An Analysis of the Presidential Elections

Whzecomjm Zhang

South University of Science and Technology of China

June 12 2013

Introduction

- Economic and Political Data Months Before the Election
- Predict Results of Election

Some Factors For Prediction (Gallup poll)

- July Popularity
- Peace Question
- Future Problem
- Leading Indicators
- GNP Change
- Second Term Indicator

Question

- Which is the best model for the prediction of Incumbent vote?
- Which variables should be chosen in the model?
- How can we have the best fit of the model?

Mainline

- Full Model Analysis (MLR)

- Reduced Model Analysis (MLR)

Full Model Analysis

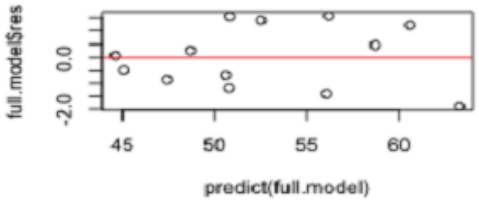
$$IncVote_i = \beta_0 + \beta_1 x_J + \beta_2 x_P + \beta_3 x_F + \beta_4 x_L + \beta_5 x_G + \beta_6 x_T + \epsilon_i$$

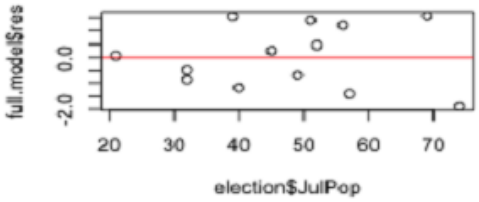
$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \\ \hat{\beta}_5 \\ \hat{\beta}_6 \end{pmatrix} = \begin{pmatrix} 29.66 \\ 0.1449 \\ 0.04638 \\ 0.1333 \\ -0.004 \\ 1.783 \\ 2.784 \end{pmatrix} \qquad p = \begin{pmatrix} 0.00109 \\ 0.04970 \\ 0.45864 \\ 0.30202 \\ 0.98477 \\ 0.00411 \\ 0.08199 \end{pmatrix}$$

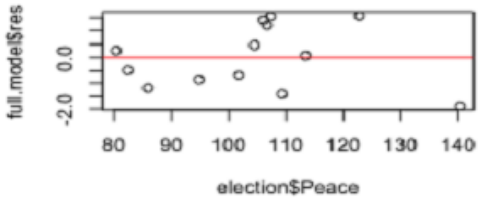
P values of beta_2, beta_3, beta_4, beta_6 are bigger than 5%

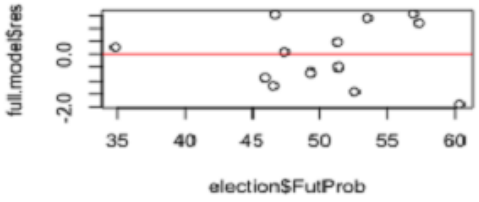
Full model Residuals Diagnostics

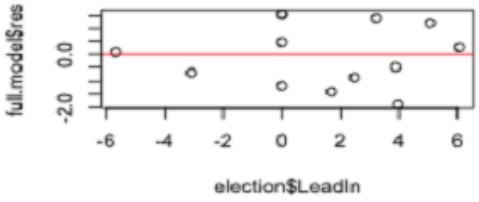
Constant Variance Assumption

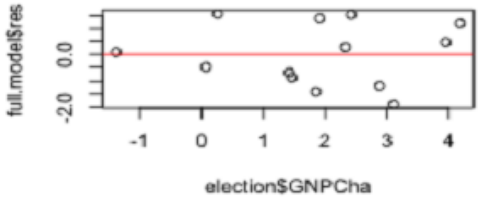


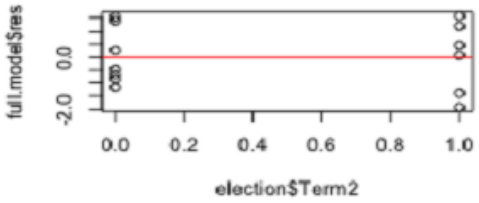






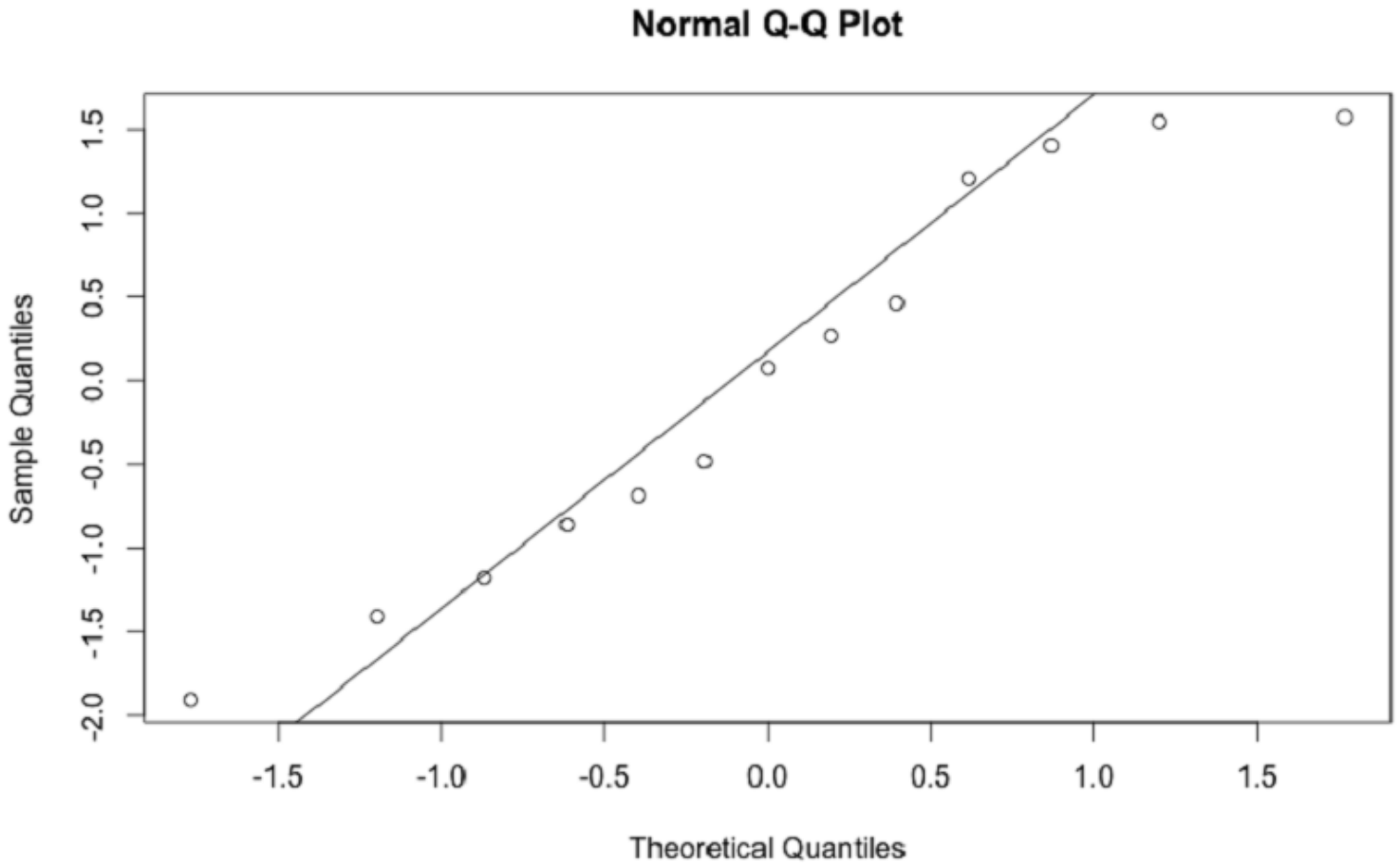






Full model Residuals Diagnostics

Residual Normality Check



Influential points or Outlier

- There is no high leverage or influential point in this data set

- Studentized residual v.s. Bonferroni value

  (Their are both equal to -5.076, hence no outliers)

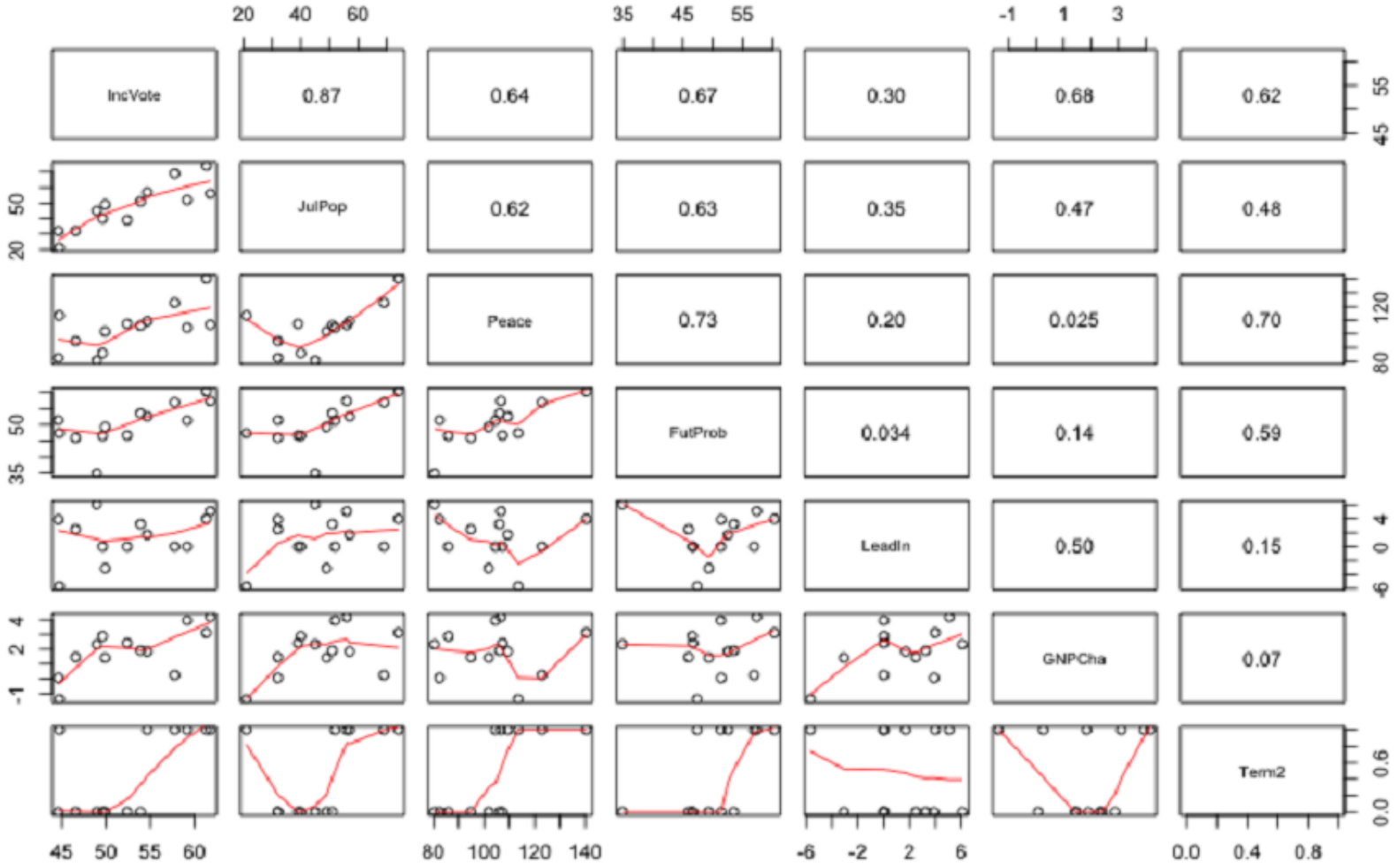
Serial Correlation of the data errors

- corr(ϵ_t,ϵ_t-1)=0.05228

  where ϵ_t is observation error

- Assumption of independence in error is proper

Colinearity



Variable selection

- Backward Elimination

backward Elimination	1st drop LeadIn	2nd drop Peace	3rd drop FutProb	4th drop Term2	5th drop JulPop	6th drop GNPCha
AIC	54.4104	53.9532	56.1634	61.84	64.4202	70.3809
BIC	58.3651	57.3429	58.9882	64.6648	67.2450	73.2056
Adjusted R-Square	0.9319	0.9329	0.9176	0.8724	0.8444	0.7539
Keep/Drop Variable	Drop	Drop	Keep	Keep	Keep	Keep

$$IncVote_i = \beta_0 + \beta_1 x_J + \beta_3 x_F + \beta_5 x_G + \beta_6 x_T + \epsilon_i$$

VARIABLE SELECTION

- Forward Elimination

Forward Elimination	1st add GNPCha	2nd add JulPop	3rd add Term2	4th add FutProb	5th add Peace	6th add LeadIn
AIC	80.323	66.0492	56.1634	53.9532	54.4104	56.4096
BIC	82.018	68.3090	58.9882	57.3429	58.3651	60.9292
Adjusted R-Square	0.4115	0.8149	0.9176	0.9329	0.9319	0.9206
Add / Drop Variable	Add	Add	Add	Add	Drop	Drop

$$IncVote_i = \beta_0 + \beta_1 x_J + \beta_3 x_F + \beta_5 x_G + \beta_6 x_T + \epsilon_i$$

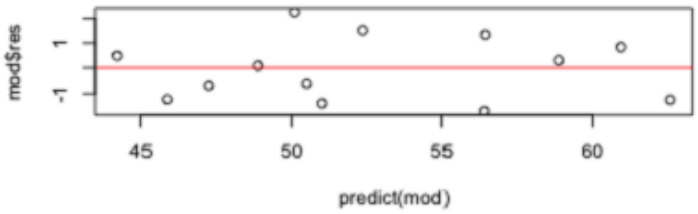
Reduced Model

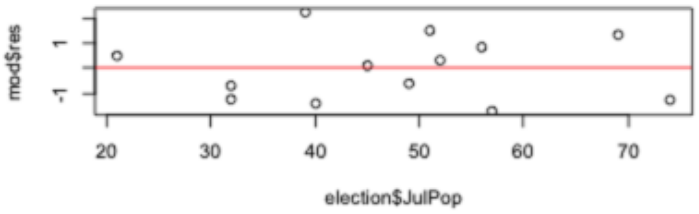
$$IncVote_i = \beta_0 + \beta_1 x_J + \beta_3 x_F + \beta_5 x_G + \beta_6 x_T + \epsilon_i$$

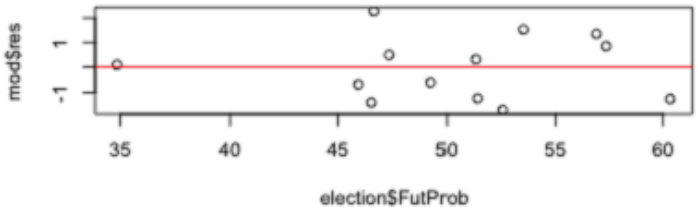
$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_3 \\ \beta_5 \\ \beta_6 \end{pmatrix} = \begin{pmatrix} 31.6122 \\ 0.1624 \\ 0.1733 \\ 1.6761 \\ 3.2973 \end{pmatrix} \qquad p = \begin{pmatrix} 5.86 \times 10^{-5} \\ 0.00711 \\ 0.11838 \\ 0.00105 \\ 0.01658 \end{pmatrix}$$

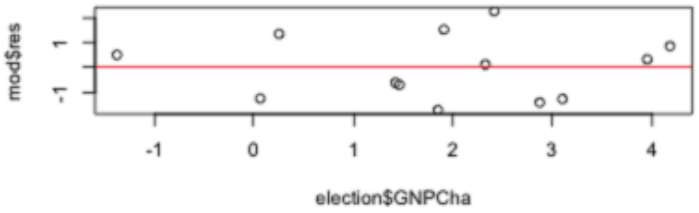
Reduced Model Residuals Diagnostics

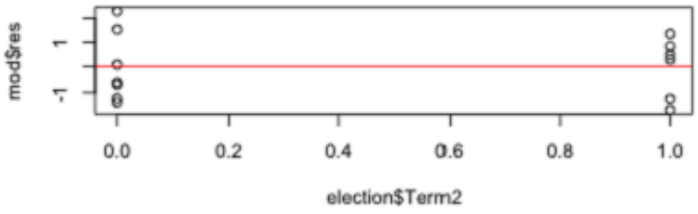
Constant Variance Assumption





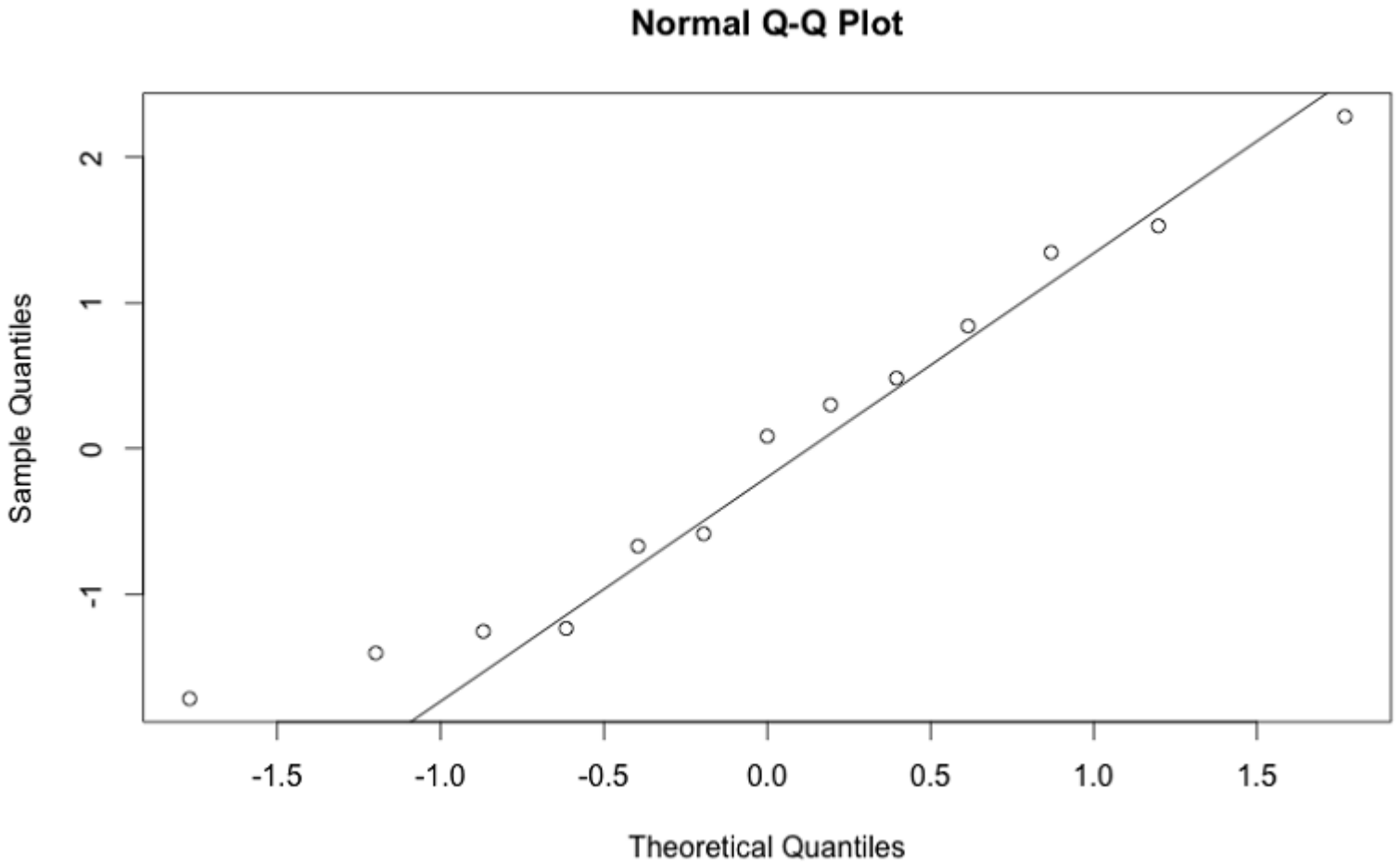






Reduced Model Residuals Diagnostics

Residual Normality Check



Influential points or Outlier

- There is no high leverage or influential point in this data set

- Studentized residual v.s. Bonferroni value

  (Their are both equal to -4.239, hence no outliers)

Serial Correlation of the data errors

- corr(ϵ_t,ϵ_t-1)= −0.1888

  where ϵ_t is observation error

- Assumption of independence in error is proper

Prediction Example

- We consider 1996 IncVote Value for example.

$$IncVote_i^{(1996)} = \beta_0 + \beta_1 x_J^{(1996)} + \beta_3 x_F^{(1996)} + \beta_5 x_G^{(1996)} + \beta_6 x_T^{(1996)} + \epsilon_i$$

$$IncVote_{i,1996} = 56.79064$$

$\hat{Y} \pm s \times t_{n-k,\frac{\alpha}{2}}$, which s is the standard error=0.7312829, and have DF=7

$$\hat{Y} = \{55.06143 \leq \hat{Y} \leq 58.51985\}$$

the true value of 1996 IncVote equal to 54.66

Reference

- [1] Ray C. Fair, Presidential and Congressional Vote-Share Equations: November 2010 Update
- [2] Audic, S. and J. M. Claverie (1997). The significance of digital gene expression profiles. Genome Res 7(10): 986-95.
- [3] Akaike, Hirotsugu. A new look at the statistical model identification. IEEE Transactions on Automatic Control. 1974, 19 (6): 716–723.
- [4] Neath, A. A. and Cavanaugh, J. E. (2012). The Bayesian information criterion: Background, derivation, and applications. WIREs Computational Statistics4, 199.203.

Thank You !