Chapter 3

Gender Classification for Fair Lending Analyses

Jason Dietrich
March 2026

I. Introduction

For any fair lending analysis of gender disparities, applications first need to be classified into gender groups. There is no legal definition of the appropriate classification strategy for fair lending analyses, so the choice of strategy becomes an empirical issue. For analyses of mortgages, Home Mortgage Disclosure Act (HMDA) data is the primary source of data on gender. HMDA data contains two gender variables, one for the primary applicant and one for the co-applicant. Using these two variables, there are several possible strategies for classifying applications into gender groups for analysis. However, there is no one clear theoretically optimal approach, so decisions on which strategy to use need to consider other factors, such as empirical patterns and legal perspectives.

This report explores some of the issues Economists need to consider when choosing a specific classification strategy, as well as the tradeoffs of these choices. We begin by characterizing the gender variables available in the HMDA data, paying particular attention to applications with multiple genders reported, since those applications are the most difficult to classify. We then discuss five general issues that need to be considered when using HMDA data to classify applications into gender groups. Finally, we analyze how volumes, underwriting disparities, and pricing disparities vary across seven different classification strategies. For all analyses we use 2021 HMDA data and focus on applications for 1st lien, owner-occupied, conventional, 1-4 family, closed-end home purchase loans that were originated, approved but NA, or denied. To make the sample more homogenous we also exclude all applications for reverse mortgages and commercial purpose mortgages.[1] These filters resulted in 3,376,935 total

[1] The sample also excludes all applications for which Financial Institutions were not required to report gender data per the Economic Growth, Regulatory Relief, and Consumer Protection Act (EGRRCPA).

applications for the underwriting analysis and 3,052,612 total loans for the pricing analysis. The overall objective of this work is to understand the tradeoffs across classification strategies to inform specific choices on strategies for fair lending analyses.

There are four main takeaways from this report,

- Classifying applications into gender groups for fair lending analyses of mortgages is challenging given the extensive gender data available in HMDA.
- A legal definition of the appropriate classification strategy to use for fair lending analyses does not exist, so the choice of strategy is an empirical one.
- There are several possible classification strategies that can be used, and the quantitative and qualitative results from fair lending analyses can vary across classification strategies.
- When deciding on which classification strategy to use, it is important to consider how to treat primary applicants versus co-applicants, same-sex joint applications, single-applicant applications versus joint applications, applicants who report both male and female, and joint applications that report gender for one applicant, but not the other.

II. General Classification Strategy Considerations

HMDA data contain two gender variables, one for the primary applicant and one for the co-applicant. These variables can take on the following values:[2]

Primary Applicant

1. Male
2. Female
3. Information not provided by applicant in mail, internet, or telephone application
4. Not applicable
5. Not a relevant value
6. Applicant selected both male and female

[2] See 2024 FIG (Filing Instructions Guide) | HMDA Documentation for details about HMDA's reporting requirements for these gender variables.

Co-Applicant

1. Male
2. Female
3. Information not provided by applicant in mail, internet, or telephone application
4. Not applicable
5. No co-applicant
6. Co-applicant selected both male and female

Table 1 presents summary statistics of key lending variables available in HMDA data (denial rate, rate spread, FICO score, combined loan-to-value (CLTV), debt-to-income (DTI), and loan amount) for all possible combinations of values for primary applicant and co-applicant gender.[3] As the table shows, single-applicant applications where the applicant reports male, single-applicant applications where the applicant reports female, and joint applications where the primary applicant is male and the co-applicant is female and vice versa show the largest volumes by far. The first two of these groups are easy to classify since they contain only one gender, while the last two groups are more challenging to classify since they contain two genders. We discuss these challenges, as well as other interesting results in Table 1, below.

With the two gender variables in HMDA, there are several possible classification strategies available. When choosing a specific classification strategy to use for fair lending analyses, there are five general issues to consider: 1) should the gender of the primary applicant be given different weight than the gender of the co-applicant, 2) how should same-sex joint applications be classified, 3) should single-applicant applications be combined with joint applications for analysis, 4) how should applicants who report "6. selected both male and

[3] For Table 1, the number of applications reflects the number of applications for action taken, which is never missing. The number of applications used to construct the summary statistics for each of the variables might be lower due to non-applicable scenarios and outlier values that were changed to missing. As one example, the sample sizes for rate spread medians is lower, since rate spread is not reported for denied applications.

female” be classified, and 5) how should joint applications that report gender for one applicant, but not the other, be classified? We address each of these considerations in turn.

Table 1: Summary Statistics of Key Lending Variables by Gender of Primary Applicant and Co-Applicant Reported

Gender Combination	# Applications*	Average Denial Rate	Median Rate Spread	Median FICO	Median CLTV	Median DTI	Median Loan Amount
Male, Male	56,824	8.48	0.28	744	85.00	38.10	341,993
Male, Female	891,700	5.26	0.18	766	80.00	34.93	358,650
Male, Indirect	16,514	9.57	0.20	756	80.00	36.33	370,500
Male, NA	240	6.67	-0.12	780	76.92	38.32	455,625
Male, No Co-Applicant	1,033,481	7.87	0.25	760	87.09	36.39	294,405
Male, Male/Female	522	7.85	0.17	760	80.00	37.00	368,880
Female, Male	394,523	5.71	0.23	757	84.97	34.99	333,600
Female, Female	50,286	8.18	0.32	741	87.23	38.90	300,000
Female, Indirect	9,672	10.59	0.25	747	85.00	37.17	329,625
Female, NA	118	4.24	0.05	777	80.00	35.72	401,999
Female, No Co-Applicant	723,982	8.04	0.27	761	88.00	38.00	247,000
Female, Male/Female	253	9.09	0.20	756	84.90	38.51	360,040
Indirect, Male	3,584	9.35	0.26	748	85.00	36.77	329,410
Indirect, Female	5,251	7.88	0.22	757	80.00	36.49	356,000
Indirect, Indirect	81,140	8.06	0.16	767	80.00	34.90	380,000
Indirect, NA	52	5.77	-0.31	773	75.00	36.76	2,483,750
Indirect, No Co-applicant	106,348	11.58	0.22	764	85.00	36.69	295,000
Indirect, Male/Female	30	6.67	0.16	766	80.00	32.39	323,700
NA, Male	269	3.35	-0.01	780	75.00	38.38	960,000
NA, Female	70	4.29	0.11	666	80.00	34.90	325,882
NA, Indirect	72	1.39	-0.27	778	74.82	39.00	2,706,598
NA, NA	39	7.69	-0.21	769	75.00	26.32	548,000
NA, No Co-Applicant	425	5.18	0.09	734	69.99	38.32	1,500,000
Male/Female, Male	314	9.87	0.21	755	80.00	36.11	320,500
Male/Female, Female	449	8.91	0.23	754	80.00	36.66	360,000
Male/Female, Indirect	43	16.28	0.18	770	80.00	34 93	415,625
Female/Maie, No Co-applicant	664	7.08	0.15	765	80.00	38.23	320,000
Male/Female, Male/Female	169	8.88	0.13	766	80.00	36.80	400,000

* A small number of applications had invalid values for gender, so the application volumes sum to slightly less than the total sample size of 3,376,935.

3.1 Primary Applicant vs Co-Applicant

When two consumers apply jointly for a mortgage, one must be designated as the primary applicant and the other as the co-applicant. Given the U.S.' long history as a patriarchal society, who is listed as the primary applicant could potentially affect treatment of the application. Reflective of this, Table 1 shows clear differences between joint applications where the primary applicant is male and the co-applicant is female (Male, Female) and joint applications where the primary applicant is female and the co-applicant is male (Female, Male). As compared to "Male, Female" joint applications, the denial rate and median rate spread for "Female, Male" joint applications are 0.45 percentage points (pps) and 5 basis points (bps) higher, respectively. Differences in applicant characteristics likely drive some of these differences as median FICO score is lower and median CLTV is higher for "Female, Male" joint applications. Alternatively, discrimination could potentially drive some of these differences as well. To better understand these impacts and disentangle the two possible effects, in the analysis below we include a classification strategy that classifies applications based solely on the gender of the primary applicant. More details on this approach are included below.

3.2 Same-Sex Joint Applications

In our sample of 3,376,935 applications, just under 57,000 are "Male, Male" joint applications, and just over 50,000 are "Female, Female" joint applications, or about 3.17 percent of all applications in total. Although it is very unlikely that all of these applications are from LGBTQ+ applicants,[4] some percentage likely are, so these applications are of particular interest due to the potentially higher risk of discrimination against LGBTQ+ applicants. This potential

[4] As on example, some of these joint applications are likely from a mother and daughter.

risk is borne out in Table 1, which shows that the denial rates for “Male, Male” and “Female, Female” joint applications are approximately 3 pps higher than for “Male, Female” and “Female, Male” joint applications. Median rate spreads are higher for same-gender joint applications as well. Some of these differences are likely driven by lower median FICO scores, and higher median CLTVs and DTIs for “Male, Male” and “Female, Female” joint applications. However, discrimination could potentially drive some of these differences as well. These differences could have important implications for fair lending analyses, and therefore need to be considered when choosing a classification strategy.

Although the volumes of these applications are not extremely large, especially at the lender level, some lenders do have sufficient numbers of these applications to conduct separate disparity analyses. Effectively and accurately capturing all of the nuances of these applications is complicated and beyond the scope of this report, so we do not conduct an in-depth analysis of the treatment of these applications here. However, we do provide some general evidence of the treatment of these applications by analyzing two versions of some classification strategies, one including these applications and one excluding them. This approach provides some indication of how these applications impact disparity estimates. Details on this approach are included below. Given that “Female, Female” joint applications are classified as female and “Male, Male” joint applications are classified as male for every classification strategy we analyze in this report, our expectation is that dropping these applications will have limited impact on disparity estimates since the volumes of these two subsets of applications are somewhat similar (50,000 to 57,000) and the disproportionately worse outcomes for each of these two subsets of joint applications might somewhat cancel each other out.

3.3 Single Applications vs Joint Applications

Differences between single-applicant applications and joint applications, specifically related to reporting of multiple genders, need to be considered when classifying applications into gender groups for fair lending analyses.[5] Classifying single-applicant applications is straightforward since there is only one applicant and therefore only one gender value. Classifying joint applications on the other hand is more complicated since there are multiple applicants, and potentially multiple gender values. If there are differences between these sets of applications, especially in the treatment of these sets of applications, this could impact fair lending analyses.

In our sample, just under 45% of applications are joint applications and approximately 85% of these joint applications reported two genders. Table 1 provides evidence of some of the underlying differences between single-applicant applications and joint applications. Single-applicant applications tend to have much higher denial rates and median rate spreads compared to joint applications. As one example, the denial rates for “Male, No Co-applicant” and “Female, No Co-applicant” applications are 7.87% and 8.04%, respectively. These compare to 5.26% and 5.71% for “Male, Female” and “Female, Male” joint applications, respectively. Applicant characteristics also differ between single-applicant applications and joint applications as well. As one example, the median DTIs for “Male, No Co-applicant” and “Female, No Co-applicant” applications are 36.39% and 38.00%, respectively. These compare to 34.93% and 34.99% for “Male, Female” and “Female, Male” joint applications, respectively. To better understand the impacts of these differences, in the analysis below we include a classification strategy that uses

[5] Single applications and joint applications often have systematically different financial and credit characteristics, and FIs often use different underwriting and pricing criteria for each group. These differences are completely independent of the impacts that differences between single and joint applications have for classifying applications into gender groups, but fair lending analyses need to address these differences as well.

only single-applicant applications, along with a variety of strategies that incorporate gender information from joint applications. More details on these approaches are included below.

3.4 Applicants who reported "6. Applicant selected both male and female"

HMDA allows lenders to report gender as "6. Applicant selected both female and male" for applicants who select both genders on the application. Among the 3,376,935 applications in our sample, 2,444 (0.07%) reported "6. Applicant selected both female and male" for gender for at least one applicant. Approximately 73 percent of these 2,444 applications are joint applications, and a majority of these included either one applicant who reported "male" or one applicant that reported "female."

These applications are of particular interest due to the higher likelihood that these applicants belong to LGBTQ+ groups, and the potentially higher risk of discrimination against applicants in these groups. This potential risk is borne out in Table 1, which shows that applications where at least one applicant reported "6. Applicant selected both male and female" (labeled in the Table 1 as "Male/Female") tend to have significantly higher denial rates, even though median FICO scores and CLTVs are generally similar to, and median DTIs only slightly higher than, the corresponding values for "Male, Female" and "Female, Male" joint applications. Unfortunately, the volumes of these applications are too small for separate statistical analyses here, especially at the lender level, which is typically the focus of fair lending analyses. Given the evidence in Table 1 that these applications have different characteristics and may have been treated differently, we exclude them from all analyses below to avoid their potential impact on the disparities for the other gender groupings we analyze in this report. In general, if these applications have sufficient volume for a given fair lending analysis, they should be analyzed.

3.5 Applications with Some Known and Some Unknown Gender

Just under 36,000 applications in our sample (1.06%) are joint applications that reported a gender for one of the applicants, but either "3. Information not provided by applicant in mail, internet, or telephone application" or "4. Not applicable" for the other applicant. Since the gender of the second applicant on these applications might be known by the lender, but is unknown to the Economist, it is unclear how to classify these applications into gender groups. Unlike for same-sex joint applications and for applications where "6. Applicant selected both male and female" was reported for at least one applicant, there is no clear and obvious potential fair lending risk for these applications.

The evidence from Table 1 for these applications is somewhat mixed. For example, the denial rate for the 269 joint applications that reported "NA" for the primary applicant and "Male" for the co-applicant was 3.35%, which was lower than for most other gender combinations. Alternatively, the denial rate for the 16,514 joint applications that reported "Male" for the primary applicant and "3. Information not provided by applicant in mail, internet, or telephone application" for the co-applicant was 9.57%, which was one of the highest denial rates reported across gender combinations. The results for the applicant characteristics are also similarly mixed. To further understand the potential impact of these applications, we conducted the disparity analysis discussed below including these applications for each specific classification strategy and then excluding them for each specific classification strategy. The disparity estimates are very similar regardless of whether these applications were included or excluded. Because of this evidence, a desire to retain sample size, and no clear historical fair lending risk for these applications, we keep these applications for all analyses below.

III. Classification Strategies

This section presents the classification strategies we analyze in this report. Following discussions from above on general classification strategy considerations, we include strategies to address issues related to primary applicants versus co-applicants (issue 3.1), same-sex joint applications (issue 3.2), and single-applicant applications versus joint applications (issue 3.3). For all analyses, we drop applications that report “6. Applicant selected both male and female” (issue 3.4) and keep all applications that report gender for one applicant, but not the other (issue 3.5). The seven specific classification strategies we analyze are:

Strategy 1: Single-applicant females vs single-applicant males

Strategy 2: Any females vs no females and at least one male

Strategy 3: Any females vs no females and at least one male, same-sex applications excluded

Strategy 4: Only females vs at least one male

Strategy 5: Only females vs at least one male, same-sex applications excluded

Strategy 6: Female as the primary applicant vs male as the primary applicant

Strategy 7: Female as the primary applicant vs male as the primary applicant, same-sex applications excluded

Tables 2a and 2b show how each of these seven strategies classify each of the unique combinations of primary applicant and co-applicant gender values in HMDA data. The "Mean" columns show the denial rates for applications with each gender combination. As an example for how to read this table, with classification strategy 1, which is single-applicant females vs

Table 2a: Composition of Classification Strategies and Denial Rates

	Strategy 1		Strategy 2		Strategy 3		Strategy 4	
	Gender	Mean	Gender	Mean	Gender	Mean	Gender	Mean
Male, Male	.	.	Male	8.48	.	.	Male	8.48
Male, Female	.	.	Female	5.26	Female	5.26	Male	5.26
Male, Indirect	.	.	Male	9.57	Male	9.57	Male	9.57
Male, NA	.	.	Male	6.67	Male	6.67	Male	6.67
Male, No Co-Applicant	Male	7.87	Male	7.87	Male	7.87	Male	7.87
Male, Male/Female	.	.	.	.	.	.	.	.
Female, Male	.	.	Female	5.71	Female	5.71	Male	5.71
Female, Female	.	.	Female	8.18	.	.	Female	8.18
Female, Indirect	.	.	Female	10.59	Female	10.59	Female	10.59
Female, NA	.	.	Female	4.24	Female	4.24	Female	4.24
Female, No Co-Applicant	Female	8.04	Female	8.04	Female	8.04	Female	8.04
Female, Male/Female	.	.	.	.	.	.	.	.
Indirect, Male	.	.	Male	9.35	Male	9.35	Male	9.35
Indirect, Female	.	.	Female	7.88	Female	7.88	Female	7.88
Indirect, Indirect	.	.	.	.	.	.	.	.
Indirect, NA	.	.	.	.	.	.	.	.
Indirect, No Co-Applicant	.	.	.	.	.	.	.	.
Indirect, Male/Female	.	.	.	.	.	.	.	.
NA, Male	.	.	Male	3.35	Male	3.35	Male	3.35
NA, Female	.	.	Female	4.29	Female	4.29	Female	4.29
NA, Indirect	.	.	.	.	.	.	.	.
NA, NA	.	.	.	.	.	.	.	.
NA, No Co-Applicant	.	.	.	.	.	.	.	.
Male/Female, Male	.	.	.	.	.	.	.	.
Male/Female, Female	.	.	.	.	.	.	.	.
Male/Female, Indirect	.	.	.	.	.	.	.	.
Male/Female, No Co-Applicant	.	.	.	.	.	.	.	.
Male/Female, Male/Female	.	.	.	.	.	.	.	.

Table 2b: Composition of Classification Strategies and Denial Rates, Continued

	Strategy 5		Strategy 6		Strategy 7		Not Assigned
	Gender	Mean	Gender	Mean	Gender	Mean	
Male, Male	.	.	Male	8.48	.	.	.
Male, Female	Male	5.26	Male	5.26	Male	5.26	.
Male, Indirect	Male	9.57	Male	9.57	Male	9.57	.
Male, NA	Male	6.67	Male	6.67	Male	6.67	.
Male, No Co-Applicant	Male	7.87	Male	7.87	Male	7.87	.
Male, Male/Female	.	.	.	.	.	.	7.85
Female, Male	Male	5.71	Female	5.71	Female	5.71	.
Female, Female	.	.	Female	8.18	.	.	.
Female, Indirect	Female	10.59	Female	10.59	Female	10.59	.
Female, NA	Female	4.24	Female	4.24	Female	4.24	.
Female, No Co-Applicant	Female	8.04	Female	8.04	Female	8.04	.
Female, Male/Female	.	.	.	.	.	.	9.09
Indirect, Male	Male	9.35	.	.	.	.	.
Indirect, Female	Female	7.88	.	.	.	.	.
Indirect, Indirect	.	.	.	.	.	.	8.06
Indirect, NA	.	.	.	.	.	.	5.77
Indirect, No Co-Applicant	.	.	.	.	.	.	11.58
Indirect, Male/Female	.	.	.	.	.	.	6.67
NA, Male	Male	3.35	.	.	.	.	.
NA, Female	Female	4.29	.	.	.	.	.
NA, Indirect	.	.	.	.	.	.	1.39
NA, NA	.	.	.	.	.	.	7.69
NA, No Co-Applicant	.	.	.	.	.	.	5.18
Male/Female, Male	.	.	.	.	.	.	9.87
Male/Female, Female	.	.	.	.	.	.	8.91
Male/Female, Indirect	.	.	.	.	.	.	16.28
Male/Female, No Co-Applicant	.	.	.	.	.	.	7.08
Male/Female, Male/Female	.	.	.	.	.	.	8.88

single-applicant males, single-applicant applications reporting male for the applicant (Male, No Co-Applicant) are coded as "Male," single-applicant applications reporting female for the applicant (Female, No Co-Applicant) are coded as "Female," and applications for all other gender combination values are excluded from the analysis. Overall, the results in these tables show significant variation in the subsets of applications included by each strategy, as well as considerable differences in how each strategy treats specific applications. Not surprisingly, strategy 1, which includes only single-applicant applications, is the least extensive strategy, while strategies 2 and 4 include the most sets of gender combinations. In addition, there is significant variation in denial rates across the combinations of gender values as well. This variation suggests that applicant characteristics, and possibly treatment of these applications, differ across combinations. These differences need to be considered when choosing a classification strategy for fair lending analyses.

IV. Analysis and Results

This section analyzes how the seven classification strategies presented in the last section impact results from fair lending analyses of underwriting (denial rates) and pricing (rate spread) disparities for the largest HMDA reporters in 2021. For the underwriting analysis we use the 200 HMDA reporters with the largest volumes of applications and for the pricing analysis we use the 200 HMDA reporters with the largest volume of originations, so the list of FIs for the two analyses is slightly different. Consistent with standard practice, all analyses are at the financial institution (FI) level. We begin by comparing volumes of applications and loans available for analysis across the different classification strategies. We define sufficient volume for the underwriting (pricing) analysis as at least 50 applications (loans) from the treatment group and at least 50 applications (loans) from the control group for the given FI. All-else-equal, a classification strategy is less useful if it is less likely to yield sufficient volumes of applications or loans for meaningful statistical analyses. We then analyze underwriting and pricing disparities by gender across classification strategies. To keep the analyses focused we filter the data to just applications for 1st lien, owner-occupied, conventional, 1-4 family, closed-end home purchase loans, excluding reverse mortgages and commercial purpose mortgages. For underwriting analyses, we further subset the data to applications that were originated, approved but NA, or denied, and for pricing analyses we further subset the data to applications that were originated.

Appendix A presents all results. The first set of results (Tables A1-A2) presents volume comparisons and includes two tables, one for underwriting and one for pricing. In these volume tables, each row represents a classification strategy. Strategy 1, which includes just single-applicant applications, is used as a benchmark of comparison for all other strategies. Strategy 1 provides a good benchmark, since it is the most conservative strategy resulting in the most homogenous gender groups. Presenting results relative to a benchmark provides evidence on how the disparity analysis conclusions differ across classification strategies. The tradeoff is that all results are benchmark-specific and will differ for other benchmarks. For each table, the first column shows the classification strategy, treatment group, and control group. Column 2 shows the number of FIs with sufficient volumes using both the strategy for that row and the benchmark strategy. Column 3 shows the number of FIs with sufficient volume for only the strategy for that row, and column 4 shows the number of FIs with sufficient volume for only the benchmark strategy. The remaining columns in these tables present average, minimum, and maximum sample sizes for both the treatment group and the control group. Only FIs with sufficient volumes for both the strategy for the given row and the benchmark strategy are used to generate these summary statistics.

As an example for how to read these volume tables, for Table Al, which contains volumes for underwriting analyses, row 5 shows that 199 of the 200 FIs we analyze had sufficient volumes for analysis for both classification strategy 5 and strategy 1 (the benchmark strategy). No FIs had sufficient volume for only strategy 5 and no FIs had sufficient volume for only the benchmark strategy. These counts do not sum to 200, because one FI did not have sufficient volumes for either classification strategy. Using the 199 FIs with sufficient volumes for both strategy 5 and the benchmark strategy, the average sample size for Females based on classification strategy 5 was 2,609; the minimum sample size for Females was 164; the maximum sample size for Females was 37,173; the average sample size for Males was 8,181; the minimum sample size for Males was 1,857; and the maximum sample size for Males was 112,906. We discuss main findings from these tables below.

The second set of results (Tables A3-A4) presents disparity results and includes two tables, one for underwriting and one for pricing. All disparities in these tables are conditional disparities estimated using regression techniques that control for key factors that typically impact underwriting and pricing decisions, and that are available in HMDA data. For analyses of approval/denial decisions, we control for FICO score, CLTV, DTI, and loan amount, and for analyses of rate spread, we control for FICO score, CLTV, and loan amount. Each row in the tables represents a classification strategy. Strategy 1 is again used as the benchmark of comparison for all other strategies. For each table, the first column shows the classification strategy, treatment group, and control group. Column 2 shows the number of FIs with sufficient volumes using both the strategy for that row and the benchmark strategy, which is the same information provided in column 2 of Tables A1-A2. Columns 3-5 show the average, minimum, and maximum disparity estimates using the classification strategy of the given row. Only FIs with sufficient volumes for both the strategy for the given row and the benchmark strategy are used to generate these summary statistics. Column 6 shows the number of FIs where the estimated disparity using the strategy of the given row is larger than the estimated disparity using the benchmark strategy. Columns 7 and 8 show the number of these FIs where only the estimated disparity using the classification strategy of the given row was statistically significant and the number of FIs where estimated disparities using both strategies were statistically significant. The results in column 7 are particularly interesting, because they show the number of FIs where different conclusions would potentially be made and actions taken. The final three columns show similar results, but for FIs where the estimated disparity using the classification strategy of the given row is smaller than the estimated disparity using the benchmark strategy. Again, the results in the second to last column are particularly interesting, because they show the number of FIs where different conclusions would potentially be made and actions taken.

As an example for how to read these tables, for Table A3, which contains conditional denial rate disparity estimates from underwriting analyses of gender, row 5 shows results for the analysis of classification strategy 5 where the treatment group is females and the control group is males. A total of 199 of the 200 FIs we analyzed had sufficient volumes for analysis for both classification strategy 5 and strategy 1 (the benchmark strategy). The average, minimum, and maximum conditional denial rate disparity estimates across these 199 FIs are 0.59, -2.21, and 5.02 percentage points (pps). For 186 of these FIs, the estimated denial rate disparity was larger for strategy 5 than for the benchmark strategy. For 41 of these 186 FIs, only the estimated disparity for strategy 5 was statistically significant, and for 22 of these FIs both estimated disparities were statistically significant. Therefore, for 41 FIs, the results based on classification strategy 5 would indicate potential fair lending risk, while the results based on the benchmark strategy would not. Similarly, for 13 of the 199 FIs with sufficient volumes using strategies 5 and the benchmark strategy, the conditional denial rate disparity estimate was smaller for strategy 5 than for the benchmark strategy. For none of these 13 FIs, was only the estimated disparity for the benchmark strategy statistically significant, and for 4 of these FIs both estimated disparities were statistically significant. Therefore, for no FIs would the benchmark strategy indicate potential fair lending risk, but strategy 5 would not. We discuss main findings from these tables below.

As an additional perspective of the conditional disparity results, Tables A5-A6 present results summarizing the rank-order of conditional disparity estimates across classification strategies within FIs. There is one table for denial rate disparities and one table for rate spread disparities. Using the underwriting analysis as an example, the first step of this analysis is to estimate, separately for each FI, conditional denial rate disparities using each classification strategy. This analysis generates 7 denial rate disparity estimates for females for each FI. As a second step, we then sort, separately for each FI, the denial rate disparity estimates from largest to smallest. This creates, for each FI, a rank-ordered list of classification strategies, with the strategy resulting in the largest estimated disparity having a rank value of 1 and the strategy resulting in the smallest estimated disparity having a rank value of 7. For example, classification strategy 1 might have a rank value of 2 for the first FI, 5 for the second FI, 1 for the third FI, 1 for the fourth FI, and so forth. As a final step, we then generate the minimum, maximum, and median of each classification strategies’ rank values across the 199 FIs that had sufficient volumes for analysis.

As a specific example for how to read these two tables, we focus on the results for in Table A5, which provides denial rate disparity estimate results by gender. The minimum, maximum, and median ranks for classification strategy 1 are 1, 7, and 5, respectively. This means that for at least one of the 199 FIs the conditional denial rate disparity estimate for females based on classification strategy 1 was the largest estimated disparity (min rank = 1), for at least one of the 199 FIs the conditional denial rate disparity estimate for females based on classification strategy 1 was the highest disparity estimate (max rank = 7), and the estimated denial rate disparities for females based on classification strategy 1 tended to be lower on average compared to the other strategies (median rank = 5).

With the large volume of results in Appendix A it is not possible to discuss every interesting finding. Instead, we highlight just a few of the most interesting themes. First, volumes are not a constraint for any of the seven strategies, since for each strategy the volumes are insufficient for only 1 of the 200 FIs in our sample. This is not surprising given that our sample includes just the 200 largest HMDA reporters in 2021, and that percentages across genders tend to be more similar than across races or ethnicities. For smaller volume FIs beyond the FIs in our sample, volume would likely be a constraint for some strategies, specifically strategy 1 which uses only single-applicant applications. Second, there are clear patterns to the magnitudes of disparity estimates across classification strategies. The two strategies that classified applications as female if female was reported anywhere on the application (strategies 2 and 3) consistently showed the smallest conditional denial rate and rate spread disparity estimates. Conversely, the two strategies that classified applications as female if female was the only gender reported on the application (strategies 4 and 5) consistently showed the largest conditional denial rate and rate spread disparity estimates. In other words, when applications that included a female and a male applicant are classified as female, disparity estimates tended to be lower than when these applications are classified as male. Third, excluding same-sex applications appeared to have little impact on the denial rate or rate spread disparity estimates. Specifically, the disparity results for strategies 2 and 3 are similar, the disparity results for strategies 4 and 5 are similar, and the disparity results for strategies 6 and 7 are similar. This meets our expectations that worse potential outcomes separately for female/female and male/male joint applications in general might somewhat cancel each other out as these applications are classified as female and male, respectively for all strategies. Fourth, single-applicant applications, which is the cleanest classification strategy, tended to be in the middle of the disparity results for both underwriting and pricing. Finally, there is a fair amount of variation in potential conclusions and actions across classification strategies and FIs. For both underwriting and pricing, each classification strategy had some FIs where the disparity was higher and more likely to be statistically significant compared to the benchmark strategy and some FIs where the disparity was lower and less likely to be statistically significant compared to the benchmark strategy. This suggests that the appropriate classification to use might be analysis-specific, and depend on other information and perspectives, such as legal interpretation and policy objectives.

V. Conclusion

This report explores the gender variables available in HMDA data. While these data provide considerable flexibility and opportunities to conduct a wide variety of analyses of gender, it also creates significant challenges when choosing an approach to classifying applications into gender groups for fair lending analyses. Five specifically challenging issues, which we discuss in this report, are primary applicants versus co-applicants, same-sex joint applications, single-applicant applications versus joint applications, applications that report "6. Applicant selected both male and female," and applications that report gender for one applicant, but not the other. In addition to the analytical and coding complexities that these issues create, different approaches to addressing these issues can lead to different results and conclusions from fair lending disparity analyses.

Focusing on just the first three of the five issues, we provided empirical evidence using 2021 HMDA data showing how different choices about how to address these three issues generated subsets of applications with different underlying characteristics, and potentially different treatment. We then developed seven classification strategies that address these three issues in different ways, and showed how volumes of applications and loans, as well as estimated underwriting and pricing disparities, varied across these classification strategies. There are four items of note from these analyses: 1) volumes are not a constraint for any of the seven strategies, 2) there are clear patterns to the magnitudes of disparity estimates across classification strategies with strategies 2 and 3 consistently showing the smallest disparity estimates and strategies 4 and 5 consistently showing the largest disparity estimates, 3) excluding same-sex applications appeared to have little impact on the denial rate or rate spread disparity estimates, and 4) there was a fair amount of variation in potential conclusions and actions across classification strategies and FIs.

Overall, given the complexity of the gender data in HMDA, the lack of a clear legal definition of the appropriate classification strategy, and the empirical evidence in this report, it is important to consider each of the five specific issues discussed in this report, as well as any other issues or information that may be relevant, when choosing an approach to classifying applications into gender groups for fair lending analyses. Most importantly, it is vital to document the choices made and the rationale for those choices, as well as any supporting analyses showing how different strategies may impact the results from fair lending analyses.

Appendix A: Underwriting and Pricing Disparity Results

Table A1: Underwriting Counts (Benchmark is Strategy 1)

Classification Strategy, Treatment Group, Control Group	# FIs where the given strategy and benchmark strategy both have volume	# FIs where only the given strategy has volume	# FIs where only the benchmark strategy has volume	Avg # of Females	Min # of Females	Max # of Females	Avg # of Males	Min # of Males	Max # of Males
1, females vs males	199	0	0	2,554	152	36,562	3,661	646	56,144
2, females vs males	199	0	0	7,236	1,202	96,360	3,940	677	61,199
3, females vs males	199	0	0	7,054	1,189	93,144	3,735	650	56,935
4, females vs males	199	0	0	2,790	187	40,389	8,385	1,898	117,170
5, females vs males	199	0	0	2,609	164	37,173	8,181	1,857	112,906
6, females vs males	199	0	0	4,165	604	55,943	6,978	1,438	101,161
7, females vs males	199	0	0	3,984	590	52,751	6,775	1,411	96,996

Table A2: Pricing Counts (Benchmark is Strategy 1)

Classification Strategy, Treatment Group, Control Group	# FIs where the given strategy and benchmark strategy both have volume	# FIs where only the given strategy has volume	# FIs where only the benchmark strategy has volume	Avg # of Females	Min # of Females	Max # of Females	Avg # of Males	Min # of Males	Max # of Males
1, females vs males	199	0	0	2,295	109	33,349	3,289	541	51,260
2, females vs males	199	0	0	6,597	947	89,375	3,537	645	55,853
3, females vs males	199	0	0	6,435	923	86,474	3,355	562	51,979
4, females vs males	199	0	0	2,505	138	36,811	7,629	1,415	108,417
5, females vs males	199	0	0	2,343	119	33,910	7,447	1,361	104,543
6, females vs males	199	0	0	3,771	516	51,378	6,333	1,272	93,427
7, females vs males	199	0	0	3,610	498	48,501	6,153	1,219	89,641

Table A3: Conditional Denial Rate Disparities (Benchmark is Strategy 1)

Classification Strategy, Treatment Group, Control Group	# FIs where row strategy and benchmark strategy both have volume	Avg Disparity (pps)	Min Disparity (pps)	Max Disparity (pps)	# FIs where row disparity > benchmark disparity	# FIs where row disparity > benchmark disparity and only row disparity is s.s.*	# FIs where row disparity > benchmark disparity and both disparities are s.s.	# FIs where row disparity < benchmark disparity	# FIs where row disparity < benchmark disparity and only benchmark disparity is s.s.	# FIs where row disparity < benchmark disparity and both disparities are s.s.
1, females vs males	199	-0.23	-4.20	5.22	0	0	0	0	0	0
2, females vs males	199	-0.99	-6.79	2.45	33	8	8	166	6	24
3, females vs males	199	-1.02	-6.71	2.58	27	5	6	172	6	26
4, females vs males	199	0.55	-2.28	4.70	181	38	20	18	2	3
5, females vs males	199	0.59	-2.21	5.02	186	41	22	13	0	4
6, females vs males	199	-0.05	-3.92	2.70	125	11	20	74	4	10
7, females vs males	199	-0.04	-3.46	2.61	131	13	20	68	5	8

* s.s. means statistically significant at the 95 percent confidence level in this table and for all subsequent tables.

Table A4: Conditional Rate Spread Disparities (Benchmark is Strategy 1)

Classification Strategy, Treatment Group, Control Group	# FIs where row strategy and benchmark strategy both have volume	Avg Disparity (bps)	Min Disparity (bps)	Max Disparity (bps)	# FIs where row disparity > benchmark disparity	# FIs where row disparity > benchmark disparity and only row disparity is s.s.*	# FIs where row disparity > benchmark disparity and both disparities are s.s.	# FIs where row disparity < benchmark disparity	# FIs where row disparity < benchmark disparity and only benchmark disparity is s.s.	# FIs where row disparity < benchmark disparity and both disparities are s.s.
1, females vs males	199	0.31	-39.79	16.83	0	0	0	0	0	0
2, females vs males	199	-0.47	-38.54	6.53	81	17	16	118	21	23
3, females vs males	199	-0.40	-40.35	6.83	83	18	16	116	19	25
4, females vs males	199	1.19	-16.17	26.24	125	35	31	74	11	17
5, females vs males	199	1.01	-17.69	26.36	113	30	30	86	12	18
6, females vs males	199	1.06	-18.94	22.83	148	36	37	51	3	19
7, females vs males	199	1.01	-19.98	22.92	147	35	37	52	4	18

Table A5: Rankings of Conditional Denial Rate Disparities

Classification Strategy, Treatment Group, Control Group	Min Rank	Max Rank	Median Rank
1, females vs males	1	7	5.00
2, females vs males	1	7	6.00
3, females vs males	1	7	7.00
4, females vs males	1	7	2.00
5, females vs males	1	7	2.00
6, females vs males	1	7	4.00
7, females vs males	1	7	4.00

Table A6: Rankings of Conditional Rate Spread Disparities

Classification Strategy, Treatment Group, Control Group	Min Rank	Max Rank	Median Rank
1, females vs males	1	7	5.00
2, females vs males	1	7	6.00
3, females vs males	1	7	6.00
4, females vs males	1	7	3.00
5, females vs males	1	7	4.00
6, females vs males	1	7	3.00
7, females vs males	1	7	4.00