A COMPARISON OF BAYESIAN AND SAMPLING THEORY INFERENCES IN A PROBIT MODEL

by William E. Griffiths, R. Carter Hill, and Christopher J. O'Donnell*

THE PROBIT MODEL IS A POPULAR DEVICE for explaining binary choice decisions in econometrics. It has been used to describe choices such as labor force participation, travel mode, home ownership, and type of education. These and many more examples can be found in papers by Amemiya (1981) and Maddala (1983). Given the contribution of economics towards explaining such choices, and given the nature of data that are collected, prior information on the relationship between a choice probability and several explanatory variables frequently exists. Bayesian inference is a convenient vehicle for including such prior information. Given the increasing popularity of Bayesian inference it is useful to ask whether inferences from a probit model are sensitive to a choice between Bayesian and sampling theory techniques. Of interest is the sensitivity of inference on coefficients, probabilities, and elasticities. We consider these issues in a model designed to explain choice between fixed and variable interest rate mortgages. Two Bayesian priors are employed: a uniform prior on the coefficients, designed to be noninformative for the coefficients, and an inequality restricted prior on the signs of the coefficients. We often know, a priori, whether increasing the value of a particular explanatory variable will have a positive or negative effect on a choice probability. This knowledge can be captured by using a prior probability density function (pdf) that is truncated to be positive or negative. Thus, three sets of results are compared: those from maximum likelihood (ML) estimation, those from Bayesian estimation with an unrestricted uniform prior on the coefficients, and those from Bayesian estimation with a uniform prior truncated to accommodate inequality restrictions on the coefficients.

The first Bayesian analysis of binary choice models in the econometrics literature was that of Zellner and Rossi (1984). They derived a normal approximation to the posterior pdf of the coefficients, and, focusing mainly on the logit model, showed how importance sampling can be used to find posterior pdfs for coefficients, probabilities, and elasticities. In line with the explosion of work using Markov Chain Monte Carlo methods (see, for example, Hill 1996), Albert and Chib (1993) show how data augmentation, in conjunction with the Gibbs sampler, can be used to

* William E. Griffiths is Professor in the Department of Economics, University of Melbourne. R. Carter Hill is Professor in the Department of Economics, Louisiana State University, Baton Rouge. Christopher J. O'Donnell is Senior Lecturer in the School of Economics, University of Queensland, Brisbane.

estimate posterior pdfs of interest for the probit model. They generalize the analysis to a *t*-distribution, showing how a Gibbs sampler can be set up in the context of a scale mixture of normal distributions. Extensions to the multinomial probit model were also considered by Albert and Chib, and later by Geweke, Keane, and Runkle (1994, 1997). Other related work includes Koop and Poirier (1993, 1994), McCulloch and Rossi (1994), Poirier (1994), and Wood and Kohn (1998). Inequality restrictions on the coefficients of logit models have been imposed by Allenby, Arora, and Ginter (1995). Prior pdfs for probabilities in probit models have been considered in the more general context of conditional mean priors in generalized linear models by Bedrick, Christiensen, and Johnson (1996).

The model for explaining choice between fixed and variable rate mortgages is described in the next section, and the results from ML estimation are used to motivate the likely consequences of Bayesian estimation. We then proceed to compare the results from Bayesian estimation with those from ML estimation. We find that the difference in results can be quite dramatic, particularly for the estimation of probabilities and elasticities.

Mortgage Example: Model and Preliminary Estimates

DHILLON, SHILLING, AND SIRMANS (1987) estimate a probit model designed to explain the choice by homebuyers of fixed versus adjustable rate mortgages. They use 78 observations from a bank in Baton Rouge, Louisiana, taken over the period January 1983 to February 1984. In this data set 46 fixed rate and 38 adjustable rate mortgages were chosen. Dhillon et al. used both financial measures and personal characteristics as explanatory variables in their model, and did not reject a hypothesis that the personal characteristics have no impact on the choice probability. We focus on the financial measures and introduce sign constraints in the form of inequality restrictions on the coefficients, using the signs implied by the discussion in Dhillon et al. The data are taken from Lott and Ray (1992).

The probit model can be written as

$$P_i = \Phi(x_i'\beta)\,, \tag{1}$$

where $\Phi(.)$ is the standard normal cumulative distribution function, β is a vector of unknown coefficients to be estimated, and, in the context of our example, P_i is the probability of choosing an adjustable rate mortgage. The vector of explanatory variables x_i is of dimension 7. Its components and the expected signs of the corresponding coefficients are:

$x_{1i} =$ 1

$x_{2i} =$ fixed interest rate $(\beta_2 > 0)$

$x_{3i} =$ margin = the variable rate less the fixed rate $(\beta_3 < 0)$

$x_{4i} =$ yield = the 10-year treasury rate less the one-year treasury rate $(\beta_4 < 0)$

$x_{5i} =$ points = ratio of points paid on adjustable rates to those paid on fixed rates $(\beta_5 < 0)$

$x_{6i} =$ maturity = ratio of maturities on adjustable to fixed rates $(\beta_6 < 0)$

$x_{7i} =$ net worth of borrower $(\beta_7 > 0)$

In general, the effect of a change in one of the explanatory variables (say the *k*-th variable) on the choice probability is given by the derivative

$$\frac{\partial P_i}{\partial x_{ki}} = \frac{\partial \Phi(x_i'\beta)}{\partial x_{ki}} = \beta_k \phi(x_i'\beta), \tag{2}$$

where $\phi(\cdot)$ is the standard normal probability density function. Since $\phi(\cdot)$ is always positive, an increase in x_k leads to an increase or a decrease in P_i depending on the sign of β_k. The fixed rate and the margin are designed to pick up cross-price and own-price effects, respectively, and hence their coefficients β_2 and β_3 are expected to be positive and negative, respectively. The yield variable represents a risk variable. The larger the yield the more likely it is that the adjustable rate will increase, and hence the less attractive is the adjustable rate mortgage $(\beta_4 < 0)$. Other things equal, the greater the relative points, the less likely an adjustable rate will be chosen $(\beta_5 < 0)$. Assuming shorter maturities are more desirable than longer ones, we have $\beta_6 < 0$. Finally, the greater the net worth of the borrower, the more likely is the borrower to take the risk of an adjustable rate $(\beta_7 > 0)$.

Unrestricted maximum likelihood estimates of the coefficients are given in Table 1. All estimates have the expected signs. However, 95 percent confidence intervals for some of the coefficients will include both positive and negative values and will hence have regions that are infeasible, in the sense that endpoints will have the wrong sign. The coefficients where this happens are those with *p*-values greater than 0.05, namely, those for fixed rate, points, and maturity. Thus, although Bayesian inequality-restricted estimation is unlikely to change the signs of estimated coefficients, it will have an impact on interval estimation, producing interval estimates without infeasible regions.

Table 1. Unrestricted Maximum Likelihood Estimates for Mortgage Data

Variable	Estimate	St. Error	t	p-value
constant	–1.877	4.225	–0.444	0.657
fixrate	0.499	0.277	1.799	0.072
margin	–0.431	0.174	–2.483	0.013
yield	–2.384	1.088	–2.191	0.028
points	–0.300	0.241	–1.242	0.214
maturity	–0.059	0.615	–0.096	0.923
networth	0.084	0.042	1.988	0.047

Mortgage Example: A Comparison of Bayesian and ML Estimates

TWO SETS OF BAYESIAN ESTIMATES ARE obtained: those from a uniform prior with no inequality constraints and those from a truncated uniform prior, truncated to accommodate the sign restrictions described above. Both of these prior pdfs can be described using the notation

$$f(\beta) \propto I_{\beta}(R), \tag{3}$$

where $I_{\beta}(R)$ denotes an indicator function, always equal to 1 when the coefficients are unconstrained, and, for the inequality-restricted case, equal to 1 when β is such that all the sign constraints are satisfied (β belongs to the feasible region R) and zero otherwise.

The likelihood function is given by

$$f(y \mid \beta) = \prod_{i=1}^{N} [\Phi(x_i'\beta)]^{y_i} [1 - \Phi(x_i'\beta)]^{1-y_i}, \tag{4}$$

where $y = (y_1, y_2, \ldots, y_N)'$ is a vector of binary variables, with $y_i = 1$ if the i-th observation is a variable rate mortgage and $y_i = 0$ if the i-th observation is a fixed rate mortgage. Using Bayes' Theorem, the posterior pdf for β can be written as

$$\begin{aligned} f(\beta \mid y) &\propto f(\beta) f(y \mid \beta) \\ &\propto I_{\beta}(R) \prod_{i=1}^{N} [\Phi(x_i'\beta)]^{y_i} [1 - \Phi(x_i'\beta)]^{1-y_i}. \end{aligned} \tag{5}$$

A random-walk Metropolis-Hastings algorithm (see, for example, Koop 2003, p. 92) was used to draw observations from the posterior pdf in equation (5). The ML estimate for β was chosen as the starting value, and a scalar multiple of the ML covariance matrix estimate was used as the covariance matrix in a normal candidate-generating distribution. The scalar was set such that candidates were accepted 40-50 percent of the time. A total of 200,000 draws were made, with the first 40,000 being discarded for a "burn in." Various tests for convergence were carried out; there was no evidence to suggest the chain had not achieved stationarity.

The posterior means and standard deviations for the coefficients appear in Table 2, alongside their maximum likelihood counterparts. The Bayesian estimates without the inequality restrictions imposed are very similar to those from ML. Also, with the exception of the coefficient on maturity, imposing the sign constraints has had only a small effect on the coefficient estimates. As expected, this effect is to increase the absolute value of the point estimates and to reduce estimation uncertainty (measured by the standard error in the case of maximum likelihood estimation and the posterior standard deviation in the case of Bayesian estimation). The dotted pdfs in Figure 1 (see page 11) are normal pdfs centered at the maximum likelihood estimates and with the corresponding standard errors as standard deviations. Given that these pdfs are used for sampling-theory interval estimation, they can be viewed as the sampling

Table 2. Maximum Likelihood and Bayesian Estimates for Mortgage Data[a]

Variable	ML	Bayes (unrestricted)	Bayes (restricted)
constant	–1.877	–1.819	–1.323
	(4.225)	(4.130)	(3.996)
fixrate	0.499	0.524	0.561
	(0.277)	(0.264)	(0.253)
margin	–0.431	–0.457	–0.494
	(0.174)	(0.178)	(0.176)
yield	–2.384	–2.554	–2.767
	(1.088)	(1.097)	(1.078)
points	–0.300	–0.330	–0.379
	(0.241)	(0.248)	(0.216)
maturity	–0.059	–0.098	–0.557
	(0.615)	(0.641)	(0.421)
networth	0.084	0.087	0.085
	(0.042)	(0.038)	(0.037)

[a] The numbers in parentheses are ML standard errors and posterior standard deviations.

theorists' posterior pdfs. Because the unrestricted coefficient for maturity is almost zero, with a large standard error, truncation has a big impact in this case; the mode of the Bayesian posterior pdf is close to zero, and the mean is almost 10 times larger than the maximum likelihood estimate. Interestingly, the other truncations have little effect, even for the coefficient of "points" where the ML pdf has noticeable probability above zero. The posterior pdfs from the unrestricted uniform-β prior were omitted from Figure 1 to reduce congestion. These posterior pdfs were almost identical to the "ML posterior pdfs."

The coefficients are useful for examining directions of probability changes that result from changes in the explanatory variables, but their magnitudes by themselves are not very informative. One is usually more interested in elasticities and probabilities evaluated at a particular point, x_*. These quantities are given, respectively, by

$$E_{k*} = \frac{\partial P_*}{\partial x_{k*}} \frac{x_{k*}}{P_*} = \beta_k x_{k*} \frac{\phi(x_*'\beta)}{\Phi(x_*'\beta)} \tag{6}$$

and

$$P_* = \Phi(x_*'\beta)\,. \tag{7}$$

To compare the results from ML estimation with those from unrestricted and inequality-restricted Bayesian estimation, two values for x_* were chosen, namely, observations 13 and 29 in the data set. Also, for the elasticities we focused on two of the more important explanatory variables: margin and yield. Observations 13 and 29 were chosen because they led to quite different estimated probabilities, one about 0.9 and the other about 0.05. Their characteristics and how they stand relative to the whole sample are given in Table 3. Note that the major difference between the two observations is in net worth of the borrower.

Table 3. Characteristics of Explanatory Variables for Mortgage Data

Variable	Mean	Minimum	Maximum	Obs. 13	Obs. 29
fixrate	13.25	11.76	14.50	13.5	12.13
margin	2.292	–0.90	5.50	2.5	3.36
yield	1.606	1.38	2.04	1.59	1.60
points	1.498	0.00	4.34	1.00	1.66
maturity	1.058	0.42	2.38	1.00	0.85
networth	3.504	–0.056	17.86	17.86	0.118

Table 4. Estimated Probabilities and Elasticities for Mortgage Data[a]

	Obs. 13			Obs. 29		
	ML	Bayes (unrest'd)	Bayes (rest'd))	ML	Bayes (unrest'd)	Bayes (rest'd))
Probability	0.879 (0.130)	0.856 (0.118)	0.865 (0.105)	0.052 (0.041)	0.056 (.043)	0.058 (0.043)
Elasticity (Margin)	–0.237 (0.054)	–0.281 (0.239)	–0.286 (0.230)	–2.966 (2.917)	–3.327 (1.572)	–3.566 (1.578)
Elasticity (Yield)	–0.869 (0.206)	–1.050 (0.938)	–1.070 (0.898)	–7.814 (8.183)	–8.767 (4.287)	–9.387 (4.266)

[a] The numbers in parentheses are ML standard errors and posterior standard deviations.

Table 4 contains ML estimates and Bayesian posterior means for the elasticities and probabilities at observations 13 and 29, along with their corresponding standard errors and posterior standard deviations. The ML standard errors are obtained using the conventional first-order approximation for the asymptotic variance of a nonlinear function of the maximum likelihood estimator. See, for example, Judge et al. (1985, p. 160). If $\hat{\beta}$ denotes the maximum likelihood estimator, then $\hat{P}_* = \Phi(x_*'\hat{\beta})$ and $\hat{E}_{k*} = \hat{\beta}_k x_{k*} \phi(x_*'\hat{\beta}) / \Phi(x_*'\hat{\beta})$, and the asymptotic variances can be derived as

$$\mathrm{var}(\hat{P}_*) = [\phi(x_*'\hat{\beta})]^2 x_*' V x_* \tag{8}$$

and

$$\mathrm{var}(\hat{E}_{k*}) = x_{k*}^2 \, \mathrm{var}[\hat{\beta}_k \phi(x_*'\hat{\beta}) \big/ \Phi(x_*'\hat{\beta})], \tag{9}$$

where V is the covariance matrix for $\hat{\beta}$, and the variance term in equation (9) is given by the k-th diagonal element of

$$\mathrm{cov}\left[\hat{\beta}\, \frac{\phi(x_*'\hat{\beta})}{\Phi(x_*'\hat{\beta})}\right] = \left[\frac{\phi(x_*'\beta)}{\Phi(x_*'\beta)}\right]^2 QVQ' \tag{10}$$

with

$$Q = I - \left[x_*'\beta + \phi(x_*'\beta)/\Phi(x_*'\beta)\right]\beta x_*'. \tag{11}$$

For the Bayesian posterior means and standard deviations we computed the sample means and standard deviations of the 160,000 MCMC generated observations, calculated using the expressions in equations (6) and (7).

Looking first at the ML results in Table 4, we find that conventional 95 percent confidence intervals for the probabilities will include infeasible values. The one for observation 13 will contain values greater than one, while that for observation 29 will contain values less than zero. Confidence intervals for the elasticities include only the expected negative values in the case of observation 13, but the large standard errors in the case of observation 29 will lead to confidence intervals that contain a substantial region of positive values. When we talk of confidence intervals, we are assuming that the usual large-sample practice of deriving intervals from the normal distribution is being employed. The higher standard errors for $\hat{E}_{29}$ relative to those of $\hat{E}_{13}$ are likely to be a consequence of values for $\Phi(x'_{29}\beta)$ close to zero. The appearance of $\Phi(x'_{*}\beta)$ in the denominator of equation (6) means that values close to zero are likely to cause more instability.

A comparison of the Bayesian and ML estimates of the probabilities in Table 4 suggests little difference between the results. Making this judgment on the basis of posterior means and standard deviations is, however, misleading. The pdfs graphed in Figure 2 (see page 12) for the ML estimates and for the posteriors from the unrestricted prior show that there is a considerable difference. If one blindly uses the normal distribution to construct interval estimates for P_{13} and P_{29} on the basis of the maximum likelihood results, the interval estimates will include negative probabilities and probabilities that exceed one. By using the empirical pdf from the MCMC-generated observations, Bayesian estimation overcomes this problem. The posterior pdfs for P_{13} and P_{29} are truncated at 1 and 0, respectively, and are far from being symmetrical like the ML pdfs. The standard deviations of the ML and posterior pdfs are similar, despite the truncations, because of the longer tails of the posterior pdfs. Graphs of the posterior pdfs for P_{13} and P_{29} derived from the inequality-restricted prior were very similar to those from the uniform prior and hence were omitted from Figure 2. Thus, imposing inequality restrictions on the coefficients had little effect on the posterior pdfs for the probabilities. The large differences between the ML and Bayesian results can be attributed to the small-sample inference properties of Bayesian estimation relative to the asymptotic approximation used for ML estimation, not to the prior information utilized by Bayesian inference.

Inference about elasticities is also very sensitive to whether one opts for large-sample ML inference or finite-sample Bayesian inference. Note the dramatic differences in the standard errors (deviations) in Table 4 and the differences in spread in Figure 3 (see page 13). It would appear that, when P_i is close to one, ML estimation overstates the precision with which the elasticities are estimated; when P_i is close to zero, ML understates this precision. We conjecture that the term $\phi(x_i'\beta)/\Phi(x_i'\beta)$ in equations (10) and (11) helps explain this phenomenon. It could be too small when evaluated at a point estimate where its denominator P_i is close to one, and too large when evaluated at a point estimate where its denominator P_i is close to zero. There was very little difference between the two sets of Bayesian results for the margin and yield elasticities. The posterior pdfs from the inequality-restricted prior are truncated at zero from above, but the effect of this truncation is minimal. As is evident from Figure 3, the posterior pdfs from the unrestricted prior have small probability content above zero.

Overall, we conclude that, when using a uniform unrestricted prior on the coefficients, Bayesian and ML estimation lead to almost identical inferences about the coefficients, but can produce vastly different inferences about probabilities and elasticities. Inequality restrictions have the expected effect on coefficient estimation, truncating the posterior pdfs of coefficients with low *t*-values. The effect of the inequality restrictions on inferences about probabilities was minimal. In line with their impact on the coefficients, the inequality restrictions will have a similar effect on the elasticities; the two variables whose elasticities we examined in detail were not largely affected by the restrictions.

Conclusion

The probit model is commonly used for modeling binary choice decisions in economics. It is also one for which prior information, in one form or another, is likely to exist. Bayesian estimation provides a convenient vehicle for including such information, but it does raise questions about the sensitivity of results to the choice of estimation technique and to the specification of alternative forms of prior information. We have examined the sensitivity of results to a uniform prior on the coefficients and a uniform but inequality restricted prior on the coefficients. We also considered ML estimates and we examined the sensitivity of inferences on coefficients, probabilities, and elasticities. Our main findings are summarized as follows:

- When a uniform prior on the coefficients is specified, Bayesian and ML estimation lead to similar inferences about the coefficients, but quite different inferences about probabilities and elasticities can occur. The likely reason for any differences is poor asymptotic approximations to the distributions of ML estimators for probabilities and elasticities. In particular, those distributions may not preclude probability values outside the (0,1) interval.

- Using a prior that imposes inequality restrictions on the coefficients truncates the posterior pdfs on the coefficients and the elasticities in the expected way, but does not necessarily impact on the posterior pdfs for probabilities.

References

Albert, J., and S. Chib. "Bayesian Analysis of Binary and Polychotomous Response Data." *Journal of the American Statistical Association* 88(1993): 669-679.

Allenby, G.M., N. Arora, and J.L. Ginter. "Incorporating Prior Knowledge into the Analysis of Conjoint Studies." *Journal of Marketing Research* 32(1995): 152-162.

Amemiya, T. "Qualitative Response Models: A Survey." *Journal of Economic Literature* 19(1981): 1483-1536.

Bedrick, E.J., R. Christensen, and W. Johnson. "A New Perspective on Priors for Generalized Linear Models." *Journal of the American Statistical Association* 91(1996): 1450-1460.

Dhillon, U.S., J.D. Shilling, and C.F. Sirmans. "Choosing Between Fixed and Adjustable Rate Mortgages." *Journal of Money, Credit and Banking* 19(1987): 260-267.

Geweke, J., M. Keane, and D. Runkle. "Alternative Computational Approaches to Inference in the Multinomial Probit Model." *Review of Economics and Statistics* 76(1994): 609-632.

Geweke, J., M. Keane, and D. Runkle. "Statistical Inference in the Multinomial Multiperiod Probit Model." *Journal of Econometrics* 80(1997): 125-166.

Hill, R.C., ed. *Advances in Econometrics Volume 11A: Bayesian Computational Methods and Applications*. Greenwich: JAI Press, 1996.

Judge, G.G., W.E. Griffiths, R.C. Hill, H. Lütkepohl, and T.-C. Lee. *The Theory and Practice of Econometrics*, 2nd edition. New York: John Wiley and Sons, 1985.

Koop, G. *Bayesian Econometrics*. New York: John Wiley and Sons, 2003.

Koop, G., and D. Poirier. "Bayesian Analysis of Logit Models Using Natural Conjugate Priors." *Journal of Econometrics* 56(1993): 323-340.

Koop, G., and D. Poirier. "Rank-Ordered Logit Models: An Empirical Analysis of Ontario Voter Preferences before the 1988 Canadian Federal Election." *Journal of Applied Econometrics* 9(1994): 369-388.

Lott, W.F., and S.C. Ray. *Applied Econometrics: Problems with Data Sets.* Orlando: Dryden Press, 1992.

Maddala, G.S. *Limited Dependent and Qualitative Variables in Econometrics.* Cambridge: Cambridge University Press, 1983.

McCulloch, R., and P.E. Rossi. "An Exact Likelihood Analysis of the Multinomial Probit Model." *Journal of Econometrics* 64(1994): 207-240.

Poirier, D. "Jeffreys' Prior for Logit Models." *Journal of Econometrics* 63(1994): 327-339.

Wood, S., and R. Kohn. "A Bayesian Approach to Robust Binary Nonparametric Regression." *Journal of the American Statistical Association* 93(1998): 203-213.

Zellner, A., and P.E. Rossi. "Bayesian Analysis of Dichotomous Quantal Response Models." *Journal of Econometrics* 25(1984): 365-393.

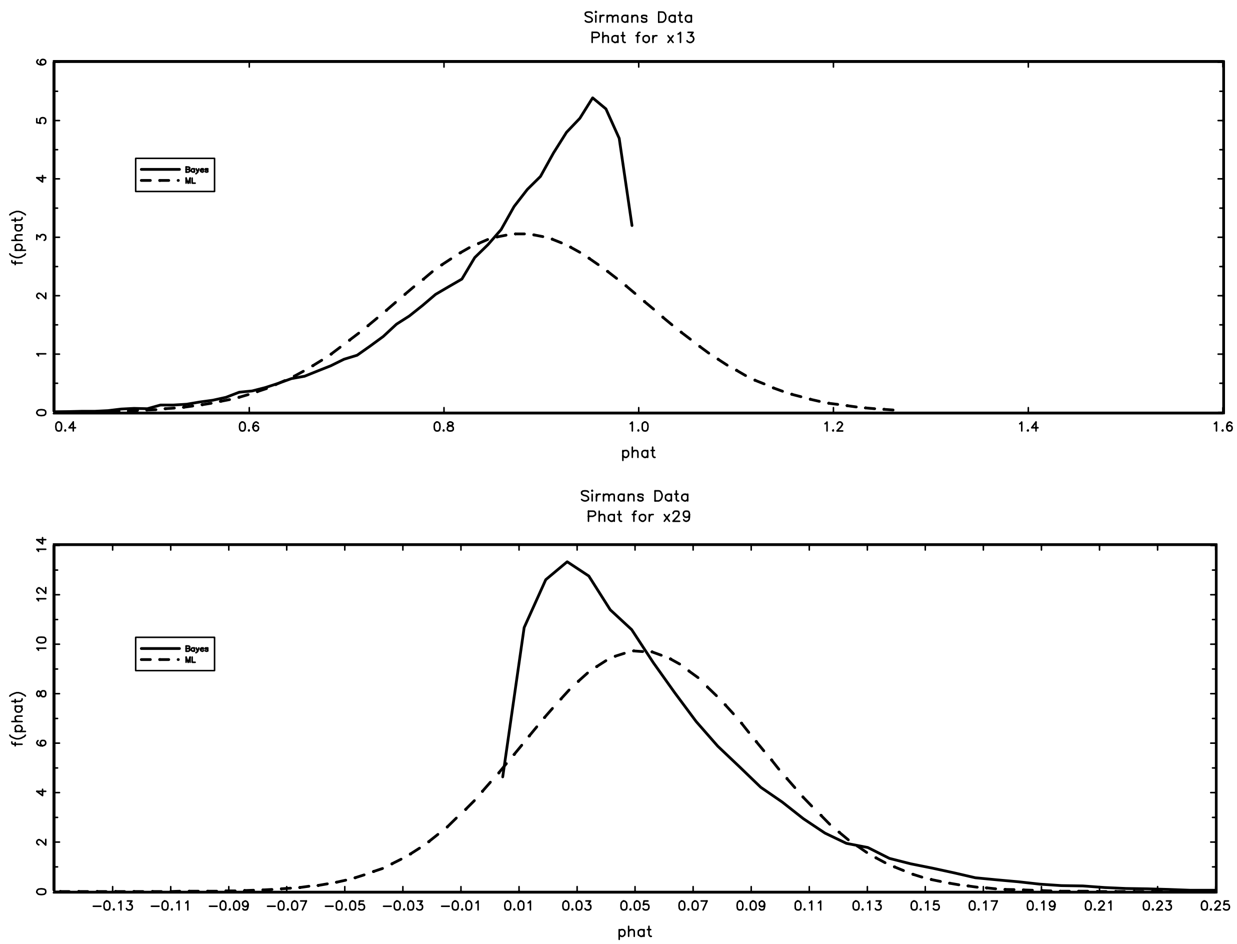

Figure 2. Posterior PDFs for Probabilities for Mortgage Data: ML and Bayes from Uniform Prior

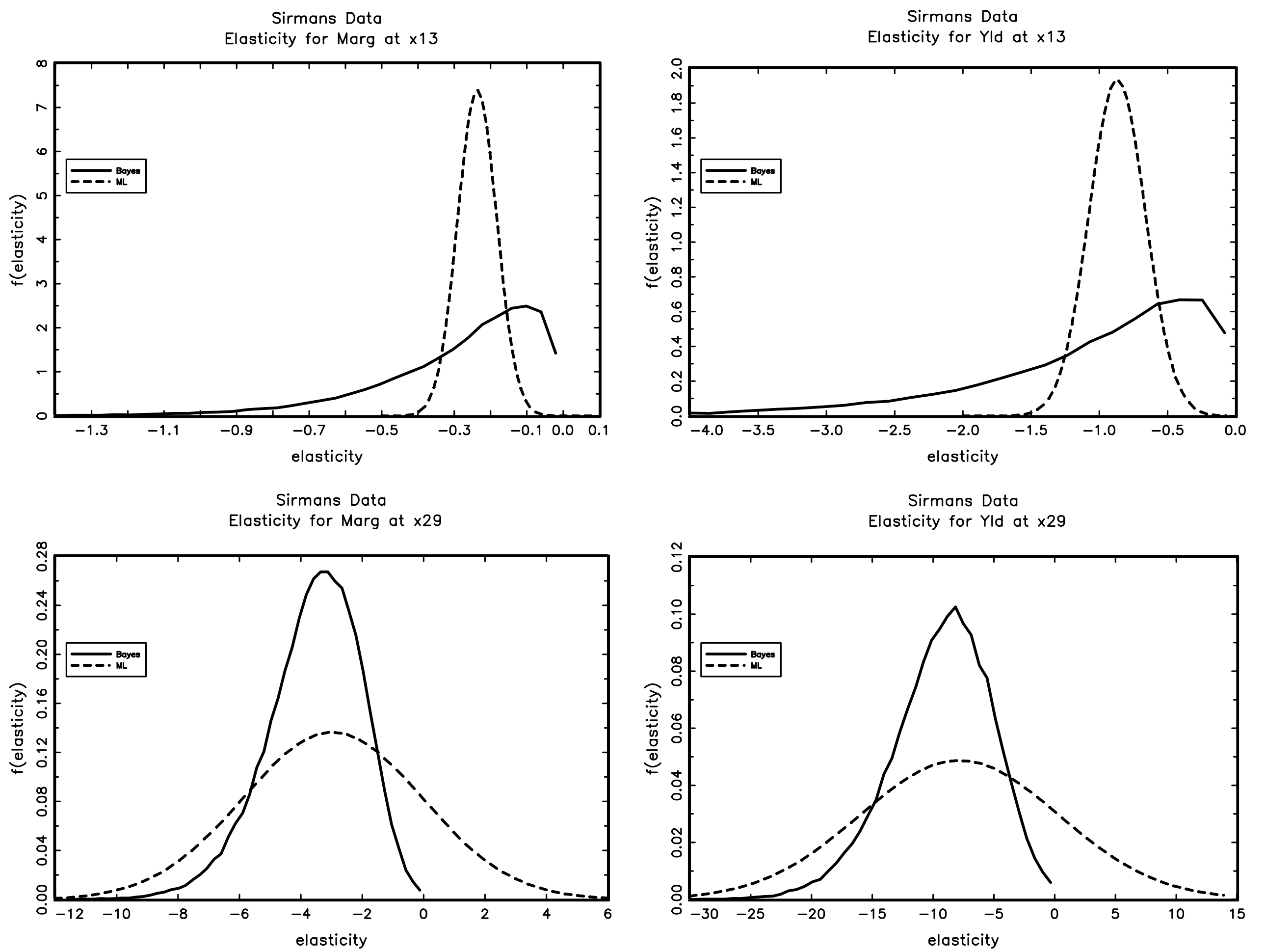

Figure 3. Posterior PDFs for Elasticities for Mortgage Data: ML and Bayes from Uniform Prior