STATE SPACE MODELS FOR NEUROPSYCHOLOGICAL COGNITIVE SCORES

BY ZACHARY BAUCOM[1,*], YORGHOS TRIPODIS[1,†], MICHAEL ALOSCO[2] AND EVAN JOHNSON[3]

[1]*Department of Biostatistics, Boston University,* [*]*zhbaucom@bu.edu;* [†]*yorghos@bu.edu*

[2]*Alzheimer's Disease Research Center and CTE Center, Boston University, malosoco@bu.edu*

[3]*Department of Medicine, Boston University, wej@bu.edu*

Alzheimer's disease, and other related dementia diseases, are a worsening issue with an acceleration in today's aging population. Longitudinal cognitive assessment of those suffering from dementia offers vital insight into disease progression and allows for assessment of possible disease interventions. Difficulty in modeling such data arises as there are often non-linear and heterogenous patterns of decline from patient to patient. We propose the use of state space models (SSM), specifically a Local Linear Trend (LLT) model, as an alternative to the commonly used linear mixed effect models (LMEM) for longitudinal assessments. The proposed model includes the estimation of interpretable population linear effects on the outcome, while also allowing for subject-specific non-linearities in cognitive trajectories. To fit the LLT model, we utilize the traditional full likelihood estimation using the Kalman Filter and Kalman Smoother. We also compare the use of a partitioned LLT and a Bayesian LLT for computational efficiency. In two separate simulation analyses, we show the advantages of the LLT models over the predominant techniques. We go on to show that of the LLT methods, they Bayesian LLT excels. Lastly, we use the LLT models to estimate the effect of the APOE e4 allele on cognitive trajectory.

1. Introduction. According to the World Health Organization, dementia affects around 50 million people in the world today, with 60-70% of those due to Alzheimer's Disease ("Dementia" 2020). The number of those suffering from Alzheimer's disease (AD) and Alzheimer's disease related dementia (ADRD) is only accelerating due to an increase in the aging population. Pathologically, AD is characterized by beta-amyloid neuritic plaques and accumulation of hyperphosphorylated tau (p-tau). These physical changes in the brain correspond with episodic memory and expressive language difficulties in the early stages, which then progresses to executive dysfunction, agitation, and an inability to care for one's self in the later stages ("2021 alzheimer's disease facts and figures" 2021). In addition to suffering experienced by those with AD, there is an added physical, emotional, and financial burden on family members and caretakers.

Much research has been done to understand how and why AD progresses in order to facilitate early detection and diagnosis, which is critical for timely implementation of intervention strategies. In particular, the National Alzheimer's Coordinating Center (NACC) is a publicly available, National Institute on Aging funded, centralized repository of harmonized data from approximately 30 Alzheimer's Disease Research Centers (ADRC) in the United States (Beekly et al. 2007). The overarching goal of the NACC is to facilitate research on AD and related diseases. Each ADRC follows a cohort of participants with and without cognitive impairment and approximately annually conducts harmonized cognitive, neuropsychiatric, and neurological evaluations. The ADRCs then deposit the harmonized data to the NACC which is compiled

into large amounts of serial neuropsychological measurements. Analyzing these longitudinal outcomes provides vital insight into the disease course and presentation.

A number of challenges exist when accurately estimating effect modifiers on the neuropsychological outcomes presented in the NACC data. For interpretable results, we often rely on estimating linear effect modifiers. However, participant level cognition data are often non-linear over time and heterogeneous from subject to subject. Many participants may show a similar overall trajectory, but a given participant often shows unique patterns in their cognitive course (Rongve et al. 2016; Vik-Mo et al. 2018). This heterogeneity can arise from unmeasured changes over time in social behavior, treatments, routines, or biologic processes that have an effect on the neuropsychological outcome.

Subject level heterogeneity in cognitive trajectories often goes unaccounted for, as is the case in studying the effect of the APOE e4 allele on cognitive decline. The APOE e4 gene has been known to be associated with Alzheimer's disease (Raber et al. 2004), but there have been conflicting publications on the effect of the allele on the rate of cognitive decline, as discussed by (Martins et al. 2005). Some studies claim no effect of the e4 allele on cognitive decline, some conclude a faster rate of decline, and a few even indicate a protective effect on the rate of decline. None of these studies seek to control for heterogeneous trajectories, which may contribute to underestimated parameter standard errors, leading to a mis-specified type I error and too many significant tests.

To better accommodate data characteristics common to neuropsychological measurements, and to estimate the effect of the APOE e4 allele on cognitive trajectory, we propose the use of a Local Linear Trend model (LLT). The LLT is a class of State Space Model (SSM) and is proposed due to its flexibility in measuring population linear effects while allowing for heterogeneous participant trajectory patterns. The SSM is traditionally a time series technique employed in situations with only a few number of participants and a large series of time points (Durbin and Koopman 2012; Fildes et al. 1991). The SSM framework has been successfully used in numerous fields such as engineering, econometrics, and health sciences. We propose its use in the situation more common to observational cognition data of a large number of participants with only a few repeated measures.

The Linear Mixed Effect Model (LMEM), which has an intuitive comparison to the LLT model, is very common when modeling neuropsychological outcomes (Feng et al. 2009; Rolland et al. 2009; Sanders et al. 2010; Stott et al. 2008). The LMEM is used due to its simple implementation and interpretation (Laird and Ware 1982), but the LMEM is typically only meant to model low order polynomials of trajectory. The LLT shares an equivalent linear effect interpretation as the LMEM, but with an added flexibility for non-linear changes in participant cognition. Some studies using the LMEM have enforced an AR(1) auto-correlation structure when modeling cognitive data to help counteract heterogeneous non-linearities (Dueck et al. 2009; Mitchell et al. 2020). Autocorrelation can be considered the effect of unobserved time-varying variables (UTV) on an outcome. Both the LMEM with an AR(1) autocorrelation and LLT make different assumptions on the behavior of the UTV. The AR(1) holds the strict assumption that the UTV effect will revert to the levels they held at an often arbitrary time of baseline. The LLT allows more flexibility in letting the UTV effect vary freely over time following a random walk process. Even with the added flexibility of the LLT, we show that it still produces accurate effect parameter estimates. Despite the advantages of the SSM, it has not been used for neuropsychological data.

Fitting a model with the traditional full likelihood LLT estimation process on a large data set can lead to extensive computation time. Estimation of the LLT model is an iterative process that includes difficult to avoid matrix inversions which increases computation cubicly as the sample size increases (Durbin and Koopman 2012). To address the computational burden of the brute force full likelihood estimation, we propose and compare a partitioned LLT and a Bayesian LLT.

The partitioned LLT randomly partitions the participants into a pre-specified number of groups then runs the traditional estimation process independently on each group. Population parameter estimates from each group are combined for more accurate population inference. By pre-specifying the maximum number of participants in a group, computation time increases linearly as the sample size increases.

The Bayesian estimation approach to the LLT model allows us to avoid large matrix inversions and instead relies on a Markov Chain Monte Carlo (MCMC) Gibb's sampling algorithm. Accurate estimation using MCMC is typically more computationally expensive for a small number of participants, but maintains a linear increase in computation, allowing it to quickly outperform the full likelihood method. Along with being the most accurate of the LLT estimation procedures, the Bayesian method can also be useful in identifying model misspecification through visual inspection of the MCMC draws.

To assess model performance we show two simulation analyses comparing the LMEM and our proposed LLTs: 1) on fully simulated data controlling the underlying data generation process and 2) on real data where the underlying data generating process is unknown.

By controlling the underlying data generation process we are able to demonstrate the LLT effectiveness under correct and incorrect model specification. The fully simulated data analysis show that the LMEMs and LLTs are unbiased, but only the LLTs maintain proper parameter coverage, even when the underlying data generating model follows that of a traditional LMEM. Maintaining proper coverage is essential for correct parameter inference.

The effectiveness of the LLT models is further showcased when we simulate using real data with an unknown underlying data generation process. Utilizing real data from the NACC we are able to show the LLT allows for proper inference of a simulated linear effect on the Animals test outcome. The LMEM with an AR(1) correlation structure fails to maintain proper parameter coverage, yielding poor results consistent with the aforementioned fully simulated analysis. The real data simulation validates that the LLTs intuitive interpretation is more realistic for neuropsycholgical outcomes.

After the proficiency of the proposed LLT is shown, we use the LMEM models and LLT models to estimate the effect of APOE e4 allele on the Animals test outcome over time. The Animals outcome, in which a participant names as many animals as possible within a minute, was chosen due its continual decline throughout dementia disease progression (Ferris and Farlow 2013).

2. Methods.

2.1. *Data and Model of Interest.*

2.1.1. *Data.* Data collected by the National Alzheimer's Coordinating Center (NACC) Uniform Data Set Version 2.0 (UDS, September 2014) was used to test the proposed LLT model validity. Criteria for entry into the analysis requires Alzheimer's disease participants to have transitioned from cognitively normal to mild cognitive impairment (MCI) or Dementia during the NACC follow-up period. This results in 1,269 participants with an average of 6.6 return visits (SD = 3.1) seen over a period of 6.7 years (SD = 3.4). The average transition from cognitively normal to MCI or Dementia occurs at 4.9 visits (SD = 2.8). In this sample 36.0% of the subjects carry at least 1 APOE e4 allele. The average study participants are predominantly female (63.9%) and of white ethnicity (87.3%). Participants are primarily older with an average age of 77.1 years old (SD = 8.2) and have an average of 16.0 years of education (SD = 6.5).

2.1.2. *Model of Interest.* The Animals test outcome, in which participants have one minute to name as many animals as possible, was used as the cognitive outcome of interest (mean = 16.6, SD = 5.7). After controlling for dementia status (1 = diagnosed with MCI or dementia, 0 = otherwise), sex (1 = female, 0 = male), race (1 = white, 0 = other), age (mean centered), and education (mean centered), we wish to accurately estimate the effect of having an APOE e4 allele (1 = has at least 1 APOE e4 allele, 0 = otherwise) on the Animals test trajectory. Previous research suggests that the effect of APOE e4 status differs between males and females, therefore an interaction between e4 status and sex is included in the model. To measure the effect of the dependent variables on the Animals test trajectory, all dependent variables are put into the model as an interaction with time.

2.2. *Proposed Model For Neuropsychological Outcomes.* In the context of neuropsychological outcomes, the vector μ_{ij} represents the components of subject i's underlying cognition at observation j for $i \in \{1, 2, ..., n\}$ and $j \in \{1, 2, ..., J\}$. The observation y_{ij} are scores from a test aimed to measure the true underlying cognition of μ_{ij}. We propose modeling the underlying cognition μ_{ij} broken into two components, 1.) $\boldsymbol{\beta}$, which is a population parameter that captures the linear trajectory based on a subject's measured characteristics and 2.) α_{ij}, which captures subject specific random variation in cognition. This proposed model is a special case of the SMM called a Local Linear Trend Model (LLT) and can be written as,

$$y_{ij} = \alpha_{ij} + \boldsymbol{x}_{ij}\boldsymbol{\beta}_j + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$$

(1)
$$\mu_{ij} = \begin{bmatrix} \alpha_{ij} \\ \boldsymbol{\beta}_j \end{bmatrix} = \begin{bmatrix} \alpha_{i(j-1)} \\ \boldsymbol{\beta}_{j-1} \end{bmatrix} + \begin{bmatrix} \eta_{ij} \\ 0_{p \times 1} \end{bmatrix}, \quad \eta_{ij} \sim N(0, \delta_{ij}\sigma_\eta^2)$$

$$\alpha_{i0} \sim N(a_0, P_0), \quad \boldsymbol{\beta}_0 \sim N(\boldsymbol{\beta}, 0)$$

The row vector $\boldsymbol{x}_{ij} \in R^p$ are the measured independent time varying-variables for subject i at the jth observation. We assume the column vector $\boldsymbol{\beta}_j \in R^p$ remains fixed over time and follows the same linear effect interpretation common with other linear regression methods. The linear effects of $\boldsymbol{\beta}_j$ is a population parameter, as there is no index for subject i.

The proposed model differs from traditional models in that there is a randomly varying subject specific α_{ij} that follow a random walk through time. With respect to y_{ij}, α_{ij} follows the random walk about $\boldsymbol{x}_{ij}\boldsymbol{\beta}_j$ and therefore can be interpreted as random variation in cognition not accounted for by the predictors $\boldsymbol{x}_{ij}$. Notice, $E(\alpha_{ij}|\alpha_{i(j-1)} = \tilde{a}_{i(j-1)}) = \tilde{a}_{i(j-1)}$. This indicates that the underlying variations in one's cognitive state at observation j, not accounted for by the predictors $\boldsymbol{x}_{ij}$, is centered on their underlying cognitive state at observation $j - 1$ and can freely vary up or down from that state.

The variance of the underlying state α_{ij} is time dependent as the jth observation can occur at any time. To accommodate unequal observation times we can adjust the variance of α_{ij}. Let t_{ij} be the observation time for the jth observation. To adjust the variance we multiply the population variance σ_η^2 with $\delta_{ij} = t_{ij} - t_{i(j-1)}$ for observation j. This adjustment allows for the estimation of the underlying variance σ_η^2, even when each subject's α_{ij} variance is different due to heterogeneous observation times. The freedom of variation in the subject specific cognition is much more flexible than restriction enforced in commonly used LMEM models.

A simple comparison to the LLT is a random intercept LMEM with an AR(1) autocorrelation in the errors. The LMEM AR(1) can be generalized as,

$$y_{ij} = b_{i0} + \boldsymbol{x}_{ij}\boldsymbol{\beta} + e_{ij}, \quad b_{i0} \sim N(0, \sigma_b^2)$$

$$e_{ij} = \rho e_{i(j-1)} + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \sigma_\epsilon^2), \quad \rho \in (-1, 1)$$

The linear effect is represented by $\boldsymbol{\beta}$ and $\boldsymbol{x}_{ij}$ are the covariates of subject i at observation j. The value e_{ij} can be considered the underlying cognitive state at observation j. If $\rho = 1$ then e_{ij} can vary freely and this model mirrors the LLT without a measurement error on the observation equation. However, if $|\rho| < 1$ then $|E(e_{ij}|e_{i(j-1)} = \tilde{e}_{i(j-1)})| = |\rho\tilde{e}_{i(j-1)}| \leq |\tilde{e}_{i(j-1)}|$, meaning the underlying state is reverting back to the level it held at time 0. The LLT relaxes the restrictive mean reverting assumption.

2.2.1. *Autocorrelation.* Our proposed model implies a flexible dynamic moving average auto-correlation structure. Suppose $\delta_{ij} = 1$ for all j. The correlation between any two time points for subject i is defined as,

$$corr(y_{ij}, y_{i(j+\tau)}) = \frac{j\sigma_\eta^2 + P_0}{\sqrt{\sigma_\varepsilon^2 + j\sigma_\eta^2 + P_0}\sqrt{\sigma_\varepsilon^2 + (j+\tau)\sigma_\eta^2 + P_0}}$$

If there is no variation in α over time ($\sigma_\eta^2 = 0$) then $corr(y_{ij}, y_{i(j+\tau)}) = \frac{P_0}{\sigma_\varepsilon^2 + P_0}$. This auto-correlation structure provides an observationally equivalent parallel to a random intercept LMEM. Next, consider the reduced model of the LLT,

$$y_{ij} = \alpha_{i0} + \sum_{j=1}^{t} \eta_{ij} + \boldsymbol{x}_{ij}\boldsymbol{\beta}_0 + \varepsilon_{ij}$$

If $\sigma_\eta^2 = 0$ then $\eta_j = 0$ for $j \in \{1, 2, ..., J\}$ then the model further reduces to,

$$y_{ij} = \alpha_{i0} + \boldsymbol{x}_{ij}\boldsymbol{\beta}_0 + \varepsilon_{ij}$$

Where $\alpha_{i0} \sim N(a_0, P_0)$, which is directly comparable to a linear mixed effect model with a random intercept. This highlights that the proposed model can accommodate the simplistic LMEM while also accommodating more complex temporal auto-correlation.

2.3. *General Linear Gaussian State Space Model.* The proposed LLT model is derived from the general linear state space model. For the general linear SSM we use matrix notation with bold face vector $\boldsymbol{y}_j = [y_{1j}, y_{2j}, ..., y_{nj}]'$ as the outcome and $\boldsymbol{\mu}_j = [\mu_{1j}, \mu_{2j}, ..., \mu_{qj}]'$ as the latent state. A general linear state space model can be denoted as:

(2)
$$\boldsymbol{y_j} = \boldsymbol{F_j}\boldsymbol{\mu_j} + \boldsymbol{v_j}, \quad \boldsymbol{v_j} \sim N(\boldsymbol{0}, \boldsymbol{V_j})$$
$$\boldsymbol{\mu_j} = \boldsymbol{G_j}\boldsymbol{\mu_{j-1}} + \boldsymbol{w_j}, \quad \boldsymbol{w_j} \sim N(\boldsymbol{0}, \boldsymbol{W_j}) \ \& \ \boldsymbol{\mu_0} \sim N(\boldsymbol{u}, \boldsymbol{P})$$

Both $\boldsymbol{v_j}$ and $\boldsymbol{w_j}$ are independent and mutually uncorrelated. The vector $\boldsymbol{\mu_j} \in R^q$ represents the unobserved true state of the process. To estimate $\boldsymbol{\mu_j}$ we use the observations $\boldsymbol{y_j}$. The observations $\boldsymbol{y_j}$ are assumed to be a linear combination of $\boldsymbol{\mu_j}$ after being transformed by the observation matrix $\boldsymbol{F_j}$, which is fixed by design, plus the random noise $\boldsymbol{v_j}$. The latent states follow a Markov Chain process where $\boldsymbol{\mu_j}$ is a linear function of $\boldsymbol{\mu_{j-1}}$ in the form $\boldsymbol{G_j}\boldsymbol{\mu_{j-1}}$ with the random noise $\boldsymbol{w_j}$. The matrix $\boldsymbol{G_j}$ is the state transition matrix and is also assumed fixed by design. This state space model can be regarded as a Hidden Markov model with continuous latent states and continuous observations.

Here, we denote $\hat{\boldsymbol{\mu}}_{a|b} = E(\boldsymbol{\mu_a}|\boldsymbol{y_1}, \boldsymbol{y_2}, ..., \boldsymbol{y_b})$ and $\boldsymbol{P}_{a|b} = \text{Var}(\boldsymbol{\mu_a}|\boldsymbol{y_1}, \boldsymbol{y_2}, ..., \boldsymbol{y_b})$ for a and b in $\{1, 2, ..., J\}$. With known initial parameters of underlying state $\boldsymbol{\mu_0}$ (mean $\boldsymbol{u}$ and variance $\boldsymbol{P}$) and variance parameters ($\boldsymbol{V_j}$ and $\boldsymbol{W_j}$) we can utilize the Kalman Filter to calculate

---

Algorithm 1 Kalman Filter and Smoother

Kalman Filter

For j in $1, 2, ..., J$:

1. Predicted state: $\hat{\boldsymbol{\mu}}_{\boldsymbol{j}|\boldsymbol{j-1}} = \boldsymbol{G_j}\hat{\boldsymbol{\mu}}_{\boldsymbol{j-1}|\boldsymbol{j-1}}$
2. Predicted state covariance: $\boldsymbol{P_{j|j-1}} = \boldsymbol{G_j}\boldsymbol{P_{j-1|j-1}}\boldsymbol{G'_j} + W$
3. Innovation covariance: $\boldsymbol{S_j} = \boldsymbol{F_j}\boldsymbol{P_{j|j-1}}\boldsymbol{F'_j} + \boldsymbol{V}$
4. Kalman Gain: $\boldsymbol{K}_j = P_{j|j-1}F'_j S_j^{-1}$
5. Innovation: $\tilde{\boldsymbol{f}}_{\boldsymbol{j}} = \boldsymbol{y_j} - \boldsymbol{F_j}\hat{\mu}_{j|j-1}$
6. Updated state estimate: $\hat{\boldsymbol{\mu}}_{\boldsymbol{j}|\boldsymbol{j}} = \hat{\mu}_{j|j-1} + K_j\tilde{f}_j$
7. Updated state covariance: $\boldsymbol{P_{j|j}} = (I - K_j F_j)P_{j|j-1}$
8. Updated innovation: $\tilde{\boldsymbol{f}}_{j|j} = y_j - F_j\hat{\mu}_{j|j}$

Where $\hat{\boldsymbol{\mu}}_{\boldsymbol{0|0}} = \boldsymbol{u}$ and $\boldsymbol{P_{0|0}} = \boldsymbol{P}$.

Kalman Smoother

For j^* in J, J-1, ..., 1:

1. Smoothed predicted state: $\hat{\boldsymbol{\mu}}_{\boldsymbol{j^*}|\boldsymbol{J}} = \hat{\boldsymbol{\mu}}_{\boldsymbol{j^*}|\boldsymbol{j^*}} + \boldsymbol{L}^*_j(\hat{\boldsymbol{\mu}}_{(j^*+1)|J} - \hat{\boldsymbol{\mu}}_{(j^*+1)|j^*})$
2. Smoothed predicted state covariance: $\boldsymbol{P_{j^*|J}} = \boldsymbol{P_{j^*|j^*}} - \boldsymbol{L_{j^*}}\boldsymbol{G_{(j^*+1)}}\boldsymbol{P_{j^*|j^*}}$

Where $\boldsymbol{L_{j^*}} = \boldsymbol{P_{j^*|j^*}}\boldsymbol{G'_{(j^*+1)}} + \boldsymbol{P^{-1}_{(j^*+1)|j^*}}$.

$\hat{\boldsymbol{\mu}}_{\boldsymbol{j}|\boldsymbol{j}} = E(\boldsymbol{\mu_j}|\boldsymbol{y_1}, \boldsymbol{y_2}, ..., \boldsymbol{y_j})$ and $\boldsymbol{P_{j|j}} = \text{Var}(\boldsymbol{\mu_j}|\boldsymbol{y_1}, \boldsymbol{y_2}, ..., \boldsymbol{y_j})$ for j in $\{1, 2, ..., J\}$ (Durbin and Koopman 2012). The expected value and variance of the unobserved states conditioned on all the data can then be calculated by iterating backwards through the Kalman Smoother. The Kalman Filter and Smoother are maximum likelihood recursive algorithms that proceed as shown in algorithm 1.

Proper estimates for $\boldsymbol{\mu_j}$ depend on correctly specifying the variance parameters $\boldsymbol{V_j}$ and $\boldsymbol{W_j}$. When $\boldsymbol{V_j}$ and $\boldsymbol{W_j}$ are unknown, these parameters can be estimated by maximizing the joint log-likelihood using the limited memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) method Zhou and Li (2007).

$$\ell(V_j, W_j) = -\frac{np}{2}log(2\pi) - \frac{1}{2}\sum_{i=1}^{t}\left(log|\tilde{S}_i| + \tilde{f}_i\textrm{`}S_i^{-1}f_i\right)$$

To get a final estimate of the SSM model, 1.) run the Kalman Filter and Smoother, 2.) compute and maximize the log-likelihood to estimate variance parameters, 3.) iterate between 1.) and 2.) until convergence criteria is met.

Another important aspect is correctly specifying the initial parameters of $\boldsymbol{\mu_0}$ (mean $\boldsymbol{u}$ and variance $\boldsymbol{P}$), which are often unknown. However, by giving $\boldsymbol{\mu_0}$ an "infinite prior" of $\boldsymbol{P} = \infty$ and $\boldsymbol{u} = 0$, after only a few iterations of the Kalman Filter, our estimates $\boldsymbol{P_{i|j}}$ converge to the values they would have taken if we would have started with the true $\boldsymbol{P}$ and $\boldsymbol{u}$ (Durbin and Koopman 2012).

2.3.1. *Formulating Proposed Model Into State Space Form.* The proposed model in equation 1 can be rewritten to fit in the general matrix state space model framework from equation 2,

$$y_j = \begin{bmatrix} I_n \ X_j \end{bmatrix} \begin{bmatrix} \alpha_j \\ \beta_j \end{bmatrix} + \varepsilon_j$$

$$\begin{bmatrix} \alpha_j \\ \beta_j \end{bmatrix} = \begin{bmatrix} I_{(n+p)} \end{bmatrix} \begin{bmatrix} \alpha_{j-1} \\ \beta_{j-1} \end{bmatrix} + \begin{bmatrix} \eta_j \\ 0_{p\times 1} \end{bmatrix}$$

Where,

$$F_j = \begin{bmatrix} I_n \ X_j \end{bmatrix}, \ \ \mu_j = \begin{bmatrix} \alpha_j \\ \beta_j \end{bmatrix}, \ \ V_j = \sigma_\varepsilon^2 I_n, \ \ G_j = I_{n+p}, \ \ W_j = \sigma_\eta^2 \begin{bmatrix} diag(\delta_j) & 0 \\ 0 & 0 \end{bmatrix}$$

The Kalman Filter and Kalman Smoother along with the variance optimization as described in algorithm 1 may be calculated to get an estimate of $\mu_{j|J} = [\alpha'_{j|J} \ \beta'_{j|J}]'$. Recall, we assume that β_j remains constant over time, so all $\beta_{j|J}$ for $j \in \{1, 2, ..., J\}$ are equal. Our estimate for the linear effect of our predictors is $\hat{\beta} = \beta_{J|J}$ which has variance $P_{\hat{\beta}} = [P_{J|J}]_{(n+1:n+p),(n+1:n+p)}$. If modeling assumptions are met, $\hat{\beta} \sim N(B, P_{\hat{\beta}})$ which can be used for hypothesis testing (Durbin and Koopman 2012; Kalman 1960).

2.4. *Computational Considerations.* Although putting a diffuse prior on $\mu_0 = [\alpha_0' \ \beta_0']'$ results in theoretical niceties, it comes at the cost of computational burden. Each subject shares β, which is now assumed random as we let $\beta_0 \sim N(0, \infty)$, therefore the observations in the vector y_j are no longer treated as independent in the Kalman Filter and Kalman Smoother. This leads to the inverse of a possibly large non-sparse matrix $\text{Var}(Y_j|y_{1:j-1}) = S_j \in R^{n\times n}$ in the Kalman Filter process for $j \in \{1, 2, ..., J\}$. The S_j is also needed in the calculation of the log-likelihood of the variance parameters σ_ε^2 and σ_η^2, which makes it difficult to avoid the inversion at each iteration. This computational burden is then amplified as the Kalman Filter needs to be run multiple iterations for the maximum likelihood estimation calculation of σ_ε^2 and σ_η^2 to converge. In addition to considerable computation time, inverting a large matrix S_j can also lead to numerical inaccuracies. To overcome the issues of the full likelihood estimation process, we propose either partitioning or a Bayesian Gibb's Sampling approach.

2.5. *Partitioning.* To decrease the computational burden, we propose randomly and equally partitioning the n subjects into k groups then run the Kalman Filter and Smoother on each group independently. This will result in $\hat{\beta}^{(i)}$ and $P_{\hat{\beta}}^{(i)}$ for $i \in \{1, 2, ..., k\}$ independent groups. We then use $\bar{\beta} = k^{-1} \sum_{i=1}^{k} \hat{\beta}^{(i)}$ as our estimate for B. If modeling assumptions are met, $\bar{\beta} \sim N(B, k^{-2} \sum_{i=1}^{k} P_{\hat{\beta}^{(i)}})$ which can be used for hypothesis testing.

The partitioning method enforces a linear increase in the Kalman Filter computation time as n increases for a fixed partition group size.

2.6. *Bayesian Gibb's Sampling.* Computational issues arise in the full likelihood estimation because β is not treated as fixed in the Kalman Filter. This issue can be avoided by treating β as fixed when estimating α, then estimating β outside of the Kalman Filter using a Bayesian Gibb's Sampler. By following this methodology, the elements of y_j will be treated as

independent in the Kalman Filter and Kalman Smoother process. Using the Bayesian Gibb's Sampling framework, under regularity conditions, a sample from the joint distribution of the unknown parameters can be created. Inference can then be made using the samples of the respective parameters.

2.6.1. *The Bayesian LLT Model.* In this representation we no longer index the linear effect $\boldsymbol{\beta}$ with a time component.

$$y_{ij} = \alpha_{ij} + \boldsymbol{x}_{ij}\boldsymbol{\beta} + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$$

$$\alpha_{ij} = \alpha_{i(j-1)} + \eta_{ij}, \quad \eta_{ij} \sim N(0, \delta_{ij}\sigma_\eta^2)$$

We also set the following prior distributions for the unknown parameters,

$$\boldsymbol{\alpha_0} \sim N(\boldsymbol{u}, \boldsymbol{P_0})$$

$$\boldsymbol{\beta} \sim N(\boldsymbol{\theta}, \boldsymbol{I}_p\sigma_\beta^2)$$

$$\sigma_\eta^2 \sim IG(a_0/2, b_0/2)$$

$$\sigma_\varepsilon^2 \sim IG(c_0/2, d_0/2)$$

2.6.2. *Posterior Distributions.* The conditional distribution $\boldsymbol{\alpha_{1:n}}|\boldsymbol{y_{1:n}}, \boldsymbol{X_{1:n}}, \boldsymbol{\beta}, \sigma_\eta^2, \sigma_\varepsilon^2$ can be estimated directly from the the Kalman Filter. By conditioning on $\boldsymbol{x}_{ij}$ and $\boldsymbol{\beta}$, each y_{ij} is independent for $i \in \{1, 2, ..., n\}$ and fixed j. Thus, we can run the Kalman Filter with $\tilde{y}_{ij} = y_{ij} - \boldsymbol{x}_{ij}\boldsymbol{\beta}$ as the outcome in the model,

$$\tilde{y}_{ij} = \alpha_{ij} + \varepsilon_{ij}$$

$$\alpha_{ij} = \alpha_{i(j-1)} + \eta_{ij}$$

Because each element of $\tilde{\boldsymbol{y}}_{\boldsymbol{j}}$ are treated as independent, it is equivalent to estimating each element $\boldsymbol{\alpha}_j$ independently for each subject, resulting in a computational efficient Kalman Filter and Smoother found in algorithm 2.

Let $\psi = \{\sigma_\eta^2, \sigma_\varepsilon^2, \boldsymbol{\beta}\}$, the vector of the other unknown parameters. For the posterior of $\boldsymbol{\alpha}$ we seek to find the distribution defined by $P_\psi(\boldsymbol{\alpha}_{0:J}|\boldsymbol{y}_{1:J})$. Which, because of the Markov structure, can be defined as,

$$P_\psi(\boldsymbol{\alpha}_{0:J}|\boldsymbol{y}_{1:J}) = P_\psi(\boldsymbol{\alpha}_J|\boldsymbol{y}_{1:J})P_\psi(\boldsymbol{\alpha}_{J-1}|\boldsymbol{\alpha}_J, \boldsymbol{y}_{1:(J-1)})...P_\psi(\boldsymbol{\alpha}_0|\boldsymbol{\alpha}_1)$$

Therefore we need the following densities for j in $1, 2, ..., J-1$:

$$P_\psi(\boldsymbol{\alpha}_j|\boldsymbol{\alpha}_{j+1}, \boldsymbol{y}_{1:j}) \propto P_\psi(\boldsymbol{\alpha}_j|\boldsymbol{y}_{1:j})P_\psi(\boldsymbol{\alpha}_{j+1}|\boldsymbol{\alpha}_j)$$

From the Kalman Filter we calculated $\boldsymbol{\alpha}_j|\boldsymbol{y}_{1:j} \sim N_\psi(\boldsymbol{\alpha}_{j|j}, \boldsymbol{P}_{j|j})$ and $\boldsymbol{\alpha}_{j+1}|\boldsymbol{\alpha}_j \sim N_\psi(\boldsymbol{\alpha}_j, \sigma_\eta^2)$. After combining the two densities $\boldsymbol{m}_j = E_\psi(\boldsymbol{\alpha}_j|\boldsymbol{\alpha}_{j+1}, \boldsymbol{y}_{1:j}) = \boldsymbol{\alpha}_{j|j} + \boldsymbol{L}_j(\boldsymbol{\alpha}_{j+1} - \boldsymbol{\alpha}_{j+1|j})$ and $\boldsymbol{R}_j = \text{Var}_\psi(\boldsymbol{\alpha}_j|\boldsymbol{\alpha}_{j+1}, \boldsymbol{y}_{1:j}) = \boldsymbol{P}_{j|j} - \boldsymbol{L}_j^2\boldsymbol{P}_{j+1|j}$ (Shumway, Rober H. and Stoffer, David S. 2017). Because of normality, the posterior distribution for $\boldsymbol{\alpha}_j$ is $N(\boldsymbol{m}_j, \boldsymbol{R}_j)$.

For the backward sampling procedure we start by sampling a $\boldsymbol{\alpha}_J^*$ from a $N_\psi(m_J, \boldsymbol{R}_J)$, then setting $\boldsymbol{\alpha}_J^* = \boldsymbol{\alpha}_J$ for the calculation of $\boldsymbol{m}_{T-1}$ to then sample $\boldsymbol{\alpha}_{T-1}^*$ from a $N_\psi(\boldsymbol{m}_{T-1}, \boldsymbol{R}_{T-1})$. This process continues until a whole chain $\boldsymbol{\alpha}_{0:T}^*$ has been sampled.

Algorithm 2 Kalman Filter and Smoother for Bayesian Estimation

---

Kalman Filter

For j in $1, 2, ..., J$:

1. Predicted state: $\boldsymbol{\alpha}_{j|j-1} = \boldsymbol{\alpha}_{j-1|j-1}$
2. Predicted state variance: $\boldsymbol{P}_{j|j-1} = \boldsymbol{P}_{j-1|j-1} + \sigma_\eta^2$
3. Kalman Gain: $\boldsymbol{K}_j = \dfrac{\boldsymbol{P}_{j|j-1}}{\boldsymbol{P}_{j|j-1} + \sigma_\varepsilon^2}$
4. Updated state estimate: $\boldsymbol{\alpha}_{j|j} = \boldsymbol{\alpha}_{j|j-1} + \boldsymbol{K}_j(\tilde{\boldsymbol{y}}_j - \boldsymbol{\alpha}_{j|j-1})$
5. Updated state covariance: $\boldsymbol{P}_{j|j} = (1 - \boldsymbol{K}_j)\boldsymbol{P}_{j|j-1}$

Kalman Smoother

For j^* in J, J-1, ..., 1:

1. Smoothed predicted state: $\boldsymbol{\alpha}_{J|j^*-1} = \boldsymbol{\alpha}_{j^*-1|j^*-1} + \boldsymbol{L}_{j^*-1}(\boldsymbol{\alpha}_{j^*|J} - \boldsymbol{\alpha}_{j^*|j-1})$
2. Smoothed predicted state variance: $\boldsymbol{P}_{j^*-1|J} = \boldsymbol{P}_{j^*-1|j^*-1} - \boldsymbol{L}_{j^*-1}^2(\boldsymbol{P}_{j^*|J} - \boldsymbol{P}_{j^*|j^*-1})$

Where $\boldsymbol{L}_{j^*-1} = \boldsymbol{P}_{j^*-1|j^*-1}\boldsymbol{P}_{j^*|j^*-1}^{-1}$.

NOTE: Although we are using vectors, because of the independence each operation is done element-wise.

The posterior distribution $\boldsymbol{\beta}$ follows a pattern more similar to multivariate Bayesian regression. With all the other parameters fixed,

$$P(\boldsymbol{\beta}|Y, \boldsymbol{\alpha}, \sigma_\eta^2, \sigma_\varepsilon^2) = \frac{P(Y, \boldsymbol{\alpha}, \sigma_\eta^2, \sigma_\varepsilon^2|\boldsymbol{\beta})P(\boldsymbol{\beta})}{P(Y, \boldsymbol{\alpha}, \sigma_\eta^2, \sigma_\varepsilon^2)}$$

Because $P(Y, \boldsymbol{\alpha}, \sigma_\eta^2, \sigma_\varepsilon^2)$ is constant with respect to $\boldsymbol{\beta}$, and $P(\boldsymbol{\beta})$ is already defined, we focus our attention on $P(Y, \boldsymbol{\alpha}, \sigma_\eta^2, \sigma_\varepsilon^2|\boldsymbol{\beta})$.

$$
\begin{aligned}
P(Y, \boldsymbol{\alpha}, \sigma_\eta^2, \sigma_\varepsilon^2|\boldsymbol{\beta}) =& P(\boldsymbol{y}_1, ..., \boldsymbol{y}_J, \boldsymbol{\alpha}_1, ..., \boldsymbol{\alpha}_J, \sigma_\eta^2, \sigma_\varepsilon^2|\boldsymbol{\beta}) \\
=& P(\boldsymbol{y}_J|\boldsymbol{y}_1, ..., \boldsymbol{y}_{J-1}, \boldsymbol{\alpha}_1, ..., \boldsymbol{\alpha}_J, \sigma_\eta^2, \sigma_\varepsilon^2\boldsymbol{\beta}) \\
& \times P(\boldsymbol{y}_1, ..., \boldsymbol{y}_{J-1}, \boldsymbol{\alpha}_1, ..., \boldsymbol{\alpha}_J, \sigma_\eta^2, \sigma_\varepsilon^2|\boldsymbol{\beta}) \\
=& P(\boldsymbol{y}_J|\boldsymbol{\alpha}_J, \sigma_\varepsilon^2\boldsymbol{\beta})P(\boldsymbol{\alpha}_{J-1}|\boldsymbol{y}_1, ..., \boldsymbol{y}_{J-1}, \boldsymbol{\alpha}_1, ..., \boldsymbol{\alpha}_{J-1}, \sigma_\eta^2, \sigma_\varepsilon^2, \boldsymbol{\beta}) \\
& \times P(\boldsymbol{y}_1, ..., \boldsymbol{y}_{J-1}, \boldsymbol{\alpha}_1, ..., \boldsymbol{\alpha}_{J-1}, \sigma_\eta^2, \sigma_\varepsilon^2|\boldsymbol{\beta}) \\
=& P(\boldsymbol{y}_J|\boldsymbol{\alpha}_J, \sigma_\varepsilon^2\boldsymbol{\beta})P(\boldsymbol{\alpha}_J|\boldsymbol{\alpha}_{J-1}, \sigma_\eta^2) \\
& \times P(\boldsymbol{y}_1, ..., \boldsymbol{y}_{J-1}, \boldsymbol{\alpha}_1, ..., \boldsymbol{\alpha}_{J-1}, \sigma_\eta^2, \sigma_\varepsilon^2|\boldsymbol{\beta}) \\
=& P(\sigma_\varepsilon^2)P(\sigma_\eta^2)\left(\prod_{k=0}^{J} P(\boldsymbol{\alpha}_k|\boldsymbol{\alpha}_{k-1}, \sigma_\eta^2)\right)\prod_{j=1}^{J} P(\boldsymbol{y}_j|\boldsymbol{\alpha}_j, \sigma_\varepsilon^2, \boldsymbol{\beta}) \\
\propto& \prod_{j=1}^{J} P(\boldsymbol{y}_j|\boldsymbol{\alpha}_j, \sigma_\varepsilon^2, \boldsymbol{\beta})
\end{aligned}
$$

For simplicity we can further write,

$$-2logP(Y, \boldsymbol{\alpha}, \sigma_\eta^2, \sigma_\varepsilon^2 | \boldsymbol{\beta}) \propto \sum_{j=1}^{J} \frac{(\boldsymbol{y}_j - \boldsymbol{\alpha}_j - \boldsymbol{X}_j \boldsymbol{\beta})'(\boldsymbol{y}_j - \boldsymbol{\alpha}_j - \boldsymbol{X}_j \boldsymbol{\beta})}{\sigma_\varepsilon^2}$$

$$\propto \sum_{j=1}^{J} \frac{-2\boldsymbol{y}_j' \boldsymbol{X}_j \boldsymbol{\beta} + 2\boldsymbol{\alpha}_j' \boldsymbol{X}_j \boldsymbol{\beta} + \boldsymbol{\beta}' \boldsymbol{X}_j' \boldsymbol{X}_j \boldsymbol{\beta}}{\sigma_\varepsilon^2}$$

$$\propto \frac{\boldsymbol{\beta}' \left(\sum_{j=1}^{J} \boldsymbol{X}_j' \boldsymbol{X}_j \right) \boldsymbol{\beta} - 2 \left(\sum_{j=1}^{J} \boldsymbol{y}_j - \boldsymbol{\alpha}_j \right)' \boldsymbol{X}_j \boldsymbol{\beta}}{\sigma_\varepsilon^2}$$

We can then find the proportionality of the prior of $\boldsymbol{\beta}$,

$$-2logP(\boldsymbol{\beta}) \propto \frac{(\boldsymbol{\beta} - \boldsymbol{\theta})'(\boldsymbol{\beta} - \boldsymbol{\theta})}{\sigma_\beta^2}$$

$$\propto \frac{\boldsymbol{\beta}' \boldsymbol{\beta} - 2\boldsymbol{\theta}\boldsymbol{\beta}}{\sigma_\beta^2}$$

Thus the -2log posterior of $\boldsymbol{\beta}$ is proportional to,

$$-2logP(\boldsymbol{\beta}|Y, \boldsymbol{\alpha}, \sigma_\eta^2, \sigma_\varepsilon^2) \propto \frac{\boldsymbol{\beta}' \left(\sum_{j=1}^{J} \boldsymbol{X}_j' \boldsymbol{X}_j \right) \boldsymbol{\beta} - 2 \left(\sum_{j=1}^{J} \boldsymbol{y}_j - \boldsymbol{\alpha}_j \right)' \boldsymbol{X}_j \boldsymbol{\beta}}{\sigma_\varepsilon^2} + \frac{\boldsymbol{\beta}' \boldsymbol{\beta} - 2\boldsymbol{\theta}\boldsymbol{\beta}}{\sigma_\beta^2}$$

$$\propto \frac{\boldsymbol{\beta}' \left(\sigma_\beta^2 \sum_{j=1}^{J} \boldsymbol{X}_j' \boldsymbol{X}_j \right) \boldsymbol{\beta} - 2\sigma_\beta^2 \left(\sum_{j=1}^{J} \boldsymbol{y}_j - \boldsymbol{\alpha}_j \right)' \boldsymbol{X}_j \boldsymbol{\beta} + \boldsymbol{\beta}' \sigma_\varepsilon^2 \boldsymbol{I}_p \boldsymbol{\beta} - 2\sigma_\varepsilon^2 \boldsymbol{\theta}\boldsymbol{\beta}}{\sigma_\varepsilon^2 \sigma_\beta^2}$$

$$\propto \frac{\boldsymbol{\beta}' \left(\sigma_\beta^2 \sum_{j=1}^{J} \boldsymbol{X}_j' \boldsymbol{X}_j \right) \boldsymbol{\beta} + \boldsymbol{\beta}' \sigma_\varepsilon^2 \boldsymbol{I}_p \boldsymbol{\beta} - 2\sigma_\beta^2 \left(\sum_{j=1}^{J} \boldsymbol{y}_j - \boldsymbol{\alpha}_j \right)' \boldsymbol{X}_j \boldsymbol{\beta} - 2\sigma_\varepsilon^2 \boldsymbol{\theta}\boldsymbol{\beta}}{\sigma_\varepsilon^2 \sigma_\beta^2}$$

$$\propto \frac{\boldsymbol{\beta}' \left(\left(\sigma_\beta^2 \sum_{j=1}^{J} \boldsymbol{X}_j' \boldsymbol{X}_j \right) + \sigma_\varepsilon^2 \boldsymbol{I}_p \right) \boldsymbol{\beta} - 2 \left(\sigma_\beta^2 \left(\sum_{j=1}^{J} \boldsymbol{y}_j - \boldsymbol{\alpha}_j \right)' \boldsymbol{X}_j - \sigma_\varepsilon^2 \boldsymbol{\theta}' \right) \boldsymbol{\beta}}{\sigma_\varepsilon^2 \sigma_\beta^2}$$

$$\propto \frac{(\boldsymbol{\beta} - \boldsymbol{\Sigma}^{-1} B)' \boldsymbol{\Sigma} (\boldsymbol{\beta} - \boldsymbol{\Sigma}^{-1} B)}{\sigma_\varepsilon^2 \sigma_\beta^2}$$

Where $B = \sigma_\beta^2 \left(\sum_{j=1}^{J} \boldsymbol{y}_j - \boldsymbol{\alpha}_j \right)' \boldsymbol{X}_j - \sigma_\varepsilon^2 \boldsymbol{\theta}'$ and $\boldsymbol{\Sigma} = \left(\sigma_\beta^2 \sum_{j=1}^{J} \boldsymbol{X}_j' \boldsymbol{X}_j \right) + \sigma_\varepsilon^2 \boldsymbol{I}_p$. Therefore, $\boldsymbol{\beta}|Y, \boldsymbol{\alpha}, \sigma_\eta^2, \sigma_\varepsilon^2 \sim N(\boldsymbol{\Sigma}^{-1} B, \sigma_\varepsilon^2 \sigma_\beta^2 \boldsymbol{\Sigma}^{-1})$.

Note, $\boldsymbol{\Sigma}$ is a $p \times p$ matrix that needs to be inverted. If p is large, this can greatly slow down the Gibb's Sampler, especially when considering we may do several thousand iterations. However, $\boldsymbol{\Sigma}$ can be broken down to increase computation speed. Recall, $\boldsymbol{\Sigma} = \left(\left(\sigma_\beta^2 \sum_{j=1}^{J} \boldsymbol{X}_j' \boldsymbol{X}_j \right) + \sigma_\varepsilon^2 \boldsymbol{I}_p \right)$. The term $\left(\sigma_\beta^2 \sum_{j=1}^{J} \boldsymbol{X}_j' \boldsymbol{X}_j \right)$ will not change at each iteration because it does not contain unknown parameters. By calculating the eigenvalue decomposition on $\sigma_\beta^2 \sum_{j=1}^{J} \boldsymbol{X}_j' \boldsymbol{X}_j$ we can rewrite $\boldsymbol{\Sigma}$ as follows,

$$\left(\left(\sigma_\beta^2 \sum_{j=1}^{J} \boldsymbol{X}_j' \boldsymbol{X}_j \right) + \sigma_\varepsilon^2 \boldsymbol{I} \right) = (\boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}' + \sigma_\varepsilon^2 \boldsymbol{I})$$

$$= (\boldsymbol{Q\Lambda Q'} + \sigma_\varepsilon^2 \boldsymbol{QQ'})$$

$$= \boldsymbol{Q}(\boldsymbol{\Lambda} + \sigma_\varepsilon^2 \boldsymbol{I})\boldsymbol{Q'}$$

then,

$$((\sigma_\beta^2 \sum_{j=1}^J \boldsymbol{X}_j'\boldsymbol{X}_j) + \sigma_\varepsilon^2 \boldsymbol{I})^{-1} = \boldsymbol{Q}\big(1/(\boldsymbol{\Lambda} + \sigma_\varepsilon^2 \boldsymbol{I})\big)\boldsymbol{Q'}$$

We only need to calculate the eigen vectors $\boldsymbol{Q}$ and eigen values $\boldsymbol{\Lambda}$ of $\sigma_\beta^2 \sum_{j=1}^J \boldsymbol{X}_j'\boldsymbol{X}_j$ once, then simply update σ_ε^2 before calculating the inverse.

For the variance parameter σ_η^2, we apply the same Bayes' rule rationale,

$$P(\sigma_\eta^2|Y,\boldsymbol{\beta},\boldsymbol{\alpha},\sigma_\varepsilon^2) = \frac{P(Y,\boldsymbol{\beta},\boldsymbol{\alpha},\sigma_\varepsilon^2|\sigma_\eta^2)P(\sigma_\eta^2)}{P(Y,\boldsymbol{\beta},\boldsymbol{\alpha},\sigma_\varepsilon^2)}$$

$$\propto P(Y,\boldsymbol{\beta},\boldsymbol{\alpha},\sigma_\varepsilon^2|\sigma_\eta^2)P(\sigma_\eta^2)$$

$$\propto (\delta_j\sigma_\eta^2)^{-nJ/2} e^{\sum_{j=1}^J (\boldsymbol{\alpha}_j - \boldsymbol{\alpha}_{j-1})^2/2\delta_j\sigma_\eta^2} (\sigma_\eta^2)^{-a_0/2-1} e^{-b_0/2\sigma_\eta^2}$$

$$\propto (\sigma_\eta^2)^{-(nJ+a_0)/2-1} e^{(\sum_{j=1}^J (\boldsymbol{\alpha}_j - \boldsymbol{\alpha}_{j-1})^2/\delta_j + b_0)/2\sigma_\eta^2}$$

Therefore, $\sigma_\eta^2|Y,\boldsymbol{\beta},\boldsymbol{\alpha},\sigma_\varepsilon^2 \sim IG(\frac{nJ+a_0}{2}, \frac{\sum_{j=1}^J (\boldsymbol{\alpha}_j - \boldsymbol{\alpha}_{j-1})^2/\delta_j + b_0}{2})$.

In a very similar fashion we can show $\sigma_\varepsilon^2|Y,\boldsymbol{\beta},\boldsymbol{\alpha},\sigma_\eta^2 \sim IG(\frac{nJ+c_0}{2}, \frac{d_0 + \sum_{j=1}^J (\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\beta} - \boldsymbol{\alpha}_j)^2}{2})$.

Using the posterior distributions, the Gibb's sampling can be carried out as shown in algorithm 3. After removing the initial burn-in samples, to allow for posterior sample convergence, the rest of the samples can be used to make inference on any parameter of interest. We use the mean of the samples of $\boldsymbol{\beta}$ as a point estimate for $\boldsymbol{\beta}$ and take the 2.5 and 97.5 percentiles of the posterior draws as a pseudo 95% confidence interval.

Algorithm 3 Gibb's Sampling Algorithm
___

1. Select prior parameters $\boldsymbol{\theta}, \sigma_\beta^2, a_0, b_0, c_0, d_0$.

2. Let $\boldsymbol{\beta}^{(0)} = \boldsymbol{\theta}$, $\sigma_\eta^{2(0)} = \frac{d_0/2}{1+c_0/2}$, and $\sigma_\varepsilon^{2(0)} = \frac{b_0/2}{1+a_0/2}$.

3. Repeat steps 4-7 for $i \in \{1, 2, ..., M\}$

4. Run a forward-filtering backward sampling procedure, conditioning on $\boldsymbol{\beta} = \boldsymbol{\beta}^{i-1}, \sigma_\eta^2 = \sigma_\eta^{2(i-1)}, \sigma_\varepsilon^2 = \sigma_\varepsilon^{2(i-1)}$ to get samples $\boldsymbol{\alpha}^*$, then set $\boldsymbol{\alpha}^{(i)} = \boldsymbol{\alpha}^*$.

5. Sample σ_η^{2*} from $IG(\frac{nJ+a_0}{2}, \frac{\sum_{j=1}^J (\boldsymbol{\alpha}_j^{(i)} - \boldsymbol{\alpha}_{j-1}^{(i)})^2 + b_0}{2})$ and set $\sigma_\eta^{2(i)} = \sigma_\eta^{2*}$.

6. Sample σ_ε^{2*} from $IG(\frac{nJ+c_0}{2}, \frac{d_0 + \sum_{j=1}^J (\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\beta}^{(i-1)} - \boldsymbol{\alpha}_j^{(i)})^2}{2})$ and set $\sigma_\varepsilon^{2(i)} = \sigma_\varepsilon^{2*}$.

7. Sample $\boldsymbol{\beta}^*$ from $N(\boldsymbol{\Sigma}^{-1}B, \sigma_\varepsilon^2 \sigma_\beta^2 \boldsymbol{\Sigma}^{-1})$ where $\boldsymbol{\alpha} = \boldsymbol{\alpha}^{(i)}, \sigma_\eta^2 = \sigma_\eta^{2(i)}, \sigma_\varepsilon^2 = \sigma_\varepsilon^{2(i)}$ and set $\boldsymbol{\beta}^{(i)} = \boldsymbol{\beta}^*$.

After steps 1-7 are completed posterior samples can be used as an empirical distribution to make parameter inference.
___

3. Model Validation.

3.1. *Simulation Controlling Underlying Data Generation Process.* The fully simulated data analysis allows for insight into model behavior under different levels of model misspecification as well as confirming model accuracy. For this study we compare the proposed LLT methods to an LMEM with a random intercept and a random intercept LMEM with an AR(1) auto-correlation structure. Data for this simulation emulates the model of interest (section 2.1.2), using a simulated Animals test score as the outcome and simulated independent variables time, sex, education, race, age, transition to MCI or dementia, APOE status, and an APOE status by sex interaction. For each simulation, 100 participants are created. Each subject has 2 to 10 cognitive scores at possible observation times 0 through 9. The "true" linear effects are $\boldsymbol{\beta} = [\text{-}0.343, \text{-}0.109, 0.006, 0.247, \text{-}0.031, \text{-}1.025, \text{-}0.132, 0.038]^T$. The values for $\boldsymbol{\beta}$ in the simulations are the same as the model of interest (section 2.1.2) linear effect estimates from a multiple linear regression using the NACC data set. Under each simulation scenario the linear effects are estimated using an LMEM with random intercept, an LMEM with an AR(1) auto-correlation structure, the full likelihood LLT (FL LLT), partitioned LLT with 2 groups (Part 2), partitioned LLT with 4 groups (Part 4), partitioned LLT with 10 groups (Part 10), and the Bayesian LLT (BLLT).

When the underlying data generating process follows that of the LLT it has the form,

$$y_{ij} = \alpha_{ij} + \boldsymbol{x}_{ij}\boldsymbol{\beta} + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2 \boldsymbol{I}_n)$$
$$\alpha_{ij} = \alpha_{i(j-1)} + \eta_{ij}, \quad \eta_{ij} \sim N(0, \sigma_\eta^2 \boldsymbol{I}_n)$$

for $i \in \{1, 2, ..., 100\}$. Multiple simulation scenarios are created by altering σ_ε^2 and σ_η^2 which will either increase or decrease the amount of auto-correlation in the data. For the LLT data generation, the variance parameters $(\sigma_\varepsilon^2, \sigma_\eta^2)$ take on the values (3, 0), (3, 1), (3, 2), (3, 3), (30, 10), and (60, 20).

When the underlying data generation process is an LMEM with an AR(1) variance structure it has the form,

$$y_{ij} = b_0 + \boldsymbol{x}_{ij}\boldsymbol{\beta} + \epsilon_{ij}, \quad b_0 \sim N(0, \boldsymbol{I}_n)$$
$$\epsilon_{ij} = \rho \, \epsilon_{i(j-1)} + e_{ij}, \quad e_{ij} \sim N(0, \boldsymbol{I}_n)$$

Under this scheme there are three scenarios of varying $\rho = 0, 0.1,$ and 0.5.

Each simulation scenario is repeated 1000 times. The main metrics of interest are 95% confidence interval coverage, bias evaluation, and confidence interval length. In order to make proper inference, the 95% confidence intervals should cover the parameter of interest approximately 95% of the time. Proper coverage indicates proper probability of type I error, or probability of rejecting a null hypothesis given the null hypothesis is true. Once type I error is established, we typically wish to minimize the probability of type II error, or the probability of not rejecting a null hypothesis given the null hypothesis is false. To minimize probability of type II error we desire low parameter variance which corresponds to shorter confidence interval length. If type I error is maintained for an unbiased estimator, than a smaller confidence interval leads to higher power. Note, confidence interval length was chosen over the more common parameter variance because the Bayesian LLT uses an empirical pseudo 95% confidence interval from the parameter posterior draws rather than the traditional Frequentist variance estimation.

The Bayesian LLT uses the following prior distributions,

$$\alpha_0 \sim N(0, 10), \qquad \boldsymbol{\beta} \sim N(0, 10)$$
$$\sigma_\varepsilon^2 \sim IG(0.005, 0.005), \quad \sigma_\eta^2 \sim IG((0.005, (0.005))$$

After 2,000 posterior draws for the unknown parameters, samples 1 through 1,000 are discarded as a "burn-in" sample and samples 1,001 through 2000 are used to make parameter inference.

3.1.1. *Fully Simulated Data Results.* The LME and LMEM with an AR(1) temporal correlation structure both maintain 95% parameter coverage when the model is correctly specified. However, as the variance of the latent state (σ_η^2) increases there is evident under coverage. The LLTs maintain near 95% coverage even when the underlying data generating process is that of an AR(1) with $\rho = 0.5$, showing the LLT models are more robust to model misspecification (see table 1).

All methods show unbiasedness under each simulation scenario (see figure 1). Of the state space models the Bayesian estimation process has the smallest variability in effect estimates and closely rivals that of the LMEMs. The partitioned state space model with 10 groups shows high estimate volatility, however, when there are 2 or 4 groups the estimate resemble the full data model. If the number of estimated parameters is reasonable compared to the group size, then partitioning will decrease computation time without sacrificing much in terms of estimation.

Among the LLT methods, the Bayesian LLT estimation also maintains the smallest confidence interval length for the linear effect parameters, which translates to higher power (see figure 1). The LMEM methods have an even smaller confidence interval length, but because the LMEM methods do not have the correct 0.05 probability of type I error the comparison is inconsequential.

Table 1: *Combined linear effect coverage proportion. The LMEM methods fail to maintain 95% coverage when the true data generation process is that of an LLT. The LLT estimation procedures, however, maintain near 95% coverage even under model misspecification.*

Simulation Model	Variance Parameters		Traditional Methods		State Space Methods				
			LME	AR(1)	FL LLT	Part 2	Part 4	Part 10	BLLT
AR(1)	$\sigma^2 = 1$	$\rho = 0$	0.954	0.953	0.947	0.952	0.963	0.983	0.964
AR(1)	$\sigma^2 = 1$	$\rho = 0.1$	0.938	0.947	0.945	0.953	0.963	0.981	0.955
AR(1)	$\sigma^2 = 1$	$\rho = 0.5$	0.889	0.944	0.965	0.967	0.977	0.987	0.962
LLT	$\sigma_\varepsilon^2 = 3$	$\sigma_\eta^2 = 0$	0.941	0.940	0.944	0.965	0.969	0.978	0.958
LLT	$\sigma_\varepsilon^2 = 3$	$\sigma_\eta^2 = 1$	0.790	0.847	0.947	0.948	0.956	0.972	0.940
LLT	$\sigma_\varepsilon^2 = 3$	$\sigma_\eta^2 = 2$	0.736	0.836	0.948	0.944	0.950	0.970	0.932
LLT	$\sigma_\varepsilon^2 = 3$	$\sigma_\eta^2 = 3$	0.715	0.838	0.946	0.940	0.945	0.969	0.928
LLT	$\sigma_\varepsilon^2 = 30$	$\sigma_\eta^2 = 10$	0.781	0.840	0.947	0.941	0.945	0.970	0.946
LLT	$\sigma_\varepsilon^2 = 60$	$\sigma_\eta^2 = 20$	0.780	0.840	0.946	0.944	0.945	0.967	0.943

FIG 1. *Bias and confidence interval length with imperical 95% confidence interval. All estimation methods show unbiasedness. Of the LLT procedures, the Bayesian LLT has the least amount of variablity in the estimates. The LMEM methods have the smallest confidence interval length, but do not maintain 95% coverage. Among the methods that do cover at 95% (the LLTs), the Bayesian LLT has the smallest confidence interval length across the different scenarios.*
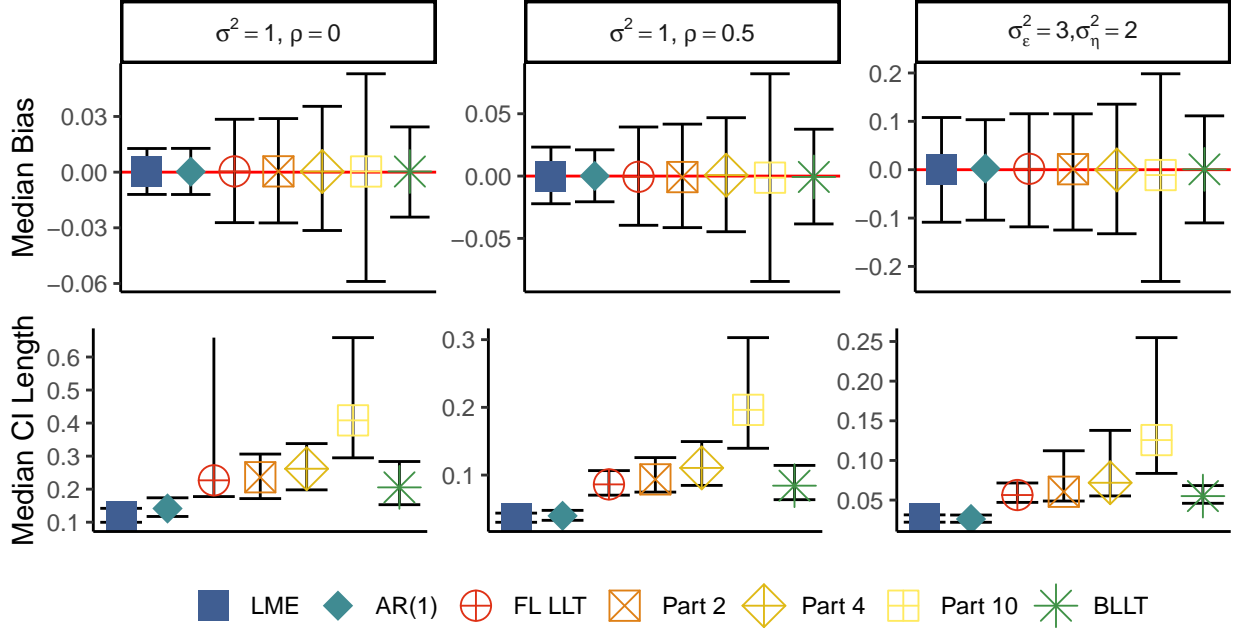
3.2. *Real Data Simulation.* Although assessing model accuracy under correct and incorrect model specification is important, of more importance is evaluation of the models when the underlying data generation of the neuropsychological outcome is unknown. To compare the proposed LLT models and commonly used LMEMs we conduct a simulation study using the real NACC data. For each of the 1,000 simulations, half of the NACC participants are randomly selected and given a linear group effect to the existing Animals test outcome. The LLT and LME models are then used to estimate the original model of interest (section 2.1.2) with the additional simulated group effect. If the unknown temporal covariance structure is correctly specified by the model, we expect to accurately estimate this simulated group effect by showing unbiasedness and proper 95% coverage. The models used to estimate the simulated group effect are a linear mixed effect model with a random intercept, a linear mixed effect model with a random intercept and AR(1) auto-correlation, partitioned LLT model with k = 50 groups, and a Bayesian LLT model.

For the NACC simulation, the Bayesian LLT model unknown parameters have the same prior distribution as is done in the full data simulation (section 3.1). We use the same 1,000 sample "burn-in" and 1,000 samples for inference as well.

3.2.1. *Real Data Simulation Results.* When estimating the randomly prescribed group effect on the Animals outcome, the partitioned LLT and Bayesian LLT methods maintain unbiasedness and proper 95% confidence interval coverage (93.5% and 94.0% respectively). Both linear mixed effect models, while unbiased (see figure 2), did not maintain proper type I error of 0.05 (see table 2). The random intercept LMEM has poor coverage at 79.5% and the random intercept LMEM with an AR(1) temporal variance structure has coverage of 88.9%.

Table 2: *Simulated effect parameter coverage proportion. The LMEM methods do not maintain 95% coverage, while the LLT estimation procedures do.*

LMEM	LMEM AR(1)	Partition LLT	Bayesian LLT
0.795	0.889	0.935	0.940

FIG 2. Simulated effect parameter bias with empirical 95% confidence interval. All estimation methods are unbiased. The Bayesian LLT has less variability in the estimates when compard to the Partition LLT.
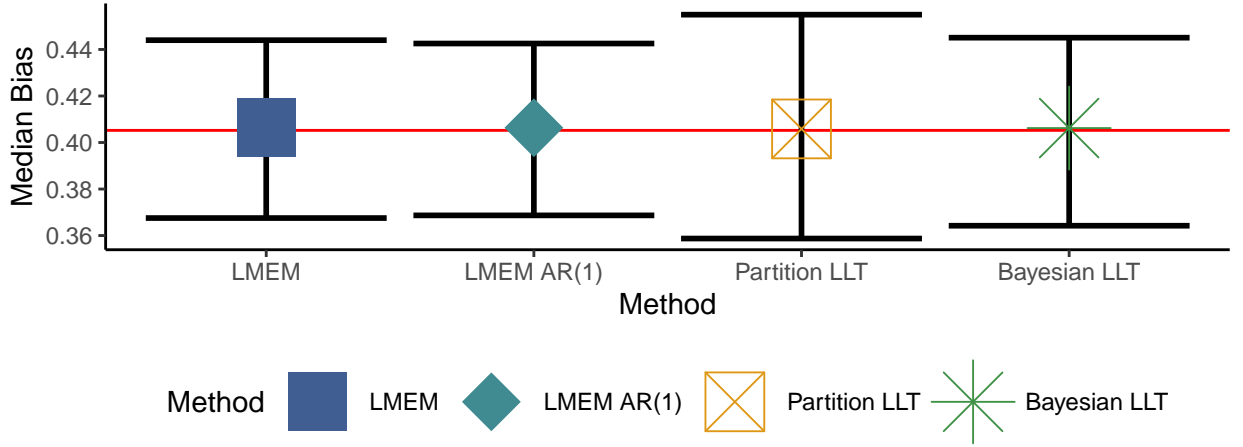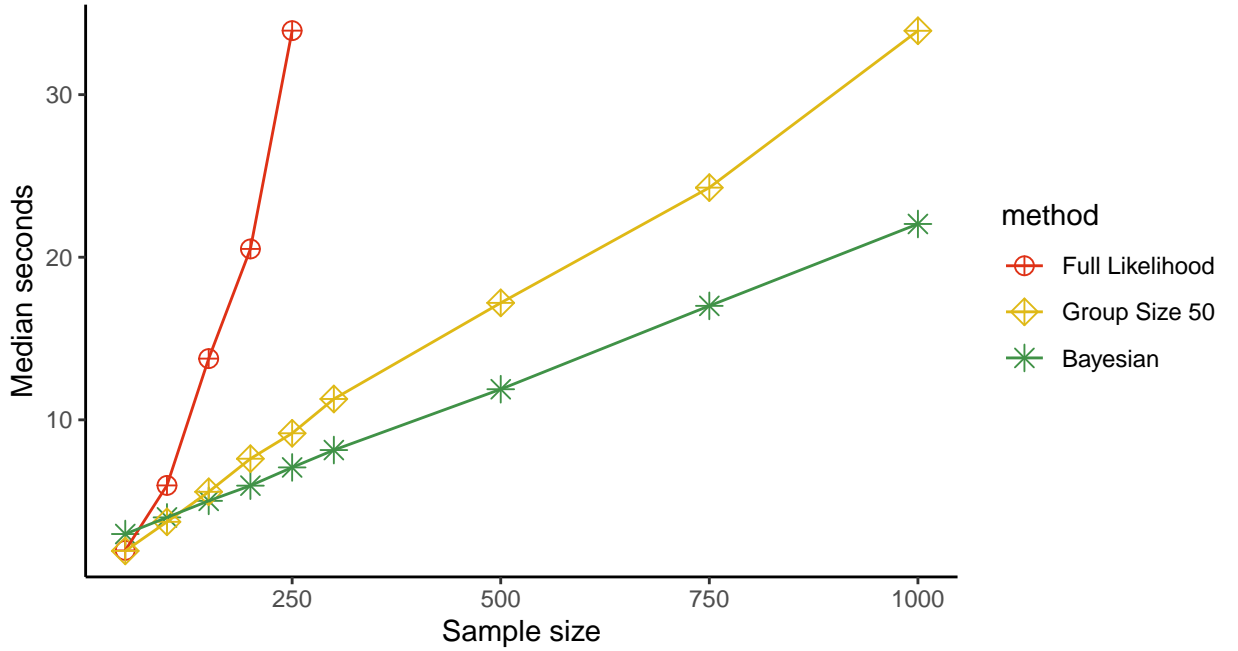

The LMEM with the AR(1) variance structure did better than the standard LMEM on coverage, but the probability of type I error is estimated as 0.11 which is more than double the prescribed 0.05.

The coverage results from the real data simulation are consistent with results found in the full data simulation. The variance parameter estimates on the NACC data set for the Bayesian LLT are $\hat{\sigma}_\varepsilon^2 = 7.86$ and $\hat{\sigma}_\eta^2 = 2.19$ which is a 3.59 to 1 ratio. The fully simulated data coverage for the standard LMEM, the LMEM with an AR(1), and the Bayesian LLT under the scenario $\sigma_\varepsilon^2 = 3, \sigma_\eta^2 = 1$ are 79.0%, 84.7%, and 94.0% respectively (see section 3.1.1).

3.3. *Computation Time Comparison.* To compare the computation time of the full likelihood, partitioned LLT, and Bayesian LLT we simulate data with two continuous predictors, 5 repeated measurements, and a sample size of 50, 100, 200, 250, 300, 500, 750, and 1000 subjects. Estimation is carried out 1000 times for each scenario and the computation time is tracked for the 3 different model estimation processes. For the partitioned LLT we set the group size equal to 50 i.e. for the sample size of 250, we will have 5 groups, each with size of 50. The partition of group size 50 was chosen as it showed the fastest computation time among other group sizes. The Bayesian LLT estimation used 2,000 Gibb's sampling iterations.

3.3.1. *Computation Time Results.* The full likelihood LLT estimation procedure yields an exponential increase in computation time as the sample size increases. This is due to repeatedly inverting large matrices in the Kalman Filter and Kalman Smoother algorithm. Sample sizes greater than 250 resulted in estimation failures for the full likelihood, and for this reason we only show up to n = 250 in figure 3.

Also shown in figure 3, the partitioned LLT and the Bayesian LLT estimation procedures show a linear increase in computation time as the sample size increases. At the sample size of 50, the Bayesian LLT estimation has the longest median computation time (2.98 seconds) when compared to the full likelihood LLT (1.97 seconds) and the partitioned LLT (1.94 seconds). However, when compared to the other methods, the Bayesian LLT also has the lowest rate of change in computation time as the sample size increases. At the sample size of 150, the Bayesian method maintains faster computation time (5.02 seconds) than the full likelihood (13.77 seconds) and partitioned (5.57 seconds) LLTs. For each subsequent sample size, the Bayesian method outperforms the other methods.

4. NACC APOE e4 Data Analysis. Using the described data from the NACC, we fit the model of interest (section 2.1.2) with a random intercept LMEM, an LMEM with a random intercept and an AR(1) temporal variance structure, a partitioned LLT with a group size of 50 in each partition, and a Bayesian LLT. The primary question is whether those with the presence of the APOE e4 allele have a different rate of decline than those without an APOE e4 allele. As sex may have an interactive effect with APOE e4 on cognition, the interaction variable is also a parameter of interest. We test the null hypothesis of no effect of the APOE e4 allele on cognitive trajectory at the 0.05 significance level.

4.1. *NACC Analysis Results.* The effect estimates of the linear effect of APOE e4 on the Animals test trajectory are consistent with both simulation studies. The male participant APOE e4 effect estimate confidence intervals are smaller for the random intercept LMEM (0.085) and the AR(1) random intercept LMEM (0.219) when compared to the partitioned LLT (0.360) and the Bayesian LLT (0.280). As the Real Data Simulation suggests (section 3.2), the smaller confidence interval for the LMEMs are indicative of improper Type I error.

The LMEM models indicate a significant decrease in cognitive trajectory for those who carry the APOE e4 allele (see table 3). The Bayesian LLT, which can be considered the

Table 3: Effect estimates for APOE and APOE x Sex interaction. As was shown in the simulation anayses, the LMEM methods have a smaller confidence interval on the effect estimate compared to the LLT methods. This could lead to infering a significant effect, when one does not necessarily exist.

	APOE	APOE x Sex
LMEM	-0.143 (-0.229, -0.058)	-0.023 (-0.128, 0.082)
LMEM AR(1)	-0.135 (-0.244, -0.025)	-0.049 (-0.182, 0.085)
Partition LLT	-0.810 (-0.360, 0.000)	-0.047 (-0.261, 0.167)
Bayesian LLT	-0.136 (-0.269, 0.011)	-0.054 (-0.23, 0.122)

most trusted of the estimation processes with respect to inference, indicates that there is not a significant effect. Referring to the results in section 3.2.1, the LMEMs likely show significance due to an underestimated standard error.

5. Discussion. A Local Linear Trend State Space Model was developed to model cognition trajectory in the aging population. The proposed LLT models offer an intuitive representation of population effects on cognition while also allowing for heterogeneous trajectories from subject to subject. Model performance was validated using a fully simulated data analysis and a real data simulation analysis. The Partitioned and Bayesian LLT models were then used to estimate the linear effect of the APOE e4 allele on the Animals test trajectory using data from the NACC.

The fully simulated data simulation shows the LLT methods are more robust to model misspecification than the LMEM counterparts. The LLT models maintain 95% coverage when the true underlying data generation process follows that of a traditional LMEM, whereas the LMEM models failed when the the true underlying data generation process followed an LLT. Of the LLT models, the Bayesian LLT was the top performer as it maintained proper coverage, was unbiased, and had the tightest confidence intervals, meaning higher power.

The real data simulation allowed for assessment of the LLT using the actual NACC cognition trajectory data. The partitioned LLT and Bayesian LLT were both unbiased and maintained proper 95% coverage, unlike the commonly used LMEMs. Once again, the Bayesian LLT had less parameter variance and therefore maintains greater power than the partitioned LLT.

Of the LLT estimation processes, the Bayesian LLT stands out. Not only did it meet all the requirements of 1.) 95% coverage, 2.) unbiasedness, and 3.) small confidence interval length, it also did not suffer from issues of non-convergence in test statistic variances. The full likelihood LLT can be prone to numerical inaccuracy and non-convergence in the variance when inverting large matrices as part of the Kalman Filter. The partitioned LLT can help with this issue, but if there is not enough participants in each partitioned group to estimate the number of parameters there will also be a lack of convergence in the variance estimation. There is also an issue of choosing the right partition size which needs to balance group size to number of parameters, matrix inversion complexity, and number of iterations it takes the variance estimation to converge. By using the Bayesian methodology we are able to bypass numerical inaccuracy due to matrix inversion without needing to lose any information by partitioning, which is another reason the Bayesian LLT is preferred over the other LLT estimation strategies.

Lastly, the NACC analysis shows consistent results with what was found in the simulation analyses. The partitioned LLT and Bayesian LLT show wider confidence intervals on the effect estimate of APOE e4 on the Animals test trajectory. The Bayesian LLT effect estimate suggests

having an APOE e4 does not lead to a significant impact on cognitive trajectory among those with AD while the LMEM models show a significant effect. The simulation results indicate that a potential reason for conflicting effects of the APOE e4 allele on cognitive trajectory may be due to improperly accounting for the data structure leading to under estimation of effect standard errors for the LMEMs.

The growing aging population emphasizes the need to better understand dementia and, more specifically, Alzheimer's Disease. The local linear trend model, using the Bayesian estimation process, is shown to have many advantages over the commonly used linear mixed effect models. The Bayesian LLT proves to be a promising tool to understand how and why AD progresses in order to facilitate early detection and diagnosis, which is critical for timely implementation of intervention strategies.

Bibliography. 10 "2021 alzheimer's disease facts and figures" (2021), *Alzheimer's & dementia*, United States, 17, 327–406.

Beekly, D. L., Ramos, E. M., Lee, W. W., Deitrich, W. D., Jacka, M. E., Wu, J., Hubbard, J. L., Koepsell, T. D., Morris, J. C., and Kukull, W. A. (2007), "The national alzheimer's coordinating center (NACC) database : The uniform data set," *Alzheimer disease and associated disorders*, Hagerstown, MD: NIA Alzheimer's Disease Centers; Lippincott Williams & Wilkins, 21, 249–258.

Byrd, R. H., Lu, P., Nocedal, J., and Zhu, C. (1995), "A limited memory algorithm for bound constrained optimization," *Scientific Computing*, 16, 1190–1208.

"Dementia" (2020), *World Health Organization*, World Health Organization.

Dueck, A. C., Osborne, D., Snyder, C. H., Sabbagh, M. N., Shi, J., Locke, D. E. C., Rademakers, R., Alexander, G. E., Caselli, R. J., Rapcsak, S. Z., Ahern, G. L., Woodruff, B. K., Baxter, L. C., Reiman, E. M., and Connor, D. J. (2009), "Longitudinal modeling of age-related memory decline and the APOE e4 effect," *The New England journal of medicine*, Massachusetts Medical Society, 361, 255.

Durbin, J., and Koopman, S. J. (2012), *Time series analysis by state space methods*, Oxford University Press.

Feng, L., Li, J., Yap, K.-B., Kua, E.-H., and Ng, T.-P. (2009), "Vitamin b-12, apolipoprotein e genotype, and cognitive performance in community-living older adults : Evidence of a gene-micronutrient interaction," *The American journal of clinical nutrition*, Bethesda, MD: American Society for Nutrition, 89, 1263–1268.

Ferris, S. H., and Farlow, M. (2013), "Language impairment in alzheimer's disease and benefits of acetylcholinesterase inhibitors," *Clinical Interventions in Aging*, 8, 1007–1014.

Fildes, R., Harvey, A. C., West, M., and Harrison, J. (1991), "Forecasting, structural time series models and the kalman filter," *The Journal of the Operational Research Society*, Abingdon: Macmillan Press, 42, 1031–1033.

Kalman, R. E. (1960), "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering*, 82, 35–45.

Laird, N. M., and Ware, J. H. (1982), "Random-effects models for longitudinal data," *Biometrics*, 38, 963–974.

Martins, C. A. R., Oulhaj, A., De Jager, C. A., and Wwilliams, J. H. (2005), "APOE alleles predict the rate of cognitive decline in alzheimer disease : A nonlinear model," *Neurology*, Hagerstown, MD: Lippincott Williams & Wilkins, 65, 1888–1893.

Mitchell, D. J., Dujon, A. M., Beckmann, C., and Biro, P. A. (2020), "Temporal autocorrelation: A neglected factor in the study of behavioral repeatability and plasticity," *Behavorial Ecology*, 31, 222–231.

Raber, J., Huang, Y., and Ashford, J. W. (2004), "ApoE genotype accounts for the vast majority of AD risk and AD pathology," *Neurobiology of Aging*, 25, 641–650. https://doi.org/https://doi.org/10.1016/j.neurobiolaging.2003.12.023.

Rolland, Y., Abellan van Kan, G., Nourhashemi, F., Andrieu, S., Cantet, C., Guyonnet-Gillette, S., and Vellas, B. (2009), "An abnormal "one-leg balance" test predicts cognitive decline during alzheimer's disease," *Journal of Alzheimer's disease*, Netherlands, 16, 525–531.

Rongve, A., Soennesyn, H., Skogseth, R., Oesterhus, R., Hortobágyi, T., Ballard, C., Auestad, B. H., and Aarsland, D. (2016), "Cognitive decline in dementia with lewy bodies: A 5-year prospective cohort study," *BMJ open*, England: BMJ Publishing Group LTD, 6, e010357–e010357.

Sanders, A. E., Wang, C., Katz, M., Derby, C. A., Barzilai, N., Ozelius, L., and Lipton, R. B. (2010), "Association of a functional polymorphism in the cholesteryl ester transfer protein (CETP) gene with memory decline and incidence of dementia," *JAMA : the journal of the American Medical Association*, Chicago, IL: American Medical Association, 303, 150–158.

Shumway, Rober H. and Stoffer, David S. (2017), *Time series analysis and its applications: With r examples*, Springer.

Stott, D. J., Falconer, A., Kerr, G. D., Murray, H. M., Trompet, S., Westendorp, R. G. J., Buckley, B., De Craen, A. J. M., Sattar, N., and Ford, I. (2008), "Does low to moderate alcohol intake protect against cognitive decline in older people?" *Journal of the American Geriatrics Society (JAGS)*, Malden, USA: Blackwell Publishing Inc, 56, 2217–2224.

Vik-Mo, A. O., Giil, L. M., Ballard, C., and Aarsland, D. (2018), "Course of neuropsychiatric symptoms in dementia: 5-year longitudinal study," *International journal of geriatric psychiatry*, England: Wiley Subscription Services, Inc, 33, 1361–1369.

Zhou, W., and Li, D. (2007), "Limited memory BFGS method for nonlinear monotone equations," *Journal of computational mathematics*, Chinese Academy of Mathematices; System Sciences (AMSS) Chinese Academy of Sciences, 25, 89–96.