Economics Letters 127 (2015) 54–57

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

# Semiparametric estimation of default probability: Evidence from the Prosper online credit market

Xiaofeng Li, Ying Shang *, Zhi Su

*International School of Economics & Management, Capital University of Economics & Business, Beijing, PR China*

## HIGHLIGHTS

- We examine the effects of a person's past financial characteristics on his likelihood to default.
- Borrowers with higher credit score rankings usually have lower probability to default.
- A borrower with score ranking B is less likely to default than a ranking A borrower.
- The semiparametric estimator outperforms the Probit estimator.
- A model specification test rejects the null hypothesis of the Probit specification at 5% level.

ARTICLE INFO

*Article history:*
Received 13 June 2014
Received in revised form
7 November 2014
Accepted 30 November 2014
Available online 6 December 2014

*JEL classification:*
C14

*Keywords:*
Semiparametric method
Single index model
Default probability

ABSTRACT

This paper examines the effects of a person's past financial characteristics on his likelihood to default in ex-post loan performance using both Probit and a semiparametric single-index estimator proposed by Klein and Spady (1993). The data used in the paper are a sample of individual loans generated on Prosper, a large US online lending market. The out of sample predictions and the model specification test suggest a misspecification of the Probit model due to the violation of the normality assumption. Estimation results suggest that a borrower's past financial credit score is a reasonably good indicator of one's loan performance. In general, the higher one's credit score ranking, the lower the probability that one would default. One exceptional finding is that a borrower with score ranking B is less likely to default than a borrower with score ranking A. Such a finding suggests that individuals who are in the middle range of credit grades may be more financially credit-dependent than those with higher rankings. As a result, they are more willing to keep their loans in good standings.



## 1. Introduction

Loan performance is a crucial factor for the returns to investors in the credit market. Bankruptcy and delinquency can create enormous losses for lenders and raise the cost of credit for borrowers. Most empirical studies of default focus on why the default rate has increased over time even in strong economic periods. These studies explore two explanations. One is the possibility of an increase in the number of less credit worthy borrowers due to the increased accessibility to credit (e.g. Moss and Johnson, 1999 and Mian and Sufi, 2009). The second explanation is the possibility of decreased default costs including social, information and legal costs due to the increase in the number of bankruptcy lawyers and the increased availability of the relevant legal information. See White (1998), Fay et al. (1998), Gross and Souleles (2002), among others.

In this paper, we examine the effects of a person's past financial characteristics on his likelihood to default in ex-post loan performance using a semiparametric single-index estimator proposed by Klein and Spady (1993). We also compare the estimates from the semiparametric model with the parametric Probit model. We find that estimates for both the semiparametric and parametric models have the same signs as expected. The relative absolute magnitude of estimates is larger for semiparametric estimators. On the other hand, the standard errors are much smaller in the semiparametric model and therefore the estimates are more efficient than in the standard Probit model. The out-of-sample prediction shows that the semiparametric model has better predictive performance than the Probit model. The differences in predicted default probability from the two models with same fitted index values imply that the standard Probit model may have a misspecification problem. To test the validity of the Probit model for this data, we

* Corresponding author.
*E-mail address:* shangying@cueb.edu.cn (Y. Shang).

http://dx.doi.org/10.1016/j.econlet.2014.11.033
0165-1765/

performed a model specification test using a wild bootstrap method and the result rejects the Probit specification at 5% level. Moreover, the empirical estimates suggest that a borrower's past financial credit score is a reasonably good indicator for one's ex-post loan performance. In general, the higher one's credit score ranking, the lower the probability that one would default. One exceptional finding is that a borrower with score ranking B is less likely to default than a borrower with score ranking A. Such a result may indicate that borrowers with score B may face tighter financial constraints than those with score A. As a result, they want to improve their credit ratings by avoiding default.

## 2. Estimation methodology

Our estimation methodology is based on the semiparametric estimator proposed by Klein and Spady (1993). Their estimator for discrete choice models makes no parametric assumption on the form of the distribution generating the error terms. However, the estimator does assume that the default probability function depends on a parametrically specified index function. Consider the following binary choice model where the dependent variable $Y$ is a vector of binary random variables that take on value one when there is a default and zero otherwise. The matrix $X$ is composed as $k$ vectors of explanatory variables of individuals' available financial characteristics. The parameter $\beta$ is a $k \times 1$ vector, estimated from data $\{Y_i, X_i\}$, $i = 1, \ldots, n$. Assume the default probability function depends on a single index, the conditional probability of $Y = 1$ given $X = x$ is

$$P(Y = 1|X = x) = E(Y|X) = G(X'\beta_0), \tag{2.1}$$

where $G$ is an unknown continuous function and the last equality holds because of the index specification. Index restriction permits heteroscedasticity of unknown form if it depends only on the index. To estimate the unknown function $G$, Klein and Spady (1993) propose two types of kernel estimation. One is for *Bias Reduction* where bandwidth is a fixed window and a further condition is imposed to the kernel selection. The other is *Adaptive with Local Smoothing* where the bandwidth is data dependent and adjusted by the standard deviation of the fitted values.[1] We adopt the first method where $G$ is estimated by the nonparametric estimator $G_n$, which is a kernel regression given by

$$G_n(X_i'\beta) = \frac{\sum_{j=1}^{n} Y_j K\left[(X_i - X_j)'\beta/h_n\right]}{\sum_{j=1}^{n} K\left[(X_i - X_j)'\beta/h_n\right]}. \tag{2.2}$$

The semiparametric estimates of $\beta$ are then obtained by maximizing the following quasi-loglikelihood function

$$LogL_n(\beta) = n^{-1}\sum_{i=1}^{n}\left\{Y_i \log G_n(X_i'\beta) + (1 - Y_i)\log\left[1 - G_n(X_i'\beta)\right]\right\}. \tag{2.3}$$

Replacing $G_n(\cdot)$ with the cumulative normal distribution function $\Phi(\cdot)$ in the log likelihood function, results in the standard Probit estimation.

Klein and Spady (1993) show that the asymptotic distribution of $n^{1/2}(\hat{\beta} - \beta_0)$ is $N(0, \Sigma)$, where the explicit expression of the variance–covariance matrix $\Sigma$ can also be found in their paper.

**Table 1**
Summary statistics.

| Variables | Mean | Std.Dev | Min | Max |
|---|---|---|---|---|
| Default (dependent variable) | 0.078 | 0.269 | 0 | 1 |
| Borrower maximum rate | 0.198 | 0.066 | 0.05 | 0.30 |
| Amount funded | 7194.514 | 6123.126 | 1000 | 25,000 |
| Close when funded (dummy) | 0.242 | 0.428 | 0 | 1 |
| Credit grade AA (dummy) | 0.123 | 0.329 | 0 | 1 |
| Credit grade A (dummy) | 0.097 | 0.296 | 0 | 1 |
| Credit grade B (dummy) | 0.116 | 0.320 | 0 | 1 |
| Credit grade C (dummy) | 0.189 | 0.391 | 0 | 1 |
| Credit grade D (dummy) | 0.197 | 0.398 | 0 | 1 |
| Bankcard utilization rate | 0.545 | 0.400 | 0 | 5.95 |

## 3. Data set and empirical results

### 3.1. Data

The data set is generated from Prosper, a US online lending market with over 2,210,000 members and 907 million dollars in personal loans funded so far since its inception in February 2006. The lending process is through an online auction procedure. Anyone with a US social security number can become a borrower or lender on the website. Borrowers create loan listings with a maximum amount of $25,000 and set maximum interest rates they are willing to pay a lender. Prosper then provides the lenders with credit information such as credit scores, credit histories, debt levels, income, employment status, etc. Then the auction begins as lenders can bid down the interest rate. Lenders bid in increments of $50 to $25,000 with the minimum interest rates they are willing to receive on loan listings they select. If enough bids are made such that the amount requested is fully funded before the listing expires, a loan is generated at the lowest rates that clear the market. Prosper then transfers the money to the borrowers and they have the obligation to repay the loan in 36 monthly payments. Prosper handles all on-going loan administration tasks including loan repayment and collections on behalf of the matched borrower and lenders. Once a loan is generated, Prosper reports it to the credit bureau. As a result, the delinquency and default on such a loan can affect the credit score of the borrower. If a payment is late, Prosper charges the borrower a late fee. If delinquency lasts for more than 4 months, the loan is considered to be in default.[2] Prosper then sells the loan to a collection agency through an auction and the lenders get the proceeds of the sale. Prosper generates revenue by collecting a one-time 1%–3% fee on funded loans from borrowers, and assessing a 1% per annual loan servicing fee to lenders.

Our sample shown in Table 1 consists of 2377 loan listings that have been fully funded and become loans in the years 2007 and 2008. All the loans have already been in the payment process for at least 8 months up to 10 months. There are 186 defaults in the sample, constituting an average of 7.82% of default rate. The average borrower's maximum interest rate is 19.82%, which is considered very high compared to bank rates. Borrowers require an average amount of $7200 in loans. About 24.19% of borrowers take the option of closing the listings until the amounts requested are fully funded. Among all borrowers in the sample, about 12.33% of them have credit grade AA with credit scores greater than 760. A further 9.72% have credit grade A with credit scores between 720 and 759. Another 11.61% have credit grade B with credit scores between 680 and 719. Then, 18.85% have credit grade C with credit scores between 640 and 679. Next, 19.73% have credit grade D with credit score between 600 and 639. The remaining 27.76% of the borrowers have credit grade E and HR (high risk) with scores between 520 and 599. On average, a borrower has more than half of available revolving credit that one could use at the time the listing was created.

1 See Klein and Spady (1993) assumptions (C.8a) and (C.8b).

2 We also use this criteria to construct the default variable in our analysis.