Harvard Department of Government 2017
Chapter 5, The Bayesian Linear Model

JEFF GILL

Visiting Professor, Spring 2025

Bayesian Linear Regression Model

- $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\mathbf{X}$ is an $n \times k$, rank k matrix of explanatory variables with a leading vector of ones for the constant, $\boldsymbol{\beta}$ is a $k \times 1$ vector of coefficients to be estimated, $\mathbf{y}$ is an $n \times 1$ vector of outcome variable values, and $\boldsymbol{\epsilon}$ is a $n \times 1$ vector of errors distributed $\mathcal{N}(0, \sigma^2 I)$ for a constant σ^2.

- The likelihood function for a sample of size n is:

$$L(\boldsymbol{\beta}, \sigma^2 | \mathbf{X}, \mathbf{y}) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right].$$

- The values for which this form is at its maximum from standard maximum likelihood theory and ordinary least squares principles:

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, \qquad \hat{\sigma}^2 = \frac{(\mathbf{y} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\mathbf{b})}{(n-k)},$$

Bayesian Linear Regression Model (cont.)

- Now plug these in and simplify:

$$
\begin{aligned}
L(\boldsymbol{\beta},\sigma^2|\mathbf{X},\mathbf{y}) &\propto \sigma^{-n}\exp\left[-\frac{1}{2\sigma^2}(\mathbf{y}'\mathbf{y}-2\boldsymbol{\beta}'\mathbf{X}'\mathbf{y}+\boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta})\right]\\
&= \sigma^{-n}\exp\Big[-\frac{1}{2\sigma^2}(\mathbf{y}'\mathbf{y}-2\boldsymbol{\beta}'\mathbf{X}'\mathbf{y}+\boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta})\\
&\qquad \underbrace{-2((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y})'\mathbf{X}'\mathbf{y}+2((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y})'\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y})}_{\text{sums to zero}}\Big]\\
&= \sigma^{-n}\exp\Big[-\frac{1}{2\sigma^2}(\mathbf{y}'\mathbf{y}-2\underbrace{\boldsymbol{\beta}'\mathbf{X}'\mathbf{y}}_{\boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\mathbf{b}}+\boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}\\
&\qquad -2(\underbrace{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}}_{\mathbf{b}})'\mathbf{X}'\mathbf{y}+2(\underbrace{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}}_{\mathbf{b}})'\mathbf{X}'\mathbf{X}\underbrace{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}}_{\mathbf{b}})\Big]\\
&= \sigma^{-n}\exp\left[-\frac{1}{2\sigma^2}((\mathbf{y}-\mathbf{X}\mathbf{b})'(\mathbf{y}-\mathbf{X}\mathbf{b})+\mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b}+\boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}-2\boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\mathbf{b})\right]\\
&= \sigma^{-n}\exp\left[-\frac{1}{2\sigma^2}(\hat{\sigma}^2(n-k)+(\boldsymbol{\beta}-\mathbf{b})'\mathbf{X}'\mathbf{X}(\boldsymbol{\beta}-\mathbf{b}))\right].
\end{aligned}
$$

Bayesian Linear Regression Model (cont.)

- Tricks used here:
  - completing the square after inserting a quantity that is simultaneously subtracted and added,

$$\underbrace{-2((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y})'\mathbf{X}'\mathbf{y} + 2((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y})'\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y})}_{\text{sums to zero}}$$

  - splitting up $+2\mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b}$,
  - using the property $\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = I$ to rearrange terms,
  - replacement of $-2\boldsymbol{\beta}'\mathbf{X}'\mathbf{y}$ with $-2\boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\mathbf{b}$ from the normal equation ($\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{y}$).

Bayesian Linear Regression Model (cont.)

- **Use Uninformative Priors** $p(\boldsymbol{\beta}) \propto c$ and $p(\sigma^2) = \frac{1}{\sigma}$ over the support $[-\infty{:}\infty]$ and $[0{:}\infty]$.

- Therefore the joint posterior from the likelihood function is provided by:

$$\begin{aligned}\pi(\boldsymbol{\beta}, \sigma^2 | \mathbf{X}, \mathbf{y}) &\propto L(\boldsymbol{\beta}, \sigma^2 | \mathbf{X}, \mathbf{y}) p(\boldsymbol{\beta}) p(\sigma^2) \\ &\propto \sigma^{-n-1} \exp\left[-\frac{1}{2\sigma^2} (\hat{\sigma}^2 (n-k) + (\boldsymbol{\beta} - \mathbf{b}) \mathbf{X}' \mathbf{X} (\boldsymbol{\beta} - \mathbf{b})) \right].\end{aligned}$$

- Note that the c constant from $p(\boldsymbol{\beta}) \propto c$ drops out with proportionality.

Bayesian Linear Regression Model (cont.)

- To obtain the desired marginal distribution of $\boldsymbol{\beta}$, first make the transformation: $s = \sigma^{-2}$, with a required Jacobian:

$$J = \left|\frac{\partial}{\partial s}\sigma\right| = \left|\frac{\partial}{\partial s}s^{-\frac{1}{2}}\right| = \frac{1}{2}s^{-\frac{3}{2}}.$$

  Note that this is a re-expression of the variance as a precision.

- Reexpressing in terms of s gives the joint posterior distribution:

$$\pi(\boldsymbol{\beta}, s|\mathbf{X}, \mathbf{y}) \propto (s^{-\frac{1}{2}})^{-n-1}\exp\left[-\frac{1}{2}s(\hat{\sigma}^2(n-k) + (\boldsymbol{\beta}-\mathbf{b})\mathbf{X}'\mathbf{X}(\boldsymbol{\beta}-\mathbf{b}))\right]\left(\frac{1}{2}s^{-\frac{3}{2}}\right)$$

$$\propto s^{\frac{n}{2}-1}\exp\left[-\frac{1}{2}s(\hat{\sigma}^2(n-k) + (\boldsymbol{\beta}-\mathbf{b})\mathbf{X}'\mathbf{X}(\boldsymbol{\beta}-\mathbf{b}))\right].$$

- Now our job is to integrate this two different ways to get the desired marginals.

Bayesian Linear Regression Model (cont.)

- Now integrate with respect to s to get the marginal for $\boldsymbol{\beta}$:

$$\pi(\boldsymbol{\beta}|\mathbf{X},\mathbf{y}) = \int_0^\infty s^{\frac{n}{2}-1}\exp\left[-\frac{1}{2}s(\hat{\sigma}^2(n-k)+(\boldsymbol{\beta}-\mathbf{b})\mathbf{X}'\mathbf{X}(\boldsymbol{\beta}-\mathbf{b}))\right]ds.$$

Inside the integral is a gamma PDF kernel, with the integration performed over the appropriate support. So use the following substitution:

$$1 = \int_0^\infty \frac{q^{p+1}}{\Gamma(p+1)}s^p e^{-qs}ds, \qquad \Longrightarrow \qquad \frac{\Gamma(p+1)}{q^{p+1}} = \int_0^\infty s^p e^{-qs}ds.$$

Setting $p = \frac{n}{2} - 1$ and $q = \frac{1}{2}(\hat{\sigma}^2(n-k)+(\boldsymbol{\beta}-\mathbf{b})\mathbf{X}'\mathbf{X}(\boldsymbol{\beta}-\mathbf{b}))$:

$$\pi(\boldsymbol{\beta}|p,q) \propto q^{-(p+1)} = q^{-(n/2-1)+1} = q^{-\frac{n}{2}},$$

Bayesian Linear Regression Model (cont.)

- Now substitute $q = \frac{1}{2}(\hat{\sigma}^2(n-k) + (\boldsymbol{\beta} - \mathbf{b})\mathbf{X}'\mathbf{X}(\boldsymbol{\beta} - \mathbf{b}))$ backing into $\pi(\boldsymbol{\beta}|p,q) \propto q^{-\frac{n}{2}}$ produces:

$$\pi(\boldsymbol{\beta}|\mathbf{X},\mathbf{y}) \propto \left[(n-k) + (\boldsymbol{\beta} - \mathbf{b})'\hat{\sigma}^{-2}\mathbf{X}'\mathbf{X}(\boldsymbol{\beta} - \mathbf{b})\right]^{-\frac{n}{2}}.$$

  which is the kernel of a multivariate-t distribution with $n-k$ degrees of freedom for $\boldsymbol{\beta} - \mathbf{b}$.

- This is valid only provided that the covariance matrix, $\mathbf{R} = \frac{(n-k)\hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}}{n-k-2}$, is positive definite.

Bayesian Linear Regression Model (cont.)

- The marginal distribution for σ^2 starts with the defining integral and separate terms in the exponent:

$$\pi(\sigma^2|\mathbf{X},\mathbf{y}) \propto \int_{-\infty}^{\infty} \sigma^{-n-1} \exp\left[-\frac{1}{2\sigma^2}(\hat{\sigma}^2(n-k) + (\boldsymbol{\beta}-\mathbf{b})'\mathbf{X}'\mathbf{X}(\boldsymbol{\beta}-\mathbf{b}))\right] d\boldsymbol{\beta}.$$

$$= \sigma^{-n-1} \exp\left[-\frac{1}{2\sigma^2}\hat{\sigma}^2(n-k)\right] \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{\beta}-\mathbf{b})'\mathbf{X}'\mathbf{X}(\boldsymbol{\beta}-\mathbf{b})\right] d\boldsymbol{\beta}.$$

The second exponential term is a k-dimensional kernel of a multivariate normal distribution providing the following substitution and simplification:

$$\pi(\sigma^2|\mathbf{X},\mathbf{y}) \propto \sigma^{-n-1} \exp\left[-\frac{1}{2\sigma^2}\hat{\sigma}^2(n-k)\right] (2\pi\sigma^2)^{\frac{k}{2}}$$

$$\propto (\sigma^2)^{-\frac{1}{2}(n-k+1)} \exp\left[-\frac{1}{2\sigma^2}\hat{\sigma}^2(n-k)\right].$$

Bayesian Linear Regression Model (cont.)

- The exponent on σ^2 can be rewritten:

$$-\frac{1}{2}(n-k+1) = -\frac{1}{2}(n-k) - \frac{1}{2} = -\frac{1}{2}(n-k) - 1 + \frac{1}{2} = -\frac{1}{2}(n-k-1) - 1$$

- This means we can apply the substitutions: $\alpha = \frac{1}{2}(n-k-1)$, $\beta = \frac{1}{2}\hat{\sigma}^2(n-k)$, so that:

$$\pi(\sigma^2|\mathbf{X},\mathbf{y}) \propto (\sigma^2)^{-\frac{1}{2}(n-k+1)} \exp\left[-\frac{1}{2\sigma^2}\hat{\sigma}^2(n-k)\right]$$

  becomes:

$$\pi(\sigma^2|\alpha,\beta) \propto (\sigma^2)^{-(\alpha+1)} \exp[-\beta/\sigma^2]$$

  which shows that this the inverse gamma: $\mathcal{IG}(\alpha,\beta)$.

- Standard IG PDF: $\mathcal{IG}(x|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}(x)^{-(\alpha+1)} \exp[-\beta/x], \quad 0 < x, \alpha, \beta.$

Bayesian Linear Regression Model (cont.)

- **Using Conjugate Priors**

$$p(\boldsymbol{\beta}, \sigma^2) = p(\boldsymbol{\beta}|\sigma^2)p(\sigma^2),$$

where the two component priors are specified by:

$$p(\boldsymbol{\beta}|\sigma^2) = (2\pi)^{-\frac{k}{2}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}}\exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{\beta} - \mathbb{B})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta} - \mathbb{B})\right],$$

$$p(\sigma^2) \propto \sigma^{-(a-k)}\exp\left[-\frac{b}{\sigma^2}\right].$$

and $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}$ by assumption.

Bayesian Linear Regression Model (cont.)

- Summarizing terms:

$n \times k$	size of the $\mathbf{X}$ matrix
$\boldsymbol{\beta}$	the unknown linear model coefficient vector
$\mathbb{B}$	prior mean vector for $\boldsymbol{\beta}$
σ^2	prior variance for $\boldsymbol{\beta}$, collected in the diagonal matrix $\boldsymbol{\Sigma}$
$\mathbf{b}$	$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$
$\hat{\sigma}^2$	$\dfrac{(\mathbf{y}-\mathbf{X}\mathbf{b})'(\mathbf{y}-\mathbf{X}\mathbf{b})}{(n-k)}$
$a-k$	assigned first parameter for σ^2 prior
b	assigned second parameter for σ^2 prior.

Bayesian Linear Regression Model (cont.)

- Introduce two completely unintuitive transformations figured out by Zellner (1971):

$$\tilde{\boldsymbol{\beta}} = (\boldsymbol{\Sigma}^{-1} + \mathbf{X}'\mathbf{X})^{-1}(\boldsymbol{\Sigma}^{-1}\mathbb{B} + \mathbf{X}'\mathbf{X}\mathbf{b})$$

$$\tilde{s} = 2b + \hat{\sigma}^2(n-k) + (\mathbb{B} - \tilde{\boldsymbol{\beta}})'\boldsymbol{\Sigma}^{-1}\mathbb{B} + (\mathbf{b} - \tilde{\boldsymbol{\beta}})'\mathbf{X}'\mathbf{X}\mathbf{b}.$$

which leads to:

$$\pi(\boldsymbol{\beta}, \sigma^2|\mathbf{X}, \mathbf{y}) \propto \sigma^{-n-a} \exp\left[-\frac{1}{2\sigma^2}\left(\tilde{s} + (\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})'(\boldsymbol{\Sigma}^{-1} + \mathbf{X}'\mathbf{X})(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})\right)\right]$$

which is easily integrated using the tricks from before produce the marginal posteriors of interest:

$$\pi(\boldsymbol{\beta}|\mathbf{X}, \mathbf{y}) \propto \left[\tilde{s} + (\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})'(\boldsymbol{\Sigma}^{-1} + \mathbf{X}'\mathbf{X})(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})\right]^{-\frac{n+a}{2}},$$

$$\pi(\sigma^2|\mathbf{X}, \mathbf{y}) \propto \sigma^{-n-a+k-1} \exp\left[-\frac{1}{2\sigma^2}\hat{\sigma}^2(n+a-k)\right].$$

which are correspondingly $t(n+a-k)$ and $\mathcal{IG}(n+a-k, \frac{1}{2}\hat{\sigma}^2(n+a-k))$.

Bayesian Linear Regression Model Summary

Setup	Prior	Posterior
Conjugate	$\boldsymbol{\beta}\|\sigma^2 \sim \mathcal{N}(\mathbb{B}, \sigma^2)$	$\boldsymbol{\beta}\|\mathbf{X} \sim t(n+a-k)$
	$\sigma^2 \sim \mathcal{IG}(a, b)$	$\sigma^2\|\mathbf{X} \sim \mathcal{IG}(n+a-k, \frac{1}{2}\hat{\sigma}^2(n+a-k))$
Uninformative	$\boldsymbol{\beta} \propto c$ over $[-\infty{:}\infty]$	$\boldsymbol{\beta}\|\mathbf{X} \sim t(n-k)$
	$\sigma^2 = \frac{1}{\sigma}$ over $[0{:}\infty]$	$\sigma\|\mathbf{X} \sim \mathcal{IG}(\frac{1}{2}(n-k-1), \frac{1}{2}\hat{\sigma}^2(n-k))$

recall that $\hat{\sigma}^2 = \frac{(\mathbf{y}-\mathbf{X}\mathbf{b})'(\mathbf{y}-\mathbf{X}\mathbf{b})}{(n-k)}$

Example: A Controversy in Education Policy

- Chubb and Moe argue that the institutional structure of the schools, especially the overhead democratic control, cause poor student performance.

- Meier and his coauthors: bureaucracy is an adaptation to poor performance and not the cause.

- Meier, Polinard and Wrinkle develop a linear model based on panel dataset from more than 1000 school districts for a seven-year period to test organizational theory and educational policy. The question asked is whether there is a causal relationship between bureaucracy and poor performance by public schools.

- The linear regression model proposed by the Meier et al. is affected by both serial correlation and heteroscedasticity. Meier et al. address these concerns through a set of six dummy variables for each year as well as through the use of weighted least squares.

Example: A Controversy in Education Policy

- The Meier et al. results are obtained by specifying diffuse normal priors on the unknown parameters. These are Gaussian normal specifications centered at zero with small precision. The model is summarized in "stacked" notation that shows the distributional assumptions (priors and likelihood):

$$
\begin{aligned}
Y[i] &\sim \mathcal{N}(\lambda[i], \sigma^2),\\
\lambda[i] &= \beta_0 + \beta_1 x_1[i] + \ldots + \beta_k x_k[i] + \epsilon[i]\\
\beta[i] &\sim \mathcal{N}(0.0, 10)\\
\epsilon[i] &\sim \mathcal{N}(0.0, \tau)\\
\tau &\sim \mathcal{G}(16, 6)
\end{aligned}
$$

Example: A Controversy in Education Policy

- The results are:

Explanatory Variables	Mean	SE	95% HPD
Constant Term	9.172	1.358	[6.510:11.840]
Low Income Students	-0.108	0.006	[-0.119:-0.097]
Teacher Salaries	0.073	0.053	[-0.035: 0.181]
Teacher Experience	-0.009	0.046	[-0.099: 0.082]
Gifted Classes	0.097	0.023	[0.054: 0.139]
Class Size	-0.220	0.052	[-0.322:-0.118]
State Aid Percentage	-0.002	0.004	[-0.010: 0.006]
Funding Per Student ($\times$1000)	0.065	0.174	[-0.276: 0.406]
Lag of Student Pass Rate	0.677	0.008	[0.661: 0.693]
Lag of Bureaucrats	-0.081	0.262	[-0.595: 0.431]

Posterior standard error of $\tau = 0.00072$

Example: A Controversy in Education Policy

- To expand on the Meier et al. model, it is possible to include non-sample information for the creation of the Bayesian prior drawn from Meier's previous work on school bureaucracy and school performance with Kevin Smith (1995).

- The Smith and Meier work includes data and inference on the impact of funding and other institutional variables on student achievement in Florida. These include district level data for all of the public schools in Florida. Smith and Meier note also that the Florida data provides a diverse group of students with constant measures over time.

- "Meier Priors:"

$$\begin{array}{rlrlrl}
\beta[0] & \sim \mathcal{N}(0.0, 10) & \beta[1] & \sim \mathcal{N}(-0.025, 10) & \beta[2] & \sim \mathcal{N}(0.0, 10) \\
\beta[3] & \sim \mathcal{N}(0.23, 10) & \beta[4] & \sim \mathcal{N}(0.615, 10) & \beta[5] & \sim \mathcal{N}(-0.068, 10) \\
\beta[6] & \sim \mathcal{N}(0.0, 10) & \beta[7] & \sim \mathcal{N}(-0.033, 10) & \beta[8] & \sim \mathcal{N}(0.299, 10) \\
\beta[9] & \sim \mathcal{N}(0.0, 10) & \beta[10] & \sim \mathcal{N}(0.0, 10) & \beta[11] & \sim \mathcal{N}(0.0, 10) \\
\beta[12] & \sim \mathcal{N}(0.0, 10) & \beta[13] & \sim \mathcal{N}(0.0, 10) & \beta[14] & \sim \mathcal{N}(0.0, 10) \\
\beta[15] & \sim \mathcal{N}(0.0, 10) & & & &
\end{array}$$

Example: A Controversy in Education Policy

Explanatory Variables	Mean	SE	95% HPD
Constant Term	4.799	2.373	[0.165: 9.516]
Low Income Students	-0.105	0.006	[-0.117:-0.094]
Teacher Salaries	0.382	0.099	[0.189: 0.575]
Teacher Experience	-0.066	0.046	[-0.156: 0.025]
Gifted Classes	0.096	0.021	[0.054: 0.138]
Class Size	0.196	0.191	[-0.180: 0.569]
State Aid Percentage	0.002	0.004	[-0.006: 0.010]
Funding Per Student ($\times$1000)	0.049	0.175	[-0.294: 0.392]
Lag of Student Pass Rate	0.684	0.008	[0.667: 0.699]
Lag of Bureaucrats	-0.042	0.261	[-0.557: 0.469]
Class Size $\times$ Teacher Salaries	-0.015	0.007	[-0.029:-0.002]

Posterior standard error of $\tau = 0.00071$

The Bayesian Linear Regression Model Accounting For Heteroscedasticity

- Instead of the usual assumption about the distribution of $\mathbf{y}$ given $\mathbf{X}$, we now assert that:

$$\mathbf{y}_i|\mathbf{X}_i \sim \mathcal{N}(\mathbf{X}_i\boldsymbol{\beta}, \sigma^2\omega_i),$$

  where $\omega = (\omega_1, \omega_2, \ldots, \omega_n)$ is a vector of unknown regression weights (parameters) which we can also organize along the diagonal of a $n \times n$ matrix Ω for convenience.

- The linear model is now defined as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \qquad \mathrm{Var}[\boldsymbol{\epsilon}] = \sigma^2\Omega,$$

  which implies from the conditional distribution of the $\mathbf{y}_i$ that $\boldsymbol{\epsilon}_i \sim N(0, \sigma^2\omega_i)$.

- Thus the likelihood function from before becomes:

$$L(\boldsymbol{\beta}, \sigma^2|\mathbf{X}, \mathbf{y}) = \sigma^{-n}|\Omega|^{-\frac{1}{2}} \exp\left[-\frac{1}{2\sigma^2}(\hat{\sigma}^2(n-k) + (\boldsymbol{\beta} - \mathbf{b})'\mathbf{X}'\Omega^{-1}\mathbf{X}(\boldsymbol{\beta} - \mathbf{b}))\right].$$

The Bayesian Linear Regression Model Accounting For Heteroscedasticity

- Use the uninformed priors:

$$p(\boldsymbol{\beta}) \propto c, \qquad \boldsymbol{\beta} \in [-\infty:\infty]$$

$$p(\sigma^2) = \frac{1}{\sigma}, \qquad \sigma^2 \in [0:\infty]$$

- Geweke (1993) suggests independent $\omega|\nu \sim \chi^2(df = \nu)$ for the weights, which is also expressible as the gamma distribution $\mathcal{G}(\nu/2, 1/2)$.

- The resulting joint posterior distribution is:

$$\pi(\boldsymbol{\beta}, \sigma^2, \Omega|\mathbf{X}, \mathbf{y}) \propto L(\boldsymbol{\beta}, \sigma^2|\mathbf{X}, \mathbf{y})p(\boldsymbol{\beta})p(\sigma^2)p(\Omega)$$

$$\propto \sigma^{-n-1}|\Omega|^{-\frac{\nu+3}{2}} \exp\left[-\frac{1}{2\sigma^2}\left((\hat{\sigma}^2(n-k) + (\boldsymbol{\beta}-\mathbf{b})'\mathbf{X}'\Omega^{-1}\mathbf{X}(\boldsymbol{\beta}-\mathbf{b}) + \nu\mathrm{tr}(\Omega)^{-1}\right)\right].$$

- This means that the posterior distribution of the residuals are independent Student's -t with $n-k$ degrees of freedom not $\boldsymbol{\epsilon}_i \sim N(0, \sigma^2\omega_i)$.

The Bayesian Linear Regression Model Accounting For Heteroscedasticity

- Conditional posterior distributions:

$$\pi(\boldsymbol{\beta}|\sigma^2,\Omega) \propto \exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{\beta}-\mathbf{b}^*)'\mathbf{X}'\Omega^{-1}\mathbf{X}(\boldsymbol{\beta}-\mathbf{b}^*)\right],$$

where $\mathbf{b}^* = (\mathbf{X}'\Omega\mathbf{X})^{-1}\mathbf{X}'\Omega\mathbf{y}$. So $\boldsymbol{\beta}|\sigma^2,\Omega \sim \mathcal{N}(\mathbf{b}^*, \sigma^2(\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1})$.

$$\pi(\sigma^2|\Omega) \propto (\sigma^2)^{-\frac{n+1}{2}} \exp\left[-\frac{1}{2\sigma^2}\hat{\sigma}^{2*}\right],$$

where $\hat{\sigma}^{2*} = (\mathbf{y}-\mathbf{X}\mathbf{b})'\Omega^{-1}(\mathbf{y}-\mathbf{X}\mathbf{b})$. So $\sigma^2|\boldsymbol{\beta},\Omega \sim \mathcal{IG}(\sigma^2|(n-1)/2, \hat{\sigma}^{2*}/2$. The conditional posterior distribution of Ω:

$$\pi(\Omega|\boldsymbol{\beta},\sigma^2) \propto |\Omega|^{-\frac{\nu+3}{2}} \exp\left[-\frac{1}{2}\left(\hat{\sigma}^{2*}/\sigma^2 + \nu\mathrm{tr}(\Omega)^{-1}\right)\right]$$

An individual diagonal element of Ω is conditionally distributed $\omega_i|\boldsymbol{\beta},\sigma^2 \propto \omega_i^{-\frac{\nu+3}{2}} \exp\left[-(u_i^2/2\sigma^2 + \nu/2)/\omega_i\right]$ where $u_i = y_i - \mathbf{x}_i\boldsymbol{\beta}$. Which gives an inverse gamma PDF with parameters $\frac{\nu+1}{2}$ and $(u_i^2/\sigma^2 + \nu)/2$.

The Bayesian Linear Regression Model Accounting For Heteroscedasticity

- The Gibbs sampler iterates at the $j+1$ step according to:

$$\begin{aligned}\boldsymbol{\beta}^{[j+1]} &\sim \pi(\boldsymbol{\beta}|\sigma^{2[j]}, \Omega^{[j]})\\ \sigma^{2[j+1]} &\sim \pi(\sigma^2|\Omega^{[j+1]})\\ \Omega^{[j+1]} &\sim \pi(\Omega|\boldsymbol{\beta}^{[j+1]}, \sigma^{2[j+1]})\end{aligned}$$

- The $\boldsymbol{\beta}$ vector and the Ω matrix (which is really a vector's worth of information) can be sampled individually ($\boldsymbol{\beta}_1^{[j]}, \ldots, \boldsymbol{\beta}_k^{[j]}, \omega_1^{[j]}, \ldots, \omega_n^{[j]}$) or as a block.

War In Ancient China

- This is conflict data from West Asia for events taking place between 2700 BC to 722 BC.
- Cioffi-Revilla and Lai (1995, 2001) coded documents from multiple epigraphic and archaeological sources on war and politics in ancient China covering the Xia (Hsia), Shang, and Western Zhou (Chou) periods.
- These data ($n = 104$ conflicts) are available via the Murray Archive.
- Cioffi-Revilla and Lai use the modern *Long-Range Analysis of War* definitions.

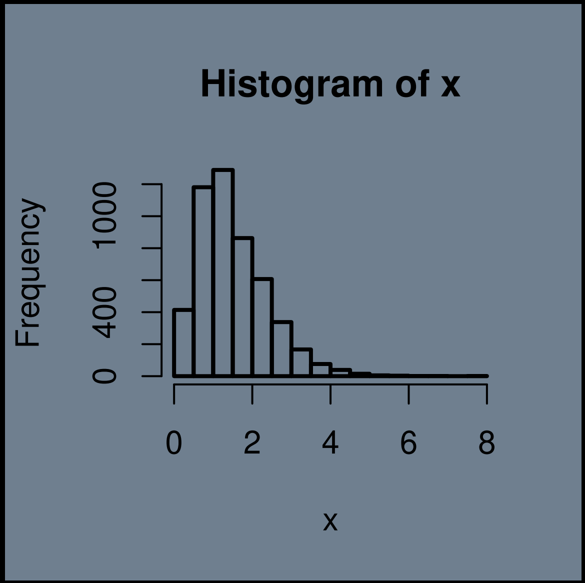
War In Ancient China

- OUTCOME VARIABLE: combination of two of the coded variables: *Political Level* (1 for internal war, 2 for interstate war) and ten times *Political Complexity* (governmental level of the warring parties).
- We are looking to explain the *political scope* of conflicts in terms of governmental units affected.
- EXPLANATORY VARIABLES: *Extent* (number of belligerents), *Diverse* (number ethnic groups participating as belligerents), *Alliance* (total number of alliances among belligerents), *Dyads* (number of alliance pairs), *Tempor* (type of war: protracted rivalry, integrative conquest, disintegrative/fracturing conflict, sporadic event), and *Duration* measured in years.
- No intercept (zero levels do not make sense here).

War In Ancient China, Model Features

- Specify uninformed priors: $p(\boldsymbol{\beta}) \propto c$ and $p(\sigma^2) = \frac{1}{\sigma}$.
- Furthermore, fix $\nu = 12$ (recall that ν is the df parameter on the prior for the ω_is. This improves model fit, and also make the Gibbs sampler more efficient.
- Gibbs sampler run for $10,000$ iterations and throwing away the first $5,000$, giving $5,000$ samples for each of the $\boldsymbol{\beta}$'s, σ^2, and Ω.

```
> x <- rgamma(5000,3,2)
> mean(x)
[1] 1.5079
> var(x)
[1] 0.7483
> hist(x)
```



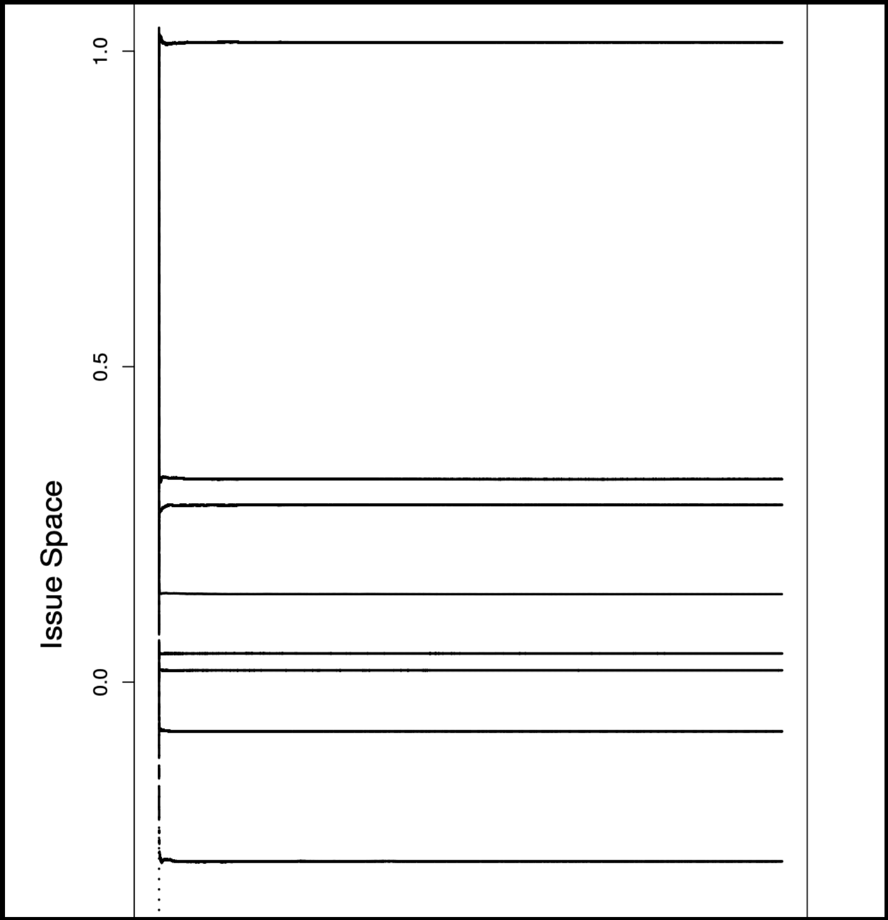
War In Ancient China

Table 3: Heteroscedastic Model, Ancient Chinese Conflicts

Explanatory Variables	Mean	SE	95% HPD
EXTENT	1.4156	0.1299	[1.1610: 1.6703]
DIVERSE	0.2279	0.1061	[0.0199: 0.4358]
ALLIANCE	-0.4597	0.1182	[-0.6914:-0.2280]
DYADS	-0.8860	0.1192	[-1.1197:-0.6524]
TEMPOR	0.0889	0.0504	[-0.0100: 0.1877]
DURATION	-0.0276	0.0460	[-0.1177:0.0625]

Mean of $\sigma = 0.0736$

War In Ancient China, MCMC Running Means



Example: Poverty Among the Elderly, Europe

- Governments often worry about the economic condition of senior citizens for political and social reasons.
- Typically in a large industrialized society, a substantial portion of these people obtain the bulk of their income from government pensions.
- An important question is whether there is enough support through these payments to provide subsistence above the poverty rate.
- To see if this is a concern, the European Union (EU) looked at this question in 1998 for the (then) 15 member countries with two variables:
  1. the median (EU standardized) income of individuals age 65 and older as a percentage of the population age 0–64,
  2. the percentage with income below 60% of the median (EU standardized) income of the national population.

Example: Poverty Among the Elderly, Europe

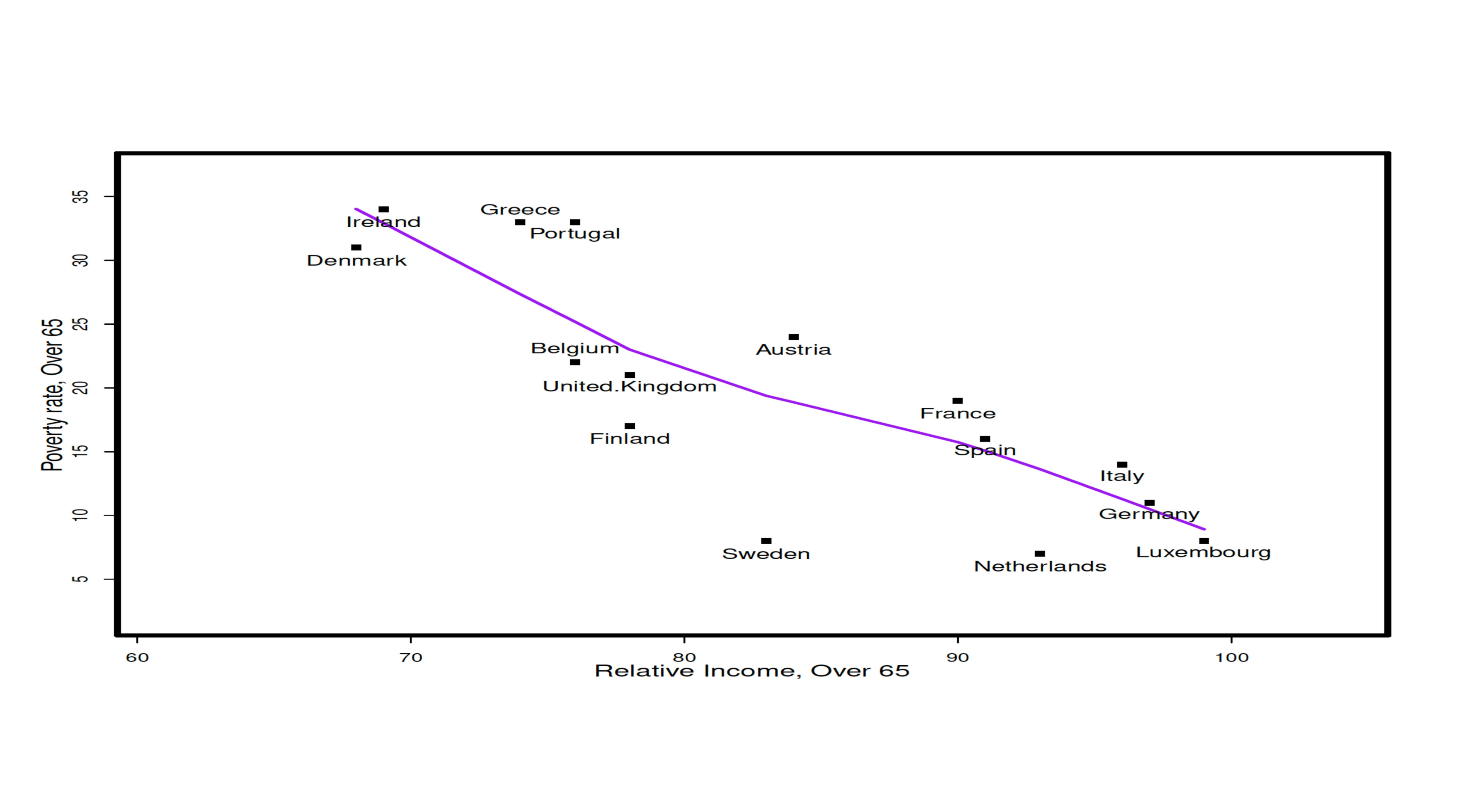
- The data from the European Household Community Panel Survey are:

Nation	Relative Income	Poverty Rate
Netherlands	93.00	7.00
Luxembourg	99.00	8.00
Sweden	83.00	8.00
Germany	97.00	11.00
Italy	96.00	14.00
Spain	91.00	16.00
Finland	78.00	17.00
France	90.00	19.00
United.Kingdom	78.00	21.00
Belgium	76.00	22.00
Austria	84.00	24.00
Denmark	68.00	31.00
Portugal	76.00	33.00
Greece	74.00	33.00
Ireland	69.00	34.00

Lowess Smooth: Poverty Among the Elderly, Europe

```
eu.pov <- read.table("http://jeffgill.org/files/jeffgill/files/inc.pov_.dat_.txt",
    row.names=1)
names(eu.pov) <- c("relative income", "poverty rate")
eu.pov <- eu.pov[-1,]

par(mfrow=c(1,1),mar=c(4,4,2,2),lwd=5)
plot(eu.pov,pch=15,xlab="",ylab="",ylim=c(2,37),xlim=c(61,104))
lines(lowess(eu.pov),col="purple",lwd=3)
text.loc <- cbind(eu.pov[,1],(eu.pov[,2]-1))
text.loc[14,2] <- text.loc[14,2] +2
text.loc[10,2] <- text.loc[10,2] +2
text(text.loc,dimnames(eu.pov)[[1]],cex=1.2)
mtext(side=1,cex=1.3,line=2,"Relative Income, Over 65")
mtext(side=2,cex=1.3,line=2,"Poverty rate, Over 65")
```

Example: Poverty Among the Elderly, Europe



Simple Linear Bayesian Specification: Poverty Among the Elderly, Europe

- For basic regressions the `arm` package by Andrew Gelman, Yu-Sung Su, Daniel Lee, and Aleks Jakulin works nicely.
- Load the package and run a linear regression at the defaults:
- A useful function in this package is `bayesglm`, which is functionally equivalent to the regular `glm`.

```
library(arm)
eu.pov.out <- bayesglm(eu.pov[,2] ~ eu.pov[,1], prior.mean=0, prior.df=1,
    prior.df.for.intercept=1, prior.mean.for.intercept=0)
summary(eu.pov.out)
```

Simple Linear Bayesian Specification: Poverty Among the Elderly, Europe

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-12.218   -3.306    1.488    3.929    7.429

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  83.6859    12.2525   6.830 1.21e-05
eu.pov[, 1]  -0.7647     0.1458  -5.246 0.000158

(Dispersion parameter for gaussian family taken to be 31.48377)

    Null deviance: 1275.73  on 14  degrees of freedom
Residual deviance:  409.29  on 13  degrees of freedom
AIC: 98.164
```

Non-Bayesian Specification: Poverty Among the Elderly, Europe

```
x.y.fit <- lm(eu.pov[,2] ~ eu.pov[,1])
summary(x.y.fit)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  83.6928    12.2526   6.831 1.21e-05
eu.pov[, 1]  -0.7647     0.1458  -5.246 0.000158

Residual standard error: 5.611 on 13 degrees of freedom
Multiple R-Squared: 0.6792,     Adjusted R-squared: 0.6545
F-statistic: 27.52 on 1 and 13 DF,  p-value: 0.0001580
```

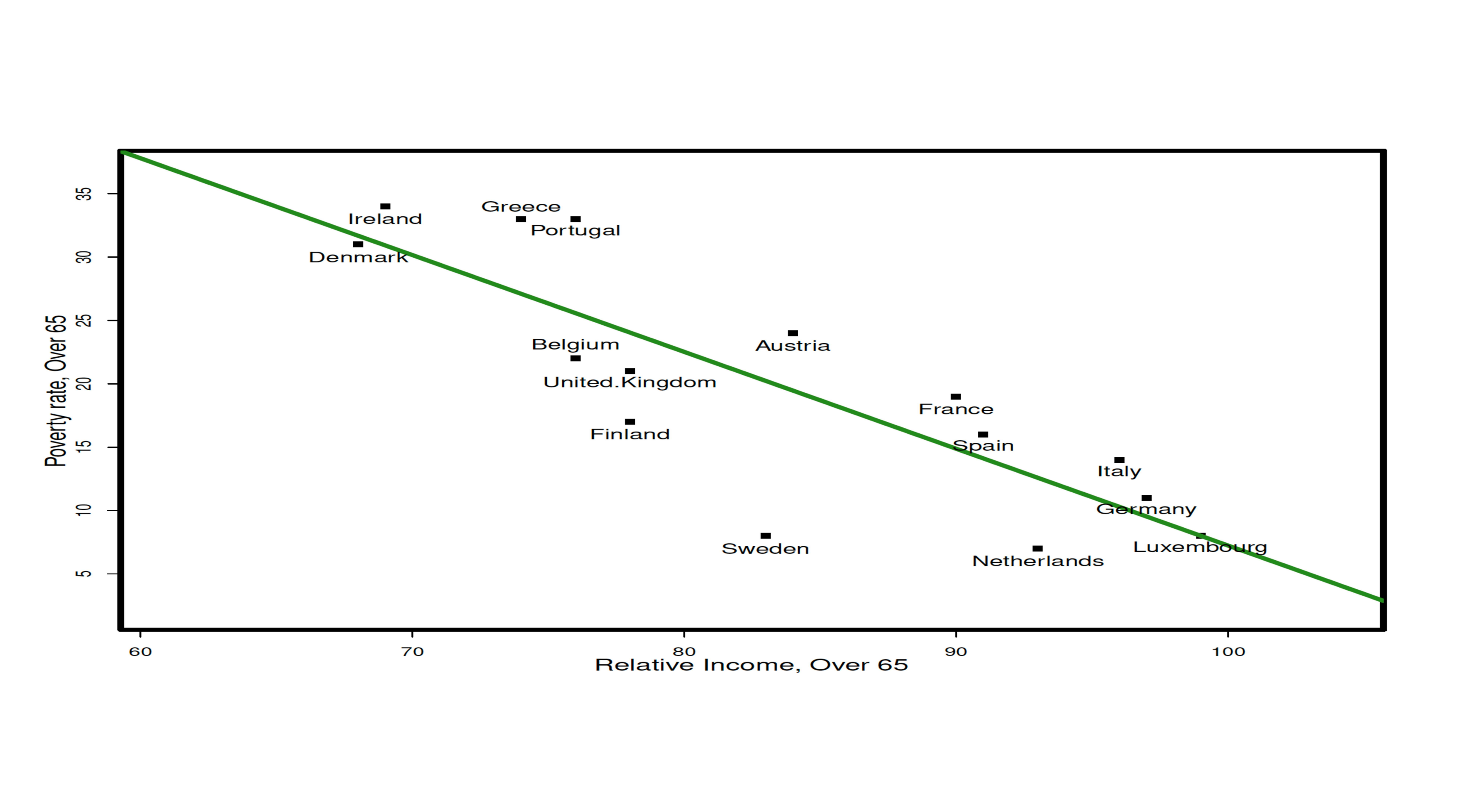
Prior Sensitivity: Poverty Among the Elderly, Europe

```
eu.pov.out2 <- bayesglm(eu.pov[,2] ~ eu.pov[,1], prior.mean=50, prior.df=100,
    prior.df.for.intercept=-50, prior.mean.for.intercept=100)
summary(eu.pov.out2)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  83.6578    12.2526   6.828 1.21e-05
eu.pov[, 1]  -0.7642     0.1458  -5.243 0.000159

(Dispersion parameter for gaussian family taken to be 31.48379)

Null deviance: 1275.73  on 14  degrees of freedom
Residual deviance:  409.29  on 13  degrees of freedom
AIC: 98.164
```

Graphing: Poverty Among the Elderly, Europe

```
par(mfrow=c(1,1),mar=c(4,4,2,2),lwd=5)
plot(eu.pov,pch=15,xlab="",ylab="",ylim=c(2,37),xlim=c(61,104))
abline(eu.pov.out2$coefficients,col="forest green")
text.loc <- cbind(eu.pov[,1],(eu.pov[,2]-1))
text.loc[14,2] <- text.loc[14,2] +2
text.loc[10,2] <- text.loc[10,2] +2
text(text.loc,dimnames(eu.pov)[[1]],cex=1.2)
mtext(side=1,cex=1.3,line=2,"Relative Income, Over 65")
mtext(side=2,cex=1.3,line=2,"Poverty rate, Over 65")
```

Graphing: Poverty Among the Elderly, Europe



ANES Data from 2012

- Suppose we are interested in survey "mode effects" such as whether face-to-face versus internet responses differ.
- To analyze the potential consequences of mode effects on the uncertainty that surrounds public opinion data, we examine the American National Election Studies 2012 Time Series Study.
- The ANES 2012 study is the 29th installment in a longstanding series of election studies that go back to 1948.
- The 2012 edition differs from its predecessors significantly, and lends itself exceptionally well to our analysis because it is the first ANES study that implements a dual-mode design by incorporating a traditional ANES face-to-face sample as well as a separate sample interviewed on the Internet.
- Both samples were independently drawn and data collection was conducted independently in the two modes as well.

ANES Data from 2012

- Download the data in `R`(or go to the class page and click):

```
f2f.anes <-
    read.table("http://jeffgill.org/files/jeffgill/files/f2f.anes_.1.dat_.txt")
inet.anes <-
    read.table("http://jeffgill.org/files/jeffgill/files/inet.anes_.1.dat_.txt")

dim(f2f.anes)
[1] 2054   54
dim(inet.anes)
[1] 3860   54
```

ANES Data from 2012

- Look at the variables:

```
names(f2f.anes)
 [1] "weight_ftf"            "presvote2012_x"        "gender_respondent_x"   "dem_edu"
 [5] "dem_birthyr"           "dem_racecps_white"     "dem_racecps_black"     "dem_hisp"
 [9] "dem_marital"           "pid_x"                 "libcpre_self"          "interest_attention"
[13] "candrel_dpc"           "cses_econ"             "ftcasi_illegal"        "egal_worryless"
[17] "interest_voted2008"    "prmedia_atinews"       "prmedia_wktvnws"       "prmedia_attvnews"
[21] "prmedia_atpprnews"     "prevote_regist_addr"   "prevote_intpreswho"    "prevote_intpresst"
[25] "congapp_job"           "presapp_track"         "presapp_job"           "presapp_econ"
[29] "presapp_foreign"       "presapp_health"        "presapp_war"           "ft_dpc"
[33] "ft_rpc"                "ft_dvpc"               "ft_rvpc"               "ft_dem"
[37] "ft_rep"                "finance_finfam"        "finance_finpast"       "finance_finnext"
[41] "health_insured"        "health_2010hcr"        "health_self"           "health_smokeamt"
[45] "likelypct_whatpct1"    "campfin_limcorp"       "ineq_incgap_x"         "effic_complicrev"
[49] "effic_carerev"         "econ_ecpast_x"         "econ_unpast_x"         "preswin_dutychoice_x"
[53] "war_terror"            "gun_control"
```

Some 2 and 3 Point Items

Variable Name	Variable Description
`cses_econ`	State of Economy
`campfin_banads`	Ban Corporate/Union Ads
`ineqinc_ineqreduc`	Gov't Reducing Income Inequality
`econ_ecpast`	National Economy: Better/Worse
`econ_unpast`	Unemployment: Better/Worse
`mip_prob2pty`	Best Party to Handle MIP #2
`iran_nuksite`	Bombing Iran's Nuclear Sites
`auth_consid`	Important for Child: Considerate or Well-Behaved
`finance_finpast`	Better/Worse Off Than Year Ago
`interest_wherevote`	Know Where to Vote
`tea_suppln`	Tea Party: Leaning Towards Support/Opposition
`preswin_dutyst`	Voting as Duty: Feeling Strength
`fedspend_schools`	Public Schools: More or Less Spending

Some 4 and 5 Point Items

Variable Name	Variable Description
`resent_deserve`	Blacks: Gotten Less Than Deserved
`cses_govtact`	Gov't Reducing Income Inequality
`resent_try`	Blacks: Must Try Harder
`ecperil_payhlthcst`	Able to Pay Health Care
`egal_worryless`	Worry Less About Equality
`ctrait_dpccare`	Dem Cand: Cares About People Like Me
`ecblame_pres`	Blame President for Economy
`ctrait_rpclead`	Rep Cand: Strong Leadership
`ctrait_dpcmoral`	Dem Cand: Is Moral
`ctrait_rpcmoral`	Rep Cand: Is Moral
`finance_finpast_x`	Better/Worse Off Than Year Ago (5 point scale)
`likelypct_howlikvt1`	Likelihood of Voting
`trustgov_trustgstd`	Trust Gov't in Washington
`cses_diffvote`	Vote Makes a Difference
`gayrt_discstd_x`	Favor Laws Against Gays/Lesbian Job Discrim
`egal_equal`	Provide Equal Opportunities

Some 7 and 11 Point Items

Variable Name	Variable Description
cses_dptylike	Democratic Party Like (0-10)
cses_rptylike	Republican Party Like (0-10)
cses_rpclike	Republican Pres Cand Like (0-10)
cses_rptyleft	Left-Right Republican Party (0-10)
cses_dpclike	Democratic Pres Cand Like (0-10)
cses_rpclike	Republican Pres Cand Like (0-10)
wpres_gdbd_x	Good/Bad: Woman Pres
women_bond_x	Working Mother's Bond with Child
abort_sex_x	Legal Abortion to Select Child Gender
budget_deficit_x	Favor Reducing Budget Deficit
scourt_remove_x	Possibility to Remove Sup Court Judges

Some 101 Point Items

Variable Name	Variable Description
`ftgr_unions`	FT: Unions
`ftgr_fedgov`	FT: Federal Government
`ftcasi_illegal`	FT: Illegal Immigrants
`ft_dpc`	FT: Democratic Presidential Candidate
`ft_rpc`	FT: Republican Presidential Candidate
`ft_dvpc`	FT: Democratic Vice Presidential Candidate
`ft_rvpc`	FT: Republican Vice Presidential Candidate
`ft_dpcsp`	FT: Spouse of Democratic Presidential Candidate

ANES Data, Linear Model, Face-to-Face

- ```
  f2f.linear.out <- bayesglm(ft_dpc ~ health_insured + interest_voted2008
      + gun_control + congapp_job + dem_racecps_black,
      prior.mean=0, prior.df=1, prior.df.for.intercept=0,
      prior.mean.for.intercept=1, data=f2f.anes, weights=weight_ftf)
  summary(f2f.linear.out)
  ```

- Face-to-face results:

	Estimate	Std. Error	t value	$\Pr(>\lvert t\rvert)$
(Intercept)	84.9985	4.2595	19.955	<2e-16
health_insured	4.0757	1.7892	2.278	0.0228
interest_voted2008	-0.2391	1.4213	-0.168	0.8664
gun_control	-9.0931	0.6507	-13.974	<2e-16
congapp_job	-9.9348	1.4983	-6.631	4.26e-11
dem_racecps_black	30.1378	1.8654	16.157	<2e-16

(Dispersion parameter for gaussian family taken to be 805.6735)

Null deviance: 2173744 on 2053 degrees of freedom

Residual deviance: 1650019 on 2048 degrees of freedom

AIC: 20579
  ```

ANES Data, Linear Model, Internet

- ```
  inet.linear.out <- bayesglm(ft_dpc ~ health_insured + interest_voted2008
      + gun_control + congapp_job + dem_racecps_black,
      prior.mean=0, prior.df=1, prior.df.for.intercept=0,
      prior.mean.for.intercept=1, data=inet.anes, weights=weight_web)
  summary(inet.linear.out)
  ```

- Internet subset results:

	Estimate	Std. Error	t value	Pr($>\|t\|$)
(Intercept)	82.8155	3.2436	25.532	<2e-16
health_insured	2.8211	1.3852	2.037	0.0418
interest_voted2008	0.4189	1.1946	0.351	0.7259
gun_control	-11.5795	0.5036	-22.995	<2e-16
congapp_job	-8.5299	1.1919	-7.156	9.87e-13
dem_racecps_black	34.2749	1.4854	23.074	<2e-16

(Dispersion parameter for gaussian family taken to be 904.4944)

Null deviance: 4674189 on 3859 degrees of freedom

Residual deviance: 3485921 on 3854 degrees of freedom

AIC: 38634
  ```

Logit Model for Survey Responses in Scotland

- These data come from the British General Election Study, Scottish Election Survey, 1997 (ICPSR Study Number 2617).
- These data contain 880 valid cases, each from an interview with a Scottish national after the election.
- Our outcome variable of interest is their party choice in the UK general election for Parliament where we collapse all non-Conservative party choices (abstention, Labour, Liberal Democrat, Scottish National, Plaid Cymru, Green, Other, Referendum) to one category, which produces 104 Conservative votes.
- For a logit model the `prior.scale` is `2.5`, and for a probit model the prior scale is `2.5*1.6` (typical assumptions are `1` and `1.6`).

Logit Model for Survey Responses in Scotland, Explanatory Variables

- `POLITICS`, which asks how much interest the respondent has in political events (increasing scale: none at all, not very much, some, quite a lot, a great deal).
- `READPAP`, which asks about daily morning reading of the newspapers (yes=1 or no=0).
- `PTYTHNK`, how strong that party affiliation is for the respondent (categorical by party name).
- `IDSTRNG` (increasing scale: not very strong, fairly strong, very strong).
- `TAXLESS` asks if "it would be better if everyone paid less tax and had to pay more towards their own healthcare, schools and the like" (measured on a five point increasing Likert scale).
- `DEATHPEN` asks whether the UK should bring back the death penalty ((measured on a five point increasing Likert scale).
- `LORDS` queries whether the House of Lords should be reformed (asked as *remain as is* coded as zero and *change is needed* coded as one).
- `SCENGBEN` asks how economic benefits are distributed between England and Scotland with the choices: England benefits more $= -1$, neither/both lose $= 0$, Scotland benefits more $= 1$.

Logit Model for Survey Responses in Scotland, Explanatory Variables

- INDPAR asks which of the following represents the respondent's view on the role of the Scottish government in light of the new parliament: (1) Scotland should become independent, separate from the UK and the European Union, (2) Scotland should become independent, separate from the UK but part of the European Union, (3) Scotland should remain part of the UK, with its own elected parliament which has some taxation powers, (4) Scotland should remain part of the UK, with its own elected parliament which has no taxation powers, and (5) Scotland should remain part of the UK without an elected parliament.

- SCOTPREF1 asks "should there be a Scottish parliament within the UK? (yes=1, no=0).

- RSEX, the respondent's sex.

- RAGE, the respondent's age.

- RSOCCLA2, the respondents social class (7 category ascending scale).

- TENURE1, whether the respondent rents (0) or owns (1) their household.

- PRESB, a categorical variable for church affiliation, measurement of religion is collapsed down to one for the dominant historical religion of Scotland (Church of Scotland/Presbyterian) and zero otherwise and designated

Logit Model for Survey Responses in Scotland

- Run a probit model for the conservative/not-conservative outcome with these covariates:

- Results given across two slides...

```
scot.mat <- read.table(
    "http://jeffgill.org/files/jeffgill/files/scotland.dat_.txt",
    sep=",",header=TRUE)
scot.out <- bayesglm(VOTE ~ POLITICS + READPAP + PTYTHNK + IDSTRNG + TAXLESS
                                + DEATHPEN + LORDS + SCENGBEN + SCOPREF1
                                + RSEX + RAGE + RSOCCLA2 + TENURE1 + PRESB
                                + IND.PAR,
                                data=scot.mat, family=binomial(link="logit"))
summary(scot.out)
```

- Presenting in non-Bayesian fashion...

Logit Model for Survey Responses in Scotland, Results

	Estimate	Std. Error	z value	$\Pr(>\|z\|)$
(Intercept)	-1.520579	1.026848	-1.481	0.13865
POLITICS	0.369950	0.141111	2.622	0.00875
READPAP	0.412143	0.333149	1.237	0.21604
PTYTHNK	-0.987682	0.171090	-5.773	7.79e-09
IDSTRNG	0.384849	0.141462	2.721	0.00652
TAXLESS	0.219942	0.134571	1.634	0.10218
DEATHPEN	0.168560	0.104639	1.611	0.10721
LORDS	-0.798360	0.287346	-2.778	0.00546
SCENGBEN	0.586385	0.200590	2.923	0.00346
SCOPREF1	-1.651397	0.339557	-4.863	1.15e-06
RSEX	0.705948	0.310589	2.273	0.02303
RAGE	0.019811	0.007722	2.566	0.01030
RSOCCLA2	-0.246593	0.108432	-2.274	0.02296
TENURE1	0.851254	0.336873	2.527	0.01151
PRESB	-0.225592	0.304528	-0.741	0.45882
IND.PAR	0.568041	0.349059	1.627	0.10366

Logit Model for Survey Responses in Scotland, Results

```
(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 639.38  on 879  degrees of freedom
Residual deviance: 340.64  on 864  degrees of freedom
AIC: 372.64

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.2572  -0.2931  -0.1577  -0.0627   3.3411
```

Objects Produced

```
names(scot.out)
"coefficients"  "residuals"      "fitted.values"  "effects"            "R"
"rank"          "qr"             "family"         "linear.predictors"  "deviance"
"aic"           "null.deviance"  "iter"           "weights"            "prior.weights"
"df.residual"   "df.null"        "y"              "converged"          "boundary"
"prior.mean"    "prior.scale"    "prior.df"       "prior.sd"           "dispersion"
"model"         "call"           "formula"        "terms"              "data"
"offset"        "control"        "method"         "contrasts"          "xlevels"
"keep.order"    "drop.baseline"
```

Percent Predicted Correctly

```
scot.pred <- scot.out$fitted.values
scot.pred[scot.pred < 0.5] <- 0
scot.pred[scot.pred > 0.5] <- 1
table(scot.pred,scot.mat$VOTE)

scot.pred   0   1
        0 750  49
        1  26  55

sum(diag(table(scot.pred,scot.mat$VOTE)))/nrow(scot.mat)
[1] 0.9147727
```

Percent Predicted Correctly

```
mean(scot.pred)
[1] 0.09204545
scot.pred <- scot.out$fitted.values
scot.pred[scot.pred < mean(scot.pred)] <- 0
scot.pred[scot.pred > mean(scot.pred)] <- 1
table(scot.pred,scot.mat$VOTE)

scot.pred   0   1
        0 667  12
        1 109  92

sum(diag(table(scot.pred,scot.mat$VOTE)))/nrow(scot.mat)
[1] 0.8625
```

Application: Poisson Model of Military Coups.

- Sub-Saharan Africa has experienced a disproportionately high proportion of regime changes due to the military takeover of government for a variety of reasons, including ethnic fragmentation, arbitrary borders, economic problems, outside intervention, and poorly developed governmental institutions.

- These data, selected from a larger set given by Bratton and Van De Walle (1994), look at potential causal factors for counts of military coups (ranging from 0 to 6 events) in 33 sub-Saharan countries over the period from each country's colonial independence to 1989.

- Included are 99 variables describing governmental, economic, and social conditions for the 47 cases. Also provided are data from 106 presidential and 185 parliamentary elections, including information about parties, turnout, and political openness.

- Seven explanatory variables are chosen here to model the count of military coups: **Military Oligarchy** (the number of years of this type of rule); **Political Liberalization** (0 for no observable civil rights for political expression, 1 for limited, and 2 for extensive); **Parties** (number of legally registered political parties); **Percent Legislative Voting**; **Percent Registered Voting**; **Size** (in one thousand square kilometer units); and **Population** (given in millions).

Application: Poisson Model of Military Coups.

- The focus here is an outcome variable included in Bratton and Van De Walle's work (1994, p. 479), but not featured as a modeled result: regime change through military coups.

- This is a well-studied issue (Bienen 1979; Decalo 1976a and 1976b; Feit 1968; Jackman, et al. 1986; Johnson, et al. 1984), but not necessarily so from a *statistical perspective*.

- **Military Coups** is operationalized as the successful number of military coups for a country over the period from independence to 1989 (ranging from zero to six events). This outcome variable is defined only over a positive integer sample space and therefore requires a generalized linear model link function appropriate to counts.

Application: Poisson Model of Military Coups.

- A generalized linear model for these data with the Poisson link function is specified as:

$$g^{-1}(\boldsymbol{\theta}) = g^{-1}(\mathbf{X}\boldsymbol{\beta}) = \exp\left[\mathbf{X}\boldsymbol{\beta}\right] = \mathbb{E}[\mathbf{Y}] = \mathbb{E}[\textbf{Military Coups}].$$

- In this specification, the systematic component is $\mathbf{X}\boldsymbol{\beta}$, the stochastic component is $\mathbf{Y} = \textbf{Military Coups}$, and the link function is $\boldsymbol{\theta} = \log(\boldsymbol{\mu})$.

- We can re-express this model by moving the link function to the left-hand side exposing the linear predictor: $g(\boldsymbol{\mu}) = \log(\mathbb{E}[\mathbf{Y}]) = \mathbf{X}\boldsymbol{\beta}$ (although this is now a less intuitive form for understanding the outcome variable).

- Left out of the discussion above are the prior distributions: again we will assume Student's-t for the coefficients.

- Get the data:

```
africa.dat <- read.table("http://jeffgill.org/files/jeffgill/files/africa.dat__1.txt")
```

Looking At the Variables

```
names(africa.dat)
 [1] "CNTRYCDE" "REGION"   "POP"      "SIZE"     "COLONIAL" "BORDERS"
 [7] "DATEINDP" "GNPPC"    "GROWTH"   "ENERGY"   "MANUF"    "AGLABOR"
[13] "INFLATN"  "ADJPROGS" "AIDFLOWS" "DEBT"     "SERVICE"  "URBAN"
[19] "ETHNIC"   "PCTTRAD"  "PCTCATH"  "PCTPROT"  "PCTMUSL"  "RADIOS89"
[25] "TELEV89"  "PARTY75"  "PARTY89"  "PARTY93"  "UNION89"  "BUSIN75"
[31] "BUSIN89"  "CHURSCH"  "CHURMED"  "DAILY75"  "DAILY89"  "DAILY93"
[37] "PERIOD75" "PERIOD89" "PERIOD93" "PUBLIS75" "PUBLIS89" "PUBLIS93"
[43] "CONSTIT"  "DICTATOR" "MILITARY" "PLBSCTRY" "COMPTIVE" "SETTLER"
[49] "POLYACHY" "NUMREGIM" "REGCHANG" "MILTCOUP" "NUMELEC"  "NUMLEGS"
[55] "NUMPRES"  "COMPELEC" "YRSCOMP"  "DATELAST" "PARTYLEG" "PCTSEAT"
[61] "MEANSEAT" "PCTTURN"  "MEANTURN" "PCTVOTE"  "MEANVOTE" "MEANPRES"
[67] "PROTFREQ" "PROTEST"  "REPRESS"  "POLLIB"   "MANIP"    "OPPCOH"
[73] "GOVTCOH"  "MILTROLE" "INTLPR"   "NATCON"   "PRESELEC" "PRESDATE"
[79] "PRESCAND" "PRESVOTS" "PRESTURN" "LEGSELEC" "LEGSDATE" "LEGSCAND"
[85] "LEGSPRTY" "LEGSEATS" "LEGSTURN" "FREEFAIR" "INCBOUST" "LOSERACC"
[91] "PROTBEG"  "POLLIBEG" "TRANSEND" "LIBCHANG" "BACKSLID" "OUTCOME"
[97] "DEMCHANG" "DEMLEVEL"
```

Results: Poisson Model of Military Coups.

- The R language GLM call for this model is:

```
africa.out <- bayesglm(MILTCOUP ~ MILITARY + POLLIB + PARTY93 + PCTVOTE + PCTTURN
                       + SIZE*POP + NUMREGIM*NUMELEC, family=poisson, data=africa.dat)
```

	Coef	Std.Err.	0.95 Lower	0.95 Upper	CIs:ZE+RO
(Intercept)	1.739	1.177	-0.566	4.045	|--o--|
MILITARY	0.113	0.038	0.038	0.187	|o|
POLLIB	-0.465	0.298	-1.049	0.120	|o|
PARTY93	0.024	0.011	0.003	0.045	|o|
PCTVOTE	0.039	0.017	0.006	0.072	|o|
PCTTURN	-0.025	0.011	-0.047	-0.002	|o|
SIZE	-0.001	0.000	-0.002	0.000	|o|
POP	-0.055	0.025	-0.105	-0.006	|o|
SIZE:POP	0.000	0.000	0.000	0.000	|o|
NUMREGIM	-0.427	0.363	-1.138	0.285	|o|
NUMELEC	-0.272	0.150	-0.566	0.022	|o|
NUMREGIM:NUMELEC	0.104	0.050	0.007	0.201	|o|

Comments: Poisson Model of Military Coups.

- Note that the two interaction terms are specified by using the multiplication character. The iteratively weighted least squares algorithm converged in only four iterations using Fisher scoring, and the results are provided in the table.

- The model appears to fit the data quite well:

  - an improvement from the null deviance of 62 on 32 degrees of freedom to a residual deviance of 7.5 on 21 degrees of freedom

  - evidence that the model does not fit would be supplied by a model deviance value in the tail of a χ^2_{n-k} distribution

  - and nearly all the coefficients have 95% confidence intervals bounded away from zero and therefore appear reliable in the model.