This paper takes Gary Becker's theory of marriage seriously. In his seminal paper, Becker [1973] proposes an invisible hand theorem for the marriage market. He argues that the competition for spouse leads men and women to be matched in such a way that maximizes the sum total of marital output. Applications of the economic approach to marriage are now commonplace. Becker et al. [1977] study the effect of imperfect information on divorce. Benham [1974], Scully [1979] and Wong [1986] estimate the effect of wife's education on husband's earnings. Grossbard-Shechtman [1993] and Rao [1993] focus on the effect of sex ratio on (implicit or explicit) bride prices or dowries. The literature, however, has virtually ignored Becker's hypothesis that marriage markets maximize total marital output, i.e., that marriage markets are efficient.

Mistakes are no doubt made in marriage decisions. The marriage market is not a textbook example of perfect competition because of elaborate social norms, substantial search costs, and room for bargaining and opportunistic behavior. Nevertheless, we intend to push Becker's efficient marriage market model at face value, and directly confront it with data. The hypothesis that the marriage market maximizes marital output provides a framework that allows estimating a model of spouse selection which recovers some of the parameters of the production function for marital output. Using the estimated parameters, we can use programming techniques to find the optimal male-female pairing that maximizes the sum of marital output. The degree to which the actual pairing of husbands and wives corresponds to the optimal pairing provides a goodness-of-fit test of the efficient marriage market hypothesis. Our method is applied to census data in Hong Kong.

A model of spouse selection (and other matching models) differs from a discrete choice model (e.g., McFadden [1984]) in two important ways. First, the choice of a spouse is mutual. Not only does each man choose a woman that yields him the highest net utility, but each woman also picks a man that gives her the highest net utility. Second, in a discrete choice model, several individuals can choose the same alternative. In marriage markets, this is prohibited under monogamy. Thus new methods have to be developed to analyze choice in marriage markets. Although marital output and the shadow price of husbands and wives are not observable, the theory of optimal matching allows us to write down the likelihood function for the observed choice of marital partners. It turns out that this model of spouse selection reduces to a version of the Tobit model under the assumption that error terms are independently and identically normal. In multinomial logit models the interpretation of parameters is cumbersome, especially when the number of alternatives is large. In our model of spouse selection, the estimated parameters are directly related to the production function of marital output. The signs of these parameters can be used to address the issue of positive versus negative assortative mating. The model can also be used to test marginal product pricing in the marriage market.

With the estimated production function from the model of spouse selection, we can impute the marital output for all possible male-female pairs. Using these imputed output figures, the optimal pairing between men and women that maximizes the sum of marital output can be found using an algorithm that solves the assignment problem. The degree to which this optimal pairing corresponds to the observed pairing then provides a natural way of assessing marriage market efficiency. Given the limited range of information contained in the census, we find that our model of spouse selection fits the observed choices better than other models (such as one based on ranking method). Following Varian's [1990] proposal for goodness-of-fit tests for optimizing models, we derive a measure of efficiency which is based on the distance between total imputed marital output under observed matching and that under optimal matching. This efficiency index is calculated to be about 80 percent, which is 27 standard deviations greater than the mean efficiency index under random assignment.

Previous empirical work on the choice of marriage partners are mostly descriptive in nature. The issue of positive versus negative assortative mating has attracted attention from demographers, psychologists, sociologists, and economists alike (e.g., Winch [1958]; Epstein and Guttman [1984]; Mare [1991]). Most of these studies rely on cross tabulations (when the personal traits are qualitative) or on simple correlations (when the traits are quantitative). Our approach can deal with qualitative and quantitative traits, as well as same-trait interactions and cross-trait interactions, under a unified framework. More importantly, tools such as tabulations and correlations are *ad hoc* and are inadequate for hypothesis testing. By focusing on the correlation of traits between married couples, these methods ignore the matrix of potential couples in a marriage market. The fact that two people can—but choose not to—marry each other contains information about the underlying production function of marital output. Our model makes use of the full information contained in the data as specified by the theory of efficient matching. Interpretation of parameters is straightforward and hypothesis testing is well integrated with the underlying theory.

The theory behind this paper is not new. We consider the matching of husbands and wives as an optimal assignment problem (see Koopmans and Beckmann [1957] and Becker [1973]). In taking this track, we implicitly assume price taking behavior and ignore the bargaining issues associated with matching. Alvin Roth and others (e.g., Roth and Sotomayor [1990]; Shapley and Shubik [1972]; Gale and Shapley [1962]) have studied extensively the game theoretic aspect of matching problems with and without transferable utility. Manser and Brown [1980], McElroy and Horney [1981] and Lundberg and Pollak [1993] discuss intrafamily allocation problems from a Nash bargaining perspective. Cohen [1987] and Allen [1992] analyze the problems that arise when transaction costs within marriage are significant. The extent to which such problems affect the conclusions of the competitive model, however, is an empirical issue. This paper intends to determine how well a pure competitive model corresponds to observed marital patterns.

Research on marriage using the optimal assignment approach is sparse. Bergstrom and Lam [1994] were the first to apply linear programming methods to the empirical study of marriage. Their paper focuses on matching by age and its relationship to the "marriage squeeze." The full implications of an efficient marriage market are not explored. Foster and Khan [1994] and Foster [1996] develop an empirical model of marriage market equilibrium that relies on the notion that couples in a marriage market do not want to exchange partners. Their papers are similar in spirit to ours, but they do not address issues relating to pricing and efficiency in the marriage market.

I. AN EMPIRICAL MODEL OF SPOUSE SELECTION

In this section we briefly review the theory of efficient marriage markets, with an eye on empirical implementation. The theory is a direct application of the optimal assignment problem in operations research, which is in turn a special case of linear programming. More detailed discussions can be found in Koopmans and Beckmann [1957], Shapley and Shubik [1972], Becker [1973], and in most texts on linear programming. The main innovation in this paper is in the econometric implementation of the theory, rather than in the theory as such.

We focus on the selection of spouse rather than the decision to get married. Suppose there are n women and n men in the marriage market. The ith woman is characterized by a vector of characteristics $F_i = (F_{i1}, \ldots, F_{iK})$, and the ith man has characteristics $M_i = (M_{i1}, \ldots, M_{iK})$. If woman i and man j get married, their marital output will be given by the production function:

$$Z_{ij} = Z(F_i, M_j, u_{ij}).$$

Here u_{ij} refers to the effect of random disturbances that are unobservable by the econometrician but are known to the participants in the marriage market (e.g., "personal chemistry"). We assume the u_{ij}'s are independent for all i and j.[1] Marital output Z_{ij} need not be material output, and it is unobservable.

To be more specific, let the production function take the form:

$$Z_{ij} = \alpha + \sum\nolimits_k \beta_{0k} F_{ik} + \sum\nolimits_k \gamma_{0k} M_{jk} + \sum\nolimits_k \delta_k F_{ik} M_{jk} + u_{ij}.$$

This production function exhibits complementarities if the δ coefficients are positive. When this is the case, people with similar characteristics will be matched together in an efficient market (see also Sattinger [1975] and Kremer [1993]). For example, let characteristic k refer to years of schooling, and let δ_k be positive, normalized to one. Suppose there are two men in the market, one with 16 years of schooling and the other with 12 years. Similarly, let the two women in the market have 16 and 12 years of schooling respectively. If men and women with similar education marry each other (i.e., positive assortative matching), total marital output will be $256 + 144 = 400$. If the college educated woman marries the high school man (negative assortative matching), total marital output will be $192 + 192 = 384$. An efficient marriage market will then match men and women with similar characteristics. Conversely, if $\partial^2 Z_{ij}/\partial F_{ik}\partial M_{jk} = \delta_k$ is negative, corresponding male and female traits are substitutes in the production of marital output. One would then expect to see negative assortative matching.

An efficient marriage market is one which maximizes the sum of marital output subject to the constraint that each woman marries only one man and each man marries only one woman.[2] Without loss of generality, we arrange the women and men in the sample in such a way that the ith woman and the ith man are married couples. A necessary and sufficient condition for efficiency is that there exist vectors $P^m = (P_1^m, \ldots, P_n^m)$ and $P^f = (P_1^f, \ldots, P_n^f)$ such that

$$P_i^f + P_i^m = Z_{ii}, \qquad i = 1, \ldots, n;$$
$$P_i^f + P_j^m \geq Z_{ij}, \qquad i \neq j.$$

The P^f and P^m vectors are the shadow values for women and men in the competitive marriage market.[3] The equality restrictions state that the sum of shadow values for a couple is equal to their marital output. The inequalities imply that no "blocking coalition" exists in the marriage market equilibrium. If the inequality restriction is not satisfied for some pair of man and woman who are not a couple, they can be made better off by marrying to each other and dividing the marital output in such a way that both get more income than their existing shadow values.

We assume the distribution of personal traits is continuous and sufficiently dense so that the shadow prices are uniquely determined.[4] Given that woman i commands a price of P_i^f, the net marital output for man i who wants to marry her will be $Z_{ii} - P_i^f = P_i^m$. If man i marries another woman, say j, his net marital output will be $Z_{ji} - P_j^f < P_i^m$. Thus, given the efficiency conditions shown above, man i will maximize his net income by choosing woman i as his spouse. Similarly the ith woman will maximize her net marital output by marrying the ith man. The competitive marriage market with its set of shadow prices works like an invisible hand that ties men and women together in such a way that maximizes total marital output.

The shadow prices are unobservable to the econometrician. One way of estimating them is to treat them as individual fixed effects. This will be attempted in Section IV. However, estimating a large number of fixed effect parameters is not always practical. One way of simplifying the model is to treat the shadow prices as hedonic functions of individual characteristics. Specifically, we let

$$P_i^f = \alpha^f + \sum_k \beta_{1k} F_{ik} + \sum_k \beta_{2k} F_{ik}^2;$$
$$P_i^m = \alpha^m + \sum_k \gamma_{1k} M_{ik} + \sum_k \gamma_{2k} M_{ik}^2.$$

The quadratic terms are designed to capture the non-linear effects arising from assortative matching (e.g., Rosen [1982]; Kremer [1993]). For example, if there is positive assortative matching in education, shadow prices will be convex in years of schooling.

Notice that the shadow price of a person in the marriage market only depends on his or her personal characteristics, but not on the identity of his or her spouse.[5] As in Rosen's [1974] model of competitive market for differentiated products, the price of a product depends on its attributes but not on the identity of the buyer.[6] Bargaining and small numbers problems are assumed away. Also note that we have not included a random term in the hedonic price functions. Doing so will introduce complicated interdependence into the model so that the resulting log-likelihood function will not be manageable. This is remedied in Section IV, where we estimate fixed effects models for random sub-samples of smaller size.

Given our specification of the production function and of the price functions, the conditions for market efficiency can be rewritten as:

$$-(a + b_1 F_i + b_2 F_i^2 + c_1 M_i + c_2 M_i^2 + dFM_{ii}) = u_{ii}, \qquad i = 1, \ldots, n;$$
$$-(a + b_1 F_i + b_2 F_i^2 + c_1 M_j + c_2 M_j^2 + dFM_{ij}) > u_{ij}, \qquad i \neq j.$$

In the above expression, $a = \alpha - \alpha^f - \alpha^m$, $b_1 = \beta_0 - \beta_1$, $b_2 = -\beta_2$,$c_1 = \gamma_0 - \gamma_1$, $c_2 = -\gamma_2$, and $d = \delta$. F_i^2 and M_i^2 are the vectors of the squares of female and male attributes. FM_{ij} refers to the vector of interaction terms. The products are understood to be inner products.

We assume the u_{ij}'s are independently distributed with density function $\phi(\cdot)$ and distribution function $\Phi(\cdot)$. Thus the log-likelihood function for the sample observations is:

$$\sum_i \log \phi(-(a + b_1 F_i + b_2 F_i^2 + c_1 M_i + c_2 M_i^2 + dFM_{ii})) + \sum_{i \neq j} \log \Phi(-(a + b_1 F_i + b_2 F_i^2 + c_1 M_j + c_2 M_j^2 + dFM_{ij})).$$

If u is normally distributed, this log-likelihood function is equivalent to that of a Tobit model. Let the dependent variable be $y_{ij} = \max\{0, a + b_1 F_i + b_2 F_i^2 + c_1 M_j + c_2 M_j^2 + dFM_{ij} + u_{ij}\}$. If y_{ij} is censored for $i \neq j$ and is uncensored for $i = j$, and if y_{ij} is identically zero, the resulting log-likelihood function is identical to our model of spouse selection. Thus our model can be estimated by simply running a Tobit regression with zero as the dependent variable and with F_i, M_j, F_i^2, M_j^2 and FM_{ij} as independent variables.

There are two differences from the conventional Tobit model, however. First, since the uncensored variables are identically zero, the variance of the error distribution cannot be identified. In other words, all the estimated parameters are unique up to a multiplicative constant. In what follows, we normalize the variance to one. Second, in a sample of n couples, the log-likelihood function consists of the sum of n^2 terms. It is as if one were to estimate the Tobit model on n^2 observations.

It can also be observed that the production function parameter β_0 and the hedonic price function parameter β_1 cannot be separately identified. Similarly only the difference $c_1 = \gamma_0 - \gamma_1$ is identifiable. The production parameter ($\delta = d$), and the quadratic terms in the price function ($\beta_2 = -b_2$, $\gamma_2 = -c_2$), on the other hand, are identifiable (up to a multiplicative constant). Thus estimators of d can be used to test hypotheses concerning positive or negative assortative matching. Estimators of b_2 and c_2 can be used to assess the concavity or convexity of the hedonic price functions.

We can further use the model to test marginal productivity theory in the marriage market. In our model, the shadow price of attribute k for the jth man is $\gamma_{1k} + 2\gamma_{2k} M_{jk}$. The marginal product of this attribute for this man is $\gamma_{0k} + \delta_k F_{jk}$. The theory of marginal productivity suggests that the price of an attribute is equal to its marginal product. Unless matching is perfect, however, this restriction is unlikely to be true for each and every couple in the sample. We test a weaker version of marginal product pricing by summing the equality restriction over all couples. Denote the sample mean values of F_{jk} and M_{jk} by $F_{.k}$ and $M_{.k}$, respectively. Then, in terms of the parameters that can be estimated, the implied restriction on the model coefficients can be written as

$$c_{1k} + 2c_{2k}M_{.k} + d_k F_{.k} = 0,$$

for all attributes k. This restriction will be tested both for men and for women.[7]

II. DESCRIPTION OF THE DATA

The empirical model of spouse selection developed in this paper is estimated using census data from Hong Kong. We choose to work on Hong Kong data mainly because of convenience and familiarity. As in any industry study, determining the appropriate extent of the market is fraught with difficulties. We simply assume the entire territory of Hong Kong constitutes the relevant marriage market.[8]

Data for this study are drawn from the ten percent random sample of the Hong Kong 1976 population by-census. More recent censuses do not contain information about the year of marriage and are less suitable for our purposes. We identify all couples who were married a year before the census and who were living in the same household at the time of the census. People who were not born in Hong Kong or China are excluded. This gives a sample of 2110 couples.

We focus on people who were married in 1975 for two reasons. First, our model assumes anyone in the marriage market can marry anyone else if he or she so chooses. If the data contain, say, a man married in 1970 and a woman married in 1975, the man would have to first obtain a divorce before he could legally marry that woman. One could argue that, *ex ante*, the man has the option to delay marriage; that is, before 1970, the man and the woman were in the same marriage market (unless the woman was a minor at that time). However, this kind of argument will bring the timing of marriage into the model and complicate the analysis.[9] We choose instead to consider the marriage market as consisting of men and women who were married within some relatively short period of time (one year). The second reason for excluding people who were married more than one year before the census is that we have information on wages at the time of the census but not at the time of marriage. Wages in 1976 are a better indicator for wages in 1975 than they are for wages in, say, 1970. Focusing on the newly married therefore has the advantage that the wage data will more accurately reflect the information available at the time marriage decisions were made.

To examine the issue of assortative matching in wage, we further select only those records with wage information for both husband and wife. The resulting group of 772 couples will form the main data set of our empirical work. The main difference between this sample and the full sample of 2110 couples is that women with wage information on average received one more year of education than women in the full sample. The labor force participation rate for newly married women (.44) is between that for all married women aged 16–30 (.35) and for single women in the same age group (.69).

Many personal characteristics that are relevant for marriage decisions are not available from the census. This study concentrates on four observable characteristics: age at marriage, years of schooling, place of birth, and market wage rate. The variables `age` and `school` measure the first two characteristics. The variable `china` is a dummy variable indicating whether the person was born in China ($\texttt{china} = 1$) or born in Hong Kong ($\texttt{china} = 0$).

The construction of the wage variable requires some explanation. There are systematic differences in wage rates during the life cycle. For example, on average, a 25 year old man is expected to earn less than a 35 year old. This does not necessarily make the 25 year old man a less desirable spouse because it is the present value of lifetime earnings that counts in marriage decisions. For this reason, the variable `wage` is obtained by standardizing the observed log wage at the mean age of marriage. For males, we run an OLS regression of log wage on schooling, place of birth and a fourth order polynomial on age. If the vector of age variables is denoted A and the vector of corresponding estimated coefficients in the wage regression is denoted $\hat{\eta}$, the `wage` variable is defined as

$$\log(\text{observed wage}) - \hat{\eta}A + \hat{\eta}\bar{A},$$

where $\bar{A}$ is the vector of age variables evaluated at the sample mean. The definition of the `wage` variable for women follows that for men, except that $\hat{\eta}$ is computed from a selection-bias corrected regression instead of by ordinary least squares.[10]

Table I shows some descriptive statistics of the various variables. Numbers in the bottom row refer to the means and standard deviations of male characteristics, and numbers in the marginal column refer to female characteristics. For example, the husbands in this sample were on average 3.3 years older than the wives, and had .7 more year of education. The main elements in the cross table are correlation coefficients between male and female characteristics. For example, the simple correlation between husbands' age and wives' age is .578. All the diagonal entries are positive, indicating that there is positive assortative matching on each attribute.

Notice, however, that correlation coefficients are reduced-form statistics. For example, suppose `f_school` and `m_school` are complementary in the production of marital output, while `f_wage` and `m_wage` are substitutes. If more educated people tend to earn more, and if the complementarity between education is much stronger than the substitutability between wage, we will observe that more educated (higher wage) men are married to more educated (higher wage) women. The observed correlation between `f_wage` and `m_wage` would then be positive even though these two attributes are substitutes. The table of correlation coefficients in Table I, therefore, does not reveal the underlying substitutability and complementarity relationships in the production function for marital output.

III. TESTING ASSORTATIVE MATCHING AND MARGINAL PRODUCT PRICING

Our model of spouse selection is estimated with (Model W) and without (Model NW) the wage variable. The parameter estimates are displayed in the first two columns of Table II. The coefficients on female traits and on male traits represent linear combinations of production function and price function parameters. Theory does not place any restrictions on their sign or statistical significance. The main parameters of interest are those associated with the interaction terms. In each specification, the coefficients for the same-trait interaction terms are positive and highly significant. This indicates that corresponding male and female traits are complements in the production function for marital output. Such results are consistent with positive assortative matching in age, schooling, place of birth, and wage.

Comparing Model W to Model NW, the model that includes wage variables gives a significantly higher log-likelihood ratio. Earnings power seems to be an important consideration in marriage decisions. Although Becker [1973] predicts negative assortative matching in wage as a result of the household division of labor, our findings do not support his hypothesis. Lam [1988] offers a theoretical reason why Becker's prediction may fail in the presence of household public goods. Our empirical results are also consistent with Watkins and Meredith [1981], who report a positive correlation between husband's and wife's income.[11]

The coefficients on the squared terms in Table II are all negative and statistically significant. This means the shadow prices are convex functions of the personal attributes (since $b_2 = -\beta_2$ and $c_2 = -\gamma_2$). The convexity of the price functions is consistent with complementarities in the production function of marital output.

Cross-Trait Interactions

The specification of the marital output production function need not be confined to same-trait interactions. There may exist complementarity relations between husband's education and wife's age, or between wife's education and husband's age. Such cross-trait interactions need not be symmetric either. In fact the set of female characteristics used in the estimation does not have to be the same as the set of male characteristics. In other applications of the assignment model, such as the matching of houses to homeowners, positive or negative assortative matching is not an issue because the characteristics of a house cannot be directly compared to the characteristics of a homeowner. Yet the matching model developed here is still applicable. If there are K_1 relevant attributes on one side and K_2 relevant attributes on the other side, a full model will involve $K_1 K_2$ parameters to be estimated for the interaction effects.

The last two columns of Table II shows the estimated coefficients of the full interactions model with and without wage variables. For Model NW-C, 4 out of 6 coefficients for the cross-trait interaction terms are statistically insignificant (at the .01 level). The likelihood ratio test fails to reject the hypothesis that the cross-interaction effects are jointly zero. Although the hypothesis that the cross effects are jointly zero is rejected for Model W-C, 9 out of 12 coefficients for the cross-trait terms are not significant individually. Introducing cross-interaction effects does not substantially improve the fit of the model.[12] Moreover, the coefficient estimates for same-trait interaction terms do not change much across specifications; it is only the coefficients that are imprecisely estimated that change quite a bit. We therefore prefer to work with the simpler same-trait interactions model. In the remainder of the paper, the empirical work will focus only on Models W and NW.

Robustness

Our sample of 772 couples does not include those for which wage information for either husband or wife is not available. To make sure that our results are robust to sample selection rule, we have re-estimated Model NW using the records of all 2110 newly married couples. The estimated coefficients are not materially different from those obtained from the smaller sample. For example, the coefficients (standard errors) associated with the same-trait interaction terms for `age`, `china`, and `school` are 0.010 (0.001), 0.288 (0.030), and 0.020 (0.001), respectively. These estimates are fairly close to those reported in in first column of Table II.

The `wage` variable used in this study has been adjusted for life-cycle differences in wage. We have also estimated Model W where the observed log wage is used instead of the standardized wage. It turns out that using the raw wage instead of the adjusted wage does not change any qualitative results. In particular, our conclusion about positive assortative matching in wage still holds.

Finally we have checked the robustness of the model against alternative functional forms for the error distribution. If the error terms are not normal, the model of spouse selection can no longer be estimated by a Tobit regression. Nevertheless the likelihood function can still be regarded as one arising from a survival type model (e.g., Cox and Oakes [1984]) involving censored and uncensored observations. We have estimated the model using the extreme value distribution, the logistic distribution and the gamma distribution. Again, the results regarding positive assortative matching continue to hold. Our model of spouse selection is not very sensitive to the assumed distribution of the error term.[13]

Marginal Product Pricing

The theory of marginal product pricing requires that the marginal product of a personal attribute to be equal to its marginal value in the marriage market. In terms of the parameters of the model, the theory applied to women requires that $b_{1k} + 2b_{2k}F_{.k} + d_k M_{.k} = 0$, and the theory applied to men requires $c_{1k} + 2c_{2k}M_{.k} + d_k F_{.k} = 0$. We impose these restrictions separately and jointly to our model of spouse selection. Table III shows the results.

The coefficient estimates shown in Table III do not differ much from those in Table II. Formal statistical tests do not reject the equality restrictions imposed by the theory of marginal productivity. Under the null hypothesis, two times the difference in the value of the log-likelihood function is distributed as a chi-squared random variable with degrees of freedom equal to the number of equality restrictions imposed. For example, the χ^2-statistic for models W-M&F is 2.15 with 8 degrees of freedom. The critical χ^2-statistic at the .01 significance level is 20.09. As shown in the bottom row of Table III, the χ^2-statistics are all smaller than the corresponding critical values at conventional levels of statistical significance. The hypothesis that personal traits are correctly priced in the marriage market is not rejected.

IV. FIXED EFFECTS MODEL

One problem with the empirical model of spouse selection estimated in the previous section is that it ignores unobserved individual fixed effects. This problem is particularly important when census data is used, because the census only contains very limited information relevant to marriage decisions. Individual characteristics such as personality, wealth, family background and "looks" are probably more important determinants of marital choice than the basic demographic variables (age, schooling, place of birth, wage) used in our estimation. Such characteristics are unobservable to the econometrician but are observable to participants in the marriage market, and they affect marital output as well as shadow prices. In this section we attempt to control for unobserved heterogeneity using a fixed effects model.

We let marital output be a function of observable characteristics as well as individual fixed effects:

$$Z_{ij} = \mu_i^{fz} + \mu_j^{mz} + \beta F_i + \gamma M_j + \delta FM_{ij} + u_{ij}.$$

Shadow prices $(P_1^f, \ldots, P_n^f)$ and $(P_1^m, \ldots, P_n^m)$ are simply specified as fixed constants. The condition for marriage market efficiency becomes:

$$\begin{aligned} -(\mu_i^f + \mu_i^m + \beta F_i + \gamma M_i + \delta FM_{ii}) &= u_{ii}, \qquad i = 1, \ldots, n; \\ -(\mu_i^f + \mu_j^m + \beta F_i + \gamma M_j + \delta FM_{ij}) &> u_{ij}, \qquad i \neq j; \end{aligned}$$

where $\mu_i^f = \mu_i^{fz} - P_i^f$ and $\mu_j^m = \mu_j^{mz} - P_j^m$. With these fixed effects, the parameters β and γ will not be identified and the log likelihood function can be reduced to:

$$\sum_i \log \phi(-(\mu_i^f + \mu_i^m + \delta FM_{ii})) + \sum_{i \neq j} \log \Phi(-(\mu_i^f + \mu_j^m + \delta FM_{ij})).$$

This log likelihood function can again be maximized using a Tobit regression.

With a sample of 772 couples, estimating a fixed effects model would require estimating more than 1500 parameters. To deal with the computational problem, we resort to a bootstrap procedure (see, for example, Efron and Tibshirani [1993]). We randomly draw (without replacement) a sub-sample of 100 couples from the original sample.[14] A Tobit model is estimated for this sub-sample, which involves 200 rather than 1544 fixed effects. The coefficient estimates of δ are saved. This procedure is repeated one thousand times. The means of the saved values of δ will give the point estimates of δ and the standard deviations of these saved values will give the standard errors.

The second and fifth columns in Table IV shows the bootstrap estimates of the parameters in the fixed effects model with (Model W-FE) and without (Model NW-FE) the wage variable. The coefficients for the same-trait interaction terms are all positive and significant, indicating that age, schooling, place of birth, and wage of husband and wife are complementary inputs to the production of marital output. The introduction of individual fixed effects therefore does not change our conclusions about assortative matching.

Our original hedonic pricing model estimated in Section III (i.e., without fixed effects) can be regarded as a special case of the fixed effects model where the individual fixed effects are constrained to be of the form:

$$\mu_i^f = a^f + \sum_k b_{1k} F_{ik} + \sum_k b_{2k} F_{il}^2;$$
$$\mu_j^m = a^m + \sum_k c_{1k} M_{jk} + \sum_k c_{2k} M_{jk}^2;$$

where $a^f + a^m = a$. We estimate the hedonic pricing model on the same 1000 random sub-samples and the results are displayed in the first and fourth column of Table IV. Comparing Model NW to Model NW-FE, the largest χ^2-statistic in these 1000 sub-samples is 9.54. The critical χ^2-statistic (189 degrees of freedom) at the .01 level is 237.15. In other words, in none of the 1000 random sub-samples does the fixed effects model give a significantly better fit to the data than the hedonic price model. Similarly, comparing Model W to Model W-FE, two times the difference in log-likelihood never exceeds 17; while the corresponding critical χ^2 value is 233 (185 degrees of freedom). This indicates our original hedonic pricing model is a good and parsimonious way of representing the marriage market; introducing fixed effects into the model does not substantially improve the fit.

We also use the bootstrap method to estimate the model with restrictions imposed by marginal productivity theory. The results are shown in columns 3 and 6 of Table IV. In all the 1000 bootstrap runs, marginal product pricing is not rejected. This agrees with our earlier conclusion reached using the full sample of 772 couples.

V. PREDICTING MARRIAGE PARTNERS

Is the efficient marriage market model useful in predicting actual marriage patterns? We approach this question by using the estimated spouse selection model to compute the optimal pairing of men and women that maximizes the sum of imputed marital output. Such optimal pairing is then compared to the observed pairing. We find that the efficient market model correctly predicts more married couples than either random pairing or pairing based on ranking.

In our model of spouse selection, the estimated coefficients cannot be used to give an imputed marital output, $\hat{Z}_{ij}$, because some of the production function parameters are not identified. Fortunately the theory of optimal assignment (see Koopmans and Beckmann [1957]) establishes that any assignment that maximizes $\sum \hat{Z}_{ij}$ also maximizes $\sum(\hat{Z}_{ij} + \lambda_i^f + \lambda_j^m)$, for any fixed vectors $(\lambda_1^f, \ldots, \lambda_n^f)$ and $(\lambda_1^m, \ldots, \lambda_n^m)$. Since the index variable, I_{ij}, from the Tobit regressions (with or without fixed effects) only differs from $\hat{Z}_{ij}$ by two fixed vectors, we can solve the optimal assignment problem by maximizing $\sum I_{ij}$ instead of $\sum \hat{Z}_{ij}$.[15]

The actual solution of an optimal assignment problem is non-trivial, but there are efficient computer algorithms that can be used. We rely on the Fortran subroutine in Burkard and Derigs [1980]. We save the predicted index variable I_{ij} from the Tobit regressions and use Burkard and Derigs's program to find the optimal pairing between men and women that maximizes $\sum I_{ij}$. Of the n optimal pairs computed by the program, if k pairs are actual married couples, the model is said to make k correct predictions.

One way to assess a model which gives k correct matches out of n pairs is to find the probability that such an event will occur by pure chance. Let $p(n, k)$ be the probability of having k married couples when n husbands and wives are randomly re-matched.[16] When n is moderately large (greater than 20, say), the distribution of k is well approximated by a Poisson distribution with parameter 1. That is,

$$p(n, k) \approx e^{-1}/k!.$$

Two features about the distribution of k are worth mentioning. First, for large n, the approximate distribution of k is independent of n. For example, the probability of predicting one married couple from one hundred randomly re-shuffled couples is approximately the same as the probability of predicting one married couple from one thousand randomly re-shuffled couples. Second, the mean and variance of k are both equal to 1. On average, therefore, randomly matching n pairs will give 1 correct prediction regardless of n.

The first column in Table 5 gives the exact distribution of k for $n = 100$. Note that if the matching of men and women is random, the probability of finding no correct match is non-trivial (.368), while the probability of finding 10 or more matches is almost nil. In the remaining four columns of Table V, we display the empirical distributions of the number of correct predictions using the efficient market model. For each Tobit regression that we run on the 1000 sub-samples discussed in Section 4, we record the number of correct predictions generated by Burkard and Derigs's optimal assignment program. For example, the model using the variables `age`, `school` and `china` without fixed effects yields on average 5.34 correct predictions. Under random matching the probability of giving 5 or more correct predictions is less than 1 percent, while the model gives more than 5 correct matches in 605 out of 1000 trials. The fixed effects model produces similar results as the model without fixed effects. When the variable `wage` is added to the spouse selection model, the average number of correct predictions increases to 6.70 (without fixed effects) and to 6.75 (with fixed effects), respectively. The spouse selection model clearly gives more correct predictions than what pure chance would suggest.

While predicting 6 married couples out of 100 may seem trivial, it is not clear what the appropriate metric is for assessing whether a certain number of correct predictions is "large" or "small." With 100 men and 100 women, there are $100! \approx 9.3 \times 10^{157}$ possible permutations. Table V shows that the chance of making 6 or more correct predictions is less than one in a thousand. Predicting 6 correct matches on average is not so easy as it may first seem.

Another way of assessing the spouse selection model is to compare the number of correct predictions with that generated by an alternative model. In the literature on assortative matching (e.g., Epstein and Guttmann 1984), positive or negative sorting is indicated by the sign of the correlation coefficient between the traits of husbands and wives. Thus one method to predict marriage partners is to match men and women by their rank in the distribution of traits. For example, if there is positive assortative matching in wage, we can match the woman with the rth highest female wage to the man with the rth highest wage among men.

To implement the ranking method to the case involving more than one trait, we rely on canonical correlation analysis.[17] Suppose men and women are arranged in such a way that the ith man is married to the ith woman. Let $F_{i1}, \ldots, F_{iK}$ be relevant traits of the wife, and let $M_{i1}, \ldots, M_{iK}$ be the husband's characteristics. Instead of examining the simple correlation coefficient between each trait separately, canonical correlation analysis examines the correlation between the K traits combined. More specifically, the analysis finds two vectors of weights, $(\theta_1^f, \ldots, \theta_K^f)$ and $(\theta_1^m, \ldots, \theta_K^m)$, such that the linear combinations $x_i = \sum_k \theta_k^f F_{ik}$ and $y_i = \sum_k \theta_k^m M_{ik}$ have the maximal correlation.

In our 1000 random sub-samples, the canonical correlation between husbands and wives averages to .713 for the model using `age`, `school` and `china`. The average canonical correlation increases to .758 when `wage` is included. For each sub-sample, we also save the canonical latent variables x and y. The woman with the rth highest x value is matched to the man with the rth highest y. The number of correct married couples predicted by this ranking method is recorded.

Figure 1 compares the distributions of the number of correct matches generated by the spouse selection model to that generated by the ranking method using the four variables `age`, `school`, `china` and `wage`. The ranking method produces an average of 3.2 correct predictions. On the other hand, the efficient market models give an average of 6.7 correct matches. In 856 of the 1000 bootstrap samples, the spouse selection model (Model W) gives more correct matches than the ranking method, while the ranking method gives more correct matches in only 83 of the 1000 samples. The frequency distributions in Figure 1 clearly show that the efficient market model first-order stochastically dominates the ranking method. Comparison of the ranking method to the spouse selection model without the wage variable yields essentially the same results.

VI. HOW EFFICIENT IS THE MARRIAGE MARKET?

In optimizing models, Varian [1990] argues that the economically relevant measure of goodness-of-fit is the distance between the predicted and the observed value of the objective function. Even if the observed pairing of men and women bears little resemblance to the pattern predicted by the efficient matching model, it is possible that the resulting loss in total marital output is economically insignificant. In our model, let $Z^o = \sum \hat{Z}_{ii}$ denote the imputed total marital output. If the optimal assignment matches man $\pi^*(i)$ to woman i, let $Z^* = \sum \hat{Z}_{i,\pi^*(i)}$ denote the maximum marital output. Then a natural measure of the degree of marriage market efficiency is Z^o/Z^*.

Assuming the estimated coefficients in our spouse selection model are the true production function parameters, information on total marital output can be partially recovered even though marital output is not directly observable. Since some of the production function parameters are unidentified, however, the index variable from our Tobit regressions, I_{ij}, will differ from imputed marital output, $\hat{Z}_{ij}$, by two fixed but unknown vectors, $(\lambda_1^f, \ldots, \lambda_n^f)$ and $(\lambda_1^m, \ldots, \lambda_n^m)$. Thus, for any assignment π that assigns man $\pi(i)$ to woman i, the sum of the index variables will differ from the sum of marital output by an additive constant: $\sum I_{i,\pi(i)} = \Lambda + \sum \hat{Z}_{i,\pi(i)}$, where $\Lambda = \sum_i(\lambda_i^f + \lambda_i^m)$. Because Λ is unknown, Z^o/Z^* cannot be recovered from $\sum I_{ii}$ and $\sum I_{i,\pi^*(i)}$. We therefore use an alternative measure of marriage market efficiency:

$$E = (Z^o - Z_*)/(Z^* - Z_*),$$

where Z_* is the value of the objective function given by the assignment π_* that *minimizes* the sum of imputed marital output.[18]

Denote $I^o = \sum I_{ii}$, $I^* = \sum I_{i,\pi^*(i)}$, and $I_* = \sum I_{i,\pi_*(i)}$, then $(I^o - I_*)/(I^* - I_*) = (Z^o - Z_*)/(Z^* - Z_*)$ regardless of the magnitude of the additive constant Λ. Therefore the efficiency measure E can be computed even though the marital output production function is not fully identified. This measure of efficiency ranges between zero and one: If the marriage market minimizes marital output, the efficiency measure is equal to zero; if it maximizes output, the efficiency measure is one.

Using the sample of 772 couples, and using the spouse selection model based on the variables `age`, `school`, `china` and `wage`, (i.e., Model W in Table 2), we find that

$E = .801$. Thus, compared to a full information and perfectly competitive market, the inefficiency loss in the Hong Kong marriage market is about twenty percent.[19]

One way to assess whether the calculated efficiency measure is "large" or "small" is to compare it to the efficiency measures generated from other possible assignments. With 772 couples, there are a total of $772! \approx 6.4 \times 10^{1895}$ possible assignments. Clearly it is impossible to exhaust all possibilities. We instead draw one million random assignments and compute the value of E for each case. Under Model W, the one million values of E have a mean of .476 and a standard deviation of .012. The maximum value of the efficiency index from these one million random assignments is .533 (4.7 standard deviations away from the mean). In contrast, the actual efficiency measure of is .801, which is 26.7 standard deviations greater than the mean. In other words, none of the one million random assignments produces a total marital output above that in the marriage market.[20] Such an event will occur with a probability of less than 3.7×10^{-44} if Z^o is not in the top 1/10000 quantile of the distribution of $\sum \hat{Z}_{i,\pi(i)}$. This exercise lead us to conclude that the marriage market is not grossly inefficient.

VII. CONCLUDING REMARKS

With information only on the basic demographic characteristics available from the census, we do not expect the spouse selection model to explain every detail of the marriage market. On the whole, however, the empirical model derived from the efficient marriage market hypothesis fits the data well. We find that corresponding male and female attributes (age, schooling, place of birth, and wage) are complementary in the production function for marital output. The hedonic pricing function is found to be a good representation of shadow prices in the marriage market, and marginal product pricing is not rejected. The empirical model also suggests that the marriage market is relatively efficient: total marital output is greater than that under generated by most other assignments.

One obvious way to improve on the model is to utilize data sets that contain richer information on personal characteristics. For example, if panel data is used, pre-marriage wage and occupation can safely be assumed to be independent of marital decisions. Such information can be exploited to help settle the question of whether there is positive or negative assortative matching in wage.

Conventional large scale surveys, however, are not expected to contain many of the variables pertinent to marriage decisions. More fruitful use of our model probably depend on new applications and more novel data sets. For example, the matching of CEOs to corporations can be regarded as an optimal assignment problem and estimated using the model developed here. With information on corporate characteristics and the personal characteristics of the CEOs, the model will help determine the complementarity relationships between such characteristics. Another example is the matching of houses to homeowners. Data on the characteristics of houses and of homeowners are readily available. By deriving an empirically matching model that is fully consistent with maximizing behavior, we hope this paper will stimulate the empirical research on such and other matching problems.

References

Allen, D. W. "'What Does She Sees in Him?': The Effect of Sharing on the Choice of Spouse." *Economic Inquiry* , January 1992, 57–67.

Becker, G. S. "A Theory of Marriage: Part I." *Journal of Political Economy* , August 1973, 813–46.

——. "Human Capital, Effort, and the Sexual Division of Labor." *Journal of Labor Economics* , 1985 Supplement, S33–S58.

——. *A Treatise on the Family,* enlarged edition. Cambridge: Harvard University Press, 1991.

——, Landes, E. M., and R. T. Michael. "An Economic Analysis of Marital Instability." *Journal of Political Economy* , December 1977, 1141–87.

Benham, L. "Benefits of Women's Education within Marriage." *Journal of Political Economy* , 1974 Supplement, S57–S71.

Bennett, E. "Consistent Bargaining Conjectures in Marriage and Matching." *Journal of Economic Theory* , August 1988, 392–407.

Bergstrom, T., and D. Lam. "The Effects of Cohort Size on Marriage Markets in Twentieth Century Sweden," in *The Family, the Market and the State of Aging Societies,* edited by J. Ermisch and N. Ogawa. Oxford: Oxford University Press, 1994.

Burkard, R. E., and U. Derigs. *Assignment and Matching Problems: Solution Methods with FORTRAN-Programs.* Berlin: Springer-Verlag, 1980.

Cochrane, J. "The Sensitivity of Tests of Intertemporal Allocation of Consumption to Near-Rational Alternatives." *American Economic Review* , June 1989, 319–37.

Cohen, L. "Marriage, Divorce, and Quasi Rents; Or, 'I Gave Him the Best Years of My Life.'" *Journal of Legal Studies*, June 1987, 267–303.

Cox, D.R., and D. Oakes. *Analysis of Survival Data.* London: Chapman and Hall, 1984.

Efron, B., and R. J. Tibshirani. *An Introduction to the Bootstrap.* New York: Chapman-Hall, 1993.

Epstein, E., and R. Guttman. "Mate Selection in Man: Evidence, Theory and Outcome." *Social Biology*, Fall 1984, 243–78.

Feller, W. *Introduction to Probability Theory and Its Applications,* Vol. 1, 3d. ed., New York: Wiley, 1968.

Foster, A. D. "Analysis of Household Behavior when Households Choose Their Members: Marriage Market Selection and Human Capital Allocations in Rural Bangladesh." Working paper, University of Pennsylvania, 1996.

——, and N. U. Khan. "Equilibrating the Marriage Market in a Rapidly Growing Population: Evidence from Rural Bangladesh." Working paper, University of Pennsylvania, 1994.

Gale, D., and L. S. Shapley. "College Admissions and the Stability of Marriage." *American Mathematical Monthly*, January 1962, 9–15.

Grossbard-Shechtman, S. *On the Economics of Marriage: A Theory of Marriage, Labor, and Divorce.* Boulder: Westview Press, 1993.

Heckman, J. J. "Sample Selection Bias as a Specification Error." *Econometrica* , January 1979, 153–62.

Hotelling, H. "The Most Predictable Criterion." *Journal of Educational Psychology*, 1935, 139–42.

Keeley, Michael C. "The Economics of Family Formation." *Economic Inquiry* , April 1977, 238–50.

Koopmans, T. C., and M. Beckmann. "Assignment Problems and the Location of Economic Activities." *Econometrica* , January 1957, 53–76.

Kremer, M. "The O-Ring Theory of Economic Development." *Quarterly Journal of Economics* , August 1993, 551–75.

Lam, D. "Marriage Markets and Assortative Mating with Household Public Goods." *Journal of Human Resources*, Fall 1988, 462–487.

Lundberg, S., and R. A. Pollak. "Separate Spheres Bargaining and the Marriage Market." *Journal of Political Economy* , December 1993, 988–1010.

Manly, B. F. J. *Multivariate Analysis: A Primer.* London: Chapman-Hall, 1986.

Mare, R. D. "Five Decades of Educational Assortative Mating." *American Sociological Review*, February 1991, 15–32.

Manser, M., and M. Brown. "Marriage and Household Decision Theory—A Bargaining Analysis." *International Economic Review* , February 1980, 31–44.

McElroy, M. B., and M. J. Horney. "Nash-Bargained Household Decisions: Toward a Generalization of the Theory of Demand." *International Economic Review* , June 1981, 333–49.

McFadden, D. "Econometric Analysis of Qualitative Response Models." in *Handbook of Econometrics,* Vol. 2, edited by Z. Griliches and M. Intrilligator. Amsterdam: North Holland, 1984.

McLaughlin, K. J. "Rent Sharing in an Equilibrium Model of Matching and Turnover." *Journal of Labor Economics* , October 1994, 499–523.

Rao, V. "The Rising Price of Husbands: A Hedonic Analysis of Dowry Increases in Rural India." *Journal of Political Economy* , August 1993, 666–77.

Rosen, S. "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition." *Journal of Political Economy* , February 1974, 34–55.

——. "Authority, Control, and the Distribution of Earnings." *Bell Journal of Economics* , Autumn 1982, 311–23.

Roth, A. E., and M. A. O. Sotomayor. *Two-sided Matching: A Study in Game Theoretic Modeling and Analysis.* New York: Cambridge University Press, 1990.

Sattinger, M. "Comparative Advantage and the Distributions of Earnings and Abilities." *Econometrica* , May 1975, 455–68.

Scully, G. W. "Mullahs, Muslims, and Marital Sorting." *Journal of Political Economy* , October 1979, 1139–43.

Shapley, L., and M. Shubik. "The Assignment Game I: The Core." *International Journal of Game Theory*, 1972, 111–30.

Varian, Hal. "Goodness-of-Fit in Optimizing Models." *Journal of Econometrics*, October 1990, 125–40.

Watkins, M. P., and W. Meredith. "Spouse Similarity in Newlyweds with respect to Specific Cognitive Abilities, Socioeconomic Status, and Education." *Behavior Genetics*, January 1981, 1–21.

Winch, R. F. *Mate-Selection.* New York: Harper, 1958.

Wong, Y. C. "Entrepreneurship, Marriage, and Earnings." *Review of Economics and Statistics* , November 1986, 693–99.

Notes

* We thank Dr. Ira Hammerman for helping us locate a computer algorithm that solves the optimal assignment problem. Seminar participants at the University of Hong Kong, the Chinese University of Hong Kong, Simon Fraser University, the University of Washington, the Western Economic Association International Conference, and the Australasian Meeting of Econometric Society offered many useful comments. We are particularly indebted to Doug Allen, Yoram Barzel, William Chan, Rick Harris, Bob Pollak, Gene Silberberg and Alan Siu for their helpful advice.

[1] The problem of unobserved individual heterogeneity will be addressed in a fixed effects model in Section IV.

[2] We ignore polygamy and polyandry.

[3] The optimal matching of women and men can be described as a linear program that maximizes $\sum_{ij} x_{ij} Z_{ij}$ subject to $\sum_i x_{ij} = 1$, $\sum_j x_{ij} = 1$, and $x_{ij} \geq 0$. The P^f and P^m vectors are the Lagrange multipliers of the equality constraints. As usual these Lagrange multipliers can be interpreted as shadow values.

[4] See, for example, Roth and Sotomayer [1990], Bennett [1988] and McLaughlin [1994] for discussions of matching models involving thin markets.

[5] The shadow value of a certain male attribute will depend on the distribution of the corresponding female attribute in the market. For example, if male and female education are complementary, the shadow value of male education will be relatively high in a market where women are well educated, and the shadow value will be relatively low in a market where women are less well educated. There is an extensive literature on how the age distribution of the population produces a “marriage squeeze” (e.g., Grossbard-Shechtman [1993]; Rao [1993]). In a large market, however, each individual woman will have negligible effect on the shadow value of male characteristics.

[6] The shadow price of a woman is her net utility from marriage. If a woman marries an otherwise undesirable man, she will be compensated by a higher material output (see Grossbard-Shechtman [1993]), but her net utility—i.e., her shadow price—remains

unchanged.

[7] If the production function and pricing function parameters were known rather than estimated, we can also test marginal productivity theory by performing a paired t-test between $\gamma_{1k} + 2\gamma_{2k}M_{jk}$ and $\gamma_{0k} + \delta_k F_{jk}$.

[8] Assuming the entire territory constitutes one market is less unrealistic for a city such as Hong Kong than for larger countries. Researchers interested in marriage markets elsewhere are advised to analyze the data at the city or county level rather than at the national level.

[9] See, for example, Keeley [1977] for a discussion on the optimal timing of marriage.

[10] We use Heckman's [1979] two-stage estimation procedures to correct for this self selectivity. In the first stage probit regression, we use household size, schooling (quadratic), place of birth and age (third order polynomial) as independent variables. The inverse Mills ratio computed from the probit estimates together with the above independent variables except household size are entered into the second stage wage regression. The estimated coefficient on the self-selectivity term is negative but statistically insignificant.

[11] The wage observations we have are post-marriage wages. Since the division of labor within the family may affect the choice of occupations and labor effort (see Becker [1985; 1991]), observed wages may be correlated with marital choice. Using pre-marriage wages would be superior to post-marriage wages if such data are available.

[12] Including wage variables into Model NW uses 5 degrees of freedom and increases the log-likelihood by 51.3. Including cross-interaction terms into Model W uses 12 degrees of freedom and increases the log-likelihood by 17.2.

[13] Detailed tables showing the coefficient estimates under the alternative specifications discussed in this sub-section are available from the authors.

[14] If sampling were done with replacement, a couple could appear more than once in a sub-sample and equilibrium matching would not be unique.

[15] In effect this means that only interaction effects are important in predicting marriage partners. Therefore the ability to identify δ (subject to a multiplicative constant)

is sufficient to generate a prediction for optimal pairing. Knowledge about α, β and γ are not required.

[16] The problem of finding $p(n,k)$ is a classic problem in probability. One version of this problem goes with the following story: A secretary has prepared n different letters and envelopes addressed to different persons. He puts his letters randomly into the n envelopes. What is the probability that k letters will go to the intended destination? This problem was solved more than two centuries ago by the mathematician Montmort, and the formula for $p(n,k)$ can be found in probability texts such as Feller [1968].

[17] Canonical correlation analysis was introduced by Hotelling [1935]. Manly [1986] contains an elementary exposition of the subject.

[18] Both the minimum and the maximum assignment problem can be solved by Burkard and Derigs's program with a simple change of signs.

[19] Models of individual maximization (e.g., Cochrane [1989]; Varian [1990]) typically find an efficiency loss of less than five percent. However, since the envelope theorem does not apply to the optimal assignment problem, the order of magnitude of the efficiency index in these two types of models are not directly comparable.

[20] We also perform a similar exercise for Model NW. The maximum E from one million random assignments is .537 while the actual E is .806. Again the actual marriage pattern is more efficient than any of the one million random assignments.

Table I

Sample Means, Standard Deviations, and Correlation Coefficients

	m_age	m_school	m_china	m_wage	mean (female)
f_age	.578	.208	.148	.144	23.108 (3.660)
f_school	.043	.669	−.008	.520	9.220 (3.908)
f_china	.152	−.007	.249	−.086	.266 (.442)
f_wage	−.110	.421	−.121	.552	2.855 (.606)
mean (male)	26.439 (4.470)	9.877 (3.685)	.380 (.486)	3.161 (.626)	

The bottom row contains the means and standard deviations (in parentheses) of male characteristics. The last column refer to female characteristics. Main elements in the cross table are simple correlation coefficients between the corresponding male and female attributes.

Table II

Parameter Estimates for Spouse Selection Model

Variable	NW	W	NW-C	W-C
intercept	1.0550 (0.4017)	2.8599 (0.5499)	0.9748 (0.4450)	3.1975 (0.7899)
Male traits				
m_age	−0.1502 (0.0240)	−0.1496 (0.0244)	−0.1329 (0.0253)	−0.1050 (0.0321)
m_china	−0.0784 0.0301	−0.0745 (0.0304)	−0.0793 (0.1819)	−0.0133 (0.2445)
m_school	−0.0595 (0.0158)	−0.0527 (0.0162)	−0.1107 (0.0272)	−0.0261 (0.0413)
m_wage		−0.4682 (0.1514)		−0.9684 (0.2390)
m_age^2	−0.0027 (0.0005)	−0.0026 (0.0005)	−0.0027 (0.0005)	−0.0027 (0.0005)
m_school^2	−0.0076 (0.0011)	−0.0066 (0.0011)	−0.0079 (0.0011)	−0.0065 (0.0011)
m_wage^2		−0.1133 (0.0267)		−0.1429 (0.0298)
Female traits				
f_age	−0.1064 (0.0244)	−0.1185 (0.0250)	−0.1215 (0.0257)	−0.1841 (0.0322)
f_china	−0.1234 (0.0379)	−0.1157 (0.0382)	−0.0579 (0.1808)	0.1576 (0.2278)
f_school	−0.1015 (0.0141)	−0.0811 (0.0146)	−0.0423 (0.0249)	−0.0497 (0.0346)
f_wage		−0.6888 (0.1451)		−0.4884 (0.2261)
f_age^2	−0.0050 (0.0008)	−0.0047 (0.0008)	−0.0053 (0.0008)	−0.0049 (0.0008)
f_school^2	−0.0067 (0.0010)	−0.0062 (0.0010)	−0.0067 (0.0010)	−0.0064 (0.0010)
f_wage^2		−0.1094 (0.0276)		−0.1055 (0.0288)
Same-trait interactions				
m_age*f_age	0.0128 (0.0011)	0.0126 (0.0011)	0.0131 (0.0011)	0.0127 (0.0012)
m_china*f_china	0.2897 (0.0546)	0.2829 (0.0549)	0.3055 (0.0577)	0.2876 (0.0587)
m_school*f_school	0.0226 (0.0015)	0.0199 (0.0015)	0.0229 (0.0015)	0.0211 (0.0018)
m_wage*f_wage		0.4116 (0.0456)		0.4789 (0.0570)

continued on next page

Table II (continued)

Variable	NW	W	NW-C	W-C
Cross-trait interactions				
m_age*f_china			−0.0033	−0.0036
			(0.0066)	(0.0066)
m_age*f_school			−0.0024	−0.0018
			(0.0008)	(0.0009)
m_age*f_wage				−0.0088
				(0.0063)
m_china*f_age			−0.0033	−0.0045
			(0.0080)	(0.0084)
m_china*f_school			0.0078	0.0160
			(0.0068)	(0.0086)
m_china*f_wage				−0.0373
				(0.0552)
m_school*f_age			0.0024	0.0006
			(0.0011)	(0.0013)
m_school*f_china			0.0017	0.0111
			(0.0077)	(0.0094)
m_school*f_wage				−0.0196
				(0.0091)
m_wage*f_age				0.0227
				(0.0085)
m_wage*f_china				−0.0927
				(0.0579)
m_wage*f_school				0.0004
				(0.0089)
log-likelihood	−4660.51	−4609.22	−4653.92	−4592.02

Standard errors are in parentheses. Models labeled by "W" includes the wage variable, and models labeled by "NW" does not include the wage variable. The suffix "C" refer to models that include cross-interaction terms.

Table III

Spouse Selection Model: Restricted Estimates

Variable	NW-M	NW-F	NW-M&F	W-M	W-F	W-M&F
intercept	1.2239	1.0453	1.2774	3.0062	2.6810	2.8265
	(0.3697)	(0.3767)	(0.3120)	(0.4982)	(0.4974)	(0.3923)
Male traits						
m_age	−0.1637	−0.1504	−0.1642	−0.1631	−0.1504	−0.1624
	(0.0453)	(0.0240)	(0.0450)	(0.0453)	(0.0243)	(0.0452)
m_china	−0.0771	−0.0791	−0.0762	−0.0754	−0.0750	−0.0749
	(0.0145)	(0.0302)	(0.0141)	(0.0146)	(0.0305)	(0.0143)
m_school	−0.0604	−0.0603	−0.0606	−0.0533	−0.0530	−0.0532
	(0.0319)	(0.0157)	(0.0319)	(0.0321)	(0.0161)	(0.0320)
m_wage				−0.4553	−0.4683	−0.4383
				(0.2615)	(0.1513)	(0.2604)
m_age^2	−0.0024	−0.0027	−0.0024	−0.0024	−0.0026	−0.0024
	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)
m_school^2	−0.0075	−0.0076	−0.0075	−0.0066	−0.0066	−0.0066
	(0.0011)	(0.0011)	(0.0011)	(0.0011)	(0.0011)	(0.0011)
m_wage^2				−0.1137	−0.1128	−0.1159
				(0.0255)	(0.0267)	(0.0252)
Female traits						
f_age	−0.1030	−0.1062	−0.1071	−0.1151	−0.1097	−0.1105
	(0.0242)	(0.0352)	(0.0348)	(0.0249)	(0.0354)	(0.0351)
f_china	−0.1226	−0.1082	−0.1090	−0.1145	−0.1060	−0.1071
	(0.0376)	(0.0202)	(0.0202)	(0.0378)	(0.0204)	(0.0204)
f_school	−0.1017	−0.0998	−0.1005	−0.0810	−0.0798	−0.0801
	(0.0139)	(0.0295)	(0.0295)	(0.0145)	(0.0298)	(0.0631)
f_wage				−0.6919	−0.6548	−0.6568
				(0.1449)	(0.2680)	(0.2679)
f_age^2	−0.0050	−0.0050	−0.0049	−0.0047	−0.0048	−0.0048
	(0.0008)	(0.0008)	(0.0008)	(0.0008)	(0.0008)	(0.0008)
f_school^2	−0.0067	−0.0067	−0.0066	−0.0062	−0.0063	−0.0063
	(0.0010)	(0.0009)	(0.0009)	(0.0010)	(0.0010)	(0.0010)
f_wage^2				−0.1092	−0.1122	−0.1119
				(0.0276)	(0.0275)	(0.0274)
Same-trait interactions						
m_age*f_age	0.0126	0.0127	0.0126	0.0125	0.0126	0.0125
	(0.0011)	(0.0011)	(0.0011)	(0.0011)	(0.0011)	(0.0011)
m_china*f_china	0.2904	0.2850	0.2871	0.2840	0.2792	0.2821
	(0.0545)	(0.0533)	(0.0532)	(0.0549)	(0.0538)	(0.0537)
m_school*f_school	0.0226	0.0226	0.0226	0.0199	0.0199	0.0198
	(0.0015)	(0.0015)	(0.0015)	(0.0015)	(0.0015)	(0.0015)
m_wage*f_wage				0.4113	0.4099	0.4100
				(0.0456)	(0.0455)	(0.0455)
log-likelihood	−4661.2	−4660.7	−4661.4	−4610.0	−4609.6	−4610.3
χ^2-statistic (d.f.)	1.33 (3)	0.37 (4)	1.79 (6)	1.59 (4)	0.75 (4)	2.15 (8)

Models marked with "F" indicate that marginal productivity restrictions are applied to females. Models marked with "M" impose such restrictions to males. Models marked with "F&M" impose the restrictions to both sexes. Standard errors are shown in parentheses.

Table IV

Spouse Selection Model: Bootstrap Estimates

Variable	NW	NW-FE	NW-M&F	W	W-FE	W-M&F
intercept	3.1329		3.3351	5.4508		5.4078
	(1.1342)		(1.1573)	(1.3396)		(1.4184)
Male traits						
m_age	−0.2027		−0.2169	−0.2020		−0.2154
	(0.0551)		(0.0581)	(0.0568)		(0.0594)
m_china	−0.0913		−0.0871	−0.0887		−0.0855
	(0.0513)		(0.0426)	(0.0524)		(0.0431)
m_school	−0.0734		−0.0735	−0.0650		−0.0645
	(0.0249)		(0.0237)	(0.0267)		(0.0259)
m_wage				−0.5943		−0.5708
				(0.2170)		(0.2358)
$\texttt{m\_age}^2$	−0.0036		−0.0032	−0.0036		−0.0032
	(0.0018)		(0.0018)	(0.0018)		(0.0018)
$\texttt{m\_school}^2$	−0.0096		−0.0093	−0.0085		−0.0083
	(0.0036)		(0.0035)	(0.0035)		(0.0035)
$\texttt{m\_wage}^2$				−0.1580		−0.1532
				(0.0789)		(0.0768)
Female traits						
f_age	−0.1534		−0.1510	−0.1736		−0.1625
	(0.0617)		(0.0629)	(0.0690)		(0.0700)
f_china	−0.1492		−0.1245	−0.1401		−0.1221
	(0.0867)		(0.0608)	(0.0089)		(0.0616)
f_school	−0.1248		−0.1222	−0.0999		−0.0982
	(0.0241)		(0.0226)	(0.0245)		(0.0236)
f_wage				−0.8547		−0.8210
				(0.2654)		(0.2666)
$\texttt{f\_age}^2$	−0.0066		−0.0063	−0.0062		−0.0060
	(0.0031)		(0.0030)	(0.0032)		(0.0030)
$\texttt{f\_school}^2$	−0.0085		−0.0083	−0.0081		−0.0079
	(0.0031)		(0.0030)	(0.0031)		(0.0030)
$\texttt{f\_wage}^2$				−0.1579		−0.1546
				(0.0909)		(0.0905)
Same-trait interactions						
m_age*f_age	0.0173	0.0192	0.0168	0.0172	0.0193	0.0167
	(0.0054)	(0.0060)	(0.0052)	(0.0054)	(0.0061)	(0.0052)
m_china*f_china	0.3409	0.3691	0.3281	0.3336	0.3653	0.3218
	(0.1718)	(0.1831)	(0.1603)	(0.1733)	(0.1862)	(0.1624)
m_school*f_school	0.0285	0.0304	0.0278	0.0253	0.0277	0.0247
	(0.0068)	(0.0073)	(0.0067)	(0.0069)	(0.0074)	(0.0067)
m_wage*f_wage				0.5527	0.6509	0.5391
				(0.1994)	(0.2296)	(0.1945)
mean log-likelihood	−420.56	−418.72	−421.29	−413.71	−410.86	−414.55

Columns marked with "FE" refer to the fixed effect models. Estimates of the individual fixed effects are not shown in this table. Columns marked with "M&F" indicate that marginal productivity restrictions are applied. Bootstrap standard errors are shown in parentheses.

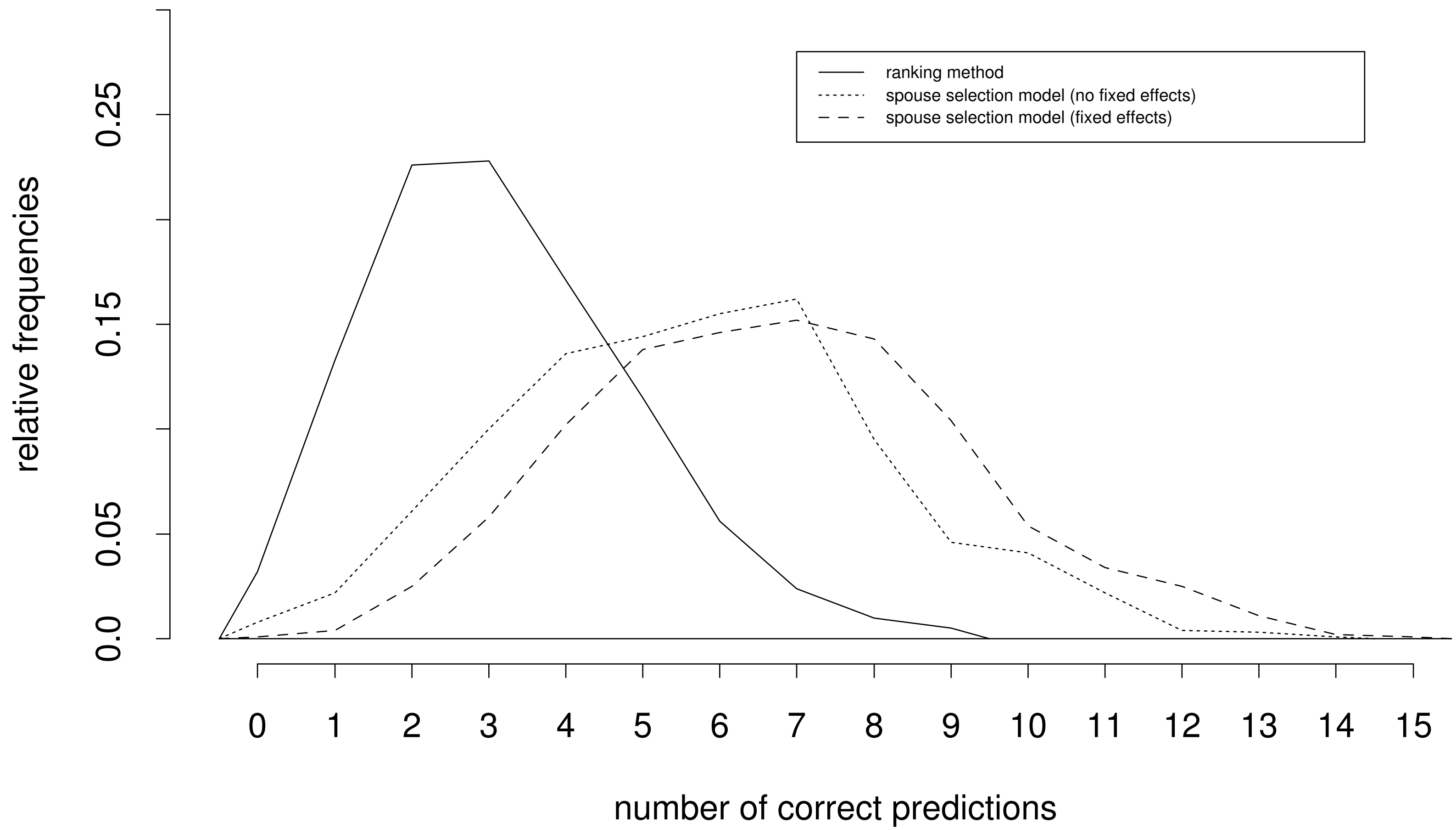
Table V

Theoretical and Empirical Distribution of Correct Matches

k ($n = 100$)	random matching	NW	NW-FE	W	W-FE
$k \geq 1$	0.6321	0.996	0.996	1.000	0.999
$k \geq 2$	0.2642	0.972	0.969	0.992	0.995
$k \geq 3$	0.0803	0.908	0.891	0.970	0.970
$k \geq 4$	0.0190	0.791	0.774	0.909	0.912
$k \geq 5$	0.0037	0.605	0.618	0.809	0.810
$k \geq 6$	0.0006	0.432	0.444	0.673	0.672
$k \geq 7$	0.0001	0.294	0.273	0.529	0.526
$k \geq 8$	1.0e-05	0.177	0.163	0.374	0.374
$k \geq 9$	1.1e-06	0.091	0.097	0.212	0.231
$k \geq 10$	1.1e-07	0.043	0.048	0.117	0.127
$k \geq 11$	1.0e-08	0.017	0.024	0.071	0.073
$k \geq 12$	8.3e-10	0.009	0.007	0.030	0.039
$k \geq 13$	6.4e-11	0.003	0.003	0.008	0.014
$k \geq 14$	4.5e-12	0.000	0.000	0.004	0.003
$k \geq 15$	3.0e-13	0.000	0.000	0.001	0.001
mean	1	5.338	5.307	6.699	6.746

The theoretical distribution is calculated by using the formula in Feller (1968). The empirical distribution is obtained by solving the optimal assignment problem for the predicted index variables from each of the 1000 sub-samples.

Figure 1
Frequency Distributions of Correct Predictions: Model W



A Direct Test of the Efficient Marriage Market Hypothesis

Wing Suen

and

Hon-Kwong Lui*

JEL C51, C61, C78, J12

Abstract. This paper takes Becker's efficient marriage market hypothesis at face value, and directly confront it with data from Hong Kong. The theory of optimal assignment is used to develop an empirical model of spouse selection, which resembles a Tobit model. This model can address positive or negative assortative matching as well as marginal product pricing in marriage markets. We also use a computer algorithm to solve the assignment problem for imputed marital output. The degree to which the actual pairing of husbands and wives corresponds to the optimal pairing provides a goodness-of-fit test of the efficient marriage market hypothesis.

Suen: Senior Lecturer, School of Economics and Finance, The University of Hong Kong, Hong Kong

Phone 852-2859-1052, Fax 852-2548-1152

E-mail hrneswc@hkusua.hku.hk

Lui: Assistant Professor, Department of Marketing and International Business, Lingnan College, Hong Kong

Phone 852-2616-8233, Fax 852-2467-3049

E-mail hklui@ln.edu.hk