\boldsymbol{y}}_1 = \bar{y}_k \text{ and } \bar{\boldsymbol{y}}_2 = \begin{bmatrix} \bar{y}_{k+1} \\ \vdots \\ \bar{y}_N \end{bmatrix}$$

$$\boldsymbol{\omega}_R = \begin{bmatrix} \omega_3 \\ \omega_4 \end{bmatrix}.$$

The estimates based on splitting at the kth observation are found by

$\hat{\boldsymbol{\omega}}_R = (\boldsymbol{U}_2^T\boldsymbol{U}_2)^{-1}\boldsymbol{U}_2^T\bar{\boldsymbol{y}}_2$. Splitting at the jth observation gives

$$\begin{aligned}
\tilde{\boldsymbol{\omega}}_R &= (\boldsymbol{U}_R^T\boldsymbol{U}_R)^{-1}\boldsymbol{U}_R^T\bar{\boldsymbol{y}}_R \\
&= (\boldsymbol{U}_1^T\boldsymbol{U}_1 + \boldsymbol{U}_2^T\boldsymbol{U}_2)^{-1}(\boldsymbol{U}_1^T\bar{\boldsymbol{y}}_1 + \boldsymbol{U}_2^T\bar{\boldsymbol{y}}_2) \\
&= [(\boldsymbol{U}_2^T\boldsymbol{U}_2)((\boldsymbol{U}_2^T\boldsymbol{U}_2)^{-1}\boldsymbol{U}_1^T\boldsymbol{U}_1 + \boldsymbol{I})]^{-1}(\boldsymbol{U}_1^T\bar{\boldsymbol{y}}_1 + \boldsymbol{U}_2^T\bar{\boldsymbol{y}}_2) \\
&= \boldsymbol{C}(\boldsymbol{U}_2^T\boldsymbol{U}_2)^{-1}(\boldsymbol{U}_1^T\bar{\boldsymbol{y}}_1 + \boldsymbol{U}_2^T\bar{\boldsymbol{y}}_2) \\
&= \boldsymbol{C}(\boldsymbol{U}_2^T\boldsymbol{U}_2)^{-1}\boldsymbol{U}_1^T\bar{\boldsymbol{y}}_1 + \boldsymbol{C}\hat{\boldsymbol{\omega}}_R
\end{aligned}$$

where $\boldsymbol{C} = ((\boldsymbol{U}_2^T\boldsymbol{U}_2)^{-1}\boldsymbol{U}_1^T\boldsymbol{U}_1 + \boldsymbol{I})^{-1}$. The Sherman-Morrison-Woodbury formula (Golub and Van Loan, 1996) can be used to rewrite $\boldsymbol{C}$.

$$\boldsymbol{C} = \boldsymbol{I} - \frac{(\boldsymbol{U}_2^T\boldsymbol{U}_2)^{-1}\boldsymbol{U}_1^T\boldsymbol{U}_1}{d} \text{ where } d = 1 + \boldsymbol{U}_1(\boldsymbol{U}_2^T\boldsymbol{U}_2)^{-1}\boldsymbol{U}_1^T$$

Thus, $\tilde{\boldsymbol{\omega}}_R$ can be simplified as follows. (See Appendix 1.1(a) for the detailed simplification.)

$$\tilde{\boldsymbol{\omega}}_R = \hat{\boldsymbol{\omega}}_R - \frac{(\boldsymbol{U}_2^T\boldsymbol{U}_2)^{-1}\boldsymbol{U}_1^T\boldsymbol{U}_1}{d}\hat{\boldsymbol{\omega}}_R + \left(\boldsymbol{I} - \frac{(\boldsymbol{U}_2^T\boldsymbol{U}_2)^{-1}\boldsymbol{U}_1^T\boldsymbol{U}_1}{d}\right)(\boldsymbol{U}_2^T\boldsymbol{U}_2)^{-1}\boldsymbol{U}_1^T\bar{\boldsymbol{y}}_1$$

$$\begin{bmatrix} \tilde{\omega}_3 \\ \tilde{\omega}_4 \end{bmatrix} = \begin{bmatrix} \hat{\omega}_3 \\ \hat{\omega}_4 \end{bmatrix} - \frac{\hat{\omega}_3 + u_k\hat{\omega}_4 - \bar{y}_k}{A + \sum_{i=k+1}^{N}(u_i - u_k)^2} \begin{bmatrix} \sum_{i=k+1}^{N} u_i(u_i - u_k) \\ \sum_{i=k+1}^{N}(u_k - u_i) \end{bmatrix}$$

where $A = \det(\boldsymbol{U}_2^T\boldsymbol{U}_2) = (N-k)\sum_{i=k+1}^{N} u_i^2 - \left(\sum_{i=k+1}^{N} u_i\right)^2$.

It is important to note that $A > 0$. This is due to the fact that $\boldsymbol{U}_2^T\boldsymbol{U}_2$ is positive definite. By definition, $\boldsymbol{U}_2^T\boldsymbol{U}_2$ is positive definite if, for any $\boldsymbol{x} \in \mathbb{R}$ where

$\boldsymbol{x} \neq 0$, $\boldsymbol{x}^T(\boldsymbol{U}_2^T\boldsymbol{U}_2)\boldsymbol{x} > 0$. Let $\boldsymbol{x} = [x_1, x_2]^T$. Then,

$$\begin{aligned}
\boldsymbol{x}^T\boldsymbol{U}_2^T\boldsymbol{U}_2\boldsymbol{x} &= (\boldsymbol{U}_2\boldsymbol{x})^T(\boldsymbol{U}_2\boldsymbol{x}) \\
&= \left(\begin{bmatrix} 1 & u_{k+1} \\ \vdots & \vdots \\ 1 & u_N \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)^T\left(\begin{bmatrix} 1 & u_{k+1} \\ \vdots & \vdots \\ 1 & u_N \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) \\
&= \begin{bmatrix} (x_1 + u_{k+1}x_2) & \ldots & (x_1 + u_N x_2) \end{bmatrix}\begin{bmatrix} (x_1 + u_{k+1}x_2) \\ \vdots \\ (x_1 + u_N x_2) \end{bmatrix} \\
&= (x_1 + u_{k+1}x_2)^2 + \ldots + (x_1 + u_N x_2)^2 \geq 0
\end{aligned}$$

Note that $\boldsymbol{x}^T(\boldsymbol{U}_2^T\boldsymbol{U}_2)\boldsymbol{x} = 0$ only if both elements of $\boldsymbol{x}$ are equal to 0. Thus, $\boldsymbol{x}^T(\boldsymbol{U}_2^T\boldsymbol{U}_2)\boldsymbol{x} > 0$ for all nonzero $\boldsymbol{x}$. This implies that $A > 0$.

For this case, the left side is correctly specified regardless of whether the split was at the jth or kth observation. Thus, the attention will be focused on how the regression line on the right changes depending on where the data is split. First, compare the fitted values at u_k from the right regression line to the actual observation, $(u_k, \bar{y}_k)$.

$$\begin{aligned}
\bar{y}_k &= \hat{\omega}_1 + u_k\hat{\omega}_2 \rightarrow_{a.s.} \omega_{01} + u_k\omega_{02} \\
\hat{y}_k &= \hat{\omega}_3 + u_k\hat{\omega}_4 \rightarrow_{a.s.} \omega_{03} + u_k\omega_{04} \\
\tilde{y}_k &= \tilde{\omega}_3 + u_k\tilde{\omega}_4
\end{aligned}$$

Then, $\tilde{\boldsymbol{\omega}}_R$ can be re-written as follows:

$$\begin{aligned}
\tilde{\boldsymbol{\omega}}_R &= \hat{\boldsymbol{\omega}}_R - \frac{\hat{\omega}_3 + u_k\hat{\omega}_4 - \bar{y}_k}{A + \sum_{i=k+1}^{N}(u_i - u_k)^2}\begin{bmatrix} \sum_{i=k+1}^{N} u_i(u_i - u_k) \\ \sum_{i=k+1}^{N}(u_k - u_i) \end{bmatrix} \\
\boldsymbol{U}_1\tilde{\boldsymbol{\omega}}_R &= \boldsymbol{U}_1\hat{\boldsymbol{\omega}}_R - \frac{\hat{y}_k - \bar{y}_k}{A + \sum_{i=k+1}^{N}(u_i - u_k)^2}\boldsymbol{U}_1\begin{bmatrix} \sum_{i=k+1}^{N} u_i(u_i - u_k) \\ \sum_{i=k+1}^{N}(u_k - u_i) \end{bmatrix} \\
\tilde{y}_k &= \hat{y}_k - \frac{\hat{y}_k - \bar{y}_k}{A + \sum_{i=k+1}^{N}(u_i - u_k)^2}\sum_{i=k+1}^{N}(u_i - u_k)^2 \\
\tilde{y}_k &= \hat{y}_k - c(\hat{y}_k - \bar{y}_k)
\end{aligned}$$

where $c = \frac{\sum_{i=k+1}^{N}(u_i-u_k)^2}{A+\sum_{i=k+1}^{N}(u_i-u_k)^2}$. Since $A > 0$, the value of c is between 0 and 1. Thus, $\tilde{y}_k = \hat{y}_k(1-c) + c \cdot \bar{y}_k$ implies that $\hat{y}_k < \tilde{y}_k < \bar{y}_k$ or $\bar{y}_k < \tilde{y}_k < \hat{y}_k$. Solving for c,

$$0 < \frac{\tilde{y}_k - \hat{y}_k}{\bar{y}_k - \hat{y}_k} < 1.$$

Without loss of generality, assume $\omega_{02} < \omega_{04}$. (The argument is similar for $\omega_{02} > \omega_{04}$.) Then $\hat{y}_k < \tilde{y}_k < \bar{y}_k$. By the Strong Law of Large Numbers, $\bar{y}_i \rightarrow_{a.s.} \mu_{0,i}$ for $i = 1, \ldots, N$. Thus, when m is large, the values of $\bar{y}_i$ will be very close to $\mu_{0,i}$ for all i. Figure A.1 illustrates the means and the fitted values for the various regression lines. Based on the graph in Figure A.1, it is clear, geometrically, that $\hat{\omega}_2 < \tilde{\omega}_4 < \hat{\omega}_4$.

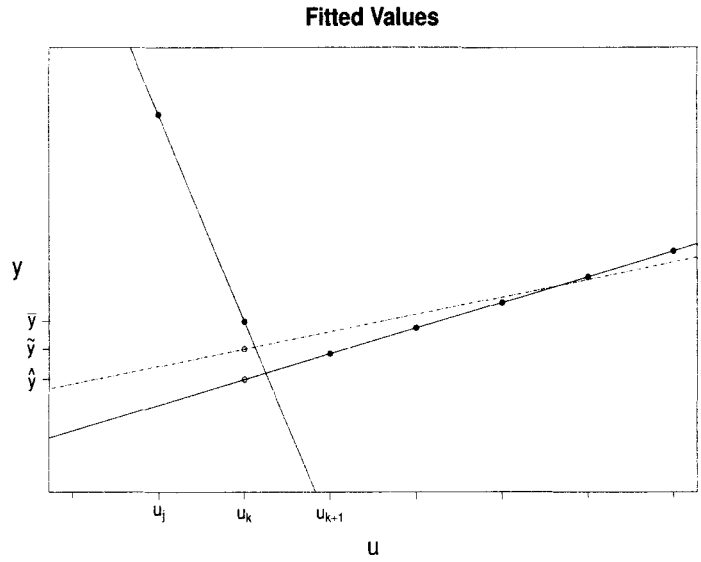


Figure A.1: Illustration of fitted values from right regression line when m is large.

Then,

$$\begin{aligned}
\frac{\tilde{y}_k - \hat{y}_k}{\bar{y}_k - \hat{y}_k} &< 1 \\
\tilde{y}_k - \hat{y}_k &< \bar{y}_k - \hat{y}_k \\
\tilde{y}_k &< \bar{y}_k \\
\tilde{\omega}_3 + u_k\tilde{\omega}_4 &< \hat{\omega}_1 + u_k\hat{\omega}_2 \\
\tilde{\omega}_3 - \hat{\omega}_1 &< (\hat{\omega}_2 - \tilde{\omega}_4)u_k \\
\left(\frac{\tilde{\omega}_3 - \hat{\omega}_1}{\hat{\omega}_2 - \tilde{\omega}_4}\right) &> u_k \\
\tilde{\tau} &> u_k.
\end{aligned}$$

Due to the nature of the regression line on the right, $\hat{\omega}_3 + \hat{\omega}_4\hat{\tau} < \tilde{\omega}_3 + \tilde{\omega}_4\hat{\tau}$. To clarify this, the assumption that $\hat{y}_k < \tilde{y}_k < \bar{y}_k$ means that there is a positive residual at u_k. By the nature of regression, in particular least squares, there must also be at least one negative residual on the right side. This must occur at u_{k+1}. (If the residual at u_{k+1} was positive, then all of the residuals on the right would be positive. This would contradict the fact that this is a regression line.) This can be seen graphically

in Figure A.2 below.

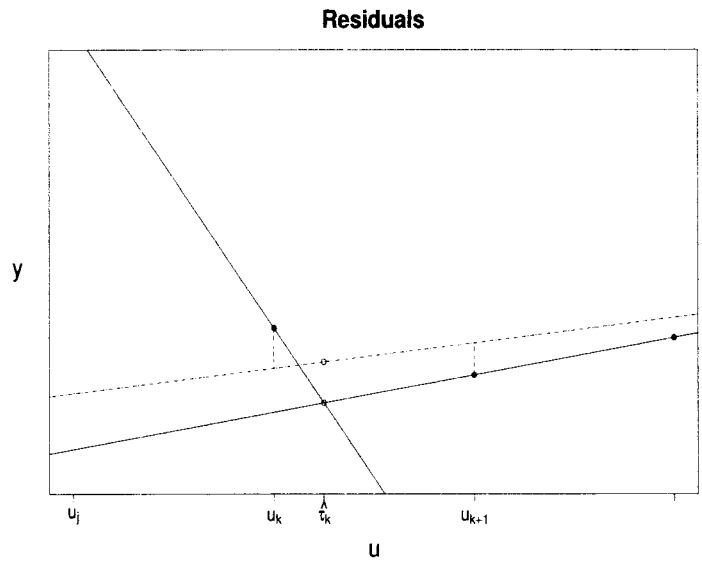


Figure A.2: Illustration of residuals from right regression line based on jth split when m is large.

Thus, the assumption that $\hat{y}_k < \tilde{y}_k < \bar{y}_k$ implies $\hat{\omega}_3 + \hat{\omega}_4\hat{\tau} < \tilde{\omega}_3 + \tilde{\omega}_4\hat{\tau}$. Using this, the following holds:

$$\begin{aligned}
\hat{\omega}_1 + \hat{\omega}_2\hat{\tau} &< \tilde{\omega}_3 + \tilde{\omega}_4\hat{\tau} \\
\hat{\omega}_1 - \tilde{\omega}_3 &< (\tilde{\omega}_4 - \hat{\omega}_2)\hat{\tau} \\
\left(\frac{\hat{\omega}_1 - \tilde{\omega}_3}{\tilde{\omega}_4 - \hat{\omega}_2}\right) &< \hat{\tau} \\
\left(\frac{\tilde{\omega}_3 - \hat{\omega}_1}{\hat{\omega}_2 - \tilde{\omega}_4}\right) &< \hat{\tau} \\
\tilde{\tau} &< \hat{\tau}
\end{aligned}$$

Formally, we have

$$\begin{aligned}
\tilde{\tau} &= \frac{\tilde{\omega}_3 - \tilde{\omega}_1}{\tilde{\omega}_2 - \tilde{\omega}_4} \\
&= \frac{(\tilde{\omega}_3 - \hat{\omega}_1) + (\hat{\omega}_1 - \tilde{\omega}_1)}{(\tilde{\omega}_2 - \hat{\omega}_2) + (\hat{\omega}_2 - \tilde{\omega}_4)}.
\end{aligned}$$

Now, it follows that

$$\tilde{\omega}_4 = \hat{\omega}_4 + \frac{\sum_{i=k+1}^{N}(u_i - u_k)(\hat{y}_k - \bar{y}_k)}{A + \sum_{i=k+1}^{N}(u_i - u_k)^2}.$$

Since

$$\begin{aligned}
\hat{y}_k - \bar{y}_k &\rightarrow_{a.s.} (\omega_{03} - \omega_{04}u_k) - (\omega_{01} + \omega_{02}u_k) \\
&= (\omega_{03} - \omega_{01}) + (\omega_{04} - \omega_{02})u_k \\
&= -\tau_0(\omega_{04} - \omega_{02}) + (\omega_{04} - \omega_{02})u_k \\
&= -(\omega_{04} - \omega_{02})(\tau_0 - u_k),
\end{aligned}$$

we have

$$\begin{aligned}
\tilde{\omega}_4 &\rightarrow_{a.s.} \omega_{04} - c^*(\omega_{04} - \omega_{02}) \\
&= c^*\omega_{02} + (1 - c^*)\omega_{04}
\end{aligned}$$

where

$$c^* = \frac{(\tau_0 - u_k)\sum_{i=k+1}^{N}(u_i - u_k)}{A + \sum_{i=k+1}^{N}(u_i - u_k)^2} \in (0, 1).$$

Hence, $\hat{\omega}_2 - \tilde{\omega}_4 \rightarrow_{a.s.} (1 - c^*)(\omega_{02} - \omega_{04})$. Also, it follows that

$$\begin{aligned}
\tilde{\omega}_3 - \hat{\omega}_1 &= (\tilde{y}_k - u_k\tilde{\omega}_4) - (\bar{y}_k - u_k\hat{\omega}_2) \\
&= (\tilde{y}_k - \bar{y}_k) + u_k(\hat{\omega}_2 - \tilde{\omega}_4) \\
&= (1 - c)(\hat{y}_k - \bar{y}_k) + u_k(\hat{\omega}_2 - \tilde{\omega}_4) \\
&\rightarrow_{a.s.} (1 - c)(\omega_{02} - \omega_{04})(\tau_0 - u_k) + u_k(1 - c^*)(\omega_{02} - \omega_{04}).
\end{aligned}$$

So, we have

$$\begin{aligned}
\tilde{\tau} &\rightarrow_{a.s.} \frac{(1 - c)(\omega_{02} - \omega_{04})(\tau_0 - u_k) + u_k(1 - c^*)(\omega_{02} - \omega_{04}) + 0}{0 + (1 - c^*)(\omega_{02} - \omega_{04})} \\
&= \frac{1 - c^*}{1 - c}(\tau_0 - u_k) + u_k \in (u_k, \tau_0)
\end{aligned}$$

since

$$c > c^* \Rightarrow \frac{1-c}{1-c^*} < 1 \Rightarrow \frac{1-c}{1-c^*}(\tau_0 - u_k) + u_k < (\tau_0 - u_k) + u_k = \tau_0.$$

Putting this all together. $u_k < \tilde{\tau} < \tau_0$ wp1 as $m \to \infty$. The same argument can be reversed for the assumption that $\omega_{02} > \omega_{04}$. Either way, the conclusion is the same, i.e., that $\tilde{\tau} \in (u_k, \tau_0)$ wp1 as $m \to \infty$.

To summarize this argument, if $\tau_0 \in (u_l, u_{l+1})$ for any $l = 2, \ldots, N-2$ with τ_0 closer to u_l, then for sufficiently large m,

$$\begin{aligned} \hat{\tau} \to_{a.s} \tau_0 \text{ and } \hat{\sigma}^2 \to_{a.s.} \sigma_0^2 \text{ as } m \to \infty \quad &\Rightarrow \quad \hat{\tau} \in (u_k, u_{k+1}) \text{ and } \sqrt{\frac{\hat{\sigma}^2 \tilde{a}^T \widetilde{\Sigma} \tilde{a}}{m}} \to 0 \\ &\Rightarrow \quad C_{\alpha,k} \subset (u_k, u_{k+1}) \end{aligned}$$

and

$$\begin{aligned} \tilde{y}_k \text{ is between } \hat{y}_k \text{ and } \bar{y}_k \text{ wp1 as } m \to \infty \quad &\Rightarrow \quad u_k < \tilde{\tau} < \tau_0 \\ &\Rightarrow \quad C_{\alpha,j} \subset (u_k, u_{k+1}). \end{aligned}$$

Thus, Scenario A2 occurs.

A.2 Proof for Lemma 4.3

This argument is similar to that for Lemma 4.2. However, in this case, the left side of the data is mis-specified if the split occurs at the wrong observation. The rest of the argument can be followed similarly to show that if $\tau_0 \in (u_l, u_{l+1})$ for any $l = 2, \ldots, N-2$ with τ_0 closer to u_{l+1}, then for sufficiently large m, Scenario A1 occurs.

A.3 Proof for Lemma 4.4

This argument is similar to that for Lemma 4.2 and Lemma 4.3. In this case, it is random which side is mis-specified. If m is sufficiently large, then Scenario A1 occurs when the left side is mis-specified, while Scenario A2 occurs when the right side is mis-specified.

A.1(a): Simplification of $\tilde{\omega}_R$

$$\boldsymbol{U}_2^T\boldsymbol{U}_2 = \begin{bmatrix} N-k & \sum_{i=k+1}^{N} u_i \\ \sum_{i=k+1}^{N} u_i & \sum_{i=k+1}^{N} u_i^2 \end{bmatrix} \Rightarrow (\boldsymbol{U}_2^T\boldsymbol{U}_2)^{-1} = \frac{1}{A}\begin{bmatrix} \sum_{i=k+1}^{N} u_i^2 & -\sum_{i=k+1}^{N} u_i \\ -\sum_{i=k+1}^{N} u_i & N-k \end{bmatrix}.$$

$$\begin{aligned} \boldsymbol{U}_1(\boldsymbol{U}_2^T\boldsymbol{U}_2)^{-1}\boldsymbol{U}_1^T &= \frac{1}{A}\begin{bmatrix} 1 & u_k \end{bmatrix}\begin{bmatrix} \sum_{i=k+1}^{N} u_i^2 & -\sum_{i=k+1}^{N} u_i \\ -\sum_{i=k+1}^{N} u_i & N-k \end{bmatrix}\begin{bmatrix} 1 \\ u_k \end{bmatrix} \\ &= \frac{1}{A}\sum_{i=k+1}^{N}(u_i-u_k)^2 \end{aligned}$$

$$d = \boldsymbol{I} + \boldsymbol{U}_1(\boldsymbol{U}_2^T\boldsymbol{U}_2)^{-1}\boldsymbol{U}_1^T = \frac{A+\sum_{i=k+1}^{N}(u_i-u_k)^2}{A}$$

$$\frac{(\boldsymbol{U}_2^T\boldsymbol{U}_2)^{-1}\boldsymbol{U}_1^T\boldsymbol{U}_1}{d} = \frac{1}{A+\sum_{i=k+1}^{N}(u_i-u_k)^2}\begin{bmatrix} \sum_{i=k+1}^{N} u_i(u_i-u_k) & \sum_{i=k+1}^{N} u_k u_i(u_i-u_k) \\ \sum_{i=k+1}^{N}(u_k-u_i) & \sum_{i=k+1}^{N} u_k(u_k-u_i) \end{bmatrix}$$

$$\boldsymbol{I} - \frac{(\boldsymbol{U}_2^T\boldsymbol{U}_2)^{-1}\boldsymbol{U}_1^T\boldsymbol{U}_1}{d} = \frac{1}{A+\sum_{i=k-1}^{N}(u_i-u_k)^2}\begin{bmatrix} A+\sum_{i=k+1}^{N} u_k(u_k-u_i) & -\sum_{i=k+1}^{N} u_k u_i(u_i-u_k) \\ -\sum_{i=k+1}^{N}(u_k-u_i) & A+\sum_{i=k+1}^{N} u_i(u_i-u_k) \end{bmatrix}$$

$$(\boldsymbol{U}_2^T\boldsymbol{U}_2)^{-1}\boldsymbol{U}_1^T\bar{\boldsymbol{y}}_1 = \frac{1}{A}\begin{bmatrix} \sum_{i=k+1}^{N} u_i^2 & -\sum_{i=k+1}^{N} u_i \\ -\sum_{i=k+1}^{N} u_i & N-k \end{bmatrix}\begin{bmatrix} 1 \\ u_k \end{bmatrix}\bar{\boldsymbol{y}}_k = \frac{\bar{\boldsymbol{y}}_k}{A}\begin{bmatrix} \sum_{i=k+1}^{N} u_i(u_i-u_k) \\ \sum_{i=k+1}^{N}(u_k-u_i) \end{bmatrix}$$

So, $\left(\boldsymbol{I}-\frac{(\boldsymbol{U}_2^T\boldsymbol{U}_2)^{-1}\boldsymbol{U}_1^T\boldsymbol{U}_1}{d}\right)(\boldsymbol{U}_2^T\boldsymbol{U}_2)^{-1}\boldsymbol{U}_1^T\bar{\boldsymbol{y}}_1 = \frac{\bar{y}_k}{A+\sum_{i=k+1}^{N}(u_i-u_k)^2}\begin{bmatrix}\sum_{i=k+1}^{N}u_i(u_i-u_k)\\ \sum_{i=k+1}^{N}(u_k-u_i)\end{bmatrix}$

and $\frac{(\boldsymbol{U}_2^T\boldsymbol{U}_2)^{-1}\boldsymbol{U}_1^T\boldsymbol{U}_1}{d}\hat{\boldsymbol{\omega}}_R = (\hat{\omega}_3+u_k\hat{\omega}_4)\begin{bmatrix}\sum_{i=k+1}^{N}u_i(u_i-u_k)\\ \sum_{i=k+1}^{N}(u_k-u_i)\end{bmatrix}$.

B.1 Bivariate Normal Calculation for Theorem 4.2

Suppose that $\tau_0 = u_l$ for some $l = 1,\ldots,N$. To be consistent with the statement of the comprehensive algorithm, let $l = k$. Then, as $m \to \infty$, the asymptotic normality is described in Theorem 3.7 as

$$\begin{bmatrix}\hat{\tau}_j\\ \hat{\tau}_k\end{bmatrix} \to_d \mathcal{N}\left(\begin{bmatrix}\tau_0\\ \tau_0\end{bmatrix}, \frac{\sigma_0^2}{m}(\boldsymbol{a}^*)^T\Sigma^*\boldsymbol{a}^*\right),$$

where $\frac{\sigma_0^2}{m}(\boldsymbol{a}^*)^T\Sigma^*\boldsymbol{a}^* = \begin{bmatrix}\sigma_j^2 & \mathrm{Cov}(\hat{\tau}_j,\hat{\tau}_k)\\ \mathrm{Cov}(\hat{\tau}_k,\hat{\tau}_j) & \sigma_k^2\end{bmatrix}$ and $\rho = \frac{\mathrm{Cov}(\hat{\tau}_j,\hat{\tau}_k)}{\sqrt{\sigma_j^2\sigma_k^2}}$.

$$\begin{aligned}
P(B1) &= P(u_k \in C_{\alpha,j} \;\&\; u_k \in C_{\alpha,k})\\
&= P\big[(u_k - z_{\frac{\alpha}{2}}\sigma_j < \hat{\tau}_j < u_k + z_{\frac{\alpha}{2}}\sigma_j) \;\&\; (u_k - z_{\frac{\alpha}{2}}\sigma_k < \hat{\tau}_k < u_k + z_{\frac{\alpha}{2}}\sigma_k)\big]\\
P(B3) &= P(u_k \in C_{\alpha,j} \;\&\; C_{\alpha,k} \subset (u_j, u_k))\\
&= P\big[(u_k - z_{\frac{\alpha}{2}}\sigma_j < \hat{\tau}_j < u_k + z_{\frac{\alpha}{2}}\sigma_j) \;\&\; (u_j + z_{\frac{\alpha}{2}}\sigma_k < \hat{\tau}_k < u_k - z_{\frac{\alpha}{2}}\sigma_k)\big]\\
P(B4) &= P(C_{\alpha,j} \subset (u_k, u_{k+1}) \;\&\; u_k \in C_{\alpha,k})\\
&= P\big[(u_k + z_{\frac{\alpha}{2}}\sigma_j < \hat{\tau}_j < u_{k+1} - z_{\frac{\alpha}{2}}\sigma_j) \;\&\; (u_k - z_{\frac{\alpha}{2}}\sigma_k < \hat{\tau}_k < u_k + z_{\frac{\alpha}{2}}\sigma_k)\big]
\end{aligned}$$

Graphically, these probabilities are the same as the volume under the bivariate normal curve for the shaded regions shown in Figure B.3.

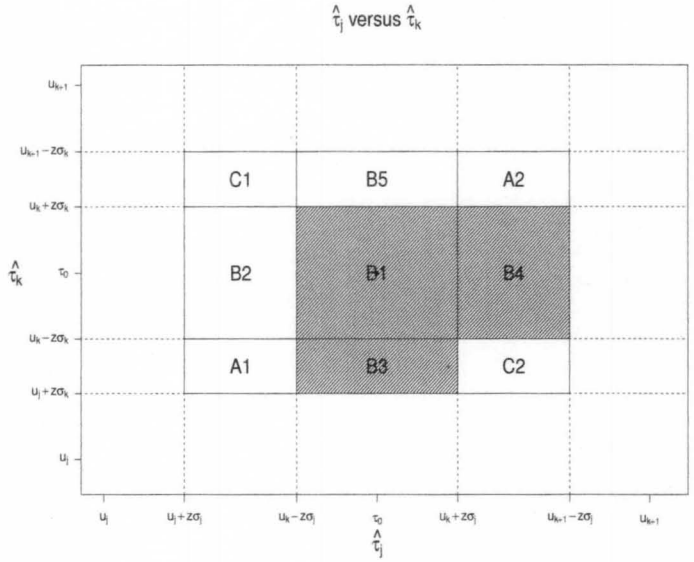


Figure B.3: Shaded regions for integration.

Since the 9 regions shown in the graph above are the only possibilities for the location of the estimators, the volume of the bivariate normal curve over all 9 regions is equal to 1. Let p_1 be the volume over the regions B1, B3, B4, and C1 and let p_2 be the volume over the region C1. Thus, the region of interest is $p_1 - p_2$, which can be calculated using a series of transformations and rotations of the curve. For the sake of simplicity, denote $z_{\frac{\alpha}{2}} = z$. Then, p_1 and p_2 are equal to the following integrals, respectively.

$$\int_{-\infty}^{u_k+z\sigma_k} \int_{u_k-z\sigma_j}^{\infty} \frac{1}{2\pi\sigma_j\sigma_k\sqrt{1-\rho^2}} e^{\frac{-1}{2(1-\rho^2)}\left(\frac{(\hat{\tau}_j-\tau_0)^2}{\sigma_j^2} - \frac{2\rho(\hat{\tau}_j-\tau_0)(\hat{\tau}_k-\tau_0)}{\sigma_j\sigma_k} + \frac{(\hat{\tau}_k-\tau_0)^2}{\sigma_k^2}\right)} \mathrm{d}\hat{\tau}_j \mathrm{d}\hat{\tau}_k$$

$$\int_{-\infty}^{u_k-z\sigma_k} \int_{u_k+z\sigma_j}^{\infty} \frac{1}{2\pi\sigma_j\sigma_k\sqrt{1-\rho^2}} e^{\frac{-1}{2(1-\rho^2)}\left(\frac{(\hat{\tau}_j-\tau_0)^2}{\sigma_j^2} - \frac{2\rho(\hat{\tau}_j-\tau_0)(\hat{\tau}_k-\tau_0)}{\sigma_j\sigma_k} + \frac{(\hat{\tau}_k-\tau_0)^2}{\sigma_k^2}\right)} \mathrm{d}\hat{\tau}_j \mathrm{d}\hat{\tau}_k$$

The first transformation will be made by letting $x_1 = \frac{\hat{\tau}_j-\tau_0}{\sigma_j}$ and $x_2 = \frac{\hat{\tau}_k-\tau_0}{\sigma_k}$. Then, $\mathrm{d}x_1 = \frac{1}{\sigma_j}\mathrm{d}\hat{\tau}_j$ and $\mathrm{d}x_2 = \frac{1}{\sigma_k}\mathrm{d}\hat{\tau}_k$. The limits of integration for p_1 are transformed as follows:

$$\begin{aligned} u_k - x\sigma_j < \hat{\tau}_j < \infty \quad &\Rightarrow \quad u_k - z\sigma_j < x_1\sigma_j + \tau_0 < \infty \\ &\Rightarrow \quad -z < x_1 < \infty \end{aligned}$$

$$-\infty < \hat{\tau}_k < u_k + z\sigma_k \;\Rightarrow\; -\infty < x_2\sigma_k + \tau_0 < u_k + z\sigma_k$$
$$\Rightarrow\; -\infty < x_2 < z$$

The limits of integration for p_2 can be found similarly and the integrals can be rewritten as follows.

$$p_1 = \int_{-\infty}^{z}\int_{-z}^{\infty} \frac{1}{2\pi\sqrt{1-\rho^2}} e^{\frac{-1}{2(1-\rho^2)}(x_1^2 - 2\rho x_1 x_2 + x_2^2)}\, \mathrm{d}x_1 \mathrm{d}x_2$$
$$p_2 = \int_{-\infty}^{-z}\int_{z}^{\infty} \frac{1}{2\pi\sqrt{1-\rho^2}} e^{\frac{-1}{2(1-\rho^2)}(x_1^2 - 2\rho x_1 x_2 + x_2^2)}\, \mathrm{d}x_1 \mathrm{d}x_2$$

This can be visualized by the graph shown in Figure B.4.

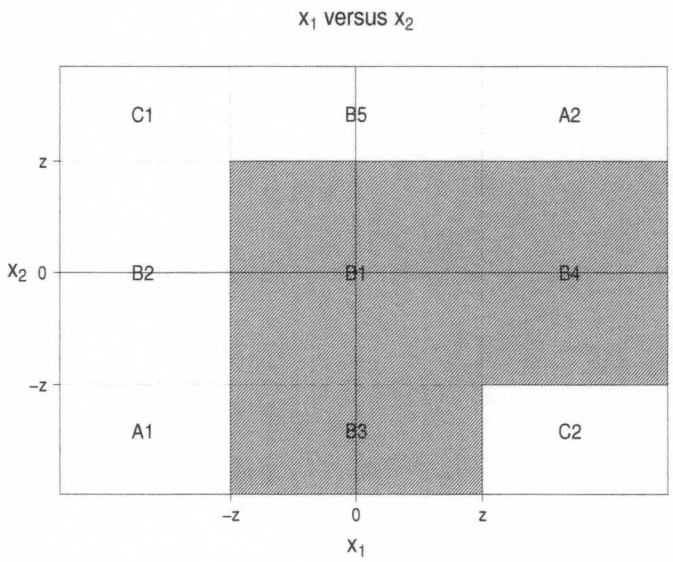


Figure B.4: First transformation

The distribution for x_1 and x_2 can be found as follows:

$$E(x_1) = E(\tfrac{\hat{\tau}_j - \tau_0}{\sigma_j}) = \tfrac{1}{\sigma_j}E(\hat{\tau}_j) - \tfrac{\tau_0}{\sigma_j} = \tfrac{\tau_0}{\sigma_j} - \tfrac{\tau_0}{\sigma_j} = 0$$
$$\mathrm{var}(x_1) = \mathrm{var}(\tfrac{\hat{\tau}_j - \tau_0}{\sigma_j}) = \tfrac{1}{\sigma_j^2}\mathrm{var}(\hat{\tau}_j) = \tfrac{1}{\sigma_j^2}\sigma_j^2 = 1$$

Similarly, $E(x_2) = 0$ and $\mathrm{var}(x_2) = 1$. Since x_1 and x_2 are not independent,

$$\mathrm{Cov}(x_1, x_2) = \mathrm{Cov}(\tfrac{\hat{\tau}_j - \tau_0}{\sigma_j}, \tfrac{\hat{\tau}_k - \tau_0}{\sigma_k}) = \tfrac{1}{\sigma_j \sigma_k}\mathrm{Cov}(\hat{\tau}_j, \hat{\tau}_k) = \tfrac{1}{\sigma_j \sigma_k}\sigma_j \sigma_k \rho = \rho$$

Thus, $\boldsymbol{X} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, where $\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$. Since $|\boldsymbol{\Sigma}| = 1 - \rho^2 \Rightarrow |\boldsymbol{\Sigma}|^{1/2} = \sqrt{1-\rho^2}$ and

$$\boldsymbol{\Sigma}^{-1} = \tfrac{1}{1-\rho^2}\begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix} \Rightarrow$$

$$\begin{aligned} \boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X} &= \frac{1}{1-\rho^2}\begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= \frac{1}{1-\rho^2}\begin{bmatrix} x_1 - \rho x_2 & -\rho x_1 + x_2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= \frac{1}{1-\rho^2}(x_1^2 - 2\rho x_1 x_2 + x_2^2), \end{aligned}$$

then p_1 and p_2 can be re-written as follows.

$$p_1 = \int_{-\infty}^{z}\int_{-z}^{\infty} \frac{1}{2\pi|\boldsymbol{\Sigma}|^{1/2}} e^{-\frac{1}{2}\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X}}\,\mathrm{d}\boldsymbol{X}$$

$$p_2 = \int_{-\infty}^{-z}\int_{z}^{\infty} \frac{1}{2\pi|\boldsymbol{\Sigma}|^{1/2}} e^{-\frac{1}{2}\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X}}\,\mathrm{d}\boldsymbol{X}$$

In order to make the second transformation in the form of a rotation, the eigenvalues and eigenvectors of $\boldsymbol{\Sigma}$ must be found. The eigenvalues are found by the following calculation.

$$\begin{aligned} |\boldsymbol{\Sigma} - \lambda\boldsymbol{I}| &= 0 \\ \begin{vmatrix} 1-\lambda & \rho \\ \rho & 1-\lambda \end{vmatrix} &= 0 \\ (1-\lambda)^2 - \rho^2 &= 0 \\ \lambda^2 - 2\lambda + (1-\rho)^2 &= 0 \\ \lambda = \tfrac{2 \pm \sqrt{4-4(1-\rho^2)}}{2} &= 1 \pm \rho \end{aligned}$$

The eigenvector corresponding to the first eigenvalue, $1+\rho$, can be found by considering

$$\begin{aligned}
(\boldsymbol{\Sigma} - (1+\rho)\boldsymbol{I})\boldsymbol{V} &= \mathbf{0} \\
\left(\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} - \begin{bmatrix} 1+\rho & 0 \\ 0 & 1+\rho \end{bmatrix} \right) \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} &= \mathbf{0} \\
\begin{bmatrix} -\rho & \rho \\ \rho & -\rho \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} &= \mathbf{0} \\
\rho \begin{bmatrix} v_2 - v_1 \\ v_1 - v_2 \end{bmatrix} &= \mathbf{0}.
\end{aligned}$$

This implies that $v_1 = v_2$. So for any constant $c \in \mathbb{R}$, the eigenvector corresponding to the first eigenvalue is $c \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Similarly, the eigenvector corresponding to the second eigenvalue, $1-\rho$, is found by considering

$$\begin{aligned}
(\boldsymbol{\Sigma} - (1-\rho)\boldsymbol{I})\boldsymbol{V} &= \mathbf{0} \\
\left(\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} - \begin{bmatrix} 1-\rho & 0 \\ 0 & 1-\rho \end{bmatrix} \right) \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} &= \mathbf{0} \\
\begin{bmatrix} \rho & \rho \\ \rho & \rho \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} &= \mathbf{0} \\
\rho \begin{bmatrix} v_1 + v_2 \\ v_1 + v_2 \end{bmatrix} &= \mathbf{0}
\end{aligned}$$

which implies that $v_1 = -v_2$. The eigenvector that corresponds to the second eigenvalue is $c \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ for any constant $c \in \mathbb{R}$.

This is used to find the Singular Value Decomposition (SVD) of $\boldsymbol{\Sigma}$. Letting

$\boldsymbol{U} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$ and $\boldsymbol{D} = \begin{bmatrix} 1+\rho & 0 \\ 0 & 1-\rho \end{bmatrix}$, the SVD is $\boldsymbol{\Sigma} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^T$ because

$$\begin{aligned} \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^T &= \frac{1}{2} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1+\rho & 0 \\ 0 & 1-\rho \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \\ &= \frac{1}{2} \begin{bmatrix} 1+\rho & \rho-1 \\ 1+\rho & 1-\rho \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \\ &= \frac{1}{2} \begin{bmatrix} 2 & 2\rho \\ 2\rho & 2 \end{bmatrix} \\ &= \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \\ &= \boldsymbol{\Sigma}. \end{aligned}$$

Using the fact the $\boldsymbol{D}$ is diagonal and $\boldsymbol{U}$ is orthogonal, $\boldsymbol{\Sigma}$ can be written as

$$\begin{aligned} \boldsymbol{\Sigma} &= \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^T \\ &= \boldsymbol{U}\boldsymbol{D}^{1/2}\boldsymbol{D}^{1/2}\boldsymbol{U}^T \\ &= \boldsymbol{U}\boldsymbol{D}^{1/2}\boldsymbol{U}^T\boldsymbol{U}\boldsymbol{D}^{1/2}\boldsymbol{U}^T \\ &= \boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{1/2} \end{aligned}$$

where $\boldsymbol{\Sigma}^{1/2} = \boldsymbol{U}\boldsymbol{D}^{1/2}\boldsymbol{U}^T$. Finally, the second transformation can be made by letting

$\boldsymbol{W} = \boldsymbol{D}^{-1/2}\boldsymbol{U}^T\boldsymbol{X}$. This means that $\boldsymbol{W} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ because

$$\begin{aligned}
E(\boldsymbol{W}) &= \boldsymbol{D}^{1/2}\boldsymbol{U}^T E(\boldsymbol{X}) = \boldsymbol{0} \\
Var(\boldsymbol{W}) &= (\boldsymbol{D}^{-1/2}\boldsymbol{U}^T)Var(\boldsymbol{X})(\boldsymbol{D}^{-1/2}\boldsymbol{U}^T)^T \\
&= \boldsymbol{D}^{-1/2}\boldsymbol{U}^T\boldsymbol{\Sigma}\boldsymbol{U}\boldsymbol{D}^{-1/2} \\
&= \boldsymbol{U}^T\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{U} \\
&= \boldsymbol{U}^T\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{U} \\
&= \boldsymbol{U}^T\boldsymbol{U} \\
&= \boldsymbol{I}.
\end{aligned}$$

This transformation is useful because the integrand is now $f(w_1, w_2) = \frac{1}{2\pi}e^{-\frac{1}{2}(w_1^2+w_2^2)}$. The region of integration is transformed as follows. It can easily be seen that $w_1 = \frac{x_1+x_2}{\sqrt{2}\sqrt{1+\rho}}$ and $w_2 = \frac{x_2-x_1}{\sqrt{2}\sqrt{1-\rho}}$ by

$$\begin{aligned}
\boldsymbol{W} &= \boldsymbol{D}^{-1/2}\boldsymbol{U}^T\boldsymbol{X} \\
\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} &= \frac{1}{\sqrt{2}} \begin{bmatrix} \frac{1}{\sqrt{1+\rho}} & 0 \\ 0 & \frac{1}{\sqrt{1-\rho}} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\
&= \frac{1}{\sqrt{2}} \begin{bmatrix} \frac{1}{\sqrt{1+\rho}} & \frac{1}{\sqrt{1+\rho}} \\ -\frac{1}{\sqrt{1-\rho}} & \frac{1}{\sqrt{1-\rho}} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\
&= \begin{bmatrix} \frac{x_1+x_2}{\sqrt{2}\sqrt{1+\rho}} \\ \frac{x_2-x_1}{\sqrt{2}\sqrt{1-\rho}} \end{bmatrix}.
\end{aligned}$$

However, the transformation for the region of integration, $p_1 - p_2$, is more complicated than before. First, consider mapping the region p_1 from the $\boldsymbol{X}$-plane to the $\boldsymbol{W}$-plane by letting $x_1 = t$ for $t \geq -z$ and let $x_2 = s$ for $s \leq z$. Then, $\forall t \geq -z \; \forall s \leq z$, $w_1 = \frac{t+s}{\sqrt{2}\sqrt{1+\rho}}$ and $w_2 = \frac{s-t}{\sqrt{2}\sqrt{1-\rho}}$. For a fixed value of s, it is useful to map all points $t \geq -z$ from the $\boldsymbol{X}$-plane to the $\boldsymbol{W}$-plane. To visualize how these points will map to the new plane, consider in detail how two such points are mapped for a fixed value of s.

For example, fix $s = z$. Then, $w_1 = \frac{t+z}{\sqrt{2}\sqrt{1+\rho}}$ and $w_2 = \frac{z-t}{\sqrt{2}\sqrt{1-\rho}}$. Then the value $t = -z$ gives $w_1 = 0$ and $w_2 = \frac{z\sqrt{2}}{\sqrt{1-\rho}}$. This means that the point $(x_1 = -z, x_2 = z)$ in the $\boldsymbol{X}$-plane maps to $(w_1 = 0, w_2 = \frac{z\sqrt{w}}{\sqrt{1-\rho}})$ in the $\boldsymbol{W}$-plane. To map a second point for this fixed value of s, let $t = z$. Then, the point (z, z) in the $\boldsymbol{X}$-plane maps to the point $(\frac{z\sqrt{2}}{\sqrt{1+\rho}}, 0)$ in the $\boldsymbol{W}$-plane. Consider the visualization of this in Figures B.5(a) and B.5(b) shown below.

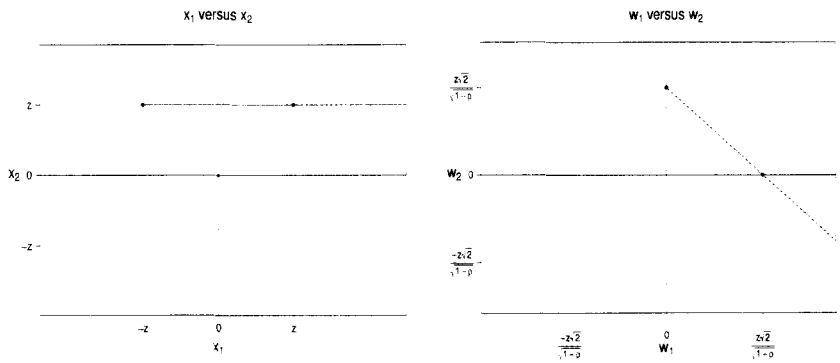


Figure B.5: (a) original image in $\boldsymbol{X}$-plane (b) transformed image in $\boldsymbol{W}$-plane

Now, if $s = -z$, the mapping of the following two points from the $\boldsymbol{X}$-plane to the $\boldsymbol{W}$-plane, respectively, can be found similarly.

$$(-z, z) \text{ maps to } (\tfrac{-z\sqrt{2}}{\sqrt{1+\rho}}, 0)$$
$$(z, -z) \text{ maps to } (0, \tfrac{-z\sqrt{2}}{\sqrt{1-\rho}})$$

This can be seen visually in Figures B.6(a) and B.6(b) below.

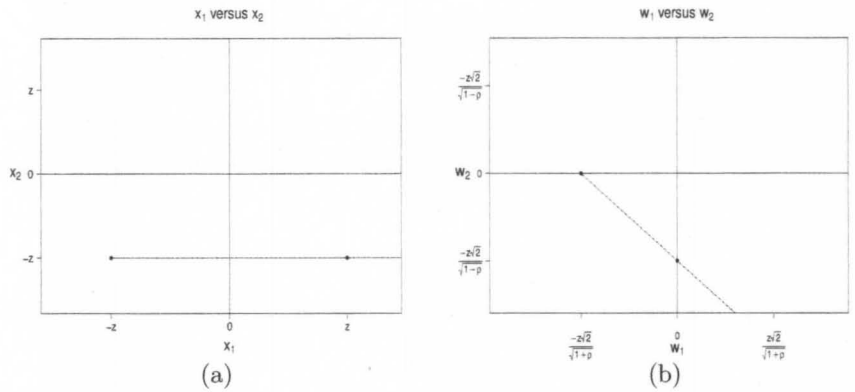


Figure B.6: (a) original image in $\boldsymbol{X}$-plane (b) transformed image in $\boldsymbol{W}$-plane

Consider doing this for all values of $s \leq z$. Then, the transformation of region p_1 is shown in Figures B.7(a) and B.7(b) below.

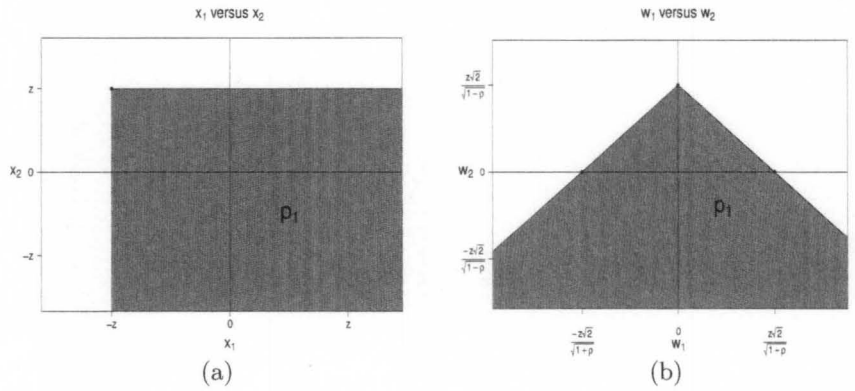
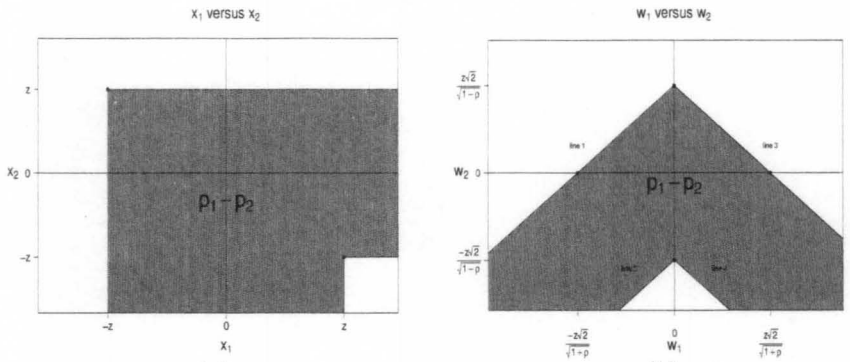


Figure B.7: (a) original p_1 in $\boldsymbol{X}$-plane (b) original p_1 in $\boldsymbol{W}$-plane

The same method can be used to transform the region p_2. The region of integration, $p_1 - p_2$, is shown shaded in Figures B.8(a) and B.8(b).



(a) (b)

Figure B.8: (a) $p_1 - p_2$ in $\boldsymbol{X}$-plane (b) $p_1 - p_2$ in $\boldsymbol{W}$-plane

Letting $r = \sqrt{\frac{1-\rho}{1+\rho}}$, $s = \frac{z\sqrt{2}}{\sqrt{1+\rho}}$, and $t = \frac{z\sqrt{2}}{\sqrt{1-\rho}}$, the labeled lines in figure B.8(b) can be written in the following way.

$$\begin{aligned}
\text{Line } 1 : w_1 &= rw_2 - s \\
\text{Line } 2 : w_1 &= rw_2 + s \\
\text{Line } 3 : w_1 &= -rw_2 + s \\
\text{Line } 4 : w_1 &= -rw_2 - s
\end{aligned}$$

Now the integrals can be simplified as shown below.

$$\begin{aligned}
p_1 &= \int_{-\infty}^{t} \int_{rw_2 - s}^{-rw_2 + s} \frac{1}{2\pi} e^{-\frac{1}{2}(w_1^2 + w_2^2)} \mathrm{d}w_1 \mathrm{d}w_2 \\
p_2 &= \int_{-\infty}^{-t} \int_{rw_2 + s}^{-rw_2 - s} \frac{1}{2\pi} e^{-\frac{1}{2}(w_1^2 + w_2^2)} \mathrm{d}w_1 \mathrm{d}w_2
\end{aligned}$$

Since it is useful for the difference between p_1 and p_2 to simplify nicely, make the following substitution on p_2. Let $\boldsymbol{W} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \mathbf{V}$. This means that $w_1 = v_1$ and $w_2 = -v_2$ and

$$\begin{aligned}
p_2 &= \int_{\infty}^{t} \int_{-rv_2 + s}^{rv_2 - s} \frac{1}{2\pi} e^{\frac{1}{2}(v_1^2 + v_2^2)} \mathrm{d}v_1 (-\mathrm{d}v_2) \\
&= (-)(-)(-) \int_{t}^{\infty} \int_{rv_2 - s}^{-rv_2 + s} \frac{1}{2\pi} e^{\frac{1}{2}(v_1^2 + v_2^2)} \mathrm{d}v_1 \mathrm{d}v_2 \\
&= -\int_{t}^{\infty} \int_{rv_2 - s}^{-rv_2 + s} \frac{1}{2\pi} e^{\frac{1}{2}(v_1^2 + v_2^2)} \mathrm{d}v_1 \mathrm{d}v_2
\end{aligned}$$

Substituting back to $\mathbf{V} = \boldsymbol{W}$ gives $p_2 = -\int_t^{\infty} \int_{rw_2-s}^{-rw_2+s} \frac{1}{2\pi} e^{\frac{1}{2}(w_1^2+w_2^2)} \mathrm{d}w_1 \mathrm{d}w_2$. This means $p_1 - p_2$ is equal to the following.

$$
\begin{aligned}
&= \int_{-\infty}^{t} \int_{rw_2-s}^{-rw_2+s} \frac{1}{2\pi} e^{\frac{1}{2}(w_1^2+w_2^2)} \mathrm{d}w_1 \mathrm{d}w_2 + \int_{t}^{\infty} \int_{rw_2-s}^{-rw_2+s} \frac{1}{2\pi} e^{\frac{1}{2}(w_1^2+w_2^2)} \mathrm{d}w_1 \mathrm{d}w_2 \\
&= \int_{-\infty}^{\infty} \int_{rw_2-s}^{-rw_2+s} \frac{1}{2\pi} e^{\frac{1}{2}(w_1^2+w_2^2)} \mathrm{d}w_1 \mathrm{d}w_2 \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}w_2^2} \left[\int_{rw_2-s}^{-rw_2+s} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}w_1^2} \mathrm{d}w_1 \right] \mathrm{d}w_2 \\
&= \int_{-\infty}^{\infty} \Phi'(w_2) \Big[1 - 2\Phi(rw_2 - s) \Big] \mathrm{d}w_2 \\
&= \int_{-\infty}^{\infty} \Phi'(w_2) \mathrm{d}w_2 - 2 \int_{-\infty}^{\infty} \Phi'(w_2) \Phi(rw_2 - s) \mathrm{d}w_2
\end{aligned}
$$

where Φ is the CDF (cumulative distribution function) of the standard normal distribution. Thus, $p_1 - p_2 = 1 - 2I$ where $I = \int_{\infty}^{\infty} \Phi'(w_2)\Phi(rw_2 - s)\mathrm{d}w_2$. The integral I will be considered in more detail by substituting the original values back in for r and s.

$$
\begin{aligned}
I &= \int_{-\infty}^{\infty} \Phi'(w_2) \Phi\left(\sqrt{\frac{1-\rho}{1+\rho}} w_2 - \frac{z\sqrt{2}}{\sqrt{1+\rho}} \right) \mathrm{d}w_2 \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}w_2^2} \int_{-\infty}^{\sqrt{\frac{1-\rho}{1+\rho}} w_2 - \frac{z\sqrt{2}}{\sqrt{1+\rho}}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}w_1^2} \mathrm{d}w_1 \mathrm{d}w_2
\end{aligned}
$$

The integral, I, is the volume of the bivariate normal curve above the shaded region shown below in Figure B.9.

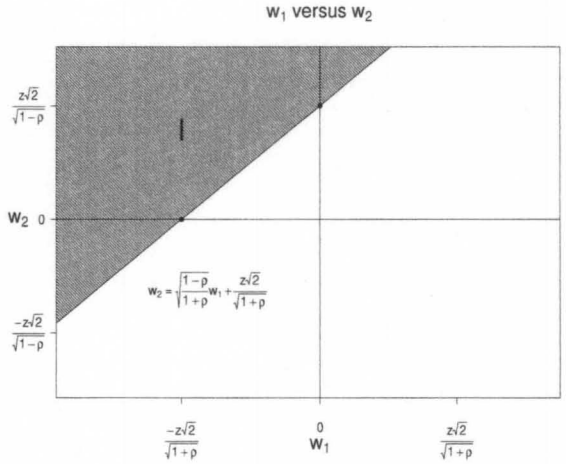


Figure B.9: Region of integration for I.

To transform the region of integration shown in Figure B.9, consider $\boldsymbol{U} = \boldsymbol{AW}$ where

$$\boldsymbol{A} = \begin{bmatrix} \sin\theta & -\cos\theta \\ \cos\theta & \sin\theta \end{bmatrix},$$

using θ as the angle between the w_1-axis and the line $w_2 = \sqrt{\frac{1-\rho}{1+\rho}}w_1 + \frac{z\sqrt{2}}{\sqrt{1+\rho}}$. This is shown in Figure B.10 below.

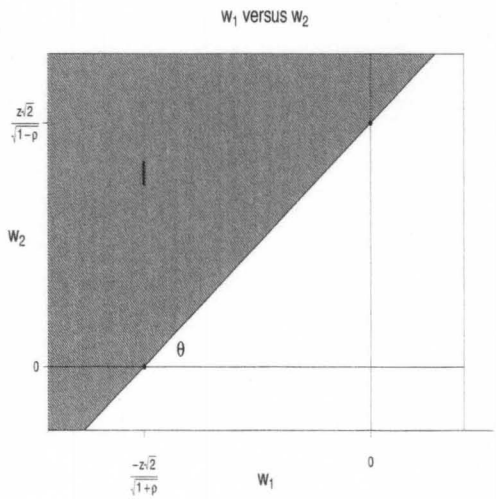


Figure B.10: Angle θ.

Then, $\boldsymbol{A}$ can be simplified as follows.

$$\boldsymbol{A} = \frac{1}{\sqrt{2}} \begin{bmatrix} \sqrt{1+\rho} & -\sqrt{1-\rho} \\ \sqrt{1-\rho} & \sqrt{1+\rho} \end{bmatrix}$$

Now, it can be shown that $\boldsymbol{U} \sim N(\boldsymbol{0}, \boldsymbol{I})$ because $E(\boldsymbol{U}) = \boldsymbol{A} \cdot E(\boldsymbol{W}) = \boldsymbol{0}$ and

$$\begin{aligned} Var(\boldsymbol{U}) = \boldsymbol{A} Var(\boldsymbol{W}) \boldsymbol{A}^T &= \boldsymbol{A}\boldsymbol{A}^T \\ &= \frac{1}{2} \begin{bmatrix} \sqrt{1+\rho} & -\sqrt{1-\rho} \\ \sqrt{1-\rho} & \sqrt{1+\rho} \end{bmatrix} \begin{bmatrix} \sqrt{1+\rho} & \sqrt{1-\rho} \\ -\sqrt{1-\rho} & \sqrt{1+\rho} \end{bmatrix} \\ &= \frac{1}{2} \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \\ &= \boldsymbol{I}. \end{aligned}$$

Since $\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} w_1\sqrt{1+\rho} - w_2\sqrt{1-\rho} \\ w_1\sqrt{1-\rho} + w_2\sqrt{1+\rho} \end{bmatrix}$, the transformation can be made from the $\boldsymbol{W}$-plane to the $\boldsymbol{U}$-plane similar to what was done before. To form the region of integration, I, shown in Figure B.9, let $w_1 = t$ for $\infty < t < \infty$ and $w_2 = \sqrt{\frac{1+\rho}{1-\rho}} \cdot t + s$ for $s \geq \frac{z\sqrt{2}}{\sqrt{1-\rho}}$. By substitution, for $t \in (-\infty, \infty)$ and $s \in [\frac{z\sqrt{2}}{\sqrt{1-\rho}}, \infty)$,

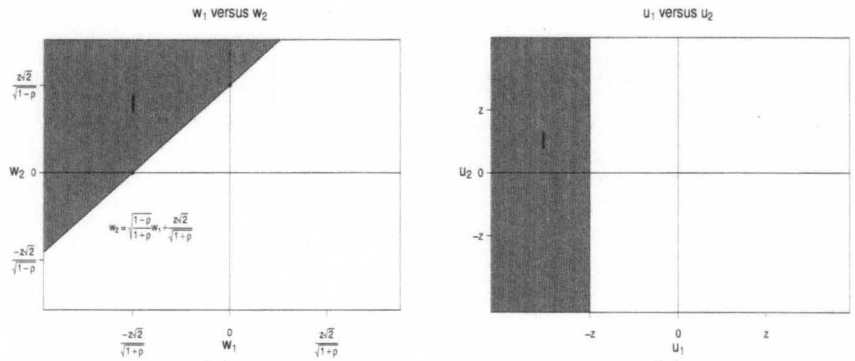
$$\begin{aligned} u_1 &= -\sqrt{\frac{1-\rho}{2}} \cdot s \\ u_2 &= \sqrt{\frac{2}{1-\rho}} \cdot t + \sqrt{\frac{1-\rho}{2}} \cdot s \end{aligned}$$

The shaded region, I, shown in Figure B.9 can be mapped to the $\boldsymbol{U}$-plane by considering how the line $w_2 = \sqrt{\frac{1+\rho}{1-\rho}} \cdot t + s$ is mapped for a fixed value of $s \in (\frac{z\sqrt{2}}{\sqrt{1-\rho}}, \infty)$ as t ranges from $-\infty$ to ∞. For example, fix $s = \frac{z\sqrt{2}}{\sqrt{1-\rho}}$. Then,

$$\begin{aligned} u_1 &= -z \\ u_2 &= \sqrt{\frac{2}{1-\rho}} \cdot t + z \end{aligned}$$

As t ranges from $-\infty$ to ∞, u_1 is fixed at $-z$ and u_2 takes on all values from $-\infty$ to ∞. So the line $w_2 = \sqrt{\frac{1+\rho}{1-\rho}} \cdot t + \frac{z\sqrt{2}}{\sqrt{1-\rho}}$ in the $\boldsymbol{W}$-plane is mapped to the vertical

line, $u_1 = z$, in the $\boldsymbol{U}$-plane. This can be done for all values of $s \in [\frac{z\sqrt{2}}{\sqrt{1-\rho}}, \infty)$. This maps the region I from the $\boldsymbol{W}$-plane to the $\boldsymbol{U}$-plane as shown in Figures B.11(a) and B.11(b) below.



(a) (b)

Figure B.11: (a) I in $\boldsymbol{W}$-plane (b) I in $\boldsymbol{U}$-plane

Then, the integral I can be calculated as follows.

$$\begin{aligned} I &= \int_{-\infty}^{\infty}\int_{-\infty}^{-z} \frac{1}{2\pi} e^{-\frac{1}{2}(u_1^2+u_2^2)} \mathrm{d}u_1 \mathrm{d}u_2 \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u_2^2} \left(\int_{-\infty}^{-z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u_1^2} \mathrm{d}u_1 \right) \mathrm{d}u_2 \\ &= \frac{\alpha}{2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u_2^2} \mathrm{d}u_2 \\ &= \frac{\alpha}{2} \end{aligned}$$

Finally, substituting this value for I into the equation we have for $p_1 - p_2$ gives

$$\begin{aligned} p_1 - p_2 &= 1 - 2I \\ &= 1 - 2\left(\frac{\alpha}{2}\right) \\ &= 1 - \alpha. \end{aligned}$$

CURRICULUM VITAE

Rebekah Ann Robinson

Education

University of Louisville, Louisville, Kentucky
Ph.D., Applied and Industrial Mathematics, expected Spring 2012
M.A., Mathematics, May 2008

University of Evansville, Evansville, Indiana
B.S., Pre-doctoral Mathematics, May 2002

Professional Experience

University of Louisville, Louisville, Kentucky
Dual Credit Coordinator, Department of Mathematics, 2011–2012

University of Louisville, Louisville, Kentucky
Graduate Teaching Assistant, Department of Mathematics, 2006–2007, 2009–2011

University of Louisville, Louisville, Kentucky
National Science Foundation GK-12 Fellow, 2007–2009

Scholarly Activities

Graduate Dean's Citation, University of Louisville, 2012

Credit for Exam FM, Society of Actuaries, 2011

Credit for Exam P, Society of Actuaries, 2010

Ken F. and Sandra S. Hohman Fellowship, University of Louisville, 2008

Faculty Favorite Nomination, University of Louisville, 2007

Trustee Scholarship, University of Evansville, 2002–2006