Chapter 5

Statistical Analyses Testing for Pricing Discrimination

Jason Dietrich
March 2026

I. Introduction

For nearly 30 years, Regulatory Agencies have used statistical analyses during fair lending supervisory examinations to identify areas of risk and test for discrimination. The typical approach Economists use is to build a statistical model that reflects the lender's formal decision-making process per policy and then test whether prohibited bases factors explain any remaining variation in the outcome variable. Although this is a standard and straightforward application of regression techniques, there has historically been uncertainty within industry, especially among non-quantitative staff, about how regulators conduct and use regression analyses during fair lending exams. In addition, there is long-standing disagreement among stakeholders, both within industry and within Regulatory Agencies, about whether regression estimates can capture causal relationships and provide conclusive proof of discrimination. This report attempts to clarify how regulators use regression analyses during fair lending exams, how regression analysis works (and does not work), and how to interpret and use the results.

To make the discussion more concrete, we focus on one specific question, whether a lender is discriminating when making mortgage pricing decisions. Analyzing mortgage pricing for discrimination is one of the most challenging fair lending statistical analyses regulators conduct, so in addition to providing general analytical details, using this example also provides valuable insight into strategies for analyzing mortgage pricing decisions for discrimination. Within this framework, we rely heavily on simulated policies and data. By using simulated policies and data, we control exactly how the data are generated, so we know exactly what results the regression analysis should generate. This is a very effective approach for understanding how regression analysis works, and how different analytical scenarios and assumptions impact the results. To ease concerns about the generalizability of the results, we use simulated policies and data that generally reflect what mortgage lenders use for real-world applications. In addition, Appendix A and B contain all of the STATA code and statistical results for all analyses so readers can explore alternative assumptions and scenarios to extend their understanding of the concepts covered.

This report covers a broad set of concepts relevant to statistical analyses of discrimination including,

- How mortgage pricing works, strategies for analyzing mortgage pricing for discrimination, and the importance of focusing on pricing exceptions for these analyses
- What it means when regulators say they build statistical models that reflect the lender's formal decision-making processes per policy
- Whether regression analyses can provide causal estimates that provide conclusive proof of discrimination
- How model mis-specification can impact regression results
- How correlations among variables can impact regression results
- How sample size can impact regression results
- How discretion can impact regression results

Throughout the report, we employ a learning-by-doing approach, relying heavily on a variety of interactive, empirical exercises to highlight the underlying intuition of these concepts for end users. We leave it to the reader to independently explore all of the formal econometric theories behind these concepts.

II. Overview of Mortgage Pricing

Modeling pricing disparities for mortgages is particularly challenging since there are multiple components of price (rate, points, fees), as well as tradeoffs between these components. In addition, lenders often give loan officers discretion to apply pricing exceptions to deviate from the policy rate, points, and fees. The details of these pricing components can also vary across several dimensions such as product, program, lender type, and time. All of these complexities create multiple opportunities for lenders to inject discrimination into pricing outcomes, making it difficult for regulators to detect.

For this report we focus on a retail lender who uses risk-based pricing to price standard, conventional, conforming, mortgage loans.[1] Risk-based pricing is typically based on a rate sheet. The primary component of the rate sheet is a rate/point menu, which presents the number of points the lender charges for a variety of interest rates.[2] The expected return to the lender is the same across rate/point combinations, so the lender is indifferent between combinations. Rate sheets also typically include applicant-level, risk-related adjustments to rate and points, called loan level price adjusters (LLPAs). For example, adjustments are often included for FICO scores and LTV values. Many lenders also charge origination points to cover processing costs. Typically, lenders charge 1 origination point. However, some lenders incorporate origination points into tradeoffs with rate and/or points, for example charging higher origination points to make up for a lower rate offer. In addition to rate, discount/rebate points, and origination points, lenders also typically charge several fees to cover third-party vendor costs, such as pulling a credit bureau report and conducting an appraisal. Lenders often simply pass on all third-party vendor charges directly to the customer.

If a lender applies its pricing policies in a completely deterministic way, there is no risk of disparate treatment. There may be disparate impact risk however, as some components of the lender's rate sheets or pricing policies might disproportionately harm certain demographic

[1] Wholesale and Correspondent lending create an additional layer of complexity as the number of decision-makers impacting pricing outcomes increases dramatically. See El-Anshasy et. al. (2004), Calem and Longhofer (1999), and Kleiner and Todd (2007) for more information on analyzing mortgage pricing by brokers.

[2] A point is one percent of the loan amount. Discount points are points an applicant pays the lender to buy down the rate, rebate points are points the lender pays to an applicant to accept a higher rate, and origination points are separate points a lender charges to cover processing fees. For most of this report we ignore origination points and focus on the tradeoffs between rate and discount/rebate points. We use the general term, “points,” to refer to both discount points (denoted by positive values) and rebate points (denoted by negative values).

groups. If pricing is not completely deterministic, then disparate treatment can occur. There are two primary ways for disparate treatment to occur: 1) Loan officers do not apply the stated policies fairly and consistently across all applicants or 2) The lender gives loan officers discretion to deviate from formal pricing policies. For this report, we focus primarily on disparate treatment.

Table 1 provides a simple example rate sheet illustrating the basic components of mortgage pricing. It contains only three rate/point combinations and two LLPAs, one for FICO score and one for property type. Both LLPAs adjust points. The rate/point menu is slightly convex, since applicants typically must pay an increasing number of points for the same additional reduction in rate. Although very simple, this rate sheet reflects most of the general components of a typical rate sheet mortgage lenders use to price mortgages. To keep the example manageable, we ignore all origination points and fees here and throughout the report.

Table 1: Example Rate Sheet

Rate	Points
4.375	-1
4	0
3.875	1

Loan Level Price Adjusters (LLPAs: Points)		
FICO		
	350-660	+1
	661-720	+0.5
Property Type		
	Manufactured Home	+2

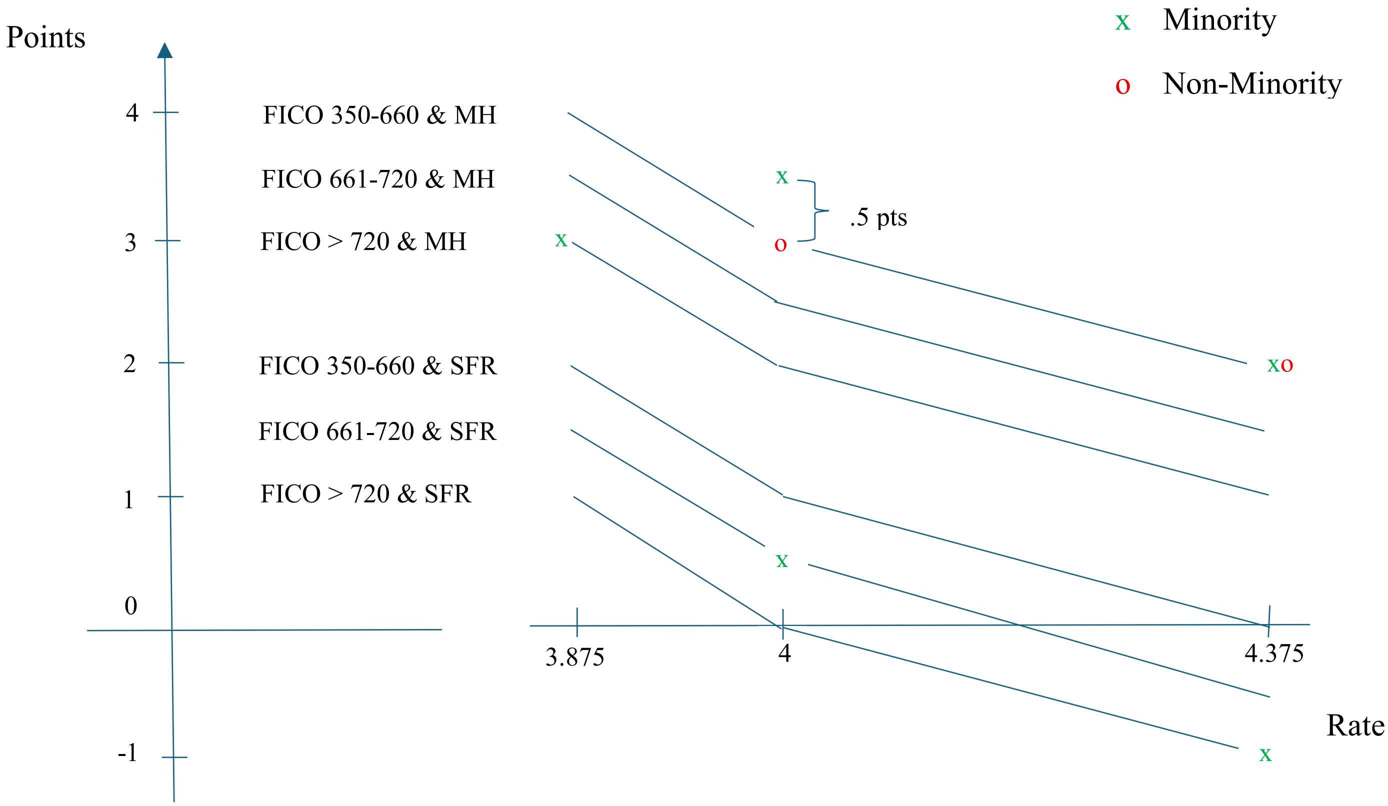
When using a rate sheet to price a mortgage loan, the starting point is the base rate/point menu. Based on our example rate sheet, an applicant for credit can choose between three rate/point combinations. Ignoring the LLPAs for now, they can choose a rate of 4.375% and receive one rebate point, a rate of 4% and pay no discount points and receive no rebate points, or a rate of 3.875% and pay one discount point. Applicants with shorter time horizons, and who need help with closing costs, typically prefer to receive rebate points and are willing to accept a higher rate, and applicants with longer time horizons, and who are not liquidity constrained, typically prefer to pay discount points to buy down the rate.

The LLPAs are credit risk adjustments to the base rate/point menu, which create applicant-specific menus of pricing options based on the specific applicant's credit characteristics. In our example rate sheet, all LLPAs adjust points, although in real-world applications rate sheets can include LLPAs that adjust points, rate, or both. As one example of how to interpret the LLPAs, per the rate sheet in Table 1, an applicant with a FICO score between 350 and 660 has an LLPA of +1. This means that, all-else-equal, she faces the pricing menu, 4.375:0, 4:1, and 3.875:2, instead of the base rate/point menu. In other words, she has to pay an extra point for each of the three rates the lender offers for this product, because she is a higher credit risk.

Graph 1 presents the rate sheet from Table 1 in graphical form to explicitly show the pricing menus applicants with different credit characteristics face. Often, rate is on the vertical axis and points are on the horizontal axis, but they are reversed in Graph 1. For this report, all LLPAs are point adjusters, and we inject discrimination and discretion into pricing only by adjusting points. Under these assumptions, the relevant question for analysis is, how many points does an applicant have to pay for a given rate, conditional on their creditworthiness. With this framework, it is more appropriate to put points on the vertical axis and rate on the horizontal axis, as in Graph 1. When LLPAs adjust rate, or discretion and discrimination impact rate, then rate would typically be on the vertical axis.

Since LLPAs are all positive adjustments and only impact points in our pricing example, all menus are represented by an upward shift from the base rate/point menu (the bottom line in Graph 1). Depending on their FICO score and property type, each applicant will face one

Graph1: Example Pricing Menus



specific menu of rate/price combinations (i.e., one line on Graph 1). For example, all applicants with a FICO score between 350 and 660 who applied for a mortgage secured by a manufactured home would face the pricing menu on the top line of Graph 1. If pricing is deterministic and every applicant receives a rate/point combination on their particular pricing menu per the rate sheet and their credit characteristics, then there are no disparate treatment concerns. Essentially, every applicant received the pricing per the rate sheet. This is represented by all but one of the x and o points in Graph 1, all of which are on their appropriate rate/point menus. Disparate treatment becomes a potential issue when the lender allows discretionary pricing exceptions, or pricing deviates from the formal rate sheet for other reasons. To show this, Graph 1 includes one minority applicant who received a rate/point combination that is 0.5 points above their rate/point menu, because of a pricing exception. For our current example, pricing exceptions only impact points, so the deviation from the rate/point menu is a vertical shift of 0.5 points. If the pricing exception affected rate, then the adjustment due to the exception would be a rightward movement in the graph.

All of the simulation analyses below use the rate sheet in Table 1 as the lender's base set of mortgage pricing policies. In real-world applications, mortgage rate sheets are typically much more complex than the rate sheet in Table 1. However, the general analytical approach to conducting statistical analyses to test for discrimination in mortgage pricing will be similar. Regardless of how complex the rate sheet is, if every applicant's pricing is on their credit-risk-adjusted pricing menu per the rate sheet, then the statistical analysis should focus on whether the rate sheet creates a disparate impact. If the pricing for some applicants deviates from their credit-risk-adjusted pricing menu, then the statistical analysis should also test for disparate treatment. Specifically, are these deviations more likely to occur, and be larger in magnitude, for some demographic groups, and what are the reasons for these deviations?

III. Simulation Analyses

This section details the simulation analyses we conduct in this report. The foundation of the simulation analyses here, and of all statistical analyses in general, is the data generating process (DGP), which is the true, underlying process that generates the data available to the Economist for analysis. One particularly important component of the overall DGP is the true, underlying process that generates the outcome variables in the data. The primary goal of a statistical analysis is to build a regression model that accurately reflects this specific process. For fair lending statistical analyses, the lender's decision-making processes generate the outcome variables available for the analysis, so it is critical for Economists to fully understand the lender's policies and procedures when conducting the statistical analysis. Since a lender's formal policies and procedures will never state that the lender discriminates, the question for Economists is then, is prohibited basis group information actually a component of the DGP, i.e., does the data suggest that the lender considered prohibited basis group information when making credit decisions?

We conduct statistical analyses testing for discrimination using simulated data generated from a variety of DGPs. Knowing the DGP, and most importantly knowing whether discrimination is part of the DGP, is a very effective way to assess the properties and effectiveness of regression analyses under a variety of scenarios, which is the primary goal of this report. The general analytical approach we use consists of three steps: 1) generate simulated data on the credit characteristics of a sample of mortgage applicants, 2) apply a lender's pricing policies to determine what price to charge each applicant conditional on their credit characteristics, and 3) use regression tools to analyze the data for discrimination. We start with a Base DGP based on the rate sheet in Table 1 and then apply several modifications to this Base DGP to assess how these modifications impact the regression results.

Base DGP

The first step of the analytical approach is to simulate data for a sample of mortgage applicants. We begin by simulating data for minority status, FICO score, property type (single-family residence (SFR) vs manufactured housing (MH)), and the preferred rate for 500 applicants using the following specifications,

- Minority Flag
    - = 1 if a draw from a uniform distribution < 0.25; 0 otherwise

- FICO Score
  - Random draw from a N(660, 60) distribution[3]

- Manufactured Housing Flag (MH)
  - = 1 if a draw from a uniform distribution < 0.1; 0 otherwise

- Preferred Rate
  - = 4.375 if a draw from a uniform distribution < 0.15
  - = 4.000 if a draw from a uniform distribution >= 0.15 and <= 0.65
  - = 3.875 if a draw from uniform distribution > 0.65

At this point, we have data on race, FICO score, property type, and preferred rate for 500 applicants. Although simulated, if we were to look at these data, or summary statistics for these data, it would look just like any actual sample of potential applicants for mortgages.

The second step of the analytical approach is to price each loan as if these applicants had applied for a mortgage. For the lender's pricing policies, we use the rate sheet in Table 1, ignore all fees and origination points, and do not allow loan officers any discretion to deviate from the rate sheet. Specifically, given each applicant's FICO score, property type, and preferred rate, the rate sheet determines the points charged. The Base DGP is now complete, and we have a simulated dataset with 500 applicants for analysis.

The third and final step of the analytical approach is to use regression tools to analyze these data and test for discrimination. The standard approach regulators use when conducting a statistical analysis during exams is to build a statistical model that reflects the lender's decision-making process. The starting point for this analysis is therefore understanding the lender's decision-making process. For the Base DGP, this consists of reviewing the rate sheet in Table 1. Next, the dependent variable, which reflects the lender's decision that is of interest to the

[3] N(660, 60) stands for a normal distribution with a mean of 660 and a standard deviation of 60.

Economist, must be specifically defined. For analyses of mortgage pricing, Economists analyze a variety of dependent variables, including rate, points, fees, rate spread, APR, exceptions, and others. Given that all LLPAs in our rate sheet affect points, and the discrimination and discretion we inject into pricing decisions below only impact points, we use points as our dependent variable. Essentially, we are focusing on the question, how many points do applicants pay for a given rate, conditional on their creditworthiness. We can now specify the econometric model for the statistical analysis,

$$PointsPaid_i = \beta_0 + \beta_1 Minority_i + \beta_2 RatePaid4375_i + \beta_3 RatePaid3875_i + \beta_4 MH_i + \beta_5 FICO350660_i + \beta_6 FICO661720_i + \varepsilon_i \quad (1)$$

where,

Minority	= 1 if the applicant is a minority; 0 otherwise
RatePaid4375	= 1 if the applicant prefers a rate of 4.375%; 0 otherwise
RatePaid3875	= 1 if the applicant prefers a rate of 3.875%; 0 otherwise
MH	= 1 if the property securing the mortgage is a manufactured home; 0 otherwise
FICO350660	= 1 if FICO score is between 350 and 660; 0 otherwise
FICO661720	= 1 if FICO score is between 661 and 720; 0 otherwise
ε	= 0 for the Base DGP

Except for the error term (ε), which we will discuss in more detail below, equation (1) provides a good example of how Economists tie the statistical analysis to the lender's policies. Looking at the rate sheet in Table 1, points charged are determined by the rate the applicant prefers, whether the property type is a manufactured home, and whether the FICO score is between 350 and 660 or between 661 and 720. The regression model represented by equation (1) includes all of these factors exactly as specified on the lender's rate sheet.[4] We included the Minority variable in equation (1), because that is our test for discrimination. For the Base DGP, the lender applies the rate sheet in Table 1 exactly as given, so the true, underlying β values are, β_0 = 0, β_1 = 0, β_2 = -1, β_3 = 1, β_4 = 2, β_5 = 1, and β_6 = .5. With these policy parameter values, using equation (1) to calculate the points charged to each applicant in the simulated dataset would generate the exact same points charged values that we generated by applying the lender's rate sheet in step 2 of the analytical approach discussed above.

For real-world analyses, Economists have expectations about the true, underlying β values based on their understanding of the lender's decision-making processes, but do not know these values with certainty. Therefore, they use the applicant data and regression analysis to estimate the β values in equation (1). If the regression analysis is accurate, the estimated β values should match the expectations of the true, underlying β values. If the estimated β values differ from expectations, then either the regression analysis is not accurately reflecting the DGP, or the lender is not applying the DGP as stated. An advantage of using a simulation approach is that we know the true, underlying β values of the DGP, so we can disentangle these two effects, and assess when regression analysis is accurate and when it is not (and why).

Modifications to the Base DGP

The Base DGP and equation (1) provide the starting point for our analyses. From that foundation, we explore how five modifications to the Base DGP impact the regression results.

---

[4] Note that when an econometric model includes a constant term (β_0 in equation (1)), one category in each set of mutually exclusive indicator variables must be excluded from the model. For example, rate can take on three values per the rate sheet, so we only include two 0/1 indicator variables for rate in the model. The 0/1 indicator variable for a rate of 4.000 is excluded. In equation (1), the constant reflects the number of points charged when all other 0/1 indicator variables are 0, i.e., rate = 4.000, FICO score > 720, and property type is a SFR.

We describe each of the modifications here, and provide the STATA code to generate these five modifications in Appendix B.

DGP1: Errors in Pricing Loans

The Base DGP is completely deterministic since pricing follows the rate sheet exactly with no deviations. In reality, however, errors can occur during decision-making, especially when there are large numbers of decision-makers, large numbers of loans, and complex policies and procedures. To show the types of impacts these errors can have on the regression results, we add a random component to the Base DGP. This is the error term (ε) in equation (1) above. Specifically, for each loan, we draw a random number from a uniform distribution. If the random number is greater than a specified threshold, that loan is flagged as having a processing error. An adjustment is then added to the points charged for those loans, where the adjustment amount is based on a draw from a normal distribution. We then assess how different rates of processing errors, as well as different variances for error amounts, impact the regression results. Here, and throughout all analyses below, we assume that these processing error rates and amounts are randomly distributed across applicants and not correlated with minority status.

DGP2: Sample Size

Next, we vary the sample size to show how different sample sizes can impact the regression results. We use two different sample size values, the relatively small sample size from the Base DGP (500), and a relatively large sample size (500,000), to somewhat bound the types of impacts sample size can have on the regression results.

DGP3: Correlations Between Race and Policy Factors

In real-world data, most policy factors are correlated with race to some degree, with minority applicants often having worse credit characteristics on average. To show the potential impacts these real-world correlations can have on regression results, we analyze different versions of the FICO score and property type variables that are correlated with race. Specifically, in place of the approach to generating FICO score and property type in the Base DGP above, we use,

- FICO Score
  - Random draw from a N(660, 60) distribution if minority = 1
  - Random draw from a N(700, 45) distribution if minority = 0

- MH
  - = 1 if a draw from a uniform distribution < .1 if minority = 1; 0 otherwise
  - = 1 if a draw from a uniform distribution < .05 if minority = 0; 0 otherwise

DGP4: Non-discriminatory Pricing Exceptions

Many lenders give loan officers discretion to deviate from the rate sheet pricing. One of the most common types of discretion is a reduction in points charged to match a competitor's offer. To show the potential impacts that non-discriminatory pricing exceptions can have on regression results we modify the Base DGP to include a pricing exceptions adjustment (PE) when generating points charged for each loan. We generate PE for each loan using,

PE = -1 * (draw from a uniform distribution / 2)

The expected value of PE is -0.25, which suggests that pricing exceptions reduce points charged across all loans by 25 basis points (bps) on average. Constructed and applied in this way, pricing exceptions are equally likely to occur, and have the same average amount, for both minority and non-minority applicants, so the lender is not discriminating against any group via pricing exceptions in this modified DGP.

As discussed in more detail below we analyze the impacts of non-discriminatory pricing exceptions on the regression results assuming the policy factors are uncorrelated with race as in the Base DGP and then separately assuming the policy factors are correlated with race as in DGP3.

DGP5: Discriminatory Pricing Exceptions

Finally, to show the potential impacts that discriminatory pricing exceptions can have on regression results, we modify the Base DGP to include a pricing exceptions adjustment that is correlated with race. Specifically, to determine points charged for each loan when generating the simulated dataset, we add a PE adjustment where PE is constructed as,

PE = 0 if minority = 1

PE = -1 * (draw from a uniform distribution / 2) if minority = 0

Constructed in this way, PE is beneficial to applicants and has an expected value of 0 bps for minority applicants and -25 bps for non-minority applicants.

IV. Results

Base DGP

The first analysis focuses on the Base DGP where, using the rate sheet in Table 1, the lender determines points charged based solely on FICO score, property type, and the applicant's preferred rate. The data contain no processing errors, no correlations between policy factors and race, no pricing exceptions, and no discrimination. The sample size is 500 loans. To help illustrate how regression analysis works, we generate a variety of statistical results including summary statistics and t-tests, as well as regressions using equation (1) and several variations of equation (1). The section labeled "Base DGP" in Appendix A contains the following statistical results,

0.1 Summary statistics for all variables
0.2 T-test of the difference of mean points charged by minority status
0.3 Regression: points charged = f(minority)
0.4 Regression: points charged = f(rate_3_875, rate_4_375, fico_350_660, fico_661_720, MH)
0.5 Regression: points charged = f(rate_3_875, rate_4_375, fico_350_660, fico_661_720, MH, minority)
0.6 Regression: points charged = f(rate, fico_350_660, fico_661_720, MH)
0.7 Regression: points charged = f(rate, fico_350_660, fico_661_720, MH, minority)

The section labeled "Base DGP" in Appendix B contains the STATA code that generates these results.[5]

There are several take-aways from these results that provide an initial picture of how regression analysis works, and that we will build on with the various modifications to the Base DGP discussed above. First, the unconditional disparity in points charged to minority and non-minority applicants is the same (0.036 points) whether we conduct a t-test of difference of means (0.2) or run a regression with just the minority flag (0.3). Since many people are comfortable with comparing means, but not with regression analyses, this purely mathematical point just illustrates that a regression approach yields the same unconditional disparity as taking the difference in means. Going forward, we provide either the regression results or t-tests of difference of means, but not both.

[5] Since most data are generated using random draws from probability distributions, running the code will create data that is different than that used for this report. As a result, the results will not be exactly the same as in Appendix A. However, across runs, the quantitative results should be very similar and the qualitative results should be the same.

Second, the coefficient estimates in regression (0.4), which is just equation (1) but without the minority flag, match the policy parameters in the rate sheet in Table 1 exactly. This shows that the regression analysis is accurately capturing the lender's decision-making process. When we add the minority flag to the model (regression (0.5)), all of the coefficient estimates still match the policy parameters exactly and the coefficient estimate on the minority flag is exactly 0, which is expected given there is no discrimination in the Base DGP. These results provide a simple example showing that regression analysis can capture causal relationships, i.e., generate the exact same relationships as in the lender's decision-making process.

Third, the results for regressions (0.4) and (0.5), our two preferred model specifications since they exactly reflect the lender's decision-making process, look odd with many missing values, root Mean Square Error (MSE) values of 0, and R-square values of 1. These odd results occur because the lender's decision-making process for the Base DGP is completely deterministic with no error. When the decision-making process is completely deterministic, regression analysis is not needed, since we could just simply calculate (instead of estimate) what the points charged should be for each applicant based on the rate sheet, along with their FICO score, property type, and preferred rate. The calculated points charged should exactly equal the actual points that the lender charged.

Finally, the results for regressions (0.6) and (0.7) provide an illustration of model mis-specification, which is our first example of when regression analysis does not work as well. As Graph 1 above shows, per the lender's policies, there is a non-linear relationship between points charged and applicants' preferred rate. Regressions (0.4) and (0.5) accurately capture these non-linearities with the two indicator variables, rate_3_875 and rate_4_375. Regressions (0.6) and (0.7), on the other hand, impose a linear relationship between points charged and preferred rate by including rate as a continuous variable. Comparing the results for regressions (0.6) and (0.7) to the results for regressions (0.4) and (0.5) shows how this model mis-specification reduces the accuracy of all of the coefficient estimates. As one example, per the rate sheet in Table 1, the lender charges 2 additional points to loans secured by a manufactured home (i.e., the coefficient estimate on the MH variable should be 2), all-else-equal. Due to the model mis-specification related to the rate variable, the coefficient estimates on the MH variable are instead 1.975 and 1.971. Although the impacts in this simple example are small, in real-world applications model mis-specification can have large impacts, so it is important to build statistical models that reflect the DGP (i.e., the lender's policy) as accurately as possible.

DGP1: Errors in Pricing Loans

For this analysis we modify the Base DGP to allow for the possibility that the lender makes some random errors when making pricing decisions, and then explore the impacts of different error assumptions on the regression results. Specifically, we analyze four modified DGPs with different likelihoods of a pricing error and different variances of the pricing error amount,

1) 1% chance of a pricing error, and low pricing error variance
    a. Loans with a random draw from a uniform distribution > 0.99 receive a pricing error; pricing error amount ~ N(0,1)

2) 1% chance of a pricing error, and high pricing error variance
    a. Loans with a random draw from a uniform distribution > 0.99 receive a pricing error; pricing error amount ~ N(0,4)

3) 50% chance of a pricing error, and low pricing error variance
    a. Loans with a random draw from a uniform distribution > 0.50 receive a pricing error; pricing error amount ~ N(0,1)

4) 50% chance of a pricing error, and high pricing error variance
    a. Loans with a random draw from a uniform distribution > 0.50 receive a pricing error; pricing error amount ~ N(0,4)

Each of these modifications yields a new measure of the outcome variable, which we label Points1-Points4. To assess the impact of these modifications, we estimate model specification (0.5) from the Base DGP above separately using each of the four modified Points variables. For all of these analyses, we keep the sample size at 500 loans. The section labeled "DGP1: Errors in Pricing Loans" in Appendix A contains summary statistics for all variables, as well as four sets of regression results based on the four modified DGPs, which we label (1.5.1) – (1.5.4). The section labeled "DGP1: Errors in Pricing Loans" in Appendix B contains the STATA code used to generate these results.

There are four main takeaways from these results. First, unlike the results for regression (0.5) using the deterministic Base DGP, all four regression results here look reasonable in that there are no missing values, Root MSE values do not equal 0, and the R-square values do not equal 1. Adding noise to the DGP means it is no longer deterministic. As a result, we are now appropriately estimating relationships instead of calculating relationships, and that is the purpose of regression analysis.

Second, related to the first takeaway, since we are now estimating relationships, there will be some uncertainty in the estimated coefficients. Therefore, instead of the estimated coefficients being exactly equal to the policy parameters from the rate sheet, they now differ, slightly in some instances and more in others. Fortunately, one of the advantages of regression analysis is that it provides standard error estimates (see the Std. Err. columns), which quantify the level of uncertainty in the estimated coefficients. These standard error estimates are used to conduct hypothesis testing and to quantify the likelihood that each variable in the regression model (policy factors and the minority flag in our example) has a true, underlying effect on the outcome variable.

Third, the precision of the coefficient estimates, as reflected by the standard error estimates in the Std. Err. columns, is negatively correlated to the likelihood of a pricing error occurring. Comparing the results for regression (1.5.1) to regression (1.5.3) and regression (1.5.2) to regression (1.5.4), the estimated standard errors of the coefficient estimates are approximately 7 times larger when the likelihood of a pricing error is 50% than when the likelihood of a pricing error is 1%. As one example, the estimated standard error for the minority flag is 0.011 in regression (1.51), which is based on a DGP with a 1% error rate, and 0.074 in regression (1.5.3), which is based on a DGP with a 50% error rate. All-else-equal, less precise coefficient estimates are less likely to be statistically significant. This suggests that increased noise in the outcome variable due to increased error rates, whether it is from random errors the lender makes as in our example or other sources such as data errors, reduces the likelihood of statistically significant results. Data errors are another instance where regression analysis is less accurate and does not work as well.

Finally, the precision of the coefficient estimates, as shown by the standard error estimates in the Std. Err. columns, is also negatively correlated to the variance of the pricing error amount. Comparing the results for regression (1.5.1) to regression (1.5.2) and regression (1.5.3) to regression (1.5.4), the estimated standard errors of the coefficient estimates are approximately 4 times larger when the standard deviation of the pricing error amount is 4 than when the standard deviation of the pricing error amount is 1. As one example, the estimated standard error for the minority flag is 0.074 in regression (1.5.3), which is based on a DGP with a standard deviation of the pricing error amount of 1, and 0.295 in regression (1.5.4), which is based on a DGP with a standard deviation of the pricing error amount is 4. Similar to the previous takeaway, increased noise in the outcome variable due to larger error amounts reduces the likelihood of statistically significant results.

DGP2: Sample Size

For this analysis we investigate the impact of different sample sizes on the regression results. Specifically, we analyze the same four modifications from DGP1 but with a sample size of 500,000 instead of 500. The section labeled "DGP2: Sample Size" in Appendix A contains the summary statistics and regression results. The section labeled "DGP2: Sample Size" in Appendix B contains the STATA code used to generate these results.

There are two main takeaways from these results. First, all four of the main takeaways from the previous section (DGP1: Errors in Pricing Loans) hold here as well. Specifically, all four regressions look reasonable in that there are no missing values, the Root MSE values do not equal 0, and the R-square values do not equal 1; instead of the estimated coefficients being exactly equal to the policy parameters from the rate sheet, as with the deterministic Base DGP, they now differ slightly; the coefficient estimates are much less precise and less likely to be statistically significant when the likelihood of pricing errors increases; and the coefficient estimates are much less precise and less likely to be statistically significant when the variance of the pricing error amount is higher.

Second, compared to the regression results from the modified DGPs in the section, DGP1: Errors in Pricing Loans, which were based on a sample size of 500, the regression results here based on a sample size of 500,000 show much lower estimated standard errors overall, and coefficient estimates that are much more aligned with the policy parameters in the rate sheet from Table 1. These results show that the loss of precision of the estimates from higher likelihoods of pricing errors and higher variances in pricing error amounts can be significantly reduced with large enough samples. We encourage readers to use the STATA code in Appendix B to explore different combinations of pricing error assumptions and sample sizes to further understand these impacts. The results also provide another example showing that regression analysis can capture causal relationships, i.e., generate the relationships as in the lender's decision-making process with a high degree of accuracy.

DGP3: Correlations Between Race and Policy Factors

For the next analysis we use the modified versions of FICO score and property type detailed above for DGP3 in the section on modified DGPs, to incorporate correlations between policy factors and race similar to those found in real-world mortgage applications. For all analyses in this section, we set the sample size to 500,000 and assume that 1% of loans receive a pricing error, the pricing error amount is ~N(0,1), and there is no discrimination. The section labeled "DGP3: Correlations Between Race and Policy Factors" in Appendix A contains the following results for DGP3:

3.1 Summary statistics for all variables, for all loans and by minority status
3.3 Regression: points charged = f(minority status)
3.5 Regression: points charged = f(rate_3_875, rate_4_375, fico_350_660, fico_661_720, MH, minority)

The section labeled "DGP3: Correlations Between Race and Policy Factors" in Appendix B contains the STATA code used to generate these results.

There are two main takeaways from these results. First, unlike for the Base DGP, the unconditional disparity in points, as shown by regression (3.3), is now large and statistically significant, indicating that minority applicants pay 0.342 points more on average than non-minority applicants. DGP3 does not include any discrimination, so we know that this unconditional disparity is not capturing discrimination. Instead, this result provides a classic example of omitted variable bias, which is another scenario when regression analysis does not work as well. Per DGP3, we know that the lender considers both FICO score and property type when determining how many points to charge. Given the correlations we incorporated into DGP3, we know that minorities have lower FICO scores on average and are more likely to use a manufactured home to secure the mortgage. The summary statistics (3.1) reflect these correlations. Since a lower FICO score and securing a loan with a manufactured home both increase points charged via the LLPAs, and minority applicants tend to have lower FICO scores and are more likely to secure loans with a manufactured home, we would expect minority applicants to be charged higher points on average. This is exactly what the unconditional disparity shows. As shown in regression (3.5), once we account for the differences in FICO score and property type across applicants, the minority coefficient estimate is approximately 0 and not statistically significant, as expected for DGP3.

Second, looking at regression (3.5) the coefficient estimates again very closely match the parameters from the rate sheet. Following results from above, the pricing errors we incorporated into DGP3 added some noise into the coefficient estimates and increased the standard errors, but the large sample size (500,000) significantly offset this noise resulting in highly accurate and precise coefficient estimates. The results from regression (3.5) provide another example that regression analyses can capture causal relationships with a high degree of certainty, even when there is noise from pricing errors and correlations between policy factors and race.

DGP4.1: Non-Discriminatory Exceptions, No Correlations Between Race and Policy Factors

The next extension to the statistical analysis is to allow discretionary, but non-discriminatory, pricing exceptions. The amount of the pricing exception is randomly drawn and non-discriminatory across all loans. We assume that pricing exceptions are always beneficial to applicants, so the randomly drawn exception amount for each loan is subtracted from the initial points charged amount for that loan based on the lender's policy, i.e., the rate sheet from Table 1.[6] We continue to set the sample size to 500,000 loans and assume that 1% of loans receive a pricing error, the pricing error amount is ~N(0,1), and there is no correlation between policy factors and race. The section labeled "DGP4.1: Non-Discriminatory Exceptions, No Correlations" in Appendix A contains the following results for this DGP,

4.1.1 Summary statistics for all variables
4.1.3 Regression: points charged = f(minority status)
4.1.5 Regression: points charged = f(rate_3_875, rate_4_375, fico_350_660, fico_661_720, mh, minority)
4.1.9 Regression: points charged = f(rate_3_875, rate_4_375, fico_350_660, fico_660_721, mh, PE, minority)
4.1.10 T-test of difference of means of pricing exceptions amount

The section labeled "DGP4.1: Non-Discriminatory Exceptions, No Correlations" in Appendix B contains the STATA code used to generate these results.

There are several takeaways from these results. First, as shown in regression (4.1.3) the unconditional disparity in points is approximately 0. This is as expected given that the pricing errors we included in the DGP are randomly distributed and not correlated with any policy factors or race; there is no omitted variable bias since FICO score, property type, and PE are not correlated with race; and the DGP includes no discrimination.

---

[6] For real-world applications, exceptions may be beneficial for some applicants and harmful to others. In these instances, care must be taken to analyze beneficial exceptions and harmful exceptions separately, especially when analyzing the likelihood of receiving an exception.

Second, in regression (4.1.5), which includes the policy factors from the rate sheet along with a minority flag, the coefficient estimates on all of the policy factors match very closely to the policy parameters from the rate sheet. In addition, the estimated coefficient on the minority flag is approximately 0, which is as expected given the DGP incorporates no discrimination. These results provide another example of the regression analysis accurately capturing the lender's pricing policies and decision-making processes, and testing for the presence of discrimination. It is worth noting that the estimated coefficients are not exactly equal to the policy parameters as they were for the Base DGP analysis above. Similar to the results for DGP1 above, these deviations are in part a result of the random pricing errors we included in the DGP here. They are also in part a result of the randomly drawn pricing exceptions for this DGP. Overall, the deviations from the policy parameters are small because the very large sample size dominates the impact of the noise from the pricing errors and exceptions. We encourage readers to use the code in Appendix B to explore how the results change with different sample sizes.

Third, the results from regression (4.1.9) where we add the pricing exceptions variable provide valuable insight into the role of PE in DGP4.1 and the regression analysis. Including PE in the regression model seems like a reasonable choice because our goal is to control for all factors the lender considers when making pricing decisions, and whether to grant a pricing exception is one factor the lender considers in DGP4.1. As the results for regression (4.1.9) show, the coefficient estimate on the PE variable is essentially -1, which makes sense since the randomly drawn pricing exceptions were simply subtracted from the points charged. In addition, the coefficient estimate on the minority variable is again essentially 0 and not statistically significant. This also makes sense, since in DGP4.1, the lender did not directly consider race when determining the number of points to charge. For this DGP, where PE is randomly determined and uncorrelated with race, it did not matter whether the PE variable was included in the model (4.1.10) or not (4.1.10), since the estimated coefficient on the minority flag is approximately 0 in both cases. In the discussion of DGP5 below, we show how this result and these explanations change significantly when the pricing exceptions are discriminatory.

Finally, the t-test of difference of means of the pricing exceptions amount (4.1.10) shows that the average pricing exception amount is approximately the same for both minority and non-minority applicants. Again, this result is consistent with how PE was randomly generated and applied in DGP4.1.

DGP4.2: Non-Discriminatory Exceptions, Correlations Between Race and Policy Factors

Building off the analyses of DGP4.1, we now use the modified versions of FICO score and property type detailed above for DGP3 in the section on modified DGPs, to incorporate correlations between policy factors and race similar to those found in real-world mortgage applications. All other aspects of the DGP here are the same as for DGP4.1. The section labeled "DGP4.2: Non-Discriminatory Exceptions, Correlations" in Appendix A contains the following results for this DGP,

4.2.1 Summary statistics for all variables
4.2.3 Regression: points charged = f(minority status)
4.2.5 Regression: points charged = f(rate_3_875, rate_4_375, fico_350_660, fico_661_720, mh, minority)
4.2.9 Regression: points charged = f(rate_3_875, rate_4_375, fico_350_660, fico_660_721, mh, PE, minority)
4.2.10 T-test of difference of means of pricing exceptions amount

The section labeled "DGP4.2: Non-Discriminatory Exceptions, Correlations" in Appendix B contains the STATA code used to generate these results.

There are two main takeaways from these results. First, the correlations between policy factors and race create omitted variable bias. As shown in regression (4.2.3), the unconditional disparity in points is now 0.337 points. This disparity suggests that, on average, minority applicants pay 0.337 points more than non-minority applicants. For the exact same reasons presented above in the discussion of correlations between policy factors and race for DGP3, this unconditional disparity is another example of omitted variable bias. In the DGP here, we know that the lender considers both FICO score and property type when determining how many points to charge. We also know that minorities have lower FICO scores on average and are more likely to use a manufactured home to secure a mortgage. Since a lower FICO score and using a manufactured home to secure a loan both increase points paid via the LLPAs, and minorities tend to have lower FICO scores and are more likely to use a manufactured home to secure a loan, we would expect their points charged to be higher on average. As shown in regression (4.2.5), once we account for the differences in FICO score and property type across applicants, the minority coefficient estimate is near 0 and not statistically significant.

Second, all of the other main results for DGP4.1 (takeaways 2-4 from the previous section) hold in exactly the same way for DGP4.2. Stated another way, for DGP4.1 and DGP4.2, the only impact that adding correlations between policy factors and race has is to change the unconditional disparity in points charged. If we accurately control for all policy factors, or if we focus on disparities in pricing exceptions, we get the same results with or without the correlations.

DGP5.1: Discriminatory Exceptions, No Correlations Between Race and Policy Factors

The next extension to the analysis is to allow discretionary, and discriminatory, pricing exceptions. For non-minority applicants, the pricing exception amount for each loan is generated as ½ times a random draw from a uniform distribution. For minority applicants, the pricing exception amount is set to 0 for all loans. With this construction, pricing exceptions are 25 bps higher on average for non-minority applicants than for minority applicants. We assume that pricing exceptions are always beneficial to applicants, so the exception amount for each loan is subtracted from the initial points charged amount for that loan based on the lender's policy, i.e., the rate sheet from Table 1. We continue to set the sample size to 500,000 loans and assume that 1% of loans receive a pricing error, the pricing error amount is ~N(0,1), and there is no correlation between policy factors and race. The section labeled "DGP5.1: Discriminatory Exceptions, No Correlations" in Appendix A contains the following results for this DGP,

- 5.1.1 Summary statistics for all variables
- 5.1.3 Regression: points charged = f(minority status)
- 5.1.5 Regression: points charged = f(rate_3_875, rate_4_375, fico_350_660, fico_661_720, mh, minority)
- 5.1.9 Regression: points charged = f(rate_3_875, rate_4_375, fico_350_660, fico_661_720, mh, PE, minority)
- 5.1.10 T-test of difference of means of pricing exception amounts

The section labeled "DGP5.1: Discriminatory Exceptions, No Correlations" in Appendix B contains the STATA code used to generate these results.

There are several takeaways in these results. First, the results provide insight into how to interpret the impact of pricing exceptions on unconditional disparities. As shown in regression (5.1.3), the unconditional disparity in points charged is 0.253 points. Following discussions above for DGP3 and DGP4.2, this disparity might just reflect omitted variable bias, since we have not accounted for any differences in pricing errors or policy factors across loans. We can rule out this explanation, however, since neither the pricing errors nor any of the policy factors are correlated with race. We know this given the DGP here, and in addition, this could be verified in the data. That leaves two possible drivers of the unconditional disparity, the lender applied a discriminatory adjustment directly to points charged or the disparity reflects omitted variable bias stemming from pricing exceptions. Although we can rule out the first of these drivers here since we know that direct discrimination is not part of this DGP, in real-world applications we rarely know with certainty if the lender applied a direct discriminatory adjustment to points charged. However, the statistical analysis and results provide useful insights here. We know that pricing exceptions impact points charged per the lender's policies, and as t-test (5.1.10) shows, pricing exceptions are correlated with race. These two results suggest that pricing exceptions generate at least some omitted variable bias in the unconditional disparity. Further, as t-test (5.1.10) shows, the average pricing exceptions amount for non-minority applicants is 0.25 points higher than for minority applicants, which is approximately the same as the unconditional disparity in points charged of 0.253 points. Taken together, all of these results suggest that omitted variable bias from pricing exceptions is the primary driver of the unconditional disparity. If one or more of these relationships did not occur in the lender's policies or in the data, this would suggest that direct discrimination may be the driver of some or all of the unconditional disparity.

Second, in regression (5.1.5), which includes the policy factors from the rate sheet along with a minority flag, the coefficient estimates on all of the policy factors match very closely to the policy parameters from the rate sheet. This provides another example of how regression analysis can accurately capture a lender's policies and decision-making process, which in turn leads to accurate statistical tests for discrimination. Again, we note that there is a very small amount of variation in these estimates, i.e., they are not exactly equal to the policy parameters like they were for the Base DGP analysis above. As a reminder, these small deviations are a result of the pricing errors and pricing exceptions we included in the DGP here, and how the very large sample size dominates the impact of the noise from the pricing errors and exceptions.

Third, in regression (5.1.5), the estimated conditional disparity is 0.2504 points (i.e., the coefficient on the minority flag), which is approximately equal to both the unconditional disparity for points charged and the disparity in the pricing exceptions amount. Therefore, adding the policy factors to the regression model had little to no impact on the disparity, which further supports the argument above that omitted variable bias from the policy factors was not driving the unconditional disparity. This result also supports the argument that the unconditional disparity is driven by either direct discrimination or by the pricing exceptions.

Fourth, the results provide insight into how to interpret the impact of pricing exceptions on conditional disparity estimates. Regression (5.1.9) extends regression (5.1.5) by adding the PE variable to the model. Adding the PE variable is a reasonable choice because we want to control for all factors the lender considers when making pricing decisions, and whether to offer a pricing exception is one factor the lender considers in this DGP. As the results for regression (5.1.9) show, the coefficient estimate on the PE variable is essentially -1, which makes sense since in this DGP pricing exceptions were simply subtracted from the points charged. In addition, the coefficient estimate on the minority variable is now essentially 0 and not statistically significant. This also makes sense, since regression (5.1.9) includes every factor the lender considered when determining the number of points to charge in this DGP, and the lender did not directly consider race when determining the number of points to charge. Since excluding the PE variable resulted in a conclusion of discrimination (minority coefficient of 0.2504 in regression (5.1.5)), and including the PE variable resulted in a conclusion of no discrimination (minority coefficient estimate of 0 in regression (5.1.9)), the question then is whether it is appropriate to include the PE variable in the regression model. On one hand, it could be completely legitimate for a lender to offer pricing exceptions and for these offers to vary across loans. For example, for legitimate market and economic reasons, some applicants might be less likely to have competing offers, which means it is less likely that the lender will offer them a pricing exception. This scenario is outside of the lender's control and justifies including the PE variable as a control variable in the regression model. Very importantly, disparate impact risk related to the lender's pricing exception policy would be a significant concern here since minority applicants often are less likely to have competing offers. Alternatively, the pricing exceptions could be capturing discrimination. For example, loan officers might proactively suggest to non-minority applicants that obtaining a competing offer increases their chances of receiving a pricing exception, but not provide similar guidance to minority applicants. This scenario is within the lender's control and justifies excluding the PE variable as a control variable, since including it would mask potential discrimination. To disentangle these two scenarios, it is important to estimate two separate regression models, one with and one without, the PE variable to assess the extent that pricing exceptions are driving any estimates of discrimination. When the results indicate that exceptions are driving estimates of discrimination (i.e., the disparity declines or disappears when PE is added to the model), a follow-up review of the lender is needed to determine whether the exceptions are legitimate or discriminatory.

Finally, t-test (5.1.10) shows that minority applicants receive a pricing exception that is 0.25 points smaller on average than for non-minority applicants. This is one of the most important results in this entire report. In regressions (5.1.5) and (5.1.9) we took the standard approach of first controlling for all policy factors exactly how the lender considered them, and then adding a minority flag to estimate the amount of discrimination. In our simulated example, including every factor the lender considered, and including them exactly how the lender considered them, was easy since the policies are not complex and we have data for all relevant policy factors. As a result, we were able to estimate the amount of potential discrimination very accurately in regressions (5.1.5) and (5.1.9). In real-world applications, however, policies are typically much more complex, and we often do not have electronic data for every factor a lender considers when making credit decisions. As a result, issues such as omitted variable bias and model mis-specification, which reduce the reliability of regression results are often a significant challenge. As t-test (5.1.10) shows, however, we can get a highly accurate estimate of potential discrimination with a very simple statistical analysis without needing to accurately account for the lender's policies, procedures, and decision-making process. Very importantly, the validity of this result hinges on discretionary pricing exceptions being the only source of potential discrimination, which is true for the DGP here. While this is typically the primary source of discrimination in real-world applications, there are other sources of discrimination as well, such as the lender applying its rate sheets differently for applicants of different groups. Therefore, care must be taken when generalizing the results here to other analyses.

DGP 5.2: Discriminatory Exceptions, Correlations Between Race and Policy Factors

Building off the analyses of DGP5.1, we now use the modified versions of FICO score and property type detailed above for DGP3 in the section on modified DGPs, to incorporate correlations between policy factors and race similar to those found in real-world mortgage applications. All other aspects of the DGP here are the same as for DGP5.1. The section labeled

"DGP5.2: Discriminatory Exceptions, Correlations" in Appendix A contains the following results for this DGP,

5.2.1 Summary statistics for all variables
5.2.3 Regression: points charged = f(minority status)
5.2.5 Regression: points charged = f(rate_3_875, rate_4_375, fico_350_660, fico_661_720, mh, minority)
5.2.9 Regression: points charged = f(rate_3_875, rate_4_375, fico_350_660, fico_661_720, mh, PE, minority)
5.2.10 T-test of difference of means of pricing exception amounts

The section labeled "DGP5.2: Discriminatory Exceptions, Correlations" in Appendix B contains the STATA code used to generate these results.

There are three main takeaways from these results. First, omitted variable bias from multiple sources is now driving the unconditional disparity. As regression (5.2.3) shows, the unconditional disparity is now 0.593 points, which suggests that, on average, points charged is 0.593 points higher for minority applicants than for non-minority applicants. Following the same arguments discussed in the first takeaway for DGP5.1 above, part of this unconditional disparity is driven by omitted variable bias stemming from pricing exceptions. By construction in this DGP, PE is correlated with race and impacts points charged, so the unconditional disparity is indirectly capturing the impact of pricing exceptions on points charged. In addition, following the same arguments discussed in the first takeaway for DGP3 above, part of this unconditional disparity is also driven by omitted variable bias stemming from the policy factors, which are correlated with race in this DGP. Omitted variable bias from multiple sources explains why the unconditional disparity here (0.593 points) is higher than the unconditional disparity for DGP5.1 (0.253 points).

Second, although we know direct discrimination is not driving any part of the unconditional disparity here given the DGP, in real-world applications that is one possibility that needs to be explored in the data. As regression (5.2.9) shows, once we control for differences in FICO score, property type, and pricing exceptions across loans, the estimated conditional disparity is approximately 0 (the coefficient on the minority flag). This suggests that there is no direct discrimination, and that the unconditional disparity is driven solely by omitted variable bias stemming from pricing exceptions and the policy factors.

Third, all of the other main results for DGP5.1 (takeaways 2-4 from the previous section) hold in exactly the same way for DGP5.2. Stated another way, comparing the results for DGP5.1 and DGP5.2, the only impact of adding correlations between policy factors and race to the data was a change to the unconditional disparity in points charged. If we accurately control for all policy factors, or if we focus on disparities in pricing exceptions amounts, we get the same results with or without the correlations.

V. APR Analyses

To this point, we have made only limited mention of analyses of APR disparities. Since focusing on APR disparities is a common and standard approach Economists use to analyze pricing decisions for discrimination, we end with a brief discussion of these analyses here. Focusing on APR as the measure of price is appealing since TILA/RESPA requires lenders to disclose APR to applicants on most mortgage loans, and it simplifies the analysis since it combines the various components of mortgage pricing into one overall measure of price. Unfortunately, APR disparities are difficult to interpret, and it can be difficult to accurately estimate discrimination in mortgage pricing with APR disparities. We mention three of the most important challenges here. First, although APR is a non-linear, net present value formula, Economists typically estimate conditional APR disparities using a linear regression model. As illustrated in the discussion of the Base DGP above, this model mis-specification impacts all coefficient estimates, including the disparity estimate. Second, APR is a deterministic formula that combines various applicant and loan characteristics into one pricing measure, so there is no uncertainty or error term. Again, as discussed above, when the DGP is deterministic there is no need to use regression analysis to estimate relationships between policy factors, race, and outcomes, since we can just calculate these relationships directly. Finally, omitted variable bias is a significant challenge for analyses of APR disparities. For any given rate/point menu, APR will be lower for rate/point combinations with lower rates and higher points, all-else-equal. Because of this, rate and points should both be included as control variables in any regression model with APR as the outcome measure. If they are not included, which is typically the approach Economists take, the estimated conditional disparities are impacted by omitted variable bias. However, including rate and points in these regression models will likely mask potential discrimination, since adjustments to rate or points via direct discrimination or indirectly through exceptions are a primary opportunity for lenders to discriminate. Disentangling all of these effects is difficult.

Because of these challenges, one approach to consider when focusing on APR is to simply calculate APR disparities directly. Specifically, for each loan, generate an APR after first zeroing out all pricing exceptions and all discretionary deviations from the rate sheet and formal pricing policies, and then compute the difference between the adjusted APR and actual APR for each loan. Generating the average of this difference measure separately for two groups and then taking the difference of the two averages provides a measure of an APR disparity. Similar to the analyses of pricing exceptions above, if an APR disparity exists, further review is needed to identify the source of the disparity and whether it is legitimate or discriminatory.

VI. Conclusion

In this report we have attempted to shed light on the general approach Economists use to conduct fair lending statistical analyses. Using simplified pricing policies and simulated data, we showed empirically that regression analysis can capture causal relationships, i.e., generate the exact same relationships as in the lender's decision-making process, and generate highly accurate estimates of discrimination. We also provided many examples of when regression analyses do not work as well. Specifically, we showed how sample size, model mis-specification, omitted variables bias, pricing errors, and pricing exceptions can impact estimated disparities from regression analyses. Throughout, we focused on understanding the intuition behind all of these concepts using several empirical examples, and minimized any discussion or reference to the underlying econometric theory to make the results and conclusions as accessible as possible to end users. In addition, we provided all of the code in Appendix B, so readers can replicate the results and then extend the analyses in any direction of interest.

Most likely, none of the analyses in this report will exactly mirror any analyses in real world applications. Hopefully, however, the general approach, as well as the foundational principles and concepts covered here will provide the necessary foundation and understanding to conduct a broad array of fair lending statistical analyses that are highly accurate and reliable.

References

El Anshasy, Amany, Yoshiaki Shimazaki, and Gregory E. Elliehausen. 2005. "Mortgage Brokers and the Subprime Mortgage Market," Proceedings 953, Federal Reserve Bank of Chicago.

Calem, Paul S., and Stanley D. Longhofer. 1999. "Mortgage Brokers and Fair Lending." *Economic Commentary*, Federal Reserve Bank of Cleveland, May 15th, 1999 http://webs.wichita.edu/longhofer/Vita/mortbrok.pdf

Kleiner, Morris M. and Richard M. Todd. 2007. "Mortgage Broker Regulations that Matter: Analyzing Earnings, Employment, and Outcomes for Consumers," NBER Working Paper No. w13684, December 2007.

Appendix A: Statistical Results

Base DGP

0.1 Summary statistics for all variables

Variable	Obs	Mean	SD	Min	Max
Minority	500	.250	.433	0	1
FICO	500	659.134	60.296	493	827
FICO_350_660	500	.508	.500	0	1
FICO_661_720	500	.342	.475	0	1
MH	500	.094	.292	0	1
Rate	500	4.004	.156	3.875	4.375
Rate_4_375	500	.132	.339	0	1
Rate_4	500	.504	.501	0	1
Rate_3_875	500	.364	.482	0	1
Points	500	1.099	.959	-1	4

0.2 T-test of the difference of mean points charged by minority status

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% conf.	Interval]
0	375	1.108	.049	.945	1.012	1.204
1	125	1.072	.090	1.003	.894	1.250
Combined	500	1.099	.043	.959	1.015	1.183
diff		.036	.0991		-.159	.231

```
Diff = mean(0) - mean(1)                                  t = 0.363
H0: diff = 0                                degrees of freedom = 498
Ha: diff < 0            Ha: diff != 0           Ha: diff > 0
Pr(T < t) = 0.642       Pr(|T| > |t|) = 0.717   Pr (T > t) = 0.358
```

0.3 Regression: points charged = f(minority)

Source	SS	df	MS
Model	.122	1	.122
Residual	458.728	498	.921
Total	458.850	499	.920

```
Number of obs =     500
F(1, 498) =       0.130
Prob > F =        0.717
R-squared =       0.000
Adj R-squared=   -0.002
Root MSE =        0.960
```

Points	Coefficient	Std. Err.	t	P>|t|	[95% conf.	Interval]
Minority	-.036	.099	-0.36	0.717	-.231	.159
_cons	1.108	.050	22.36	0.000	1.010	1.205

0.4 Regression: points charged = f(rate_3_875, rate_4_375, fico_350_660, fico_661_720, MH)

Source	SS	df	MS		
				Number of obs =	500
				F(5, 494) =	.
Model	458.850	5	91.770	Prob > F =	.
Residual	0	494	0	R-squared =	1.000
				Adj R-squared=	1.000
Total	458.850	499	.920	Root MSE =	0.000

Points	Coefficient	Std. Err.	t	P>|t|	[95% conf.	Interval]
Rate_4_375	-1	.	.	.	.	.
Rate_3_875	1	.	.	.	.	.
FICO_350_660	1	.	.	.	.	.
FICO_661_720	.5	.	.	.	.	.
MH	2	.	.	.	.	.
_cons	0	.	.	.	.	.

0.5 Regression: points charged = f(rate_3_875, rate_4_375, fico_350_660, fico_661_720, MH, minority)

Source	SS	df	MS		
				Number of obs =	500
				F(6, 493) =	.
Model	458.850	6	76.475	Prob > F =	.
Residual	0	493	0	R-squared =	1.000
				Adj R-squared=	1.000
Total	458.850	499	.920	Root MSE =	0.000

Points	Coefficient	Std. Err.	t	P>|t|	[95% conf.	Interval]
Rate_4_375	-1	.	.	.	.	.
Rate_3_875	1	.	.	.	.	.
FICO_350_660	1	.	.	.	.	.
FICO_661_720	.5	.	.	.	.	.
MH	2	.	.	.	.	.
Minority	0	.	.	.	.	.
_cons	0	.	.	.	.	.

0.6 <u>Regression: points charged = f(rate, fico_350_660, fico_661_720, MH)</u>

Source	SS	df	MS		
				Number of obs =	500
				F(4, 495) =	1703.600
Model	427.776	4	106.094	Prob > F =	0.000
Residual	31.073	495	.063	R-squared =	0.932
				Adj R-squared=	0.932
Total	458.850	499	.920	Root MSE =	0.251

Points	Coefficient	Std. Err.	t	P>|t|	[95% conf.	Interval]
Rate	-3.955	.072	-54.75	0.000	-4.097	-3.813
FICO_350_660	.999	.033	30.22	0.000	.934	1.064
FICO_661_720	.532	.035	15.27	0.000	.463	.600
MH	1.971	.039	51.25	0.000	1.895	2.046
_cons	16.061	.293	54.86	0.000	15.486	16.636

0.7 <u>Regression: points charged = f(rate, fico_350_660, fico_661_720, MH, minority)</u>

Source	SS	df	MS		
				Number of obs =	500
				F(5, 494) =	1365.530
Model	427.890	5	85.578	Prob > F =	0.000
Residual	30.959	494	.063	R-squared =	0.933
				Adj R-squared=	0.931
Total	458.850	499	.920	Root MSE =	0.250

Points	Coefficient	Std. Err.	t	P>|t|	[95% conf.	Interval]
Rate	-3.955	.072	-54.79	0.000	-4.097	-3.814
FICO_350_660	.997	.033	30.16	0.000	.932	1.062
FICO_661_720	.529	.035	15.18	0.000	.461	.598
MH	1.975	.039	51.24	0.000	1.899	2.051
Minority	.035	.026	1.35	0.077	-.016	.086
_cons	16.055	.293	54.88	0.000	15.480	16.630

DGP1: Errors in Pricing Loans

1.1 Summary statistics for all variables

Variable	Obs	Mean	SD	Min	Max
Minority	500	.274	.447	0	1
FICO	500	663.040	57.632	493	830
FICO_350_660	500	.476	.500	0	1
FICO_661_720	500	.368	.483	0	1
MH	500	.076	.265	0	1
Rate	500	4.019	.164	3.875	4.375
Rate_4_375	500	.158	.365	0	1
Rate_4	500	.520	.500	0	1
Rate_3_875	500	.322	.468	0	1
Points	500	.976	.913	-1	4
Pricing_Error1	500	.018	.133	0	1
Points1	500	.976	.923	-1	4
Points2	500	.998	.994	-3.756	5.467
Pricing_Error2	500	.510	.500	0	1
Points3	500	.985	1.181	-2.106	4.870
Points4	500	1.123	3.042	-9.070	16.736

1.5.1 Regression: Points1 = f(rate_3_875, rate_4_375, fico_350_660, fico_661_720, MH, minority)

```
Source      |     SS          df          MS          Number of obs =     500
--------------------------------------------------    F(6, 493)    =  6440.780
Model       |     419.426      6       69.904         Prob > F =         0.000
Residual    |       5.351     493        .011         R-squared =        0.987
--------------------------------------------------    Adj R-squared=     0.987
Total       |     424.776     499        .851         Root MSE =         0.104
-------------------------------------------------------------------------
Points1     | Coefficient   Std. Err.    t       P>|t| [95% conf. Interval]
---------------------------------------------------------------------------
Rate_4_375  |     -.983       .013    -73.39     0.000   -1.009      -.957
Rate_3_875  |     1.013       .011     96.68     0.000     .992      1.034
FICO_350_660|      .986       .014     72.27     0.000     .959      1.012
FICO_661_720|      .489       .014     34.68     0.000     .462       .517
MH          |     2.039       .018    115.54     0.000    2.004      2.074
Minority    |     -.000       .011     -0.02     0.987    -.021       .020
_cons       |      .001       .013      0.09     0.931    -.025       .027
---------------------------------------------------------------------------
```

1.5.2 Regression: Points2 = f(rate_3_875, rate_4_375, fico_350_660, fico_661_720, MH, minority)

```
Source       |     SS           df           MS          Number of obs =   500
---------------------------------------------------------- F(6, 493)    = 421.260
Model        |     412.560       6       68.760          Prob > F =      0.000
Residual     |      80.469     493         .163          R-squared =     0.837
---------------------------------------------------------- Adj R-squared=  0.835
Total        |     493.029     499         .988          Root MSE =      0.404
-------------------------------------------------------------------------------
Points2      | Coefficient    Std. Err.     t      P>|t|  [95% conf. Interval]
--------------------------------------------------------------------------------
Rate_4_375   |    -.999         .052      -19.22   0.000   -1.101    -.897
Rate_3_875   |     .980         .041       24.11   0.000     .900    1.059
FICO_350_660|     .990         .053       18.72   0.000     .886    1.094
FICO_661_720|     .487         .055        8.90   0.000     .380     .595
MH           |   2.026         .066       29.60   0.000    1.892    2.161
Minority     |    -.001         .041       -0.01   0.988    -.080     .079
_cons        |     .036         .051        0.71   0.478    -.063     .135
--------------------------------------------------------------------------------
```

1.5.3 Regression: Points3 = f(rate_3_875, rate_4_375, fico_350_660, fico_661_720, MH, minority)

```
Source       |     SS           df           MS          Number of obs =   500
---------------------------------------------------------- F(6, 493)    = 134.240
Model        |     431.624       6       71.937          Prob > F =      0.000
Residual     |     264.201     493         .536          R-squared =     0.620
---------------------------------------------------------- Adj R-squared=  0.616
Total        |     695.826     499        1.394          Root MSE =      0.732
-------------------------------------------------------------------------------
Points3      | Coefficient    Std. Err.     t      P>|t| [95% conf. Interval]
--------------------------------------------------------------------------------
Rate_4_375   |     -.928        .094      -9.86    0.000   -1.113    -.743
Rate_3_875   |     1.068        .074      14.51    0.000     .924    1.213
FICO_350_660|      .928        .096       9.68    0.000     .740    1.116
FICO_661_720|      .498        .099       5.02    0.000     .303     .693
MH           |     2.160        .124      17.42    0.000    1.917    2.404
Minority     |      .052        .074       0.70    0.484    -.093     .196
_cons        |     -.016        .092      -0.18    0.861    -.196     .164
--------------------------------------------------------------------------------
```

1.5.4 Regression: Points4 = f(rate_3_875, rate_4_375, fico_350_660, fico_661_720, MH, minority)

Source	SS	df	MS	
				Number of obs = 500
				F(6, 493) = 7.140
Model	369.182	6	61.530	Prob > F = 0.000
Residual	4247.272	493	8.615	R-squared = 0.080
				Adj R-squared= 0.069
Total	4616.454	499	9.251	Root MSE = 2.935

Points4	Coefficient	Std. Err.	t	P>|t|	[95% conf.	Interval]
Rate_4_375	-.715	.377	-1.89	0.059	-1.457	.026
Rate_3_875	.667	.295	2.26	0.024	.087	1.247
FICO_350_660	1.304	.384	3.39	0.001	.549	2.059
FICO_661_720	.617	.398	1.55	0.121	-.164	1.398
MH	2.161	.497	4.35	0.000	1.184	3.138
Minority	.172	.295	0.58	0.560	-.408	.752
_cons	-.038	.367	-0.10	0.918	-.758	.683

DGP2: Sample Size

2.1 Summary statistics for all variables

Variable	Obs	Mean	SD	Min	Max
Minority	500,000	.250	.433	0	1
FICO	500,000	659.400	59.942	389	925
FICO_350_660	500,000	.507	.500	0	1
FICO_661_720	500,000	.339	.473	0	1
MH	500,000	.100	.300	0	1
Rate	500,000	4.012	.162	3.875	4.375
Rate_4_375	500,000	.149	.356	0	1
Rate_4	500,000	.501	.500	0	1
Rate_3_875	500,000	.349	.477	0	1
Points	500,000	1.076	.976	-1	4
Error1	500,000	.010	.099	0	1
Points1	500,000	1.076	.980	-3.559	5.967
Points2	500,000	1.076	1.050	-11.946	13.691
Error2	500,000	.501	.500	0	1
Points3	500,000	1.075	1.205	-4.885	8.068
Points4	500,000	1.066	2.999	-18.815	18.888

2.5.1 Regression: Points1 = f(rate_3_875, rate_4_375, fico_350_660, fico_661_720, MH, minority)

Source	SS	df	MS	
				F(6, 499993)=99999.00
Model	475826.651	6	79304.442	Prob > F = 0.000
Residual	4810.812	499,993	.010	R-squared = 0.990
				Adj R-squared= 0.990
Total	480637.464	499,999	.961	Root MSE = 0.098

Number of obs=500,000

Points1	Coefficient	Std. Err.	t	P>|t|	[95% conf.	Interval]
Rate_4_375	-1.0001	.0004	-2445.84	0.000	-1.0009	-.9993
Rate_3_875	.9997	.0003	3296.67	0.000	.9991	1.0003
FICO_350_660	.9997	.0004	2478.71	0.000	.9989	1.0005
FICO_661_720	.5001	.0004	1173.51	0.000	.4992	.5009
MH	2.0001	.0005	4324.44	0.000	1.9992	2.0010
Minority	-.0001	.0003	-0.39	0.699	-.0008	.0005
_cons	.0004	.0004	1.00	0.316	-.0004	.0012

2.5.2 Regression: Points2 = f(rate_3_875, rate_4_375, fico_350_660, fico_661_720, MH, minority)

```
Source       |     SS           df           MS          Number of obs= 500,000
---------------------------------------------------------- F(6, 499993)=99999.000
Model        |   475260.012         6   79210.002        Prob > F =        0.000
Residual     |    75462.693   499,993        .151        R-squared =       0.863
---------------------------------------------------------- Adj R-squared=   0.863
Total        |   550722.705   499,999       1.101        Root MSE =        0.389
------------------------------------------------------------------------------
Points2      | Coefficient    Std. Err.     t      P>|t| [95% conf. Interval]
--------------------------------------------------------------------------------
Rate_4_375  |     -.9967       .0016    -615.43   0.000  -.9999      -.9935
Rate_3_875  |     1.0007       .0012     826.32   0.000   .9983      1.0030
FICO_350_660|      .9987       .0016     625.21   0.000   .9956      1.0019
FICO_661_720|      .5000       .0017     296.28   0.000   .4967       .5033
MH          |     1.9998       .0018    1091.74   0.000  1.9962      2.0034
Minority    |     -.0006       .0013      -0.50   0.620  -.0031       .0019
_cons       |      .0004       .0015       0.25   0.800  -.0026       .0034
--------------------------------------------------------------------------------
```

2.5.3 Regression: Points3 = f(rate_3_875, rate_4_375, fico_350_660, fico_661_720, MH, minority)

```
Source       |     SS           df           MS          Number of obs=  500,000
---------------------------------------------------------- F(6, 499993)= 99999.000
Model        |    475974.668         6  79329.111        Prob > F =         0.000
Residual     |    250073.898   499,993       .500        R-squared =        0.656
---------------------------------------------------------- Adj R-squared=    0.656
Total        |    726048.567   499,999      1.452        Root MSE =         0.707
------------------------------------------------------------------------------
Points3      |   Coefficient  Std. Err.     t      P>|t| [95% conf. Interval]
--------------------------------------------------------------------------------
Rate_4_375  |     -1.0009      .0029    -339.48   0.000  -1.0066     -.9951
Rate_3_875  |      1.0017      .0022     454.39   0.000    .9974     1.0060
FICO_350_660|       .9962      .0029     342.59   0.000    .9905     1.0019
FICO_661_720|       .4998      .0031     162.69   0.000    .4938      .5059
MH          |      2.0003      .0033     599.86   0.000   1.9937     2.0068
Minority    |      -.0016      .0023      -0.68   0.498   -.0061      .0030
_cons       |       .0005      .0028       0.19   0.846   -.0050      .0061
--------------------------------------------------------------------------------
```

2.5.4 Regression: Points4 = f(rate_3_875, rate_4_375, fico_350_660, fico_661_720, MH, minority)

```
Source       |     SS          df          MS          Number of obs=  500,000
------------------------------------------------------ F(6, 499993)= 10064.160
Model        |     484592.861        6   80765.477     Prob > F =        0.000
Residual     |    4012474.780  499,993       8.025     R-squared =       0.108
------------------------------------------------------ Adj R-squared=    0.108
Total        |    4497067.640  499,999       8.994     Root MSE =        2.833
-------------------------------------------------------------------------
Points4      | Coefficient    Std. Err.     t      P>|t| [95% conf. Interval]
---------------------------------------------------------------------------
Rate_4_375  |      -.9951       .0118     -84.26   0.000  -1.0183    -.9720
Rate_3_875  |      1.0094       .0088     114.31   0.000     .9921    1.0267
FICO_350_660|      1.0150       .0116      87.14   0.000     .9922    1.0379
FICO_661_720|       .5005       .0123      40.67   0.000     .4763     .5246
MH          |      2.0247       .0134     151.58   0.000    1.9985    2.0509
Minority    |       .0016       .0093       0.18   0.860    -.0165     .0198
_cons       |      -.0247       .0113      -2.19   0.028    -.0468    -.0026
---------------------------------------------------------------------------
```

DGP3: Correlations Between Race and Policy Factors

3.1 Summary statistics for all variables and loans

Variable	Obs	Mean	SD	Min	Max
Minority	500,000	.250	.433	0	1
FICO	500,000	689.540	52.185	393	903
FICO_350_660	500,000	.271	.444	0	1
FICO_661_720	500,000	.451	.498	0	1
MH	500,000	.063	.242	0	1
Rate	500,000	4.013	.163	3.875	4.375
Rate_4_375	500,000	.151	.358	0	1
Rate_4	500,000	.499	.500	0	1
Rate_3_875	500,000	.350	.477	0	1
Points	500,000	.821	.932	-4.061	6.043
Error	500,000	.010	.100	0	1

3.1 Summary statistics for all variables for loans to minority applicants

Variable	Obs	Mean	SD	Min	Max
Minority	125,155	1	0	0	1
FICO	125,155	659.360	60.016	393	903
FICO_350_660	125,155	.507	.500	0	1
FICO_661_720	125,155	.339	.474	0	1
MH	125,155	.100	.300	0	1
Rate	125,155	4.013	.163	3.875	4.375
Rate_4_375	125,155	.151	.358	0	1
Rate_4	125,155	.499	.500	0	1
Rate_3_875	125,155	.351	.477	0	1
Points	125,155	1.077	.981	-2.852	5.768
Error	125,155	.010	.101	0	1

3.1 Summary statistics for all variables for loans to non-minority applicants

Variable	Obs	Mean	SD	Min	Max
Minority	374,155	0	0	0	1
FICO	374,155	699.62	44.992	496	901
FICO_350_660	374,155	.192	.394	0	1
FICO_661_720	374,155	.488	.500	0	1
MH	374,155	.050	.218	0	1
Rate	374,155	4.013	.163	3.875	4.375
Rate_4_375	374,155	.151	.358	0	1
Rate_4	374,155	.499	.500	0	1
Rate_3_875	374,155	.350	.477	0	1
Points	374,155	.736	.887	-4.061	6.043
Error	374,155	.010	.100	0	1

3.3 Regression: points charged = f(minority)

```
Source       |      SS           df           MS          Number of obs=  500,000
----------------------------------------------------- F(1, 499998)= 13178.050
Model        |      10941.006          1  10941.006       Prob > F =         0.000
Residual     |     415120.526   499,998       .830       R-squared =        0.026
----------------------------------------------------- Adj R-squared=    0.026
Total        |     426061.531   499,999       .852       Root MSE =         0.911
-----------------------------------------------------------------------------
Points       | Coefficient   Std. Err.      t     P>|t| [95% conf. Interval]
-------------------------------------------------------------------------------
Minority     |      .342        .003       114.80  0.000     .338     .347
_cons        |      .736        .002       494.41  0.000     .733     .739
-------------------------------------------------------------------------------
```

3.5 Regression: Points = f(rate_3_875, rate_4_375, fico_350_660, fico_661_720, MH, minority)

```
Source       |      SS           df           MS          Number of obs=  500,000
----------------------------------------------------- F(6, 499993)= 99999.000
Model        |    420968.731          6   70161.455      Prob > F =         0.000
Residual     |      5092.800    499,993        .010      R-squared =        0.988
----------------------------------------------------- Adj R-squared=    0.988
Total        |    426061.531    499,999        .852      Root MSE =         0.101
-----------------------------------------------------------------------------
Points       | Coefficient    Std. Err.     t      P>|t| [95% conf. Interval]
-------------------------------------------------------------------------------
Rate_4_375   |      -.9999       .0004   -2382.92   0.000  -1.0007    -.9990
Rate_3_875   |      1.0004       .0003    3180.07   0.000    .9998    1.0010
FICO_350_660|        .9998       .0004    2487.57   0.000    .9990    1.0006
FICO_661_720|        .4997       .0003    1450.85   0.000    .4991     .5004
MH           |      2.0000       .0006    3382.82   0.000   1.9988    2.0012
Minority     |      -.0001       .0003      -0.31   0.755   -.0008     .0006
_cons        |       .0000       .0003       0.06   0.952   -.0006     .0006
-------------------------------------------------------------------------------
```

DGP4.1: Non-Discriminatory Exceptions, No Correlations

4.1.1 Summary statistics for all variables

Variable	Obs	Mean	SD	Min	Max
Minority	500,000	.251	.434	0	1
FICO	500,000	659.541	59.940	390	942
FICO_350_660	500,000	.506	.500	0	1
FICO_661_720	500,000	.339	.473	0	1
MH	500,000	.100	.300	0	1
Rate	500,000	4.012	.163	3.875	4.375
Rate_4_375	500,000	.150	.357	0	1
Rate_4	500,000	.499	.500	0	1
Rate_3_875	500,000	.351	.477	0	1
Points	500,000	.826	.991	-3.849	6.534
Error	500,000	.010	.100	0	1
PE	500,000	.250	.144	0	0.500

4.1.3 Regression: points charged = f(minority status)

Source	SS	df	MS		
				Number of obs=	500,000
				F(1, 499998) =	0.500
Model	.495	1	.495	Prob > F =	0.478
Residual	491188.691	499,998	.982	R-squared =	0.000
				Adj R-squared=	-0.000
Total	491189.185	499,999	.982	Root MSE =	0.991

Points	Coefficient	Std. Err.	t	P>|t|	[95% conf.	Interval]
Minority	-.002	.003	-0.71	0.478	-.009	.004
_cons	.826	.002	510.19	0.000	.823	.830

4.1.5 Regression: points charged = f(rate_3_875, rate_4_375, fico_350_660, fico_661_720, mh, minority)

Source	SS	df	MS		
				Number of obs=	500,000
				F(6, 499993)=	99999.000
Model	475779.163	6	79296.527	Prob > F =	0.000
Residual	15410.022	499,993	.031	R-squared =	0.969
				Adj R-squared=	0.969
Total	491189.185	499,999	.982	Root MSE =	0.176

Points	Coefficient	Std. Err.	t	P>|t|	[95% conf.	Interval]
Rate_4_375	-1.0012	.0007	-1368.82	0.000	-1.0026	-.9997
Rate_3_875	1.0001	.0005	1828.64	0.000	.9991	1.0012
FICO_350_660	1.0007	.0007	1388.05	0.000	.9992	1.0021
FICO_661_720	.5001	.0008	656.79	0.000	.4986	.5016
MH	2.0000	.0008	2412.32	0.000	1.9984	2.0016
Minority	.0004	.0006	0.65	0.516	-.0008	.0015
_cons	-.2505	.0007	-358.67	0.000	-.2519	-.2491

4.1.9 Regression: points charged = f(rate_3_875, rate_4_375, fico_350_660, fico_661_720, mh, PE, minority)

Source	SS	df	MS		
				Number of obs=	500,000
				F(7, 499992)=	99999.000
Model	486209.436	7	69458.491	Prob > F =	0.000
Residual	4979.750	499,992	.010	R-squared =	0.990
				Adj R-squared=	0.990
Total	491189.185	499,999	.982	Root MSE =	0.100

Points	Coefficient	Std. Err.	t	P>|t|	[95% conf.	Interval]
Rate_4_375	-1.0000	.0004	-2405.15	0.000	-1.0008	-.9992
Rate_3_875	1.0001	.0003	3216.84	0.000	.9995	1.0007
FICO_350_660	1.0004	.0004	2441.23	0.000	.9996	1.0012
FICO_661_720	.5003	.0004	1155.72	0.000	.4994	.5011
MH	1.9999	.0005	4243.32	0.000	1.9989	2.0008
PE	-1.0016	.0010	-1023.35	0.000	-1.0035	-.9997
Minority	-.0003	.0003	-0.81	0.419	-.0009	.0004
_cons	.0003	.0005	0.55	0.580	-.0007	-.0012

4.1.10 T-test of difference of means of pricing exceptions amount

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% conf.	Interval]
0	374,488	.2505	.0002	.1442	.2500	.2510
1	125,512	.2499	.0004	.1442	.2491	.2507
Combined	500,000	.2503	.0002	.1442	.2499	.2507
diff		.0006	.0005		-.0003	.0016

```
Diff = mean(0) - mean(1)                        t=1.3465
H0: diff = 0                                    degrees of freedom = 499998
Ha: diff < 0            Ha: diff != 0           Ha: diff > 0
Pr(T < t) = 0.9109      Pr(|T| > |t|) = 0.1781  Pr (T > t) = 0.0891
```

DGP4.2: Non-Discriminatory Exceptions, Correlations

4.2.1 Summary statistics for all variables

Variable	Obs	Mean	SD	Min	Max
Minority	500,000	.250	.433	0	1
FICO	500,000	689.368	52.173	412	934
FICO_350_660	500,000	.273	.445	0	1
FICO_661_720	500,000	.449	.497	0	1
MH	500,000	.063	.243	0	1
Rate	500,000	4.013	.163	3.875	4.375
Rate_4_375	500,000	.150	.358	0	1
Rate_4	500,000	.500	.500	0	1
Rate_3_875	500,000	.349	.477	0	1
Points	500,000	.571	.936	-3.994	5.326
Error	500,000	.010	.100	0	1
PE	500,000	.250	.144	0	0.500

4.2.3 Regression: points charged = f(minority status)

```
Source      |     SS          df          MS          Number of obs=  500,000
--------------------------------------------------- F(1, 499998)= 12446.450
Model       |    10635.028         1  10635.028       Prob > F =        0.000
Residual    |   427229.822   499,998       .855       R-squared =       0.024
--------------------------------------------------- Adj R-squared=    0.024
Total       |   437864.850   499,999       .876       Root MSE =        0.924
--------------------------------------------------------------------------
Points      | Coefficient   Std. Err.     t      P>|t| [95% conf. Interval]
--------------------------------------------------------------------------
Minority    |     .337        .003      111.56   0.000      .331    .343
_cons       |     .487        .002      322.93   0.000      .484    .490
--------------------------------------------------------------------------
```

4.2.5 Regression: points charged = f(rate_3_875, rate_4_375, fico_350_660, fico_661_720, mh, minority)

```
Source      |     SS          df          MS          Number of obs=  500,000
--------------------------------------------------- F(6, 499993)= 99999.000
Model       |   422566.350         6   70427.725      Prob > F =        0.000
Residual    |    15298.500   499,993        .031      R-squared =       0.965
--------------------------------------------------- Adj R-squared=    0.965
Total       |   437864.850   499,999        .876      Root MSE =        0.175
-------------------------------------------------------------------------
Points      | Coefficient    Std. Err.     t       P>|t|[95% conf. Interval]
--------------------------------------------------------------------------
Rate_4_375  |     -.9996       .0007   -1374.38   0.000  -1.0011    -.9982
Rate_3_875  |     1.0004       .0005    1834.03   0.000    .9993    1.0015
FICO_350_660|     1.0009       .0007    1441.42   0.000    .9996    1.0023
FICO_661_720|      .5007       .0006     838.23   0.000    .4996     .5019
MH          |     2.0011       .0010    1954.58   0.000   1.9991    2.0031
Minority    |     -.0008       .0006      -1.35   0.176   -.0020     .0004
_cons       |     -.2507       .0005    -465.18   0.000   -.2518    -.2496
--------------------------------------------------------------------------
```

4.2.9 Regression: points charged = f(rate_3_875, rate_4_375, fico_350_660, fico_661_720, mh, PE, minority)

Source	SS	df	MS		
				F(7, 499992)=	99999.000
Model	432941.419	7	61848.774	Prob > F =	0.000
Residual	4923.431	499,992	.010	R-squared =	0.989
				Adj R-squared=	0.989
Total	437864.850	499,999	.876	Root MSE =	0.099

Number of obs= 500,000

Points	Coefficient	Std. Err.	t	P>|t|	[95% conf.	Interval]
Rate_4_375	-.9999	.0004	-2423.15	0.000	-1.0006	-.9990
Rate_3_875	.9999	.0003	3231.50	0.000	.9993	1.0006
FICO_350_660	1.0011	.0004	2541.25	0.000	1.0003	1.0018
FICO_661_720	.5011	.0003	1478.57	0.000	.5004	.5017
MH	1.9998	.0006	3443.26	0.000	1.9987	2.0010
PE	-.9981	.0010	-1026.46	0.000	-1.0000	-.9962
Minority	-.0003	.0003	-0.73	0.463	-.0009	.0004
_cons	-.0011	.0004	-2.80	0.005	-.0019	-.0003

4.2.10 T-test of difference of means of pricing exceptions amount

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% conf.	Interval]
0	375,236	.2500	.0002	.1444	.2496	.2505
1	124,764	.2505	.0004	.1442	.2497	.2513
Combined	500,000	.2502	.0002	.1443	.2498	.2506
diff		-.0005	.0005		-.0001	.0004

Diff = mean(0) - mean(1) t=-1.0636

H0: diff = 0 degrees of freedom = 499998

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0

Pr(T < t) = 0.1438 Pr(|T| > |t|) = 0.2875 Pr (T > t) = 0.8562

DGP5.1: Discriminatory Exceptions, No Correlations

5.1.1 Summary statistics for all variables

Variable	Obs	Mean	SD	Min	Max
Minority	500,000	.250	.433	0	1
FICO	500,000	659.667	60.050	362	937
FICO_350_660	500,000	.506	.500	0	1
FICO_661_720	500,000	.338	.473	0	1
MH	500,000	.100	.301	0	1
Rate	500,000	4.012	.162	3.875	4.375
Rate_4_375	500,000	.149	.356	0	1
Rate_4	500,000	.500	.500	0	1
Rate_3_875	500,000	.351	.477	0	1
Points	500,000	.890	.998	-4.209	6.380
Error	500,000	.010	.100	0	1
PE	500,000	.187	.165	0	0.500

5.1.3 Regression: points charged = f(minority status)

Source	SS	df	MS		
				Number of obs=	500,000
				F(1, 499998)=	6109.010
Model	6006.936	1	6006.936	Prob > F =	0.000
Residual	491644.046	499,998	.983	R-squared =	0.012
				Adj R-squared=	0.012
Total	497650.982	499,999	.995	Root MSE =	0.992

Points	Coefficient	Std. Err.	t	P>|t|	[95% conf.	Interval]
Minority	.253	.003	78.16	0.000	.247	.260
_cons	.827	.002	510.44	0.000	.823	.830

5.1.5 Regression: points charged = f(rate_3_875, rate_4_375, fico_350_660, fico_661_720, mh, minority)

Source	SS	df	MS		
				Number of obs=	500,000
				F(6, 499993)=	99999.000
Model	484625.438	6	80770.906	Prob > F =	0.000
Residual	13025.544	499,993	.026	R-squared =	0.974
				Adj R-squared=	0.974
Total	497650.982	499,999	.995	Root MSE =	0.161

Points	Coefficient	Std. Err.	t	P>|t|	[95% conf.	Interval]
Rate_4_375	-1.0001	.0007	-1484.59	0.000	-1.0014	-.9987
Rate_3_875	.9997	.0005	1988.58	0.000	.9987	1.0007
FICO_350_660	.9990	.0007	1511.11	0.000	.9977	1.0003
FICO_661_720	.4989	.0007	714.06	0.000	.4975	.5003
MH	2.0005	.0008	2634.22	0.000	1.9990	2.0020
Minority	.2504	.0005	475.03	0.000	.2494	.2514
_cons	-.2493	.0006	-389.72	0.000	-.2506	-.2481

5.1.9 Regression: points charged = f(rate_3_875, rate_4_375, fico_350_660, fico_661_720, mh, PE, minority)

Source	SS	df	MS		
Model	492464.784	7	70352.112	Number of obs=	500,000
Residual	5186.199	499,992	.010	F(7, 499992)=	99999.000
Total	497650.982	499,999	.995	Prob > F =	0.000
				R-squared =	0.990
				Adj R-squared=	0.990
				Root MSE =	0.102

Points	Coefficient	Std. Err.	t	P>|t|	[95% conf.	Interval]
Rate_4_375	-.9998	.0004	-2352.21	0.000	-1.0006	-.9990
Rate_3_875	.9999	.0003	3152.05	0.000	.9993	1.0005
FICO_350_660	.9993	.0004	2395.65	0.000	.9985	1.0001
FICO_661_720	.4993	.0004	1132.59	0.000	.4985	.5002
MH	2.0000	.0005	4173.61	0.000	1.9990	2.0009
PE	-1.0016	.0012	-869.35	0.000	-1.0039	-.9994
Minority	.0000	.0004	0.03	0.975	-.0008	.0009
_cons	.0007	.0005	1.46	0.145	-.0002	.0017

5.1.10 T-test of difference of means of pricing exceptions amount

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% conf.	Interval]
0	375,004	.2500	.0002	.1444	.2495	.2505
1	124,996	0	0	0	0	0
Combined	500,000	.1875	.0002	.1654	.1870	.1880
diff		.2500	.0004		.2492	.2508

```
Diff = mean(0) - mean(1)                          t=612.2926
H0: diff = 0                                      degrees of freedom = 499998
Ha: diff < 0             Ha: diff != 0            Ha: diff > 0
Pr(T < t) = 1.0000       Pr(|T| > |t|) = 0.0000   Pr (T > t) = 0.0000
```

DGP5.2: Discriminatory Exceptions, Correlations

5.2.1 Summary statistics for all variables

Variable	Obs	Mean	SD	Min	Max
Minority	500,000	.251	.4335	0	1
FICO	500,000	689.345	52.1702	375	912
FICO_350_660	500,000	.272	.4450	0	1
FICO_661_720	500,000	.450	.4975	0	1
MH	500,000	.062	.2416	0	1
Rate	500,000	4.012	.1623	3.875	4.375
Rate_4_375	500,000	.149	.3564	0	1
Rate_4	500,000	.500	.5000	0	1
Rate_3_875	500,000	.351	.4772	0	1
Points	500,000	.636	.9534	-5.164	6.215
Error	500,000	.010	.1001	0	1
PE	500,000	.187	.1654	0	0.500

5.2.3 Regression: points charged = f(minority status)

Source	SS	df	MS		
				Number of obs=	500,000
				F(1, 499998)=	39195.500
Model	33038.625	1	33038.625	Prob > F =	0.000
Residual	421457.741	499,998	.843	R-squared =	0.073
				Adj R-squared=	0.073
Total	454496.366	499,999	.909	Root MSE =	0.918

Points	Coefficient	Std. Err.	t	P>|t|	[95% conf.	Interval]
Minority	.593	.003	197.98	0.000	.587	.599
_cons	.487	.002	324.49	0.000	.484	.490

5.2.5 Regression: points charged = f(rate_3_875, rate_4_375, fico_350_660, fico_661_720, mh, minority)

Source	SS	df	MS		
				Number of obs=	500,000
				F(6, 499993)=	99999.000
Model	441800.006	6	73633.334	Prob > F =	0.000
Residual	12696.360	499,993	.025	R-squared =	0.972
				Adj R-squared=	0.972
Total	454496.366	499,999	.909	Root MSE =	0.159

Points	Coefficient	Std. Err.	t	P>|t|	[95% conf.	Interval]
Rate_4_375	-1.0009	.0007	-1505.99	0.000	-1.0022	-.9996
Rate_3_875	.9998	.0005	2014.20	0.000	.9989	1.0008
FICO_350_660	1.0004	.0006	1577.24	0.000	.9991	1.0020
FICO_661_720	.4996	.0005	917.85	0.000	.4986	.5007
MH	2.0011	.0009	2136.88	0.000	1.9992	2.0029
Minority	.2494	.0005	454.34	0.000	.2484	.2505
_cons	-.2496	.0005	-507.93	0.000	-.2506	-.2487

5.2.9 Regression: points charged = f(rate_3_875, rate_4_375, fico_350_660, fico_661_720, mh, PE, minority)

```
Source      |     SS          df          MS          Number of obs=  500,000
----------------------------------------------------- F(7, 499992)= 99999.000
Model       |   449620.981          7   64231.569     Prob > F =        0.000
Residual    |     4875.385    499,992        .010     R-squared =       0.989
----------------------------------------------------- Adj R-squared=    0.989
Total       |   454496.366    499,999        .909     Root MSE =        0.099
-------------------------------------------------------------------------
Points      | Coefficient    Std. Err.     t      P>|t|  [95% conf. Interval]
---------------------------------------------------------------------------
Rate_4_375  |    -1.0008       .0004    -2430.08  0.000   -1.0016   -1.0000
Rate_3_875  |     1.0000       .0003     3250.86  0.000     .9994    1.0006
FICO_350_660|     1.0002       .0004     2544.95  0.000     .9995    1.0010
FICO_661_720|      .5000       .0003     1482.36  0.000     .4994     .5007
MH          |     2.0005       .0006     3447.39  0.000    1.9994    2.0016
PE          |    -1.0007       .0011     -895.59  0.000   -1.0029    -.9985
Minority    |     -.0005       .0004       -1.11  0.265    -.0014     .0004
_cons       |      .0002       .0004        0.52  0.605    -.0006     .0010
---------------------------------------------------------------------------
```

5.2.10 T-test of difference of means of pricing exceptions amount

```
Two-sample t test with equal variances
----------------------------------------------------------------------------
Group |     Obs          Mean   Std. Err.  Std. Dev.   [95% conf. Interval]
----------------------------------------------------------------------------
0     |     374,534     .2499    .0002      .1444       .2494       .2503
1     |     125,466       0        0          0           0           0
----------------------------------------------------------------------------
Combined    500,000     .1872    .0002      .1654       .1867       .1876
----------------------------------------------------------------------------
diff  |                 .2499    .0004                  .2491       .2507
----------------------------------------------------------------------------
Diff = mean(0) - mean(1)                         t=612.9074
H0: diff = 0                                     degrees of freedom = 499998
Ha: diff < 0            Ha: diff != 0            Ha: diff > 0
Pr(T < t) = 1.0000      Pr(|T| > |t|) = 0.0000   Pr (T > t) = 0.0000
```

Appendix B: STATA Code

Base DGP

```
clear
set obs 500

*** Generate simulated data for applicant characteristics ***

gen minority = uniform() < .25

gen fico = int(660 + (60*rnormal()))
replace fico = 350 if fico < 350
gen fico_350_660 = (fico>=350 & fico<=660)
gen fico_661_720 = (fico>=661 & fico<=720)

gen mh = uniform() < .1

gen temp_cat = uniform()
gen rate = 4.375 if temp_cat < .15
replace rate = 4 if temp_cat >= .15 & temp_cat <= .65
replace rate = 3.875 if temp_cat > .65

gen rate_4_375 = (rate == 4.375)
gen rate_4 = (rate == 4)
gen rate_3_875 = (rate == 3.875)

*** Generate points charged ***

gen points = -1 if rate == 4.375
replace points = 0 if rate == 4
replace points = 1 if rate == 3.875

*** Make LLPA adjustments to points charged ***

replace points = points + 1 if fico_350_660 == 1 & mh ==0
replace points = points + .5 if fico_661_720 == 1 & mh == 0
replace points = points + 3 if fico_350_660 == 1 & mh ==1
replace points = points + 2.5 if fico_661_720 == 1 & mh == 1
replace points = points + 2 if fico_350_660 == 0 & fico_661_720 == 0 & mh == 1
```

```
*** Generate statistical results ***

summarize
ttest points, by(minority)
regress points minority
regress points rate_4_375 rate_3_875 fico_350_660 fico_661_720 mh
regress points rate_4_375 rate_3_875 fico_350_660 fico_661_720 mh minority
regress points rate fico_350_660 fico_661_720 mh
regress points rate fico_350_660 fico_661_720 mh minonity
```

DGP1: Errors in Pricing Loans

```
clear
set obs 500

*** Generate simulated data for applicant characteristics ***

gen minority = uniform() < .25

gen fico = int (660 + (60*rnormal()))
replace fico = 350 if fico < 350
gen fico_350_660 = (fico>=350 & fico<=660)
gen fico_661_720 = (fico>=661 & fico<=720)

gen mh = uniform() < .1

gen temp_cat = uniform()
gen rate = 4.375 if temp_cat < .15
replace rate = 4 if temp_cat >= 15 & temp_cat <= .65
replace rate = 3.875 if temp_cat > .65

gen rate_4_375 = (rate == 4.375)
gen rate_4 = (rate == 4)
gen rate_3_875 = (rate == 3.875)

*** Generate points charged ***

gen points = -1 if rate == 4.375
replace points = 0 if rate == 4
replace points = 1 if rate == 3.875

*** Make LLPA adjustments to points charged ***

replace points = points + 1 if fico_350_660 == 1 & mh ==0
replace points = points + .5 if fico_661_720 == 1 & mh == 0
replace points = points + 3 if fico_350_660 == 1 & mh ==1
replace points = points + 2.5 if fico_661_720 == 1 & mh == 1
replace points = points + 2 if fico_350_660 == 0 & fico_661_720 == 0 & mh == 1

*** Add in pricing errors ***

gen pricing_error1 = (uniform() >.99)
gen points1 = points
replace points1 = pointsl + (1 * rnormal()) if pricing_error1 == 1
gen points2 = points
```

```
replace points2 = points2 + (4* rnormal()) if pricing_error1 == 1

gen pricing_error2 = (uniform() > .50)
gen points3 = points
replace points3 = points3 + (1 * rnormal()) if pricing_error2 == 1
gen points4 = points
replace points4 = points4 + (4 * rnormal()) if pricing_error2 == 1

*** Generate statistical results ***

summarize
regress points1 rate_4_375 rate_3_875 fico_350_660 fico_661_720 mh minority
regress points2 rate_4_375 rate_3_875 fico_350_660 fico_661_720 mh minority
regress points3 rate_4_375 rate_3_875 fico_350_660 fico_661_720 mh minority
regress points4 rate_4_375 rate_3_875 fico_350_660 fico_661_720 mh minority
```

DGP2: Sample Size

```
clear
set obs 500000

*** Generate simulated data for applicant characteristics ***

gen minority = uniform() < .25

gen fico = int 660 + (60*rnormal()))
replace fico = 350 if fico < 350
gen fico_350_660 = (fico>=350 & fico<=660)
gen fico_661_720 = (fico>=661 & fico<=720)

gen mh = uniform() < .1

gen temp_cat = uniform()
gen rate = 4.375 if temp_cat < .15
replace rate = 4 if temp_cat >= .15 & temp_cat <= .65
replace rate = 3.875 if temp_cat > .65

gen rate_4_375 = (rate == 4.375)
gen rate_4 = (rate == 4)
gen rate_3_875 = (rate == 3.875)

*** Generate points charged ***

gen points = -1 if rate == 4.375
replace points = 0 if rate == 4
replace points = 1 if rate == 3.875

*** Make LLPA adjustments to points charged ***

replace points = points + 1 if fico_350_660 == 1 & mh ==0
replace points = points + .5 if fico_661_720 == 1 & mh == 0
replace points = points + 3 if fico_350_660 == 1 & mh ==1
replace points = points + 2.5 if fico_661_720 == 1 & mh == 1
replace points = points + 2 if fico_350_660 == 0 & fico_661_720 == 0 & mh == 1

*** Add in pricing errors ***

gen pricing_errorl = (uniform() >.99)
gen points1 = points
replace points1 = pointsl + (1 * rnormal()) if pricing_error1 == 1
gen points2 = points
```

```
replace points2 = points2 + (4* rnormal()) if pricing_error1 ==1

gen pricing_error2 = (uniform() > .50)
gen points3 = points
replace points3 = points3 + (1 * rnormal()) if pricing_error2 == 1
gen points4 = points
replace points4 = points4 + (4 * rnormal()) if pricing_error2 == 1

*** Generate statistical results ***

summarize
regress points1 rate_4_375 rate_3_875 fico_350_660 fico_661_720 mh minority
regress points2 rate_4_375 rate_3_875 fico_350_660 fico_661_720 mh minority
regress points3 rate_4_375 rate_3_875 fico_350_660 fico_661_720 mh minority
regress points4 rate_4_375 rate_3_875 fico_350_660 fico_661_720 mh minority
```

DGP3: Correlations Between Race and Policy Factors

```
clear
set obs 500000

*** Generate simulated data for applicant characteristics ***

gen minority = uniform() < .25

gen fico = int(660 + (60*rnormal())) if minority == 1
replace fico = int(700 + (45*rnormal())) if minority == 0
replace fico = 350 if fico < 350
gen fico_350_660 = (fico >= 350 & fico <= 660)
gen fico_661_720= (fico >= 661 & fico <= 720)

gen mh = uniform() < .10 if minority == 1
replace mh = uniform() < .05 if minority == 0

gen temp_cat = uniform()
gen rate = 4.375 1f temp_cat < .15
replace rate = 4 if temp_cat >= .15 & temp_cat <= .65
replace rate = 3.875 if temp_cat > .65

gen rate_4_375 = (rate == 4.375)
gen rate_4 = (rate == 4)
gen rate_3_875 = (rate == 3.875)

*** Generate points charged ***

gen points = -1 if rate == 4.375
replace points = 0 if rate == 4
replace points = 1 if rate == 3.875

*** Make LLPA adjustments to points charged ***

replace points = points + 1 if fico_350_660 == 1 & mh ==0
replace points = points + .5 if fico_661_720 == 1 & mh == 0
replace points = points + 3 if fico_350_660 == 1 & mh ==1
replace points = points + 2.5 if fico_661_720 == 1 & mh == 1
replace points = points + 2 if fico_350_660 == 0 & fico_661_720 == 0 & mh == 1
```

```
*** Add in pricing errors ***

gen pricing_error = (uniform() >.99)
replace points = points + (1 * rnormal()) if pricing_error == 1

*** Generate statistical results ***

summarize
summarize if minority == 1
summarize if minority == 0

regress points minority
regress points rate_4_375 rate_3_875 fico_350_660 fico_661_720 mh minority
```

DGP4.1: Non-Discriminatory Exceptions, No Correlations

```
clear
set obs 500000

*** Generate simulated data for applicant characteristics ***

gen minority = (uniform() < .25)

gen fico = int(660 + (60*rnormal()))
replace fico = 350 if fico < 350
gen fico_350_660 = (fico>=350 & fico<=660)
gen fico_661_720 = (fico>=661 & fico<=720)

gen mh = uniform() < .1

gen temp_cat = uniform()
gen rate = 4.375 if temp_cat < .15
replace rate = 4 if temp_cat >= .15 & temp_cat <= .65
replace rate = 3.875 if temp_cat > .65

gen rate_4_375 = (rate == 4.375)
gen rate_4 = (rate == 4)
gen rate_3_875 = (rate == 3.875)

*** Generate points charged ***

gen points = -1 if rate == 4.375
replace points = 0 if rate == 4
replace points = 1 if rate == 3.875

*** Make LLPA adjustments to points charged ***

replace points = points + 1 if fico_350_660 == 1 & mh == 0
replace points = points + .5 if fico_661_720 == 1 & mh == 0
replace points = points + 3 if fico_350_660 == 1 & mh == 1
replace points = points + 2.5 if fico_661_720 == 1 & mh == 1
replace points = points + 2 if fico_350_660 == 0 & fico_661_720 == 0 & mh == 1

*** Add in pricing errors ***

gen pricing_error = (uniform() >.99)
replace points = points + (1 * rnormal()) if pricing_error == 1
```

```
*** Add in pricing exceptions ***

gen pe = uniform() / 2
replace points = points – pe

*** Generate statistical results ***

summarize

regress points minority
regress points rate_4_375 rate_3_875 fico_350_660 fico_661_720 mh minority
regress points rate_4_375 rate_3_875 fico_350_660 fico_661_720 mh pe minority
ttest pe, by(minority)
```

DGP4.2: Non-Discriminatory Exceptions, Correlations

```
clear
set obs 500000

*** Generate simulated data for applicant characteristics ***

gen minority = uniform() < .25

gen fico = int(660 + (60*rnormal())) if minority == 1
replace fico = int(700 + (45*rnormal())) if minority == 0
replace fico = 350 if fico < 350
gen fico_350_660 = (fico>=350 & fico<=660)
gen fico_661_720 = (fico>=661 & fico<=720)

gen mh = uniform() < .10 if minority == 1
replace mh = uniform() < .05 if minority == 0

gen temp_cat = uniform()
gen rate = 4.375 if temp_cat < .15
replace rate = 4 if temp_cat >= .15 & temp_cat <= .65
replace rate = 3.875 if temp_cat > .65

gen rate_4_375 = (rate == 4.375)
gen rate_4 = (rate == 4)
gen rate_3_875 = (rate == 3.875)

*** Generate points charged ***

gen points = -1 if rate == 4.375
replace points = 0 if rate == 4
replace points = 1 if rate == 3.875

*** Make LLPA adjustments to points charged ***

replace points = points + 1 if fico_350_660 == 1 & mh == 0
replace points = points + .5 if fico_661_720 == 1 & mh == 0
replace points = points + 3 if fico_350_660 == 1 & mh == 1
replace points = points + 2.5 if fico_661_720 == 1 & mh == 1
replace points = points + 2 if fico_350_660 == 0 & fico_661_720 == 0 & mh == 1

*** Add in pricing errors ***

gen pricing_error  = (uniform() >.99)
replace points = points + (1 * rnormal()) if pricing_error == 1
```

```
*** Add in pricing exceptions ***

gen pe = uniform() / 2
replace points = points – pe

*** Generate statistical results ***

summarize

regress points minority
regress points rate_4_375 rate_3_875 fico_350_660 fico_661_720 mh minority
regress points rate_4_375 rate_3_875 fico_350_660 fico_661_720 mh pe minority
ttest pe, by(minority)
```

DGP5.1: Discriminatory Exceptions, No Correlations

```
clear
set obs 500000

*** Generate simulated data for applicant characteristics ***

gen minority = (uniform() < .25)

gen fico = int(660 + (60*rnormal()))
replace fico = 350 if fico < 350
gen fico_350_660 = (fico>=350 & fico<=660)
gen fico_661_720 = (fico>=661 & fico<=720)

gen mh = uniform() < .1

gen temp_cat = uniform()
gen rate = 4.375 if temp_cat < .15
replace rate = 4 if temp_cat >= .15 & temp_cat <= .65
replace rate = 3.875 if temp_cat > .65

gen rate_4_375 = (rate == 4.375)
gen rate_4 = (rate == 4)
gen rate_3_875 = (rate == 3.875)

*** Generate points charged ***

gen points = -1 if rate == 4.375
replace points = 0 if rate == 4
replace points = 1 if rate == 3.875

*** Make LLPA adjustments to points charged ***

replace points = points + 1 if fico_350_660 == 1 & mh == 0
replace points = points + .5 if fico_661_720 == 1 & mh == 0
replace points = points + 3 if fico_350_660 == 1 & mh == 1
replace points = points + 2.5 if fico_661_720 == 1 & mh == 1
replace points = points + 2 if fico_350_660 == 0 & fico_661_720 == 0 & mh == 1

*** Add in pricing errors ***

gen pricing_error = (uniform() >.99)
replace points = points + (1 * rnormal()) if pricing_error == 1
```

```
*** Add in pricing exceptions ***

gen pe = uniform() / 2 if minority == 0
replace pe = 0 if minority == 1

replace points = points – pe

*** Generate statistical results ***

summarize

regress points minority
regress points rate_4_375 rate_3_875 fico_350_660 fico_661_720 mh minority
regress points rate_4_375 rate_3_875 fico_350_660 fico_661_720 mh pe minority
ttest pe, by(minority)
```

DGP5.2: Discriminatory Exceptions, Correlations

```
clear
set obs 500000

*** Generate simulated data for applicant characteristics ***

gen minority = uniform() < .25

gen fico = int(660 + (60*rnormal())) if minority == 1
replace fico = int(700 + (45*rnormal())) if minority == 0
replace fico = 350 if fico < 350
gen fico_350_660 = (fico>=350 & fico<=660)
gen fico_661_720 = (fico>=661 & fico<=720)

gen mh = uniform() < .10 if minority ==1
replace mh = uniform() < .05 if minority == 0

gen temp_cat = uniform()
gen rate = 4.375 if temp_cat < .15
replace rate = 4 if temp_cat >= .15 & temp_cat <= .65
replace rate = 3.875 if temp_cat > .65

gen rate_4_375 = (rate == 4.375)
gen rate_4 = (rate == 4)
gen rate_3_875 = (rate == 3.875)

*** Generate points charged ***

gen points = -1 if rate == 4.375
replace points = 0 if rate == 4
replace points = 1 if rate == 3.875

*** Make LLPA adjustments to points charged ***

replace points = points + 1 if fico_350_660 == 1 & mh == 0
replace points = points + .5 if fico_661_720 == 1 & mh == 0
replace points = points + 3 if fico_350_660 == 1 & mh == 1
replace points = points + 2.5 if fico_661_720 == 1 & mh == 1
replace points = points + 2 if fico_350_660 == 0 & fico_661_720 == 0 & mh == 1

*** Add in pricing errors ***

gen pricing_error = (uniform() >.99)
replace points = points + (1 * rnormal()) if pricing_error == 1
```

```
*** Add in pricing exceptions ***

gen pe = uniform() / 2 if minority == 0
replace pe = 0 if minority == 1

replace points = points – pe

*** Generate statistical results ***

summarize

regress points minority
regress points rate_4_375 rate_3_875 fico_350_660 fico_661_720 mh minority
regress points rate_4_375 rate_3_875 fico_350_660 fico_661_720 mh pe minority
ttest pe, by(minority)
```