# MTS Preliminary Simulation Results

*Daniel M. Swan*

*Monday, March 16, 2015*

This document provides a summary of the preliminary simulation of the partial interval recording (PIR) estimator using the penalized log likelihood function. I will summarize the results of the simulation by characterizing:

- The Monte Carlo error for phi and zeta in both their natural parameterizations and the logit of phi and log of zeta
- The error in the estimates of logit phi in terms of the median and maximum bias and RMSE
- The error in the estimates of log zeta in terms of the median and maximum bias and RMSE

Initial examination of plots of the bias of the parameters indicates that intervals of K $<$ 90 are insufficient for unbiased estimates even in a restricted range of the parameter space of phi and zeta.

## Monte Carlo error

```
##        stat       90      120
## 1 logit phi 1.49e-03 1.37e-03
## 2  log zeta 1.47e-04 1.40e-04
## 3       phi 5.47e-06 4.19e-06
## 4      zeta 2.11e-05 2.03e-05
```

The maximum Monte Carlo error is quite small in all cases.

# Logit phi results

## Median bias

```
##   K_intervals k_priors          5         10         20         40
## 1          90    1.001 0.06754532 0.07022720 0.10331763 0.11835742
## 2          90    1.010 0.06248269 0.06805353 0.09027454 0.12251406
## 3          90    1.050 0.05757788 0.06007610 0.09553825 0.11775306
## 4          90    1.100 0.05201071 0.05349718 0.08694612 0.10171494
## 5         120    1.001 0.05337032 0.04548150 0.06371120 0.08602858
## 6         120    1.010 0.04680430 0.04824247 0.06447111 0.08227274
## 7         120    1.050 0.04723531 0.04432245 0.06527968 0.08393933
## 8         120    1.100 0.04326989 0.04311350 0.06373274 0.07546744
```

The median absolute bias for logit phi suggests that in general, increasing the value of shape parameter decreases the bias and increasing the value of the scale parameter increases the bias. Higher values of K are also better.

## Maximum bias

```
##   K_intervals k_priors        5       10       20       40
## 1          90    1.001 1.584654 2.208144 2.812872 3.412707
## 2          90    1.010 1.341529 1.945813 2.523420 3.104500
## 3          90    1.050 1.200360 1.737153 2.286724 2.815735
## 4          90    1.100 1.172079 1.673529 2.174533 2.682276
## 5         120    1.001 1.589124 2.195453 2.803611 3.394343
## 6         120    1.010 1.369206 1.969105 2.545275 3.119739
## 7         120    1.050 1.181515 1.772641 2.320788 2.852123
## 8         120    1.100 1.145363 1.693329 2.233270 2.737632
```

The values for maximum absolute bias are all quite large, more than 100% maximum absolute bias in all cases. The same general relationship between shape, scale, number of intervals, and the magnitude of the absolute bias is observed here as above.

## Median RMSE

```
##   K_intervals k_priors         5        10        20        40
## 1          90    1.001 0.6325564 0.6688699 0.7126661 0.7307377
## 2          90    1.010 0.6031233 0.6379347 0.6506289 0.6619241
## 3          90    1.050 0.5335489 0.5709688 0.6042273 0.6087262
## 4          90    1.100 0.5059727 0.5569598 0.5812868 0.5971107
## 5         120    1.001 0.5245830 0.5634733 0.5846706 0.5942339
## 6         120    1.010 0.5097454 0.5283602 0.5369372 0.5505067
## 7         120    1.050 0.4790010 0.4889398 0.5069148 0.5275748
## 8         120    1.100 0.4541635 0.4802856 0.4912206 0.5006791
```

The relationship between RMSE and shape, scale, and number of intervals is the same as above. Increasing K and increasing the shape parameter decreases the median RMSE. Increasing the value of the shape parameter increases median RMSE.

## Maximum RMSE

```
##   K_intervals k_priors        5       10       20       40
## 1          90    1.001 3.014163 2.934679 2.858386 3.457570
## 2          90    1.010 1.943850 2.040325 2.561093 3.139176
## 3          90    1.050 1.360037 1.872509 2.372828 2.845138
## 4          90    1.100 1.308144 1.799893 2.283612 2.725269
## 5         120    1.001 2.873893 2.731987 2.857155 3.450488
## 6         120    1.010 1.846194 2.009339 2.586808 3.162044
## 7         120    1.050 1.323301 1.832545 2.354411 2.887743
## 8         120    1.100 1.284240 1.773602 2.260011 2.766603
```

The same relationship is observed here except that when scale = 1.001, the lowest values of the maximum RMSE occur at shape = 10 or 20, rather than 5.

Take together, this suggests that "good" estimates might be found when K = 120, shape >= 1.01, and scale <= 10.

# Log zeta resuts

## Median bias

```
##    K_intervals k_priors          5         10         20         40
## 9           90    1.001 0.05959162 0.06744245 0.07865529 0.11186660
## 10          90    1.010 0.06327601 0.06987241 0.08650544 0.11286766
## 11          90    1.050 0.07008267 0.08483202 0.09365256 0.12432170
## 12          90    1.100 0.07635785 0.09246211 0.10633305 0.13942413
## 13         120    1.001 0.04610412 0.05272184 0.05903420 0.08375811
## 14         120    1.010 0.04521209 0.05352984 0.06200197 0.08354188
## 15         120    1.050 0.05034048 0.06297795 0.06807125 0.09708849
## 16         120    1.100 0.05386303 0.06985423 0.07914999 0.10912990
```

For log zeta, the median absolute bias increases as both shape and scale increase, while increasing the number of intervals decreases the median absolute bias.

## Maximum bias

```
##    K_intervals k_priors        5       10       20       40
## 9           90    1.001 1.314942 1.610277 2.303521 2.996794
## 10          90    1.010 1.301835 1.620936 2.313728 3.007245
## 11          90    1.050 1.261308 1.666593 2.360157 3.054160
## 12          90    1.100 1.198658 1.722275 2.417701 3.113335
## 13         120    1.001 1.261332 1.610347 2.303819 2.996557
## 14         120    1.010 1.258074 1.620945 2.314122 3.006726
## 15         120    1.050 1.211867 1.665834 2.360032 3.053581
## 16         120    1.100 1.155280 1.721932 2.415878 3.111691
```

As with logit phi, the values for maximum absolute bias are all quite large. The relationship between bias and shape, scale, and intervals is the same as with median absolute bias.

### Median RMSE

```
##    K_intervals k_priors         5        10        20        40
## 9           90    1.001 0.3654061 0.3562483 0.3710361 0.4185133
## 10          90    1.010 0.3663205 0.3519171 0.3733957 0.4143586
## 11          90    1.050 0.3613494 0.3518462 0.3715042 0.4207069
## 12          90    1.100 0.3565163 0.3503199 0.3731234 0.4242671
## 13         120    1.001 0.3195476 0.3096926 0.3306055 0.3598567
## 14         120    1.010 0.3190805 0.3126668 0.3266423 0.3583929
## 15         120    1.050 0.3158147 0.3115416 0.3304555 0.3633310
## 16         120    1.100 0.3149784 0.3112431 0.3318905 0.3643657
```

As the number of intervals increases, the median RMSE decreases. Increasing the value of the sahe causes the RMSE to decrease very slightly, although the magnitude of the differences is very small. In some, but not all, cases the difference might be attributed to Monte Carlo error. Here the relationship between the value of the scale and the median RMSE is not monotonic. The smallest values occur when scale = 10, and increase when the scale increases or decreases.

### Maximum RMSE

```
##    K_intervals k_priors        5       10       20       40
## 9           90    1.001 1.350155 1.610282 2.303524 2.996794
## 10          90    1.010 1.336581 1.620936 2.313737 3.007245
## 11          90    1.050 1.294637 1.666593 2.360157 3.054160
## 12          90    1.100 1.232268 1.722277 2.417704 3.113339
## 13         120    1.001 1.308561 1.610350 2.303823 2.996557
## 14         120    1.010 1.304391 1.620946 2.314126 3.006726
## 15         120    1.050 1.258255 1.665836 2.360034 3.053581
## 16         120    1.100 1.200352 1.721935 2.415884 3.111694
```

The monotonic relationship seen in earlier tables reasserts itself here. Increasing the number of intervals and the value of the shape decreases the maximum RMSE. Increasing the scale increases the value of the maximum RMSE.

Taken together, this suggests that "good" estimates might be found when K = 120, shape = 1.001 or 1.01 and scale = 5.

## Further investigation

I had originally intended to include exploratory plots, however because I cannot control the orientation of plots in simple R-markdown and need latex to do that I decided to leave it out in the interest of time. Please take a look at 'bias plots.R' in the simulation 5 folder - they are part of the reasoning for the investigation outlined here.

Between the plots and tables of summary statistics, it's possible that a more granular view of the space beween scale = 5 and scale = 10 might be worth investigating. Phi is "ok" when the scale is higher than 10, but when scale = 20 zete is has considerable bias so future PIR simulations should probably stick to smaller values of scale.