Biostat 234 Final DAP: Traffic in Los Angeles County

Hubert Jenq

Introduction:

Traffic in Los Angeles County has been increasing and commute times for many of its residents are becoming longer. In order to determine where highways should be expanded or how many highways should be developed in a new area, we are interested city level measurements that could correlate with traffic in Los Angeles County. The goal of this analysis is to look for variables that are linearly associated with a measure of traffic.

Data:

The data in this analysis was taken from the U.S. Census Bureau and the California Traffic Census Program. Census statistics that relate to people, housing, education, geography, income and poverty, and economy of cities in Los Angeles County were obtained from https://www.census.gov for the years between 2010 and 2015. Traffic data for two years, 2014 and 2013, was taken from http://www.dot.ca.gov/trafficops/census/ and had the form of annual average daily traffic volumes for segments of the highway. The traffic data had a count for average number of daily cars in front and behind of the postmile marker on the highway. Vehicle miles traveled (VMT) was calculated by using the average forward and behind count of the respective postmile markers multiplied by the length of the segment between the two. A city had its total VMT summed from all the highways that exist in the city. The VMT per capita in a city was calculated from the average total daily vehicle miles traveled in the city divided by the population in the city. A total of 15 covariates were chosen from the total census data based on intuition of how they would affect VMT per capita and based on their completeness of values. The covariates were centered and standardized before any analysis was performed.

Summary statistics of the collected variables of the data can be found in table 1. Of note is the VMT per capita, which has a continuous outcome and a range of 102.9. An outlier in number of housing units, and retail sales can be observed which corresponds to Los Angeles. Inclusion or removal of the outlier does not seem to significantly affect its linearity. The outcome variable VMT per capita and individual covariates are plotted with a LOESS curve in figure 1 to assess linearity and to determine whether transformations will be necessary. No transformations were found to improve the visual linearity of the data.

Model Specification

$$
\begin{align}
y_{ij}|\mu_{ij} &\sim No(\mu_{ij}, \tau^2) \\
\mu_{ij}|x_{ij}, \sigma_i &= \boldsymbol{BX}[i + 68 * j,]^T + \sigma_i \\
\tau^2 &\sim \text{Gamma}(a, b) \\
\beta_k &\sim No(M[k], P[k]) \\
\sigma_i|\psi^2 &\sim No(0, \psi^2) \\
\psi^2 &\sim \text{Gamma}(c, d) \\
&
\end{align}
$$

A linear regression model with random effects was chosen for the set of data. The counter i goes from 1 to 68, j goes from 0 to 1, and k goes from 1 to 15. y_{ij} is the VMT per capita for city i in year 2013 (j=0) or 2014 (j=1). x is the covariate matrix, where the first 68 rows are values for the 68 cities in 2013. The next 68 rows are the same cities but the 2014 values. μ_{ij} is the expected VMT per capita. τ^2 is the linear regression error term which follows a gamma distribution with shape parameters a and b. The β_k's are the slope coefficients which have a normal prior with mean M[k] and precision P[k]. ϵ_i is the random effect for the city i which follows a normal prior with mean 0 and precision ψ^2. ψ^2 follows a gamma distribution with shape parameters c and d.

Prior Specification

Priors for the β's will be specified based on intuition. The mean will be set to the average VMT per capita divided by the covariate of interest if there is an anticipated correlation between the two. The sign will be set based on predicted positive or negative association. The precision of the coefficient will be set such that 0 will be within one standard deviation of the mean. If there is no correlation, the mean will be set to zero and the precision will be set such that the average VMT per capita divided by the covariate of interest is within one standard deviation of zero. Values for the prior means and precision can be found in the table 2 and the rationale in the three paragraphs below.

I anticipate that a higher owned housing percentage, median value of house, median household income, and per capita income will all positively contribute to VMT per capita since all of these metrics can indicate a more wealthy neighborhood neighborhood. A more wealthy neighborhood can be more attractive for other people to commute to for work or entertainment. I believe a higher working percentage would also contribute to VMT per capita since those residents will need to commute to work daily. I belive both high school graduate percentage and bachelor percentage will increase VMT per capita since graduates are more likely to own a car and be employed.

I think population density will be negatively correlated with VMT per capita since less travel is necessary in areas that are more population dense. Entertainment and workplaces are more likely to be close since the density is increased. I speculate mean travel time to work will negatively correlated since those travelling long distances will be in cities that are less busy with fewer entertainment and work options. Since fewer people are working and going to those cities, there will be less driving within the city. I anticipate poverty percentage to be negatively correlated with VMT per acpita since those in poverty will be less likely to be employed and own a car which decreases their VMT.

I do not anticipiate race having any effect on VMT per capita so the means for black percentage, asian percentage and hispanic percentage. I believe the number of housing units or retail sales will not have an effect on the VMT per capita since the VMT is normalized by capita.

Variance for the random effect of the city is unknown but I will estimate it using the range method on the VMT per capita. VMT per capita has an approximate range of 100 which yields a standard deviation of 25, or a precision of 1/625. The shape and rate parameters, c and d, for the random effect precision ψ^2 will be set to 1 and 625 to yield a mean of 1/625. Error for the model τ^2 is estimated with 120 degrees of freedom and precision equal to the variance of the VMT per capita which equals 1/458. This yields shape and rate parameters a, b, of 60 and 27480.

Convergence

Convergence of the model was assessed by examining the autocorrelation plots and the time series plots. The autocorrelation plots for the β_i go to zero immediately as shown in figure 3. The time series plots for β_i also look completely random indicating convergence of the chain as shown in figure 2. The autocorrelation plot for the random effect precision ψ^2 converges to zero in about 15 iterations as shown in figure 4. The model error τ^2's autocorrelation quicky goes to zero in about five iterations. The corresponding time series plot of τ^2 also looks completely random indicating it has converged. The random effects σ_i autocorrelation and time series plots are not included but all converge in less than 5 iterations. In the simulations below, three chains of length of 50,000 are used.

Results

The posteriors of the model parameters of interest are shown in table 3. Two parameters are significant with a p-value less than 0.05. Hispanic percentage has a mean of 0.56 and a 95% credible interval of 0.02 to 1.09. This can be interpreted as for each percentage increase in hispanics in a city holding all the other covariates constant, we can expect an 0.56 increase in the average VMT per capita. The other significant parameter is population density which has a mean of -0.65, and a 95% credible interval of -1.26 to -0.03. The population density parameter can be interpreted as for each increase in 1,000 people per square mile holding all other covariates constant, we can expect a decrease of 0.65 in the average VMT per capita. This follows our original intuition that the more dense the city, the less likely the residents within are to drive in it. Asian percentage is significant with a p-value of less than 0.10 with mean 0.27 and 95% credible interval -0.12 to 0.66. This can be interpreted as for each percentage increase of asians in a city, we can expect the average VMT per capita to increase by 0.27.

Sensitivity Analysis

A sensitvitiy analysis of our priors is performed to see how our prior assumptions affect the final result. Priors on the regression coefficients are varied by halving or doubling their respective precisions. Additionally, the prior regression coefficients means are set to an uninformed 0 to see if the intuition on the coefficients affects the final result. The change in significance can be seen in table 4. Here we see that hispanic percentage remains significant despite changes in its prior precision. The respective means are 0.49, 0.52, and 0.46 for the no mean, half precision, and double precision models.The population density changes significance with varying the prior assumptions indicating that it is affected by our prior belief. The means for the population density parameter varys from -0.34, -0.63, and -0.18 for the no mean, half precision, and double precision model. In all three cases, the parameters are not significant at the 0.05 level. The rest of the posterior parameter estimates can be seen in tables 5, 6, and 7.

Discussion

From the linear regression analysis above, we can conclude that both hispanic percentage and population density has a significant linear relationship with VMT per capita while holding all other covariates constant. When planning a new city, the expected population density and percentage of hispanics can be used to predict how busy the highways will be after adjusting for population.

Appendix

Table 1: Summary Statistics

	n	mean	sd	median	min	max	range
VMTperCapita	136	21.62	21.39	13.82	0.63	103.62	102.99
BlackPercentage	136	6.04	8.05	3.00	0.40	43.90	43.50
AsianPercentage	136	18.20	18.21	11.25	0.30	66.90	66.60
HispanicPercentage	136	41.05	26.74	34.10	6.10	94.80	88.70
NumberHousingUnits(1,000)	136	43.45	169.80	16.02	3.38	1413.99	1410.61
OwnedHousesPercentage	136	56.09	14.89	54.95	25.50	88.10	62.60
MedianValueOfHouse($100,000)	136	5.08	2.24	4.26	1.60	10.00	8.40
HighSchoolGradPercentage	136	81.20	13.89	82.95	49.20	99.30	50.10
BachelorPercentage	136	33.22	19.49	29.85	5.50	74.20	68.70
WorkingPercentage	136	63.78	4.67	62.80	54.20	75.40	21.20
RetailSales($100,000)	136	15.33	48.33	6.00	0.37	401.57	401.20
MeanTravelTimeToWork	136	29.26	2.92	29.05	22.80	40.20	17.40
MedianHouseholdIncome($1,000)	136	70.58	25.38	64.69	41.93	151.79	109.86
PerCapitaIncome($1,000)	136	32.38	17.43	27.52	12.58	95.21	82.63
PovertyPercentage	136	13.04	6.10	11.55	2.30	26.70	24.40
PopulationDensity(1,000)	136	7.02	3.84	7.08	0.64	16.60	15.96

Table 2: Linear Regression Prior parameter specification

	Mean	Precision
BlackPercentage	0.00	12.81
AsianPercentage	0.00	1.41
HispanicPercentage	0.00	0.28
NumberHousingUnits(1,000)	0.00	0.25
OwnedHousesPercentage	2.59	0.15
MedianValueOfHouse($100,000)	0.23	18.15
HighSchoolGradPercentage	3.76	0.07
BachelorPercentage	1.54	0.42
WorkingPercentage	2.95	0.11
RetailSales($100,000)	0.00	1.99
MeanTravelTimeToWork	-1.35	0.55
MedianHouseholdIncome($1,000)	3.26	0.09
PerCapitaIncome($1,000)	1.50	0.45
PovertyPercentage	-0.60	2.75
PopulationDensity(1,000)	-0.32	9.50

Table 3: Posterior Estimates of model parameters

	Mean	SD	2.5%	97.5%	P>0	Significance
BlackPercentage	0.04	0.23	-0.42	0.49	0.56	
AsianPercentage	0.32	0.20	-0.06	0.72	0.95	*
HispanicPercentage	0.56	0.27	0.02	1.09	0.98	**
NumberHousingUnits(1,000)	-0.02	0.11	-0.22	0.19	0.44	

	Mean	SD	2.5%	97.5%	P>0	Significance
OwnedHousesPercentage	0.07	0.44	-0.76	0.96	0.55	
MedianValueOfHouse($100,000)	0.23	0.23	-0.23	0.70	0.85	
HighSchoolGradPercentage	-0.28	0.55	-1.34	0.81	0.30	
BachelorPercentage	0.15	0.54	-0.91	1.21	0.61	
WorkingPercentage	0.39	0.67	-0.93	1.71	0.72	
RetailSales($100,000)	0.03	0.37	-0.71	0.76	0.53	
MeanTravelTimeToWork	-0.75	0.82	-2.35	0.85	0.18	
MedianHouseholdIncome($1,000)	0.01	0.52	-1.02	1.01	0.51	
PerCapitaIncome($1,000)	0.39	0.59	-0.75	1.55	0.75	
PovertyPercentage	-0.51	0.49	-1.45	0.47	0.15	
PopulationDensity(1,000)	-0.65	0.32	-1.26	-0.03	0.02	**
intercept	0.03	1.00	-1.92	1.99	0.51	

Table 4: Sensitivity analysis on parameters. *** = p<0.01, ** = p<0.05, * = p<0.10

	Original Model	No Prior Mean	Double Precision	Half Precision
BlackPercentage				
AsianPercentage	*	*	*	*
HispanicPercentage	**	**	**	**
NumberHousingUnits(1,000)				
OwnedHousesPercentage				
MedianValueOfHouse($100,000)				
HighSchoolGradPercentage				
BachelorPercentage				
WorkingPercentage				
RetailSales($100,000)				
MeanTravelTimeToWork				
MedianHouseholdIncome($1,000)				
PerCapitaIncome($1,000)				
PovertyPercentage				
PopulationDensity(1,000)	**			*
intercept				

Table 5: Posterior Estimates of no prior mean model parameters

	Mean	SD	2.5%	97.5%	P>0	Significance
BlackPercentage	-0.02	0.23	-0.46	0.42	0.47	
AsianPercentage	0.27	0.20	-0.12	0.66	0.91	*
HispanicPercentage	0.49	0.27	-0.06	1.03	0.96	**
NumberHousingUnits(1,000)	-0.05	0.11	-0.25	0.16	0.33	
OwnedHousesPercentage	-0.14	0.44	-0.99	0.72	0.37	
MedianValueOfHouse($100,000)	0.00	0.23	-0.46	0.47	0.50	
HighSchoolGradPercentage	-0.19	0.55	-1.28	0.89	0.36	
BachelorPercentage	0.14	0.55	-0.95	1.25	0.60	
WorkingPercentage	0.05	0.67	-1.27	1.34	0.53	
RetailSales($100,000)	0.13	0.37	-0.60	0.86	0.64	
MeanTravelTimeToWork	-0.25	0.84	-1.90	1.41	0.38	
MedianHouseholdIncome($1,000)	0.31	0.52	-0.70	1.31	0.72	

	Mean	SD	2.5%	97.5%	P>0	Significance
PerCapitaIncome($1,000)	0.07	0.60	-1.13	1.23	0.55	
PovertyPercentage	-0.14	0.50	-1.12	0.83	0.38	
PopulationDensity(1,000)	-0.34	0.31	-0.95	0.29	0.14	
intercept	0.03	0.99	-1.89	1.98	0.52	

Table 6: Posterior Estimates of half precision model parameters

	Mean	SD	2.5%	97.5%	P>0	Significance
BlackPercentage	0.01	0.29	-0.57	0.57	0.52	
AsianPercentage	0.31	0.21	-0.10	0.72	0.93	*
HispanicPercentage	0.52	0.29	-0.06	1.07	0.96	**
NumberHousingUnits(1,000)	-0.04	0.11	-0.26	0.19	0.35	
OwnedHousesPercentage	-0.17	0.46	-1.05	0.72	0.35	
MedianValueOfHouse($100,000)	0.01	0.33	-0.65	0.66	0.51	
HighSchoolGradPercentage	-0.21	0.57	-1.35	0.90	0.36	
BachelorPercentage	0.11	0.58	-1.05	1.23	0.58	
WorkingPercentage	0.14	0.70	-1.21	1.55	0.58	
RetailSales($100,000)	0.12	0.40	-0.68	0.90	0.62	
MeanTravelTimeToWork	-0.31	0.94	-2.19	1.55	0.37	
MedianHouseholdIncome($1,000)	0.29	0.54	-0.79	1.35	0.71	
PerCapitaIncome($1,000)	0.14	0.63	-1.08	1.40	0.58	
PovertyPercentage	-0.22	0.62	-1.43	1.00	0.36	
PopulationDensity(1,000)	-0.63	0.43	-1.47	0.22	0.07	*
intercept	0.03	0.99	-1.89	1.96	0.52	

Table 7: Posterior Estimates of double precision model parameters

	Mean	SD	2.5%	97.5%	P>0	Significance
BlackPercentage	-0.02	0.18	-0.37	0.32	0.46	
AsianPercentage	0.24	0.19	-0.13	0.60	0.91	*
HispanicPercentage	0.46	0.26	-0.05	0.96	0.96	**
NumberHousingUnits(1,000)	-0.04	0.09	-0.22	0.14	0.33	
OwnedHousesPercentage	-0.13	0.41	-0.93	0.69	0.38	
MedianValueOfHouse($100,000)	0.00	0.17	-0.32	0.33	0.50	
HighSchoolGradPercentage	-0.18	0.52	-1.20	0.83	0.37	
BachelorPercentage	0.15	0.50	-0.81	1.14	0.61	
WorkingPercentage	0.00	0.62	-1.22	1.20	0.49	
RetailSales($100,000)	0.11	0.32	-0.53	0.75	0.65	
MeanTravelTimeToWork	-0.18	0.70	-1.55	1.21	0.40	
MedianHouseholdIncome($1,000)	0.31	0.49	-0.62	1.28	0.74	
PerCapitaIncome($1,000)	0.05	0.54	-1.01	1.05	0.54	
PovertyPercentage	-0.10	0.37	-0.81	0.64	0.39	
PopulationDensity(1,000)	-0.18	0.23	-0.61	0.27	0.21	
intercept	0.01	1.01	-1.99	1.97	0.50	

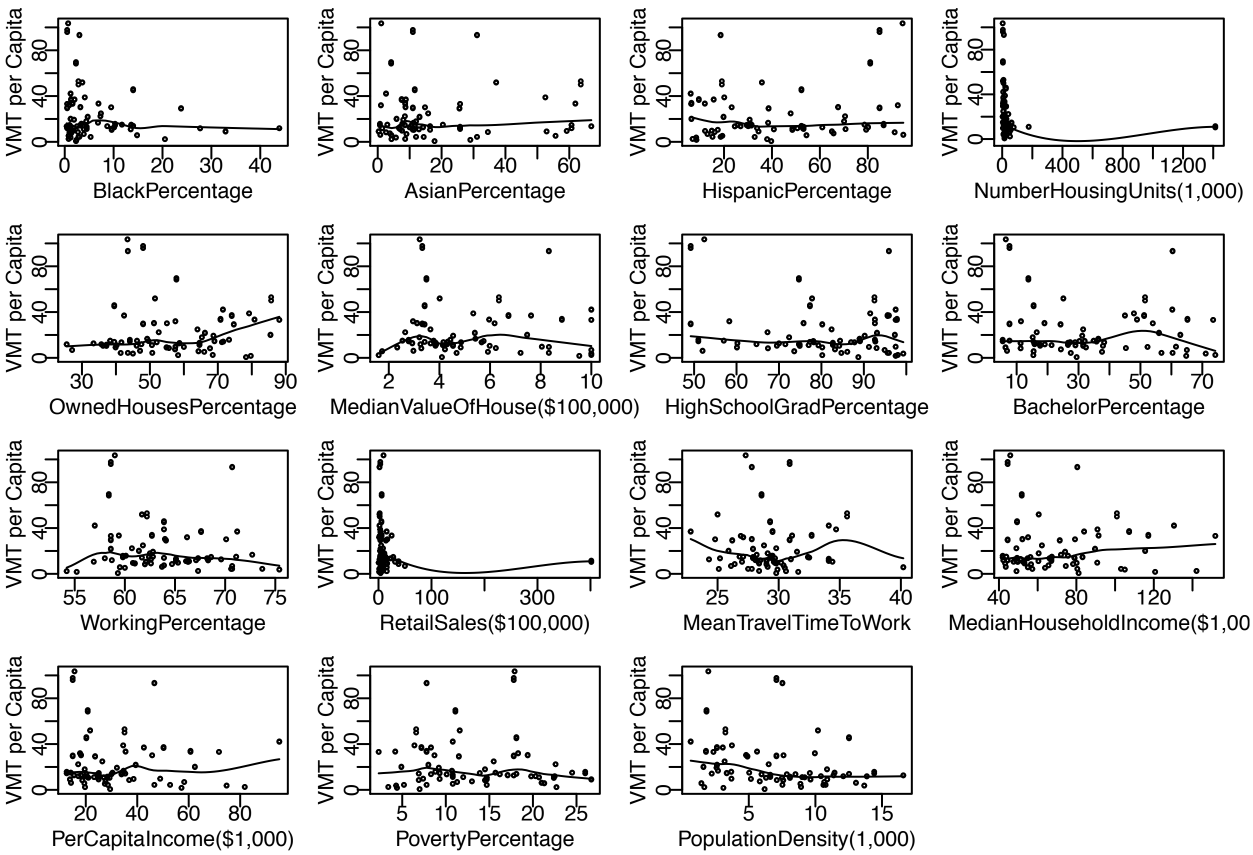

Figure 1: Scatterplots of VMT per capita vs. covariates with LOESS curve, span=.4

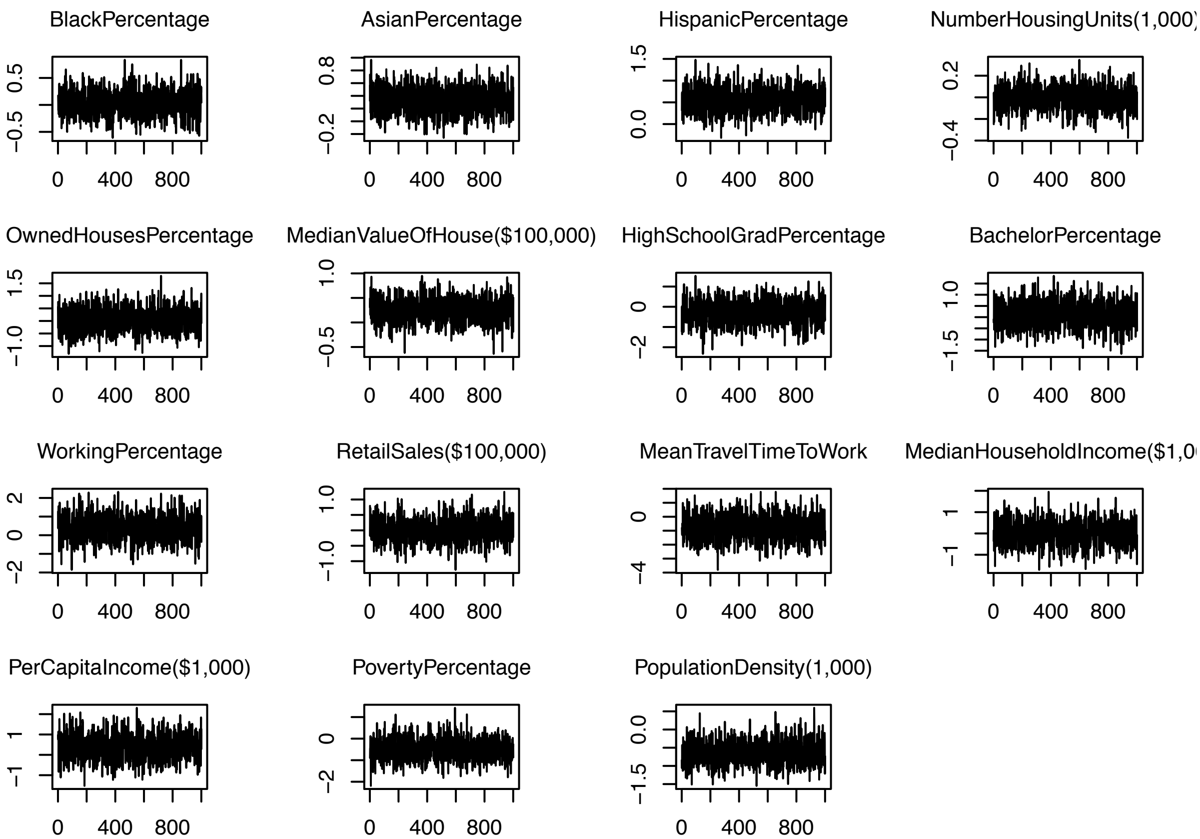

Figure 2: Time series plots for beta

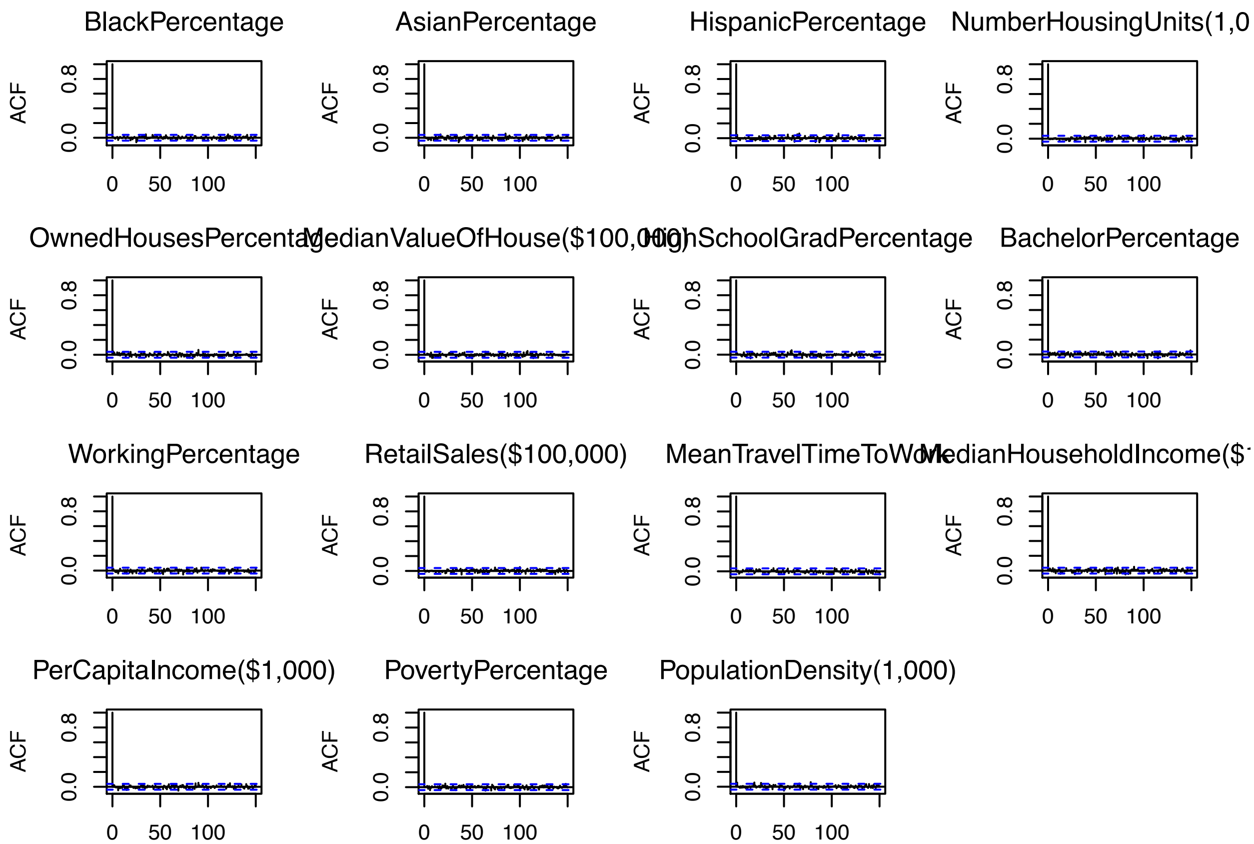

Figure 3: Autocorrelation plots for beta

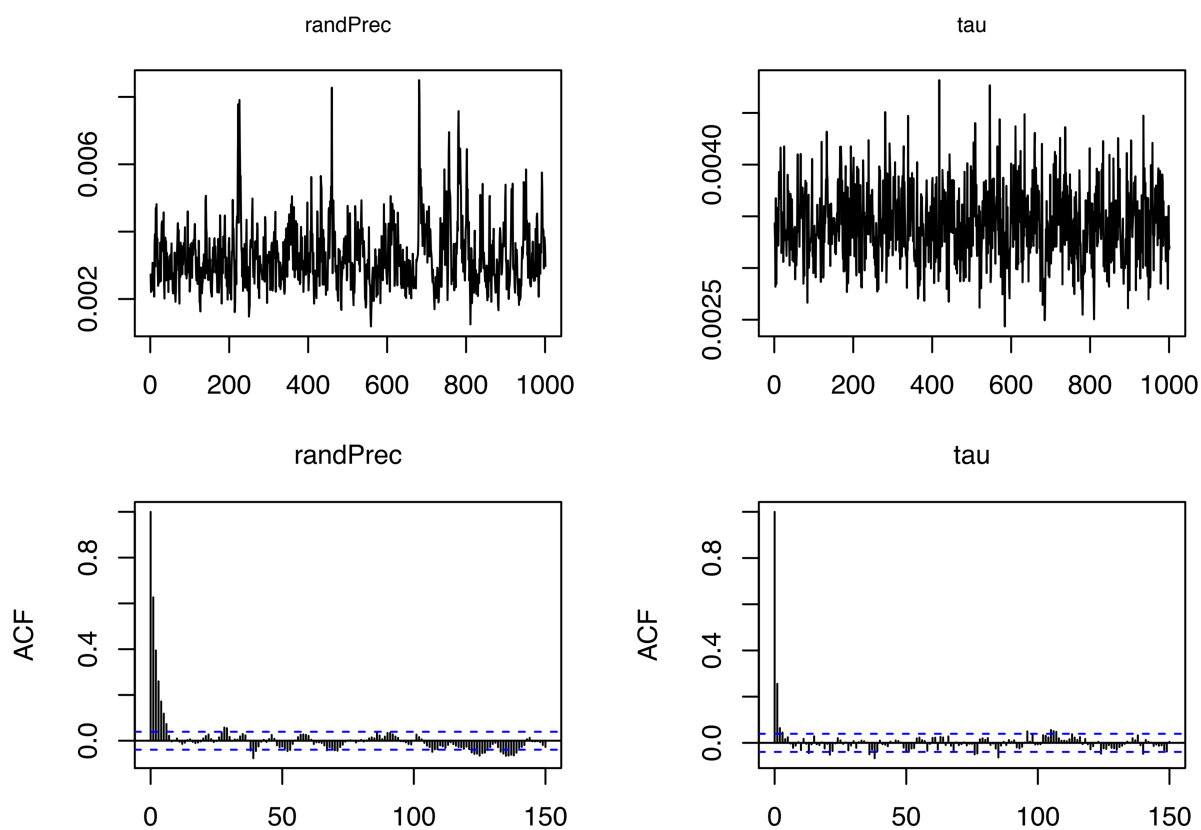

Figure 4: Time series and autocorrelation plots for random effects precision and tau^2

Jags Model

```
model<-function()
{
  for(i in 1:N) {
    for(k in 1:2){
      y[i+(k-1)*68] ~ dnorm( mu[i,k] , tau )
      mu[i,k] <- beta0 + inprod(x[i+(k-1)*68,] , beta[] ) + randEff[i]
    }
    randEff[i] ~ dnorm(randMu,randPrec)
  }

  beta0 ~ dnorm( 0 , 1)

  for (j in 1:K) {
    beta[j] ~ dnorm( betapriormean[j] , betapriorprec[j] )
  }

  tau ~ dgamma( tau.a , tau.b )
  randPrec ~dgamma(randPrec.a,randPrec.b)
  sigma <- 1 / sqrt( tau )
}
```