Stan implementation of "How many X's" model by Zheng et al. (2006)

This is the Stan implementation of the model in Zheng et al. (2006):

```
data {
  int<lower=0> I;                          // number of respondents
  int<lower=0> K;                          // number of subpopulations
  vector[K] mu_beta;                       // prior mean of beta
  vector<lower=0>[K] sigma_beta;           // prior variance of beta
  int  y[I,K];                             // number known by respondent i in
                                           // subpopulation k
  }

parameters {
  vector[I] alpha;                         // log degree
  vector[K] beta;                          // log prevalence of group in population
  vector<lower=0, upper=1>[K] inv_omega; // ineverse overdispersion; implies the
                                           // uniform prior
  real mu_alpha;                           // prior mean for alpha
  real<lower=0> sigma_alpha;               // prior scale for alpha
  }

model {
// priors
  alpha ~ normal(mu_alpha, sigma_alpha);
  beta ~ normal(mu_beta, sigma_beta);     // informative prior on beta: location
                                           // and scale are specified

// hyperpriors
  mu_alpha ~ normal(0,25);                 // weakly informative (no prior in paper)
  sigma_alpha ~ normal(0,5);               // weakly informative (no prior in paper)


  for (k in 1:K) {
    real omega_k_m1;
    omega_k_m1 <- inv(inv(inv_omega[k]) - 1) ;
    for (i in 1:I) {
      real xi_i_k;
      xi_i_k <- omega_k_m1 * exp(alpha[i] + beta[k])  ;
      y[i,k] ~ neg_binomial(xi_i_k, omega_k_m1);
      }
    }
  }
```

Data are simulated from the model:

```
I <- 100;
K <- 32;

mu_alpha <- 5;
mu_beta <- -5;
```

```
sigma_alpha <- 1;
sigma_beta <- 1;

alpha <- rnorm(I, mu_alpha, sigma_alpha);
beta <- rnorm(K, mu_beta, sigma_beta);

omega_inv <- runif(K, 0.1, 0.95);
omega <- 1 / omega_inv;

y <- array(dim = c(I,K))
for (i in 1:I) {
  for (k in 1:K) {
    xi_i_k <- exp(alpha[i] + beta[k]) / (omega[k] - 1);
    y[i,k] <- rnbinom(1,
                      size = xi_i_k,
                      prob = 1 / omega[k]);
  }
}
```

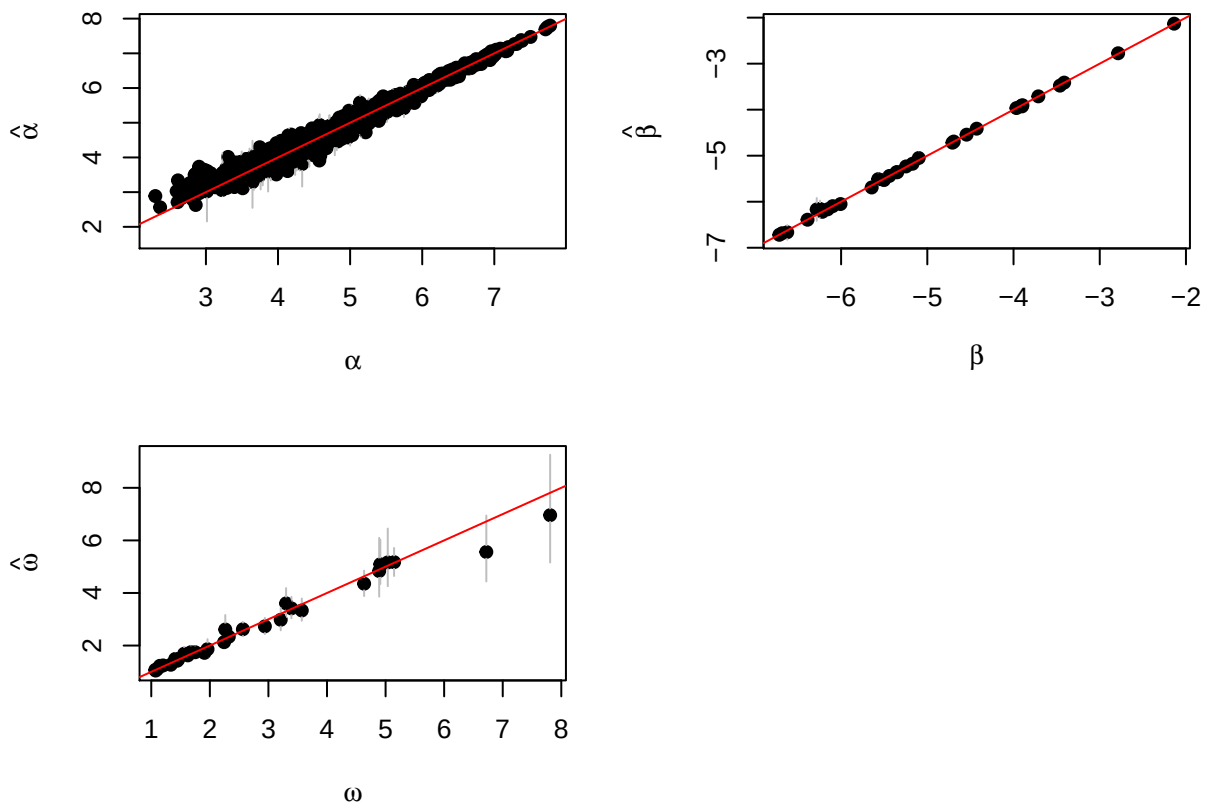
We eliminate responses with zero variance as is done in Zheng et al. (2006):

```
var0 <- apply(y, 1, var)
var0 <- which(var0 == 0)
if (length(var0) > 0) {
  y <- y[- var0, ]
  alpha <- alpha[- var0]
  I <- nrow(y)
  }
```

Hyper-parameter specification for informative priors; this is an alternative to the renormalization step in the MCMC by Zheng et al. (2006):

```
mu_beta <- c(beta[1:12], rep(mean(beta[1:12]), K - 12))
sigma_beta <- c(rep(.01, 12), rep(10, K - 12))
```

The results are presented as the plots of simulated versus estimated values of model parameters.



Zheng, T., Salganik, M. J. and Gelman, A. (2006) How Many People Do You Know in Prison?: Using Overdispersion in Count Data to Estimate Social Structure in Networks. Journal of the American Statistical Association. 101:474, 409-423.