# Bayesian Analysis of Choice Data

Simon Jackman

Stanford University
http://jackman.stanford.edu/BASS

February 3, 2012

# Discrete Choice

- binary (e.g., probit model; we looked at with data augmentation)

# Discrete Choice

- binary (e.g., probit model; we looked at with data augmentation)
- ordinal (ordinal logit or probit)

# Discrete Choice

- binary (e.g., probit model; we looked at with data augmentation)
- ordinal (ordinal logit or probit)
- multinomial models for unordered choices: e.g., multinomial logit (MNL), multinomial probit (MNP).

# Discrete Choice

- binary (e.g., probit model; we looked at with data augmentation)
- ordinal (ordinal logit or probit)
- multinomial models for unordered choices: e.g., multinomial logit (MNL), multinomial probit (MNP). We won't consider models for "tree-like" choice structures (nested logit, GEV, etc).

# Binary Choices: logit or probit

- for "standard" models (e.g., no "fancy" hierarchical structure, no concerns re missing data etc), other avenues besides `BUGS`/`JAGS`
- e.g., `MCMCpack`
- implementations in `BUGS`/`JAGS`: don't use data augmentation *a la* Albert & Chib (1991).
- `dbern` or `dbin` and sample from the conditional distributions using Metropolis-within-Gibbs, slice sampling

# Binary Choices: logit or probit

## Voter turnout example.

JAGS code

```
model{
  for (i in 1:N){                    ## loop over observations
    y[i] ~ dbern(p[i])               ## binary outcome
    logit(p[i]) <- ystar[i]          ## logit link
    ystar[i] <- beta[1]              ## regression structure for covariates
              + beta[2]*educ[i]
              + beta[3]*(educ[i]*educ[i])
              + beta[4]*age[i]
              + beta[5]*(age[i]*age[i])
              + beta[6]*south[i]
              + beta[7]*govelec[i]
              + beta[8]*closing[i]
              + beta[9]*(closing[i]*educ[i])
              + beta[10]*(educ[i]*educ[i]*closing[i])
   }

  ## priors
  beta[1:10] ~ dmnorm(mu[] , B[ , ])       # diffuse multivariate Normal prior
                                           # see data file
}
```

# Binary Data Is Binomial Data when Grouped (§8.1.4)

- big, micro-level data sets with binary data (e.g., CPS)
- MCMC gets slow
- collapse the data into *covariate classes*, treat as *binomial* data; much smaller data set, much shorter run-times
- $y_i|\mathbf{x}_i \sim \text{Bernoulli}(F[\mathbf{x}_i\boldsymbol{\beta}])$, where $\mathbf{x}_i$ is a vector of covariates.
- *Covariate classes*: a set $\mathcal{C} = \{i : \mathbf{x}_i = \mathbf{x}_{\mathcal{C}}\}$ i.e., the set of respondents who have covariate vector $\mathbf{x}_{\mathcal{C}}$.
- probability assignments over $y_i \; \forall i \in \mathcal{C}$ are conditionally exchangeable given their common $\mathbf{x}_i$ and $\boldsymbol{\beta}$.
- binomial model $r_{\mathcal{C}} \sim \text{Binomial}(p_{\mathcal{C}}; n_{\mathcal{C}})$, where $p_{\mathcal{C}} = F(\mathbf{x}_{\mathcal{C}}\boldsymbol{\beta})$, $r_{\mathcal{C}} = \sum_{i \in \mathcal{C}} y_i$ is the number of "successes" in $\mathcal{C}$ and $n_{\mathcal{C}}$ is the cardinality of $\mathcal{C}$.

# Example 8.5; binomial model for grouped binary data

Form covariate classes, and `groupedData` object; original data set $n \approx 99000$; only 636 unique covariates classes.

R code

```
## collapse by covariate classes
X <- cbind(nagler$age,nagler$educYrs)
X <- apply(X,1,paste,collapse=":")
covClasses <- match(X,unique(X))
covX <- matrix(unlist(strsplit(unique(X),":")),ncol=2,byrow=TRUE)
r <- tapply(nagler$turnout,covClasses,sum)
n <- tapply(nagler$turnout,covClasses,length)
groupedData <- list(n=n,r=r,
                    age=as.numeric(covX[,1]),
                    educYrs=as.numeric(covX[,2]),
                    NOBS=length(n))
```

# Example 8.5; binomial model for grouped binary data

We can then pass the `groupedData` data frame to `JAGS`. We specify the binomial model $r_i \sim \text{Binomial}(p_i; n_i)$ with $p_i = F(\mathbf{x}_i \boldsymbol{\beta})$ and vague normal priors on $\boldsymbol{\beta}$ with the following code:

JAGS code

```
model{
  for (i in 1:NOBS){
    logit(p[i]) <- beta[1] + age[i]*beta[2]
                          + pow(age[i],2)*beta[3]
                          + educYrs[i]*beta[4]
                          + pow(educYrs[i],2)*beta[5]
    r[i] ~ dbin(p[i],n[i])  ## binomial model for each covariate class
  }


  beta[1:5] ~ dmnorm(b0[],B0[,])
}
```

# Ordinal Responses

- e.g., 7-point scale when measuring party identification in the U.S., assigning the numerals $y_i \in \{0, \ldots, 6\}$ to the categories {"Strong Republican", "Weak Republican", . . ., "Strong Democrat"}.
- Censored, latent variable representation:

$$\begin{aligned} y_i^* &= \mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2), \quad i = 1, \ldots, n. \\ y_i = 0 &\iff y_i^* < \tau_1 \\ y_i = j &\iff \tau_j < y_i^* \leq \tau_{j+1}, \quad j = 1, \ldots, J - 1 \\ y_i = J &\iff y_i^* > \tau_J \end{aligned}$$

- threshold parameters obey the ordering constraint $\tau_1 < \tau_2 < \ldots < \tau_J$.
- The assumption of normality for $\varepsilon_i$ generates the probit version of the model; a logistic density generates the ordinal logistic model.
- Bayesian analysis: we want $p(\boldsymbol{\beta}, \boldsymbol{\tau}|\mathbf{y}, \mathbf{X}) \propto p(\mathbf{y}|\mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\tau})p(\boldsymbol{\beta}, \boldsymbol{\tau})$.

# Ordinal responses, $y_i^* \sim N(\mathbf{x}_i\boldsymbol{\beta}, \sigma^2)$

$$\Pr[y_i = j] = \Phi[(\tau_{j+1} - \mathbf{x}_i\boldsymbol{\beta})/\sigma] - \Phi[(\tau_j - \mathbf{x}_i\boldsymbol{\beta})/\sigma]$$

# Identification

$$
\begin{aligned}
y_i^* &= \mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i, \quad \varepsilon_i \sim N(0,\sigma^2), \quad i = 1,\ldots,n. \\
y_i = 0 &\iff y_i^* < \tau_1 \\
y_i = j &\iff \tau_j < y_i^* \leq \tau_{j+1}, \quad j = 1,\ldots,J-1 \\
y_i = J &\iff y_i^* > \tau_J
\end{aligned}
$$

- Model needs identification constraints
- Set one of the $\tau$ to a point (zero); set $\sigma$ to a constant (one)
- Drop the intercept and fix $\sigma$
- Fix two of the $\tau$ parameters.

# Priors on thresholds

- $\tau_j \sim N(0, 10^2)$, subject to ordering constraint $\tau_j > \tau_{j\text{-}1}, \forall\, j = 2, \ldots, J$. In `JAGS` only, use nifty `sort` function:

JAGS code

```
for(j in 1:4){
      tau0[j] ~ dnorm(0,.01)
}
tau[1:4] <- sort(tau0)    ## JAGS only, not in WinBUGS!
```

- BUGS:

$$
\begin{aligned}
\tau_1 &\sim N(t_1, T_1) \\
\delta_j &\sim \text{Exponential}(d), \qquad j = 2, \ldots, J, \\
\tau_j &= \tau_{j\text{-}1} + \delta_j, \qquad j = 2, \ldots, J,
\end{aligned}
$$

BUGS code

```
        tau[1] ~ dnorm(0,.01)
        for(j in 1:3){
                delta[j] ~ dexp(2)
                tau[j+1] <- tau[j] + delta[j]
        }
```

# Example 8.6, interviewer ratings of respondents

- 5 point rating scale used by interviewers in assessing respondents' levels of political information
- In 2000 ANES:

| Label | $y$ | $n$ | % |
|---|---|---|---|
| Very Low | 0 | 105 | 6 |
| Fairly Low | 1 | 334 | 19 |
| Average | 2 | 586 | 33 |
| Fairly High | 3 | 450 | 25 |
| Very High | 4 | 325 | 18 |

- covariates: education, gender, age, home-owner, public sector employment

# Ordinal Logistic Model

JAGS code

```
model{
        for(i in 1:N){  ## loop over observations
              ## form the linear predictor (no intercept)
              mu[i] <- x[i,1]*beta[1] +
                       x[i,2]*beta[2] +
                       x[i,3]*beta[3] +
                       x[i,4]*beta[4] +
                       x[i,5]*beta[5] +
                       x[i,6]*beta[6]

              ## cumulative logistic probabilities
              logit(Q[i,1]) <- tau[1]-mu[i]
              p[i,1] <- Q[i,1]
              for(j in 2:4){
                   logit(Q[i,j]) <- tau[j]-mu[i]
                   ## trick to get slice of the cdf we need
                   p[i,j] <- Q[i,j] - Q[i,j-1]
              }
              p[i,5] <- 1 - Q[i,4]
              y[i] ~ dcat(p[i,1:5])  ## p[i,] sums to 1 for each i
        }

        ## priors over betas
        beta[1:6] ~ dmnorm(b0[],B0[,])

        ## thresholds
        for(j in 1:4){
              tau0[j] ~ dnorm(0, .01)
        }
        tau[1:4] <- sort(tau0)  ## JAGS only not in BUGS!
}
```

# Redundant Parameterization

- exploit lack of identification
- run the MCMC algorithm deployed in the space of *unidentified parameters*
- *post-processing*: map MCMC output back mixes better than the MCMC algorithm in the space of the *identified* parameters
- get a better mixing Markov chain
- in ordinal model case, exploit lack of identification between thresholds and intercept parameters
- take care!

# Interviewer heterogeneity in scale-use

- Different interviewers use the rating scale differently: e.g., interviewer $k$ is a tougher grader than interviewer $k'$.
- We tap this with a set of interviewer terms, varying over interviewers $k = 1, \ldots, K$
- We augment the usual ordinal model as follows:

$$\begin{aligned}
\Pr(y_i \geq j) &= F(\tau_j - \mu_i), \quad j = 0, \ldots, J - 1 \\
\Pr(y_i = J) &= 1 - F(\tau_{J-1} - \mu_i) \\
\mu_i &= \mathbf{x}_i \boldsymbol{\beta} + \eta_k \\
\eta_k &\sim N(0, \sigma^2) \quad k = 1, \ldots, K
\end{aligned}$$

- A positive $\eta_k$ is equivalent to the thresholds being shifted down (i.e., interviewer $k$ is an easier-than-average grader).
- Zero-mean restriction on $\eta_k$: why?
- Alternative model: each interviewer gets their own set of thresholds, perhaps fit these hierarchically.

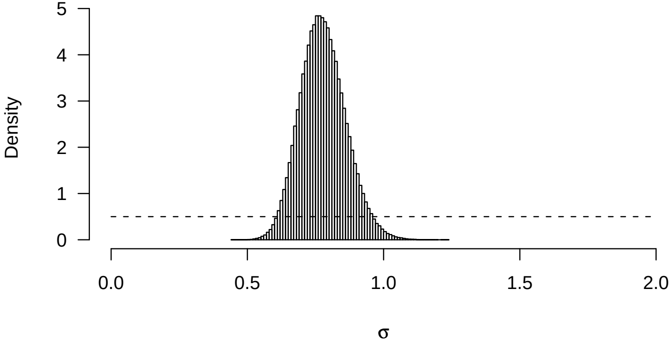
# JAGS code for hierarchical model

JAGS code

```
model{
        for(i in 1:N){  ## loop over observations
              ## form the linear predictor
              mu[i] <- x[i,1]*beta[1] + x[i,2]*beta[2] + x[i,3]*beta[3] +
                       x[i,4]*beta[4] + x[i,5]*beta[5] + x[i,6]*beta[6] + eta[id[i]]

              ## cumulative logistic probabilities
              logit(Q[i,1]) <- tau[1]-mu[i]
              p[i,1] <- Q[i,1]
              for(j in 2:4){
                logit(Q[i,j]) <- tau[j]-mu[i]
                p[i,j] <- Q[i,j] - Q[i,j-1]
              }
              p[i,5] <- 1 - Q[i,4]
              y[i] ~ dcat(p[i,1:5])  ## p[i,] sums to 1 for each i
        }
        ## priors over betas
        beta[1:6] ~ dmnorm(b0[],B0[,])
        ## hierarchical model over etas, note zero mean restriction
        for(k in 1:NID){
              eta[k] ~ dnorm(0.0,eta.tau)
        }
        eta.tau <- 1/pow(sigma,2)  ## convert stddev to precision
        sigma ~ dunif(0,2)
        ## priors over thresholds
        for(j in 1:4){
              tau0[j] ~ dnorm(0,.01)
        }
        tau[1:4] <- sort(tau0)   ## JAGS only, not in WinBUGS!
}
```

# $\sigma$, prior and posterior densities




- since $\eta_k \sim N(0, \sigma^2)$, if we set $\sigma$ to its posterior mean of .77, then half of the interviewer effects will lie more than 1.35 $\sigma \approx 1.04$ "logits" away from zero.

# Tabular summary of results

| | Non-Hierarchical | Hierarchical |
|---|---|---|
| College Degree | 1.46 | 1.61 |
| | (.10) | (.10) |
| Female | -.66 | -.76 |
| | (.09) | (.09) |
| log(Age) | .47 | .42 |
| | (.12) | (.13) |
| Home Owner | .45 | .48 |
| | (.10) | (.10) |
| Government Employee | .17 | .16 |
| | (.14) | (.14) |
| log(Interview Length) | 1.13 | 1.45 |
| | (.15) | (.18) |
| $\sigma$ | 0 | .77 |
| | | (.08) |
| Threshold parameters: | | |
| $\tau_0$ | 3.85 | 4.69 |
| | (.67) | (.75) |
| $\tau_1$ | 5.66 | 6.60 |
| | (.67) | (.75) |
| $\tau_2$ | 7.37 | 8.46 |
| | (.68) | (.76) |
| $\tau_3$ | 8.83 | 10.08 |
| | (.69) | (.77) |

# Models for Multinomial Choices, §8.3

- multinomial logit (MNL), §8.3.1
- multinomial probit (MNP) §8.3.2

# Multinomial logit (MNL), §8.3.1

- Random utility rationale: utility to decision-maker $i$ of choice $j$ is linear in some predictors, plus a random component,

$$U_{ij} = \mathbf{x}_i \boldsymbol{\beta}_j + \varepsilon_{ij}, j = 0, \ldots, J$$

- $\varepsilon_{ij}$ are drawn a distribution whose cumulative distribution function is a Type-1 extreme value distribution with functional form $F(\varepsilon_{ij}) = \exp[-\exp(-\varepsilon_{ij})]$ and hence $\varepsilon_{ij}$ has density

$$p(\varepsilon_{ij}) = \exp(-\varepsilon_{ij})\exp[-\exp(-\varepsilon_{ij})].$$

- Decision-maker $i$ chooses option $j$ with probability

$$\pi_{ij} = \Pr(y_i = j) = \Pr[U_{ij} > U_{ik}], \quad \forall\, k \neq j.$$

# Multinomial logit (MNL), §8.3.1

- Consider a choice set with 3 elements, {"0", "1", "2"}.
- Suppose we observe $y_i = 2$:

$$\begin{aligned}
\Pr(y_i = 2) &= \Pr(U_{i2} > U_{i1}, U_{i2} > U_{i0}) \\
&= \Pr[\mathbf{x}_i\boldsymbol{\beta}_2 + \varepsilon_{i2} > \mathbf{x}_i\boldsymbol{\beta}_1 + \varepsilon_{i1},\ \mathbf{x}_i\boldsymbol{\beta}_2 + \varepsilon_{i2} > \mathbf{x}_i\boldsymbol{\beta}_0 + \varepsilon_{i0}], \\
&= \Pr[\varepsilon_{i2} + \mathbf{x}_i\boldsymbol{\beta}_2 - \mathbf{x}_i\boldsymbol{\beta}_1 > \varepsilon_{i1},\ \varepsilon_{i2} + \mathbf{x}_i\boldsymbol{\beta}_2 - \mathbf{x}_i\boldsymbol{\beta}_0 > \varepsilon_{i0}], \\
&= \int_{-\infty}^{\infty} f(\varepsilon_2) \left[ \int_{-\infty}^{\varepsilon_{i2} + \mathbf{x}_i\boldsymbol{\beta}_2 - \mathbf{x}_i\boldsymbol{\beta}_1} f(\varepsilon_1)\, d\varepsilon_1 \cdot \int_{-\infty}^{\varepsilon_{i2} + \mathbf{x}_i\boldsymbol{\beta}_2 - \mathbf{x}_i\boldsymbol{\beta}_0} f(\varepsilon_0)\, d\varepsilon_0 \right] d\varepsilon_2, \\
&= \int_{-\infty}^{\infty} f(\varepsilon_2) \times \exp[-\exp(-\varepsilon_{i2} - \mathbf{x}_i\boldsymbol{\beta}_2 + \mathbf{x}_i\boldsymbol{\beta}_1)] \times \exp[-\exp(-\varepsilon_{i2} - \mathbf{x}_i\boldsymbol{\beta}_2 + \mathbf{x}_i\boldsymbol{\beta}_0)] \\
&= \frac{\exp(\mathbf{x}_i\boldsymbol{\beta}_2)}{\exp(\mathbf{x}_i\boldsymbol{\beta}_0) + \exp(\mathbf{x}_i\boldsymbol{\beta}_1) + \exp(\mathbf{x}_i\boldsymbol{\beta}_2)}.
\end{aligned}$$

- Thus:

$$\pi_{ij} = \Pr(y_i = j) = \frac{\exp(\mathbf{x}_i\boldsymbol{\beta}_j)}{\sum_{k=0}^{J} \exp(\mathbf{x}_i\boldsymbol{\beta}_k)}.$$

# Multinomial logit (MNL), §8.3.1

$$\pi_{ij} = \Pr(y_i = j) = \frac{\exp(\mathbf{x}_i \boldsymbol{\beta}_j)}{\sum_{k=0}^{J} \exp(\mathbf{x}_i \boldsymbol{\beta}_k)}.$$

- Identification by normalizing on a "baseline outcome", e.g., $\boldsymbol{\beta}_0 = \mathbf{0}$.
- Independence of irrelevant alternatives §8.3.2

# Example 8.7

- Vote choice in the 1992 U.S. Presidential election
- ANES data; choices are Clinton, George H.W. Bush, Perot. $n = 909$.
- Original analysis by Alvarez and Nagler (1995), who used MNP.
- Predictors: dummies for Dem or Rep party-id, dummy for gender, retrospective evaluations of the national economy (-1, 0, 1), and $z_{ij}$, square of the distance of respondent $i$ from candidate $j$.
- $$\begin{aligned} \Pr(U_{ij} > U_{ik}) &= \Pr(\mathbf{x}_i\boldsymbol{\beta}_j + z_{ij}\gamma + \varepsilon_{ij} - \mathbf{x}_i\boldsymbol{\beta}_k - z_{ij}\gamma - \varepsilon_{ik} > 0) \\ &= \Pr(\mathbf{x}_i[\boldsymbol{\beta}_j - \boldsymbol{\beta}_k] + [z_{ij} - z_{ik}]\gamma > \varepsilon_{ik} - \varepsilon_{ij}). \end{aligned}$$

# Example 8.7, using dcat

JAGS code

```
model{
        for(i in 1:NOBS){
              for(j in 1:3){    ## loop over choices
                          mu[i,j] <- beta[j,1]
                             + beta[j,2]*dem[i]
                             + beta[j,3]*ind[i]
                             + beta[j,4]*rep[i]
                             + beta[j,5]*female[i]
                             + beta[j,6]*natlecon[i]
                             + gamma*dist[i,j]
                    emu[i,j] <- exp(mu[i,j])
                    p[i,j] <- emu[i,j]/sum(emu[i,1:3])
              }
              y[i] ~ dcat(p[i,1:3])
        }

        ## priors
        for(k in 1:6){
              beta[1,k] <- 0  ## identifying restriction
        }
        for(j in 2:3){
              beta[j,1:6] ~ dmnorm(b0,B0)  ## b0, B0 passed as data from R
        }
        gamma ~ dnorm(0,.01)

        ## plus code for mapping to identified parameters, see book
}
```

# Multinomial Probit, §8.4

- same random utility rationale:

$$U_{ij} = \mathbf{r}_{ij}\boldsymbol{\beta} + v_{ij}, \quad j = 0, 1, \ldots, J; i = 1, \ldots, n$$

- MNP for MVN model for un-modelled sources of utility:

$$\mathbf{v}_i = (v_{i1}, \ldots, v_{iJ})' \overset{\text{iid}}{\sim} N(\mathbf{0}, \mathbf{V})$$

  where $\mathbf{V}$ is a $(J+1)$-by-$(J+1)$ covariance matrix.
- But probabilities are difficult to compute:

$$\begin{aligned}\pi_{ij} &= \Pr(y_i = j) = \Pr(U_{ij} > U_{ik}), \ \forall\, k \neq j \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{U_{ij}} \ldots \int_{-\infty}^{U_{ij}} f(U_0, U_1, \ldots, U_J)\, dU_0\, dU_1 \ldots dU_J\end{aligned}$$

# Multinomial Probit, MCMC via data augmentation

- if choice $j$ is observed for person $i$, we know that $U_{ij} - U_{ik} > 0 \, \forall \, j \neq k$.
- Without loss of generality choose a "baseline" outcome, $j = 0$, and define the utility differences $\mathbf{w}_i = (w_{i1}, \ldots, w_{iJ})'$ with $w_{ij} = U_{ij} - U_{i0}, j = 1, \ldots, J$:

$$w_{ij} = (\mathbf{r}_{ij} - \mathbf{r}_{i0})\boldsymbol{\beta} + v_{ij} - v_{i0} = \mathbf{x}_{ij}\boldsymbol{\beta} + \varepsilon_{ij}.$$

  where $\boldsymbol{\varepsilon}_i \overset{\text{iid}}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma})$
- mapping from latent variables to observed choices:

$$y_i = h(\mathbf{w}_i) \equiv \begin{cases} 0 & \text{if } \max(\mathbf{w}_i) < 0 \\ j & \text{if } \max(\mathbf{w}_i) = w_{ij} > 0 \end{cases}$$

- Identification: the distribution of $\mathbf{y}|\mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\Sigma}$ is the same as the distribution of $\mathbf{y}|\mathbf{X}, c\boldsymbol{\beta}, c^2\boldsymbol{\Sigma}$
- solution: set $\sigma_{11} = 1$.

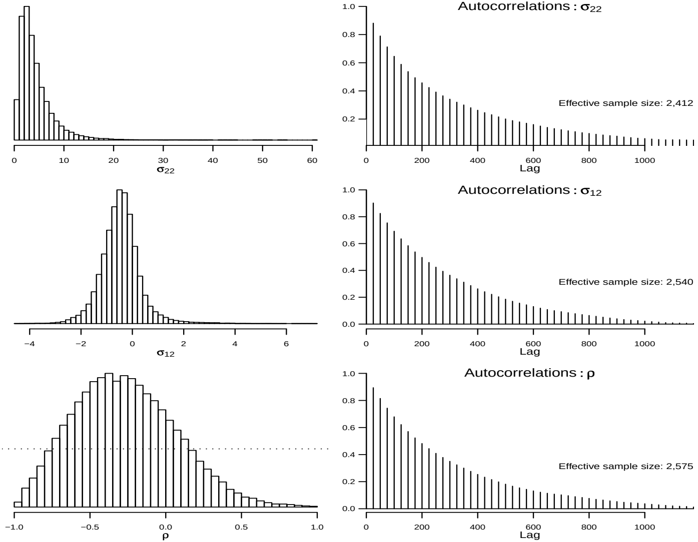
# Multinomial Probit, MCMC via data augmentation

- posterior density $p(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\mathbf{y}, \mathbf{X})$
  1. sample $\mathbf{w}_i^{(t)}$ from $p(\mathbf{w}_i|\boldsymbol{\beta}^{(t\text{-}1)}, \boldsymbol{\Sigma}^{(t\text{-}1)}, \mathbf{y}, \mathbf{X})$, $i = 1, \ldots, n$, the data-augmentation step
  2. sample $\boldsymbol{\beta}^{(t)}$ from $p(\boldsymbol{\beta}|\boldsymbol{\Sigma}^{(t\text{-}1)}, \mathbf{W}^{(t)}, \mathbf{y}, \mathbf{X})$.
  3. sample $\boldsymbol{\Sigma}^{(t)}$ from $p(\boldsymbol{\Sigma}|\boldsymbol{\beta}^{(t)}, \mathbf{W}^{(t)}, \mathbf{y}, \mathbf{X})$.
- Conditional on the latent $\mathbf{w}_i$, we have a very simple multivariate normal regression (McCulloch and Rossi 1994; Chib and Greenberg 1997; McCulloch, Polson and Rossi 1998).
- For step 3, the prior and the conditional distribution for $\boldsymbol{\Sigma}$ is complicated by the identifying constraint $\sigma_{11} = 1$.
- Implemented in `MNP` package in `R` (Imai and van Dyk 2005).

# Example 8.8, 1992 U.S. Presidential election

- $n = 909$, $j \in \{\text{Perot}, \text{Bush}, \text{Clinton}\}$
- mix of individual (party-id, gender, evaluations of the economy) and choice-specific covariates (squared ideological distance from candidates)
- `MNP` in `R`, 1.5M iterations, extremely inefficient exploration of the posterior densities for some parameters

# Example 8.8, 1992 U.S. Presidential election



# Example 8.8, 1992 U.S. Presidential election

1.5 million iterations:

| | $z$ | $p$ | $\rho$ | $N$ | $I$ | EffSamp |
|---|---|---|---|---|---|---|
| $\beta_{11}$, Intercept, Perot | -0.82 | 0.62 | 0.37 | 303,375 | 81.00 | 5,908 |
| $\beta_{21}$, Intercept, Clinton | -0.36 | 0.36 | 0.46 | 547,875 | 146.00 | 6,211 |
| $\beta_{12}$, Dem Id, Perot | -1.12 | 0.49 | 0.48 | 320,025 | 85.40 | 4,076 |
| $\beta_{22}$, Dem Id, Clinton | -0.27 | 0.69 | 0.73 | 860,850 | 230.00 | 3,175 |
| $\beta_{13}$, Repub Id, Perot | 0.66 | 0.72 | 0.34 | 659,850 | 176.00 | 6,599 |
| $\beta_{23}$, Repub Id, Clinton | 0.42 | 0.72 | 0.75 | 958,400 | 256.00 | 2,763 |
| $\beta_{14}$, Female, Perot | -0.07 | 0.32 | 0.01 | 94,025 | 25.10 | 58,544 |
| $\beta_{24}$, Female, Clinton | 1.04 | 0.51 | 0.02 | 99,850 | 26.70 | 50,304 |
| $\beta_{15}$, Econ Retro, Perot | 0.51 | 0.97 | 0.18 | 198,950 | 53.10 | 11,272 |
| $\beta_{25}$, Econ Retro, Clinton | 0.15 | 0.57 | 0.59 | 607,750 | 162.00 | 3,709 |
| $\gamma$, Ideological Distance | 0.21 | 0.57 | 0.72 | 967,950 | 258.00 | 2,778 |
| $\sigma_{12}$ | -1.46 | 0.09 | 0.90 | 975,375 | 260.00 | 2,540 |
| $\sigma_{22}$ | -0.78 | 0.98 | 0.88 | 902,500 | 241.00 | 2,412 |
| $\rho$ | -1.16 | 0.33 | 0.89 | 1,159,350 | 309.00 | 2,575 |

# References


Alvarez, R. Michael and Jonathan Nagler. 1995. "Economics, Issues and the Perot Candidacy: Voter Choice in the 1992 Presidential Election." *American Journal of Political Science* 39:714--44.

Chib, Siddhartha and Edward Greenberg. 1997. "Analysis of Multivariate Probit Models." *Biometrika* 85:347--361.

Imai, Kosuke and David A. van Dyk. 2005. "MNP: R Package for Fitting the Multinomial Probit Model." *Journal of Statistical Software* 14:1--32.

McCulloch, Robert E., Nicholas G. Polson and Peter E. Rossi. 1998. "A Bayesian Analysis of the Multinomial Probit Model with Fully Identified Parameters." Typescript. Graduate School of Business, University of Chicago.

McCulloch, Robert E. and Peter E. Rossi. 1994. "An exact likelihood analysis of the multinomial probit model." *Journal of Econometrics* 64:207--40.