Joint Modeling of Time Series Measures and Recurrent Events and Analysis of the Effects of Air Quality on Respiratory Symptoms

Heping Zhang, Yale University

heping.zhang@yale.edu

Coauthors: Yuanqing Ye, Peter Diggle, and Jian Shi

http://c2s2.yale.edu

Background

Significance

- Exposure to ambient pollutants at concentrations above current US Environmental Protection Agency standards is a risk factor for respiratory symptoms, especially in sensitive children.

Significance

- Exposure to ambient pollutants at concentrations above current US Environmental Protection Agency standards is a risk factor for respiratory symptoms, especially in sensitive children.
- Major components of the pollutant mix of health concern are suspended particulates and ozone.

Significance

- Exposure to ambient pollutants at concentrations above current US Environmental Protection Agency standards is a risk factor for respiratory symptoms, especially in sensitive children.
- Major components of the pollutant mix of health concern are suspended particulates and ozone.
  - Suspended particles are of varying size and chemical composition. Of particular health interest are particles of mass ≤ 10 microns in diameter (PM_{10}), particles of mass ≤ 2.5 microns in diameter ($PM_{2.5}$), and sulfate (SO_4^{2-}).

Existing Studies of Air Quality

Clinical and epidemiologic studies have documented that exposure to atmospheric particulate matter and ozone increases risk in

- hospital admissions for respiratory diseases

Existing Studies of Air Quality

Clinical and epidemiologic studies have documented that exposure to atmospheric particulate matter and ozone increases risk in

- hospital admissions for respiratory diseases
- chronic respiratory diseases

Existing Studies of Air Quality

Clinical and epidemiologic studies have documented that exposure to atmospheric particulate matter and ozone increases risk in

- hospital admissions for respiratory diseases
- chronic respiratory diseases
- and lower and upper respiratory illness

Existing Studies of Air Quality

Clinical and epidemiologic studies have documented that exposure to atmospheric particulate matter and ozone increases risk in

- hospital admissions for respiratory diseases
  - Schwartz et al. (1994a, 1994b) and Thurston et al. (1994)
- chronic respiratory diseases
- and lower and upper respiratory illness

Existing Studies of Air Quality

Clinical and epidemiologic studies have documented that exposure to atmospheric particulate matter and ozone increases risk in

- hospital admissions for respiratory diseases
  - Schwartz et al. (1994a, 1994b) and Thurston et al. (1994)
- chronic respiratory diseases
  - Neas et al. (1995)
- and lower and upper respiratory illness

Existing Studies of Air Quality

Clinical and epidemiologic studies have documented that exposure to atmospheric particulate matter and ozone increases risk in

- hospital admissions for respiratory diseases
  - Schwartz et al. (1994a, 1994b) and Thurston et al. (1994)
- chronic respiratory diseases
  - Neas et al. (1995)
- and lower and upper respiratory illness
  - Schwartz et al. (1994a, 1994b), Thurston et al. (1994), and Peters et al., (1997a, 1997b)

Limitations of Existing Studies

Despite the large volume of studies for the effect of ambient pollutants on respiratory diseases, there is only a limited literature examining the effects of ambient pollutant concentrations on daily respiratory symptoms whilst taking account of daily meteorologic changes.

Limitations of Existing Studies

Despite the large volume of studies for the effect of ambient pollutants on respiratory diseases, there is only a limited literature examining the effects of ambient pollutant concentrations on daily respiratory symptoms whilst taking account of daily meteorologic changes.

One of the major difficulties is the lack of interpretable models that can incorporate such diverse information.

Limitations of Existing Studies

Despite the large volume of studies for the effect of ambient pollutants on respiratory diseases, there is only a limited literature examining the effects of ambient pollutant concentrations on daily respiratory symptoms whilst taking account of daily meteorologic changes.

- Zhang et al. (2000) and Gent et al. (2003)

One of the major difficulties is the lack of interpretable models that can incorporate such diverse information.

Yale Mothers and Infants Health (YMIH) Study
PI: Brian Leaderer, Ph.D.

General Information

- The purpose of the YMIH study was to investigate the health effects of air quality on respiratory symptoms.

General Information

- The purpose of the YMIH study was to investigate the health effects of air quality on respiratory symptoms.
- Data were collected from 237 mothers and their infants in Southwest Virginia for a summer period from June 10 to August 31, 1995.

General Information

- The purpose of the YMIH study was to investigate the health effects of air quality on respiratory symptoms.
- Data were collected from 237 mothers and their infants in Southwest Virginia for a summer period from June 10 to August 31, 1995.
- Symptoms recorded daily include runny or stuffy nose.

General Information

- The purpose of the YMIH study was to investigate the health effects of air quality on respiratory symptoms.
- Data were collected from 237 mothers and their infants in Southwest Virginia for a summer period from June 10 to August 31, 1995.
- Symptoms recorded daily include runny or stuffy nose.
- A general hypothesis is that symptom prevalence is related to air quality as well as to non-specific personal characteristics.

Variables in Our Analysis

Air quality measures include the highest daily temperature (MTMP), humidity (MHUM), COARSE (the difference between PM_{10} and $PM_{2.5}$), and SO_4^{2-} (SO4).

Variables in Our Analysis

Air quality measures include the highest daily temperature (MTMP), humidity (MHUM), COARSE (the difference between PM_{10} and $PM_{2.5}$), and SO_4^{2-} (SO4).

- We denote these four measures by Y_1, Y_2, Y_3, Y_4, respectively.

Variables in Our Analysis

Air quality measures include the highest daily temperature (MTMP), humidity (MHUM), COARSE (the difference between PM_{10} and $PM_{2.5}$), and SO_4^{2-} (SO4).

- We denote these four measures by Y_1, Y_2, Y_3, Y_4, respectively.

We consider three symptom variables for mothers (i.e., runny nose, cough, sore throat) and three for infants (runny nose, cough, general sickness).

Variables in Our Analysis

- Air quality measures include the highest daily temperature (MTMP), humidity (MHUM), COARSE (the difference between PM_{10} and $PM_{2.5}$), and SO_4^{2-} (SO4).
  - We denote these four measures by Y_1, Y_2, Y_3, Y_4, respectively.
- We consider three symptom variables for mothers (i.e., runny nose, cough, sore throat) and three for infants (runny nose, cough, general sickness).
  - These events are denoted by Z, indexed by individual symptom.

Variables in Our Analysis

Air quality measures include the highest daily temperature (MTMP), humidity (MHUM), COARSE (the difference between PM_{10} and $\text{PM}_{2.5}$), and SO_4^{2-} (SO4).

- We denote these four measures by Y_1, Y_2, Y_3, Y_4, respectively.

We consider three symptom variables for mothers (i.e., runny nose, cough, sore throat) and three for infants (runny nose, cough, general sickness).

- These events are denoted by Z, indexed by individual symptom.

Personal characteristics include allergy (ALL), household pets (PETS), number of children (or siblings) in day care (CHDC), and mother's marital status (MS).

Variables in Our Analysis

Background

Yale Mothers and Infants Health (YMIH) Study
PI: Brian Leaderer, Ph.D.
- General Information
- Variables in Our Analysis
- Sample Data
- Data Summary

Literature

Model

Estimation

Simulation Study

Application

Air quality measures include the highest daily temperature (MTMP), humidity (MHUM), COARSE (the difference between PM_{10} and $\mathrm{PM}_{2.5}$), and SO_4^{2-} (SO4).

- We denote these four measures by $Y_1, Y_2, Y_3, Y_4,$ respectively.

We consider three symptom variables for mothers (i.e., runny nose, cough, sore throat) and three for infants (runny nose, cough, general sickness).

- These events are denoted by Z, indexed by individual symptom.

Personal characteristics include allergy (ALL), household pets (PETS), number of children (or siblings) in day care (CHDC), and mother's marital status (MS).

- These variables are denoted by $x_1, \ldots, x_4$, indexed by individual symptom.

Sample Data

DAY	SYMP	MTMP	MHUM	COARSE	SO4	ALL	PETS	CHDC	MS
1	0	86	97	10.30	130.24	1	1	0	1
2	0	88	100	8.00	35.99	1	1	0	1
3	1	69	100	5.94	23.42	1	1	0	1
4	1	72	75	4.74	46.42	1	1	0	1
5	1	80	77	6.98	38.65	1	1	0	1
6	1	80	76	4.81	35.48	1	1	0	1
7	1	81	93	7.87	69.11	1	1	0	1
8	1	80	100	6.66	100.37	1	1	0	1
9	1	81	96	2.85	91.74	1	1	0	1
10	1	78	90	3.82	104.12	1	1	0	1
⋮	⋮		⋮	⋮		⋮	⋮	⋮	⋮
80	0	87	93	8.12	66.01	1	1	0	1
81	1	90	97	7.49	181.98	1	1	0	1
82	0	91	93	10.78	208.98	1	1	0	1
83	1	92	93	7.41	208.44	1	1	0	1

Data Summary

Variable Label	Description	Range	Summary
MTMP	Maximum 24-hour temperature	$69\text{-}100^{\mathrm{o}}$F	85.8 ± 6.9
MHUM	Maximum 24-hour Humidity	$79\text{-}100$	92.3 ± 5.6
COARSE	Coarse mode particles (PM_{10}-$PM_{2.5}$)	1.41-19.79μg/m^3	7.5 ± 3.3
SO4	24-hour sample sulfate level	6.34-306.89nm/m^3	98.3 ± 66.4
ALLERGY	Allergies diagnosed or treated by a doctor	0,1	42%(1.3%)
PETS	Fur-bearing pets kept in the home within the past year	0, 1	46%(1.3%)
CHDC	Number of children in day care(index child excluded)	0-5	45%*(1.3%)
MS	Mother's marital status	0,1	83%(4%)

* for CHDC > 0.

Literature

Limitations of Existing Models

Existing models for the data described above are generally restrictive and sometimes involve somewhat arbitrary decisions.

Limitations of Existing Models

Existing models for the data described above are generally restrictive and sometimes involve somewhat arbitrary decisions.

- Gent et al. (2003) used logistic regression in the context of repeated measures. They used each subject to serve as his or her own control; as a result, personal variables that remained constant during the study could not be included. They also categorized the air quality exposure variables into quintiles for modeling purposes.

Limitations of Existing Models

Existing models for the data described above are generally restrictive and sometimes involve somewhat arbitrary decisions.

- Gent et al. (2003) used logistic regression in the context of repeated measures. They used each subject to serve as his or her own control; as a result, personal variables that remained constant during the study could not be included. They also categorized the air quality exposure variables into quintiles for modeling purposes.
- Zhang et al. (2000) introduced a simple model that uses a binary time series for each individual as the response variable against a battery of covariates.

Zhang et al. (2000) Model

✓ is simple

Zhang et al. (2000) Model

- ✓ is simple
- ✓ enables separate analyses for incidence data, prevalence data, and symptom duration, which are usually difficult to incorporate in a single model

Zhang et al. (2000) Model

- ✓ is simple
- ✓ enables separate analyses for incidence data, prevalence data, and symptom duration, which are usually difficult to incorporate in a single model
- × air quality measures were included as time-varying covariates ignoring the uncertainties in those repeated measures.

Zhang et al. (2000) Model

✓ is simple

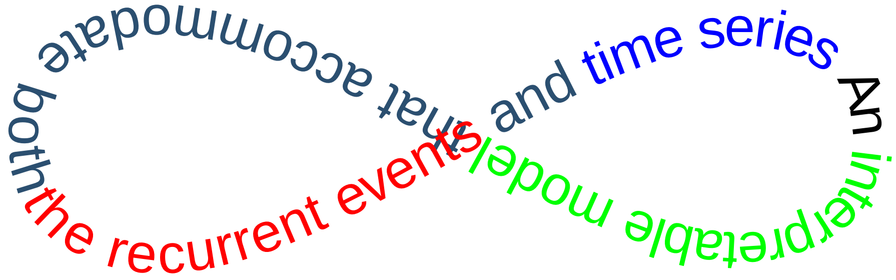
✓ enables separate analyses for incidence data, prevalence data, and symptom duration, which are usually difficult to incorporate in a single model

× air quality measures were included as time-varying covariates ignoring the uncertainties in those repeated measures.

× characterization of binary time series is difficult due to the discrete nature of the series and this limits our ability to conduct rigorous statistical inference.

Zhang et al. (2000) Model

- ✓ is simple
- ✓ enables separate analyses for incidence data, prevalence data, and symptom duration, which are usually difficult to incorporate in a single model
- × air quality measures were included as time-varying covariates ignoring the uncertainties in those repeated measures.
- × characterization of binary time series is difficult due to the discrete nature of the series and this limits our ability to conduct rigorous statistical inference.



Joint Models

Tsiatis, Degruttola and Wulfsohn (1995): evaluate the relationship between the repeated measures of CD4 counts and survival. No recurrent event and no multiple repeated measures.

Joint Models

Tsiatis, Degruttola and Wulfsohn (1995): evaluate the relationship between the repeated measures of CD4 counts and survival. No recurrent event and no multiple repeated measures.

Additional work: Faucett and Thomas (1996), Wulfsohn and Tsiatis (1997), Hogan and Laird (1997a, b), Faucett, Schenker and Elashoff (1998), Finkelstein and Schoenfeld (1999), Vaida and Xu (2000), Henderson, Diggle and Dobson (2000), Xu and Zeger (2001), Wang and Taylor (2001), and Ibrahim et al. (2004)

Joint Models

Tsiatis, Degruttola and Wulfsohn (1995): evaluate the relationship between the repeated measures of CD4 counts and survival. No recurrent event and no multiple repeated measures.

Additional work: Faucett and Thomas (1996), Wulfsohn and Tsiatis (1997), Hogan and Laird (1997a, b), Faucett, Schenker and Elashoff (1998), Finkelstein and Schoenfeld (1999), Vaida and Xu (2000), Henderson, Diggle and Dobson (2000), Xu and Zeger (2001), Wang and Taylor (2001), and Ibrahim et al. (2004)

Joint Models

Background

Yale Mothers and Infants Health (YMIH) Study
PI: Brian Leaderer, Ph.D.

Literature
- Limitations of Existing Models
- Zhang et al. (2000) Model
- Joint Models

Model

Estimation

Simulation Study

Application

Tsiatis, Degruttola and Wulfsohn (1995): evaluate the relationship between the repeated measures of CD4 counts and survival. No recurrent event and no multiple repeated measures.

Additional work: Faucett and Thomas (1996), Wulfsohn and Tsiatis (1997), Hogan and Laird (1997a, b), Faucett, Schenker and Elashoff (1998), Finkelstein and Schoenfeld (1999), Vaida and Xu (2000), Henderson, Diggle and Dobson (2000), Xu and Zeger (2001), Wang and Taylor (2001), and Ibrahim et al. (2004)

Excellent review: Tsiatis and Davidian (2004)

Henderson, Diggle and Dobson (2000): a latent bivariate Gaussian process affects both a repeated measurement sequence and the hazard for an associated event-time.

Model

Decomposition of Time Series

$$Y_k(t) = \mu_k(t) + W_k(t) \tag{1}$$

where $W(t) = \{W_1(t), \ldots, W_m(t)\}$ is a multivariate zero-mean Gaussian process. Thus, $W_k(t)$ is specific to $Y_k(t)$.

Decomposition of Time Series

$$Y_k(t) = \mu_k(t) + W_k(t) \tag{1}$$

where $W(t) = \{W_1(t), \ldots, W_m(t)\}$ is a multivariate zero-mean Gaussian process. Thus, $W_k(t)$ is specific to $Y_k(t)$.

$$W_k(t) = q_k Q(t) + \sigma_k \mathcal{E}_k(t) \tag{2}$$

where $Q(t)$ and $\mathcal{E}(t) = \{\mathcal{E}_1(t), ..., \mathcal{E}_m(t)\}$ are independent Gaussian processes with mean zero and unit variance, and $q_k(\geq 0)$ and $\sigma_k(\geq 0)$ are coefficient parameters.

Decomposition of Time Series

$$Y_k(t) = \mu_k(t) + W_k(t) \tag{1}$$

where $W(t) = \{W_1(t), \ldots, W_m(t)\}$ is a multivariate zero-mean Gaussian process. Thus, $W_k(t)$ is specific to $Y_k(t)$.

$$W_k(t) = q_k Q(t) + \sigma_k \mathcal{E}_k(t) \tag{2}$$

where $Q(t)$ and $\mathcal{E}(t) = \{\mathcal{E}_1(t), ..., \mathcal{E}_m(t)\}$ are independent Gaussian processes with mean zero and unit variance, and $q_k(\geq 0)$ and $\sigma_k(\geq 0)$ are coefficient parameters.

All of the independence conditions are imposed to ensure the uniqueness of the decomposition.

Two Types of Event Transition

$\Uparrow$ Transition from a normal state ($Z = 0$) to an abnormal state ($Z = 1$), denoted by $0 \rightarrow 1$. We assume that the event intensity (hazard rate) for this transition is $\lambda_1(t)$.

Two Types of Event Transition

$\Uparrow$ Transition from a normal state ($Z = 0$) to an abnormal state ($Z = 1$), denoted by $0 \rightarrow 1$. We assume that the event intensity (hazard rate) for this transition is $\lambda_1(t)$.

$\Downarrow$ The reverse $1 \rightarrow 0$, with event intensity $\lambda_2(t)$.

Proportional Hazards

For any individual i,

$$\lambda_{is}(t) = \exp\{X_i(t)^T \beta_s + \mathcal{B}_{is}(t)\}\lambda_s, \tag{3}$$

where

$$\mathcal{B}_{is}(t) = \gamma_{0s} U_i + \gamma_s Q(t), \tag{4}$$

and $\{U_i\}_{i=1}^n$ are subject-specific frailties which follow the standard normal distribution and are independent of $Q(t)$ and $\mathcal{E}(t)$.

Correlation

We write the u-lag correlation functions for $Q(t)$ and $\mathcal{E}_k(t)$ as $\rho_1(\alpha_1, u)$ and $\rho_{2k}(\alpha_{2k}, u)$, respectively.

Correlation

We write the u-lag correlation functions for $Q(t)$ and $\mathcal{E}_k(t)$ as $\rho_1(\alpha_1, u)$ and $\rho_{2k}(\alpha_{2k}, u)$, respectively.

Many different correlation structures have been proposed in the geostatistical literature (see, for example, Matérn, 1960, p.16; Cressie, 1993, pp. 85-86; Chilès and Delfiner, 1999, Section 2.5).

Correlation

We write the u-lag correlation functions for $Q(t)$ and $\mathcal{E}_k(t)$ as $\rho_1(\alpha_1, u)$ and $\rho_{2k}(\alpha_{2k}, u)$, respectively.

We use the powered exponential correlation function:

$$\rho(\alpha, u) = \exp(-\alpha|u|^{\delta}) : 0 < \delta \leq 2. \tag{5}$$

Covariance-Stationarity

Let
$V_1 = \big(\rho_1(\alpha_1, |i-j|)\big)_{d\times d}$, where $\rho_1(\alpha_1, u)$ is defined by (5).

Covariance-Stationarity

Let
$V_1 = \big(\rho_1(\alpha_1, |i-j|)\big)_{d\times d}$, where $\rho_1(\alpha_1, u)$ is defined by (5).

$$V_{2k} = \big(\rho_{2k}(\alpha_{2k}, |i-j|)\big)_{d\times d}.$$

Covariance-Stationarity

Let
$V_1 = \big(\rho_1(\alpha_1, |i-j|)\big)_{d\times d}$, where $\rho_1(\alpha_1, u)$ is defined by (5).

$\diamond$ $Q \overset{d}{\sim} N(0, V_1)$.

$V_{2k} = \big(\rho_{2k}(\alpha_{2k}, |i-j|)\big)_{d\times d}$.

Covariance-Stationarity

Let

$V_1 = \big(\rho_1(\alpha_1, |i-j|)\big)_{d\times d}$, where $\rho_1(\alpha_1, u)$ is defined by (5).

$\diamond$ $Q \overset{d}{\sim} N(0, V_1)$.

$V_{2k} = \big(\rho_{2k}(\alpha_{2k}, |i-j|)\big)_{d\times d}$.

$\diamond$ $\mathcal{E}_k = (\mathcal{E}_k(1), \cdots, \mathcal{E}_k(d))^T \overset{d}{\sim} N(0, V_{2k})$.

Time Series

Let

$$\begin{cases} Y &= (Y_1(1), \cdots, Y_1(d), \cdots, Y_m(1), \cdots, Y_m(d))^T, \\ \mu &= (\mu_1(1), \cdots, \mu_1(d), \cdots, \mu_m(1), \cdots, \mu_m(d))^T. \end{cases}$$

Time Series

Let

$$\begin{cases} Y &= (Y_1(1), \cdots, Y_1(d), \cdots, Y_m(1), \cdots, Y_m(d))^T, \\ \mu &= (\mu_1(1), \cdots, \mu_1(d), \cdots, \mu_m(1), \cdots, \mu_m(d))^T. \end{cases}$$

$Y \overset{d}{\sim} N(\mu, V)$ with

$$V = \begin{pmatrix} q_1^2 V_1 + \sigma_{21}^2 V_{21} & q_1 q_2 V_1 & \cdots & q_1 q_m V_1 \\ q_2 q_1 V_1 & q_2^2 V_1 + \sigma_{22}^2 V_{22} & \cdots & q_2 q_m V_1 \\ \cdots & \cdots & \cdots & \cdots \\ q_m q_1 V_1 & q_m q_2 V_1 & \cdots & q_m^2 V_1 + \sigma_{2m}^2 V_{2m} \end{pmatrix}_{q \times q},$$

where $q = d \times m$.

Counting Processes

$$\begin{cases} N_i^{(1)}(t) = \#\{0 < u \leq t : \ Z_i(u) = 1, \ Z_i(u-) = 0\}, \\ N_i^{(1)}(0) = 0, \end{cases}$$

Counting Processes

$$\begin{cases} N_i^{(1)}(t) = \#\{0 < u \leq t: \; Z_i(u) = 1, \; Z_i(u-) = 0\}, \\ N_i^{(1)}(0) = 0, \end{cases}$$

and

$$\begin{cases} N_i^{(2)}(t) = \#\{0 < u \leq t: \; Z_i(u) = 0, \; Z_i(u-) = 1\}, \\ N_i^{(2)}(0) = 0. \end{cases}$$

Counting Processes

$$\begin{cases} N_i^{(1)}(t) = \#\{0 < u \le t : \; Z_i(u) = 1, \; Z_i(u-) = 0\}, \\ N_i^{(1)}(0) = 0 \; \clubsuit\clubsuit\clubsuit \end{cases}$$

and

$$\begin{cases} N_i^{(2)}(t) = \#\{0 < u \le t : \; Z_i(u) = 0, \; Z_i(u-) = 1\}, \\ N_i^{(2)}(0) = 0 \; \clubsuit\clubsuit\clubsuit \end{cases}$$

Intensities

It follows from (3) that $E\left[dN_i^{(s)}(t) \,|\, Q(t), U_i\right] = \lambda_{is}(t)\,dt$ is given by the model

$$\lambda_{is}(t)\,dt = \exp\{X_i(t)^T\beta_s + \mathcal{B}_{is}(t)\}\lambda_s\,dt, \tag{6}$$

$s = 1, 2$ and $1 \leq i \leq n$.

Stopping Times

$$\begin{aligned}
\tau_{ij}^{(1)} &= \inf\{0 \le t \le T : N_i^{(1)}(t) = j\} \text{ for } 1 \le j \le N_i^{(1)}, \\
\tau_{ij}^{(2)} &= \inf\{0 \le t \le T : N_i^{(2)}(t) = j\} \text{ for } 1 \le j \le N_i^{(2)}.
\end{aligned}$$

- $N_i^{(1)} = N_i^{(2)} + 1$.
  - $0 = \tau_{i0}^{(2)} \le \tau_{i1}^{(1)} \le \tau_{i1}^{(2)} \le \cdots \le \tau_{iN_i^{(2)}}^{(1)} \le \tau_{iN_i^{(2)}}^{(2)} \le \tau_{iN_i^{(1)}}^{(1)} \le T$
- $N_i^{(2)} = N_i^{(1)}$.
  - $0 = \tau_{i0}^{(2)} \le \tau_{i1}^{(1)} \le \tau_{i1}^{(2)} \le \cdots \le \tau_{iN_i^{(2)}}^{(1)} \le \tau_{iN_i^{(2)}}^{(2)} \le T$

Partitioning of the Time Interval

$$C_{i1} \triangleq \begin{cases} \bigcup_{j=1}^{N_i^{(1)}} (\tau^{(2)}_{i(j-1)}, \tau^{(1)}_{ij}] \bigcup \{0\} \bigcup (\tau^{(2)}_{iN_i^{(1)}}, T] & \text{if } N_i^{(1)} = N_i^{(2)}, \\ \bigcup_{j=1}^{N_i^{(1)}} (\tau^{(2)}_{i(j-1)}, \tau^{(1)}_{ij}] \bigcup \{0\} & \text{if } N_i^{(1)} = N_i^{(2)} + 1, \end{cases}$$

Partitioning of the Time Interval

$$C_{i1} \triangleq \begin{cases} \bigcup_{j=1}^{N_i^{(1)}} (\tau_{i(j-1)}^{(2)}, \tau_{ij}^{(1)}] \bigcup \{0\} \bigcup (\tau_{iN_i^{(1)}}^{(2)}, T] & \text{if } N_i^{(1)} = N_i^{(2)}, \\ \bigcup_{j=1}^{N_i^{(1)}} (\tau_{i(j-1)}^{(2)}, \tau_{ij}^{(1)}] \bigcup \{0\} & \text{if } N_i^{(1)} = N_i^{(2)} + 1, \end{cases}$$

and

$$C_{i2} \triangleq \begin{cases} \bigcup_{j=1}^{N_i^{(1)}} (\tau_{ij}^{(1)}, \tau_{ij}^{(2)}] & \text{if } N_i^{(1)} = N_i^{(2)}, \\ \bigcup_{j=1}^{N_i^{(2)}} (\tau_{ij}^{(1)}, \tau_{ij}^{(2)}] \bigcup (\tau_{iN_i^{(1)}}^{(1)}, T] & \text{if } N_i^{(1)} = N_i^{(2)} + 1. \end{cases}$$

Partitioning of the Time Interval

$$C_{i1} \triangleq \begin{cases} \bigcup_{j=1}^{N_i^{(1)}} (\tau_{i(j-1)}^{(2)}, \tau_{ij}^{(1)}] \bigcup \{0\} \bigcup (\tau_{iN_i^{(1)}}^{(2)}, T] & \text{if } N_i^{(1)} = N_i^{(2)}, \\ \bigcup_{j=1}^{N_i^{(1)}} (\tau_{i(j-1)}^{(2)}, \tau_{ij}^{(1)}] \bigcup \{0\} & \text{if } N_i^{(1)} = N_i^{(2)} + 1, \end{cases}$$

and

$$C_{i2} \triangleq \begin{cases} \bigcup_{j=1}^{N_i^{(1)}} (\tau_{ij}^{(1)}, \tau_{ij}^{(2)}] & \text{if } N_i^{(1)} = N_i^{(2)}, \\ \bigcup_{j=1}^{N_i^{(2)}} (\tau_{ij}^{(1)}, \tau_{ij}^{(2)}] \bigcup (\tau_{iN_i^{(1)}}^{(1)}, T] & \text{if } N_i^{(1)} = N_i^{(2)} + 1. \end{cases}$$

$C_{i1} \cup C_{i2} = [0, T]$ and $C_{i1} \cap C_{i2} = \emptyset$

Partitioning of the Time Interval

$$C_{i1} \triangleq \begin{cases} \bigcup_{j=1}^{N_i^{(1)}} (\tau_{i(j-1)}^{(2)}, \tau_{ij}^{(1)}] \bigcup \{0\} \bigcup (\tau_{iN_i^{(1)}}^{(2)}, T] & \text{if } N_i^{(1)} = N_i^{(2)}, \\ \bigcup_{j=1}^{N_i^{(1)}} (\tau_{i(j-1)}^{(2)}, \tau_{ij}^{(1)}] \bigcup \{0\} & \text{if } N_i^{(1)} = N_i^{(2)} + 1, \end{cases}$$

and

$$C_{i2} \triangleq \begin{cases} \bigcup_{j=1}^{N_i^{(1)}} (\tau_{ij}^{(1)}, \tau_{ij}^{(2)}] & \text{if } N_i^{(1)} = N_i^{(2)}, \\ \bigcup_{j=1}^{N_i^{(2)}} (\tau_{ij}^{(1)}, \tau_{ij}^{(2)}] \bigcup (\tau_{iN_i^{(1)}}^{(1)}, T] & \text{if } N_i^{(1)} = N_i^{(2)} + 1. \end{cases}$$

$C_{i1} \cup C_{i2} = [0, T]$ and $C_{i1} \cap C_{i2} = \emptyset$

$N_i^{(1)}$ and $N_i^{(2)}$ jump on C_{i1} and C_{i2}, respectively.

Likelihood Function

$$L(\theta) = L_1(\theta, Y) E_{(Q,U)|Y}\left[L_2(\theta, N \mid Q, U)\right], \tag{7}$$

where

Likelihood Function

$$L(\theta) = L_1(\theta, Y) E_{(Q,U)|Y} \left[L_2(\theta, N \,|\, Q, U) \right], \tag{7}$$

where

- θ contains all parameters

Likelihood Function

$$L(\theta) = L_1(\theta, Y) E_{(Q,U)|Y} \left[L_2(\theta, N \mid Q, U) \right], \tag{7}$$

where

- θ contains all parameters
- $L_1(\theta, Y)$ is the likelihood from the marginal multivariate normal distribution of Y

Likelihood Function

$$L(\theta) = L_1(\theta, Y) E_{(Q,U)|Y} \left[L_2(\theta, N \mid Q, U) \right], \tag{7}$$

where

- θ contains all parameters
- $L_1(\theta, Y)$ is the likelihood from the marginal multivariate normal distribution of Y
- $N = \{(N_i^{(1)}(t),\, N_i^{(2)}(t)) :\, 0 < t \leq T\}_{i=1}^n$

Likelihood Function

$$L(\theta) = L_1(\theta, Y) E_{(Q,U)|Y}\left[L_2(\theta, N \mid Q, U)\right], \qquad (7)$$

where

- θ contains all parameters
- $L_1(\theta, Y)$ is the likelihood from the marginal multivariate normal distribution of Y
- $N = \{(N_i^{(1)}(t),\, N_i^{(2)}(t)) :\, 0 < t \leq T\}_{i=1}^{n}$
- $U = (U_1, U_2, \cdots, U_n)^T$

Conditional Likelihood

$$
\begin{aligned}
& L_2(\theta, N \,|\, Q, U) \\
= & \left(\prod_{i=1}^{n} \prod_{s=1}^{2} \prod_{t \in C_{is}} \lambda_{is}(t)^{\Delta N_i^{(s)}(t)} \right) \times \\
& \exp\left[- \sum_{i=1}^{n} \sum_{s=1}^{2} \int_0^T \lambda_{is}(t) I(u \in C_{is})\, du \right] \\
= & \left(\prod_{i=1}^{n} \prod_{s=1}^{2} \prod_{t \in C_{is}} \left[\exp\{X_i^T(t)\beta_s + \mathcal{B}_{is}(t)\}\lambda_s \right]^{\Delta N_i^{(s)}(t)} \right) \times \\
& \exp\left[- \sum_{i=1}^{n} \sum_{s=1}^{2} \int_0^T \exp\{X_i^T(u)\beta_s + \mathcal{B}_{is}(u)\}\lambda_s I(u \in C_{is})\, du \right],
\end{aligned}
$$

where $I(\cdot)$ is an indicator function, and $\Delta N_i^{(s)}(t) = N_i^{(s)}(t) - N_i^{(s)}(t-)$.

Estimation

Two-stage Procedure

1. Estimate parameters α_l, α_{2k}, q_k and σ_{2k} associated with the time series data Y by maximizing the likelihood function $L_1(\theta, Y)$ in (7).

Two-stage Procedure

1. Estimate parameters α_l, α_{2k}, q_k and σ_{2k} associated with the time series data Y by maximizing the likelihood function $L_1(\theta, Y)$ in (7).
2. Treat the maximum likelihood estimates from Stage 1 as if they are known and use the counting processes model (6) to estimate parameters $\beta_s, \gamma_{0s}, \gamma_s, \lambda_s \ (s = 1, 2)$ by maximizing the likelihood function $E_{(Q,U)|Y}\left[L_2(\theta, N \mid Q, U)\right]$.

Stage 1

We have $Y \overset{d}{\sim} N(\mu, V)$.

Stage 1

We have $Y \stackrel{d}{\sim} N(\mu, V)$. Then,

$$L_1(\theta, Y) = (2\pi)^{-q}\left[\det(V)\right]^{-1/2}\exp\{-\frac{1}{2}(Y-\mu)^T V^{-1}(Y-\mu)\},$$

Stage 1

We have $Y \overset{d}{\sim} N(\mu, V)$. Then,

$$L_1(\theta, Y) = (2\pi)^{-q}\left[\det(V)\right]^{-1/2}\exp\{-\frac{1}{2}(Y-\mu)^T V^{-1}(Y-\mu)\},$$

To reduce computational complexity, we can pre-estimate μ by a weighted moving average,

$$\widehat{\mu}_k(t) = \sum_{s=-m_0}^{m_0} w(s) Y_k(t+s) \tag{8}$$

for pre-specified non-zero weights $\{w(s):\ s = -m_0, -m_0+1, \cdots, 0, \cdots, m_0-1, m_0\}$.

Stage 2

We use the EM algorithm (Dempster, Laird and Rubin, 1977) to maximize

$$E_{(Q,U)|Y}\left[L_2(\theta, N \,|\, Q, U)\right].$$

Stage 2

We use the EM algorithm (Dempster, Laird and Rubin, 1977) to maximize

$$E_{(Q,U)|Y}\left[L_2(\theta, N \,|\, Q, U)\right].$$

Q and U are the unobserved data and N is observed, so the complete likelihood is the joint density of (N, Q, U).

Stage 2

We use the EM algorithm (Dempster, Laird and Rubin, 1977) to maximize

$$E_{(Q,U)|Y}\left[L_2(\theta, N \,|\, Q, U)\right].$$

Q and U are the unobserved data and N is observed, so the complete likelihood is the joint density of (N, Q, U).

The EM algorithm starts with an initial value $\theta^{(0)}$, and then evaluates the expectation of the log likelihood of (Q, U) conditional on N, denoted by $E_{\theta^{(0)}}[l_2(\theta, N, Q, U)|N]$.

Stage 2

We use the EM algorithm (Dempster, Laird and Rubin, 1977) to maximize

$$E_{(Q,U)|Y}\left[L_2(\theta, N\,|\,Q, U)\right].$$

Q and U are the unobserved data and N is observed, so the complete likelihood is the joint density of (N, Q, U).

The EM algorithm starts with an initial value $\theta^{(0)}$, and then evaluates the expectation of the log likelihood of (Q, U) conditional on N, denoted by $E_{\theta^{(0)}}[l_2(\theta, N, Q, U)|N]$.

- This expectation involves integral of $U = \{U_i\}_{i=1}^{83}$ and Q, where U is subject specific frailty and Q is random process.

Stage 2

We use the EM algorithm (Dempster, Laird and Rubin, 1977) to maximize

$$E_{(Q,U)|Y}\left[L_2(\theta, N \,|\, Q, U)\right].$$

Q and U are the unobserved data and N is observed, so the complete likelihood is the joint density of (N, Q, U).

The EM algorithm starts with an initial value $\theta^{(0)}$, and then evaluates the expectation of the log likelihood of (Q, U) conditional on N, denoted by $E_{\theta^{(0)}}[l_2(\theta, N, Q, U)|N]$.

- This expectation involves integral of $U = \{U_i\}_{i=1}^{83}$ and Q, where U is subject specific frailty and Q is random process.
- Gibbs sampler is used to approximate this high dimensional integral.

Stage 2

We use the EM algorithm (Dempster, Laird and Rubin, 1977) to maximize

$$E_{(Q,U)|Y}\left[L_2(\theta, N \,|\, Q, U)\right].$$

Q and U are the unobserved data and N is observed, so the complete likelihood is the joint density of (N, Q, U).

The EM algorithm starts with an initial value $\theta^{(0)}$, and then evaluates the expectation of the log likelihood of (Q, U) conditional on N, denoted by $E_{\theta^{(0)}}[l_2(\theta, N, Q, U)|N]$.

- This expectation involves integral of $U = \{U_i\}_{i=1}^{83}$ and Q, where U is subject specific frailty and Q is random process.
- Gibbs sampler is used to approximate this high dimensional integral.

In the maximization step, we use a Newton-Raphson algorithm to maximize $E_{\theta^{(0)}}[l_2(\theta, N, Q, U)|N]$ and obtain an updated point estimate for θ.

Simulation Study

Stage 1: Time Series Model

- Using model (2) and assuming $\sigma_k = q_k$, we generated a two dimensional time series Y, i.e., $\{Y(t) = (Y_1(t), Y_2(t))^T\}_{t=1}^d$ for d days, where d was chosen to be either 30 or 50.

Stage 1: Time Series Model

- Using model (2) and assuming $\sigma_k = q_k$, we generated a two dimensional time series Y, i.e., $\{Y(t) = (Y_1(t), Y_2(t))^T\}_{t=1}^d$ for d days, where d was chosen to be either 30 or 50.
- The model for Y_k is $Y_k(t) = \mu_k(t) + q_k Q(t) + q_k \mathcal{E}_k(t)$.

Stage 1: Time Series Model

- Using model (2) and assuming $\sigma_k = q_k$, we generated a two dimensional time series Y, i.e., $\{Y(t) = (Y_1(t), Y_2(t))^T\}_{t=1}^d$ for d days, where d was chosen to be either 30 or 50.
- The model for Y_k is $Y_k(t) = \mu_k(t) + q_k Q(t) + q_k \mathcal{E}_k(t)$.
- We used the correlation families (5). To demonstrate that assuming $\delta = 1$ for the modeling has only a small effect on the estimation, we generated data with the true δ taking values 0.5 and 2.0.

Stage 1: Time Series Model

- Using model (2) and assuming $\sigma_k = q_k$, we generated a two dimensional time series Y, i.e., $\{Y(t) = (Y_1(t), Y_2(t))^T\}_{t=1}^d$ for d days, where d was chosen to be either 30 or 50.
- The model for Y_k is $Y_k(t) = \mu_k(t) + q_k Q(t) + q_k \mathcal{E}_k(t)$.
- We used the correlation families (5). To demonstrate that assuming $\delta = 1$ for the modeling has only a small effect on the estimation, we generated data with the true δ taking values 0.5 and 2.0.
- Each simulation was replicated 1000 times.

Effect of Correlation Parameter $\delta = .5$

Parameter	True Value	d=30		d=50	
		Estimate	S.E.	Estimate	S.E.
α	.81	1.375	1.732	1.081	0.622
q_1	1.0	0.916	0.130	0.935	0.102
q_2	1.0	0.908	0.131	0.940	0.104

Effect of Correlation Parameter $\delta = 2$

Parameter	True Value	d=30		d=50	
		Estimate	S.E.	Estimate	S.E.
α	0.81	0.971	0.334	0.897	0.208
q_1	1.0	0.962	0.138	0.975	0.099
q_2	1.0	0.956	0.132	0.975	0.104

Effect of Nonstationarity

We used the following model to simulate a non-stationary process

$$Y_k(t) = \mu_k(t) + q_k Q(t) + \sigma(t)\mathcal{E}_k(t), \tag{9}$$

where $Q(t)$ and $\mathcal{E}_k(t)$ are independent stationary Gaussian processes, whilst the function $\sigma(t)$ was generated from the χ^2_1 distribution at the discrete time points to introduce the nonstationarity for $Y_k(t)$.

Effect of Nonstationarity

We used the following model to simulate a non-stationary process

$$Y_k(t) = \mu_k(t) + q_k Q(t) + \sigma(t)\mathcal{E}_k(t), \tag{9}$$

where $Q(t)$ and $\mathcal{E}_k(t)$ are independent stationary Gaussian processes, whilst the function $\sigma(t)$ was generated from the χ_1^2 distribution at the discrete time points to introduce the nonstationarity for $Y_k(t)$.

When $d = 30$, in roughly 10% of the simulations our estimation procedure failed to converge. When $d = 50$, the estimation procedure failed to converge in about 4% of the simulations. This computational problem is due to the difficulty of estimating α under the stationary assumption.

Parameter Estimates under Nonstationarity

Parameter	d=30		d=50	
	Estimate	S.E.	Estimate	S.E.
α	1.846	0.885	1.775	0.689
q_1	0.937	0.459	0.940	0.376
q_2	0.903	0.443	0.916	0.374
σ_1	1.536	0.659	1.586	0.551
σ_2	1.498	0.619	1.578	0.539

($\delta = 0.5$)

Stage 2: Counting Processes

- $X_1 \equiv 1$ and $X_2 \overset{d}{\sim} Uniform(0,1)$.

Stage 2: Counting Processes

- $X_1 \equiv 1$ and $X_2 \overset{d}{\sim} Uniform(0, 1)$.
- The counting processes $N^{(1)}$ and $N^{(2)}$ were generated with intensities $\lambda_1(t)$ and $\lambda_2(t)$ defined by (3) and (4), respectively.

Stage 2: Counting Processes

- $X_1 \equiv 1$ and $X_2 \stackrel{d}{\sim} Uniform(0,1)$.
- The counting processes $N^{(1)}$ and $N^{(2)}$ were generated with intensities $\lambda_1(t)$ and $\lambda_2(t)$ defined by (3) and (4), respectively.
- The autocorrelation was again $\rho(1,t)$.

Stage 2: Counting Processes

- $X_1 \equiv 1$ and $X_2 \stackrel{d}{\sim} Uniform(0, 1)$.
- The counting processes $N^{(1)}$ and $N^{(2)}$ were generated with intensities $\lambda_1(t)$ and $\lambda_2(t)$ defined by (3) and (4), respectively.
- The autocorrelation was again $\rho(1, t)$.
- To generate stopping times $\{\tau_{i1}^{(1)}, \tau_{i1}^{(2)}, \tau_{i2}^{(1)}, \tau_{i2}^{(2)}, \cdots, \tau_{ij}^{(1)}, \tau_{ij}^{(2)}, \cdots\}$, we first generated $\tau_{i1}^{(1)}$ based on the conditional distribution of $\tau_{i1}^{(1)} | \tau_{i0}^{(2)}$, then generated $\tau_{i1}^{(2)}$ based on the conditional distribution $\tau_{i1}^{(2)} | \tau_{i1}^{(1)}$, and so on, stopping when the last value was larger than or equal to d.

Other Settings

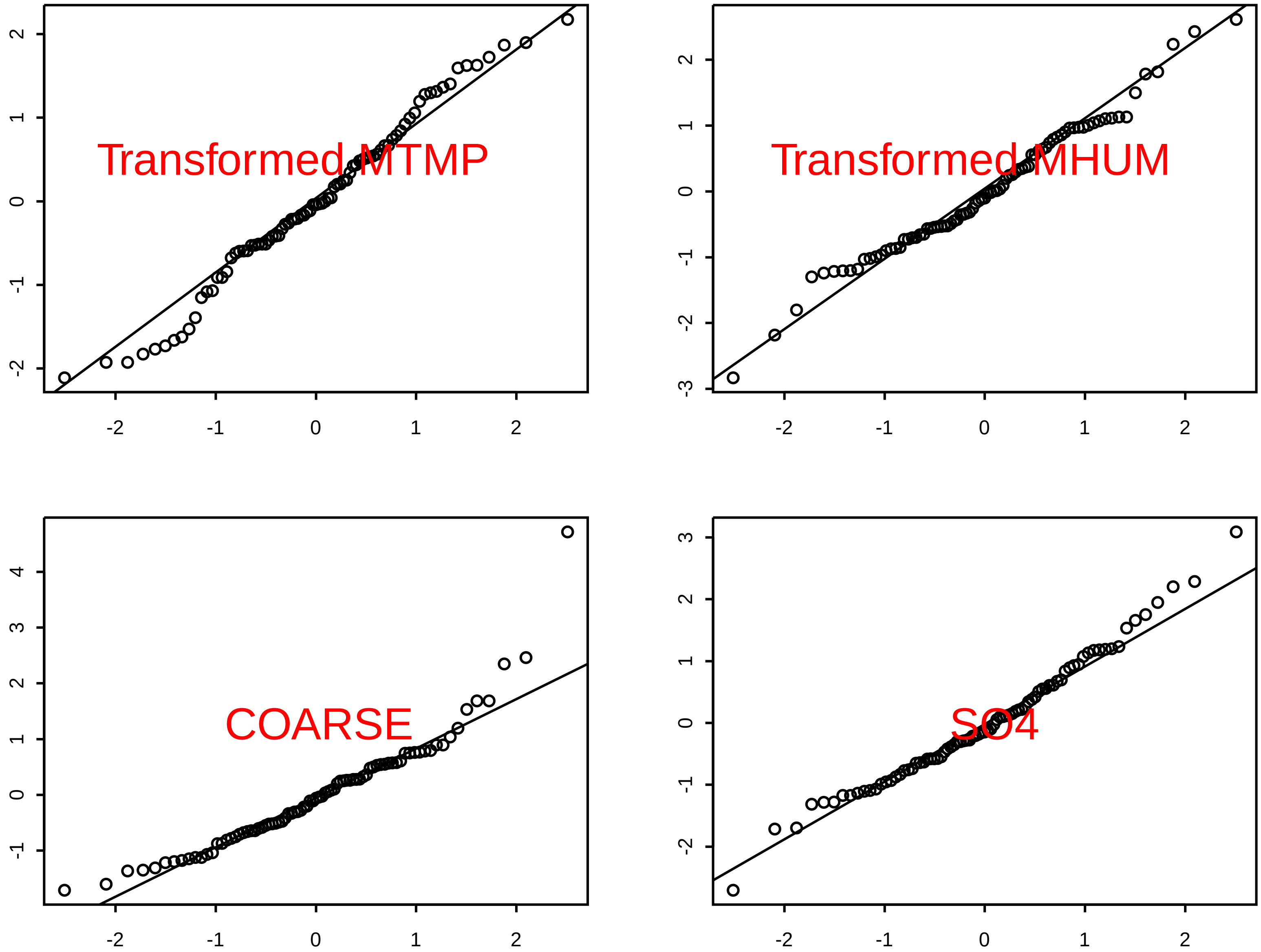
- The simulation was replicated 100 times.
- In each simulation, we used $n = 100$ subjects.
- The number of Gibbs samples depended on the EM iteration and was chosen large enough to minimize numerical differences.
  - It was set at 500, 2000 and 10000 for iterations from 1 to 20, from 20 to 40, and over 40, respectively (Booth and Hobert 1999, McCulloch 1997).
  - The maximum number of EM iterations was set at 100.
- The standard errors of the estimated parameters were calculated using the observed information matrix, based on the formula given by Louis (1982).

Estimation of Covariate Effects

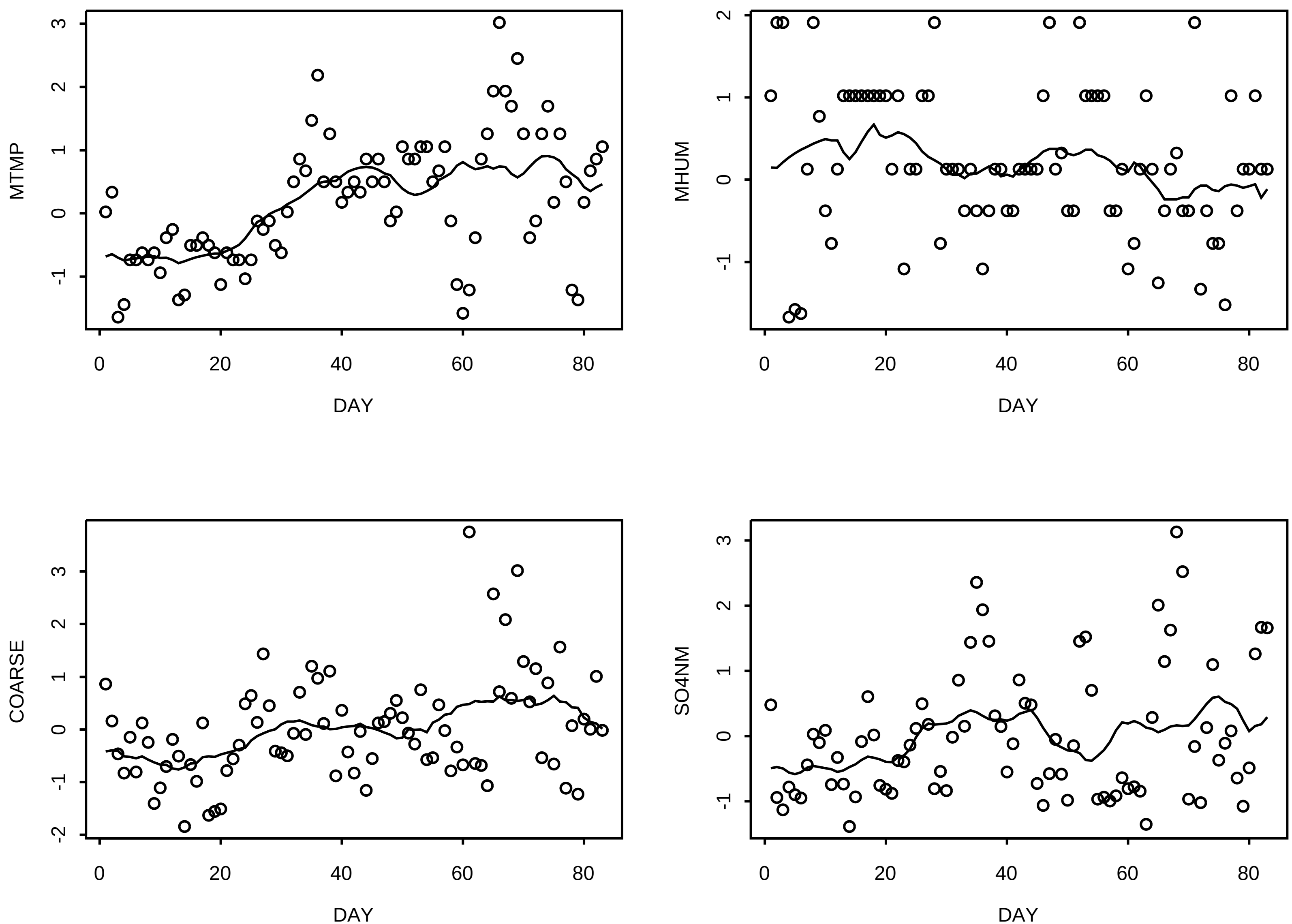
Parameter	True Value	Average	S.E.
γ_{11}	.5	0.43	0.089
γ_{12}	1.0	0.88	0.151
γ_{01}	1.0	0.75	0.095
γ_{02}	1.0	0.81	0.130
β_{11}	-2.5	-2.55	0.236
β_{12}	1.0	0.88	0.334
β_{21}	-4.0	-3.86	0.311
β_{22}	1.5	1.35	0.357

Application

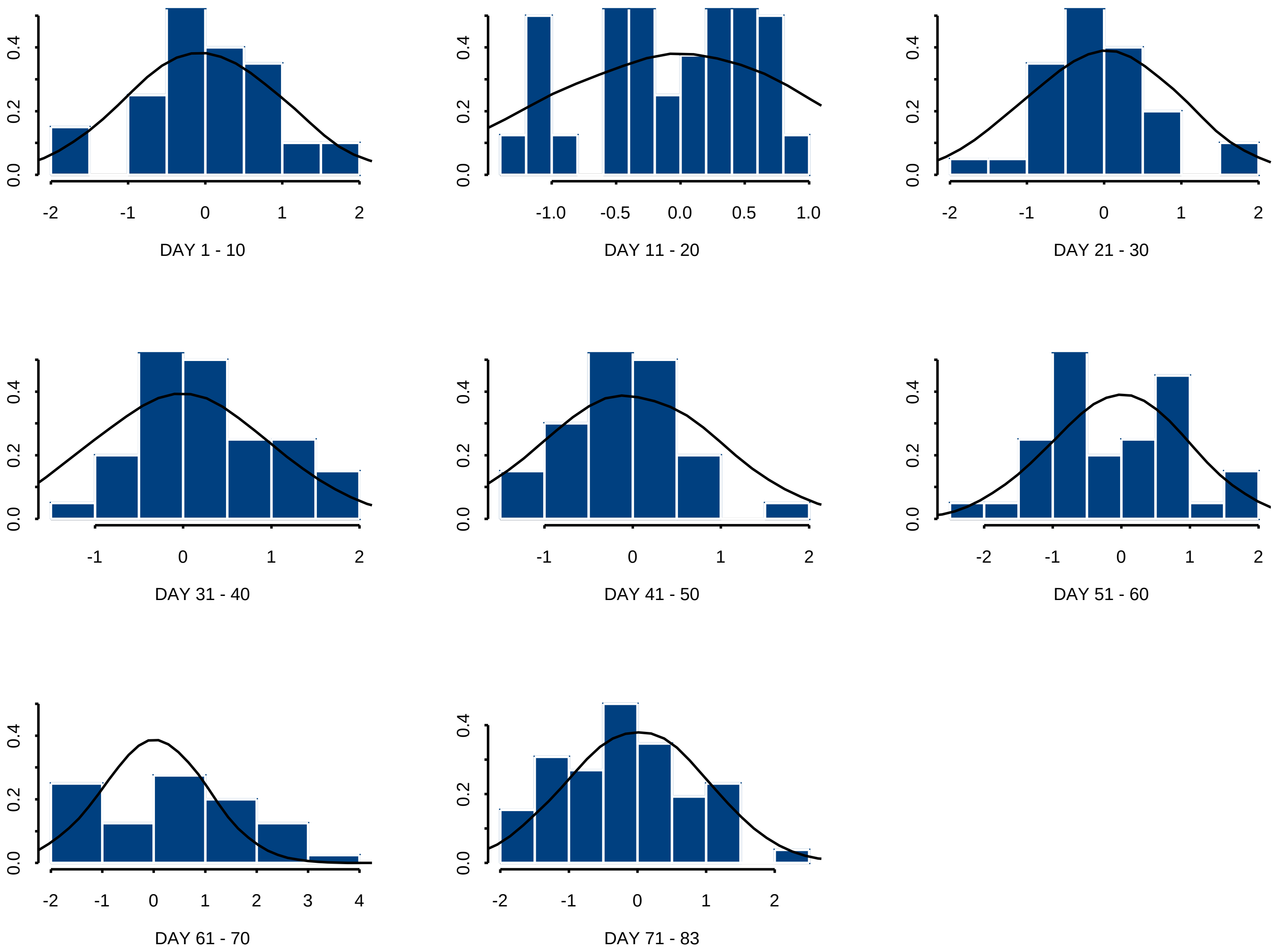
Normality



Air Quality Measures



Residual Plots



Mothers' Predictors for $\lambda_1(t)$

Variable	Runny Nose		Cough	
	Coeff.	SE	Coeff.	SE
$Q_1(t)$	0.025	0.082	-0.043	0.119
U_i	1.092	0.135	1.571	0.199
COARSE	0.404	0.202	0.595	0.285
MTMP	0.146	0.140	0.195	0.195
SO4	0.226	0.238	0.642	0.334
MHUM	-0.644	0.356	-1.029	0.504
ALLERGY	0.598	0.241	0.444	0.354
PETS	0.526	0.244	0.245	0.377
MS	0.584	0.379	0.080	0.515
CHDC	-0.252	0.154	-0.366	0.241

Mothers' Predictors for $\lambda_2(t)$

Variable	Runny Nose		Cough	
	Coeff.	SE	Coeff.	SE
$Q_1(t)$	0.065	0.082	0.074	0.111
U_i	0.004	0.139	0.115	0.163
COARSE	-0.267	0.202	0.228	0.301
MTMP	-0.185	0.147	0.109	0.226
SO4	-0.231	0.252	0.194	0.382
MHUM	0.544	0.358	-0.480	0.527
ALLERGY	-0.255	0.182	0.032	0.262
PETS	0.209	0.172	0.163	0.307
MS	-0.576	0.312	-0.290	0.423
CHDC	0.046	0.133	-0.009	0.214

Infants' Predictors for $\lambda_1(t)$

Variable	Runny Nose		Cough	
	Coeff.	SE	Coeff.	SE
$Q_1(t)$	-0.188	0.081	0.038	0.109
U	0.811	0.107	1.000	0.153
COARSE	-0.159	0.157	-0.425	0.222
MTMP	-0.220	0.107	-0.321	0.151
SO4	-0.419	0.188	-0.653	0.266
MHUM	-0.025	0.284	0.676	0.401
PETS	-0.018	0.176	0.092	0.248
MS	0.372	0.254	-0.341	0.311
CHDC	-0.110	0.109	-0.245	0.159

Infants' Predictors for $\lambda_2(t)$

Variable	Runny Nose		Cough	
	Coeff.	SE	Coeff.	SE
$Q_1(t)$	0.033	0.070	-0.131	0.109
U	0.152	0.076	0.031	0.131
COARSE	0.170	0.156	0.167	0.225
MTMP	0.105	0.110	0.112	0.156
SO4	0.101	0.189	0.079	0.269
MHUM	-0.023	0.285	0.143	0.430
PETS	-0.169	0.138	0.150	0.199
MS	-0.361	0.199	-0.246	0.235
CHDC	0.038	0.098	0.059	0.144

Conclusion

There are differences in the etiology of respiratory symptoms between mothers and infants.

Conclusion

There are differences in the etiology of respiratory symptoms between mothers and infants.

- Coarse particles of mass between 2.5 and 10 microns in diameter increased the risks of mothers' runny nose and cough symptoms, but not on infants' symptoms.

Conclusion

There are differences in the etiology of respiratory symptoms between mothers and infants.

- Coarse particles of mass between 2.5 and 10 microns in diameter increased the risks of mothers' runny nose and cough symptoms, but not on infants' symptoms.
- The sulfate level was negatively associated with the risk of infants' runny nose and cough symptoms, but not on the mothers' symptoms.

Conclusion

There are differences in the etiology of respiratory symptoms between mothers and infants.

- Coarse particles of mass between 2.5 and 10 microns in diameter increased the risks of mothers' runny nose and cough symptoms, but not on infants' symptoms.
- The sulfate level was negatively associated with the risk of infants' runny nose and cough symptoms, but not on the mothers' symptoms.
- High level of humidity is negatively associated with the mothers' cough incidence, but not on infants' symptoms.

Conclusion

There are differences in the etiology of respiratory symptoms between mothers and infants.

- Coarse particles of mass between 2.5 and 10 microns in diameter increased the risks of mothers' runny nose and cough symptoms, but not on infants' symptoms.
- The sulfate level was negatively associated with the risk of infants' runny nose and cough symptoms, but not on the mothers' symptoms.
- High level of humidity is negatively associated with the mothers' cough incidence, but not on infants' symptoms.

Such differences reveal not only the sensitivity of the mothers and infants to the air quality, but also call for further understanding of the differences.

Conclusion

There are differences in the etiology of respiratory symptoms between mothers and infants.

- Coarse particles of mass between 2.5 and 10 microns in diameter increased the risks of mothers' runny nose and cough symptoms, but not on infants' symptoms.
- The sulfate level was negatively associated with the risk of infants' runny nose and cough symptoms, but not on the mothers' symptoms.
- High level of humidity is negatively associated with the mothers' cough incidence, but not on infants' symptoms.

Such differences reveal not only the sensitivity of the mothers and infants to the air quality, but also call for further understanding of the differences.
It is possible that actions taken to overcome humidity by mothers may inadvertently affect the infants.