5-2014

An Exercise in Bayesian Econometric Analysis Probit and Linear Probability Models

Brooke Jeneane Siler

Utah State University

5-1-2014

An Exercise in Bayesian Econometric Analysis Probit and Linear Probability Models

Brooke Jeneane Siler
Utah State University

AN EXERCISE IN BAYESIAN ECONOMETRIC ANALYSIS
PROBIT AND LINEAR PROBABILITY MODELS

by

Brooke Jeneane Siler

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

of

MASTER OF SCIENCE

in

Financial Economics

Approved:

Tyler Brough Major Professor	Benjamin Blau Committee Member

Ryan Whitby
Committee Member

UTAH STATE UNIVERSITY
Logan, Utah

2014

ABSTRACT

An Exercise in Bayesian Econometric Analysis

Probit and Linear Probability Modelsl

by

Brooke Jeneane Siler, Master of Science

Utah State University, 2014

Major Professor: Tyler Brough

Department: Finance and Economics

The aim of this paper is to carry out a Bayesian econometric application. Using a dataset obtained from Wooldridge's *Introductory Econometrics* textbook, each step in conducting a Bayesian econometric analysis is performed and explained. For illustrative and comparative purposes, two limited dependent variable regression forms were used: a linear probability model and a probit model. This paper covers the benefits of Bayesian methodology, including selection of distributions for the prior and the likelihood. Additionally, a series of diagnostic checks are done after the models are computed.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The probability of the occurrence of an event is rarely static. For instance, I may state that I believe the probability of a sports team winning a competition is .5, but soon after I make this statement it is revealed that the star player of the team can no longer play due to an injury. As a rational person, I would update the probability of winning to less than .5, which is significantly different than the probability stated before. Bayes' rule is a method to compute probabilities in a changing world. Mathematically stated as $P(\theta|Y) \propto P(Y|\ \theta)P(\theta)$, the theorem relies on the use of conditional probabilities. To compute a posterior belief of probability, a prior belief, $P(\theta)$, is multiplied by a likelihood, $P(Y|\theta)$, which is an expression for the distribution of the data observed. Bayes' theorem and Bayesian inference has increasingly been used as a substitute to frequentist methods in the creation and evaluation of econometric models.

In this paper, I will give a brief overview of the Bayesian algorithm and apply it to a data set obtained from *Introductory Economics: A Modern Approach*, written by Jeffrey Wooldridge. The data set consists of married women's labor force participation rates and is used in example 17.1 of the textbook. Wooldridge uses a number of variables to describe the probability of a woman participating in the labor force and computes the same model using both a linear probability model as well as a probit model. In my analysis, I will duplicate this study using Bayesian econometric techniques and numerical methods such as the Gibbs Sampler. After a description of the techniques, I present the results of the study and compare it to the frequentist econometric models published by Wooldridge. Following the results, I discuss methods of checking the Bayesian models both formally and informally and present the results of the checks on the models presented in the paper.

UTILIZING BAYES' THEOREM

A. OVERVIEW

There are two dominant approaches to statistics, the frequentist method and the Bayesian method. Frequentist statisticians treat a hypothesis as fixed and the data as random. That is, a frequentist believes in an unknown or unseen true population distribution of data that can be repeatedly sampled in order to produce estimates of the population distribution. Because statisticians using this approach are concerned with the frequency of which the sample data is observed given the true population, this method has been labeled the frequentist method. The other dominant statistical approach is the Bayesian approach. Bayesians treat the data as fixed (you only have one set of data) and test their theories by treating the hypothesis as variable. Given a set of data, Bayesians seek to test the probability of a hypothesis to be true. The Bayesian approach utilizes Bayes theorem to calculate the probability of the truth of hypotheses.

Bayesian inference is a method to evaluate uncertainty. Given a set of data, any statistician can easily compute the relative frequencies. However, this is simply describing the relative frequencies of events of interest that occurred in the past. These frequencies are not yet probabilities and additional assumptions must be made to treat them as probabilities. An economist is interested in the probabilities associated with the occurrence of events in the future. Up until this point, both Bayesian and frequentist statisticians agree on methods and techniques. But the parties have differing views on how to turn these descriptive statistics into probabilities. A frequentist makes the assumption that the sample he has collected is a good representation of the population in general. In the future, the frequentist expects the same distribution as the past sample. A Bayesian, however, believes that given the fact that we don't know if the sample is truly representative, we should accept our current belief (start with a belief about the probability without seeing the data), and then let the data update our belief to create the posterior belief. Bayesian econometrics uses distributions for each of these: the prior, the likelihood (representation of the sample), and the posterior because it yields a more complete picture of the structure of the uncertainty. The posterior distribution allows us to estimate the error in our computed probabilities.

Because Bayesian inference conditions on the data as fixed, it expands the ability of the econometrician to estimate the probabilities of hypotheses that are unrepeatable. Well known probabilities that have been estimated using Bayes theorem are: (1) Is the moon made of green cheese? (2) What is the probability that God exists? (3) Is the United States in a recession?

Bayesian probability and inference are valuable tools to use in econometric model building. In addition to the previously mentioned merits, Bayesian analysis can serve as a robustness check to a model that has been calculated using frequentist methods. Because Bayesian econometrics allows for the checking of diverse priors, models can be shown to be robust to a variety of previous beliefs about a variable. In the following sections, I will detail the Bayesian algorithm and each of its components.

The Bayesian method can be easily turned into an algorithm for computation:

1. Formulate the economic model to be tested by collecting probability distributions conditional on values for the model parameters.

2. Form a prior beliefs into a probability distribution.

3. Collect the data and insert them into the distributions determined in step 1.

4. Use Bayes theorem by multiplying the distributions of the prior and the likelihood.

5. Check your model for robustness and criticize.

B. *PRIOR*

Formally expressed as P(θ), the prior is a distribution based on one's subjective beliefs about the parameters before observing data. Because it is subjective, one can choose any prior she would like. A variety of prior beliefs can be checked in multiple models to observe the effect of the prior distribution on the posterior distribution. If one is likely to present the model to other researchers, it may be useful to ensure that the prior does not conflict with the beliefs of your reviewers. For this reason, it may be wise to choose a prior distribution that is consistent with the beliefs of the industry that one is studying. It would not be wise to choose a prior that a coin has a 0.25 probability of producing a heads if one is presenting to a society that believes every coin is fair. To avoid this inconsistency, it is common for researchers to choose a vague prior that is not inconsistent with many beliefs.

Prior distributions, when multiplied by the likelihood, produce a posterior distribution. Many Bayesian econometricians prefer to choose a prior that produces a natural conjugate posterior distribution. This requires the posterior distribution to be in the same family of distributions as the prior distribution. Before advanced numerical techniques of integration, natural conjugate priors were very popular. Today, any prior can be chosen and the posterior estimated using simulation.

Other popular priors are called improper priors, which are named improper because the integral over the distribution does not integrate to unity. In fact, because it does not integrate to unity, it is technically not a probability distribution. It is possible to have a proper posterior distribution given an improper prior distribution. During preliminary analysis (or simply to save time), improper or vague prior distributions are often used to test the posterior results without having to specify exact beliefs of the prior.

In summary, a prior can be of any form one chooses. The distribution represents a belief of the probability before seeing data. It can be proper, improper, a natural conjugate, informative, dogmatic, or vague. It is common to use a vague prior in preliminary model testing or when the econometrician is unsure about what his/her prior ought to be. In the example analysis of the Wooldridge data below, a normal-gamma prior was chosen for the linear probability model and a normal prior was chosen for the probit model.

C. LIKELIHOOD

The likelihood is expressed in Bayes' theorem as P(Y | θ). This is the mathematical expression for the function of θ with the values of the observed data serving as values of the parameters. The choice of the likelihood function can come from any distribution. One distribution is chosen for each variable in the model. To choose these distributions, it is important to consider that the likelihood function must express the inherent economic model that you are interested in observing. The likelihood function for a linear model is therefore different than the likelihood that would be chosen for a probit or logit model.

Because Bayes theorem, when implemented in a Gibbs sampler, uses the term proportional ($\propto$) rather than an equality, the full likelihood is not necessary —we only need to concern ourselves with the kernel of the likelihood. After the kernel of the posterior has been calculated, solve for the constant that must be multiplied by the kernel to integrate the posterior distribution to a value of one.

D. POSTERIOR

The posterior distribution, stated in Bayes' theorem as P(θ | Y), is the result of the Bayesian algorithm. This distribution represents ones beliefs given both the prior and the likelihood and is the final result of the study. Once the posterior distribution is calculated, it is typical to report the moments of the distribution. STypically, note the mean, median, and standard deviation of the posterior. It may also be useful to draw the posterior distribution if the distribution is a scalar value.

One useful aspect of the posterior distribution is the interpretation of the confidence interval. In frequentist statistics, the interpretation of a 95% confidence interval is that, given repeated samples, 95% of the intervals sampled given this parameter would contain the true value of the parameter of interest. Alternatively, the statistician can conclude that she is 95% confident that the true mean beta value of the parameter lies between the values stated in the 95% confidence interval. In the Bayesian sense, however, the equivalent is simply a statement of probability. The 95% highest density interval for a Bayesian is an interval such that with probability equal to .95 that the true value of the parameter lies within the interval.

Due to the distributions of the parameters chosen for the likelihood and the prior, especially if the prior is a joint prior, it can be difficult to simulate the posterior. This is because the resulting distribution from a joint prior and a likelihood after using Bayes theorem is not in standard form. To simulate from the posterior distribution, numerical integration is necessary, marginal values must be simulated, and then the distribution must be normalized with a constant to ensure integration to a value of unity. When dealing with joint priors and cumbersome numerical integration, an alternative approach may be used to simulate the posterior distribution.

As computing technology has advanced, the use of Markov Chain Monte Carlo (MCMC) has been used to simplify the calculation of the posterior distribution. One specific version of MCMC is known as the Gibbs Sampler, which allows the econometrician to replace numerical integration with repeated marginal sampling. Rather than trying to integrate, the Gibbs sampler draws a random value from the conditional distribution for each of the parameters of the function. In this case, the draws from the other parameter distributions serve as the condition for the next parameters random draw. We use the Gibbs sampler to transition from the conditional distribution to the marginal distribution.

An example of the Gibbs sampler is as follows: Consider the case where you have two parameters, θ_1 and θ_2 and a non standard joint posterior p(θ_1, θ_2 | y). To calculate the marginal distribution, complete the following steps:

1. Begin with an initial value for the parameters θ_1 and θ_2, labeled ${\theta_1}^0$ and ${\theta_2}^0$.

2. From p(θ_1 | ${\theta_2}^0$, y), draw a random value ${\theta_1}^1$.

3. Now obtain a random value ${\theta_2}^1$ from p(θ_2 | ${\theta_1}^1$, y).

4. Next, use ${\theta_2}^1$ to draw random value ${\theta_1}^2$ from p(θ_1 | ${\theta_2}^1$, y).

5. Repeat this process many times.

This process is called the Gibbs sampler and enables econometricians to obtain the marginal posterior densities for each parameter. After many replications of this process, the draws start to behave as if they are from the marginal posterior distributions instead of the conditional distributions. Because the user assigns the initial values of the parameters, the first few draws must be discarded, as the sampler has not yet begun to converge to the marginal distribution of the posterior. This is called the burn-in sample. After the burn-in, draws from the Gibbs sampler act as if they came from the marginal posterior distribution. Simple summary statistics of the distribution can then be used on the data to summarize the parameter values.

MODEL SELECTION

A. DATA

The data used in this paper is from *Introductory Economics: A Modern Approach*, written by Jeffrey Wooldridge. The data set is of married womens labor force participation rates and is used in example 17.1 of the textbook. Links to the data set can be found on the Wooldridge Introductory Economics website. Wooldridge sources the data from a paper written by Thomas A. Mroz published in 1987 titled, *Sensitivity of an Empirical Model of Married Womens Hours of Work to Economic and Statistical Assumptions*. The data used by Mroz was the Panel Study of Income Dynamics 1975 labor supply data. The data set has 753 observations and 8 variables.[1]

B. VARIABLE SELECTION

The dependent variable used in this study (and in the Wooldridge comparison study) was *inlf* (in labor force), which is equal to one if the woman sampled reported working outside the home at any point in the year, zero otherwise. The independent variables used in this study are the earnings of the husband (measured in thousands and denoted as *nwifeinc*), years of education (*educ*), previous experience (*exper*), age of the married woman (*age*), number of children below the age of six (*kidslt6*), number of kids older than six but younger than 18 (*kidsage6*). Other variables in the data set but not used are *hours* (hours worked up to 1975), *wage* (estimated wage per hour), *repwage* (self reported wage in an interview in 1976), *hushrs* (hours worked by the husband in 1975), *husband's age*, *husband's education* (measured in years), *family income*, and the *unemployment rate* (in the county of residence).

In this model, we wish to determine if the models evaluated using Bayesian econometrics tell a consistent story with the Wooldridge models and evaluate the

[1] For Summary Statistics see Econometrica 55 (1987) 765-799.

difference between different Bayesian limited dependent variable models.

C. MODEL COMPARISON

A limited dependent variable model is a model with dependent sample data that only shows a zero or a one. The econometrician is not as interested in the value of zero or one as much as they are interested in the underlying index variable that is indicated by the dependent variable. The assumptions of an OLS regression require that the distribution of the dependent variable given the independent variables to be normally distributed. In the case of a limited dependent variable model, this assumption fails. However, the econometrician is interested in the latent data (the index variable underlying the dependent variable, labeled as Y*) and assumes the latent data to be normal. In the case of the labor force model, the latent variable Y* is the underlying utility associated with working in the labor force for a married woman. Bayesian inference can be carried out on these models if we can calculate the posterior distribution of the latent data conditional on the observed data and the chosen model parameters. This is done using a Gibbs sampler. This model and the Gibbs sampler can be used, given an independent normal-gamma prior, as long as the following assumption about the latent data holds true:

$$p(h|y^*, y, \beta) = p(h|y^*, \beta) \tag{3.1}$$

The interpretation of this assumption is, if you know the latent data, knowing y as well will provide you with no additional information. Note that $\theta = \{h, \beta\}$.

In this paper, I evaluate a linear probability model and a probit model using a Bayesian methodology and compare the results with each other. Linear probability models are commonly used for their simplicity of quickly observing estimated probabilities of the occurrence of an event. Because standard ordinary least squares is used in the calculation of the model, the resulting coefficients are directly interpretable. However, the model is flawed because it can yield probabilities far outside the acceptable range of $0<P<1$. Due to constraining the probabilities of events between the values of zero and one, most models that have a limited dependent variable are nonlinear. The linear probability model does not allow for nonlinearities and is commonly criticized because it can commonly generate biased and inconsistent results.

The probit model typically is used when the dependent variable indicates an outcome in one out of two categories when the individual is making a choice. Because each choice has an underlying utility function, the choice y^* signifies the difference in the utility functions. The data observed indicates the choice, not the underlying utility. The model can be used in the place of the linear probability model because it allows for nonlinearities. The probit model uses the standard normal CDF to calculate $X'\beta$ and then maximum likelihood is used to solve for the coefficients (in frequentist methods). Probit model coefficients are not directly interpretable due to their nonlinear nature. Marginal probabilities must be calculated to interpret the model.

RESULTS

A. *PROBIT SUMMARY*

When using Bayesian techniques to calculate the posterior distribution of a probit, there exists an identification problem when you assume a normal-gamma independent prior. This is because multiple values for different model parameters will give rise to identical values for the likelihood function. Essentially, you can estimate the same model that yields different coefficients and errors each time. To solve the identification problem, it is standard to set the gamma prior equal to one.

The results observed from the Probit model are similar in sign and magnitude to those that were seen by Wooldridge in example 17.1. Summary statistics are described below in table 1.

Probit model estimates are compared to results found by Wooldridge by comparing coefficients of the Wooldridge model (point estimates) and the mean of the posterior distribution as a point estimate of the Bayesian model. To begin, an informal comparison of coefficient sign is conducted. This allows us to determine if the same relative impact is occuring for each variable. The sign of *nwifeinc* is negative. Similarly, coefficients for the *intercept*, $exper^2$, *age*, and *kidslt6* are also negative. Coefficients for *kidsage6*, *age*, *education*, and *experience* are positive.

To more accurately depict the results of the Probit model, the density plots of each coefficient were graphed. The black line shown represents the prior distribution and the red line represents the posterior distribution for each variable. As you can see, the posterior and the prior distribution do not differ dramatically in any of the distributions for the variables.

Table 1: Probit Model Results

Regression used a noninformative normal-gamma prior and a Gibbs sampler. Summary statistics report the characteristics of the distribution of each independent variable. Note that Freq is the column with the frequentist point estimates and FrTv is the t-statistic for each frequentist point estimate.

	kidslt6	*kidsage6*	*age*	*educ*	*exper*	*nwifeinc*	*exper*^{2}	*int*
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Freq	-0.2612	0.011	-0.0159	0.0394	0.0371	-0.0036	.0371	0.27
FrPv	-7.33	0.83	-6.24	5.18	6.59	-2.48	-3.14	0.53
Mean	-.388	.0265	.02178	.05194	.05966	-.00480	-.00100	-0.5011
SDev	.1009	0.03569	0.0069	0.0199	0.0160	.0041	.0005	.4153
Var	0.0102	0.0013	4.699e^{-5}	0.00039	0.00026	1.700e^{-5}	2.533e^{-7}	0.1725
Skew	-0.0452	-0.00125	0.00031	0.0208	0.05332	-0.03009	-0.0676	0.00108
Kurt	0.00107	0.04426	0.0606	0.01713	-0.0838	-0.02424	-0.04713	-0.0623
Min	-.77119	-0.1028	-0.0521	-0.0249	0.0029	-0.0217	-0.00299	-2.0463
Med	-0.3888	0.0265	-0.0218	0.0519	0.0594	-0.0048	-0.0009	-0.4999
Max	-0.0462	0.1884	0.00264	0.13795	0.12612	0.0119	0.0006	1.0264
IQR	0.1339	0.0469	0.0092	0.0269	0.02199	0.0057	0.0007	0.5666

Figure 1: Intercept

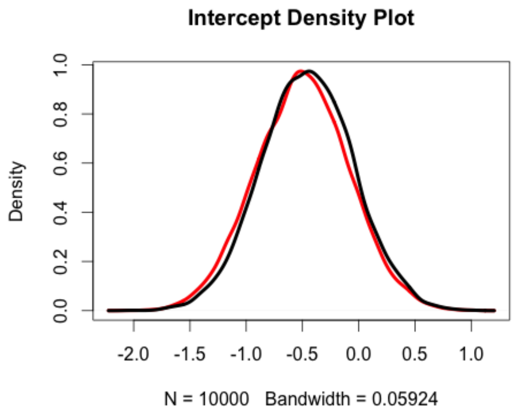


Figure 2: Kidslt6

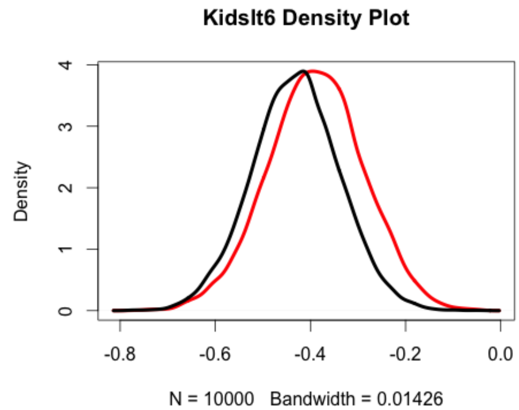


Figure 3: Kidsage6

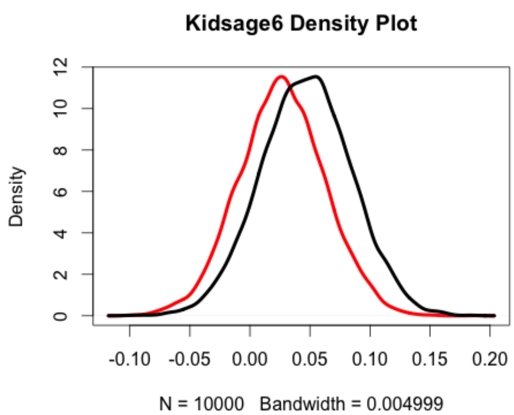


Figure 4: Age

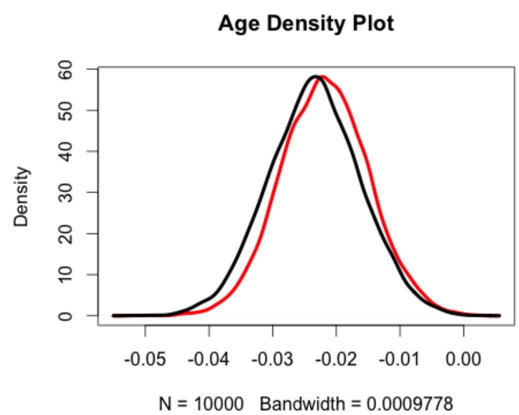


Figure 5: Education

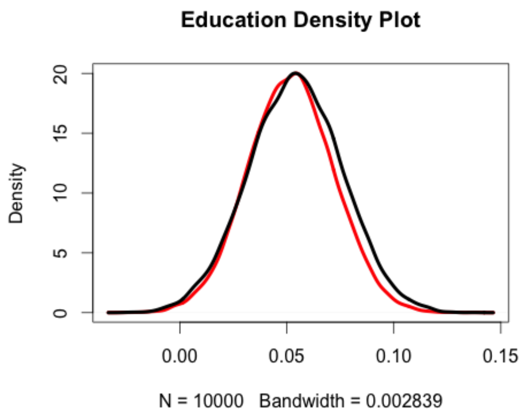


Figure 6: Experience

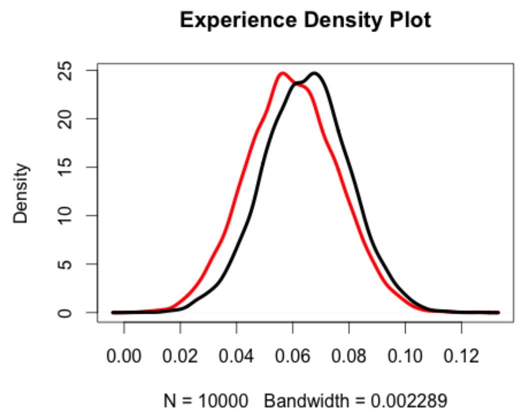


Figure 7: Nwifeinc

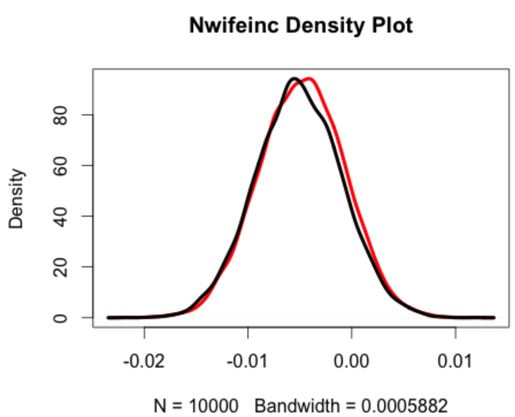
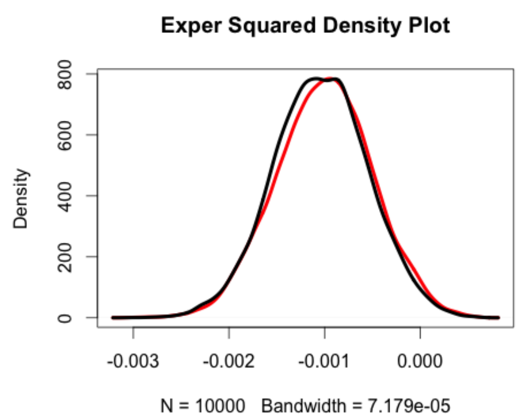


Figure 8: Exper^2



B. LPM SUMMARY

The linear probability model used a noninformative normal-gamma prior and a Gibbs sampler. Summary statistics report the characteristics of each independent variable. The linear probability model does not have the identification problem that was present in the probit model. The results observed in this model are quite similar to the probit model summarized above. See table 2 below to view distributional characteristics of each parameter. The black line represents the normal prior, the blue line represents the gamma prior, and the red line represents the posterior distribution for each variable. Similar to the results seen in the probit model, none of the distributions differ dramatically from eachother.

Table 2: Linear Probability Model Results

Results of the Linear Probability Model are reported below. Note that Freq is the column that represents the point estimates for the frequentist LPM and FrTV is the corresponding t statistic for each estimate.

	kidslt6	*kidsage6*	*age*	*educ*	*exper*	*nwifeinc*	$exper^2$	*int*
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Freq	-0.262	0.013	-0.016	0.038	0.039	-0.038	-0.0006	0.5856
FrTv	-7.81	0.99	-6.48	5.15	6.96	-2.35	-3.23	3.8
Mean	$8.339e^{-3}$	-0.1281	0.0206	0.0996	0.0105	-0.0224	-0.0006	-1.063
SDev	$4.062e^{1}$	15.955	2.994	8.873	6.811	1.747	0.222	$1.852e^{2}$
Var	$1.650e^{3}$	254.57	8.966	78.728	46.392	3.053	0.049	$3.431e^{4}$
Skew	$4.576e^{-2}$	0.0062	-0.0038	0.0223	-0.0199	-0.01425	0.0386	$02.01e^{-2}$
Kurt	$2.723e^{-2}$	-0.0066	0.1176	0.00329	0.0362	0.0051	-0.021	$1.164e^{-1}$
Min	$-1.51e^{2}$	-58.394	-10.341	-31.796	-26.819	-6.294	-0.8429	$-7.982e^{2}$
Med	$-6.194e^{-1}$	-0.0918	0.02611	0.1528	-0.0029	-0.0117	-0.0023	1.298
Max	$1.716e^{2}$	62.611	13.331	33.57	24.934	6.225	0.8454	$7.611e^{2}$
IQR	$5.402e^{1}$	62.611	13.331	33.570	24.934	7.225	0.8454	$7.611e^{2}$

Figure 9: Intercept

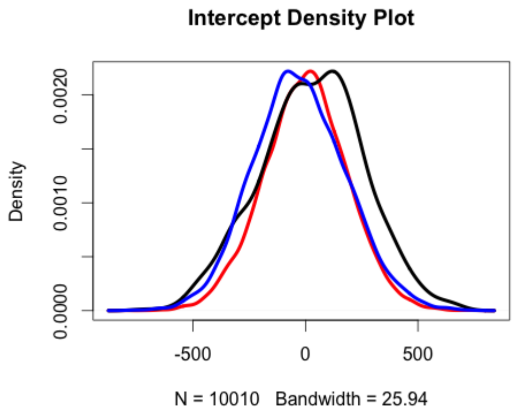


Figure 10: Kidslt6

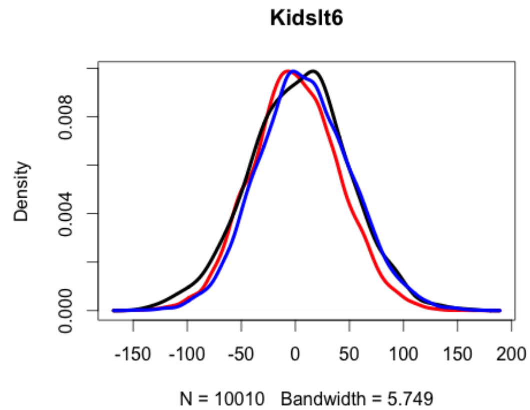


Figure 11: Kidsage6

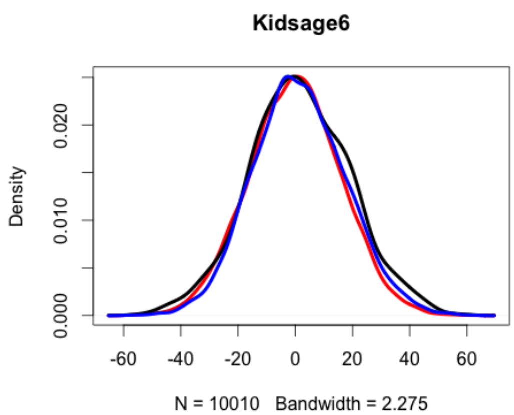


Figure 12: Age

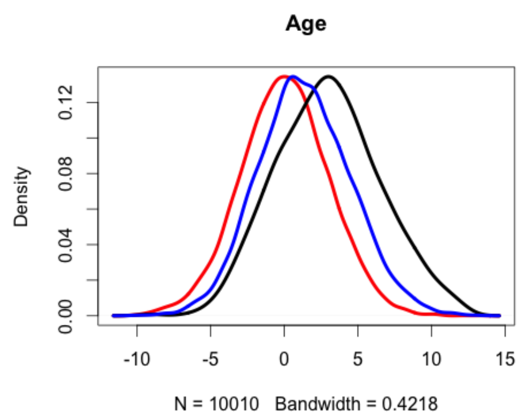


Figure 13: Education

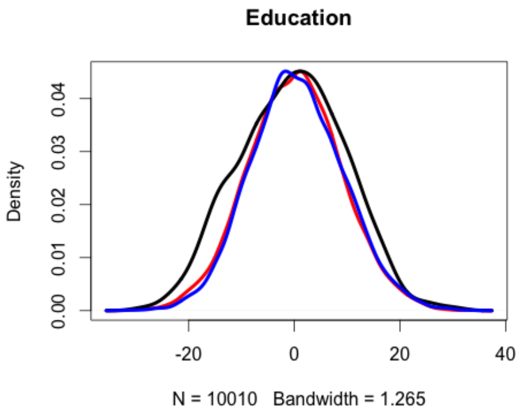


Figure 14: Experience

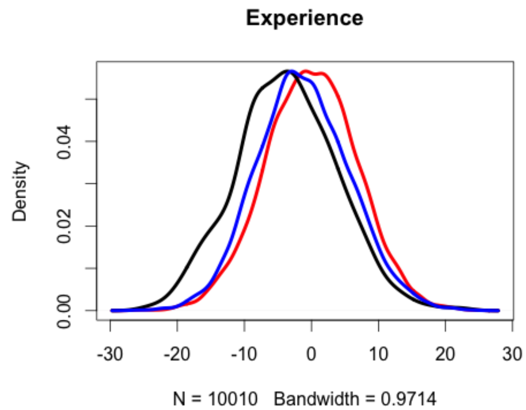


Figure 15: Nwifeinc

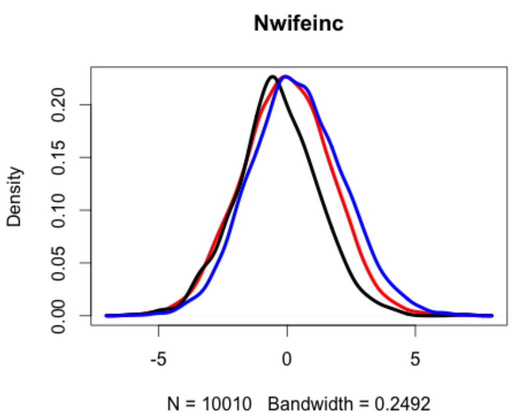
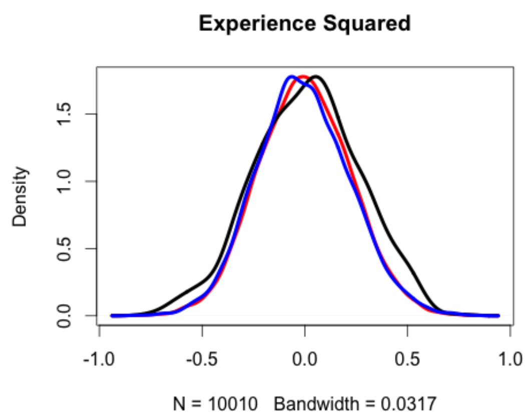


Figure 16: Exper^2



MODEL CHECKS

A. INFORMAL

Simply creating a econometric model for inference is not enough to ensure its usefulness. No sooner than the model is created, it must be criticized and checked for errors. It is common in Bayesian analysis to conduct both formal and informal model checks to ensure robustness and proper model fit. Plotting the residuals of each regression is one way to informally check a model. Residuals, e_i, are calculated from the equation $y_i = X_i\hat{\beta}_i + e_i$. Once residuals are calculated, it is common practice to plot Q-Q plots and also residual density plots to determine model fit. Q-Q plots are used to plot the sample quantiles against the theoretical quantiles. In theory, this plot should produce a straight line. Systematic deviation from this line indicates a poor fit. Residual density plots are used to determine if the residuals are normally distributed and whether or not there are two underlying distributions in the model (errors would appear to be bimodal). The results of my Q-Q plots and Residual density plots can be seen below. Both the LPM and the Probit Q-Q plots fit well in the mid range, but fit poorly at the extremes. This may be evidence of poor model fit. The triple hump seen in the LPM residuals density plot may indicate that the data do not consist of one mean, rather they are trimodal. Again, this may be evidence of poor model fit.

Figure 1: LPM Q-Q Plot

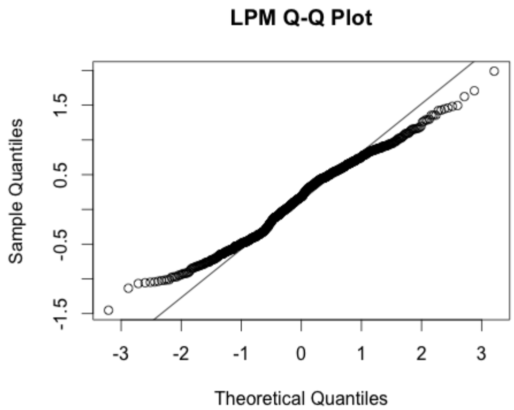


Figure 2: Probit Q-Q Plot

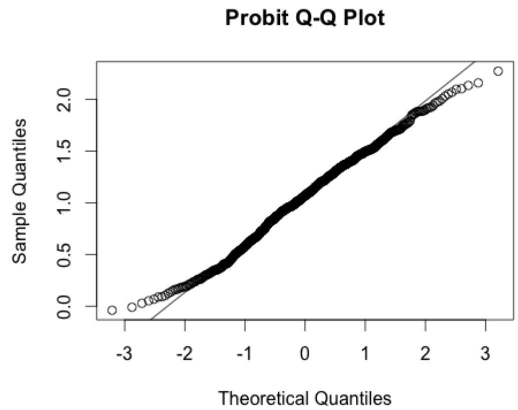


Figure 3: LPM Residuals Plot

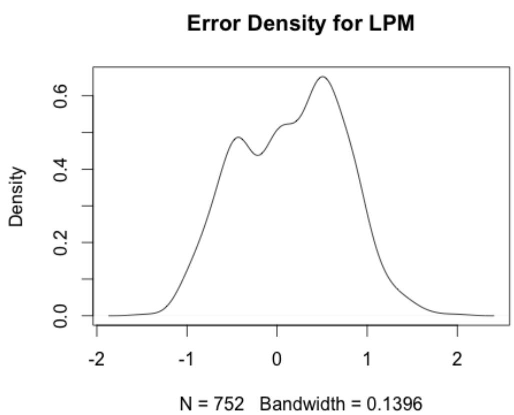
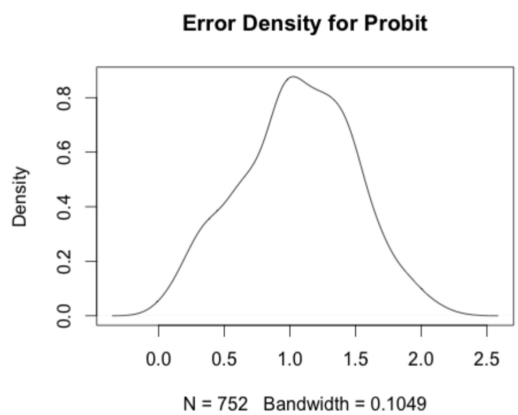


Figure 4: Probit Residuals Plot



B. FORMAL

As part of the model analysis, two formal model checks were performed. Both of these model checks involve using expected distributions to analyze the fit of the model.

The first formal model check performed is the prior predictive distribution. This distribution, given the prior distribution you have selected, is what you expect your data to look like before you have seen the data. This is also referred to as the marginal likelihood. Mathematically, the prior predictive distribution can be calculated by the equation:

$$p(y) = \int p(y|\theta)p(\theta)d\theta \tag{5.1}$$

This theoretical distribution can be compared to your data, as the prior predictive distribution hints at what the data should look like given your prior belief. If the data do not resemble the prior predictive distribution, there may be something incorrect with the model or you may stand to learn a lot from the data.

As seen in the figure below, the prior predictive distribution for the LPM and Probit center around zero. This predictive distribution has a large variance due to the vague prior distribution that was purposefully chosen. This vague prior is a good starting point to any Bayesian analysis, as choosing too specific of a prior may produce inaccurate results if the model does not have enough data to converge. The prior predictive distribution is an important part of Bayesian analysis and model checking.

The second formal model check that was performed is the posterior predictive distribution. After creating a model and seeing the data, the creation of a posterior distribution allows the econometrician to predict what another realization from another data point should indicate. The posterior predictive distribution can be mathematically summarized as:

$$p(y|y) = \int p(y|\theta)p(\theta|y)d\theta \tag{5.2}$$

Density plots of the posterior predictive probit and LPM can be seen below. These plots are quite different from eachother. The posterior predictive LPM indicates that the frequency of ones and zeros is approximately equal while the probit posterior density plot indicates that the realization of a one is much less likely than that of a zero. This indicates that there may be a better model to fit the data. Additional models with various priors must be tested before a conclusion is made of the best fitting model.

Figure 5: Prior Predictive - LPM

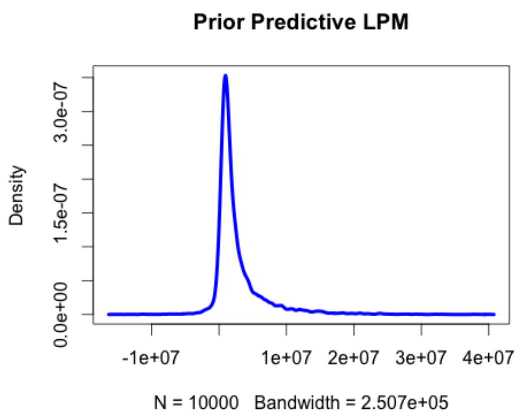


Figure 6: Prior Predictive - Probit

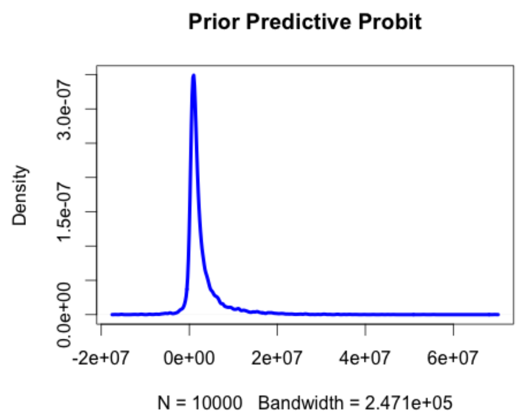


Figure 7: Posterior Predictive - LPM

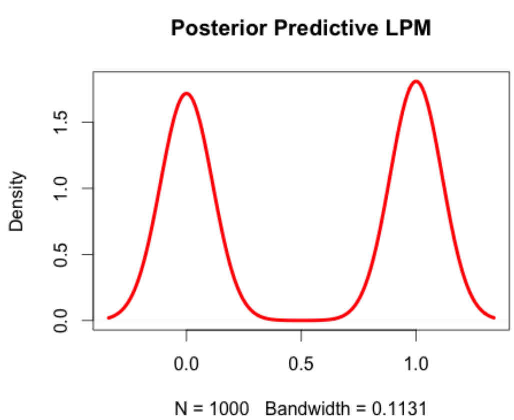
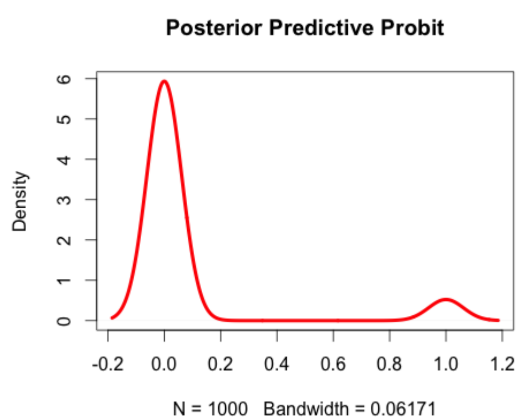


Figure 8: Posterior Predictive Probit



CONCLUSION

In this paper, Bayes' theorem was introduced in the context of econometrics. While frequentist statistical methods continue to dominate financial econometrics, Bayesian inference and models can serve an important role in estimating difficult models or adding robustness to a current frequentist model. Simply the act of attempting to understand Bayesian statistics and econometrics can make an individual a more thoughtful practitioner.

To illustrate the methods described in the paper, data found in chapter 17.1 of Wooldridge's *Introductory Econometrics* textbook was used to estimate two models. The results of the Bayesian econometric analysis were compared to the results found by Wooldridge. While the point coefficients are marginally different, the trend across the variables is the same. Formal and informal model checks showed that there were flaws in the model that need to be addressed, which may be why the estimates between Wooldridge's and my study differ.

To further this study, the models presented in this analysis would be repeated using a diverse array of priors. Further methods of model checking that would be conducted include using Bayes Factors as a model comparison technique. Bayes Factors allows for preferences between models to be determined. Additionally, model averaging may be used once other models have been calculated. This allows for many models to be used in prediction.

In conclusion, Bayesian econometrics can be beneficial for research. The requirement of stating a prior provides the technician a tool to ensure that her beliefs are coherent. Once the econometrician forms a prior and conditions the data on a likelihood, a posterior distribution is estimated. This distribution can be advantageous to having frequentist point estimates because it allows for a more robust interpretation of probability and it is relatively straightforward to conduct inference from that point.

The posterior distribution, when not of the same form as the prior distribution, can be easily calculated using numerical methods such as the Gibbs sampler rather than impossible integrals. Bayesian methods have many benefits, which have been shown in this paper, and are a valuable addition to econometric analysis and in everyday life.

REFERENCES

Wooldridge, Jeffrey M.. "Limited Dependent Variable Models and Sample Selection Corrections." Introductory Econometrics: a Modern Approach. 5th ed. Mason, Ohio: South-Western Cengage Learning, 2012. Print.

Mroz, Thomas. "The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions." Econometrica 55 (): 765-799. Print.

Lancaster, Tony. An Introduction to Modern Bayesian Econometrics. Malden: Blackwell Publishing, 2004. Print.

Koop, Gary. Bayesian Econometrics. Chichester: Wiley, 2003. Print.