

An Experiment with ‘*OrthoPanels*’

Davor Cubranic and Mark Pickup

2016-11-04

Let’s investigate the accuracy of *opm*’s parameter estimates on 200 simulated datasets.

First, let’s define the parameters used by the data-generating process:

```
rho <- .5
beta <- .5
sig2 <- 1
```

The following function generates a synthetic dataset of desired dimensions (N cases and T time points) and distribution parameters ($\rho = \text{rho}$, $\beta = \text{beta}$, and $\sigma^2 = \text{sig2}$):

```
generate <- function(N, T, rho, beta, sig2) {
  f <- runif(N, -2, 2)
  K <- length(beta)
  beta <- matrix(beta, K, 1)

  x <- array(.75*f, dim=c(N, K, T)) + rnorm(N*K*T, sd = 1)

  y <- matrix(0, N, T)
  for (t in 1:T) {
    yy <- if (t>1) y[,t-1] else 0
    y[,t] <- rho * yy + f + x[,t] %*% beta + rnorm(N, sd = sqrt(sig2))
  }

  data.frame(i = rep(seq(N), T),
             t = rep(seq(T), each = N),
             as.data.frame(matrix(aperm(x, c(1, 3, 2)), N*T, K,
                                   dimnames = list(NULL, paste0('x', seq(K))))),
             y = c(y))
}
```

Now we generate a dataset with $N=1000$ cases and $T=3$ time points and fit the model to it 200 times:

```
library(OrthoPanels)
library(knitr)

N <- 1000
T <- 3
reps <- 200

set.seed(123)
opms <- replicate(n = reps,
                  opm(y~x1,
                     data = generate(N = N, T = T,
                                     rho = rho,
                                     beta = beta,
                                     sig2 = sig2),
                     n.samp = 1000),
                  simplify = FALSE)
```

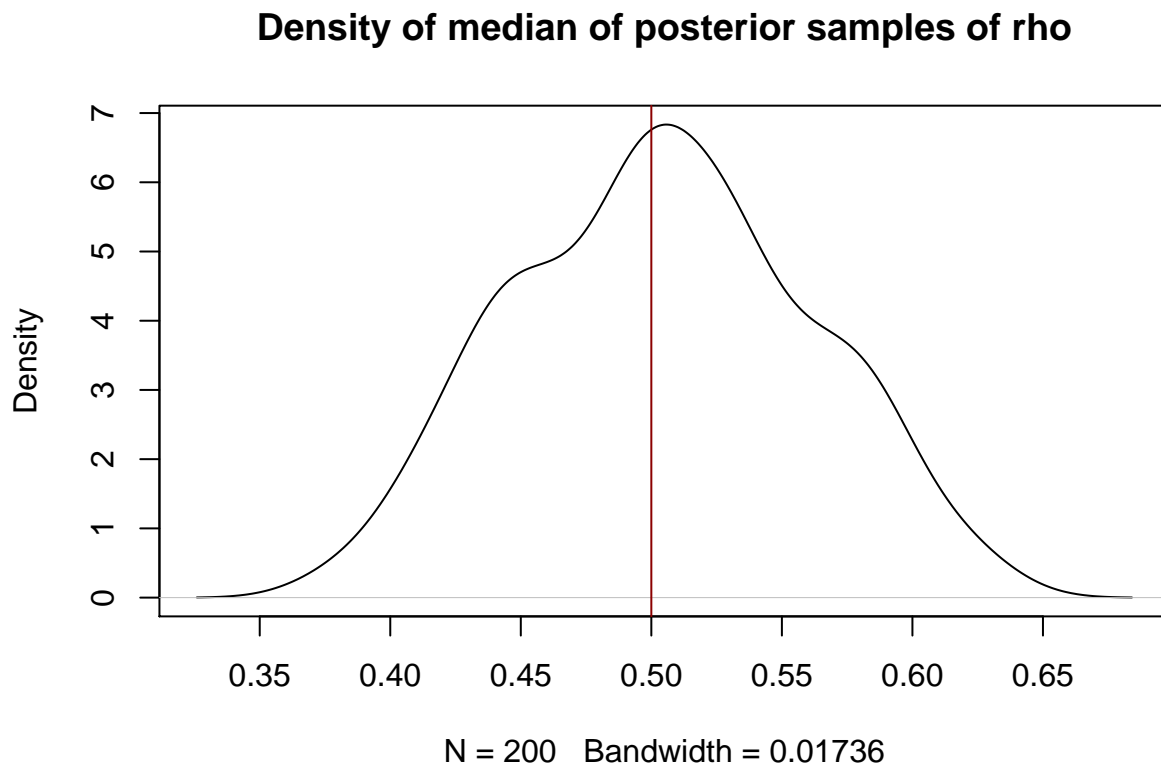
Let's check the sampled parameters:

```
true_param <- c(rho = rho, sig2 = sig2, beta = beta)
est_param <- sapply(opms, coef)
resid <- sweep(est_param, 1, true_param)
rmse <- sqrt(rowMeans(resid^2))
kable(rbind(`True` = true_param,
            `Est` = rowMeans(est_param),
            `Bias` = rowMeans(resid),
            `RMSE` = rmse))
```

	rho	sig2	beta
True	0.5000000	1.0000000	0.5000000
Est	0.5043025	1.0090909	0.5025456
Bias	0.0043025	0.0090909	0.0025456
RMSE	0.0556957	0.0704548	0.0352854

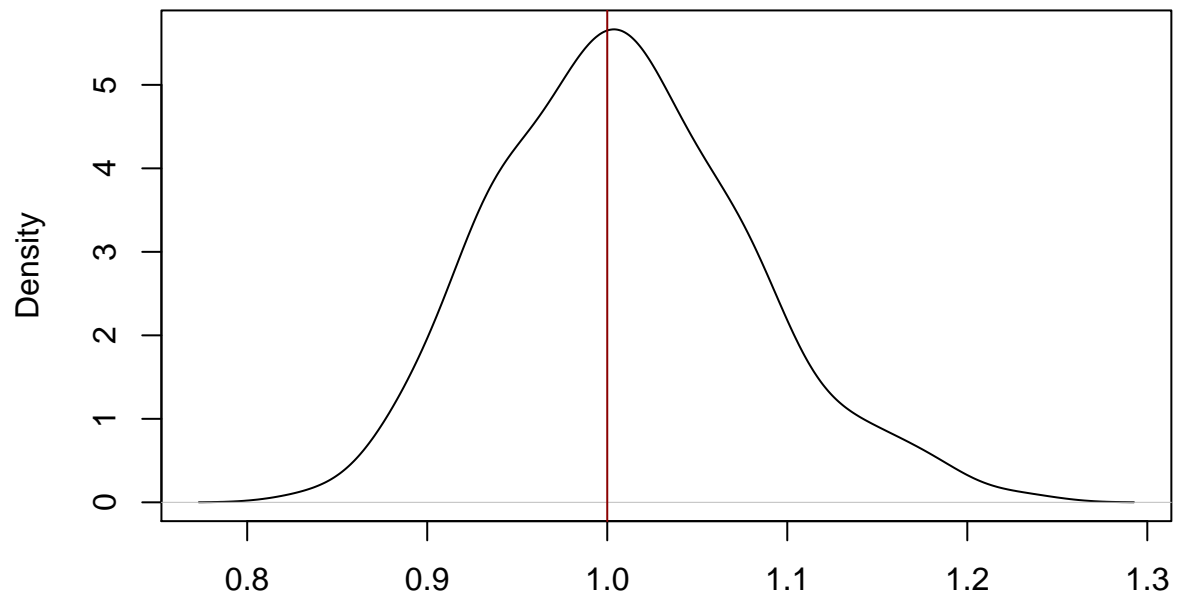
Density plot for each parameter, with true value marked with a vertical line:

```
plot(density(sapply(opms, coef)[1,]),
     main = 'Density of median of posterior samples of rho')
abline(v = rho, col='darkred')
```



```
plot(density(sapply(opms, coef)[2,]),
     main = 'Density of median of posterior samples of sig2')
abline(v = sig2, col='darkred')
```

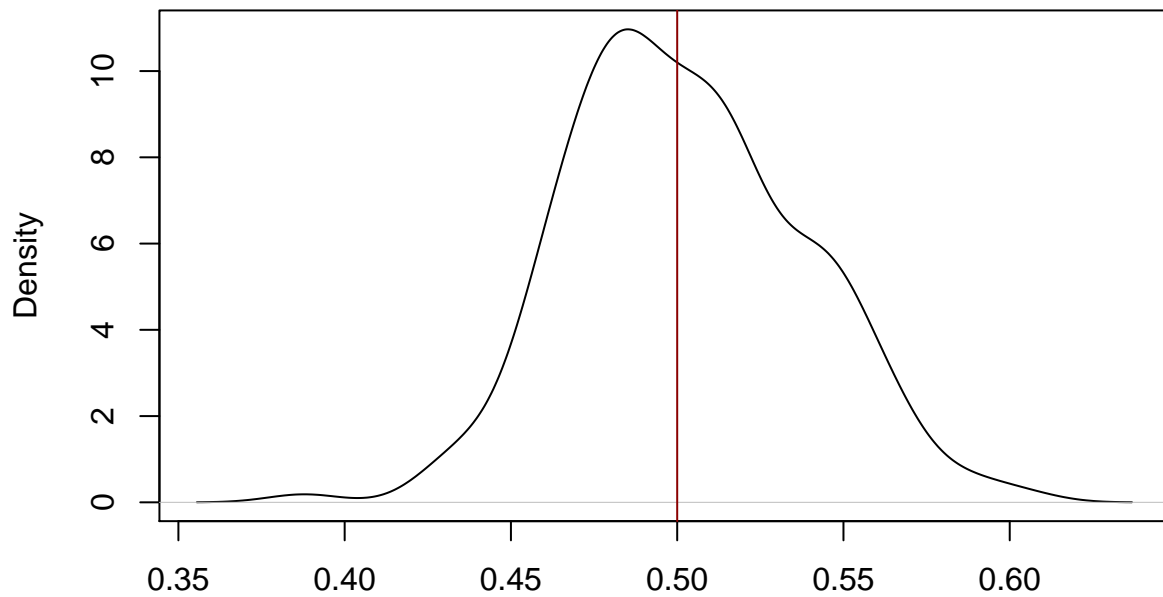
Density of median of posterior samples of sig2



N = 200 Bandwidth = 0.02147

```
plot(density(sapply(opms, coef)[3,]),  
     main = 'Density of median of posterior samples of beta')  
abline(v = beta, col='darkred')
```

Density of median of posterior samples of beta



N = 200 Bandwidth = 0.01076

The proportion of time the 95% credible interval includes the true value of the parameter:

```
cis <- sapply(lapply(opms, confint),
  function(ci) {
    ci[, '2.5%'] <= c(rho, sig2, beta) &
    ci[, '97.5%'] >= c(rho, sig2, beta)
  })
kable(rowSums(cis) / reps)
```

rho	0.895
sig2	0.920
beta	0.940