As good as married

Gu, Xinhua; Rao Sahib, Padma

As good as married: a model of long-term cohabitation, learning and marriage[1].

Xinhua Gu[2]   Padma Rao Sahib[3]

SOM Theme D: Regional Science

[1]We thank Marco Haan for useful comments.

[2]Department of Economics, University of Toronto, 150 St. George Street, Toronto, Ontario, M5S 3G7, Canada, fax +1 416 978 6713, email: xinhuagu@chass.utoronto.ca.

[3]Corresponding author, Department of International Economics and Business, Faculty of Economics, University of Groningen, PO Box 800, 9700 AV Groningen, The Netherlands, fax: +31 50 363 7337, email: p.rao.sahib@eco.rug.nl.

Abstract

This paper develops a two-sided search-matching model with imperfectly observed types and learning. Since agents do not observe one another's type accurately, they first engage in a probationary partnership to learn one another's true type. Using the metaphor of premarital cohabitation and marriage, we demonstrate that long-term cohabiting individuals eventually learn each other's true type. We also demonstrate that singles of either sex are partitioned into classes and are matched in the same class in equilibrium. We show that sequential learning reduces signalling errors so that the Bayes estimator of the true type converges almost surely to true type. As noisy information is filtered over time, the mismatch risk disappears and the aggregate matching pattern based on true types is restored.

Keywords: two-sided search-matching, marriage, premarital cohabitation, imperfect information, Bayesian learning.

JEL classification: D83, D84, J12.

1 Introduction

There is a growing literature on modelling long-term partnership formation (see Burdett and Coles 1999 for a survey). The models typically used are two-sided search-matching models (SM) which incorporate features of marriage and labour markets such as incomplete information about characteristics, bilateral meetings that occur at uncertain intervals (one does not meet prospective spouses on a daily basis) and heterogeneity across individuals or firms. Among studies in this literature are Collins and McNamara (1990), Morgan (1995, 1996), Smith (1997), Shimer and Smith (2000), Bloch and Ryder (2000), Burdett and Wright (1998) and Chade (1999, 2001). Many of such models yield positive assortative mating, a positive association of the traits of partners, as an equilibrium outcome. The decision rule in such models is simple. Each party employs a cut-off rule accepting the other as partner only if a certain minimum requirement is attained and what you see is what you get. Once such a partnership is formed, the pair leave the market forever.

While models in the search-matching literature include many features of long-term partnership formation, they do not incorporate the probationary period that precedes many long-term relationships. Workers may be on a temporary contract before getting a permanent one, couples may date or cohabit before marrying and you may want to try a sports club a few times before signing up for an entire year.

Such probationary relationships arise because both parties have limited information about each other. Long-term relationships, as anyone who has experienced a divorce will agree, can be costly to dissolve. Temporary relationships, however, are easier to dissolve and can serve the purpose of information gathering. While the success of a temporary relationship cannot guarantee the success of the long-term relationship that may follow it, it is generally believed that temporary relationships allow both parties to make a more informed choice. Among studies that incorporate imperfectly observed information of this particular type in SM models are Chade (1999) and Rao Sahib and Gu (2002a, 2002b). In Chade, agents observe and match on imperfectly observed type but never learn true type. However, Chade also incorporates the information contained in whether or not a partner is accepted which leads to strategies that are non-monotonic in the characteristics of individuals. In Rao Sahib and Gu (2002a), a simple model of premarital cohabitation and learning is developed in which couples cohabit prior to marriage for one period at the end of which their true qualities are revealed to one another. At this juncture, couples either marry or separate. Rao Sahib and Gu (2002b) explores the implications of relaxing the assumption of risk neutrality in a two-sided search-matching framework. In the current paper, we allow premarital cohabitation lasts until the true qualities of at least one partner is revealed to the other. Therefore, premarital cohabitation can last for a random number of periods. Cohabiting couples receive information shocks about each other and therefore learn over time about each other's true qualities.

In this paper, we extend a two-sided search-matching (henceforth SM) model to incorporate this frequently observed feature of long-term partnerships. We develop our model using marriage and premarital cohabitation as a metaphor, although the model with some modifications may be interpreted as a model of marriage with the option to get a divorce. Using the framework originally developed in Burdett and Coles (1997) hereafter simply Burdett and Coles, we incorporate imperfect information of a particular type and learning in the SM context. We use Bayesian learning techniques similar to those used in Jovanovic (1979) and Prescott and Townsend (1980) to address sequential learning. Our focus is on the decision making of the individual rather than the aggregate implications of marriage markets with matching.

In contrast to many studies in which an individual's type is instantly observed on contact, we assume that match quality is gradually revealed and individuals learn each others' worth over time. Therefore, in our model, couples cohabit with one another for one or more time periods before forming a long-run marital partnership. Because of learning, a match once formed may dissolve as a consequence of rational choice. Cohabitation, therefore, does not always lead to marriage.

We retain however, other assumptions often made in SM models (1) stationary market environment: the distribution of types among singles in the market is constant over time, and individuals believe that the market may be characterized this way; (2) stationary Nash strategy: the search strategy is a list of the opposite sex singles to whom a particular single will propose on contact. Singles compete for a good opposite sex partner by maximizing their utility, given the other singles' behaviour and the constraints they have to face defined by their own types; (3) rational expectation equilibrium: the realized distribution of types among singles in the market coincides with their subjective steady state distribution, which is taken as given when analyzing the individual's optimization behaviour. Therefore, our focus is on Nash equilibrium rather than market equilibrium.

We show that the equilibrium class partition result as obtained in previous studies holds even under uncertainty of information. Although the introduction of learning in the equilibrium matching model substantially complicates the SM framework, it generate new insights into issues related to the intertemporal properties of a long-run partnership. In terms of our analogy of cohabitation and marriage, this implies that long-term cohabiting couples eventually learn one another's true type.

The next section describes the marriage market with the incorporation of uncertainty. Section 3 discusses Bayesian learning in the SM context. Section 4 concludes and the Appendix provides some proofs of results in the paper.

2 The marriage market

2.1 Uncertainty and information

We assume that a large and equal number of male and female singles participate in a marriage market. Using the terminology in Burdett and Coles, the attractiveness of each individual as a marriage partner is assumed to be a real number called 'pizazz' or type x (we use type and pizazz interchangeably in the rest of the paper). Pizazz is a comprehensive index that takes into account all the factors that make up an individual's attractiveness in the marriage market. Pizazz is important because it is the if a man and woman marry, the man receives the pizazz of the woman as utility and vice versa and the instantaneous utility of being single is zero. For the population of men in the market, there is a known distribution $G_M(x)$ of men's pizazz types from which a woman draws offers. That is, before a woman meets a man with type x in the market, $G_M(x)$ is her prior probability that the man's pizazz is not greater than x. Similarly, the women's pizazz distribution, is denoted $G_W(x)$. For simplicity, we assume that $G_M(x) = G_W(x) = G(x) = P(X \leq x)$ with support $[\underline{x}, \overline{x}]$. In this sense, the SM process is random.

To ease exposition, we discuss a woman's search problem in detail. Throughout the woman is treated as decision maker and the man's search problem can be handled by symmetry. For now, learning is assumed to take place only once. Suppose that the woman meets a man of true type x (that is, she draws an offer from the distribution of X). She is unable to observe a man's true type because it is distorted by random noise, ε, where $E(\varepsilon) = 0$ and $Var(\varepsilon)$ is small. She observes y, which is a realization of the signal $Y = X + \varepsilon$, where $\varepsilon \perp X$. The conditional mean pizazz is denoted $M = E(X \mid Y)$ and its realization is denoted m. The woman is assumed to infer the man's true type x from his noisy signal y according to the noise reduction rule,

$m = m(y) = E(X \mid Y = y)$. Notice that $m = \widehat{x}$ is also the Bayes estimator of x obtained by minimizing a risk function, which is the expectation of a quadratic loss function with respect to[1] $Q(x \mid y) = P(X \leq x \mid Y = y)$. This is the woman's posterior knowledge about the distribution of the man's true pizazz given his observed pizazz. That is,

$$m = \arg\min_{\widehat{x}} E_{x|y} \frac{1}{2} (\widehat{x} - x)^2 .$$

We make two additional assumptions to address the issue of class partitions with learning. The first ($A1$) is that $m'(y) > 0$, suggesting that the higher the value of a man's observed pizazz y, the greater the woman's perception of him, reflected by a higher m. From $A1$, it can be seen that $m = m(y)$ is invertible so that $y = y(m)$ exists. It is straightforward to derive the distribution of M: $F(m) = P(M \leq m)$ $= P(Y \leq y(m)) = F_Y(y(m))$, its support denoted $[\underline{m}, \overline{m}]$.

The second assumption ($A2$) is related to $Q(x \mid y)$, whose mean is simply m by definition. Using $y = y(m)$, one can transform $Q(x \mid y)$ into $Q(x \mid m)$. We assume in $A2$ that $\partial Q(x \mid m) / \partial m < 0$. Under $A2$, the distribution of pizazz of a man with a higher level of m' first-order stochastically dominates that of another man with a lower m'' since $Q(x \mid m') < Q(x \mid m'')$. The degree of stochastic dominance increases

[1] The joint distribution density of (X, Y), $f(x, y)$, is involved in the calculation of $Q(x \mid y)$ and $F(m)$ in the following way:

$$Q(x \mid y) = \frac{\int_{\underline{x}}^{x} f(x, y)\, dy}{\int_{\underline{x}}^{\overline{x}} f(x, y)\, dx} \quad \text{and} \qquad F(m) = \int_{\underline{y}}^{y(m)} \int_{\underline{x}}^{\overline{x}} f(x, y)\, dx dy.$$

When $X \sim N(\mu, \sigma_0^2)$ and $\varepsilon \sim N(0, 1)$, the correlation coefficient ρ between X and Y is $\sigma_0 / \sqrt{\sigma_0^2 + 1}$, and hence the joint distribution density is

$$f(x, y) = \frac{1}{2\pi\sigma_0} \exp\left\{ -\frac{1}{2} \left[\frac{(x - \mu)^2}{\sigma_0^2} + (x - y)^2 \right] \right\}$$

In the case of normal distributions, $\overline{x}$ and $\overline{m}$ should be set equal to ∞ in the text of the paper.

with rising levels of m. This assumption is important and underlies the equilibrium matching.

For example, consider $x \sim N(\mu, \sigma_0^2)$, $\varepsilon \sim N(0, \sigma_\varepsilon^2)$. Let $m_0 = \mu$. Suppose[2] that after drawing x, only $y_1 = x + \varepsilon_1$ is observed where $\varepsilon_1 \perp x$. Then, we have $\widehat{x}_1 = m_1 = m(y_1)$ as a Bayes estimator for $x_1 \equiv (x \mid y_1) \sim N(m_1, \sigma_1^2)$, where

$$m_1 = w_0 m_0 + (1 - w_0) y_1, \qquad \sigma_1^2 = w_0 \sigma_0^2, \tag{1}$$

and $w_0 = \sigma_\varepsilon^2 / (\sigma_0^2 + \sigma_\varepsilon^2)$. In the above, the m_0-related term is the revision of estimating x based on y_1 through the weighted average involving the variances of two underlying distributions regarding x and ε. Clearly, $m'(y_1) = 1 - w_0 > 0$, and $A1$ is met. Since $\partial Q(x \mid m_1) / \partial m_1 = -\frac{1}{\sigma_1} \phi\left(\frac{x - m_1}{\sigma_1}\right) < 0$ where $\phi(\cdot)$ is the standard normal density, we know that $A2$ is also met. Since

$$E\widehat{x}_1 = E(m_1) = E[E(x \mid y_1)] = E(x) = \mu, \tag{2}$$

it follows that $m_1 \sim N\left(\mu, \sigma_{m_1}^2\right)$ where $\sigma_{m_1}^2 = \sigma_0^4 / (\sigma_0^2 + \sigma_\varepsilon^2)$, and $\sigma_{m_1}^2 < \sigma_0^2 < \sigma_{y_1}^2$ $(= \sigma_0^2 + \sigma_\varepsilon^2)$. Note that m_1 is not only more informative than y_1 since it is an inference about x from y_1, but also more precise in estimating a particular x since $\sigma_{m_1|x}^2 \; (= (1 - w_0)^2 \sigma_\varepsilon^2\ ) < \sigma_{y_1|x}^2 \; (= \sigma_\varepsilon^2)$.

In terms of the mechanics of the model, a parameterizing variable is needed to take into account the effect of the woman's own type on her SM opportunities in the market. In general, the higher her own type, the more easily can she can attract potential partners of higher types. Since the woman has an incentive to over-state her own pizazz, her potential partners act according to their own perception $\widetilde{m}$ of her

[2]For simplicity, we henceforth follow the convention of using lower-case letters to refer both to random variables and their realization.

true pizazz $\widetilde{x}$ based on her signal $\widetilde{y}$. Here, $\widetilde{m} = E\left(\widetilde{X} \mid \widetilde{Y} = \widetilde{y}\right)$ serves as a filter used by men to reduce part of noise in the signalled pizazz of the woman. Therefore, $\widetilde{m}$ is a feasible candidate for parameterization. On the contrary, it is illogical to use $\widetilde{x}$ as a parameterizing instrument since $\widetilde{x}$ is not observed by men in the earlier stages of SM. That is, before her true type $\widetilde{x}$ is revealed, only her filtered signal $\widetilde{m}$ matters in affecting her opportunities in contacting single men.

A complication arises however, if $\widetilde{m}$ is used for parameterization. Even if the woman knows her own type $\widetilde{x}$, she is not sure what signal $\widetilde{y}$ men receive when they meet her. This is because her noise $\widetilde{\varepsilon}$ is purely random and the market is assumed to be memoryless. Since $\widetilde{m}$ is stochastic from a woman's own standpoint, the utility $R(\widetilde{m})$ of her being single would be random if it is parameterized with $\widetilde{m}$. To eliminate the randomness arising from parameterizing R with $\widetilde{m}$, as required in a standard search model, we utilize the woman's expectation of her $\widetilde{m}$ conditional on her private information about $\widetilde{x}$ as the parameterizing variable. Denote $\widetilde{m}_e = E[m(\widetilde{y}) \mid \widetilde{x}] = E_{\widetilde{\varepsilon}}[m(\widetilde{x} + \widetilde{\varepsilon})] = m_e(\widetilde{x})$, which is, under $A1$, an increasing function of $\widetilde{x}$ after the noise have been averaged out. In the normal distribution case, $m_e(\widetilde{x}) = w_0\mu + (1 - w_0)\widetilde{x}$. This weighted average of $\widetilde{x}$ (her privately-known type) and μ (her *ex ante* estimate of the men's estimate of her type) should give this woman a reasonable estimate of the value of her $\widetilde{m}$ as perceived by men in the market. In what follows, we parameterize the men's pizazz distribution faced by a single female. We write $F(m \mid \widetilde{m}_e)$ as the distribution of conditional mean pizazz among single men who will propose to cohabit with an $\widetilde{m}_e$-type woman when they meet. We next describe the process and decisions related to meeting and matching.

2.2 Meeting and Matching

Since singles in the market have difficulty contacting each other, a Poisson process is used to characterize a market participant's opportunities of encountering the opposite sex. Let α be the total arrival rate of offers of cohabitation that are made by all opposite sex singles and received by any single of either sex. Since singles of different types face different opportunities, only part of these offers are received by a $\widetilde{m}_e$-type woman. We denote the offer arrival rate faced by an $\widetilde{m}_e$-type woman by $\alpha(\widetilde{m}_e)$ and $\alpha(\widetilde{m}_e) \leq \alpha$. It is intuitive to assume that the arrival rate faced by an $\widetilde{m}_e$-type woman is positively related to her type. That is, the more attractive the woman, the more offers of cohabitation she receives.

Since cohabitation unions are initiated on the basis of imperfect information, they may dissolve when one or both parties find the other's type too low. Therefore, we introduce another Poisson process to model the probability that a woman learns the true type of her cohabiting partner. Therefore, λ is the rate of arrival of new signals that leads to the revelation of the partner's true type. When this occurs, a cohabitation relationship may be transformed into marriage or dissolve.

Learning about match quality is assumed to take place in three stages in the context of the Poisson process approximation. This is similar to the exposition in Sargent (1987) of the three-stage model matching model from Jovanovic (1979). The separation into three stages is for the sake of analytic convenience, although the division into stages is not very clear in this continuous-time setting in which offers arrive at random intervals.

Recall that if a man and a woman decide to marry, the woman's utility from the match is assumed to equal the man's pizazz and vice versa. Similar to other studies in this literature, we rule out narcissism. The advantage to a woman of possessing high pizazz ($m_e(\widetilde{x})$) is that it enables her to attract (captured by $\alpha(\widetilde{m}_e)$) men of higher types (captured by $F(m \mid \widetilde{m}_e)$).

In this setting, a marriage search model is nested in a cohabitation search model. The two models are interrelated by the parameterizing variable $\widetilde{m}_e$.

In stage 1, the pre-draw stage, a woman is in the singles market looking for a partner. At this moment she only knows $F(m)$. Only nature determines whether she will meet a man who will propose to cohabit with her and this occurs at random. Hence, $R(\widetilde{m}_e)$, the lifetime utility of the $\widetilde{m}_e$-type woman being single and receiving $u_{\widetilde{m}_e}(m)$ in the current period and behaving optimally later on is:

$$R(\widetilde{m}_e) = \frac{1}{1+rh}\left[u_{\widetilde{m}_e}(m)\,h + \alpha(\widetilde{m}_e)\,hE_m V_{\widetilde{m}_e}(m) + (1-\alpha(\widetilde{m}_e)\,h)\,R(\widetilde{m}_e)\right] + o(h)\,. \tag{3}$$

In the above, $V_{\widetilde{m}_e}(m)$ is the value function of this $\widetilde{m}_e$-type woman having an offer of cohabitation from a m-type man in the next stage and behaving optimally later on. The discount rate is r, $\alpha(\widetilde{m}_e)\,h$ is the probability of receiving an offer of cohabitation within a time interval of short length h, and $(1-\alpha(\widetilde{m}_e)\,h)$ the probability of receiving no offers during this interval. Given that we have already ruled out narcissism, we simplify our calculations further by assuming $u_{\widetilde{m}_e}(m) = u(m) = m$. Since there is no partner at this time, $m = 0$. Noting that $o(h)/h \to 0$ as $h \to 0$, and rewriting eqn (3), we obtain

$$R(\widetilde{m}_e) = \frac{\alpha(\widetilde{m}_e)}{r+\alpha(\widetilde{m}_e)} E_m V_{\widetilde{m}_e}(m)\,. \tag{4}$$

Stage 2 marks the stage when the woman has received an offer of cohabitation from an x-type man with a signal y. She must then decide whether to accept him as a cohabiting partner. This woman's value function in this stage is derived from the cohabitation search model as follows:

$$V_{\widetilde{m}_e}(m) = \max\left\{\varphi_{\widetilde{m}_e}(m), R(\widetilde{m}_e)\right\}.$$

In the above, $\varphi_{\widetilde{m}_e}(m)$ is the utility of the $\widetilde{m}_e$-type woman accepting an offer from an m-type man in this stage and behaving optimally later on. If she decides to cohabit with him, she receives as instantaneous utility the expected value of his true pizazz conditional on observed pizazz, $u(m)h$. Therefore,

$$\varphi_{\widetilde{m}_e}(m) = \frac{1}{1+rh}\left[u(m)h + \lambda h E_{x|y} J_{\widetilde{m}_e}(x) + (1-\lambda h)\varphi_{\widetilde{m}_e}(m)\right] + o(h). \quad (5)$$

Here, λh is the probability in the next stage h, that the woman receives a new signal that reveals her cohabiting partner's true type (the woman then decides whether or not to marry him. $(1-\lambda h)$ is the probability of not receiving any new signals and continuing in the current cohabiting union. In our model, cohabiting couples are not allowed to transit directly to another cohabiting union without a intervening period of singlehood. $J_{\widetilde{m}_e}(x)$ is the value function of finding out her partner's true pizazz x and having to choose between marrying him and leaving him. The expectation of this value function is taken with respect to x conditional on its observation y $(= y(m))$ or m. As $h \to 0$, eqn (5) reduces to

$$\varphi_{\widetilde{m}_e}(m) = \frac{m + \lambda E_{x|m} J_{\widetilde{m}_e}(x)}{r+\lambda}. \quad (6)$$

In stage 3, the true pizazz of her cohabiting partner is revealed to the woman and she must decide whether to marry her cohabiting partner after having found out his true type x. Her value function in this stage can be characterized by the marriage search model as follows:

$$J_{\widetilde{m}_e}(x) = \max\left\{\psi_{\widetilde{m}_e}(x), R(\widetilde{m}_e)\right\},$$

where $\psi_{\widetilde{m}_e}(x) = x/r = \psi(x)$ is the utility of accepting an offer for marriage in this stage and staying married forever.

The above inter-dependent search models maybe solved treating them as a dynamic programming problem by working backwards. The optimal strategy for the $\widetilde{m}_e$-type woman in stage 3 is a reservation pizazz policy: marry the man if his true pizazz $x \geqslant x_r$, where $x_r = rR(\widetilde{m}_e) = \psi^{-1}(R(\widetilde{m}_e))$ is the minimum type of a spouse acceptable to this woman. Similarly, the woman's optimal strategy in stage 2 is also a reservation strategy: cohabit with a man if his conditional mean pizazz $m \geqslant m_r$, where $m_r = \varphi_{\widetilde{m}_e}^{-1}(R(\widetilde{m}_e))$ is the minimum conditional mean pizazz acceptable to this woman in a cohabiting partner.

To solve for the utility of the woman being single in stage 1, we first calculate the expectation of her stage 2 value function (see Appendix (5.1)):

$$E_m V_{\widetilde{m}_e}(m) = \int_{\varphi_{\widetilde{m}_e}^{-1}(R(\widetilde{m}_e))}^{\overline{m}} [1 - F(m \mid \widetilde{m}_e)]\, d_m \varphi_{\widetilde{m}_e}(m) + R(\widetilde{m}_e). \tag{7}$$

Substituting this into eqn (4) yields $R(\widetilde{m}_e)$ for this $\widetilde{m}_e$-type woman:

$$R(\widetilde{m}_e) = \frac{\alpha(\widetilde{m}_e)}{r(r+\lambda)} \int_{\varphi_{\widetilde{m}_e}^{-1}(R(\widetilde{m}_e))}^{\overline{m}} [1 - F(m \mid \widetilde{m}_e)] \left[1 - \frac{\lambda}{r} \int_{\psi^{-1}(R(\widetilde{m}_e))}^{\overline{x}} \frac{\partial Q(x \mid m)}{\partial m} dx\right] dm. \tag{8}$$

This is also the utility of being single for the whole population in the market if $\widetilde{m}_e = m_e(\widetilde{x})$ changes with different true types $\widetilde{x}$ across single individuals. Blackwell's sufficient conditions ensure that there exists a unique solution to this functional equation in $R(\widetilde{m}_e)$ which is continuous in $\left[\varphi_{\widetilde{m}_e}^{-1}(R(\widetilde{m}_e)), \overline{m}\right]$. By symmetry, we have another functional equation in $R(m_e)$ for a m_e-type man as the decision-maker, which is identical to eqn (8) except that m_e and $\widetilde{m}_e$ are interchanged. The use of symmetry is justified by the assumption that the underlying distributions for both sexes are the same: $G_M(x) = G_W(x)$. From these two functional equations class partitioning can be obtained.

It should be noted that the optimal reservation policy as an *ex ante* decision rule derived from a search model must be determined based on the underlying offer distribution and prior to the realization of any offers. Any specific offer realization only matters in *ex post* decision making but has no impact on the formulation of the decision rule. Therefore, to be precise, we should treat x_r as being determined in stage 2; we do so only to highlight the information updating related to the realization of y at stage 2. However, y is still random in the first stage 1, and hence $Q(x \mid y)$ or $Q(x \mid m)$ is unknown at stage 1 (though the distribution of m is prior knowledge). In fact, the model implies that the woman as a decision-maker, does not derive her reservation pizazz policy for marriage from a single signal issued by a particular man she meets. Also, this is a continuous-time setting with what are in fact two interrelated search models. In theory, the woman has to determine, *ex ante,* x_r as well as m_r by solving the above setting based on her knowledge of the prior distributions, $G(x)$ and $F(m)$.

Using truncated distributions and discounted arrival rates the class partition at the aggregate level may be derived. It can be shown that the utility of being single, for either a man or woman, is a non-decreasing step function, and this result ultimately leads to the class partition (see Rao Sahib and Gu 2002a for a detailed exposition). Therefore, market participants can be split up into n distinct classes in equilibrium so that matches, cohabitation or marital, take place only between the two sexes in the same class. The class partition for marriage is consistent with that for cohabitation on a one-to-one basis in the symmetric equilibrium, consistent with our assumption that cohabitation is a probationary period that precedes marriage. Individuals set higher cut-off levels for marriage than for cohabitation and this follows from the assumption that observed pizazz is a noisy version of true pizazz (Rao Sahib and Gu 2002a). Since ε is a pure random noise having zero mean and trivial variance, and therefore causes no systematic distortions, the two classes are unlikely to deviate much from each other, and may coincide in the extreme case when ε converges almost surely ($a.s.$) to zero.

The macro pattern of matching induced by class partition is not Pareto optimal due to imperfect information, in the sense that a low-type single can increase his or her welfare through mismatch. This is a situation faced by a single whose partner is qualified for cohabitation but not for marriage, which is revealed in the later stage of cohabitation. The mismatch can be achieved by mis-signalling at the partner's expense if strategic interaction is allowed. The cohabiting individual of higher type will suffer a utility loss as opposed to what she or he would otherwise receive when matched with a partner of compatible type under perfect information.

3 Bayesian sequential learning

In the case of one-period learning, it is possible, though not very likely (because σ_ε^2 is assumed to be small), for a person with very high (or low) pizazz x to generate a very low (or high) signal y. In this case, even though m filters the noise in observed pizazz and the variance of m is less than that of y or even x, the imprecision still remains in cohabitation and causes the risk of mismatch. We now discuss the case of multiple-period learning, in which information updating takes place in each period.

Of course, one must keep in mind that updating of information does not pertain only to the signal received in the first period but also to subsequent signals which may contain errors. As before, it is assumed that there is no opportunistic behaviour or deliberate mis-signalling.

To make the mechanism of noise reduction clear we change now to the discrete-time setting but still maintain the assumption of a stationary market environment in which $G(x,t) = G(x)$ where t is time index. We first discuss certain issues related to the individual's optimization problem. Regarding the probability structure of this setting, we keep the analytics simple by assuming that the underlying distributions are normal as in the normal example presented earlier.

In the pre-draw stage, the information set is $I_0 = \{\phi\}$. In period 1, a single woman contacting a single man is in fact sampling from $x_0 \equiv (x \mid I_0) = x \sim N(m_0, \sigma_0^2)$. Upon drawing from the observed pizazz distribution, only $y_1 = x_0 + \varepsilon_1$ is observed with $\varepsilon_1 \perp \{I_0, x_0\}$. The woman then has a new information set $I_1 = \{y_1\} \cup I_0$ in this period, and $x_1 \equiv (x_0 \mid y_1) = (x \mid I_1) \sim N(m_1, \sigma_1^2)$, where its moments were stated in the normal example (as shown in (1)). From this time on, Bayesian learning is used to consecutively update information as the woman cohabits with the same man to learn more and more about him. The information set in period t is $I_t = \{y_1, \cdots, y_t\}$, the set of signals received until this period. In period $t+1$, the woman samples from $x_t \sim N(m_t, \sigma_t^2)$. She observes signal $y_{t+1} = x_t + \varepsilon_{t+1}$ with $\varepsilon_{t+1} \perp \{I_t, x_t\}$. This yields $I_{t+1} = \{y_{t+1}\} \cup I_t$ and $x_{t+1} \equiv (x_t \mid y_{t+1}) = (x \mid I_{t+1}) \sim N\left(m_{t+1}, \sigma_{t+1}^2\right)$, where

$$m_{t+1} = w_t m_t + (1 - w_t)\, y_{t+1}, \qquad \sigma_{t+1}^2 = \sigma_t^2 w_t < \sigma_t^2, \tag{9}$$

and $w_t = \sigma_\varepsilon^2 / \left(\sigma_t^2 + \sigma_\varepsilon^2\right)$ (see Appendix (5.2)).

This recursive relation is the same as the Kalman filtering mechanism for noise reduction in signal extraction. The updated inference m_{t+1} is the weighted average of the last inference m_t and the new signal y_{t+1}. Since $Cov\left[(m_t, y_{t+1}) \mid m_t\right] = 0$ (see Appendix (5.3)), we then refer to m_t and y_{t+1} as conditionally uncorrelated. So, this new inference is a linear combination of the two conditionally uncorrelated elements. The variance σ_t^2 of this kind of inference decreases with the updating of information. The coefficient w_t is the fraction of σ_ε^2 that causes a noise effect. The larger is this fraction, the smaller is the weight placed on y_{t+1} in revising m_t to form m_{t+1}, and the more reliable it is to rely upon m_t to predict m_{t+1} than to use y_{t+1}. It is more likely in this case that y_{t+1} reflects the noise effect rather than the true type, for the informational content of $y_{t+1} = (x \mid I_t) + \varepsilon_{t+1}$ decreases with rising σ_ε^2.

Noting that $m_{t+1} = E\left(x_{t+1}\right) = E\left(x \mid I_{t+1}\right)$, we see that m_{t+1} is just the Bayes estimator of $(x \mid I_{t+1})$. In other words, $m_{t+1} = \widehat{x}_{t+1}$ is an observable, workable decision variable; whereas x_{t+1} is a latent, underlying process of sequential learning about true type x from all previous signals in I_{t+1}. Letting $h_t = 1/\sigma_t^2$ be the precision of x_t, setting $\sigma_\varepsilon^2 = 1$ for simplicity, and iterating (9), yields

$$h_t = h_0 + t \to \infty, \quad m_t = \frac{\mu h_0/t + x + \overline{\varepsilon}_t}{h_0/t + 1} \to x, \quad \text{as } t \to \infty. \tag{10}$$

This limiting result is due to the fact that $\{\varepsilon_t\}$ is a sequence of i.i.d. noises and hence $\overline{\varepsilon}_t = \sum_{i=1}^t \varepsilon_i/t \overset{a.s.}{\to} E(\varepsilon) = 0$ by the Kolmogorov law of large numbers. Since $E(x_t) = m_t \overset{a.s.}{\to} x$ and $Var(x_t) = \sigma_t^2 = \frac{\sigma_0^2}{\sigma_0^2 t + 1} \to 0$ as $t \to \infty$, we know from asymptotic theory, that $x_t \overset{p}{\to} x$ (convergence in probability) and hence $x_t \overset{d}{\to} x$ (convergence in distribution)[3]. That is, $Q_t \equiv Q(x \mid I_t) = Q(x_t) \to G(x)$ as $t \to \infty$. This distribution known at t is referred to as the posterior distribution because its mean m_t becomes

[3]Since x is stochastic *ex ante,* we know that $x_t \overset{p}{\to} x \Rightarrow x_t \overset{d}{\to} x$ prior to the realization of x. If x is realized, then we have only $x_t \overset{p}{\to} x$, where x is a constant *ex post.* Mathematically, for $m_t \overset{a.s.}{\to} x$, a true type x can only be revealed after an infinite amount of time: $t = \infty$.

known only after y_t has been realized even if its variance is known in advance at the pre-draw stage.

Although m_t is known *ex post* and unknown *ex ante,* its distribution is known *ex ante* with $m_t \xrightarrow{d} x$. In fact, it follows from (10) that $m_t = \widehat{x}_t \sim N\left(\mu, \sigma^2_{m_t}\right)$, where

$$E\left(m_t\right) = E\left[E\left(x \mid I_t\right)\right] = E\left(x\right) = \mu, \qquad \sigma^2_{m_t} = \frac{\sigma_0^4}{\sigma_0^2 + 1/t} \to \sigma_0^2. \tag{11}$$

$F_t \equiv F\left(m_t\right)$, like $G\left(x\right)$, should be viewed as a prior distribution even if its argument m_t involves information updating. This is because both the mean and variance of $F\left(m_t\right)$ are known at the pre-draw stage, and a normal distribution is entirely determined by its mean and variance. Note that $\sigma^2_{m_t}$, monotonic with t, reaches its minimum at $t = 1$ and then increases all way to σ_0^2 at a diminishing rate as $t \to \infty$. We term this the monotone property of $\sigma^2_{m_t}$ for later use. This implies that from the *ex ante* perspective at the pre-draw stage, $F\left(m_t\right)$ is riskier than $F\left(m_{t-1}\right)$ and $G\left(x\right)$ is riskier than $F\left(m_t\right)$ for any t (the farther into the future, the more the uncertainty). The decision-makers' prior knowledge of the riskiness of these distributions determines the amount of risk premia accruing to them, as shown in the next section. The distribution of m_t, although ultimately approaching x, is more concentrated around μ for smaller values of t. This suggests that the estimation of a specific true type x using filtered signals m_t in earlier periods relies heavily upon μ, the mean of the population X from which this particular x is drawn.

Next, we explore the asymptotic property of a new random variable $(m_{t+1} \mid m_t)$. We use this random variable in the intermediate stage of the multi-period learning search model which will be discussed next. Utilizing the conditional uncorrelation property mentioned above, and the serial independence of noises, one can derive, from

(9) (see Appendix (5.3)), that as $t \to \infty$,

$$E\left(m_{t+1} \mid m_t\right) = m_t \to x, \qquad Var\left(m_{t+1} \mid m_t\right) = \sigma_t^2 \left(1 - w_t\right) \to 0. \tag{12}$$

It then follows that $\left(m_{t+1} \mid m_t\right) \sim N\left[m_t, \sigma_t^2 \left(1 - w_t\right)\right]$ which is known at t, and $Plim\left(m_{t+1} \mid m_t\right) = x$. Thus, $\left(m_{t+1} \mid m_t\right) \xrightarrow{d} x$. That is, the (second sort of) posterior distribution $K_t \equiv K\left(m_{t+1} \mid m_t\right) \to G(x)$ as $t \to \infty$. It is now obvious that the magnitude of σ_ε^2 becomes less important once multiple-period learning is incorporated and a larger σ_ε^2 only slows down the speed of convergence. Therefore, we arrive at

Theorem 1 *Sequential sampling from $x_t = (x \mid I_t)$ and its Bayes estimator $m_t = \widehat{x}_t$, along with $(m_{t+1} \mid m_t)$, converge in probability or almost surely to the true type x. Therefore, all the sampling distributions (Q_t, F_t, K_t) must converge to the population distribution (G) of true types. The greater the learning, the lower the uncertainty; the imprecision σ_t^2 (prior) or $\sigma_t'^2$ (posterior) associated with a cohabiting union (given its survival), approaches zero through learning over infinitely many time periods.*

In what follows, ζ is the probability of receiving a new signal from the current cohabiting partner, λ is the probability of receiving a signal leading to revealing the partner's true type and $\beta = 1/\left(1 + r\right)$ is the discount factor. The posterior distribution $K\left(m_{t+1} \mid m_t\right)$ for $t \geq 1$ is not truncated, and only $K\left(m_1 \mid m_0, \widetilde{m}_e\right) = F\left(m_1 \mid \widetilde{m}_e\right)$ involves the parameterization by $\widetilde{m}_e$. Building upon the above probability structure, a SM model with t_1-period cohabitation can be specified as follows:

$$\begin{aligned} R\left(\widetilde{m}_e\right) &= \frac{\beta\alpha\left(\widetilde{m}_e\right)}{1 - \beta\left[1 - \alpha\left(\widetilde{m}_e\right)\right]} E_{m_1 \mid \widetilde{m}_e} V_{\widetilde{m}_e}\left(m_1\right), \qquad (13) \\ V_{\widetilde{m}_e}\left(m_t\right) &= \max\left\{\frac{m_t + \beta\zeta E_{m_{t+1} \mid m_t} V_{\widetilde{m}_e}\left(m_{t+1}\right)}{1 - \beta\left(1 - \zeta\right)},\ R\left(\widetilde{m}_e\right)\right\}; \quad t = 1, \cdots, \left(t_1 - 1\right), \\ V_{\widetilde{m}_e}\left(m_{t_1}\right) &= \max\left\{\frac{m_t + \beta\lambda E_{x \mid I_{t_1}} J_{\widetilde{m}_e}\left(x\right)}{1 - \beta\left(1 - \lambda\right)},\quad R\left(\widetilde{m}_e\right)\right\}, \end{aligned}$$

$$J_{\widetilde{m}_e}(x) \quad = \quad \max\{x + \beta J_{\widetilde{m}_e}(x), \quad R(\widetilde{m}_e)\}.$$

The reservation pizazz m_{rt} as a decision rule derived from the posterior distribution $K(m_{t+1} \mid m_t)$ in period-t search model depends upon the realization of m_t. However, m_t is unknown *ex ante,* and any decision rule must be established at $t = 0$, which provides a role for the prior distribution $F(m_t)$ to play in determining the decision rule *ex ante.* This, along with the last theorem, implies

Corollary 2 *Given learning over an infinite horizon, the updated inference* $\{m_t\}$ *does indeed lead to the revelation of a particular partner's true type* x*. This results in an individual making the correct decision about whether or not to accept a partner's offer for marriage. The posterior distributions* (Q_t, K_t) *upon which a realization-dependent decision rule for each period of cohabitation is based, approach the prior distribution* (G) *of true pizazz among singles in the market. The realization-dependent decision rules are governed by the prior distribution* (F_t) *of future realizations* m_t*.*

The solution to the above set of interrelated search models with multiple-period learning, and its implications for matching and partitioning are presented next.

The setting in (13) can be used by a woman as decision-maker to carry out her search for a partnership. Suppose this t_1-period cohabitation setting produces a sequence of reservation demands for her cohabiting partner's and spouse's types in all periods: $(m_{r1}, \cdots, m_{rt_1}, x_r)$. It is implied that she and her partner would have to return to the singles pool at any time t $(\leq t_1)$ if their cohabiting union has survived $(t-1)$ periods but at least one party fails to meet the other party's reservation demand m_{rt} at period t. This model artificially stipulates that a partner's true type must be revealed through gradual learning after t_1-period cohabitation. If both sides survive the t_1-time selection determined by the other side's updated optimal reservation strategies, then they will be in a position to decide on whether to marry, which is specified by the last functional equation in $J_{\widetilde{m}_e}(x)$.

By symmetry, the same setting (13) can also be treated as the decision-making model for the man. It is the interaction of this type between both parties that leads to matching and partitioning, such as $(m_{r1}^j, \cdots, m_{rt_1}^j, x_r^j)$ for $1 \leq j \leq n$, where j is the class index, and n the number of classes. As j changes, the partitioning bounds form a $n \times (t_1 + 1)$ matrix. As claimed by Burdett and Coles (1999), the uncertainty of a long-term partnership (or the mismatch risk in our model) can be reduced but not eliminated by updating information unless an infinite-period cohabiting scheme sets in. In this case, the last two equations in (13) should be dropped and t_1 is allowed to go to infinity. The class partition for the multiple learning setting can be proven to exist in a way similar to how the class partition is proven to exist in the absence of learning (see Rao Sahib and Gu 2002a) with certain modifications. It is obvious that the class partition in period $t \leq t_1$, $(m_{rt}^1, \cdots, m_{rt}^n)$ with one more time-period of learning, is Pareto superior to its period-$(t-1)$'s counterpart, $(m_{r,t-1}^1, \cdots, m_{r,t-1}^n)$. Also, the Pareto efficiency will be almost surely ($a.s.$) attained at a stable pattern of macro mating $(x_r^1, \cdots, x_r^n)$. In this sense, the equilibrium matching pattern in BC is the limiting case of our model. We next examine a 3-period cohabitation model as a heuristic case.

3.1 Example: A three-period cohabitation model

In this example, we continue to consider a 3-stage framework. That is, the example consists of a pre-draw stage, a cohabiting stage (lasting three periods), and marital stage (this last stage can continue indefinitely). The calculation involved in solving the model (13) for $t_1 = 3$ is still quite tedious. Since this is a multi-period model, we should solve the last-period problem (marital search) first and then work backwards up to the pre-draw stage. To solve the search problem in each period, we must first derive the decision-maker's optimal reservation strategy and finally the utility of being single. For convenience, we assume that $\zeta = \lambda = 1$. That is, receiving a new signal and information updating occurs with certainty in each period. We use integration by parts repeatedly in the calculation (see Appendix (5.4)). The solution to the interrelated search problems, beginning with the marital stage and working backwards through cohabiting to pre-draw stages, is characterized by

$$\begin{aligned} x_r &= (1-\beta)\, R\left(\widetilde{m}_e\right), \\ x_r - m_{r3} &= \frac{1}{r}\int_{x_r}^{\overline{x}} \left[1 - Q\left(x \mid m_{r3}\right)\right] dx, \\ x_r - m_{r2} &= \beta \int_{m_{r3}}^{\overline{m}} \left[1 - K\left(m_3 \mid m_{r2}\right)\right] \left[1 + \frac{1}{r}\int_{x_r}^{\overline{x}} \frac{-\partial Q\left(x \mid m_3\right)}{\partial m_3} dx\right] dm_3, \\ x_r - m_{r1} &= \beta \int_{m_{r2}}^{\overline{m}} \left[1 - K\left(m_2 \mid m_{r1}\right)\right] \left\{1 + \beta \int_{m_{r3}}^{\overline{m}} \frac{-\partial K\left(m_3 \mid m_2\right)}{\partial m_2} \times \right. \\ & \left. \left[1 + \frac{1}{r}\int_{x_r}^{\overline{x}} \frac{-\partial Q\left(x \mid m_3\right)}{\partial m_3} dx\right] dm_3 \right\} dm_2, \end{aligned} \tag{14}$$

and

$$\begin{aligned} R\left(\widetilde{m}_e\right) &= \frac{\alpha\left(\widetilde{m}_e\right)}{r} \int_{m_{r1}}^{\overline{m}} \left[1 - F\left(m_1 \mid \widetilde{m}_e\right)\right] \left\{1 + \beta \int_{m_{r2}}^{\overline{m}} \frac{-\partial K\left(m_2 \mid m_1\right)}{\partial m_1} \times \right. \\ & \left. \left[1 + \beta \int_{m_{r3}}^{\overline{m}} \frac{-\partial K\left(m_3 \mid m_2\right)}{\partial m_2} \left\langle 1 + \frac{1}{r}\int_{x_r}^{\overline{x}} \frac{-\partial Q\left(x \mid m_3\right)}{\partial m_3} dx \right\rangle dm_3\right] dm_2 \right\} dm_1. \end{aligned} \tag{15}$$

Obviously, $x_r > m_{rt}$ $\forall t$, and $R(\widetilde{m}_e) > 0$ as desired, where

$$\frac{\partial K(m_{t+1} \mid m_t)}{\partial m_t} = -\frac{1}{\sigma_t'}\phi\left(\frac{m_{t+1}-m_t}{\sigma_t'}\right), \qquad \frac{\partial Q(x \mid m_{t_1})}{\partial m_{t_1}} = -\frac{1}{\sigma_{t_1}}\phi\left(\frac{x-m_{t_1}}{\sigma_{t_1}}\right).$$

Denoting $\sigma_t' = \sqrt{1-w_t}\sigma_t$ and $\sigma_0' = \sigma_{m_1}$and replacing term $\frac{1}{r}\int_{x_r}^{\overline{x}}(-\frac{\partial Q}{\partial m_{t_1}})dx$ with a term similar to $\beta(\cdots,\cdots)$, eqn (15) can then be viewed as indicating the infinite horizon case where $t_1 \to \infty$. In this case, eqns in (14) need to be modified accordingly.

By substituting (15) into (14), a set of four equations in four unknowns $(x_r, m_{r3}, m_{r2}, m_{r1})$ can be obtained as the optimality conditions which characterize the solution. This set defines $(x_r, m_{r3}, m_{r2}, m_{r1})$ to be a function of $(\sigma_0', \sigma_1', \sigma_2', \sigma_3)$. Differentiating this set and applying the implicit function theorem suggests that the reservation pizazz in each period and the utility of being single depend on the variances of the prior and posterior distributions. In class partitioning, the upper limits of the integrals with respect to m_t (for $t > 1$) and x remain unchanged, whereas the upper limit of the integral with respect to m_1 keeps changing. If a woman as the decision-maker believes that she is perceived as having $\widetilde{m}_e = \overline{m}$, then substituting this into (15) yields m_{r1} so that $(m_{r1}^1, \overline{m}]$ is identified as the top class of cohabitation in period 1. Then m_{r1}^1 will be used as the upper limit of the integral with respect to m_1 in identifying the second class of cohabitation $(m_{r1}^2, m_{r1}^1]$ for period 1. This procedure is continued for the third class, fourth class and so on. The procedure for the class partition in the case of the other periods can be carried out in the same manner.

An intuitive explanation for why $x_r > m_{rt}$ $\forall t$ is as follows. At the pre-draw stage, the woman has two types of prior information that matter in establishing, *ex ante,* her decision rules. The first is $F(m_t)$ which is related to cohabitation search, and the second is $G(x)$ which is related to marital search. With the same mean μ for these distributions and the monotone property $\sigma^2_{m_t} \leq \sigma^2$, $G(x)$ is a mean-preserving spread of $F(m_t)$. Her rational response to the riskier distribution $G(x)$ as opposed to $F(m_t)$, is to become more exacting at the pre-draw stage in establishing the decision rule for a marital over a cohabiting union to be formed later on. Therefore, her reservation demand for the marital partner's pizazz is higher than her reservation demand for the cohabiting partner' pizazz. It follows then that $(x_r - m_{rt}) \equiv RP_t$ is simply a kind of risk premium to compensate her for the higher risk she must bear in search of marriage relative to cohabitation.

From the above, we can establish (see Appendix (5.5)) that for any class j,

(i) the reservation demand rises with the tenure of a partnership such that $m^j_{r1} \leq m^j_{r2} \leq \cdots \leq x^j_r$. That is, the risk premium always exists across the periods of cohabitation until the marital stage;

(ii) the risk premium is shrinking with learning over time such that $RP^j_t \leq RP^j_{t-1}$ $\forall t$, where $RP^j_t = m^j_{rt} - m^j_{r,t-1}$; and (iii) $RP^j_{t_1} \to 0$ as $t_1 \to \infty$, in which case the marital class partition is restored.

The underlying reason for point (i) is that $\mu_{m_t} = \mu \; \forall t$ and $\sigma^2_{m_t}$ displays a monotone property across cohabiting periods, as shown in (11). Therefore, the prior distribution for period t, $F(m_t)$ is riskier than the prior distribution for period $t-1$, $F(m_{t-1})$. As a result, the decision-maker must demand a risk premium for compensation embedded in the *ex ante* decision rules, that is why $m^j_{r,t-1} \leq m^j_{rt}$. Similarly, $m^j_{rt} \leq x^j_r$. Point (ii) is in agreement with the fact in (12) that $\sigma^2_{(m_{t+1}|m_t)}$ decreases with information updating over time. Even if $\mu_{(m_{t+1}|m_t)} = m_t$ is still stochastic from the pre-draw standpoint, this favorable property concerning the precision of the posterior distribution $K(m_{t+1} \mid m_t)$ shows that signalling errors continue to reduce under sequential learning. Therefore, the risk premium decreases period by period. Point (iii) is consistent with the result of $(m_{t+1} \mid m_t) \xrightarrow{p} x$. This implies that an ideal class partition by the market participants' true types can be achieved only in the limiting case, when all noises have been eliminated through infinite learning. These points are summarized in

Theorem 3 *The greater the learning, the more selective are individuals in search of a partnership and the lower will be the mismatch risk in class partition. The risk premium, though non-zero during the entire process, decreases with sequential learning and approaches zero once the correct pattern of class partition based on true types is restored.*

It is important to note that random noise as the source of mismatch accompanies each contact between the sexes, and that people have to learn from their mistakes under uncertainty of information. The possibility of mismatch is therefore allowed for in cohabitation. This may particularly be the case in the early stages of SM when substantial signalling error may arise because there may be a large discrepancy between true type $\widetilde{x}$ and parameterizing type $m_e(\widetilde{x})$ for a single individual. New noises continue to arrive over the course of a partnership. This is reduced through Bayesian sequential learning. The persistence of mismatch risk during all premarital periods of a relationship is the reason why cohabiting individuals raise their reservation demands for their partners' type. Similarly, marriage is associated with a higher reservation demand than cohabiting partnerships.

4 Conclusion

This paper incorporates imperfect information and learning into a two-sided search matching model. We use the analogy of premarital cohabitation and marriage, and assume that couples initially cohabit to find out more about each other before getting married. During cohabitation, couples continue to receive signals about one another although these signals are noisy. They are assumed to use Bayesian updating to infer the true type of their partner. We show that couples whose cohabiting relationships survive eventually learn one another's true type. Equilibrium is characterized by a class partition: couples cohabitation and marriages occur among individuals within the same class. With learning, the class partition corresponding to cohabiting relationships then approaches the class partition of marital unions which are formed based on true type. That is, the mismatch risk continues to decrease as the couple learn more and more about one another. The continued updating of information during cohabitation restores the pattern of matching at the aggregate level that is based on true types. Concurrently, the risk premia borne by individuals diminishes with sequential learning and finally disappears when true type is revealed. We leave market equilibrium determination in the context of imperfect information and learning as a topic for further research.

5 Appendix

5.1 Derivation of (7)

Expanding the expectation of the value function in stage 3, we obtain

$$\begin{aligned} E_{x|m}J &= E_{x|m}\max\left(\psi, R\right) = \Pr(\psi \geq R)E_{x|m}\left[\psi \mid \psi \geq R\right] + \Pr(\psi < R)R \\ &= \frac{1}{r}\Pr(x \geq x_r \mid y)\left[E\left(x \mid x \geq x_r, y\right) - x_r\right] + R(\widetilde{m}_e) \\ &= \frac{1}{r}\int_{x_r}^{\overline{x}}\left(x - x_r\right)dQ(x \mid m) + R = -\frac{1}{r}\int_{x_r}^{\overline{x}}\left(x - x_r\right)d\left[1 - Q(x \mid m)\right] + R \end{aligned}$$

$$= \frac{1}{r} \int_{\psi^{-1}(R(\widetilde{m}_e))}^{\overline{x}} [1 - Q(x \mid m)]\, dx + R(\widetilde{m}_e). \tag{16}$$

We have used $x_r = rR(\widetilde{m}_e)$ in the third equality and applied the technique of integration by parts and $x_r = \psi^{-1}(R(\widetilde{m}_e))$ to the last equality. We omit $\widetilde{m}_e$ and other arguments of a function here and below to keep notations simple.

Recalling that $\varphi_{\widetilde{m}_e}(m) = R(\widetilde{m}_e)$, and applying the same procedure as above to the expectation of the value function in stage 2: $EV(m) = E \max\{\varphi, R\}$, one can derive (7) in the text of the paper.

5.2 Derivation of (9)

A random variable $x \sim N(\mu_x, \sigma_x^2)$ conditional on another $y \sim N(\mu_y, \sigma_y^2)$ follows a posterior distribution such as $(x \mid y) \sim N\left(\mu_{x|y}, \sigma_{x|y}^2\right)$, where

$$\mu_{x|y} = \mu_x + \rho \frac{\sigma_x}{\sigma_y}(y - \mu_y), \quad \sigma_{x|y}^2 = \sigma_x^2(1 - \rho^2), \quad \rho = \frac{Cov(x, y)}{\sigma_x \sigma_y}. \tag{17}$$

The calculation in the text is performed by sequentially employing this formula for $t = 1, 2, 3, \ldots$. Here we show how to do this in the general case. Consider a cohabiting woman who has $I_t = \{y_1, \cdots, y_t\}$ in hand. She samples x_t from $(x \mid I_t) \sim N(m_t, \sigma_t^2)$ but obtains $y_{t+1} = x_t + \varepsilon_{t+1}$ with $\varepsilon_{t+1} \perp \{I_t, x_t\}$. Then

$$\begin{aligned}
\mu_{y_{t+1}} &= E(x_t \mid I_t) + E(\varepsilon_{t+1} \mid I_t) = E[(x \mid I_t) \mid I_t] = E(x \mid I_t) = m_t, \\
\sigma_{y_{t+1}}^2 &= \sigma^2(x_t \mid I_t) + \sigma^2(\varepsilon_{t+1} \mid I_t) = \sigma^2(x_t) + \sigma^2(\varepsilon_{t+1}) = \sigma_t^2 + \sigma_\varepsilon^2.
\end{aligned}$$

According to (17), she will face a new posterior distribution

$$(x_t \mid y_{t+1}) \sim N\left[\mu_{x_t} + \rho_{x_t y_{t+1}} \frac{\sigma_{x_t}}{\sigma_{y_{t+1}}} \left(y_{t+1} - \mu_{y_{t+1}}\right), \sigma_{x_t}^2 \left(1 - \rho_{x_t y_{t+1}}^2\right)\right] \equiv N\left(m_{t+1}, \sigma_{t+1}^2\right). \tag{18}$$

Using $\varepsilon_{t+1} \perp \{I_t, x_t\}$ again yields

$$cov\left[(x_t, y_{t+1}) \mid I_t\right] = \sigma^2 (x_t \mid I_t) + cov\left[(x_t, \varepsilon_{t+1}) \mid I_t\right] = \sigma_t^2,$$

Substituting from the above yields:

$$\begin{aligned} &\mu_{x_t} = E(x_t) = E(x \mid I_t) = m_t, && \rho_{x_t y_{t+1}} = \frac{cov\left[(x_t, y_{t+1}) \mid I_t\right]}{\sigma_{x_t}\sigma_{y_{t+1}}} = \frac{\sigma_t}{\sqrt{\sigma_t^2 + \sigma_\varepsilon^2}}, \\ &\rho_{x_t y_{t+1}} \frac{\sigma_{x_t}}{\sigma_{y_{t+1}}} = \frac{\sigma_t^2}{\sigma_t^2 + \sigma_\varepsilon^2} = 1 - w_t, && \sigma_{x_t}^2 \left(1 - \rho_{x_t y_{t+1}}^2\right) = \frac{\sigma_t^2 \sigma_\varepsilon^2}{\sigma_t^2 + \sigma_\varepsilon^2} = \sigma_t^2 w_t. \end{aligned} \tag{19}$$

Therefore, it follows from (18) and (19) that

$$m_{t+1} = m_t + (1 - w_t)(y_{t+1} - m_t), \qquad \sigma_{t+1}^2 = \sigma_t^2 w_t.$$

This gives (9) in the text.

5.3 Derivation of (12)

Noticing that for $x = z$, $E\left[(xy) \mid z\right] = zE(y \mid z)$ and $E(x \mid z) = z$, and using

$$cov\left[(x, y) \mid z\right] = E\left[(xy) \mid z\right] - E(x \mid z) E(y \mid z),$$

we know that $cov\left[(m_t, y_{t+1}) \mid m_t\right] = 0$.

Using this and (9), recalling that $E(\varepsilon_{t+1} \mid y_t) = 0$, and noting that $\left[(x \mid I_t) \mid y_t\right]$

$= (x \mid I_t)$, the moments of $(m_{t+1} \mid m_t)$ are calculated as follows:

$$\begin{aligned}
E(m_{t+1} \mid m_t) &= w_t E(m_t \mid m_t) + (1-w_t) E(y_{t+1} \mid y_t) \\
&= w_t m_t + (1-w_t) E[(x \mid I_t) \mid y_t] \\
&= w_t m_t + (1-w_t) E(x_t) = m_t, \\
\sigma^2(m_{t+1} \mid m_t) &= w_t^2 \sigma^2(m_t \mid m_t) + (1-w_t)^2 \sigma^2(y_{t+1} \mid y_t) \\
&\quad +2w_t(1-w_t)\, cov\,[(m_t, y_{t+1}) \mid m_t] \\
&= (1-w_t)^2 \left\{\sigma^2[(x \mid I_t) \mid y_t] + \sigma^2(\varepsilon_{t+1} \mid y_t)\right\} \\
&= (1-w_t)^2 \left(\sigma_t^2 + \sigma_\varepsilon^2\right) = (1-w_t)\,\sigma_t^2.
\end{aligned}$$

Then, (12) obtains.

5.4 Derivation of (14) and (15)

To obtain the solution to (13), work backwards. Solve the period-4 problem first. Evaluating its value function at x_r, and calculating the period-3 expectation of the period-4 value function, $E_{x|I_3} J(x)$ (or interchangeably, $E_{Q(x|m_3)} J$), in the same way as in (16), we have

$$x_r = (1-\beta) R, \qquad E_{x|I_3} J(x) = \frac{1}{1-\beta} \int_{x_r}^{\overline{x}} (x - x_r)\, dQ(x \mid m_3) + R.$$

Evaluating the period-3 value function at m_{r3} to have $m_{r3} + \beta E_{x|m_{r3}} J(x) = R$, and calculating the period-2 expectation of the period-3 value function, $E_{m_3|m_2} V(m_3)$, in the same manner as in (16), yields the following:

$$x_r - m_{r3} = \frac{1}{r}\int_{x_r}^{\overline{x}} (x - x_r)\, dQ\,(x \mid m_{r3})\,, \tag{20}$$

$$E_{m_3|m_2} V\,(m_3) = \int_{m_{r3}}^{\overline{m}} \left[m_3 - x_r + \frac{1}{r}\int_{x_r}^{\overline{x}} (x - x_r)\, dQ\,(x \mid m_3) \right] dK\,(m_3 \mid m_2) + R.$$

Repeating the above procedure for period 2 to calculate

$$x_r - m_{r2} = \beta \int_{m_{r3}}^{\overline{m}} \left[m_3 - x_r + \frac{1}{r}\int_{x_r}^{\overline{x}} (x - x_r)\, dQ\,(x \mid m_3) \right] dK\,(m_3 \mid m_{r2})\,,$$

$$E_{m_2|m_1} V\,(m_2) = \int_{m_{r2}}^{\overline{m}} \left\{ m_2 - x_r + \beta \int_{m_{r3}}^{\overline{m}} \left[m_3 - x_r + \frac{1}{r}\int_{x_r}^{\overline{x}} (x - x_r)\, dQ\,(x \mid m_3) \right] dK\,(m_3 \mid m_2) \right\} dK\,(m_2 \mid m_1) + R. \tag{21}$$

Applying the same calculation to period 1 results in

$$x_r - m_{r1} = \beta \int_{m_{r2}}^{\overline{m}} \left\{ m_2 - x_r + \beta \int_{m_{r3}}^{\overline{m}} \left[m_3 - x_r + \frac{1}{r}\int_{x_r}^{\overline{x}} (x - x_r)\, dQ\,(x \mid m_3) \right] dK\,(m_3 \mid m_2) \right\} dK\,(m_2 \mid m_{r1})\,, \tag{22}$$

$$E_{m_1} V\,(m_1) = \int_{m_{r1}}^{\overline{m}} \left\langle m_1 - x_r + \beta \int_{m_{r2}}^{\overline{m}} \left\{ m_2 - x_r + \beta \int_{m_{r3}}^{\overline{m}} \left[m_3 - x_r + \frac{1}{r}\int_{x_r}^{\overline{x}} (x - x_r)\, dQ\,(x \mid m_3) \right] dK\,(m_3 \mid m_2) \right\} dK\,(m_2 \mid m_1) \right\rangle dF\,(m_1) + R.$$

Substituting the expression for $E_{m_1} V\,(m_1)$ into the first equation in (13) to yield the utility of being single in period 0 (the pre-draw stage)

$$R = \frac{\alpha}{r} \int_{m_{r1}}^{\overline{m}} \left\langle m_1 - x_r + \beta \int_{m_{r2}}^{\overline{m}} \left\{ m_2 - x_r + \beta \int_{m_{r3}}^{\overline{m}} \left[m_3 - x_r + \frac{1}{r}\int_{x_r}^{\overline{x}} (x - x_r) \right.\right.\right.$$

$$dQ\left(x \mid m_3\right)\right] dK\left(m_3 \mid m_2\right)\right\} dK\left(m_2 \mid m_1\right)\right\rangle dF\left(m_1\right). \tag{23}$$

First applying the technique of integration by parts to these formulae in the same fashion as in the treatment of (16), and then substituting from (20), (21) and (22), we obtain (14) and (15) in the text.

5.5 Reservation demand rises with tenure

To prove that reservation pizazz types m_{rt} rise with tenure t, we need to simplify and modify the notation used in (13) by omitting $\widetilde{m}_e$, letting $\zeta = \lambda = 1$, and assuming that cohabiting individuals never directly observe x in any finite horizon.

$$\begin{aligned} R &= \frac{\alpha}{r+\alpha} E_{F(m')} V_1\left(m'\right), \\ V_t\left(m\right) &= \max\left\{m + \beta E_{K_t(m'|m)} V_{t+1}\left(m'\right), \quad R\right\}, \qquad t = 1, 2, 3, \ldots \end{aligned} \tag{24}$$

$V_t\left(m\right)$ replaces $V\left(m_t\right)$, and $K_t\left(m' \mid m\right) = K\left(m' \mid m, \sigma_t'^2\right)$ represents $K\left(m_{t+1} \mid m_t\right)$. The above notational changes do not alter the nature of the original model.

Several steps are taken to complete the proof. (i) First, Jensen's inequality is used to develop the property of second-order stochastic dominance of $K_{t+1}\left(m'\right)$ over $K_t\left(m'\right)$, where $\sigma_{t+1}'^2 < \sigma_t'^2$. In what follows, we sometimes drop m (and in some cases m'), for simplicity. Considering $K_t\left(m''\right)$ as a mean-preserving spread of $K_{t+1}\left(m'\right)$, in the sense that each possible outcome m' is randomized further so as to have m'' $= m' + z$, where $z \sim H_{z|m'}\left(z \mid 0, \sigma_z^2\right)$. Since $E\left[\Psi\left(m\right)\right] \geqslant \Psi\left[E\left(m\right)\right]$ when $\Psi\left(m\right)$ is convex, we use the iterated expectations theorem, to obtain

$$\begin{aligned} E_{K_t}\Psi\left(m''\right) &= E_{(m',z)}\Psi\left(m'+z\right) = E_{m'}\left\{E_{z|m'}\left[\Psi\left(m'+z\right)\right]\right\} \\ &\geqslant E_{m'}\Psi\left[E_{z|m'}\left(m'+z\right)\right] = E_{m'}\Psi\left(m'\right) = E_{K_{t+1}}\Psi\left(m'\right). \end{aligned} \tag{25}$$

(ii) We apply mathematical induction to compare the relative magnitude of the value functions between different periods of cohabitation. Since this model is forward-looking, we should adopt the backward induction approach. By this we show that $V_t(m) \geqslant V_{t+1}(m)$. (a) We first check whether this is true for the two limiting cases $t_1 = \infty$ and $t_1 - 1$. If the decision maker's cohabiting with any particular partner of x type survives all periods up to t_1 where $m_{t_1} = x$, then the continuous random variable $(m_{t_1+1} \mid m_{t_1})$ degenerates into this realized type x at t_1 (recalling (12)) and thus $V_{t_1}(x) = x$, in which case R (the utility of being single) is not involved owing to the cohabitation. However, one can imagine that x is less clear at period $t_1 - 1$ than it is at t_1, and hence $(m_{t_1} \mid m_{t_1-1})$ is still contains uncertainty. Then, $E_{K_{t_1-1}(x|m)}V_{t_1}(x)$ $= m$, and so

$$V_{t_1-1}(m) = \max\left\{m + \beta E_{K_{t_1-1}(x|m)}V_{t_1}(x)\,,\ R\right\} \geqslant m(1+\beta) \geqslant m = V_{t_1}(m)\,. \quad (26)$$

(b) Suppose a similar result holds for previous periods $t+1$ and $t+2$. That is, $V_{t+1}(m) \geqslant V_{t+2}(m)$. (c) Then we prove that this is also the case for periods t and $t+1$. From (b) and (25), it follows that if $V_t(m)$ is convex, then

$$E_{K_t}V_{t+1}(m'') \geqslant E_{K_t}V_{t+2}(m'') \geqslant E_{K_{t+1}}V_{t+2}(m')\,,$$

which, under (24), implies that $V_t(m) \geqslant V_{t+1}(m)$ for any possible m.

(iii) Since $m + \beta E_{K_t(m'|m)}V_{t+1}(m')$ is larger than $m + \beta E_{K_{t+1}(m'|m)}V_{t+2}(m')$ for every m, and m_{rt} and $m_{r,t+1}$ are the values of m at which the individual is indifferent between being single (and receiving R) and continuing in a relationship, it follows that $m_{rt} > m_{r,t+1}$. Finally, letting t change across periods yields the monotonic property of $\{m_{rt}\}$.

(iv) We prove by induction that $V_t(m)$ is nondecreasing. Due to (26), we know that $V_{t_1}(x)$ and $V_{t_1-1}(m)$ are nondecreasing. Suppose $V_{t+1}(m)$ is also nondecreasing. Then, recalling the first-order stochastic dominance of a normal distribution with increasing mean, we have

$$E_{K_t(m'|m_A)}V_{t+1}(m') \geqslant E_{K_t(m'|m_B)}V_{t+1}(m'), \qquad \text{for} \quad m_A \geqslant m_B,$$

which, by (24), yields that $V_t(m_A) \geqslant V_t(m_B)$ for $m_A \geqslant m_B$.

(v) Proving that $V_t(m)$ is convex is also done by induction. Obviously, $V_{t_1}(x)$ and $V_{t_1-1}(m)$ are convex. Assuming that $V_{t+1}(m)$ is convex, define $\overline{V}_{t+1}(m) = E_{K_t(S|m)}V_{t+1}(S)$. Then, it is to be shown that $\overline{V}_{t+1}(m)$ is convex, as well. For any $\lambda \in (0,1)$ and any (m_1, m_2) (assuming $m_1 < m_2$ without losing generality), consider

$$\begin{aligned}
& \overline{V}_{t+1}\left[\lambda m_1 + (1-\lambda)m_2\right] \\
= & \int V_{t+1}(S)\, d_S\Phi\left(\frac{S - \lambda m_1 - (1-\lambda)m_2}{\sigma_t'}\right) \\
= & \int V_{t+1}\left[\lambda(S' + m_1) + (1-\lambda)(S' + m_2)\right] d\Phi\left(\frac{S'}{\sigma_t'}\right) \\
\leq & \int \left[\lambda V_{t+1}(S' + m_1) + (1-\lambda)V_{t+1}(S' + m_2)\right] d\Phi\left(\frac{S'}{\sigma_t'}\right) \\
= & \lambda \int V_{t+1}(S'')\, d_{S''}\Phi\left(\frac{S'' - m_1}{\sigma_t'}\right) + (1-\lambda)\int V_{t+1}(S''')\, d_{S'''}\Phi\left(\frac{S''' - m_2}{\sigma_t'}\right) \\
= & \lambda E_{K_t(S|m_1)}V_{t+1}(S) + (1-\lambda)E_{K_t(S|m_2)}V_{t+1}(S) = \lambda\overline{V}_{t+1}(m_1) + (1-\lambda)\overline{V}_{t+1}(m_2).
\end{aligned}$$

The second equality involves the change of variables: $S' = S - \lambda m_1 - (1-\lambda)m_2$; the inequality is the application of Jensen's inequality. The third equality uses once again the change of variables: $S'' = S' + m_1$, $S''' = S' + m_2$. Finally, for proving that

$V_t(m)$ is convex, it suffices to show

$$\begin{aligned} V_t[\lambda m_1 + (1-\lambda) m_2] &= [\lambda m_1 + (1-\lambda) m_2] + \overline{V}_{t+1}[\lambda m_1 + (1-\lambda) m_2] \\ &\leq \lambda m_1 + (1-\lambda) m_2 + \lambda \overline{V}_{t+1}(m_1) + (1-\lambda) \overline{V}_{t+1}(m_2) \\ &= \lambda \left[m_1 + \overline{V}_{t+1}(m_1)\right] + (1-\lambda)\left[m_2 + \overline{V}_{t+1}(m_2)\right] \\ &= \lambda V_t(m_1) + (1-\lambda) V_t(m_2). \end{aligned}$$

The first equality assumes $\lambda m_1 + (1-\lambda) m_2 \geqslant m_{rt}$ so that max operator and R can be dropped; the inequality uses the convexity of $\overline{V}_{t+1}(m)$; the third equality assumes m_2, $m_1 \geqslant m_{rt}$. The proof is similar and may be simpler for the other cases of (m_1, $m_2 < m_{rt}$) and ($m_1 \leq m_{r,t+1}$, $m_2 > m_{rt}$) in which R must be considered.

The fundamental difference between the two value functions $V_t(m)$ in model (24) lies in the difference in the variances $\sigma_t'^2$ of their underlying posterior distributions. Informally, although any difference in reservation pizazz $(m_{r,t+1} - m_{rt})= RP_t$ is an increasing function of $(\sigma_1'^2, \sigma_2'^2, \sigma_3'^2, \cdots\cdots)$, it is directly affected by $\sigma_t'^2$. Since $\sigma_t'^2$ decreases with time, the risk premium must be also falling. Alternatively, the variances σ_t^2 of prior distributions $F(m_t)$ increase at a decreasing rate, m_{rt} (as an increasing function of $(\sigma_1^2, \sigma_2^2, \sigma_3^2, \cdots\cdots)$) must be increasing at a decreasing rate.

References

[1] Bloch, F. and Ryder, H. (2000): "Two-Sided Search, Marriages and Matchmakers," International Economic Review, February, 41(1), 93-115.

[2] Burdett, K. and Coles, M.G. (1999): "Long-Term Partnership Formation: Marriage and Employment," The Economic Journal, No. 109, June, 307-34.

[3] ——— and ——— (1997): "Marriage and Class," Quarterly Journal of Economics, Vol. 112, February, 141-68.

[4] ——— and Wright, R. (1998): "Two-sided Search with Nontransferable Utility," Review of Economic Dynamics, Vol. 1, 220-45.

[5] Chade, H. (1999): "Two-sided Search with Noisy Observable Types," mimeo, Department of Economics, Arizona State University.

[6] Chade, H. (2001): "Two-Sided Search and Perfect Segregation with Fixed Search Costs," Mathematical Social Sciences, July, Vol. 42(1), 31-51.

[7] Collins, R.J. and McNamara, J.M. (1990): "The Job-search Problem with Competition: an Evolution Stable Dynamic Strategy," Advances in Applied Probability, Vol. 25, 314-33.

[8] Jovanovic, B. (1979): "Job Matching and Theory of Turnover" Journal of Political Economy, Vol. 87(5), 972-990.

[9] Morgan, P. (1995): "A model of search, coordination and market segmentation," Department of Economics, SUNY-Buffalo, working paper.

[10] Morgan, P. (1996): Two-sided search and matching. Department of Economics, SUNY-Buffalo, mimeo.

[11] Rao Sahib, P. and Gu, X. (2002a): "Living in Sin" and Marriage: a Matching Model," Journal of Population Economics, Vol. 15, 261-282.

[12] Rao Sahib, P. and Gu, X. (2002b): "To Tie the Knot or Not: Risk, Cohabitation and Individuals' Attitudes to Risk" Demographic Research, Vol. 6, issue 13, 353-382.

[13] Shimer R. and Smith, L. (2000): “Assortative Matching and Search,” Econometrica, March,68(2), 343-69.

[14] Sargent, T. J (1987): Dynamic Macroeconomic Theory, Harvard University Press, Cambridge, USA.

[15] Smith, L. (1997): “The marriage model with search frictions,” Working paper, Department of Economics, MIT, USA.

[16] Prescott, E. C. and Townsend, P. (1980): “Equilibrium under Uncertainty: Multi-Agent Statistical Decision Theory,” in Studies in Bayesian Econometrics and Statistics in Honor of Harold Jeffreys, Arnold Zellner, ed., North-Holland, 169-194.