Untitled

Zach

12/16/2020

Methods

Data

1,643 subjects from the NACC who started cognitively normal and transitioned to either MCI or Dementia. An average of 6.25 visits per subjects. There was a minimum of 2 visits and a maximum of 15 IQR 8-4.63.8% male, 85.8% white, 34.0% APOE e4 carriers, average starting age of 76.3, average education of 15.9 years.

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. The NACC was established by the National Institute of Aging in 1990 is an effort to centralize AD-related clinical and neuropsychological data from 34 different research facilities (Beekly 2007). Participants from the varying research facilities are offer a battery of neuropyschological tests to gain insight into underlying cognitive ability.

Criteria for entry into this analysis requires the subject to have transitioned from cognitively normal to mild cognitive impairment (MCI) or Dementia during the subject's follow-up period in the NACC. Study participants are predominantly male (63.8%)!!!!DOUBLE CHECK CODING!!!!!! and of white ethnicity (85.8%). Participants are primarily older with an average age of 76.7 years old (SD = 9.3) and have an average of 15.9 years of education (SD = 6.9). A primary research question is whether cognitive trajectory differs between APOE carrier. 34.0% of the subjects carry at least 1 APOE e4 allele.

The Animals test, in which subjects name as many animals as possible in one minute, is the primary outcome of interest (mean = 16.8, SD = 5.7). This test is one of the few offered in the battery without a truncated distribution. Truncation in the outcome can lead to biased estimates when treating an outcome as continuous, as is done with the standard LMEM. For the real data simulation analysis a linear effect generated for half the subjects was added to the Animals test. By randomly adding a linear effect to the outcome we are able to show model appropriateness when the underlying data generating distribution is unknown.

Fully simulated data was generated to compare general estimation accuracy under incorrect and correct model specification. Each simulation consisted of 100 subjects with 2-10 observations. Each predictor was simulated to roughly resemble predictors use in the real data analysis.

General Linear Gaussian State Space Model

To understand our proposed model we first introduce the general linear state space model and the estimation process. Suppose you have observations for n subjects at times $t \in \{1, 2, ..., T\}$. We denote the outcome vector for the n subjects at time t as y_t and for our purposes assume that y_t is continuous. A general linear state space model can be denoted as:

$$y_t = F_t\mu_t + v_t, \quad v_t \sim N(0, V_t)$$
$$\mu_t = G_t\mu_{t-1} + w_t, \quad w_t \sim N(0, W_t) \ \ \& \ \ \mu_0 \sim N(u, P)$$

Both v_t and w_t are independent. The $q \times 1$ vector μ_t is the unobserved latent state vector where q is the number of latent states. From this formulation we see that the observed value of y_t is a linear combination of μ_t after being transformed by the observation matrix F_t, which we assume is known, plus some random noise v_t. Furthermore, we assume μ_t follows a random walk through time where μ_t is a linear function of μ_{t-1} in the form $G_t\mu_{t-1}$ plus some random noise w_t. The matrix G_t is the state transition matrix and assumed to be known. This state space model can be regarded as a Hidden Markov model with continuous latent states and continuous observations.

For our purposes, we can think of μ_t as the components of someones underlying cognition. One's underlying cognitive ability may vary randomly through time where they have longer periods of being well and other periods of not doing so well. For an observational study we then come up with a test to measure the underlying cognition μ_t. This test takes into account different parts of cognition through the transformation F_t plus some random error. In this way we are simply tapping into the underlying cognition to get our observation vector y_t.

For a fixed F_t, G_t, V_t, W_t, u, and P we can utilize the popular Kalman Filter and Kalman Smoother to calculate $\hat{\mu}_{t|T} = E(\mu_t|y_{1:T})$ and $P_{t|T} = \text{Var}(\mu_t|y_{1:T})$. The Kalman filter is a recursive algorithm to estimate the unobserved states conditioned on the observed data $y_{1:T} = \{y_1, y_2, ..., y_T\}$ (Kalman, 1960; Durbin and Koopman, 2012). Let $\hat{\mu}_{i|j} = E(\mu_i|y_{1:j})$ and $P_{i|j} = \text{Var}(\mu_i|y_{1:j})$. The filter proceeds as follows:

- Predicted state: $\hat{\mu}_{t|t-1} = G_t\hat{\mu}_{t-1|t-1}$
- Predicted state covariance: $P_{t|t-1} = G_t P_{t-1|t-1} G_t' + W$
- Innovation covariance: $S_t = F_t P_{t|t-1} F_t' + V$
- Kalman Gain: $K_t = P_{t|t-1} F_t' S_t^{-1}$
- Innovation: $\tilde{f}_t = y_t - F_t\hat{\mu}_{t|t-1}$
- Updated state estimate: $\hat{\mu}_{t|t} = \hat{\mu}_{t|t-1} + K_t\tilde{f}_t$
- Updated state covariance: $P_{t|t} = (I - K_t F_t)P_{t|t-1}$
- Updated innovation: $\tilde{f}_{t|t} = y_t - F_t\hat{\mu}_{t|t}$

The Kalman Filter results in $\hat{\mu}_{t|t}$ and $P_{t|t}$ for $t \in \{1, 2, 3, ..., T\}$. To find the expected value and variance of the unobserved states conditioned on all the data we iterate backwards through the Kalman Smoother. By letting $J_t = P_{t|t}G_{t+1}' + P_{t+1|t}^{-1}$ we can compute the following.

$$\hat{\mu}_{t|T} = \hat{\mu}_{t|t} + J_t(\hat{\mu}_{t+1|T} - \hat{\mu}_{t+1|t})$$
$$P_{t|T} = P_{t|t} - J_t G_{t+1} P_{t|t}$$

Proper estimates for μ_t depend on correctly specifying parameters, including V_t, W_t, u, and P. For our purposes, the matrices F_t and G_t are fixed by design. V_t and W_t can be estimated by maximizing their joint log-likelihood.

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

Typically, (1) the Kalman Filter and Smoother are computed, (2) the log-likelihood is computed then maximized, then we iterate through (1) and (2) until a convergence criteria is met.

Dealing with the initial parameters μ_0, namely the mean u and variance P, tends to be more simple. It can be shown that by giving μ_0 an "infinite prior" of $P = \infty$ and $u = 0$, after only a few iterations our estimates $P_{i|j}$ converge to the values they would have taken if we would have started with the true P and u.

2

Proposed Model For Neuropsychological Outcomes

We propose the following as a flexible alternative to a linear mixed effect model,

$$y_t = \alpha_t + X_t\beta_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2 I_n)$$
$$\alpha_t = \alpha_{t-1} + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2 I_n) \ \& \ \alpha_0 \sim N(a_0, P_0)$$
$$\beta_t = \beta_{t-1}, \quad \beta_0 \sim N(\beta, 0)$$

The X_t is an $n \times p$ matrix of predictors at time t, where the ith row is the ith subject's predictors. We assume β_t is a $p \times 1$ vector that remains fixed over time and follows the same linear fixed effect interpretation as a similar linear regression or linear mixed effect model. The proposed model differs from traditional models in that there is a randomly varying underlying α_t that follow a random walks through time. With respect to y_t, α_t follows a random walk about $X_t\beta_t$ and therefore can be interpreted as random variation in cognition not accounted for by the predictors in X_t.

Notice, $E(\alpha_t|\alpha_{t-1}) = \alpha_{t-1}$. This indicates that the underlying variations in one's cognitive state at time t, not accounted for by the predictors X_t, depends on their underlying cognitive state at time $t - 1$. This creates a Markov Chain trend in a subject's underlying cognition.

We also make the assumption that $\alpha_{i,t}$ is independent from $\alpha_{i',t}$ for $i \neq i'$ and $i, i' \in \{1, 2, ..., n\}$, meaning the underlying variation about the trend $X_t\beta$ is independent between subjects. Therefore, the chain of $\alpha_{i,1:T}$ is a random walk specific to subject i. The parameter β_t, however, is shared between each subject.

Autocorrelation

Our proposed model implies a flexible dynamic moving average auto-correlation structure. The correlation between any two time points for subject i is defined as,

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

If there is no variation in α over time ($\sigma_\eta^2 = 0$) then $corr(y_{it}, y_{i(t+\tau)}) = 0$. In this situation we are able to draw a parallel to a linear mixed effect model. To further understand this consider our reduced model,

$$y_t = \alpha_0 + \sum_{j=1}^{t} \eta_j + X_t\beta_0 + \varepsilon_t$$

If $\sigma_\eta^2 = 0$ then $\eta_t = 0$ for $t \in \{1, 2, ..., T\}$. The model reduces to,

$$y_t = \alpha_0 + X_t\beta_0 + \varepsilon_t$$

Where $\alpha_0 \sim N(a_0, P_0)$, 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, but also accommodate more complex temporal auto-correlation.

Formulating Proposed Model Into State Space Form

By rewriting the proposed model the following way we show that it fits a state space model,

$$y_t = \begin{bmatrix} I_n & X_t \end{bmatrix} \begin{bmatrix} \alpha_t \\ \beta_t \end{bmatrix} + \varepsilon_t$$

$$\begin{bmatrix} \alpha_t \\ \beta_t \end{bmatrix} = \begin{bmatrix} I_{(n+p)} \end{bmatrix} \begin{bmatrix} \alpha_{t-1} \\ \beta_{t-1} \end{bmatrix} + \begin{bmatrix} \eta_t \\ 0_{p \times 1} \end{bmatrix}$$

where,

$$F_t = \begin{bmatrix} I_n & X_t \end{bmatrix}, \quad \mu_t = \begin{bmatrix} \alpha_t \\ \beta_t \end{bmatrix}, \quad V_t = \sigma_\varepsilon^2 I_n, \quad G_t = I_{n+p}, \quad W_t = \sigma_\eta^2 \begin{bmatrix} I_n & 0 \\ 0 & 0 \end{bmatrix}$$

The Kalman Filter and Kalman Smoother along with the variance optimization as described above may be calculated to get an estimate of $\mu_{t|T} = [\alpha'_{t|T} \ \beta'_{t|T}]'$. Recall, we assume that β_t remains constant over time, so all $\beta_{j|T}$ for $j \in \{1, 2, ..., T\}$ are equal. Our estimate for the linear effect of our predictors is $\hat{\beta} = \beta_{T|T}$ which has variance $P_{\hat{\beta}} = [P_{T|T}]_{(n+1:n+p),(n+1:n+p)}$. If modeling assumptions are met, $\hat{\beta} \sim N(B, P_{\hat{\beta}})$ which can be used for hypothesis testing.

Missing Data and Continuous Time Implementation

Observational data often comes with unequally spaced observations. So far we have introduced t as an indicator for time, but it can also indicate the tth observation. In the proposed formulation, $V_t = \text{Var}(\alpha_t | \alpha_{t-1}) = \sigma_\eta^2 I_n$. Suppose we are missing observations $t - 1 : t - \tau$ then we would be interested in the following,

$$\begin{aligned}
\text{Var}(\alpha_t | \alpha_{t-\tau-1}) &= \text{Var}(\alpha_{t-1} + \eta_t | \alpha_{t-\tau-1}) \\
&= \sigma_\eta^2 + \text{Var}(\alpha_{t-2} + \eta_{t-1} | \alpha_{t-\tau-1}) \\
&= \tau \sigma_\eta^2
\end{aligned}$$

Using this logic, we can set t to indicate the tth observation and let $V_t = \delta_t \sigma_\eta^2 I_n$ where δ_t is the time difference between t^{th} and $(t-1)^{th}$ observations. Instead of a scalar, δ_t can also be an $n \times 1$ vector where δ_{ti} is the time difference between the tth and $(t-1)^{th}$ observation for the ith subject. Adjusting the variance matrix V_t allows for continuous measurement times for each subject.

For situations where the number of observations are not equal between subjects we can define W_t as a subset of rows of I_n corresponding to those with t^{th} observations present. Then let,

- $y^* = W_t y_t$.
- $F_t^* = W_t F_t$
- $\varepsilon_t^* = W_t \varepsilon_t$

Carrying out the same Kalman filter and smoother replacing y with y^*, Z with Z^*, and ε_t^* with ε_t allows for computation of all the smoothed values for α_t and β_t.

Computational Considerations

Recall, state vector initial parameters are unknown, putting a diffuse prior on the variances will quickly converge to the same variances as if we correctly specified the initial conditions. To estimate our proposed

model we set a diffuse prior on the vector $\mu_0 = [\alpha_0' \ \beta_0']'$. Each subject shares β, which is now random as we assume $\beta_0 \sim N(0, \infty)$, therefore the observations y_t are no longer treated as independent in the Kalman Filter and Kalman Smoother. This leads to the inverse of a possibly large non-sparse $n \times n$ matrix $\text{Var}(Y_t | y_{1:t-1}) = S_t$ in the Kalman Filter process for $t \in \{1, 2, ..., T\}$. The S_t 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 inefficiency is then amplified as the Kalman Filter needs to be run multiple times for the maximum likelihood estimation of σ_ε^2 and σ_η^2. We show, as expected, that the computation time increases exponentially as n increases.

Partitioning

A simple way to decrease the computation time is to randomly and equally partition 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 computation time as n increases.

Bayesian Gibb's Sampling

As shown above, the computational issues arise because β is not treated as fixed in the Kalman Filter. This issue can be avoided by estimating β outside of the Kalman Filter using a Bayesian Gibb's Sampler which will fix the previous inversion obstacle as the elements of Y_t will be treated as independent. Given some regularity conditions, after an adequate burn-in sample, the joint sample from the posterior distributions will represent a sample from the joint distribution of the unknown parameters. Inference can then be made using the samples of the respective parameters.

The Model

In this representation we no longer index the linear effect β with a time component.

$$
\begin{aligned}
y_t &= \alpha_t + X_t \beta + \varepsilon_t \\
\alpha_t &= \alpha_{t-1} + \eta_t
\end{aligned}
\tag{1}
$$

where $\varepsilon_t \sim N(0, \sigma_\varepsilon^2 I_n)$, $\eta_t \sim N(0, \sigma_\eta^2 I_n)$, and $\alpha_0 \sim N(u_0, P_0)$ for some vector α_0 and some positive valued diagonal matrix P_0.

We also have the following prior distributions,

$$
\begin{aligned}
\beta &\sim N(\theta, \sigma_\beta^2) \\
\sigma_\eta^2 &\sim IG(a_0/2, b_0/2) \\
\sigma_\varepsilon^2 &\sim IG(c_0/2, d_0/2)
\end{aligned}
\tag{2}
$$

For notional ease, let $Y = [y_1, ..., y_T]$, $\alpha = [\alpha_1, ..., \alpha_T]$, $X = [X_1, ..., X_T]$

Posterior Distributions

$\alpha|Y, X, \beta, \sigma_\eta^2, \sigma_\varepsilon^2$

The conditional distribution $\alpha|Y, X, \beta, \sigma_\eta^2, \sigma_\varepsilon^2$ can be estimated directly from the the Kalman Filter. By conditioning on X_t and β each y_{ti} for $i \in \{1, 2, ..., n\}$ is independent at each observation t. Thus, we can run the Kalman Filter with $\tilde{y}_t = y_t - X_t\beta$ as the outcome in the model,

$$\begin{aligned} \tilde{y}_t &= \alpha_t + \varepsilon_t \\ \alpha_t &= \alpha_{t-1} + \eta_t \end{aligned} \tag{3}$$

Because the $\tilde{y}_t$ are independent it is equivalent to estimating the α_{ti} independently, resulting in a computational efficient Kalman Filter where,

$$\begin{aligned} \alpha_{t|t-1} &= \alpha_{t-1|t-1}, \quad P_{t|t-1} = P_{t-1|t-1} + \sigma_\eta^2 \\ \alpha_{t|t} &= \alpha_{t|t-1} + K_t(\tilde{y}_t - \alpha_t^{t-1}), \quad P_{t|t} = (1 - K_t)P_{t|t-1} \\ K_t &= \frac{P_{t|t-1}}{P_{t|t-1} + \sigma_\varepsilon^2} \end{aligned} \tag{4}$$

Although we are using vectors, because of the independence each operation is done element-wise. Let $\psi = \{\sigma_\eta^2, \sigma_\varepsilon^2, \beta\}$, the vector of the other unknown parameters. To get $\alpha_{t|T} = E_\psi(\alpha_t|y_{1:T})$ and $P_{t|T} = \text{Var}_\psi(\alpha_t|y_{1:T})$ we need the backward recursions starting with $J_T = 0$,

$$\begin{aligned} \alpha_{t-1|T} &= \alpha_{t-1|t-1} + J_{t-1}(\alpha_{t|T} - \alpha_{t|t-1}) \\ P_{t-1|T} &= P_{t-1|t-1} + J_{t-1}^2(P_{t|T} - P_{t|t-1}) \\ J_{t-1} &= \frac{P_{t-1|t-1}}{P_{t|t-1}} \end{aligned} \tag{5}$$

For the posterior of α we want,

$$\begin{aligned} P_\psi(\alpha_{0:T}|y_{1:T}) &= P_\psi(\alpha_T|y_{1:T})P_\psi(\alpha_{T-1}|\alpha_T, y_{1:T})...P_\psi(\alpha_0|\alpha_{1:T}, y_{1:T}.) \\ &= P_\psi(\alpha_T|y_{1:T})P_\psi(\alpha_{T-1}|\alpha_n, y_{1:(T-1)})...P_\psi(\alpha_0|\alpha_1) \end{aligned} \tag{6}$$

Therefore we need the densities,

$$p_\psi(\alpha_t|\alpha_{t+1}, y_{1:t}) \propto p_\psi(\alpha_t|y_{1:t})p_\psi(\alpha_{t+1}|\alpha_t) \tag{7}$$

Because of the normality assumption we know that $\alpha_t|y_{1:t} \sim N_\psi(\alpha_{t|t}, P_{t|t})$ and $\alpha_{t+1}|\alpha_t \sim N_\psi(\alpha_t, \sigma_\eta^2)$. We can show that after combining the two densities $m_t = E_\psi(\alpha_t|\alpha_{t+1}, y_{1:t}) = \alpha_{t|t} + J_t(\alpha_{t+1} - \alpha_{t+1|t})$ and $R_t = \text{Var}_\psi(\alpha_t|\alpha_{t+1}, y_{1:t}) = P_{t|t} - J_t^2 P_{t+1|t}$. Therefore, the posterior distribution for α_t is $N(m_t, R_t)$.

For the backward sampling procedure we start by sampling a α_T^* from a $N_\psi(m_T, R_T)$, then setting $\alpha_T^* = \alpha_T$ for the calculation of m_{T-1} to then sample α_{T-1}^* from a $N_\psi(m_{T-1}, R_{T-1})$. This process continues until a whole chain $\alpha_{0:T}^*$ has been sampled.

$\beta|Y, \alpha, \sigma_\eta^2, \sigma_\varepsilon^2$

With all the other parameters fixed,

$$P(\beta|Y, \alpha, \sigma_\eta^2, \sigma_\varepsilon^2) = \frac{P(Y, \alpha, \sigma_\eta^2, \sigma_\varepsilon^2|\beta)P(\beta)}{P(Y, \alpha, \sigma_\eta^2, \sigma_\varepsilon^2)} \tag{8}$$

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

$$
\begin{aligned}
P(Y, \alpha, \sigma_\eta^2, \sigma_\varepsilon^2|\beta) =& P(y_1, ..., y_T, \alpha_1, ..., \alpha_T, \sigma_\eta^2, \sigma_\varepsilon^2|\beta) \\
=& P(y_T|y_1, ..., y_{T-1}, \alpha_1, ..., \alpha_T, \sigma_\eta^2, \sigma_\varepsilon^2\beta) \\
& \times P(y_1, ..., y_{T-1}, \alpha_1, ..., \alpha_T, \sigma_\eta^2, \sigma_\varepsilon^2|\beta) \\
=& P(y_T|\alpha_T, \sigma_\varepsilon^2\beta)P(\alpha_{T-1}|y_1, ..., y_{T-1}, \alpha_1, ..., \alpha_{T-1}, \sigma_\eta^2, \sigma_\varepsilon^2, \beta) \\
& \times P(y_1, ..., y_{T-1}, \alpha_1, ..., \alpha_{T-1}, \sigma_\eta^2, \sigma_\varepsilon^2|\beta) \\
=& P(y_T|\alpha_T, \sigma_\varepsilon^2\beta)P(\alpha_{T-1}|\alpha_{T-1}, \sigma_\eta^2) \\
& \times P(y_1, ..., y_{T-1}, \alpha_1, ..., \alpha_{T-1}, \sigma_\eta^2, \sigma_\varepsilon^2|\beta) \\
=& P(\sigma_\varepsilon^2)P(\sigma_\eta^2)\left(\prod_{k=0}^{T} P(\alpha_k|\alpha_{k-1}, \sigma_\eta^2) \right) \prod_{t=1}^{T} P(y_t|\alpha_t, \sigma_\varepsilon^2, \beta) \\
\propto& \prod_{t=1}^{T} P(y_t|\alpha_t, \sigma_\varepsilon^2, \beta)
\end{aligned}
\tag{9}
$$

For simplicity we can further write,

$$
\begin{aligned}
-2logP(Y, \alpha, \sigma_\eta^2, \sigma_\varepsilon^2|\beta) \propto& \sum_{t=1}^{T} \frac{(y_t - \alpha_t - X_t\beta)'(y_t - \alpha_t - X_t\beta)}{\sigma_\varepsilon^2} \\
\propto& \sum_{t=1}^{T} \frac{-2y_t'X_t\beta + 2\alpha_t'X_t\beta + \beta'X_t'X_t\beta}{\sigma_\varepsilon^2} \\
\propto& \frac{\beta'\left(\sum_{t=1}^{T} X_t'X_t\right)\beta - 2\left(\sum_{t=1}^{T} y_t - \alpha_t\right)'X_t\beta}{\sigma_\varepsilon^2}
\end{aligned}
\tag{10}
$$

We can then find the proportionality of the prior of β,

$$
\begin{aligned}
-2logP(\beta) \propto& \frac{(\beta - \theta)'(\beta - \theta)}{\sigma_\beta^2} \\
\propto& \frac{\beta'\beta - 2\theta\beta}{\sigma_\beta^2}
\end{aligned}
\tag{11}
$$

Thus the -2log posterior of β is proportional to,

$$
\begin{aligned}
-2logP(\beta|Y, \alpha, \sigma_\eta^2, \sigma_\varepsilon^2) \propto& \frac{\beta'\left(\sum_{t=1}^{T} X_t'X_t\right)\beta - 2\left(\sum_{t=1}^{T} y_t - \alpha_t\right)'X_t\beta}{\sigma_\varepsilon^2} + \frac{\beta'\beta - 2\theta\beta}{\sigma_\beta^2} \\
\propto& \frac{\beta'\left(\sigma_\beta^2 \sum_{t=1}^{T} X_t'X_t\right)\beta - 2\sigma_\beta^2\left(\sum_{t=1}^{T} y_t - \alpha_t\right)'X_t\beta + \beta'\sigma_\varepsilon^2 I_p\beta - 2\sigma_\varepsilon^2\theta\beta}{\sigma_\varepsilon^2\sigma_\beta^2} \\
\propto& \frac{\beta'\left(\sigma_\beta^2 \sum_{t=1}^{T} X_t'X_t\right)\beta + \beta'\sigma_\varepsilon^2 I_p\beta - 2\sigma_\beta^2\left(\sum_{t=1}^{T} y_t - \alpha_t\right)'X_t\beta - 2\sigma_\varepsilon^2\theta\beta}{\sigma_\varepsilon^2\sigma_\beta^2} \\
\propto& \frac{\beta'\left(\left(\sigma_\beta^2 \sum_{t=1}^{T} X_t'X_t\right) + \sigma_\varepsilon^2 I_p\right)\beta - 2\left(\sigma_\beta^2\left(\sum_{t=1}^{T} y_t - \alpha_t\right)'X_t - \sigma_\varepsilon^2\theta'\right)\beta}{\sigma_\varepsilon^2\sigma_\beta^2} \\
\propto& \frac{(\beta - \Sigma^{-1}B)'\Sigma(\beta - \Sigma^{-1}B)}{\sigma_\varepsilon^2\sigma_\beta^2}
\end{aligned}
\tag{12}
$$

7

Where $B = \sigma_\beta^2 \left(\sum_{t=1}^T y_t - \alpha_t \right)' X_t - \sigma_\varepsilon^2 \theta'$ and $\Sigma = \left(\sigma_\beta^2 \sum_{t=1}^T X_t'X_t \right) + \sigma_\varepsilon^2 I_p$. Therefore, $\beta | Y, \alpha, \sigma_\eta^2, \sigma_\varepsilon^2 \sim N(\Sigma^{-1} B, \sigma_\varepsilon^2 \sigma_\beta^2 \Sigma^{-1})$.

Note here that Σ 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, Σ can be broken down to increase computation speed. Recall, $\Sigma = \left(\left(\sigma_\beta^2 \sum_{t=1}^T X_t'X_t \right) + \sigma_\varepsilon^2 I_p \right)$ and $\left(\sigma_\beta^2 \sum_{t=1}^T X_t'X_t \right)$ will not change at each iteration because it does not contain unknown parameters. By calculating the eigenvalue decomposition on $sigma_\beta^2 \sum_{t=1}^T X_t'X_t$ we can rewrite Σ as follows,

$$\left(\left(\sigma_\beta^2 \sum_{t=1}^T X_t'X_t \right) + \sigma_\varepsilon^2 I \right) = (Q\Lambda Q' + \sigma_\varepsilon^2 I)$$

$$= (Q\Lambda Q' + \sigma_\varepsilon^2 Q Q')$$

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

then,

$$\left(\left(\sigma_\beta^2 \sum_{t=1}^T X_t'X_t \right) + \sigma_\varepsilon^2 I \right)^{-1} = Q\left(1/(\Lambda + \sigma_\varepsilon^2 I) \right)Q'$$

We only need to calculate the eigen vectors Q and eigen values Λ of $sigma_\beta^2 \sum_{t=1}^T X_t'X_t$ once, then simply update σ_ε^2 before calculating the inverse.

$\sigma_\eta^2 | Y, \beta, \alpha, \sigma_\varepsilon^2$ **and** $\sigma_\varepsilon^2 | Y, \alpha, \sigma_\eta^2$

By applying the same Bayes' rule rationale,

$$
\begin{aligned}
P(\sigma_\eta^2 | Y, \beta, \alpha, \sigma_\varepsilon^2) &= \frac{P(Y, \beta, \alpha, \sigma_\varepsilon^2 | \sigma_\eta^2) P(\sigma_\eta^2)}{P(Y, \beta, \alpha, \sigma_\varepsilon^2)} \\
&\propto P(Y, \beta, \alpha, \sigma_\varepsilon^2 | \sigma_\eta^2) P(\sigma_\eta^2) \\
&\propto (\sigma_\eta^2)^{-nT/2} e^{\sum_{t=1}^T (\alpha_t - \alpha_{t-1})^2 / 2\sigma_\eta^2} (\sigma_\eta^2)^{-a_0/2-1} e^{-b_0/2\sigma_\eta^2} \\
&\propto (\sigma_\eta^2)^{-(nT+a_0)/2-1} e^{\left(\sum_{t=1}^T (\alpha_t - \alpha_{t-1})^2 + b_0 \right)/2\sigma_\eta^2}
\end{aligned}
\tag{13}
$$

Therefore, $\sigma_\eta^2 | Y, \beta, \alpha, \sigma_\varepsilon^2 \sim IG\left(\frac{nT+a_0}{2}, \frac{\sum_{t=1}^T (\alpha_t - \alpha_{t-1})^2 + b_0}{2} \right)$.

In a very similar fashion we can show $\sigma_\varepsilon^2 | Y, \beta, \alpha, \sigma_\eta^2 \sim IG\left(\frac{nT+c_0}{2}, \frac{d_0 + \sum_{t=1}^T (y_t - X_t\beta - \alpha_t)^2}{2} \right)$.

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.