ECONOMICS BULLETIN

Are credit constraints in Italy really more binding in the South?

Claudio Lupi
University of Molise, Dept. SEGeS

Abstract

This paper is motivated by a very practical question: are there significant geographical differences in the accessibility to the credit market on the part of Italian households? The investigation is carried using robust probit model. Estimation is carried out in a Bayesian framework. The results are somewhat surprising, showing that the area where households are more likely to be credit constrained is not the South, as could be easily imagined, but rather the highly developed and industrialized North−West.

I would like to thank,without implicating, Agostino Nobile and Sergio de Nardis for comments and suggestions. This paper uses K.Abadir and J.Magnus' (2002) notation and LaTeX commands.
Citation: Lupi, Claudio, (2005) "Are credit constraints in Italy really more binding in the South?." *Economics Bulletin*, Vol. 3, No. 35 pp. 1−6
Submitted: March 29, 2005. **Accepted:** July 20, 2005.
URL: http://www.economicsbulletin.com/2005/volume3/EB−05C10004A.pdf

1. Introduction

This paper is motivated by a very practical question: are there significant geographical differences in the accessibility to the credit market on the part of Italian households? The interest in this question is related to the implications for growth and the labor market (see, among others, Brunello *et al.*, 2005; Guiso *et al.*, 2004).

A closely related (yet different) issue has been briefly considered in Guiso *et al.* (2004). However, Guiso *et al.*'s main goal is rather to build an indicator of financial development that takes account of other credit-related regional features in the framework of a linear probability model.

In the present paper I want to investigate if households with identical profiles experience significantly different chances of being credit constrained in different areas of the country. The focus is therefore on households' characteristics, and any extra-household feature is considered as part of the regional economic context that may influence the probability, on the part of the households, of being credit-constrained.

The investigation is carried out in a probit framework that allows for some households heterogeneity. As is well known, while heteroscedasticity in linear regression affects estimated standard errors only, in discrete choice models the presence of heteroscedasticity not only adversely affects estimated standard errors, but also makes the estimated parameters biased. In order to allow for some heterogeneity among the households, I consider a robust probit (sometimes called *robit*) model. The model is based on the idea that the link function has thicker tails than the normal distribution.

Estimation is carried out in a Bayesian framework. The advantage of this choice is threefold:

1. Estimation can be carried out in a rather natural framework using Gibbs sampling and data augmentation (see Albert and Chib, 1993);

2. Parameters densities can be easily obtained and there is no need to rely on asymptotic results;

3. Inference can be extended to derived quantities that combine parameters and data.

The paper is organized as follows: Section 2 is devoted to a brief description of the data and the statistical model. The empirical results are illustrated in Section 3.

2. The data and the model

Data are from the 2002 Bank of Italy's Survey of Households Income and Wealth (SHIW).[1] The sample is constituted by the 434 (out of 8011) households that declared having applied for a loan or a mortgage in 2002 or being interested in applying, but having not applied because discouraged, thinking that the application was going to be refused.[2] The dichotomous dependent variable is 1 (*i.e.* the household is credit constrained) if the application was refused or the household discouraged from applying, 0 otherwise.

A number of household characteristics are observed and used to model the probability of being credit constrained (see Table 1).

[1] At the time of writing, this is the most recent available survey.

[2] Unfortunately, there are no information about the amount of the loan.

Parameter	Variable	
b1	HEADY	Head of household's income
b2	HFAGE	Head of household's age
b3	LOWEDU	The head of household has at most an elementary education
b4	MIDEDU	The head of household has a middle school education
b5	PROEDU	The head of household has a professional education
b6	HSCEDU	The head of household has a high school diploma
b7	INDWOR	The head of household is an independent worker
b8	RETWOR	The head of household is retired
b9	UNEMPL	The head of household is unemployed
b10	NOWOR	The head of household is out of the labor force
b11	HOWNER	The family owns the house since before 2002
b12	SMALLC	The family lives in a small city (less than 20,000 inhabitants)
b13	LARGEC	The family lives in a large city (more than 40,000 inhabitants)
b14	NOBANK	None in the family has a bank account
g1	NEAST	The family lives in the North East
g2	CENTR	The family lives in the Centre
g3	SOUTH	The family lives in the South

Table 1: Variable names, definitions, and associated parameters. The reference household lives in an average city (20,000-40,000 inhabitants) of the North West, possesses one or more bank accounts, did not own the house where it lives: the household's head has a university degree and works as an employee. All the variables are from the Bank of Italy's Survey of Households Income and Wealth (SHIW 2002).

The main goal of this paper is to find an interval estimate of the *mean marginal effect*[3] of the regional dummy variables.[4] In the present context, the mean marginal effect of these variables is taken to represent the strength of the bindings on credit availability.

The statistical model is fully described in Albert and Chib (1993) and only a brief account is given here. Assume that $y_i = 1$ $(i = 1, 2, \ldots, n)$ if the continuous latent variable $y_i^* = \boldsymbol{x}_i'\boldsymbol{\beta} + \epsilon_i > 0$ and $y_i = 0$ otherwise. In the standard probit framework one assumes $\boldsymbol{\varepsilon}|\boldsymbol{X}, \boldsymbol{\beta} \sim \mathrm{N}(0, \boldsymbol{I}_n)$ so that $\Pr(y_i = 1) = \Phi(\boldsymbol{x}_i'\boldsymbol{\beta})$, with $\Phi(\cdot)$ denoting the standard normal cumulative distribution function. In order to allow for some heterogeneity it is possible to assume instead $\boldsymbol{\varepsilon}|\boldsymbol{X}, \boldsymbol{\beta}, \boldsymbol{\Lambda} \sim \mathrm{N}(0, \boldsymbol{\Lambda})$ with $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$. This is equivalent to assuming $\epsilon_i|\boldsymbol{X}, \boldsymbol{\beta}, \nu \sim \mathrm{t}(\nu, \boldsymbol{x}_i'\boldsymbol{\beta})$. Indeed, $z_i \sim \mathrm{t}(\nu, \mu)$ can be viewed in terms of a mixture of normals as $z_i|V_i \sim \mathrm{N}(\mu, V_i)$ with $V_i \sim \text{Inv-}\chi^2(\nu, \sigma^2)$ (see Gelman *et al.*, 2004). Albert and Chib (1993) show that such a model can be estimated using the Gibbs sampler with data augmentation. Both the latent variables and the individual variances are treated as missing data.

Albert and Chib's algorithm in principle allows one to leave the degrees of freedom parameter ν unspecified and let the procedure to estimate it. However, this proves often to be impractical (Gelman *et al.*, 2004) so that a low value for this parameter is usually chosen *a priori* to ensure robustness. Some tentative exploratory analysis on the present data set indicated that a reasonable value might be 4, but convergence problems tended

[3]In a probit model, the mean marginal effect of the j-th explanatory variable over a sample of n households is defined as $n^{-1}\sum_{i=1}^{n} \partial \Pr(y_i = 1)/\partial x_{ji}$ and involves both estimated parameters and data. This complicates inference in a standard frequentist setting. The mean marginal effect should not be confused with the so-called *mean derivative*.

[4]In this paper, the Italian National Statistical Institute's official classification is used. North-West is composed of Piemonte, Valle d'Aosta, Lombardia and Liguria (130 observations in the sample); North-East is composed of Trentino, Veneto, Friuli Venezia Giulia and Emilia Romagna (91 observations); Centre includes Toscana, Umbria, Marche and Lazio (108 observations); the South coincides with Abruzzo, Molise, Campania, Puglia, Basilicata, Calabria, Sicilia and Sardegna (105 observations).

Parameter	$\hat{R}$	Mean	SD	Batch SE	Batch ACF	Q(0.05)	Q(0.5)	Q(0.95)
b1	0.9998	0.0090	0.0608	0.0009	-0.1882	-0.0914	0.0102	0.1063
b2	0.9997	-0.0152	0.0095	0.0001	0.1879	-0.0311	-0.0152	0.0003
b3	0.9998	0.7653	0.4170	0.0045	-0.3375	0.0852	0.7615	1.4685
b4	0.9999	0.5408	0.3781	0.0041	-0.2061	-0.0648	0.5331	1.1776
b5	1.0005	0.0254	0.4781	0.0048	-0.3966	-0.7466	0.0132	0.8209
b6	0.9998	0.2576	0.3974	0.0043	-0.3057	-0.3759	0.2466	0.9352
b7	0.9998	0.2600	0.6024	0.0083	-0.2430	-0.7392	0.2739	1.2177
b8	1.0003	0.6162	0.2876	0.0039	0.0645	0.1472	0.6132	1.0925
b9	1.0004	2.0964	0.8846	0.0154	-0.0277	0.8673	1.9828	3.6818
b10	0.9999	0.3128	0.6260	0.0079	-0.1299	-0.7248	0.3248	1.3171
b11	0.9996	-0.5096	0.1754	0.0022	-0.0677	-0.8047	-0.5085	-0.2253
b12	0.9998	-0.2979	0.2107	0.0028	-0.0893	-0.6476	-0.2974	0.0419
b13	1.0005	-1.1143	0.3762	0.0053	-0.1261	-1.7689	-1.0982	-0.5348
b14	1.0000	1.0623	0.2619	0.0029	-0.2107	0.6364	1.0601	1.4928
c	0.9995	0.3218	0.7319	0.0095	-0.2767	-0.8636	0.3237	1.5181
g1	1.0001	-1.0200	0.2444	0.0030	-0.1805	-1.4294	-1.0162	-0.6215
g2	1.0003	-0.3699	0.2125	0.0031	-0.0723	-0.7205	-0.3694	-0.0211
g3	0.9999	-0.9986	0.2515	0.0028	-0.1915	-1.4228	-0.9942	-0.5957
Mult. $\hat{R}$	1.0888							

Table 2: Descriptive statistics are based on 6000 iterations (6 independent chains of 1000 iterations each). $\hat{R}$ and Mult. $\hat{R}$ are the potential scale reduction (Gelman & Rubin, 1992) and the multivariate potential scale reduction (Brooks & Gelman, 1998), respectively. Batch SE is an estimate of the standard error calculated as the sample standard deviation of the means from consecutive batches of size 50 divided by the square root of the number of batches. Batch ACF is the sample autocorrelation between batch means. Q(0.05), Q(0.5), and Q(0.95) are the 5% quantile, the median, and the 95% quantile of the distribution.

to arise for some parameters. Therefore in the final estimation the degrees of freedom parameter has been set at $\nu = 4$.

3. Empirical results

Six independent chains of length 1000 (excluding the burn-in replications) have been used to derive the posterior distributions of the parameters.[5] The starting values of the chains were randomly selected using perturbed OLS estimates. In order to reduce autocorrelation, a thin factor of 10 has also been used. Convergence was fairly slow for some parameters, but it was reached for all the parameters of the model, as can be seen from Table 2 that lists the potential scale reduction factor $\hat{R}$ (Gelman and Rubin, 1992; Gelman *et al.*, 2004) for each parameter. The multivariate potential scale reduction factor (Brooks and Gelman, 1998) is also reported.

Table 2 also shows the main descriptive statistics for all the estimated parameters. From this Table it is possible to verify that various household's characteristics are significant in shaping the probability of being credit constrained. Note in particular the role of characteristics related to the head of the household. Young heads of household are favored in obtaining credit. Low educated, retired, or unemployed heads of household experience instead a significantly higher probability of being credit constrained. The same applies to families whose components do not have any bank account, while being

[5]Post-estimation output analysis has been performed using BOA 1.1.4 (Smith, 2005).

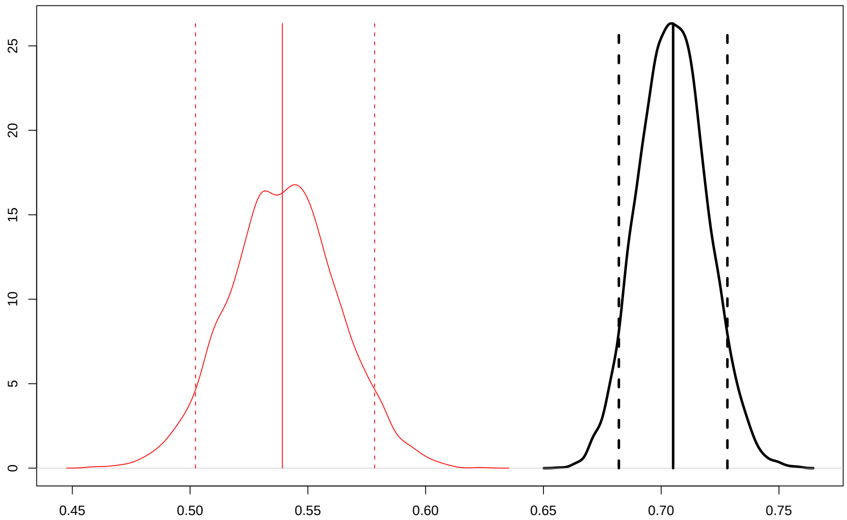

Figure 1: The distribution of the model's hit rate (thick black line) compared to the distribution of the hit rate simulated from a Bernoulli distribution with parameter $\theta = n^{-1}\sum_{i=1}^{n} y_i$ (thin red line). Solid vertical lines are the medians of the distributions; dashed vertical lines are the 5% and 95% quantiles.

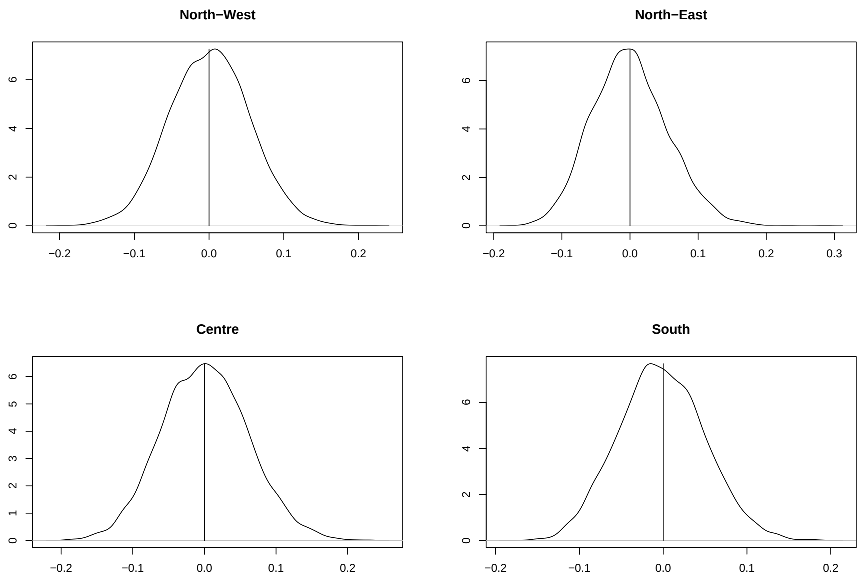

Figure 2: Posterior predictive distributions of the proportion of credit constrained households minus the sample proportion.

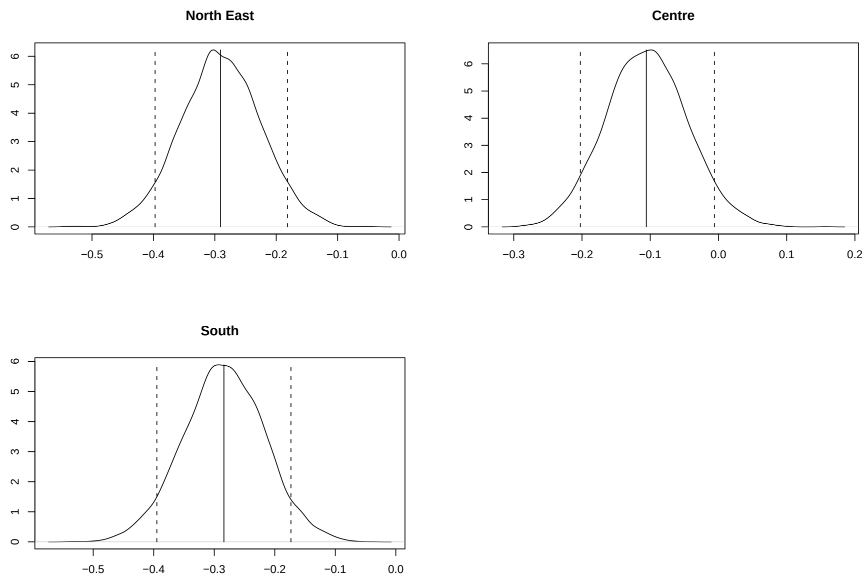

Figure 3: Distributions of the mean marginal effects of the three geographical dummies. Solid vertical lines are the medians of the distributions; dashed vertical lines are the 5% and 95% quantiles.

a house owner of course tends to lessen the binding. The dimension of the city where the family lives also seems to influence the probability of obtaining a loan on the part of the households, with small and very large cities offering more favorable environments. As expected, regional dummies are also significant.

The hit rate of the model is on average slightly above 70%, which is significantly higher than what could be obtained from random draws from i.i.d. Bernoulli random variables with parameter θ equal to the observed sample average of the dependent variable (see Figure 1).

Furthermore, the model seems able to replicate the proportion of constrained households in the four areas of the country (see Figure 2).

The main result of the paper is summarized in Figure 3 that reports the distributions of the mean marginal effects of the three geographical dummies.

Households in the North-East and in the South are 20 to 40 per cent less likely to be credit rationed than those living in the industrialized North-West: those in the Centre are up to 20% less likely to be credit constrained. This is a somewhat surprising result. However, this does not necessarily mean that the credit market is more developed in the South than in the North-West (See Guiso *et al.* 2004 on this aspect). Nor it does mean that the conditions to which household borrow in the South are more convenient than those applied in the North-West.[6]

Indeed, banks in the North-West seem to have suffered more from the recent Italian economic difficulties. This might have suggested a particularly prudent loan policy.

References

[1] Abadir K.M. & Magnus J.R. (2002) "Notation in Econometrics: A Proposal for a Standard", *Econometrics Journal*, **5**, 76-90.

[6] However, according to SHIW, the median overdraft rate is slightly higher in the North-West (9%), than it is in the North-East (7.5%) or in the Centre and the South (8%).

[2] Albert J.H. & Chib S. (1993) "Bayesian Analysis of Binary and Polychotomous Response Data", *Journal of the American Statistical Association*, **86**, 669-679.

[3] Brooks S. & Gelman A. (1998) "General Methods for Monitoring Convergence of Iterative Simulations", *Journal of Computational and Graphical Statistics*, **7**, 434-455.

[4] Brunello G., Lupi C. & Ordine P. (2005) "Labor Taxes in Local Labor Markets: Evidence from Italian Regions", *Giornale degli economisti e annali di economia*, forthcoming.

[5] Gelman A., Carlin J.B., Stern H.S & Rubin D.B. (2004) *Bayesian Data Analysis.* Boca Raton, Chapman & Hall / CRC.

[6] Gelman A. & Rubin D.B. (1992) "Inference From Iterative Simulation Using Multiple Sequences", *Statistical Science*, **7**, 457-511.

[7] Guiso L., Sapienza P. & Zingales L. (2004) "Does Local Financial Development Matter?", *Quarterly Journal of Economics*, **119**, 929-969.

[8] R Development Core Team (2004) *R: A language and environment for statistical computing.* Vienna, R Foundation for Statistical Computing. http://www.R-project.org.

[9] Smith B.J. (2005) "Bayesian Output Analysis Program (BOA) for MCMC", R package version 1.1.4-2. http://www.public-health.uiowa.edu/boa