2006

Penalized likelihood estimation of a fixed-effect and a mixed-effect transfer function model

Elizabeth Ann Hansen
University of Iowa

PENALIZED LIKELIHOOD ESTIMATION OF A FIXED-EFFECT AND A MIXED-EFFECT TRANSFER FUNCTION MODEL

by

Elizabeth Ann Hansen

An Abstract

Of a thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy
degree in Statistics
in the Graduate College of
The University of Iowa

July 2006

Thesis Supervisor: Professor Kung-Sik Chan

ABSTRACT

Motivated by the need of estimating the main spawning period of North Sea cod, we develop a common transfer function model with a panel of contemporaneously correlated times series data. This model incorporates (i) the smoothness on the parameters by assuming that the second differences are small and (ii) the contemporaneous correlation by assuming that the errors have a general variance-covariance matrix. Penalized likelihood estimation of this model requires an iterative procedure that is developed in this work. We develop three methods for determining confidence bands: frequentist, Bayesian, and bootstrap (both nonparametric and parametric). A simulation study on the frequentist and Bayesian confidence bands motivated by the cod spawning data is conducted and the results of those simulations are compared. The model is then used on the cod spawning data, with all confidence bands computed. The results of this analysis are discussed. We then delve further into our model by discussing the theory behind this model. We prove a theorem that shows that the estimated regression parameter vector is a consistent estimate of the true regression parameter. We further prove that this estimated regression parameter vector has an asymptotic normal distribution. Both theorems are proved while assuming mild conditions.

We further develop our model by incorporating between-series variation in the transfer function, with the random effect assumed to have a normal distribution with a smooth" mean vector. We implement the EM algorithm to do the penalized likelihood

estimation. We consider five different specifications of the variance-covariance matrix of the random transfer function model, namely, a general variance-covariance matrix, a diagonal matrix, a multiple of the identity matrix, an autoregressive matrix of order one, and a multiplicative error specification. Since the computation of confidence bands would lead to numerical problems, we introduce a bootstrap approach for estimating the confidence bands. We consider both the nonparametric and parametric bootstrap approaches. We then apply this model to estimate the cod spawning period, while also looking into the different specifications of the variance-covariance matrix of the random effect, the two types of bootstrapped confidence bands, and model checking.

Abstract approved: _____
Thesis Supervisor

_____
Title and Department

_____
Date

PENALIZED LIKELIHOOD ESTIMATION OF A FIXED-EFFECT AND A MIXED-EFFECT TRANSFER FUNCTION MODEL

by

Elizabeth Ann Hansen

A thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy
degree in Statistics
in the Graduate College of
The University of Iowa

July 2006

Thesis Supervisor: Professor Kung-Sik Chan

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph.D. thesis of

Elizabeth Ann Hansen

has been approved by the Examining Committee for the thesis requirement for the Doctor of Philosophy degree in Statistics at the July 2006 graduation.

Thesis committee: _____
Kung-Sik Chan, Thesis Supervisor

_____
Richard Dykstra

_____
Johannes Ledolter

_____
Osnat Stramer

_____
Jane Pendergast

ABSTRACT

Motivated by the need of estimating the main spawning period of North Sea cod, we develop a common transfer function model with a panel of contemporaneously correlated times series data. This model incorporates (i) the smoothness on the parameters by assuming that the second differences are small and (ii) the contemporaneous correlation by assuming that the errors have a general variance-covariance matrix. Penalized likelihood estimation of this model requires an iterative procedure that is developed in this work. We develop three methods for determining confidence bands: frequentist, Bayesian, and bootstrap (both nonparametric and parametric). A simulation study on the frequentist and Bayesian confidence bands motivated by the cod spawning data is conducted and the results of those simulations are compared. The model is then used on the cod spawning data, with all confidence bands computed. The results of this analysis are discussed. We then delve further into our model by discussing the theory behind this model. We prove a theorem that shows that the estimated regression parameter vector is a consistent estimate of the true regression parameter. We further prove that this estimated regression parameter vector has an asymptotic normal distribution. Both theorems are proved while assuming mild conditions.

We further develop our model by incorporating between-series variation in the transfer function, with the random effect assumed to have a normal distribution with a smooth" mean vector. We implement the EM algorithm to do the penalized likelihood

estimation. We consider five different specifications of the variance-covariance matrix of the random transfer function model, namely, a general variance-covariance matrix, a diagonal matrix, a multiple of the identity matrix, an autoregressive matrix of order one, and a multiplicative error specification. Since the computation of confidence bands would lead to numerical problems, we introduce a bootstrap approach for estimating the confidence bands. We consider both the nonparametric and parametric bootstrap approaches. We then apply this model to estimate the cod spawning period, while also looking into the different specifications of the variance-covariance matrix of the random effect, the two types of bootstrapped confidence bands, and model checking.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

CHAPTER 1
INTRODUCTION

Recent genetic analysis has suggested that the young (half-year old) cod sampled in some fjords in the Skagerrak, Norway, resembled adult cod in the North Sea in year 2001 but less so in year 2000 [13]. It was, furthermore, found that in 2001 when the sampled young cod of Skagerrak were genetically similar to the adult cod of North Sea, there was higher than average inflow of sea current from the North Sea to the Skagerrak, but not so in 2000 when the resemblance switched to the local adult cod. Thus, it was suggested the hypothesis that the North Sea cod stock might have contributed to the local cod population in the Skagerrak via transportation of cod eggs by sea current from North Sea into the Skagerrak. This hypothesis was tested using a long-term monitoring beach seine data on the annual counts of young cod, the (annual) spawning biomass of North Sea cod and daily inflow of sea current from North Sea to Skagerrak as pictured in Figures 1.1 and 1.2 [18]. It is believed that the cod spawn, or breed, in the months of March and April, but it is not known specifically when the majority of the spawning took place. The average daily inflow (from North Sea to the Skagerrak) was computed over several windows of two-week period between March and April, and tested the transportation hypothesis using a regression model with a covariate that is the product of average sea influx times log spawning biomass, a proxy for the transportable amount of cod eggs, the coefficient of which is non-zero under the transportation hypothesis and zero otherwise. It was found that the transportation hypothesis is consistent with the data, with stronger,

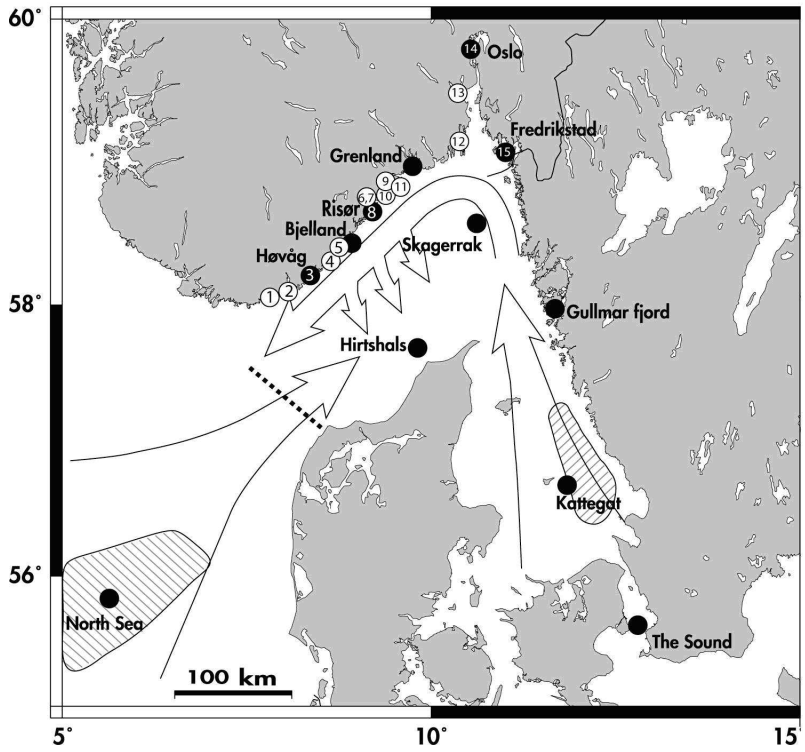

Figure 1.1: Map of Skagerrak area of North Sea. Fjords are denoted as numbered dots and currents as arrows coming into and out of the Skagerrak. Solid dots denote different bodies of water.

significant result when the mean inflow is computed over the second half of March. Clearly, which two-week period over which the mean inflow is computed is critical as the test can be made more powerful by aligning the period with the main period when the cod spawned.

Consider the following stochastic regression model describing how the response depends on the aggregate effects of a covariate:

$$Y_t = \alpha^T W_t + \sum_{j=m_1}^{m_2} \psi_j X_{t,j} + e_t, t = 1, 2, \cdots, T, \tag{1.1}$$

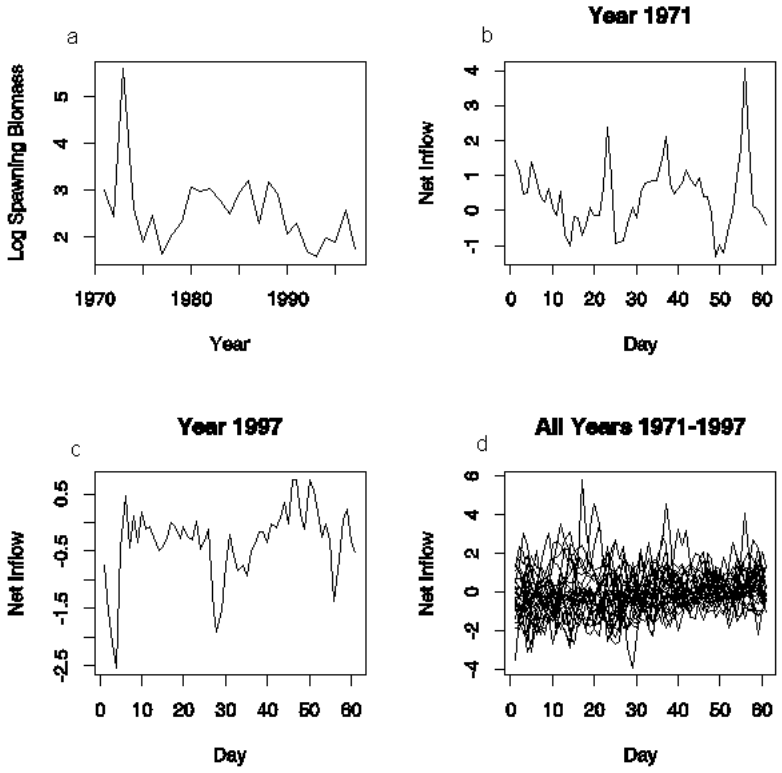

Figure 1.2: a) Time plot of the annual spawning biomass under the log transformation. b) Time plot of the daily net inflow into the Skagerrak in 1971. c) Time plot of the daily net inflow into the Skagerrak in 1997. d) Time plots of the daily net inflow into the Skagerrak for all years.

where Y_t are the responses, W_t and $X_{t,j}$ are vector-valued and scalar-valued covariates, α and ψ's are parameters and $\{e_t\}$ is a sequence of independent and identically distributed random variables of zero mean and finite variance; the superscript T denotes the transpose. The sampling scheme is as follows. Both Y and W are measured over regular, basic sampling intervals. Each basic sampling interval is further sub-divided into, say M, equal intervals over each of which X is measured. In the biological application to be discussed in section 3.3, Y and W were measured annually whereas X was measured daily. The measurement of X in the jth sub-interval of the tth basic sampling interval is denoted by $X_{t,j}$. The summation limits m_1 and m_2 are assumed to be known integers. The model defined by (1.1) is also known as a transfer-function model [3] or distributed-lag model [2].

The main interest is to estimate ψ_j as a function of j. Often, the functional form of ψ is unknown. Empirical parametric models such as the rational transfer function model and the Almon polynomial lag model, which are briefly reviewed in Chapter 2, are popular methods for estimating ψ, but they are less useful with complex functional forms. For example, in our biological application, ψ may be a multimodal function, in which case both the rational transfer function model and the Almon polynomial lag model require many parameters for providing an adequate description of ψ. Shiller introduces a nonparametric approach for estimating a smooth ψ function by postulating a smoothness prior on the second difference of ψ, but otherwise putting no constraints on ψ [17] [12]. (Shiller also discussed briefly the use of higher differences, but we shall not pursue this point here.) Specifically, it is

assumed that

$$(1-B)^2\psi_j = \eta_j, \tag{1.2}$$

where B is the backshift operator defined by $B\psi_j = \psi_{j-1}$, and η_j are independent and identically distributed normal random variables with zero mean and variance $\sigma_\eta^2 > 0$. That is, a hierarchical model is employed. Shiller discussed both the use of a fully Bayesian analysis with non-informative priors as well as a sort of empirical Bayes approach where σ_η^2 is specified by some rule of thumb. Note that the approach introduced by Shiller is similar to spline smoothing [20] [21] [9]. The nonparametric approach of Shiller can cope with complex functional form of ψ.

However, even the nonparametric approach fails if the number of data cases is small compared to $m_2 - m_1 + 1$, the number of lags of X appearing in the model. This problem may be circumvented if there exist a panel of time series that share the same transfer function so that information can be pooled across series for estimating ψ. Here, we consider this situation so that the sth series is generated by the model:

$$Y_{t,s} \;=\; \alpha^T W_{t,s} + \sum_{j=m_1}^{m_2} \psi_j X_{t,j} + e_{t,s}, t = 1, 2, \cdots, T, \tag{1.3}$$

where we note that the same X's enter into the equation for each component series, but W and e may vary across series. For the panel data, the errors are often contemporaneously correlated although they may be serially independent. Here, we "extend" Shiller's approach to a multivariate stochastic regression model with contemporaneously correlated errors that subsumes the common transfer function model

defined by (1.3). However, our approach differs from Shiller's approach in that we use a penalized likelihood approach [8]. In Chapter 3 we will discuss the fixed effect transfer function model, then present some asymptotic results based on this model. Theoretical results for the estimate for the regression parameter such as consistency and asymptotic distribution of the estimate are discussed. In this chapter we will discuss different methods for determining confidence bands for the parameter estimates. We will conduct simulations for some of these approaches to determine how well these confidence bands will detect the number of modes a distribution has. We will then apply this model to the biological data that was described above. Using that analysis, we have determined that there is a significant spawning in time span of March 1st and March 24th. We also determine that the mature cod spawned outside of the Skagerrak and the current transported the eggs into the Skagerrak.

The common transfer function model imposes a very strong assumption that the ψ_j's are fixed across series. In practice, this assumption may not hold. For example, in the biological application, the amount of cod eggs transported from North Sea into a fjord in the Skagerrak clearly depends on the geographical characteristics of a fjord, and hence a random transportation effect model is more plausible. We thus further extend the common transfer function model defined by (1.3) to a mixed-effect model that allows the ψ_j's to vary across series: we shall define this model by assuming that

$$Y_{t,s} = \alpha_s^T W_{t,s} + \sum_{j=m_1}^{m_2} \psi_{s,j} X_{t,j} + e_{t,s}, t = 1, 2, \cdots, T, \tag{1.4}$$

where we will assume that $\psi_s = \left[\alpha_s^T,\ \psi_{s,1},\ \ldots,\ \psi_{s,m_2-m_1+1}\right]^T \sim N(\phi, \Gamma)$, where $\phi = \left[\alpha^T,\ \phi_1,\ \ldots,\ \phi_{m_2-m_1+1}\right]^T$ is such that the ϕ_s's are assumed to form a smooth function. The EM algorithm can be implemented to estimate the common mean value of the ψ_s's and their between-series variation, as well as the variance-covariance matrix for the errors.

The between-series variation in the ψ_s's can be studied in a number of ways. We will consider five cases: a multiple of the identity matrix, a diagonal matrix, an AR(1) specification, a multiplicative error specification, or as a general covariance matrix which will be discussed in Section 4.4. Each specification has possible attractions and detractions. For example, the general covariance matrix is the most flexible specification, and, furthermore, may allow more smoothing to occur than if we were to use a multiple of the identity matrix or a diagonal matrix. However the general matrix has many more parameters to estimate, and thus may reduce the power to detect fine features in the mean function of the ψ_s's.

In order to determine confidence bands for the mean value of the ψ_s's, a bootstrap approach can be taken. We will implement both the nonparametric and parametric approaches to determining these confidence intervals.

In Chapter 4 we will elaborate the mixed-effect model and efforts to estimate this model and its standard error. These efforts include some different approaches to estimation under various specifications of the between-series variation in the ψ's. The

model is fitted to the biological data. The results, though less significant, lead us to believe that the spawning occurs at two different times during the months of March and April: one time period in the middle of March and one at the end of March. This is consistent with the beliefs of the biologists who drove the analysis. We will draw some conclusions in Chapter 5. Computer code using the R statistical programming language can be found in an appendix.

CHAPTER 2
PREVIOUS WORK

In situations in which the number of parameters in a regression model is large as compared to the number of observations, multicollinearity can be a problem. There have been several approaches that have been introduced in the past that can be used in these instances. We will discuss the Almon polynomials [2] in Section 2.1 and the transfer function model [3] in Section 2.2.

2.1 Almon polynomials

Consider the situation where the response variable Y_t can be defined as

$$Y_t = \sum_{i=0}^{\ell-1} w(i) X_{t-i} + \epsilon_t, \tag{2.1}$$

where ϵ_t are independent and identically distributed errors. Extension to correlated errors is straight forward. With relatively large ℓ, one specification of the $w(i)$ can be some low-degree polynomial of degree, say, $q+1$. Moreover, with a suitable parameterization of the polynomial coefficients, the model becomes a linear model. The required parameterization can be derived from the Lagrangian interpolation formula: let $w(x_i) = b_i$, where $x_i, i = 1, \ldots, q+1$ are $q+1$ distinct numbers. Then

$$w(i) = \sum_{j=0}^{q+1} \phi_j(i) b_j, \tag{2.2}$$

where $\phi_j(i)$ are Lagrangian interpolation polynomials, i.e.

$$\phi_j(x) = \frac{\prod_{\substack{i=1 \\ i\neq j}}^{q+1}(x - x_i)}{\prod_{\substack{i=1 \\ i\neq j}}^{q+1}(x_j - x_i)},\ j = 0, \ldots, q+1. \tag{2.3}$$

From (2.1) and (2.2), then we can write

$$Y_t = \sum_{j=0}^{q+1} b_j \sum_{i=0}^{\ell-1} \phi_j(i) X_{t-i} + \epsilon_t$$

So now, instead of estimating ℓ parameters, we would estimate $q+2$ parameters by regressing Y_t on Z_t, where $Z_t = [z_{t,0}\ z_{t,1}\ z_{t,2}\ \cdots\ z_{t,q+1}]$ and

$$z_{t,j} = \sum_{i=0}^{\ell-1} \phi_j(i) X_{t-i},\ j = 0, \ldots, q+1$$

In practice ℓ may be determined by subject matter considerations. However q is often unknown, although it can determined by minimizing the Akaike Information Criterion (AIC) [1] or by using some other criterion such as the Bayesian Information Criterion (BIC) [16].

2.2 The transfer function model

An ARMA model can be written as

$$Y_t = \mu + \frac{\Theta(B)}{\Phi(B)} a_t, \tag{2.4}$$

where Y_t is a random variable over time, μ is a constant mean value, B is the backshift operator, meaning that $BZ_t = Z_{t-1}$,

$$\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$$

where $\Theta(B)$ the moving average operator,

$$\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$$

which makes $\Phi(B)$ the autoregressive operator, and a_t is white noise. We can rewrite (2.4) as

$$\Phi(B)(Y_t - \mu) = \Theta(B)a_t.$$

Hence if we let $\gamma = \Phi(B)\mu = \mu - \phi_1\mu - \cdots \phi_p\mu$, then

$$\Phi(B)Z_t = \gamma + \Theta(B)a_t.$$

We can further incorporate an input series X_t and define a general transfer function model as

$$Y_t = \mu + \sum_{i=d}^{\infty} w(i)X_{t-i} + \frac{\Theta(B)}{\Phi(B)}a_t, \tag{2.5}$$

where d is an integer. If d is positive, then X_t has a delayed effect on Y_t. Note that if $\Theta(B) = \Phi(B) = 1$, $d = 0$ and $w(i) = 0$ for all $i \geq \ell$, then (2.5) becomes (2.1). Equation (2.5) can be rewritten as

$$Y_t = \mu + \sum_{i=d}^{\infty} w(i) B^i X_t + \frac{\Theta(B)}{\Phi(B)} a_t.$$

The rational transfer function model is obtained by letting $\sum_{i=d}^{\infty} w(i) B^i = \frac{\Omega(B)}{\Delta(B)}$, where $\Omega(B)$ and $\Delta(B)$ are polynomials of finite order. In particular, an AR(1) specification entails that $w(i) = w^i$ for some w. See Chapter 11 of Box et al. for a review of the rational transfer function model [3].

CHAPTER 3
A COMMON TRANSFER FUNCTION MODEL

In this chapter, we introduce a common transfer function model for modeling the common (lagged) effects of a covariate time series on a panel of response time series. To alleviate the multicollinearity problem, the common transfer function is assumed to be smooth in that its second differences are small. The smoothness of the common transfer function is controlled by a smoothing parameter. We consider in Section 3.1 the penalized likelihood approach to estimate the new model. The proposed estimation scheme is iterative. In Section 3.2 we study some methods of computing confidence bands for our model. In Section 3.3 we fit the model to the cod data for investigating the relationship between the number of mature cod and the mean inflow into the Skagerrak. We present in Section 3.4 some simulation results for our model and for some of the methods of computing confidence intervals. In Section 3.5 we derive some theoretical properties of the penalized likelihood. Consistency results for the regression parameter is found in Section 3.6, and their limiting distributional results in Section 3.7.

3.1 A multivariate stochastic regression model

Consider the following general regression model with multivariate response and covariate.

$$\mathbf{Y}_t = X_t\beta + \mathbf{e}_t;\ t = 1, \ldots, T \tag{3.1}$$

where the dimension of $\mathbf{Y}_t$ is $n \times 1$, X_t is $n \times k$, and the coefficient vector β is $k \times 1$. The $\mathbf{e}_t$'s are independent and identically distributed as normal with mean zero and variance-covariance matrix Ω, and $\mathbf{e}_t$ is independent of X_t. (Here, we restrict the errors to be normally distributed for convenience; extension to non-normality is straight-forward but it will complicate the iterative estimation procedure below.) This model is rather general and includes the common transfer-function model discussed in Chapter 1. Consider the case that the dimension of β is high compared to the sample size resulting in multicollinearity. The multicollinearity problem can be mitigated by exploiting some known "smoothness" property of β. Suppose that the roughness of β can be quantified by the Euclidean norm of $\eta = A\beta$ where A is a known $m \times k$ matrix. (In the case of the common transfer function model defined by (1.3), $A\beta$ is the vector of second differences of ψ.) We can now construct a penalized log-likelihood where a quadratic penalty for $A\beta$ is used:

$$\ell(\beta) = -\frac{T}{2}\log|\Omega| - \frac{1}{2}\sum_{t=1}^{T}(\mathbf{Y}_t - X_t\beta)^T\Omega^{-1}(\mathbf{Y}_t - X_t\beta) - \frac{1}{2}\frac{\beta^T A^T A\beta}{\sigma_\eta^2}, \qquad (3.2)$$

where the coefficient $\sigma_\eta^2 > 0$ quantifies the trade-off between badness of fit and roughness of the parameter; σ_η^2 will be determined by the method of cross-validation. Here, the penalized log-likelihood has the Bayesian interpretation that the components of η have joint prior independent and identical normal distribution of zero mean and variance σ_η^2.

Notice that if $\Omega = [\Omega_{ij}]$ and the smoothness parameter $\lambda = \frac{\Omega_{11}}{\sigma_\eta^2}$ are known, where Ω_{11} is the first element of Ω, β can be estimated by minimizing the following

expression.

$$\begin{aligned}
&\Omega_{11}\left\{\sum_{t=1}^{T}(\mathbf{Y}_t - X_t\beta)^T\Omega^{-1}(\mathbf{Y}_t - X_t\beta) + \frac{\beta^T A^T A\beta}{\sigma_\eta^2}\right\} \\
= &\sum_{t=1}^{T}\left\{\left(\frac{\Omega}{\Omega_{11}}\right)^{-\frac{1}{2}}\mathbf{Y}_t - \left(\frac{\Omega}{\Omega_{11}}\right)^{-\frac{1}{2}}X_t\beta\right\}^T\left\{\left(\frac{\Omega}{\Omega_{11}}\right)^{-\frac{1}{2}}\mathbf{Y}_t - \left(\frac{\Omega}{\Omega_{11}}\right)^{-\frac{1}{2}}X_t\beta\right\} \\
&+\lambda\beta^T A^T A\beta. \qquad (3.3)
\end{aligned}$$

The introduction of Ω_{11} in the definition of λ^2 is to make the latter interpretable as the signal-to-noise ratio; alternatively, we could use $\mathrm{tr}(\Omega)/n$, i.e. the average variance, instead of Ω_{11}, where for any square matrix A, $\mathrm{tr}(A)$ is the sum of diagonal elements of A. Assume for the moment that the smoothness parameter λ is known. Define $L = (\Omega/\Omega_{11})^{-\frac{1}{2}}$. Then the preceding penalized sum of least squares can be minimized by an iterative process. The iterative procedure bears resemblance to the method of seemingly unrelated regression technique [24] [10]. We first find $\widehat{\beta}^{(0)}$ by regressing $\mathbf{Y}_t$ on X_t. For $i = 0, 1, 2, \ldots$, compute $\widehat{\Omega}^{(i)} = \frac{1}{T}\sum_{t=1}^{T}(\mathbf{Y}_t - X_t^T\widehat{\beta}^{(i)})(\mathbf{Y}_t - X_t^T\widehat{\beta}^{(i)})^T$ and $\widehat{L}^{(i)}$ obtained via the Cholesky decomposition of $\widehat{\Omega}^{(i)}$. Define

$$\mathbf{Y}^{(i)} = \begin{pmatrix} \widehat{L}^{(i)}\mathbf{Y}_1 \\ \widehat{L}^{(i)}\mathbf{Y}_2 \\ \vdots \\ \widehat{L}^{(i)}\mathbf{Y}_T \\ \mathbf{0} \end{pmatrix}, X^{(i)} = \begin{pmatrix} \widehat{L}^{(i)}X_1 \\ \widehat{L}^{(i)}X_2 \\ \vdots \\ \widehat{L}^{(i)}X_T \\ \lambda A \end{pmatrix}.$$

Next, we update $\widehat{\beta}^{(i+1)}$ by regressing $\mathbf{Y}^{(i)}$ on $X^{(i)}$. The iterative procedure can be stopped by using some stopping criteria, e.g., when the relative change in the L^1-norm of the $\beta^{(i)}$ or the objective function defined in (3.2) is smaller than some prespecified

tolerance level. At the end of the iteration and letting $\widehat{\Omega}$ be the estimator of Ω, it can be readily checked that

$$\widehat{\beta} \;=\; (A^T A/\widehat{\sigma}_\eta^2 + \sum_t X_t^T \widehat{\Omega}^{-1} X_t)^{-1} \sum_t X_t^T \widehat{\Omega}^{-1} Y_t, \tag{3.4}$$

where $\widehat{\sigma}_\eta^2 = \lambda/\widehat{\Omega}_{11}$. Hence, the asymptotic covariance matrix of $\widehat{\beta}$ is approximately given by

$$(A^T A/\widehat{\sigma}_\eta^2 + \sum_t X_t^T \widehat{\Omega}^{-1} X_t)^{-1} (\sum_t X_t^T \widehat{\Omega}^{-1} X_t)(A^T A/\widehat{\sigma}_\eta^2 + \sum_t X_t^T \widehat{\Omega}^{-1} X_t)^{-1}. \tag{3.5}$$

The smoothness parameter λ can be determined by minimizing the cross-validation (CV) or the generalized cross-validation (GCV) [20] [8]. In order to visualize the calculation of CV or GCV, let $\mathbf{Z}$ be the vector obtained by stacking up the $\mathbf{Z}_t = \widehat{\Omega}^{-1/2}\mathbf{Y}_t$'s, $\tilde{\mathbf{X}}$ the corresponding design matrix found by stacking the $\widehat{\Omega}^{-1/2}\mathbf{X}_t$s, and $\widehat{\mathbf{Z}}$ the fitted values so that

$$\widehat{\mathbf{Z}} = \tilde{\mathbf{X}}\widehat{\beta} = H\mathbf{Z}, \tag{3.6}$$

where the matrix $H = [h_{ij}]$ is the hat matrix equal to $\tilde{\mathbf{X}}W$ where W is implicitly defined by (3.4) in the vector form $\widehat{\beta} = W\mathbf{Z}$. The incorporation of $\mathbf{Z}$ is necessary to the calculation of CV and GCV to account for the contemporaneous correlation. Let $\mathbf{Z} = [Z_1 \;\; Z_2 \;\; \cdots \;\; Z_{TS}]$ be the result of stacking the $\mathbf{Z}_t$, where Z_i are scalars. The value of CV $= CV(\lambda)$ can be found as follows: denote $\widehat{\mathbf{Y}}_{-t}$ as the fitted value for observation t based on the model with the parameters estimated without using

observation t. Then

$$\begin{aligned} CV(\lambda) &= \sum_{t=1}^{T}(\mathbf{Y}_t - \widehat{\mathbf{Y}}_{-t})^T\Omega^{-1}(\mathbf{Y}_t - \widehat{\mathbf{Y}}_{-t}) \\ &= \sum_{t=1}^{T}(\widehat{\Omega}^{-1/2}\mathbf{Y}_t - \widehat{\Omega}^{-1/2}\widehat{\mathbf{Y}}_{-t})^T(\widehat{\Omega}^{-1/2}\mathbf{Y}_t - \widehat{\Omega}^{-1/2}\widehat{\mathbf{Y}}_{-t}) \\ &= \sum_{i=1}^{TS}(Z_i - \widehat{Z}_{-i})^2 \\ &= \sum_{i=1}^{TS}\frac{(Z_i - \widehat{Z}_i)^2}{(1-h_{ii})^2}. \end{aligned} \tag{3.7}$$

GCV can also be defined similarly by replacing $1 - h_{ii}$ by the average.

$$GCV(\lambda) = \sum_{i=1}^{TS}\frac{(Z_i - \widehat{Z}_i)^2}{(1 - \frac{1}{TS}\sum_{j=1}^{TS}h_{jj})^2}. \tag{3.8}$$

In order to estimate λ using this technique, we will start the optimization at a value that "equates" the information from the data with that from the smoothness "prior", i.e. the initial value $\tilde{\lambda}$ is set to be

$$\tilde{\lambda} \doteq \frac{\mathrm{tr}(\sum_t X_t^T X_t)}{\mathrm{tr}(A^T A)}. \tag{3.9}$$

In other words, the initial value corresponds to a probably over-smoothed model with the amount of smoothing having as much weight as the data information. Since the value of λ is nonnegative, the optimization can be more easily done by applying Newton's method on $\log(\lambda)$ with a starting value of $\log(\tilde{\lambda})$ to find a minimum on either the CV or GCV functions. Once an optimum value of λ is found, it can be used in the procedure shown above. Alternatively, the estimation can be implemented via the `magic` function of the `mgcv` library of R, which is done in the simulation and data analysis reported below [22]. The function `magic` is a quick algorithm to find the

smoothing parameter that minimized the GCV function. The code for the estimation of this fixed effect model can be found in Appendix B.1.

3.2 Other methods for computing confidence bands

Along with the frequentist approach to estimating an individual confidence band for β, other methods of estimation are possible. In Section 3.2.1 we will consider the Bayesian approach of computing a confidence band [23]. In Section 3.2.2 we will elaborate two approaches to computing the confidence band using bootstrap approaches [7].

3.2.1 Bayesian approach

Using a Bayesian approach, the smoothness in β may be formulated in terms of putting a prior on β as

$$f_\beta(\beta) \propto e^{-\frac{\lambda}{2}\beta^T A^T A\beta}.$$

This prior is typically improper because A may not be of full rank. Now the conditional distribution of $\mathbf{Y}$ given the parameter β can then be stated as

$$f(\mathbf{Y}|\beta) \propto e^{-\frac{1}{2}(\mathbf{Y}-\mathbf{X}\beta)\Omega^{-1}(\mathbf{Y}-\mathbf{X}\beta)}.$$

So now we can compute the posterior distribution $f(\beta|\mathbf{Y})$.

$$f(\beta|\mathbf{Y}) \propto e^{-\frac{1}{2}(-2\beta^T\mathbf{X}^T\Omega^{-1}\mathbf{Y}+\beta^T(\mathbf{X}^T\Omega^{-1}\mathbf{X}+\lambda A^T A)\beta}$$

This posterior distribution turns out to be a normal distribution with variance-covariance matrix

$$(\lambda A^T A + \sum_t X_t^T \widehat{\Omega}^{-1} X_t)^{-1}. \tag{3.10}$$

We can then use the square root of the diagonal elements of (3.10) as standard errors for β, which can then be used for constructing confidence bands for β. The confidence band can then be computed by using the square roots of the diagonals of (3.10) as standard errors.

3.2.2 Bootstrap confidence bands

In order to compute the bootstrap confidence bands, we will have to use the fitted values of $\widehat{\mathbf{Y}}_t$ to compute the estimated residuals.

$$\widehat{\epsilon}_t = \mathbf{Y}_t - \widehat{\mathbf{Y}}_t,\ t = 1, \ldots, T \tag{3.11}$$

These $\widehat{\epsilon}_t$'s would keep the property that they approximately preserve the underlying contemporaneous correlation. From that, we can repeat the following steps N times.

1. Take a random sample of T $\widehat{\epsilon}_t$'s with replacement, where $\widehat{\epsilon}_t$ are defined in Equation (3.11). These sampled values will be called $\epsilon_1^*, \epsilon_2^*, \ldots, \epsilon_T^*$.

2. Compute the bootstrap response vector $\mathbf{Y}_t^*$ as

$$\mathbf{Y}_t^* = \widehat{\mathbf{Y}}_t + \epsilon_t^* \tag{3.12}$$

3. Fit $\mathbf{Y}_t^*$ from Equation (3.12) onto X_t with Ω and λ fixed at the values estimated from the original analysis to obtain β^*.

From these N bootstrapped samples, we can construct $(1-\alpha)\times 100\%$ (individual) confidence band for β; for each β_i, find the $(\alpha/2)\times 100$th percentile and the $(1-\alpha/2)\times 100$th percentile of β_i^* [7].

Similarly, we can define steps for computing a parametric bootstrap confidence band for β. Repeat the following N times.

1. Simulate $\epsilon_t^*, t = 1, \ldots, T$ from $N(0, \widehat{\Omega})$, where $\widehat{\Omega}$ is the variance-covariance matrix estimated from the original analysis.

2. Compute the bootstrap response vector $\mathbf{Y}_t^*$ as

$$\mathbf{Y}_t^* = \widehat{\mathbf{Y}}_t + \epsilon_t^* \tag{3.13}$$

3. Fit $\mathbf{Y}_t^*$ from Equation (3.13) onto X_t with Ω and λ fixed at the values estimated from the original analysis to obtain β^*.

Parametric bootstrapping is suitable in situations where the form of the underlying error distribution is known. On the other hand, the idea of bootstrapping is to be able to compute standard errors without restricting ourselves to a known distribution. So nonparametric bootstrap standard errors would be preferred when an assumption on the error distributions is uncertain and is considered more robust than parametric bootstrap confidence bands. It should be noted that all methods discussed so far do not account for the variations in the smoothing parameters, i.e.

the confidence bands so constructed are conditional on the estimated smoothing parameter.

Code for the estimation of the bootstrapped samples can be found in Appendix B.2.1. The nonparametric bootstrapped samples are generated in Appendix B.2.2, while the parametric counterparts are generated in Appendix B.2.3.

3.3 Inflow of larvae cod as an example

The developed method allows us to study the problem from a different perspective. Instead of searching for an optimal window for averaging the daily inflow, we consider the distribution of the cod spawning date. Let S be the day counted starting from the beginning of March of each year, when a randomly selected adult cod spawns. Let ψ_j be the probability that $S = j$. The daily contribution of North Sea cod to the Skagerrak is postulated to additively contribute, on the logarithmic scale, to the young cod counts by an amount proportional to $\psi_j c_{t,j} b_t$ where b_t is the log spawning biomass in the tth year and $c_{t,j}$ be the mean inflow on the jth (counted starting March 1st) day of the tth year; for simplicity of notation, the proportional constant is absorbed into ψ_j so that they need not sum to 1. In other words, the total annual North-Sea-cod contribution equaled $\sum_{j=1}^{61} \psi_j c_{t,j} b_t$, under the transportation hypothesis.

Let $n^0_{t,s}$ be the logarithm of the number of young cod caught in fjord s in year t. We confine the analysis to eight fjords in the Southern Norway, over the period from 1971 to 1997 over which we have complete data. These eight fjords are reported

in the earlier analysis to admit significant transportation effects. In fact, $n^0_{t,s}$ are part of a longer residual series from a stochastic regression model using a longer database that has adjusted for the intra-specific and the inter-specific effects, as well as the environmental effects on the local cod in the Skagerrak [4] [5] [18]. We now state the model.

$$n^0_{t,s} = \kappa_s + \zeta b_t + \sum_{j=1}^{61} \psi_j c_{t,j} b_t + e_{t,s};\ t = 1, \ldots, 26;\ s = 1, \ldots, 8. \tag{3.14}$$

The κ_s's can be interpreted as the fjord-specific effect on the cod population, and they may be expected to be close to zero because $n^0_{t,s}$ are part of a long residual series. The term ζb_t can be interpreted as the contributions of the North Sea adult cod by directly swimming to the Skagerrak and spawning there.

Our prior knowledge of the amount of spawning that occurs outside of the months of April and March (see, e.g., [13]) allows us to impose the following end constraints:

$$\begin{aligned} \psi_1 &= \eta_1 \\ -2\psi_1 + \psi_2 &= \eta_2 \\ \psi_{D-1} - 2\psi_D &= \eta_{D+1} \\ \psi_D &= \eta_{D+2}, \end{aligned} \tag{3.15}$$

where $D = 61$. These end constraints merely incorporate the prior assumption that ψ is zero beyond March and April, and maintain the constraint of small roughness across the boundaries. We estimated the model using the method proposed in Section 3.1.

Figure 3.4 plots the ψ function where the central solid line is the estimated curve and the other two dashed lines enclose the individual 95% confidence limits. For the frequentist confidence band, there is a significant spike in the spawning that begins on March 1st and ends March 24th. For the Bayesian confidence band, the spike is between March 16th and March 21st. The 90% nonparametric bootstrap confidence band indicates an interval of March 12th to March 29th, and the 90% parametric bootstrap confidence band indicates March 18th to March 27th. Certainly all of these approaches are asserting that the spawning is occuring in and around the second half of March. Table 3.1 reports the rest of the parameter estimates. In particular, all intercept terms are non-significant, and so is the coefficient estimate of ζ.

Model checking procedures are done on our data by drawing a residual plot, which can be found in Figure 3.1, a residual plot over year in Figure 3.2, and a normal Q-Q plot of the residuals, which was drawn in Figure 3.3. These model diagnostics suggest that the model assumptions seem reasonable.

In conclusion, there is clear evidence that sea current transported the North Sea cod eggs to the Skagerrak, mainly over the second half of March. Furthermore, the data suggest that the North Sea adult cod did not swim to the Skagerrak to spawn there. These conclusions are consistent with the previous findings that is obtained by assuming constant ψ over 2-week periods [18]. However, our new method allows for much more refined conclusions of great importance to the field of marine ecology.

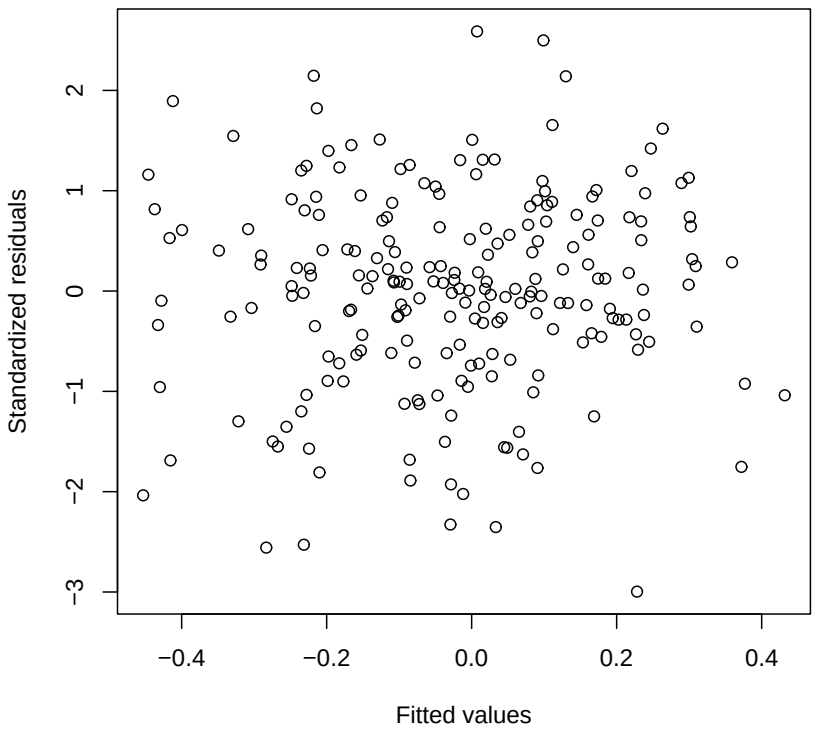

Figure 3.1: Residual plot for the model fitted with the fixed-effect model.

Figure 3.2: Residual plot of the residuals of the fixed effect model by year (from 1971 to 1997).

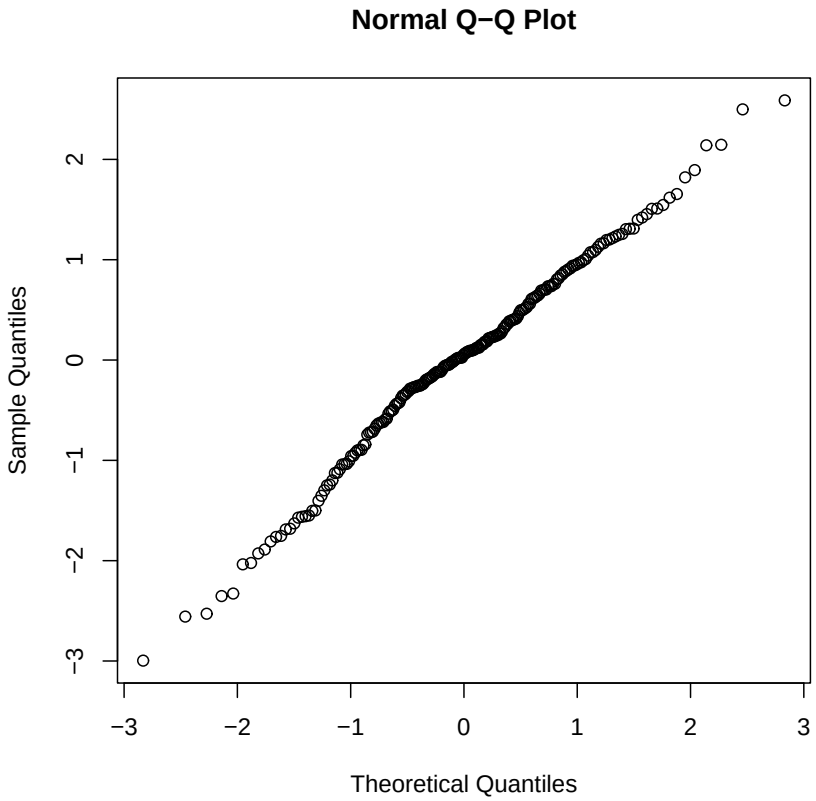

Figure 3.3: Normal Q-Q plot of the errors for the fixed effect model.

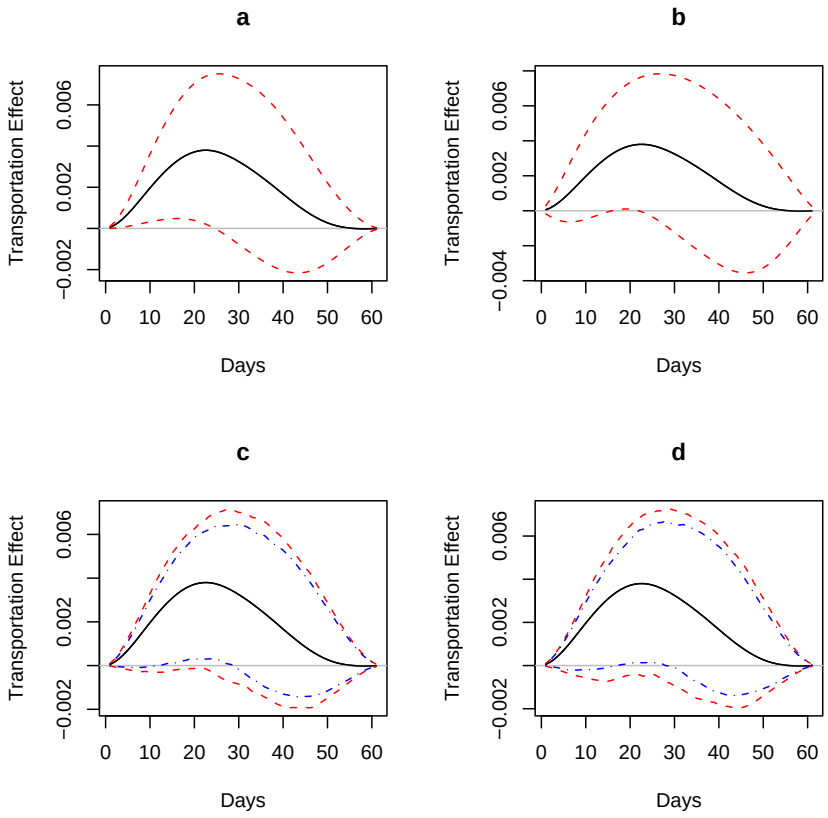

Figure 3.4: The plot of $\widehat{\psi}_j$ versus j for the North Sea cod – the central solid line in each graph. The confidence bands are shown for a) frequentist (dashed line is 95%), b) Bayesian (dashed line is 95%), c) nonparametric bootstrap (dashed line is 95%, dotted-dashed line is 90%), and d) parametric bootstrap (dashed line is 95%, dotted-dashed line is 90%) methods.

Parameter	Estimate	SE
κ_1	0.219	0.261
κ_2	0.206	0.267
κ_3	0.418	0.325
κ_4	0.199	0.224
κ_5	0.017	0.237
κ_6	0.001	0.261
κ_7	0.126	0.265
κ_8	0.345	0.255
ζ	-0.119	0.083

Table 3.1: Estimates of model parameters for the Skagerrak cod.

3.4 Simulation

We investigate the empirical performance of the proposed method by simulations. We will study two matters: the number of significant modes detected and the error involved in the estimates. The simulation model is motivated by the cod example:

$$\mathbf{Y}_t = \kappa + \zeta b_t + \psi^T \mathbf{c}_t b_t + \mathbf{e}_t; t = 1, \ldots, T, \tag{3.16}$$

where $\mathbf{Y}_t$ is the response vector of dimension $S \times 1$, κ, ζ, and ϕ are unknown parameters, and b_t and $\mathbf{c}_t$ are covariates, where the dimension of the vectors ψ and $\mathbf{c}_t$ is $D \times 1$ for both. The error vector $\mathbf{e}_t$ has a multivariate normal distribution with mean zero and variance-covariance matrix $\sigma_e^2 P$, where P is a correlation matrix.

In the examples below, $T = 50$, $S = 3$, $D = 61$, $\kappa = [1, 0, -1]^T$, and $\zeta = 1$. The value of b_t is determined by taking a random number from a normal distribution with mean zero and standard deviation one, while the value of $\mathbf{c}_t$ are determined by creating a vector of sixty-one random numbers from a normal distribution with zero mean and unit variance. There will be two levels of the error standard deviation: either 0.05 or 0.1. We will use a correlation matrix that is compound symmetric, meaning that

$$P = \begin{bmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{bmatrix}$$

where there are four levels of ρ: -0.2, 0.2, 0.5, and 0.9. The ψ_j equal the probability density function at j of an equal mixture of two normal distributions, namely $N(30-\Delta, 9)$ and $N(30+\Delta, 4), \Delta = 10, 5, 0$, making three levels of the ψ function. Hence ψ has two modes that are separated by either 20 units, 10 units, or 0 units (thus making one mode). Each case was simulated 1,000 times. The plots of the three different sets of ψ's used can be found in Figure 3.5.

Recall the simulation was used to study two aspects of the problem: the number of significant modes detected and the error involved in the estimates. Results of both of these can be found in Tables 3.1, 3.2, 3.3, 3.4, 3.5, 3.6, 3.7, 3.8, and 3.9 using standard errors determined by the frequentist approach, i.e. using Formula (3.5). Recent works have made a comparison between the Bayesian and frequentist approaches for the case of curves with continuous arguments [23]. In the estimation, we impose the smoothness constraints

$$\psi_i - 2\psi_{i-1} + \psi_{i-2} = \eta_i;\ i = 3, \ldots, D,$$

as well as end constraints:

$$\begin{aligned} \psi_1 &= \eta_1 \\ \psi_2 - 2\psi_1 &= \eta_2 \\ -2\psi_D + \psi_{D-1} &= \eta_{D+1} \\ \psi_D &= \eta_{D+2}, \end{aligned}$$

to ensure that the ψ function estimates are smooth across the boundaries beyond which they are zero. The simulation results based on the Bayesian confidence bands

can be found in Tables 3.10, 3.11, 3.12, 3.13, 3.14, 3.15, 3.16, and 3.17. In regard to the first aspect, the simulation catches unimodality and multimodality fairly well when the correlation is high and the error variance is low. (Note that we count the number of modes only for the ψ_j's that are significantly different from zero, hence there could be no mode in the curve if none of the ψ_j's are significant.) The Bayesian confidence bands tend to catch unimodality and bimodality better than the frequentist confidence bands when the correlation is closer to zero. It seems that the two methods give similar results with a correlation around 0.5. The frequentist confidence bands tend to catch unimodality and bimodality better than the Bayesian confidence bands when the correlation is around 0.9. As far as the second aspect, the mean absolute deviation and mean deviation were small as compared to the maximum value of the ψ's being estimated. Their standard deviations were small as well and depended proportionately on the error variance.

3.5 Penalized maximum likelihood estimation

We will now explore the theory behind our model. Let's assume $\{(X_t, Y_t)^T, t \in Z\}$ is weakly stationary with absolutely summable autocorrelations and have the same distribution of (X, Y). Recall the penalized log-likelihood.

$$\ell(\beta, \Omega) = -\frac{T}{2}\log|\Omega| - \frac{1}{2}\sum_{t=1}^{T}(Y_t - X_t\beta)^T\,\Omega^{-1}\,(Y_t - X_t\beta) - \lambda\beta^T A^T A\beta \tag{3.17}$$

Let $\Delta = \Omega^{-1}$ and $U = \sum_t (Y_t - X_t\beta)(Y_t - X_t\beta)^T$. Adopting the Δ parameter

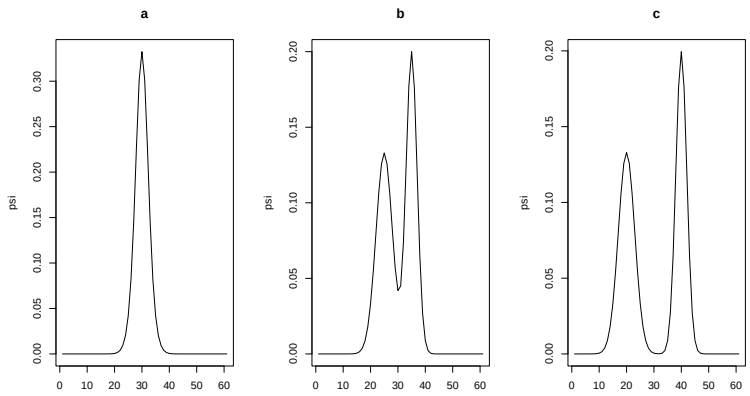

Figure 3.5: Plots of ψ_j versus j for the simulation model; ψ_j is the probability density function of the an equal mixture of $N(30-\Delta, 9)$ and $N(30+\Delta, 4)$, where $\Delta = 0, 5, 10$, from left to right.

noise standard deviation = 0.05				
		Δ		
		0	5	10
no. of modes		% of significant modes		
0		0.0%	0.0%	0.0%
1		10.2%	0.0%	0.0%
2		18.2%	14.1%	21.2%
≥ 3		71.6%	85.9%	78.8%
$\frac{1}{D}\sum_j \lvert\widehat{\psi}_j - \psi_j\rvert$	mean	0.0285	0.0263	0.0258
	SD	0.00767	0.00726	0.00713
$\frac{1}{D}\sum_j \widehat{\psi}_j - \psi_j$	mean	0.0163	0.0164	0.0163
	SD	0.00434	0.00435	0.00425

Table 3.2: Simulation results for the model defined by (3.16) for $\rho = -0.2$ using the frequentist confidence bands.

noise standard deviation = 0.1				
		Δ		
		0	5	10
no. of modes		% of significant modes		
0		0.0%	0.0%	0.0%
1		20.3%	10.9%	0.1%
2		26.2%	23.2%	38.6%
≥ 3		53.5%	65.9%	61.3%
$\frac{1}{D}\sum_j \lvert\widehat{\psi}_j - \psi_j\rvert$	mean	0.0360	0.0311	0.0310
	SD	0.0139	0.0117	0.0123
$\frac{1}{D}\sum_j \widehat{\psi}_j - \psi_j$	mean	0.0163	0.0166	0.0164
	SD	0.00786	0.00778	0.00782

Table 3.3: Simulation results for the model defined by (3.16) for $\rho = -0.2$ using the frequentist confidence bands.

noise standard deviation = 0.05				
		Δ		
		0	5	10
no. of modes		% of significant modes		
0		0.0%	0.0%	0.0%
1		41.0%	1.2%	0.0%
2		22.2%	43.5%	48.1%
≥ 3		36.8%	55.3%	51.9%
$\frac{1}{D}\sum_j \lvert\widehat{\psi}_j - \psi_j\rvert$	mean	0.0357	0.0328	0.0318
	SD	0.0154	0.0153	0.0151
$\frac{1}{D}\sum_j \widehat{\psi}_j - \psi_j$	mean	0.0165	0.0165	0.0165
	SD	0.00669	0.00690	0.00674

Table 3.4: Simulation results for the model defined by (3.16) for $\rho = 0.2$ using the frequentist confidence bands.

noise standard deviation = 0.1			
	Δ		
	0	5	10
no. of modes	% of significant modes		
0	0.0%	0.1%	0.1%
1	51.4%	36.8%	6.7%
2	21.5%	30.3%	57.4%
≥ 3	27.1%	32.8%	35.8%
$\frac{1}{D}\sum_j \lvert\widehat{\psi}_j - \psi_j\rvert$ mean	0.0489	0.0411	0.0423
SD	0.0300	0.0284	0.0296
$\frac{1}{D}\sum_j \widehat{\psi}_j - \psi_j$ mean	0.0165	0.0168	0.0165
SD	0.0125	0.0121	0.0121

Table 3.5: Simulation results for the model defined by (3.16) for $\rho = 0.2$ using the frequentist confidence bands.

noise standard deviation = 0.05				
		Δ		
		0	5	10
no. of modes		% of significant modes		
0		0.0%	0.0%	0.0%
1		47.9%	3.6%	0.0%
2		15.9%	45.0%	50.5%
≥ 3		36.2%	51.4%	49.5%
$\frac{1}{D}\sum_j \lvert\widehat{\psi}_j - \psi_j\rvert$	mean	0.0465	0.0429	0.0423
	SD	0.0260	0.0258	0.0257
$\frac{1}{D}\sum_j \widehat{\psi}_j - \psi_j$	mean	0.0166	0.0167	0.0165
	SD	0.00913	0.00909	0.00898

Table 3.6: Simulation results for the model defined by (3.16) for $\rho = 0.5$ using the frequentist confidence bands.

noise standard deviation = 0.1			
	Δ		
	0	5	10
no. of modes	% of significant modes		
0	0.1%	0.6%	1.7%
1	56.3%	45.1%	19.5%
2	12.6%	21.8%	42.0%
≥ 3	31.0%	32.5%	36.8%
$\frac{1}{D}\sum_j \lvert\widehat{\psi}_j - \psi_j\rvert$ mean	0.0712	0.0602	0.0644
SD	0.0549	0.0533	0.0559
$\frac{1}{D}\sum_j \widehat{\psi}_j - \psi_j$ mean	0.0169	0.0169	0.0165
SD	0.0172	0.0164	0.0167

Table 3.7: Simulation results for the model defined by (3.16) for $\rho = 0.5$ using the frequentist confidence bands.

noise standard deviation = 0.05			
	Δ		
	0	5	10
no. of modes	% of significant modes		
0	0.0%	0.1%	0.1%
1	61.4%	9.5%	1.8%
2	11.5%	51.8%	58.8%
≥ 3	27.1%	38.6%	39.3%
$\frac{1}{D}\sum_j \lvert\widehat{\psi}_j - \psi_j\rvert$ mean	0.0524	0.0491	0.0486
$\frac{1}{D}\sum_j \lvert\widehat{\psi}_j - \psi_j\rvert$ SD	0.0321	0.0329	0.0330
$\frac{1}{D}\sum_j \widehat{\psi}_j - \psi_j$ mean	0.0165	0.0167	0.0166
$\frac{1}{D}\sum_j \widehat{\psi}_j - \psi_j$ SD	0.0104	0.0105	0.0105

Table 3.8: Simulation results for the model defined by (3.16) for $\rho = 0.9$ using the frequentist confidence bands.

noise standard deviation = 0.1			
	Δ		
	0	5	10
no. of modes	% of significant modes		
0	1.0%	10.1%	13.6%
1	64.6%	51.1%	36.1%
2	12.0%	15.7%	25.9%
≥ 3	22.4%	23.1%	24.4%
$\frac{1}{D}\sum_j \lvert\widehat{\psi}_j - \psi_j\rvert$ mean	0.0842	0.0750	0.0794
$\frac{1}{D}\sum_j \lvert\widehat{\psi}_j - \psi_j\rvert$ SD	0.0689	0.0707	0.0722
$\frac{1}{D}\sum_j \widehat{\psi}_j - \psi_j$ mean	0.0167	0.0169	0.0163
$\frac{1}{D}\sum_j \widehat{\psi}_j - \psi_j$ SD	0.0202	0.0195	0.0202

Table 3.9: Simulation results for the model defined by (3.16) for $\rho = 0.9$ using the frequentist confidence bands.

noise standard deviation = 0.05				
		Δ		
		0	5	10
no. of modes		% of significant modes		
0		0.0%	0.0%	0.0%
1		56.4%	0.0%	0.0%
2		27.0%	60.7%	63.7%
≥ 3		16.6%	39.3%	36.3%
$\frac{1}{D}\sum_j \lvert\widehat{\psi}_j - \psi_j\rvert$	mean	0.0285	0.0263	0.0258
	SD	0.00767	0.00726	0.00713
$\frac{1}{D}\sum_j \widehat{\psi}_j - \psi_j$	mean	0.0163	0.0164	0.0163
	SD	0.00434	0.00435	0.00425

Table 3.10: Simulation results for the model defined by (3.16) for $\rho = -0.2$ using the Bayesian confidence bands.

noise standard deviation = 0.1				
		Δ		
		0	5	10
no. of modes		% of significant modes		
0		0.0%	0.0%	0.0%
1		65.6%	13.1%	0.2%
2		22.3%	58.0%	72.8%
≥ 3		12.1%	28.9%	27.0%
$\frac{1}{D}\sum_j \lvert\widehat{\psi}_j - \psi_j\rvert$	mean	0.0360	0.0311	0.0310
	SD	0.0139	0.0117	0.0123
$\frac{1}{D}\sum_j \widehat{\psi}_j - \psi_j$	mean	0.0163	0.0166	0.0164
	SD	0.00786	0.00778	0.00782

Table 3.11: Simulation results for the model defined by (3.16) for $\rho = -0.2$ using the Bayesian confidence bands.

noise standard deviation = 0.05				
		Δ		
		0	5	10
no. of modes		% of significant modes		
0		0.0%	0.0%	0.0%
1		56.3%	1.3%	0.0%
2		22.4%	54.9%	59.7%
≥ 3		21.3%	43.8%	40.3%
$\frac{1}{D}\sum_j \lvert\widehat{\psi}_j - \psi_j\rvert$	mean	0.0357	0.0328	0.0318
	SD	0.0154	0.0153	0.0151
$\frac{1}{D}\sum_j \widehat{\psi}_j - \psi_j$	mean	0.0165	0.0165	0.0165
	SD	0.00669	0.00690	0.00674

Table 3.12: Simulation results for the model defined by (3.16) for $\rho = 0.2$ using the Bayesian confidence bands.

noise standard deviation = 0.1			
	Δ		
	0	5	10
no. of modes	% of significant modes		
0	0.0%	0.2%	0.2%
1	63.7%	38.9%	7.2%
2	18.5%	36.6%	64.6%
≥ 3	17.8%	24.3%	28.0%
$\frac{1}{D}\sum_j \lvert\widehat{\psi}_j - \psi_j\rvert$ mean	0.0489	0.0411	0.0423
$\frac{1}{D}\sum_j \lvert\widehat{\psi}_j - \psi_j\rvert$ SD	0.0300	0.0284	0.0296
$\frac{1}{D}\sum_j \widehat{\psi}_j - \psi_j$ mean	0.0165	0.0168	0.0165
$\frac{1}{D}\sum_j \widehat{\psi}_j - \psi_j$ SD	0.0125	0.0121	0.0121

Table 3.13: Simulation results for the model defined by (3.16) for $\rho = 0.2$ using the Bayesian confidence bands.

noise standard deviation = 0.05				
		Δ		
		0	5	10
no. of modes		% of significant modes		
0		0.0%	0.0%	0.0%
1		47.8%	3.5%	0.0%
2		19.9%	44.5%	49.6%
≥ 3		32.3%	52.0%	50.4%
$\frac{1}{D}\sum_j \lvert\widehat{\psi}_j - \psi_j\rvert$	mean	0.0465	0.0429	0.0423
	SD	0.0260	0.0258	0.0257
$\frac{1}{D}\sum_j \widehat{\psi}_j - \psi_j$	mean	0.0166	0.0167	0.0165
	SD	0.00913	0.00909	0.00898

Table 3.14: Simulation results for the model defined by (3.16) for $\rho = 0.5$ using the Bayesian confidence bands.

noise standard deviation = 0.1			
	Δ		
	0	5	10
no. of modes	% of significant modes		
0	0.0%	0.4%	1.1%
1	53.9%	43.3%	14.2%
2	16.1%	23.5%	48.3%
≥ 3	30.0%	32.8%	36.4%
$\frac{1}{D}\sum_j \lvert\widehat{\psi}_j - \psi_j\rvert$ mean	0.0712	0.0602	0.0644
SD	0.0549	0.0533	0.0559
$\frac{1}{D}\sum_j \widehat{\psi}_j - \psi_j$ mean	0.0169	0.0169	0.0165
SD	0.0172	0.0164	0.0167

Table 3.15: Simulation results for the model defined by (3.16) for $\rho = 0.5$ using the Bayesian confidence bands.

noise standard deviation = 0.05			
	Δ		
	0	5	10
no. of modes	% of significant modes		
0	0.0%	0.0%	0.0%
1	47.3%	7.1%	0.1%
2	18.4%	43.2%	50.6%
≥ 3	34.3%	49.7%	49.3%
$\frac{1}{D}\sum_j \lvert\widehat{\psi}_j - \psi_j\rvert$ mean	0.0524	0.0491	0.0486
$\frac{1}{D}\sum_j \lvert\widehat{\psi}_j - \psi_j\rvert$ SD	0.0321	0.0329	0.0330
$\frac{1}{D}\sum_j \widehat{\psi}_j - \psi_j$ mean	0.0165	0.0167	0.0166
$\frac{1}{D}\sum_j \widehat{\psi}_j - \psi_j$ SD	0.0104	0.0105	0.0105

Table 3.16: Simulation results for the model defined by (3.16) for $\rho = 0.9$ using the Bayesian confidence bands.

noise standard deviation = 0.1			
	Δ		
	0	5	10
no. of modes	% of significant modes		
0	0.1%	1.9%	2.8%
1	51.6%	45.3%	21.7%
2	16.8%	18.2%	37.8%
≥ 3	31.5%	34.6%	37.7%
$\frac{1}{D}\sum_j \lvert\widehat{\psi}_j - \psi_j\rvert$ mean	0.0842	0.0750	0.0794
SD	0.0689	0.0707	0.0722
$\frac{1}{D}\sum_j \widehat{\psi}_j - \psi_j$ mean	0.0167	0.0169	0.0163
SD	0.0202	0.0195	0.0202

Table 3.17: Simulation results for the model defined by (3.16) for $\rho = -0.2$ using the Bayesian confidence bands.

instead of Ω will greatly simplify many of the following formulas. Rewrite the log-likelihood as follows.

$$\ell(\beta, \Delta) \;=\; \frac{T}{2}\log|\Delta| - \frac{1}{2}\mathrm{tr}(\Delta U) - \frac{\lambda}{2}\beta^T A^T A\beta$$

We compute the first differential to find the Penalized Maximum Likelihood (PML) estimates for β and $\mathrm{v}(\Delta)$ [15], assuming λ is fixed, where $\mathrm{v}(\Delta)$ denotes the vector obtained by stacking the upper triangular part (including the diagonal) of the symmetric matrix Δ. For any matrix M, $\mathrm{vec}(M)$ vectorizes M by stacking up its column vectors. We shall make heavy use of the result that $\mathrm{vec}(ABC) = (C^T \otimes A)\mathrm{vec}(B)$, where for any two matrices P and Q, $P \otimes Q = (P_{i,j}Q)$, where $P = (P_{i,j})$, making the $\otimes$ operator the (right) Kronecker product [15].

$$\begin{aligned}
d\ell &= \frac{T}{2}\mathrm{tr}(\Delta^{-1}d\Delta) - \frac{1}{2}\mathrm{tr}(d\Delta U) - \frac{1}{2}\mathrm{tr}(\Delta dU) - \frac{\lambda}{2}d\beta^T A^T A\beta - \frac{\lambda}{2}\beta^T A^T Ad\beta \\
&= \frac{1}{2}\mathrm{vec}(d\Delta)^T\mathrm{vec}(T\Delta^{-1}) - \frac{1}{2}\mathrm{vec}(d\Delta)^T\mathrm{vec}(U) + \\
&\quad \frac{1}{2}\sum_t (X_t d\beta)^T\Delta(Y_t - X_t\beta) + \frac{1}{2}\sum_t (Y_t - X_t\beta)^T\Delta X_t d\beta) - \lambda d\beta^T A^T A\beta \\
&= \frac{1}{2}\mathrm{vec}(d\Delta)^T\mathrm{vec}(T\Delta^{-1} - U) + d\beta^T\left(\sum_t X_t^T\Delta(Y_t - X_t\beta) - \lambda A^T A\beta\right) \\
&= \frac{1}{2}d\mathrm{v}(\Delta)^T D_S^T\mathrm{vec}(T\Delta^{-1} - U) + d\beta^T\left(\sum_t X_t^T\Delta(Y_t - X_t\beta) - \lambda A^T A\beta\right) \\
&= \left[\begin{array}{cc} d\beta^T & \mathrm{v}(\Delta)^T \end{array}\right]\left[\begin{array}{c} \sum_t X_t^T\Delta(Y_t - X_t\beta) - \lambda A^T A\beta \\ \\ \frac{1}{2}D_S^T\mathrm{vec}(T\Delta^{-1} - U) \end{array}\right]
\end{aligned}$$

From this expression, we can read off $\partial\ell/\partial\beta$ and $\partial\ell/\partial\mathrm{v}(\Delta)$. Setting the first

order partial derivatives to be zero, we can get a system of equations for $\widehat{\beta}$ and $\widehat{\Delta}$. For the ML estimator for β and with a fixed Δ, we have

$$\frac{\partial \ell}{\partial \beta} = \sum_t X_t^T \Delta (Y_t - X_t \beta) - \lambda A^T A \beta$$

Setting this expression equal to zero, we get

$$\begin{array}{rcl} \sum_t X_t^T \Delta Y_t & = & \left\{\sum_t (X_t^T \Delta X_t) + \lambda A^T A\right\} \widehat{\beta} \\ \Rightarrow \qquad \widehat{\beta} & = & \left\{\sum_t (X_t^T \Delta X_t) + \lambda A^T A\right\}^{-1} \sum_t X_t^T \Delta Y_t. \end{array} \tag{3.18}$$

Likewise for $\mathrm{v}(\Delta)$ and if β is known, we have

$$\frac{\partial \ell}{\partial \mathrm{v}(\Delta)} = \frac{1}{2} D_S^T \mathrm{vec}(T\Delta^{-1} - U) = 0$$

Again, setting this to be zero, we get

$$\begin{array}{rcl} \mathrm{vec}(T\widehat{\Delta}^{-1}) & = & \mathrm{vec}(U) \\ \Rightarrow \qquad T\widehat{\Delta}^{-1} & = & U \\ \Rightarrow \qquad \widehat{\Delta}^{-1} & = & \frac{U}{T}. \end{array} \tag{3.19}$$

These equations can be solved iteratively to get the penalized estimator of β and Δ. We shall show below that under some mild regularity conditions, Equations (3.18) and (3.19) admit a solution. To see if the solutions maximize the penalized likelihood, we must compute the Hessian through the second differential of the log-likelihood. Below, D_S is a constant matrix such that $\Delta = D_s \mathrm{v}(\Delta)$.

$$
\begin{aligned}
d^2\ell &= \frac{1}{2}d\mathrm{v}(\Delta)^T D_S^T \mathrm{vec}(T d\Delta^{-1}) - \frac{1}{2}d\mathrm{v}(\Delta)^T D_S^T \mathrm{vec}(dU) + \\
&\quad d\beta^T \sum_t X_t^T d\Delta (Y_t - X_t\beta) - d\beta^T \sum_t X_t^T \Delta X_t d\beta - \lambda d\beta^T A^T A d\beta \\
&= \frac{1}{2}d\mathrm{v}(\Delta)^T D_S^T \mathrm{vec}(-T\Delta^{-1} d\Delta \Delta^{-1}) + \frac{1}{2}d\mathrm{v}(\Delta)^T D_S^T \sum_t \left\{(Y_t - X_t\beta)\otimes X_t\right\} d\beta \\
&\quad + \frac{1}{2}d\mathrm{v}(\Delta)^T D_S^T \sum_t \left\{X_t \otimes (Y_t - X_t\beta)\right\} d\beta + d\beta^T \sum_t X_t^T d\Delta (Y_t - X_t\beta) - \\
&\quad d\beta^T \sum_t X_t^T \Delta X_t d\beta - \lambda d\beta^T A^T A d\beta \\
&= -\frac{T}{2}d\mathrm{v}(\Delta)^T D_S^T (\Delta^{-1}\otimes\Delta^{-1}) d\mathrm{vec}(\Delta) + \\
&\quad \frac{1}{2}d\mathrm{v}(\Delta)^T D_S^T \sum_t \left\{(Y_t - X_t\beta)\otimes X_t\right\} d\beta + \\
&\quad \frac{1}{2}d\mathrm{v}(\Delta)^T D_S^T \sum_t \left\{X_t \otimes (Y_t - X_t\beta)\right\} d\beta + d\beta^T \sum_t X_t^T d\Delta (Y_t - X_t\beta) - \\
&\quad d\beta^T \sum_t X_t^T \Delta X_t d\beta - \lambda d\beta^T A^T A d\beta \\
&= -\frac{T}{2}d\mathrm{v}(\Delta)^T D_S^T (\Delta^{-1}\otimes\Delta^{-1}) D_S d\mathrm{v}(\Delta) + \\
&\quad \frac{1}{2}d\mathrm{v}(\Delta)^T D_S^T \sum_t \left\{(Y_t - X_t\beta)\otimes X_t\right\} d\beta + \\
&\quad \frac{1}{2}d\mathrm{v}(\Delta)^T D_S^T \sum_t \left\{X_t \otimes (Y_t - X_t\beta)\right\} d\beta + \\
&\quad \frac{1}{2}d\mathrm{v}(\Delta)^T D_S^T \sum_t \left\{(Y_t - X_t\beta)\otimes X_t\right\} d\beta + \\
&\quad \frac{1}{2}d\mathrm{v}(\Delta)^T D_S^T \sum_t \left\{X_t \otimes (Y_t - X_t\beta)\right\} d\beta - d\beta^T \sum_t X_t^T \Delta X_t d\beta - \lambda d\beta^T A^T A d\beta \\
&= -\frac{T}{2}d\mathrm{v}(\Delta)^T D_S^T (\Delta^{-1}\otimes\Delta^{-1}) D_S d\mathrm{v}(\Delta) + \\
&\quad d\mathrm{v}(\Delta)^T D_S^T \sum_t \{X_t \otimes (Y_t - X_t\beta)\} d\beta + \\
&\quad d\beta^T \sum_t \{X_t^T \otimes (Y_t - X_t\beta)^T\} D_S d\mathrm{v}(\Delta)^T - \\
&\quad d\beta^T \sum_t X_t^T \Delta X_t d\beta - \lambda d\beta^T A^T A d\beta
\end{aligned}
\tag{3.20}
$$

$$= \begin{bmatrix} d\beta^T & d\mathrm{v}(\Delta)^T \end{bmatrix} H_\ell \begin{bmatrix} d\beta \\ d\mathrm{v}(\Delta)^T \end{bmatrix}, \text{ where}$$

$$H_\ell = \begin{bmatrix} -\sum_t X_t^T \Delta X_t - \lambda A^T A & \sum_t \{X_t^T \otimes (Y_t - X_t\beta)^T\} D_S \\ D_S^T \sum_t \{X_t \otimes (Y_t - X_t\beta)\} & -\frac{T}{2} D_S^T (\Delta^{-1} \otimes \Delta^{-1}) D_S \end{bmatrix}$$

is the Hessian matrix. This is because the fourth term in (3.20) can be written

$$\begin{aligned} d\beta^T \sum_t \left\{ X_t^T d\Delta (Y_t - X_t\beta) \right\} &= \frac{1}{2} d\beta^T \sum_t X_t^T d\Delta (Y_t - X_t\beta) + \\ &\quad \frac{1}{2} \sum_t (Y_t - X_t\beta) d\Delta X_t d\beta \\ &= \frac{1}{2} \mathrm{tr} \left\{ \sum_t d\Delta X_t d\beta (Y_t - X_t\beta)^T \right\} + \\ &\quad \frac{1}{2} \mathrm{tr} \left\{ \sum_t d\beta (Y_t - X_t\beta)^T d\Delta X_t \right\} \\ &= \frac{1}{2} d\mathrm{v}(\Delta)^T D_S^T \sum_t \left\{ (Y_t - X_t\beta) \otimes X_t \right\} d\beta + \\ &\quad \frac{1}{2} d\mathrm{v}(\Delta)^T D_S^T \sum_t \left\{ X_t \otimes (Y_t - X_t\beta) \right\} d\beta. \end{aligned}$$

3.6 Consistency

Theorem 3.6.1. *Assume*

1. *the vector process* $\left\{ (X_t, Y_t)^T, t \in Z \right\}$ *is weakly stationary with absolutely summable autocorrelations, and*

2. λ *in (3.17) is* $o_P(T)$.

Then Equations (3.18) and (3.19) admit a solution around the true parameter value, denoted by $\widehat{\theta} = (\widehat{\beta}, \widehat{\Delta})$. *Moreover* $\widehat{\theta}$ *is a consistent estimator of* $\theta_0 = (\beta_0, \Delta_0)$.

Proof: Because of assumption 2, the penalty term turns out to be negligible and will be suppressed within the proof (See remark at the end of the section.). Define $\ell_T(\theta) = (1/T)\ell(\theta)$. Let $\theta_0 = (\beta_0, \mathrm{v}(\Delta_0))$, which is the true value of the parameters $\theta = (\beta, \mathrm{v}(\Delta))$. Consider a ball of radius ϵ,

$$B(\theta_0, \epsilon) = \{\theta; |\theta - \theta_0| < \epsilon\},$$

and its boundary

$$\partial B = \partial B(\theta_0, \epsilon) = \{\theta; |\theta - \theta_0| = \epsilon\}.$$

Below all expectations are taken with respect to the true model. We first wish to show that

$$E(\ell_T(\theta_0)) > \sup\{E(\ell_T(\theta));\ \theta \in \partial B\}. \tag{3.21}$$

We know from the property of the Kullback-Leibler divergence [19] that

$$E(\ell_T(\theta_0)) - E(\ell_T(\theta)) > 0,\ \forall\ \theta \neq \theta_0. \tag{3.22}$$

Since ℓ is a continuous function and $\partial B(\theta_0, \epsilon)$ is a compact set, then (3.22) implies (3.21).

Below we shall show that for any compact set K, we have the following convergence result.

$$\sup\{|\ell_T(\theta_0) - \ell_T(\theta) - \{E(\ell_T(\theta_0)) - E(\ell_T(\theta))\}|;\ \theta \in K\} \overset{a.s.}{\to} 0 \tag{3.23}$$

If we consider $K = \partial B(\theta_0, \epsilon)$ then K is a closed and bounded set, making it compact. By continuity, $\ell_T(\cdot)$ attains a maximum on $\overline{B(\theta_0, \epsilon)}$, the closure of $B(\theta_0, \epsilon)$. However (3.21) and (3.23) imply that $\ell_T(\cdot)$ attains a local maximum in $B(\theta_0, \epsilon)$ with probability tending to 1 as $T \to \infty$. This local maximum will be taken as $\widehat{\theta}$.

To complete the proof, we now verify (3.23) by first noticing that

$$\ell_T(\theta_0) - \ell_T(\theta) - \{E(\ell_T(\theta_0)) - E(\ell_T(\theta))\}$$

$$\begin{aligned} = \quad & -\frac{1}{2}\mathrm{tr}\left\{\frac{1}{T}\Delta\sum_t (Y_t - X_t\beta)(Y_t - X_t\beta)^T\right\} + \\ & \frac{1}{2}\mathrm{tr}\left\{\frac{1}{T}\Delta_0\sum_t (Y_t - X_t\beta_0)(Y_t - X_t\beta_0)^T\right\} - \{E(\ell_T(\theta_0)) - E(\ell_T(\theta))\}. \end{aligned} \tag{3.24}$$

The first term in the right side of this expression can be rewritten this way.

$$-\frac{1}{2T}\sum_t \left(Y_t^T\Delta Y_t - 2Y_t^T\Delta X_t\beta + \beta^T X_t^T\Delta X_t\beta\right)$$

We can then determine if the terms in this expression will converge to a finite limit. If Z_t is a covariance-stationary process with moments $E(Z_t) = \mu$ and $E(Z_t - \mu)(Z_{t-j} - \mu) = \gamma_j$ for all t, where the $\gamma_j,\ j = 0, 1, 2, ...$ are such that $\sum_{j=0}^{\infty}|\gamma_j| < \infty$, then $\bar{Z_T} \overset{m.s.}{\to} \mu$ [10]. If we make these assumptions, then note that

$$\begin{aligned}\frac{1}{T}\sum_t Y_t^T \Delta Y_T &= \frac{1}{T}\sum_t \mathrm{tr}(Y_t^T \Delta Y_t) \\ &= \sum_t \left(\frac{Y_t^T \otimes Y_t^T}{T}\right) \mathrm{vec}(\Delta),\end{aligned}$$

and $\sum_t (Y_t^T \otimes Y_t^T)/T$ will converge almost surely by the Law of Large Numbers to $E(Y^T \otimes Y^T)$. Secondly,

$$\begin{aligned}\frac{1}{T}\sum_t Y_t^T \Delta X_t \beta &= \frac{1}{T}\sum_t \mathrm{tr}(\beta Y_t^T \Delta X_t) \\ &= \beta^T \sum_t \left(\frac{X_t^T \otimes Y_t^T}{T}\right) \mathrm{vec}(\Delta),\end{aligned}$$

and $\sum_t (X_t^T \otimes Y_t^T)/T$ will converge almost surely by the Law of Large Numbers to $E(X^T \otimes Y^T)$. Thirdly,

$$\begin{aligned}\frac{1}{T}\sum_t \beta^T X_t^T \Delta X_t \beta &= \frac{1}{T}\sum_t \mathrm{tr}(\beta\beta^T X_t^T \Delta X_t) \\ &= \mathrm{vec}^T(\beta\beta^T) \sum_t \left(\frac{X_t^T \otimes X_t^T}{T}\right) \mathrm{vec}(\Delta),\end{aligned}$$

and $\sum_t (X_t^T \otimes X_t^T)/T$ will converge almost surely by the Law of Large Numbers to $E(X^T \otimes X^T)$. We can now express (3.24) in this way:

$$\frac{1}{2}\Bigg[\mathrm{vec}^T(I)\left\{\frac{\sum_t Y_t^T\otimes Y_t^T}{T}-E(Y^T\otimes Y^T)\right\}\mathrm{vec}(\Delta_0-\Delta)$$
$$+2\beta_0^T\left\{\frac{\sum_t X_t^T\otimes Y_t^T}{T}-E(X^T\otimes Y^T)\right\}\mathrm{vec}(\Delta_0)$$
$$-2\beta^T\left\{\frac{\sum_t X_t^T\otimes Y_t^T}{T}-E(X^T\otimes Y^T)\right\}\mathrm{vec}(\Delta)$$
$$+\mathrm{vec}^T(\beta_0\beta_0^T)\left\{\frac{\sum_t X_t^T\otimes X_t^T}{T}-E(X^T\otimes X^T)\right\}\mathrm{vec}(\Delta_0)$$
$$-\mathrm{vec}^T(\beta\beta^T)\left\{\frac{\sum_t X_t^T\otimes X_t^T}{T}-E(X^T\otimes X^T)\right\}\mathrm{vec}(\Delta)\Bigg]\ \overset{a.s.}{\to}\ 0$$

This completes the proof.

Remark: The penalty term would add $\left\{(\lambda/T)(\beta^TA^TA\beta-\beta_0^TA^TA\beta_0\right\}$ to the left side of (3.21) and the right side of (3.24), which is uniformly $o_P(1)$ as $T\to\infty$ for $\theta\in K$ and hence negligible.

3.7 Asymptotic distribution of $\widehat{\beta}$

Let's now have a look at $\widehat{\beta}$.

$$\begin{aligned}\widehat{\beta} &= \left\{\sum_t(X_t^T\widehat{\Delta}X_t)+\lambda A^TA\right\}^{-1}\sum_t X_t^T\widehat{\Delta}Y_t\\ &= \left\{\frac{\sum_t X_t^T\widehat{\Delta}X_t}{T}+\frac{\lambda}{T}A^TA\right\}^{-1}\frac{\sum_t X_t^T\widehat{\Delta}X_t}{T}\beta+\\ &\quad\left\{\frac{\sum_t X_t^T\widehat{\Delta}X_t}{T}+\frac{\lambda}{T}A^TA\right\}\frac{\sum_t X_t^T\widehat{\Delta}\epsilon_t}{T}\end{aligned}$$

Consider

$$\begin{aligned}\frac{\sum_t X_t^T \widehat{\Delta} X_t}{T} &= \frac{\sum_t X_t^T (\widehat{\Delta} - \Delta + \Delta) X_t}{T} \\ &= \frac{\sum_t X_t^T (\widehat{\Delta} - \Delta) X_t}{T} + \frac{\sum_t X_t^T \Delta X_t}{T}.\end{aligned}$$

The first term on the right side of this expression, $\frac{1}{T}\text{vec}(\sum_t X_t^T(\widehat{\Delta} - \Delta)X_t) = \frac{1}{T}\sum_t X_t^T \otimes X_t^T \text{vec}(\widehat{\Delta} - \Delta)$ tends to zero in probability because $\frac{1}{T}\sum_t X_t^T \otimes X_t^T$ approaches some limit if T is large enough almost surely and $\text{vec}(\widehat{\Delta} - \Delta)$ goes to zero in probability. Along those same lines, the second term in the right side converges to some matrix Q almost surely. If we allow $\lambda/T = o_p(1)$, we can write

$$\sqrt{T}\left\{\widehat{\beta} - \left(Q + \frac{\lambda}{T}A^T A\right)^{-1} Q\beta_0\right\} = \left(Q + \frac{\lambda}{T}A^T A\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_t X_t^T \Delta \epsilon_t\right) + o_p(1).$$

Using the Martingale Central Limit Theorem [10], we can state this resulting theorem.

Theorem 3.7.1. *$\sqrt{T}(\widehat{\beta} - \beta_0)$ is asymptotically normally distributed with mean zero and variance-covariance matrix $\Gamma = E(X^T \Delta X)$ if*

1. *$\{X_t\}$ is stationary with finite fourth moments and $vec(vec(X_t) \otimes vec(X_t))$ has finite absolutely summable autocorrelations,*

2. *ϵ_t are independent of $X_t, X_{t-1}, ...$ and $\epsilon_{t-1}, \epsilon_{t-2}, ...$ and of zero mean and covariance matrix Ω, and*

3. *$\lambda = o_P(T)$.*

The asymptotic bias is $\left\{\left(Q + \frac{\lambda}{T}A^T A\right)^{-1} Q - I\right\}\beta_0$. This bias is zero if $A\beta_0 = 0$.

To show that the asymptotic bias is zero if $A\beta_0 = 0$, We first show that

$$\left(Q + \frac{\lambda}{T}A^TA\right)^{-1} = Q^{-1} - Q^{-1}\left(I + \frac{\lambda}{T}A^TAQ^{-1}\right)^{-1}\frac{\lambda}{T}A^TAQ^{-1}. \quad (3.25)$$

Inserting (3.25) into the bias gives us

$$\begin{aligned}\left\{\left(Q + \frac{\lambda}{T}A^TA\right)^{-1}Q - I\right\}\beta_0 &= \left\{I - Q^{-1}\left(I + \frac{\lambda}{T}A^TAQ^{-1}\right)^{-1}\frac{\lambda}{T}A^TA - I\right\}\beta_0 \\ &= Q^{-1}\left(I + \frac{\lambda}{T}A^TAQ^{-1}\right)^{-1}\frac{\lambda}{T}A^TA\beta_0, \quad (3.26)\end{aligned}$$

which is zero if $A\beta_0 = 0$. Also, it follows from (3.26) that the bias is $O(1/T)$.

Remark: Γ can be estimated by $\widehat{\Gamma} = \left(Q + \frac{\lambda}{T}A^TA\right)^{-1}\frac{\sum_t X_t^T\widehat{\Delta}X_t}{T}\left(Q + \frac{\lambda}{T}A^TA\right)^{-1}$, which may be more accurate for the finite sample case.

The distribution of $\widehat{\Delta}^{-1} = \widehat{\Omega}$ can also be derived under the conditions of Theorem 3.7.1.

$$
\begin{aligned}
\widehat{\Omega} &= \frac{1}{T}\sum_{t=1}^{T}(Y_t - X_t\widehat{\beta})(Y_t - X_t\widehat{\beta})^T \\
&= \frac{1}{T}\sum_{t=1}^{T}(X_t\beta_0 + \epsilon_t - X_t\widehat{\beta})(X_t\beta + \epsilon_t - X_t\widehat{\beta})^T \\
&= \frac{1}{T}\sum_{t=1}^{T}\Big\{\epsilon_t\epsilon_t^T + (X_t\beta_0 - X_t\widehat{\beta})\epsilon_t^T + \epsilon_t(X_t\beta_0 - X_t\widehat{\beta})^T + \\
&\quad (X_t\beta_0 - X_t\widehat{\beta})(X_t\beta_0 - X_t\widehat{\beta})^T\Big\} \\
\Rightarrow \sqrt{T}(\widehat{\Omega} - \Omega) &= \sqrt{T}\sum_{t=1}^{T}\left\{\left(\frac{\epsilon_t\epsilon_t^T}{T} - \Omega\right) + \frac{(X_t\beta_0 - X_t\widehat{\beta})\epsilon_t^T}{T} + \frac{\epsilon(X_t\beta_0 - X_t\widehat{\beta})^T}{T}\right. \\
&\quad \left. + \frac{(X_t\beta_0 - X_t\widehat{\beta})(X_t\beta_0 - X_t\widehat{\beta})^T}{T}\right\}
\end{aligned}
$$

The second term on the right side of the final expression is rewritten under a vec operation as

$$
\begin{aligned}
\text{vec}\left\{\sqrt{T}\sum_{t=1}^{T}\frac{(X_t\beta_0 - X_t\widehat{\beta})\epsilon_t^T}{T}\right\} &= \sqrt{T}\sum_{t=1}^{T}\text{vec}\left\{\frac{X_t(\beta_0 - \widehat{\beta})\epsilon_t^T}{T}\right\} \\
&= \left(\frac{\sum_{t=1}^{T}\epsilon_t \otimes X_t}{T}\right)\left\{\sqrt{T}(\beta_0 - \widehat{\beta})\right\} \\
&= \left\{\mathbf{J}_{SN\times S}o_p(1)\right\}\left\{\mathbf{J}_{S\times 1}O_p(1)\right\} \\
&= \mathbf{J}_{SN\times 1}o_p(1),
\end{aligned}
$$

where $\mathbf{J}_{m\times n}$ is a $m \times n$ matrix of ones and N is the number of rows in X_t. The third term is just the transpose of the of the second term. The fourth term can be rewritten under a vec operation as

$$
\text{vec}\left\{\sqrt{T}\sum_{t=1}^{T}\frac{(X_t\beta_0 - X_t\widehat{\beta})(X_t\beta_0 - X_t\widehat{\beta})^T}{T}\right\}
$$

$$
\begin{aligned}
&= \sqrt{T}\sum_{t=1}^{T}\text{vec}\left\{\frac{X_t(\beta_0-\widehat{\beta})(\beta_0-\widehat{\beta})^T X_t^T}{T}\right\} \\
&= \left\{\frac{\sum_{t=1}^{T} X_t\otimes X_t}{T}\right\}\left[\sqrt{T}\text{vec}\left\{(\beta_0-\widehat{\beta})(\beta_0-\widehat{\beta})^T\right\}\right] \\
&= \{\mathbf{J}_{N^2\times S^2}O_p(1)\}\left\{\mathbf{J}_{S^2\times 1}O_p\left(1/\sqrt{T}\right)\right\} \\
&= \mathbf{J}_{N^2\times 1}O_p\left(1/\sqrt{T}\right) \\
&= \mathbf{J}_{N^2\times 1}o_p(1).
\end{aligned}
$$

So now we can state that

$$\sqrt{T}(\widehat{\Omega}-\Omega) = \sqrt{T}\sum_{t=1}^{T}\left(\frac{\epsilon_t\epsilon_t^T}{T}-\Omega\right)+o_p(\mathbf{J}_{S\times S}),$$

and

$$\widehat{\Omega} = \frac{\sum_t \epsilon_t\epsilon^T}{T}+o_P\left(\frac{1}{\sqrt{T}}\right).$$

Therefore the correlations derived from $\widehat{\Omega}$ have the same asymptotic distribution as if they were derived directly from ϵ_t. In particular, the correlation estimates were asymptotically normal with zero mean and variance $\frac{1}{T}$ under the null hypothesis that Ω is diagonal.

CHAPTER 4
A MIXED-EFFECT TRANSFER FUNCTION MODEL

In the cod spawning analysis, the fixed effect common transfer function model assumes that the ψ's are fixed for all fjords. Because the fjords are located differently and are of variable geographic characteristics, the ψ's are generally fjord-specific. However introducing a current effect for every day and for each fjord will cause multicollinearity problem, so it may be the case that we believe that the ψ's are normally distributed with mean ϕ which is a smooth function. That would facilitate the need for a mixed-effect model. In Section 4.1 we will introduce a hierarchical regression model that subsumes the mixed-effect transfer function model. We shall develop the penalized likelihood estimation of the new model in Section 4.2. We shall give other specifications of our model along with their respective estimations in Section 4.3. In Section 4.4 we will define another specification called the multiplicative error model. In Section 4.5 we will develop the estimation of confidence bands. Lastly, in Section 4.6 we will fit the mixed-effect transfer function model to the cod spawning data using the various specifications that were defined in earlier sections.

4.1 A hierarchical regression model

We consider the following mixed-effect regression model

$$Y_t = X_t\psi + \epsilon_t, t = 1, \ldots, T, \tag{4.1}$$

where Y_t is the response vector, X_t is the covariate vector, $\psi = [\psi_1^T \ \ldots \ \psi_S^T]^T$ with the

ψ_s's being independent and identically distributed normal random vectors with mean vector ϕ and variance covariance matrix Γ, and ϵ_t are independent and identically distributed normal random vectors with mean 0 and variance-covariance matrix Ω. (Note that some components of ψ_s can be made into fixed-effect parameters by formally specifying infinite variance, i.e. a flat prior, for these components.) This model subsumes the mixed-effect transfer function model defined by (1.4). For example, let $Y_{t,s} = \kappa_s + \zeta_s b_t + \sum_{d=1}^{D} \psi_{s,d} b_t c_{d,t} + \epsilon_{t,s}$, $t = 1, \ldots, T$ and $s = 1, \ldots, S$. We can combine each intercepts term, b_t's and $c_{d,t}$'s to make $X_{t,s}$, Define Y_t and X_t where $t = 1, \ldots, T$ as

$$Y_t = \begin{pmatrix} Y_{t,1} \\ Y_{t,2} \\ \vdots \\ Y_{t,S} \end{pmatrix}, \text{ and}$$

$$X_t = \begin{pmatrix} X_{t,1} & 0 & \cdots & 0 \\ 0 & X_{t,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_{t,S} \end{pmatrix}.$$

Then $Y_t = X_t \psi + \epsilon_t$, where $\psi = [\psi_1^T \ \psi_2^T \ \cdots \ \psi_S^T]^T$ and $\psi_s = [\kappa_s \ \zeta_s \ \psi_{s,1} \ \cdots \psi_{s,D}]^T$.

The likelihood of the regression model defined by (4.1) is intractable as it involves high-dimensional integration needed to integrate out the unknown ψ's. How-

ever if the ψ's are known, the likelihood is tractable. This motivates us to treat ψ as a missing variable, and make use of the complete-incomplete data framework, specifically the EM-algorithm, to do maximum likelihood estimation. Another problem with this model is that ϕ may be high dimensional. The latter problem can be circumvented by imposing some smoothness constraints on ϕ. Specifically, let $||A\phi|| = \phi^T A^T A\phi$ measure the roughness of ϕ. The penalized complete data log-likelihood is given by the equation below.

$$\begin{aligned}
\ell_{\mathbf{Y},\mathbf{X},\psi}\left(\phi,\Omega,\Gamma,\lambda\right) &= -\frac{T}{2}\log|\Omega| - \frac{S}{2}\log|\Gamma| \\
&- \frac{1}{2}\sum_{t=1}^{T}\left(Y_t - X_t\psi\right)^T\Omega^{-1}\left(Y_t - X_t\psi\right) \\
&- \frac{1}{2}\sum_{s=1}^{S}\left(\psi_s - \phi\right)^T\Gamma^{-1}\left(\psi_s - \phi\right) \\
&- \frac{\lambda}{2}\phi^T A^T A\phi,
\end{aligned}$$

where λ is the smoothing parameter and the analysis is conditional on $\mathbf{X}$. Hence we sometimes write $\ell_{\mathbf{Y},\psi}$ for $\ell_{\mathbf{Y},\mathbf{X},\psi}$.

We can adapt the EM algorithm [6] to maximize the (observed) likelihood and find the maximum likelihood estimates. To implement the EM algorithm. We will first need to find the conditional distribution of the ψ_j's given the data. The conditional distribution of ψ_t given the data will be derived by an iterative scheme, which is essentially the Kalman filter algorithm. Let $\mathcal{F}_i = \{X_t;\ t = 1,\ldots,T\}\bigcup\{Y_t;\ t = 1,\ldots,i\}$, where $i = 1,\ldots,T$ and $\mathcal{F}_0 = \{X_t;\ t = 1,\ldots,T\}$. We shall denote the conditional distribution of the ψ_i's given $\mathcal{F}_i$ by $N(\widehat{\mu}_i,\widehat{\Sigma}_i)$, where $\widehat{\mu}_0 = 1_S\otimes\phi$ and $\widehat{\Sigma}_0 = I_S\otimes\Gamma$.

Then, given $\mathcal{F}_{i-1}$,

$$\begin{pmatrix} Y_i \\ \\ \psi, \end{pmatrix} \sim N\left\{ \begin{pmatrix} X_i\widehat{\mu}_{i-1} \\ \\ \widehat{\mu}_{i-1} \end{pmatrix}, \begin{pmatrix} \Omega + X_i\widehat{\Sigma}_{i-1}X_i^T & X_i\widehat{\Sigma}_{i-1} \\ \\ \widehat{\Sigma}_{i-1}X_i^T & \widehat{\Sigma}_{i-1} \end{pmatrix} \right\}$$

The $\widehat{\mu}_i$ and $\widehat{\Sigma}_i$ can be updated by computing the mean and variance of the conditional distribution of $\psi|\mathcal{F}_i$, which is again normal with mean

$$\widehat{\mu}_i = \widehat{\mu}_{i-1} + \widehat{\Sigma}_{i-1}X_i^T\left(\Omega + X_i\widehat{\Sigma}_{i-1}X_i^T\right)^{-1}(Y_i - X_i\widehat{\mu}_{i-1})$$

and variance-covariance matrix

$$\widehat{\Sigma}_i = \widehat{\Sigma}_{i-1} - \widehat{\Sigma}_{i-1}X_i^T\left(\Omega + X_i\widehat{\Sigma}_{i-1}X_i^T\right)^{-1}X_i\widehat{\Sigma}_{i-1}.$$

After doing this for all of the Y_t's, we can get the distribution of ψ given the Y's and X's, i.e.

$$\psi|\mathcal{F}_T \sim N\left(\widehat{\mu}_T, \widehat{\Sigma}_T\right)$$

Let $\mathbf{Y}$ be the set of all Y_t and $\mathbf{X}$ be the set of all X_t.

4.2 The EM algorithm

The EM algorithm is a method that can be used to maximize an intractable likelihood, e.g. due to the fact some data is unobserved. The EM requires two steps:

(1) impute the complete data log-likelihood by $Q\left(\theta|\theta^{(k)}\right) = E(\ell(\theta)|\mathbf{Y}, \mathbf{X}, \psi)$, where θ is a vector containing the parameters and $\theta^{(k)}$ is the kth iterate of θ, and (2) update $\theta^{(k)}$ by the θ which maximizes $Q\left(\cdot|\theta^{(k)}\right)$. These steps would be repeated until some sort of stopping criteria has been reached. The EM algorithm has the property that $\ell(\theta^{(k+1)}) \geq \ell(\theta^{(k)})$, meaning that the likelihood gets larger for each iteration [6].

Let $\theta = (\phi, \Gamma, \Omega)$ and write

$$\ell(\theta) = \ell_{\mathbf{Y},\mathbf{X},\phi}(\phi, \Omega, \Gamma; \lambda),$$

where $\ell_{\mathbf{Y},\mathbf{X},\phi}(\phi, \Omega, \Gamma; \lambda)$ is the complete data log-likelihood. Both steps of the EM algorithm can be easily done by updating θ component-wise, i.e. we first update ϕ and Γ given that Ω is fixed, and then we update Ω. To get estimates for ϕ and Γ, we must maximize the imputed log-likelihood. Write $E(\cdot)$ for $E(\cdot|\mathbf{Y}, \mathbf{X}, \psi)$. First,

$$\begin{aligned}
E\left\{\sum_s (\psi_s - \phi)^T \Gamma^{-1} (\psi_s - \phi)\right\} &= \sum E\left\{(\psi_s - \phi)^T \Gamma^{-1} (\psi_s - \phi)\right\} \\
&= \sum E\left[\mathrm{tr}\left\{\Gamma^{-1}(\psi_s - \phi)(\psi_s - \phi)^T\right\}\right] \\
&= \sum \mathrm{tr}\left[\Gamma^{-1} E\left\{(\psi_s - \phi)(\psi_s - \phi)^T\right\}\right] \\
&= \sum \mathrm{tr}\left[\Gamma^{-1}\left\{(\widehat{\mu}_{s,T} - \phi)\left(\widehat{\mu}_{s.T} - \phi\right)^T + \widehat{\Sigma}_{s,T}\right\}\right] \\
&= \sum (\widehat{\mu}_{s,T} - \phi)^T \Gamma^{-1} (\widehat{\mu}_{s,T} - \phi) + \sum \mathrm{tr}(\Gamma^{-1} \widehat{\Sigma}_{s,T})
\end{aligned}$$

where $\widehat{\mu}_{s,T}$ is the sth component of the $\widehat{\mu}_T$ vector and $\widehat{\Sigma}_{s,T}$ is the sth block diagonal of the $\widehat{\Sigma}_T$ matrix. To estimate ϕ with known Γ and Ω, first note that the complete-data log-likelihood is $Q\left(\theta|\theta^{(k)}\right) = E(\ell(\theta)|\theta^{(k)}, Y, X)$ equals up to an additive constant,

$$Q(\theta|\theta^{(k)}) = -\frac{1}{2}\sum \left(\widehat{\mu}_{s,T}^{(k)} - \phi\right)^T \Gamma^{-1} \left(\widehat{\mu}_{s,T}^{(k)} - \phi\right) - \frac{\lambda}{2}\phi^T A^T A\phi$$

So in order to compute $\phi^{(k+1)}$, we compute a regression of $\Gamma^{-1/2}\widehat{\mu}_{s,T}^{(k)}$ on $\Gamma^{-1/2}$ with smoothness imposed, where the smoothing parameter λ is determined by GCV.

In a similar fashion, we will update Γ, where up to an additive constant,

$$\begin{aligned}
Q(\theta|\theta^{(k)}) &= -\frac{S}{2}\log|\Gamma| - \frac{1}{2}\sum_s \left(\widehat{\mu}_{s,T}^{(k)} - \phi\right)^T \Gamma^{-1} \left(\widehat{\mu}_{s,T}^{(k)} - \phi\right) - \frac{1}{2}\sum_s \mathrm{tr}\left(\Gamma^{-1}\widehat{\Sigma}_{s,T}^{(k)}\right) \\
&= -\frac{S}{2}\log|\Gamma| - \frac{1}{2}\sum_s \mathrm{tr}\left[\Gamma^{-1}\left\{\widehat{\Sigma}_{s,T}^{(k)} + \left(\widehat{\mu}_{s,T}^{(k)} - \phi\right)\left(\widehat{\mu}_{s,T}^{(k)} - \phi\right)^T\right\}\right] \\
&= -\frac{S}{2}\log|\Gamma| - \frac{1}{2}\mathrm{tr}\left[\Gamma^{-1}\sum_s \left\{\widehat{\Sigma}_{s,T}^{(k)} + \left(\widehat{\mu}_{s,T}^{(k)} - \phi\right)\left(\widehat{\mu}_{s,T}^{(k)} - \phi\right)^T\right\}\right].
\end{aligned}$$

So our estimate of Γ is

$$\widehat{\Gamma}^{(k+1)} = \frac{\sum_s \left\{\widehat{\Sigma}_{s,T}^{(k)} + \left(\widehat{\mu}_{s,T}^{(k)} - \phi^{(k+1)}\right)\left(\widehat{\mu}_{s,T}^{(k)} - \phi^{(k+1)}\right)^T\right\}}{S}.$$

We will also estimate Ω in this fashion. Up to an additive constant,

$$\begin{aligned}
Q(\theta|\theta^{(k)}) &= E\left\{\sum_t (Y_t - X_t\psi)^T\Omega^{-1}(Y_t - X_t\psi)\right\} \\
&= \sum E\left[\mathrm{tr}\left\{\Omega^{-1}(Y_t - X_t\psi)(Y_t - X_t\psi)^T\right\}\right] \\
&= \sum \mathrm{tr}\left[\Omega^{-1}E\left\{(Y_t - X_t\psi)\left(Y_t - X_t\psi\right)^T\right)\right\}] \\
&= \sum \mathrm{tr}\left[\Omega^{-1}\left\{\left(Y_t - X_t\widehat{\mu}_T^{(k)}\right)\left(Y_t - X_t\widehat{\mu}_T^{(k)}\right)^T + X_t\widehat{\Sigma}_T^{(k)}X_t^T\right\}\right] \\
&= \sum \left\{\left(Y_t - X_t\widehat{\mu}_T^{(k)}\right)^T\Omega^{-1}\left(Y_t - X_t\widehat{\mu}_T^{(k)}\right) + \mathrm{tr}\left(\Omega^{-1}X_t\widehat{\Sigma}_T^{(k)}X_t^T\right)\right\}.
\end{aligned}$$

The estimate of Ω is

$$\widehat{\Omega}^{(k+1)} = \frac{\sum_t \left\{ X_t \widehat{\Sigma}_T^{(k)} X_t^T + \left(Y_t - X_t \widehat{\mu}_T^{(k)} \right) \left(Y_t - X_t \widehat{\mu}_T^{(k)})^T \right) \right\}}{T}$$

We will need initial values for θ.

1. Ω will be estimated by computing the smooth regression model series by series and finding the estimate for the residual covariance matrix.

2. ϕ will be estimated by taking the mean of $S-1$ estimates found from estimating a smooth regression when each Y_s and X_s, $s = 1. \ldots, S$ is removed.

3. Γ will be estimated in several ways. First we could use a general structure by computing the covariance matrix of the ψ_s's computed in 2 and multiplying by $S-1$. Second, we can just use the diagonal of the aforementioned matrix. Third, we can use a common value for the diagonal.

4.3 Specification of Γ

As stated before, there are several ways to define Γ, the covariance matrix of the ψ_s's.

1. Γ could be a general covariance matrix. Code for the estimation assuming a general Γ can be found in Appendix B.3.1.

2. Γ could be diagonal. We would then update Γ by taking the diagonal of the general covariance matrix and fill the rest of the matrix with zeros. Code for the estimation assuming a diagonal Γ can be found in Appendix B.3.2.

3. Γ could be a multiple of the identity matrix. We would then update Γ by estimating a common value of the variance.

$$\widehat{\sigma}_{\psi}^{2(k+1)} = \frac{\sum_s \left\{ \text{tr}\left(\widehat{\Sigma}_{s,T}\right)^{(k)} + \left(\widehat{\mu}_{s,T}^{(k)} - \phi^{(k+1)}\right)^T \left(\widehat{\mu}_{s,T}^{(k)} - \phi^{(k+1)}\right) \right\}}{DS}$$

Code for the estimation assuming Γ that is a multiple of the identity matrix can be found in Appendix B.3.3.

4. Γ could have a AR(1) format.

Let's elaborate on 4. If we assume that the ψ_s's have an autoregressive relationship of order one, the Γ could have an AR(1) form. Specifically,

$$(1 - \rho B)(\psi_{s,d} - \phi_d) = a_{s,d}$$

where B is the backshift operator over the d's and $a_{s,d}$ are independent and identically distributed $N(0, \sigma_a^2)$. If that is the case, then we can modify the penalized complete likelihood as follows.

$$\begin{aligned}\ell\left(\phi,\Sigma,\rho,\sigma_a^2|\mathbf{X},\mathbf{Y},\psi\right) &= -\frac{T}{2}\log|\Omega| - \frac{DS}{2}\log\sigma_a^2 - \\ &\frac{1}{2}\sum_{t=1}^{T}\left(Y_t - X_t\psi\right)^T\Omega^{-1}\left(Y_t - X_t\psi\right) - \\ &\frac{1}{2\sigma_a^2}\sum_{s=2}^{S}\left\{\left(\psi_{s,d}-\phi_d\right)-\rho\left(\psi_{s,d-1}-\phi_{d-1}\right)\right\}^2 - \\ &\frac{1}{2\sigma_a^2}(1-\rho^2)(\psi_{s,1}-\phi_1)^2 - \\ &\frac{\lambda}{2}\phi^T A^T A\phi\end{aligned}$$

To estimate ρ and σ_a^2 with other parameter being known, then first we need to compute the E-step. Up to an additive constant and letting the (i,j)th element of the sth block of $\widehat{\Sigma}_T$ be denoted as $\widehat{\Sigma}_{s,i,j}$,

$$\begin{aligned}-Q(\theta|\theta^{(k)}) &= E\left\{\frac{1}{2\sigma_a^2}\sum_{s=1}^{S}\sum_{d=2}^{D}\left(\psi_{s,d}-\phi_d\right)-\rho\left(\psi_{s,d-1}-\phi_{d-1}\right)\right\}+\frac{DS}{2}\log\sigma_a^2 \\ &= \frac{1}{2\sigma_a^2}\sum_{s=1}^{S}\sum_{d=2}^{D}\left[\left\{\widehat{\psi}_{s,d}^{(k)}-\phi_d-\rho\left(\widehat{\psi}_{s,d-1}^{(k)}-\phi_{d-1}\right)\right\}^2+\right. \\ &\quad \left.\mathrm{Var}(\psi_{s,d})+\rho^2\mathrm{Var}(\psi_{s,d-1})-2\rho\mathrm{Cov}(\psi_{s,d},\psi_{s,d-1})\right]+\frac{DS}{2}\log\sigma_a^2 \\ &= \frac{1}{2\sigma_a^2}\sum_{s=1}^{S}\sum_{d=2}^{D}\left[\left\{\left(\widehat{\psi}_{s,d}^{(k)}-\phi_d\right)^2-2\rho\left(\widehat{\psi}_{s,d}^{(k)}-\phi_d\right)\left(\widehat{\psi}_{s,d-1}^{(k)}-\phi_{d-1}\right)+\right.\right. \\ &\quad \left.\left.\rho^2\left(\widehat{\psi}_{s,d-1}^{(k)}-\phi_{d-1}\right)^2+\widehat{\Sigma}_{s,d,d}^{(k)}-2\rho\widehat{\Sigma}_{s,d,d-1}^{(k)}+\rho^2\widehat{\Sigma}_{s,d-1,d-1}^{(k)}\right\}\right]+ \\ &\quad \frac{DS}{2}\log\sigma_a^2 \\ &= \frac{1}{2\sigma_a^2}\sum_{s=1}^{S}\sum_{d=2}^{D}\left[\rho^2\left\{\left(\widehat{\psi}_{s,d-1}^{(k)}-\phi_{d-1}\right)^2+\widehat{\Sigma}_{s,d-1,d-1}\right\}-\right. \\ &\quad 2\rho\left\{\left(\widehat{\psi}_{s,d}^{(k)}-\phi_d\right)\left(\widehat{\psi}_{s,d-1}^{(k)}-\phi_{d-1}\right)+\right. \\ &\quad \left.\left.\widehat{\Sigma}_{s,d,d-1}^{(k)}\right\}+\left\{\left(\widehat{\psi}_{s,d}^{(k)}-\phi_d\right)^2+\widehat{\Sigma}_{s,d,d}^{(k)}\right\}\right]+\frac{DS}{2}\log\sigma_a^2 \qquad (4.2)\end{aligned}$$

Write the right hand side of 4.2 by $\frac{1}{2\sigma_a^2}RSS(\rho)+\frac{DS}{2}\log(\sigma_a^2)$. It can be readily checked that

$$\begin{aligned}\widehat{\rho}^{(k+1)} &= \frac{\sum_{s=1}^{S}\sum_{d=2}^{D}\left\{\left(\widehat{\psi}_{s,d}^{(k)}-\widehat{\phi}_{d}^{(k+1)}\right)\left(\widehat{\psi}_{s,d-1}^{(k)}-\widehat{\phi}_{d-1}^{(k+1)}\right)+\widehat{\Sigma}_{s,d,d-1}^{(k)}\right\}}{\sum_{s=1}^{S}\sum_{d=2}^{D}\left\{\left(\widehat{\psi}_{s,d-1}^{(k)}-\widehat{\phi}_{d-1}^{(k+1)}\right)^2+\widehat{\Sigma}_{s,d-1,d-1}^{(k)}\right\}} \\ (\widehat{\sigma}_a^2)^{(k+1)} &= \frac{1}{DS}RSS(\widehat{\rho}^{(k+1)})\end{aligned}$$

4.4 A multiplicative error model

One potential problem with the preceding specification of Γ is that $\psi_{s,d}$ need not be small even if its corresponding mean ϕ_d is close to zero. Consequently, the $\psi_{s,d}$ may fluctuate too much. To alleviate this problem, we consider another specification of Γ which constrains ψ_s to vary less when ϕ is closer to zero. This can be done by using a multiplicative specification:

$$\psi_{s,d}=\phi_d e^{\eta_{s,d}};\, s=1,\ldots,S;\, d=1,\ldots,D \tag{4.3}$$

where

$$\eta_{s,d}\overset{iid}{\sim}N(0,\sigma_\eta^2).$$

Since this specification would be difficult to implement using the EM algorithm, we can approximate Equation (4.3) by assuming that $\eta_{s,d}$ is small:

$$\begin{aligned}\psi_{s,d} &\approx \phi_d(1+\eta_{s,d}) \\ &= \phi_d+\phi_d\eta_{s,d} \end{aligned}\tag{4.4}$$

From this approximation, we can state that

$$\psi_{s,d} \overset{iid}{\sim} N(\phi_d, \phi_d^2 \sigma_\eta^2).$$

Recall that $\psi_s = [\alpha_s^T \ \psi_{s,1} \ \cdots \ \psi_{s,D}]^T$. For simplicity, we will assume that α is two-dimensional (as in the cod specification), and

$$\psi_s \overset{iid}{\sim} N(\phi, \Gamma(\phi, \sigma_1^2, \sigma_2^2, \sigma_\eta^2)),$$

where $\phi = [\kappa \ \zeta \ \phi_1 \ \cdots \ \phi_D]^T$,

$$\Gamma(\phi, \sigma_1^2, \sigma_2^2, \sigma_\eta^2) = \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_\eta^2 \phi_1^2 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \sigma_\eta^2 \phi_2^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \sigma_\eta^2 \phi_D^2 \end{bmatrix}$$

So now we can display the penalized log-likelihood function.

$$\begin{aligned} \ell(\phi, \sigma_\eta^2, \Omega) \ &= \ -\frac{T}{2} \log |\Omega| - \frac{1}{2} \sum_{t=1}^{T} (Y_t - X_t \psi)^T \Omega^{-1} (Y_t - X_t \psi) - \frac{S}{2} \log |\Gamma| - \\ &\quad \frac{1}{2} \sum_{s=1}^{S} (\psi_s - \phi)^T \Gamma^{-1} (\psi_s - \phi) - \frac{1}{2} \lambda \phi^T A^T A \phi \end{aligned} \tag{4.5}$$

We shall develop an EM-like algorithm to estimate the parameters ϕ, $\Gamma(\phi, \sigma_1^2, \sigma_2^2, \sigma_\eta^2)$, and Ω. The estimation of ϕ strictly using the EM algorithm would be difficult since ϕ occurs in both the mean function and the variance-covariance matrix. However, updating ϕ with Γ and Ω fixed at their previous values will be feasible because then ϕ only occurs in the mean function, and hence the updating formula for ϕ is similar to that derived earlier. The updating of Ω presents no new difficulty. Now to estimate Γ, we have to compute estimates for σ_1^2, σ_2^2, and σ_η^2. First, we will compute the expectation of the relevant parts of the likelihood function given the observed data.

$$\begin{aligned} Q(\theta|\theta^{(k)}) &= E\left\{-\frac{S}{2}\log|\Gamma| - \frac{1}{2}\sum_s(\psi_s-\phi)^T\Gamma^{-1}(\psi_s-\phi)\right\} \\ &= -\frac{S}{2}\log|\Gamma| - \frac{1}{2}\sum_s\left\{(\widehat{\mu}_{s,T}^{(k)}-\phi)^T\Gamma^{-1}(\widehat{\mu}_{s,T}^{(k)}-\phi) - \frac{1}{2}\mathrm{tr}(\Gamma^{-1}\widehat{\Sigma}_{s,T}^{(k)})\right\} \end{aligned}$$

Then we will maximize $Q(\theta|\theta^{(k)})$. To do so, we will compute the first differential and set it equal to zero. Note that

$$\Gamma^{-1} = \begin{bmatrix} 1/\sigma_1^2 & 0 & 0 & \cdots & 0 \\ 0 & 1/\sigma_2^2 & 0 & \cdots & 0 \\ 0 & 0 & 1/\phi_1^2\sigma_\eta^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1/\phi_D^2\sigma_\eta^2 \end{bmatrix}, \text{ and}$$

$$d\Gamma = \begin{bmatrix} d\sigma_1^2 & 0 & 0 & \cdots & 0 \\ 0 & d\sigma_2^2 & 0 & \cdots & 0 \\ 0 & 0 & 2\phi_1 d\phi_1 \sigma_\eta^2 + \phi_1^2 d\sigma_\eta^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 2\phi_D d\phi_D \sigma_\eta^2 + \phi_D^2 d\sigma_\eta^2 \end{bmatrix}.$$

Thus, up to an additive constant,

$$dQ(\theta|\theta^{(k)})$$

$$
\begin{aligned}
&= -\frac{S}{2}\mathrm{tr}(\Gamma^{-1}d\Gamma) + d\phi^T\Gamma^{-1}\sum_s(\widehat{\mu}_{s,T}-\phi) + \\
&\quad \frac{1}{2}\sum_s(\widehat{\mu}_{s,T}-\phi)^T\Gamma^{-1}d\Gamma\Gamma^{-1}(\widehat{\mu}_{s,T}-\phi) + \frac{1}{2}\mathrm{tr}(\Gamma^{-1}d\Gamma\Gamma^{-1}\widehat{\Sigma}_{s,T}) \\
&= -\frac{S}{2}\mathrm{tr}\left(\begin{bmatrix} \frac{d\sigma_1^2}{\sigma_1^2} & 0 & 0 & \cdots & 0 \\ 0 & \frac{d\sigma_2^2}{\sigma_2^2} & 0 & \cdots & 0 \\ 0 & 0 & \frac{2\phi_1 d\phi_1\sigma_\eta^2+\phi_1^2 d\sigma_\eta^2}{\phi_1^2\sigma_\eta^2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{2\phi_D d\phi_D\sigma_\eta^2+\phi_D^2 d\sigma_\eta^2}{\phi_D^2\sigma_\eta^2} \end{bmatrix}\right) + \\
&\quad \begin{bmatrix} \frac{d\kappa}{\sigma_1^2} & \frac{d\zeta}{\sigma_2^2} & \frac{d\phi_1}{\phi_1^2\sigma_\eta^2} & \cdots & \frac{d\phi_D}{\phi_D^2\sigma_\eta^2} \end{bmatrix}\sum_s(\widehat{\mu}_{s,T}-\phi) + \\
&\quad \frac{1}{2}\sum_s(\widehat{\mu}_{s,T}-\phi)^T\begin{bmatrix} \frac{d\sigma_1^2}{\sigma_1^4} & 0 & 0 & \cdots & 0 \\ 0 & \frac{d\sigma_2^2}{\sigma_2^4} & 0 & \cdots & 0 \\ 0 & 0 & \frac{2\phi_1 d\phi_1\sigma_\eta^2+\phi_1^2 d\sigma_\eta^2}{\phi_1^4\sigma_\eta^4} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{2\phi_D d\phi_D\sigma_\eta^2+\phi_D^2 d\sigma_\eta^2}{\phi_D^4\sigma_\eta^4} \end{bmatrix}(\widehat{\mu}_{s,T}-\phi) \\
&\quad + \frac{1}{2}\mathrm{tr}\left(\begin{bmatrix} \frac{d\sigma_1^2}{\sigma_1^4} & 0 & 0 & \cdots & 0 \\ 0 & \frac{d\sigma_2^2}{\sigma_2^4} & 0 & \cdots & 0 \\ 0 & 0 & \frac{2\phi_1 d\phi_1\sigma_\eta^2+\phi_1^2 d\sigma_\eta^2}{\phi_1^4\sigma_\eta^4} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{2\phi_D d\phi_D\sigma_\eta^2+\phi_D^2 d\sigma_\eta^2}{\phi_D^4\sigma_\eta^4} \end{bmatrix}\widehat{\Sigma}_{s,T}\right).
\end{aligned}
$$

Now, only looking at the parts that contain $d\sigma_1^2$, $d\sigma_2^2$, and $d\sigma_\eta^2$, we come up with the following.

$$
\begin{aligned}
dQ(\theta|\theta^{(k)}) &= -\frac{S}{2}\left(\frac{d\sigma_1^2}{\sigma_1^2}+\frac{d\sigma_2^2}{\sigma_2^2}+\frac{Dd\sigma_\eta^2}{\sigma_\eta^2}\right)+ \\
&\quad \frac{1}{2}\left[\frac{d\sigma_1^2\sum_s\left\{(\widehat{\mu}_{s,1,T}^{(k)}-\kappa)^2+\widehat{\Sigma}_{s,1,T}^{(k)}\right\}}{\sigma_1^4}+\right. \\
&\quad \frac{d\sigma_2^2\sum_s\left\{(\widehat{\mu}_{s,d,T}^{(k)}-\zeta)^2+\widehat{\Sigma}_{s,2,T}^{(k)}\right\}}{\sigma_2^4}+ \\
&\quad \left.\sum_d\frac{d\sigma_\eta^2\sum_s\left\{(\widehat{\mu}_{s,d,T}^{(k)}-\phi_d)^2+\widehat{\Sigma}_{s,d,T}^{(k)}\right\}}{\phi_d^2\sigma_1^4}\right],
\end{aligned}
$$

where $\widehat{\mu}_{s,d,T}^{(k)}$ is the dth element of $\widehat{\mu}_{s,T}^{(k)}$, ϕ_d is the dth element of ϕ, and $\widehat{\Sigma}_{s,d,T}^{(k)}$ is the dth diagonal element in the sth block of $\widehat{\Sigma}_T^{(k)}$. This can also be derived directly from routine analysis.

First, we will find the estimate for σ_1^2.

$$
\frac{dQ}{d\sigma_1^2}=-\frac{S}{2\sigma_1^2}+\frac{\sum_s\left\{(\widehat{\mu}_{1,s,T}^{(k)}-\kappa)^2+\widehat{\Sigma}_{1,s,T}^{(k)}\right\}}{2\sigma_1^4}
$$

Setting this to be zero, we get

$$
(\widehat{\sigma}_1^2)^{(k+1)}=\frac{1}{S}\sum_{s=1}^{S}\left\{(\widehat{\mu}_{s,1,T}^{(k)}-\widehat{\kappa}^{(k+1)})^2+\widehat{\Sigma}_{s,1,T}^{(k)}\right\}.
$$

Similarly for σ_2^2, we will estimate this as

$$
(\widehat{\sigma}_2^2)^{(k+1)}=\frac{1}{S}\sum_{s=1}^{S}\left\{(\widehat{\mu}_{s,2,T}^{(k)}-\zeta^{(k+1)})^2+\widehat{\Sigma}_{s,2,T}^{(k)}\right\}.
$$

Now for σ_η^2.

$$\frac{dQ}{d\sigma_\eta^2} = -\frac{SD}{2\sigma_\eta^2} + \sum_d \sum_s \frac{(\widehat{\mu}_{s,d,T}^{(k)} - \phi_d)^2 + \widehat{\Sigma}_{s,d,T}^{(k)}}{2\phi_d^2 \sigma_\eta^2}$$

Setting this to be zero, we get

$$(\widehat{\sigma}_\eta^2)^{(k+1)} = \frac{1}{SD} \sum_{d=1}^{D} \sum_{s=1}^{S} \left\{ (\widehat{\mu}_{s,d,T}^{(k)} - \widehat{\phi}_d^{(k+1)})^2 + \widehat{\Sigma}_{s,d,T}^{(k)} \right\} / (\widehat{\phi}_d^{(k+1)})^2$$

Code used in estimation with Γ specified in the exponential error model can be found in Appendix B.3.4.

4.5 Bootstrap approach to estimating confidence bands

We now would like to compute individual confidence bands for the mean function ϕ. One method of doing this is to compute the estimated asymptotic variance-covariance matrix of $\widehat{\phi}$ [14]. This method was discussed in Appendix A. This computation will require a large number of computations, making the possibility of computational error high. Also, further theoretical investigation is needed to extend the results from Louis [14] to a situation with dependent data where there is a smoothing penalty. Another approach is to use the bootstrap to obtain the confidence bands. There are two methods that can be used: the parametric approach, which would assume the our data follows a known distribution, and the nonparametric approach, which would not make that assumption.

Recall that $\widehat{\psi}_{s,T}$ is the estimated mean effect for fjord s given all of the observed information (X_t and Y_t for $t = 1, \ldots, T$). From these we can define the fitted values

$$\widehat{Y}_s = X_s \widehat{\psi}_{s,T},\ s = 1, \ldots, S.$$

Recall that $\widehat{Y}_s$ is the vector of fitted values $(\widehat{Y}_{t,s}, t = 1, \ldots, T)$, which is the estimated response for each fjord. We can then rearrange $\widehat{Y}_s, s = 1, \ldots, S$ to $\widehat{Y}_t, t = 1, \ldots, T$, which is the estimated response for each year $(\widehat{Y}_{t,s}, s = 1, \ldots, S)$. From that we can define the vector of yearly residuals

$$\widehat{\epsilon}_t = Y_t - \widehat{Y}_t,\ t = 1, \ldots, T. \tag{4.6}$$

These $\widehat{\epsilon}_t$ would keep the spatial correlation in each year and can be used in a (partially) nonparametric bootstrap procedure defined in this way:
Repeat the following B times.

1. Take a random sample of T $\widehat{\epsilon}_t$'s with replacement, where $\widehat{\epsilon}_t$ are defined in Equation (4.6). These sampled values will be called $\epsilon_1^*, \epsilon_2^*, \ldots, \epsilon_T^*$.

2. Compute the bootstrap response vector Y_t^* as

$$Y_t^* = \widehat{Y}_t + \epsilon_t^* \tag{4.7}$$

3. Fit Y_t^* from Equation (4.7) onto X_t with Ω, Γ, and λ fixed at the values estimated from the original analysis to obtain ϕ^*.

Now that we have B bootstrapped values for ϕ, we can make a $(1-\alpha) \times 100\%$ (individual) confidence band for ϕ, i.e. for each element ϕ_i, we compute the $(\alpha/2) \times$ 100th and $(1 - \alpha/2) \times 100$th percentile of $\phi_i^*, i = 1, \ldots, B$ [7].

We can also define a parametric bootstrap in this way:

Repeat the following B times.

1. Simulate $\epsilon_t^*, t = 1, \ldots, T$ from $N(0, \widehat{\Omega})$, where $\widehat{\Omega}$ is the variance-covariance matrix estimated from the original analysis.

2. Compute the bootstrap response vector Y_t^* as

$$Y_t^* = \widehat{Y}_t + \epsilon_t^* \tag{4.8}$$

3. Fit Y_t^* from Equation (4.8) onto X_t with Ω, Γ, and λ fixed at the values estimated from the original analysis to obtain ϕ^*.

Code used to estimate model parameters for the bootstrap samples can be found in Appendix B.4.1. The nonparametric bootstrap samples are created using the code in Appendix B.4.2. The parametric bootstrap samples are created using the code in Appendix B.4.3.

4.6 Inflow of larvae cod as an example, revisited

Consider fitting the mixed-effect transfer function model for the cod spawning data analyzed in Section 3.3,

$$Y_{t,s} = \kappa_s + \zeta_s b_t + \sum_{d=1}^{D} \psi_{s,d} b_t c_{d,t} + \epsilon_{t,s}, \ t = 1, \ldots, T,$$

where b_t is the natural logarithm of the spawning biomass for year t, $c_{d,t}$ is the net inflow into the Skagerrak on day d in year t counted starting March 1st, $\psi_s =$

$[\kappa_s \ \zeta_s \ \psi_{s,1} \ \cdots \ \psi_{s,D}]$ are independent and identically distributed normal random vectors with mean ϕ and variance-covariance matrix Γ, and $\epsilon_t = [\epsilon_{t,1} \ \cdots \ \epsilon_{t,S}]$ are independent and identically distributed normal random vectors with zero mean and variance-covariance matrix Ω.

We fitted this model to the data presented in Section 3.3 using three different estimates for Γ: as a multiple of the identity matrix, as a diagonal matrix, and as a general covariance matrix. Graphs of $\widehat{\phi}$ using these three different estimates for Γ can be found in Figure 4.1. From these graphs, restricting Γ to a multiple of the identity matrix or as diagonal matrix does not lead to a smooth $\widehat{\phi}$. However, allowing Γ to be an unrestricted general variance-covariance matrix does result in $\widehat{\phi}$ as a smooth function, similar to that of Figure 3.4. This could possibly be explained by the fact that allowing Γ to be general yields an almost singular $\widehat{\Gamma}$. This fact would impose smoothness on the ϕ's. The possible drawback of using a general Γ is that the number of parameters that needs to be estimated could be quite large as compared to the sample size. However, when we computed $\widehat{\Omega}$ for the general covariance matrix model, it was nearly singular, so model checking such as residual plots and normal Q-Q plots could not be done on the standardized residuals.

The results of the confidence band based on 1,000 nonparametric bootstrap samples fitted with the estimate for the general variance-covariance matrix Γ are shown in Figure 4.2. The estimate of ϕ was hardly significant, as it was only significant at 50% confidence. The results of the confidence band based on 1,000 parametric bootstrap samples fitted with the estimate for the general variance-covariance matrix

Γ are shown in Figure 4.3. The estimate of ϕ is determined to be significant at 60% confidence. The results of the two bootstrap approaches are very similar.

Since $\widehat{\Omega}$ was nonsingular in the multiplicative error model, we were able to do some model checking. A (standardized) residual versus fitted values plot for the model fitted with the multiplicative error model can be found in Figure 4.4. Similar to the findings with the common transfer function model, both the estimates of κ and ζ are not significantly different from zero. In particular, this supports the earlier result that North Sea cod did not swim to Skagerrak to spawn there. The fitted values for year 1973 were all very small, thereby creating a gap in Figure 4.4. Otherwise, no other pattern can be detected in this plot. A residual plot for the model fitted with the multiplicative error model where residuals are plotted over year is shown in Figure 4.5. There is no apparent pattern of the residuals over the years. A normal Q-Q plot of the residuals for the model fitted with multiplicative error can be found in Figure 4.6. This plot shows no strong deviation from normality. We report the estimates for the parameters for this model in able 4.1. Estimation of σ_η^2 was sensitive to the starting value; if the value was too large, then Γ became very large in magnitude. We selected different starting values until $\widehat{\sigma}_\eta^2$ changed a small amount. Note that the between-series variance in the transfer function model is very small, suggesting that a common transfer function model may hold for this set of data. Indeed, this also raises the interesting question of testing the hypothesis of the common transfer function model, by testing whether or not $\sigma_\eta^2 = 0$, which is a hard problem because the hypothesized value is at the boundary of the parameter space.

Parameter	Estimate	95% Confidence Limits	
		Nonparametric	Parametric
κ	0.197	$(-0.221, 0.857)$	$(-0.057, 0.717)$
ζ	-0.118	$(-0.379, 0.107)$	$(-0.304, 0.025)$
σ_1^2	4.15×10^{-5}		
σ_2^2	5.26×10^{-6}		
σ_η^2	0.00499		

Table 4.1: Mixed-effect parameter estimates

The results of the confidence band based on 1,000 nonparametric bootstrap samples fitted with the estimate for the multiplicative error variance-covariance matrix Γ can be found in Figure 4.7. The mean function ϕ was determined to be significant at 70% confidence. The results of the confidence band based on 1,000 parametric bootstrap samples fitted with the estimate for the multiplicative error variance-covariance matrix Γ can be found in Figure 4.8. The mean function ϕ was determined to be significant at 90% confidence.

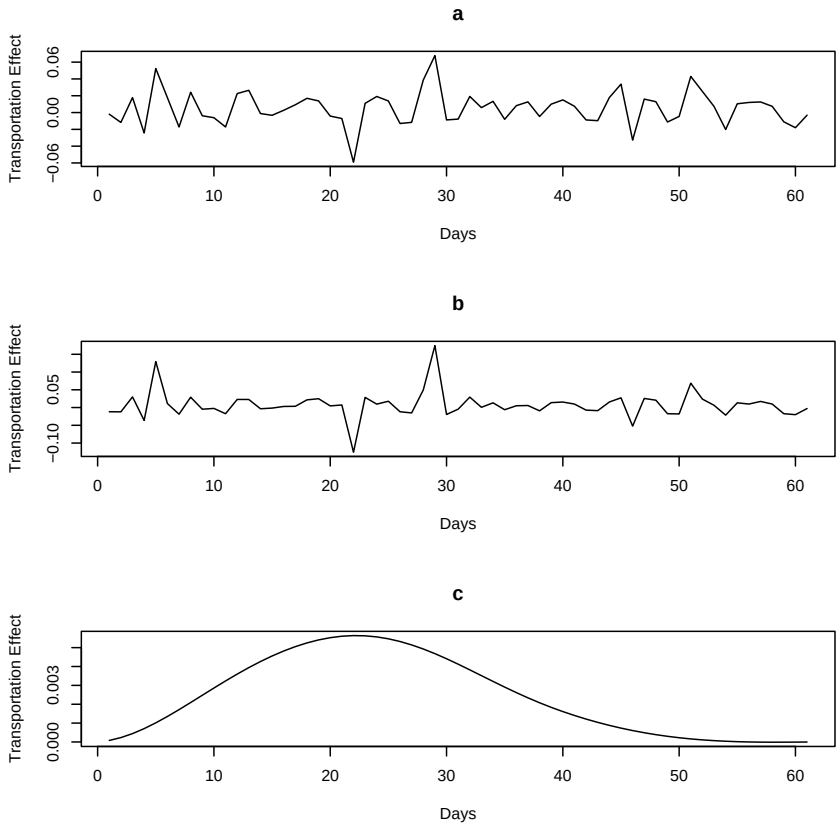

Figure 4.1: The plot of $\widehat{\phi}_j$ versus j for the North Sea cod for (a) Γ is a multiple of the identity matrix, (b) Γ is a diagonal matrix, and (c) Γ is a general variance-covariance matrix.

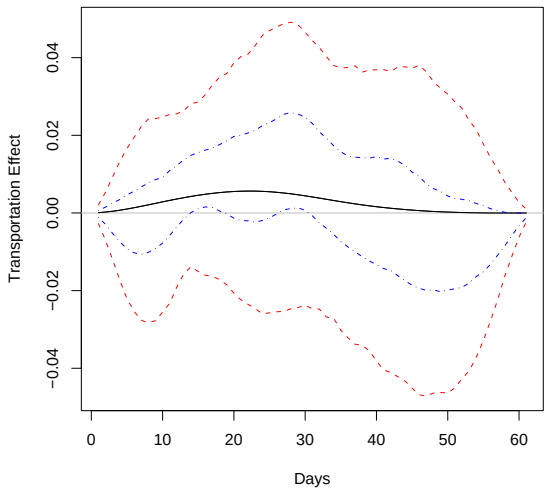

Figure 4.2: Plot of estimate for ϕ along with 95% nonparametric bootstrap confidence band (dashed line) and 50% confidence band (dotted-dashed line) assuming a general variance-covariance matrix Γ.

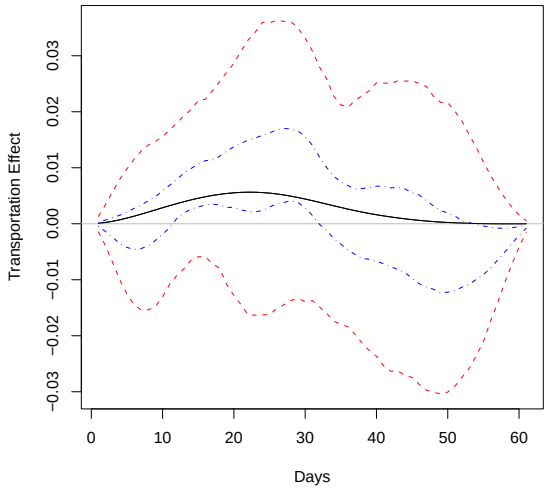

Figure 4.3: Plot of estimate for ϕ along with 95% parametric bootstrap confidence band (dashed line) and 60% confidence band (dotted-dashed) assuming a general variance-covariance matrix Γ.

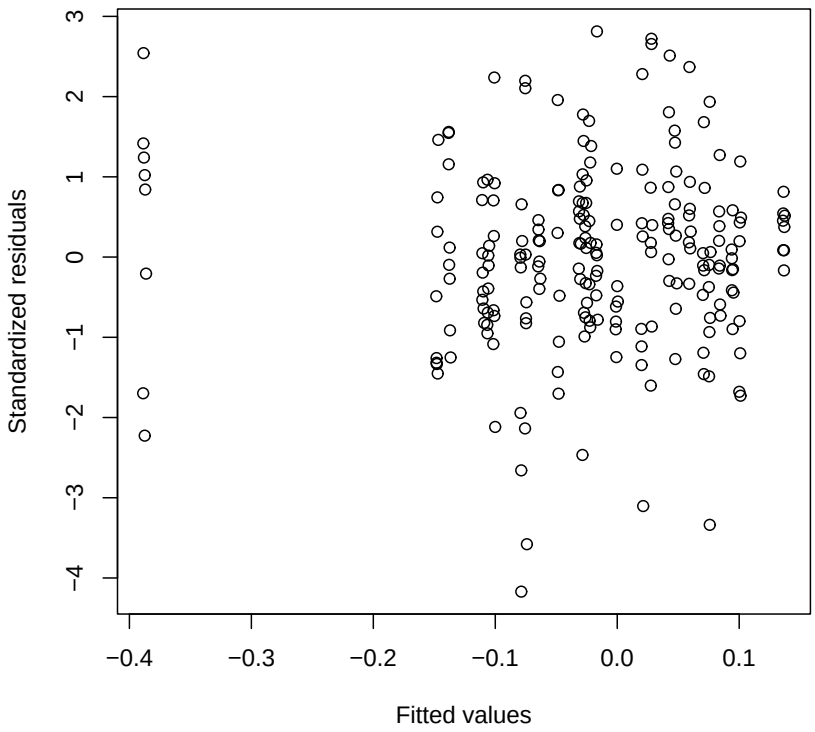

Figure 4.4: Residual plot for the model fitted with the multiplicative error model.

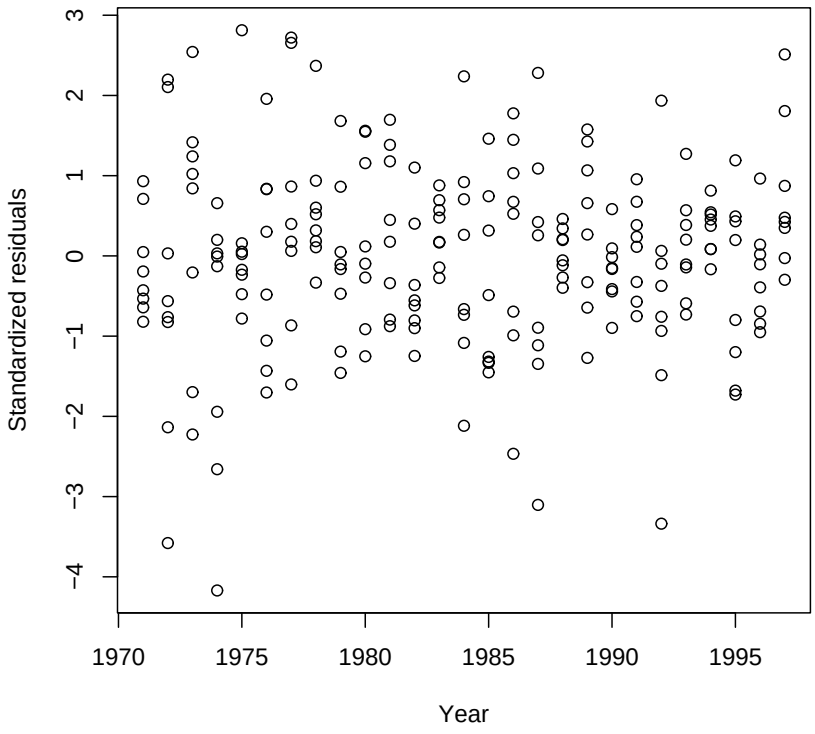

Figure 4.5: Residual plot of the residuals of the model fitted with multiplicative error by year (from 1971 to 1997).

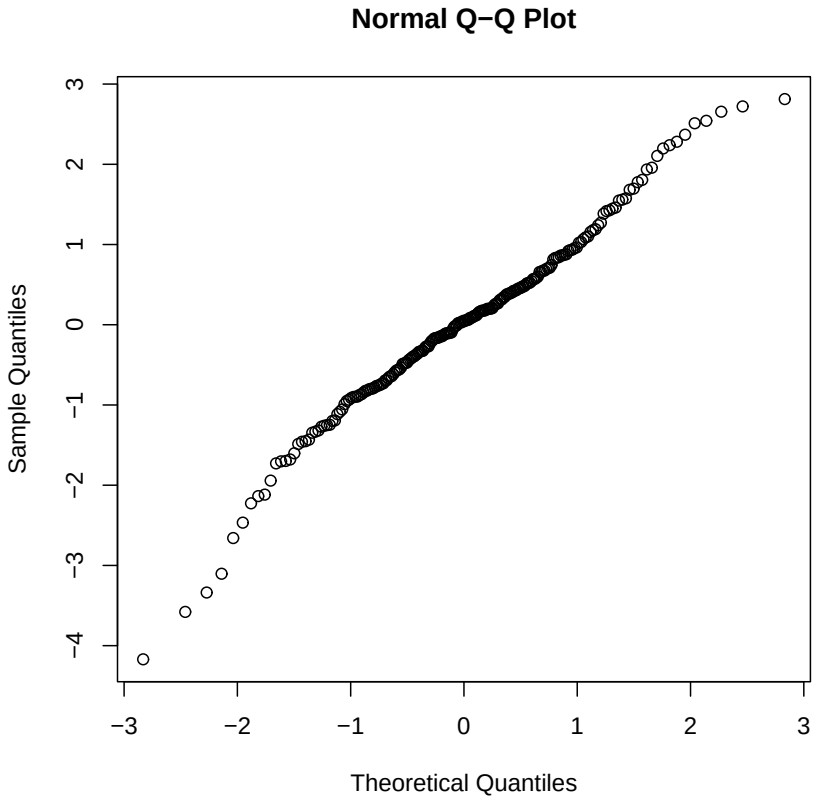

Figure 4.6: Normal Q-Q plot of the errors for the model fitted with multiplicative error.

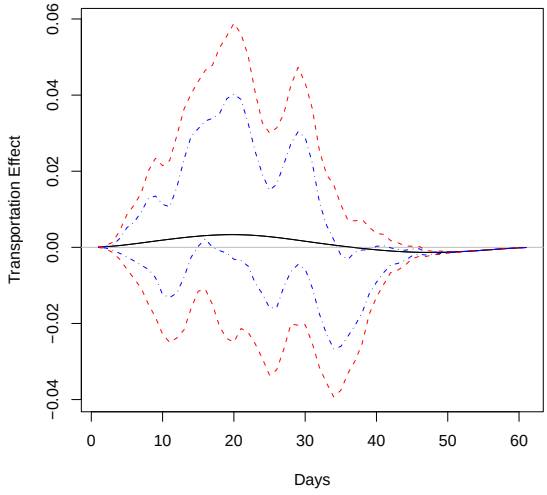

Figure 4.7: Plot of estimate for ϕ along with 95% nonparametric bootstrap confidence band (dashed line) and 70% confidence band (dotted-dashed line) assuming the multiplicative error model.

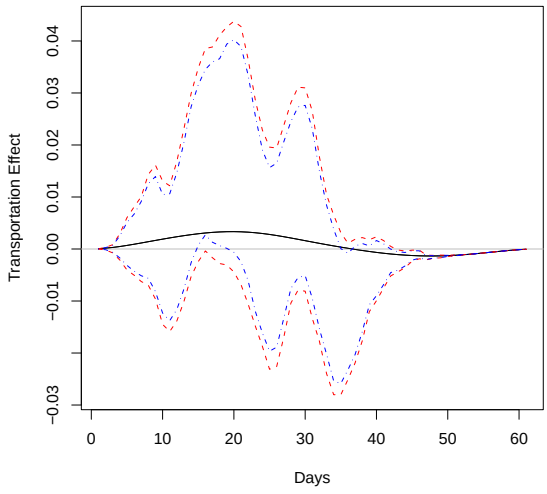

Figure 4.8: Plot of estimate for ϕ along with 95% parametric bootstrap confidence band (dashed line) and 90% confidence band (dotted-dashed line) assuming the multiplicative error model.

CHAPTER 5
CONCLUSION

The transfer function model is useful for assessing the cumulative, say daily, effects of a covariate on a response, say measured annually. We have developed a method of estimating a fixed-effect smooth transfer function model that incorporates contemporaneous correlation. We have shown that the estimated regression parameter is a consistent estimator for the true value, and that estimator has an asymptotic normal distribution. We have shown that this method can be extended to a mixed-effect model that can incorporate many different assumed between-series variance-covariance matrices. We illustrated both models with a cod spawning data that included spawning biomass and the daily net inflow into the Skagerrak. The proposed methods can be broadly applicable to problems in other areas. One case could be using the depth of a lake as a response, where covariates such as lake depth, rainfall, and other meteorological measurements are taken at different stations around the lake. Another is to determine yield from different corn fields, using meteorological or geographic measurements.

Our models are flexible in the sense that we can incorporate contemporaneous correlation and between-series correlation in the mixed-effect case. Although we only estimated a general contemporaneous variance-covariance matrix, certainly other specifications can be use, such as compound symmetric. Both models are relatively quick in the computation of estimates, and computation of confidence intervals in the fixed-effect case is also simple regardless of the type of confidence interval. However,

mixed-effect confidence intervals are time and computer intensive since it does require an iterative procedure in both the nonparametric and parametric approaches. Particular types of between-series variance-covariance matrices have both strengths and weaknesses. While a general variance-covariance matrix can impose smoothness by being computationally singular due to the number of parameters estimated, it could also lack the power to detect significance for that very same reason.

There are a few interesting future research problems. First, it is of interest to work out the case of non-normal errors in greater details. Further development of the computation of standard errors of the mean transfer function could be explored. We can study methods for speeding up the EM algorithm for the sake of the bootstrap confidence intervals and standard errors. This would also be prudent in possible calibration of the bootstrap confidence intervals. We could develop a test that can be used to determine if the true model is fixed-effect versus mixed-effect. We would like to generalize the R programs found in the appendix to more general situations and develop a package.

APPENDIX A
EXTRA RESULTS

A.1 Penalized log-likelihood for mixed effect model

An analytical approach to computing the standard error of the parameter estimates consists of adapting the asymptotics of M-estimation to our case using Theorem 3.1 from Van der Vaart [19]. This requires the derivation of the first and second derivative of the likelihood function. Let $U = \sum_t (Y_t - X_t\psi)(Y_t - X_t\psi)^T$ and $V = \sum_s (\psi_s - \phi)(\psi_s - \phi)^T$. Then the penalized complete-data log-likelihood can be written as

$$
\begin{aligned}
\ell(\phi,\Omega,\Gamma;\lambda) &= -\frac{T}{2}\log|\Omega| - \frac{S}{2}\log|\Gamma| - \frac{1}{2}\sum_t (Y_t - X_t\psi)^T\Omega^{-1}(Y_t - X_t\psi) - \\
&\quad \frac{1}{2}\sum_s (\psi_s - \phi)^T\Gamma^{-1}(\psi_s - \phi) - \frac{\lambda}{2}\phi^T A^T A\phi \\
\Rightarrow d\ell &= \frac{T}{2}\mathrm{tr}(\Omega^{-1}d\Omega) - \frac{S}{2}\mathrm{tr}(\Gamma^{-1}d\Gamma) + \\
&\quad \frac{1}{2}\sum_t (Y_t - X_t\psi)^T\Omega^{-1}d\Omega\Omega^{-1}(Y_t - X_t\psi) + \frac{1}{2}\sum_s d\phi^T\Gamma^{-1}(\psi_s - \phi) + \\
&\quad \frac{1}{2}\sum_s (\psi_s - \phi)^T\Gamma^{-1}d\Gamma\Gamma^{-1}(\psi_s - \phi) + \\
&\quad \frac{1}{2}\sum_s (\psi_s - \phi)^T\Gamma^{-1}d\phi - \frac{\lambda}{2}d\phi^T A^T A\phi - \frac{\lambda}{2}\phi^T A^T Ad\phi \\
&= -\frac{1}{2}\mathrm{tr}(T\Omega^{-1}d\Omega) - \frac{1}{2}\mathrm{tr}(S\Gamma^{-1}d\Gamma) + \frac{1}{2}\mathrm{tr}(d\Omega\Omega^{-1}U\Omega^{-1}) + \\
&\quad d\phi^T\Gamma^{-1}\sum_s (\psi_s - \phi) + \frac{1}{2}\mathrm{tr}(d\Gamma\Gamma^{-1}V\Gamma^{-1}) - \lambda d\phi^T A^T A\phi \\
&= -\frac{1}{2}\mathrm{vec}(d\Omega)^T\mathrm{vec}(T\Omega^{-1} - \Omega^{-1}U\Omega^{-1}) + d\phi^T\Gamma^{-1}\sum_s (\psi_s - \phi) - \\
&\quad \frac{1}{2}\mathrm{vec}(d\Gamma)^T\mathrm{vec}(S\Gamma^{-1} - \Gamma^{-1}V\Gamma^{-1}) - \lambda d\phi^T A^T A\phi \\
&= -\frac{1}{2}\mathrm{v}(d\Omega)^T D_S^T\mathrm{vec}(T\Omega^{-1} - \Omega^{-1}U\Omega^{-1}) - \\
&\quad \frac{1}{2}\mathrm{v}(d\Gamma)^T D_D^T\mathrm{vec}(S\Gamma^{-1} - \Gamma^{-1}V\Gamma^{-1}) + \\
&\quad d\phi^T(\Gamma^{-1}\sum_s (\psi_s - \phi) - \lambda A^T A\phi).
\end{aligned}
$$

The second differential will be computed for each part of the first differential.

$$\begin{aligned}
d-\frac{1}{2}\mathrm{v}(d\Omega)^T D_S^T \mathrm{vec}(T\Omega^{-1}-\Omega^{-1}U\Omega^{-1}) &= -\frac{1}{2}\mathrm{v}(d\Omega)^T D_S^T \mathrm{vec}(Td\Omega^{-1}) \\
&\quad +\frac{1}{2}\mathrm{v}(d\Omega)^T D_S^T \mathrm{vec}(d\Omega^{-1}U\Omega^{-1}) \\
&\quad +\frac{1}{2}\mathrm{v}(d\Omega)^T D_S^T \mathrm{vec}(\Omega^{-1}Ud\Omega^{-1}) \\
&= \frac{T}{2}\mathrm{v}(d\Omega)^T D_S^T \mathrm{vec}(\Omega^{-1}d\Omega\Omega^{-1}) \\
&\quad -\frac{1}{2}\mathrm{v}(d\Omega)^T D_S^T \mathrm{vec}(\Omega^{-1}d\Omega\Omega^{-1}U\Omega^{-1}) \\
&\quad -\frac{1}{2}\mathrm{v}(d\Omega)^T D_S^T \mathrm{vec}(\Omega^{-1}U\Omega^{-1}d\Omega\Omega^{-1}) \\
&= \frac{T}{2}\mathrm{v}(d\Omega)^T D_S^T (\Omega^{-1}\otimes\Omega^{-1}) D_S \mathrm{v}(d\Omega) \\
&\quad -\frac{1}{2}\mathrm{v}(d\Omega)^T (\Omega^{-1}U\Omega^{-1}\otimes\Omega^{-1}) D_S \mathrm{v}(d\Omega) \\
&\quad -\frac{1}{2}\mathrm{v}(d\Omega)^T (\Omega^{-1}\otimes\Omega^{-1}U\Omega^{-1}) D_S \mathrm{v}(d\Omega) \\
&= \frac{1}{2}(d\Omega)^T D_S^T \left\{ T\Omega^{-1}\otimes\Omega^{-1} \right. \\
&\quad -(\Omega^{-1}U\Omega^{-1})\otimes\Omega^{-1} \\
&\quad \left. -\Omega^{-1}\otimes(\Omega^{-1}U\Omega^{-1}) \right\} D_S \mathrm{v}(d\Omega)
\end{aligned}$$

Now

$$-\frac{1}{2}\mathrm{v}(d\Omega)^T D_S^T \mathrm{vec}(\Omega^{-1}d\Omega\Omega^{-1}U\Omega^{-1})$$

$$\begin{aligned}
&= -\frac{1}{2}\mathrm{tr}(d\Omega\Omega^{-1}d\Omega\Omega^{-1}U\Omega^{-1}) \\
&= -\frac{1}{2}\mathrm{tr}(d\Omega\Omega^{-1}U\Omega^{-1}d\Omega\Omega^{-1}) \\
&= -\frac{1}{2}\mathrm{v}(d\Omega)^T D_S^T \left\{ \Omega^{-1}\otimes(\Omega^{-1}U\Omega^{-1}) \right\} D_S \mathrm{v}(d\Omega).
\end{aligned}$$

So

$$d - \frac{1}{2}\mathrm{v}(d\Omega)^T D_S^T \mathrm{vec}(T\Omega^{-1} - \Omega^{-1}U\Omega^{-1}) = \frac{1}{2}\mathrm{v}(d\Omega)^T D_S^T \left\{T\Omega^{-1} \otimes \Omega^{-1} - 2\Omega^{-1} \otimes (\Omega^{-1}U\Omega^{-1})\right\} D_S \mathrm{v}(d\Omega).$$

Also we can state that

$$d - \frac{1}{2}\mathrm{v}(d\Gamma)^T D_D^T \mathrm{vec}(S\Gamma^{-1} - \Gamma^{-1}V\Gamma^{-1})$$

$$\begin{aligned}
&= \frac{1}{2}\mathrm{v}(d\Gamma)^T D_D^T \left\{S\Gamma^{-1} - (\Gamma^{-1}V\Gamma^{-1}) \otimes \Gamma^{-1} - \Gamma^{-1} \otimes (\Gamma^{-1}V\Gamma^{-1})\right\} D_D \mathrm{v}(d\Gamma) + \\
&\quad \frac{1}{2}\mathrm{v}(d\Gamma)^T D_D^T \mathrm{vec}(\Gamma^{-1}dV\Gamma^{-1})\mathrm{vec}(dV) \\
&= \frac{1}{2}\mathrm{v}(d\Gamma)^T D_D^T \left\{S\Gamma^{-1} - (\Gamma^{-1}V\Gamma^{-1}) \otimes \Gamma^{-1} - \Gamma^{-1} \otimes (\Gamma^{-1}V\Gamma^{-1})\right\} D_D \mathrm{v}(d\Gamma) - \\
&\quad \frac{1}{2}\mathrm{v}(d\Gamma) D_D^T \mathrm{vec}\left\{\Gamma^{-1}d\phi \sum_s (\psi_s - \phi)^T \Gamma^{-1}\right\} - \\
&\quad \frac{1}{2}\mathrm{v}(d\Gamma) D_D^T \mathrm{vec}\left\{\Gamma^{-1} \sum_s (\psi_s - \phi) d\phi^T \Gamma^{-1}\right\} \\
&= \frac{1}{2}\mathrm{v}(d\Gamma)^T D_D^T \left\{S\Gamma^{-1} - (\Gamma^{-1}V\Gamma^{-1}) \otimes \Gamma^{-1} - \Gamma^{-1} \otimes (\Gamma^{-1}V\Gamma^{-1})\right\} D_D \mathrm{v}(d\Gamma) - \\
&\quad \frac{1}{2}\mathrm{v}(d\Gamma) D_D^T \left\{\Gamma^{-1} \sum_s (\psi_s - \phi) \otimes \Gamma^{-1} + \Gamma^{-1} \otimes \Gamma^{-1} \sum_s (\psi_s - \phi)\right\} d\phi
\end{aligned}$$

So in the same way as above

$$d - \frac{1}{2}\mathrm{v}(d\Gamma)^T D_D^T \mathrm{vec}(S\Gamma^{-1} - \Gamma^{-1}V\Gamma^{-1})$$

$$= -\mathrm{v}(d\Gamma)^T D_D^T \left[\Gamma^{-1} \otimes \left\{\Gamma^{-1} \sum_s (\psi_s - \phi)\right\}\right] d\phi,$$

$$d\left\{d\phi^T \sum_s \Gamma^{-1}(\psi_s - \phi)\right\}$$

$$\begin{aligned}
&= d\phi^T d\Gamma^{-1} \sum_s (\psi_s - \phi) - d\phi^T \Gamma^{-1} \sum_s d\phi \\
&= -d\phi^T \Gamma^{-1} d\Gamma \Gamma^{-1} \sum_s (\psi_s - \phi) - S d\phi \Gamma^{-1} d\phi \\
&= -\mathrm{v}(d\Gamma)^T D_D^T \left[\Gamma^{-1} \otimes \left\{\Gamma^{-1} \sum_s (\psi_s - \phi)\right\}\right] d\phi - S d\phi^T \Gamma^{-1} d\phi, \text{ and}
\end{aligned}$$

$$d\left\{-\lambda d\phi^T A^T A \phi\right\} = -\lambda d\phi^T A^T A d\phi.$$

Consequently,

$$d^2 \ell(\phi, \Gamma, \Omega; \lambda)$$

$$\begin{aligned}
= \; & \frac{1}{2}\mathrm{v}(d\Omega)^T D_S^T \left\{T\Omega^{-1} \otimes \Omega^{-1} - 2\Omega^{-1} \otimes (\Omega^{-1} U \Omega^{-1})\right\} D_S \mathrm{v}(d\Omega) + \\
& \frac{1}{2}\mathrm{v}(d\Gamma)^T D_D^T \left\{S\Gamma^{-1} \otimes \Gamma^{-1} - 2\Gamma^{-1} \otimes (\Gamma^{-1} V \Gamma^{-1})\right\} D_D \mathrm{v}(d\Gamma) - \\
& d\phi^T \left(S\Gamma^{-1} + \lambda A^T A\right) d\phi - \\
& 2\mathrm{v}(d\Gamma)^T D_D^T \left[\Gamma^{-1} \otimes \left\{\Gamma^{-1} \sum_s (\psi_s - \phi)\right\}\right] d\phi. \qquad \text{(A.1)}
\end{aligned}$$

Let $\tilde{\ell}_Y(\theta)$ be the log-likelihood of the observed data and $\tilde{\ell}(\theta)$ be the complete data log-likelihood. Then we can state that

$$-d^2 \tilde{\ell}_Y(\theta) = E_\theta \left\{-d^2 \tilde{\ell}(\theta) | Y\right\} - E\left\{d\tilde{\ell}(\theta) d^T \ell_X(\theta) | Y\right\} + d\tilde{\ell}_Y(\theta) d^T \tilde{\ell}_Y(\theta) [14]. \quad \text{(A.2)}$$

The first term in the right side of (A.2) is

$$E_\theta \left\{ -d^2 \tilde{\ell}(\theta) | Y \right\}$$

$$= \begin{bmatrix} d\phi^T & \mathrm{v}(d\Gamma)^T & \mathrm{v}(d\Omega)^T \end{bmatrix} \begin{bmatrix} E_1 & E_2 & 0 \\ E_2^T & E_3 & 0 \\ 0 & 0 & E_4 \end{bmatrix} \begin{bmatrix} d\phi \\ \mathrm{v}(d\Gamma) \\ \mathrm{v}(d\Omega) \end{bmatrix}, \text{ where}$$

$$E_1 = S\Gamma^{-1},$$

$$E_2 = \left[\Gamma^{-1} \otimes \left\{ \sum_s (\widehat{\psi}_{s,T} - \phi)^T \Gamma^{-1} \right\} \right] D_D,$$

$$E_3 = D_D^T \left\{ \Gamma^{-1} \otimes \left(\Gamma^{-1} \left[2 \sum_s \left\{ (\widehat{\psi}_{s,T} - \phi)(\widehat{\psi}_{s,T} - \phi)^T + \widehat{\Sigma}_{s,s,T} \right\} \Gamma^{-1} - SI \right] \right) \right\} D_D,$$

$$\text{and } E_4 = D_S^T \left\{ \Omega^{-1} \otimes \left(\Omega^{-1} \left[2 \sum_t \left\{ (Y_t - X_t \widehat{\psi}_T)(Y_t - X_t \widehat{\psi}_T)^T + \widehat{\Sigma}_T \right\} \Omega^{-1} - TI \right] \right) \right\} D_S.$$

The second term in the right side of (A.2) is not easily computed. That is because it requires that computation of the expectation of a non-central fourth moment of the normal distribution. We can state that if

$$\begin{bmatrix} W \\ X \\ Y \\ Z \end{bmatrix} \sim N \left(\begin{bmatrix} \mu_W \\ \mu_X \\ \mu_Y \\ \mu_Z \end{bmatrix}, \begin{bmatrix} \sigma_W^2 & \sigma_{W,X} & \sigma_{W,Y} & \sigma_{W,Z} \\ \sigma_{W,X} & \sigma_X^2 & \sigma_{X,Y} & \sigma_{X,Z} \\ \sigma_{W,Y} & \sigma_{X,Y} & \sigma_Y^2 & \sigma_{Y,Z} \\ \sigma_{W,Z} & \sigma_{X,Z} & \sigma_{Y,Z} & \sigma_Z^2 \end{bmatrix} \right), \text{ then}$$

$$
\begin{aligned}
E(XYZ) &= \mu_X\mu_Y\mu_Z + \mu_X\sigma_{Y,Z} + \mu_Y\sigma_{X,Z} + \mu_Z\sigma_{X,Y}, \text{ and} \\
E(WXYZ) &= \mu_W\mu_X\mu_Y\mu_Z + \mu_W\mu_X\sigma_{Y,Z} + \mu_W\mu_Y\sigma_{X,Z} + \mu_W\mu_Z\sigma_{X,Y} + \\
&\quad \mu_X\mu_Y\sigma_{W,Z} + \mu_X\mu_Z\sigma_{W,Y} + \mu_Y\mu_Z\sigma_{W,X} + \sigma_{W,X}\sigma_{Y,Z} + \sigma_{W,Y}\sigma_{X,Z} + \\
&\quad \sigma_{W,Z}\sigma_{X,Y}[11].
\end{aligned}
$$

The third term on the right side of (A.2) is

$$
\begin{aligned}
d\tilde{\ell}_Y(\theta)d^T\tilde{\ell}_Y(\theta) &= \begin{bmatrix} L_1 \\ L_2 \\ L_3 \end{bmatrix} \begin{bmatrix} L_1^T & L_2^T & L_3^T \end{bmatrix} \\
&= \begin{bmatrix} L_1L_1^T & L_1L_2^T & L_1L_3^T \\ L_2L_1^T & L_2L_2^T & L_2L_3^T \\ L_3L_1^T & L_3L_2^T & L_3L_3^T \end{bmatrix}, \text{ where} \\
L_1 &= \Gamma^{-1}\sum_s(\widehat{\psi}_{s,T} - \phi), \\
L_2 &= D_D^T\text{vec}\left[S\Gamma^{-1} - \Gamma^{-1}\left\{\sum_s(\widehat{\psi}_{s,T} - \phi)(\widehat{\psi}_{s,T} - \phi)^T + \widehat{\Sigma}_{s,s,T}\right\}\Gamma^{-1}\right], \text{ and} \\
L_3 &= D_S^T\text{vec}\left[T\Omega^{-1} - \Omega^{-1}\left\{\sum_s(Y_t - X_t\widehat{\psi}_T)(Y_t - X_t\widehat{\psi}_T)^T + \widehat{\Sigma}_T\right\}\Omega^{-1}\right]
\end{aligned}
$$

Because of the difficulty in computing (A.2) and its inverse in this particular case makes computing the standard errors difficult. In Section 4.5 we will develop another method of computing confidence intervals for this model.

A.2 Multiplicative error model

In looking at the second differential of Equation (4.5), Equation (A.1) is now

$$d^2\ell\{\phi, \Gamma(\phi, \sigma_1^2, \sigma_2^2, \sigma_\eta^2)\Omega; \lambda\}$$

$$\begin{aligned}
= \quad & \frac{1}{2}\mathrm{v}(d\Omega)^T D_S^T \left\{T\Omega^{-1} \otimes \Omega^{-1} - 2\Omega^{-1} \otimes (\Omega^{-1}U\Omega^{-1})\right\} D_S \mathrm{v}(d\Omega) + \frac{Sd^2\sigma_1^2}{\sigma_1^4} + \\
& \frac{Sd^2\sigma_2^2}{\sigma_2^2} + 2S\sum_d \frac{d^2\phi_d}{\phi_d^4} + \frac{2Sd\sigma_\eta^2}{\sigma_\eta^2}\sum_d \frac{d\phi_d}{\phi_d} + \frac{DSd^2\sigma_\eta^2}{2\sigma_\eta^4} - \frac{d^2\sigma_1^2\sum_s(\psi_{s,1}-\kappa)^2}{\sigma_1^6} - \\
& \frac{d^2\sigma_2^2\sum_s(\psi_{s,2}-\zeta)^2}{\sigma_2^6} - \frac{4}{\sigma_\eta^2}\sum_d\sum_s \frac{d^2\phi_d(\psi_{s,d}-\phi_d)^2}{\phi_d^4} - \\
& \frac{4d\sigma_\eta^2}{\sigma_\eta^4}\sum_d\sum_s \frac{d\phi_d(\psi_{s,d}-\phi_d)^2}{\phi_d^3\sigma_\eta^2} - \frac{d^2\sigma_\eta^2}{\sigma_\eta^6}\sum_d\sum_s(\psi_{s,d}-\phi_d)^2 - \\
& d\phi^T(S\Gamma^{-1} + \lambda A^T A)d\phi - \frac{2\sum_s(\psi_{s,1}-\kappa)d\kappa d\sigma_1^2}{\sigma_1^2} - \frac{2\sum_s(\psi_{s,2}-\zeta)d\zeta d\sigma_2^2}{\sigma_2^2} - \\
& \frac{4}{\sigma_\eta^2}\sum_d\sum_s \frac{d^2\phi_d(\psi_{s,d}-\phi_d)}{\phi_d^3} - \frac{2}{\sigma_\eta^2}\sum_d\sum_s \frac{d\phi_d d\sigma_\eta^2(\psi_{s,d}-\phi_d)}{\phi_d^2} \\
= \quad & \frac{1}{2}\mathrm{v}(d\Omega)^T D_S^T \left\{T\Omega^{-1} \otimes \Omega^{-1} - 2\Omega^{-1} \otimes (\Omega^{-1}U\Omega^{-1})\right\} D_S \mathrm{v}(d\Omega) - \\
& d\phi^T(S\Gamma^{-1} + \lambda A^T A)d\phi + d^2\sigma_1^2\left\{\frac{S}{\sigma_1^4} - \frac{(\psi_{s,1}-\kappa)^2}{\sigma_1^6}\right\} + d^2\sigma_2^2\left\{\frac{S}{\sigma_2^4} - \right. \\
& \left.\frac{(\psi_{s,1}-\zeta)^2}{\sigma_2^6}\right\} + d^2\sigma_\eta^2\left\{\frac{DS}{2\sigma_\eta^4} - \frac{\sum_d\sum_s(\psi_{s,d}-\phi_d)^2}{\sigma_\eta^6}\right\} + \\
& \sum_d d^2\phi_d\left\{\frac{2S}{\phi_d^4} - \frac{4\sum_s(\psi_{s,d}-\phi_d)^2}{\sigma_\eta^2\phi_d^4} - \frac{2\sum_s(\psi_{s,d}-\phi_d)}{\phi_d^3}\right\} - \\
& d\kappa d\sigma_1^2\left\{\frac{2}{\sigma_1^2}\sum_s(\psi_{s,1}-\kappa)\right\} - d\zeta d\sigma_2^2\left\{\frac{2}{\sigma_2^2}\sum_s(\psi_{s,2}-\zeta)\right\} + \\
& \sum_d d\phi_d d\sigma_\eta^2\left\{\frac{2S}{\sigma_\eta^2\phi_d} - \sum_s \frac{4(\psi_{s,d}-\phi_d)^2}{\sigma_\eta^2\phi_d^3} - \sum_s \frac{2(\psi_{s,d}-\phi_d)}{\sigma_\eta^2\phi_d^2}\right\}
\end{aligned}$$

$$
\begin{aligned}
&= \begin{bmatrix} d\phi^T & d\sigma_1^2 & d\sigma_2^2 & d\sigma_\eta^2 & \mathrm{v}(d\Omega)^T \end{bmatrix} \times \\
&\quad \begin{bmatrix} F_1 & F_2 & F_3 & F_4 & 0 \\ F_2^T & F_5 & 0 & 0 & 0 \\ F_3^T & 0 & F_6 & 0 & 0 \\ F_4^T & 0 & 0 & F_7 & 0 \\ 0 & 0 & 0 & 0 & F_8 \end{bmatrix} \begin{bmatrix} d\phi \\ d\sigma_1^2 \\ d\sigma_2^2 \\ d\sigma_\eta^2 \\ \mathrm{v}(d\Omega) \end{bmatrix}, \text{ where} \\
F_1 &= -S\Gamma^{-1} - \lambda A^T A + \mathrm{diag}\left(0,\ 0,\ \frac{S}{\phi_1^4} - \frac{4\sum_s(\psi_{s,3}-\phi_1)^2}{\sigma_\eta^2\phi_1} - \frac{2\sum_s(\psi_{s,3}-\phi_1)}{\phi_1},\right. \\
&\quad \left. \ldots,\ \frac{S}{\phi_D^4} - \frac{4\sum_s(\psi_{s,D}-\phi_D)^2}{\sigma_\eta^2\phi_D} - \frac{2\sum_s(\psi_{s,D}-\phi_D)}{\phi_D}\right), \\
F_2 &= \begin{bmatrix} -\frac{\sum_s(\psi_{s,1}-\kappa)}{\sigma_1^2},\ 0,\ 0,\ \ldots,\ 0 \end{bmatrix}^T, \\
F_3 &= \begin{bmatrix} 0,\ -\frac{\sum_s(\psi_{s,2}-\zeta)}{\sigma_2^2},\ 0,\ \ldots,\ 0 \end{bmatrix}^T, \\
F_4 &= \begin{bmatrix} 0,\ 0,\ \frac{S}{\sigma_\eta^2\phi_1} - \frac{2\sum_s(\psi_{s,3}-\phi_1)^2}{\sigma_\eta^2\phi_1^3} - \frac{\sum_s(\psi_{s,3}-\phi_1)}{\sigma_\eta^2\phi_1},\ \ldots \\ \frac{S}{\sigma_\eta^2\phi_D} - \frac{2\sum_s(\psi_{s,D}-\phi_D)^2}{\sigma_\eta^2\phi_D^3} - \frac{\sum_s(\psi_{s,D}-\phi_D)}{\sigma_\eta^2\phi_D} \end{bmatrix}^T, \\
F_5 &= \frac{S}{\sigma_1^4} - \frac{(\psi_{s,1}-\kappa)^2}{\sigma_1^6}, \\
F_6 &= \frac{S}{\sigma_2^4} - \frac{(\psi_{s,2}-\zeta)^2}{\sigma_2^6}, \\
F_7 &= \frac{DS}{2\sigma_\eta^4} - \frac{\sum_d\sum_s(\psi_{s,d}-\phi_d)^2}{\sigma_\eta^2}, \text{ and} \\
F_8 &= D_S^T\left\{T\Omega^{-1}\otimes\Omega^{-1} - 2\Omega^{-1}\otimes(\Omega^{-1}U\Omega^{-1})\right\}D_S
\end{aligned}
$$

APPENDIX B
R CODE

B.1 Fixed effect model

```
skagerrak <- function(y, X, nint, nw, maxiter = 100, end.criteria
= 0.001, sp, off)
{
# Used in estimation of the fixed effect model.
# Arguments:
#    y - The response vector sorted by year
#    X - The covariate matrix sorted by year
#    nint - The number of intercepts
#    nw - The number of days per year where measurements are taken
#    maxiter - The maximum number of iterations
#    end.criteria - The value for which if the relative change in
#       the regression parameter estimate between the previous
#       iteration and the current estimate is below this value,
#       then convergence is assumed.
#    sp - the starting value of the smoothing parameter
#    off - The number of fjords + 2.  If missing it is set to the
#       number of intercepts + 2.
# Details:
```

```
#    Uses magic from the mgcv library to estimate the common
#    transfer function model with contemporaneous correlation
#    between the fjords.
# Values:
#    model - Output from the implementation of magic, including
#       b - The regression parameter estimates
#       sp - The estimate for the smoothing parameter
#    residuals - The residuals for each observation
#    Sigma - The estimated covariance matrix
#    hat.matrix - The hat matrix, or the matrix such that
#       if post-multiplied by the response vector, the
#       result is the fitted values
if(missing(off)){off<-nint+2}
r <- rep(0, (nw - 2)*nw)
R <- matrix(r, ncol = nw)
for(i in 1:(nw-2))
{
        R[i, i] <- 1
        R[i, i+1] <- -2
        R[i, i+2] <- 1
}
rtop <- rep(0, nw*2)
```

```
Rtop <- matrix(rtop, ncol = nw)

Rtop[1, 1] <- 1

Rtop[2, 1] <- -2

Rtop[2, 2] <- 1

R <- rbind(Rtop, R)

rbottom <- rep(0, nw*2)

Rbottom <- matrix(rbottom, ncol = nw)

Rbottom[1, nw - 1] <- 1

Rbottom[1, nw] <- -2

Rbottom[2, nw] <- 1

R <- rbind(R, Rbottom)

i <- 1

gcv <- 0

library(mgcv)



while(T)

{

if(i == 1)

{

blah <- lm(y~X-1)

omega<-diag(rep(1,nint))
```

```
} else
{
residuals<-y-X%*%blah$b
omega <- cov(matrix(residuals, ncol = nint,byrow=T))


}
Omega <- omega/omega[1,1]
msqrt<-function(x){solve(t(chol(x)))}
Omega.5inv <- msqrt(Omega)
T.length <- length(y)/nint
X.new <- NULL
y.new <- NULL
for(j in 1:T.length)
{
        X.new <- rbind(X.new, Omega.5inv %*% X[((j - 1) * nint + 1):
(j * nint),])
        y.new <- rbind(y.new, Omega.5inv %*% y[((j - 1) * nint + 1):
(j * nint)])
}
blah <- magic(y = y.new, X = X.new, sp = sp, S = list(t(R)%*%R),
off=off)
criteria <- abs(gcv-blah$score) <= end.criteria*abs(gcv)
```

```
if(criteria)

{

hat.matrix <- X.new%*%solve(t(X.new)%*%X.new)%*%t(X.new)

return(list(model=blah,residuals=residuals, Sigma=omega,

hat.matrix=hat.matrix))

break

} else

{

i <- i+1

gcv <- blah$score

}

if(i==maxiter)

{

print('Too many iterations')

return(list(model=blah,residuals=residuals,Sigma=omega))

break

}

}

}
```

B.2 Fixed effect bootstrap procedures for confidence intervals

B.2.1 Estimation for bootstrap samples

```
skagerrak1 <- function(y, X, nint, nw, Omega, sp, off)
{
# Used in estimation of the fixed effect model used on the bootstrap
# samples.
# Arguments:
#    y - The response vector sorted by year
#    X - The covariate matrix sorted by year
#    nint - The number of intercepts
#    nw - The number of days per year where measurements are taken
#    Omega - The assumed matrix of the covariance between
#       the fjords
#    sp - the starting value of the smoothing parameter
#    off - The number of fjords + 2.  If missing it is set to the
#       number of intercepts + 2.
# Details:
#    Uses magic from the mgcv library to estimate the common
#    transfer function model with contemporaneous correlation
#    between the fjords.  Does not require iteration.
# Values:
#    model - Output from the implementation of magic, including
```

```
#       b - The regression parameter estimates

#       sp - The estimate for the smoothing parameter

#    Sigma - The estimated covariance matrix

if(missing(off)){off<-nint+2}

r <- rep(0, (nw - 2)*nw)

R <- matrix(r, ncol = nw)

for(i in 1:(nw-2))

{

        R[i, i] <- 1

        R[i, i+1] <- -2

        R[i, i+2] <- 1

}

rtop <- rep(0, nw*2)

Rtop <- matrix(rtop, ncol = nw)

Rtop[1, 1] <- 1

Rtop[2, 1] <- -2

Rtop[2, 2] <- 1

R <- rbind(Rtop, R)

rbottom <- rep(0, nw*2)

Rbottom <- matrix(rbottom, ncol = nw)

Rbottom[1, nw - 1] <- 1

Rbottom[1, nw] <- -2
```

```
Rbottom[2, nw] <- 1

R <- rbind(R, Rbottom)

gcv <- 0

library(mgcv)

msqrt<-function(x){solve(t(chol(x)))}

Omega.5inv <- msqrt(Omega)

T.length <- length(y)/nint

X.new <- NULL

y.new <- NULL

for(j in 1:T.length)

{

        X.new <- rbind(X.new, Omega.5inv %*% X[((j - 1) * nint + 1):

(j * nint),])

        y.new <- rbind(y.new, Omega.5inv %*% y[((j - 1) * nint + 1):

(j * nint)])

}

rows <- nrow(t(R)%*%R)

zeros <- matrix(0, nrow = rows, ncol = nint+1)

Sm <- cbind(zeros,t(R)%*%R)

cols <- ncol(Sm)

zeros <- matrix(0, ncol = cols, nrow = nint+1)

Sm <- rbind(zeros, Sm)
```

```
blah <- magic(y = y.new, X = X.new, sp = rep(0,0), H = sp*Sm,
S = list(), off=rep(0,0))
return(list(model=blah, Sigma=omega))
}
```

B.2.2 Nonparametric approach

```
nonparbootstrap <- function(n, S, D, times, Y, X, phi, Omega, sp)
{
# Arguments:
#    n - The number of bootstrap samples needed
#    S - The number of fjords
#    D - The number of days per year where measurements are taken
#    times - The number of years
#    Y - The response vector sorted by year
#    X - The covariate matrix sorted by year
#    phi - The assumed regression parameter
#    Omega - The assumed matrix of the covariance between
#       the fjords
#    sp - The assumed value of the smoothing parameter
# Details:
#    Computes n nonparametric bootstrap samples
# Values:
```

```
#    phis - The estimated regression parameters for each

#       of the n bootstrap samples

phis <- NULL

sps <- NULL

Yf <- X%*%phi

for(j in 1:n)

{

samps <- sample(1:times, times, replace = TRUE)

Xnew <- matrix(0, nrow = nrow(X), ncol = ncol(X))

Ynew <- rep(0, length(Y))

for(i in 1:times)

{

Yf1 <- Yf[((i-1)*S+1):(i*S)]

X1 <- X[((i-1)*S+1):(i*S),]

Y2 <- Y[((samps[i]-1)*S+1):(samps[i]*S)]

Yf2 <- Yf[((samps[i]-1)*S+1):(samps[i]*S)]

E2 <- Y2 - Yf2

Y1 <- Yf1 + E2

Xnew[((i-1)*S+1):(i*S),] <- X1

Ynew[((i-1)*S+1):(i*S)] <- Y1

}

blah <- skagerrak1(y = Ynew, X = Xnew, Omega = Omega, nint = S,
```

```
nw = D, sp = sp)

phis <- cbind(phis, blah$model$b)

sps <- c(sps, blah$model$sp)

}

return(phis)

}
```

B.2.3 Parametric bootstrap

```
parbootstrap <- function(n, S, D, times, Y, X, phi, Omega, sp)

{

# Arguments:

#    n - The number of bootstrap samples needed

#    S - The number of fjords

#    D - The number of days per year where measurements are taken

#    times - The number of years

#    Y - The response vector sorted by year

#    X - The covariate matrix sorted by year

#    phi - The assumed regression parameter

#    Omega - The assumed matrix of the covariance between

#       the fjords

#    sp - The assumed value of the smoothing parameter
```

```
# Details:

#    Computes n parametric bootstrap samples

# Values:

#    phis - The estimated regression parameters for each

#       of the n bootstrap samples

library(MASS)

phis <- NULL

sps <- NULL

Yf <- X%*%phi

for(j in 1:n)

{

Xnew <- matrix(0, nrow = nrow(X), ncol = ncol(X))

Ynew <- rep(0, length(Y))

for(i in 1:times)

{

Yf1 <- Yf[((i-1)*S+1):(i*S)]

X1 <- X[((i-1)*S+1):(i*S),]

E2 <- mvrnorm(n = 1, mu = rep(0, length(Yf1)), Sigma = Omega)

Y1 <- Yf1 + E2

Xnew[((i-1)*S+1):(i*S),] <- X1

Ynew[((i-1)*S+1):(i*S)] <- Y1

}
```

```
blah <- skagerrak1(y = Ynew, X = Xnew, Omega = Omega, nint = S,
nw = D, sp = sp)
phis <- cbind(phis, blah$model$b)
sps <- c(sps, blah$model$sp)
}
return(phis)
}
```

B.3 Mixed effect model

B.3.1 Γ general

```
em <- function(Y, X, phi, Gamma, Omega, S, ints, times, nw, sp,
end.crit = 0.001, maxiter = 100)
{
# Used in estimation of the mixed effect model.
# Arguments:
#    Y - The response vector sorted by fjord
#    X - The covariate matrix sorted by fjord
#    phi - The initial estimate for the mean of the psi's,
#       where the psi's are the random effects for each
#       fjord
#    Gamma - The initial estimate for the covariance matrix
#       of the psi's
```

```
#    Omega - The initial estimate for the contemporaneous
#       covariance matrix
#    S - The number of fjords
#    ints - The number of intercepts; if missing is set to be the
#       same as S
#    nw - The number of days per year where measurements are taken
#    sp - The starting value of the smoothing parameter
#    end.crit - The value for which if the relative change in the
#       regression parameter estimate between the previous
#       iteration and the current estimate is below this value,
#       then convergence is assumed.
#    maxiter - The maximum number of iterations
# Details:
#    Uses the EM algorithm and magic from the mgcv library
#    to estimate the common transfer function model with
#    contemporaneous correlation between the fjords in the
#    mixed effect case.  Assumes that Gamma is a general
#    covariance matrix.
# Values:
#    phi - The estimate of the mean of the random effects psi_s
#    Gamma - The estimate for the covariance matrix of the psi_s
#    Omega - The estimate for the contemporaneous covariance
```

```
#       matrix
#    iterations - The number of iterations needed for convergence
#    sp - The estimate for the smoothing parameter
#    gcv - The value of the gcv function at the value of sp
#    mu - The mean vector estimated from the imputation
#    Sigma - The covariance matrix estimated from the imputation
if(missing(ints)){ints <- S}
sp.old <- sp
library(mgcv)
r <- rep(0, (nw - 2)*nw)
R <- matrix(r, ncol = nw)
for(i in 1:(nw-2))
{
        R[i, i] <- 1
        R[i, i+1] <- -2
        R[i, i+2] <- 1
}
rtop <- rep(0, nw*2)
Rtop <- matrix(rtop, ncol = nw)
Rtop[1, 1] <- 1
Rtop[2, 1] <- -2
Rtop[2, 2] <- 1
```

```
R <- rbind(Rtop, R)

rbottom <- rep(0, nw*2)

Rbottom <- matrix(rbottom, ncol = nw)

Rbottom[1, nw - 1] <- 1

Rbottom[1, nw] <- -2

Rbottom[2, nw] <- 1

R <- rbind(R, Rbottom)


oneS <- rep(1,S)

IS <- diag(oneS)

lphi <- length(phi)

psidist <- function(Y, X, phi, Gamma, Omega, S, times)

{

oneS <- rep(1,S)

IS <- diag(oneS)

mu <- oneS%x%phi

Sigma <- IS%x%Gamma

lphi <- length(phi)

for(i in 1:times)

{

Yt <- Y[((i-1)*S+1):(i*S)]

xt <- X[((i-1)*S+1):(i*S),]
```

```
Xt <- NULL
for(k in 1:S)
{
Xt <-rbind(Xt, c(rep(0, (k-1)*lphi), xt[k,], rep(0, (S-k)*lphi)))
}
op <- solve(Omega+Xt%*%Sigma%*%t(Xt))
XSig <- Xt%*%Sigma
mu <- mu + drop(t(XSig)%*%op%*%(Yt-Xt%*%mu))
Sigma <- Sigma - t(XSig)%*%op%*%XSig
}
return(list(mu = mu, Sigma = Sigma))
}
msqrt <- function(x){solve(t(chol(x)))}
j <- 1
while(T)
{
phi.old <- phi
Gamma.old <- Gamma
Omega.old <- Omega
hats <- psidist(Y=Y, X=X, phi = phi, Gamma = Gamma, Omega = Omega,
S = S, times = times)
mu <- hats$mu
```

```
Sigma <- hats$Sigma

Gamma.5inv <- msqrt(Gamma)

blah <- NULL

for(i in 1:S)

{

blah <- rbind(blah, Gamma.5inv%*%

hats$mu[((i-1)*dim(Gamma)[1]+1):(i*dim(Gamma)[1])])

}

phi.comp <- magic(y = blah, X = oneS%x%Gamma.5inv, sp = sp,

S = list(t(R)%*%R), off = ints+2)

phi <- phi.comp$b

sum.Sigma <- 0

sum.diff <- 0

for(i in 1:S)

{

sum.Sigma <- sum.Sigma + Sigma[((i-1)*lphi+1):(i*lphi),

((i-1)*lphi+1):(i*lphi)]

sum.diff <- sum.diff + (mu[((i-1)*lphi+1):(i*lphi)]-phi)%*%

t(mu[((i-1)*lphi+1):(i*lphi)]-phi)

}

Gamma <- (sum.Sigma+sum.diff)/S

d <- length(mu)
```

```
sum.XSigmaX <- 0

sum.diff <- 0

for(i in 1:times)

{

Yt <- Y[((i-1)*S+1):(i*S)]

xt <- X[((i-1)*S+1):(i*S),]

Xt <- NULL

for(k in 1:S)

{

Xt <-rbind(Xt, c(rep(0, (k-1)*lphi), xt[k,], rep(0, (S-k)*lphi)))

}

sum.XSigmaX <- sum.XSigmaX + Xt%*%Sigma%*%t(Xt)

sum.diff <- sum.diff + (Yt-Xt%*%mu)%*%t(Yt-Xt%*%mu)

}

Omega <- (sum.XSigmaX + sum.diff)/times

crit1 <- sum(abs(phi-phi.old))/sum(abs(phi.old))

crit2 <- sum(abs(diag(Gamma - Gamma.old)))/sum(diag(Gamma.old))

crit3 <- sum(abs(diag(Omega - Omega.old)))/sum(diag(Omega.old))

if(crit1 < end.crit & crit2 < end.crit & crit3 < end.crit)

{

return(list(phi = phi, Gamma = Gamma, Omega = Omega, iterations = j,

sp = phi.comp$sp, gcv = phi.comp$score, mu = mu, Sigma = Sigma))
```

```
break

} else

{

j <- j+1

sp <- max(phi.comp$sp, sp.old)

}

if(j >= maxiter)

{

print("Stopped: Maximum number of iterations reached")

return(list(phi = phi, Gamma = Gamma, Omega = Omega,

crit = c(crit1, crit2, crit3), sp = phi.comp$sp,

gcv = phi.comp$score, mu = mu, Sigma = Sigma))

break

}

}

}
```

B.3.2 Γ is diagonal

```
em <- function(Y, X, phi, Gamma, Omega, S, ints, times, nw, sp,

end.crit = 0.001, maxiter = 100)

{

# Used in estimation of the mixed effect model.
```

```
# Arguments:
#    Y - The response vector sorted by fjord
#    X - The covariate matrix sorted by fjord
#    phi - The initial estimate for the mean of the psi's,
#       where the psi's are the random effects for each
#       fjord
#    Gamma - The initial estimate for the covariance matrix
#       of the psi's
#    Omega - The initial estimate for the contemporaneous
#       covariance matrix
#    S - The number of fjords
#    ints - The number of intercepts; if missing is set to be the
#       same as S
#    nw - The number of days per year where measurements are taken
#    sp - The starting value of the smoothing parameter
#    end.crit - The value for which if the relative change in the
#       regression parameter estimate between the previous
#       iteration and the current estimate is below this value,
#       then convergence is assumed.
#    maxiter - The maximum number of iterations
# Details:
#    Uses the EM algorithm and magic from the mgcv library
```

```
#    to estimate the common transfer function model with
#    contemporaneous correlation between the fjords in the
#    mixed effect case.  Assumes that Gamma is a diagonal
#    matrix.
# Values:
#    phi - The estimate of the mean of the random effects psi_s
#    Gamma - The estimate for the covariance matrix of the psi_s
#    Omega - The estimate for the contemporaneous covariance
#       matrix
#    iterations - The number of iterations needed for convergence
#    sp - The estimate for the smoothing parameter
#    gcv - The value of the gcv function at the value of sp
if(missing(ints)){ints <- S}
sp.old <- sp
library(mgcv)
r <- rep(0, (nw - 2)*nw)
R <- matrix(r, ncol = nw)
for(i in 1:(nw-2))
{
        R[i, i] <- 1
        R[i, i+1] <- -2
        R[i, i+2] <- 1
```

```
}

rtop <- rep(0, nw*2)

Rtop <- matrix(rtop, ncol = nw)

Rtop[1, 1] <- 1

Rtop[2, 1] <- -2

Rtop[2, 2] <- 1

R <- rbind(Rtop, R)

rbottom <- rep(0, nw*2)

Rbottom <- matrix(rbottom, ncol = nw)

Rbottom[1, nw - 1] <- 1

Rbottom[1, nw] <- -2

Rbottom[2, nw] <- 1

R <- rbind(R, Rbottom)


oneS <- rep(1,S)

IS <- diag(oneS)

lphi <- length(phi)

psidist <- function(Y, X, phi, Gamma, Omega, S, times)

{

oneS <- rep(1,S)

IS <- diag(oneS)

mu <- oneS%x%phi
```

```
Sigma <- IS%x%Gamma

lphi <- length(phi)

for(i in 1:times)

{

Yt <- Y[((i-1)*S+1):(i*S)]

xt <- X[((i-1)*S+1):(i*S),]

Xt <- NULL

for(k in 1:S)

{

Xt <-rbind(Xt, c(rep(0, (k-1)*lphi), xt[k,], rep(0, (S-k)*lphi)))

}

op <- solve(Omega+Xt%*%Sigma%*%t(Xt))

XSig <- Xt%*%Sigma

mu <- mu + drop(t(XSig)%*%op%*%(Yt-Xt%*%mu))

Sigma <- Sigma - t(XSig)%*%op%*%XSig

}

return(list(mu = mu, Sigma = Sigma))

}

msqrt <- function(x){solve(t(chol(x)))}

j <- 1

while(T)

{
```

```
phi.old <- phi

Gamma.old <- Gamma

Omega.old <- Omega

hats <- psidist(Y=Y, X=X, phi = phi, Gamma = Gamma, Omega = Omega,

S = S, times = times)

mu <- hats$mu

Sigma <- hats$Sigma

Gamma.5inv <- msqrt(Gamma)

blah <- NULL

for(i in 1:S)

{

blah <- rbind(blah, Gamma.5inv%*%

hats$mu[((i-1)*dim(Gamma)[1]+1):(i*dim(Gamma)[1])])

}

phi.comp <- magic(y = blah, X = oneS%x%Gamma.5inv, sp = sp,

S = list(t(R)%*%R), off = ints+2)

phi <- phi.comp$b

sum.Sigma <- 0

sum.diff <- 0

for(i in 1:S)

{

sum.Sigma <- sum.Sigma + Sigma[((i-1)*lphi+1):(i*lphi),
```

```
((i-1)*lphi+1):(i*lphi)]

sum.diff <- sum.diff + (mu[((i-1)*lphi+1):(i*lphi)]-phi)%*%

t(mu[((i-1)*lphi+1):(i*lphi)]-phi)

}

Gamma <- diag(rep(mean(diag((sum.Sigma+sum.diff)/S)), lphi))


d <- length(mu)

sum.XSigmaX <- 0

sum.diff <- 0

for(i in 1:times)

{

Yt <- Y[((i-1)*S+1):(i*S)]

xt <- X[((i-1)*S+1):(i*S),]

Xt <- NULL

for(k in 1:S)

{

Xt <-rbind(Xt, c(rep(0, (k-1)*lphi), xt[k,], rep(0, (S-k)*lphi)))

}

sum.XSigmaX <- sum.XSigmaX + Xt%*%Sigma%*%t(Xt)

sum.diff <- sum.diff + (Yt-Xt%*%mu)%*%t(Yt-Xt%*%mu)

}

Omega <- (sum.XSigmaX + sum.diff)/times
```

```
crit1 <- sum(abs(phi-phi.old))/sum(abs(phi.old))

crit2 <- sum(abs(diag(Gamma - Gamma.old)))/sum(diag(Gamma.old))

crit3 <- sum(abs(diag(Omega - Omega.old)))/sum(diag(Omega.old))

if(crit1 < end.crit & crit2 < end.crit & crit3 < end.crit)

{

return(list(phi = phi, Gamma = Gamma, Omega = Omega, iterations = j,

sp = phi.comp$sp, gcv = phi.comp$score))

break

} else

{

j <- j+1

sp <- max(phi.comp$sp, sp.old)

}

if(j >= maxiter)

{

print("Stopped: Maximum number of iterations reached")

return(list(phi = phi, Gamma = Gamma, Omega = Omega,

crit = c(crit1, crit2, crit3), sp = phi.comp$sp,

gcv = phi.comp$score))

break

}

}
```

```
}
```

B.3.3 Γ is multiple of the identity matrix

```
em <- function(Y, X, phi, Gamma, Omega, S, ints, times, nw, sp,
end.crit = 0.001, maxiter = 100)
{
# Used in estimation of the mixed effect model.
# Arguments:
#    Y - The response vector sorted by fjord
#    X - The covariate matrix sorted by fjord
#    phi - The initial estimate for the mean of the psi's,
#       where the psi's are the random effects for each
#       fjord
#    Gamma - The initial estimate for the covariance matrix
#       of the psi's
#    Omega - The initial estimate for the contemporaneous
#       covariance matrix
#    S - The number of fjords
#    ints - The number of intercepts; if missing is set to be the
#       same as S
#    nw - The number of days per year where measurements are taken
#    sp - The starting value of the smoothing parameter
```

```
#    end.crit - The value for which if the relative change in the
#       regression parameter estimate between the previous
#       iteration and the current estimate is below this value,
#       then convergence is assumed.
#    maxiter - The maximum number of iterations
# Details:
#    Uses the EM algorithm and magic from the mgcv library
#    to estimate the common transfer function model with
#    contemporaneous correlation between the fjords in the
#    mixed effect case.  Assumes that Gamma is a multiple
#    of the identity matrix.
# Values:
#    phi - The estimate of the mean of the random effects psi_s
#    Gamma - The estimate for the covariance matrix of the psi_s
#    Omega - The estimate for the contemporaneous covariance
#       matrix
#    iterations - The number of iterations needed for convergence
#    sp - The estimate for the smoothing parameter
#    gcv - The value of the gcv function at the value of sp
if(missing(ints)){ints <- S}
sp.old <- sp
library(mgcv)
```

```
r <- rep(0, (nw - 2)*nw)

R <- matrix(r, ncol = nw)

for(i in 1:(nw-2))

{

        R[i, i] <- 1

        R[i, i+1] <- -2

        R[i, i+2] <- 1

}

rtop <- rep(0, nw*2)

Rtop <- matrix(rtop, ncol = nw)

Rtop[1, 1] <- 1

Rtop[2, 1] <- -2

Rtop[2, 2] <- 1

R <- rbind(Rtop, R)

rbottom <- rep(0, nw*2)

Rbottom <- matrix(rbottom, ncol = nw)

Rbottom[1, nw - 1] <- 1

Rbottom[1, nw] <- -2

Rbottom[2, nw] <- 1

R <- rbind(R, Rbottom)


oneS <- rep(1,S)
```

```
IS <- diag(oneS)

lphi <- length(phi)

psidist <- function(Y, X, phi, Gamma, Omega, S, times)

{

oneS <- rep(1,S)

IS <- diag(oneS)

mu <- oneS%x%phi

Sigma <- IS%x%Gamma

lphi <- length(phi)

for(i in 1:times)

{

Yt <- Y[((i-1)*S+1):(i*S)]

xt <- X[((i-1)*S+1):(i*S),]

Xt <- NULL

for(k in 1:S)

{

Xt <-rbind(Xt, c(rep(0, (k-1)*lphi), xt[k,], rep(0, (S-k)*lphi)))

}

op <- solve(Omega+Xt%*%Sigma%*%t(Xt))

XSig <- Xt%*%Sigma

mu <- mu + drop(t(XSig)%*%op%*%(Yt-Xt%*%mu))

Sigma <- Sigma - t(XSig)%*%op%*%XSig
```

```
}

return(list(mu = mu, Sigma = Sigma))

}

msqrt <- function(x){solve(t(chol(x)))}

j <- 1

while(T)

{

phi.old <- phi

Gamma.old <- Gamma

Omega.old <- Omega

hats <- psidist(Y=Y, X=X, phi = phi, Gamma = Gamma, Omega = Omega,

S = S, times = times)

mu <- hats$mu

Sigma <- hats$Sigma

Gamma.5inv <- msqrt(Gamma)

blah <- NULL

for(i in 1:S)

{

blah <- rbind(blah, Gamma.5inv%*%

hats$mu[((i-1)*dim(Gamma)[1]+1):(i*dim(Gamma)[1])])

}

phi.comp <- magic(y = blah, X = oneS%x%Gamma.5inv, sp = sp,
```

```
S = list(t(R)%*%R), off = ints+2)

phi <- phi.comp$b

sum.Sigma <- 0

sum.diff <- 0

for(i in 1:S)

{

sum.Sigma <- sum.Sigma + Sigma[((i-1)*lphi+1):(i*lphi),

((i-1)*lphi+1):(i*lphi)]

sum.diff <- sum.diff + (mu[((i-1)*lphi+1):(i*lphi)]-phi)%*%

t(mu[((i-1)*lphi+1):(i*lphi)]-phi)

}

Gamma <- diag(rep(mean(diag((sum.Sigma+sum.diff)/S)), lphi))


d <- length(mu)

sum.XSigmaX <- 0

sum.diff <- 0

for(i in 1:times)

{

Yt <- Y[((i-1)*S+1):(i*S)]

xt <- X[((i-1)*S+1):(i*S),]

Xt <- NULL

for(k in 1:S)
```

```
{
Xt <-rbind(Xt, c(rep(0, (k-1)*lphi), xt[k,], rep(0, (S-k)*lphi)))
}
sum.XSigmaX <- sum.XSigmaX + Xt%*%Sigma%*%t(Xt)
sum.diff <- sum.diff + (Yt-Xt%*%mu)%*%t(Yt-Xt%*%mu)
}
Omega <- (sum.XSigmaX + sum.diff)/times
crit1 <- sum(abs(phi-phi.old))/sum(abs(phi.old))
crit2 <- sum(abs(diag(Gamma - Gamma.old)))/sum(diag(Gamma.old))
crit3 <- sum(abs(diag(Omega - Omega.old)))/sum(diag(Omega.old))
if(crit1 < end.crit & crit2 < end.crit & crit3 < end.crit)
{
return(list(phi = phi, Gamma = Gamma, Omega = Omega, iterations = j,
sp = phi.comp$sp, gcv = phi.comp$score))
break
} else
{
j <- j+1
sp <- max(phi.comp$sp, sp.old)
}
if(j >= maxiter)
{
```

```
print("Stopped: Maximum number of iterations reached")
return(list(phi = phi, Gamma = Gamma, Omega = Omega,
crit = c(crit1, crit2, crit3), sp = phi.comp$sp,
gcv = phi.comp$score))
break
}
}
}
```

B.3.4 Γ for multiplicative error model

```
em <- function(Y, X, phi, var1, var2, var.eta, Omega, S, ints,
times, nw, sp, end.crit = 0.001, maxiter = 100)
{
# Used in estimation of the mixed effect model.
# Arguments:
#    Y - The response vector sorted by fjord
#    X - The covariate matrix sorted by fjord
#    phi - The initial estimate for the mean of the psi's,
#       where the psi's are the random effects for each
#       fjord
#    var1 - The initial value of the variance of the intercept term
#    var2 - The initial value of the variance of the biomass term
```

```
#    Var.eta - The initial value of the variance of the eta's,
#       which appear in the exponents to e
#    Omega - The initial estimate for the contemporaneous
#       covariance matrix
#    S - The number of fjords
#    ints - The number of intercepts; if missing is set to be the
#       same as S
#    nw - The number of days per year where measurements are taken
#    sp - The starting value of the smoothing parameter
#    end.crit - The value for which if the relative change in the
#       regression parameter estimate between the previous
#       iteration and the current estimate is below this value,
#       then convergence is assumed.
#    maxiter - The maximum number of iterations
# Details:
#    Uses the EM algorithm and magic from the mgcv library
#    to estimate the common transfer function model with
#    contemporaneous correlation between the fjords in the
#    mixed effect case.  Assumes that Gamma has an multiplicative
#    error specification.
# Values:
#    phi - The estimate of the mean of the random effects psi_s
```

```
#    var1 - The estimate of the variance for the intercept term
#    var2 - The estimate of the variance for the biomass term
#    var.eta - The estimate of the variance of the eta's
#    Omega - The estimate for the contemporaneous covariance
#       matrix
#    iterations - The number of iterations needed for convergence
#    sp - The estimate for the smoothing parameter
#    gcv - The value of the gcv function at the value of sp
#    mu - The mean vector estimated from the imputation
#    Sigma - The covariance matrix estimated from the imputation
if(missing(ints)){ints <- S}
sp.old <- sp
library(mgcv)
r <- rep(0, (nw - 2)*nw)
R <- matrix(r, ncol = nw)
for(i in 1:(nw-2))
{
        R[i, i] <- 1
        R[i, i+1] <- -2
        R[i, i+2] <- 1
}
rtop <- rep(0, nw*2)
```

```
Rtop <- matrix(rtop, ncol = nw)

Rtop[1, 1] <- 1

Rtop[2, 1] <- -2

Rtop[2, 2] <- 1

R <- rbind(Rtop, R)

rbottom <- rep(0, nw*2)

Rbottom <- matrix(rbottom, ncol = nw)

Rbottom[1, nw - 1] <- 1

Rbottom[1, nw] <- -2

Rbottom[2, nw] <- 1

R <- rbind(R, Rbottom)


oneS <- rep(1,S)

IS <- diag(oneS)

lphi <- length(phi)

psidist <- function(Y, X, phi, Gamma, Omega, S, times)

{

oneS <- rep(1,S)

IS <- diag(oneS)

mu <- oneS%x%phi

Sigma <- IS%x%Gamma

lphi <- length(phi)
```

```
for(i in 1:times)

{

Yt <- Y[((i-1)*S+1):(i*S)]

xt <- X[((i-1)*S+1):(i*S),]

Xt <- NULL

for(k in 1:S)

{

Xt <-rbind(Xt, c(rep(0, (k-1)*lphi), xt[k,], rep(0, (S-k)*lphi)))

}

op <- solve(Omega+Xt%*%Sigma%*%t(Xt))

XSig <- Xt%*%Sigma

mu <- mu + drop(t(XSig)%*%op%*%(Yt-Xt%*%mu))

Sigma <- Sigma - t(XSig)%*%op%*%XSig

}

return(list(mu = mu, Sigma = Sigma))

}

# msqrt <- function(x){solve(t(chol(x)))}

msqrt <- function(x)

{

inv <- 1/diag(x)

sqrtinv <- sqrt(inv)

return(diag(sqrtinv))
```

```
}

j <- 1

Gamma <- diag(c(var1,var2,var.eta*phi[(ints+2):lphi]^2))

while(T)

{

phi.old <- phi

var1.old <- var1

var2.old <- var2

var.eta.old <- var.eta

Omega.old <- Omega

hats <- psidist(Y=Y, X=X, phi = phi, Gamma = Gamma, Omega = Omega,

S = S, times = times)

mu <- hats$mu

Sigma <- hats$Sigma

Gamma.5inv <- msqrt(Gamma)

blah <- NULL

for(i in 1:S)

{

blah <- rbind(blah, Gamma.5inv%*%

hats$mu[((i-1)*dim(Gamma)[1]+1):(i*dim(Gamma)[1])])

}

phi.comp <- magic(y = blah, X = oneS%x%Gamma.5inv, sp = sp,
```

```
S = list(t(R)%*%R), off = ints+2)

phi <- phi.comp$b

sum.Sigma <- 0

sum.diff <- 0

for(i in 1:S)

{

sum.Sigma <- sum.Sigma + Sigma[((i-1)*lphi+1):(i*lphi),

((i-1)*lphi+1):(i*lphi)]

sum.diff <- sum.diff + (mu[((i-1)*lphi+1):(i*lphi)]-phi)%*%

t(mu[((i-1)*lphi+1):(i*lphi)]-phi)

}


# Gamma <- (sum.Sigma+sum.diff)/S

var1 <- (sum.Sigma+sum.diff)[1,1]/S

var2 <- (sum.Sigma+sum.diff)[2,2]/S

sum.eta <- 0

for(i in 1:nw)

{

sum.eta <- sum.eta + (sum.Sigma+sum.diff)[i+1+ints,i+1+ints]/(phi[i+

1+ints])^2

}

var.eta <- sum.eta/(S*nw)
```

```
Gamma <- diag(c(var1,var2,var.eta*phi[(ints+2):lphi]^2))

d <- length(mu)

sum.XSigmaX <- 0

sum.diff <- 0

for(i in 1:times)

{

Yt <- Y[((i-1)*S+1):(i*S)]

xt <- X[((i-1)*S+1):(i*S),]

Xt <- NULL

for(k in 1:S)

{

Xt <-rbind(Xt, c(rep(0, (k-1)*lphi), xt[k,], rep(0, (S-k)*lphi)))

}

sum.XSigmaX <- sum.XSigmaX + Xt%*%Sigma%*%t(Xt)

sum.diff <- sum.diff + (Yt-Xt%*%mu)%*%t(Yt-Xt%*%mu)

}

Omega <- (sum.XSigmaX + sum.diff)/times

crit1 <- sum(abs(phi-phi.old))/sum(abs(phi.old))

crit2 <- sum(abs(diag(Omega - Omega.old)))/sum(diag(Omega.old))

crit3 <- abs(var1-var1.old)/var1.old

crit4 <- abs(var2-var2.old)/var2.old

crit5 <- abs(var.eta-var.eta.old)/var.eta.old
```

```
if(crit1 < end.crit & crit2 < end.crit & crit3 < end.crit &
 crit4 < end.crit & crit5 < end.crit)
{
return(list(phi = phi, var1 = var1, var2 = var2, var.eta = var.eta,
Omega = Omega, iterations = j, sp = phi.comp$sp,
gcv = phi.comp$score, mu = mu, Sigma = Sigma))
break
} else
{
j <- j+1
sp <- max(phi.comp$sp, sp.old)
}
if(j >= maxiter)
{
print("Stopped: Maximum number of iterations reached")
return(list(phi = phi, var1 = var1, var2 = var2, var.eta = var.eta,
Omega = Omega, crit = c(crit1, crit2, crit3, crit4, crit5),
sp = phi.comp$sp, gcv = phi.comp$score, mu = mu, Sigma = Sigma))
break
}
}
}
```

B.4 Bootstrapping

B.4.1 EM estimation used for bootstrapping

```
em1 <- function(Y, X, phi, Gamma, Omega, S, ints, times, nw, sp,
end.crit = 0.001, maxiter = 100)
{
# Used in estimation of the mixed effect model used on the bootstrap
# samples.
# Arguments:
#    Y - The response vector sorted by fjord
#    X - The covariate matrix sorted by fjord
#    phi - The assumed mean of the psi's
#    Gamma - The assumed covariance matrix for the psi's
#    Omega - The assumed contemporaneous covariance matrix
#    S - The number of fjords
#    ints - The number of intercepts; if missing, it is set to
#       be the same as S
#    times - The number of years that measurements were taken
#    nw - The number of days per year where measurements are taken
#    sp - the starting value of the smoothing parameter
#    end.crit - The value for which if the relative change in the
#       regression parameter estimate between the previous
#       iteration and the current estimate is below this value,
```

```
#       then convergence is assumed.
#    maxiter - The maximum number of iterations
# Details:
#    Uses magic from the mgcv library to estimate the common
#    transfer function model with contemporaneous correlation
#    between the fjords.  Does not require iteration.
# Values:
#    model - Output from the implementation of magic, including
#       b - The regression parameter estimates
#       sp - The estimate for the smoothing parameter
#    Sigma - The estimated covariance matrix
if(missing(ints)){ints <- S}
library(mgcv)
r <- rep(0, (nw - 2)*nw)
R <- matrix(r, ncol = nw)
for(i in 1:(nw-2))
{
        R[i, i] <- 1
        R[i, i+1] <- -2
        R[i, i+2] <- 1
}
rtop <- rep(0, nw*2)
```

```
Rtop <- matrix(rtop, ncol = nw)

Rtop[1, 1] <- 1

Rtop[2, 1] <- -2

Rtop[2, 2] <- 1

R <- rbind(Rtop, R)

rbottom <- rep(0, nw*2)

Rbottom <- matrix(rbottom, ncol = nw)

Rbottom[1, nw - 1] <- 1

Rbottom[1, nw] <- -2

Rbottom[2, nw] <- 1

R <- rbind(R, Rbottom)


oneS <- rep(1,S)

IS <- diag(oneS)

lphi <- length(phi)

psidist <- function(Y, X, phi, Gamma, Omega, S, times)

{

oneS <- rep(1,S)

IS <- diag(oneS)

mu <- oneS%x%phi

Sigma <- IS%x%Gamma

lphi <- length(phi)
```

```
for(i in 1:times)

{

Yt <- Y[((i-1)*S+1):(i*S)]

xt <- X[((i-1)*S+1):(i*S),]

Xt <- NULL

for(k in 1:S)

{

Xt <-rbind(Xt, c(rep(0, (k-1)*lphi), xt[k,], rep(0, (S-k)*lphi)))

}

op <- solve(Omega+Xt%*%Sigma%*%t(Xt))

XSig <- Xt%*%Sigma

mu <- mu + drop(t(XSig)%*%op%*%(Yt-Xt%*%mu))

Sigma <- Sigma - t(XSig)%*%op%*%XSig

}

return(list(mu = mu, Sigma = Sigma))

}

msqrt <- function(x){solve(t(chol(x)))}

j <- 1

while(T)

{

phi.old <- phi

hats <- psidist(Y=Y, X=X, phi = phi, Gamma = Gamma, Omega = Omega,
```

```
S = S, times = times)

mu <- hats$mu

Sigma <- hats$Sigma

Gamma.5inv <- msqrt(Gamma)

blah <- NULL

for(i in 1:S)

{

blah <- rbind(blah, Gamma.5inv%*%

hats$mu[((i-1)*dim(Gamma)[1]+1):(i*dim(Gamma)[1])])

}

rows <- nrow(t(R)%*%R)

zeros <- matrix(0, nrow = rows, ncol = ints+1)

Sm <- cbind(zeros,t(R)%*%R)

cols <- ncol(Sm)

zeros <- matrix(0, ncol = cols, nrow = ints+1)

Sm <- rbind(zeros, Sm)

phi.comp <- magic(y = blah, X = oneS%x%Gamma.5inv, H = sp*Sm,

S = list(), sp = rep(0,0), off = rep(0,0))

phi <- phi.comp$b

crit1 <- sum(abs(phi-phi.old))/sum(abs(phi.old))

if(crit1 < end.crit)

{
```

```
return(list(phi = phi, sp = phi.comp$sp))
break
} else
{
j <- j+1
}
if(j >= maxiter)
{
print("Stopped: Maximum number of iterations reached")
return(list(phi = phi, sp = phi.comp$sp))
break
}
}
}
```

B.4.2 Nonparametric bootstrap

```
nonparbootstrap <- function(n, S, D, times, Y, X, mu, Gamma, Omega,
phi.init, sp, maxiter=50)
{
# Arguments:
#    n - The number of bootstrap samples needed
#    S - The number of fjords
```

```
#    D - The number of days per year where measurements are taken
#    times - The number of years
#    Y - The response vector sorted by year
#    X - The covariate matrix sorted by year
#    mu - The assumed imputed mean vector
#    Gamma - The assumed covariance matrix of the psi's
#    Omega - The assumed matrix of the covariance between
#       the fjords
#    phi.init - The initial value of phi for each estimation
#       for the bootstrap sample
#    sp - The assumed value of the smoothing parameter
#    maxiter - The maximum number of iterations for each EM
#       estimation
# Details:
#    Computes n nonparametric bootstrap samples
# Values:
#    phis - The estimated regression parameters for each
#       of the n bootstrap samples
phis <- NULL
sps <- NULL
Yf <- NULL
for(i in 1:S)
```

```
{
Yhati <- X[(1+(i-1)*times):(times*i),]%*%mu[(1+(i-1)*(D+2)):
((D+2)*i)]
Yf <- c(Yf, Yhati)
}
for(j in 1:n)
{
samps <- sample(1:times, times, replace = TRUE)
Xnew <- matrix(rep(0,times*S*(D+2)), nrow = times*S)
Ynew <- rep(0, times*S)
for(i in 1:times)
{
Yf1 <- Yf[i+(0:(S-1))*times]
X1 <- X[i+(0:(S-1))*times,]
Y2 <- Y[samps[i]+(0:(S-1))*times]
Yf2 <- Yf[samps[i]+(0:(S-1))*times]
E2 <- Y2 - Yf2
Y1 <- Yf1 + E2
Xnew[i+(0:(S-1))*times,] <- X1
Ynew[i+(0:(S-1))*times] <- Y1
}
blah <- em1(Y = Ynew, X = Xnew, phi = phi.init, Gamma = Gamma,
```

```
Omega = Omega, S = S, ints = 1, times = times, nw = D,
sp = sp, maxiter = maxiter)
phis <- cbind(phis, blah$phi)
sps <- c(sps, blah$sp)
}
return(phis)
}
```

B.4.3 Parametric bootstrap

```
parbootstrap <- function(n, S, D, times, Y, X, mu, Gamma, Omega,
phi.init, sp, maxiter=50)
{
# Arguments:
#    n - The number of bootstrap samples needed
#    S - The number of fjords
#    D - The number of days per year where measurements are taken
#    times - The number of years
#    Y - The response vector sorted by year
#    X - The covariate matrix sorted by year
#    mu - The assumed imputed mean vector
#    Gamma - The assumed covariance matrix of the psi's
#    Omega - The assumed matrix of the covariance between
```

```
#       the fjords
#    phi.init - The initial value of phi for each estimation
#       for the bootstrap sample
#    sp - The assumed value of the smoothing parameter
#    maxiter - The maximum number of iterations for each EM
#       estimation
# Details:
#    Computes n nonparametric bootstrap samples
# Values:
#    phis - The estimated regression parameters for each
#       of the n bootstrap samples
library(MASS)
phis <- NULL
sps <- NULL
Yf <- NULL
for(i in 1:S)
{
Yhati <- X[(1+(i-1)*times):(times*i),]%*%mu[(1+(i-1)*(D+2)):
((D+2)*i)]
Yf <- c(Yf, Yhati)
}
for(j in 1:n)
```

```
{
Xnew <- matrix(rep(0,times*S*(D+2)), nrow = times*S)
Ynew <- rep(0, times*S)
for(i in 1:times)
{
Yf1 <- Yf[i+(0:(S-1))*times]
X1 <- X[i+(0:(S-1))*times,]
E2 <- mvrnorm(n = 1, mu = rep(0, length(Yf1)), Sigma = Omega)
Y1 <- Yf1 + E2
Xnew[i+(0:(S-1))*times,] <- X1
Ynew[i+(0:(S-1))*times] <- Y1
}
blah <- em1(Y = Ynew, X = Xnew, phi = phi.init, Gamma = Gamma,
Omega = Omega, S = S, ints = 1, times = times, nw = D,
sp = sp, maxiter = maxiter)
phis <- cbind(phis, blah$phi)
sps <- c(sps, blah$sp)
}
return(phis)
}
```

REFERENCES

[1] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19:716–723, 1974.

[2] S. Almon. The distributed lag between capital appropriations and expenditures. *Econometrica*, 33:178 – 196, 1965.

[3] G.E.P. Box, G.M. Jenkins, and G.C. Reinsel. *Time Series Analysis, Forcasting and Control*. Pearson Education, San Fransisco, California, USA, 1994.

[4] K.S. Chan, N.C. Stenseth, M.O. Kittilsen, J. Gjosæter, K. Lkve, T. Smith, S. Tveite, and D. Danielssen. Assessing the effectiveness of releasing cod larvae for stock improvement: using long-term beach-seine monitoring data to resolve a century-long controversy. *Ecological Applications*, 13:3 – 22, 2003.

[5] K.S. Chan, N.C. Stenseth, K. Lekve, and J. Gjosæter. Modeling pulse disturbance impact on cod population dynamics; the 1988 algae bloom of skagerrak, norway. *Ecological Monographs*, 73:151 – 171, 2003.

[6] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society*, B 39:1–38, 1977.

[7] B. Efron and R. Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall/CRC, Boca Raton, 1998.

[8] P.J. Green and B.W. Silverman. *Nonparametric Regression and Generalized Linear Models*. Chapman and Hall, London, England, 1994.

[9] C. Gu. *Smoothing Spline ANOVA Models*. Sringer-Verlag, New York, USA, 2002.

[10] J. D. Hamilton. *Time Series Analysis*. Princeton University Press, Princeton, New Jersey, USA, 1994.

[11] L Isserlis. On certain probable errors and correlation coefficients of multiple frequency distributions with skew regression. *Biometrika*, 11:185–190, 1916.

[12] G. Kitagawa and W. Gersch. *Smoothness Priors Analysis of Time Series*. Springer-Verlag, New York, USA, 1996.

[13] H. Knutsen, C. André, P. Jorde, M. Skogen, E. Thuroczy, and N.C. Stenseth. Transport of north sea cod larva into the skagerrak coastal populations. *Proceedings of the Royal Society of London, B*, 2004.

[14] T.A. Louis. Finding the observed information matrix when using the em algorithm. *Journal of the Royal Statistical Society*, B 44:226–233, 1982.

[15] J. R. Schott. *Matrix Analysis for Statistics*. Wiley Interscience, New York, USA, 1997.

[16] G. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6:461–464, 1978.

[17] R.J. Shiller. A distributed lag estimator derived from smoothness priors. *Econometrica*, 41:775 – 788, 1973.

[18] N.C. Stenseth, P.E. Jorde, K.S. Chan, E. Hansen, H. Knutsen, C. André, M.D. Skogen, and K. Lekve. Ecological and genetic impact of atlantic cod larval drift in the skagerrak. *Proceedings of the Royal Society of London, B*, 2006, in press.

[19] A.W. Van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998.

[20] G. Wahba. Spline models for observational data. *Philadelphia: SIAM*, 1990.

[21] S. Wood. Modelling and smoothing parameter estimation with multiple quadratic penalties. *Journal of Royal Statistical Society*, B62:413–428, 2000.

[22] S. Wood. Stable and efficient multiple smoothing parameter estimation for generalized additive models. *Journal of the American Statistical Association*, 99:673–686, 2004a.

[23] S. Wood. On confidence intervals fo gams based on penalized regression splines. *Technical Report 4-12 Department of Statistics, University of Glasgow*, 2004b.

[24] A. Zellner. An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*, 57:348–368, 1962.