



Comptroller of the Currency
Administrator of National Banks

US Department of the Treasury

OCC Fair Lending Conference 2008

Statistical Analysis and Modeling for Risk Assessment

September 10, 2008

The opinions expressed in this paper reflect those of the author only and do not necessarily reflect those of the Office of the Comptroller of the Currency or the Treasury Department.

ENSURING A SAFE AND SOUND
NATIONAL BANKING SYSTEM
FOR ALL AMERICANS

OCC Fair Lending Conference 2008

Fair Lending Modeling of Pricing Decisions

Mortgage Pricing, Fair Lending and Statistics

Jason Dietrich

Economist

Risk Analysis Division

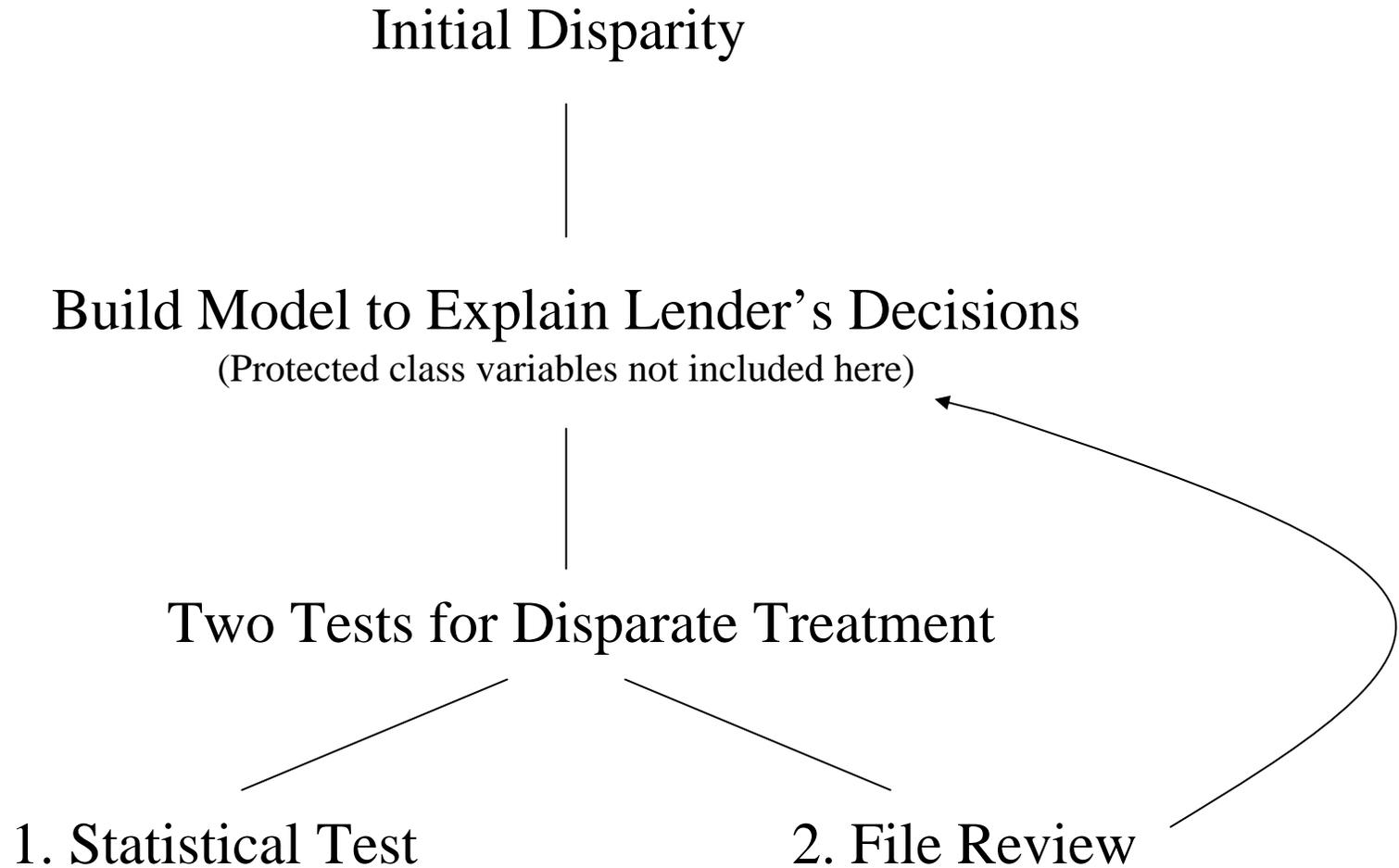
OCC



Comptroller of the Currency
Administrator of National Banks

US Department of the Treasury

ENSURING A SAFE AND SOUND
NATIONAL BANKING SYSTEM
FOR ALL AMERICANS





- Average APR
 - Blacks: 8.0
 - Whites: 6.5
 - Disparity = $8.0 - 6.5 = 1.5$
- Meaningful?
 - May be legitimate reasons why average APRs differ
- Manual file review can be difficult
 - Geography
 - Timing
 - Product characteristics
 - Applicant characteristics
- Modeling makes all originations comparable



Example: Just 2 originations

- Suppose we have just two originations:

	<u>Credit Score</u>	<u>LTV</u>	<u>Loan Amount</u>	<u>APR</u>
Applicant 1	640	95	\$75k	8.0
Applicant 2	720	80	\$125k	6.5

- Use Model to predict APR given risk profile
- Objective: Make originations comparable



Example: Statistical Model with 3 Factors

<u>Factor</u>	<u>Weight</u>
Constant	5.38
Credit Score	
missing	4.35
300-600	3.05
601-660	1.78
661-720	0.64
721 +	0.00
LTV	
0-79	0.00
80-90	0.46
91-95	0.72
96 +	2.15
Loan amount	
0-100	0.37
100+	0.00

The Black Box:

Where do the weights come from?

Similar to building a credit score

OLS

Logit



Does Model Make Sense?

<u>Factor</u>	<u>Weight</u>
Constant	5.38
Credit Score	
missing	4.35
300-600	3.05
601-660	1.78
661-720	0.64
721 +	0.00
LTV	
0-79	0.00
80-90	0.46
91-95	0.72
96 +	2.15
Loan amount	
0-100	0.37
100+	0.00

Lender's Policy

Lower credit score means higher APR

Higher LTV means higher APR

Lower loan amount means higher APR



Score the Two Applicants

Factor	Weight	Applicant 1 Actual Data	Applicant 2 Actual Data
Constant	5.38	--	--
Credit Score			
missing	4.35	0	0
300-600	3.05	0	0
601-660	1.78	1	0
661-720	0.64	0	1
721 +	0.00	0	0
LTV			
0-79	0.00	0	0
80-90	0.46	0	1
91-95	0.72	1	0
95 +	2.15	0	0
Loan amount			
0-100	0.37	1	0
100+	0.00	0	1
actual APR		8.00	6.50



Score the Two Applicants

Factor	Weight	Applicant 1 Actual Data	Applicant 2 Actual Data
Constant	5.38	--	5.38
Credit Score			
missing	4.35	0	0
300-600	3.05	0	0
601-660	1.78	1	1.78
661-720	0.64	0	1
721 +	0.00	0	0
LTV			
0-79	0.00	0	0
80-90	0.46	0	1
91-95	0.72	1	0.72
95 +	2.15	0	0
Loan amount			
0-100	0.37	1	0.37
100+	0.00	0	1
actual APR		8.00	6.50



Score the Two Applicants

Factor	Weight	Applicant 1 Actual Data	Applicant 2 Actual Data
Constant	5.38	--	5.38
Credit Score			
missing	4.35	0	0
300-600	3.05	0	0
601-660	1.78	1	1.78
661-720	0.64	0	1
721 +	0.00	0	0
LTV			
0-79	0.00	0	0
80-90	0.46	0	1
91-95	0.72	1	0.72
95 +	2.15	0	0
Loan amount			
0-100	0.37	1	0.37
100+	0.00	0	1
predicted APR		8.25	
actual APR		8.00	6.50
over/under paid		-0.25	



Score the Two Applicants

Factor	Weight	Applicant 1 Actual Data	Applicant 2 Actual Data
Constant	5.38	--	5.38
Credit Score			
missing	4.35	0	0
300-600	3.05	0	0
601-660	1.78	1	1.78
661-720	0.64	0	1
721 +	0.00	0	0
LTV			
0-79	0.00	0	0
80-90	0.46	0	1
91-95	0.72	1	0
95 +	2.15	0	0
Loan amount			
0-100	0.37	1	0
100+	0.00	0	1
predicted APR		8.25	6.48
actual APR		8.00	6.50
over/under paid		-0.25	0.02



- Use Model to Calculate Amount Over/Under Paid for All Originations
- Average Over/Under Paid
 - Blacks = 0.50
 - Whites = -0.15
- Disparity in Average Over/Under Paid = 0.65
 - Raw disparity was 1.50



Actual Statistical Test for Disparate Treatment

<u>Factor</u>	<u>Weight</u>
Constant	5.35
Credit Score	
missing	4.15
300-600	3.09
601-660	2.24
661-720	0.91
721 +	0.00
LTV	
0-79	0.00
80-90	0.52
91-95	0.71
96 +	2.11
Loan amount	
0-100	0.50
100+	0.00
Race	
Black	0.65
White	0.00



1. Statistically significant?

- Yes

2. Robust?

- Model specification: Should form of factors differ?
- Omitted variables: Should other factors be included?
- File Review: Review outliers to test robustness of statistical results
 - Review matched pairs as test of disparate treatment



Model Mis-Specification Example

Model 1		Model 2	
<u>Factor</u>	<u>Weight</u>	<u>Factor</u>	<u>Weight</u>
Constant	5.35	Constant	5.36
Credit Score		Credit Score	
missing	4.15	missing	4.10
300-600	3.09	300-600	3.15
601-660	2.24	601-660	2.20
661-720	0.91	661-720	1.01
721 +	0.00	721 +	0.00
LTV		LTV	
0-79	0.00	0-79	0.00
80-90	0.52	80-90	0.52
91-95	0.71	91-95	0.73
96 +	2.11	96 +	2.06
Loan amount		Loan Amount	
0-100	0.50	0-50	0.95
100+	0.00	50-100	0.38
		100-150	0.11
		150+	0.00
Race		Race	
Black	0.65	Black	0.36
White	0.00	White	0.00



Omitted Variable Example

Model 1

<u>Factor</u>	<u>Weight</u>
Constant	6.50

Model 2

<u>Factor</u>	<u>Weight</u>
Constant	5.35
Credit Score	
missing	4.15
300-600	3.09
601-660	2.24
661-720	0.91
721 +	0.00

Model 3

<u>Factor</u>	<u>Weight</u>
Constant	5.38
Credit Score	
missing	4.06
300-600	3.55
601-660	2.21
661-720	1.11
721 +	0.00

LTV

0-79	0.00
80-90	0.52
91-95	0.71
96 +	2.11

LTV

0-79	0.00
80-90	0.49
91-95	0.82
96 +	2.01

Loan amount

0-100	0.50
100+	0.00

Loan Amount

0-100	0.35
100+	0.00

Geography *

Race

Black	1.50
White	0.00

Race

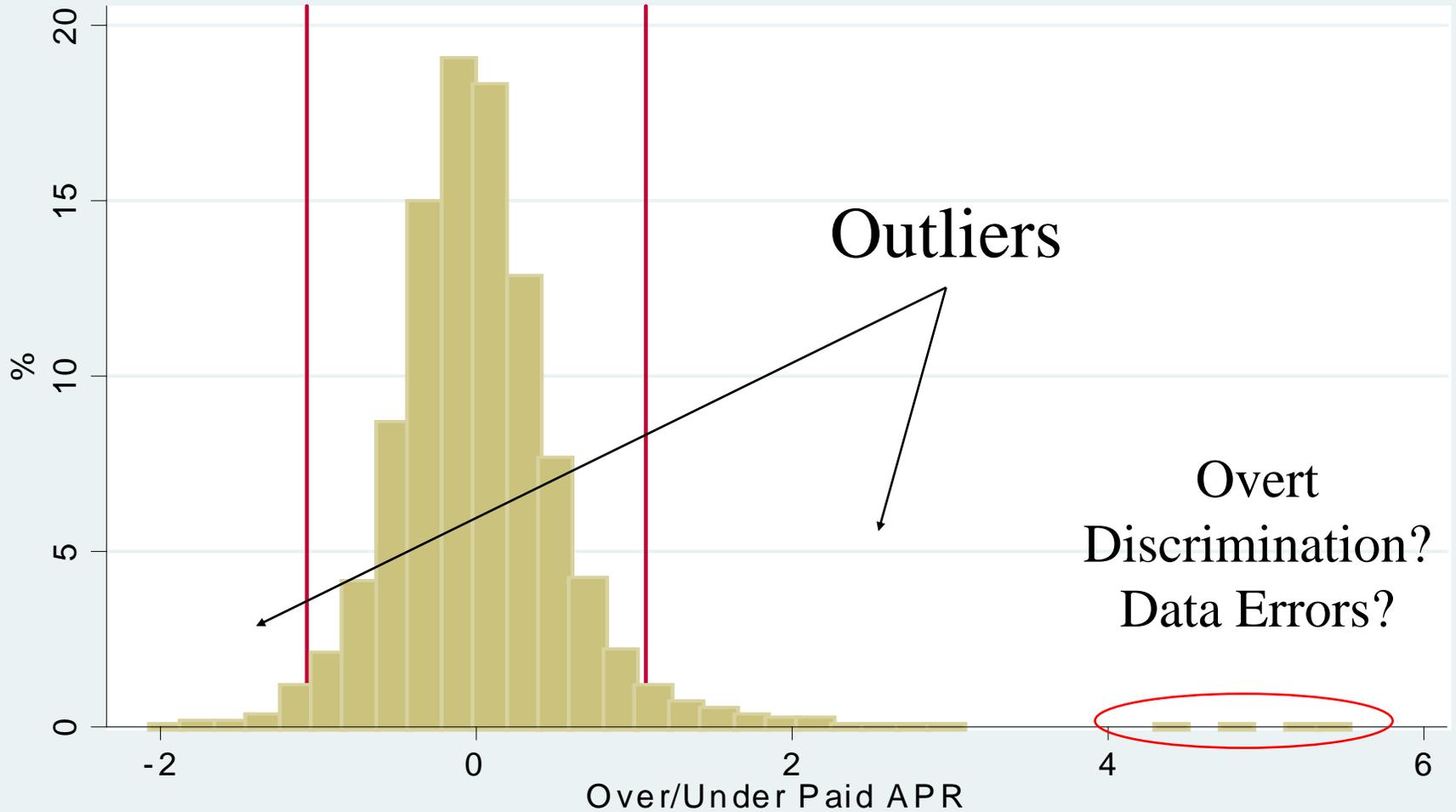
Black	0.65
White	0.00

Race

Black	0.23
White	0.00

Identification of Outliers for Review

Distribution of Over/Under Paid APR





Statistical models may be black boxes, but they should make sense

A statistical model is one component of an overall statistical analysis