Empirical Priors for Measurement Model Parameters

Lecture 4g

https://jonathantemplin.com/bayesian-psychometric-modeling-fall-2022/

Today's Lecture Objectives

1. Show differing choices of prior distributions for varying parameters

Example Data: Conspiracy Theories

Today's example is from a bootstrap resample of 177 undergraduate students at a large state university in the Midwest. The survey was a measure of 10 questions about their beliefs in various conspiracy theories that were being passed around the internet in the early 2010s. Additionally, gender was included in the survey. All items responses were on a 5-point Likert scale with:

1. Strongly Disagree
2. Disagree
3. Neither Agree or Disagree
4. Agree
5. Strongly Agree

Please note, the purpose of this survey was to study individual beliefs regarding conspiracies. The questions can provoke some strong emotions given the world we live in currently. All questions were approved by university IRB prior to their use.

Our purpose in using this instrument is to provide a context that we all may find relevant as many of these conspiracy theories are still prevalent today.

Conspiracy Theory Questions 1-5

Questions:

1. The U.S. invasion of Iraq was not part of a campaign to fight terrorism, but was driven by oil companies and Jews in the U.S. and Israel.
2. Certain U.S. government officials planned the attacks of September 11, 2001 because they wanted the United States to go to war in the Middle East.
3. President Barack Obama was not really born in the United States and does not have an authentic Hawaiian birth certificate.
4. The current financial crisis was secretly orchestrated by a small group of Wall Street bankers to extend the power of the Federal Reserve and further their control of the world's economy.
5. Vapor trails left by aircraft are actually chemical agents deliberately sprayed in a clandestine program directed by government officials.

Conspiracy Theory Questions 6-10

Questions:

6. Billionaire George Soros is behind a hidden plot to destabilize the American government, take control of the media, and put the world under his control.
7. The U.S. government is mandating the switch to compact fluorescent light bulbs because such lights make people more obedient and easier to control.
8. Government officials are covertly Building a 12-lane "NAFTA superhighway" that runs from Mexico to Canada through America’s heartland.
9. Government officials purposely developed and spread drugs like crack-cocaine and diseases like AIDS in order to destroy the African American community.
10. God sent Hurricane Katrina to punish America for its sins.

Model Setup Today

Today, we will revert back to the CFA model assumptions to discuss the impact of different priors

- I chose CFA as it is very clear from non-Bayesian analyses what minimal identification constraints are needed
- Additionally, we will use a single latent variable/factor for this lecture

For an item i the model is:

$$Y_{pi} = \mu_i + \lambda_i \theta_p + e_{p,i}; \quad e_{p,i} \sim N\left(0, \psi_i^2\right)$$

Empirical Priors for Item Parameters

Empirical Priors

In many Bayesian references, the use of so-called “empirical priors” for various model parameters is suggested

- An empirical prior is one where the hyper parameters of the prior distribution are estimated and are not fixed
- For example:
  - $\lambda_i \sim N(\mu_\lambda, \sigma_\lambda)$; previously we specified $\lambda_i \sim N(0, \sqrt{1000})$
  - $\mu_i \sim N(\mu_\mu, \sigma_\mu)$; previously we specified $\mu_i \sim N(0, \sqrt{1000})$
  - $\psi_i \sim \text{exponential}(\text{rate}_\psi)$; previously we specified $\psi_i \sim \text{exponential}(.1)$
- Note: we aren’t including θ just yet…
  - Scale identification discussion is needed

Empirical Priors in Psychometric Models

- For psychometric models, the choice of empirical priors can have several pitfalls:
  - Not all model parameters can use empirical priors (see the last example and the next lecture on identification)
  - The use of such priors can make some parameter estimates move toward values that would indicate more information for θ than what is present in the data
  - Empirical priors may be inappropriate when some observed variables have widely different scales (or use different distributions)
- Overall, I do not recommend the use of empirical priors in psychometric analyses
  - I show them to sync class with other Bayesian texts

Empirical Priors in Stan: Model Block

```
model {

  meanLambda ~ normal(meanLambdaMean, meanLambdaSD);
  sdLambda ~ exponential(sdLambdaRate);
  lambda ~ normal(meanLambda, sdLambda);

  meanMu ~ normal(meanMuMean, meanMuSD);
  sdMu ~ exponential(sdMuRate);
  mu ~ normal(meanMu, sdMu);

  psiRate ~ exponential(ratePsiRate);
  psi ~ exponential(psiRate);

  theta ~ normal(0, 1);

  for (item in 1:nItems){
    Y[,item] ~ normal(mu[item] + lambda[item]*theta, psi[item]);
  }

}
```

Notes:

- $\lambda_i \sim N(\text{meanLambda}, \text{sdLambda})$
  - `meanLambda` is the estimated hyperparameter for the mean of the factor loadings with prior distribution N (meanLambdaMean, meanLambdaSD)
  - `` `sdLambda `` is the estimated hyper parameter for the standard deviation of the factor loadings with prior distribution exponential(sdLambdaRate)

Additional Model Block Notes

- $\mu_i \sim N(\text{meanMu}, \text{sdMu})$
  - `meanMu` is the estimated hyperparameter for the mean of the factor loadings with prior distribution N (meanMuMean, meanMuSD)
  - `` `sdMu `` is the estimated hyper parameter for the standard deviation of the factor loadings with prior distribution exponential(sdMuRate)
- $\psi_i \sim \text{exponential}(\text{psiRate})$
  - `psiRate` is the estimated rate parameter for the unique standard deviations with prior distribution exponential(ratePsiRate)

Stan Parameters Block

```
parameters {
  vector[nObs] theta;
  vector[nItems] mu;
  vector[nItems] lambda;
  vector<lower=0>[nItems] psi;
  real meanLambda;
  real<lower=0> sdLambda;

  real meanMu;
  real<lower=0> sdMu;
  real<lower=0> psiRate;
}
```

Notes:

- The rate parameters are constrained to be positive (as needed for the PDF of the exponential distribution)

Stan Data Block

```
data {
  int<lower=0> nObs;                // number of observations
  int<lower=0> nItems;              // number of items
  matrix[nObs, nItems] Y;           // item responses in a matrix

  real meanLambdaMean;
  real<lower=0> meanLambdaSD;
  real<lower=0> sdLambdaRate;

  real meanMuMean;
  real<lower=0> meanMuSD;
  real<lower=0> sdMuRate;

  real<lower=0> ratePsiRate;
}
```

Notes:

- We can import values for the hyperparameters for the prior distributions of each

R Data List

```r
modelCFA2_data = list(
  nObs = nObs,
  nItems = nItems,
  Y = conspiracyItems,
  meanLambdaMean = 0,
  meanLambdaSD = 1,
  sdLambdaRate = .1,
  meanMuMean = 0,
  meanMuSD = 1,
  sdMuRate = .1,
  ratePsiRate = .1
)
```

Notes:

- We are setting the hyperparameters for the loading mean and intercept mean to $N(0, 1)$
- The hyperparameters for each rate are set to .1

Connection of Empirical Priors to Multilevel Models

The empirical priors on the parameters are similar to specifying multilevel models for each

- $\lambda_i \sim N(\mu_\lambda, \sigma_\lambda)$ can be reparameterized as $\lambda_i^* = \mu_\lambda + e_{\lambda_i}$ with $e_{\lambda_i} \sim N(0, \sigma_\lambda)$
- $\mu_i \sim N(\mu_\mu, \sigma_\mu)$ can be reparameterized as $\mu_i^* = \mu_\mu + e_{\mu_i}$ with $e_{\mu_i} \sim N(0, \sigma_\mu)$
- The rate is a bit trickier, but also can be reparameterized similarly

  Moreover, these reparameterizations can lead to predicting what each parameter should be based on item-specific predictors

- The basis of explanatory item response models

Model Results

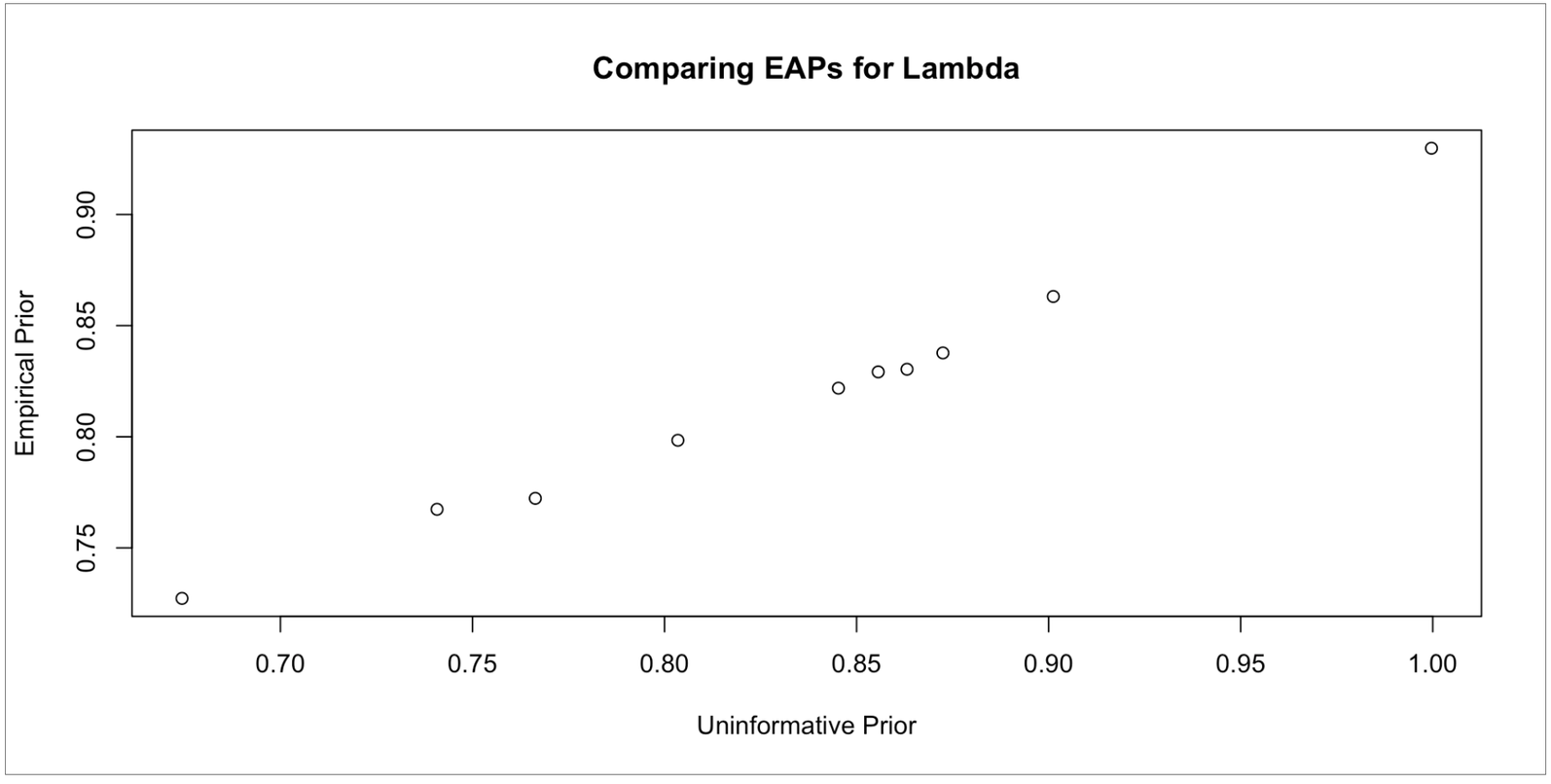
```
# checking convergence
max(modelCFA2_samples$summary()$rhat, na.rm = TRUE)
```

```
[1] 1.017071
```

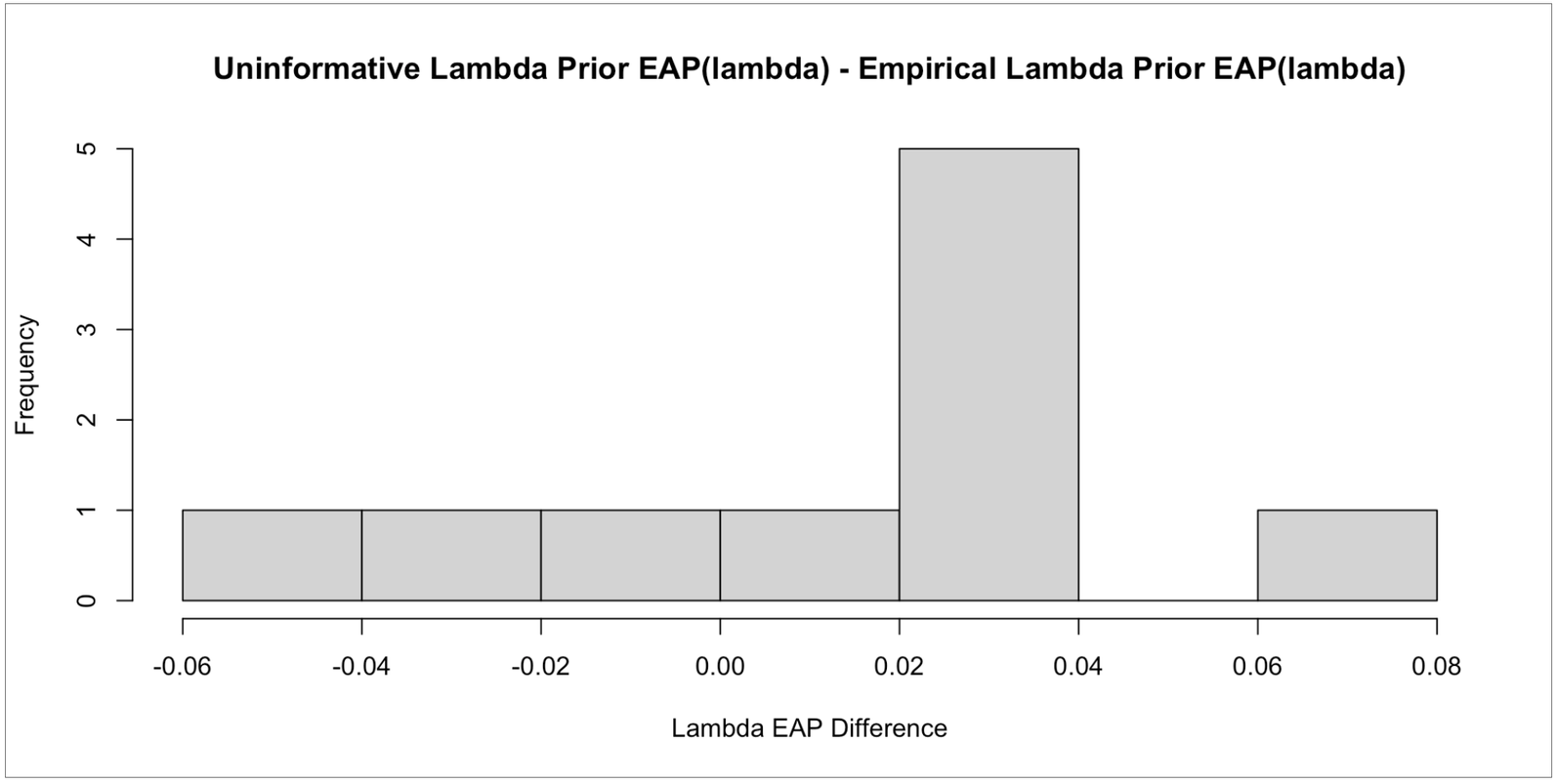
```
# item parameter results
print(
  modelCFA2_samples$summary(
    variables = c("mu", "meanMu", "sdMu", "lambda", "meanLambda", "sdLambda", "psi", "psiRate")
    ),
  n=Inf
)
```

```
# A tibble: 35 × 10
   variable     mean median     sd    mad     q5   q95  rhat ess_bulk ess_tail
   <chr>       <dbl>  <dbl>  <dbl>  <dbl>  <dbl> <dbl> <dbl>    <dbl>    <dbl>
 1 mu[1]      2.31   2.31   0.0891 0.0876 2.16   2.45   1.01     609.   1613.
 2 mu[2]      1.94   1.94   0.0837 0.0856 1.80   2.07   1.01     423.   1136.
 3 mu[3]      1.86   1.86   0.0828 0.0838 1.72   1.99   1.01     431.   1370.
 4 mu[4]      1.99   1.99   0.0815 0.0783 1.85   2.12   1.01     479.   1103.
 5 mu[5]      1.96   1.97   0.0801 0.0800 1.83   2.09   1.01     363.    825.
 6 mu[6]      1.88   1.88   0.0748 0.0737 1.75   2.00   1.01     388.    922.
 7 mu[7]      1.72   1.72   0.0777 0.0774 1.59   1.84   1.01     387.   1865.
 8 mu[8]      1.83   1.83   0.0701 0.0678 1.71   1.94   1.01     382.    805.
 9 mu[9]      1.80   1.80   0.0841 0.0816 1.66   1.93   1.01     487.   1131.
10 mu[10]     1.54   1.54   0.0844 0.0836 1.40   1.67   1.01     393.   1871.
11 meanMu     1.87   1.87   0.0998 0.0960 1.71   2.03   1.01     758.   2895.
12 sdMu       0.241  0.230  0.0734 0.0603 0.149  0.375  1.00    6825.   5291.
13 lambda[1]  0.767  0.769  0.0695 0.0692 0.653  0.880  1.00    1632.   2833.
14 lambda[2]  0.838  0.835  0.0652 0.0670 0.736  0.950  1.01    1092.   2945.
15 lambda[3]  0.798  0.797  0.0644 0.0618 0.696  0.908  1.00    1332.   2906.
16 lambda[4]  0.822  0.820  0.0640 0.0651 0.721  0.928  1.00    1227.   2949.
17 lambda[5]  0.930  0.930  0.0699 0.0695 0.808  1.05   1.01     429.     64.5
18 lambda[6]  0.863  0.861  0.0592 0.0607 0.772  0.964  1.01     689.   1936.
19 lambda[7]  0.772  0.771  0.0612 0.0600 0.673  0.875  1.00    1406.   2633.
20 lambda[8]  0.829  0.828  0.0559 0.0579 0.743  0.925  1.00     879.   2358.
21 lambda[9]  0.830  0.827  0.0647 0.0658 0.727  0.941  1.00    1192.   3095.
22 lambda[10] 0.727  0.729  0.0699 0.0694 0.610  0.841  1.00    1684.   3555.
23 meanLambda 0.817  0.815  0.0545 0.0555 0.731  0.908  1.01     888.   2226.
```

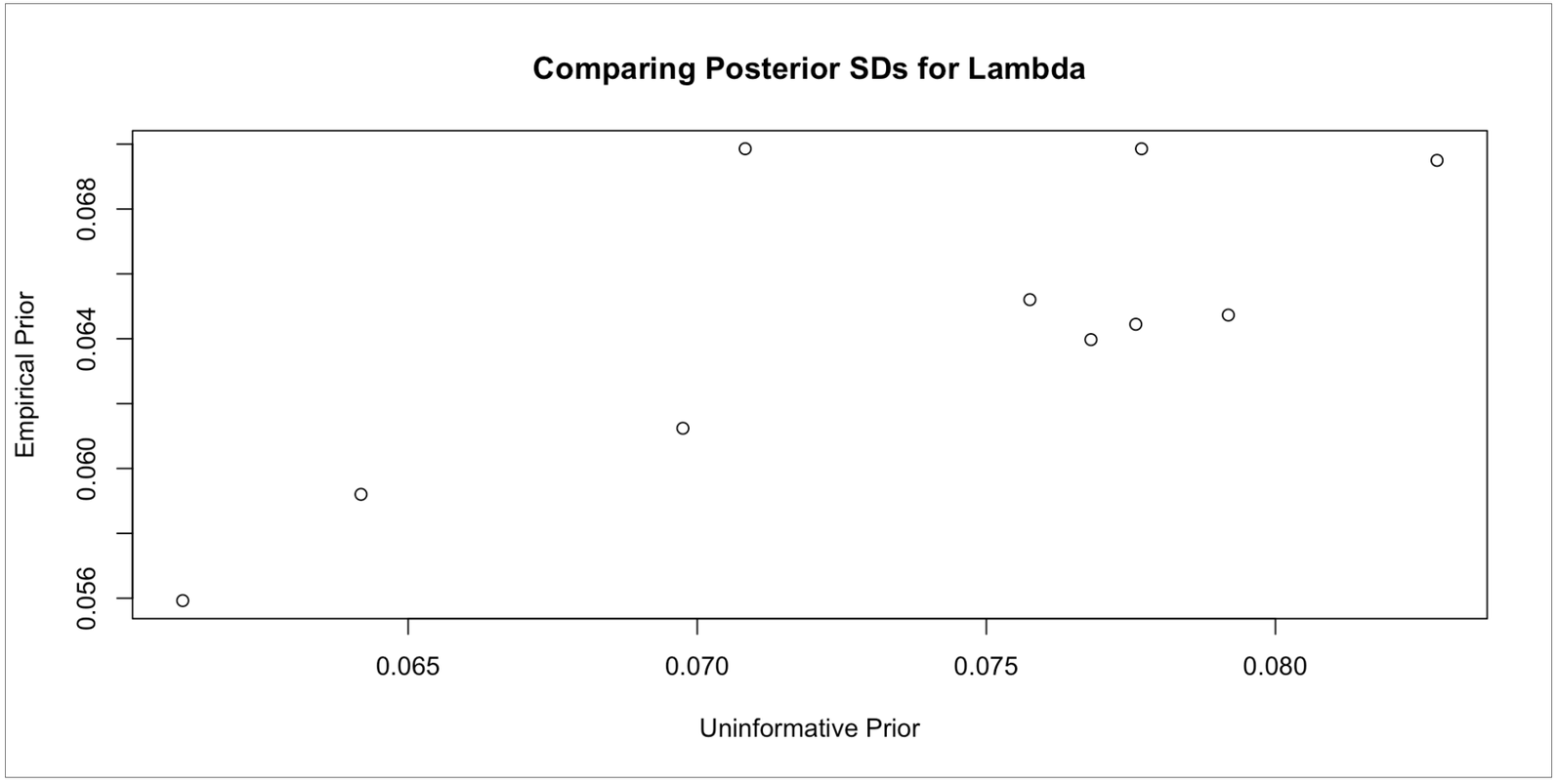
Comparisons with Non-Empirical Priors: λ



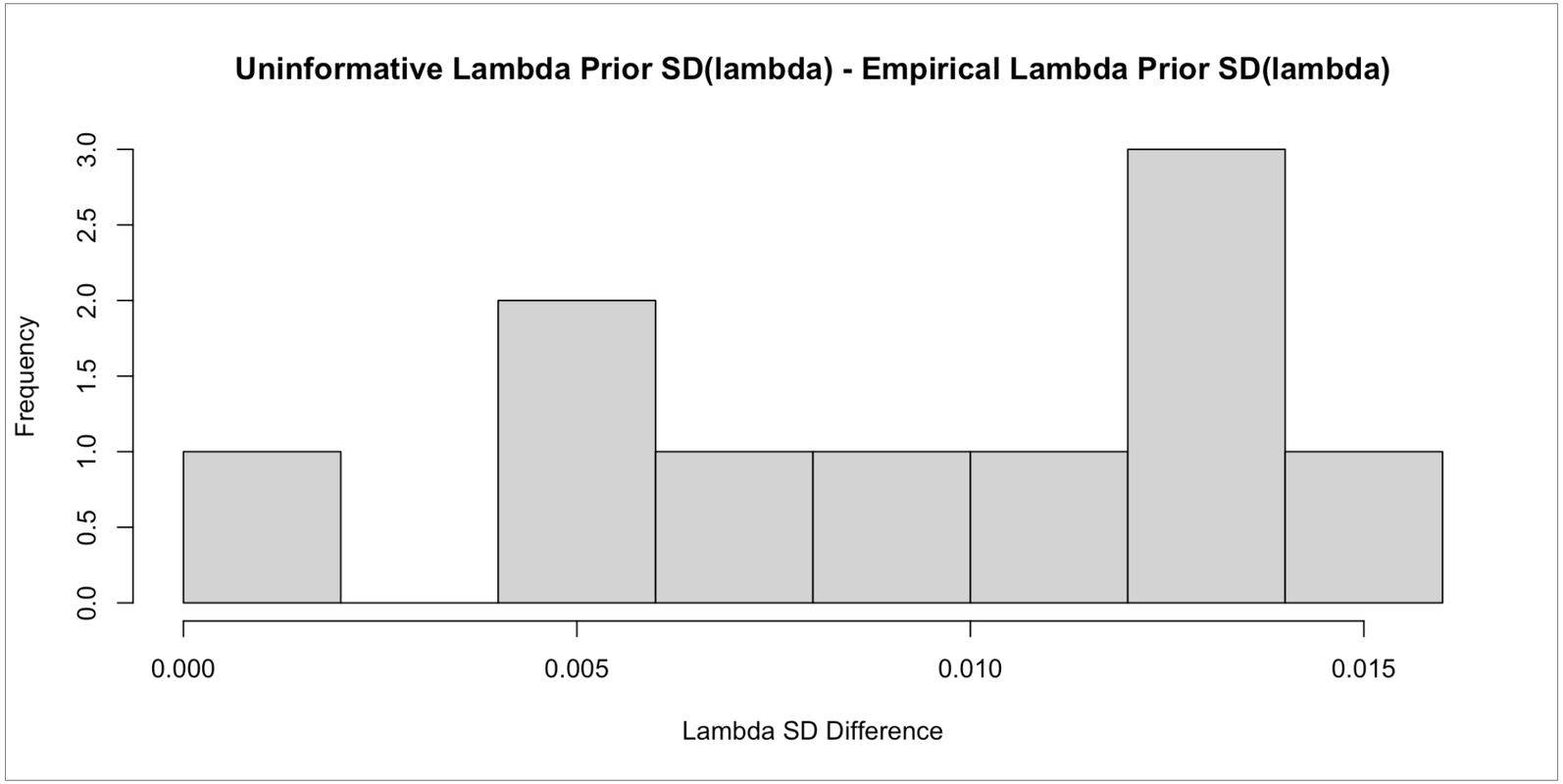
Comparisons with Non-Empirical Priors: λ



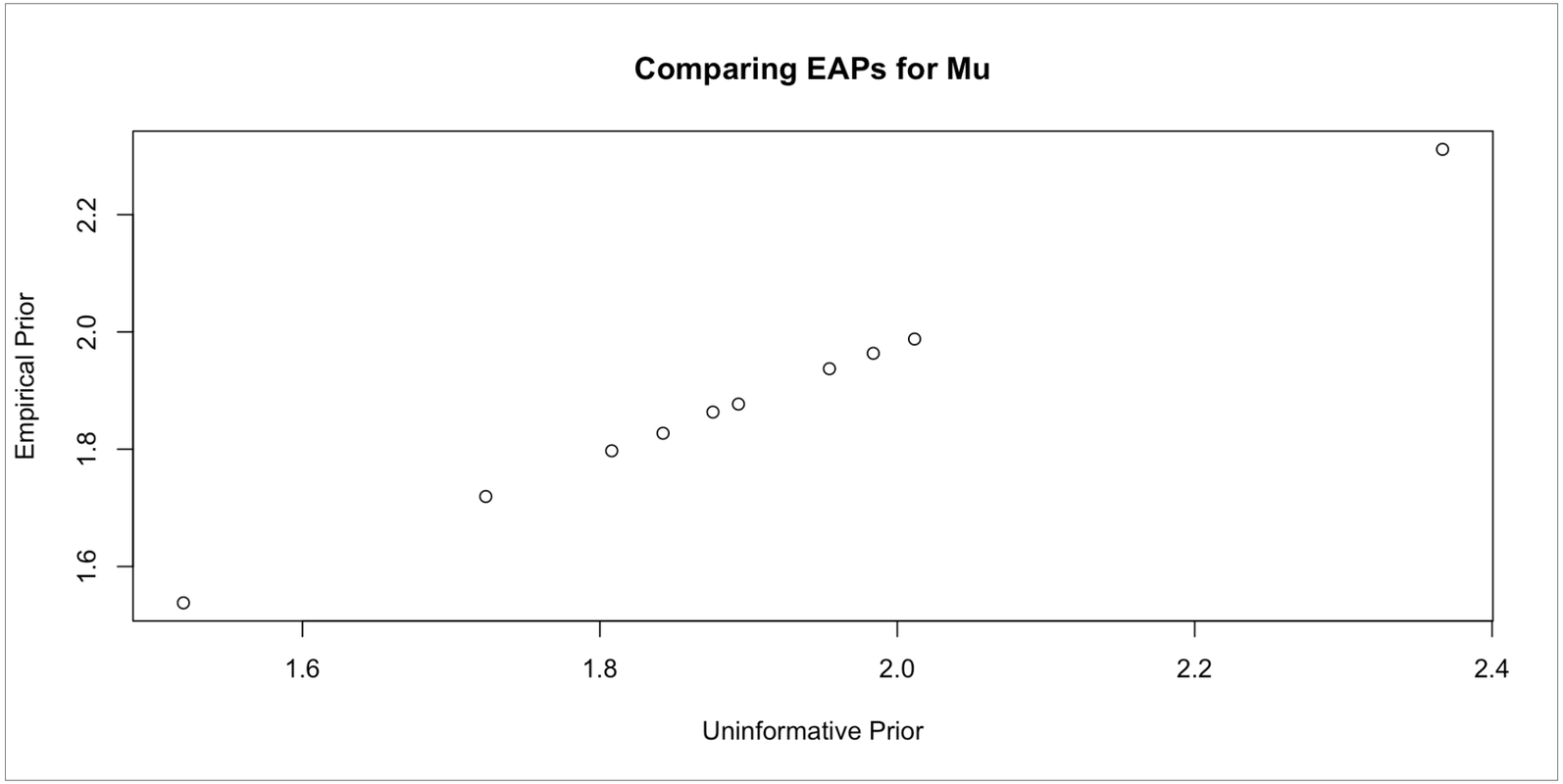
Comparisons with Non-Empirical Priors: λ



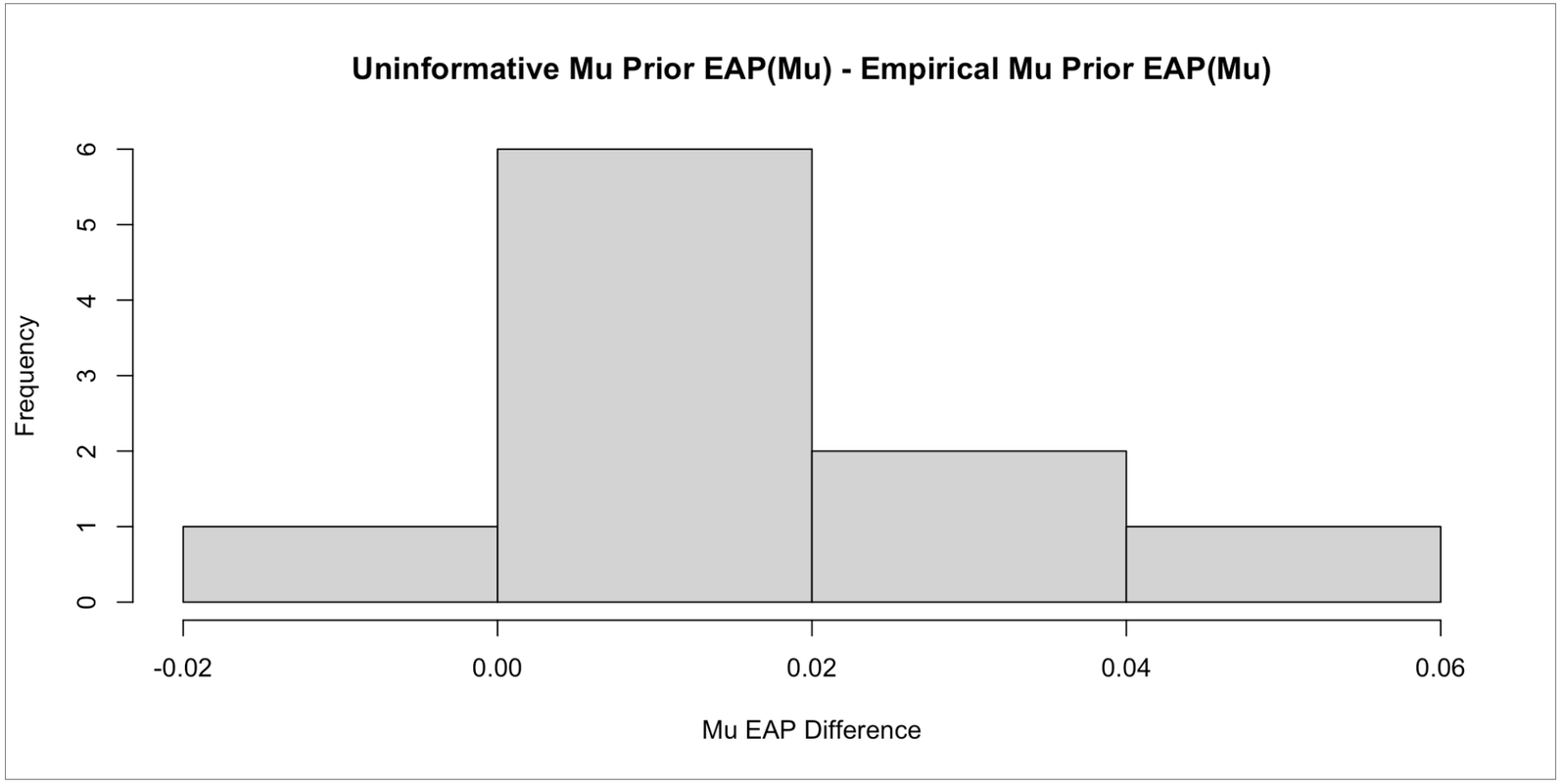
Comparisons with Non-Empirical Priors: λ



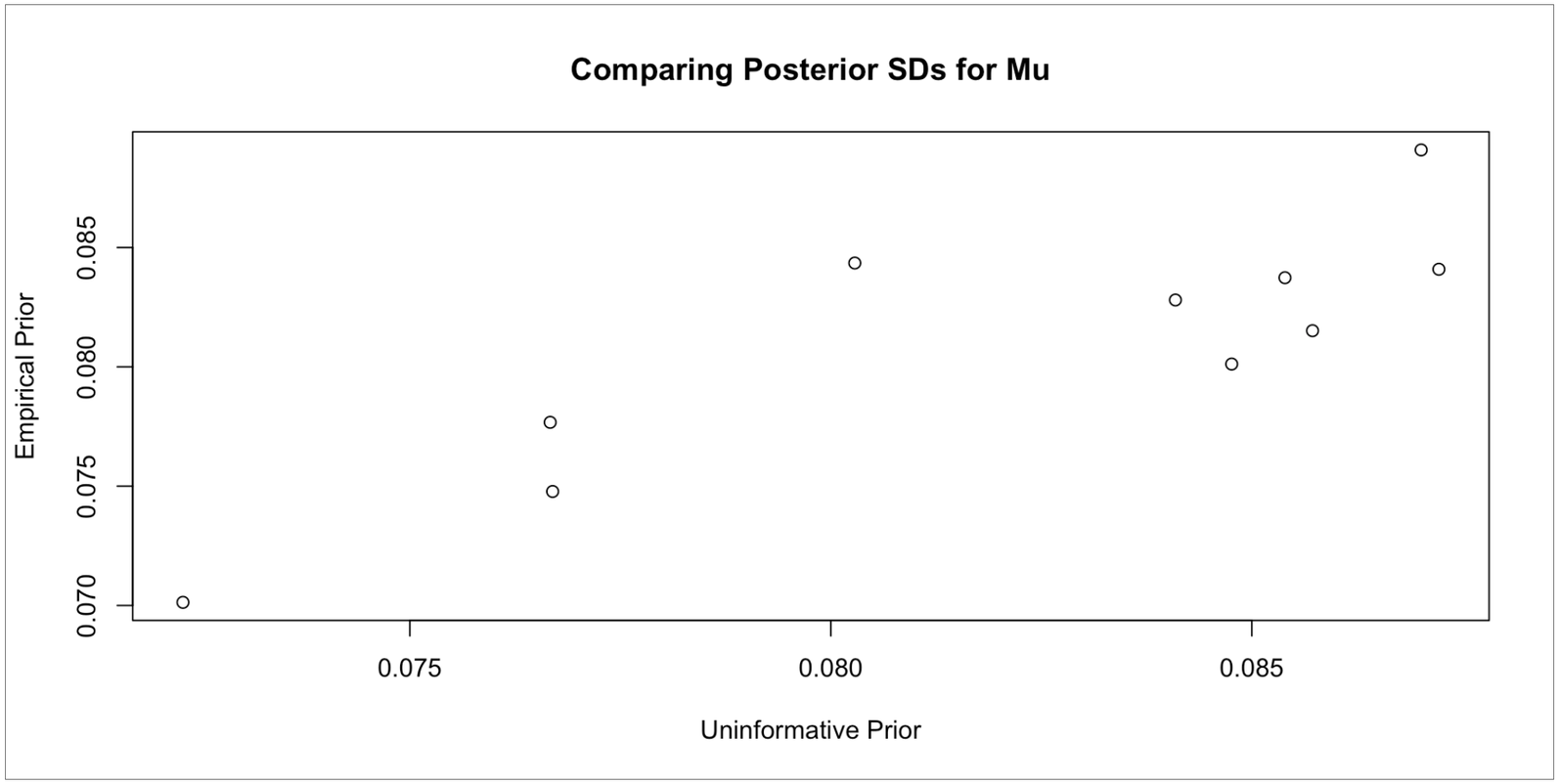
Comparisons with Non-Empirical Priors: μ



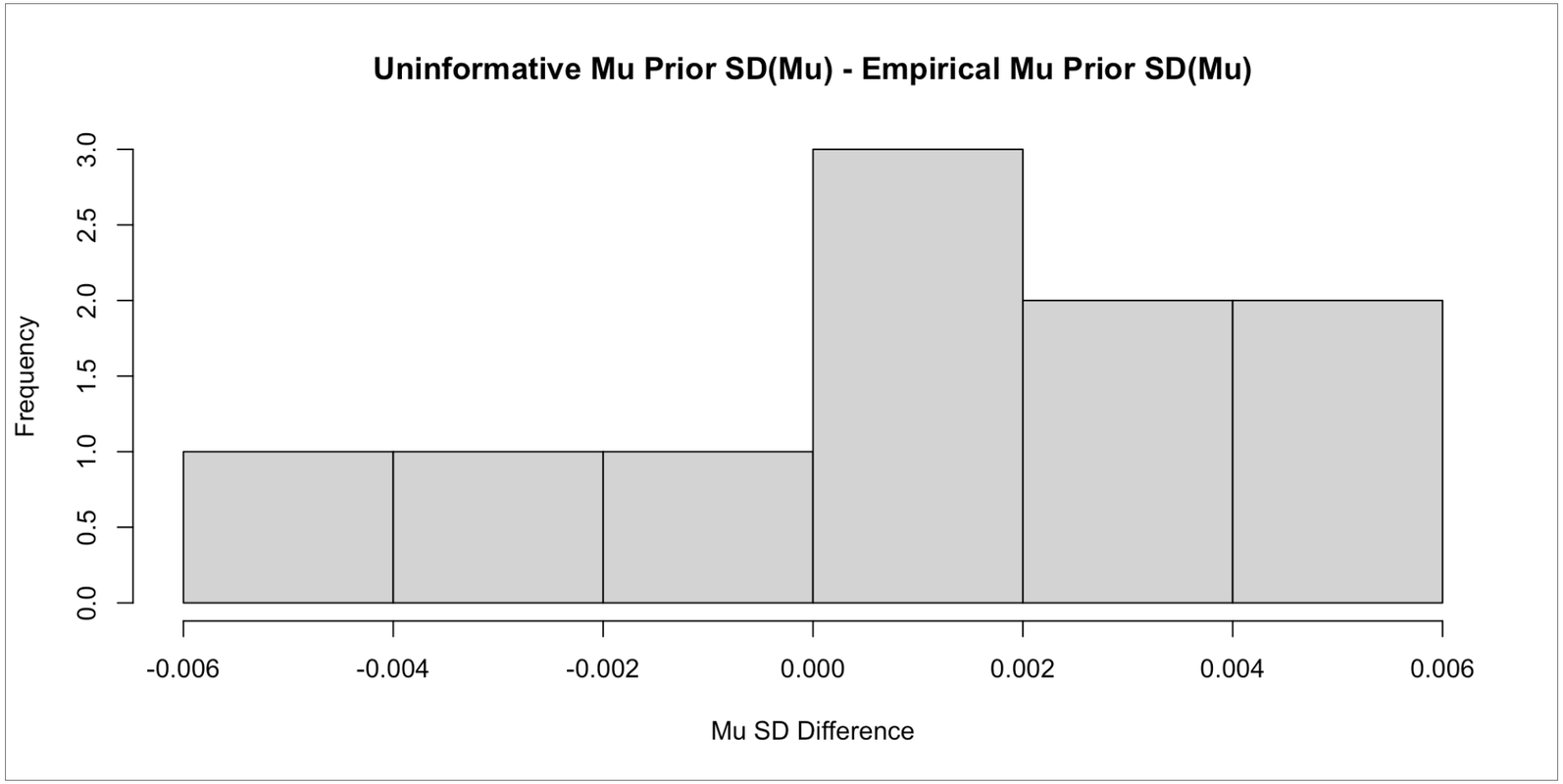
Comparisons with Non-Empirical Priors: μ



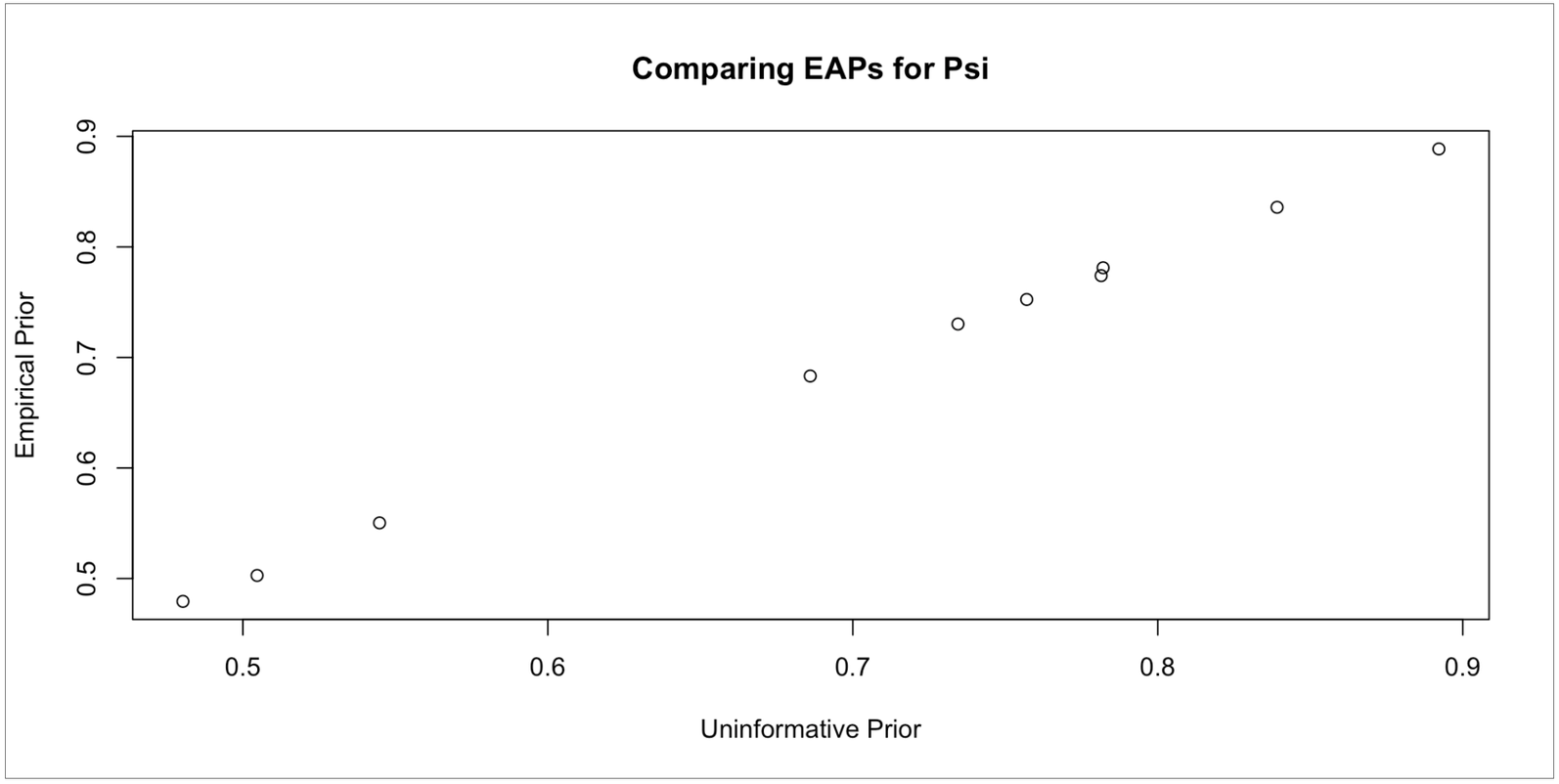
Comparisons with Non-Empirical Priors: μ



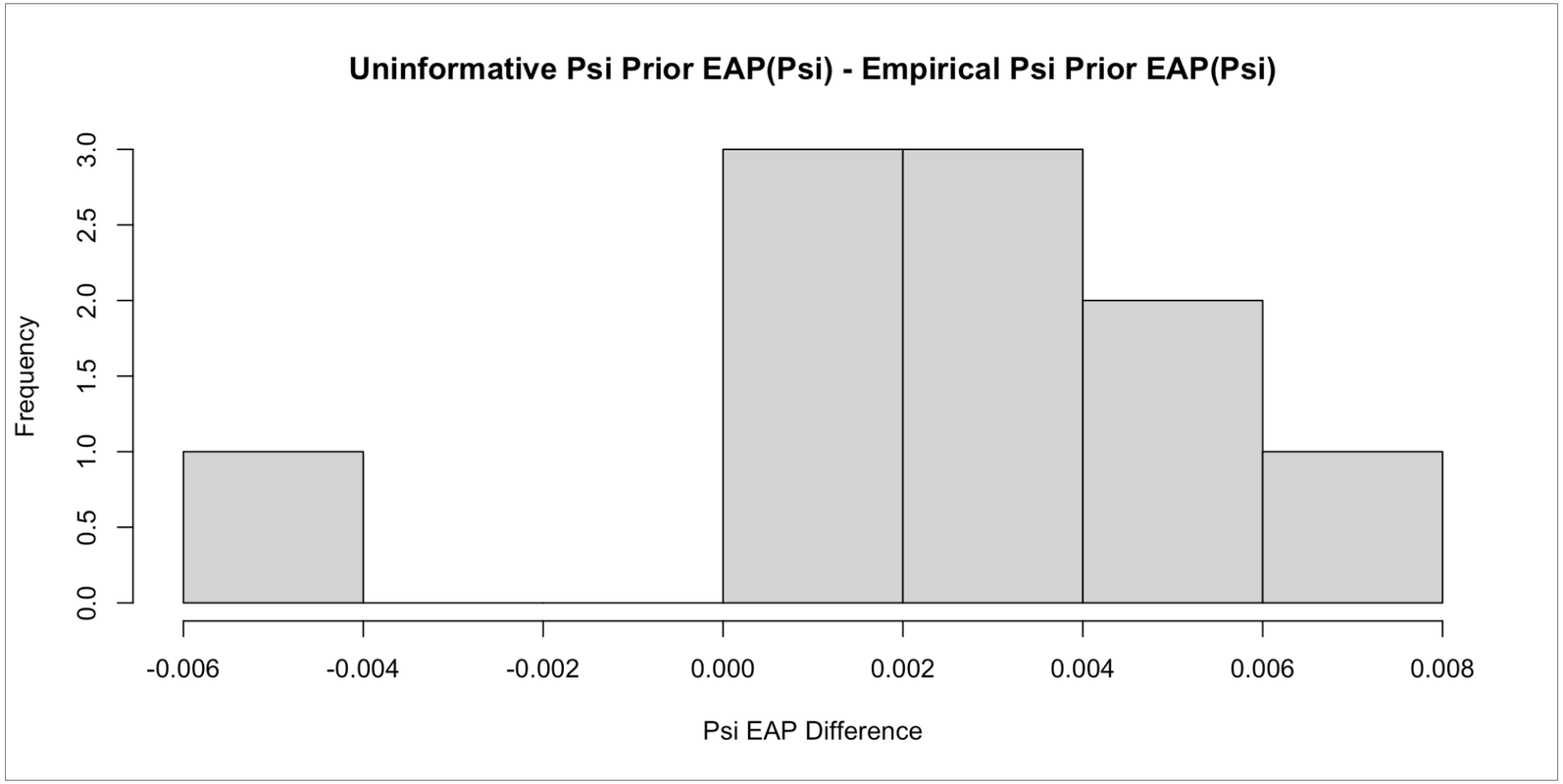
Comparisons with Non-Empirical Priors: μ



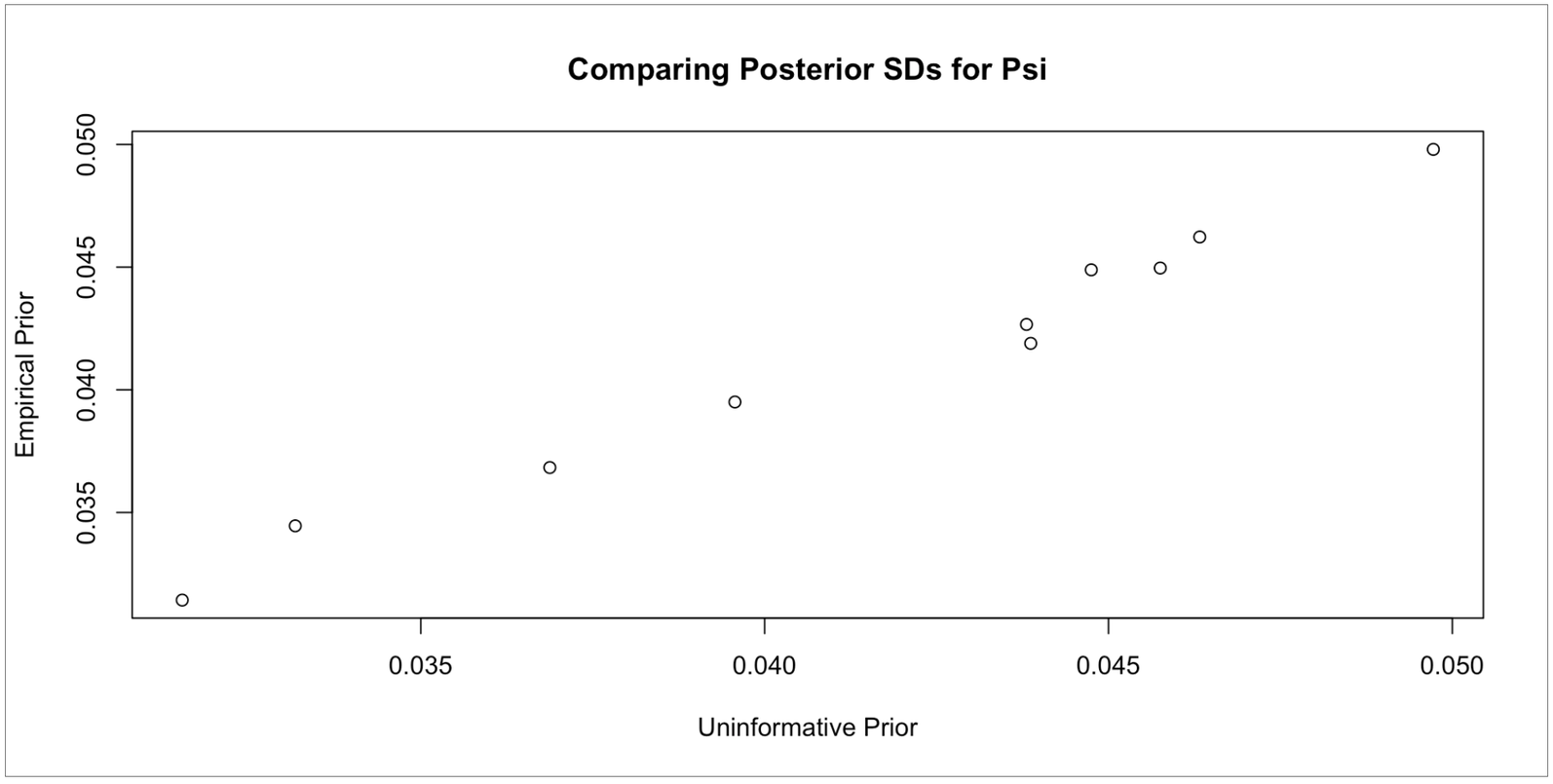
Comparisons with Non-Empirical Priors: ψ



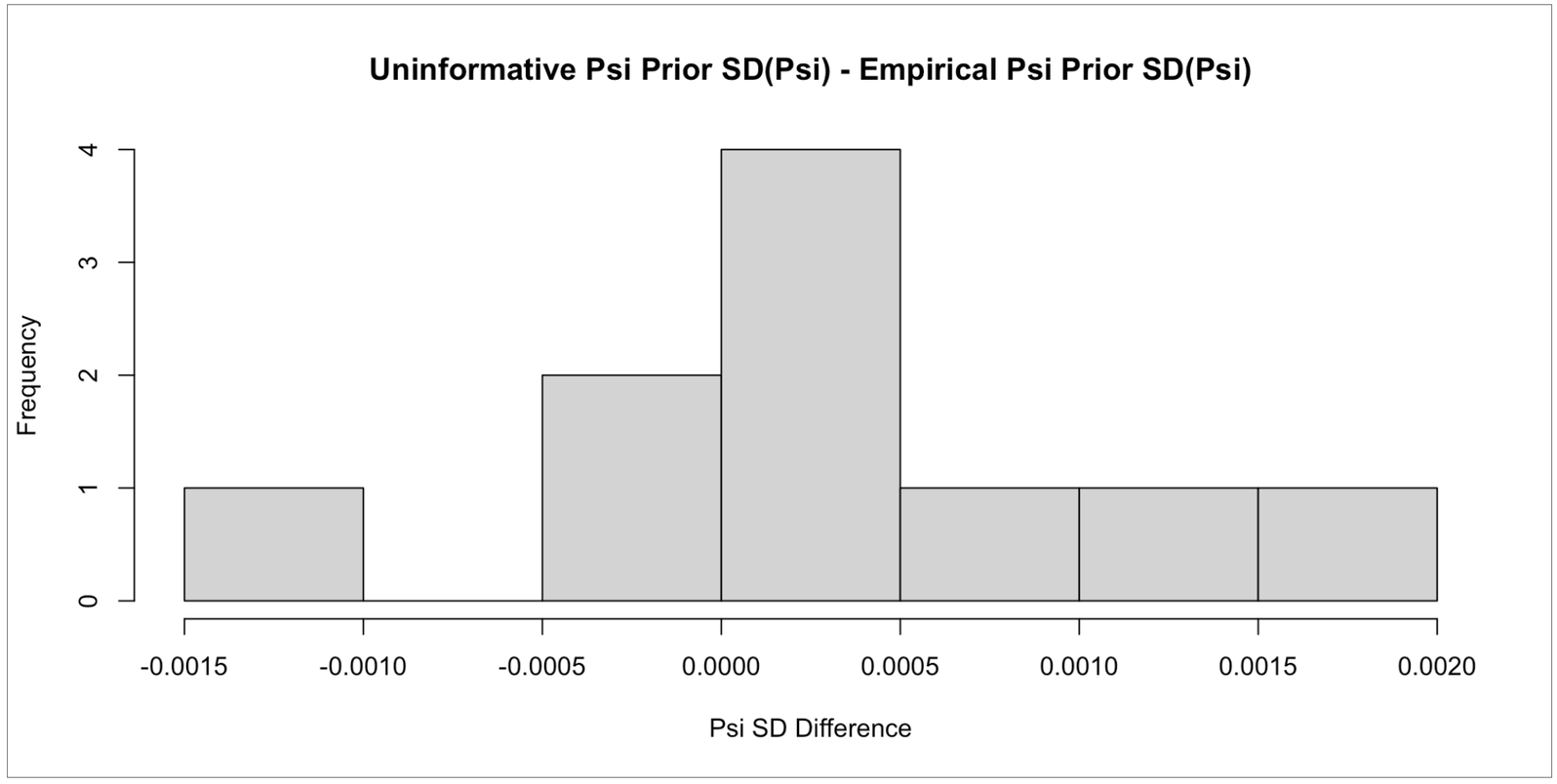
Comparisons with Non-Empirical Priors: ψ



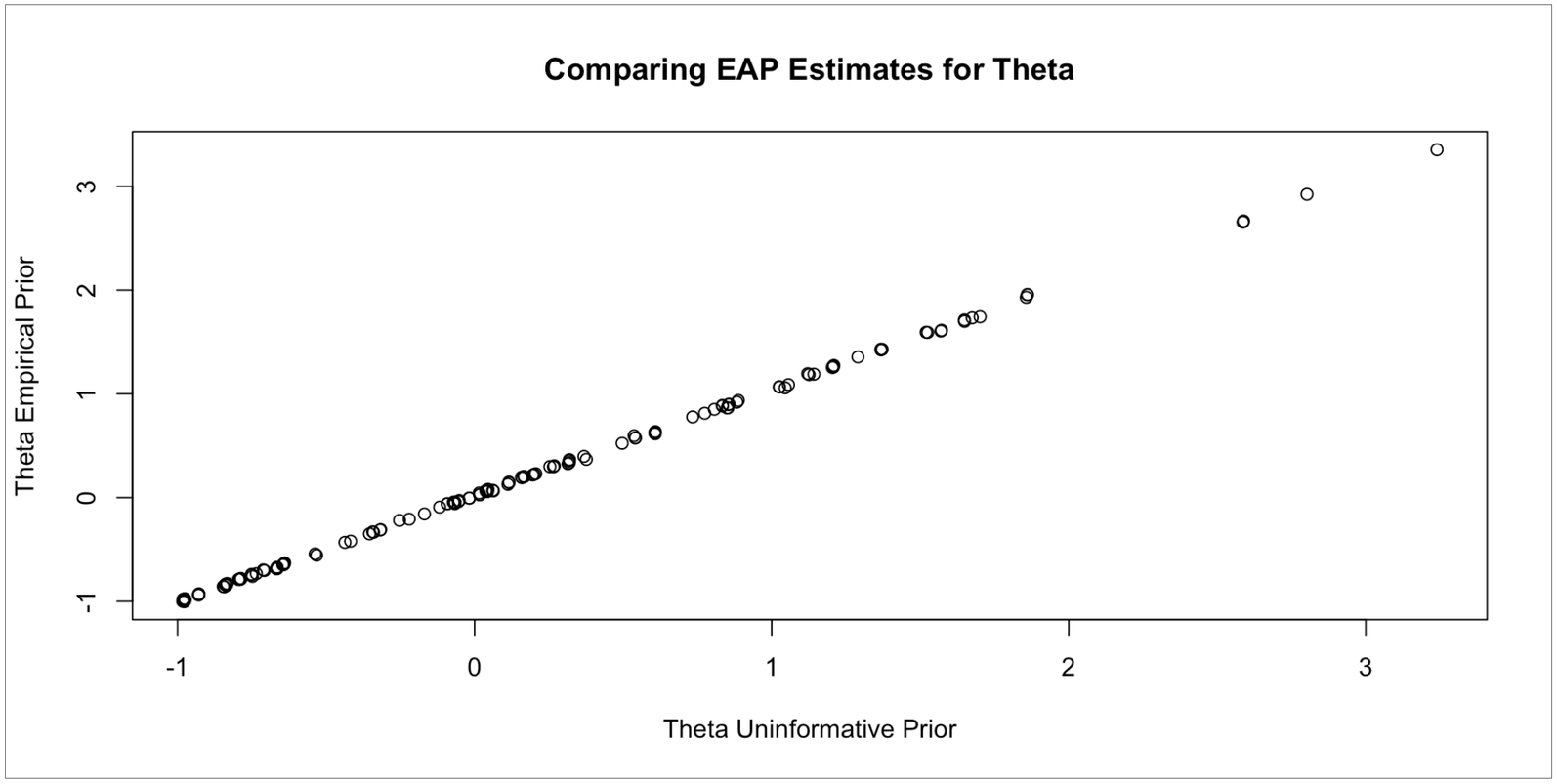
Comparisons with Non-Empirical Priors: ψ



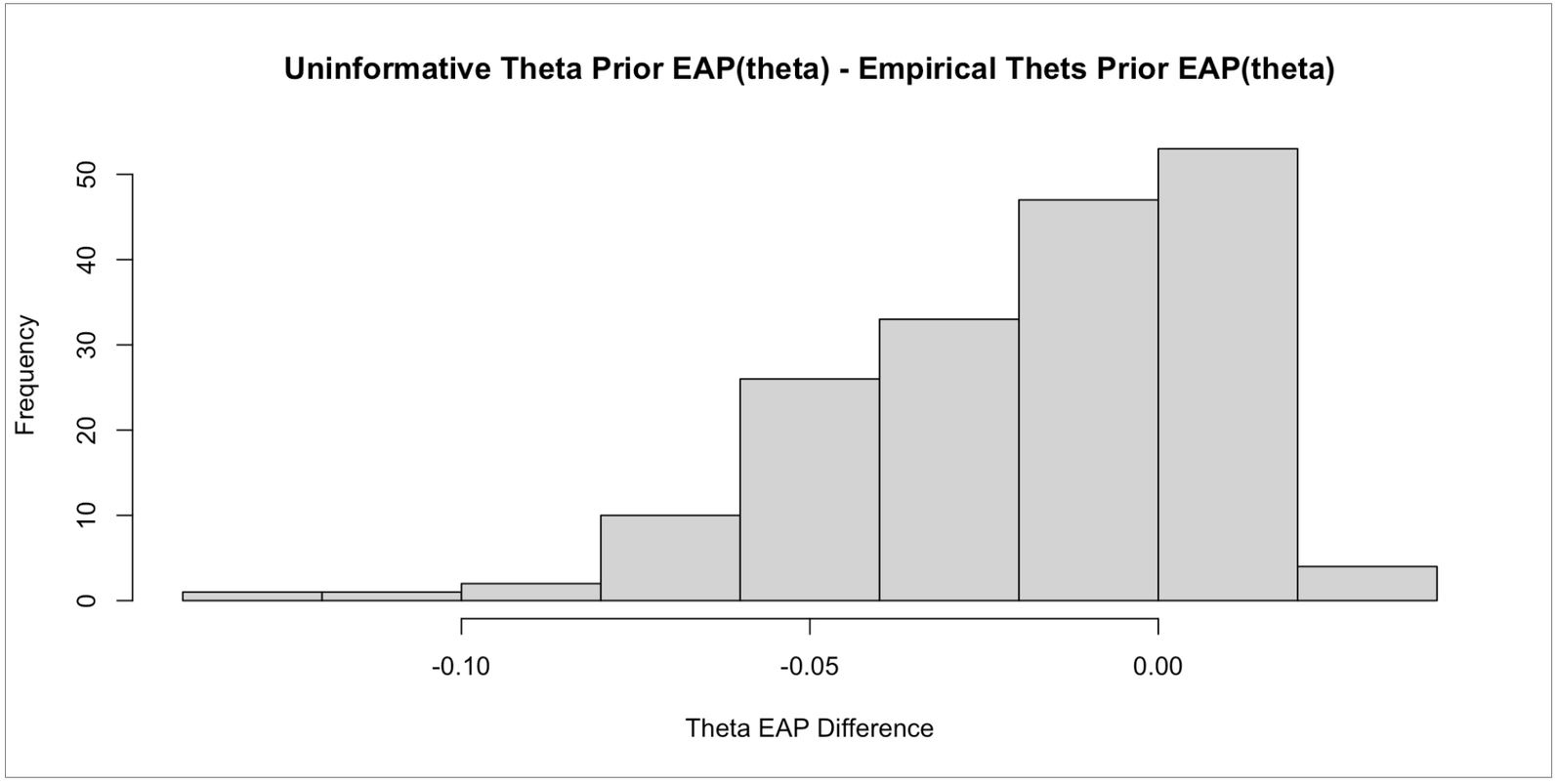
Comparisons with Non-Empirical Priors: ψ



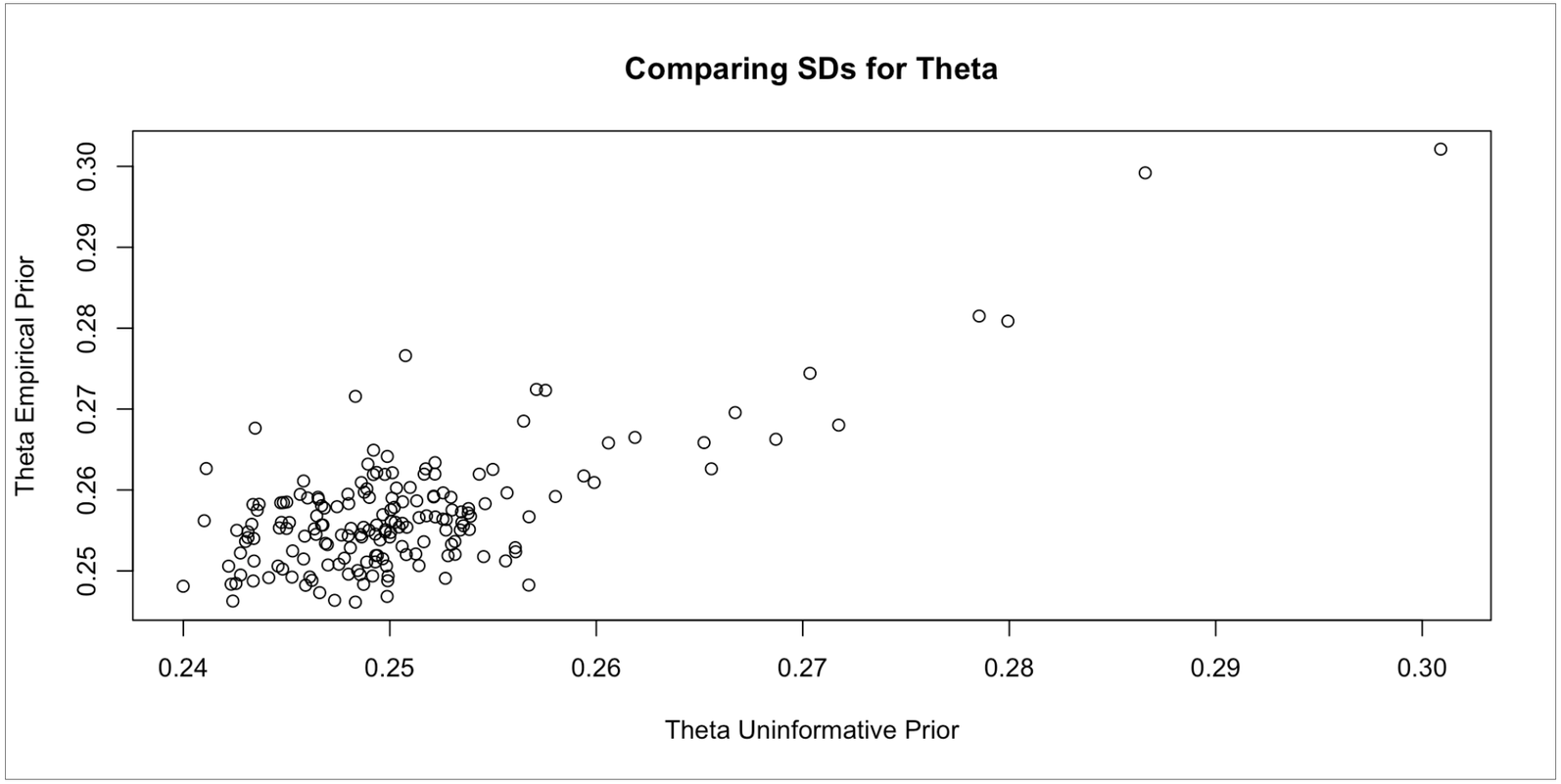
Comparisons with Non-Empirical Priors: θ



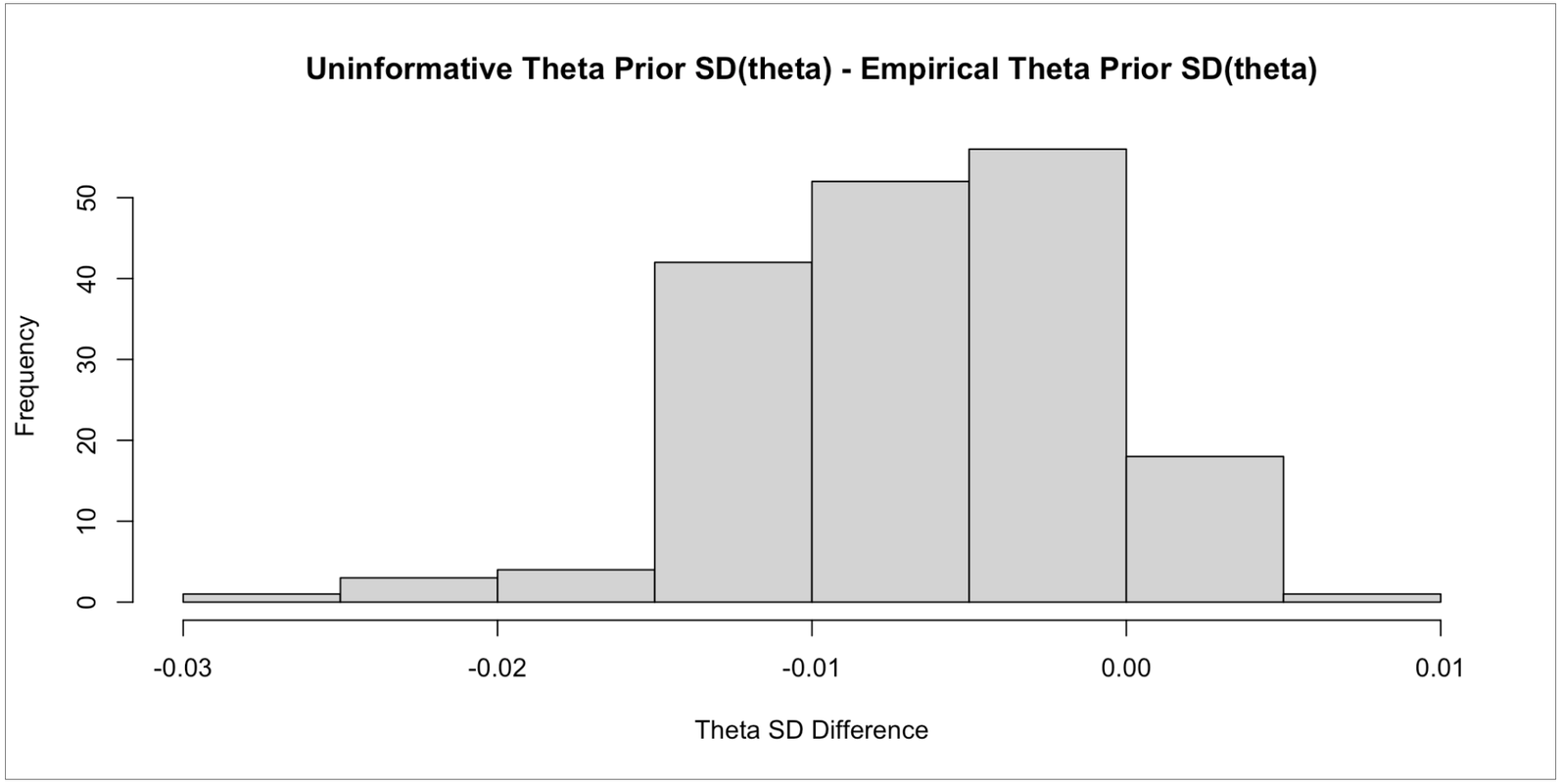
Comparisons with Non-Empirical Priors: θ



Comparisons with Non-Empirical Priors: θ



Comparisons with Non-Empirical Priors: θ



Empirical Priors for θ (?)

Empirical Priors for θ

If empirical priors can work for the item parameters, can we use empirical priors to estimate the mean/standard deviation of the latent variable θ?

- In short: No!
- The reason: Empirical priors for θ change the mean/standard deviation of the latent variable
  - These quantities have to be set for identification
- We've set them to mean = 0 and standard deviation =1 (standardized factor) throughout this class
  - We would have to fix other parameters to estimate these values (see next lecture)
- For now, let me show you what would happen
  - No syntax here, just results

First Attempt: Everything Empirical

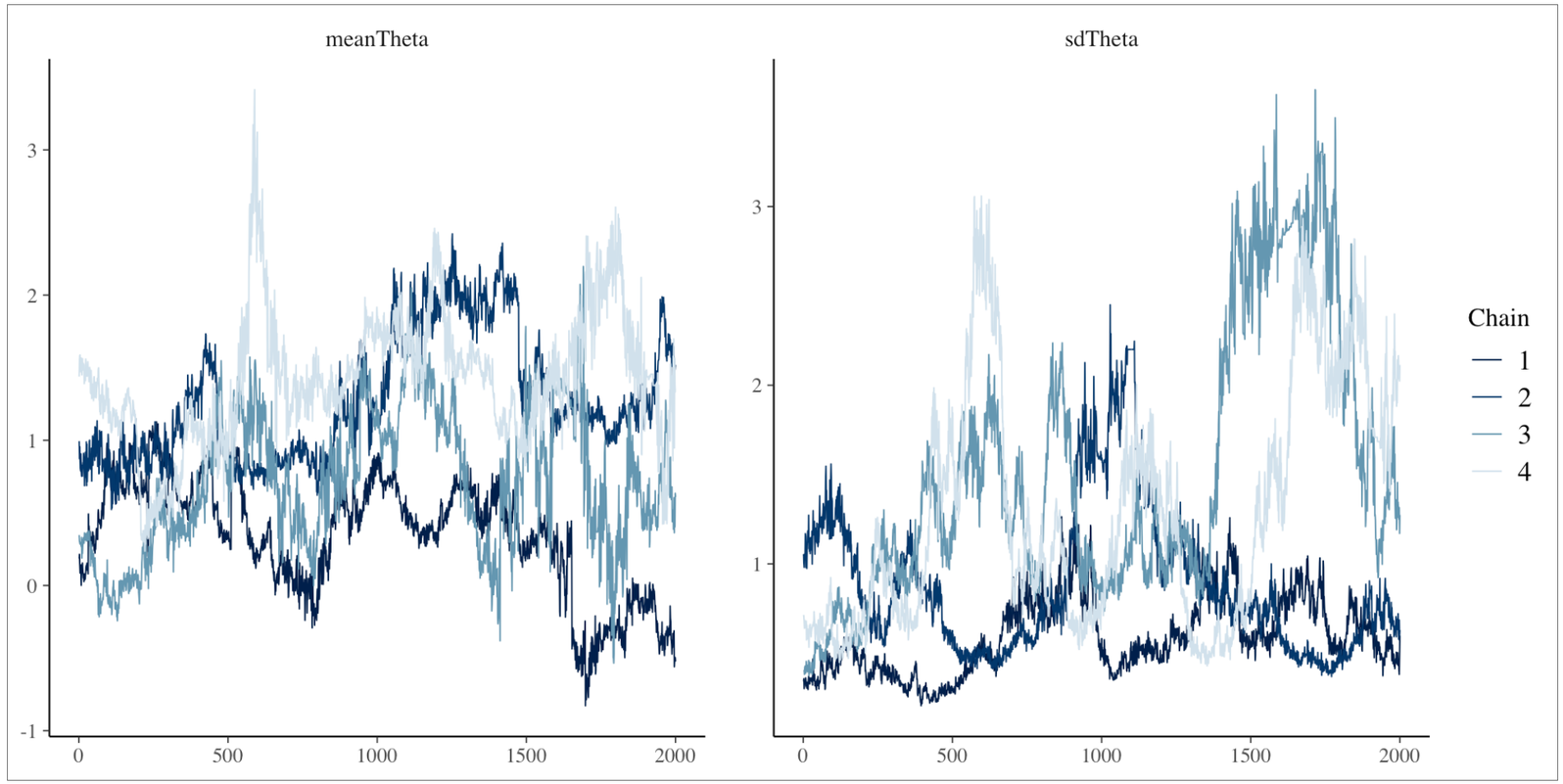
```
# checking convergence
max(modelCFA3_samples$summary()$rhat, na.rm = TRUE)
```

```
[1] 1.603846
```

```
# item parameter results
print(
  modelCFA3_samples$summary(
    variables = c("meanTheta", "sdTheta", "mu", "meanMu", "sdMu", "lambda", "meanLambda", "sdLambda", "psi", "psiRate")
  ),
  n=Inf
)
```

```
# A tibble: 37 × 10
   variable    mean median     sd    mad      q5   q95  rhat ess_bulk ess_tail
   <chr>      <dbl>  <dbl>  <dbl>  <dbl>   <dbl> <dbl> <dbl>    <dbl>    <dbl>
 1 meanTheta  0.937 0.925  0.617  0.641  -0.0902 1.98   1.60     6.88     19.4
 2 sdTheta    1.07  0.870  0.656  0.487   0.393  2.58   1.37     9.72     15.5
 3 mu[1]      1.45  1.57   0.677  0.659   0.240  2.42   1.34     9.89     19.2
 4 mu[2]      1.06  1.17   0.673  0.657  -0.137  2.04   1.34     9.72     18.9
 5 mu[3]      1.03  1.13   0.647  0.623  -0.135  1.98   1.34     9.72     19.6
 6 mu[4]      1.12  1.23   0.668  0.644  -0.0690 2.09   1.34     9.75     18.1
 7 mu[5]      1.00  1.10   0.729  0.722  -0.289  2.08   1.34     9.67     18.5
 8 mu[6]      0.982 1.08   0.685  0.675  -0.235  1.98   1.35     9.64     18.2
 9 mu[7]      0.915 1.01   0.623  0.598  -0.217  1.82   1.34     9.80     18.7
10 mu[8]      0.966 1.07   0.661  0.640  -0.209  1.93   1.35     9.68     18.2
11 mu[9]      0.943 1.03   0.658  0.643  -0.232  1.91   1.34     9.73     19.1
12 mu[10]     0.788 0.875  0.584  0.559  -0.254  1.64   1.34     9.80     18.8
13 meanMu     1.02  1.11   0.657  0.645  -0.142  1.99   1.34     9.74     19.5
14 sdMu       0.225 0.212  0.0780 0.0655  0.124  0.367  1.01   800.     1280.
15 lambda[1]  1.02  0.909  0.573  0.556   0.300  2.06   1.35    10.3      16.3
16 lambda[2]  1.07  0.962  0.585  0.570   0.323  2.12   1.37     9.74     14.5
17 lambda[3]  1.02  0.911  0.556  0.548   0.305  2.01   1.37     9.79     15.1
18 lambda[4]  1.05  0.949  0.576  0.565   0.319  2.09   1.36     9.89     15.3
19 lambda[5]  1.18  1.06   0.639  0.632   0.360  2.32   1.38     9.54     15.3
20 lambda[6]  1.10  0.984  0.596  0.595   0.332  2.16   1.38     9.64     15.3
21 lambda[7]  0.977 0.879  0.534  0.522   0.296  1.95   1.37     9.84     15.5
22 lambda[8]  1.05  0.943  0.574  0.563   0.319  2.07   1.38     9.68     14.4
23 lambda[9]  1.05  0.947  0.571  0.564   0.319  2.07   1.38     9.67     14.9
```

First Attempt: Everything Empirical



Second Attempt: Only Empirical for θ

Here, I set the prior distributions for each type of parameter to be much more informative:

- $\lambda_i \sim N(0, 1)$
- $\mu_i \sim N(0, 1)$
- $\psi_i \sim \text{exponential1}$

Then, I set the prior values for the mean and SD of θ:

- $\mu_\theta \sim N(0, 1)$
- $\sigma_\theta \sim \text{exponential1}$

Second Attempt: Only Empirical for θ

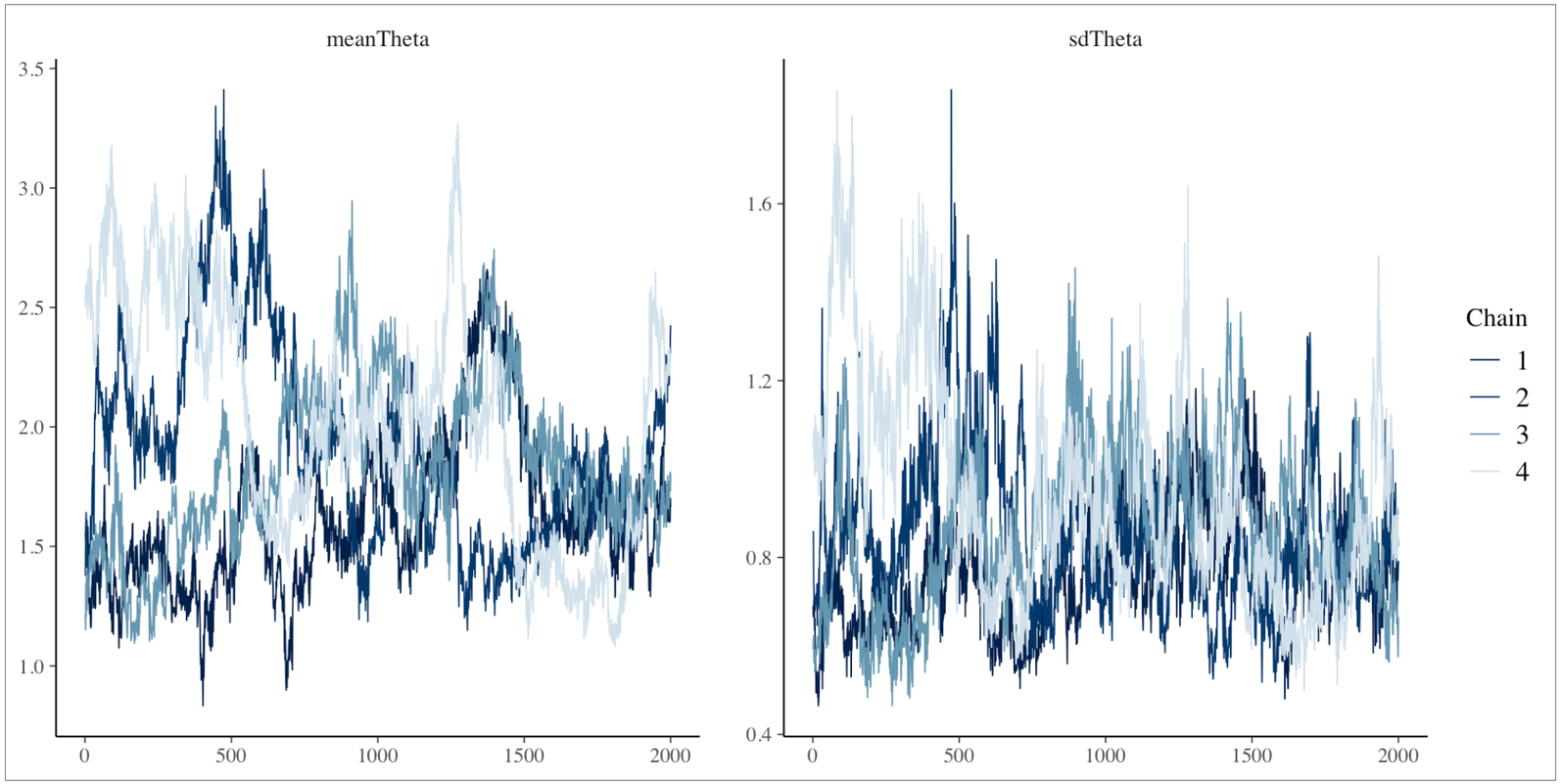
```
# checking convergence
max(modelCFA4_samples$summary()$rhat, na.rm = TRUE)
```

```
[1] 1.234166
```

```
# item parameter results
print(modelCFA4_samples$summary(variables = c("meanTheta", "sdTheta", "mu", "lambda", "psi")) ,n=Inf)
```

```
# A tibble: 32 × 10
   variable     mean   median     sd    mad     q5   q95  rhat ess_bulk ess_tail
   <chr>       <dbl>    <dbl>  <dbl>  <dbl>  <dbl> <dbl> <dbl>    <dbl>    <dbl>
 1 meanTheta  1.86    1.79e+0 0.427  0.427   1.28  2.66   1.23     14.6     68.9
 2 sdTheta    0.858   8.25e-1 0.196  0.170   0.602 1.23   1.15     19.0     48.2
 3 mu[1]      0.786   7.99e-1 0.288  0.293   0.295 1.24   1.03    168.     452.
 4 mu[2]      0.111   1.23e-1 0.306  0.308  -0.422 0.596  1.03    150.     354.
 5 mu[3]      0.175   1.83e-1 0.293  0.294  -0.329 0.630  1.03    160.     412.
 6 mu[4]      0.221   2.32e-1 0.302  0.303  -0.297 0.695  1.03    152.     370.
 7 mu[5]     -0.133  -1.23e-1 0.336  0.339  -0.704 0.393  1.04    132.     313.
 8 mu[6]     -0.0165 -7.93e-3 0.304  0.304  -0.534 0.457  1.04    132.     291.
 9 mu[7]      0.100   1.09e-1 0.273  0.277  -0.367 0.526  1.04    150.     386.
10 mu[8]      0.0289  3.53e-2 0.288  0.292  -0.459 0.478  1.04    132.     280.
11 mu[9]     -0.0130  3.94e-4 0.308  0.308  -0.538 0.473  1.03    154.     379.
12 mu[10]     0.0989  1.13e-1 0.264  0.260  -0.365 0.502  1.03    178.     448.
13 lambda[1]  0.875   8.67e-1 0.192  0.191   0.571 1.21   1.13     22.2     58.0
14 lambda[2]  1.02    1.02e+0 0.218  0.218   0.679 1.40   1.14     19.9     45.5
15 lambda[3]  0.944   9.34e-1 0.205  0.203   0.617 1.30   1.14     20.8     48.8
16 lambda[4]  0.993   9.83e-1 0.213  0.211   0.655 1.37   1.14     19.9     50.7
17 lambda[5]  1.17    1.17e+0 0.242  0.239   0.785 1.59   1.15     18.6     44.5
18 lambda[6]  1.06    1.06e+0 0.221  0.215   0.704 1.44   1.16     18.3     43.4
19 lambda[7]  0.901   8.93e-1 0.193  0.192   0.596 1.24   1.14     20.5     49.6
20 lambda[8]  1.01    9.99e-1 0.209  0.204   0.670 1.37   1.15     19.0     45.5
21 lambda[9]  1.01    1.00e+0 0.217  0.212   0.669 1.39   1.14     20.1     48.4
22 lambda[1…  0.789   7.79e-1 0.178  0.180   0.514 1.10   1.13     22.2     48.3
23 psi[1]     0.892   8.90e-1 0.0517 0.0516  0.812 0.983  1.00   3336.     4216.
```

Second Attempt: Everything Empirical



Wrapping Up

Wrapping Up

Empirical priors tend to be somewhat common in Bayesian analyses

- But, they can be difficult to implement fully in psychometric models
  - Item parameters can work fine – but may end up putting too much weight on some parameters
  - Different types of models will have differing results
- Empirical priors for θ require more work
  - Need more constraints to identify the model