Random Effects in Point Processes: Adding Flexibility to Ecological Momentary Assessment Analysis

by

Benjamin B. Neustifter

(Under the direction of Stephen L. Rathbun)

Abstract

Models based upon the research done by Rathbun, Shiffman, & Gwaltney (2007) and Waagepetersen (2008) are developed for modeling repeated behavioral events impacted by time-varying covariates. Using forms similar to Poisson-Gamma Hierarchical Generalized Linear Models and Generalized Linear Mixed Models, two modulated Poisson process with random effects are introduced to allow for inter-subject variability. The first allows for random baseline event rates across subjects, and the second is a mixed-effect model with normally distributed random components. Estimation of the parameters for each model is discussed in the case that the covariates are only known at certain random event and non-event assessment times. Integral estimation methods developed by Rathbun et al. (2007) and Waagepetsersen (2008) are utilized to calculate computationally efficient estimating functions. Prediction of the subject-level effects is also discussed. The approaches are illustrated using data from an Ecological Momentary Assessment of smoking.

Index words: Point process, random effect, Ecological Momentary Assessment, smoking, integral estimation, hierarchical generalized linear model, generalized linear mixed model, Poisson process, intensity estimation

Random Effects in Point Processes:
Adding Flexibility to Ecological Momentary Assessment Analysis

by

Benjamin B. Neustifter

B.S., Indiana University Bloomington, 2004

B.A., Indiana University Bloomington, 2004

M.S., The University of Georgia, 2007

A Dissertation Submitted to the Graduate Faculty
of The University of Georgia in Partial Fulfillment
of the
Requirements for the Degree

Doctor of Philosophy

Athens, Georgia

2009

Random Effects in Point Processes: Adding Flexibility to Ecological Momentary Assessment Analysis

by

Benjamin B. Neustifter

Approved:

Major Professor: Stephen L. Rathbun

Committee: Nicole Lazar
William P. McCormick
Lynne Seymour
T. N. Sriram

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2009

Dedication

To Ruth. Even 12-hour days spent typing are bearable when spent with you.

Acknowledgments

First, I am endlessly grateful to Dr. Stephen Rathbun for his hands-on guidance throughout the research and dissertation process. A good major professor is a priceless resource, and your tireless work and constant availability and supervision made this work possible. My thanks also go out to the rest of my wonderful committee. Dr. Seymour's excellent programming classes made it possible for me to develop the software necessary to run my estimators, and her encylopedic knowledge of university requirements kept me from ever running afoul of paperwork and bureacracy. Dr. McCormick's useful class on point processes, Dr. Sriram's classes on asymptotics, Dr. Lazar's flexibility and suggestions for future research, and Dr. Hall's expertise in the area of estimating equations and helpful revisions to the final draft were all other essential resources for which I am very appreciative.

My partner, Ruth, has been an endless source of energy, support, and love during the dissertation process. I am so glad we could help each other through the doctoral process, and am so lucky to have you. I am looking forward to our new start together in D.C.

My parents have encouraged my love of math since a very early age, and raised me to embrace my education and pursue it to its fullest. Their love and support, even when I am busy and unavailable, will always be important to me.

Bob and Rita also deserve thanks for their boundless support during this process. The monthly care packages kept us going even through the toughest times, and we always had someone to turn to with our stresses and troubles. Thank you!

Finally, my pets deserve thanks for being so patient with their boring parents, who spent far too much time typing on the couch and not nearly enough time taking you to the park. We're getting you a new, bigger back yard soon, we promise!

This work was supported by NSF grant SES-0720195. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Table of Contents

Page

Acknowledgments v

List of Figures ix

List of Tables x

Chapter

1 Introduction 1

1.1 Introduction 1

1.2 Ecological Momentary Assessment 2

1.3 Motivating Dataset 3

1.4 Previous Research on Modulated Poisson Processes and EMA Data Analysis 4

1.5 Proposed Models 9

1.6 References 11

2 Estimator of the Intensity of a Modulated Poisson Process with a Gamma Prior 14

2.1 Introduction 14

2.2 Modulated Poisson Processes with Probability Sampling 16

2.3 Random Baseline Intensity Model 18

2.4 Results 22

2.5 References 34

3 Baseline Intensity in Modulated Poisson Processes: A Technique for Analyzing Ecological Momentary Assessment Data 38

3.1 Introduction 38
3.2 Random Baseline Intensity Model 40
3.3 Asymptotic Properties 46
3.4 Estimation of Baseline Intensity Rates 50
3.5 Results 51
3.6 References 61

4 Modulated Poisson Processes with Random Coefficients: Adding Flexibility to EMA Analysis 64
4.1 Introduction 64
4.2 Random Coefficient Modulated Poisson Process 67
4.3 Asymptotic Properties of the Proposed Estimators 74
4.4 Estimation of the $\boldsymbol{\beta}_i$s 78
4.5 EMA of Smoking 79
4.6 References 98

5 Conclusion 102
5.1 Findings 102
5.2 Future Research 105
5.3 Conclusions 107
5.4 References 107

List of Figures

3.1 Baseline Estimates using Rathbun et al. (2007) method and gamma distribution 54

3.2 Baseline Estimates using Waagepetersen (2008) method and gamma distribution 55

4.1 Estimates for the Subject-Level Effects of "Other Smokers" 84

List of Tables

2.1 Results of Rathbun et al. (2007) 24

2.2 Results of Random Baseline Intensity Model 25

2.3 Estimated Risk Ratios and 95% CIs Using Current Method 25

3.1 Comparison between Rathbun et al. (2007) method and current method . . . 53

3.2 Comparison between Waagepetersen (2008) method and current method . . 53

4.1 Fixed/Mean-effect estimates for first model 81

4.2 Variance/covariance parameters for random effects, first model 81

4.3 Fixed/Mean-effect estimates for second model 82

4.4 Variance parameter for random effect, second model 83

Chapter 1

Introduction

1.1 Introduction

Ecological Momentary Assessment (EMA) is an emerging method of data collection in psychology that is focused upon obtaining information about human behavior in a subject's environment (Shiffman & Stone 1998; Stone & Shiffman 2002). This increasingly utilized method uses electronic devices to assess study subjects repeatedly over time. These devices may also be used to record discrete behavioral events, as well as the subjects' psychological state at the time. Doing so can allow researchers to explore the relationship between subjects' psychological states and the timing of relevant events.

While there is a body of literature addressing the analysis of EMA data (see Shiffman & Stone 1998; Stone & Shiffman 2002; Rathbun, Shiffman, & Gwaltney 2007), it is an area of statistical procedures that has not yet received significant attention. The current methods of EMA data analysis do not provide adequate room for inter-subject variability, as will be addressed in section 1.4. This thesis proposes two different models that expand upon previous research for the use of Poisson processes to analyze repeated event data (Rathbun et al. 2007; Waagepetersen 2008). These models, which are written in a general form to incorporate the integral estimation techniques of Rathbun et al. and Waagepetersen, will be shown to have the desirable asymptotic properties of consistency and asymptotic normality.

To establish the need for this research the literature regarding EMA, particularly the analysis of repeated event data (such as that obtained by EMA studies), will be reviewed. The motivating dataset for this work will then be discussed. A brief overview of modulated Poisson processes will be given, followed by summaries of Rathbun et al. (2007) and Waagepetersen's

(2008) works on these processes and their restrictions for inter-subject variability will be demonstrated. This chapter will conclude with an overview of the contents of the rest of the dissertation.

1.2 Ecological Momentary Assessment

EMA is a method of data collection that is primarily concerned with gathering data about human behavior in the subject's environment (Shiffman & Stone 1998; Stone & Shiffman 2002). While Stone and Shiffman (1994) are responsible for providing the name EMA in psychology and a strong body of literature to support the usage of this method, there is some prior literature that addresses this area. Most of these papers refer to the method as the "experience sampling method," (Csikszentmihalyi & Larson 1987, DeVries 1992) which is its name in the public health and biomedicinal fields. However, since Shiffman and Stone are involved in much of the current research on this topic, this dissertation will use the name EMA.

The collection of event data as it occurs is often accomplished through the usage of electronic devices, such as handheld personal digital assistants (PDAs), that are provided to participants. By allowing subjects to record event data as it occurs naturally in their environments, EMA eliminates some of the biases associated with retrospective recall and artificial laboratory environments. Reducing these biases accordingly increases the "ecological validity" of the data (Shiffman & Stone 1998, p. 117). Stone et al.'s (1998) study on coping showed that subjects consistently appeared to underreport cognitive coping items and overreport behavioral coping items by comparing subjects' EMA reports on coping versus their retrospective recall. They also found that subjects tended to omit reporting certain cognitive processes in the EMA collection. Despite this, a strong argument can be made for the use of EMA data collection, particularly regarding the occurrence of discrete behavioral events. EMA has been used to study numerous phenomena, including stress coping (Stone et al. 1998), eating disorders (Smyth et al. 2001), emotional adaptation to medical treatments (Riis et al. 2005), and pain (Stone, Broderick, Porter & Kaell 1997; Stone, Broderick, Shiffman & Schwartz 2004).

As will be addressed in Section 1.4, some models have previously been suggested for the analysis of EMA data. However, they have either been unable to accurately model the timing of repeated behavioral events or have not provided adequate flexibility for the modeling of inter-subject variability. The models that will be proposed in this dissertation will bridge this gap to provide a better grounding for statistical analysis of EMA data.

1.3 Motivating Dataset

The motivating dataset for this research comes from Shiffman et al.'s (2002) multi-phase EMA study on smoking habits. A sample of 304 smokers used electronic diaries to record the exact times that they smoked each cigarette. According to a Bernoulli process with known probability, the subjects were prompted by an audible cue to answer a series of assessment questions about their mood and environment at selected smoking "events." The probability for the Bernoulli process was chosen such that there were an expected 5 event assessments each day for every subject.

Subjects were also prompted to complete assessments at random times throughout the day. These random, or "dummy," assessments followed a stratified sampling design, by subject and day. An expected 5 random assessments were prompted over the time each day that the diary was on and available for assessing. Subjects were allowed to turn off the diaries or put them on stand-by for sleep, driving, or situations such as business meetings or movies. The random assessments were restricted to when the diary was on, not on stand-by, and not within 10 minutes after an event. Among the variables recorded during assessments were Negative Affect (based on reactions to negative mood adjectives), Arousal (reactions to words such as "energetic"), Attention (based on reports of difficulty concentrating), Restlessness (based on a single item that did not load into these other factors), and whether Other Smokers were present.

It is important to note that the random deletion of points from a Poisson process by independent Bernoulli trials results in a Poisson process (Cressie 1991). This result ensures that, even though the theory that follows in Section 1.4 and Chapters 2 through 4 assumes that every event in the Poisson process has been assessed, it is still applicable to situations such as the Shiffman et al. (2002) dataset.

This is the same dataset that was analyzed by Rathbun et al. (2007). With their techniques, they found that Restlessness was the only one of the four variables that was found to have a significant impact on the timing of smoking events. Negative Affect had a questionably significant ($.05 < p < .10$) effect, as well. The following chapters will use the Rathbun et al. (2007) analysis as a baseline against which to compare the results of the methods in Chapters 2 through 4.

1.4 Previous Research on Modulated Poisson Processes and EMA Data Analysis

1.4.1 Modulated Poisson Processes

Previous researchers have suggested Hierarchical Linear Models (HLMs) for the analysis of EMA data (Affleck et al. 1999; Bolger, Davis, & Rafaeli 2003; Schwartz & Stone 2007). HLMs allow researchers to separate within- and between-subjects variation in the data. These suggestions, however, were based upon the analysis of EMA data without data on behavioral events, instead focusing upon the randomly sampled covariate states. HLMs are not designed for the analysis or modeling of the *timing* of discrete, repeated events based upon the values of covariates.

Point processes are frequently used to model the occurrence of events in time, space, or space-time. They have been used to model such events as the timing and location of rainfall (Northrop 1998), neural activity (Johnson 1996), or the location and diameter of trees (Stoyan & Penttinen 2000). Rathbun et al. (2007) suggested the use of these processes to analyze data from EMA studies where the timing of repeated behavioral events were available, as well as data regarding covariates at these and other times throughout the study interval.

Point processes can be partially described by a function termed the intensity, which is a measure of the expected rate of events over time for a specific instant. In the one-dimensional case, if N is a random measure such that $N([t_1, t_2])$ is the number of events occurring in the interval $[t_1, t_2]$, then the intensity $\lambda(t)$ is defined as

$$\lambda(t) = \lim_{\delta \to 0} \frac{E(N[t, t+\delta])}{\delta}.$$

The definition in the multi-dimensional case is analogous. The Poisson process (Cox 1972) is a specific type of point process where the counting measure N is such that for any interval A, $N(A) \sim Poisson(\Lambda(A))$, where $\Lambda(A) = \int_A \lambda(t)\, dt$, and where, conditional on $N(A)$, the event locations are independently sampled from the probability density function proportional to $\lambda(t)$. Poisson processes have several desirable properties that enhance their use in the analysis of repeated events: their relatively simple form, and their independence on disjoint intervals. In other words, Poisson processes have no "memory;" the occurrence of events in the time interval $[t_1, t_2]$ does not affect the occurrence of events in $[t_3, t_4]$ for $t_3 > t_2$.

A modulated Poisson process (Cox 1972), is a special case of a Poisson process where the intensity function is

$$\lambda(t; \boldsymbol{\beta}) = \exp\{\boldsymbol{\beta}^T \mathbf{x}(t)\}, t \in A. \tag{1.4.1}$$

Here $\mathbf{x}(t)$ is a vector of time-varying covariates (e.g. ratings of mood), and $\boldsymbol{\beta}$ is a (generally unknown) vector of coefficient parameters for these covariates. Thus, $\boldsymbol{\beta}$ describes the effects of the covariates on the intensity of the process, and so on the frequency of events.

For the Poisson process with intensity (1.4.1), the log-likelihood of the parameter vector $\boldsymbol{\beta}$ for a single individual i over the study interval A_i is

$$\ell_i(\boldsymbol{\beta}) = \sum_{t \in X_i} \boldsymbol{\beta}^T \mathbf{x}_i(t) - \Lambda_i(A_i; \boldsymbol{\beta}),$$

where X_i is the set of events in the Poisson process, $\mathbf{x}_i(t)$ is the covariate vector for the ith individual, and $\Lambda_i(A_i; \boldsymbol{\beta}) = \int_{A_i} \exp\{\boldsymbol{\beta}^T \mathbf{x}_i(t)\}\, dt$. The score equation for the parameter vector $\boldsymbol{\beta}$ in this case, then, is

$$\Psi_i(\boldsymbol{\beta}) = \sum_{t \in X_I} \mathbf{x}(t) - \Lambda_i^{(1)}(A_i; \boldsymbol{\beta}), \tag{1.4.2}$$

where $\Lambda_i^{(1)}(A_i; \boldsymbol{\beta}) = (\partial / \partial \boldsymbol{\beta}) \Lambda_i(A_i; \boldsymbol{\beta})$.

1.4.2 Integral Estimation

Obtaining the solution to (1.4.2) requires knowing the values of the covariates at all time points in the study period in order to evaluate $\Lambda_i^{(1)}(A_i; \boldsymbol{\beta})$. In most applications constant observation of the covariates is unrealistic. Thus, methods of estimating this integral are necessary.

In addition to assuming that assessments of the covariates were taken at all event times $\{t : t \in X_{ni}\}$, if one assumes that random assessments were taken according to a known probabilistic sampling scheme, one can unbiasedly estimate these integrals. Recent literature by Rathbun et al. (2007) and Waagepetersen (2008) give techniques that allow the estimation of (1.4.2) and similar integrals in this case.

Take a set of assessment times $\{u : u \in D_i\}$, where D_i is a set of random, or "dummy," assessments chosen according to a known probability-based sampling scheme. Define $\pi_i(u)$ as the inclusion density (Cordy 1993) at time u for subject i according to this scheme. Rathbun et al. (2007) proposed the unbiased integral estimator

$$\hat{\Lambda}_{i,R}^{(1)}(A; \boldsymbol{\beta}) = \sum_{u \in D_i} \frac{\mathbf{x}_i(t) \exp\{\boldsymbol{\beta}^T \mathbf{x}_i(t)\}}{\pi_i(u)}. \tag{1.4.3}$$

Waagepetersen (2008), though not specifically writing about the application to EMA, suggested another unbiased integral estimator that appears to be more efficient, as it uses all of the available covariate information:

$$\hat{\Lambda}^{(1)}_{i,w}(A;\boldsymbol{\beta}) = \sum_{t\in X_i} \frac{\mathbf{x}_i(t)\exp\{\boldsymbol{\beta}^T\mathbf{x}_i(t)\}}{\pi_i(t)+\exp\{\boldsymbol{\beta}^T\mathbf{x}_i(t)\}} + \sum_{u\in D_i} \frac{\mathbf{x}_i(t)\exp\{\boldsymbol{\beta}^T\mathbf{x}_i(t)\}}{\pi_i(u)+\exp\{\boldsymbol{\beta}^T\mathbf{x}_i(u)\}}. \tag{1.4.4}$$

These integral estimators, when used in (1.4.2) with appropriate normalization, provide consistent and asymptotically normal estimates of $\boldsymbol{\beta}$, under increasing domain and infill asymptotics, respectively. These researchers have provided methods through which EMA data can be collected and analyzed to model repeated behavioral events as impacted by subject covariates.

1.4.3 Adding Inter-Subject Variability

If one estimates $\boldsymbol{\beta}$ using the estimating equation (with some normalization)

$$\hat{\Psi}(\boldsymbol{\beta}) = \sum_{i=1}^{n}\left[\sum_{t\in X_i}\mathbf{x}_i(t) - \hat{\Lambda}^{(1)}_i(A_i;\boldsymbol{\beta})\right],$$

for either Rathbun et al. (2007) or Waagepetersen's (2008) integral estimator, one can obtain consistent and asymptotically normal estimates for $\boldsymbol{\beta}$, as stated in the previous section. Doing so requires the explicit assumption that all subjects should have identical intensities if their covariate vectors are equal at some point in time. In the context of the motivating dataset, for example, this implies that two subjects who are demographically alike and in the same mood and location should be smoking at the same rate. This assumption may be unrealistic in many applications.

The purpose of the current research is to develop models for evaluating repeated event data that model event timings based on a modulated Poisson process, but allow for inter-subject variability by adding random effect terms. In order to justify adding such terms, as well as prompt some of this research's modeling and analysis decisions, this dissertation will briefly review the relevent literature on Poisson models with random effects.

It is important to note that little literature exists on Poisson processes with random effects, if any. Much of the previous research focused on an extension of Generalized Linear Models (GLMs) (Nelder & Wedderburn 1972). In GLMs, the response vector $\mathbf{y}$ has expected value $E(\mathbf{y}) = \boldsymbol{\mu}$, and this mean is related to the covariates by a "link function" $g(\cdot)$, such that

$$g(\boldsymbol{\mu}) = \boldsymbol{\beta}^T\mathbf{x}. \tag{1.4.5}$$

Furthermore, the distribution of $\mathbf{y}$ comes from the exponential family of distributions (Lehmann 1959). This general set of models covers various forms of regression such as probit, logistic, and Poisson regression.

Generalized Linear Mixed Models (GLMMs) are an extension of GLMs that add a normal random effect upon the terms $\boldsymbol{\beta}$ in (1.4.5) in order to allow covariate effects to vary among subjects (see Breslow & Clayton 1993). Generally, GLMMs where the elements of $\mathbf{y}$ are Poisson distributed and g is the logarithmic link (which is the case closest to this dissertation's modulated Poisson process model) are evaluated using computationally intensive methods, such as Monte Carlo Expectation-Maximization (MCEM) (see McCulloch 1997), since the log-likelihood and the resulting score equations have no closed form. These methods are not adapted to the current model due to their computationally intensive nature and lack of closed-form solutions. GLMMs will provide the basis for the model definition for adding random effects to the covariate coefficient parameters in Chapter 4.

Hierarchical Generalized Liear Models (HGLMs) are a further extension of GLMs that are generally used to model multi-level data where the responses may be clustered (Lee & Nelder 1996). Rather than assuming a normal distribution, HGLMs allow any distribution to be placed upon the random effect terms. Traditionally, the conjugate to the response distribution is chosen; in the case that the responses are Poisson-distributed the random effects are often Gamma-distributed. This allows for a closed-form likelihood and score equations. Lee and Nelder (1996) also suggest the h-likelihood as an alternative form of parameter estimation when the random effect parameter is a function of the canonical parameter.

When it is not, this is instead Extended Likelihood estimation, which Lee and Nelder warn may provide consistent or unbiased results. Although HGLMs, like GLMMs, are regression-type models their approach can be somewhat adapted to the modulated Poisson case. This approach largely informs the model-definition direction in Chapters 2 and 3, where the model is adapted to allow for more inter-subject variability by placing a random baseline term in the intensity.

1.5 Proposed Models

In the following chapters of this dissertation, two different methods of adding inter-subject variability to the modulated Poisson process model will be proposed. Chapters 2 and 3 modify the model to allow subjects to have a randomly-distributed baseline event occurrence rate while holding the effects of the covariates fixed. The intensity (1.4.1) is changed to the form

$$\lambda_i(t) = u_i \exp\{\boldsymbol{\beta}^T \mathbf{x}_i(t)\}; t \in A_{ni}, i = 1, \ldots, n, \tag{1.5.1}$$

where A_{ni} is the ith subject's study interval when the sample size is n. The u_i terms are assumed to be Gamma distributed. This model provides a closed-form full likelihood and score equations.

Chapter 2 discusses the effects of utilizing Rathbun et al.'s (2007) integral estimators of the form (1.4.3) on these score equations. Since the score equations involve a ratio of integrals, both of which need to be estimated, Waagepetersen's (2008) estimators of the form (1.4.4) prove to greatly complicate the proofs. They are therefore bypassed for this chapter's method. Under the Rathbun et al. integral estimation scheme and increasing sample and domain asymptotics, it will be shown that the estimators for both $\boldsymbol{\beta}$ and the variance of the u_i terms are consistent and asymptotically normal.

In chapter 3, the model (1.5.1) is addressed in a more straightforward and computationally efficient manner. This is particularly important since the score equations used in Chapter 2 may provide unstable estimates due to the ratio of estimated integrals. In Chapter 3, simplified forms of the estimating equations used in Chapter 2 are developed. These simplified equations are still unbiased, and equal in expected value to those in Chapter 2, but no longer contain integral ratios. Therefore, they are more computationally efficient and provide more stable estimates for the parameters. Under this simplified form, consistency and asymptotic normality are proven under sample asymptotics (without the domain restrictions in Chapter 2). This is done in a general manner that encompasses both Rathbun et al. (2007) and Waagepetersen's (2008) estimation methods. It is also noteworthy that the coefficient parameter estimates obtained in Chapter 3 are consistent even when the modulated process is not Poisson, as long as it follows some restrictions upon its first four moments. The baseline spread parameter estimates, however, require a Poisson process in order to be consistent. Finally, Chapter 3 will discuss estimating the baseline parameter u_i for each subject using Lee and Nelder's (1996) h-likelihood.

Chapter 4 contains a more flexible model that allows for inter-subject variability by placing random effects upon some of the covariate coefficients. In this case, the mixed-effect intensity is defined as

$$\lambda_i(t) = \exp\{\boldsymbol{\alpha}^T \mathbf{x}_i(t) + \boldsymbol{\beta}_i^T \mathbf{z}_i(t)\}; t \in A_{ni}, i = 1, \ldots, n,$$

where $\boldsymbol{\alpha}$ is the parameter vector of coefficients for the fixed effects and/or the mean random effects, $\boldsymbol{\beta}_i$ is parameter vector of the subject-specific random effects and is assumed to be $MVN(\mathbf{0}, \boldsymbol{\Sigma})$, and $\mathbf{z}_i(t)$ is a subset of the covariates from $\mathbf{x}_i(t)$ that have random effects by subject. A set of estimating equations that are related to Laplace approximations of the log-likelihood are derived and shown to provide consistent and asymptotically normal (in sample size) estimates for $\boldsymbol{\alpha}$ and $\boldsymbol{\Sigma}$ under both Rathbun et al. (2007) and Waagepetersen's (2008) estimation methods. Also, Chapter 4 will discuss estimation of the $\boldsymbol{\beta}_i$ terms using methods similar to Breslow and Clayton's (1993). The asymptotic properties of the $\boldsymbol{\beta}_i$ terms are also examined.

Chapter 5 will summarize the methods and findings of this dissertation. It will also provide an outline of future research that is called for in this area.

This dissertation seeks to fill a gap in the literature regarding hiearchical point processes by extending current methods of analyzing modulated Poisson process models to allow for more inter-subject variability. Furthermore, this research should improve statistical methods of analyzing EMA and other forms of repeated-event data. This knowledge could also be expanded to numerous other applications of point processes such as epidemiology and meteorology.

1.6 References

Affleck, G., Zautra, A., Tennen, H. & Armelia, S. (1999). Multilevel daily proces designs for consulting and clinical psychology: A preface for the perplexed. *Journal of Consulting and Clinical Psychology, 67*, 746–754.

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology, 54*, 579–616.

Breslow, N. E. & Clayton, D. G. (1993). Approximate inference in Generalized Linear Mixed Models. *Journal of the American Statistical Association, 88*(421), 9–25.

Cox, D. R. (1972). The statistical analysis of dependencies in point processes. In *Stochastic Point Processes*, P. A. W. Lewis (Ed.), 55–66. New York: Wiley.

Cressie, N. (1991). *Statistics for spatial data*. New York: Wiley.

Csikszentmihalyi, M., & Larson, R. (1987). Validity and reliability of the experience-sampling method. *Journal of Nervous and Mental Disease, 175*(9), 526–536.

DeVries, M. (1992). *The experience of psychopathology.* Cambridge: Cambridge University Press.

Johnson, D. H. (1996). Point process models of single-neuron discharges. *Journal of Computational Neuroscience, 3*(4), 275–299.

Lee, Y. & Nelder, J. A. (1996). Hierarchical generalized linear models. *Journal of the Royal Statistical Society, Series B (Methodological), 58*, 619–678.

Lehmann, E.L. (1959). *Testing Statistical Hypotheses*. New York: Wiley.

McCulloch, C. E. (1997). Maximimum likelihood algorithms for Generalized Linear Mixed Models. *Journal of the American Statistical Association, 92*(437), 162–170.

Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society: Series A (General), 135*(3), 370–384.

Northrop, P. (1998). A clustered spatial-temporal model of rainfall. *Proceedings of the Royal Society: Mathematical, Physical and Engineering Sciences, 454*(1975), 1875–1888.

Rathbun, S. L., Shiffman, S., & Gwaltney, C. J. (2007). Modelling the effects of partially observed covariates on Poisson process intensity. *Biometrika, 94*, 153–165.

Riis, J. et al. (2005). Ignorance of hedonic adaptation to hemodialysis: A study using ecological momentary assessment. *Journal of Experimental Psychology, 134*, 3–9.

Schwartz, J. E. & Stone, A. A. (2007). The analysis of real-time momentary data: A practical guide. In *The Science of Real-time Data Capture: Self-report in Health Research*, Stone, A. A., Shiffman, S. S., Atienza, A., & Nebeling, L. (Eds.), 76–113. New York: Oxford University Press USA.

Shiffman, S. et al. (2002). Immediate antecedents of cigarette smoking: An analysis from ecological momentary assessment. *Journal of Abnormal Psychology, 111*, 531–545.

Shiffman, S. & Stone, A. A. (1998). Ecological Momentary Assessment: A new tool for behavioral medicine research. In *Technology and Methods in Behavioral Medicine*, Krantz, D. S. & Baum, A. (Eds.), 117–132. Philadelphia: Lawrence Erlbaum Associates.

Smyth, J. et al. (2001). The use of ecological momentary assessment approaches in eating disorder research. *International Journal of Eating Disorders, 30*, 83–95.

Stone, A. A., Broderick, J. B., Porter, L., & Kaell, A. T. (1997). The experience of rheumatoid arthritis pain and fatigue: Examining momentary reports and correlates over one week. *Arthritis Care and Research, 10*, 185-193.

Stone, A. A., Broderick, J. E., Shiffman, S. S. & Schwartz, J.E. (2004). Understanding recall of weekly pain from a momentary assessment perspective: Absolute agreement, between- and within-person consistency, and judged change in weekly pain. *Pain, 107*(1–2), 61–69.

Stone, A. A. & Shiffman, S. (1994). Ecological Momentary Assessment (EMA) in behavioral medicine. *Annals of Behavioral Medicine, 16,* 203–209.

Stone, A. A. & Shiffman, S. (2002). Capturing momentary, self-report data: A proposal for reporting guidelines. *Annals of Behavioral Medicine, 24*, 236–243.

Stoyan, D. & Penttinen, A. (2000). Recent applications of point process methods in forestry statistics. *Statistical Science, 15*(1), 61–78.

Waagepetersen, R. (2008). Estimating functions for inhomogenous spatial point processes with incomplete covariate data. *Biometrika, 95*(2), 351–363.

Chapter 2

Estimator of the Intensity of a Modulated Poisson Process with a Gamma Prior[1]

2.1 Introduction

Ecological Momentary Assessment (EMA) is an emerging method of data collection that is focused upon obtaining information about events as they occur naturally in a subject's environment (Shiffman & Stone 1998; Stone & Shiffman 2002). It has been used to study topics such as coping (Stone et al. 1998), eating disorders (Smyth et al. 2001), and emotional adaptation to hemodialysis (Riis et al. 2005). This technique uses electronic devices to take repeated assessments over time of the immediate states of subjects in their natural environments. In doing so, it eliminates some of the biases associated with recall and studying subjects in a laboratory setting, and accordingly increases the "ecological validity" of the data (Shiffman & Stone 1998, p. 117). EMA allows for unbiased estimation of the mean psychological states of subjects through probability-based sampling of their mood and other variables over time. It also makes it possible to record the times of repeated behavioral events. By observing psychological states at the event times as well as times selected from a probabilitistic sampling scheme, the effects of one's state on event timing may be investigated. Poisson processes provide an appropriate model for the timing of discrete behavioral events, and this paper will expand upon their use by allowing for inter-subject variation that is not accounted for in current models.

The motivating dataset for this research comes from Shiffman et al.'s (2002) EMA study of smoking habits. In this study, a sample of 304 smokers used electronic diaries to record the

[1]Neustifter, B., S. L. Rathbun & S. Shiffman. Unsubmitted as of 6/15/09

exact times that they smoked. At randomly chosen smoking times, as well as other random times throughout the day, the subjects were audibly prompted by the electronic diaries to answer questions related to their mood and environment. Both subject-level covariates such as gender and nicotine dependence and time-varying covariates such as restlessness were assessed during this study.

It's been generally recommended that Hierarchical Linear Models (HLMs) be used to analyze EMA data (Affleck, Zautra, Tennen, & Armeli 1999; Bolger, Davis, & Rafaeli 2003; Schwartz & Stone 2007). Such multilevel modeling can capture the variability among subjects in baseline rates or response to time-varying covariate values; HLMs allow researchers to separate variation into within- and between-subject components. This can help us understand the interactions between subject-level and momentary covariates in the occurrence of events. However, HLMs are not designed to model the timing of repeated discrete events over an interval.

The times of repeated behavioral events may be viewed as a realization of a temporal point process (Rathbun, Shiffman, & Gwaltney 2006). Recently, Rathbun, Shiffman, and Gwaltney (2007) proposed models for the analysis of EMA longitudinal repeated-event data using a Poisson process. Their method, however, has the limitation that it treats all subjects as having equal baseline rates of event occurrence and equal changes in event rate based on changes in covariates. By adapting the methodology of hierarchical modeling, it is possible to add random effect terms that allow for such inter-subject variability while using point processes to model temporal events. The current paper intends to address the identical-baseline assumption of previous research by using the techniques of Hierarchical Generalized Linear Models (HGLMs) to place Gamma-distributed random effects upon the baseline rates of event occurrence among subjects. We will show that the paramter estimates from the estimated score equations of this model are consistent and asymptotically normal under some weak restrictions.

Section 2 will briefly describe the modulated Poisson process model for discrete behavioral events and how to estimate its parameters using random covariate sampling. Then the proposed estimators will be described and the conditions under which they are consistent and asymptotically normal specified in Section 3. Section 4 will discuss Shiffman et al.'s (2002) data in more detail and compare the results of this method with those of Rathbun et al. (2007) on the same subset of variables. Proofs of consistency and normality and the associated Lemmas, as well as expanded expressions for the variance of the estimates, are included in the Appendices.

2.2 Modulated Poisson Processes with Probability Sampling

A temporal point process models the occurrences of repeated events over time, and may be described by its *intensity*, defined as

$$\lambda(t) = \lim_{\delta \to 0} \frac{E(N[t, t+\delta])}{\delta},$$

where $N[t, t+\delta]$ is the number of events in the interval $[t, t+\delta]$. For a Poisson process (Cox 1972), the measure N is such that for any interval A, $N(A) \sim Poisson(\Lambda(A))$, where $\Lambda(A) = \int_A \lambda(t)dt$. Then, conditional on $N(A)$, the event locations are independently sampled from a distribution with density proportional to $\lambda(t)$.

A modulated Poisson process (Cox 1972) is a specific form of the Poisson process that is particularly useful for examining the impact of time-varying covariates on the rate of event occurrence. Here, the intensity $\lambda(t; \boldsymbol{\beta})$ has the log-linear form

$$\lambda(t; \boldsymbol{\beta}) = \exp\{\boldsymbol{\beta}^T \mathbf{x}(t)\},\ t \in A,$$

where $\mathbf{x}(t)$ is a vector of covariates, and $\boldsymbol{\beta}$ is the p-dimensional vector of coefficient parameters for these covariates. Thus, $\boldsymbol{\beta}$ describes the effects of the covariates on the intensity of the process, and, therefore, the rate of events. Specifically, one may interpret $\exp\{\boldsymbol{\beta}_i\}$ as a risk ratio, akin to the interpretation in survival analysis.

For the modulated Poisson process, the log-likelihood for a single individual observed on a study interval A is

$$\ell(\boldsymbol{\beta}) = \sum_{t \in X} \boldsymbol{\beta}^T \mathbf{x}(t) - \Lambda(A; \boldsymbol{\beta}), \tag{2.2.1}$$

where X is the set of event times, and

$$\Lambda(A; \boldsymbol{\beta}) = \int_A \lambda(t; \boldsymbol{\beta})\, dt. \tag{2.2.2}$$

Based on (2.2.1), the score equation is

$$\Psi_A(\boldsymbol{\beta}) = \sum_{t \in X} \mathbf{x}(t) - \Lambda^{(1)}(A; \boldsymbol{\beta}),$$

where

$$\Lambda^{(1)}(A; \boldsymbol{\beta}) = \int_A \mathbf{x}(t) \exp\{\boldsymbol{\beta}^T \mathbf{x}(t)\} dt. \tag{2.2.3}$$

To obtain the MLEs for the elements of $\boldsymbol{\beta}$, the values of $\mathbf{x}(t)$ must be known for the entire interval A. In many cases, this knowledge is unrealistic and/or impossible to obtain.

To address the problem of estimating the integral in (2.2.3), Rathbun et al. (2007) and Waagepetersen (2008) propose approaches based on sampling the covariates at random times $\{u \in D\}$, where D is the set of "dummy" assessments selected from a probability-based sampling design with inclusion densities $\pi(u), u \in A$. Note that for the purposes of this article, as in Rathbun et al. and Waagepetersen, we are treating the covariates as deterministic. This treatment is equivalent to conditioning upon the realized values of the covariates. Then an unbiased estimator for (2.2.2) may be substituted into the log-likelihood, yielding an estimated score function of the form

$$\hat{\Psi}_A(\boldsymbol{\beta}) = \sum_{t \in X} \mathbf{x}(t) - \hat{\Lambda}^{(1)}(A; \boldsymbol{\beta}),$$

where $\hat{\Lambda}^{(1)}(A; \boldsymbol{\beta})$ is a design-unbiased estimator of (2.2.3). Rathbun et al. (2007) take

$$\hat{\Lambda}^{(1)}(A; \boldsymbol{\beta}) = \sum_{u \in D} \frac{\mathbf{x}(u) \exp\{\boldsymbol{\beta}^T \mathbf{x}(u)\}}{\pi(u)}, \tag{2.2.4}$$

while Waagepetersen (2008) uses

$$\hat{\Lambda}^{(1)}(A;\boldsymbol{\beta}) = \sum_{t\in X} \frac{\mathbf{x}(t)\exp\{\boldsymbol{\beta}^T\mathbf{x}(t)\}}{\left[\pi(t)+\exp\{\boldsymbol{\beta}^T\mathbf{x}(t)\}\right]} + \sum_{u\in D} \frac{\mathbf{x}(u)\exp\{\boldsymbol{\beta}^T\mathbf{x}(u)\}}{\left[\pi(u)+\exp\{\boldsymbol{\beta}^T\mathbf{x}(u)\}\right]}.$$

The authors each show that the estimators obtained from solving these equations are consistent and have asymptotically normal distributions. Waagepetersen's results suggest that his estimator may be more efficient than Rathbun et al.'s since it uses more of the available information for integral estimation. However, Waagepetersen's estimator is correlated with the number of events $N(A)$. When a random effect term is introduced, this correlation leads to lengthy variance terms that are difficult to write in closed form, and a similarly complicated consistency proof. For this reason, Waagepetersen's estimator will not be used in this article, though its use in the proposed setting may be the subject of future research.

2.3 Random Baseline Intensity Model

Rathbun et al.'s approach (2007) assumes that the intercept or baseline intensity is identical across subjects. To relax this assumption, the following modified model is proposed for the intensity of the Poisson process for each subject i:

$$\lambda_i(t) = u_i \exp\{\boldsymbol{\beta}^T\mathbf{x}_i(t)\}; t \in A_{ni}, i = 1,\ldots,n,$$

where u_i is a random baseline smoking intensity and A_{ni} is the study region for subject i with a sample of size n. This random intensity $\lambda_i(t)$ yields a simple Cox process (Cox 1955). We assume that the u_i are independent and identically distributed as Gamma(α,γ), where

$$f(u_i;\alpha,\gamma) = \frac{\gamma^\alpha}{\Gamma(\alpha)} u_i^{\alpha-1}\exp(-\gamma u_i)$$

is used as the definition of the Gamma density. The Gamma is chosen as the baseline density since it is the conjugate for the Poisson process and thus allows a closed-form likelihood, as in the Poisson-Gamma model for HGLMs (see Lee & Nelder 1996). Note that if an intercept term is incuded in the covariate parameter vector $\boldsymbol{\beta}$, then we may assume an arbitrary mean for the density of the u_i's. In many similar cases (see e.g. Lawless 1987), the mean is assumed to be 1. However, in this particular application, it greatly simplifies the consistency proofs to assume the mean is γ^{-1} by setting α equal to 1.

Let $|T_n| = \sum_{i=1}^{n} |A_{ni}|$, where $|\cdot|$ denotes interval length. Then under this model, the log-likelihood for $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \gamma)^T$ (divided by $|T_n|$ for technical reasons) is

$$\begin{aligned} \ell_n(\boldsymbol{\theta}) &= \frac{n}{|T_n|} \ln\ \gamma + \sum_{i=1}^{n} \frac{1}{|T_n|} \Bigg[\sum_{t \in X_{ni}} \boldsymbol{\beta}^T \mathbf{x}_i(t) \\ &- (1 + N_i) \ln\{\gamma + \Lambda_i(A_{ni}; \boldsymbol{\beta})\} + \ln(1 + N_i) \Bigg], \end{aligned}$$

where N_i is the number of events assessed for the i^{th} subject in her/his respective study interval, X_{ni} is the set of event times for the ith individual when the sample size is n, and

$$\Lambda_i(A_{ni}; \boldsymbol{\beta}) = \int_{A_{ni}} \exp\{\boldsymbol{\beta}^T \mathbf{x}_i(t)\} dt.$$

Define the matrices

$$\begin{aligned} \mathbf{Z}_{ni}(\boldsymbol{\theta}) &= (\Lambda_i^{(1)}(A_{ni}, \boldsymbol{\beta})^T, 1)^T;\ i = 1, \ldots, n, \\ \mathbf{R}_{ni}(\boldsymbol{\theta}) &= \frac{1 + N_i}{\gamma + \Lambda_i(A_{ni}; \boldsymbol{\beta})} \mathbf{Z}_{ni}(\boldsymbol{\theta}), \end{aligned} \tag{2.3.1}$$

and

$$\mathbf{Q}_{ni}(\boldsymbol{\theta}) = \left(\sum_{t \in X_{ni}} \mathbf{x}_i(t)^T, \frac{1}{\gamma} \right)^T.$$

Then the score equations are

$$\mathbf{g}_n(\boldsymbol{\theta}) = \frac{1}{|T_n|} \sum_{i=1}^{n} \{\mathbf{Q}_{ni}(\boldsymbol{\theta}) - \mathbf{R}_{ni}(\boldsymbol{\theta})\}. \tag{2.3.2}$$

Since it is often not the case that the covariate vectors $\mathbf{x}_i(t)$ are known across the entire study interval, it is desirable to be able to estimate the integrals in (2.3.1). We will demonstrate that, under some weak restrictions, substituting the integral estimators used by Rathbun et al. (2007), of the form found in (2.2.4), into (2.3.2) will still provide consistent estimates of the true parameters $\boldsymbol{\theta}_0$ as sample size $n \to \infty$. Specifically, with random non-event assessments $\{u \in D_{ni}; i = 1, \ldots, n,\ \}$, where D_{ni} is the set of random assessments for subject i, selected according to the inclusion densities $\pi(u)$, we will show the consistency and asymptotic normality of the solution $\tilde{\boldsymbol{\theta}}_n$ to the approximated score equations

$$\hat{\mathbf{g}}_n(\boldsymbol{\theta}) = \frac{1}{|T_n|} \sum_{i=1}^{n} \{\mathbf{Q}_{ni}(\boldsymbol{\theta}) - \hat{\mathbf{R}}_{ni}(\boldsymbol{\theta})\}, \tag{2.3.3}$$

where $\hat{\mathbf{R}}_{ni}(\boldsymbol{\theta})$ is as defined in (2.3.1), but with $\Lambda_i(A_{ni};\boldsymbol{\beta})$ and $\Lambda_i^{(1)}(A_{ni};\boldsymbol{\beta})$ replaced by the estimators, respectively,

$$\hat{\Lambda}_i(A_{ni};\boldsymbol{\beta}) = \sum_{u \in D_{ni}} \frac{\exp\{\boldsymbol{\beta}^T \mathbf{x}_i(u)\}}{\pi(u)}$$

and $\hat{\Lambda}_i^{(1)}(A_{ni};\boldsymbol{\beta}) = (\partial/\partial\boldsymbol{\beta})\hat{\Lambda}_i(A_{ni};\boldsymbol{\beta})$.

The following assumptions are required to prove the consistency of the proposed estimator $\tilde{\boldsymbol{\theta}}_n$:

Assumption 1. *The subjects $i = 1, \ldots, n$ are independently sampled, and all event times are independent of all random assessment times, both within and between subjects.*

Assumption 2. *There exists a constant M such that $|\mathbf{x}_i(t)| < M$ for almost all $t > 0$ and all $i = 1, \ldots, n$.*

Assumption 3. *Assume $\min_i |A_{ni}| \to \infty$ as $n \to \infty$, and there exists a constant $r > 0$ such that*

$$\frac{\min_i |A_{ni}|}{\max_i |A_{ni}|} \geq r$$

for all n.

Assumption 4.

$$\frac{1}{|T_n|} \sum_{i=1}^{n} \int_{A_{ni}} \mathbf{x}_i(t)\mathbf{x}_i(t)^T dt \to B$$

as $|T_n| \to \infty$, where the minimum eigenvalue of B is strictly positive, for all n, $i = 1, \ldots, n$.

Assumption 5. *The random covariate assessments $\{u \in D_{ni}; i = 1, \ldots, n\}$ are obtained according to a known probability-based sampling scheme with single and pairwise inclusion densities $\pi_i(u)$ and $\pi_i(u, v)$ (Cordy 1993), where the sampling intensity is such that*

$$\lim_{n\to\infty} \frac{m_i}{|A_{ni}|} = \rho$$

for some m_i *is the number of points in* D_{ni}, $0 < \rho < \infty$ *and* $i = 1, \ldots, n$.

Assumption 6.

$$\lim_{|A_{ni}|\to\infty} \inf_{s\in A_{ni}} \pi_i(s) = c$$

for some constant $c > 0$, *for any* n, $i = 1, \ldots, n$.

Assumption 7.

$$\frac{1}{|A_{ni}|^2} \int_{A_{ni}} \int_{A_{ni}} \left| \frac{\pi_i(s,u) - \pi_i(s)\pi_i(u)}{\pi_i(s)\pi_i(u)} \right| ds\, du \to 0$$

as $|A_{ni}| \to \infty$ *for any* n, $i = 1, \ldots, n$.

These assumptions are very similar to those posed by Rathbun et al. (2007). Assumption 1 is necessary to ensure that there is zero inter-subject correlation and independence between event assessments and random assessments. Assumption 2 is necessary to ensure that the log-likelihood is well-defined, and that $\hat{\Lambda}_i(A_{ni}; \boldsymbol{\beta})$ is an unbiased estimator of $\Lambda_i(A_{ni}; \boldsymbol{\beta})$ (Cordy 1993). This assumption merely requires that the covariates do not expand off to infinitely large values across subjects; this is a reasonable assumption for most realistic applications. Assumption 3 requires that all the study intervals increase to infinity in a well-behaved fashion, so that the results are not dominated by a finite number of subjects with too-large study intervals. Assumption 4 ensures that $(\partial^2/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T)\Lambda_i(A_{ni}; \boldsymbol{\beta})$ is asymptotically positive definite for all i.

For Assumption 5, it is sufficient for the random assessment times to be determined by a probabilistic sampling scheme that has an equal expected frequencies of assessments per subject (or per subject/stratum combination, for a stratified sampling design) per unit time. For example, taking random assessments from a simple random uniform distribution over the assessment time with some set number c of expected assessments per subject(/stratum) satisfies this requirement. Assumption 6 is required for the random assessments to have adequate coverage of the assessment regions. Assumption 7 (with Assumption 6) is necessary to ensure the variance of $\hat{\Lambda}_i(A_{ni}; \boldsymbol{\beta})$ is bounded. As mentioned by Rathbun et al. (2007), one sampling design that meets these assumptions is simple random sampling, using a uniform distribution over the assessment region.

Under these assumptions, we can prove that $\tilde{\boldsymbol{\theta}}_n$ is consistent:

Theorem 1. *Let $\tilde{\boldsymbol{\theta}}_n$ be the vector of coefficients that solves (2.3.3). Then under Assumptions 1 through 7, $\tilde{\boldsymbol{\theta}}_n$ is a consistent estimator of $\boldsymbol{\theta_0} = (\boldsymbol{\beta_0}^T, \gamma_0)^T$ as $n \to \infty$.*

The proof of this Theorem may be found in Appendix A. Moreover, Theorem 2 demonstrates the asymptotic normality of $\tilde{\boldsymbol{\theta}}_n$.

Theorem 2. *Under Assumptions 1 through 7,*

$$|T_n|^{\frac{1}{2}}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0}) \xrightarrow{D} \mathbf{N}(\mathbf{0}, \Gamma(\boldsymbol{\theta_0})),$$

as $n \to \infty$, where

$$\Gamma(\boldsymbol{\theta_0}) = \{\mathbf{J}_\infty(\boldsymbol{\theta_0})\}^{-1}\mathbf{I}_\infty(\boldsymbol{\theta_0})\{\mathbf{J}_\infty(\boldsymbol{\theta_0})\}^{-1} + \{\mathbf{J}_\infty(\boldsymbol{\theta_0})\}^{-1}\boldsymbol{\Sigma}_\infty(\boldsymbol{\theta_0})\{\mathbf{J}_\infty(\boldsymbol{\theta_0})\}^{-1},$$

with $\mathbf{J}_\infty(\boldsymbol{\theta}) = \lim_{n\to\infty} E_{\boldsymbol{\theta_0}}\{(\partial/\partial\boldsymbol{\theta}^T)\mathbf{g}_n(\boldsymbol{\theta})\}$, $\mathbf{I}_\infty(\boldsymbol{\theta})$ is the limit as $n \to \infty$ of the variance/covariance matrix of $|T_n|^{\frac{1}{2}}\mathbf{g}_n(\boldsymbol{\theta})$, and $\boldsymbol{\Sigma}_\infty(\boldsymbol{\theta})$ is the limit as $n \to \infty$ of the variance/covariance matrix of $|T_n|^{-\frac{1}{2}}\sum_{i=1}^n\{\mathbf{R}_{ni}(\boldsymbol{\theta}) - \hat{\mathbf{R}}_{ni}(\boldsymbol{\theta})\}$.

The proof of this theorem may be found in Appendix B.

As in Rathbun et al. (2007), the variance of our estimator $\tilde{\boldsymbol{\theta}}_n$ can be decomposed into two parts: $\{\mathbf{J}_\infty(\boldsymbol{\theta_0})\}^{-1}\mathbf{I}_\infty(\boldsymbol{\theta_0})\{\mathbf{J}_\infty(\boldsymbol{\theta_0})\}^{-1}$, the variance due to the model; and the variance due to the random sampling and integral estimation, $\{\mathbf{J}_\infty(\boldsymbol{\theta_0})\}^{-1}\boldsymbol{\Sigma}_\infty(\boldsymbol{\theta_0})\{\mathbf{J}_\infty(\boldsymbol{\theta_0})\}^{-1}$. The exact expressions for $\mathbf{J}_\infty(\boldsymbol{\theta}), \mathbf{I}_\infty(\boldsymbol{\theta})$, and $\boldsymbol{\Sigma}_\infty(\boldsymbol{\theta})$, and their estimation are detailed in Appendix C.

2.4 Results

We will illustrate the proposed method using data from an EMA of smoking by Shiffman et al. (2002). These data were collected from 304 smokers. Subjects were given small electronic diaries and were instructed to record on the diary any time that they smoked a cigarette (hereafter referred to as an “event”). At randomly selected events, as well as random times sampled according to a probability-based sampling design, the diary prompted subjects to complete assessments. Among the covariates sampled in this assessment were the variables of interest in this study: Negative Affect, Arousal, Attention, and Restlessness. Negative Affect, Arousal, and Attention were calculated according to a factor analysis: Negative Affect was based on subjects’ responses to various negative mood adjectives such as “irritable,” “miserable,” and “frustrated;” Arousal was based on adjectives such as “tired” and “energetic.” and Attention was determined by subjects’ reports of difficulty concentrating. Restlessness was based on a single item of the assessment which did not factor into Negative Affect or Arousal. These are the same four variables considered by Rathbun et al. (2007).

The non-event random assessments were sampled according to a stratified (by day) sampling design, with the restriction that random assessments could not occur within 10 minutes after an event and could only occur when the diary was on and active (not in stand-by while the subject was driving or in a meeting, for example).

Whenever an event was logged in the diary, the diary randomly prompted for an assessment according to a Bernoulli process with a probability that was constant throughout the day. An important note regarding Poisson processes is the following well-known result: The series of points obtained by randomly deleting each point of a Poisson process via a Bernoulli trial with known probability of deletion is also a Poisson process (Cressie 1991). This result is of particular importance for the Shiffman et al. (2002) data, as each event was assessed according to a certain known probability. This thinned process of points, then, is also a modulated Poisson process, and so the results of Rathbun et al. (2007) and Waagepetersen (2008), as well as those obtained in this paper, are applicable.

The probability of an event being assessed was determined by the number of events for the subject on the previous day, and was calculated to target an expected five event assessments per day. Specifically, for subject i, each of the events on day j was chosen to be assessed with probability

$$p_{ij} = \min\{5N_{i,j-1}^{-1}, 1\},$$

where $N_{i,j-1}$ was the number of events for subject i on day $j-1$.

Estimates of the parameters $\tilde{\boldsymbol{\theta}}_n$ were obtained by solving the estimated score equations $\hat{\mathbf{g}}_n(\boldsymbol{\theta}) = 0$, where $\hat{\mathbf{g}}_n(\boldsymbol{\theta})$ are as defined in (2.3.3), with

$$\mathbf{Q}_{ni}(\boldsymbol{\theta}) = \sum_{t \in X_{nij}} \left(\mathbf{x}_i(t), \frac{1}{\gamma}\right)^T,$$

and

$$\hat{\mathbf{R}}_{ni}(\boldsymbol{\theta}) = \sum_{j=1}^{d_i} \frac{1 + N_{ij}}{\gamma + \hat{\Lambda}_{ij}(\boldsymbol{\beta})} \left(\hat{\Lambda}_{ij}^{(1)}(\boldsymbol{\beta}), 1\right)^T,$$

with $n = 304$ subjects, d_i being the number of days for which subject i participated in the study, N_{ij} the number of events for subject i on day j, and X_{nij} the set of event assessments for subject i on day j. In this case, we define

$$\hat{\Lambda}_{ij}(\boldsymbol{\beta}) = \frac{p_{ij}|L_{ij}|}{m_{ij}} \sum_{u \in D_{nij}} \exp\{\boldsymbol{\beta}^T \mathbf{x}_i(u)\}$$

and $\hat{\Lambda}_{ij}^{(1)}(\boldsymbol{\beta}) = (\partial/\partial\boldsymbol{\beta})\hat{\Lambda}_{ij}(\boldsymbol{\beta})$, with D_{nij} the set of random assessments for subject i on day j and $|L_{ij}|$ the length of time that the electronic diary was active. Variances were estimated using the formulae in Appendix C.

Table 2.1 provides a summary of the results obtained by using the method of Rathbun et al. (2007), as reported in their article. Table 2.2 provides a summary of the results when the current random baseline intensity method is used.

Table 2.1: Results of Rathbun et al. (2007)

Covariate	Estimate	Model Variation	Sampling Variation	SE	z	p-value
Intercept	-0.05924	0.000069	0.000001	0.00839		
Negative Affect	0.01950	0.000075	0.000041	0.01077	1.81	0.0702
Arousal	-0.01594	0.000070	0.000046	0.01078	-1.48	0.1392
Attention	-0.01787	0.000093	0.000051	0.01198	-1.49	0.1358
Restlessness	0.21017	0.000127	0.000122	0.01577	13.33	<0.0001

Table 2.2: Results of Random Baseline Intensity Model

Covariate	Estimate	Model Variation	Sampling Variation	SE	z	p-value
Intercept	0.20434	0.006083	0.000174	0.07910		
Negative Affect	0.056038	0.001057	0.003777	0.06953	0.81	0.2101
Arousal	-0.02442	0.000957	0.002442	0.05830	-0.42	0.3377
Attention	-0.03566	0.000810	0.002310	0.05586	-0.64	0.2616
Restlessness	0.28190	0.004553	0.013816	0.13553	2.08	0.0188
Gamma	0.99935	0.042186	0.000027	0.20546		

Table 2.3: Estimated Risk Ratios and 95% CIs Using Current Method

Covariate	Estimated Risk Ratio	95% Confidence Interval
Negative Affect	1.05764	(0.92290, 1.21205)
Arousal	0.97588	(0.87051, 1.09400)
Attention	0.96311	(0.85819, 1.08086)
Restlessness	1.32565	(1.01639, 1.72901)

Clearly, the random baseline intensity method yields estimates of greater magnitude for the variables of interest, but with increased variance, as well. As with Rathbun et al.'s (2007) model, only Restlessness is shown to have a significant effect on the occurrence of smoking events. However, unlike Rathbun et al.'s model, Negative Affect no longer has a marginally significant effect, but rather has similar magnitude to Arousal and Attention.

The estimated risk ratios for all variables and the associated 95% confidence intervals, all calculated using the current method, are shown in Table 2.3. These estimates may be interpreted similarly to survival analysis. For example, an increase of 1 unit in a subject's Restlessness is associated with an increased smoking risk of 32.6%.

Since γ_0 is inversely related to the variance of baseline rate occurrence among subjects, the estimate of nearly 1 for $\tilde{\gamma}_n$ suggests that there is moderate variance in smoking rates among the subjects in the sample.

These results agree with those of Rathbun et al. (2007) and Shiffman et al. (2002) in that they find Restlessness to be one of the most significant predictors of smoking rates. However, it appears as though the marginal significance of Negative Affect in Rathbun et al. may have been a result of inter-subject noise, as it is much less significant and smaller in magnitude under this model. These results may be taken as indicative of the over-restrictive nature of Rathbun et al.'s (2007) model, as a model that ignores the overdispersion of data relative to the assumed Poisson process would underestimate the standard errors of the estimates, inflating their perceived significance.

Appendix A

The following Lemmas are necessary to prove Theorem 1:

Lemma 1. *Under Assumptions 1 through 7,* $|T_n|^{-1}\sum_{i=1}^{n}\{\mathbf{R}_{ni}(\boldsymbol{\theta})-\hat{\mathbf{R}}_{ni}(\boldsymbol{\theta})\}\xrightarrow{p}\mathbf{0}$ *as* $n\to\infty$*.*

Proof of Lemma 1. We can write

$$|A_{ni}|^{-1}\mathbf{R}_{ni}(\boldsymbol{\theta})=\frac{|A_{ni}|^{-1}\{1+N_i\}}{|A_{ni}|^{-1}\{\gamma+\hat{\Lambda}_i(A_{ni};\boldsymbol{\beta})\}}|A_{ni}|^{-1}\hat{\mathbf{Z}}_{ni}(\boldsymbol{\theta}).$$

Then note that $|A_{ni}|^{-1}\{1+N_i\}$, $|A_{ni}|^{-1}\{\gamma+\hat{\Lambda}_i(A_{ni};\boldsymbol{\beta})\}$, and $|A_{ni}|^{-1}\hat{\mathbf{Z}}_{ni}(\boldsymbol{\theta})$ have respective expected values $|A_{ni}|^{-1}\{1+\frac{1}{\gamma_0}\Lambda_i(A_{ni};\boldsymbol{\beta_0})\}$, $|A_{ni}|^{-1}\{\gamma+\Lambda_i(A_{ni};\boldsymbol{\beta})\}$, and $|A_{ni}|^{-1}\mathbf{Z}_{ni}(\boldsymbol{\theta})$, and variances that converge pointwise to 0 as $n\to\infty$ by Assumptions 1, 2, 4, 5, 6, and 7. So $\hat{\mathbf{R}}_{ni}(\boldsymbol{\theta})$ converges to a constant by Slutsky's Theorem. It can similarly be shown that $\mathbf{R}_{ni}(\boldsymbol{\theta})$ converges in a probability to the same constant. Then $|A_{ni}|^{-1}\{\mathbf{R}_{ni}(\boldsymbol{\theta})-\hat{\mathbf{R}}_{ni}(\boldsymbol{\theta})\}$ is $o_p(1)$ for all i, and so $n^{-1}\sum_{i=1}^{n}|A_{ni}|^{-1}\{\hat{\mathbf{R}}_{ni}(\boldsymbol{\theta})-\mathbf{R}_{ni}(\boldsymbol{\theta})\}$ is also $o_p(1)$. Then since $|T_n|^{-1}\leq\frac{1}{rn}\sum_{i=1}^{n}\frac{1}{max_i|A_{ni}|}$ by Assumption 3, the Lemma immediately follows. $\square$

Lemma 2. *Let* $\mathbf{A}$ *and* $\mathbf{C}$ *be* $p\times p$ *matrices and* $b>0$ *a scalar. If* $\mathbf{A}$ *is positive definite, then for any real* $\gamma>0$ *and any* $\mathbf{v}\in\boldsymbol{R}^p$*,*

$$\mathbf{v}^T\left(\frac{\mathbf{A}}{b}-\frac{\mathbf{C}\mathbf{C}^T}{b^2}\right)\mathbf{v}>0\quad\textit{implies}\quad\mathbf{v}^T\left(\frac{\mathbf{A}}{b+\gamma}-\frac{\mathbf{C}\mathbf{C}^T}{(b+\gamma)^2}\right)\mathbf{v}>0.$$

Proof of Lemma 2.

$$\mathbf{v}^T\left(\frac{\mathbf{A}}{b}-\frac{\mathbf{C}\mathbf{C}^T}{b^2}\right)\mathbf{v}>0 \quad \text{implies} \quad \mathbf{v}^T(b\mathbf{A}-\mathbf{C}\mathbf{C}^T)\mathbf{v}>0. \tag{A.1}$$

Then

$$\mathbf{v}^T\left(\frac{\mathbf{A}}{b+\gamma}-\frac{\mathbf{C}\mathbf{C}^T}{(b+\gamma)^2}\right)\mathbf{v}=(b+\gamma)^{-2}\mathbf{v}^T(b\mathbf{A}-\mathbf{C}\mathbf{C}^T+\gamma\mathbf{A})\mathbf{v}, \tag{A.2}$$

with $\mathbf{v}^T(b\mathbf{A}-\mathbf{C}\mathbf{C}^T)\mathbf{v}>0$ from (A.1), and $\gamma(\mathbf{v}^T\mathbf{A}\mathbf{v})>0$ since $\gamma>0$ and $\mathbf{A}$ is positive definite. Therefore, (A.2) is non-negative, completing the proof. □

To prove Theorem 1, we will use Crowder's (1986) results that if estimating equations $\mathbf{g}_n(\boldsymbol{\theta})$ satisfy the following three conditions, then the solution $\hat{\boldsymbol{\theta}}_n$ to $\mathbf{g}_n(\boldsymbol{\theta})=0$ is a consistent estimator of the true value $\boldsymbol{\theta}_\mathbf{0}$:

Condition 1. $\mathbf{g}_n(\boldsymbol{\theta})$ *are continuous*

Condition 2. *For n sufficiently large and some $\delta>0$,*
$\inf_{\partial S(\boldsymbol{\theta}_\mathbf{0},\,\epsilon)}(\boldsymbol{\theta}_\mathbf{0}-\boldsymbol{\theta})^T E_{\boldsymbol{\theta}_\mathbf{0}}\{\mathbf{g}_n(\boldsymbol{\theta})\}\geq\delta$

Condition 3. $\sup_{\partial S(\boldsymbol{\theta}_\mathbf{0},\,\epsilon)}\|\mathbf{g}_n(\boldsymbol{\theta})-E_{\boldsymbol{\theta}_\mathbf{0}}\{\mathbf{g}_n(\boldsymbol{\theta})\}\|\xrightarrow{p}0$ *as $n\to\infty$.*

Proof of Theorem 1. Clearly, (2.3.3) satisfies Condition 1.

For Condition 2, note that

$$\hat{\mathbf{g}}_n(\boldsymbol{\theta})=\mathbf{g}_n(\boldsymbol{\theta})+|T_n|^{-1}\sum_{i=1}^n\{\mathbf{R}_{ni}(\boldsymbol{\theta})-\hat{\mathbf{R}}_{ni}(\boldsymbol{\theta})\}.$$

From Lemma 1, $|T_n|^{-1}\sum_{i=1}^n\{\mathbf{R}_{ni}(\boldsymbol{\theta})-\hat{\mathbf{R}}_{ni}(\boldsymbol{\theta})\}\xrightarrow{p}\mathbf{0}$ as $n\to\infty$. Also, from Assumptions 2 and 3, $|T_n|^{-1}\sum_{i=1}^n\{\mathbf{R}_{ni}(\boldsymbol{\theta})-\hat{\mathbf{R}}_{ni}(\boldsymbol{\theta})\}$ is bounded. Then, by Theorem 4.1.4 of Chung (1968), for all $\boldsymbol{\theta}\in\partial S(\boldsymbol{\theta}_\mathbf{0}\,\epsilon)$ and all $\epsilon>0$, $E_{\boldsymbol{\theta}_\mathbf{0}}\left[|T_n|^{-1}\sum_{i=1}^n\{\mathbf{R}_{ni}(\boldsymbol{\theta})-\hat{\mathbf{R}}_{ni}(\boldsymbol{\theta})\}\right]\to\mathbf{0}$ as $n\to\infty$, implying $E_{\boldsymbol{\theta}_\mathbf{0}}(\hat{\mathbf{g}}_n(\boldsymbol{\theta})-\mathbf{g}_n(\boldsymbol{\theta}))\to\mathbf{0}$ as $n\to\infty$.

Note that (see e.g. Sándor 1996) $\inf_x\sum_{i=1}^n f_i(x)\geq\sum_{i=1}^n\inf_x f_i(x)$. Then

$$\inf_{\partial S(\boldsymbol{\theta}_\mathbf{0},\epsilon)}(\boldsymbol{\theta}_\mathbf{0}-\boldsymbol{\theta})^T E_{\boldsymbol{\theta}_\mathbf{0}}\{\mathbf{g}_n(\boldsymbol{\theta})\geq\inf_{\partial S(\boldsymbol{\theta}_\mathbf{0},\epsilon)}(\boldsymbol{\beta}_\mathbf{0}-\boldsymbol{\beta}^T)E_{\boldsymbol{\theta}_\mathbf{0}}\{\mathbf{A}_n(\boldsymbol{\theta})\}+\inf_{\partial S(\boldsymbol{\theta}_\mathbf{0},\epsilon)}(\gamma_0-\gamma)E_{\boldsymbol{\theta}_\mathbf{0}}\{B_n(\boldsymbol{\theta})\}, \tag{A.3}$$

where

$$\mathbf{A}_n(\boldsymbol{\theta}) = \frac{1}{|T_n|}\sum_{i=1}^{n}\left(\sum_{t\in X_{ni}}\mathbf{x}_i(t) - \frac{1+N_i}{\gamma+\Lambda_i(A_{ni};\boldsymbol{\beta})}\Lambda_i^{(1)}(A_{ni};\boldsymbol{\beta})\right)$$

and

$$B_n(\boldsymbol{\theta}) = \frac{1}{|T_n|}\sum_{i=1}^{n}\left(\frac{1}{\gamma} - \frac{1+N_i}{\gamma+\Lambda_i(A_{ni};\boldsymbol{\beta})}\right).$$

So the condition holds for $\mathbf{g}_n(\boldsymbol{\theta})$ if it holds for $\mathbf{A}_n(\boldsymbol{\theta})$ and $B_n(\boldsymbol{\theta})$ individually. Since $E_{\boldsymbol{\theta}_0}\{\mathbf{A}_n(\boldsymbol{\theta}_0)\} = 0$, we can take a Taylor Expansion of $\mathbf{A}_n(\boldsymbol{\theta})$ around $\boldsymbol{\theta}_0$ to obtain

$$\begin{aligned}
& (\boldsymbol{\beta}_0 - \boldsymbol{\beta})^T E_{\boldsymbol{\theta}_0}\{\mathbf{A}_n(\boldsymbol{\theta})\} \\
= \; & (\boldsymbol{\beta}_0 - \boldsymbol{\beta})^T \left\{\frac{\partial}{\partial \boldsymbol{\beta}^T} E_{\boldsymbol{\theta}_0}\{\mathbf{A}_n(\boldsymbol{\theta}_0)\}\right\}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + o\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|^2 \\
= \; & (\boldsymbol{\beta}_0 - \boldsymbol{\beta})^T \left[\frac{1}{|T_n|}\sum_{i=1}^{n}\frac{\gamma_0 + \Lambda_i(A_{ni};\boldsymbol{\beta}_0)}{\gamma_0}\left\{\frac{\Lambda_i^{(2)}(A_{ni};\boldsymbol{\beta}_0)}{\gamma_0 + \Lambda_i(A_{ni};\boldsymbol{\beta}_0)}\right.\right. \\
- \; & \left.\left.\frac{\Lambda_i^{(1)}(A_{ni};\boldsymbol{\beta}_0)\Lambda_i^{(1)}(A_{ni};\boldsymbol{\beta}_0)^T}{\{\gamma_0 + \Lambda_i(A_{ni};\boldsymbol{\beta}_0)\}^2}\right\}\right](\boldsymbol{\beta}_0 - \boldsymbol{\beta}) + o\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|^2
\end{aligned} \tag{A.4}$$

To show that the right-hand side of (A.4) is positive, we note that

$$\frac{\Lambda_i^{(2)}(A_{ni};\boldsymbol{\beta}_0)}{\Lambda_i(A_{ni};\boldsymbol{\beta}_0)} - \left\{\frac{\Lambda_i^{(1)}(A_{ni};\boldsymbol{\beta}_0)}{\Lambda_i(A_{ni};\boldsymbol{\beta}_0)}\right\}\left\{\frac{\Lambda_i^{(1)}(A_{ni};\boldsymbol{\beta}_0)}{\Lambda_i(A_{ni};\boldsymbol{\beta}_0)}\right\}^T$$

is positive definite under Assumption 4, since it is the variance/covariance matrix of $\mathbf{x}_i(t)$, where t is sampled from a distribution with density proportional to $\exp\{{\boldsymbol{\beta}_0}^T\mathbf{x}_i(t)\}$. For sufficiently large n, $\Lambda_i^{(2)}(A_{ni};\boldsymbol{\beta}_0)$ is positive definite by Assumption 4. The fact that the right-hand side of (A.4) is positive, then, follows from Lemma 2.

Taking a similar Taylor Expansion of $B_n(\boldsymbol{\theta})$ gives

$$(\gamma_0-\gamma)E_{\boldsymbol{\theta}_0}\{B_n(\boldsymbol{\theta})\} = (\gamma_0-\gamma)^2\left[\frac{1}{|T_n|}\sum_{i=1}^{n}\left\{-\frac{1}{\gamma_0^2} + \frac{1}{\gamma_0(\gamma_0 + \Lambda_i(A_{ni};\boldsymbol{\beta}_0))}\right\}\right] + o\|\gamma_0-\gamma\|^2. \tag{A.5}$$

The right-hand side of (A.5) is positive since $\Lambda_i(A_{ni};\boldsymbol{\beta}_0) \geq 0$, then $\gamma_0^{-2} \geq [\gamma_0\{\gamma_0 + \Lambda_i(A_{ni};\boldsymbol{\beta}_0)\}]^{-1}$ and so the second term must be non-negative. Then since we have shown that the condition is satisfied for $\mathbf{A}_n(\boldsymbol{\theta})$ and $B_n(\boldsymbol{\theta})$, then by (A.3), $\mathbf{g}_n(\boldsymbol{\theta})$ also satisfies Crowder's second condition. We have shown that $E_{\boldsymbol{\theta}_0}\{\hat{\mathbf{g}}_n(\boldsymbol{\theta}) - \mathbf{g}_n(\boldsymbol{\theta})\} \to \mathbf{0}$; then for n large enough, we can substitute $E_{\boldsymbol{\theta}_0}(\hat{\mathbf{g}}_n(\boldsymbol{\theta}))$ into (A.3), and Crowder's second condition follows.

To show that the third condition is also satisfied, we will first show that $\hat{\mathbf{g}}_n(\boldsymbol{\theta}) - E_{\boldsymbol{\theta}_0}\{\hat{\mathbf{g}}_n(\boldsymbol{\theta})\}$ converges pointwise to $\mathbf{0}$. Then we will show that the derivatives of this expression are bounded in probability, implying that the expression converges in probability. The third condition will then follow from Lemmas 2.2 and 3.2 of Crowder (1986), giving the theorem.

We use a Taylor Expansion to obtain

$$\begin{aligned}
& \hat{\mathbf{g}}_n(\boldsymbol{\theta}) - E_{\boldsymbol{\theta}_0}(\hat{\mathbf{g}}_n(\boldsymbol{\theta})) \\
= & \frac{1}{|T_n|}\sum_{i=1}^{n}[\{\mathbf{Q}_{ni}(\boldsymbol{\theta}) - \hat{\mathbf{R}}_{ni}(\boldsymbol{\theta})\} - E_{\boldsymbol{\theta}_0}\{\mathbf{Q}_{ni}(\boldsymbol{\theta})\} + E_{\boldsymbol{\theta}_0}\{\mathbf{R}_{ni}(\boldsymbol{\theta})\}] + o_p(1) \\
= & \frac{1}{|T_n|}\sum_{i=1}^{n}[\{\mathbf{Q}_{ni}(\boldsymbol{\theta}) - E_{\boldsymbol{\theta}_0}(\mathbf{Q}_{ni}(\boldsymbol{\theta}))\} + \{\mathbf{R}_{ni}(\boldsymbol{\theta}) - \hat{\mathbf{R}}_{ni}(\boldsymbol{\theta})\} \\
- & \{\mathbf{R}_{ni}(\boldsymbol{\theta}) - E_{\boldsymbol{\theta}_0}(\mathbf{R}_{ni}(\boldsymbol{\theta}))\}] - o_p(1).
\end{aligned}$$

Then

$$\begin{aligned}
\|\hat{\mathbf{g}}_n(\boldsymbol{\theta}) - E_{\boldsymbol{\theta}_0}(\hat{\mathbf{g}}_n(\boldsymbol{\theta}))\| \leq & \left\|\frac{1}{|T_n|}\sum_{i=1}^{n}[\mathbf{Q}_{ni}(\boldsymbol{\theta}) - E_{\boldsymbol{\theta}_0}\{\mathbf{Q}_{ni}(\boldsymbol{\theta})\}]\right\| \\
+ & \left\|\frac{1}{|T_n|}\sum_{i=1}^{n}\{\mathbf{R}_{ni}(\boldsymbol{\theta}) - \hat{\mathbf{R}}_{ni}(\boldsymbol{\theta})\}\right\| \\
+ & \left\|\frac{1}{|T_n|}\sum_{i=1}^{n}[\mathbf{R}_{ni}(\boldsymbol{\theta}) - E_{\boldsymbol{\theta}_0}\{\mathbf{R}_{ni}(\boldsymbol{\theta})\}]\right\| + o_p(1) \qquad \text{(A.6)}
\end{aligned}$$

Without respect to the norms, Lemma 1 shows that the second term converges in probability to 0. Then since, without respect to norms, the first and third terms of (A.6) have expected value 0 and so converge pointwise to 0, we only must show the derivatives of these terms are bounded in probability. The first term equals

$$\frac{1}{|T_n|}\sum_{i=1}^{n}\left\{\sum_{t\in X_{ni}}\left(\mathbf{x}_i(t), \frac{1}{\gamma}\right)^T - \left(\frac{1}{\gamma_0}\Lambda_i^{(1)}(A_{ni};\boldsymbol{\beta}_0), \frac{1}{\gamma}\right)^T\right\}.$$

The derivatives of this expression are zero, and so the suprimum of the norm of the first term of (A.6) over the boundary of $S(\boldsymbol{\theta}_0, \epsilon)$ converges in probability to 0. The third term has derivatives

$$\frac{\partial}{\partial\boldsymbol{\beta}}\frac{1}{|T_n|}\sum_{i=1}^{n}[\mathbf{R}_{ni}(\boldsymbol{\theta}) - E_{\boldsymbol{\theta}_0}\{\mathbf{R}_{ni}(\boldsymbol{\theta})\}] = \frac{1}{|T_n|}\sum_{i=1}^{n}\frac{N_i - \frac{1}{\gamma_0}\Lambda_i(A_{ni};\boldsymbol{\beta}_0)}{\{\gamma + \Lambda_i(A_{ni};\boldsymbol{\beta})\}^2}\mathbf{Z}_{ni}^{(\boldsymbol{\beta})}(\boldsymbol{\theta}) \qquad \text{(A.7)}$$

and

$$\frac{\partial}{\partial\gamma}\frac{1}{|T_n|}\sum_{i=1}^{n}[\mathbf{R}_{ni}(\boldsymbol{\theta}) - E_{\boldsymbol{\theta}_\mathbf{0}}\{\mathbf{R}_{ni}(\boldsymbol{\theta})\}] = \frac{\frac{1}{\gamma_0}\Lambda_i(A_{ni};\boldsymbol{\beta}_\mathbf{0}) - N_i}{\{\gamma + \Lambda_i(A_{ni};\boldsymbol{\beta})\}^2}\mathbf{Z}_{ni}(\boldsymbol{\theta}), \tag{A.8}$$

where

$$\begin{aligned}\mathbf{Z}_{ni}^{(\boldsymbol{\beta})}(\boldsymbol{\theta}) &= \Big(\Lambda_i^{(2)}(A_{ni};\boldsymbol{\beta})\{\gamma + \Lambda_i(A_{ni};\boldsymbol{\beta})\} - \Lambda_i^{(1)}(A_{ni};\boldsymbol{\beta})\Lambda_i^{(1)}(A_{ni};\boldsymbol{\beta})^T, \\ &\quad -\Lambda_i^{(1)}(A_{ni};\boldsymbol{\beta})^T\Big)^T.\end{aligned}$$

We can see that (A.7) satisfies Condition 3 by noting that, under Assumption 2,

$$\begin{aligned}&\frac{1}{|T_n|}\sum_{i=1}^{n}\frac{N_i - \frac{1}{\gamma_0}\Lambda_i(A_{ni};\boldsymbol{\beta}_\mathbf{0})}{\{\gamma + \Lambda_i(A_{ni};\boldsymbol{\beta}_\mathbf{0})\}^2}\mathbf{Z}_{ni,\boldsymbol{\beta}}^{(\boldsymbol{\beta})}(\boldsymbol{\theta}) \\ \leq\ &\frac{B_{u1}}{|T_n|}\sum_{i=1}^{n}\{N_i - \frac{1}{\gamma_0}\Lambda_i(A_{ni};\boldsymbol{\beta}_\mathbf{0})\} \\ +\ &\frac{B_{u2}}{|T_n|}\sum_{i=1}^{n}\{N_i - \frac{1}{\gamma_0}\Lambda_i(A_{ni};\boldsymbol{\beta}_\mathbf{0})\}\end{aligned} \tag{A.9}$$

for some constant matrices B_{u1} and B_{u2}. Then, since $E_{\boldsymbol{\beta}_\mathbf{0}}(N_i) - \gamma_0^{-1}\Lambda_i(A_{ni};\boldsymbol{\beta}_\mathbf{0}) = 0$ and by Assumptions 1 and 3 and the Weak Law of Large Numbers, the right-hand side of (A.9) converges in probability to 0. We can likewise show that (A.7) is bounded below by a quantity which also converges in probability to 0, proving that (A.7) must so converge. (A.8) can be shown to converge in probability to 0 in a similar manner. □

Appendix B

To prove asymptotic normality of the estimator $\tilde{\boldsymbol{\theta}}_n$, we require the following lemma:

Lemma 3. *Under Assumptions 1 through 7,*

$$|T_n|^{-\frac{1}{2}}\sum_{i=1}^{n}\{\mathbf{R}_{ni}(\boldsymbol{\theta}) - \hat{\mathbf{R}}_{ni}(\boldsymbol{\theta})\} \xrightarrow{D} \boldsymbol{N}(\mathbf{0}, \boldsymbol{\Sigma}_\infty(\boldsymbol{\theta})),$$

as n approaches infinity, where $\boldsymbol{\Sigma}_\infty(\boldsymbol{\theta})$ is a nonnegative definite matrix.

Similarly to Rathbun et al. (2007), it can be shown (omitted here) that under Assumptions 1 through 7, $\mathbf{R}_{ni}(\boldsymbol{\theta}) - \hat{\mathbf{R}}_{ni}(\boldsymbol{\theta})$ satisfies the uniformly bounded case of the CLT (see Ash 1972, p. 387), and thus Lemma 3 is true by the Lindeberg-Levy Central Limit Theorem.

Proof of Theorem 2. We can take a Taylor Expansion, giving

$$\hat{\mathbf{g}}_n(\tilde{\boldsymbol{\theta}}_n) = \hat{\mathbf{g}}_n(\boldsymbol{\theta}_0) + \hat{\mathbf{J}}_n(\boldsymbol{\theta}_0)(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{\mathbf{0}}) + \hat{\mathbf{K}}_n(\boldsymbol{\theta}^*)(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{\mathbf{0}}), \tag{B.1}$$

where

$$\hat{\mathbf{J}}_n(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}^T}\hat{\mathbf{g}}_n(\boldsymbol{\theta}), \tag{B.2}$$

$$\begin{aligned} \hat{\mathbf{K}}_n(\boldsymbol{\theta}) &= \left\{ \frac{1}{2}\sum_{l=1}^{p}(\tilde{\beta}_{n,l} - \beta_{0,l})\left(\frac{\partial^2}{\partial \beta_l \partial \boldsymbol{\beta}^T}\hat{\mathbf{g}}_n(\boldsymbol{\theta})\right)^T \right. \\ &+ \left. (\tilde{\gamma}_n - \gamma_0)\left(\frac{\partial^2}{\partial \gamma \partial \boldsymbol{\beta}^T}\hat{\mathbf{g}}_n(\boldsymbol{\theta})\right)^T, \frac{1}{2}(\tilde{\gamma}_n - \gamma_0)\left(\frac{\partial^2}{\partial \gamma^2}\hat{\mathbf{g}}_n(\boldsymbol{\theta})\right)^T \right\}^T, \end{aligned}$$

$\boldsymbol{\theta}^* = \alpha\tilde{\boldsymbol{\theta}}_n + (1-\alpha)\boldsymbol{\theta}_{\mathbf{0}}$ for some real $\alpha \in (0,1)$, $\tilde{\beta}_{n,l}$ is the lth element of $\tilde{\boldsymbol{\beta}}$ and likewise for $\beta_{0,l}$. Since $\tilde{\boldsymbol{\theta}}_n$ is the zero for $\hat{\mathbf{g}}_n$, we can rearrange and rewrite (B.1) to obtain

$$|T_n|^{\frac{1}{2}}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{\mathbf{0}}) = -\{\hat{\mathbf{J}}_n(\boldsymbol{\theta}_{\mathbf{0}}) + \hat{\mathbf{K}}_n(\boldsymbol{\theta}^*)\}^{-1}|T_n|^{\frac{1}{2}}\hat{\mathbf{g}}_n(\boldsymbol{\theta}_0) \tag{B.3}$$

Under Assumptions 1 through 4 and 6, all of the partial derivatives in $\hat{\mathbf{K}}(\boldsymbol{\theta}^*)$ are bounded by some finite constant for all l and n. Furthermore, under Assumptions 1 through 7, $\tilde{\boldsymbol{\theta}}_n$ is a consistent estimator of $\boldsymbol{\theta}_0$ as n approaches infinity. Then as $n \to \infty$,

$$\hat{\mathbf{K}}_n(\boldsymbol{\theta}^*) \xrightarrow{p} 0.$$

Note that, for $\mathbf{J}_n(\boldsymbol{\theta})$ being defined as $E_{\boldsymbol{\theta}_{\mathbf{0}}}\{(\partial/\partial\boldsymbol{\theta}^T)\mathbf{g}_n(\boldsymbol{\theta})\}$, we can prove that

$$\hat{\mathbf{J}}_n(\boldsymbol{\theta}_{\mathbf{0}}) - \mathbf{J}_n(\boldsymbol{\theta}_{\mathbf{0}}) \xrightarrow{p} 0 \tag{B.4}$$

as $n \to \infty$ by similar methods to Lemma 1. Then $\hat{\mathbf{J}}_n(\boldsymbol{\theta}_{\mathbf{0}}) + \hat{\mathbf{K}}_n(\boldsymbol{\theta}^*) \xrightarrow{p} \mathbf{J}_\infty(\boldsymbol{\theta}_{\mathbf{0}})$, where

$$\mathbf{J}_\infty(\boldsymbol{\theta}_{\mathbf{0}}) = \lim_{n\to\infty} \mathbf{J}_n(\boldsymbol{\theta}_{\mathbf{0}}).$$

This limit exists by Assumptions 1 through 3 and 5 through 7.

Also note that

$$|T_n|^{\frac{1}{2}}\hat{\mathbf{g}}_n(\boldsymbol{\theta}_0) = |T_n|^{\frac{1}{2}}\mathbf{g}_n(\boldsymbol{\theta}_0) + |T_n|^{-\frac{1}{2}}\sum_{i=1}^{n}\{\mathbf{R}_{ni}(\boldsymbol{\theta}_{\mathbf{0}}) - \hat{\mathbf{R}}_{ni}(\boldsymbol{\theta}_{\mathbf{0}})\}. \tag{B.5}$$

By the Lindeberg-Feller Central Limit Theorem, whose conditions are satisfied by Assumptions 1 through 3,

$$|T_n|^{\frac{1}{2}}\mathbf{g}_n(\boldsymbol{\theta}_0) \xrightarrow{D} \mathbf{N}(0, \mathbf{I}_\infty(\boldsymbol{\theta}_\mathbf{0})), \tag{B.6}$$

where $\mathbf{I}_\infty(\boldsymbol{\theta}_\mathbf{0})$ is as defined in the theorem.

Expressions (B.5) and (B.6), combined with Lemma 3 and the fact that the two terms on the right-hand side of (B.5) have zero asymptotic covariance, give

$$|T_n|^{\frac{1}{2}}\hat{\mathbf{g}}_n(\boldsymbol{\theta}_\mathbf{0}) \xrightarrow{D} \mathbf{N}(\mathbf{0}, \mathbf{I}_\infty(\boldsymbol{\theta}_\mathbf{0}) + \boldsymbol{\Sigma}_\infty(\boldsymbol{\theta}_\mathbf{0})).$$

This result, with Lemma 3, (B.3) and (B.4), implies the Theorem. □

Appendix C

It can be shown that $\mathbf{J}_\infty(\boldsymbol{\theta}_\mathbf{0}) = \lim_{n\to\infty} E_{\boldsymbol{\theta}_\mathbf{0}}\{(\partial/\partial\boldsymbol{\theta}^T)\mathbf{g}_n(\boldsymbol{\theta}_\mathbf{0})\}$ is of the form

$$\mathbf{J}_\infty(\boldsymbol{\theta}_\mathbf{0}) = \left(\begin{array}{c|c} J_{11,\infty}(\boldsymbol{\theta}_\mathbf{0}) & J_{12,\infty}(\boldsymbol{\theta}_\mathbf{0}) \\ \hline J_{12,\infty}(\boldsymbol{\theta}_\mathbf{0})^T & J_{22,\infty}(\boldsymbol{\theta}_\mathbf{0}) \end{array}\right),$$

with

$$J_{11,\infty}(\boldsymbol{\theta}_\mathbf{0}) = \lim_{n\to\infty}\frac{1}{|T_n|}\sum_{i=1}^{n}\left(\frac{\Lambda_i^{(1)}(A_{ni};\boldsymbol{\beta}_\mathbf{0})\Lambda_i^{(1)}(A_{ni};\boldsymbol{\beta}_\mathbf{0})^T}{\gamma_0\{\gamma_0 + \Lambda_i(A_{ni};\boldsymbol{\beta}_\mathbf{0})\}} - \frac{\Lambda_i^{(2)}(A_{ni};\boldsymbol{\beta}_\mathbf{0})}{\gamma_0}\right), \tag{C.1}$$

$$J_{12,\infty}(\boldsymbol{\theta}_\mathbf{0}) = \lim_{n\to\infty}\frac{1}{|T_n|}\sum_{i=1}^{n}\frac{\Lambda_i^{(1)}(A_{ni};\boldsymbol{\beta}_\mathbf{0})}{\gamma_0\{\gamma_0 + \Lambda_i(A_{ni};\boldsymbol{\beta}_\mathbf{0})\}}, \tag{C.2}$$

and

$$J_{22,\infty}(\boldsymbol{\theta}_\mathbf{0}) = \lim_{n\to\infty}\frac{1}{|T_n|}\sum_{i=1}^{n}\left(-\frac{1}{\gamma_0^2} + \frac{1}{\gamma_0\{\gamma_0 + \Lambda_i(A_{ni};\boldsymbol{\beta}_\mathbf{0})\}}\right). \tag{C.3}$$

By Assumptions 2 and 3, $|T_n|^{-1}\Lambda_i(A_{ni};\boldsymbol{\beta})$ is bounded above and below by some real constants B_u and B_l, respectively, for all $i = 1, \ldots, n$. Similarly, all of its derivatives, and continuous functions of Λ_i (such as its reciprocal) can likewise bounded with weighting by area length. Then by Assumption 1, the Weak Law of Large Numbers, and the fact that events are independent of random assessments for each subject, all the above limits converge to finite quantities.

Similarly, $\mathbf{I}_\infty(\boldsymbol{\theta_0}) = \lim_{n\to\infty} Var_{\boldsymbol{\theta_0}}(|T_n|^{\frac{1}{2}}\mathbf{g}_n(\boldsymbol{\theta_0}))$, can be written

$$\mathbf{I}_\infty(\boldsymbol{\theta_0}) = \left(\begin{array}{c|c} I_{11,\infty}(\boldsymbol{\theta_0}) & I_{12,\infty}(\boldsymbol{\theta_0}) \\ \hline I_{12,\infty}(\boldsymbol{\theta_0})^T & I_{22,\infty}(\boldsymbol{\theta_0}) \end{array}\right),$$

with

$$I_{11,\infty}(\boldsymbol{\theta_0}) = \lim_{n\to\infty} \frac{1}{|T_n|}\sum_{i=1}^{n}\left[\frac{\Lambda_i^{(2)}(A_{ni};\boldsymbol{\beta_0})}{\gamma_0} - \frac{\Lambda_i^{(1)}(A_{ni};\boldsymbol{\beta_0})\Lambda_i^{(1)}(A_{ni};\boldsymbol{\beta_0})^T}{\gamma_0\{\gamma_0+\Lambda_i(A_{ni};\boldsymbol{\beta_0})\}}\right],$$

$$I_{12,\infty}(\boldsymbol{\theta_0}) = \lim_{n\to\infty} -\frac{1}{|T_n|}\sum_{i=1}^{n}\frac{\Lambda_i^{(1)}(A_{ni};\boldsymbol{\beta_0})}{\gamma_0\{\gamma_0+\Lambda_i(A_{ni};\boldsymbol{\beta_0})\}},$$

and

$$I_{22,\infty}(\boldsymbol{\theta_0}) = \lim_{n\to\infty} \frac{1}{|T_n|}\sum_{i=1}^{n}\frac{\Lambda_i(A_{ni};\boldsymbol{\beta_0})}{\gamma_0^2\{\gamma_0+\Lambda_i(A_{ni};\boldsymbol{\beta_0})\}}.$$

These converge to finite quantities for the same reasons as (C.1) through (C.3).

Finally, consider the matrix

$$\boldsymbol{\Sigma}_\infty(\boldsymbol{\theta_0}) = \lim_{n\to\infty} Var_{\boldsymbol{\theta_0}}\left[|T_n|^{-\frac{1}{2}}\sum_{i=1}^{n}\{\mathbf{R}_{ni}(\boldsymbol{\theta_0}) - \hat{\mathbf{R}}_{ni}(\boldsymbol{\theta_0})\}\right].$$

First, define $L_{ni}(\boldsymbol{\beta}) = Var\{\hat{\Lambda}_{ni}(A_{ni};\boldsymbol{\beta})\}$ and $L_{ni}^{(1)}(\boldsymbol{\beta}) = Var\{\hat{\Lambda}_{ni}^{(1)}(A_{ni};\boldsymbol{\beta})\}$. From Cordy (1993), we can compute

$$\begin{aligned} L_{ni}^{(1)}(\boldsymbol{\beta}) &= \int_{A_{ni}} \frac{\mathbf{x}_i(s)\mathbf{x}_i(s)^T\exp\{2\boldsymbol{\beta}^T\mathbf{x}_i(s)\}}{\pi_i(s)}\,ds \\ &+ \int_{A_{ni}}\int_{A_{ni}} \mathbf{x}_i(s)\mathbf{x}_i(u)^T\exp\{\boldsymbol{\beta}^T(\mathbf{x}_i(s)+\mathbf{x}_i(u))\}\frac{\pi_i(s,u)-\pi_i(s)\pi_i(u)}{\pi_i(s)\pi_i(u)}\,ds\,du. \end{aligned}$$

$L_{ni}(\boldsymbol{\beta})$ is the upper left-hand element of $L_{ni}^{(1)}(\boldsymbol{\beta})$. Then, using Taylor approximations of the ratios of estimated integrals (see e.g. van Kempen & van Vlient 1999), we can then write, for n large enough,

$$\begin{aligned}
\boldsymbol{\Sigma}_\infty(\boldsymbol{\theta_0}) &= \lim_{n\to\infty}\sum_{i=1}^{n}\Bigg[E_{\boldsymbol{\theta_0}}\{(1+N_i)^2\}Var_{\boldsymbol{\theta_0}}\bigg(\frac{\hat{\mathbf{Z}}_{ni}(\boldsymbol{\theta_0})}{\gamma_0+\hat{\Lambda}_i(A_{ni};\boldsymbol{\beta_0})} \\
&- \frac{\mathbf{Z}_{ni}(\boldsymbol{\theta_0})}{\gamma_0+\Lambda_i(A_{ni};\boldsymbol{\beta_0})}\bigg)\Bigg] \\
&= \lim_{n\to\infty}\sum_{i=1}^{n}\Bigg[\bigg(\frac{\Lambda_i(A_{ni};\boldsymbol{\beta_0})}{\gamma_0}+2\frac{\Lambda_i(A_{ni};\boldsymbol{\beta_0})^2}{\gamma_0^2}\bigg)\mathbf{V}_{ni}(\boldsymbol{\theta_0})\Bigg],
\end{aligned}$$

where

$$\mathbf{V}_{ni}(\boldsymbol{\theta_0}) = \left(\begin{array}{c|c} V_{ni}^{11}(\boldsymbol{\theta_0}) & V_{ni}^{12}(\boldsymbol{\theta_0}) \\ \hline V_{ni}^{12}(\boldsymbol{\theta_0})^T & V_{ni}^{22}(\boldsymbol{\theta_0}), \end{array}\right)$$

with

$$\begin{aligned}
V_{ni}^{11}(\boldsymbol{\theta}) &= \frac{L_{ni}^{(1)}(\boldsymbol{\beta_0})}{(\gamma_0+\Lambda_i(A_{ni};\boldsymbol{\beta_0}))^2}+\frac{L_{ni}(\boldsymbol{\beta_0})\Lambda_i^{(1)}(A_{ni};\boldsymbol{\beta_0})\Lambda_i^{(1)}(A_{ni};\boldsymbol{\beta_0})^T}{(\gamma_0+\Lambda_i(A_{ni};\boldsymbol{\beta_0}))^4} \\
&- 2\frac{\Lambda_i^{(1)}(A_{ni};\boldsymbol{\beta_0})L_{ni}^C(\boldsymbol{\beta_0})}{(\gamma_0+\Lambda_i(A_{ni};\boldsymbol{\beta_0}))^3},
\end{aligned}$$

$$V_{ni}^{12}(\boldsymbol{\theta}) = \frac{\Lambda_i(A_{ni};\boldsymbol{\beta_0})}{L_{ni}(\boldsymbol{\beta_0})+(\gamma_0+\Lambda_i(A_{ni};\boldsymbol{\beta_0}))^2}-\frac{\Lambda_i(A_{ni};\boldsymbol{\beta_0})}{(\gamma_0+\Lambda_i(A_{ni};\boldsymbol{\beta_0}))^2}$$

$$V_{ni}^{22}(\boldsymbol{\theta}) = \frac{L_{ni}(\boldsymbol{\theta})}{(\gamma+\Lambda_i(A_{ni};\boldsymbol{\beta}))^4}.$$

One can estimate $\mathbf{J}_\infty(\boldsymbol{\theta_0})$, $\mathbf{I}_\infty(\boldsymbol{\theta_0})$ and $\boldsymbol{\Sigma}_\infty(\boldsymbol{\theta_0})$ by replacing $\boldsymbol{\theta_0}$ with $\tilde{\boldsymbol{\theta}}_n$, removing the limits, and using the estimates $\hat{\Lambda}_i(A_{ni};\tilde{\boldsymbol{\beta}})$ and derivatives for $\Lambda_i(A_{ni};\boldsymbol{\beta_0})$ and derivatives. Care must be taken to adjust for bias when estimating the $\Lambda_i(A_{ni};\boldsymbol{\beta_0})$ and derivative terms multiplied together or raised to powers. For example,

$$E_{\boldsymbol{\beta_0}}[\{\tilde{\gamma}_n+\hat{\Lambda}_i(A_{ni};\tilde{\boldsymbol{\beta}}_n)\}^2] = L_{ni}(\boldsymbol{\beta_0})+\{\gamma_0+\Lambda_i(A_{ni};\boldsymbol{\beta_0})\}^2.$$

One should then use the unbiased estimators $[\{\tilde{\gamma}_n+\hat{\Lambda}_i(A_{ni};\tilde{\boldsymbol{\beta}}_n)\}^2-L_{ni}(\tilde{\boldsymbol{\beta}}_n)]$ or $[\tilde{\gamma}^2\{1+\frac{3}{2}N_i+\frac{1}{2}(N_i)^2\}]$ for $\{\gamma_0+\Lambda_i(A_{ni};\boldsymbol{\beta_0})\}^2$ rather than simply substituting $\{\tilde{\gamma}+\hat{\Lambda}_i(A_{ni};\tilde{\boldsymbol{\beta}}_n)\}^2$.

2.5 References

Affleck, G., Zautra, A., Tennen, H. & Armelia, S. (1999). Multilevel daily proces designs for consulting and clinical psychology: A preface for the perplexed. *Journal of Consulting and Clinical Psychology, 67*, 746–754.

Ash, R. B. (1972). *Real Analysis and Probability.* New York: Academic Press.

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology, 54*, 579–616.

Chung, K. L. (1968). *A course in probability theory.* New York: Harcourt, Brace, & World.

Cordy, C. B. (1993). An extension of the Horvitz-Thompson theorem to point sampling from a continuous universe. *Statistics & Probability Letters, 18*, 353–362.

Cox, D. R. (1955). Some statistical methods connected with series of events. *Journal of the Royal Statistical Society B, 17*, 129–164.

Cox, D. R. (1972). The statistical analysis of dependencies in point processes. In *Stochastic Point Processes*, P. A. W. Lewis (Ed.), 55–66. New York: Wiley.

Cressie, N. (1991). *Statistics for spatial data*. New York: Wiley.

Crowder, M. (1986). On consistency and inconsistency of estimating equations. *Econometric Theory, 2*(3), 305–330.

Lawless, J. F. (1987). Regression methods for poisson process data. *Journal of the American Statistical Association, 82*, 808–815.

Lee, Y. & Nelder, J. A. (1996). Hierarchical generalized linear models. *Journal of the Royal Statistical Society, Series B (Methodological), 58*, 619–678.

Rathbun, S. L., Shiffman, S., & Gwaltney, C. J. (2006). Point process models for event history data: Applications in behavioral science. In *Models for intensive longitudinal data*, T. A. Walls & J. L. Schafer (Eds.), 219–253. New York: Oxford University Press.

Rathbun, S. L., Shiffman, S., & Gwaltney, C. J. (2007). Modelling the effects of partially observed covariates on Poisson process intensity. *Biometrika, 94*, 153–165.

Riis, J. et al. (2005). Ignorance of hedonic adaptation to hemodialysis: A study using ecological momentary assessment. *Journal of Experimental Psychology, 134*, 3–9.

Sándor, J. (1996). On an inequality for the sum of infimums of functions. *Journal of Mathematical Analysis and Applications, 204*, 646–654.

Schwartz, J. E. & Stone, A. A. (2007). The analysis of real-time momentary data: A practical guide. In A. A. Stone, J. E. Shiffman, A. Atienza & L. Nebeling (Eds.), *The science of real-time data capture: Self-report in health research* (76–113). New York: Oxford University Press USA.

Shiffman, S. et al. (2002). Immediate antecedents of cigarette smoking: An analysis from ecological momentary assessment. *Journal of Abnormal Psychology, 111*, 531–545.

Shiffman, S. & Stone, A. A. (1998). Ecological Momentary Assessment: A new tool for behavioral medicine research. In *Technology and Methods in Behavioral Medicine*, Krantz, D. S. & Baum, A. (Eds.), 117–132. Philadelphia: Lawrence Erlbaum Associates.

Smyth, J. et al. (2001). The use of ecological momentary assessment approaches in eating disorder research. *International Journal of Eating Disorders, 30*, 83–95.

Stone, A. A. et al. (1998). A comparison of coping assessed by ecological momentary assessment and retrospective recall. *Journal of Personality and Social Psychology, 74*, 1670–1680.

Stone, A. A. & Shiffman, S. (1994). Ecological Momentary Assessment (EMA) in behavioral medicine. *Annals of Behavioral Medicine, 16*, 203–209.

Stone, A. A. & Shiffman, S. (2002). Capturing momentary, self-report data: A proposal for reporting guidelines. *Annals of Behavioral Medicine, 24*, 236–243.

van Kempen, G. M. P. & van Vliet, L. J. (2000). Mean and variance of ratio estimators used in florescence ratio imaging. *Cytometry, 39*, 300–305.

Waagepetersen, R. (2008). Estimating functions for inhomogenous spatial point processes with incomplete covariate data. *Biometrika, 95*(2), 351–363.

Chapter 3

Baseline Intensity in Modulated Poisson Processes: A Technique for Analyzing Ecological Momentary Assessment Data[1]

3.1 Introduction

Ecological Momentary Assessment (EMA) is an emerging method of data collection in psychological research (Shiffman & Stone 1998, Stone & Shiffman 2002) that uses electronic devices to record information about each subject's psychological states and repeated behavioral events as they occur in the subject's natural environment. EMA reduces recall biases in results and avoids effects introduced by an artificial laboratory environment, thereby increasing the "ecological validity" of the data (Shiffman & Stone 1998, p. 117). Previous uses of EMA include the study of eating disorders (Smyth et al. 2001), stress coping (Stone et al. 1998), and pain (Stone, Broderick, Porter & Kaell 1997; Stone, Broderick, Shiffman & Schwartz 2004), to list a few. By utilizing probability-based sampling of subjects' psychological states and other variables, EMA allows for the unbiased estimation of these covariates' mean states. Furthermore, by additionally assessing subjects at event times, this method makes it possible to estimate the effects of the mean states of the covariates upon the timing of the events of interest.

The motivating dataset comes from Shiffman et al.'s (2002) study of the smoking behavior of a sample of 304 smokers. Subjects recorded the exact times that they smoked on electronic diaries and answered questions regarding their mood (e.g. restlessness, boredom) and environment (e.g. whether other smokers were present) at randomly selected smoking times

[1]Neustifter, B., S. L. Rathbun & S. Shiffman. Unsubmitted as of 6/15/09

and other randomly selected times during the day. In addition to the time-varying covariates of mood and environment, subject-level covariates such as gender and age were also recorded. By analyzing these data, we hope to gain insight regarding the impact of covariates on cigarette consumption rates in smokers; this will contribute to our understanding of what prompts smoking and may assist in designing programs to help smokers quit. As one can view repeated behavioral events as being timed according to a temporal point process (Rathbun, Shiffman, & Gwaltney 2006), EMA repeated-event data may be modeled using a modulated Poisson process (Rathbun, Shiffman, & Gwaltney 2007). However, this method implicitly assumes that all subjects have the same baseline rate of event occurrence. Therefore, the current article proposes a method to allow for inter-subject variability in baseline smoking rates while maintaining simple estimation of the covariate parameters and their associated variance. This is accomplished using the techniques of Hierarchical Generalized Linear Models (HGLMs).

HGLMs are an extension of Generalized Linear Models (GLMs) (Nelder & Wedderburn 1972) that are used to model multi-level data, and so allow researchers to model several levels of variation within a model. Hierarchical Linear Models (HLMs) are a particular type of HGLMs, and have been recommended for analyzing EMA data (Affleck et al. 1999; Bolger, Davis, & Rafaeli 2003; Schwartz & Stone 2007), as they allow researchers to separate within- and between-subject variation. HLMs make the assumption of Normally distributed random components, however, while HGLMs allow the random copmonents to follow any exponential-family distribution. A particular HGLM of interest in this case is one in which the responses are Poisson distributed, the random effects are gamma distributed, and the link between the expected value of the responses and the linear model is the canonical log link. This conjugate HGLM is called the Poisson-Gamma model (Lee, Nelder, & Pawitan 2006). Since the focus of the current article is modeling repeated event times and not predicting responses, and HGLMs are not designed to model the timing of repeated discrete events, the theory of

HGLMs is not directly applicable. However, much of the work in the area of HGLMs can be carried over to this setting.

In Section 3.2, the modulated Poisson process will be described, and the current model with random baseline event terms will be defined. Preliminarily, the equations will be constructed in the case that the covariates are fully observed over all study intervals. Simple estimating equations will then be derived for the coefficient and baseline rate spread paramters. Finally, estimation of the equations and parameters using random assessments of covariates, rather than fully observing them, will be addressed in a general manner that encompasses the techniques of both Rathbun et al. (2007) and Waagepetersen (2008). Section 3.3 will give the conditions under which the estimators are consistent and asymptotically normal. Section 3.4 will discuss how to estimate the subjects' baseline rates using the h-likelihood (Lee & Nelder 1996). Section 5 will describe the motivating dataset in detail and demonstrate the use of the derived estimators on these data, comparing results with those received from Rathbun et al. (2007) and Waagepetersen's (2008) estimators. The Appendices contain the proofs of consistency and asymptotic normality of the estimators.

3.2 Random Baseline Intensity Model

A temporal point process is a model for generating the times of repeated events over a study interval. Such a process has a random counting measure N, where $N(A)$ is the number of events in the interval A, and can be partially described by its *intensity*,

$$\lambda(t) = \lim_{\delta \to 0} \frac{E(N[t, t+\delta])}{\delta}.$$

When the process is Poisson (Cox 1972), N is such that for any interval A, $N(A)$ is Poisson distributed with mean $\Lambda(A) = \int_A \lambda(t)dt$. Conditional upon $N(A)$, the events of a Poisson process are independently sampled from a probability density function proportional to $\lambda(t)$. In the case of a modulated Poisson process (Cox 1972), $\lambda(t)$ is a function of some vector of covariates $\mathbf{x}(t)$, some or all of which may vary over time. The intensity takes the form

$$\lambda(t; \boldsymbol{\beta}) = \exp\{\boldsymbol{\beta}^T \mathbf{x}(t)\}; t \in A, \tag{C.1}$$

where $\boldsymbol{\beta}$ is the p-dimensional vector of coefficient parameters for the covariates. The parameter $\boldsymbol{\beta}$, then, describes the effects of the covariates on the rate of events by its impact upon the intensity. Similar to survival analysis, one may interpret $\exp\{\beta_i\}$ as a risk ratio; a unit increase of the ith element of $\mathbf{x}(t)$ is associated with an increase or decrease in risk of immediate event occurrence of size $\exp\{\beta_i\}$.

We can generalize this form of modeling temporal series of events to a study such as the Shiffman et al. (2002) dataset by modeling the events for each subject by (C.1). Doing so, however, assumes that the subjects have equal baseline event occurrence rates and identical reactions to changes in the covariate vector $\mathbf{x}(t)$. Obviously, such an assumption is likely unrealistic, particularly when the events are behavioral; subjects in the study may have inherently (due to genetic or other factors) different rates of event occurrence.

To permit inter-subject variability in the model, we will allow the baseline event occurence rate to vary randomly among individuals. Specifically, we will take the intensity of the Poisson process for subject i to be

$$\lambda_i(t) = u_i \exp\{\boldsymbol{\beta}^T \mathbf{x}_i(t)\}; t \in A_{ni}, i = 1, \ldots, n, \tag{C.2}$$

where u_i is the random baseline intensity and A_{ni} is the study region for the ith subject when the sample size is n. Note that this random intensity defines a simple Cox process (Cox 1955). We assume that the u_i in (C.2) are independently sampled from a distribution with mean 1, variance σ^2, and finite third and fourth moments. Note that, if an intercept term is included in the parameter vector $\boldsymbol{\beta}$, the first moment, if not equal to 1, can be absorbed into the intercept term. A particular distribution which satisfies these assumptions would be a gamma distribution with shape parameter σ^{-2} and scale parameter σ^2. A gamma distribution is particularly appropriate to this setting, as it is conjugate for the Poisson process and therefore allows for a closed-form marginal likelihood. However, as the consistency of our estimator depends only upon the first four moments of u_i, any distribution meeting the above assumptions will suffice.

3.2.1 Statistical Inference: Fully-Observed Covariates

To motivate the methods for statistical inference, we will temporarily assume for this subsection that the covariates $\mathbf{x}_i(t)$ are known for all $t \in A_{ni}$ and all $i = 1, \dots, n$; that is, that they are some known function of time. For the moment, also assume that the baseline rates u_i follow a gamma distribution. Then we obtain the marginal log-likelihood for $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2)^T$ (divided by n for technical reasons) to be

$$\begin{aligned} \ell_n(\boldsymbol{\theta}) &= n^{-1}\Bigg[n\sigma^{-2}\ln\sigma^{-2} - n\ln\Gamma(\sigma^{-2}) + \sum_{i=1}^{n}\Bigg\{\sum_{t\in X_{ni}}\boldsymbol{\beta}^T\mathbf{x}_i(t) \\ &- (\sigma^{-2}+N_i)\ln(\sigma^{-2}+\Lambda_i(A_{ni};\boldsymbol{\beta})) + \ln(\sigma^{-2}+N_i)\Bigg\}\Bigg], \end{aligned} \tag{C.3}$$

where

$$\Lambda_i(A_{ni};\boldsymbol{\beta}) = \int_{A_{ni}} \exp\{\boldsymbol{\beta}^T\mathbf{x}_i(t)\}\,dt,$$

N_i is the number of events occurring in the ith study interval, and X_{ni} is the set of all the times t where an event occurred for the ith subject. Other distributions may be assumed for the u_i terms, but the gamma allows for a closed-form marginal log-likelihood and gives score equations that help guide toward the suggested estimating equations. Based on this log-likelihood, the score equation for $\boldsymbol{\beta}$ is

$$\Psi_n(\boldsymbol{\theta}) = n^{-1}\sum_{i=1}^{n}\left\{\sum_{t\in X_{ni}}\mathbf{x}_i(t) - \frac{(\sigma^{-2}+N_i)\Lambda_i^{(1)}(A_{ni};\boldsymbol{\beta})}{\sigma^{-2}+\Lambda_i(A_{ni};\boldsymbol{\beta})}\right\}, \tag{C.4}$$

where $\Lambda_i^{(1)}(A_{ni};\boldsymbol{\beta}) = (\partial/\partial\boldsymbol{\beta}^T)\Lambda_i(A_{ni};\boldsymbol{\beta})$. Note that this score equation is the same as the equation that would be obtained from the marginal log-likelihood for the point process once the MHLE for $v_i = \ln(u_i)$ was inserted into the joint likelihood (see Section C.1 for more detail on h-likelihoods). The score equation for σ^2, while writable in closed-form, will not be used due to its complexity and so is omitted here.

Under some weak restrictions, slightly less general assumptions about the distribution of the u_i terms, and if (C.3) is divided by $|T_n| = \sum_{i=1}^{n} |A_{ni}|$ rather than n, it can be shown that the estimators obtained by substituting integral estimators of the type discussed in the following subsection into the score equations of (C.3) are consistent and asymptotically normal as n and $|T_n|$ approach infinity. The proofs and particularly the variance estimation, however, become quite complicated, due to the resulting ratios of estimated integrals in the score equations, such as the term $(\sigma^{-2} + N_i)\hat{\Lambda}_i^{(1)}(A_{ni}; \boldsymbol{\beta})[\sigma^{-2} + \hat{\Lambda}_i(A_{ni}; \boldsymbol{\beta})]^{-1}$ in (C.4). These integral ratios may also cause instability problems with the estimates. Therefore, instead of using the score equations from (C.3), we will now give an unbiased simplification of the equations for $\boldsymbol{\beta}$ and an alternate, more computationally efficient estimating equation for σ^2 that will avoid such ratios of integrals and the associated problems.

To obtain a simpler estimating equation for $\boldsymbol{\beta}$ that is still unbiased, we note that since $E_{\boldsymbol{\theta_0}}(N_i) = \Lambda_i(A_{ni}; \boldsymbol{\beta_0})$, the ratio term of (C.4), evaluated at the true parameter $\boldsymbol{\theta_0}$, has expected value

$$E_{\boldsymbol{\theta_0}}\left\{ \frac{(\sigma_0^{-2} + N_i)\Lambda_i^{(1)}(A_{ni}; \boldsymbol{\beta_0})}{\sigma_0^{-2} + \Lambda_i(A_{ni}; \boldsymbol{\beta_0})} \right\} = \Lambda_i^{(1)}(A_{ni}; \boldsymbol{\beta_0}).$$

Then an unbiased estimating equation for $\boldsymbol{\beta}$ can be obtained by substituting $\Lambda_i^{(1)}(A_{ni}; \boldsymbol{\beta})$ for the ratio in (C.4), yielding

$$\hat{\Psi}_n(\boldsymbol{\beta}) = n^{-1} \sum_{i=1}^{n} \left\{ \sum_{t \in X_{ni}} \mathbf{x}_i(t) - \Lambda_i^{(1)}(A_{ni}; \boldsymbol{\beta}) \right\}. \tag{C.5}$$

Note that this estimating equation does not involve the variance parameter σ^2, and that it is unbiased regardless of the distribution of u_i. The former is particularly helpful, as it allows us to treat σ^2 as a nuisance parameter to be estimated after a final estimate for $\boldsymbol{\beta}$ has been obtained, rather than simultaneously with $\boldsymbol{\beta}$. Furthermore, this equation allows us to have relaxed distributional assumptions on u_i, as this term no longer needs to come from a gamma distribution in order for our estimates of $\boldsymbol{\beta}$ to be consistent.

To obtain an unbiased estimate for σ^2, define

$$\phi_i(\boldsymbol{\beta_0}, t) = \exp\{-\boldsymbol{\beta_0}^T \mathbf{x}_i(t)\}, \tag{C.6}$$

and note

$$\begin{aligned}
& E_{\boldsymbol{\theta_0}}\left[n^{-1}\sum_{i=1}^{n}\left\{|A_{ni}|^{-2}\sum_{s\neq t\in X_{ni}}\phi_i(\boldsymbol{\beta_0}, s)\phi_i(\boldsymbol{\beta_0}, t) - 1\right\}\right] \\
= \; & n^{-1}\sum_{i=1}^{n}\left[\left(E_{\boldsymbol{\theta_0}}\{N_i^2 - N_i\}|A_{ni}|^{-2}\Lambda_i(A_{ni}; \boldsymbol{\beta_0})^{-2}\right)\right. \\
* \; & \left.\left(\int_{A_{ni}}\int_{A_{ni}}\phi_i(\boldsymbol{\beta_0}, s)\phi_i(\boldsymbol{\beta_0}, t)\exp\{\boldsymbol{\beta_0}^T\mathbf{x}_i(s)\}\exp\{\boldsymbol{\beta_0}^T\mathbf{x}_i(t)\}\, ds\, dt\right) - 1\right] \\
= \; & n^{-1}\sum_{i=1}^{n}\left[(1+\sigma_0^2) - 1\right] \\
= \; & \sigma_0^2,
\end{aligned}$$

where $|\cdot|$ denotes interval length. Then an unbiased estimator of σ^2 is

$$\hat{\sigma}_n^2 = n^{-1}\sum_{i=1}^{n}\left\{|A_{ni}|^{-2}\sum_{s\neq t\in X_{ni}}\phi_i(\boldsymbol{\beta_0}, s)\phi_i(\boldsymbol{\beta_0}, t) - 1\right\}. \tag{C.7}$$

Since we are treating σ^2 as a nuisance parameter, and the estimating equation for $\boldsymbol{\beta}$ does not involve σ^2, we may first estimate $\boldsymbol{\beta}$ separately, obtaining $\tilde{\boldsymbol{\beta}}_n$ and then use this estimate in place of $\boldsymbol{\beta_0}$ in (C.7). Note that, since $\hat{\sigma}_n^2$ depends upon the value of $\boldsymbol{\beta}$ used, it should technically be written as $\hat{\sigma}_n^2(\boldsymbol{\beta})$ to emphasize this fact; however, for notational simplicity we will not do so here. Care will be taken to clarify the value of $\boldsymbol{\beta}$ used in situations where confusion may arise. Finally, recognize that while the estimating equation for $\boldsymbol{\beta}$ remains unbiased even if the data are not realized from a Poisson process (Schoenberg 2005), $\hat{\sigma}_n^2$ is unbiased for σ^2 if and only if the second order intensity (Diggle 2003) satisfies $\lambda_2(s,t) = \lambda(s)\lambda(t)$.

3.2.2 Statistical Inference: Partially-Observed Covariates

To evaluate the function $\Lambda_i(A_{ni}; \boldsymbol{\beta})$ and derivatives necessary for estimating $\boldsymbol{\theta}$, $\mathbf{x}_i(t)$ must be a known function of $t \in A_{ni}$. This requirement is difficult or impossible in many applied settings. While the theory established in the previous subsection assumes that the covariates are a known function of time, we will now address the problem of obtaining estimates of the parameters with desirable asymptotic properties when this is not the case.

If we consider the study interval A_{ni} to be a population of points, then it should be possible to unbiasedly estimate the mean covariate state over the study interval by sampling from this point population according to some known probability (Rathbun et al. 2007). This random sampling of times for covariate assessment is a traditional feature of EMA, and so it is particularly useful to approach the problem from this perspective when analyzing EMA data. Rathbun et al. (2007) and Waagepetersen (2008) proposed estimators based upon values of $\mathbf{x}_i(t)$ for such randomly sampled times $\{t \in D_{ni} : i = 1, \ldots, n\}$, where D_{ni} is a set of "dummy," randomly sampled non-event time points from A_{ni} chosen according to the known inclusion densities $\pi_{ni}(t), t \in A_{ni}$. Note that, as in Rathbun et al. and Waagepetersen, we are treating the covariates as deterministic, which is equivalent to conditioning upon their realized values. By such random sampling from A_{ni}, we can obtain design-unbiased estimators for $\Lambda_i^{(1)}(A_{ni}; \boldsymbol{\beta})$ of the form

$$\hat{\Lambda}_i^{(1)}(A_{ni}; \boldsymbol{\beta}) = \sum_{t \in D_{ni}} \frac{\mathbf{x}_i(t) \exp\{\boldsymbol{\beta}^T \mathbf{x}_i(t)\}}{\pi_{ni}(t)} \tag{C.8}$$

for Rathbun et al. (2007), and

$$\hat{\Lambda}_i^{(1)}(A_{ni}; \boldsymbol{\beta}) = \sum_{t \in X_{ni} \cup D_{ni}} \frac{\mathbf{x}_i(t) \exp\{\boldsymbol{\beta}^T \mathbf{x}_i(t)\}}{\pi_{ni}(t) + \exp\{\boldsymbol{\beta}^T \mathbf{x}_i(t)\}} \tag{C.9}$$

for Waagepetersen (2008). Note that substituting either (C.8) or (C.9) into (C.5) gives weighted estimating equations of the form

$$\tilde{\Psi}_n(\boldsymbol{\beta}) = n^{-1} \sum_{i=1}^{n} \left\{ \sum_{t \in X_{ni}} \mathbf{x}_i(t) f(t; \boldsymbol{\beta}) - \sum_{t \in D_{ni}} \frac{\mathbf{x}_i(t) f(t; \boldsymbol{\beta}) \exp\{\boldsymbol{\beta}^T \mathbf{x}_i(t)\}}{\pi_{ni}(t)} \right\}, \tag{C.10}$$

where $f(t; \boldsymbol{\beta}) = 1$ for Rathbun et al.'s estimator, and $f(t; \boldsymbol{\beta}) = \pi_{ni}(t)[\pi_{ni}(t) + \exp\{\boldsymbol{\beta}^T \mathbf{x}_i(t)\}]^{-1}$ for Waagepetersen's.

As stated in subsection 3.2.1, it is possible to show that substituting integral estimators of the form (C.8) into (C.3) can yield consistent and asymptotically normal estimators of

$\boldsymbol{\beta_0}$ and σ_0^2 under certain conditions, but these estimators have complicated variance and difficult proofs due to ratios of estimated integrals. Furthermore, proving the asymptotic properties for substitutions using Waagepetersen's estimator is made difficult due to covariance between $\hat{\Lambda}_i(A_{ni};\boldsymbol{\beta})$, calculated similarly to (C.9), and the number of event occurrences N_i. The proposed estimating equations (C.10) and (C.7) (with $\tilde{\boldsymbol{\beta}}_n$ substituted) simplify the proofs necessary to show these properties for the Rathbun et al. (2007) substitution, and will also allow consistency and normality to be shown for the Waageptersen (2008) substitution case.

3.3 Asymptotic Properties

We must make some weak assumptions regarding the covariates, event and dummy points, and the inclusion densities π_{ni} to prove that the estimator $\tilde{\boldsymbol{\theta}}_n$ is consistent as $n \to \infty$ and that $\tilde{\boldsymbol{\beta}}_n$ is asymptotically normal.

Assumption 8. *The subjects $i = 1, \ldots, n$ are independently sampled, and all event times are independent of all random assessment times, both within and between subjects.*

Assumption 9. *There exists a constant M such that $|\,\mathbf{x}_i(t)| < M$ for almost all $t > 0$ and all $i = 1, \ldots, n$.*

Assumption 10.

$$n^{-1} \sum_{i=1}^{n} \int_{A_{ni}} \mathbf{x}_i(t)\mathbf{x}_i(t)^T dt \to B$$

as $n \to \infty$, where the minimum eigenvalue of B is strictly positive.

Assumption 11. *The random covariate assessments $\{u \in D_{ni};\, i = 1, \ldots, n\}$ are obtained according to a known probability-based sampling scheme with single and pairwise inclusion densities $\pi_{ni}(u)$ and $\pi_{ni}(u, v)$ (Cordy 1993), where the sampling intensity is such that*

$$\lim_{n\to\infty} \frac{m_i}{|A_{ni}|} = \rho$$

for some $0 < \rho < \infty$ and $i = 1, \ldots, n$.

Assumption 12.

$$\lim_{n\to\infty} \inf_{t\in A_{ni}} \pi_{ni}(t) = c$$

for some constant $c > 0$, for any $i = 1, \ldots, n$.

These assumptions are similar to those posed by Rathbun et al. (2007). Assumption 8 is necessary to ensure that there is no correlation either among subjects, or between events and dummy points within subjects. Assumption 9 ensures that the estimating equations are bounded, and also that $E\{\hat{\Lambda}_i^{(1)}(A_{ni};\boldsymbol{\beta})\} = \Lambda_i^{(1)}(A_{ni};\boldsymbol{\beta})$ (Cordy 1993). Under Assumption 10, $(\partial/\partial\boldsymbol{\beta}^T)\Lambda_i(A_{ni};\boldsymbol{\beta})$ will be positive definite for sufficiently large n; this assumption also ensures that there is no redundancy among the covariates. Assumptions 11 and 12 place some weak restrictions on the density for the sampling of the random dummy points. It is sufficient for Assumption 11 that the random assessments are realized from a probability sampling design where the expected number of assessments per unit time is equal across subjects or subject/stratum combinations. Assumption 12 ensures that the random assessments adequately cover the study region, and allows for unbiased esitmation of $\Lambda_i^{(1)}(A_{ni};\boldsymbol{\beta})$.

Under these assumptions, we can prove the consistency of the vector of solutions $\tilde{\boldsymbol{\theta}}_n$ to (C.10) and

$$\tilde{\sigma}_n^2(\tilde{\boldsymbol{\beta}}_{\boldsymbol{n}}) = n^{-1}\sum_{i=1}^{n}\left\{|A_{ni}|^{-2}\sum_{s\neq t\in X_{ni}}\phi_i(\tilde{\boldsymbol{\beta}}_n, s)\phi_i(\tilde{\boldsymbol{\beta}}_n, t) - 1\right\} \tag{C.1}$$

where $\phi_{ni}(\tilde{\boldsymbol{\beta}}_n, t)$ is defined as in (C.6). We can further show that the solutions $\tilde{\boldsymbol{\beta}}_n$ to (C.10) are asymptotically normal.

Theorem 3. *Under Assumptions 8 through 12, the solution $\tilde{\boldsymbol{\theta}}_n$ to (C.5) and (C.1) is a consistent estimator of the true parameter values $\boldsymbol{\theta}_{\mathbf{0}} = (\boldsymbol{\beta}_{\mathbf{0}}{}^T, \sigma_0^2)^T$ as $n \to \infty$.*

This proof may be found in Appendix A. We can further prove that $\tilde{\boldsymbol{\beta}}_n$ is asymptotically normal under these same conditions. To discuss the asymptotic variance of this estimator, define

$$\mathbf{J}_\infty(\boldsymbol{\beta}_{\mathbf{0}}) = \lim_{n\to\infty}\mathbf{J}_n(\boldsymbol{\beta}_{\mathbf{0}}), \tag{C.2}$$

$$\mathbf{\Sigma}_\infty^Q(\boldsymbol{\theta_0}) = \lim_{n\to\infty} \mathbf{\Sigma}_n^Q(\boldsymbol{\theta_0}),$$

and

$$\mathbf{\Sigma}_\infty^R(\boldsymbol{\beta_0}) = \lim_{n\to\infty} \mathbf{\Sigma}_n^R(\boldsymbol{\beta_0}),$$

where

$$\mathbf{J}_n(\boldsymbol{\beta}) = -n^{-1}\sum_{i=1}^n \left\{ \int_{A_{ni}} \mathbf{x}_i(t)\mathbf{x}_i(t)^T f(t;\boldsymbol{\beta}) \exp\{\boldsymbol{\beta}^T \mathbf{x}_i(t)\}\, dt \right\}, \tag{C.3}$$

$$\begin{aligned} & \mathbf{\Sigma}_n^Q(\boldsymbol{\theta}) \\ = \; & n^{-1}\sum_{i=1}^n \left\{ \sigma^2 \int_{A_{ni}} \int_{A_{ni}} \left(\mathbf{x}_i(s)\mathbf{x}_i(t)^T f(s;\boldsymbol{\beta}) f(t;\boldsymbol{\beta}) \exp\{\boldsymbol{\beta}^T(\mathbf{x}_i(s)+\mathbf{x}_i(t))\} \right) ds\, dt \right. \\ + \; & \left. \int_{A_{ni}} \mathbf{x}_i(t)\mathbf{x}_i(t)^T f(t;\boldsymbol{\beta})^2 \exp\{\boldsymbol{\beta}^T \mathbf{x}_i(t)\}\, dt \right\}, \end{aligned} \tag{C.4}$$

and

$$\begin{aligned} & \mathbf{\Sigma}_n^R(\boldsymbol{\beta}) \\ = \; & n^{-1}\sum_{i=1}^n \left\{ \int_{A_{ni}} \int_{A_{ni}} \mathbf{x}_i(s)\mathbf{x}_i(t)^T f(s;\boldsymbol{\beta}) f(t;\boldsymbol{\beta}) \exp\{\boldsymbol{\beta}^T(\mathbf{x}_i(s)+\mathbf{x}_i(t))\} \right. \\ * \; & \left(\frac{\pi_{ni}(s,t) - \pi_{ni}(s)\pi_{ni}(t)}{\pi_{ni}(s)\pi_{ni}(t)} \right) ds\, dt \\ + \; & \left. \int_{A_{ni}} \mathbf{x}_i(t)\mathbf{x}_i(t)^T f(t;\boldsymbol{\beta_0})^2 \exp\{2\boldsymbol{\beta}^T \mathbf{x}_i(t)\} \pi_{ni}(t)^{-1}\ dt \right\}. \end{aligned} \tag{C.5}$$

Here, $\mathbf{J}_\infty(\boldsymbol{\beta_0})$ is the asymptotic information matrix evaluated at the true parameter value, and $\mathbf{\Sigma}_\infty^Q(\boldsymbol{\theta_0})$ and $\mathbf{\Sigma}_\infty^R(\boldsymbol{\beta_0})$ are the asymptotic variances of the estimating equation due to the event assessments and the random assessments, respectively. With these defined, we can give the asymptotic distribution of $\tilde{\boldsymbol{\beta}}_n$.

Theorem 4. *Under Assumptions 8 through 12,*

$$n^{\frac{1}{2}}(\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta_0}) \xrightarrow{D} \mathbf{N}(\mathbf{0}, \mathbf{V}(\boldsymbol{\theta_0})),$$

where

$$\mathbf{V}(\boldsymbol{\theta_0}) = \{\mathbf{J}_\infty(\boldsymbol{\beta_0})\}^{-1}\mathbf{\Sigma}_\infty^Q(\boldsymbol{\theta_0})\{\mathbf{J}_\infty(\boldsymbol{\beta_0})\}^{-1} + \{\mathbf{J}_\infty(\boldsymbol{\beta_0})\}^{-1}\mathbf{\Sigma}_\infty^R(\boldsymbol{\beta_0})\{\mathbf{J}_\infty(\boldsymbol{\beta_0})\}^{-1},$$

with $\mathbf{J}_\infty(\boldsymbol{\beta_0})$, $\mathbf{\Sigma}_\infty^Q(\boldsymbol{\theta_0})$, and $\mathbf{\Sigma}_\infty^R(\boldsymbol{\beta_0})$ defined as in (C.3), (C.4), and (C.5), respectively.

The proof of this theorem may be found in Appendix B.

Let $\tilde{\boldsymbol{\theta}}_n^R$ and $\tilde{\boldsymbol{\theta}}_n^W$ be the estimates $\tilde{\boldsymbol{\theta}}_n$ obtained from, respectively, Rathbun et al. (2007) and Waagepetersen's (2008) estimation methods. Then under Rathbun et al.'s estimation technique, (C.3) and (C.4) may be approximated by

$$\hat{\mathbf{J}}_\infty(\tilde{\boldsymbol{\beta}}_n^R) = -n^{-1} \sum_{i=1}^n \left\{ \sum_{t \in D_{ni}} \frac{\mathbf{x}_i(t)\mathbf{x}_i(t)^T \exp\{(\tilde{\boldsymbol{\beta}}_n^R)^T \mathbf{x}_i(t)\}}{\pi_{ni}(t)} \right\}$$

and

$$\begin{aligned}\hat{\boldsymbol{\Sigma}}_\infty^Q(\tilde{\boldsymbol{\theta}}_n^R) &= n^{-1} \sum_{i=1}^n \left\{ \tilde{\sigma}_n^2 \sum_{s \neq t \in D_{ni}} \frac{\mathbf{x}_i(s)\mathbf{x}_i(t)^T \exp\{(\tilde{\boldsymbol{\beta}}_n^R)^T (\mathbf{x}_i(s) + \mathbf{x}_i(t))\}}{\pi_{ni}(s,t)} \right. \\ &+ \left. \sum_{t \in D_{ni}} \frac{\mathbf{x}_i(t)\mathbf{x}_i(t)^T \exp\{(\tilde{\boldsymbol{\beta}}_n^R)^T \mathbf{x}_i(t)\}}{\pi_{ni}(t)} \right\}.\end{aligned}$$

(C.5) is estimated similarly to (C.4). For the case of Rathbun et al.'s (2007) estimator, these estimations may be simplified. For example,

$$\tilde{\mathbf{J}}_\infty(\tilde{\boldsymbol{\beta}}_n^R) = -n^{-1} \sum_{i=1}^n \left\{ \sum_{t \in X_{ni}} \mathbf{x}_i(t)\mathbf{x}_i(t)^T \right\} \tag{C.6}$$

is an unbiased estimator of $\mathbf{J}_\infty(\boldsymbol{\beta}_\mathbf{0})$; this estimate does not depend upon the estimated value $\tilde{\boldsymbol{\beta}}_n^R$, which may lead to more stable estimates of the variance. Similarly, we can estimate $\boldsymbol{\Sigma}_\infty^Q(\boldsymbol{\theta}_\mathbf{0})$ and $\boldsymbol{\Sigma}_\infty^R(\boldsymbol{\beta}_\mathbf{0})$ with

$$\tilde{\boldsymbol{\Sigma}}_\infty^Q(\tilde{\boldsymbol{\theta}}_n^R) = n^{-1} \sum_{i=1}^n \left\{ \frac{(\tilde{\sigma}_n^R)^2}{1 + (\tilde{\sigma}_n^R)^2} \sum_{s \neq t \in X_{ni}} \mathbf{x}_i(s)\mathbf{x}_i(t)^T + \sum_{t \in X_{ni}} \mathbf{x}_i(t)\mathbf{x}_i(t)^T \right\}$$

and

$$\begin{aligned}\tilde{\boldsymbol{\Sigma}}_\infty^R(\tilde{\boldsymbol{\beta}}_n^R) &= n^{-1} \sum_{i=1}^n \left\{ \left(\sum_{t \in D_{ni}} \frac{\mathbf{x}_i(t) \exp\{(\tilde{\boldsymbol{\beta}}_n^R)^T \mathbf{x}_i(t)\}}{\pi_{ni}(t)} \right) \left(\sum_{t \in D_{ni}} \frac{\mathbf{x}_i(t) \exp\{(\tilde{\boldsymbol{\beta}}_n^R)^T \mathbf{x}_i(t)\}}{\pi_{ni}(t)} \right)^T \right. \\ &- \left. (1 + (\tilde{\sigma}_n^R)^2)^{-1} \sum_{s \neq t \in X_{ni}} \mathbf{x}_i(s)\mathbf{x}_i(t)^T \right\}.\end{aligned}$$

Under Waagepetersen's (2008) estimation technique, one could estimate (C.3) and (C.4) by

$$\hat{\mathbf{J}}_\infty(\tilde{\boldsymbol{\beta}}_n^W) = -n^{-1} \sum_{i=1}^n \left\{ \sum_{t \in X_{ni} \cup D_{ni}} \frac{\mathbf{x}_i(t)\mathbf{x}_i(t)^T \pi_{ni}(t) \exp\{(\tilde{\boldsymbol{\beta}}_n^W)^T \mathbf{x}_i(t)\}}{(\pi_{ni}(t) + \exp\{(\tilde{\boldsymbol{\beta}}_n^W)^T \mathbf{x}_i(t)\})^2} \right\}, \tag{C.7}$$

$$
\begin{aligned}
\tilde{\boldsymbol{\Sigma}}_{\infty}^{Q}(\tilde{\boldsymbol{\theta}}_n^W) &= n^{-1}\sum_{i=1}^{n}\Bigg\{(1+(\tilde{\sigma}_n^W)^2)\sum_{s\neq t\in X_{ni}}(\pi_{ni}(s,t)+\exp\{(\tilde{\boldsymbol{\beta}}_n^W)^T(\mathbf{x}_i(s)+\mathbf{x}_i(t))\})^{-1}\\
&* \frac{\pi_{ni}(s)\pi_{ni}(t)\mathbf{x}_i(s)\mathbf{x}_i(t)^T\exp\{(\tilde{\boldsymbol{\beta}}_n^W)^T(\mathbf{x}_i(s)+\mathbf{x}_i(t))\}}{(\pi_{ni}(s)+\exp\{(\tilde{\boldsymbol{\beta}}_n^W)^T\mathbf{x}_i(s)\})(\pi_{ni}(t)+\exp\{(\tilde{\boldsymbol{\beta}}_n^W)^T\mathbf{x}_i(t)\})}\\
&+ \sum_{s\neq t\in D_{ni}}(\pi_{ni}(s,t)+\exp\{(\tilde{\boldsymbol{\beta}}_n^W)^T(\mathbf{x}_i(s)+\mathbf{x}_i(t))\})^{-1}\\
&* \frac{\pi_{ni}(s)\pi_{ni}(t)\mathbf{x}_i(s)\mathbf{x}_i(t)^T\exp\{(\tilde{\boldsymbol{\beta}}_n^W)^T(\mathbf{x}_i(s)+\mathbf{x}_i(t))\}}{(\pi_{ni}(s)+\exp\{(\tilde{\boldsymbol{\beta}}_n^W)^T\mathbf{x}_i(s)\})(\pi_{ni}(t)+\exp\{(\tilde{\boldsymbol{\beta}}_n^W)^T\mathbf{x}_i(t)\})}\Bigg\}.
\end{aligned}
$$

Again, (C.5) is estimated similarly to (C.4) and so is omitted here. Simplifications of these equations are not possible as they were with Rathbun et al.'s (2007) estimation method because of the more complicated form for Waagepetersen's (2008) estimators.

3.4 Estimation of Baseline Intensity Rates

It is desirable to estimate the u_i terms in (C.2) to assess subjects' baseline event occurrence rates. If u_i is gamma distributed, it is conjugate to the Poisson distribution, and it is therefore possible to estimate it using the h-likelihood (Lee & Nelder 1996, 2001; Lee, Nelder & Pawitan 2006). Specifically, we can estimate $v_i = \ln(u_i)$, which is the canonical parameter for a Poisson-gamma HGLM (Lee, Nelder & Pawitan 2006). Then the h-likelihood has the form

$$
\begin{aligned}
h_n(\boldsymbol{\theta},\mathbf{v}_n) &= \ell_n(\boldsymbol{\beta}|\mathbf{v}_n,\sigma^2)+\ell_n(\mathbf{v}_n|\sigma^2)\\
&= \sum_{i=1}^{n}\Bigg[\Bigg(\sum_{t\in X_{ni}}\boldsymbol{\beta}^T\mathbf{x}_i(t)\Bigg)+v_iN_i-\Lambda_i(A_{ni};\boldsymbol{\beta})\exp\{v_i\}\\
&+ v_i-\sigma^{-2}\exp\{v_i\}-\ln\Gamma(\sigma^{-2})-\sigma^{-2}\ln(\sigma^2)\Bigg] \qquad \text{(C.1)}
\end{aligned}
$$

where $\mathbf{v}_n = (v_1,\ldots,v_n)^T$.

Given parameters $\boldsymbol{\beta}$ and σ^2, the score equation with respect to v_i is

$$
\frac{\partial h}{\partial v_i} = (N_i+1)-(\Lambda_i(A_{ni})+\sigma^{-2})\exp\{v_i\}
$$

giving a maximum h-likelihood estimate (MHLE) of

$$\hat{v}_i = \ln\left[\frac{\sigma^2(N_i+1)}{\sigma^2\Lambda_i(A_{ni})+1}\right];\ i = 1,\ldots,n,$$

which one can estimate by

$$\tilde{v}_i = \ln\left[\frac{\tilde{\sigma}_n^2(N_i+1)}{\tilde{\sigma}_n^2\hat{\Lambda}_i(A_{ni})+1}\right];\ i = 1,\ldots,n.$$

It is possible to assume other distributions can be assumed for the u_is by substituting the appropriate log-likelihood into the first line of (C.1); however, the resulting log-likelihood would be an extended log-likelihood, rather than an h-liklihood, and the resulting estimator may be biased (Lee, Nelder & Pawitan 2006). Often the solutions to such extended likelihoods take the form of estimating equations for non-conjugate distributions. For example, if the u_i terms are taken from a lognormal distribution, the baseline estimates solve the equations

$$\begin{aligned} 0 &= -u_i^{-2}\left\{1 - \frac{\ln(u_i) + \frac{1}{2}\ln(\tilde{\sigma}_n^2+1)}{\ln(\tilde{\sigma}_n^2+1)}\right\}\exp\left[-\frac{(\ln(u_i)+\frac{1}{2}\ln(\tilde{\sigma}_n^2+1))^2}{2\ln(\tilde{\sigma}_n^2+1)}\right] \\ &+ \frac{N_i}{u_i} - \hat{\Lambda}_i(A_{ni};\tilde{\boldsymbol{\beta}}_n);\ i = 1,\ldots,n. \end{aligned}$$

3.5 Results

The methods developed here are illustrated using Shiffman et al.'s (2002) data from an EMA study of smoking. 304 smokers were given electronic diaries, upon which they were instructed to record each cigarette smoked (hereafter referred to as an "event"). The diary prompted subjects to complete assessments at randomly selected event times and random times chosen according to a probability-based sampling design. The assessments sampled numerous covariates, including Negative Affect, Arousal, Attention, and Restlessness, which were the covariates analyzed by Rathbun et al. (2007) and will be the coviarates of interest in this analysis. Negative Affect was a composite of subjects' responses to negative adjectives such as "miserable" and "irritated," and Arousal was similarly calculated from responses to words such as "tired" and "energetic." Attention was based on subjects' responses to questions regarding their ease or difficulty concentrating. Restlessness was based on a single assessment item that did not load into the other three variables.

Both the non-event and event times were chosen by the diary to be prompted for assessment according to know sampling designs. The non-event assessments were chosen from a stratified (by day within subject) sampling design such that the expected number of non-event random assessments was five for each day.

Events were chosen to be assessed based on a Bernoulli sampling scheme, with a constant probability that a given event would be assessed throughout each day of

$$p_{ij} = \min\{5N_{i,j-1}^{-1}, 1\}, \tag{C.1}$$

where $N_{i,j-1}$ is the number of events for subject i on day $j-1$.

The series of points that results from randomly deleting each point of a Poisson process via a Bernoulli trial with known probability is a Poisson process (Cressie 1991). Therefore, conditional on $N_{i,j-1}$ and u_i, the assessed events come from a modulated Poisson process with intensity $p_{ij}u_i \exp\{\boldsymbol{\beta}^T \mathbf{x}_i(t)\}$.

Estimates of the parameters $\tilde{\boldsymbol{\theta}}_n$ were obtained by solving the equations (C.10) and (C.1), modified to account for the stratification by day. Specifically, the estimates solved

$$\tilde{\Psi}_n(\boldsymbol{\theta}) = n^{-1} \sum_{i=1}^{n} \left[\sum_{j=1}^{d_i} \left\{ \sum_{t \in X_{ij}} \mathbf{x}_i(t) - p_{ij} \hat{\Lambda}_{ij}^{(1)}(A_{ni}; \boldsymbol{\beta}) \right\} \right]$$

and

$$\tilde{\sigma}_n^2 = n^{-1} \sum_{i=1}^{n} \left\{ \left(\sum_{j=1}^{d_i} |L_{ij}|^2 \right)^{-1} \sum_{j=1}^{d_i} \sum_{s \neq t \in X_{ij}} p_{ij}^{-2} \phi_i(\tilde{\boldsymbol{\beta}}_n, s) \phi_i(\tilde{\boldsymbol{\beta}}_n, t) - 1 \right\}$$

where d_i is the number of days that subject i was involved in the study, X_{ij} is the set of event times for the ith subject on the jth day, $|L_{ij}|$ the length of time the diary was active for the ith subject on the jth day, p_{ij} as defined in (C.1), and, using Rathbun et al.'s (2007) estimation methods as an example,

$$\hat{\Lambda}_{ij}^{(1)}(\tilde{\boldsymbol{\beta}}_n, t) = \frac{|L_{ij}|}{m_{ij}} \sum_{t \in D_{ij}} \exp\{\tilde{\boldsymbol{\beta}}_n^T \mathbf{x}_i(t)\}, \tag{C.2}$$

with D_{ij} representing the set of non-event assessments for the ith subject on the jth day and m_{ij} the number of elements in D_{ij}. Waagepetersen's method of estimation modified (C.2) accordingly, similar to (C.9).

Table 3.1: Comparison between Rathbun et al. (2007) method and current method

		Rathbun		Current	
Covariate	Estimate	Std. Error	p-value	Std. error	p-value
Intercept	0.0258	0.0147	0.0624	0.0664	0.3488
Negative Affect	0.0062	0.0165	0.3545	0.0418	0.4414
Arousal	0.0006	0.0161	0.4853	0.0379	0.4937
Attention	0.0248	0.0184	0.0886	0.0509	0.3131
Restlessness	0.0808	0.0082	< 0.0001	0.0400	0.0216
σ^2	0.0914				

Table 3.2: Comparison between Waagepetersen (2008) method and current method

		Waagepetersen		Current	
Covariate	Estimate	Std. Error	p-value	Std. error	p-value
Intercept	-0.0196	0.0145	0.0880	0.0367	0.2963
Negative Affect	-0.0025	0.0164	0.4390	0.0306	0.4673
Arousal	0.0111	0.0158	0.2409	0.0276	0.3430
Attention	0.0296	0.0177	0.0470	0.0347	0.1966
Restlessness	0.0628	0.0099	< 0.0001	0.0240	0.0045
σ^2	0.1728				

Table 3.1 compares the results from Rathbun et al.'s (2007) estimation of the coefficient parameters with the current method, using Rathbun et al.'s value for $f(t;\boldsymbol{\beta})$. Note that both methods obtain the same point estimates for these parameters, as they use the same estimating equations. Table 3.2 provides a similar comparison of the results from Waagepetersen's (2008) method against the current method, using Waagepetersen's value for $f(t;\boldsymbol{\beta})$. The estimates for σ^2 only apply to the current method, as previous methods did not estimate this parameter.

From these results, we can see that the current model does cause a significant trade-off in terms of increased variance. For both methods, the standard errors of the estimates were increased by several times, with appropriate reductions in significance. It is worth noting, however, that Restlessness was still determined to have a significant positive relationship

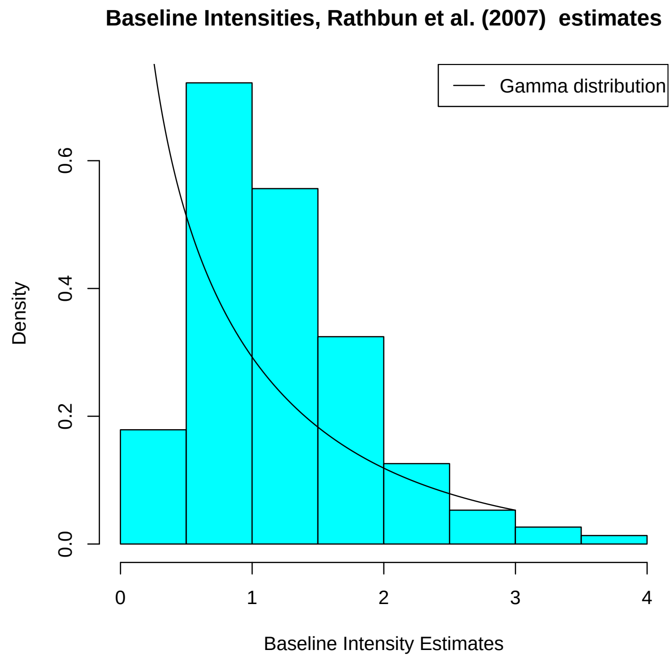

Figure 3.1: Baseline Estimates using Rathbun et al. (2007) method and gamma distribution

with smoking rate using the current method under both choices of f; previous literature and analyses suggest that this variable should be the most important of those studied (Shiffman et al. 1996, Shiffman et al. 2002).

Figures 3.1 and 3.2 show histograms of the estimated baseline intensities u_i for both values of f and the corresponding estimate of σ^2, assuming a gamma distribution for the u_i terms; these were obtained by estimating v_i as detailed in Section 3.4, and taking $\hat{u}_i = \exp\{\hat{v}_i\}$. The true Gamma$(\tilde{\sigma}_n^{-2}, \tilde{\sigma}_n^2)$ distribution is superimposed on each histogram. Clearly the u_i estimates based on Waagepetersen's method of estimating $\tilde{\sigma}_n^2$ and $\hat{\Lambda}_i(A_{ni}; \tilde{\boldsymbol{\beta}}_n)$ do not follow a gamma distribution; they are more heavily clustered around a cetral point and right-skewed than this distribution. While Rathbun et al.'s estimation methods provide a distribution of u_is that is more closely gamma, it still does not appear to fit this distribution well. This lack of fit for both estimates may be due to a misspecification of the u_i distribution, or also due

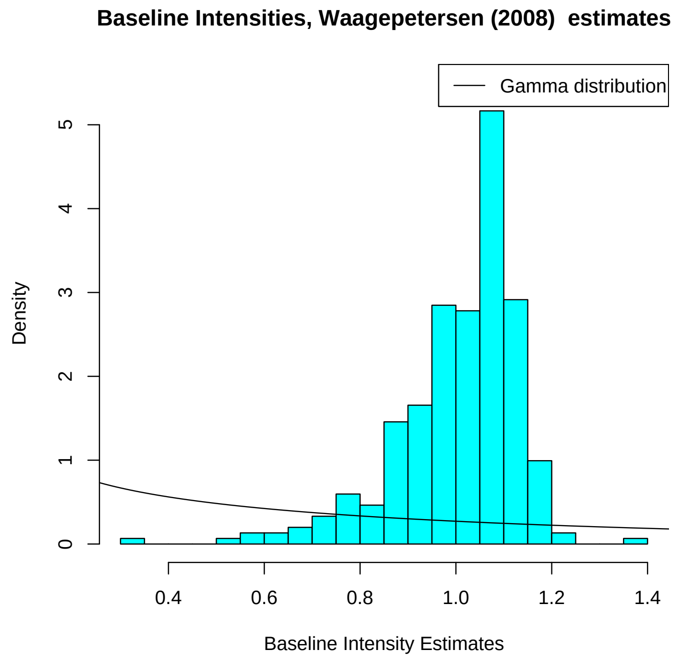

Figure 3.2: Baseline Estimates using Waagepetersen (2008) method and gamma distribution

to the exclusion of important covariates from the model, which would provide an incorrect estimate of $\tilde{\sigma}_n^2$ and thus the baseline terms. One area of future research on these models should concern developing diagnostics for the fit of the baseline intensity terms to a given distribution, and possibly nonparametric estimation of these terms.

Figures 3.1 and 3.2 show histograms of the estimated baseline intensities u_i for both values of f and the corresponding estimate of σ^2, assuming a gamma distribution for the u_i terms; these were obtained by estimating v_i as detailed in Section 3.4, and taking $\hat{u}_i = \exp\{\hat{v}_i\}$. The true Gamma$(\tilde{\sigma}_n^{-2}, \tilde{\sigma}_n^2)$ distribution is superimposed on each histogram. Clearly the u_i estimates based on Waagepetersen's method of estimating $\tilde{\sigma}_n^2$ and $\hat{\Lambda}_i(A_{ni}; \tilde{\boldsymbol{\beta}}_n)$ do not follow a gamma distribution; they are more heavily clustered around a central point and right-skewed than this distribution. While Rathbun et al.'s estimation methods provide a distribution of u_is that is more closely gamma, it still does not appear to fit this distribution well. This lack of fit for both estimates may be due to a misspecification of the u_i distribution, or also due to the exclusion of important covariates from the model, which would provide an incorrect estimate of $\tilde{\sigma}_n^2$ and thus the baseline terms. One area of future research on these models should concern developing diagnostics for the fit of the baseline intensity terms to a given distribution, and possibly nonparametric estimation of these terms.

Appendix A

Since $\tilde{\sigma}_n^2$ may be treated as nuisance parameter, we will first prove the consistency of $\tilde{\boldsymbol{\beta}}_n$. The following Lemmas are necessary to prove this:

Lemma 4. *Under Assumptions 8 through 12, $n^{\frac{1}{2}}\tilde{\Psi}_n(\boldsymbol{\beta_0})$ is asymptotically normal with mean 0 and constant variance (detailed in Appendix B).*

Proof of Lemma 4. Under Assumption 8, the indepdence between event and random assessments implies that showing the asymptotic normality of $n^{\frac{1}{2}}\tilde{\Psi}_n(\boldsymbol{\beta_0})$ is equivalent to proving the asymptotic normality of

$$Q_n(\boldsymbol{\beta}) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \left\{ \sum_{t \in X_{ni}} \mathbf{x}_i(t) f(t; \boldsymbol{\beta}) \right\}$$

and

$$R_n(\boldsymbol{\beta}) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \left\{ \sum_{t \in D_{ni}} \frac{\mathbf{x}_i(t) f(t; \boldsymbol{\beta}) \exp\{\boldsymbol{\beta}^T \mathbf{x}_i(t)\}}{\pi_{ni}(t)} \right\}$$

under both Rathbun et al. and Waagepetersen's choices for $f(t; \boldsymbol{\beta})$. Briefly, under Assumptions 8, 9, 10, and 12, each of these equations satisfies the conditions of Lyapunov's CLT. A similar proof can be found in Guan and Loh's (2007) proof of their Theorem 1, but in this setting the spatial partitioning is unnecessary, since by Assumption 8 these equations are sums of independent variables. Then $n^{\frac{1}{2}}\tilde{\Psi}_n(\boldsymbol{\beta})$ is asymptotically the sum of independent normal variables and is therefore asymptotically normal. □

Lemma 5. *Under Assumptions 8 through 12,*

$$\frac{\partial}{\partial \boldsymbol{\beta}^T} \tilde{\Psi}_n(\boldsymbol{\beta}) \rightarrow -\mathbf{M}$$

as $n \to \infty$ for some positive definite matrix $\mathbf{M}$.

Proof of Lemma 5. First note that

$$\frac{\partial}{\partial \boldsymbol{\beta}^T} \tilde{\Psi}_n(\boldsymbol{\beta}) = -n^{-1} \sum_{i=1}^{n} \frac{\partial}{\partial \boldsymbol{\beta}^T} \hat{\Lambda}_i^{(1)}(A_{ni}; \boldsymbol{\beta}).$$

It can be shown that for both Rathbun et al.'s (2007) and Waagepetersen's (2008) estimators, this quantity equals

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\beta}^T} \tilde{\Psi}_n(\boldsymbol{\beta}) &= -n^{-1} \sum_{i=1}^{n} \Bigg\{ \sum_{t \in X_{ni}} \mathbf{x}_i(t) \mathbf{x}_i(t)^T f(t; \boldsymbol{\beta})(1 - f(t; \boldsymbol{\beta})) \\ &+ \sum_{t \in D_{ni}} \frac{\mathbf{x}_i(t) \mathbf{x}_i(t)^T f(t; \boldsymbol{\beta})^2 \exp\{\boldsymbol{\beta}^T \mathbf{x}_i(t)\}}{\pi_{ni}(t)} \Bigg\} \end{aligned} \tag{A.1}$$

Note that for Rathbun et al.'s (2007),definition of $f(t; \boldsymbol{\beta})$, (A.1) is an unbiased estimator of

$$-\Lambda_i^{(2)}(A_{ni}; \boldsymbol{\beta}) = -\int_{A_{ni}} \mathbf{x}_i(t) \mathbf{x}_i(t)^T \exp\{\boldsymbol{\beta}^T \mathbf{x}_i(t)\}\, dt,$$

and for Waagepetersen's (2008) definition, it is an unbiased estimator of

$$-\int_{A_{ni}} \mathbf{x}_i(t) \mathbf{x}_i(t)^T \exp\{\boldsymbol{\beta}^T \mathbf{x}_i(t)\} \frac{\pi_{ni}(t)}{\pi_{ni}(t) + \exp\{\boldsymbol{\beta}^T \mathbf{x}_i(t)\}}\, dt.$$

In both cases, by the Assumptions 8 through 12, (A.1) converges to some matrix $-\mathbf{M}$ as $n \to \infty$, where $\mathbf{M}$ has a strictly positive minimum eigenvalue and is thus positive definite. $\square$

Proof of Theorem 3. To prove this, we follow a similar approach to Lin and Clayton (2005, Lemma 2.3). We show that the estimating equation $n^{\frac{1}{2}} \tilde{\Psi}_n(\boldsymbol{\beta_0})$ converges to a zero-mean normally distributed random vector (Lemma 1), and that $(\partial/\partial \boldsymbol{\beta}^T)\tilde{\Psi}_n(\boldsymbol{\beta})$ converges to a positive definite matrix as $n \to \infty$ (Lemma 2). Then by using the inverse function theorem (see Apostol 1974), and the result that a sequence converges in probability iff every subsequence contains a subsubsequence that converges almost surely (see Chung 2001, p. 79), it follows that, for every subsequence $\tilde{\Psi}_{n,j}(\boldsymbol{\beta})$, there exists an r and a subsubsequence $\tilde{\Psi}_{n,j,k}(\boldsymbol{\beta})$ such that the probability that $\tilde{\Psi}_{n,j,k}(\boldsymbol{\beta})$ is one-to-one with the open ball $B(\boldsymbol{\beta_0}, r)$ approaches 1 as n approaches infinity. The asymptotic normality of $\tilde{\Psi}_{n,j,k}(\boldsymbol{\beta})$ ensures that

$n^{\frac{1}{2}}\tilde{\Psi}_{n,j,k}(\boldsymbol{\beta_0}) \rightarrow E_{\boldsymbol{\theta_0}}\{\tilde{\Psi}_{n,j,k}(\boldsymbol{\beta_0})\} = \mathbf{0}$. Since there exists an r^* for which $B(n^{\frac{1}{2}}\tilde{\Psi}_{n,j,k}(\boldsymbol{\beta_0}), r^*) \subseteq n^{\frac{1}{2}}B(\tilde{\Psi}_{n,j,k}(\boldsymbol{\beta_0}), r)$, and $P(\|n^{\frac{1}{2}}\tilde{\Psi}_{n,j,k}(\boldsymbol{\beta_0}) - \mathbf{0}\| < r^*) \rightarrow 1$ as $n \rightarrow \infty$, the probability that $\mathbf{0} \in n^{\frac{1}{2}}\tilde{\Psi}_{n,j,k}\{B(\boldsymbol{\beta_0}, r)\}$ goes to 1 as $n \rightarrow \infty$. Since $\tilde{\Psi}_{n,j,k}(\boldsymbol{\beta})$ is one-to-one on $B(\boldsymbol{\beta_0}, r)$, $B(n^{\frac{1}{2}}\tilde{\Psi}_{n,j,k}(\boldsymbol{\beta_0}), r^*) \subseteq n^{\frac{1}{2}}B(\tilde{\Psi}_{n,j,k}(\boldsymbol{\beta_0}), r)$, and r can be made infinitely small, $\tilde{\Psi}_{n,j,k}(\tilde{\boldsymbol{\beta}}_n) = \mathbf{0}$ implies the subsubsequence converges almost surely, giving the theorem.

The consistency of $\tilde{\sigma}_n^2$ may be proven in a more straightforward manner, using Chebychev's Theorem. All that must be shown is that $Var_{\boldsymbol{\theta_0}}(\tilde{\sigma}_n^2) \rightarrow 0$ as $n \rightarrow \infty$. By Assumption 8, we have

$$\begin{aligned}
Var_{\boldsymbol{\theta_0}}(\tilde{\sigma}_n^2) &= n^{-2}\sum_{i=1}^{n}|A_{ni}|^{-4}Var_{\boldsymbol{\theta_0}}\Big\{\sum_{s\in X_{ni}}\sum_{t\neq s\in X_{ni}}\phi_i(\tilde{\boldsymbol{\beta}}_n, s)\phi_i(\tilde{\boldsymbol{\beta}}_n, t) - 1\Big\} \\
&= n^{-2}\sum_{i=1}^{n}|A_{ni}|^{-4}\Big[E_{\boldsymbol{\theta_0}}\Big\{\Big(\sum_{s\in X_{ni}}\sum_{t\neq s\in X_{ni}}\phi_i(\tilde{\boldsymbol{\beta}}_n, s)\phi_i(\tilde{\boldsymbol{\beta}}_n, t)\Big)^2\Big\} \\
&- E_{\boldsymbol{\theta_0}}\Big\{\sum_{s\in X_{ni}}\sum_{t\neq s\in X_{ni}}\phi_i(\tilde{\boldsymbol{\beta}}_n, s)\phi_i(\tilde{\boldsymbol{\beta}}_n, t)\Big\}^2\Big] \\
&= n^{-2}\sum_{i=1}^{n}|A_{ni}|^{-4}\{V_{i1} - V_{i2}\},
\end{aligned}$$

where $\phi_i(\tilde{\boldsymbol{\beta}}_n, t)$ is as defined in (C.6). We obtain

$$\begin{aligned}
V_{i1} &= E_{\boldsymbol{\theta_0}}\Big[\Big\{\sum_{s\in X_{ni}}\sum_{t\neq s\in X_{ni}}\phi_i(\tilde{\boldsymbol{\beta}}_n, s)\phi_i(\tilde{\boldsymbol{\beta}}_n, t)\Big\}^2\Big] \\
&= E_{\boldsymbol{\theta_0}}\Big[\sum_{s\in X_{ni}}\sum_{t\neq s\in X_{ni}}\phi_i(\tilde{\boldsymbol{\beta}}_n, s)^2\phi_i(\tilde{\boldsymbol{\beta}}_n, t)^2 \\
&+ 4\sum_{s\in X_{ni}}\sum_{t\neq s\in X_{ni}}\sum_{u\neq s,t\in X_{ni}}\phi_i(\tilde{\boldsymbol{\beta}}_n, s)^2\phi_i(\tilde{\boldsymbol{\beta}}_n, t)\phi_i(\tilde{\boldsymbol{\beta}}_n, u) \\
&+ \sum_{s\in X_{ni}}\sum_{t\neq s\in X_{ni}}\sum_{u\neq s,t\in X_{ni}}\sum_{v\neq s,t,u\in X_{ni}}\phi_i(\tilde{\boldsymbol{\beta}}_n, s)\phi_i(\tilde{\boldsymbol{\beta}}_n, t)\phi_i(\tilde{\boldsymbol{\beta}}_n, u)\phi_i(\tilde{\boldsymbol{\beta}}_n, v)\Big]
\end{aligned}$$

$$
\begin{aligned}
&= E_{\boldsymbol{\theta_0}}\{N_i(N_i-1))\}\Lambda_i(A_{ni};\boldsymbol{\beta})^{-2} \\
&* \left(\int_{A_{ni}}\int_{A_{ni}} \phi_i(\tilde{\boldsymbol{\beta}}_n, s)^2\phi_i(\tilde{\boldsymbol{\beta}}_n, t)^2 \exp\{\boldsymbol{\beta_0}^T(\mathbf{x}_i(s)+\mathbf{x}_i(t))\}\, ds\, dt\right) \\
&+ 4E_{\boldsymbol{\theta_0}}\{N_i(N_i-1)(N_i-2)\}\Lambda_i(A_{ni};\boldsymbol{\beta})^{-3}|A_{ni}|^2\left(\int_{A_{ni}}\int_{A_{ni}}\int_{A_{ni}} \phi_i(\tilde{\boldsymbol{\beta}}_n, s)^2\phi_i(\tilde{\boldsymbol{\beta}}_n, t)\right. \\
&* \left.\phi_i(\tilde{\boldsymbol{\beta}}_n, u)\exp\{\boldsymbol{\beta_0}^T(\mathbf{x}_i(s)+\mathbf{x}_i(t)+\mathbf{x}_i(u))\}\, ds\, dt\, du\right) \\
&+ E_{\boldsymbol{\theta_0}}\{N_i(N_i-1)(N_i-2)(N_i-3)\}\Lambda_i(A_{ni};\boldsymbol{\beta})^{-4}|A_{ni}|^4 \\
&* \left(\int_{A_{ni}}\int_{A_{ni}}\int_{A_{ni}}\int_{A_{ni}} \phi_i(\tilde{\boldsymbol{\beta}}_n, s)\phi_i(\tilde{\boldsymbol{\beta}}_n, t)\phi_i(\tilde{\boldsymbol{\beta}}_n, u)\phi_i(\tilde{\boldsymbol{\beta}}_n, v)\right. \\
&* \left.\exp\{\boldsymbol{\beta_0}^T(\mathbf{x}_i(s)+\mathbf{x}_i(t)+\mathbf{x}_i(u)+\mathbf{x}_i(v))\}\, ds\, dt\, du\, dv\right) \\
&= (1+\sigma_0^2)\left(\int_{A_{ni}}\int_{A_{ni}} \phi_i(\tilde{\boldsymbol{\beta}}_n, 2s+2t)\phi_i(\boldsymbol{\beta_0}, s+t)^{-1}\, ds\, dt\right) \\
&+ 4M_3\left(\int_{A_{ni}}\int_{A_{ni}}\int_{A_{ni}} \phi_i(\tilde{\boldsymbol{\beta}}_n, 2s+t+u)\phi_i(\boldsymbol{\beta_0}, s+t+u)^{-1}\, ds\, dt\, du\right) \\
&+ M_4\left(\int_{A_{ni}}\int_{A_{ni}}\int_{A_{ni}}\int_{A_{ni}} \phi_i(\tilde{\boldsymbol{\beta}}_n, s+t+u+v)\phi_i(\boldsymbol{\beta_0}, s+t+u+v)^{-1}\, ds\, dt\, du\, dv\right),
\end{aligned}
$$

where M_3 and M_4 are the third and forth moments of u_i, respectively, and

$$
\begin{aligned}
V_{i2} &= \left[(1+\sigma_0^2)\left(\int_{A_{ni}}\int_{A_{ni}} \phi_i(\tilde{\boldsymbol{\beta}}_n, s+t)\phi_i(\boldsymbol{\beta_0}, s+t)^{-1}\, ds\, dt\right)\right]^2 \\
&= (1+\sigma_0^2)^2\left(\int_{A_{ni}}\int_{A_{ni}}\int_{A_{ni}}\int_{A_{ni}} \phi_i(\tilde{\boldsymbol{\beta}}_n, s+t+u+v)\right. \\
&* \left.\phi_i(\boldsymbol{\beta_0}, s+t+u+v)^{-1}\, ds\, dt\, du\, dv\right).
\end{aligned}
$$

Then

$$
\begin{aligned}
Var_{\boldsymbol{\theta_0}}(\tilde{\sigma}_n^2) &= n^{-2}\sum_{i=1}^{n}|A_{ni}|^{-4} \\
&* \left[(1+\sigma_0^2)\left(\int_{A_{ni}}\int_{A_{ni}} \phi_i(\tilde{\boldsymbol{\beta}}_n, 2s+2t)\phi_i(\boldsymbol{\beta_0}, s+t)^{-1}\, ds\, dt\right)\right. \\
&+ 4M_3\left(\int_{A_{ni}}\int_{A_{ni}}\int_{A_{ni}} \phi_i(\tilde{\boldsymbol{\beta}}_n, 2s+t+u)\phi_i(\boldsymbol{\beta_0}, s+t+u)^{-1}\, ds\, dt\, du\right) \\
&+ (M_4-(1+\sigma_0^2)^2)\left(\int_{A_{ni}}\int_{A_{ni}}\int_{A_{ni}}\int_{A_{ni}} \phi_i(\tilde{\boldsymbol{\beta}}_n, s+t+u+v)\right. \\
&* \left.\left.\phi_i(\boldsymbol{\beta_0}, s+t+u+v)^{-1}\, ds\, dt\, du\, dv\right)\right]. \qquad \text{(A.2)}
\end{aligned}
$$

Under Assumption 9, (A.2) can be shown to converge to 0 as $n \to \infty$. Then, since $\tilde{\sigma}_n^2$, evaluted at $\boldsymbol{\beta_0}$, is unbiased by design and $\tilde{\boldsymbol{\beta}}_{\boldsymbol{n}}$ is consistent as $n \to \infty$, $\tilde{\sigma}_n^2$, evaluated at $\tilde{\boldsymbol{\beta}}_n$ approaches σ_0^2 in probability as $n \to \infty$, completing the proof. □

Appendix B

Proof of Theorem 4. We take a Taylor Expansion of $\tilde{\Psi}_n(\tilde{\boldsymbol{\beta}}_n)$ about $\boldsymbol{\beta_0}$ to obtain

$$\tilde{\Psi}_n(\tilde{\boldsymbol{\beta}}_n) = \tilde{\Psi}_n(\boldsymbol{\beta_0}) + \hat{\mathbf{J}}_n(\boldsymbol{\beta_0})(\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta_0}) + \frac{1}{2}\hat{\mathbf{K}}_n(\boldsymbol{\beta^*})(\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta_0}), \tag{B.1}$$

where

$$\hat{\mathbf{J}}_n(\boldsymbol{\beta}) = \frac{\partial}{\partial \boldsymbol{\beta}^T}\tilde{\Psi}_n(\boldsymbol{\beta}),$$

$$\hat{\mathbf{K}}_n(\boldsymbol{\beta}) = \sum_{l=1}^{p}(\beta_l - \beta_{0,l})\left(\frac{\partial^2}{\partial \beta_l \partial \boldsymbol{\beta}^T}\tilde{\Psi}_n(\boldsymbol{\beta})\right),$$

and $\boldsymbol{\beta^*} = c\boldsymbol{\beta} + (1-c)\boldsymbol{\beta_0}$ for some real $c \in (0,1)$. Then, since $\tilde{\boldsymbol{\beta}}_n$ is a zero for $\tilde{\Psi}_n$, we can rearrange (B.1) to obtain

$$n^{\frac{1}{2}}(\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta_0}) = -\left\{\hat{\mathbf{J}}_n(\boldsymbol{\beta_0}) + \hat{\mathbf{K}}_n(\boldsymbol{\beta^*})\right\}^{-1} n^{\frac{1}{2}}\tilde{\Psi}_n(\boldsymbol{\beta_0}).$$

Under Assumptions 8 through 10 and 12, $(\partial^2/\partial\beta_l\partial\boldsymbol{\beta}^T)\tilde{\Psi}_n(\boldsymbol{\beta_0})$ is bounded above and below for any $l = 1, \ldots, p$ and all n. Then since, by Theorem 3, $\tilde{\boldsymbol{\beta}}_n \xrightarrow{p} \boldsymbol{\beta_0}$ as $n \to \infty$ and the derivative term of $\hat{\mathbf{K}}_n(\boldsymbol{\beta^*})$ is bounded,

$$\hat{\mathbf{K}}_n(\boldsymbol{\beta^*}) \xrightarrow{p} \mathbf{0}$$

as $n \to \infty$.

Define

$$\begin{aligned}
\mathbf{J}_n(\boldsymbol{\beta}) &= n^{-1}\sum_{i=1}^{n} E_{\boldsymbol{\theta_0}}\{\hat{\mathbf{J}}_n(\boldsymbol{\beta})\} \\
&= n^{-1}\sum_{i=1}^{n} E_{\boldsymbol{\theta_0}}\Bigg\{\sum_{t\in X_{ni}} \mathbf{x}_i(t)\mathbf{f}^{(1)}(t;\boldsymbol{\beta})^T \\
&- \sum_{t\in D_{ni}}\left(\frac{\mathbf{x}_i(t)\mathbf{x}_i(t)^T f(u;\boldsymbol{\beta})\exp\{\boldsymbol{\beta}^T\mathbf{x}_i(t)\} + \mathbf{x}_i(t)\mathbf{f}^{(1)}(u;\boldsymbol{\beta})\exp\{\boldsymbol{\beta}^T\mathbf{x}_i(t)\}}{\pi_{ni}(t)}\right)\Bigg\},
\end{aligned}$$

where $\mathbf{f}^{(1)}(t;\boldsymbol{\beta}) = (\partial/\partial\boldsymbol{\beta}^T)f(t;\boldsymbol{\beta})$. It can be shown then that (C.2) equals (C.3) and

$$\hat{\mathbf{J}}_n(\boldsymbol{\beta}_\mathbf{0}) - \mathbf{J}_\infty(\boldsymbol{\beta}_\mathbf{0}) \xrightarrow{p} 0$$

as $n \to \infty$, under Assumptions 9 and 12. It follows that

$$\hat{\mathbf{J}}_n(\boldsymbol{\beta}_\mathbf{0}) + \hat{\mathbf{K}}_n(\boldsymbol{\beta}^*) \xrightarrow{p} \mathbf{J}_\infty(\boldsymbol{\beta}_\mathbf{0}) \tag{B.2}$$

as $n \to \infty$.

It was already proven in Appendix A that $n^{\frac{1}{2}}\tilde{\Psi}_n(\boldsymbol{\beta}_\mathbf{0})$ is asymptotically normal. Its distribution must have mean $\mathbf{0}$ (since the equation is unbiased by design), and we define the asymptotic variance as in (C.4) and (C.5). Then $n^{\frac{1}{2}}\tilde{\Psi}_n(\boldsymbol{\beta}_\mathbf{0}) \overset{a}{\sim} \mathbf{N}(\mathbf{0}, \boldsymbol{\Sigma}^Q_\infty(\boldsymbol{\beta}_\mathbf{0}) + \boldsymbol{\Sigma}^R_\infty(\boldsymbol{\beta}_\mathbf{0}))$, which, with (B.2), gives the theorem. $\square$

3.6 References

Affleck, G., Zautra, A., Tennen, H. & Armelia, S. (1999). Multilevel daily proces designs for consulting and clinical psychology: A preface for the perplexed. *Journal of Consulting and Clinical Psychology, 67*, 746–754.

Apostol, T. M. (1974). *Mathematical Analysis* (2nd ed.). Reading, Massachusetts: Addison-Wesley.

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology, 54*, 579–616.

Chung, K. L. (2001). *A Course in Probability Theory* (Revised 2nd ed.). New York: Academic Press.

Cordy, C. B. (1993). An extension of the Horvitz-Thompson theorem to point sampling from a continuous universe. *Statistics & Probability Letters, 18*, 353–362.

Cox, D. R. (1955). Some statistical methods connected with series of events. *Journal of the Royal Statistical Society B, 17*, 129–164.

Cox, D. R. (1972). The statistical analysis of dependencies in point processes. In *Stochastic Point Processes*, P. A. W. Lewis (Ed.), 55–66. New York: Wiley.

Cressie, N. (1991). *Statistics for spatial data*. New York: Wiley.

Diggle, P. J. (2003). *Statistical analysis of spatial point patterns* (2nd ed.). New York: Oxford University Press.

Guan, Y. & Loh, J. M. (2007). A thinned block bootstrap variance estimation procedure for inhomogenous spatial point patterns. *Journal of the American Statistical Association, 102*(480), 1377–1386.t

Lee, Y. & Nelder, J. A. (1996). Hierarchical generalized linear models. *Journal of the Royal Statistical Society, Series B (Methodological), 58*, 619–678.

Lee, Y. & Nelder, J. A. (2001). Hierarchical generalized linear models: A synthesis of generalised linear models, random-effect models and structured dispersions. *Biometrika, 88*(4), 987–1006.

Lee, Y., Nelder, J. A. & Pawitan, Y. (2006). *Generalized Linear Models with random effects: Unified analysis via H-likelihood.* New York: Chapman & Hall.

Lin, P. & Clayton, M. K. (2005). Analysis of binary spatial data by quasi-likelihood estimating functions. *Annals of Statistics, 33*(2), 542–555.

Nelder, J.A. & Wedderburn, R. W. M. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society, Series A (General), 135*(3), 370–384.

Rathbun, S. L., Shiffman, S., & Gwaltney, C. J. (2006). Point process models for event history data: Applications in behavioral science. In *Models for intensive longitudinal data*, T. A. Walls & J. L. Schafer (Eds.), 219–253. New York: Oxford University Press.

Rathbun, S. L., Shiffman, S., & Gwaltney, C. J. (2007). Modelling the effects of partially observed covariates on Poisson process intensity. *Biometrika, 94*, 153–165.

Schwartz, J. E. & Stone, A. A. (2007). The analysis of real-time momentary data: A practical guide. In *The Science of Real-time Data Capture: Self-report in Health Research*, Stone, A. A., Shiffman, S. S., Atienza, A., & Nebeling, L. (Eds.), 76–113. New York: Oxford University Press USA.

Shiffman, S. et al. (2002). Immediate antecedents of cigarette smoking: An analysis from ecological momentary assessment. *Journal of Abnormal Psychology, 111*, 531–545.

Shiffman, S. & Stone, A. A. (1998). Ecological Momentary Assessment: A new tool for behavioral medicine research. In *Technology and Methods in Behavioral Medicine*, Krantz, D. S. & Baum, A. (Eds.), 117–132. Philadelphia: Lawrence Erlbaum Associates.

Smyth, J. et al. (2001). The use of ecological momentary assessment approaches in eating disorder research. *International Journal of Eating Disorders, 30*, 83–95.

Stone, A. A., Broderick, J. B., Porter, L., & Kaell, A. T. (1997). The experience of rheumatoid arthritis pain and fatigue: Examining momentary reports and correlates over one week. *Arthritis Care and Research, 10*, 185-193.

Stone, A. A., Broderick, J. E., Shiffman, S. S. & Schwartz, J.E. (2004). Understanding recall of weekly pain from a momentary assessment perspective: Absolute agreement, between- and within-person consistency, and judged change in weekly pain. *Pain, 107*(1–2), 61–69.

Stone, A. A. et al. (1998). A comparison of coping assessed by ecological momentary assessment and retrospective recall. *Journal of Personality and Social Psychology, 74*, 1670–1680.

Stone, A. A. & Shiffman, S. (2002). Capturing momentary, self-report data: A proposal for reporting guidelines. *Annals of Behavioral Medicine, 24*, 236–243.

Waagepetersen, R. (2008). Estimating functions for inhomogenous spatial point processes with incomplete covariate data. *Biometrika, 95*(2) 351–363.

Chapter 4

Modulated Poisson Processes with Random Coefficients: Adding Flexibility to EMA Analysis[1]

4.1 Introduction

The availability of portable electronic devices within the past decade has promoted the development of Ecological Momentary Assessment (EMA) (Shiffman & Stone 1998, Stone & Shiffman 2002), an emerging method in psychological research that uses electronic devices to record information about subjects' current psychological states in their natural environments. These electronic devices may also be used to record the times of discrete behavioral events (e.g. cigarettes smoked, incidences of purging in bulimic patients) as they occur. Among the benefits of EMA are a reduction in recall bias and an increase in the "ecological validity" of the data by not placing subjects in an artificial clinical environment (Shiffman & Stone 1998, p. 117). EMA has been used to study numerous behavioral phenomena, including eating disorders (Smyth et al. 2001), stress coping (Stone et al. 1998), and pain (Stone, Broderick, Porter & Kaell 1997; Stone, Broderick, Shiffman & Schwartz 2004). By sampling subjects' psychological states using a probability-based scheme, unbiased estimates of the mean psychological states may be computed. If subjects are also assessed at event times, this method further allows one to estimate the effects of the covariates' mean states upon the timing of these events (Rathbun, Shiffman, & Gwaltney 2007).

Hierarchical Linear Models (HLMs) have been recommended for analyzing EMA data (Affleck et al. 1999; Bolger, Davis, & Rafaeli 2003; Schwartz & Stone 2007) because of the nested structure of the data, including momentary time-varying covariates that are measured

[1]Neustifter, B., S. L. Rathbun & S. Shiffman. Unsubmitted as of 6/15/09

within subjects, as well as subject-level variables (e.g. gender, education level, etc.) that vary among subjects. Random variation among subjects in their responses to momentary covariates may be described through the inclusion of random effects in the within-subjects level of the model. HLMs, however, assume that the response is normally distributed. To relax this assumption, Hierarchical Generalized Linear Models (HGLMs) (Lee & Nelder 1996) may be considered, which model the responses by Generalized Linear Model (GLM) (Nelder & Wedderburn 1972) with some random components. These random components may be assumed to have a variety of distributions (Lee & Nelder 1996). Like HLMs, HGLMs allow researchers to analyze and model two or more levels of variation. Generalized Linear Mixed Models (GLMMs) (Breslow & Clayton 1993) are a particular type of HGLM (Lee & Nelder 1996), where the random components are assumed to follow a normal distribution.

The motivating dataset comes from Shiffman et al.'s (2002) EMA study of smoking behavior, which consisted of 304 smokers who were given electronic diaries. Subjects recorded each smoked cigarette (hereafter referred to as an "event") at the time it was smoked, and filled out assessments regarding their mood (e.g. ratings of boredom or restlessness) and environment (e.g. whether other smokers were present) when prompted by the diary at randomly selected event and non-event times. Subject-level covariates such as age and gender were included in addition to the time-varying covariates. The response of interest was the rate and timing of smoking events, and their relationship to the measured covariates. Analyzing these data will provide insight into how mood and environment variables prompt smoking. This understanding will lead to improvements in intervention programs designed for smoking cessation.

The times of repeated behavioral events can be viewed as being a realization of a temporal point process (Rathbun, Shiffman, & Gwaltney 2006). Rathbun, Shiffman, and Gwaltney (2007) used a Poisson process (Cox 1972) to model the smoking data and found that smoking rate increased with the increasing restlessness of the smoker. However, their model implicitly assumed that all subjects have the same baseline rate of event occurrence and that all subjects' event occurrence rates are impacted equally by the covariates. Thus, taking inspiration from the form and techniques of GLMMs, we will modify such models to allow the covariate coefficients in the model to vary randomly among subjects. In this paper, we will describe a mixed-effects version of a modulated Poisson process, where the random effect terms vary according to a Normal distribution to create a model with more flexibility and the ability to model inter-subject variation.

We construct a simple set of estimating equations for the fixed effects and covariance components of the random effects. Our method bypasses the computationally intensive MCEM, Penalized Quasi-Likelihood or other likelihood-based estimation techniques of Poisson GLMMs (see McCulloch 1997, Breslow & Clayton 1993). The initial version of these estimating equations involve integrating a function of the time-varying covariates over the study domain comprised to the time intervals over wihch the subjects are observed, and so requires that the covariates be known functions of time. Using the approaches of Rathbun et al. (2007) and Waagepetersen (2008), we replace those integrals by design-unbiased estimators computed using covariate data obtained from the events of the process, and from the probability-sampling design implemented by the EMA. We demonstrate that the solution to the resulting estimating equations is consistent and asymptotically normally distributed.

In Section 4.2, we will briefly describe a mixed effects version of the modulated Poisson process. A set of simple, unbiased estimating equations will be proposed for the fixed effect parameters and the variance-covariance matrix of the random effects. Finally, this section will also discuss the estimation of these equations and parameters using partially observed covariates. Estimation will be addressed in a general manner that encompasses the techniques for integral estimation proposed by both Rathbun et al. (2007) and Waagepetersen (2008). Section 4.3 will provide the necessary assumptions and theorems to prove the consistency and asymptotic normality of the estimators. Section 4.4 will discuss the post-hoc estimation of the random effects for each subject based on the Penalized Quasi-Likelihood (Breslow &

Clayton 1993). Section 4.5 will illustrate our methods using data from an EMA study of smoking.

4.2 Random Coefficient Modulated Poisson Process

The timing of repeated behavioral events may be modeled as the realization of a temporal point process (Rathbun et al. 2006). However, the techniques presented in this paper may also be applied to spatial or spatio-temporal processes.

A point process may be at least partially described by its *intensity*

$$\lambda(t) = \lim_{\delta \to 0} \left\{ \frac{E\{N([t, t+\delta])\}}{\delta} \right\},$$

where $N([t_1, t_2])$ represents the number of events that occur in the time interval $[t_1, t_2]$. Thus, the intensity measures the instantaneous expected rate of event occurrence as a function of time. When the point process is a Poisson process (Cox 1972), $N(A) \sim Poisson(\Lambda(A))$ for any interval A, where $\Lambda(A) = \int_A \lambda(t)\, dt$, and conditional on $N(A)$, the event locations are independently sampled from the probability density function proportional to $\lambda(t)$. Poisson processes are well-suited to modeling the impact of time-varying covariates on the timing of repeated behavioral events.

For the modulated Poisson process (Cox 1972), the log intensity takes the log-linear form

$$\log \lambda(t; \boldsymbol{\alpha}) = \boldsymbol{\alpha}^T \mathbf{x}(t), \tag{B.1}$$

where $\boldsymbol{\alpha}$ is a p-dimensional vector of coefficient paramters that describe the effects of the covariates on the intensity. Note that, as in survival analysis, $\exp\{\alpha_j\}$ may be interpreted as a risk ratio, quantifying the rate at which event risk increases with every unit increase in the jth covariate.

This model assumes that all subjects have identical baseline rates of event occurrence and that changes in the covariates affect all subjects' event rates identically. Obviously, such an assumption may be untenable, particularly when dealing with behavioral data. Some subjects may have inherent genetic factors that lead to higher or lower baseline smoking rates, and some subjects may be more susceptible to some factors (e.g. restlessness) than others (e.g. the presence of other smokers).

To create a flexible model that allows for such inter-subject variability, we consider a mixed effects version of the model (B.1). Specifically, we will define the intensity for subject i to be

$$\lambda_i(t; \boldsymbol{\alpha}, \boldsymbol{\beta}_i) = \exp\{\boldsymbol{\alpha}^T \mathbf{x}_i(t) + \boldsymbol{\beta}_i^T \mathbf{z}_i(t)\}; t \in A_i, \tag{B.2}$$

where A_i is the study interval for subject i, $\mathbf{x}_i(t)$ and $\mathbf{z}_i(t)$ are p- and q-dimensional vectors of covariates for subject i, $\boldsymbol{\alpha}$ is a vector of constant fixed-effect parameters, and $\boldsymbol{\beta}_i$ is a vector of random parameters that vary by subject. We include separate covariate and parameter vectors to construct a mixed-effects model and therefore allow for the existence of universal effects that may impact event rates equally across subjects.

We assume that the $\boldsymbol{\beta}_i$ are independently sampled from a multivariate normal distribution with mean 0 and covariance matrix $\boldsymbol{\Sigma}$; doing so makes (B.2) similar to a GLMM. Note that if we require $\mathbf{x}_i(t)$ to contain the vector $\mathbf{z}_i(t)$, as well as possible additional covariates, the assumption of a zero mean for the random effects is not restrictive, as the mean covariate effects can be absorbed into $\boldsymbol{\alpha}$.

4.2.1 Parameter Estimation with Fully-Observed Covariates

We will first discuss estimating the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\Sigma}$ for the case where the covariates are fully observed for all subjects. By this we mean that the values of the covariate vectors $\mathbf{x}_i(t)$ and $\mathbf{z}_i(t)$ are known for all $t \in A_{ni}$, $i = 1, \ldots, n$; since our results will with regards to increasing sample size, we will note the study intervals as A_{ni} to represent the ith study interval out of n subjects.

With these assumptions, the likelihood for $(\boldsymbol{\alpha}, \boldsymbol{\Sigma})$ is of the form

$$L_n(\boldsymbol{\alpha}, \boldsymbol{\Sigma}) = \prod_{i=1}^{n} \int_{-\infty}^{\infty} p(N_i; \boldsymbol{\alpha}, \boldsymbol{\beta}_i) p(\boldsymbol{\beta}_i; \boldsymbol{\Sigma}) \, d\boldsymbol{\beta}_i,$$

where $p(\boldsymbol{\beta}_i;\boldsymbol{\Sigma})$ is the pdf for a multivariate normal with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$,

$$p(N_i;\boldsymbol{\alpha},\boldsymbol{\beta}_i)=\exp\left[\left\{\sum_{t\in X_{ni}}\boldsymbol{\alpha}^T\mathbf{x}_i(t)+\boldsymbol{\beta}_i^T\mathbf{z}_i(t)\right\}-\int_{A_{ni}}\exp\{\boldsymbol{\alpha}^T\mathbf{x}_i(t)+\boldsymbol{\beta}_i^T\mathbf{z}_i(t)\}\,dt\right],$$

is the Janossy density (see Borodin & Soshnikov 2003), and X_{ni} is the set of event times. While this likelihood could be calculated using numerical integration, it has no closed form and thus makes it impossible to attain simple score equations. Laplace approximation (e.g. Breslow & Clayton 1993) and quadrature evaluation (e.g. Jansen 1990) are methods of evaluating such a likelihood, but they do not necessarily obtain the simple, closed-form solutions we desire. Monte Carlo Expectation Maximization (MCEM) (e.g. McCulloch 1997) can be applied, but is computationally inefficient, particularly with large datasets, such as those that one would obtain from an EMA study. Therefore, instead of these current methods of evaluation, we will propose alternative estimating equations that are not directly derived from the likelihood. This decision will have a trade-off in an increase in the variance of the estimators for $\boldsymbol{\alpha}$ and $\boldsymbol{\Sigma}$, but will be more computationally efficient.

To obtain estimating equations for $\boldsymbol{\alpha}$, we first consider similar methods to Breslow and Clayton's (1993) version of the Penalized Quasi-Likelihood (Green 1987). We note that our likelihood is of the form

$$L_n(\boldsymbol{\alpha},\boldsymbol{\Sigma})=\prod_{i=1}^n c|\boldsymbol{\Sigma}|^{-1/2}\int_{-\infty}^{\infty}e^{-f_i(\boldsymbol{\alpha},\boldsymbol{\beta}_i,\boldsymbol{\Sigma})}\,d\boldsymbol{\beta}_i,$$

where

$$f_i(\boldsymbol{\alpha},\boldsymbol{\beta}_i,\boldsymbol{\Sigma})=-\sum_{t\in X_{ni}}\{\boldsymbol{\alpha}^T\mathbf{x}_i(t)+\boldsymbol{\beta}_i^T\mathbf{z}_i(t)\}+\int_{A_{ni}}\exp\{\boldsymbol{\alpha}^T\mathbf{x}_i(t)+\boldsymbol{\beta}_i^T\mathbf{z}_i(t)\}\,dt+\frac{1}{2}\boldsymbol{\beta}_i^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}_i,$$

We can take the Laplace approximation (see Tierney & Kadane 1986) to the marginal log-likelihood to obtain

$$L_n(\boldsymbol{\alpha},\boldsymbol{\Sigma})\approx-\frac{n}{2}\log|\boldsymbol{\Sigma}|-\sum_{i=1}^n\frac{1}{2}\log|f_i^{(2)}(\hat{\boldsymbol{\beta}}_i)|-\sum_{i=1}^n f_i(\hat{\boldsymbol{\beta}}_i)+c, \tag{B.3}$$

where c is a constant that does not depend on the model parameters, $\hat{\boldsymbol{\beta}}_i$ is the solution to $f_i^{(1)}(\boldsymbol{\beta}_i)=0$,

$$f_i^{(1)}(\boldsymbol{\beta}_i)=\frac{\partial}{\partial\boldsymbol{\beta}_i^T}f_i(\boldsymbol{\beta}_i)=-\sum_{t\in X_{ni}}\mathbf{z}_i(t)+\int_{A_{ni}}\mathbf{z}_i(t)\exp\{\boldsymbol{\alpha}^T\mathbf{x}_i(t)+\boldsymbol{\beta}_i^T\mathbf{z}_i(t)\}\,dt+\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}_i,$$

and

$$f_i^{(2)}(\boldsymbol{\beta}_i) = \int_{A_{ni}} \mathbf{z}_i(t)\mathbf{z}_i(t)^T \exp\{\boldsymbol{\alpha}^T\mathbf{x}_i(t) + \boldsymbol{\beta}_i^T\mathbf{z}_i(t)\}\, dt + \boldsymbol{\Sigma}^{-1}.$$

Then, substituting into (B.3), we obtain the marginal log-likelihood

$$\begin{aligned} L_n(\boldsymbol{\alpha}, \boldsymbol{\Sigma}) &\approx \sum_{i=1}^{n}\Big\{-\frac{1}{2}\log\Big|I + \boldsymbol{\Sigma}\int_{A_{ni}} \mathbf{z}_i(t)\mathbf{z}_i(t)^T \exp \boldsymbol{\alpha}^T\mathbf{x}_i(t) + \hat{\boldsymbol{\beta}}_i^T\mathbf{z}_i(t)\}\, dt\Big| \\ &+ \sum_{t\in X_{ni}}\Big(\boldsymbol{\alpha}^T\mathbf{x}_i(t) + \hat{\boldsymbol{\beta}}_i^T\mathbf{z}_i(t)\Big) - \int_{A_{ni}} \exp\{\boldsymbol{\alpha}^T\mathbf{x}_i(t) + \hat{\boldsymbol{\beta}}_i^T\mathbf{z}_i(t)\}\, dt \\ &- \frac{1}{2}\hat{\boldsymbol{\beta}}_i^T\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\beta}}_i\Big\} + c. \end{aligned} \tag{B.4}$$

Ignoring the first term in the summand in (B.4), define $\hat{\boldsymbol{\alpha}}$ as the value of $\boldsymbol{\alpha}$ that maximizes the remaining terms in the summand. Thus, we could estimate $(\boldsymbol{\alpha}, \boldsymbol{\beta}_i)$ with $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}_i)$, obtained by jointly maximizing Green's (1987) Penalized Quasi-Likelihood (PQL):

$$\sum_{i=1}^{n}\Big\{\sum_{t\in X_{ni}}\{\boldsymbol{\alpha}^T\mathbf{x}_i(t) + \boldsymbol{\beta}_i^T\mathbf{z}_i(t)\} - \int_{A_{ni}} \exp\{\boldsymbol{\alpha}^T\mathbf{x}_i(t) + \boldsymbol{\beta}_i^T\mathbf{z}_i(t)\}\, dt - \frac{1}{2}\boldsymbol{\beta}_i^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}_i\Big\}. \tag{B.5}$$

Differentiating (B.5) with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}_i$, respectively. gives the biased score functions

$$\sum_{i=1}^{n}\Big[\sum_{t\in X_{ni}}\mathbf{x}_i(t) - \int_{A_{ni}} \mathbf{x}_i(t)\exp\{\boldsymbol{\alpha}^T\mathbf{x}_i(t) + \boldsymbol{\beta}_i^T\mathbf{z}_i(t)\}\, dt\Big] \tag{B.6}$$

and

$$\sum_{i=1}^{n}\Big[\sum_{t\in X_{ni}}\mathbf{z}_i(t) - \int_{A_{ni}} \mathbf{z}_i(t)\exp\{\boldsymbol{\alpha}^T\mathbf{x}_i(t) + \boldsymbol{\beta}_i^T\mathbf{z}_i(t)\}\, dt - \boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}_i\Big]. \tag{B.7}$$

The estimating function (B.6) is a function of the unobserved random effects $\boldsymbol{\beta}_i$. Since estimates of these random effects are based on the limited information from individual subjects, estimates of $\boldsymbol{\alpha}$ based on the joint solution (B.6) and (B.7) are inefficient. Therefore it is desirable to obtain an estimating function for $\boldsymbol{\alpha}$ that is free of $\boldsymbol{\beta}_i$; $i = 1, \ldots, n$.

Taking the first term of (B.6) minus its expected value yields an estimating function that is free of the random effects and is unbiased in the sense that it hs expected value zero under the true value of $\boldsymbol{\theta}$. To obtain an expression for this estimating function, define $\lambda_i^*(t; \boldsymbol{\alpha}) = \exp\{\boldsymbol{\alpha}^T\mathbf{x}_i(t)\}$ and the moment generating function for the random effect

$$M_{\boldsymbol{\beta}_i}(\mathbf{z}_i(t); \boldsymbol{\Sigma}_{\mathbf{0}}) = E[\exp\{\boldsymbol{\beta}_i^T\mathbf{z}_i(t)\}] = \exp\{\frac{1}{2}\mathbf{z}_i(t)^T\boldsymbol{\Sigma}_{\mathbf{0}}\mathbf{z}_i(t)\}. \tag{B.8}$$

Then our unbiased estimating function for $\boldsymbol{\alpha}$ is

$$\mathbf{A}_n(\boldsymbol{\theta}) = n^{-1}\sum_{i=1}^{n}\left[\sum_{t\in X_{ni}}\mathbf{x}_i(t) - \boldsymbol{\Phi}_{ni}(\boldsymbol{\theta})\right], \tag{B.9}$$

where the sum is divided by sample size for technical reasons related to the consistency of the estimators,

$$\boldsymbol{\Phi}_{ni}(\boldsymbol{\theta}) = E\left\{\sum_{t\in X_{ni}}\mathbf{x}_i(t)\right\} = \int_{A_{ni}}\phi_i(t;\boldsymbol{\theta})\,dt, \tag{B.10}$$

and

$$\phi_i(t;\boldsymbol{\theta}) = \mathbf{x}_i(t)\lambda_i^*(t;\boldsymbol{\alpha})M_{\boldsymbol{\beta}_i}(\mathbf{z}_i(t);\boldsymbol{\Sigma}).$$

Since our focus at the moment is on the fixed parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\Sigma}$, we will return to a discussion of equation (B.7) in Section 4.4.

4.2.2 Estimation of Covariance Components

In order to discuss the estimating equations for $\boldsymbol{\Sigma}$, it is simpler if we redefine $\boldsymbol{\theta}$ as $(\boldsymbol{\alpha}^T, \mathbf{v}^T)$, where $\mathbf{v} = \mathrm{vech}(\boldsymbol{\Sigma})$, with vech as the "vector-half" function that stacks the upper or lower-triangular columns of a symmetric matrix to create a non-redundant vector (see Henderson & Searle 1979). Define

$$\mathbf{y}_i(s,t) = \mathrm{vech}\{\mathbf{z}_i(s)\mathbf{z}_i(t)^T\},$$

so that

$$\mathbf{y}_i(s,t) = (z_{i,1}(s)z_{i,1}(t), \ldots, z_{i,q}(s)z_{i,1}(t), z_{i,2}(s)z_{i,2}(t), \ldots, z_{i,q}(s)z_{i,2}(t), \ldots, z_{i,q}(s)z_{i,q}(t))^T.$$

Note that $\mathbf{y}_i(s,t)$ is not an application of the *vech* operation to a necessarily symmetric matrix, but it is notationally convenient. Then a similar set of non-redundant estimating equations for $\mathbf{v}$ are

$$\mathbf{B}_n(\boldsymbol{\theta}) = n^{-1}\sum_{i=1}^{n}\left[\sum_{s\neq t\in X_{ni}}\mathbf{y}_i(s,t) - \boldsymbol{\Omega}_{ni}(\boldsymbol{\theta})\right], \tag{B.11}$$

where

$$\boldsymbol{\Omega}_{ni}(\boldsymbol{\theta}) = E\left\{\sum_{s\neq t\in X_{ni}}\mathbf{y}_i(s,t)\right\} = \int_{A_{ni}}\int_{A_{ni}}\omega_i(s,t;\boldsymbol{\theta})\,ds\,dt, \tag{B.12}$$

and

$$\omega_i(s,t;\boldsymbol{\theta}) = \mathbf{y}_i(s,t)\lambda_i^*(s;\boldsymbol{\alpha})\lambda_i^*(t;\boldsymbol{\alpha})M_{\boldsymbol{\beta}_i}(\mathbf{z}_i(s)+\mathbf{z}_i(t);\boldsymbol{\Sigma}).$$

The estimating functions (B.11) are chosen due to their simple form and associated computational efficiency, and their similarity in form to (B.9) and the variance estimation equation in Neustifter, Rathbun, and Shiffman (2009). We exclude the cases where $s = t$ in the sum in (B.11) to eliminate an additional integral term resulting from computing the expected value of $\sum_{s\in X_{ni}} \mathbf{y}_i(s,s)\mathbf{y}_i^T(s,s)$.

Note that while (B.9), when conditioned upon the realization of the random effect, remains unbiased even if the data are realized from a non-Poisson point process, the unbiased nature of $\mathbf{B}_n(\boldsymbol{\theta})$ depends on the independent increments property of Poisson processes. If the assumption of a Poisson process is violated, unbiased estimating equations for $\boldsymbol{\Sigma}$ can be constructed, but they will be functions of the second-order intensity of the point process (Diggle 2003, p. 43). Extending our research to this more general case would add undue complication to already complex results without adding to the understanding of the approach.

To simply the notation, we will call $\boldsymbol{\Psi}_n(\boldsymbol{\theta}) = (\mathbf{A}_n(\boldsymbol{\theta})^T, \mathbf{B}(\boldsymbol{\theta})^T)^T$. Note that we can write this estimating function as

$$\boldsymbol{\Psi}_n(\boldsymbol{\theta}) = n^{-1}\sum_{i=1}^{n}\{\mathbf{X}_{ni}(\boldsymbol{\theta}) - \boldsymbol{\Lambda}_{ni}(\boldsymbol{\theta})\}, \tag{B.13}$$

where

$$\mathbf{X}_{ni}(\boldsymbol{\theta}) = \left(\sum_{t\in X_{ni}} \mathbf{x}_i(t)^T, \sum_{s\neq t\in X_{ni}} \mathbf{y}_i(s,t)^T\right)^T$$

and

$$\boldsymbol{\Lambda}_{ni}(\boldsymbol{\theta}) = (\boldsymbol{\Phi}_{ni}(\boldsymbol{\theta}), \boldsymbol{\Omega}_{ni}(\boldsymbol{\theta}))^T.$$

4.2.3 Parameter Estimation with Partially-Observed Covariates

The estimating equations proposed in Sections 4.2.1 and 4.2.2 involve terms of the form (B.10) and (B.12), where the integrands are functions of the covariates. Thus, to implement the above parameter estimation procedures, the covariates must be known functions of time.

However, we can also regard $\boldsymbol{\Phi}_{ni}(\boldsymbol{\theta})$ and $\boldsymbol{\Omega}_{ni}(\boldsymbol{\theta})$ as population totals of $\phi_i(t;\boldsymbol{\theta})$ and $\omega_i(s,t;\boldsymbol{\theta})$, respectively, over the respective populations of points in A_{ni} and $A_{ni} \times A_{ni}$. Suppose for each subject i, the covariates are independently sampled at sites $\{u; u \in D_{ni}\}$, where D_{ni} is the total set of such sampled "dummy" points, and that these come from a known probability-based sampling design with first and second-order inclusion densities $\pi_i(t)$ and $\pi_i(s,t)$, as defined in Cordy (1993). In this case, we can form unbiased estimators of the integrals $\boldsymbol{\Phi}_{ni}(\boldsymbol{\theta})$ and $\boldsymbol{\Omega}_{ni}(\boldsymbol{\theta})$.

Rathbun et al. (2007) suggest the Horvitz-Thompson estimator for $\boldsymbol{\Phi}_{ni}(\boldsymbol{\theta})$

$$\hat{\boldsymbol{\Phi}}_{ni}(\boldsymbol{\theta}) = \sum_{u \in D_{ni}} \frac{\phi_i(u;\boldsymbol{\theta})}{\pi_{ni}(u)}. \tag{B.14}$$

An analogous estimator for $\boldsymbol{\Omega}_{ni}(\boldsymbol{\theta})$ is

$$\hat{\boldsymbol{\Omega}}_{ni}(\boldsymbol{\theta}) = \sum_{u \neq v \in D_{ni}} \frac{\omega_i(u,v;\boldsymbol{\theta})}{\pi_{ni}(u,v)}. \tag{B.15}$$

Both (B.14) and (B.15) are design-unbiased for their respective terms, but each only uses information from the random covariate samples, not from the event times. We use the phrase *design-unbiased* (Cassel et al. 1997) to refer to an estimator that has expectation equal to its associated parameter under the sampling design's induced probability distribution. Waagepetersen (2008) suggests a more precise estimator for $\boldsymbol{\Phi}_{ni}(\boldsymbol{\theta})$ that is a function of the event as well as the dummy assessments:

$$\hat{\boldsymbol{\Phi}}_{ni}(\boldsymbol{\theta}) = \sum_{t \in X_{ni} \cup D_{ni}} \frac{\phi_i(t;\boldsymbol{\theta})}{\pi_{ni}(t) + \lambda_i^*(t;\boldsymbol{\alpha}) M_{\boldsymbol{\beta}_i}(\mathbf{z}_i(t);\boldsymbol{\Sigma})}. \tag{B.16}$$

An analogous estimator for $\boldsymbol{\Omega}_{ni}(\boldsymbol{\theta})$ is

$$\hat{\boldsymbol{\Omega}}_{ni}(\boldsymbol{\theta}) = \sum_{s \neq t \in X_{ni} \cup D_{ni}} \frac{\omega_i(s,t;\boldsymbol{\theta})}{\pi_{ni}(s,t) + \lambda_i^*(s;\boldsymbol{\alpha})\lambda_i^*(t;\boldsymbol{\alpha}) M_{\boldsymbol{\beta}_i}(\mathbf{z}_i(s) + \mathbf{z}_i(t);\boldsymbol{\Sigma})}.$$

Note that use of either of Rathbun et al. or Waagepetersen's estimators gives estimating equations of the form

$$\tilde{\mathbf{A}}_n(\boldsymbol{\theta}) = n^{-1} \sum_{i=1}^{n} \left[\sum_{t \in X_{ni}} f(t;\boldsymbol{\theta}) \mathbf{x}_i(t) - \sum_{u \in D_{ni}} \frac{f(u;\boldsymbol{\theta})\phi_i(u;\boldsymbol{\theta})}{\pi_{ni}(u)} \right] \tag{B.17}$$

and

$$\tilde{\mathbf{B}}_n(\boldsymbol{\theta}) = n^{-1}\sum_{i=1}^{n}\Bigg[\sum_{s\neq t\in X_{ni}} g(s,t;\boldsymbol{\theta})\mathbf{y}_i(s,t) - \sum_{u\neq v\in D_{ni}} \frac{g(u,v;\boldsymbol{\theta})\omega_i(u,v;\boldsymbol{\theta})}{\pi_{ni}(u,v)}\Bigg], \tag{B.18}$$

where $f(t;\boldsymbol{\theta}) = 1$ for Rathbun et al. (2007), $f(t;\boldsymbol{\theta}) = \pi_{ni}(t)[\pi_{ni}(t) + \lambda_i(t;\boldsymbol{\alpha})M_{\boldsymbol{\beta}_i}(\mathbf{z}_i(t);\boldsymbol{\Sigma})]^{-1}$ for Waagepetersen (2008), $g(s,t;\boldsymbol{\theta}) = 1$ for Rathbun et al., and Waagepetersen (2008) takes

$$g(s,t;\boldsymbol{\theta}) = \frac{\pi_{ni}(s,t)}{\pi_{ni}(s,t) + \lambda_i(s;\boldsymbol{\alpha})\lambda_i(t;\boldsymbol{\alpha})M_{\boldsymbol{\beta}_i}(\mathbf{z}_i(s)+\mathbf{z}_i(t);\boldsymbol{\Sigma})}.$$

As in (B.13), we can define $\tilde{\boldsymbol{\Psi}}_n(\boldsymbol{\theta}) = (\tilde{\mathbf{A}}_n(\boldsymbol{\theta})^T, \tilde{\mathbf{B}}_n(\boldsymbol{\theta})^T)^T$ and rewrite it as

$$\tilde{\boldsymbol{\Psi}}_n(\boldsymbol{\theta}) = n^{-1}\sum_{i=1}^{n}\{\mathbf{X}^*_{ni}(\boldsymbol{\theta}) - \boldsymbol{\Lambda}^*_{ni}(\boldsymbol{\theta})\}, \tag{B.19}$$

where

$$\mathbf{X}^*_{ni}(\boldsymbol{\theta}) = \Bigg(\sum_{t\in X_{ni}} f(t;\boldsymbol{\theta})\mathbf{x}_i(t)^T, \sum_{s\neq t\in X_{ni}} g(s,t;\boldsymbol{\theta})\mathbf{y}_i(s,t)^T\Bigg)^T$$

and

$$\boldsymbol{\Lambda}^*_{ni}(\boldsymbol{\theta}) = \Bigg(\sum_{u\in D_{ni}} \frac{f(u;\boldsymbol{\theta})\phi_i(u;\boldsymbol{\theta})^T}{\pi_{ni}(u)}, \sum_{u\neq v\in D_{ni}} \frac{g(u,v;\boldsymbol{\theta})\omega_i(u,v;\boldsymbol{\theta})^T}{\pi_{ni}(u,v)}\Bigg)^T.$$

These equations can be considered weighted versions of $\mathbf{X}_{ni}(\boldsymbol{\theta})$ and $\boldsymbol{\Lambda}_{ni}(\boldsymbol{\theta})$; note that $\mathbf{X}^*_{ni}(\boldsymbol{\theta})$ and $\boldsymbol{\Lambda}^*_{ni}(\boldsymbol{\theta})$ are not, however, unbiased estimators of $\mathbf{X}_{ni}(\boldsymbol{\theta})$ and $\boldsymbol{\Lambda}_{ni}(\boldsymbol{\theta})$ if f and g are chosen to be anything other than uniformly equal to 1.

In the following section, we will show that the solution to $\tilde{\boldsymbol{\Psi}}_n(\boldsymbol{\theta}) = \mathbf{0}$ yields an estimator $\tilde{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta}_{\mathbf{0}}$ that is consistent and asymptotically normal as $n \to \infty$.

4.3 Asymptotic Properties of the Proposed Estimators

To prove that $\tilde{\boldsymbol{\theta}}_n \xrightarrow{P} \boldsymbol{\theta}_{\mathbf{0}}$ as $n \to \infty$, we must make some weak assumptions regarding the covariates, event and random assessments, and the inclusion densities $\pi_{ni}(t)$.

Assumption 13. *The subjects $i = 1, \ldots, n$ are independently sampled, and all event times are independent of all dummy assessment times, both within and among subjects.*

Assumption 14. *There exists a constant M such that $|\,\mathbf{x}_i(t)\,| < M$ for almost all $t > 0$ and all $i = 1, \ldots, n$.*

Assumption 15. *There exists a constant C such that $|A_{ni}| < C$ for all $i = 1, \ldots, n$.*

Assumption 16.

$$n^{-1} \sum_{i=1}^{n} \int_{A_{ni}} \mathbf{x}_i(t)\mathbf{x}_i(t)^T dt \xrightarrow{P} B_1$$

as $n \to \infty$, where the minimum eigenvalue of B_1 is strictly positive.

Assumption 17. *Define*

$$\mathbf{d}_i(s,t) = vech(\mathbf{D}_i(s,t)), \tag{B.1}$$

where $\mathbf{D}_i(s,t)$ is a matrix with elements

$$[\mathbf{D}_i(s,t)]_{j,k} = (1 + I[j \neq k])\{(z_{i,j}(s) + z_{i,j}(t))(z_{i,k}(s) + z_{i,k}(t))\}; j, k = 1, \ldots, q,$$

and $z_{i,j}(t)$ is the jth element of the vector $\mathbf{z}_i(t)$. Then

$$\frac{1}{2n} \sum_{i=1}^{n} \int_{A_{ni}} \int_{A_{ni}} [\mathbf{y}_i(s,t)\mathbf{d}_i(s,t)^T + \mathbf{d}_i(s,t)\mathbf{y}_i(s,t)^T]\, ds\, dt \xrightarrow{P} B_2$$

as $n \to \infty$, where the minimum eigenvalue of B_2 is strictly positive.

Assumption 18. *The random covariate assessments $\{u \in D_{ni};\ i = 1, \ldots, n\}$ are obtained according to a known probability-based sampling scheme with single and pairwise inclusion densities $\pi_{ni}(u)$ and $\pi_{ni}(u,v)$ (Cordy 1993), where the sampling intensity is such that*

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} m_i}{\sum_{i=1}^{n} |A_{ni}|} = \rho$$

for some $0 < \rho < \infty$ and $i = 1, \ldots, n$.

Assumption 19.

$$\lim_{n\to\infty} \inf_{t \in A_{ni}} \pi_{ni}(t) = c_1$$

for some constant $c_1 > 0$, for any $i = 1, \ldots, n$.

Assumption 20.

$$\lim_{n\to\infty} \inf_{s,t\in A_{ni}} \pi_{ni}(s,t) = c_2$$

for some constant $c_2 > 0$, for any $i = 1, \ldots, n$.

Assumption 21. *Define*

$$\lim_{n\to\infty} E_{\boldsymbol{\theta_0}} \left\{ \frac{\partial}{\partial \boldsymbol{\theta}^T} \tilde{\boldsymbol{\Psi}}_n(\boldsymbol{\theta_0}) \right\} = \mathbf{J}_\infty(\boldsymbol{\theta_0}), \tag{B.2}$$

where $\mathbf{J}_\infty(\boldsymbol{\theta_0})$ is nonsingular.

These assumptions are related to those made by Rathbun et al. (2007). Under Assumption 13, $\mathbf{X}^*_{ni}(\boldsymbol{\theta})$ and $\boldsymbol{\Lambda}^*_{ni}(\boldsymbol{\theta})$ in (B.19) are independent sequences of mutually independent random vectors. Assumption 14 prevents the estimating functions from diverging, and ensures that the integral estimators in Section 4.2.3 are unbiased (Cordy 1993). Assumption 15 keeps the estimating functions from diverging and prevents a subset of subjects with infinitely large study intervals from dominating the results. Assumptions 16 and 17 are necessary to ensure that the derivatives of the estimating equations are positive definite. Assumption 16 also guarantees that the covariate vectors do not contain redundant information. Assumption 17 should be true for most covariate vectors $\mathbf{z}_i(t)$ that satisfy Assumption 16 and do not behave in a pathological fashion. Assumptions 18 through 20 make some weak restrictions upon the probability sampling design used to sample the covariates. Assumption 18 requires that the sampling intensity approaches a constant as sample size increases. Assumptions 19 and 20 are necessary to ensure that the dummy assessments provide adequate coverage of the study region, and for the unbiased estimation of (B.10) and (B.12). Assumption 21 is required for the information matrix to be invertible, to ensure the estimators are asymptotically normal.

Under these assumptions, it is possible to prove that the solution to (B.19) is consistent as sample size n increases.

Theorem 5. *Under Assumptions 13 through 20, the solution $\tilde{\boldsymbol{\theta}}_n$ to (B.19) is a consistent estimator of the true parameter vector $\boldsymbol{\theta_0} = (\boldsymbol{\alpha_0}, \mathbf{v_0})^T$ as $n \to \infty$.*

The proof of this theorem may be found in Appendix A.

We can also show that, under Assumptions 13 through 21, $\tilde{\boldsymbol{\theta}}_n$ is asymptotically normal. To discuss the asymptotic variance of this estimator, we define

$$\mathbf{S}^X_\infty(\boldsymbol{\theta_0}) = \lim_{n\to\infty} \mathbf{S}^X_n(\boldsymbol{\theta_0}), \text{ where } \mathbf{S}^X_n(\boldsymbol{\theta_0}) = n^{-1} Var_{\boldsymbol{\theta_0}}\left\{\sum_{i=1}^n \mathbf{X}^*_{ni}(\boldsymbol{\theta_0})\right\}, \tag{B.3}$$

and

$$\mathbf{S}^\Lambda_\infty(\boldsymbol{\theta_0}) = \lim_{n\to\infty} \mathbf{S}^\Lambda_n(\boldsymbol{\theta_0}), \text{ where } \mathbf{S}^\Lambda_n(\boldsymbol{\theta_0}) = n^{-1} Var_{\boldsymbol{\theta_0}}\left\{\sum_{i=1}^n \boldsymbol{\Lambda}^*_{ni}(\boldsymbol{\theta_0})\right\}. \tag{B.4}$$

The existence of these limits in the neighborhood of $\boldsymbol{\theta_0}$ is guaranteed under Assumption 14. The full form of $\mathbf{S}^X_n(\boldsymbol{\theta_0})$ and $\mathbf{S}^\Lambda_n(\boldsymbol{\theta_0})$ are provided in Appendix C. Note that while $\mathbf{J}_\infty(\boldsymbol{\theta_0})$ is non-symmetric, it is a 2×2 block matrix that can be inverted by applying the common formula for the inverse of a block matrix.

Theorem 6. *Under Assumptions 13 through 21,*

$$n^{\frac{1}{2}}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0}) \xrightarrow{D} \mathbf{N}(\mathbf{0}, \mathbf{V}(\boldsymbol{\theta_0})),$$

where

$$\mathbf{V}(\boldsymbol{\theta_0}) = \{\mathbf{J}_\infty(\boldsymbol{\theta_0})^{-1}\}\mathbf{S}^X_\infty(\boldsymbol{\theta_0})\{\mathbf{J}_\infty(\boldsymbol{\theta_0})^{-1}\}^T + \{\mathbf{J}_\infty(\boldsymbol{\theta_0})^{-1}\}\mathbf{S}^\Lambda_\infty(\boldsymbol{\theta_0})\{\mathbf{J}_\infty(\boldsymbol{\theta_0})^{-1}\}^T, \tag{B.5}$$

where $\mathbf{J}_\infty(\boldsymbol{\theta_0})$, $\mathbf{S}^X_\infty(\boldsymbol{\theta_0})$, *and* $\mathbf{S}^\Lambda_\infty(\boldsymbol{\theta_0})$ *are as defined in (B.2), (B.3), and (B.4), respectively.*

As in Rathbun et al. (2007), the variance of $\tilde{\boldsymbol{\theta}}_n$ can be partitioned into two parts: $\{\mathbf{J}_\infty(\boldsymbol{\theta_0})^{-1}\}\mathbf{S}^X_\infty(\boldsymbol{\theta_0})\{\mathbf{J}_\infty(\boldsymbol{\theta_0})^{-1}\}^T$ and $\{\mathbf{J}_\infty(\boldsymbol{\theta_0})^{-1}\}\mathbf{S}^\Lambda_\infty(\boldsymbol{\theta_0})\{\mathbf{J}_\infty(\boldsymbol{\theta_0})^{-1}\}^T$.

Note that many of the terms in (B.2), (B.3), and (B.4) do not need to be estimated by integral estimators such as (B.14) and (B.16). Instead, one may obtain simpler unbiased estimators by utilizing that

$$E_{\boldsymbol{\theta_0}}\left\{\sum_{t\in X_{ni}} \mathbf{x}_i(t)\right\} = \int_{A_{ni}} \phi_i(t;\boldsymbol{\theta_0})\,dt.$$

For example, to estimate (C.1), the upper-left block of $\mathbf{J}_\infty(\boldsymbol{\theta_0})$, one can use the sum

$$\hat{J}_{11,\infty}(\tilde{\boldsymbol{\theta}}_n) = -n^{-1}\sum_{i=1}^n\left\{\sum_{t\in X_{ni}} f(t;\tilde{\boldsymbol{\theta}}_n)\mathbf{x}_i(t)\mathbf{x}_i(t)^T\right\}.$$

The other portions of $\mathbf{J}_\infty(\boldsymbol{\theta}_\mathbf{0})$, as well much of the $\mathbf{S}_\infty$ matrices, may be estimated in a similar manner by exploiting the expected intensity of the process.

4.4 Estimation of the $\boldsymbol{\beta}_i$s

For model diagnostics, as well as in many research situations, it is desirable to estimate the random effects $\boldsymbol{\beta}_i$ for each subject i. Doing so will allow us to observe estimates for deviations from the assumed Multivariate Normal distribution, as well as to identify subjects for whom random effects have particularly strong or weak effects on event timing.

One way of estimating the subject-level terms would be maximize the Extended log-likelihood (ELLH) (Lee, Nelder, & Pawitan 2006)

$$L_n(\boldsymbol{\theta}, \boldsymbol{\beta_n}) = \ell_n(\boldsymbol{\alpha}|\boldsymbol{\beta_n}.\boldsymbol{\Sigma}) + \ell_n(\boldsymbol{\beta_n}|\boldsymbol{\Sigma}). \tag{B.1}$$

However, since $\boldsymbol{\beta}_i$ is not the canonical parameter for the Poisson process, (B.1) would be an extended likelihood (rather than an h-likelihood), and extended likelihood maximization may provide biased estimates of random effect terms (see Lee, Nelder & Pawitan 2006). Instead, we estimate the random effects using the PQL equation (B.7), defined in Section 4.2.1.

As with (B.17) and (B.18), we can estimate the integrals in (B.7) using the dummy assessments. By doing so, as well as plugging in our estimates $\tilde{\boldsymbol{\alpha}}_n$ and $\tilde{\boldsymbol{\Sigma}}$, we obtain the estimating functions for the $\boldsymbol{\beta}_i$s

$$\tilde{\boldsymbol{\kappa}}(\boldsymbol{\beta}_i, \tilde{\boldsymbol{\theta}}_n) = |A_{ni}|^{-1}\Bigg(\sum_{t \in X_{ni}} h(t; \tilde{\boldsymbol{\alpha}}_n, \boldsymbol{\beta}_i)\mathbf{z}_i(t)$$
$$- \sum_{u \in D_{ni}} \frac{h(u; \tilde{\boldsymbol{\alpha}}_n, \boldsymbol{\beta}_i)\mathbf{z}_i(u)\exp\{\tilde{\boldsymbol{\alpha}}_n^T \mathbf{x}_i(u) + \boldsymbol{\beta}_i^T \mathbf{z}_i(u)\}}{\pi_{ni}(u)} - \tilde{\boldsymbol{\Sigma}}_n^{-1}\boldsymbol{\beta}_i\Bigg)\ ;\ i = 1, \ldots, n, \tag{B.2}$$

where $h(t; \tilde{\boldsymbol{\alpha}}_n, \boldsymbol{\beta}_i) = 1$ for Rathbun et al.'s estimation method, and

$$h(t; \tilde{\boldsymbol{\alpha}}_n, \boldsymbol{\beta}_i) = \pi_{ni}(t)[\pi_{ni}(t) + \exp\{\tilde{\boldsymbol{\alpha}}_n^T \mathbf{x}_i(t) + \boldsymbol{\beta}_i \mathbf{z}_i(t)]^{-1}$$

for Waagepetersen's. The asymptotic properties for the solution $\tilde{\boldsymbol{\beta}}_i$ to $\tilde{\boldsymbol{\kappa}}(\boldsymbol{\beta}_i, \tilde{\boldsymbol{\theta}}) = 0$ are investigated under increasing domain asymptotics, under which the lengths $|A_{ni}|$ of the study intervals for each subject increase. Division by $|A_{ni}|$ in (B.2) is done for technical reasons related to our investigation of these asymptotic properties. Note that, if we consider $\boldsymbol{\alpha}$ and $\boldsymbol{\Sigma}$ as nuisance parameters, (B.2) is simply a penalized version of the estimating functions obtained by Rathbun et al. (2007) and Waagepetersen (2008). The proof of the asymptotic consistency, as $|A_{ni}| \to \infty$ for all i, of the solutions $\tilde{\boldsymbol{\beta}}_{\boldsymbol{n}}$ to (B.2) is similar to the consistency proofs in these articles. Regarding the asymptotic normality of $|A_{ni}|^{1/2}(\tilde{\boldsymbol{\beta}}_{\boldsymbol{n}} - \tilde{\boldsymbol{\beta}}_{\boldsymbol{n}0})$, we must make some assumptions about the rate at which the lengths of the study intervals $|A_{ni}|$ approach infinity. First, we need that, for all n,

$$\frac{\min_i |A_{ni}|}{\max_i |A_{ni}|} \geq r$$

for some $r > 0$. This prevents a single individual from dominating the others and, therefore, that individual's covariates having disproportionate control over the estimates of $\tilde{\boldsymbol{\alpha}}_n$ and $\tilde{\boldsymbol{\Sigma}}_n$. Second, we must assume that $\min_i |A_{ni}| \propto n$ for all n. This ensures that the variance of $\tilde{\boldsymbol{\alpha}}_n$ converges when $\tilde{\boldsymbol{\alpha}}_n$ is multiplied by $|A_{ni}|^{1/2}$. Under these assumptions, then, (B.2) satisfies the conditions of Theorem 1 in Yuan and Jennrich's (2000), weighting by $|A_{ni}|^{1/2}$ rather than $n^{1/2}$, which gives the asymptotic normality of $\tilde{\boldsymbol{\beta}}_{\boldsymbol{n}}$. The asymptotic variance is calculated directly, using Taylor Approximations to $(\tilde{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_{\mathbf{0}i})$ and $(\tilde{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_{\mathbf{0}})$. The process of obtaining this variance is lengthy, but its final form can be found in Appendix E.

4.5 EMA of Smoking

Mixed effects versions of the modulated Poisson process are illustrated using data from an EMA of smoking (Shiffman et al. 2002). In this study, 304 smokers were instructed to record on electronic diaries every cigarette they smoked. At randomly selected "events" of a cigarette smoking and randomly selected times, the diaries prompted subjects to fill out assessments. These diary assessments included questions regarding mood and environment. Negative Affect, Arousal, Attention, and Restlessness are the four of these variables analyzed by Rathbun et al. (2007). Negative Affect, Arousal, and Attention were all variables created from a factor analysis of questions asking subjects to respond to descriptive adjectives. Restlessness was assessed from subjects' responses to an adjective for which the data did no load into the three main factors of Negative Affect, Arousal, or Attention in the Shiffman et al. data. Restlessness was the only one of these four variables that, by Rathbun et al.'s analysis, had a significant impact on the timing of smoking events. In addition, we will consider the time-varying environmental variable, Other Smokers, and the subject-level variable Age (in years), modeled as a fixed effect. Other Smokers was a binary variable marking whether the subject indicated that other smokers were present at the time of the assessment.

The prompts for both non-event and event assessments were chosen randomly by the diary according to known sampling designs. For the non-event assessments, these times were selected according to a stratified scheme, where days were the within-subject strata. This stratified scheme was placed under some restrictions: these dummy assessments could not occur within the 10 minutes following an event, and the diary could not prompt for assessments when the diary was off or placed into a "stand-by" mode for situations such as meetings, driving a vehicle, etc. The stratified probabilities were chosen such that there were an expected number of five dummy assessments per day per subject.

The event assessments were randomly chosen from a Bernoulli sampling scheme, where each day had a constant probability of an event being assessed. This probability was based upon the number of events for that subject on the previous day, such that the expected number of events for the current day would be five. Specifically,

$$p_{ij} = \min\{5N_{i,j-1}^{-1}, 1\}, \tag{B.1}$$

where p_{ij} is the probability of event assessment for subject i on day j, $N_{i,j-1}$ is the number of events for the ith subject on day $j-1$.

The well-known result that if one takes a Poisson process and random deletes each point within that process according to a Bernoulli trial with known probability, the resultant series of points is also a Poisson process (Cressie 1991) ensures that the methods and results of this paper may be applied to data such as these.

Table 4.1: Fixed/Mean-effect estimates for first model
(Note: Starred variables treated as random effects)

Covariate	Estimate	Std. Error	p-value
Intercept	-0.2642	0.2242	0.1194
Other Smokers*	0.4655	0.0491	<0.0001
Restlessness*	0.1164	0.8130	0.4431
Age	0.0039	0.0032	0.1108

Table 4.2: Variance/covariance parameters for random effects, first model

Covariate 1	Covariate 2	Estimate	Std. Error	p-value
Other Smokers	Other Smokers	0.2496	0.2242	0.1328
Other Smokers	Restlessness	-0.07417	0.0491	0.0654
Restlessness	Restlessness	0.0127	0.8130	0.4938

Estimates of $\tilde{\boldsymbol{\theta}}_n$ were obtained by solving modified versions of (B.17) and (B.18) that accounted for the stratification by day and the effects upon the process intensity from the Bernoulli sampling of events. These modified estimating equations were

$$\tilde{\mathbf{A}}_n(\boldsymbol{\theta}) = n^{-1}\sum_{i=1}^{n}\left[\sum_{j=1}^{d_i}\left\{\sum_{t\in X_{ni}} f(t;\boldsymbol{\theta})\mathbf{x}_i(t) - p_{ij}\sum_{u\in D_{ni}}\frac{f(u;\boldsymbol{\theta})\phi_i(u;\boldsymbol{\theta})}{\pi_{ni}(u)}\right\}\right]$$

and

$$\tilde{\mathbf{B}}_n(\boldsymbol{\theta}) = n^{-1}\sum_{i=1}^{n}\left[\sum_{j=1}^{d_i}\left\{\sum_{s\neq t\in X_{ij}} g(s,t;\boldsymbol{\theta})\mathbf{y}_i(s,t) - p_{ij}\sum_{u\neq v\in D_{ij}}\frac{g(u,v;\boldsymbol{\theta})\omega_i(u,v;\boldsymbol{\theta})}{\pi_{ni}(u,v)}\right\}\right],$$

where d_i is the number of days the ith subject participated in the study, X_{ij} and D_{ij} are the sets of event and dummy assessment times, respectively, for the ith subject on the jth day, and p_{ij} is as defined in (B.1). For the purposes of this paper, only Rathbun et al.'s (2007) estimation method was utilized, due to its greater computational simplicity.

Tables 4.1 and 4.2 contain the results for the initial analysis of the data, using “Age” as a fixed-effect covariate, and “Other Smokers” and “Restlessness” as random-effect covariates.

Table 4.3: Fixed/Mean-effect estimates for second model
(Note: Starred variables treated as random effects)

Covariate	Estimate	Std. Error	p-value
Intercept	-0.1523	0.0660	0.0105
Other Smokers*	0.4119	0.0432	$<$0.0001
Restlessness	0.0808	0.0139	$<$0.0001
Age	0.0033	0.0014	0.0095

Table 4.1 contains the estimates, estimated standard errors, and significance tests (versus $\alpha_i = 0$) for the covariate effects; these include the effect of "Age" and the mean effects of "Other Smokers" and "Restlessness." Table 4.2 contains the estimates of the elements of $\mathbf{\Sigma}$ and the associated estimated standard errors and significance tests (versus $\Sigma_{ij} = 0$). The mixed effects model conflicts with Rathbun et al.'s (2007) results, which found increasing Restlessness to have a significant positive effect on the timing of cigarette-smoking events, increasing their frequency. The estimates of the elements of $\mathbf{\Sigma}$ are all found to not be significantly different from 0, as well. "Age," which, as a subject-level covariate, would affect the baseline rate of smoking, also appears to be non-significant. Since, from $\mathbf{\Sigma}$, the random effect for Restlessness appeared to have little variance, it was decided to model Restlessness as a fixed effect, instead. This near-zero variance for Restlessness may be the reason for an abnormality in the $\mathbf{\Sigma}$ estimates; note from Table 4.2 that the correlation between Other Smokers and Restlessness is estimated to be approximately -1.32. This was taken as further evidence that the variance for Restlessness is too small, as it is the most likely reason that this correlation is larger than 1 in magnitude. The decision to model Restlessness as fixed-effect affected the standard errors and, resultantly, the significance of the other variables, as well as for Restlessness.

Tables 4.3 and 4.4 contain the results for the second analysis, where both "Age" and "Restlessness" are treated as fixed-effect variables, and "Other Smokers" as a random-effect variable. Now all variables are found to significantly impact the timing of events, and the

Table 4.4: Variance parameter for random effect, second model

Covariate 1	Covariate 2	Estimate	Std. Error	p-value
Other Smokers	Other Smokers	0.0977	0.0660	.0694

estimate for the variance of the distribution of the effects of Other Smokers is marginally significantly different from 0. These results agree with Rathbun et al.'s results as well as previous literature (Shiffman et al. 1996, Shiffman et al. 2002) by having Restlessness significantly impact smoking timing in a positive manner. This model further builds upon these results, suggesting the Restlessness has a universal positive effect on smoking times, rather than varying among subjects in its effects. In summary, Restlessness was shown to have a significant relationship to increased smoking rate, as did Age and the presence of Other Smokers.

Figure 4.1 shows a histogram of the estimated β_i values for "Other Smokers." Superimposed upon the graph is the true normal distribution based upon the estimate of $\tilde{\boldsymbol{\Sigma}}_n$. Clearly the $\tilde{\beta}_i$s do not follow this normal curve; they are strongly bimodal with a tendency toward right skewness. One reason is that the model for the $\boldsymbol{\beta}_i$ terms may be misspecified. It may also be that important covariates were left out of the model, which would affect the estimates of $\tilde{\boldsymbol{\alpha}}_n$ and $\tilde{\boldsymbol{\Sigma}}_n$, and thus the $\tilde{\beta}_i$ estimates. One area of future research in this area should concern developing fit diagnostics for these estimates and nonparametric techniques that are not affected by the estimates of $\tilde{\boldsymbol{\alpha}}_n$ and $\tilde{\boldsymbol{\Sigma}}_n$.

While these results are encouraging, it is clear that more research must be done in this area on model checking; the inappropriate inclusion of Restlessness as a random effect greatly affected the variances of several terms, including the incorrectly-modeled one, and would have led to inappropriate conclusions regarding which covariates affect smoking timing.

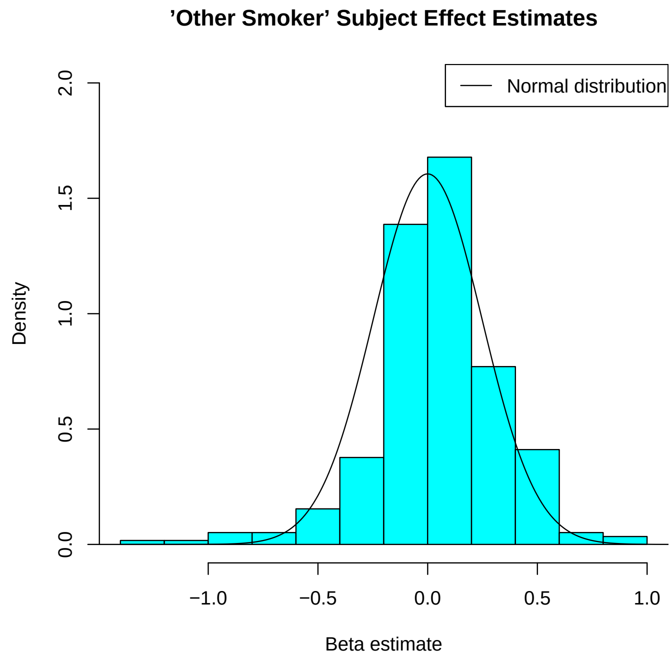

Figure 4.1: Estimates for the Subject-Level Effects of "Other Smokers"

Appendix A

Crowder (1986) proved that if the estimating equations $\mathbf{g}_n(\boldsymbol{\theta})$ satisfied the following three conditions, then the solution $\hat{\boldsymbol{\theta}}_n$ to $\mathbf{g}_n(\boldsymbol{\theta}) = 0$ is a consistent estimator of the true value $\boldsymbol{\theta}_\mathbf{0}$:

Condition 4. $\mathbf{g}_n(\boldsymbol{\theta})$ *are continuous*

Condition 5. *For n sufficiently large and some $\delta > 0$,*
$\inf_{\partial S(\boldsymbol{\theta}_\mathbf{0},\,\epsilon)} (\boldsymbol{\theta}_\mathbf{0} - \boldsymbol{\theta})^T E_{\boldsymbol{\theta}_\mathbf{0}}\{\mathbf{g}_n(\boldsymbol{\theta})\} \geq \delta$

Condition 6. $\sup_{\partial S(\boldsymbol{\theta}_\mathbf{0},\,\epsilon)} \|\mathbf{g}_n(\boldsymbol{\theta}) - E_{\boldsymbol{\theta}_\mathbf{0}}\{\mathbf{g}_n(\boldsymbol{\theta})\}\| \xrightarrow{p} 0$ *as* $n \to \infty$.

Proof of Theorem 5. $\tilde{\boldsymbol{\Psi}}_n(\boldsymbol{\theta})$ clearly satisfies Condition 4.

For Condition 5, first recall (see e.g. Sándor 1996) that $\inf_x \sum_{i=1}^n f_i(x) \geq \sum_{i=1}^n \inf_x f_i(x)$. Then

$$\begin{aligned} \inf_{\partial S(\boldsymbol{\theta_0},\,\epsilon)} (\boldsymbol{\theta_0} - \boldsymbol{\theta})^T E_{\boldsymbol{\theta_0}}\{\tilde{\boldsymbol{\Psi}}_n(\boldsymbol{\theta})\} &\geq \inf_{\partial S(\boldsymbol{\theta_0},\,\epsilon)} (\boldsymbol{\alpha_0} - \boldsymbol{\alpha})^T E_{\boldsymbol{\theta_0}}\{\tilde{\mathbf{A}}_n(\boldsymbol{\theta})\} \\ &+ \inf_{\partial S(\boldsymbol{\theta_0},\,\epsilon)} (\boldsymbol{\Sigma_0}_{vec} - \mathbf{v}) E_{\boldsymbol{\theta_0}}\{\tilde{\mathbf{B}}_n(\boldsymbol{\theta})\}, \end{aligned} \tag{A.1}$$

so it is sufficient to prove that the condition holds for $\tilde{\mathbf{A}}_n(\boldsymbol{\theta})$ and $\tilde{\mathbf{B}}_n(\boldsymbol{\theta})$ individually. Define

$$\mathbf{A}_n^*(\boldsymbol{\theta}) = n^{-1} \sum_{i=1}^n \left[\sum_{t \in X_{ni}} f(t;\boldsymbol{\theta}) \mathbf{x}_i(t) - \int_{A_{ni}} f(t;\boldsymbol{\theta}) \phi_i(t;\boldsymbol{\theta}) \right]$$

and

$$\mathbf{B}_n^*(\boldsymbol{\theta}) = n^{-1} \sum_{i=1}^n \left[\sum_{s \neq t \in X_{ni}} g(s,t;\boldsymbol{\theta}) \mathbf{y}_i(s,t) - \int_{A_{ni}} \int_{A_{ni}} g(s,t;\boldsymbol{\theta}) \omega_i(s,t;\boldsymbol{\theta})\, ds\, dt \right]$$

as weighted versions of $\mathbf{A}_n(\boldsymbol{\theta})$ and $\mathbf{B}_n(\boldsymbol{\theta})$. Note that $E_{\boldsymbol{\theta_0}}\{\tilde{\mathbf{A}}_n(\boldsymbol{\theta})\} = E_{\boldsymbol{\theta_0}}\{\mathbf{A}_n^*(\boldsymbol{\theta})\}$ and likewise $E_{\boldsymbol{\theta_0}}\{\tilde{\mathbf{B}}_n(\boldsymbol{\theta})\} = E_{\boldsymbol{\theta_0}}\{\mathbf{B}_n^*(\boldsymbol{\theta})\}$. We can show that for both Rathbun et al. and Waagepetersen's choices of $f(t;\boldsymbol{\theta})$ and $g(s,t;\boldsymbol{\theta})$,

$$\frac{\partial}{\partial \boldsymbol{\alpha}} f(t;\boldsymbol{\theta}) = -\mathbf{x}_i(t)^T f(t;\boldsymbol{\theta})(1 - f(t;\boldsymbol{\theta})) \tag{A.2}$$

and

$$\frac{\partial}{\partial \mathbf{v}} g(s,t;\boldsymbol{\theta}) = -\frac{1}{2} \mathbf{d}_i(s,t)^T g(s,t;\boldsymbol{\theta})(1 - g(s,t;\boldsymbol{\theta})), \tag{A.3}$$

where $\mathbf{d}_i(s,t)$ is as defined in (B.1). Finally, note that since $E_{\boldsymbol{\theta_0}}(\mathbf{A}_n^*(\boldsymbol{\theta_0})) = 0$, we can take a Taylor Expansion of $\mathbf{A}_n^*(\boldsymbol{\theta})$ around $\boldsymbol{\theta_0}$ to obtain

$$\begin{aligned} & (\boldsymbol{\alpha_0} - \boldsymbol{\alpha})^T E_{\boldsymbol{\theta_0}}\{\mathbf{A}_n^*(\boldsymbol{\theta})\} \\ = \; & (\boldsymbol{\alpha_0} - \boldsymbol{\alpha})^T \left\{ \frac{\partial}{\partial \boldsymbol{\alpha}^T} E_{\boldsymbol{\theta_0}}\{\mathbf{A}_n^*(\boldsymbol{\theta_0})\} \right\} (\boldsymbol{\alpha} - \boldsymbol{\alpha_0}) + o \left\| \boldsymbol{\alpha} - \boldsymbol{\alpha_0} \right\|^2 \\ = \; & (\boldsymbol{\alpha_0} - \boldsymbol{\alpha})^T \left[n^{-1} \sum_{i=1}^n \frac{\partial}{\partial \boldsymbol{\alpha}} \left\{ \int_{A_{ni}} f(t;\boldsymbol{\theta}) \{\phi_i(t;\boldsymbol{\theta_0}) - \phi_i(t;\boldsymbol{\theta})\} dt \right\} \right] (\boldsymbol{\alpha} - \boldsymbol{\alpha_0}) + o \left\| \boldsymbol{\alpha} - \boldsymbol{\alpha_0} \right\|^2 \end{aligned}$$

$$\begin{aligned}
&= (\boldsymbol{\alpha_0} - \boldsymbol{\alpha})^T \Bigg[n^{-1} \sum_{i=1}^{n} \Bigg\{ \int_{A_{ni}} f(t;\boldsymbol{\theta})(1 - f(t;\boldsymbol{\theta}))\mathbf{x}_i(t)\{\phi_i(t;\boldsymbol{\theta_0}) - \phi_i(t;\boldsymbol{\theta})\}^T \\
&+ f(t;\boldsymbol{\theta})\mathbf{x}_i(t)\phi_i(t;\boldsymbol{\theta})^T\, dt \Bigg\} \Bigg] (\boldsymbol{\alpha_0} - \boldsymbol{\alpha}) + o\, \|\boldsymbol{\alpha} - \boldsymbol{\alpha_0}\|^2 \\
&= (\boldsymbol{\alpha_0} - \boldsymbol{\alpha})^T \Bigg[n^{-1} \sum_{i=1}^{n} \int_{A_{ni}} f(t;\boldsymbol{\theta})(1 - f(t;\boldsymbol{\theta}))\mathbf{x}_i(t)\phi_i(t;\boldsymbol{\theta_0})^T\, dt \\
&+ n^{-1} \sum_{i=1}^{n} \int_{A_{ni}} f(t;\boldsymbol{\theta})^2 \mathbf{x}_i(t)\phi_i(t;\boldsymbol{\theta})^T\, dt \Bigg] (\boldsymbol{\alpha_0} - \boldsymbol{\alpha}) + o\, \|\boldsymbol{\alpha} - \boldsymbol{\alpha_0}\|^2 . \qquad \text{(A.4)}
\end{aligned}$$

Under Assumptions 14 and 16, the first sum on the right-hand side of (A.4) is non-negative definite, and the second sum is positive definite. Then the first term on the right-hand side of (A.4) is bounded below by some constant $c > 0$ for n sufficiently large, and so satisfies Condition 5. Then, since $E_{\boldsymbol{\theta_0}}\{\tilde{\mathbf{A}}_n(\boldsymbol{\theta})\} = E_{\boldsymbol{\theta_0}}\{\mathbf{A}_n^*(\boldsymbol{\theta})\}$, we can replace the expectation on the left-hand side of (A.4) with $E_{\boldsymbol{\theta_0}}\{\tilde{\mathbf{A}}_n(\boldsymbol{\theta})\}$, giving that the first term of (A.1) is bounded below by a positive constant. We can similarly find that

$$\begin{aligned}
&(\mathbf{v_0} - \mathbf{v})^T E_{\boldsymbol{\theta_0}}\{\tilde{\mathbf{B}}_n(\boldsymbol{\theta})\} = (\mathbf{v_0} - \mathbf{v})^T \Bigg\{ \frac{\partial}{\partial \mathbf{v}^T} E_{\boldsymbol{\theta_0}}\{\mathbf{B}_n^*(\boldsymbol{\theta_0})\} \Bigg\} (\mathbf{v} - \mathbf{v_0}) \\
&+ o\, \|\mathbf{v} - \mathbf{v_0}\|^2 \\
&= (\mathbf{v_0} - \mathbf{v})^T \Bigg[(2n)^{-1} \sum_{i=1}^{n} \int_{A_{ni}} \int_{A_{ni}} g_i(s,t;\boldsymbol{\theta})(1 - g_i(s,t;\boldsymbol{\theta}))\omega_i(s,t;\boldsymbol{\theta_0})\mathbf{d}_i(s,t)^T \\
&+ (2n)^{-1} \sum_{i=1}^{n} \int_{A_{ni}} \int_{A_{ni}} g_i(s,t;\boldsymbol{\theta})^2 \omega_i(s,t;\boldsymbol{\theta})\mathbf{d}_i(s,t)^T\, ds\, dt \Bigg] (\mathbf{v_0} - \mathbf{v}) \\
&+ o\, \|\mathbf{v} - \mathbf{v_0}\|^2 . \qquad \text{(A.5)}
\end{aligned}$$

Note that one can write (Kolman 1970, p. 20)

$$\begin{aligned}
&\omega_i(s,t;\boldsymbol{\theta})\mathbf{d}_i(s,t)^T = \mathbf{S} + \mathbf{K} \\
&= \frac{1}{2}\lambda_i^*(s;\boldsymbol{\alpha})\lambda_i^*(t;\boldsymbol{\alpha}) M_{\boldsymbol{\beta}_i}(\mathbf{z}_i(s) + \mathbf{z}_i(t))[\mathbf{y}_i(s,t)\mathbf{d}_i(s,t)^T + \mathbf{d}_i(s,t)\mathbf{y}_i(s,t)^T] \\
&+ \frac{1}{2}\lambda_i^*(s;\boldsymbol{\alpha})\lambda_i^*(t;\boldsymbol{\alpha}) M_{\boldsymbol{\beta}_i}(\mathbf{z}_i(s) + \mathbf{z}_i(t))[\mathbf{y}_i(s,t)\mathbf{d}_i(s,t)^T - \mathbf{d}_i(s,t)\mathbf{y}_i(s,t)^T]. \qquad \text{(A.6)}
\end{aligned}$$

The second term of (A.6), $\mathbf{K}$ is skew-symmetric, and thus $\mathbf{x}^T\mathbf{K}\mathbf{x} = 0$ for any vector $\mathbf{x}$. Then $\omega_i(s,t;\boldsymbol{\theta})\mathbf{d}_i(s,t)^T$ iff $\mathbf{S}$ is positive definite. Under Assumptions 14 and 17, $\mathbf{S}$ will be positive definite, and so the first two terms on the right-hand side of (A.5) are bounded below by a positive constant. Thus, the second term of (A.1) also satisfies Crowder's second condition. Therefore, $\tilde{\boldsymbol{\Psi}}_n(\boldsymbol{\theta})$ satisfies Condition 5 with both Rathbun et al. and Waagepetersen's estimators.

To prove Crowder's third condition, we will first show that $\tilde{\boldsymbol{\Psi}}_n(\boldsymbol{\theta}) - E_{\boldsymbol{\theta}_\mathbf{0}}\{\tilde{\boldsymbol{\Psi}}_n(\boldsymbol{\theta})\} \to 0$ pointwise as $n \to \infty$. Then we will prove that the derivatives of this difference are bounded in probability. This pointwise convergence and boundedness of the derivatives, with Lemmas 2.2 and 3.2 of Crowder (1986), will give that $\tilde{\boldsymbol{\Psi}}_n(\boldsymbol{\theta}) - E_{\boldsymbol{\theta}_\mathbf{0}}\{\tilde{\boldsymbol{\Psi}}_n(\boldsymbol{\theta})\} \to 0$ uniformly as $n \to \infty$, and so will give Crowder's third condition and thus the theorem. Note that by the triangle inequality, for some vectors $\mathbf{x}$ and $\mathbf{y}$,

$$\sup \|(\mathbf{x}^T, \mathbf{y}^T)^T\| \leq \sup \|\mathbf{x}\| + \sup \|\mathbf{y}\|.$$

Then we can obtain

$$\begin{aligned} \sup_{\partial S(\boldsymbol{\theta}_\mathbf{0},\,\epsilon)} \|\tilde{\boldsymbol{\Psi}}_n(\boldsymbol{\theta}) - E_{\boldsymbol{\theta}_\mathbf{0}}\{\tilde{\boldsymbol{\Psi}}_n(\boldsymbol{\theta})\}\| \quad &\leq \quad \sup_{\partial S(\boldsymbol{\theta}_\mathbf{0},\,\epsilon)} \left\| \tilde{\mathbf{A}}_n(\boldsymbol{\theta}) - E_{\boldsymbol{\theta}_\mathbf{0}}\{\tilde{\mathbf{A}}_n(\boldsymbol{\theta})\} \right\| \\ &+ \quad \sup_{\partial S(\boldsymbol{\theta}_\mathbf{0},\,\epsilon)} \left\| \tilde{\mathbf{B}}_n(\boldsymbol{\theta}) - E_{\boldsymbol{\theta}_\mathbf{0}}\{\tilde{\mathbf{B}}_n(\boldsymbol{\theta})\} \right\|. \end{aligned} \tag{A.7}$$

Thus, it is sufficient to show that the two terms on the right-hand side of (A.7) satisfy Crowder's third condition separately. Without respect to the norms, both of the terms on the right-hand side of (A.7) have expected value $\mathbf{0}$ and respective variances $n^{-1}\mathbf{S}_n^X(\boldsymbol{\theta})$ and $n^{-1}\mathbf{S}_n^\Lambda(\boldsymbol{\theta})$; expressions for $\mathbf{S}_n^X(\boldsymbol{\theta})$ and $\mathbf{S}_n^\Lambda(\boldsymbol{\theta})$ may be found in Appendix C. Since the limits of $\mathbf{S}_n^X(\boldsymbol{\theta}_\mathbf{0})$ and $\mathbf{S}_n^\Lambda(\boldsymbol{\theta}_\mathbf{0})$ exist as $n \to \infty$, and these terms are continuous $\boldsymbol{\theta}$, both terms on the right-hand side of (A.7) converge pointwise to zero as $n \to \infty$. To demonstrate uniform convergence, it suffices to demonstrate that the partial derivatives of $\tilde{\mathbf{A}}_n(\boldsymbol{\theta}) - E_{\boldsymbol{\theta}_\mathbf{0}}\{\tilde{\mathbf{A}}_n(\boldsymbol{\theta})\}$ and $\tilde{\mathbf{B}}_n(\boldsymbol{\theta}) - E_{\boldsymbol{\theta}_\mathbf{0}}\{\tilde{\mathbf{B}}_n(\boldsymbol{\theta})\}$ are bounded in probability; see Lemmas 2.2 and 3.2 of Crowder (1986). For the first term of the right-hand side, we note

$$\begin{aligned} \tilde{\mathbf{A}}_n(\boldsymbol{\theta}) - E_{\boldsymbol{\theta}_\mathbf{0}}(\tilde{\mathbf{A}}_n(\boldsymbol{\theta})) \quad &= \quad n^{-1}\sum_{i=1}^{n}\Bigg[\sum_{t\in X_{ni}} f(t;\boldsymbol{\theta})\mathbf{x}_i(t) - \sum_{u\in D_{ni}} \frac{f(u;\boldsymbol{\theta})\phi_i(u;\boldsymbol{\theta})}{\pi_{ni}(u)} \\ &- \quad \left(\int_{A_{ni}} f(t;\boldsymbol{\theta})\{\phi_i(t;\boldsymbol{\theta}_\mathbf{0}) - \phi_i(t;\boldsymbol{\theta})\}\,dt\right)\Bigg], \end{aligned}$$

which has derivative

$$\begin{aligned}
&\frac{\partial}{\partial \boldsymbol{\alpha}^T}\Big(\tilde{\mathbf{A}}_n(\boldsymbol{\theta}) - E_{\boldsymbol{\theta}_0}(\tilde{\mathbf{A}}_n(\boldsymbol{\theta}))\Big)\\
= \;& n^{-1}\sum_{i=1}^{n}\Bigg[\sum_{t\in X_{ni}} f^{(1)}_{\boldsymbol{\alpha}}(t;\boldsymbol{\theta})\mathbf{x}_i(t)^T - \sum_{u\in D_{ni}} \frac{\{f^{(1)}_{\boldsymbol{\alpha}}(u;\boldsymbol{\theta}) + f(u;\boldsymbol{\theta})\mathbf{x}_i(u)\}\phi_i(u;\boldsymbol{\theta})^T}{\pi_{ni}(u)}\Bigg]\\
- \;& \Bigg(\int_{A_{ni}} f^{(1)}_{\boldsymbol{\alpha}}(t;\boldsymbol{\theta})\phi_i(t;\boldsymbol{\theta_0})^T - \{f^{(1)}_{\boldsymbol{\alpha}}(t;\boldsymbol{\theta}) + f(t;\boldsymbol{\theta})\mathbf{x}_i(t)\}\phi_i(t;\boldsymbol{\theta})^T\,dt\Bigg). \qquad \text{(A.8)}
\end{aligned}$$

The form of $f^{(1)}_{\boldsymbol{\alpha}}(t;\boldsymbol{\theta})$ was given in (A.2). Under Assumptions 14 and 19 and $|A_{ni}|$ bounded, $f(t;\boldsymbol{\theta})$, $f^{(1)}_{\boldsymbol{\alpha}}(t;\boldsymbol{\theta})$, and $\phi_i(t;\boldsymbol{\theta})$ are bounded for all t in $|A_{ni}|$, which implies that (A.8) is bounded in probability, and thus the first term of (A.7) converges to 0 in probability as $n\to\infty$. The second term in (A.7) equals

$$\begin{aligned}
&\tilde{\mathbf{B}}_n(\boldsymbol{\theta}) - E_{\boldsymbol{\theta}_0}(\tilde{\mathbf{B}}_n(\boldsymbol{\theta})) = n^{-1}\sum_{i=1}^{n}\Bigg[\sum_{s\neq t\in X_{ni}} g(s,t;\boldsymbol{\theta})\mathbf{y}_i(s,t)\\
- \;& \sum_{u\neq v\in D_{ni}} \frac{g(u,v;\boldsymbol{\theta})\omega_i(u,v;\boldsymbol{\theta})}{\pi_{ni}(u,v)} - \Bigg(\int_{A_{ni}} g(s,t;\boldsymbol{\theta})\{\omega_i(s,t;\boldsymbol{\theta_0}) - \omega_i(s,t;\boldsymbol{\theta})\}\,ds\,dt\Bigg)\Bigg],
\end{aligned}$$

with derivative

$$\begin{aligned}
&\frac{\partial}{\partial \mathbf{v}^T}\Big(\tilde{\mathbf{B}}_n(\boldsymbol{\theta}) - E_{\boldsymbol{\theta}_0}(\tilde{\mathbf{B}}_n(\boldsymbol{\theta}))\Big) = n^{-1}\sum_{i=1}^{n}\Bigg[\sum_{s\neq t\in X_{ni}} g^{(1)}_{\boldsymbol{\Sigma}}(s,t;\boldsymbol{\theta})\mathbf{y}_i(s,t)^T\\
- \;& \sum_{u\neq v\in D_{ni}} \{g^{(1)}_{\boldsymbol{\Sigma}}(u,v;\boldsymbol{\theta}) + g(u,v;\boldsymbol{\theta})\mathbf{d}_i(u,v)\}\frac{\omega_i(u,v;\boldsymbol{\theta})^T}{\pi_{ni}(u,v)}\Bigg]\\
- \;& \Bigg(\int_{A_{ni}} g^{(1)}_{\boldsymbol{\Sigma}}(s,t;\boldsymbol{\theta})\omega_i(s,t;\boldsymbol{\theta_0})^T - \{g^{(1)}_{\boldsymbol{\Sigma}}(s,t;\boldsymbol{\theta}) + g(s,t;\boldsymbol{\theta})\mathbf{d}_i(s,t)\}\omega_i(s,t;\boldsymbol{\theta})\,ds\,dt\Bigg),
\end{aligned} \qquad \text{(A.9)}$$

where $g^{(1)}_{\boldsymbol{\Sigma}}(s,t;\boldsymbol{\theta}) = (\partial/\partial\mathbf{v}^T)g(s,t;\boldsymbol{\theta})$. Under Assumptions 14 and 20 and $|A_{ni}|$ bounded, $g(s,t;\boldsymbol{\theta})$, $g^{(1)}_{\boldsymbol{\Sigma}}(s,t;\boldsymbol{\theta})$, and $\omega_i(s,t;\boldsymbol{\theta})$ are bounded and so (A.9) is bounded in probability. Thus the second term of (A.7) converges in 0 to probability as $n\to\infty$, giving Crowder's third condition and, thus, the theorem. □

Appendix B

The following Lemma is necesary to prove the theorem.

Lemma 6. *Under Assumptions 13 through 20, $n^{\frac{1}{2}}\tilde{\boldsymbol{\Psi}}_n(\boldsymbol{\theta_0})$ is asymptotically normal with mean $\mathbf{0}$ and variance $\mathbf{S}^X_\infty(\boldsymbol{\theta_0}) + \mathbf{S}^\Lambda_\infty(\boldsymbol{\theta_0})$, where $\mathbf{S}^X_\infty(\boldsymbol{\theta_0})$ and $\mathbf{S}^\Lambda_\infty(\boldsymbol{\theta_0})$ are as defined in (B.3) and (B.4).*

Proof of Lemma 6. The independence of event and dummy assessments, given by Assumption 13, and the form of $\tilde{\boldsymbol{\Psi}}_n(\boldsymbol{\theta})$ given in (B.19) imply that it is sufficient to show that $n^{-\frac{1}{2}}\sum_{i=1}^n \tilde{\mathbf{X}}_{ni}(\boldsymbol{\theta_0})$ and $n^{-\frac{1}{2}}\sum_{i=1}^n \tilde{\boldsymbol{\Lambda}}_{ni}(\boldsymbol{\theta_0})$ are asymptotically normal with mean $\mathbf{0}$ and respective variances $\mathbf{S}^X_\infty(\boldsymbol{\theta_0})$ and $\mathbf{S}^\Lambda_\infty(\boldsymbol{\theta_0})$. It can be shown that each of these sums satisfies the conditions for Lyapunov's CLT under Assumptions 13 through 17 and 19. The demonstration of this fact is lengthy and unenlightening and so is omitted here; Guan and Loh's (2007) proof of their Theorem 1 contains a similar example. Note that Guan and Loh partition the study space in order to create independent sampled sections for the independence requirement of the CLT. Such partitioning is unnecessary in our case due to Assumption 13. It has already been shown in Section 4.3 that the variances of these sums are $\mathbf{S}^X_\infty(\boldsymbol{\theta_0})$ and $\mathbf{S}^\Lambda_\infty(\boldsymbol{\theta_0})$. Then by the Lyapunov CLT and (B.19), $n^{\frac{1}{2}}\tilde{\boldsymbol{\Psi}}_n(\boldsymbol{\theta_0})$ is asymptotically $\mathbf{N}(\mathbf{0}, \mathbf{S}^X_\infty(\boldsymbol{\theta_0}) + \mathbf{S}^\Lambda_\infty(\boldsymbol{\theta_0}))$. $\square$

Proof of theorem 6. By taking a Taylor Expansion, we can write

$$\tilde{\boldsymbol{\Psi}}_n(\tilde{\boldsymbol{\theta}}_n) = \tilde{\boldsymbol{\Psi}}_n(\boldsymbol{\theta_0}) + \tilde{\mathbf{J}}_n(\boldsymbol{\theta_0})(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0}) + \tilde{\mathbf{K}}_n(\boldsymbol{\theta}^*)(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0}),$$

where

$$\tilde{\mathbf{J}}_n(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}^T}\tilde{\boldsymbol{\Psi}}_n(\boldsymbol{\theta}),$$

$$\begin{aligned}\tilde{\mathbf{K}}_n(\boldsymbol{\theta}) &= \frac{1}{2n}\sum_{i=1}^n\Bigg\{\sum_{l=1}^p(\tilde{\alpha}_{n,l} - \alpha_{0,l})\left(\frac{\partial^2}{\partial\alpha_l\partial\boldsymbol{\theta}^T}\tilde{\boldsymbol{\Psi}}_n(\boldsymbol{\theta})\right)^T \\ &+ \sum_{l=1}^{q(q+1)/2}(\tilde{v}_{n,l} - v_{0,l})\left(\frac{\partial^2}{\partial v_l\partial\boldsymbol{\theta}^T}\tilde{\boldsymbol{\Psi}}_n(\boldsymbol{\theta})\right)^T\Bigg\},\end{aligned}$$

and $\boldsymbol{\theta}^* = c\tilde{\boldsymbol{\theta}}_n + (1-c)\boldsymbol{\theta_0}$ for some real $c \in (0,1)$. Since $\tilde{\boldsymbol{\Psi}}_n(\tilde{\boldsymbol{\theta}}_n) = 0$ by definition, we can rearrange this equation to obtain

$$n^{\frac{1}{2}}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0}) = -n^{\frac{1}{2}}\{\tilde{\mathbf{J}}_n(\boldsymbol{\theta_0}) + \tilde{\mathbf{K}}_n(\boldsymbol{\theta}^*)\}^{-1}\tilde{\boldsymbol{\Psi}}_n(\boldsymbol{\theta_0}).$$

Under Assumptions 13 through 17 and 19, all of the elements of the partial derivatives in $\tilde{\mathbf{K}}_n(\boldsymbol{\theta}^*)$ are bounded by some finite constant for all l and n. Furthermore, by Theorem 5, under Assumptions 13 through 19, $\tilde{\boldsymbol{\theta}}_n \xrightarrow{p} \boldsymbol{\theta_0}$ as $n \to \infty$. Thus as $n \to \infty$,

$$\tilde{\mathbf{K}}_n(\boldsymbol{\theta}^*) \xrightarrow{p} \mathbf{0}.$$

Define

$$\mathbf{J}_n(\boldsymbol{\theta_0}) = E_{\boldsymbol{\theta_0}}\left\{\frac{\partial}{\partial \boldsymbol{\theta}^T}\tilde{\boldsymbol{\Psi}}_n(\boldsymbol{\theta_0})\right\}$$

and $\mathbf{J}_\infty(\boldsymbol{\theta_0}) = \lim_{n\to\infty} \mathbf{J}_n(\boldsymbol{\theta_0})$. It can be shown that, under Assumptions 14, 19, and 20,

$$\tilde{\mathbf{J}}_n(\boldsymbol{\theta_0}) - \mathbf{J}_\infty(\boldsymbol{\theta_0}) \xrightarrow{p} \mathbf{0},$$

and thus

$$\tilde{\mathbf{J}}_n(\boldsymbol{\theta_0}) + \tilde{\mathbf{K}}_n(\boldsymbol{\theta}^*) \xrightarrow{p} \mathbf{J}_\infty(\boldsymbol{\theta_0})$$

as $n \to \infty$. It was proven in Lemma 6 that $n^{\frac{1}{2}}\tilde{\boldsymbol{\Psi}}_n(\boldsymbol{\theta_0})$ is asymptotically $\mathbf{N}(\mathbf{0}, \mathbf{S}^X_\infty(\boldsymbol{\theta_0}) + \mathbf{S}^\Lambda_\infty(\boldsymbol{\theta_0}))$ and the theorem follows. □

Appendix C

It can be shown that $\mathbf{J}_\infty(\boldsymbol{\theta_0})$, as defined in (B.2), is of the form

$$\mathbf{J}_\infty(\boldsymbol{\theta_0}) = \left(\begin{array}{c|c} J_{11,\infty}(\boldsymbol{\theta_0}) & J_{12,\infty}(\boldsymbol{\theta_0}) \\ \hline J_{21,\infty}(\boldsymbol{\theta_0}) & J_{22,\infty}(\boldsymbol{\theta_0}) \end{array}\right),$$

with

$$J_{11,\infty}(\boldsymbol{\theta_0}) = \lim_{n\to\infty} -n^{-1}\sum_{i=1}^{n}\left\{\int_{A_{ni}} f(t;\boldsymbol{\theta_0})\mathbf{x}_i(t)\phi_i(t;\boldsymbol{\theta_0})^T\,dt\right\}, \tag{C.1}$$

$$J_{12,\infty}(\boldsymbol{\theta_0}) = \lim_{n\to\infty} -(2n)^{-1}\sum_{i=1}^{n}\left\{\int_{A_{ni}} f(t;\boldsymbol{\theta_0})\mathbf{c}_i(t)\phi_i(t;\boldsymbol{\theta_0})^T\,dt\right\},$$

$$J_{21,\infty}(\boldsymbol{\theta_0}) = \lim_{n\to\infty} -n^{-1}\sum_{i=1}^{n}\left\{\int_{A_{ni}}\int_{A_{ni}} g(s,t;\boldsymbol{\theta_0})\omega_i(s,t;\boldsymbol{\theta_0})\{\mathbf{x}_i(s)+\mathbf{x}_i(t)\}^T\,ds\,dt\right\},$$

and

$$J_{22,\infty}(\boldsymbol{\theta_0}) = \lim_{n\to\infty} -(2n)^{-1} \sum_{i=1}^{n} \left\{ \int_{A_{ni}} \int_{A_{ni}} g(s,t;\boldsymbol{\theta_0})\omega_i(s,t;\boldsymbol{\theta_0})\mathbf{d}_i(t)^T \, ds \, dt \right\},$$

where

$$\mathbf{c}_i(t) = vech(\mathbf{C}_i(t))$$

for the matrix $\mathbf{C}_i(t)$ with elements

$$[\mathbf{C}_i(t)]_{j\,k} = (1 + I[j \neq k])\{z_{i,j}(t)z_{i,k}(t)\}; j,k = 1,\ldots,q.$$

Similarly, we can write the two variance matrices as

$$\mathbf{S}_n^X(\boldsymbol{\theta_0}) = \left(\begin{array}{c|c} S_{11,n}^X(\boldsymbol{\theta_0}) & S_{12,n}^X(\boldsymbol{\theta_0}) \\ \hline S_{12,n}^X(\boldsymbol{\theta_0})^T & S_{22,n}^X(\boldsymbol{\theta_0}) \end{array} \right)$$

and

$$\mathbf{S}_n^\Lambda(\boldsymbol{\theta_0}) = \left(\begin{array}{c|c} S_{11,n}^\Lambda(\boldsymbol{\theta_0}) & S_{12,n}^\Lambda(\boldsymbol{\theta_0}) \\ \hline S_{12,n}^\Lambda(\boldsymbol{\theta_0})^T & S_{22,n}^\Lambda(\boldsymbol{\theta_0}) \end{array} \right),$$

where

$$\begin{aligned} S_{11,n}^X(\boldsymbol{\theta_0}) &= n^{-1} \sum_{i=1}^{n} \Bigg[\int_{A_{ni}} f(t;\boldsymbol{\theta_0})^2 \mathbf{x}_i(t)\phi_i(t;\boldsymbol{\theta_0})^T \, dt \\ &+ \int_{A_{ni}} \int_{A_{ni}} f(s;\boldsymbol{\theta_0}) f(t;\boldsymbol{\theta_0})\mathbf{x}_i(s)\mathbf{x}_i(t)^T \lambda_i^*(s;\boldsymbol{\alpha_0})\lambda_i^*(t;\boldsymbol{\alpha_0}) \\ &* \{M_{\boldsymbol{\beta}_i}(\mathbf{z}_i(s) + \mathbf{z}_i(t);\boldsymbol{\Sigma_0}) - M_{\boldsymbol{\beta}_i}(\mathbf{z}_i(s);\boldsymbol{\Sigma_0})M_{\boldsymbol{\beta}_i}(\mathbf{z}_i(t);\boldsymbol{\Sigma_0})\} \, ds \, dt \Bigg], \end{aligned}$$

$$\begin{aligned} S_{12,n}^X(\boldsymbol{\theta_0}) &= n^{-1} \sum_{i=1}^{n} \Bigg[\int_{A_{ni}} \int_{A_{ni}} \{f(s;\boldsymbol{\theta_0})\mathbf{x}_i(s) + f(t;\boldsymbol{\theta_0})\mathbf{x}_i(t)\} g(s,t;\boldsymbol{\theta_0})\omega_i(s,t;\boldsymbol{\theta_0})^T \, ds \, dt \\ &+ \int_{A_{ni}} \int_{A_{ni}} \int_{A_{ni}} f(r;\boldsymbol{\theta_0}) g(s,t;\boldsymbol{\theta_0})\mathbf{x}_i(r)\mathbf{y}_i(s,t)^T \lambda_i^*(r;\boldsymbol{\alpha_0})\lambda_i^*(s;\boldsymbol{\alpha_0})\lambda_i^*(t;\boldsymbol{\alpha_0}) \\ &* \{M_{\boldsymbol{\beta}_i}(\mathbf{z}_i(r) + \mathbf{z}_i(s) + \mathbf{z}_i(t);\boldsymbol{\Sigma_0}) - M_{\boldsymbol{\beta}_i}(\mathbf{z}_i(r);\boldsymbol{\Sigma_0})M_{\boldsymbol{\beta}_i}(\mathbf{z}_i(s) + \mathbf{z}_i(t);\boldsymbol{\Sigma_0})\} \, dr \, ds \, dt \Bigg] \end{aligned}$$

$$
\begin{aligned}
S^{X}_{22,n}(\boldsymbol{\theta_0}) &= n^{-1}\sum_{i=1}^{n}\Big[\int_{A_{ni}}\int_{A_{ni}} g(s,t;\boldsymbol{\theta_0})^2\omega_i(s,t;\boldsymbol{\theta_0})\{\mathbf{y}_i(s,t)+\mathbf{y}_i(t,s)\}^T\,ds\,dt\\
&+ \int_{A_{ni}}\int_{A_{ni}}\int_{A_{ni}} g(r,s;\boldsymbol{\theta_0})\mathbf{y}_i(r,s)[g(r,t;\boldsymbol{\theta_0})\{\mathbf{y}_i(r,t)+\mathbf{y}_i(t,r)\}\\
&+ g(s,t;\boldsymbol{\theta_0})\{\mathbf{y}_i(s,t)+\mathbf{y}_i(t,s)\}]^T\lambda_i^*(r;\boldsymbol{\alpha_0})\lambda_i^*(s;\boldsymbol{\alpha_0})\lambda_i^*(t;\boldsymbol{\alpha_0})\\
&* M_{\boldsymbol{\beta}_i}(\mathbf{z}_i(r)+\mathbf{z}_i(s)+\mathbf{z}_i(t);\boldsymbol{\Sigma_0})\,dr\,ds\,dt\\
&+ \int_{A_{ni}}\int_{A_{ni}}\int_{A_{ni}}\int_{A_{ni}} g(r,s;\boldsymbol{\theta_0})g(t,u;\boldsymbol{\theta_0})\mathbf{y}_i(r,s)\mathbf{y}_i(t,u)^T\lambda_i^*(r;\boldsymbol{\alpha_0})\lambda_i^*(s;\boldsymbol{\alpha_0})\\
&* \lambda_i^*(t;\boldsymbol{\alpha_0})\lambda_i^*(u;\boldsymbol{\alpha_0})M_{\boldsymbol{\beta}_i}(\mathbf{z}_i(r)+\mathbf{z}_i(s)+\mathbf{z}_i(t)+\mathbf{z}_i(u);\boldsymbol{\Sigma_0})\,dr\,ds\,dt\,du\Big],
\end{aligned}
$$

Then

$$
\begin{aligned}
S^{\Lambda}_{11,n}(\boldsymbol{\theta_0}) &= n^{-1}\sum_{i=1}^{n}\Big[\int_{A_{ni}}\pi_{ni}(t)^{-1}f(t;\boldsymbol{\theta_0})\phi_i(t;\boldsymbol{\theta_0})\phi_i(t;\boldsymbol{\theta_0})^T\,dt\\
&+ \int_{A_{ni}}\int_{A_{ni}}\frac{\pi_{ni}(s,t)-\pi_{ni}(s)\pi_{ni}(t)}{\pi_{ni}(s)\pi_{ni}(t)}f(s;\boldsymbol{\theta_0})f(t;\boldsymbol{\theta_0})\phi_i(s;\boldsymbol{\theta_0})\phi_i(t;\boldsymbol{\theta_0})\,ds\,dt\Big]
\end{aligned}
$$

$$
\begin{aligned}
S^{\Lambda}_{12,n}(\boldsymbol{\theta_0}) &= n^{-1}\sum_{i=1}^{n}\Big[\int_{A_{ni}}\int_{A_{ni}}\{\pi_{ni}(s)^{-1}f(s;\boldsymbol{\theta_0})\phi_i(s;\boldsymbol{\theta_0})+\pi_{ni}(t)^{-1}f(t;\boldsymbol{\theta_0})\phi_i(t;\boldsymbol{\theta_0})\}\\
&* g(s,t;\boldsymbol{\theta_0})\omega_i(s,t;\boldsymbol{\theta_0})^T\,ds\,dt\\
&+ \int_{A_{ni}}\int_{A_{ni}}\int_{A_{ni}}\frac{\pi_{ni}(r,s,t)-\pi_{ni}(r)\pi_{ni}(s,t)}{\pi_{ni}(r)\pi_{ni}(s,t)}f(r;\boldsymbol{\theta_0})g(s,t;\boldsymbol{\theta_0})\\
&* \phi_i(r;\boldsymbol{\theta_0})\omega_i(s,t;\boldsymbol{\theta_0})^T\,dr\,ds\,dt\Big]
\end{aligned}
$$

and

$$
\begin{aligned}
S^{\Lambda}_{22,n}(\boldsymbol{\theta_0}) &= n^{-1}\sum_{i=1}^{n}\Big[\int_{A_{ni}}\int_{A_{ni}}\pi_{ni}(s,t)^{-1}g(s,t;\boldsymbol{\theta_0})^2\omega_i(s,t;\boldsymbol{\theta_0})\omega_i(s,t;\boldsymbol{\theta_0})^T\,ds\,dt\\
&+ \int_{A_{ni}}\int_{A_{ni}}\int_{A_{ni}}\frac{\pi_{ni}(r,s,t)}{\pi_{ni}(r,s)}g(r,s;\boldsymbol{\theta_0})\omega_i(r,s;\boldsymbol{\theta_0})\Big[\frac{\omega_i(r,t;\boldsymbol{\theta_0})^T+\omega_i(t,r;\boldsymbol{\theta_0})^T}{\pi_{ni}(r,t)}\\
&+ \frac{\omega_i(s,t;\boldsymbol{\theta_0})^T+\omega_i(t,s;\boldsymbol{\theta_0})^T}{\pi_{ni}(s,t)}\Big]\,dr\,ds\,dt\\
&+ \int_{A_{ni}}\int_{A_{ni}}\int_{A_{ni}}\int_{A_{ni}}\frac{\pi_{ni}(r,s,t,u)-\pi_{ni}(r,s)\pi_{ni}(t,u)}{\pi_{ni}(r,s)\pi_{ni}(t,u)}g(r,s;\boldsymbol{\theta_0})g(t,u;\boldsymbol{\theta_0})\\
&* \omega_i(r,s;\boldsymbol{\theta_0})\omega_i(t,u;\boldsymbol{\theta_0})\,dr\,ds\,dt\,du\Big].
\end{aligned}
$$

A more complete derivation of these variance terms can be found in Appendix D.

Appendix D

Below, find the more complete expansions of the variance terms (B.3) and (B.4).

$$
\begin{aligned}
& S^X_{11,\infty}(\boldsymbol{\theta_0}) = n^{-1}\sum_{i=1}^{n} Var_{\boldsymbol{\theta_0}}\Big\{\sum_{t\in X_{ni}} f(t;\boldsymbol{\theta_0})\mathbf{x}_i(t)\Big\} \\
= \;& n^{-1}\sum_{i=1}^{n}\Big[E_{\boldsymbol{\theta_0}}\Big\{\sum_{t\in X_{ni}} f(t;\boldsymbol{\theta_0})^2\mathbf{x}_i(t)\mathbf{x}_i(t)^T + \sum_{s\neq t\in X_{ni}} f(s;\boldsymbol{\theta_0})f(t;\boldsymbol{\theta_0})\mathbf{x}_i(s)\mathbf{x}_i(t)^T\Big\} \\
- \;& E_{\boldsymbol{\theta_0}}\Big\{\sum_{t\in X_{ni}} f(t;\boldsymbol{\theta_0})\mathbf{x}_i(t)\Big\}E_{\boldsymbol{\theta_0}}\Big\{\sum_{t\in X_{ni}} f(t;\boldsymbol{\theta_0})\mathbf{x}_i(t)\Big\}^T\Big] \\
= \;& n^{-1}\sum_{i=1}^{n}\Big[\int_{A_{ni}} f(t;\boldsymbol{\theta_0})^2\mathbf{x}_i(t)\phi_i(t;\boldsymbol{\theta_0})^T\,dt \\
+ \;& \int_{A_{ni}}\int_{A_{ni}} f(s;\boldsymbol{\theta_0})f(t;\boldsymbol{\theta_0})\mathbf{x}_i(s)\mathbf{x}_i(t)^T\lambda_i(s;\boldsymbol{\alpha_0})\lambda_i(t;\boldsymbol{\alpha_0})M_{\boldsymbol{\beta}_i}(\mathbf{z}_i(s)+\mathbf{z}_i(t);\boldsymbol{\Sigma_0})\,ds\,dt \\
- \;& \int_{A_{ni}}\int_{A_{ni}} f(s;\boldsymbol{\theta_0})f(t;\boldsymbol{\theta_0})\phi_i(s;\boldsymbol{\theta_0})\phi_i(t;\boldsymbol{\theta_0})^T\,ds\,dt\Big] \\
= \;& n^{-1}\sum_{i=1}^{n}\Big[\int_{A_{ni}} f(t;\boldsymbol{\theta_0})^2\mathbf{x}_i(t)\phi_i(t;\boldsymbol{\theta_0})^T\,dt \\
+ \;& \int_{A_{ni}}\int_{A_{ni}} f(s;\boldsymbol{\theta_0})f(t;\boldsymbol{\theta_0})\mathbf{x}_i(s)\mathbf{x}_i(t)^T\lambda_i(s;\boldsymbol{\alpha_0})\lambda_i(t;\boldsymbol{\alpha_0}) \\
* \;& \{M_{\boldsymbol{\beta}_i}(\mathbf{z}_i(s)+\mathbf{z}_i(t);\boldsymbol{\Sigma_0}) - M_{\boldsymbol{\beta}_i}(\mathbf{z}_i(s);\boldsymbol{\Sigma_0})M_{\boldsymbol{\beta}_i}(\mathbf{z}_i(t);\boldsymbol{\Sigma_0})\}\,ds\,dt\Big]
\end{aligned}
$$

$$
\begin{aligned}
& S^X_{12,\infty}(\boldsymbol{\theta_0}) = n^{-1}\sum_{i=1}^{n} E_{\boldsymbol{\theta_0}}\Big[\Big\{\sum_{t\in X_{ni}} f(t;\boldsymbol{\theta_0})\mathbf{x}_i(t) - \int_{A_{ni}} f(t;\boldsymbol{\theta_0})\phi_i(t;\boldsymbol{\theta_0})\,dt\Big\} \\
* \;& \Big\{\sum_{s\neq t\in X_{ni}} g(s,t;\boldsymbol{\theta_0})\mathbf{y}_i(s,t) - \int_{A_{ni}}\int_{A_{ni}} g(s,t;\boldsymbol{\theta_0})\omega_i(s,t;\boldsymbol{\theta_0})\,ds\,dt\Big\}^T\Big] \\
= \;& n^{-1}\sum_{i=1}^{n}\Big[E_{\boldsymbol{\theta_0}}\Big\{\sum_{s\neq t\in X_{ni}} (f(s;\boldsymbol{\theta_0})\mathbf{x}_i(s) + f(t;\boldsymbol{\theta_0})\mathbf{x}_i(t))g(s,t;\boldsymbol{\theta_0})\mathbf{y}_i(s,t)^T \\
+ \;& \sum_{r\neq s\neq t\in X_{ni}} f(r;\boldsymbol{\theta_0})g(s,t;\boldsymbol{\theta_0})\mathbf{x}_i(r)\mathbf{y}_i(s,t)^T\Big\}
\end{aligned}
$$

$$
\begin{aligned}
- & \int_{A_{ni}}\int_{A_{ni}}\int_{A_{ni}} f(r;\boldsymbol{\theta_0})g(s,t;\boldsymbol{\theta_0})\phi_i(r;\boldsymbol{\theta_0})\omega_i(s,t;\boldsymbol{\theta_0})^T\,dr\,ds\,dt\Big] \\
= & \; n^{-1}\sum_{i=1}^{n}\Big[\int_{A_{ni}}\int_{A_{ni}}\{f(s;\boldsymbol{\theta_0})\mathbf{x}_i(s)+f(t;\boldsymbol{\theta_0})\mathbf{x}_i(t)\}g(s,t;\boldsymbol{\theta_0})\omega_i(s,t;\boldsymbol{\theta_0})^T\,ds\,dt \\
+ & \int_{A_{ni}}\int_{A_{ni}}\int_{A_{ni}} f(r;\boldsymbol{\theta_0})g(s,t;\boldsymbol{\theta_0})\mathbf{x}_i(r)\mathbf{y}_i(s,t)^T\lambda_i(r;\boldsymbol{\alpha_0})\lambda_i(s;\boldsymbol{\alpha_0})\lambda_i(t;\boldsymbol{\alpha_0}) \\
* & \;\{M_{\boldsymbol{\beta}_i}(\mathbf{z}_i(r)+\mathbf{z}_i(s)+\mathbf{z}_i(t);\boldsymbol{\Sigma_0})-M_{\boldsymbol{\beta}_i}(\mathbf{z}_i(r);\boldsymbol{\Sigma_0})M_{\boldsymbol{\beta}_i}(\mathbf{z}_i(s)+\mathbf{z}_i(t);\boldsymbol{\Sigma_0})\}\,dr\,ds\,dt\Big]
\end{aligned}
$$

$$
\begin{aligned}
& S^X_{22,\infty}(\boldsymbol{\theta_0}) = n^{-1}\sum_{i=1}^{n} Var_{\boldsymbol{\theta_0}}\Big\{\sum_{s\neq t\in X_{ni}} g(s,t;\boldsymbol{\theta_0})\mathbf{y}_i(s,t)\Big\} \\
= & \; n^{-1}\sum_{i=1}^{n}\Big[E_{\boldsymbol{\theta_0}}\Big\{\sum_{s\neq t\in X_{ni}} g(s,t;\boldsymbol{\theta_0})^2\mathbf{y}_i(s,t)\mathbf{y}_i(s,t)^T+\sum_{s\neq t\in X_{ni}} g(s,t;\boldsymbol{\theta_0})^2\mathbf{y}_i(s,t)\mathbf{y}_i(t,s)^T \\
+ & \sum_{r\neq s\neq t\in X_{ni}} g(r,s;\boldsymbol{\theta_0})g(r,t;\boldsymbol{\theta_0})\mathbf{y}_i(r,s)\mathbf{y}_i(r,t)^T \\
+ & \sum_{r\neq s\neq t\in X_{ni}} g(r,s;\boldsymbol{\theta_0})g(r,t;\boldsymbol{\theta_0})\mathbf{y}_i(r,s)\mathbf{y}_i(t,r)^T \\
+ & \sum_{r\neq s\neq t\in X_{ni}} g(r,s;\boldsymbol{\theta_0})g(s,t;\boldsymbol{\theta_0})\mathbf{y}_i(r,s)\mathbf{y}_i(s,t)^T \\
+ & \sum_{r\neq s\neq t\in X_{ni}} g(r,s;\boldsymbol{\theta_0})g(s,t;\boldsymbol{\theta_0})\mathbf{y}_i(r,s)\mathbf{y}_i(t,s)^T \\
+ & \sum_{r\neq s\neq t\neq u\in X_{ni}} g(r,s;\boldsymbol{\theta_0})g(t,u;\boldsymbol{\theta_0})\mathbf{y}_i(r,s)\mathbf{y}_i(t,u)^T\Big\} \\
- & \; E_{\boldsymbol{\theta_0}}\Big\{\sum_{s\neq t\in X_{ni}} g(s,t;\boldsymbol{\theta_0})\mathbf{y}_i(s,t)\Big\}E_{\boldsymbol{\theta_0}}\Big\{\sum_{s\neq t\in X_{ni}} g(s,t;\boldsymbol{\theta_0})\mathbf{y}_i(s,t)\Big\}^T\Big] \\
= & \; n^{-1}\sum_{i=1}^{n}\Big[\int_{A_{ni}}\int_{A_{ni}} g(s,t;\boldsymbol{\theta_0})^2\omega_i(s,t;\boldsymbol{\theta_0})\{\mathbf{y}_i(s,t)+\mathbf{y}_i(t,s)\}^T\,ds\,dt \\
+ & \int_{A_{ni}}\int_{A_{ni}}\int_{A_{ni}} g(r,s;\boldsymbol{\theta_0})\mathbf{y}_i(r,s)[g(r,t;\boldsymbol{\theta_0})\{\mathbf{y}_i(r,t)+\mathbf{y}_i(t,r)\}^T \\
+ & \; g(s,t;\boldsymbol{\theta_0})\{\mathbf{y}_i(s,t)+\mathbf{y}_i(t,s)\}^T]\lambda_i(r;\boldsymbol{\alpha_0})\lambda_i(s;\boldsymbol{\alpha_0})\lambda_i(t;\boldsymbol{\alpha_0}) \\
* & \; M_{\boldsymbol{\beta}_i}(\mathbf{z}_i(r)+\mathbf{z}_i(s)+\mathbf{z}_i(t);\boldsymbol{\Sigma_0})\,dr\,ds\,dt \\
+ & \int_{A_{ni}}\int_{A_{ni}}\int_{A_{ni}}\int_{A_{ni}} g(r,s;\boldsymbol{\theta_0})g(t,u;\boldsymbol{\theta_0})\mathbf{y}_i(r,s)\mathbf{y}_i(t,u)^T\lambda_i(r;\boldsymbol{\alpha_0})\lambda_i(s;\boldsymbol{\alpha_0})
\end{aligned}
$$

$$
\begin{aligned}
* \quad & \lambda_i(t;\boldsymbol{\alpha_0})\lambda_i(u;\boldsymbol{\alpha_0})\{M_{\boldsymbol{\beta}_i}(\mathbf{z}_i(r)+\mathbf{z}_i(s)+\mathbf{z}_i(t)+\mathbf{z}_i(u);\boldsymbol{\Sigma_0}) \\
- \quad & M_{\boldsymbol{\beta}_i}(\mathbf{z}_i(r)+\mathbf{z}_i(s);\boldsymbol{\Sigma_0})M_{\boldsymbol{\beta}_i}(\mathbf{z}_i(t)+\mathbf{z}_i(u);\boldsymbol{\Sigma_0})\}\, dr\, ds\, dt\, du\Big],
\end{aligned}
$$

$$
\begin{aligned}
& S^{\Lambda}_{11,\infty}(\boldsymbol{\theta_0}) = n^{-1}\sum_{i=1}^{n} Var_{\boldsymbol{\theta_0}}\Big\{\sum_{u\in D_{ni}} \pi_{ni}(u)^{-1} f(u;\boldsymbol{\theta_0})\phi_i(u;\boldsymbol{\theta_0})\Big\} \\
= \quad & n^{-1}\sum_{i=1}^{n}\Big[E_{\boldsymbol{\theta_0}}\Big\{\sum_{u\in D_{ni}} f(u;\boldsymbol{\theta_0})^2\phi_i(u;\boldsymbol{\theta_0})\phi_i(u;\boldsymbol{\theta_0})^T \\
+ \quad & \sum_{u\neq v\in D_{ni}} (\pi_{ni}(u)\pi_{ni}(v))^{-1} f(u;\boldsymbol{\theta_0})f(v;\boldsymbol{\theta_0})\phi_i(u;\boldsymbol{\theta_0})\phi_i(v;\boldsymbol{\theta_0})^T\Big\} \\
- \quad & E_{\boldsymbol{\theta_0}}\Big\{\sum_{u\in D_{ni}} \pi_{ni}(u)^{-1} f(u;\boldsymbol{\theta_0})\phi_i(u;\boldsymbol{\theta_0})\Big\} E_{\boldsymbol{\theta_0}}\Big\{\sum_{u\in D_{ni}} \pi_{ni}(u)^{-1} f(u;\boldsymbol{\theta_0})\phi_i(u;\boldsymbol{\theta_0})\Big\}^T\Big] \\
= \quad & n^{-1}\sum_{i=1}^{n}\Big[\int_{A_{ni}} \pi_{ni}(t)^{-1} f(t;\boldsymbol{\theta_0})^2\phi_i(t;\boldsymbol{\theta_0})\phi_i(t;\boldsymbol{\theta_0})^T\, dt \\
+ \quad & \int_{A_{ni}}\int_{A_{ni}} \frac{\pi_{ni}(s,t)-\pi_{ni}(s)\pi_{ni}(t)}{\pi_{ni}(s)\pi_{ni}(t)} f(s;\boldsymbol{\theta_0})f(t;\boldsymbol{\theta_0})\phi_i(s;\boldsymbol{\theta_0})\phi_i(t;\boldsymbol{\theta_0})\, ds\, dt\Big]
\end{aligned}
$$

$$
\begin{aligned}
& S^{\Lambda}_{12,\infty}(\boldsymbol{\theta_0}) = n^{-1}\sum_{i=1}^{n} E_{\boldsymbol{\theta_0}}\Big[\Big\{\sum_{u\in D_{ni}} \pi_{ni}(u)^{-1} f(u;\boldsymbol{\theta_0})\phi_i(u;\boldsymbol{\theta_0}) - \int_{A_{ni}} f(t;\boldsymbol{\theta_0})\phi_i(t;\boldsymbol{\theta_0})\, dt\Big\} \\
* \quad & \Big\{\sum_{u\neq v\in D_{ni}} \pi_{ni}(u,v)^{-1} g(u,v;\boldsymbol{\theta_0})\omega_i(u,v;\boldsymbol{\theta_0}) - \int_{A_{ni}}\int_{A_{ni}} g(s,t;\boldsymbol{\theta_0})\omega_i(s,t;\boldsymbol{\theta_0})\, ds\, dt\Big\}^T\Big] \\
= \quad & n^{-1}\sum_{i=1}^{n}\Big[E_{\boldsymbol{\theta_0}}\Big\{\sum_{u\neq v\in D_{ni}} \{\pi_{ni}(u)^{-1} f(u;\boldsymbol{\theta_0})\phi_i(u;\boldsymbol{\theta_0}) + \pi_{ni}(v)^{-1} f(v;\boldsymbol{\theta_0})\phi_i(v;\boldsymbol{\theta_0})\} \\
* \quad & \pi_{ni}(u,v)^{-1} g(u,v;\boldsymbol{\theta_0})\omega_i(u,v;\boldsymbol{\theta_0}) \\
+ \quad & \sum_{u\neq v\neq w\in D_{ni}} \pi_{ni}(u)^{-1}\pi_{ni}(v,w)^{-1} f(u;\boldsymbol{\theta_0})g(v,w;\boldsymbol{\theta_0})\phi_i(u;\boldsymbol{\theta_0})\omega_i(v,w;\boldsymbol{\theta_0}) \\
- \quad & \int_{A_{ni}}\int_{A_{ni}}\int_{A_{ni}} f(r;\boldsymbol{\theta_0})g(s,t;\boldsymbol{\theta_0})\phi_i(r;\boldsymbol{\theta_0})\omega_i(s,t;\boldsymbol{\theta_0})\, dr\, ds\, dt\Big]
\end{aligned}
$$

$$
\begin{aligned}
&= n^{-1}\sum_{i=1}^{n}\Bigg[\int_{A_{ni}}\int_{A_{ni}}\{\pi_{ni}(s)^{-1}f(s;\boldsymbol{\theta_0})\phi_i(s;\boldsymbol{\theta_0})+\pi_{ni}(t)^{-1}f(t;\boldsymbol{\theta_0})\phi_i(t;\boldsymbol{\theta_0})\}\\
&* \;\; g(s,t;\boldsymbol{\theta_0})\omega_i(s,t;\boldsymbol{\theta_0})\,ds\,dt\\
&+ \;\; \int_{A_{ni}}\int_{A_{ni}}\int_{A_{ni}}\frac{\pi_{ni}(r,s,t)-\pi_{ni}(r)\pi_{ni}(s,t)}{\pi_{ni}(r)\pi_{ni}(s,t)}f(r;\boldsymbol{\theta_0})g(s,t;\boldsymbol{\theta_0})\\
&* \;\; \phi_i(r;\boldsymbol{\theta_0})\omega_i(s,t;\boldsymbol{\theta_0})\,dr\,ds\,dt\Bigg]
\end{aligned}
$$

$$
S^{\Lambda}_{22,\infty}(\boldsymbol{\theta_0}) = n^{-1}\sum_{i=1}^{n}Var_{\boldsymbol{\theta_0}}\Bigg\{\sum_{u\neq v\in D_{ni}}\pi_{ni}(u,v)^{-1}g(u,v;\boldsymbol{\theta_0})\omega_i(u,v;\boldsymbol{\theta_0})\Bigg\}
$$

$$
\begin{aligned}
&= n^{-1}\sum_{i=1}^{n}\Bigg[E_{\boldsymbol{\theta_0}}\Bigg\{\sum_{u\neq v\in D_{ni}}\pi_{ni}(u,v)^{-2}g(u,v;\boldsymbol{\theta_0})^2\omega_i(u,v;\boldsymbol{\theta_0})\omega_i(u,v;\boldsymbol{\theta_0})^T\\
&+ \sum_{t\neq u\neq v\in D_{ni}}\pi_{ni}(t,u)^{-1}\pi_{ni}(t,v)^{-1}g(t,u;\boldsymbol{\theta_0})g(t,v;\boldsymbol{\theta_0})\omega_i(t,u;\boldsymbol{\theta_0})\\
&* \;\; \{\omega_i(t,v;\boldsymbol{\theta_0})+\omega_i(v,t;\boldsymbol{\theta_0})\}^T\\
&+ \sum_{t\neq u\neq v\in D_{ni}}\pi_{ni}(t,u)^{-1}\pi_{ni}(u,v)^{-1}g(t,u;\boldsymbol{\theta_0})g(u,v;\boldsymbol{\theta_0})\omega_i(t,u;\boldsymbol{\theta_0})\\
&* \;\; \{\omega_i(u,v;\boldsymbol{\theta_0})+\omega_i(v,u;\boldsymbol{\theta_0})\}^T\\
&+ \sum_{t\neq u\neq v\neq w\in D_{ni}}\pi_{ni}(t,u)^{-1}\pi_{ni}(v,w)^{-1}g(t,u;\boldsymbol{\theta_0})g(v,w;\boldsymbol{\theta_0})\omega_i(t,u;\boldsymbol{\theta_0})\omega_i(v,w;\boldsymbol{\theta_0})^T\Bigg\}\\
&- \Bigg\{\int_{A_{ni}}\int_{A_{ni}}g(s,t;\boldsymbol{\theta_0})\omega_i(s,t;\boldsymbol{\theta_0})\,ds\,dt\Bigg\}\Bigg\{\int_{A_{ni}}\int_{A_{ni}}g(s,t;\boldsymbol{\theta_0})\omega_i(s,t;\boldsymbol{\theta_0})\,ds\,dt\Bigg\}^T\Bigg]\\
&= n^{-1}\sum_{i=1}^{n}\Bigg[\int_{A_{ni}}\int_{A_{ni}}\pi_{ni}(s,t)^{-1}g(s,t;\boldsymbol{\theta_0})^2\omega_i(s,t;\boldsymbol{\theta_0})\omega_i(s,t;\boldsymbol{\theta_0})^T\,ds\,dt\\
&+ \int_{A_{ni}}\int_{A_{ni}}\int_{A_{ni}}\frac{\pi_{ni}(r,s,t)}{\pi_{ni}(r,s)}g(r,s;\boldsymbol{\theta_0})\omega_i(r,s;\boldsymbol{\theta_0})\Bigg[\frac{\omega_i(r,t;\boldsymbol{\theta_0})^T+\omega_i(t,r;\boldsymbol{\theta_0})^T}{\pi_{ni}(r,t)}\\
&+ \frac{\omega_i(s,t;\boldsymbol{\theta_0})^T+\omega_i(t,s;\boldsymbol{\theta_0})^T}{\pi_{ni}(s,t)}\Bigg]\,dr\,ds\,dt\\
&+ \int_{A_{ni}}\int_{A_{ni}}\int_{A_{ni}}\int_{A_{ni}}\frac{\pi_{ni}(r,s,t,u)-\pi_{ni}(r,s)\pi_{ni}(t,u)}{\pi_{ni}(r,s)\pi_{ni}(t,u)}g(r,s;\boldsymbol{\theta_0})g(t,u;\boldsymbol{\theta_0})\\
&* \;\; \omega_i(r,s;\boldsymbol{\theta_0})\omega_i(t,u;\boldsymbol{\theta_0})\,dr\,ds\,dt\,du\Bigg]
\end{aligned}
$$

APPENDIX E

It can be shown (omitted, see Yuan & Jennrich 2000) that $Var_{\boldsymbol{\theta_0},\boldsymbol{\beta}_i}\{|A_{ni}|^{1/2}(\tilde{\boldsymbol{\beta}}_i - \boldsymbol{\beta_0})\} = \boldsymbol{\Xi}_n(\boldsymbol{\theta_0}, \boldsymbol{\beta}_i)$, where

$$\boldsymbol{\Xi}_n(\boldsymbol{\theta_0}, \boldsymbol{\beta}_i) = |A_{ni}| \left(\begin{array}{c|c} \Xi_{n,11}(\boldsymbol{\theta_0}, \boldsymbol{\beta}_i) & \Xi_{n,12}(\boldsymbol{\theta_0}, \boldsymbol{\beta}_i) \\ \hline \Xi_{n,12}(\boldsymbol{\theta_0}, \boldsymbol{\beta}_i)^T & \Xi_{n,22}(\boldsymbol{\theta_0}, \boldsymbol{\beta}_i) \end{array} \right),$$

where

$$\Xi_{n,11}(\boldsymbol{\theta_0}, \boldsymbol{\beta}_i) = \left(\frac{\partial}{\partial \boldsymbol{\beta}_i^T} \tilde{\boldsymbol{\kappa}}_i(\boldsymbol{\beta_0}_i, \boldsymbol{\theta_0}) \right)^{-1} Var_{\boldsymbol{\alpha_0},\boldsymbol{\beta_0}_i} \left\{ \tilde{\boldsymbol{\kappa}}_i(\boldsymbol{\beta_0}_i, \boldsymbol{\theta_0}) \right\} \left(\frac{\partial}{\partial \boldsymbol{\beta}_i^T} \tilde{\boldsymbol{\kappa}}_i(\boldsymbol{\beta_0}_i, \boldsymbol{\theta_0}) \right)^{-1},$$

$$\Xi_{n,12}(\boldsymbol{\theta_0}, \boldsymbol{\beta}_i) = \left(\frac{\partial}{\partial \boldsymbol{\beta}_i^T} \tilde{\boldsymbol{\kappa}}_i(\boldsymbol{\beta_0}_i, \boldsymbol{\theta_0}) \right)^{-1} Cov \boldsymbol{\alpha_0}, \boldsymbol{\beta_0}_i \left\{ \tilde{\boldsymbol{\kappa}}_i(\boldsymbol{\beta_0}_i, \boldsymbol{\theta_0}), \tilde{\boldsymbol{\alpha}}_n \right\} \left(\frac{\partial}{\partial \boldsymbol{\beta}_i^T} \tilde{\boldsymbol{\kappa}}_i(\boldsymbol{\beta_0}_i, \boldsymbol{\theta_0}) \right)^{-1},$$

and

$$\Xi_{n,22}(\boldsymbol{\theta_0}, \boldsymbol{\beta}_i) = \left(\frac{\partial}{\partial \boldsymbol{\beta}_i^T} \tilde{\boldsymbol{\kappa}}_i(\boldsymbol{\beta_0}_i, \boldsymbol{\theta_0}) \right)^{-1} Var_{\boldsymbol{\alpha_0},\boldsymbol{\beta_0}_i} \left\{ \tilde{\boldsymbol{\alpha}}_n \right\} \left(\frac{\partial}{\partial \boldsymbol{\beta}_i^T} \tilde{\boldsymbol{\kappa}}_i(\boldsymbol{\beta_0}_i, \boldsymbol{\theta_0}) \right)^{-1}.$$

The quantities in these expressions are, for $|A_{ni}|$ and n large enough,

$$\frac{\partial}{\partial \boldsymbol{\beta}_i^T} \tilde{\boldsymbol{\kappa}}_i(\boldsymbol{\beta_0}_i, \boldsymbol{\theta_0}) \approx -|A_{ni}|^{-1} \int_{A_{ni}} h(t; \boldsymbol{\alpha_0}, \boldsymbol{\beta_0}_i) \mathbf{z}_i(t) \mathbf{z}_i(t)^T \lambda_i(t; \boldsymbol{\alpha_0}) \tau_i(t; \boldsymbol{\beta_0}_i)\, dt,$$

$$\begin{aligned}
& Var_{\boldsymbol{\alpha_0},\boldsymbol{\beta_0}_i}\{\tilde{\boldsymbol{\kappa}}_i(\boldsymbol{\beta_0}_i, \boldsymbol{\theta_0})\} \\
= \;& |A_{ni}|^{-2} \Bigg[\int_{A_{ni}} h(t; \boldsymbol{\alpha_0}, \boldsymbol{\beta_0}_i)^2 \mathbf{z}_i(t) \mathbf{z}_i(t)^T \lambda_i(t; \boldsymbol{\alpha_0}) \tau_i(t; \boldsymbol{\beta_0}_i) \left(1 + \frac{\lambda_i(t; \boldsymbol{\alpha_0}) \tau_i(t; \boldsymbol{\beta_0}_i)}{\pi_{ni}(t)} \right) dt \\
+ \;& \int_{A_{ni}} \int_{A_{ni}} h(s; \boldsymbol{\alpha_0}, \boldsymbol{\beta_0}_i) h(t; \boldsymbol{\alpha_0}, \boldsymbol{\beta_0}_i) \mathbf{z}_i(s) \mathbf{z}_i(t)^T \lambda_i(s; \boldsymbol{\alpha_0}) \lambda_i(t; \boldsymbol{\alpha_0}) \tau_i(s; \boldsymbol{\beta_0}_i) \tau_i(t; \boldsymbol{\beta_0}_i) \\
* \;& \left(\frac{\pi_{ni}(s,t) - \pi_{ni}(s) \pi_{ni}(t)}{\pi_{ni}(s) \pi_{ni}(t)} \right) ds\, dt \Bigg],
\end{aligned}$$

and

$$
\begin{aligned}
& Cov_{\boldsymbol{\alpha_0},\boldsymbol{\beta_0}_i}\{\tilde{\boldsymbol{\kappa}}_i(\boldsymbol{\beta_0}_i,\boldsymbol{\theta_0}),\tilde{\boldsymbol{\alpha}}_n\} \\
\approx\ & (n|A_{ni}|)^{-1}\Bigg[\sum_{i=1}^{n}\Bigg\{\Bigg(\int_{A_{ni}} f(t;\boldsymbol{\theta_0})(1-f(t;\boldsymbol{\theta_0}))\mathbf{x}_i(t)\mathbf{x}_i(t)^T\lambda_i(t;\boldsymbol{\alpha_0})\{\tau_i(t;\boldsymbol{\beta_0}_i) \\
-\ & M_{\boldsymbol{\beta}_i}(\mathbf{z}_i(t);\boldsymbol{\Sigma_0})\}\,dt\Bigg) - \int_{A_{ni}} f(t;\boldsymbol{\theta_0})\mathbf{x}_i(t)\phi_i(t;\boldsymbol{\theta_0})^T dt\Bigg\}\Bigg]^{-1} \\
*\ & \Bigg[\int_{A_{ni}} h(t;\boldsymbol{\alpha_0},\boldsymbol{\beta_0}_i)f(t;\boldsymbol{\theta_0})\mathbf{z}_i(t)\mathbf{x}_i(t)^T\lambda_i(t;\boldsymbol{\alpha_0})\tau_i(t;\boldsymbol{\beta_0}_i)\left(1-\frac{\lambda_i(t;\boldsymbol{\alpha_0})M_{\boldsymbol{\beta}_i}(\mathbf{z}_i(t);\boldsymbol{\Sigma_0})}{\pi_{ni}(t)}\right)dt \\
+\ & \int_{A_{ni}}\int_{A_{ni}}\left(\frac{\pi_{ni}(s,t)-\pi_{ni}(s)\pi_{ni}(t)}{\pi_{ni}(s)\pi_{ni}(t)}\right)h(s;\boldsymbol{\alpha_0},\boldsymbol{\beta_0}_i)f(t;\boldsymbol{\theta_0})\mathbf{z}_i(s)\phi_i(t;\boldsymbol{\theta_0})^T \\
*\ & \lambda_i(s;\boldsymbol{\alpha_0})\tau_i(s;\boldsymbol{\beta_0}_i)\,ds\,dt\Bigg]
\end{aligned}
$$

The variance of $\tilde{\boldsymbol{\alpha}}_n$ may be obtained from the top-left block of (B.5). Note that this quantity must be divided by n, since (B.5) provides the variance of $n^{1/2}\tilde{\boldsymbol{\theta}}_n$.

4.6 References

Affleck, G., Zautra, A., Tennen, H. & Armelia, S. (1999). Multilevel daily proces designs for consulting and clinical psychology: A preface for the perplexed. *Journal of Consulting and Clinical Psychology, 67*, 746–754.

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology, 54*, 579–616.

Borodin, A., & Soshnikov, A. (2003). Janossy densities I: Determinantal ensembles. *Journal of Statistical Physics, 113*, 595–610.

Breslow, N. E. & Clayton, D. G. (1993). Approximate inference in Generalized Linear Mixed Models. *Journal of the American Statistical Association, 88*(421), 9–25.

Cassel, C.-M., Särndal, C.-E., & Wretman, J. H. (1977). *Foundations of inference in survey sampling.* New York: Wiley.

Cordy, C. B. (1993). An extension of the Horvitz-Thompson theorem to point sampling from a continuous universe. *Statistics & Probability Letters, 18*, 353–362.

Cox, D. R. (1972). The statistical analysis of dependencies in point processes. In *Stochastic Point Processes*, P. A. W. Lewis (Ed.), 55–66. New York: Wiley.

Cressie, N. (1991). *Statistics for spatial data*. New York: Wiley.

Crowder, M. (1986). On consistency and inconsistency of estimating equations. *Econometric Theory* **2**(3), 305–330.

Diggle, P. J. (2003). *Statistical analysis of spatial point patterns* (2nd ed.). New York: Oxford University Press.

Green, P. J. (1987). Penalized likelihood for general semi-parametric regression models. *International Statistical Review, 55*(3), 245–259.

Guan, Y. & Loh, J. M. (2007). A thinned block bootstrap variance estimation procedure for inhomogenous spatial point patterns. *Journal of the American Statistical Association, 102*(480), 1377–1386.

Henderson, H. V. & Searle, S. R. (1979). Vec and vech operators for matrices, with some uses in Jacobians and multivariate statistics. *Canadian Journal of Statistics, 7*(1), 65–81.

Jansen, J. (1990). On the statistical analysis of ordinal data when extravariation is present. *Applied Statistics, 39*(1), 75–84.

Kolman, B. (1970). *Elementary linear algebra.* New York: McMillan Publishing Company.

Lee, Y. & Nelder, J. A. (1996). Hierarchical generalized linear models. *Journal of the Royal Statistical Society, Series B (Methodological), 58*, 619–678.

Lee, Y., Nelder, J. A. & Pawitan, Y. (2006). *Generalized Linear Models with random effects: Unified analysis via H-likelihood.* New York: Chapman & Hall.

McCulloch, C. E. (1997). Maximimum likelihood algorithms for Generalized Linear Mixed Models. *Journal of the American Statistical Association, 92*(437), 162–170.

Nelder, J.A. & Wedderburn, R. W. M. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society, Series A (General), 135*(3), 370–384.

Neustifter, B. (2009). *Random effects in point processes: Adding flexibility to Ecological Momentary Assessment analysis.* Unpublished doctoral dissertation, University of Georgia.

Rathbun, S. L., Shiffman, S., & Gwaltney, C. J. (2006). Point process models for event history data: Applications in behavioral science. In *Models for intensive longitudinal data*, T. A. Walls & J. L. Schafer (Eds.), 219–253. New York: Oxford University Press.

Rathbun, S. L., Shiffman, S., & Gwaltney, C. J. (2007). Modelling the effects of partially observed covariates on Poisson process intensity. *Biometrika, 94*, 153–165.

Schwartz, J. E. & Stone, A. A. (2007). The analysis of real-time momentary data: A practical guide. In *The Science of Real-time Data Capture: Self-report in Health Research*, Stone, A. A., Shiffman, S. S., Atienza, A., & Nebeling, L. (Eds.), 76–113. New York: Oxford University Press USA.

Shiffman, S. et al. (2002). Immediate antecedents of cigarette smoking: An analysis from ecological momentary assessment. *Journal of Abnormal Psychology, 111*, 531–545.

Shiffman, S. & Stone, A. A. (1998). Ecological Momentary Assessment: A new tool for behavioral medicine research. In *Technology and Methods in Behavioral Medicine*, Krantz, D. S. & Baum, A. (Eds.), 117–132. Philadelphia: Lawrence Erlbaum Associates.

Smyth, J. et al. (2001). The use of ecological momentary assessment approaches in eating disorder research. *International Journal of Eating Disorders, 30*, 83–95.

Stone, A. A., Broderick, J. B., Porter, L., & Kaell, A. T. (1997). The experience of rheumatoid arthritis pain and fatigue: Examining momentary reports and correlates over one week. *Arthritis Care and Research, 10*, 185-193.

Stone, A. A., Broderick, J. E., Shiffman, S. S. & Schwartz, J.E. (2004). Understanding recall of weekly pain from a momentary assessment perspective: Absolute agreement, between- and within-person consistency, and judged change in weekly pain. *Pain, 107*(1–2), 61–69.

Stone, A. A. et al. (1998). A comparison of coping assessed by ecological momentary assessment and retrospective recall. *Journal of Personality and Social Psychology, 74*, 1670–1680.

Stone, A. A. & Shiffman, S. (2002). Capturing momentary, self-report data: A proposal for reporting guidelines. *Annals of Behavioral Medicine, 24*, 236–243.

Tierney, L. & Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association, 81*(393), 82–86.

Waagepetersen, R. (2008). Estimating functions for inhomogenous spatial point processes with incomplete covariate data. *Biometrika, 95*(2), 351–363.

Yuan, K. & Jennrich, R. I. (2000). Estimating equations with nuisance parameters: Theory and applications. *Annals of the Institute of Statistical Mathematics, 52*(2), 343–350.

Chapter 5

Conclusion

5.1 Findings

Ecological Momentary Assessment (EMA) presents an excellent opportunity for numerous fields. By randomly sampling subject states in their natural environments EMA reduces bias in findings and allows for unbiased estimation of covariates' mean states across the study interval. EMA can also be used to gather data about the timing of, and covariates related to, repeated behavioral events in subjects. While some methods have been suggested for the analysis of EMA data (Affleck et al. 1999; Bolger, Davis, & Rafaeli 2003; Schwartz & Stone 2007), such techniques are not appropriate for modeling the timing of these repeated events, because they are based on regression-type models that are not appropriate for the timing of longitudinal occurrences.

This dissertation extends the current methods for the analysis of repeated event data. While modulated Poisson processes provide a useful method for modeling such data (Rathbun, Shiffman, & Gwaltney 2006), recently proposed methods (Rathbun, Shiffman, & Gwaltney 2007; Waagepetersen 2008) contain the implicit assumption that all subjects have identical intensity functions. For psychological and social science research, where Ecological Momentary Assessment (EMA) data collection is a growing trend, such assumptions are often unrealistic; subjects often have different baseline rates of event occurrence or different reactions to changes in covariates.

In order to address this limitation of the literature, this dissertation proposed two different models that allow for inter-subject variability. An extension of the modulated Poisson process inspired by Hierarchical Generalized Linear Models (HGLMs) was developed in Chapter 2 that included a randomly distributed baseline term. This random-baseline model resembled the Poisson-Gamma HGLM. The random-baseline model was proposed with two separate sets of estimating equations and assumptions in Chapters 2 and 3. In the first, more restrictive setting, it was shown that the solution to the approximated maximum likelihood (ML) estimating equations for a Gamma-distributed baseline, provided estimators for the covariate coefficient and baseline parameters that were consistent and asymptotically normal. The ML equations were estimated using Rathbun et al.'s (2007) integral estimation method, and both increasing sample and domain asymptotics were required for the results. In Chapter 3, the assumption that baseline rates are gamma-distributed was relaxed by using simpler estimating equations that are equal in expected value to the ML estimating equations. These revised equations both improved the computational efficiency of the estimated ML equations and avoided possible instability problems in the original estimating equations. The revised estimating equations were generalized to encompass both Rathbun et al. (2007) and Waagepetersen's (2008) integral estimation methods. This dissertation showed that the coefficient and baseline variance estimators for obtained by these equations are consistent under increasing sample size for a general random-baseline point process with some fourth-moment assumptions. These estimators were also shown to be asymptotically normal under the assumption of a Poisson process. It is noteworthy that the estimators obtained in Chapter 3 have the desirable asymptotic properties under increasing sample size, rather than both sample and domain size as the original ML-based estimators required.

The second model developed, in Chapter 4, was more flexible than the random-baseline model. Rather than focusing upon baseline rates it allowed for a mixed-effect model for covariate effects. This mixed-effect model allowed some covariates to have fixed effects on event occurrence rates, while others could have varying impacts among subjects. This model more closely resembled the Poisson Generalized Linear Mixed Model (GLMM) because random effect terms were assumed to be normally distributed with a zero mean and unknown general covariance matrix. While the mixed-effect model had more flexibility, its likelihood had no closed form. Unbiased estimating equations for the fixed effects were thus developed that were simple in form and related to a Laplace approximation of the likelihood. This dissertation also proposed unbiased estimating equations for the covariance matrix that, while not related to the likelihood, were similar in form to those obtained for variance estimation in Chapter 3. It was proven that the estimators obtained from these equations, using either Rathbun et al. (2007) or Waagepetersen's (2008) estimation methods, were consistent and asymptotically normal under increasing sample size. Furthermore, this chapter provided estimating equations for the subject-level random effect parameters. Finally, it was shown that under increasing domain asymptotics the estimators obtained for these terms were also consistent and asymptotically normal.

All of the proposed methods were demonstrated using Shiffman et al.'s (2002) EMA study of smoking data. The results were compared with Rathbun et al.'s (2007) results, and when applicable, with Waagepetersen's (2008) as well. These comparisons showed that the proposed estimation methods increased the variance of the estimates of fixed effects, but yielded with the same significant variables. The techniques proposed in this dissertation appeared to be reliably analyzing the data, since the results coincided both with Rahtbun et al. (2007) and Waagepetersen's (2008) methods, as well as literature on smoking. For the random-baseline model in Chapters 2 and 3, it was not readily apparent from the obtained estimates whether the Gamma-distributed baseline was appropriate. For the mixed-effect model in Chatper 4, however, the results seemed to strongly suggest that one of the random effects (Other Smokers) was appropriately modeled by the normally distributed random effect term, rather than a fixed effect.

Overall, the two models proposed provide an important step forward for the field of point process and repeated event analysis. They demonstrate a manner in which one may add flexibiity to such models and allow for partially-observed covariates such as those obtained in EMA studies. The majority of EMA literature has focused upon analyzing event counts or other related quantities, yet art of the usefulness of EMA's immediately obtained data is that they record the timing of events. Rathbun et al.'s (2007) recommendation of, and initial research into, the use of point processes to analyze the timing of these events as related to covariates provided the groundwork for the more flexible models presented in this dissertation. Waagepetersen's (2008) improvements upon Rathbun et al.'s (2007) integral estimation methods were incorporated into the models in Chapters 3 and 4, allowing individual researchers to choose between the computational efficiency of Rathbun et al.'s methods and the informational efficiency of Waagepetersen's. The models proposed in this dissertation provide new, more flexible methods of analyzing EMA and other repeated-event data. With regard to the Shiffman et al. (2002) dataset, these models may help us to better understand how covariates impact smoking rates and which covariates have universal effects rather than varying among smokers. Such conclusions can lead to the development of more focused and effective smoking intervention programs. This dissertation's methods may be extended beyond EMA data to numerous other applications of point processes, such as epidemiology, forestry, or any other field that models spatio-temporal data with point processes.

5.2 Future Research

While this dissertation provides important steps toward the development of new, more general models for the analysis of event timing in repeated-event data, there are further areas for research that must be addressed. The need to develop diagnostic techniques and tests for lack of fit for the proposed models is of primary importance. Techniques such as the K function (Ripley 1976, 1977, 1981) are used to analyze the second-order properties of data and therefore make conclusions regarding whether those data are derived from a Poisson process. More evaluation is necessary to determine whether K functions and related diagnostics may be extended to this dissertation's models, and how to best do so. Since the consistency and asymptotic normality of the mixed-effects model and the asymptotic normality of random-baseline model in Chapter 3 depend upon the point process being Poisson, a method of verifying whether this assumption is satisfied must be developed.

Diagnostic measures are also needed for the fit of the subject-level parameter estimates of the baseline or random effects. In the estimation of the subject-level parameters, significant differences in the histograms of these parameters and their expected distributions were apparent. This suggests that incorrect models may be being fitted to the data or that further research on the estimation of subject-level parameters is needed. A conclusive test for the appropriateness of a baseline or random effect term would allow researchers to spare the extra variation in estimation when it is unnecessary. Such a test could also address concerns regarding the mis-fit of the subject-level parameters. With regard to the mixed-effect model in Chapter 4, such tests would provide a more firm theoretical position for researchers to determine whether a variable should be modeled by a fixed or random effect. It was clear in the data analysis of the mixed-effect model that the addition or removal of a random effect for one covariate could strongly change the variances and significances of other covariates. These changes included whether the random effects for those covariates appeared significant according to the asymptotic normal approximation. Such strong effects make it difficult to accurately determine which variables should have their effects modeled as fixed and which as random. More conclusive and stable tests would be useful in fitting the most accurate possible model.

Finally, tests for variable-selection would be useful with these models. While the asymptotic normality provides approximate significance tests for single parameters, this study does not provide an equivalent to tests such as the F-test for subset models in regression. When dealing with EMA data, which may contain numerous covariates, it would be useful to have a test that allows us to eliminate groups of variables from the model without the need for multiple-testing correction.

The application of a Poisson process with random elements to repeated event data, particularly in the case that covariates are partially observed, is a new branch of point process research. As such, this area has a great deal of unexplored applications, and also room for theoretical scrutiny. Currently, these new models require diagnostic criteria, so that researchers applying this dissertation's techniques may check for lack of fit or other statistical anomalies. Research is needed on techniques for checking the Poisson process assumption, the fit of subject-level parameters, the appopriatness of random effects, and model comparison in order to make these models more soundly applicable.

5.3 Conclusions

The models and estimation methods proposed in this dissertation should extend the field of repeated-event analysis, particularly for Ecological Momentary Assessment (EMA) data. This dissertation extends current methods of repeated-event analysis by proposing models with inter-subject variability. These models are more flexible and provide room for more realistic and accurate conclusions to be drawn from EMA data. This dissertation has shown that the models proposed in Chapters 2 through 4 provide estimators with desirable asymptotic properties. Each model has also been demonstrated on an EMA dataset of smokers. While EMA is used in the social sciences, the theories presented here are applicable to a much wider array of fields. The models presented here provide a flexible and theoretically sound basis for analyzing the data and determining the effects of covariates upon the rate or spacing of event occurrence for any studies in which covariates can be measured at the times or locations of events as well as other random times or spaces.

5.4 References

Affleck, G., Zautra, A., Tennen, H. & Armelia, S. (1999). Multilevel daily proces designs for consulting and clinical psychology: A preface for the perplexed. *Journal of Consulting and Clinical Psychology, 67*, 746–754.

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology, 54*, 579–616.

Schwartz, J. E. & Stone, A. A. (2007). The analysis of real-time momentary data: A practical guide. In *The Science of Real-time Data Capture: Self-report in Health Research*, Stone, A. A., Shiffman, S. S., Atienza, A., & Nebeling, L. (Eds.), 76–113. New York: Oxford University Press USA.

Rathbun, S. L., Shiffman, S., & Gwaltney, C. J. (2006). Point process models for event history data: Applications in behavioral science. In *Models for intensive longitudinal data*, T. A. Walls & J. L. Schafer (Eds.), 219–253. New York: Oxford University Press.

Rathbun, S. L., Shiffman, S., & Gwaltney, C. J. (2007). Modelling the effects of partially observed covariates on Poisson process intensity. *Biometrika, 94*, 153–165.

Ripley, B. D. (1976). The second-order analysis of stationary point processes. *Journal of Applied Probability, 13*, 255–266.

Ripley, B. D. (1977). Modelling spatial statistics. *Journal of Royal Statistical Society (Series B), 39*, 172–212.

Ripley, B. D. (1981). *Spatial Statistics.* New York: Wiley.

Shiffman, S. et al. (2002). Immediate antecedents of cigarette smoking: An analysis from ecological momentary assessment. *Journal of Abnormal Psychology, 111*, 531–545.

Waagepetersen, R. (2008). Estimating functions for inhomogenous spatial point processes with incomplete covariate data. *Biometrika, 95*(2), 351–363.