

ANALYSIS OF LONGITUDINAL DATA OF MIXED TYPES USING A
STATE SPACE MODEL APPROACH

By

Billy K. S. Ching

B. Sc. University of Hong Kong

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES
DEPARTMENT OF STATISTICS

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

June 1997

© Billy K. S. Ching, 1997

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of STATISTICS

The University of British Columbia
Vancouver, Canada

Date JULY 29, 1997

Abstract

A new method for multivariate regression analysis of longitudinal data of mixed types is applied to the data from a sub-study of the Betaseron multicenter clinical trial in relapsing-remitting multiple sclerosis (MS) (The IFNB Multiple Sclerosis Study Group, 1993). The sub-study is based on a cohort of 52 patients at one center (University of British Columbia) for frequent magnetic resonance imagings (MRIs) for analysis of disease activity over the first two years of the trial (Paty, Li, the UBC MS/MRI Study Group and the IFNB Multiple Sclerosis Study Group, 1993). We consider a bivariate response vector with two different data types as components. The first component is a positive continuous variable and the second one is a count variable. We use a state space model approach based on the Tweedie class of exponential dispersion models assuming conditional independence of the two components given a latent gamma Markov process. The latent process is interpreted as the underlying severity of the disease whereas the observations reflect the symptoms. One advantage the new method offers is that it enables the examination of patterns over time. Not only can it identify the presence of treatment effect, but also the nature of the effect. It has well been established that Betaseron has substantially altered the natural history of MS in a properly controlled clinical trial (The IFNB Multiple Sclerosis Study Group, 1993). The main objective of this thesis is to illustrate the utilization of the new method using this data set and to extract additional valuable information from the data.

Table of Contents

Abstract	ii
List of Tables	vi
List of Figures	vii
Acknowledgments	xi
1 Introduction	1
2 Review of literature	4
2.1 Marginal models	5
2.2 Random effects models	6
2.3 Transition models	8
2.4 State space models	9
3 Tweedie state space models	11
3.1 The model	11
3.2 The Kalman filter and smoother	15
3.3 Parameter estimation	17
3.3.1 The Kalman estimating equation	17
3.3.2 Estimation of dispersion parameters	20
3.4 Residual analysis	21
4 Analysis of Betaseron data	25

4.1	Background	25
4.1.1	Description of the experiment	25
4.1.2	Description of the data	27
4.2	Initial data analysis	29
4.2.1	Burden of disease	29
4.2.2	Active lesions	35
4.2.3	Covariates	37
4.3	Model identification	40
4.4	Model checking	47
4.4.1	Full data set	47
4.4.2	Data set with 446* removed	59
4.4.3	Data set with 446* and 545* removed	73
4.5	The final model	73
5	Discussion	81
	Bibliography	84
	Appendix A Exponential dispersion models	87
	Appendix B Godambe information matrix	90
	B.1 Moment calculations	90
	B.2 Derivatives of Kalman filter and smoother	91
	B.3 Variability matrix	93
	B.4 Sensitivity matrix	94
	Appendix C Residuals plots for each patient in the placebo group	96

Appendix D	Residuals plots for each patient in the low dose group	112
Appendix E	Residuals plots for each patient in the high dose group	128

List of Tables

4.1	Descriptive statistics for average <i>relative burden</i> values	30
4.2	Descriptive statistics for average <i>active lesions</i> values	35
4.3	Descriptive statistics for <i>age</i> , <i>duration</i> , <i>EDSS</i> and <i>log area</i>	37
4.4	Counts for <i>origin</i> and <i>gender</i>	39
4.5	Estimates and standard errors for baseline covariate effects	43
4.6	Estimates and standard errors for the long-term covariate (<i>time</i>) effects for patients in the placebo group	44
4.7	Estimates and standard errors for the long-term covariate effects (<i>time</i>) for patients in the low dose group	45
4.8	Estimates and standard errors for the long-term covariate effects (<i>time</i>) for patients in the high dose group	46
4.9	Empirical variance-covariance matrix for standardized residuals	71
4.10	Estimates for the long-term covariates effects	77
4.11	Estimates for the baseline covariates effects	77
4.12	Estimates and standard errors for the time effects for each patient; placebo (PL), low dose (LD) and high dose (HD).	79
4.13	Estimates and standard errors for baseline covariate effects	79
A.14	Summary of the Tweedie exponential dispersion models	88

List of Figures

4.1	Boxplots of average <i>relative burden</i> values by treatment group	31
4.2	Boxplots of <i>relative burden</i> values for each patient	32
4.3	Plots of <i>relative burden</i> values versus time for each patient	33
4.4	Boxplots of average <i>active lesions</i> values across treatment group	34
4.5	Plots of <i>active lesions</i> values versus time	36
4.6	Boxplots of covariates values across treatment group	38
4.7	Residuals plots for patient #446 in the placebo group	48
4.8	Residuals plots for patient #545 in the placebo group	49
4.9	Residuals plots for patient #502 in the low dose group	51
4.10	Residuals plots for patient #500 in the high dose group	52
4.11	Residuals against log-linear predictors for the two categories	55
4.12	Residuals against log-linear predictors for the latent process	56
4.13	Residuals against lag 1 residuals for the two categories	57
4.14	Residuals against lag 1 residuals for the latent process	58
4.15	Residuals plots for patient #446 in the placebo group	60
4.16	Residuals plots for patient #545 in the placebo group	61
4.17	Residuals plots for patient #502 in the low dose group	62
4.18	Residuals plots for patient #500 in the high dose group	63
4.19	Residuals against log-linear predictors for the two categories	65
4.20	Residuals against log-linear predictors for the latent process	65
4.21	Residuals against lag 1 residuals for the two categories	66

4.22	Residuals against lag 1 residuals for the latent process	67
4.23	Residuals against time for the two categories	67
4.24	Residuals against time for the latent process	68
4.25	Smoother residuals against log-linear predictors for the random effects . .	69
4.26	Smoother residuals against baseline covariates	70
4.27	Autocorrelation functions for average filter residuals: (a) to (c) are the au- torecorrelations functions; (d) to (e) are the partial autocorrelations functions	72
4.28	Estimated latent processes for each patient	75
C.29	Residuals plots for patient #420 in the placebo group	97
C.30	Residuals plots for patient #421 in the placebo group	98
C.31	Residuals plots for patient #443 in the placebo group	99
C.32	Residuals plots for patient #449 in the placebo group	100
C.33	Residuals plots for patient #451 in the placebo group	101
C.34	Residuals plots for patient #498 in the placebo group	102
C.35	Residuals plots for patient #501 in the placebo group	103
C.36	Residuals plots for patient #504 in the placebo group	104
C.37	Residuals plots for patient #507 in the placebo group	105
C.38	Residuals plots for patient #522 in the placebo group	106
C.39	Residuals plots for patient #523 in the placebo group	107
C.40	Residuals plots for patient #539 in the placebo group	108
C.41	Residuals plots for patient #540 in the placebo group	109
C.42	Residuals plots for patient #550 in the placebo group	110
C.43	Residuals plots for patient #565 in the placebo group	111
D.44	Residuals plots for patient #419 in the low dose group	113
D.45	Residuals plots for patient #424 in the low dose group	114

D.46 Residuals plots for patient #448 in the low dose group	115
D.47 Residuals plots for patient #450 in the low dose group	116
D.48 Residuals plots for patient #452 in the low dose group	117
D.49 Residuals plots for patient #499 in the low dose group	118
D.50 Residuals plots for patient #508 in the low dose group	119
D.51 Residuals plots for patient #521 in the low dose group	120
D.52 Residuals plots for patient #525 in the low dose group	121
D.53 Residuals plots for patient #542 in the low dose group	122
D.54 Residuals plots for patient #544 in the low dose group	123
D.55 Residuals plots for patient #547 in the low dose group	124
D.56 Residuals plots for patient #548 in the low dose group	125
D.57 Residuals plots for patient #564 in the low dose group	126
D.58 Residuals plots for patient #568 in the low dose group	127
E.59 Residuals plots for patient #422 in the high dose group	129
E.60 Residuals plots for patient #444 in the high dose group	130
E.61 Residuals plots for patient #445 in the high dose group	131
E.62 Residuals plots for patient #453 in the high dose group	132
E.63 Residuals plots for patient #454 in the high dose group	133
E.64 Residuals plots for patient #497 in the high dose group	134
E.65 Residuals plots for patient #503 in the high dose group	135
E.66 Residuals plots for patient #506 in the high dose group	136
E.67 Residuals plots for patient #524 in the high dose group	137
E.68 Residuals plots for patient #526 in the high dose group	138
E.69 Residuals plots for patient #541 in the high dose group	139
E.70 Residuals plots for patient #543 in the high dose group	140

E.71 Residuals plots for patient #546 in the high dose group	141
E.72 Residuals plots for patient #549 in the high dose group	142
E.73 Residuals plots for patient #566 in the high dose group	143

Acknowledgments

I would like to thank Dr. Bent Jørgensen for his guidance and supervision throughout the development of this thesis. The data set used in this thesis were collected at the UBC MS Clinic as part of the UBC Frequent MRI Sub-study of the Betaseron Clinical Trial sponsored by Berlex Laboratories, Richmond, California. I would like to thank Berlex Laboratories for making the data available to us. I am grateful to Dr. John Petkau for introducing us to the data, his careful reading of the manuscript and subsequent helpful suggestions. The assistance from Dr. Peter X.K. Song concerning computing issues is also gratefully acknowledged.

Chapter 1

Introduction

This thesis focuses on the Tweedie state space models proposed by Jørgensen, Lundbye-Christensen, Song and Sun (1995c) and the application of these models to a multivariate longitudinal data set from the UBC 6-weekly frequent Magnetic Resonance Imaging (MRI) sub-study of the Betaseron clinical trial in relapsing-remitting multiple sclerosis (MS) (Paty, Li, the UBC MS/MRI Study Group and the IFNB Multiple Sclerosis Study Group, 1993). The data will be referred to simply as the Betaseron Data.

Models for analyzing longitudinal data have been developed extensively recently. These models can be classified as marginal, random effects, transition and state space models. The marginal models focus on inferences about the population average. A key feature of these models is the estimation of regression parameters by the generalized estimation equation method. Among these are the models proposed by Liang and Zeger (1986) and Liang, Zeger, and Qaqish (1992). The random effects models consider situations in which the regression coefficients vary among individuals. Some examples are Laird and Ware (1982) and Zeger and Karim (1991). Transition models are particularly useful when the correlation structure is investigated. The models developed by Korn and Whittemore (1979) and Zeger, Liang and Self (1985) are specific examples of transition models for binary data.

The state space models have their origin in aerospace engineering and have been

widely applied in modern systems theory. These kind of models are well developed within the context of the Gaussian distribution. The concept has been extended by several authors to include non-Gaussian distributions. In particular, Zeger (1988) and Harvey and Fernandes (1989) have proposed state space models for Poisson counts. The use of the Kalman smoother in predicting the latent process is a key feature of these models. Details of the models for longitudinal data, as mentioned above, will be discussed in Chapter 2.

The new method proposed by Jørgensen *et al.* (1995c) is based on distributions in the class of Tweedie exponential dispersion models (Jørgensen, 1987). The distributions included in the Tweedie class are the Gaussian, Poisson, compound Poisson and gamma, among others. Therefore, the method can handle a wide variety of data types, including continuous, count, mixed and positive continuous data. The method has been developed for regression analysis of multivariate longitudinal data of mixed types. That is, the response vector can have components (called categories) with distributions in the Tweedie class; these distributions may not all be the same. The categories are assumed to all reflect the same underlying latent process; in particular, an order 1 Markov process is used in the analysis in Chapter 4. The observations are assumed conditionally independent given the latent process, both across categories and over time. This results in serial correlation for each category as well as correlation between the categories because of the shared latent process. A realistic correlation structure of the general sort found in many types of data is hence achieved.

The state space model approach provides an intuitively appealing conceptual framework for longitudinal data analysis and a more practical interpretation of the data-generating mechanism. A new feature offered by the new method is that the time-varying

covariates can enter the model either via the latent process or via the observation model. The covariates that enter via the observation model are termed short-term covariates. They have immediate effects on the categories but such effects are short-lived. The covariates that enter via the latent process are termed long-term covariates. They affect the underlying trend of the observations and have a carry-over effect to the next few data points. Residual analysis may help to determine whether a covariate is long-term or short-term. Detailed discussion of this issue is presented in Chapter 3. Also, the latent process may be interpreted as a measure of the severity of the disease. The response variables, in turn, reflect the symptoms caused by the given severity. This provides a convenient framework for interpretation of the model and the statistical analysis.

This thesis has five chapters. A literature review is presented in Chapter 2. The new model is introduced in Chapter 3. In Chapter 4, we will discuss the analysis of the Betaseron data. The conclusions of the data analysis and the comparisons with the previous analyses of the data are presented in Chapter 5.

Chapter 2

Review of literature

This chapter surveys the current literature on approaches used in analyzing discrete and continuous longitudinal data. We will present the marginal, random effects, transition and state space models. Following Diggle, Liang and Zeger (1994), the marginal, random effects and transition models are extensions of generalized linear models (GLM). Traditional maximum likelihood or maximum conditional likelihood methods are normally employed in estimating the parameters in both the random effects and transition models. In marginal models, the marginal means and covariance structure of the observations are modelled separately. These models do not require specification of the joint likelihood of the data. Liang and Zeger (1986) and Zeger and Liang (1986) proposed the generalized estimating equations (GEE) approach for estimating regression parameters in these models. In the GEE approach, the “working likelihood” is used, instead of the actual likelihood, for generating estimating equations for the regression parameters. The applications of these extensions to GLM are confined to univariate response variables and depend on the objective of the scientific research. To be specific, the marginal models are suitable when the population average is the primary focus, whereas the random effects models are appropriate in situations where subject-specific effects are to be studied. The transition models are particularly useful in investigating the dependence of the response on both the explanatory variables and previous responses.

The key idea underlying state space models is that the longitudinal data (observed)

is related to a latent process (unobserved) by means of assumptions on the conditional distribution of the former given the value of the latter. Moreover, a transition model is assumed for the latent process. One advantage of this approach over the three GLM extensions is its ability to handle a multivariate response. In addition, it gives an appealing intuition regarding the data-generating mechanism.

We will present some key features and ideas of these models in the following sections. For more detailed descriptions, we refer to the original papers. We denote the longitudinal data by Y_{jt} , $j = 1, \dots, h$, $t = 1, \dots, n_j$; the subject is indexed by j and n_j is the number of observations from subject j . This notation will be used throughout the rest of this chapter.

2.1 Marginal models

In a marginal model, the regression of the response on explanatory variables is modelled separately from within-subject correlation. In the regression, the average response over a sub-population (such as patients from a treatment group) is modelled as a function of explanatory variables. Liang and Zeger (1986) and Zeger and Liang (1986) proposed one such approach based upon generalized linear models (GLM). These marginal models for GLMs aim at addressing scientific interest concerning the population average. In such a marginal model for longitudinal data Y_{jt} , the marginal expectation $E(Y_{jt}) = \mu_{jt}$ is assumed to be a function (link) h of the explanatory variables \mathbf{x}_{jt} . The marginal variance $\text{Var}(Y_{jt})$ is related to the marginal mean through a known variance function V and an unknown dispersion parameter σ^2 . Furthermore, the correlation between Y_{jt} and Y_{jk} is a known function ρ of the marginal means and, perhaps, an additional parameter, α .

That is,

- $h(\mu_{jt}) = \mathbf{x}_{jt}^\top \boldsymbol{\beta}$;
- $\text{Var}(Y_{jt}) = \sigma^2 V(\mu_{jt})$;
- $\text{Corr}(Y_{jt}, Y_{jk}) = \rho(\mu_{jt}, \mu_{jk}; \boldsymbol{\alpha})$.

The regression parameter $\boldsymbol{\beta}$ is assumed to be the same across subjects and it represents the overall effects of the explanatory variables on the response. One advantage of the marginal models is their simplicity of interpretation. They serve as natural analogues for correlated data of generalized linear models for independent data. Note that assumptions are made only on the first two moments and the likelihood function is not specified. On the one hand, these make them suitable for a wide range of discrete and continuous data. On the other hand, traditional maximum likelihood methods cannot be used in parameter estimation. The GEE approach was proposed to estimate the regression parameters and it was found to be robust to misspecifications of the correlation structure. For details of the GEE approach, see Liang and Zeger (1986) and Zeger and Liang (1986).

2.2 Random effects models

When the scientific objective concerns the individual subjects and not only the population average, the random effects models are extremely useful. In a random effects model, the regression coefficients can vary across subjects and this variability takes the natural heterogeneity among subjects into account. One way of specifying random effects model is as follows:

- the responses Y_{j1}, \dots, Y_{jn_j} are conditionally independent given the random effects, U_j , $j = 1, \dots, h$; and the conditional densities are of the form

$$f(y_{jt}|U_j) = \exp\left\{\frac{y_{jt}\theta_{jt} - b(\theta_{jt})}{\sigma^2} + c(y_{jt}, \sigma^2)\right\},$$

where b and c are known real functions and θ_{jt} and σ^2 are unknown natural and dispersion parameters respectively;

- the conditional mean and variance satisfy

$$h(E(Y_{jt}|U_j)) = \mathbf{x}_{jt}^\top \boldsymbol{\beta} + \mathbf{d}_{jt}^\top U_j$$

and

$$\text{Var}(Y_{jt}|U_j) = \sigma^2 V(E(Y_{jt}|U_j))$$

where h and V are known link and variance functions respectively, \mathbf{x}_{jt} is the t th row of the design matrix for the fixed population effects, $\boldsymbol{\beta}$ is the vector of unknown regression parameters, \mathbf{d}_{jt} is the t th row of the design matrix for the random individual effects and U_j is the vector of unknown individual effects.

- the U_j 's are mutually independent.

Note that the heterogeneity among subjects is represented by a probability distribution and the serial correlations between observations from the same subjects are assumed to have been caused by their having the same U_j . Moreover, the inclusion of \mathbf{d}_{jt} in the model permits some explanatory variables to have random coefficients while the remaining ones do not. This allows greater flexibility in the applications of these models. Since the likelihood in these models is explicitly stated, regression parameters are usually estimated by maximum likelihood or restricted maximum likelihood. For more details, see Laird and Ware (1982) and Zeger and Karim (1991).

2.3 Transition models

The transition models are very useful when the dependence of the response on both the explanatory variables and the past are investigated. In a transition model, the conditional distribution of Y_{jt} is assumed to depend on both the explanatory variables and the past responses, $H_{jt} = \{y_{jk}, k = 1, \dots, t-1\}$. A transition model is specified by:

- the conditional distribution of Y_{jt} is assumed to be of the form

$$f(y_{jt}|H_{jt}) = \exp\left\{\frac{y_{jt}\theta_{jt} - b(\theta_{jt})}{\sigma^2} + c(y_{jt}, \sigma^2)\right\},$$

where b and c are known real functions and θ_{jt} and σ^2 are unknown natural and dispersion parameters respectively;

- the conditional mean and variance satisfy

$$h(E(Y_{jt}|H_{jt})) = \mathbf{x}_{jt}^T \boldsymbol{\beta} + \psi_{jt}(H_{jt}; \boldsymbol{\beta}, \boldsymbol{\alpha})$$

and

$$\text{Var}(Y_{jt}|H_{jt}) = \sigma^2 V(E(Y_{jt}|H_{jt}))$$

where ψ_{jt} is a known function, $\boldsymbol{\beta}$ is the regression parameter, $\boldsymbol{\alpha}$ is the additional parameter relating Y_{jt} to H_{jt} , and h and V are known link and variance functions respectively.

A Markov chain assumption for the series of responses is commonly used. For example, in a order 1 Markov chain, the response Y_{jt} is assumed to depend only on the immediate previous response, $Y_{j,t-1}$. Transition models with a Markov assumption are routinely used to analyze binary, categorical and count data. Note that the regression parameter $\boldsymbol{\beta}$ is assumed to be the same across subjects. It leaves the natural heterogeneity of subjects to be explained by the past responses, H_{jt} . A shortcoming of the

transition models is the possible confounding of the effects of the explanatory variables with those of the past responses. For specific examples, see Zeger and Qaqish (1988) and Kaufmann (1987).

2.4 State space models

The state space models for longitudinal data proposed in the literature are diverse both in approach and in method of estimation. In these models, the response is related to the latent process (unobserved) and assumptions are made on the conditional distribution of the response given the latent process. The Kalman filter and smoother are commonly used to predict the latent process, see Durbin (1990), Fahrmeir and Kaufmann (1991), Jones (1993) and Fahrmeir and Tutz (1994). Models for Gaussian data are well developed and the idea has recently been extended to include non-Gaussian data. For example, models for univariate Poisson counts have been proposed by Azzalini (1982), Zeger (1988), Harvey and Fernandes (1989) and Chan and Ledolter (1995). Jørgensen, Lundbye-Christensen, Song and Sun (1995a) developed models for multivariate count data.

Most state space models do not explicitly specify the distribution for the latent process but rather its first and second moments only. The model proposed by West, Harrison and Migon (1985) is one such model. In the model developed by Zeger (1988) for Poisson counts, only second moment assumptions were made for the latent process. In these cases where the probability model is not fully specified, the quasi-likelihood method is used in estimating the regression parameters. One shortcoming of the quasi-likelihood approach is that it is difficult to interpret the results in the absence of full distributional

assumptions. Jørgensen *et al.* (1995a) overcame this by explicitly modelling both the latent process and the data; these models are further developed into our current approach to handle various types of multivariate longitudinal data.

Chapter 3

Tweedie state space models

These models are generalizations of the Poisson-gamma models considered by Jørgensen *et al.* (1995a) in which the observations are multivariate Poisson counts and the latent process is assumed to follow the gamma distribution. The new method is based on the class of Tweedie exponential dispersion models and can be used for regression analysis of multivariate longitudinal data where the components (called categories) of the response vector can be of different data types: discrete, continuous and mixed. For more details, see Jørgensen *et al.* (1995c).

3.1 The model

For the case where a d -dimensional vector of observations is recorded at equally spaced times t for each of the h subjects, we have h independent series of observations indexed by $j = 1, \dots, h$; the j th series has length n_j . We let Y_{ijt} denote the observation for category i , series j and time t ; $i = 1, \dots, d$, $j = 1, \dots, h$ and $t = 1, \dots, n_j$. The vector of observations and the value of the latent process at time t for series j are denoted by \mathbf{Y}_{jt} and θ_{jt} respectively. Let θ_{j0} denote the initial value of the latent process for series j . The variation of these initial values, θ_{j0} s, is assumed to explain the natural heterogeneity across subjects. To summarize, the model consists of three parts: the observation model, the latent process and the initial values.

- The observation model:

The categories of \mathbf{Y}_{jt} are assumed to be conditionally independent given θ_{jt} , and the conditional distribution for the i th category of \mathbf{Y}_{jt} is assumed to follow a Tweedie exponential dispersion model,

$$Y_{ijt}|\theta_{jt} \sim Tw_{p_i} \left(a_{ijt}\theta_{jt}, \frac{\nu_i^2}{\theta_{jt}^{p_i-1}} \right), \quad (3.1)$$

with

$$a_{ijt} = \exp(\mathbf{x}_{jt}^\top \boldsymbol{\alpha}_{ij}), \quad (3.2)$$

where $\mathbf{x}_{jt} \in \mathbf{R}^k$ are the time-varying short-term covariates modifying the conditional mean of the observations given the latent process θ_{jt} , $\boldsymbol{\alpha}_{ij} \in \mathbf{R}^k$ are the regression parameters which can vary across categories, p_i is the known shape parameter which determines the distributional form of the category and ν_i^2 is the dispersion parameter which can be different even if all the p_i s are the same. We consider the case where the dispersion parameter depends on the category i only but not on patient j . However, the model allows the dispersion parameter to be different across categories and across patients. The notation $Tw_{p_i}(\mu, \delta^2)$ denotes the Tweedie model with known shape parameter p_i , $p_i \geq 1$ or $p_i \leq 0$. The case $p_i = 1$ and $\nu_i^2 = 1$ gives the Poisson distribution and the case $p_i = 2$ corresponds to the gamma distribution. Note that the dispersion parameter for the Poisson distribution is equal to 1. A summary of these exponential dispersion models is given in Appendix A.

The conditional expectation and variance are given

$$E(Y_{ijt}|\theta_{jt}) = a_{ijt}\theta_{jt} \text{ and } \text{Var}(Y_{ijt}|\theta_{jt}) = \nu_i^2 a_{ijt}^{p_i} \theta_{jt},$$

respectively.

For example, in the Betaseron data which we will analyze in Chapter 4, the latent process θ_{jt} may be interpreted as the underlying severity of multiple sclerosis and the observations Y_{ijt} may be interpreted as the symptoms reflecting this underlying severity. We assume that the conditional distribution for the first and second categories are Poisson and gamma distributions, respectively. Moreover, the dispersion parameter for category i is assumed to be the same across patients.

- The latent process:

The latent process for each series is assumed to be an order 1 Markov chain with transition distribution of the form

$$\theta_{jt}|\theta_{jt-1} \sim Tw_q \left(b_{jt}\theta_{jt-1}, \frac{\sigma^2}{\theta_{jt-1}^{q-1}} \right) \quad (3.3)$$

with

$$b_{jt} = \exp(\Delta \mathbf{z}_{jt}^\top \boldsymbol{\beta}_j), \quad (3.4)$$

where $\mathbf{z}_{jt} \in \mathbf{R}^l$ are the time-varying long-term covariates with increments $\Delta \mathbf{z}_t = \mathbf{z}_t - \mathbf{z}_{t-1}$, $\boldsymbol{\beta}_j \in \mathbf{R}^l$ are the regression parameters and σ^2 is the dispersion parameter. Moreover, we assume $\mathbf{z}_0 = 0$.

The conditional expectation and variance are, respectively,

$$E(\theta_{jt}|\theta_{jt-1}) = b_{jt}\theta_{jt-1} \text{ and } \text{Var}(\theta_{jt}|\theta_{jt-1}) = \sigma^2 b_{jt}^q \theta_{jt-1}$$

- The initial values:

The marginal distribution of the initial values is assumed to be

$$\theta_{j0} \sim Tw_r(g_j, \omega^2), \quad (3.5)$$

with

$$g_j = \exp(\mathbf{w}_j^\top \boldsymbol{\gamma}) = E(\theta_{j0}), \quad (3.6)$$

where $\mathbf{w}_j \in \mathbf{R}^m$ are the constant baseline covariates which reflect systematic variation between series, $\boldsymbol{\gamma} \in \mathbf{R}^m$ are the regression parameters and ω^2 is the dispersion parameter reflecting the random variation between series.

For example, the treatment groups and the baseline covariates present in the Betaseron study are treated as constant baseline covariates.

The correlations between the categories Y_{ijt} , $i = 1, \dots, d$, are assumed to be caused by their shared θ_{jt} . Short-term covariates have a short-lived effects on the observations whereas the effects of the long-term covariates can be carried over to the next few data points. The marginal expectations and variances in model (3.1)–(3.6) are now presented. For the initial values, the variance of θ_{j0} is

$$\text{Var}(\theta_{j0}) = \omega^2 g_j^r$$

For the latent process, the marginal expectation of θ_{jt} is

$$E(\theta_{jt}) = \tau_{jt} = g_j b_{j1} \cdots b_{jt} = \exp(\mathbf{z}_{jt}^\top \boldsymbol{\beta}_j + \mathbf{w}_j^\top \boldsymbol{\gamma}). \quad (3.7)$$

and the marginal variance of θ_{jt} is

$$\text{Var}(\theta_{jt}) = \sigma^2 \phi_{jt} \tau_{jt} + \omega^2 \tau_{jt}^2,$$

where

$$\phi_{jt} = b_{jt}^{q-1} + b_{jt} b_{jt-1}^{q-1} + \cdots + b_{jt} b_{jt-1} \cdots b_{j1}^{q-1}$$

For example, in the Betaseron study, time is the only long-term covariate we will consider. Since the latent process is log-linear in the covariates as seen in (3.7), the time effect may be interpreted as the rate of change of the log of the underlying severity of the disease. A positive estimate for the time effect suggests that the disease is increasing in severity.

For the observation model, the marginal expectations of the observations are

$$E(Y_{ijt}) = \exp(\mathbf{x}_{jt}^\top \boldsymbol{\alpha}_{ij} + \mathbf{z}_{jt}^\top \boldsymbol{\beta}_j + \mathbf{w}_j^\top \boldsymbol{\gamma}).$$

Note that the observation and latent process means are log-linear in the covariates and so are the initial values. Furthermore, the marginal variance of \mathbf{Y}_{jt} has the form

$$\text{Var}(\mathbf{Y}_{jt}) = \boldsymbol{\Lambda}_{jt} \tau_{jt} + \mathbf{a}_{jt} \mathbf{a}_{jt}^\top \text{Var}(\theta_{jt})$$

where $\boldsymbol{\Lambda}_{jt} = \text{diag}(\nu_1^2 a_{1jt}^{p_1}, \dots, \nu_d^2 a_{djt}^{p_d})$ and $\mathbf{a}_{jt} = (a_{1jt}, \dots, a_{djt})^\top$. The above expression shows that, at time t , the categories are positively correlated since the second term is always positive. Finally, for observation vectors separated by a lag of $s > 0$, the covariance of \mathbf{Y}_{jt} and \mathbf{Y}_{jt+s} is given by

$$\text{Cov}(\mathbf{Y}_{jt}, \mathbf{Y}_{jt+s}) = \mathbf{a}_{jt} \mathbf{a}_{jt+s}^\top b_{jt+1} \cdots b_{jt+s} \text{Var}(\theta_{jt}).$$

3.2 The Kalman filter and smoother

For each series j , the Kalman filter predicts a future observation \mathbf{Y}_{jt+1} and the current value of the latent process θ_{jt} based on past and current observations $\mathbf{Y}_{j1}, \dots, \mathbf{Y}_{jt}$ whereas the smoother predicts the latent process using all the observations in the series, $\mathbf{Y}_{j1}, \dots, \mathbf{Y}_{jn_j}$. These filtered and smoothed values are used to estimate, or more precisely, predict the values of θ_{jt} in model (3.1)–(3.6). From now on, we consider filtering and smoothing for one series at a time. The index j is suppressed for notational convenience.

We define the *innovations* for the latent process by

$$\xi_t = \theta_t - E(\theta_t | \boldsymbol{\theta}^{t-1}) = \theta_t - b_t \theta_{t-1}$$

and for the observation model by

$$\varepsilon_t = \mathbf{Y}_t - E(\mathbf{Y}_t | \boldsymbol{\theta}^t) = \mathbf{Y}_t - \mathbf{a}_t \theta_t,$$

where $\boldsymbol{\theta}^t = (\theta_0, \dots, \theta_t)^\top$, $t = 1, \dots, n$. The model can be expressed in the additive form as follows,

$$\mathbf{Y}_t = \mathbf{a}_t \boldsymbol{\theta}_t + \boldsymbol{\varepsilon}_t, \quad \theta_t = b_t \theta_{t-1} + \xi_t,$$

where $\boldsymbol{\varepsilon}_t$ and θ_t are uncorrelated, and so are ξ_t and θ_{t-1} . These $2n_j$ innovations are uncorrelated. The model has the same correlation structure as that of an ordinary Gaussian state space model. The variances of the innovations are given by

$$\text{Var}(\xi_t) = \text{E}(\xi_t^2) = \sigma^2 b_t^{q-1} \tau_t$$

and

$$\text{Var}(\boldsymbol{\varepsilon}_t) = \text{E}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top) = \boldsymbol{\Lambda}_t \tau_t.$$

The forward recursive formula for the filtered values, m_t , and their corresponding mean square errors, C_t , for the latent process are given by, for $t = 1, \dots, n$,

$$m_t = b_t \{m_{t-1} + D_t \mathbf{a}_t^\top \mathbf{Q}_t^{-1} (\mathbf{Y}_t - \mathbf{f}_t)\} \quad (3.8)$$

and

$$C_t = b_t D_t (1 - b_t D_t \mathbf{a}_t^\top \mathbf{Q}_t^{-1} \mathbf{a}_t). \quad (3.9)$$

where

$$D_t = b_t C_{t-1} + \sigma^2 b_t^{q-1} \tau_{t-1},$$

$$\mathbf{f}_t = \mathbf{a}_t b_t m_{t-1} \quad \text{and} \quad \mathbf{Q}_t = b_t D_t \mathbf{a}_t \mathbf{a}_t^\top + \boldsymbol{\Lambda}_t \tau_t.$$

The recursion starts with $m_0 = g$ and $C_0 = \omega^2 g^r$.

When all the n observations are known, the smoother values, m_t^* , and their corresponding mean square errors, C_t^* , for the latent process can be obtained by the following backward recursion for $t = n-1, \dots, 1, 0$; starting with $m_n^* = m_n$ and $C_n^* = C_n$:

$$m_t^* = m_t + \frac{C_t}{D_{t+1}} (m_{t+1}^* - b_{t+1} m_t) \quad (3.10)$$

and

$$C_t^* = \sigma^2 \frac{C_t}{D_{t+1}} b_{t+1}^{q-2} + \frac{C_t^2}{D_{t+1}^2} C_{t+1}^*. \quad (3.11)$$

For detailed arguments, see Harvey (1981, Chapter 4), or Jørgensen *et al.* (1995a).

3.3 Parameter estimation

Let the full data vector of the $N = n_1 + \dots + n_h$ observations \mathbf{Y}_{jt} be

$$\mathbf{Y} = (\mathbf{Y}_{11}^\top, \dots, \mathbf{Y}_{1n_1}^\top, \dots, \mathbf{Y}_{h1}^\top, \dots, \mathbf{Y}_{hn_h}^\top)^\top.$$

and the vector of the smoother values be

$$\mathbf{m}^* = (m_{10}^*, \dots, m_{1n_1}^*, \dots, m_{h0}^*, \dots, m_{hn_h}^*)^\top.$$

The two regression parameter vectors are defined by

$$\boldsymbol{\alpha} = (\boldsymbol{\alpha}_{11}^\top, \dots, \boldsymbol{\alpha}_{d1}^\top, \dots, \boldsymbol{\alpha}_{1h}^\top, \dots, \boldsymbol{\alpha}_{dh}^\top)^\top, \quad \boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \dots, \boldsymbol{\beta}_h^\top)^\top.$$

The *Kalman estimating equation* method is used to estimate the regression parameters $\boldsymbol{\eta} = (\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$. The dispersion parameters ν_i^2 , σ^2 and ω^2 are estimated by the *adjusted Pearson estimates*.

3.3.1 The Kalman estimating equation

The *Kalman estimating equation* method corresponds to replacing the E-step of the EM algorithm by the Kalman smoother; but instead of the EM algorithm, the *Newton scoring algorithm* is used which will be defined later. When the values of the dispersion parameters ν_i^2 , σ^2 and ω^2 are known, the regression parameter $\boldsymbol{\eta}$ is estimated by $\hat{\boldsymbol{\eta}}$, the solution to the *Kalman estimating equation* $\boldsymbol{\psi}(\boldsymbol{\eta}) = 0$ where the vector $\boldsymbol{\psi}$ has components

$$\boldsymbol{\psi}_1(\boldsymbol{\eta}) = \mathbf{X}^\top \mathbf{K}_a^{-1} (\mathbf{Y} - \mathbf{A} \mathbf{m}^*), \quad (3.12)$$

$$\psi_2(\boldsymbol{\eta}) = \boldsymbol{\Delta Z}^\top \mathbf{K}_b^{-1} \mathbf{B} \mathbf{m}^*, \quad (3.13)$$

and

$$\psi_3(\boldsymbol{\eta}) = \mathbf{W}^\top \mathbf{K}_g^{-1} (\mathbf{m}_0^* - \mathbf{g}), \quad (3.14)$$

where \mathbf{A} and \mathbf{B} are suitably defined matrices such that the elements of $\mathbf{Y} - \mathbf{A} \mathbf{m}^*$ are $\mathbf{Y}_t - \mathbf{a}_t m_t^*$ and those of $\mathbf{B} \mathbf{m}^*$ are $m_t^* - b_t m_{t-1}^*$; the matrices \mathbf{X} , $\boldsymbol{\Delta Z}$ and \mathbf{W} are the design matrices of the generalized linear models (3.1–2), (3.3–4) and (3.5–6), respectively;

$$\mathbf{K}_a = \text{diag}(\nu_1^2 a_{111}^{p_1-1}, \dots, \nu_d^2 a_{dhn_h}^{p_d-1}),$$

$$\mathbf{K}_b = \text{diag}(b_{11}^{q-1}, \dots, b_{1n_1}^{q-1}, \dots, b_{h1}^{q-1}, \dots, b_{hn_h}^{q-1}),$$

$$\mathbf{K}_g = \text{diag}(g_1^{r-1}, \dots, g_h^{r-1}),$$

$$\mathbf{g} = (g_1, \dots, g_h)^\top, \text{ and } \mathbf{m}_0^* = (m_{10}^*, \dots, m_{h0}^*)^\top.$$

The estimating equation $\psi(\boldsymbol{\eta}) = 0$ is unbiased. The estimator $\hat{\boldsymbol{\eta}}$ is asymptotically normal and its standard errors can be obtained from the inverse of the *Godambe information matrix*,

$$\mathbf{J}(\boldsymbol{\eta}) = \mathbf{S}^\top(\boldsymbol{\eta}) \mathbf{V}^{-1}(\boldsymbol{\eta}) \mathbf{S}(\boldsymbol{\eta}),$$

where $\mathbf{S}(\boldsymbol{\eta})$ is the *sensitivity matrix*,

$$\mathbf{S}(\boldsymbol{\eta}) = \mathbf{E}_{\boldsymbol{\eta}} \left(\frac{\partial \psi}{\partial \boldsymbol{\eta}^\top} \right),$$

and

$$\mathbf{V}(\boldsymbol{\eta}) = \mathbf{E}_{\boldsymbol{\eta}} [\psi(\boldsymbol{\eta}) \psi^\top(\boldsymbol{\eta})]$$

is the *variability matrix*. The matrices $\mathbf{S}(\boldsymbol{\eta})$ and $\mathbf{V}(\boldsymbol{\eta})$ are calculated in Appendix B.

The joint significance of a subset of regression parameters is tested using the statistic W defined by

$$W = \hat{\boldsymbol{\eta}}_1^\top [\mathbf{J}^{11}(\hat{\boldsymbol{\eta}})]^{-1} \hat{\boldsymbol{\eta}}_1,$$

where $\boldsymbol{\eta}_1$ is the subvector whose significance we want to test and $\mathbf{J}^{11}(\hat{\boldsymbol{\eta}})$ is the corresponding block of the asymptotic covariance matrix of $\hat{\boldsymbol{\eta}}$. The test statistic W is an analogue of Wald's statistic and follows a $\chi^2(u)$ -distribution asymptotically, where u is the dimensionality of the subvector $\boldsymbol{\eta}_1$.

The *Newton scoring algorithm*, a parallel to Fisher's scoring method, is defined as the Newton algorithm applied to the equation $\boldsymbol{\psi}(\boldsymbol{\eta}) = 0$, but with the derivative of $\boldsymbol{\psi}$ replaced by its expectation $\mathbf{S}(\boldsymbol{\eta})$. This new algorithm gives the updated value $\boldsymbol{\eta}^*$ for $\boldsymbol{\eta}$ according to

$$\boldsymbol{\eta}^* = \boldsymbol{\eta} - \mathbf{S}^{-1}(\boldsymbol{\eta})\boldsymbol{\psi}(\boldsymbol{\eta}). \quad (3.15)$$

An advantage of this algorithm is that the calculation of $\mathbf{S}(\boldsymbol{\eta})$ is done recursively in parallel with the calculation of the Kalman smoother, cf. Appendix B.

The initial parameter estimates are obtained from the data by fitting the generalized linear models (3.1)–(3.6) with the θ_{jt} s replaced by their initial estimates which can be found by averaging the d components of \mathbf{Y}_{jt} . If $d = 1$, some smoothing is required, for example, using a moving average of the series. The algorithm is then started by calculating the smoothers and so on.

When the dispersion parameters are unknown, they are estimated by the unbiased estimating equations (3.16), (3.17) and (3.18) in Section 3.3.2. These three equations together with the equation $\boldsymbol{\psi}(\boldsymbol{\eta}) = 0$ form a set of unbiased estimating equations. Hence, the estimator obtained by solving this set of equations is asymptotically unbiased. The *Newton scoring algorithm* obtains updated values for the regression parameters $\boldsymbol{\eta}$ and the dispersion parameters σ^2 , ν_i^2 and ω^2 in two stages. First, it updates $\boldsymbol{\eta}$ via (3.15).

Second, it updates σ^2 , ν_i^2 and ω^2 via (3.16), (3.17) and (3.18). This extended algorithm is also referred as *Newton scoring algorithm*. The correct asymptotic standard errors for the estimator $\hat{\boldsymbol{\eta}}$ could be calculated from the *Godambe information matrix* for the full set of estimating equations, but this involves third and fourth moments of the Kalman smoother, which we have not calculated. Instead, we calculate the asymptotic standard errors from the *Godambe information matrix* for the estimating equation $\boldsymbol{\psi}(\boldsymbol{\eta}) = 0$; this method is outlined earlier in this section. If the vector $\boldsymbol{\psi}$ depends little on ν_i^2 , σ^2 and ω^2 , this method will give approximately correct asymptotic standard errors. Further investigation on how the dependence of the vector $\boldsymbol{\psi}$ on ν_i^2 , σ^2 and ω^2 affects the asymptotic standard errors for the estimator is still pending.

3.3.2 Estimation of dispersion parameters

Due to the substitution of the smoother value m_{jt}^* for the θ_{jt} , the simple Pearson estimates for the dispersion parameters are biased downward. Following Jørgensen *et al.* (1995a), the *adjusted Pearson estimates* for the dispersion parameters σ^2 , ν_i^2 and ω^2 are used instead:

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{j=1}^h \sum_{t=1}^{n_j} \frac{\hat{\xi}_{jt}^{*2}}{b_{jt}^{q-1} \tau_{jt}} + \frac{1}{N} \sum_{j=1}^h \sum_{t=1}^{n_j} \frac{C_{jt}^* + b_{jt}^2 C_{jt-1}^* - 2b_{jt} C_{jt-1} C_{jt}^* / D_{jt}}{b_{jt}^{q-1} \tau_{jt}}, \quad (3.16)$$

$$\hat{\nu}_i^2 = \frac{1}{N} \sum_{j=1}^h \sum_{t=1}^{n_j} \frac{\hat{\varepsilon}_{ijt}^{*2}}{\tau_{jt} a_{ijt}^{p_i}} + \frac{1}{N} \sum_{j=1}^h \sum_{t=1}^{n_j} \frac{a_{ijt}^2 C_{jt}^*}{\tau_{jt} a_{ijt}^{p_i}} \quad (3.17)$$

and

$$\hat{\omega}^2 = \frac{1}{h} \sum_{j=1}^h \frac{(m_{j0}^* - g_j)^2}{g_j^r} + \frac{1}{h} \sum_{j=1}^h \frac{C_{j0}^*}{g_j^r}. \quad (3.18)$$

The equations for $\hat{\sigma}^2$, $\hat{\nu}_i^2$ and $\hat{\omega}^2$ are unbiased. The second term in (3.16) corrects the bias of the simple Pearson estimate and so do those in (3.17) and (3.18). In most cases, the dispersion parameters are unknown. Hence, the *Newton scoring algorithm* updates

$\boldsymbol{\eta}$ via (3.15) and then σ^2 , ν_i^2 and ω^2 via (3.16), (3.17) and (3.18), respectively.

3.4 Residual analysis

The main model assumptions for both the observed and unobserved parts of the model are checked by plots of the standardized residuals (that is, residuals divided by their standard deviations). Techniques commonly used in the generalized linear models are utilized to check the assumptions on the distribution as well as the regression of the model. The assumptions on the correlation structure of the data are examined by methods available from time series analysis; in particular, the autocorrelation function (ACF) and the partial autocorrelation function (PACF). When we check the assumptions (such as about the correlation structure) on the observations from a patient, we consider only the residuals from a given series. However, for the distributional assumptions of the observation model and the latent process, they are checked by plots of the residuals from all the h series.

The two main types of residuals considered are the filter and smoother residuals. First, the filter residuals are defined as the predictor errors based on the Kalman filter and are given by

$$\hat{\boldsymbol{\epsilon}}_{jt} = \mathbf{Y}_{jt} - \mathbf{f}_{jt} \quad \text{and} \quad \hat{\xi}_{jt} = m_{jt} - b_{jt}m_{jt-1}.$$

Their variances are given by

$$\text{Var}(\hat{\boldsymbol{\epsilon}}_{jt}) = \mathbf{Q}_{jt} \quad \text{and} \quad \text{Var}(\hat{\xi}_{jt}) = b_{jt}D_{jt} - C_{jt}.$$

The $\hat{\boldsymbol{\epsilon}}_{jt}$ s are mutually uncorrelated over time and so are the $\hat{\xi}_{jt}$ s. Moreover, $\hat{\xi}_{jt}$ is uncorrelated with f_{ijt} and $\hat{\xi}_{jt}$ is uncorrelated with $b_{jt}m_{jt-1}$. These properties make the $\hat{\boldsymbol{\epsilon}}_{jt}$ s

and $\hat{\xi}_{jt}$ s very useful in the residual analysis. In particular, they are useful in checking the correlation structure of the model. For example, the ACF plot of the filter residuals (from a given series) showing significant serial correlation may suggest a violation of the assumption that the observations are conditional independent across time. However, these residuals may have relatively large variances for the first few observations of each series.

Second, the smoother residuals are given by

$$\hat{\epsilon}_{jt}^* = Y_{jt} - \mathbf{a}_{jt}m_{jt}^* \quad \text{and} \quad \hat{\xi}_{jt}^* = m_{jt}^* - b_{jt}m_{jt-1}^*.$$

Their corresponding variances are

$$\text{Var}(\hat{\epsilon}_{jt}^*) = \mathbf{\Lambda}_{jt}\tau_{jt} - \mathbf{a}_{jt}\mathbf{a}_{jt}^\top C_{jt}^*$$

and

$$\text{Var}(\hat{\xi}_{jt}^*) = \sigma^2 b_{jt}^{q-1} \tau_{jt} - C_{jt}^* - b_{jt}^2 C_{jt-1}^* + 2b_{jt} \frac{C_{jt-1}}{D_{jt}} C_{jt}^*.$$

Contrary to the filter residuals, the smoother residuals are correlated over time. In general, they have smaller variances than the corresponding filter residuals do.

All these residuals have mean 0 but different variances. Note that the properties mentioned in this section do not take into account the effects of substituting estimates for the regression parameters.

Standardized filter and smoother residuals (having unit variance) from the latent process are denoted by r_{jt} and r_{jt}^* respectively. The residuals $\hat{\epsilon}_{jt}$ are standardized componentwise. For each category, we use R_{ijt} and R_{ijt}^* to denote the standardized filter and

smoother residuals, respectively.

For the assumptions on the observation model, we may plot R_{ijt} against $R_{ij,t-1}$ to check the conditional independence of the observations over time. For each category i of the j th series, the plots of the ACF and PACF of R_{ijt} are useful to see whether the assumed correlation structure is appropriate. We can check the distributional assumptions (to be specific, the form of the variance functions) by plotting the standardized filter residuals against the log fitted values, $\log(\mathbf{a}_{ijt}m_{jt}^*)$. Any pattern in this plot would indicate that the variance function chosen is incorrect; for example, a “megaphone” shape would indicate that the variance of the residuals increases with the log fitted values. Due to the fact that the observation means are log-linear in the covariates in the model, the log link assumption can be checked by plotting $\log Y_{ijt}$ against $\log(\mathbf{a}_{ijt}m_{jt}^*)$. If the log link assumption is incorrect, this plot should show a curvature or an unusual shape. The linearity of the regression model can be checked by plotting the smoother residuals R_{ijt}^* against each covariate. The empirical variance-covariance matrix, H , of the vector of the standardized filter residuals given by

$$H = \frac{1}{n_j} \sum_{t=1}^{n_j} \mathbf{Q}_{jt}^{-1/2} \hat{\boldsymbol{\varepsilon}}_{jt} \hat{\boldsymbol{\varepsilon}}_{jt}^{\top} \mathbf{Q}_{jt}^{-1/2}$$

is used to check the assumption that the categories are conditionally independent given the latent process. Since the expectation of H is the $d \times d$ identity matrix, the observed H should not be very different from the identity matrix. One criterion can be used is that the off-diagonal elements should not exceed their respective asymptotic standard errors. Note that the off-diagonal elements have asymptotic standard errors $n_j^{-1/2}$ and the standardization depends on the version of the square-root matrix $\mathbf{Q}_{jt}^{1/2}$ chosen.

Regarding the latent process, we can check the Markov assumption by plotting r_{jt}

against r_{jt-1} . The ACF and PACF plots of r_{jt} are also useful for checking the assumptions on the correlation structure of the latent process. The distributional assumptions can be checked by plotting r_{jt} against the log fitted values, $\log(b_{jt}m_{jt}^*)$. This plot can indicate whether the variance of the residuals depends on the log fitted values. No particular pattern in the plot should be observed; this suggests no evidence against the chosen variance function. Non-linearity of the model can be detected by plotting r_{jt}^* against each covariate. Moreover, we can check the log link assumption by plotting $\log(m_{jt}^*)$ against $\log(b_{jt}m_{jt-1}^*)$. A pattern in this plot may suggest that the log link assumption is inappropriate.

We now consider the assumptions on the initial values. Similarly, we can plot the standardized smoother residuals $m_{j0}^* - g_j$ against the log fitted values $\log(g_j)$ to check the form of the random effects distribution. Plots of the standardized smoother residuals against each covariate are useful for detecting nonlinearity of the model. Furthermore, the log link assumption can be checked by plotting $\log(m_{j0}^*)$ against $\log(g_j)$.

An additional feature this model offers is that residual plots can help to determine whether a covariate is long-term or short-term. The basic idea is that a short-term covariate would show an association with the observation residuals (R_{ijt} or R_{ijt}^*) if it had been incorrectly fitted as a long-term covariate. Similarly, a long-term covariate would show association with the latent process residuals (r_{jt} or r_{jt}^*) if it had been incorrectly fitted as a short-term covariate.

Chapter 4

Analysis of Betaseron data

4.1 Background

4.1.1 Description of the experiment

A multicenter, randomized, double-blind, placebo-controlled trial of interferon beta-1b in 372 ambulatory patients with relapsing-remitting multiple sclerosis (MS) established that interferon beta-1b has substantially altered the natural history of MS (The IFNB Multiple Sclerosis Study Group, 1993). The interferon beta-1b was manufactured by Chiron Corporation, Emeryville, CA, and supplied to doctors as Betaseron by Berlex Laboratories, Richmond, CA. As a sub-study of this trial, a cohort of 52 patients at the University of British Columbia also had magnetic resonance imagings (MRIs) every 6 weeks during the initial two years of the trial (Paty, Li, the UBC MS/MRI Study Group and the IFNB Multiple Sclerosis Study Group, 1993). The main purpose of this frequent MRI sub-study was to learn more about the nature of MS lesions.

We consider data from this UBC 6-weekly frequent MRI sub-study of the Betaseron clinical trial in relapsing-remitting multiple sclerosis (MS). The data is hereafter referred to as the Betaseron data. In this sub-study, the values of the response variables as well as the time-varying covariates for each patient were scheduled to be observed at baseline (the first day of treatment) and every 6 weeks subsequently in the initial two years of the trial. Although patients were scheduled to have the MRIs (responses were observed based

on these MRIs) every 6 weeks, the actual times when these MRIs were taken were somewhat different from this schedule. Due to a variety of reasons, several patients missed one or more of their scheduled MRI scans. Two of the total of 52 patients dropped out very early and they contributed a very small amount of data to this MRI sub-study. These patients were withheld from the analysis in this thesis. All analyses we report here are based on a total of only 50 patients. Moreover, several patients missed one or two isolated MRI scans. These missed scans are denoted by NA's. The only time-varying covariate we investigate in this thesis is time itself. Each patient was scheduled to take 17 MRI scans in addition to the baseline scan. These scans were critically interpreted and the observed values of the responses were based on these interpretations. The time was arbitrarily set to 0 at baseline and increased by 1 for each subsequent scan. That is, the time was equal to 1 for the first subsequent scan, 2 for the second subsequent scan and so on. Several baseline covariates were also present. For more detailed descriptions of this study, we refer to Paty, Li, the UBC MS/MRI Study Group and the IFNB Multiple Sclerosis Study Group (1993); see also Petkau and White (1995).

The main objective of this thesis is to extensively illustrate the utilization of the Tweedie state space models described in Chapter 3 to analyze multivariate longitudinal data. We also aim at extracting additional information from the data using these new models. In this thesis, the relationship between Betaseron and two of the response variables is considered simultaneously and the patterns of these response variables over time will be investigated utilizing the Tweedie state space models.

4.1.2 Description of the data

The two response variables that we considered were *active lesions* (lesions) and *burden of disease* (burden). For each patient, each subsequent scan was compared to the immediate previous one to determine the number of lesions that were found to be new, recurrent or enlarging. The value of *active lesions* is the sum of the number of new, recurrent and enlarging lesions at each scan. For each subsequent scan, the areas of MS lesions on all slices of the MRI scans were summed up. The total area (in units of mm^2) is defined as the *burden of disease*. Therefore, the response which we studied is a bivariate vector with 2 components (called categories). The first category is *active lesions* which is a count variable whereas the second category is *burden of disease* which is a continuous variable.

The data consist of 50 time series, one for each patient, for each response vector. As mentioned in section 4.1.1, there were deviations from the target dates of this 6-weekly schedule. The practical difficulties in maintaining this schedule is understandable. The most notable of these deviations was for the final scan which was delayed by a period of up to about two weeks for most patients. Three patients had even longer delays to their final scans (approximately 3, 4 and 5 weeks respectively) and two patients had final scans approximately 2 weeks before the target data according to the 6-weekly schedule. Most other deviations from the 6-weekly schedule were minor. Due to methodological constraints, in the analyses presented here, we consider each scan to be taken at the target date for which it was intended, irrespective of the deviation of the actual date of the scan from that target date.

Due to early drop-outs, the length of these time series varies across patients. For example, one patient in the placebo group dropped out after time 14; the length of this

time series is 14 versus the targeted length of 17. These patients with early drop-outs do not complicate the analysis since the new method allows different lengths across patients. However, the isolated missing scans give rise to some difficulties. There are two possible approaches to handle them. Either we can withhold these NA's from the analyses or we can replace them with some imputed data (for example, the smoothed values of the latent process). For convenience, the former approach is used in the analyses to follow. There are all together 5 of these isolated missing scans. This number is relatively small as compared to the total number of scans taken in the sub-study. It is worth noting that when an isolated missing scan is withheld, the two scans originally adjacent to this missing scan will be treated as consecutive scans. For example, one patient in the placebo group missed the scan at time 7. When the NA at time 7 is withheld, the scans at time 6 and 8 will be treated as consecutive scans and the length of the time series becomes 16.

The treatment has three levels: placebo, low dose and high dose. Patients were randomized to either the placebo, the low dose or the high dose group. The baseline covariates present are:

- *age* (in years),
- *duration of disease* (in years),
- *initial EDSS score*,
- *origin* (B.C. or Washington State),
- *gender* (Male or Female),
- *initial burden of disease* (in mm^2).

Since the values of the *initial burden of disease* vary from 14 to 12300 and because of the log-linear nature of the baseline covariates effects for our model, the log of the *initial burden of disease* will be used as a covariate in the analysis to follow. For convenience, *duration of disease*, *initial EDSS score* and the log of the *initial burden of disease* hereafter will be referred to simply as *duration*, *EDSS* and *log area* respectively. *Time*, the only time-varying covariate present, will be included in the model as long-term effect.

The Betaseron data were previously analysed by Petkau and White (1995) who used the generalized estimating equation (GEE) approach developed by Liang and Zeger (1986) and investigated one variable at a time. Among the variables that they analysed was the *burden of disease*. D'yachkova (1997) analysed the whole data set from the Betaseron clinical trial. Among the responses, she investigated *active lesions* using the GEE approach. Later, the results from Petkau and White (1995) and D'yachkova (1997) will be compared to those obtained from our approach.

4.2 Initial data analysis

4.2.1 Burden of disease

Following the ideas of Petkau and White (1995), the response was transformed into the variable *relative burden* (the ratio of the values of subsequent scans to that at baseline). Moreover, patient #505 was removed from the data set for further analysis as the transformation was not defined for this patient, who had zero values at baseline and for all subsequent scans except at times 3 and 4. Further analysis will be based on the transformed variable: *relative burden*.

Table 4.1: Descriptive statistics for average *relative burden* values

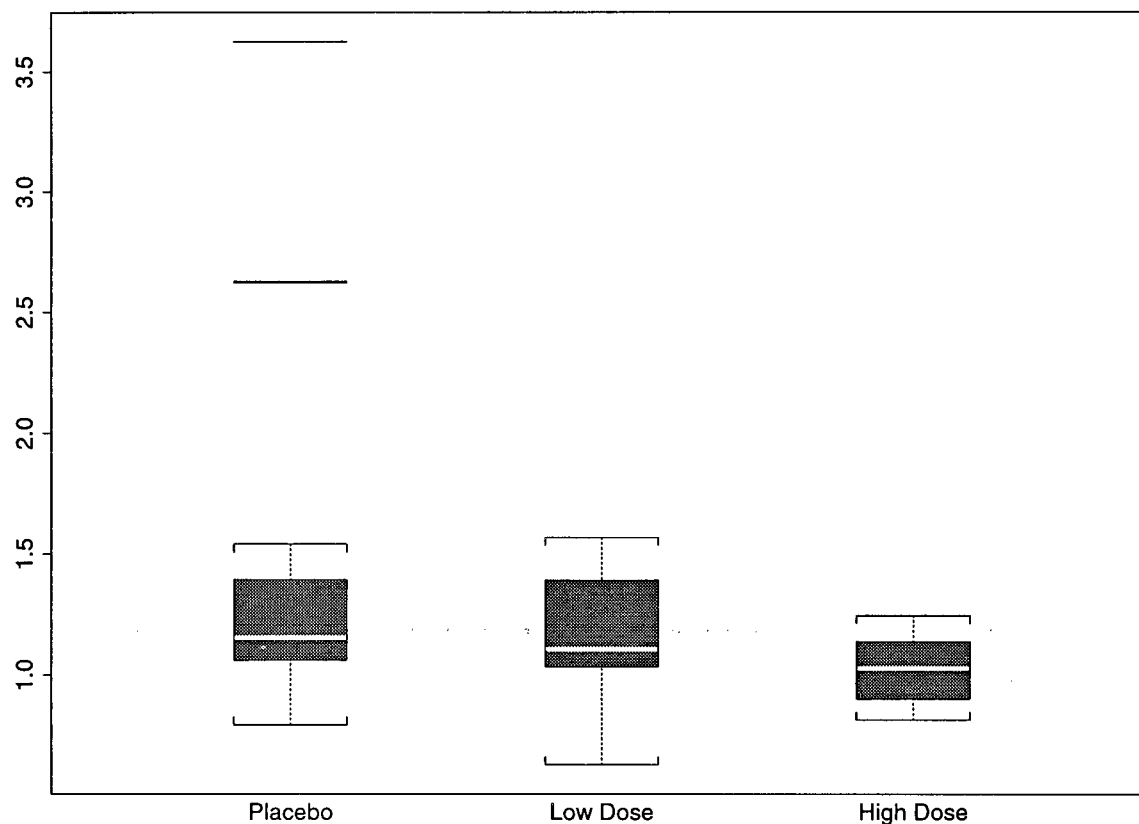
Statistic	Placebo	Low Dose	High Dose
Number of patients	17	16	16
Minimum	0.793	0.629	0.814
Maximum	3.625	1.569	1.246
Median	1.153	1.108	1.028
Mean	1.474	1.167	1.023
Standard deviation	0.760	0.255	0.138

The descriptive statistics for the average of the *relative burden* values for each patient are summarized in Table 4.1. Boxplots of the average *relative burden* values by treatment groups presented in Figure 4.1 do not suggest a strong relationship with treatment groups. However, it appears that the average *relative burden* values tend to decrease with an increase in dosage of Betaseron. Three isolated extreme values (relative to the values of the same treatment group) are seen and they are:

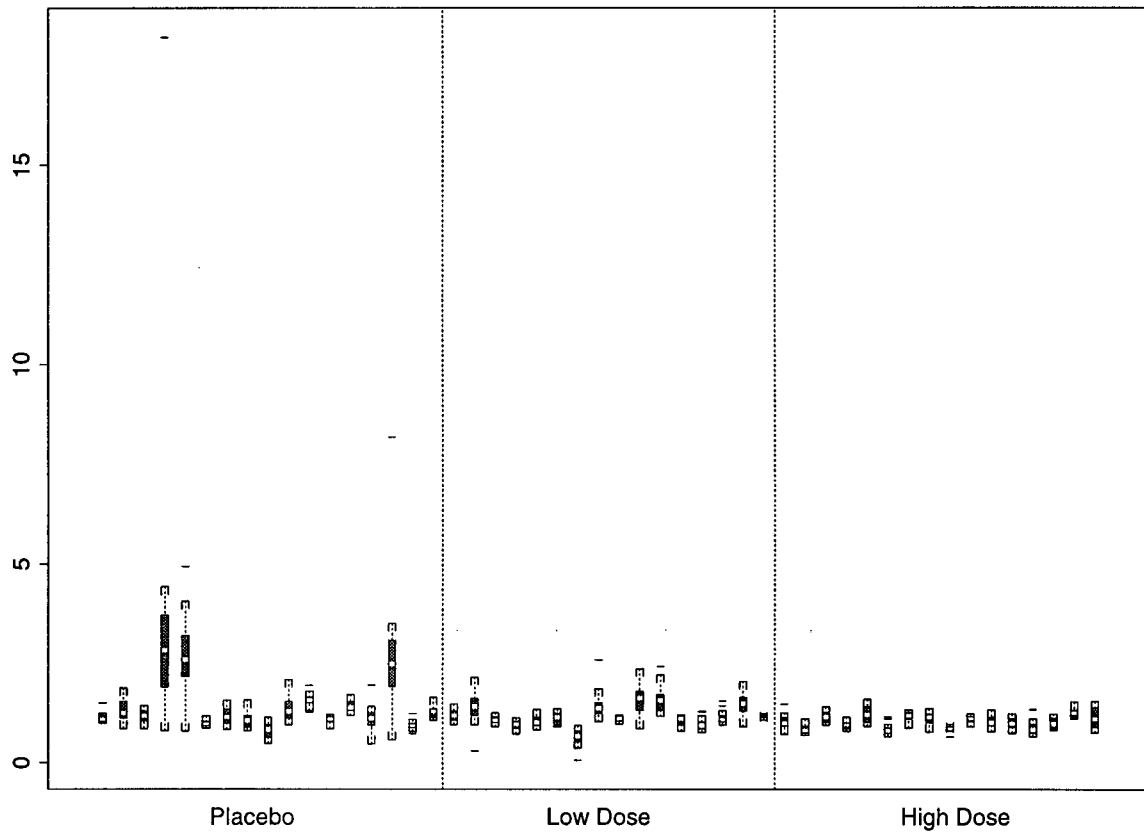
- patient #446 in the placebo group: average relative burden value is 3.625,
- patient #449 in the placebo group: average relative burden value is 2.627,
- patient #545 in the placebo group: average relative burden value is 2.623.

Boxplots of the *relative burden* values for each patient shown in Figure 4.2 indicate that three patients (#446, #449 and #545) have higher values than those for others in the placebo group. Two very extreme values (relative to the values of the same patient) are seen in Figure 4.2 and they are:

- patient #446 at time 17: *relative burden* value is 18.214,
- patient #545 at time 6: *relative burden* value is 8.179.

Figure 4.1: Boxplots of average *relative burden* values by treatment group

These two extreme values will be referred to as 446* and 545* respectively. The extreme average *relative burden* values found in three patients (#446, #449 and #545) are not solely due to one or two isolated values. Instead, Figure 4.2 seems to indicate that the collections of values are larger for these three patients with extreme average *relative burden* values. In addition, the two extreme *relative burden* values 446* and 545* are found in patients #446 and #545, respectively. The collections of values for these two patients are already larger than those of other patients in the placebo group. Special attention will be paid to investigating the influence of these two extreme values, 446* and 545*, on

Figure 4.2: Boxplots of *relative burden* values for each patient

the analysis.

Plots of the *relative burden* values versus time for each patient presented in Figure 4.3 show that three patients (#446, #449 and #545) have larger values relative to those of patients in the placebo group. It also appears that the values and their variations with time tend to decrease with an increase in dosage of Betaseron.

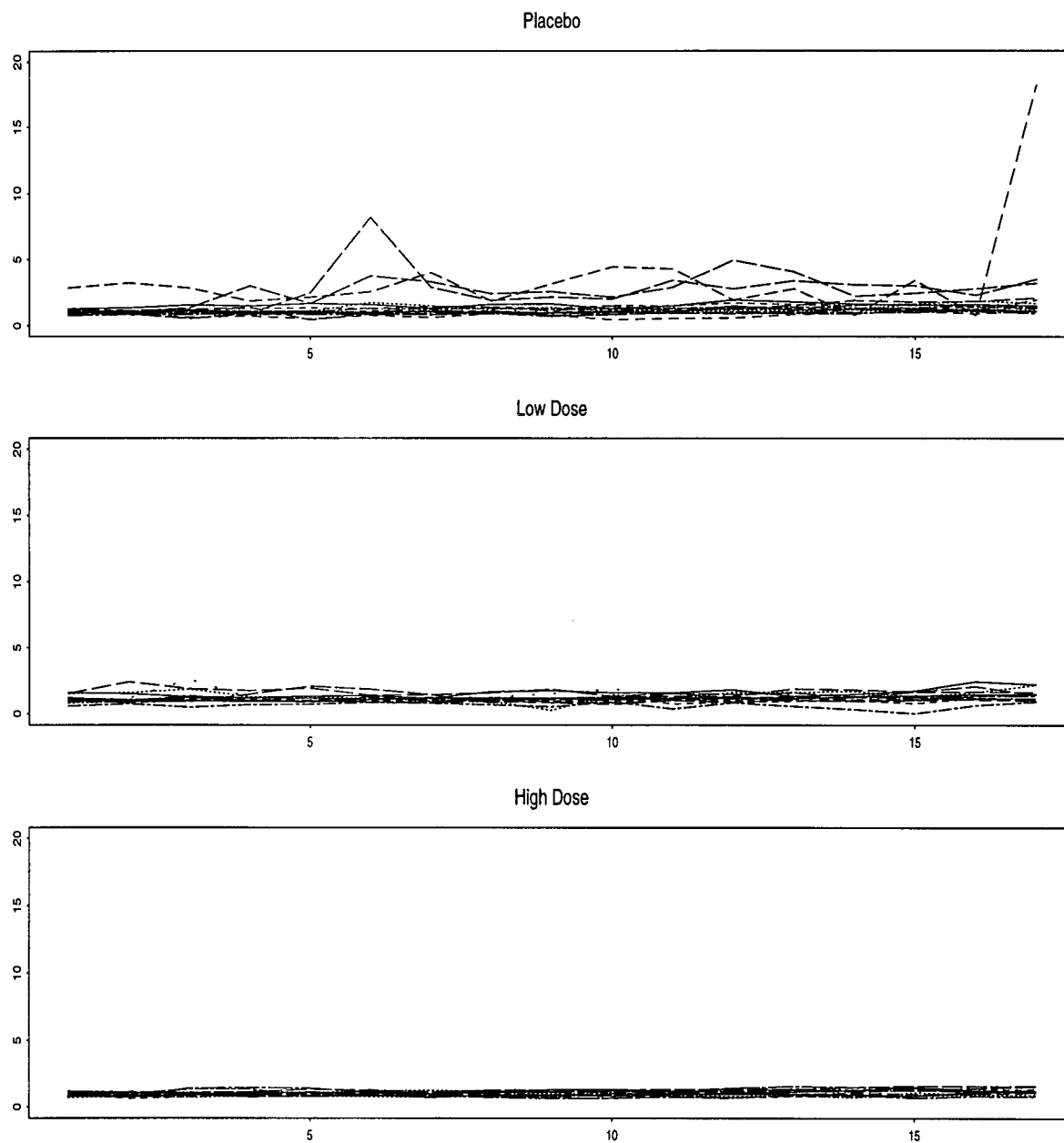
Figure 4.3: Plots of *relative burden* values versus time for each patient

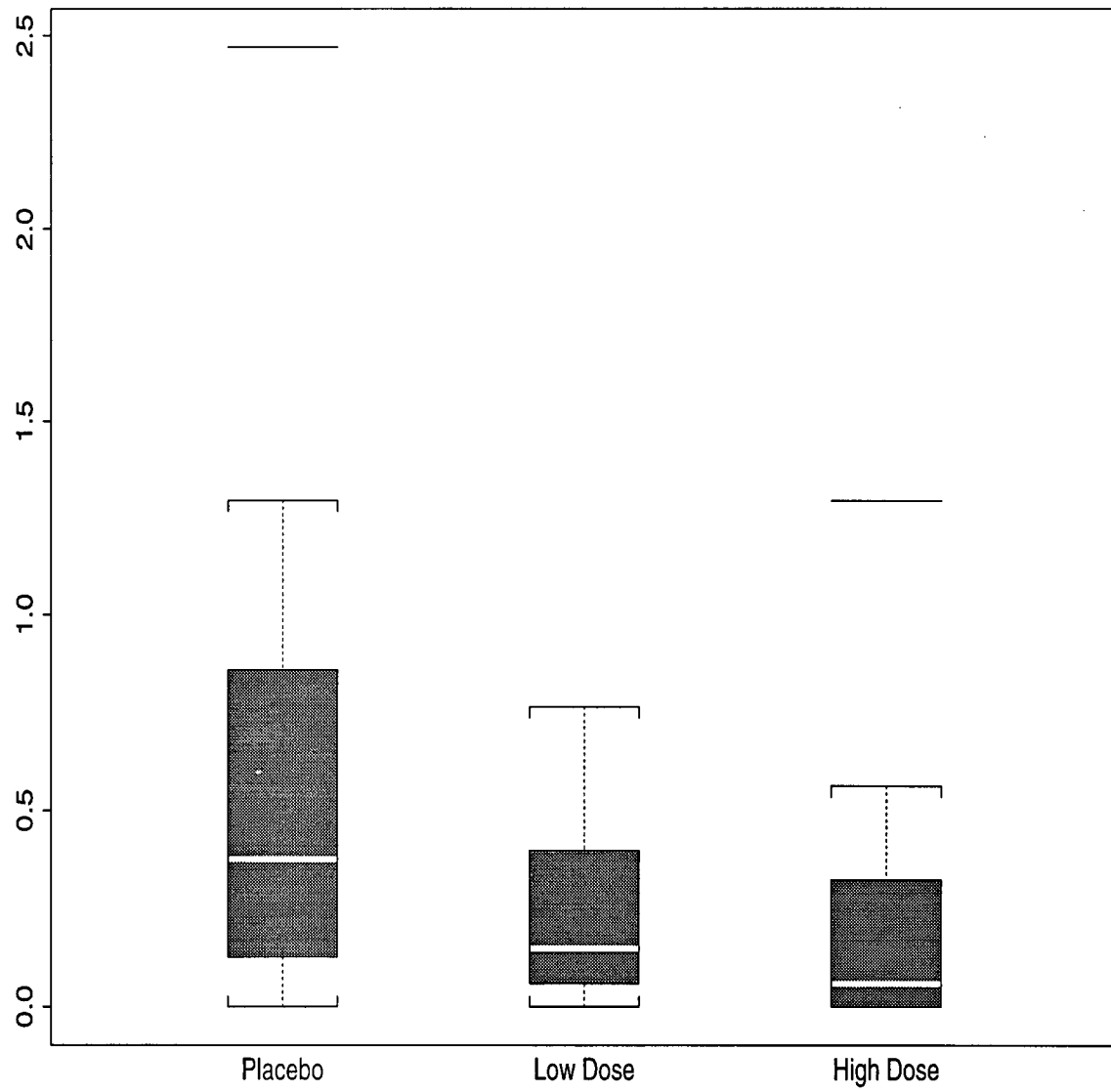
Figure 4.4: Boxplots of average *active lesions* values across treatment group

Table 4.2: Descriptive statistics for average *active lesions* values

Statistic	Placebo	Low dose	High dose
Number of patients	17	16	16
Minimum	0	0	0
Maximum	2.471	0.765	1.294
Median	0.375	0.147	0.059
Mean	0.579	0.225	0.225
Standard deviation	0.618	0.217	0.342

4.2.2 Active lesions

As an initial step, the average of the *active lesions* values for each patient was calculated and the descriptive statistics for these averages are presented in Table 4.2.

Table 4.2 suggests that there is a relationship with treatment groups. The average value in the placebo group is higher than the averages in the low and high dose groups. Boxplots of the average *active lesion* values by treatment groups shown in Figure 4.4 also indicate that there is a relationship with treatment groups. Two patients are seen to have extreme average *active lesions* values (relative to the values of the treatment group they are in) and they are listed below:

- patient #545 in the placebo group: average *active lesion* value is 2.471,
- patient #497 in the high dose group: average *active lesion* value is 1.294.

These two extreme values can be explained by the low successive counts (mostly 0 or 1) of the majority of the patients. As most patients have average *active lesions* values close to 0, this makes patients with a few high counts stand out readily.

Figure 4.5: Plots of *active lesions* values versus time

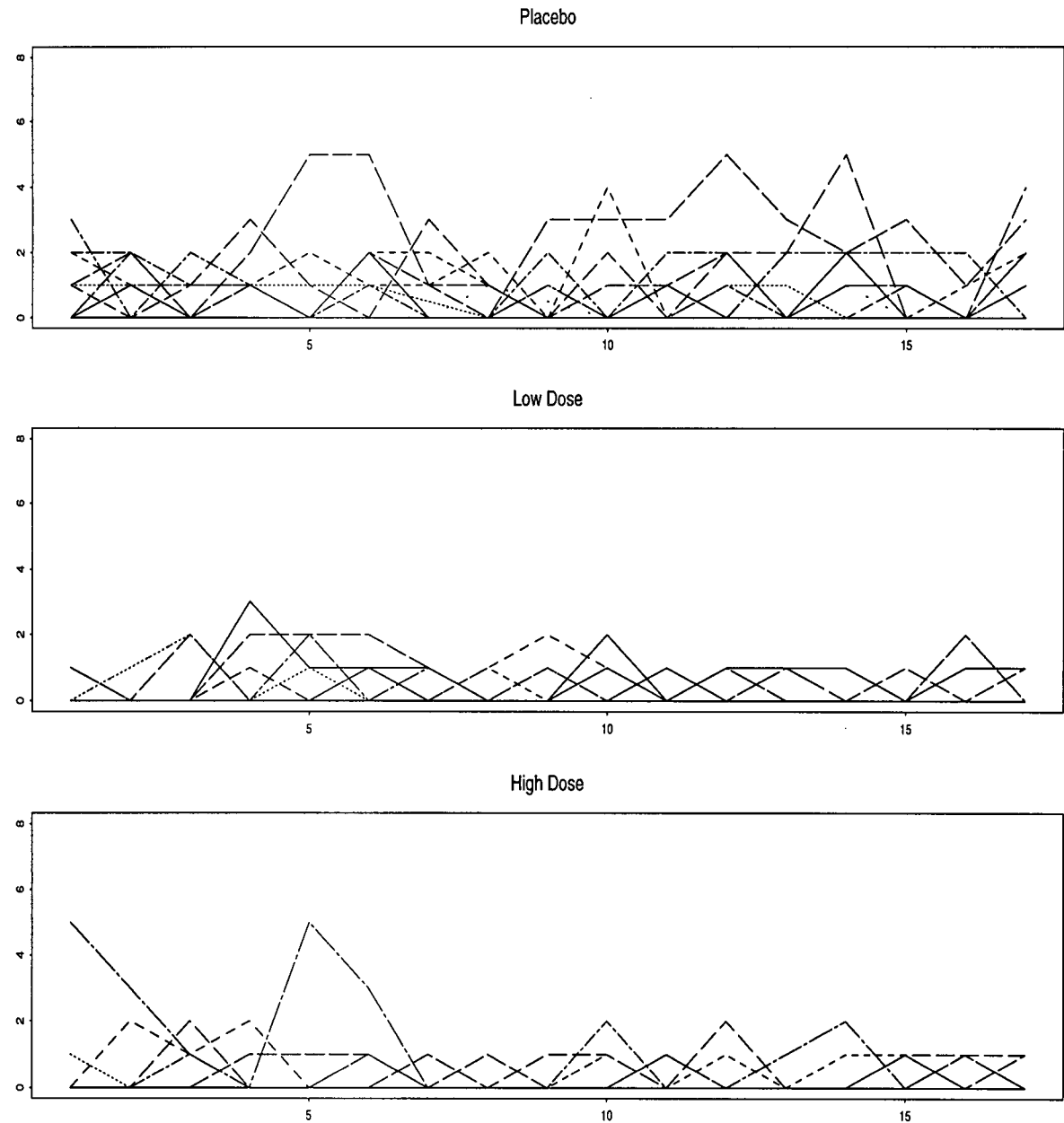


Table 4.3: Descriptive statistics for *age*, *duration*, *EDSS* and *log area*

Variable	Statistic	Placebo	Low dose	High dose
<i>age</i>	Median	35.0	36.0	36.0
	Mean	34.6	37.5	37.2
	Standard deviation	4.5	8.8	8.5
<i>duration</i>	Median	6.9	6.1	10.1
	Mean	7.4	9.5	11.0
	Standard deviation	5.0	7.2	6.4
<i>EDSS</i>	Median	1.5	2.0	2.5
	Mean	1.9	2.1	2.5
	Standard deviation	0.9	1.2	1.1
<i>log area</i>	Median	7.5	7.0	7.5
	Mean	6.9	6.9	7.5
	Standard deviation	1.6	1.6	0.8

Plots of *active lesions* values versus time for each patient are presented in Figure 4.5. Patient #545 in the placebo group appears to have a collection of larger values of *active lesions* than those of other patients in the placebo group. Figure 4.5 also seems to indicate that patient #497 only has larger number of active lesions in the first 6 MRI scans; the number of active lesions becomes smaller from time 7 onwards. In general, it seems that the number of active lesions decreases with an increase in dosage of the drug Betaseron.

4.2.3 Covariates

The descriptive statistics for *age*, *duration*, *EDSS* and *log area* are shown in Table 4.3. These covariates are also summarized in the boxplots in Figure 4.6. The average age seems to be roughly the same across the treatment groups. However, the variability of age in the placebo group appears to be less than those in the low and high dose groups.

Figure 4.6: Boxplots of covariates values across treatment group

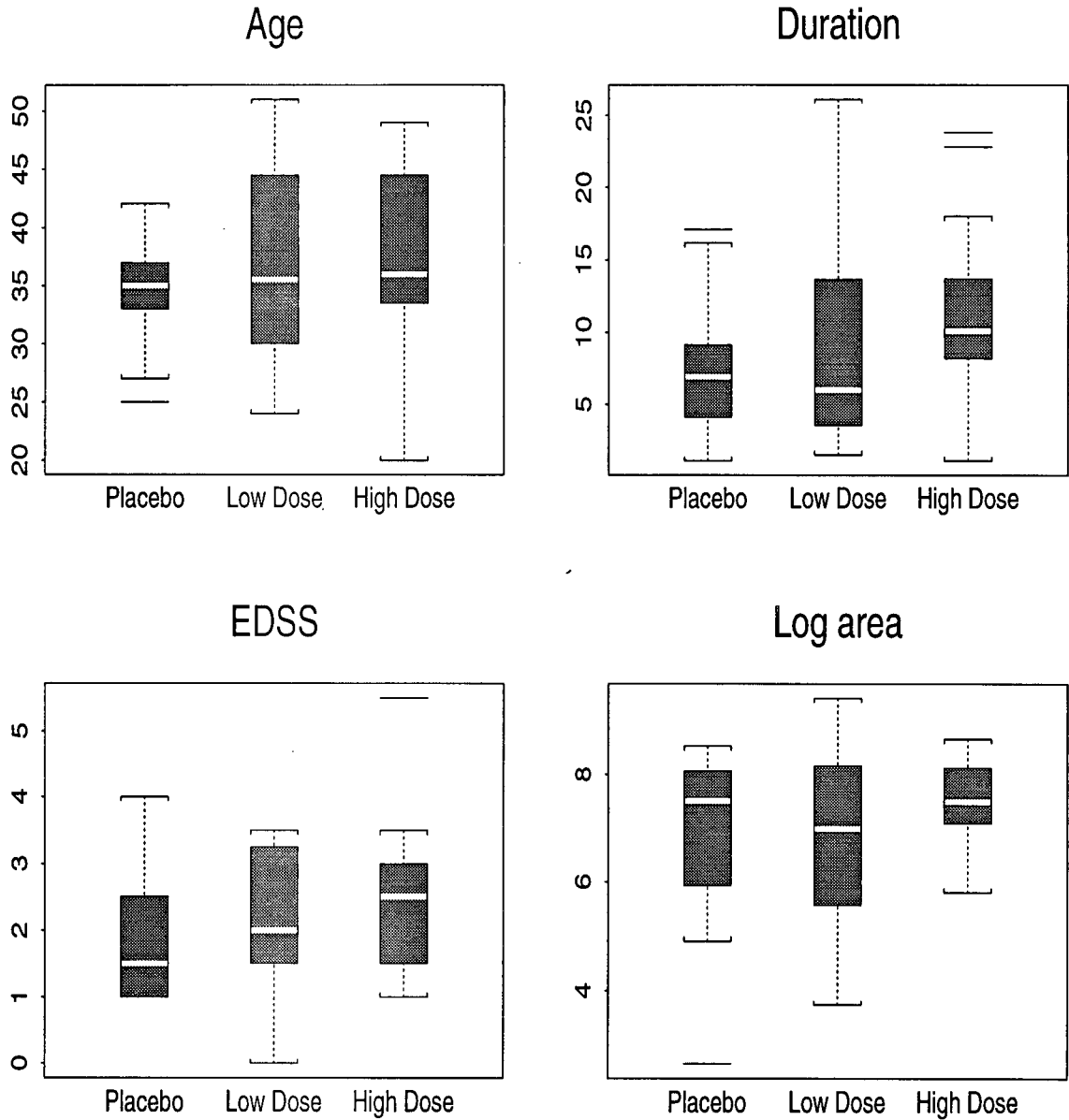


Table 4.4: Counts for *origin* and *gender*

Variable	Category	Placebo	Low Dose	High Dose
<i>origin</i>	B.C.	15	10	14
	Washington	2	6	2
<i>gender</i>	Male	7	2	3
	Female	10	14	13

The patients in the high dose group appear to have longer duration of disease than those for patients in the other groups and roughly the same variability as those in the placebo group. The patients in the low dose group seem to have the highest variability in the duration of disease. It is worth noting that *EDSS* is an ordinal variable. For convenience, we treat it as a continuous variable in the analyses which we report here. However, we should keep this in mind and explore the possibility in future to include *EDSS* as an ordinal variable in the analysis. The initial *EDSS* scores appear to be roughly comparable in mean and in variability across the treatment groups. The averages of *log area* values seem to be roughly equal across the treatment groups but the high dose patients appear to have the least spread of *log area* values among patients in other treatment groups.

The counts for *origin* and *gender* in each treatment group is presented in Table 4.4. As seen in Table 4.4, it appears that the origin of patients in the placebo and high dose group is highly unbalanced. Only 2 out of the 17 patients from the placebo group and only 2 out of the 16 patients in the high dose group were from Washington. Imbalances in the gender distribution of the patients were also observed in the low and high dose groups. The patients in the low dose and high dose group were mostly male (2 out of 16 in the low dose group and 3 out of 16 in the high dose group).

This descriptive information about the covariates indicate that there are imbalances in covariates across treatment groups which have to be taken into account in the analysis to follow.

4.3 Model identification

The only time-varying covariate that we consider in this thesis is time itself, which is used as a long-term covariate. All the baseline covariates together with the three treatment groups are included as constant baseline covariates in the initial values part of the model in Section 3.1. Hence, our preliminary model for the Betaseron data is as follows.

Long-term covariate:

time.

Baseline covariates:

intercept;

low dose;

high dose;

age;

EDSS;

duration;

origin;

gender;

log area.

Let \mathbf{Y}_{jt} be the bivariate response vector of observations at time t for patient j with

Y_{ijt} , $i = 1$ (*relative burden*), 2 (*active lesions*) as its components (called categories). Note that the notation in Section 3.1 is used here. We assume *relative burden* to follow a gamma distribution, which corresponds to $p_1 = 2$. We assume *active lesions* to be Poisson counts, which corresponds to $p_2 = 1$ and $\nu_2^2 = 1$. The latent processes for all the patients are assumed to follow a gamma distribution; that is, we choose $q = 2$. For the initial values, the random effects are assumed to follow a gamma distribution, which corresponds to $r = 2$. The dispersion parameter, ν_i^2 , is assumed to be the same across patients. In addition, all the latent processes are assumed to have the same dispersion parameter, σ^2 . Note that the dispersion parameter for a Poisson distribution is known and is equal to 1. The remaining unknown dispersion parameters are ν_1^2 , σ^2 and ω^2 and they are estimated from the data.

In the initial values part of the model, for identifiability, the effect for the placebo group is set to zero. In this way, the intercept represents the placebo level, and the effects for *low dose* and *high dose* represent the effects relative to *placebo*. The fitted model for the initial values becomes:

$$\begin{aligned} \log \hat{g}_{kj} = & \hat{\alpha} + \hat{\tau}_k + \hat{\gamma}_1 \text{age} + \hat{\gamma}_2 \text{EDSS} + \hat{\gamma}_3 \text{duration} + \\ & \hat{\gamma}_4 \text{origin} + \hat{\gamma}_5 \text{gender} + \hat{\gamma}_6 \log \text{area} \end{aligned}$$

where \hat{g}_{kj} is the expected initial value of the latent process for the j th patient in the k th treatment group; $\hat{\tau}_k$ are the treatment effects for the k th treatment group; $k = 0$ (placebo), 1 (low dose), 2 (high dose). Note that the effect for the placebo group is $\tau_0 \equiv 0$.

As mentioned in section 4.2.1, there are two extreme values, 446* and 545*, found in the variable *relative burden*. To investigate their effects on the analysis, we fit the

model separately to the full data set, the data set with 446* removed and the data set with both 446* and 545* removed. The estimates and their asymptotic standard errors, shown in brackets, for the baseline covariates are reported in Table 4.5 and those for the long-term covariate are presented in Table 4.6, 4.7 and 4.8. Our model allows the long-term effect for each patient to be different. There are all together 49 effects for *time*, one for each patient. The regression parameter for *time* can be interpreted as the rate of change of the log of the underlying severity per 6-week period. That is, each patient is assumed to have a different development of the disease. This is reasonable as there is natural heterogeneity among patients. Moreover, patients may react differently to the drug applied. A positive estimate for the time effect suggests an increasing trend over time of the severity. As the observations reflect the underlying severity, their patterns over time can well be addressed by investigating the trends of the latent processes.

The baseline covariates influence the initial values of the latent process with time taken into account. The main focus is on the effects for the treatment groups. As explained in the beginning of this section, the estimate for low dose represents the effect of the low dose group relative to that of the placebo group. A negative estimate suggests a possible lowering of the initial values, thus hinting a possible improvement in the disease. The estimate for *high dose* has a similar interpretation.

The two extreme values exerted great influences on the estimates for the effects for *intercept*, *low dose*, *high dose* and *log area*. For example, the estimates for the effects of the low dose group varies from 0.081 (full data set) to -0.118 (with 446* removed) and then to -0.067 (with 446* and 545* removed). Similarly, the estimates for the effects of the high dose group changes from -0.078 (full data set), -0.173 (with 446* removed) and then to -0.131 (with 446* and 545* removed). The dominant effects of these extreme

Table 4.5: Estimates and standard errors for baseline covariate effects

	full data set	with 446* removed	with 446* and 545* removed
Parameter	estimate (s.e.)	estimate (s.e.)	estimate (s.e.)
<i>intercept</i>	-0.131 (0.280)	1.048 (0.253)	0.924 (0.276)
<i>low dose</i>	0.081 (0.075)	-0.118 (0.066)	-0.067 (0.073)
<i>high dose</i>	-0.078 (0.081)	-0.173 (0.072)	-0.131 (0.078)
<i>age</i>	-0.006 (0.004)	-0.009 (0.004)	-0.006 (0.004)
<i>EDSS</i>	0.066 (0.032)	0.048 (0.029)	0.055 (0.031)
<i>duration</i>	-0.009 (0.005)	-0.007 (0.005)	-0.009 (0.005)
<i>origin</i>	-0.057 (0.080)	-0.082 (0.071)	-0.087 (0.078)
<i>gender</i>	0.051 (0.078)	0.053 (0.071)	0.043 (0.076)
<i>log area</i>	0.033 (0.021)	-0.095 (0.019)	-0.091 (0.021)

values on the estimates are not unexpected. Although it is not desirable to have the results of the analysis strongly influenced by a few extreme values, it requires further investigations before removing them from the data set. Provided that the model gives a reasonable fit to the data, we will retain as many data points as possible to preserve valuable information. In the following section, we will see that the extreme value 446* causes some violations of model assumptions. As it is the last data point of the time series for patient #446, it causes a strong upward trend of the latent process from time 7 onwards. When 446* is removed from the data set, the model fits reasonably well. The second extreme value 545* does not have much effect on the adequacy of the model. It may be due to the fact that it is in the middle of the time series. However, it stands out as an outlier in the plots of residuals. On the one hand, we need to keep in mind the effects of these extreme values on model fitting. On the other hand, as these two large extreme values are found in the placebo patients, they may suggest evidence of a treatment effect.

Table 4.6: Estimates and standard errors for the long-term covariate (*time*) effects for patients in the placebo group

	full data set	with 446* removed	with 446* and 545* removed
Patient #	estimate (s.e.)	estimate (s.e.)	estimate (s.e.)
420	0.011 (0.032)	0.010 (0.011)	0.011 (0.012)
421	0.031 (0.029)	0.017 (0.010)	0.022 (0.010)
443	0.017 (0.030)	0.013 (0.010)	0.020 (0.011)
446	0.935 (0.000)	0.007 (0.010)	-0.005 (0.011)
449	0.053 (0.026)	0.051 (0.009)	0.047 (0.009)
451	0.007 (0.033)	0.006 (0.012)	0.010 (0.012)
498	0.016 (0.029)	0.017 (0.011)	0.022 (0.011)
501	0.020 (0.031)	0.001 (0.011)	0.012 (0.011)
504	0.010 (0.034)	-0.020 (0.012)	-0.001 (0.012)
507	0.046 (0.028)	0.038 (0.010)	0.041 (0.010)
522	0.020 (0.029)	0.018 (0.011)	0.014 (0.011)
523	0.005 (0.031)	0.005 (0.011)	0.008 (0.011)
539	0.030 (0.033)	0.025 (0.013)	0.025 (0.013)
540	0.030 (0.029)	-0.012 (0.010)	0.005 (0.010)
545	0.056 (0.027)	0.057 (0.008)	0.068 (0.008)
550	0.000 (0.032)	0.002 (0.012)	0.006 (0.012)
565	0.025 (0.031)	0.028 (0.011)	0.026 (0.012)

Table 4.7: Estimates and standard errors for the long-term covariate effects (*time*) for patients in the low dose group

	full data set	with 446* removed	with 446* and 545* removed
Patient #	estimate (s.e.)	estimate (s.e.)	estimate (s.e.)
419	0.022 (0.030)	0.029 (0.011)	0.026 (0.012)
424	0.030 (0.028)	-0.002 (0.010)	0.007 (0.010)
448	0.008 (0.031)	0.015 (0.013)	0.010 (0.013)
450	-0.001 (0.033)	0.010 (0.013)	0.010 (0.013)
452	0.009 (0.033)	0.003 (0.012)	0.001 (0.012)
499	0.006 (0.029)	0.025 (0.011)	0.026 (0.012)
502	-0.022 (0.039)	-0.037 (0.014)	-0.027 (0.014)
508	0.009 (0.029)	0.011 (0.012)	0.007 (0.011)
521	-0.005 (0.030)	0.008 (0.011)	0.008 (0.011)
525	0.021 (0.029)	0.002 (0.010)	-0.002 (0.011)
542	0.043 (0.028)	0.030 (0.010)	0.030 (0.010)
544	-0.002 (0.037)	0.022 (0.016)	0.023 (0.016)
547	-0.009 (0.033)	-0.018 (0.012)	-0.009 (0.011)
548	0.017 (0.032)	0.008 (0.011)	0.013 (0.011)
564	0.013 (0.028)	0.014 (0.010)	0.014 (0.010)
568	-0.005 (0.036)	0.019 (0.016)	0.011 (0.016)

Table 4.8: Estimates and standard errors for the long-term covariate effects (*time*) for patients in the high dose group

	full data set	with 446* removed	with 446* and 545* removed
Patient #	estimate (s.e.)	estimate (s.e.)	estimate (s.e.)
422	0.020 (0.032)	0.015 (0.011)	0.018 (0.012)
444	0.006 (0.036)	-0.004 (0.013)	0.002 (0.013)
445	0.026 (0.032)	0.030 (0.011)	0.026 (0.012)
453	0.011 (0.034)	0.008 (0.012)	0.012 (0.012)
454	0.035 (0.030)	0.042 (0.011)	0.040 (0.011)
497	-0.031 (0.033)	-0.026 (0.013)	-0.025 (0.013)
500	0.015 (0.033)	0.015 (0.011)	0.012 (0.012)
503	0.009 (0.031)	0.009 (0.011)	0.013 (0.011)
506	-0.038 (0.042)	-0.029 (0.019)	-0.021 (0.018)
524	0.010 (0.031)	0.020 (0.012)	0.018 (0.012)
526	0.014 (0.034)	0.014 (0.012)	0.011 (0.012)
541	-0.003 (0.034)	-0.011 (0.012)	-0.011 (0.012)
543	0.025 (0.035)	0.008 (0.013)	0.012 (0.013)
546	0.005 (0.041)	0.020 (0.018)	0.021 (0.018)
549	0.023 (0.031)	0.024 (0.011)	0.020 (0.012)
566	0.017 (0.032)	0.008 (0.011)	0.012 (0.011)

4.4 Model checking

Following the notation used in Section 3.4, we refer the standardized filter residuals as R_{1t} (from the active lesion process), R_{2t} (from the relative burden process) and r_t (from the latent process). The corresponding standardized smoother residuals are denoted by R_{1t}^* , R_{2t}^* and r_t^* . For notational convenience, the index j for a given series is suppressed. We now analyse the residuals from the fitted models on the full data set, on the data set with 446* removed and on the data set with both 446* and 545* removed.

4.4.1 Full data set

For the fitted model using the full data set, we will first analyze the residuals from each patient one at a time and then consider the residuals from the whole data set to investigate their overall pattern. Among the 49 patients in this study, we will only discuss our findings on four (2 from the placebo group, one from the low dose group and one from the high dose group) of the patients. Patients with extreme values are of particular importance. Therefore, we will present our findings on the residuals analysis for patient #446 and #545 in the placebo group. Patient #502 and #500 in the low dose and high dose group respectively are also selected. The residual plots for patient #497, who has extreme average *active lesions*, are available in Appendix C for reference.

Plots for the residuals from patient #446 are shown in Figure 4.7. The autocorrelations function (ACF) plot for R_{1t} in Figure 4.7(a) clearly indicates that R_{1t} are correlated as the autocorrelations for lag 1 and 2 fall significantly outside the asymptotic 95% confidence bands; thus violating the conditional independence assumption of *active lesions*

Figure 4.7: Residuals plots for patient #446 in the placebo group

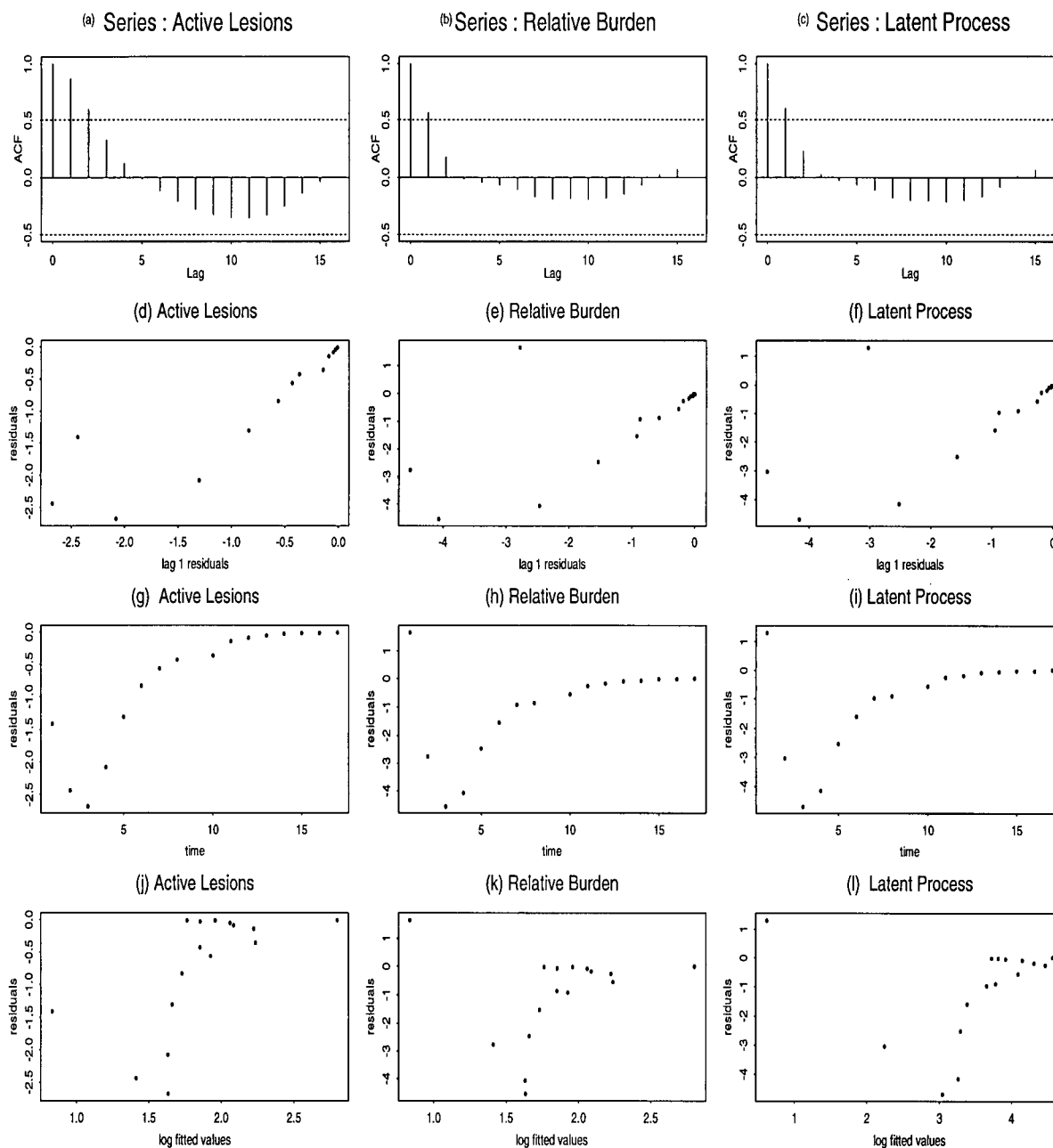
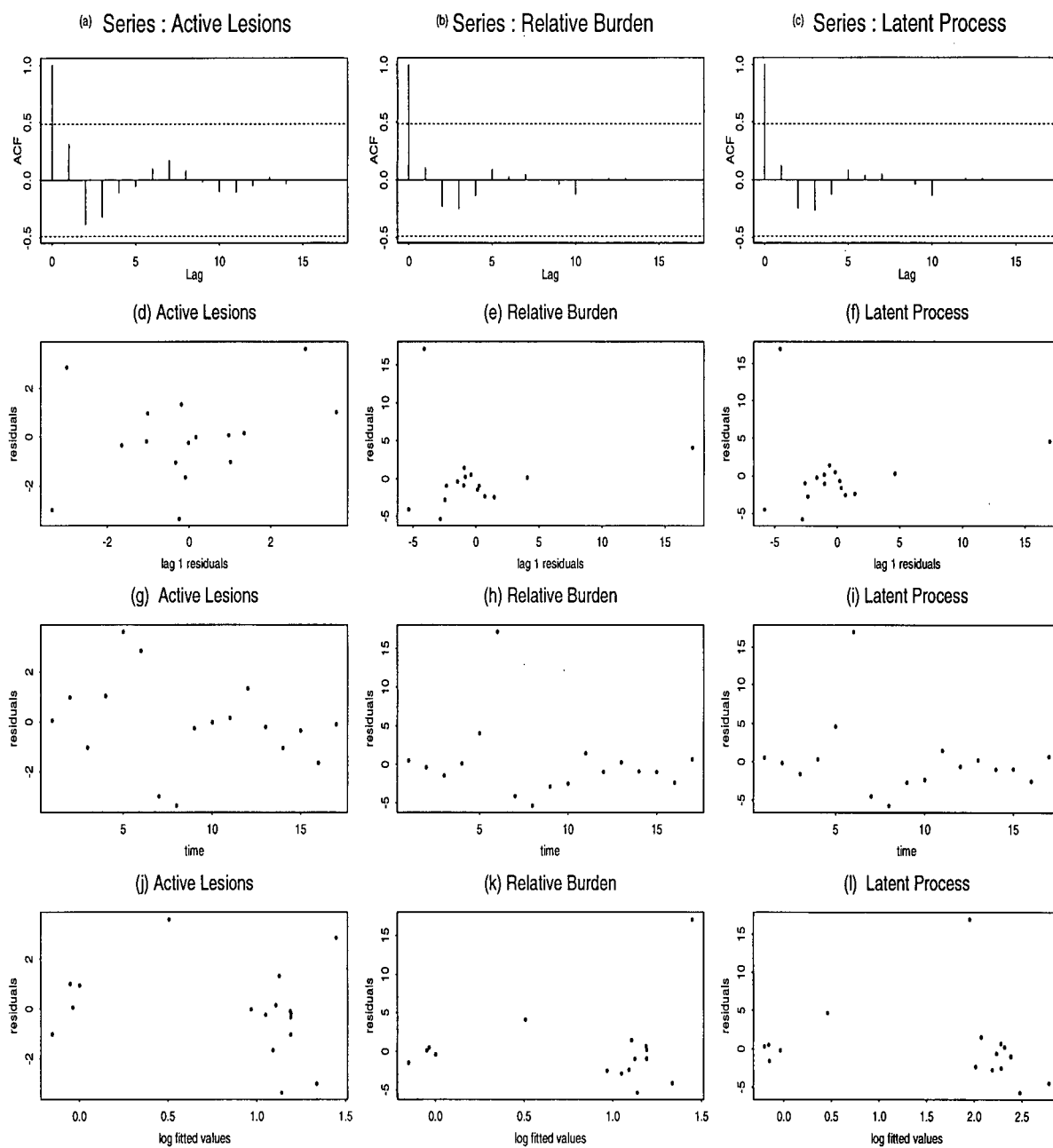


Figure 4.8: Residuals plots for patient #545 in the placebo group



values over time. Moreover, both the ACF plot for R_{2t} (Figure 4.7(b)) and for r_t (Figure 4.7(c)) have lag 1 correlations found outside the asymptotic 95% confidence bands. The plots of R_{1t} against R_{1t-1} (Figure 4.7(d)) and R_{2t} against R_{2t-1} (Figure 4.7(e)) show a linear pattern in the right most part of each plot. This gives further evidence against the assumptions that values from the two observation processes are conditionally independent over time, given the latent process. Evidence against the Markov assumption of the latent process is seen in the plot of r_t against r_{t-1} (Figure 4.7(f)) which shows a linear relationship in the right part of the plot. Curved patterns are observed in the plots of R_{1t} against time (Figure 4.7(g)) and R_{2t} against time (Figure 4.7(h)). This means that the values of the observation process are not conditionally independent over time, given the latent process. The same pattern is also visible in the plots of r_t against time (Figure 4.7(i)). It confirms the inadequacy of the Markov assumption for the latent process. The plots for the residuals against the log-linear predictor presented in Figure 4.7(j), (k) and (l) all have a peculiar structure which should not be expected. The distributional assumptions for the observation and the latent processes can be considered inappropriate. Note that an outlier corresponding to the observation at time 17 (the extreme value 446* mentioned in section 4.2.1) is visible in plots Figure 4.7(e), (f), (h), (i), (k) and (l). When there is something wrong with the model assumptions, it manifests itself in various ways which are seen in the residual plots. All the residuals plots suggest inadequacy of the model. The residuals analysis for patient #446 gives convincing evidence that the model assumptions are inappropriate. We suspect that the extreme value 446* causes some serious violations of main model assumptions. Although this is not entirely obvious from the plots in Figure 4.7, 446* at the last MRI scan (time 17) for that patient may pull the estimated latent process upwards to give rise to some patterns observed in those plots.

Figure 4.9: Residuals plots for patient #502 in the low dose group

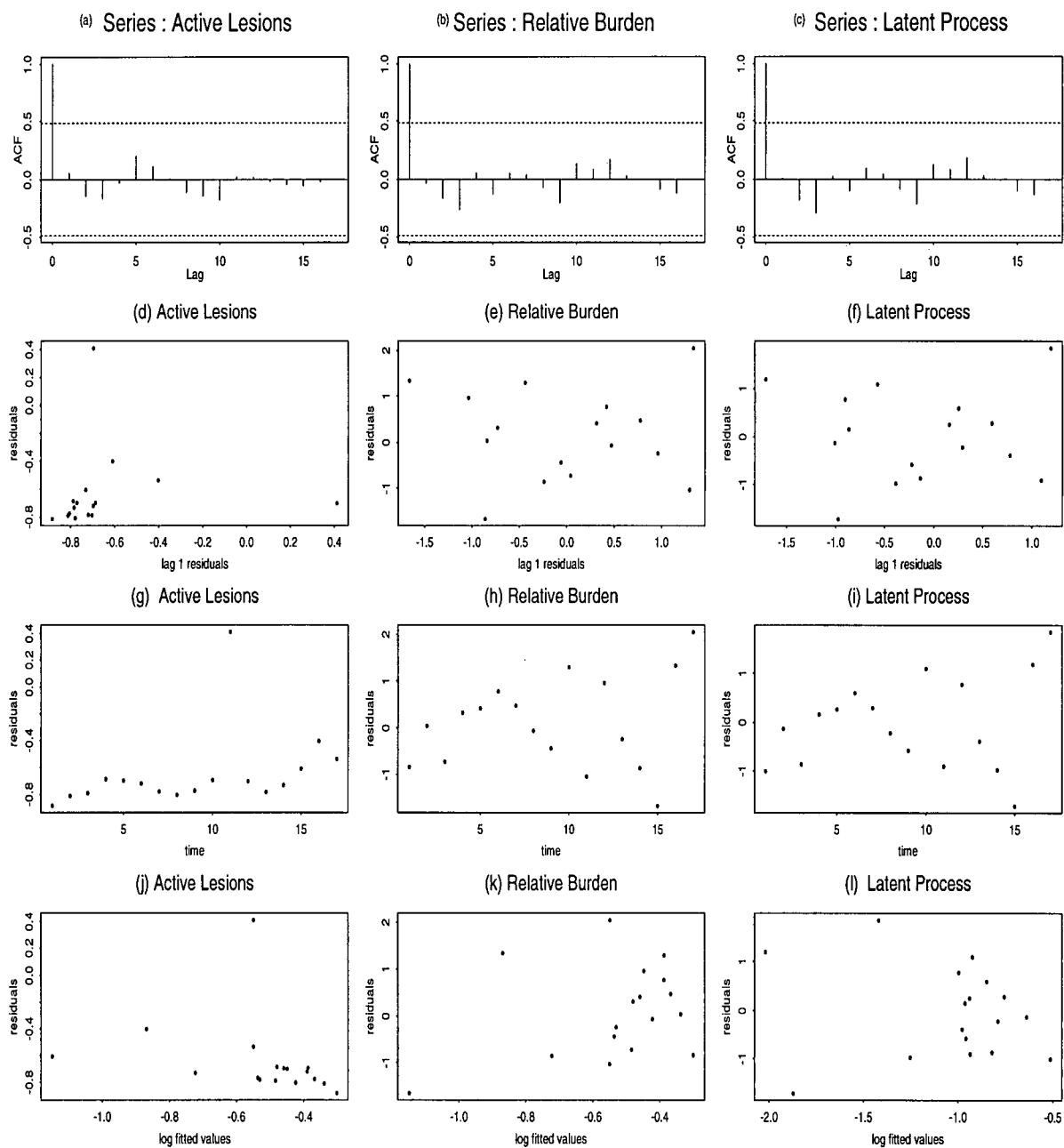


Figure 4.10: Residuals plots for patient #500 in the high dose group

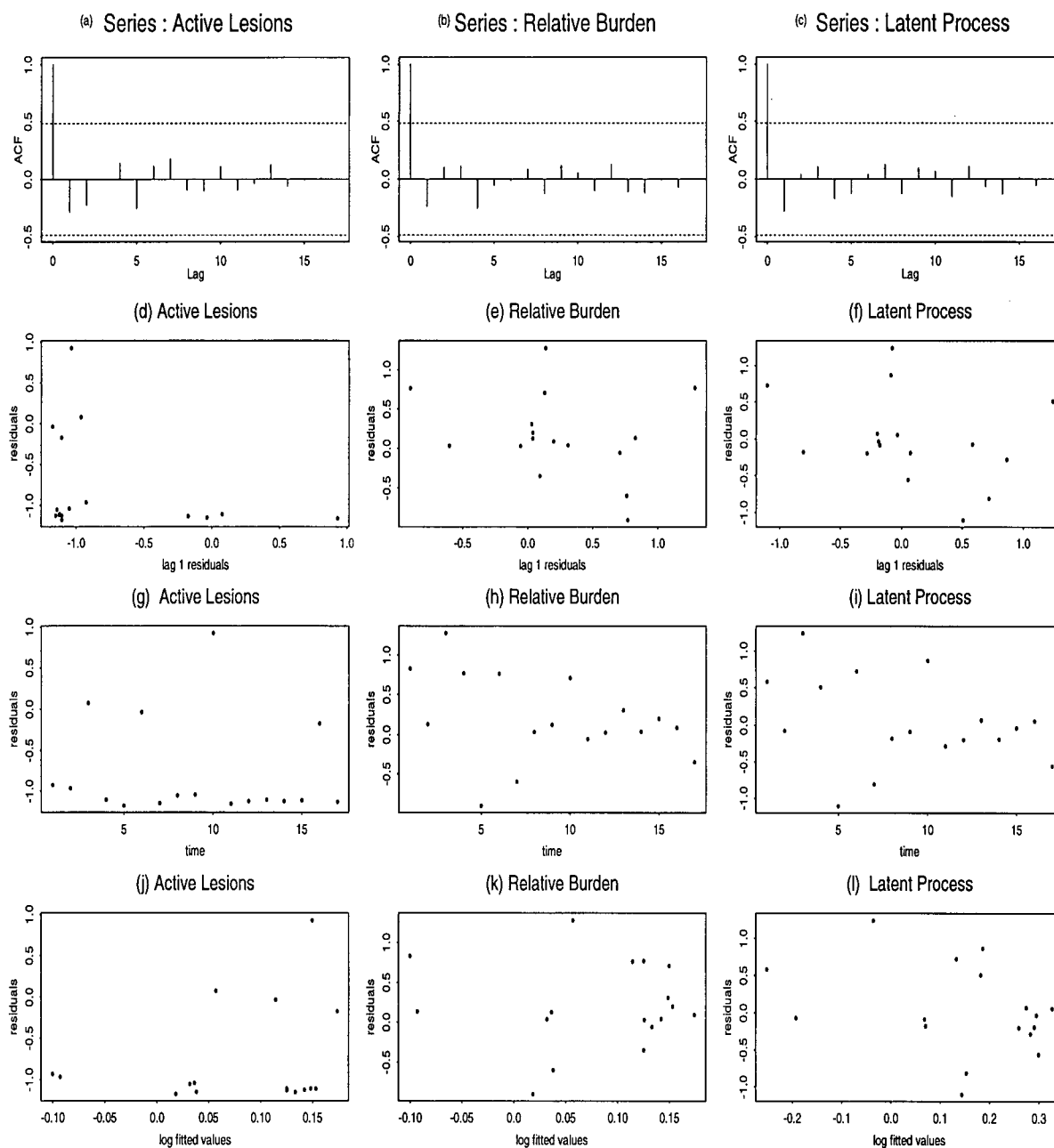


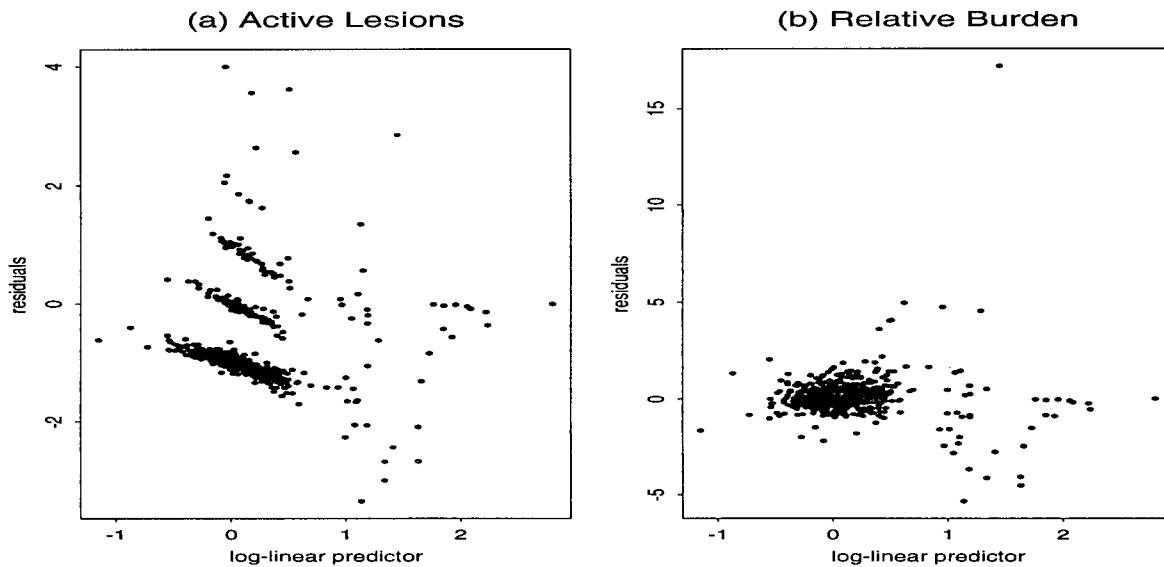
Figure 4.8 consists of residuals plots for patient #545. The ACF plot for R_{1t} (Figure 4.8(a)) does not show significant autocorrelation and neither does the ACF plot for R_{2t} (Figure 4.8(b)). These give no evidence against the conditional independence assumption of the observation processes. No significant autocorrelation is seen in the ACF plot for r_t (Figure 4.8(c)) giving no evidence against the Markov assumption of the latent process. No specific structure is observed in the plot of R_{1t} against R_{1t-1} as shown in Figure 4.8(d). However, the plot for R_{2t} against R_{2t-1} presented in Figure 4.8(e) shows a weak linear pattern with 2 outliers. It is worth noting that these outliers are due to the extreme value at time 6, that is, 545* (mentioned in section 4.2.1). Hence, Figure 4.8(e) suggests a mild violation of the conditional independence assumption of the observation process. A similar weak linear pattern is also apparent in the plot for r_t against r_{t-1} as shown in Figure 4.8(f); it gives evidence against the Markov assumption of the latent process. Two outliers are also seen and they are due to the extreme value 545*. An outlier is observed in both Figure 4.8(h) and 4.8(i) which correspond to 545*. Besides that, nothing unusual is seen. No particular pattern is observed in the plot of R_{1t} against the log-linear predictor (Figure 4.8(j)) and neither for R_{2t} against the log-linear predictor (Figure 4.8(k)) and r_t against the log fitted values (Figure 4.8(l)) except that an outlier is present. The outliers that are found in Figure 4.8(k) and in Figure 4.8(l) are caused by the extreme value 545*. From these plots, the extreme value 545* is confirmed to be an outlier and mild violations of the conditional independence assumptions are evident.

Plots for the residuals from patient #502 in the low dose group are presented in Figure 4.9. None of the ACF plots for R_{1t} , R_{2t} and r_t have autocorrelations falling outside the 95% confidence bands. There are two possible outliers in the plots of R_{1t} against R_{1t-1} shown in Figure 4.9(d). A closer look reveals that this feature is due to the discreteness of the data. The values of *active lesions* are all 0 except at time 11 when the

number of active lesions is 1. The point in the far right corresponds to the case where the data value of 0 is immediately preceded by a value of 1. The reverse order gives the point in the upper left corner. The cluster of residuals in the lower left corner correspond to successive data values of 0. Although the residual at time 11 stands out from the rest, its value is 0.41 which is not large by any means. The residual from time 11 is perceived as an outlier in both Figure 4.9(g) and (j). Note that this feature caused by the discreteness of the data will also be observed in other plots. No particular pattern is observed in the plot for R_{2t} against R_{2t-1} presented in Figure 4.9(e) and r_t against r_{t-1} in Figure 4.9(f). The Markov assumption of the latent process appears to be acceptable as the plot of r_t against r_{t-1} in Figure 4.9(f) reveals no pattern. Again, nothing specific is observed for plots in Figure 4.9(g), (h), (i), (j), (k) and (l). In Figure 4.10(g), it is worth pointing out that the residuals in the lower band corresponding to data values of 0 and the isolated residual appears in the top corresponding to data value of 1. The same explanation applies to Figure 4.9(j).

Plots for the residuals from patient #500 in the high dose group are presented in Figure 4.10. None of these plots give evidence against the main model assumptions. We now comment on some features found in these plots which are caused by the discrete nature of active lesions counts. In Figure 4.10(g) where R_{1t} is plotted against time, all the residuals appear to align themselves in three bands. The residuals in the lowest band correspond to the counts of 0. Residuals from the counts of 1 are in the second lowest band whereas the residual in the upper left hand corner comes from the observation with 2 active lesions. This feature is also observed in Figure 4.10(j). The plot of R_{1t} against R_{1t-1} presented in Figure 4.10(d) also reveals an interesting feature due to the discreteness of the data. The residuals appear in clusters. The cluster in the lower left corner corresponds to two consecutive 0 observations. The cluster in the center-left portion are

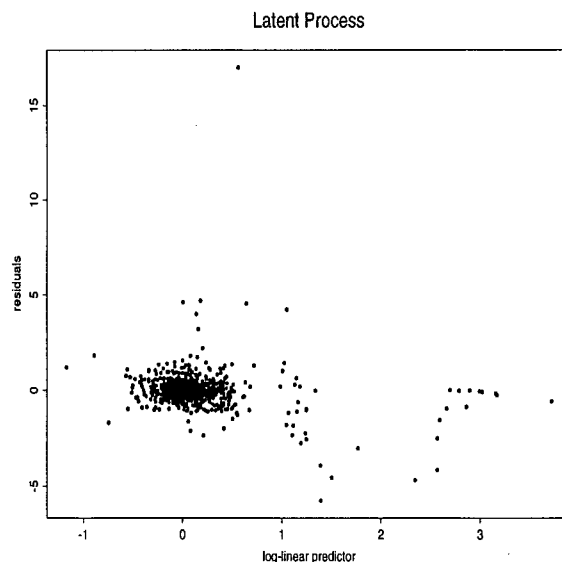
Figure 4.11: Residuals against log-linear predictors for the two categories



due to a data value of 1 immediately preceded by a data value of 0 whereas a count of 2 immediately preceded by a count of 0 gives the isolated point in the upper-left corner. The features caused by the discrete nature of the observation are highly visible in plots where residuals from the entire data set are considered together.

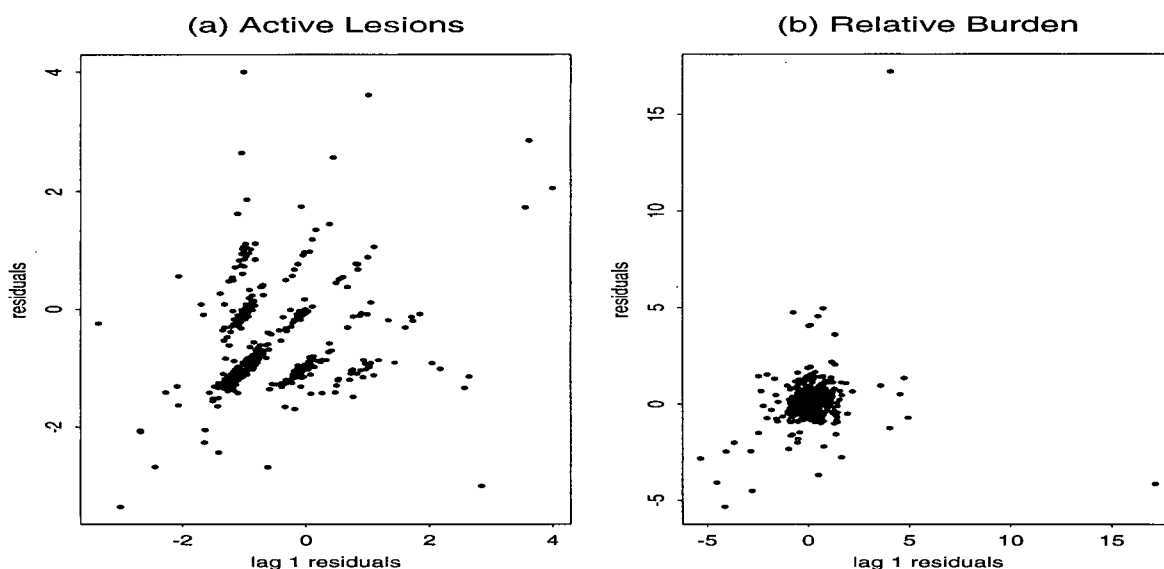
We now consider residual plots using the entire data set. Plots for R_{it} against log predicted values of the observed processes based on the Kalman filter are shown in Figure 4.11. The discreteness of the active lesions values, coupled with the low expected counts, accounts for the curved set of bands with negative slope found in Figure 4.11(a). The lowest band are the filter residuals for data values of 0, the next lowest band for data values of 1 and so on. In particular, it is impossible to have large negative residuals for small fitted values. This feature, together with the negative slope of the curves on which the residuals lie, makes the residual plots show an apparent downward trend. These bands dominate the plot for log expected counts less than 1 but seem to disappear

Figure 4.12: Residuals against log-linear predictors for the latent process



above this value. The Poisson residuals become approximately normal for log expected counts greater than 1. Taking this into account, Figure 4.11(a) is compatible with the Poisson assumption. In particular, it gives no evidence against the use of the Poisson variance function. The downward trend in Figure 4.11(a) would be apparent in any plot for Poisson residuals against fitted values, making it difficult to interpret some plots. In Figure 4.11(b), one outlier (near the top) corresponding to the extreme value 545* stands out. Nearly all the residuals are clustered in the lower left corner with a few of them scattered around. The residuals with a peculiar pattern in the lower right corner are from patient #446. This plot does not give clear evidence of serious violation of the gamma assumption for the relative burden process. However, it is important to note the peculiar pattern of the residuals from patient #446. The plot of r_t against log fitted values ($\log(b_t m_{t-1})$) presented in Figure 4.12 also has an outlier corresponding to 545*. Moreover, there are a set of residuals in the far right having a peculiar pattern. These

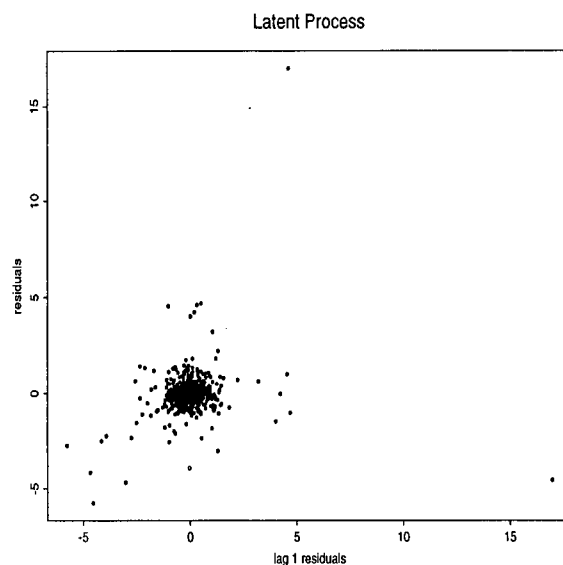
Figure 4.13: Residuals against lag 1 residuals for the two categories



residuals are from patient #446. When these residuals are not taken in consideration, the gamma assumption for the latent process seems to be appropriate. Up to now, the residuals plots all suggest that patient #446 gives somewhat peculiar residuals. The strange pattern of these residuals are observed in various plots and is believed to have caused by the extreme value 446*.

Figure 4.13 shows the plots for R_{1t} against R_{1t-1} and R_{2t} against R_{2t-1} . A band-like structure is detected in Figure 4.13(a). All the bands run from the lower left corner to the upper right corner and are less obvious when away from the center-portion of the plot. These bands are caused by the discreteness of the data. The residuals in the center-most band correspond to two consecutive zero observations. The two bands adjacent to the center-most correspond to either data values of 0 immediately preceded by 1 or data values of 1 immediately preceded by 0. Figure 4.13(a) gives no evidence against the conditional independence assumption of the *active lesions* values over time, given the

Figure 4.14: Residuals against lag 1 residuals for the latent process



latent process. There are two outliers in Figure 4.13(b) which are due to the extreme value at time 6 of patient #545, 545*. It appears that there is a very weak relationship between R_{2t} and R_{2t-1} which can be treated as negligible. As a significant portion of the residuals form a cloud, the perceived weak relationship is caused by a few residuals in the lower left corner of the plot. It is not surprising to see that these residuals are from patient #446 and #545. A similar very weak linear relationship between r_t and r_{t-1} is also found in Figure 4.14. This weak linear relationship is considered negligible using the same arguments discussed above. Moreover, two outliers corresponding to 545* are seen in Figure 4.13.

In conclusion, the fitted model on the full data set is inadequate. There are serious violations of some of the main model assumptions. For example, the conditional independence assumption of the *relative burden* values over time and the Markov assumption

of the latent process are clearly inappropriate. These violations are mainly caused by the extreme observation of *relative burden* at time 17 for patient number 446.

4.4.2 Data set with 446* removed

We now consider the residual analysis for the fitted model on the data set with the extreme value 446* removed. As in the previous sub-section, we will discuss the residual plots from the same four patients as well as the residual plots from the entire data set. We will later confirm that the fitted model is reasonable. We provide the residual plots for the remaining patients in appendix C for reference.

Plots for residuals from patient #446 are presented in Figure 4.15. None of the ACF plots in Figure 4.15(a), (b) and (c) have autocorrelations falling outside the asymptotic 95% confidence bands. No pattern is observed in Figure 4.15(d), (e) and (f). In Figure 4.15(g), the slight downward trend may be of concern. It suggests that the *active lesions* values may not be dependent via the latent process only. However, the band structure of R_{1t} (as discussed in the previous sub-section) may help explain part of this feature. As all the residuals corresponding to data values of 0 lie in the lower band, any slight relationship between R_{1t} and time will be magnified. We will address this issue later by plotting residuals from all patients against time to see if such pattern persists. Figure 4.15(h) and (i) has no particular pattern, confirming the conditional independence assumption of the relative burden and the Markov assumption of the latent process. None of Figure 4.15(j), (k) and (l) reveal any unexpected patterns. Hence, they give no evidence against the distributional assumptions for the observation and latent processes. It comes as no surprise that there are great improvements of the fit for patient #446. When the extreme value 446* is removed, the estimated latent process for patient #446

Figure 4.15: Residuals plots for patient #446 in the placebo group

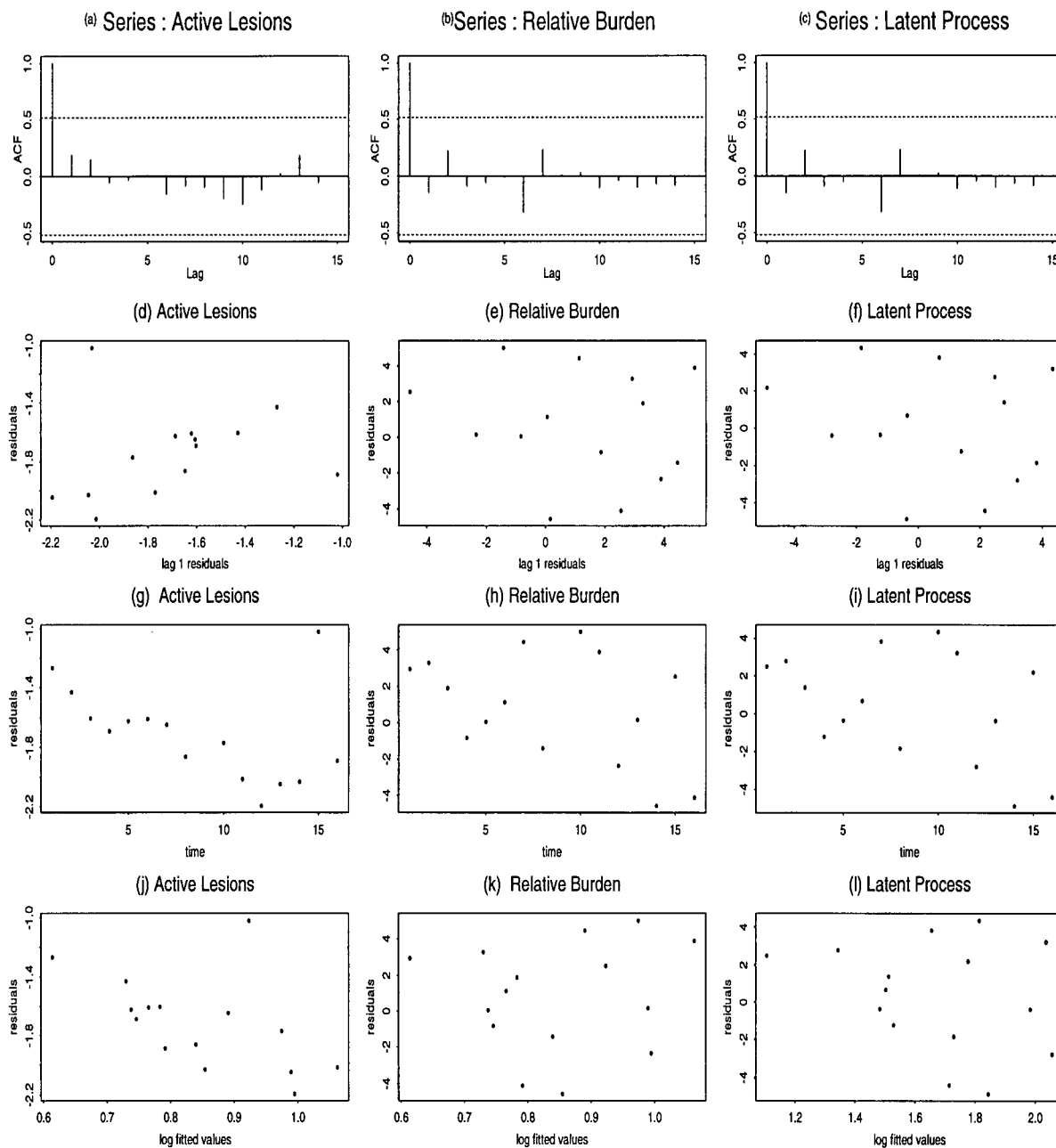


Figure 4.16: Residuals plots for patient #545 in the placebo group

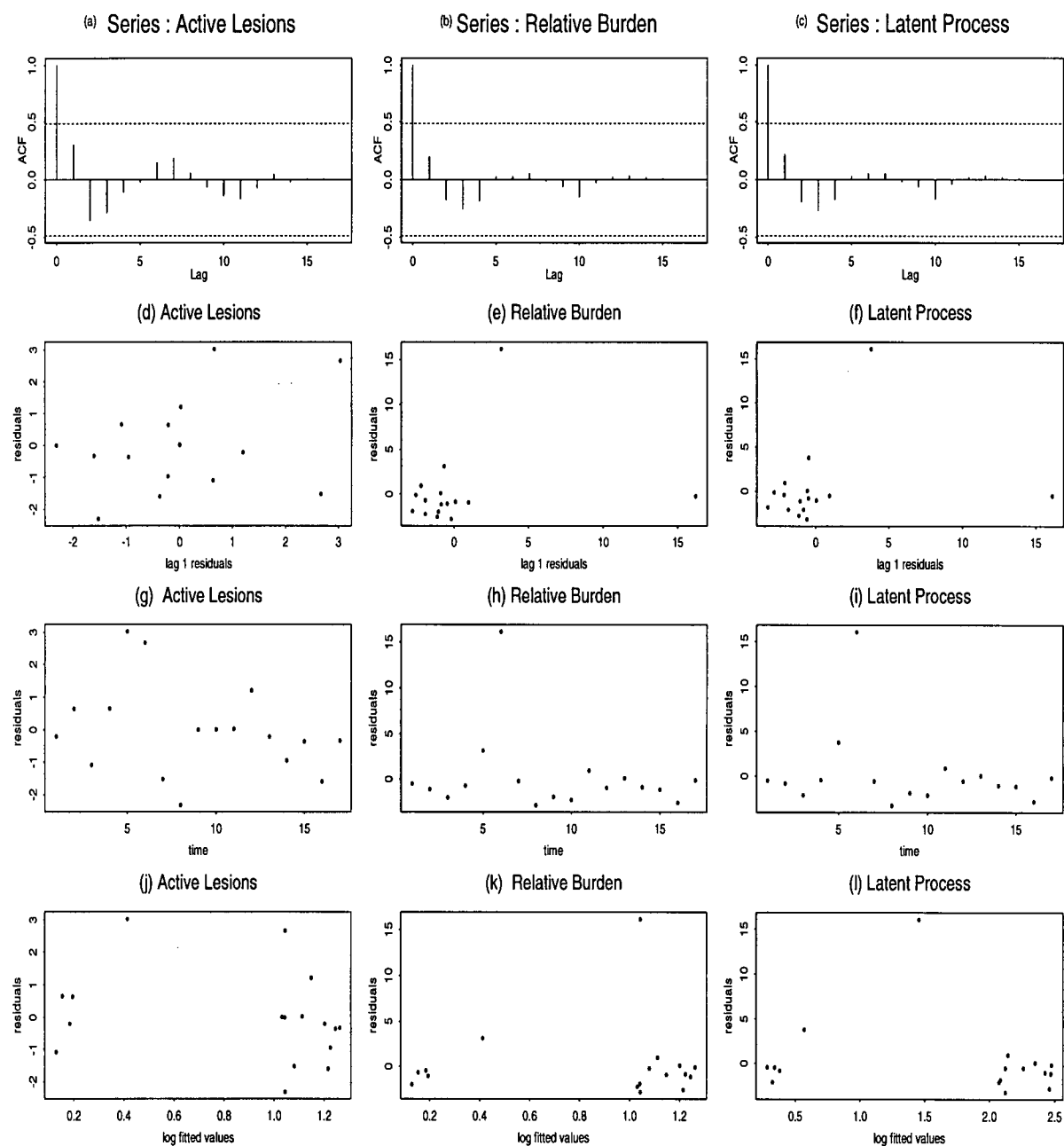


Figure 4.17: Residuals plots for patient #502 in the low dose group

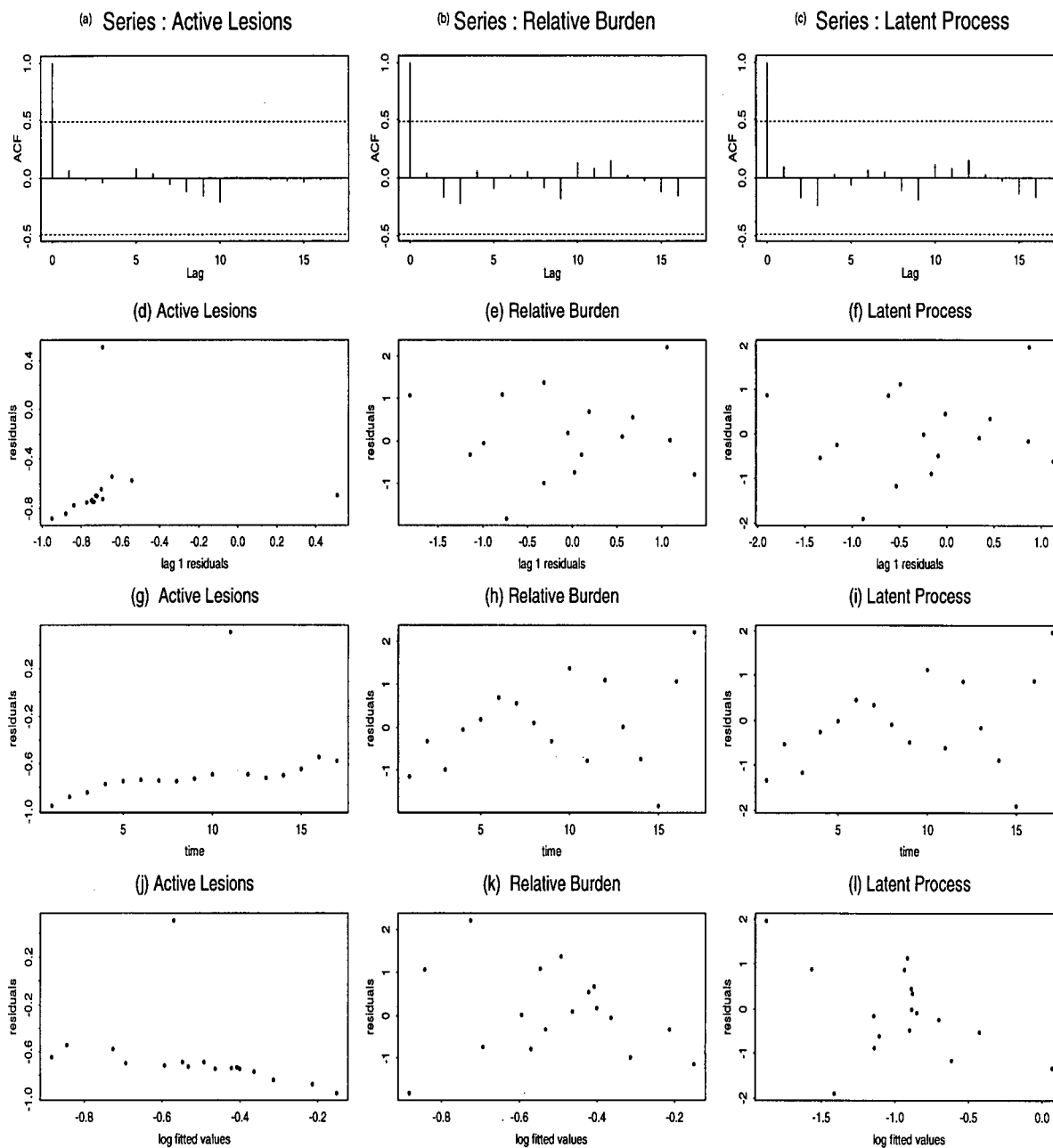
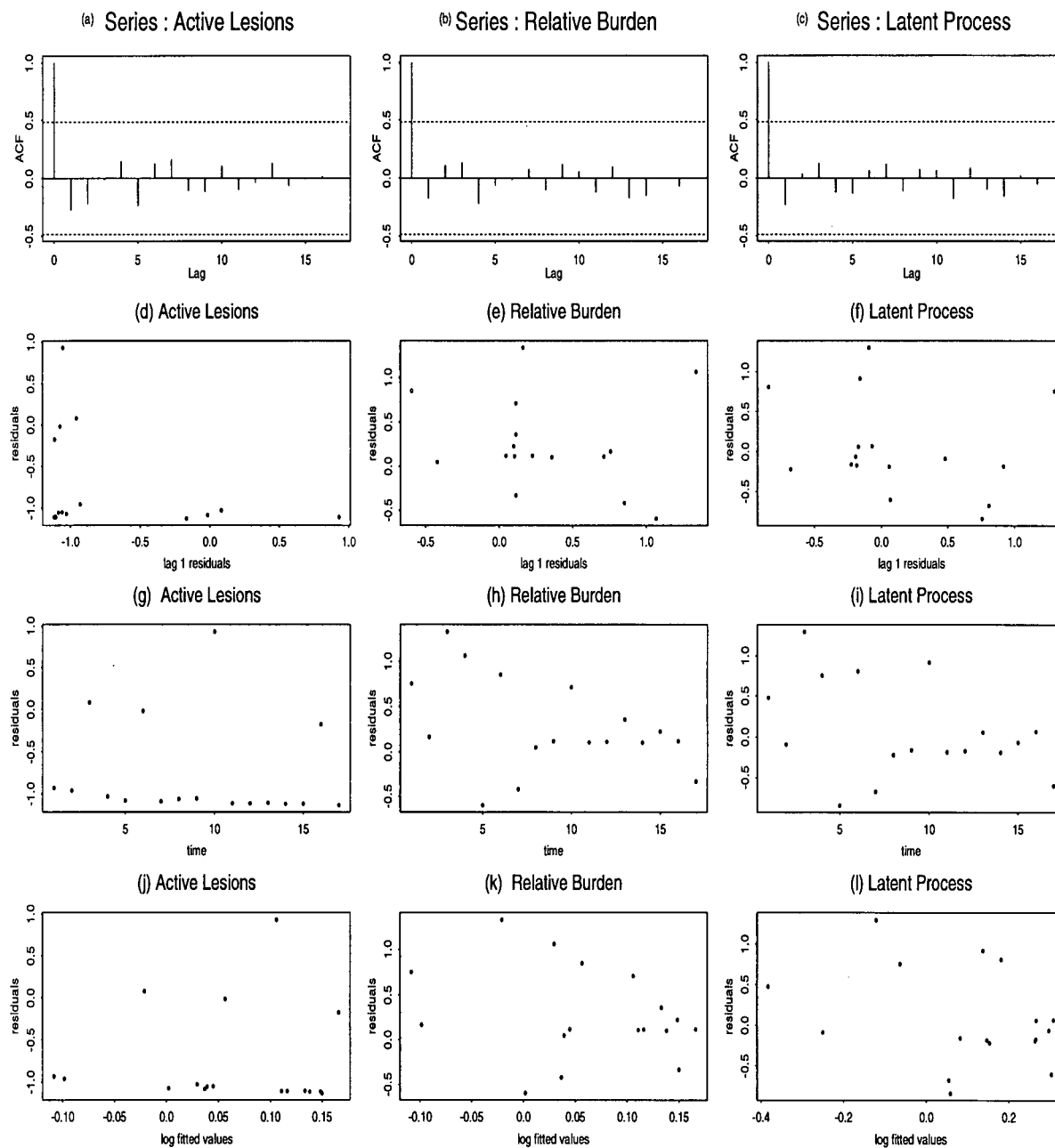


Figure 4.18: Residuals plots for patient #500 in the high dose group



is no longer forced to go up from time 7 onwards as in the fitted model on the full data set.

Plots for residuals from patient #545 are shown in Figure 4.16. Two outliers are seen in Figure 4.16(e) and (f). These two outliers are due to the extreme value 545*. The outliers found in Figure 4.17(h), (i), (k) and (l) are also due to this extreme value. It is not surprising to see the frequent appearance of outliers due to 545* in various plots. Following the procedures on plots in Figure 4.8, there is no evidence against the model assumptions.

Figure 4.17 and Figure 4.18 show the plots for residuals from patient #502 and #500 respectively. Note that the features observed in Figure 4.17(d), (g) and (j) have been discussed when we analyse Figure 4.9. Regarding the features seen in Figure 4.18(d), (g) and (j), we follow similar arguments when we discuss Figure 4.10. Again, no evidence against the main model assumptions are found.

We now analyse residuals from all the patients. Plots for R_{it} against log fitted values are shown in Figure 4.19. The downward banded structure found in Figure 4.19(a) has been explained previously. The Poisson assumption for the *active lesions* values appears to be acceptable. In Figure 4.19(b), there is an outlier corresponding to the extreme value 545* in the top right corner. A slight “megaphone” shape is also observed in Figure 4.19(b); the shape is caused by a very few points to the far right. As most of the residuals form a cloud in the center of the plot, it seems that the gamma variance function assumption is justified. When the outlier is withheld from the plot, the new plot (not shown here) has a similar shape as Figure 4.19(b). In the new plot, only a very few points to the right cause a slight “megaphone” shape. There is no clear evidence against the gamma assumption. Figure 4.20 shows the plot for r_t against log fitted

Figure 4.19: Residuals against log-linear predictors for the two categories

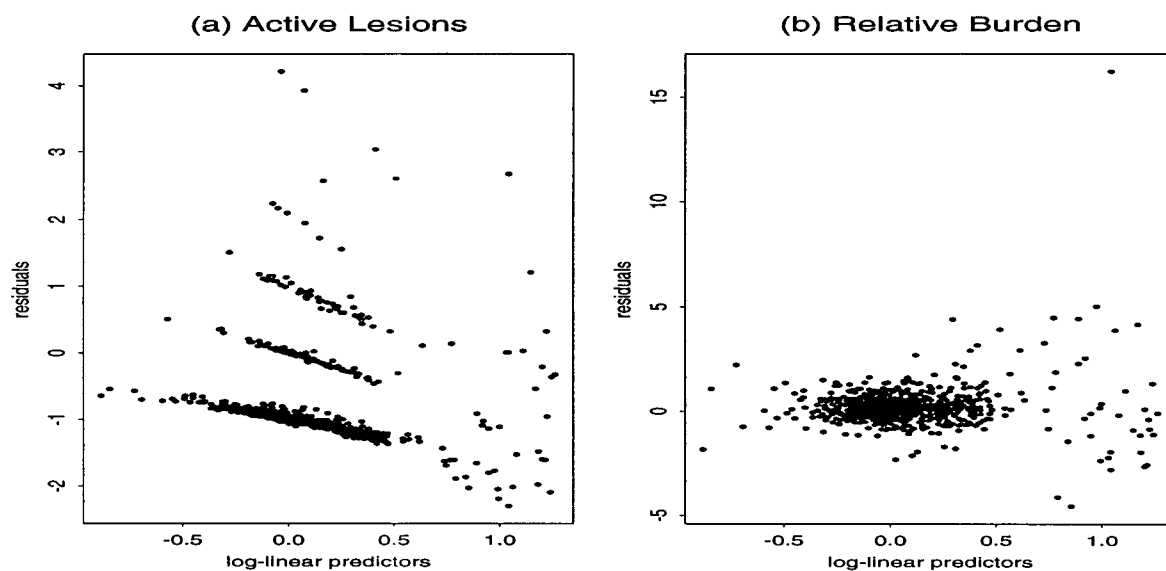


Figure 4.20: Residuals against log-linear predictors for the latent process

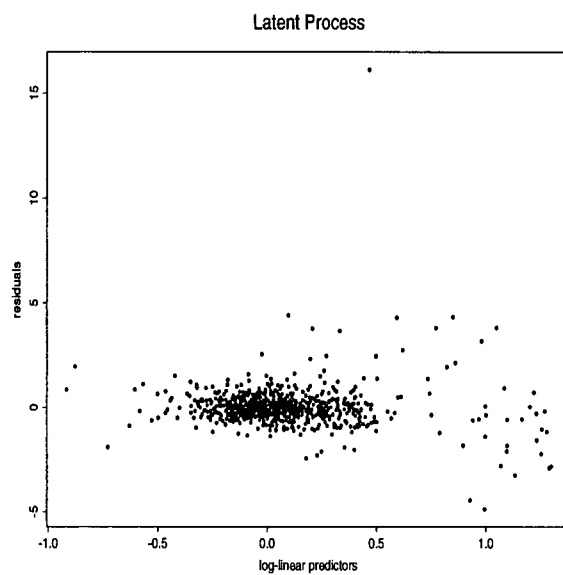
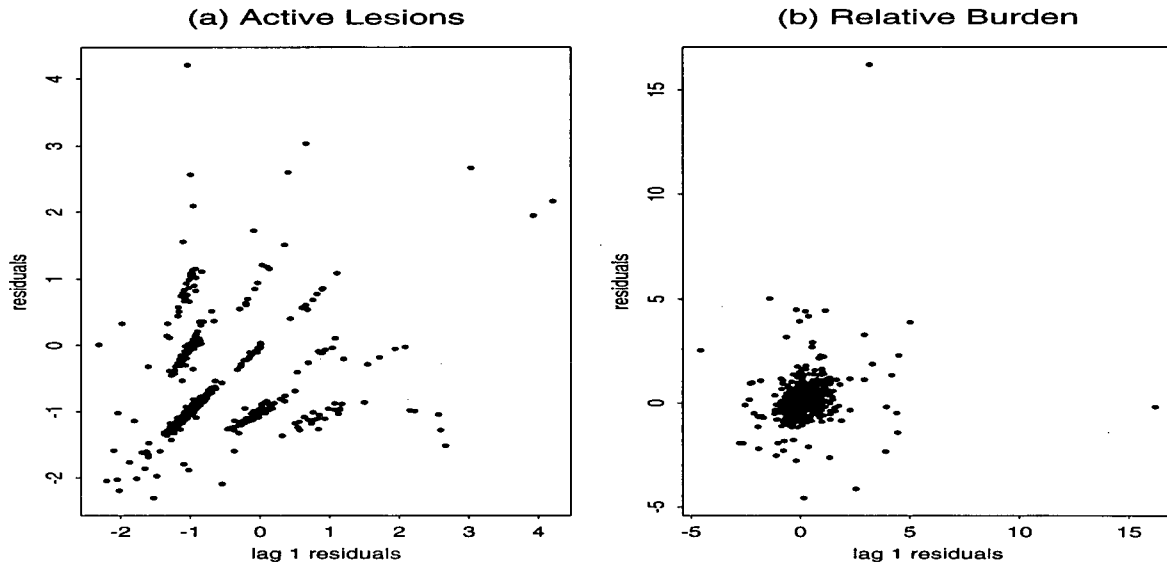


Figure 4.21: Residuals against lag 1 residuals for the two categories



values ($\log(b_t m_{t-1})$). There is an outlier in the upper part; corresponds to 545*. A slight “megaphone” shape caused by a small percentage of residual is also observed. The new plot (not shown here) with the outlier removed also gives a similar shape. Only a very small portion of points cause the perceived shape. Note that there are 804 observation vectors in the data. Following the argument used for Figure 4.19(b), we consider the gamma variance function assumption for the latent process acceptable. Note that the extreme value 545* consistently causes an outlier in some of the plots mentioned in this sub-section and is expected to cause outliers in some of the following plots, too.

The band-like structure for R_{1t} against R_{1t-1} in Figure 4.21(a) has been previously explained. It is due to the discreteness of the data. The plot for R_{2t} against R_{2t-1} in Figure 4.21(b) does not show any particular structure except the presence of two outliers. The plot for r_t against r_{t-1} shown in Figure 4.22 reveals no specific structure; the two outliers are as expected. Plots for R_{it} against time are presented in Figure 4.23. Both

Figure 4.22: Residuals against lag 1 residuals for the latent process

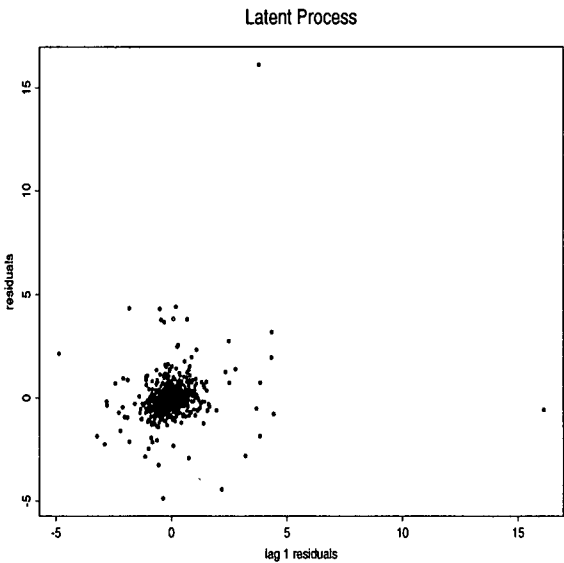


Figure 4.23: Residuals against time for the two categories

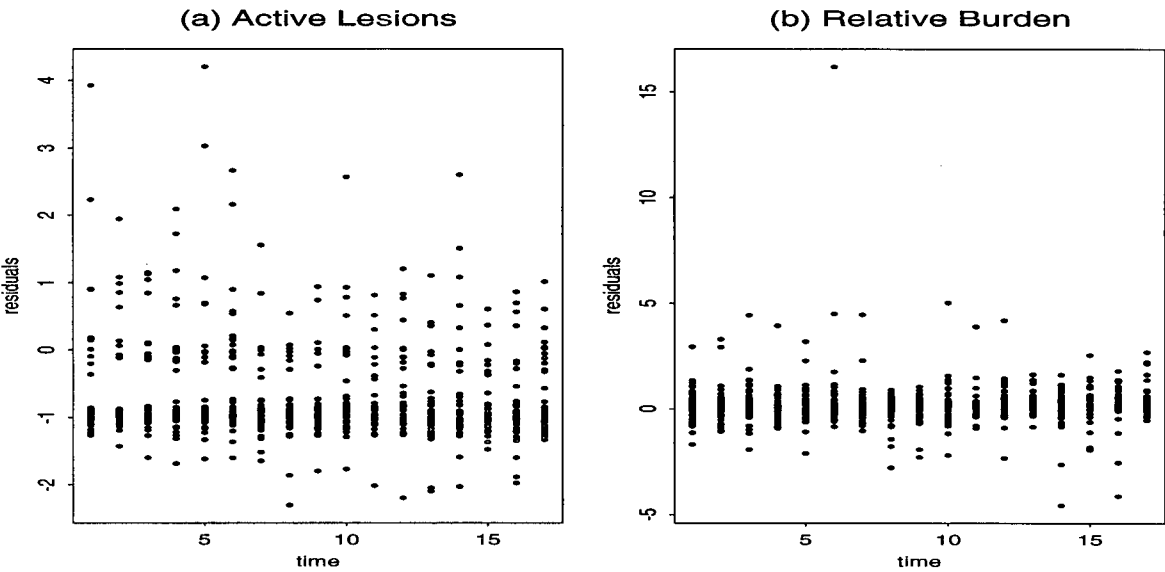
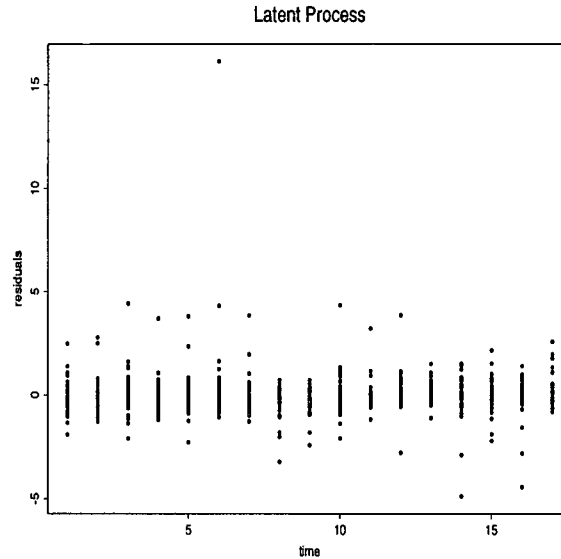


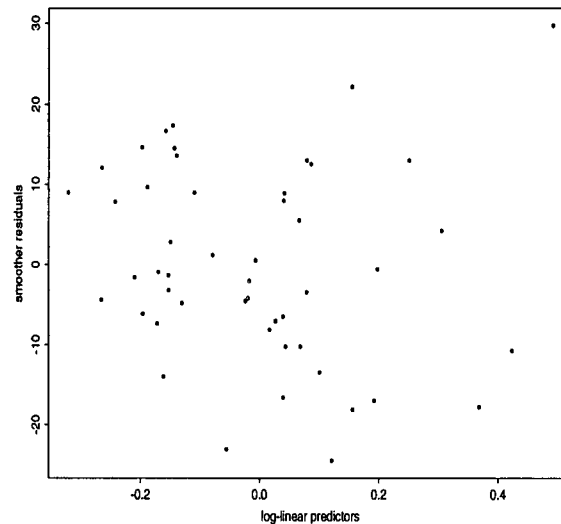
Figure 4.24: Residuals against time for the latent process



plots demonstrate a cloud structure as expected. If *time* is incorrectly classified as a long-term covariate, the residuals R_{it} should show some relationship with *time*. Since these two plots have no particular pattern, there is no evidence against *time* being a long-term covariate. Nothing unexpected is observed in the plot of r_t against time shown in Figure 4.24. Note that the outlier found in both Figure 4.23(b) and Figure 4.24 corresponds to 545*.

We now check the model assumptions for the initial state values. The plot for the smoother residuals $(m_{j0} - g_j)$ against the log fitted values $(\log g_j)$ presented in Figure 4.25 does not reveal any particular pattern. Therefore, the gamma variance function assumption for the random effects distribution is acceptable. The linearity assumption is checked by plotting the smoother residuals $(m_{j0} - g_j)$ against the covariates; as presented in Figure 4.26. Nothing unexpected is seen in Figure 4.26(a), (b), (c) and (f). Note that as

Figure 4.25: Smoother residuals against log-linear predictors for the random effects



mentioned in Section 4.2.3, *EDSS* is an ordinal variable but is treated as a continuous variable in the analysis here for computational convenience. Due to the considerable larger number of patients from BC, it is not entirely clear that the greater range of the residuals for the BC patients seen in Figure 4.26(d) indicates larger variability than residuals for the Washington patients. Moreover, the greater range of the residuals for the female patients observed in Figure 4.26(e) does not clearly indicate greater variability than residuals for male patients because of the substantial larger number of female patients in the study. The plots in Figure 4.26 do not suggest violations of the model assumptions.

Finally, we check the correlations between the two categories by calculating the empirical variance-covariance matrix. This matrix is obtained by first standardizing the vector consisting of R_{it} , $i = 1$ and 2 , by the square-root of its variance-covariance matrix;

Figure 4.26: Smoother residuals against baseline covariates

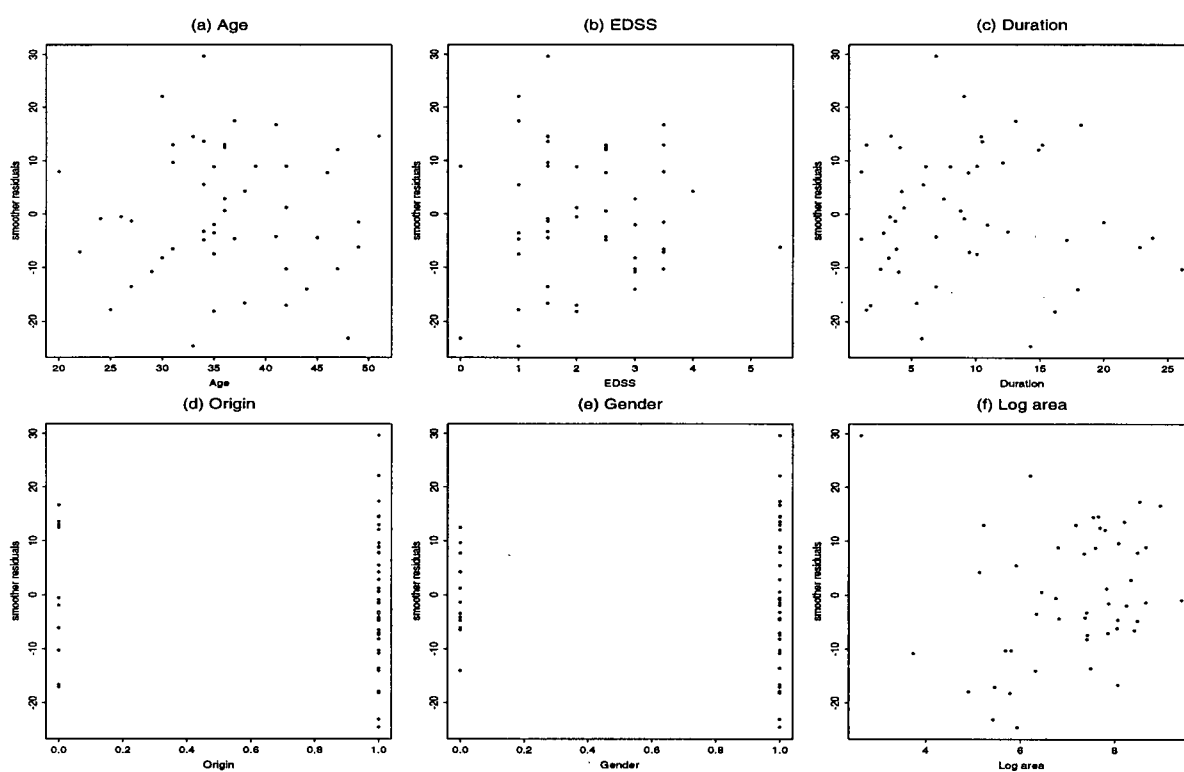


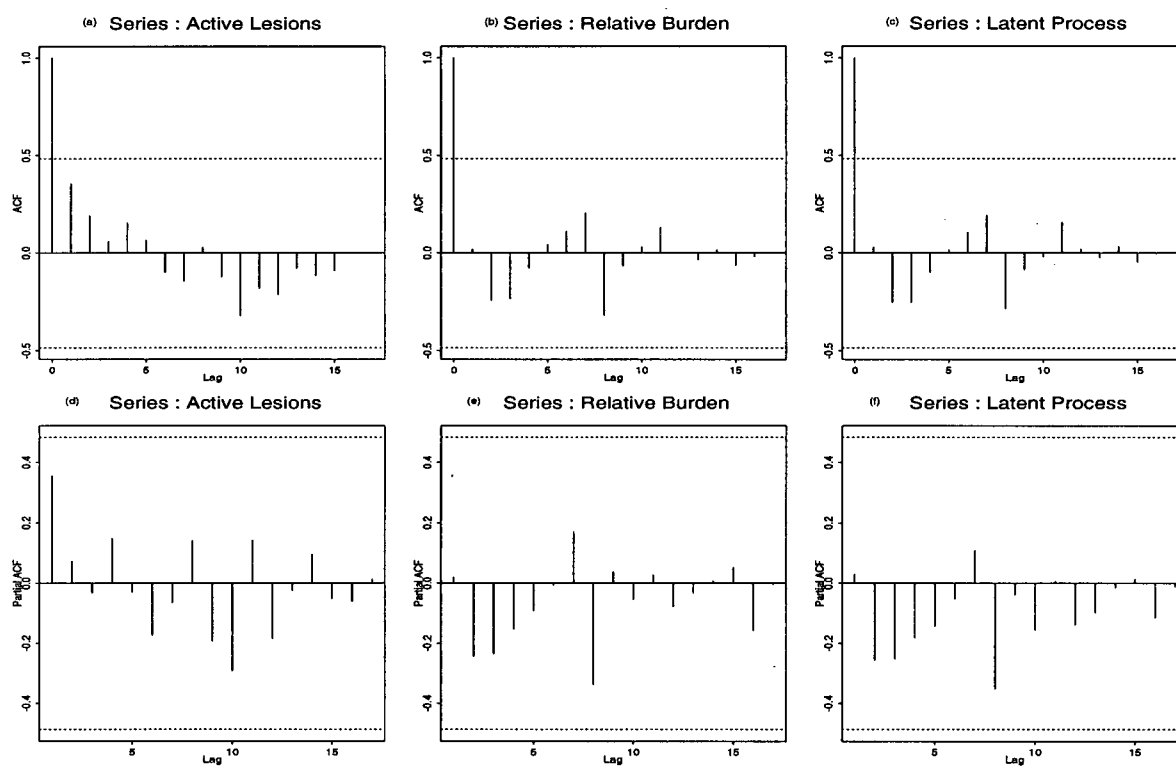
Table 4.9: Empirical variance-covariance matrix for standardized residuals

Category	1	2
1	0.993	-0.152
2	-0.152	0.685

and then calculating the corresponding empirical variance-covariance matrix. The results are given in Table 4.9. Ideally, an identity matrix is expected if the two categories are uncorrelated. The interpretation of Table 4.9 is complicated by the fact that it depends on the choice of the square-root matrix used for the standardization. The matrix in Table 4.9 is much different from the identity matrix. However, such deviation may not suggest serious violation of the conditional independence assumption of the two categories. If we take the asymptotic standard error of the off-diagonal elements of the matrix to be $1/\sqrt{804} = 0.035$, the off-diagonal elements are much larger than predicted by the model. The empirical variance of 0.685 in the second category is much lower than it should be. Although there is evidence that the two categories are slightly conditionally correlated, this does not seem to indicate any serious shortcoming of our model. However, we should keep in mind the effect of this possible correlation between the two categories on the asymptotic standard errors of the regression parameter estimates. These standard errors will be under-estimated if the categories are correlated instead of independent as assumed in our model. None of the ACF and PACF plots for the average filter residuals presented in Figure 4.28 indicate significant correlations.

As a conclusion, the main model assumptions for the fitted model on the data set with 446* removed seems to be appropriate; no serious violations are evident from the data. The model gives a very reasonable fit to this data set.

Figure 4.27: Autocorrelation functions for average filter residuals: (a) to (c) are the autocorrelations functions; (d) to (e) are the partial autocorrelations functions



4.4.3 Data set with 446* and 545* removed

The residual analyses for the fitted model on the data set with 446* and 545* removed have also been carried out. As expected, all the main assumptions for the fitted are confirmed to be appropriate. Due to space limitation, these plots are not included in this thesis.

4.5 The final model

We have now confirmed that the fitted model on the data set with extreme value 446* removed is reasonable. As mentioned previously in Section 4.3, we will retain as many data points as possible in the data for analysis provided that the fit is acceptable. Therefore, we will start our search for the final model with the fitted model on the data set with 446* removed.

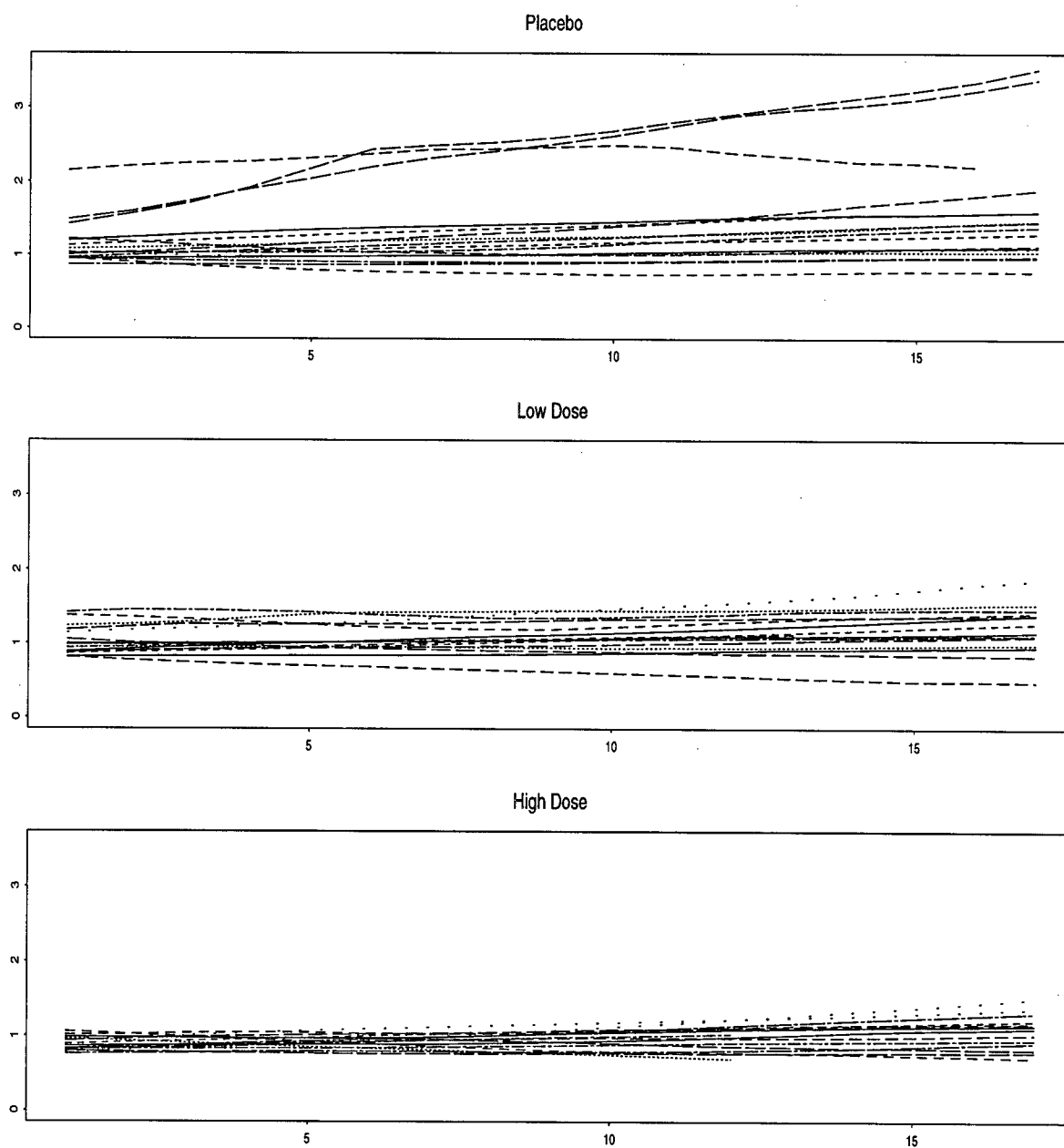
Since the latent process may be interpreted as the underlying severity of the disease, the effects for *time*, the long-term covariate, is discussed first. The regression parameter for *time* may be interpreted as the rate of increase of the log of the underlying severity ($\log \theta_t$) per 6-week period. A positive estimate means that the underlying severity of MS is getting worse over the period of this study. Our model allows each patient to have a different regression parameter. There are all together 49 estimates for the long-term covariate effects, one for each patient. These estimates are summarized in Table 4.6 on P.44, Table 4.7 on P.45 and Table 4.8 on P.46, with asymptotic standard errors of the estimates shown in brackets.

In our model, the latent process is estimated by the Kalman smoother. The estimated latent processes for each patient presented in Figure 4.28 can be used to guide our discussion on the pattern of the response over time. It appears that these estimated processes are not so variable. The estimate for the dispersion parameter ($\hat{\sigma}^2 = 0.00038$) shows that the stochastic variation of the latent morbidity process θ_t itself is small. Among the three treatment groups, patients in the placebo group seems to have the greatest upward trend. The estimated latent processes of three patients (#449, #507 and #545) are very different from those for other patients in the placebo group. For the low dose group, nearly all the estimated processes are mostly horizontal except that for patient #502. This patient has an estimated latent process with a relatively large negative slope.

Since one of the objectives of this study is the pattern of the response over time, it is desirable to see whether patients in a given treatment group share the same regression parameters and then compare these common effects across treatment groups. The test for patients in the placebo group having the same regression parameter has the Wald's statistic (as mentioned in section 3.3.1.) $W = 27.31$, which gives the p-value of 0.038 compared with a $\chi^2(16)$ -distribution. That is, there is weak evidence that the long-term covariate effects are not the same. The same test for the low dose group gives $W = 12.64$ with 15 degrees of freedom; the p-value is 0.630. Therefore, there is no evidence against the assumption that patients in the low dose group share the same regression parameter. The test applied to the patients in the high dose group has $W=13.68$ with 15 degrees of freedom and the p-value is 0.550. The test suggests that there are no significant differences of the long-term covariate effects among patients in the high dose group.

We now discuss the effects for the baseline covariates. The estimates for the regression

Figure 4.28: Estimated latent processes for each patient



parameters are summarized in Table 4.5 on P.43. The Wald's test for the joint significance of *low dose* and *high dose* has $W = 6.27$ with 2 degrees of freedom, which gives the p-value of 0.043. Hence, the data give weak evidence that the effects for the treatment groups are significant. Since there are significant treatment effects, it is appropriate to compare the treatment groups. The effect for the low dose group is mildly significant (p-value of 0.073) and that for the high dose group is highly significant (p-value of 0.016). No significant difference between the low dose and the high dose group (p-value of 0.430) is found. That is, the log of the expected initial value for patients in the placebo group is 0.118 higher than those for patients in the low dose group and 0.173 higher than those for patients in the high dose group. Although the estimated effect for the high dose group is 0.055 less than that for the low dose group, such a difference should be not taken seriously as it is not statistically significant.

The other baseline covariates found to have significant effects are *age* (p-value of 0.012) and *log area* (p-value of 0.000). For every unit increase in *age* and *log area*, the log of the expected initial value would be reduced by 0.009 and 0.095 respectively. That is, the analysis suggests that older patients with larger values of initial burden of disease seem to have smaller initial values for the latent process.

We now consider the case where patients in a given treatment group share the same regression parameter for the long-term covariate. Although the data give some evidence that patients in the placebo group do not share a common time effect, it will be interesting to discuss the results under this assumption as one of the objective of this MRI sub-study is to know more about the "average" effects of Betaseron on patients. These common time effects may be interpreted as the "average" time effects in a treatment group, thus lending themselves to address the objective of the sub-study. The estimates for the

Table 4.10: Estimates for the long-term covariates effects

Parameter	estimate	standard error	z
β_{pl}	0.021	0.0055	3.76
β_{ld}	0.011	0.0063	1.67
β_{hd}	0.011	0.0068	1.65

Table 4.11: Estimates for the baseline covariates effects

Parameter	estimate	standard error	z
<i>intercept</i>	0.861	0.299	2.88
<i>low dose</i>	-0.075	0.085	-0.88
<i>high dose</i>	-0.163	0.091	-1.79
<i>age</i>	-0.008	0.005	1.60
<i>EDSS</i>	0.049	0.033	1.48
<i>duration</i>	-0.005	0.006	0.83
<i>origin</i>	-0.049	0.084	0.58
<i>gender</i>	0.041	0.083	0.49
<i>log area</i>	-0.080	0.022	3.64

effects are summarized in Table 4.10 and Table 4.11. We denote the common regression parameters by β_{pl} (in the placebo group), β_{ld} (in the low dose group) and β_{hd} (in the high dose group).

The test for equality of the β_{pl} , β_{ld} and β_{hd} gives the Wald's statistic $W = 1.79$ with 2 degrees of freedom (p-value of 0.410). Note that neither β_{ld} nor β_{hd} are significantly different from zero. It suggests that the time effects across the treatment groups are the same. Moreover, only two baseline covariates are found to have significant effects. The effect for *high dose* is marginally significant (p-value of 0.073) and for *log area* is highly significant (p-value of 0.000). The effects for *low dose* and *age* are no longer significant. Note that, in general, the magnitude of the estimates are smaller and the standard errors

are larger than those for the full model. For example, the estimate for the effect for the high dose group is -0.163 versus -0.173 in the full model.

The Tweedie models possess an advantage that subject-specific effects can be investigated. This advantage is fully utilized to extract more valuable information from the data set. First, the time effects for patients are allowed to be different. By investigating these time effects, information on how individual patients respond to the treatment can be extracted. There is some evidence that the time effects for patients in the placebo group are somewhat different. If the objective is on studying the “average” time effects across the treatment groups, the analysis can be carried out with the assumption that patients in a treatment group share a common time effect. The common time effect may be interpreted as the “average” effect for patients in a treatment group.

To conclude, we start with the full model and then reduce it to a more parsimonious one. The final model identified is:

Long-term covariate:

time.

Baseline covariates:

low dose;

high dose;

age;

log area.

The estimates for this final model is presented in Table 4.12 and Table 4.13. The Wald’s test for common time effects for patients in a treatment group is carried out. For

Table 4.12: Estimates and standard errors for the time effects for each patient; placebo (PL), low dose (LD) and high dose (HD).

PL	estimate (s.e.)	LD	estimate (s.e.)	HD	estimate (s.e.)
#420	0.010 (0.011)	#419	0.021 (0.011)	#422	0.011 (0.011)
#421	0.014 (0.010)	#424	0.001 (0.010)	#444	-0.007 (0.012)
#443	0.013 (0.010)	#448	0.015 (0.012)	#445	0.029 (0.011)
#446	0.003 (0.010)	#450	0.010 (0.012)	#453	0.005 (0.011)
#449	0.054 (0.008)	#452	-0.003 (0.011)	#454	0.040 (0.011)
#451	0.002 (0.011)	#499	0.021 (0.011)	#497	-0.021 (0.013)
#498	0.021 (0.010)	#502	-0.041 (0.013)	#500	0.012 (0.011)
#501	0.001 (0.010)	#508	0.012 (0.011)	#503	0.013 (0.011)
#504	-0.021 (0.011)	#521	0.013 (0.011)	#506	-0.026 (0.019)
#507	0.039 (0.010)	#525	0.005 (0.010)	#524	0.023 (0.012)
#522	0.022 (0.010)	#542	0.027 (0.010)	#526	0.012 (0.012)
#523	0.007 (0.011)	#544	0.016 (0.015)	#541	-0.011 (0.012)
#539	0.025 (0.013)	#547	-0.011 (0.011)	#543	0.002 (0.012)
#540	-0.009 (0.010)	#548	0.010 (0.011)	#546	0.030 (0.017)
#545	0.053 (0.008)	#546	0.017 (0.010)	#549	0.024 (0.011)
#550	0.005 (0.011)	#568	0.023 (0.016)	#566	0.013 (0.011)
#565	0.024 (0.011)				

Table 4.13: Estimates and standard errors for baseline covariate effects

Parameter	estimate (s.e.)
<i>intercept</i>	1.123 (0.190)
<i>low dose</i>	-0.074 (0.063)
<i>high dose</i>	-0.146 (0.068)
<i>age</i>	-0.010 (0.004)
<i>log area</i>	-0.097 (0.019)

the placebo group, the test statistic is $W = 74.71$, which gives a p-value of 0.000 when compared with a $\chi^2(16)$ -distribution. For patients in the low dose group, the p-value is 0.036 ($W = 26.20$ with 15 degrees of freedom). As for the high dose group, the test gives a p-value of 0.011 ($W = 30.14$ with 15 degrees of freedom). Hence, there is clear evidence that the time effects are different for patients in a treatment group. These tests all have smaller p-values as compared to the same tests in the full model. This may be due to the smaller standard errors for the time effects in the final model. Note that in the full model, the same test for common time effects leads to test statistics of $W = 27.31$ (for the placebo group), $W = 12.64$ (for the low dose group) and $W = 13.68$ (for the high dose group).

Finally, since there is marginal significant treatment effects (p-value of 0.089), it is quite reasonable to compare the treatment groups. As seen in Table 4.13, only the high dose group significantly (p-value of 0.032) influences the initial values of the underlying severity; the effect of the low dose group is not significant (p-value of 0.240).

Chapter 5

Discussion

It is helpful to use the results from the previous analyses of the data by Petkau and White (1995) and D'yachkava (1997) for reference. Comparing their results with those obtained in this thesis may not be so appropriate as the details in the previous analyses are quite different. Petkau and White (1995) used an univariate approach and considered a similar form of transformation of the response variable *burden of disease*. The transformation they considered was the log of *relative burden*. They found significant treatment effects and differences among the linear time effects for the different treatment group. D'yachkava (1997) used an univariate approach and looked at *active lesions*. She found significant treatment effects and no differences among the linear time effects for the different treatment group. Both Petkau and White (1995) and D'yachkava (1997) used the GEE approach in their analyses and focused on the sub-population (such as patients in a treatment group) average time effects but not subject-specific effects.

We use a multivariate approach in our analysis and considered a bivariate response vector with *relative burden* and *active lesions* as its components. The possible effects of the treatment are addressed by investigating the time effects for patients across treatment group. Since the latent process reflects the underlying severity of the disease, a large positive time effect for a patient suggests that the disease is getting worse over time. A time effect close to zero suggests that the disease does not change much during the sub-study. This may indicate a benefit of the treatment. The constant covariate effects

influence the initial value of the time series for a given patient, thus accounting for the natural heterogeneity of the patients. These constant covariate effects do not have the same interpretations as in the previous analyses.

Regarding the pattern of the response over time, our model allows the time effects to be different across patients. The equality of these effects in a given treatment group can then be tested. There is clear evidence that patients in a treatment group do not share a common time effect. The estimated latent processes shown in Figure 4.28 indicate that patients in the placebo group have increasing severity of disease while those in the treatment group do not have so much change in disease severity. When the sub-population “average” time effects are of focus, as seen in Table 4.10, the time effects for the treatment group are smaller than that for the placebo group, perhaps even zero, suggesting a clear benefit of the treatment.

In the analysis presented in this thesis, we avoid a potential computational complication by treating an ordinal variable (*EDSS*) simply as a continuous variable. Moreover, only one long-term covariate (*time*) is considered here. It may be interesting to include other time-varying covariates present in the MRI sub-study, for example, the EDSS scores at the time of subsequent scans, in future analysis. Our models currently can only handle longitudinal data collected at equally spaced time intervals. Due to this methodological constraint, we consider each scan to be taken at the target date for which it was intended, irrespective of the deviation of the actual date of the scan from that target date. Further development of the new method remains to be conducted in order to handle categorical (binary or ordinal) response data and to accommodate longitudinal data collected at unequally spaced intervals.

Despite having some methodological constraints, the new method possesses various attractive features in analyzing multivariate longitudinal data. The Tweedie state space models provide a unified approach to analysis of multivariate longitudinal data of mixed types. Since the Tweedie class exponential dispersion models cover a wide range of discrete and continuous distributions, the new method can handle a variety of response vectors with components of different data types. It offers an intuitively appealing conceptual framework for the data-generating mechanism of longitudinal data. A realistic correlation structure is achieved under the model assumptions. In contrast to the quasi-likelihood approach, the method is based on the specification of the joint likelihood of the longitudinal data, thereby facilitating a wide variety of inferences and diagnostic procedures. The new method is flexible and easy to use. It has an extra feature that the covariates can enter the model via the short-term observation process or via the long-term latent process. At present, the new methodology cannot handle binary or categorical data. The method needs further development in order to accommodate these types of data. One limitation is that it is a new methodology. The justification of its use is based on asymptotics. Its properties are not yet fully understood. It would be desirable to provide a better understanding of the new methodology through further theoretical investigations and simulation studies.

Bibliography

- [1] Azzalini, A. (1982). Approximate filtering of parameter driven processes. *Journal of Time Series Analysis*, **3**, 38–44.
- [2] Chan, K.S. and Ledolter, J. (1995). Monte Carlo EM estimation for time series models involving observations. *Journal of American Statistical Association*, **90**, 242–252.
- [3] D'yachkova, Y. (1997). *Analysis of longitudinal data from the Betaseron multiple sclerosis clinical trial*. M.Sc. Thesis, Department of Statistics, University of British Columbia.
- [4] Diggle, P.J., Liang, K.-Y. and Zeger, S.L. (1994). *The analysis of Longitudinal Data*. Oxford: Oxford University Press.
- [5] Durbin, J. (1990). Extensions of Kalman modelling to non-Gaussian observations. *Quaderni di Statistica e Matematica Applicata alle Scienze Economico-Sociali*, **12**, 3–12.
- [6] Fahrmeir, L. and Kaufmann, H. (1991). On Kalman filtering, posterior mode estimation and Fisher scoring in dynamic exponential family regression. *Metrika*, **38**, 37–60.
- [7] Fahrmeir, L. and Tutz, G. (1994). *Multivariate Statistical Modelling Based on Generalized Linear Models*. New York: Springer-Verlag.
- [8] Harvey, A.C. (1981). *Time Series Models*. Oxford: Allan.
- [9] Harvey, A.C. and Fernandes, C. (1989). Time series models for counts or qualitative observations (with discussion). *Journal of Business Economic Statistics*, **7**, 407–422.
- [10] The IFNB Multiple Sclerosis Study Group (1993). Interferon beta-1b is effective in relapsing-remitting multiple sclerosis: clinical results of a multicenter, randomized, double-blind, placebo-controlled trial. *Neurology*, **43**, 655–661.
- [11] Jones, R.H. (1993). *Longitudinal Data with Serial Correlation: A State-Space Approach*. London: Chapman and Hall.
- [12] Jørgensen, B. (1986). Some properties of exponential dispersion models. *Scandinavian Journal of Statistics*, **13**, 187–197.

- [13] Jørgensen, B. (1987). Exponential dispersion models (with discussion). *Journal of Royal Statistical Society B*, **49**, 127–162.
- [14] Jørgensen, B., Lundbye-Christensen, S., Song, X.-K. and Sun, L. (1995a). A state space model for multivariate longitudinal count data. *Technical Report #148*, Department of Statistics, University of British Columbia.
- [15] Jørgensen, B., Lundbye-Christensen, S., Song, X.-K. and Sun, L. (1995b). A longitudinal study of emergency room visits and air pollution for Prince George, British Columbia. *Biostatistics Research Report #6*, Biostatistics Research Group, University of British Columbia. To appear in *Statistics in Medicine*.
- [16] Jørgensen, B., Lundbye-Christensen, S., Song, X.-K. and Sun, L. (1995c). State space models for multivariate longitudinal data of mixed types. *Canadian Journal of Statistics*, **24**, 385–402.
- [17] Jørgensen, B., Martinez, J.R. and Tsao, M. (1994). Asymptotic behaviour of the variance function. *Scandinavian Journal of Statistics*, **21**, 223–243.
- [18] Korn, E.L. and Whittemore, A.S. (1979). Methods for analyzing panel studies of acute health effects of air pollution. *Biometrics*, **35**, 795–802.
- [19] Kaufmann, H. (1987). Regression models for nonstationary categorical time series: asymptotic estimation theory. *Annals of Statistics*, **15**, 863–871.
- [20] Laird, N.M. and Ware, J.H. (1982). Random-effects models for longitudinal data. *Biometrics*, **38**, 963–974.
- [21] Liang, K.-Y. and Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13–22.
- [22] Liang, K.-Y., Zeger, S.L. and Qaqish, B. (1992). Multivariate regression analyses for categorical data (with Discussion). *Journal of Royal Statistical Society, B* **54**, 3–40.
- [23] Paty, D.W., Li, D.K.B., the UBC MS/MRI Study Group, and the IFNB Multiple Sclerosis Study Group (1993). Interferon beta-1b is effective in relapsing-remitting multiple sclerosis. II. MRI analysis results of a multicenter, randomized, double-blind, placebo-controlled trial. *Neurology*, **43**, 662–667.
- [24] Petkau, J. and White R. (1995). Longitudinal analyses for the UBC 6-weekly frequent MRI sub-study of the Betaseron multiple sclerosis clinical trial. *Biostatistics Research Report #10*, Biostatistics Research Group, University of British Columbia.
- [25] Shephard, N. (1994). Partial non-Gaussian state space. *Biometrika*, **81**, 115–131.

- [26] West, M., Harrison, P.J. and Migon, H.S. (1985). Dynamic generalized linear models and Bayesian forecasting. *Journal of American Statistical Association*, **80**, 73–79.
- [27] Zeger, S.L., Liang, K.-Y. and Self, S.G. (1985). The analysis of binary longitudinal data with time-independent covariates. *Biometrika*, **72**, 31–38.
- [28] Zeger, S.L. and Karim, M.R. (1991). Generalized linear models with random effects: a Gibbs sampling approach. *Journal of American Statistical Association*, **86**, 79–86.
- [29] Zeger, S.L. (1988). A regression model for time series of counts. *Biometrika*, **75**, 621–629.
- [30] Zeger, S.L. and Liang, K.-Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, **42**, 121–130.
- [31] Zeger, S.L. and Qaqish, B. (1988). Markov regression models for time series: a quasi-likelihood approach. *Biometrics*, **44**, 1019–1031.

Appendix A

Exponential dispersion models

The following is a review of the basic properties of exponential dispersion models from Jørgensen (1986, 1987). An exponential dispersion model for a random variable \mathbf{Y} is given by the probability density function

$$p(y; \theta, \lambda) = c^*(y; \lambda) \exp\{\theta y - \lambda \kappa(\theta)\} \quad (\text{A.19})$$

for $y \in \mathbf{R}$, with respect to a suitable measure, and for appropriate functions c^* and κ . The distribution (A.19) is denoted by $Y \sim \text{ED}^*(\theta, \lambda)$, and θ is called the *canonical parameter* and λ the *index parameter*. The maximum possible domain for θ , denoted by $\Theta \subseteq \mathbf{R}$, is called the *canonical parameter domain*. The domain for λ , called the *index set*, is denoted by Λ , and is a subset of \mathbf{R}_+ . We have $\Lambda = \mathbf{R}_+$ if and only if (A.19) is infinitely divisible (Jørgensen 1986). The function κ is called the *cumulant generator* of the model. The mean and variance of Y are

$$\text{E}(Y) = \mu = \lambda \tau(\theta) \quad \text{and} \quad \text{Var}(Y) = \lambda V(\theta/\lambda), \quad (\text{A.20})$$

where

$$\tau(\theta) = \kappa'(\theta)$$

is the mean-value mapping and

$$V(\mu) = \kappa''(\tau^{-1}(\mu))$$

is the unit variance function on $\Omega = \tau(\text{int}\Theta)$.

Table A.14: Summary of the Tweedie exponential dispersion models

Distribution	p	S	Ω	Θ
Extreme stable	$p < 0$	\mathbf{R}	\mathbf{R}_+	\mathbf{R}_0
Normal	$p = 0$	\mathbf{R}	\mathbf{R}	\mathbf{R}
{Do not exist}	$0 < p < 1$	—	—	—
Poisson	$p = 1$	\mathbf{N}_0	\mathbf{R}_+	\mathbf{R}
Compound Poisson	$1 < p < 2$	\mathbf{R}_0	\mathbf{R}_+	\mathbf{R}_-
Gamma	$p = 2$	\mathbf{R}_+	\mathbf{R}_+	\mathbf{R}_-
Positive stable	$2 < p < 3$	\mathbf{R}_+	\mathbf{R}_+	$-\mathbf{R}_0$
Inverse Gaussian	$p = 3$	\mathbf{R}_+	\mathbf{R}_+	$-\mathbf{R}_0$
Positive stable	$p > 3$	\mathbf{R}_+	\mathbf{R}_+	$-\mathbf{R}_0$

Notation: S = support, $-\mathbf{R}_0 = (-\infty, 0]$.

An exponential dispersion model satisfies the convolution formula

$$\text{ED}^*(\theta, \lambda_1) \star \text{ED}^*(\theta, \lambda_2) = \text{ED}^*(\theta, \lambda_1 + \lambda_2). \quad (\text{A.21})$$

The Tweedie class is an important class of univariate exponential dispersion models, defined as the special case of the distribution (A.19) with unit-variance functions of the form

$$V(\mu) = \mu^p$$

for some $p \in (-\infty] \cup [1, \infty)$, denoted by $\text{Tw}_p(\mu, \sigma^2)$, where $\sigma^2 = \lambda^{1-p}$ is called the *dispersion parameter*; the variance is $\sigma^2 \mu^p$. The Tweedie models may be characterized as the only exponential dispersion models that satisfy the following scale transformation property:

$$c\text{Tw}_p(\mu, \sigma^2) \sim \text{Tw}_p(c\mu, c^{2-p}\sigma^2) \quad \forall c > 0. \quad (\text{A.22})$$

Table A.14 shows a summary of the Tweedie class. The case $p = 1$, $\sigma^2 = 1$ gives the Poisson distribution, and $1 < p < 2$ gives the compound Poisson distributions [in the terminology of, for example, Feller (1971)], which have a positive probability mass at zero and are continuous for $y > 0$. The cases $p = 0, 2, 3$ correspond to the normal,

gamma and inverse Gaussian distributions, respectively.

For $p \geq 2$ or $p \leq 0$, the Tweedie models correspond to the exponential families generated by the positive stable distributions and the extreme stable distributions, respectively. We refer to these models as the stable Tweedie models. These models are not stable as such, but all the Tweedie models appear as the limiting distributions for the exponential dispersion models according to the convergence theorem by Jørgensen, Martínez and Tsao (1994), which extends the stable generalization of the central limit theorem. Many exponential dispersion models have variance functions that are asymptotically of the Tweedie form, and may be approximated by a member of the Tweedie family by means of the result of Jørgensen, Martínez and Tsao (1994). For this reason, the Tweedie models occupy a central position among exponential dispersion models.

The density function of the Tweedie class has the following form:

$$f(y; \mu, \sigma^2, p) = c_p(y; \sigma^2) \exp\left[\frac{1}{\sigma^2} \{y\tau_p^{-1}(\mu) - \kappa_p(\tau_p^{-1}(\mu))\}\right], \quad (\text{A.23})$$

where

$$\kappa_p(\theta) = \alpha^{-1}(\alpha - 1)\{\theta/(\alpha - 1)\}^\alpha$$

and

$$\tau_p(\theta) = \{\theta/(\alpha - 1)\}^{\alpha-1}, \text{ with } \alpha = (p - 2)/(p - 1).$$

Appendix B

Godambe information matrix

In this appendix, the calculation of the sensitivity matrix \mathbf{S} and the variability matrix \mathbf{V} mentioned in section 3.3.1 is briefly outlined. These matrices enter in the Godambe information matrix $\mathbf{J} = \mathbf{S}^\top \mathbf{V}^{-1} \mathbf{S}$ and in the Newton scoring algorithm. The first two sections contain some intermediate results and the final results are given in the last two sections.

B.1 Moment calculations

The variance-covariance matrices of $\hat{\boldsymbol{\varepsilon}}_t^*$, $\hat{\xi}_t^*$ and m_0^* needed for calculating the variability matrix \mathbf{V} are obtained as follows. We consider one series indexed by j at a time. For convenience, we suppress the index j in the notation. Let the matrix \mathbf{C}^* be defined by

$$\mathbf{C}^* = \mathbf{E}\{(\boldsymbol{\theta} - \mathbf{m}^*)(\boldsymbol{\theta} - \mathbf{m}^*)^\top\}$$

with diagonal elements C_t^* and off-diagonal elements C_{tt+s}^* given by

$$C_{tt+s}^* = C_{t+s}^* \prod_{i=1}^s \frac{C_{t+i-1}}{D_{t+i}}.$$

Since $\hat{\boldsymbol{\varepsilon}}_t^* = \boldsymbol{\varepsilon}_t - \mathbf{a}_t(m_t^* - \theta_t)$, where $\mathbf{a}_t \theta_t$ is the conditional expectation of \mathbf{Y}_t given θ_t , we have for all $s, t > 0$,

$$\text{Cov}(\boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_s) = \text{Cov}(\hat{\boldsymbol{\varepsilon}}_t^*, \hat{\boldsymbol{\varepsilon}}_s^*) + \mathbf{a}_t \mathbf{a}_s^\top C_{ts}^*.$$

According to the results in section 3.2, the $\hat{\epsilon}_t^*$ s are mutually uncorrelated, and hence we obtain

$$E(\hat{\epsilon}_t^*, \hat{\epsilon}_s^{*\top}) = \begin{cases} \Lambda_t \tau_t - \mathbf{a}_t \mathbf{a}_t^\top C_t^* & \text{for } t = s \\ -\mathbf{a}_t \mathbf{a}_s^\top C_{ts}^* & \text{for } t \neq s. \end{cases}$$

A special case of this result is

$$\text{Cov}(\hat{\epsilon}_t^*, m_0^*) = \mathbf{a}_t C_{t0}^*.$$

From section 3.2, we have $\text{Cov}(\epsilon_t, \theta) = 0$, and using $\theta_t = m_t^* - (m_t^* - \theta_t)$, we obtain

$$\begin{aligned} \text{Cov}(\hat{\epsilon}_t^*, \hat{\xi}_s^*) &= \mathbf{a}_t \text{Cov}(m_t^* - \theta_t, m_s^* - \theta_s - b_s(m_{s-1}^* - \theta_{s-1})) \\ &= \mathbf{a}_t C_{ts}^* - \mathbf{a}_t b_s C_{ts-1}^*. \end{aligned}$$

To obtain the covariance matrix of $\xi^* = (\hat{\xi}_1^*, \dots, \hat{\xi}_n^*)$, note that the innovation ξ_t may be written as a sum of two uncorrelated terms

$$\xi_t = \hat{\xi}_t^* + \{(\theta_t - m_t^*) - b_t(\theta_{t-1} - m_{t-1}^*)\}.$$

The innovations are mutually uncorrelated with variance $\sigma^2 b_t^{q-1} \tau_t$, and since $\text{Var}(\theta - \mathbf{m}^*) = \mathbf{C}^*$, we have

$$E(\hat{\xi}_t^*, \hat{\xi}_{t+s}^*) = \begin{cases} \sigma^2 b_t^{q-1} \tau_t - C_t^* + 2b_t C_{tt-1}^* - b_t^2 C_{t-1}^* & \text{for } s = 0 \\ -C_{tt+s}^* + b_t C_{t-1t+s}^* + b_{t+s} C_{tt+s-1}^* - b_t b_{t+s} C_{t-1t+s-1}^* & \text{for } s > 0 \end{cases}$$

A special case of this formula is

$$E\{(m_0^* - \tau_0) \hat{\xi}_t^*\} = -C_{0t}^* + b_t C_{0t-1}^*.$$

Finally, the variance of m_0^* is given by

$$\text{Var}(m_0^*) = g_j^2 \omega^2 - C_{j0}^*.$$

B.2 Derivatives of Kalman filter and smoother

The expected values of the partial derivatives of m_t and m_t^* with respect to α are, respectively, given by

$$\mathbf{D}\alpha(t) = \mathbb{E} \left(\frac{\partial m_t}{\partial \alpha^\top} \right)$$

and

$$\mathbf{D}\alpha^*(t) = \mathbb{E} \left(\frac{\partial m_t^*}{\partial \alpha^\top} \right)$$

The expected values of the partial derivatives of m_t and m_t^* with respect to β and γ are defined similarly.

As mentioned in section 3.2, for the Kalman filter, we have

$$m_0 = g = \exp(\mathbf{w}^\top \gamma) \text{ and } m_t = b_t \{m_{t-1} + D_t \mathbf{a}_t^\top \mathbf{Q}_t^{-1} (\mathbf{Y}_t - \mathbf{f}_t)\},$$

where \mathbf{a}_t is a function of α , b_t is a function of β and D_t and \mathbf{Q}_t are functions of α and β . Since $\mathbb{E}(\mathbf{Y}_t - \mathbf{f}_t) = 0$, we obtain the recursion, for $t = 1, \dots, n$,

$$\begin{aligned} \mathbf{D}\alpha(t) &= b_t \mathbf{D}\alpha(t-1) - b_t^2 D_t \mathbf{a}_t^\top \mathbf{Q}_t^{-1} \mathbb{E} \left(\frac{\partial}{\partial \alpha^\top} \mathbf{a}_t m_{t-1} \right) \\ &= b_t (1 - b_t D_t \mathbf{a}_t^\top \mathbf{Q}_t^{-1} \mathbf{a}_t) \mathbf{D}\alpha(t-1) - b_t D_t \tau_t \mathbf{a}_t^\top \mathbf{Q}_t^{-1} \frac{\partial}{\partial \alpha^\top} \mathbf{a}_t \\ &= \frac{C_t}{D_t} \mathbf{D}\alpha(t-1) - b_t D_t \tau_t \mathbf{a}_t^\top \mathbf{Q}_t^{-1} \frac{\partial}{\partial \alpha^\top} \mathbf{a}_t \end{aligned}$$

starting with $\mathbf{D}\alpha(0) = 0$. Note that

$$\frac{\partial}{\partial \alpha^\top} \mathbf{a}_t = \text{blockdiag} (\mathbf{x}_t^\top a_{1t}, \dots, \mathbf{x}_t^\top a_{dt}).$$

Analogously, we obtain the recursions

$$\mathbf{D}\beta(t) = \frac{C_t}{b_t D_t} \{ \Delta \mathbf{z}_t^\top \tau_t + b_t \mathbf{D}\beta(t-1) \},$$

starting with $\mathbf{D}_\beta(0) = 0$, and

$$\mathbf{D}_\gamma(t) = \frac{C_t}{D_t} \mathbf{D}_\gamma(t-1),$$

starting with $\mathbf{D}_\gamma(0) = \mathbf{w}^\top g$.

For the smoother, we immediately obtain the following backward recursions for, $t = n, \dots, 0$,

$$\begin{aligned} \mathbf{D}_\alpha^*(t) &= \mathbf{D}_\alpha(t) + \frac{C_t}{D_{t+1}} \{\mathbf{D}_\alpha^*(t+1) - b_{t+1} \mathbf{D}_\alpha(t)\}, \\ \mathbf{D}_\beta^*(t) &= \mathbf{D}_\beta(t) + \frac{C_t}{D_{t+1}} \{\mathbf{D}_\beta^*(t+1) - b_{t+1} \mathbf{D}_\beta(t) - \Delta \mathbf{z}_{t+1}^\top \tau_{t+1}\} \end{aligned}$$

and

$$\mathbf{D}_\gamma^*(t) = \mathbf{D}_\gamma(t) + \frac{C_t}{D_{t+1}} \{\mathbf{D}_\gamma^*(t+1) - b_{t+1} \mathbf{D}_\gamma(t)\},$$

starting with $\mathbf{D}_\alpha^*(n) = \mathbf{D}_\alpha(n)$, $\mathbf{D}_\beta^*(n) = \mathbf{D}_\beta(n)$ and $\mathbf{D}_\gamma^*(n) = \mathbf{D}_\gamma(n)$.

B.3 Variability matrix

By definition, the variability matrix for the estimating function ψ is

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} & \mathbf{V}_{13} \\ \mathbf{V}_{21} & \mathbf{V}_{22} & \mathbf{V}_{23} \\ \mathbf{V}_{31} & \mathbf{V}_{32} & \mathbf{V}_{33} \end{pmatrix},$$

where $\mathbf{V}_{ij} = E[\psi_i \psi_j^\top]$. From section B.1, we have the variance-covariance structure of $\mathbf{Y} - \mathbf{A}\mathbf{m}^*$, $\mathbf{B}\mathbf{m}^*$ and \mathbf{m}_0^* . Straightforward calculations give

$$\mathbf{V}_{11} = \mathbf{X}^\top \mathbf{K}_a^{-1} \text{Var}(\mathbf{Y} - \mathbf{A}\mathbf{m}^*) \mathbf{K}_a^{-1} \mathbf{X},$$

$$\mathbf{V}_{12} = \mathbf{X}^\top \mathbf{K}_a^{-1} \text{Cov}(\mathbf{Y} - \mathbf{A}\mathbf{m}^*, \mathbf{B}\mathbf{m}^*) \mathbf{K}_b^{-1} \Delta \mathbf{Z},$$

$$\mathbf{V}_{13} = \mathbf{X}^\top \mathbf{K}_a^{-1} \text{Cov}(\mathbf{Y} - \mathbf{A}\mathbf{m}^*, \mathbf{m}_0^*) \mathbf{K}_g^{-1} \mathbf{W},$$

$$\mathbf{V}_{22} = \Delta \mathbf{Z}^\top \mathbf{K}_b^{-1} \text{Var}(\mathbf{B}\mathbf{m}^*) \mathbf{K}_b^{-1} \Delta \mathbf{Z},$$

$$\mathbf{V}_{23} = \Delta \mathbf{Z}^\top \mathbf{K}_b^{-1} \text{Cov}(\mathbf{B}\mathbf{m}^*, \mathbf{m}_0^*) \mathbf{K}_g^{-1} \mathbf{W}$$

and

$$\mathbf{V}_{33} = \mathbf{W}^\top \mathbf{K}_g^{-1} \text{Var}(\mathbf{m}_0^*) \mathbf{K}_g^{-1} \mathbf{W}.$$

The variances and covariances involved in these expressions may be calculated from the results in section B.1 by noting that $\mathbf{Y} - \mathbf{A}\mathbf{m}^*$ is the vector of the predicted innovations $\hat{\boldsymbol{\varepsilon}}_{jt}^*$ and $\mathbf{B}\mathbf{m}^*$ is the vector of predicted innovations $\hat{\boldsymbol{\xi}}_{jt}^*$.

B.4 Sensitivity matrix

As in the previously section, we may write the sensitivity matrix for $\boldsymbol{\psi}$ in the form

$$\mathbf{S} = \begin{pmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} & \mathbf{S}_{13} \\ \mathbf{S}_{21} & \mathbf{S}_{22} & \mathbf{S}_{23} \\ \mathbf{S}_{31} & \mathbf{S}_{32} & \mathbf{S}_{33} \end{pmatrix},$$

where $\mathbf{S}_{ij} = \text{E}[\partial \boldsymbol{\psi}_i / \partial \boldsymbol{\eta}_j^\top]$, with $j = 1, 2, 3$ referring to $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ respectively. Using the derivatives of the Kalman smoother from section B.2, we obtain

$$\mathbf{S}_{11} = -\mathbf{X}^\top \mathbf{K}_a^{-1} \left(\frac{\partial \mathbf{A}}{\partial \boldsymbol{\alpha}^\top} \boldsymbol{\tau} + \mathbf{A} \mathbf{D}_{\boldsymbol{\alpha}}^* \right),$$

where $\mathbf{D}_{\boldsymbol{\alpha}}^*$ denotes the vectors $\mathbf{D}_{\boldsymbol{\alpha}}^*(t)$ stacked in the appropriate order, and

$$\left(\frac{\partial \mathbf{A}}{\partial \boldsymbol{\alpha}^\top} \right)_{ij} = \sum_{t=1}^n \frac{\partial a_{ij}}{\partial \alpha_{ij}} \tau_t.$$

The remaining blocks of \mathbf{S} are, using a similar notation,

$$\mathbf{S}_{12} = -\mathbf{X}^\top \mathbf{K}_a^{-1} \mathbf{A} \mathbf{D}_{\boldsymbol{\beta}}^*,$$

$$\mathbf{S}_{13} = -\mathbf{X}^\top \mathbf{K}_a^{-1} \mathbf{A} \mathbf{D}_\gamma^*,$$

$$\mathbf{S}_{21} = -\Delta \mathbf{Z}^\top \mathbf{K}_b^{-1} \mathbf{B} \mathbf{D}_\alpha^*,$$

$$\mathbf{S}_{22} = \Delta \mathbf{Z}^\top \mathbf{K}_b^{-1} \left(\frac{\partial \mathbf{B}}{\partial \beta^\top} \boldsymbol{\tau} + \mathbf{B} \mathbf{D}_\beta^* \right),$$

$$\mathbf{S}_{23} = -\Delta \mathbf{Z}^\top \mathbf{K}_b^{-1} \mathbf{B} \mathbf{D}_\gamma^*,$$

$$\mathbf{S}_{31} = \mathbf{W}^\top \mathbf{K}_g^{-1} \mathbf{D}_{0\alpha}^*,$$

$$\mathbf{S}_{32} = \mathbf{W}^\top \mathbf{K}_g^{-1} \mathbf{D}_{0\beta}^*$$

and

$$\mathbf{S}_{33} = \mathbf{W}^\top \mathbf{K}_g^{-1} \left(\mathbf{D}_{0\gamma}^* - \frac{\partial \mathbf{g}}{\partial \gamma^\top} \right).$$

Here the subscript 0 refers to differentiation of m_0^* .

Appendix C

Residuals plots for each patient in the placebo group

Figure C.29: Residuals plots for patient #420 in the placebo group

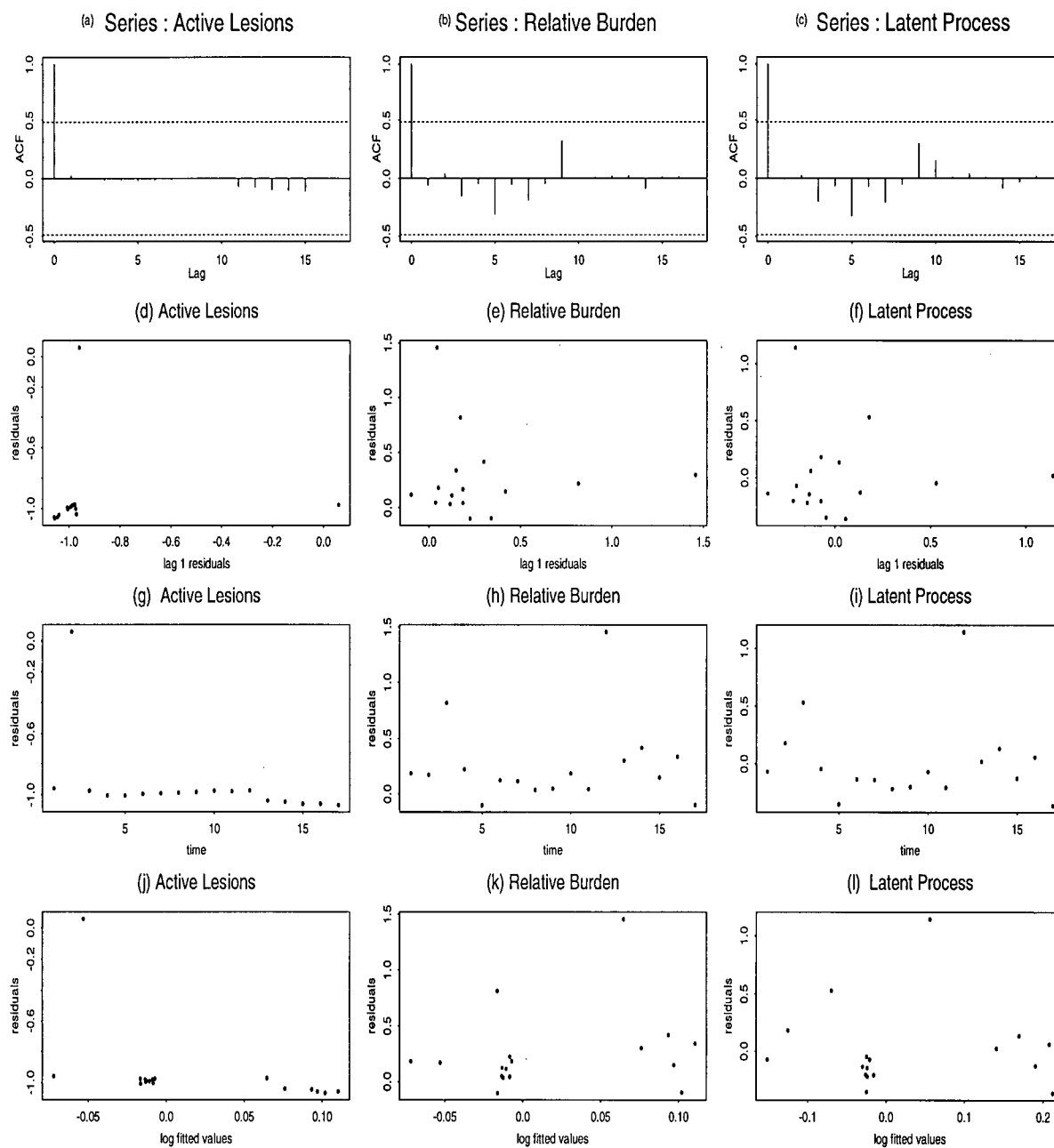


Figure C.30: Residuals plots for patient #421 in the placebo group

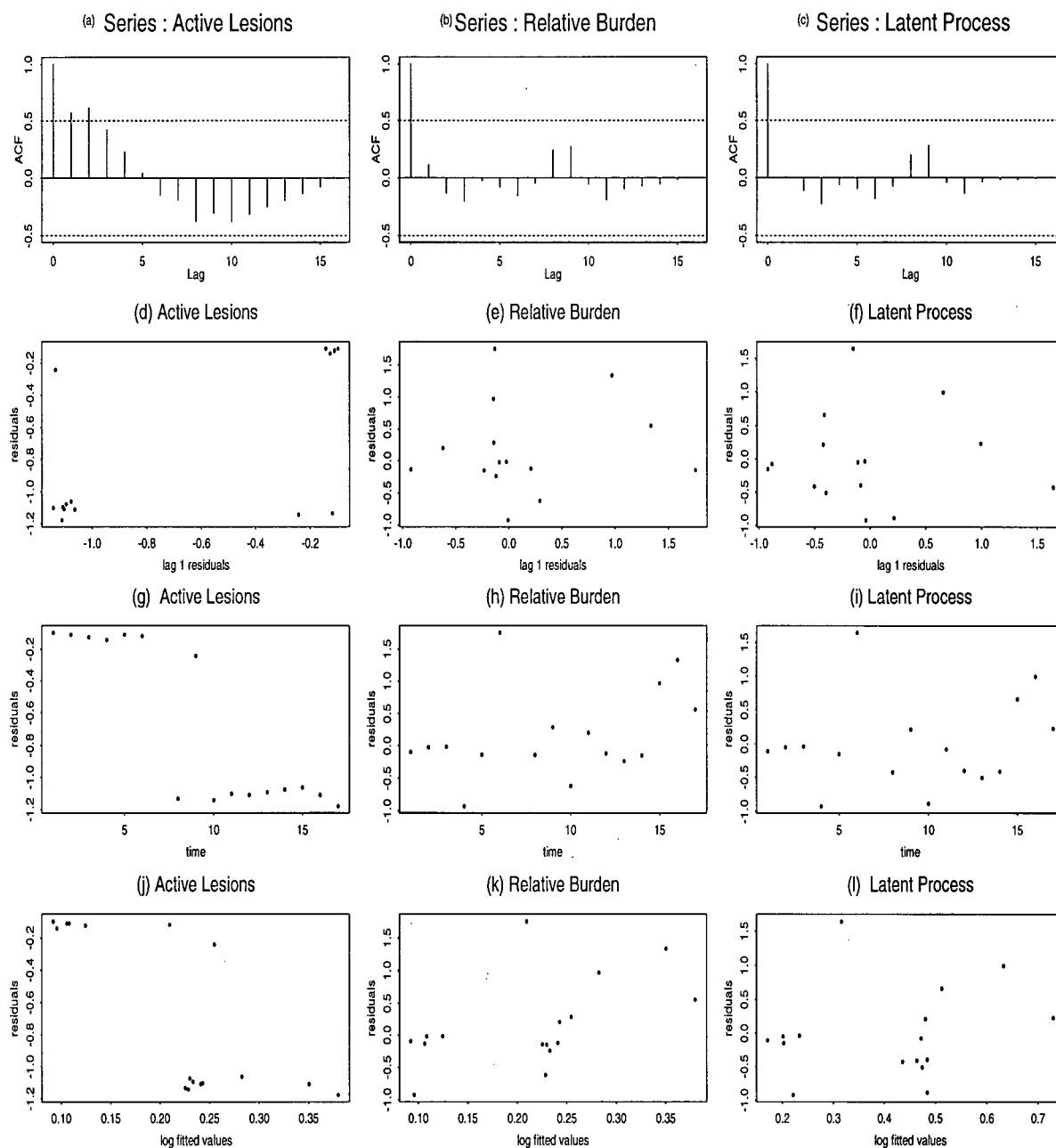


Figure C.31: Residuals plots for patient #443 in the placebo group

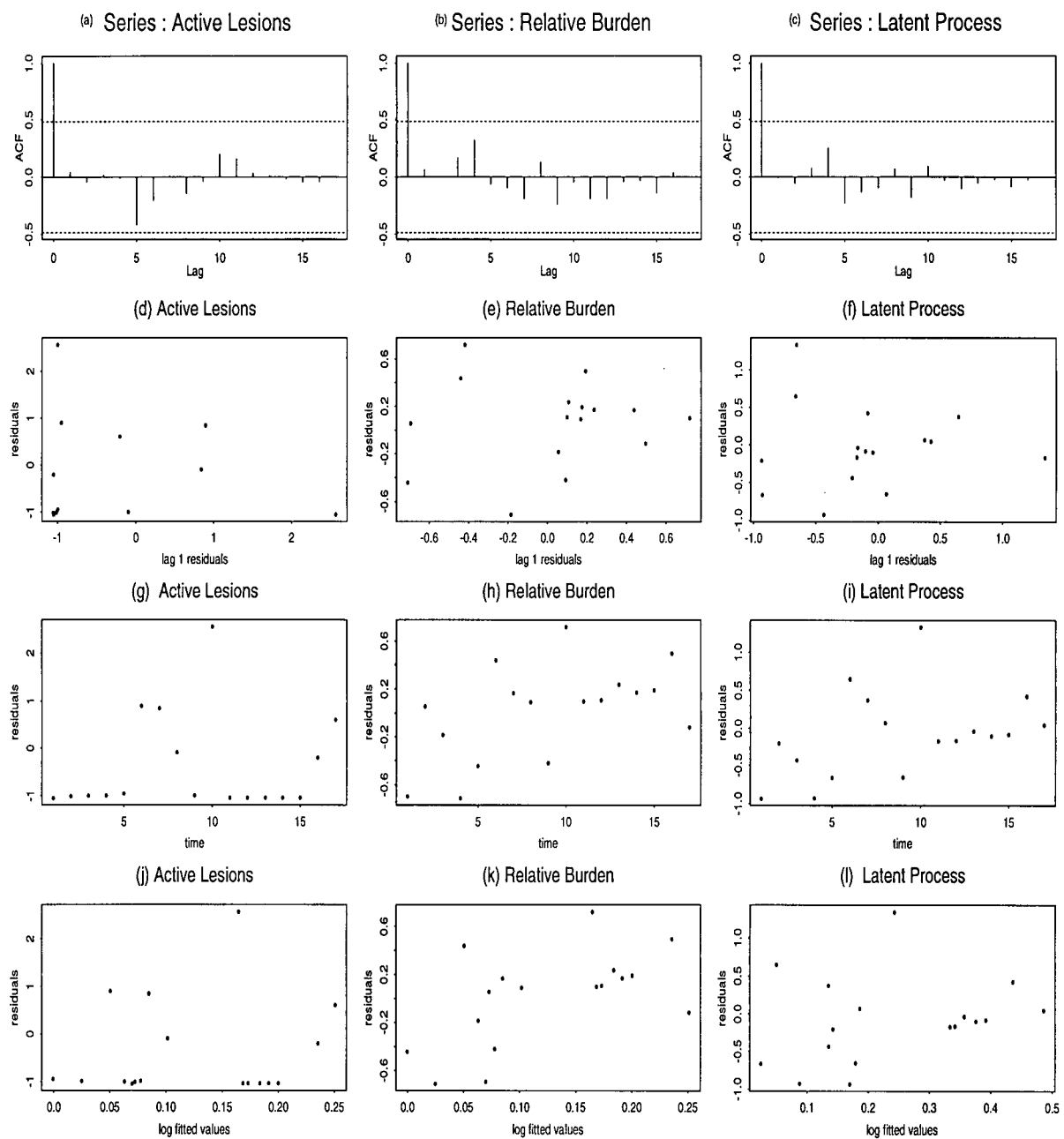


Figure C.32: Residuals plots for patient #449 in the placebo group

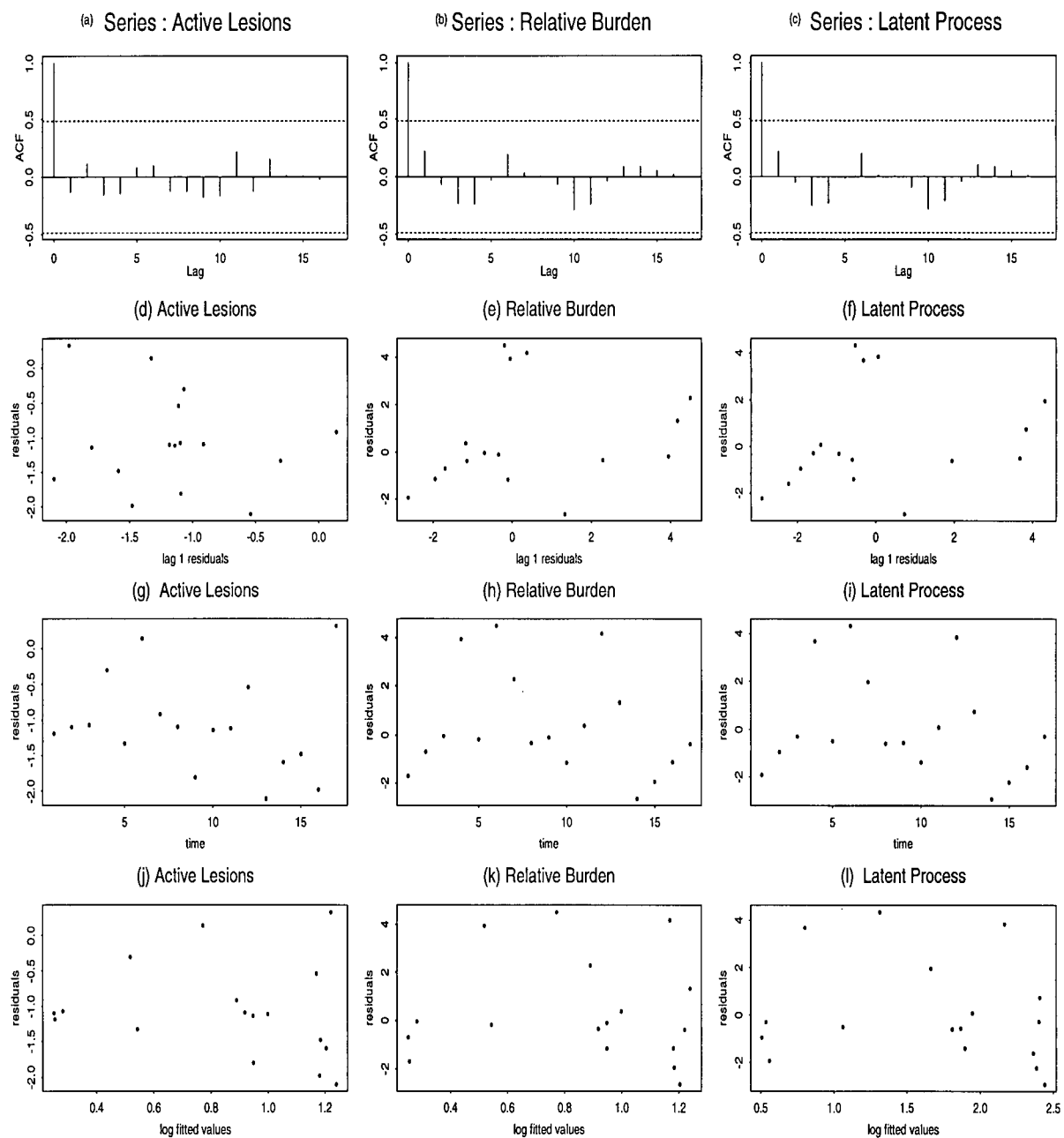


Figure C.33: Residuals plots for patient #451 in the placebo group

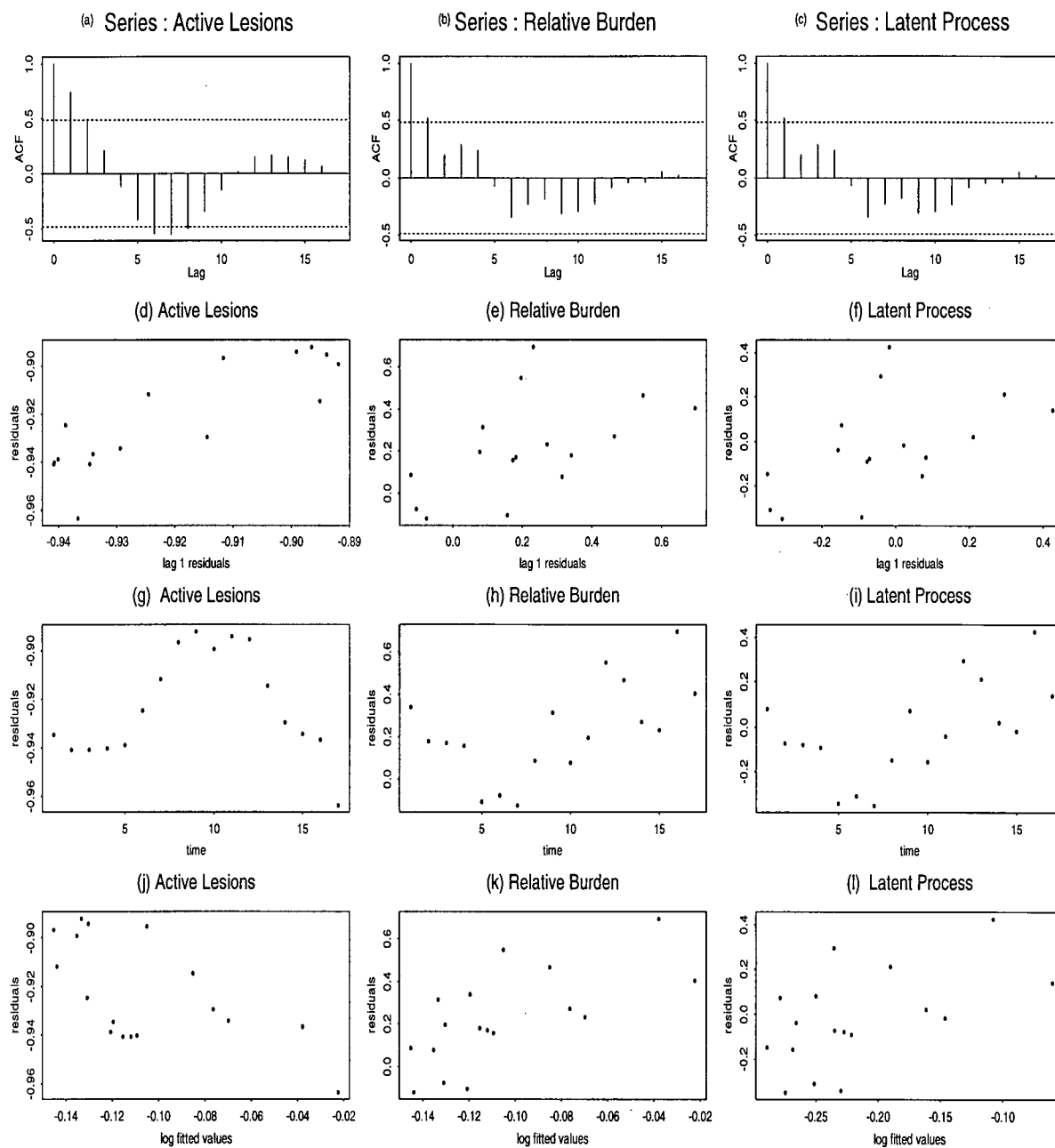


Figure C.34: Residuals plots for patient #498 in the placebo group

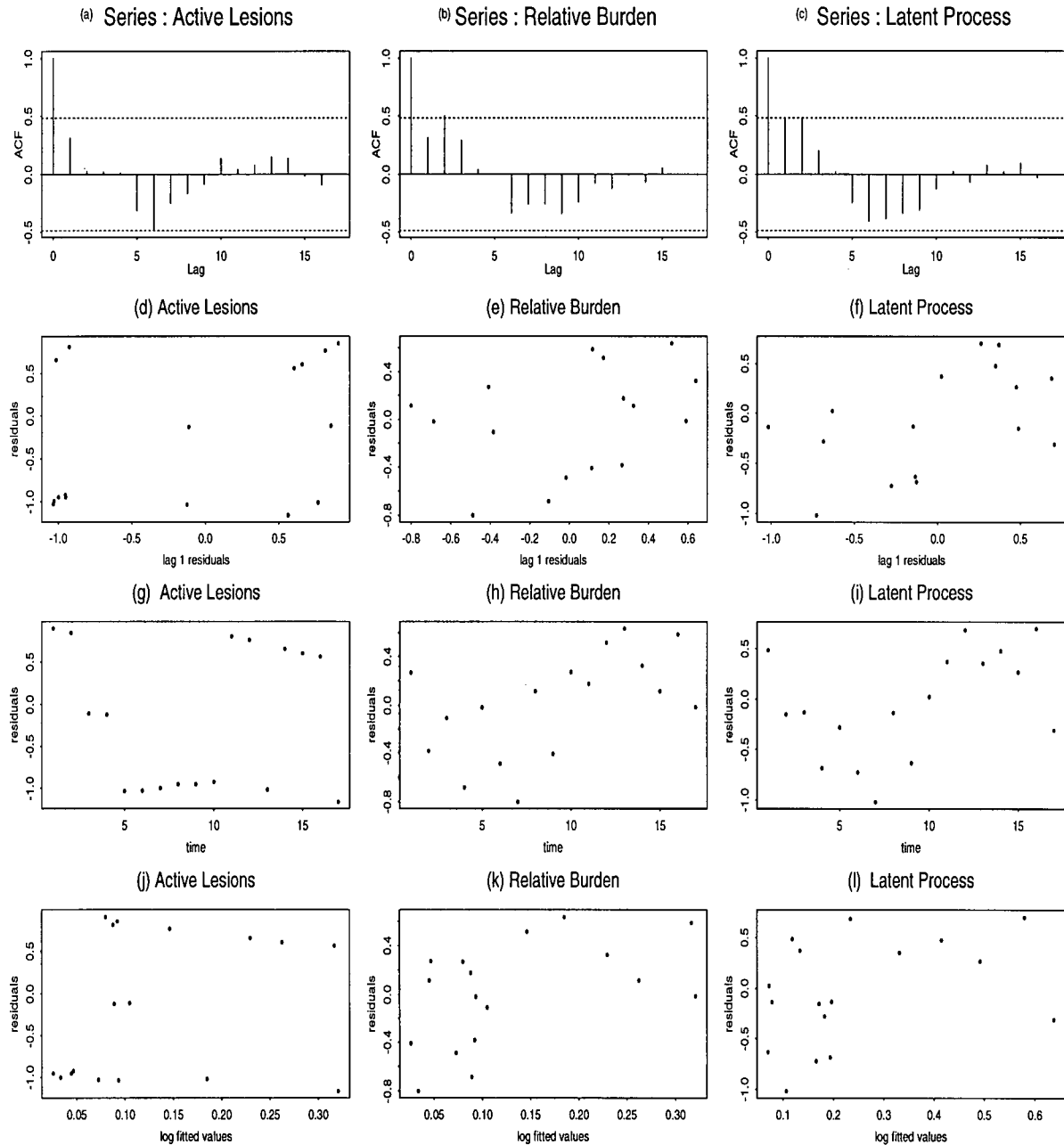


Figure C.35: Residuals plots for patient #501 in the placebo group

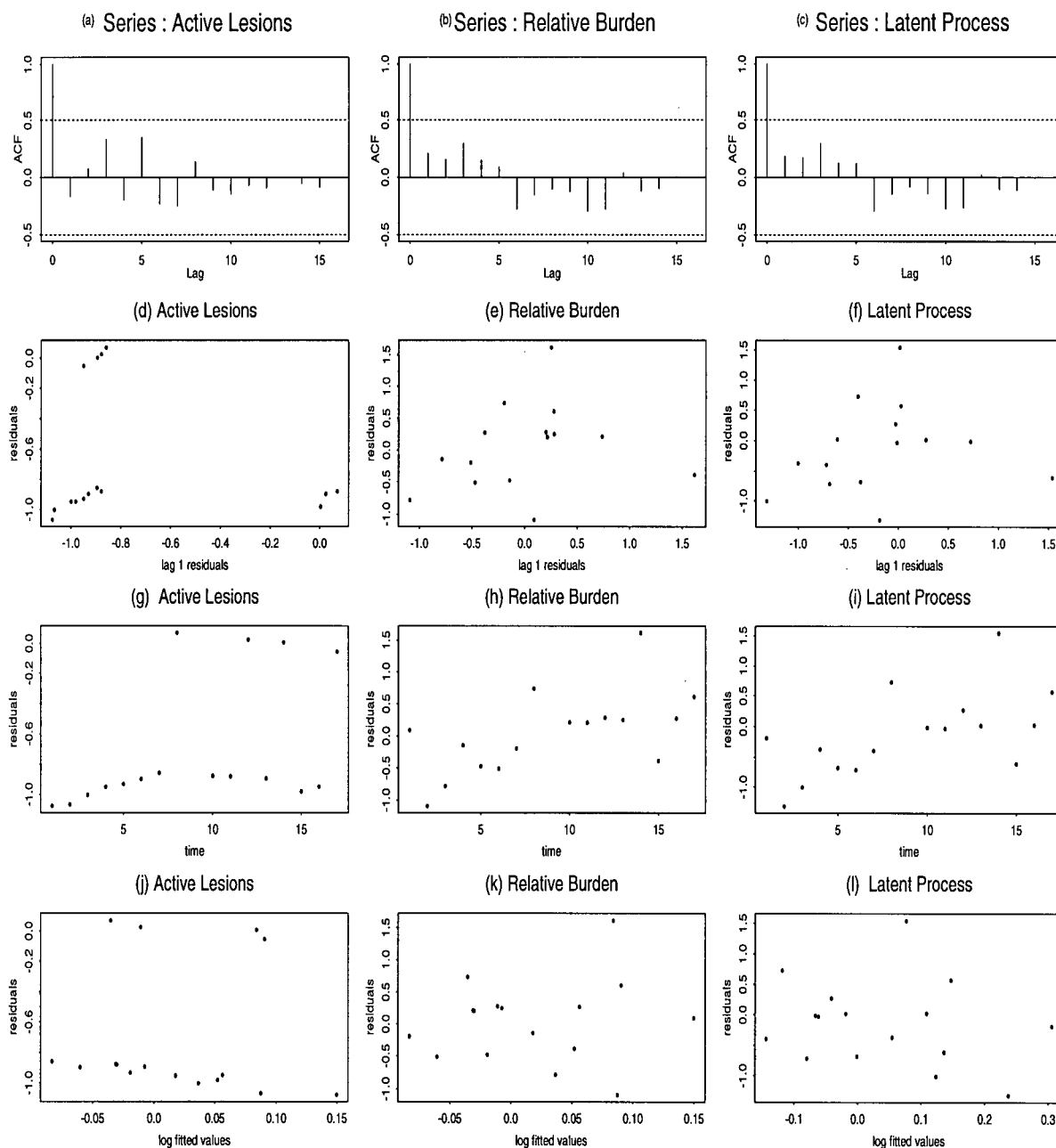


Figure C.36: Residuals plots for patient #504 in the placebo group

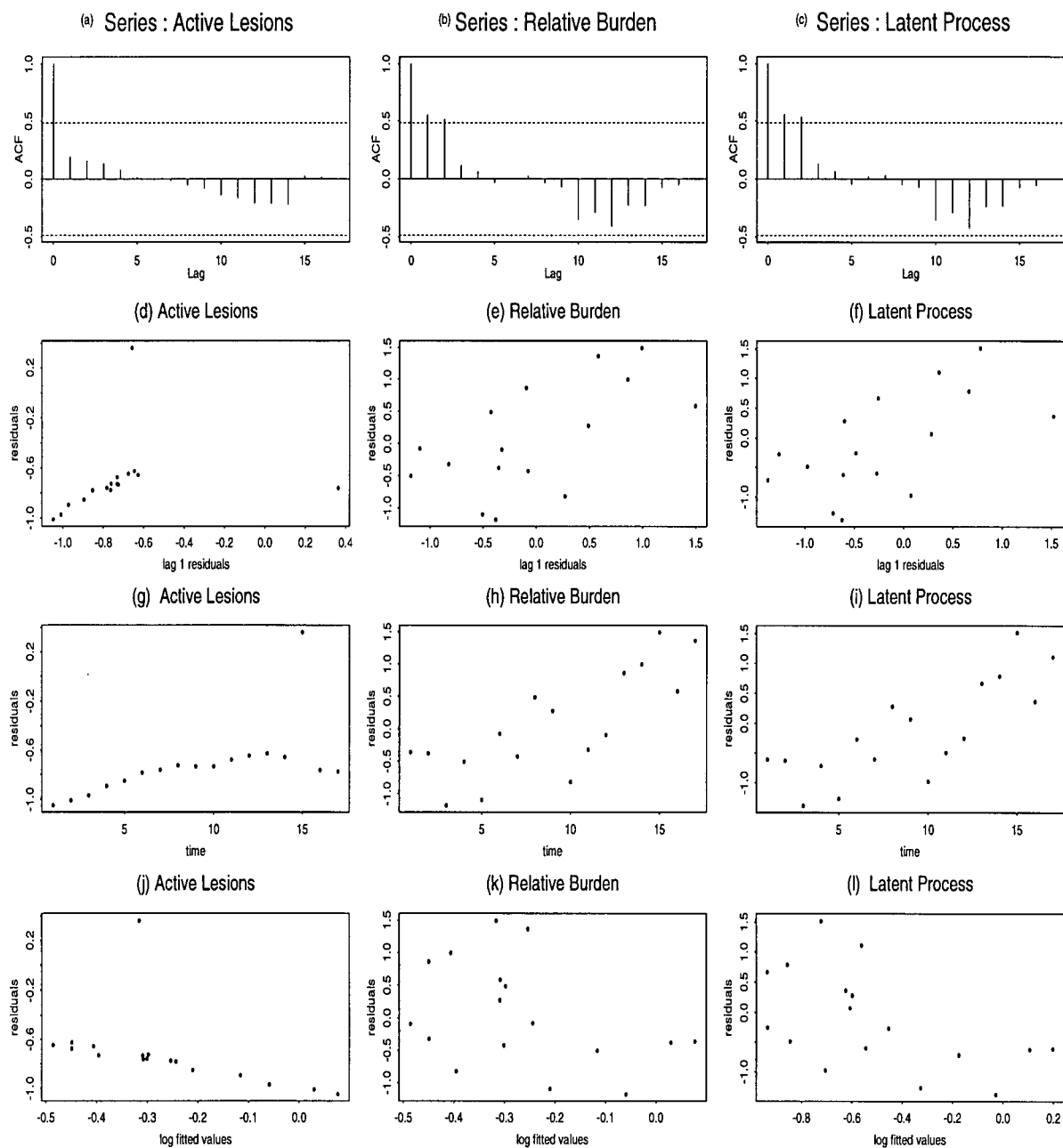


Figure C.37: Residuals plots for patient #507 in the placebo group

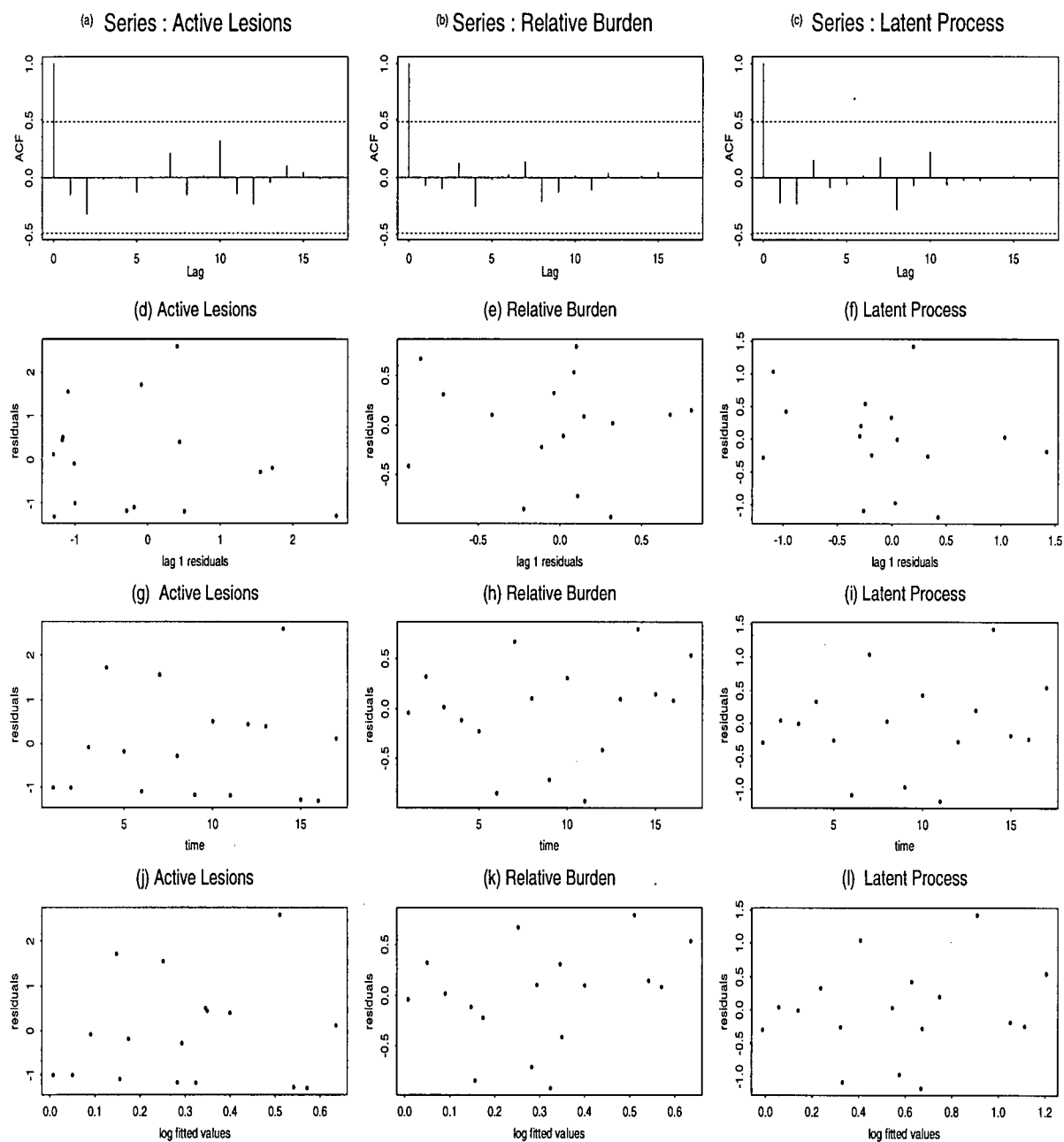


Figure C.38: Residuals plots for patient #522 in the placebo group

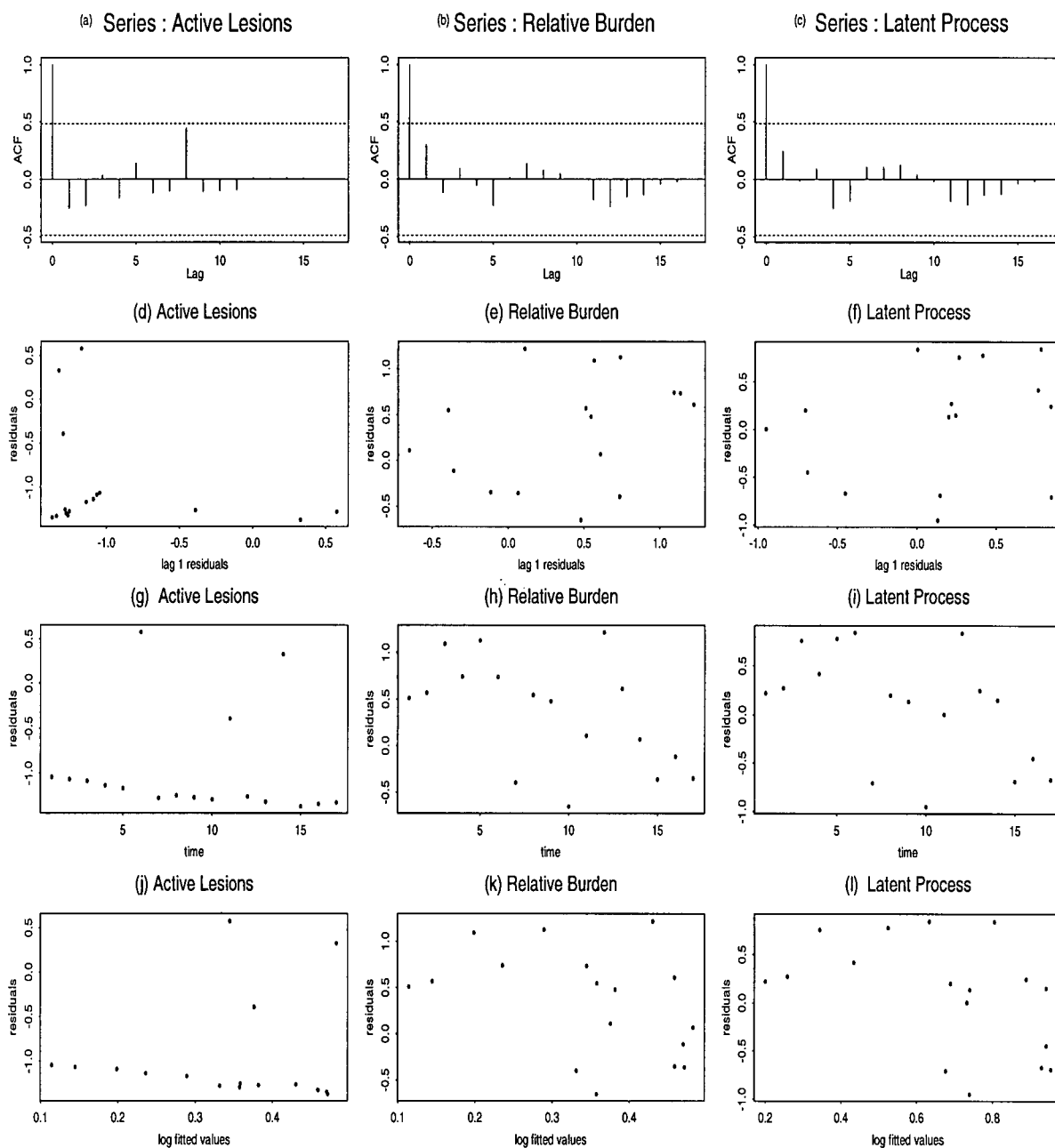


Figure C.39: Residuals plots for patient #523 in the placebo group

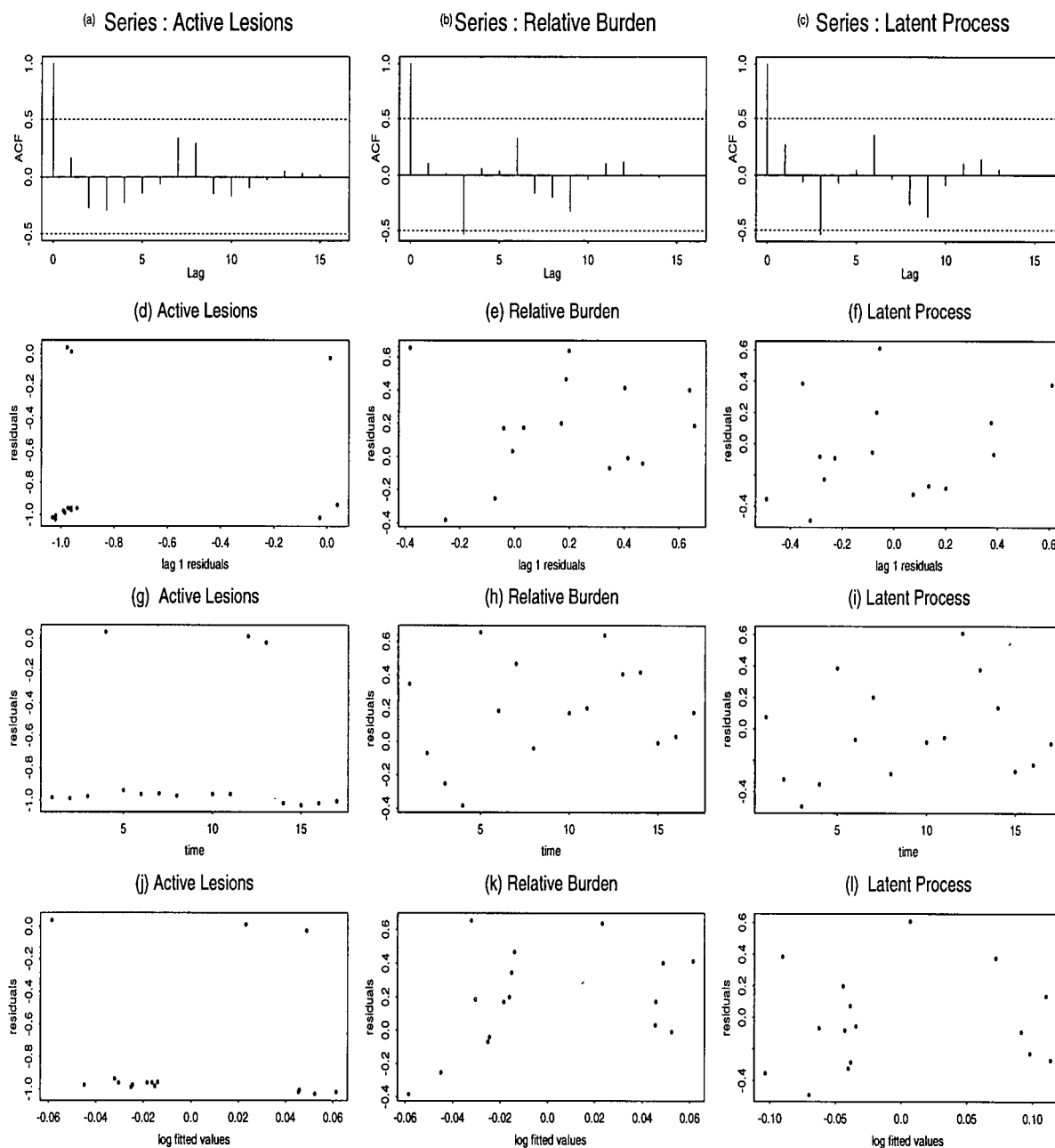


Figure C.40: Residuals plots for patient #539 in the placebo group

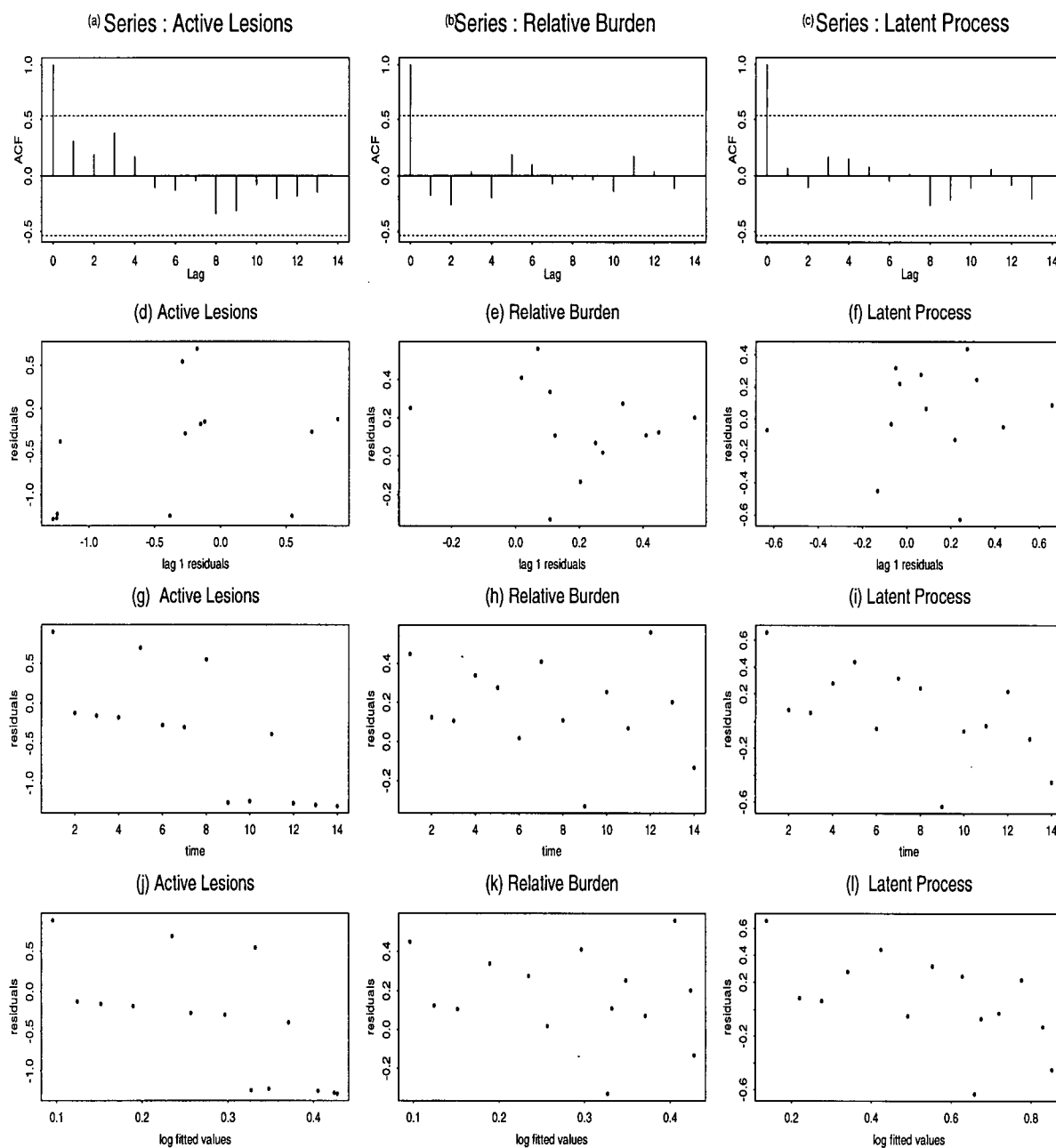


Figure C.41: Residuals plots for patient #540 in the placebo group

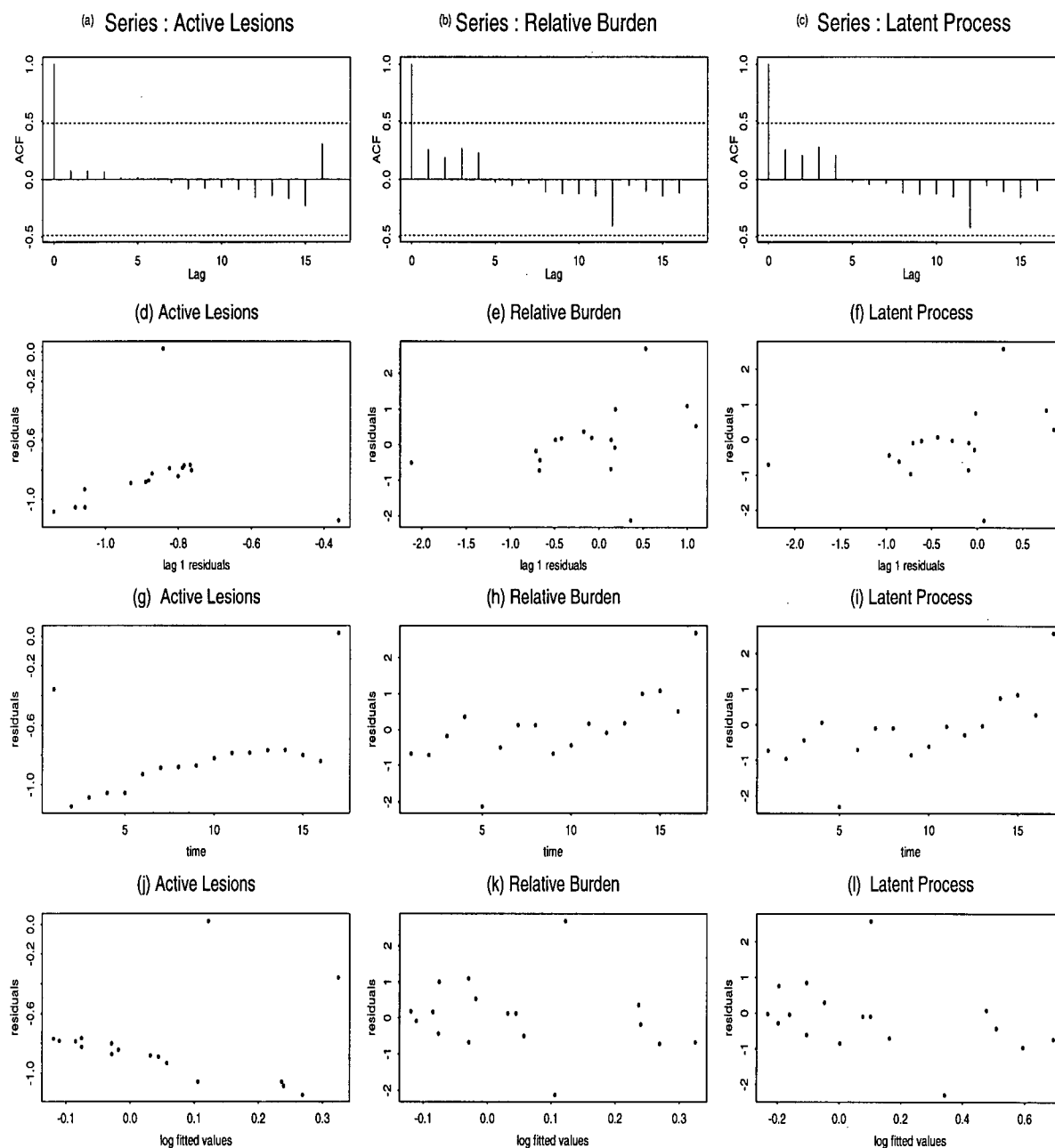


Figure C.42: Residuals plots for patient #550 in the placebo group.

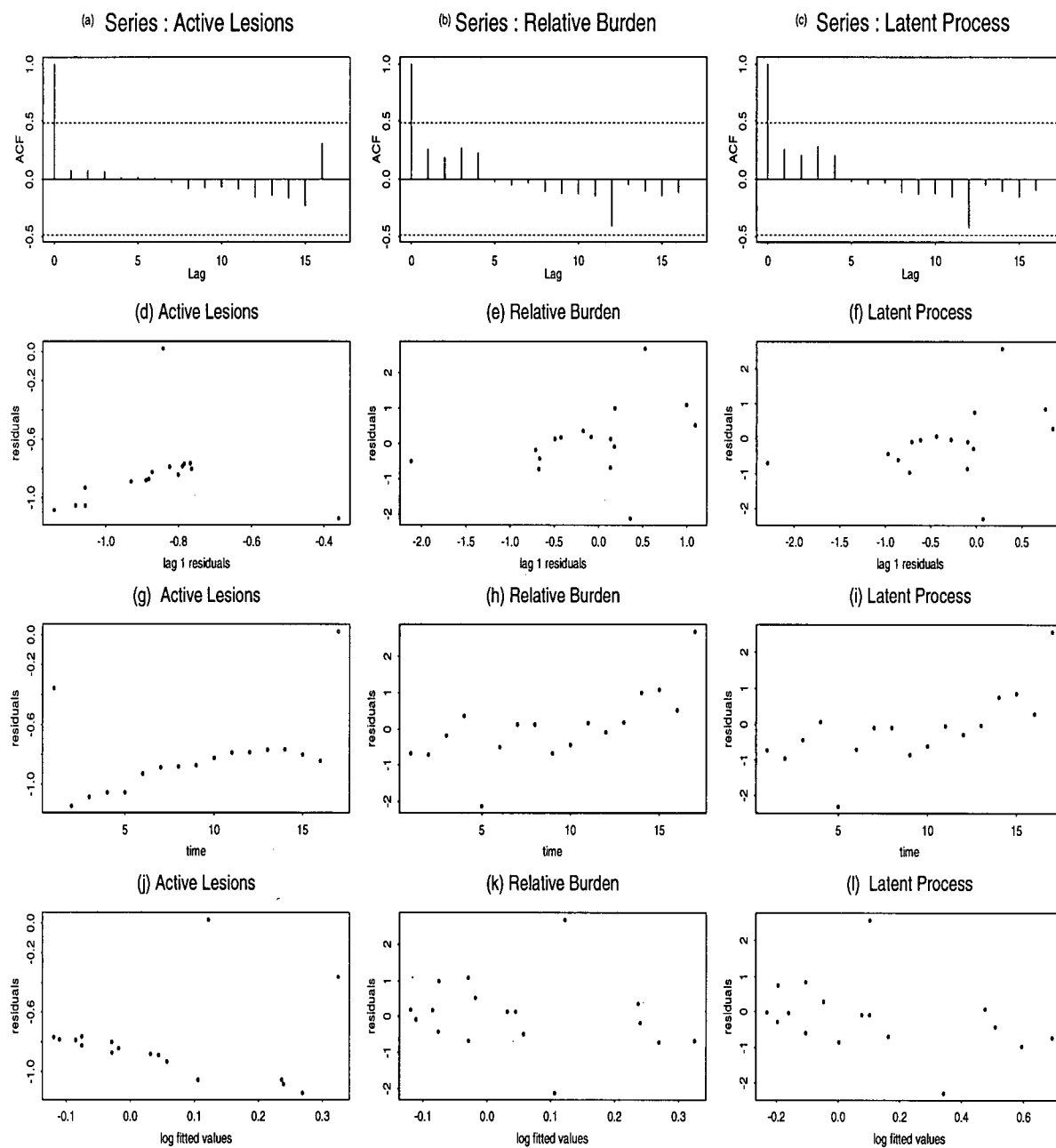
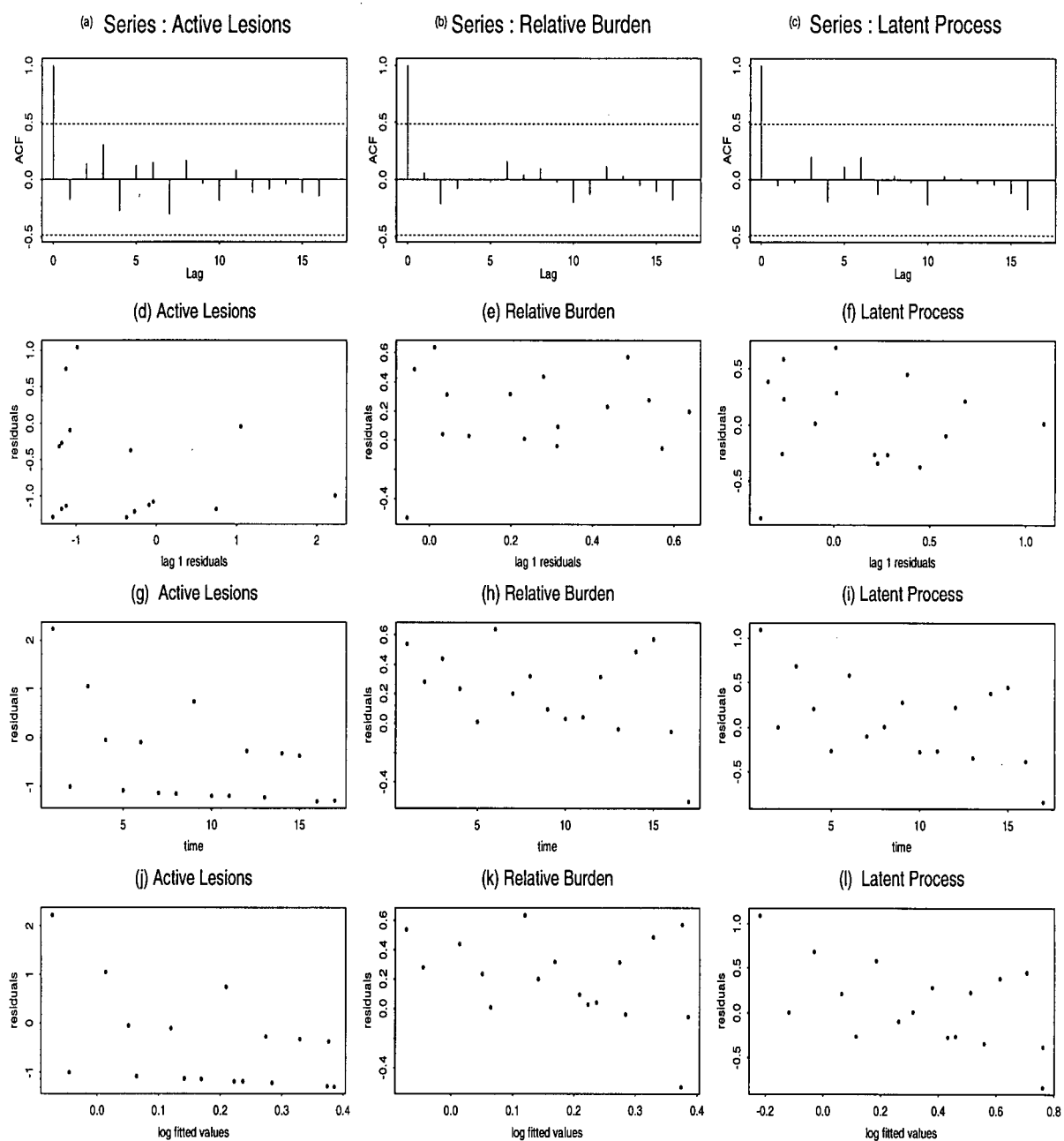


Figure C.43: Residuals plots for patient #565 in the placebo group



Appendix D

Residuals plots for each patient in the low dose group

Figure D.44: Residuals plots for patient #419 in the low dose group

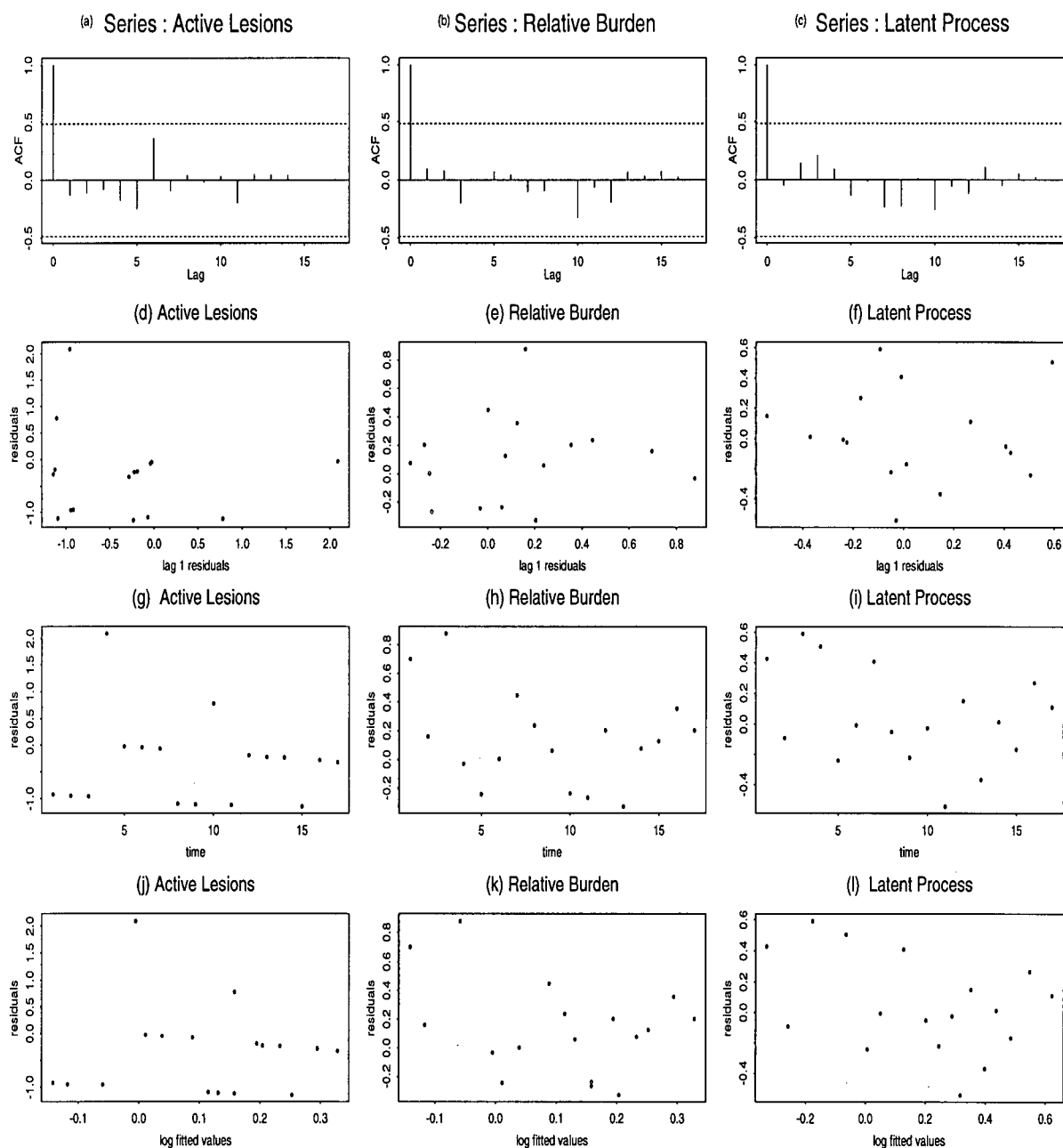


Figure D.45: Residuals plots for patient #424 in the low dose group

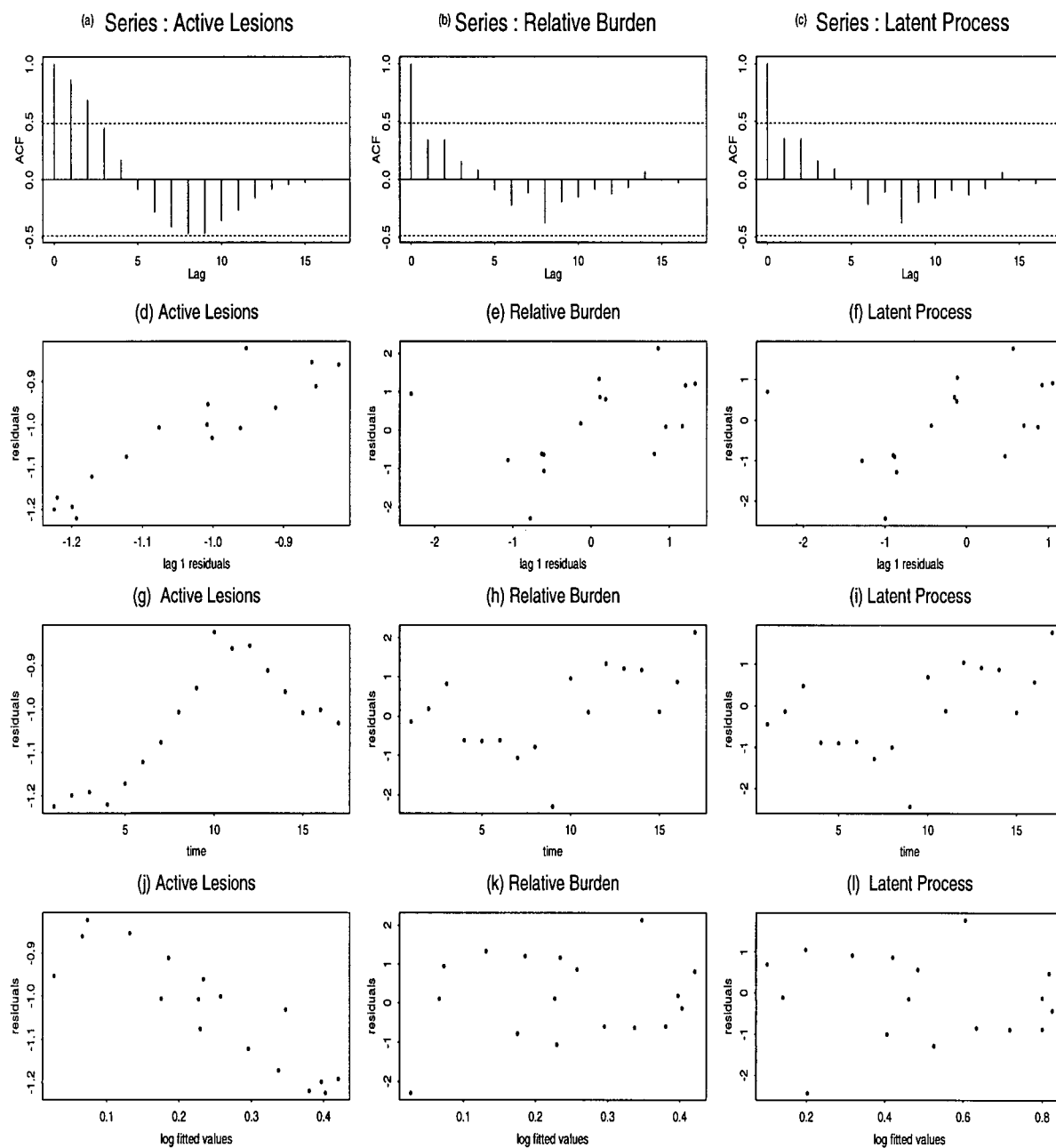


Figure D.46: Residuals plots for patient #448 in the low dose group

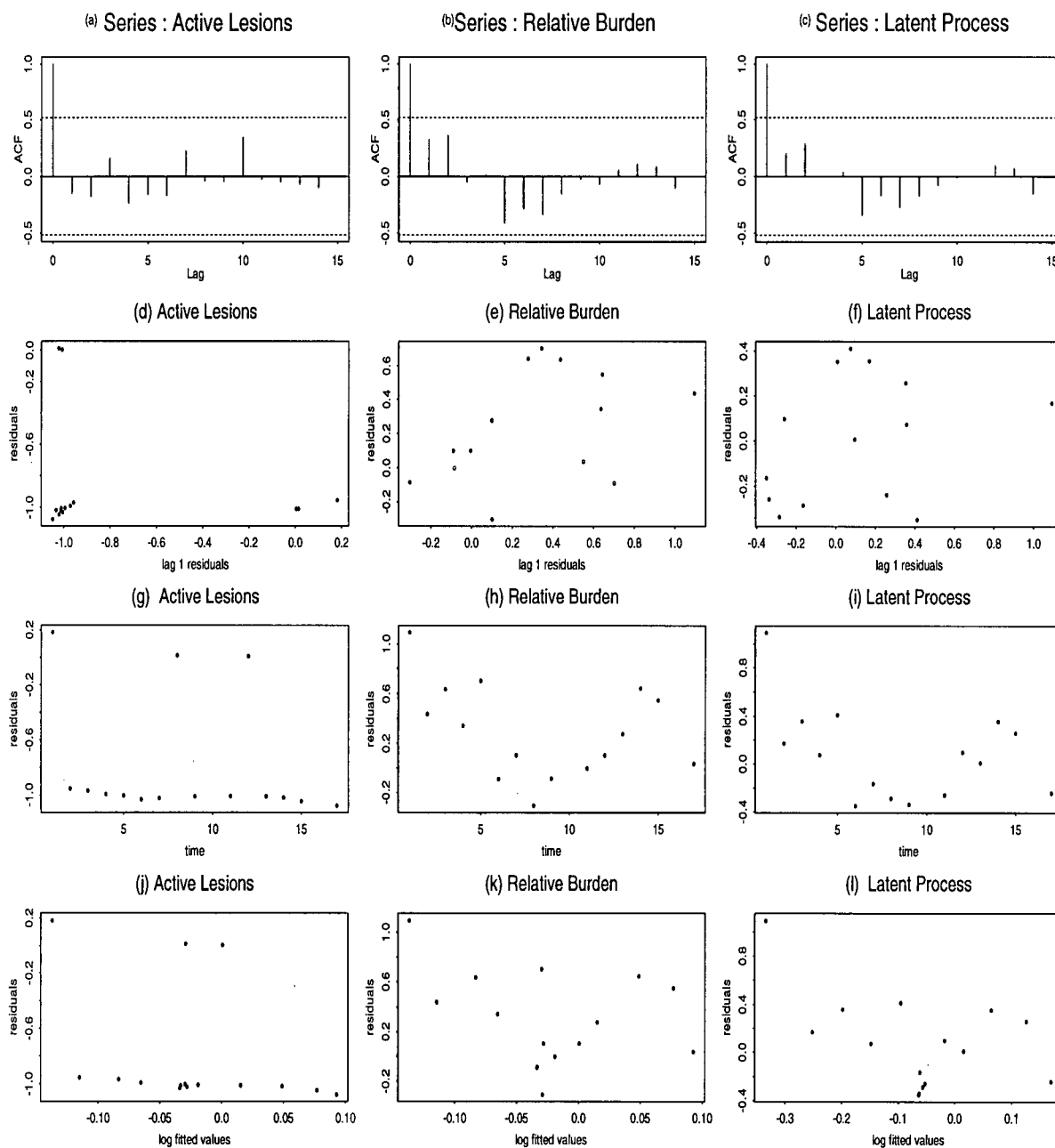


Figure D.47: Residuals plots for patient #450 in the low dose group

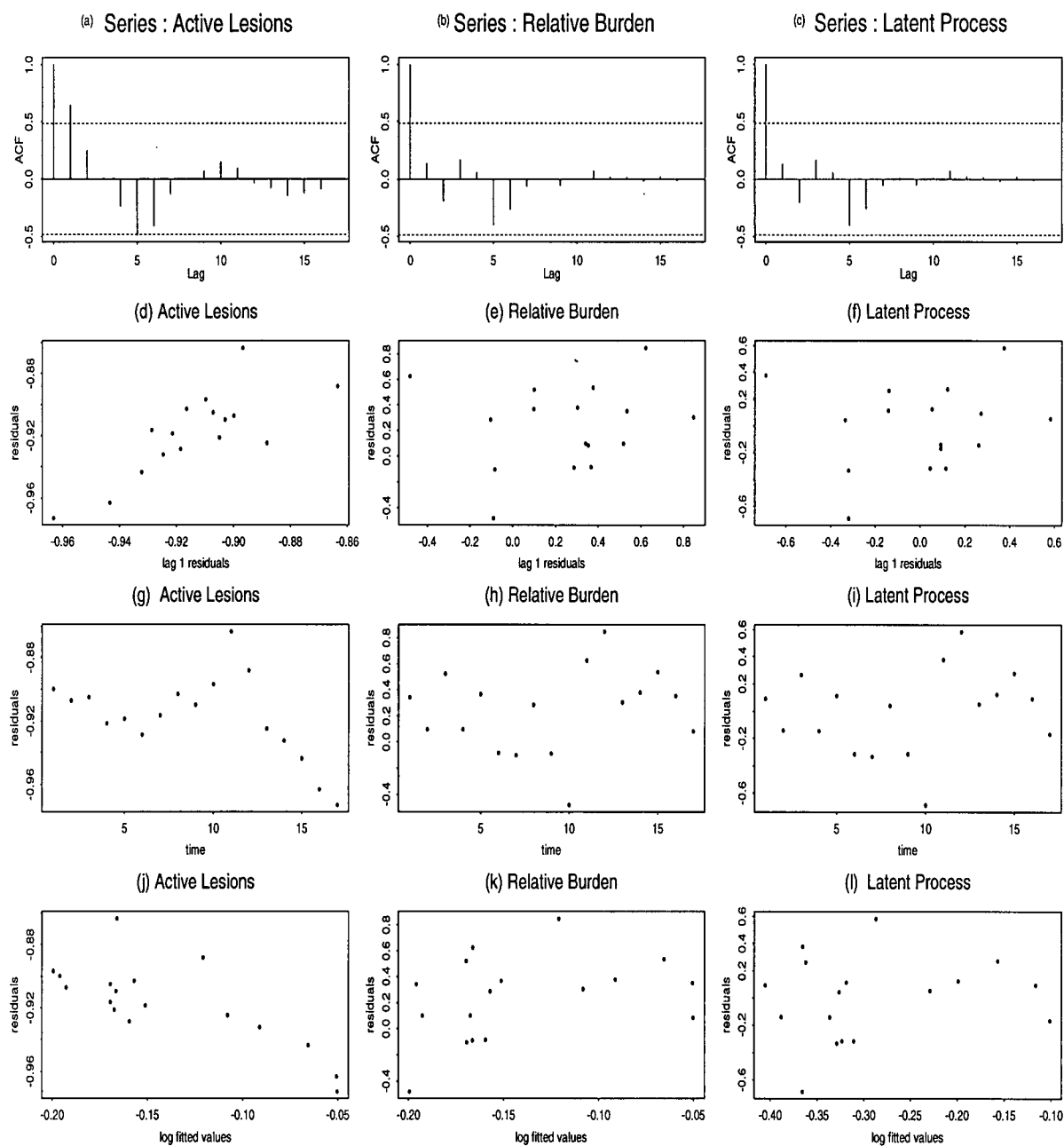


Figure D.48: Residuals plots for patient #452 in the low dose group

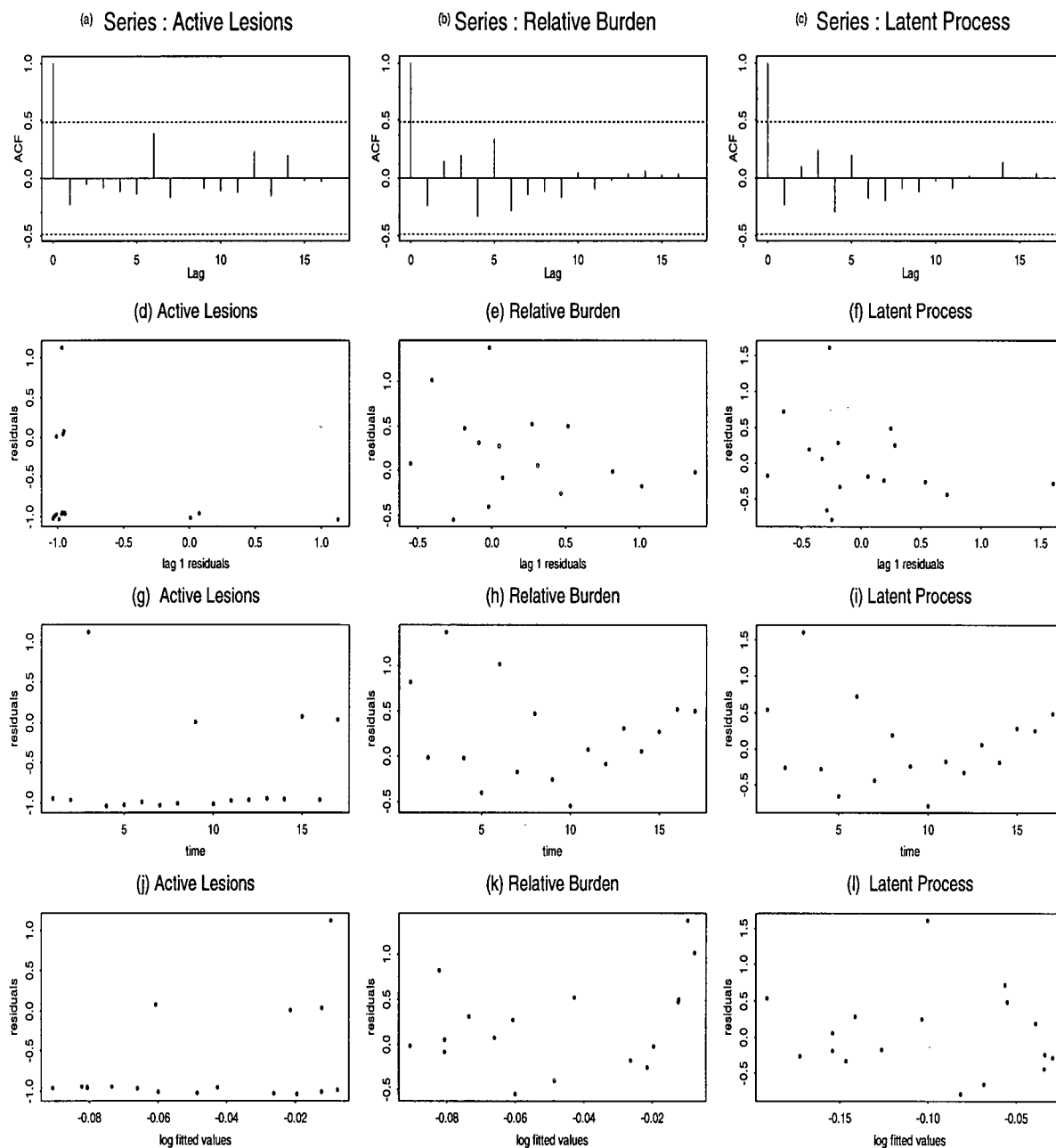


Figure D.49: Residuals plots for patient #499 in the low dose group

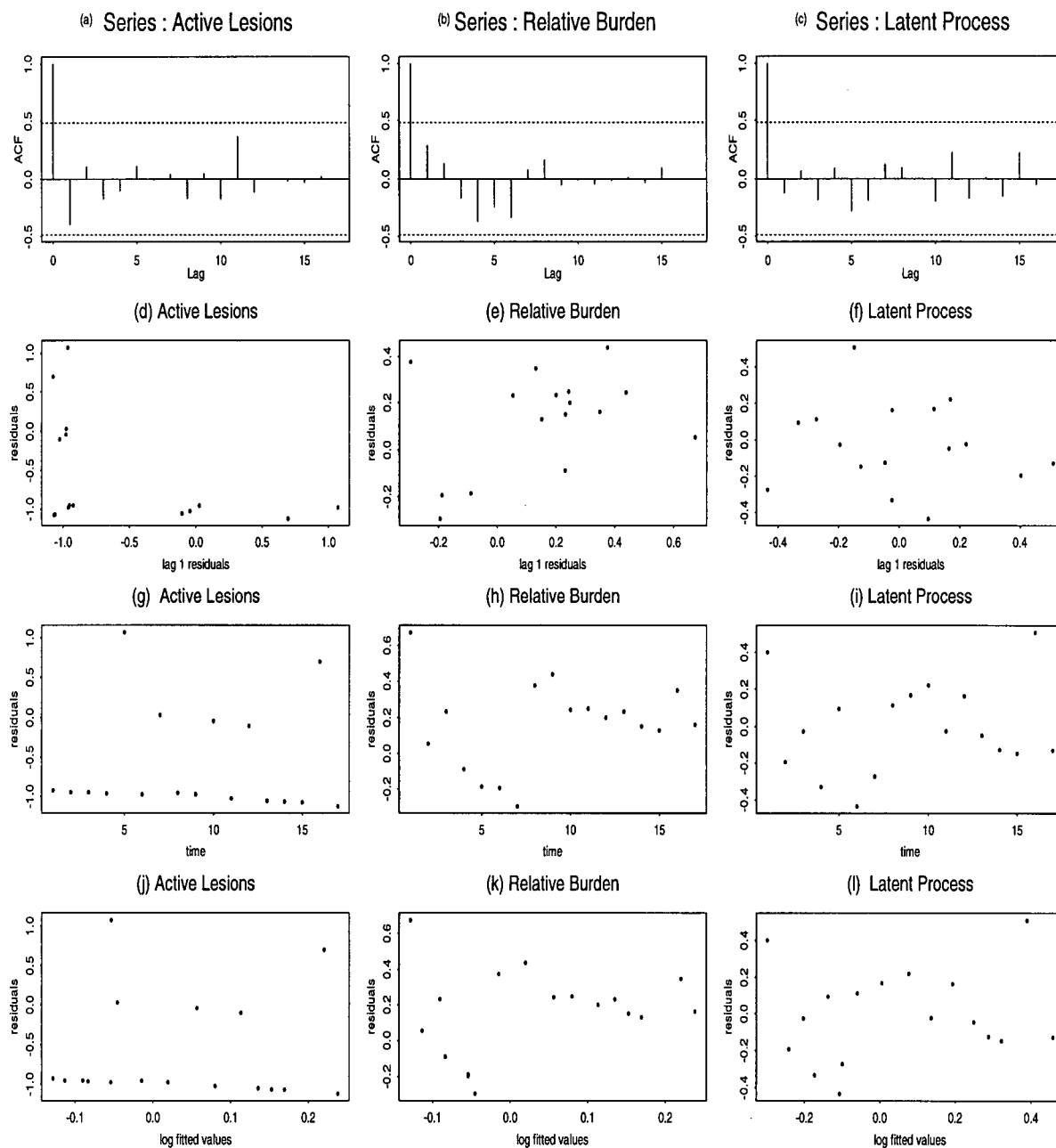


Figure D.50: Residuals plots for patient #508 in the low dose group

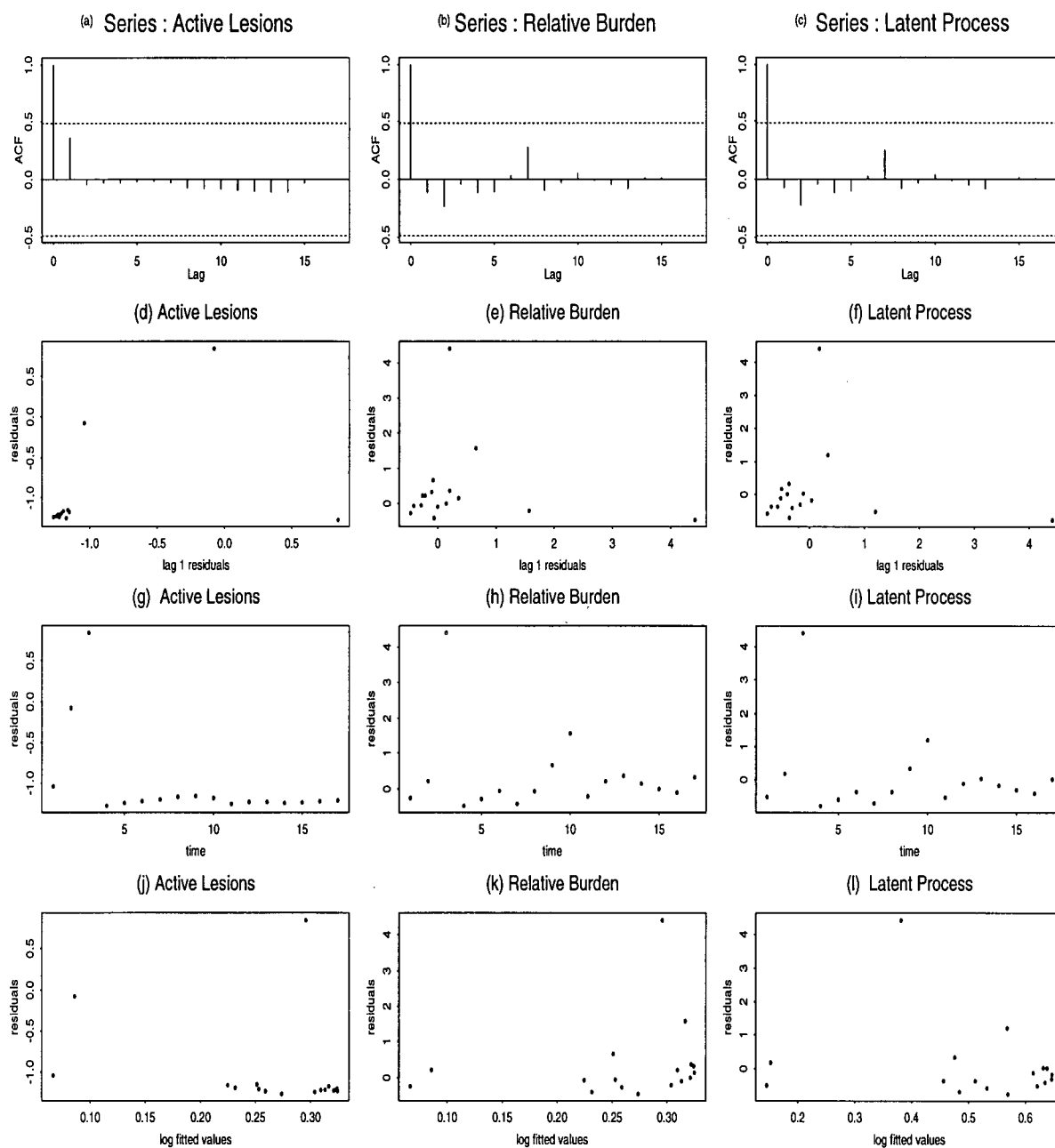


Figure D.51: Residuals plots for patient #521 in the low dose group

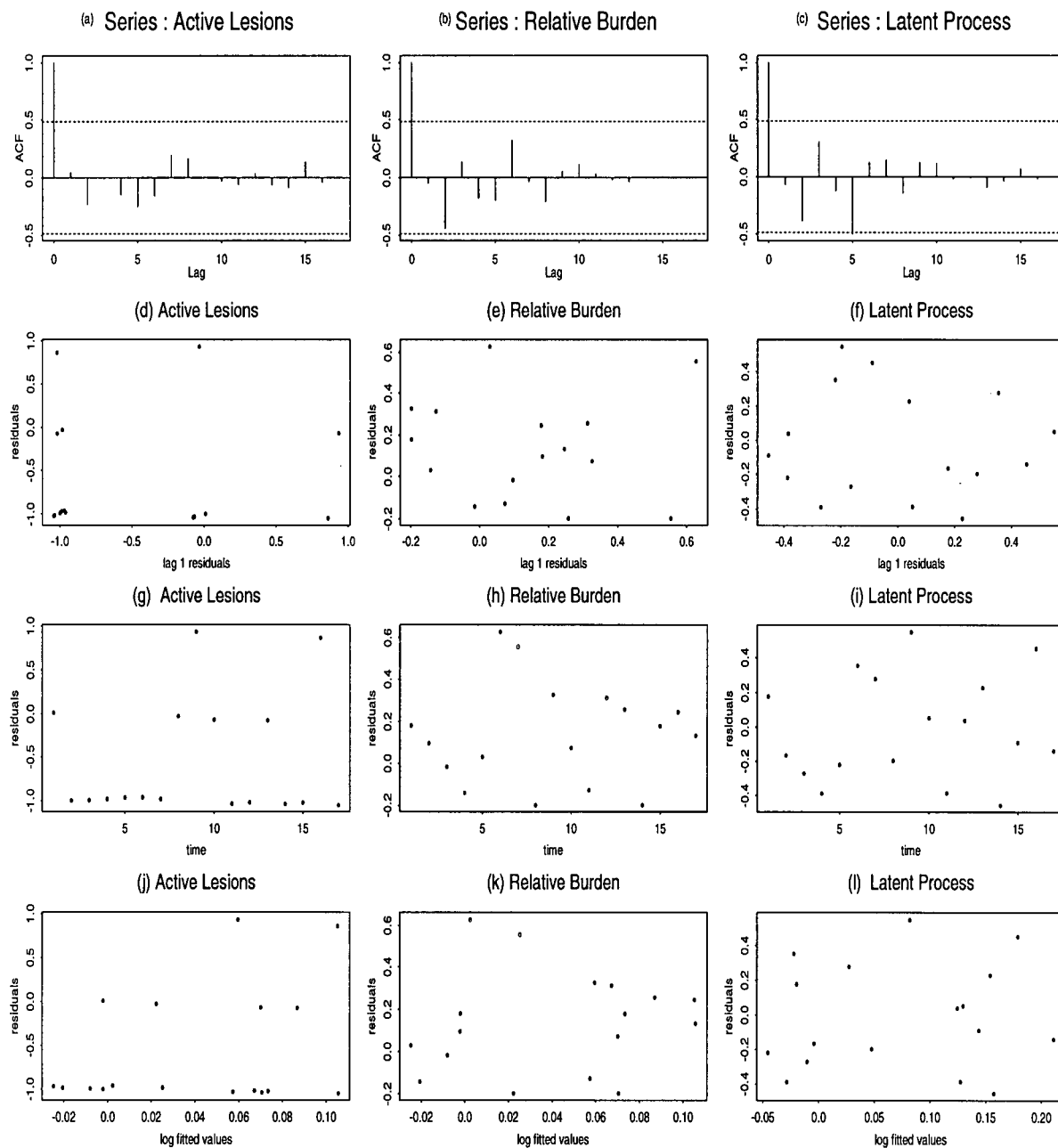


Figure D.52: Residuals plots for patient #525 in the low dose group

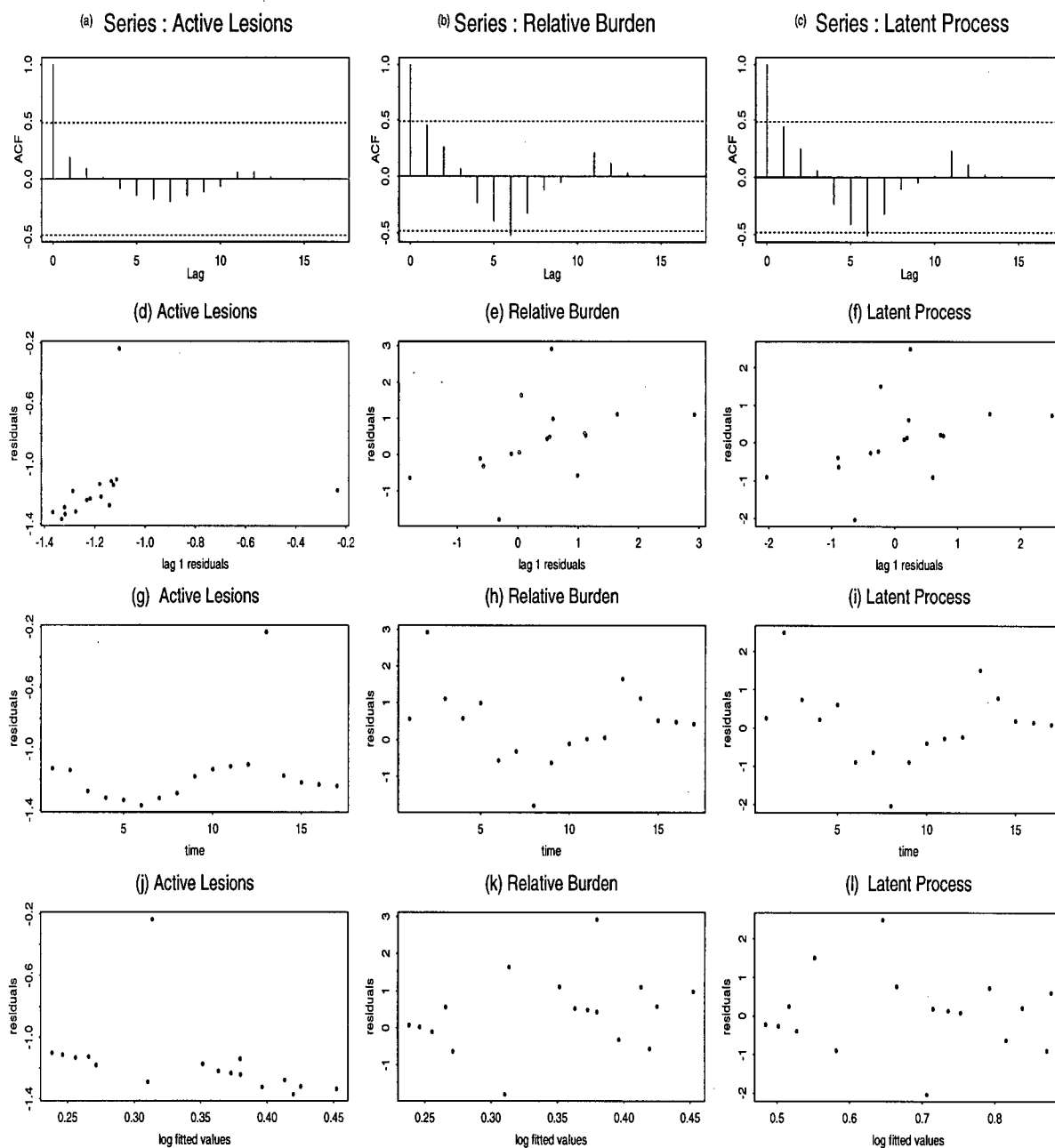


Figure D.53: Residuals plots for patient #542 in the low dose group

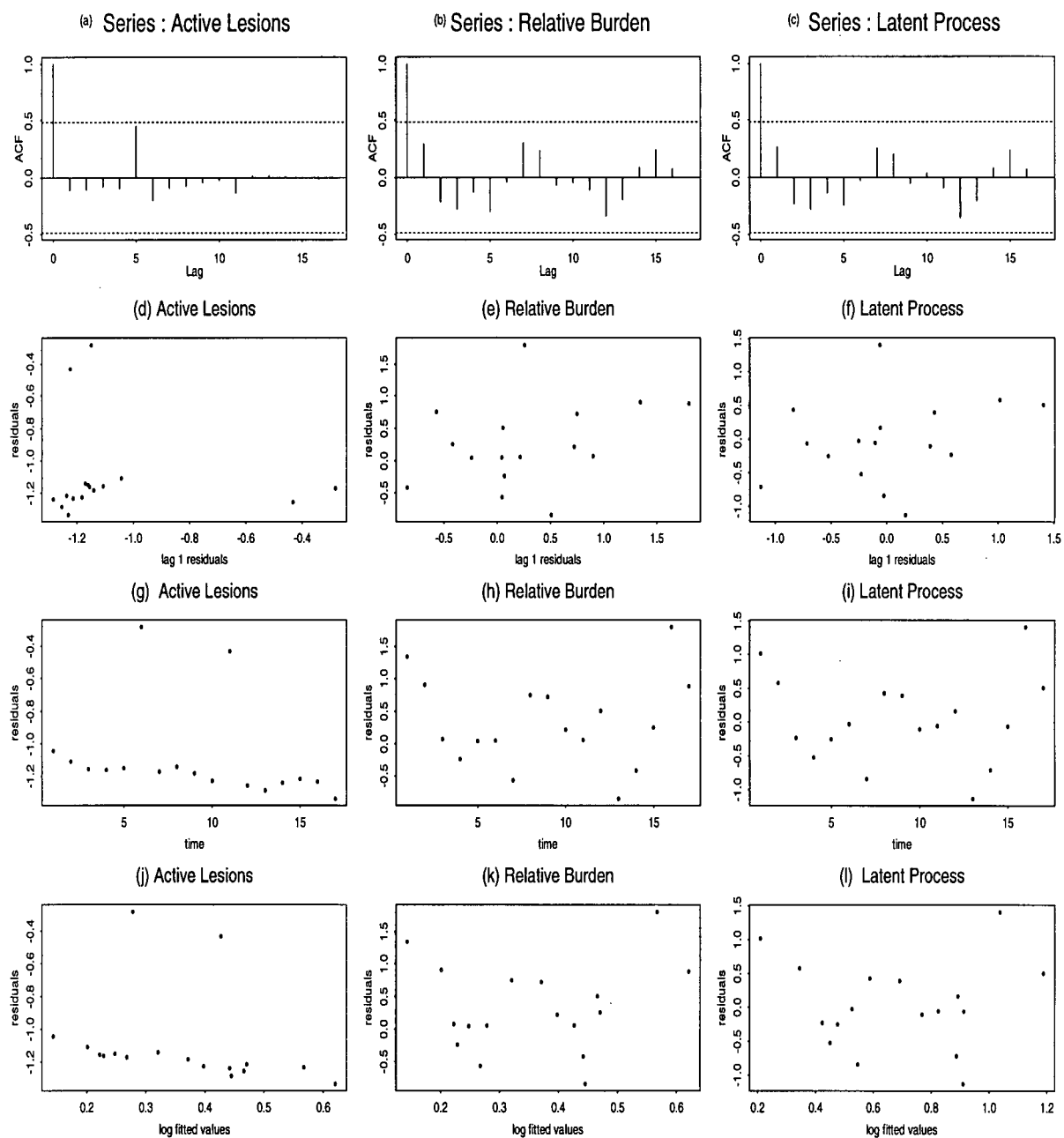


Figure D.54: Residuals plots for patient #544 in the low dose group

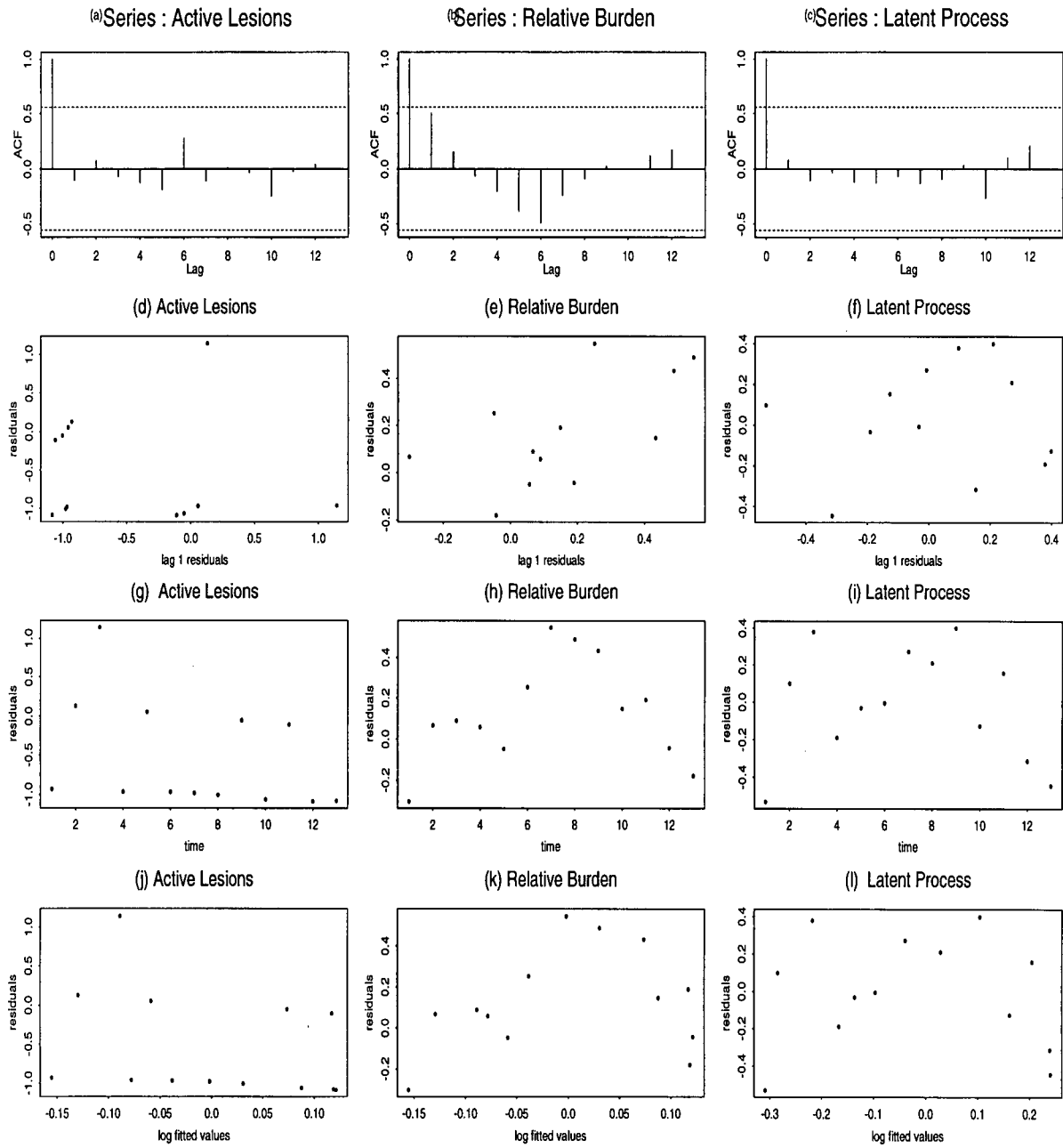


Figure D.55: Residuals plots for patient #547 in the low dose group

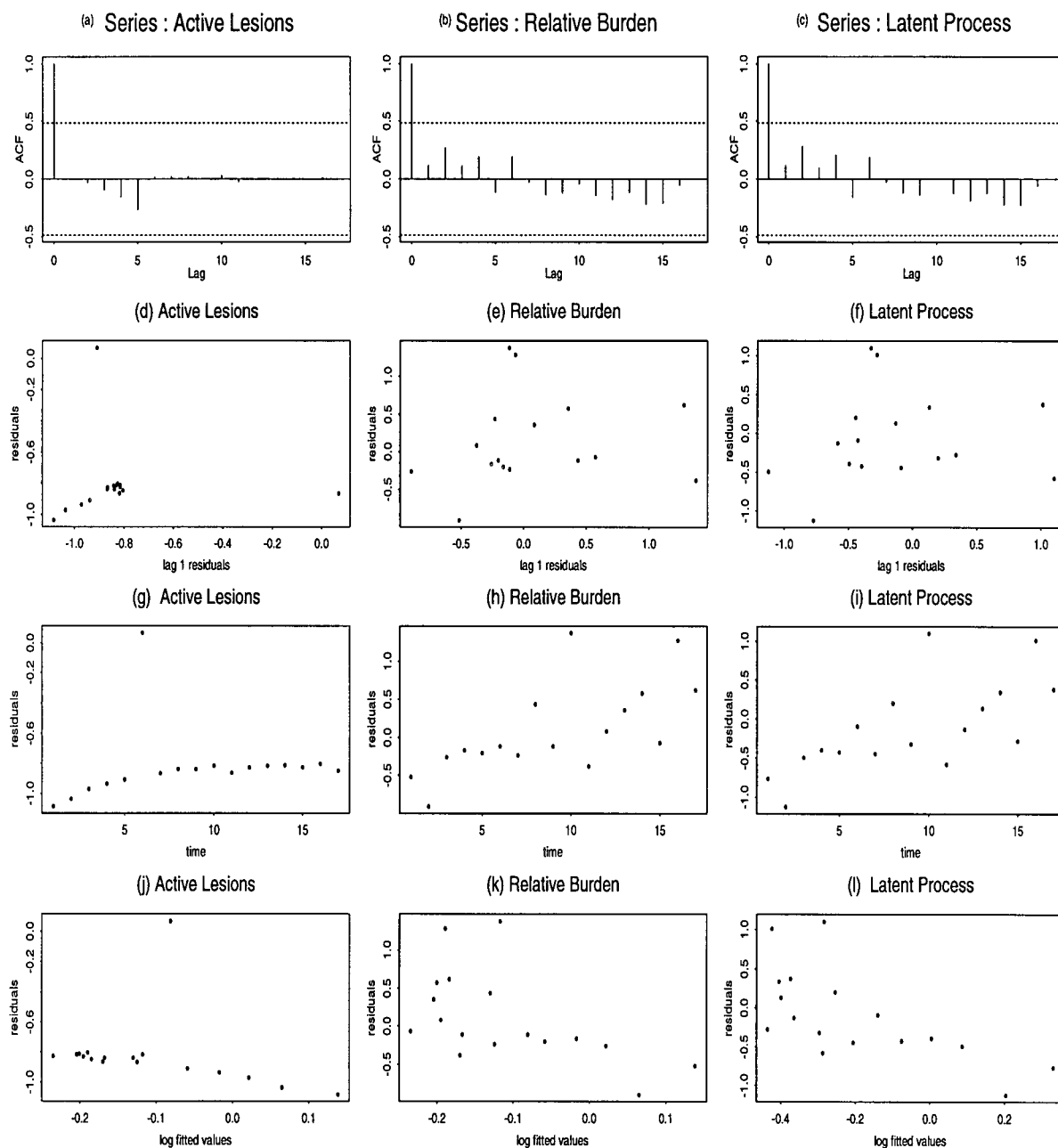


Figure D.56: Residuals plots for patient #548 in the low dose group

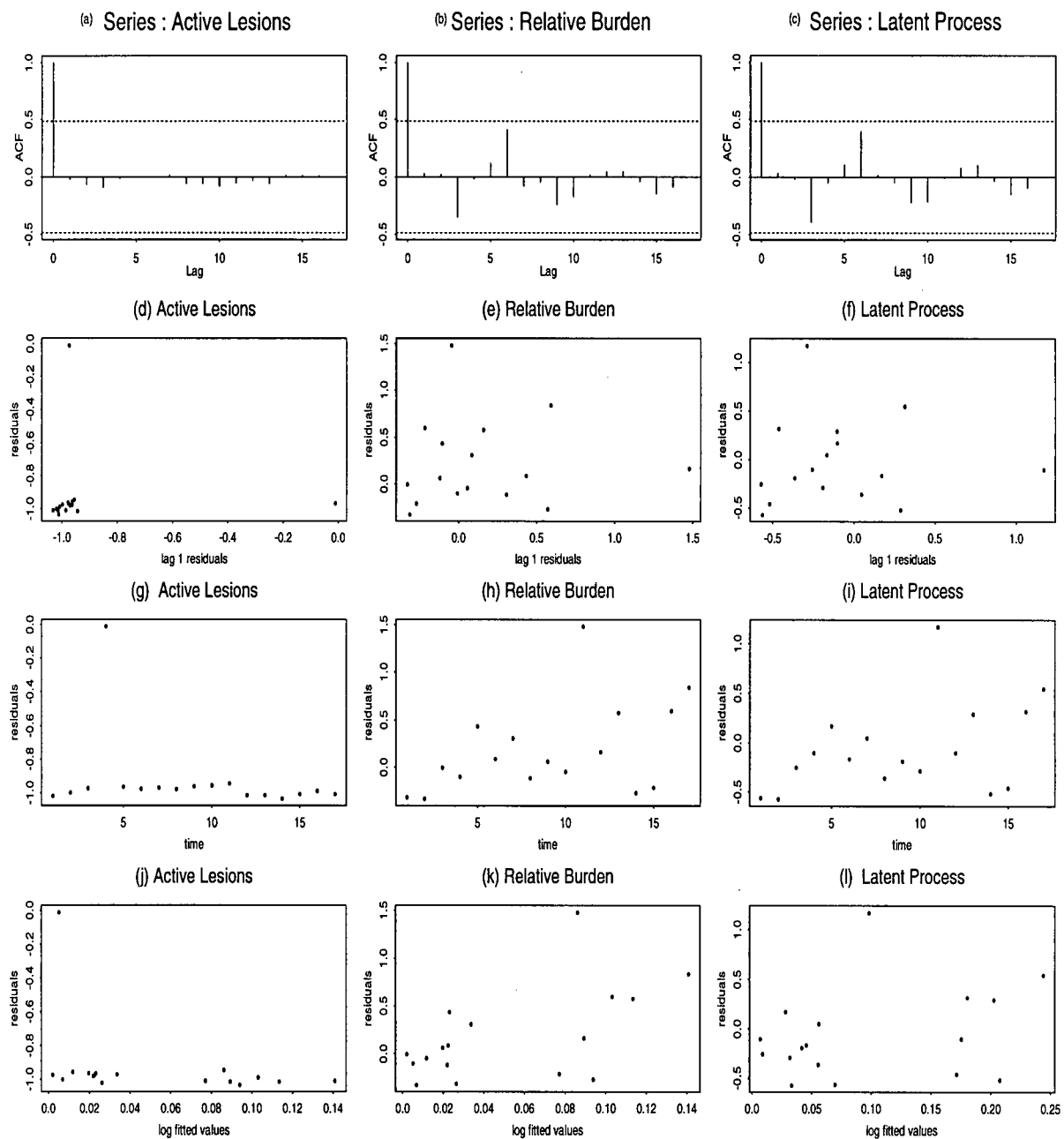


Figure D.57: Residuals plots for patient #564 in the low dose group

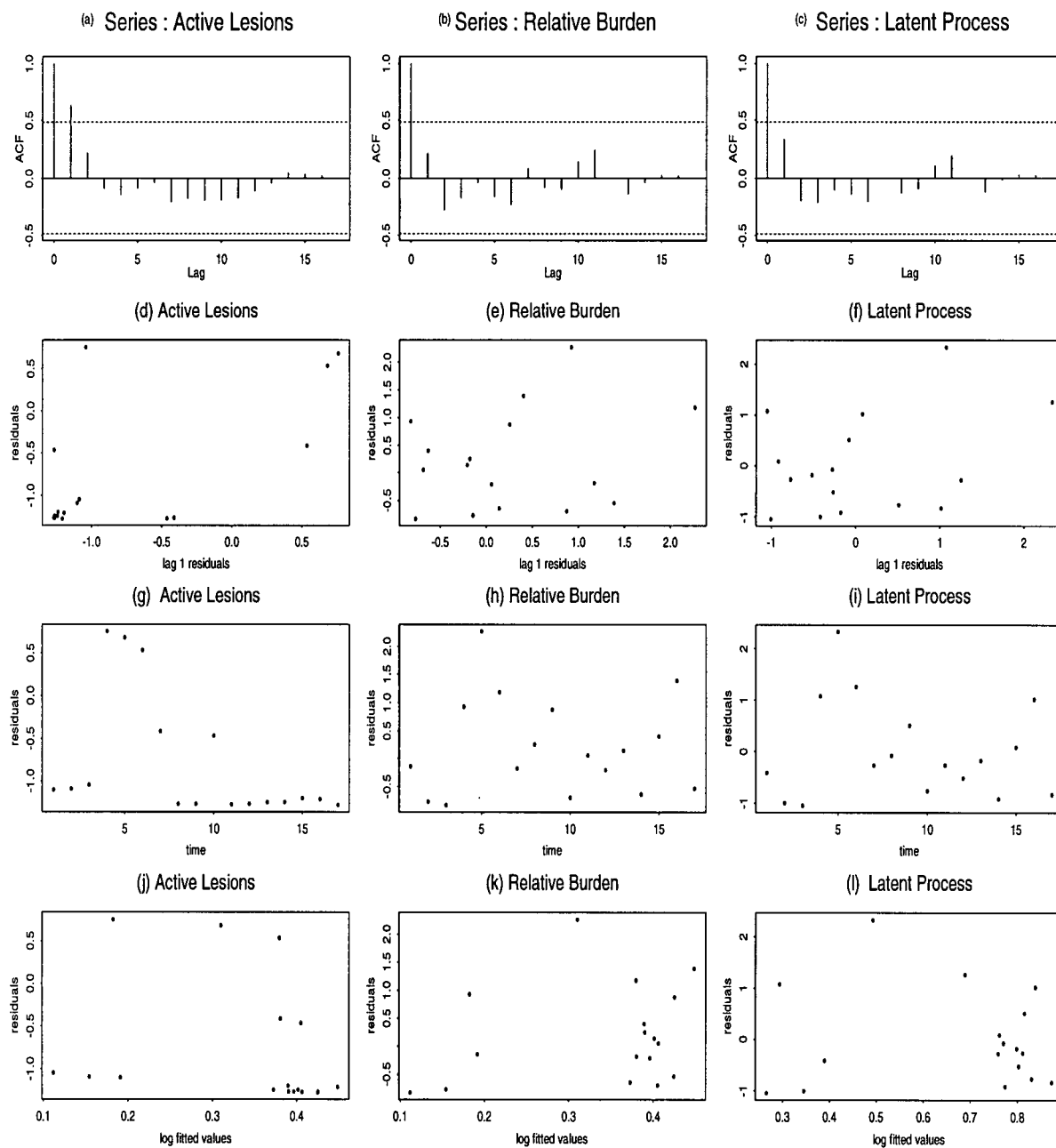
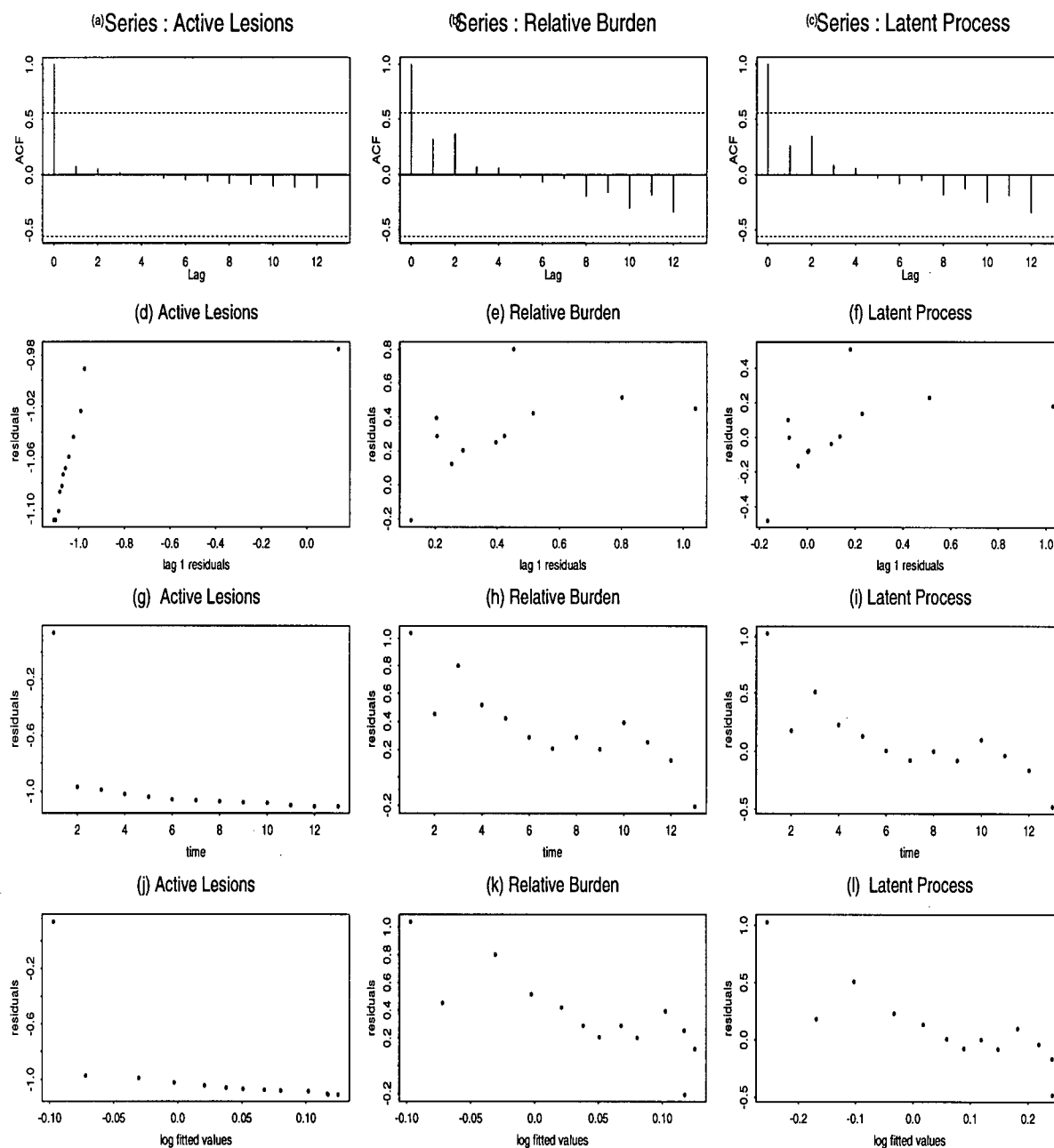


Figure D.58: Residuals plots for patient #568 in the low dose group



Appendix E

Residuals plots for each patient in the high dose group

Figure E.59: Residuals plots for patient #422 in the high dose group

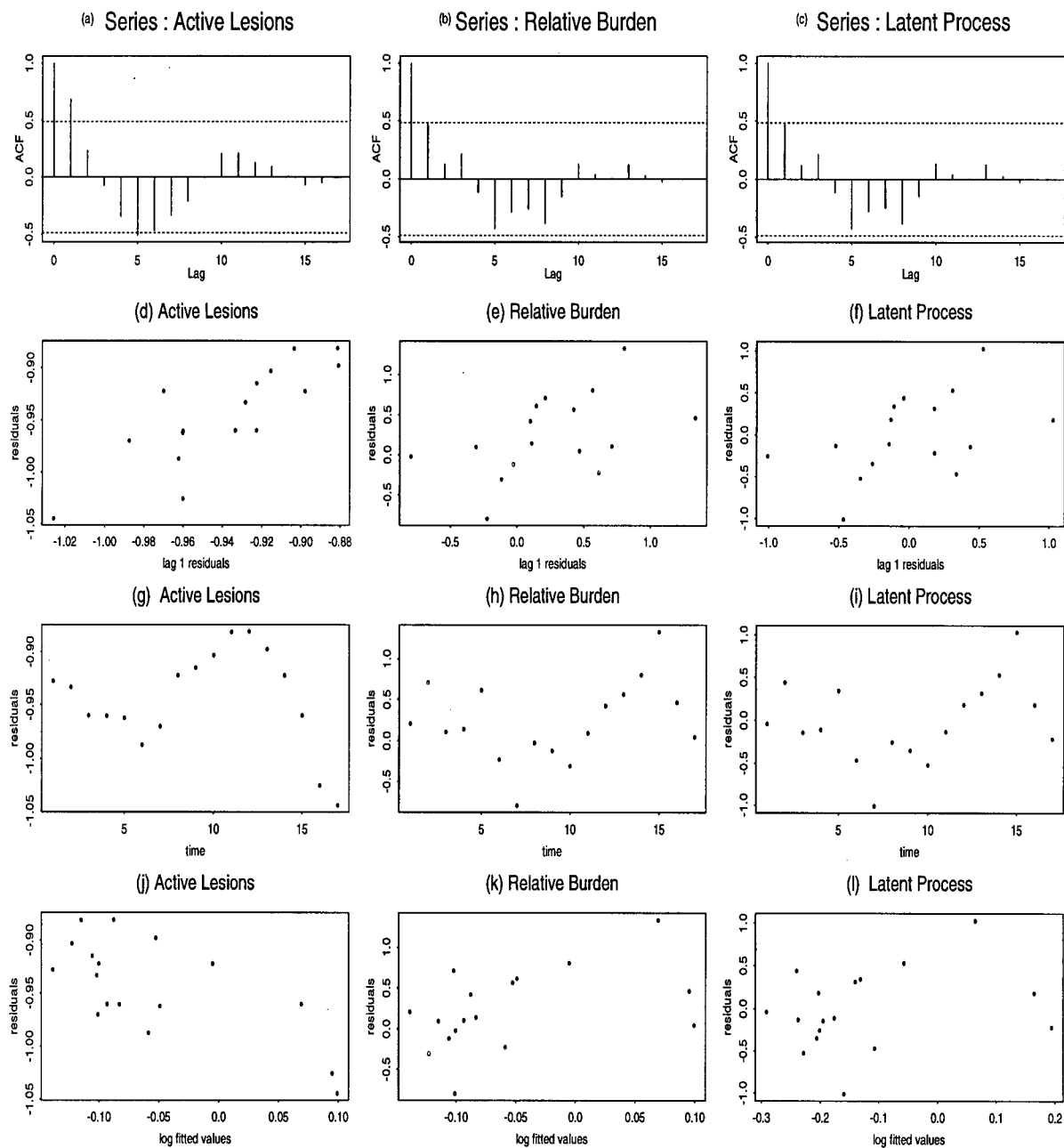


Figure E.60: Residuals plots for patient #444 in the high dose group

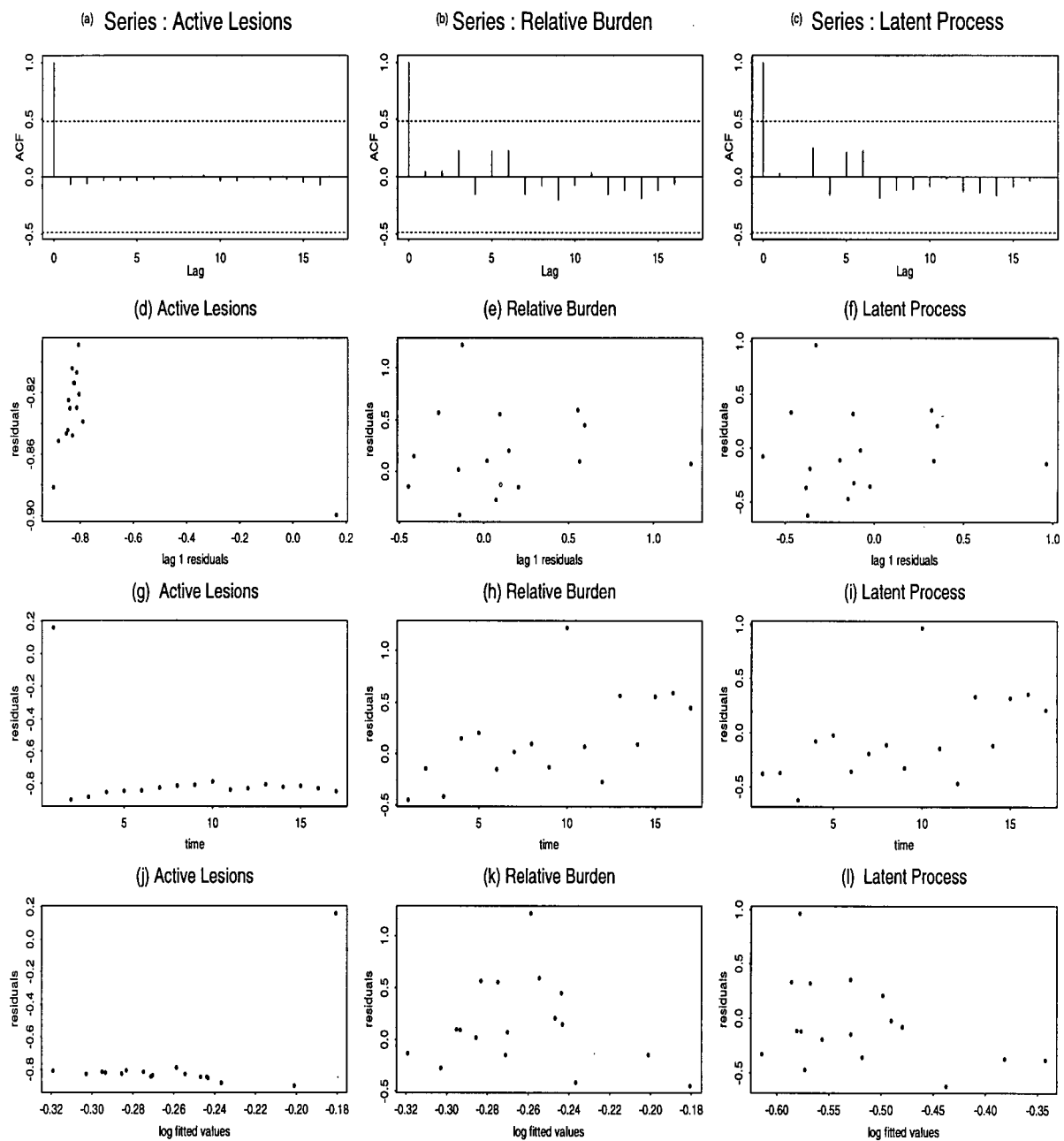


Figure E.61: Residuals plots for patient #445 in the high dose group

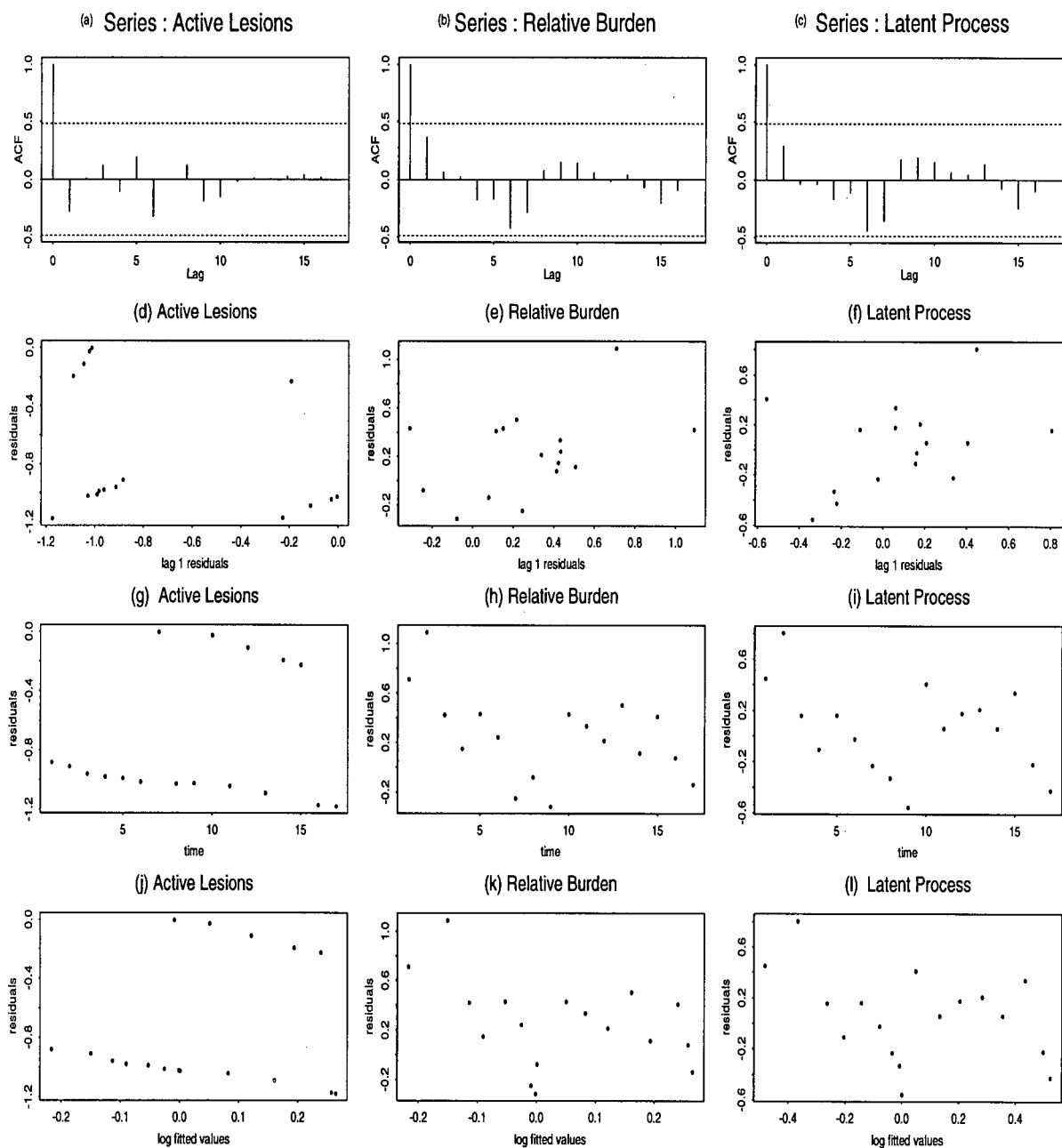


Figure E.62: Residuals plots for patient #453 in the high dose group

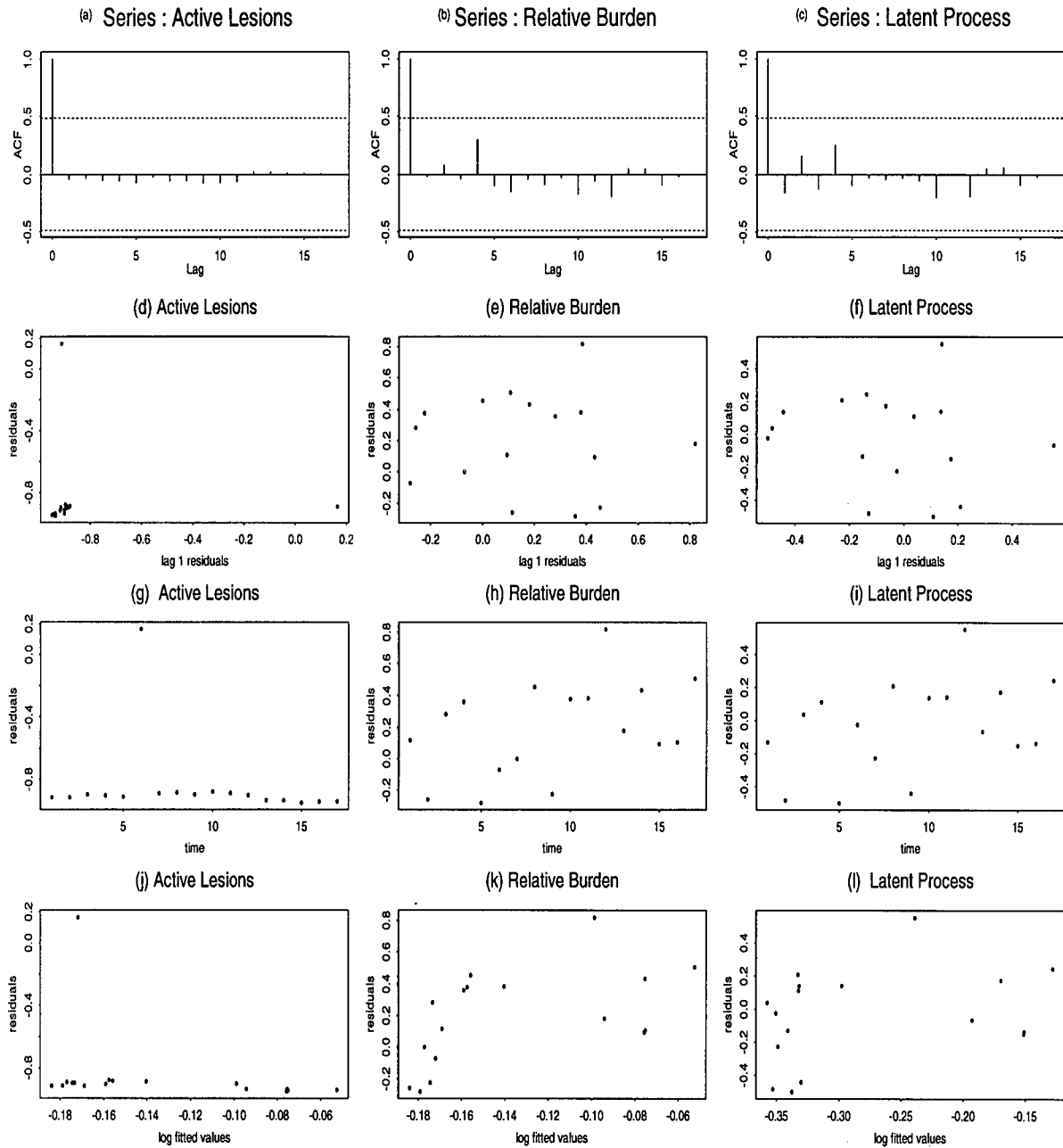


Figure E.63: Residuals plots for patient #454 in the high dose group

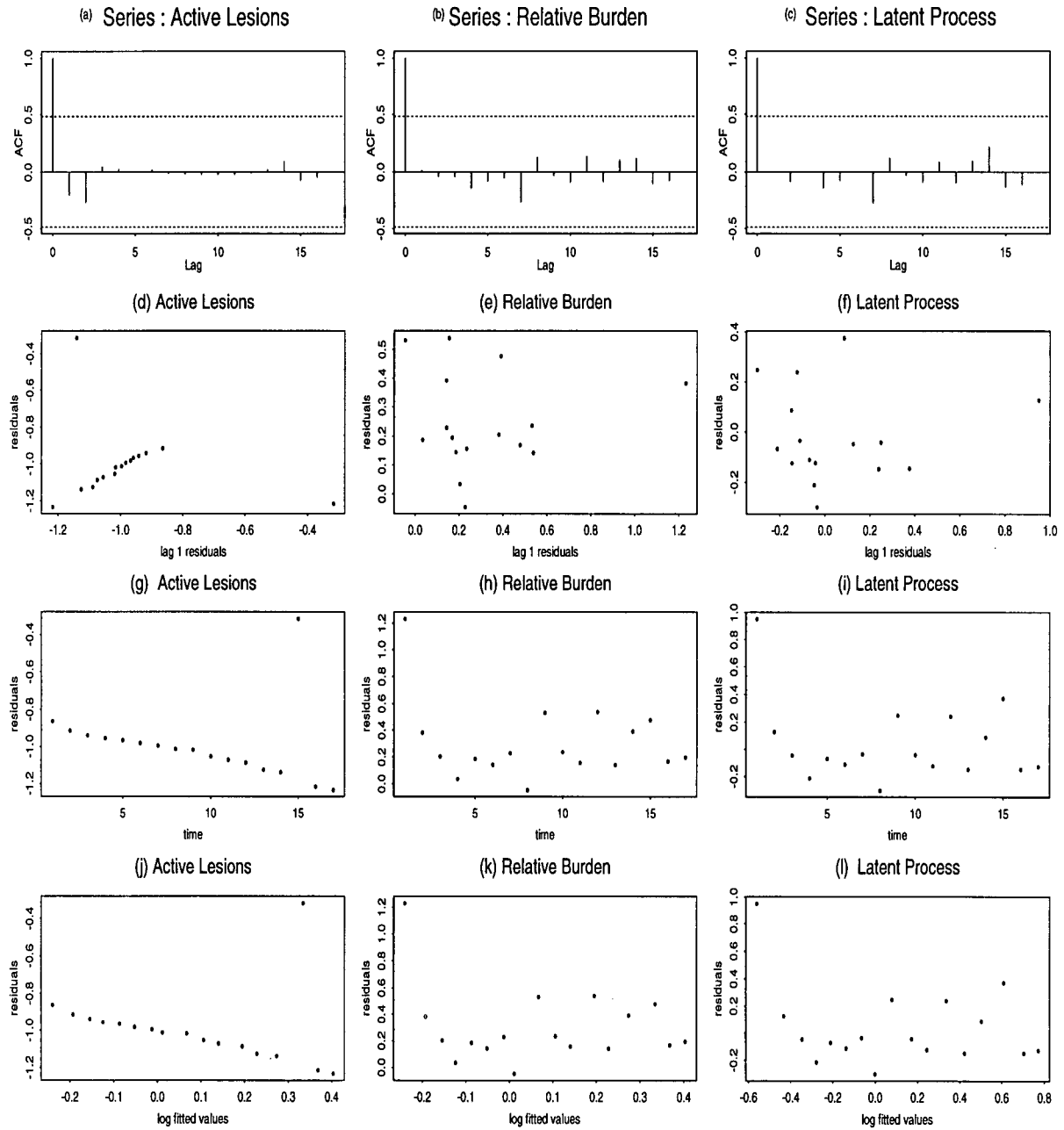


Figure E.64: Residuals plots for patient #497 in the high dose group

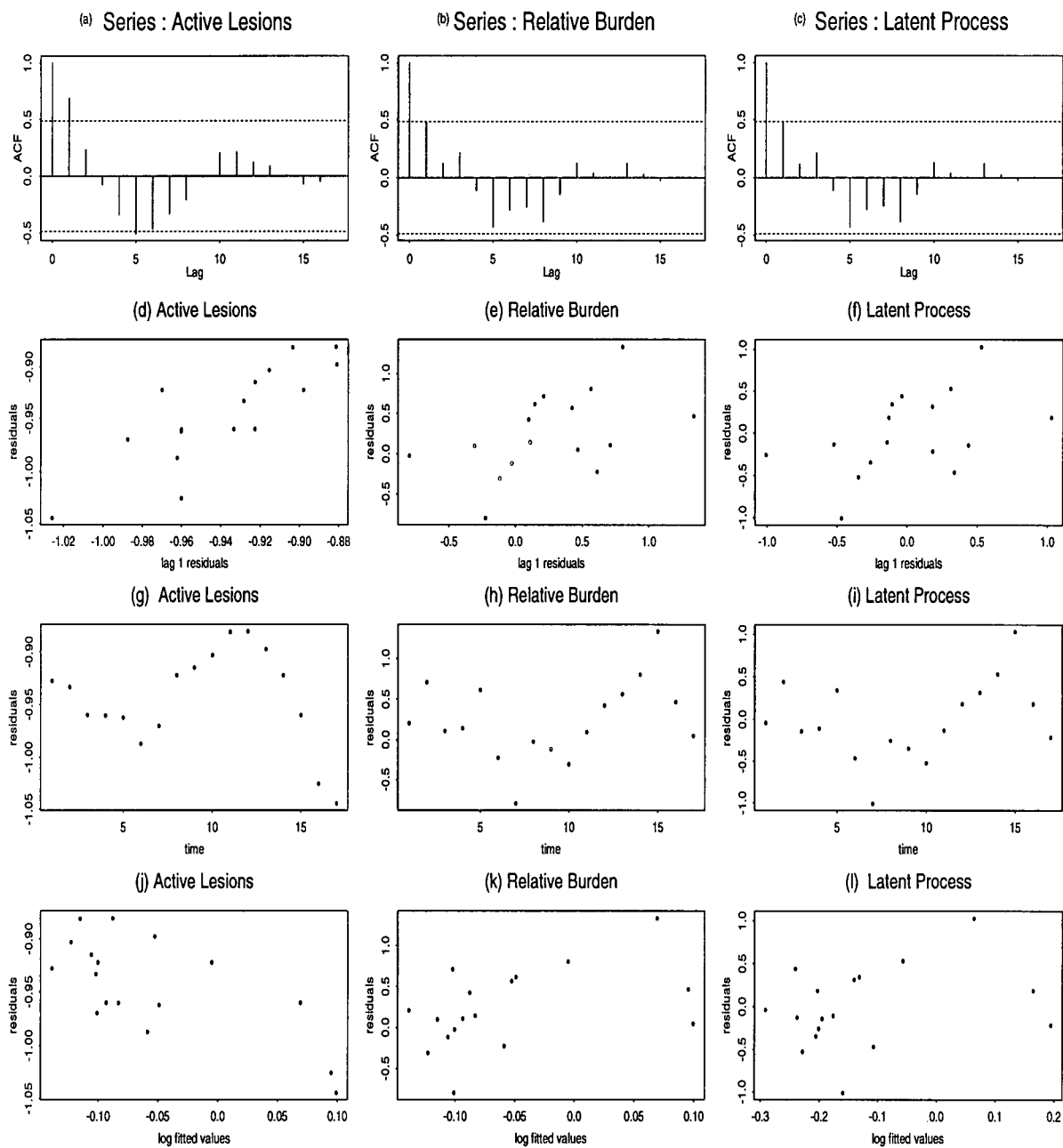


Figure E.65: Residuals plots for patient #503 in the high dose group

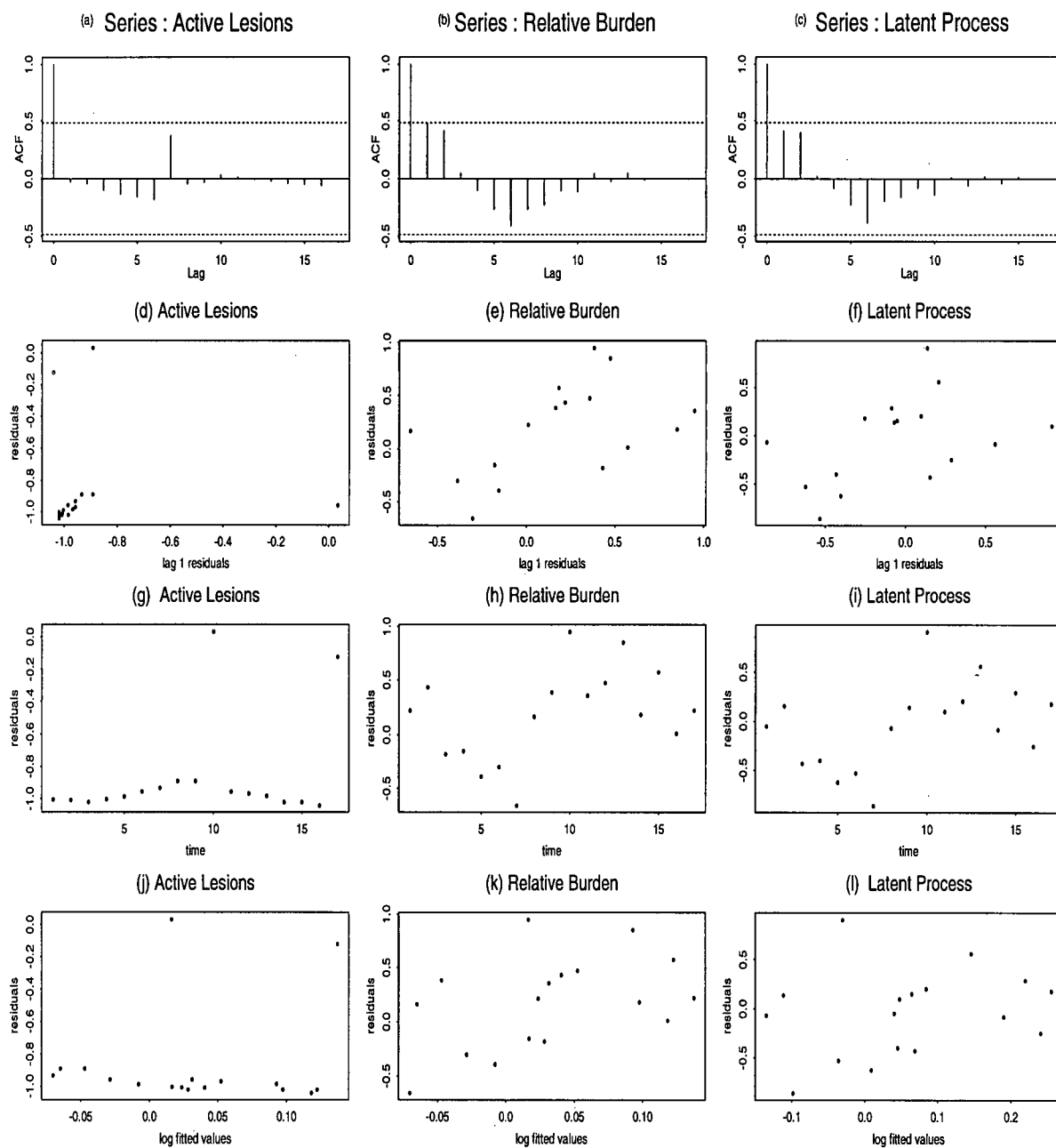


Figure E.66: Residuals plots for patient #506 in the high dose group

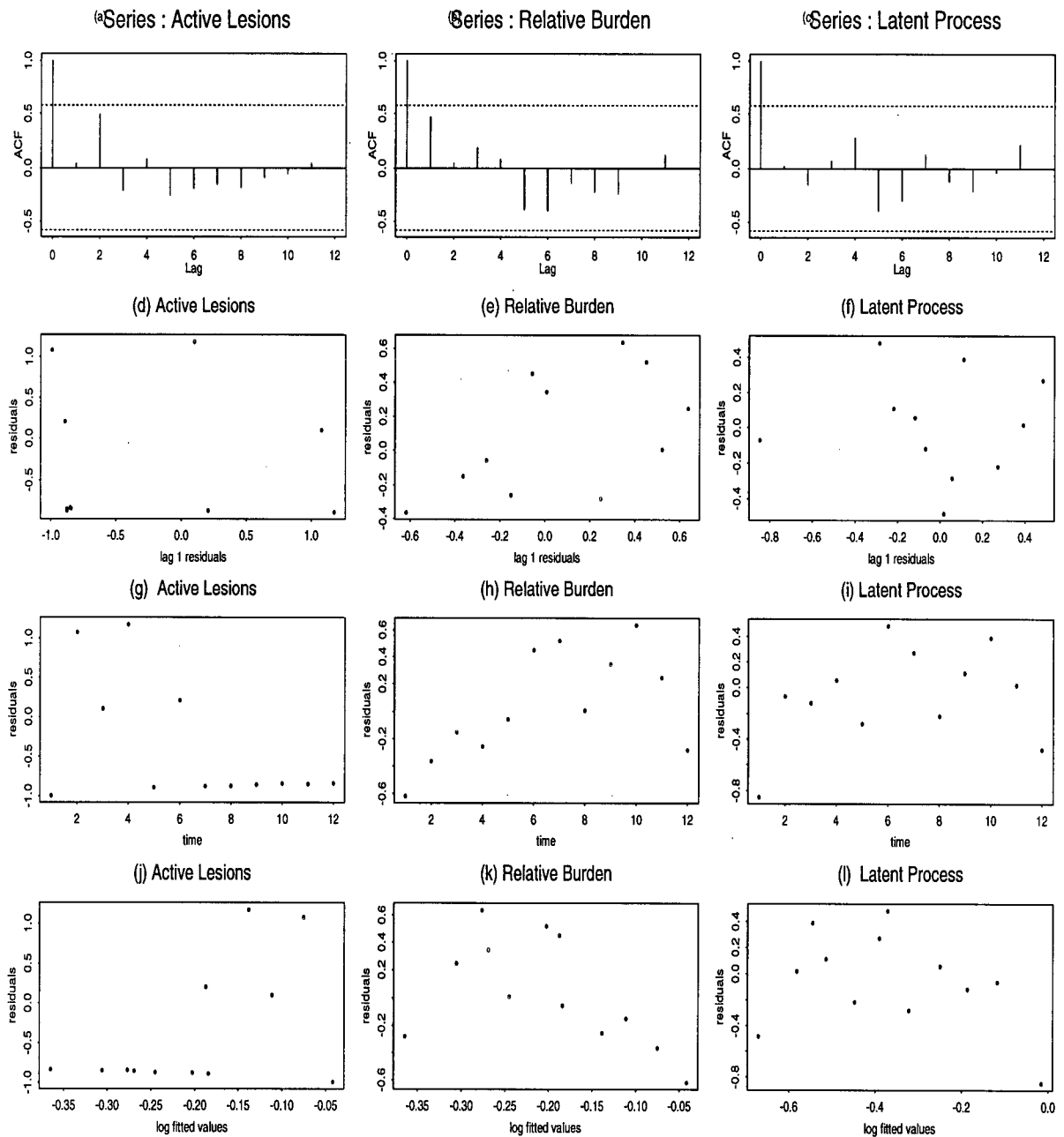


Figure E.67: Residuals plots for patient #524 in the high dose group

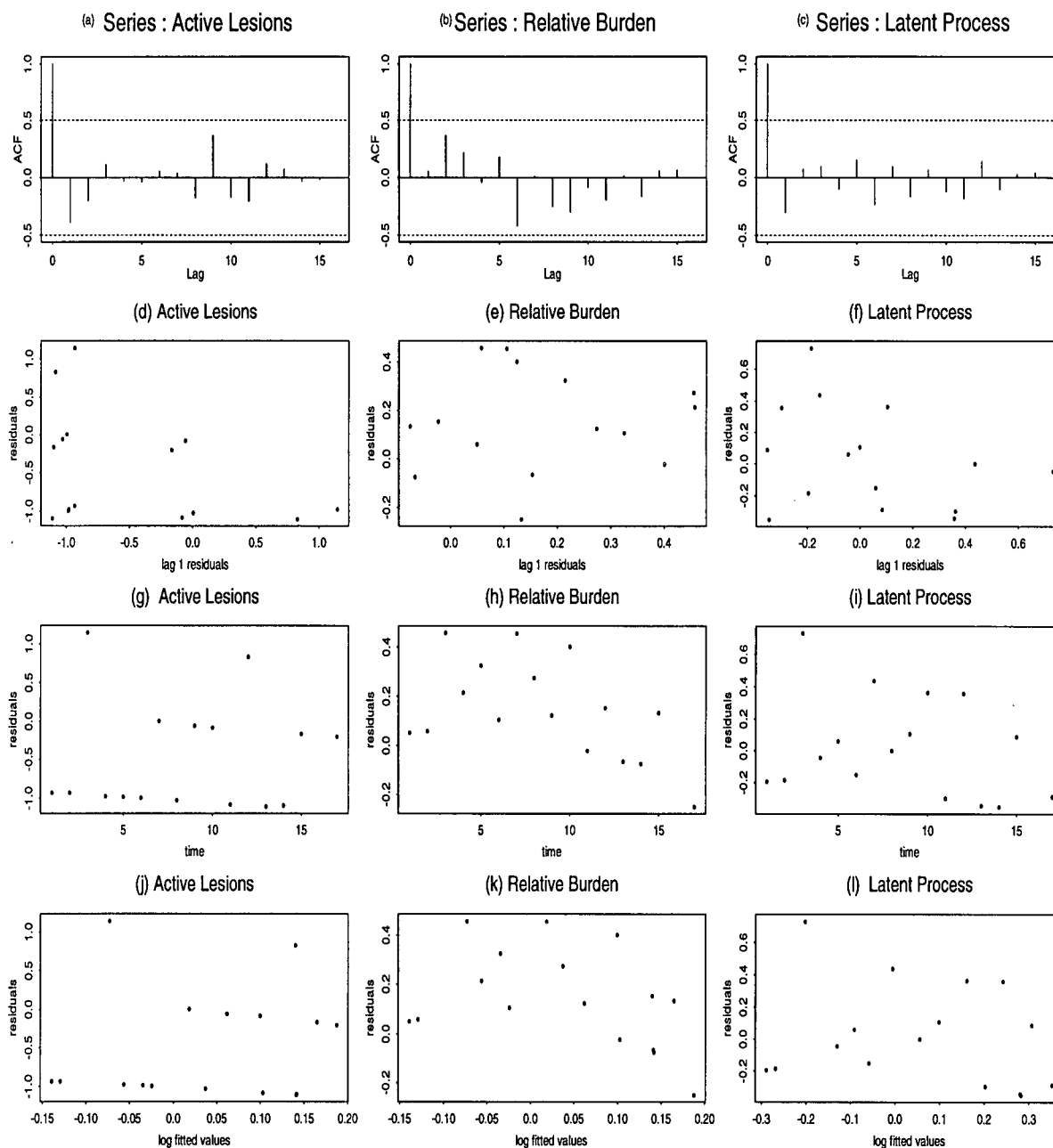


Figure E.68: Residuals plots for patient #526 in the high dose group

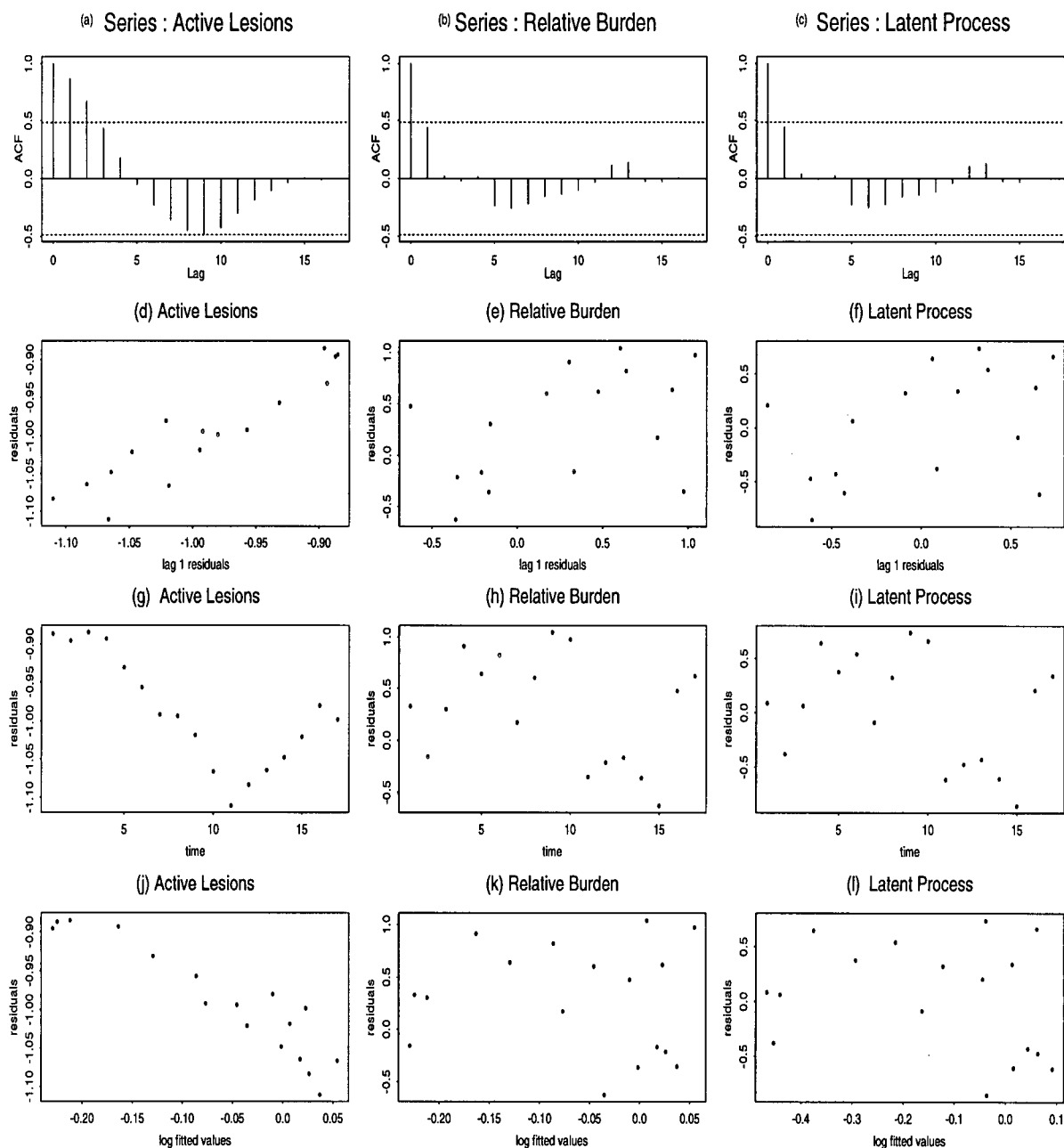


Figure E.69: Residuals plots for patient #541 in the high dose group

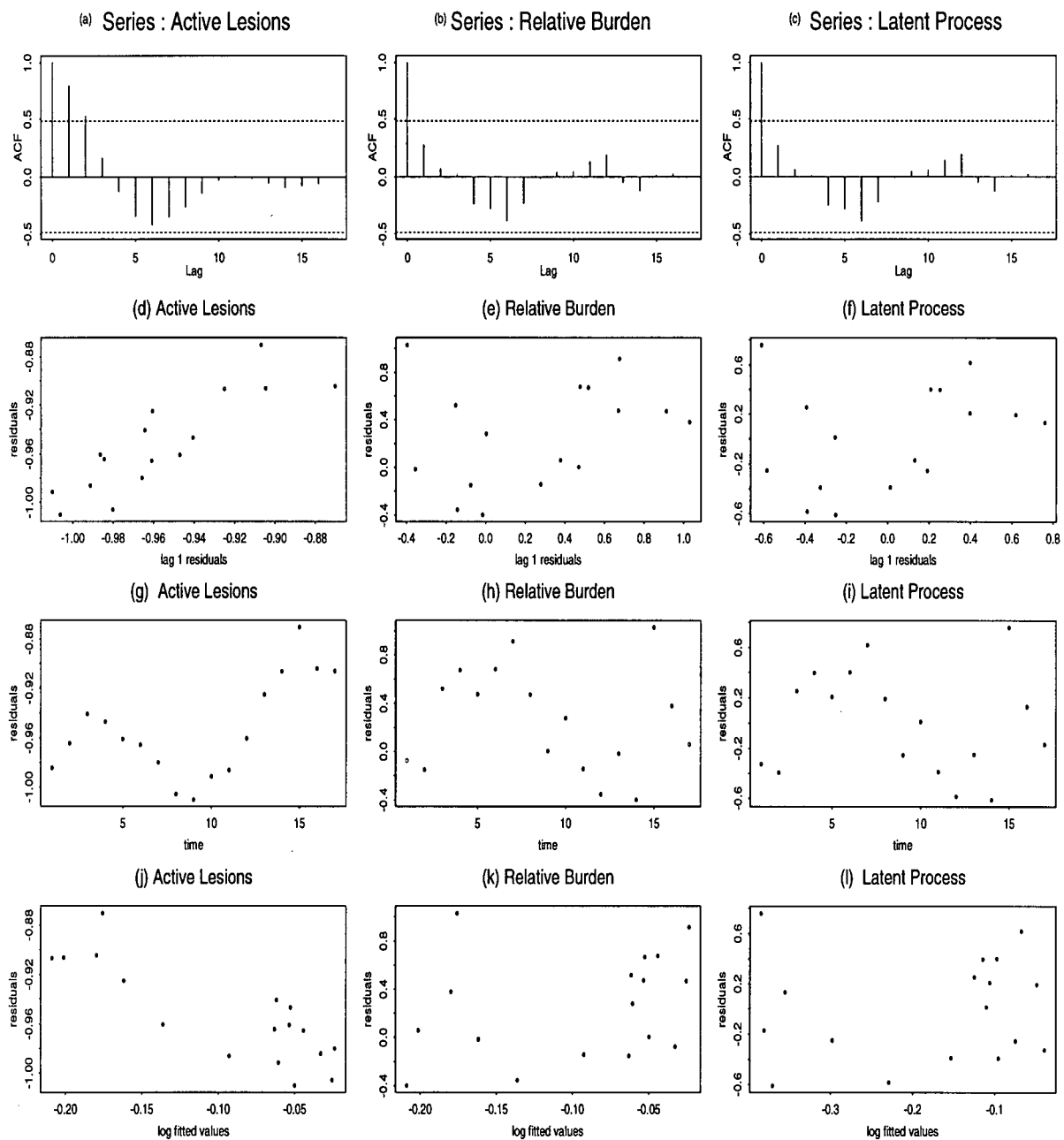


Figure E.70: Residuals plots for patient #543 in the high dose group

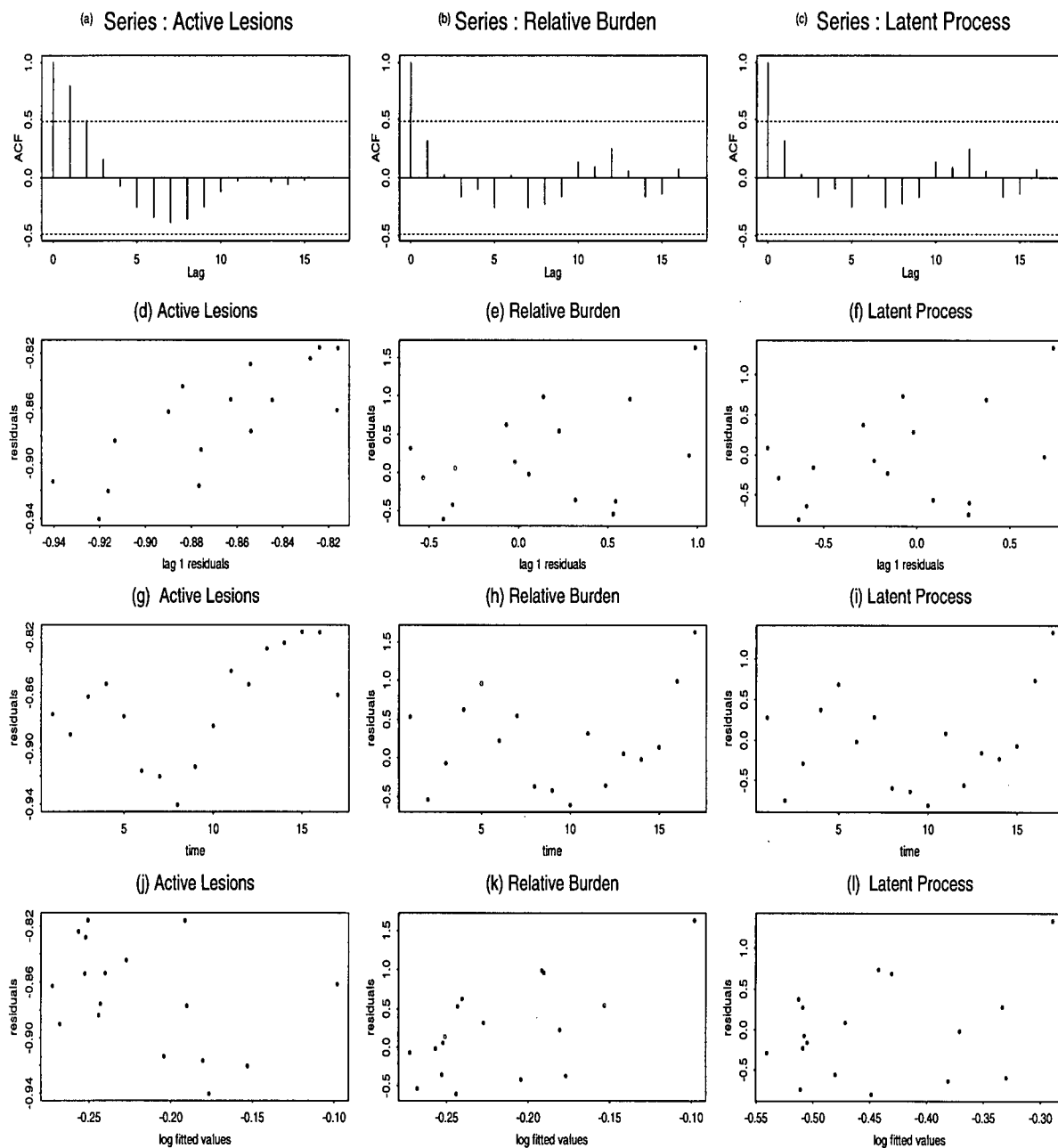


Figure E.71: Residuals plots for patient #546 in the high dose group

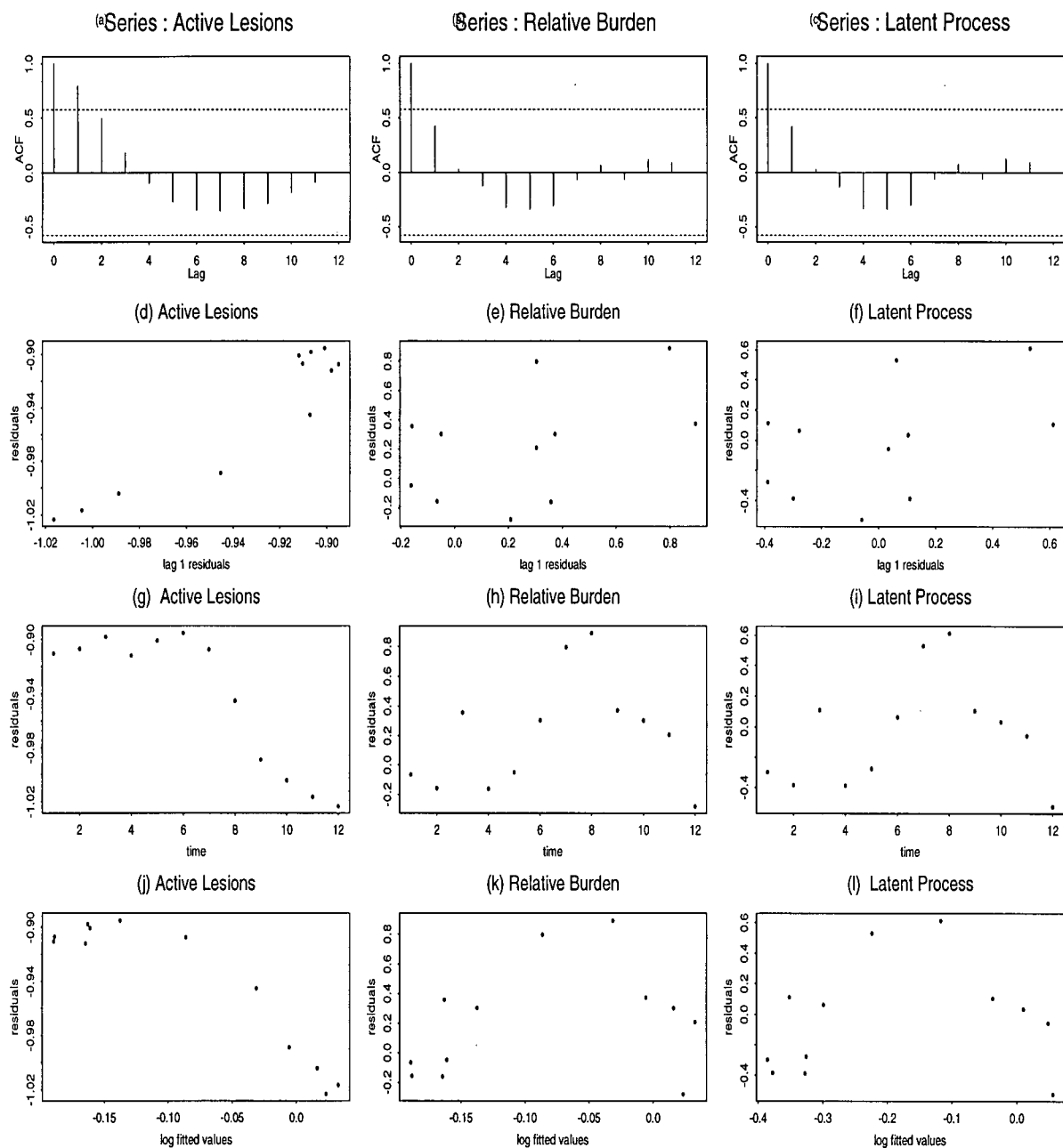


Figure E.72: Residuals plots for patient #549 in the high dose group

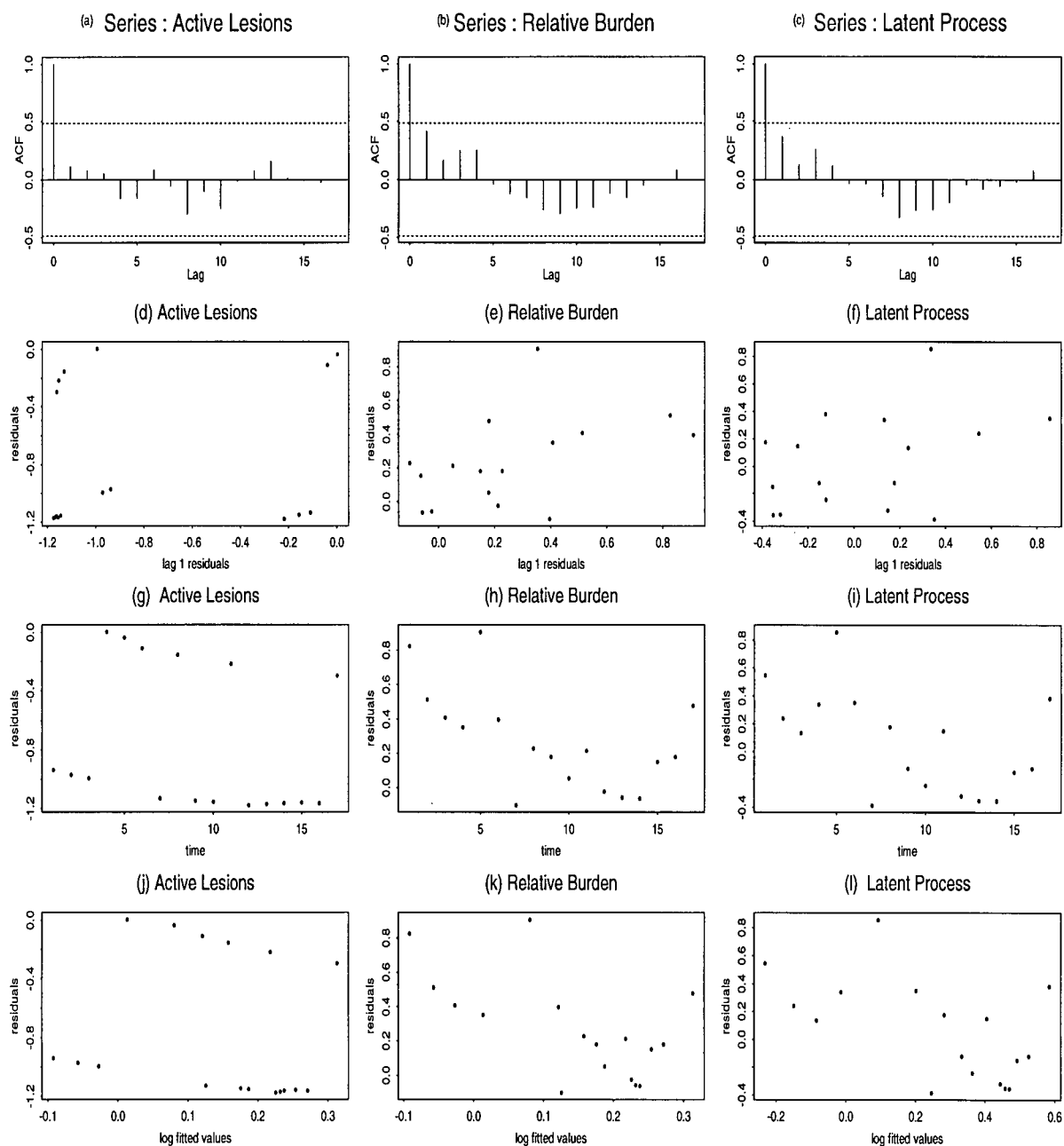


Figure E.73: Residuals plots for patient #566 in the high dose group

