University of Connecticut
DigitalCommons@UConn

Economics Working Papers | Department of Economics

5-1-2007

Marriage Matching with Correlated Preferences

Onur B. Celik
University of Connecticut

Vicki Knoblauch
University of Connecticut

Follow this and additional works at: http://digitalcommons.uconn.edu/econ_wpapers

Recommended Citation
Celik, Onur B. and Knoblauch, Vicki, "Marriage Matching with Correlated Preferences" (2007). *Economics Working Papers.* Paper 200716.
http://digitalcommons.uconn.edu/econ_wpapers/200716

This is brought to you for free and open access by the Department of Economics at DigitalCommons@UConn. It has been accepted for inclusion in Economics Working Papers by an authorized administrator of DigitalCommons@UConn. For more information, please contact digitalcommons@uconn.edu.

University of Connecticut

Department of Economics Working Paper Series

Marriage Matching with Correlated Preferences

Onur B. Celik
University of Connecticut

Vicki Knoblauch
University of Connecticut

Working Paper 2007-16

May 2007

341 Mansfield Road, Unit 1063
Storrs, CT 06269–1063
Phone: (860) 486–3022
Fax: (860) 486–4463
http://www.econ.uconn.edu/

This working paper is indexed on RePEc, http://repec.org/

Abstract

Authors of experimental, empirical, theoretical and computational studies of two-sided matching markets have recognized the importance of correlated preferences. We develop a general method for the study of the effect of correlation of preferences on the outcomes generated by two-sided matching mechanisms. We then illustrate our method by using it to quantify the effect of correlation of preferences on satisfaction with the men-propose Gale-Shapley matching for a simple one-to-one matching problem.

Journal of Economic Literature Classification: C78, D63, B41

Keywords: Two-Sided Matching, Correlated preferences, Gale-Shapley algorithm

The authors would like to thank Fusun Yaman for programming assistance and James Boudreau for helpful comments and suggestions.

1. Introduction.

Correlation of preferences affects every two-sided market, as has been recognized for sorority rushes, (Mongell and Roth, 1991), the market for law clerks (Haruvy, Roth and Unver, 2006) and New York City high school admissions (Abdulkadiroğlu, Pathak and Roth, 2006). In particular, correlation of preferences is a strong factor in determining men's and women's aggregate levels of satisfaction with the outcome of the Gale-Shapley (1962) marriage matching algorithm. For example, for the men-propose G-S algorithm, if women's preferences over men are highly correlated and men's preferences over women are random, we would expect a high total of rankings by women of their assigned mates (that is, a low level of aggregate satisfaction among women) because the high correlation of women's preferences means they are competing for the same men. The answer to the question of how men fare in the same scenario is less obvious. To answer this and related questions, and to increase our general understanding of both the efficacy and workings of the G-S algorithm, our goal is to quantify the relationship between men's and women's satisfaction with the G-S outcome on the one hand and correlation of men's preferences and of women's preferences on the other. We will find that, despite the complex interaction between men's and women's preferences that takes place during a run of the G-S algorithm, aggregate levels of satisfaction with the outcome can be predicted with some degree of accuracy from just two numbers, one derived from men's preferences and one from women's preferences.

The contribution of this study is two-fold. First there is the above-mentioned quantification of the relationship between correlation of preferences and aggregate satisfaction with the men-propose G-S matching. Second there is our four step methodology described in the following paragraph. Since we define a measure of correlation that can be applied to the preference profile of any group, our methodology has a wide range of application in the field of two-sided matching. Although we limit ourselves to one-to-one matching, our methodology applies to many-to-one matching. Furthermore, it could be used to study mechanisms other than the G-S algorithm, and to study the effects of correlated preferences on measures of social welfare other than the one we use, the sum of a group's rankings of their mates. Finally, designers of a matching mechanism could use our methodology to

determine to what extent the effects of strategic behavior on social welfare are ameliorated or enhanced by correlation of preferences.

Our quantification of the relationship between satisfaction with the G-S outcome on the one hand and correlation of preferences on the other involves four steps.

Step 1. We define a measure of correlation for any preference profile, where a preference profile consists of n ranking lists formed when each of n men ranks n women (or each of n women ranks n men).

Step 2. We generate a variety of marriage matching problems; that is, we generate a variety of pairs of preference profiles, each pair consisting of a men's preference profile and a women's preference profile.

Step 3. We run the men-propose Gale-Shapley algorithm for each problem generated in Step 2.

Step 4. For each problem generated we calculate the correlation (as defined in Step 1) of each of the two preference profiles, the sum of men's rankings of their assigned mates and the sum of women's rankings of their assigned mates (for the assignment made in Step 3). Then we examine the relationship between correlation and the expected sum of a gender's rankings of their assigned mates.

Caldarelli and Capocci (2001) were the first to endow men and women with correlated preferences over members of the opposite sex, then run the Gale-Shapley algorithm using these preferences. In their study each man m gives each woman w a score $S = \eta_{mw} + UI_w$, where η_{mw} is chosen randomly from $(0, 1)$ and represents a personal dimension of m's opinion of w, I_w is chosen randomly from $(0, 1)$ and represents a consensus view of w and $U \geq 0$ is a weighting factor. Then each man uses the scores he has assigned to form a preference list. Women form preferences similarly. The G-S algorithm was run using $U = 0$, $U = .1$ and $U = 1$. They found that more desirable individuals do better; that is, the rankings of their assigned mates are lower, and this effect is greater for $U = 1$ than for $U = .1$. Our study differs from Caldarelli and Capocci in at least four ways: 1) We define a measure of correlation for *any* preference profile. Caldarelli and Capocci's measure of

correlation U is defined only for preferences constructed using their method. 2) Even when applicable , U is a less direct measure of correlation than ours (defined in Section 3 below), since U can be assigned before the random selections of η_{mw} and I_w. Therefore U is really an expected correlation. 3) We use a different method of generating correlated preference profiles for reasons explained in Section 7. 4) Our goals are different from Caldarelli and Capocci's. We are trying to quantify the relationship between a gender's aggregate level of satisfaction with the G-S outcome on the one hand and the levels of correlation of men's preferences and of women's preferences on the other; and to produce a methodology with a wide range of applications.

Similarly, Chen and Sönmez (2006) endow experimental subjects portraying students with correlated preferences over schools in a study comparing the efficency of three matching mechanisms in a student/school matching problem. The four comments immediately above distinguishing our work from that of Calderelli and Capocci also apply to Chen and Sönmez.

Wilson (1972) proved that for any profile of women's preferences, if men's preferences are random, then the expected sum of men's rankings of their mates as assigned by the men-propose G-S algorithm is bounded above by $n(1 + 1/2 + \ldots + \frac{1}{n})$. Knoblauch (2006) showed that this is also an approximate lower bound in the sense that the ratio of the expected sum of men's rankings of their assigned mates and $n(1 + 1/2 + \ldots + \frac{1}{n})$ has limit 1 as n goes to ∞. However, for the broader goal in this paper–quantification of the relationship between satisfaction with assigned mates and correlation of preferences–simulation is more promising than theoretical work: Roth and Peranson (1999) point out that economists need to develop "an engineering design literature," and show through example that "Theoretical computation can be a big help in this effort."

The paper is organized as follows. Section 2 contains preliminaries. Section 3 introduces a measure of correlation of preferences. Section 4 describes a method for generating preferences with varying levels of correlation. Sections 5 and 6 gauge the effects of correlation of preferences on aggregate satisfaction when one gender's preferences are correlated, the other's random, and when both gender's preferences are correlated, respectively. Section 7 concludes.

2. Preliminaries.

Our stated goal is to study via simulation the effect of correlation of preferences on aggregate satisfaction with the G-S men-propose matching in the context of the simplest version of the marriage matching problem. We therefore begin with n men, $m_1, m_2, \ldots, m_n$ and n women $w_1, w_2, \ldots w_n$. Each woman ranks the men 1st through n^{th}. We denote w_j's ranking of the men by a one-to-one, onto function r_{w_j}: $\{m_1, m_2, \ldots, m_n\} \to \{1, 2, \ldots, n\}$. She prefers m_k to m_l if $r_{w_j}(m_k) < r_{w_j}(m_l)$. We denote the men's ranking of the women analogously.

The G-S men-propose algorithm produces a matching, that is a 1-1 onto function μ: $\{m_1, m_2, \ldots m_n\} \to \{w_1, w_2, \ldots, w_n\}$, as follows. In round 1 each man proposes to his most preferred woman. Each woman proposed to becomes tentatively engaged to the man she prefers among those who have proposed to her if any. In each subsequent round, each man not engaged proposes to his most preferred woman among those he has not yet proposed to. Each woman becomes tentatively engaged to the man she prefers among her new proposers, if any, and her current fianceé, if she has one, or remains single if she has neither. The procedure ends, and engagements become marriages, when every woman is engaged.

The most important property of the G-S algorithm is that it always produces a stable matching. A matching is *stable* if there exists no man and woman each of whom prefers the other to his or her own spouse.

In the following sections we will define a measure of correlation of preferences, describe our method for generating correlated preferences, and then proceed to study via simulation the effect of correlated preferences on aggregate satisfaction with the G-S men-propose matching.

3. A Measure of Correlation for Preferences.

Our definition of a measure of correlation of preferences is motivated by the observation that if a women's preference profile is highly correlated, then *Ave* 1, *Ave* 2, $\ldots$, *Ave* n tend to be widely spaced, where *Ave* i is the average ranking of m_i by the n women.

Given a women's preference profile $\{r_{w_j}(m_i)\}$ for $1 \leq i, j \leq n$, for $1 \leq i \leq n$ let $Ave\ i = \frac{1}{n}\sum_{j=1}^{n} r_{w_j}(m_i)$. Then fix integer $k \geq 2$ and define the correlation ρ of the preference profile by

$$\rho = \frac{\sum_{i=1}^{n}(Ave\ i)^k - n(\frac{n+1}{2})^k}{\sum_{i=1}^{n} i^k - n(\frac{n+1}{2})^k} \tag{1}$$

Notice that we have a family of measures of correlation indexed by k, and we can choose one that predicts well the outcome of the men-propose G-S algorithm. The first term in the numerator of (1) measures the spread of the $Ave\ i$'s. The denominator and the second term in the numerator normalize ρ. The second term of the numerator is the spread of a perfectly uncorrelated profile in which $Ave\ i = \frac{n+1}{2}$ for every i. The first term of the denominator is the spread of a perfectly correlated profile, that is, one in which all n women have identical preferences. Thus, $\rho = 0$ for perfectly uncorrelated preferences and $\rho = 1$ for perfectly correlated preferences.

For men's preferences over women, ρ is defined similarly.

One caveat: there is one type of correlation that our measure does not detect. In the following example we might describe the preferences as locally determined. Suppose n is even and $w_1, w_2, \ldots, w_{\frac{n}{2}}$ rank the men in order $m_1, m_2, \ldots, m_n$, while women $w_{\frac{n}{2}+1}, w_{\frac{n}{2}+2}, \ldots, w_n$ rank them in order $m_n, m_{n-1}, \ldots, m_1$. Then $\rho = 0$ despite the obvious "local" correlation. It is possible to devise measures of correlation that detect local correlation, but for this study we will stay with our simple definition of ρ.

4. Generating Correlated Preferences.

Here is a very simple method that uses groupings of men to generate correlated women's preference profiles. If $n = 100$, we can write, for example (20, 10, 70) to indicate that the men are grouped into three sets: the first 20 men, the next 10 men and the final 70 men. Then for each woman men $m_1, m_2, \ldots m_{20}$ are randomly assigned rankings 1 through 20, men $m_{21}, m_{22}, \ldots, m_{30}$ are randomly assigned rankings 21 through 30, and men $m_{31}, m_{32}, \ldots, m_{100}$ are randomly assigned rankings 31 through 100. Then ρ can be calculated for the resulting preference profile using (1). In Table 1 we have calculated ρ with $k = 9$ for ten groupings. The results will be used in the next section in the comparison of satisfaction levels and correlation. The groupings were chosen to provide a reasonable

spread of correlations. Concerning the choice of $k = 9$, recall that our goal in (1) was to define a measure of correlation of preferences that could be used to predict men's and women's levels of satisfaction with the outcome of the G-S algorithm. Through trial and error we found that the performance of correlation as a predictor of satisfaction improves as k increases from its minimum value, 2, and this improvement is substantial until k reaches about 9, after which it is slight.

5. One Gender's Preferences Correlated, the other's Random.

In this section we investigate the effect of correlated preferences for one gender on the satisfaction levels of men and women. This special case is of interest because it is a transition between the simple and well-studied random preferences case and the more realistic correlated preferences case, and because, as mentioned in the introduction, there has been theoretical work done in this case to which our simulation results can be compared. With n fixed at 100, in this section we generated 200 correlated preference profiles for women for each grouping listed in Table 1. Then for each of the women's profiles generated we found ρ (with $k = 9$ in (1)) and also ran the men-propose G-S algorithm for that profile paired with a randomly generated men's profile. We ran a quadratic regression for the sum of women's rankings of their assigned mates averaged over each grouping against ρ averaged over each grouping. Results are given in the first column of Table 2. The result is illustrated in Figure 1. Figure 2 gives the result for sum of men's rankings of their assigned mates.

The intuition for Figure 1 is clear. Women's overall level of satisfaction decreases with increasing correlation of women's preferences because highly correlated preferences mean women are competing for the same men. It is harder to provide intuition for Figure 2, since increased competition among women does not translate in any obvious way to the observed (slight) increase in overall satisfaction for men. On the other hand, Figure 2 gives us further insight into Figure 1. The sharp decrease in women's overall satisfaction evident in Figure 1 is due in small part to a slight decrease in the number of proposals (see Figure 2 and notice that MS equals the total number of proposals made) and therefore in large part to decrease in quality of proposals–with high correlation a proposal by a man

who has been rejected previously is unlikely to be welcomed.

Next the entire process was repeated with men's preferences correlated and women's random yielding Figures 3 and 4. The intuition for Figure 4 is identical to that for Figure 1. Men's overall level of satisfaction decreases with increasing correlation because highly correlated preferences mean men are competing for the same women. The increasing overall level of satisfaction for women in Figure 3 is explained by the increase in the number of proposals (see Figure 4) unmitigated by a decrease in quality of proposals, since women's preferences are random in Figure 3.

6. Both Genders' Preferences Correlated.

Now we investigate the effect of correlation when both men's preferences and women's preferences are correlated. We again used $n = 100$. For each of the 36 ordered pairs of groupings from table 3, we constructed a women's preference profile using the first grouping and a men's preference profile using the second grouping. Then we ran the G-S algorithm for the profile pair. We carried out 200 trials for each ordered pair of groupings. We then ran a regression for the sum of women's rankings of their assigned mate averaged over each grouping against ρ_m and ρ_w averaged over each grouping. The results appear in column 2 of Table 4, which yields the regression equation $WS = 2,069 + 7,634\rho_w - 4,780\rho_w^2 - 661\rho_m$.

The results are illustrated in Figure 5 where the 36 data points are projected from the ρ_m-ρ_w-WS space into the ρ_m-WS plane, as are the six lines from the regression surface. For example the line labeled $\rho_w = .14$ is the line of intersection of the plane $\rho_w = .14$ with the regression surface, projected into the ρ_m-WS plane. This is the line that the regression surface gives as an approximation for the six data points represented by plus signs in Figure 5. One can also think of Figure 5 as the regression surface and data viewed from the ρ_w-axis.

Column 3 of Table 4 and Figure 6 illustrate the results of the regression of MS against ρ_m and ρ_w.

Notice that the approximations in Figures 5 and 6 are not as tight as in Figures 1-4, which is only natural since in Figure 5 for example we are trying to summarize a more gen-

eral phenomenon−−G-S matching when both genders have correlated preferences−−than in Figure 1.

Next notice that the findings in this section are consistent with the findings and intuitive explanations of the previous section.

Finally notice that the only coefficient that is not significant at the 10% level in Table 4 is the coefficient for whose sign we had no intuitive explanation in the previous section.

7. Concluding Remarks.

We use a method other than that of Calderelli and Capocci for generating correlated preference profiles because our method yields a wider variety of preference profiles, in the sense that two different groupings might have the same expected correlation, but different expected levels of overall satisfaction. In fact, after generating our figures using arbitrarily chosen groupings, we searched for a grouping that would provide data points well off our regression curves. We found that a grouping with a few very unpopular individuals, such as (97, 3), would result in significantly less satisfying outcomes for the gender so grouped than the regression curves predicted.

Given the complexity of the G-S algorithm, it is not surprising that two numbers, one for each gender, are not perfect predictors of aggregate satisfaction allowing no anomalies. We have mentioned two such anomalies, groupings with a few very unpopular individuals and locally determined preferences. Other properties of preferences that affect aggregate satisfaction will be dealt with in a later study.

References

1. Abdulkadiroğlu, A., Pathak, P. Roth, A. (2005) The New York City high school match. AEA Papers and Proceedings 95: 364-367.

2. Caldarelli, G., Capocci A. (2001) Beauty and distance in the stable matching problem. Physica A 300: 325-331.

3. Chen, Y., Sönmez, T. (2006) School choice: an experimental study. Journal of Economic Theory 127: 202-231.

4. Gale, D., Shapley, L.S. (1962) College admissions and the stability of marriage. American Mathematical Monthly 69: 9-15.

5. Haruvy, E., Roth, A. Unver, M.U. (2006) The dynamics of law clerk matching: An experimental and computational investigation of proposals for reform of the market. Journal of Economic Dynamics and Control 30: 457-486.

6. Knoblauch, V. (2006) Marriage matching: a conjecture of Donald Knuth. University of Connecticut working paper.

7. Mongell, S., Roth, A. (1991) Sorority rush as a two-sided matching mechanism. The American Economic Review 81: 441-464.

8. Roth, A., Peranson, E. (1999) The redesign of the matching market for American physicians: some engineering aspects of economic design. The American Economic Review 89: 748-780.

9. Wilson, L. B. (1972) An analysis of the marriage matching assignment algorithm. BIT 12: 569-75.

Table 1. Groupings used when only men's or women's lists are correlated.*

Grouping	Correlation
Random	0.00255
3,97	0.00817
10,90	0.02644
15,85	0.04458
10,10,80	0.06804
3,7,20,70	0.13535
10,10,20,60	0.23556
5,15,30,50	0.37171
10,20,30,40	0.53749
5,10,15,20,20,30	0.71528
15,15,15,15,15,15,10	0.94444
Perfect Corr.	1

*Correlation values are averages over two hundred profiles.

Figure 1. Women's Preferences Correlated, Men's Preferences Random

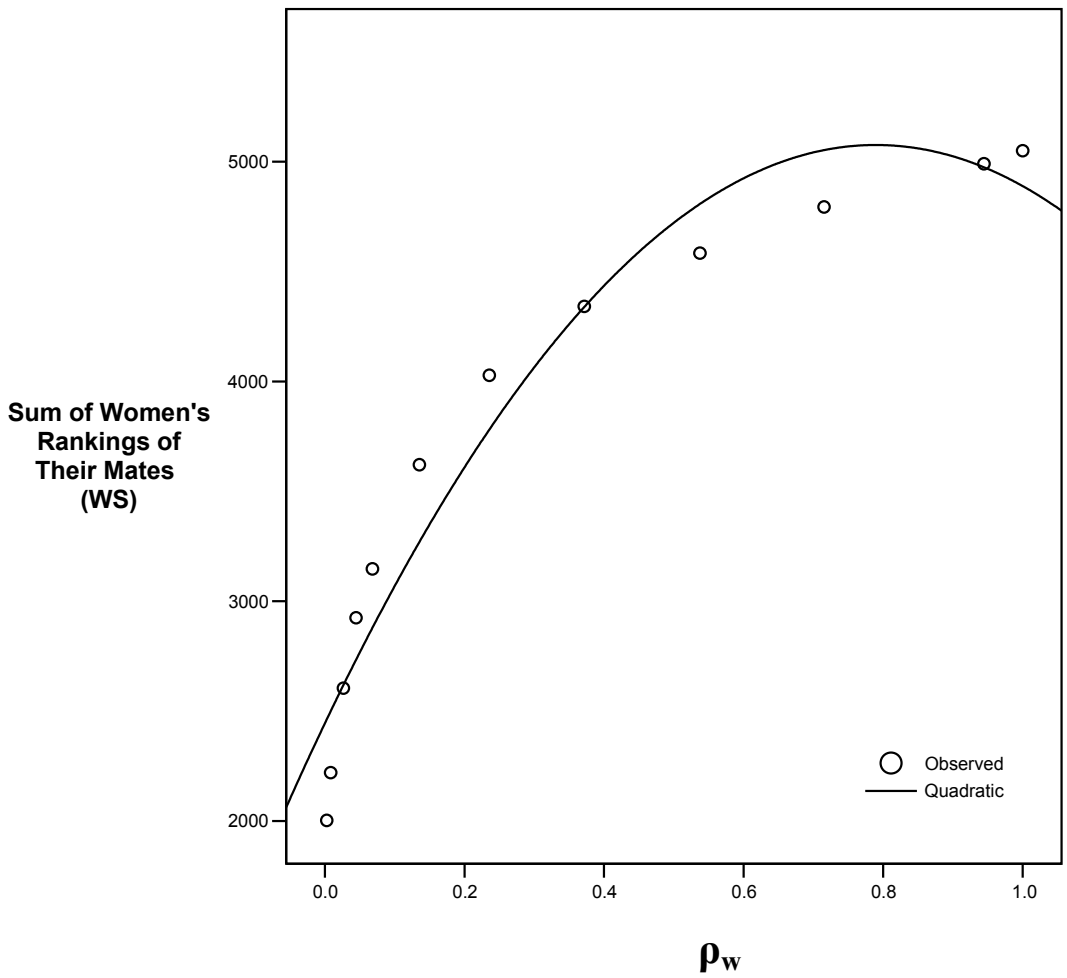

Figure 2. Women's Preferences Correlated, Men's Preferences Random

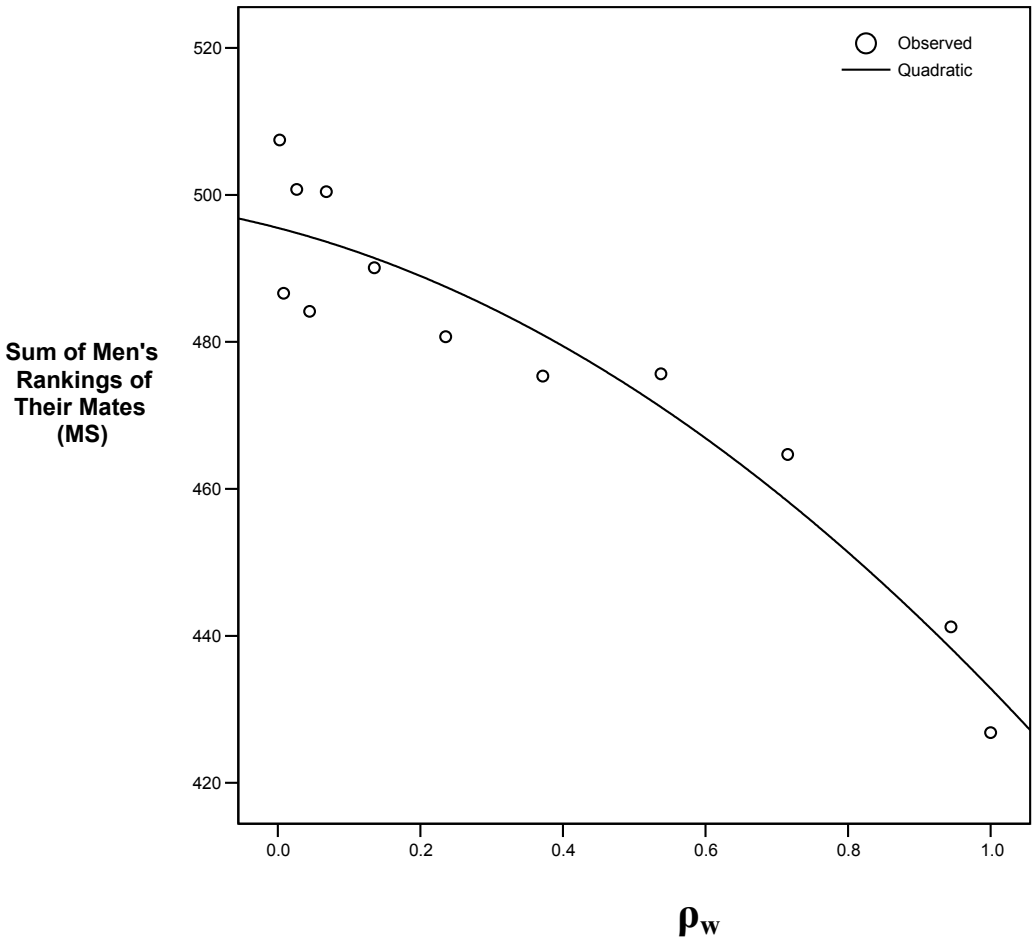

Figure 3. Men's Preferences Correlated, Women's Preferences Random

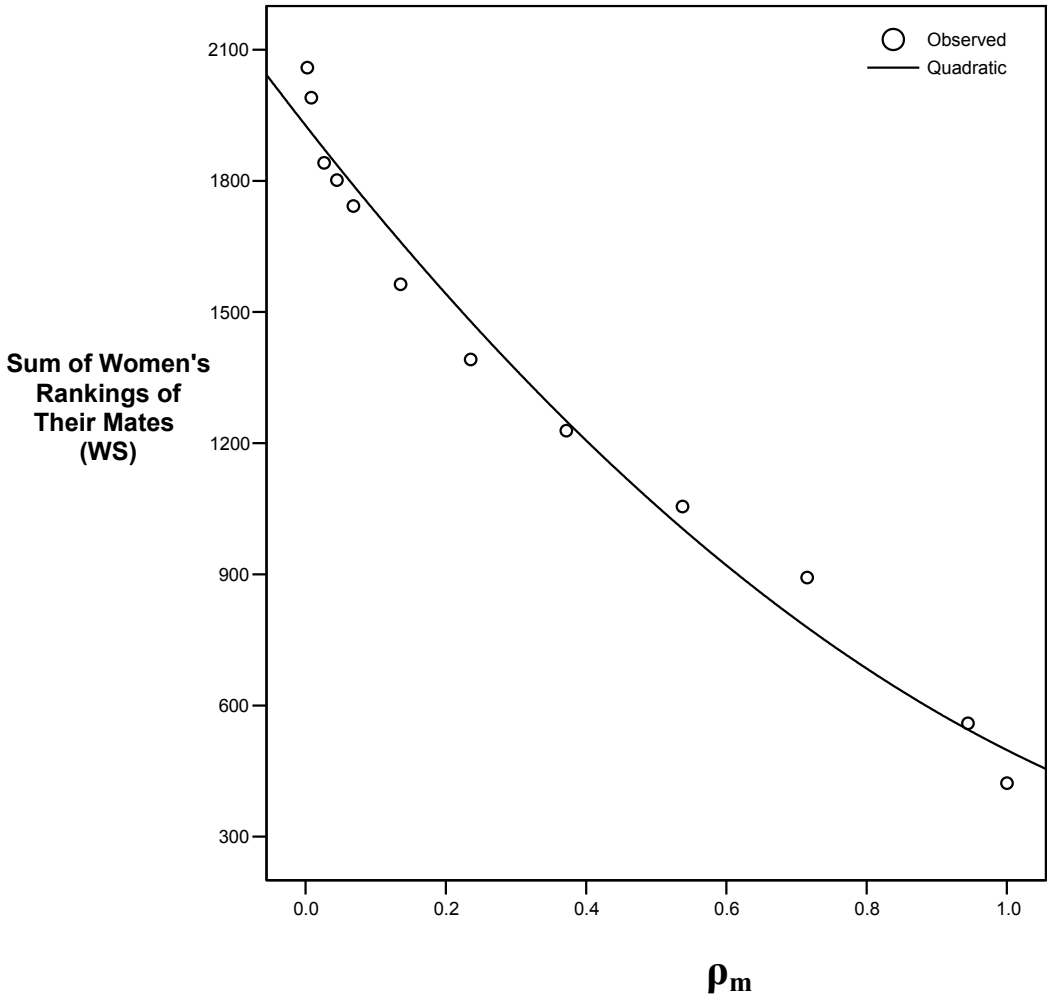

Figure 4. Men's Preferences Correlated, Women's Preferences Random

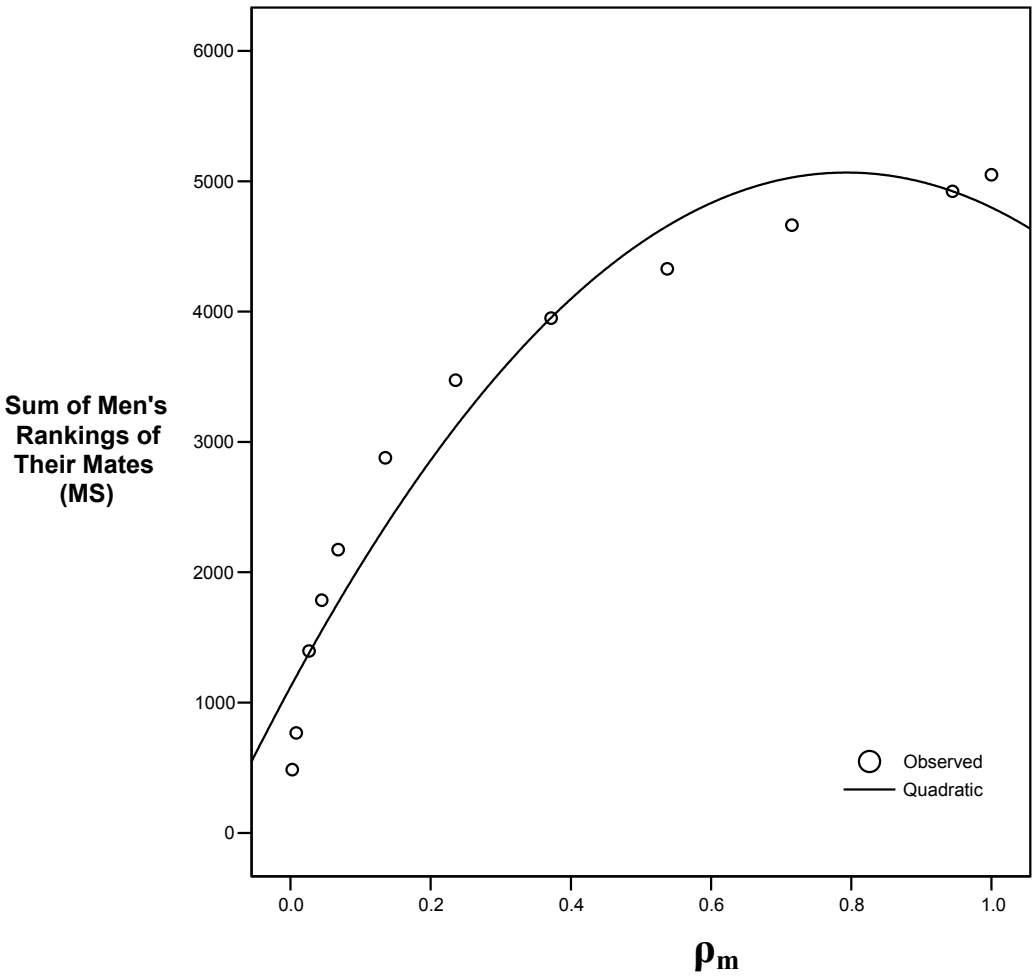

Table 2. Regression results

	Women's Preferences Correlated		Men's Preferences Correlated	
	WS	**MS**	**WS**	**MS**
Constant	2445.228*** (135.73)	495.410*** (3.83)	1926.587*** (41.86)	1120.161*** (198.64)
ρ	6663.873*** (909.01)	-24.498 (25.65)	-2049.251*** (280.38)	9951.800*** (1330.37)
ρ^2	-4219.784*** (927.94)	-38.029 (26.19)	620.435* (286.22)	-6273.645*** (1358.08)
R^2	.945	.908	.979	.948

*Significant at 10%
**Significant at 5%
***Significant at 1 %

Table 3. Groupings used when both men's and women's lists are correlated*

Grouping	Correlation
3,97	0.00817
3,7,20,70	0.13535
5,15,30,50	0.37171
10,20,30,40	0.53749
5,10,15,20,20,30	0.71528
15,15,15,15,15,15,10	0.94444

*Correlation values are averages over two hundred profiles.

Table 4. Regression results

Variable	WS	MS
Constant	2068.724*** (143.727)	1228.616*** (148.020)
ρ_w	7633.702*** (613.912)	- 295.806 (179.058)
ρ_w^2	- 4779.664*** (630.528)	-
ρ_m	- 660.538*** (173.865)	8672.615*** (632.251)
ρ_m^2	-	- 5231.072*** (649.363)
R^2	0.927	0.942

* significant at 10%
** significant at 5%
*** significant at 1%

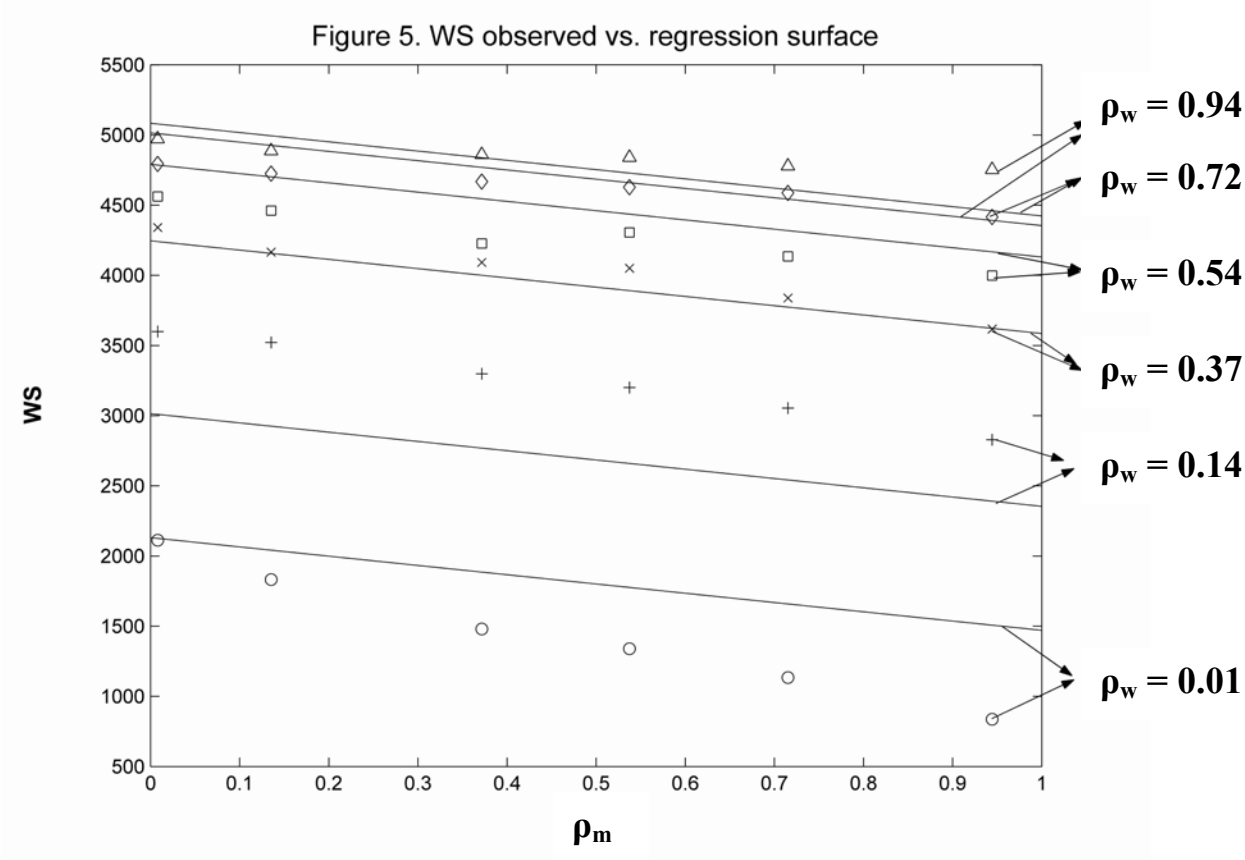
Figure 5. WS observed vs. regression surface
WS
5500
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
ρm
ρw = 0.94
ρw = 0.72
ρw = 0.54
ρw = 0.37
ρw = 0.14
ρw = 0.01
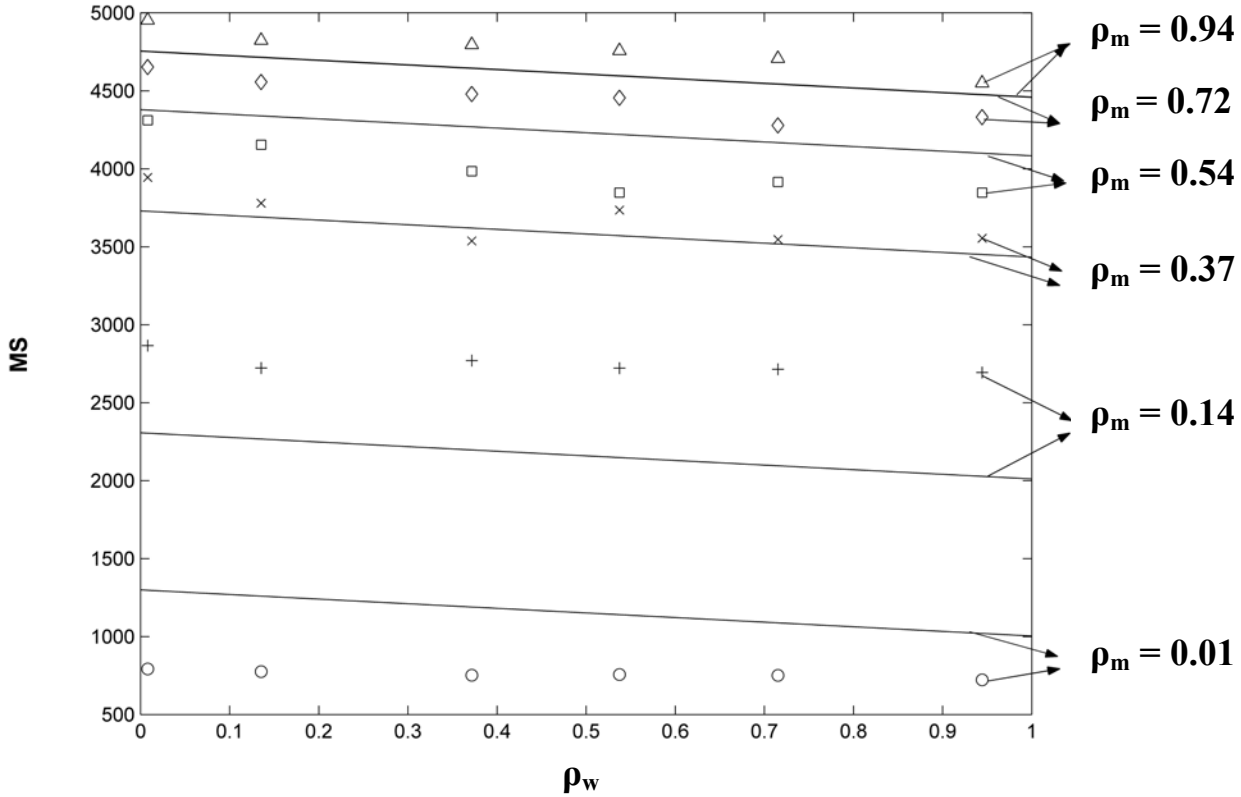
Figure 6. MS observed vs. regression surface