Introduction

The well established association of the APOE e4 gene on risk of Alzheimer's Disease (AD) (E H Corder 1993) has led to numerous studies analyzing the cognitive performance trajectory of APOE e4 carriers. Of these studies, the Linear Mixed Effect Model (LMEM) is commonly utilized due to its simple implementation and interpretation. Rarely do these models attempt to take into account the temporal autocorrelation of subject specific repeated cognitive measurements (David J Mitchell 2020). Although modeling temporal autocorrelation can be a difficult task in the LMEM framework, the AR(1) covariance structure has been successful in many applications and is commonly used when the structure is unknown (Chi 1989), even if the intuition behind such use is unclear.

We propose the use of a State Space Model (SSM) for modeling cognitive trajectory due to its dynamic temporal autocorrelation structure among other benefits.The SSM framework has been successfully used in numerous fields such as_______. SSMs are a time series technique commonly used in situations with few subjects and a large number of repeated measurements. We wish to employ the SSM in situations more common to neuropsychological outcomes of a large number of subjects with only few observations.

The SSM autocorrelation structures proves to be more intuitive than the commonly accepted AR(1) when considering multiple measurements of neuropysch scores.There is an added flexibility in this intuition that also accommodates non-linear subject specific predictions of a given outcome. Even with this flexibility, our proposed SSM is able to accurately estimate population level linear effects on cognitive ability with the same general interpretation as the LMEM.

A drawback of SSM models is the relative computation time when compared to LMEM estimation. In addition to a brute force full likelihood SSM, we propose and contrast the use of a partitioned SSM and a Bayesian SSM. Both of which significantly reduce computation time.

We then conduct two simulation analyses: 1) comparing LMEM models and our proposed SSMs on fully simulated data controlling the underlying data generation process and 2) on real data where the underlying data gerating process is unknown. The fully simulated data analysis shows that the LMEMs and SSMs, while unbiased, are sensitive to effect variance estimation under different data generating scenarios. The Bayesian SSM has an additional benefit of easy to see signs of autocorrelation misspecification.

Utilizing real data from the NACC (Beekly 2007) we are able to show the SSMs properly estimate the variance of a simulated linear effect on the Animals outcome. The LMEM with an AR(1) correlation fails to correctly estimate the effect variance yielding results consistent with the full simulated analysis. These validate the SSMs intuitive interpretation is more realistic for neuropsycholgical outcomes.

The proposed Bayesian SSM stands out in both simulation studies due to its unbiasedness, accurate variance estimation, signalling of model misspecification, and computation time. We go on to discuss potential extensions of the Bayesian SSM as a generalized linear model and non-linear model.

A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A A

This autocorrelation structure is shown to be more intuitive and realistic than the commonly accepted AR(1) structure.

We show, through simulation on real data, that our SSM model is more effective at estimating effect on cognitive trajectory. The SSM model p allowance for non-linear subject predictions and

Our proposed model estimates population level trajectories like the LMEM, but allows for non-linear subject specific predictions

The SSM is traditionally a time series model that is been used in various fields such as _______. This modeling technique has been primarily utilized in situations with many repeated measurements and only a few number subject. We shows the SSM effectiveness in situation more common to neuropsychological data of a large number of subjects with only few time points.

Many studies have been conducted showing an increased risk of Alzheimer's Disease (AD) for those carrying the APOE e4 gene (E H Corder 1993). This well documented association has led to various studies on cognitive trajectory of those suffering from AD (Richard J. Caselli 2009), with varying results.

Many studies have utilized repeated neuropsychological measurements to assess predictors of cognitive ability trajectory. The APOE e4 gene has also been studied extensively for its relation to Alzheimer's Disease (AD) prevalence (E H Corder 1993) and cognitive ability trajectory (Richard J. Caselli 2009).

Repeated neuropsychological measurements are commonly used to compare individual and population trajectory differences in cognitive ability. This data allows for inference on possible determinants of cognitive function which can be used for interventions and further study. Linear mixed effect models (LMEM) have become a common go-to when analyzing longitudinal data due to simple implementation and interpretation_____.

The APOE e4 allele is known to be associated with dementia and in turn poorer cognitive ability _____. Although this association is well documented, little has been done analyzing cognition trajectory for those with APOE e4. To describe differences in cognition trajectory we used the NACC data set. _____NACC DESCRIPTION_____.

We started by using the common linear mixed effect model with a subject specific random intercept controlling for time interactions with _____.

After estimating the model, we found the residuals had a lag 1 autocorrelation of _____. Unaccounted for temporal autocorrelation can lead to improper effect estimates and effect standard errors which invalidates model inference. Correctly specifying temporal autocorrelation can be a difficult process and many methods within the LMEM framework have been implemented_***. A common choice when the true autocorrelation process is unknown is the AR(1) covariance process which has been proven effective in many applications***___. However, the AR(1) autocorrelation lacks simple intuition when applied to neuropsychological outcomes and the true underlying cognitive ability.

To address shortcomings of the LMEMs we propose the use of State Space Models (SSM). Using the SSM framework we are able to flexibly fit an intuitive autocorrelation process. SSM allows for data driven disease trajectories rather than the deterministic LMEM AR(1) approach.

(David J Mitchell 2020)

In both synthetic simulation and real data simulation, we show that correct specification of temporal autocorrelation is vital for correct parameter inference.

Novel additions

Applying SSM towards real data (small t large n)

Doing everything in matrix form. Partitioned SSM. Bayesian SSM.

Beekly, Ramos, D. L. 2007. "The National Alzheimer's Coordinating Center (NACC) Database: The Uniform Data Set." *Alzheimer Disease and Associated Disorders* 21: 249–58.

Chi, & Reinsel, E. M. 1989. "Models for Longitudinal Data with Random Effects and AR (1) Errors." *Journal of the American Statistical Association* 84: 452–59.

David J Mitchell, Christa Beckmann, Antoine M Dujon. 2020. "Temporal Autocorrelation: A Neglected Factor in the Study of Behavioral Repeatability and Plasticity." *Behavorial Ecology* 31: 222–31.

E H Corder, W J Strittmatter, A M Saunders. 1993. "Gene Dose of Apolipoprotein e Type 4 Allele and the Risk of Alzheimer's Disease in Late Onset Families." *Science* 261: 921.

Richard J. Caselli, Amylou C. Dueck, M. D. 2009. "Longitudinal Modeling of Age-Related Memory Decline and the APOE E4 Effect." *New England Journal of Medicine* 361: 255–63.