STAT 365 Project

Identifying Stock Market Regimes Using State-Space and Classification Models

Richard Chen

12/9/2016

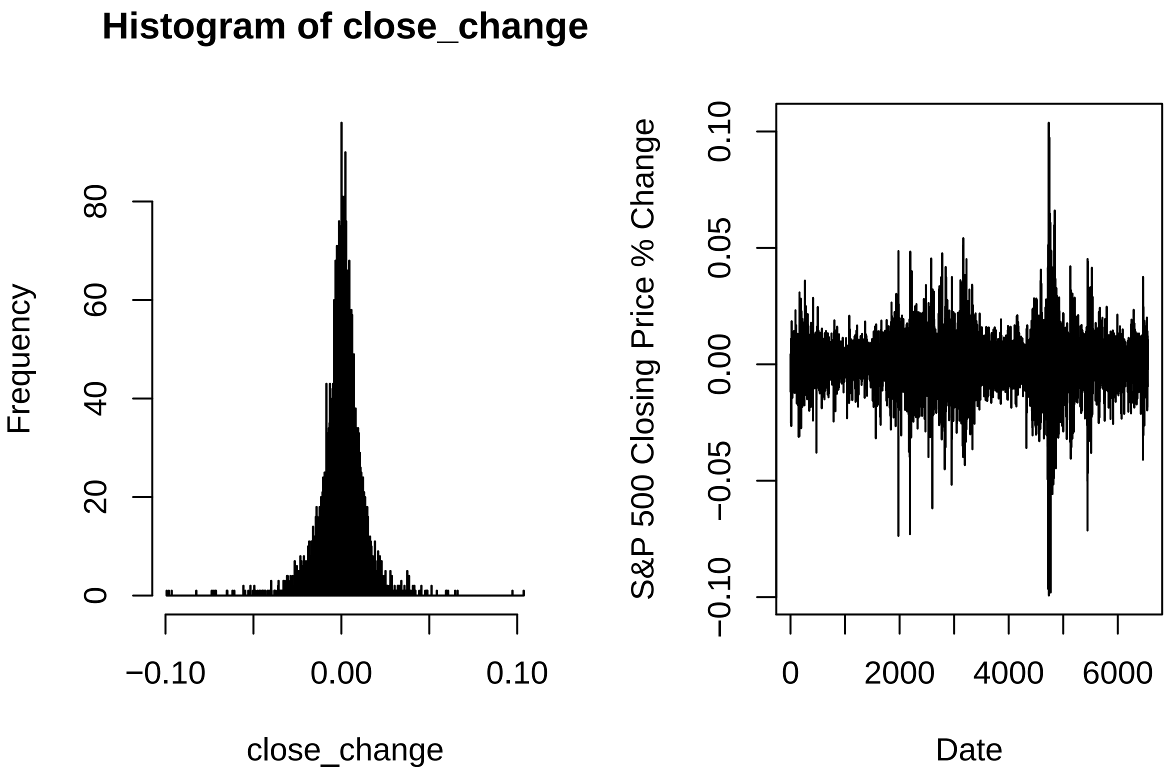
```
## Warning: package 'nimble' was built under R version 3.3.2
```

Introduction

In this report, we classify the stock market from 1990 to the end of 2015 as either a bull or bear market. We define a state-space model that assumes there are two possible states, bull and bear, as the hidden states of the stock market. We then use it to estimate the transition probabilities of changing from one state to the other and also of staying in the same state. Using these probabilities, we then define a classification model that estimates the probability of being in a bull market.

This motivation is that of this analysis is that it is often useful and interesting to know whether the current state of the stock market is bullish or bearish. A bull market is generally understood as an upward trend in stock prices. A bear market is the opposite where there is a downward trend in stock prices [1]. The volatilities under each regime may also be different. Investors are interested in knowing the state of the market because if the state is a bull market they can expect that if that stock prices will go up and they can buy accordingly and sell later at higher prices. If the current state is a bear market, investors know that it is an unprofitable time to buy stocks because because the stock prices are expected to fall. They will not buy, but rather can seize the opportunity to make a profit by shorting stock and selling selling at the current price and buying it later at a lower price in the bear market. Either way, identifying the current market regime, bull or bear, can help investors manage strategies and inform the public about the broad state of the economy.

Data



The data used for this report are daily values of the S&P 500 index (ticker: ˆGSPC) from January 2, 1990 to December 31, 2015. They were downloaded from Yahoo Finance[4]. Using the S&P 500 data makes sense because it reflects the prices of the 500 largest companies in the U.S. and we are trying to identify broad market regimes in the U.S. It is also commonly used as the benchmark for the overall market. The specific data used in the models is the daily percent change of the S&P 500 closing prices. The percent difference is used because compared to the overall level of the index, it captures the behavior of the market, doesn't depend on its previous values, and is able to be fit with a normal distribution. Also, with percent changes can more easily spot the volatility of the market, which is often used as a measure of the risk of stock prices and a factor that some consider when identifying market regimes because sometimes volatility increases during bearish periods. Now we can classify based on the day to day behavior of the broad stock market. Also, the start date is chosen so that there are a substantial number of data points to draw inference from, as well as to allow us to easily identify the state of the initial value in the series of hidden states. Several sources, including Yardeni Research Inc.[4] identify this as a bull market. It is also close to the beginning of the historic 90's bull market that ended in the Dot Com bubble bursting around 2000. We will see how much of this we can capture in our classification model results.

State Space Model

The first model we build is a state space model that is used to estimate the transition probabilities of going from each state to the other states. We assume there are two hidden states of the stock market, bull and bear. This is a simplification, but remains a reasonable assumption given that these two are the most general states of the market and is usually reported as such in the news. The data we can observe that tells us something about the hidden states is the S&P 500 returns, which is the daily percent change.

The transition matrix is:

$$\begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix}$$

The subscripts represent the states transitioned from and the state transitioned to (i.e. p_{01} is the probability of transition to state 1 from state 0, bull to bear market).

Model

The model is define below. The parameters' subscripts correspond to the states: 0 for bear market, 1 for bull market. p is the probability of a bull market. σ's are the standard deviations under bull or bear markets. We assume they are different because it is somtimes said that bear markets are associated with higher volatility. The x_is are the hidden states and y_is are the S&P 500 percent changes.

Top level parameters:

$$p_{01} \sim Uniform(0,1),\ p_{11} \sim Uniform(0,1),\ p_{00} = 1 - p_{01},\ p_{10} = 1 - p_{11}$$
$$p \sim Uniform(0,1)$$

$$\beta_0 \sim Normal(0, sd = 10000),\ \beta_1 \sim Normal(0, sd = 10000)$$
$$\sigma_0 \sim Uniform(0, 10000),\ \sigma_1 \sim Uniform(0, 10000)$$

$$x_1 \sim Bernoulli(p)$$

Latent states:

$$x_t \sim Bernoulli(p_{11} x_{t-1} + (1 - x_{t-1}) p_{01}),\ \text{for } t > 1$$

Likelihood:

$$y_t \sim Normal(\beta_0 + \beta_1 x_t)$$

```
code <- nimbleCode({
  #Priors
  #transition probabilities for going from 1 to 1, 0 to 0.
  #State 0 is bear market. State 1 is bull market
  p01 ~ dunif(0,1)
  p11 ~ dunif(0,1)
  p00 <- 1 - p01
  p10 <- 1- p11
  p ~ dunif(0,1)

  b0 ~ dnorm(0, sd = 10000)
  b1 ~ dnorm(0, sd = 10000)

  sigma0 ~ dunif(0,10000)
  sigma1 ~ dunif(0,10000)

  x[1] ~ dbinom(size=1, prob = p)
  y[1] ~ dnorm(b0 + bbull*x[1], sd = sigma1*x[1] + (1-x[1])*sigma0)

    for (t in 2:days){
      x[t] ~ dbinom(size = 1, prob = p11*x[t-1] + (1-x[t-1])*p01)
      y[t] ~ dnorm(b0 + b1*x[t], sd = sigma1*x[t] + (1-x[t])*sigma0)
```

```
    }
})
constants <- list(days = days)
data <- list(y = sp500_pct_change$close_pct_change) # data is percent change in closing price
inits <- list(x = array(1, days), p01 = .5, p11 = .5, b0 = 0, b1 =0, sigma0 = 1, sigma1 = 1, p = .5)
Rmodel <- nimbleModel(code, constants, data, inits)
```

I then configure an MCMC for the model and run three chains for it, each of which samples 120,000 times and burns-in the first 20,000 samples. Below are the summary statistics of the top-level parameters.

Results

Parameter	Posterior mean	Posterior std. dev.	Bayesian 95% CI	Effective sample size	Gelman diagnost estimate (Upper
p_{01}	0.0111	0.00214	(0.0187,0.0394)	1070	1 (1.01)
p_{11}	0.972	0.00529	(0.961,0.981)	1365	1 (1.02)
p_{00}	0.989	0.00214	(0.984,0.993)	1070	NA (derived qua
p_{10}	0.0283	0.00529	(0.0188,0.0394)	1365	NA (derived qua
p	0.544	0.286	(0.0351,0.981)	3230	1 (1.00)
β_0	0.000659	0.000110	(0.000445,0.000875)	8047	1 (1.00)
β_1	-0.00162	0.000452	(-0.002511, -0.000743)	13712	1(1.00)
σ_0	0.00706	0.000116	(0.00683,0.01743)	406	1 (1.00)
σ_1	0.0182	0.000395	(0.00728,0.01896)	720	1 (1.00)

The mixing for each top level parameters looks very good, with the exceptions of β_0, β_1, both of which display similarly poor mixing when looked at on shorter intervals of samples. An example of the poor mixing can be found in the appendix. The values of the βs, are both statistically significant, but they are counter intuitive in their interpretation due to the results indicating that under bull markets, returns are on average less. The same is true for the σs. However, this is not too important, since the transition probabilities and the probability of a bull market are the main parameters of interest in this section, since they are the only essential part for use in the classification. However, all the parameters converge and have large effective sample sizes, except the σs are a lower, but still not too small. Also, since we do not use them in the remainder of the analysis for classification, we do not worry about them. The transition probabilities show that it is quite unlikely to transition from one state to another, but that about 97-98% of the time, the state will stay the same. The probability of a bull market is also slightly above 50%, but still quite close, which shows the market doesn't strongly favor one over the other.

Convergence

Now we asses the convergence of the MCMC chains using the Gelman diagnostic. For each of the variables, the point estimate of the Gelman diagnostic are 1, and the upper C.I's are 1 or very close to 1. This shows that the chains have converged.

Classification

Next, using the transition probabilities and the probability of being in state 1, we define a classification model to identify the state at each day in the original data from January 1, 1990 to December 31, 2015. The data is again the percentage change in the closing prices because this captures the volatility and changes from day to day, instead of the level of the S&P 500, which also includes other trends such as inflation, and is not mean-reverting, so you cannot identify bull and bear markets simply based on the level, because as seen in the early 1990's the level was substantially lower than from 2000 and onward. This will miss the spikes in volatility and price changes.

The likelihood of the observed is normally distributed around one of two means, μ_{bear}, μ_{bull}, with standard deviation $\sigma_{bear}, \sigma_{bull}$, corresponding to the average change of the S&P 500 and standard deviation under bear and bull markets, respectively. We assume these are different because by their definitions, returns should be positive in bull markets and negative in bear markets. Also, standard deviation (risk) is generally thought to be lower in bull markets, so we use different standard deviations for the two regimes. The transition probabilities are now no longer uniformly distributed, but distributed around the posterior means of estimated in the previous model, with standard deviation corresponding to their standard deviations. These are distributed as Beta distribution, but parameterized in mean and standard deviation. This is to account for the constraints that they are probabilities and cannot be greater than 1 or less than 0, or else they would be normally distributed around the estimated mean and standard deviations. The parameter $regime_t$ identifies the state the observed data, y_t belongs to. p_bull is the probability that $regime_t$ belonges to state 1, bull market.

Model

Top level parameters:

$$\mu_{bull} \sim Normal(0, sd = 10000),\ \ \mu_{bear} \sim Normal(0, sd = 10000)$$

$$\sigma_{bull} \sim Uniform(0, 10000), \sigma_{bear} \sim Uniform(0, 10000)$$

$$p_{01} \sim Beta(mean = \hat{p_{01}}, sd = \hat{\sigma}_{p_{01}}),\ \ p_{11} \sim Beta(mean = \hat{p_{11}}, sd = \hat{\sigma}_{p_{11}}),\ \ p_{00} = 1 - p01,\ \ p_{10} = 1 - p_{11}$$

$$p_{bull} \sim Beta(mean = \hat{p}, sd = \hat{\sigma}_p)$$

$$\sigma_{bull} \sim Uniform(0, 10000),\ \ \sigma_{bear} \sim Uniform(0, 10000)$$

Latent states:

$$regime_t \sim Bernoulli(p_bull) regime_t = \begin{cases} 1 \text{ if } y_t \text{ is in group 1} \\ 0 \text{ if } y_t \text{ is in group 0} \end{cases}$$

$$\mu_t \sim \begin{cases} \mu_{bull} \text{ if } regime_t \text{ equals 1} \\ \mu_{bear} \text{ if } regime_t \text{ equals 0} \end{cases}$$

$$\sigma_t \sim \begin{cases} \sigma_{bull} \text{ if } regime_t \text{ equals 1} \\ \sigma_{bear} \text{ if } regime_t \text{ equals 0} \end{cases}$$

Likelihood

$$y_t \sim Normal(\mu_t, sd = \sigma_t)$$

In the NIMBLE code, mu_t is means[t] and σ_t is sigmas[t].

```
code2 <- nimbleCode({
  #the priors for S&P500 prices
  mu_bull ~ dnorm(0, sd = 10000) #bull
  mu_bear ~ dnorm(0, sd = 10000) #bear market

  sigma_bull ~ dunif(0, 10000) #bull
```

```r
  sigma_bear ~ dunif(0, 10000) #bear

  #transition probabilities for going from 1 to 1, 0 to 1.
  p01 ~ dbeta(mean = p01_mean, sd = p01_sd)
  p11 ~ dbeta(mean = p11_mean, sd = p11_sd)
  prob_bull ~ dbeta(mean = p_mean, sd = p_sd)

  regime[1] ~ dbern(prob_bull)
  means[1] <- equals(z[1],0)*mu_bear + equals(z[1],1)*mu_bull
  sigmas[1] <- equals(z[1],0)*sigma_bear + equals(z[1],1)*sigma_bull
  y[1] ~ dnorm(means[1], sd = sigmas[1])

  for (i in 2:days){
    regime[i] ~ dbern(prob = (1-z[i-1])*p01 + z[i-1]*p11)

    means[i] <- equals(regime[i],0)*mu_bear + equals(regime[i],1)*mu_bull
    sigmas[i] <- equals(regime[i],0)*sigma_bear + equals(regime[i],1)*sigma_bull

    y[i] ~ dnorm(means[i], sd = sigmas[i])
  }
})
constants2 <- list(days = days, p01_mean= p01_mean, p11_mean= p11_mean, p01_sd= p01_sd, p11_sd= p11_sd,
data2 <- list(y = close_change, pcon1 = 1, pcon2 = 1)
inits2 <- list(regime = array(1,days), mu_bull = 0.01, mu_bear = 0, sigma_bull = 1, sigma_bear = 1, prol

Rmodel2 <- nimbleModel(code = code2, constants = constants2, data = data2, inits = inits2)
```

I then configure an MCMC for the model and run three chains for it, each of which samples 120,000 times and burns-in the first 20,000 samples.

Below are the summary statistics and plot of the classification. ###Results

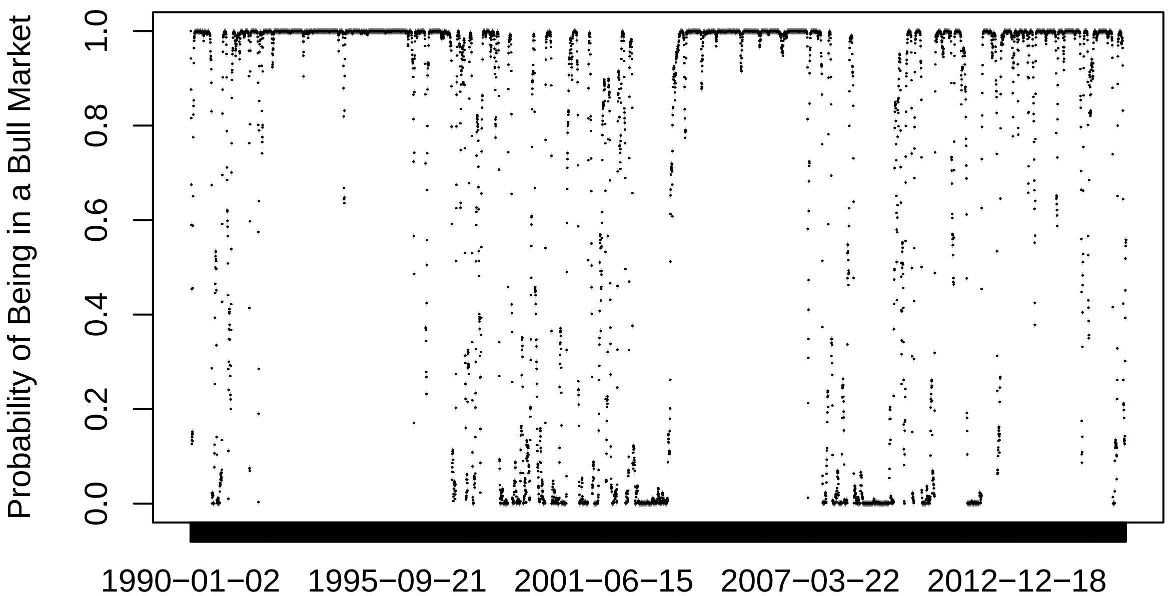
Parameter	Posterior mean	Posterior std. dev.	Bayesian 95% CI	Effective Sample Size	Gelman Diagnost est. (Upper C.I.)
p_{01}	0.0171	0.00227	(0.0130, 0.0219)	3597	1 (1.00)
p_{11}	0.987	0.00181	(0.983, 0.990)	2249	1 (1.00)
p_{00}	0.983	0.00227	(0.978,0.987)	3597	(NA, since derive
p_{10}	0.0135	0.0187	(0.0101,0.0172)	2249	(NA, since derive
p_{bull}	0.545	0.2844	(0.0376, 0.9835)	19238	1 (1.00)
μ_{bull}	.000656	0.000109	(0.000444, 0.000871)	9159	1 (1.00)
μ_{bear}	-0.000917	0.000426	(-0.00176,-0.0000878)	19086	1 (1.00)
σ_{bull}	0.00703	0.000116	(0.0173, 0.0188)	380	1 (1.01)
σ_{bear}	0.0180	0.000392	(0.00680, 0.00725)	637	1 (1.01)

We see that the transition probabilities remain relatively stable as from before, which is to be expected due to the restrictive priors enforced on them, and also due to the low standard error of them from the estimation in the state-space model. The effective sizes are all sufficiently large, with the exception of σ_{bull}, but it is not terribly small and it still converges as shown in the Gelman diagnostic point estimate. The rest of the parameters also converge very well since the Gelman diagnostics are all 1 with upper C.I.'s very close to 1. Now, interestingly we see that the average percent change in the S&P 500 returns slightly greater than 0 and is statistically significant since 0 isn't in its credible interval, which means that the returns of the market are on average positive under bull markets. Under bull markets, we see the average return is negative and also statistically significant, showing that under bear markets, returns are on average negative. Another pair of interesting statistics from the results are the standard deviations of returns under each market regime. We see both are statistically significant, and that the standard deviation under bull markets is almost two thirds less than the standard deviation of returns under bear markets. Since standard deviation of stock returns is usually associated with risk, this shows that under bull markets, risk is lower than in bear markets. These results are consistent with conventional thought and definitions of the characterisitcs of bull and bear markets.

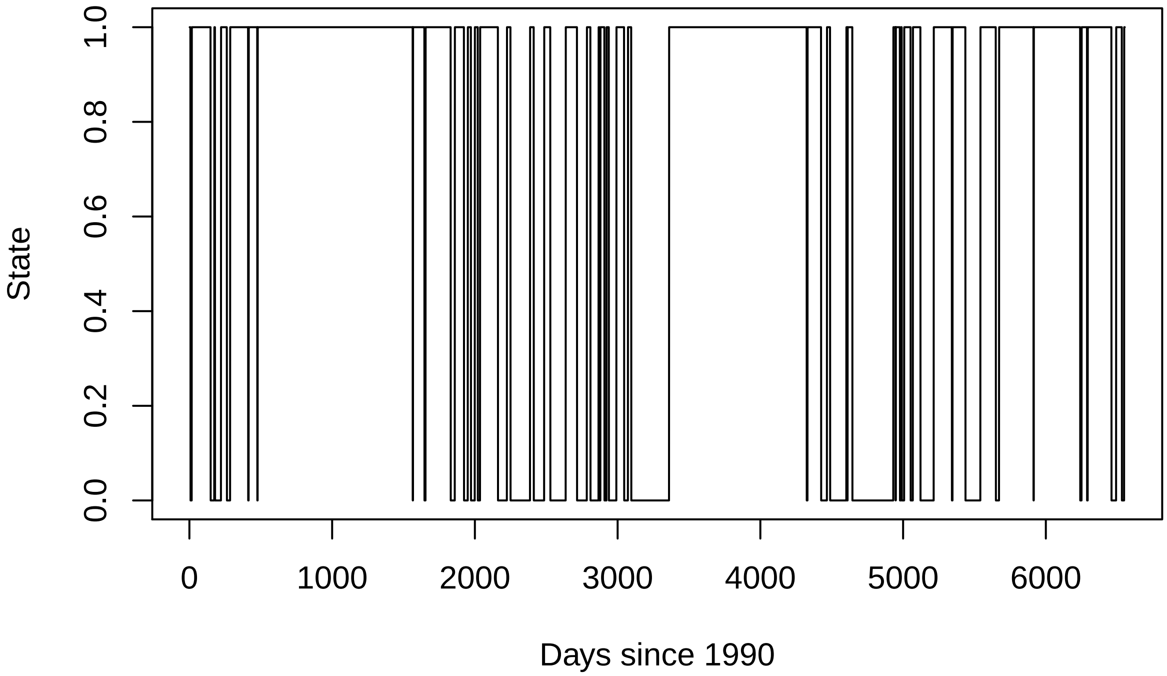
Convergence

Now we asses the convergence of the MCMC chains using the Gelman diagnostic. For each of the variables, the point estimate of the Gelman diagnostic are 1, and the upper C.I's are 1 or very close to 1. This shows that the chains have converged.

```
## Warning in rm(codaSamples): object 'codaSamples' not found
```




```
## Warning in plot.xy(xy, type, ...): plot type 'line' will be truncated to
## first character
```



Above is a graph with the probabilities of being classified as a bull market. The results of the classification show us that the market since 1990 has been a bull market for slightly greater than half of the period. There are two substantial periods that bounce around in probabilities and are most likely classified as bear markets, around 1999-2002 and around 2007. This is very good for our results because it is consistent with the historical occurrences of the 2000 stock market crash due to the Dot Com bubble and the 2008 financial crisis. The shorter periods of bearish markets are around the end of 1990 and latter half of 2015, as shown in the dipping probabilities in the graph. This is also consistent with results from other sources, such as Bloomberg[2], Yardeni Research [3], and others. Below is a table from Yardeni Research, Inc. classifying bear markets and corrections (which are essentially less severe bear markets) since 1990. We can assess the performance of our classification by comparing the results and it looks fairly consistent, showing the classification performed quite well.

7/16/1990	10/11/1990	368.95	295.46	-19.9	87
7/17/1998	8/31/1998	1186.75	957.28	-19.3	45
3/24/2000	10/9/2002	1527.46	776.76	-49.1	929
11/27/2002	3/11/2003	938.87	800.73	-14.7	104
10/9/2007	3/9/2009	1565.15	676.53	-56.8	517
4/23/2010	7/2/2010	1217.28	1022.58	-16.0	70
4/29/2011	10/3/2011	1363.61	1099.23	-19.4	157
5/21/2015	8/25/2015	2130.82	1867.61	-12.4	96
11/3/2015	2/11/2016	2109.79	1829.08	-13.3	100

Figure 1: Bear Markets and Corrections (https://www.yardeni.com/pub/sp500corrbeartables.pdf)

Conclusion

Using the transition probabilities from the state-space model to classify the states as bear or bull markets appears to be approximately consistent with past research. From the state-space model, we attained estimates of the probabilities of moving from one state to another and saw that it is quite unlikely to change regimes, given the respective probabilities of going from bull to bear and bear to bull are about 1% and 3%. The MCMC chains all converged nicely and the parameters we used mixed well for the most part. The parameters also had high effective sample sizes. The model could be extended to include states representing stagnant markets, or bull and bear markets with high or low volatility.

Reference

1. Marino, Matt. 19 October 2015. Stock Market Basics: What Are Bear and Bull Markets?. http://www.nasdaq.com/article/stock-market-basics-what-are-bear-and-bull-markets-cm531696

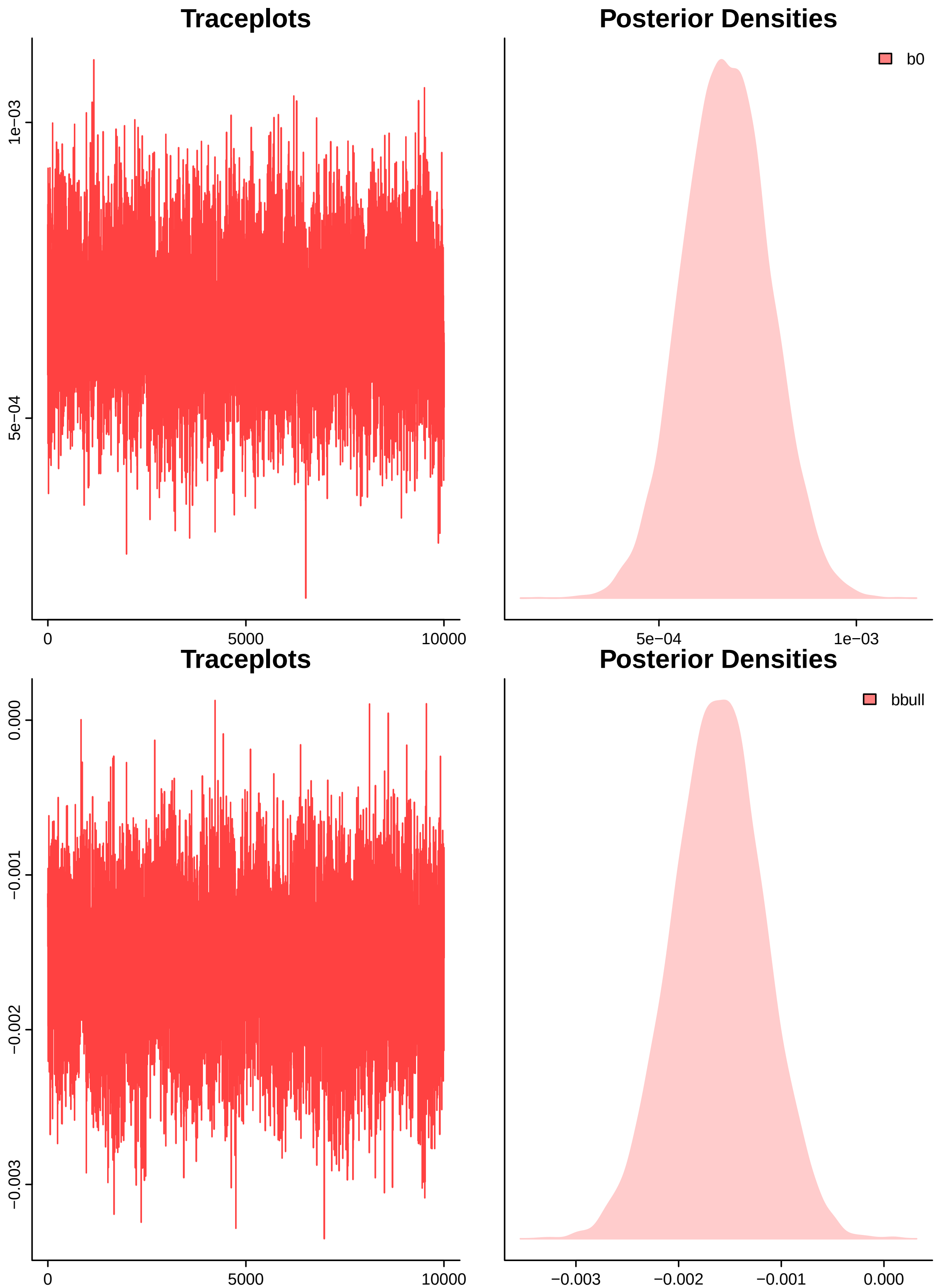
2. Verhage, Julie. 2016. Morgan Stanley Analyzed 43 Bear Markets and Here's What It Found. https://www.bloomberg.com/news/articles/2016-01-26/morgan-stanley-analyzed-43-bear-markets-and-here-s-what-it-found

3. Yardeni E., et al. 5 November 2016. Stock Market Briefing:S&P 500 Bear & Bull Market Tables. https://www.yardeni.com/pub/sp500corrbeartables.pdf

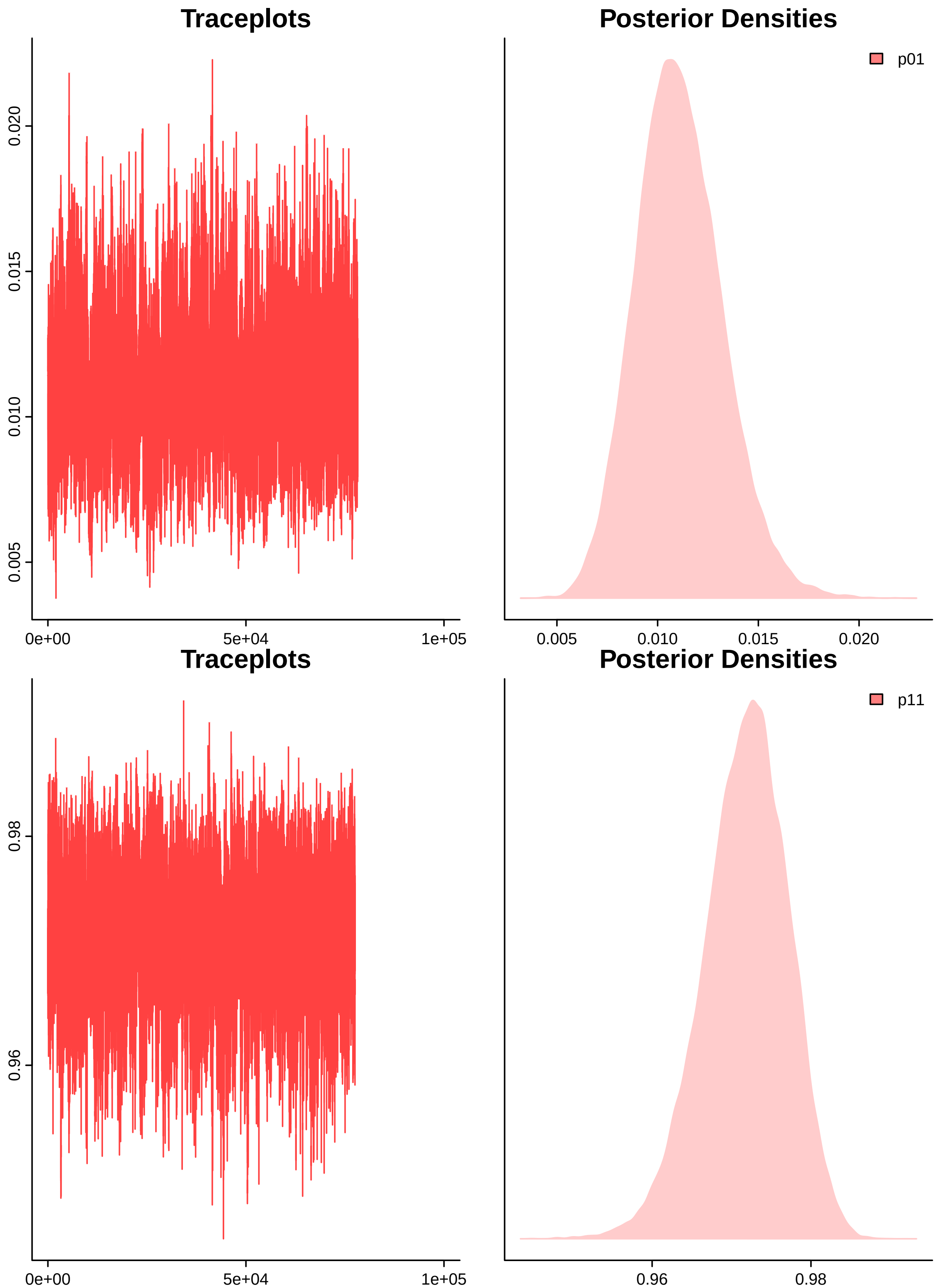
4. https://finance.yahoo.com/quote/%5EGSPC/history?period1=631170000&period2=1451538000&interval=1d&filter=history&frequency=1d

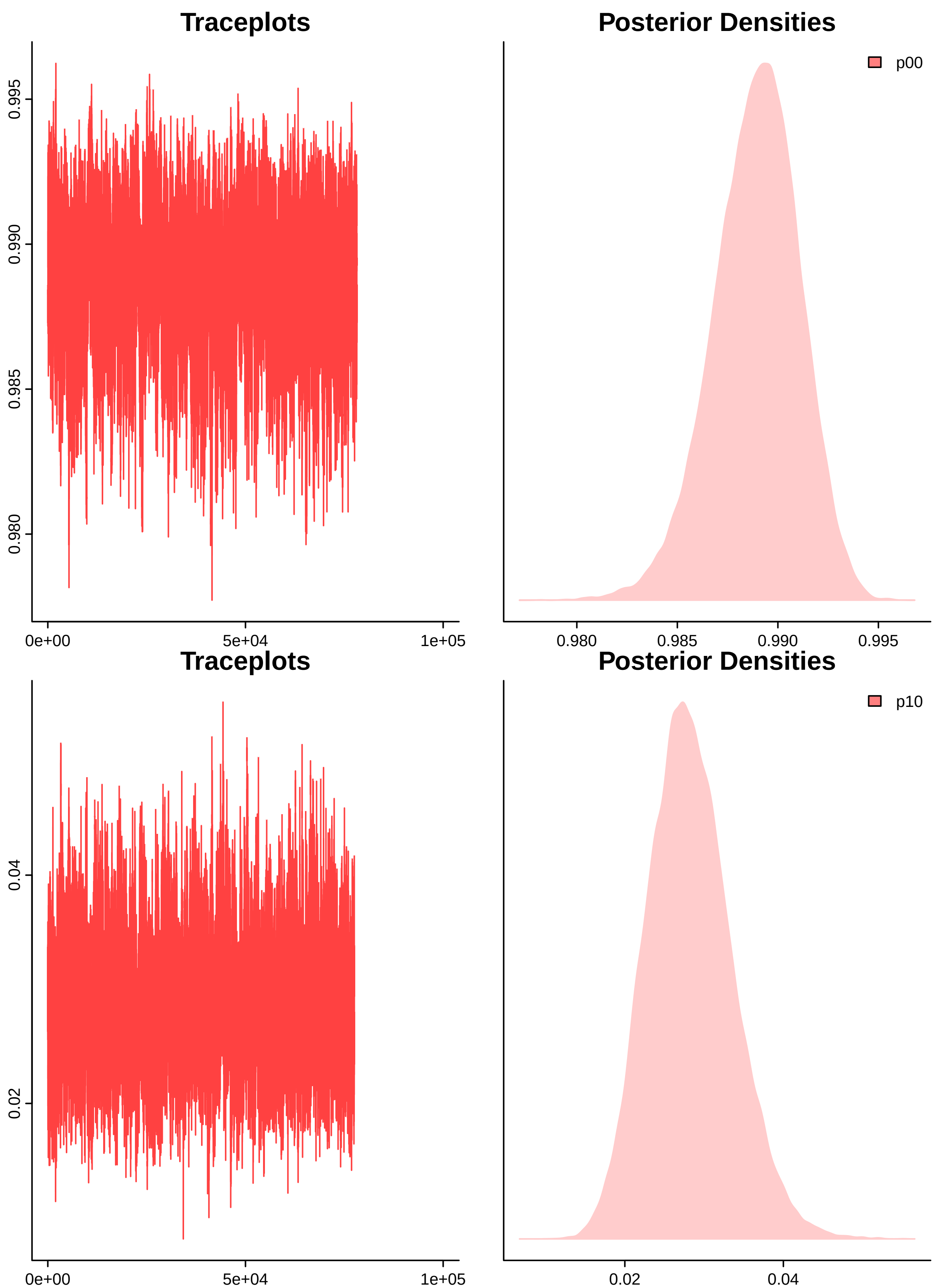
Appendix

Bad mixing of samples plots for βs in state-space model

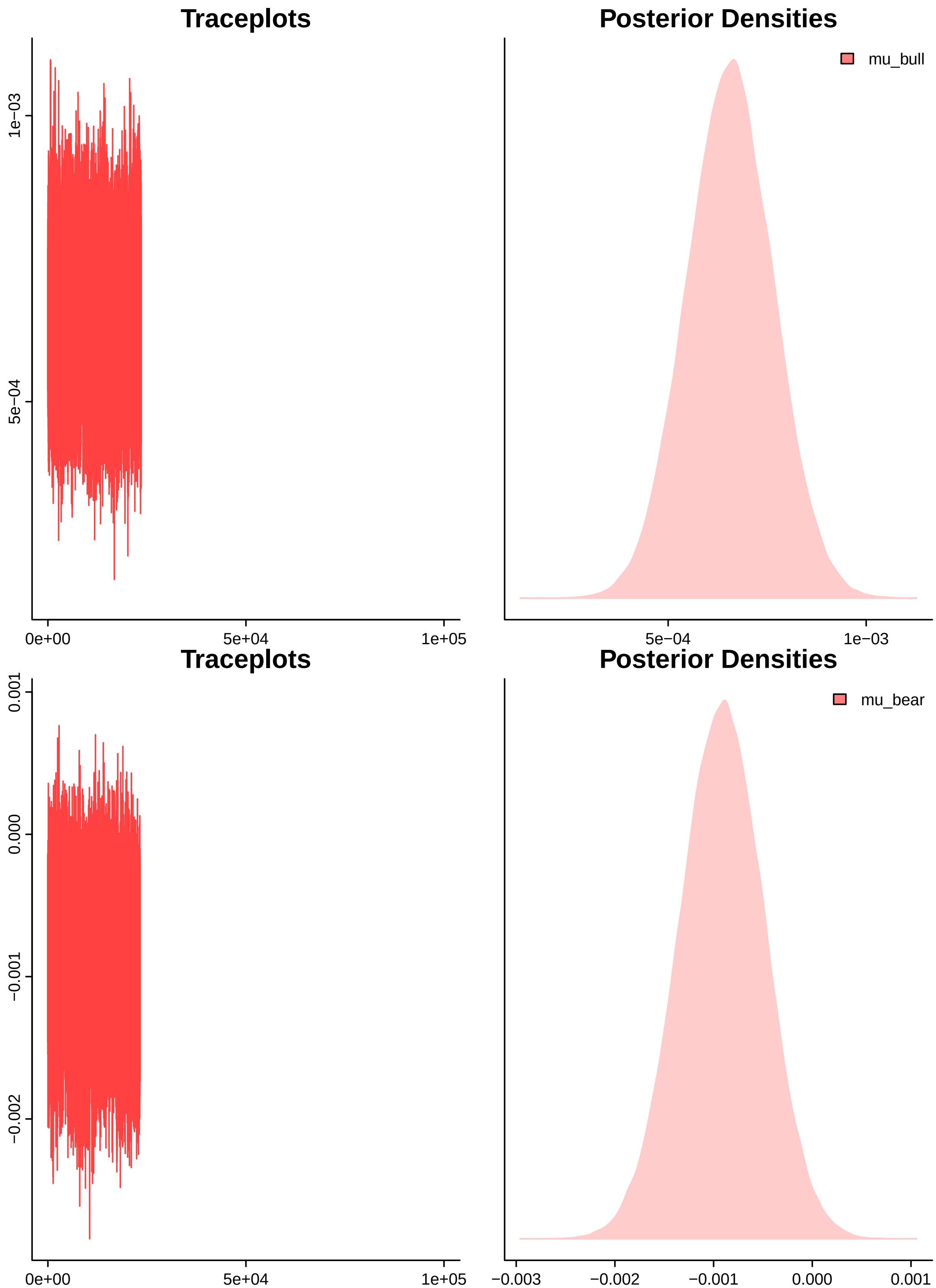

Traceplots
Posterior Densities
b0
1e−03
5e−04
0
5000
10000
5e−04
1e−03
Traceplots
Posterior Densities
bbull
0.000
−0.001
−0.002
−0.003
0
5000
10000
−0.003
−0.002
−0.001
0.000

Posterior distributions of the transitions probabilities to state 1. The other two are derived, so not included.

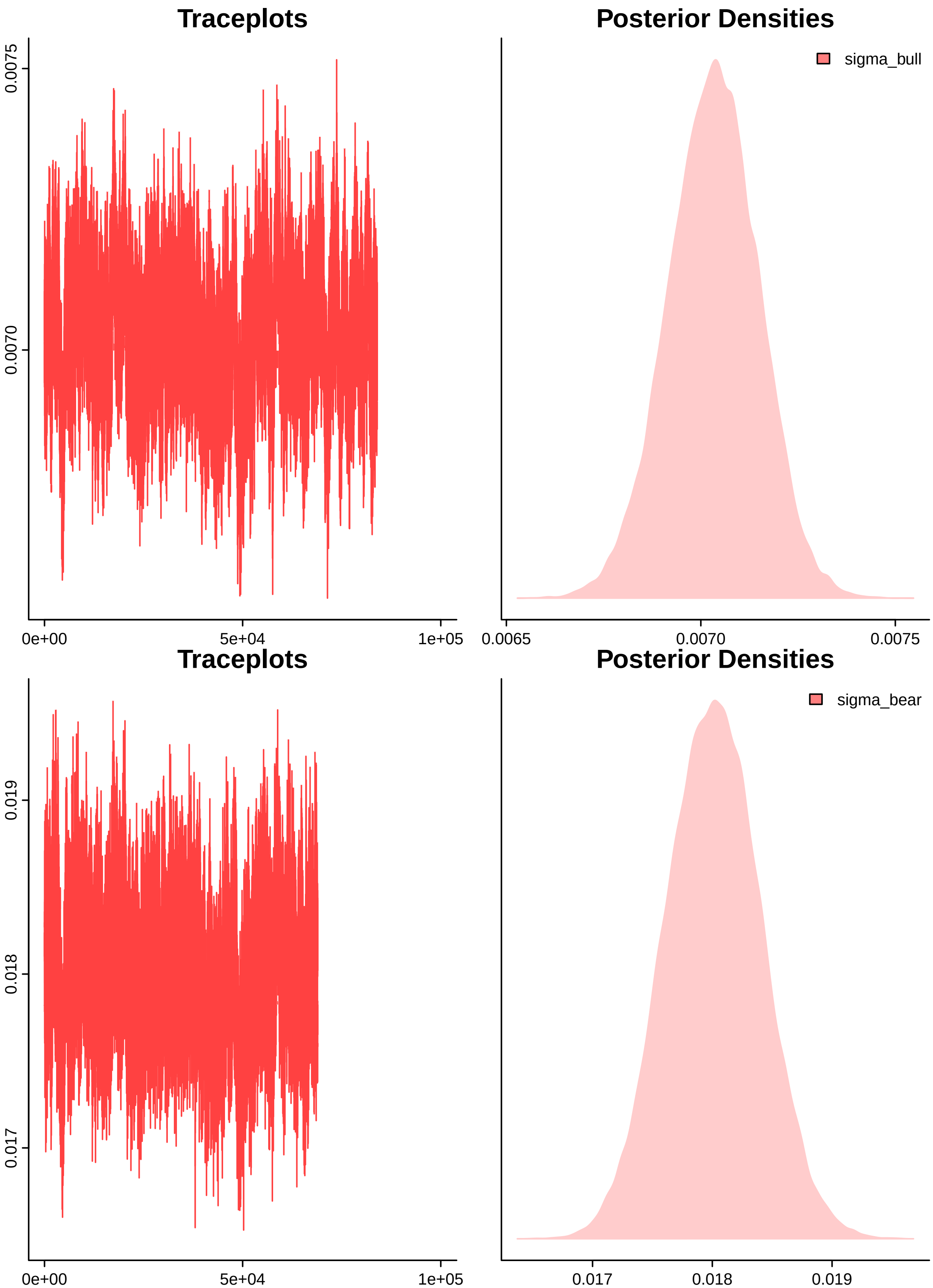

Traceplots
Posterior Densities
p01
0.005
0.010
0.015
0.020
0e+00
5e+04
1e+05
Traceplots
Posterior Densities
p11
0.96
0.98
0e+00
5e+04
1e+05


Traceplots
Posterior Densities
p00
0.995
0.990
0.985
0.980
0e+00
5e+04
1e+05
0.980
0.985
0.990
0.995
Traceplots
Posterior Densities
p10
0.04
0.02
0e+00
5e+04
1e+05
0.02
0.04

Posterior plots of the average returns of the S&P 500 under bull and bear markets


Traceplots
Posterior Densities
mu_bull
1e−03
5e−04
0e+00
5e+04
1e+05
5e−04
1e−03
Traceplots
Posterior Densities
mu_bear
0.001
0.000
−0.001
−0.002
0e+00
5e+04
1e+05
−0.003
−0.002
−0.001
0.000
0.001

Posterior plots of the average returns of the S&P 500 under bull and bear markets


Traceplots
Posterior Densities
sigma_bull
0.0075
0.0070
0e+00
5e+04
1e+05
0.0065
0.0070
0.0075
Traceplots
Posterior Densities
sigma_bear
0.019
0.018
0.017
0e+00
5e+04
1e+05
0.017
0.018
0.019

Gelman diagnostic for state-space model output:

```
## Potential scale reduction factors:
##
##            Point est. Upper C.I.
## mu_bear             1       1.00
## mu_bull             1       1.00
## p01                 1       1.00
## p11                 1       1.00
## prob_bull           1       1.00
## sigma_bear          1       1.01
## sigma_bull          1       1.01
##
## Multivariate psrf
##
## 1
```

Gelman diagnostic for classification model output:

```
## Potential scale reduction factors:
##
##            Point est. Upper C.I.
## mu_bear             1       1.00
## mu_bull             1       1.00
## p01                 1       1.00
## p11                 1       1.00
## prob_bull           1       1.00
## sigma_bear          1       1.01
## sigma_bull          1       1.01
##
## Multivariate psrf
##
## 1
```