Machine Learning in Finance

Overview

Overview

1. Classification and Discrete Choice
2. Classical Approaches, Part I
3. Classical Approaches, Part II
4. Neural Net Extensions, Part I
5. Neural Net Extensions, Part II
6. Credit Card Default in Germany, Part I
7. Credit Card Default in Germany, Part II
8. Credit Card Default in Germany, Part III
9. Banking Intervention in Texas, Part I
10. Banking Intervention in Texas, Part II
11. Banking Intervention in Texas, Part III

FORDHAM
THE JESUIT UNIVERSITY OF NEW YORK

Gabelli School of Business

Classification and Discrete Choice

Classification and Discrete Choice

- Basically we have a set of characteristics on a client or customer
- We wish to predict from these characteristics if the client falls into a high or low risk category. Let's start with a binary dependent variable.
- We have past data on clients who have been both low risk and high risk (default) clients and we have their characteristics, eg. age, income, homeowner, car owner, professional status, education, etc.
- We wish to know how we can use these data sets to make classify new clients as high or low risk clients, based on their characteristics. We are interested in prediction, we are not interested in the significance of particular coefficients.

FORDHAM
THE JESUIT UNIVERSITY OF NEW YORK

Gabelli School of Business

Classical Approaches, Part I

Classical Approaches, Part I

- An early approach is **Discriminant Analysis**, pioneered by Edward Altman at New York University
- This approach takes a set of k-dimensional characteristics from observed data falling into two groups, for example, a group which paid its loans on schedule and another which became arrears in loan payments.
- We first define the matrices X_1, X_2, where the rows of each X_i represents a series of k-different characteristics of the members of each group, such as a "low risk" or a ":high risk" group. The relevant characteristics may be age, income, marital status, and years in current employment.
- It is quick and easy to use.

Classical Approaches, Part I

- We calculate the same means of the two groups: X_1, X_2, as well as the variance-covariance matrices, Σ_1, Σ_2
- Compute the **pooled** variance, $\Sigma = (((n_1 - 1)/(n_1 + n_2 - 2)))\Sigma_1 + (((n_2 - 1)/(n_1 + n_2 - 2)))\Sigma_2$, where n_1, n_2 represent the population sizes in groups 1 and 2;
- Estimate the coefficient vector, $\beta = \Sigma^{-1}[X_1 - X_2]$
- With the vector β, one simply examines the characteristics of a new set, for classification in either the "low risk" or "high risk" sets, X_1, X_2.
- Defining the set net set of characteristics,x_i, we simply calculate the value: βx_i.
- If this value is closer to βX_1 than to βX_2, then we classify x_i as belonging to the low-risk group X_1. Otherwise, X_2

Classical Approaches, Part I

- Discriminant is a simple linear method
- It does not take into account any assumptions about the distribution of the dependent variable used in the classification.
- We often use binary variables as characteristics. When the binary variable is a dependent variable, it no longer has a normal distribution.
- Discriminant analysis classifies a set of characteristics X.as belonging to group 1 or 2 simply by a **distance** measure.
- For this reason it has been replaced by the more commonly used logistic (also called logit) or probit regression.

FORDHAM
THE JESUIT UNIVERSITY OF NEW YORK

Gabelli School of Business

Classical Approaches, Part II

Classical Approaches, Part II

Logit analysis assumes the relation between probability p_i of the binary dependent variable y_i, and the k explanatory variables x:

$$p_i = (1/(1 + e^{-[x_i\beta+\beta_0]}))$$

- To estimate the parameters β and β_0, we simply maximize the following likelihood (or log-likelihood) function Λ with respect to the parameter vector β for the n observations. The symbol $\prod_{i=1}^{n}$ is the product operator.

$$\max_{\{\beta\}} \Lambda = \prod_{i=1}^{n} (p_i)^{y_i} (1 - p_i)^{1-y_i}$$

$$s.t : p_i = (1/(1 + e^{-[x_i\beta+\beta_0]}))$$

- Log-likelihood objective function

$$\boldsymbol{ln(\Lambda) = y_i ln(p_i) + (1 - y_i) ln(1 - p_i)}$$

Classical Approaches, Part II

- The usual way to evaluate this logistic model is to examine the percentage of correct predictions, both "true" and "false", set at 1 and 0, on the basis of the expected value.
- Setting the estimated p_i at 0 or 1 depends of course on the choice of an appropriate "threshold value". If the estimated probability or expected value p_i.is greater than .5, then p_i is rounded to 1, and expected to take place. Otherwise, it is not expected to occur.
- Of course, the cut-off change change. For example if p_i is the risk of a serious illness, a value of $p_i \geq .1$ would be a reasonable cutoff.

Classical Approaches, Part II

- Probit models are also used: these models simply use the cumulative Gaussian normal distribution rather than the logistic function for calculating the probability of being in one category or not:

$$p_i = \Phi(x_i\beta + \beta_0) = \int_{-\infty}^{x_i\beta+\beta_0} \varphi(i)di$$

- where the symbol Φ is simply the cumulative standard distribution, while the lower case symbol, φ,represents the standard normal density function. We maximize the same log-likelihood function.
- The logistic distribution is similar to the normal or probit one, except in the tails.
- However, it is difficult to justify the choice of one distribution or another on theoretical grounds
- For most cases, it seems not to make much difference

Classical Approaches, Part II

- The Weibull distribution is an asymmetric distribution, strongly negatively skewed.
- It approaches zero only slowly, and goes to one, much more rapidly than the probit and logit models:

$$p_i = 1 - exp(-exp(x_i\beta + \beta_0))$$

- The Weibull regression is used in **survival analysis** and comes from **extreme value theory.**

FORDHAM
THE JESUIT UNIVERSITY OF NEW YORK

Gabelli School of Business

Neural Net Extensions, Part I

Neural Net Extensions, Part I

- Neural network regression for binary choice
- Predicting probability p_i, for a network with k^* input characteristics and j^*neurons:

$$n_{j,i} = \omega_{j,0} + \sum_{k=1}^{k*} \omega_{j,k} x_{k,i}$$

$$N_{j,i} = (1/(1 + e^{-n_{j,i}}))$$

$$p_i = \sum_{j=1}^{j^*} \gamma_j N_{j,i}$$

$$\sum_{j=1}^{j^*} \gamma_j = 1, \gamma_j \geq 0$$

FORDHAM
THE JESUIT UNIVERSITY OF NEW YORK

Gabelli School of Business

Neural Net Extensions, Part II

Neural Net Extensions, Part II

- It is straightforward to extend the logit and neural network models to the case of multiple discrete choices, or classification into three or more outcomes.
- For example, a credit officer may wish to classify potential customers into "safe", "low risk", and "high risk" categories, based on a net of characteristics, x_k.
- One direct approach for such a classification is a "nested" classification. One can use the logistic or neural network model to separate the "normal" categories from the absolute "default" or high risk categories, with a first-stage estimation.
- Then, with the remaining "normal" data, one can separate the categories into a "low risk" and "higher risk" categories. However, there are many cases in financial decision making where there are multiple categories.

Neural Net Extensions, Part II

- Thus, one might wish to neural network classification to predict which type of category a particular firm's bond may fall into, given the characteristics of the particular firm, from observable market data and current market classifications or bond ratings.
- In this case, using the example of three outcomes, we use the **softmax** function to compute p_1, p_2, p_3 for each observation i:

$$P_{1,i} = (1/(1 + e^{-[x_i\beta_1 + \beta_{10}]}))$$

$$P_{2,i} = (1/(1 + e^{-[x_i\beta_2 + \beta_{20}]}))$$

$$P_{3i} = (1/(1 + e^{-[x_i\beta_3 + \beta_{30}]}))$$

Neural Net Extensions, Part II

- The probabilities of falling in category 1, 2 and 3 come from weighting each probability by the cumulative probabilities:

$$p_{1,i} = ((P_{1,i})/(\sum_{j=1}^{3} P_{j,i}))$$

$$p_{2,i} = ((P_{2,i})/(\sum_{j=1}^{3} P_{j,i}))$$

$$p_{3,i} = ((P_{3})/(\sum_{j=1}^{3} P_{j,i}))$$

Neural Net Extensions, Part II

- The parameters of both the logistic and neural network models are estimated by maximizing a similar likelihood function:

$$\Lambda = \prod_{i=0}^{i=i^*} (p_{1,i})^{y1,i} (p_{2,i})^{y_{2,i}} (p_{3,i})^{y_{3,i}}$$

- The success of these alternative models is readily tabulated by the percentage of correct predictions for particular categories.

FORDHAM
THE JESUIT UNIVERSITY OF NEW YORK

Gabelli School of Business

Credit Card Default in Germany, Part I

Credit Card Default in Germany, Part I

- We list 20 arguments, a mix of "categorical" and continuous variables.
- We show the maximum minimum, and median values of each of the variables.
- y it takes on a value of 0 if there is no default and a value of 1 if there is a default.
- The are 300 cases of defaults in this sample of 1000, with y=1 for default, and y = 0 for no default.
- As we can see in the mix of variables, there is considerable discretion about how to categorize the information.

Credit Card Default in Germany, Part I

Attributes for German Credit Data Set

Variable	*Definition*	*Type/Explanation*	Max	Min	Median
1	Checking account	Categorical, 0 to 3	3	0	1
2	Term	Continuous	72	4	18
3	Credit history	Categorical, 0 to 4, from no history to delays	4	0	2
4	Purpose	Categorical, 0 to 9, based on type of purchase	10	0	2
5	Credit amount	Continuous	18424	250	2319.5
6	Savings account	Categorical, 0 to 4, lower to higher to unknown	4	0	1
7	Yrs in present employment	Categorical, 0 to 4, 1 unemployment, to longer years	4	0	2
8	Installment rate	Continuous	4	1	3
9	Personal status and gender	Categorical, 0 to 5, 1 male, divorced, 5 female, single	3	0	2
10	Other parties	Categorical, 0 to 2, 1 none, 2 co-applicant, 3 guarantor	2	0	0
11	Yrs in present residence	Continuous	4	1	3
12	Property type	Categorical, 0 to 3, 0 real estate, 3 no property or unknown	3	0	2
13	Age	Continuous	75	19	33
14	Other installment plans	Categorical, 0 to 2, 0 bank, 1 stores, 2 none	2	0	0
15	Housing status	Categorical: 0 to 2: 0 rent, 1 own, 2 for free	2	0	2
16	Number of exiting credits	Continuous	4	1	1
17	Job status	Categorical, 0 to 3, 0 unemployed, 3 management	3	0	2
18	Numer of dependents	Continuous	2	1	1
19	Telephone	Categorical, 0 to 1, 0 none, 1, yes, under customer name	1	0	0
20	Foreign worker	Categroical, 0 to 1, 0 yes, 1 no	1	0	0

FORDHAM
THE JESUIT UNIVERSITY OF NEW YORK

Gabelli School of Business

Credit Card Default in Germany, Part II

Credit Card Default in Germany, Part II

- We first examine **in-sample** statistics.
- We show the likelihood functions for the four nonlinear alternatives to the discriminant analysis (Logit, Probit, Weibull, and Neural Net)
- We show the error percentages of all five methods.
- There are two types of errors, as taught from statistical decision theory.

Credit Card Default in Germany, Part II

- False positives take place when we incorrectly label the dependent variables as one, with y=1, when in truth y=0.
- False negatives occur when we have y=0, when the true y=1.
- The overall "error ratio" in the table is simply a weighted average of the two error percentages, with the weight set at .5.
- The neural network alternative to the logit, probit, and Weibull methods is a network with three neurons.

Credit Card Default in Germany, Part II

- In-sample performance:

Method	*Likelihood Fn.*	ERROR PERCENTAGES *FALSE POSITIVES*	*FALSE NEGATIVES*	*WEIGHTED AVERAGE*
Discriminant Analysis	na	0.207	0.091	0.149
Neural Network	519.8657	0.062	0.197	0.1295
Logit	519.8657	0.062	0.197	0.1295
Probit	519.1029	0.062	0.199	0.1305
Weibull	516.507	0.072	0.189	0.1305

Credit Card Default in Germany, Part II

- We see a familiar trade-off.
- Discriminant analysis has lower false negatives
- But a much much higher percentage (by more than a factor of three) of false positives.

FORDHAM
THE JESUIT UNIVERSITY OF NEW YORK

Gabelli School of Business

Credit Card Default in Germany, Part III

Credit Card Default in Germany, Part III

- For out-of-sample performance: we used the "0.632 Bootstrap Method" .
- To summarize this method, we simply took 1000 random draws of data from the original sample, with replacement, to do an estimation.
- We used the "excluded data" from the original sample to evaluate the out-of-sample forecast performance.
- We measured the out-of-sample forecast performance by the error percentages of false positives or false negatives.
- We repeated this process 100 times and examined the mean and distribution of the error-percentages of the alternative models.

Credit Card Default in Germany, Part III

	OUT-OF-SAMPLE FORECASTING: 100 DRAWS MEAN ERROR PERCENTAGES (0.632 Bootstrap)		
Method	*FALSE POSITIVES*	*FALSE NEGATIVES*	*WEIGHTED AVERAGE*
Discriminant Analysis	0.000	0.763	0.382
Neural Network	0.095	0.196	0.146
Logit	0.095	0.196	0.146
Probit	0.702	0.003	0.352
Weibull	0.708	0.000	0.354

Credit Card Default in Germany, Part III

- We see that the neural network and logit models given identical performance, in terms of out-of-sample accuracy.
- We also see that discriminant analysis and the probit/Weibull methods are almost mirror images of each other.
- Whereas discriminant analysis is perfectly accurate in terms of false positives, it is extremely imprecise (with an error rate of more than 75 percent) in terms of false negatives.
- The probit and Weibull are quite accurate in terms of false negatives, but highly imprecise in terms of false positives.
- The better choice would be to use logit or the neural network method.

FORDHAM
THE JESUIT UNIVERSITY OF NEW YORK

Gabelli School of Business

Banking Intervention in Texas, Part I

Banking Intervention in Texas, Part I

- Banking intervention, the need to close or to put a private bank under state management, more extensive supervision, or to impose a change of management, is, unfortunately, common enough.
- We use the same binary or classification methods to examine how well key "characteristics" of banks may serve as "early warning signals" for a "crisis" or intervention of a particular bank.
- The data were obtained from the Federal Reserve Bank of Dallas using banking records from the last two decades prior to 2001.
- The total percentage of banks which required intervention, either by state or federal authorities, was 16.7.
- We use twelve variables as arguments. The capital-asset ratio, of course, is the key component of the Basel accord for international banking standards.

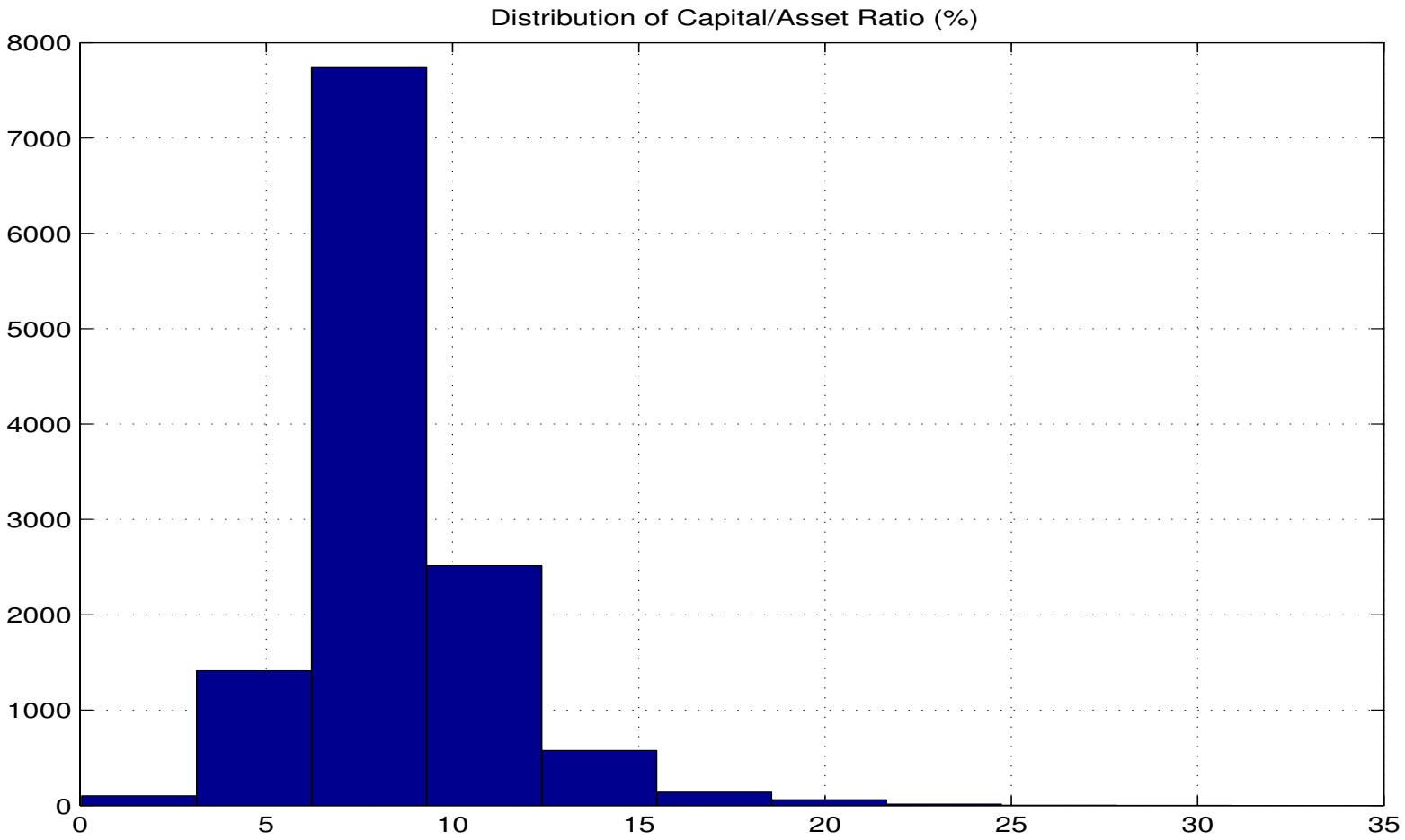
Banking Intervention in Texas, Part I

TEXAS BANKING DATA

		Max	Min	Median
1	Charter	1	0	0
2	Federal Reserve	1	0	1
3	Capital/asset %	30.9	-77.71	7.89
4	Agricultural loan/total loan ratio	0.822371	0	0.013794
5	Consumer loan/total loan ratio	0.982775	0	0.173709
6	Credit card loan/total loan ratio	0.322974	0	0
7	Installment loan/total loan ratio	0.903586	0	0.123526
8	Nonperforming loan/total loan - %	35.99	0	1.91
9	Return on assets - %	10.06	-36.05	0.97
10	Interest margin - %	10.53	-2.27	3.73
11	Liquid assets/total assets - %	96.54	3.55	52.35
12	US total loans/US gdp ratio	2.21	0.99	1.27

Dependent Variables: Bank closing or intervention
No observations: 12605
% of Interventions/closings: 16.7

Banking Intervention in Texas, Part I



Banking Intervention in Texas, Part I

- While the negative number for the minimum of the capital/asset ratio may seem surprising, the data set includes both "sound" and "unsound" banks.
- When we remove the observations having negative capital/asset ratios, the distribution of this variable shows that the ratio is between five and ten percent for most of the banks in the sample.
- Some of the variables come from CAMEL ratings of banks (Captal, Asset quality, Management, Earnings, and Liquidity)

FORDHAM
THE JESUIT UNIVERSITY OF NEW YORK

Gabelli School of Business

Banking Intervention in Texas, Part II

Banking Intervention in Texas, Part II

- In-sample criteria:

Method	*Likelihood Fn.*	ERROR PERCENTAGES *FALSE POSITIVES*	*FALSE NEGATIVES*	*WEIGHTED AVERAGE*
Discriminant Analysis	na	0.205	0.038	0.122
Neural Network	65535	0.032	0.117	0.075
Logit	65535	0.092	0.092	0.092
Probit	4041.349	0.026	0.122	0.074
Weibull	65535	0.040	0.111	0.075

Banking Intervention in Texas, Part II

- Out-of-sample criteria with .636 Bootstrapping

OUT-OF-SAMPLE FORECASTING: 40 DRAWS
MEAN ERROR PERCENTAGES (0.632 Bootstrap)

Method	*FALSE POSITIVES*	*FALSE NEGATIVES*	*WEIGHTED AVERAGE*
Discriminant Analysis	0.000	0.802	0.401
Neural Network	0.035	0.111	0.073
Logit	0.035	0.089	0.107
Probit	0.829	0.000	0.415
Weibull	0.638	0.041	0.340

Banking Intervention in Texas, Part II

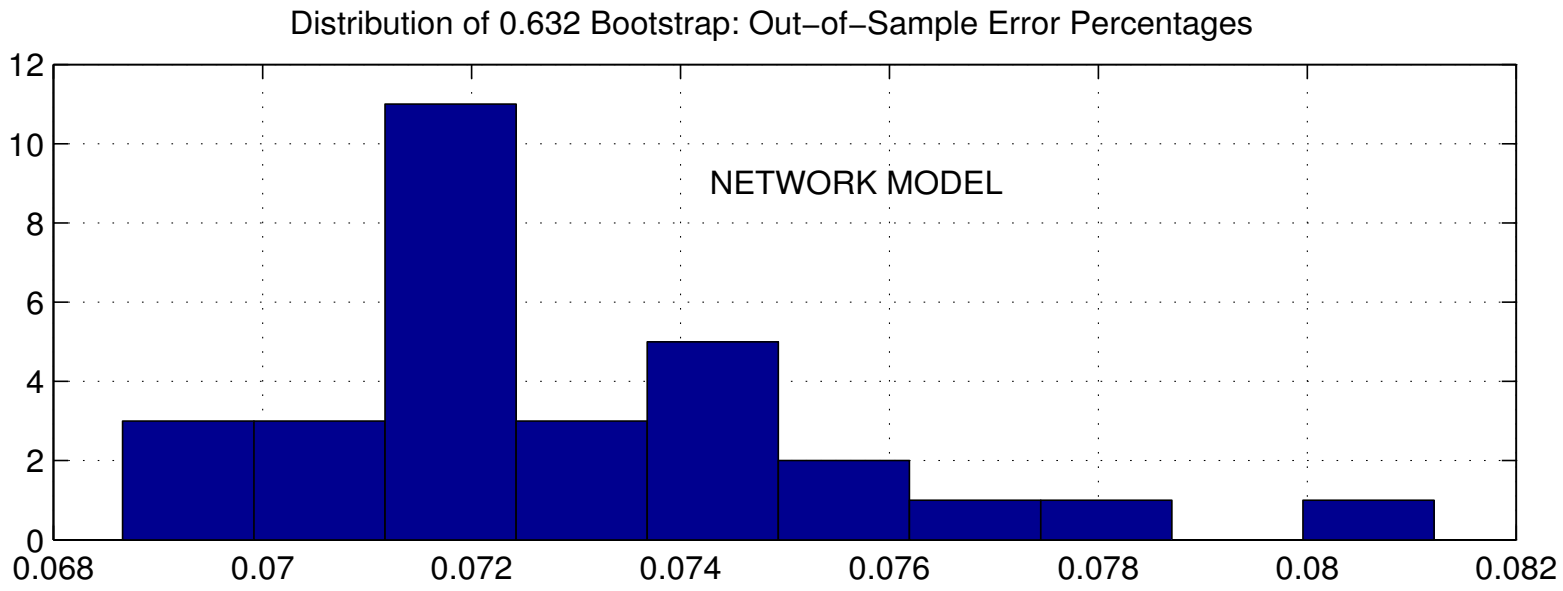
- We see that discriminant analysis has a perfect score, zero percent, on false positives, but has a score of over 80 percent, on false negatives.
- The overall best performance in this experiment is by the neural network, with a 7.3 percent "weighted average" error-percent score.
- The logit model is next, with a 10 percent weighted average score.
- The neural network family outperforms the other methods in terms of out-of-sample accuracy.

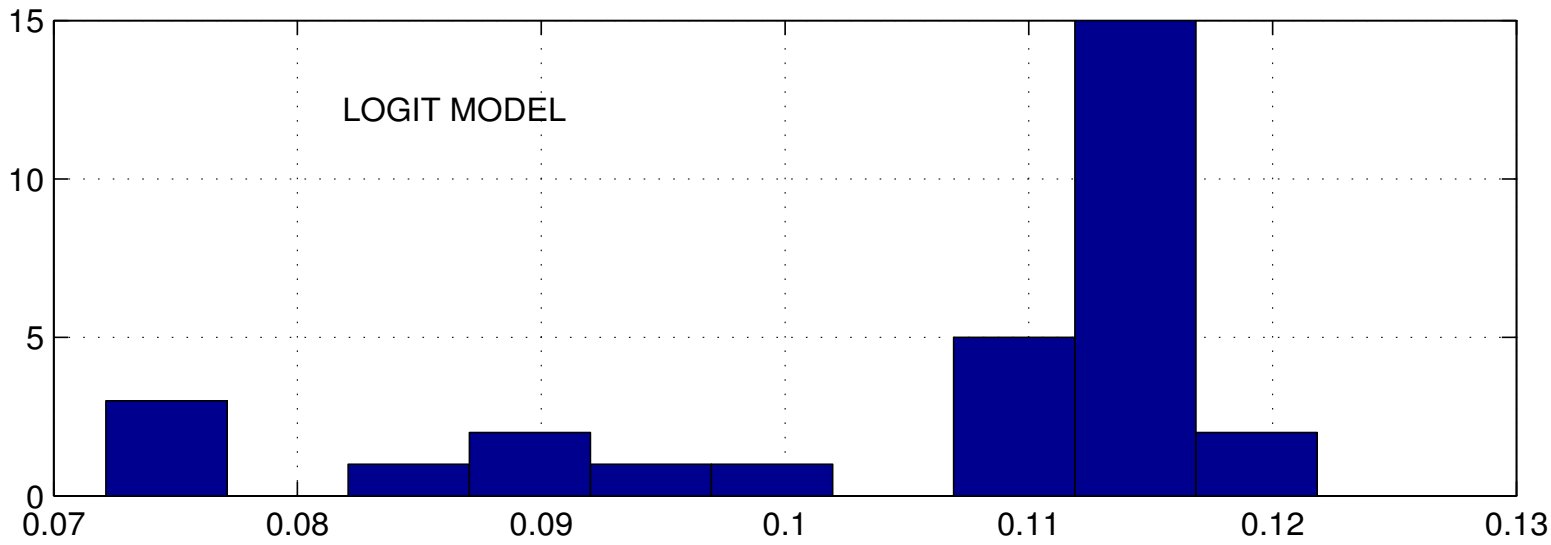
FORDHAM
THE JESUIT UNIVERSITY OF NEW YORK

Gabelli School of Business

Banking Intervention in Texas, Part III

Banking Intervention in Texas, Part III





FORDHAM
THE JESUIT UNIVERSITY OF NEW YORK

Gabelli School of Business