

# Defining Pseudoresiduals for Multivariate Gaussian HMMs

Tazman Libson

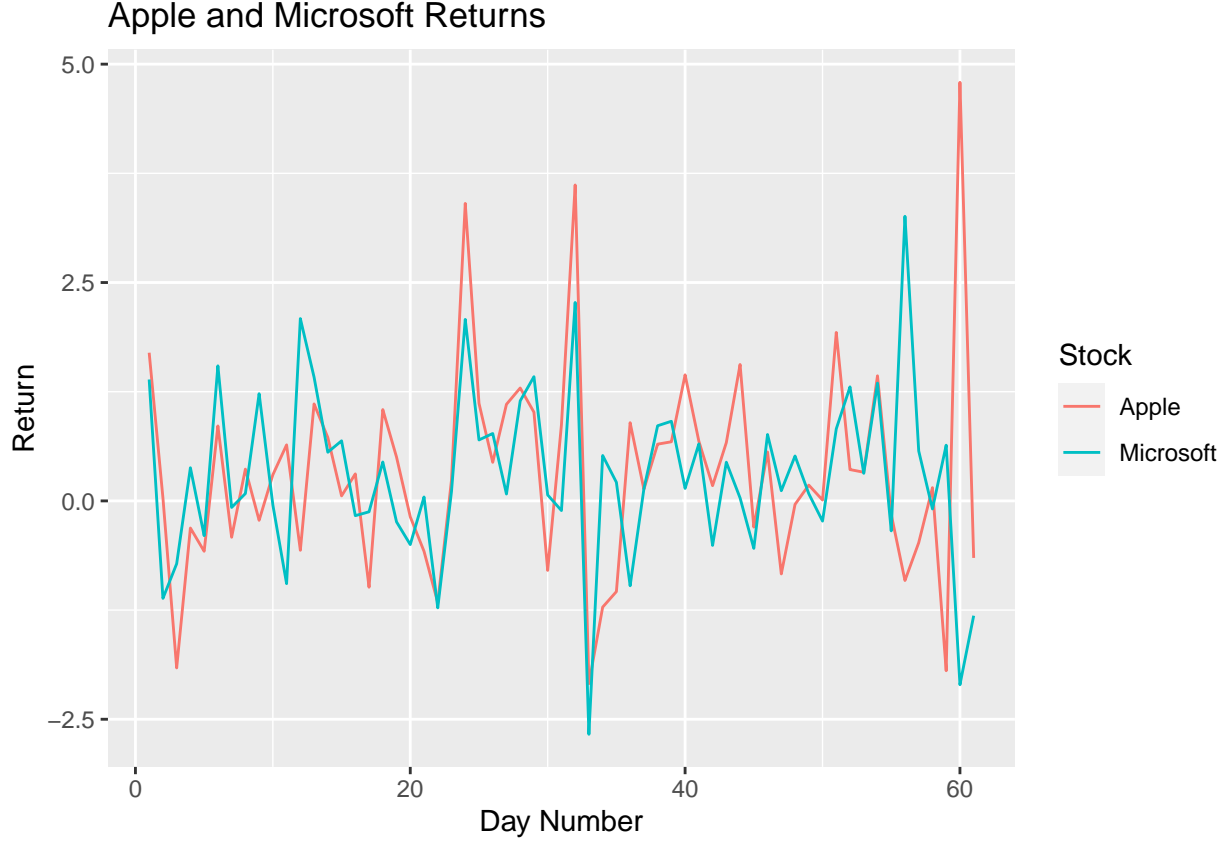
2024-03-25

## Introduction

Hidden Markov models (HMMs) are a class of stochastic models with a wide range of applications in Ecology[1], Voice Recognition, Finance[2], and many other fields. They are most often used with either temporal or spacial data. There are two main components for a hidden Markov model, an unobserved or hidden process, and an observed variable. The unobserved process is a discrete set of states, more specifically described by a Markov chain. The interpretation of these states depend on the application of the model. For example, in finance, an interpretation for a 3 state model would be a bearish/growth state, a bullish/decay state, and a mixed middle state. The key feature of these states is that they cannot be directly observed.

Each of these states have an associated state dependent distribution. These distributions are what determine the values of the observed variable. Again in a financial context, these observed variable could be the returns of a stock, interest rates, or any value one could consistently measure over a period of time.

For this report, daily returns data for 4 companies: Apple, Microsoft, Meta, and Intel between Feb 1st 2019 and Feb 1st 2024 were used. For the data, the returns are defined as  $100 \log(s_t/s_{t-1})$ , where  $s_t$  is the price on day  $t$ . This data was downloaded from <https://www.nasdaq.com/market-activity/stocks/> on Feb 1st, 2024. The Hidden Markov models fitted will be fitting multivariate data. So they will be modelling all four of these stocks simultaneously. This will allow the models to take into account correlations between stocks. As seen in the small slice of the returns for Microsoft and Apple plotted below, the two return values rise and fall at similar days. In financial modelling, stocks are often correlated, and have changing correlation over time.[3] Multivariate hidden markov models will accommodate for this.



Using HMMs in finance certainly is not a new development, however, using HMMs and the typical HMM model diagnostics to fit multivariate data leads to some issues. Specifically, this report is going to investigate how using multivariate data affects the pseudoresiduals for HMMs.

## Model Definition

### Unobserved Markov Chain

The unobserved process described by a Markov chain. This chain will have  $m$  states. The state at time  $t$  will be written as  $c_t$ . These states are discrete. The probability from going to state  $j$  from state  $i$  at time  $t$  will be referred to as  $p_{ij}$ . These transition probability matrix are used to make the one step transition probability matrix.

$$\Gamma = [p_{ij}]; \quad p_{ij} = \mathbb{P}[c_t = j | c_{t-1} = i]$$

For all our cases, the markov chains will all be homogeneous, meaning that  $\Gamma$  is independent of time. This assumption is to simplify the model and is required for the likelihood calculation later, and thus the entire model fitting process.

The transition probability matrix will allow for the expected evolution of states to be calculated given an initial state distribution. A state distribution  $\mathbf{u}_t$  is a vector of length  $m$  where  $u_i$  is the probability at time  $t$  to for  $c_t = i$ .

$$\mathbf{u}_t = \{u_1, u_2, \dots, u_m\}, \quad u_i = \mathbb{P}[c_t = i]$$

To get the state distribution for the following time, one takes the product of the the current state distribution and the transition probability matrix.

$$\mathbf{u}_t \Gamma = \mathbf{u}_{t+1}$$

The stationary distribution of a markov chain is a state distribution where, when multiplied with the transition probability matrix, remains unchanged. It is defined as follows:

$$\mathbf{u}\Gamma = \mathbf{u}$$

It can be shown that every markov chain that will be used in our models has a stationary distribution.

For model specification, one has to either indicate a specific initial distribution, or indicate that the model is stationary, and the stationary distribution will be used as the initial distribution.

### State Dependent Distributions

In our case, the state dependent distributions will be an  $n$ -dimension multivariate Gaussian. For every state  $i \in (1, \dots, m)$ , the state-dependent distribution will be defined as follows:

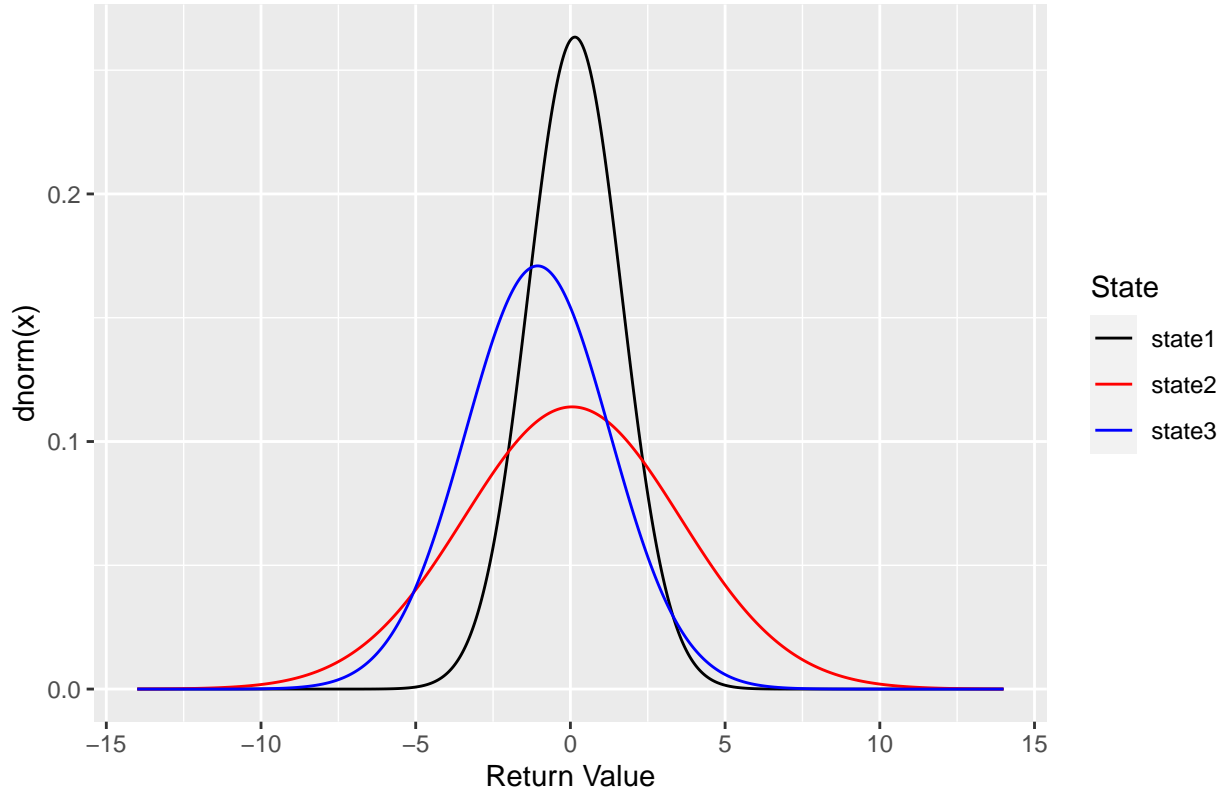
$$p_i(\mathbf{x}_t) = \mathbb{P}[\mathbf{X}_t = \mathbf{x}_t | c_t = i]$$

For every state  $i \in \{1, \dots, m\}$  the probability density function will have a  $n \times n$  variance covariance matrix  $\Sigma_i$  and a vector of means of length  $n$   $\mu_i$ .

$$\mathbf{X}_t | c_t = i \sim N(\mu_i, \Sigma_i)$$

Below one can see the marginalized probability density function for the apple returns of the fitted model. Although each state has a normal distribution, each one has a different mean and variance. For the technology stocks that were fitted for the model, each state has a different 4 dimensional normal distribution.

### Marinalized PDFs for Apple Returns



To summarise, each model needs to have the following parameters specified:

\* Transition Probability Matrix  $\Gamma$

- \* Initial Distribution/Stationary Distribution  $\delta$
- \* Means  $\mu_i$  and Variance Covariance Matrix  $\Sigma_i$  for each state  $i \in \{1, \dots, m\}$

Now let us go through how one can find values for these parameters.

## Model Fitting

Model fitting was done by maximum likelihood estimation, where the likelihood was calculated for an initial model, then using the nonlinear maximization function `nlm`, one finds the parameter values which finds the model with the largest likelihood.

## Likelihood Calculation

It can be shown that the likelihood of a series of observations from a HMM can be calculated recursively through matrix multiplication. Here  $\mathbf{X}^{(T)}$  indicates all the observations from  $t = 1, \dots, T$ .

$$L_T = \mathbb{P}[\mathbf{X}^{(T)} = \mathbf{x}^{(T)}] = \delta P(\mathbf{x}_1) \Gamma P(\mathbf{x}_2) \Gamma P(\mathbf{x}_3) \dots \Gamma P(\mathbf{x}_T) \mathbf{1}'$$

Where  $P(\mathbf{x}_i)$  is a diagonal matrix with entries:

$$p_i(\mathbf{x}_t) = \mathbb{P}[\mathbf{x} = \mathbf{x}_t | c_t = i]$$

The calculation takes into account the probability of each observation under each state-based distribution and is scaled by the probability of being in each state at the time of each observation. To avoid underflow, the log-likelihood is calculated.

## Reparameterization

The optimization functions `optim` and `nlm` both need unrestricted parameters. The means  $\mu$  are unrestricted, but otherwise, the other parameters, the variance covariance matrices and the transition probability matrix all have several restrictions. These parameters will have to be transformed. In the following descriptions, working parameters will be the reparameterized values, natural parameters will be the non transformed value.

**Transition Probability Matrix** The following method of reparameterization has been previously described [cite Z and M]. The values of the transition probability matrix are all between 0 and 1 inclusive. The rows also must all sum to 1. So there are only  $m(m - 1)$  free parameters. For the following, the diagonal elements will be considered to be 1 minus the sum of the other elements of their row, so the non diagonal elements are the ones being estimated. These elements are first transformed to be on the non-negative reals by dividing them by the diagonal elements. These values are then mapped to the entire real line by the log function. Here the natural transition probabilities will be  $p_{ij}$  and the working transition probabilities will be  $\tau_{ij}$ .

$$\tau_{ij} = \log\left(\frac{p_{ij}}{p_{ii}}\right), \quad i \neq j; \quad \tau_{ii} = 1 \quad \forall i \in \{1, \dots, m\}$$

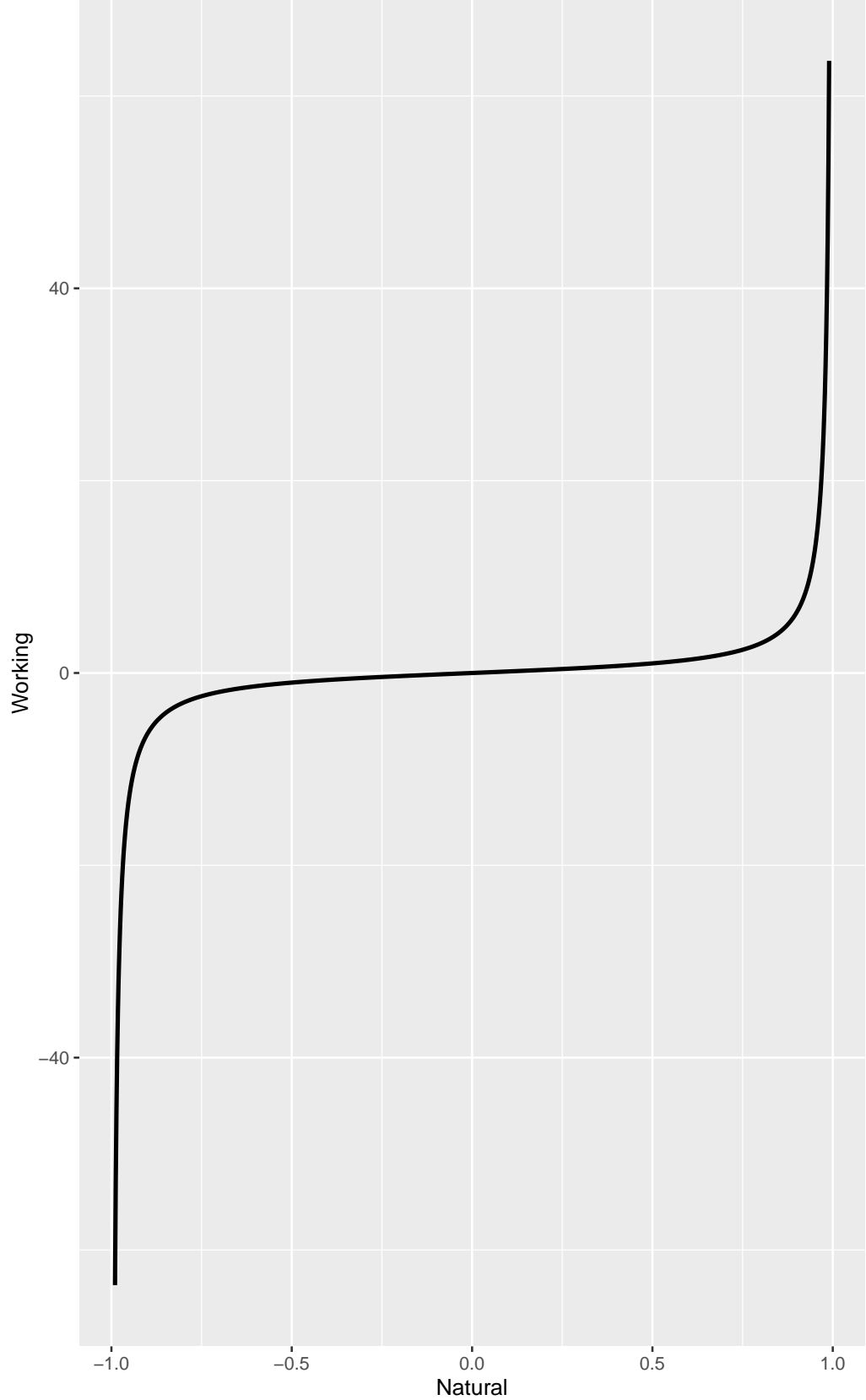
$$p_{ij} = \frac{\exp(\tau_{ij})}{1 + \sum_{i \neq k} \exp(\tau_{ik})}$$

**Variance Covariance Matrix** The restrictions on the variance covariance matrices make it inconvenient to directly transform the matrix into working parameters. The matrices is positive definite, meaning they are symmetric and its eigenvalues are positive. In addition, the square of all of the non-diagonal elements cannot be larger then the product of both of the variances of its row or column number (i.e. for covariance  $c_{ij} = \text{Cov}(X_i, X_j)$ ,  $[c_{ij}]^2 \leq \text{Var}(X_i)\text{Var}(X_j)$ ). This comes from the Cauchy-Schwartz inequality [Citation Needed]. Thus every non diagonal element has a unique range of possible values depending on the values of variances for each of the variables of the Gaussian. This problem is made easier by instead, estimating the variances and correlation matrices separately and then getting the variance covariance matrices from there.

The correlation matrices are all positive definite. In addition, the values on the diagonal are all 1, since each return always has a correlation of 1 with itself. The non-diagonal values all have values between -1 and 1 inclusive. With all of these restrictions, only the upper triangular values of the correlation matrix need to be estimated. A scaled tan function is used to map the correlation values onto the reals. Here the natural values will be  $k_{jh}^h$  and the working values will be  $\kappa_{ij}^h$  with  $ij$  indicating the position in the correlation matrix and  $i$  indicating the state for which the correlation is used.

$$\kappa_{jh}^i = \tan\left(\frac{\pi c_{jh}^i}{2}\right), \quad k_{jh}^i = \frac{2 \arctan(\kappa_{jh}^i)}{\pi}; h \neq j$$

Correlation Matrix Reparameterization



The variances are all strictly positive so they can be reparameterized by using a log function. The working variance will be  $\sigma_{ij}^2$ , the natural variance will be  $s_{ij}^2$ . Here  $i$  is one of the  $m$  states and  $j$  is one of the  $n$  observed variables.

$$\sigma_{ij} = \log(s_{ij}^2), \quad s_{ij}^2 = \exp(\sigma_{ij})$$

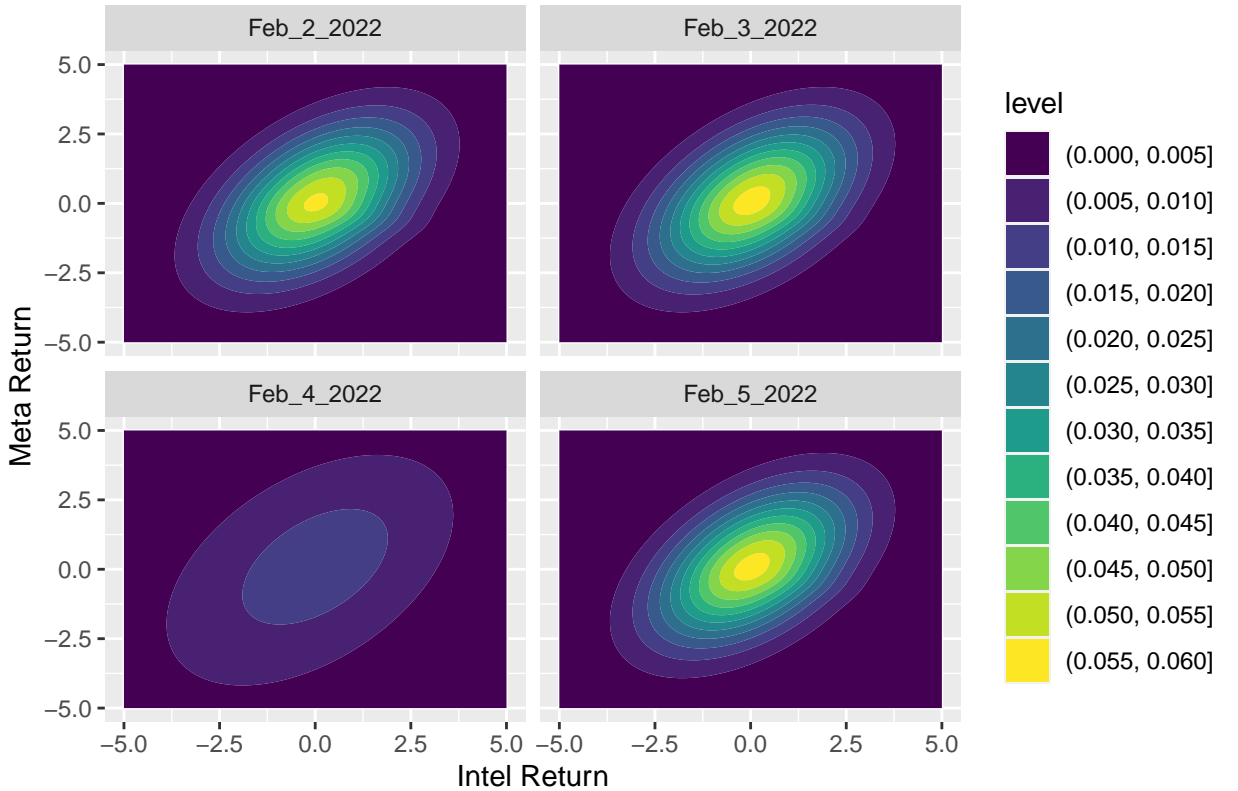
## Conditional Distributions

In order to describe pseudoresiduals, we first must define conditional distributions. First a to define a small bit of notation. The observation vector  $\mathbf{X}^{(-t)}$  will be defined as all of the observations from time 1,  $\dots, T$  except  $t$ . So it represents  $\{X_1, X_2, \dots, X_{t-1}, X_{t+1}, X_{t+2}, \dots, X_T\}$ . With that conditional distributions are defined as the following probability distribution:

$$\mathbb{P}[\mathbf{X}_t = \mathbf{x} | \mathbf{X}^{(-t)} = \mathbf{x}^{(-t)}]$$

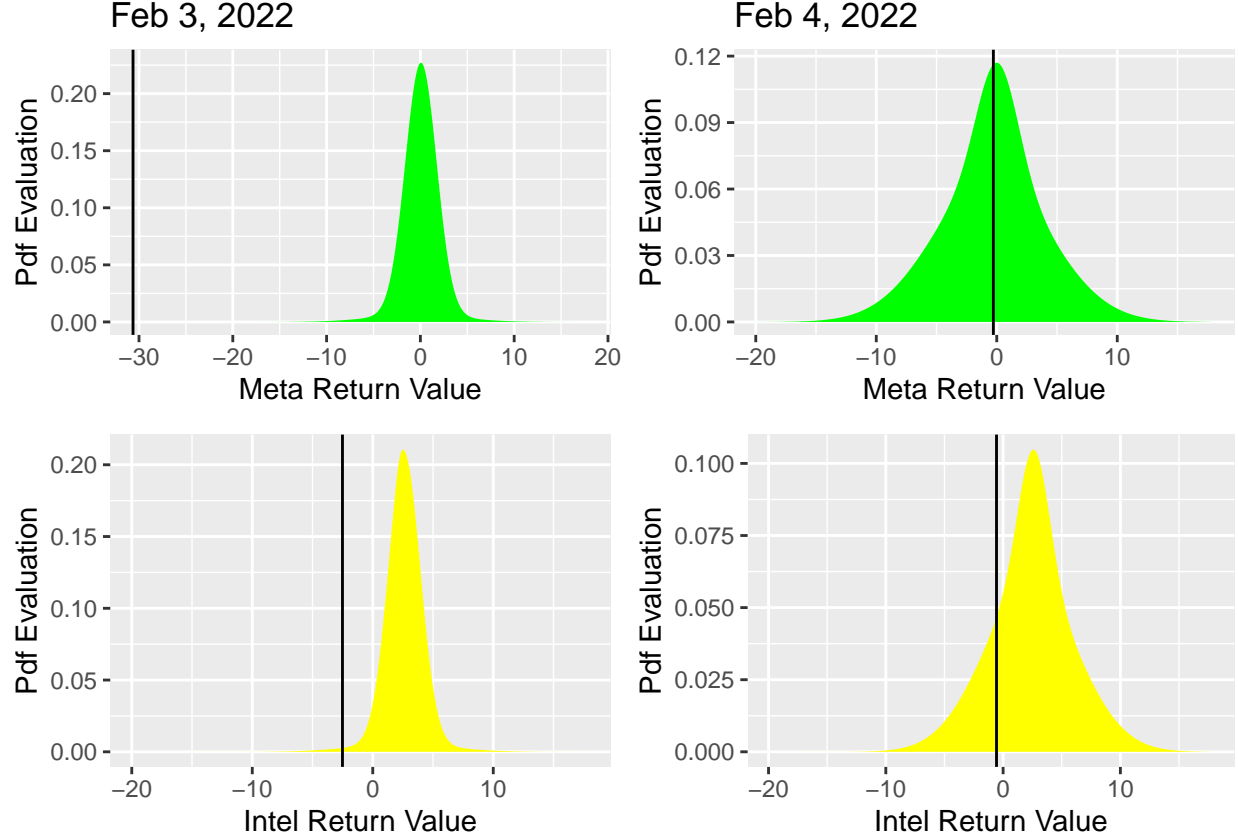
So conditional distributions is the probability distribution of  $\mathbf{X}_t | \mathbf{X}^{(-t)} = \mathbf{x}^{(-t)}$  or the probability of observing a specific value under the model at time  $t$  given every other observation. These distributions can act as the model's predicted distribution for each time. As seen below, the conditional distribution is different for each

### Maringal Conditional PDFs on Different Days



time.

Unfortunately the full conditional distributions for a given time cannot be easily displayed because of the dimensionality of the distribution. For the distributions above, the other two stocks have been marginalized out. In the figure, one can see that the conditional distributions change day by day. This is very noticeable for the distribution for February 4th where the distribution becomes significantly more disperse. This effect is seen for all stocks that were used in the fitted model.



So far, only conditional probability density functions have been shown. For pseudoresiduals, the conditional cumulative density functions will be needed. These will be indicated by  $F(X_t)$  and will be defined as follows:

$$F(X_t) = \mathbb{P}[X_t \leq x_t | \mathbf{X}^{(-t)} = \mathbf{x}^{(-t)}]$$

## PseudoResiduals

With these conditional distributions, we can now define pseudoresiduals. Pseudoresiduals are a tool for outlier checking and checking model fit. The definition that will be used here will be from the Zucchini and Macdonald textbook on Hidden Markov Models, which has been widely used in the use of HMMs.

First we will define uniform pseudoresiduals. Uniform pseudoresiduals are the conditional cumulative distribution function evaluated for every observation in the observation dataset. This will generate values that are between 0 and 1. The utility here is that for a well fitted model, these uniform pseudoresiduals should be uniformly distributed for a well fitted model.

$$F(X_t) \sim U(0, 1) \quad (\text{for a well fitted model})$$

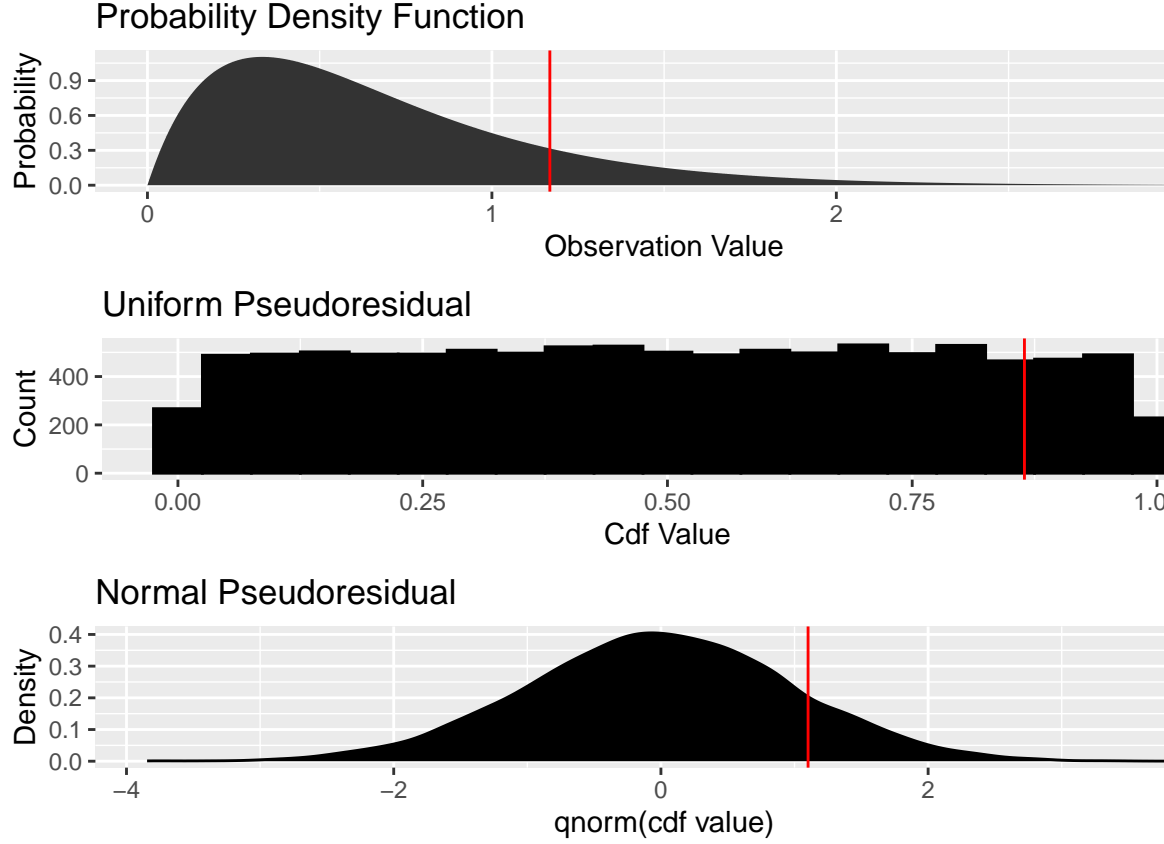
There are some issues with these values, firstly they are not well suited for outlier detection. Outliers here would be defined as values on the fringe of their respective conditional distributions. For uniform pseudoresiduals, outliers would be values close to 0 or 1, however values that would be considered outliers would be hard to distinguish from one another (e.g. distinguishing between 0.99 and 0.999)[\*].

With these uniform pseudoresiduals we can define normal pseudoresiduals. By taking the inverse standard normal function,  $\Phi^{-1}$ , of the uniform pseudoresiduals we get normal pseudoresiduals. This solves the outlier problem, where outliers are clearer, and tests for conformation to the standard normal are well defined[5].



$$\Phi^{-1}(F(X_t)) \sim N(0, 1) \quad (\text{for a well fitted model})$$

Below is a graphical representation of the process to get from a conditional distribution to a normal pseudoresidual. The red line indicates how an observation is transformed through the path to a normal pseudoresidual. The observation value is first inputted into a conditional cumulative density function. This yields a normal pseudoresidual. The inverse standard normal is then applied to the uniform pseudoresidual to get a



normal pseudoresidual.

From now on we will only work with normal pseudoresiduals, so for convenience they will be referred to as only pseudoresiduals. This definition has been assuming univariate data. Let us now proceed to defining pseudoresiduals for multivariate data.

The above definition has been clearly established by Zucchini and Macdonald. This has been widely used in the application of HMMs. This definition becomes ambiguous when applied to multivariate data. The crux of the issue is how the data is inputted to the conditional cumulative density function. I have come up with two separate metrics which will be defined as vector and element pseudoresiduals

## Vector Pseudoresiduals

Vector pseudoresiduals are acquired by inputting the entire vector (hence the name) of observations for each time into the conditional cdf. Here  $\mathbf{X}_t \leq \mathbf{x}_t$  will indicate  $\{X_t^1, \dots, X_t^n\} \leq \{x_t^1, \dots, x_t^n\}$ . In other words, the probability that each separate variable  $X_t^j$  is less than the observed value for that variable at that time  $x_t^j$  for  $j \in \{1, \dots, n\}$ .

$$F(\mathbf{X}_t) = \mathbb{P}[\mathbf{X}_t \leq \mathbf{x}_t | \mathbf{X}^{(-t)} = \mathbf{x}^{(-t)}]$$

This will yield a single pseudoresidual value for each time, so there will be a total of  $T$  vector pseudoresiduals

## Element Pseudoresiduals

Element pseudoresiduals are acquired by inputting each element of the observation vector into a marginalized conditional cdf individually. Again if  $X_t^j \in \mathbf{X}_t$  indicates an observation of one of the variables  $j$  at time  $t$  then the marginalized conditional cdf for  $X_j$  will be defined as

$$F_j(X_t^j) = \mathbb{P}[X_t^j \leq x_t^j | \mathbf{X}^{(-t)} = \mathbf{x}^{(-t)}]$$

This process will yield an element pseudoresidual for each observation of each variable. So there will be  $nT$  element pseudoresiduals.

## Generating Data

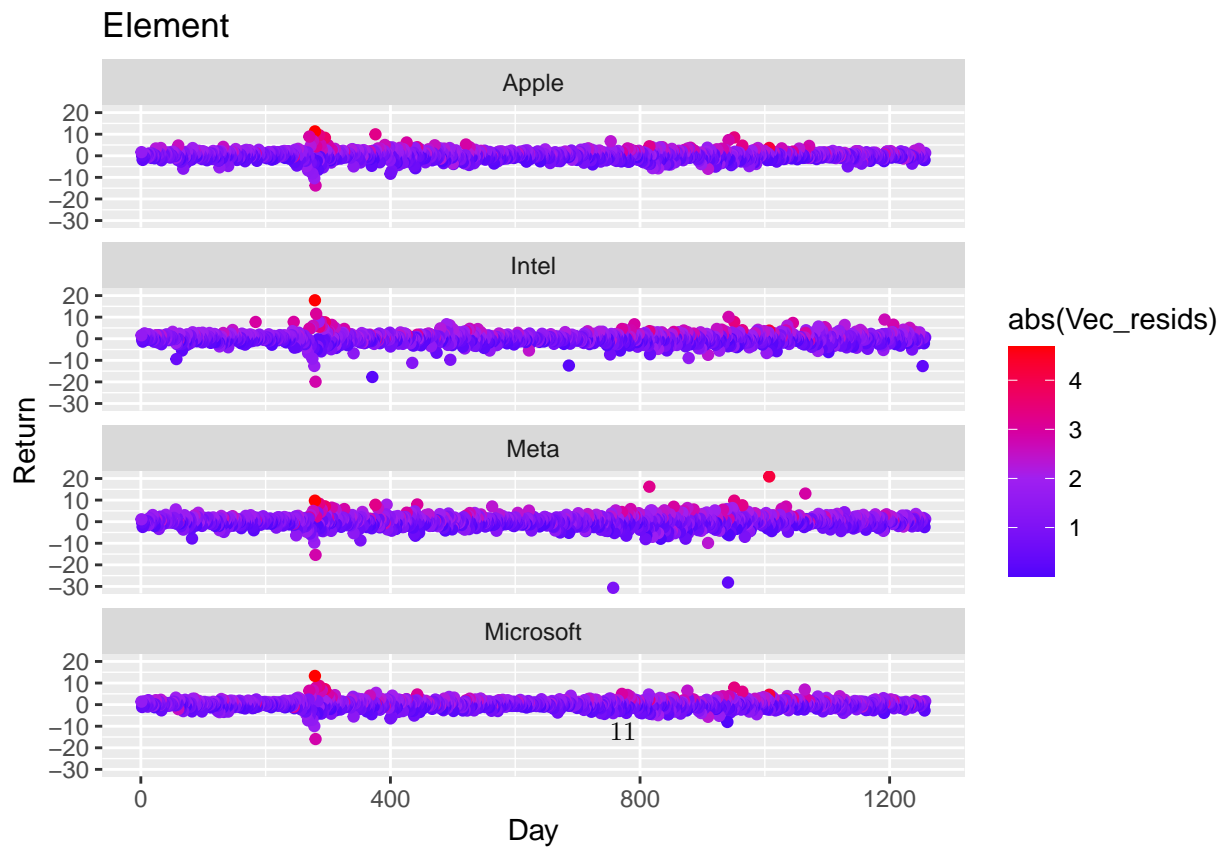
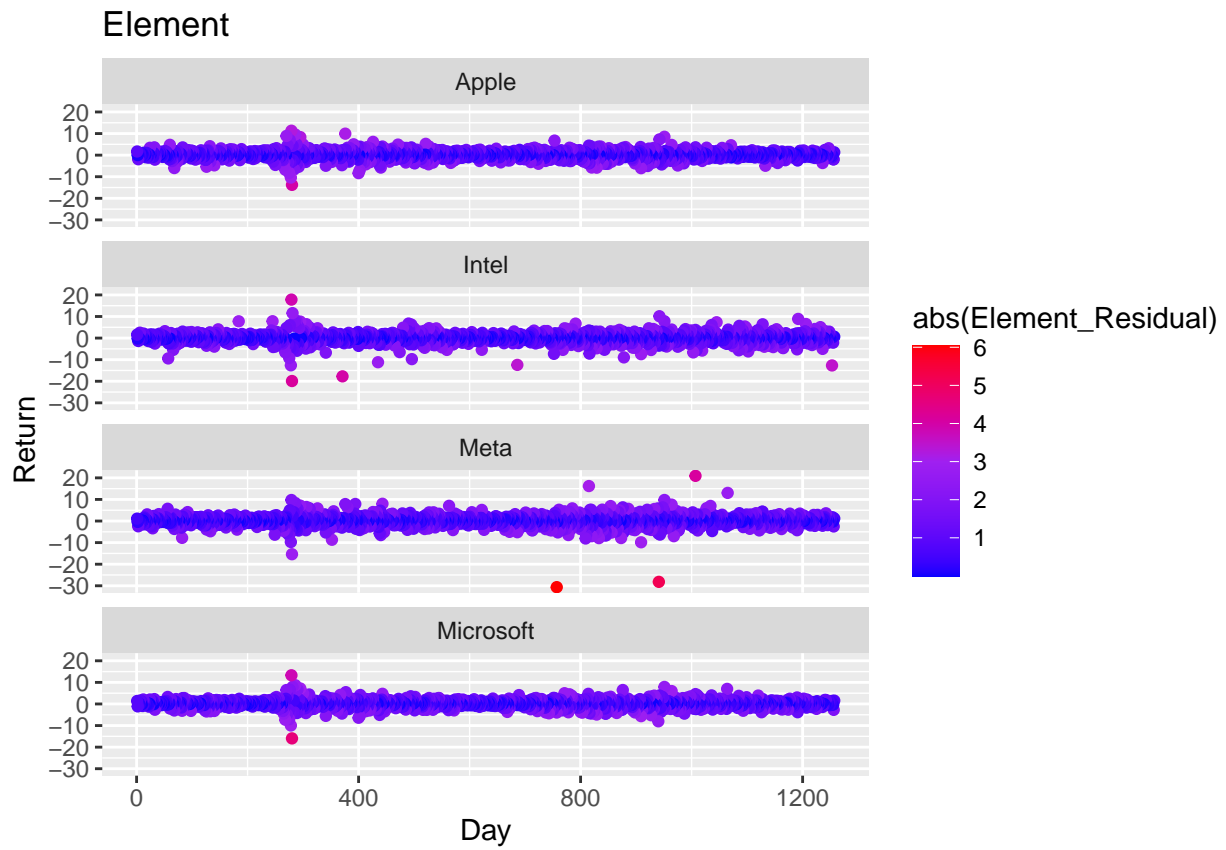
In order to further investigate vector and element pseudoresiduals, randomly generated data was used in order to eliminate the possibility that the returns data wasn't causing the issue. This is to confirm that element pseudoresiduals are the extension of the previously described normal pseudoresiduals for univariate data, and that vector pseudoresiduals

2-Dimensional Gaussians were used in order to reduce the number of parameters to be fitted for each trial. The means were set to be close to zero with variances much greater than the means, which was observed in the returns data. For each model, 200 observations were generated, then a second random model was used as an initial condition for fitting a new model to the generated data. For each fitted model, the `mvn.cdf_vector` and `mvn.cdf_element` functions were used. The inverse standard normal function `qnorm` was then used to get the two kinds of normal pseudoresiduals as described above. The mean and variance of the pseudoresiduals was then stored. Note, for the element normal pseudoresiduals, the values for each element were combined into a single vector and that was used for the variance calculation. (NOTE: If this is wrong to do, I can run the numbers again and separate the variances and means for the separate variables).

# Results

## Pseudoresiduals of Example Dataset

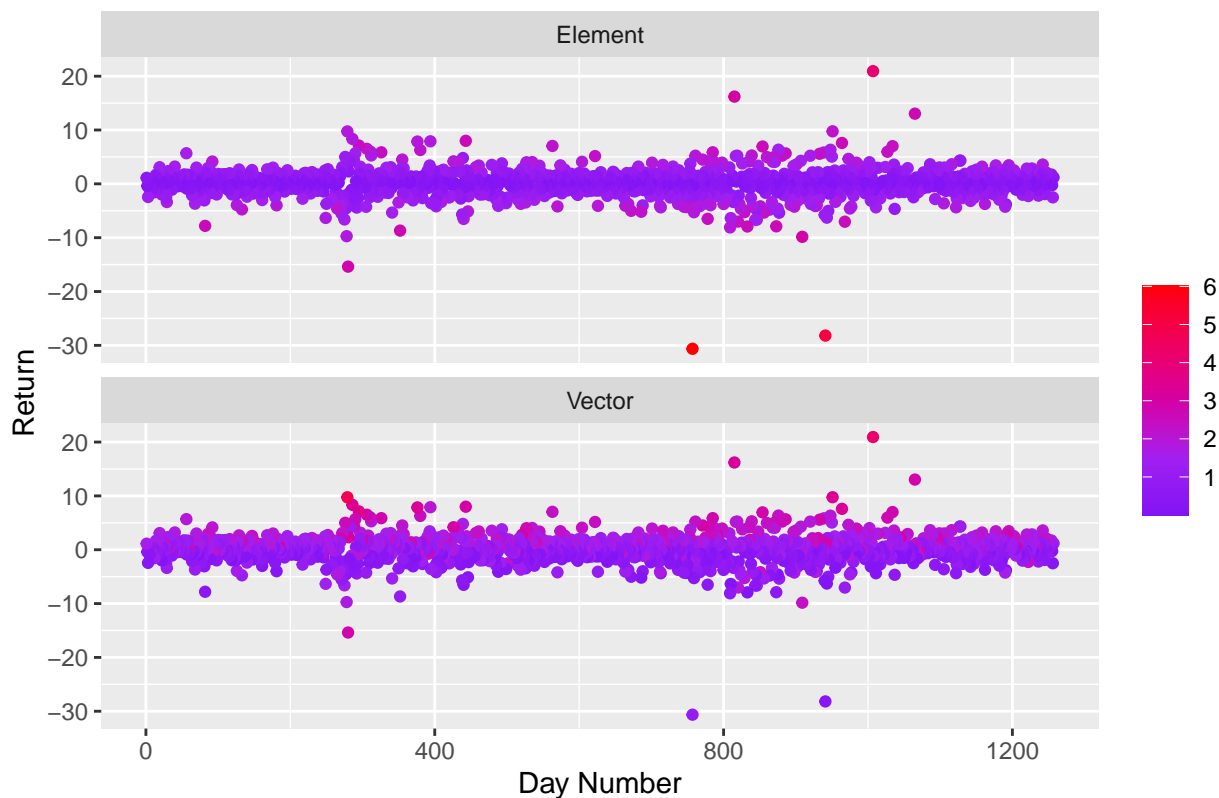
Element



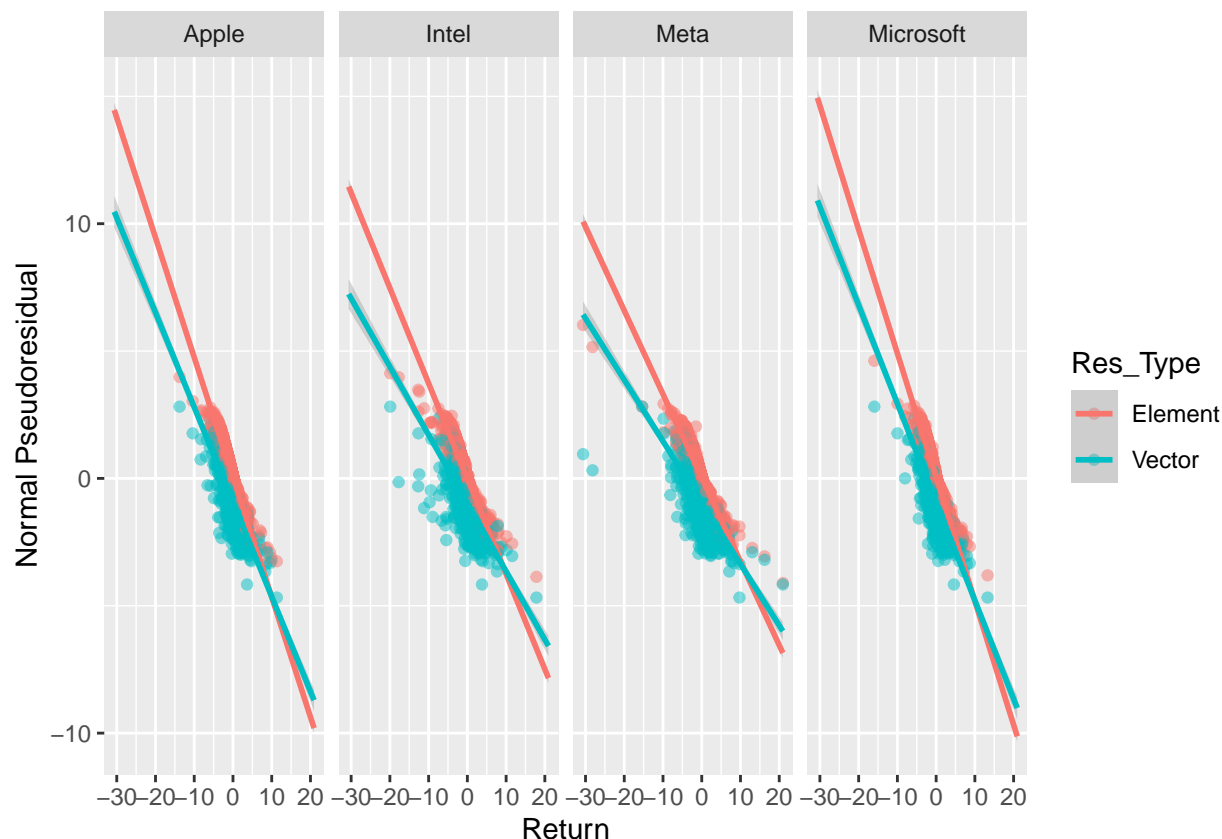
## Comparing Element and Vector Pseudoresiduals for Outlier Detection

One of the primary uses for normal pseudoresiduals is that of outlier detection. For the returns dataset, the element pseudoresiduals act as expected. The means for each state, and thus the means for every conditional distribution, are close to zero. Returns that are far from zero are at the tails of the conditional distributions so are flagged as outliers. The element pseudoresiduals consistently show this property for the model fitted to the returns dataset.

Meta Returns Over Time with Normal Pseudoresiduals



When directly comparing the element and vector pseudoresiduals for a single stock, one can see that clear outliers, most notably the lowest return value at around day 750. This is the lowest return that Meta has had in market history. It has the largest magnitude of the element pseudoresiduals, whereas the vector pseudoresiduals do not identify it in this way. This is just a single case. Let us further investigate the difference between vector and element pseudoresiduals by directly finding their relationship with returns. The behaviour from the returns over time figures appear to indicate that the farther away the return is from zero, the larger the magnitude of the pseudoresidual. It is equivalent to say that the pseudoresiduals and returns have an inverse relationship (i.e. as the return value increases, the value of the pseudoresidual decreases). Seen below, this is the case for both vector and element pseudoresiduals.



Clearly the negative relationship between pseudoresidual and return is stronger for element pseudoresiduals. In addition, the linear fit is better for the element pseudoresiduals. The r squared value for all stocks is higher for element then for vector pseudoresiduals.

Stock	R Squared
Apple	0.9307971
Intel	0.8836330
Meta	0.8642703
Microsoft	0.9119906

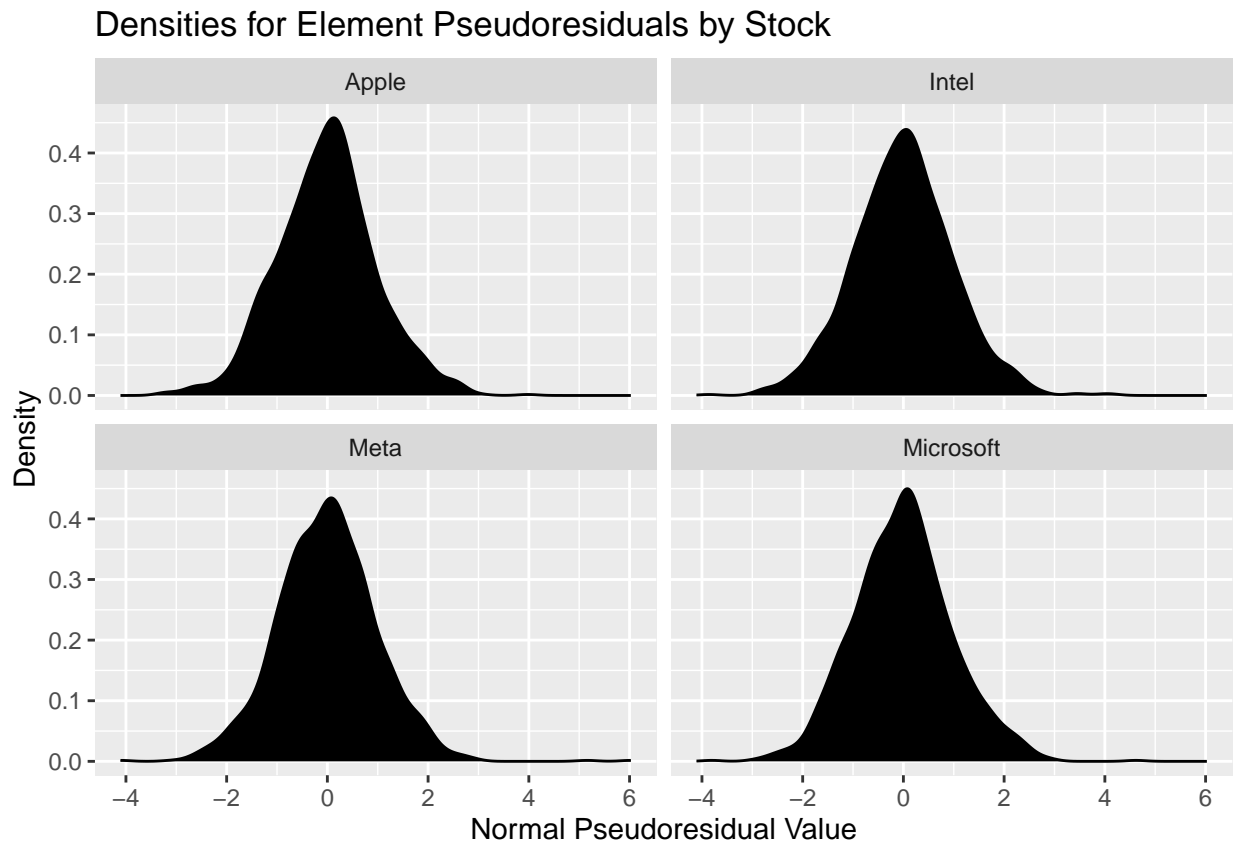
Stock	R Squared
Apple	0.6433921
Intel	0.5107222
Meta	0.5173959
Microsoft	0.6339389

The purpose of this analysis is two fold: firstly, the element pseudoresiduals are acting as expected for outlier detection. Secondly, the vector pseudoresiduals are not acting as well in this capacity. The is to reinforce the point that the element pseudoresiduals are a natural extension of the previously described pseudoresiduals, and vector pseudoresiduals are something else entirely.

The outlier detection for pseudoresiduals has been examined, let us now examine testing for goodness of fit.

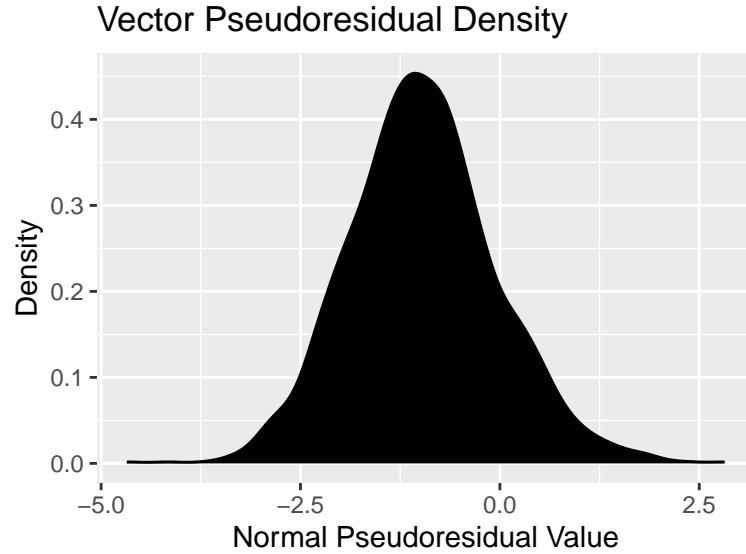
As previously said, for univariate data, normal pseudoresiduals are expected to have a standard normal distribution for a well fitted model.

First let us examine element pseudoresiduals.



Stock	Mean	Variance
Apple	-0.0094	0.9515
Intel	-0.0048	0.9688
Meta	-0.0010	0.9557
Microsoft	-0.0042	0.9457

Once again, element pseudoresiduals behave like the pseudoresiduals of their univariate counterparts. The means for all four stocks are close to 0, and the variances are close to 1. This is not to say that the model fitted to the returns data set is perfectly fitted. This is to demonstrate that element pseudoresiduals have the behaviour expected of normal pseudoresiduals. To contrast, let us now look at the density of the vector pseudoresiduals. SOME SORT OF HYPOTHESIS TEST HERE. GENERAL LIKELIHOOD TEST?



Mean	Variance
-0.9674	0.8590992

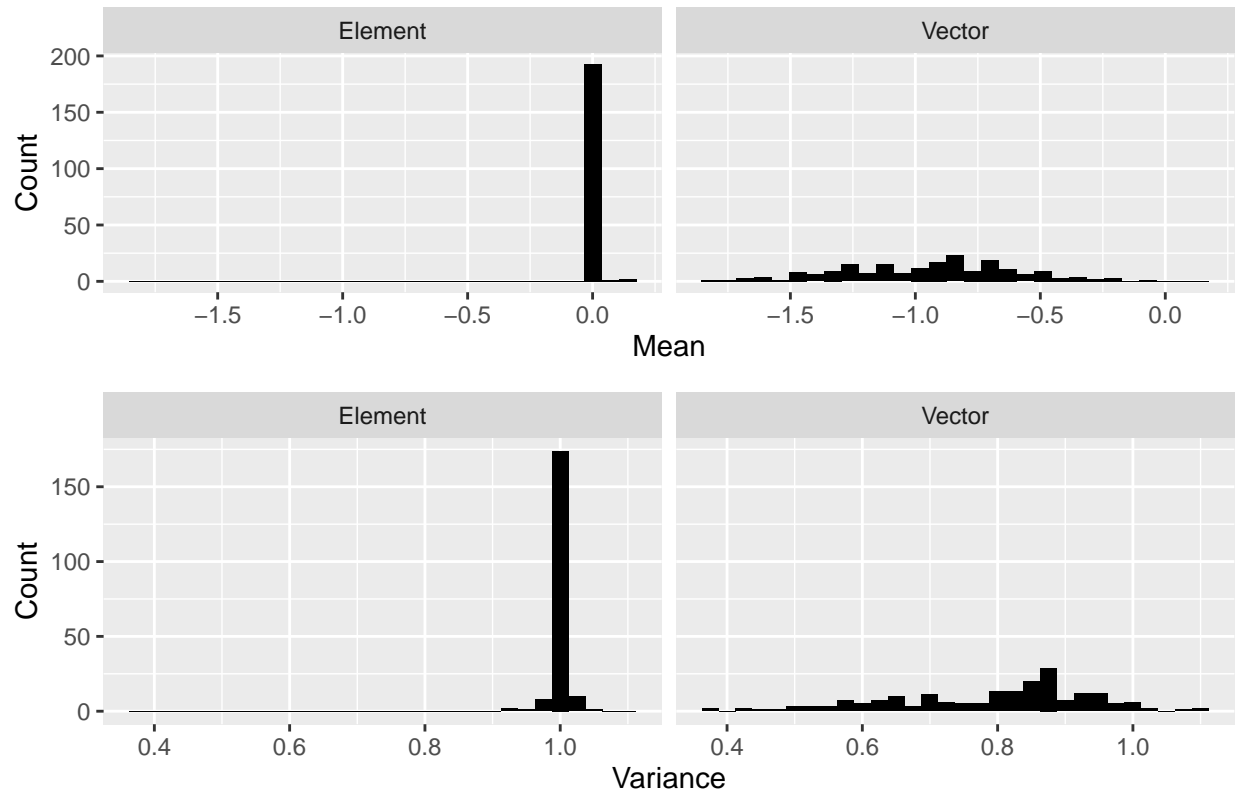
For the vector pseudoresiduals once again there is unexpected behaviour for normal pseudoresiduals. The vector pseudoresiduals are clearly not standard normal distributed. (MAYBE SOME CLASSICAL STATS BS STUFF HERE)

So far, element pseudoresiduals have behaved as expected of normal pseudoresiduals. They identify outliers, and they are close to standard normal. Vector pseudoresiduals have consistently diverged from these behaviours. The question remains, is this a consequence of the returns dataset that has been used thus far. To eliminate this possibility, we will now work with simulated data.

As mentioned above, for each trial for the results below, element and vector pseudoresiduals were calculated for a model fitted to randomly generated model from a different random model. The results are from the trials where the pseudoresidual function were able to run properly.

## Generated Data Results

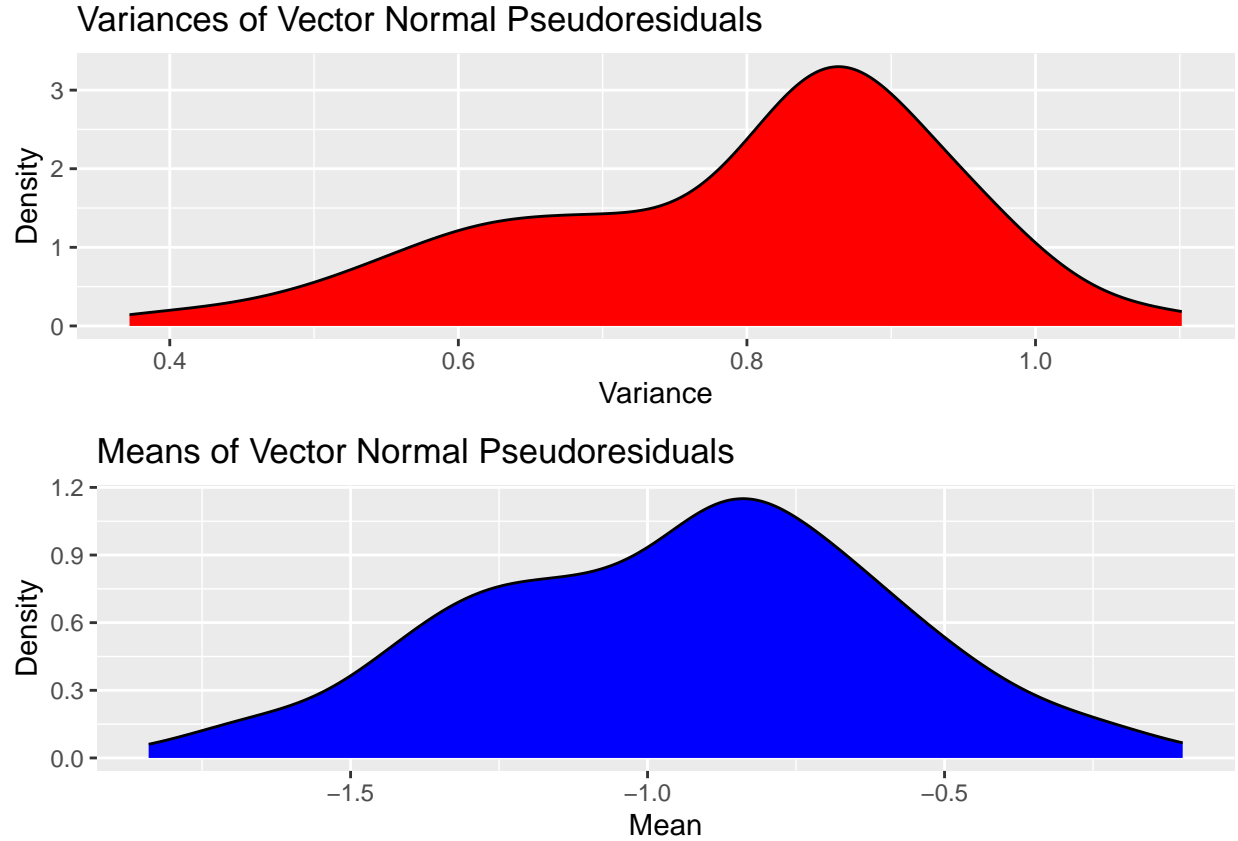
### Generated Data Pseudoresidual Statistics



Type	Avg Mean	Avg Variance	Med Mean	Med Variance
Element	0.0020	1.0010	0.0000	1.0025
Vector	-0.9449	0.7924	-0.8996	0.8286

The results dismiss any notions that it was the returns data set that was causing the behaviour of the vector pseudoresiduals. The element pseudoresiduals behave the same as seen with the example dataset. The vast majority of the means and variances are in line with those of the standard normal. Unfortunately a clear pattern did not arise from the vector pseudoresiduals. The density of the means and variances of the vector pseudoresiduals for each trial are shown below. The only thing that can definitely be said is that they are not in line with the standard normal.





## Conclusion

### Taking Stock of Vector Pseudoresiduals

There is still a lot of investigation needed to bring more clarity into the uses of vector pseudoresiduals. They do not conform to the distributions previously described for pseudoresiduals of univariate models. If their expected distribution for a well fitting model could be described, then vector pseudoresiduals could act as another model diagnostic.

In order to perhaps identify an expected distribution for the vector pseudoresiduals there are many avenues of investigation that are clear at this point. Firstly, investigating the effect of the state dependent distribution. There may be a relationship between dimensionality of the state dependent distribution and the evaluations of the conditional distributions. The means of all of the vector pseudoresiduals are all less than that of the standard normal. This seems to imply that when the observation vectors are inputted into the conditional distributions, they are consistently less likely. Simultaneously observing several values is always going to be less likely than observing a single return. Not only will it be less likely, the 'area' underneath the curve is also going to be smaller from the increase in dimensionality. Perhaps there is a way to scale the vector pseudoresidual values to get them to an expected standard normal distribution. In addition to dimensionality, there could be some relationship with the specific multivariate distribution impacting the distribution of the vector pseudoresiduals. Perhaps changing to different multivariate continuous distributions like a multivariate t-distribution will cause different behaviours to arise. However, for univariate distributions, there the specific state dependent distribution didn't have to be taken into account when checking the pseudoresiduals.

## Extensions

More generally, the techniques outlined here are not at all restricted for financial modelling. Multivariate hidden markov models have a wide range of applications. Finding a use for vector pseudoresiduals has the potential to have an impact for any field for which these models are applicable.

## Using Outlier Detection to Improve Models

Outlier detection was used here to point out differences between the vector and element pseudoresiduals. However, there could be some interesting uses for vector pseudoresiduals in identifying entire vectors of outliers.

## Appendix

### A: Results from Model Fitting

For completeness sake, here is the resulting model that resulted from the model fitting described above.

Table 6: Means

Apple	Microsoft	Meta	Intel
0.1458	0.1056	0.0548	0.1363
0.0567	0.1014	-0.4038	-0.2420
-1.0593	0.2534	1.0226	-0.4097

Table 7: Variances

Apple	Microsoft	Meta	Intel
1.5149	1.4020	1.6733	1.8198
3.4999	3.4949	4.7567	5.4408
2.3339	0.6887	1.2559	0.4435

Table 8: Correlation State 1

Apple	Microsoft	Meta	Intel
1.0000	0.7677	0.5921	0.6122
0.7677	1.0000	0.5783	0.6758
0.5921	0.5783	1.0000	0.4941
0.6122	0.6758	0.4941	1.0000

Table 9: Correlation State 2

Apple	Microsoft	Meta	Intel
1.0000	0.7640	0.5671	0.5879
0.7640	1.0000	0.5894	0.5916
0.5671	0.5894	1.0000	0.4320

Apple	Microsoft	Meta	Intel
0.5879	0.5916	0.4320	1.0000

Table 10: Correlation State 3

Apple	Microsoft	Meta	Intel
1.0000	0.6752	-0.3192	0.0301
0.6752	1.0000	-0.7581	0.0097
-0.3192	-0.7581	1.0000	-0.4641
0.0301	0.0097	-0.4641	1.0000

Table 11: Transition Probability Matrix

State 1	State 2	State 3
0.90234	0.09031	0.00735
0.44961	0.55039	0.00000
0.52896	0.00000	0.47104

Draw particular attention to correlation matrix of state 3, which has negative correlations.

## B: Formula for Conditional Distributions

### Forward Probabilities:

Forward probabilities are defined as follows:

$$\alpha_t = \delta P(\mathbf{x}_1) \Gamma P(\mathbf{x}_2) \Gamma P(\mathbf{x}_3) \dots \Gamma P(\mathbf{x}_t)$$

Essentially each  $\alpha_t$  a vector of the probabilities of observing  $\mathbf{x}^{(t)}$  given the state at time  $t$  is  $1, \dots, m$ .

$$\alpha_t(j) = \mathbb{P}(\mathbf{X}^{(t)} = \mathbf{x}^{(t)} | c_t = j)$$

### Backward Probabilities

Whereas forward probabilities are defined as the probability of observing all the observations up to and including time  $t$ , backward probabilities are the probabilities of observing all the observations after time  $t$ . The backward probability vector is defined as follows:

$$\beta_t = \Gamma P(\mathbf{x}_{t+1}) \Gamma P(\mathbf{x}_{t+2}) \dots \Gamma P(\mathbf{x}_T) \mathbb{1}$$

$$\beta_T = \mathbb{1}$$

These values will be used to define the conditional probabilities for each observation. The pdf for each observation is going to be the probability of an observation conditioned on the rest of the observations. To be explicit:

$$\mathbb{P}[X_t = \mathbf{x} | \mathbf{X}^{(-t)} = \mathbf{x}^{(-t)}] = \mathbb{P}\left[\frac{X_t = \mathbf{x}, \mathbf{X}^{(-t)} = \mathbf{x}^{(-t)}}{\mathbf{X}^{(-t)} = \mathbf{x}^{(-t)}}\right]$$

The numerator is just the likelihood as described above, just with the observation at time  $t$  replaced with the generic observation vector  $\mathbf{x}$ . The above probability can be calculated using the following values:

$$\mathbb{P}(X_t = \mathbf{x} | \mathbf{X}^{(-t)} = \mathbf{x}^{(-t)}) = \frac{\delta P(\mathbf{x}_1) \Gamma P(\mathbf{x}_2) \dots \Gamma P(\mathbf{x}_{t-1}) \Gamma P(\mathbf{x}) \dots \Gamma P(\mathbf{x}_T) \mathbf{1}'}{\delta P(\mathbf{x}_1) \Gamma P(\mathbf{x}_2) \Gamma P(\mathbf{x}_3) \dots \Gamma P(\mathbf{x}_{t-1}) \Gamma P(\mathbf{x}_{t+1}) \dots \Gamma P(\mathbf{x}_T) \mathbf{1}'}$$

This can be simplified using our notation from above for forward and backward probabilities to then get the pdf for an observation:

$$\mathbb{P}(\mathbf{X}_t = \mathbf{x} | \mathbf{X}^{(-t)} = \mathbf{x}^{(-t)}) = \frac{\alpha_{t-1} \Gamma P(\mathbf{x}) \beta'_t}{\alpha_{t-1} \Gamma \beta'_t}$$

From here, getting the conditional cumulative distribution function is trivial, one need only calculate the cdf for  $P(\mathbf{x})$  instead of the pdf (Essentially `pmvnorm` instead of `dmvnorm`). Here  $\mathbf{P}(\mathbf{x})$  indicates the use of the cdf instead of the pdf.

$$\mathbb{P}(\mathbf{X} \leq \mathbf{x} | \mathbf{X}^{(-t)} = \mathbf{x}^{(-t)}) = \frac{\alpha_{t-1} \Gamma \mathbf{P}(\mathbf{x}) \beta'_t}{\alpha_{t-1} \Gamma \beta'_t}$$

## Code