Bayesian GBTM Model Selection

In this vignette we will demonstrate how to use BayesTraj to use Bayesian Model Averaging to estimate a Bayesian GBTM with a normal likelihood. We will use simulated data in order to verify that the estimation routine can select the correct functional forms, recover the true parameters, and to demonstrate how the data should be formatted before calling the estimation routines.

Begin by loading the BayesTraj library:

```
library(BayesTraj)
```

Simulating Data

First, we will simulate data. This will not be necessary in your own projects, but it is useful both for testing the package and for using as a template for formatting your own datasets.

```
N=1000 #number of units
T=9 #time periods
pi=c(0.5,0.2,0.3) #group membership probabilities
K = length(pi) #number of groups
#coefficients
beta=matrix(c(110,5,-0.5,
              111,-2,0,
              118,0,0),nrow=3,ncol=3,byrow=TRUE)
sigma=2 #standard deviation of outcomes

set.seed(1)
data = gen_data(N=N,
                T=T,
                pi=pi,
                beta=beta,
                sigma=sigma,
                poly = 2 #degree of polynomial
                )
```

In this example we have simulated data for 1000 units with 9 time periods each, for a total of 9000 observations. We have chosen the group-membership probabilities to be 50%, 20%, and 30%. From this, the `gen_data` function can infer that there should be three groups.

Each row of the `beta` matrix defines the trajectory coefficients. For example, the expected value at time t in Group 1 is $110 + 5t - 0.5t^2$. Sigma, defines the standard deviation of the outcomes.

When calling the `gen_data` function, we also specify `poly=2` in order to tell model to use a second-degree polynomial for time. If there are more non-intercept columns of `beta` than `poly`, `gen_data` will generate random covariates corresponding to the remaining columns. In general, the last `poly` columns of the `beta` matrix correspond to the polynomial coefficients.

Please note that we have selected `beta` coefficients corresponding to a second-degree polynomial in Group 1, a first-degree polynomial in Group 2, and a 0-degree (constant) polynomial in Group 3.

Now let's take a look at the generated data. We can unpack the individual attributes from the data object.

```
X=data$X
y=data$Y
```

The first 18 rows of `X` are:

```
print(head(X,18))
#>       [,1] [,2] [,3]
#>  [1,]    1    1    1
#>  [2,]    1    2    4
#>  [3,]    1    3    9
#>  [4,]    1    4   16
#>  [5,]    1    5   25
#>  [6,]    1    6   36
#>  [7,]    1    7   49
#>  [8,]    1    8   64
#>  [9,]    1    9   81
#> [10,]    2    1    1
#> [11,]    2    2    4
#> [12,]    2    3    9
#> [13,]    2    4   16
#> [14,]    2    5   25
#> [15,]    2    6   36
#> [16,]    2    7   49
#> [17,]    2    8   64
#> [18,]    2    9   81
```

The first column identifies the unit. For example, the first 9 rows correspond to unit 1, the second 9 rows correspond to unit 2, and so forth. The second column is the time variable. Rows 1 and 10 correspond to time 1, rows 2 and 11 correspond to time 2, and so forth. Similarly, the third column is the square of the time column.

Now we take a look at `y`. These are the outcomes. `y[1]` corresponds to the outcome for unit 1 at time 1. `y[2]` corresponds to the outcome for unit 1 at time 2, and so forth. The values of `y` must correspond with the rows of `X`. Therefore `X` and `y` should have the same length.

```
print(head(y,18))
#>  [1] 114.6093 117.5802 118.8266 122.0160 123.9023 124.2542 118.5587
#>  [8] 117.6470 116.1397 112.9243 114.4242 119.1764 120.6321 122.5672
#> [15] 121.4290 120.8274 117.4027 115.0291
```

Estimating the model

We now turn our attention toward estimating the model. We can do this by calling the `trajMS` function. This function uses the following weakly-informative hyperparameters: a uniform prior on group membership probabilities, Jeffery's prior on the variances, and a unit-information g-prior on the regression coefficients.

```
iter = 5000
thin = 1
model = trajMS(X=X, #data matrix
               y=y, #outcomes
               K=K, #number of groups
               time_index=2, #column of X corresponding to time
               iterations=iter, #number of iterations
               thin=thin, #thinning
               displter=1000) #Print a message every 1000 iterations
#> [1] 1000
```

```
#> [1] 2000
#> [1] 3000
#> [1] 4000
#> [1] 5000
```

First, let's clarify the model specification. `trajMS` will sample the polynomial degree in the MCMC samples, rather than independently choose whether or not to include each covariate. For example, if 2 is sampled in an MCMC iteration, both the main-effect and the squared coefficients will be sampled. If 1 is selected, only the main effect will be sampled. If 0 is selected, neither the main effect nor the squared term will be sampled (only the intercept will remain). Users wishing to average over more complicated functions than polynomials, or to estimate models in which high order polynomials can be included even if a lower polynomial was selected out, will need make corresponding edits to the `trajMS` function.

Here we run the model for 5000 MCMC iterations. In practice, more iterations may be desirable to ensure the posterior results are valid. Setting the `thin` parameter to 1 tells us to keep every sample. We can set `thin=10`, for example, to only keep 1 out of every 10 samples. Thinning is not necessary unless your computer has memory limitations. We also set `dispIter=1000` to tell the program to send us a message every 1000 MCMC iterations. This will help us monitor the progress of the program.

The only argument we have not touched on yet is `time_index`. This parameter specified which column of `X` corresponds to the time variable. If the data does not contain any covariates, this should be the second column of `X`. If, for example, we were using a dataset with additional covariates in columns 2 and 3, time in column 4, and time-squared in column 5, we would set `time_index=4`.

Analyzing the Model

The model object contains the MCMC samples for each of the model's parameters. We can access the MCMC samples as follows, where each row represent an iteration of the MCMC:

```
head(model$beta[[1]]) #group 1's coefficients
#>          [,1]     [,2]       [,3]
#> [1,] 111.5110 2.550531 -0.2769183
#> [2,] 112.7637 3.302669 -0.3356828
#> [3,] 112.7638 3.329113 -0.3356322
#> [4,] 111.5664 4.044737 -0.4068073
#> [5,] 110.2673 4.924220 -0.4957123
#> [6,] 110.1238 4.967789 -0.4992188
head(model$beta[[2]]) #group 2's coefficients
#>          [,1]      [,2]        [,3]
#> [1,] 111.7733 2.5765635 -0.28952430
#> [2,] 112.0807 3.5241792 -0.35236787
#> [3,] 112.9858 3.1548973 -0.31676085
#> [4,] 114.0404 2.4470950 -0.24411107
#> [5,] 116.9405 0.6147143 -0.06050324
#> [6,] 117.9631 0.0000000  0.00000000
head(model$beta[[3]]) #group 3's coefficients
#>          [,1]      [,2]       [,3]
#> [1,] 112.2581  2.200616 -0.2569627
#> [2,] 111.7156  1.453005 -0.2222034
#> [3,] 110.8080 -1.964840  0.0000000
#> [4,] 110.8511 -1.967305  0.0000000
#> [5,] 110.7582 -1.953631  0.0000000
#> [6,] 110.8729 -1.969829  0.0000000
head(model$sigma) #variance - NOT THE STANDARD DEVIATION
#>           [,1]      [,2]      [,3]
```

```
#> [1,] 45.625533 54.686719 64.041952
#> [2,]  5.695543  7.282856 80.670000
#> [3,]  3.777737  4.699675  2.006411
#> [4,]  3.242428  4.607336  2.054425
#> [5,]  2.082802  3.189741  2.161093
#> [6,]  2.066434  1.965535  2.113664
model$c[1:6,1:10] #unit-level group memberships
#>      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
#> [1,]    3    3    3    3    2    3    3    3    1     3
#> [2,]    2    2    3    3    3    3    3    3    2     1
#> [3,]    2    1    2    3    2    3    3    2    1     2
#> [4,]    1    2    1    3    2    3    3    2    2     1
#> [5,]    1    1    2    3    1    3    3    2    2     1
#> [6,]    1    1    2    3    1    3    3    2    2     1
head(model$pi) #group-membership probabilities
#>           [,1]      [,2]      [,3]
#> [1,] 0.1025598 0.2397216 0.6577186
#> [2,] 0.1812220 0.2949648 0.5238132
#> [3,] 0.3912114 0.4272833 0.1815053
#> [4,] 0.4031879 0.4139962 0.1828159
#> [5,] 0.4999599 0.3159259 0.1841142
#> [6,] 0.5086566 0.2948827 0.1964607
```

A conveniant way to summarize the posterior is with the `summary_single_MS` function:

```
burn = 0.9
summary = summary_single_MS(model,X,y,burn)
```

The `burn` parameter specifies the fraction of draws to keep. In this example, we keep the last 90% of MCMC samples. The first 10% are discarded as the burn-in period.

We can now print out a posterior summary to obtain the posterior mean, standard deviation, and 95% credible interval, and parameter inclusion probabilities as follows:

```
print(summary$estimates)
#>                Estimate Standard Deviation       2.5%         50%
#> beta_1[1]  1.100884e+02       0.0819803143 109.929292 110.0892637
#> beta_1[2]  4.986598e+00       0.0381446263   4.914285   4.9864336
#> beta_1[3] -5.008085e-01       0.0037272377  -0.507893  -0.5007788
#> sigma_1    1.442101e+00       0.0150261150   1.413178   1.4421379
#> beta_2[1]  1.179858e+02       0.0306771467 117.928052 117.9866335
#> beta_2[2]  2.647580e-04       0.0044192545   0.000000   0.0000000
#> beta_2[3] -1.063148e-05       0.0003589971   0.000000   0.0000000
#> sigma_2    1.403376e+00       0.0198774368   1.365231   1.4032296
#> beta_3[1]  1.108209e+02       0.0772629260 110.668565 110.8190970
#> beta_3[2] -1.970339e+00       0.0167806917  -1.998095  -1.9697196
#> beta_3[3]  7.608147e-05       0.0010268160   0.000000   0.0000000
#> sigma_3    1.456561e+00       0.0241136670   1.410595   1.4563111
#> pi[1]      5.192242e+01       1.5638733254  48.793494  51.9140058
#> pi[2]      2.824211e+01       1.4121998816  25.461145  28.2454984
#> pi[3]      1.983547e+01       1.2574759141  17.398846  19.8095623
#>                97.5% Inclusion Prob.
#> beta_1[1] 110.247101          1.0000
#> beta_1[2]   5.059099          1.0000
#> beta_1[3]  -0.493701          1.0000
```

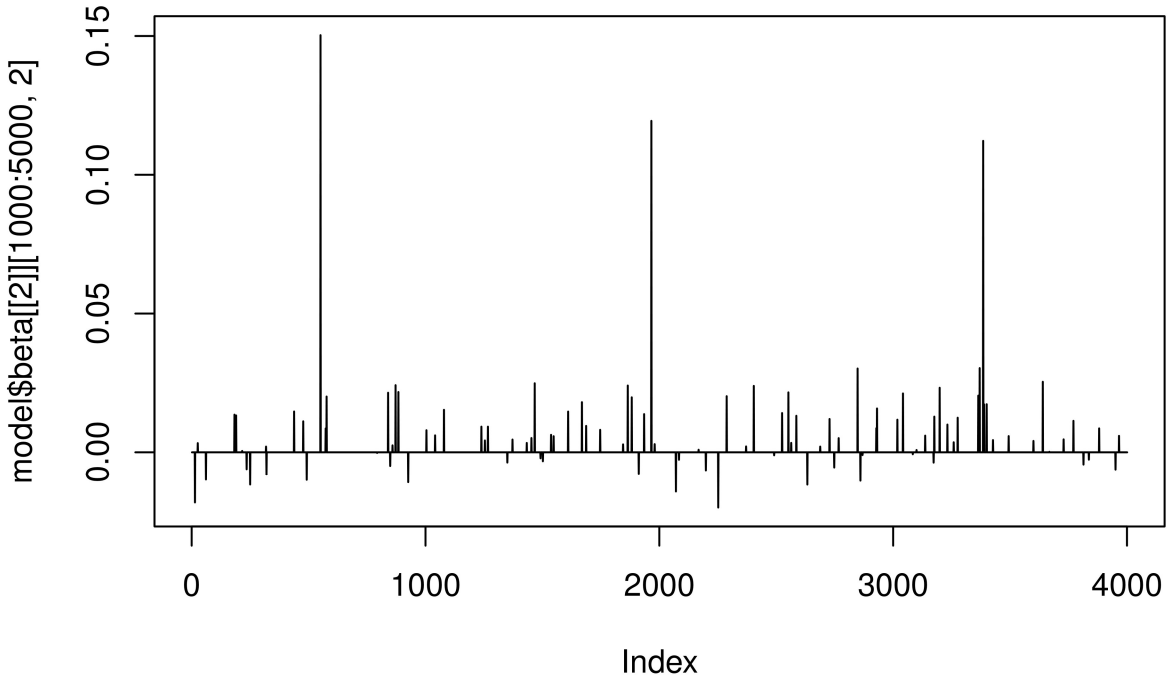
```
#> sigma_1     1.471797              NA
#> beta_2[1] 118.042765          1.0000
#> beta_2[2]   0.000000          0.0256
#> beta_2[3]   0.000000          0.0022
#> sigma_2     1.442843              NA
#> beta_3[1] 110.972779          1.0000
#> beta_3[2]  -1.943187          1.0000
#> beta_3[3]   0.000000          0.0234
#> sigma_3     1.504750              NA
#> pi[1]      55.001249              NA
#> pi[2]      30.989493              NA
#> pi[3]      22.398426              NA
```

The inclusion probability corresponds to the proportion of non-zero posterior samples. In our simulated dataset, the posterior correctly selects out the zero coefficients from the model. However, real datasets are unlikely to be generated from such simple functional forms. As a result, the inclusion probabilities may not be so close to 0 or 1 in practice. This is fine - it simply means the model is incorporating uncertainty in the optimal functional form.

MCMC samples for variables selected out of the model.

The MCMC samples can be plotted to see how the variable selection works. In the example below, we see that all but about 2.5% of the draws for the time main-effect in Group 2 are set ot zero, effectively selecting this parameter out of the model.

```
plot(model$beta[[2]][1000:5000,2],type='l')
```



Checking for Label Switching and Local Modes

One issue with GBTMs is the tendency for estimation routines to find a local mode which is not globally optimal. This is a problem for GBTMs estimated using maximum likelihood as well. To increase the probability that we are in a global optimum rather than a local optimum, we often run the Gibbs sampler using several seeds and print out the likelihood at the posterior mean:

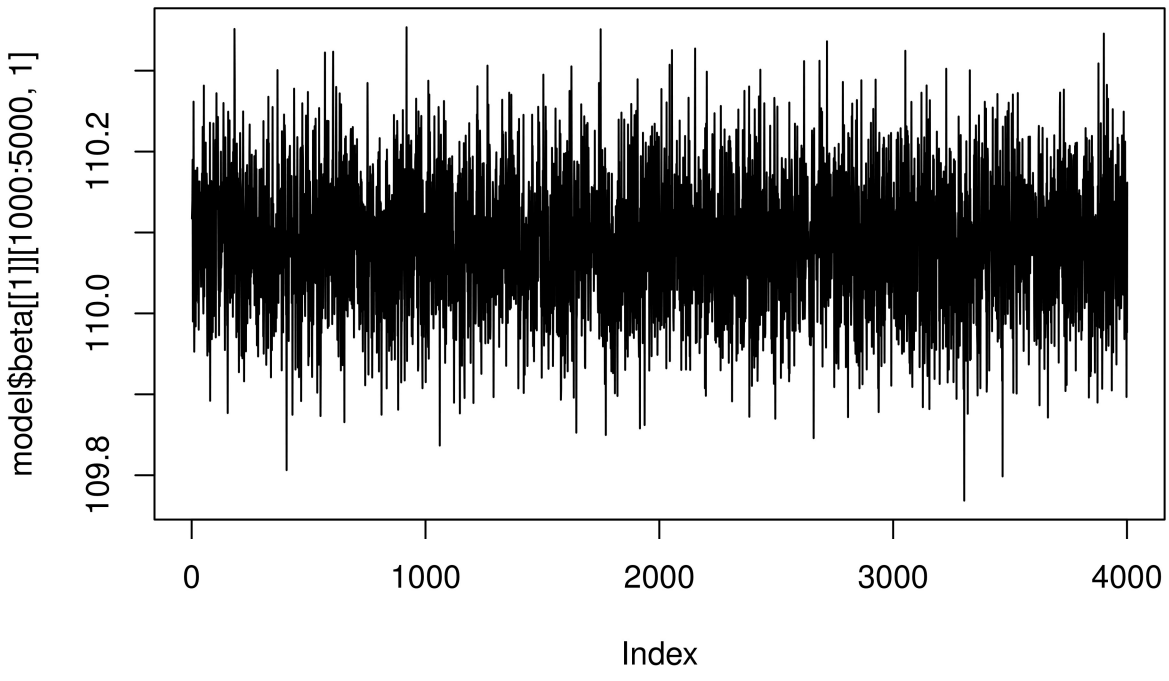
```
print(summary$log.likelihood)
#> [1] -14865.28
```

We then use the seed which maximize the likelihood. This solution has no optimality guarantees, but we have found that we can often reach better optimas this way than other existing packages using maximum likelihood.

The main drawback of the Bayesian approach is the tendency for label-switching and mode-switching. In the label-switching problem, the group labels switch in the middle of the algorithm. As a consequence, the group labeled "1" for the first 1000 draws may be labeled "2" in the second 1000 draws and vice versa. This would render and posterior summary of these coefficients meaningless. In our experience, label switching has not been a problem. However, switching between local-modes during the sampling process has occasionally been an issue.

There is no consensus for the best way to deal with label and mode switching. Either problem can be easily observed by plotting the draws sequentially and checking for sudden and sustained breaks in the trend. For example, the plot below looks consistent throughout the post-burn-in samples:

```
plot(model$beta[[1]][1000:5000,1],type='l')
```



This indicates that neither label-switching nor mode-siwtching occured.

If we do observe a sudden break, there are multiple possible solutions. From our experience, we usually find

that re-estimating the model using a different seed will solve the problem with least amount of effort.