440Independent

Ziyang Ding

Introduction

Real estate, as a major part of the economy, has been constanly and closely monitored by investors and researchers. Since 1970 {[1]}, when most countries' statistical offices or central banks began to collect data on house prices, interest in how to predict and forecast house prices have gradually augmented, bringing out more and more sophisticated modling techniques. Due to the increased ability of modern society to collect and store more data, attempts to make prediction on real estate prices have therefore shifted to data-driven, which further improved modeling precision.

Being such a sophisticated product, real estate prices are also impacted by many factors. While most of the factors helpful in predicting the house are observable and descriptive to the house itself, such as house' size, number of bathrooms, and whether it possesses a swimming pool etc., there are also inobservable factors that also impact house prices, such as the underlying real state market economy, cyclicality of real estate prices, and so on.

Many previous researches have already proposed multiple ways of predicting house prices. From the most simple regression methods as proposed in [2], to those that account for repeated sells of houses [3], and to those that take temporal effects into consideration, such as [4]. Though these studies are drastically different and are definitly other researches proposing more sophisticated models , each study has a different but clear focus. Thus, it is important to make certain of the research question before creating model.

Therefore, we proposes our goal of this study. The only type of house that we'll be researching into is house in Durham, NC, due to our better familiarity of the terrain. The goals include 1) understand how descriptive and observable variables affect housing prices, 2) understand how temporal effect affect housing price, 3) extract cyclical, trend, and mean-shift effect of past real estate, and 3) make short term forecast of housing prices.

EDA

Data Discription

The Dataset is scraped from redfin official set [5]. Redfin is a real estate brokerage that was founded in 2004. It's website consist of historical purchase record of the past 3 years. We therefore scrapped these 3 years of data, ranging from 2017 April to 2020 May. This dataset contains 6962 observation. Thanks to redfin's meticulous data record, no missing value in any field was presented. Each observation is a recorded deal of house purchase. Therefore, the price is the deal price between customer and seller, which is objective enough for us to fit on.

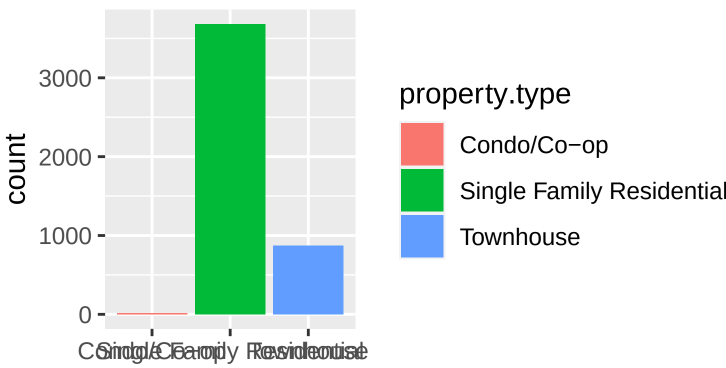
The data set contains many covariates. Among which, there are some non-process-able string information, such as name of the community, or geographical information which is beyond the scope of our interests. Therefore, to simplify our research, we introduce the following covariates of our interest

As we've already indicated above, we're interested only in houses and apartments in Durham. Therefore, after we filter out the data, the `city` variable no longer exist.

Name	Description	Missing
`Price`	the deal price of the house	0%
`beds`	number of beds in the house	0%
`sold.dat`	the date on which the deal is settled	0%
`baths`	the number of bathrooms the house has	0%
`square.feet`	usable area (ft^2) measured in square feet	0%
`lot.size`	total area (ft^2) of the lot	0%
`house.age`	age (years) of the house when purchased	0%
`property.type`	Townhouse, or Single family residential	0%
`latitude`	latitude of the house	0%
`longitude`	longitude of the house	0%
`city`	Durham, Chapel Hill, or Morresville	0%

The real EDA

The following figure indicates the extreme uneven number of observation for 3 types of houses: Conda/Co-op, Townhouse, and Single Family Residential. A complete pairwise-plot has been attached in the Appendix. Below are 3 main observations (concerns) and their solutions



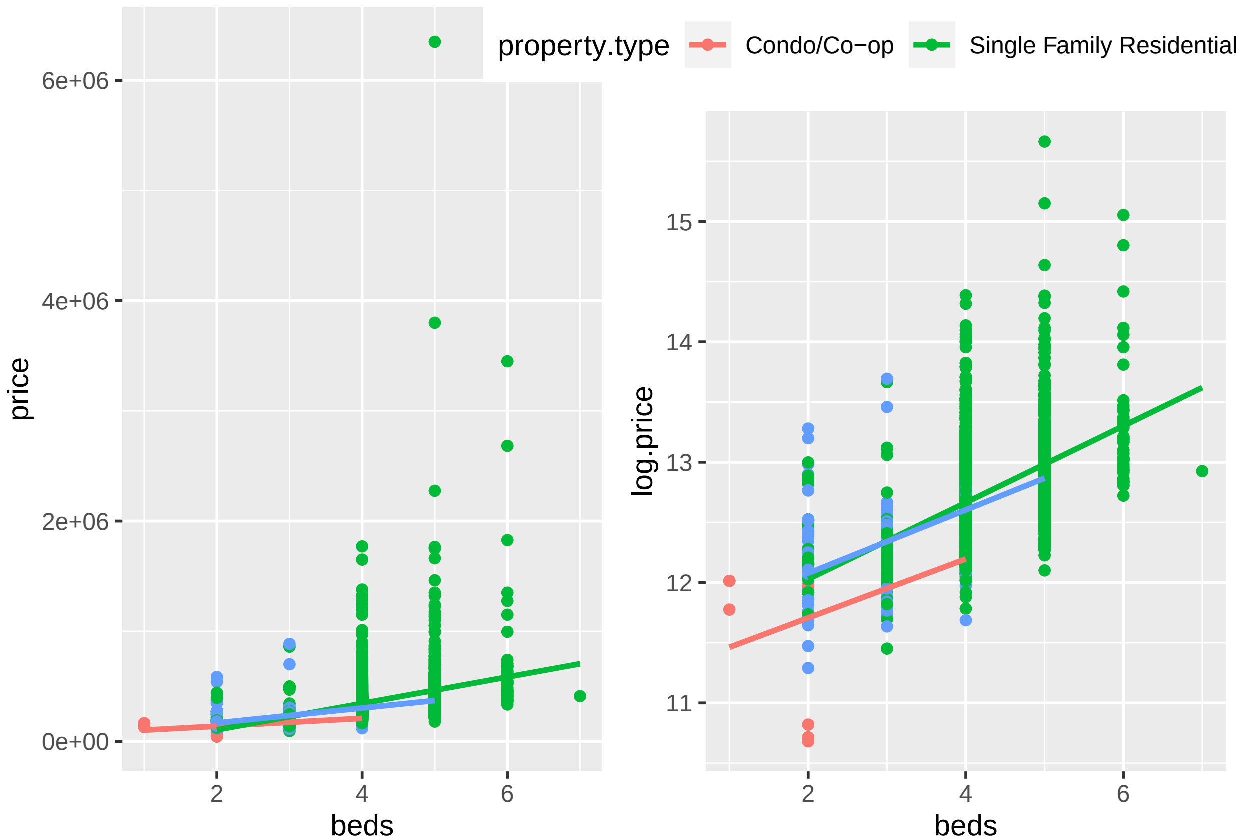
Multi-collinearity

Though increased number of beds in the house need not imply the increase of square feet, increasing number of baths in the house does imply the increase square feet more directly. Notice that in figure(________), a strong collinearity is shown between the number of baths and square feet of the house, achieving an correlation of 0.7843. Thus, we should be careful in the final model output for these highly correlated covariates More pairwise distribution between variables can be found in pair plot shown at appendix (________)

```
## [1] 0.7843887
```

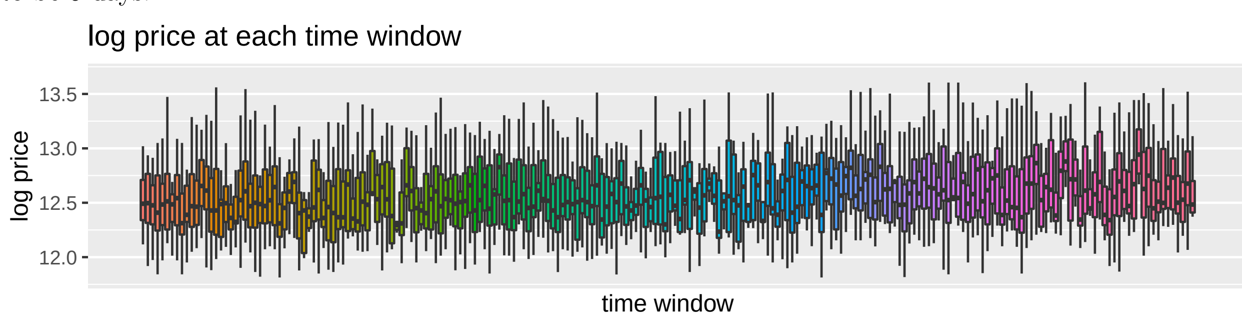
Heteroskedasticity

While some strong linearity and positive correlation is evident between some predictor variables, such as `beds`, the number of beds, and `square.feet`, the usable area of the house, accompanied with the increase in these predictor variables is the increase of variance. This violates the linear regression monoskedasticity assumption. To address this, we perform log-transformation on response variable and create response variable `log.price`. Shown in the last line of the pair-plot, heteroskdasticity problem is sufficiently solved without harming positive correlation between the original response and predictors. Furthermore, distribution supports a stronger linearity becomes transformed `prices`, which is `log.price`, and its predictors.



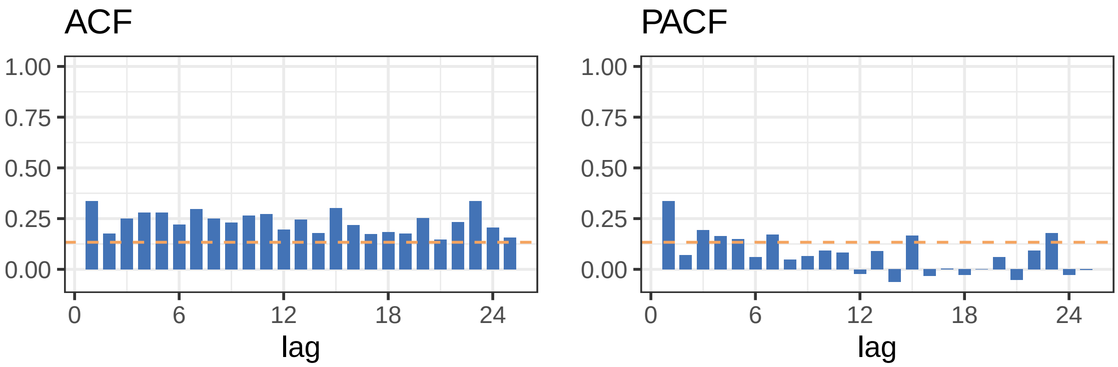
Stationarity

When determining the model for temporal effect, it is imperative to verify whether the stationarity assumption has been satisfied. In our case, due to the fact it is very unlikely that there are houses are sold everyday, we're create out time series by window period – that is, we treat all the real estate deals that lie in a prespesified window as deals happening at the same timestep. By specifying the width of our window, we can modulate and balance the amount of information to make better inference of exogenou variables (regression coefficients) versus the flexibility and variability of temporal effect. At this point, we choose window width to be 5 days.



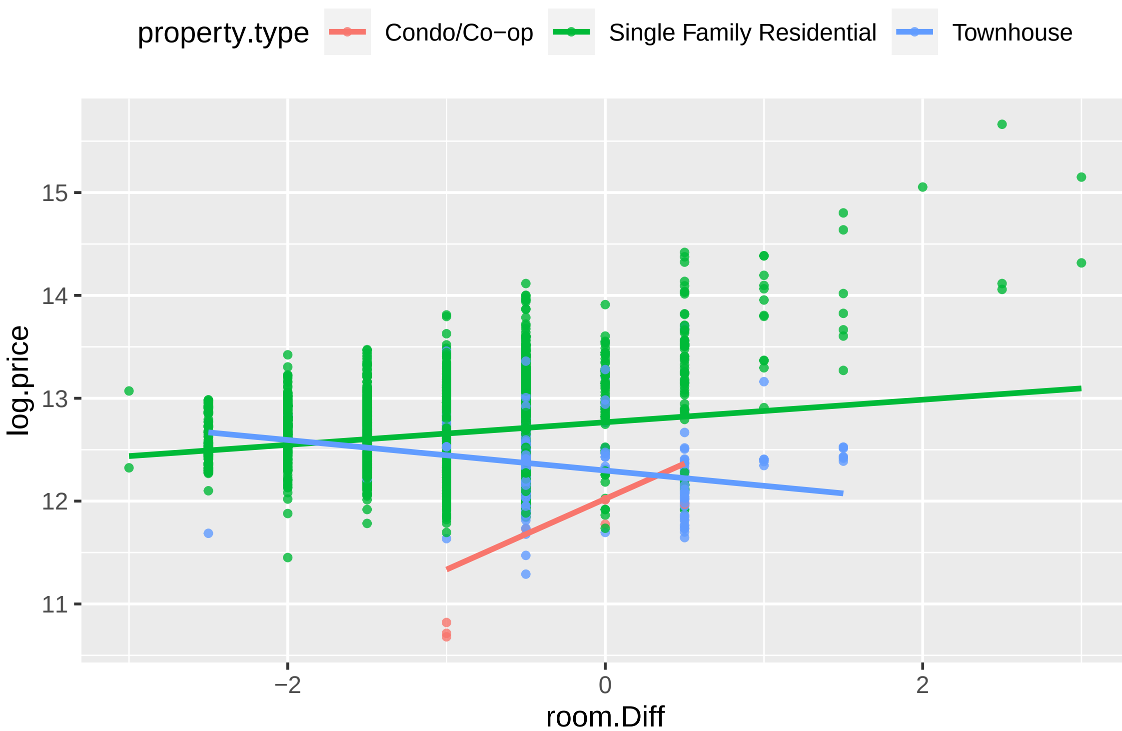
From the plot in figure (________), we found that there is a slight upward trend across in the distribution of logarithm of house prices. Though this distribution is marginalized by other exogenous variable, but it is helpful for us identify that the time the house is sold: `sold.dat` should also be considered as a potential variable in the regression. Note that it is also possible to add in moving average (MA) term in the termporal effect to model such trend, we choose to simplify this by problem by include it into the regression effect, and only including Autoregressive factors in the temporal effect.

Besides, it would be very helpful to identify the AR(p) terms by looking checking ACF and PACF plots. Below in figure (______) shows the plots mentioned. We observe that lag 1,3,4,5,7 are statistically significantly correlated with the term. Therefore, we will start incorporating 7 autocorrelation terms in our model. Incorporating less or more terms will incorporated in the model validation section.

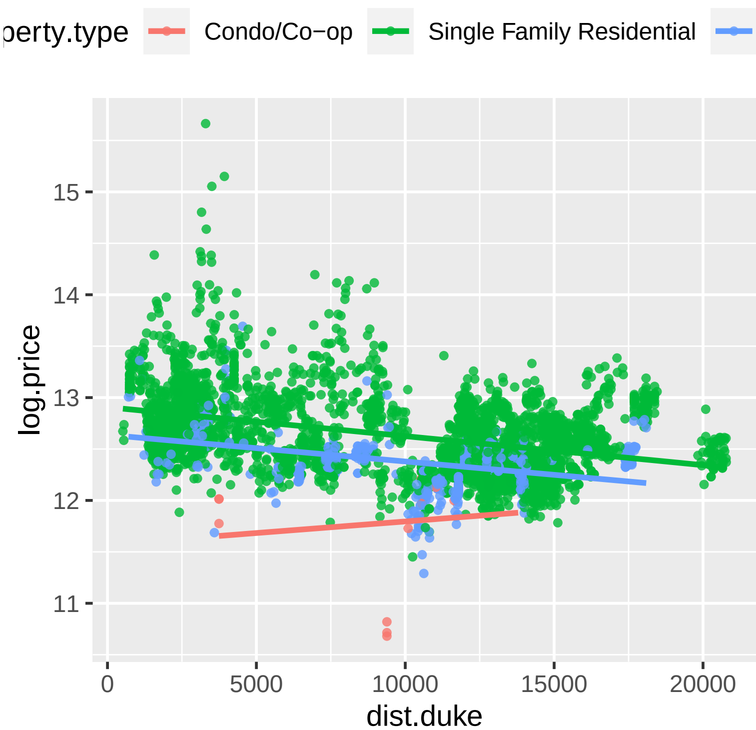


Engineered Feature and Interaction

We're interested in the difference between number of beds and baths and its relationship with logarithm of house price `log.price`. By creating new variable `room.Diff`, which means how much more baths does the house have than beds, we found such feature creates distinct effects across Single Family Residential, Townhouse, and Condo in affecting log of price. In figure (________), we can observe heterogeneity across 3 type of houses. Therefore, interaction between the difference and house type should also be added to our model.



Besides, we're interested in incorporating simple spatial information for additional prediction power. Therefore, we engineered the new covariate `dist.duke`, indicating the distance of the house to Duke. This is approximately calculated via shortest distance of 2 points on a ellipsoid indicated by latitude and longtitude of the house and Duke University. Then, as in figure (_________) shows, we anticipate longer distance to Duke will result in a lower housing price.



Methodology

We define our model as a regression model on top of a AR(P) model.

Regression (Observation) Model

After EDA, we determines to use response as `log.price`: the logarithm of the house deal price. The covariate predictors are `beds`, `sold.dat`, `baths`, `square.feet`, `lot.size`, `house.age`, `property.type`, `city`, `room.Diff`, `dist.Duke`, and finally `room.Diff` interact with `property.type`. The regression model is

$$
\begin{aligned}
y_{t_i} &= \beta_{\text{beds}}\ \textbf{beds}_{t_i} + \beta_{\text{sold date}}\ \textbf{sold date}_{t_i} + \beta_{\text{beds}}\ \textbf{baths}_{t_i} + \beta_{\text{square feet}}\ \textbf{square feet}_{t_i} + & (1)\\
&\beta_{\text{lot size}}\ \textbf{lot size}_{t_i} + \beta_{\text{house age}}\ \textbf{house age}_{t_i} + \beta_{\text{property type}}\ \textbf{property type}_{t_i} + & (2)\\
&\beta_{\text{city}}\ \textbf{city}_{t_i} + \beta_{\text{room difference}}\ \textbf{room Diff}_{t_i} + \beta_{\text{dist. to Duke}}\ \textbf{dist. to Duke}_{t_i} + & (3)\\
&\beta_{\text{interaction}}\ \textbf{room Diff}_{t_i} \times \textbf{property type}_{t_i} + & (4)\\
&\alpha_t + \nu_t & (5)\\
\nu_t &\sim \mathcal{N}(0, v) & (6)
\end{aligned}
$$

Where y_{t_i} is the response of the i^{th} house sold on the t^th window date, which is its logarithm of deal price. (notice that suppose for each window $t \in \{1, 2, 3, T\}$, there are n_t sold houses in the t^{th} window. Then n_t need not equal for all t). The rests are predictive variables. α_t is an time varying intercept which will be modeled by the AR(P) model described in the next session. ν_t is an additional observation uncertainty. To simply our modeling process, we take ν_t to have constant variance. Also, to simplify our notation, we write vectorized equation by merging line $(1), (2), (3), (4)$. The compact form is denoted as

$$
\begin{aligned}
\boldsymbol{y}_t &= \alpha_t \mathbf{1}_{n_t} + \boldsymbol{X}_t \boldsymbol{\beta} + \boldsymbol{\nu}_t\\
\boldsymbol{\nu}_t &\sim \mathcal{N}(\mathbf{0}, v\boldsymbol{I}_{n_t})
\end{aligned}
$$

Time Series Model

We construct the AR(p) model to model the underlying intercept α_t as described above in the regression model. As we've already indicated in EDA section, we'll choose $p = 7$ for our

$$\begin{aligned}\alpha_t &= \sum_{i=1}^{7} \theta_i \alpha_{t-i} + \omega_t \\ \omega_t &\sim \mathcal{N}(0, w)\end{aligned}$$

However, the above parametrization requires us to take-in many timesteps value to predict α_t, we may simply it by vectorizing the expression into the following:

$$\boldsymbol{\alpha}_t = \begin{bmatrix} \alpha_t \\ \alpha_{t-1} \\ \alpha_{t-2} \\ \vdots \\ \alpha_{t-7} \end{bmatrix} = \begin{bmatrix} \theta_1 & \theta_2 & \theta_3 & \dots & \theta_7 \\ 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} \alpha_{t-1} \\ \alpha_{t-2} \\ \alpha_{t-3} \\ \vdots \\ \alpha_{t-7} \end{bmatrix} = \boldsymbol{\Theta} \boldsymbol{\alpha}_{t-1}$$

In this way, transition becomes easy, as dependency rely on only the past one timestep.

Combined Model

We end up with the model as the following

$$\begin{aligned}\boldsymbol{\alpha}_t &= \boldsymbol{\Theta} \boldsymbol{\alpha}_{t-1} + \boldsymbol{W}_t \\ \boldsymbol{y}_t &= \mathbf{1} \boldsymbol{\alpha}_t + \boldsymbol{\beta} \boldsymbol{X}_t + \boldsymbol{\nu}_t \\ \boldsymbol{W}_t &\sim \mathcal{N}(\mathbf{0}, w\boldsymbol{I}) \\ \boldsymbol{\nu}_t &\sim \mathcal{N}(\mathbf{0}, v\boldsymbol{I}) \\ \mathbf{1} &:= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}\end{aligned}$$

Parameter Inference

The Model requires statistical inference uppon the following parameters:

$$\{w, v, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\alpha}_{1:T}\}$$

Algorithm 1: parameter inference algorithm

Result: A short algorithm of initializing reservoir weights with insured Echo State Property. Through empirical experiments, we recommend setting $\eta_1 = 0.97, \eta_2 = 0.85, \boldsymbol{\mu} = -2, \epsilon = 1$.

Initialize: $\mathbb{P}((\theta_1, \ldots, \theta_7)^\top), \mathbb{P}(w), \mathbb{P}(v)$ via pre-set prior distribution

while *not converged* **do**

  Calculate: posterior mean and covariance for $\boldsymbol{\alpha}_t|\mathcal{D}_t$: $m_t, C_t \ \forall t \in 1:T$ via forward filtering algorithm in Appendix

  Sample: $\boldsymbol{\alpha}_t|\mathcal{D}_T$ from $m_t^*, C_t^*, \forall t \in 1:T$ by backward smoothing

  Sample: $\boldsymbol{\theta}, \phi = w^{-1}|\boldsymbol{X}, \mathcal{D}_T, \boldsymbol{\beta}$ by first sampling $\phi = w^{-1}|\boldsymbol{X}, \mathcal{D}_T, \boldsymbol{\beta}$ and then $\boldsymbol{\theta}|\boldsymbol{X}, \mathcal{D}_T, \boldsymbol{\beta}, \phi$

  Sample: $\boldsymbol{\beta}|\boldsymbol{X}, \kappa, \boldsymbol{y}_t, \tau = v^{-1}$ and then sample $\tau = v^{-1}|\boldsymbol{X}, \kappa, \boldsymbol{y}_t, \boldsymbol{\beta}$

end

Return: $w, v, \boldsymbol{\beta}, \boldsymbol{\theta}_{1:T}$

Bibliography

[1] https://www.dallasfed.org/-/media/documents/institute/wpapers/2014/0208.pdf

[2], https://www.kaggle.com/manisaurabh/house-prices-advanced-regression-technique

[3], https://rady.ucsd.edu/faculty/directory/valkanov/pub/docs/HandRE_GPTV.pdf

[4]. https://medium.com/@feraguilari/time-series-analysis-modfinalproyect-b9fb23c28309

[5]. https://www.redfin.com/

Appendix

Parameter Inference

Forward Filtering

$$
\begin{aligned}
\boldsymbol{\alpha}_t &= \boldsymbol{\Theta}\boldsymbol{\alpha}_{t-1} + \boldsymbol{W}_t \\
\boldsymbol{y}_t &= \mathbf{1}\boldsymbol{\alpha}_t + \boldsymbol{\beta}\boldsymbol{X}_t + \boldsymbol{\nu}_t \\
\boldsymbol{W}_t &\sim \mathcal{N}(\mathbf{0}, w\boldsymbol{I}) \\
\boldsymbol{\nu}_t &\sim \mathcal{N}(\mathbf{0}, v\boldsymbol{I}) \\
\mathbf{1} &:= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}
\end{aligned}
$$

First, denote that

$$
\begin{aligned}
\boldsymbol{\alpha}_t|\mathcal{D}_t, - &\sim \mathcal{N}(m_t, C_t) \\
\boldsymbol{\alpha}_{t+1}|\mathcal{D}_t, - &\sim \mathcal{N}(a_{t+1}, R_{t+1})
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{\alpha}_t|\mathcal{D}_{t-1}, - &\sim \mathcal{N}(\boldsymbol{\Theta}m_{t-1}, \boldsymbol{\Theta}^T C_{t-1}\boldsymbol{\Theta} + w\boldsymbol{I}) = \mathcal{N}(a_t, R_t) \\
\boldsymbol{y}_t|\boldsymbol{\alpha}_t, \mathcal{D}_{t-1}, - &\sim \mathcal{N}(\mathbf{1}\boldsymbol{\alpha}_t + \boldsymbol{X}_t\boldsymbol{\beta}, v\boldsymbol{I}) \\
\mathbb{P}(\boldsymbol{\alpha}_t|\mathcal{D}_t) &\propto \mathbb{P}(\boldsymbol{\alpha}_t|\mathcal{D}_{t-1}, -)\mathbb{P}(\boldsymbol{y}_t|\boldsymbol{\alpha}_t, \mathcal{D}_{t-1}, -) \\
&\propto \exp\left\{-\frac{1}{2}\left[\boldsymbol{\alpha}_t^T(R_t^{-1} + v^{-1}\mathbf{1}^T\mathbf{1})\boldsymbol{\alpha} - 2\boldsymbol{\alpha}_t^T(R_t^{-1}a_t + v^{-1}\mathbf{1}^T(\boldsymbol{y}_t - \boldsymbol{X}_t\boldsymbol{\beta}))\right]\right\} \\
&\sim \mathcal{N}\left(\left(R_t^{-1} + v^{-1}\mathbf{1}^T\mathbf{1}\right)^{-1}(R_t^{-1}a_t + v^{-1}\mathbf{1}^T(\boldsymbol{y}_t - \boldsymbol{X}_t\boldsymbol{\beta})), \left(R_t^{-1} + v^{-1}\mathbf{1}^T\mathbf{1}\right)^{-1}\right) \\
&= \mathcal{N}(m_t, C_t)
\end{aligned}
$$

$$
\begin{aligned}
a_t &= \boldsymbol{\Theta}m_{t-1} \\
R_t &= \boldsymbol{\Theta}^T C_{t-1}\boldsymbol{\Theta} + w\boldsymbol{I} \\
m_t &= \left(R_t^{-1} + v^{-1}\mathbf{1}^T\mathbf{1}\right)^{-1}(R_t^{-1}a_t + v^{-1}\mathbf{1}^T(\boldsymbol{y}_t - \boldsymbol{X}_t\boldsymbol{\beta})) \\
C_t &= \left(R_t^{-1} + v^{-1}\mathbf{1}^T\mathbf{1}\right)^{-1}
\end{aligned}
$$

Use this equation to update

Backward Smoothing

Sh*t, I hate this...

Suppose we already know that

$$
\mathbb{P}(\boldsymbol{\alpha}_{t+1}|\mathcal{D}_T) \sim \mathcal{N}(m_{t+1}^*, R_{t+1}^*)
$$

Let's look at log likelihood of $\boldsymbol{\alpha}_t, \boldsymbol{\alpha}_{t+1}|\mathcal{D}_T$. Using conditional independence, we have

$$
\begin{aligned}
-\frac{1}{2}\ell(\boldsymbol{\alpha}_t, \boldsymbol{\alpha}_{t+1}; \mathcal{D}_T) &= \log \mathbb{P}\left(\boldsymbol{\alpha}_{t+1} \mid \boldsymbol{\alpha}_t\right) + \log \mathbb{P}\left(\boldsymbol{\alpha}_t \mid \mathcal{D}_t\right) - \log \mathbb{P}\left(\boldsymbol{\alpha}_{t+1} \mid \mathcal{D}_t\right) + \log \mathbb{P}\left(\boldsymbol{\alpha}_{t+1} \mid \mathcal{D}_T\right) \\
&= (w)^{-1}(\boldsymbol{\alpha}_{t+1} - \boldsymbol{\Theta}\boldsymbol{\alpha}_t)^T(\boldsymbol{\alpha}_{t+1} - \boldsymbol{\Theta}\boldsymbol{\alpha}_t) + (\boldsymbol{\alpha}_t - m_t)^T(C_t)^{-1}(\boldsymbol{\alpha}_t - m_t) - \\
&\quad (\boldsymbol{\alpha}_{t+1} - a_{t+1})^T(R_{t+1})^{-1}(\boldsymbol{\alpha}_{t+1} - a_{t+1}) + \\
&\quad (\boldsymbol{\alpha}_{t+1} - m^*_{t+1})^T(C^*_{t+1})^{-1}(\boldsymbol{\alpha}_{t+1} - m^*_{t+1}) + \text{constant} \\
&= \boldsymbol{\alpha}_{t+1}^T({C^*_{t+1}}^{-1} + w^{-1}\boldsymbol{I} + R_{t+1}^{-1})\boldsymbol{\alpha}_{t+1} + \boldsymbol{\alpha}_t^T(w^{-1}\boldsymbol{\Theta}^T\boldsymbol{\Theta} + C_t^{-1})\boldsymbol{\alpha}_t + \\
&\quad 2\boldsymbol{\alpha}_{t+1}^T(-w^{-1}\boldsymbol{\Theta})\boldsymbol{\alpha}_t - 2\boldsymbol{\alpha}_t^T(C_t^{-1}m_t) - 2\boldsymbol{\alpha}_{t+1}^T(R_{t+1}^{-1}a_{t+1} + {C^*_{t+1}}^{-1}m^*_{t+1}) + \text{constant}
\end{aligned}
$$

One eternity of calculation later, we end up with:

$$
\begin{aligned}
J_t &= C_t\boldsymbol{\Theta}(\boldsymbol{\Theta}C_t\boldsymbol{\Theta}^T + w\boldsymbol{I})^{-1} \\
m^*_t &= m_t + J_t(m^*_{t+1} - \boldsymbol{\Theta}m_t) \\
C^*_t &= C_t + J_t(C^*_{t+1} - \boldsymbol{\Theta}C_t\boldsymbol{\Theta}^T - w\boldsymbol{I})J_t^T
\end{aligned}
$$

Dynamic model sampling: $(\theta_1, \cdots, \theta_7)^\intercal$**,** $w = \phi^{-1}$

This is a simple linear regression $\boldsymbol{\alpha} = \mathbf{X}\boldsymbol{\theta} + w_t$ with design matrices as

$$
\boldsymbol{y} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \\ \vdots \\ \alpha_T \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} \alpha_{1-1} & \cdots & \alpha_{1-7} \\ \alpha_{2-1} & \cdots & \alpha_{2-7} \\ \alpha_{3-1} & \cdots & \alpha_{3-7} \\ \alpha_{4-1} & \cdots & \alpha_{4-7} \\ \alpha_{5-1} & \cdots & \alpha_{5-7} \\ \vdots & \vdots & \vdots \\ \alpha_{T-1} & \cdots & \alpha_{T-7} \end{bmatrix} \qquad \boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ \theta_5 \\ \theta_6 \\ \theta_7 \end{bmatrix}
$$

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{y}; \boldsymbol{\theta}, \mathbf{X}) &\propto \phi^{\frac{T}{2}} \exp\{-\frac{1}{2}\phi(\boldsymbol{y} - \mathbf{X}\boldsymbol{\theta})^\intercal(\boldsymbol{y} - \mathbf{X}\boldsymbol{\theta})\} \\
\boldsymbol{\theta}|\phi, \mathcal{D}_T, \boldsymbol{\beta}, v &\sim \mathcal{N}\left(\boldsymbol{\mu}_0, \boldsymbol{\Lambda}_0^{-1}/\phi\right) = \mathcal{N}((0.5, 0.5, 0.5)^T, 1/3\phi^{-1}\mathbf{I}) \\
\phi|\mathcal{D}_T, \boldsymbol{\beta}, v &\sim \mathbf{G}\left(a_0 = \frac{v_0}{2}, b_0 = \frac{v_0 s_0^2}{2}\right) = \mathbf{G}\left(\frac{1}{2}, \frac{1}{2}\right) \\
\boldsymbol{\mu}_n &= \left(\mathbf{X}^{\mathrm{T}}\mathbf{X} + \boldsymbol{\Lambda}_0\right)^{-1}\left(\boldsymbol{\Lambda}_0\boldsymbol{\mu}_0 + \mathbf{X}^{\mathrm{T}}\boldsymbol{y}\right) \\
\boldsymbol{\Lambda}_n &= \left(\mathbf{X}^{\mathrm{T}}\mathbf{X} + \boldsymbol{\Lambda}_0\right) \\
a_n &= a_0 + \frac{T}{2} \\
b_n &= b_0 + \frac{1}{2}(\boldsymbol{y}^\intercal\boldsymbol{y} + \boldsymbol{\mu}_0^\intercal\boldsymbol{\Lambda}_0\boldsymbol{\mu}_0 - \boldsymbol{\mu}_n^\intercal\boldsymbol{\Lambda}_n\boldsymbol{\mu}_n) \\
\boldsymbol{\theta}|\phi, \boldsymbol{X}, \mathcal{D}_T, \boldsymbol{\beta}, v &\sim \mathcal{N}\left(\boldsymbol{\mu}_n, \boldsymbol{\Lambda}_n^{-1}/\phi\right) \\
\phi|\boldsymbol{X}, \mathcal{D}_T, \boldsymbol{\beta}, v &\sim \mathbf{G}\left(a_n, b_n\right)
\end{aligned}
$$

Observation Model Sampling $(\beta_1 \cdots)^\intercal$**,** $v = \tau^{-1}$

Very similar as above, this is also a linear model. Besides, it is possible to apply Bayesian Ridge here. let's create the Bayesian ridge model

$$
\begin{aligned}
\boldsymbol{y}_t &= \boldsymbol{\alpha}_t \mathbf{1} + \boldsymbol{X}\boldsymbol{\beta} + \nu_t \\
\boldsymbol{z}_t = (\boldsymbol{y}_t - \boldsymbol{\alpha}_t \mathbf{1}) \mid \boldsymbol{\alpha}_t, \boldsymbol{\beta}, \tau &\sim \mathrm{N}\left(\mathbf{X}\boldsymbol{\beta}, \mathbf{l}_n/\tau\right) \\
\boldsymbol{\beta} \mid \tau, \kappa &\sim \mathrm{N}\left(\mathbf{0}, \mathbf{l}(\tau\kappa)^{-1}\right) \\
p(\tau \mid \kappa) &\propto 1/\tau
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{P}(\boldsymbol{y}_t - \boldsymbol{\alpha}_t \mathbf{1} | \boldsymbol{X}, \boldsymbol{\beta}, \tau, \boldsymbol{\alpha}) &\propto \tau^{\frac{n}{2}} \exp\{-\frac{\tau}{2}(\boldsymbol{z}_t - \boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{z}_t - \boldsymbol{X}\boldsymbol{\beta})\} \\
\mathbb{P}(\boldsymbol{\beta}|\boldsymbol{X}, \boldsymbol{z}_t, \tau, \boldsymbol{\alpha}) &\propto \mathbb{P}(\boldsymbol{z}_t|\boldsymbol{X}, \boldsymbol{\beta}, \tau, \boldsymbol{\alpha})\mathbb{P}(\boldsymbol{\beta}|\tau, \kappa, \boldsymbol{\alpha})\mathbb{P}(\tau|\kappa, \boldsymbol{\alpha}) \\
&\propto \tau^{\frac{n}{2}} \exp\{-\frac{\tau}{2}(\boldsymbol{z}_t - \boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{z}_t - \boldsymbol{X}\boldsymbol{\beta})\}(\tau\kappa)^{\frac{p}{2}} \exp\left\{-\frac{\tau\kappa}{2}\boldsymbol{\beta}^T\boldsymbol{\beta}\right\}\tau^{-1} \\
&\propto \exp\left\{-\frac{1}{2}\left[\boldsymbol{\beta}^T(\tau\boldsymbol{X}^T\boldsymbol{X} + \tau\kappa\mathbf{1})\boldsymbol{\beta} - 2\tau\boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{z}_t\right]\right\} \\
&\sim N\left((\boldsymbol{X}^T\boldsymbol{X} + \kappa\mathbf{1}_p)^{-1}\boldsymbol{X}^T\boldsymbol{z}_t, \tau^{-1}(\boldsymbol{X}^T\boldsymbol{X} + \kappa\mathbf{1}_p)^{-1}\right) \\
\boldsymbol{\beta}|\boldsymbol{X}, \boldsymbol{z}_t, \tau, \boldsymbol{\alpha} &= \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \\
\mathbb{P}(\tau|\boldsymbol{\beta}, \kappa, \boldsymbol{z}_t, \boldsymbol{\alpha}) &\propto \tau^{\frac{n+p}{2}-1} \exp\left\{-\frac{\tau\kappa}{2}\boldsymbol{\beta}^T\boldsymbol{\beta}\right\} \\
\tau|\boldsymbol{\beta}, \kappa, \boldsymbol{z}_t, \boldsymbol{\alpha} &\sim \boldsymbol{G}(\frac{n+p}{2}, \frac{\kappa}{2}\boldsymbol{\beta}^T\boldsymbol{\beta})
\end{aligned}
$$

{Last Revised: September 13, 2020}