5-2014

Bayesian Inference: Probit and Linear Probability Models

Nate Rex Reasch

Utah State University

5-1-2014

Bayesian Inference: Probit and Linear Probability Models

Nate Rex Reasch
Utah State University

BAYESIAN INFERENCE:

PROBIT AND LINEAR PROBABILITY MODELS

by

Nate Rex Reasch

A report submitted in partial fulfillment
of the requirements for the degree

of

MASTER OF SCIENCE

in

Financial Economics

Approved:

Tyler Brough Major Professor	Jason Smith Committee Member

Alan Stephens
Committee Member

UTAH STATE UNIVERSITY
Logan, Utah

2014

ABSTRACT

Bayesian Model Comparison

Probit Vs. Linear Probability Model

by

Nate Rex Reasch, Master of Science

Utah State University, 2014

Major Professor: Tyler Brough

Department: Finance and Economics

The following paper analyzes the benefits of Bayes' theorem in applied econometrics. This is accomplished by demonstrating each step in conducting Bayesian inference. This includes the prior selection, the likelihood function, posterior simulation, and model diagnostics. To provide a concrete example I replicate, by Bayesian inference, the main model of Blau, Brough, and Thomas.(2013) This model is found in their research paper titled, *Corporate lobbying, Political Connections, and the Bailout of Banks.* The analysis focuses on two different forms of limited dependent variable regressions, the probit and linear probability model. The benefits of Bayesian econometrics were extensive and serve as a testament to Bayesian methodology.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Bayesian inference is a method for formulating beliefs about uncertainty. When approaching uncertainty, individuals innately form decisions based on prior beliefs. These beliefs are unique to each individual and can be as vague or specific as the individual chooses. When new information is observed, a rational individual will update their prior belief. This new belief, called the posterior belief, is based on the prior belief conditioned on the new information. The ability to update beliefs is fundamental to Bayesian methodology. This approach is portrayed in Bayes' Rule. Given an event A and evidence B, Bayes' theorem states: $P(A|B) \propto P(A) * P(B|A)$. This is read as: The posterior belief of event A is proportional to the prior belief updated by the likelihood of evidence B. This simple equation is extremely powerful and is unknowingly used in every day life. From basic decision making to approaching complex issues, Bayes' Theorem is an indispensable tool.

This paper analyzes the benefits of Bayesian techniques and demonstrates commonly used practices to illustrate Bayesian inference. The data to conduct this analysis were adopted from the Journal of Banking and Finance. The essay is titled *Corporate Lobbying, Political Connections, and the Bailout of Banks.* This analysis was written by Benjamin Blau, Tyler Brough, and Diana Thomas (BBT) on the possible bias of the Troubled Asset Relief Program. In their research they find that firms with political connections had a high probability of receiving the TARP issued bailout funds. In addition, they found that these same firms were also likely to receive the TARP issued funds more quickly. Testing their models will provide additional evidence to support their theory and be useful in illustrating Bayesian inference. The analysis will begin with the formulation of priors and the simulation of the posterior. Their results will then be verified by a probit regression, and extended with a linear probability model. Finally, informal and formal model checks will be demonstrated.

BAYES THEOREM

A. PRIOR

Theta, θ, will represent a set of parameters (β, h). The prior, P(θ), is an innate belief of the parameters, in the form of a probability distribution, prior to seeing data or evidence. Hence, it is called the prior. The prior can reflect an informative or non-informative belief. An informative prior reflects a subjective and specific prior knowledge. A non-informative prior, called a diffuse prior, represents ignorance. Critics of the prior emphasize the subjectivity it introduces. In order to reduce this bias it is a common practice to use diffuse priors in order to produce convincing arguments. The prior is a powerful tool and can facilitate even the most complex questions.

Prior selection is unique to the individual asking the question. One important aspect to remember is that one has the ability to choose different priors and compare. Model comparison is encouraged and increases the robustness of your argument. Even though different priors may be chosen, Bayesian methodology relies on the idea of convergence. This implies that given an infinite amount of time and data, individuals with different priors should converge on the same inference. This increases the defense against subjectivity. In this analysis, Normal-Gamma priors are assumed. It is important to remember that a prior belief is a personal belief and is represented by a probability distribution.

B. LIKELIHOOD

The likelihood, P(Y | θ), is a joint probability function of the data and parameters. In Bayesian analysis it is viewed as a function of just the parameters. This is due to the assumption that the data are fixed and represent all available information. The likelihood function is chosen based on the probability distributions assumed for the parameters. For example, in the probit regression, the likelihood of a normal distribution was selected. In the linear probability model, a combination of the normal and gamma likelihood functions was selected. Likelihood selection has a direct impact on the posterior distribution in two ways. First, the prior only affects the posterior through the likelihood. Second, the likelihood function expresses all available information regarding the data. Different likelihood functions can be used, but to conduct this analysis commonly agreed upon likelihood functions were chosen. The following represents a Normal-Gamma likelihood function:

$$P(Y|\theta) = \frac{1}{(2\pi)^{\frac{N}{2}}} h^{\frac{1}{2}} exp[-\frac{h}{2}(\beta - \widehat{\beta})^T X^T X(\beta - \widehat{\beta})] h^{\frac{v}{2}} exp[-\frac{hv}{2s^{-2}}] \qquad (2.1)$$

C. *POSTERIOR*

The posterior, $P(Y|\theta)$, is proportional to the prior, $P(\theta)$, multiplied by the likelihood, $P(\theta|Y)$. Instead of simple point estimates, commonly calculated in frequentist theory, an entire distribution is formed for each parameter of interest. The posterior distribution provides abundant information to conduct inference. It also allows for straightforward interpretations and vast amounts of summary statistics. Some important statistics to summarize the posterior distribution are the mean, mode, standard deviation, variance, skewness, kurtosis, and quantiles. The posterior distribution can also be used to conduct various types of model diagnostics, which are discussed later. The benefits of the posterior distribution are extremely advantageous in conducting inference in regression analysis.

While the posterior has numerous advantages, it can be difficult to calculate. To demonstrate this, suppose you have a model with two parameters, X_1 and X_2. Assuming the prior for X_1 and X_2 is P(X_1, X_2), and the likelihood is L(X_1, X_2| Y), Bayes theorem states the following: P($X_1, X_2|Y$) $\propto$ P(X_1, X_2) * L(X_1, X_2| Y). In order to calculate the posterior distribution, numerical integration will be required on each independent distribution. The resulting distribution will have to be normalized to ensure the posterior distribution integrates to one. To simulate from the posterior will require more numerical integration. This numerical integration can be very difficult even with just two parameters. Adding additional parameters would require third and fourth dimensional integration. This can be cumbersome and extremely difficult to calculate.

The introduction of Markov Chain Monte Carlo reduced the calculation of the posterior substantially. A version of MCMC, called the Gibbs Sampler, is used to replace the numerical integration and make use of modern day computers. Using the same example stated above with two parameters X_1 and X_2. Suppose the posterior distribution $\mathrm{P}(X_1, X_2|\theta\ )$ is based on two independent conjugate priors. Assuming independent priors with standard forms, the Gibbs sampler is computed as follows:

1. Fix initial values for X_1^0 and X_2^0.
2. Draw a random value X_1^1 from the conditional distribution $\mathrm{L}(X_1|\ X_2^0, \theta\ )$
3. Then draw a random value X_2^1 from the conditional distribution $\mathrm{L}(X_2|\ X_1^1, \theta)$

4 Continue this process thousands of times.

After a few repetitions, called the burn-in period, these random conditional distribution draws begin to behave like random draws from the marginal distribution. Posterior simulation, which replaces difficult numerical integration, reduces the computational burden of Bayesian econometrics and allows for a wide variety of models to be computed and evaluated. Straightforward interpretations, reduced computation, and an abundance of summarizing statistics makes the posterior distribution an important benefit of Bayesian econometrics.

DATA VARIABLE AND MODEL SELECTION

A. DATA

The data begins with 237 financial firms that received Troubled Asset Relief Program support. (TARP) This data was acquired from the Center for Responsive Politics.(CRP) The data set reports quarterly statements of lobbying expenditures. Using the center for Research on Security Prices(CRSP), the sample size was extended to include all firms with the Standard Industrial Classification Codes (SIC) 60, 61, and 62. These SIC codes represent the majority of financial entities, most importantly banks. This extension led to including 334 firms that did not receive TARP support. The final sample size includes 571 firms along with their matching characteristics.[1]

B. VARIABLE SELECTION

The dependent variable will be $TARPDUM_i$. This variable is equal to one if a firm received TARP support, and zero otherwise. To control for firm characteristics, the following explanatory variables were chosen: $Price_i$, $ln(Size_i)$, $ln(TotAssets_i)$, D/E_i, $Turn_i$, $Volt_i$, and $Dummy_i$. The variable of interest is $Dummy_i$ which is equal to one if a firm had positive lobbying expenditures prior to TARP, and zero otherwise. The coefficient for the variable $Dummy_i$ will be the focus of the analysis. The estimated sign and magnitude will indicate the association between lobbying and the probability of receiving TARP.

C. MODEL COMPARISON

Limited dependent variables normally represent a qualitative choice with an underlying latent utility. Typical ordinary least squares regressions require the distribution of Y given X to be normal. This is usually not the case with a dependent

[1] For Summary Statistics see Journal of Banking & Finance 37 (2013) 3007-3017.

variable of zeros and ones. The assumption can be made that there is an underlying latent variable Y* that is normally distributed. The objective of a limited dependent variable is to derive the posterior of the latent data conditional on the observed data Y. The latent variable Y* is unobserved, Y is observed. The following condition must be met:

$$P(\beta, h|Y*, Y) = P(\beta, h|Y*) \tag{3.1}$$

This infers that if you knew Y*, there would be no additional information added by knowing Y as well. In other words, if Y* was observed, normal regression techniques would be implemented using Y* as the dependent variable. If this condition is met, and Y* is unobserved, we can use the Gibbs Sampler with data augmentation to simulate the posterior distribution of the latent data Y*. The latent data is assumed to be normally distributed in both models. The variable Y is assumed to demonstrate a nonlinear relationship in the probit, and a linear relationship in the linear probability model. The attributes of the probit and linear probability models are described below.

The first technique will be to estimate the same probit regression used by BBT, to verify their results. Probit models have unique assumptions that can capture effects that normal regressions cannot. A probit regression assumes the binary dependent variable is not normally distributed. It also binds the probabilities between zero and one. The assumption is then made that the underlying latent data is normally distributed. The resulting regression is nonlinear. This can capture nonlinear effects when the observed values are only zeros and ones. One weakness of the probit model, due to its nonlinear function, is the interpretation of the coefficients can be difficult. To directly interpret a coefficient, the marginal effect must be calculated. Marginal effects can be calculated in various ways. In this analysis the marginal effects are demonstrated by selecting three individuals with independent utilities and presenting their unique probabilities. Despite the difficult interpretation, the probit regression is very useful and produces nonlinear relationships.

The second technique is the linear probability model. This regression method was chosen to extend the results estimated by the probit regression. The linear probability model is estimated using normal regression techniques to fit a linear regression through the observed zeros and ones. Since the relationship between Y and X is assumed to be linear, the interpretations are straightforward. The weakness associated with a linear relationship is that the estimated coefficients can imply irrational probabilities beyond the interval [0,1]. Testing this model will be useful in providing additional evidence and demonstrating the usefulness of Bayesian techniques.

RESULTS

A. *PROBIT SUMMARY*

The probit model was assumed to have a diffuse Normal-Gamma prior. The posterior was simulated by using MCMC in the form of the Gibbs Sampler. The mean was assumed to represent point estimates of the coefficients. Summary statistics are portrayed in the following table followed by density plots.

Table 1: Probit Model Results

Regression was based on a diffuse Normal-Gamma prior. Each column represents summary statistics for the posterior distributions of each estimated coefficient.

	Intercept	*Price*	*DE*	*Turn*	*Idiovolt*	*Dummy*
Mean	-0.19	-0.002	0.004	-0.50	-0.13	0.49
Std. Dev.	0.18	0.004	0.005	0.39	0.04	0.23
Variance	0.03	.00002	0.00003	0.15	0.002	0.05
Skewness	0.07	-0.04	0.69	-0.66	-0.13	-0.03
Kurtosis	3.03	3.04	3.47	3.29	3.01	3.01
0%	-0.91	-0.02	-0.01	-2.37	-0.28	-0.42
25%	-0.31	-0.004	-0.000	-0.75	-0.16	0.34
50%	-0.19	-0.001	0.001	-0.45	-0.13	0.49
75%	-0.08	0.001	0.007	-0.20	-0.11	0.64
100%	0.45	0.01	0.03	0.36	0.01	1.41
MargEff						25.49%

Figure 1: Intercept Distribution

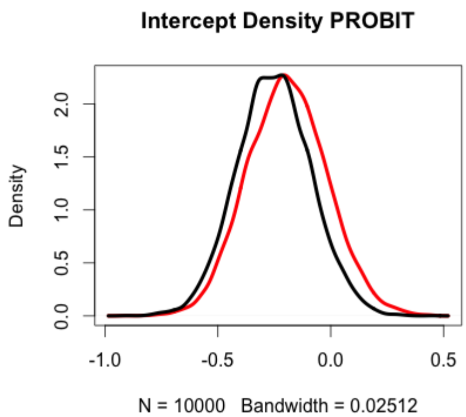


Figure 2: Price Distribution

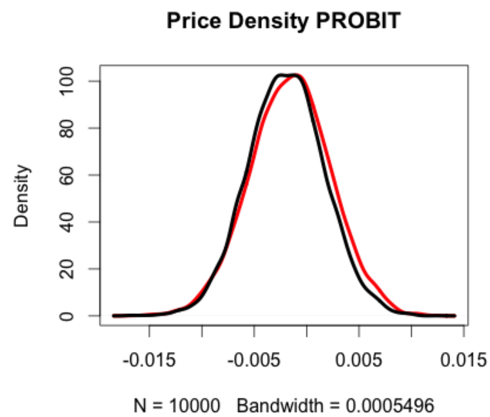


Figure 3: DE Distribution

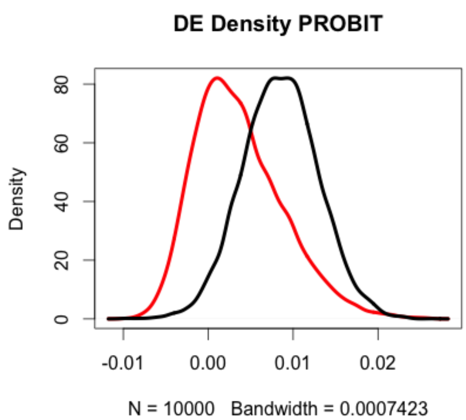


Figure 4: Turn Distribution

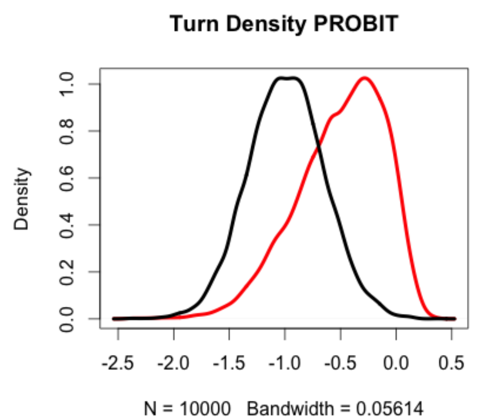


Figure 5: Idiovolt Distribution

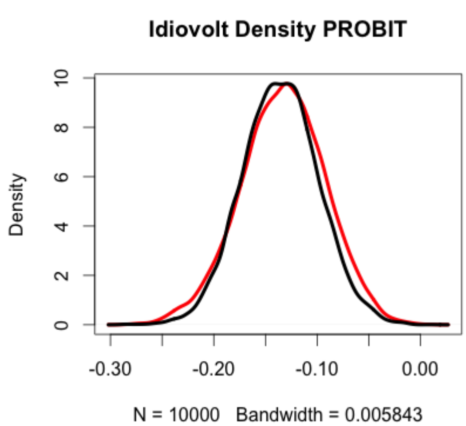
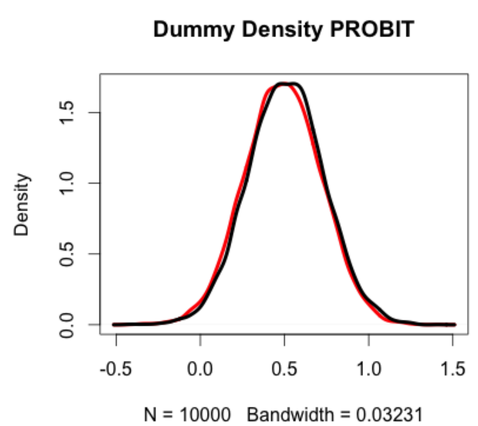


Figure 6: Dummy Distribution



Using the mean as a point estimate of each posterior distribution, the probit estimates are similar to the results found by BBT. As assumed, the coefficients roughly follow the prior distribution. Interpretations are provided in two ways. First, the coefficients are informally interpreted to measure sign impact. This assumes, that in large samples and holding everything else constant, each variable will have a positive or negative impact on the probability of receiving TARP. $Price_i$ in large levels has a negative impact. DE_i has a positive impact. $Turn_i$ has a negative impact. $Dummy_i$, which is the focus of the regression, has a positive impact. This informal interpretation shows that having positive lobbying expenditures prior to 2008, has a positive impact on receiving TARP. Second, the Marginal effect for $Dummy_i$ was portrayed by representative individuals with utilities of zero and one. The mean of the probability distribution was observed to represent the marginal effect of $Dummy_i$. The individual with a utility of one shows a 25.49% increase in probability of receiving TARP. This is strong evidence in favor of the theory that BBT find.

To provide a graphical view of the posterior distributions, density plots were graphed. As assumed, the coefficients roughly follow the prior distribution. There exists a positive skewness in the variable DE_i and negative skewness in the variable $Turn_i$. This is noted but not deemed too important since the focus of the analysis lies on the variable $Dummy_i$. The posterior distribution of the variable $Dummy_i$ is very similar to that of the prior. This is evidence of correct model assumptions for the variable $Dummy_i$. It also provides evidence in favor of the bias found in political contributions.

B. LPM SUMMARY

The results found by the probit regression were extended by using a linear probability model to find any existing linear relationships. The regression was assumed to have a diffuse Normal-Gamma prior. The Gibbs sampler was used to simulate the posterior distributions. The following table and figures represent the results found.

Table 2: Linear Probability Model Results

Regression was based on a diffuse Normal-Gamma prior. Each column represents summary statistics for the posterior distributions of each estimated coefficient.

	Intercept	*Price*	*DE*	*Turn*	*Idiovolt*	*Dummy*
Mean	0.62	-0.01	0.01	0.61	-0.03	2.12
Std. Dev.	62.12	1.61	1.16	56.80	12.02	107.01
Variance	3858.34	2.58	1.34	3226.40	144.47	11450.62
Skewness	-0.02	-0.03	-0.02	0.03	0.05	0.02
Kurtosis	2.97	3.03	2.91	3.00	3.04	3.07
0%	-216.67	-6.13	-4.24	-199.43	-39.72	-420.41
25%	-41.02	-1.07	-0.76	-37.55	-8.08	-69.98
50%	0.36	0.01	-0.01	0.65	-0.08	2.53
75%	43.02	1.07	0.82	39.19	7.83	74.38
100%	239.38	7.02	4.09	221.36	42.51	400.72

Figure 7: Intercept Distribution

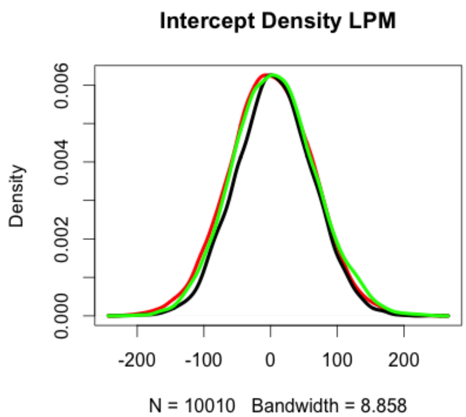


Figure 8: Price Distribution

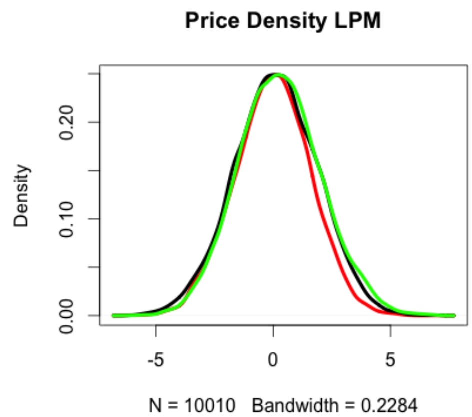


Figure 9: DE Distribution

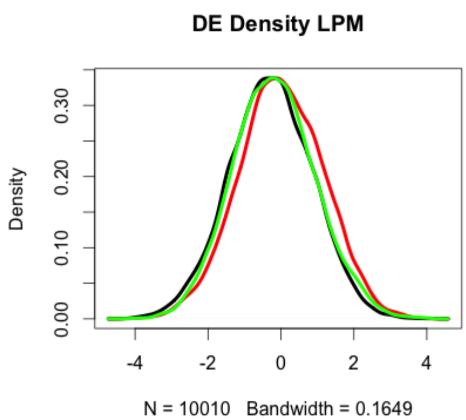


Figure 10: Turn Distribution

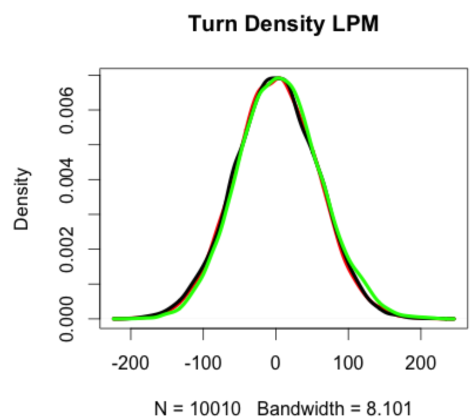


Figure 11: Idiovolt Distribution

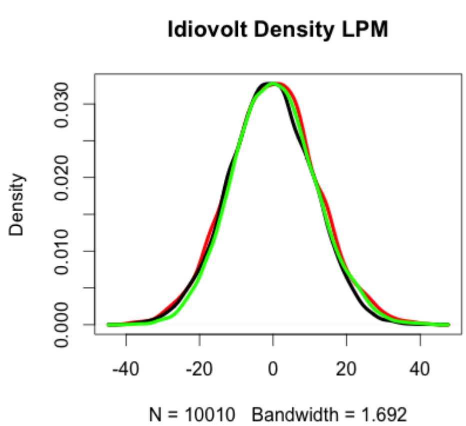
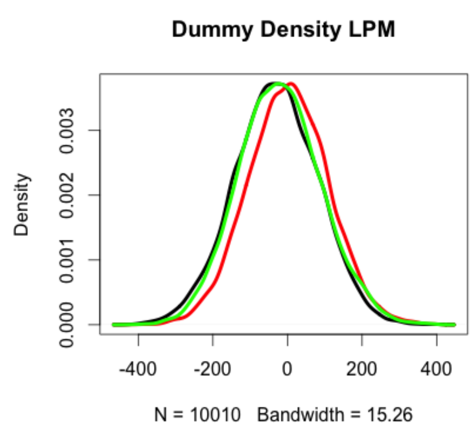


Figure 12: Dummy Distribution



The results estimated by the linear probability model had some flaws. These model estimated irrational probabilities. As discussed before, this is an issue in the linear probability model. Direct interpretation states that an increase in one unit of the variable $Dummy_i$ is associated with a 212% increase in the probability of receiving TARP. This is an unrealistic probability. This leads to the assumption that our prior was not correctly specified, or the relationship is actually nonlinear and using a linear probability model was not the correct assumption. This does not deter from the results found by BBT. It serves as an example for further model comparison with different priors.

MODEL CHECKS

A. INFORMAL

One informal diagnostic technique is to plot the residuals of the regression. The residuals are indicated by the symbol ε_i. The residuals were estimated using the following equation: $\varepsilon_i = Y_i - X_i\widehat{\beta}_i$. The results are shown below in the form of QQ and density plots. The nonlinear QQ plots and the rough density plots may indicate model assumption errors. This leads to the conclusion that additional priors should be tested to better fit the residuals.

Figure 1: Linear Probability Model Residuals Plot

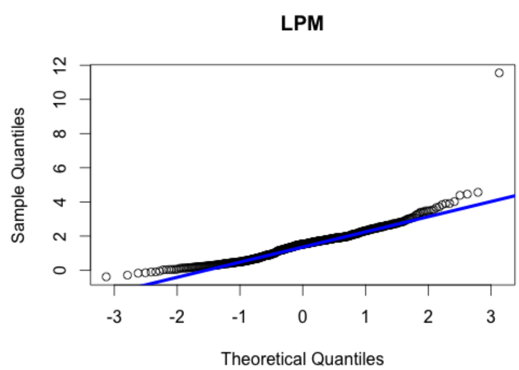


Figure 2: Probit Residuals Plot

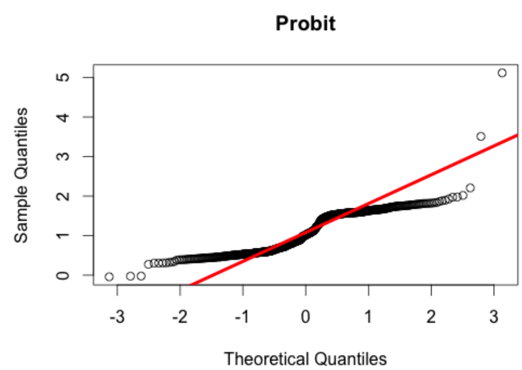


Figure 3: Linear Probability Density Plot

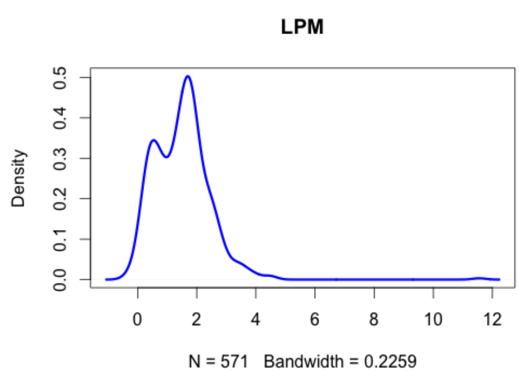
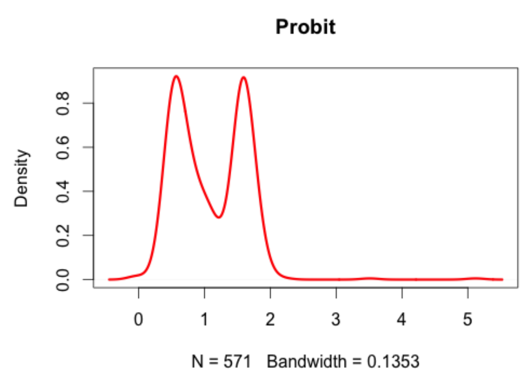


Figure 4: Probit Density Plot



B. PRIOR PREDICTIVE

The first formal approach of model diagnostics is called the prior predictive distribution. It is based on creating a distribution prior to seeing any evidence. This creates a theoretical distribution based on the prior belief. The following two figures are density plots which represent the prior predictive distributions. These portray vague priors that are centered around the hypermean. As observed in the results section, this prior placed on the probit was a good assumption that corresponds to similar results found by BBT. The linear probability model also represents vague priors centered around a hyper mean. The results section estimates irrational probabilities indicating this prior may be misstated. The prior predictive provides beneficial insight on what the prior belief is predicting and should be part of every analysis.

Figure 5: Linear Probability Prior Predictive

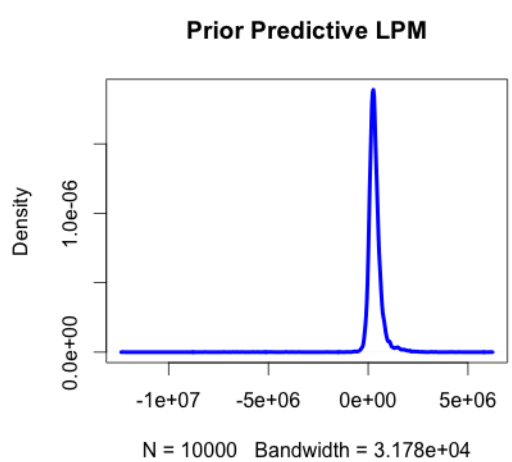
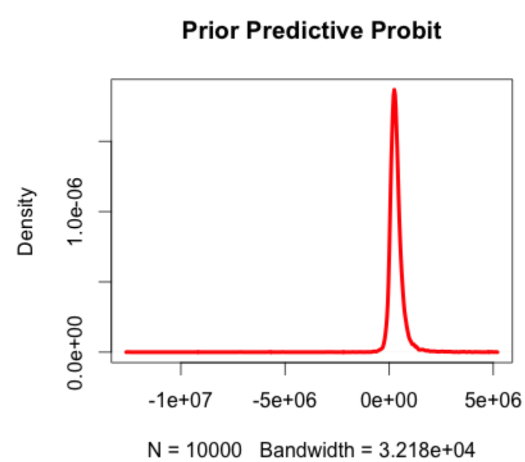


Figure 6: Probit Prior Predictive



C. POSTERIOR PREDICTIVE

Another formal approach is to simulate a predictive posterior distribution. This creates a hypothetical distribution if additional observations were found. This is simulated by creating a hypothetical series of zeros and ones given the assumptions of a model. A ratio of the number of ones predicted divided by number of simulations is calculated. This provides a method of comparing the predictive distribution to the posterior distribution. This ratio represents the number of firms that received TARP. The data has a ratio of $\frac{237}{571} = 0.415$. The probit model has a predictive distribution ratio of $\frac{148}{10000} = 0.0148$. The linear probability model has predictive distribution ratio of $\frac{4989}{10010} = 0.489$. The density plots are shown below. These ratios find evidence that the probit does not have a good predictive posterior distribution and the linear probability model does. This indicates that there may be a better model that is more compatible with the data. Additional models with different priors should be estimated and compared to find the best predictive model.

Figure 7: Linear Probability Posterior Predictive

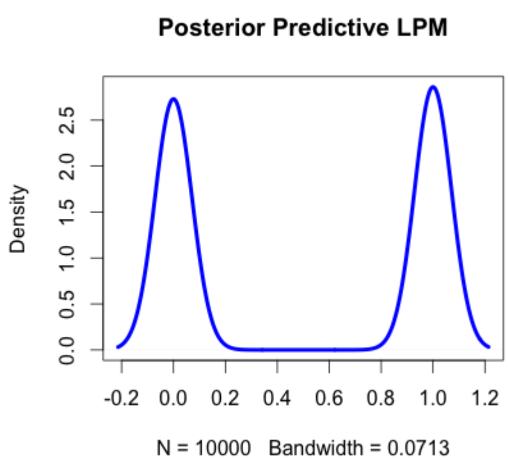
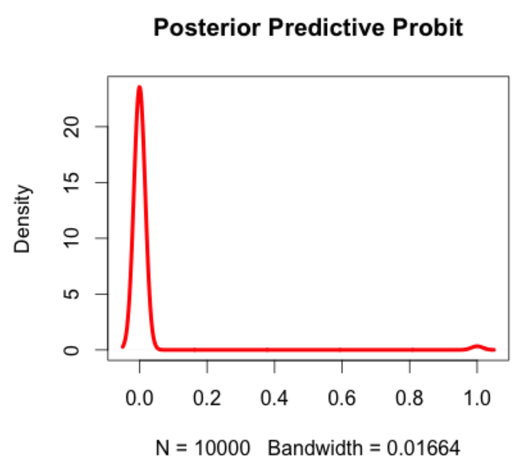


Figure 8: Probit Posterior Predictive



CONCLUSION

To extend the results found by BBT, a linear probability model was estimated. The model was found to estimate irrational results indicating possible errors in the prior. The model diagnostics were clear and consistent and lead to the conclusion of a nonlinear relationship. This does not deter from the results found by BBT, instead it indicates additional models need to be estimated using different priors. The results estimated by the probit regression were similar to the results found by BBT. While the posterior predictive and residuals were not consistent, the results did indicate that political lobbying had a positive impact on the probability of receiving TARP. This serves as additional evidence to support their theory.

Further research would include Bayes' Factor and model averaging. These could resolve the issues found during this analysis. Bayes' Factor involves directly comparing two different models. Model averaging allows an econometrician to conduct inference from more than one model. These additional topics could extend the research and increase the robustness of the argument.

Bayesian econometrics is a beneficial method that can be applied in any situation. It allows an econometrician to form a prior, condition on the likelihood, and form a posterior distribution. The posterior distribution provides straightforward interpretations with an abundance of information to conduct inference. Applying Monte Carlo methods allows for posterior simulation. The benefits of Bayesian econometrics were demonstrated and found to be very useful in providing additional evidence to support BBT. Bayesian methodology is an indispensable learning process that should be applied in everyday life and in any econometric analysis.

REFERENCES

Benjamin Blau, Tyler Brough, Diana Thomas, *Corporate Lobbying, Political Connections, and the Bailout of Banks.* Journal of Banking & Finance 37 (2013) 3007-3017

Yu, F., X, Yu., 2011. *Corporate Lobbying and Fraud Protection.* Journal of Financial and Quantitative Analysis 46, 18651891.

Tony Lancaster (2004) *An Introduction to Modern Bayesian Econometrics.* Blackwell Publishing. ISBN 1-4051-1720-6

Kevin Boone (2013) *Kevin Boones Blog.* Copyright 1994-2013 www.kevinboone.net

Gary Koop (2003) *Bayesian Econometrics.* John Wiley & Sons, Ltd. ISBN 0470-84567-8-448pp