A STATISTICAL ANALYSIS OF BIAS IN A PERSONNEL ASSIGNMENT PROBLEM: STABLE SOLUTIONS VS. MULTIPLICATIVE UTILITY SOLUTIONS

by

Yufei Yuan, Abraham Mehrez, Amiram Gafni
McMaster University
Faculty of Business

WORKING PAPER NO. 311

October, 1988
A STATISTICAL ANALYSIS OF BIAS IN A PERSONNEL ASSIGNMENT PROBLEM: STABLE SOLUTIONS VS. MULTIPLICATIVE UTILITY SOLUTIONS

by

Yufei Yuan, Abraham Mehrez, Amiram Gafni
McMaster University
Faculty of Business

WORKING PAPER NO. 311

October, 1988
A STATISTICAL ANALYSIS OF BIAS IN A PERSONNEL ASSIGNMENT PROBLEM:

STABLE SOLUTIONS VS. MULTIPLICATIVE UTILITY SOLUTIONS

Yufei Yuan*, Abraham Mehrez**, Amiram Gafni***

* Assistant Professor, Dept. of Management Science, Faculty of Business, McMaster University, Hamilton, Ontario, Canada.
** Visiting Professor, Dept. of Management Science, Faculty of Business, McMaster University, Hamilton, Ontario, Canada, and the Dept. of Industrial Engineering and Management, Ben-Gurion University of the Negev, Beer Sheva, Israel
*** Associate Professor, Center for Health Economics and Policy Analysis, Dept. of C.E. & B. Faculty of Health Sciences, McMaster University, Hamilton, Ontario, Canada.

Acknowledgement:
Financial support for this research was provided by the Natural Sciences and Engineering Research Council of Canada.
ABSTRACT

Fairness or bias in selecting employees is an important issue which is widely discussed in the literature dealing with human resources. In this paper we study a different type of bias. This bias stems from the type of mathematical algorithm used to determine an optimal match between two groups. We compare two different solution concepts for the matching assignment problem: the stable solution vs. the multiplicative utility approach. For a very small scale problem the multiplicative utility approach was found by Mehrez, Yuan and Gafni (1988) to be more fair compared to the stable approach. Using a simulation model we study the following questions: (a) Does the size of the problem affect the degree of the bias when using different approaches to solve the problem? (b) If yes, in what direction? Our main findings are: With respect to all sizes compared in our experiment the outcome was always more fair when using the multiplicative utility approach compared to the stable approach. When using an absolute measure to determine the scope of these discrepancies we find a size effect -- the bigger the size of the problem the bigger is the performance discrepancy between two parties when using the stable approach. No such size effects were found when the multiplicative utility approach was used. When using a relative measure to determine the scope of the discrepancies, no size effects were found for both approaches.
INTRODUCTION

Recently Mehrez, Yuan and Gafni (1988) have studied an important personnel management problem -- the matching assignment problem, which has many applications in particular in personnel and manpower planning. The authors concentrated on studying two different solution concepts for this problem. The first one is known as the stable solution and was applied for example, by the National Resident Matching Program (NRMP) to assign graduates of medical schools to hospital internship positions (see Graettinger and Peranson (1981), Roth (1984a)). This type of solution is accepted by many authors due to the fact that it results in a match which is Pareto optimal and its computational simplicity (see McVitie and Wilson (1970), Roth (1985)).

The second approach to solve the matching assignment problem is the multiplicative utility approach which does not necessarily lead to a Pareto-optimal solution and may require complex computations (see Keeney and Raiffa (1976, ch. 10), Mehrez and Shinhar (1982), or Mehrez, Yuan and Gafni (1988)). The advantage of this approach is that unlike the stable approach, it allows to treat both sides (students and hospitals in the case of the NRMP) equally. This is an important feature since for example in the case of the NRMP, a major concern was raised about the fairness of the algorithm (stable solution) which "treats students and programs differently, in a way that favors programs whenever desires conflict" (Williams et al (1981)).

In the paper by Mehrez, Yuan and Gafni (1988) the matching assignment problem is presented with the two-attribute multiplicative utility function. Then its solutions are compared both conceptually and numerically with stable solutions using McVitie and Wilson's (1971) small size example of matching eight men with eight women. The paper's conclusion is that utility
maximization solutions do not seem, on numerical basis, to perform worse than the stable solutions. Furthermore, as expected, the multiplicative utility method results in a more equitable (or fair) outcome compared with those generated by using the stable assignment algorithm.

The field of personnel selection is one of the most dynamic areas within the larger field of human resource management. The specific topic of fairness in selecting employees has been particularly volatile due to the constant flow of court decisions (see for example Arvey and Faley (1988)). Thus the "fairness aspect" of the results of personnel assignment problems is a very important one. It is important at the outset to clarify that the selection procedure may differ among organizations. However, an essential ingredient in each of these procedures is to obtain information to be used in making selection decisions. In this paper we do not deal with the process of information gathering by organizations or how individuals decide to apply for a particular job (which is known as self selection). The reader who is interested in these topics can consult the following references: Arvey and Faley (1988) or Wanons (1979). In this paper we deal with the "fairness aspect" of what can be seen as two personnel selection algorithms (or methods) assuming a given process of information gathering by participants.

Unfortunately, it is not altogether clear what is meant by the terms "fair" versus "unfair" discrimination. There are several definitions of these concepts, all of which make some sense, but at present, there is no universally acceptable definition of "unfair" or "biased" discrimination. The approach that we use in this paper (as well as the previous one -- Mehrez et al (1988)) is to define bias as statistically measured performance difference between two groups. In spite of some criticism raised this is a
very common approach in the literature which looks at discrimination in a statistical sense (rather than legal or moral sense). More about this approach and its use in the literature can be found in Arvey and Faley (1988).

Another reason why the "fairness aspect" of the result of assignment problems is important is the following: In many cases the matching process is a voluntary one (namely, both sides have to agree to participate in the process). Thus, the feeling of both sides about the fairness of the process is an important element in determining the willingness to participate and hence the success of such system. The National Resident Matching Program (NRMP) is a good example. It was established to bring order and fairness to a previously chaotic application process for internship and residency positions. The success of this program was partly due to the feeling that both sides were treated equally. However, in recent years many reservations were raised about the lack of fairness of the process which some claim have led to the existence of "official" and "unofficial" matching processes (Polk (1986)).

In light of the importance of the "Fairness aspect" of the outcome of a personnel assignment problem the finding in Mehrez et al (1988) study seems to be an important one. However, this finding is based on a small size example. It is thus important to find (a) does the size of the problem affect the degree of the bias? (b) if yes, in what direction? In this paper we describe the results of a simulation model which is used to analyze the effect of the size of the problem on the performance of the participants when using different assignment algorithms (stable solution vs. multiplicative utility solution). In the second section we describe the
experiment. In the third section we describe the major results. The final section of the paper deals with the policy implications of our findings.

THE EXPERIMENT

We use a simulation model to analyze the effect of the size of the problem on the relative performance of the participants when using different assignment algorithms. The assignment problem used in our experiment is the classical matching assignment problem which deals with finding the optimal match for n men and n women when their preferences towards all tentative marriages are given. This example was chosen for illustrative purposes only. The results of our analysis are general and hold in many other cases as well, such as matching residents and hospitals, students and universities etc.

The three methods compared are: the multiplicative utility solution (denoted as method 1), the male-optimal stable solution (denoted as method 2), and the female-optimal stable solution (denoted as method 3). A description of the multiplicative solution can be found in Mehrez, Yuan and Gafni (1988). A description of the stable solution can be found in Gale and Shapley (1962) or McVitie and Wilson (1970, 1971). For the convenience of the reader we provide a brief description of the above methods in the appendix. We chose only the male-optimal and female-optimal stable solutions since the stable algorithm does not have any mechanism to prevent bias toward one side in the assignment process. Thus the male-optimal and the female-optimal solutions are the most commonly used algorithms. It should be mentioned that all possible stable solutions can be found (see McVitie and Wilson (1971)) but in practice this has not been done.
The sample preference data was randomly generated based on the following assumptions: (1) all participants have equal probabilities of receiving any rank; (2) participants' ranks are independent of each others; (3) for each random run the increase in sample size does not change the relative preference ranking of the participants (however, it might change their absolute rank order).

Six different sizes were chosen for the analysis: 8x8, 16x16, 32x32, 64x64, 128x128 and 256x256. With each size we analyzed the solutions generated from 50 random runs. Note that the smallest size problem in our analysis is 8x8, which is equal to the size of the example used in Mehrez, Yuan and Gafni (1988) to compare these assignment algorithms. We have restricted our analysis to these six size levels from the following reasons: (1) As can be seen in the results section these six size levels are sufficient for our purposes to analyze the effect of the size of the problem on the performance of the participants. (2) Our largest problem size level 256x256 can represent the case of a typical medium size firm. We realize that large scale problems do exist and later in the paper we discuss the relevance of our findings to such large scale problems.

For the case of the stable algorithm (method 2 -- male-optimal; method 3 -- female-optimal) the rank order provided by the different participants was utilized to determine the match. For the case of the multiplicative utility algorithm (method 3) the ranks provided were converted into utility values. The conversion function that was used is a linear function which converts the i-th rank order to a utility value which equals to \(1 - \frac{i-1}{n}\), where n is the number of all possible partners. We realize that in real life the conversion function is not necessarily a linear one. However, in
this example for the sake of simplicity, a linear conversion function was chosen.

THE RESULTS

To compare the three assignment algorithms (multiplicative utility solution, male-optimal stable solution and female-optimal stable solution) we use the average rank order received by each side, resulting from the match outcome, as the measure of their performance. The difference between the average rank order of males and females is used as a measure of "fairness" or bias of the match results. Since our sample preference data are symmetric, if any bias exists it can only be attributed to the algorithm used to generate the match.

In Table 1 the average ranks received by each side (males and females) for the different size levels and for the different methods used to determine the match are presented. This information is also presented in a graphical way in Figures 1-3. Comparing the three methods one can see that the male performance is always the best (regardless the size) when using method 2 (male-optimal stable solutions) and the female performance is always the best when using method 3 (female-optimal stable solution) to determine the match. This is not surprising since method 2 is biased in favour of males and method 3 is biased in favour of females. It is also easy to see that the female performance is the worst (regardless of size) when using method 2 and male performance is the worst when using method 3. (The theoretical discussion and mathematical proving of the male or female optimality can be found in Roth (1984b) but Roth did not study to which extent the bias might be.)

(Table 1 and Figures 1-3 about here)
The discrepancies between males' and females' performances lead to several questions. The first one is whether for each method of matching used and for each size of the problem the difference in performance (denoted as the bias of the result) is statistically significant from zero. In Table 2 the results of an analysis aimed at answering this question are presented. For method 1 (multiplicative utility solution) for the first three size levels compared (8x8, 16x16, 32x32) the differences in performance are not found to be statistically significant non zero. For the other three size levels compared (64x64, 128x128, 256x256) small, but statistically significant differences between males' and females' performance are found. Theoretically, there should be no differences between male and female because they are treated equally. The small differences might be due to the inadequate number of sample runs for the larger size problem. Since the calculations for the large size personnel assignment problem were quite time-consuming, we did not further increase the number of sample runs. In both methods 2 and 3 (male-optimal stable solution and female-optimal stable solution respectively) statistically significant differences in performance are found for all 6 size levels compared. In method 2 the bias is in favour of males and in method 3 the bias is in favour of females. Also, as can be seen in Table 2, the differences in performance between participants for all sizes compared, are much bigger when using the stable solution method (method 2 and 3) than when using the multiplicative utility method.

(Table 2 about here)

A general linear regression model and F tests are used to examine the impact of the size of the match results and on the degree of the match bias. The results are presented in tables 3 and 4. From table 3 we learn that the size of the problem is an important factor in determining the average rank
received by the matched males and females for all three methods. However, for method 2 (male-optimal stable solution) we find that the size of the problem has a smaller effect on the average rank received by matched males \( R^2 = 0.2382 \) compared to matched females \( R^2 = 0.4930 \), in other words, the performance of males is less influenced by the size. For method 3 (female-optimal stable solution) we find that the size of the problem has a smaller effect on the average rank received by females \( R^2 = 0.1347 \) compared to males \( R^2 = 0.9509 \).

(Table 3 about here)

We have also investigated the effect of the problem size on the degree of the match bias. Two criteria are used: an absolute measure of bias and a relative measure of bias. The first one compares the difference between the ranks of the matched males and the matched females, the second one compares the difference between the relative rank order of the matched males and females. From Table 4 we learn that using the absolute measure the size of problem has little effect on the degree of the match bias in the case of method 1, the multiplicative utility approach \( R^2 = 0.663 \). However, in the cases of method 2, the male-optimal stable solution and method 3, the female-optimal stable solution the size of the problem affects the degree of the match bias significantly \( R^2 = 0.9134 \) and \( 0.9127 \) respectively). For these two methods, the larger the problem is, the larger is the bias. When using the relative measure of bias, the size of the problem does not affect the size of the match bias in all cases.

(Table 4 about here)
CONCLUSIONS AND POLICY IMPLICATIONS

Fairness or bias in selecting employees is an important issue which is widely discussed in the literature dealing with human resource management (for a good and detailed review of this literature see a recent book by Arvey and Faley (1988)). In general the most commonly studied topics in this field are: legal discrimination, unfair test discrimination and discrimination in the employment interview. In this paper we study a different type of bias. This bias stems from the type of mathematical algorithm used to determine an optimal match between two groups. This type of bias characterizes the era of computers where the assignment (or selection) task is "left to computers".

In this paper we have compared two different solution concepts for the matching assignment problem: the stable solution vs the multiplicative utility approach. We show that regardless of the size of the problem, the multiplicative utility algorithm always results in a more equitable (fair) outcome compared with the stable algorithm. Furthermore, when using an absolute measure of bias we find that the size of the problem affects the size of the bias when using the stable algorithm. The size of the problem does not have any effect on the size of the bias when using a relative measure of outcome.

It is important to note that in this paper we deal with personnel selection algorithms (or methods) where the process of information gathering by the participants results in a complete data set (by which every male ranks every female and vice versa). In very large scale problems this process might be very expensive and thus not feasible. Therefore, in such large scale problems we are likely to end up with incomplete data set (due to incomplete ranking). This might mean in practice that some participants
will be left without a match. This might also affect the degree of bias created. However, prior to measuring the degree of bias in such systems one has to cope with the problem of assessing the performance of participants who will be left without a match which is not an easy one to solve.

An important question is which algorithm should be used in matching assignment problems. From a fairness perspective, as demonstrated in this paper, the multiplicative utility approach is the preferred option which enables us to treat both sides in a more equitable way and results in a much smaller bias of the match outcome. But using the stable method it is difficult to treat both parties equally. However, in some cases such bias may be acceptable due to the unequal demand and supply in the market or other organizational considerations. Other criteria may also affect our decision on selecting an assignment algorithm. For example, when dealing with large scale problems the cost of information gathering and the computational efficiency are important factors to be considered. The stable algorithm requires only ordinal preference information and can easily handle a very large scale assignment problem. The multiplicative utility approach however, requires cardinal utility information which is difficult to collect. The multiplicative utility solution although can be derived by using coded assignment algorithm, it requires large memory and consumes much more CPU time (see Klingman and Phillips (1984) for the description of such algorithms for very large scale problems). The decision maker, who has to choose the algorithm to be used, is faced with a classical trade-off between the equity aspect of the solution and the computational efficiency aspect of the algorithm.
REFERENCES


1) The Multiplicative utility approach

The following describes the multiplicative utility algorithm suggested by Mehrez, Yuan and Gafni (1988). The procedure is based on solving the following assignment problem:

Maximize \( Z = \sum_{i=1}^{n} \sum_{j=1}^{n} U(i,j) x_{ij} \)

s.t. \( \sum_{i=1}^{n} x_{ij} = 1 \)

\( \sum_{j=1}^{n} x_{ij} = 1 \)

\( x_{ij} \) are all 0 or 1

where \( i, j = 1, \ldots, n \)

\( x_{ij} =
\begin{cases}
1 & \text{if the ith man is assigned to the jth woman} \\
0 & \text{otherwise}
\end{cases}
\)

This is a personnel assignment problem with utilities that are derived by a heuristic approach. The \( U(i,j) \) are evaluated using the following relations:

\( U(i,j) = \frac{[(1+K_M(i)K(i,j)U_M(i,j))(1+K_F(j)K(i,j)U_F(i,j))-1]}{K(i,j)} \)

where \( U_M(i,j) \) and \( U_F(i,j) \) are the utilities for the pair \((i,j)\) as evaluated by the \( i \)-th man and the \( j \)-th woman respectively,

\( K_M(i) = \frac{\sum_{j=1}^{n} U_M(i,j)}{\sum_{i=1}^{n} \sum_{j=1}^{n} U_M(i,j)} \), and

\( K_F(j) = \frac{\sum_{i=1}^{n} U_F(i,j)}{\sum_{i=1}^{n} \sum_{j=1}^{n} U_F(i,j)} \)

represent the overall attractiveness of the \( i \)-th man and the \( j \)-th woman respectively, and

\( K(i,j) = \frac{[1-K_M(i)K_F(j)]}{[K_M(i)K_F(j)]} \)

is used to formalize the two-attribute multiplicative utility function for the pair \((i,j)\) (see for example Keeney and Raiffa (1976)).
2) The stable assignment approach

The following describes the stable marriage algorithm suggested by Gale and Shapley (1962):

A certain community consists of n men and n women. Each man ranks each woman in accordance with his preference for a marriage partner and each woman ranks each man in accordance with her preference for a marriage partner. To start the assignment, let each man propose to his most favorite woman. Each woman who receives more than one proposal rejects all but her favorite from among those who have proposed to her. Those men who were rejected then propose to their second choices. If a woman receives new proposals, she compares them with the proposal she has held, selects the most favorite one and rejects the rest. Those men who were rejected then propose to their next choices. The process continues in the same manner until every woman receives a proposal and no further rejection happens. Then, each woman finally accepts the proposal she holds as the solution of the matching. This resulting solution is male-optimal. If women propose and men decide to hold or reject the proposals, the solution is female-optimal.
Table 1
Average Rank Received for the Matched Pairs

<table>
<thead>
<tr>
<th>Problem Size</th>
<th>Method 1</th>
<th></th>
<th>Method 2</th>
<th>Method 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R-Male</td>
<td>R-Female</td>
<td>R-Male</td>
<td>R-Female</td>
</tr>
<tr>
<td>8 x 8</td>
<td>2.468</td>
<td>2.415</td>
<td>2.180</td>
<td>3.180</td>
</tr>
<tr>
<td>16 x 16</td>
<td>3.381</td>
<td>3.191</td>
<td>2.715</td>
<td>5.349</td>
</tr>
<tr>
<td>32 x 32</td>
<td>4.718</td>
<td>4.631</td>
<td>3.583</td>
<td>8.158</td>
</tr>
<tr>
<td>64 x 64</td>
<td>6.883</td>
<td>5.924</td>
<td>4.092</td>
<td>15.253</td>
</tr>
<tr>
<td>128 x 128</td>
<td>8.843</td>
<td>8.209</td>
<td>5.172</td>
<td>22.606</td>
</tr>
<tr>
<td>256 x 256</td>
<td>11.295</td>
<td>10.649</td>
<td>4.524</td>
<td>48.557</td>
</tr>
</tbody>
</table>

Method 1: Multiplicative utility solution
Method 2: Male-optimal stable solution
Method 3: Female-optimal stable solution
R-Male: Average rank received by male
R-Female: Average rank received by female
<table>
<thead>
<tr>
<th>Problem Size</th>
<th>Method 1</th>
<th></th>
<th>Method 2</th>
<th></th>
<th>Method 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R-Bias</td>
<td>R-Bias</td>
<td>R-Bias</td>
<td>R-Bias</td>
<td>R-Bias</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>Std.</td>
<td>Mean</td>
<td>Std.</td>
<td>Mean</td>
</tr>
<tr>
<td>8 x 8</td>
<td>0.053*</td>
<td>0.874</td>
<td>-1.000</td>
<td>1.215</td>
<td>0.928</td>
</tr>
<tr>
<td></td>
<td>(0.7628)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16 x 16</td>
<td>-0.010*</td>
<td>0.794</td>
<td>-2.634</td>
<td>1.821</td>
<td>3.014</td>
</tr>
<tr>
<td></td>
<td>(0.9295)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32 x 32</td>
<td>0.086*</td>
<td>0.792</td>
<td>-4.575</td>
<td>2.355</td>
<td>3.157</td>
</tr>
<tr>
<td></td>
<td>(0.4448)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>64 x 64</td>
<td>0.958</td>
<td>0.812</td>
<td>-11.161</td>
<td>3.953</td>
<td>7.071</td>
</tr>
<tr>
<td>128 x 128</td>
<td>0.634</td>
<td>0.547</td>
<td>-17.434</td>
<td>5.438</td>
<td>15.577</td>
</tr>
<tr>
<td>256 x 256</td>
<td>0.646</td>
<td>0.698</td>
<td>-44.032</td>
<td>7.284</td>
<td>38.471</td>
</tr>
</tbody>
</table>

A T-test is used to test the bias of three solution methods.

Method 1: Multiplicative utility solution
Method 2: Male-optimal stable solution
Method 3: Female-optimal stable solution
R-Bias: Rank received by male - Rank received by female
* : T test of Mean ≠ 0 can not be rejected. The significant level of the T test is presented in brackets. In all other cases Mean ≠ 0 is rejected at at the significant level 0.0001.
Table 3

The Size Effect On the Match Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Method 1</th>
<th>Method 2</th>
<th>Method 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R-Male</td>
<td>R-Female</td>
<td>R-Male</td>
</tr>
<tr>
<td>Rank=Ax+size+B</td>
<td>0.0341</td>
<td>0.0315</td>
<td>0.0086</td>
</tr>
<tr>
<td></td>
<td>3.4075</td>
<td>3.2220</td>
<td>2.9886</td>
</tr>
<tr>
<td>R²</td>
<td>0.8907</td>
<td>0.9129</td>
<td>0.2382</td>
</tr>
<tr>
<td>F Value(DF=1)</td>
<td>2429.00</td>
<td>3122.66</td>
<td>145.61</td>
</tr>
<tr>
<td>Prob &gt; F</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

A general linear regression model is used to test the size effect on the average rank received by males and females.

Method 1: Mutliplicative utility solution
Method 2: Male-optimal stable solution
Method 3: Female-optimal stable solution
R-Male: Average rank received by male
R-Female: Average rank received by female
Table 4
The Size Effect On Match Bias

<table>
<thead>
<tr>
<th>Model</th>
<th>Absolute measure of bias</th>
<th>Relative measure of bias</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Method 1</td>
<td>Method 2</td>
</tr>
<tr>
<td>Bias=A\times size+B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0.00249</td>
<td>-0.1693</td>
</tr>
<tr>
<td>B</td>
<td>0.1855</td>
<td>0.7521</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.0663</td>
<td>0.9134</td>
</tr>
<tr>
<td>F value (DF=1)</td>
<td>21.149</td>
<td>3124.544</td>
</tr>
<tr>
<td>Prob &gt; F</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

A general linear regression model is used to test the size effect on the degree of bias between males and females.

Method 1: Multiplicative utility solution  
Method 2: Male-optimal stable solution  
Method 3: Female-optimal stable solution

Absolute measure of bias = Average rank received by male - Average rank received by female
Relative measure of bias = Average relative rank received by male - Average relative rank received by female
Figure 1

Match Performance For the Multiplicative Utility Solution
Figure 2
Match Performance For the Male-Optimal Solution
Figure 3

Match Performance For the Female-Optimal Solution

![Graph showing the match performance for the female-optimal solution.](image-url)
Faculty of Business
McMaster University

WORKING PAPERS - RECENT RELEASES


