# vbsr: Variational Bayes Spike regression

Benjamin A. Logsdon

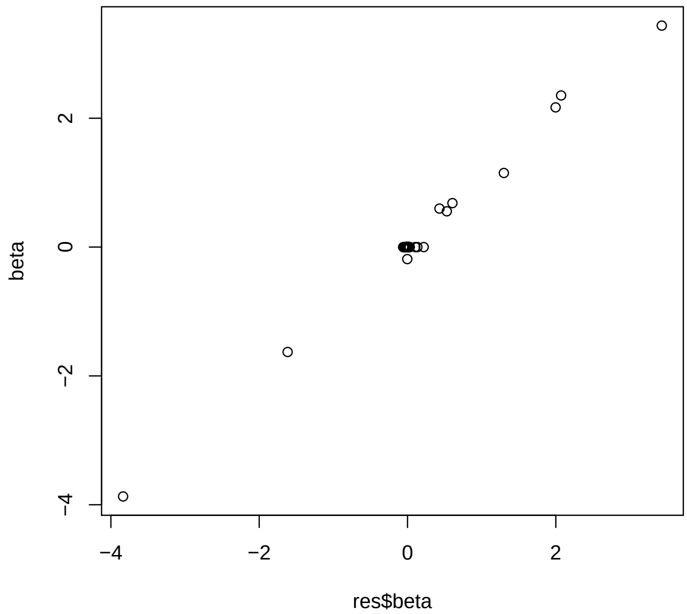
June 5, 2014

## 1 Example 1

We first consider the case of uncorrelated features, and a linear response, with a sparse true model with 100 observations, 95 variables, and 10 true variables:

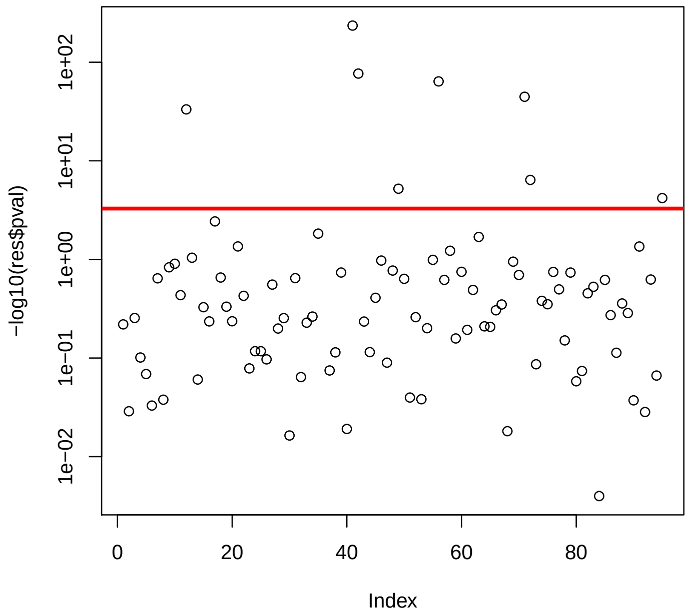
```
> library(vbsr)
> set.seed(2)
> n <- 100
> m <- 95
> ntrue <- 10
> e <- rnorm(n)
> X <- matrix(rnorm(n*m),n,m)
> tbeta <- sample(1:m,ntrue)
> beta <- rep(0,m)
> beta[tbeta]<- rnorm(ntrue,0,2)
> y <- X%*%beta+e
> res<- vbsr(y,X,family='normal')

> plot(res$beta,beta)
```




And the -log10 p-values:

```
> plot(-log10(res$pval),log='y')
> lines(c(-10,m+10),c(-log10(0.05/m),-log10(0.05/m)),col='red',lwd=3)
```




True features v.s. features significant in vbsr:

```
> cat('True variables:',sort(tbeta),'\n');
```

```
True variables: 12 34 36 41 42 49 56 71 72 95
```

```
> cat('Vbsr variables:',which(res$pval<0.05/m),'\n');
```

```
Vbsr variables: 12 36 41 42 49 56 71 72 95
```

Compare this to the OLS estimates

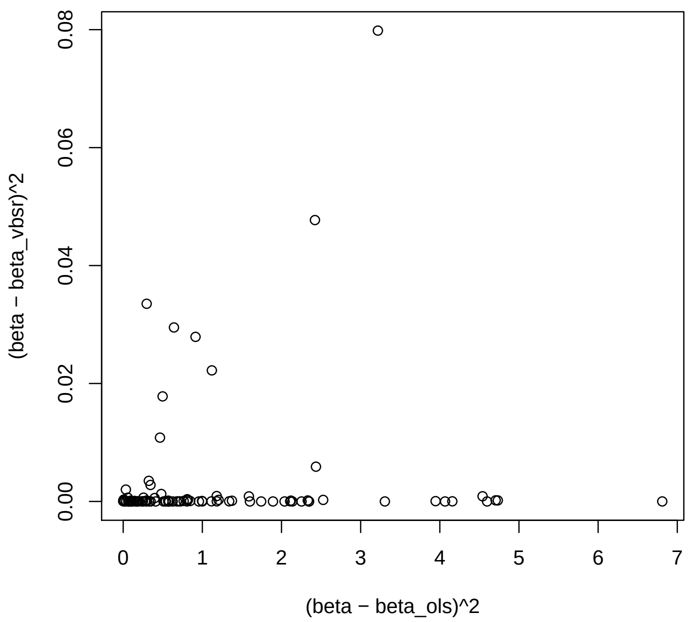
```
> ols <- lm(y~X);
> beta_ols <- summary(ols)$coef[-1,1];
> beta_vbsr <- res$beta;
> cat('OLS MSE:',mean((beta-beta_ols)^2),'\n');
```

```
OLS MSE: 1.13003
```

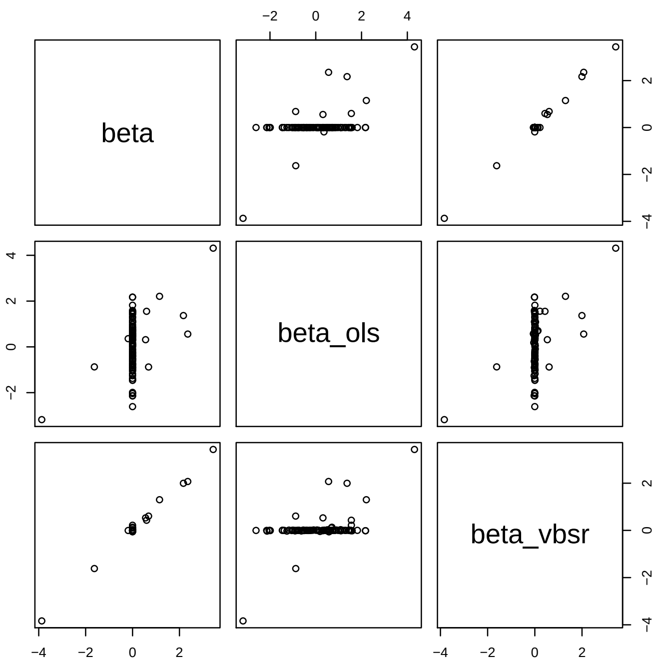
```
> cat('VBSR MSE:',mean((beta-beta_vbsr)^2),'\n');
```

```
VBSR MSE: 0.003087738
```

```
> #barplot(t(cbind(beta[tbeta],summary(ols)$coef[-1,1][tbeta],res$beta[tbeta])),beside=T,c
> #legend('topleft',c('beta','beta_ols','beta_vbsr'),fill=c('blue','red','green'))
> plot((beta-beta_ols)^2,(beta-beta_vbsr)^2)
```




```
> pairs(cbind(beta,beta_ols,beta_vbsr));
```




Compare to univariate estimates

```
> lmfun <- function(x,y){return(summary(lm(y~x))$coef[2,1]);}
> beta_uni <- apply(X,2,lmfun,y);
> cat('UNI MSE:',mean((beta-beta_uni)^2),'\n');
```

```
UNI MSE: 0.4133284
```

```
> cat('VBSR MSE:',mean((beta-beta_vbsr)^2),'\n');
```

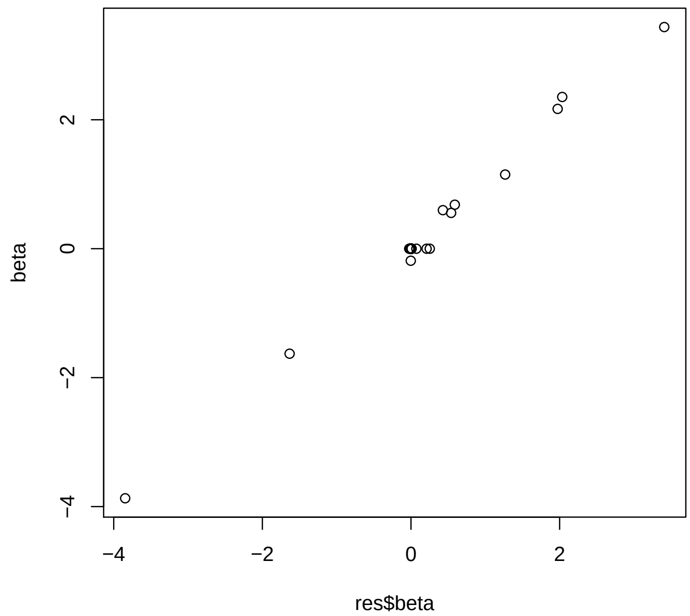
```
VBSR MSE: 0.003087738
```

# 2 Example 2

We next consider the case of moderately correlated features.

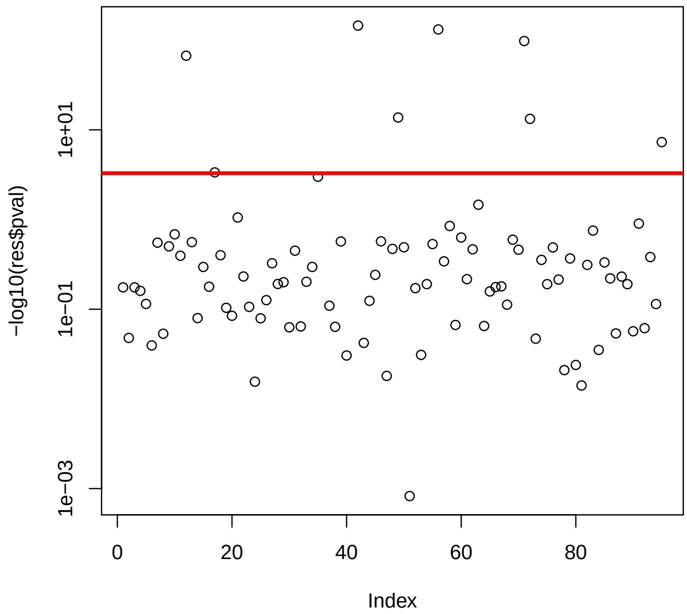
```
> g <- rnorm(n);
> X <- X+g;
> y <- X%*%beta+e
> res<- vbsr(y,X,family='normal')
```

```
> plot(res$beta,beta)
```




And the -log10 p-values:

```
> plot(-log10(res$pval),log='y')
> lines(c(-10,m+10),c(-log10(0.05/m),-log10(0.05/m)),col='red',lwd=3)
```




True features v.s. features significant in vbsr:

```
> cat('True variables:',sort(tbeta),'\n');
```

```
True variables: 12 34 36 41 42 49 56 71 72 95
```

```
> cat('Vbsr variables:',which(res$pval<0.05/m),'\n');
```

```
Vbsr variables: 12 17 36 41 42 49 56 71 72 95
```

Compare this to the OLS estimates

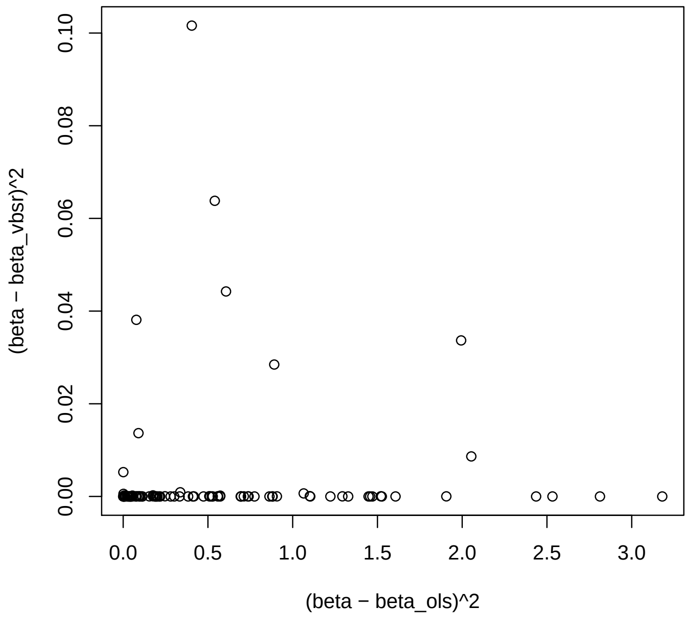
```
> ols <- lm(y~X);
> beta_ols <- summary(ols)$coef[-1,1];
> beta_vbsr <- res$beta;
> cat('OLS MSE:',mean((beta-beta_ols)^2),'\n');
```

```
OLS MSE: 0.579314
```

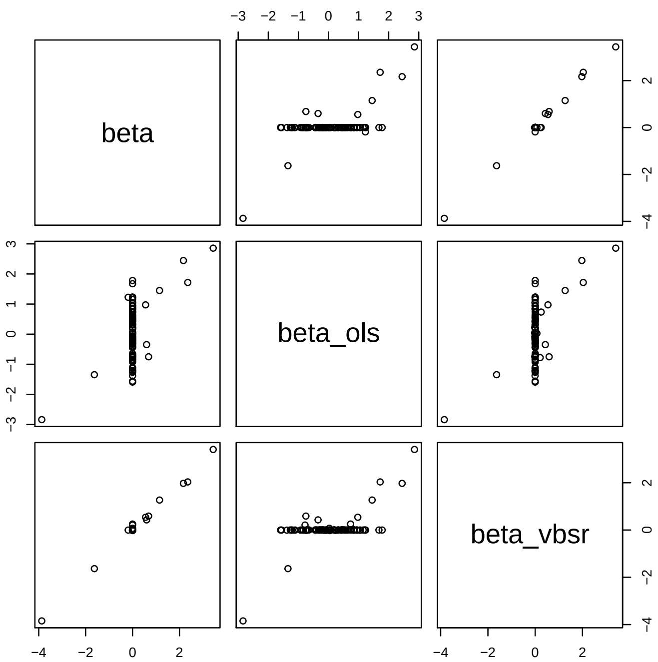
```
> cat('VBSR MSE:',mean((beta-beta_vbsr)^2),'\n');
```

```
VBSR MSE: 0.003592266
```

```
> #barplot(t(cbind(beta[tbeta],summary(ols)$coef[-1,1][tbeta],res$beta[tbeta])),beside=T,c
> #legend('topleft',c('beta','beta_ols','beta_vbsr'),fill=c('blue','red','green'))
> plot((beta-beta_ols)^2,(beta-beta_vbsr)^2)
```




```
> pairs(cbind(beta,beta_ols,beta_vbsr));
```




Compare to univariate estimates

```
> lmfun <- function(x,y){return(summary(lm(y~x))$coef[2,1]);}
> beta_uni <- apply(X,2,lmfun,y);
> cat('UNI MSE:',mean((beta-beta_uni)^2),'\n');
```

```
UNI MSE: 6.379615
```

```
> cat('VBSR MSE:',mean((beta-beta_vbsr)^2),'\n');
```

```
VBSR MSE: 0.003592266
```