Selection Problems for Application of Probit, Tobit, Logit & Maximum Likelihood Estimation: A Methodological Issue

BY

[1,]Dr Debasis Patnaik , [2,]Mr Nikunj Sunil Sharma

[1],(Asst Professor, Department Of Economics, BITS Pilani, K K Birla Goa Campus, Goa, India)
[2],(B Tech, Department Of EEE, BITS Pilani, K K Birla Goa Campus, Goa, India)

ABSTRACT:

The application of probabilistic models to economics and finance study poses a problem in the sense of which model is more appropriate. A brief discussion using case studies by authors is undertaken to assess a realistic level of difficulty in the discipline. Then we take recourse to data on women's wages and distribution thereof to assess equity in the system is assess the appropriability of use of a probabilistic model. Assessment of student scorecard is also done to show the relative degree of successful prediction achieved. Stata and SPSS softwares were used for filling in data, testing hypothesis and deriving results to nullify software specificity in result efficiency. Finally a model is built to decide whether an individual decides to pay the requisite amount of taxes or not.

I. INTRODUCTION

The tobit and probit models are similar in many ways. Each have the same structural model, just different measurement models i.e. how the *y*¤ is translated into the observed *y* is different. In the tobit model, we know the value of *y*¤ when *y*¤ > 0, while in the probit model we only know if *y*¤ > 0. Since there is more information in the tobit model, the estimates of the ¯s should be more efficient. The logistic has slightly flatter tails i.e., the normal or probit curve approaches the axes more quickly than the logistic curve.Qualitatively, Logit and Probit Models give similar results but the estimates of parameters of the two models are not directly comparable. The likelihood function is the joint probability (density) function of observable random variables but it is viewed as the function of the parameters given the realized random variables.

A brief survey of literature on related study of probability models and its applications reveal the following:Carmen Cote and Joseph Farhan (2002) in the paper 'Application of the Two-Stage Bivariate Probit-Tobit Model to Corporate Financing Decisions' used a simulated model aiming to study the factors affecting firms' choice of the form of financing and the size of issue using a two stage *Bivariate Probit – Tobit model. The first stage examines the factors affecting the firms' choice of the form of financing using a Bivariate-Probit model. They used use a two-stage *Bivariate Probit-Tobit model to examine the corporate financing decisions. In this model, managers make three sequential financing decisions that are not necessarily independent. They are: whether to use internal or external source of funding; if external source of funding is the choice, whether to issue debt or equity; make the decision about the size of the debt (equity) issue. The simulation is based on random draws corresponding to 100 years of data for 1,000 firms. The results show that even all firms follow the pecking order behavior, only 85% of the internal and external issuance decisions and less than 70% of the debt and equity issuance decisions are accurately identified. The results show that the correlation coefficients between the Bivariate-Probit equations and those between the Bivariate-Probit and issue size equations (Tobit) are different from zero. This implies that using the Bivariate-Probit model is more appropriate than two independent Probit when studying corporate financing choices. An examination of factors that affect the choice of financing form and the size of issue support the predictions of both trade-off and pecking order theory. Trade-off factors have a significant impact on the debt-equity choice as well as on the size of issue. Firm size and Z-score have a negative impact on the likelihood of using external funding.

1- Carmen Cote And Joseph Farhat , Application of the Two-Stage Bivariate Probit-Tobit Model to Corporate Financing Decisions' in Baker, M., and Wurgler J.(2002), Market Timing and Capital Structure, Journal of Finance 57, 1-32 Henry W. Chappell Jr.(1982) in 'Campaign Contributions and Congressional Voting: A Simultaneous Probit-Tobit Model' aimed to analyze the financial relationships between Interest groups and policymakers empirically. In the analysis of voting on a particular issue, the principal economic agents of concern are congressmen and a single interest group. Congressmen's voting decisions are presumed to be motivated by a desire to be reelected, while the interest group is assumed to allocate campaign funds to various candidates in an attempt to influence the legislative outcome of the issue. A "simultaneous probit-Tobit model" (hence referred to as model SPT) has been hypothesized to explain voting decisions made by congressmen and contribution decisions made by the interest groups. The probit equation is hypothesized to explain votes on the issue. According to the model, a "yes" vote occurs when the unobserved latent variable exceeds a threshold level of zero, and a "no" vote occurs otherwise. This unobserved variable can be interpreted as the candidate's "propensity to vote in favor of the interest group." Interest group contributions are explained by the Tobit equation.

The preceding theoretical discussion provides a basis for the empirical analysis of interest group campaign contributions and roll call voting by members of the U.S. House of Representatives in the 1974-1977 period. Several criteria were used to guide the selection of the seven issues analyzed in the study. First, an effort was made to avoid issues of concern to numerous diverse competing interest groups. Ideally, just one group should be associated with each issue. Issues in regulatory policy often conform to this criterion. It also attempted to select issues for which close votes were recorded in the House, since congressmen may behave differently in their decision- making when voting on issues of certain versus those of doubtful outcomes. Issues for which a congressman must seriously consider the possibility that his vote could influence the ultimate outcome of legislation are preferred. Finally, it was also necessary to choose issues for which an associated interest group made substantial contributions. The seven issues chosen for study include mortgage disclosure requirements for lenders, milk price supports, truck weight limits, tax rebates for oil companies, funding for the B1 bomber, auto emissions controls, and a maritime cargo preference bill. FIML estimates of the simultaneous probit-Tobit model suggest that he effects of campaign Contributions on voting are smaller than single equation probit estimates would indicate. We are generally unable to conclude that contributions have a significant impact on voting decisions; apparently votes are most often decided on the basis of personal ideology or the preferences of constituents. Despite the lack of significance according to model SPT, it would not, however, be appropriate to unambiguously conclude that contributions have no effects on voting. The FIML estimates of the contribution coefficients are not very precise. It is probable that rather poor overall explanatory power in the equations explaining contributions leads to imprecision of these estimates in the voting equation. If better models to explain contributions are developed in the future, this might result in greater precision in estimating the effects of contributions on voting.

2- Henry W. Chappell Jr.(1981) in 'Campaign Contributions and Congressional Voting: A Simultaneous Probit-Tobit Model'*Review of Economics and Statistics*, Volume 64, Issue 1, 1982, pages 77-83. http://www.mitpressjournals.org/loi/rest. *Received for publication December 29, 1980. Revision accepted for publication May 27, 1981. * Universityo f South Carolina.
This paper is based on my Ph.D. dissertationi n economics.

3- Lee C. Adkins in 'An Instrumental Variables Probit Estimator using gretl' aimed at Application of Probit Estimation using gretl (Gnu Regression, Econometrics and Time-series Library) script.And to estimate endogenous probit models Stata 10 was used. Two estimators of this model: a simple `two-step' estimator and a maximum likelihood estimator. Adkins (2008a) compares these estimators to several others in a Monte Carlo study and finds that the two-step estimator performs reasonably well in some circumstances. Gretl script is used to estimate the parameters of an dichotomous choice model that contains endogenous regressors. The routine is simple and yields the same results as the two-step option in the commercially available Stata 10 software. The next step is to duplicate the maximum likelihood estimator, a considerably more challenging undertaking given the multitude of ways the mle can be computed. It should be noted that the only other commercial software that estimates this model via mle is Limdep; [1] Limdep and [2]Stata use different algorithms and yield different results.

Jay Stewart (2009) in ‘Tobit or No Tobit?’ aim to decide whether to use a Tobit-biased model or not.

*[1]Limdep & [2]Stata are statistical softwares for the estimation of linear and nonlinear regression models and qualitative dependent variable models for cross-section, time-series and panel data. * The GNU General Public License (GNU GPL or simply GPL) is the most widely used free software license. Adkins, Lee C. (2008a), Small sample performance of instrumental variables probit estimators: A monte carlo investigation.Adkins, Lee C. (2008b), `Small sample performance of instrumental variables probit estimators: A monte carlo investigation', Department of Economics, Oklahoma State University, Stillwater OK 74078. available at http://www.learneconometrics.com/pdf/JSM2008.pdf. --Cragg (1971) proposes a double-hurdle model, where the first hurdle is the decision to ever spend money on the good. Since I am restricting my attention to situations where this decision is taken as given, the double-hurdle model reduces to a two-part model. In the first part of the two-part model, a probit is estimated over all observations to determine the probability that individuals purchase the good during the reference period. In the second part, an OLS regression is estimated over the non-zero-value observations. The estimated average probability from the probit is combined with the coefficients from the OLS regression to arrive at unconditional marginal effects.

3-It is published as an IZA Discussion Paper No. 4588 November 2009

Greene Willams[4](2004) in the **‘**The behavior of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects’.The general results for the probit and logit models appear to be mimicked by the ordered probit model. Heckman’s widely cited result for the probit model appears to be incorrect, however. The differences observed here do not appear to be a function of the mechanism used to generate the exogenous variables. The marginal effects in these binary choice models are overestimated by a factor closer to 50% .A result which has not been considered previously is the incidental parameters effect on estimates of the standard errors of the MLEs. We find that while the coefficients are uniformly overestimated, the asymptotic variances are generally underestimated. This result seems to be general, carrying across a variety of models, independently of whether the biases in the coefficient estimators are towards or away from zero. The ML estimator shows essentially no bias in the coefficient estimators of the tobit model. But the small sample bias appears to show up in the estimate of the disturbance variance. The truncated regression and Weibull [1] models are contradictory, and strongly suggest that the direction of bias in the fixedeffects model is model specific.

5- H. E. RAUCH, F. TUNG AND C. T. STRIEBEL(1965) in ‘Maximum Likelihood Estimates of Linear Dynamic Systems‘considers the problem of estimating the states of linear dynamic systems in the presence of additive Gaussian noise.

4- Econometrics Journal (2004), volume 7, pp. 98–119. In probability theory and statistics, the **Weibull distribution** is a continuous probability distribution. It is named after Waloddi Weibull, who described it in detail in 1951, although it was first identified by Fréchet (1927) and first applied by Rosin & Rammler (1933) to describe the size distribution of particles.

5-Publication Info: AIAA JOURNAL VOL. 3, NO.-8, AUGUST 1965

Difference equations relating the estimates for the problems of filtering and smoothing are derived as well as a similar set of equations relates the covariance of the errors. The derivation is based on the method of maximum likelihood and depends primarily on the simple manipulation of the probability density functions. The solutions are mechanized on a digital computer. The MLE of the states with continuous observations can be obtained formally from the MLE of the discrete system. The method used here depends primarily on the simple manipulation of the probability density functions and hence leads immediately to recursion equations. The results are also different. The derivation leads directly to a smoothing solution that uses processed data instead of the original measurements.The solution to the discrete version of the filtering and smoothing problem has been derived using the principal of maximum likelihood and simple manipulation of the probabilitydensity function. The filtered estimate is calculated forward point by point as a linear combination of the previous filtered estimate and the current observation. The smoothing solution starts with the filtered estimate at the last point and calculates backward point by point determining the smoothed estimate as a linear combination of the filtered estimate at that point and the smoothed estimate at the previous point. A numerical example has been presented to illustrate the advantage of smoothing in reducing the error in the estimate.

[1] Wiener, N., The Extrapolation, Interpolation and Smoothing of Stationary Time Series (John Wiley & Sons, Inc., New York,
[2] 1949).
[3] Parzen, E., "An approach to time series analysis," Ann. Math. Statist. **32,** 951-989 (1961).

METHODOLOGY:

This paper uses Adkins (2008b,a) method to produce a simple routine using the free gretl software. The gretl results are compared to those produced by Stata 10 using data on bank holding companies. The gretl and Stata 10 results are virtually identical. The method of **instrumental variables** (**IV**) is used to estimate causal relationships when controlled experiments are not feasible. Instrumental variable methods allow consistent estimation when the explanatory variables (covariates) are correlated with the error terms of a regression relationship. Time-use surveys collect very detailed information about individuals' activities over a short period of time, typically one day. As a result, a large fraction of observations have values of zero for the time spent in many activities, even for individuals who do the activity on a regular basis. For example, it is safe to assume that all parents do at least some childcare, but a relatively large fraction report no time spent in childcare on their diary day. Tobit seems to be the natural approach. However, it is important to recognize that the zeros in time-use data arise from a mismatch between the reference period of the data (the diary day) and the period of interest, which is typically much longer. Then Tobit doesn't seem appropriate. The bias associated with alternative estimation procedures for estimating the marginal effects of covariates on time use is thus noticed. The bias is often large, and that the extent of the bias increases as the fraction of zero observations increases. It seems likely that one of the main reasons for this poor performance is that the Tobit model assumes that the process that determines whether an individual engages in an activity is the same one that governs how much time is spent in that activity. It adapts the infrequency of purchase model to time-diary data and showing that OLS estimates are unbiased. Next, using simulated data, the bias associated with three procedures that are commonly used to analyze time-diary data – Tobit, the Cragg (1971) two-part model, and OLS under a number of alternative assumptions about the data-generating process. The estimated marginal effects from Tobits are found to be biased and that the extent of the bias varies with the fraction of zero-value observations. The two-part model performs significantly better, but generates biased estimated in certain circumstances. Only OLS generates unbiased estimates in all of the simulations considered here.

METHODOLOGY:

a. **Log likelihood -** This is the log likelihood of the fitted model. It is used in the Likelihood Ratio Chi-Square test of whether all predictors' regression coefficients in the model are simultaneously zero.

b. **Number of obs -** This is the number of observations in the dataset for which all of the response and predictor variables are non-missing.

c. **LR chi2(3) -** This is the Likelihood Ratio (LR) Chi-Square test that at least one of the predictors' regression coefficient is not equal to zero. The number in the parentheses indicates the degrees of freedom of the Chi-Square distribution used to test the LR Chi-Square statistic and is defined by the number of predictors in the model (3).

d. **Prob > chi2 -** This is the probability of getting a LR test statistic as extreme as, or more so, than the observed statistic under the null hypothesis; the null hypothesis is that all of the regression coefficients are simultaneously equal to zero. In other words, this is the probability of obtaining this chi-square statistic (22.09) or one more extreme if there is in fact no effect of the predictor variables. This p-value is compared to a specified alpha level, our willingness to accept a type I error, which is typically set at 0.05 or 0.01. The small p-value from the LR test, 0.0001, would lead us to conclude that at least one of the regression coefficients in the model is not equal to zero. The parameter of the chi-square distribution used to test the null hypothesis is defined by the degrees of freedom in the prior line, **chi2(3).**

FIRST MODEL
Effect of Education on Women's Wages

we want to estimate the effect of education on women's wages. The OLS regression for this would be

$yi = xi^{-} + {}^{2}i$ ----(1)

where *yi* is the woman's wage and *xi* is her education. The basic selection problem arises in that the sample consists only of women who choose to work and these women may differ in important *unmeasured* ways from women who do not work. For example, women who are smarter may be more likely to enter the labor market. The 'selection equation' for entering the labor market might be:

$Ui = wi° + ui$ ----(2)

where *Ui* represents the utility to woman *i* of entering the labor market and *wi* is a vector of factors known to influence a woman's decision to work such as her education level. *ui* is assumed to be jointly normally distributed with ${}^{2}i$ and contains any unmeasured characteristics in the selection equation. We don't actually observe *Ui*. All we observe is a dichotomous variable *Zi* with a value of 1 if the woman enters the labor force (*Ui* > 0) and 0 otherwise. So, where does the selection problem actually come from? Well, there are two selection effects.

1. Women with higher levels of education will be more likely to enter the labor force and so we will have a sample of educated women. As Sartori (2003, 114) notes, this non-random aspect of the sample is what is commonly *misunderstood* to be the problem of 'selection bias'. But this on its own does not bias the estimation of the outcome equation in (1).

2. The second selection effect, which is the most important, *is that some uneducated women will go to work*. This is because these women decide that work is worthwhile because they have a high value on some unmeasured variable which is captured in *ui* . In other words, these women get into our sample not because they have high education (they have low values of *wi*°), but because they have large error terms. In contrast, those women who get into our sample because they have high education (large values of *wi*°) will have a more normal range of errors. The problem is that whether or not education (or independent variables of interest in the outcome equation) is correlated with the unmeasured intelligence (our unmeasured variable) in the overall population, these two variables will be correlated in the selected sample. If intelligence does lead to higher wages, then we will underestimate the effect of education on wages because in the selected sample women with little education are unusually smart.

Many dependent variables of interest take only two values (a dichotomous variable), denoting an event or non-event and coded as 1 and 0 respectively. Some

The Logit Model

• When the transformation function *F* is the logistic function, the response probabilities are given by

$$P(y_i = 1 \mid x_i) = \frac{e^{x_i' \beta}}{1 + e^{x_i' \beta}}$$

• And, when the transformation *F* is the cumulative density function (cdf) of the standard normal distribution, the response probabilities are given by

$$P(y_i = 1 \mid x_i) = \Phi(x_i' \beta) = \int_{-\infty}^{x_i'\beta} \Phi(s)\, ds \;\; = \int_{-\infty}^{x_i'\beta} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} \;\; ds$$

• The Logit and Probit models are almost identical and the choice of the model is arbitrary, although logit model has certain advantages (simplicity and ease of interpretation)

However, the parameters of the two models are scaled differently. The parameter estimates in a logistic regression tend to be 1.6 to 1.8 times higher than they are in a corresponding probit model.

The probit and logit models are estimated by maximum likelihood (ML), assuming independence across observations. The ML estimator of β is consistent and asymptotically normally distributed. However, the estimation rests on the strong assumption that the latent error term is normally distributed and homoscedastic. If homoscedasticity is violated, no easy solution is found.

In the probit model, use the *Z*-score terminology. For every unit increase in X, the *Z*-score (or the Probit of “success”) increases by *b* units. [Or, we can also say that an increase in X changes *Z* by *b* standard deviation units.]

One can convert the z-score to probabilities using the normal table.
• The Tobit model uses all of the information, including info on censoring and provides consistent estimates.
• It is also a nonlinear model and similar to the probit model. It is estimated using maximum likelihood estimation techniques. The likelihood function for the tobit model takes the form:

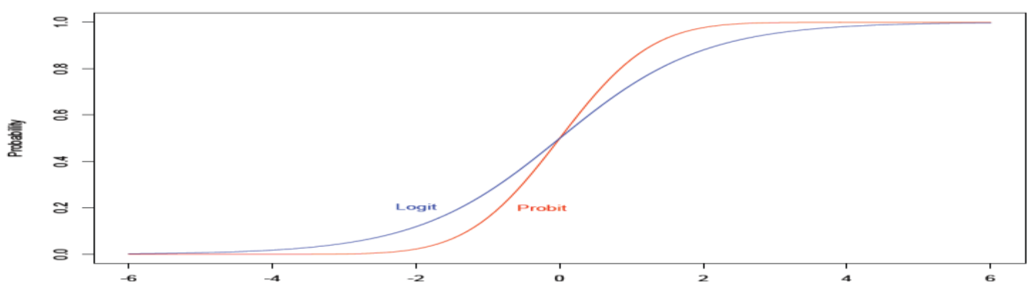
$$\log L = \sum_{Y_i > 0} -\frac{1}{2}\left[\log(2\pi) + \log\sigma^2 + \frac{(Y_i - \beta X_i)^2}{\sigma^2}\right] + \sum_{Y_i = 0} \log\left[1 - F\left(\frac{\beta X_i}{\sigma}\right)\right]$$

• This is an unusual function, it consists of two terms, the first for non-censored observations (it is the pdf), and the second for censored observations (it is the cdf).

• The estimated tobit coefficients are the marginal effects of a change in *xj* on y*, the unobservable latent variable and can be interpreted in the same way as in a
linear regression model.But such an interpretation may not be useful since we are interested in the effect of *X* on the observable *y* (or change in the censored outcome).
J. Scott Long, 1997 (translated 2002), *Regression Models for Categorical and Limited Dependent Variables*. It can be shown that change in *y* is found by multiplying the coefficient with $Pr(a<y^*<b)$, that is, the probability of being uncensored. Since this probability is a fraction, the marginal effect is actually attenuated. In the above, *a* and *b* denote lower and upper censoring points. For example, in left censoring, the limits will be: *a* =0, *b*=+∞.

SECOND MODEL

Logit versus probit



Effect of GRE Scores on Grades in Graduate School

Suppose that an admissions committee want to know how GRE scores affect the likelihood of success in
graduate school. The problem is that information about success in graduate school (grades) is only available for those students who were admitted. The admissions committee wish to forecast outcomes in the whole pool of applicants but are forced to rely solely on experience with a non-random subset of them. Let’s assume that we have the following model. The selection equation for getting admitted might be

Admission Rating = °0 + °1GRE + *ui* ----(3)
Admission = 1 if Admission Rating > 0
0 if Admission Rating < 0

where ADMISSION RATING is the latent variable measuring the underlying propensity to be admitted, GRE represents a student’s GRE score, and Admission is a dichotomous variable indicating whether the student was admitted or not. The outcome equation is

Success = ½ ¯0 + ¯1GRE + ²*i* if Admission=1

Unobserved if Admission = 0

Admitted graduate students are not representative of applicants generally as the admission equation makes clear. There are many college graduates with low grades who attempt to enroll in graduate school; only a few succeed. These exceptions usually owe their success to favorable (unmeasured) personal characteristics other than grades. While many of these personal characteristics will have no affect on their success in graduate school, it seems reasonable to think that some of them probably will. As a result, there will be some students with low grades who make into graduate school because they have large error terms (they have strong personal characteristics). As a result, this low-grade subset of students will perform above the level of other applicants with the same college grades and so they are no longer representative. Now suppose the admissions committee examine graduate student grades to compare the performance of those who entered with low GREs to those who entered with high GREs. The group of students who were admitted because they had strong GREs will be representative of the group of applicants with strong GREs. However, the subset of admitted students with low GREs will not be representative of the group of applicants with low GREs - they will perform better in graduate school (because of their large disturbance terms due to personal characteristics) than applicants with low GREs that were not admitted. Ultimately, it may appear that students with high GREs do not outperform students with low GREs in graduate school. The admissions committee might be tempted to conclude that GREs do not predict success. However, intuition makes it clear that this result does not extend to the applicant pool where students with low GREs would, in general, perform quite poorly had they been admitted. In effect, if a random sample of applicants were admitted to graduate school, GREs would be a good predictor of their success.

THIRD MODEL

Questionnaire :

- On what data is the model being applied.
- Finding the factors affecting the dataset.
- What model to use.
- Calculating the model Coefficients.
- Estimations using either software or hand calculations.
- What has been concluded.

To decide whether an individual decides to pay the requisite amount of taxes or not. And , thus also decide the model to be used for the same.

Below is the dataset for individuals in 17 Latin American countries.

Table 1. Reasons why individuals evade taxes

Why do people not pay their taxes?	Arg	Bol	Braz	Col	Cos	Chi	Ecu	El	Gua	Hon	Mex	Nic	Pan	Par	Per	Uru	Ven	Average
Lack of honesty	17.7	47.0	45.5	31.3	54.0	54.7	53.8	58.5	49.6	53.5	39.2	36.0	49.8	47.3	41.6	20.3	57.5	44.5
Because nationals are quick-witted and sly	14.8	17.6	31.8	17.8	29.2	44.4	47.2	25.8	12.8	28.3	25.4	16.2	32.7	8.9	25.6	30.5	39.9	26.4
They don't see the point in paying taxes	19.7	28.5	25.9	24.4	21.2	30.3	37.8	44.8	15.3	41.3	49.9	30.4	26.9	29.9	21.4	23.4	29.2	29.4
Lack of civic conscience	15.3	35.3	32.0	28.9	24.9	39.5	49.3	40.4	20.2	49.3	38.3	33.4	41.1	37.4	34.2	24.2	40.7	34.4
Because those that evade taxes go unpunished	26.0	23.0	24.3	16.6	19.2	18.1	31.3	36.4	13.4	24.6	36.5	21.0	18.6	19.9	14.6	26.5	22.3	23.1
Because the taxes are ill-spent	26.7	40.4	29.7	40.4	27.8	22.6	45.8	46.4	20.1	35.1	50.3	33.7	27.5	29.9	23.2	25.1	26.6	32.4
Because the taxes are too high	65.6	37.1	50.0	62.8	37.6	32.0	50.8	54.3	24.1	47.2	55.8	57.6	38.5	42.9	50.2	63.7	25.3	46.8
Because there is corruption	32.0	42.4	48.9	48.7	43.7	32.5	59.0	52.5	43.2	44.4	54.6	41.9	40.5	47.0	32.8	41.0	45.7	44.2

Note. Percentage of Individuals that mentioned reasons why people do not pay their taxes.

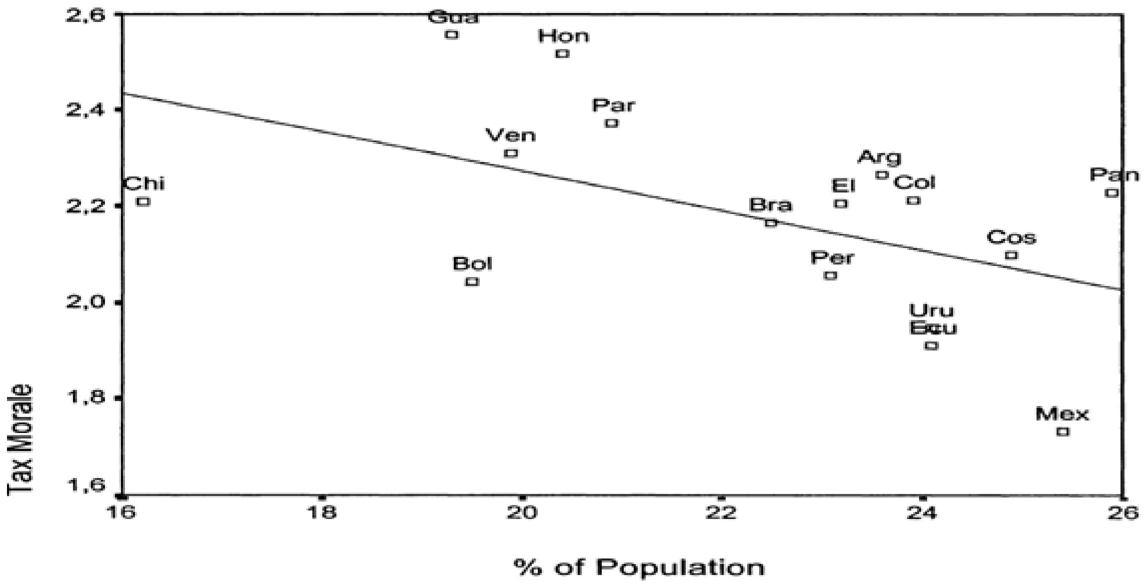
Using the the two data sets Latinobarometro (data from 1998) and World Values Survey , We apply the standard Probit Model for which $Yi^* = 1$ (for an individual paying his taxes) and $Yi^* = 0$ otherwise for the following set of equations :

Yi^* is unobservable but $Yi = 0$ if $Yi^* < 0$

1 if $Yi^* >= 0$

Wherein, $P(Yi=1) = P(Yi^*>=0) = P(ui >= -B1 - \ldots\ldots -Bk^*xki)$

= F(B1 + B2*x2i +….+Bk*xki)
Here, F is the cumulative distributive Function of ui. We are assuming that the probability density function to be symmetric.
All the factors accounted for are listed in the above table.
Now , using Finney's table ,
Thus transforming percentages to probits**.Through Hand Calculations or using Computer Software such as SPSS, SAS , R, or S we can convert the percent responded to probits automatically.
The following figure has been obtained by computing the values in the form of a graph. Its been done using hand calculations.



'igure 1. Correlation between tax morale and the size of shadow economy

lotes. Arg = Argentina, Bol = Bolivia, Bra = Brazil, Col = Columbia, Cos = Costa ica, Chi = Chile, Ecu = Ecuador, El = El Salvador, Gua = Guatemala, Hon = Honduras, Mex Mexico, Nic = Nicaragua, Pan = Panama, Par = Paraguay, Per = Peru, Uru = Uruguay, Ven Venezuela.

** The Conversion of percentages (w.r.t the various factors affecting the tax morale in various latin American countries) to probits is carried out with the help Finney's Table.
* Finney's Table is shown in Appendix [A] as follows.

Finney's Table :

%	0	1	2	3	4	5	6	7	8	9
0	—	2.67	2.95	3.12	3.25	3.36	3.45	3.52	3.59	3.66
10	3.72	3.77	3.82	3.87	3.92	3.96	4.01	4.05	4.08	4.12
20	4.16	4.19	4.23	4.26	4.29	4.33	4.36	4.39	4.42	4.45
30	4.48	4.50	4.53	4.56	4.59	4.61	4.64	4.67	4.69	4.72
40	4.75	4.77	4.80	4.82	4.85	4.87	4.90	4.92	4.95	4.97
50	5.00	5.03	5.05	5.08	5.10	5.13	5.15	5.18	5.20	5.23
60	5.25	5.28	5.31	5.33	5.36	5.39	5.41	5.44	5.47	5.50
70	5.52	5.55	5.58	5.61	5.64	5.67	5.71	5.74	5.77	5.81
80	5.84	5.88	5.92	5.95	5.99	6.04	6.08	6.13	6.18	6.23
90	6.28	6.34	6.41	6.48	6.55	6.64	6.75	6.88	7.05	7.33
—	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
99	7.33	7.37	7.41	7.46	7.51	7.58	7.65	7.75	7.88	8.09

Finney (1948) has given a table which may be used to test the significance of the deviation from proportionality. As in this case, its been used for converting thefactors affecting tax morale percentages into probits.

The Probit Model

Index function

$$y_i^* = \beta_1 + \beta_2 x_{2i} + \ldots\ldots + \beta_k x_{ki} + u_i$$

$$y_i^* \text{ is unobservable but } y_i = \begin{cases} 0 \text{ if } y_i^* < 0 \\ 1 \text{ if } y_i^* \geq 0 \end{cases}$$

$$P(y_i = 1) = P(y_i^* \geq 0) = P(u_i \geq -\beta_1 - \beta_2 x_{2i} - \ldots\ldots - \beta_k x_{ki})$$

$$= F(\beta_1 + \beta_2 x_{2i} + \ldots\ldots + \beta_k x_{ki}) \text{ where } F \text{ is the cumulative distribution function}$$

of u_i. We are assuming that the probability density function of u_i is symmetric.

The Logit Model

Its very similar to the probit model.

$$y_i^* = \beta_1 + \beta_2 x_{2i} + \ldots\ldots + \beta_k x_{ki} + u_i$$

$$y_i^* \text{ is unobservable but } y_i = \begin{cases} 0 \text{ if } y_i^* < 0 \\ 1 \text{ if } y_i^* \geq 0 \end{cases}$$

$$P(y_i = 1) = P(y_i^* \geq 0) = P(u_i \geq -\beta_1 - \beta_2 x_{2i} - \ldots\ldots - \beta_k x_{ki}$$

$$= F(\beta_1 + \beta_2 x_{2i} + \ldots\ldots + \beta_k x_{ki}) \text{ where } F \text{ is the cumulative distribution function}$$

of u_i. We are assuming that the probability density function of u_i is symmetric.

In the probit model we assumed that $u_i \sim N(0, \sigma_u^2)$. In the logit model we assumed that has what is known as a logistic distribution. The pdf of is given by $f(u_i) = \dfrac{e^{u_i}}{(1+e^{u_i})^2}$

The model is estimated by MLE.

The Censored Regression (Tobit) Model

The Tobit Model arises when the y variable is limited (or censored) from above or below.

$$y_i^* = \beta_1 + \beta_2 x_{2i} + \ldots\ldots + \beta_k x_{ki} + u_i$$

$$y_i^* \text{ is unobservable but } y_i = \begin{cases} 0 \text{ if } y_i^* < 0 \\ y_i^* \text{ if } y_i^* \geq 0 \end{cases}$$

```
Binary Probit Regression (in SPSS, use the ordinal regression menu and select probi
link function. Ignore the test of parallel lines, etc.)
```

Model Fitting Information

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	1645.024			
Final	1166.702	478.322	4	.000

Link function: Probit.

Parameter Estimates

		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Threshold	[work = 0]	2.037	.209	94.664	1	.000	1.626	2.447
Location	age	.035	.004	67.301	1	.000	.026	.043
	education	.058	.011	28.061	1	.000	.037	.080
	children	.447	.029	243.907	1	.000	.391	.503
	[married=0]	-.431	.074	33.618	1	.000	-.577	-.285
	[married=1]	0[a]	.	.	0	.	.	.

Link function: Probit.

a. This parameter is set to zero because it is redundant.

```
Tobit regression cannot be done in SPSS. Use Stata. Here are the Stata commands.
First, fit simple OLS Regression of the variable lwf (just to check)

. regress lwf age married children education

      Source |       SS       df       MS              Number of obs =    2000
-------------+------------------------------           F(  4,  1995) =  134.21
       Model |  937.873188     4  234.468297           Prob > F      =  0.0000
    Residual |  3485.34135  1995  1.74703827           R-squared     =  0.2120
-------------+------------------------------           Adj R-squared =  0.2105
       Total |  4423.21454  1999  2.21271363           Root MSE      =  1.3218

------------------------------------------------------------------------------
         lwf |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0363624    .003862     9.42   0.000      .0287885    .0439362
     married |   .3188214   .0690834     4.62   0.000      .1833381    .4543046
    children |   .3305009   .0213143    15.51   0.000      .2887004    .3723015
   education |   .0843345   .0102295     8.24   0.000      .0642729    .1043961
       _cons |  -1.077738   .1703218    -6.33   0.000     -1.411765   -.7437105
------------------------------------------------------------------------------

. tobit lwf age married children education, ll(0)
```

```
Tobit regression                                  Number of obs   =       2000
                                                  LR chi2(4)      =     461.85
                                                  Prob > chi2     =     0.0000
Log likelihood = -3349.9685                       Pseudo R2       =     0.0645

------------------------------------------------------------------------------
         lwf |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |    .052157   .0057457     9.08   0.000     .0408888    .0634252
     married |   .4841801   .1035188     4.68   0.000     .2811639    .6871964
    children |   .4860021   .0317054    15.33   0.000     .4238229    .5481812
   education |   .1149492   .0150913     7.62   0.000     .0853529    .1445454
       _cons |  -2.807696   .2632565   -10.67   0.000    -3.323982   -2.291409
-------------+----------------------------------------------------------------
      /sigma |   1.872811    .040014                      1.794337    1.951285
------------------------------------------------------------------------------
  Obs. summary:        657  left-censored observations at lwf<=0
                      1343     uncensored observations
                         0 right-censored observations


. mfx compute, predict(pr(0,.))

Marginal effects after tobit
      y  = Pr(lwf>0) (predict, pr(0,.))
         =  .81920975
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
     age |   .0073278      .00083    8.84   0.000   .005703  .008952   36.208
married*|   .0706994      .01576    4.48   0.000   .039803  .101596    .6705
children |   .0682813      .00479   14.26   0.000   .058899  .077663   1.6445
educat~n |   .0161499      .00216    7.48   0.000   .011918  .020382   13.084
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1

. mfx compute, predict(e(0,.))

Marginal effects after tobit
      y  = E(lwf|lwf>0) (predict, e(0,.))
         =  2.3102021
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
     age |   .0314922      .00347    9.08   0.000   .024695   .03829   36.208
married*|   .2861047      .05982    4.78   0.000   .168855  .403354    .6705
children |   .2934463      .01908   15.38   0.000   .256041  .330852   1.6445
educat~n |   .0694059      .00912    7.61   0.000   .051531  .087281   13.084
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

A researcher is interested in how variables, such as GRE (Graduate Record Exam scores), GPA (grade point average) and prestige of the undergraduate institution, effect admission into graduate school. The response variable, admit/don't admit, is a binary variable.

We have generated hypothetical data, which can be obtained from the URL :
http://www.ats.ucla.edu/stat/stata/dae/probit.htm
This data set has a binary response (outcome, dependent) variable called **admit**. There are three predictor variables: **gre**, **gpa** and **rank**. We will treat the variables **gre** and **gpa** as continuous. The variable **rank** is ordinal, it takes on the values 1 through 4. Institutions with a rank of 1 have the highest prestige, while those with a rank of 4 have the lowest. We will treat **rank** as categorical.

summarize gre gpa

Variable	Obs	Mean	Std. Dev.	Min	Max
gre	400	587.7	115.5165	220	800
gpa	400	3.3899	.3805668	2.26	4

tab rank

rank	Freq.	Percent	Cum.
1	61	15.25	15.25
2	151	37.75	53.00
3	121	30.25	83.25
4	67	16.75	100.00
Total	400	100.00	

tab admit

admit	Freq.	Percent	Cum.
0	273	68.25	68.25
1	127	31.75	100.00
Total	400	100.00	

To run the model in Stata, we first give the response variable (**admit**), followed by our predictors (**gre, topnotch** and **gpa**).

tab admit rank

	rank				
admit	1	2	3	4	Total
0	28	97	93	55	273
1	33	54	28	12	127
Total	61	151	121	67	400

Analysis methods you might consider

Below is a list of some analysis methods you may have encountered. Some of the methods listed are quite reasonable while others have either fallen out of favor or have limitations.

- Probit regression.
- Logistic regression. A logit model will produce results similar probit regression. The choice of probit versus logit depends largely on individual preferences.
- OLS regression. When used with a binary response variable, this model is known as a linear probability model and can be used as a way to describe conditional probabilities. However, the errors (i.e., residuals) from the linear probability model violate the homoskedasticity and normality of errors assumptions of OLS regression, resulting in invalid standard errors and hypothesis tests.

Probit regression

Below we use the **probit** command to estimate a probit regression model. The **i.** before **rank** indicates that **rank** is a factor variable (i.e., categorical variable), and that it should be included in the model as a series of indicator variables. Note that this syntax was introduced in Stata 11.

probit admit gre gpa i.rank

```
Iteration 0:   log likelihood = -249.98826
Iteration 1:   log likelihood = -229.29667
Iteration 2:   log likelihood = -229.20659
Iteration 3:   log likelihood = -229.20658

Probit regression                        Number of obs   =        400
                                    LR chi2(5)      =      41.56
                                    Prob > chi2     =     0.0000
Log likelihood = -229.20658              Pseudo R2       =     0.0831

------------------------------------------------------------------------------
admit |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
gre |   .0013756   .0006489     2.12   0.034     .0001038    .0026473
gpa |   .4777302   .1954625     2.44   0.015     .0946308    .8608297

|------------------------------------------------------------------------------
rank |
2  |  -.4153992   .1953769    -2.13   0.033    -.7983308   -.0324675
3  |   -.812138   .2085956    -3.89   0.000    -1.220978   -.4032981
4  |   -.935899   .2456339    -3.81   0.000    -1.417333   -.4544654
|
_cons |  -2.386838   .6740879    -3.54   0.000    -3.708026   -1.065649
------------------------------------------------------------------------------
```

- In the output above, we first see the iteration log, indicating how quickly the model converged. The log likelihood (-229.20658) can be used in comparisons of nested models, but we won't show an example of that here.
- Also at the top of the output we see that all 400 observations in our data set were used in the analysis (fewer observations would have been used if any of our variables had missing values).
- The likelihood ratio chi-square of 41.56 with a p-value of 0.0001 tells us that our model as a whole is statistically significant, that is, it fits significantly better than a model with no predictors.
- In the table we see the coefficients, their standard errors, the z-statistic, associated p-values, and the 95% confidence interval of the coefficients. Both **gre**, **gpa**, and the three indicator variables for **rank** are statistically significant. The probit regression coefficients give the change in the z-score or probit index for a one unit change in the predictor.
  - For a one unit increase in **gre**, the z-score increases by 0.001.
  - For each one unit increase in **gpa**, the z-score increases by 0.478.
  - The indicator variables for **rank** have a slightly different interpretation. For example, having attended an undergraduate institution of **rank** of 2, versus an institution with a **rank** of 1 (the reference group), decreases the z-score by 0.415.

A test for an overall effect of **rank** using the **test** command can be done. Below we see that the overall effect of **rank** is statistically significant.
test 2.rank 3.rank 4.rank

```
 ( 1)  [admit]2.rank = 0
 ( 2)  [admit]3.rank = 0
 ( 3)  [admit]4.rank = 0

           chi2(  3) =   21.32
         Prob > chi2 =    0.0001
```

We can also test additional hypotheses about the differences in the coefficients for different levels of rank. Below we test that the coefficient for **rank**=2 is equal to the coefficient for **rank**=3.
test 2.rank = 3.rank

(1) [admit]2.rank - [admit]3.rank = 0

chi2(1) = 5.60
Prob > chi2 = 0.0179

Model Summary

Parameter Estimates

admit[g]	Coef.[h]	Std. Err.[i]	z[j]	P>|z|[k]	[95% Conf. Interval][l]	
gre	.0015244	.0006382	2.39	0.017	.0002736	.0027752
topnotch	.2730334	.1795984	1.52	0.128	-.078973	.6250398
gpa	.4009853	.1931077	2.08	0.038	.0225012	.7794694
_cons	-2.797884	.6475363	-4.32	0.000	-4.067032	-1.528736

e. **admit -** This is the binary response variable predicted by the model.
gre - The coefficient of **gre** is 0.0015244. This means that an increase in GRE score increases the predicted probability of admission.
topnotch - The coefficient of **topnotch** is 0.2730334. This means attending a top notch institution as an undergraduate increases the predicted probability of admission.
gpa - The coefficient of **gpa** is 0.4009853. This means that an increase in GPA increases the predicted probability of admission.
_cons - The constant term is -2.797884. This means that if all of the predictors (**gre**, **topnotch** and **gpa**) are evaluated at zero, the predicted probability of admission is
F(-2.797884) =0.002571929. So, as expected, the predicted probability of a student with a GRE score of zero and a GPA of zero from a non-topnotch school has an extremely low predicted probability of admission.
To generate values from *F* in Stata, use the **normal** function. For example,
display normal(0)
will display .5, indicating that *F*(0) = .5 (i.e., half of the area under the standard normal distribution curve falls to the left of zero). The first student in our dataset has a GRE score of 380, a GPA of 3.61, and a topnotch indicator value of 0. We could multiply these values by their corresponding coefficients,
display -2.797884 +(.0015244*380) + (.2730334*0) + (.4009853*3.61)
to determine that the predicted probability of admittance is *F*(-0.77105507). To find this value, we type
display normal(-0.77105507)
and arrive at a predicted probability of 0.22033715.
f. **Std. Err. -** These are the standard errors of the individual regression coefficients. They are used in both the calculation of the **z** test statistic, superscript j, and the confidence interval of the regression coefficient, superscript l.
g. **z -** The test statistic **z** is the ratio of the **Coef.** to the **Std. Err.** of the respective predictor. The z value follows a standard normal distribution which is used to test against a two-sided alternative hypothesis that the **Coef.** is not equal to zero.
h. **P>|z| -** This is the probability the **z** test statistic (or a more extreme test statistic) would be observed under the null hypothesis that a particular predictor's regression coefficient is zero, given that the rest of the predictors are in the model. For a given alpha level, **P>|z|** determines whether or not the null hypothesis can be rejected. If **P>|z|** is less than alpha, then the null hypothesis can be rejected and the parameter estimate is considered statistically significant at that alpha level.
gre - The **z** test statistic for the predictor **gre** is (0.0015244/0.0006382) = 2.39 with an associated p-value of 0.017. If we set our alpha level to 0.05, we would reject the null hypothesis and conclude that the regression coefficient for **gre** has been found to be statistically different from zero given **topnotch** and **gpa** are in the model.
topnotch - The **z** test statistic for the predictor **topnotch** is (0.2730334/0.1795984) =1.52 with an associated p-value of 0.128. If we set our alpha level to 0.05, we would fail to reject the null hypothesis and conclude that the regression coefficient for **topnotch** has not been found to be statistically different from zero given **gre** and **gpa** are in the model.

gpa - The **z** test statistic for the predictor **gpa** is (0.4009853/0.1931077) = 2.08 with an associated p-value of 0.038. If we set our alpha level to 0.05, we would reject the null hypothesis and conclude that the regression coefficient for **gpa** has been found to be statistically different from zero given **gre** and **topnotch** are in the model.
_cons -The **z** test statistic for the intercept, **_cons,** is (-2.797884/0.6475363) = -4.32 with an associated p-value of < 0.001. With an alpha level of 0.05, we would reject the null hypothesis and conclude that **_cons** has been found to be statistically different from zero given **gre**, **topnotch** and **gpa** are in the model and evaluated at zero.

[95% Conf. Interval] - This is the Confidence Interval (CI) for an individual coefficient given that the other predictors are in the model. For a given predictor with a level of 95% confidence, we'd say that we are 95% confident that the "true" coefficient lies between the lower and upper limit of the interval. It is calculated as the **Coef.** ± ($z_{\alpha/2}$)*(**Std.Err.**), where $z_{\alpha/2}$ is a critical value on the standard normal distribution. The CI is equivalent to the **z** test statistic: if the CI includes zero, we'd fail to reject the null hypothesis that a particular regression coefficient is zero given the other predictors are in the model. An advantage of a CI is that it is illustrative; it provides a range where the "true" parameter may lie.

Illustration (2) :

Example on MLE –

Estimations using R (Statististical Computer Software).

- Goal: Try to find the parameter value $\tilde{\beta}$ that makes $E(Y|X, \beta)$ as close as possible to the observed Y
- For Bernoulli: Let $p_i = P(Y_i = 1|X_i)$ which implies $P(Y_i = 0|X_i) = 1 - P_i$. The probability of observing Y_i is then

$$P(Y_i|X_i) = P_i^{Y_i}(1 - P_i)^{1-Y_i}$$

- Since the observations can be assumed independent events, then

$$P(Y_i|X_i) = \prod_{i=1}^{N} P_i^{Y_i}(1 - P_i)^{1-Y_i}$$

- When evaluated, this expression yields a result on the interval $(0, 1)$ that represents the likelihood of observing this sample Y given X *if* $\hat{\beta}$ *were the "true" value*
- The MLE is denoted as $\tilde{\beta}$ for β that maximizes $L(Y|X, b) = \max L(Y|X, b)$

MLE example: what π for a tossed coin?

Y_i	P^yi (0.5)	(1-P)^(1-yi)	L	ln L
0	1	0.5	0.5	-0.693147
1	0.5	1	0.5	-0.693147
1	0.5	1	0.5	-0.693147
0	1	0.5	0.5	-0.693147
1	0.5	1	0.5	-0.693147
1	0.5	1	0.5	-0.693147
0	1	0.5	0.5	-0.693147
1	0.5	1	0.5	-0.693147
1	0.5	1	0.5	-0.693147
1	0.5	1	0.5	-0.693147
Likelihood			0.0009766	
Log-Likelihood				-6.931472

Y_i	P^yi (0.6)	(1-P)^(1-yi)	L	ln L
0	1	0.4	0.4	-0.916291
1	0.6	1	0.6	-0.510826
1	0.6	1	0.6	-0.510826
0	1	0.4	0.4	-0.916291
1	0.6	1	0.6	-0.510826
1	0.6	1	0.6	-0.510826
0	1	0.4	0.4	-0.916291
1	0.6	1	0.6	-0.510826
1	0.6	1	0.6	-0.510826
1	0.6	1	0.6	-0.510826
Likelihood			0.0017916	
Log-Likelihood				-6.324652

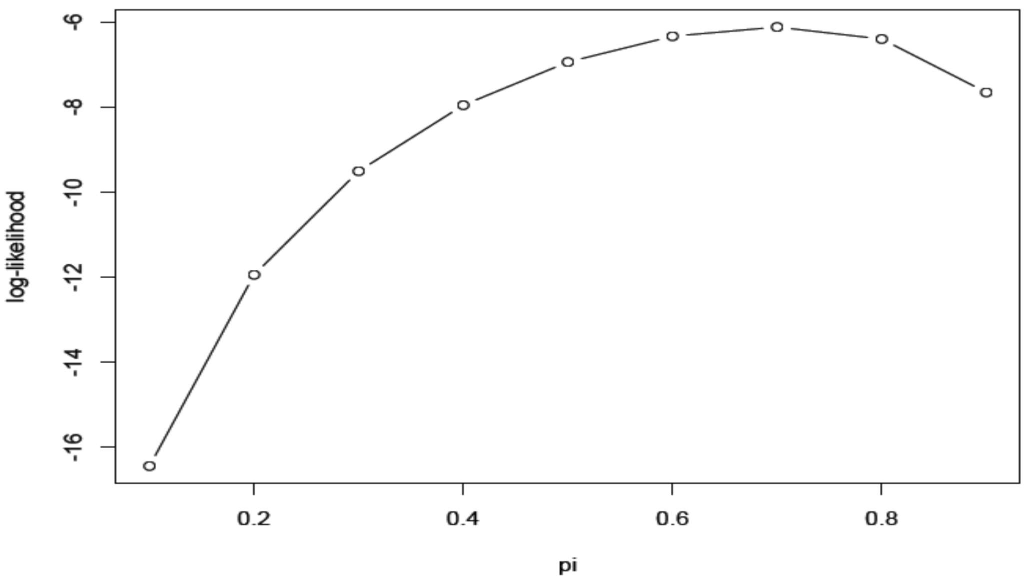
Y_i	P^yi	(1-P)^(1-yi)	L	ln L
	0.7			
0	1	0.3	0.3	-1.203973
1	0.7	1	0.7	-0.356675
1	0.7	1	0.7	-0.356675
0	1	0.3	0.3	-1.203973
1	0.7	1	0.7	-0.356675
1	0.7	1	0.7	-0.356675
0	1	0.3	0.3	-1.203973
1	0.7	1	0.7	-0.356675
1	0.7	1	0.7	-0.356675
1	0.7	1	0.7	-0.356675
Likelihood			0.0022236	
Log-Likelihood				-6.108643

Y_i	P^yi	(1-P)^(1-yi)	L	ln L
	0.8			
0	1	0.2	0.2	-1.609438
1	0.8	1	0.8	-0.223144
1	0.8	1	0.8	-0.223144
0	1	0.2	0.2	-1.609438
1	0.8	1	0.8	-0.223144
1	0.8	1	0.8	-0.223144
0	1	0.2	0.2	-1.609438
1	0.8	1	0.8	-0.223144
1	0.8	1	0.8	-0.223144
1	0.8	1	0.8	-0.223144
Likelihood			0.0016777	
Log-Likelihood				-6.390319

MLE example in R

```
> ## MLE example
> y <- c(0,1,1,0,1,1,0,1,1,1)
> coin.mle <- function(y, pi) {
+   lik <- pi^y * (1-pi)^(1-y)
+   loglik <- log(lik)
+   cat("prod L = ", prod(lik), ",  sum ln(L) = ", sum(loglik), "\n")
+   (mle <- list(L=prod(lik), lnL=sum(loglik)))
+ }
> ll <- numeric(9)
> pi <- seq(.1,.9,.1)
> for (i in 1:9) (ll[i] <- coin.mle(y, pi[i])$lnL)
prod L =  7.29e-08 ,  sum ln(L) =  -16.43418
prod L =  6.5536e-06 ,  sum ln(L) =  -11.93550
prod L =  7.50141e-05 ,  sum ln(L) =  -9.497834
prod L =  0.0003538944 ,  sum ln(L) =  -7.946512
prod L =  0.0009765625 ,  sum ln(L) =  -6.931472
prod L =  0.001791590 ,  sum ln(L) =  -6.324652
prod L =  0.002223566 ,  sum ln(L) =  -6.108643
prod L =  0.001677722 ,  sum ln(L) =  -6.390319
prod L =  0.0004782969 ,  sum ln(L) =  -7.645279
> plot(pi, ll, type="b")
```

MLE example in R: plot



REFERENCES

[1] Almeida, H. and Campello, M., 2008, Financing Frictions and the Substitution between Internal and External funds, *Journal of Financial and Quantitative Analysis* forthcoming

[2] Chen, L. and Zhao S., 2004, The modified pecking order theory: New empirical evidence from corporate financing decisions, *Working Paper*, Michigan State University

[3] Gatchev, V. A., Spindt, P. A., and Tarhan V., 2009, How Do Firms Finance Their Investments? The Relative Importance of Equity Issuance and Debt Contracting Costs, *Journal of Corporate Finance* forthcoming

[4] Gomes, A. and Phillips G., 2004, Why do public firms issue private and public equity, convertibles and debt?, *Working Paper*, University of Pennsylvania

[5] Abrams, Burton A., "Legislative Profitsa nd the Economic Theory of RepresentativeV oting:A n Empirical Investigation," Public Choice 31 (Fall 1977), 111-128.

[6] Alexander, Herbert E., Money in Politics (Washington, D.C.: Public Affairs Press, 1972).

[7] Chappell, Henry W., "The Economics of Campaign Finance: An Empirical Study," unpublishedP h.D. dissertation, Yale University, 1979.

[8] McDonald, John and Robert Moffitt (1980) "The Uses of Tobit Analysis." *The Review of*

[9] *Economics and Statistics* 62(2), May 1980, pp. 318-321.

[10] Meng, C., and Schmidt P., 1985, On the Cost of Partial Observability in the Bivariate Probit Model, *International Economic Review* 26, 71-85

[11] Heckman, James J., "Simultaneous Equation Models with Continuous and Discrete Endogenous Variables and Structural Shifts," in Stephen M. Goldfeld and Richard E. Quandt (eds.), Studies in Nonlinear Estiination (Cambridge, MA: Ballinger Publishing Co., 1976).

URL : http://www.crcnetbase.com/action/showPublications?display=bySubject
http://www.ats.ucla.edu/stat/stata/dae/probit.htm
http://www.stata.com/help.cgi?mfx