Ethan Weiland
12 December 2023

Bayesian Unicorns

Executive Summary

Privately held companies worth $1 billion are of special interest in business and finance because their success has been achieved without the need to raise capital on public markets. Additionally, if they do plan to go public, their initial public offering (IPO) will be worth hundreds of millions of dollars. This report explores the adequacy of the “unicorn” label by modeling the data generating process producing unicorn company valuation. If this underlying process can best be modeled by a single distribution, with one set of parameters, the label “unicorn” is adequate. If this underlying process is instead better modeled as a *mixture* of different distributions, each with different parameters, then it would be better to break unicorns down into different subspecies (so to speak). To answer this, I implement a Bayesian mixture model using the gibbsnorm() function in the bayess package in R. I then determine the optimal number of component distributions (one, two, three, or four) using model selection criteria and visualization techniques.

The analysis concludes that the three component mixture model is best. This model’s predicted density of logged valuation outperforms the predicted density of the two component and four component model. The predicted density of the one component model, which implies one set of parameters to describe all unicorns, performs poorly. The “unicorn” label is inadequate, and business and finance professionals should disaggregate unicorns into three groups to better understand valuations. The below report broadly describes these three groups, but future researchers should explore them in more detail. The chain for all the parameters

converges and this analysis is robust to prior specification. All code is available at: https://github.com/ethanphilipweiland/bayesian-unicorns

Data

The data for this project come from DataCamp, originally from https://www.cbinsights.com/research-unicorn-companies. The data contain information on unicorn companies as of November 2021. There were 917 companies that qualified as unicorns as of this date and there are thus 917 companies in the dataset. The following variables are included: company name, valuation (in $B), date the company became a unicorn, country of origin, sector, and a few select investors.

Figure 1: Distribution of Unicorn Valuation (Untransformed)

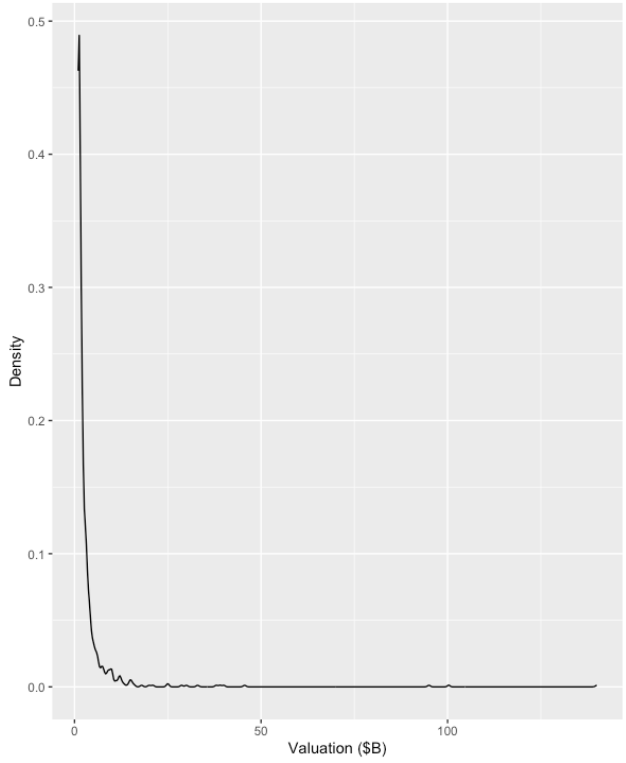
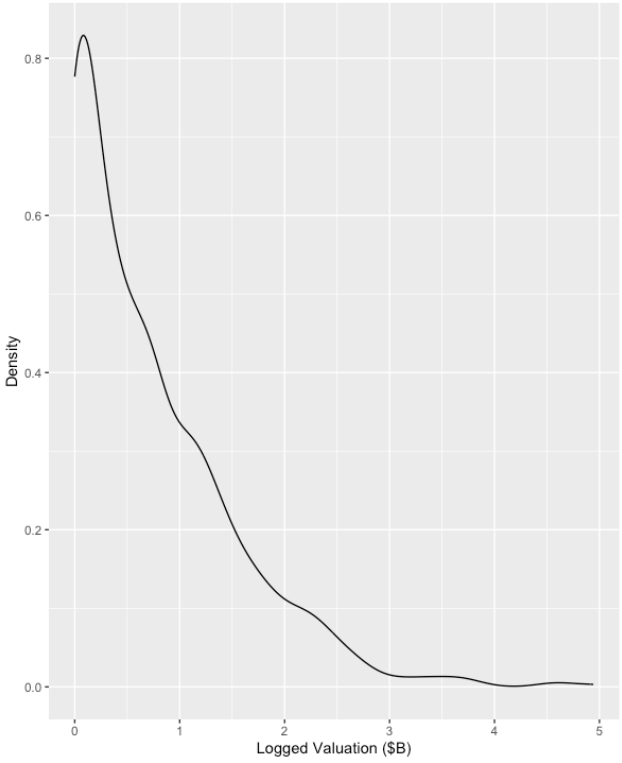


Figure 2: Distribution of Unicorn Valuation (Logged)



This analysis focuses on one variable: valuation. The median valuation is $1.6B, with an interquartile range of $1.93B. The mean valuation ($3.3B) is more than double the median, dragged upward by outliers. The most extreme are Bytedance ($140B), SpaceX ($100B), Stripe ($95B), and Klarna ($45.6B). The presence of these outliers results in a highly right skewed distribution (Figure 1). To adjust this, I log transform valuation (Figure 2). This transformation results in a more Normal distribution, although one that is still right skewed. It should be noted that this distribution is abruptly cut off when valuation is less than $1B, as this is the cutoff to be considered a unicorn.

Analysis

To determine if the data generating mechanism underlying the (logged) valuation of unicorns is a single Normal distribution or a mixture of Normal distributions, I implement a Bayesian mixture model for k = {1, 2, 3, 4}. The algorithm underlying this model is a Gibbs sampler, which is a type of Markov Chain Monte Carlo (MCMC). To illustrate how it works, I will use the case of two component distributions (k=2). First, the chain is initialized by choosing p, the probability of belonging to the first component, and $\boldsymbol{\theta}$, a vector containing the parameters of each component Normal distribution. Then, for each data point, the probability of belonging to each distribution is calculating using p (or 1-p) multiplied by the likelihood given the previous (initial) $\boldsymbol{\theta}$ values. Then each data point is sampled into component 1 or component 2 using the calculated probabilities. New $\boldsymbol{\theta}$ values are then sampled using the information from the data points in each component and the prior information. Next, a new p parameter is sampled taking into consideration the number of data points in each group and prior information. This process is repeated for the desired number of iterations, where at each step the chain only considers

information from the previous step. Posterior statistics can be approximated from the chain values; for example, the posterior expectation for the first component distribution can be calculated by averaging the samples across each iteration. A formal treatment of this process can be found in Marin and Robert (2014).

Of course, since this is a Bayesian analysis, the specification of priors is important. The prior on p is specified using a Dirichlet distribution with a hyperparameter of 0.5 (Marin and Robert 2014). The hyperparameter of 0.5 results in a uniform prior mass (see Figures A1 – A5 in the Appendix). The priors for the component distribution precisions follow a Gamma(10, σ^2) distribution, where σ^2 is the sample variance (see Figure A6). The priors for the component means follow a Normal(xbar, σ^2_i) distribution, where xbar is the sample mean and σ^2_i is the precision parameter (see Figure A7). Three things should be noted. First, the priors are not specified completely *a priori* but depend on the data. While some would object to this, Marin and Robert argue that this is "relatively harmless" and akin to standardizing a variable before an analysis (2014:187). Second, each component distribution is specified using the same prior information. This is important because this analysis tests whether the components are different and specifying that they are in the priors would bias the results. Third, the prior for the mean is conditional on the variance rather than having a fixed variance. Therefore, Figure A7 plots the prior conditional on ten randomly sampled precisions to give an idea of the various values the prior mean can take.

The Bayesian mixture model is implemented using the function gibbsnorm() in the package bayess in R (Marin and Robert 2014). Four different models are run: one with one component Normal distribution, a second with two component Normal distributions, a third with three component Normal distributions, and a fourth with four component Normal distributions.

The number of iterations for each model is 10,000. The gibbsnorm() function outputs a sequence of log likelihood values along Gibbs iterations. These are averaged and reported below in Table 1 as a (rough) measure of model fit. To best determine model fit, for each model, I randomly generate 1,000 predicted values and then plot this predicted density against the original data density (see Figure 3 – Figure 6).

Table 1: Model Fit by Number of Component Distributions

Components	logLik
1	-1,065.984
2	-839.283
3	-756.395
4	-720.347

Figure 3: One Component Bayesian Mixture Model

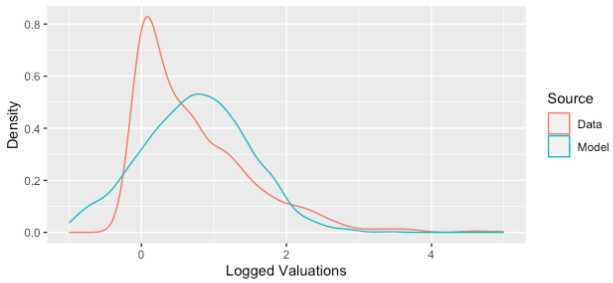


Figure 4: Two Component Bayesian Mixture Model

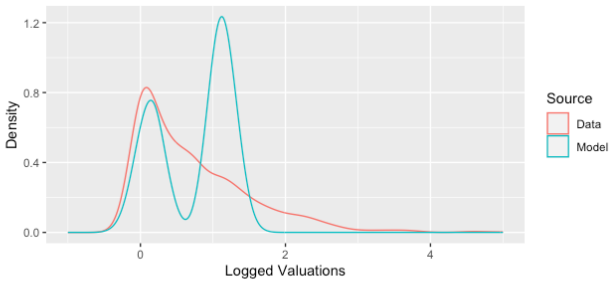


Figure 5: Three Component Bayesian Mixture Model

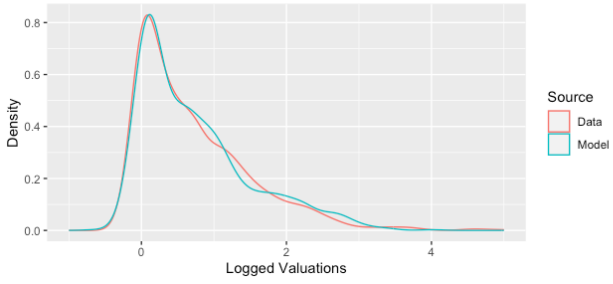
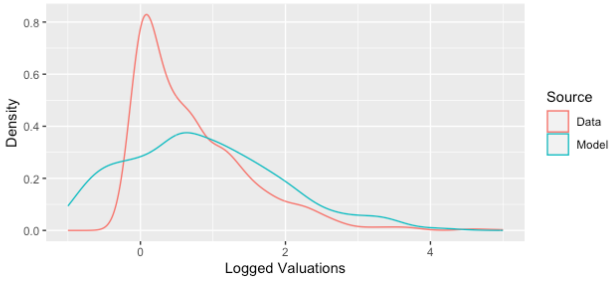


Figure 6: Four Component Bayesian Mixture Model



The four component Bayesian mixture model has the greatest (average) log likelihood, indicating that it is the best model. However, comparing the density of 1,000 predicted values

from the models against the original data density tells a different story. Both the four component model and the one component model are not able to capture the high density at log(0) valuations (representing companies that just qualify as a unicorn) and are too flat compared to the original data density. The two component model is able to capture the density at log(0) valuation, but greatly overestimates the density at log(1) valuation due to the inflexibility of having just two components. The three component model is clearly the best fit. It captures the high density at log(0) valuation as well as the density in the long right tail. Due to this close fit and the second-highest log likelihood value, I conclude that the three-component mixture model is the best model to describe unicorn company valuation.

Figure 5: Three Component Bayesian Mixture Model

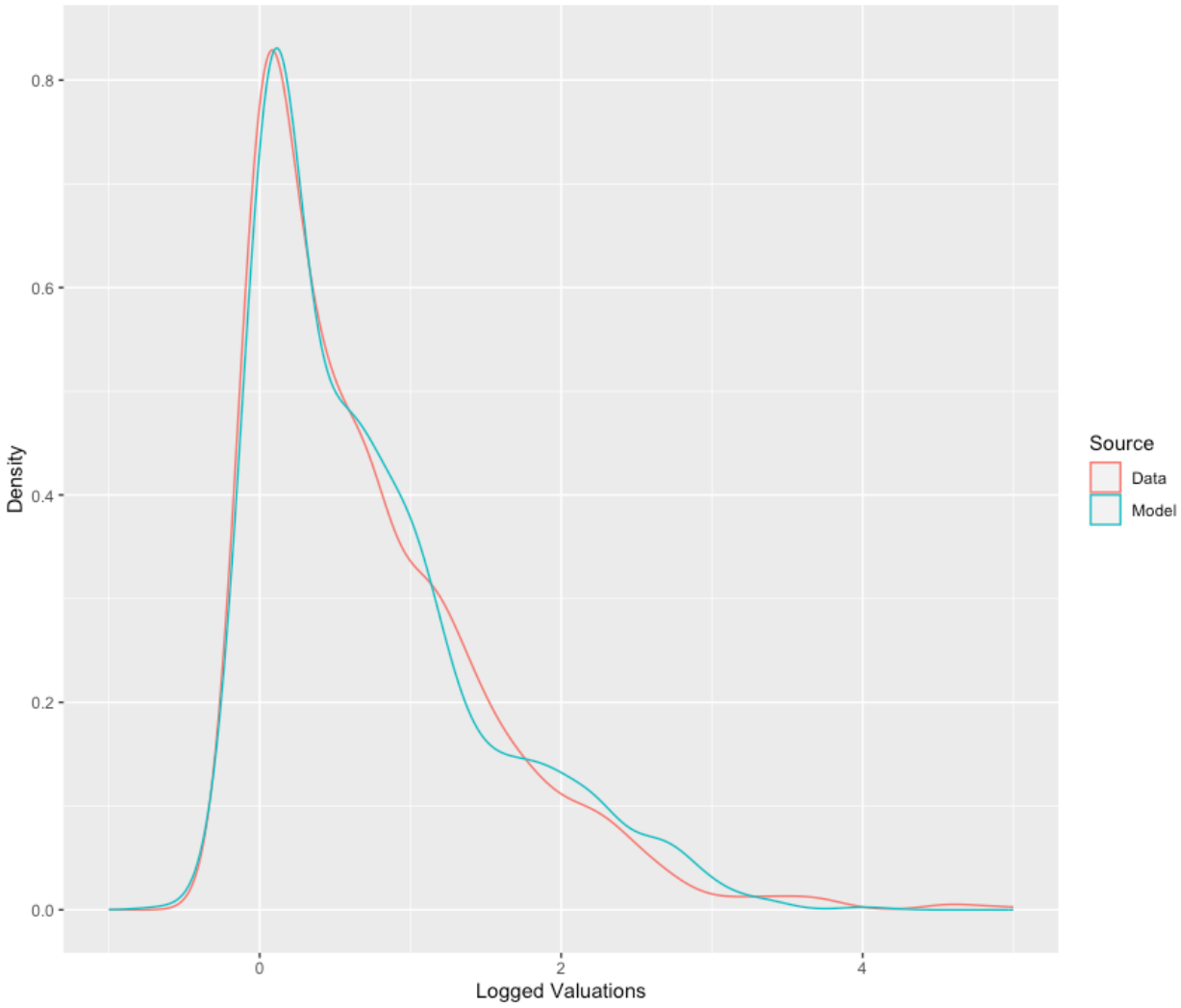


Table 2: Estimated Parameter Expectations of Three Components

	Mean	SD	Probability
Component 1	0.667	0.338	0.382
Component 2	1.645	0.777	0.271
Component 3	0.066	0.119	0.348

Table 3: Estimated Parameter Standard Deviations of Three Components

	Mean	SD	Probability
Component 1	0.047	0.023	0.040
Component 2	0.111	0.064	0.041
Component 3	0.019	0.011	0.022

Above, I reproduce Figure 5 so that it is easier to examine the model's predicted density against the actual data density. Table 2 and Table 3 detail the approximate expectation and approximate standard deviation of the posterior distributions of the parameters estimated in the mixture model (the mean, the standard deviation, and the probability of component membership) for each of the three components. These approximations are calculated from the output of the MCMC. For example, the approximate expectation of the mean parameter is the average of the means across all the iterations and the approximate standard deviation of the means is the standard deviation of these values. Examining these tables reveals three clusters of unicorns. The first cluster (Component 3) is centered around a log(0.066) valuation, corresponding to those companies that just qualify as a unicorn (just above the $1B threshold). The second cluster (Component 2) centers around a log(1.645) valuation and captures the most valuable unicorns, like ByteDance. Unsurprisingly, since this cluster contains the outliers, this cluster has the greatest posterior distribution expectation for the standard deviation parameter (0.777). The last component (Component 1) contains the remaining unicorn companies. The parameter posterior distributions are tight; the standard deviation for all these distributions is less than 0.15.

Table 4. Effective Sample Sizes Across Parameters and Components

	Mean	Variance	p
Component 1	330.600	444.238	343.746
Component 2	367.930	3,882.450	311.081
Component 3	851.514	1,088.437	1,547.289

To evaluate the Gibbs sampler, I first examine trace plots of the samples for each parameter (μ, σ^2, and p) across all three component distributions (see Appendix Figure A7 – Figure A15). There are no jumps in any of the nine trace plots, indicating that the chain has converged. However, cyclicality appears in many of these plots (see Figure A14 for a clear example). ACF plots (see Appendix Figure A16 – Figure A24) confirm that this is not just an appearance; there is high autocorrelation for all nine parameters. For example, the p parameter for the second component distribution has significant autocorrelation at the 40^{th} lag. Table 4 displays the effective sample sizes for the nine parameters. Even though the chain was run for 10,000 iterations, this is only equivalent to 330 independent samples for the first component mean, due to the high autocorrelation. Five of the parameters have effective sample sizes below 500 and six have effective sample sizes below 1,000. Future analysts should run the chain for many more iterations to increase the effective sample sizes.

Figure 7: Unicorn Valuation vs. Predicted Valuation Across Different Prior Configurations (Faceted)

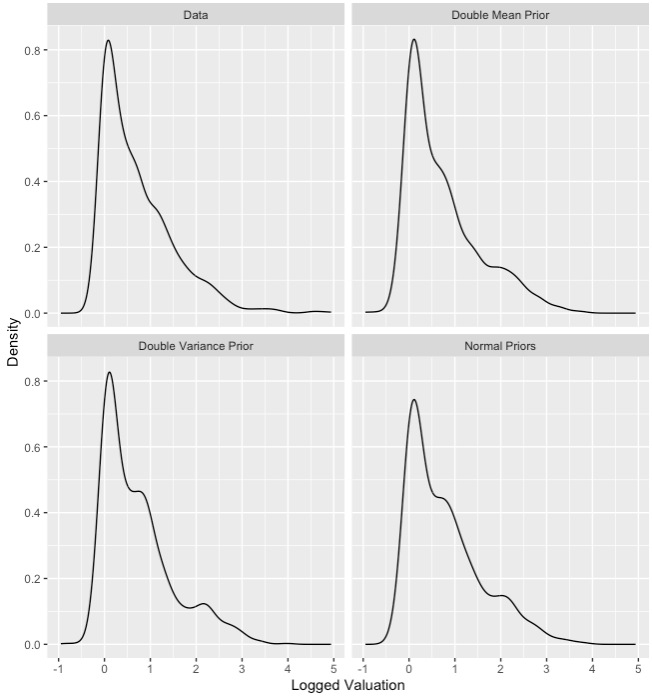


Figure 8. Unicorn Valuation vs. Predicted Valuation Across Different Prior Configurations

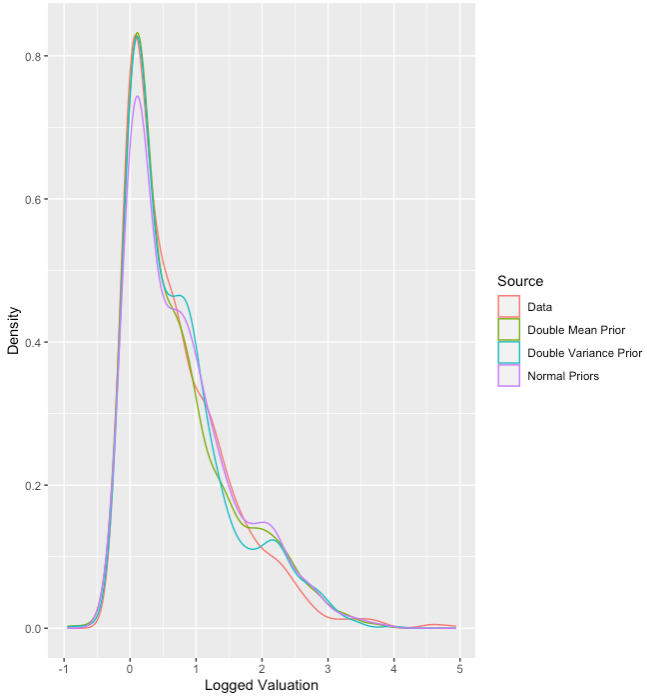


Table 5: Log Likelihoods Across Different Prior Specifications

Prior	logLik
Unadjusted Prior	-756.395
Double Mean Prior	-756.339
Double Variance Prior	-756.314

Finally, I test the sensitivity of this analysis to prior configuration. The original priors were a uniform mass on p, a Gamma(10, σ^2) distribution on the precision (where σ^2 is the sample variance) and a Normal(xbar, σ^2_i) distribution on the mean (where xbar is the sample mean and σ^2_i is the precision parameter). I rerun the three component model with two more prior configurations: one in which I double the sample mean and a second in which I double the sample variance (note that since the mean prior is conditional on the variance, this also affects

that mean). Figure 7 and Figure 8 display the density of the original logged unicorn valuation and the predicted density from the models with different priors. All three densities overlap substantially. In fact, the densities for the double mean prior and the double variance prior better capture the peak density of the original data. The average of log likelihood values along Gibbs iterations for these three prior specifications are nearly identical (Table 5). The model is robust to prior specification.

Implications & Future Research

Business and finance professionals label privately held companies valued at $1 billion or more “unicorns”. But how adequate is this label? In other words, are unicorn companies a homogenous group or not? To determine this, I implemented a Bayesian Normal mixture model on (logged) unicorn company valuations. I implemented this model for one component distribution, two component distributions, three component distributions, and four component distributions. Then, I used visualization techniques and model fit criteria to determine the optimal number of components.

I conclude that the three component model is best at describing logged unicorn company valuation. This model’s predicted density most closely aligns with the actual density of the logged valuations. The model is able to capture the high density at a valuation of log(0) (those companies that just qualify as unicorns) and also the right-skew. The one component model is a poor fit to the data, indicating that one set of parameters is inadequate at describing unicorns. Rather, there are three types of unicorns described by three different sets of parameters: those that just barely qualify as unicorns, the most successful unicorns (like ByteDance), and the rest. The valuations of these three types center around different values and have different spreads.

Professionals in business and finance should disaggregate unicorn companies into more precise groups to better understand their valuations.

Future researchers should address three primary concerns. First, future work should explore these three groups of unicorns in greater depth. Is the component membership a proxy for an already known variable like country or industry? Or, perhaps, the model is uncovering something more enticing like future success or failure. Second, future work should reproduce this analysis with different data. This analysis was conducted on unicorn companies from November 2021. The list of unicorn companies changes as unicorns go public or new privately held companies achieve a valuation of $1 billion. Researchers should use updated data to determine if the three component mixture model is still the most optimal model. Finally, future work should utilize greater processing power to increase the number of chain iterations. As discussed in the analysis, autocorrelation is high which reduces the effective sample size. To account for this, future researchers should run the chain for millions of iterations to ensure the most accurate results.

References

Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables. R package version 5.2.3. https://CRAN.R-project.org/package=stargazer

Marin, J. M. and Robert, C. (2013), Bayesian Essentials with R. New York: Springer. ISBN 978-1-4614-8686-2.

Appendix

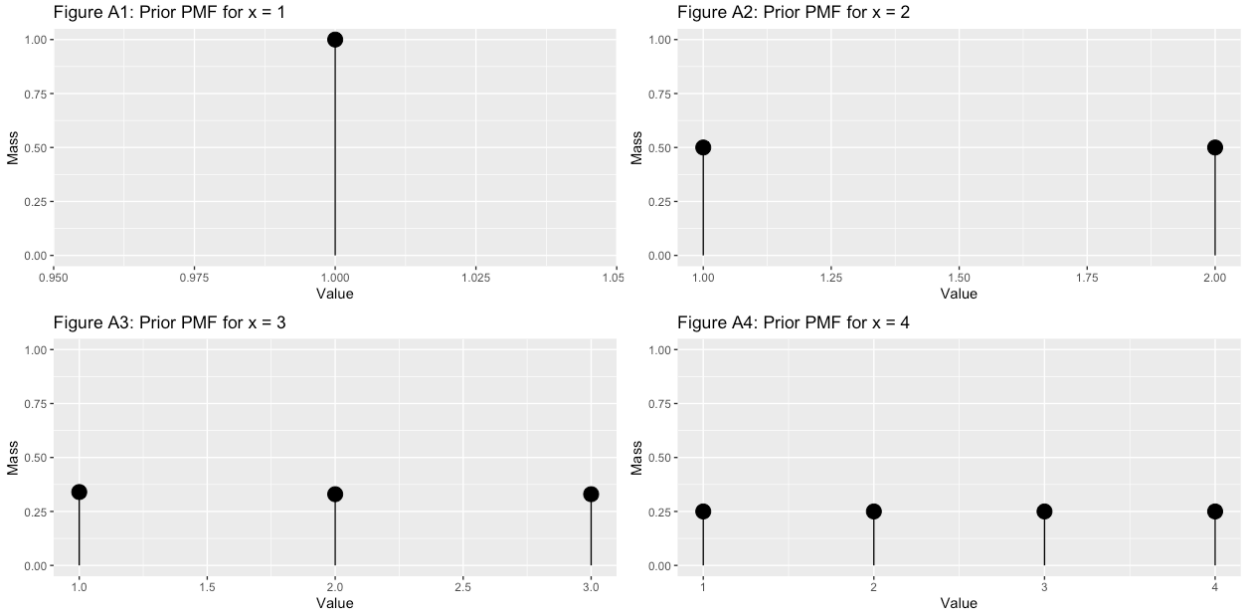


Figure A1: Prior PMF for x = 1
Mass
1.00
0.75
0.50
0.25
0.00
0.950
0.975
1.000
1.025
1.05
Value
Figure A2: Prior PMF for x = 2
Mass
1.00
0.75
0.50
0.25
0.00
1.00
1.25
1.50
1.75
2.00
Value
Figure A3: Prior PMF for x = 3
Mass
1.00
0.75
0.50
0.25
0.00
1.0
1.5
2.0
2.5
3.0
Value
Figure A4: Prior PMF for x = 4
Mass
1.00
0.75
0.50
0.25
0.00
1
2
3
4
Value

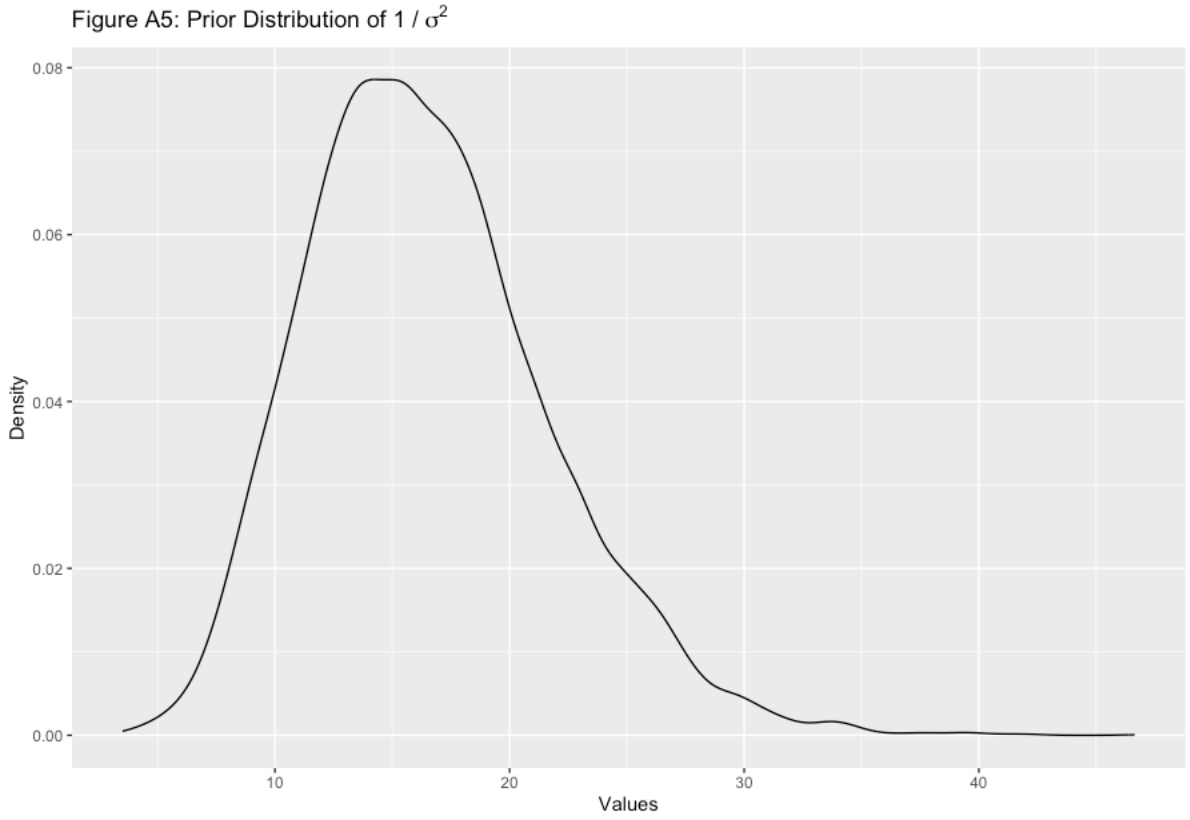


Figure A5: Prior Distribution of 1 / σ2
Density
0.08
0.06
0.04
0.02
0.00
10
20
30
40
Values

Figure A6: Prior Plot for Mean Conditional on Various Precisions

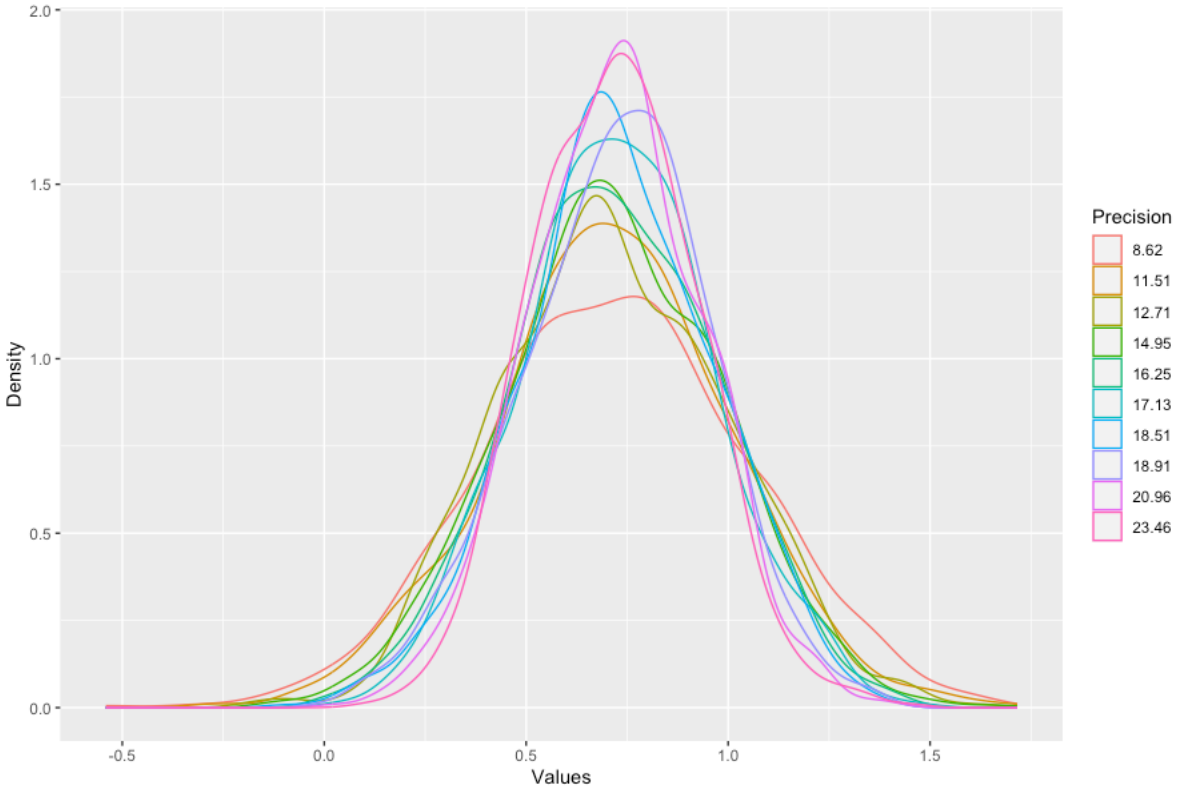


Figure A7: Trace Plot of First Component μ

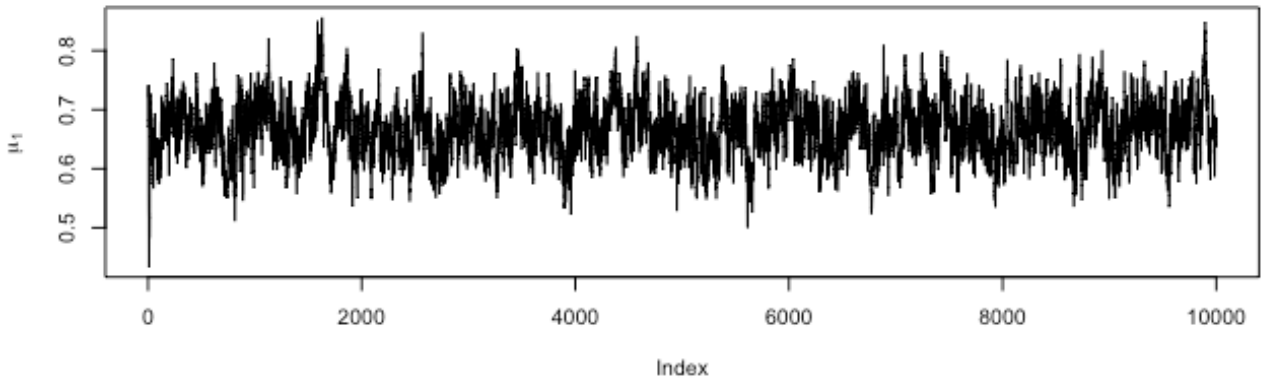


Figure A8: Trace Plot of Second Component μ

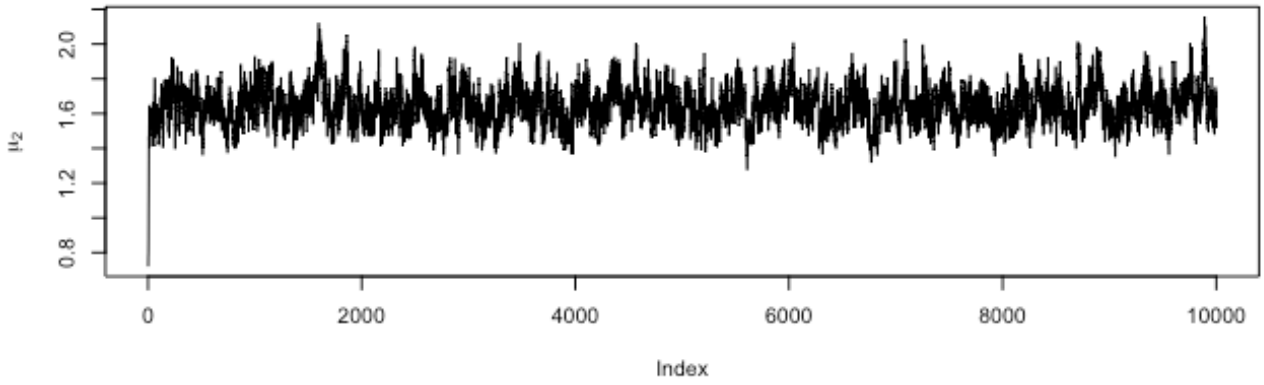


Figure A9: Trace Plot of Third Component μ

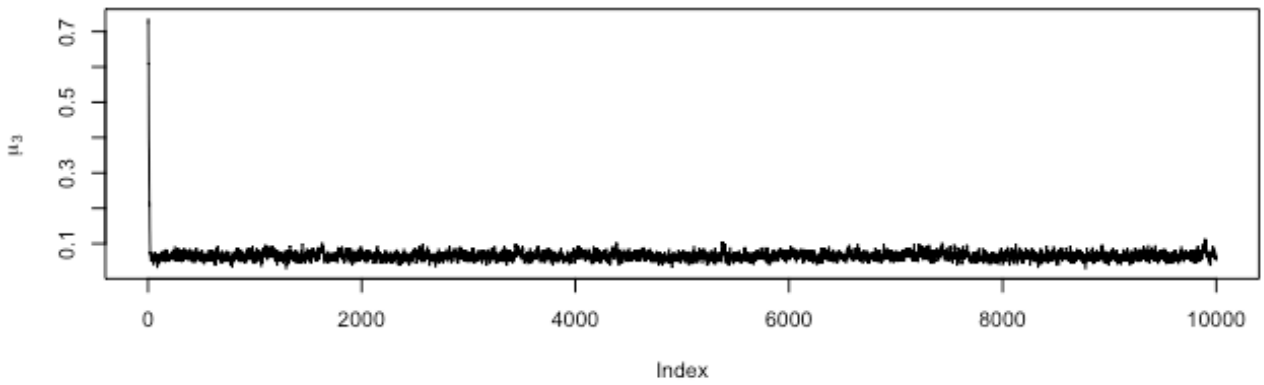


Figure A10: Trace Plot of First Component σ^2

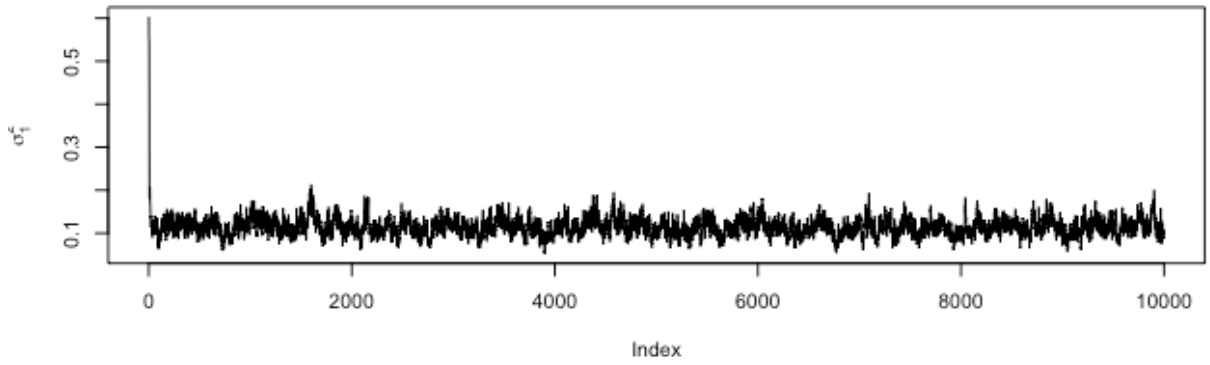


Figure A11: Trace Plot of Second Component σ^2

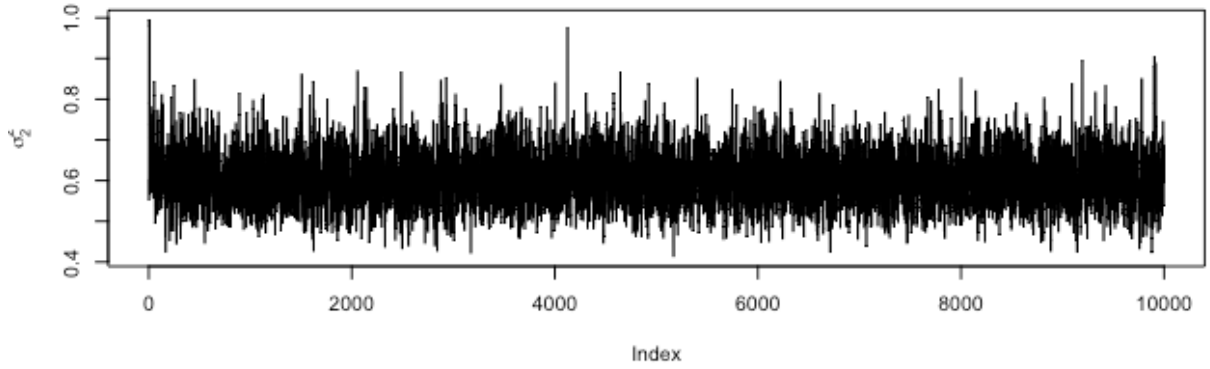


Figure A12: Trace Plot of Third Component σ^{2_3}

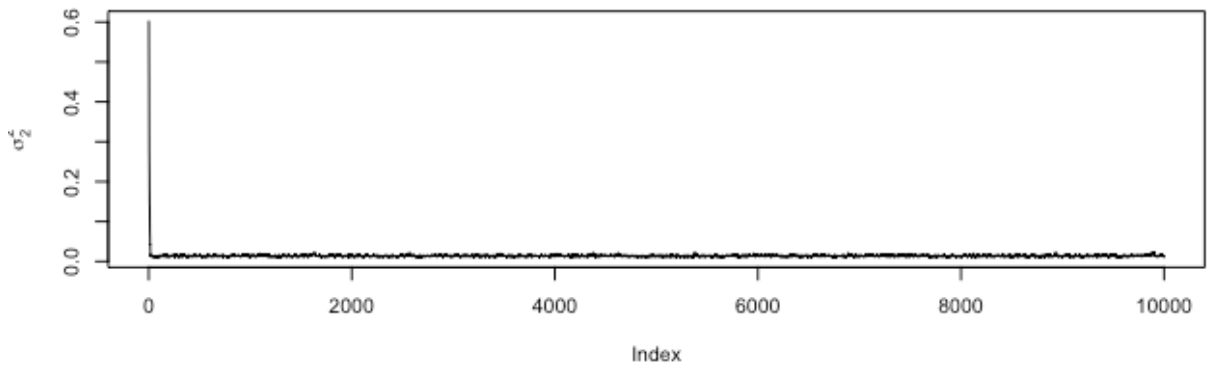


Figure A13: Trace Plot of First Component p

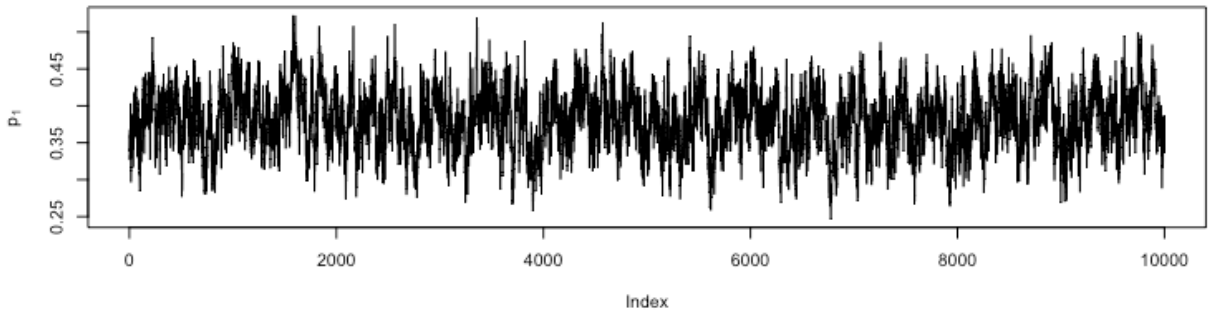


Figure A14: Trace Plot of Second Component p

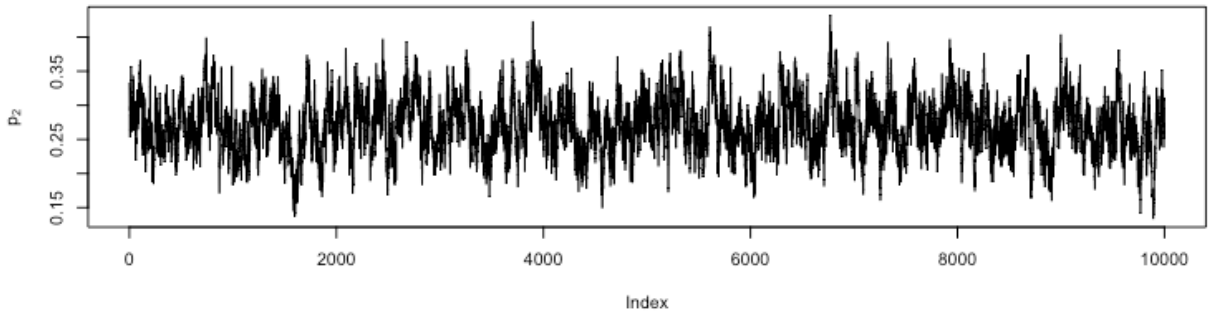


Figure A15: Trace Plot of Third Component p

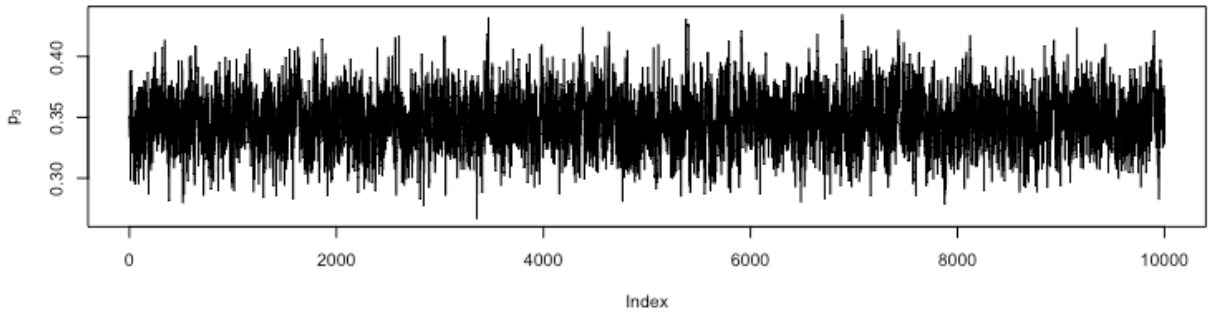


Figure A16: ACF Plot of First Component μ

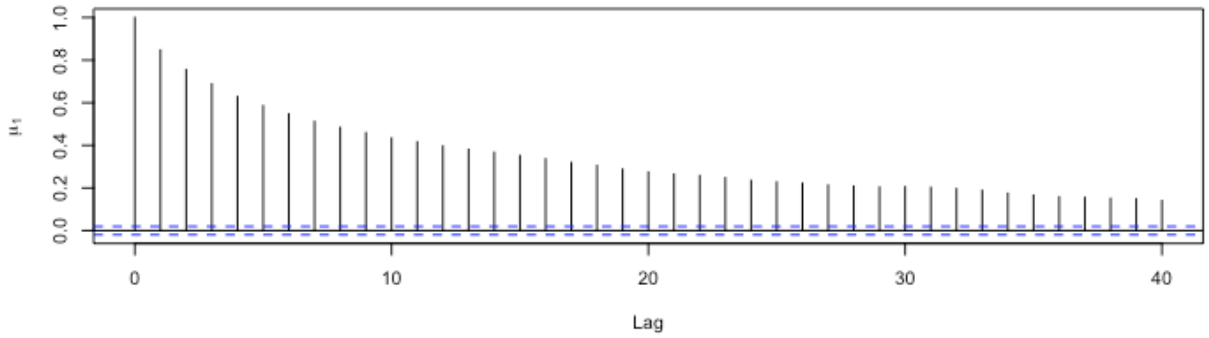


Figure A17: ACF Plot of Second Component μ

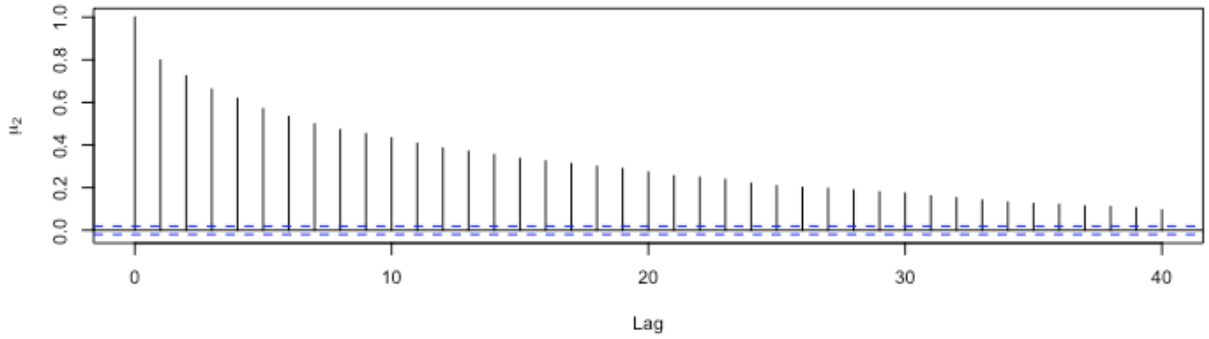


Figure A18: ACF Plot of Third Component μ

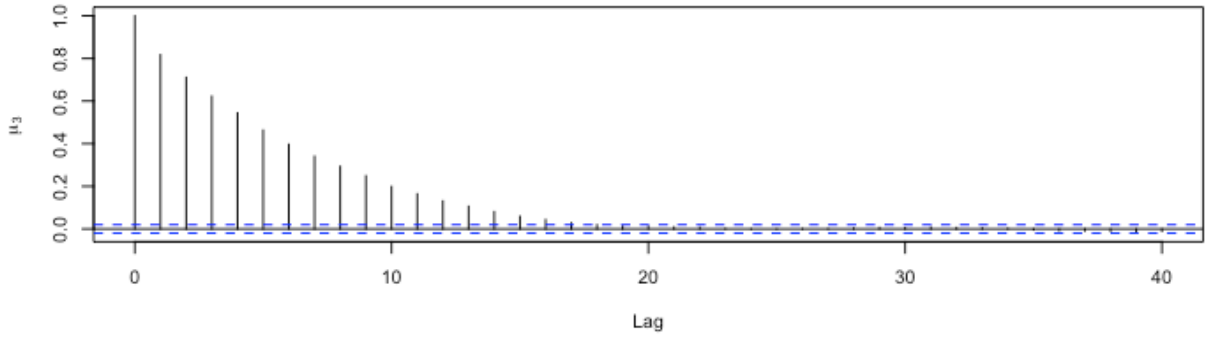


Figure A19: ACF Plot of First Component σ^2

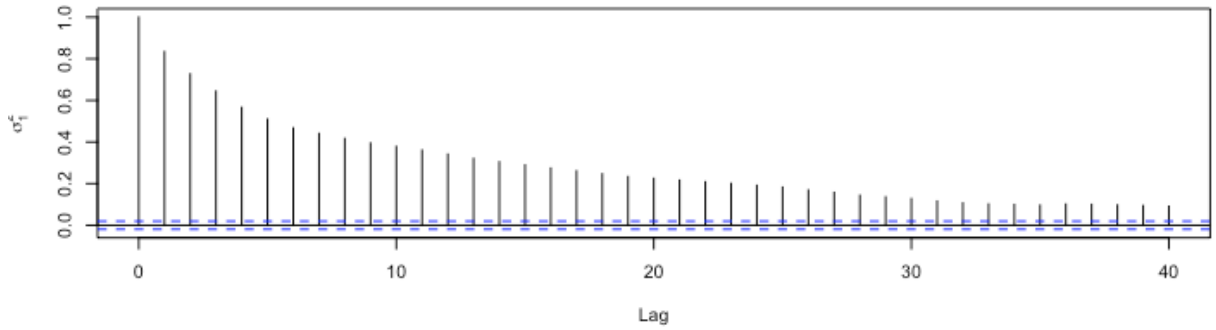


Figure A20: ACF Plot of Second Component σ^2

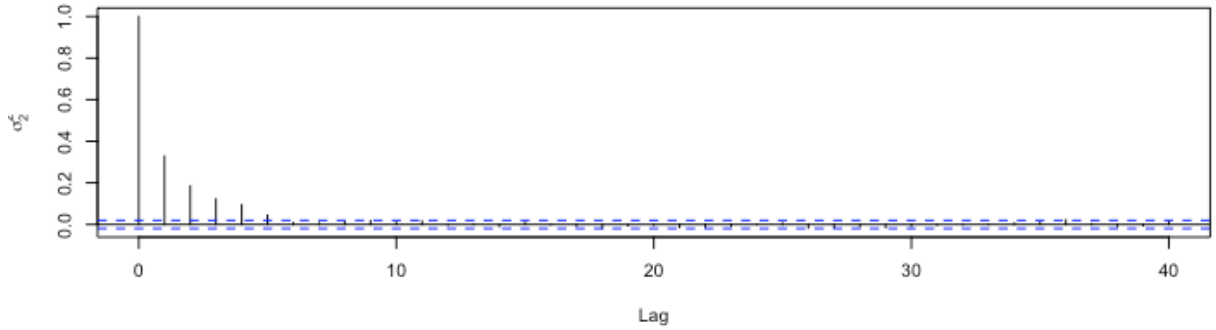


Figure A21: ACF Plot of Third Component σ^2

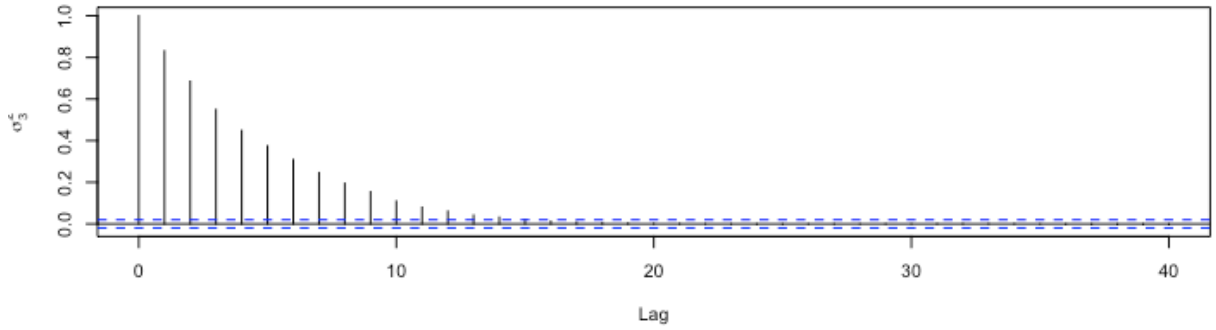


Figure A22: ACF Plot of First Component p

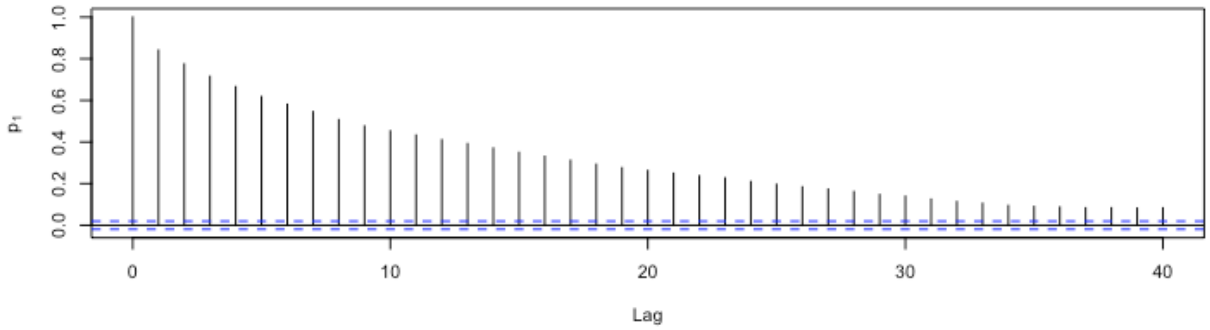


Figure A23: ACF Plot of Second Component p

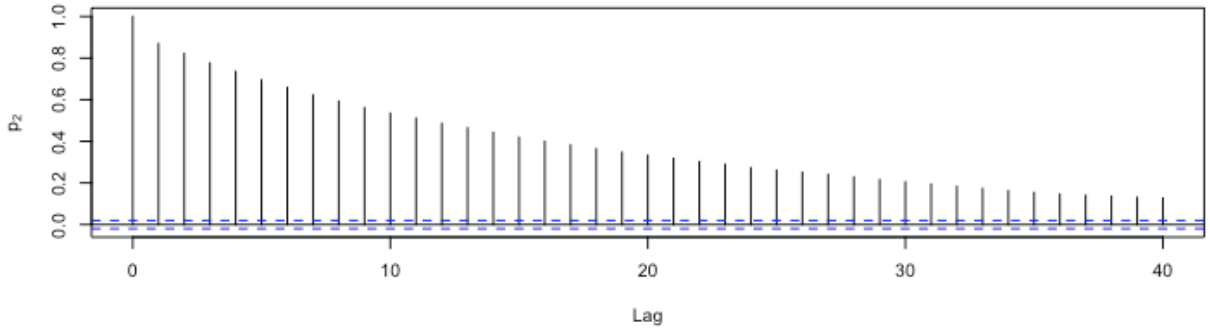


Figure A24: ACF Plot of Third Component p