Bayesian Modeling in Personalized Medicine with Applications to N-of-1 Trials

Ziwei Liao

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2021

Abstract

Bayesian Modeling in Personalized Medicine with Applications to N-of-1 Trials

Ziwei Liao

The ultimate goal of personalized or precision medicine is to identify the best treatment for each patient. An N-of-1 trial is a multiple-period crossover trial performed within a single individual, which focuses on individual outcome instead of population or group mean responses. As in a conventional crossover trial, it is critical to understand carryover effects of the treatment in an N-of-1 trial, especially in situations where there are no washout periods between treatment periods and high volume of measurements are made during the study. Existing statistical methods for analyzing N-of-1 trials include nonparametric tests, mixed effect models and autoregressive models. These methods may fail to simultaneously handle measurements autocorrelation and adjust for potential carryover effects. Distributed lag model is a regression model that uses lagged predictors to model the lag structure of exposure effects. In the dissertation, we first introduce a novel Bayesian distributed lag model that facilitates the estimation of carryover effects for single N-of-1 trial, while accounting for temporal correlations using an autoregressive model. In the second part, we extend the single N-of-1 trial model to multiple N-of-1 trials scenarios. In the third part, we again focus on single N-of-1 trials. But instead of modeling comparison with one treatment and one placebo (or active control), multiple treatments and one placebo (or active control) is considered.

In the first part, we propose a Bayesian distributed lag model with autocorrelated errors (BDLM-AR) that integrate prior knowledge on the shape of distributed lag coefficients and explicitly model the magnitude and duration of carryover effect. Theoretically, we show the connection between the proposed prior structure in BDLM-AR and frequentist regularization approaches. Simulation studies were conducted to compare the performance of our proposed BDLM-AR model with other methods and the proposed model is shown to have better performance in estimating total treatment effect, carryover effect and the whole treatment effect coefficient curve under most of the simu-

lation scenarios. Data from two patients in the light therapy study was utilized to illustrate our method.

In the second part, we extend the single N-of-1 trial model to multiple N-of-1 trials model and focus on estimating population level treatment effect and carryover effect. A Bayesian hierarchical distributed lag model (BHDLM-AR) is proposed to model the nested structure of multiple N-of-1 trials within the same study. The Bayesian hierarchical structure also improves estimates for individual level parameters by borrowing strength from the N-of-1 trials of others. We show through simulation studies that BHDLM-AR model has best average performance in terms of estimating both population level and individual level parameters. The light therapy study is revisited and we applied the proposed model to all patients' data.

In the third part, we extend BDLM-AR model to multiple treatments and one placebo (or active control) scenario. We designed prior precision matrix on each treatment. We demonstrated the application of the proposed method using a hypertension study, where multiple guideline-recommended medications were involved in each single N-of-1 trial.

Table of Contents

List of Figures

List of Tables

Acknowledgements

There are many who helped me along the way on this journey. I want to take this opportunity to thank them. Foremost, I would like to express my upmost gratitude to my advisor Prof. Ying Kuen K. Cheung for his guidance, patience, enthusiasm, and immense knowledge. His plentiful experience and immense knowledge have helped me in every building block of this dissertation and encouraged me in all the time of my academic research and daily life.

I would also like to thank the rest of my committee members: Professor Bin Cheng, Professor Min Qian, Professor Jianhua Hu, Professor Prof. Ian M. Kronish, for kindly agreeing to serve on my committee and offering tremendous suggestions on model improvement, applications and interpretations in the real data examples.

My sincere thanks also go to Dr. Matilde Kam, Dr. Mat Soukup, and Dr. Jiawen Zhu, for offering me the summer internship opportunities in their groups and leading me working on diverse exciting research projects in industry.

Last but not the least, I would like to thank my family and friends for their consistent encouragement and supporting me spiritually throughout my life. This dissertation is dedicated to them.

Chapter 1: Introduction

1.1 Overview of N-of-1 Trial

Traditional parallel-group randomized controlled trials (RCTs) are the most prevalent study design to identify the population-level treatment effect in evidence-based medicine. However, clinical evidence generated from RCTs often has limited generalizability of the results (Greenfield et al., 2007), especially if efficacy is indicated from a "successful" trial. In the United States, the top ten highest-grossing drugs will only benefit between 1 in 25 and 1 in 4 of the patients who used them (Schork, 2015). One of the main reasons of ineffective treatment is the existence of heterogeneity of treatment effects (HTE), i.e., the variability in the direction and magnitude of treatment effects for single patient (Kent et al., 2010; R. L. Kravitz et al., 2004). RCTs provide estimate of average treatment effect for whole participants in the study rather than individual treatment effect (ITE) for each participant. The variation in treatment effect may trigger a mix of enhanced, weakened or even negative treatment benefits.

This issue has led growing interest in different types of clinical trials that focuses on individual, not average, responses to treatment. A lot of research that focuses on personalizing medicine have been done, and regulatory agencies such as the US Food and Drug Administration (FDA) has evolved its regulatory processes in response to scientific developments that are critical for the development of personalized therapeutics and diagnostics (FDA, 2013; Hamburg and Collins, 2010). There are mainly two approaches to match patients with potential effective treatment. The first one is subgroup analysis, where large population is partitioned or stratified into several smaller groups based on a shared characteristic, such that patients in certain stratified categories are more likely

to benefit from a given treatment (Rothwell, 2005; Tsapas and Matthews, 2009). The alternative method is N-of-1 trial, where individual patient is the entire trial. The patient receives multiple interventions and the outcome is compared within the patient to identify the optimal treatment for that patient. Both subgroup analysis and N-of-1 trials are data driven method to address treatment heterogeneity problem, but the former aimed at selecting the most promising subgroups for a given intervention while the latter targeted the most promising course of therapy for a given participant.

N-of-1 trials were first proposed by Guyatt in 1986 (G. Guyatt et al., 1988; G. Guyatt et al., 1986) to address the difficulty in extrapolating the results obtained from traditional multi-patient, double-blinded, randomized trials to individual patients. N-of-1 trials are prospectively multiple-period crossover trials comparing two or more interventions within single participant. Unlike subgroup analysis, N-of-1 trials can estimate ITE directly for each patient, making it in accord with the ultimate goal of personalizing medical care. In N-of-1 trials, most participants will immediately benefit from the study if one of the treatments stands out. This is an important merit of N-of-1 trials since participants in population based trials may only be assigned to placebo group during the entire study.

1.1.1 Application of N-of-1 Trials

N-of-1 trials are applicable to evaluating treatments for chronic conditions with stable treatment outcomes and quick onset of treatment effect. In addition, N-of-1 trials are especially useful in studying rare diseases, where participants recruitment is usually challenging (Stunnenberg et al., 2018). However, similarly as standard parallel-group crossover trials, they may fail to work on acute or unrelentingly progressive, permanent or slowly reversible treatment outcomes (Duan et al., 2013; S. S. Senn, 2002). Mirza et al. (2017) summarized the situations where N-of-1 is more appropriate than conventional randomized trials:

- Quickly acting symptomatic treatment with large response variability (e.g. Samuel et al., 2019).
- Rare diseases which is infeasible for large parallel group RCTs (e.g. Stunnenberg et al.,

2018).

- Patients who are substantially different from subjects in existing trials.
- Patients undergoing long-period of treatment and lack of information for for further treatment (e.g. Liu et al., 2016).

One major concern of N-of-1 trials is that the costs and logistical efforts are probably too high for individualized patient care (R. L. Kravitz et al., 2009). But with the advent of efficient and cheap medical monitoring devices, such as non-invasive/wearable sensors for medical metrics (Topol, 2010) and mobile applications (R. L. Kravitz et al., 2018), data collection process for N-of-1 trials is getting more feasible and reliable.

Another emerging application area of N-of-1 trials is the use of mobile and wireless health (mHealth) technologies to increase access to health services and lower health cost (Nilsen et al., 2012). Smartphones have been utilized to track patients' outcomes and physical activity (Glynn et al., 2013; Luxton et al., 2011) and deliver patient education through software applications (apps, Wang et al., 2017). N-of-1 trials have been used in several mHealth studies to enhance a more personalized health care (Whitney et al., 2018). From the population perspective, mHealth-based N-of-1 trial has the potential to achieve faster therapeutic success and more convenient health services access (Barr et al., 2015).

1.1.2 Design of N-of-1 Trials

Design features of N-of-1 trials are similarly as those of grouped randomized crossover trials, but on an individual level. The unit of treatment assignment is a prespecified treatment period block, during which same treatment is assigned to the patient. The duration of treatment period block is selected to enable that each treatment can reach full effect (Duan et al., 2013). Within each treatment period block, clinical outcomes are expected to be collected in a frequent way, such as hourly, daily or other regular time period collection, which renders that data from N-of-1 trial has a time-series structure. The sequence of assignment can either be randomized or predetermined. For

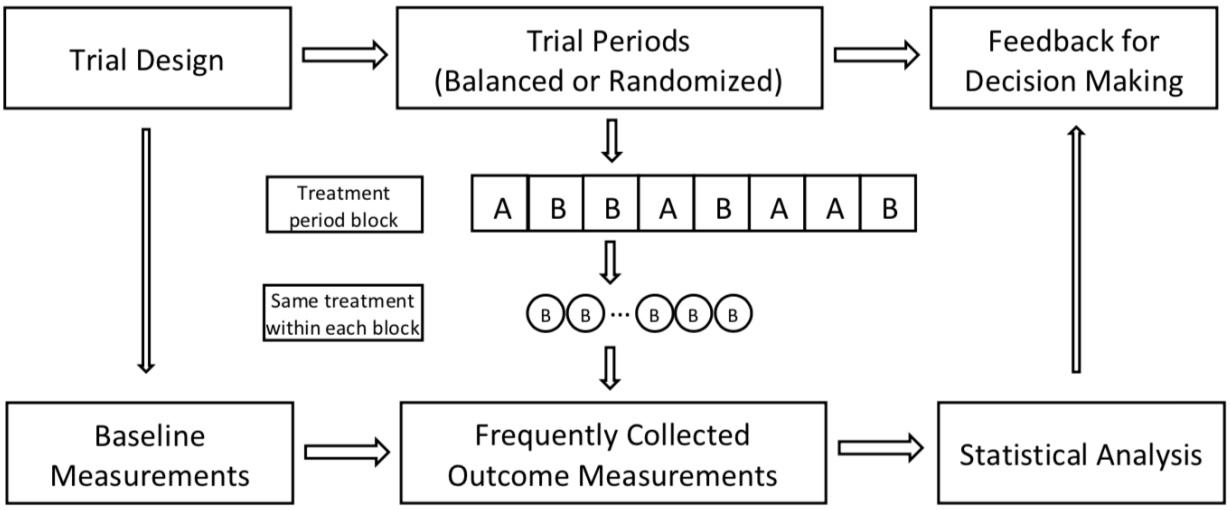


Figure 1.1: Scheme of N-of-1 trials (modified from R. Kravitz et al. (2014))

example, the paired design ABABABAB and the singly counterbalanced design ABBAABBA can protect against temporally linear confounders, while doubly counterbalanced design ABBABAAB offer better protection against both linear and nonlinear confounders (R. Kravitz et al., 2014). An example of N-of-1 trial procedures is illustrated in Figure 1.1.

Unlike standard RCTs, where their sample size is counted as the number of participants in each treatment group, the sample size of N-of-1 trials is the total number of measurements taken on each participant. Total number of measurements is further determined by the number of treatment periods and number of measurement within each period. Adding number of treatment periods or extending the duration of treatment periods can increase study sample size.

Washout period can be used between treatment periods to mitigate the potential carryover effect of treatment used in the previous time period. But the length of washout period is hard to determine. Short length of washout period can not grantee the elimination of carryover effect, while long length of washout period may compromise patient safety since patients are temporarily off all treatments during the trial (Lillie et al., 2011).

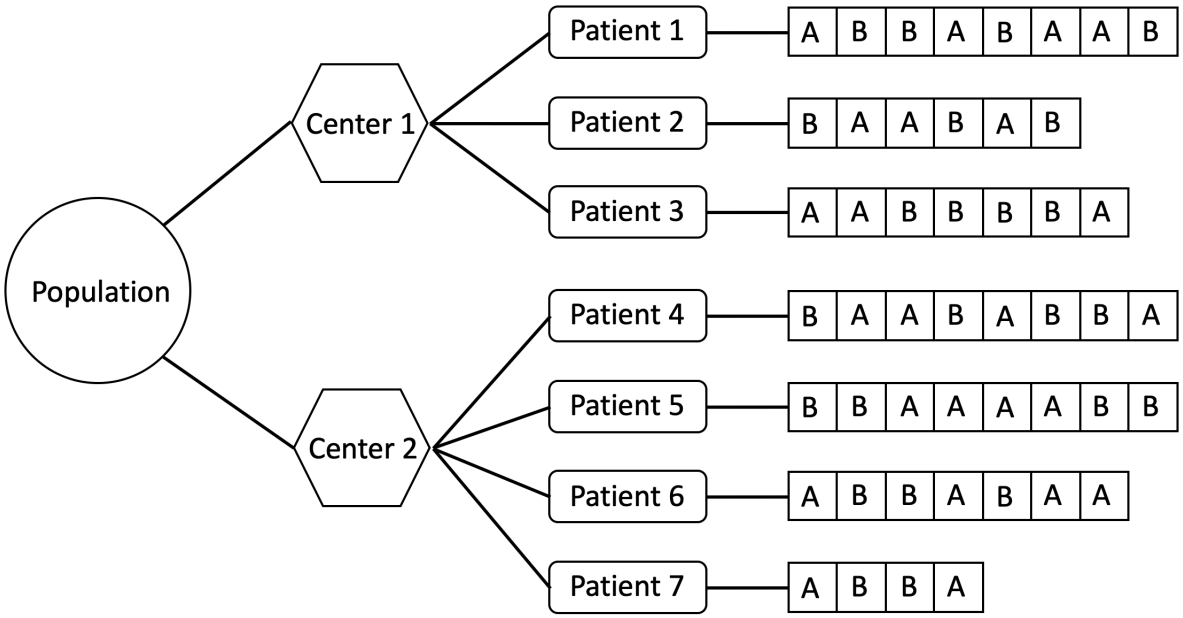


Figure 1.2: Scheme of multiple N-of-1 trials

1.1.3 Combining Multiple N-of-1 Trials

Individual N-of-1 trials are designed to compare interventions within specific patient and make suggestion on personal health service. Traditional randomized clinical trials are utilized to obtain information on differences between interventions in a study population. When data from multiple individual N-of-1 trials are available, they can be aggregated together to simultaneously estimate the population and individual level treatment effect. From population perspective, multiple N-of-1 trials can provide a population estimate on treatment effectiveness while adjusting for heterogeneity of treatment effect. From individual patient's perspective, the combined N-of-1 trials can learn about within and between patient variation and improve the clinical decision on individuals. It is especially useful when individual patient just starts a trial and has only a few observations or even does not participate in the trial, since clinicians can "borrow strength" from other patients in the trial if one believes the existence of similarities between patients. The schematic representation of population groupings for multiple N-of-1 trials is illustrated in Figure 1.2.

1.1.4 Current Methods to Analyze N-of-1 Trials

Individual N-of-1 Trials

Many statistical methods have been used in analyzing N-of-1 trials, but no clear consensus has been reached so far. Gabler et al. (2011) conducted a systematic review of N-of-1 trials published between 1985 and 2010. Among a total 108 trials reporting on 2,154 participants, most trials reported more than 1 method of data analysis. Approximately half of the trials reported using a t-test (44%), 52% reported using a visual/graphical comparison with no statistical comparison. Only 17% of the trials applied any type of regression analysis and 22% trials reported using nonparametric statistics. Three statistical methods will be introduced bellow.

(1). Nonparametric tests

Early stage statistical analysis of N-of-1 trials used nonparametric tests such as sign test (G. H. Guyatt et al., 1990), Wilcoxcon signed rank test, chi-square test, et al. Treatment information is gathered from day level to treatment block/period level. For each treatment block, we will only have a binary response, representing whether one treatment is better than the other and omit the size of difference in treatment effect.

(2). Analysis of variance (ANOVA) type tests (G. H. Guyatt et al., 1990)

ANOVA type tests (t test as a special case) are also used in analyzing N-of-1 trials, which utilize the information of actual size of treatment effects. These type of tests also use treatment block level information and assume treatment blocks are mutually independent, neglecting potential carryover effect and autocorrelation between consecutive treatment blocks. The normality assumption of ANOVA type tests is also hard to hold since measurement unit is block instead of day. In order to maintain adequate sample size, long trial and frequent treatment switching are expected, which may be impractical.

(3). Interrupted time series analysis (Cortina et al., 2013; Lane-Brown and Tate, 2010)

Interrupted time series analysis (ITS), or also called quasi-experimental time series analysis, is a study design for assessing the effectiveness of population-level interventions that have been implemented at a clearly defined point in time (Bernal et al., 2017). It can be used to evaluate the intervention's effect by tracking a long-term period before and after a given intervention point. The model can be written as:

$$Y_t = \beta_0 + \beta_1 t + \beta_2 X_t + \beta_3 t X_t + e_t$$

where β_0 and β_1 represent pre-intervention intercept and slope. β_2 and β_3 represent post-intervention changes in intercept and slope. e_t follows an ARIMA(p,q,d) process:

$$\left(1 - \sum_{i=1}^{p} \phi_i B^i\right)(1 - B)^d e_t = \left(1 + \sum_{i=1}^{q} \theta_i B^i\right) w_t$$

where B denotes back shift operator and w_t is independent Gaussian white noise. Effects of the intervention are estimated by changes in the intercept and slope of the time series before and after the intervention. ITS can also be viewed as a segmented regression model with ARIMA error and it takes autocorrelation in measurement into consideration, however, only one intervention change point is allowed in ITS, which restricts its usage in N-of-1 trials.

Combine N-of-1 Trials

(1). Meta-analysis (Martin and Whyte, 2007; Punja et al., 2016)

In meta-analysis, the estimated treatment effect in each individual N-of-1 trial is pooled together to provide a weighted average of multiple trials. Denote $\hat{\mu}$ as the estimated total treatment effect for the study population, then within a total of N individual trials,

$$\hat{\mu} = \frac{\sum_{i=1}^{N} w_i y_i}{\sum_{i=1}^{N} w_i}$$

where y_i is the observed mean treatment effect in individual trial and w_i is the weight of trial. A common choice of weight function is $w_i = 1/\sigma_i^2$, that is, the inverse of variances of each individual trial.

(2). Regression based tests (Araujo et al., 2016; X. Chen and Chen, 2014)

When a group of participants' information is available and treatment block level information is used, multiple N-of-1 trial can be regarded as grouped crossover design with fixed treatment sequence. Therefore, linear/generalized mixed effect models can be utilized to analyze data of N-of-1 trials. If we assume data are normally distributed, then the model can be written as:

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + \beta_2 P_j + \lambda_A Z_{Aij} + \lambda_B Z_{Bij} + b_{0i} + b_{1i} X_{ij} + \varepsilon_{ij}$$

where Y_{ij} denotes the i-th (i = 1,...,n) subject's outcome for the j-th period. X_{ij} denotes the i-th (i = 1,...,n) subject's treatment for the j-th period. P_j denotes the period effect. β_0 represents the intercept of the model. β_1 is the mean of treatment difference. β_2 is the slope of period effect. λ_A and λ_B are the carryover effect of treatment A and B. Z_{Aij} and Z_{Bij} are the indicator variables for two treatments. b_{0i} and b_{1i} are the random intercept and random slope of subject i, $b_{0i} \sim N(0, \sigma_0^2)$, $b_{1i} \sim N(0, \sigma_1^2)$. ε_{ij} is the Gaussian random error of subject i in period j. Especially, when we want to model for single subject, the model will be reduced to:

$$Y_j = \beta_0 + \beta_1 X_j + \beta_2 P_j + \lambda_A Z_{Aj} + \lambda_B Z_{Bj} + \varepsilon_j$$

which is similar to ANOVA type tests for single subject while adjusting for period and carryover effect.

(3). Bayesian hierarchical model (Schluter and Ware, 2005; D. Zucker et al., 1997)

Bayesian hierarchical model has been used to combine the results from individual N-of-1 trials and posterior estimates of population, individual level and between-patient variance can be obtained

after specify the prior distribution and the mean model. Use the same notation as the previous section, the Bayesian hierarchical model has the following structure

$$Y_{ij} \sim N(\beta_i X_{ij}, \sigma^2)$$
$$\beta_i \sim N(\theta, \sigma_0^2)$$
$$\theta \sim N(\mu_0, \sigma_{00}^2)$$
$$\sigma^2 \propto 1/\sigma^2$$

where β_i is the treatment effect in individual trial and θ is the treatment effect at population level. Each patient is assumed to have some underlying treatment effect β_i and $\beta_1, \beta_2, ..., \beta_N$ are assumed to be exchangeable between patients and arise from the same population distribution of θ. μ_0 and σ_{00}^2 are parameters in the hyperprior distribution, which are usually given as fixed value. This structure can also be applied to binary outcomes with a change of likelihood distribution and prior distribution.

There are two important features that need to be accounted for in the analysis of N-of-1 trial data. One is temporal correlation between measurements. Data from N-of-1 trials resemble a time series since repeated measurements are taken from the same individual. Measurements collected at adjacent time points will exhibit autocorrelation or temporal dependence. Most of current methods assume that observations are independent. However, ignoring autocorrelation can lead to underestimates or overestimates of standard error, depending on the existence of positive or negative autocorrelation, which will inflate the probability of type I error and type II error respectively (Vieira et al., 2017).

Carryover effects have also long been an issue in validity of N-of-1 trials. Even if washout periods are included in a study, we are still not guaranteed the elimination of carryover effects. With the existence of treatment carryover effect, the study results are likely to be biased. In the extreme, carryover may extend to multiple or all treatment periods and contaminate most of the study measurement (R. Kravitz et al., 2014).

Hardly any current methods can simultaneously account for temporal correlation between mea-

surements and carryover effects. Thus, innovative analytic strategy to separate these effects from the true treatment effect is expected.

1.1.5 Carryover Effects in Clinical Trials

Carryover effect is the effect of a treatment that persists after the treatment period and carryover effects which last up to and including k periods after the treatment has been stopped are known as kth-order carryover effects (B. Jones and Kenward, 1989). Carryover effects are usually detected in two-period crossover designs in clinical trials. Grizzle (1965) proposed the following model for cross-over type of design:

$$Y_{ijk} = \mu + b_{ij} + \pi_k + \phi_m + \lambda_m + \epsilon_{ijk}$$

where Y_{ijk} is the observed outcome on the j-th subject in the i-th treatment sequence during period k, μ is the overall mean, b_{ij} is the effect of the j-th subject in the ith treatment sequence, π_k is the effect of k-th period, ϕ_m is the effect of m-th drug, λ_m is the carryover effect of m-th drug and ϵ_{ijk} is the random error.

In a two interventions setting (Brown Jr, 1980), if there is no carryover effect, then the treatment difference $\delta = \phi_A - \phi_B$ can be estimated by

$$\hat{\delta} = \frac{1}{2}[(\bar{Y}_{1.2} - \bar{Y}_{1.1}) + (\bar{Y}_{2.2} - \bar{Y}_{2.1})]$$

where $\bar{Y}_{i.k} = \sum_{j=1}^{n_j} Y_{ijk}$, with variance

$$\text{var}(\hat{\delta}) = \frac{\sigma_\epsilon^2}{2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

To estimate the carryover effect, let $\lambda_A + \lambda_B = 0$ and $\gamma = \lambda_A - \lambda_B$, then an unbiased estimate of γ, $\hat{\gamma}$ is

$$\hat{\gamma} = \frac{1}{2}[(\bar{Y}_{2.1} + \bar{Y}_{2.2}) - (\bar{Y}_{1.1} + \bar{Y}_{1.2})]$$

with variance

$$\text{var}(\hat{\gamma}) = (4\sigma_b^2 + 2\sigma_\epsilon^2)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

If the difference in carryover effects is not zero, an unbiased estimate of δ can be obtained by using the data from first period only,

$$\hat{\delta}^* = \bar{Y}_{2.1} - \bar{Y}_{1.1}$$

with variance

$$\text{var}(\hat{\delta}^*) = (\sigma_b^2 + \sigma_\epsilon^2)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

To sum up, the analysis of the crossover design is more complicated than parallel group design. If carryover effects are negligible, then data from both periods can be utilized to estimate and test for a treatment effect. Otherwise, the analysis will be based on only the first-period data.

1.1.6 Motivating Examples of N-of-1 Trials

To motivate our work, consider an N-of-1 trial series that compare bright white light (10,000 lux) and dim red light (50 lux) in cancer patients with depressive symptoms, where light therapy was delivered by portable light boxes with instructions and reminders given by a smartphone application (I. M. Kronish et al., 2020). Briefly, each individual would use one of two light boxes each morning for 30 minutes per day over a 12 weeks. Along with the light boxes, a smartphone application would be used to give treatment reminders and to assess daily depressive symptoms, fatigue level, and affectivity over the entire 12-week period. While theory suggests bright white light may reduce cancer-related depression and fatigue, its effects may vary from individual to individual (Johnson et al., 2018). Thus, the primary analytical goal in I. M. Kronish et al. (2020) is to identify for each individual whether bright white light is superior in terms of symptom control. Figure 1.3 shows the daily assessments of two patients during the study course.

Another example is from a hypertension treatment study (I. M. Kronish et al., 2019). 7 patients with a history of mild hypertension were included and assigned three blood pressure medications

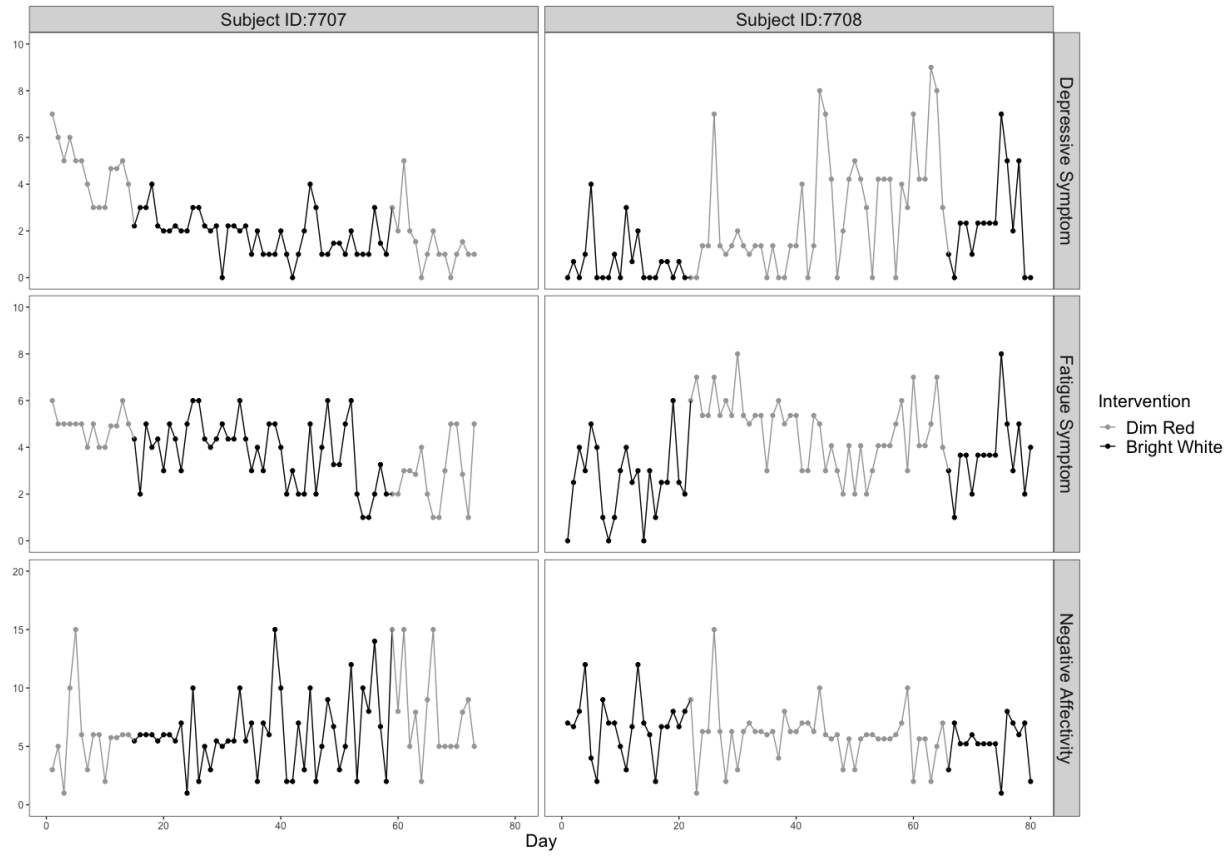


Figure 1.3: Daily assessments of two patients id 7707 and 7708. Black line represents bright white light intervention, and grey line represents dim red light.

from three different first-line blood pressure classes: hydrochlorothiazide (HCTZ), amlodipine and losartan. Each trial was designed to last 12 weeks with a counterbalanced sequence ABCCBA. The outcome of interest is systolic blood pressure (SBP), which was measured by patients twice in the morning and twice in at night by home blood pressure device. The primary analytical goal is to determine the personalized selection of blood pressure medications for each patient.

In Section 2.5 and 4.4, the light therapy study dataset and hypertension study dataset are used to estimate treatment effect of single patient. In Section 3.5, the light therapy study dataset is revisited to estimate the treatment effect of the entire study population.

1.2 Distributed Lag Models

The distributed lag model (DLM), used widely in economics (Almon, 1965; Koyck, 1954), advertising (Bass and Clarke, 1972) and environment health studies (Welty et al., 2009; Zanobetti et al., 2000), is a model in which current value of a dependent variable is not only associated with current value of an explanatory variable, but also its lagged values. This method allows the effect of an exposure to be distributed over a specific period of time, and providing a better understanding of the exposure-outcome relationship, which is especially useful in evaluating the delayed effect. Compared to existing methods which assume carryover effect is constant within each treatment period, DLMs are more informative in evaluating the whole time course of treatment effect. Furthermore, the number of observations measured in each treatment period are not required to be same within or between participants, which makes it more flexible to analyze N-of-1 trials with varying treatment period length. The general form of a DLM is:

$$y_t = \mu + \sum_{i=0}^{\infty} \beta_i x_{t-i} + u_t \tag{1.1}$$

where $t = 1, 2, ..., n$ denotes different time periods, y_t denotes a response variable of interest at time t and x_t denotes a time-varying explanatory variable, which has some influence on all outcome after y_t. u_t are independent Gaussian errors with mean 0 and variance σ_u^2, and the lagged coefficient

β_i is usually assumed to satisfy the condition that $\lim_{i\to\infty}\beta_i$ and $\sum_i \beta_i < \infty$. In practice, the lag length is truncated to some finite length L since the lag coefficient vanishes to zero after L periods. Then we can substitute the upper limit in (1.1) with the maximum number of lags L.

In DLMs, the parameters of interest are lag coefficients and the overall impact of exposure, which is the summation of all lag coefficients. One difficulty of DLM is the multicollinearity problem of lagged explanatory variables.

1.2.1 Polynomial Distributed Lag Model

Almon (1965) proposed a technique to add constraint on the lagged coefficients by assuming these coefficients β_i can be approximated by a d-th degree polynomial function of i, where d is usually much smaller than L, i.e.,

$$\beta_i = \sum_{j=0}^{d} \alpha_j i^j, \quad i = 0, \ldots L \text{ and } 0 < d \leq L \tag{1.2}$$

Or written in matrix notation:

$$\boldsymbol{\beta} = \boldsymbol{R}\boldsymbol{\alpha} \tag{1.3}$$

where

$$\boldsymbol{R} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & i & i^2 & \cdots & i^d \end{bmatrix}, \ \boldsymbol{\alpha} = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_d \end{bmatrix} \tag{1.4}$$

Then (1.1) can be written as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{R}\boldsymbol{\alpha} + \boldsymbol{u} \tag{1.5}$$

The OLS estimator of α is:

$$\hat{\boldsymbol{\alpha}}_{\text{OLS}} = (\boldsymbol{R}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{R})^{-1}\boldsymbol{R}'\boldsymbol{X}'\boldsymbol{y} \tag{1.6}$$

and Almon estimator of $\boldsymbol{\beta}$ is:

$$\hat{\boldsymbol{\beta}}_{\text{Almon}} = \boldsymbol{R}(\boldsymbol{R}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{R})^{-1}\boldsymbol{R}'\boldsymbol{X}'\boldsymbol{y} \tag{1.7}$$

If the error term is homoscedastic, it is the best linear unbiased estimator (BLUE) of (Judge, 1982). The Almon technique reduces the effect of multicollinearity because there are fewer explanatory variables in the transformed model as compared to the actual DLM. However, in the presence of heteroscedasticity, $\hat{\boldsymbol{\alpha}}_{\text{OLS}}$ does not remain efficient and consequently $\hat{\boldsymbol{\beta}}_{\text{Almon}}$ also becomes inefficient.

The polynomial distributed lag model alleviates the multicollinearity issue by requiring the lag weights to fall on a smooth curve. While the polynomial DLM is flexible, it is still a very strong assumption to make about the curve of lag coefficients.

1.2.2 Geometric Distributed Lag Model

Koyck (1954) proposed an infinite distributed lag model by adding the following constrain on lag coefficients in (1.1):

$$\beta_i = \beta_0\lambda^i, \quad \text{and } 0 < \lambda < 1 \tag{1.8}$$

In this model, lag coefficients are assumed to decrease geometrically. In other words, recent past explanatory variable are more influential than distant past explanatory variable. λ is the coefficient decaying rate. By replacing β_i into (1.1), we can have:

$$y_t = \mu + \beta_0 \sum_{i=0}^{\infty} \lambda^i x_{t-i} + u_t \tag{1.9}$$

where β_0 is the immediate effect and the long term effect is defined as:

$$\sum_{k=0}^{\infty} \beta_k = \frac{\beta_0}{1-\lambda} \tag{1.10}$$

As Koyck DLM contains infinite parameters, it is usually transformed from an infinite dis-

tributed lag model into an autoregressive model with only three observable variables by subtracting λy_{t-1} from (1.9):

$$y_t = (1-\lambda)\mu + \lambda y_{t-1} + \beta_0 x_t + (u_t - \lambda u_{t-1}) \tag{1.11}$$

There are several approaches to estimate the parameters in the Koyck DLM, like the methods proposed by Klein (1962) and Liviatan (1963). Here we regard Koyck DLM as an autoregressive-moving average with exogenous terms (ARMAX) model where $(y_t - \lambda y_{t-1})$ represents the autoregressive part and $u_t - \lambda u_{t-1}$ represents the moving average part. Maximum likelihood method is used to estimate the parameters. The likelihood function is:

$$L(\mu, \beta_0, \lambda, \sigma^2) = \prod_{t=1}^{N} \left\{ \left(2\pi\sigma^2\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2} u_t^2\right) \right\} \tag{1.12}$$

where $u_1, , u_2, ..., u_n$ is:

$$\begin{aligned} u_1 &= 0 \\ u_t &= y_t - (1-\lambda)\mu - \lambda y_{t-1} - \beta_0 x_t + \lambda u_{t-1}, \quad \text{for } t = 2, ..., n \end{aligned} \tag{1.13}$$

Klein (1962) suggested a method to obtain the maximum likelihood estimation of the parameters in the model by first writing (2.4.4) as:

$$\begin{aligned} y_t &= \beta_0 \sum_{i=0}^{t-1} \lambda^i x_{t-i} + \beta_0 \sum_{i=t}^{\infty} \lambda^i x_{t-i} + u_t \\ &= \beta_0 \eta_{1t} + \gamma \eta_{2t} + u_t \end{aligned} \tag{1.14}$$

where $\eta_{1t} = \sum_{i=0}^{t-1} \lambda^i x_{t-i}$ and $\eta_{2t} = \lambda^t/(1-\lambda)$.

Given λ, we can easily generate η_{1t} and η_{2t}, then β_0 and γ can be obtained by ordinary least squares. If we further assume that u_t independently follow $N(0, \sigma^2)$, then searching over a fine grid for λ that returns the minimal residual sum of squares will be the maximum likelihood estimate of λ and same for the corresponding β_0 and γ.

Iterative numeric methods, such as Gradient Decent algorithm, BFGS algorithm and Newton-

Raphson algorithm can be applied to obtain the maximum likelihood estimate of parameters.

Liviatan (1963) proposed to use instrumental variables approach to estimate coefficients of Koyck DLM, which can avoid above iterative procedure and computational burden. Equation (1.11) is reparameterized as:

$$y_t = \delta_0 + \delta_1 y_{t-1} + \delta_2 x_t + \epsilon_t \tag{1.15}$$

where $\delta_0 = (1-\lambda)\mu$, $\delta_1 = \lambda$, $\delta_2 = \beta_0$ and $\epsilon_t = u_t - \lambda u_{t-1}$. The first lag of the dependent series y_{t-1} and the error term ϵ_t are correlated with each other, since y_{t-1} depends on u_{t-1} and the error term ϵ_t depends on both u_t and u_{t-1}. The OLS coefficient estimator is therefore biased and inconsistent, unless there is no association between y_{t-1} and ϵ_t.

From the assumption of DLM, x_t and x_{t-1} are exogenous, i.e., $\text{cov}(x_t, \epsilon_t) = \text{cov}(x_{t-1}, \epsilon_t) = 0$. Furthermore, x_{t-1} is correlated with change in y_{t-1}. Then x_{t-1} can be utilized as instrument variable for y_{t-1} in formula (1.15). Note that x_t in the model is exogenous, so it is both an independent and instrumental variable.

Consider formula (1.15) in matrix form:

$$\boldsymbol{y} = \tilde{\boldsymbol{X}}\boldsymbol{\delta} + \boldsymbol{\epsilon} \tag{1.16}$$

where

$$\tilde{\boldsymbol{X}} = \begin{bmatrix} 1 & y_1 & x_2 \\ 1 & y_2 & x_3 \\ \vdots & \vdots & \vdots \\ 1 & y_{n-1} & x_n \end{bmatrix}, \ \boldsymbol{\delta} = \begin{bmatrix} \delta_0 \\ \delta_1 \\ \delta_2 \end{bmatrix}, \ \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_2 \\ \epsilon_3 \\ \vdots \\ \epsilon_n \end{bmatrix} \tag{1.17}$$

Let $\boldsymbol{Z}$ be an $n \times 2$ matrix with i^{th} row $z_i = (x_i, x_{i-1})$, which denotes the instrumental variables. Then the instrumental variables estimator is:

$$\hat{\boldsymbol{\delta}}_{\text{IV}} = (\boldsymbol{Z}'\tilde{\boldsymbol{X}})^{-1}\boldsymbol{Z}'\boldsymbol{y} \tag{1.18}$$

Under homoskedasticity, the asymptotic distribution of instrumental variables estimator is normal with mean $\boldsymbol{\delta}$ and the asymptotic covariance matrix is:

$$\widehat{\mathrm{V}}\left[\widehat{\boldsymbol{\delta}}_{\mathrm{IV}}\right] = \frac{\hat{\sigma}^2}{n}\left(\mathbf{Z}'\mathbf{X}\right)^{-1}\mathbf{Z}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{X}\right)^{-1} \tag{1.19}$$

where $\hat{\sigma}^2 = \frac{1}{n-3}(\boldsymbol{y} - \tilde{\boldsymbol{X}}\hat{\boldsymbol{\delta}}_{\mathrm{IV}})'(\boldsymbol{y} - \tilde{\boldsymbol{X}}\hat{\boldsymbol{\delta}}_{\mathrm{IV}})$.

The original parameters of Koyck model can be viewed as a function of $\boldsymbol{\delta}$, i.e., $(\mu, \lambda, \beta_0) = g(\delta_0, \delta_1, \delta_2)$, Using multivariate delta method, we can derive with estimation $(\hat{\mu}, \hat{\lambda}, \hat{\beta_0}) = (\frac{\hat{\delta}_0}{1-\hat{\delta}_1}, \hat{\delta}_1, \hat{\delta}_2)$ and its estimated covariance matrix $\nabla g(\boldsymbol{\delta})'\widehat{\mathrm{V}}\left[\widehat{\boldsymbol{\delta}}_{\mathrm{IV}}\right]\nabla g(\boldsymbol{\delta})$.

If we are interested in single distributed lag coefficient up to lag L, then $\hat{\boldsymbol{\beta}} = (\hat{\beta_0}, \hat{\beta_1}, ..., \hat{\beta_L}) = (\hat{\delta}_2, \hat{\delta}_2\hat{\delta}_1, ..., \hat{\delta}_2\hat{\delta}_1^L)$, with estimated covariance matrix:

$$\widehat{\mathrm{V}}\left[\widehat{\boldsymbol{\beta}}\right] = \begin{bmatrix} 0 & 1 \\ \hat{\delta}_2 & \hat{\delta}_1 \\ \vdots & \vdots \\ L\hat{\delta}_2\hat{\delta}_1^{L-1} & \hat{\delta}_1^L \end{bmatrix} \widehat{\mathrm{V}}(\delta_1, \delta_2) \begin{bmatrix} 0 & \hat{\delta}_2 & \dots & L\hat{\delta}_2\hat{\delta}_1^{L-1} \\ 1 & \hat{\delta}_1 & \dots & \hat{\delta}_1^L \end{bmatrix} \tag{1.20}$$

where $\widehat{\mathrm{V}}(\delta_1, \delta_2)$ is the lower right block matrix of $\widehat{\mathrm{V}}\left[\widehat{\boldsymbol{\delta}}_{\mathrm{IV}}\right]$.

Koyck DLM has a very specific lag coefficient shape which is hard to be verified before data analysis. Due to its parsimonious modelling structure, it is hard to allow more freedom to the estimation of lag coefficient shape. However, the advantage of Koyck DLM is that it allows for coefficient estimation for an infinite time span, and does not require to truncate the model to an arbitrary finite lag length.

1.2.3 Bayesian Distributed Lag Model

Instead of adding parametric constraint directly on the lag coefficient, Welty et al. (2009) proposed to incorporate prior knowledge of the shape of lag coefficient curve into DLM through a Bayesian framework. For a normal linear distributed lag model, the Bayesian DLM has the fol-

lowing hierarchical structure:

$$\begin{aligned} \boldsymbol{Y}|\boldsymbol{X},\boldsymbol{\beta},\sigma^2 &\sim N(\boldsymbol{X}^T\boldsymbol{\beta},\sigma^2\mathbf{I}_n) \\ \boldsymbol{\beta}|\sigma^2_\beta,\eta &\sim N(0,\sigma^2_\beta\boldsymbol{\Omega}(\eta)) \\ \sigma^2_\beta &= 10\text{Var}(\hat{\beta}_0) \\ \boldsymbol{\Omega}(\eta) &= \boldsymbol{V}(\eta_1)\boldsymbol{W}(\eta_2)\boldsymbol{V}(\eta_1) \end{aligned} \tag{1.21}$$

where $\boldsymbol{V}(\eta_1) = \text{diag}(1, \exp(\eta_1)^{1/2}, ..., \exp(\eta_1 L)^{1/2})$, $\boldsymbol{W}(\eta_2)$ is the correlation matrix derived from the covariance matrix $\boldsymbol{V}(\eta_2)\boldsymbol{V}(\eta_2)' + \{\mathbf{I}_{L+1} - \mathbf{V}(\eta_2)\}\mathbf{1}_{L+1}\mathbf{1}'_{L+1}\{\mathbf{I}_{L+1} - \mathbf{V}(\eta_2)\}'$, with $\mathbf{1}_{L+1}$ denotes a $(L+1)\times 1$ vectors of ones.

The prior covariance matrix of $\boldsymbol{\beta}$ can also be written component-wisely:

$$\text{var}(\beta_i) = \sigma^2_\beta \exp(-\eta_1 i)$$

$$\text{cov}\left(\beta_i,\beta_j\right) = \frac{\sigma^2_\beta\left\{1-\exp\left(-\eta_2 i\right)\right\}\left\{1-\exp\left(-\eta_2 j\right)\right\}\exp\left\{-\eta_1\left(i+j\right)/2\right\}}{\sqrt{\left(\left[\left\{1-\exp\left(-\eta_2 i\right)\right\}^2+\exp\left(-2\eta_2 i\right)\right]\left[\left\{1-\exp\left(-\eta_2 j\right)\right\}^2+\exp\left(-2\eta_2 j\right)\right]\right)}}$$

for $i, j = 0, 1, ...L-1$ and $i \neq j$. η_1 controls the rate at which the variance of $\boldsymbol{\beta}$ approach zero and η_2 controls the rate at which neighbouring coefficients become more correlated.

Let $\hat{\boldsymbol{\beta}}$ be the ML estimate of the unconstrained distributed lag coefficients and let $\boldsymbol{\Sigma}$ be the sample covariance matrix, then the full conditional posterior distribution for $\boldsymbol{\beta}$ is:

$$\begin{aligned} \boldsymbol{\beta}|\hat{\boldsymbol{\beta}},\boldsymbol{\eta},\sigma^2_\beta \sim N(&\left\{1/\sigma^2_\beta\boldsymbol{\Omega}(\boldsymbol{\eta})^{-1}+\boldsymbol{\Sigma}^{-1}\right\}^{-1}\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\beta}} \\ &\left\{1/\sigma^2_\beta\boldsymbol{\Omega}(\boldsymbol{\eta})^{-1}+\boldsymbol{\Sigma}^{-1}\right\}^{-1}) \end{aligned} \tag{1.22}$$

For a general linear DLM, the posterior distribution of $\boldsymbol{\beta}$ has no closed form, but it can be computed through Gibbs sampling or other Markov chain Monte Carlo methods. Also, we can use Bayesian statistics software, like Stan or Winbugs to obtain MCMC sampling from posterior distribution.

Bayesian DLM adds the constrain of lag coefficient curve through prior information and allows the degree of smoothness of the lag curve to be decided by the data. When prior assumption that lag effects smoothly approach zero as lag increases is valid, the performance of Bayesian DLM is better than other DLMs in terms of coefficient estimation.

Chapter 2: Analysis of Single N-of-1 Trial Using Bayesian Distributed Lag Model with Autocorrelated Errors

2.1 Introduction

N-of-1 trials are multi-period crossover studies that compare two or more interventions in single individuals, and are suitable for evaluating personalized treatment effects in those with chronic conditions where the outcome is relatively stable (R. Kravitz et al., 2014). While treatment delivery and data collection involved in N-of-1 trials are more intensive and lengthier than those in usual care, advances in mobile and sensor technology (Topol, 2010) and better understanding of patient preferences (Cheung et al., 2020) have improved the implementation and uptake of this trial methodology.

In a systematic review of 108 N-of-1 trial series between 1985 and 2010, Gabler et al. (2011) report the use of graphical comparison, hypothesis tests (e.g., t-test, nonparametric tests), and regression models in the analysis of N-of-1 data. While there is no single agreed upon analysis method, these methods ignore two key features of experimental N-of-1 data. First, most methods do not account for temporal dependence between assessments. Second, the methods do not capture the carryover effects of an intervention. The second data feature can be partly addressed by using a distributed lagged model (DLM), which is widely used in economics (Almon, 1965; Koyck, 1954), advertising (Bass and Clarke, 1972), and environment health studies (Welty et al., 2009; Zanobetti

et al., 2000). A DLM postulates that the current value of the outcome variable depends on the previous values (lags) of an exposure as well as the current exposure value, thus allowing the total exposure effect to be distributed over a time period and facilitating explicit modeling of carryover effects. A potential challenge in fitting a DLM is collinearity of the exposure lags. The N-of-1 trial design will further aggravate the problem: as illustrated in Figure 1.3, the exposure (light box) often remains the same as in the previous day in order to avoid switching intervention too frequently during a trial. A strategy to handle collinearity in DLM is by putting parametric constraints on the lag coefficients such as polynomial lags (Almon, 1965), or geometric lags (Koyck, 1954). Alternatively, one may consider putting informative prior on the coefficients in a Bayesian framework (Welty et al., 2009).

In this chapter, we adopt the Bayesian framework and propose a Bayesian distributed lag model with autocorrelated errors (BDLM-AR) as an extension of DLMs for N-of-1 trial data. The model is novel in several ways. First, we propose a prior distribution that constraints the lag coefficients with shrinkage factors that increase over time. Second, we impose a fused ridge-type penalty to address collinearity, which may be viewed as a variant of the fused lasso method (Tibshirani and Wang, 2008). Third, while current DLM methods assume independent error terms, we incorporate temporal correlations using an autoregressive error model. We will introduce the proposed BDLM-AR with details in Section 2.2, and describe the posterior computations in Section 2.3. The performance of BDLM-AR will be evaluated and compared with other methods by simulation studies presented in Section 2.4. We will apply the proposed method to the light therapy data in Section 2.5, and will conclude this article with a discussion in Section 2.6. Technical details are given in the Appendix A.

2.2 Bayesian Distributed Lag Model with Autocorrelated Errors

2.2.1 Proposed Model

Suppose we observe data from a patient on n consecutive days. On day $t = 1, \ldots, n$, let X_t and Y_t denote the binary treatment indicator and the outcome of interest, respectively. We consider a distributed lag autoregressive model for Y, described as follows:

$$Y_t = \mu + \sum_{l=0}^{L} \beta_l X_{t-l} + \epsilon_t \tag{2.1}$$

for $t = p + 1, ..., n$, where the error term ϵ_t follows an autoregressive process,

$$\epsilon_t = \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + ... + \phi_p \epsilon_{t-p} + w_t \tag{2.2}$$

w_t is a white Gaussian noise with mean zero and unknown variance $\sigma^2 > 0$, and L and p are pre-specified. Note that for $t < L$, the maximum lag effect is of order $t - 1$, and terms involving X with non-positive subscript are not included in the model.

Model (2.1) is composed of two parts. First, for the structural component, the mean model is specified by lag coefficients $\boldsymbol{\beta} = (\beta_0, ..., \beta_L)'$ and control mean μ. The immediate treatment effect is measured by β_0, and the carryover effect due to treatment on l days ago is measured by β_l for $l > 0$. In the model, we assume the carryover effect beyond day L is zero. As such, the total carryover treatment effect is captured as

$$\delta \triangleq \sum_{l=1}^{L} \beta_l = E(Y_t | X_{t-1} = 1, ...X_{t-L} = 1, X_t) - E(Y_t | X_{t-1} = 0, ..., X_{t-L} = 0, X_t).$$

Hence, the model naturally breaks down total treatment effect into β_0 and δ.

Second, for the stochastic component, temporal dependency between errors is specified using an order-p autoregressive error model with autoregression coefficient $\boldsymbol{\phi} = (\phi_1, ..., \phi_p)'$. Let B denote the backshift operator, that is, having $\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p$ so that the

autoregression model for the error terms can be written as $\Phi(B)\epsilon_t = w_t$. It is often convenient to work with the transformed data $Y_t^* = \Phi(B)Y_t$ and $X_t^* = \Phi(B)X_t$ in the estimation steps. Thus, applying $\Phi(B)$ to both sides of model (2.1), we will rewrite the model

$$Y_t^* = \mu^* + \sum_{l=0}^{L} \beta_l X_{t-l}^* + w_t, \tag{2.3}$$

for $t = p+1, ..., n$, where $\mu^* = \Phi(B)\mu$. To stack the data in vector form, we have

$$(\boldsymbol{Y}^* \mid \boldsymbol{X}^*, \mu^*, \boldsymbol{\beta}) \sim N(\mu^* \mathbf{1}_{n-p} + \boldsymbol{X}^* \boldsymbol{\beta}, \sigma^2 \mathbf{I}_{n-p}) \tag{2.4}$$

where $\boldsymbol{Y}^* = (Y_{p+1}^*, ..., Y_n^*)'$, $\boldsymbol{X}^*$ is a $(n-p)\times(L+1)$ matrix with $X_{k-l+p+1}^*$ being the (k,l)-th element of $\boldsymbol{X}^*$, $\mathbf{1}_{n-p}$ is a 1-vector of length $n-p$, and $\mathbf{I}_{n-p}$ is the identity matrix of dimension $n-p$. We denote $\tilde{\boldsymbol{\beta}} = (\mu, \boldsymbol{\beta}')'$ and $\tilde{\boldsymbol{X}}^* = (\Phi(B)\mathbf{1}_{n-p}, \boldsymbol{X}^*)$, so that $\tilde{\boldsymbol{X}}^*\tilde{\boldsymbol{\beta}} = \mu^*\mathbf{1}_{n-p} + \boldsymbol{X}^*\boldsymbol{\beta}$.

2.2.2 Prior Distribution on the Mean Model

We consider normal prior distribution for $\tilde{\boldsymbol{\beta}}$, that is, having

$$\tilde{\boldsymbol{\beta}} \sim N(\mathbf{0}, \sigma^2 \tilde{\boldsymbol{\Omega}}^{-1}), \tag{2.5}$$

where $\tilde{\boldsymbol{\Omega}} = diag(c_0, \boldsymbol{\Omega})$ so that the prior variance of μ is $\sigma^2 c_0^{-1}$ and the prior variance-covariance matrix of $\boldsymbol{\beta}$ is $\sigma^2\boldsymbol{\Omega}^{-1}$. We note that the prior variance depends on the variance σ^2 of the observations: such dependence renders a fused ridge penalized estimation procedure that is free of the variance parameters, resulting in computational stability; see Equation (2.7) below.

We will postulate a non-informative prior on μ by setting c_0 to be a small number, and we will

consider $\boldsymbol{\Omega}$ of the following form:

$$\begin{pmatrix} \lambda_0+\lambda_0^* & -\lambda_0^* & 0 & \dots & \dots & 0 \\ -\lambda_0^* & \lambda_1+\lambda_0^*+\lambda_1^* & -\lambda_1^* & \dots & \dots & 0 \\ 0 & -\lambda_1^* & \lambda_2+\lambda_1^*+\lambda_2^* & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \lambda_{L-1}+\lambda_{L-2}^*+\lambda_{L-1}^* & -\lambda_{L-1}^* \\ 0 & 0 & 0 & \dots & -\lambda_{L-1}^* & \lambda_L+\lambda_{L-1}^*+\lambda_L^* \end{pmatrix}, \tag{2.6}$$

where the hyperparameters $\lambda_l, \lambda_l^* > 0$, for $l = 0, \dots, L$, are constrained to increase over l. As a result of the monotonicity constraint, a lag coefficient β_l at a greater lag l is associated with a larger diagonal element in precision $\boldsymbol{\Omega}$, thus shrinking β_l toward the prior mean (zero) to a greater extent. This effectively addresses collinearity of the lag coefficients without imposing strong parametric structure to $\boldsymbol{\beta}$. In addition, using the normal prior (2.5) with precision matrix (2.6), we can show the maximum *a posteriori* probability estimate of $\tilde{\boldsymbol{\beta}}$ minimizes a fused ridge-type penalty:

$$(\boldsymbol{Y}^* - \tilde{\boldsymbol{X}}^*\tilde{\boldsymbol{\beta}})^T(\boldsymbol{Y}^* - \tilde{\boldsymbol{X}}^*\tilde{\boldsymbol{\beta}}) + c_0\mu^2 + \sum_{l=0}^{L}\lambda_l\beta_l^2 + \sum_{l=0}^{L}\lambda_l^*(\beta_l - \beta_{l+1})^2 \tag{2.7}$$

where $\beta_{L+1} \triangleq 0$, thus giving insights on how the proposed prior constrains the lag coefficients: it regularizes not only the ℓ_2-norm of the coefficients but also their successive differences, thereby enhancing local smoothness. The equivalence between the Bayesian inference and the fused ridge regularization (2.7) is proved in Appendix A.1.

There are many ways to specify λ_l and λ_l^* to meet the monotonicity constraints. In this article, we consider $\lambda_l = \exp\{\gamma_1(l+1)\} - 1$ and $\lambda_l^* = \exp\{\gamma_2(l+1)\} - 1$ for $\gamma_1, \gamma_2 > 0$, so that γ_1 controls the rate at which the ridge penalty in (2.7) increases, and γ_2 controls the increasing rate of smoothness of the coefficient curve $\boldsymbol{\beta}$. Instead of treating these hyperparameters as fixed, we postulate a truncated standard exponential hyperprior on (γ_1, γ_2), that is, having probability density

function

$$\pi(\gamma_1, \gamma_2) \propto \exp(-\gamma_1 - \gamma_2)\mathbf{1}_{S_\gamma}(\gamma_1, \gamma_2) \tag{2.8}$$

where the support S_γ includes all pairs (γ_1, γ_2) with which the precision matrix $\mathbf{\Omega}$ is positive definite. As such, the degree of ridge and smooth penalization can be determined according to the posterior distribution of the pair.

2.2.3 Prior Distribution on the Error Model

We put the Jeffreys prior for the error variance σ^2, that is, having density function

$$\pi(\sigma^2) \propto 1/\sigma^2 \tag{2.9}$$

Note that any inverse-gamma prior for σ^2 would maintain conjugacy, and the Jeffreys prior can be regarded as an improper limit of inverse-gamma prior distribution.

For the autoregressive process, we consider a truncated normal prior for $\boldsymbol{\phi}$ subject to the constraint that the error process is stationary. Specifically, we postulate

$$\boldsymbol{\phi} \sim N_p\left(0_p, 200 \times \mathbf{I}_p\right)\mathbf{1}_{S_{\boldsymbol{\phi}}}(\boldsymbol{\phi}) \tag{2.10}$$

where $S_{\boldsymbol{\phi}}(\boldsymbol{\phi})$ denotes the support where all roots of the polynomial $\Phi(B) = 1 - \sum_{l=1}^{p} \phi_l B^l$ are outside the unit circle. Following Chib (1993), we can show that the process $\{\epsilon_t : t = 1, 2, \ldots\}$ is stationary when $\boldsymbol{\phi} \in S_{\boldsymbol{\phi}}(\boldsymbol{\phi})$. Note that the range of each ϕ_i is $(-1, 1)$; thus, a prior variance of 200 in (2.10) essentially amounts to a flat prior.

2.3 Conditional Posterior Distributions

The proposed Bayesian model includes several conditionally conjugate priors, which facilitate posterior computations via a hybrid Metropolis-Hastings/Gibbs algorithm. We describe the conditional posterior distributions in this section; the details of derivation can be found in Appendix

A.2.

Working with the likelihood (2.4) based on the transformed data Y_t^*, we obtain that $\tilde{\boldsymbol{\beta}}$ is conditionally normally distributed *a posteriori*:

$$\tilde{\boldsymbol{\beta}} \mid \boldsymbol{Y}, \tilde{\boldsymbol{X}}, \sigma^2, \boldsymbol{\phi}, \boldsymbol{\gamma} \sim N_{L+1}\left\{[\tilde{\boldsymbol{X}}^{*'}\tilde{\boldsymbol{X}}^* + \tilde{\boldsymbol{\Omega}}(\boldsymbol{\gamma})]^{-1}\tilde{\boldsymbol{X}}^{*'}\boldsymbol{Y}^*, \sigma^2[\tilde{\boldsymbol{X}}^{*'}\tilde{\boldsymbol{X}}^* + \tilde{\boldsymbol{\Omega}}(\boldsymbol{\gamma})]^{-1}\right\} \tag{2.11}$$

and that σ^2 has an inverse-gamma conditional posterior:

$$\sigma^2 \mid \boldsymbol{Y}, \tilde{\boldsymbol{X}}, \tilde{\boldsymbol{\beta}}, \boldsymbol{\phi}, \boldsymbol{\gamma} \sim \text{IG}\left[\frac{n-p+L+1}{2}, \frac{(\boldsymbol{Y}^* - \tilde{\boldsymbol{X}}^*\tilde{\boldsymbol{\beta}})'(\boldsymbol{Y}^* - \tilde{\boldsymbol{X}}^*\tilde{\boldsymbol{\beta}}) + \tilde{\boldsymbol{\beta}}'\tilde{\boldsymbol{\Omega}}(\boldsymbol{\gamma})\tilde{\boldsymbol{\beta}}}{2}\right] \tag{2.12}$$

Note that the dependence of (2.11) and (2.12) on $\boldsymbol{\phi}$ is via the transformed data $\boldsymbol{Y^*}$.

Working with model (2.2) and (2.10), we obtain the conditional posterior distribution of $\boldsymbol{\phi}$ is truncated multivariate normal:

$$\boldsymbol{\phi} \mid \boldsymbol{Y}, \tilde{\boldsymbol{X}}, \tilde{\boldsymbol{\beta}}, \sigma^2, \boldsymbol{\gamma} \sim N_p\left[\left(\sigma^{-2}\boldsymbol{E}^{*'}\boldsymbol{E}^* + \sigma_{\boldsymbol{\phi}}^{-2}\mathbf{I}\right)^{-1}\sigma^{-2}\boldsymbol{E}^{*'}\boldsymbol{\epsilon}^*, \left(\sigma^{-2}\boldsymbol{E}^{*'}\boldsymbol{E}^* + \sigma_{\boldsymbol{\phi}}^{-2}\mathbf{I}\right)^{-1}\right]\mathbf{1}_{S_{\boldsymbol{\phi}}}(\boldsymbol{\phi}) \tag{2.13}$$

where $\boldsymbol{\epsilon}^* = (\epsilon_{p+1}^*, \ldots, \epsilon_n^*)'$, $\epsilon_t^* = Y_t - \mu - \sum_{l=0}^{L}\beta_l X_{t-l}$, and $\boldsymbol{E}^*$ is a $(n-p)\times p$ matrix with ϵ_{p+k-j}^* being the (k, j)-th element.

Because of conjugacy, the parameters $\tilde{\boldsymbol{\beta}}, \sigma^2$, and $\boldsymbol{\phi}$ can be easily updated in a Gibbs sampling fashion.

Using the likelihood (2.4) and prior of $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\beta}}$, the conditional posterior distribution can be expressed as

$$\pi(\boldsymbol{\gamma} \mid \boldsymbol{Y}, \tilde{\boldsymbol{X}}, \tilde{\boldsymbol{\beta}}, \boldsymbol{\phi}, \sigma^2) \propto |\sigma^{-2}\tilde{\boldsymbol{\Omega}}(\boldsymbol{\gamma})|^{\frac{1}{2}}\exp\left[-\frac{1}{2\sigma^2}\tilde{\boldsymbol{\beta}}'\tilde{\boldsymbol{\Omega}}(\boldsymbol{\gamma})\tilde{\boldsymbol{\beta}}\right]\exp(-\gamma_1-\gamma_2)\mathbf{1}_{S_{\boldsymbol{\gamma}}}(\gamma_1, \gamma_2). \tag{2.14}$$

We propose to sample $\boldsymbol{\gamma}$ using a Metropolis-Hastings (MH) step with a uniform $U(-a, a)$ proposal distribution. That is, $\gamma_{i,new} = \gamma_i + U(-a, a)$, for $i = 1, 2$ where the tuning parameter a is chosen such that the acceptance rate of proposed sample is around 50% (Gelman et al., 1996).

Note that updating the hyperparameter $\boldsymbol{\gamma}$ involves the calculation of the precision matrix $\tilde{\boldsymbol{\Omega}}(\boldsymbol{\gamma})$, which needs to be positive definite. The $(L+2) \times (L+2)$ precision matrix $\tilde{\boldsymbol{\Omega}}(\boldsymbol{\gamma})$ is a special case of tridiagonal matrix. The computational cost of regular algorithms for checking whether a given matrix is positive definite is at most $O(L^3)$. In this article, we implement an efficient computation algorithm with cost of $O(L)$. Specifically, define an $(L+2)$-dimensional vector $\mathbf{c} = (c_0, c_2, ..., c_{L+1})$ by

$$c_l = \begin{cases} \lambda_0 + \lambda_0^*, & l = 0, \\ (\lambda_l + \lambda_l^*) - \frac{1}{c_{l-1}}, & l = 1, 2, ..., L+1 \end{cases}$$

El-Mikkawy (2004) showed that the $\tilde{\boldsymbol{\Omega}}(\boldsymbol{\gamma})$ is positive definite if and only if $c_l > 0$ for $l = 0, 1, ...L+1$. Thus the problem boils down to checking the positiveness of elements in $\mathbf{c}$.

Having the full conditional distribution of all parameters in BDLM-AR, we use a hybrid Metropolis-Hastings/Gibbs algorithm to generate samples of $(\tilde{\boldsymbol{\beta}}, \boldsymbol{\phi}, \sigma^2, \boldsymbol{\gamma})$ from the posterior distribution.

Algorithm 1: The hybrid Metropolis-Hastings/Gibbs algorithm

Step 1. Set initial values for $\tilde{\boldsymbol{\beta}}$, σ^2, $\boldsymbol{\phi}$ and $\boldsymbol{\gamma}$;

for $i \leftarrow 1$ *to* $n_{iteration}$ **do**

Step 2. Given current value of $\boldsymbol{\phi}$, transform $\boldsymbol{Y}$, $\tilde{\boldsymbol{X}}$ to $\boldsymbol{Y}^*$, $\tilde{\boldsymbol{X}}^*$ as described in equation (3) of Section 2.1; Also construct precision matrix $\tilde{\boldsymbol{\Omega}}(\boldsymbol{\gamma})$ based on $\boldsymbol{\gamma}$ as described in Section 2.1;

Step 3. Conditional on current values of $\boldsymbol{Y}^*$, $\tilde{\boldsymbol{X}}^*$, σ^2 and $\tilde{\boldsymbol{\Omega}}(\boldsymbol{\gamma})$, update $\tilde{\boldsymbol{\beta}}$ based on $\pi(\tilde{\boldsymbol{\beta}} \mid \boldsymbol{Y}^*, \tilde{\boldsymbol{X}}^*, \sigma^2, \boldsymbol{\phi}, \boldsymbol{\gamma})$;

Step 4. Conditional on current values of $\boldsymbol{Y}^*$, $\tilde{\boldsymbol{X}}^*$, $\tilde{\boldsymbol{\beta}}$, $\boldsymbol{\phi}$ and $\tilde{\boldsymbol{\Omega}}(\boldsymbol{\gamma})$, update σ^2 based on $\pi(\sigma^2 \mid \boldsymbol{Y}^*, \tilde{\boldsymbol{X}}^*, \tilde{\boldsymbol{\beta}}, \boldsymbol{\phi}, \boldsymbol{\gamma})$;

Step 5. Update $\boldsymbol{\epsilon}$ conditional on current value of $\tilde{\boldsymbol{\beta}}$ and $\boldsymbol{Y}$, $\tilde{\boldsymbol{X}}$. Then update $\boldsymbol{\phi}$ based on $\pi(\boldsymbol{\phi} \mid \boldsymbol{Y}, \tilde{\boldsymbol{X}}, \tilde{\boldsymbol{\beta}}, \sigma^2, \boldsymbol{\gamma})$. Reject samples if the roots of $\boldsymbol{\phi}(L)$ lie outside the unit circle;

Step 6. Update (γ_1, γ_2) based on $\pi(\boldsymbol{\gamma} \mid \tilde{\boldsymbol{\beta}}, \sigma^2)$. Sample a proposal γ_i^* by $\gamma_i^* = \gamma_i + a * U(-1, 1)$ for $i = 1, 2$. a is an adjustable step size. Compute the ratio

$$R_\gamma = \frac{\pi(\boldsymbol{\gamma}^* \mid \tilde{\boldsymbol{\beta}}, \sigma^2)}{\pi(\boldsymbol{\gamma} \mid \tilde{\boldsymbol{\beta}}, \sigma^2)}$$

if $\tilde{\boldsymbol{\Omega}}(\boldsymbol{\gamma}^*)$ *is positive definite* **then**

update $\boldsymbol{\gamma} = \boldsymbol{\gamma}^*$ with probability $\min(1, R_\gamma)$;

end

end

2.4 Simulation Study

2.4.1 Performance in Estimating Lag Coefficients

In this section, we evaluate the performance of the proposed BDLM-AR using simulation studies. At the end of each simulated trial, we fitted BDLM-AR with lag $L = 7$ and AR(1),

that is, having

$$Y_t = \mu + \sum_{l=0}^{7} \beta_l X_{t-l} + \epsilon_t \tag{2.15}$$

where $\epsilon_t = \phi\epsilon_{t-1} + w_t$ and $w_t \sim N(0, \sigma^2)$. Posterior distributions were derived using the hybrid Metropolis Hastings/Gibbs algorithm described in the previous section with 50,000 iterations, a burn-in period of 25,000, and $a = 0.2$ for sampling γ in the MH step.

We compared BDLM-AR with some existing methods including the Bayesian distributed lag model (BDLagM; Welty et al., 2009), Bayesian ridge DLM (BR-DLM) with a mean zero normal prior for $\tilde{\boldsymbol{\beta}}$, and a non-informative prior Bayesian DLM (NB-DLM) with a flat improper priors on each parameter in $\tilde{\boldsymbol{\beta}}$. These existing methods would use the same mean model (2.15) but assume independent errors without accounting for autocorrelation.

In addition, as a benchmark, we include the parametric Koyck's DLM (Koyck, 1954) which assumes the knowledge of the true autoregressive coefficients. Details of the model specifications of the competing methods are given in Section 2.4.4.

We consider N-of-1 trials where measurements are collected daily for 120 days, and consider two different treatment sequences. In the first sequence, a participant would receive $x_t = 1$ on the first 30 days and the last 30 days, and receive $x_t = 0$ between days 31 and 90; that is,

$$x_t^{(1)} = \begin{cases} 1 & t = 30s + 1, ..., 30s + 30 \text{ for } s = 0, 3 \\ 0 & t = 30s + 1, ..., 30s + 30 \text{ for } s = 1, 2. \end{cases}$$

In the second treatment sequence, a participant would switch treatments more frequently; specifically,

$$x_t^{(2)} = \begin{cases} 1 & t = 15s + 1, ..., 15s + 15 \text{ for } s = 0, 3, 5, 6 \\ 0 & t = 15s + 1, ..., 15s + 15 \text{ for } s = 1, 2, 4, 7. \end{cases}$$

For each given treatment sequence, the data are generated according to model (2.15) under five sets of lag coefficients (lag curves) reflecting different patterns of carryover treatment effects:

1. Exponential decay curve: $\boldsymbol{\beta} = (5, 2.5, 1.25, 0.625, 0.3125, 0, 0, 0)^T$;

2. Exponential decay curve with oscillation: $\boldsymbol{\beta} = (5, 2.5, -1.25, -0.625, 0.3125, 0, 0, 0)^T$;

3. Slow absorption curve: $\boldsymbol{\beta} = (1.51, 2.75, 3.36, 2.03, 0.34, 0, 0, 0)^T$;

4. Slow absorption curve with oscillation: $\boldsymbol{\beta} = (1.51, 2.75, -3.36, -2.03, 0.34, 0, 0, 0)^T$;

5. No carryover effect: $\boldsymbol{\beta} = (10, 0, 0, 0, 0, 0, 0, 0)^T$.

The exponential decay curves (1) and (2) specify coefficients that diminish in magnitude as lag lengthens. The slow absorption curves (3) and (4) reflect scenarios where there is a delay for the carryover effects to peak. The last scenario (5) is the null case where there is no carryover effect. The coefficients in exponential curve (1) decrease geometrically, which is aligned with the assumption of Koyck DLM. The lag curves of all five settings are given in Figure 2.1. The total effect in each scenario is 10 and total carryover effects (δ) are 4.69, 0.94, 8.48, -2.30 and 0 respectively.

We consider the two values of the standard deviation σ to mimic two signal to noise ratios (SNR = $\sum_{l=0}^{7} |\beta_l|/\sigma$): $\sigma = 10$ so that SNR= 1 for strong treatment effect, and $\sigma = 20$ so that SNR = 0.5 for weak treatment effect. Two different autoregressive coefficients are used: $\phi = 0.5$ for strong serial correlation and $\phi = 0.2$ for weak serial correlation. In the simulation below, we focus on the setting of bimonthly switch treatment sequence, strong treatment effect (SNR = 1) and strong serial correlation ($\phi = 0.5$). We will then compare this setting with settings that alter only one of the three specifications (i.e., treatment generating sequence, σ or ϕ) while keeping the other two unchanged.

For each simulation scenario, 100 data sets were generated. We fit the proposed BDLM-AR model with lag $L = 7$ and AR(1) for the error terms using the hybrid Metropolis-Hastings/Gibbs algorithm in section 2.3 with 50,000 iterations, a burn-in period of 25,000 and tuning parameter $a = 0.2$. We check the convergence of all the MCMC simulations using both trace plots and the GelmanRubin diagnostics (Gelman, Rubin, et al., 1992). The potential scale reduction factor for

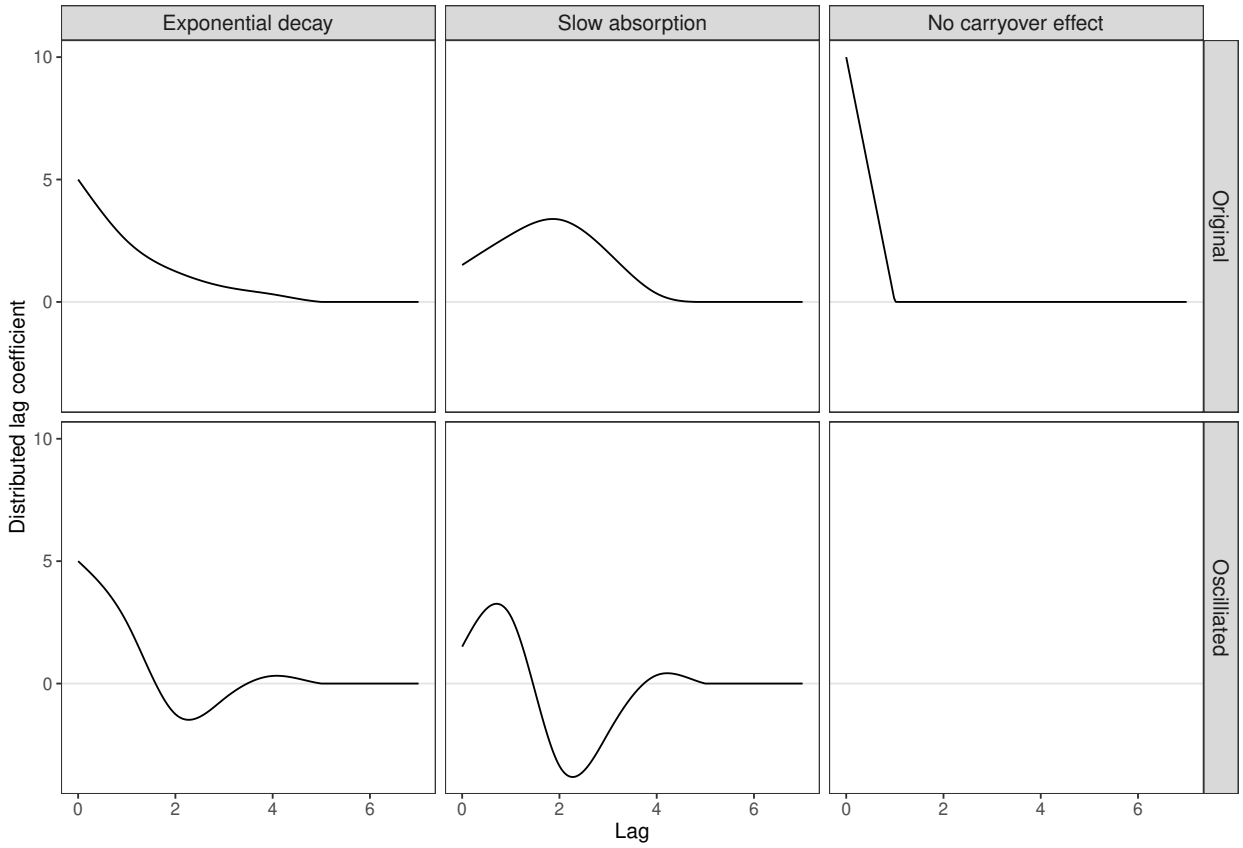


Figure 2.1: Hypothetical true lag coefficients curve in the simulation. Columns represent shape of lag coefficients curve. Rows represent variation based on the original shape, from top to bottom are: no variation and variate with oscillation (every 2 days).

lag coefficients, autoregressive coefficients and model variance all are smaller than 1.2, as Brooks and Gelman (1998) have suggested, indicating the convergence of posterior samples.

Evaluation of the lag coefficient estimation performance are based on the bias and root mean square error (RMSE) for each lag parameter, the total treatment effect ($\sum_{l=0}^{L} \beta_l$) and the total carryover effect ($\sum_{l=1}^{L} \beta_l$) over 100 simulation data sets. To compare the overall performance of each method in fitting the true lag curve, Euclidean distance is used to measure the estimation property of estimated lag coefficient vector $\hat{\boldsymbol{\beta}}$ to the true lag coefficient vector $\boldsymbol{\beta}$ by using formula: Distance = $||\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}||_2$.

First, we investigate the performance of proposed BDLM-AR model in estimating individual lag coefficients. Figure 2.2 shows the bias and RMSE of each estimated lag coefficient of five models under different scenarios. Figure 2.3 shows the estimated lag curve of five models under different scenarios. The 90% posterior mean band represents the distribution of posterior mean of each lag coefficient under 100 simulations, rather than 90% credible interval of each coefficient estimation. The NB-DLM method has the largest RMSE and relatively low bias in most parameters. This is expected since no regularization is imposed on the lag coefficients due to the use of improper priors, which yields low bias, but the variance is inflated due to collinearity. The normal prior in BR-DLM method acts as an ℓ_2 penalty on the lag coefficients, and thus the RMSE of BR-DLM is consistently smaller than that of NB-DLM across different lag coefficients with slightly inflated bias. Both BDLagM method and the proposed BDLM-AR method place increasing constraint on lag coefficients so that the bias and variance of posterior mean estimator decrease as lag increases, providing small RMSE at large lags. In addition, compared to BDLagM method, the proposed BDLM-AR method explicitly incorporates ridge type of regularization on lag coefficients, which will achieve much smaller RMSE at first several lag coefficients. This is the most common situation in N-of-1 trials since immediate response from patients usually accounts for a large proportion of total treatment effect. However, as a trade-off, the bias of BDLM-AR for early lag coefficients will be slightly inflated as compared to BDLagM, BR-DLM and NB-DLM, especially when true lag curve has frequent fluctuation. The benchmark method, Koyck DLM,

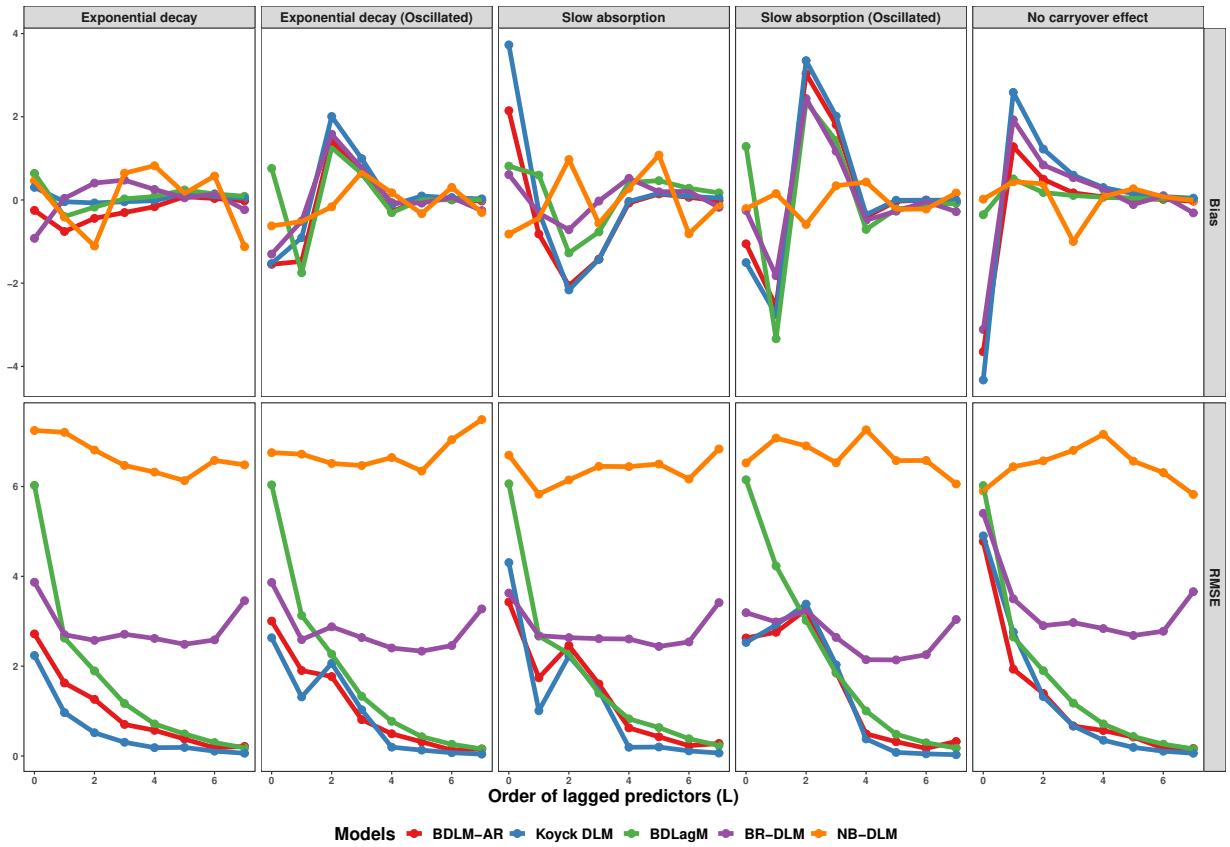


Figure 2.2: Bias and RMSE of posterior mean lag coefficients under five different true lag curves. Bimonthly switch treatment sequence, strong treatment effect (SNR = 1) and strong serial correlation ($\phi = 0.5$) are used to generate simulated data.

performs best in the exponential decay case, where the true coefficients of autoregressive error (ϕ) are assumed to be known. The proposed BDLM-AR has very similar performance as Koyck DLM. Note that the autoregressive errors are estimated directly from proposed BDLM-AR model, which is more practical in real application. In summary, the proposed BDLM-AR has the smallest RMSE in nearly all lag coefficients.

Table 2.1 summarizes the results of total effect ($\sum_{l=0}^{7}\beta_l$), total carryover effect ($\sum_{l=i}^{7}\beta_l$) and immediate effect (β_0) across different lag curves. Our proposed BDLM-AR model has comparable performance as other models in estimating total effect, and outperforms BDLagM and NB-DLM in terms of RMSE of total carryover effect in most scenarios, especially when the true lag curve does not contain carryover effect. This scenario indicates that our proposed model is less likely to report false positive carryover effect. A summary of the evaluation metrics of all parameters

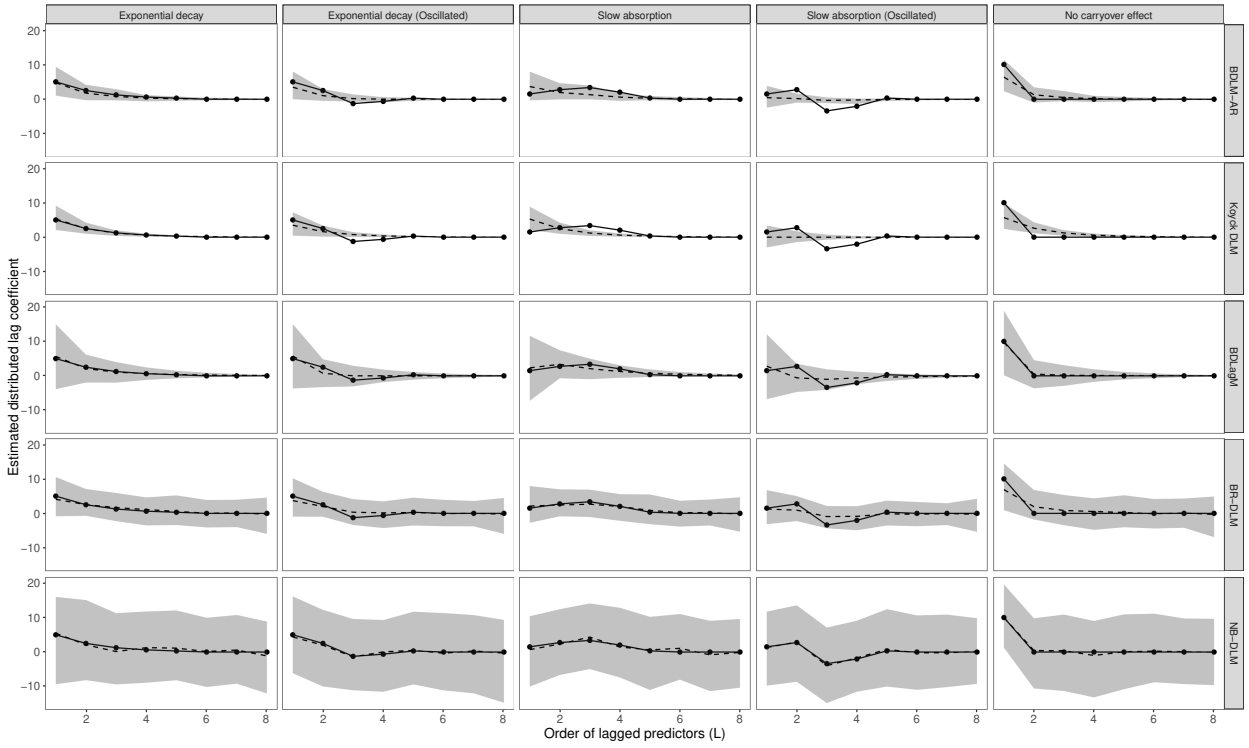


Figure 2.3: Posterior mean estimated lag curve (dashed) and true lag curve (solid) with 90% posterior bands (grey) under five lag curves, estimated by five models. Treatment sequence $x_t^{(1)}$, strong treatment effect (SNR = 1) and strong serial correlation (ϕ = 0.5) are used to generate simulated data.

		BDLM-AR	Koyck DLM	BDLagM	BR-DLM	NB-DLM
Bias	**Total Effect**					
	Exponential decay	-1.82	0.41	0.65	0.21	**0.02**
	Exponential decay (Oscillated)	-1.24	0.61	0.61	**0.20**	-0.79
	Slow absorption	-2.04	**0.12**	0.71	0.30	-0.42
	Slow absorption (Oscillated)	0.73	0.73	0.59	0.46	**-0.15**
	No carryover	-1.60	0.65	0.58	**0.14**	0.25
	Total Carryover Effect					
	Exponential decay	-1.57	0.10	**0.01**	1.13	-0.44
	Exponential decay (Oscillated)	0.31	2.14	**-0.15**	1.50	-0.17
	Slow absorption	-4.19	-3.61	**-0.11**	-0.31	0.40
	Slow absorption (Oscillated)	1.78	2.23	-0.70	0.71	**0.06**
	No carryover	2.05	4.98	0.94	3.25	**0.23**
	Immediate Effect					
	Exponential decay	**-0.25**	0.30	0.64	-0.92	0.46
	Exponential decay (Oscillated)	-1.55	-1.53	0.76	-1.30	**-0.62**
	Slow absorption	2.15	3.73	0.81	**0.61**	-0.82
	Slow absorption (Oscillated)	-1.05	-1.51	1.29	-0.25	**-0.21**
	No carryover	-3.65	-4.33	-0.35	-3.11	**0.02**
RMSE	**Total Effect**					
	Exponential decay	**3.61**	3.87	4.05	4.17	3.96
	Exponential decay (Oscillated)	**2.95**	3.90	4.05	4.11	3.93
	Slow absorption	**3.81**	3.85	4.06	4.16	4.40
	Slow absorption (Oscillated)	**2.28**	3.93	4.04	4.01	3.99
	No carryover	**3.54**	3.91	4.04	4.18	3.93
	Total Carryover Effect					
	Exponential decay	3.03	**2.01**	6.57	4.83	7.61
	Exponential decay (Oscillated)	**2.25**	2.87	6.56	4.83	6.31
	Slow absorption	5.13	**4.15**	6.58	4.64	6.94
	Slow absorption (Oscillated)	2.96	**2.94**	6.62	4.44	6.39
	No carryover	**3.39**	5.36	6.65	6.06	6.56
	Immediate Effect					
	Exponential decay	2.71	**2.23**	6.00	3.85	7.21
	Exponential decay (Oscillated)	2.99	**2.62**	6.01	3.85	6.72
	Slow absorption	**3.42**	4.30	6.03	3.61	6.67
	Slow absorption (Oscillated)	2.62	**2.52**	6.12	3.18	6.49
	No carryover	**4.77**	4.90	5.99	5.39	5.87

Table 2.1: Summary of evaluation metrics (best values in bold) of total effect, total carryover effect and immediate effect(β_0) under five lag curves, estimated by five models. Treatment sequence $x_t^{(1)}$, strong treatment effect (SNR = 1) and strong serial correlation (ϕ = 0.5) are used to generate simulated data.

	Bias						RMSE				
	Truth	BDLM-AR	Koyck DLM	BDLagM	BR-DLM	NB-DLM	BDLM-AR	Koyck DLM	BDLagM	BR-DLM	NB-DLM
μ	10	1.19	-0.11	-0.34	-0.11	-0.41	2.31	2.71	2.66	2.68	2.35
Total effect	9.69	-1.82	0.41	0.65	0.21	0.02	3.61	3.87	4.05	4.17	3.96
Total carryover effect	4.69	-1.57	0.10	0.01	1.13	-0.44	3.03	2.01	6.57	4.83	7.61
β_0	5	-0.25	0.30	0.64	-0.92	0.46	2.71	2.23	6.00	3.85	7.21
β_1	2.5	-0.76	-0.04	-0.40	0.04	-0.41	1.62	0.96	2.61	2.69	7.17
β_2	1.25	-0.44	-0.07	-0.17	0.41	-1.11	1.26	0.52	1.89	2.56	6.78
β_3	0.62	-0.30	-0.04	0.03	0.48	0.64	0.70	0.30	1.16	2.70	6.44
β_4	0.31	-0.16	-0.02	0.08	0.25	0.82	0.57	0.18	0.71	2.60	6.29
β_5	0	0.08	0.15	0.24	0.05	0.16	0.38	0.19	0.49	2.47	6.10
β_6	0	0.03	0.08	0.15	0.14	0.58	0.18	0.11	0.30	2.57	6.55
β_7	0	-0.02	0.04	0.09	-0.23	-1.12	0.21	0.06	0.18	3.44	6.45
ϕ	0.5	0.00	-	-	-	-	0.08	-	-	-	-
σ	10	-0.04	-0.08	1.47	1.21	1.27	0.63	0.66	1.79	1.57	1.59

Table 2.2: Summary of evaluation metrics of total effect($\sum_{l=0}^{L} \beta_l$), total carryover effect ($\sum_{l=1}^{L} \beta_l$), lag coefficients, autoregressive coefficient(ϕ) and model standard deviation(σ) under exponential decay lag curve, estimated by five models. Treatment sequence $x_t^{(1)}$, strong treatment effect (SNR = 1) and strong serial correlation ($\phi = 0.5$) are used to generate simulated data.

	Bias						RMSE				
	Truth	BDLM-AR	Koyck DLM	BDLagM	BR-DLM	NB-DLM	BDLM-AR	Koyck DLM	BDLagM	BR-DLM	NB-DLM
μ	10	0.91	-0.19	-0.33	-0.11	0.38	2.07	2.72	2.66	2.66	2.71
Total effect	5.94	-1.24	0.61	0.61	0.20	-0.79	2.95	3.90	4.05	4.11	3.93
Total carryover effect	0.94	0.31	2.14	-0.15	1.50	-0.17	2.25	2.87	6.56	4.83	6.31
β_0	5	-1.55	-1.53	0.76	-1.30	-0.62	2.99	2.62	6.01	3.85	6.72
β_1	2.5	-1.48	-0.91	-1.75	-0.51	-0.52	1.90	1.31	3.12	2.58	6.69
β_2	-1.25	1.42	2.01	1.26	1.58	-0.16	1.77	2.06	2.26	2.87	6.48
β_3	-0.62	0.65	1.00	0.64	0.77	0.67	0.81	1.03	1.32	2.63	6.44
β_4	0.31	-0.28	-0.13	-0.31	-0.06	0.17	0.49	0.20	0.77	2.39	6.61
β_5	0	0.01	0.10	0.00	-0.08	-0.33	0.31	0.13	0.43	2.32	6.31
β_6	0	0.01	0.05	0.00	0.06	0.30	0.13	0.07	0.26	2.45	7.01
β_7	0	-0.02	0.03	0.00	-0.26	-0.30	0.16	0.04	0.16	3.26	7.45
ϕ	0.5	-0.01	-	-	-	-	0.08	-	-	-	-
σ	10	-0.05	-0.07	1.47	1.22	1.21	0.63	0.67	1.79	1.57	1.54

Table 2.3: Summary of evaluation metrics under exponential decay (oscillated) lag curve.

	Bias						RMSE				
	Truth	BDLM-AR	Koyck DLM	BDLagM	BR-DLM	NB-DLM	BDLM-AR	Koyck DLM	BDLagM	BR-DLM	NB-DLM
μ	10	1.30	0.02	-0.36	-0.17	0.04	2.39	2.71	2.67	2.68	3.04
Total effect	9.99	-2.04	0.12	0.71	0.30	-0.42	3.81	3.85	4.06	4.16	4.40
Total carryover effect	8.49	-4.19	-3.61	-0.11	-0.31	0.40	5.13	4.15	6.58	4.64	6.94
β_0	1.51	2.15	3.73	0.81	0.61	-0.82	3.42	4.30	6.03	3.61	6.67
β_1	2.75	-0.82	-0.28	0.60	-0.32	-0.43	1.73	1.00	2.66	2.67	5.80
β_2	3.36	-2.06	-2.16	-1.27	-0.72	0.97	2.46	2.23	2.27	2.62	6.12
β_3	2.03	-1.42	-1.43	-0.77	-0.03	-0.56	1.60	1.46	1.40	2.60	6.42
β_4	0.34	-0.07	-0.04	0.42	0.52	0.28	0.62	0.19	0.82	2.59	6.41
β_5	0	0.14	0.16	0.46	0.20	1.08	0.43	0.20	0.63	2.43	6.47
β_6	0	0.06	0.09	0.28	0.21	-0.81	0.23	0.11	0.38	2.53	6.14
β_7	0	-0.02	0.05	0.17	-0.18	-0.12	0.28	0.07	0.23	3.40	6.80
ϕ	0.5	0.00	-	-	-	-	0.08	-	-	-	-
σ	10	-0.03	-0.06	1.48	1.21	1.35	0.62	0.65	1.79	1.57	1.63

Table 2.4: Summary of evaluation metrics under slow absorption lag curve.

		Bias					RMSE				
	Truth	BDLM-AR	Koyck DLM	BDLagM	BR-DLM	NB-DLM	BDLM-AR	Koyck DLM	BDLagM	BR-DLM	NB-DLM
μ	10	-0.05	-0.22	-0.32	-0.25	0.09	1.75	2.73	2.66	2.63	2.61
Total effect	-0.79	0.73	0.73	0.59	0.46	-0.15	2.28	3.93	4.04	4.01	3.99
Total carryover effect	-2.3	1.78	2.23	-0.70	0.71	0.06	2.96	2.94	6.62	4.44	6.39
β_0	1.51	-1.05	-1.51	1.29	-0.25	-0.21	2.62	2.52	6.12	3.18	6.49
β_1	2.75	-2.56	-2.76	-3.34	-1.82	0.15	2.75	2.92	4.22	2.98	7.04
β_2	-3.36	3.03	3.35	2.35	2.44	-0.59	3.26	3.38	3.01	3.24	6.87
β_3	-2.03	1.81	2.01	1.43	1.17	0.34	1.87	2.03	1.84	2.63	6.50
β_4	0.34	-0.41	-0.35	-0.71	-0.49	0.43	0.49	0.38	1.00	2.13	7.23
β_5	0	-0.03	-0.01	-0.22	-0.27	-0.22	0.32	0.08	0.48	2.13	6.54
β_6	0	0.00	-0.01	-0.13	-0.04	-0.22	0.17	0.05	0.29	2.24	6.55
β_7	0	-0.05	0.00	-0.08	-0.28	0.17	0.32	0.03	0.18	3.02	6.03
ϕ	0.5	-0.01	-	-	-	-	0.08	-	-	-	-
σ	10	-0.07	-0.06	1.48	1.21	1.41	0.64	0.67	1.79	1.57	1.73

Table 2.5: Summary of evaluation metrics under slow absorption (oscillated) lag curve.

		Bias					RMSE				
	Truth	BDLM-AR	Koyck DLM	BDLagM	BR-DLM	NB-DLM	BDLM-AR	Koyck DLM	BDLagM	BR-DLM	NB-DLM
μ	10	1.07	-0.25	-0.31	-0.07	0.12	2.25	2.72	2.66	2.69	2.54
Total effect	10	-1.60	0.65	0.58	0.14	0.25	3.54	3.91	4.04	4.18	3.93
Total carryover effect	0	2.05	4.98	0.94	3.25	0.23	3.39	5.36	6.65	6.06	6.56
β_0	10	-3.65	-4.33	-0.35	-3.11	0.02	4.77	4.90	5.99	5.39	5.87
β_1	0	1.28	2.59	0.51	1.93	0.44	1.93	2.76	2.64	3.49	6.41
β_2	0	0.50	1.22	0.18	0.84	0.39	1.39	1.32	1.89	2.89	6.54
β_3	0	0.17	0.60	0.11	0.54	-1.00	0.66	0.67	1.17	2.96	6.77
β_4	0	0.07	0.30	0.07	0.27	0.08	0.56	0.35	0.71	2.82	7.13
β_5	0	0.04	0.15	0.04	-0.11	0.27	0.42	0.19	0.43	2.67	6.53
β_6	0	0.01	0.08	0.02	0.10	0.08	0.17	0.11	0.26	2.76	6.28
β_7	0	-0.03	0.04	0.01	-0.31	-0.02	0.17	0.06	0.16	3.64	5.79
ϕ	0.5	-0.01	-	-	-	-	0.08	-	-	-	-
σ	10	-0.03	-0.06	1.47	1.23	1.49	0.63	0.67	1.79	1.59	1.77

Table 2.6: Summary of evaluation metrics under no carryover effect lag curve.

comparing five models under different lag curves can be found in Table 2.2 to 2.6.

Treatment sequence, signal to noise ratio and magnitude of autocorrelation are other simulation parameters of interest. Figure 2.4 to 2.6 show comparison of RMSE of total effect and total carryover effect between different models under different simulation design settings.

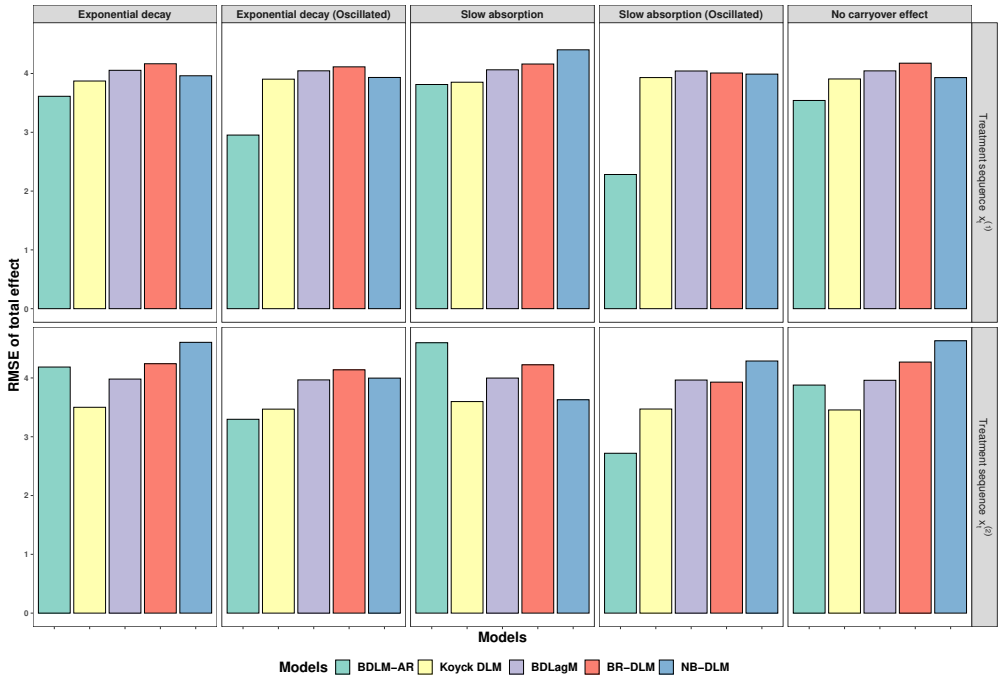
The overall performance in estimating the lag coefficient vector is summarized in Figure 2.7. Two rows represent treatment sequence $x_t^{(1)}$ and $x_t^{(2)}$ respectively while adjusting for signal to noise ratio and magnitude of autocorrelation. The distance $||\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}||_2$ between the posterior means/MLE and the truth under $x_t^{(2)}$ (bottom panel) is smaller than that under $x_t^{(1)}$ (top panel) suggesting frequently switching treatments will help improve the information content in N-of-1 trial data. This is in line with what we expect because collinearity of exposure lags will be lessened as treatments change frequently, while the total duration is held fixed. An implication in practice is that we should alternate intervention as frequent as it is feasible. The relative performance among methods is the same regardless of the treatment sequence, that is, the proposed BDLM-AR yields the shortest distance from the true lag coefficients $\boldsymbol{\beta}$. Comparison of distance to true lag curves under other simulation parameters can be found in Figure 2.8 and 2.9.

2.4.2 Effects of Model Misspecification

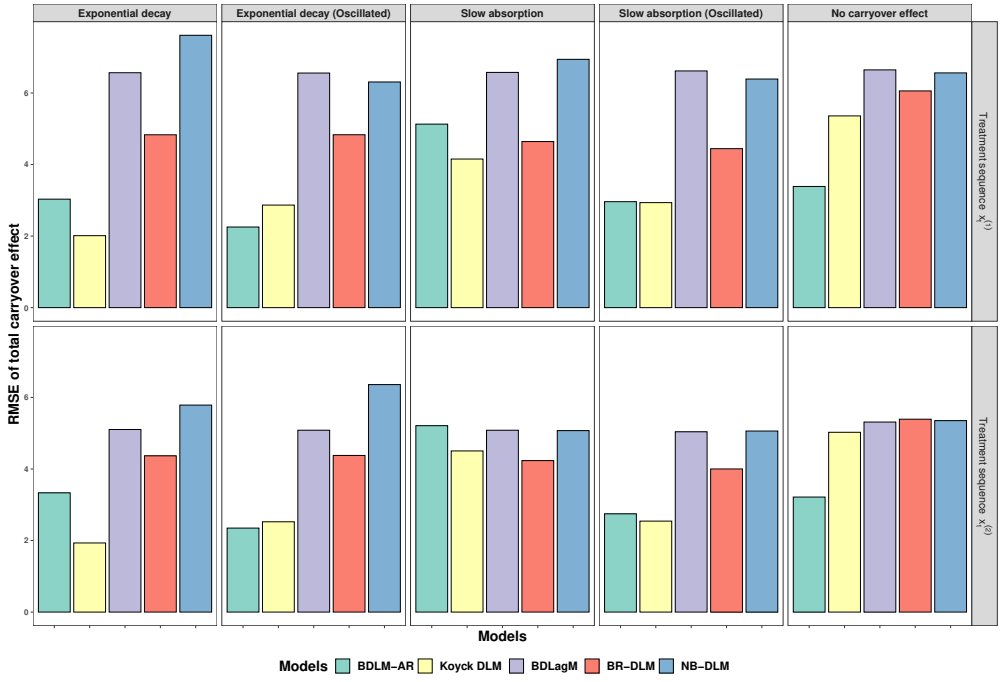
In the previous subsections, BDLM-AR and other methods use a working mean model with $L = 7$ and an AR(1) model for autoregressive errors. These working models correctly specify (or include the data generation model) in the previous simulation study. In this subsection, we investigate the robustness of BDLM-AR under model mis-specifications. Specifically, we will consider (1) the working mean model with $L = 0, 1, \ldots, 7$; (2) the stochastic components that assumes autoregressive error order of $p = 0, 1, 7$. That is, we consider a total of 24 BDLM-AR models.

In data generation, we use exponential decay curve as the true mean model, where $\beta_l > 0$ for $l = 0, 1, 2, 3, 4$, and we consider true scenarios for the errors:

1. AR(1) with $\phi = 0.5$

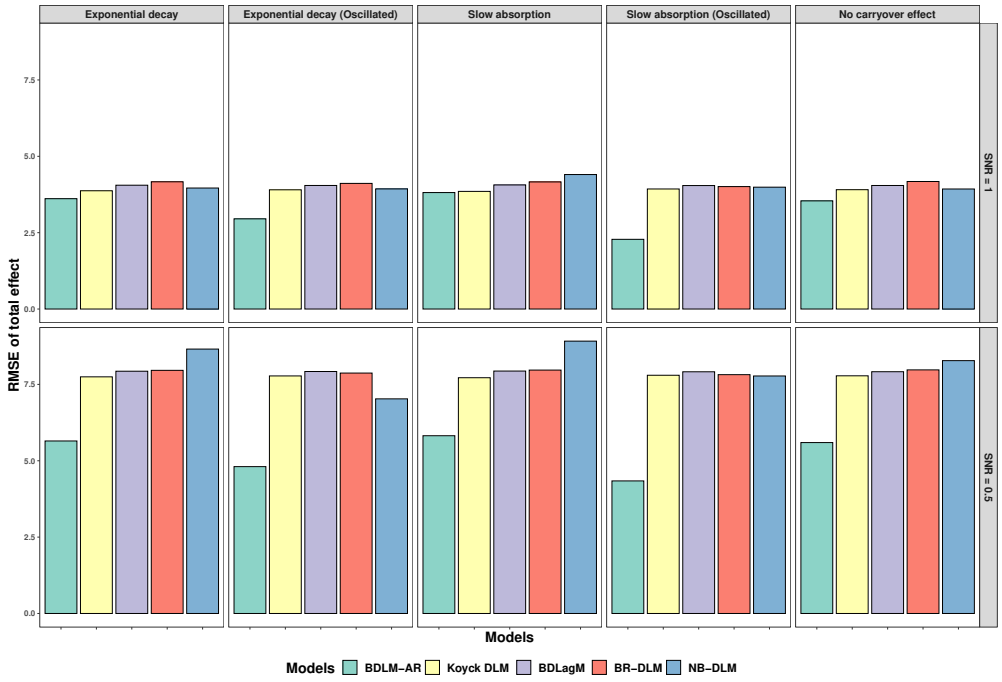


(a)

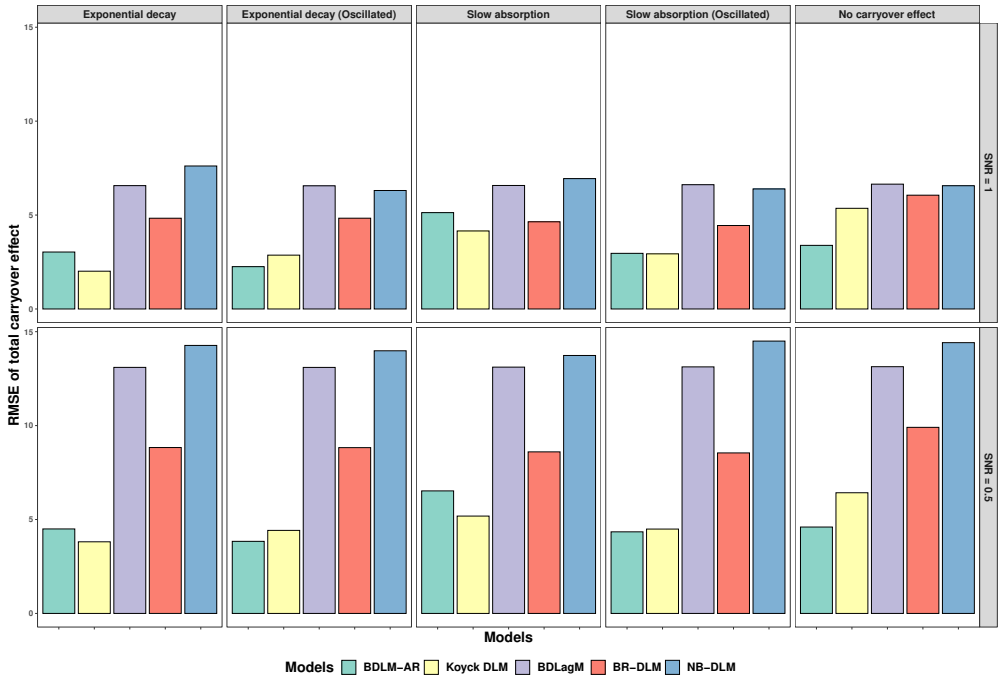


(b)

Figure 2.4: RMSE of (a) Total effect (b) Total carryover effect under: Treatment sequence $x_t^{(1)}$ vs. Treatment sequence $x_t^{(2)}$.

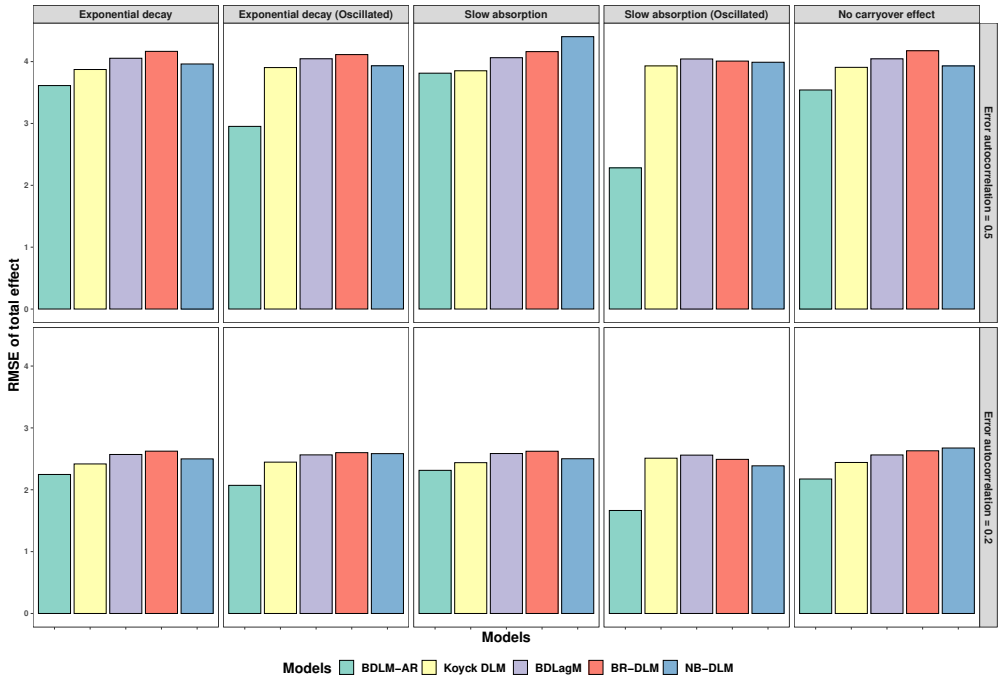


(a)

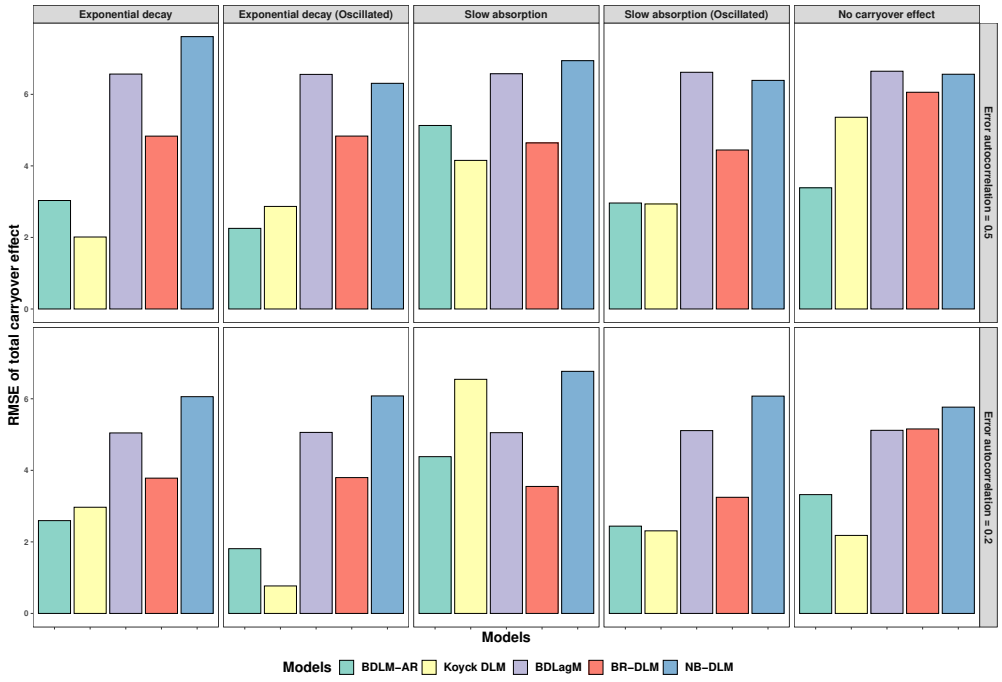


(b)

Figure 2.5: RMSE of (a) Total effect (b) Total carryover effect under: strong signal to noise ratio (SNR = 1) vs. weak signal to noise ratio (SNR = 0.5).



(a)



(b)

Figure 2.6: RMSE of (a) Total effect (b) Total carryover effect under: strong serial correlation (ϕ = 0.5) vs. weak serial correlation (ϕ = 0.2).

2. Autoregressive model with $\phi_1 = 0.5, \phi_2 = 0, \phi_3 = 0, \phi_4 = 0.3, \phi_5 = 0, \phi_6 = 0.2, \phi_7 = 0$

Note that, under the scenario 1, a working model with $L < 4$ or $p = 0$ under-specifies the true model. Likewise, under the scenario 2, a working model with $L < 4$ or $p = 0, 1$ under-specifies the true model.

Table 2.7 and 2.8 summarize the bias and RMSE of these 24 models under the two scenarios with $\sigma = 10$ under $x_t^{(1)}$ It can be seen that misspecified lag length has little influence on estimating total effect, total carryover effect and immediate effect, while under-specified error AR order will increase RMSE of parameters to a higher level than over-specified error AR order. Note that when choosing a small lag length value, we can hardly acquire estimation about the whole lag curve, as well as the information on the duration of carryover effect. Therefore, when the lag length is unknown, we suggest to fit data with a reasonable long lag length. For error autoregressive order, when the true orders are unknown, it is also suggested to fit a model with high autoregressive order.

2.4.3 Performance in Estimating Autoregressive Coefficients

To demonstrate the performance of proposed model under high order autoregressive errors, we conducted simulation to compare proposed BDLM-AR model with RegAR model. The data generation model is $y_t = \mu + \sum_{l=0}^{7} \beta_l x_{t-l} + \epsilon_t$, where $\epsilon_t = \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + \cdots + \phi_7 \epsilon_{t-7} + w_t$ and $w_t \sim N(0, \sigma^2)$ for $t = 8, 3, ..., 120$. Both models are fitted using AR(7) errors and AR(1) errors respectively, in order to evaluate the effect of neglecting high order autoregressive errors. In this simulation, we used exponential decay lag curve, strong SNR and treatment sequence $x_t^{(1)}$ across different models. Serial correlation $\boldsymbol{\phi}$ is set to be $\phi_1 = 0.5, \phi_2 = 0, \phi_3 = 0, \phi_4 = 0.3, \phi_5 = 0, \phi_6 = 0.2, \phi_7 = 0$.

Table 2.9 summarize the performance of proposed BDLM-AR model in estimating individual model parameter, averaged across 100 simulation data sets. BDLM-AR performs comparably to RegAR in estimating autoregressive coefficients and total effect. Comparing BDLM-AR model with AR(7) and AR(1) error, we observe when higher order positive autoregressive coefficients are omitted in estimation, the model variance will be under estimated. This will further affect the

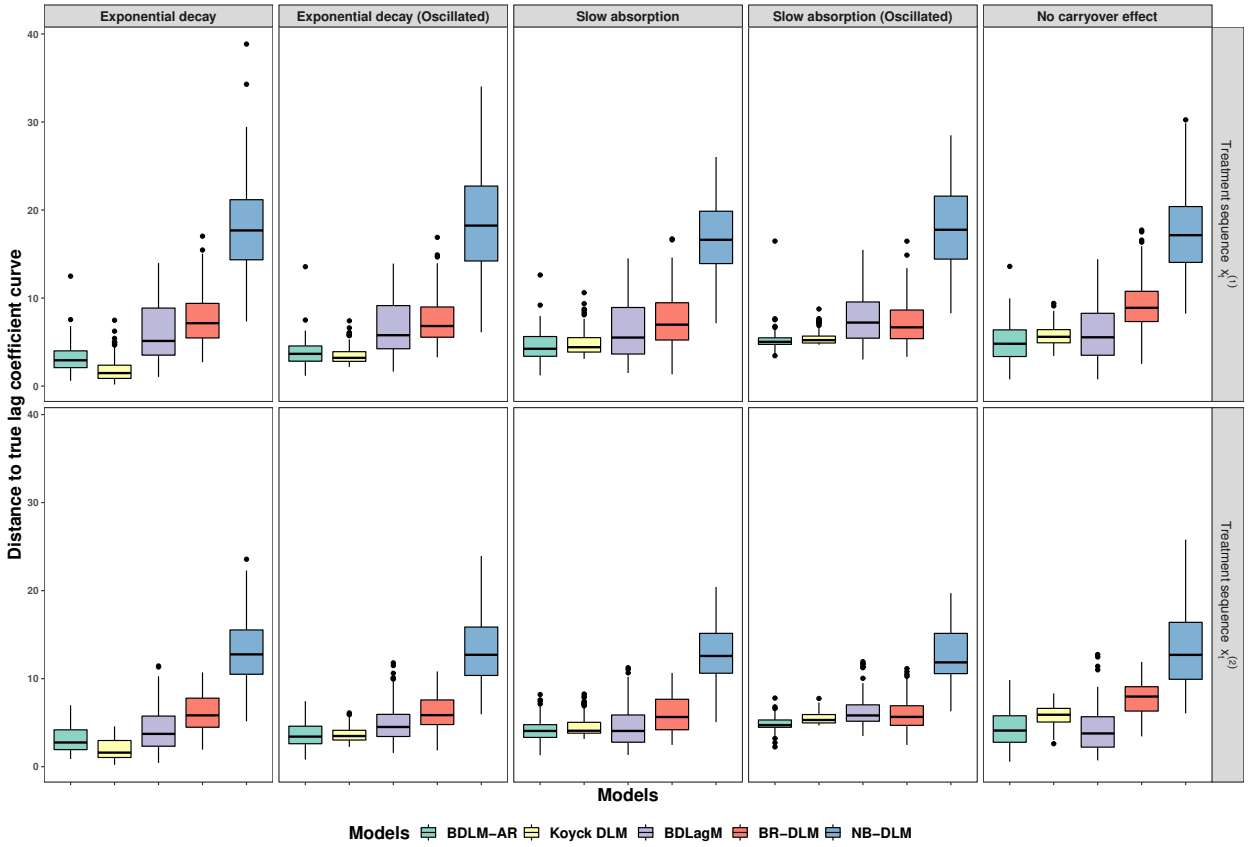


Figure 2.7: Distance to true lag curves under: Treatment sequence $x_t^{(1)}$ vs. Treatment sequence $x_t^{(2)}$.

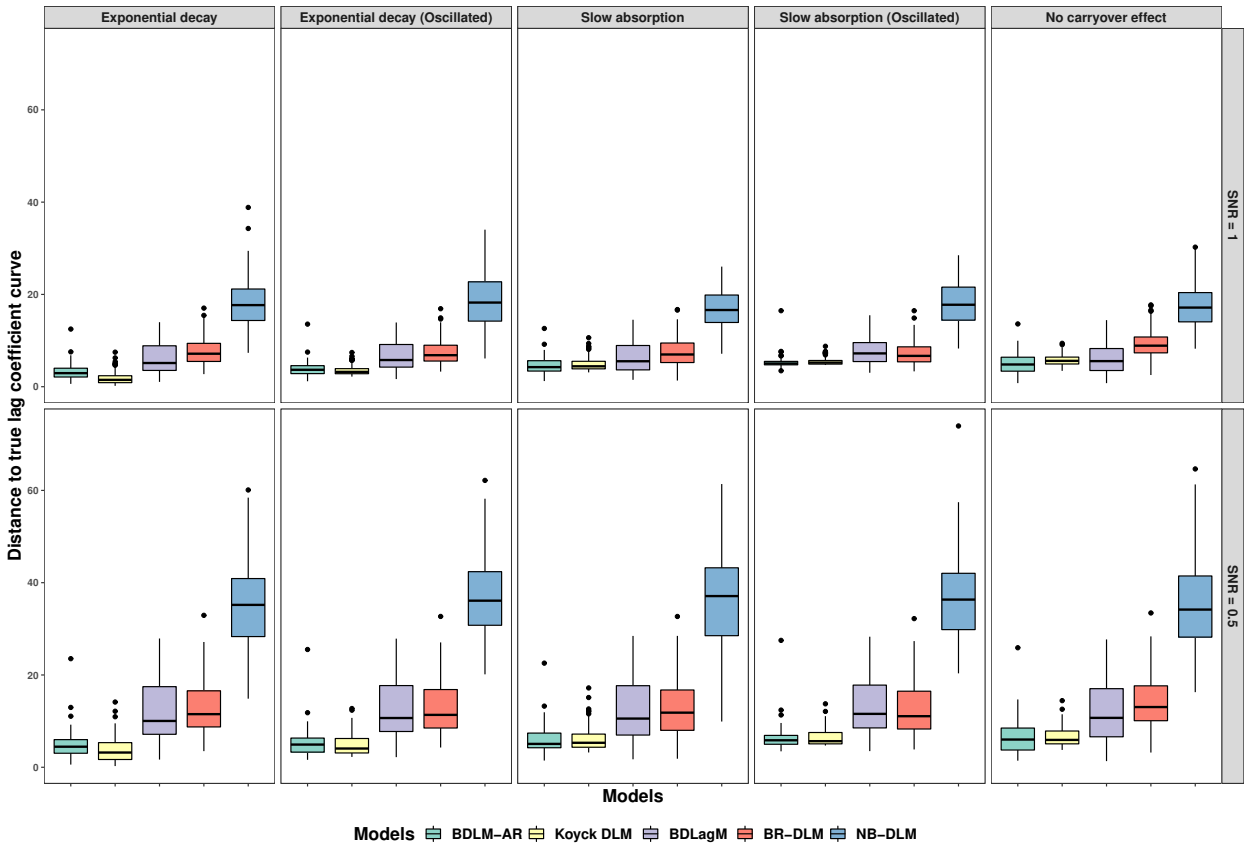


Figure 2.8: Distance to true lag coefficients curves under: strong signal to noise ratio (SNR = 1) vs. weak signal to noise ratio (SNR = 0.5).

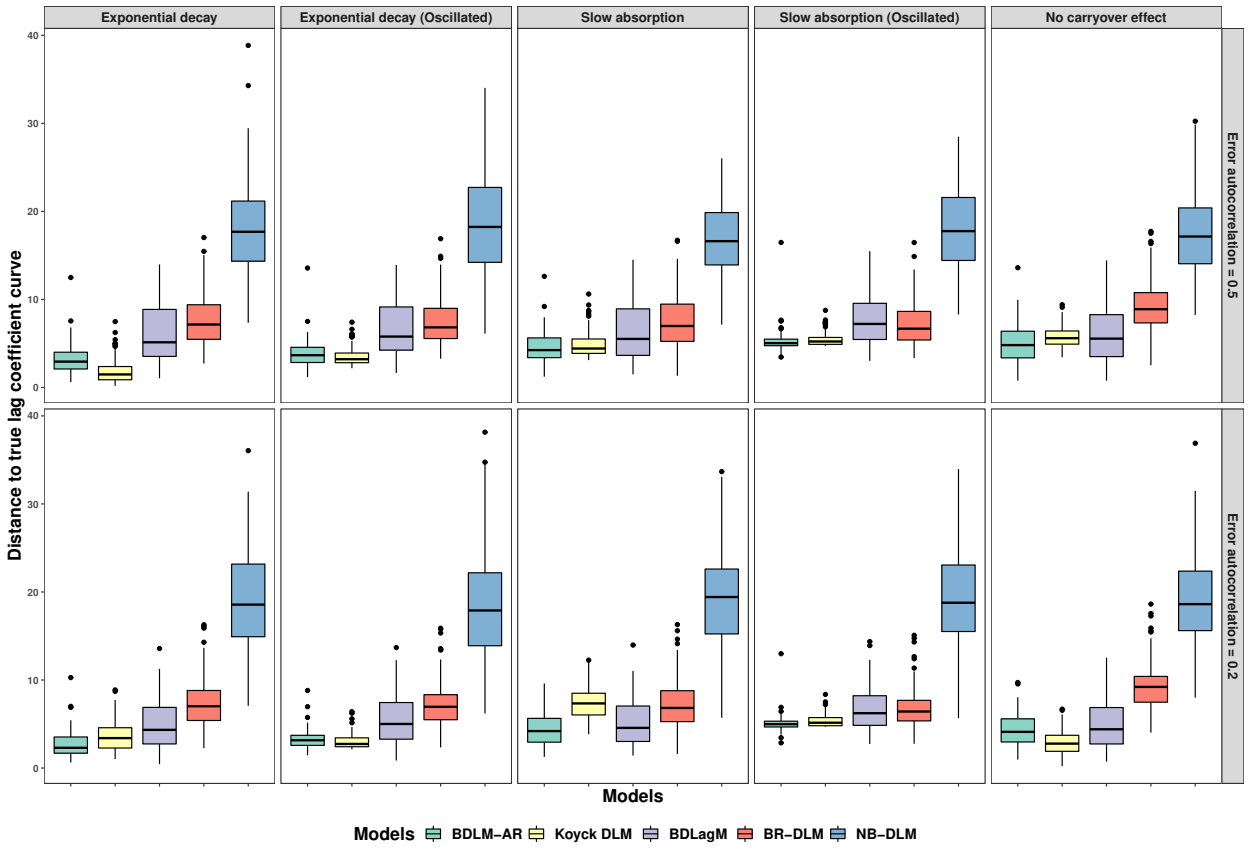


Figure 2.9: Distance to true lag coefficients curves under: strong serial correlation ($\phi = 0.5$) vs. weak serial correlation ($\phi = 0.2$).

		Bias			RMSE		
	Lag	AR(7)	AR(1)	AR(0)	AR(7)	AR(1)	AR(0)
Total Effect	7	-1.64	-1.25	-0.09	4.32	3.85	3.95
	6	-1.75	-1.31	-0.12	4.37	3.88	3.92
	5	-1.93	-1.42	-0.14	4.44	3.91	3.93
	4	-2.03	-1.55	-0.19	4.49	3.98	3.94
	3	-2.18	-1.67	-0.25	4.59	4.02	3.93
	2	-2.39	-1.85	-0.32	4.63	4.07	3.91
	1	-2.74	-2.13	-0.45	4.75	4.15	3.87
	0	-0.96	-0.44	-0.02	4.01	3.71	3.77
Total Carryover Effect	7	-1.40	-1.29	-0.53	3.56	3.18	3.85
	6	-1.44	-1.35	-0.60	3.51	3.17	3.73
	5	-1.59	-1.44	-0.65	3.42	3.15	3.61
	4	-1.64	-1.51	-0.73	3.41	3.12	3.49
	3	-1.78	-1.68	-0.81	3.41	3.12	3.36
	2	-2.12	-2.01	-1.05	3.38	3.10	3.15
	1	-2.86	-2.77	-1.87	3.58	3.33	2.92
	0	-	-	-	-	-	-
Immediate Effect	7	-0.24	0.04	0.44	3.09	2.91	3.54
	6	-0.31	0.04	0.48	3.02	2.91	3.47
	5	-0.34	0.02	0.51	3.03	2.87	3.38
	4	-0.39	-0.03	0.54	3.04	2.85	3.31
	3	-0.40	0.02	0.57	3.03	2.88	3.28
	2	-0.27	0.17	0.74	3.09	2.97	3.29
	1	0.12	0.64	1.42	3.15	3.15	3.49
	0	3.73	4.25	4.67	5.39	5.62	6.00

Table 2.7: Summary of evaluation metrics of total effect, total carryover effect and immediate effect(β_0) fitted using BDLM-AR model with different lag length and error autoregressive order. Exponential decay curve and serial correlation $\phi = 0.5$ are used to generate simulated data.

		Bias			RMSE		
	Lag	AR(7)	AR(1)	AR(0)	AR(7)	AR(1)	AR(0)
Total Effect	7	-2.46	-1.63	0.71	5.85	8.32	10.72
	6	-2.58	-1.84	0.66	5.92	8.25	10.62
	5	-2.86	-1.93	0.63	5.89	8.21	10.55
	4	-2.89	-2.08	0.54	6.01	8.06	10.48
	3	-3.24	-2.41	0.45	6.02	7.81	10.40
	2	-3.56	-2.75	0.33	6.08	7.68	10.34
	1	-4.03	-3.45	0.16	6.12	7.43	10.26
	0	-1.66	-1.52	0.57	5.82	8.58	10.52
Total Carryover Effect	7	-1.70	-0.97	0.06	3.55	5.24	6.69
	6	-1.71	-1.10	-0.05	3.60	5.15	6.38
	5	-1.86	-1.23	-0.08	3.60	4.95	6.08
	4	-1.94	-1.37	-0.24	3.73	4.67	5.68
	3	-2.19	-1.71	-0.38	3.70	4.29	5.33
	2	-2.51	-2.19	-0.73	3.73	4.00	4.62
	1	-3.24	-3.11	-1.80	3.86	3.88	3.54
	0	-	-	-	-	-	-
Immediate Effect	7	-0.75	-0.65	0.66	3.61	4.54	8.52
	6	-0.87	-0.73	0.71	3.56	4.50	8.46
	5	-1.00	-0.70	0.72	3.58	4.52	8.40
	4	-0.95	-0.71	0.79	3.57	4.54	8.29
	3	-1.05	-0.70	0.84	3.54	4.66	8.25
	2	-1.06	-0.56	1.06	3.64	4.91	8.28
	1	-0.78	-0.34	1.96	3.71	5.21	8.48
	0	3.03	3.17	5.26	6.35	9.02	11.75

Table 2.8: Summary of evaluation metrics of total effect, total carryover effect and immediate effect(β_0) fitted using BDLM-AR model with different lag length and error autoregressive order. Exponential decay curve and serial correlation $\phi_1 = 0.5, \phi_2 = 0, \phi_3 = 0, \phi_4 = 0.3, \phi_5 = 0, \phi_6 = 0.2, \phi_7 = 0$ are used to generate simulated data.

estimation of lag coefficients as we can find wider 90% credible interval and bigger RMSE for each lag coefficient.

2.4.4 Competing Methods Specification

BDLagM model full specifications

Welty et al. (2009) proposed to incorporate prior knowledge of the shape of lag curve into DLM through a Bayesian framework. For a normal linear distributed lag model, the Bayesian DLM (BDLagM) has the following hierarchical structure:

$$\begin{aligned}
\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{\beta}, \sigma^2 &\sim N(\boldsymbol{X}^T\boldsymbol{\beta}, \sigma^2\mathbf{I}_n) \\
\boldsymbol{\beta}|\sigma^2_\beta, \eta &\sim N(0, \sigma^2_\beta\boldsymbol{\Omega}(\eta)) \\
\sigma^2_\beta &= 10\text{Var}(\hat{\beta}_0) \\
\boldsymbol{\Omega}(\eta) &= \boldsymbol{V}(\eta_1)\boldsymbol{W}(\eta_2)\boldsymbol{V}(\eta_1)
\end{aligned}$$

where $\hat{\beta}_0$ is the ML estimate of β_0, using the unconstrained distributed lag coefficients, $\boldsymbol{V}(\eta_1) = \text{diag}(1, \exp(\eta_1)^{1/2}, ..., \exp(\eta_1 L)^{1/2})$, $\boldsymbol{W}(\eta_2)$ is the correlation matrix derived from the covariance matrix $\boldsymbol{V}(\eta_2)\boldsymbol{V}(\eta_2)' + \{\mathbf{I}_{L+1} - \mathbf{V}(\eta_2)\}\mathbf{1}_{L+1}\mathbf{1}'_{L+1}\{\mathbf{I}_{L+1} - \mathbf{V}(\eta_2)\}'$, with $\mathbf{1}_{L+1}$ denotes a $(L+1)\times 1$ vectors of ones.

The prior covariance matrix of $\boldsymbol{\beta}$ can also be written component-wisely:

$$\begin{aligned}
\text{var}(\beta_i) &= \sigma^2_\beta \exp(-\eta_1 i) \\
\text{cov}\left(\beta_i, \beta_j\right) &= \frac{\sigma^2_\beta \left\{1 - \exp\left(-\eta_2 i\right)\right\}\left\{1 - \exp\left(-\eta_2 j\right)\right\}\exp\left\{-\eta_1\left(i+j\right)/2\right\}}{\sqrt{\left(\left[\left\{1 - \exp\left(-\eta_2 i\right)\right\}^2 + \exp\left(-2\eta_2 i\right)\right]\left[\left\{1 - \exp\left(-\eta_2 j\right)\right\}^2 + \exp\left(-2\eta_2 j\right)\right]\right)}}
\end{aligned}$$

for $i, j = 0, 1, ...L-1$ and $i \neq j$. η_1 controls the rate at which the variance of $\boldsymbol{\beta}$ approach zero and η_2 controls the rate at which neighbouring coefficients become more correlated.

Bayesian ridge DLM full specifications

Frequentist ridge estimate can be regarded as the posterior mean of regression coefficients with a prior normal distribution with mean zero and an scalar variance-covariance matrix (Hoerl and Kennard, 1970; Lindley and Smith, 1972). Bayesian ridge regression can be written as:

$$(\boldsymbol{Y} \mid \boldsymbol{X}, \tilde{\boldsymbol{\beta}}) \sim N(\boldsymbol{X}\tilde{\boldsymbol{\beta}}, \sigma^2\mathbf{I})$$

$$\tilde{\boldsymbol{\beta}} \sim N(\mathbf{0}, \sigma^2/\lambda_{\text{ridge}}\mathbf{I})$$

$$\lambda_{\text{ridge}} \sim \text{unif}(0, \infty)$$

$$\pi(\sigma^2) \propto 1/\sigma^2$$

where λ_{ridge} is a hyperparameter that controls the amount of ridge penalty.

Non-informative prior Bayesian DLM full specifications

To impose a non-informative prior on the lag coefficients, we used flat improper priors on each parameter. Non-informative prior Bayesian DLM can be considered as the counterpart of distributed lag model in Bayesian framework.

$$(\boldsymbol{Y} \mid \boldsymbol{X}, \tilde{\boldsymbol{\beta}}) \sim N(\boldsymbol{X}\tilde{\boldsymbol{\beta}}, \sigma^2\mathbf{I})$$

$$\pi(\tilde{\boldsymbol{\beta}}) = \pi(\mu, \beta_0, \beta_1, ..., \beta_L) \propto 1$$

$$\pi(\sigma^2) \propto 1/\sigma^2$$

Koyck DLM full specifications

Koyck (1954) proposed an infinite distributed lag model by adding the following constrain on lag coefficients in distributed lag model:

$$\beta_i = \beta_0\lambda^i, \quad \text{and } 0 < \lambda < 1$$

Table 2.9: Summary of evaluation metrics of total effect($\sum_{i=0}^{L}\beta_i$), total carryover effect ($\sum_{i=1}^{L}\beta_i$), lag coefficients, autoregressive coefficient(ϕ) and model standard deviation(σ) under exponential decay lag curve, estimated by BDLM-AR and RegAR models. Treatment sequence $x_t^{(1)}$ and strong treatment effect (SNR = 1) are used to generate simulated data.

		Bias				MSE			
	Truth	BDLM-AR(7)	RegAR(7)	BDLM-AR(1)	RegAR(1)	BDLM-AR(7)	RegAR(7)	BDLM-AR(1)	RegAR(1)
Total effect	9.69	-3.22	-2.73	-3.14	-2.84	26.12	21.09	26.02	26.85
Total carryover effect	4.69	-1.89	-	-1.91	-	10.93	-	10.09	-
β_0	5	-1.34	1.95	-1.23	1.85	6.51	17.43	7.27	22.22
β_1	2.50	-0.96	-	-0.92	-	2.44	-	2.50	-
β_2	1.25	-0.57	-	-0.58	-	1.14	-	1.06	-
β_3	0.62	-0.36	-	-0.40	-	0.45	-	0.34	-
β_4	0.31	-0.16	-	-0.19	-	0.13	-	0.13	-
β_5	0	0.10	-	0.10	-	0.07	-	0.09	-
β_6	0	0.05	-	0.06	-	0.03	-	0.04	-
β_7	0	0.01	-	0.02	-	0.01	-	0.01	-
ϕ_1	0.5	7.83E-04	-9.60E-03	1.20E-01	9.77E-02	1.11E-02	1.14E-02	3.19E-02	2.69E-02
ϕ_2	0	6.15E-05	-9.95E-03	-	-	1.08E-02	1.08E-02	-	-
ϕ_3	0	-1.22E-02	-1.59E-02	-	-	1.12E-02	1.13E-02	-	-
ϕ_4	0.3	-2.28E-02	-3.67E-02	-	-	1.08E-02	1.16E-02	-	-
ϕ_5	0	-1.81E-02	-2.70E-02	-	-	1.06E-02	1.12E-02	-	-
ϕ_6	0.2	2.00E-03	-4.27E-03	-	-	1.32E-02	1.27E-02	-	-
ϕ_7	0	-7.45E-03	-1.50E-02	-	-	8.28E-03	8.61E-03	-	-
σ	10	-0.02	-0.09	0.57	0.49	0.46	0.45	0.95	0.83

In this model, lag coefficients are assumed to decrease geometrically. In other words, recent past explanatory variables are more influential than distant past explanatory variables. λ is the coefficient decaying rate. By replacing β_i into distributed lag model, we can have:

$$Y_t = \mu + \beta_0 \sum_{i=0}^{\infty} \lambda^i X_{t-i} + u_t$$

where u_t are independent Gaussian errors with mean 0 and variance σ^2, β_0 is the immediate effect and the long term effect is defined as:

$$\sum_{k=0}^{\infty} \beta_k = \frac{\beta_0}{1-\lambda}$$

2.5 Application to Light Therapy Study

The data set we used is from I. M. Kronish et al. (2020), which studies the effectiveness of bright white light therapy for depressive symptoms within cancer survivors. Besides bright white intervention (10,000 lux), dim red (50 lux) was used as a control intervention, which lacks sufficient light intensity to affect cells from retina. Patients received light therapy through mobile phone application and were instructed to use one of two portable lightboxes each morning for 30 minutes per day. For each patient, the whole study duration is 12 weeks. One intervention was assigned on the first three weeks and last three weeks and the other intervention was assigned between the fourth week and the ninth week. The initial intervention was randomized, either bright white lightbox or dim red lightbox. The collected outcomes are depressive symptom, fatigue symptom and negative affectivity. The first two are measured by patient's self-reported standard single-item visual analog scale from 0-not at all depressed (tired) to 10-extremely depressed (tired). Other patients' characteristics (age, gender, race, ethnicity, years of education, employment status, health insurance status) were also collected at baseline. Some occasional missing outcomes were imputed using average non-missing value in the corresponding treatment block. Principal component analysis method is used to generate a composite score of collected outcomes. To be specific, depressive

symptom, fatigue symptom and negative affectivity are standardized and principal components are constructed as linear combinations of these three variables. The first principal component which maximize the variance of original data is used as response variable of data analysis. Final composite score is constructed as: Score = $0.65 \times$ depressive $+ 0.57 \times$ fatigue $+ 0.50 \times$ negative affectivity.

As an illustrative example, We present results from two subjects' data (ID:7707 and 7708) with measurement on daily composite score. Subjects 7707 and 7708 started with dim red light intervention and bright white light intervention respectively. We fit the data with the proposed BDLM-AR model. The maximum number of lags L is set to 7, since a week's period is long enough to exhibit the intervention effect on negative symptoms and intervention effect beyond one week can hardly have substantial influence on negative symptoms. Two autoregressive orders of BDLM-AR model AR(1) and AR(7) were used. As a comparison, besides other Bayesian distributed lag models, we also fit a classical linear regression model with autoregressive errors (RegAR) with fixed autoregressive (AR) order 1 and 7. Note that RegAR can be viewed as special case of distributed lag models with L setting to 0.

Table 2.10 shows the posterior mean/MLE of both distributed lag and autoregressive coefficients. For subject 7707, the RegAR and other Bayesian DLM models indicate a strong total effect of bright white intervention in relieving negative symptoms. However, only weak effect of bright white intervention is identified by BDLM-AR model. The distinct conclusion drawn from these two models is potentially triggered by the strong autocorrelation between outcomes. When outcome serial correlation is adjusted, the true intervention effect is much smaller than those estimated from models using white noise. To check the fitness of each model, we used LjungBox test (Ljung and Box, 1978) to examine autocorrelation of the residuals, and the corresponding p-values of χ^2-test are also shown in Table 2.10. No statistically significant autocorrelation is found in residuals of BDLM-AR model. For subject 7708, we observe a similar estimation between different models in terms of total effect. Treatment total effect estimated from BDLM-AR model is -0.64 (90% CI: -1.23, -0.18), which is slightly lower than that from other models. Extra information obtained from BDLM-AR model is that white light intervention effect lasts for around two days.

	Subject ID: 7707						
	BDLM-AR(1)	BDLM-AR(7)	RegAR(1)	RegAR(7)	BDLagM	BR-DLM	NB-DLM
μ	-0.34 (-0.66,-0.02)	-1.07 (-3.19,-0.31)	-0.21 (-0.52,0.10)	-0.19 (-0.80, 0.42)	-0.23 (-0.50,0.03)	-0.22 (-0.49,0.04)	-0.17 (-0.44,0.10)
Total effect	-0.34 (-0.75,0.04)	-0.20 (-0.65, 0.23)	-0.52 (-0.91,-0.12)	-0.53 (-0.97, -0.09)	-0.48 (-0.83,-0.12)	-0.5 (-0.85,-0.14)	-0.59 (-0.95,-0.23)
Total carryover effect	-0.01 (-0.40,0.43)	0.07 (-0.27, 0.56)	-	-	0.21 (-0.74,1.16)	-0.19 (-0.74,0.45)	0.22 (-0.76,1.20)
β_0	-0.33 (-0.85,0.09)	-0.27 (-0.76, 0.14)	-0.99 (-2.30, 0.31)	-0.53 (-0.97, -0.09)	-0.68 (-1.62,0.24)	-0.31 (-0.97,0.17)	-0.81 (-1.78,0.17)
β_1	0 (-0.31,0.36)	-0.02 (-0.32, 0.29)	-	-	0.38 (-0.72,1.48)	0.01 (-0.58,0.67)	0.52 (-0.84,1.88)
β_2	-0.01 (-0.24,0.22)	0.02 (-0.17, 0.24)	-	-	-0.07 (-0.55,0.4)	-0.11 (-0.79,0.48)	-0.69 (-2.05,0.66)
β_3	0.02 (-0.12,0.23)	0.03 (-0.08, 0.23)	-	-	-0.04 (-0.26,0.19)	0.28 (-0.29,1.11)	1.17 (-0.20,2.52)
β_4	0 (-0.12,0.12)	0.02 (-0.06, 0.17)	-	-	-0.03 (-0.16,0.10)	-0.01 (-0.64,0.61)	-0.37 (-1.74,0.98)
β_5	0 (-0.09,0.08)	0.01 (-0.04,0.11)	-	-	-0.02 (-0.09,0.06)	0.05 (-0.54,0.72)	0.35 (-1.01,1.71)
β_6	-0.01 (-0.09,0.04)	0 (-0.04,0.07)	-	-	-0.01 (-0.06,0.04)	-0.13 (-0.76,0.46)	-0.16 (-1.51,1.19)
β_7	-0.01 (-0.07,0.03)	0 (-0.03,0.05)	-	-	-0.01 (-0.03,0.02)	-0.28 (-0.90,0.20)	-0.58 (-1.56,0.39)
ϕ_1	0.22 (0.02,0.42)	0.06 (-0.14,0.27)	0.22 (0.02,0.41)	0.14 (-0.04, 0.32)	-	-	-
ϕ_2	-	0.02 (-0.20,0.24)	-	0.02 (-0.16, 0.20)	-	-	-
ϕ_3	-	-0.10 (-0.30,0.10)	-	-0.06 (-0.24, 0.12)	-	-	-
ϕ_4	-	0.16 (-0.04,0.35)	-	0.21 (0.04, 0.39)	-	-	-
ϕ_5	-	0.16 (-0.06,0.36)	-	0.17 (-0.02, 0.35)	-	-	-
ϕ_6	-	0.01 (-0.17,0.20)	-	0.01 (-0.18, 0.20)	-	-	-
ϕ_7	-	0.32 (0.11,0.52)	-	0.31 (0.12, 0.51)	-	-	-
p-value	0.55	0.76	0.10	0.85	0.12	0.15	0.09
	Subject ID: 7708						
	BDLM-AR(1)	BDLM-AR(7)	RegAR(1)	RegAR(7)	BDLagM	BR-DLM	NB-DLM
μ	-0.30 (-0.60,-0.01)	-0.16 (-1.50,1.83)	-0.23 (-0.51, 0.05)	-0.21 (-0.50, 0.07)	-0.25 (-0.47,-0.03)	-0.29 (-0.51,-0.07)	-0.26 (-0.48,-0.03)
Total effect	-0.64 (-1.08,-0.19)	-0.64 (-1.23, -0.18)	-0.84 (-1.24, -0.43)	-0.85 (-1.27, -0.43)	-0.76 (-1.11,-0.41)	-0.68 (-1.05,-0.31)	-0.74 (-1.12,-0.36)
Total carryover effect	-0.11 (-0.56,0.32)	-0.18 (-0.65, 0.18)	-	-	0.44 (-0.36,1.25)	0.02 (-0.61,0.74)	0.52 (-0.32,1.36)
β_0	-0.53 (-1.04,-0.06)	-0.46 (-0.94, -0.04)	-0.84 (-1.24, -0.43)	-0.85 (-1.27, -0.43)	-1.21 (-2.00,-0.42)	-0.69 (-1.43,-0.09)	-1.26 (-2.09,-0.42)
β_1	-0.07 (-0.41,0.28)	-0.15 (-0.50, 0.14)	-	-	0.43 (-0.50,1.35)	0.04 (-0.60,0.81)	0.48 (-0.67,1.64)
β_2	-0.02 (-0.25,0.22)	-0.03 (-0.27, 0.16)	-	-	0.02 (-0.38,0.42)	0.07 (-0.57,0.81)	0.22 (-0.95,1.38)
β_3	-0.01 (-0.19,0.14)	-0.02 (-0.19, 0.12)	-	-	0 (-0.20,0.19)	-0.01 (-0.67,0.69)	0.21 (-0.95,1.37)
β_4	-0.02 (-0.17,0.08)	-0.02 (-0.16, 0.08)	-	-	0 (-0.11,0.11)	-0.47 (-1.35,0.19)	-1.24 (-2.40,-0.09)
β_5	0 (-0.07,0.10)	0.02 (-0.05, 0.14)	-	-	0 (-0.07,0.07)	0.24 (-0.40,1.03)	0.75 (-0.41,1.91)
β_6	0 (-0.05,0.07)	0.01 (-0.03, 0.11)	-	-	0 (-0.04,0.04)	0.17 (-0.47,0.88)	0.28 (-0.87,1.44)
β_7	0 (-0.04,0.04)	0 (-0.03, 0.05)	-	-	0 (-0.02,0.03)	-0.03 (-0.60,0.51)	-0.18 (-1.02,0.66)
ϕ_1	0.27 (0.09,0.46)	0.28 (0.09, 0.49)	0.27 (0.08, 0.45)	0.25 (0.07, 0.44)	-	-	-
ϕ_2	-	0.06 (-0.17, 0.25)	-	0.02 (-0.17, 0.21)	-	-	-
ϕ_3	-	-0.18 (-0.36, 0.06)	-	-0.18 (-0.38, 0.02)	-	-	-
ϕ_4	-	0.15 (-0.07, 0.37)	-	0.09 (-0.11, 0.29)	-	-	-
ϕ_5	-	0.07 (-0.17, 0.25)	-	0.05 (-0.15, 0.25)	-	-	-
ϕ_6	-	-0.06 (-0.25, 0.21)	-	-0.02 (-0.24, 0.19)	-	-	-
ϕ_7	-	0.16 (-0.09, 0.36)	-	0.08 (-0.13, 0.28)	-	-	-
p-value	0.77	0.75	0.03	0.99	0.03	0.02	0.01

Table 2.10: Posterior mean/MLE of distributed lag and autoregressive coefficients for light therapy study. 90% credible intervals/confidence intervals are in brackets. P-value of Ljung-Box test for each models is on the last row.

For RegAR(1) model, BDLagM, BR-DLM and NB-DLM, statistically significant autocorrelation is found in residuals, indicating an inadequacy of model fitting.

2.6 Discussion

In this paper, we provide a novel method to analyze data from N-of-1 trials. The method handles temporal correlation between measurements and carryover effects via distributed lag structure and parameters are estimated using Bayesian approach with (fused) ridge type regularization. From the design perspective, N-of-1 trial can be viewed as a special crossover trial for one person. Tra-

ditional crossover trial requires physical washout period to prevent carryover effect. Our proposed method directly models the carryover effect, thus shortens the entire duration of trial.

To alleviate multicollinearity issue in explanatory variables and incorporate prior knowledge on the shape of lag curve, we designed a prior precision matrix on lag coefficient and showed its MAP estimate is equivalent to fused ridge regression in Appendix A.1. The shrinkage and smoothness penalty terms are determined in a data-driven manner, which provides a more systematic way to choose regularization tuning parameters in time series data analysis. In Section 2.2.2, we let hyperparameters follow exponential function, but any function that decay to zero will also work. In practice, if available, population level pharmacokinetics information can be utilized to select the functional form of hyperparameters.

We adopt a Bayesian estimation procedure, while demonstrating a connection to a fused ridge penalized estimation procedure (2.7). Cross validation is often a method of choice in choosing the penalties (λ_i and λ_i^* via γ_1 and γ_2). However, in our application, it is not feasible to split the sample at random time points because of the temporal order. Therefore, Bayesian formulation provides a natural way to "estimate" the penalties.

Through simulations, we showed that our proposed BDLM-AR model outperforms Koyck DLM, BDLagM, BR-DLM and NB-DLM in estimating total effect, total carryover effect and single lag coefficients in most scenarios. Furthermore, we showed that BDLM-AR can simultaneously estimate autoregressive error. The advantage of BDLM-AR model increases when strong serial correlation exists or signal to noise ratio is small.

We have applied our model to a real application of using white light therapy for depressive symptoms. For subject 7707, we found white light therapy nearly has no effect on resolving depressive symptoms, which is different from the results in the original literature. For subject 7708, our findings are consistent with the literature. However, the proposed BDLM-AR model can provide more information on patient's whole time course of treatment effect.

Chapter 3: Bayesian Hierarchical Distributed Lag Model with Autocorrelated Errors for Combining and Evaluating Multiple N-of-1 Trials

3.1 Introduction

N-of-1 trials are multiple-period crossover trials conducted within single subject comparing two or more interventions (Duan et al., 2013; G. Guyatt et al., 1986). Each subject will receive treatment assignment in randomized sequence and same treatment will be administrated in a period. There is an increasing interest in N-of-1 trials since it is compatible with the goal of patient-centered outcomes research (Selby et al., 2012) and comparative effectiveness research (Sox and Greenfield, 2009). A lot of conditions that N-of-1 trials have been utilized include hypertension (Chatellier et al., 1995), depression (I. M. Kronish et al., 2018), and chronic airflow limitation (Mahon et al., 1999; Patel et al., 1991).

When a series of N-of-1 trials are conducted within similar subjects, besides treatment effect at individual level, it is also possible to pursue estimates of treatment effect at population level. Due to the existence of heterogeneity of treatment effects (Kent et al., 2010), any statistical methods that aim to estimate treatment effect for whole subjects in the study is expected to take the variation in individual treatment effect into consideration. Typical methods that address multiple-subject

data in a model-base approach that analytically incorporate heterogeneity are mixed effect model (Berkey et al., 1995; A. P. Jones et al., 2009; D. R. Zucker et al., 2010) and Bayesian hierarchical model (D. R. Zucker et al., 2010; D. Zucker et al., 1997). The benefit of multilevel modeling structure is that between and within patient variation can be estimated separately.

Like traditional cross-over type of clinical trials, in most N-of-1 trials, a paired unit randomization for paired periods is usually used in study design. Data are collected or summarized at period level, which means one data point will represent the treatment effect in each period and the only way to increase sample size within each trial is to increase the number of periods and times of switching treatment in the trial. But additional treatment switches may experience practical difficulties. Another problem in analyzing N-of-1 trials is the potential existence of carryover effect, which may compromise the validity of N-of-1 trials (Duan et al., 2013). The usual statistical models only allows for first-order carryover effect, that is the carryover effect lasting no more than one period after the end of previous treatment period. Take a two treatments with two periods crossover trial as an example, if significant carryover is detected, only data from the first period will be used. S. Senn (1992) points out that using designs and models adjusting for first-order carryover effect can hardly give protection against real carryover effects.

Distributed lag models (DLM) postulates that the current value of the outcome variable depends on the previous values (lags) of an exposure as well as the current exposure value, thus allowing the total exposure effect to be distributed over a time period and facilitating explicit modeling of carryover effects. It has wide applications in in economics (Almon, 1965; Koyck, 1954), advertising (Bass and Clarke, 1972) and environment health studies (Peng et al., 2009; Welty et al., 2009; Zanobetti et al., 2000). Chapter 2 proposed a Bayesian framework (BDLM-AR) to extend DLMs for N-of-1 trial data that explicitly model the carryover effect and correlation within subject.

In this chapter, we further extend the BDLM-AR model to Bayesian hierarchical model to leverage the individual N-of-1 trials data to obtain population-level treatment effect estimation. Additionally, Bayesian hierarchical model can improve the accuracy and precision of individual-level parameters estimates since the posterior distribution at population level parameters can be

considered as prior distribution of individual-level parameters, and these priors are themselves generated by the data (Gelman et al., 2013).

This chapter is organized as follows. In Section 3.2, we will formally introduce the proposed Bayesian hierarchical distributed lag model with autocorrelated errors (BHDLM-AR). We then show the posterior computations in Section 3.3. Next, in Section 3.4, we will perform numerical simulations to evaluate the performance of BHDLM-AR in estimating both population and individual level parameters and compare with other methods. In Section 3.5, the proposed model is applied to a light therapy study and will conclude this chapter with a discussion in Section 3.6.

3.2 Methods

3.2.1 Proposed Model

Suppose that we obtain data from a total of S single N-of-1 trials, each with n_i observations, where $i = 1, 2, ..., S$. For individual N-of-1 trial i, let $Y_{i,t}$ denote a response variable of interest at time t of subject i and $X_{i,t}$ denote a time-varying explanatory binary variable, which has some influence on $Y_{i,t}$ up to some pre-determined maximum number of lags L. The distributed lag model can be described as follows:

$$Y_{i,t} = \mu_i + \sum_{l=0}^{L} \beta_{i,l} X_{i,t-l} + \epsilon_{i,t} \tag{3.1}$$

where the error term $\epsilon_{i,t}$ follows an autoregressive process,

$$\epsilon_{i,t} = \phi_{i,1}\epsilon_{i,t-1} + \phi_{i,2}\epsilon_{i,t-2} + ... + \phi_{i,p}\epsilon_{i,t-p} + w_{i,t} \tag{3.2}$$

$w_{i,t}$ is a white Gaussian noise with mean zero and variance $\sigma^2 > 0$. Note that when $t < L$, those terms with negative subscript will not be included in the model.

The advantage of using above distributed lag model (3.1) to analyze N-of-1 trials data is (1). The mean model specified by $\boldsymbol{\beta_i} = (\beta_{i,0}, ..., \beta_{i,L})'$ reflects the overall time course of the effect that explanatory variable $X_{i,t}$ have on outcome $Y_{i,t}$. To be specific, each single coefficient $\beta_{i,l}$ represents

treatment effect on the previous L-th day, $\beta_{i,0}$ represents current treatment effect and total carryover effect δ_i can be defined as:

$$\delta_i \triangleq \sum_{l=1}^{L} \beta_{i,l} = E(Y_{i,t}|X_{i,t-1} = 1, ...X_{i,t-L} = 1, X_{i,t}) - E(Y_{i,t}|X_{i,t-1} = 0, ..., X_{i,t-L} = 0, X_{i,t}).$$

$\sum_{l=0}^{L} \beta_{i,l}$ represents the total treatment effect in the i-th N-of-1 trial, which is the summation of current effect and total carryover effect. (2). The temporal dependency between errors is modeled by an order-p autoregressive model with autoregressive coefficient $\boldsymbol{\phi_i} = (\phi_{i,1}, ..., \phi_{i,p})'$.

Let $\Phi_i(B) = 1 - \phi_{i,1}B - \phi_{i,2}B^2 - ... - \phi_{i,p}B^p$, where B is the backshift operator. Similarly as Chapter 2, utilizing the relationship that $\Phi_i(B)\epsilon_{i,t} = w_{i,t}$, we can apply $\Phi_i(B)$ to both sides of model (3.1) and rewrite the model,

$$Y^*_{i,t} = \mu^* + \sum_{l=0}^{L} \beta_{i,l} X^*_{i,t-l} + w_{i,t}, \tag{3.3}$$

for $t = p + 1, ..., n_i$, where $\mu^*_i = \Phi(B)\mu_i$, $Y^*_{i,t} = \Phi_i(B)Y_{i,t}$, $X^*_{i,t} = \Phi_i(B)X_{i,t}$ and $\mu^*_i = \Phi_i(B)\mu_i$. To stack the data in vector form, we have

$$\boldsymbol{Y}^*_i \sim N(\mu^*_i \mathbf{1}_{n_i-p} + \boldsymbol{X}^*_i \boldsymbol{\beta_i}, \sigma^2 \mathbf{I}_{n_i-p}) \tag{3.4}$$

We denote $\tilde{\boldsymbol{\beta}}_i = (\mu_i, \boldsymbol{\beta}'_i)'$ and $\tilde{\boldsymbol{X}}^*_i = (\Phi(B)\mathbf{1}_{n-p}, \boldsymbol{X}^*_i)$, so that $\tilde{\boldsymbol{X}}^*_i \tilde{\boldsymbol{\beta}}_i = \mu^*_i \mathbf{1}_{n-p} + \boldsymbol{X}^*_i \boldsymbol{\beta}_i$.

3.2.2 Prior Distribution on the Mean Model

To combine multiple N-of-1 trials and simultaneously estimating study specific and overall population level treatment effect, we regard the distributed lag coefficient $\boldsymbol{\beta}_i$ in each N-of-1 trial are related and connected and consider the following Bayesian hierarchical structure,

$$\begin{aligned} \tilde{\boldsymbol{\beta}}_{\boldsymbol{i}} &\sim N(\tilde{\boldsymbol{\theta}}, \sigma^2 \tilde{\boldsymbol{\Omega}}^{-1}_{\beta}(\gamma)) \\ \tilde{\boldsymbol{\theta}} &\sim N(\mathbf{0}, \sigma^2_{\theta} \tilde{\boldsymbol{\Omega}}^{-1}_{\theta}(\tau)) \end{aligned} \tag{3.5}$$

for $i = 1, 2, ..., S$, where $\tilde{\boldsymbol{\Omega}}_\beta(\gamma) = diag(c_0, \boldsymbol{\Omega}_\beta(\gamma))$ so that the prior variance of μ_i is $\sigma^2 c_0^{-1}$ and the prior variance-covariance matrix of $\boldsymbol{\beta}_i$ is $\sigma^2 \tilde{\boldsymbol{\Omega}}_\beta^{-1}(\gamma)$. $\tilde{\boldsymbol{\theta}} = (\mu_\theta, \boldsymbol{\theta}) = (\mu_\theta, \theta_0, ..., \theta_L)$ is a hyperparameter vector with length $L + 2$, representing the population distributed lag coefficients. Similarly as single distributed lag model, each coefficient θ_l represents the treatment effect on l days ago, but at population level. $\boldsymbol{\Omega}_\beta(\gamma)$ is constructed in the following form:

$$\begin{pmatrix} \lambda_0 + \lambda_0^* & -\lambda_0^* & 0 & \dots & \dots & 0 \\ -\lambda_0^* & \lambda_1 + \lambda_0^* + \lambda_1^* & -\lambda_1^* & \dots & \dots & 0 \\ 0 & -\lambda_1^* & \lambda_2 + \lambda_1^* + \lambda_2^* & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \lambda_{L-1} + \lambda_{L-2}^* + \lambda_{L-1}^* & -\lambda_{L-1}^* \\ 0 & 0 & 0 & \dots & -\lambda_{L-1}^* & \lambda_L + \lambda_{L-1}^* + \lambda_L^* \end{pmatrix}, \tag{3.6}$$

where the hyperparameters $\lambda_l = \exp\{\gamma_1(l+1)\} - 1$ and $\lambda_l^* = \exp\{\gamma_2(l+1)\} - 1$. The function form of (λ_l, λ_l^*) is not unique. Any other monotone increasing function of γ_1 and γ_2 will work. As the lag index increases, $\boldsymbol{\beta_i}$ will be shrunk towards the population mean $\boldsymbol{\theta}$. As mentioned in Chapter 2, an alternative view of $\tilde{\boldsymbol{\Omega}}_\beta$ is a fused ridge-type penalty, the maximum *a posteriori* probability estimate of $\tilde{\boldsymbol{\beta}}$ minimizes a fused ridge-type penalty:

$$(\boldsymbol{Y}^* - \tilde{\boldsymbol{X}}^* \tilde{\boldsymbol{\beta}})^T (\boldsymbol{Y}^* - \tilde{\boldsymbol{X}}^* \tilde{\boldsymbol{\beta}}) + c_0 \mu^2 + \sum_{l=0}^{L} \lambda_l (\beta_l - \theta_l)^2 + \sum_{l=0}^{L} \lambda_l^* (\beta_l - \beta_{l+1})^2 \tag{3.7}$$

where $\beta_{L+1} \triangleq 0$, λ_l is an increasing ℓ_2 penalty of the individual level coefficients towards population level coefficients, and λ_l^* is an increasing smoothness penalty on the difference of adjacent individual coefficients.

To choose the amount of fused ridge-type penalty at individual level, we consider a truncated standard exponential hyperprior on (γ_1, γ_2) of the form

$$\pi(\gamma_1, \gamma_2) \propto \exp(-\gamma_1 - \gamma_2) \mathbf{1}_{S_\gamma}(\gamma_1, \gamma_2) \tag{3.8}$$

where the support $S_{\boldsymbol{\gamma}}$ includes all pairs (γ_1, γ_2) with which the precision matrix $\boldsymbol{\Omega}$ is positive definite.

$\tilde{\boldsymbol{\Omega}}_\theta(\tau)$ has the same form as $\tilde{\boldsymbol{\Omega}}_\beta(\boldsymbol{\gamma})$ but is parameterized by the vector $\tau = (\tau_1, \tau_2)$. To reduce the size of model space, the parameters in the hyperprior is set as fixed value. From the experience of previous works, $\tau = (\tau_1, \tau_2)$ are both set to 1, and σ_θ is set to be a thousand times of the average model variance estimated from single N-of-1 trials.

3.2.3 Prior Distribution on the Error Model

We use the Jeffreys prior for the error variance σ^2 and for the autoregressive process, a normal prior for $\boldsymbol{\phi}_i$ is used, with the constraint that the error process is stationary.

$$\begin{aligned} \pi(\sigma^2) &\propto 1/\sigma^2 \\ \boldsymbol{\phi}_i &\sim N_p\left(\mathbf{0}_p, 200 \times \mathbf{I}_p\right) \mathbf{1}_{S_{\boldsymbol{\phi}}}(\boldsymbol{\phi}_i) \end{aligned} \tag{3.9}$$

for i = 1,2,...,S, where $S_{\boldsymbol{\phi}}(\boldsymbol{\phi}_i)$ denotes the support where all roots of the polynomial $\Phi(B) = 1 - \sum_{l=1}^{p} \phi_l B^l$ are outside the unit circle. Following Chib (1993), we can show that the process $\{\epsilon_t : t = 1, 2, \ldots\}$ is stationary when $\boldsymbol{\phi}_i \in S_{\boldsymbol{\phi}}(\boldsymbol{\phi}_i)$. Note that the range of each element in ϕ is $(-1, 1)$; thus, a prior variance of 200 in (3.9) essentially amounts to a flat prior.

3.3 Conditional Posterior Distributions

In this part, we show the full conditional distributions of all parameters discussed in the previous parts. Details of derivation of full conditional distributions for each parameter can be found in Appendix B.1. For notational convenience, let $\boldsymbol{Y} = (\boldsymbol{Y}_1, ..., \boldsymbol{Y}_S)$, $\tilde{\boldsymbol{X}} = (\tilde{\boldsymbol{X}}_1, ..., \tilde{\boldsymbol{X}}_S)$, $\tilde{\boldsymbol{B}} = (\tilde{\boldsymbol{\beta}}_1, ..., \tilde{\boldsymbol{\beta}}_S)$, and $\boldsymbol{\Phi} = (\boldsymbol{\phi}_1, ..., \boldsymbol{\phi}_S)$, for $i = 1, 2, ..., S$.

Given the likelihood function of transformed data (3.4) and prior distribution of $\boldsymbol{\beta}_i$, for $i =$

$1, 2, ..., S$, the full conditional distribution of distributed lag coefficient in each individual trial $\tilde{\boldsymbol{\beta}}_i$ is

$$\tilde{\boldsymbol{\beta}}_i \mid \boldsymbol{Y}, \tilde{\boldsymbol{X}}, \sigma^2, \boldsymbol{\Phi}, \boldsymbol{\gamma}, \tilde{\boldsymbol{\theta}} \sim N\left\{[\tilde{\boldsymbol{X}}_i^{*'}\tilde{\boldsymbol{X}}_i^{*} + \tilde{\boldsymbol{\Omega}}_\beta(\boldsymbol{\gamma})]^{-1}\left[\boldsymbol{X}_i^{*'}\boldsymbol{Y}_i^{*} + \boldsymbol{\theta}\tilde{\boldsymbol{\Omega}}_\beta(\boldsymbol{\gamma})\right], \sigma^2[\tilde{\boldsymbol{X}}_i^{*'}\tilde{\boldsymbol{X}}_i^{*} + \tilde{\boldsymbol{\Omega}}_\beta(\boldsymbol{\gamma})]^{-1}\right\} \tag{3.10}$$

The posterior distribution of hyperparameter $\boldsymbol{\theta}$ can be derived from likelihood function of transformed data (3.4), prior distribution of each individual lag coefficient $\boldsymbol{\beta}_i$ and hyperprior on $\boldsymbol{\theta}$ (3.5),

$$\tilde{\boldsymbol{\theta}} \mid \boldsymbol{Y}, \tilde{\boldsymbol{X}}, \sigma^2, \tilde{\boldsymbol{B}}, \boldsymbol{\Phi}, \boldsymbol{\gamma} \sim N\left\{\left[\frac{\tilde{\boldsymbol{\Omega}}_\theta(\tau)}{\sigma_\theta^2} + \frac{\tilde{\boldsymbol{\Omega}}_\beta(\boldsymbol{\gamma})}{\sigma^2/S}\right]^{-1}\frac{\tilde{\boldsymbol{\Omega}}_\beta(\boldsymbol{\gamma})}{\sigma^2}\sum_{i=1}^{S}\boldsymbol{\beta}_i, \left[\frac{\tilde{\boldsymbol{\Omega}}_\theta(\tau)}{\sigma_\theta^2} + \frac{\tilde{\boldsymbol{\Omega}}_\beta(\boldsymbol{\gamma})}{\sigma^2/S}\right]^{-1}\right\} \tag{3.11}$$

Working with model (3.2) and (3.9), the full condition distribution of σ^2 is an inverse-gamma distribution,

$$\sigma^2 \mid \boldsymbol{Y}, \tilde{\boldsymbol{X}}, \tilde{\boldsymbol{B}}, \tilde{\boldsymbol{\theta}}, \boldsymbol{\Phi}, \boldsymbol{\gamma} \sim \text{IG}\left\{\frac{\sum_{i=1}^{S} n_i - S(p-L-1)}{2}, \frac{\sum_{i=1}^{S}\left[(\tilde{\boldsymbol{\beta}}_i - \tilde{\boldsymbol{\theta}})'\tilde{\boldsymbol{\Omega}}_\beta(\boldsymbol{\gamma})(\tilde{\boldsymbol{\beta}}_i - \tilde{\boldsymbol{\theta}}) + (\boldsymbol{Y}_i^{*} - \tilde{\boldsymbol{X}}_i^{*}\tilde{\boldsymbol{\beta}}_i)'(\boldsymbol{Y}_i^{*} - \tilde{\boldsymbol{X}}_i^{*}\tilde{\boldsymbol{\beta}}_i)\right]}{2}\right\} \tag{3.12}$$

and $\boldsymbol{\phi}_i$ has a truncated multivariate normal conditional posterior:

$$\boldsymbol{\phi}_i \mid \boldsymbol{Y}, \tilde{\boldsymbol{X}}, \tilde{\boldsymbol{B}}, \tilde{\boldsymbol{\theta}}, \sigma^2, \boldsymbol{\gamma} \sim N\left[\left(\sigma^{-2}\boldsymbol{E}_i^{*'}\boldsymbol{E}_i + \sigma_{\boldsymbol{\phi}}^{-2}\mathbf{I}\right)^{-1}\sigma^{-2}\boldsymbol{E}_i^{*'}\boldsymbol{\epsilon}_i^{*}, \left(\sigma^{-2}\boldsymbol{E}_i^{*'}\boldsymbol{E}_i + \sigma_{\boldsymbol{\phi}}^{-2}\mathbf{I}\right)^{-1}\right]\mathbf{1}_{S_\phi}(\boldsymbol{\phi}_i) \tag{3.13}$$

where $\boldsymbol{\epsilon}_i^{*} = (\epsilon_{i,p+1}^{*}, \ldots, \epsilon_{i,n}^{*})'$, $\epsilon_{i,t}^{*} = Y_{i,t} - \mu_i - \sum_{l=0}^{L}\beta_{i,l}X_{i,t-l}$, and $\boldsymbol{E}_i^{*}$ is a $(n-p)\times p$ matrix with $\epsilon_{i,p+k-j}^{*}$ being the (k,j)-th element.

Using the likelihood (3.4) and prior of $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{B}}$, the full conditional posterior distribution can not be expressed explicitly, but the density function is proportional to

$$\pi(\boldsymbol{\gamma} \mid \boldsymbol{Y}, \tilde{\boldsymbol{X}}, \sigma^2, \tilde{\boldsymbol{B}}, \tilde{\boldsymbol{\theta}}, \boldsymbol{\Phi}) \propto \exp(-\gamma_1-\gamma_2)\mathbf{1}_{S_\gamma}(\gamma_1,\gamma_2)\prod_{i=1}^{S}|\sigma^{-2}\tilde{\boldsymbol{\Omega}}_\beta(\boldsymbol{\gamma})|^{\frac{1}{2}}\exp\left[-\frac{1}{2\sigma^2}(\tilde{\boldsymbol{\beta}}_i - \tilde{\boldsymbol{\theta}})'\tilde{\boldsymbol{\Omega}}_\beta(\boldsymbol{\gamma})(\tilde{\boldsymbol{\beta}}_i - \tilde{\boldsymbol{\theta}})\right] \tag{3.14}$$

Based on the model specifies in section (3.2), as long as the conditional posterior distributions

of the parameters, we proposed a hybrid Metropolis-Hastings (MH)/Gibbs sampler for posterior sampling. The Gibbs sampling steps will update $\tilde{\boldsymbol{B}}$, $\tilde{\boldsymbol{\theta}}$, σ^2 and $\boldsymbol{\Phi}$, whose full conditional distributions are available. And MH algorithm will update $\boldsymbol{\gamma}$ with a uniform $U(-a, a)$ proposal distribution. That is, $\gamma_{i,new} = \gamma_i + U(-a, a)$, where the tuning parameter a is chosen such that the acceptance rate of proposed sample is around 50% (Gelman et al., 1996).

Having the full conditional distribution of all parameters in BHDLM-AR, we use a hybrid Metropolis-Hastings/Gibbs algorithm to generate samples of $(\tilde{\boldsymbol{B}}, \tilde{\boldsymbol{\theta}}, \boldsymbol{\Phi}, \sigma^2, \boldsymbol{\gamma})$ from the posterior distribution.

Algorithm 2: The hybrid Metropolis-Hastings/Gibbs algorithm for Bayesian Hierarchical Distributed Lag Model with Autocorrelated Error

---

Step 1. Set initial values for $\tilde{\boldsymbol{B}}$, $\tilde{\boldsymbol{\theta}}$, σ^2, $\boldsymbol{\Phi}$ and $\boldsymbol{\gamma}$;

for $j \leftarrow 1$ **to** $n_{\text{iteration}}$ **do**

  for $i \leftarrow 1$ *to* S **do**

    Step 2. Given current value of $\boldsymbol{\phi}_i$, transform $\boldsymbol{Y}_i$, $\tilde{\boldsymbol{X}}_i$ to $\boldsymbol{Y}_i^*$, $\tilde{\boldsymbol{X}}_i^*$ as described in equation (3.4); Also construct precision matrix $\tilde{\boldsymbol{\Omega}}_\beta(\boldsymbol{\gamma})$ based on $\boldsymbol{\gamma}$ as described in Section 3.2.2;

    Step 3. Conditional on current values of $\boldsymbol{Y}_i^*$, $\tilde{\boldsymbol{X}}_i^*$, $\tilde{\boldsymbol{\theta}}$, σ^2 and $\tilde{\boldsymbol{\Omega}}_\beta(\boldsymbol{\gamma})$, update $\tilde{\boldsymbol{\beta}}_i$ based on $\pi(\tilde{\boldsymbol{\beta}}_i \mid \boldsymbol{Y}^*, \tilde{\boldsymbol{X}}^*, \tilde{\boldsymbol{\theta}}, \sigma^2, \boldsymbol{\Phi}, \boldsymbol{\gamma})$;

  end

  Step 4. Conditional on current values of $\boldsymbol{Y}^*$, $\tilde{\boldsymbol{X}}^*$, $\tilde{\boldsymbol{B}}$, σ^2, $\boldsymbol{\Phi}$ and $\tilde{\boldsymbol{\Omega}}_\beta(\boldsymbol{\gamma})$, update $\tilde{\boldsymbol{\theta}}$ based on $\pi(\tilde{\boldsymbol{\theta}} \mid \boldsymbol{Y}^*, \tilde{\boldsymbol{X}}^*, \tilde{\boldsymbol{B}}, \sigma^2, \boldsymbol{\Phi}, \boldsymbol{\gamma})$;

  Step 5. Conditional on current values of $\boldsymbol{Y}^*$, $\tilde{\boldsymbol{X}}^*$, $\tilde{\boldsymbol{B}}$, $\tilde{\boldsymbol{\theta}}$, $\boldsymbol{\Phi}$ and $\tilde{\boldsymbol{\Omega}}_\beta(\boldsymbol{\gamma})$, update σ^2 based on $\pi(\sigma^2 \mid \boldsymbol{Y}^*, \tilde{\boldsymbol{X}}^*, \tilde{\boldsymbol{B}}, \tilde{\boldsymbol{\theta}}, \boldsymbol{\Phi}, \boldsymbol{\gamma})$;

Step 6. for $i \leftarrow 1$ *to* S **do**

Update $\boldsymbol{\epsilon}_i^*$ conditional on current value of $\tilde{\boldsymbol{\beta}}_i$ and $\boldsymbol{Y}_i$, $\tilde{\boldsymbol{X}}_i$. Then update $\boldsymbol{\phi}_i$ based on $\pi(\boldsymbol{\phi}_i \mid \boldsymbol{Y}, \tilde{\boldsymbol{X}}, \tilde{\boldsymbol{B}}, \tilde{\boldsymbol{\theta}}, \sigma^2, \boldsymbol{\gamma})$. Reject samples if the roots of $\boldsymbol{\phi}_i(L)$ lie outside the unit circle;

end

Step 7. Update (γ_1, γ_2) based on $\pi(\boldsymbol{\gamma} \mid \tilde{\boldsymbol{B}}, \tilde{\boldsymbol{\theta}}, \sigma^2)$. Sample a proposal γ_i^* by $\gamma_i^* = \gamma_i + a * U(-1, 1)$ for $i = 1, 2$. a is an adjustable step size. Compute the ratio

$$R_\gamma = \frac{\pi(\boldsymbol{\gamma}^* \mid \tilde{\boldsymbol{B}}, \tilde{\boldsymbol{\theta}}, \sigma^2)}{\pi(\boldsymbol{\gamma} \mid \tilde{\boldsymbol{B}}, \tilde{\boldsymbol{\theta}}, \sigma^2)}$$

if $\tilde{\boldsymbol{\Omega}}_\beta(\boldsymbol{\gamma}^*)$ *is positive definite* **then**

update $\boldsymbol{\gamma} = \boldsymbol{\gamma}^*$ with probability $\min(1, R_\gamma)$;

end

3.4 Simulation Study

3.4.1 Simulation Scenarios and Data Generation

In this section, we present a simulation study to compare the proposed BHDLM-AR model with other methods for combining N-of-1 trials. Multiple individual N-of-1 trials are regarded as a simulation group trial. For each individual trial, we generate 84 observations which represent 12 weeks, a reasonable length for N-of-1 trial in real practice. Two numbers of individual trials were used: $S = 10, 20$. For each simulation group trial, we generate data under the model:

$$\begin{aligned} \tilde{\boldsymbol{\beta}}_i &\sim N(\tilde{\boldsymbol{\theta}}, \sigma^2 \tilde{\boldsymbol{\Omega}}_\beta^{-1}(\gamma_0)) \\ Y_{i,t} &= \mu_i + \sum_{l=0}^{7} \beta_{i,l} X_{i,t-l} + \epsilon_{i,t} \end{aligned} \tag{3.15}$$

for $i = 1, 2, ...S$, where $\epsilon_{i,t} = \phi_i \epsilon_{i,t-1} + w_{i,t}$ and $w_{i,t} \sim N(0, \sigma^2)$. $\tilde{\boldsymbol{\Omega}}_{\beta}^{-1}(\gamma_0)$ is defined as equation (3.6), but γ_0 is given and set to both be 1. Five sets of population level lag coefficients (lag curves) $\boldsymbol{\theta}$ were used, reflecting different patterns of carryover treatment effects:

1. Exponential decay curve: $\boldsymbol{\theta} = (5, 2.5, 1.25, 0.625, 0.3125, 0, 0, 0)^T$;
2. Exponential decay curve with oscillation: $\boldsymbol{\theta} = (5, 2.5, -1.25, -0.625, 0.3125, 0, 0, 0)^T$;
3. Slow absorption curve: $\boldsymbol{\theta} = (1.51, 2.75, 3.36, 2.03, 0.34, 0, 0, 0)^T$;
4. Slow absorption curve with oscillation: $\boldsymbol{\theta} = (1.51, 2.75, -3.36, -2.03, 0.34, 0, 0, 0)^T$;
5. No carryover effect: $\boldsymbol{\theta} = (10, 0, 0, 0, 0, 0, 0, 0)^T$.

The exponential decay curves (1) and (2) specify coefficients that diminish in magnitude as lag lengthens. The slow absorption curves (3) and (4) reflect scenarios where there is a delay for the carryover effects to peak. The last scenario (5) is the null case where there is no carryover effect. The total effect in each scenario is 10 and total carryover effects (δ) are 4.69, 0.94, 8.48, -2.30 and 0 respectively.

In each individual N-of-1 trial, participants would receive same treatment sequence: $x_t = 1$ on the first 30 days and the last 30 days, and receive $x_t = 0$ between days 31 and 90; that is,

$$x_t = \begin{cases} 1 & t = 30s + 1, ..., 30s + 30 \text{ for } s = 0, 3 \\ 0 & t = 30s + 1, ..., 30s + 30 \text{ for } s = 1, 2. \end{cases}$$

For the stochastic component in data generation, we set $\sigma = 10$ and consider $\phi = 0.5, 0.2$, representing strong and weak serial correlation respectively.

3.4.2 Comparison Methods

For each simulation scenario, 100 data sets were generated. We fit the proposed BHDLM-AR model with lag $L = 7$ and AR(1) for the error terms. Posterior distributions were derived using

the hybrid Metropolis Hastings/Gibbs algorithm described in the previous section with 50,000 iterations, a burn-in period of 25,000, and $a = 0.2$ for sampling $\boldsymbol{\gamma}$ in the MH step.

We compare estimates obtained using BDLM-AR in Chapter 2, linear mixed effect model with first-order autoregressive structure covariance matrix for observations within subjects (LMM-AR; D. R. Zucker et al., 2010) and an extension of regular distributed lag model with random effects, following the same first-order autoregressive covariance structure (DLMM-AR). DLMM-AR can be regarded as a frequentist counterpart of Bayesian hierarchical model with non-informative prior on parameters of both population and individual level. Details of the model specifications of the competing methods are shown below.

Linear mixed effect model full specifications

D. R. Zucker et al. (2010) proposed to use linear mixed effect model to account for correlations within the study clusters and combine multiple N-of-1 trials. They use data aggregated to the patient-period level, so the time series treatments of N-of-1 trial will be grouped into different periods, where treatments in each period will be same. Let $Y_{i,j}$ denote an average response variable of interest for the i-th individual N-of-1 trial in the j-th period ($j = 1, 2, ...J$) and $\boldsymbol{Y_i} = (Y_{i,1}, ..., Y_{i,j})$ denote the collection of response variable for the i-th individual N-of-1 trial. Let $X_{i,j}$ denote the treatment indicator for the i-th individual N-of-1 trial in the j-th period ($j = 1, 2, ...J$) and $\boldsymbol{X_i} = (X_{i,1}, ..., X_{i,j})$ denote the period-level treatment sequence of the i-th individual N-of-1 trial. The linear mixed effect model can be written as:

$$\boldsymbol{Y_i} = \boldsymbol{\mu 1} + \boldsymbol{X_i}\beta + \boldsymbol{Z_i b_i} + \boldsymbol{\epsilon_i}$$

$$\boldsymbol{b_i} \sim N(\boldsymbol{0}, \boldsymbol{G})$$

$$\boldsymbol{\epsilon_i} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma})$$

where $\boldsymbol{\beta}$ is a coefficient for treatment effect representing the fixed effect. Z_i is the design matrix for the random effects. The $\boldsymbol{b_i}$ are the random effects. Random intercept and random slope are included in random effects, thus the dimension for $\boldsymbol{b_i}$ is also 2. Here $\boldsymbol{G}$ is the covariance matrix

between random intercept and random slope. We assume they are independent, and hence $\boldsymbol{G}$ will be a diagonal matrix. $\boldsymbol{\Sigma}$ is the within-patient covariance matrix. There are several forms for $\boldsymbol{\Sigma}$ and for the sake of comparison with other methods, first-order autoregressive structure is used. To be specific,

$$\boldsymbol{\Sigma} = \frac{\sigma^2}{1-\rho^2}\begin{pmatrix} 1 & \rho & \dots & \rho^{J-1} \\ \rho & 1 & \dots & \rho^{J-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{J-1} & \rho^{J-2} & \dots & 1 \end{pmatrix}$$

Distributed lag model with mixed effect full specifications

An extension of distributed lag model that incorporates correlation within individual N-of-1 trial is using distributed lag model as the mean model and adding random effects of lag coefficients. The data is collected at day level and let X_t and Y_t denote the binary treatment indicator and the outcome of interest on day t for the i-th individual N-of-1 trial respectively, where $t = 1, 2, ..., n_i$ and $i = 1, 2, ..., S$. Let $\tilde{\boldsymbol{X}}_i$ be a $n_i \times (L+1)$ matrix with $X^*_{i,k-l+1}$ being the (k,l)-th element of $\tilde{\boldsymbol{X}}_i$, then the distributed lag model with mixed effect is:

$$\boldsymbol{Y_i} = \boldsymbol{\mu 1} + \tilde{\boldsymbol{X}}_i\boldsymbol{\beta} + \boldsymbol{Z_i b_i} + \boldsymbol{\epsilon_i}$$

$$\boldsymbol{b_i} \sim N(\boldsymbol{0}, \boldsymbol{G})$$

$$\boldsymbol{\epsilon_i} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma})$$

The model parameters are same as those in linear mixed effect model, except that $\boldsymbol{\beta}$ is a $l+1$ by 1 vector, and the dimension for random effects $\boldsymbol{b_i}$ and its covariance matrix $\boldsymbol{G}$ are also $l+1$ to model for random effects of every lag coefficient. The within-patient covariance matrix $\boldsymbol{\Sigma}$ follows first-order autoregressive structure.

3.4.3 Simulation Results

First, we investigate the performance of the proposed BHDLM-AR model in estimating population level parameter $\boldsymbol{\theta}$. BDLM-AR model is designed for individual N-of-1 trial analysis, therefore it cannot obtain estimates for population level parameters. Table 3.1 gives the summary of bias and root mean squared error, RMSE $= [1/n_{\text{sim}} \times \sum_{j=1}^{n_{\text{sim}}} (\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j)^2]^{1/2}$ under different lag curves with $S = 10$ and $\phi = 0.5$, where $\hat{\boldsymbol{\theta}}$ is the posterior mean / maximum likelihood estimate. LMM-AR only involves one treatment indicator in the model, which means it can only estimate total treatment effect and is not able to estimate the whole treatment effect curve over time. As we observe, compared with DLMM-AR, due to the ridge type of shrinkage at both individual and population level parameter, the proposed BHDLM-AR model has the lowest RMSE in estimating population level total effect, total carryover effect and immediate effect in most of the scenarios, but at the cost of larger bias in estimation. So the advantage of the proposed model is the ability of our method in breaking down the total effect and improving the estimation properties of carryover effect. Detailed summary of evaluation metrics for each lag coefficient, model standard deviation σ and autoregressive coefficient ϕ can be found in Table 3.2 to 3.6.

For individual level parameters, we estimate the average bias of estimation across individual trials and root of the squared error between the posterior mean and the true value of j-th simulation iteration of the i-th individual trial, and denoted as average root mean squared error (RAMSE), which is similar as the metric proposed by Katahira (2016) for evaluating estimation accuarcy at individual level:

$$\begin{aligned} \text{Average bias} &= \frac{1}{S \times n_{\text{sim}}} \sum_{i=1}^{S} \sum_{j=1}^{n_{\text{sim}}} (\hat{\boldsymbol{\beta}}_{ij} - \boldsymbol{\beta}_{ij}) \\ \text{RAMSE} &= \sqrt{\frac{1}{S \times n_{\text{sim}}} \sum_{i=1}^{S} \sum_{j=1}^{n_{\text{sim}}} (\hat{\boldsymbol{\beta}}_{ij} - \boldsymbol{\beta}_{ij})^2} \end{aligned} \tag{3.16}$$

For LMM-AR and DLMM-AR, best linear unbiased prediction (BLUP, Robinson et al., 1991) is used for the estimation of random effects. Table 3.7 summarizes the performance of each method in estimating individual level parameters under different lag curves with $S = 10$ and $\phi = 0.5$.

		BHDLM-AR	LMM-AR	DLMM-AR
Bias	**Total Effect**			
	Exponential decay	-0.98	**-0.24**	0.76
	Exponential decay (Oscillated)	-0.31	**0.29**	0.30
	Slow absorption	-1.29	-0.89	**0.37**
	Slow absorption (Oscillated)	0.66	0.51	**-0.12**
	No carryover	0.13	**-0.02**	-0.05
	Total Carryover Effect			
	Exponential decay	**-0.92**	-	-4.57
	Exponential decay (Oscillated)	0.32	-	**0.12**
	Slow absorption	-2.57	-	**0.23**
	Slow absorption (Oscillated)	1.03	-	**-0.17**
	No carryover	1.24	-	**-0.14**
	Immediate Effect			
	Exponential decay	**-0.05**	-	5.33
	Exponential decay (Oscillated)	-0.63	-	**0.17**
	Slow absorption	1.28	-	**0.14**
	Slow absorption (Oscillated)	-0.37	-	**0.05**
	No carryover	-1.11	-	**0.09**
RMSE	**Total Effect**			
	Exponential decay	2.28	**2.25**	2.47
	Exponential decay (Oscillated)	**2.08**	2.32	2.52
	Slow absorption	**2.44**	2.49	2.49
	Slow absorption (Oscillated)	**2.15**	2.73	2.90
	No carryover	2.22	**2.19**	2.52
	Total Carryover Effect			
	Exponential decay	**1.95**	-	5.17
	Exponential decay (Oscillated)	**1.74**	-	2.21
	Slow absorption	3.09	-	**2.58**
	Slow absorption (Oscillated)	**1.99**	-	2.53
	No carryover	**2.06**	-	2.37
	Immediate Effect			
	Exponential decay	**1.94**	-	5.88
	Exponential decay (Oscillated)	**2.04**	-	2.23
	Slow absorption	**2.32**	-	2.45
	Slow absorption (Oscillated)	**1.97**	-	2.28
	No carryover	2.75	-	**2.66**

Table 3.1: Summary of bias and RMSE (best values in bold) of population level total effect, total carryover effect and immediate effect (θ_0) under five lag coefficient curves, estimated by three models.

		Bias			RMSE		
	Truth	BHDLM-AR	LMM-AR	DLMM-AR	BHDLM-AR	LMM-AR	DLMM-AR
μ	10	-1.25	-0.39	-0.62	3.42	3.51	3.55
Total effect	9.69	-0.98	-0.24	0.76	2.28	2.25	2.47
Total carryover effect	4.69	-0.92	-	-4.57	1.95	-	5.17
θ_0	5	-0.05	-	5.33	1.94	-	5.88
θ_1	2.50	-0.43	-	-2.82	1.33	-	3.73
θ_2	1.25	-0.23	-	-1.16	0.76	-	2.42
θ_3	0.62	-0.21	-	-0.62	0.48	-	2.17
θ_4	0.31	-0.15	-	-0.39	0.28	-	2.14
θ_5	0	0.07	-	0.21	0.14	-	2.13
θ_6	0	0.02	-	0.16	0.00	-	2.03
θ_7	0	0.01	-	0.05	0.00	-	1.81
ϕ	0.5	-	-0.38	-0.01	-	0.47	0.10
σ	10	-0.02	-5.75	1.32	0.26	5.82	1.38

Table 3.2: Summary of evaluation metrics for population level parameters under exponential decay lag coefficient curve.

		Bias			RMSE		
	Truth	BHDLM-AR	LMM-AR	DLMM-AR	BHDLM-AR	LMM-AR	DLMM-AR
μ	10	-1.39	0.02	0.01	3.50	3.71	3.69
Total effect	5.94	-0.31	0.29	0.30	2.08	2.32	2.52
Total carryover effect	0.94	0.32	-	0.12	1.74	-	2.21
θ_0	5	-0.63	-	0.17	2.04	-	2.23
θ_1	2.50	-1.03	-	-0.20	1.61	-	2.02
θ_2	-1.25	1.09	-	0.39	1.31	-	2.16
θ_3	-0.63	0.54	-	-0.12	0.69	-	2.07
θ_4	0.31	-0.30	-	0.04	0.38	-	2.04
θ_5	0	0.01	-	-0.06	0.11	-	1.93
θ_6	0	0.01	-	-0.20	0.05	-	2.03
θ_7	0	0.00	-	0.27	0.02	-	1.67
ϕ	0.5	-	-0.46	-0.01	-	0.54	0.10
σ	10	-0.02	-5.61	1.28	0.26	5.77	1.33

Table 3.3: Summary of evaluation metrics for population level parameters under exponential decay (oscillated) lag coefficient curve.

		Bias			RMSE		
	Truth	BHDLM-AR	LMM-AR	DLMM-AR	BHDLM-AR	LMM-AR	DLMM-AR
μ	10	-1.13	0.99	0.57	3.38	3.11	3.00
Total effect	9.99	-1.29	-0.89	0.37	2.44	2.49	2.49
Total carryover effect	8.49	-2.57	-	0.23	3.09	-	2.58
θ_0	1.51	1.28	-	0.14	2.32	-	2.45
θ_1	2.75	-0.26	-	0.15	1.28	-	2.38
θ_2	3.36	-1.33	-	-0.05	1.51	-	2.35
θ_3	2.03	-1.11	-	0.09	1.19	-	2.03
θ_4	0.34	-0.03	-	-0.19	0.24	-	2.12
θ_5	0	0.11	-	-0.10	0.14	-	2.19
θ_6	0	0.04	-	0.18	0.00	-	2.21
θ_7	0	0.01	-	0.14	0.00	-	1.80
ϕ	0.5	-	-0.46	-0.01	-	0.55	0.10
σ	10	-0.02	-5.62	1.31	0.26	5.71	1.36

Table 3.4: Summary of evaluation metrics for population level parameters under slow absorption lag coefficient curve.

		Bias			RMSE		
	Truth	BHDLM-AR	LMM-AR	DLMM-AR	BHDLM-AR	LMM-AR	DLMM-AR
μ	10	-1.56	-0.65	-0.45	3.57	3.56	3.50
Total effect	-0.79	0.66	0.51	-0.12	2.15	2.73	2.90
Total carryover effect	-2.30	1.03	-	-0.17	1.99	-	2.53
θ_0	1.51	-0.37	-	0.05	1.97	-	2.28
θ_1	2.75	-1.95	-	-0.08	2.32	-	2.12
θ_2	-3.36	2.09	-	-0.23	2.21	-	2.01
θ_3	-2.03	1.42	-	0.12	1.49	-	2.06
θ_4	0.34	-0.48	-	-0.01	0.54	-	2.05
θ_5	0	-0.03	-	0.00	0.10	-	1.90
θ_6	0	-0.01	-	0.19	0.00	-	1.83
θ_7	0	0.00	-	-0.17	0.00	-	1.92
ϕ	0.5	-	-0.45	-0.01	-	0.54	0.00
σ	10	-0.01	-5.82	1.22	0.26	5.89	1.28

Table 3.5: Summary of evaluation metrics for population level parameters under slow absorption (oscillated) lag coefficient curve.

		Bias			RMSE		
	Truth	BHDLM-AR	LMM-AR	DLMM-AR	BHDLM-AR	LMM-AR	DLMM-AR
μ	10	-0.50	-0.09	-0.09	3.48	3.10	3.13
Total effect	10	0.13	-0.02	-0.05	2.22	2.19	2.52
Total carryover effect	0	1.24	-	-0.14	2.06	-	2.37
θ_0	10	-1.11	-	0.09	2.75	-	2.66
θ_1	0	0.72	-	-0.01	1.41	-	2.07
θ_2	0	0.28	-	0.17	0.77	-	2.11
θ_3	0	0.13	-	-0.16	0.45	-	2.04
θ_4	0	0.06	-	-0.27	0.24	-	2.32
θ_5	0	0.03	-	0.28	0.10	-	1.89
θ_6	0	0.01	-	0.02	0.00	-	1.92
θ_7	0	0.01	-	-0.17	0.00	-	1.68
ϕ	0.5	-	-0.46	-0.01	-	0.56	0.00
σ	10	-0.02	-5.70	1.20	0.26	5.85	1.27

Table 3.6: Summary of evaluation metrics for population level parameters under no carryover effect lag coefficient curve.

As we can see, the proposed model has the lowest RAMSE in all of the scenarios. Compared to fitting separate individual BDLM-AR model, the prior structure in BHDLM-AR pulls the estimates of individual level parameters towards the population mean, thus yielding the estimation to be less sensitive to noise. Comparison results under different number of subjects and model error autocorrelation are similar (Table 3.9 to 3.14).

Figure 3.1 further shows the performance of estimating each individual level lag coefficient in terms of average bias and RAMSE. DLMM-AR has the smallest average bias in almost all lag coefficients, but largest RAMSE since it behaves similarly as Bayesian hierarchical distributed lag model with non-informative prior. BHDLM-AR and BDLM-AR model have large average bias at early lag coefficients but average bias and RAMSE decreases as lag increases due to the designed prior structure on lag coefficients with increasing ridge and smoothness penalties. Compared with fitting individual N-of-1 trials separately with BDLM-AR model, BHDLM-AR further reduces average bias and RAMSE. This could be because the hierarchical structure of BHDLM-AR enables the estimation of individual level parameters not only depends on its own observations, but also "borrow strength" from other N-of-1 trials through utilizing the information at population level parameters. Overall, the proposed BHDLM-AR model has the smallest RAMSE in all lag

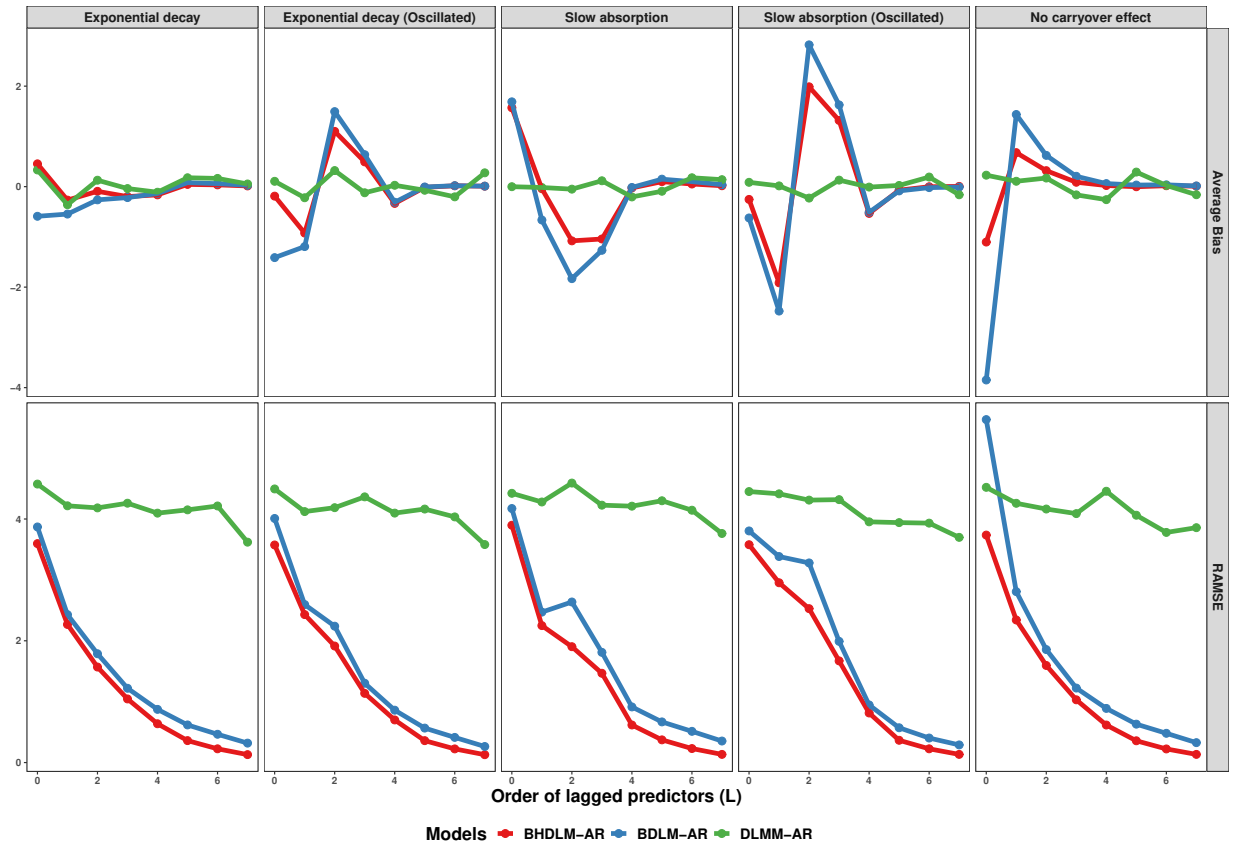


Figure 3.1: Average bias and RAMSE of posterior mean (Maximum likelihood) estimated lag coefficients under five different true lag coefficient curve with 10 subjects and $\phi = 0.5$.

coefficients under different scenarios.

3.5 Revisiting Light Therapy Study

We used data from I. M. Kronish et al. (2020), which studies the effectiveness of bright white light therapy for depressive symptoms within cancer survivors. Eight eligible patients completed a 12-week N-of-1 trial using both bright white light (10,000 lux) and dim red light (50 lux) intervention, but in different sequence. The outcome of interest is a composite depressive score calculated through a combination of depressive symptom, fatigue symptom and negative affectivity. Additional details regarding the data can be found in Section 2.5. Figure 3.2 shows the daily assessments of all patients in the study.

The proposed BHDLM-AR model was applied to the light therapy data. We choose number of

		BHDLM-AR	BDLM-AR	LMM-AR	DLMM-AR
Average Bias	**Total Effect**				
	Exponential decay	**-0.18**	-1.59	-0.34	0.34
	Exponential decay (Oscillated)	0.16	-0.77	**0.10**	0.11
	Slow absorption	-0.44	-1.80	-1.19	**0.07**
	Slow absorption (Oscillated)	0.53	0.73	0.68	**0.05**
	No carryover	**0.03**	-1.44	0.26	0.24
	Total Carryover Effect				
	Exponential decay	-0.63	-1.00	-	**0.01**
	Exponential decay (Oscillated)	0.35	0.65	-	**0.00**
	Slow absorption	-2.02	-3.49	-	**0.07**
	Slow absorption (Oscillated)	0.79	1.35	-	**-0.04**
	No carryover	1.13	2.41	-	**0.01**
	Immediate Effect				
	Exponential decay	0.45	-0.59	-	**0.33**
	Exponential decay (Oscillated)	-0.19	-1.41	-	**0.11**
	Slow absorption	1.57	1.69	-	**0.00**
	Slow absorption (Oscillated)	-0.25	-0.62	-	**0.09**
	No carryover	-1.10	-3.85	-	**0.23**
RAMSE	**Total Effect**				
	Exponential decay	**3.75**	4.43	3.91	4.04
	Exponential decay (Oscillated)	**3.75**	4.00	4.06	4.23
	Slow absorption	**3.78**	4.56	3.95	4.02
	Slow absorption (Oscillated)	**3.79**	3.92	3.99	4.17
	No carryover	**3.74**	4.37	3.85	4.00
	Total Carryover Effect				
	Exponential decay	**3.33**	4.00	-	4.66
	Exponential decay (Oscillated)	**3.28**	3.68	-	4.69
	Slow absorption	**3.85**	5.43	-	4.83
	Slow absorption (Oscillated)	**3.35**	3.85	-	4.69
	No carryover	**3.45**	4.55	-	4.72
	Immediate Effect				
	Exponential decay	**3.60**	3.87	-	4.57
	Exponential decay (Oscillated)	**3.57**	4.01	-	4.49
	Slow absorption	**3.90**	4.17	-	4.42
	Slow absorption (Oscillated)	**3.58**	3.80	-	4.45
	No carryover	**3.74**	5.63	-	4.52

Table 3.7: Summary of average bias and RAMSE (best values in bold) of individual level total effect, total carryover effect and immediate effect (β_0) under five lag coefficient curves, estimated by four models.

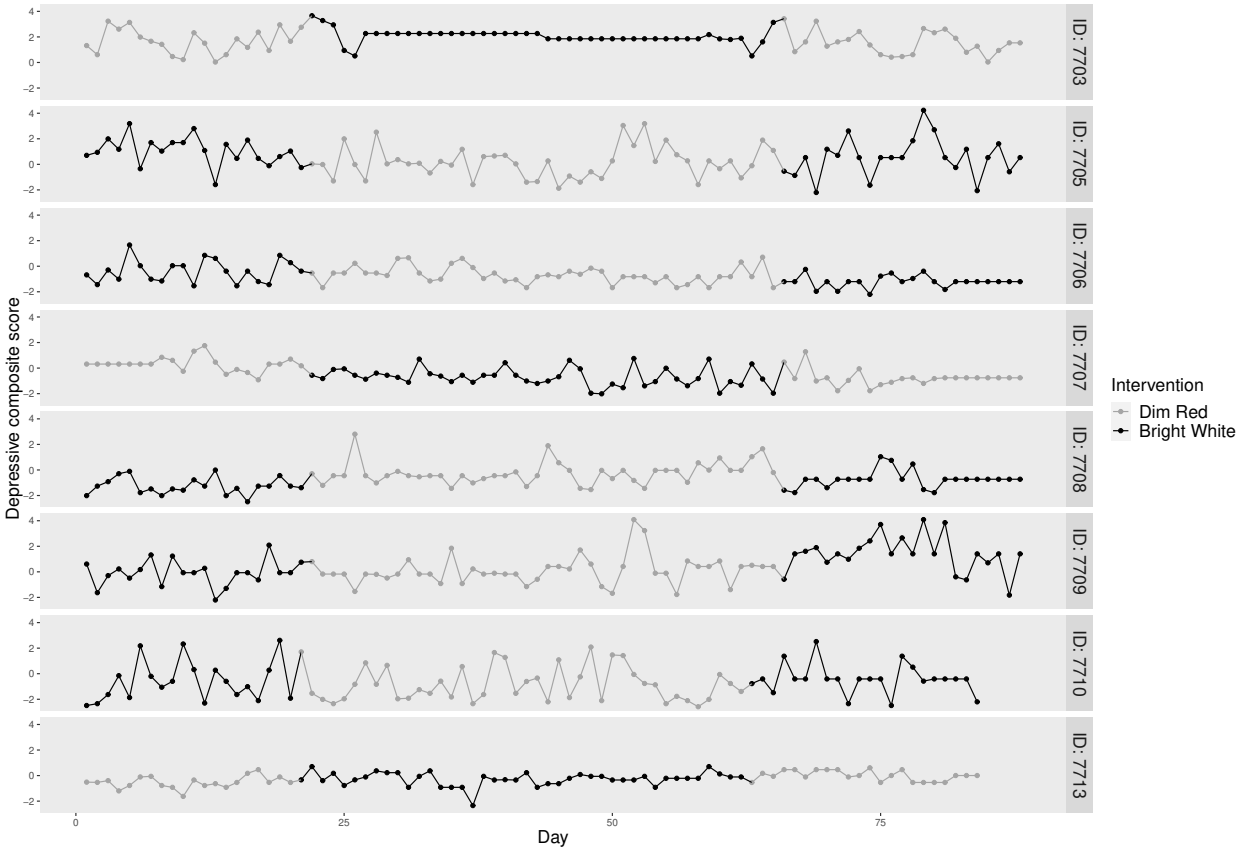


Figure 3.2: Daily assessments of eight subjects in the light therapy study. Black line represents bright white light intervention, and grey line represents dim red light.

lags $L = 7$ so that the model can capture the lagged treatment effect up to a week, which is long enough to exhibit the treatment effect on composite depressive score. As a comparison, we also applied BDLM-AR model that fitted data from individual N-of-1 trial separately. Convergence of all the MCMC were checked using both trace plots and the GelmanRubin diagnostics (Gelman, Rubin, et al., 1992). To be specific, the potential scale reduction factor for lag coefficients, autoregressive coefficients and model variance all are smaller than 1.2, as Brooks and Gelman, 1998 have suggested, indicating the convergence of posterior samples.

The posterior means of each lag coefficient at both population and individual level are shown in Figure 3.3 as a function of lag. Results fitted by BHDLM-AR and BDLM-AR model are presented in solid and dashed line respectively. Point-wise 90% posterior credible intervals for each lag coefficients, total effect, total carryover effect and autoregressive coefficient are summarized in

Table 3.8. At population level, BHDLM-AR suggests a positive immediate treatment effect in decreasing composite depressive score (-0.4, 90% CI: -0.90, 0.09) while a negative carryover effect at lag 1 (0.44, 90% CI: -0.02, 0.91) and lag 2 (-0.16, 90% CI: -0.23, 0.56), indicating patients receiving light therapy will have immediate benefit from the treatment but the adverse effect will follow for around 2 days. In the Bayesian hierarchical model, we also observed estimated lag coefficient curve of single subject follows the trend of population level estimated lag coefficient curve, which is known as the "shrinkage" effect of the individual parameters estimates towards the population mean induced by Bayesian hierarchical structure (Efron and Morris, 1977). At individual level, patient 7707 and 7708 are the only two who have positive total treatment effect. The composite depressive score decreases 0.37 (90% CI: -0.80, 0.20) point in patient 7707 and 0.58 (90% CI: -1.07, -0.06) point in patient 7708. This result is aligned with their positive feedback after their individual N-of-1 trial. Another interesting finding is that three patients (ID: 7705, 7709 and 7710) have been found significant carryover effect by the hierarchical model, which provides extra information in understanding the light therapy treatment effect over the time for each individual patient.

3.6 Discussion

In this chapter, we introduce a Bayesian approach to combine results from multiple N-of-1 trials and estimate lag coefficients both at population and individual level. BHDLM-AR is an extension to the BDLM-AR model as described in Chapter 2, and preserve the ability to handle temporal correlation between observations and carryover effect through distributed lag structure. In addition, the proposed method uses Bayesian hierarchical structure to account for both within and between-patient variation, which can improve the estimates for individuals.

Compared with analyzing each single N-of-1 trial separately by BDLAR model, the hierarchical model reflects the sampling regime more appropriately in the design since the heterogeneity of treatment effect can be regard as a kind of random effects. In the simulation, we showed BHDLM-AR outperforms single BDLAR model in estimating total treatment effect, total carryover effect

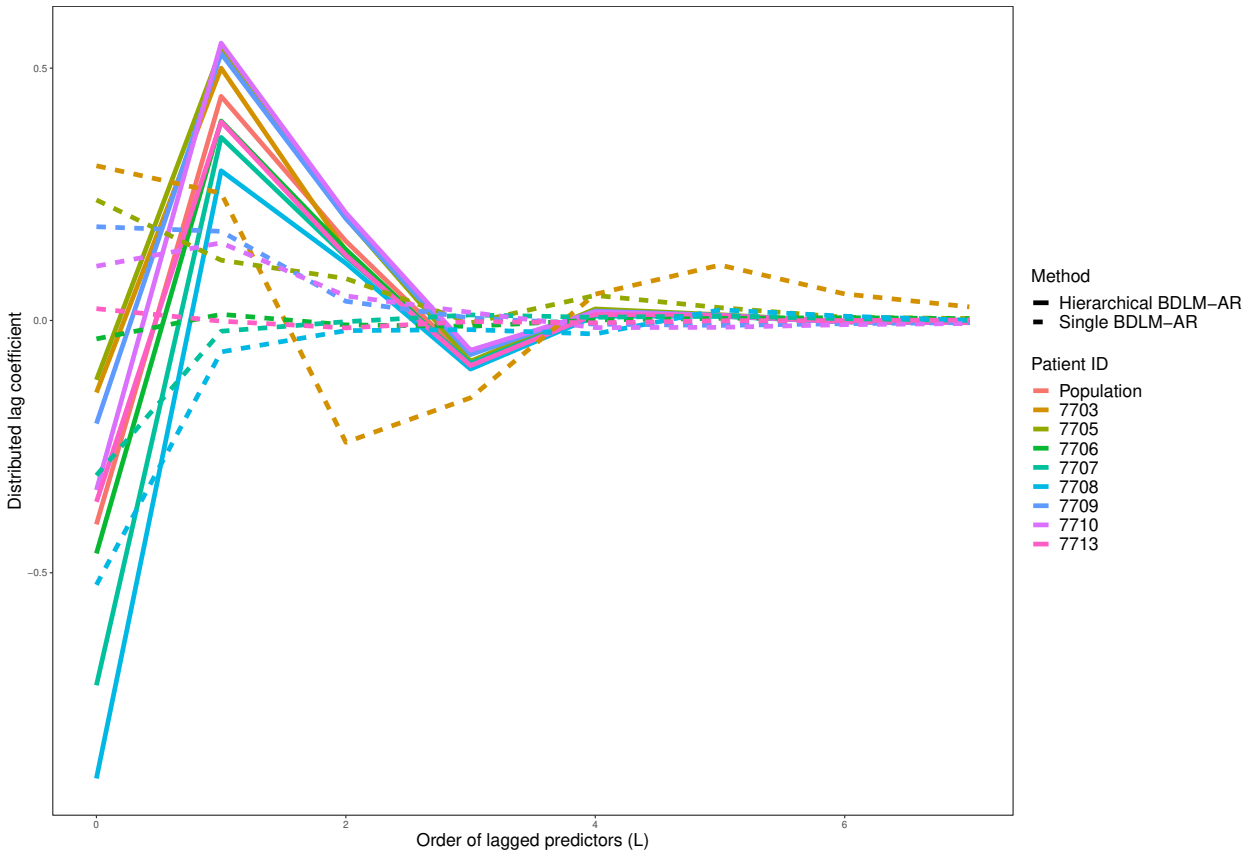


Figure 3.3: Posterior mean of the lag coefficients of eight patients in light therapy study. Solid and dashed line represent data fitted by BHDLM-AR and single BDLM-AR.

Table 3.8: Posterior means of total effect, total carryover effect, lag coefficients and autoregressive coefficients for light therapy study, fitted by BHDLM-AR and single BDLM-AR. 90% credible intervals are in brackets. Significant carryover effect estimated from BHDLM-AR are indicated in bold.

	Bayesian Hierarchical Distributed Lag Model with Autocorrelated Error								
	ID: 7703	ID: 7705	ID: 7706	ID: 7707	ID: 7708	ID: 7709	ID: 7710	ID: 7713	Population
Intercept	1.62 (1.2,2.04)	0.18 (-0.1,0.46)	-0.71 (-1.04,-0.39)	-0.37 (-0.79,0.04)	-0.32 (-0.69,0.02)	0.15 (-0.19,0.49)	-0.82 (-1.06,-0.57)	-0.32 (-0.89,0.2)	-0.1 (-5.83,5.66)
Total effect	0.42 (-0.12,0.94)	0.57 (0.16,0.98)	0 (-0.45,0.47)	-0.31 (-0.8,0.2)	-0.58 (-1.07,-0.06)	0.48 (0.01,0.94)	0.39 (0.03,0.74)	0.09 (-0.48,0.72)	0.13 (-0.3,0.56)
Total carryover effect	0.56 (-0.02,1.14)	**0.69 (0.14,1.24)**	0.46 (-0.09,1.02)	0.42 (-0.15,0.99)	0.32 (-0.24,0.88)	**0.69 (0.12,1.25)**	**0.72 (0.18,1.28)**	0.45 (-0.13,1.03)	**0.54 (0.09,0.99)**
Lag 0	-0.14 (-0.77,0.49)	-0.12 (-0.7,0.46)	-0.46 (-1.06,0.14)	-0.72 (-1.33,-0.11)	-0.91 (-1.5,-0.3)	-0.2 (-0.8,0.4)	-0.34 (-0.91,0.23)	-0.36 (-0.99,0.3)	-0.4 (-0.9,0.09)
Lag 1	0.5 (-0.04,1.05)	0.54 (0,1.07)	0.4 (-0.14,0.93)	0.36 (-0.18,0.9)	0.3 (-0.24,0.84)	0.53 (0,1.07)	0.55 (0.01,1.09)	0.39 (-0.15,0.93)	0.44 (-0.02,0.91)
Lag 2	0.14 (-0.28,0.58)	0.2 (-0.22,0.65)	0.14 (-0.28,0.58)	0.13 (-0.3,0.58)	0.11 (-0.3,0.54)	0.2 (-0.22,0.64)	0.21 (-0.21,0.65)	0.13 (-0.29,0.57)	0.16 (-0.23,0.56)
Lag 3	-0.1 (-0.41,0.24)	-0.08 (-0.4,0.25)	-0.09 (-0.4,0.24)	-0.08 (-0.4,0.24)	-0.1 (-0.41,0.23)	-0.07 (-0.38,0.26)	-0.06 (-0.37,0.27)	-0.09 (-0.41,0.24)	-0.08 (-0.38,0.23)
Lag 4	0.01 (-0.23,0.26)	0.02 (-0.22,0.26)	0.02 (-0.23,0.26)	0.01 (-0.23,0.26)	0.01 (-0.24,0.25)	0.02 (-0.22,0.26)	0.02 (-0.23,0.26)	0.01 (-0.23,0.25)	0.02 (-0.22,0.25)
Lag 5	0.01 (-0.12,0.13)	0.01 (-0.12,0.14)	0.01 (-0.12,0.13)	0.01 (-0.12,0.13)	0.01 (-0.12,0.13)	0.01 (-0.11,0.13)	0.01 (-0.12,0.13)	0.01 (-0.12,0.13)	0.01 (-0.11,0.13)
Lag 6	0 (-0.06,0.06)	0 (-0.06,0.06)	0 (-0.06,0.06)	0 (-0.06,0.06)	0 (-0.06,0.06)	0 (-0.06,0.06)	0 (-0.06,0.06)	0 (-0.06,0.06)	0 (-0.05,0.06)
Lag 7	0 (-0.03,0.03)	0 (-0.03,0.03)	0 (-0.03,0.03)	0 (-0.03,0.03)	0 (-0.03,0.03)	0 (-0.03,0.03)	0 (-0.03,0.03)	0 (-0.03,0.03)	0 (-0.03,0.03)
ϕ	0.4 (0.16,0.64)	0.15 (0.01,0.29)	0.24 (-0.01,0.49)	0.25 (0.01,0.5)	0.29 (0.07,0.52)	0.3 (0.16,0.44)	-0.02 (-0.16,0.11)	0.42 (0.05,0.82)	-
	Bayesian Distributed Lag Model with Autocorrelated Error								
	ID: 7703	ID: 7705	ID: 7706	ID: 7707	ID: 7708	ID: 7709	ID: 7710	ID: 7713	Population
Intercept	2.04 (1.74,2.33)	0.73 (0.34,1.13)	-0.73 (-0.97,-0.49)	-0.67 (-0.92,-0.42)	-0.92 (-1.26,-0.58)	0.57 (0.11,1.03)	-0.48 (-0.82,-0.14)	-0.27 (-0.44,-0.09)	-
Total effect	0.42 (0,0.86)	0.54 (-0.01,1.1)	-0.03 (-0.34,0.27)	-0.3 (-0.7,0.06)	-0.62 (-1.07,-0.12)	0.37 (-0.2,0.99)	0.28 (-0.15,0.77)	0 (-0.22,0.21)	-
Total carryover effect	0.1 (-0.41,0.67)	0.28 (-0.27,1.1)	0 (-0.26,0.28)	0.01 (-0.33,0.44)	-0.1 (-0.57,0.36)	0.19 (-0.28,0.92)	0.17 (-0.24,0.87)	-0.02 (-0.25,0.15)	-
Lag 0	0.31 (-0.18,0.89)	0.26 (-0.47,0.9)	-0.03 (-0.37,0.28)	-0.32 (-0.8,0.07)	-0.52 (-1.07,-0.03)	0.18 (-0.49,0.81)	0.11 (-0.54,0.66)	0.02 (-0.2,0.27)	-
Lag 1	0.25 (-0.14,0.87)	0.12 (-0.35,0.66)	0.01 (-0.2,0.23)	-0.02 (-0.3,0.27)	-0.07 (-0.43,0.3)	0.18 (-0.2,0.8)	0.15 (-0.2,0.74)	0 (-0.15,0.15)	-
Lag 2	-0.22 (-0.91,0.11)	0.08 (-0.21,0.5)	-0.01 (-0.15,0.12)	0 (-0.19,0.19)	-0.02 (-0.28,0.23)	0.04 (-0.24,0.38)	0.05 (-0.18,0.39)	-0.01 (-0.13,0.07)	-
Lag 3	-0.14 (-0.68,0.1)	0 (-0.26,0.25)	-0.01 (-0.1,0.06)	0.01 (-0.1,0.17)	-0.02 (-0.22,0.15)	0 (-0.2,0.22)	0.02 (-0.14,0.22)	0 (-0.07,0.06)	-
Lag 4	0.05 (-0.15,0.36)	0.05 (-0.08,0.31)	0 (-0.05,0.05)	0.01 (-0.07,0.12)	-0.03 (-0.21,0.09)	-0.01 (-0.17,0.12)	-0.01 (-0.15,0.08)	0 (-0.06,0.03)	-
Lag 5	0.1 (-0.05,0.5)	0.02 (-0.07,0.19)	0 (-0.02,0.04)	0.01 (-0.04,0.1)	0.02 (-0.06,0.17)	-0.01 (-0.13,0.07)	-0.01 (-0.13,0.05)	0 (-0.03,0.02)	-
Lag 6	0.05 (-0.07,0.32)	0.01 (-0.07,0.11)	0 (-0.01,0.03)	0 (-0.03,0.06)	0.01 (-0.05,0.1)	-0.01 (-0.1,0.05)	-0.01 (-0.08,0.03)	0 (-0.02,0.02)	-
Lag 7	0.02 (-0.09,0.22)	0 (-0.05,0.07)	0 (-0.01,0.02)	0 (-0.03,0.03)	0 (-0.05,0.06)	0 (-0.06,0.03)	-0.01 (-0.05,0.02)	0 (-0.01,0.01)	-
ϕ	0.35 (0.17,0.53)	0.16 (-0.02,0.34)	0.22 (0.03,0.41)	0.22 (0.02,0.41)	0.28 (0.09,0.47)	0.32 (0.14,0.5)	0 (-0.19,0.19)	0.32 (0.13,0.5)	-

and each lag coefficient. When number of observations in some of the single N-of-1 trail is small, the hierarchical structure allows these trials to borrow strength from other trials to improve estimation.

In BHDLM-AR model, both population and individual level lag coefficients are added with a designed prior structure so that both lag coefficient curves are shrunk to behave in a monotone decreasing trend. The trend at large lag is usually right since treatment effect will finally diminish to zero, but the trend at first several lag varies treatment to treatment. If more informative prior can be obtained from clinician's experience or pharmacokinetic/pharmacodynamic research, we may have more accurate estimation.

The use of BHDLM-AR model to combined N-of-1 trials has practical meanings. From a regulatory perspective, the population level estimated lag coefficient curve is of interest since it can provide a more precise general association between the treatment and outcome while adjusting for treatment heterogeneity across patients. From a patient perspective, the posterior distribution of population level parameters can be utilized for future light therapy treatment recommendation and improve the estimation at individual level.

The proposed model is applied to a real application of using white light therapy for depressive symptoms. We find an overall positive immediate treatment effect while a negative carryover is identified for around 2 days. The results on light therapy study should be interpreted carefully. The 90% credible interval of a lot of lag coefficients cover zero, indicating substantial uncertainty in parameters estimation. But in view of the small number of observations of each patient in the study, compared to other applications of distributed lag model such as Peng et al. (2009) used 4 years' day level data and Zanobetti and Schwartz (2008) made use of 12 years of data, we could expect an increasing in precision of parameters estimation if more data is collected. Although the finding of light therapy study is interesting, considering the small number of patients in the study, the results may not be generalized to the overall population. However, further study on the biological mechanism of light therapy can be performed to obtain a better understanding of its relationship to depressive symptoms.

		BHDLM-AR	LMM-AR	DLMM-AR
Bias	**Total Effect**			
	Exponential decay	-0.63	-0.61	**0.10**
	Exponential decay (Oscillated)	-0.28	**0.15**	**0.15**
	Slow absorption	-0.79	-1.21	**-0.02**
	Slow absorption (Oscillated)	0.21	0.50	**-0.14**
	No carryover	-0.53	**0.12**	0.13
	Total Carryover Effect			
	Exponential decay	-0.56	-	**0.05**
	Exponential decay (Oscillated)	0.17	-	**0.04**
	Slow absorption	-1.56	-	**-0.01**
	Slow absorption (Oscillated)	0.54	-	**-0.21**
	No carryover	0.69	-	**0.12**
	Immediate Effect			
	Exponential decay	-0.07	-	**0.06**
	Exponential decay (Oscillated)	-0.45	-	**0.11**
	Slow absorption	0.76	-	**-0.01**
	Slow absorption (Oscillated)	-0.34	-	**0.07**
	No carryover	-1.23	-	**0.02**
RMSE	**Total Effect**			
	Exponential decay	**1.56**	1.59	1.59
	Exponential decay (Oscillated)	**1.45**	1.52	1.60
	Slow absorption	**1.63**	2.02	1.73
	Slow absorption (Oscillated)	**1.45**	1.93	1.99
	No carryover	**1.52**	1.78	1.95
	Total Carryover Effect			
	Exponential decay	**1.65**	-	1.75
	Exponential decay (Oscillated)	**1.57**	-	1.66
	Slow absorption	2.20	-	**1.71**
	Slow absorption (Oscillated)	**1.65**	-	1.71
	No carryover	**1.71**	-	1.87
	Immediate Effect			
	Exponential decay	**1.62**	-	1.73
	Exponential decay (Oscillated)	1.68	-	**1.53**
	Slow absorption	1.79	-	**1.77**
	Slow absorption (Oscillated)	**1.66**	-	1.69
	No carryover	2.04	-	**1.71**

Table 3.9: Summary of bias and RMSE (best values in bold) of population level total effect, total carryover effect and immediate effect (θ_0) under five lag coefficient curves, estimated by three models with 20 subjects and $\phi = 0.5$.

		BHDLM-AR	LMM-AR	DLMM-AR
Bias	**Total Effect**			
	Exponential decay	**0.00**	-0.34	0.02
	Exponential decay (Oscillated)	-0.32	**0.26**	0.27
	Slow absorption	-0.95	-1.36	**-0.23**
	Slow absorption (Oscillated)	0.43	0.49	**-0.15**
	No carryover	-0.74	**-0.18**	-0.25
	Total Carryover Effect			
	Exponential decay	-0.40	-	**0.11**
	Exponential decay (Oscillated)	0.15	-	**0.10**
	Slow absorption	-1.99	-	**-0.30**
	Slow absorption (Oscillated)	0.45	-	**-0.18**
	No carryover	1.22	-	**-0.34**
	Immediate Effect			
	Exponential decay	0.39	-	**-0.09**
	Exponential decay (Oscillated)	-0.47	-	**0.16**
	Slow absorption	1.04	-	**0.07**
	Slow absorption (Oscillated)	**-0.02**	-	0.03
	No carryover	-1.96	-	**0.09**
RMSE	**Total Effect**			
	Exponential decay	2.06	**2.04**	2.36
	Exponential decay (Oscillated)	**1.92**	2.12	2.27
	Slow absorption	**2.15**	2.47	2.26
	Slow absorption (Oscillated)	**1.93**	2.40	2.51
	No carryover	**2.05**	2.27	2.43
	Total Carryover Effect			
	Exponential decay	**1.88**	-	2.12
	Exponential decay (Oscillated)	**1.75**	-	2.07
	Slow absorption	2.66	-	**2.16**
	Slow absorption (Oscillated)	**1.79**	-	2.32
	No carryover	**2.13**	-	2.37
	Immediate Effect			
	Exponential decay	**1.86**	-	2.37
	Exponential decay (Oscillated)	**1.92**	-	2.24
	Slow absorption	**2.13**	-	2.35
	Slow absorption (Oscillated)	**1.85**	-	2.21
	No carryover	2.71	-	**2.39**

Table 3.10: Summary of bias and RMSE (best values in bold) of population level total effect, total carryover effect and immediate effect (θ_0) under five lag coefficient curves, estimated by three models with 10 subjects and $\phi = 0.2$.

		BHDLM-AR	LMM-AR	DLMM-AR
Bias	**Total Effect**			
	Exponential decay	-0.52	-0.64	**0.02**
	Exponential decay (Oscillated)	-0.31	0.15	**0.14**
	Slow absorption	-0.57	-1.25	**-0.07**
	Slow absorption (Oscillated)	**0.05**	0.48	-0.13
	No carryover	-0.52	**0.12**	**0.12**
	Total Carryover Effect			
	Exponential decay	-0.41	-	**0.03**
	Exponential decay (Oscillated)	**0.00**	-	0.05
	Slow absorption	-1.12	-	**-0.02**
	Slow absorption (Oscillated)	**0.06**	-	-0.20
	No carryover	0.78	-	**0.11**
	Immediate Effect			
	Exponential decay	-0.12	-	**-0.01**
	Exponential decay (Oscillated)	-0.31	-	**0.09**
	Slow absorption	0.54	-	**-0.05**
	Slow absorption (Oscillated)	**-0.02**	-	0.06
	No carryover	-1.30	-	**0.02**
RMSE	**Total Effect**			
	Exponential decay	**1.42**	1.48	**1.42**
	Exponential decay (Oscillated)	**1.36**	**1.36**	1.45
	Slow absorption	**1.45**	1.92	1.56
	Slow absorption (Oscillated)	**1.32**	1.72	1.74
	No carryover	**1.42**	1.60	1.75
	Total Carryover Effect			
	Exponential decay	**1.59**	-	1.68
	Exponential decay (Oscillated)	**1.54**	-	**1.54**
	Slow absorption	1.91	-	**1.57**
	Slow absorption (Oscillated)	**1.54**	-	1.56
	No carryover	**1.73**	-	1.74
	Immediate Effect			
	Exponential decay	**1.53**	-	1.75
	Exponential decay (Oscillated)	1.57	-	**1.51**
	Slow absorption	**1.62**	-	1.74
	Slow absorption (Oscillated)	**1.53**	-	1.62
	No carryover	2.01	-	**1.66**

Table 3.11: Summary of bias and RMSE (best values in bold) of population level total effect, total carryover effect and immediate effect (θ_0) under five lag coefficient curves, estimated by three models with 20 subjects and $\phi = 0.2$.

		BHDLM-AR	BDLM-AR	LMM-AR	DLMM-AR
Average Bias	**Total Effect**				
	Exponential decay	**-0.10**	-1.54	-0.49	0.20
	Exponential decay (Oscillated)	0.09	-0.76	**0.05**	**0.05**
	Slow absorption	-0.27	-1.81	-1.10	**0.08**
	Slow absorption (Oscillated)	0.31	0.58	0.67	**0.04**
	No carryover	0.03	-1.31	**0.00**	0.01
	Total Carryover Effect				
	Exponential decay	-0.36	-1.11	-	**0.07**
	Exponential decay (Oscillated)	0.24	0.61	-	**-0.03**
	Slow absorption	-1.23	-3.54	-	**0.06**
	Slow absorption (Oscillated)	0.50	1.45	-	**-0.13**
	No carryover	0.75	2.23	-	**0.07**
	Immediate Effect				
	Exponential decay	0.26	-0.43	-	**0.13**
	Exponential decay (Oscillated)	-0.15	-1.37	-	**0.08**
	Slow absorption	0.96	1.73	-	**0.02**
	Slow absorption (Oscillated)	-0.19	-0.87	-	**0.18**
	No carryover	-0.72	-3.53	-	**-0.06**
RAMSE	**Total Effect**				
	Exponential decay	**3.73**	4.30	3.86	4.41
	Exponential decay (Oscillated)	**3.73**	3.92	3.87	4.71
	Slow absorption	**3.74**	4.47	3.97	4.84
	Slow absorption (Oscillated)	**3.74**	3.84	3.84	4.39
	No carryover	**3.73**	4.21	3.89	4.40
	Total Carryover Effect				
	Exponential decay	**3.23**	3.80	-	4.03
	Exponential decay (Oscillated)	**3.21**	3.49	-	4.12
	Slow absorption	**3.44**	5.31	-	4.44
	Slow absorption (Oscillated)	**3.24**	3.77	-	4.04
	No carryover	**3.29**	4.25	-	4.12
	Immediate Effect				
	Exponential decay	**3.48**	3.69	-	4.11
	Exponential decay (Oscillated)	**3.48**	3.88	-	4.29
	Slow absorption	**3.61**	4.11	-	4.33
	Slow absorption (Oscillated)	**3.48**	3.73	-	4.09
	No carryover	**3.55**	5.21	-	4.06

Table 3.12: Summary of average bias and RAMSE (best values in bold) of individual level total effect, total carryover effect and immediate effect (β_0) under five lag coefficient curves, estimated by four models with 20 subjects and $\phi = 0.5$.

		BHDLM-AR	BDLM-AR	LMM-AR	DLMM-AR
Average Bias	**Total Effect**				
	Exponential decay	**0.00**	-0.73	-0.44	0.05
	Exponential decay (Oscillated)	0.13	-0.32	**0.07**	0.08
	Slow absorption	-0.11	-0.76	-1.06	**0.07**
	Slow absorption (Oscillated)	0.29	0.39	0.67	**0.03**
	No carryover	0.07	-0.57	**-0.03**	-0.10
	Total Carryover Effect				
	Exponential decay	-0.40	-0.57	-	**0.08**
	Exponential decay (Oscillated)	0.15	0.57	-	**-0.02**
	Slow absorption	-1.42	-2.88	-	**-0.19**
	Slow absorption (Oscillated)	0.16	1.03	-	**-0.05**
	No carryover	1.23	2.45	-	**-0.24**
	Immediate Effect				
	Exponential decay	0.39	-0.16	-	**-0.03**
	Exponential decay (Oscillated)	-0.02	-0.89	-	**0.10**
	Slow absorption	1.31	2.13	-	**0.26**
	Slow absorption (Oscillated)	0.13	-0.64	-	**0.08**
	No carryover	-1.16	-3.02	-	**0.15**
RAMSE	**Total Effect**				
	Exponential decay	**2.63**	2.88	2.78	2.91
	Exponential decay (Oscillated)	**2.63**	2.70	2.78	2.88
	Slow absorption	**2.63**	2.91	2.91	2.80
	Slow absorption (Oscillated)	**2.64**	2.69	2.78	2.83
	No carryover	**2.63**	2.82	2.79	2.99
	Total Carryover Effect				
	Exponential decay	**3.13**	3.57	-	4.43
	Exponential decay (Oscillated)	**3.10**	3.23	-	4.36
	Slow absorption	**3.42**	4.66	-	4.43
	Slow absorption (Oscillated)	**3.10**	3.43	-	4.28
	No carryover	**3.34**	4.11	-	4.59
	Immediate Effect				
	Exponential decay	**3.25**	3.49	-	4.38
	Exponential decay (Oscillated)	**3.22**	3.33	-	4.36
	Slow absorption	**3.49**	4.01	-	4.39
	Slow absorption (Oscillated)	**3.23**	3.31	-	4.28
	No carryover	**3.43**	4.51	-	4.34

Table 3.13: Summary of average bias and RAMSE (best values in bold) of individual level total effect, total carryover effect and immediate effect (β_0) under five lag coefficient curves, estimated by four models with 10 subjects and $\phi = 0.2$.

		BHDLM-AR	BDLM-AR	LMM-AR	DLMM-AR
Average Bias	**Total Effect**				
	Exponential decay	**0.00**	-0.73	-0.53	0.14
	Exponential decay (Oscillated)	0.07	-0.38	0.05	**0.04**
	Slow absorption	-0.05	-0.83	-1.14	**0.04**
	Slow absorption (Oscillated)	0.15	0.35	0.65	**0.03**
	No carryover	0.04	-0.64	**0.00**	**0.00**
	Total Carryover Effect				
	Exponential decay	-0.21	-0.78	-	**0.06**
	Exponential decay (Oscillated)	0.07	0.60	-	**-0.02**
	Slow absorption	-0.78	-2.91	-	**0.05**
	Slow absorption (Oscillated)	**0.00**	1.09	-	-0.13
	No carryover	0.83	2.48	-	**0.06**
	Immediate Effect				
	Exponential decay	**0.20**	0.05	-	0.08
	Exponential decay (Oscillated)	**0.00**	-0.98	-	0.06
	Slow absorption	0.73	2.07	-	**-0.02**
	Slow absorption (Oscillated)	**0.15**	-0.74	-	0.16
	No carryover	-0.80	-3.11	-	**-0.06**
RAMSE	**Total Effect**				
	Exponential decay	**2.60**	2.80	2.74	2.75
	Exponential decay (Oscillated)	**2.60**	2.67	2.70	2.77
	Slow absorption	**2.61**	2.88	2.91	2.67
	Slow absorption (Oscillated)	**2.61**	2.65	2.74	2.72
	No carryover	**2.60**	2.78	2.76	2.89
	Total Carryover Effect				
	Exponential decay	**2.98**	3.35	-	3.67
	Exponential decay (Oscillated)	**2.98**	3.15	-	3.71
	Slow absorption	**3.08**	4.60	-	3.88
	Slow absorption (Oscillated)	**2.97**	3.37	-	3.64
	No carryover	**3.09**	4.07	-	3.77
	Immediate Effect				
	Exponential decay	**3.10**	3.23	-	3.77
	Exponential decay (Oscillated)	**3.10**	3.32	-	3.74
	Slow absorption	**3.18**	3.94	-	3.83
	Slow absorption (Oscillated)	**3.10**	3.29	-	3.66
	No carryover	**3.20**	4.55	-	3.73

Table 3.14: Summary of average bias and RAMSE (best values in bold) of individual level total effect, total carryover effect and immediate effect (β_0) under five lag coefficient curves, estimated by four models with 20 subjects and $\phi = 0.2$.

Chapter 4: Extension of Bayesian Distributed Lag Model with Multiple Interventions

4.1 Introduction

Most N-of-1 trials so far considered two interventions within each patient in the study, either comparison of drug and placebo or drug and active control (Coxeter et al., 2003). But a lot of therapeutic areas involve multiple guideline-recommended or first line drug classes, such as hypertension and oncology studies (I. M. Kronish et al., 2019), the only way to know the best treatment for a patient is to have the patient try the drugs first, if no additional medical diagnose or genetic information is available. When multiple interventions are used in the N-of-1 trial, a straightforward approach to analyze the data is to evaluate their effects one at a time by fitting two interventions model. However, the potential carryover effect from several previous interventions may be hard to evaluate in multiple single models.

Distributed lag models (DLMs) were first developed in econometric area (Almon, 1965; Koyck, 1954), have in recent years been applied widely used in advertising (Bass and Clarke, 1972), and environment health studies (Welty et al., 2009; Zanobetti et al., 2000). They are regression models for time series data when current value of a dependent variable is not only associated with current value of an explanatory variable, but also its lagged values. The advantage of DLMs is that it allows the effect of an exposure to be distributed over a specific period of time, and provides a better understanding of the exposure-outcome relationship. Very few methods so far consider multiple interventions in N-of-1 trials and their potential carryover effect over the study time period. In this

chapter, we extended the Bayesian distributed lag model with autocorrelated error (BDLM-AR) in Chapter 2 to multiple interventions with one placebo (or active control) scenario.

As a motivating example, we consider a study that uses N-of-1 trials to find personalized blood pressure medications. For each hypertensive patient, several medications from three different first-line blood pressure medications classes were assigned in 12 weeks with a counterbalanced sequence (i.e., ABCCBA). The outcome of interest is systolic blood pressure which is measured twice in the morning and twice at night by patients themselves.

The rest of this chapter is organized as follows. In Section 4.2, we introduce the proposed Bayesian distributed lag model with autocorrelated errors (BDLM-AR) that works for multiple interventions. In Section 4.3, the posterior distributions and MCMC algorithms are presented. We then illustrate our models by analyzing data from hypertension study to estimate both total treatment effect and carryover effect of each intervention. And finally, We conclude with a discussion in Section 4.5.

4.2 Methods

4.2.1 Proposed Model

Suppose we observe data from a patient on n consecutive days with a total of C interventions during the entire trial. On day $t = 1, \ldots, n$, let Y_t denote the outcome of interest and $X_{t,c}$ denote c-th binary treatment indicator, where $c = 1, 2, ...C$. The distributed lag autoregressive model for multiple interventions can be described as follows:

$$Y_t = \mu + \sum_{l=0}^{L} \sum_{c=0}^{C-1} \beta_{l,c} X_{t-l,c} + \epsilon_t \tag{4.1}$$

for $t = p + 1, ..., n$, where the error term ϵ_t follows an autoregressive process,

$$\epsilon_t = \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + ... + \phi_p \epsilon_{t-p} + w_t \tag{4.2}$$

for $t = p+1, ..., n$, where the error term $\epsilon_t = \phi_1\epsilon_{t-1} + \phi_2\epsilon_{t-2} + ... + \phi_p\epsilon_{t-p} + w_t$, w_t is a white Gaussian noise with mean zero and unknown variance $\sigma^2 > 0$, and μ is the intercept. Note that for $t < L$, the maximum lag effect is of order $t-1$, and terms involve X with non-positive subscript are not included in the model.

For intervention c, $\beta_{0,c}$ measures the current treatment effect, and $\beta_{i,c}, i > 0$, measures the carryover effect of lag-i intervention. Unlike group-level crossover design where carryover effects are usually assumed to exist only in the immediate past treatment period, the distributed lag models consider the whole time course of carryover treatment effect. In particular, $\sum_{i=1}^{L}\beta_{i,c} = E(Y_t|X_{t-1,c} = 1, ...X_{t-L,c} = 1, X_{t,c}) - E(Y_t|X_{t-1,c} = 0, ..., X_{t-L,c} = 0, X_{t,c})$ measures the cumulative carryover treatment effect of intervention c from past L lags.

As mentioned previously, it could be challenging to estimate lag coefficient $\boldsymbol{\beta}_c = (\beta_{0,c}, ..., \beta_{L,c})$ due to the potential collinearity of the lagged treatment indicators. Therefore, we used the same prior in Chapter 2 on $\boldsymbol{\beta}_c$, which has little constraint on first several lag coefficients and gradually increased constraint on following coefficients. At the same time, the constraint on $\boldsymbol{\beta}_c$ is expected to alleviate multicollinearity problem.

To estimate the parameters, we reformat model (4.1) as follows. Let $\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p$, where B is the backshift operator. Then the autoregression model for the errors can be written as $\Phi(B)\epsilon_t = w_t$. Denote $Y_t^* = \Phi(B)Y_t$, $X_{t,c}^* = \Phi(B)X_{t,c}$ and $\mu^* = \Phi(B)\mu$. Applying $\Phi(B)$ to both sides of equation (4.1), we obtain

$$Y_t^* = \mu^* + \sum_{i=0}^{L}\sum_{c=1}^{C-1}\beta_{i,c}X_{t-i,c}^* + w_t,$$

for $t = p+1, ..., n$. This implies that

$$(\boldsymbol{Y}^* \mid \boldsymbol{X}^*, \mu^*, \boldsymbol{\beta}) \sim N(\mu^*\mathbf{1}_{n-p} + \sum_{c=1}^{C-1}\boldsymbol{X}_c^*\boldsymbol{\beta}_c, \sigma^2\mathbf{I}_{n-p}) \tag{4.3}$$

where $\boldsymbol{Y}^* = (Y_{p+1}^*, ..., Y_n^*)'$, $\boldsymbol{X}^* = (\boldsymbol{X}_1^*, ..., \boldsymbol{X}_C^*)$ and $\mathbf{I}$ is the identity matrix. Each element in $\boldsymbol{X}^*$, $\boldsymbol{X}_c^*$ is a $(n-p)\times(L+1)$ matrix with $X_{k-i+p+1,c}^*$ being the (k,i)-th element of $\boldsymbol{X}_c^*$, In following

sections, we denote $\tilde{\boldsymbol{\beta}}_c = (\mu, \boldsymbol{\beta}_c')'$ and $\tilde{\boldsymbol{X}}_c^* = (\Phi(B)\mathbf{1}_{n-p}, \boldsymbol{X}_c^*)$ for $c = 1$. Otherwise, $\tilde{\boldsymbol{\beta}}_c = \boldsymbol{\beta}_c$ and $\tilde{\boldsymbol{X}}_c^* = \boldsymbol{X}_c^*$.

4.2.2 Prior Distribution on the Mean Model

Prior on $\tilde{\boldsymbol{\beta}}_c$. The same normal prior distribution as Chapter 2 for $\tilde{\boldsymbol{\beta}}_c$. One advantage of using normal prior is that the mode of the posterior distribution is equivalent to Ridge regression estimate (i.e., l_2-penalty) in the frequentist setting, which deals with the collinearity problem. In addition, the banded structure variance covariance matrix in the prior distribution has the property that 1) the ℓ_2 penalty increases with lag i, such that coefficients at large lag are more likely to be shrink to zero, which is in accordance with the fact that delayed treatment effect will decrease and diminish eventually; 2) additional regularization is imposed to the difference between adjacent coefficients to smooth the lag coefficient curve. Specifically, the prior on $\tilde{\boldsymbol{\beta}}_c$ is

$$\tilde{\boldsymbol{\beta}}_c \sim N(\mathbf{0}, \sigma^2 \tilde{\boldsymbol{\Omega}}_c^{-1}), \tag{4.4}$$

where $\tilde{\boldsymbol{\Omega}}_c = diag(c_0, \boldsymbol{\Omega}_c)$ is a precision matrix if $c = 1$ and otherwise, $\tilde{\boldsymbol{\Omega}}_c = \boldsymbol{\Omega}_c$. c_0 is a scalar, and $\boldsymbol{\Omega}_c$ is a $(L+1) \times (L+1)$ matrix. Note that c_0^{-1} is the variance of the prior on intercept μ. The detailed form of $\boldsymbol{\Omega}_c$ can be found in equation (2.6). We need to note that in equation (2.6), the precision matrix $\boldsymbol{\Omega}$ is a function of $\boldsymbol{\lambda} = (\lambda_0, ..., \lambda_L)$ and $\boldsymbol{\lambda}^* = (\lambda_0^*, ..., \lambda_L^*)$ and when multiple interventions are used, $\boldsymbol{\Omega}_c$ is governed by index c, and $\boldsymbol{\Omega}_c$ is a function of $\boldsymbol{\lambda}_c = (\lambda_{0,c}, ..., \lambda_{L,c})$ and $\boldsymbol{\lambda}_c^* = (\lambda_{0,c}^*, ..., \lambda_{L,c}^*)$ for $c = 1, 2, ..., C$.

With prior distribution (2.5), the resulting maximum a posteriori probability (MAP) estimate of $\tilde{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\beta}}$ is of the fused Ridge estimate form:

$$\hat{\boldsymbol{\beta}} = \underset{\mu, \boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_C}{\operatorname{argmin}} (\boldsymbol{Y}^* - \sum_{c=0}^{C-1} \tilde{\boldsymbol{X}}_c^* \tilde{\boldsymbol{\beta}}_c)' (\boldsymbol{Y}^* - \sum_{c=0}^{C-1} \tilde{\boldsymbol{X}}_c^* \tilde{\boldsymbol{\beta}}_c) + c_0 \mu^2 + \sum_{c=1}^{C-1} \sum_{i=0}^{L} \lambda_{i,c} \beta_{i,c}^2 + \sum_{c=1}^{C-1} \sum_{i=0}^{L} \lambda_{i,c}^* (\beta_{i,c} - \beta_{i+1,c})^2$$

where $\beta_{L+1,c} \triangleq 0$. The fused ridge regularization penalizes not only the ℓ_2-norm of the coefficients but also their successive differences, thereby encouraging local smoothness.

We further parameterize $\boldsymbol{\lambda}_c$ and $\boldsymbol{\lambda}_c^*$ by $\lambda_{i,c} = \exp[\gamma_{1,c}(i+1)]$ and $\lambda_{i,c}^* = \exp[\gamma_{2,c}(i+1)]$ for $i = 0, 1, 2, ...L$. Here $\gamma_{1,c}, \gamma_{2,c} \geq 0$ are intervention specific hyper-parameters. $\gamma_{1,c}$ controls the rate at which the ridge penalty increases, hence also controls the rate that coefficients taper to 0. $\gamma_{2,c}$ controls the increasing rate of smoothness of the coefficients curve. Under such parameterization, coefficients at large lags are more likely to shrink to zero. In the following, we rewrite $\tilde{\boldsymbol{\Omega}}_c$ as $\tilde{\boldsymbol{\Omega}}_c(\boldsymbol{\gamma}_c)$ to indicate the dependence of $\tilde{\boldsymbol{\Omega}}_c$ on $\boldsymbol{\gamma}_c = (\gamma_{1,c}, \gamma_{2,c})$.

Prior on hyperparameters $\boldsymbol{\gamma}_c$. We use standard exponential hyperprior for $\boldsymbol{\gamma}_c = (\gamma_{1,c}, \gamma_{2,c})$, which will provide a natural non-negative support. The prior distribution of $\boldsymbol{\gamma}_c$ is truncated to the region $S_{\boldsymbol{\gamma}}$ to ensure that precision matrix $\boldsymbol{\Omega}_c$ is positive definite,

$$\pi(\gamma_{1,c}, \gamma_{2,c}) \propto \exp\left(-\gamma_{1,c} - \gamma_{2,c}\right) \mathbf{1}_{S_\gamma}(\gamma_{1,c}, \gamma_{2,c}) \tag{4.5}$$

Then the amount of ridge and smooth penalization can be determined in the posterior inference procedure.

4.2.3 Prior Distribution on the Error Model

Prior on σ^2. We use Jeffreys prior to model the error variance σ^2. That is,

$$\pi(\sigma^2) \propto 1/\sigma^2 \tag{4.6}$$

Note that any inverse-gamma prior for σ^2 would maintain conjugacy, and the Jeffreys prior can be regarded as an improper limit of inverse-gamma prior distribution.

Prior on $\boldsymbol{\phi}$. We consider a truncated normal prior for $\boldsymbol{\phi}$ so that the error process $\{\epsilon_t, t = 1, 2, \ldots\}$ is stationary. Following Chib (1993), we can construct a region $S_{\boldsymbol{\phi}}$, where error series $\boldsymbol{\epsilon}$ is stationary. To be specific, all roots of the polynomial $\Phi(z) = 1 - \sum_{i=1}^{p} \phi_i z^i$ are outside the unit circle. Diffuse multivariate normal prior is used for $\boldsymbol{\phi}$ and truncated to the region $S_{\boldsymbol{\phi}}$,

$$\boldsymbol{\phi} \sim N_p\left(0_p, \sigma_{\boldsymbol{\phi}}^2 \mathbf{I}_p\right) \mathbf{1}_{S_\phi}(\boldsymbol{\phi}) \tag{4.7}$$

where $\mathbf{1}_{S_{\boldsymbol{\phi}}}(\boldsymbol{\phi})$ denotes the indicator function of the region $S_{\boldsymbol{\phi}}$. $\sigma^2_{\boldsymbol{\phi}}$ is assumed to be 200, which is a hundred times the range of single autoregressive coefficient, so that there will be little influence on the autoregressive coefficients.

4.3 Conditional Posterior Distributions

In this section, we provide the full conditional distributions of all parameters discussed above, and propose a hybrid Metropolis-Hastings/Gibbs algorithm to estimate the parameters. Details of derivation of full conditional distribution for each parameter can be found in Appendix C.1.

Given the likelihood function (4.3) and prior distribution of $\tilde{\boldsymbol{\beta}}_c$, the full conditional distribution of $\tilde{\boldsymbol{\beta}}_c$ is

$$\begin{aligned}\tilde{\boldsymbol{\beta}}_c \mid \boldsymbol{Y}, \tilde{\boldsymbol{X}}_c, \tilde{\boldsymbol{\beta}}_{-c}, \sigma^2, \boldsymbol{\phi}, \boldsymbol{\gamma}_c \sim N_{L+1}\{&[\tilde{\boldsymbol{X}}_c^{*'}\tilde{\boldsymbol{X}}_c^* + \tilde{\boldsymbol{\Omega}}_c(\boldsymbol{\gamma}_c)]^{-1}\tilde{\boldsymbol{X}}_c^{*'}(\boldsymbol{Y}^* + \boldsymbol{X}_c^*\boldsymbol{\beta}_c - \sum_{k=1}^{C-1}\boldsymbol{X}_k^*\boldsymbol{\beta}_k),\\ &\sigma^2[\tilde{\boldsymbol{X}}_c^{*'}\tilde{\boldsymbol{X}}_c^* + \tilde{\boldsymbol{\Omega}}_c(\boldsymbol{\gamma}_c)]^{-1}\}\end{aligned} \tag{4.8}$$

Where $\tilde{\boldsymbol{\beta}}_{-c}$ denotes all elements in $\tilde{\boldsymbol{\beta}}$ except for $\tilde{\boldsymbol{\beta}}_c$. Similarly, it can be shown that the full condition distribution of $\boldsymbol{\phi}$ is truncated multivariate normal distribution,

$$\boldsymbol{\phi} \mid \boldsymbol{Y}, \tilde{\boldsymbol{X}}, \tilde{\boldsymbol{\beta}}, \sigma^2, \boldsymbol{\gamma} \sim N_p\left[\left(\sigma^{-2}\boldsymbol{E}^{*'}\boldsymbol{E}^* + \sigma_{\boldsymbol{\phi}}^{-2}\mathbf{I}\right)^{-1}\sigma^{-2}\boldsymbol{E}^{*'}\boldsymbol{\epsilon}^*, \left(\sigma^{-2}\boldsymbol{E}^{*'}\boldsymbol{E}^* + \sigma_{\boldsymbol{\phi}}^{-2}\mathbf{I}\right)^{-1}\right]\mathbf{1}_{S_{\boldsymbol{\phi}}}(\boldsymbol{\phi}) \tag{4.9}$$

where $\boldsymbol{\epsilon}^* = (\epsilon^*_{p+1}, \ldots, \epsilon^*_n)'$, $\epsilon^*_t = Y_t - \mu - \sum_{l=0}^{L}\sum_{c=1}^{C-1}\beta_{l,c}X_{t-l,c}$, and $\boldsymbol{E}^*$ is a $(n-p)\times p$ matrix with ϵ^*_{p+k-j} being the (k,j)-th element.

The prior distribution of $\tilde{\boldsymbol{\beta}}$ is conditional on σ^2. Therefore, the posterior distribution of σ^2 depends on the conditional distribution $\pi(\sigma^2|\tilde{\boldsymbol{\beta}})$. From (4.3), (4.4) and (4.6), the full conditional distribution of σ^2 is inverse-gamma distribution:

$$\sigma^2 \mid \boldsymbol{Y}, \tilde{\boldsymbol{X}}, \tilde{\boldsymbol{\beta}}, \boldsymbol{\phi}, \boldsymbol{\gamma} \sim \mathrm{IG}\left[\frac{n-p+(C-1)(L+1)}{2}, \frac{(\boldsymbol{Y}^* - \sum_{c=1}^{C-1}\tilde{\boldsymbol{X}}_c^*\tilde{\boldsymbol{\beta}}_c)'(\boldsymbol{Y}^* - \sum_{c=1}^{C-1}\tilde{\boldsymbol{X}}_c^*\tilde{\boldsymbol{\beta}}_c) + \tilde{\boldsymbol{\beta}}_c'\tilde{\boldsymbol{\Omega}}_c(\boldsymbol{\gamma}_c)\tilde{\boldsymbol{\beta}}_c}{2}\right] \tag{4.10}$$

The full conditional distribution of $\boldsymbol{\gamma}_c$ is proportional to:

$$\pi(\boldsymbol{\gamma} \mid \boldsymbol{Y}, \tilde{\boldsymbol{X}}_c, \tilde{\boldsymbol{\beta}}_c, \boldsymbol{\phi}, \sigma^2) \propto |\sigma^{-2}\tilde{\boldsymbol{\Omega}}_c(\boldsymbol{\gamma}_c)|^{\frac{1}{2}} \exp\left[-\frac{1}{2\sigma^2}\tilde{\boldsymbol{\beta}}_c'\tilde{\boldsymbol{\Omega}}_c(\boldsymbol{\gamma}_c)\tilde{\boldsymbol{\beta}}_c\right] \exp(-\gamma_{1,c}-\gamma_{2,c})\mathbf{1}_{S_{\gamma_c}}(\gamma_{1,c}, \gamma_{2,c}) \tag{4.11}$$

Since there is no closed form of the full conditional distribution, we propose to generate $\boldsymbol{\gamma}$ using Metropolis-Hastings (MH) algorithm with proposal distribution as a uniform distribution $U(-a, a)$. We tune the parameter a such that the acceptance rate of proposed sample is around 50% (Roberts, Gelman and Gilks, 1997). Recall that $\tilde{\boldsymbol{\Omega}}_c(\boldsymbol{\gamma}_c)$ is a precision matrix of $\tilde{\boldsymbol{\beta}}_c$, and therefore is positive definite. Samples of $\boldsymbol{\gamma}_c$ causing a non positive definite precision matrix will not be accepted in the MH algorithm.

Having the full conditional distribution of all parameters in the proposed model, we use a hybrid Metropolis-Hastings/Gibbs algorithm to generate samples of $(\tilde{\boldsymbol{\beta}}, \boldsymbol{\phi}, \sigma^2, \boldsymbol{\gamma})$ from the posterior distribution. The detailed algorithm is given below.

Algorithm 3: The hybrid Metropolis-Hastings/Gibbs algorithm

Step 1. Set initial values for $\tilde{\boldsymbol{\beta}}$, σ^2, $\boldsymbol{\phi}$ and $\boldsymbol{\gamma}$;

for $i \leftarrow 1$ *to* $n_{iteration}$ **do**

  for $c \leftarrow 1$ *to* $C-1$ **do**

    Step 2. Given current value of $\boldsymbol{\phi}$, transform $\boldsymbol{Y}$, $\tilde{\boldsymbol{X}}$ to $\boldsymbol{Y}^*$, $\tilde{\boldsymbol{X}}^*$ as described in equation (4.3); Also construct precision matrix $\tilde{\boldsymbol{\Omega}}_c(\boldsymbol{\gamma}_c)$ based on $\boldsymbol{\gamma}_c$ as described in Section 4.4;

    Step 3. Conditional on current values of $\boldsymbol{Y}^*$, $\tilde{\boldsymbol{X}}_c^*$, σ^2 and $\tilde{\boldsymbol{\Omega}}_c(\boldsymbol{\gamma}_c)$, update $\tilde{\boldsymbol{\beta}}_c$ based on $\pi(\tilde{\boldsymbol{\beta}}_c \mid \boldsymbol{Y}^*, \tilde{\boldsymbol{X}}_c^*, \tilde{\boldsymbol{\beta}}_{-c}, \sigma^2, \boldsymbol{\phi}, \boldsymbol{\gamma}_c)$;

    Step 4. Update $(\gamma_{1,c}, \gamma_{2,c})$ based on $\pi(\boldsymbol{\gamma}_c \mid \tilde{\boldsymbol{\beta}}_c, \sigma^2)$. Sample a proposal $\gamma_{i,c}^*$ by $\gamma_{i,c}^* = \gamma_{i,c} + a * U(-1,1)$ for $i = 1, 2$. a is an adjustable step size. Compute the ratio

$$R_\gamma = \frac{\pi(\boldsymbol{\gamma}_c^* \mid \tilde{\boldsymbol{\beta}}_c, \sigma^2)}{\pi(\boldsymbol{\gamma}_c \mid \tilde{\boldsymbol{\beta}}_c, \sigma^2)}$$

    if $\tilde{\boldsymbol{\Omega}}_c(\boldsymbol{\gamma}_c^*)$ *is positive definite* **then**

      update $\boldsymbol{\gamma}_c = \boldsymbol{\gamma}_c^*$ with probability $\min(1, R_\gamma)$;

    end

  end

  Step 4. Conditional on current values of $\boldsymbol{Y}^*$, $\tilde{\boldsymbol{X}}^*$, $\tilde{\boldsymbol{\beta}}$, $\boldsymbol{\phi}$ and $\tilde{\boldsymbol{\Omega}}(\boldsymbol{\gamma})$, update σ^2 based on $\pi(\sigma^2 \mid \boldsymbol{Y}^*, \tilde{\boldsymbol{X}}^*, \tilde{\boldsymbol{\beta}}, \boldsymbol{\phi}, \boldsymbol{\gamma})$;

  Step 5. Update $\boldsymbol{\epsilon}^*$ conditional on current value of $\tilde{\boldsymbol{\beta}}$ and $\boldsymbol{Y}$, $\tilde{\boldsymbol{X}}$. Then update $\boldsymbol{\phi}$ based on $\pi(\boldsymbol{\phi} \mid \boldsymbol{Y}, \tilde{\boldsymbol{X}}, \tilde{\boldsymbol{\beta}}, \sigma^2, \boldsymbol{\gamma})$. Reject samples if the roots of $\boldsymbol{\phi}(L)$ lie outside the unit circle;

end

4.4 Hypertension Treatment Study

We utilize the dataset from a hypertension treatment study (I. M. Kronish et al., 2019) that use N-of-1 trials to determine the personalized selection of blood pressure medications. Traditional clinical approach to treating hypertension is to start with selecting one out of four guideline-recommended drug classes by clinicians without using too much information from patients. Multiple studies (Mancia et al., 2011; Materson, 2007) have shown that antihypertension medications may have different treatment effect on patients, while existing techniques to precisely make personalized medications recommendation through genetic tests are still immature. N-of-1 trials provide an alternative way to select the medications.

In this study, 7 patients with a history of mild hypertension were included. Each patient was assigned three blood pressure medications from three different first-line blood pressure classes: hydrochlorothiazide (HCTZ), amlodipine and losartan. Each trial was designed to last 12 weeks with a counterbalanced sequence ABCCBA. Some patients were allowed to include other medications if they was already taking one, for example Dyazide (triamterene and HCTZ). One-week's washout period was allowed between different medications. The outcome of interest is systolic blood pressure (SBP), which was measured by patients twice in the morning and twice in at night by home blood pressure device. Example time series curve of outcome collected from two patients was presented in Figure 4.1.

We fit the data with the proposed BDLM-AR model and the maximum number of lags L is set to 7 since a week's period is long enough to reflect the difference in treatment effect between medications and treatment effect after a week will hardly impact the outcome. The error autoregressive (AR) order is fixed to 1. Multiple observations in the morning and at night will be combined separately by taking average. Therefore, a patient will have two observations everyday. Some occasional missing outcomes were imputed using average non-missing value in the corresponding treatment block. Convergence of all the MCMC was checked using both trace plots and the GelmanRubin diagnostics (Gelman, Rubin, et al., 1992). We also fit a classical linear regression

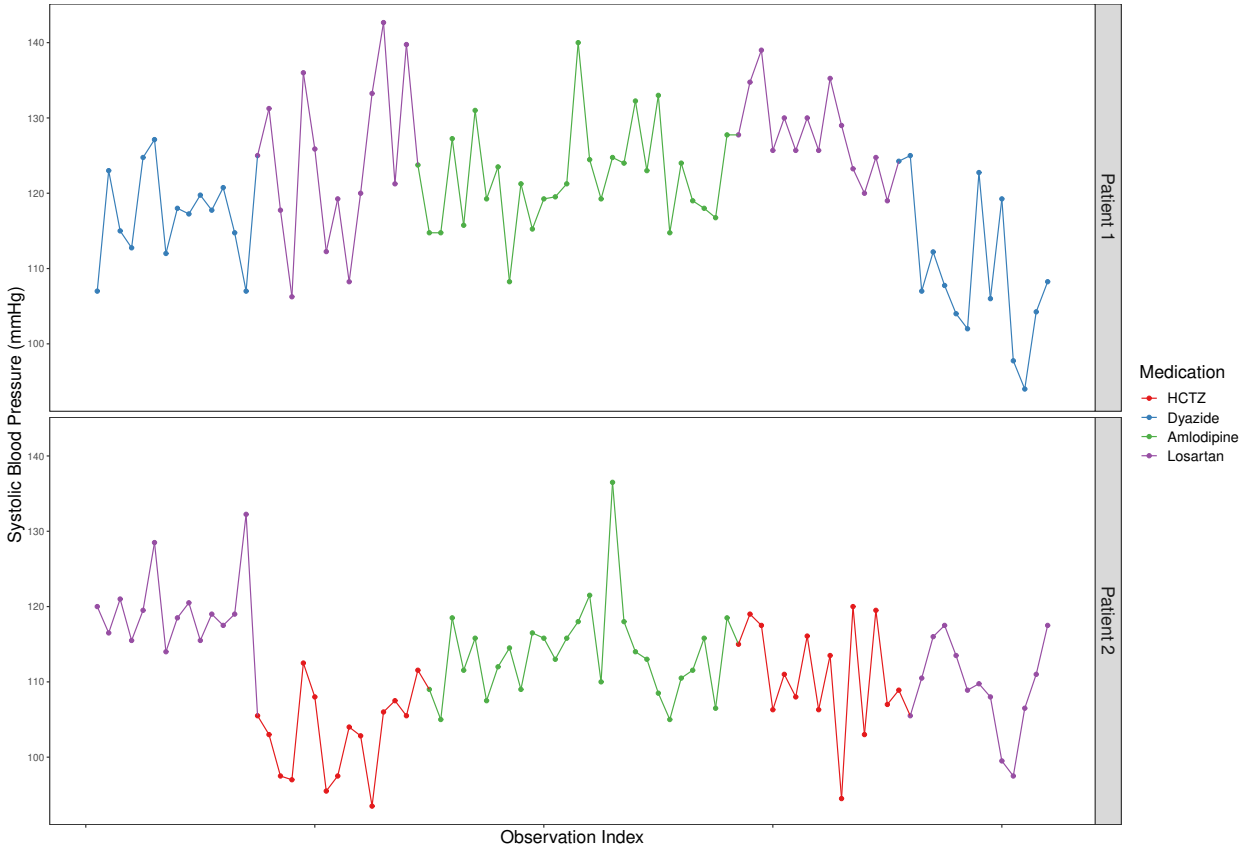


Figure 4.1: Systolic blood pressure measurements of two subjects in the hypertension study.

model with autoregressive errors (RegAR) with AR order 1. Note that RegAR can be viewed as special case of distributed lag models with L setting to 0.

Table 4.1 shows the posterior mean/MLE and 90% credible/confidence interval of both distributed lag and autoregressive coefficients. For patient 1, with respect to total treatment effect, both RegAR and the proposed BDLM-AR model find that compared to Dyazide, Losartan and Amlodipine will increase SBP. The estimated increases of Losartan by RegAR and BDLM-AR were 12.68 (90% CI: 8.34, 17.01) and 10.41 (90% CI: 5.49, 15.44) respectively. The estimated increases of Amlodipine were 8.62 (90% CI: 4.22, 13.01) and 7.07 (90% CI: 2.12, 11.87) respectively. From BDLM-AR model, we also find treatment difference between Dyazide and Losartan is quite substantial at the first 3 days, while the difference between Dyazide and Amlodipine reaches to the maximum when patient 1 took the medication and then keep a flat trend. For patient 2, with

		Patient 1		Patient 2	
		RegAR(1)	BDLM-AR(1)	RegAR(1)	BDLM-AR(1)
Losartan vs. HCTZ	Total effect	12.68 (8.34,17.01)	10.41 (5.49,15.44)	7.86 (3.78,11.94)	5.82 (1.75,10.24)
	Total carryover effect	-	3.62 (-1.19,9.04)	-	1.47 (-2.88,5.84)
	β_0	-	6.78 (1.88,11.52)	-	4.35 (0.09,8.78)
	β_1	-	2.95 (-0.76,7.29)	-	1.48 (-1.51,5.02)
	β_2	-	0.77 (-2.13,3.87)	-	0.74 (-1.26,3.81)
	β_3	-	-0.22 (-2.8,1.81)	-	-0.11 (-2.18,1.45)
	β_4	-	0.38 (-1.15,2.65)	-	-0.3 (-2.21,0.75)
	β_5	-	-0.09 (-1.58,1.15)	-	-0.12 (-1.3,0.65)
	β_6	-	-0.08 (-1.34,0.88)	-	-0.13 (-1.31,0.49)
	β_7	-	-0.08 (-1.06,0.62)	-	-0.09 (-0.97,0.37)
Amlodipine vs. HCTZ	Total effect	8.62 (4.22,13.01)	7.07 (2.12,11.87)	6.34 (2.23,10.46)	5.51 (0.91,10.34)
	Total carryover effect	-	4.29 (-0.9,11.55)	-	3.48 (-0.67,9.83)
	β_0	-	2.78 (-3.38,8.12)	-	2.03 (-2.59,6.28)
	β_1	-	0.87 (-3.8,4.85)	-	0.9 (-2.57,4.25)
	β_2	-	1.04 (-1.81,4.6)	-	1.17 (-0.97,4.47)
	β_3	-	1.2 (-0.81,4.86)	-	0.64 (-0.85,3.13)
	β_4	-	0.51 (-1.03,2.96)	-	0.32 (-0.87,2.06)
	β_5	-	0.43 (-0.74,2.57)	-	0.16 (-0.76,1.53)
	β_6	-	0.13 (-0.97,1.57)	-	0.15 (-0.58,1.37)
	β_7	-	0.11 (-0.71,1.27)	-	0.14 (-0.4,1.13)
	μ	113.3 (110.2,116.39)	114.71 (111.32,118.24)	107.17 (104.31,110.03)	107.93 (104.85,111.21)
	ϕ	0.18 (0,0.36)	0.2 (0,0.4)	0.28 (0.09,0.45)	0.27 (0.07,0.47)
	σ	8.31 (7.26,9.53)	8.34 (7.3,9.57)	7.3 (6.33,8.43)	7.1 (6.19,8.11)

Table 4.1: Posterior mean/maximum likelihood estimates of total effect, total carryover effect, lag coefficients and autoregressive coefficients for hypertension study, fitted by BDLM-AR and RegAR. 90% credible/confidence intervals are in brackets.

respect to total treatment effect, both methods also find that Losartan and Amlodipine will increase SBP compared to HCTZ. The estimated increases of Losartan by RegAR and BDLM-AR were 7.86 (90% CI: 3.78, 11.94) and 5.82 (90% CI: 1.75, 10.24) respectively. The estimated increases of Amlodipine were 6.34 (90% CI: 2.23, 10.46) and 5.51 (90% CI: 0.91, 10.34) respectively. All of the 90% credible/confidence intervals do not contain zero, suggesting a significant difference in treatment effect.

4.5 Discussion

In this chapter, we have proposed a novel method to analyze N-of-1 trials when multiple interventions were used. As an extension to the method in Chapter 2, the proposed method in this chapter also handles temporal correlation between observations and estimate carryover effect through a Bayesian distributed lag structure. To alleviate multicollinearity problem in the explanatory variables and add prior knowledge on the shape of distributed lag curve, we introduce a designed prior

precision matrix on each intervention to shrink the treatment effect to zero along the time.

DLMs have been used widely in econometric, advertising and environmental research area but have not been applied in N-of-1 trials. They have the advantage of decomposing the total treatment effects over a specific period of time. Many researches have been conducted on DLMs to estimate distributed lag coefficients including parametric method (Almon, 1965; Koyck, 1954), penalized regression splines (Zanobetti et al., 2000) and Bayesian framework (Welty et al., 2009). However, most of these studies focus on single continuous explanatory variable. In N-of-1 trials, using multiple interventions in different periods within patients are sometimes common, since there are several first-line medications available with different mechanisms of action. The proposed BDLM-AR method provides a direct and scientific way to compare interventions within the patient.

One assumption of the proposed model is that the interactions effects between interventions will not substantially affect the outcome. In N-of-1 trials, the interventions are assigned in predetermined order and each patient will repeatedly receive one intervention for several days. Switching treatment will not happened frequently for practical considerations. However, if the treatment period of the same intervention is much shorter than its elimination rate/half-life estimated by pharmacokinetic/pharmacodynamic (PK/PD) modeling, then the interaction term of different interventions may be added into the model due to the potential influence of synergistic or antagonistic effects from treatment combinations (Y.-H. Chen et al., 2019).

In analyzing hypertension study data, we estimated the treatment effect difference in three guideline-recommended medications. HCTZ (or Dyazide) is the best treatment for both patients in reducing blood pressure. These findings are consistent with patients' preference at the end of trial. The largest treatment effect difference between HCTZ and Losartan occurs on the same day of administration of medication while Amlodipine has consistently smaller treatment effect as HCTZ along the time.

BDLM-AR in this chapter is designed to analyze individual N-of-1 trial time series data with multiple interventions. A future direction is to combine it with the Bayesian hierarchical structure in Chapter 3. The multi-trial framework can provide treatment effect estimation at population level,

while adjusted for treatment heterogeneity, and potentially improve individual treatment effect estimation by borrowing strength from other similar trials. When two treatments are not compared directly in any single N-of-1 trial, but each of them compared directly to a third treatment in common or through a network of relative effect, another advantage of combining multiple N-of-1 trials with multiple intervention is that it can estimate the treatment effect between drugs with indirect comparison.

Chapter 5: Conclusion

N-of-1 trials have been poised to emerge as an important part of personalized medicine in health care practice. Most of the existing models for analyzing data from N-of-1 trials still regard them as a kind of conventional multiple-period crossover trials, thus data are usually aggregated to several time periods in the trial. However, with the advance of modern technology, such as wearable devices and mobile apps, the abundant data collected not only provide more information about treatment response, but also require different statistical methods to work with these granular data. In this dissertation, our main goal is to appropriately estimate the treatment effect in N-of-1 trials using densely collected time series type of data.

To achieve this purpose, in Chapter 2, we first proposed Bayesian DLM to analytically model autocorrelation between measurements and carryover effect. One typical issue of DLM is the problem of multicollinearity, which triggers unreliable coefficient estimates with large variances. To alleviate the adverse effects of multicollinearity, we use Bayesian regularization approach, that is, impose specific prior distributions on the model parameters. The advantage of using Bayesian model to analyze N-of-1 trial data is that the degree of regularization is data-driven through putting prior distribution on tuning parameters. This is especially useful since regular approach to choose tuning parameters such as cross validation might be problematic in time series data analysis. As demonstrated in simulation studies, our method outperforms other models in estimating total treatment effect, carryover effect and lag coefficient curve under most of simulation scenarios. The advantage of BDLM-AR in estimating lag coefficient curve can bring more insights in understanding the dynamic treatment effect within each participant. Analyzing data from two subjects in the light therapy study, we find two patients have distinct optimal treatment, indicating the existence

of heterogeneity of treatment effect.

In Chapter 3, we considered Bayesian hierarchical structure to combine data from multiple N-of-1 trials, focusing on inference of treatment effect and carryover effect at population level and between patients heterogeneity. In the simulation, we demonstrate the superiority of the proposed model for population level parameters estimation. Besides that, the advantage of partial pooling among patients also further improves single N-of-1 trial model in individual level parameters estimation. The Bayesian structure that allows to borrow strength from the N-of-1 trials of others is substantial for patients with only a small number of observations.

In Chapter 4, we extended BDLM-AR model from one treatment and one placebo (or active control) to multiple treatments and one placebo (or active control) scenario. By applying the proposed model to hypertension study data, we identify the optimal medications for each hypertensive patient among several available first-line treatments.

From N-of-1 trial design perspective, in individual N-of-1 trial, when the total number of observations is fixed, increasing treatment switching will reduce the severity of multicollinearity, which is encouraged. However, frequent treatment change may affect coherence in patients and cause additional difficulties in administration. This is the trade off that clinicians need to consider before the trial. Another design element in N-of-1 trial is washout period between different treatments. A washout period is theoretically important to eliminate lingering effects left from previous intervention. Instead of using physical washout period, our proposed methods provide an alternative to address the carryover effects analytically, which can be applied to N-of-1 trial either with or without washout period. When washout period is included, it is important to collect the measurements during this period in order to apply our proposed methods.

In the dissertation, all the proposed models discussed so far use binary predictors, but it can be easily extended to continuous predictors if one has more information on the interventions, e.g. dosage, and these methods can potentially be applicable to a wide range of problems. We hope this dissertation contributes to statistical methods for analyzing N-of-1 trials.

Bibliography

Almon, S. (1965). The distributed lag between capital appropriations and expenditures. *Econometrica: Journal of the Econometric Society*, 178–196.

Araujo, A., Julious, S., & Senn, S. (2016). Understanding variation in sets of n-of-1 trials. *PloS one*, *11*(12), 1–24.

Barr, C., Marois, M., Sim, I., Schmid, C. H., Wilsey, B., Ward, D., Duan, N., Hays, R. D., Selsky, J., Servadio, J., et al. (2015). The preempt study-evaluating smartphone-assisted n-of-1 trials in patients with chronic pain: Study protocol for a randomized controlled trial. *Trials*, *16*(1), 1–11.

Bass, F. M., & Clarke, D. G. (1972). Testing distributed lag models of advertising effect. *Journal of Marketing Research*, *9*(3), 298–308.

Berkey, C. S., Hoaglin, D. C., Mosteller, F., & Colditz, G. A. (1995). A random-effects regression model for meta-analysis. *Statistics in medicine*, *14*(4), 395–411.

Bernal, J. L., Cummins, S., & Gasparrini, A. (2017). Interrupted time series regression for the evaluation of public health interventions: A tutorial. *International journal of epidemiology*, *46*(1), 348–355.

Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of computational and graphical statistics*, *7*(4), 434–455.

Brown Jr, B. W. (1980). The crossover experiment for clinical trials. *Biometrics*, 69–79.

Chatellier, G., Day, M., Bobrie, G., & Menard, J. (1995). Feasibility study of n-of-1 trials with blood pressure self-monitoring in hypertension. *Hypertension*, *25*(2), 294–301.

Chen, X., & Chen, P. (2014). A comparison of four methods for the analysis of n-of-1 trials. *PloS one*, *9*(2).

Chen, Y.-H., Mukherjee, B., & Berrocal, V. J. (2019). Distributed lag interaction models with two pollutants. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *68*(1), 79–97.

Cheung, Y., Wood, D., Zhang, K., Ridenour, T., Derby, L., St Onge, T., Duan, N., Duer-Hefele, J., Davidson, K., Kronish, I., et al. (2020). Personal preferences for personalised trials among

patients with chronic diseases: An empirical bayesian analysis of a conjoint survey. *BMJ Open*, *10*(6).

Chib, S. (1993). Bayes regression with autoregressive errors: A gibbs sampling approach. *Journal of Econometrics*, *58*(3), 275–294.

Cortina, S., Somers, M., Rohan, J. M., & Drotar, D. (2013). Clinical effectiveness of comprehensive psychological intervention for nonadherence to medical treatment: A case series. *Journal of Pediatric Psychology*, *38*(6), 649–663.

Coxeter, P., Schluter, P., Eastwood, H., Nikles, C., & Glasziou, P. (2003). Valerian does not appear to reduce symptoms for patients with chronic insomnia in general practice using a series of randomised n-of-1 trials. *Complementary therapies in medicine*, *11*(4), 215–222.

Duan, N., Kravitz, R. L., & Schmid, C. H. (2013). Single-patient (n-of-1) trials: A pragmatic clinical decision methodology for patient-centered comparative effectiveness research. *Journal of clinical epidemiology*, *66*(8), S21–S28.

Efron, B., & Morris, C. (1977). Stein's paradox in statistics. *Scientific American*, *236*(5), 119–127.

El-Mikkawy, M. E. (2004). A fast algorithm for evaluating nth order tri-diagonal determinants. *Journal of computational and applied mathematics*, *166*(2), 581–584.

FDA. (2013). Paving the way for personalized medicine: Fdas role in a new era of medical product development. *Silver Spring, MD: US Food and Drug Administration*.

Gabler, N. B., Duan, N., Vohra, S., & Kravitz, R. L. (2011). N-of-1 trials in the medical literature: A systematic review. *Medical care*, 761–768.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. CRC press.

Gelman, A., Roberts, G. O., Gilks, W. R., et al. (1996). Efficient metropolis jumping rules. *Bayesian statistics*, *5*(599-608), 42.

Gelman, A., Rubin, D. B. et al. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, *7*(4), 457–472.

Glynn, L. G., Hayes, P. S., Casey, M., Glynn, F., Alvarez-Iglesias, A., Newell, J., ÓLaighin, G., Heaney, D., & Murphy, A. W. (2013). Smart move-a smartphone-based intervention to promote physical activity in primary care: Study protocol for a randomized controlled trial. *Trials*, *14*(1), 1–7.

Greenfield, S., Kravitz, R., Duan, N., & Kaplan, S. H. (2007). Heterogeneity of treatment effects: Implications for guidelines, payment, and quality assessment. *The American journal of medicine*, *120*(4), S3–S9.

Grizzle, J. E. (1965). The two-period change-over design and its use in clinical trials. *Biometrics*, 467–480.

Guyatt, G., Sackett, D., Adachi, J., Roberts, R., Chong, J., Rosenbloom, D., & Keller, J. (1988). A clinician's guide for conducting randomized trials in individual patients. *CMAJ: Canadian Medical Association Journal*, *139*(6), 497.

Guyatt, G., Sackett, D., Taylor, D. W., Ghong, J., Roberts, R., & Pugsley, S. (1986). Determining optimal therapyrandomized trials in individual patients. *New England Journal of Medicine*, *314*(14), 889–892.

Guyatt, G. H., Heyting, A., Jaeschke, R., Keller, J., Adachi, J. D., & Roberts, R. S. (1990). N of 1 randomized trials for investigating new drugs. *Controlled clinical trials*, *11*(2), 88–100.

Hamburg, M. A., & Collins, F. S. (2010). The path to personalized medicine. *New England Journal of Medicine*, *363*(4), 301–304.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55–67.

Johnson, J. A., Garland, S. N., Carlson, L. E., Savard, J., Simpson, J. S. A., Ancoli-Israel, S., & Campbell, T. S. (2018). Bright light therapy improves cancer-related fatigue in cancer survivors: A randomized controlled trial. *Journal of Cancer Survivorship*, *12*(2), 206–215.

Jones, A. P., Riley, R. D., Williamson, P. R., & Whitehead, A. (2009). Meta-analysis of individual patient data versus aggregate data from longitudinal clinical trials. *Clinical Trials*, *6*(1), 16–27.

Jones, B., & Kenward, M. G. (1989). *Design and analysis of cross-over trials*. Chapman; Hall/CRC.

Judge, G. G. (1982). Introduction to the theory and practice of econometrics.

Katahira, K. (2016). How hierarchical models improve point estimates of model parameters at the individual level. *Journal of Mathematical Psychology*, *73*, 37–58.

Kent, D. M., Rothwell, P. M., Ioannidis, J. P., Altman, D. G., & Hayward, R. A. (2010). Assessing and reporting heterogeneity in treatment effects in clinical trials: A proposal. *Trials*, *11*(1), 85.

Klein, L. R. (1962). Introduction to econometrics.

Koyck, L. M. (1954). *Distributed lags and investment analysis* (Vol. 4). North-Holland Publishing Company.

Kravitz, R. L., Duan, N., & Braslow, J. (2004). Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages. *The Milbank Quarterly*, *82*(4), 661–687.

Kravitz, R. L., Paterniti, D. A., Hay, M. C., Subramanian, S., Dean, D. E., Weisner, T., Vohra, S., & Duan, N. (2009). Marketing therapeutic precision: Potential facilitators and barriers to adoption of n-of-1 trials. *Contemporary clinical trials*, *30*(5), 436–445.

Kravitz, R. L., Schmid, C. H., Marois, M., Wilsey, B., Ward, D., Hays, R. D., Duan, N., Wang, Y., MacDonald, S., Jerant, A., et al. (2018). Effect of mobile device–supported single-patient multi-crossover trials on treatment of chronic musculoskeletal pain: A randomized clinical trial. *JAMA internal medicine*, *178*(10), 1368–1377.

Kravitz, R., Duan, N., Eslick, I., Gabler, N., Kaplan, H., et al. (2014). Design and implementation of n-of-1 trials: A users guide. *Agency for healthcare research and quality, US Department of Health and Human Services*.

Kronish, I. M., Cheung, Y. K., Julian, J., Parsons, F., Lee, J., Yoon, S., Valdimarsdottir, H., Green, P., Suls, J., Hershman, D. L., et al. (2020). Clinical usefulness of bright white light therapy for depressive symptoms in cancer survivors: Results from a series of personalized (n-of-1) trials. *Healthcare*, *8*(1), 10.

Kronish, I. M., Cheung, Y. K., Shimbo, D., Julian, J., Gallagher, B., Parsons, F., & Davidson, K. W. (2019). Increasing the precision of hypertension treatment through personalized trials: A pilot study. *Journal of general internal medicine*, *34*(6), 839–845.

Kronish, I. M., Hampsey, M., Falzon, L., Konrad, B., & Davidson, K. W. (2018). Personalized (n-of-1) trials for depression: A systematic review. *Journal of clinical psychopharmacology*, *38*(3), 218.

Lane-Brown, A., & Tate, R. (2010). Evaluation of an intervention for apathy after traumatic brain injury: A multiple-baseline, single-case experimental design. *The Journal of head trauma rehabilitation*, *25*(6), 459–469.

Lillie, E. O., Patay, B., Diamant, J., Issell, B., Topol, E. J., & Schork, N. J. (2011). The n-of-1 clinical trial: The ultimate strategy for individualizing medicine? *Personalized medicine*, *8*(2), 161–173.

Lindley, D. V., & Smith, A. F. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society: Series B (Methodological)*, *34*(1), 1–18.

Liu, C., Wang, H., & Liu, K. (2016). Meta-analysis of the efficacy of proton pump inhibitors for the symptoms of laryngopharyngeal reflux. *Brazilian Journal of Medical and Biological Research*, *49*.

Liviatan, N. (1963). Consistent estimation of distributed lags. *International Economic Review*, *4*(1), 44–52.

Ljung, G. M., & Box, G. E. (1978). On a measure of lack of fit in time series models. *Biometrika*, *65*(2), 297–303.

Luxton, D. D., McCann, R. A., Bush, N. E., Mishkind, M. C., & Reger, G. M. (2011). Mhealth for mental health: Integrating smartphone technology in behavioral healthcare. *Professional Psychology: Research and Practice*, *42*(6), 505.

Mahon, J. L., Laupacis, A., Hodder, R. V., Mc Kim, D. A., Paterson, N. A., Wood, T. E., & Donner, A. (1999). Theophylline for irreversible chronic airflow limitation: A randomized study comparing n of 1 trials to standard practice. *Chest*, *115*(1), 38–48.

Mancia, G., Parodi, A., Merlino, L., & Corrao, G. (2011). Heterogeneity in antihypertensive treatment discontinuation between drugs belonging to the same class. *Journal of hypertension*, *29*(5), 1012–1018.

Martin, R. T., & Whyte, J. (2007). The effects of methylphenidate on command following and yes/no communication in persons with severe disorders of consciousness: A meta-analysis of n-of-1 studies. *American journal of physical medicine & rehabilitation*, *86*(8), 613–620.

Materson, B. J. (2007). Variability in response to antihypertensive drugs. *The American journal of medicine*, *120*(4), S10–S20.

Mirza, R., Punja, S., Vohra, S., & Guyatt, G. (2017). The history and development of n-of-1 trials. *Journal of the Royal Society of Medicine*, *110*(8), 330–340.

Nilsen, W., Kumar, S., Shar, A., Varoquiers, C., Wiley, T., Riley, W. T., Pavel, M., & Atienza, A. A. (2012). Advancing the science of mhealth. *Journal of health communication*, *17*(sup1), 5–10.

Patel, A., Jaeschke, R., Guyatt, G. H., Keller, J. L., & Newhouse, M. T. (1991). Clinical usefulness of n-of-1 randomized controlled trials in patients with nonreversible chronic airflow limitation. *Am Rev Respir Dis*, *144*(4), 962–964.

Peng, R. D., Dominici, F., & Welty, L. J. (2009). A bayesian hierarchical distributed lag model for estimating the time course of risk of hospitalization associated with particulate matter air pollution. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *58*(1), 3–24.

Punja, S., Bukutu, C., Shamseer, L., Sampson, M., Hartling, L., Urichuk, L., & Vohra, S. (2016). N-of-1 trials are a tapestry of heterogeneity. *Journal of Clinical Epidemiology*, *76*, 47–56.

Robinson, G. K. et al. (1991). That blup is a good thing: The estimation of random effects. *Statistical science*, *6*(1), 15–32.

Rothwell, P. M. (2005). Subgroup analysis in randomised controlled trials: Importance, indications, and interpretation. *The Lancet*, *365*(9454), 176–186.

Samuel, J. P., Tyson, J. E., Green, C., Bell, C. S., Pedroza, C., Molony, D., & Samuels, J. (2019). Treating hypertension in children with n-of-1 trials. *Pediatrics*, *143*(4).

Schluter, P., & Ware, R. (2005). Single patient (n-of-1) trials with binary treatment preference. *Statistics in medicine*, *24*(17), 2625–2636.

Schork, N. J. (2015). Personalized medicine: Time for one-person trials. *Nature*, *520*(7549), 609–611.

Selby, J. V., Beal, A. C., & Frank, L. (2012). The patient-centered outcomes research institute (pcori) national priorities for research and initial research agenda. *Jama*, *307*(15), 1583–1584.

Senn, S. (1992). Is the simple carry-overmodel useful? *Statistics in Medicine*, *11*(6), 715–726.

Senn, S. S. (2002). *Cross-over trials in clinical research* (Vol. 5). John Wiley & Sons.

Sox, H. C., & Greenfield, S. (2009). Comparative effectiveness research: A report from the institute of medicine. *Annals of internal medicine*, *151*(3), 203–205.

Stunnenberg, B. C., Raaphorst, J., Groenewoud, H. M., Statland, J. M., Griggs, R. C., Woertman, W., Stegeman, D. F., Timmermans, J., Trivedi, J., Matthews, E., et al. (2018). Effect of mexiletine on muscle stiffness in patients with nondystrophic myotonia evaluated using aggregated n-of-1 trials. *Jama*, *320*(22), 2344–2353.

Tibshirani, R., & Wang, P. (2008). Spatial smoothing and hot spot detection for cgh data using the fused lasso. *Biostatistics*, *9*(1), 18–29.

Topol, E. J. (2010). Transforming medicine via digital innovation. *Science translational medicine*, *2*(16), 16cm4–16cm4.

Tsapas, A., & Matthews, D. R. (2009). Using n-of-1 trials in evidence-based clinical practice. *Jama*, *301*(10), 1022–1023.

Vieira, R., McDonald, S., Araújo-Soares, V., Sniehotta, F. F., & Henderson, R. (2017). Dynamic modelling of n-of-1 data: Powerful and flexible data analytics applied to individualised studies. *Health Psychology Review*, *11*(3), 222–234.

Wang, Y., Xue, H., Huang, Y., Huang, L., & Zhang, D. (2017). A systematic review of application and effectiveness of mhealth interventions for obesity and diabetes treatment and self-management. *Advances in Nutrition*, *8*(3), 449–462.

Welty, L. J., Peng, R., Zeger, S., & Dominici, F. (2009). Bayesian distributed lag models: Estimating effects of particulate matter air pollution on daily mortality. *Biometrics*, *65*(1), 282–291.

Whitney, R. L., Ward, D. H., Marois, M. T., Schmid, C. H., Sim, I., & Kravitz, R. L. (2018). Patient perceptions of their own data in mhealth technology–enabled n-of-1 trials for chronic pain: Qualitative study. *JMIR mHealth and uHealth*, *6*(10), e10291.

Zanobetti, A., & Schwartz, J. (2008). Mortality displacement in the association of ozone with mortality: An analysis of 48 cities in the united states. *American Journal of Respiratory and Critical Care Medicine*, *177*(2), 184–189.

Zanobetti, A., Wand, M., Schwartz, J., & Ryan, L. (2000). Generalized additive distributed lag models: Quantifying mortality displacement. *Biostatistics*, *1*(3), 279–292.

Zucker, D. R., Ruthazer, R., & Schmid, C. H. (2010). Individual (n-of-1) trials can be combined to give population comparative treatment effect estimates: Methodologic considerations. *Journal of clinical epidemiology*, *63*(12), 1312–1323.

Zucker, D., Schmid, C., McIntosh, M., D'agostino, R., Selker, H., & Lau, J. (1997). Combining single patient (n-of-1) trials to estimate population treatment effects and to evaluate individual patient responses to treatment. *Journal of clinical epidemiology*, *50*(4), 401–410.

Appendix A: Appendices to Chapter 2

A.1 Relationship between prior precision matrix on lag coefficient and fused ridge regression

Let the notation "$\pi(A|\cdot)$" represents conditional distribution of A given all other variables in the model. The posterior distribution of $\tilde{\boldsymbol{\beta}}$ is:

$$\begin{aligned}
\pi(\tilde{\boldsymbol{\beta}} \mid \cdot) &\propto \pi(\boldsymbol{Y}|\tilde{\boldsymbol{X}}, \tilde{\boldsymbol{\beta}}, \sigma^2, \boldsymbol{\phi}, \boldsymbol{\gamma})\pi(\tilde{\boldsymbol{\beta}}|\sigma^2, \boldsymbol{\gamma}) \\
&\propto \exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{Y}^* - \tilde{\boldsymbol{X}}^*\tilde{\boldsymbol{\beta}})'(\boldsymbol{Y}^* - \tilde{\boldsymbol{X}}^*\tilde{\boldsymbol{\beta}})\right] \exp\left[-\frac{1}{2\sigma^2}\tilde{\boldsymbol{\beta}}'\tilde{\boldsymbol{\Omega}}(\boldsymbol{\gamma})\tilde{\boldsymbol{\beta}}\right] \\
&\propto \exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{Y}^* - \tilde{\boldsymbol{X}}^*\boldsymbol{\beta})'(\boldsymbol{Y}^* - \tilde{\boldsymbol{X}}^*\boldsymbol{\beta}) - \frac{1}{2\sigma^2}c_0\mu^2 - \frac{1}{2\sigma^2}\sum_{l=0}^{L}\lambda_l\beta_l^2 - \frac{1}{2\sigma^2}\sum_{l=0}^{L}\lambda_l^*(\beta_l - \beta_{l+1})^2\right]
\end{aligned}$$

We now compute the MAP estimate of $\tilde{\boldsymbol{\beta}}$:

$$\begin{aligned}
\hat{\beta}_{\text{MAP}} &= \underset{\mu,\beta}{\text{argmax}} \exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{Y}^* - \tilde{\boldsymbol{X}}^*\tilde{\boldsymbol{\beta}})'(\boldsymbol{Y}^* - \tilde{\boldsymbol{X}}^*\tilde{\boldsymbol{\beta}}) - \frac{1}{2\sigma^2}c_0\mu^2 - \frac{1}{2\sigma^2}\sum_{l=0}^{L}\lambda_l\beta_l^2 - \frac{1}{2\sigma^2}\sum_{l=0}^{L}\lambda_l^*(\beta_l - \beta_{l+1})^2\right] \\
&= \underset{\mu,\beta}{\text{argmin}}(\boldsymbol{Y}^* - \tilde{\boldsymbol{X}}^*\tilde{\boldsymbol{\beta}})^T(\boldsymbol{Y}^* - \tilde{\boldsymbol{X}}^*\tilde{\boldsymbol{\beta}}) + c_0\mu^2 + \sum_{l=0}^{L}\lambda_l\beta_l^2 + \sum_{l=0}^{L}\lambda_l^*(\beta_l - \beta_{l+1})^2
\end{aligned}$$

It is important for the prior distribution of $\tilde{\boldsymbol{\beta}}$ to condition on σ^2. Without this, the ridge and smoothness penalties will not only depend on λ_l and λ_l^*, but depend on model variance and prior covariance matrix on $\tilde{\boldsymbol{\beta}}$ as well. Conditioning on σ^2 provides an explicit way to control the penalty tuning parameters.

A.2 Derivation of full conditional posterior distributions of parameters

Given the likelihood function and prior distribution in Section 2.2, full conditional posterior distributions of each variable are as follows:

(1) Posterior of $\tilde{\boldsymbol{\beta}}$

$$\begin{aligned}\pi(\tilde{\boldsymbol{\beta}}|\cdot) &\propto \pi(\boldsymbol{Y}|\tilde{\boldsymbol{X}}, \tilde{\boldsymbol{\beta}}, \sigma^2, \boldsymbol{\phi}, \boldsymbol{\gamma})\pi(\boldsymbol{\beta}|\sigma^2, \boldsymbol{\gamma}) \\ &\propto \exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{Y}^* - \tilde{\boldsymbol{X}}^*\tilde{\boldsymbol{\beta}})'(\boldsymbol{Y}^* - \tilde{\boldsymbol{X}}^*\tilde{\boldsymbol{\beta}})\right] \exp\left[-\frac{1}{2\sigma^2}\tilde{\boldsymbol{\beta}}'\tilde{\boldsymbol{\Omega}}(\boldsymbol{\gamma})\tilde{\boldsymbol{\beta}}\right]\end{aligned}$$

where $\boldsymbol{Y}^*$ and $\tilde{\boldsymbol{X}}^*$ are transformed $\boldsymbol{Y}$ and $\tilde{\boldsymbol{X}}$ as described in equation (2.4).
Denote $\hat{\boldsymbol{\beta}} = \left[\tilde{\boldsymbol{X}}^{*'}\tilde{\boldsymbol{X}}^* + \tilde{\boldsymbol{\Omega}}(\boldsymbol{\gamma})\right]^{-1}\tilde{\boldsymbol{X}}^{*'}\boldsymbol{Y}^*$ and using the equation

$$\begin{aligned}(\boldsymbol{Y}^* - \tilde{\boldsymbol{X}}^*\tilde{\boldsymbol{\beta}})'(\boldsymbol{Y}^* - \tilde{\boldsymbol{X}}^*\tilde{\boldsymbol{\beta}}) + \tilde{\boldsymbol{\beta}}'\tilde{\boldsymbol{\Omega}}(\boldsymbol{\gamma})\tilde{\boldsymbol{\beta}} = (\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}})'\left[\tilde{\boldsymbol{X}}^{*'}\tilde{\boldsymbol{X}}^* + \tilde{\boldsymbol{\Omega}}(\boldsymbol{\gamma})\right](\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}) \\ + \boldsymbol{Y}^{*'}\boldsymbol{Y}^* - \hat{\boldsymbol{\beta}}'\left[\tilde{\boldsymbol{X}}^{*'}\tilde{\boldsymbol{X}}^* + \tilde{\boldsymbol{\Omega}}(\boldsymbol{\gamma})\right]\hat{\boldsymbol{\beta}}\end{aligned}$$

Then the full conditional posterior distribution of $\tilde{\boldsymbol{\beta}}$ is:

$$\begin{aligned}\pi(\tilde{\boldsymbol{\beta}}|\cdot) &\propto \exp\left[(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}})'\left[\tilde{\boldsymbol{X}}^{*'}\tilde{\boldsymbol{X}}^* + \tilde{\boldsymbol{\Omega}}(\boldsymbol{\gamma})\right](\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}})\right] \\ \tilde{\boldsymbol{\beta}} \mid \cdot &\sim N_{L+1}\left\{[\tilde{\boldsymbol{X}}^{*'}\tilde{\boldsymbol{X}}^* + \tilde{\boldsymbol{\Omega}}(\boldsymbol{\gamma})]^{-1}\tilde{\boldsymbol{X}}^{*'}\boldsymbol{Y}^*, \sigma^2[\tilde{\boldsymbol{X}}^{*'}\tilde{\boldsymbol{X}}^* + \tilde{\boldsymbol{\Omega}}(\boldsymbol{\gamma})]^{-1}\right\}\end{aligned}$$

(2) Posterior of σ^2

$$\begin{aligned}\pi(\sigma^2|\cdot) &\propto \pi(\boldsymbol{Y}|\tilde{\boldsymbol{X}}, \tilde{\boldsymbol{\beta}}, \sigma^2, \boldsymbol{\phi}, \boldsymbol{\gamma})\pi(\tilde{\boldsymbol{\beta}}|\sigma^2, \boldsymbol{\gamma})\pi(\sigma^2) \\ &\propto \left(\sigma^2\right)^{-\frac{n-p}{2}} \exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{Y}^* - \tilde{\boldsymbol{X}}^*\tilde{\boldsymbol{\beta}})'(\boldsymbol{Y}^* - \tilde{\boldsymbol{X}}^*\tilde{\boldsymbol{\beta}})\right] |\sigma^2\tilde{\boldsymbol{\Omega}}^{-1}(\boldsymbol{\gamma})|^{-\frac{1}{2}} \exp\left[-\frac{1}{2\sigma^2}\tilde{\boldsymbol{\beta}}'\tilde{\boldsymbol{\Omega}}(\boldsymbol{\gamma})\tilde{\boldsymbol{\beta}}\right] \frac{1}{\sigma^2} \\ &\propto \left(\sigma^2\right)^{-\frac{n-p+L+3}{2}} \exp\left\{-\frac{1}{2\sigma^2}\left[(\boldsymbol{Y}^* - \tilde{\boldsymbol{X}}^*\tilde{\boldsymbol{\beta}})'(\boldsymbol{Y}^* - \tilde{\boldsymbol{X}}^*\tilde{\boldsymbol{\beta}}) + \tilde{\boldsymbol{\beta}}'\tilde{\boldsymbol{\Omega}}(\boldsymbol{\gamma})\tilde{\boldsymbol{\beta}}\right]\right\}\end{aligned}$$

where L is the maximum number of lags, and p is the order of autoregressive error.

Then the full conditional posterior distribution of σ^2 is:

$$\sigma^2 \mid \cdot \sim \text{Inv-Gamma}\left[\frac{n-p+L+1}{2}, \frac{(\boldsymbol{Y}^* - \tilde{\boldsymbol{X}}^*\tilde{\boldsymbol{\beta}})'(\boldsymbol{Y}^* - \tilde{\boldsymbol{X}}^*\tilde{\boldsymbol{\beta}}) + \tilde{\boldsymbol{\beta}}'\tilde{\boldsymbol{\Omega}}(\boldsymbol{\gamma})\tilde{\boldsymbol{\beta}}}{2}\right]$$

(3) Posterior of $\boldsymbol{\phi}$

Given $\boldsymbol{Y}$, $\tilde{\boldsymbol{X}}$, $\tilde{\boldsymbol{\beta}}$ and σ^2, the conditional joint distribution of $\epsilon^*_{p+1}, \epsilon^*_{p+2}, ..., \epsilon^*_n$ can be regarded as a sequence of one-sided conditional distributions:

$$\pi(\epsilon^*_{p+1}, \epsilon^*_{p+2}, ..., \epsilon^*_n) = \pi(\epsilon^*_{p+1}|\epsilon^*_p, ..., \epsilon^*_1) \cdots \pi(\epsilon^*_n|\epsilon^*_{n-1}, ..., \epsilon^*_{n-p})$$

where $\epsilon^*_{p+1}, \epsilon^*_{p+2}, ..., \epsilon^*_n$ are described in equation (2.13) of Section 2.3.

$$\begin{aligned}
\pi(\boldsymbol{\phi}|\cdot) &\propto \pi(\epsilon^*_{p+1}, \epsilon^*_{p+2}, ..., \epsilon^*_n|\boldsymbol{\phi}, \sigma^2)\pi(\boldsymbol{\phi}|\sigma^2_{\boldsymbol{\phi}}) \\
&\propto \prod_{t=p+1}^{n} \pi(\epsilon^*_t|\epsilon^*_{t-1}, ..., \epsilon^*_{t-p}, \boldsymbol{\phi}, \sigma^2)\pi(\boldsymbol{\phi}|\sigma^2_{\boldsymbol{\phi}}) \\
&\propto \exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{\epsilon}^* - \boldsymbol{E}^*\boldsymbol{\phi})'(\boldsymbol{\epsilon}^* - \boldsymbol{E}^*\boldsymbol{\phi})\right] \exp\left(-\frac{1}{2\sigma^2_{\boldsymbol{\phi}}}\boldsymbol{\phi}'\boldsymbol{\phi}\right) \mathbf{1}_{S_\phi}(\boldsymbol{\phi}) \\
&\propto \exp\left[(\boldsymbol{\phi} - \hat{\boldsymbol{\phi}})'\left(\sigma^{-2}\boldsymbol{E}^{*'}\boldsymbol{E}^* + \sigma^{-2}_{\boldsymbol{\phi}}\mathbf{I}\right)(\boldsymbol{\phi} - \hat{\boldsymbol{\phi}})\right] \mathbf{1}_{S_\phi}(\boldsymbol{\phi})
\end{aligned}$$

where $\hat{\boldsymbol{\phi}} = \left(\sigma^{-2}\boldsymbol{E}^{*'}\boldsymbol{E}^* + \sigma^{-2}_{\boldsymbol{\phi}}\mathbf{I}\right)^{-1} \sigma^{-2}\boldsymbol{E}^{*'}\boldsymbol{\epsilon}^*$, $\sigma^2_{\boldsymbol{\phi}}$ is known and set to be 200.

Then the full conditional posterior distribution of $\boldsymbol{\phi}$ is:

$$\boldsymbol{\phi} \mid \cdot \sim N_p\left[\left(\sigma^{-2}\boldsymbol{E}^{*'}\boldsymbol{E}^* + \sigma^{-2}_{\boldsymbol{\phi}}\mathbf{I}\right)^{-1} \sigma^{-2}\boldsymbol{E}^{*'}\boldsymbol{\epsilon}^*, \left(\sigma^{-2}\boldsymbol{E}^{*'}\boldsymbol{E}^* + \sigma^{-2}_{\boldsymbol{\phi}}\mathbf{I}\right)^{-1}\right] \mathbf{1}_{S_\phi}(\boldsymbol{\phi})$$

(4) Posterior of $\boldsymbol{\gamma}$

$$
\begin{aligned}
\pi(\boldsymbol{\gamma} \mid \cdot) &\propto \pi(\boldsymbol{Y}|\tilde{\boldsymbol{X}}, \tilde{\boldsymbol{\beta}}, \sigma^2, \boldsymbol{\phi}, \boldsymbol{\gamma})\pi(\boldsymbol{\gamma} \mid \tilde{\boldsymbol{\beta}}, \sigma^2, \boldsymbol{\phi}) \\
&\propto \pi(\boldsymbol{Y}|\tilde{\boldsymbol{X}}, \tilde{\boldsymbol{\beta}}, \sigma^2, \boldsymbol{\phi}, \boldsymbol{\gamma})\pi(\tilde{\boldsymbol{\beta}} \mid \sigma^2, \boldsymbol{\gamma})\pi(\boldsymbol{\gamma}) \\
&\propto |\sigma^{-2}\tilde{\boldsymbol{\Omega}}(\boldsymbol{\gamma})|^{\frac{1}{2}} \exp\left[-\frac{1}{2\sigma^2}\tilde{\boldsymbol{\beta}}'\tilde{\boldsymbol{\Omega}}(\boldsymbol{\gamma})\tilde{\boldsymbol{\beta}}\right] \exp(-\gamma_1 - \gamma_2)\mathbf{1}_{S_\gamma}(\gamma_1, \gamma_2)
\end{aligned}
$$

Appendix B: Appendices to Chapter 3

B.1 Derivation of full conditional posterior distribution of parameters

Let the notation "$\pi(A|\cdot)$" represents conditional distribution of A given all other variables in the model.

Given the likelihood function and prior distribution in Section 3.2, full conditional posterior distributions of each variable are as follows:

(1) Posterior of $\tilde{\boldsymbol{\beta}}_i$

$$\begin{aligned}\pi(\tilde{\boldsymbol{\beta}}_i|\cdot) &\propto \pi(\boldsymbol{Y}_i|\tilde{\boldsymbol{X}}_i, \tilde{\boldsymbol{\beta}}_i, \tilde{\boldsymbol{\theta}}, \sigma^2, \boldsymbol{\phi}_i, \boldsymbol{\gamma})\pi(\tilde{\boldsymbol{\beta}}_i|\tilde{\boldsymbol{\theta}}, \sigma^2, \boldsymbol{\gamma}) \\ &\propto \exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{Y}_i^* - \tilde{\boldsymbol{X}}_i^*\tilde{\boldsymbol{\beta}}_i)'(\boldsymbol{Y}_i^* - \tilde{\boldsymbol{X}}_i^*\tilde{\boldsymbol{\beta}}_i)\right] \exp\left[-\frac{1}{2\sigma^2}(\tilde{\boldsymbol{\beta}}_i - \tilde{\boldsymbol{\theta}})'\tilde{\boldsymbol{\Omega}}_\beta(\boldsymbol{\gamma})(\tilde{\boldsymbol{\beta}}_i - \tilde{\boldsymbol{\theta}})\right]\end{aligned}$$

where $\boldsymbol{Y}_i^*$ and $\tilde{\boldsymbol{X}}_i^*$ are transformed $\boldsymbol{Y}_i$ and $\tilde{\boldsymbol{X}}_i$ as described in equation (3.4).
Denote $\hat{\boldsymbol{\beta}}_i = \left[\tilde{\boldsymbol{X}}_i^{*'}\tilde{\boldsymbol{X}}_i^* + \tilde{\boldsymbol{\Omega}}_\beta(\boldsymbol{\gamma})\right]^{-1}\left[\tilde{\boldsymbol{X}}_i^{*'}\boldsymbol{Y}_i^* + \tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\Omega}}_\beta(\boldsymbol{\gamma})\right]$ and using the equation

$$\begin{aligned}(\boldsymbol{Y}_i^* - \tilde{\boldsymbol{X}}_i^*\tilde{\boldsymbol{\beta}}_i)'(\boldsymbol{Y}_i^* - \tilde{\boldsymbol{X}}_i^*\tilde{\boldsymbol{\beta}}_i) + (\tilde{\boldsymbol{\beta}}_i - \tilde{\boldsymbol{\theta}})'\tilde{\boldsymbol{\Omega}}_\beta(\boldsymbol{\gamma})(\tilde{\boldsymbol{\beta}}_i - \tilde{\boldsymbol{\theta}}) &= (\tilde{\boldsymbol{\beta}}_i - \hat{\boldsymbol{\beta}}_i)'\left[\tilde{\boldsymbol{X}}_i^{*'}\tilde{\boldsymbol{X}}_i^* + \tilde{\boldsymbol{\Omega}}_\beta(\boldsymbol{\gamma})\right](\tilde{\boldsymbol{\beta}}_i - \hat{\boldsymbol{\beta}}_i) \\ &\quad + \boldsymbol{Y}_i^{*'}\boldsymbol{Y}_i^* - \hat{\boldsymbol{\beta}}_i'\left[\tilde{\boldsymbol{X}}_i^{*'}\tilde{\boldsymbol{X}}_i^* + \tilde{\boldsymbol{\Omega}}_\beta(\boldsymbol{\gamma})\right]\hat{\boldsymbol{\beta}}_i + \tilde{\theta}'\tilde{\boldsymbol{\Omega}}_\beta(\boldsymbol{\gamma})\tilde{\theta}\end{aligned}$$

Then the full conditional posterior distribution of $\tilde{\boldsymbol{\beta}}$ is:

$$\pi(\tilde{\boldsymbol{\beta}}_i|\cdot) \propto \exp\left\{-\frac{1}{2\sigma^2}(\tilde{\boldsymbol{\beta}}_i - \hat{\boldsymbol{\beta}}_i)'\left[\tilde{\boldsymbol{X}}^{*'}\tilde{\boldsymbol{X}}^* + \tilde{\boldsymbol{\Omega}}_\beta(\boldsymbol{\gamma})\right](\tilde{\boldsymbol{\beta}}_i - \hat{\boldsymbol{\beta}}_i)\right\}$$

$$\tilde{\boldsymbol{\beta}}_i \mid \cdot \sim N_{L+1}\left\{\left[\tilde{\boldsymbol{X}}_i^{*'}\tilde{\boldsymbol{X}}_i^* + \tilde{\boldsymbol{\Omega}}(\boldsymbol{\gamma})\right]^{-1}\left[\tilde{\boldsymbol{X}}_i^{*'}\boldsymbol{Y}_i^* + \tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\Omega}}_\beta(\boldsymbol{\gamma})\right], \sigma^2\left[\tilde{\boldsymbol{X}}_i^{*'}\tilde{\boldsymbol{X}}_i^* + \tilde{\boldsymbol{\Omega}}_\beta(\boldsymbol{\gamma})\right]^{-1}\right\}$$

(2) Posterior of $\tilde{\boldsymbol{\theta}}$

$$
\begin{aligned}
\pi(\tilde{\boldsymbol{\theta}} \mid \cdot) &\propto \prod_{i=1}^{S} \pi(\boldsymbol{Y}_i|\tilde{\boldsymbol{X}}_i, \tilde{\boldsymbol{\beta}}_i, \tilde{\boldsymbol{\theta}}, \sigma^2, \boldsymbol{\phi}_i, \boldsymbol{\gamma}) \pi(\tilde{\boldsymbol{\theta}} \mid \boldsymbol{\beta}_i, \sigma^2, \boldsymbol{\phi}_i, \boldsymbol{\gamma}) \\
&\propto \pi(\tilde{\boldsymbol{\theta}}) \prod_{i=1}^{S} \pi(\boldsymbol{Y}_i|\tilde{\boldsymbol{X}}_i, \tilde{\boldsymbol{\beta}}_i, \tilde{\boldsymbol{\theta}}, \sigma^2, \boldsymbol{\phi}_i, \boldsymbol{\gamma}) \pi(\tilde{\boldsymbol{\beta}}_i \mid \tilde{\boldsymbol{\theta}}, \sigma^2, \boldsymbol{\gamma}) \\
&\propto \exp\left[-\frac{1}{2\sigma_\theta^2} \tilde{\boldsymbol{\theta}}' \tilde{\boldsymbol{\Omega}}_\theta(\boldsymbol{\tau}) \tilde{\boldsymbol{\theta}}\right] \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{S} (\tilde{\boldsymbol{\beta}}_i - \tilde{\boldsymbol{\theta}})' \tilde{\boldsymbol{\Omega}}_\beta(\boldsymbol{\gamma}) (\tilde{\boldsymbol{\beta}}_i - \tilde{\boldsymbol{\theta}})\right] \\
&\propto \exp\left[-\frac{1}{2\sigma_\theta^2} \tilde{\boldsymbol{\theta}}' \tilde{\boldsymbol{\Omega}}_\theta(\boldsymbol{\tau}) \tilde{\boldsymbol{\theta}}\right] \exp\left[-\frac{S}{2\sigma^2} (\tilde{\boldsymbol{\theta}} - \frac{\sum_{i=1}^{S} \boldsymbol{\beta}_i}{S})' \tilde{\boldsymbol{\Omega}}_\beta(\boldsymbol{\gamma}) (\tilde{\boldsymbol{\theta}} - \frac{\sum_{i=1}^{S} \boldsymbol{\beta}_i}{S})\right] \\
&\propto \exp\left[-\frac{1}{2} (\tilde{\boldsymbol{\theta}} - \boldsymbol{V}^{-1}\boldsymbol{m})' \boldsymbol{V} (\tilde{\boldsymbol{\theta}} - \boldsymbol{V}^{-1}\boldsymbol{m})\right]
\end{aligned}
$$

where $\boldsymbol{V} = \frac{\tilde{\boldsymbol{\Omega}}_\theta(\boldsymbol{\tau})}{\sigma_\theta^2} + \frac{\tilde{\boldsymbol{\Omega}}_\beta(\boldsymbol{\gamma})}{\sigma^2/S}$, $\boldsymbol{m} = \frac{\tilde{\boldsymbol{\Omega}}_\beta(\boldsymbol{\gamma})}{\sigma^2} \sum_{i=1}^{S} \tilde{\boldsymbol{\beta}}_i$.

Then the full conditional posterior distribution of $\tilde{\boldsymbol{\theta}}$ is:

$$
\tilde{\boldsymbol{\theta}} \mid \cdot \sim N_{L+1}\left\{\left[\frac{\tilde{\boldsymbol{\Omega}}_\theta(\boldsymbol{\tau})}{\sigma_\theta^2} + \frac{\tilde{\boldsymbol{\Omega}}_\beta(\boldsymbol{\gamma})}{\sigma^2/S}\right]^{-1} \frac{\tilde{\boldsymbol{\Omega}}_\beta(\boldsymbol{\gamma})}{\sigma^2} \sum_{i=1}^{S} \tilde{\boldsymbol{\beta}}_i, \left[\frac{\tilde{\boldsymbol{\Omega}}_\theta(\boldsymbol{\tau})}{\sigma_\theta^2} + \frac{\tilde{\boldsymbol{\Omega}}_\beta(\boldsymbol{\gamma})}{\sigma^2/S}\right]^{-1}\right\}
$$

(3) Posterior of σ^2

$$
\begin{aligned}
\pi(\sigma^2|\cdot) &\propto \prod_{i=1}^{S} \pi(\boldsymbol{Y}_i|\tilde{\boldsymbol{X}}_i, \tilde{\boldsymbol{\beta}}_i, \tilde{\boldsymbol{\theta}}, \sigma^2, \boldsymbol{\phi}_i, \boldsymbol{\gamma}) \pi(\tilde{\boldsymbol{\beta}}_i|\tilde{\boldsymbol{\theta}}, \sigma^2, \boldsymbol{\gamma}) \pi(\sigma^2) \\
&\propto \prod_{i=1}^{S} \left\{\left(\sigma^2\right)^{-\frac{n_i - p}{2}} \exp\left[-\frac{1}{2\sigma^2} (\boldsymbol{Y}_i^* - \tilde{\boldsymbol{X}}_i^* \tilde{\boldsymbol{\beta}}_i)' (\boldsymbol{Y}_i^* - \tilde{\boldsymbol{X}}_i^* \tilde{\boldsymbol{\beta}}_i)\right] |\sigma^2 \tilde{\boldsymbol{\Omega}}_\beta^{-1}(\boldsymbol{\gamma})|^{-\frac{1}{2}}\right\} \\
&\quad \exp\left[-\frac{1}{2\sigma^2} (\tilde{\boldsymbol{\beta}}_i - \tilde{\boldsymbol{\theta}})' \tilde{\boldsymbol{\Omega}}_\beta(\boldsymbol{\gamma}) (\tilde{\boldsymbol{\beta}}_i - \tilde{\boldsymbol{\theta}})\right] \frac{1}{\sigma^2} \\
&\propto \left(\sigma^2\right)^{-\frac{\sum n_i - S(p-L-1) - 2}{2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^{S} \left[(\tilde{\boldsymbol{\beta}}_i - \tilde{\boldsymbol{\theta}})' \tilde{\boldsymbol{\Omega}}_\beta(\boldsymbol{\gamma}) (\tilde{\boldsymbol{\beta}}_i - \tilde{\boldsymbol{\theta}}) + (\boldsymbol{Y}_i^* - \tilde{\boldsymbol{X}}_i^* \tilde{\boldsymbol{\beta}}_i)' (\boldsymbol{Y}_i^* - \tilde{\boldsymbol{X}}_i^* \tilde{\boldsymbol{\beta}}_i)\right]\right\}
\end{aligned}
$$

where L is the maximum number of lags, and p is the order of autoregressive error.

Then the full conditional posterior distribution of σ^2 is:

$$\sigma^2 \mid \cdot \sim \text{Inv-Gamma}\left[\frac{\sum n_i - S(p-L-1)}{2}, \frac{\sum_{i=1}^{S}\left[(\tilde{\boldsymbol{\beta}}_i - \tilde{\boldsymbol{\theta}})'\tilde{\boldsymbol{\Omega}}_\beta(\boldsymbol{\gamma})(\tilde{\boldsymbol{\beta}}_i - \tilde{\boldsymbol{\theta}}) + (\boldsymbol{Y}_i^* - \tilde{\boldsymbol{X}}_i^*\tilde{\boldsymbol{\beta}}_i)'(\boldsymbol{Y}_i^* - \tilde{\boldsymbol{X}}_i^*\tilde{\boldsymbol{\beta}}_i)\right]}{2}\right]$$

(4) Posterior of $\boldsymbol{\phi}$

For each individual N-of-1 trial, given $\boldsymbol{Y}_i$, $\tilde{\boldsymbol{X}}_i$, $\tilde{\boldsymbol{\beta}}_i$ and σ^2, the conditional joint distribution of $\epsilon^*_{i,p+1}, \epsilon^*_{i,p+2}, ..., \epsilon^*_{i,n}$ can be regarded as a sequence of one-sided conditional distributions:

$$\pi(\epsilon^*_{i,p+1}, \epsilon^*_{i,p+2}, ..., \epsilon^*_{i,n}) = \pi(\epsilon^*_{i,p+1}|\epsilon^*_{i,p}, ..., \epsilon^*_{i,1}) \cdots \pi(\epsilon^*_{i,n}|\epsilon^*_{i,n-1}, ..., \epsilon^*_{i,n-p})$$

where $\epsilon^*_{i,p+1}, \epsilon^*_{i,p+2}, ..., \epsilon^*_{i,n}$ are described in equation (3.13) of Section 3.3.

$$\begin{aligned}
\pi(\boldsymbol{\phi}_i|\cdot) &\propto \pi(\epsilon^*_{i,p+1}, \epsilon^*_{i,p+2}, ..., \epsilon^*_{i,n}|\boldsymbol{\phi}_i, \sigma^2)\pi(\boldsymbol{\phi}_i|\sigma^2_{\boldsymbol{\phi}}) \\
&\propto \prod_{t=p+1}^{n} \pi(\epsilon^*_{i,t}|\epsilon^*_{i,t-1}, ..., \epsilon^*_{i,t-p}, \boldsymbol{\phi}_i, \sigma^2)\pi(\boldsymbol{\phi}_i|\sigma^2_{\boldsymbol{\phi}}) \\
&\propto \exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{\epsilon}^*_i - \boldsymbol{E}^*_i\boldsymbol{\phi}_i)'(\boldsymbol{\epsilon}^*_i - \boldsymbol{E}^*_i\boldsymbol{\phi}_i)\right] \exp\left(-\frac{1}{2\sigma^2_{\boldsymbol{\phi}}}\boldsymbol{\phi}_i'\boldsymbol{\phi}_i\right) \mathbf{1}_{S_{\boldsymbol{\phi}}}(\boldsymbol{\phi}_i) \\
&\propto \exp\left[(\boldsymbol{\phi}_i - \hat{\boldsymbol{\phi}}_i)'\left(\sigma^{-2}\boldsymbol{E}_i^{*'}\boldsymbol{E}^*_i + \sigma^{-2}_{\boldsymbol{\phi}}\mathbf{I}\right)(\boldsymbol{\phi}_i - \hat{\boldsymbol{\phi}}_i)\right] \mathbf{1}_{S_{\boldsymbol{\phi}}}(\boldsymbol{\phi}_i)
\end{aligned}$$

where $\hat{\boldsymbol{\phi}}_i = \left(\sigma^{-2}\boldsymbol{E}_i^{*'}\boldsymbol{E}^*_i + \sigma^{-2}_{\boldsymbol{\phi}}\mathbf{I}\right)^{-1}\sigma^{-2}\boldsymbol{E}_i^{*'}\boldsymbol{\epsilon}^*_i$, $\sigma^2_{\boldsymbol{\phi}}$ is known and set to be 200.

Then the full conditional posterior distribution of $\boldsymbol{\phi}_i$ is:

$$\boldsymbol{\phi}_i \mid \cdot \sim N_p\left[\left(\sigma^{-2}\boldsymbol{E}_i^{*'}\boldsymbol{E}^*_i + \sigma^{-2}_{\boldsymbol{\phi}}\mathbf{I}\right)^{-1}\sigma^{-2}\boldsymbol{E}_i^{*'}\boldsymbol{\epsilon}^*_i, \left(\sigma^{-2}\boldsymbol{E}_i^{*'}\boldsymbol{E}^*_i + \sigma^{-2}_{\boldsymbol{\phi}}\mathbf{I}\right)^{-1}\right] \mathbf{1}_{S_{\boldsymbol{\phi}}}(\boldsymbol{\phi}_i)$$

(5) Posterior of $\boldsymbol{\gamma}$

$$
\begin{aligned}
\pi(\boldsymbol{\gamma} \mid \cdot) &\propto \prod_{i=1}^{S} \pi(\boldsymbol{Y}_i | \tilde{\boldsymbol{X}}_i, \tilde{\boldsymbol{\beta}}_i, \tilde{\boldsymbol{\theta}}, \sigma^2, \boldsymbol{\phi}_i, \boldsymbol{\gamma}) \pi(\boldsymbol{\gamma} \mid \tilde{\boldsymbol{\beta}}_i, \tilde{\boldsymbol{\theta}}, \sigma^2, \boldsymbol{\phi}_i) \\
&\propto \pi(\boldsymbol{\gamma}) \prod_{i=1}^{S} \pi(\boldsymbol{Y}_i | \tilde{\boldsymbol{X}}_i, \tilde{\boldsymbol{\beta}}_i, \tilde{\boldsymbol{\theta}}, \sigma^2, \boldsymbol{\phi}_i, \boldsymbol{\gamma}) \pi(\tilde{\boldsymbol{\beta}}_i \mid \tilde{\boldsymbol{\theta}}, \sigma^2, \boldsymbol{\gamma}) \\
&\propto \exp(-\gamma_1 - \gamma_2) \mathbf{1}_{S_\gamma}(\gamma_1, \gamma_2) \prod_{i=1}^{S} |\sigma^{-2} \tilde{\boldsymbol{\Omega}}_\beta(\boldsymbol{\gamma})|^{\frac{1}{2}} \exp\left[-\frac{1}{2\sigma^2} (\tilde{\boldsymbol{\beta}}_i - \tilde{\boldsymbol{\theta}})' \tilde{\boldsymbol{\Omega}}_\beta(\boldsymbol{\gamma}) (\tilde{\boldsymbol{\beta}}_i - \tilde{\boldsymbol{\theta}}) \right]
\end{aligned}
$$

Appendix C: Appendices to Chapter 4

C.1 Derivation of full conditional posterior distributions of parameters

Let the notation "$\pi(A|\cdot)$" represents conditional distribution of A given all other variables in the model.

Given the likelihood function and prior distribution in Section 4.2, full conditional posterior distributions of each variable are as follows:

(1) Posterior of $\tilde{\boldsymbol{\beta}}_c$

$$\begin{aligned}\pi(\tilde{\boldsymbol{\beta}}_c|\cdot) &\propto \pi(\boldsymbol{Y}|\tilde{\boldsymbol{X}}, \tilde{\boldsymbol{\beta}}, \sigma^2, \boldsymbol{\phi}, \boldsymbol{\gamma})\pi(\boldsymbol{\beta}_c|\sigma^2, \boldsymbol{\gamma}) \\ &\propto \exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{Y}^* - \sum_{c=1}^{C-1}\tilde{\boldsymbol{X}}_c^*\tilde{\boldsymbol{\beta}}_c)'(\boldsymbol{Y}^* - \sum_{c=1}^{C-1}\tilde{\boldsymbol{X}}_c^*\tilde{\boldsymbol{\beta}}_c)\right]\exp\left[-\frac{1}{2\sigma^2}\tilde{\boldsymbol{\beta}}_c'\tilde{\boldsymbol{\Omega}}_c(\boldsymbol{\gamma}_c)\tilde{\boldsymbol{\beta}}_c\right]\end{aligned}$$

where $\boldsymbol{Y}^*$ and $\tilde{\boldsymbol{X}}^*$ are transformed $\boldsymbol{Y}$ and $\tilde{\boldsymbol{X}}$ as described in equation (4.3).
Denote $\hat{\boldsymbol{\beta}}_c = \left[\tilde{\boldsymbol{X}}_c^{*'}\tilde{\boldsymbol{X}}_c^* + \tilde{\boldsymbol{\Omega}}_c(\boldsymbol{\gamma}_c)\right]^{-1}\tilde{\boldsymbol{X}}_c^{*'}(\boldsymbol{Y}^* + \tilde{\boldsymbol{X}}_c^*\tilde{\boldsymbol{\beta}}_c - \sum_{k=1}^{C-1}\tilde{\boldsymbol{X}}_k^*\tilde{\boldsymbol{\beta}}_k)$ and using the equation

$$\begin{aligned}(\boldsymbol{Y}^* - \tilde{\boldsymbol{X}}_c^*\tilde{\boldsymbol{\beta}})_c'(\boldsymbol{Y}^* - \tilde{\boldsymbol{X}}_c^*\tilde{\boldsymbol{\beta}})_c + \tilde{\boldsymbol{\beta}}_c'\tilde{\boldsymbol{\Omega}}_c(\boldsymbol{\gamma}_c)\tilde{\boldsymbol{\beta}}_c &= (\tilde{\boldsymbol{\beta}}_c - \hat{\boldsymbol{\beta}}_c)'\left[\tilde{\boldsymbol{X}}_c^{*'}\tilde{\boldsymbol{X}}_c^* + \tilde{\boldsymbol{\Omega}}_c(\boldsymbol{\gamma}_c)\right](\tilde{\boldsymbol{\beta}}_c - \hat{\boldsymbol{\beta}}_c) \\ &+(\boldsymbol{Y}^* + \tilde{\boldsymbol{X}}_c^*\tilde{\boldsymbol{\beta}}_c - \sum_{k=1}^{C-1}\tilde{\boldsymbol{X}}_k^*\tilde{\boldsymbol{\beta}}_k)'(\boldsymbol{Y}^* + \tilde{\boldsymbol{X}}_c^*\tilde{\boldsymbol{\beta}}_c - \sum_{k=1}^{C-1}\tilde{\boldsymbol{X}}_k^*\tilde{\boldsymbol{\beta}}_k) - \hat{\boldsymbol{\beta}}_c'\left[\tilde{\boldsymbol{X}}_c^{*'}\tilde{\boldsymbol{X}}_c^* + \tilde{\boldsymbol{\Omega}}_c(\boldsymbol{\gamma}_c)\right]\hat{\boldsymbol{\beta}}_c\end{aligned}$$

Then the full conditional posterior distribution of $\tilde{\boldsymbol{\beta}}_c$ is:

$$\pi(\tilde{\boldsymbol{\beta}}_c|\cdot) \propto \exp\left[(\tilde{\boldsymbol{\beta}}_c - \hat{\boldsymbol{\beta}})_c'\left[\tilde{\boldsymbol{X}}_c^{*'}\tilde{\boldsymbol{X}}_c^* + \tilde{\boldsymbol{\Omega}}_c(\boldsymbol{\gamma}_c)\right](\tilde{\boldsymbol{\beta}}_c - \hat{\boldsymbol{\beta}}_c)\right]$$

$$\tilde{\boldsymbol{\beta}}_c \mid \cdot \sim N_{L+1}\left\{[\tilde{\boldsymbol{X}}_c^{*'}\tilde{\boldsymbol{X}}_c^* + \tilde{\boldsymbol{\Omega}}_c(\boldsymbol{\gamma}_c)]^{-1}\tilde{\boldsymbol{X}}_c^{*'}(\boldsymbol{Y}^* + \tilde{\boldsymbol{X}}_c^*\tilde{\boldsymbol{\beta}}_c - \sum_{k=1}^{C-1}\tilde{\boldsymbol{X}}_k^*\tilde{\boldsymbol{\beta}}_k), \sigma^2[\tilde{\boldsymbol{X}}_c^{*'}\tilde{\boldsymbol{X}}_c^* + \tilde{\boldsymbol{\Omega}}_c(\boldsymbol{\gamma}_c)]^{-1}\right\}$$

(2) Posterior of σ^2

$$\begin{aligned}
\pi(\sigma^2|\cdot) &\propto \pi(\boldsymbol{Y}|\tilde{\boldsymbol{X}}, \tilde{\boldsymbol{\beta}}, \sigma^2, \boldsymbol{\phi}, \boldsymbol{\gamma})\pi(\sigma^2) \prod_{c=1}^{C-1} \pi(\tilde{\boldsymbol{\beta}}_c|\sigma^2, \boldsymbol{\gamma}) \\
&\propto \left(\sigma^2\right)^{-\frac{n-p}{2}} \exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{Y}^* - \sum_{c=1}^{C-1} \tilde{\boldsymbol{X}}_c^* \tilde{\boldsymbol{\beta}}_c)'(\boldsymbol{Y}^* - \sum_{c=1}^{C-1} \tilde{\boldsymbol{X}}_c^* \tilde{\boldsymbol{\beta}}_c)\right] \frac{1}{\sigma^2} \\
&\quad \prod_{c=1}^{C-1} |\sigma^2 \tilde{\boldsymbol{\Omega}}_c^{-1}(\boldsymbol{\gamma}_c)|^{-\frac{1}{2}} \exp\left[-\frac{1}{2\sigma^2} \tilde{\boldsymbol{\beta}}_c' \tilde{\boldsymbol{\Omega}}_c(\boldsymbol{\gamma}_c) \tilde{\boldsymbol{\beta}}_c\right] \\
&\propto \left(\sigma^2\right)^{-\frac{n-p+(L+1)(C-1)+2}{2}} \exp\left\{-\frac{1}{2\sigma^2}\left[(\boldsymbol{Y}^* - \sum_{c=1}^{C-1} \tilde{\boldsymbol{X}}^* \tilde{\boldsymbol{\beta}})'(\boldsymbol{Y}^* - \sum_{c=1}^{C-1} \tilde{\boldsymbol{X}}^* \tilde{\boldsymbol{\beta}}) + \sum_{c=1}^{C-1} \tilde{\boldsymbol{\beta}}_c' \tilde{\boldsymbol{\Omega}}_c(\boldsymbol{\gamma}_c) \tilde{\boldsymbol{\beta}}_c\right]\right\}
\end{aligned}$$

where L is the maximum number of lags, and p is the order of autoregressive error.

Then the full conditional posterior distribution of σ^2 is:

$$\sigma^2 \mid \cdot \sim \text{Inv-Gamma}\left[\frac{n-p+(C-1)(L+1)}{2}, \frac{(\boldsymbol{Y}^* - \sum_{c=1}^{C-1} \tilde{\boldsymbol{X}}_c^* \tilde{\boldsymbol{\beta}}_c)'(\boldsymbol{Y}^* - \sum_{c=1}^{C-1} \tilde{\boldsymbol{X}}_c^* \tilde{\boldsymbol{\beta}}_c) + \tilde{\boldsymbol{\beta}}_c' \tilde{\boldsymbol{\Omega}}_c(\boldsymbol{\gamma}_c) \tilde{\boldsymbol{\beta}}_c}{2}\right]$$

(3) Posterior of $\boldsymbol{\phi}$

Given $\boldsymbol{Y}$, $\tilde{\boldsymbol{X}}$, $\tilde{\boldsymbol{\beta}}$ and σ^2, the conditional joint distribution of $\epsilon_{p+1}^*, \epsilon_{p+2}^*, ..., \epsilon_n^*$ can be regarded as a sequence of one-sided conditional distributions:

$$\pi(\epsilon_{p+1}^*, \epsilon_{p+2}^*, ..., \epsilon_n^*) = \pi(\epsilon_{p+1}^*|\epsilon_p^*, ..., \epsilon_1^*) \cdots \pi(\epsilon_n^*|\epsilon_{n-1}^*, ..., \epsilon_{n-p}^*)$$

where $\epsilon_{p+1}^*, \epsilon_{p+2}^*, ..., \epsilon_n^*$ are described in equation (4.9) of Section 4.3.

$$\begin{aligned}
\pi(\boldsymbol{\phi}|\cdot) &\propto \pi(\epsilon_{p+1}^*, \epsilon_{p+2}^*, ..., \epsilon_n^*|\boldsymbol{\phi}, \sigma^2)\pi(\boldsymbol{\phi}|\sigma_{\boldsymbol{\phi}}^2) \\
&\propto \prod_{t=p+1}^{n} \pi(\epsilon_t^*|\epsilon_{t-1}^*, ..., \epsilon_{t-p}^*, \boldsymbol{\phi}, \sigma^2)\pi(\boldsymbol{\phi}|\sigma_{\boldsymbol{\phi}}^2) \\
&\propto \exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{\epsilon}^* - \boldsymbol{E}^*\boldsymbol{\phi})'(\boldsymbol{\epsilon}^* - \boldsymbol{E}^*\boldsymbol{\phi})\right] \exp\left(-\frac{1}{2\sigma_{\boldsymbol{\phi}}^2}\boldsymbol{\phi}'\boldsymbol{\phi}\right) \mathbf{1}_{S_{\boldsymbol{\phi}}}(\boldsymbol{\phi}) \\
&\propto \exp\left[(\boldsymbol{\phi} - \hat{\boldsymbol{\phi}})'\left(\sigma^{-2}\boldsymbol{E}^{*'}\boldsymbol{E}^* + \sigma_{\boldsymbol{\phi}}^{-2}\mathbf{I}\right)(\boldsymbol{\phi} - \hat{\boldsymbol{\phi}})\right] \mathbf{1}_{S_{\boldsymbol{\phi}}}(\boldsymbol{\phi})
\end{aligned}$$

where $\hat{\boldsymbol{\phi}} = \left(\sigma^{-2}\boldsymbol{E}^{*'}\boldsymbol{E}^{*} + \sigma_{\boldsymbol{\phi}}^{-2}\mathbf{I}\right)^{-1}\sigma^{-2}\boldsymbol{E}^{*'}\boldsymbol{\epsilon}^{*}$, $\sigma_{\boldsymbol{\phi}}^{2}$ is known and set to be 200.

Then the full conditional posterior distribution of $\boldsymbol{\phi}$ is:

$$\boldsymbol{\phi} \mid \cdot \sim N_p\left[\left(\sigma^{-2}\boldsymbol{E}^{*'}\boldsymbol{E}^{*} + \sigma_{\boldsymbol{\phi}}^{-2}\mathbf{I}\right)^{-1}\sigma^{-2}\boldsymbol{E}^{*'}\boldsymbol{\epsilon}^{*}, \left(\sigma^{-2}\boldsymbol{E}^{*'}\boldsymbol{E}^{*} + \sigma_{\boldsymbol{\phi}}^{-2}\mathbf{I}\right)^{-1}\right]\mathbf{1}_{S_{\boldsymbol{\phi}}}(\boldsymbol{\phi})$$

(4) Posterior of $\boldsymbol{\gamma}$

$$\begin{aligned}
\pi(\boldsymbol{\gamma}_c \mid \cdot) &\propto \pi(\boldsymbol{Y}|\tilde{\boldsymbol{X}}, \tilde{\boldsymbol{\beta}}, \sigma^2, \boldsymbol{\phi}, \boldsymbol{\gamma})\pi(\boldsymbol{\gamma}_c \mid \tilde{\boldsymbol{\beta}}_c, \sigma^2, \boldsymbol{\phi}) \\
&\propto \pi(\boldsymbol{Y}|\tilde{\boldsymbol{X}}, \tilde{\boldsymbol{\beta}}, \sigma^2, \boldsymbol{\phi}, \boldsymbol{\gamma})\pi(\tilde{\boldsymbol{\beta}}_c \mid \sigma^2, \boldsymbol{\gamma}_c)\pi(\boldsymbol{\gamma}_c) \\
&\propto |\sigma^{-2}\tilde{\boldsymbol{\Omega}}_c(\boldsymbol{\gamma}_c)|^{\frac{1}{2}} \exp\left[-\frac{1}{2\sigma^2}\tilde{\boldsymbol{\beta}}_c'\tilde{\boldsymbol{\Omega}}_c(\boldsymbol{\gamma}_c)\tilde{\boldsymbol{\beta}}_c\right] \exp(-\gamma_{1,c} - \gamma_{2,c})\mathbf{1}_{S_{\boldsymbol{\gamma}}}(\gamma_{1,c}, \gamma_{2,c})
\end{aligned}$$