Gender discrimination in the age of Big Data: the case of auto loans

Galina Andreeva, University of Edinburgh, UK. E-mail: Galina.Andreeva@ed.ac.uk
Anna Matuszyk, New York University, USA; Warsaw School of Economics, Poland

1. Introduction
We investigate the consequences of restrictions on information in automated decision-making, using a specific example when the applicant’s gender cannot be used as a factor in risk assessment/credit scoring. The results apply generally to situations of algorithmic decisions based on empirical data. Credit scoring is a collection of mathematical and statistical models that predict the probability of a borrower’s default, using historic data that may include personal characteristics such as age, income, residential status. Gender is prohibited by law from use in decision-making in the majority of developed countries. The prohibition follows from anti-discrimination provisions, e.g. the European Equal Treatment in Goods and Services Directive.

We would like to test empirically the effectiveness of law in application to automatic decision-making, to highlight potential inconsistencies and inspire further research into better legal solutions. We do it by analyzing a unique proprietary dataset on car loans from an EU bank, which contains gender, other application characteristics and observed credit performance. Our investigation consists in following a standard credit scoring methodology that is used by banks in practice to construct a model based on credit application variables (with and without Gender) and to observe changes in parameter estimates and predictive accuracy.

Table 1. Training and test samples.

<table>
<thead>
<tr>
<th>Training</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good</td>
<td>Bad</td>
</tr>
<tr>
<td>Female % by column</td>
<td>16746</td>
</tr>
<tr>
<td>Male % by column</td>
<td>45696</td>
</tr>
<tr>
<td>Total % by column</td>
<td>62442</td>
</tr>
</tbody>
</table>

4. Predictive accuracy
Parameter estimates and model fit measures of the four models are available on request. Gender is statistically significant (p-value <0.001), its removal leads to slightly worse model fit and some changes in parameter estimates. The biggest changes in parameter estimates are observed for ‘female’ model.

Predictive accuracy is measured by Area under the Roc curve (AUC) and is given in Table 2. Although AUC is higher when separate models are used for both sexes, for males it does not matter much which model is used, whereas for females the uplift is more pronounced. Whist there is little benefit to women from a simple inclusion of gender into the model, the segmentation does allow capturing unique features of female risk profiles.

Figure 1. Impact on rejection by gender when different models are used, % men/women rejected v overall reject rate.

5. Chances of being accepted/rejected
The ultimate question is how the chances to be accepted for credit are affected. Hand (2012) outlines the following scenarios when talking of potential solutions in achieving discrimination-free credit decisions:

- **a) Current situation** – prohibit the use of gender. One can also consider removing variables with Gender, but the question arises, what level of correlation would be acceptable and how many variables would be left for model building.
- **b) Ensure equal outcome** – accept the same proportion of men and women.

Note that only scenario A is legal under the existing regulations, since B requires the use of a legal solution.

To assess the impact on access to credit, the proportions of men and women have been compared for different cut-off levels that would correspond to a range of rejection/acceptance rates: from 10% to 90% in 10% increments (Figure 1). E.g., if a lender rejects 60% of the population (sample) and uses PDs from the unisex model (Model 2) as scores, 58.74% of all men in the sample would be rejected as compared to 63.48% of all women, thus the latter segment is not being rewarded for being better credit risks. However, if Model 1 is used for the same cut-off (60% overall rejection), the corresponding percentages become 61.27% for men and 56.52% for women, thus rewarding women for being better credit risks.

Overall, men, being less creditworthy, benefit from unisex model. On the contrary, women would benefit from including Gender, since more females would be accepted for credit. However, the removal of Gender does not make the reject rates equal for both genders, it almost reverses them, disproportionately punishes women as a more creditworthy class.

6. Conclusion
The results are indicative of the law of unintended consequences. Surely, the main objective of the equality provisions is to protect consumers. Yet, it has been shown that the regulations do not ensure equality of outcome. More creditworthy groups subsidise worse risks, but this subsidy is disproportionately punishing women as a more creditworthy class.

References