Introduction II

According to the World Health Organization dementia effects around 50 million people in the world today with 60-70% of those due to Alzheimer's disease (AD). The number of those suffering with dementia and AD is only accelerating due to an increase in population life expectancy. Dementia is characterized by physical changes in the brain leading to deterioration in cognitive function, emotional control, social behavior, and/or motivation. Not only do those diagnosed with dementia suffer, but there is also an additional physical, emotional, and financial burden on family members and caretakers.

Much research has been done to understand how and why dementia progresses in order to assist in early diagnosis and treatments. Organizations, like the Nation Alzheimer's Coordinating Center (NACC), have collected large amounts of repeated measurements on participants functional ability to aid in this research. Repeatedly measuring a subject's cognitive ability through various tests works as a valid proxy to physical changes in the brain. We can estimate possible effect modifiers of the cognitive tests to understand what drives dementia trajectory as subjects get older. Analyzing these test outcomes can gain vital insight into disease identification and treatment.

Estimating effect modifiers in cognition data comes with a number of challenges. For interpretable results, we often rely on estimating linear effect modifiers. However, subject level cognition data are often non-linear over time. Many subjects show a similar overall trajectory, but also show random oscillations in cognition along the way. Non-linear methods and non-parametric have been used on such data, but it often comes at the cost of simple interpretation. **OTHERS???**

To address these issues we propose the use of a Local Linear Trend (LLT) State Space Model (SSM) for modeling cognitive trajectory due to its flexibility in measuring linear effects in data that is not necessarily linear. The SSM is traditionally a time series technique employed in situations with only a few number of subjects and a large series of time points. 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 subjects with only a few time points.

An intuitive comparison to the LLT model is the simple Linear Mixed Effect Model (LMEM), which is a subset of the LLT. The LMEM is common in modeling cognitive outcomes due to its simple implementation and interpretation. The LLT shares an equivalent linear effect interpretation as the LMEM, with an added flexibility for random changes in subject cognition. Some studies using the LMEM have enforced an AR(1) auto-correlation structure when modeling cognitive data to help counteract non-linearities. Autocorrelation can be thought of as the effect of unobserved time-varying variables (UTV) on the 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 has more flexibility in letting the UTV effect vary freely over time following a stochastic process. Even with the added flexibility of the LLT, we show that it still produces accurate effect parameter estimates.

A possible drawback of the traditional likelihood LLT estimation process on a large data set is the computation time. Estimation of this model is an iterative process that includes difficult to avoid matrix inversions which increase computation exponentially as the sample size increases. To address the computational inefficeinty of the brute force full likelihood estimation, we propose and compare a partioned LLT and a Bayesian LLT.

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

By using a Bayesian approach on parameters of interest, we are able to avoid large matrix inversion and instead rely on Markov Chain Monte Carlo (MCMC) sampling. Accurate estimation using MCMC is typically more computationally expensive for a small number of subjects, but maintains a linear increase in computation which allows it to quickly out perform the full likelihood method. Along with being the most accurate

of the LLT estimation procedures, the Bayesian method is also useful in identifying model mispecification through visual inspection of the MCMC draws.

To assess model performace we show 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 generating process is unknown.

By controlling the underlying data generating process we are able to demonstrate the LLT effectiveness when model specification are met and when they are not. We also demonstrate simple diagnostics to identify model mispecification using the Bayesian LLT. The fully simulated data analysis shows that the LMEMs and SSMs, while unbiased, are sensitive to parameter variance estimation under different data generating scenarios.

The effectiveness of the LLT models is showcased when we simulate using real data with an unknown true underlying data generation process. Utilizing real data from the NACC [@NACC] we are able to show the LLT models 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 parameter variance yielding results consistent with the full simulated analysis. The real data simulation validates that the SSMs intuitive interpretation is more realistic for neuropsycholgical outcomes.

==

The UTV effect assumption of the LMEM assumes a deterministic long term trajectory if you know a subjects random intercept and covariates. Additional data on how the subject is performing cognitively does not change the long term trajectory. The SSM, however, allows for the long term cognitive trajectory to change based on observed data.

Simple Methods Description. . . . .

Go through simulations. . . . .

Go through results. . . . .