

NONPARAMETRIC IDENTIFICATION AND ESTIMATION  
OF A BINARY CHOICE MODEL OF LOAN APPROVAL  
USING ONLY APPROVED LOANS

Wei Jiang

Columbia Business School, Finance and Economics Division  
wj2006@columbia.edu

Ashlyn Aiko Nelson,

Indiana University School of Public and Environmental Affairs  
ashlyn@indiana.edu

Edward J. Vytlačil

Department of Economics, Yale University  
edward.vytlacil@yale.edu

March 19, 2010

## Abstract

We consider the identification and estimation of latent-index, threshold crossing models of loan approval when the econometrician observes all information observed by the decision maker for the approved loans, but observes no information on the loans that were not approved. For example, in our application to mortgage lending, we have all information collected by the bank at the time of loan origination for all loans originated by the bank, so that the bank's decision to approve the loan is a deterministic function of covariates observed by the econometrician, while we have no information on loans that were not approved by the bank. Under smoothness conditions, we use the boundary of the support the covariates in the approved sample and whether the density of the covariates at this boundary is strictly positive to infer a level set for the approval process. We then use that level set to identify and estimate the parameters of the loan approval process. We consider identification and estimation of both nonparametric and semiparametric models of loan approval. We apply our results to study the mortgage loan approval process for a major national bank, using a data set comprised of all information collected by the bank at the time of loan origination on all 721,767 loans funded by the bank between January 2004 and February 2008.

JEL Codes: C24, C25

KEYWORDS: binary choice model, nonparametric identification, choice based sampling