

106th HOUSE EXAMPLE

. d

Contains data from C:\docs_c_summer_course\hdmg106_2009_fixed.dta

obs: 432
vars: 11 22 Oct 2009 12:17
size: 16,848 (99.9% of memory free)

variable name	storage type	display format	value label	variable label
bush00	byte	%8.0g		bush % 2000
gore00	byte	%8.0g		gore % 2000
black00	float	%9.0g		black pct 2000
south	byte	%8.0g		1=SOUTH
hispanic00	float	%9.0g		hispanic 2000
income	float	%9.0g		median family income
owner00	float	%9.0g		pct owner occupied housing 2000
dwnom1n	float	%9.0g		dwnominate 1st dim
dwnom2n	float	%9.0g		dwnominate 2nd dim
ybush	float	%9.0g		
ygore	float	%9.0g		

Sorted by:

. summ

Variable	Obs	Mean	Std. Dev.	Min	Max
bush00	432	46.81019	14.36873	6	80
gore00	432	49.71296	14.005	20	93
black00	432	12.68333	16.25699	.3	96
south	432	.3101852	.4631056	0	1
hispanic00	432	12.31597	16.25281	.6	86
income	432	30.75224	8.319547	15.06	57.219
owner00	432	65.86875	11.92404	8	84.3
dwnom1n	432	.0586875	.4671698	-.815	1.269
dwnom2n	432	-.0381065	.4706545	-1.123	1.27
ybush	432	.4861111	.5003865	0	1
ygore	432	.4375	.4966535	0	1

. probit ybush black00 south hispanic00 income owner00 dwnom1n dwnom2n

Iteration 0: log likelihood = -299.27289
Iteration 1: log likelihood = -154.89847
Iteration 2: log likelihood = -134.46169
Iteration 3: log likelihood = -130.43351
Iteration 4: log likelihood = -130.12954
Iteration 5: log likelihood = -130.12789
Iteration 6: log likelihood = -130.12789

Probit estimates

Number of obs	=	432
LR chi2(7)	=	338.29
Prob > chi2	=	0.0000
Pseudo R2	=	0.5652

Log likelihood = -130.12789

ybush	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
-------	-------	-----------	---	------	----------------------

black00		-.0285973	.0097913	-2.92	0.003	-.0477878	-.0094067
south		.7695862	.2545783	3.02	0.003	.2706219	1.268551
hispanic00		-.0089458	.0069163	-1.29	0.196	-.0225015	.0046099
income		-.0241489	.0126394	-1.91	0.056	-.0489217	.0006239
owner00		.0235461	.0143687	1.64	0.101	-.0046161	.0517083
dwnom1n		2.761974	.2619813	10.54	0.000	2.2485	3.275448
dwnom2n		1.136417	.2339829	4.86	0.000	.6778186	1.595015
_cons		-.9697998	1.077738	-0.90	0.368	-3.082127	1.142528

FROM R USING GLM(.) FUNCTION

```
> summary(probit106)
```

Call:

```
glm(formula = TT[, 1] ~ TT[, 5] + TT[, 6] + TT[, 7] + TT[, 8] +  
    TT[, 9] + TT[, 10] + TT[, 11], family = binomial(link = probit))
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.608421	-0.303689	-0.001876	0.420137	2.796934

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-0.969796	1.089591	-0.890	0.37344	
TT[, 5]	-0.028597	0.009875	-2.896	0.00378	**
TT[, 6]	0.769584	0.255495	3.012	0.00259	**
TT[, 7]	-0.008946	0.006698	-1.336	0.18165	
TT[, 8]	-0.024149	0.012560	-1.923	0.05452	.
TT[, 9]	0.023546	0.014383	1.637	0.10162	
TT[, 10]	2.761973	0.261545	10.560	< 2e-16	***
TT[, 11]	1.136416	0.235248	4.831	1.36e-06	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null Deviance is the deviation of a model that contains only the intercept term, that is, a fixed probability for all observations. In this instance:

$$-2 \times 210 \times [\text{LOG}(210/432)] - 2 \times 222 \times [\text{LOG}(222/432)] = 302.95 + 295.59 = 598.54$$

Residual Deviance corresponds to the residual sum of squares in OLS which is used to estimate the standard deviation around the regression line. Here is simply:

$$-2 \times [\text{LOG LIKELIHOOD}] = -2 \times (-130.12789) = 260.26$$

Null deviance: 598.55 on 431 degrees of freedom

Residual deviance: 260.26 on 424 degrees of freedom

The Akaike Information Criterion (AIC) is:

$$276.26 = -2 \times (\text{LOG LIKELIHOOD}) + 2 \times K = -2 \times (-130.12789) + 2 \times 8$$

Where k=# of betas and the LOG LIKELIHOOD is given by STATA above.

AIC: 276.26

This is simply the number of iterations to estimate the model. Note that it is the same as the number of iterations in STATA.

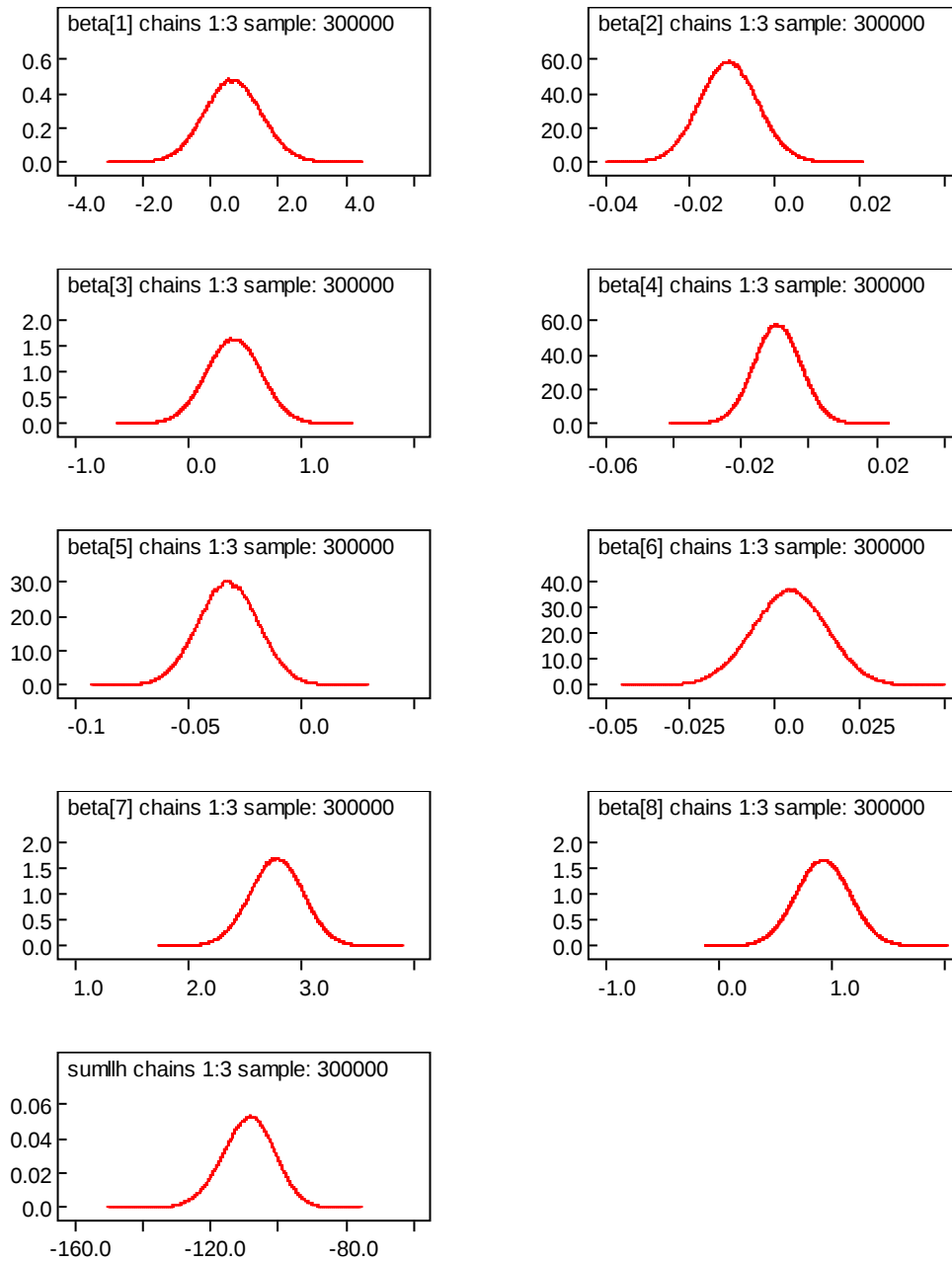
Number of Fisher Scoring iterations: 7

WINBUGS MODEL

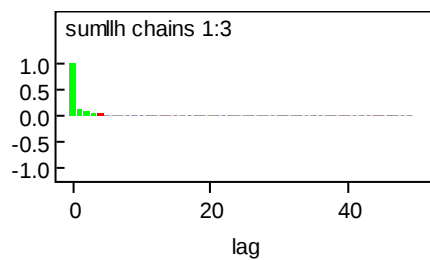
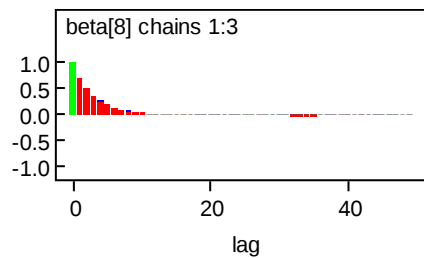
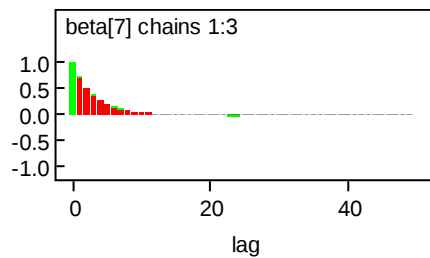
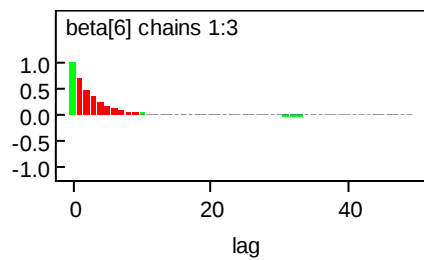
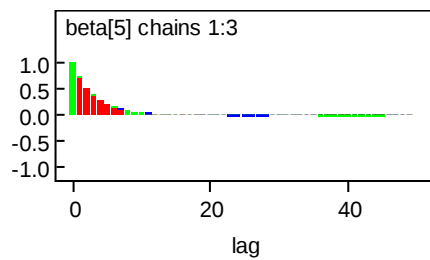
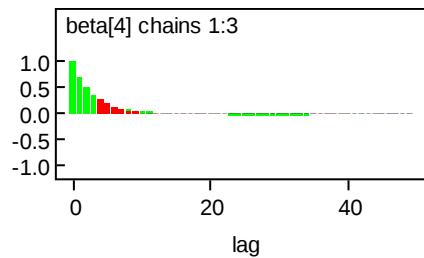
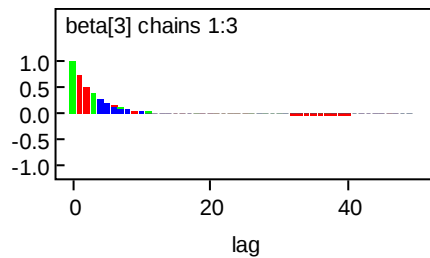
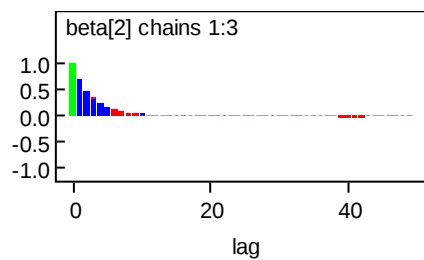
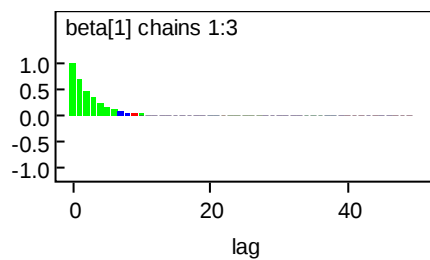
```
model
{
# X[,1] = 1 if bush vote >= 50%
# X[,2] = 1 if Gore vote >= 50%
# X[,3] = Bush Percentage in CD
# X[,4] = Gore Percentage in CD
# X[,5] = Black Percentage in CD
# X[,6] = 1 if Southern State (11 states of Confederacy + OK + KY
# X[,7] = Hispanic Percentage in CD
# X[,8] = Median Family Income (in thousands) in CD
# X[,9] = Percent Owner-Occupied Housing
# X[,10] = DW-NOMINATE 1st Dimension
# X[,11] = DW-NOMINATE 2nd Dimension
#
# PRIORS
#
#       for (k in 1 : 8) { beta[k] ~ dnorm(0,0.001)} # vague priors
#
# LIKELIHOOD
#
#       for (i in 1 : 432) # loop over congressional districts
#       {
#
#           X[i,1] ~ dbern(p[i]);
#           probit(p[i]) <- delta[i]
#           delta[i] ~ dnorm(mu[i], 1.0)I(-4, 4)   TRICK SHOWN ON p.48
#           mu[i] <-
# beta[1]+X[i,5]*beta[2]+X[i,6]*beta[3]+X[i,7]*beta[4]+X[i,8]*beta[5]+X[i,9]*beta[6]+X[i,10]
# ]*beta[7]+X[i,11]*beta[8]
#
#       Borrowed From Simon Jackman
#
#           llh[i] <- X[i,1]*log(p[i]) + (1-X[i,1])*log(1-p[i]);
#       }
#       sumllh <- sum(llh[]);
#
#
# }
```

	node	mean	sd	MC error	2.5%	median	97.5%	start	sample
const	beta[1]	0.6527	0.8329	0.003467	-0.9636	0.6471	2.3	1001	300000
black	beta[2]	-0.01095	0.006828	2.886E-5	-0.02418	-0.011	0.002597	1001	300000
south	beta[3]	0.4021	0.2405	0.001108	-0.0689	0.4023	0.8726	1001	300000
hispan	beta[4]	-0.00937	0.006909	3.034E-5	-0.02291	-0.0093	0.004186	1001	300000
income	beta[5]	-0.03229	0.01332	6.236E-5	-0.05833	-0.0323	-0.00621	1001	300000
owner	beta[6]	0.00421	0.0109	4.678E-5	-0.01739	0.004275	0.02536	1001	300000
dwnom1	beta[7]	2.781	0.2343	0.001081	2.32	2.782	3.236	1001	300000
dwnom2	beta[8]	0.9224	0.2388	0.001079	0.4556	0.9222	1.391	1001	300000
	sumllh	-109.0	7.639	0.01882	-124.5	-108.8	-94.56	1001	300000

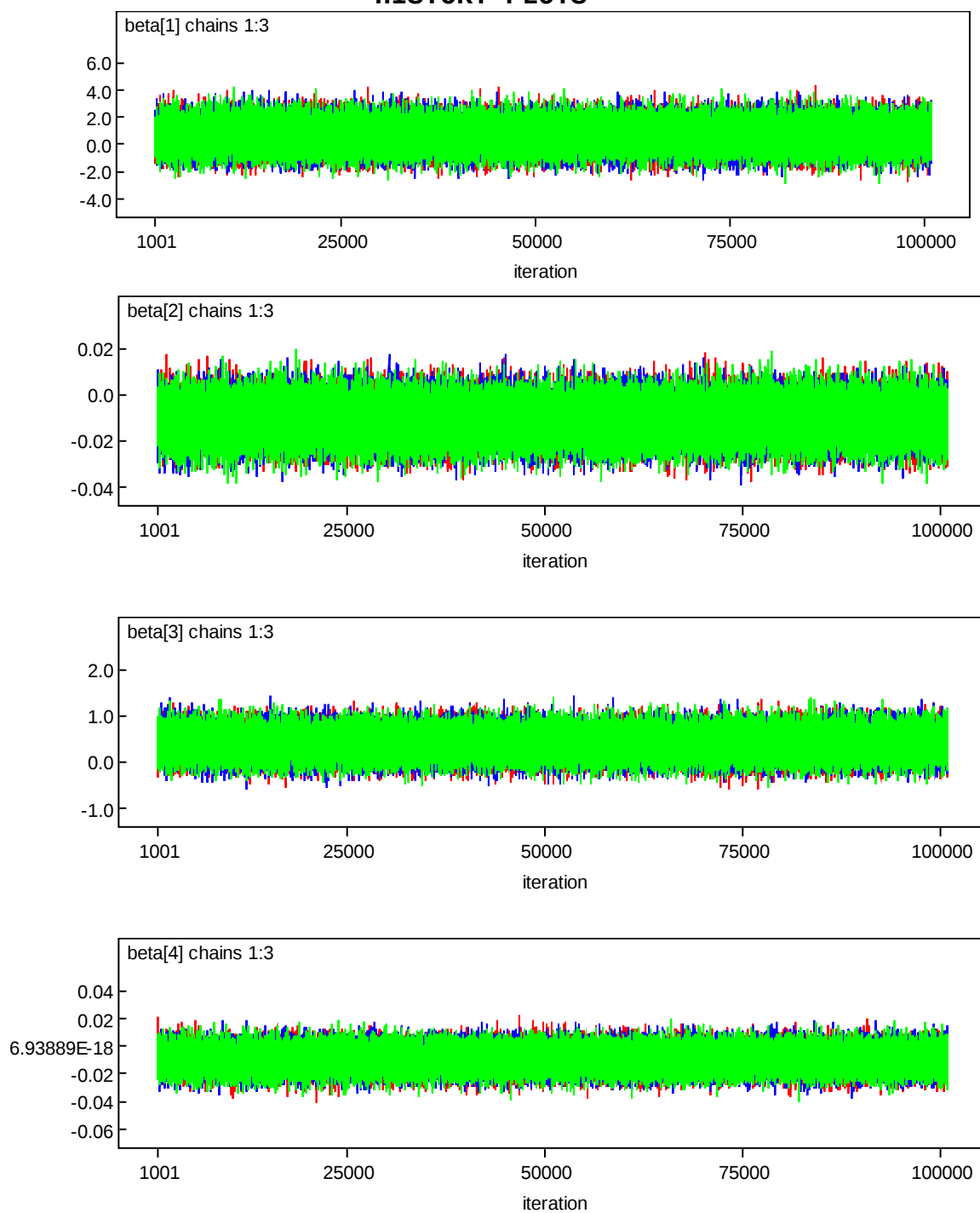
DENSITY PLOTS

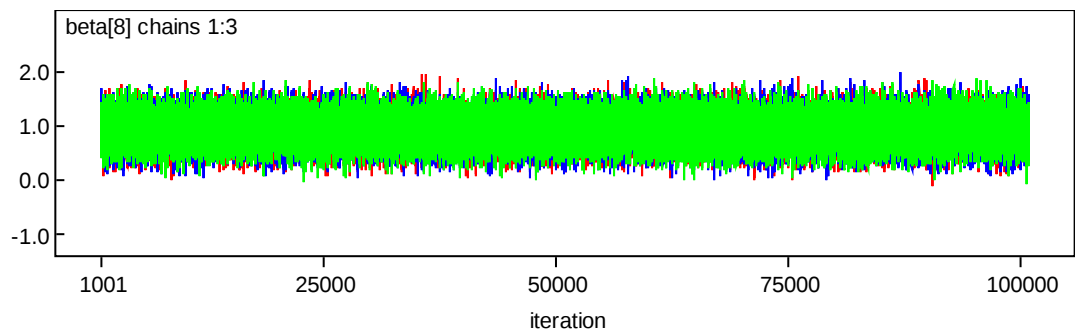
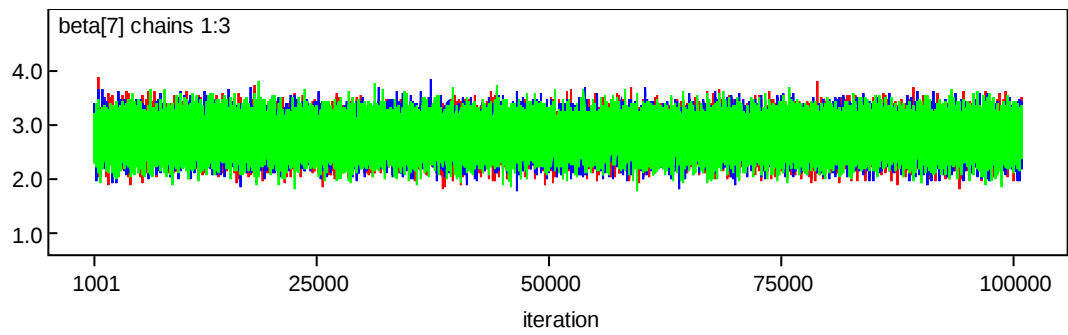
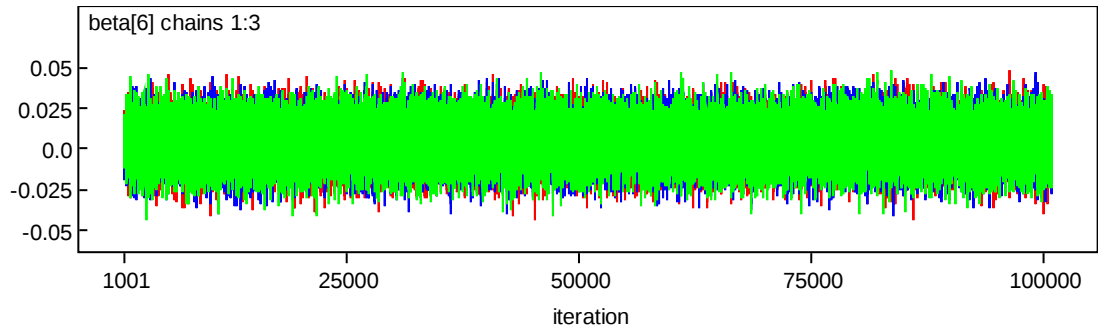
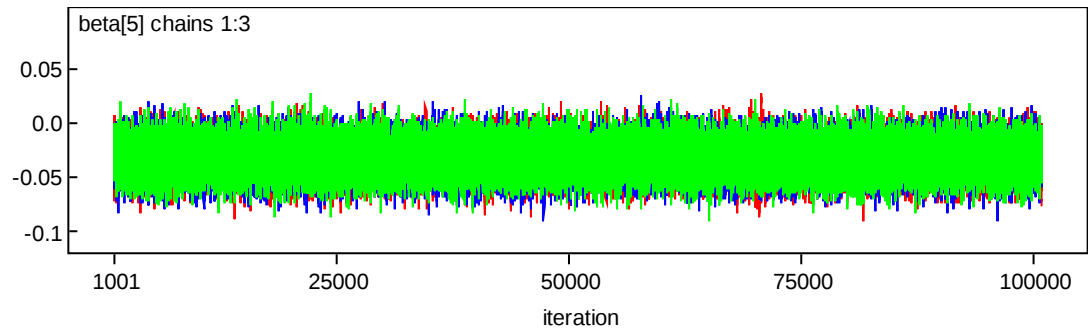


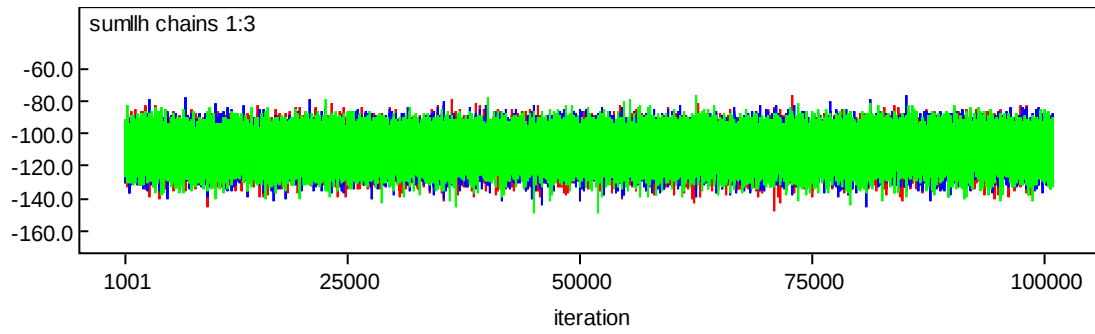
AUTOCORRELATION PLOTS



HISTORY PLOTS







Brooks, Gelman, and Rubin Diagnostic

