

ECON 899b - Problem Set 1

Due: November 8

Consider the following random utility model for loan repayment:

$$Y_i = 1 \text{ If } \beta_0 + X_i\beta + \epsilon_{i1} > \epsilon_{i0} \quad (1)$$

where X_i includes a vector borrower characteristics, and $\epsilon_{ij} \sim \text{T1EV}(0, 1)$.

The data for this problem is a sample from the National Survey of Mortgage Originations Public Use File. A version of this data-set is available in STATA format: `Mortgage_performance_data.dta`. The data-set includes 16,401 individual residential 30-year mortgages originated between 2013 and 2017. The characteristics (X_i) include:

```
i_large_loan i_medium_loan rate_spread i_refinance age_r cltv dti  
cu first_mort_r score_0 score_1 i_FHA i_open_year2-i_open_year5
```

The main outcome variable is an indicator variable equal to one if the loan is pre-paid within the first year: $Y_i = \text{i.close.first.year}$. The Stata file contains the variable labels for most variables. The do-file “PS1.do” describes the main variables, and estimate a logit model of the probability of repayment.

1. Using your favorite software (not Stata or R), write a routine that evaluates the following three functions: log-likelihood (conditional on β), score of the log-likelihood function, and Hessian. Evaluate the three functions at $\beta_0 = -1$ and $\beta = 0$.
2. Compare the score and hessian obtained from (1) with the numerical first and second derivative of the log-likelihood.
3. Write a routine that solves the maximum likelihood problem using a Newton algorithm.
4. Compare the solution and numerical speed with two optimization packages provided by your favorite software: BFGS and Simplex.