POLYNOMIAL SPLINE REGRESSION WITH UNKNOWN KNOTS
AND AR(1) ERRORS

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the Graduate
School of the Ohio State University

By

Shih-huang Chan, BBA, M.S.

* * * * *
The Ohio State University
1989

Dissertation Committee:
Professor Sue Leurgans
Professor Joe Verducci
Professor Elizbeth Stasny

Approved by

Sue Leurgans
Adviser
Department of Statistics
Copyright by
Shih-huang Chan
1989
To My Wife and Children
ACKNOWLEDGMENTS

I wish to express my deep appreciation to my advisor, Dr. Sue Leurgans, for her guidance and patience during the researching and writing of the dissertation. Thanks also go to the other members of my advisory committee, Dr. Joseph Verducci and Dr. Elizabeth Stasny, for their valuable suggestions and comments. To my wife, Suchen, I offer sincere thanks for her unshakable faith in me. Her encouragement helps me go through the past four years. To my children, Ruby and Peter, I thank them for letting me enjoy the delight and share the warmth they bring to the family.
VITA

Dec. 18, 1955 .......................... Born-Touliu, Taiwan, Republic of China

1978 ................................. B.B.A. National Taiwan University,
Taipei, Taiwan, R.O.C.

1983 ................................. M.S. National Cheng-chi University,
Taipei, Taiwan, R.O.C.

1984-1988 ............................ Teaching Associate
Department of Statistics
The Ohio State University
Columbus, Ohio

1988-Present .......................... Research Associate
Department of Statistics
The Ohio State University
Columbus, Ohio

FIELDS OF STUDY

Major Field: Mathematical Statistics

Study in Statistical Inference: Dr. Jason Hsu

Study in Probability Theory: Dr. R. Bartoszyński

Study in Multivariate Analysis: Dr. Haikady N. Nagaraja

Study in Nonparametric Statistics: Dr. James C. Aubuchon
TABLE OF CONTENTS

DEDICATION .................................................................................. ii
ACKNOWLEDGMENTS .................................................................... iii
VITA ................................................................................................. iv
LIST OF TABLES ........................................................................... vii
LIST OF FIGURES ........................................................................... viii

CHAPTER

I. INTRODUCTION ............................................................................. 1
  §1.1 Parametric spline regression and smoothing spline ........... 2
  §1.2 Applications of spline function ......................................... 7
  §1.3 Literature review ............................................................... 11
  §1.4 Contents of the dissertation ............................................... 13

II. THE MODEL AND PRELIMINARY RESULTS .......................... 15
  §2.1 Model and assumptions ..................................................... 15
  §2.2 Preliminary results ............................................................ 24
  §2.3 Notation ............................................................................... 29

III. ASYMPTOTIC DISTRIBUTIONS : ρ KNOWN ...................... 33
  §3.1 Example ............................................................................... 33
  §3.2 Asymptotic distributions of θ ............................................ 38
    §3.2.1 Deleting process of the pseudo-problem .................. 38
    §3.2.2 Proof that ( ˆθ - θ(0) ) is of order O_P (1/√n) .......... 40
    §3.2.3 Proof that √n ( ˆθ - θ(0) ) is asymptotically normal .... 59
    §3.2.4 Proof that ˆθ is close to θ with the order
of $o_p\left(\frac{1}{\sqrt{n}}\right)$ ................................................. 62

§3.3 Asymptotic distribution of $\hat{\theta}$ ............................................. 74
§3.4 Example ......................................................................... 79
§3.5 Asymptotic distribution of $\hat{\sigma}^2$ ............................................. 86

IV ASYMPTOTIC DISTRIBUTIONS : $\rho$ UNKNOWN ..................... 95

§4.1 Asymptotic distribution of $\hat{\theta}$ and $\hat{\rho}$ ......................... 95
  §4.1.1 Asymptotic distribution of $\hat{\theta}$ ........................................ 95
  §4.1.2 Asymptotic distribution of $\hat{\rho}$ .................................... 105
§4.2 Asymptotic distribution of $\hat{\rho}$ .............................................. 107
§4.3 Asymptotic distribution of $\hat{\sigma}^2$ ............................................. 119

V DATA ANALYSIS ................................................................. 124

§5.1 Angular motion data ............................................................... 124
§5.2 Preliminary analysis ............................................................... 126
§5.3 The procedure used to find the estimates .................................... 137
§5.4 Justification of the approach used in §5.3 .................................... 140
§5.5 Data analysis and the advantage of the approach used in §5.3 .... 146
  §5.5.1 Data analysis and comparisons ........................................... 149
  §5.5.2 Advantage of the approach used in §5.3 .............................. 160

VI CONCLUSIONS AND SUGGESTIONS ..................................... 161

§6.1 Conclusions ........................................................................ 161
§6.2 Suggestions for future work .................................................. 163

APPENDIX ................................................................................. 167

A SAS code for data analysis .................................................... 167
B SAS code for selecting representative runs ............................ 174

REFERENCES ........................................................................... 179
LIST OF TABLES

5.1 Estimates of \( \hat{\theta}_{ij} \)'s : Known Knots at Reported Times and i.i.d. Model ..............................................................132

5.2 Estimates of \( \hat{\theta}_{ij} \)'s : Unknown Knot Points and i.i.d. Model ........133

5.3 Flow Chart of the SAS Code .......................................................138

5.4 Estimates of Regression Coefficients and Asymptotic Standard Errors of \( \hat{\theta}_{ij} \)'s : Unknown Knot Points ...............................150

5.5 Values of \( \hat{\theta} \) for Different Models.............................................155
LIST OF FIGURES

1.1 Plot of Original Data and Predicted Values -- Cubic Splines with Known Knots.................................................................9

1.2 Residual Plot Cubic Splines with Known Knots.................................10

3.1 Mean Function..............................................................................37

3.2 Likelihood Function.....................................................................37

3.3 \( \hat{\theta} \) as Projection of \( e \)..................................................92

5.1 Plot of Difference of Two Mean Functions..................................129

5.2 Plot of Raw Data and Predicted Values -- Generalized Cubic Splines with Unknown Knots..................................................134

5.3 Residual Plot -- Generalized Cubic Splines with Unknown Knots....135

5.4 Lag Plot of Residuals and Predicted Residuals -- I.I.D. Model.......136

5.5 Residual Plot -- Unknown Knots with AR(1) Model and I.I.D. Model................................................................................147

5.6 Lag Plot of Residuals and Predicted Values -- AR(1) Model........148

5.7 Plot of Predicted Mean Functions................................................157

5.8 Plot of Standard Deviations of Y hat for Different Models.........158
Chapter I

Introduction

In this dissertation, we extend the polynomial spline regression model with free knots and independent identically distributed (i.i.d.) errors to the polynomial spline regression model with free knots and first-order autoregressive (AR(1)) errors. The asymptotic distributions of the estimated regression coefficients, free knots, autoregressive parameter and variance are shown, after suitable standardization, to be normal under some mild conditions. We also use our model to fit motion data collected at Children’s Hospital in San Diego. The fit is found to be satisfactory. However, analyses on the estimated regression coefficients and residual plot suggest that model selection techniques and more complicated autocorrelated error models may need to be considered in order to obtain a better model for the motion data.

We start in this introductory chapter by explaining the meaning of a spline and by describing two different kinds of splines in §1.1. The application of spline functions is discussed in §1.2. We give two graphs of the original data and the estimated cubic spline and the residual plot, coming from a motion data set fitted with cubic spline with continuous second derivatives and known knot points. The topic of the dissertation is motivated by the observation that the residual plot for this model has a cyclical pattern. Section 1.3 is our literature review. Work on linear models and smooth nonlinear models with i.i.d. or correlated errors are presented in this section. Feder’s results for polynomial spline regression with unknown knots and i.i.d. errors, and the Bayesian approaches to the parametric spline model are also discussed. From this section, we know that the estimation theory for polynomial spline regression with unknown knots and correlated errors is still open. The contents of the dissertation are stated in §1.4.
§1.1. Parametric spline regression and smoothing spline

A spline, in its original meaning, is a tool made of wood and used in construction. A draftman uses a spline to draw a smooth curve passing through some specific points. In some sense, the curve drawn is optimal under the constraint that the curve must pass through these specific points. In mathematical terminology, a spline is the solution to a constrained optimization problem. Because under some smoothness constraints, the optimal spline functions always appear to be piecewise polynomials, some authors define spline functions as "piecewise polynomials with degree m join in the knots obeying continuity condition for the function itself and its first m-1 derivatives" (Wold (1974)). Smith (1979) also had a similar definition, except that she represented splines as truncated polynomials.

From above, we can see that there can be two different definitions of a spline. One class of splines is the optimal solution to some constrained problem; the other definition can be seen in Wold(1974). We call the first type of splines smoothing splines, and the second type of splines parametric or regression splines. We are primarily interested in regression splines here.

Suppose we have data \((y_1,t_1),(y_2,t_2),\ldots,(y_n,t_n)\) coming from the following model

\[ y = f(t) + e. \tag{1.1} \]

Here \(f(t)\) is the unknown true mean function, \(e_i\) is an unknown error term with \(E(e_i) = 0\), \(\text{Var}(e_i) = \sigma^2\), \(\text{Cov}(e_i, e_j) = 0\) for \(i \neq j\), and \(t_i \in [a, b]\), \(a, b \in \mathbb{R}\). We know from the Stone-Weierstrass theorem that we can approximate an arbitrary continuous function \(f(t)\)
uniformly on a compact set \([a, b]\) by a polynomial \(p(t)\) with degree \(m\). Denoting the remainder term by \(r(t)\), we can write (1.1) as

\[ y = p(t) + r(t) + e. \]  

(1.2)

If we think that \(r(t)\) is small, we can discard it or assume that \(r(t) + e = e\) has similar behavior to \(e\) and use the standard polynomial regression \(p(t)\) to fit our data. However, sometimes we hope our model can adapt to local "irregular" behavior, so instead of eliminating \(r(t)\), for suitable choice of knots \(a < \tau_1 < \tau_2 < \ldots < \tau_{r-1} < b\), and constants \(\beta_1, \beta_2, \ldots, \beta_{r-1}\), which depend on \(\tau_i\)'s, but not \(t\), we can approximate \(r(t)\) by

\[ \sum_{j=1}^{r-1} \beta_j (t - \tau_j)_+^m, \]

where \(x_+ = x\) or 0 depending on whether \(x > 0\) or not. Then we obtain a polynomial regression spline model with knots \(\tau_1, \tau_2, \ldots, \tau_{r-1}\), which is a special case of parametric spline model.

The above polynomial regression spline model has the following form

\[ y = \sum_{i=0}^{m} \alpha_i t^i + \sum_{j=1}^{r-1} \beta_j (t - \tau_j)_+^m + e, \]

(1.3)

and allows continuous derivatives up to order \(m-1\). The model has the advantage that it is easy to formulate statistical hypothesis tests about the regression coefficients. (See Smith (1979)). However, if the design matrix of the model is poorly set, for example, the columns of the design matrix are nearly linear dependent, estimation of the parameters may result in numerical and statistical inaccuracy. In order to avoid the possibility of singularity in computation, piecewise polynomials were suggested. de Boor (1978, page 101-103) showed that the deterministic functional form in (1.3) can be represented as piecewise polynomials:
\[ f(t) = \sum_{k=0}^{m} \theta_{jk} t^k, \quad t \in [\tau_j, \tau_{j+1}], \quad j=1,2,\ldots,r-1, \quad (1.4) \]

where \( \tau_0 = a, \tau_r = b \) and \( f(t) \) satisfies the linear continuity constraints

\[ f^{(k)}(\tau_j^-) = f^{(k)}(\tau_j^+); \quad k=0,1,2,\ldots,m-1, \quad j=1,2,\ldots,r-1. \]

See §2.3 for the definitions of \( f^{(k)}(\tau_j^+) \) and \( f^{(k)}(\tau_j^-) \). Although the model (1.4) has advantages when compared to (1.3), it adds another estimation problem. In model (1.3), we have only \( m+r \) parameters to be estimated; but in (1.4), we have \( r(m+1) \) parameters subjected to continuity constraints.

One method of avoiding both of the problems above is to use B-splines. As an example, for cubic splines we can define

\[ B_j(t) = \sum_{i=j-2}^{j+2} \left\{ \frac{(t-\tau_i)^3}{\prod_{k=j-2}^{j+2} (\tau_i - \tau_k)} \right\}, \]

where

\[ \tau_i = \begin{cases} \tau_1 - (1-i)[\tau_1 - \min(t_j)] & \text{if } i \leq 0 \\ \tau_{r-1} + (i-r+1)[\max(t_j) - \tau_{r-1}] & \text{if } i \geq r. \end{cases} \]

Then \( \{B_{-1}, B_0, \ldots, B_{r+1}\} \) is a basis for the space of functions \( f(t) \) in (1.4), and there exist constants \( \omega_{-1}, \omega_0, \ldots, \omega_{r+1} \) such that
\[ f(t) = \sum_{j=-1}^{r+1} \omega_j B_j(t) \quad t \in [a, b]. \quad (1.5) \]

Note that due to the definitions of \( B_j(t) \) and of \( \tau_i \), and to (1.5), \( B_j(t) \) has the property that \( B_j(t) = 0 \) for \( t > \tau_{j+2} \) or \( t < \tau_{j-2} \). The number of unknown parameters in (1.5) is \( r+3 \), which is the same as the number of unknown parameters in (1.3). The use of the B-spline representation provides a well-conditioned design matrix and overcomes the possibility of singularity on the inverse of the covariance matrix (Wold (1974)).

Sometimes the models discussed above are too smooth. In this case, we can fit the data with a less smooth function

\[ y = \sum_{i=0}^{m} \alpha_i t^i + \sum_{j=1}^{r-1} \sum_{k=\nu_r}^{m} \beta_{kj} (t - \tau_j)^k, \quad \nu_r \in \{0,1,2,...,m\}. \quad (1.6) \]

Note that this model is a generalization of model (1.3). With \( \nu_r = m \), it reduces to (1.3). Its representation as piecewise polynomials and B-splines is similar to the case we discussed for model (1.3) (Eubank(1984)).

To complete the discussion of the spline function, we digress here from the polynomial splines to the smoothing splines. Suppose we have model (1.1). A smoothing spline is a function \( f(t) \) that minimizes the following criterion

\[ Q(f) = \sum_{i=1}^{n} [y_i - f(t_i)]^2 + \sigma^2 \int_{a}^{b} [D^m f(x)]^2 dx \]
subject to \( f \in W_m, \sigma^2 > 0. \) \hfill (1.7)

Here \( W_m \) is the set of functions on \([a, b]\) such that \( D^j f, j \leq m-1, \) is absolutely continuous and \( D^m f \) is in \( L_2, \) where \( L_2 \) is the set of measurable square-integrable functions on \([a, b]\). The solution to the problem is a polynomial spline with degree \( 2m-1 \) and knots at each data point. The objective function contains two parts: the least squares term and the smoothness term. The least squares term measures the fidelity of the model \( f(t) \) to the data and the smoothing term measures the smoothness of the function. If the smoothing parameter \( \sigma^2 \) goes to zero, we will have a function passing through all the data points so that the objective function can be minimized. In this case, we overfit the data and obtain an interpolating spline. If \( \sigma^2 \) is large, then the smoothing term overwhelms least squares term and offsets the information the data carries. Therefore, choosing a suitable smoothing parameter \( \sigma^2 \) is very important in fitting smoothing splines. Wahba and Wold (1975a,1975b) discussed the method of cross-validation for choosing \( \sigma^2 \) in their 1975 papers.

We now review the differences between parametric splines and smoothing splines. The first is the setting. For parametric splines, we do not have an objective function. The number of unknown parameters is finite. And we already have the specific functional form of the model in mind, which is chosen to be a piecewise polynomial between each two consecutive knots. The question left for parametric splines is how well the model will fit. As to smoothing splines, they are the optimal solution to some constrained problem. The objective function of the problem contains two terms: a least-squares term and smoothness term. The second difference between the two kinds of splines is the positions of knots. For parametric splines, the positions of the knots need not be data points, even though some authors (Wold (1974), Stone (1985)) recommended that knot points be placed at data points or close to some specific data points. As to smoothing splines, each data point is a knot. So, each point has equal "weight" and the
knot positions of smoothing splines have already been set by the
data, not like the knots of parametric splines which have to be
determined by estimation or by judgement. Besides, we have the
freedom to choose the degrees of each piecewise polynomial to be
odd or even and to choose smoothness constraints in parametric
spline regression, but for smoothing splines the degree of
polynomials is always \(2m-1\) and the \(f\) is always in \(W_m\), for some
positive integer \(m\).

§1.2. Applications of spline function

Spline functions have been widely applied to several fields of
statistics. Regression, as a tool of data analysis, is one of the most
popular ones. Depending on the availability and the quality of data,
Wold (1974) suggested several rules in choosing between
parametric spline regression and smoothing splines. Splines,
especially cubic splines, have been further applied to non-
parametric regression (Silverman, 1985), multiple regression (Stone
(1985) and Stone and Koo (1985)), and projection pursuit
regression (Friedman and Stuetzle (1981)).

For parametric spline regression, Gallant (1977) used the
quadratic-quadratic polynomial model to fit the data on specific
retention volume of methylene chloride in polyethylene
terephthalate. Wold (1974) used cubic splines to fit transformed
pharmacokinetics data on the concentration of a drug in blood and
found that the results are more reliable than those obtained from
fitting a classical compartment model. Also, Gallant and Fuller
(1973) used a two-segment quadratic-linear polynomial to model
preschool boys' weight/height ratios against age, but with an
estimated knot point.

Smoothing spline functions have been successfully applied to
biomechanical data also. To name but a few, McLaughlin et al.
(1977) used cubic splines to fit the acceleration of lower leg in
running and used cubic splines in fitting first and second
derivatives of forearm motion in elbow flexion. Wood and Jennings
(1979) used a cubic spline and a quintic spline to fit vertical jump data. Zernicke et al. (1976) used a cubic spline in modelling force platform data (vertical reaction forces). These authors, after comparing results from fitting splines with other methods; like orthogonal polynomial, finite difference methods etc.; concluded that fitting by spline functions is a promising method in modelling biomechanical data.

In 1987, the author used Smith's (1979) parametric cubic spline regression model with continuous second derivatives and with known knot points to fit nine angular motion (hip flexion- extension) data sets from Helen Hayes Hospital, Children's Hospital in San Diego and Newington Children's Hospital. Two pictures illustrate two of the fitted results. Figure 1.1 shows the estimated cubic spline and the original data. Figure 1.2 is the residual plot. From Figure 1.1, we can see that the cubic spline model fits the data quite well (with $R^2 > 0.9$). However, if we examine Figure 1.2, we can see that the residual plot has a cyclical pattern across time, indicating that there are some characteristics in the human motion data that have not been incorporated in the parametric cubic spline regression model.

The investigation of the residual plots of the hip flexion-extension data motivated the addition of correlated errors to the parametric spline regression model. We conjecture that adding AR(1) errors to the parametric spline regression model will make the residuals more like i.i.d. errors. This leads us to study the asymptotic behaviors of the estimators of the regression coefficients and knot points for the parametric spline regression model with AR(1) errors. For the case that knot points are known, we just have another version of the general linear model. We will explore the asymptotic distributions for the polynomial spline regression model with unknown knots and AR(1) errors.
Figure 1.1: Plot of Original Data and Predicted Values -- Cubic Splines with Known Knots
1 = Original Data. Solid Curve = Predicted Values
$\tau_1 = 0.17308$. $\tau_2 = 0.53846$. $\tau_3 = 0.69231$
Left Hip Flexion-Extension

Figure 1.2: Residual Plot -- Cubic Splines with Known Knots
§1.3. Literature review

Many authors have investigated the asymptotic properties of the general linear model with uncorrelated or correlated errors. In 1963, Eicker examined the asymptotic properties for linear regression with independent errors. Hildreth (1969) derived the asymptotic distributions for linear model with AR(1) errors. Pierce (1971) generalized Hildreth's model to ARMA(p,q) errors, but used a conditional least-squares estimators approach. Pierce found that the conditional least-squares estimator, for which the difference is asymptotically negligible as compared to the usual least-squares estimator, is also asymptotically normal. Recognizing the difficulty in finding the inverse of the covariance matrix of disturbances as we try to compute the estimates, Harvey and Phillips (1979) suggested Kalman filtering to compute generalized least squares estimates. Sarkar (1985) discussed inference for linear model with error terms further decomposed into two AR(1) errors.

Much work has also been done in the field of nonlinear regression with correlated or uncorrelated errors. However, all the work assumed that the nonlinear regression function is smooth in the unknown parameters. Hartley and Booker (1965) developed an interative method to compute the nonlinear least squares estimator (l.s.e.) and showed that the resulting estimator is consistent. Jennrich (1969) derived the asymptotic distribution of the least squares estimator for nonlinear model with i.i.d. errors. For nonlinear model with correlated errors, Gallant and Goebe! (1976) examined AR(1) errors. They first estimated the covariance matrix and then used the estimated covariance matrix to obtain nonlinear least squares estimators. They showed that the estimator has an asymptotic normal distribution. Greenhouse et al. (1987) allowed ARMA(p,q) errors and gave asymptotic properties of the estimators. A unified theoretical discussion on the asymptotic properties of nonlinear regression with correlated errors can be found in the book of Gallant and White (1988).
If a parametric spline regression model has known knots and independent errors, the asymptotic properties of the regression coefficients are similar to those of the usual general linear model, since we have a linear form in each region. For correlated errors, the results on the general linear model with correlated errors can be applied directly. However, if the positions of the knot points are unknown, the model is no longer linear in all the parameters and the deviation of the asymptotic properties of the estimators becomes difficult.

Several authors have worked on inference in the case of the parametric spline regression with unknown knots and independent errors. Quandt (1958) and Robison (1964) discussed the estimation of the shift point. Hudson (1966) discussed estimation of three types of join points. The join point classification depends on the positions of the estimated join point and the value of the derivative at the estimated join. Motivated by Hudson's work, Gallant and Fuller (1973) investigated Hudson's type three model and showed that under suitable conditions the estimator obtained by using a modified Gauss-Newton method is consistent. Hinkley (1969,1971) derived the asymptotic distribution for the estimator of an unknown knot with the condition that the mean function is continuous at the shift point. In a more general setting, Feder (1975a) showed that the asymptotic distributions of the estimators of regression coefficients and knot points are normal.

A Bayesian approach was also adopted by some authors to make inference about the regression coefficients and the change points of the parametric spline regression with independent or correlated errors. For independent errors, Halpern (1973) considered the case in which the positions of possible knots are known, but the number of knots is unknown. Smith and Cook (1980) investigated straight lines with one knot point. Norton (1982) performed a Monte Carlo study for joint locations in segmented linear models. As to correlated errors, Salazar (1982), Salazar et. al. (1981) and Ohtani (1982) considered AR(1) errors and used a discrete uniform prior for the shift point. Ilmakunnas
and Tsuruni (1984) considered different AR(1) errors for different regions.

Although much work has been done on linear and nonlinear models with independent or correlated errors, the asymptotic distributions of parametric spline regression with unknown knots and correlated errors is still unknown. As discussed in Feder (1975a), the presence of unknown knot positions dramatically increases the difficulty and complexity as we try to explore the asymptotic properties of the spline regression model. The model is not linear in all the parameters and the likelihood (or least-squares) function will have differentiability problems if we want to use classical methods to find the asymptotic distributions of the estimators of the regression coefficients and knot points. The addition of correlated errors terms to Feder's model will make the analysis more difficult. We will see this in later chapters.

§1.4. Contents of the dissertation

The dissertation consists of six chapters. Chapter 2 describes the model we are going to investigate and some preliminary results about consistency. In that chapter, we also define some notation that will be used throughout the dissertation.

In Chapter 3, we derive the asymptotic distributions of the estimators of a polynomial spline regression model with unknown knots and AR(1) errors, assuming that the correlation coefficient between two consecutive observations is known. The asymptotic distribution of $\hat{\theta}^2$ is derived also. We find that the asymptotic distributions of the estimators, including regression coefficients, knot points and $\sigma^2$, with additional assumptions than in the general linear model, asymptotically follow normal distributions. In Chapter 4, we extend to the case where the correlation coefficient is unknown. We find that, under the same assumptions as in the case
that $\rho$ is known, the asymptotic distributions of the estimators also have the same asymptotic distributions as their counterparts in Chapter 3. We also find that the asymptotic distribution of $\sqrt{n} (\hat{\rho} - \rho^{(o)})$ converges to a normal distribution as $n$ diverges to infinity.

We give the data analysis in Chapter 5. In this chapter, we use generalized cubic spline regression model with unknown knots and AR(1) errors to fit a motion data set collected at Children's Hospital in San Diego. An approach different from the one discussed in Chapter 2 is used to find the interesting estimates. We prove that the two methods are asymptotically equivalent. The fitness of the model used is discussed in this chapter also.

Chapter 6 presents our conclusions and suggestions for future work. We summarize the results derived in Chapters 3 and 4, and findings in Chapter 5 in §6.1. As we will see from the derivation of the asymptotic distributions and from the data analysis chapter, there are several practical problems and theoretical problems still to be investigated. These problems, discussed in §6.2, include the determination of sample size and model selection. Besides, theoretical problems such as different $\rho$ values in different regions and more complicated time series models like ARMA($p,q$) errors also need to be explored.
Chapter II

The Model and Preliminary Results

This chapter contains three sections. In §2.1, we describe the model to be investigated and list the assumptions used to derive the asymptotic distributions of estimators we are interested in. Some preliminary results, basically due to Feder (1975), and the approach used to find the maximum likelihood estimators are stated in §2.2. The concepts of 'bounded in probability' and 'convergence in probability' are defined in §2.3. We use the traditional notation $O_p(*)$ and $o_p(*)$ to denote these two probability concepts. Some notation used throughout this dissertation is also defined in §2.3.

§2.1. Model and Assumptions

We are interested in the polynomial regression spline model with unknown knots and AR(1) errors. The polynomial regression spline model has the following form:
\[ y_t = \sum_{j=0}^{m} \alpha_j t^j + \sum_{k=1}^{r-1} \sum_{j=1}^{m} \beta_{jk} (t - \tau_k)^j_+ + u_t, \quad (2.1) \]

where \( \tau_1 < \tau_2 < \ldots < \tau_{r-1} \), \( m \) is the maximum degree of the polynomials, and \( r \) is the number of polynomial pieces, \( r \geq 2 \), and \( (t - \tau_j)^j_+ = (t - \tau_j)^j \) if \( t \geq \tau_j \); and = 0, otherwise. Also \( u_t \) follows a first-order autoregressive process. As we will see, this model is a special case of (1.6) with \( v_r = 1 \). Without loss of generality, we will take \( 0 \leq t \leq 1 \) and \( \tau_0 = 0, \tau_r = 1 \). As stated in Chapter 1, model (2.1) has the property that its mean function is continuous and the deterministic part of (2.1) can be represented as a piecewise polynomial:

\[ \sum_{j=0}^{m} \alpha_j t^j + \sum_{k=1}^{r-1} \sum_{j=1}^{m} \beta_{jk} (t - \tau_k)^j_+ \]

\[ = \sum_{j=1}^{r} \sum_{k=0}^{m} \theta_{jk} t^k I_j(t) \]

\[ = \sum_{j=1}^{r} f_j(\theta_j, t) I_j(t) \]

\[ = f(\theta, \tau, t), \quad (2.2) \]
Where \( f_j(\theta_j, t) \) is the polynomial on the \( j \)th interval with the form
\[
\sum_{k=0}^{m} \theta_{jk} t^k,
\]
and \( I_j(t) \) is defined as
\[
I_j(t) = \begin{cases} 
1 & \text{if } \tau_{j-1} \leq t < \tau_j \\
0 & \text{otherwise} 
\end{cases}
\] (2.3)

Also, the mean function satisfies the continuity constraints
\[
f_j(\theta_j; \tau_j) = f_{j+1}(\theta_{j+1}; \tau_j), \quad j = 1, 2, \ldots, r - 1. \quad (2.4)
\]

Therefore under conditions (2.4), we can write model (2.1) as
\[
y_t = \sum_{k=0}^{m} \theta_{jk} t^k + u_t \quad \text{if } \tau_{j-1} \leq t < \tau_j, \quad j = 1, 2, \ldots, r. \quad (2.5)
\]

If the \( u_t \)'s are uncorrelated, then model (2.5) is a special case of Feder's model (1975a). In his paper, Feder showed that if we have i.i.d. errors, then under suitable conditions, the least squares estimator \( (\hat{\theta}, \hat{\tau}) \) converges in probability to the true parameter \( (\theta(0), \tau(0)) \). Furthermore, Feder showed that the asymptotic distributions of \( \hat{\theta} \) and \( \hat{\tau} \), after suitable standardization, are normal. However, in many cases, the error terms are not i. i. d., but correlated, especially if we have time-related data. Also, from the discussion in Chapter 1, we know that a polynomial regression spline can fit data well. So, if the residuals obtained from fitting polynomial regression spline model appear to have cyclical pattern, it will be more meaningful and more suitable if we refit it by a
polynomial regression spline model with AR(1) errors rather than i.i.d. errors.

Suppose that the data are observed at discrete time points. For a given sample size \( n \), let us first write the mean function of our model in matrix form. From (2.1) and (2.2), we have

\[
y_{ni} = f(\theta, \tau; t_{ni}) + u_{ni}, \quad i = 1, 2, \ldots, n. \tag{2.6}
\]

where

\[
f(\theta, \tau; t_{ni}) = \sum_{j=1}^{r} f_j(\theta, t_{ni}) I_j(t_{ni})
\]

\[
= \sum_{j=1}^{r} \sum_{k=0}^{m} \theta_{jk} t_{ni}^k I_j(t_{ni}).
\]

and \( I_j(t_{ni}) \) is defined as above. Writing \( f_j(\theta, t) \) in matrix form,

\[
f_j(\theta, t) = \sum_{k=0}^{m} \theta_{jk} t_{ni}^k = (\theta_{j0}, \theta_{j1}, \ldots, \theta_{jm})^T
\]

Letting \( \theta_j^* = [\theta_j 0, \theta_j 1, \ldots, \theta_j m] \) and \( t = [1, t, t^2, \ldots, t^m] \), then
\[ f_j(\theta, t) = \theta_j t. \]

Therefore,

\[
f(\theta, \tau, t_{ni}) = \sum_{j=1}^{r} (\theta_{j0}, \theta_{j1}, \ldots, \theta_{jm}) \cdot I_j(t_{ni})
\]

\[
= \sum_{j=1}^{r} \theta_j t_{ni} I_j(t_{ni})
\]

\[
= (\theta_1', \theta_2', \ldots, \theta_r') \begin{pmatrix}
  t_{ni}I_1(t_{ni}) \\
  t_{ni}I_2(t_{ni}) \\
  \vdots \\
  t_{ni}I_r(t_{ni})
\end{pmatrix}
\]

\[ = \theta' [t_{ni} \otimes I(t_{ni})] \]

Here \( I(t_{ni}) = [I_1(t_{ni}), I_2(t_{ni}), \ldots, I_r(t_{ni})]' \); and \( t_{ni} \otimes I(t_{ni}) \), unlike the usual definition of Kronecker product, denotes the left direct product of the vectors \( t_{ni} \) and \( I(t_{ni}) \) (Graybill (1983), p.216), which is defined as
\[
\mathbf{t}_{ni} \otimes \mathbf{I}(t_{ni}) = \begin{pmatrix}
t_{ni}I_1(t_{ni}) \\
t_{ni}I_2(t_{ni}) \\
\vdots \\
t_{ni}I_r(t_{ni})
\end{pmatrix}.
\]

So, in matrix form, we have

\[
\begin{pmatrix}
y_{n1} \\
y_{n2} \\
\vdots \\
y_{nn}
\end{pmatrix} = \begin{pmatrix}
\theta'[t_{n1} \otimes \mathbf{I}(t_{n1})] \\
\theta'[t_{n2} \otimes \mathbf{I}(t_{n2})] \\
\vdots \\
\theta'[t_{nn} \otimes \mathbf{I}(t_{nn})]
\end{pmatrix} + \begin{pmatrix}
u_{n1} \\
u_{n2} \\
\vdots \\
u_{nn}
\end{pmatrix}.
\]

Letting \( \mathbf{\tilde{t}}_{ni} = \mathbf{t}_{ni} \otimes \mathbf{I}(t_{ni}), \) \( \mathbf{T}_n = (\mathbf{\tilde{t}}_{n1}, \mathbf{\tilde{t}}_{n2}, \ldots, \mathbf{\tilde{t}}_{nn}), \) \( \mathbf{y} = (y_{n1}, y_{n2}, \ldots, y_{nn})' \)
and \( \mathbf{u} = (u_{n1}, u_{n2}, \ldots, u_{nn})' \), we have

\[
\begin{pmatrix}
\mathbf{\tilde{t}}'_{n1} \\
\mathbf{\tilde{t}}'_{n2} \\
\vdots \\
\mathbf{\tilde{t}}'_{nn}
\end{pmatrix} + \mathbf{u} = \begin{pmatrix}
\mathbf{\tilde{t}}'_{n1} \\
\mathbf{\tilde{t}}'_{n2} \\
\vdots \\
\mathbf{\tilde{t}}'_{nn}
\end{pmatrix} + \mathbf{u}
\]

\[
= \mathbf{T}_n \mathbf{\theta} + \mathbf{u} = \mu(\mathbf{\theta}, \tau, \mathbf{t}_n) + \mathbf{u}. \quad (2.7)
\]
Notice that $\theta$ is the vector of regression coefficients, $\tau$ is the vector of unknown knot points, $t_n = [t_{n1}, t_{n2}, \ldots, t_{nn}]'$ is the vector of observed times, $u$ is the vector of AR(1) error terms and $\mu(\theta, \tau, t_n)$ is the vector of mean function observed at times $t_n$. Also, the matrix $\tilde{T}_n$ depends on the unknown parameter $\tau$ because $I(t_{ni})$ involves $\tau$. Since the form in (2.7) is more compact than that in (2.2), we will use it for later analysis.

For the above model, we make the following seven assumptions:

1. The stochastic part in (2.4) follows an AR(1) model with $u_{ni} = \rho u_{ni-1} + e_{ni}, i = 2, 3, \ldots, n$. Here the $e_{ni}$'s are i. i. d. normal with $E(e_{ni}) = 0$, $\text{var}(e_{ni}) = \sigma^2$ and $u_{ni}$ has a stationary distribution.

2. In each interval $[\tau_{j-1}, \tau_j)$, there exists at least $m+1$ different design points, $j=1, 2, \ldots, r$.

3. The functions $f_i(\theta_i; t)$ and $f_{i+1}(\theta_{i+1}; t)$ are not equal, and if $f_i(\theta_i; \tau) = f_{i+1}(\theta_{i+1}; \tau)$, then only one value of $\tau$ is a solution to the equation.

4. For given $n$, let $H_n$ be the empirical distribution function of $t_{ni}$. We assume that the sequence $\{t_{ni}\}$ is selected such that $H_n(\cdot) \rightarrow H(\cdot)$ in distribution as $n \rightarrow \infty$, where $H(\cdot)$ is a continuous distribution with $H(0) = 0$ and $H(1) = 1$.

5. For each design point, if $N$ is a neighborhood of the design point, then $H(N) > 0$. 
6. Feder's technical assumption is satisfied. That is, letting

(a). \( e_{n1}, e_{n2}, \ldots, e_{nn} \) be i. i. d. random variables,
(b). \( \{ N_n \} \) be a sequence of random variables such that \( 1 \leq N_n \leq n \) and \( N_n = O_p(g(n)) \), where \( 0 < g(n) \leq n, g(n) \to \infty \) as \( n \to \infty \),
(c). \( \{ \Delta_i \}_{i=1}^\infty \) be a fixed set of constants and \( a \geq 0 \), and \( \Delta_{n1}, \Delta_{n2}, \ldots, \Delta_{nn} \) be constants such that

\[
\sup_n \max_{1 \leq i \leq n} |\Delta_{ni}| < \infty, \quad n^a \Delta_{ni} = \Delta_i (1 + b_{ni}),
\]

where

\[
\max_{1 \leq i \leq n} |b_{ni}| = o(g(n)^{-\frac{1}{2}}) \quad \text{as} \quad n \to \infty,
\]

(d). \( h(i) = o(1) \) as \( i \to \infty \),

then for every \( K > 0 \) such that \( Kg(n) < n \) for \( n \) sufficiently large, if for \( \delta > 0 \),

\[
\lim_{n \to \infty} \frac{Kg(n)}{\sum_{i=1}^{\infty} \sum_{j=1}^{i} \Delta_i \log \log \sum_{j=1}^{i} \Delta_j^2} < \infty,
\]

we have as \( n \to \infty \),
\[ T_{N_n} = \langle \Delta_{N_n}, e_{N_n} \rangle / \| \Delta_{N_n} \| = O_p((\log \log n)^{\frac{1}{2}}). \]

Here \( \Delta_{N_n} = [\Delta_{n1}, \Delta_{n2}, \ldots, \Delta_{nn}]', e_{N_n} = [e_{n1}, e_{n2}, \ldots, e_{nn}]' \), \( \langle *, * \rangle \) denotes the inner product and \( \| * \| \) is the Euclidean norm.

7. \( \lim_{n \to \infty} \frac{1}{n} \left( \frac{\partial \mu(\theta, \tau; t_n)}{\partial \theta} \right) \Sigma_n^{-1} \left( \frac{\partial \mu(\theta, \tau; t_n)}{\partial \theta} \right)' = G(\rho) \), where \( \Sigma_n \) is the covariance matrix of \( u \) and \( G(\rho) \) is a positive definite matrix.

Assumption 1 says that the \( u_{ni} \)'s follow an AR(1) model, which is a simple and popular model for time-related data. Assumption 2 guarantees that \( \theta \) is well identified. If we have at least \( (m+1) \) different values of the \( t_{ni} \)'s in each region, then the corresponding block in the design matrix is an extended Vandermonde matrix. This means that the design matrix is of full rank, \( r(m+1) \), and hence we have a nonsingular covariance matrix. Assumption 3 says that no two consecutive polynomials are identical and \( \tau \) is uniquely identified if \( \theta \) is known. It is our intention that the knot points really indicate structural change. If assumption 3 is violated, that is, \( f_i = f_{i+1} \), then we can merge the intervals \([ \tau_i, \tau_{i+1} \]) and \([ \tau_{i+1}, \tau_{i+2} \]) into \([ \tau_i, \tau_{i+2} \]) and still have the same functional form \( f_i = f_{i+1} = f_i^* \) for \([ \tau_i, \tau_{i+2} \]). That we require assumption 3 to be satisfied in our model should be reasonable, since sudden structural change in the join of the mean function can occur in many cases.

Assumptions 4 and 5 allow us to delete \( g(n) = o(n/\log \log n) \) of the observations around the knot points without affecting the asymptotic behavior of the maximum likelihood estimators.
Assumption 6 is a technical assumption and is rather opaque. Feder (1975) used this assumption in the proof of a theorem that accounts for the influence on the observations near the knot points on the estimators. By this assumption, we have that \( \hat{\theta} \) converges to \( \theta^{(0)} \), the true value of \( \theta \), at the rate of \( \sqrt{(\log \log n)/n} \). Assumption 7 guarantees that the asymptotic distributions of \( \hat{\theta} \) and \( \hat{\sigma}^2 \) are not degenerate.

Our main interest is to obtain the asymptotic properties of model (2.6) under assumptions 1 through 7. In other words, we are aiming to prove the following Proposition:

**Proposition 1**: Suppose \( y_{ni} \) s are generated from model (2.6) and suppose that assumptions 1 through 7 are satisfied. Then the maximum likelihood estimators \( \hat{\theta}, \hat{\sigma}^2 \) and \( \hat{\sigma}^2 \), after suitable normalization, asymptotically follow normal distributions.

§2.2 Preliminary Results

To be able to compute the maximum likelihood estimates and to derive their asymptotic distributions, we need to compute the form of the covariance matrix of the \( u_{ni} \) first. Observe that \( u_{ni} = \rho u_{n(i-1)} + e_{ni} \). \( u_{ni} \) is a stationary AR(1) process, so \( u \) can be rewritten in terms of i.i.d. random variables. Now \( e_{ni} \)'s are i.i.d. normal with mean \( 0 \), we have \( E(u) = 0 \). Also,

\[
E(u_{ni}^2) = \text{Var}(u_{ni}) = \frac{\sigma^2}{1-\rho^2},
\]

\[
E(u_{ni}u_{n(i-1)}) = \frac{\rho}{1-\rho^2} \sigma^2,
\]
and

\[ E(u_n u_{n(i-2)}) = \rho E(u_{n(i-1)} u_{n(i-2)}) = \rho^2 E(u_{n(i-2)}) = \frac{\rho^2}{1 - \rho^2} \sigma^2. \]

In general,

\[ E(u_{n(i-1)} u_{n(i-k)}) = \rho^{\lvert i-k \rvert} \frac{\sigma^2}{1 - \rho^2}. \]

Therefore the covariance matrix of \( u \) is

\[
\text{Cov} (u) = \sigma^2 \Sigma_n = \frac{\sigma^2}{1 - \rho^2} \begin{pmatrix}
1 & \rho & \rho^2 & \ldots & \rho^{n-1} \\
\rho & 1 & \rho & \ldots & \rho^{n-2} \\
\rho^2 & \rho & 1 & \ldots & \rho^{n-3} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \ldots & 1
\end{pmatrix}.
\]

Since \( \Sigma_n \) is symmetric and nonsingular for \( \lvert \rho \rvert < 1 \), by matrix theory, there exists a nonsingular matrix \( \Psi_n \) such that \( \Psi_n \Psi_n' = \Sigma_n \) (Graybill (1976), Theorem 1.4.2). To claim that \( \Psi_n^{-1} y \) is a normal vector with mean \( \theta \) and covariance matrix \( I_n \sigma^2 \), we let \( \tilde{y} = \Psi_n^{-1} y \). Then

\[
\tilde{y} = \Psi_n^{-1} T_n \theta + \Psi_n^{-1} u
\]

\[
= \tilde{T}_n \theta + \tilde{u}
\]

\[
= \hat{\mu}(\theta, \tau, t_n) + \hat{u}
\]
\[ \hat{u} = \hat{\mu}(\theta, \tau, t_n) + \hat{u} \]

where \( T_n' = \Psi_n^{-1} T_n \), \( \hat{u} = \Psi_n^{-1} u \), and \( \hat{\mu}(\theta, \tau; t_n) = \hat{T}_n \theta \). Now

\[ E(\hat{u}) = \Psi_n^{-1} E(u) = 0 \]

\[ \text{Cov}(\hat{u}) = \text{Cov}(\Psi_n^{-1} u) = \Psi_n^{-1} \text{Cov}(u)(\Psi_n^{-1})' \]

\[ = \Psi_n^{-1} \Psi_n \Psi_n'(\Psi_n^{-1})' = I_n \sigma^2 \]

Since \( u \) is a normal vector with mean 0 and covariance matrix \( \Sigma_n \), we have that \( \hat{u} \) follows a multivariate normal distribution with mean 0 and covariance matrix \( I_n \sigma^2 \), which implies that components of \( \hat{u} \) are i.i.d. normal variates with mean 0 and variance \( \sigma^2 \). Therefore, we have the following log likelihood function for the observed data:

\[ L(\theta, \tau; \rho, \sigma^2, y) = \log L(\theta, \tau; \rho, \sigma^2, y) \]

\[ = \frac{n}{2} \log(2\pi \sigma^2) - \frac{1}{2} \log|\Sigma_n| - \frac{1}{2\sigma^2} (y - \mu(\theta, \tau; t_n)\)' \Sigma_n^{-1} (y - \mu(\theta, \tau; t_n)) \]

\[ = \frac{n}{2} \log(2\pi \sigma^2) - \frac{1}{2} \log|\Sigma_n| - \frac{1}{2\sigma^2} (\hat{y} - \hat{\mu}(\theta, \tau; \rho, t_n)\)' \hat{\Sigma}_n^{-1} (\hat{y} - \hat{\mu}(\theta, \tau; \rho, t_n)) \]

(2.8)

When \( \rho \) is known, the maximum likelihood estimates of \( \theta, \tau \) and \( \sigma^2 \) are the values of \( \hat{\theta}, \hat{\tau} \) and \( \hat{\sigma}^2 \) that maximize (2.8). For \( \tau \), the vector of knot points, unknown, the mean function is not linear in \( (\theta, \tau) \). An iterative method must be used to find the m.le.'s. In most
take the advantage of this information while we compute the m.l.e.'s. The algorithm is as follows: for a fixed \( \tau \), our model, with unknown parameters \( \theta \) and \( \sigma^2 \) only, is a linear model. The m.l.e.'s of \( \theta \) and \( \sigma^2 \) then can be obtained by maximizing \( \log L(\theta, \tau, \sigma^2, y) \) by using the Lagrange method subject to continuity constraints. Then, for the given \( \hat{\theta} \) and \( \hat{\sigma}^2 \), consider \( \log L(\hat{\theta}, \tau; \hat{\sigma}^2, y) \) as a function of \( \tau \) and find the value of \( \hat{\tau} \) that maximizes \( \log L(\hat{\theta}, \tau; \hat{\sigma}^2, y) \). This process continues until we obtain satisfactory estimates. If in addition that \( \rho \) is also unknown, the computation is similar, but more involved. However, the above computational difficulty can be overcome by using conditional least squares method. In Chapter 5, we give an alternative method to find the estimates of \( \hat{\theta} \) and \( \hat{\tau} \). The estimates found by this method is proved, also given in Chapter 5, to be asymptotically equivalent to the estimates obtained by the above algorithm.

It is interesting to notice that the mean function is not disturbed by the use of autocorrelated error terms. Therefore, the continuity constraints at the knot points about the mean function are not destroyed and hence we can write \( \theta = \theta(\mu, t^*) \), where \( \theta(\cdot, \cdot) \) is a continuous function in its arguments. Since \( \theta \) is identified by the design points, and \( t^k \), \( k=0,1,2,\ldots,m \), are functionally independent, so following Feder's argument, the mean function satisfies the Lipschitz condition \( \| \theta_1 - \theta_2 \| \leq c \| \mu_1 - \mu_2 \|, \ c > 0 \), whenever \( t_n \) belongs to \( T \) and \( \mu_1 = \mu(\theta_1, \tau_1; t_n) \), \( \mu_2 = \mu(\theta_2, \tau_2; t_n) \) are both in \( M \), where \( T \) is a neighborhood of \( t_n \) and \( M \) is a neighborhood of \( \mu^{(0)} \). Furthermore, directly applying Feder's approach, with the vectors that span the column space of the design matrix slightly modified by multiplying by the matrix \( \Psi^{-1}_n \), we can obtain the following theorems.
**Theorem 2.1.** Suppose we have model (2.5) with assumptions 1 through 7 satisfied. Then

\[ \theta - \theta^{(o)} = o_p \left( \frac{-1}{n^2} + \frac{1}{(\log \log n)^2} \right) \]

**Theorem 2.2.** Under model (2.5) and with the assumptions in Theorem 2.1, if

\[ D_j^+ = D_j^- \text{, then} \]

\[ (\hat{\tau}_j - \tau_j^{(o)}) = o_p \left( \frac{-1}{n^2} + \frac{1}{(\log \log n)^2} \right), \quad j = 1, 2, ..., r-1. \]

where

\[ D_j^+ = f_{j+1}(\theta_j^{(o)}; \tau_j^{(o)+}) - f_j(\theta_j^{(o)}; \tau_j^{(o)+}) \]

\[ D_j^- = f_{j+1}(\theta_j^{(o)}; \tau_j^{(o)-}) - f_j(\theta_j^{(o)}; \tau_j^{(o)-}) \]

and \( f_{j+1}(\theta_j^{(o)}; \tau_j^{(o)+}) \) and \( f_{j+1}(\theta_j^{(o)}; \tau_j^{(o)-}) \) are as defined in Sec. 2.3.

We observe that the rate of convergence of \( \hat{\theta} \) and \( \hat{\tau} \) in the case with AR(1) errors is the same as in the case with i.i.d. errors. We should not be surprised by this fact. Because, as shown in (2.8), the mean function is not changed and the appearance of autocorrelated
error terms only affects the covariance structure. As a result, the rate of convergence is not disturbed but the variability of the estimator is affected. We will discuss this further in Section 3.3.

§2.3. Notation

This section is used to define two probability concepts to provide and some of the notation used throughout this dissertation. First, we define two probability concepts:

**Definition 2.1.** (Serfling (1980), page 8). A sequence of random variables \( \{ Y_n \} \), with distribution functions \( \{ F_n \} \), is said to be bounded in probability if for every \( \varepsilon > 0 \), there exists \( M_\varepsilon \) and \( N_\varepsilon \) such that if \( n > N_\varepsilon \), then

\[
P( |Y_n| < M_\varepsilon ) > 1 - \varepsilon.
\]

*If* \( Y_n \) *is bounded in probability, we write* \( Y_n = O_p(1) \).

From the above definition, it is easy to see that if \( Y_n \) converges weakly to \( Y \), then \( Y_n = O_p(1) \).

**Definition 2.2.** (Serfling (1980), page 6). A sequence of random variables \( \{ Y_n \} \) is said to converge in probability to 0 if for every \( \varepsilon > 0 \),

\[
\lim_{n \to \infty} P( |Y_n| < \varepsilon ) = 1.
\]

*If* \( Y_n \) *converges to 0 in probability, we write* \( Y_n = o_p(1) \).
To evaluate whether or not a function of two sequences of random variables \( \{ Y_n \} \) and \( \{ Z_n \} \) converges, we use the continuity of the function considered. For example, if \( \{ Y_n \} = o_p(1) \) and \( \{ Z_n \} = o_p(1) \), then \( Y_n Z_n = o_p(1) o_p(1) = o_p(1) \), and \( Z_n / Y_n = o_p(1)/o_p(1) = o_p(1) \). For more discussion, see Chernoff (1956).

By convention, we will use bold capital characters to denote matrices, and use bold lower-case characters to denote vectors. The following are the symbols used very often in the dissertation. We list them here for easy reference:

1. \( \theta, \tau, \rho, \sigma^2 \) : parameters that we are interested in. Here \( \theta \) is the vector of regression coefficients, \( \tau \) is the vector of knot points, \( \rho \) is the correlation of two consecutive errors and \( \sigma^2 \) is the variance of elements of the vector \( \Psi_n^{-1} u \). We will let \( \theta^{(o)}, \tau^{(o)}, \rho^{(o)} \) and \( \sigma_o^2 \) denote the true values of the parameters.

2. \( t, t_{ni}, \hat{t}_{ni} \) and \( t_n \): \( t \) is an \((m+1)\)-dimensional vector defined by \( t = [1, t, t^2, ..., t^m] \); \( t_{ni} \) is an \((m+1)\)-dimensional vector if \( t \) is observed at \( t_{ni} \), therefore,
\[
t_{ni} = [1, t_{ni}, t_{ni}^2, ..., t_{ni}^m].
\]
Also, \( t_n \) is the \( n \)-dimensional vector of the \( n \) design points, that is,
\[
t_n = [t_{n1}, t_{n2}, ..., t_{nn}].
\]
Finally, we let
\[
\hat{t}_{ni} = [0, 0, ..., 1, t_{ni}, ..., t_{ni}^m, 0, ..., 0] = [0, 0, ..., t_{ni}, 0, ..., 0] \]
be an \( r(m+1) \)-dimensional vector, where the subvector
\[
t_{ni} = [1, t_{ni}, ..., t_{ni}^m] \]
is in the \( k \)th block if \( t_{ni} \) is in the \( k \)th time interval.
3. \( \mu(\theta, \tau, t) \) and \( \mu(\theta, \tau, t_\mathbf{n}) \) : \( \mu(\theta, \tau, t) \) is the mean function of our model observed at time \( t \); \( \mu(\theta, \tau, t_\mathbf{n}) \) is the vector of the mean function \( \mu(\theta, \tau, t) \) observed at time \( t_\mathbf{n} \). We denote the true value of the mean function by \( \mu^{(o)} = \mu(\theta^{(o)}, \tau^{(o)}, t) \), or \( \mu^{(o)} = \mu(\theta^{(o)}, \tau^{(o)}, t_\mathbf{n}) \) if \( \mu(\theta^{(o)}, \tau^{(o)}, t) \) is observed at \( t_\mathbf{n} \). We will write \( v = \mu^{(o)} - \mu \) if observed at time \( t \), and \( v = \mu^{(o)} - \mu \) if observed at time vector \( t_\mathbf{n} \).

4. \( \Omega \): the collection of all vectors \( \theta \) such that the continuity constraints (2.4) are satisfied.

5. \( \frac{\partial^k}{\partial \theta^k} f(\theta) \): kth partial derivative of \( f \) with respect to \( \theta \) evaluated at \( \theta = \hat{\theta} \). Also, \( f^{(k)}(\theta^{(o)}, \tau^{(o)}) \) is the derivative of \( f \) with respect to \( t \) if evaluated at \( (\theta^{(o)}, \tau^{(o)}) \); \( f^{(k)}(\theta^{(o)}, \tau^{(o)}+) \) is the right kth derivative of \( f \) with respect to \( t \) if evaluated at \( (\theta^{(o)}, \tau^{(o)}) \). The left kth derivative of \( f \) respect to \( t \) evaluated at \( (\theta^{(o)}, \tau^{(o)}) \) is denoted as \( f^{(k)}(\theta^{(o)}, \tau^{(o)}-) \).

6. * and ** : The notation of * and ** is used in a pseudo-problem to be discussed in Chapter 3. For functions with *, we mean that only those design points present in the pseudo-problem are considered in this function; If a function is with **, then only the deleted points are considered. Similar definitions can be applied to vectors, matrices and sample sizes (like \( n^* \) and \( n^{**} \)).

7. \( \frac{\partial \Sigma}{\partial \rho} \): derivative of matrix \( \Sigma \). This is defined as
\[
\frac{\partial \Sigma}{\partial \rho} = \left( \frac{\partial \sigma_{ij}}{\partial \rho} \right), \text{ where } \sigma_{ij} \text{ is the } (i,j)\text{th entry of matrix } \Sigma.
\]
Chapter III

Asymptotic distributions: \( \rho \) known

In this chapter, we will derive the asymptotic distributions of the m.l.e.'s \( \hat{\theta}, \hat{\tau} \) and \( \hat{\delta}^2 \) in the case where \( \rho \) is known. The asymptotic distributions in the case of unknown \( \rho \) will be derived in chapter 4. There are four sections in Chapter 3. In §3.1, we give an example illustrating why classical methods can not be applied to derive the asymptotic distributions. Then, we derive the asymptotic distributions of \( \hat{\theta} \) and \( \hat{\tau} \) when \( \rho \) is known in §3.2 and §3.3, respectively. In §3.4, we illustrate how to find the limiting covariance matrices of \( \sqrt{n} (\hat{\theta} - \theta^{(o)}) \) and \( \sqrt{n} (\hat{\tau} - \tau^{(o)}) \). The asymptotic distribution of \( \hat{\delta}^2 \) in the case where \( \rho \) is known is discussed in §3.5.

§3.1. Example

From (2.8) in §2.2, we know that the maximum likelihood estimators of \( \theta^{(o)} \) and \( \tau^{(o)} \) are the values of \( \theta \) and \( \tau \) that minimize

\[
S(\theta, \tau; y) = \frac{1}{n} \left( \hat{y} - \hat{\mu}(\theta, \tau, \rho, t_n) \right)' \left( \hat{y} - \hat{\mu}(\theta, \tau, \rho, t_n) \right) \\
= \frac{1}{n} \left( y - \mu(\theta, \tau; t_n) \right)' \Sigma_n^{-1} \left( y - \mu(\theta, \tau; t_n) \right),
\]

(3.1)

where \( \Sigma_n \) is the covariance matrix of the vector \( u \). Notation \( S(\theta, \tau; y) \) is used here to represent the generalized mean squared function. The second derivatives of \( S(\theta, \tau, y) \) with respect to the unknown
points $\tau$ are not continuous at the observed times. The following example illustrates this point.

Suppose we have the following model:

$$\mu(\tau, t) = \begin{cases} f_1(\tau, t) = 1 + t & \text{if } 0 < t \leq \tau \\ f_2(\tau, t) = 1 + 2\tau - t & \text{if } \tau < t \leq 1 \end{cases}$$

with $\sigma_0^2 = 1$. Suppose we take observations at three different times, say $t_1 < t_2 < t_3$. Then the covariance matrix of $[y_1, y_2, y_3]$ is

$$\mathbf{\Sigma}_3 = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{pmatrix}$$

and hence the inverse matrix of $\mathbf{\Sigma}_3$ is

$$\mathbf{\Sigma}_3^{-1} = \begin{pmatrix} 1 & -\rho & 0 \\ -\rho & 1+\rho^2 & -\rho \\ 0 & -\rho & 1 \end{pmatrix}.$$ 

So,

$$S(\tau, \mathbf{y}, \mathbf{t}_n) = \frac{1}{3} \left[ y_1 - \mu(\tau, t_1), y_2 - \mu(\tau, t_2), y_3 - \mu(\tau, t_3) \right] \times \begin{pmatrix} 1 & -\rho & 0 \\ -\rho & 1+\rho^2 & -\rho \\ 0 & -\rho & 1 \end{pmatrix} \begin{pmatrix} y_1 - \mu(\tau, t_1) \\ y_2 - \mu(\tau, t_2) \\ y_3 - \mu(\tau, t_3) \end{pmatrix}.$$ 

If $0 < \tau < t_1$, then we have
\[ S(\tau) = \frac{1}{3} \left[ y_1 - (1+2\tau-t_1), y_2 - (1+2\tau-t_2), y_3 - (1+2\tau-t_3) \right] \times \begin{pmatrix} 1 & -\rho & 0 \\ -\rho & 1+\rho^2 & -\rho \\ 0 & -\rho & 1 \end{pmatrix} \times \begin{pmatrix} y_1 - (1+2\tau-t_1) \\ y_2 - (1+2\tau-t_2) \\ y_3 - (1+2\tau-t_3) \end{pmatrix} \]

\[ = \frac{1}{3} \left\{ [y_1 - (1+2\tau-t_1)]^2 + (1+\rho^2) [y_2 - (1+2\tau-t_2)]^2 + [y_3 - (1+2\tau-t_3)]^2 - 2\rho [y_1 - (1+2\tau-t_1)] \times [y_2 - (1+2\tau-t_2)] - 2\rho [y_2 - (1+2\tau-t_2)] \times [y_3 - (1+2\tau-t_3)] \right\} \]

and

\[ \frac{dS(\tau)}{d\tau} = \frac{1}{3} \left\{ 4(\rho - 1)[y_1 - (1+2\tau-t_1)] + 4(\rho - 1)[y_3 - (1+2\tau-t_3)] + 8\rho [y_2 - (1+2\tau-t_2)] - 4(1+\rho^2)[y_2 - (1+2\tau-t_2)] \right\} \]

\[ \frac{d^2S(\tau)}{d\tau^2} = \frac{1}{3} \left\{ 8\rho^2 - 32\rho + 24 \right\}. \]

Similarly, for \( t_1 < \tau < t_2 \), we have

\[ S(\tau) = \frac{1}{3} \left[ y_1 - (1+t_1), y_2 - (1+2\tau-t_2), y_3 - (1+2\tau-t_3) \right] \times \]
\[
\begin{pmatrix}
1 & -\rho & 0 \\
-\rho & 1+\rho^2 & -\rho \\
0 & -\rho & 1
\end{pmatrix}
\begin{pmatrix}
y_1 - (1+t_1) \\
y_2 - (1+2\tau-t_2) \\
y_3 - (1+2\tau-t_3)
\end{pmatrix},
\]

\[
\frac{dS(\tau)}{d\tau} = \frac{1}{3} \left\{ 4\rho[y_1 + y_2 + y_3 - (1+t_1) - (1+2\tau-2t_2) - (1+2\tau-2t_3)] \\
- 4[y_3 - (1+2\tau-2t_3)] - 4(1+\rho^2)[y_2 - (1+2\tau-2t_2)] \right\} ,
\]

\[
\frac{d^2S(\tau)}{d\tau^2} = \frac{1}{3} \left\{ 8\rho^2 - 16\rho + 16 \right\}.
\]

and for \( t_2 < \tau < t_3 \), we have

\[
\frac{d^2S(\tau)}{d\tau^2} = \frac{8}{3} ,
\]

for \( t_3 < \tau < 1 \), we have

\[
\frac{d^2S(\tau)}{d\tau^2} = 0.
\]

So for any fixed value of \( \rho \), the second derivative of the log likelihood function is discontinuous at the observed points \( t_1, t_2 \) and \( t_3 \). (In fact, the first derivative is not continuous at these three points, either.). See Fig. 3.1 and Fig. 3.2, which show the plots of mean function and negative log likelihood function.
Figure 3.1. Mean Function

Figure 3.2. Likelihood Function

This example shows that the derivatives of the log likelihood function $S(\theta, \tau)$, not only for this example but for our model, with respect to the unknown knot point and the regression coefficients are not continuous at the design points. This discontinuity comes from the unknown knot points. So unknown knot points causes
trouble if we want to use classical methods to derive the asymptotic distributions of the m.l.e.'s of \( \theta \) and \( \tau \), since classical methods assume that the log likelihood function behaves like a paraboloid in the neighborhood of \( (\theta^{(o)}, \tau^{(o)}) \). Special treatment, therefore, has to be adopted to tackle this problem.

§3.2. Asymptotic distribution of \( \hat{\theta} \).

From §3.1, we know that it is the unknown knot vector \( \tau \) that makes the function \( S(\theta, \tau) \) irregular. Therefore, special attention has to be paid to the knot points \( \tau_1, \tau_2, \ldots, \tau_{r-1} \). The principal idea is to form a pseudo-problem by deleting a relatively small number of observations around \( \tau_j \). By showing that the difference between the estimator of the regression coefficient vector \( \hat{\theta} \) for the original problem and the estimator for the pseudo-problem \( \hat{\theta}^* \) is small as \( n \) goes to infinity, we can argue that the asymptotic behavior of the estimator of the regression coefficient vector in the original problem is the same as that in the pseudo-problem.

This section is further divided into four subsections. We first describe the deleting process that forms the pseudo-problem in §3.2.1. The derivation of the asymptotic distribution of \( \hat{\theta} \) is stated in §3.2.2, §3.2.3 and §3.2.4. In §3.2.2, we show that the distance between the m.l.e. \( \hat{\theta}^* \) of \( \theta^{(o)} \) in the pseudo-problem and \( \theta^{(o)} \) is \( O_p(n^{-1/2}) \). The asymptotic distribution of \( \hat{\theta}^* \) is derived in §3.2.3. In §3.2.4, we prove that \( \sqrt{n}(\hat{\theta} - \theta^*) \) is \( O_p(1) \) and complete the derivation of the asymptotic distribution of \( \hat{\theta} \).

§3.2.1. Deleting process of the pseudo-problem
The basic reason for considering the pseudo-problem is that we wish to remove the irregular behavior of the log likelihood function $S(\theta, \tau)$ in the neighborhood of $(\theta^{(o)}, \tau^{(o)})$. By deleting a relatively small number of data points around $\tau_j^{(o)}$, the log likelihood function will look like a paraboloid in the neighborhood of $(\theta^{(o)}, \tau^{(o)})$ since there are no data points in the neighborhoods of $\tau_j^{(o)}$ and $\theta^{(o)}$. Therefore, the log likelihood function within the neighborhood of $\tau^{(o)}$ does not depend on $\tau^{(o)}$ and is twice differentiable in $\theta$. Thus, classical techniques can be used to derive the asymptotic distribution of $\hat{\theta}^*$.

The deleting process is as follows. Choose a neighborhood $L_j(n)$ centered at $\tau_j^{(o)}$ with length $l_j(n)$. The choice of $L_j(n)$ depends on the sample size $n$ such that as $n \to \infty$,

$$l_j(n) \to 0,$$

but

$$[\log \frac{n}{\log \log n}]^{\frac{1}{2}} \to \infty.$$

From Theorem 2.1, we know that this condition guarantees that $\hat{\tau}_j$ will be in $L_j(n)$ with large probability as $n \to \infty$, because $|\hat{\tau}_j - \tau_j^{(o)}|$ will be less than $l_j(n)$ with large probability. We further assume that the $t_{ni}$'s are distributed in such a way that only $o\left(\frac{n}{\log \log n}\right)$ observations are deleted in the interval $L \setminus \text{do4}(j)$. At first glance, $o\left(\frac{n}{\log \log n}\right)$ may be not appear small, but deleting $o\left(\frac{n}{\log \log n}\right)$ observations eventually only discards a small amount of information relative to the whole information as $n \to \infty$. 


We will show the following proposition in §3.2.2, §3.2.3 and §3.2.4:

**Proposition 2.** Under the conditions of Theorem 2.1, we have

1. \( \sqrt{n} ( \hat{\theta}^* - \theta^{(o)}) = O_p(1) \)
2. \( \sqrt{n} ( \hat{\theta}^* - \theta^{(o)}) \to N(0, \sigma_o^2 G(p)^{-1}) \)
3. \( \sqrt{n} ( \hat{\theta} - \theta^*) = o_p(1) \)

where \( G(p) \) is defined as in assumption 7.

The reason we want to show the above three results is that by showing that the difference between \( \hat{\theta} \) and \( \theta^* \) is very small, we can argue that \( \hat{\theta} \) and \( \theta^* \) have the same asymptotic distribution, which is normal with mean \( \theta \) and covariance matrix \( \sigma_o^2 G(p)^{-1} \).

§3.2.2 Proof that ( \( \hat{\theta}^* - \theta^{(o)} \)) is of order \( O_p \left( \frac{1}{\sqrt{n}} \right) \).

In this subsection, we will prove the first part of Proposition 2. The basic idea of this proof is from Feder except that we have a minor modification, which is due to the appearence of the covariance structure \( \Sigma^{-1} \) of the disturbance term \( u \). Since \( \Sigma_n^{-1} \) has nice banded form, we will take advantage of this fact to partition \( \Sigma_n^{-1} \) into a special pattern. This partition will make the proof that
\[
\frac{1}{n} \left( \frac{\partial \hat{\mu}(\theta^{(o)}, t_n)}{\partial \theta} \right) \Sigma_n^{-1} \left( \frac{\partial \hat{\mu}(\theta^{(o)}, t_n)}{\partial \theta} \right),
\]
and
\[
\frac{1}{n} \left( \frac{\partial \mu(\theta^{(o)}, \tau^{(o)}, t_n)}{\partial \theta} \Psi_n \right) \left( \frac{\partial \mu(\theta^{(o)}, \tau^{(o)}, t_n)}{\partial \theta} \Psi_n \right),
\]
are asymptotically equal easier and then the fact that \((\hat{\theta}^* - \theta^{(o)})\) is of order \(O_p \left( \frac{1}{\sqrt{n}} \right)\) can be established.

Now, let us prove the first part of Proposition 2. First, from the original problem we have
\[
S(\theta, \tau, y) = \frac{1}{n} \left( y - \mu(\theta, \tau; t_n) \right)' \Sigma_n^{-1} \left( y - \mu(\theta, \tau; t_n) \right). \tag{3.1}
\]
The function \(S(\theta, \tau, y)\) is not smooth enough at the \(t_n\)'s, as illustrated in §3.1. For the pseudo-problem, we let
\[
\tilde{S}(\theta, \tau) = \frac{1}{n} \left( \tilde{y} - \tilde{\mu}(\theta, \tau; t_n) \right)' \Sigma_n^{-1} \left( \tilde{y} - \tilde{\mu}(\theta, \tau; t_n) \right)
\]
where \(\tilde{y}_{ni} = \tilde{\mu}_i(\theta, \tau; t_n) = 0\) if \(t_n\) is deleted; \(\tilde{y}_{ni} = y_{ni}\) and \(\tilde{\mu}_i(\theta, \tau; t_n) = \mu(\theta, \tau; t_n)\) if \(t_n\) is not deleted. The \(\Sigma_n^{-1}\) in \(\tilde{S}(\theta, \tau)\) is the same as the \(\Sigma_n^{-1}\) in \(S(\theta, \tau)\). The reason that we let \(\tilde{S}(\theta, \tau)\) have the same "covariance structure" as \(S(\theta, \tau)\) is due to the observation that \(\Sigma_n^{-1}\) has a nice banded form. This special form of \(\Sigma_n^{-1}\) will make our later analysis easier.
We notice that the term pseudo is used because in practice we do not know the value of \( \tau^{(o)} \) and hence cannot delete design points around \( \tau^{(o)} \). The deleting process is executed as if we knew between which two consecutive times each of the \( \tau^{(o)} \) is located. Therefore, it is not hard to see that the convergence rate of \( \hat{\theta}^* - \theta^{(o)} \) will be at least the same order as that of \( \hat{\theta} - \theta^{(o)} \). Therefore we can write

\[
\hat{\theta}^* - \theta^{(o)} = O_p\left(n^{-1/2}(\log \log n)^{1/2}\right).
\]

From the continuity constraints of the mean function, we know that \( \tau \) is a continuous differentiable function of \( \theta \). So we have

\[
(\hat{\tau}^* - \tau^{(o)}) = O_p\left(\frac{(\log \log n)^{1/2}}{n}\right) \text{ also.}
\]

Therefore, as \( n \to \infty \), \( \hat{\tau}^*_j \) is in \( L_j(n) \) with large probability since then \( (\hat{\tau}^*_j, \tau^{(o)}_j) \) or \( (\tau^{(o)}_j, \hat{\tau}^*_j) \) will be contained in \( L_j(n) \). Now, by the deleting process, \( \hat{S}(\theta, \tau) \) within the \( L_j(n) \)'s does not depend on \( \tau \) and is smooth as a function of \( \theta \). Let us denote \( \hat{S}(\theta, \tau) \) as \( \hat{S}(\theta) \). Then with large probability, \( \hat{\tau}^* \) is in \( L_j(n) \) and

\[
\hat{S}(\theta) = \hat{S}(\theta^{(o)}) + \left( \frac{\partial \hat{S}(\theta^{(o)})}{\partial \theta} \right)'(\theta - \theta^{(o)})
\]

\[
+ \frac{1}{2}(\theta - \theta^{(o)})' \frac{\partial^2 \hat{S}(\theta^{(o)})}{\partial \theta \partial \theta}(\theta - \theta^{(o)}).
\]  

(3.3)

Letting \( \theta = \hat{\theta}^* \) be the m.l.e. of \( \theta^{(o)} \) in the pseudo-problem, we have
\[ n(\hat{\mathcal{S}}(\hat{\theta}^*) - \hat{\mathcal{S}}(\theta^{(o)})) = n \left( \frac{\partial \hat{\mathcal{S}}(\theta^{(o)})}{\partial \theta} \right)' (\hat{\theta}^* - \theta^{(o)}) + \frac{n}{2} \left( \hat{\theta}^* - \theta^{(o)} \right) \frac{\partial^2 \hat{\mathcal{S}}(\theta^{(o)})}{\partial \theta \partial \theta} (\hat{\theta}^* - \theta^{(o)}) \]
\[ = n^{\frac{3}{2}} \left( \frac{\partial \hat{\mathcal{S}}(\theta^{(o)})}{\partial \theta} \right)' n^{\frac{1}{2}} (\hat{\theta}^* - \theta^{(o)}) + \frac{1}{2} n^{\frac{3}{2}} (\hat{\theta}^* - \theta^{(o)}) \frac{\partial^2 \hat{\mathcal{S}}(\theta^{(o)})}{\partial \theta \partial \theta} \frac{1}{n^{\frac{1}{2}}} (\hat{\theta}^* - \theta^{(o)}) \]
\[ \leq 0. \] (3.4)

The inequality in (3.4) is due to the fact that \( \hat{\mathcal{S}}(\hat{\theta}^*) \leq \hat{\mathcal{S}}(\theta^{(o)}) \), since \( \hat{\theta}^* \) is the m.l.c. of the pseudo-problem. Now, denote \( \hat{\mu}(\theta, r, t_n) \) by \( \hat{\mu}^{(o)}(\theta, t_n) \) for \( \hat{r}^* \) in \( L_j(n) \). Then,
\[ \frac{\partial \hat{\mathcal{S}}(\theta^{(o)})}{\partial \theta} = -\frac{2}{n} \frac{\partial \hat{\mu}^{(o)}(\theta^{(o)}, t_n)}{\partial \theta} \Sigma_n^{-1} (\hat{y} - \hat{\mu}^{(o)}(\theta^{(o)}, t_n)). \] (3.5)

Define \( \hat{u} = \hat{y} - \hat{\mu}(\theta^{(o)}, t_n) \) and \( \hat{u}^* = u - \hat{u} \). Then we have
\[ \frac{\partial \hat{\mathcal{S}}(\theta^{(o)})}{\partial \theta} = -\frac{2}{n} \frac{\partial \hat{\mu}^{(o)}(\theta^{(o)}, t_n)}{\partial \theta} \Psi_n \Psi'_n \hat{u} \]
\[= \frac{-2}{n} \frac{\partial \hat{\mu}(\theta^{(0)}, t_n)}{\partial \theta} \Psi_n \Psi_n' (u - \hat{\mu}^*)
\]
\[= \frac{-2}{n} \frac{\partial \hat{\mu}(\theta^{(0)}, t_n)}{\partial \theta} \Psi_n \Psi_n' u + \frac{2}{n} \frac{\partial \hat{\mu}(\theta^{(0)}, t_n)}{\partial \theta} \Sigma_n^{-1} \hat{\mu}^*. \quad (3.6)\]

Let \(n_j^*\) be the number of \(t_{ni}\)'s that are kept in the \(j\)th interval between knot points, \(n_{ij}^*\) be the number of \(t_{ni}\)'s that are deleted around \(\tau_j^{(o)}\) and \(\hat{\ell}_{nj}^{*}\) be the \((m+1)r\) dimension column vector with the form \([0,0,...,0,1, t_{nj}^{*}, (t^2)^n_{nj}, ..., (t^m)^n_{nj}, 0,...,0]'\), where the double index \(n_{j,k}^*\) means that the observed design point \(t\) is the \(k\)th \(t\) kept in region \(j\), \(k=1,2,...,n_j^*, j=1,2,...,r\). Then

\[\frac{\partial \hat{\mu}(\theta^{(o)}, t)}{\partial \theta} = [\hat{\ell}_{n1}^*, \hat{\ell}_{n12}^*, ..., \hat{\ell}_{n1n1}^*, 0,...,0, \hat{\ell}_{n21}^*, ..., \hat{\ell}_{n2n2}^*, 0,...,0,...,0,...,\hat{\ell}_{nrn_r}^*].\]

Let \(\hat{\ell}_{ni}^* = [\hat{\ell}_{ni1}^*, ..., \hat{\ell}_{nii}^*]\) be the collection of columns in the design matrix corresponding to those design points in region \(i\) that are not deleted, where \(i=1,2,...,r\). For those deleted points around \(\tau_i^{(o)}\), we denote those \(n_i^*\) \((m+1)r\) dimension column vectors as \(0_i^* = [0,0,...,0]\). Then we can write
\[
\frac{\partial \hat{\mu}_n(\theta^{(o)}, t_n)}{\partial \theta} = [\hat{T}_1^*, 0_1, \hat{T}_2^*, 0_2, \ldots, 0_{r-1}, \hat{T}_r^*].
\] (3.6.1)

Similarly, for \( \frac{\partial \mu(\theta^{(o)}, \tau^{(o)}; t_n)}{\partial \theta} \) we can write

\[
\frac{\partial \mu(\theta^{(o)}, \tau^{(o)}; t_n)}{\partial \theta} = [\hat{\tau}_{n1}, \hat{\tau}_{n2}, \ldots, \hat{\tau}_{nn}]
\]

\[
= [\hat{T}_1^*, \hat{T}_1^*, \hat{T}_2^*, \hat{T}_2^*, \ldots, \hat{T}_{r-1}^*, \hat{T}_r^*].
\] (3.6.2)

where by the definition of \( \hat{\tau}_{nj}^* \), \( \hat{\tau}_{ni} = \hat{\tau}_{nj}^* \) if \( t_{ni} \) is the kth design point in region j in the pseudo-problem.

Notice that \( \Sigma_n^{-1} \) has the following form

\[
\Sigma_n^{-1} = \begin{pmatrix}
1 & -\rho & 0 & \ldots & 0 \\
-\rho & 1+\rho^2 & -\rho & \ldots & 0 \\
0 & -\rho & 1+\rho^2 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & 1
\end{pmatrix}
\]

Partitioning \( \Sigma_n^{-1} \) corresponding to the pseudo-problem and knot points, we have
\[
\Sigma_n^{-1} = \begin{pmatrix}
A_1^* & 0_{[1,1]} & 0 & 0 & \cdots & 0 \\
0_{(-\rho)} & \cdots & 0 & 0 & \cdots & 0 \\
0_{[1,1]} & \cdots & \cdots & \cdots & \cdots & \cdots \\
o & 0_{(-\rho)} & \cdots & \cdots & \cdots & \cdots \\
0 & 0_{(-\rho)} & \cdots & \cdots & \cdots & \cdots \\
0 & 0_{(-\rho)} & \cdots & \cdots & \cdots & \cdots \\
\end{pmatrix}
\]

Here for \(i=2,3,4,\ldots,r-1\), we define the \(n_i \times n_i\) matrix \(A_i^*\) by

\[
A_i^* = \begin{pmatrix}
1+\rho^2 & -\rho & 0 & \cdots & 0 \\
-\rho & 1+\rho^2 & -\rho & \cdots & 0 \\
0 & -\rho & 1+\rho^2 & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
0 & 0 & 0 & \cdots & 1+\rho^2 \\
\end{pmatrix},
\]

and for \(i=1\) and \(i=r\), we define
\[
A_1^* = \begin{pmatrix}
1 & -\rho & 0 & \cdots & 0 \\
-\rho & 1+\rho^2 & -\rho & \cdots & 0 \\
0 & -\rho & 1+\rho^2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & 1+\rho^2
\end{pmatrix}
\quad \text{and} \quad
A_r^* = \begin{pmatrix}
1+\rho^2 & -\rho & 0 & \cdots & 0 \\
-\rho & 1+\rho^2 & -\rho & \cdots & 0 \\
0 & -\rho & 1+\rho^2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & 1
\end{pmatrix}
\]

respectively. Also, the \( n_i \times n_i \)** matrix \( A_i^{**} \) is defined as

\[
A_i^{**} = \begin{pmatrix}
1+\rho^2 & -\rho & 0 & \cdots & 0 \\
-\rho & 1+\rho^2 & -\rho & \cdots & 0 \\
0 & -\rho & 1+\rho^2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & 1+\rho^2
\end{pmatrix}, \quad i=1,2,3,4,\ldots,r-1.
\]

Finally, we define

\[
\theta^{(-\rho)}_{[i,j]} = \begin{pmatrix}
0 & 0 & 0 & \cdots & -\rho \\
0 & 0 & 0 & \cdots & 0 \\
0 & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & 0
\end{pmatrix}
\quad \text{and} \quad
\theta^{[i,j]}_{(-\rho)} = [\theta^{(-\rho)}_{[i,j]}]' = \begin{pmatrix}
0 \\
0 \\
0 \\
\vdots \\
0
\end{pmatrix}
\]

where \( \theta^{(-\rho)}_{[i,j]} \) is an \( n_i \times n_j \) matrix if \( i=j \), and an \( n_i \times n_j \) matrix if \( i=j+1, \ i=1,2,\ldots,r \).
Now
\[ \hat{\mathbf{u}}^* = [0_{1}^*, u_1^*, 0_{2}^*, ..., u_{r-1}^*, 0_{r}^*], \]
therefore
\[ \frac{\partial \mathbf{u}^{(o)}(t_n)}{\partial \theta} \Sigma_n^{-1} \hat{\mathbf{u}}^* \]

\[ = \begin{bmatrix}
\hat{T}_1^* & 0_{11}^* & 0_{12}^* & ... & 0_{1r-1}^* & 0_{1r}^*
\hat{T}_2^* & 0_{22}^* & 0_{23}^* & ... & 0_{2r-1}^* & 0_{2r}^*
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots
\hat{T}_{r-1}^* & 0_{(r-1)(r-1)}^* & 0_{(r-1)(r-2)}^* & ... & 0_{(r-1)(r-2)}^* & 0_{r-1}^*
\hat{T}_r^* & 0_{r2}^* & 0_{r3}^* & ... & 0_{rr-1}^* & 0_{rr}^*
\end{bmatrix}
\]

\[ = \sum_{i=2}^{r-1} \hat{T}_i^* \left[ 0_{[i,i-1]}^* u_{i-1} + 0_{(-\rho)} u_i \right] + \hat{T}_1^* 0_{[1,1]}^* u_1 + \hat{T}_r^* 0_{[r,r-1]}^* u_{r-1}. \]
But we know

\[ 0_{\{-\rho\}}^{(i,i)} \mathbf{u}^* = ( -\rho \mathbf{u}^*[i], 0, 0, \ldots, 0 )' \]

\[ 0_{\{-\rho\}}^{\texttt{i,i-1}} \mathbf{u}^* = ( 0, 0, \ldots, -\rho \mathbf{u}^*[i-1] )' \]

where the random variable \( \mathbf{u}^*[i] \) corresponds to the first deleted design point before \( \tau^*_i \); \( \mathbf{u}^*[i] \) is the random variable corresponding to the last deleted design point after \( \tau^*_i \). The matrix \( \mathbf{T}_i^* \), whose columns correspond to the column vectors in the original design matrix in region \( i \) with data points kept, has the property that its only nonzero rows are in the \( i \)th block. So if we let \( [\hat{T}^*_1]' = [\mathbf{Q}', \mathbf{0}'] \), where

\[
\mathbf{Q} = \begin{pmatrix}
1 & 1 & 1 \\
\mathbf{t}^*_{n_11} & \mathbf{t}^*_{n_12} & \cdots & \mathbf{t}^*_{n_1n_1} \\
\mathbf{t}^{2*}_{n_11} & \mathbf{t}^{2*}_{n_12} & \cdots & \mathbf{t}^{2*}_{n_1n_1} \\
\cdots & \cdots & \cdots & \cdots \\
\mathbf{t}^{m*}_{n_11} & \mathbf{t}^{m*}_{n_12} & \cdots & \mathbf{t}^{m*}_{n_1n_1}
\end{pmatrix},
\]

then
\[
\hat{T}^*_{10}(-\rho)_{u_1} = \begin{pmatrix}
-pu^{**}[1] \\
-\rho t_{n1}u^{**}[1] \\
\vdots \\
-\rho t_{m1}u^{**}[1] \\
0 \\
\vdots \\
\vdots \\
0
\end{pmatrix}
\]

Similarly, we can obtain

\[
\hat{T}^*_{r0}(-\rho)_{[r_{r-1}]u_{r-1}} = \begin{pmatrix}
0 \\
\vdots \\
\vdots \\
0 \\
-\rho u^{**}[r-1] \\
-\rho t_{n,r}u^{**}[r-1] \\
\vdots \\
\vdots \\
-\rho t_{m,r}u^{**}[r-1] \\
n_{r,r}u^{**}[r-1]
\end{pmatrix}
\]

and
\[ \mathbf{T}_1^\ast \left[ \theta_{\rho, i-1}^{(\rho)} u_{i-1} + \theta_{\rho, i}^{(\rho)} u_i \right] = \begin{pmatrix} 0 \\ \\ 0 \\ \\ -\rho(u^{**}[i-1] + u^{**}[i]) \\ \\ -\rho t_{n_i n_i} u^{**}[i-1] - \rho t_{n_i 1} u^{**}[i] \\ \\ . \\ \\ . \\ \\ . \\ \\ -\rho t_{n_i n_i} u^{**}[i-1] - \rho t_{n_i 1} u^{**}[i] \\ \\ 0 \\ \\ . \\ \\ 0 \end{pmatrix} \]
Therefore, \[
\frac{\partial \mu^*(\theta^{(o)}, t_n)}{\partial \theta} \Sigma_n^{-1} \hat{u}^* = \begin{pmatrix}
-\rho u^{**}[1] \\
-\rho t_{n1} u^{**}[1] \\
\vdots \\
-\rho t_{n1} u^{**}[1] \\
-\rho (u^{**}[1] + u^{**}[2]) \\
-\rho t_{n2} u^{**}[1] - \rho t_{n2} u^{**}[2] \\
\vdots \\
-\rho t_{n2} u^{**}[1] - \rho t_{n2} u^{**}[2] \\
\vdots \\
-\rho u^{**}[r-1] \\
-\rho t_{nr} u^{**}[r-1] \\
\vdots \\
-\rho t_{nr} u^{**}[r-1]
\end{pmatrix}
\]

Since \( u_{ni} \) has fixed variance, we have \([1/\sqrt{n}] u_{ni} = o_p(1)\). Also, every component of \( n^{1/2} \frac{\partial \mu^*(\theta^{(o)}, t_n)}{\partial \theta} \Sigma_n^{-1} \hat{u}^* \) contains only finitely many \( u_{ni} \)'s, therefore each component is of \( o_p(1) \). So, we can write
\[
\frac{1}{\sqrt{n}} \sum_n^{-1} \mathbf{u}^* = o_p(1).
\]

Hence from (3.6), we have

\[
\frac{1}{n^2} \frac{\partial S(\theta^{(0)})}{\partial \theta} = -2 \frac{\partial \mu(\theta^{(0)}, t_n)}{\partial \theta} \Psi_n \Psi_n' \mathbf{u} + o_p(1)
\]

\[
= -2 \frac{\partial \mu(\theta^{(0)}, t_n)}{\partial \theta} \Psi_n \mathbf{e} + o_p(1) \quad (3.7)
\]

We know that the vector \( \mathbf{e} \) is normally distributed with mean 0 and covariance matrix \( \sigma^2 I \). Therefore the random vector

\[
\frac{1}{\sqrt{n}} \frac{\partial \mu(\theta^{(0)}, \tau^{(0)}, t_n)}{\partial \theta} \Psi_n \mathbf{e}
\]

has mean 0 and, from assumption 7 in §2.1, has uniformly bounded variance as \( n \to \infty \). So

\[
\frac{1}{\sqrt{n}} \frac{\partial \mu(\theta^{(0)}, \tau^{(0)}, t_n)}{\partial \theta} \Psi_n \mathbf{e} = o_p(1).
\]

Now if we can show that

\[
\frac{1}{n} \left( \frac{\partial \mu(\theta^{(0)}, t_n)}{\partial \theta} \right) \sum_n^{-1} \left( \frac{\partial \mu(\theta^{(0)}, t_n)}{\partial \theta} \right)' - \frac{1}{n} \left( \frac{\partial \mu(\theta^{(0)}, \tau^{(0)}, t_n)}{\partial \theta} \right) \Psi_n \left( \frac{\partial \mu(\theta^{(0)}, \tau^{(0)}, t_n)}{\partial \theta} \right) \Psi_n'
\]

converges to 0, then

\[
\frac{1}{\sqrt{n}} \frac{\partial \mu(\theta^{(0)}, t_n)}{\partial \theta} \Psi_n \quad \text{and} \quad \frac{1}{\sqrt{n}} \frac{\partial \mu(\theta^{(0)}, \tau^{(0)}, t_n)}{\partial \theta} \Psi_n
\]
have the same order and we can conclude that \( \frac{1}{2} \frac{\partial s'(\theta)}{\partial \theta} \) is of \( O_p(1) \).

Notice from (3.6.1) and (3.6.2) that we have

\[
\frac{\partial \mu^*(\theta^{(o)}, t_n)}{\partial \theta} = [\hat{\mu}^*_1, \theta^{**}_1, \hat{\mu}^*_2, \theta^{**}_2, \ldots, \theta^{**}_{r-1}, \hat{\mu}^*_r]
\]  

(3.6.1)

and

\[
\frac{\partial \mu(\theta^{(o)}, \tau^{(o)}; t_n)}{\partial \theta} = [\hat{\tau}^{**}_n, \hat{\tau}^{**}_{n-1}, \ldots, \hat{\tau}^{**}_1, \hat{\tau}^*_r]
\]

(3.6.2)

Letting

\[
R^{**} = \frac{\partial \mu(\theta^{(o)}, \tau^{(o)}; t)}{\partial \theta} - \frac{\partial \mu^*(\theta^{(o)}, t)}{\partial \theta}
\]

\[
= [\theta^{*1}, \hat{\tau}^{**}_1, \theta^{*}_1, \hat{\tau}^{**}_2, \ldots, \hat{\tau}^{**}_{r-1}, \theta^{*}_r]
\]

(3.7.1)

we can write

\[
\left( \frac{\partial \mu^*(\theta^{(o)}, t_n)}{\partial \theta} \right) \Sigma_n^{-1} \left( \frac{\partial \mu^*(\theta^{(o)}, t_n)}{\partial \theta} \right)^t.
\]
\[
\begin{align*}
&= \left( \frac{\partial \mu(\theta^{(o)}, \tau^{(o)}; t_n)}{\partial \theta} - R^{**} \right) \Sigma_n^{-1} \left( \frac{\partial \mu(\theta^{(o)}, \tau^{(o)}; t_n)}{\partial \theta} - R^{**} \right)' \\
&= \left( \frac{\partial \mu(\theta^{(o)}, \tau^{(o)}; t_n)}{\partial \theta} \right) \Sigma_n^{-1} \left( \frac{\partial \mu(\theta^{(o)}, \tau^{(o)}; t_n)}{\partial \theta} \right)' + R^{**} \Sigma_n^{-1} R^{**}, \\
&\quad \quad - 2 \left( \frac{\partial \mu(\theta^{(o)}, \tau^{(o)}; t_n)}{\partial \theta} \right) \Sigma_n^{-1} R^{**}'. 
\end{align*}
\]

(3.8)

Now

\[
R^{**} \Sigma_n^{-1} \left( \frac{\partial \mu(\theta^{(o)}, \tau^{(o)}; t_n)}{\partial \theta} \right)',
\]

\[
= \begin{bmatrix} 0^* & \hat{T}_1^* & 0^* & \hat{T}_2^* & \ldots & \hat{T}_{r-1}^*, 0^* \end{bmatrix} \Sigma_n^{-1} \begin{bmatrix} \hat{T}_1^* \\ \hat{T}_1^{**} \\ \hat{T}_2^* \\ \hat{T}_2^{**} \\ \vdots \\ \hat{T}_{r-1}^{**} \\ \hat{T}_r^* \\ \hat{T}_r^{**} \end{bmatrix}
\]
\[ \begin{align*}
&= \sum_{i=1}^{r-1} \begin{bmatrix} T_{i}^{**} & A_i & T_{i}^{**} \end{bmatrix} + \sum_{i=1}^{r-2} \begin{bmatrix} T_{i}^{**} \theta \begin{pmatrix} i+1,i \end{pmatrix} & + T_{i+1}^{**} \theta \begin{pmatrix} r+1,r+1 \end{pmatrix} \end{bmatrix} T_{i+1}^{*} \\
&\quad + T_{1}^{**} \theta \begin{pmatrix} -r \end{pmatrix} T_{1}^{*} + T_{r}^{**} \theta \begin{pmatrix} r+r-1 \end{pmatrix} T_{r}^{*}
\end{align*} \tag{3.8.1} \]

By the assumptions in Chapter 2 about the design points and the special form of \( A_i^{**} \), we can see that \( T_{i}^{**} A_i T_{i}^{**} \) is of \( O(n_i^{**}) \). Also, by the forms of \( \theta \begin{pmatrix} -r \end{pmatrix} \) and \( \theta \begin{pmatrix} i+1,i \end{pmatrix} \), we know \( T_{i}^{**} \theta \begin{pmatrix} i+1,i \end{pmatrix} T_{i+1}^{*} \) and \( T_{i+1}^{**} \theta \begin{pmatrix} i+1,i+1 \end{pmatrix} T_{i+1}^{*} \) both are at most of \( O(n_i^{**}) \). So the matrix

\[ R^{**} \Sigma_n^{-1} \left( \frac{\partial \mu(\theta^{(o)}, \tau^{(o)}; t_n)}{\partial \theta} \right)' \]

is at most of order \( O(n^{**}) \). Because of the assumptions about the number of deleted design points in the pseudo-problem, we have

\[ \frac{1}{n} R^{**} \Sigma_n^{-1} \left( \frac{\partial \mu(\theta^{(o)}, \tau^{(o)}; t_n)}{\partial \theta} \right)' = O(1). \]

To evaluate \( R^{**} \Sigma_n^{-1} R^{**}' \), we notice that

\[ R^{**} = \begin{bmatrix} 0_1^*, T_1^{**}, 0_2^*, ..., T_{r-1}^{**}, 0_r^* \end{bmatrix} \quad (3.7.1) \]

and
\[
\frac{\partial \mu(\theta^{(o)}, \tau^{(o)}; t_n)}{\partial \theta} = [\hat{T}_1^*, \hat{T}_1^*, \hat{T}_2^*, \hat{T}_2^*, \ldots, \hat{T}_{r-1}^*, \hat{T}_{r-1}^*, \hat{T}_r^*, \hat{T}_r^*]. \quad (3.6.2)
\]

Comparing \( R^{**} \) to \( \frac{\partial \mu(\theta^{(o)}, \tau^{(o)}; t_n)}{\partial \theta} \), we observe that the matrix \( R^{**} \) can be obtained by replacing the matrices \( \hat{T}_i^* \) in \( \frac{\partial \mu(\theta^{(o)}, \tau^{(o)}; t_n)}{\partial \theta} \) by \( 0_i^* \).

Making the corresponding substitution in (3.8.1), it follows that

\[
\frac{1}{n} R^{**} \Sigma_n^{-1} R^{**} = \sum_{i=1}^{r-1} \hat{T}_i^{**} A_i \hat{T}_i^{**}.
\]

But we have already established that \( \hat{T}_i^{**} A_i \hat{T}_i^{**} = o(n_i^{**}) \), so it follows that \( \frac{1}{n} R^{**} \Sigma_n^{-1} R^{**} \) is \( o(1) \).

Now we know \( \frac{1}{n} R^{**} \Sigma_n^{-1} \left( \frac{\partial \mu(\theta^{(o)}, \tau^{(o)}; t_n)}{\partial \theta} \right)' = o(1) \) and \( \frac{1}{n} R^{**} \Sigma_n^{-1} R^{**} = o(1) \). Multiplying both sides of (3.8) by \( \frac{1}{n} \), we have

\[
\frac{1}{n} \left( \frac{\partial \mu(\theta^{(o)}, t_n)}{\partial \theta} \right) \Sigma_n^{-1} \left( \frac{\partial \mu(\theta^{(o)}, t_n)}{\partial \theta} \right)'.
\]
\[
\begin{aligned}
&= \frac{1}{n} \left( \frac{\partial \mu(\theta^{(o)}, \tau^{(o)}; t_n)}{\partial \theta} \right) \Sigma_n^{-1} \left( \frac{\partial \mu(\theta^{(o)}, \tau^{(o)}; t_n)}{\partial \theta} \right)' \\
&\quad - \frac{2}{n} R^{**} \Sigma_n^{-1} \left( \frac{\partial \mu(\theta^{(o)}, \tau^{(o)}; t_n)}{\partial \theta} \right)' + \frac{1}{n} R^{**} \Sigma_n^{-1} R^{**} \\
&= \frac{1}{n} \left( \frac{\partial \mu(\theta^{(o)}, \tau^{(o)}; t_n)}{\partial \theta} \right) \Sigma_n^{-1} \left( \frac{\partial \mu(\theta^{(o)}, \tau^{(o)}; t_n)}{\partial \theta} \right)' + o(1).
\end{aligned}
\]

Now, by assumption 7 in §2.1, we have

\[
\frac{\partial \mu(\theta^{(o)}, \tau^{(o)}; t_n)}{\partial \theta} \Sigma_n^{-1} \left( \frac{\partial \mu(\theta^{(o)}, \tau^{(o)}; t_n)}{\partial \theta} \right)' \to G(\rho).
\]

Here \(G(\rho)\) is a positive definite matrix. So, as \(n \to \infty\),

\[
\frac{1}{n} \left( \frac{\partial \hat{\mu}(\theta^{(o)}; t_n)}{\partial \theta} \right) \Sigma_n^{-1} \left( \frac{\partial \hat{\mu}(\theta^{(o)}; t_n)}{\partial \theta} \right)' \to G(\rho),
\]

Therefore, \(n^2 \frac{\partial \mathcal{L}(\theta^{(o)})}{\partial \theta} \), being a linear combination of jointly normal random variables with bounded covariance as \(n \to \infty\), is of \(O_p(1)\). Now

\[
\frac{1}{2} \frac{\partial^2 \mathcal{L}(\theta^{(o)})}{\partial \theta \partial \theta} = \frac{1}{n} \left( \frac{\partial \hat{\mu}(\theta^{(o)}; t_n)}{\partial \theta} \right) \Sigma_n^{-1} \left( \frac{\partial \hat{\mu}(\theta^{(o)}; t_n)}{\partial \theta} \right)' \to G(\rho)
\]

So inequality (3.4) implies that \(\sqrt{n} \left( \hat{\theta} - \theta^{(o)} \right) = O_p(1)\).
§3.2.3 Proof that $\sqrt{n}(\hat{\Theta}^* - \theta^{(o)})$ is asymptotically normal

Now, let us show the second part of Proposition 2 by establishing that $\sqrt{n}(\hat{\Theta}^* - \theta^{(o)})$ asymptotically follows a $N(0, \sigma_0^2 G^{-1}(\rho))$ distribution. Since the proof of the joint asymptotic normality of $\sqrt{n}(\hat{\Theta}^* - \theta^{(o)})$ turns into the proof of the joint asymptotic normality of a $r(m+1)$-dimension random vector, we need a theorem due to Jennrich (1969).

In the following we let $e_i$'s be i.i.d. errors with mean 0 and finite variance $\sigma_0^2$ and $x_1, x_2, \ldots, x_p$ be $p$ n-dimension vectors. Define the tail product of $x_i$ and $x_j$ as

$$\lim_{n \to \infty} (x_i, x_j)_n = \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} x_{ki} x_{kj}, \quad i, j = 1, 2, \ldots, p.$$ 

Also, we let

$$Q = \lim_{n \to \infty} \frac{1}{n} \left( \begin{array}{c} x_1' \\ x_2' \\ \vdots \\ x_p' \end{array} \right) \times (x_1, x_2, \ldots, x_p),$$

where $Q$ is a positive definite matrix.
Theorem 3.1. (Jennrich, 1969) If $e_i$'s are i. i. d. errors with mean 0 and finite variance $\sigma_0^2$ and if all possible tail products of the p n-dimension vectors $x_1, x_2, ..., x_p$ of real numbers exist, then

$$\sqrt{n} \left( (x_1, e)_n, (x_2, e)_n, ..., (x_p, e)_n \right) \to N(\theta, \sigma_0^2 \mathbf{Q})$$

as $n \to \infty$.

Now let us prove the second part of Proposition 2. We know $\theta^{(o)}$ is a relative interior point of $\Omega$, so $\hat{\theta}^*$ is in $\Omega$ with large probability as $n \to \infty$. Since $\hat{\theta}^*$ is the m.i.e. of the pseudo-problem, we have

$$\frac{\partial S(\hat{\theta}^*)}{\partial \theta} = 0.$$  

Also, when $\hat{\tau}$ is near $\tau^{(o)}$, $\hat{\theta}^* - \theta^{(o)} = o_p\left(\frac{1}{n} \right)$. Therefore, $\frac{\partial S(\theta)}{\partial \theta}$, considered as a function of $\theta$, is linear in $\theta$. We have

$$\frac{\partial S(\hat{\theta}^*)}{\partial \theta} = \frac{\partial S(\theta^{(o)})}{\partial \theta} + \frac{\partial^2 S(\theta^{(o)})}{\partial \theta \partial \theta} (\hat{\theta}^* - \theta^{(o)}) = 0.$$  

So,

$$- \frac{\partial S(\theta^{(o)})}{\partial \theta} = \frac{\partial^2 S(\theta^{(o)})}{\partial \theta \partial \theta} (\hat{\theta}^* - \theta^{(o)})$$  

(3.9)

Multiplying both sides of (3.9) by $\frac{1}{2} \frac{1}{n^2}$, we have
- \frac{1}{2} n^2 \frac{1}{\sigma^2} \frac{\partial S(\theta^{(o)})}{\partial \theta} = \frac{1}{2} \left( \frac{\partial^2 S(\theta^{(o)})}{\partial \theta \partial \theta} \right) \frac{1}{\sigma^2} (\hat{\theta}^* - \theta^{(o)}).

But from (3.6), we have

\frac{\partial S(\theta^{(o)})}{\partial \theta} = -2 \frac{\partial \mu(\theta^{(o)}, t_n)}{\partial \theta} \frac{\partial \psi_n}{\partial \theta} \epsilon.

Also, from (3.5) we have

\frac{1}{2} \left( \frac{\partial^2 S(\theta^{(o)})}{\partial \theta \partial \theta} \right) \frac{1}{\sigma^2} (\hat{\theta}^* - \theta^{(o)})

= \frac{1}{n} \left[ \left( \frac{\partial \mu(\theta^{(o)}, \tau^{(o)}, t_n)}{\partial \theta} \right) \Sigma_n^{-1} \left( \frac{\partial \mu(\theta^{(o)}, \tau^{(o)}, t_n)}{\partial \theta} \right)' + o(1) \right] \frac{1}{\sigma^2} (\hat{\theta}^* - \theta^{(o)}) + o_p(1)

= (G(\rho) + o(1)) \frac{1}{\sigma^2} (\hat{\theta}^* - \theta^{(o)}) + o_p(1)

But we know

\frac{-1}{n^2} \left( \frac{\partial \mu(\theta^{(o)}, t_n)}{\partial \theta} \right) \psi_n + o(1) \epsilon = \frac{-1}{n^2} \left( \frac{\partial \mu(\theta^{(o)}, \tau^{(o)}, t_n)}{\partial \theta} \right) \psi_n + o(1) \epsilon

= n^2 \left( \frac{\partial \mu(\theta^{(o)}, \tau^{(o)}, t_n)}{\partial \theta} \right) \psi_n \epsilon + o_p(1).
By Theorem 3.1 and assumption 7, we know
\[ n^{-\frac{1}{2}} \mathcal{D}_{\theta}(\theta^{(o)}, \tau^{(o)}, t_n) \] asymptotically follows a \( N(0, \sigma_0^2 G(\rho)) \) distribution. So
\[ \frac{1}{n^{\frac{1}{2}}} (\hat{\theta}^* - \theta^{(o)}) = \left( G(\rho) + o(1) \right)^{-1} n^{-\frac{1}{2}} \left( -\frac{\partial \hat{\mu}(\theta^{(o)}, t_n)}{\partial \theta} \Psi_n + o(1) \right) e \] asymptotically follows a \( N(0, \sigma_0^2 G^{-1}(\rho)) \) distribution. The proof of (2) in Proposition 2 is complete.

§3.2.4 Proof that \( \hat{\theta}^* \) is close to \( \hat{\theta} \) with the order of \( o_p(1) \)

From Theorem 2.1, we know that as \( n \) diverges to infinity, \( \hat{\theta} \) converges to \( \theta^{(o)} \) at least at the rate \( (\sqrt{\log \log n} / \sqrt{n}) \). Also, \( \hat{\theta}^* \) converges to \( \theta^{(o)} \) at least at the rate \( n^{-\frac{1}{2}} \) if \( n \) is large enough. Therefore, we can expect that \( \hat{\theta} \) and \( \hat{\theta}^* \) will be very close to each other if \( n \rightarrow \infty \). To show this, we will first investigate the generalized mean squared errors \( S(\theta, \tau) \) and partition it into the generalized mean squared errors of the pseudo-problem evaluated at the same point \( S^*(\theta, \tau) \) plus another two unimportant terms. The closeness of \( \hat{\theta} \) and \( \hat{\theta}^* \) then can be established by using the properties of the m.l.e. and a Taylor's expansion of function \( S^*(\theta, \tau) \).

In §2.3, we define \( \mu^{(o)} - \mu = \nu \). Also, we know that \( \mathbf{y} - \mu^{(o)} = \mathbf{u} \). So we can write
\[
(y - \mu(\theta, \tau; t_n))^\prime \Sigma_n^{-1} (y - \mu(\theta, \tau; t_n)) = (y - \mu)^\prime \Sigma_n^{-1} (y - \mu)
\]

\[
= (y - \mu^{(o)} + \mu^{(o)} - \mu)^\prime \Sigma_n^{-1} (y - \mu^{(o)} + \mu^{(o)} - \mu)
\]

\[
= (u + v)^\prime \Sigma_n^{-1} (u + v)
\]

\[
= (u + v)^* \Sigma_n^{-1} (u + v)* + (u + v)** \Sigma_n^{-1} (u + v)**
\]

\[
+ 2(u + v)^* \Sigma_n^{-1} (u + v)**
\]

\[
= (u + v)^* \Sigma_n^{-1} (u + v)* + u** \Sigma_n^{-1} u** + v** \Sigma_n^{-1} v**
\]

\[
+ u^* \Sigma_n^{-1} u** + v^* \Sigma_n^{-1} v** + 2 v** \Sigma_n^{-1} u** + v^* \Sigma_n^{-1} u**
\]

\[
+ v** \Sigma_n^{-1} u^* .
\]

(3.11)

We will prove that \(v** \Sigma_n^{-1} v**\) and \(v^* \Sigma_n^{-1} v**\) are of \(o(1)\), and

\(v** \Sigma_n^{-1} u**, v^* \Sigma_n^{-1} u**\) and \(v** \Sigma_n^{-1} u*\) are of \(o_p(1)\). Before doing that, we will show \(v** v**\) is \(o(1)\).

Select \(a_n > 0\) such that \(a_n/[\sqrt{\log \log n}] \to \infty\), \(a_n = o(\sqrt{n/n**})\).

Define
\[ \mathcal{U}_n = \{ (\theta, \tau) : \tau = \tau(\theta), \theta \in \Omega \text{ and } ||\theta - \theta^{(0)}|| < a_n n^{-\frac{1}{2}}, \tau_j \in L_j(n), j=1,2,\ldots, r-1 \}. \]

Notice that \( n^{**} = o(n/\log \log n) \), so
\[
an^{-\frac{1}{2}} = o((n/n^{**})^{\frac{1}{2}} n^{-\frac{1}{2}}) = o((n^{**})^{-\frac{1}{2}}) = o(n^{-\frac{1}{2}} (\log \log n)^{\frac{1}{2}}).\]

As a result, \((\hat{\theta}, \hat{\tau})\) and \((\theta^*, \tau^*)\) both lie in \( \mathcal{U}_n \) with large probability as \( n \to \infty \) (from Theorems 2.1 and 2.2). Since the mean function satisfies the Lipschitz condition \( |\mu^{(o)}(t_{ni}) - \mu(t_{ni})| \leq c |\theta - \theta^{(o)}|\) for \( \theta \) in the neighborhood of \( \theta^{(o)} \) and some positive constant \( c \), then

\[
sup_{(\theta, \tau) \in \mathcal{U}_n} \max_{t_{ni} \in UL_j(n)} |\mu^{(o)}(t_{ni}) - \mu(t_{ni})| \\
\leq c \times \sup_{(\theta, \tau) \in \mathcal{U}_n} ||\theta - \theta^{(o)}|| \\
\leq c \times a_n n^{-\frac{1}{2}}. \tag{3.11.1} \]

Therefore,
\[
sup_{(\theta, \tau) \in \mathcal{U}_n} \left| \frac{1}{n} v^{**}, v^{**} \right| = \frac{1}{n} \sup_{(\theta, \tau) \in \mathcal{U}_n} \left| v^{**}, v^{**} \right| \\
\leq \frac{1}{n} \left( c \times a_n n^{-\frac{1}{2}} \right)^2 (n^{**}) \]
\[ = \frac{1}{n} c^2 a_n n^{-1}(n^{**}) \]

\[ = \frac{1}{n} c^2 o\left(\frac{1}{n} (\log \log n)\right) \times o\left(\frac{n}{(\log \log n)}\right) \]

\[ = o\left(\frac{1}{n}\right) \quad (3.12) \]

Now we want to prove that \( v^{**} \sum^{-1}_n v^{**} \) and \( v^{*} \sum^{-1}_n v^{*} \) are of \( o(1) \). Partitioning \( \sum^{-1}_n \) as before and writing

\[ v^{**} = [0', v^{*}_{1}, 0', v^{*}_{2}, 0', \ldots, 0', v^{*}_{r-1}, 0'] , \]

where \( v^{*}_{i} = [v^{**}_{n_i 1}, v^{**}_{n_i 2}, \ldots, v^{**}_{n_i n_i}] \) is a vector corresponding to the design points deleted near to \( \tau_i^{(o)} \), then

\[ v^{**} \sum^{-1}_n v^{**} \]

\[ = [v^{**}_{1} \theta^{(\rho)}_{[1,1]}, v^{**}_{1} A_1, v^{**}_{1} \theta^{(\rho)}_{[2,1]} + v^{**}_{2} \theta^{(\rho)}_{[2,2]}, \ldots, v^{**}_{r-1} \theta^{(\rho)}_{[r,r-1]}] v^{**} \]

\[ = \sum_{i=1}^{r-1} v^{**}_{i} A_i v^{**}_{i} . \quad (3.13) \]
But for a typical $i$, $A_{ii}^{**}$ has the form

$$A_{ii}^{**} = \begin{pmatrix}
1 + \rho^2 & -\rho & \ldots & 0 \\
-\rho & 1 + \rho^2 & \ldots & 0 \\
0 & -\rho & \ldots & 0 \\
0 & 0 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
\vdots & \vdots & \ddots & -\rho \\
0 & 0 & \ldots & 1 + \rho^2
\end{pmatrix}.$$ 

So

$$v_i^{**} A_{ii}^{**} v_i^{**} = [v_{n_i 1}^{**}, v_{n_i 2}^{**}, \ldots, v_{n_i n_i}^{**}] \begin{pmatrix}
1 + \rho^2 & -\rho & \ldots & 0 \\
-\rho & 1 + \rho^2 & \ldots & 0 \\
0 & -\rho & \ldots & 0 \\
0 & 0 & \ldots & 0 \\
\vdots & \vdots & \ddots & -\rho \\
\vdots & \vdots & \ddots & 1 + \rho^2 \\
0 & 0 & \ldots & 1 + \rho^2
\end{pmatrix} \begin{pmatrix}
v_{n_i 1}^{**} \\
v_{n_i 2}^{**} \\
\vdots \\
v_{n_i n_i}^{**}
\end{pmatrix}$$

$$= (1 + \rho^2) \sum_{k=1}^{n_i} (v_{n_i k}^{**})^2 - 2\rho \sum_{k=1}^{n_i -1} (v_{n_i k}^{**})(v_{n_i k+1}^{**}) \geq 0.$$ 

Therefore

$$\sum_{i=1}^{r-1} v_i^{**} A_{ii}^{**} v_i^{**} = \sum_{i=1}^{r-1} [(1 + \rho^2) \sum_{k=1}^{n_i} (v_{n_i k}^{**})^2 - 2\rho \sum_{k=1}^{n_i -1} (v_{n_i k}^{**})(v_{n_i k+1}^{**})]$$
\[
(1+\rho^2) v^{**} v^{**} - \sum_{i=1}^{r-1} \sum_{k=1}^{n_i+1} \tau_{n_i k}^{**} (v^{**}_{n_i k})(v^{**}_{n_i k+1})
\]

Now \( v^{**} v^{**} = o(1) \), and

\[
-2 v^{**}_{i i} v_i \leq \sum_{k=1}^{n_i+1} (v^{**}_{n_i k})(v^{**}_{n_i k+1}) \leq 2 v^{**}_{i i} v_i
\]

So \( v^{**} \Sigma_n^{-1} v^{**} = o(1) \).

The next step is to show that \( v^{**} \Sigma_n^{-1} v^{**} \) is \( o_p(1) \). We know that \( v^{**} = [v^*_1, 0^*, v^*_2, 0^*, ..., v^*_r] \), where \( v^*_i = [v^*_i 1, v^*_i 2, ..., v^*_i n_i] \) is a vector corresponding to those points left in region \( i \). Then

\[
v^{**} \Sigma_n^{-1} v^{**} = \sum_{i=1}^{r-1} \left[ v^{**}_{i i} 0_{(-\rho)}^{[i,i]} + v^{**}_{i i+1} 0^{(-\rho)}_{[i+1,i]} \right] v^{**}_i
\]

\[
= \sum_{i=1}^{r-1} \left[ v^{**}_{n_i n_i} (-\rho) v^{**}_{n_i 1} + v^{**}_{n_i 1} (-\rho) v^{**}_{n_i n_i} \right]
\]

\[
= o(1)
\]

(3.14)
The reason we obtain $o(1)$ is that $v_{n_{ij}}^{**}$ and $v_{n_{ij}}^{***}$ are of $o(1)$.

To show $v^{**} \Sigma^{-1} u^{**}$, $v^{*} \Sigma^{-1} u^{**}$ and $v^{*} \Sigma^{-1} u^*$ are of $o_p(1)$, we need the following lemma due to Feder (1975a, Lemma 4.10.).

**Lemma 1.** Let $\mathcal{F} = \{ f(t) \}$ be the class of functions defined on $0 \leq t \leq 1$ such that $f(t)$ is composed of at most $s$ segments, each of which is a differentiable function processing at most $z$ sign changes in its derivative. And let $e_{ni}$'s be i.i.d. random variables with mean 0, variance $\sigma_o^2 < \infty$. Also, let $\mathcal{R}_m$ denote the collection of subsets $S$ of $[0,1]$, each of which consists of at most $m$ intervals. Then

$$\left| \sum_s f(t_{ni}) e_{ni} \right| \leq \left\{ \max_s |f(t_{ni})| \right\} \frac{1}{o_p(n^{2/3})}$$

As in (3.10), we can write $v^{**} \Sigma^{-1} u^{**}$ as

$$v^{**} \Sigma^{-1} u^{**} = \sum_{i=1}^{r-1} v_i^{*} A_i^{**} u_i^{**}.$$ 

For a typical $i$, $A_i^{**}$ can be written as
\[
A_i^{**} = \begin{pmatrix}
1 & -\rho & 0 & \cdots & 0 \\
-\rho & 1+\rho^2 & 0 & \cdots & 0 \\
0 & -\rho & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & \cdots & -\rho & 0 \\
0 & 0 & \cdots & 0 & 1
\end{pmatrix} + \begin{pmatrix}
\rho^2 & 0 & 0 & \cdots & 0 \\
0 & 0 & 0 & \cdots & 0 \\
0 & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & \cdots & 0 & \rho^2
\end{pmatrix}
\]

\[= \tilde{A}_i^{**} + \theta_i^{**} (\rho^2) = \Psi_i^{**} \Psi_i^{**} + \theta_i^{**} (\rho^2),\]

where

\[
\tilde{A}_i^{**} = \Psi_i^{**} \Psi_i^{**} = \begin{pmatrix}
1 & -\rho & 0 & \cdots & 0 \\
-\rho & 1+\rho^2 & 0 & \cdots & 0 \\
0 & -\rho & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & \cdots & -\rho & 0 \\
0 & 0 & \cdots & 0 & 1
\end{pmatrix}, \quad \theta_i^{**} (\rho^2) = \begin{pmatrix}
\rho^2 & 0 & 0 & \cdots & 0 \\
0 & 0 & 0 & \cdots & 0 \\
0 & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & \cdots & 0 & \rho^2
\end{pmatrix}.
\]

Then

\[
v_i^{**} A_i \ u_i = v_i^{**} \tilde{A}_i^{**} \ u_i + v_i^{**} \theta_i^{**} (\rho^2) \ u_i
\]

\[= v_i^{**} \Psi_i^{**} \Psi_i^{**} \ u_i + v_i^{**} \theta_i^{**} (\rho^2) \ u_i
\]

\[= v_i^{**} \Psi_i^{**} \Psi_i^{**} \ u_i + \alpha_p(1)
\]

Letting \( \Psi_i^{**} \ u_i = e_i \), then
\[ v_i^{**} A_i u_i = v_i^{**} \Psi_i e_i + o_p(1). \]

Since \( v_i^{**} \Psi_i^{**} \) satisfies the conditions for \( f(t_{ni}) \) in Lemma 1, if we set \( UL_j(n) = S \), then from Lemma 1 and (3.11.1) we have

\[
\sup_{(\theta, \tau) \in U_n} | v_i^{**} \Psi_i^{**} e_i^{**} |
\leq \sup_{(\theta, \tau) \in U_n} \left\{ \max_{t_{ni} \in UL_j(n)} | v_i^{**} e_i^{**} | \right\} 0_p(\sqrt{n_i^{**}})
\leq c \cdot a_n \cdot n_i^{**} 0_p(\sqrt{n_i^{**}})
\leq o(1/\sqrt{n_i^{**}}) 0_p(\sqrt{n_i^{**}})
= o_p(1).
\]

So \( v_i^{**} \Psi_i^{**} e_i^{**} = o_p(1) \) and hence \( v_i^{**} A_i u_i^{**} = o_p(1) \). Now

\[ v^{**} \Sigma_n^{-1} u^{**} = \sum_{i=1}^{r-1} v_i^{**} A_i u_i^{**} \], so we conclude that \( v^{**} \Sigma_n^{-1} u^{**} = o_p(1) \).

Now we want to show that \( v^{**} \Sigma_n^{-1} u^{**} \) and \( v^{**} \Sigma_n^{-1} u^{**} \) are

\[ o_p(1). \]

As in (3.14), we can obtain
\[
\begin{align*}
\mathbf{v}^{**} \Sigma_n^{-1} \mathbf{u} & = \sum_{i=1}^{r-1} \left[ \mathbf{v}'_i \mathbf{0}^{[i,i]} \theta^{(-\rho)} + \mathbf{v}'_{i+1} \theta^{[i+1,i]} \right] \mathbf{u}_i^* \\
& = \sum_{i=1}^{r-1} \left[ \mathbf{v}^{** \prime}_i \mathbf{0}^{[-\rho]} \mathbf{u}^{**}_i + \mathbf{v}^{** \prime}_{i+1} \theta^{[-\rho]} \mathbf{u}^{**}_i \mathbf{n}_i \mathbf{n}_i \right] \\
& = o_p(1). \quad (3.14.1)
\end{align*}
\]

The reason we obtain \( o_p(1) \) in (3.14.1) is that \( \mathbf{v}^{**}_n \), \( \mathbf{u}^{**}_n \) is normal with \( E(\mathbf{u}^{**}_n) = 0 \), \( Var(\mathbf{u}^{**}_n) = \sigma^2/(1-\rho^2) \) and there are only a finite number of terms in the summation. Similarly, for \( \mathbf{v}^{**} \Sigma_n^{-1} \mathbf{v}^{*} \) we have

\[
\begin{align*}
(\mathbf{v}^{**} \Sigma_n^{-1} \mathbf{v}^{*})' &= \sum_{i=1}^{r-1} \left[ \mathbf{u}'_i \mathbf{0}^{[i,i]} \theta^{(-\rho)} + \mathbf{u}'_{i+1} \theta^{[i+1,i]} \right] \mathbf{v}^*_i \\
& = \sum_{i=1}^{r-1} \left[ \mathbf{u}^{** \prime}_i \mathbf{0}^{[-\rho]} \mathbf{v}^{**}_i + \mathbf{u}^{** \prime}_{i+1} \theta^{[-\rho]} \mathbf{v}^{**}_i \mathbf{n}_i \mathbf{n}_i \right] \\
& = o_p(1).
\end{align*}
\]

Now, we have that \( \mathbf{v}^{**} \Sigma_n^{-1} \mathbf{v}^{**}, \mathbf{v}^{*} \Sigma_n^{-1} \mathbf{v}^{*} \) are of \( o_p(1) \), and
\[ v**' \Sigma_n^{-1} u** \] and \[ v**' \Sigma_n^{-1} u** \] are of \( o_p(1) \). From (3.11), we have

\[ (y - \mu)' \Sigma_n^{-1} (y - \mu) \]

\[ = (u + v)' \Sigma_n^{-1} (u + v) + u**' \Sigma_n^{-1} u** + u**' \Sigma_n^{-1} u** + o_p(1). \]

Therefore,

\[ S(\theta, \tau) = S(\theta, \tau) + \frac{1}{n} \left[ u**' \Sigma_n^{-1} u** + u**' \Sigma_n^{-1} u** \right] + o_p(\frac{1}{n}) \]

\[ = S(\theta, \tau) + \frac{1}{n} u' \Sigma_n^{-1} u** + o_p(\frac{1}{n}) \]  \hspace{1cm} (3.15)

Because \((\hat{\theta}, \hat{\tau})\) is the m.l.e. in the original design and \((\hat{\theta}^*, \hat{\tau}^*)\) are the m.l.e.'s in the pseudo-problem, we have

\[ S(\hat{\theta}, \hat{\tau}) \leq S(\hat{\theta}^*, \hat{\tau}^*) \]

and

\[ S^*(\hat{\theta}^*, \hat{\tau}^*) \leq S^*(\hat{\theta}, \hat{\tau}). \]

Therefore, from (3.15), with high probability

\[ 0 \leq S(\hat{\theta}^*, \hat{\tau}^*) - S(\hat{\theta}, \hat{\tau}) \]
\[ S^* (\hat{\theta}, \hat{\epsilon}) = S^* (\theta^*, \hat{\epsilon}^*) + \frac{1}{n} u' \Sigma_n^{-1} u^{**} + o_p \left( \frac{1}{n} \right) - \left[ S^* (\theta, \hat{\epsilon}) + \frac{1}{n} u' \Sigma_n^{-1} u^{**} + o_p \left( \frac{1}{n} \right) \right] \]

\[ = S^* (\theta^*, \hat{\epsilon}^*) - S^* (\theta, \hat{\epsilon}) + o_p \left( \frac{1}{n} \right) \]

Since \( S^* (\theta^*, \hat{\epsilon}^*) - S^* (\theta, \hat{\epsilon}) \leq 0 \), we have

\[ 0 \leq S(\theta^*, \hat{\epsilon}^*) - S(\theta, \hat{\epsilon}) \leq o_p \left( \frac{1}{n} \right). \]

Therefore we can write

\[ S^* (\theta^*, \hat{\epsilon}^*) - S^* (\theta, \hat{\epsilon}) = o_p \left( \frac{1}{n} \right). \]  \hspace{1cm} (3.16)

The expansion of \( S^* (\theta, \hat{\epsilon}) \) about \((\theta^*, \hat{\epsilon}^*)\) analogous to (3.3) is

\[ S^* (\theta, \hat{\epsilon}) = S^* (\theta^*, \hat{\epsilon}^*) + \left( \frac{\partial S^* (\theta^*, \hat{\epsilon}^*)}{\partial \theta} \right) (\theta - \theta^*) \]

\[ + \frac{1}{2} (\theta - \theta^*)' \frac{\partial^2 S^* (\theta^*, \hat{\epsilon}^*)}{\partial \theta \partial \theta} (\theta - \theta^*). \]

So,

\[ S^* (\theta, \hat{\epsilon}) - S^* (\theta^*, \hat{\epsilon}^*) \]

\[ = n^2 \left( \frac{\partial S^* (\theta^*, \hat{\epsilon}^*)}{\partial \theta} \right)' \left( \frac{-1}{n} \right) n^2 \left( \theta - \theta^* \right) + \frac{1}{2} (\theta - \theta^*)' \frac{\partial^2 S^* (\theta^*, \hat{\epsilon}^*)}{\partial \theta \partial \theta} (\theta - \theta^*). \]
Now, \( S^*(\hat{\theta}, \hat{\xi}) - S^*(\hat{\theta}^*, \hat{\xi}^*) \) is of \( o_p(1) \), \( n^{-1/2} \frac{\partial S^*(\hat{\theta}^*, \hat{\xi}^*)}{\partial \theta} \) is of \( o_p(1) \) and \( \frac{\partial^2 S^*(\hat{\theta}^*, \hat{\xi}^*)}{\partial \theta \partial \theta} \) converges to the positive-definite matrix \( G(\rho) \). Also, \( \hat{\theta}^* \) is the m.l.e. in the pseudo-problem, so the directional derivative \\
\( (\frac{\partial S^*(\theta^*, \xi^*)}{\partial \theta})' (\hat{\theta} - \theta^*) > 0 \), and hence \( (\hat{\theta} - \theta^*) \sim o_p(\frac{1}{n^{1/2}}) \). The proof of (3) in Proposition 2 now is complete.

From Proposition 2, we have the following theorem.

Theorem 3.2. Under conditions in Theorem 2.1, we have \\
\[ \sqrt{n} (\hat{\theta} - \theta^{(o)}) \to N(0, \sigma_o^2 G^{-1}(\rho)). \]

§3.3. Asymptotic distribution of \( \hat{\xi} \)

From Theorem 3.2, we know that \( \sqrt{n} (\hat{\theta} - \theta^{(o)}) \) is asymptotically \( N(0, \sigma_o^2 G^{-1}(\rho)) \). In this section, we will take the advantage of the continuity constraints on the mean function to derive the asymptotic distribution of \( \hat{\xi} \).

We know that the mean function is not changed by the appearance of the correlated errors and the continuity constraints also hold in our model. For \( \theta \) belonging to \( \Omega \) and in a neighborhood of \( \theta^{(o)} \), let \( \tau_j \) be the point at which \( f_j(\tau_j; t) \) and \( f_{j+1}(\theta_{j+1}; t) \) intersect. It follows that
\[ f_j(\theta_j; \tau_j) - f_{j+1}(\theta_{j+1}; \tau_j) = 0. \]

Because \( f_j(\theta_j; t) \) and \( f_{j+1}(\theta_{j+1}; t) \) are linear in \( \theta_j \) and \( \theta_{j+1} \), respectively, we have

\[ f_j(\theta_j; \tau_j) = f_j(\theta_j^{(o)}; \tau_j) + \left[ \frac{\partial f_j(\theta_j^{(o)}; \tau_j)}{\partial \theta_j} \right]' (\theta_j - \theta_j^{(o)}) \tag{3.16.1} \]

\[ f_{j+1}(\theta_{j+1}; \tau_j) = f_{j+1}(\theta_{j+1}^{(o)}; \tau_j) + \left[ \frac{\partial f_{j+1}(\theta_{j+1}^{(o)}; \tau_j)}{\partial \theta_{j+1}} \right]' (\theta_{j+1} - \theta_{j+1}^{(o)}). \tag{3.16.2} \]

For \( \tau_j \rightarrow \tau_j^{(o)} \), we can write

\[ f_j(\theta_j^{(o)}; \tau_j) = f_j(\theta_j^{(o)}; \tau_j^{(o)}) + \left[ \frac{\partial}{\partial \tau_j} f_j(\theta_j^{(o)}; \tau_j^{(o)}) + \text{o}(1) \right] (\tau_j - \tau_j^{(o)}). \tag{3.16.3} \]

Similarly,

\[ f_{j+1}(\theta_{j+1}^{(o)}; \tau_j) = f_{j+1}(\theta_{j+1}^{(o)}; \tau_j^{(o)}) + \left[ \frac{\partial}{\partial \tau_j} f_{j+1}(\theta_{j+1}^{(o)}; \tau_j^{(o)}) + \text{o}(1) \right] (\tau_j - \tau_j^{(o)}). \tag{3.16.4} \]

By the continuity constraints of the mean function, for \( \theta_j = \theta_j^{(o)} \), \( \theta_{j+1} = \theta_{j+1}^{(o)} \) and \( \tau_j = \tau_j^{(o)} \), we have

\[ f_j(\theta_j^{(o)}; \tau_j^{(o)}) = f_{j+1}(\theta_{j+1}^{(o)}; \tau_j^{(o)}) \]
Also after some simple algebra, from (3.16.1), (3.16.2), (3.16.3) and (3.16.4) we have

\[
\left[ \frac{\partial}{\partial \tau_j} f_j(\theta_j^{(o)}; \tau_j^{(o)}) - \frac{\partial}{\partial \tau_j} f_{j+1}(\theta_{j+1}; \tau_j^{(o)}) + o(1) \right] (\tau_j - \tau_j^{(o)})
\]

\[
= \left[ \frac{\partial f_{j+1}(\theta_{j+1}, \tau_j)}{\partial \theta_{j+1}} \right]' (\theta_{j+1} - \theta_j^{(o)}) - \left[ \frac{\partial f_j(\theta_j^{(o)}, \tau_j)}{\partial \theta_j} \right]' (\tau_j - \theta_j^{(o)}).
\]

However, as \( \tau_j \) approaches \( \tau_j^{(o)} \), we can write

\[
\frac{\partial f_{j+1}(\theta_{j+1}, \tau_j)}{\partial \theta_{j+1}} = \frac{\partial f_{j+1}(\theta_{j+1}, \tau_j^{(o)})}{\partial \theta_{j+1}} + o(1)
\]

\[
\frac{\partial f_j(\theta_j^{(o)}, \tau_j)}{\partial \theta_j} = \frac{\partial f_j(\theta_j^{(o)}, \tau_j^{(o)})}{\partial \theta_j} + o(1).
\]

Also, from the local behavior of the polynomial we have

\[
\frac{\partial}{\partial \tau_j} f_j(\theta_j^{(o)}; \tau_j^{(o)}) = \theta_j^{(o)} + 2\theta_j^{(o)} \tau_j^{(o)} + \ldots + m\theta_j^{(o)} \tau_j^{(o)} m-1
\]

\[
\frac{\partial}{\partial \tau_j} f_{j+1}(\theta_{j+1}; \tau_j^{(o)}) = \theta_{j+1}^{(o)} + 2\theta_{j+1}^{(o)} \tau_j^{(o)} + \ldots + m\theta_{j+1}^{(o)} m\tau_j^{(o)} m-1
\]

Letting
\[
g^0_j = \theta^{(o)}_{j1} + 2\theta^{(o)}_{j2}\tau_j + \ldots + m\theta^{(o)}_{jm}(\tau_j)^{m-1}
\]

and

\[
h^0_j = \theta^{(o)}_{(j+1)1} + 2\theta^{(o)}_{(j+1)2}\tau_j + \ldots + m\theta^{(o)}_{(j+1)m}(\tau_j)^{m-1},
\]

we have

\[
\left[ g^0_j - h^0_j + o(1) \right] (\tau_j - \tau_j^{(o)})
\]

\[
= \partial f^{(o)}_{j+1}(\theta^{(o)}_{j+1},\tau_j^{(o)}) \partial f^{(o)}_{j}(\theta^{(o)}_{j},\tau_j^{(o)})
\]

\[
= [\frac{\partial f^{(o)}_{j+1}(\theta^{(o)}_{j+1},\tau_j^{(o)})}{\partial \theta_{j+1}} + o(1)](\theta_{j+1} - \theta^{(o)}_{j+1}) - [\frac{\partial f^{(o)}_{j}(\theta^{(o)}_{j},\tau_j^{(o)})}{\partial \theta_{j}} + o(1)](\theta_j - \theta^{(o)}_{j})
\]

(3.17)

Since the first derivative of the mean function is not continuous at \(\tau^{(o)}\), we have \(g^0_j \neq h^0_j\), for all \(j\). If we write (3.17) in matrix form, then we see that

\[
(\tau - \tau^{(o)}) = \begin{pmatrix}
\tau_1 - \tau_1^{(o)} \\
\tau_2 - \tau_2^{(o)} \\
\tau_3 - \tau_3^{(o)} \\
\vdots \\
\tau_{\tau-1} - \tau_{\tau-1}^{(o)}
\end{pmatrix}
\]
\[
\begin{pmatrix}
-a_1 & b_1 & 0 & 0 & \cdots & 0 & 0 \\
0 & -a_2 & b_2 & 0 & \cdots & 0 & 0 \\
0 & 0 & -a_3 & b_3 & \cdots & 0 & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & 0 & 0 & 0 & \cdots & -a_{r-1} & b_{r-1}
\end{pmatrix}
+ o(1)
\]
\]
\[
[M + o(1)] (\theta - \theta^{(o)}),
\]
(3.18)

where

\[
a_j = \frac{1}{\frac{g_j}{h_j} \frac{o}{o}} \left( \frac{\partial f_j(\theta_j^{(o)}, \tau_j^{(o)})}{\partial \theta_j} \right)
\]
and

\[
b_j = \frac{1}{\frac{g_j}{h_j} \frac{o}{o}} \left( \frac{\partial f_{j+1}(\theta_j^{(o)}, \tau_j^{(o)})}{\partial \theta_{j+1}} \right).
\]

Since \( \sqrt{n} (\theta - \theta^{(o)}) \) converges weakly to \( N(0, \sigma_o^2 G^{-1}(\rho)) \), we have

\[
\sqrt{n} \left( \hat{\theta} - \theta^{(o)} \right) = [M + o_p(1)] \sqrt{n} (\theta - \theta^{(o)}) \rightarrow N(0, \sigma_o^2 MG^{-1}(\rho)M').
\]

From the above, we have the following theorem.

**Theorem 3.2.** Under conditions on Theorem 3.1, we have

\[
\sqrt{n} \left( \hat{\theta} - \theta^{(o)} \right) \rightarrow N(0, \sigma_o^2 MG^{-1}(\rho)M') \text{ as } n \rightarrow \infty.
\]
§3.4. Example

In this section, we will illustrate how to compute the asymptotic covariance matrices of $\hat{\theta}$ and $\hat{\ell}$. Suppose that we have the following cubic spline regression model in which the first derivatives of the mean function change at $\tau$:

$$y_{t_i} = \alpha_0 + \alpha_1 t_i + \alpha_2 t_i^2 + \alpha_3 t_i^3 + \alpha_4 (t_i - \tau)_+ + \alpha_5 (t_i - \tau)_+^2 + \alpha_6 (t_i - \tau)_+^3 + u_{t_i},$$

where $u_{t_i}$ are AR(1) errors with $\sigma_o^2 = 1, \rho = \rho^{(o)}$, $t_i$'s are generated from $H(\cdot)$, and $H(\cdot)$ is uniformly distributed over $[0,1]$. This model can be rewritten as two cubic models on separate regions

$$y_{t_i} = \begin{cases} 
\theta_{10} + \theta_{11} t_i + \theta_{12} t_i^2 + \theta_{13} t_i^3 + u_{t_i} & \text{if } t_i \leq \tau \\
\theta_{20} + \theta_{21} t_i + \theta_{22} t_i^2 + \theta_{23} t_i^3 + u_{t_i} & \text{if } t_i > \tau
\end{cases},$$

where

$$\theta_{10} = \alpha_0, \theta_{11} = \alpha_1, \theta_{12} = \alpha_2, \theta_{13} = \alpha_3,$$

and

$$\theta_{20} = \alpha_0 - \alpha_4 \tau + \alpha_5 \tau^2 - \alpha_6 \tau^3, \theta_{21} = \alpha_1 + \alpha_4 - 2\alpha_5 \tau + 3\alpha_6 \tau^2,$$
$$\theta_{22} = \alpha_2 + \alpha_4 - 3\alpha_6, \theta_{23} = \alpha_3 + \alpha_6.$$

Let

$$f_1(\theta_1, t) = \theta_{10} + \theta_{11} t + \theta_{12} t^2 + \theta_{13} t^3.$$
\[ f_2(\theta_2, t) = \theta_{20} + \theta_{21}t + \theta_{22}t^2 + \theta_{23}t^3. \]

If assumptions 1 through 7 of §2.1 are satisfied, then we have 
\[ \sqrt{n} \left( \hat{\mathbf{\theta}} - \theta^{(o)} \right) \rightarrow N(0, \sigma_0^2 \mathbf{G}^{-1}(\rho)) \quad \text{and} \quad \sqrt{n} \left( \hat{\mathbf{\tau}} - \tau^{(o)} \right) \rightarrow N(0, \sigma_0^2 \mathbf{M}^{-1}(\rho)\mathbf{M}') \]
as \( n \rightarrow \infty \). Here \( \mathbf{G}(\rho) \) and \( \mathbf{M} \) can be found as follows:

\[
\frac{\partial \mu(\theta^{(o)}, \tau^{(o)}, \mathbf{t}_n)}{\partial \theta} = \begin{pmatrix} T_1 & 0 \\ 0 & T_2 \end{pmatrix},
\]

where

\[
T_1 = \begin{pmatrix} 1 & 1 & 1 \\ t_1 & t_2 & \ldots & t_{n_1} \\ 2 & 2 & \ldots & 2 \\ t_1 & t_2 & \ldots & t_{n_1} \\ 3 & 3 & \ldots & 3 \\ t_1 & t_2 & \ldots & t_{n_1} \end{pmatrix}
\quad \text{and} \quad T_2 = \begin{pmatrix} 1 & 1 & 1 \\ t_{n_1+1} & t_{n_1+2} & \ldots & t_n \\ 2 & 2 & \ldots & 2 \\ t_{n_1+1} & t_{n_1+2} & \ldots & t_n \\ 3 & 3 & \ldots & 3 \\ t_{n_1+1} & t_{n_1+2} & \ldots & t_n \end{pmatrix}.
\]

Also,

\[
\Sigma^{-1}_n = \begin{pmatrix} \mathbf{A}_1 & 0^{[1,1]} \\ 0^{[1,1]} & \mathbf{A}_2 \end{pmatrix},
\]

where

\[
\mathbf{A}_1 = \begin{pmatrix} 1 & -\rho & \ldots & 0 \\ -\rho & 1+\rho^2 & \ldots & 0 \\ 0 & -\rho & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \ldots & -\rho & 0 \\ 0 & 0 & \ldots & 1+\rho^2 \end{pmatrix}, \quad \mathbf{A}_2 = \begin{pmatrix} 1+\rho^2 & -\rho & \ldots & 0 \\ -\rho & 1+\rho^2 & \ldots & 0 \\ 0 & -\rho & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \ldots & -\rho & 0 \\ 0 & 0 & \ldots & 1 \end{pmatrix},
\]
and $0^{[1.1]}_\rho$ and $0^{(\rho)}_{[1.1]}$ are defined as before. Then,

$$G(\rho) = \lim_{n \to \infty} \frac{1}{n} \left( \frac{\partial \mu(\theta^{(\rho)}, \tau^{(\rho)}; \tau_n)}{\partial \theta} \right) \Sigma_n^{-1} \left( \frac{\partial \mu(\theta^{(\rho)}, \tau^{(\rho)}; \tau_n)}{\partial \theta} \right)'$$

$$= \lim_{n \to \infty} \frac{1}{n} \begin{pmatrix} T_1 & 0 \\ 0 & T_2 \end{pmatrix} \begin{pmatrix} A_1 & 0^{[1.1]}_\rho \\ 0 & A_2 \end{pmatrix} \begin{pmatrix} T_1 & 0 \\ 0 & T_2 \end{pmatrix}$$

$$= \lim_{n \to \infty} \frac{1}{n} \begin{pmatrix} T_1 A_1 T'_1 & T_1 0^{[1.1]}_\rho T'_2 \\ T_2 0^{(\rho)}_{[1.1]} T'_1 & T_2 A_2 T'_2 \end{pmatrix}$$

We will first claim that the off-diagonal submatrices of $G(\rho)$ are $0$. Notice that

$$\lim_{n \to \infty} \frac{1}{n} T_2 0^{(\rho)}_{[1.1]} T'_1$$

$$= \lim_{n \to \infty} \frac{1}{n} \begin{pmatrix} -\rho & -\rho t_n \\ -\rho t_n & -\rho t_n \\ -\rho t_n & -\rho t_n \\ -\rho t_n & -\rho t_n \end{pmatrix}$$

$$= 0.$$
Since $T_1^{0^{[1.1]}} T'_2 = (T_2^{0^{[1.1]}} T'_i)'$, we have \( \lim_{n \to \infty} \frac{1}{n} T_1^{0^{[1.1]}} T'_2 = 0 \), too. Therefore, for \( i \in \{1,2,3,4\} \), \( j \in \{5,6,7,8\} \) or \( i \in \{5,6,7,8\} \), \( j \in \{1,2,3,4\} \), we have \( G_{ij}(\rho) = 0 \).

Next, let us study the diagonal blocks of $G(\rho)$. For \( i,j \in \{1,2,3,4\} \), letting \( t'_{n1i} \) denote the \( i \)th row of $T_1$, the \( (i,j) \)th entry of $G(\rho)$ is

\[
G_{ij}(\rho) = \lim_{n \to \infty} \frac{1}{n} t'_{n1i} A_1 t_{n1j}
\]

\[
= \lim_{n \to \infty} \frac{1}{n} \left[t_1, t_2, \ldots, t_{n1}\right] \begin{pmatrix}
1 & -\rho & \cdots & 0 \\
-\rho & 1+\rho^2 & \cdots & 0 \\
0 & -\rho & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & \cdots & -\rho & 0 \\
0 & \cdots & \cdots & 1+\rho^2 \\
\end{pmatrix}
\begin{pmatrix}
t_1' \\
t_2' \\
\vdots \\
t_{n1}'
\end{pmatrix}
\]

\[
= \lim_{n \to \infty} \frac{1}{n} \left( \sum_{k=1}^{n_1} t_{k}^{i+j-2} + \rho^2 \sum_{k=2}^{n_1-1} t_k^{i+j-2} - \rho \sum_{k=1}^{n_1-1} t_k^{i-1} t_{k+1} - \rho \sum_{k=2}^{n_1} t_k^{i-1} t_{k-1} \right).
\]

Since the $t_i$'s are order statistics of independent standard uniform random variables, then

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n_1-1} t_k^{i-1} t_{k+1} = \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n_1-1} t_k (t_k+\Delta_k)^{j-1}.
\]
Letting $\delta_n = \max(\Delta_1, \Delta_2, \ldots, \Delta_n)$, then using a binomial expansion we can write

$$ (t_k + \Delta_k)^{j-1} = \sum_{l=0}^{j-1} \binom{j-1}{l} t_k^l \Delta_k^{j-1-l} $$

$$ = t_k^{j-1} + \sum_{l=0}^{j-2} \binom{j-1}{l} t_k^l \Delta_k^{j-1-l} $$

$$ \leq t_k^{j-1} + \delta_n \sum_{l=0}^{j-2} \binom{j-1}{l} t_k^l $$

$$ \leq t_k^{j-1} + 2^{j-1} \delta_n. $$

Thus,

$$ \frac{1}{n} \sum_{k=1}^{n-1} t_k^{i-1} (t_k + \Delta_k)^{j-1} \leq \frac{1}{n} \sum_{k=1}^{n-1} t_k^{i-1} (t_k^{j-1} + 2^{j-1} \delta_n) $$

$$ \leq \frac{1}{n} \sum_{k=1}^{n-1} t_k^{i+j-2} + 2^{j-1} \delta_n \frac{n_1}{n}. $$

But we know that $\delta_n \to 0$ and $\frac{n_1}{n} \to \tau(0)$ as $n \to \infty$. Also, $\Delta_k > 0$ for all $k$. So we have
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n-1} t_{k} t_{k+1}^{i-1 j-1} = \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} t_{k}^{i+j-2}.
\]

Since \( E(t_i^m) \) exists, the strong law of large numbers can be applied to \( (1-I(t_i))^m \), where \( m \) is a positive integer, \( I(t_i) = 1 \) if \( t_i > \tau^{(0)} \), \( = 0 \), otherwise. Therefore,

\[
\lim_{n \to \infty} \frac{1}{n} t_n A_1 t_{n}^{i} = \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} t_k^{i+j-2} - 2\rho \sum_{k=1}^{n} t_k^{i+j-2} \]

\[
= (1-\rho)^2 \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} t_k^{i+j-2} \]

\[
= (1-\rho)^2 \int_0^{\tau^{(0)}} t^{i+j-2} dH(t) \]

\[
= (1-\rho)^2 \left( \frac{\tau^{(0)}}{i+j-1} \right) \frac{s_{n}^{i+j-1}}{i+j-1}.
\]

Similarly, for \( i,j \in \{5,6,7,8\} \), if we let \( t_{n2i}^{'} \) be the \( i \)th row of \( T_2 \), then we have

\[
G_{ij}(\rho) = t_{n2i}^{'} A_2 t_{n2j}^{''}
\]
\[
\lim_{n \to \infty} \begin{bmatrix}
 i-1 & i-1 & \cdots & i-1 \\
 t_{n+1} & t_{n+2} & \cdots & t_n
\end{bmatrix}
\begin{pmatrix}
 1+\rho^2 & -\rho & \cdots & 0 \\
 -\rho & 1+\rho^2 & \cdots & 0 \\
 0 & -\rho & \cdots & 0 \\
 \vdots & \vdots & \ddots & \vdots \\
 0 & 0 & \cdots & 1
\end{pmatrix}
\begin{pmatrix}
 j-1 \\
 t_{n+1} \\
 j-1 \\
 t_{n+2} \\
 \vdots \\
 j-1 \\
 t_n
\end{pmatrix}
\]

\[= (1-\rho)^2 \int_{\tau^{(o)}}^{1} t^{i+j-10} \, dH(t) \]

\[= (1-\rho)^2 \frac{i-(\tau^{(o)})^{i+j-9}}{i+j-9} . \]

The elements of \( M \) are:

\[
m_{11} = \frac{1}{D_{l}} \frac{\partial f_{1}(\theta^{(o)}, \tau^{(o)})}{\partial \theta_{10}} \]

\[= \frac{1}{[\theta_{j1}^{(o)} + 2\theta_{j2}^{(o)} \tau_{j}^{(o)} + 3\theta_{j3}^{(o)} (\tau_{j}^{(o)})^{2}][\theta_{(j+1)1}^{(o)} + 2\theta_{(j+1)2}^{(o)} \tau_{j}^{(o)} + 3\theta_{(j+1)3}^{(o)} (\tau_{j}^{(o)})^{2}]} \]

\[
m_{12} = \frac{\tau^{(o)}}{[\theta_{j1}^{(o)} + 2\theta_{j2}^{(o)} \tau_{j}^{(o)} + 3\theta_{j3}^{(o)} (\tau_{j}^{(o)})^{2}][\theta_{(j+1)1}^{(o)} + 2\theta_{(j+1)2}^{(o)} \tau_{j}^{(o)} + 3\theta_{(j+1)3}^{(o)} (\tau_{j}^{(o)})^{2}]} \]

\[
m_{13} = \frac{(\tau^{(o)})^{2}}{[\theta_{j1}^{(o)} + 2\theta_{j2}^{(o)} \tau_{j}^{(o)} + 3\theta_{j3}^{(o)} (\tau_{j}^{(o)})^{2}][\theta_{(j+1)1}^{(o)} + 2\theta_{(j+1)2}^{(o)} \tau_{j}^{(o)} + 3\theta_{(j+1)3}^{(o)} (\tau_{j}^{(o)})^{2}]} \]


\[ m_{14} = \frac{(\tau^{(0)})^3}{[\theta_{j1}^{(0)} + 2\theta_{j2}^{(0)}\tau_j^{(0)} + 3\theta_{j3}^{(0)}(\tau_j^{(0)})^2] - [\theta_{(j+1)1}^{(0)} + 2\theta_{(j+1)2}\tau_j^{(0)} + 3\theta_{(j+1)3}(\tau_j^{(0)})^2]} \],

and

\[ m_{15} = -\frac{1}{D_1}\frac{\partial f_2(\theta^{(0)}, \tau^{(0)})}{\partial \theta_{20}} \]

\[ = \frac{-1}{[\theta_{j1}^{(0)} + 2\theta_{j2}^{(0)}\tau_j^{(0)} + 3\theta_{j3}^{(0)}(\tau_j^{(0)})^2] - [\theta_{(j+1)1}^{(0)} + 2\theta_{(j+1)2}\tau_j^{(0)} + 3\theta_{(j+1)3}(\tau_j^{(0)})^2]} \],

\[ m_{16} = \frac{-\tau^{(0)}}{[\theta_{j1}^{(0)} + 2\theta_{j2}^{(0)}\tau_j^{(0)} + 3\theta_{j3}^{(0)}(\tau_j^{(0)})^2] - [\theta_{(j+1)1}^{(0)} + 2\theta_{(j+1)2}\tau_j^{(0)} + 3\theta_{(j+1)3}(\tau_j^{(0)})^2]} \]

\[ m_{17} = \frac{-\tau^{(0)}(\tau^{(0)})^2}{[\theta_{j1}^{(0)} + 2\theta_{j2}^{(0)}\tau_j^{(0)} + 3\theta_{j3}^{(0)}(\tau_j^{(0)})^2] - [\theta_{(j+1)1}^{(0)} + 2\theta_{(j+1)2}\tau_j^{(0)} + 3\theta_{(j+1)3}(\tau_j^{(0)})^2]} \]

\[ m_{18} = \frac{-(\tau^{(0)})^3}{[\theta_{j1}^{(0)} + 2\theta_{j2}^{(0)}\tau_j^{(0)} + 3\theta_{j3}^{(0)}(\tau_j^{(0)})^2] - [\theta_{(j+1)1}^{(0)} + 2\theta_{(j+1)2}\tau_j^{(0)} + 3\theta_{(j+1)3}(\tau_j^{(0)})^2]} \].

§3.5. Asymptotic distribution of $\delta^2$

It is well known that for a general linear model with i.i.d. normal errors, the m.l.e. of $\sigma_0^2$ is a consistent estimator, and after being suitably standardized, say $\sqrt{n}(\delta^2 - \sigma_0^2)$, it asymptotically follows a normal distribution. This observation leads us to conjecture that the asymptotic distribution of $\sqrt{n}(\delta^2 - \sigma_0^2)$ in our model is also normal. Later we will show that the random variable
\[
\sqrt{n} \left( \delta^2 - \sigma^2 \right) \text{ asymptotically follows a normal distribution with mean 0 and variance } 2\sigma^4.
\]

In our model, the m.l.e.'s \( \hat{\theta}, \hat{\tau} \) and \( \delta^2 \) are the values of \( \theta, \tau \) and \( \sigma^2 \) that maximize the log likelihood function in (2.8):

\[
L(\theta, \tau; \sigma^2, y) = \log L(\theta, \tau; \sigma^2, y) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2} \log |\Sigma_n| + \frac{1}{2\sigma^2} (y - \mu(\theta, \tau; t_n))' \Sigma_n^{-1} (y - \mu(\theta, \tau; t_n)).
\]

(2.8)

In the general linear model, the mean function \( \mu(\theta, \tau; t_n) \) does not involve \( \tau \) and is linear in \( \theta \). The m.l.e. \( \delta^2 \) is obtained by plugging the m.l.e.'s of \( \theta \) into the normal equation of \( \sigma^2 \) (this is \( \frac{\partial \log L(\theta|y)}{\partial \sigma^2} = 0 \)) and solving for \( \sigma^2 \). The resulting \( \delta^2 \) has the following form:

\[
\delta^2 = \frac{1}{n} (y - \hat{\mu})' \Sigma_n^{-1} (y - \hat{\mu}),
\]

where \( \hat{\mu} \) is the m.l.e. of the mean function \( \mu(\theta; t_n) \). Similarly, in our model the m.l.e. of \( \sigma^2 \) is obtained by inserting the m.l.e.'s of \( \theta^{(o)} \) and \( \tau^{(o)} \) into \( \mu(\theta, \tau; t_n) \). We will use the results derived in §3.3 to find the asymptotic distribution of \( \delta^2 \).

From (3.3) and (3.9), we have
\[
\$\left(\theta^*\right) = \$\left(\theta^{(o)}\right) + \left( \frac{\partial \$\left(\theta^{(o)}\right)}{\partial \theta} \right)'(\theta^* - \theta^{(o)}) + \frac{1}{2} (\theta^* - \theta^{(o)})' \frac{\partial^2 \$\left(\theta^{(o)}\right)}{\partial \theta \partial \theta} (\theta^* - \theta^{(o)})
\]

and

\[- \frac{\partial \$\left(\theta^{(o)}\right)}{\partial \theta} = \left( \frac{\partial^2 \$\left(\theta^{(o)}\right)}{\partial \theta \partial \theta} \right)'(\theta^* - \theta^{(o)}).\]

Thus,

\[
\$\left(\theta^*\right) - \$\left(\theta^{(o)}\right) = -(\theta^* - \theta^{(o)})' \frac{\partial^2 \$\left(\theta^{(o)}\right)}{\partial \theta \partial \theta} (\theta^* - \theta^{(o)}) + \frac{1}{2} (\theta^* - \theta^{(o)})' \frac{\partial^2 \$\left(\theta^{(o)}\right)}{\partial \theta \partial \theta} (\theta^* - \theta^{(o)})
\]

\[
= -\frac{1}{2} (\theta^* - \theta^{(o)})' \frac{\partial^2 \$\left(\theta^{(o)}\right)}{\partial \theta \partial \theta} (\theta^* - \theta^{(o)})
\]

\[
= - (\theta^* - \theta^{(o)})' [G(\rho) + o(1)] (\theta^* - \theta^{(o)}). \quad (3.19)
\]

Multiplying both sides of (3.19) by \(n\), we obtain

\[
n\left[ \$\left(\theta^{(o)}\right) - \$\left(\theta^*\right) \right] = \sqrt{n} (\theta^* - \theta^{(o)})' [G(\rho) + o(1)] \sqrt{n} (\theta^* - \theta^{(o)}).
\]

Now \(\sqrt{n} (\theta^* - \theta^{(o)}) \rightarrow N(0, \sigma_0^2 G^{-1}(\rho))\), so

\[n[\$\left(\theta^{(o)}\right) - \$\left(\theta^*\right)]\]

asymptotically follows \(z'G(\rho)z\), where \(z \sim N(0, \sigma_0^2 G^{-1}(\rho))\). But
\[ z' G(\rho) z / \sigma_o^2 \sim \chi^2 (r(m+1)) \], so we have \( n[S(\theta^{(o)}) - \hat{\theta}^*] / \sigma_o^2 \) follows a \( \chi^2 (r(m+1)) \) distribution asymptotically.

The next step is to show that
\[
n[S(\theta^{(o)}) - \hat{\theta}^*] - n[S(\theta^{(o)}) - S(\hat{\theta}, \hat{\tau})] = o_p(1).
\]

From (3.15) and (3.16), we have as \( n \to \infty \),
\[
n\{ [S(\theta^{(o)}) - \hat{\theta}^*] - [S(\theta^{(o)}), \tau^{(o)}] - S(\hat{\theta}, \hat{\tau}) \}
\]
\[= n[S(\theta^{(o)}) - \hat{\theta}^*] - n\{ [S(\theta^{(o)}) + \frac{1}{n} u'u^-1 u^{**} + o_p(\frac{1}{n})]
\]
\[- [S(\hat{\theta}^*) + \frac{1}{n} u'u^-1 u^{**} + o_p(\frac{1}{n})] \}
\]
\[= n \cdot o_p(\frac{1}{n})
\]
\[= o_p(1).
\]

Therefore, the two random variables
\[ n[S(\theta^{(o)}) - \hat{\theta}^*] \quad \text{and} \quad n[S(\theta^{(o)}, \tau^{(o)}) - S(\hat{\theta}, \hat{\tau})] \]

have the same asymptotic distribution. We have the following result:

**Lemma 2.** Under the assumptions in Theorem 3.1, we have
\[
\frac{n[S(\theta^{(o)}, \tau^{(o)}) - S(\hat{\theta}, \hat{\tau})]}{\sigma_o^2} \xrightarrow{n \to \infty} \chi^2_{(r(m+1))}\]

Notice that \(E[S(\theta^{(o)}, \tau^{(o)})] = \sigma_o^2\) and \(S(\theta^{(o)}, \tau^{(o)})\) converge to \(\sigma_o^2\) almost everywhere. In practice, if \(n\) is large, by Slutsky's theorem we can replace \(S(\theta^{(o)}, \tau^{(o)})\) by \(\sigma_o^2\) and use the above random variable to perform inference for \(\sigma_o^2\). But this substitution leads to questions about the magnitude of the error as \(n\) diverges to infinity. We observe that

\[
\frac{n[S(\theta^{(o)}, \tau^{(o)}) - S(\hat{\theta}, \hat{\tau})]}{\sigma_o^2} = \frac{n[S(\theta^{(o)}, \tau^{(o)}) - \sigma_o^2]}{\sigma_o^2} + \frac{n[\sigma_o^2 - S(\hat{\theta}, \hat{\tau})]}{\sigma_o^2}.
\]

But

\[
\frac{n[S(\theta^{(o)}, \tau^{(o)}) - \sigma_o^2]}{\sigma_o^2} = \frac{nS(\theta^{(o)}, \tau^{(o)})}{\sigma_o^2} - n
\]

\[
= \chi^2_{(n)} - n.
\]
here \( \chi^2_{(n)} \) denotes a random variable following chi-square distribution with \( n \) degrees of freedom. By the Central Limit Theorem, \( \left[ \frac{\chi^2_{(n)} - n}{\sqrt{2n}} \right] \xrightarrow{\text{as } n \to \infty} N(0,1) \). So we have \( \chi^2_{(n)} - n = 0_p(\sqrt{n}) \), and hence

\[
\frac{n [S(\theta^{(o)}, \tau^{(o)}) - S(\hat{\theta}, \hat{\tau})]}{\sigma_o^2} = \frac{n [S(\theta^{(o)}, \tau^{(o)}) - \sigma_o^2]}{\sigma_o^2} + 0_p(\sqrt{n}).
\]

So the error is of \( 0_p(\sqrt{n}) \), which means that as \( n \) gets large, the replacement will not perform well.

The above observation forces us to consider another random variable which we can use to make inference about \( \sigma_o^2 \). Consider the following partition

\[
\frac{nS(\theta^{(o)}, \tau^{(o)})}{\sigma_o^2} = \frac{n [S(\theta^{(o)}, \tau^{(o)}) - S(\hat{\theta}, \hat{\tau})]}{\sigma_o^2} + \frac{nS(\hat{\theta}, \hat{\tau})}{\sigma_o^2}.
\]

Notice that \( nS(\theta^{(o)}, \tau^{(o)}) = \sum_{i=1}^n e_i^2 = ||e||^2, \frac{||e||^2}{\sigma_o^2} \sim \chi^2_{(n)} \). Also, \( S(\hat{\theta}, \hat{\tau}) \) is an estimator of \( S(\theta^{(o)}, \tau^{(o)}) \), so we can write \( nS(\hat{\theta}, \hat{\tau}) = ||\hat{\tau}||^2 \). Since m.l.e. and l.s.e. coincide in our model, so we can consider \( \hat{\tau} \) to be the projection of \( e \) (see figure 3.3).
Now the inner product $<\hat{e}, e - \hat{e}> = 0$, $\hat{e} \rightarrow e$ in probability, and $e$ is a normal vector, so $\hat{e}$ and $e - \hat{e}$ are asymptotically independent. But

$$\frac{||e - \hat{e}||^2}{\sigma_o^2} = \frac{||e||^2}{\sigma_o^2} - \frac{||\hat{e}||^2}{\sigma_o^2}$$

$$= \frac{nS(\theta^o, \tau^{(o)})}{\sigma_{(o)}^2} - \frac{nS(\hat{\theta}, \hat{\tau})}{\sigma_o^2}$$

So we have $[nS(\theta^o, \tau^{(o)})/\sigma_o^2 - nS(\hat{\theta}, \hat{\tau})/\sigma_o^2]$ and $nS(\theta, \tau)/\sigma_o^2$ are asymptotically independent. Now we know $nS(\theta^o, \tau^{(o)})/\sigma_o^2$ follows a chi-square distribution with $n$ degrees of freedom, and $n[S(\theta^o, \tau^{(o)}) - S(\hat{\theta}, \hat{\tau})]/\sigma_o^2$ asymptotically follows a chi-square
distribution with \( r(m+1) \) degrees of freedom. Therefore \( nS(\hat{\theta}, \hat{c})/\sigma_o^2 \) asymptotically follows a chi-square distribution with \( n-r(m+1) \) degrees of freedom, or equivalently

\[
\left\{ \frac{nS(\hat{\theta}, \hat{c})}{\sigma_o^2} - \left[ n-r(m+1) \right] \right\}/\sqrt{2[n-r(m+1)]}
\]

asymptotically follows the standard normal distribution. We state this result in the following lemma.

**Lemma 3.** Under assumptions in Theorem 3.1, we have

\[
\left\{ \frac{nS(\hat{\theta}, \hat{c})}{\sigma_o^2} - \left[ n-r(m+1) \right] \right\}/\sqrt{2[n-r(m+1)]}
\]

asymptotically follows the standard normal distribution.

It is interesting that the above result is similar to that for a general linear model. The degrees of freedom lost is the number of regression coefficients we want to estimate, which is exactly the result obtained in the general linear model case. Also, if \( r(m+1) \) is not too large compared to \( n \), we can simplify the above theorem further. Notice that \( S(\hat{\theta}, \hat{c}) = \delta^2 \), so we can write

\[
\left\{ \frac{nS(\hat{\theta}, \hat{c})}{\sigma_o^2} - \left[ n-r(m+1) \right] \right\}/\sqrt{2[n-r(m+1)]} = \frac{n\delta^2 - n\sigma_o^2 + r(m+1)\sigma_o^2}{\sqrt{2[n-r(m+1)]}}
\]
\[
\sqrt{n}(\delta^2 - \sigma_o^2) = \frac{\sqrt{n}(\delta^2 - \sigma_o^2)}{\sqrt{2}\sigma_o} + o(1).
\]

Therefore, if \(n\) is large compared to \(r(m+1)\), we have \(\sqrt{n}(\delta^2 - \sigma_o^2) \rightarrow N(0,2\sigma_o^4)\), asymptotically. We state this result in the following theorem.

**Theorem 3.3.** Under the assumptions in Theorem 3.1, we have

\[\sqrt{n}(\delta^2 - \sigma_o^2) \text{ asymptotically follows a } N(0,2\sigma_o^4)\text{ distribution.}\]
Chapter IV

Asymptotic Distributions : \( \rho \) unknown

In Chapter 3, we discussed the asymptotic behaviors of \( \hat{\theta} \), \( \hat{\xi} \) and \( \hat{\sigma}^2 \) and showed that the asymptotic distributions of \( \sqrt{n} ( \hat{\theta} - \theta^{(o)} ) \), \( \sqrt{n} ( \hat{\xi} - \tau^{(o)} ) \) and \( \sqrt{n} ( \hat{\sigma}^2 - \sigma^{(o)}^2 ) \) are normal when \( \rho^{(o)} \) is known. In this chapter, we will derive the asymptotic distribution of \( \hat{\rho} \) and the asymptotic distributions of \( \hat{\theta} \), \( \hat{\xi} \) and \( \hat{\sigma}^2 \) in the case that \( \rho^{(o)} \) is unknown. The asymptotic distributions of \( \hat{\theta} \) and \( \hat{\xi} \) will be investigated first in §4.1. In §4.2, we will derive the asymptotic distribution of \( \hat{\rho} \). The asymptotic distribution of \( \hat{\sigma}^2 \) will be discussed in §4.3.

§4.1 Asymptotic distributions of \( \hat{\theta} \) and \( \hat{\xi} \)

Asymptotic distributions of \( \hat{\theta} \) and \( \hat{\xi} \) in the case that \( \rho^{(o)} \) is unknown are derived in this section. In §4.1.1, we show that the asymptotic distribution of \( \sqrt{n} ( \hat{\theta} - \theta^{(o)} ) \) is normal. The asymptotic distribution of \( \sqrt{n} ( \hat{\xi} - \tau^{(o)} ) \) is investigated in §4.1.2.

§4.1.1 Asymptotic distribution of \( \hat{\theta} \)

Remember that we have the following model

\[
y_{ni} = \sum_{j=1}^{r} f_{ij}(\theta_j, t_{nj})I_{ij}(t_{nj}) + u_{ni},
\]
\[ u_{ni} = \rho u_{n(i-1)} + e_{ni}, \]

\[ I_j(t_{ni}) = \begin{cases} 
1 & \text{if } \tau_{j-1} \leq t_{ni} \leq \tau_j, \\
0 & \text{otherwise},
\end{cases} \]

and

\[ f_j(\theta_j, \tau_j) = f_{j+1}(\theta_{j+1}, \tau_j) \quad j=1,2,...,r-1, \]

where

\[ f_j(\theta_j, t_{ni}) = \sum_{k=0}^{m} \theta_{jk} t_{ni}^k \quad j=1,2,...,r. \quad (4.1) \]

From Chapter 3, we know that if \( \rho \) is known, say \( \rho^{(o)} \), then under suitable conditions,

\[ \sqrt{n} \left( \hat{\theta} - \theta^{(o)} \right) \rightarrow N(0, \sigma_0^2 G^{-1}(\rho^{(o)})). \]

Since \( \hat{\theta} \) involves \( \rho^{(o)} \), we can write it as \( \hat{\theta}(\rho^{(o)}) \). Then

\[ \sqrt{n} \left( \hat{\theta}(\rho^{(o)}) - \theta^{(o)} \right) \rightarrow N(0, \sigma_0^2 G^{-1}(\rho^{(o)})). \]

Now, we do not know the value of \( \rho^{(o)} \), but still want to find the asymptotic distribution of \( \hat{\theta} \), which involves \( \hat{\rho}, \hat{\theta} \) being an estimator of \( \rho^{(o)} \) and function of \( y_{n1}, y_{n2}, ..., y_{nn} \). Borrowing from classical theory that the sample mean and sample variance of i.i.d. normal random variables are independent, with \( \rho^{(o)} \) being in the covariance structure not in the mean function, it seems that it is natural to conjecture that \( \hat{\theta}(\hat{\rho}) \) and \( \hat{\rho} \) are asymptotically independent and hence whether or not \( \rho^{(o)} \) is known will not affect the asymptotic
distribution of \( \hat{\theta}(\hat{\theta}) \). That is, the asymptotic distributions of \( \hat{\theta}(\hat{\theta}) \) and of \( \hat{\theta}(\rho^{(0)}) \) will be the same no matter whether we know the value of \( \rho^{(0)} \) or not. As we show later, if the estimator \( \hat{\theta} \) is a \( \sqrt{n} \)-consistent estimator, then \( \hat{\theta}(\hat{\theta}) \) will have the same asymptotic distribution as \( \hat{\theta}(\rho^{(0)}) \).

We will use the following strategy to derive the asymptotic distribution of \( \hat{\theta}(\hat{\theta}) \):

1. For the case that \( \rho^{(0)} \) is unknown, we form the pseudo-problem and investigate the asymptotic behavior of \( \sqrt{n} (\hat{\theta}^*(\hat{\theta}) - \theta^{(0)}) \).

2. We claim that \( \sqrt{n} (\hat{\theta}^*(\hat{\theta}) - \theta^{(0)}) \) and \( \sqrt{n} (\theta^*(\rho^{(0)}) - \theta^{(0)}) \) have the same asymptotic distribution by showing that \( \sqrt{n} (\hat{\theta}^*(\hat{\theta}) - \theta^*(\rho^{(0)})) \rightarrow 0 \) in probability as \( n \) diverges to infinity.

3. We show \( \sqrt{n} (\theta(\hat{\theta}) - \theta^*(\hat{\theta})) \rightarrow 0 \) in probability and then claim that \( \sqrt{n} (\hat{\theta}(\hat{\theta}) - \theta^{(0)}) \) and \( \sqrt{n} (\theta(\rho^{(0)}) - \theta^{(0)}) \) have the same asymptotic distribution.

In this chapter, we will use \( S^*(\theta) \) to denote \( S^*(\theta, \rho^{(0)}) \) and \( S(\theta, \tau) \) to denote \( S(\theta, \tau; \rho^{(0)}) \). First let us write \( \sqrt{n} (\theta^*(\rho^{(0)}) - \theta^{(0)}) \) explicitly. From Chapter 3, we know that if \( \rho^{(0)} \) is known and \( \theta^{(0)} \) is an interior point of \( \Omega \) then with large probability as \( n \rightarrow \infty \), \( \theta^*(\rho^{(0)}) \) is in the neighborhood of \( \theta^{(0)} \) and is the minimun point of \( S^*(\theta, \rho^{(0)}) \). Let \( \sqrt{n} \ W_n^{*}(\rho^{(0)}) = \sqrt{n} (\theta^*(\rho^{(0)}) - \theta^{(0)}) \). Then from (3.9),

\[
\sqrt{n} \ W_n^{*}(\rho^{(0)})
\]
\[
\begin{align*}
&= \sqrt{n} \left( \hat{\theta}^* (\rho^{(o)}) - \theta^{(o)} \right) \\
&= n^{\frac{-1}{2}} \left[ \mathbf{G}(\rho^{(o)}) + o(1) \right]^{-1} \left[ \frac{\partial \mu^* (\theta^{(o)}, t_n)}{\partial \theta} \Psi_n + o(1) \right] e. \quad (4.2)
\end{align*}
\]

Now, let \( \hat{\rho} \) be a \( \sqrt{n} \)-consistent estimator of the unknown parameter \( \rho^{(o)} \). Replacing \( \rho^{(o)} \) in \( \hat{\theta}^* (\rho^{(o)}) \) by \( \hat{\rho} \), then from (4.2),

\[
\sqrt{n} \mathbf{W}_n(\hat{\rho}) = \sqrt{n} \left( \hat{\theta}^* (\hat{\rho}) - \theta^{(o)} \right) \\
= n^{\frac{-1}{2}} \left[ \mathbf{G}(\hat{\rho}) + o(1) \right]^{-1} \left[ \frac{\partial \mu^* (\theta^{(o)}, t_n)}{\partial \theta} \Psi_n + o(1) \right] e,
\]

where \( \Psi_n \) is a matrix such that the value \( \rho^{(o)} \) in \( \Psi_n \) is replaced by \( \hat{\rho} \).

Since each element \( \sigma_{ij} \) in \( \Sigma_n \) has the form

\[
\sigma_{ij} = \frac{2}{1 - (\rho^{(o)})^2} (\rho^{(o)})^{k-j} \left| \frac{\partial \mu^* (\theta^{(o)}, t_n)}{\partial \theta} \right| e, \quad i,j=1,2,\ldots,n,
\]

then \( \mathbf{G}(\rho^{(o)}) \) and \( \frac{\partial \mu^* (\theta^{(o)}, t_n)}{\partial \theta} \Psi_n \) are continuously differentiable functions of \( \rho^{(o)} \) for \(-1 < \rho^{(o)} < 1 \). Therefore \( \sqrt{n} \mathbf{W}_n(\rho) \), considering \( \rho \) as a mathematical variable, is also a continuously differentiable function of \( \rho \). Then the Mean Value Theorem yields

\[
(\sqrt{n} \mathbf{W}_n^{*} (\hat{\rho}) )_{(i)} = (\sqrt{n} \mathbf{W}_n^{*} (\rho^{(o)}) )_{(i)} + (\hat{\rho} - \rho^{(o)})(\sqrt{n} \frac{\partial \mathbf{W}_n^{*}}{\partial \rho} )_{(i)},
\]
where \( (\sqrt{n} W_n^*(\hat{\rho}))_{(i)} \) is the \( i \)-th coordinate of \( \sqrt{n} W_n^*(\hat{\rho}) \), \( (\sqrt{n} W_n^*(\rho^{(o)}))_{(i)} \) is the \( i \)-th coordinate of \( \sqrt{n} W_n^*(\rho^{(o)}) \), and \( \left( \frac{\partial W_n^*}{\partial \rho} \right)_{(i)} \) is the derivative of the \( i \)-th coordinate of \( \sqrt{n} \frac{\partial W_n^*}{\partial \rho} \) evaluated at \( \rho = \rho^{(i)} \),

for some \( \rho^{(i)} \) between \( \rho^{(o)} \) and \( \hat{\rho} \). Since

\[
(\sqrt{n} W_n^*(\hat{\rho}))_{(i)} - (\sqrt{n} W_n^*(\rho^{(o)}))_{(i)} = (\hat{\rho} - \rho^{(o)}) \left( \sqrt{n} \frac{\partial W_n^*}{\partial \rho} \right)_{(i)},
\]

(4.3)

if we can show

\[
\sqrt{n} (\hat{\rho} - \rho^{(o)}) \left( \frac{\partial W_n^*}{\partial \rho} \right)_{(i)} \to 0 \text{ in probability},
\]

for all \( i \), then we can say that \( \sqrt{n} W_n^*(\hat{\rho}) \) and \( \sqrt{n} W_n^*(\rho^{(o)}) \) have the same asymptotic distribution.

Let \( \Psi^*_n(i) \) be a matrix that the value \( \rho^{(o)} \) in \( \Psi^*_n \) is replaced by \( \rho^{(i)} \). Now
\[ \frac{\partial W_n^*(\rho_{(i)})}{\partial \rho}_{(i)} \]

\[ = \frac{\partial}{\partial \rho} \left\{ \frac{1}{n} \left[ G(\rho) + o(1) \right]^{-1} \left[ \frac{\partial \mu^*(\theta^{(o)}, t_n)}{\partial \theta} \Psi_n + o(1) \right] e \right\}_{(i)} |_{\rho = \rho_{(i)}} \]

\[ = \left\{ \frac{1}{n} \frac{\partial}{\partial \rho} \left[ G(\rho) + o(1) \right]^{-1}_{(i)} |_{\rho = \rho_{(i)}}^* \left[ \frac{\partial \mu^*(\theta^{(o)}, t_n)}{\partial \theta} \Psi_n^{(i)} + o(1) \right] e \right\} \]

\[ + \frac{1}{n} \left[ G(\rho_{(i)}) + o(1) \right]^{-1}_{(i)} \frac{\partial}{\partial \rho} \left\{ \left[ \frac{\partial \mu^*(\theta^{(o)}, t_n)}{\partial \theta} \Psi_n + o(1) \right] e \right\}_{(i)} |_{\rho = \rho_{(i)}}^* , \quad (4.4) \]

and

\[ \frac{1}{n} \left[ \frac{\partial \mu^*(\theta^{(o)}, t_n)}{\partial \theta} \Psi_n^{(i)} + o(1) \right] e = \frac{1}{n} \left[ \frac{\partial \mu^*(\theta^{(o)}, t_n)}{\partial \theta} \Psi_n + o_p(1) \right] e \to 0 \text{ a.e.} \]

as \( n \) diverges to infinity. So, if we can show that \( \frac{\partial}{\partial \rho} \left[ G(\rho_{(i)}) + o(1) \right]^{-1} \)

is bounded in probability, that is, of \( o_p(1) \), then the first term in (4.4) is of \( o_p(1) \).

First, we notice that \( G(\rho) \) is continuous differentiable with respect to \( \rho \) since every element in \( G(\rho) \) is continuous differentiable with respect to \( \rho \). Also, the inverse of \( G(\rho) \) exists for all relevant \( \rho \). Therefore, \( \left[ G(\rho) + o(1) \right]^{-1} \) is continuously differentiable with respect
differentiable with respect to \( \rho \). Now \( \hat{\rho} \) being \( \sqrt{n} \)-consistent implies that \( \hat{\rho}^* \), and hence \( \rho_{(i)}, i = 1, 2, ..., r(m+1) \), converges to \( \rho^{(o)} \) in probability. Therefore by the continuity of \( \frac{\partial}{\partial \rho} G(\rho)^{-1} \), we have

\[
\frac{\partial}{\partial \rho} G(\rho_{(i)})^{-1} = \frac{\partial}{\partial \rho} G(\rho^{(o)})^{-1} + o_p(1), \quad i=1,2,...,r(m+1).
\]

For fixed \( \rho^{(o)} \), every element of \( \frac{\partial}{\partial \rho} G(\rho^{(o)})^{-1} \) is bounded by some positive constant \( k_i \). Therefore we have that \( \frac{\partial}{\partial \rho} G(\rho_{(i)})^{-1} \) is bounded by \( k_i \), \( j_p \) is an \( l \times p \) row vector of 1's, \( p = r(m+1), i=1,2,...,r(m+1) \).

Thus we have \( \frac{\partial}{\partial \rho} G(\rho_{(i)})^{-1} = 0_p(1), i = 1, 2, ..., r(m+1) \).

The next step is to prove that the second part of (4.4) is \( o_p(1) \). It is easy to see that \( [G(\rho_{(i)}) + o(1)]^{-1} \) is bounded in probability. For the derivative part of the second term in (4.4), we see that

\[
\frac{\partial}{\partial \rho} \left[ \frac{\partial \mu^* (\theta^{(o)}, t_n) \Psi_n}{\partial \theta} \right]
\]

is continuous in \( \rho \). So the consistency of \( \hat{\rho} \), and hence of \( \rho_{(i)} \), gives

\[
\frac{\partial}{\partial \rho} \left\{ \left[ \frac{\partial \mu^* (\theta^{(o)}, t_n)}{\partial \theta} \Psi_n + o(1) \right] e \right\}_{\rho=\rho^*} \]
\[
\frac{\partial}{\partial \rho} \left\{ \frac{\partial \mu^*(\theta^{(o)}, t_n)}{\partial \theta} \Psi_n \mathbf{e} \right\} \bigg|_{\rho=\rho^{(i)}} = \left[ \frac{\partial}{\partial \rho} \frac{\partial \mu^*(\theta^{(o)}, t_n)}{\partial \theta} \Psi_n \right] \mathbf{e} = o_p(1).
\]

Therefore if the limit of each of the diagonal elements of

\[
\frac{1}{n} \left[ \frac{\partial}{\partial \rho} \frac{\partial \mu^*(\theta^{(o)}, t_n)}{\partial \theta} \Psi_n \right] \left[ \frac{\partial}{\partial \rho} \frac{\partial \mu^*(\theta^{(o)}, t_n)}{\partial \theta} \Psi_n \right],
\]

exists, then by Theorem 3.1, \( \frac{1}{n} \left[ \frac{\partial}{\partial \rho} \frac{\partial \mu^*(\theta^{(o)}, t_n)}{\partial \theta} \Psi_n \right] \mathbf{e} \) is of \( o_p(1) \).

Therefore, we have \( \left( \frac{\partial W_n^*(\rho^{(i)})}{\partial \rho} \right)_{(i)} = o_p(1) \).

Now, from (4.3), we have

\[
\sqrt{n} \left( \frac{\partial W_n^*(\rho^{(i)})}{\partial \rho} \right)_{(i)} = o_p(1).
\]

By assumption, \( \hat{\rho} \) is \( \sqrt{n} \)-consistent, so \( \sqrt{n} (\hat{\rho} - \rho^{(o)}) = o_p(1) \). Thus

\[
(\sqrt{n} \ W_n^*(\hat{\rho}))_{(i)} - (\sqrt{n} \ W_n^*(\rho^{(o)}))_{(i)} = 0_p(1) \times o_p(1) = o_p(1), \ \forall \ i.
\]
Therefore, asymptotically, $\sqrt{n} \mathbf{W}_n^*(\hat{\theta})$ and $\sqrt{n} \mathbf{W}_n^*(\rho^{(o)})$ have the same distribution.

The third step is to show that $\sqrt{n} (\hat{\theta}(\hat{\rho}) - \hat{\theta}^*(\hat{\rho})) \to 0$ in probability. From Chapter 3, we know that for $\rho = \rho^{(o)}$,

$$\hat{\theta}^*(\rho^{(o)}) - \hat{\theta}(\rho^{(o)}) = o_p(n^{-\frac{1}{2}}).$$

Consider $\rho$ as a mathematical variable. Letting $\{ a_n \}$, $\mathbf{u}_n$, and $L_j(n)$ be as before, then for fixed $\rho = \hat{\rho}$, since $-1 < \hat{\rho} < 1$ with probability 1 we have that $(\hat{\theta}(\hat{\rho}), \hat{\xi}(\hat{\rho}))$ and $\hat{\theta}^*(\hat{\rho})$ both lie in $\mathbf{u}_n$ with large probability as $n \to \infty$. Proceeding as in Chapter 3, except replacing $\rho^{(o)}$ by $\hat{\rho}$ in $\theta$, $\xi$ and $\theta^*$, we have

$$S^*(\hat{\theta}(\hat{\rho})) - S^*(\hat{\theta}^*(\hat{\rho})) = o_p(n^{-1}). \quad (4.6)$$

Expanding $S^*(\hat{\theta}(\hat{\rho}))$ with respect to $\theta^*(\hat{\rho})$, we have

$$S^*(\hat{\theta}(\hat{\rho})) = S^*(\hat{\theta}^*(\hat{\rho})) + (\hat{\theta}(\hat{\rho}) - \hat{\theta}^*(\hat{\rho})) \frac{\partial S^*(\hat{\theta}^*(\hat{\rho}))}{\partial \theta}$$

$$+ \frac{1}{2} (\hat{\theta}(\hat{\rho}) - \hat{\theta}^*(\hat{\rho})) \frac{\partial^2 S^*(\hat{\theta}^*(\hat{\rho}))}{\partial \theta \partial \theta} (\hat{\theta}(\hat{\rho}) - \hat{\theta}^*(\hat{\rho})).$$

But $\frac{\partial S^*(\hat{\theta}^*(\hat{\rho}))}{\partial \theta} = o_p(n^{-\frac{1}{2}})$, $\frac{\partial^2 S^*(\hat{\theta}^*(\hat{\rho}))}{\partial \theta \partial \theta} = \mathbf{G}(\rho^{(o)}) + o_p(1)$, so from (4.6), we have $\hat{\theta}(\hat{\rho}) - \hat{\theta}^*(\hat{\rho}) = o_p(n^{-\frac{1}{2}})$. Since $\hat{\theta}(\hat{\rho})$ and $\hat{\theta}^*(\hat{\rho})$ have the same $\hat{\rho}$, then $\hat{\theta}(\hat{\rho}) - \hat{\theta}^*(\hat{\rho}) = o_p(n^{-\frac{1}{2}})$ is true for all $-1 < \hat{\rho} < 1$. 
In conclusion, we know that \( \sqrt{n} (\hat{\theta}(\rho^{(o)}) - \theta^{(o)}) = \sqrt{n} W_n^{*}(\rho^{(o)}) \) converges weakly to \( N(0, \sigma_0^{-2} G^{-1}(\rho^{(o)})) \); \( \sqrt{n} (\hat{\theta}(\rho^{(o)}) - \hat{\theta}^{*}(\rho^{(o)})) \rightarrow 0 \) in probability. Also, if \( \hat{\theta} \) is \( \sqrt{n} \)-consistent and the limits of diagonal elements of the matrix in (4.5) exist as \( n \rightarrow \infty \), then

\[
\sqrt{n} \left( W_n^{*}(\hat{\theta}) - W_n^{*}(\rho^{(o)}) \right) = \sqrt{n} \left( \hat{\theta}^{*}(\hat{\theta}) - \hat{\theta}^{*}(\rho^{(o)}) \right) = o_p(1)
\]

and

\[
\sqrt{n} \left( \hat{\theta}(\hat{\theta}) - \hat{\theta}^{*}(\hat{\theta}) \right) = o_p(1).
\]

Hence we know that \( \sqrt{n} (\hat{\theta}(\hat{\theta}) - \theta^{(o)}) \) and \( \sqrt{n} (\hat{\theta}(\rho^{(o)}) - \theta^{(o)}) \) have the same asymptotic distribution.

Before stating the above result as a theorem, let us examine the matrix in (4.5) further. The expression in (4.5) can be written as

\[
\frac{1}{n} \left[ \frac{\partial}{\partial \rho} \frac{\partial \mu^{*}(\theta^{(o)}, t_n)}{\partial \theta} \Psi \right] \left[ \frac{\partial}{\partial \rho} \frac{\partial \mu^{*}(\theta^{(o)}, t_n)}{\partial \theta} \Psi \right]'
\]

\[
= \frac{1}{n} \left[ \frac{\partial \mu(\theta^{(o)}, t_n)}{\partial \theta} \frac{\partial}{\partial \rho} \Psi \right] \left[ \frac{\partial \mu(\theta^{(o)}, t_n)}{\partial \theta} \frac{\partial}{\partial \rho} \Psi \right]'
\]

But
\[
\Psi_n = \left( \begin{array}{cccc}
\sqrt{1-\rho^2} & -\rho & 0 & \ldots & 0 \\
0 & 1 & -\rho & \ldots & 0 \\
0 & 0 & 1 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & \ldots & \ldots & \ldots & 1 \\
0 & 0 & 0 & \ldots & 1
\end{array} \right),
\]

so

\[
\frac{\partial}{\partial \rho} \Psi_n = \left( \begin{array}{cccc}
-\rho/\sqrt{(1-\rho^2)} & -1 & 0 & \ldots & 0 \\
0 & 0 & -1 & \ldots & 0 \\
0 & 0 & 0 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & \ldots & \ldots & \ldots & 0 \\
0 & 0 & 0 & \ldots & 0
\end{array} \right)
\]

Therefore, assumption 7 in Chapter 2 implies that every limit of the diagonal elements of

\[
\frac{1}{n} \left[ \frac{\partial}{\partial \rho} \frac{\partial \mu(\theta^{(o)}, \tau^{(o)}, t_n)}{\partial \theta} \Psi_n \right] \left[ \frac{\partial}{\partial \rho} \frac{\partial \mu(\theta^{(o)}, \tau^{(o)}, t_n)}{\partial \theta} \Psi_n \right]
\]

exists. We have the following theorem.

**Theorem 4.1.** Under model (4.1) and assumptions 1 through 7 in Chapter 2, if \( \hat{\theta} \) is \( \sqrt{n} \)-conistant, then

\[
\sqrt{n} \left( \theta(\hat{\theta}) - \theta^{(o)} \right) \rightarrow N(0, \sigma_o^2 G^{-1}(\rho^{(o)})) \text{ as } n \rightarrow \infty.
\]

§4.1.2. Asymptotic distribution of \( \hat{\xi} \)
Now, we want to show that $\sqrt{n}(\hat{\theta}(\hat{\rho}) - \theta^{(o)})$ and $\sqrt{n}(\hat{\theta}(\rho^{(o)}) - \theta^{(o)})$ have the same asymptotic distribution. From (3.14), we have

$$(\hat{\theta} - \theta^{(o)}) = (M + o(1)) (\theta - \theta^{(o)}).$$

Replacing $\rho^{(o)}$ by $\hat{\rho}$ in $\hat{\theta}$, as we did for $\hat{\theta}$, we have

$$\sqrt{n}(\hat{\theta}(\hat{\rho}) - \theta^{(o)}) = (M + o(1))\sqrt{n}(\hat{\theta}(\hat{\rho}) - \theta^{(o)}).$$

Therefore,

$$\sqrt{n}(\hat{\theta}(\hat{\rho}) - \theta^{(o)}) = (M + o(1))\sqrt{n}(\hat{\theta}(\hat{\rho}) - \theta^{(o)}) = M\sqrt{n}(\hat{\theta}(\hat{\rho}) - \theta^{(o)}) + o(1)\sqrt{n}(\hat{\theta}(\hat{\rho}) - \theta^{(o)}).$$

Now $\sqrt{n}(\hat{\theta}(\hat{\rho}) - \theta^{(o)}) = 0_p(1)$, so $o(1)\sqrt{n}(\hat{\theta}(\hat{\rho}) - \theta^{(o)}) = o_p(1)$ and hence $\sqrt{n}(\hat{\theta}(\hat{\rho}) - \theta^{(o)})$ and $M\sqrt{n}(\hat{\theta}(\hat{\rho}) - \theta^{(o)})$ have the same asymptotic distribution. From Theorem 4.1, we know that $\sqrt{n}(\hat{\theta}(\hat{\rho}) - \theta^{(o)}) \rightarrow N(0, \sigma_o^2G^{-1}(\rho^{(o)}))$. Thus, $\sqrt{n}(\hat{\theta}(\hat{\rho}) - \theta^{(o)}) = M\sqrt{n}(\hat{\theta}(\hat{\rho}) - \theta^{(o)}) + o_p(1) \rightarrow N(0, \sigma_o^2MG^{-1}(\rho^{(o)})M')$. We have the following theorem.

**Theorem 4.2.** Under the model and assumptions in Theorem 4.1, $\sqrt{n}(\hat{\theta}(\hat{\rho}) - \theta^{(o)})$ and $\sqrt{n}(\hat{\theta}(\rho^{(o)}) - \theta^{(o)})$ have the same asymptotic distribution. That is,

$$\sqrt{n}(\hat{\theta}(\hat{\rho}) - \theta^{(o)}) \rightarrow N(0, \sigma_o^2MG^{-1}(\rho^{(o)})M')$$

as $n \rightarrow \infty$. 
§4.2 Asymptotic distribution of $\hat{\rho}$

In section 4.1, we assumed that there exists a $\sqrt{n}$-consistent estimator $\hat{\rho}$ of $\rho^{(o)}$ and then derived the asymptotic distributions of $\hat{\theta}(\hat{\rho})$ and $\hat{\xi}(\hat{\rho})$. In this section, we will show that such an estimator does exist and can be taken to be the m.l.e. of $\rho^{(o)}$. This is done by proving that $\sqrt{n}(\hat{\rho} - \rho^{(o)})$ asymptotically follows a normal distribution, where $\hat{\rho}$ is the maximum likelihood estimator of $\rho^{(o)}$.

From (2.8), we have the following likelihood function:

$$L(\theta, \tau, \rho, \sigma^2|y) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2} \log |\Sigma_n|$$

$$- \frac{1}{2\sigma^2} (y - \mu(\theta, \tau, t_n))^T \Sigma_n^{-1} (y - \mu(\theta, \tau, t_n)). \quad (2.8)$$

The likelihood function is not smooth in $\theta$ and $\tau$, as discussed in Chapter 3, but it is differentiable in $\rho$ for fixed values of $\theta$ and $\tau$. So we can write

$$\frac{\partial L(\theta, \tau, \rho, \sigma^2|y)}{\partial \rho} = \frac{\partial L(\theta, \tau, \rho^{(o)}, \sigma^2|y)}{\partial \rho} + (\rho - \rho^{(o)}) \frac{\partial^2 L(\theta, \tau, \rho^{(o)}, \sigma^2|y)}{\partial \rho^2}$$

$$+ \frac{1}{2} (\rho - \rho^{(o)})^2 \frac{\partial^3 L(\theta, \tau, \rho, \sigma^2|y)}{\partial \rho^3}, \quad (4.8)$$

where $\rho = \alpha \rho + (1-\alpha)\rho^{(o)}$, $0 \leq \alpha \leq 1$. We will follow Cramér's approach (Cramér (1961) page 498-504) to show that
\[ \frac{\partial L(\theta, \tau, \rho, \sigma^2|y)}{\partial \rho} = 0 \text{ has a solution which converges in probability} \]

to \(\rho^{(o)}\), the true value of \(\rho\).

First, multiplying both sides of (4.8) by \(\frac{1}{n}\), we get

\[
\frac{1}{n} \frac{\partial L(\theta, \tau, \rho, \sigma^2|y)}{\partial \rho} = V_o + \frac{1}{n} (\frac{\rho - \rho^{(o)}}{n}) V_1 + \frac{1}{2} (\frac{\rho - \rho^{(o)}}{n})^2 V_2, \quad (4.9)
\]

where

\[
V_o = \frac{1}{n} \frac{\partial L(\theta, \tau, \rho^{(o)}, \sigma^2|y)}{\partial \rho}, \quad V_1 = \frac{1}{n} \frac{\partial^2 L(\theta, \tau, \rho^{(o)}, \sigma^2|y)}{\partial \rho^2}, \quad V_2 = \frac{1}{n} \frac{\partial^3 L(\theta, \tau, \rho, \sigma^2|y)}{\partial \rho^3}.
\]

To evaluate \(V_o\), \(V_1\) and \(V_2\), we need to know \(|\Sigma_n|\). From (4.7)

\[
\Psi_n = \begin{pmatrix}
\sqrt{1-\rho^2} & 0 & 0 & \cdots & 0 & 0 \\
-\rho & 1 & 0 & \cdots & 0 & 0 \\
0 & -\rho & 1 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \ddots & \ddots & \ddots \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & 0 & 0 & \cdots & -\rho & 1
\end{pmatrix}
\]
Since $\Sigma^{-1}_n = \Psi_n \Psi_n'$ and $|\Psi_n| = \sqrt{1-\rho^2}$, then

$$|\Sigma_n| = \frac{1}{|\Sigma^{-1}_n|} = \frac{1}{|\Psi_n \Psi_n'|} = \frac{1}{(\sqrt{1-\rho^2})^2} = \frac{1}{1-\rho^2}.$$  

Also,

$$\frac{\partial \Sigma^{-1}_n}{\partial \rho} = \begin{pmatrix}
0 & -1 & 0 & \ldots & 0 & 0 \\
-1 & 2\rho & -1 & \ldots & 0 & 0 \\
0 & -1 & 2\rho & \ldots & 0 & 0 \\
& \ddots & \ddots & \ddots & \ddots & \ddots \\
& & \ddots & \ddots & \ddots & \ddots \\
0 & 0 & 0 & \ldots & 2\rho & -1 \\
0 & 0 & 0 & \ldots & -1 & 0 \\
\end{pmatrix}.$$  

Thus,

$$V_o = \frac{1}{n} \frac{\partial L(\theta, \xi, \rho^{(o)}, \sigma^2 | y)}{\partial \rho} = \frac{1}{n} \left( \frac{\rho^{(o)}}{1-(\rho^{(o)})^2} - \frac{1}{2\sigma^2} \frac{\partial}{\partial \rho} \hat{\theta} \Sigma^{-1}_n \hat{\theta} \right)$$

$$= \frac{1}{n} \frac{\rho^{(o)}}{1-(\rho^{(o)})^2} - \frac{1}{n\sigma^2} \left( \sum_{i=1}^{n-1} \hat{\theta}_{ni} \hat{\theta}_{n(i+1)} - \rho^{(o)} \sum_{i=1}^{n-1} \hat{\theta}_{ni}^2 \right)$$

$$= \frac{1}{n} \frac{\rho^{(o)}}{1-(\rho^{(o)})^2} - \frac{1}{n\sigma^2} \left( \sum_{i=1}^{n-1} \hat{\theta}_{ni}^2 \hat{\theta}_{n(i+1)} + \rho^{(o)} \hat{\theta}_{n1}^2 \right). \quad (4.10)$$

From **Theorem 2.1** and **Theorem 2.2**, we know that
\[ \theta - \theta^{(o)} = 0_p \left( n^{-\frac{1}{2}} (\log \log n)^{\frac{1}{2}} \right), \quad \hat{\theta} - \tau^{(o)} = 0_p \left( n^{-\frac{1}{2}} (\log \log n)^{\frac{1}{2}} \right), \]

so we can write

\[ \hat{u}_{ni} = u_{ni} + 0_p \left( n^{-\frac{1}{2}} (\log \log n)^{\frac{1}{2}} \right), \quad \hat{e}_{ni} = e_{ni} + 0_p \left( n^{-\frac{1}{2}} (\log \log n)^{\frac{1}{2}} \right). \]

Then,

\[ V_o = \frac{1}{n} \frac{\rho^{(o)}}{1 - (\rho^{(o)})^2} - \frac{1}{n} \sum_{i=1}^{n-1} u_{ni} e_{n(i+1)} + 0_p \left( n^{-\frac{1}{2}} (\log \log n)^{\frac{1}{2}} \right) \sum_{i=1}^{n-1} u_{ni} \]

\[ + 0_p \left( n^{-\frac{1}{2}} (\log \log n)^{\frac{1}{2}} \right) \sum_{i=1}^{n-1} e_{n(i+1)} + 0_p \left( n^{-1} (\log \log n) \right) \right]. \]

It is easy to see that as \( n \to \infty \)

\[ \frac{1}{n} \frac{\rho^{(o)}}{1 - (\rho^{(o)})^2} \to 0. \]

Also,

\[ \frac{1}{n} 0_p \left( n^{-\frac{1}{2}} (\log \log n)^{\frac{1}{2}} \right) \sum_{i=1}^{n-1} e_{n(i+1)} \to 0 \quad \text{in probability} \]

and

\[ \frac{1}{n} 0_p \left( n^{-1} (\log \log n) \right) \to 0 \quad \text{in probability}. \]
So if we can find the probability limits of
\[ \frac{1}{n} \sum_{i=1}^{n-1} u_{ni} \epsilon_{n(i+1)} \]
and
\[ \frac{1}{n} \rho^{\frac{-1}{2}} (\log \log n)^{\frac{1}{2}} \sum_{i=1}^{n-1} u_{ni}, \]
we obtain the probability limit of \( V_0 \).

To show that \( \frac{1}{n} \sum_{i=1}^{n-1} u_{ni} \) converges to 0 in probability, notice that we have
\[ \mathbb{E}(\frac{1}{n} \sum_{i=1}^{n-1} u_{ni}) = 0 \]
and
\[ \text{Var}(\frac{1}{n} \sum_{i=1}^{n-1} u_{ni}) = \frac{1}{n^2} \text{Var} (\sum_{i=1}^{n-1} u_{ni}) \]

\[ = \frac{1}{n^2} j' \Sigma_{n-1} j, \]

where \( j' = [1,1,1,\ldots,1] \) and \( \Sigma_{n-1} \) is the covariance matrix of the vector \([u_{n1}, u_{n2}, \ldots, u_{n(n-1)}]^t\). But

\[ j' \Sigma_{n-1} j = (n-1) + 2\rho(n-2) + 2\rho^2(n-3) + \ldots + 2\rho^{n-2} \]

\[ < 2(n-1) \left( 1 + \left| \rho \right| + \left| \rho^2 \right| + \ldots + \left| \rho^{n-2} \right| \right) \]

\[ < \frac{2(n-1)}{1 - \left| \rho \right|}, \]

so as \( n \to \infty \),

\[ \text{Var}(\frac{1}{n} \sum_{i=1}^{n-1} u_{ni}) = \frac{1}{n^2} j' \Sigma_{n-1} j \leq \frac{2(n-1)}{n^2(1 - \left| \rho \right|)} \to 0. \]
This implies that
\[
\frac{1}{n} \sum_{i=1}^{n-1} u_{ni} \to 0 \text{ in probability.}
\]

For \( \frac{1}{n} \sum_{i=1}^{n-1} u_{ni} e_{n(i+1)} \), we know that
\[
E(u_{ni} e_{n(i+1)}) = 0,
\]

\[
\text{Cov}(u_{ni} e_{n(i+1)}, u_{nj} e_{n(j+1)}) = E(u_{ni} e_{n(i+1)} u_{nj} e_{n(j+1)}) = 0 \text{ for } i \neq j
\]

and
\[
\text{Var}(u_{ni} e_{n(i+1)}) = \text{Var}(u_{ni}) \text{ Var}(e_{n(i+1)}) = \frac{\sigma^4}{1 - \rho^2}, \text{ for all } i.
\]

Therefore, from Chung (Theorem 5.1.1 on page 103), we have
\[
\frac{1}{n} \sum_{i=1}^{n-1} u_{ni} e_{n(i+1)} \to 0 \text{ a.e.}
\]

From the above derivation, we have the result that
\[
V_o = \frac{1}{n} \frac{\partial L(\hat{\theta}, \hat{\tau}, \rho^{(o)}, \sigma^2|y)}{\partial \rho} \to 0 \text{ in probability.}
\]

The next step is to find the probability limit of
\[
V_1 = \frac{1}{n} \frac{\partial^2 L(\hat{\theta}, \hat{\tau}, \rho^{(o)}, \sigma^2|y)}{\partial \rho^2}.
\]
Since

\[
\frac{1}{n} \frac{\partial L(\theta, \hat{\ell}, \rho, \sigma^2|y)}{\partial \rho} = \frac{1}{n} \frac{\rho}{1-\rho^2} - \frac{1}{n\sigma^2} \left( \sum_{i=1}^{n-1} \hat{u}_{ni} \hat{u}_{n(i+1)} - \rho \sum_{i=1}^{n-1} \hat{u}_{ni}^2 \right),
\]

we have that

\[
V_1 = \frac{1}{n} \frac{\partial^2 L(\theta, \hat{\ell}, \rho^{(o)}, \sigma^2|y)}{\partial \rho^2} = \frac{1}{n} \frac{\partial}{\partial \rho} \left( \frac{\rho}{1-\rho^2} \right) \bigg|_{\rho=\rho^{(o)}} + \frac{1}{n\sigma^2} \sum_{i=1}^{n-1} \hat{u}_{ni}^2
\]

converges to

\[
\frac{1}{\sigma^2} \text{Var}(u_{ni}) = \frac{1}{1-(\rho^{(o)})^2} \text{ in probability.}
\]

For \( V_2 = \frac{1}{n} \frac{\partial^3 L(\theta, \hat{\ell}, \rho, \sigma^2|y)}{\partial \rho^3} \), it is easy to see that as \( n \to \infty \),

\[
V_2 = \frac{1}{n} \frac{\partial^2}{\partial \rho^2} \left( \frac{\rho}{1-\rho^2} \right) \bigg|_{\rho=\bar{\rho}} \to 0 \text{ in probability.}
\]

Now setting the expression in (4.9) to be 0 and directly following Cramér's approach (Cramér 1961, page 502-503), we can show that there is a root of \( \rho \) in the neighborhood of \( \rho^{(o)} \) if \( n \) is sufficiently large. Therefore, the proof of the existence of a consistent sequence of roots of the likelihood equation

\[
\frac{1}{n} \frac{\partial L(\theta, \hat{\ell}, \rho, \sigma^2|y)}{\partial \rho} = 0
\]

is complete.
Letting \( \hat{\rho} \) be the root of \( \frac{1}{n} \frac{\partial L(\hat{\theta}, \hat{\xi}, \rho, \sigma^2 | y)}{\partial \rho} = 0 \), we have, from (4.8),

\[
\frac{1}{n} \frac{\partial L(\hat{\theta}, \hat{\xi}, \hat{\rho}, \sigma^2 | y)}{\partial \rho} = \frac{1}{n} \frac{\partial L(\hat{\theta}, \hat{\xi}, \rho^{(o)}, \sigma^2 | y)}{\partial \rho} + \frac{1}{n} (\hat{\rho} - \rho^{(o)}) \frac{\partial^2 L(\hat{\theta}, \hat{\xi}, \rho^{(o)}, \sigma^2 | y)}{\partial \rho^2}
\]

\[+ \frac{1}{2n} (\hat{\rho} - \rho^{(o)})^2 \frac{\partial^3 L(\hat{\theta}, \hat{\xi}, \rho^{(o)}, \sigma^2 | y)}{\partial \rho^3} = 0, \tag{4.11}\]

where \( \rho = \alpha \hat{\rho} + (1 - \alpha) \rho^{(o)} \), \( 0 \leq \alpha \leq 1 \). Therefore,

\[
\sqrt{n} (\hat{\rho} - \rho^{(o)}) = -\sqrt{n} \left[ \frac{1}{n} \frac{\partial L(\hat{\theta}, \hat{\xi}, \rho^{(o)}, \sigma^2 | y)}{\partial \rho} \right] \times \left[ \frac{1}{n} \frac{\partial L(\hat{\theta}, \hat{\xi}, \rho^{(o)}, \sigma^2 | y)}{\partial \rho} + \frac{1}{2n} (\hat{\rho} - \rho^{(o)})^2 \frac{\partial^3 L(\hat{\theta}, \hat{\xi}, \rho^{(o)}, \sigma^2 | y)}{\partial \rho^3} \right].
\]

But, we know that

\[
n \left[ \frac{\partial^2 L(\hat{\theta}, \hat{\xi}, \rho^{(o)}, \sigma^2 | y)}{\partial \rho^2} \right]^{-1} \rightarrow 1 - (\rho^{(o)})^2 \text{ in probability.}
\]

Also, since \( \hat{\rho} \) is a consistent estimator,

\[
\frac{1}{2\sqrt{n}} (\hat{\rho} - \rho^{(o)})^2 \frac{\partial^3 L(\hat{\theta}, \hat{\xi}, \rho^{(o)}, \sigma^2 | y)}{\partial \rho^3} \rightarrow 0 \text{ in probability.}
\]
So the only thing we need to find is the asymptotic distribution of
\[ \frac{1}{\sqrt{n}} \frac{\partial L(\hat{\theta}, \hat{\phi}, \rho^{(o)}, \sigma^2 \mid y)}{\partial \rho} \]
. We will apply the following lemma, which is due to Stein (1972), to obtain the asymptotic distribution of
\[ \sqrt{n} (\hat{\theta} - \theta^{(o)}) \]. Let \( \{ \alpha_k \} \) be a sequence such that if A and B are any two finite sets of natural numbers for which

\[ \inf_{i \in A, j \in B} |i - j| \geq k \]

and Y and Z are two random variables with finite variances depending on the \( \{ X_i \}_{i \in A} \) and the \( \{ X_j \}_{j \in B} \), respectively, then

\[ |\text{corr}(Y, Z)| \leq \alpha_k \]

Also, for sufficiently large k, there exists a positive number \( \lambda \) such that \( \alpha_k \leq e^{-\lambda k} \).

**Lemma 4.1 (Stein 1972)** If \( X_1, X_2, X_3, \ldots \) is a stationary sequence of random variables satisfying

(a). \( E(X_i) = 0, E(X_i^2) = 1 \)

(b). \( E(X_i^8) < \infty \)

(c). \( 0 < c = \lim_{n \to \infty} \frac{1}{n} \text{Var}( \sum_{i=1}^{n} X_i ) < \infty \)

(d). \( |\text{corr}(Y, Z)| \leq \alpha_k \)

then there exists a constant \( d \) (depending on the
distribution of $X_1, X_2, X_3, \ldots\,$, but not on $n$) such that

$$
\left\{ P\left\{ \frac{\sum_{i=1}^{n} X_i}{\sqrt{\text{Var}(\sum_{i=1}^{n} X_i)}} \leq a \right\} - \Phi(a) \right\} \leq d \frac{1}{n^2},
$$

where $\Phi(\cdot)$ is the standard normal distribution.

In our case,

$$
\frac{1}{\sqrt{n}} \frac{\partial L(\theta, \xi, \rho^{(o)}, \sigma^2 | y)}{\partial \rho} = \frac{1}{\sqrt{n}} \frac{\rho^{(o)}}{1 - (\rho^{(o)})^2} - \frac{1}{\sqrt{n} \sigma^2} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \hat{\gamma}_{ni} \hat{\gamma}_{nj} (\rho^{(o)})^2 \hat{\gamma}_{n1}^2.
$$

Since as $n \to \infty$,

$$
\frac{1}{\sqrt{n}} \frac{\rho^{(o)}}{1 - (\rho^{(o)})^2} \to 0
$$

and

$$
\frac{1}{\sqrt{n} \sigma^2} \rho^{(o)} \hat{\gamma}_{n1}^2 \to 0 \text{ in probability,}
$$

then

$$
\frac{1}{\sqrt{n}} \frac{\partial L(\theta, \xi, \rho^{(o)}, \sigma^2 | y)}{\partial \rho} \text{ and } -\frac{1}{\sqrt{n} \sigma^2} \sum_{i=1}^{n-1} \hat{\gamma}_{ni} \hat{\gamma}_{n(i+1)} \text{ have the same asymptotic distribution. However, we know}
$$
\[ n^{-\frac{1}{2}} \sum_{i=1}^{n-1} \hat{u}_{ni} e_n(i+1) = n^{-\frac{1}{2}} \sum_{i=1}^{n-1} u_{ni} e_n(i+1) + o_p(1). \]

So we only need to consider the term

\[ -\frac{1}{\sqrt{n} \sigma^2} \sum_{i=1}^{n-1} u_{ni} e_n(i+1). \]

Letting \( X_i = u_{ni} e_n(i+1) \), we have

\[ E( X_i ) = 0, \]

\[ \text{Var}( X_i ) = \sigma^2 \times \frac{\sigma^2}{1-(\rho(o))^2} = \frac{\sigma^4}{1-(\rho(o))^2}, \]

and

\[ \text{Cov}( X_i, X_j ) = 0, \text{ for } i \neq j. \]

It is easy to verify that conditions (a), (b), (c) and (d) in Lemma 4.1 are satisfied. Therefore, applying Lemma 4.1 we have that

\[ n^{-\frac{1}{2}} \sum_{i=1}^{n-1} u_{ni} e_n(i+1) \rightarrow N(0, \frac{\sigma^4}{1-(\rho(o))^2}), \]

and hence

\[ \frac{1}{\sqrt{n}} \frac{\partial L(\theta, \tau, \rho(o), \sigma^2 | y)}{\partial \rho} \rightarrow N(0, \frac{1}{1-(\rho(o))^2}). \]

Now
\[
\sqrt{n} \left( \hat{\rho} - \rho^{(o)} \right) = -n \left[ \frac{\partial^2 L(\hat{\theta}, \hat{\tau}, \rho^{(o)}, \sigma^2 | y)}{\partial \rho^2} \right]^{-1} \times \frac{1}{\sqrt{n}} \frac{\partial L(\hat{\theta}, \hat{\tau}, \rho^{(o)}, \sigma^2 | y)}{\partial \rho}
\]

and

\[
n \left[ \frac{\partial^2 L(\hat{\theta}, \hat{\tau}, \rho^{(o)}, \sigma^2 | y)}{\partial \rho^2} \right]^{-1} \rightarrow 1 - (\rho^{(o)})^2 \quad \text{in probability},
\]

so \( \sqrt{n} \left( \hat{\rho} - \rho^{(o)} \right) \rightarrow N(0, 1 - (\rho^{(o)})^2) \) asymptotically. We have the following theorem.

**Theorem 4.3.** Under model (4.1) and assumptions 1 through 7 in Chapter 2, the random variable \( \sqrt{n} \left( \hat{\rho} - \rho^{(o)} \right) \)

asymptotically follows a \( N(0, 1 - (\rho^{(o)})^2) \)

distribution, where here \( \hat{\rho} \) is the m.l.e. of \( \rho^{(o)} \).

Before going to next section, we should mention that the m.l.e., \( \hat{\rho} \), derived above is not 'really' the global m.l.e., but is just one of the roots of the likelihood equation \( \frac{1}{n} \frac{\partial L(\hat{\theta}, \hat{\tau}, \rho, \sigma^2 | y)}{\partial \rho} = 0 \). The reasons we use the term 'm.l.e.' here are that Cramér uses this word in his classical book *Mathematical Methods of Statistics* and we adopt his approach to prove the existence and consistency of the root of the likelihood equation. Investigating the likelihood equation, we need only consider the parameter \( \rho \) since the parameters \( \theta \) and \( \tau \) were replaced by their m.l.e.'s., \( \hat{\theta} \) and \( \hat{\tau} \).

Therefore, the situation in

\[
\frac{1}{n} \frac{\partial L(\hat{\theta}, \hat{\tau}, \rho, \sigma^2 | y)}{\partial \rho} = 0
\]

is a one-parameter case for the fixed \( \hat{\theta} \) and \( \hat{\tau} \). This approach makes it easier to find the asymptotic distribution of \( \hat{\rho} \). Because of the irregular behavior of the likelihood function, we avoid considering...
the joint asymptotic distribution of \( \hat{\rho}, \hat{\theta} \) and \( \hat{\ell} \), which we expect to be difficult to analyze. We conjecture that the pseudo-problem approach will work if we try to find the joint asymptotic distribution of \( \hat{\rho}, \hat{\theta} \) and \( \hat{\ell} \).

§4.3 Asymptotic distribution of \( \delta^2 \)

In §3.5, we assumed that \( \rho \) was known and then derived the asymptotic distribution of \( \delta^2 \). In this section, we will find the asymptotic distribution of \( \delta^2 \) in the case that \( \rho \) is unknown.

From §4.2, we have the likelihood function

\[
L(\theta, \tau, \rho, \sigma^2 | y) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2} \log |\Sigma_n| \\
- \frac{1}{2\sigma^2} (y - \mu(\theta, \tau, t_n))^\prime \Sigma_n^{-1} (y - \mu(\theta, \tau, t_n)). (2.8)
\]

The m.l.e. \( \delta^2 \) of \( \sigma_o^2 \) is the solution of \( \frac{\partial L(\hat{\theta}, \hat{\ell}, \hat{\rho}, \delta^2 | y)}{\partial \delta^2} = 0 \), where \( \hat{\theta}, \hat{\ell} \) and \( \hat{\rho} \) are the m.l.e.'s of \( \theta^{(o)}, \tau^{(o)} \) and \( \rho^{(o)} \), respectively. Solving the likelihood equation, we have

\[
\delta^2 = \frac{1}{n} (y - \mu(\hat{\theta}, \hat{\ell}, t_n))^\prime \hat{\Sigma}_n^{-1} (y - \mu(\hat{\theta}, \hat{\ell}, t_n)),
\]

where \( \hat{\Sigma}_n^{-1} \) is \( \Sigma_n^{-1} \) with \( \rho \) replaced by \( \hat{\rho} \). Using the representation in §3.5, we have
\[ S(\hat{\theta}, \hat{\tau}, \hat{\rho}) = \delta^2 = \frac{1}{n} \left( y - \mu(\hat{\theta}, \hat{\tau}, t_n) \right)' \Sigma_n^{-1} \left( y - \mu(\hat{\theta}, \hat{\tau}, t_n) \right). \]

For fixed values of \( \hat{\theta} \) and \( \hat{\tau} \), \( S(\hat{\theta}, \hat{\tau}, \rho) \) is a differentiable function of \( \rho \). As in (4.9), we can expand \( S(\hat{\theta}, \hat{\tau}, \rho) \) with respect to \( \rho^{(o)} \) and obtain

\[
S(\hat{\theta}, \hat{\tau}, \rho) = S(\hat{\theta}, \hat{\tau}, \rho^{(o)}) + (\hat{\rho} - \rho^{(o)}) \frac{\partial S(\hat{\theta}, \hat{\tau}, \rho^{(o)})}{\partial \rho} \\
+ \frac{1}{2} (\hat{\rho} - \rho^{(o)})^2 \frac{\partial^2 S(\hat{\theta}, \hat{\tau}, \rho^{(o)})}{\partial \rho^2}. \tag{4.12}
\]

We have equality in (4.12) is because of the fact that \( S(\hat{\theta}, \hat{\tau}, \rho) \) is a polynomial in \( \rho \) with degree 2 for fixed \( \hat{\theta} \) and \( \hat{\tau} \). Also, since \( \hat{\theta}, \hat{\tau} \) and \( \hat{\rho} \) are the m.l.e.'s of \( \theta^{(o)}, \tau^{(o)} \) and \( \rho^{(o)} \), we have

\[
\frac{\partial S(\hat{\theta}, \hat{\tau}, \rho)}{\partial \rho} = \frac{\partial S(\hat{\theta}, \hat{\tau}, \rho^{(o)})}{\partial \rho} + (\hat{\rho} - \rho^{(o)}) \frac{\partial^2 S(\hat{\theta}, \hat{\tau}, \rho^{(o)})}{\partial \rho^2}
\]

\[= 0. \]

Therefore,

\[- \frac{\partial S(\hat{\theta}, \hat{\tau}, \rho^{(o)})}{\partial \rho} = (\hat{\rho} - \rho^{(o)}) \frac{\partial^2 S(\hat{\theta}, \hat{\tau}, \rho^{(o)})}{\partial \rho^2}. \tag{4.13}\]

From (4.12) and (4.13), We have

\[
S(\hat{\theta}, \hat{\tau}, \rho) - S(\hat{\theta}, \hat{\tau}, \rho^{(o)}) \\
= (\hat{\rho} - \rho^{(o)}) \frac{\partial S(\hat{\theta}, \hat{\tau}, \rho^{(o)})}{\partial \rho} + \frac{1}{2} (\hat{\rho} - \rho^{(o)})^2 \frac{\partial^2 S(\hat{\theta}, \hat{\tau}, \rho^{(o)})}{\partial \rho^2}
\]
\[
= (\hat{\rho} - \rho^{(o)}) \frac{\partial S(\theta, \hat{\xi}, \rho^{(o)})}{\partial \rho} + \frac{1}{2} (\hat{\rho} - \rho^{(o)})^2 \frac{\partial^2 S(\theta, \hat{\xi}, \rho^{(o)})}{\partial \rho^2}
\]

\[
= - (\hat{\rho} - \rho^{(o)})^2 \frac{\partial^2 S(\theta, \hat{\xi}, \rho^{(o)})}{\partial \rho^2} + \frac{1}{2} (\hat{\rho} - \rho^{(o)})^2 \frac{\partial^2 S(\theta, \hat{\xi}, \rho^{(o)})}{\partial \rho^2}
\]

\[
= \frac{1}{2} (\hat{\rho} - \rho^{(o)})^2 \frac{\partial^2 S(\theta, \hat{\xi}, \rho^{(o)})}{\partial \rho^2}
\]

Therefore,

\[
\sqrt{n} \left( S(\theta, \hat{\xi}, \hat{\rho}) - S(\theta, \hat{\xi}, \rho^{(o)}) \right) = \frac{1}{2} \sqrt{n} (\hat{\rho} - \rho^{(o)})^2 \frac{\partial^2 S(\theta, \hat{\xi}, \rho^{(o)})}{\partial \rho^2}.
\]

From §4.2, we know that

\[
\frac{\partial^2 S(\theta, \hat{\xi}, \rho^{(o)})}{\partial \rho^2} = \frac{2}{n} \sum_{i=2}^{n-1} u_i^2 \to 2 \frac{\sigma_o^2}{1 - (\rho^{(o)})^2} \text{ in probability.}
\]

Also, \( \hat{\rho} \) is a \( \sqrt{n} \)-consistent estimator, therefore as \( n \to \infty \)

\[
\sqrt{n} \left( S(\theta, \hat{\xi}, \hat{\rho}) - S(\theta, \hat{\xi}, \rho^{(o)}) \right) \to 0 \text{ in probability.}
\]

This means that as \( n \to \infty \),

\[
\sqrt{n}(S(\theta, \hat{\xi}, \hat{\rho}) - \sigma_o^2) - \sqrt{n}(S(\theta, \hat{\xi}, \rho^{(o)}) - \sigma_o^2) \to 0 \text{ in probability.} \quad (4.14)
\]
We note here that the m.l.e.'s \( \hat{\theta} \) and \( \hat{\xi} \) in the above paragraph are computed in the case that \( \rho^{(o)} \) is unknown. To distinguish the m.l.e.'s \( \hat{\theta} \) and \( \hat{\xi} \) computed in the case that \( \rho^{(o)} \) is known from those obtained from the case that \( \rho^{(o)} \) is unknown, we will represent the former as \( \hat{\theta}(\rho^{(o)}) \) and \( \hat{\xi}(\rho^{(o)}) \) and the latter as \( \hat{\theta}(\hat{\rho}) \) and \( \hat{\xi}(\hat{\rho}) \). Now from Theorem 3.3, we know that if \( \rho^{(o)} \) is known,
\[
\sqrt{n} \left( S(\hat{\theta}(\rho^{(o)}), \hat{\xi}(\rho^{(o)}), \rho^{(o)}) - \sigma_{o}^{2} \right) \text{ asymptotically follows a } N(0, 2\sigma_{o}^{4})
\]
distribution. Although the m.l.e.'s for the two cases are different, we will claim that the asymptotic behaviors of
\[
\sqrt{n} \left[ S(\hat{\theta}(\rho^{(o)}), \hat{\xi}(\rho^{(o)}), \rho^{(o)}) - \sigma_{o}^{2} \right] \text{ and } \sqrt{n} \left[ S(\hat{\theta}(\hat{\rho}), \hat{\xi}(\hat{\rho}), \rho^{(o)}) - \sigma_{o}^{2} \right]
\]
are the same.

Notice that from Theorem 4.1 and Theorem 4.2, we have
\[
\left( \hat{\theta}(\rho^{(o)}), \hat{\xi}(\rho^{(o)}) \right) \overset{d}{=} \left( \hat{\theta}(\hat{\rho}), \hat{\xi}(\hat{\rho}) \right) \text{ asymptotically,}
\]
where \( \overset{d}{=} \) means that the two random variables are equal in distribution. Since \( S(\star,\star,\star) \) is measurable, we have
\[
S(\hat{\theta}(\rho^{(o)}), \hat{\xi}(\rho^{(o)})) \overset{d}{=} S(\hat{\theta}(\hat{\rho}), \hat{\xi}(\hat{\rho})) \text{ asymptotically,}
\]
or equivalently,
\[
\sqrt{n} \left[ S(\hat{\theta}(\rho^{(o)}), \hat{\xi}(\rho^{(o)}), \rho^{(o)}) - \sigma_{o}^{2} \right] \overset{d}{=} \sqrt{n} \left[ S(\hat{\theta}(\hat{\rho}), \hat{\xi}(\hat{\rho}), \rho^{(o)}) - \sigma_{o}^{2} \right]
\]
\[ \sqrt{n} \left[ S(\hat{\theta}(\hat{\beta}), \hat{\beta}), \hat{\beta} - \sigma_0^2 \right] \text{ and } \sqrt{n} \left[ S(\hat{\theta}(\rho^{(o)}), \hat{\theta}(\rho^{(o)}), \rho^{(o)}) - \sigma_0^2 \right] \] have the same asymptotic distribution. We obtain the following theorem.

**Theorem 4.4.** Under the assumptions of Theorem 4.3, if \( \hat{\sigma}^2 \) is the m.l.e. of \( \sigma_0^2 \), then \( \sqrt{n} \left( \hat{\sigma}^2 - \sigma_0^2 \right) \to N(0, 2\sigma_0^4) \) as \( n \to \infty \).
Chapter V

Data Analysis

In this chapter, we illustrate an alternative method of computing the maximum likelihood estimates for the regression coefficients and knot points. The method is shown to be asymptotically equivalent to the iterative procedure mentioned in Chapter 2. Angular motion data obtained from Gait Laboratory at Children's Hospital in San Diego are used in the analysis. We first describe the data set in §5.1. The reasons that we choose the cubic spline regression model with continuity constraints only at the unknown knot points and AR(1) errors are given in §5.2. This model may need to be modified, as discussed in §5.5. A flow chart of the SAS code that used to obtain the least squares estimates is given in Table 5.3 (SAS Institute, 1985). In §5.3, we describe the steps of the SAS program. The verification that the computational method described in §5.3 is asymptotically equivalent to the procedure described in Chapter 2 is given in §5.4. The asymptotic covariance matrix given by SAS and the asymptotic covariance matrix obtained from our theoretical results are given in §5.5. The discrepancy is investigated and explained.

§5.1. Angular motion data

The angular motion for cerebral palsy patients was collected by the Gait Laboratory at Children's Hospital in San Diego. When a cerebral palsy patient comes to the hospital, he is asked to remove his shoes, walk barefoot and pass down a walkway at least three times where the movements of the patient are recorded. There are four motion picture cameras and a motion analyzer to record the motion. See Sutherland (1984, page 3) for camera layout. The films taken by the cameras are displayed on a motion analyzer and are

124
viewed frame by frame. A sonic digitizer is used to determine the X and Y coordinate of the positions of the markers on the viewing screen; and from there the left hip flexion-extension measurements, "the angle between the line segment formed by the line between the hip and the knee center" (see Sutherland (1984), page 7), were computed.

A gait cycle is the movements and events that occur between successive footsteps of the same foot (Sutherland(1984)). Typically, a gait cycle contains five gait events: foot-strike, opposite-toe-off, opposite-foot-strike, toe-off and foot-strike again (Sutherland (1984) page 11). It is reasonable to assume that the angular motion between successive gait events are different, since bones and muscles are functioning differently in different periods. When patient is passing down the walkway, time-distance variables including cadence, stride, velocity, foot-strike and toe-off are measured also. According to Sutherland(1984, page 25), the joint angles of the motion measured throughout the gait cycle are weak determinants of gait. The time-distance variables characterize the gait. Since cerebral palsy patients have the difficulty in walking and cadence, stride and velocity are highly correlated, we used variables cadence, right foot-strike, right toe-off and left toe-off in the selection of a representative gait among different passes down the walkway for the same patient. Among all passes made by one patient, the one with smallest Mahalanobis distance for variables cadence, right foot-strike, right toe-off and left toe-off is choosen as the representative gait of the patient. The SAS code used to select the representative gaits was written by Professor Leurgans and is attached in Appendix B.

Our data set contains three variables: time fraction, left hip flexion-extension and gait event. For the patient we choose, 52 frames were observed. Since we want to fit hip flexion-extension against time, the independent variable is time fraction and the response variable is left hip flexion-extension in degrees. Time fraction is computed from frame number divided by total number of frames. Therefore, the variable time fraction is equally spaced.
The degree coded for the hip flexion-extension is taken 90° from the observed angle. Therefore, if the observed angle is 90°, then it is coded as 0°. Each angle of left hip flexion-extension corresponds to one value of time fraction. The variable gait event indicates the times that foot-strikes and toe-offs occur. The gait cycle starts at time 0 and ends at time 1. Therefore, three gait events occur.

§5.2. Preliminary analysis

As described in §5.1, we have three gait events within the time interval (0,1), and the positions of the gait events are reported in the data. Therefore, it seems that we can fit the motion data with the known knots model. However, we know that because of the technicians and facility used in measuring the angles, the reported times of the gait events need not be accurate. Therefore, the model with measurement errors on the gait events and the model with gait events occurring at unknown times arise naturally. However, the mean function of the measurement errors on gait events model does not coincide with the mean function of the unknown knots model, which we suppose to be the true mean function. The following example illustrates this fact.

Consider the following two models:

A. \( y_t = 1 + t + \theta(t - \tau)_+ + \varepsilon_t \)

B. \( y_t = 1 + t + \theta(t - w)_+ + \varepsilon_t \)

where \( \tau \) is unknown, \( t \) is an independent variable, the \( \varepsilon_t \)'s are i.i.d. normal with mean 0 and variance \( \sigma^2 \), and \( w \) is a random variable. Model A is a model with a unknown knot, but Model B is a model with measurement error on the knot point. For model A, the expected value of \( y_t \) can be easily found to be
\[ E_A(y_t) = 1 + t + \theta(t - \tau)_+. \]

For model B, if \( w \) has a triangular distribution on \([\tau - h, \tau + h]\), \( h > 0 \),

\[
f_\tau(w) = \begin{cases} 
\frac{1}{h^2} (w - \tau + h) & \tau - h \leq w \leq \tau \\
\frac{1}{h^2} (\tau + h - w) & \tau \leq w \leq \tau + h 
\end{cases},
\]

then the expected value \( E_B(y_t) \) of model B, after some straightforward but tedious algebra, can be found to be

\[
E_B(y_t) = \begin{cases} 
1 + t & \text{if } 0 \leq t \leq \tau - h \\
1 + t + \frac{\theta}{6h^2}(t - \tau + h)^3 & \text{if } \tau - h \leq t \leq \tau \\
1 + t + \theta \left\{ \frac{1}{6}(3\tau - h - 3t) + \frac{1}{h^2} \left[ \frac{1}{6}(t - \tau)^3 + \frac{h}{2}(t - \tau)^2 \right] \right\} & \text{if } \tau \leq t \leq \tau + h \\
1 + t + \theta(t - \tau) & \text{if } \tau + h \leq t \leq 1.
\end{cases}
\]

The difference between the two expected values \( E_A \) and \( E_B \) is

\[
D(t) = E_A(y_t) - E_B(y_t) = \begin{cases} 
0 & \text{if } 0 \leq t \leq \tau - h \\
\frac{\theta}{6h^2}(t - \tau + h)^3 & \text{if } \tau - h \leq t \leq \tau \\
\theta \left\{ \frac{1}{6}(3\tau - h - 3t) + \frac{1}{h^2} \left[ \frac{1}{6}(t - \tau)^3 + \frac{h}{2}(t - \tau)^2 \right] \right\} & \text{if } \tau \leq t \leq \tau + h \\
0 & \text{if } \tau + h \leq t \leq 1.
\end{cases}
\]
From the expression for $D(t)$, it can be seen that $D(t)$ is 0 for $0 \leq t \leq \tau - h$ and $\tau + h \leq t \leq 1$, no matter what the values of $\theta$ and $h$ are. For $\tau - h \leq t \leq \tau + h$, the magnitude of $D(t)$ depends on $\theta$ and $h$. The larger the values of $|\theta|$ and $h$ are, the bigger the discrepancy between the two mean functions is. See Figure 5.1 for $\theta = 10, -3$ and $h = 0.1, 0.3$.

The above example illustrates that mean function of the model with measurement errors on knot points depends on the values of $h$. As $h \to 0$, $E_B(y_t) \to E_A(y_t)$. This means that $\hat{E}_B(y_t)$, the estimator of $E_B(y_t)$, will not be an asymptotically unbiased estimator of $1 + t + \theta(t - \tau)_+$ in the sense of Serfling (1980, page 48) unless $h = 0$.

Since the model with measurement error in the unknown knot points leads to biased estimator of the mean function, we instead will consider a model with unknown knot points. The advantage of using the unknown knots model to fit our motion data is that we are given starting values of the positions of the gait events, which is crucial if we want to get the values computed faster with a nonlinear regression techniques.

The second thing we have to consider is the smoothness of the mean function. Since our data are obtained from cerebral palsy patients, and cerebral palsy patients may have difficulty in the velocity of their walk at the gait events, it is reasonable to assume that the mean function is continuous, but not in the first derivative at the gait events. The above two considerations lead us to consider the following four-piece generalized cubic spline model:

$$y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \alpha_3 t^3 + \sum_{j=1}^{3} \beta_{1j}(t-\tau_1)_+^j + \sum_{j=1}^{3} \beta_{2j}(t-\tau_2)_+^j + \sum_{j=1}^{3} \beta_{3j}(t-\tau_3)_+^j + u_t. \quad (5.1)$$
$D(t) = E_a[Y(t)] - E_b[Y(t)]$

**Figure 5.1: Plot of Difference of Two Mean Functions**

Solid Curve = (-3,0.1)  Dotted Curve = (-3,0.3)
Dashed Curve = (10,0.1), Intermediately Dashed Curve = (10,0.3)
Note: Theta = -3,10; h = 0.1,0.3
For model (5.1), the $\alpha_i$'s are coefficients of a cubic, and the $\beta_{ij}$'s, $i,j = 1,2,3$, are cubic adjustments to the cubic of the preceding interval. The reasons we arbitrarily choose the model to be the generalized cubic spline in (5.1) are that it ordinarily fits data well and the polynomial pieces of Chapters 3 and 4 have the same degree for every interval.

Now we use model (5.1) to fit our motion data. Assuming $u_t$ to be i.i.d. errors and the knot points to be $\tau_1 = 0.17308$, $\tau_2 = 0.53846$ and $\tau_3 = 0.69231$, which are the reported times of the gait events, we obtain

\[
\hat{\alpha}_0 = 1.8578 \quad \hat{\alpha}_1 = 27.2275 \quad \hat{\alpha}_2 = 130.9853 \quad \hat{\alpha}_3 = -334.962
\]

\[
\hat{\beta}_{11} = -16.8493 \quad \hat{\beta}_{12} = -51.0177 \quad \hat{\beta}_{13} = 451.8243
\]

\[
\hat{\beta}_{21} = -22.1907 \quad \hat{\beta}_{22} = 350.9835 \quad \hat{\beta}_{23} = -1985.7
\]

\[
\hat{\beta}_{31} = 26.1862 \quad \hat{\beta}_{32} = 335.6135 \quad \hat{\beta}_{33} = 2093.0882
\]

The above estimated values $\hat{\alpha}_0$'s, $\hat{\alpha}_i$'s and $\hat{\beta}_i$'s will be used to find the estimated regression coefficients $\hat{\theta}_{ij}$'s of the piecewise polynomial model. As stated in Chapter 1, model (5.1) can be written as the following piecewise polynomial model:

\[
f(\theta, \tau, t) = \sum_{i=1}^{4} f_i(\theta, t) \ I_i(t) + u_t, \quad (5.2)
\]
where \( f_i(\theta_i, t) = \sum_{j=0}^{3} \theta_{ij} t^j, I_i(t) = 1 \) if \( \tau_{i-1} \leq t < \tau_i; = 0 \) otherwise, \( \theta_{ij}'s \) are the coefficients of the polynomial function in the jth interval; and \( f(\theta, \tau, t) \) is continuous, but its first derivative is not continuous at the knot points. Then by the continuity constraints

\[
\begin{align*}
1. \quad & \theta_{10} + \theta_{11} \tau_1 + \theta_{12} \tau_1^2 + \theta_{13} \tau_1^3 = \theta_{20} + \theta_{21} \tau_1 + \theta_{22} \tau_1^2 + \theta_{23} \tau_1^3 \\
2. \quad & \theta_{20} + \theta_{21} \tau_2 + \theta_{22} \tau_2^2 + \theta_{23} \tau_2^3 = \theta_{30} + \theta_{31} \tau_2 + \theta_{32} \tau_2^2 + \theta_{33} \tau_2^3 \\
3. \quad & \theta_{30} + \theta_{31} \tau_3 + \theta_{32} \tau_3^2 + \theta_{33} \tau_3^3 = \theta_{40} + \theta_{41} \tau_3 + \theta_{42} \tau_3^2 + \theta_{43} \tau_3^3,
\end{align*}
\]

we can find that the \( \dot{\alpha}_i \)'s, \( \beta_{ij} \)'s and \( \dot{\theta}_{ij} \)'s have the following relationships :

\[
\begin{align*}
\dot{\theta}_{10} &= \dot{\alpha}_0 \\
\dot{\theta}_{11} &= \dot{\alpha}_1 \\
\dot{\theta}_{12} &= \dot{\alpha}_2 \\
\dot{\theta}_{13} &= \dot{\alpha}_3 \\
\dot{\theta}_{20} &= \dot{\alpha}_{11} - \beta_{11} \tau_1 + \beta_{12} \tau_1^2 - \beta_{13} \tau_1^3 \\
\dot{\theta}_{21} &= \dot{\alpha}_{11} + \beta_{11} - 2\beta_{12} \tau_1 + 3\beta_{13} \tau_1^2 \\
\dot{\theta}_{22} &= \dot{\alpha}_{12} + \beta_{12} - 3\beta_{13} \tau_1 \\
\dot{\theta}_{23} &= \dot{\alpha}_{13} + \beta_{13} \\
\dot{\theta}_{30} &= \dot{\alpha}_{21} + \beta_{21} \tau_2 + \beta_{22} \tau_2^2 - \beta_{23} \tau_2^3
\end{align*}
\]
\[ \hat{\theta}_{31} = \theta_{21} + \beta_{21} - 2\hat{\beta}_{22}\hat{\xi}_2 + 3\hat{\beta}_{23}\hat{\xi}_2^2 \]
\[ \hat{\theta}_{32} = \theta_{22} + \beta_{22} - 3\hat{\beta}_{23}\hat{\xi}_2 \]
\[ \hat{\theta}_{33} = \theta_{23} + \beta_{23} \]
\[ \hat{\theta}_{40} = \theta_{30} - \beta_{31}\hat{\xi}_3 + \hat{\theta}_{32}\hat{\xi}_3^2 - \hat{\beta}_{33}\hat{\xi}_3^3 \]
\[ \hat{\theta}_{41} = \theta_{31} + \beta_{11} - 2\hat{\beta}_{32}\hat{\xi}_3 + 3\hat{\beta}_{33}\hat{\xi}_3^2 \]
\[ \hat{\theta}_{42} = \theta_{32} + \beta_{22} - 3\hat{\beta}_{33}\hat{\xi}_3 \]
\[ \hat{\theta}_{43} = \theta_{33} + \beta_{33} \]

Therefore, we can obtain the following table:

Table 5.1: Estimates of $\hat{\theta}_{ij}$'s: Known Knots at reported Times of the Gait Events and i.i.d. Model

<table>
<thead>
<tr>
<th>i</th>
<th>j</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>1.8578</td>
<td>27.2275</td>
<td>130.9853</td>
<td>-334.962</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>0.9031</td>
<td>68.644</td>
<td>-52.6022</td>
<td>116.8623</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>424.625</td>
<td>-2058.724</td>
<td>3506.041</td>
<td>-1868.838</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>-127.162</td>
<td>512.355</td>
<td>-505.5432</td>
<td>224.2505</td>
</tr>
</tbody>
</table>

Note: i is interval, j is the degree of polynomial.
For intervals 2 to 4, the signs of the spline coefficients $\hat{\theta}_{ij}$'s within each interval alternates as degree increases, and the values of $\hat{\theta}_{ij}$'s in the quadratic and cubic terms ($j=2,3$) look wild. These indicate that the estimated mean values are offset if we add more truncated polynomial terms to intervals 2, 3 and 4, which is a sign that we overfitted the motion data. The reported times of the gait events $\tau_1 = 0.17308$, $\tau_2 = 0.53846$ and $\tau_3 = 0.69231$ given above and the values of $\hat{\theta}_{ij}$'s in Table 5.1, except $\hat{\theta}_{20}$, $\hat{\theta}_{30}$ and $\hat{\theta}_{40}$, will be used as the starting values in our generalized cubic spline regression model with unknown values and AR(1) errors.

Now, we consider the unknown knot points model. Using the values of $\hat{c}_1$, $\hat{c}_2$, $\hat{c}_3$ and $\hat{\theta}_{ij}$'s in Table 5.1 except $\hat{\theta}_{20}$, $\hat{\theta}_{30}$ and $\hat{\theta}_{40}$ as starting values, we can fit our motion data with unknown knots and i.i.d. error model. We obtain $\hat{\tau}_1 = 0.088$, $\hat{\tau}_2 = 0.5514$, $\hat{\tau}_3 = 0.7236$ and the following table:

<table>
<thead>
<tr>
<th>$i$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.9828</td>
<td>22.8846</td>
<td>64.9610</td>
<td>753.800</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>87.4346</td>
<td>-207.886</td>
<td>164.8867</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>-1972.38</td>
<td>3181.50</td>
<td>-1706.08</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>921.605</td>
<td>-1071.23</td>
<td>398.505</td>
</tr>
</tbody>
</table>

Table 5.2: Estimates of $\theta_{ij}$'s: Unknown Knot Points and i.i.d. Model

Note: $i$ is interval, $j$ is the degree of polynomial.
Hip Flexion-Extension -- Left Leg

Figure 5.2: Plot of Raw Data and Predicted Values -- Generalized Cubic Splines with Unknown Knots

1 = Raw Data, Dotted Curve = Predicted Values
Figure 5.3: Residual Plot -- Generalized Cubic Splines with Unknown Knots
Figure 5.4: Lag Plot of Residuals and Predicted Residuals -- I.I.D. Model

Residuals = 1, Predicted Residuals = Dotted Line
See Figure 5.2 for a plot of the fitted values and raw data. Figure 5.3 is the residual plot. From Figures 5.2 and 5.3, we can see that the fitted model is very good. However, there exists some cyclical pattern among the residuals, indicating that the i.i.d. assumption may be violated. Using the residuals $\hat{u}_i$ obtained from the unknown gait events and i.i.d. model to fit a first order autoregression model

$$u_{t_i} = \rho \ u_{t_i-1} + e_{t_i} \quad i = 2,3,4,\ldots,52,$$

we have $\hat{\rho} = 0.7746$ and $\hat{\sigma} = 0.1441013$, where $\hat{\rho}$ is an estimate of $\rho$. These two values will be used to compute the transformation matrix $P$, which is to be used in §5.3 to transform the AR(1) errors to i.i.d. errors. The fitted first order autoregression line is plotted in Figure 5.4, along with $u_t$, $i=2,3,\ldots,52$. From Figure 5.4, the AR(1) model is a reasonable assumption for the autocorrelated errors.

§5.3. The procedure used to find the estimates

In this section, we describe an approach that may be used to compute the estimates of regression coefficients and knot points. To be able to find the estimates easily, we adopt an approach different from the procedure we discussed in Chapter 2. The justification of this approach, which is asymptotically equivalent to our previous procedure, is given in §5.4.

In this approach, we first separate the basic structure of our code into five steps. The details will be discussed later. The SAS code is given in Appendix A. The five steps are given in Table 5.3.
Table 5.3: Flow chart of the SAS code

<table>
<thead>
<tr>
<th>STEP</th>
<th>Action</th>
</tr>
</thead>
</table>
| 1    | a. Use known knots and i.i.d. errors model to obtain $\hat{\theta}_0$.  
|      | b. SAS procedure: REG.  
|      | c. Goal: to obtain starting values for STEP 2 and STEP 5. |
| 2    | a. Use unknown knots and i.i.d. model.  
|      | b. SAS procedure: NLIN.  
|      | c. Goal: to obtain residual vector $\hat{u}$. |
| 3    | Use $\hat{u}$ to compute $\hat{r}$ and $\hat{\sigma}^2$, which are to be used to generate an $n \times 2$ transformation matrix $P$. |
| 4    | Transform the original model using $P$ so that we have i.i.d. errors. |
| 5    | Run NLIN on the transformed model with the starting values found in STEP 1. |

In Step 1, we use the SAS procedure REG to compute the regression coefficients, assuming that the positions of knot points are known and we have i.i.d. errors. The regression coefficients obtained at this step and labelled $\hat{\theta}_0$ are used as the starting values of NLIN in Steps 2 and 5. Ordinarily, the starting values chosen in this way lead to faster convergence in the NLIN procedure. In Step 2, we use the SAS procedure NLIN to find the estimated regression coefficients and to obtain the residual vector. The residual vector $\hat{u}$ is used to compute $\hat{r}$, the estimated autocorrelation, and $\hat{\sigma}^2$, the estimated variance of $\varepsilon_t$.

In Step 3, we compute an $n \times 2$ matrix $P$ which will be used to transform our original data so that we have i.i.d. errors. The matrix $P$ has the following form:
\[
P = \begin{pmatrix}
\hat{\theta} / \sqrt{\hat{\phi} (0)} & 0 \\
- \hat{\phi} & 1 \\
- \hat{\phi} & 1 \\
. & . \\
. & . \\
. & . \\
- \hat{\phi} & 1
\end{pmatrix}
\]

The formulas to compute \( \hat{\phi} \) and \( \hat{\phi}^2 \) are

\[
\hat{\phi} = \frac{\hat{\phi}(1)}{\hat{\phi}(0)} \quad \text{and} \quad \hat{\phi}^2 = \hat{\phi}(0) - \hat{\phi} \hat{\phi}(1),
\]

where \( \hat{\phi}(0) \) and \( \hat{\phi}(1) \) are defined as

\[
\hat{\phi}(0) = \frac{1}{n} \sum_{i=1}^{n} \hat{\phi}_i^2 \quad \text{and} \quad \hat{\phi}(1) = \frac{1}{n} \sum_{i=1}^{n-1} \hat{\phi}_i \hat{\phi}_{i+1}.
\]

The transformed values are computed in Step 4. Letting the transformed \( y_{t_i} \) be \( Y_{t_i} \), and the transformed \( f(\theta, \tau, t_i) \) be \( F(\theta, \tau, t_i) \), then we have

\[
Y_{t_i} = (P[i,2] * y_{t_i}) + (P[i,1] * y_{t_i-1})
\]

\[
F(\theta, \tau, t_i) = (P[i,2] * f(\theta, \tau, t_i)) + (P[i,1] * f(\theta, \tau, t_{i-1})).
\]

where \( P[i,1] \) is the \( (i,1) \)th entry of matrix \( P \) and \( P[i,2] \) is the \( (i,2) \)th entry of matrix \( P \). In this way, we obtain i.i.d. errors on the transformed model, as shown in §5.4, and hence the least squares method can be applied. In Step 5, we run NLIN on the transformed
model and use the $\hat{\theta}$ value found in Step 1 as the starting value. The algorithm DUD (doesn't use derivative) developed by Ralston and Jennrich (1978) is used in the step.

§5.4. Justification of the approach used in §5.3.

We establish in this section that the estimates found in §5.3 are asymptotically equivalent to the estimates found in Chapter 2. The steps that we will investigate include: the validation of the least squares method and $\sqrt{n}$-consistency of $\hat{\theta}$. Once these two results are established, then the estimates found in §5.3 are asymptotically equivalent to the estimates found by the procedure we stated in Chapter 2.

From §5.3, we know the transformed $y_{t_i-1}$ have the following relationship with $y_{t_i}$ and $y_{t_i-1}$:

$$Y_{t_i} = y_{t_i} - \rho y_{t_i-1}, \quad i = 2, 3, 4, ..., n. \quad (5.3)$$

Also,

$$F_{t_i} = F(\theta, \tau, t_i) = f(\theta, \tau, t_i) - \rho f(\theta, \tau, t_{i-1}). \quad (5.4)$$

Equations (5.3) and (5.4) imply

$$e_{t_i} = Y_{t_i} - F_{t_i} = y_{t_i} - \rho y_{t_i-1} - f(\theta, \tau, t_i) + \rho f(\theta, \tau, t_{i-1})$$

$$= y_{t_i} - f(\theta, \tau, t_i) - \rho \left( y_{t_i-1} - f(\theta, \tau, t_{i-1}) \right)$$

$$= u_{t_i} - \rho u_{t_{i-1}}$$
\[ = \varepsilon_{t_i} \quad i = 2, 3, 4, \ldots, n. \]

For \( i = 1 \),

\[ Y_{t_1} = y_{t_1} - \sqrt{\sigma^2 r(0)} = \sqrt{1 - \rho^2} y_{t_1} \]

\[ F(\theta, \tau, t_{1}) = \sqrt{1 - \rho^2} f(\theta, \tau, t_{1}). \]

Thus,

\[ e_{t_1} = Y_{t_1} - F_{t_1} = \sqrt{1 - \rho^2} \left( y_{t_1} - f(\theta, \tau, t_{1}) \right) \]

\[ = \sqrt{1 - \rho^2} u_{t_1}. \]

But

\[ E(u_{t_1}) = 0, \quad \text{Var}(u_{t_1}) = \frac{\sigma^2}{1 - \rho^2}. \]

Hence,

\[ E(e_{t_1}) = 0, \quad \text{Var}(e_{t_1}) = \sigma^2. \]

This means that \( e_{t_1} \) follows a \( N(0, \sigma^2) \) distribution. In addition, for \( i \neq 1 \),

\[ E\left( e_{t_1} e_{t_i} \right) = E\left\{ \sqrt{1 - \rho^2} u_{t_1} \left[ (y_{t_1} - \rho y_{t_{i-1}}) - (f(\theta, \tau, t_{i}) - \rho f(\theta, \tau, t_{i-1})) \right] \right\} \]
\[
\begin{align*}
E[\sqrt{1-\rho^2} u_{t_1} e_{t_1}] &= 0.
\end{align*}
\]

Therefore, the \(e_{t_i}\)'s are i.i.d. normal variables with mean 0 and variance \(\sigma^2\). This means that \(e = [e_{t_1}, e_{t_2}, e_{t_3}, \ldots, e_{t_n}]'\) and \(\Psi^{-1}u\) have the same distribution. Since now the \(e_{t_i}\)'s are i.i.d. errors, then the least-squares method can be used.

From Chapter 4, we know that if \(\hat{\varrho}\) is \(\sqrt{n}\)-consistent then \(\hat{\Theta}(\hat{\varrho})\) and \(\hat{\Theta}(\varrho^{(0)})\), and hence \(\hat{\theta}(\hat{\varrho})\) and \(\hat{\theta}(\varrho^{(0)})\), have the same asymptotic distribution whether we know the value of \(\varrho^{(0)}\) or not. If the value of \(\varrho^{(0)}\) is known, we know that the maximum likelihood estimators of \(\theta^{(0)}\) and \(\tau^{(0)}\) are equivalent to the least squares estimators of \(\theta^{(0)}\) and \(\tau^{(0)}\). Therefore, conditioned on \(\hat{\lambda}\), the maximum likelihood estimators of \(\theta^{(0)}\) and \(\tau^{(0)}\) are asymptotically equivalent to the least squares estimates of \(\theta^{(0)}\) and \(\tau^{(0)}\). So we conclude that if \(\hat{\lambda}\) in \$5.3\) is \(\sqrt{n}\)-consistent, the estimators found in \$5.3\) are asymptotically equivalent to the ones stated in Chapter 2.

To show that the estimator \(\hat{\lambda}\) in \$5.3\) is \(\sqrt{n}\)-consistent, we first notice that \(\hat{\Theta}\) and \(\hat{\lambda}\) are \(\sqrt{n}\)-consistent whether \(\varrho = 0\) or not. Although the autocorrelated error terms, \(u_{t_i}\), are misspecified as i.i.d. error terms, the estimators \(\hat{\Theta}\) and \(\hat{\lambda}\) are still \(\sqrt{n}\)-consistent even if the asymptotic covariance matrices of \(\hat{\Theta}\) and \(\hat{\lambda}\) under the i.i.d. error model is different from the asymptotic covariance matrices \(\Theta\) and \(\lambda\) under the AR(1) model. Therefore, we have

\[
\sqrt{n} (\hat{\mu} - \mu^{(0)}) = o_p(1).
\]
This implies that

\[ \hat{u}_{t_i} - u_{t_i} = o_p\left(\frac{1}{\sqrt{n}}\right) \]

or

\[ \hat{u}_{t_i} = u_{t_i} + o_p\left(\frac{1}{\sqrt{n}}\right). \]

Therefore,

\[ \hat{u}_{t_i}^2 = u_{t_i}^2 + 2\ u_{t_i} o_p\left(\frac{1}{\sqrt{n}}\right) + o_p\left(\frac{1}{n}\right) \]

and

\[ \sum_{i=1}^{n} \hat{u}_{t_i}^2 = \sum_{i=1}^{n} u_{t_i}^2 + 20\ o_p\left(\frac{1}{\sqrt{n}}\right) \sum_{i=1}^{n} u_{t_i} + o_p(1). \]

Similarly,

\[ \hat{u}_{t_i} \hat{u}_{t_{i+1}} = (u_{t_i} + o_p\left(\frac{1}{\sqrt{n}}\right))(u_{t_{i+1}} + o_p\left(\frac{1}{\sqrt{n}}\right)) \]

\[ = u_{t_i} u_{t_{i+1}} + (u_{t_i} + u_{t_{i+1}}) o_p\left(\frac{1}{\sqrt{n}}\right) + o_p\left(\frac{1}{n}\right). \]

Thus,

\[ \sum_{i=1}^{n-1} \hat{u}_{t_i} \hat{u}_{t_{i+1}} = \sum_{i=1}^{n-1} u_{t_i} u_{t_{i+1}} + o_p\left(\frac{1}{\sqrt{n}}\right) \sum_{i=1}^{n} (u_{t_i} + u_{t_{i+1}}) + o_p(1). \]

Now,
\[
\hat{\theta} = \frac{\sum_{i=1}^{n-1} \hat{u}_{t_i} \hat{u}_{t_{i+1}}}{\sum_{i=1}^{n} \delta_{t_i}^2}
\]

\[
= \frac{\sum_{i=1}^{n-1} u_{t_i} u_{t_{i+1}} + \mathcal{O}_p\left(\frac{1}{\sqrt{n}}\right)\sum_{i=1}^{n} (u_{t_i} + u_{t_{i+1}}) + \mathcal{O}_p(1)}{\sum_{i=1}^{n} u_{t_i}^2 + 2 \mathcal{O}_p\left(\frac{1}{\sqrt{n}}\right)\sum_{i=1}^{n} u_{t_i} + \mathcal{O}_p(1)}
\]

\[
= \frac{\sum_{i=1}^{n-1} u_{t_i} u_{t_{i+1}} + \mathcal{O}_p(1)}{\sum_{i=1}^{n} u_{t_i}^2 + \mathcal{O}_p(1)}.
\]

Therefore,

\[
\sqrt{n} \left( \hat{\theta} - \rho(0) \right) = \sqrt{n} \left( \frac{\sum_{i=1}^{n-1} u_{t_i} u_{t_{i+1}} + \mathcal{O}_p(1)}{\sum_{i=1}^{n} u_{t_i}^2 + \mathcal{O}_p(1)} - \rho(0) \right)
\]
\[
\sum_{i=1}^{n-1} u_{t_i} u_{t_{i+1}} + o_p(1) - \rho^{(o)} \sum_{i=1}^{n} u^2_{t_i} - \rho^{(o)} \sum_{i=1}^{n} u^2_{t_i} + o_p(1)
\]

\[
= \sqrt{n} \left( \frac{\sum_{i=1}^{n} u^2_{t_i} + o_p(1)}{\sum_{i=1}^{n} u^2_{t_i} + o_p(1)} \right)
\]

\[
\sum_{i=1}^{n-1} u_{t_i} u_{t_{i+1}} - \rho^{(o)} \sum_{i=1}^{n} u^2_{t_i} - o_p(1) - \sum_{i=1}^{n} u^2_{t_i} + o_p(1)
\]

\[
= \sqrt{n} \left( \frac{\sum_{i=1}^{n} u^2_{t_i} + o_p(1)}{\sum_{i=1}^{n} u^2_{t_i} + o_p(1)} \right) + \sqrt{n} \left( \frac{o_p(1)}{\sum_{i=1}^{n} u^2_{t_i} + o_p(1)} \right)
\]

\[
\sum_{i=1}^{n-1} u_{t_i} u_{t_{i+1}} - \rho^{(o)} \sum_{i=1}^{n} u^2_{t_i} - \sum_{i=1}^{n} u^2_{t_i} + o_p(1)
\]

\[
= \left[ \sqrt{n} \left( \frac{\sum_{i=1}^{n} u^2_{t_i} + o_p(1)}{\sum_{i=1}^{n} u^2_{t_i} + o_p(1)} \right) \right] + o_p(1).
\]

Now,
\[
\frac{0_p(1)}{\sum_{i=1}^{n} u_{t_i}^2} \xrightarrow{n \to \infty} 0, \quad \text{in probability}
\]

and

\[
\sqrt{n} \left( \frac{\sum_{i=1}^{n-1} u_{t_i} u_{t_{i+1}}}{\sum_{i=1}^{n} u_{t_i}^2} - \rho^{(o)} \right) = 0_p(1),
\]

so

\[
\sqrt{n} \left( \hat{\theta} - \rho^{(o)} \right) = 0_p(1).
\]

The justification for our computational procedure now is complete.

§5.5. Data analysis and the advantage of the approach used in §5.3.

In §5.5.1, we compare the results of fitting the model with unknown knots and AR(1) errors to the results of fitting the model with unknown knots and i.i.d. errors. The asymptotic correlation matrix of \( \hat{\theta} \) of our theoretical result is compared with the asymptotic correlation matrix of \( \hat{\theta} \) given in the SAS printout. The latter matrix is an approximation to Jennrich's result (1969). The advantage of the approach used in §5.3 is discussed in §5.5.2.
Figure 5.5: Residual Plot -- Unknown Knots with AR(1) Model and I.I.D. Model
Solid Curve with 1: I.I.D. Model, Dotted Curve with 2: AR(1) Model
Figure 5.6: Lag Plot of Residuals and Predicted Values -- AR(1) Model

Residuals = 1, Predicted Residuals = Dotted Line
§5.5.1 Data analysis and comparisons

First, we investigate the residual patterns from fitting generalized cubic spline model with unknown knots for the i.i.d. errors and AR(1) errors, respectively. From Figure 5.5, we see that the residuals from the AR(1) model still have a cyclical pattern across time, but the magnitude is smaller than for the i.i.d. model. However, the Durbin-Watson test statistic has a value of 0.469826, indicating that the residuals are still positively correlated. This point is supported by Figure 5.6. From Figure 5.6, we see that most of the points fall in the first and third quadrants. This suggests that a more complicated time series model, like ARMA(p,q), may need to be considered.

We know that the sum of the squared errors (SSE) for the AR(1) model must be smaller than that of the i.i.d. model. The SSE for the i.i.d. model is 2.3429, compared to 0.7425 for the AR(1) model. The estimated regression coefficients under both models, obtained from the SAS printout, are listed in Table 5.4.
Table 5.4: Estimates of Regression Coefficients and Asymptotic
Standard Errors of $\hat{\theta}_{ij}$'s: Unknown Knot Points

<table>
<thead>
<tr>
<th>Parameter</th>
<th>AR(1) model estimate (asym. std. error)</th>
<th>i.i.d model estimate (asym. std. error)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_{10}$</td>
<td>2.116 (0.2446)</td>
<td>1.9828 (2.1184)</td>
</tr>
<tr>
<td>$\theta_{11}$</td>
<td>13.224 (9.6818)</td>
<td>22.884 (168.4938)</td>
</tr>
<tr>
<td>$\theta_{12}$</td>
<td>322.664 (132.2025)</td>
<td>64.961 (3689.842)</td>
</tr>
<tr>
<td>$\theta_{13}$</td>
<td>-1078.027 (520.6867)</td>
<td>757.800 (26723.949)</td>
</tr>
<tr>
<td>$\theta_{21}$</td>
<td>75.821 (50.672)</td>
<td>87.435 (7.762)</td>
</tr>
<tr>
<td>$\theta_{22}$</td>
<td>-172.256 (122.512)</td>
<td>-207.886 (26.3344)</td>
</tr>
<tr>
<td>$\theta_{23}$</td>
<td>130.248 (97.620)</td>
<td>164.8867 (27.308)</td>
</tr>
<tr>
<td>$\theta_{31}$</td>
<td>-1724.246 (9249.025)</td>
<td>-1972.3818 (1144.043)</td>
</tr>
<tr>
<td>$\theta_{32}$</td>
<td>2682.126 (13584.39)</td>
<td>3181.504 (1807.577)</td>
</tr>
<tr>
<td>$\theta_{33}$</td>
<td>-1578.066 (6694.784)</td>
<td>-1706.086 (948.953)</td>
</tr>
<tr>
<td>$\theta_{41}$</td>
<td>289.065 (424.85)</td>
<td>921.605 (287.7654)</td>
</tr>
<tr>
<td>$\theta_{42}$</td>
<td>-335.353 (478.264)</td>
<td>-1071.235 (337.523)</td>
</tr>
<tr>
<td>$\theta_{43}$</td>
<td>114.598 (179.384)</td>
<td>398.505 (131.266)</td>
</tr>
</tbody>
</table>

The reasons that there is no $\hat{\theta}_{20}, \hat{\theta}_{30}, \hat{\theta}_{40}$ in the above table are that there are continuity constraints and we use $\tau_1, \tau_2$ and $\tau_3$ as free parameters. For the values of the $\hat{c}_i$'s, see Table 5.5. From above, we see that the estimates $\hat{\theta}_{11}, \hat{\theta}_{12}, \hat{\theta}_{13}, \hat{\theta}_{32}, \hat{\theta}_{41}, \hat{\theta}_{42}$ and $\hat{\theta}_{43}$ are dramatically changed from the i.i.d. model to the AR(1) model. In addition, the asymptotic standard errors for $\hat{\theta}_{11}, \hat{\theta}_{12}, \hat{\theta}_{13}, \hat{\theta}_{31}, \hat{\theta}_{32}, \hat{\theta}_{33}, \hat{\theta}_{41}, \hat{\theta}_{42}$ and $\hat{\theta}_{43}$ in the i.i.d. error model are very large. For the AR(1) model, the asymptotic standard errors for $\hat{\theta}_{ij}$'s, except for $\hat{\theta}_{10}$, are also very large. The instability of the regression coefficients and the large asymptotic standard errors indicate that the inverse of the
large asymptotic standard errors indicate that the inverse of the
covariance matrix of the estimated regression coefficients may be
nearly singular. Although the estimated regression coefficients and
asymptotic standard errors are computed from the approximation
to Jennrich's theoretical result, which assumes that the second
derivatives of the mean function are continuous in the unknown
parameters, these values still give us an indication of the possibility
of multicollinearity among the columns of the design matrix and the
redundancy of the estimated regression coefficients.

In fact, if we look at the plot of the raw data (Figure 5.2), we
can see that the response curve looks linear in the first region,
cubic in the second and the third intervals, and linear in the forth
interval. Thus the redundancy of $\hat{\theta}_{ij}$'s may be due to our overfitting
the angular motion data.

The possibility of the redundancy of $\hat{\theta}_{ij}$'s is supported by the
investigation of the asymptotic correlation matrix. Before we
discuss the correlation matrix of the estimated regression
coefficients, we mention in advance that correlations involving $\hat{\theta}_{20}$,
$\hat{\theta}_{30}$ and $\hat{\theta}_{40}$ are not included in the correlation matrix since we
chose $\tau_1$, $\tau_2$ and $\tau_3$ as free parameters. The estimated asymptotic
correlation matrix of the regression coefficients from SAS is as
follows:
\[
\begin{bmatrix}
\theta_1 & \theta_2 & \theta_3 & \theta_4 \\
\theta_1 & R_1 & 0 & 0 & 0 \\
\theta_2 & 0 & R_2 & 0 & 0 \\
\theta_3 & 0 & 0 & R_3 & 0 \\
\theta_4 & 0 & 0 & 0 & R_4
\end{bmatrix}
\]

where \( \theta_1 = [\theta_{10}, \theta_{11}, \theta_{12}, \theta_{13}]', \theta_2 = [\theta_{21}, \theta_{22}, \theta_{23}]', \theta_3 = [\theta_{31}, \theta_{32}, \theta_{33}]', \theta_4 = [\theta_{41}, \theta_{42}, \theta_{43}]' \) and

\[
R_1 = \begin{bmatrix}
1 & -.78 & .63 & -.54 \\
-.78 & 1 & -.95 & .88 \\
.63 & -.95 & 1 & -.98 \\
-.54 & .88 & -.98 & 1
\end{bmatrix}
\quad R_2 = \begin{bmatrix}
1 & -1 & .99 \\
-1 & 1 & -1 \\
.99 & -1 & 1
\end{bmatrix}
\]

\[
R_3 = \begin{bmatrix}
1 & -1 & 1 \\
-1 & 1 & -1 \\
1 & -1 & 1
\end{bmatrix}
\quad \text{and} \quad R_4 = \begin{bmatrix}
1 & -1 & 1 \\
-1 & 1 & -1 \\
1 & -1 & 1
\end{bmatrix}
\]

We see that the \( \hat{\theta}_{ij} \)'s are highly correlated for fixed interval \( i \), and have zero correlation for different \( i \), \( i=1,2,3,4 \). The high correlations among the \( \hat{\theta}_{ij} \)'s for fixed \( i \) support the conjecture that some of the \( \hat{\theta}_{ij} \)'s are redundant.

Since the variable time fraction is generated with equal spacing, the matrix \( G(\rho) \), defined in assumption 7, Chapter 2, has a
form similar to the one we found in the example section §3.4. The $(i,j)$th entry of $G(\rho)$, after some algebra, can be found to be

\[
G_{ij}(\rho) = \begin{cases} 
\frac{(1-\rho^2) \tau_1^{i+j-1}}{(i+j-1)} & \text{if } i,j \in \{1,2,3,4\} \\
\frac{(1-\rho^2)(\tau_2^{i+j-9} - \tau_1^{i+j-9})}{(i+j-9)} & \text{if } i,j \in \{5,6,7,8\} \\
\frac{(1-\rho^2)(\tau_3^{i+j-17} - \tau_2^{i+j-17})}{(i+j-17)} & \text{if } i,j \in \{9,10,11,12\} \\
\frac{(1-\rho^2)(1-\tau_3^{i+j-25})}{(i+j-25)} & \text{if } i,j \in \{13,14,15,16\}
\end{cases}, \quad (5.6)
\]

and $G_{ij}(\rho) = 0$, otherwise. Since $\rho$ and $\sigma^2$ have nothing to do with the computation of the theoretical correlation matrix, we only need to estimate $\tau$. Using the $\hat{\tau} = [0.173557, 0.556328, 0.692307]$ obtained from the AR(1) model as the estimate of $\tau$, the estimated theoretical correlation matrix is

\[
\begin{bmatrix}
\theta_1 & \theta_2 & \theta_3 & \theta_4 \\
\theta_1 & R_1^0 & 0 & 0 & 0 \\
\theta_2 & 0 & R_2^0 & 0 & 0 \\
\theta_3 & 0 & 0 & R_3^0 & 0 \\
\theta_4 & 0 & 0 & 0 & R_4^0
\end{bmatrix}, \quad (5.7)
\]
where $\theta_1$, $\theta_2$, $\theta_3$ and $\theta_4$ are defined as above; and the $R_i^0$'s are

$$
R_1^0 = \begin{pmatrix}
1 & -.8941 & .8078 & -.7498 \\
-.8941 & 1 & -.9794 & .9467 \\
.8078 & -.9794 & 1 & -.9914 \\
-.7498 & .9467 & -.9914 & 1
\end{pmatrix}
$$

$$
R_2^0 = \begin{pmatrix}
1 & -.9934 & .9778 \\
-.9934 & 1 & -.9950 \\
.9778 & -.9950 & 1
\end{pmatrix}
$$

$$
R_3^0 = \begin{pmatrix}
1 & -1 & .9296 \\
-1 & 1 & -.9759 \\
.9296 & -.9759 & 1
\end{pmatrix}
$$

and

$$
R_4^0 = \begin{pmatrix}
1 & -.9995 & .9980 \\
-.9995 & 1 & -.9995 \\
.9980 & -.9995 & 1
\end{pmatrix}
$$

This correlation matrix from our theoretical result has a similar pattern to the estimated correlation matrix given in the SAS printout. The matrices $R_i$ and $R_i^0$, $i=2,3,4$, are very close, and the $R_1$ and $R_1^0$ are not far apart. The closeness of $R_i$ and $R_i^0$ suggests that the correlation matrix in (5.5) is a good approximation to the correlation matrix in (5.7). However, we should be alerted by the high correlations among the $\hat{\theta}_{ij}$'s in the estimated theoretical correlation matrix. When we computed the matrix $G^{-1}(\rho)$, the computer gave a warning of the singularity of the matrix $G(\rho)$, which theoretically is nonsingular. If we examine the formula for calculating $G_{ij}(\rho)$ in (5.6), we can see that the large exponents of the $\tau_i$'s and the closeness of $\tau_i$ and $\tau_{i+1}$ suggest that the correlation matrix is ill-conditioned.

It is interesting to compare the estimates of $\tau$ for different models. The estimates $\hat{\tau}$ of $\tau$ and asymptotic standard errors for unknown knots with i.i.d. errors and unknown knots with AR(1) errors are listed in the following table:
Table 5.5: Values of $\hat{t}$ for Different Models

<table>
<thead>
<tr>
<th>$\hat{t}_1$</th>
<th>Reported Knot Points</th>
<th>Unknown Knot Points</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>I.I.D. Model</td>
</tr>
<tr>
<td>$\hat{t}_1$</td>
<td>0.17308</td>
<td>0.08878 (7.2106)</td>
</tr>
<tr>
<td>$\hat{t}_2$</td>
<td>0.53846</td>
<td>0.55143 (0.01365)</td>
</tr>
<tr>
<td>$\hat{t}_3$</td>
<td>0.69231</td>
<td>0.72359 (0.00852)</td>
</tr>
</tbody>
</table>

Note: Values in Parentheses are Estimated Asymptotic Standard Errors.

From the above table, we can see that for the i.i.d. model and the AR(1) model, the values of $\hat{t}_1$ are quite different; the values of $\hat{t}_2$ and $\hat{t}_3$, however, are not too far apart. Compared to the reported gait times $\hat{t} = [0.17356,0.55633,0.69231]$, we see that the estimates obtained from the AR(1) model are much closer to the given positions of the gait events than those obtained from the i.i.d. model. The finding of little discrepancy between the reported times and those from AR(1) model is interesting. It seems to say that the reported times of gait events are reliable and hence, for easy of calculation, the general linear model with AR(1) errors can be adopted to analyze the motion data. However, if we look at the asymptotic standard errors of $\hat{t}_1$ and $\hat{t}_2$ for the AR(1) model, we should be cautious because the asymptotic standard errors of these two estimates are large. For the i.i.d. model, the values of $\hat{t}_2$ and $\hat{t}_3$ are not too far away from the reported times, but $\hat{t}_1$ is quite different and its asymptotic standard error is weird, too.

In §5.2, we mentioned that it is the biasedness of the measurement error model that leads us to consider the unknown
functions for i.i.d. errors with known knot points, i.i.d. errors with unknown knot points and AR(1) errors with unknown knot points model. From Figure 5.7, we can see that all the three models fit the angular motion data well, with unknown knot points and AR(1) errors model being the best one. Generally speaking, the fitted values for the three models have very little discrepancy in the first and fourth intervals. There are some differences in the second and the third intervals for the three estimated mean functions, but they are still small in magnitude.

Figure 5.8 is the standard errors of the predicted values across time for the three different models. The plots are obtained by considering the knot points given in Table 5.5 as the true values and using the general linear models to find the standard errors of the three different estimated mean functions. We know that for the simple linear model the variance of the predicted mean value \( \hat{Y}_k \) at \( t_k \) has the following form

\[
\text{Var}(\hat{Y}_k) = \sigma^2 \left[ \frac{1}{n} + \frac{(t_k - \bar{t})^2}{\sum_{i=1}^{n} (t_i - \bar{t})^2} \right],
\]

therefore the standard deviation will be larger if the design point \( t_k \) is further away from the central point \( \bar{t} \). Figure 5.8 reveals this fact even if the plots within each interval do not have parabolic shape. From Figure 5.8, we can see that the curve in each interval appears symmetric about the central point for each interval and has the tendency that if \( |t_k - \bar{t}| \) is larger, the standard error of \( \hat{Y}_k \) is larger. Notice that the largest standard errors occur at the knot points. However, we cannot say that the variability at the knot
Left Hip Flexion-Extension

Figure 5.7: Plot of Predicted Mean Functions

Dotted Curve = Known Knots with I.I.D. Errors, Raw Data = 1
Dashed Curve = Unknown Knots with I.I.D. Errors
Immediately Dashed Curve = Unknown Knots with AR(1) Errors
Figure 5.8: Plot of Standard Deviations of $\hat{Y}$ for Different Models

- **Solid Curve** = I.I.D. Model with Known Knots
- **Dotted Curve** = I.I.D. Model with Unknown Knots
- **Dashed Curve** = AR(1) Model with Unknown Knots
points is larger than other points, because we cannot identify the large variation is due to the knot points or due to our considering these knot points as the known end points for the general linear models.

If we look at the range of the standard errors of the predicted values for the three different models, we can see that all of them are less than 0.24. This indicates that the estimation of the mean function is quite accurate, even though the asymptotic standard errors of most of the regression coefficients are very large. Although these standard errors of the predicted mean values are obtained from the "conditional" general linear models, which are not the true models except the i.i.d. errors with known knots, the plots in Figure 5.8 still indicate that the standard errors for the estimators of the mean functions are small, not like those of the estimated regression coefficients. Also, the standard errors of the predicted values for the unknown knot points with AR(1) model are uniformly smaller than those of the known or unknown knot points with i.i.d. errors except at its first knot point $\ell_1$. This shows that the unknown knot points with AR(1) model is the most accurate one among the three models considered.

From the above discussion, we conclude that unknown knot points with AR(1) model has the best fit to the mean function. However, we keep in mind that this model overfits the angular motion data. Overfitting the motion data causes high correlations among the $\theta_{ij}$'s; and as suggested above, a linear-cubic-cubic-linear model with discontinuous derivatives at the gait events may be more reasonable. Also, more complicated model, like AR(2) or ARMA(p,q) in the error terms need to be adopted. Since our theoretical results derived in Chapters 3 and 4 are suitable for models in which the polynomials in each interval have the same degree, we will not do further analysis here. However, we believe that our theoretical results can also be extended to the linear-cubic-cubic-linear model with unknown knots and AR(1) errors. As to
more generalized spline regression functions with AR(2) or ARMA(p,q) models, the topics still need to be explored.

§5.5.2. Advantage of the approach used in §5.3.

The computational method given in §5.3 is easier to use than the procedure we discussed in Chapter 2. We use the conditional least squares method to compute the least squares estimates. Since the estimated autocorrelation coefficient $\hat{\rho}$ is $\sqrt{n}$-consistent, we can use the SAS procedure NLIN to find the l.s.e. once the $n \times 2$ matrix $P$ is available. We do not need to write program to find the estimates of the parameters of interest if a package such as SAS is available. This advantage makes our approach very attractive.
Chapter VI
Conclusions and Suggestions

This chapter includes two sections. In §6.1, we briefly discuss
the results obtained in chapters 3 and 4 and the findings in the
data analysis chapter. Comparisons to the general linear model
with AR(1) errors are made and an investigation of the limiting
covariance matrix of $\sqrt{n} (\hat{\theta} - \theta^{(o)})$ is presented in the section also.
The discussion in the data analysis chapter suggests that a linear-
cubic-cubic-linear model with more complicated errors may need to
be considered. Some problems not explored in the dissertation are
stated in §6.2. Some of the problems mentioned are future research
questions.

§6.1. Conclusions

It is well-known that the asymptotic distributions of
$\sqrt{n} (\hat{\theta} - \theta^{(o)})$ and $\sqrt{n} (\hat{d}^{2} - \sigma_{o}^{2})$ in the general linear model with i.i.d.
errors are normal. The asymptotic normality holds even if the i.i.d.
error term assumption is replaced by the assumption of AR(1)
errors for the general linear model, see Hildreth (1969). The
asymptotic distribution of $\sqrt{n} (\hat{\theta} - \theta^{(o)})$ for a smooth nonlinear
model with i.i.d. errors or correlated errors had also been derived
previously. Feder (1975) considered the non-smooth nonlinear
model (segmented regression model with unknown knots) with i.i.d.
errors and derived the asymptotic distributions fo $\sqrt{n} (\hat{\theta} - \theta^{(o)})$ and
knot vector $\sqrt{n} (\hat{\tau} - \tau^{(o)})$.

The asymptotic distributions of the estimated regression
coefficients, free knots, autoregressive parameter and variance for
the polynomial spline regression model with unknown knots and AR(1) errors were investigated in the dissertation. To avoid complexity in deriving the asymptotic distributions for the important estimators, we assume that \( \theta^{(o)} \) is an interior point of \( \Omega \). For the case that \( \rho^{(o)} \) is known, we found, under one more technical condition than the usual linear model, that the asymptotic distribution of \( \sqrt{n} (\hat{\theta} - \theta^{(o)}) \) is similar to that derived from the general linear model with AR(1) errors. Due to the continuity restrictions on the mean function, the asymptotic distribution of \( \sqrt{n} (\hat{\tau} - \tau^{(o)}) \) can be obtained from \( \sqrt{n} (\hat{\theta} - \theta^{(o)}) \), since there is a linear relationship between \( \theta^{(o)} \) and \( \tau^{(o)} \). We should not be surprised by the result that the asymptotic distribution of \( \sqrt{n} (\delta^2 - \sigma_0^2) \) in our model is the same as that in the general linear model with AR(1) errors, because the normal sample mean and sample variance are Independent.

If the autocorrelation \( \rho^{(o)} \) is unknown, the asymptotic distributions for \( \sqrt{n} (\hat{\theta} - \theta^{(o)}) \) and \( \sqrt{n} (\hat{\tau} - \tau^{(o)}) \) are still the same as in the case that \( \rho^{(o)} \) is known. Since \( \rho^{(o)} \) does not appear in the mean function, but in the disturbance terms, it seems very natural that the asymptotic behavior of \( \sqrt{n} (\hat{\theta} - \theta^{(o)}) \) and \( \sqrt{n} (\hat{\tau} - \tau^{(o)}) \) should not be affected by whether the value of \( \rho^{(o)} \) is known or not. The asymptotic distribution of \( \sqrt{n} (\delta - \rho^{(o)}) \) is found to be \( N(0, 1 - (\rho^{(o)})^2) \), which is the same as the asymptotic distribution of \( \sqrt{n} (\delta - \rho^{(o)}) \) if we have the general linear model with AR(1) errors. The asymptotic distribution of \( \sqrt{n} (\delta^2 - \sigma_0^2) \) is not affected by whether or not we know the value of \( \rho^{(o)} \), either. We have the same limiting distribution for \( \sqrt{n} (\delta^2 - \sigma_0^2) \) for both of the two situations.

We know, from above, that the asymptotic distributions of parameters in the polynomial spline regression with unknown
knots and AR(1) errors are similar to those obtained in the general linear model with AR(1) errors. The appearance of the unknown knot vector \( \tau \) in our model, in fact, does not affect the asymptotic behavior of \( \hat{\theta} \) much. The form of the limiting covariance matrix of \( \sqrt{n} (\hat{\theta} - \theta^{(0)}) \) depends on the unknown knot vector \( \tau \) and how \( t_{ni} \)'s are generated. As illustrated in the example in §3.4, if the independent variable \( t \) is generated from a \( U[0,1] \) distribution, then the limiting covariance matrix \( G^{-1}(\rho) \) can be simplified to \( (1-\rho)^{-2} G^{-1} \), where \( G^{-1} \) is the limiting covariance matrix of \( \sqrt{n} (\hat{\theta} - \theta^{(0)}) \) in the case that we have i.i.d. errors.

In Chapter 5, we used the generalized cubic spline regression model with unknown knots and AR(1) errors to fit motion data collected at Children's Hospital in San Diego. Although the fit is satisfactory, we found that there is a serious highly correlated problem among the estimated regression coefficients, which means that we may have overfit the motion data. This point is supported by examining the plot of the raw data. Although most of the asymptotic standard errors of the estimated regression coefficients are very large, the standard error of the predicted mean function is small. This says that the estimation of the mean function is accurate. Since the curve of the raw data looks like linear-cubic-cubic-linear, the deterministic part in model (5.2) can be replaced by a linear-cubic-cubic-linear model with unknown knots. For the stochastic part of model (5.2), we know, from Figure 5.5, that there still exists cyclical pattern for the residuals even the magnitude is smaller. The Dubin-Watson statistic with value 0.46986 suggests that the residuals are still positively related. An AR(\( p \)) or ARMA(\( p,q \)) but AR(1) error model, therefore, is worthwhile to fit the motion data. We did not try to refit the suggested model, because it is beyond the scope of our theoretical work.

§6.2. Suggestions for future work
From the derivation of the asymptotic distributions and data analysis chapter, we see that several problems still are unsolved and need further research:

1. We can extend our model to allow different values of \( \rho \) in different regions, as considered by Tsuruni and Ilmakunnas (1984). This model then is more reasonable if we want to consider some phenomena with different autocorrelations at different times.

2. We considered an AR(1) time series model for the disturbance terms. As discussed in §6.1, more complicated models, like ARMA(p,q), may be considered to give a more meaningful fit. However, we expect that the manipulation of the m.l.e.'s (or l.s.e.'s) with ARMA(p,q) errors will be more difficult than our current model. In this respect, the approach given in §5.3 or Pierce's (1971) conditional m.l.e. method can be adopted to reduce the complexity.

3. The asymptotic distributions of \( \hat{\theta} \), \( \hat{\phi} \), and \( \delta^2 \) in our model are the same whether or not we know the value of the autocorrelation \( \rho^{(o)} \). Also, \( \delta^2 \) has the same asymptotic distribution as it has in the general linear model. This leads us to speculate that the random variables \( \sqrt{n} (\hat{\theta} - \theta^{(o)}) \), \( \sqrt{n} (\hat{\rho} - \rho^{(o)}) \) and \( \sqrt{n} (\delta^2 - \sigma^2_0) \) are asymptotically mutually independent, or furthermore, asymptotically jointly independent. To answer this question, however, we need to investigate the joint distribution of these random variables.

4. In the pseudo-problem, we delete \( n^{**} = o(n/\log \log n) \) points around \( \tau_i \). Recognizing that \( \log \log n = 3.96 \) if \( n = 10^{23} \), we should not expect that the number of deleted points is too small. Even if the pseudo-problem is
only for theoretical consideration and cannot be performed in practice, the deleting process and the usual considerations in asymptotic results still lead us to concern about how large n should be so that the asymptotic results obtained in chapters 3 and 4 ca be applied satisfactorily. Monte Carlo studies, which were not conducted in the dissertation, may give us the answer.

(5). In Chapter 5, we fit motion data with a generalized cubic spline model with unknown knots and AR(1) errors. We found that the data were overfitted and the residuals are still positively correlated. It was suggested in Chapter 5 that a linear-cubic-cubic-linear model might be more suitable for the mean function. However, to choose the best model we need to use model selection techniques. Also, since the residuals are still positively correlated, it is possible that an AR(p) or ARMA(p,q) model may need to be considered for the stochastic term.

(6). We showed in §4.2 that a root \( \hat{\rho} \) of the likelihood equations is a \( \sqrt{n} \)-consistent estimator of \( \rho^{(o)} \) and asymptotically normal. The \( \hat{\rho} \) we found is not necessarily the m.l.e, it is only one of the possible candidates. In this dissertation we did not prove that the \( \hat{\rho} \) obtained is the true m.l.e.

(7). In §3.4, we discussed the fact that if the \( t_{ni} \)'s are generated from H(\( \cdot \)), where H(\( \cdot \)) is uniformly distributed over \([0,1]\), then the limiting covariance matrix \( G^{-1}(\rho) \) can be written as \( (1-\rho)^{-2}G^{-1} \). In fact, assuming that \( t_{ni} \)'s are generated from uniform distribution is a relatively strong assumption, and hence restricts the application of the asymptotic results. It is conjectured that the assumption above can be weakened to the requirement that H(\( \cdot \)) is a strictly increasing continuous function of t.
such that

$$\sup_{1 \leq i \leq n-1} \left| H(t_{n(i+1)}) - H(t_{ni}) \right| \to 0$$

as \( n \to \infty \) and we still can obtain the reduced form \((1-p)^2G^{-1}\).

In this dissertation, we assume that the autoregressive parameter \( p \) is the same at different times. However, in some cases the autoregressive parameter is different in different time intervals. For this situation, AR(1) error model discussed in Ilmakunnas and Tsuruni (1984) can be adopted. Also, from the data analysis chapter, we know that more complicated model, whether or not in the deterministic term like linear-cubic-cubic-linear mean function or in the stochastic part like ARMA(p,q) errors, need to be used to obtain better results. We will be eager to see that the above theoretical and practical problems can be tackled in the near future.
Appendices

Appendix A: SAS code for data analysis.

options linesize=80;
DATA sdcsmot;
   infile motion;
   input frame gaitev $ lhipfe t;
PROC NLIN method = DUD iter = 100 convergence = 1.E-8;
parms theta01 = 1.8578 theta11=27.2275 theta21=130.9853
   theta31=-334.962 theta12=68.644 theta22=-52.6022
   theta32=116.8623 theta13=-2058.7244 theta23 =3506.041
   theta33 = -1868.3877 theta14 = 512.355 theta24 = -
505.5432
   theta34 = 224.2505 tau1 = 0.17308 tau2 = 0.53846 tau3 =
0.69231;
* The above starting values are obtained by using SMITH's method;
* The following are the CONTINUITY and SMOOTHNESS restraints;

   theta02 =theta01 + (theta11-theta12) * tau1 + (theta21-
   theta22)*tau1**2
   + (theta31 - theta32) * tau1**3;
   theta03 =theta02 + (theta12-theta13)*tau2 + (theta22-theta23)*
   tau2**2
   + (theta32 - theta33) * tau2**3;
   theta04 = theta03+(theta13-theta14)*tau3 + (theta23-theta24) *
   tau3**2
   + (theta33 - theta34) * tau3**3;

IF t < tau1 THEN DO;    * FIRST part of the model;
   MODEL lhipfe = theta01 + theta11 * t + theta21 * (t**2) + theta31 *
   (t**3);
   DER.theta01 = 1;
DER.theta11 = t;
DER.theta21 = t**2;
DER.theta31 = t**3;
END;

ELSE IF t < tau2 then DO;  * Second part of the model;
MODEL lhipfe = theta02 + theta12 * t + theta22 * (t**2) + theta32 * (t**3);
   DER.theta02 = 1;
   DER.theta12 = t;
   DER.theta22 = t**2;
   DER.theta32 = t**3;
END;

ELSE IF t < tau3 then DO;  * Third part of the model;
MODEL lhipfe = theta03 + theta13 * t + theta23 * (t**2) + theta33 * (t**3);
   DER.theta03 = 1;
   DER.theta13 = t;
   DER.theta23 = t**2;
   DER.theta33 = t**3;
END;

ELSE DO;  * Fourth part of the model;
MODEL lhipfe = theta04 + theta14 * t + theta24 * (t**2) + theta34 * (t**3);
   DER.theta04 = 1;
   DER.theta14 = t;
   DER.theta24 = t**2;
   DER.theta34 = t**3;
END;

output out = work1 residual = uhat;

RUN;

data res1;
   set work1;
keep uhat ;

proc print ;

data work2 ;
    set work1 ;
    keep uhat ;
    output ;
    IF _N_= 52 then do ;
        uhat=0 ;
        output ;
    END ;

data work3 ;
    set work2 ;
    uhat0=uhat ;
    uhat1=LAG1(uhat) ;
    IF _N_=1 then do ;
        uhat1=0 ;
    END ;
RUN ;

data temp1 ;
    set work3 ;
    uhat00 = (uhat0)**2 ;
    uhat11 = (uhat1)**2 ;
    uhat01 = uhat0 * uhat1 ;
    one = 1 ;
    zero = 0 ;

proc summary data = temp1 ;
    var uhat00 uhat01 uhat11 ;
    output out = summm sum= suhat00 suhat01 suhat11 ;

data makesum ;
    set summm ;
    keep suhat00 suhat01 suhat11 ;
data temp2;
  merge temp1 makesum;

data temp3;
  set temp2;
  if suhat00 ne . then suhat00 = suhat00;
    else suhat00 = 0;
  if suhat01 ne . then suhat01 = suhat01;
    else suhat01 = 0;
  if suhat11 ne . then suhat11 = suhat11;
    else suhat11 = 0;

data temp4;
  set temp3;
  suhat001 = sum(suhat001,suhat000,0);
  suhat011 = sum(suhat011,suhat010,0);
  suhat111 = sum(suhat111,suhat110,0);
  retain suhat001 suhat011 suhat111 0;

data temp5;
  set temp4;
  gg = suhat111/52;
  g = suhat011/52;
  h = suhat001/52;

data temp6;
  set temp5;
  a = -g/(gg);

data temp7;
  set temp6;
  ss = h - (gg) * (a**2);
  k = ss/(gg);
  spq = sqrt(k);

data temp8;
  set temp7;
  if _N_ = 1 then co1 = spq;
else col = a ;
if _N_ = 1 then col2 = zero ;
else col2 = one ;

data work4 ;
set temp8 ;
if _N_ < 53 ;
keep col col2 ;
run;

data work5 ;
set sdchmot ;
lhipfe1=LAG1(lhipfe) ;
lhipfe2=lhipfe ;
x2 = t ;
x1 = LAG1(t) ;
IF _N_=2 then do ;
output ; output ;
END ;
IF _N_ > 1 then output ;
data work6 ;
merge work4 work5 ;
drop lhipfe t ;

PROC NLIN data=work6 method = DUD iter=100 convergence = 1.E-8 ;
params theta01=1.8578 theta11 = 27.2275 theta21=130.9853 
theta31=-334.962 tau1 = 0.17308 tau2 = 0.53846 tau3 = 
0.69231
theta12=68.644 theta22=-52.6022 theta32=116.8623
theta13=-2058.7244 theta23 = 3506.041 theta33=-
1868.838377
theta14=512.355 theta24=-505.5432 theta34=224.2505 ;

* The above starting values are obtained by using SMITH's method ;
* The following are the CONTINUITY restraints ;
\[
\begin{align*}
\text{theta02} &= \text{theta01} + (\text{theta11} - \text{theta12}) \cdot (\text{tau1}) + (\text{theta21} - \text{theta22}) \cdot (\text{tau1})^2 \\
&\quad + (\text{theta31} - \text{theta32}) \cdot (\text{tau1})^3 \\
\text{theta03} &= \text{theta02} + (\text{theta12} - \text{theta13}) \cdot (\text{tau2}) + (\text{theta22} - \text{theta23}) \cdot (\text{tau2})^2 \\
&\quad + (\text{theta32} - \text{theta33}) \cdot (\text{tau2})^3 \\
\text{theta04} &= \text{theta03} + (\text{theta13} - \text{theta14}) \cdot (\text{tau3}) + (\text{theta23} - \text{theta24}) \cdot (\text{tau3})^2 \\
&\quad + (\text{theta33} - \text{theta34}) \cdot (\text{tau3})^3 \\
\text{lhipfe} &= \text{col1} \cdot \text{lhipfe1} + \text{lhipfe2} \cdot \text{col2} \\
\end{align*}
\]

IF \( x2 < \text{tau1} \) THEN DO;  
   * FIRST part of the model;  
   MODEL lhipfe = \text{col1} \cdot (\text{theta01} + \text{theta11} \cdot x1 + \text{theta21} \cdot (x1)^2 + \text{theta31} \cdot (x1)^3) + \text{col2} \\
   (\text{theta01} + \text{theta11} \cdot x2 + \text{theta21} \cdot (x2)^2 + \text{theta31} \cdot (x2)^3) \\
   \text{DER.theta01} = \text{col1} + \text{col2} \\
   \text{DER.theta11} = \text{col1} \cdot x1 + \text{col2} \cdot x2 \\
   \text{DER.theta21} = \text{col1} \cdot (x1)^2 + \text{col2} \cdot (x2)^2 \\
   \text{DER.theta31} = \text{col1} \cdot (x1)^3 + \text{col2} \cdot (x2)^2 \\
END;

ELSE IF \( x2 < \text{tau2} \) then DO;  
   * Second part of the model;  
   MODEL lhipfe = \text{col1} \cdot (\text{theta02} + \text{theta12} \cdot x1 + \text{theta22} \cdot (x1)^2 \\
   \quad + \text{theta32} \cdot (x1)^3) + \text{col2} \cdot (\text{theta02} + \text{theta12} \cdot x2 \\
   \quad + \text{theta22} \cdot (x2)^2 + \text{theta32} \cdot (x2)^3) \\
   \text{DER.theta02} = \text{col1} + \text{col2} \\
   \text{DER.theta12} = \text{col1} \cdot x1 + \text{col2} \cdot x2 \\
   \text{DER.theta22} = \text{col1} \cdot (x1)^2 + \text{col2} \cdot (x2)^2 \\
   \text{DER.theta32} = \text{col1} \cdot (x1)^3 + \text{col2} \cdot (x2)^3 \\
END;

ELSE IF \( x2 < \text{tau3} \) then DO;  
   * Third part of the model;  
   MODEL lhipfe = \text{col1} \cdot (\text{theta03} + \text{theta13} \cdot x1 + \text{theta23} \cdot (x1)^2 \\
   \quad + \text{theta33} \cdot (x1)^3) + \text{col2} \cdot (\text{theta03} + \text{theta13} \cdot x2 + \text{theta23} \cdot (x2)^2 + \text{theta33} \cdot (x2)^3) \\
\]
DER.theta03 = co1 + co2;
DER.theta13 = co1 * x1 + co2 * x2;
DER.theta23 = co1 * x1**2 + co2 * x2**2;
DER.theta33 = co1 * x1**3 + co2 * x2**3;
END;

ELSE DO;
   * Fourth part of the model;
   MODEL lhipfe = co1 * (theta04 + theta14 * x1 + theta24 * (x1**2) + theta34 * (x1**3)) + co2 * (theta04 + theta14 * x2 + theta24 * x2**2 + theta34 * x2**3);
   DER.theta04 = co1 + co2;
   DER.theta14 = co1 * x1 + co2 * x2;
   DER.theta24 = co1 * x1**2 + co2 * x2**2;
   DER.theta34 = co1 * x1**3 + co2 * x2**3;
END;

RUN;
Appendix B: SAS code for selecting representative runs.

options linesize = 80 ;
DATA allcycle ;
   input gsnum runnum hosp $ vel stride lstep rstep cad wid cycle
   lhs rto
   rhs lto lhs2 ;
   onestan = 100 - lto + ( rhs - rto ) ;
   keep cad rto rhs lto runnum gsnum ;
cards ;
95881 1 N 224.55 3.18 1.72 1.47 141.17 0.25 1.52 0.00 7.84 47.06 58.82 100.00
95517 3 N 214.19 3.09 1.52 1.58 138.46 0.36 1.56 0.00 7.69 51.92 59.62 100.00
95517 1 N 208.97 3.08 1.51 1.58 135.85 0.21 1.62 0.00 5.66 50.94 56.60 100.00
95182 5 N 247.21 3.98 2.01 1.97 124.14 0.40 0.94 0.00 6.90 51.72 58.62 100.00
95182 4 N 258.79 4.17 2.14 2.02 124.14 0.46 0.90 0.00 3.45 48.28 55.17 100.00
94833 4 N 162.19 2.61 1.33 1.27 124.14 0.74 1.80 0.00 5.17 53.45 62.07 100.00
94833 3 N 225.06 2.81 1.32 1.48 160.00 0.60 1.38 0.00 6.67 48.89 57.78 100.00
94833 2 N 188.87 2.62 1.39 1.23 144.00 0.55 1.54 0.00 4.00 48.00 56.00 100.00
94750 3 N 226.84 3.65 2.03 1.63 124.14 0.47 1.76 0.00 10.34 50.00 60.34 100.00
94750 2 N 237.89 3.57 1.93 1.63 133.34 0.52 1.68 0.00 9.26 50.00 59.26 100.00
94750 1 N 183.90 3.22 1.84 1.38 114.28 0.43 1.94 0.00 12.70 49.21 66.67 100.00
94566 4 N 191.78 3.36 1.95 1.41 114.28 0.54 1.94 0.00 3.17 49.21 55.56 100.00
94566 3 N 219.50 3.54 2.00 1.54 124.14 0.61 1.80 0.00 3.45 51.72 56.90 100.00
94566 2 N 182.01 3.34 2.10 1.24 109.10 0.57 1.94 0.00 4.55 46.97 59.09 100.00
94031 3 N 174.32 3.05 1.44 1.61 114.29 0.62 1.96 0.00 7.94 49.21 57.14 100.00
94031 1 N 151.97 2.91 1.46 1.45 104.34 0.56 2.02 0.00 4.35 44.93 56.52 100.00
93936 1 N 4.07 0.16 0.50 49.65 0.93 4.42 0.00 29.66 47.59 48.97 100.00
93936 0 N 24.68 1.01 0.56 0.46 48.65 0.71 4.52 0.00 33.78 47.97 85.81 100.00
93662 5 N 222.34 3.64 1.86 1.78 122.03 0.29 1.72 0.00 13.56 57.63 62.71 100.00
93662 4 N 238.81 3.64 1.91 1.73 130.91 0.47 1.62 0.00 7.27 54.55 60.00 100.00
93662 3 N 238.63 3.77 1.98 1.80 126.31 0.29 1.78 0.00 7.02 54.39 59.65 100.00
93521 4 N 142.41 1.90 0.98 0.92 150.00 0.50 1.50 0.00 14.58 52.08 60.42 100.00
26685 22 N 249.05 4.29 2.04 2.25 116.13 0.26 1.90 0.00 4.84 46.77 51.61 100.00
;
PROC SORT data = allcycle ;
   by gsnum runnum ;
PROC STANDARD data=allcycle mean = 0 std = 1 out = standall ;
   *Stat V page 743;
   var cad rto rhs lto ;
       by gnum;

DATA distall ;
   set standall ;
   distan = cad**2 +rto**2 + rhs**2 + lto**2 ;
   keep distan runnum gnum ;

PROC SUMMARY data = distall ; * Basics page 535 ;
   var distan ;
       output out=distmin min =mindist ;
       by gnum ;

DATA allmin ;
   merge distall distmin ;
       by gnum ;

* The next data step picks one cycle
* See page 128 SAS Basics82, chapter7 (RETAIN) ;

DATA onerunnm ;
   set allmin;
       by gnum ;
           if mindist = distan ; if _FREQ_ NE 2 ;
           runtake = runnum ;
               keep gnum runtake runnum ;

*-----------------------------------------------
GSNUM and RUNNUM are present to MERGE BY
   RUNTAKE is the runnum to be used for person GSNUM ;
* The sort that follows should be redundant ;
* The steps below show how to pick the representative run ;
*Other data could replace ALLCYCLE as long as the dataset is sorted 
       by RUNNUM within GSNUM ;
PROC SORT data = onerunnm;
   by gsnum runnum;

PROC STANDARD data =allcycle out = temp  std = 1;
   var cad rto rhs lto;

PROC STANDARD data = temp out = tempcen mean = 0;
   var cad  rto rhs lto;
   by gsnum;

DATA onecycle;
   merge temp onerunnm;
   by gsnum runnum;
   if runnum = runtake;
   drop runtake;

PROC CORR cov nocorr data=onecycle out=covmat(type=cov);
   var cad rto rhs lto;

PROC MATRIX;
   fetch x data = tempcen (keep = cad rto rhs lto);
   xtran = x';
   fetch v 4 data= covmat; *CAUTION: covmat is not SQUARE!;
   a=solve (v, xtran); at = a';
   ax= (at) # x;
   mah = ax(+);
   output mah data = mah (rename=(col1=mah) );

DATA mahrun;
   merge temp mah;

PROC SUMMARY data = mahrun;
   var mah;
   output out = mahmin min = minmahd;
   by gsnum;
DATA mah3pick mah3pair ;
    merge mahrung mahmin ;
    by gsnum ;
    if _FREQ_ NE 2 then do ;
        if minmahd = mah ;
            runtake=runnum ; output mah3pick ;
        end ;
    else do ;
        drop minmahd ;
        output mah3pair ;
    end ;

PROC MATRIX ;
    fetch x data =mah3pair (keep = cad rto rhs lto ) ;
    fetch vplus data = covmat ; * CAUTION vplus is NOT square ;
    v= vplus(1:4,) ;
    xbar = vplus(7,);
    xcen = x - j(nrow(x),1) * xbar ;
    xcentr = xcen' ;
    a = solve ( v, xcentr ) ;
    ax = (a') # xcen ;
    mah = ax(+,+);
    PRINT x xcen v a ax mah ;
    output mah data = mahpairo(rename=(col1=mah)) ;

DATA mah3run ;
    merge mah3pair mahpairo ;

PROC SUMMARY data =mah3run ;
    var mah ;
    output out = mah3min min =minmahd ;
    by gsnum ;

DATA mahppick ;
    merge mah3run mah3min ;
    by gsnum ;
    if minmahd = mah ;
    runtake=runnum ;
keep gsnum runtake runnum ;

DATA runtake ;
  set mah3pick mahppick ;
  keep gsnum runtake runnum ; run;
List of References


models with ARMR errors to biological rhythm data. Statistics in Medicine. 6, 167-183.


Hudson, Derek J. (1966). Fitting segmented curves whose join points have to be estimated. J. Amer. Statist. Assoc. 61, 1097-1125.


Stein, Charles (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent


