European Economic Review 152 (2023) 104379

Contents lists available at ScienceDirect

European Economic Review

journal homepage: www.elsevier.com/locate/eer

A marriage-market perspective on risk-taking and career choices

Hanzhe Zhang [a,*,1], Ben Zou [b,1]

[a] Department of Economics, Michigan State University, United States of America
[b] Department of Economics, Purdue University, United States of America

ARTICLE INFO

JEL classification:
C78
J12
J16

Keywords:
Gender
Marriage market
Risk-taking
Occupational choices
Differential fecundity

ABSTRACT

We document that women are less likely than men to be in "risky" occupations, i.e., those exhibit large within-occupation wage dispersion. We first demonstrate that a new theoretical channel—the competitive structure of the marriage market—may incentivize both men and women to choose riskier careers with lower wage returns. We then show that a unifying factor—women's relative inability to reap the benefits of a risky career due to their shorter reproductive span and consequent career-family trade-off—can help rationalize a set of gender differences in labor-market and marriage-market outcomes. Using data from the United States, we provide empirical evidence that supports the model predictions.

1. Introduction

Compared with men, women are systematically less likely to take risks. For example, studies in economics, psychology, and biology consistently find that women are less likely to invest in risky assets (e.g., Cárdenas et al., 2012), engage in risky behavior (e.g., Rodham et al., 2005), and choose risky college majors (e.g., Patnaik et al., 2020). Byrnes et al. (1999) and Charness and Gneezy (2012) review a broad range of studies documenting and providing explanations for the risk gender gap.

This paper focuses on a less noticed aspect of the gender difference in risk-taking behavior: career choice manifested by occupational choice. We document that women are less represented in risky occupations, where occupational riskiness is measured as the within-occupation standard deviation of wage while adjusted for observable individual characteristics such as education, gender, age, and race. Fig. 1 shows that the within-occupation standard deviation of adjusted wage is negatively associated with the share of female workers in that occupation. This negative correlation is stronger among college graduates than among non-college graduates.

It is useful to be specific about what we mean by "risk" in this paper. What we refer to as a risky career has four defining characteristics. First, when choosing a risky career, agents decide to embark on a journey in a career whose *future payoff is uncertain*. In other words, a risky career can be interpreted from an individual perspective and in the data as a career in which the variation of realized lifetime income is large among people in the older generation who have pursued this career. Second, the uncertainty in income is specifically about *cross-sectional variation* in earnings (i.e., inequality) rather than the temporal variation for an individual

* Corresponding author.
E-mail addresses: hanzhe@msu.edu (H. Zhang), zou136@purdue.edu (B. Zou).

1 We are grateful to suggestions from Gary Becker, Pierre-André Chiappori, Jon Eguia, Steven Haider, Seungjin Han, James Heckman, Ernest Siyuan Liu, Alessandro Pavan, Phil Reny, Al Roth, Larry Samuelson, Hugo Sonnenschein, Balazs Szentes, Mallika Thomas, Richard Van Weelden, and Alessandra Voena. We thank seminar audiences at the University of Chicago, Michigan State University, University of Nebraska at Lincoln, UM-MSU-UWO Labo(u)r Day Conference, Cornell University, MEA-SOLE joint session, UM H2D2 Research Day, and SOLE. Zhang acknowledges Women and Leadership Archives, MSU Faculty Initiatives Fund, MSU Diversity Research Network, and the National Science Foundation, United States for financial support.

https://doi.org/10.1016/j.euroecorev.2023.104379
Received 9 December 2021; Received in revised form 1 December 2022; Accepted 26 December 2022
Available online 13 January 2023

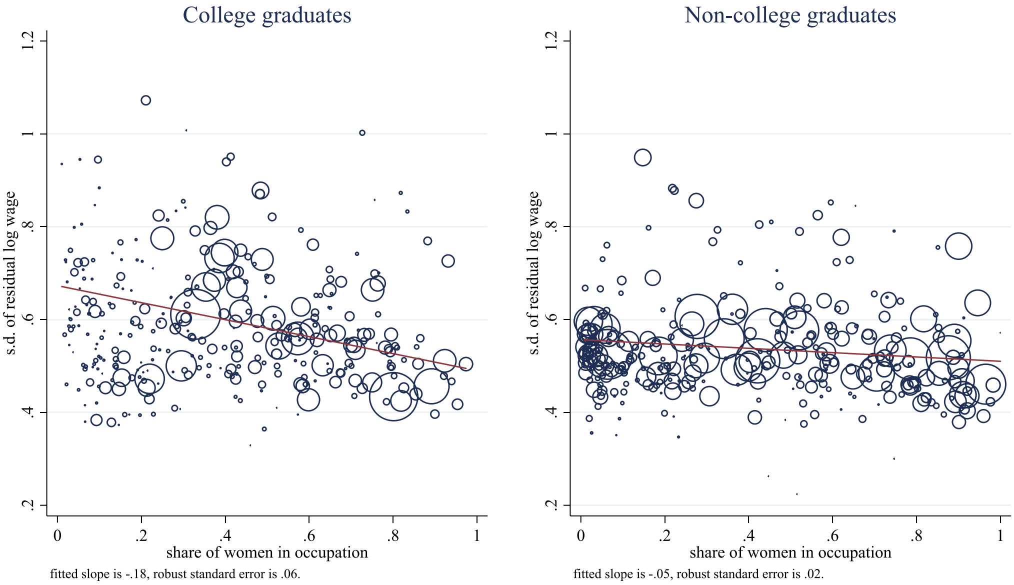


Fig. 1. Occupation Wage Riskiness and Female Ratio. Note: We use the 5-year American Community Survey ending in 2016, which provides a 5% representative sample of the US population. We restrict the sample to include those between ages 16 and 64, who are not currently enrolled in school, and report an occupation. Hourly wage is calculated by dividing the previous year's earnings by hours worked. Separately for college graduates and non-college graduates, log hourly wage is first regressed against age and age squared, an indicator for gender, a set of race indicators, as well as the exhaustive list of interactive terms of those three sets of variables. Occupation wage riskiness is measured as the standard deviation of the residual log wage within the occupation. Each bubble in the graph represents a 3-digit occupation defined by the IPUMS. The size of the bubble is proportional to the number of workers in the occupation for the specific skill group, which is also used as the weight in the linear fit regression shown in the figure. Appendix Fig. B.1 shows the same graphs by finer education groups.

(i.e., volatility). A career that could generate either a steady stream of high income or a steady stream of low income is considered a risky career before whichever steady income is realized. A career that generates a certain (discounted) lifetime income but may either front-load or back-load a large compensation is considered a safe career, since there is little uncertainty about the total future payoff. Third, the variation in realized earnings in a risky occupation, after controlling for relevant observable characteristics such as education and skills, is due to *luck*, that is, a bundle of factors that are outside one's control or uncorrelated with factors that affect the marriage market. Fourth and finally, it *takes time* for the income uncertainty from a risky career to resolve. A key insight of our paper is that, due to the competitive nature of the marriage market, it is beneficial to enter the marriage market after the uncertainty in the payoff is resolved. Because women face an additional reproductive cost from waiting, there are gender differences in career choices and marriage timing. If the uncertainty from a risky career is resolved quickly, the additional cost for women may be too small to be economically and empirically relevant.

This paper provides a novel explanation for the gender difference in risky career choices based on incentives from the marriage market. The choices of a career and a spouse are two of the most important decisions an individual could make, and the two decisions have been shown to be interlinked (e.g., Bursztyn et al., 2017; Gershoni and Low, 2021a). Numerous papers have proposed theories separately on the labor market (since Becker, 1964) and marriage market (since Becker, 1973, 1974), and many have considered simultaneous labor-market and marriage-market choices (Chiappori et al., 2009; Bhaskar and Hopkins, 2016; Fan and Zou, 2021; Zhang, 2021). Yet, few consider the active choice in the level of uncertainty to which individuals are exposed.

We first build an equilibrium marriage-market model in which occupational choices are incentivized by prospects in the marriage market. Because, all else being equal, both men and women prefer a higher-income spouse, choosing a risky career and succeeding in the chosen career is a feasible way to stand out in the competitive marriage market. We show that the gender difference in reproductive length would discourage women from embarking on a risky career whose return is uncertain and takes time to realize. This mechanism may be particularly relevant for women with college and advanced degrees—after spending many years in school, they are more likely to face a binding biological clock—consistent with the motivating evidence presented in Fig. 1.[2]

Our model generates a rich set of predictions with regard to risk-taking and career choice. First, it shows that due to the competitive nature of the marriage market, somewhat surprisingly, people may be more inclined to choose a risky career than a safe career, even if the risky career yields a *lower* expected income. Second, it predicts gender differences in premarital career

[2] Appendix Fig. B.1 plots the correlation between an occupation's riskiness and the share of women in it for five education groups: less than a high school diploma, high school graduate, some college, bachelor's degree, and advanced degree. Except for those with less than a high school diploma, women are less represented in risky occupations. The negative correlation is stronger in higher education groups. This is consistent with the hypothesis that women with more years of education face a sharper trade-off between career and family. Appendix Table B.2 shows that the main results hold for the whole sample that includes both college and non-college graduates.

choices. Women are less likely than men to choose risky careers because it is time-consuming and hence more costly—in terms of the opportunity costs in the marriage market—for them to wait for the outcome of a risky career. Third, because women are less likely to choose risky careers and are exposed to less income uncertainty, they also have a smaller within-gender income inequality. Fourth, since women systematically choose safer careers and their actual earnings are realized early, they tend to marry earlier than men. Finally, women who choose risky occupations will marry later, and due to their shorter fertility span, are less likely to have children.

We test model predictions using data from the American Community Survey (Ruggles et al., 2017). We focus on individuals with college degrees because the model's mechanisms are more relevant for this group. Throughout the paper, we use the terms "occupation" and "career" interchangeably, implicitly assuming that the choice of an occupation is informative of one's career. We measure the riskiness of an occupation as the within-occupation log wage dispersion among middle-aged workers, for whom the wage uncertainty of the occupation likely has been realized. To account for the possible correlation between an occupation's riskiness and its other features, we control for various occupation-level characteristics, such as average wage adjusted for demographics and a set of occupational amenities measures.

Empirical evidence is consistent with model predictions. First, we show that the share of people who choose risky occupations is positively associated with marriage-market competitiveness. We then provide evidence on various aspects of gender differences. The share of women in risky occupations is between 4 and 8 percentage points (pp) lower than the share of men. Consequently, wage inequality among women is between 5 and 15 log points lower. For both men and women, those who choose risky occupations tend to marry later. Choosing a risky occupation is associated with fewer children for women but not for men. We provide another set of evidence using the panel data from the National Longitudinal Study of Youth 1979 (NLSY79).

The paper makes three contributions. First, it incorporates risk-taking in an equilibrium marriage-market framework. Although many papers with an equilibrium marriage-market framework have studied how gender differences affect individuals' human capital investments and social roles (Bergstrom and Bagnoli, 1993; Siow, 1998; Iyigun and Walsh, 2007; Chiappori et al., 2009; Corinne Low, 2021; Wu and Zhang, 2021; Zhang, 2021), these papers do not consider how people voluntarily choose the *level of uncertainty* they are exposed to. This paper identifies surprising and important subtleties regarding the effects of the marriage market on risk-taking. The framework—with its endogenous determination of career choices, marriage timing, income distributions, marriage matching, and division of marriage surplus—enables us to derive a set of results that cannot be collectively explained by partial-equilibrium frameworks that are primarily constructed to empirically test specific theoretical channels.[3] Despite the model's complexity, we manage to keep it tractable and obtain closed-form solutions.

Second, the paper provides a new explanation of risk-taking due to marriage market incentives (Proposition 1). Many papers have provided explanations for why people take risks, including overconfidence and social status concerns (Smith, 1776; Becker et al., 2005), preferences for lotteries (Friedman and Savage, 1948; Friedman, 1953), subsistence concerns (Rubin and Paul, 1979), incentives to relocate (Rosen, 1997), and related to marriage market concerns, the role of polygamous marriages (Robson, 1992, 1996). This paper shows that competitiveness in the one-to-one matching market encourages risk-taking.[4]

Third, this paper uses a parsimonious assumption (the biological clock of producing offspring) to provide a unified explanation of gender differences in career choices, marriage matching, income inequality, the timing of marriage, and fertility decisions (Propositions 2 to 5). Our model, though simple and stylized, is able to shed light on a wide range of socioeconomic gender differences that are documented in existing empirical and experimental studies. Of course, we are not denying the important role of gender differences along other dimensions in explaining those phenomena. Alternative explanations include gender differences in evolutionary biology, in psychological traits such as overconfidence, taste for risk, and attitudes toward competitiveness, and different social norms toward men and women.[5] The predictions and empirical evidence of the paper should be interpreted as complementary to existing explanations.

The rest of the paper is organized as follows. Section 2 sets up the benchmark model and shows how the marriage market encourages risk-taking. Section 3 constructs a parametric generalization of the benchmark model to incorporate more gender differences and provides testable predictions. Section 4 provides empirical evidence consistent with theoretical predictions. Section 5 concludes.

[3] Previous partial-equilibrium frameworks focus on providing evidence that marital incentives affect education and labor supply decisions (Goldin and Katz, 2002; Bailey, 2006; Bertrand et al., 2010; Lafortune, 2013; Adda et al., 2017; Bronson, 2019), but few have focused on the effects of marital incentives on career choices.

[4] The channel highlighted in this paper is not restricted to marriage markets. Any competitively organized two-sided one-to-one matching market encourages risk-taking: Zhang (2020) elaborates on the theoretical argument that a competitive transferable-utilities matching market induces extreme gambles regardless of the shape of the surplus function and investigates the multiplicities and inefficiencies of equilibrium investments in two-sided matching markets. While Zhang (2020) studies the general theoretical model, this paper focuses on a more specific application.

[5] Previous studies have investigated gender differences in competitiveness (e.g., Niederle and Vesterlund, 2007; Kleinjans, 2009; Buser et al., 2014; Gill and Prowse, 2014; Wozniak et al., 2014), in risk preferences and beliefs (e.g., Altonji and Blank, 1999; Eckel and Grossman, 2002; Barbulescu and Bidwell, 2012; Koellinger et al., 2013; Zafar, 2013; Patnaik et al., 2020; Wiswall and Zafar, 2021), in genetic features (e.g., Dreber and Hoffman, 2007), due to gender social norms (e.g., Bursztyn et al., 2017; Adda et al., 2017), and due to statistical discrimination in the labor market (e.g., Xiao, 2021).

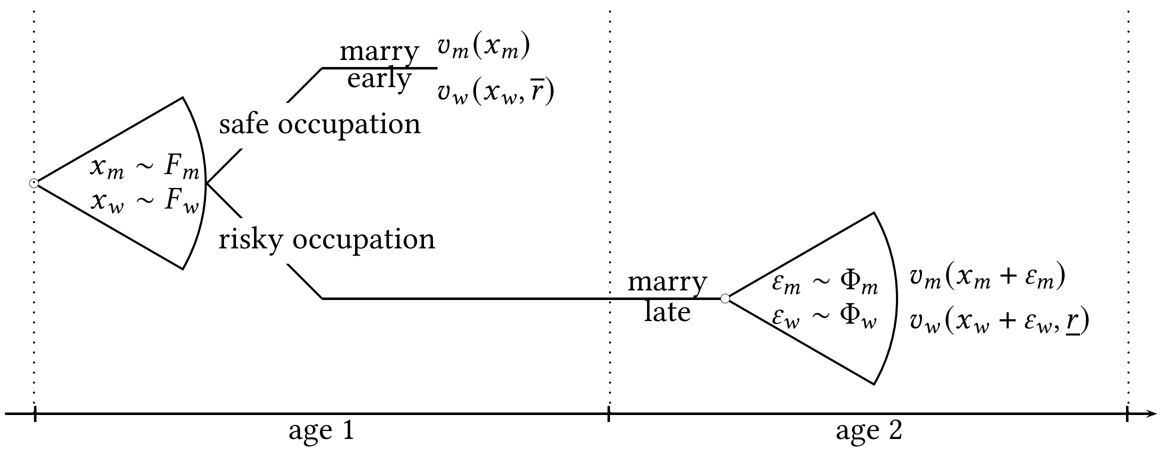


Fig. 2. An individual's occupational and marital decisions in the benchmark model.

2. The benchmark model

We start with the simplest model to demonstrate how people have a marriage-market incentive to choose a risky career. We will include additional gender differences in the more general model. Time is discrete and infinite: $t = 1, 2, \ldots$. At the beginning of each period, unit masses of men and women are born with heterogeneous *income-earning abilities* x_m and x_w, which are distributed according to continuous and strictly increasing mass distributions F_m and F_w with supports $X_m \equiv [\underline{x}_m, \overline{x}_m] \subset \mathbb{R}$ and $X_w \equiv [\underline{x}_w, \overline{x}_w] \subset \mathbb{R}$. All results will continue to hold assuming gender imbalance: Unproductive individuals can be added to the shorter side of the market to restore gender balance. Over two periods (roughly, their twenties and thirties), an agent makes career and marriage decisions, which are illustrated in Fig. 2 and elaborated next.

2.1. Model setup

2.1.1. Career choices

Each agent chooses a *safe career* or a *risky career* at the beginning of the first period of their decision-making. A safe career compensates a person's human capital with certainty: An ability-x_m man who chooses a safe career receives an income $y_m = x_m$, and an ability-x_w woman who chooses a safe career receives income $y_w = x_w$. In contrast, a risky career noisily compensates a person's human capital: An ability-x_m man who chooses a risky career receives income $y_m = x_m + \varepsilon_m$, where ε_m is distributed according to cumulative distribution function $\Phi_m(\cdot|x_m)$, and an ability-x_w woman who chooses a risky career receives income $y_w = x_w + \varepsilon_w$, where ε_w is distributed according to cumulative distribution function $\Phi_w(\cdot|x_w)$.

This setup implies that the risk we discuss is the cross-sectional inequality in income rather than its temporal fluctuation. Given one's ability (x_m and x_w), whether one succeeds or fails in a risky career (a realization of a high or low ε_m and ε_w) is uncorrelated with other individual characteristics and is solely attributed to "luck" that is outside one's control. As Fig. 2 illustrates, we also assume that the realization of such risk takes enough time to cause women's reproductive fitness to decline.

In the benchmark model, we assume that a person who chooses the safe career enters the marriage market immediately in the current period, and a person who chooses the risky career waits until the income is realized in the next period to enter the marriage market. Later we will extend the model and allow separate career and marriage timing choices, but Proposition 6 shows that it is not desirable for a person who chooses the safe career to marry late or for a person who chooses the risky career to marry early, so it is without loss of generality to assume that an individual who chooses a risky career waits to marry and an individual who chooses a safe career marries immediately. Let $p_m(x_m)$ represent the probability that an ability-x_m man chooses a risky career, and $p_w(x_w)$ the probability that an ability-x_w woman chooses a risky career. Throughout the paper, we use $p_m(\cdot)$ and $p_w(\cdot)$ to represent population strategies and p_m and p_w to represent individual strategies.

2.1.2. The marriage market

When women enter the marriage market late, they face a reproductive decline; but men do not. Namely, whereas men remain reproductively fit throughout both periods, women who enter the marriage market early are reproductively fit ($r = \overline{r}$), but women who enter the marriage market late are less fit ($r = \underline{r} < \overline{r}$). We choose to model reproductive fitness as an additional dimension of women's characteristics for generality, but it is qualitatively equivalent to assume that a woman incurs a cost when entering the marriage market late. In the parameterized version of the model we present later, women's reproductive dimension is conveniently collapsed so that each woman is represented by a single index that encompasses income and reproductive fitness.

Aggregate career choices $p_m(\cdot)$ and $p_w(\cdot)$ lead to income distributions G_m and G_w for men and women, respectively. For any y_m, men with an income below y_m include those who have an ability below y_m and choose the safe career and those who choose the risky career and realize an income below y_m:

$$G_m(y_m|p_m(\cdot)) \equiv \int_{\underline{x}_m}^{y_m} [1 - p_m(x_m)]dF_m(x_m) + \int_{\underline{x}_m}^{\overline{x}_m} \Phi_m(y_m - x_m|x_m)p_m(x_m)dF_m(x_m). \tag{1}$$

Similarly, women with an income y_w consist of two groups, due to different career choices, who are fit and less fit, respectively. Since women who choose a safe career and enter the marriage market in the first period are fit, the mass of fit women with an income below y_w is

$$G_w(y_w, \overline{r}|p_w(\cdot)) \equiv \int_{\underline{x}_w}^{y_w} [1 - p_w(x_w)]dF_w(x_w), \tag{2}$$

and the mass of less-fit women with an income below y_w is

$$G_w(y_w, \underline{r}|p_w(\cdot)) \equiv \int_{\underline{x}_w}^{\overline{x}_w} \Phi_w(y_w - x_w|x_w)p_w(x_w)dF_w(x_w). \tag{3}$$

The lifetime marriage surplus an income-y_m man and an income-y_w woman with reproductive fitness r produce is $s(y_m, y_w, r)$. The lifetime marriage surplus can be thought of as the resulting indirect utility of a household production problem in which the husband and the wife allocate their time and resources to the production of private goods and public goods, given their incomes and reproductive fitness; see Appendix A.1 for a household public good provision problem that justifies the use of specific surplus functions as well as transferable utilities. To focus on career choices, we also assume that the marriage surplus does not depend on when agents marry but only on agents' marital types. Normalize the surplus any unmarried agent produces to zero. Assume that the surplus function is twice differentiable in incomes and is strictly increasing in each of the three arguments.

Men and women match and negotiate the division of their marriage surplus to reach a stable outcome in which no pair of a man and a woman could strictly improve their payoffs.

Definition 1. A *stable outcome* of the marriage market characterized by income distributions (G_m, G_w) consists of a *matching* G and *marriage payoff functions* $(v_m(\cdot), v_w(\cdot,\cdot))$ such that (1) stable matching $G(y_m, y_w, r)$ describes the mass of couples with incomes lower than y_m and y_w such that the marginals of G are G_m and G_w; (2) stable marriage payoffs $v_m(y_m) \geq 0$ and $v_w(y_w, r) \geq 0$ satisfy the following two stability conditions: (a) Every couple divides the marriage surplus: For any (y_m, y_w, r) in the support of G, $v_m(y_m) + v_w(y_w, r) = s(y_m, y_w, r)$; (b) No division of surplus could make any unmatched pair of man and woman strictly better off: For any (y_m, y_w, r), $v_m(y_m) + v_w(y_w, r) \geq s(y_m, y_w, r)$. By Theorem 2 of Gretsky et al. (1992), a stable outcome exists.

2.1.3. Payoffs

Each person is risk-neutral and does not discount. A person's payoff is simply their marriage payoff. That is, an income-y_m man's payoff is $v_m(y_m)$, and an income-y_w woman's payoff is either $v_w(y_w, \overline{r})$ if she marries in the first period, or $v_w(y_w, \underline{r})$ if she marries in the second period. Given men's marriage payoff schedule $v_m(\cdot)$, an ability-x_m man's expected payoff from strategy p_m is

$$u_m(p_m, x_m|v_m(\cdot)) \equiv p_m\mathbb{E}\left[v_m(x_m + \varepsilon_m)|x_m\right] + (1 - p_m)v_m(x_m), \tag{4}$$

and given women's marriage payoff schedule $v_w(\cdot,\cdot)$, an ability-x_w woman's expected payoff from strategy p_w is

$$u_w(p_w, x_w|v_w(\cdot,\cdot)) \equiv p_w\mathbb{E}\left[v_w(x_w + \varepsilon_w, \underline{r})|x_w\right] + (1 - p_w)v_w(x_w, \overline{r}). \tag{5}$$

2.2. Equilibrium

In summary, the model's primitives are (1) ability distributions $F_m(\cdot)$ and $F_w(\cdot)$; (2) income distributions from a risky career for each individual, $\Phi_m(\cdot|\cdot)$ and $\Phi_w(\cdot|\cdot)$; and (3) the marriage surplus function $s(\cdot,\cdot,\cdot)$. Hence, $(F_m, F_w, \Phi_m, \Phi_w, s)$ summarizes the model. An equilibrium of the model is defined as follows. In an equilibrium, each agent chooses the career that maximizes their expected marriage payoff, and the marriage payoffs are the stable marriage payoffs in the marriage market induced by agents' career choices. Formally:

Definition 2. $(p_m^*(\cdot), p_w^*(\cdot), G_m^*, G_w^*, G^*, v_m^*, v_w^*)$ is an *equilibrium* of $(F_m, F_w, \Phi_m, \Phi_w, s)$ if (1) $p_m^*(x_m)$ maximizes an ability-x_m man's expected payoff when men's marriage payoff is $v_m^*(\cdot)$ and $p_w^*(x_w)$ maximizes an ability-x_w woman's expected payoff when women's marriage payoff is $v_w^*(\cdot)$; (2) Men's income distribution $G_m^*(\cdot)$ is induced by men's career choices $p_m^*(\cdot)$ and fit and less fit women's income distributions $G_w^*(\cdot, \overline{r})$ and $G_w^*(\cdot, \underline{r})$ are induced by women's career choices $p_w^*(\cdot)$; and (G^*, v_m^*, v_w^*) is a stable outcome of the marriage market (G_m^*, G_w^*).

Theorem 1. *An equilibrium exists.*

We apply Glicksberg's fixed-point theorem to prove equilibrium existence. See Appendix A.2 for details. Multiple equilibria may arise, with similar reasoning as in coordination problems; we provide an example in Appendix A.3. In Section 3, we use a parameterized model with a unique equilibrium to derive additional implications. The generalized model also extends the benchmark model by allowing gender-differential risk preferences and endogenous costs of choosing different careers. Before we turn to the parametric model, we use the benchmark model to demonstrate how the marriage market encourages risk-taking and to elucidate how this marriage-market incentive for risk-taking is independent of parametric assumptions.

2.3. Risk-taking due to marriage-market incentives

In the remainder of the section, we highlight an inherent force in the competitive marriage market that encourages risk-taking. When the forces against risk-taking (concretely, concavity of the surplus function in the benchmark model and risk aversion and additional costs in the extension) are not too strong, the market force that encourages risk-taking dominates and manifests in people's choice of a risky career with a low expected income and a high income variance, *without* relying on risk-loving preferences.

To understand this inherent force that drives risk-taking, we must emphasize the competitive organization of the marriage market. The key property of the competitive marriage market is that *each person marries the partner that maximizes their marriage payoff*. To understand this key property, consider how the division of marriage surplus is determined in the market.

First, when an income-y_m man marries a woman with characteristics $z_w(y_m)$, he and his partner divide up their marriage surplus: $v_m(y_m) + v_w(z_w(y_m)) = s(y_m, z_w(y_m))$. In other words, an income-y_m man gets a payoff that is the total surplus he and a type-$z_w(y_m)$ woman generate net the $z_w(y_m)$ woman's payoff:

$$v_m(y_m) = s(y_m, z_w(y_m)) - v_w(z_w(y_m)). \tag{6}$$

Second, given women's stable marriage payoff schedule $v_w(\cdot)$, no woman can form a pair with a man and improve both of their payoffs. In other words, the total hypothetical surplus the man and any other woman generate must be lower than the sum of the payoffs they are getting in their current stable outcome. Mathematically, $v_m(y_m) + v_w(z_w) \geq s(y_m, z_w) \quad \forall z_w \neq z_w(y_m)$. From a man's own perspective, his current payoff is better than the hypothetical payoff he could get by marrying any other type of woman:

$$v_m(y_m) \geq s(y_m, z_w) - v_w(z_w) \quad \forall z_w \neq z_w(y_m). \tag{7}$$

The two stability conditions (Eqs. (6) and (7)) together yield

$$v_m(y_m) = s(y_m, z_w(y_m)) - v_w(z_w(y_m)) \geq s(y_m, z_w) - v_w(z_w) \quad \forall z_w \neq z_w(y_m), \tag{8}$$

that is, each man's marriage payoff is the most he can get given women's marriage payoff schedule; a man is married to the wife that maximizes his marriage payoff. There is no restriction for a man to marry any woman as long as he is willing to give a woman her stable marriage payoff (or any payoff above it). However, the deduction above shows that he has no strictly higher incentive to marry anyone other than his partner in the stable outcome because that partner provides him the highest marriage payoff given women's stable marriage payoff schedule.

How does this property encourage risky career choices? Consider the following example to see this effect in isolation. Suppose the marriage surplus is linear in the man's income, and an ability-$\tilde{y}_m$ man chooses between a safe career and a risky career whose income realization is simply a mean-preserving spread of the income obtained from the safe career. If the man marries type $z_w = (y_w, r)$ woman regardless of his income realization y_m, his payoff is $v_m(y_m|z_w) = s(y_m, z_w) - v_w(z_w)$, depicted by a solid blue line in Fig. 3. Since the income from the risky career is assumed to be a mean-preserving spread of the income $\tilde{y}_m$ from the safe career, if a man were always to marry a z_w woman, he would be indifferent between a risky career and a safe career. However, the man marries the partner that gives him the highest payoff in the competitive marriage market, and that payoff-maximizing partner may not always be z_w. In fact, as Fig. 3 illustrates, when he realizes an income higher than $\tilde{y}_m$, he can achieve a better payoff by marrying z''_w than by marrying z_w; when he realizes an income lower than $\tilde{y}_m$, he can still achieve a better payoff by marrying z'_w than by marrying z_w. Therefore, if there are three types of women in the marriage market, z_w, z'_w, and z''_w, men's marriage payoff schedule is the upper envelope of three lines, which is piecewise linear and hence weakly convex. Furthermore, if there is a continuum of types of women, men's marriage payoff function $v_m(\cdot)$ is strictly convex, as illustrated by Fig. 4(a). When the marriage payoff is strictly convex, the man has a strict incentive to choose a risky career as long as it yields the same expected income as a safe career.

Therefore, when the marriage surplus is linear, men always choose a risky career if the risky career's income distribution is a mean-preserving spread of the safe career's, due to the incentives provided by the competitive marriage market.

Lemma 1. *If the marriage surplus is linear in the man's income and the risky career has the same expected income as the safe career, it is strictly dominant for each man to choose the risky career.*

When the surplus is linear, and there is heterogeneity in types on the other side of the market, we show that the marriage payoff is strictly convex everywhere. When the surplus is concave, the marriage payoff is not strictly convex everywhere, but in certain ranges, the marriage payoff function may still be convex and people may still have a strict incentive to choose the risky career, even if the risky career has a lower expected income than the safe career (see Fig. 4(b)).

Lemma 2. *An unmarried person might choose a career with a lower expected income and higher income uncertainty even if the marriage surplus function is strictly concave in income.*

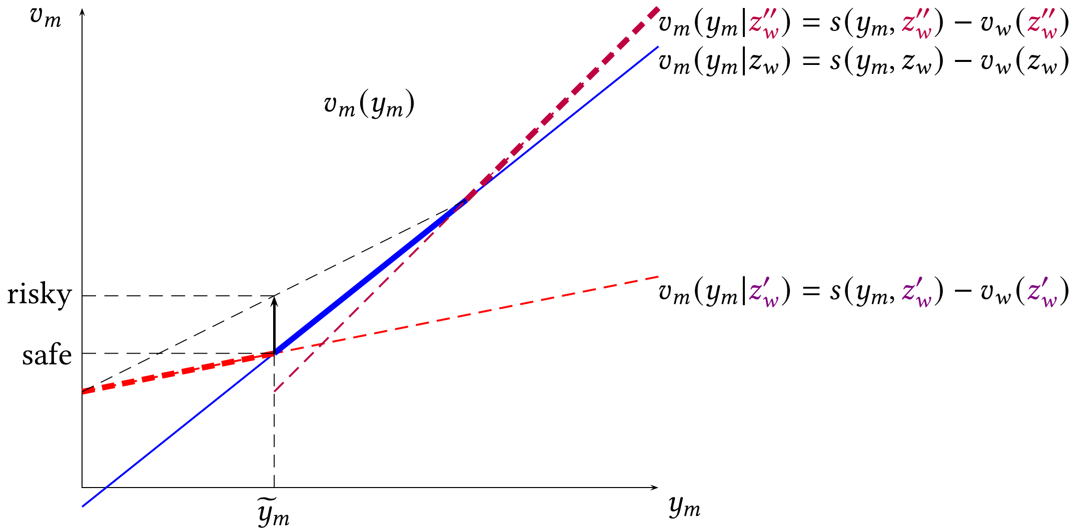
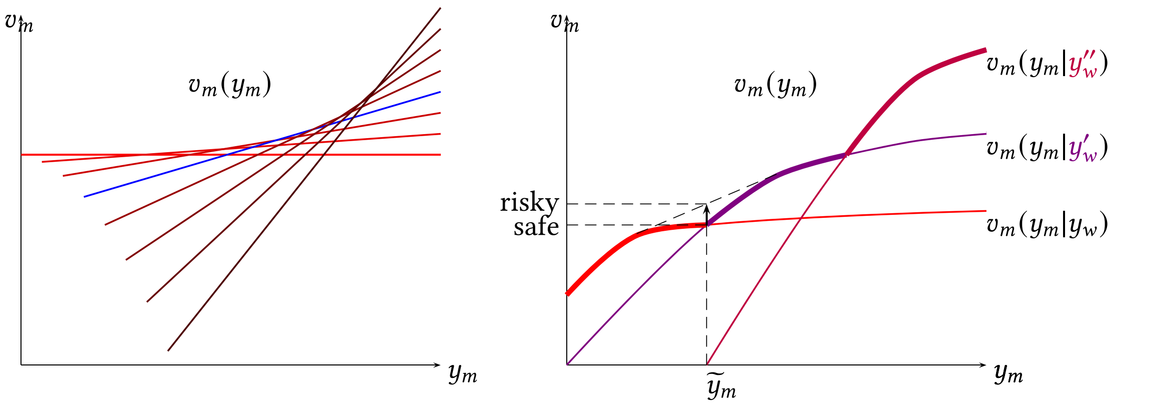


Fig. 3. A linear surplus leads to a weakly convex payoff.



(a) A linear surplus and heterogeneity lead to a strictly convex payoff.

(b) A concave surplus can lead to a partially convex payoff.

Fig. 4. Convex payoff.

This lemma helps rationalize risky career choices without relying on risk preferences. A person can be justified in choosing a risky, badly paid career that has lower expected income and higher income variance as long as he or she has marital concerns. As a result, it may be well justified for someone to choose a profession in which superstars are "overcompensated", but others are "undercompensated". Individuals facing such a wage structure may include actors, musicians, and lawyers. Previous papers have explained such risky choices as status concerns or overconfidence. This paper provides an additional explanation based on the marriage market.

Note that these arguments do not rely on surplus supermodularity. The surplus could have been strictly submodular (e.g., $s(y_m, y_w, r) = r(y_m + y_w - y_m y_w)$), and the force that encourages risk-taking would still hold. If a man chooses a risky career, compared with the women he would marry if he chooses a safe career, he would be marrying a *lower*-type woman when he realizes a higher income and a *higher*-type woman when he realizes a lower income—but nonetheless, these choices maximize his marriage payoff and convexify the payoff function.

In addition, the linearity of surplus in men's income is sufficient but not necessary for risk-taking. As Proposition 1 will show, agents may prefer risky careers even if they are risk-averse and the surplus function is strictly concave. The key ingredient that drives the strict dominance of the risky career is the competitive nature of the marriage market, and the force is stronger when the type distribution on the opposite side of the market is more heterogeneous.

Even though women have the same marriage-market incentive that encourages risk-taking, their reproductive decline acts as a cost that deters them from choosing the risky career. Exactly how much women are deterred from the risky career is captured by the following extension of the benchmark model.

3. The extended model

In this section, we extend the benchmark model. We allow people to have preferences that are not risk-neutral. We also let the income realization from choosing a risky career depend on the percentage of other people in the economy who choose the same career. Furthermore, in the next section, we separate career choice and marriage timing (i.e., someone who chooses a risky career can also marry early). To keep the model tractable and derive closed-form solutions, we impose reasonable functional form assumptions on the income-earning ability distributions, the income distributions of the risky career, and the marriage surplus function.

Time is still discrete and infinite: $t = 1, 2, \ldots$. At the beginning of each period, a unit mass of men and a unit mass of women, endowed with heterogeneous (log-income) abilities $x_m \sim N(\mu_{x_m}, \sigma^2_{x_m})$ and $x_w \sim N(\mu_{x_w}, \sigma^2_{x_w})$, choose between a safe career and a risky career. Anyone who chooses the safe career realizes their income and enters the marriage market in the current period, and anyone who chooses the risky career realizes their income and enters the marriage market in the next period.

3.1. Parameterization and the equilibrium

3.1.1. The marriage market

In the marriage market, men are distinguished by income only, and women are distinguished by income and reproductive fitness. Women who enter the marriage market in the first period are more fit than those who enter in the second period. Namely, a log-income-y_m man and a log-income-y_w and fitness-r woman produce a marriage surplus $s(y_m, y_w, r) = \exp(\alpha_m y_m + \alpha_w(y_w - 1_{r=\underline{r}}k)) \equiv \exp(\alpha_m z_m + \alpha_w z_w)$. The marriage surplus is a Cobb–Douglas function in marriage type. Note that the marriage surplus is strictly increasing and strictly supermodular in marriage types z_m and $z_w = y_w - 1_{r=\underline{r}}k$ as well as in log-incomes y_m and y_w.

In the marriage market, men and women frictionlessly match and bargain over the division of their marriage surplus until a stable outcome is reached. A stable outcome of the marriage market is described by stable matching distributions $z_m(\cdot)$ and $z_w(\cdot)$ as well as stable marriage payoff functions $v_m(\cdot)$ and $v_w(\cdot)$ such that (1) everyone gets a nonnegative payoff: $v_m(z_m) \geq 0$ and $v_w(z_w) \geq 0$ for all z_m and z_w; (2) every married couple divides the surplus: $v_m(z_m) + v_w(z_w(z_m)) = s(z_m, z_w(z_m))$ and $v_m(z_m(z_w)) + v_w(z_w) = s(z_m(z_w), z_w)$ for all z_m and z_w; and (3) no pair of man and woman who are not married to each other have a strict incentive to depart from their current marriages and marry each other instead: $v_m(z_m) + v_w(z_w) \geq s(z_m, z_w)$ for all z_m and z_w.

Stable matching is positive assortative because the surplus is strictly supermodular. Given the type distributions $N(\mu_{z_m}, \sigma^2_{z_m})$ and $N(\mu_{z_w}, \sigma^2_{z_w})$, if z_m and z_w are matched, their percentiles are the same: $(z_m - \mu_{z_m})/\sigma_{z_m} = (z_w - \mu_{z_w})/\sigma_{z_w}$. From the stability conditions, we can also derive stable marriage payoff functions, summarized as follows.

Lemma 3. *Suppose marriage types are normally distributed $N(\mu_{z_m}, \sigma^2_{z_m})$ and $N(\mu_{z_w}, \sigma^2_{z_w})$. Stable matching functions are*

$$z_m(z_w) = \frac{\sigma_{z_m}}{\sigma_{z_w}}(z_w - \mu_{z_w}) + \mu_{z_m} \text{ and } z_w(z_m) = \frac{\sigma_{z_w}}{\sigma_{z_m}}(z_m - \mu_{z_m}) + \mu_{z_w}.$$

Stable marriage payoff functions are

$$v_m(z_m) = \frac{\alpha_m}{\alpha_m + \alpha_w \frac{\sigma_{z_w}}{\sigma_{z_m}}} \exp(\alpha_m z_m + \alpha_w z_w(z_m)) \text{ and } v_w(z_w) = \frac{\alpha_w}{\alpha_w + \alpha_m \frac{\sigma_{z_m}}{\sigma_{z_w}}} \exp(\alpha_w z_w + \alpha_m z_m(z_w)).$$

Since there is an equal mass of men and women in the marriage market and marriage surplus is nonnegative, everyone marries immediately upon entering the marriage market, so from now on, it is equivalent to saying "to enter the marriage market" and "to marry". The detailed proof is provided in Appendix A.4.

3.1.2. Career choices

A safe career yields an ability-x_m man log-income $y_m = x_m$, and yields an ability-x_w woman log-income $y_w = x_w$. A risky career yields an ability-x_m man log-income $y_m = x_m - c + \varepsilon_m$, where c is a market-determined cost of taking the risky career, and $\varepsilon_m \sim N(t_m, s^2_m)$. Similarly, a risky career yields an ability-x_w woman log-income $y_w = x_w - c + \varepsilon_w$, where $\varepsilon_w \sim N(t_w, s^2_w)$. The cost c depends on the mass of people choosing the risky career. The more people choose it, the higher the cost: $c'(p_m + p_w) > 0$.

Let $p_m(x_m)$ and $p_w(x_w)$ be an ability-x_m man's and an ability-x_w woman's probability of choosing the risky career, respectively. Proportion $p_m = \int_{-\infty}^{\infty} p_m(x_m) d\Phi(\frac{x_m - \mu_{x_m}}{\sigma_{x_m}})$ of men and proportion $p_w = \int_{-\infty}^{\infty} p_w(x_w) d\Phi(\frac{x_w - \mu_{x_w}}{\sigma_{x_w}})$ of women choose the risky career, where Φ is the standard normal cumulative distribution function. Since abilities are assumed to be normally distributed, men's income distribution is $LN(\mu_{y_m}, \sigma^2_{y_m})$, where $\mu_{y_m} = \mu_{x_m} + p_m(t_m - c)$ and $\sigma^2_{y_m} = \sigma^2_{x_m} + p_m \frac{s^2_m}{2}$, and women's income distribution is $LN(\mu_{y_w}, \sigma^2_{y_w})$, where $\mu_{y_w} = \mu_{x_w} + p_w(t_w - c)$ and $\sigma^2_{y_w} = \sigma^2_{x_w} + p_w \frac{s^2_w}{2}$.

Agents derive utility from stable marriage payoffs according to $u_m(v_m) = \frac{v_m^{1-\rho_m}}{1-\rho_m}$ and $u_w(v_w) = \frac{v_w^{1-\rho_w}}{1-\rho_w}$. When $\rho_m = \rho_w = 0$, men and women are risk-neutral, and when $\rho_m > 0$ and $\rho_w > 0$, they are risk-averse. We can now derive the optimal career choices.

Lemma 4. *Suppose marriage types are normally distributed $N(\mu_{z_m}, \sigma^2_{z_m})$ and $N(\mu_{z_w}, \sigma^2_{z_w})$. Define*

$$\delta_m \equiv t_m - c + (1 - \rho_m)\left(\alpha_m + \alpha_w \frac{\sigma_{z_w}}{\sigma_{z_m}}\right)\frac{s^2_m}{2},$$

$$\delta_w \equiv t_w - c - k + (1-\rho_w)\left(\alpha_w + \alpha_m \frac{\sigma_{z_m}}{\sigma_{z_w}}\right)\frac{s_w^2}{2}.$$

Every man chooses the risky career if $\delta_m > 0$, the safe career if $\delta_m < 0$, and is indifferent between the two if $\delta_m = 0$. Every woman chooses the risky career if $\delta_w > 0$, the safe career if $\delta_w < 0$, and is indifferent the two if $\delta_w = 0$.

A proof is provided in Appendix A.5.

The terms δ_m and δ_w are key to understanding the model. First, note that δ_m and δ_w do not depend on x_m or x_w: In this parametric model, the incentive for choosing a risky career is the same for every man and for every woman. Second, decompose δ_m into the following three terms:

$$\delta_m = \left[t_m - c + \frac{s_m^2}{2}\right] + \left[((1-\rho_m)\alpha_m - 1)\frac{s_m^2}{2}\right] + \left[(1-\rho_m)\alpha_w \frac{\sigma_{z_w}}{\sigma_{z_m}}\frac{s_m^2}{2}\right].$$

The first term, $\left[t_m - c + \frac{s_m^2}{2}\right]$, is the difference in expected income between the risky career and the safe career. The second term, $\left[((1-\rho_m)\alpha_m - 1)\frac{s_m^2}{2}\right]$, is the expected gain in own payoff from choosing the risky career, *without* changing a marriage partner. If the marriage surplus is linear in the man's income and the man is risk-neutral, then this term is zero; if the marriage surplus is concave in the man's income and the man is risk-averse, then the term is negative. The third and final term, $\left[(1-\rho_m)\alpha_w \frac{\sigma_{z_w}}{\sigma_{z_m}}\frac{s_m^2}{2}\right]$, is the expected gain due to a changed partner in the marriage market. This term is positive regardless of the shape of the surplus function and the marriage-type distributions, as long as the man is not too risk-averse ($1-\rho_m > 0$) and there is some heterogeneity on the other side of the market ($\sigma_{z_w} > 0$). This third term is what drives risky career choices for moderately risk-averse agents.

The same decomposition can be done for δ_w.

$$\delta_w = -k + \left[t_w - c + \frac{s_w^2}{2}\right] + \left[((1-\rho_w)\alpha_m - 1)\frac{s_w^2}{2}\right] + \left[(1-\rho_w)\alpha_m \frac{\sigma_{z_m}}{\sigma_{z_w}}\frac{s_w^2}{2}\right].$$

It has an additional term $-k$, which is the loss in payoff associated with declined reproductive fitness.[6] All else equal, women have a lower expected gain from the risky career because of the reproductive fitness loss.

3.1.3. Equilibrium

Definition 3. $(p_m^*(\cdot), p_w^*(\cdot), v_m^*(\cdot), v_w^*(\cdot))$ is an *equilibrium* if

1. $p_m^*(x_m)$ maximizes each ability-x_m man's expected utility, and $p_w^*(x_w)$ maximizes each ability-x_w woman's expected utility, given equilibrium marriage payoff functions $v_m^*(\cdot)$ and $v_w^*(\cdot)$ as well as market-determined cost $c^* = c(p_m^* + p_w^*)$, where

$$p_m^* = \int_{-\infty}^{\infty} p_m^*(x_m) d\Phi\left(\frac{x_m - \mu_{x_m}}{\sigma_{x_m}}\right) \text{ and } p_w^* = \int_{-\infty}^{\infty} p_w^*(x_w) d\Phi\left(\frac{x_w - \mu_{x_w}}{\sigma_{x_w}}\right).$$

2. Equilibrium marriage payoff functions $v_m^*(\cdot)$ and $v_w^*(\cdot)$ are the stable marriage payoff functions of the marriage market with type distributions $N(\mu_{z_m}, \sigma_{z_m}^2)$ and $N(\mu_{z_w}, \sigma_{z_w}^2)$, where

$$\mu_{z_m} = \mu_{y_m} + p_m(t_m - c), \quad \sigma_{z_m}^2 = \sigma_{y_m}^2 = \sigma_{x_m}^2 + p_m s_m^2/2,$$

$$\mu_{z_w} = \mu_{y_w} + p_w(t_w - c - k), \quad \text{and} \quad \sigma_{z_w}^2 = \sigma_{y_w}^2 = \sigma_{x_w}^2 + p_w s_w^2/2.$$

Theorem 2. *An equilibrium exists, and exists uniquely if $\rho_m < 1$ and $\rho_w < 1$.*

3.2. Model predictions

3.2.1. Marriage-market incentives and risky career choices

The parameterized model helps determine exactly when a person prefers a risky career to a safe career, which provides sharper predictions than Lemmas 1 and 2.

[6] Note that women bearing the full reproductive cost k is an equilibrium outcome rather than an assumption. A lower reproductive fitness reduces the total marriage surplus, but who bears the reproductive cost is up to the equilibrium division of the surplus. The full bearing of reproductive cost by women is not generally true and has to do with the parameterization of the model, where marriage surplus is separable in men's income, women's income, and women's reproductive fitness. But because fitness is one of women's traits and affects their bargaining power in the marriage market, reasonably similar models would still result in women bearing the majority of the reproductive cost and are thus less likely to choose a risky career.

Proposition 1. *Between a safe career and a risky career that returns lower expected income, a man would strictly prefer the risky career if* $(1-\rho_m)(\alpha_m + \alpha_w \frac{\sigma_{y_w}}{\sigma_{y_m}}) > 1$*, and a woman would strictly prefer the risky career if* $(1-\rho_w)(\alpha_w + \alpha_m \frac{\sigma_{y_m}}{\sigma_{y_w}}) > 1 + 2k/s_w^2$*.*

A risk-neutral man ($\rho_m = 1$) would strictly prefer the risky career if the marriage surplus is linear in men's income ($\alpha_m = 1$). Furthermore, risk-averse agents may choose risky careers, and if the setting is gender-symmetric except for reproductive cost k, then women are less likely to choose the risky career. The gender difference in the two expressions is $2k/s_w^2$. That is, the higher reproductive cost and *lower* income variance of the risky career make women less inclined to choose a risky career.

The competitive nature of the marriage market encourages both men and women to engage in risk-taking behavior. The more competitive the marriage market becomes, the more likely men and women choose a risky career. In particular, the higher the variance of the income distribution of the *opposite* sex in their marriage market, the more likely they choose a risky career.

3.2.2. Gender difference in risky career choice and income dispersion

For the following propositions, we suppose ability distributions and career opportunities are gender-symmetric ($\sigma_{x_m}^2 = \sigma_{x_w}^2 \equiv \sigma_x^2$, $s_m = s_w \equiv s$, and $t_m = t_w \equiv t$).

The result whereby women are less likely than men to choose a risky career continues to hold in the general equilibrium. Note that the result is shown to hold unambiguously when there is complete gender symmetry in the model. By continuity, the result continues to hold when there is moderate gender asymmetry.

Proposition 2. *Men are more likely than women to choose the risky career.*

Because women are less likely to choose a risky career, consequently, women's income inequality is less than men's.

Proposition 3. *Variance in (log) earnings is greater for men than for women ($\sigma_{y_m}^{*2} > \sigma_{y_w}^{*2}$).*

Note that for this result, we do not need to assume that the ability distributions are gender-symmetric. For *any* underlying ability distributions, we will obtain the result whereby men's income inequality is larger than women's, as long as there is a reproductive difference. Furthermore, comparative statics results suggest that the gender disparity in income inequality increases if the reproductive cost increases, if the risk aversion increases, or if the risky career becomes more uncertain.

Remark 1. The ratio of the log-income variances, $\sigma_{y_m}^{*2}/\sigma_{y_w}^{*2}$, increases in career cost k, increases in risk aversion R, decreases in the risky career's log earnings variance s^2, increases in α_w if the ratio is smaller than 2, and decreases in α_w if the ratio is bigger than 2.

3.2.3. Marriage-market matching

Women pay an extra cost along the dimension of reproductive capital when they choose a risky career. To be equally competitive in the marriage market as those in a safe career, these women need to earn a higher realized income to compensate for the extra cost. Assuming the return to a risky career is mean-preserving—that is, the average realized income from a risky career is the same as that from a safe career—the average woman in a risky career will match with lower-ranked men, i.e., those with lower income.

Proposition 4. *Women who choose a risky career and realize income y_w marry men with a lower income than those women who choose a safe career that yields the same realized income y_w.*

This prediction is from Lemma 3: A safe-career woman with income y_w and a risky-career woman with income $y_w + k$ would marry men of the same income level. In contrast, men's spousal income does not differ by occupation. Although risky career delays marriage for men, they are minimally affected by fertility decline.

3.2.4. Gender differences in marriage timing

In the basic model, career and marital choices are connected: Risky career investors marry late, and safe career investors marry early. We separate career and marital choices in this section. There are four possible choices as a result: (1) choosing the safe career and marrying early, (2) choosing the safe career and marrying late, (3) choosing the risky career and marrying early with unresolved uncertainty in incomes, and (4) choosing the risky career and marrying late with resolved uncertainty in incomes (see Fig. 5).

Suppose men and women can choose when to marry. Namely, an ability-x_m man who chooses the safe career can choose to marry in the second period as a marriage type $z_m = x_m$, and an ability-x_w woman who chooses the safe career can choose to marry in the second period as a marriage type $z_w = x_w - k$. An ability-x_m man can choose the risky career and marry in the first period as a man who may realize a marriage type $z_m = x_m - c + \varepsilon_m$ where $\varepsilon_m \sim N(t_m, s_m^2/2)$. If he marries a marriage type z_w woman and realizes an income $x_m - c + \varepsilon_m$, they generate a marriage surplus $s(x_m - c + \varepsilon_m, z_w)$. Similarly, an ability-$x_w$ woman can choose the risky career and marry in the first period as a woman who may realize a marriage type $z_w = x_w - c + \varepsilon_w$. We show that even if marriage timing is separated from the career choice, anyone who chooses the safe career tends to marry in the first period and anyone who chooses the risky career tends to marry in the second period, as we have assumed throughout the paper.

Lemma 5. *Career choice and marriage age are related: (1) Anyone who chooses the safe career marries in the first period, and (2) anyone who chooses the risky career marries in the second period.*

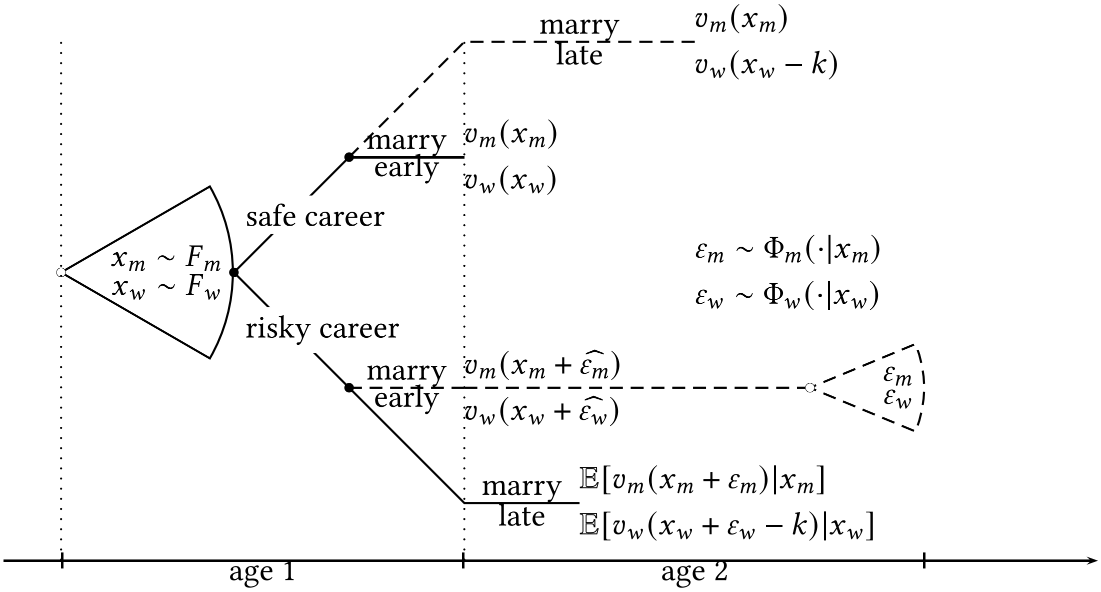


Fig. 5. An individual's career and marriage decision tree in the parametric extension.

To understand the result, realize that a person who has the incentive to choose a risky career only has the incentive to do so if they can marry to a different partner depending on the realization of their income from the risky career. For an ability-x_m man, the four choices respectively yield: (1) $v_m(x_m)$; (2) $v_m(x_m)$; (3) $v_m(x_m + \widehat{\varepsilon_m}|x_m)$; and (4) $\mathbb{E}[v_m(x_m + \varepsilon_m)|x_m]$, where $v_m(x_m + \widehat{\varepsilon_m}|x_m)$ represents the marriage payoff of an ability-x_m man who chooses the risky career. There is no advantage in choosing the safe career and marrying late over choosing the safe career and marrying early since delaying is always associated with some costs and/or discounting; when a man chooses the safe career, he might as well choose to marry early.

Proposition 5. *Women marry earlier than men on average.*

Proposition 5 predicts that men marry later than women on average. In the US, men have always married later than women on average. The same pattern holds around the world: Men have married later on average than women in every country in the world in recent decades (United Nations, 1990; Bergstrom and Bagnoli, 1993; United Nations, 2000).

Marrying at an older age means less time to have children. Since women have a shorter span of fecundity, those who marry later are more likely to be constrained by the biological clock. We thus have the following prediction from the model:

Corollary 1. *Women who choose a risky career are less likely to have children. This correlation is weaker for men.*

4. Empirical evidence

Propositions 1 to 5 and Corollary 1 summarize the model's testable predictions on gender differences in career choices, income distribution, marital and fertility outcomes, and matching patterns. While there are other explanations for why women are underrepresented in risky occupations, our model is unique in connecting labor and marriage outcomes based on a single gender difference, namely, the difference in reproductive length and consequent career-family trade-off. It is a theoretically intuitive and empirically well-documented gender difference. Explanations that are based on gender differences in preferences, beliefs, and attitudes (e.g., Altonji and Blank, 1999; Eckel and Grossman, 2002; Dreber and Hoffman, 2007; Niederle and Vesterlund, 2007; Kleinjans, 2009; Barbulescu and Bidwell, 2012; Koellinger et al., 2013; Buser et al., 2014; Gill and Prowse, 2014; Wozniak et al., 2014), whether they stem from hardwired genetic preferences or from social norms, usually do not have implications on marriage-market outcomes.

A strand of literature links the labor market and the marriage market. Bursztyn et al. (2017) find that women are less likely to show signs that may imply their ambitions in the labor market in front of men. They attribute this "acting wife" behavior to social norms—men do not like ambitious women as wives. Our model can be thought of as "microfounding" the source of such social norm: Women who are ambitious in pursuing a career are more likely to delay marriage and less likely to have children (an important public good in marriage). Corinne Low (2021) shows that women with postgraduate degrees marry men with lower incomes than those with only a college degree, suggesting that women face a trade-off between human capital and reproductive capital. We show that this trade-off also affects gender differences in occupational choices.

Another strand of literature documents labor-market responses to the competitive marriage market. Higher sex ratios (more men than women) induce higher earnings for men (Angrist, 2002) and more premarital investments (Lafortune, 2013). Potential spouses' earnings and fertility also appear to affect college students' major choices (Wiswall and Zafar, 2021). Those papers typically focus on the *level* of income instead of the *variation* and usually do not have implications for the gender differences in the timing of marriage and fertility outcomes.

Couples—after they marry—may have incentives to diversify the labor-market risk they face by choosing careers that exhibit different levels of income uncertainty (it is related to the literature on the "added worker effect", see for example, Lundberg, 1985; Stephens, 2002; Wang, 2019). If the husband is in a career that exhibits volatile income, it is beneficial for the wife to work in a job that can bring a steady stream of income. This risk-sharing explanation provides no prediction for when people choose to marry and how many children they have. It also implies *post*-marriage sorting into occupations with different levels of riskiness, while our model focuses on *pre*-marriage choices.

This section presents empirical evidence of marital and occupational choices that is consistent with our predictions. The presented evidence should be interpreted as correlations instead of causality, as we do not intend to rule out alternative explanations. Formal tests of model predictions require exogenous shocks to women's fecundity and their fertility decisions, which are beyond the scope of this paper.[7] Nonetheless, we strive to account for confounding factors that may affect the interpretation of the empirical patterns. We control for occupational characteristics that may be correlated with earnings variation and rich individual characteristics, including not only basic demographic characteristics but also aptitudes, preferences, and expectations. Put together, these pieces of evidence suggest that the mechanisms implied by our model are viable explanations of salient features of gender differences in occupational choices and marital outcomes, which are difficult to be reconciled by any single alternative explanation.

4.1. Main data and sample

We use the 5-year American Community Survey (ACS) ending in 2016 (Ruggles et al., 2017) for most of the empirical tests. The dataset is a 5% random sample of the US population. It has detailed information on respondent's occupation, work and income, education, marital outcomes, and other demographics. Although it is a 5-year sample, it should be seen as providing a snapshot of the US population.

We start by focusing on the subsample of people who have a college degree. With at least four more years of schooling than high-school graduates, the college-educated face a more acute trade-off between family and career in a shortened time span, so the mechanisms highlighted in our model are likely more relevant, as is evident from Fig. 1. In addition, college-educated workers have a larger variation in their income than less-educated workers. This is partly because college-educated workers possess a wider range of skills, which is due to different career choices earlier in life, either at the time when deciding the focus of study in college or in early career.

We further restrict the sample to individuals between ages 25 and 64 who are not currently enrolled in school or live in quarter groups. We start the sample at age 25 to allow people to finish college and settle into a stable job. Because our measure of career riskiness is based on occupation, the sample is further restricted to those who report a valid occupation code. In the ACS, this excludes people who have not worked in the past 5 years. In various cases, we need the average and the standard deviation of the wage of a certain demographic group. To calculate the wage rate, we use individuals who have positive earnings in the past year and worked at least part-time (more than 20 h in a typical week and more than 30 weeks in the past year). Hourly wage is calculated as the annual earnings in the previous year divided by the product of weeks worked last year and the usual hours worked per week.

4.2. Measuring occupational riskiness

Measuring the riskiness of a career is key to our empirical tests. Through the lens of the model, we define career riskiness as the cross-sectional variation in the realized earnings among people who have chosen the career. With the ACS data, we measure career riskiness using the within-occupation standard deviation of residual log wage among workers between ages 40 and 49. We measure one's "permanent wage" by their wage rate in their 40 s, when the uncertainty about how much one makes in a given occupation is likely resolved. The wage dispersion captures how much risk in earning potential a given occupation poses. Our results are robust to the choice of specific age groups to calculate wage dispersion. We use the hourly wage instead of annual earnings to abstract away from endogenous labor supply decisions.

We adjust the wage rate based on observable individual characteristics, so the wage dispersion we use is not affected by different demographic compositions across different occupations. Specifically, we first run the following regression:

$$\ln w_i = \mathbf{X}_i \cdot \beta + \varepsilon_i, \tag{9}$$

where $\mathbf{X}_i$ includes age and age squared, an indicator for gender, indicators for whether the person is White, Black, Asian or of other races, and full interactions among those three sets of variables. Appendix Table B.3 shows that our conclusions are robust

[7] Recent studies have used regulations on in vitro fertilization (IVF) to study the role of lengthened fecundity on women's education and career choices (e.g., Kroeger and La Mattina, 2017; Gershoni and Low, 2021a,b). Kroeger and La Mattina (2017) exploit variation in state mandates on insurance coverage of IVF and find that women in states with IVF mandates are more likely to obtain a professional degree. We tried to use the same variation but did not have enough statistical power for our outcomes of interest.

to additional tweaks to Eq. (9). Residual from the regression, $\hat{\epsilon}_i$, captures the log point deviation from the average adjusted wage. Occupational riskiness is measured as the standard deviation of the residual log wage:

$$OccRisk_o = \sqrt{\frac{1}{N_o}\sum_{i\in o}(\hat{\epsilon}_i - \bar{\hat{\epsilon}}_o)^2}, \tag{10}$$

where $\bar{\hat{\epsilon}}_o$ is the average residual log wage of the occupation, and N_o is the total number of workers in occupation o. Because $\hat{\epsilon}_i$ takes the log point form, $OccRisk_o$ is comparable across occupations with different average wage levels. Both the regression and the construction of riskiness are weighted by the ACS person weight.[8] In the full sample of occupations, we call an occupation "risky" if its $OccRisk_o$ is above the median.[9]

One issue with the definition is the measurement error induced by occupation classification. We use 3-digit occupations defined in the ACS. However, it is inevitable that some occupations are cruder and include a larger set of heterogeneous professions. For example, one may argue that "physicians" include a large set of medical professionals whose differentiated skills are valued differently in the labor market; in contrast, "primary school teachers" are more homogeneous. Hence, it may not be surprising that there is a larger wage inequality among physicians than among primary school teachers. Our model relies on the assumption that choosing an occupation bears the uncertainty of its future wage. However, while we as econometricians do not observe how a young medical student chooses to specialize in pediatrics or in anesthesiology,[10] it is arguably observable to their potential partners in the marriage market. It is difficult to gauge measurement errors in the occupation classification, but we are not aware of any evidence that systematically broader sets of skills are present in male-dominant occupations.

A risky career in the model does not only mean that the *realized* return is more dispersed, but it also requires *time* for the uncertainty to unravel. Intuitively, if the return to risky occupation is realized instantly, women's shorter span of fecundity should not impose a disadvantage. Our measure of occupational riskiness only looks at wage dispersion in midlife, but if the risk has unraveled early on, potential partners can make marriage-market decisions based on one's realized income, and there is no need to wait until the next period. Therefore, wage dispersion needs to be both large and increasing for an occupation to be regarded as risky in our setting.

Risky occupations, defined by large wage dispersion for workers of ages 40 to 49, also have larger wage dispersion for younger workers. Wage dispersion in both risky and less risky occupations grows over time, but the growth is much higher in risky occupations. As is shown in Fig. 6, the average standard deviation of residual log wage in less risky occupations increases from 0.45 for workers between ages 25 and 29 to 0.49 for workers between ages 45 and 49 (an increase of 9%); among risky occupations, average standard deviation increases from 0.53 to 0.7 (an increase of 32%). In fact, the correlation coefficient between log wage dispersion among workers between 40 and 49 and the *growth* in log wage dispersion is about 0.7. About 80% of risky occupations by our definition (i.e., above median riskiness) also have an above-median growth in log wage dispersion. Our empirical results are also robust to alternative definitions of risky occupations that take into consideration both midlife wage dispersion and growth in wage dispersion.

4.3. Other occupational characteristics

Residual wage variation can be correlated with other occupational characteristics. It is possible that women are underrepresented in some occupations not because these occupations exhibit large wage uncertainty but because they have certain disamenities that are costly for women.

While it is infeasible to measure all possible occupational characteristics, we construct proxy variables for a few that have been documented to be relevant for women. In the empirical analyses below, we control for proxies for these occupational amenities and show that residual wage variation remains to have power in explaining gender differences in occupational choice.

First, it is worth pointing out that risky occupations do not necessarily pay better or worse *on average*. The average residual log wage, $\bar{\hat{\epsilon}}_o$, and riskiness $OccRisk_o$ are only weakly correlated, with a correlation coefficient of around 0.1. We control for the average residual log wage of occupation in all empirical analyses.

Women are often the tied mover in family migration decisions (Mincer, 1978). Expecting possible future moves, it is possible for women to sort into geographically flexible jobs so that they can find similar jobs no matter where they move (Benson, 2014). Teachers and physicians are needed everywhere, while the demand for nuclear physicists is more geographically concentrated. And we definitely see women making up a larger fraction of teachers and doctors than of nuclear physicists. We can measure the geographical flexibility of an occupation using the dissimilarity index:

$$D_o = \frac{1}{2}\sum_c \left|\frac{E_{oc}}{E_o} - \frac{E_c}{E}\right|,$$

[8] Appendix Table B.1 lists the most and the least risky occupations among the 50 largest occupations of college-educated workers (there are 328 3-digit occupations classified). Men are more likely to be in more risky occupations, and women are more likely to be in less risky ones.

[9] Results are robust to alternative definitions of risky occupations. Risky occupations are alternatively defined as those with a standard deviation of residual log wage above the 75th or 90th percentiles. We also directly use the standard deviation of residual log wage as a continuous measure of occupational riskiness. Appendix Tables B.4 and B.5 report the robustness results on key outcomes.

[10] According to the Occupational Outlook Handbook published by the Bureau of Labor Statistics, general pediatricians have an average annual wage of $184,410 and anesthesiologists on average make $261,730 a year in 2021.

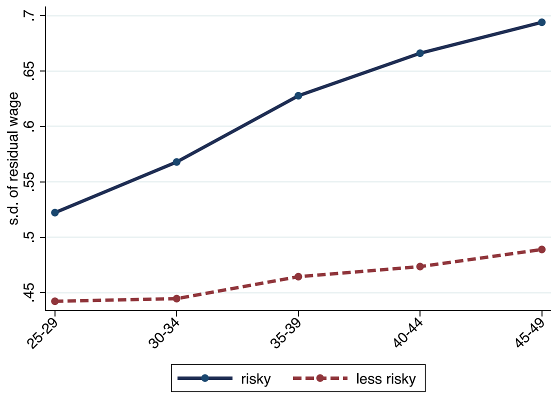


Fig. 6. Growth in Wage Dispersion by Occupational Riskiness. Note: We use the 5-year ACS ending in 2016. The sample includes workers with a college degree, report an occupation, worked at least part-time in a typical week in the previous year, and are between ages 25 and 49. Log hourly wage is first residualized from demographic characteristics. The standard deviation of residual log hourly wage is calculated for each occupation by age group. Risky and less risky occupations are categorized based on the standard deviation of residual log hourly wage for the age 45–49 group. The graph shows that while residual wage dispersion grows over time for both risky occupations and less risky occupations, the growth is larger among risky occupations.

Table 1
Correlations among occupation characteristics.

	(1) Share of female workers	(2) =1 if risky occupation
Spatial dissimilarity index	-0.49	-0.20
Long-hour premium	-0.06	0.01
“Mom-friendliness index”	0.27	-0.02

Note: Each observation is a 3-digit occupation. The data is from the 5-year ACS ending in 2016. All occupational characteristics are defined for college graduates. The spatial dissimilarity index is higher if the spatial distribution of occupation is uneven. An occupation’s long-hour premium is higher if the hourly wage is higher for those who work longer hours. The “mom-friendliness index” is a revealed preference measure on how accommodating the occupation is for mothers with young children, with higher values indicating better accommodation. See the text for detailed definitions of occupation characteristics.

where E_{oc} is the number of workers in occupation o in local labor market c (commuting zones are used to proxy for local labor markets); E_o is the total number of workers in occupation o; E_c is the total number of workers in labor market c; E is the aggregate number of workers. If occupation o is evenly distributed across space, D_o is 0. A smaller D_o indicates the occupation is more evenly distributed spatially.

Recent studies show that job flexibility is important for women’s careers (Goldin, 2014; Goldin and Katz, 2016). While job flexibility is a broad term that includes many job arrangements, one important aspect is long working hours. We measure the premium of long working hours as a proxy for job inflexibility. If a job rewards long working hours, it is likely unfavorable to women as they, shouldering a larger portion of housework, are less able to work long hours. We define long hours as working more than 50 h per week. The long-hour premium for each occupation is defined as the log wage premium of those who work long hours compared with those who work full time, controlling for other demographic characteristics.

An important factor for a woman’s career and labor force participation is the presence of young children at home. We construct a measure of an occupation’s friendliness toward female workers with young children that is based on revealed preference. If an occupation is good for women with young children, we will see a high share of women with young children relative to the share of men with young children. The measure is

$$\ln\left(\frac{E_{o,w}^{no\ kid}/E_{w}^{no\ kid}}{E_{o,w}^{kid}/E_{w}^{kid}}\right)-\ln\left(\frac{E_{o,m}^{no\ kid}/E_{m}^{no\ kid}}{E_{o,m}^{kid}/E_{m}^{kid}}\right),$$

where $E_{o,s}^{no\ kid}$, $s\in\{men, women\}$, is the employment of sex s with no young children in occupation o, while $E_{o,s}^{kid}$ is the employment of those with children. $E_{s}^{no\ kid}$ is the total employment of sex s with no children. We construct this index using working men and women between 25 and 44 years old. A young child is defined as a child under 6 years old. We call this index the “mom-friendliness index”.

Column 1 of Table 1 shows the correlation coefficients between occupational characteristics and the share of female workers in the occupation. Spatial concentration (larger spatial dissimilarity index) is negatively correlated with the female share, so is long-hour premium. Our reveal-preference-based mom-friendliness index is positively correlated with the female share. Column 2 shows the correlations between our measure of occupational riskiness and other occupation characteristics. Risky occupations tend to be less spatially concentrated, less mom-friendly, but more likely to have a long-hour premium. The correlations in Column 2 are all very small. This suggests that our occupational riskiness measure is unlikely to be confounded by other occupational features.

Table 2
Wage inequality and occupational choice.

	(1)	(2)	(3)	(4)
	Share in risky occupations ACS 2016		Δshare in risky occupations Census 2000, ACS 2016	
	male	female	male	female
Opposite-sex log res. wage dispersion	0.432**	0.308**		
	(0.138)	(0.105)		
Own-sex log res. wage dispersion	0.512**	0.293*		
	(0.157)	(0.158)		
Δ opposite sex log res. wage dispersion			0.224**	0.053
			(0.099)	(0.099)
Δ own sex log res. wage dispersion			0.028	0.163*
			(0.112)	(0.090)
Opposite-sex mean log res. wage	X	X		
Own-sex mean log res. wage	X	X		
Δ opposite sex mean log res. wage			X	X
Δ own sex mean log res. wage			X	X
N of MSAs	50	50	50	50

Note: Data include the 5-year ACS ending in 2016 and the 5% sample of the 2000 population census. The sample focuses on those with a college degree and focuses on the 50 largest metropolitan statistical areas (MSAs). The dependent variable, the share (or the change in share) of workers in risky occupations, is constructed from individuals between 25 and 34 years old. The main explanatory variables, (changes in) opposite- and own-sex wage dispersion, are constructed from those between 40 and 49 years old. Robust standard errors are in parentheses. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

4.4. *Empirical patterns consistent with model predictions*

4.4.1. *Marriage-market incentives and risky career choices*

Proposition 1 states that the competitive nature of the marriage market encourages both men and women to engage in risk-taking behavior. The more competitive the marriage market is, the more likely men and women are to choose risky careers. Consider a woman, if the variation in wage among men in the marriage market is larger, given that the marriage surplus is linear in the husband's income, the woman has a stronger incentive to choose a risky occupation. This is because a linear surplus leads to a weakly convex payoff (Lemma 1). The same is true for a man. Similarly, a larger variation in wages among those of the same sex will also lead to stronger incentive to choose a risky occupation.

To test Proposition 1, we exploit variation in the competitiveness across *local* marriage markets. We measure own-sex and opposite-sex wage dispersion among those between ages 40 and 49, and investigate their impacts on choices of risky occupations among those between ages 25 and 34. We choose to measure wage dispersion among an older generation to avoid mechanical correlations: With a larger fraction of people choosing risky occupations, wage dispersion will naturally be greater. We assume that the inequality among the older generation is the relevant information for the younger generation: Earning potential has not been realized for their contemporaries, so the younger generation evaluates the conditions of local marriage market by observing the realized wage inequality among the older generation in the same marriage market.

We use Metropolitan Statistical Areas (MSAs) as local marriage markets and focus on the 50 largest MSAs.[11] Columns 1 and 2 of Table 2 show that both own- and opposite-sex wage dispersion among workers aged between 40 and 49 are positively associated with the fraction of workers aged between 25 and 34 in risky occupations. In addition, men's occupational choice seems to be more sensitive to marriage market competitiveness.

One may be concerned that some unobserved features of local labor markets drive this correlation. Columns 3 and 4 present the correlation between *changes* in wage dispersion and *changes* in the share of risky occupational choices between 2000 and 2016. Time-invariant unobservable features of an MSA are eliminated in this log-differenced specification. The results still hold: MSAs that experienced larger increases in wage dispersion in own and the opposite sexes among the 40–49 year-olds also saw larger shares of the younger generation choosing risky occupations. The coefficients are smaller than those in Columns 1 and 2, rendering some of the coefficients not statistically significant at conventional levels. This is probably due to the fact that log difference increases the relative share of measurement error and exacerbates the attenuation bias.[12]

4.4.2. *Gender difference in risky occupational choices and wage dispersion*

Table 3 shows evidence that is consistent with Proposition 2, which states that men are more likely than women to choose risky careers. Column 1 shows that male workers are 7.6 pp more likely to be in a risky occupation. Columns 2 to 6 show that this pattern holds as we add additional occupational characteristics in the regression. Our preferred specification is Column 6, which shows that

[11] This restriction is largely due to sample size concerns. Outside the top 50 MSAs, there are typically fewer than 100 observations to calculate wage dispersion for a gender-age-group bin among the college graduates.

[12] Admittedly, there are other important caveats before one is willing to take the results in Columns 3 and 4 as causal effects. For example, a sectoral shift in an MSA's economy can contribute to both an increase in wage inequality and a larger share of workers in risky occupations.

Table 3
Risky occupational choice by sex.

Dep var: = 1 if in a risky occ.	(1)	(2)	(3)	(4)	(5)	(6)	(7)
							single
= 1 if male	0.076***	0.074***	0.114***	0.036***	0.048***	0.042***	0.039***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Occupational characteristics							
Avg. log residual wage		0.033***	-0.028***	-0.024***	-0.013***	-0.081***	-0.164***
		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Spatial dissimilarity index			-1.707***			-0.882***	-0.843***
			(0.009)			(0.010)	(0.017)
Long-hour premium				2.260***		2.087***	1.917***
				(0.007)		(0.007)	(0.012)
"Mom-friendly index"					-0.274***	-0.181***	-0.150***
					(0.003)	(0.003)	(0.005)
Individual chars.	X	X	X	X	X	X	X
N of obs.	929923	929913	929913	929913	929913	929913	338435

Note: The data is 5-year ACS ending in 2016. Each observation is a person who has a college degree, is between 25 and 39 years old, and reports an occupation. Individual characteristics include age, a set of indicators for races, and their interactions. 59% of men in the sample are in risky occupations. Robust standard errors are in parentheses. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

male workers are 4.2 pp more likely than female workers to be in a risky occupation. The pattern is similar for those who are never married (Column 7), for whom marriage-market incentives remain relevant.

The first part of Proposition 3 predicts that because men are more likely in risky occupations than women, wage inequality is greater among men. Panel A of Fig. 7 shows indeed, log residual wage dispersion is larger among men than among women at any given age, regardless of marital status. The figures also show that wage inequality for both men and women increases with age. According to the classical human capital theory, this is because of different choices in human capital accumulation: those who receive more on-the-job training have a steeper age-wage profile (Ben-Porath, 1967). Although not directly contradictory to this classical theory, our model indicates an alternative explanation: choices of riskier careers lead to larger inequality in the future. Indeed, not only men's wage inequality is larger than that of women's at any given age, the gender gap in wage inequality increases by age. This is consistent with the fact that a larger fraction of men choose risky careers whose payoffs only realize years later.

4.4.3. Occupational choices and matching patterns

Proposition 4 provides a testable prediction on marriage matching patterns. Because women pay an extra cost if they choose a risky career, in order to be equally competitive in the marriage market as women who choose a safe career, they need to earn a higher realized income to compensate for the extra cost. Assuming the return to a risky career is mean-preserving—that is, the average realized income from a risky career is the same as that from a safe career—and if marriage matching is assortative, it follows that women who choose a risky career *on average* will match with lower-ranked men, i.e., men with lower income.

The results shown in Table 4, Column 1 indicate that, among college-educated women who are currently married, there is indeed a statistically significant negative correlation between choosing a risky occupation and their husband's earnings. After controlling for individual and other occupational characteristics, it can be seen that women between the ages of 45 and 49 who are in risky occupations have husbands whose income is approximately 15% lower than those of women who choose a safe career. These findings are consistent with those of Low (2021), who found that highly educated women may have lower spousal income compared to those who only have college degrees.

4.4.4. Gender difference in marriage and fertility

Lemma 5 predicts that those who choose risky occupations marry later. Panel B of Fig. 7 plots the share of never married by age, separately for men and women, and confirms this prediction. Proposition 5 further predicts that due to a shorter reproductive span, women tend to marry earlier than men. This is a phenomenon widely observed across different cultures and over time. Panel C of Fig. 7 shows that this pattern also holds for the US data. Columns 2 and 3 of Table 4 show that by age 50, women in a risky occupation are more likely to be never married, but not among men in those occupations.

As a corollary of Proposition 5, the choice of risky occupations delays marriage, which affects the probability of having children. For men, because their fecundity span is longer, marrying at an older age has a smaller effect on the probability of having children. In contrast, the time window for women to finish college, succeed in a risky occupation, get married, and having children is tighter, given their shorter fertility period, so we would expect that women who choose risky occupations are less likely to have children. Columns 4 and 5 of Table 4 support this proposition.

4.5. Evidence from NLSY79

The ACS has a large sample and rich demographic information, but it is a cross-sectional dataset. Many of our predictions are sequential in nature—decisions today have consequences for outcomes later in life. Correlating the current occupation with current labor-market and marriage-market outcomes, as we have been doing so far, makes an implicit assumption that people do not switch

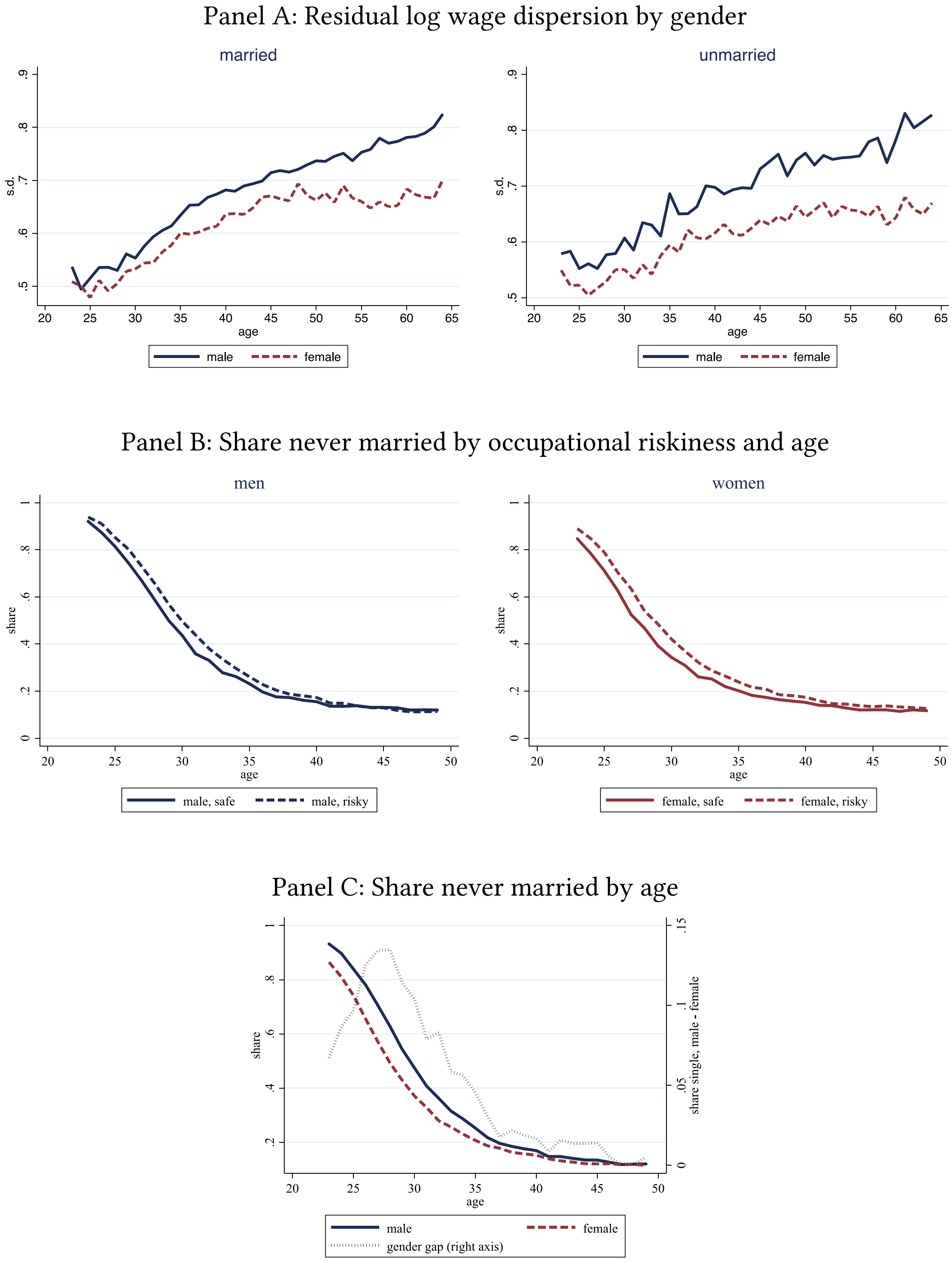


Fig. 7. Wage Dispersion, Marital Outcomes, and Risky Occupational Choices. Note: Data is from the 5-year ACS ending in 2016. Panel A includes men and women between 25 and 64 who have a college degree, are not currently enrolled in school, and were employed in the previous year, and worked at least part-time. Panels B and C include men and women between 25 and 49 who have a college degree. Panel A shows that wage inequality is larger among men than women. Panel B shows that risky occupation is correlated with later marriages for both men and women. Panel C shows that overall, women marry earlier than men.

Table 4
Risky occupational choices and marital outcomes.

	(1)	(2)	(3)	(4)	(5)
dep var:		= 1 if		= 1 if	
	invhsin(husband earnings)	never married		have children	
sample:	women	women	men	women	men
= 1 if in risky occ.	-0.155***	0.008**	0.003	-0.013**	0.005
	(0.038)	(0.003)	(0.003)	(0.004)	(0.004)
individual chars	X	X	X	X	X
occupational chars	X	X	X	X	X
husband age FE	X				
N of obs.	109,082	149,764	143,789	159,484	150,616
mean dep. var.	10.3	0.11	0.11	0.69	0.71

Note: The data is from the 5-year ACS ending in 2016. The sample includes men and women who have a college degree and are between 45 and 49 years old. Individual characteristics include age and age squared, race dummies, and interactive terms. Occupational characteristics include mean residual log wage, the spatial dissimilarity index, long-hour wage premium, and an index measuring the occupation's overall friendliness toward women with young children (the "mom-friendliness index"). Column 1 includes currently married women with spouse present and identifiable. The dependent variable is the inverse hyperbolic sine of the husband's earnings in the previous year (those with zero earnings are included in the sample). The dependent variable in Columns 2 and 3 is a binary variable for whether the person is never married. The dependent variable in Columns 4 and 5 is a binary variable for whether there are children present in the household (biological or adopted). In Columns 2 and 4, the sample includes women; in Columns 3 and 5, the sample includes men. Robust standard errors are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 5
Early-career risky occupational choice, marital and fertility outcomes: NLSY79 sample.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
dep var	=1 if in risky occ by 30	=1 if ever married by 40		# of children in hhd. by 40		biological children #	had any
sample	all	women	men	women	men	women	women
= 1 if male	0.122***						
	(0.020)						
= 1 if in risky occ by 30		-0.034	-0.024	-0.201**	-0.054	-0.165*	-0.079**
		(0.029)	(0.035)	(0.093)	(0.110)	(0.099)	(0.032)
ln income by 30	-0.016**	-0.009**	0.023**	-0.097***	0.085***	-0.084***	-0.015***
	(0.005)	(0.004)	(0.009)	(0.015)	(0.021)	(0.018)	(0.004)
year of birth FE	X	X	X	X	X	X	X
occ. chars.	X	X	X	X	X	X	X
ind. demo.	X	X	X	X	X	X	X
age last surveyed						X	X
# of ind.	3021	1265	1027	1152	940	1190	1190
mean dep. var.	0.58	0.76	0.77	1.56	1.34	1.70	0.76

Note: The sample includes men and women in NLSY79 with a college degree. The dependent variable and the sample in each column is stated in the header. All regressions include a set of year of birth fixed effects, individual demographic characteristics (Black, Hispanic, or non-Black and non-Hispanic), a set of other occupational characteristics, and are weighted by sample weights. Other occupational characteristics include mean residual log wage, the spatial dissimilarity index, long-hour wage premium, and an index measuring the occupation's overall friendliness toward women with young children (the "mom-friendliness index"). The sample in Columns 6 and 7 include female respondents in their last observed survey, conditional on the last survey being after 2000. A set of age fixed effects at the time the last survey was conducted are also controlled for. Robust standard errors in parentheses. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

occupations. To address this concern, we show the robustness of our results using the National Longitudinal Survey of Youth 1979 (NLSY79), taking advantage of its panel structure.

The NLSY79 surveys a sample of individuals born between 1957 and 1964. The survey started in 1979 and followed these individuals every year until 1994 and every other year afterward. The survey asks a wide range of questions, tailored in each wave as the respondents grew older. One drawback of the NLSY79 is its relatively small sample size. It started with 12,000 respondents, among those, about 3,500 went on to obtain a college degree. By 2000, about 60% of the respondents were still in the sample. Partly due to the much smaller sample size compared with the ACS, our estimates are less precise in some specifications.

We measure risky occupational choice by whether the person is in a risky occupation by age 30 and investigate the person's marriage and fertility outcomes later in life.[13] We always control for log earnings at age 30 (plus one to include zeros). We restrict the sample in the same way as the 2016 ACS sample.

The first column of Table 5 shows a gender gap in early-career risky occupational choice: Men are 12 pp more likely to be in a risky occupation by age 30, according to the NLSY79 sample.

[13] NLSY79 uses different occupational codes across different waves, so we map occupational codes used in the NSLY79 into IPUMS occupational codes (occ1990) following Deming (2017). We then use the occupational riskiness defined using data from the 2016 ACS. Ideally, we would like to measure riskiness by future wage dispersion given early-career occupational choice. But the sample size in NLSY79 is too small for that.

The next two columns show choosing a risky occupation in early career delays marriage, but more so for women. Women in risky occupations in early career are 3.4 pp less likely to be married by age 40, while men are 2.4 pp less likely, although the coefficients are not statistically significant. In all regressions, we control for birth year fixed effects and a set of individual demographic characteristics as well as additional occupation characteristics.

The remainder of Table 5 shows the correlation between risky career choices in early career and fertility outcomes later. Columns 4 and 5 show that choosing a risky occupation in early career is associated with fewer children in the household (both biological or non-biological) by age 40. But the effect is much larger for women (by 0.2 or about 13% of the sample mean) than for men (0.06 or about 4% of the sample mean). The effect on men is also not statistically significant.

One advantage of the NLSY79 is that in waves after 2000 (when respondents were between 36 and 43 years old in 2000), it asked female respondents how many biological children they had ever had. Columns 6 and 7 investigate the correlation between risky occupational choice and the number and probability of having biological children. To maximize the sample size, we use the information on biological children from the last year a woman was surveyed. We control for the age when the last interview was conducted to account for the fact that some women were older when they were last interviewed and are likely to report more children than those whose last interview was at an early age. Choosing a risky occupation in early career is associated with a decline of 0.165 (10% of the sample mean) biological children and a decline in the probability of having a biological child by 7.9 pp (10% from the mean). These are significant effects, both statistically and economically.[14,15]

5. Conclusion

We embed the discussion of gender-differential risky career choices in a general-equilibrium marriage-market framework. A set of outcomes is simultaneously and endogenously determined: career choices, marriage timing, income distributions, matching patterns, and the division of marriage payoffs. First, we find that the competitive marriage market inherently induces both men and women to choose risky careers. Second, we find women's shorter reproductive span, which imposes a gender-asymmetric career cost, to be a potential unifying factor to explain gender differences in premarital career choices, income inequality, marriage age, and fertility decisions.

We use both cross-sectional and longitudinal data from the United States to document patterns consistent with model predictions. We use these empirical findings to show our model's capabilities of explaining a wide range of stylized facts with regard to labor and marriage markets. Most of the patterns documented in the paper should be interpreted as correlations instead of causalities, and we do not intend to rule out alternative explanations. An interesting extension for future work is to unpack such a causal relationship.

We find that women are underrepresented in careers that exhibit substantial uncertainty in earnings and career progression. Rewarding careers such as those in medicine, law, management, and academia feature long periods of apprenticeship, which demand long hours and deep concentration on work. These careers also require many years of education. Together, they impose a high barrier for women to enter, as these careers have a larger toll on families for women than for men.

Changing cultures in the past few decades, such as delaying or forgoing marriage and desiring fewer children, have liberated women to pursue risky, challenging, yet rewarding careers. Medical progress that extends women's reproductive health is promised to help women further to make their careers without sacrificing marriage and family. Government investment and subsidies in such medical treatments have the potential to narrow the gender gap in labor market outcomes. For some occupations, restructuring the work environment and evaluation criteria could potentially shorten the period of uncertainty, which could also encourage more women to choose those careers.

Data availability

Data will be made available on request.

Appendix A. Omitted details and proofs

A.1. A microfoundation of the marriage surplus function

Consider a man with realized income y_m and a woman with realized income y_w. A single person's utility depends on the consumption of a public good and a private good: $u_m(Q, q_m) = q_m Q$ and $u_w(Q, q_w) = q_w Q$. They divide their incomes between the two goods to maximize their respective utility. The utility when they live alone are their reservation utility $z_m(y_m) = \max_Q (y_m - Q)Q = (y_m/2)^2 = {y_m}^2/4$ and $z_w(y_w) = \max_Q (y_w - Q)Q = {y_w}^2/4$, respectively. The maximal total utility when the two marry (or live together)

[14] The longitudinal nature of the NLSY79 also provided us with the opportunity to rule out several other confounding factors. Appendix B.4 summarizes these tests.

[15] Given the panel structure of NLSY79, it is tempting to test whether those who choose a risky career indeed marry after the income uncertainty is unraveled. Empirically, this would correspond to a large increase in income inequality right before marriage. Such tests, though intuitive, are difficult to conduct in reality due to at least two reasons. First, as Fig. 6 shows, uncertainty unravels gradually rather than abruptly, so it is difficult to empirically detect a structural break. Second, partners may have private information on the potential income of their future spouses that econometricians do not observe.

subject to the constraint $Q + q_m + q_w = y_m + y_w$ is

$$z(y_m, y_w) = \max_{Q,q_m,q_w} q_m Q + q_w Q = \max_{Q}(y_m + y_w - Q)Q = (y_m + y_w)^2/4.$$

The sum of private goods is determinate to be $q_m + q_w = (y_m + y_w)/2$, but the allocation of private goods q_m and q_w is indeterminate. The surplus from the marriage of a couple (y_m, y_w) is hence

$$s(y_m, y_w) = z(y_m, y_w) - z(y_m) - z(y_w) = y_m y_w/2.$$

The marital surplus is perfectly transferable between the two parties: To achieve a marital gain of v_m in combination with the reservation utility, the man consumes a private good of $q_m = [v_m + z(y_m)]/Q = 2[v_m + {y_m}^2/4]/(y_m + y_w)$. Similarly, the woman consumes a private good of $q_w = [v_w + z(y_w)]/Q = 2[v_m + {y_w}^2/4]/(y_m + y_w)$ to achieve a marital gain of v_w.

A.2. Proof of Theorem 1

Consider the following composite map

$$\Gamma : \mathcal{V} \rightrightarrows P_m \times P_w \to \mathcal{G}_m \times \mathcal{G}_w \rightrightarrows \mathcal{V},$$

where $\mathcal{V}$ is the set of stable marriage payoff functions $v_m : X_m \to \mathbb{R}_+$ and $v_w : X_w \to \mathbb{R}_+$, $P_m \times P_w$ is the set of career choice strategies $p_m : X_m \to [0,1]$ and $p_w : X_w \to [0,1]$, and $\mathcal{G}_m \times \mathcal{G}_w$ is the set of income distributions. By Glicksberg's fixed-point theorem, an equilibrium exists if $\mathcal{V}$ is non-empty, convex, and compact, and Γ is non-empty-valued, upper-hemicontinuous, convex-valued, and compact-valued. These properties are satisfied in this setting. See Zhang (2015).

A.3. An example with multiple equilibria

Suppose men and women all have abilities of 2. Each person can choose either a safe career that returns 2, or a risky career that returns 1 with probability 1/2 and 3 with probability 1/2. The surplus function is $s(y_m, y_w) = y_m y_w$. There are two equilibria in this setting. Everyone chooses the safe career in an equilibrium in which $v_m^*(1) = v_w^*(1) = 0.5$, $v_m^*(2) = v_w^*(2) = 2$, and $v_m^*(3) = v_w^*(3) = 4$. Everyone chooses the risky career in another equilibrium in which $v_m^*(2) = v_w^*(2) = 0.5$, $v_m^*(2) = v_w^*(2) = 2$, and $v_m^*(2) = v_w^*(2) = 4.5$.

A.4. Proof of Lemma 3

Fix a z_m. From the stability conditions

$$v_m(z_m) = s(z_m, z_w(z_m)) - v_w(z_w(z_m))$$

$$v_m(z_m) \geq s(z_m, z_w) - v_w(z_w) \quad \forall z_w,$$

we have

$$v_m(z_m) = \max_{z_w} s(z_m, z_w) - v_w(z_w)$$

$$z_w(z_m) \in \arg\max_{z_w} s(z_m, z_w) - v_w(z_w).$$

By the envelope theorem,

$$v_m'(z_m) = \frac{\partial s(z_m, z_w(z_m))}{\partial z_m} + \left(\frac{\partial s(z_m, z_w(z_m))}{\partial z_w} - v_w'(z_w(z_m)) \right) \frac{\partial z_w}{\partial z_m}.$$

By the fact that

$$z_w(z_m) \in \arg\max_{z_w} s(z_m, z_w) - v_w(z_w),$$

We know

$$\frac{\partial s(z_m, z_w(z_m))}{\partial z_w} - v_w'(z_w(z_m)) = 0.$$

Therefore,

$$v_m'(z_m) = \frac{\partial s(z_m, z_w(z_m))}{\partial z_m} = \frac{\partial (\exp(\alpha_m z_m + \alpha_w z_w(z_m)))}{\partial z_m} = \alpha_m \exp(\alpha_m z_m + \alpha_w z_w(z_m)).$$

Then

$$
\begin{aligned}
v_m(z_m) = & \int_{-\infty}^{z_m} \alpha_m \exp\left(\alpha_m \widetilde{z}_m + \alpha_w \left(\frac{\sigma_{z_w}}{\sigma_{z_m}}(\widetilde{z}_m - \mu_{z_m}) + \mu_{z_w}\right)\right) d\widetilde{z}_m \\
= & \frac{\alpha_m}{\alpha_m + \alpha_w \frac{\sigma_{z_w}}{\sigma_{z_m}}} \exp(\alpha_m z_m + \alpha_w z_w(z_m)).
\end{aligned}
$$

Similarly,

$$
v_w(z_w) = \frac{\alpha_w}{\alpha_w + \alpha_m \frac{\sigma_{z_m}}{\sigma_{z_w}}} \exp(\alpha_w z_w + \alpha_m z_m(z_w)).
$$

A.5. Proof of Lemma 4

An ability-x_m man's expected utility gain from the risky career over the safe career is

$$
\begin{aligned}
\Delta_m(x_m) = & \mathbb{E}_{\varepsilon_m}\left[u_m(v_m(x_m - c + \varepsilon_m))\right] - u_m(v_m(x_m)) \\
= & \mathbb{E}_{\varepsilon_m}\left[u_m(v_m(x_m - c + \varepsilon_m)) - u_m(v_m(x_m))\right],
\end{aligned}
$$

where

$$
\begin{aligned}
& u_m(v_m(x_m - c + \varepsilon_m)) - u_m(v_m(x_m)) \\
= & \frac{1}{1-\rho_m}\left(\frac{\alpha_m}{\alpha_m + \alpha_w \frac{\sigma_{z_w}}{\sigma_{z_m}}}\right)^{1-\rho_m} \times \mathbb{E}_{\varepsilon_m}[\exp((1-\rho_m)(\alpha_m(x_m - c + \varepsilon_m) + \\
& \alpha_w z_w(x_m - c + \varepsilon_m))) - \exp((1-\rho_m)(\alpha_m x_m + \alpha_w z_w(x_m)))] \\
= & \left(\frac{\alpha_m}{\alpha_m + \alpha_w \frac{\sigma_{z_w}}{\sigma_{z_m}}}\right)^{1-\rho_m} \frac{\exp[(1-\rho_m)(\alpha_m x_m + \alpha_w z_w(z_m))]}{1-\rho_m} \times \\
& \mathbb{E}_{\varepsilon_m}\left[\exp\left((1-\rho_m)(\alpha_m + \alpha_w \frac{\sigma_{z_w}}{\sigma_{z_m}})(\varepsilon_m - c)\right) - 1\right].
\end{aligned}
$$

Since $\mathbb{E}_{\varepsilon_m}[\exp(\alpha\varepsilon_m)] = \exp(\alpha t_m + \alpha^2 s_m^2/2)$, the expected value of the term in the square brackets becomes

$$
\exp\left((1-\rho_m)\left(\alpha_m + \alpha_w \frac{\sigma_{z_w}}{\sigma_{z_m}}\right)(t_m - c) + \left((1-\rho_m)\left(\alpha_m + \alpha_w \frac{\sigma_{z_w}}{\sigma_{z_m}}\right)\right)^2 \frac{s_m^2}{2}\right) - 1.
$$

Therefore,

$$
\begin{aligned}
& \Delta_m(x_m) \\
= & \left(\frac{\alpha_m}{\alpha_m + \alpha_w \frac{\sigma_{z_w}}{\sigma_{z_m}}}\right)^{1-\rho_m} \frac{\exp[(1-\rho_m)(\alpha_m x_m + \alpha_w z_w(x_m))]}{1-\rho_m} \times \\
& \left(\exp\left((1-\rho_m)\left(\alpha_m + \alpha_w \frac{\sigma_{z_w}}{\sigma_{z_m}}\right)\left(t_m - c + (1-\rho_m)\left(\alpha_m + \alpha_w \frac{\sigma_{z_w}}{\sigma_{z_m}}\right)\frac{s_m^2}{2}\right)\right) - 1\right),
\end{aligned}
$$

and $\Delta_m(x_m)$ has the same sign as

$$
\delta_m \equiv t_m - c + (1-\rho_m)\left(\alpha_m + \alpha_w \frac{\sigma_{z_w}}{\sigma_{z_m}}\right)\frac{s_m^2}{2}.
$$

Similarly, an ability-x_w woman's expected utility gain from the risky career over the safe career, $\Delta_w(x_w) \equiv \mathbb{E}_{\varepsilon_w}[u_w(v_w(x_w - c - k + \varepsilon_w))] - u_w(v_w(x_w))$, is

$$
\begin{aligned}
& \left(\frac{\alpha_w}{\alpha_w + \alpha_m \frac{\sigma_{z_m}}{\sigma_{z_w}}}\right)^{1-\rho_w} \frac{\exp[(1-\rho_w)(\alpha_w x_w + \alpha_m z_m(x_w))]}{1-\rho_w} \times \\
& \left(\exp\left((1-\rho_w)\left(\alpha_w + \alpha_m \frac{\sigma_{z_m}}{\sigma_{z_w}}\right)\left(t_w - c - k + (1-\rho_w)\left(\alpha_w + \alpha_m \frac{\sigma_{z_m}}{\sigma_{z_w}}\right)\frac{s_w^2}{2}\right)\right) - 1\right),
\end{aligned}
$$

and $\Delta_w(x_w)$ has the same sign as

$$\delta_w \equiv t_w - c - k + (1-\rho_w)\left(\alpha_w + \alpha_m \frac{\sigma_{z_m}}{\sigma_{z_w}}\right)\frac{s_w^2}{2}.$$

Without the marriage market. An ability x_m man's expected utility gain from the risky career over the safe career, *without the marriage market,* is

$$\Delta_m(x_m) = \mathbb{E}_{\varepsilon_m}[u_m(s(x_m - c + \varepsilon_m, z_w(x_m))) - v_w(z_w(x_m)) - u_m(v_m(x_m))],$$

where

$$\begin{aligned}
& u_m(s(x_m - c + \varepsilon_m, z_w(x_m))) - v_w(z_w(x_m)) - u_m(v_m(x_m)) \\
=& \frac{[s(x_m - c + \varepsilon_m, z_w(x_m)) - v_w(z_w(x_m))_+]^{1-\rho_m}}{1-\rho_m} - \frac{[v_m(x_m)]^{1-\rho_m}}{1-\rho_m} \\
=& \frac{\left[\exp[\alpha_m(x_m - c + \varepsilon_m) + \alpha_w z_w(x_m)] - \frac{\alpha_w}{\alpha_w + \alpha_m \frac{\sigma_{z_m}}{\sigma_{z_w}}}\exp[\alpha_m x_m + \alpha_w z_w(x_m)]\right]^{1-\rho_m}}{1-\rho_m} \\
& - \frac{\left[\frac{\alpha_m}{\alpha_m + \alpha_w \frac{\sigma_{z_w}}{\sigma_{z_m}}}\exp(\alpha_m x_m + \alpha_w z_w(x_m))\right]^{1-\rho_m}}{1-\rho_m} \\
=& \frac{\left[\left[\exp[\alpha_m(\varepsilon_m - c)] - \frac{\alpha_w}{\alpha_w + \alpha_m \frac{\sigma_{z_m}}{\sigma_{z_w}}}\right]\exp[\alpha_m x_m + \alpha_w z_w(x_m)]\right]^{1-\rho_m}}{1-\rho_m} \\
& - \frac{\left[\frac{\alpha_m}{\alpha_m + \alpha_w \frac{\sigma_{z_w}}{\sigma_{z_m}}}\exp(\alpha_m x_m) + \alpha_w z_w(x_m)\right]^{1-\rho_m}}{1-\rho_m} \\
=& \frac{\exp[(1-\rho_m)(\alpha_m x_m + \alpha_w z_w(x_m))]}{1-\rho_m}\left[(\exp[\alpha_m\varepsilon_m - \alpha_m c] - K)_+^{1-\rho_m} - (1-K)^{1-\rho_m}\right],
\end{aligned}$$

where $K = \frac{\alpha_w \sigma_{z_w}}{\alpha_w\sigma_{z_w} + \alpha_m \sigma_{z_m}}$. It remains to calculate

$$\begin{aligned}
& \mathbb{E}_{\varepsilon_m}\left[(\exp[\alpha_m(\varepsilon_m - c)] - K)^{1-\rho_m} - (1-K)^{1-\rho_m}\right] \\
=& \int [\exp(\alpha_m\varepsilon_m - \alpha_m c)]^{1-\rho_m}\frac{1}{\sqrt{2\pi s_m^2}}\exp\left(-\frac{(\varepsilon_m - t_m)^2}{2s_m^2}\right)d\varepsilon_m \\
=& \mathbb{E}_{\varepsilon_m}[\exp[(1-\rho_m)\alpha_m(\varepsilon_m - c)]] \\
=& \exp\left[(1-\rho_m)\alpha_m(t_m - c) + \frac{(1-\rho_m)^2{\alpha_m}^2 s_m^2}{2}\right] \\
=& \exp\left[(1-\rho_m)\alpha_m\left(t_m - c + (1-\rho_m)\alpha_m\frac{s_m^2}{2}\right)\right].
\end{aligned}$$

Since $\varepsilon_m \sim \mathcal{N}(t_m, s_m^2)$, $\alpha_m\varepsilon_m - \alpha_m c \sim \mathcal{N}(\alpha_m(t_m - c), {\alpha_m}^2 s_m^2)$.

$$z_w(z_m) = \frac{\sigma_{z_m}}{\sigma_{z_w}}(z_m - \mu_{z_m}) + \mu_{z_w}. \quad \square$$

With the marriage market. An ability-x_m man's expected utility gain from the risky career over the safe career, *with the marriage market,* is

$$\begin{aligned}
& \mathbb{E}_{\varepsilon_m}\left[u_m(v_m(x_m - c + \varepsilon_m)) - u_m(v_m(x_m))\right] \\
=& \mathbb{E}_{\varepsilon_m}\left[s(x_m - c + \varepsilon_m, z_w(x_m - c + \varepsilon_m)) - v_w(z_w(x_m - c + \varepsilon_m)) - u_m(v_m(x_m))\right]. \quad \square
\end{aligned}$$

A.6. Proof of Theorem 2

Define

$$\delta_m(p_m, p_w) \equiv t_m - c(p_m, p_w) + (1-\rho_m)\left(\alpha_m + \alpha_w\sqrt{\frac{\sigma_{x_w}^2 + p_w s_w^2}{\sigma_{x_m}^2 + p_m s_m^2}}\right)\frac{s_m^2}{2},$$

and

$$\delta_w(p_m,p_w) \equiv t_w - c(p_m,p_w) - k + (1-\rho_w)\left(\alpha_w + \alpha_m\sqrt{\frac{\sigma^2_{x_m} + p_m s^2_m}{\sigma^2_{x_w} + p_w s^2_w}}\right)\frac{s^2_w}{2}.$$

Economically, $\delta_m(p_m,p_w)$ and $\delta_w(p_m,p_w)$ have the same sign as Δ_m and Δ_w where the proportion p_m of men and proportion p_w of women choose the risky career. Define correspondences

$$\rho_m(p_m,p_w) \equiv \begin{cases} 0 & \text{if } \delta_m(p_m,p_w) < 0 \\ [0,1] & \text{if } \delta_m(p_m,p_w) = 0 \\ 1 & \text{if } \delta_m(p_m,p_w) > 0, \end{cases}$$

and

$$\rho_w(p_m,p_w) \equiv \begin{cases} 0 & \text{if } \delta_w(p_m,p_w) < 0 \\ [0,1] & \text{if } \delta_w(p_m,p_w) = 0 \\ 1 & \text{if } \delta_w(p_m,p_w) > 0. \end{cases}$$

Economically, $\rho_m(p_m,p_w)$ and $\rho_w(p_m,p_w)$ represent the optimal probabilities of men and women who would choose the risky career if proportion ρ_m of men and proportion ρ_w of women choose the risky career. An equilibrium exists if the mapping

$$\rho(p_m,p_w) = (\rho_m(p_m,p_w), \rho_w(p_m,p_w))$$

has a fixed point. Since $\rho : [0,1]^2 \to [0,1]^2$ is upper hemicontinuous, convex-valued, and nonempty, by Kakutani's fixed-point theorem, a fixed point exists.

Furthermore, equilibrium exists uniquely if $1-\rho_m > 0$ and $1-\rho_m > 0$. Note that because $1-\rho_m > 0$, $\delta_m(p_m,p_w)$ decreases in p_m. Fix p_w. Define $p_m(p_w)$ as follows:

$$p_m(p_w) = \begin{cases} 1 & \text{if } \delta_m(0,p_w) > \delta_m(1,p_w) > 0 \\ \text{solution of } \delta_m(p_m,p_w) = 0 & \text{if } \delta_m(0,p_w) > 0 > \delta_m(1,p_w) \\ 0 & \text{if } 0 > \delta_m(0,p_w) > \delta_m(1,p_w). \end{cases}$$

The function $p_m(p_w)$ is continuous and differentiable. We will show that $\delta_w(p_m(p_w),p_w)$ strictly decreases in p_w.

$$\delta_w(p_m(p_w),p_w) = t_w - c(p_m(p_w),p_w) - k + \frac{s^2_w}{2}(1-\rho_w)\left(\alpha_w + \alpha_m\sqrt{\frac{\sigma^2_{x_m} + p_m(p_w)s^2_m}{\sigma^2_{x_w} + p_w s^2_w}}\right).$$

First note that

$$\begin{aligned}\frac{d\sqrt{\frac{\sigma^2_{x_m}+p_m(p_w)s^2_m}{\sigma^2_{x_w}+p_w s^2_w}}}{dp_w} &= \sqrt{\frac{\sigma^2_{x_m} + p_m(p_w)s^2_m}{\sigma^2_{x_w} + p_w s^2_w}}\left(\frac{s^2_m/2}{\sigma^2_{x_m} + p_m(p_w)s^2_m}p_m{}'(p_w) - \frac{s^2_w/2}{\sigma^2_{x_w} + p_w s^2_w}\right) \\ &\equiv r_m(A_m p_m{}'(p_w) - A_w).\end{aligned}$$

Denote $\frac{s^2_w}{2}(1-\rho_w)$ by B_w. Hence,

$$\delta_w(p_m(p_w),p_w) = t_w - c(p_m(p_w),p_w) - k + B_w(\alpha_w + \alpha_m r_m),$$

and

$$\begin{aligned}\frac{d\delta_w(p_m(p_w),p_w)}{dp_w} &= -c'(p)(p_m{}'(p_w)+1) + B_w\alpha_m r_m(A_m p_m{}'(p_w) - A_w) \\ &= (B_w\alpha_m r_m A_m - c'(p))p_m{}'(p_w) - (A_w B_w \alpha_m r_m + c'(p)).\end{aligned}$$

If $p_m{}'(p_w) = 0$, then we are done. If $p_m{}'(p_w) \neq 0$, then by the implicit function theorem,

$$\delta_m(p_m(p_w),p_w) = t_m - c(p_m(p_w),p_w) + B_m(\alpha_m\alpha_w r_w) = 0,$$

where $B_w = (1-\rho_m)\frac{s^2_m}{2}$ and $r_w = 1/r_m$. Since

$$\frac{dr_w}{dp_w} = r_w(A_w - A_m p_m{}'(p_w)),$$

We have

$$0 = \frac{d\delta_m(p_m(p_w),p_w)}{dp_w} = -c'(p)(p_m{}'(p_w)+1) + B_m\alpha_w r_w(A_w - A_m p_m{}'(p_w)).$$

Rearrange to get

$$p_m{}'(p_w) = \frac{A_w B_m \alpha_w r_w - c'(p)}{A_m B_m \alpha_w r_w + c'(p)}.$$

Plugging into $\frac{d\delta_w}{dp_w}$, we get

$$\begin{aligned}
&\frac{d\delta_w(p_m(p_w), p_w)}{dp_w}\\
=&(A_m B_w \alpha_m r_m - c'(p))\frac{A_w B_m \alpha_w r_w - c'(p)}{A_m B_m \alpha_w r_w + c'(p)} - (A_w B_w \alpha_m r_m + c'(p))\\
=&\frac{(A_m B_w \alpha_m r_m - c'(p))(A_w B_m \alpha_w r_w - c'(p)) - (A_m B_m \alpha_w r_w + c'(p))(A_w B_w \alpha_m r_m + c'(p))}{A_m B_m \alpha_w r_w + c'(p)}\\
=&-\frac{(A_m B_w \alpha_m r_m + A_w B_m \alpha_w r_w + A_m B_m \alpha_w r_w + A_w B_w \alpha_m r_m)c'(p)}{A_m B_m \alpha_w r_w + c'(p)} < 0.
\end{aligned}$$

Since $d\delta_w/dp_w < 0$, define p_w^* as follows:

$$p_w^* = \begin{cases} 1 & \text{if } \delta_w(p_m(0), 0) > \delta_w(p_m(1), 1) > 0 \\ \text{solution of } \delta_w(p_m(p_w), p_w) = 0 & \text{if } \delta_w(p_m(0), 0) > 0 > \delta_w(p_m(1), 1) \\ 0 & \text{if } 0 > \delta_w(p_m(0), 0) > \delta_w(p_m(1), 1). \end{cases}$$

Then $(p_m(p_w^*), p_w^*)$ characterizes the unique equilibrium.

A.7. Proof of Proposition 1

A man would strictly prefer the risky career if

$$\delta_m = t_m - c + (1-\rho_m)(\alpha_m + \alpha_w \frac{\sigma_{y_w}}{\sigma_{y_m}})\frac{s_m^2}{2} > 0.$$

If $t_m - c = -\frac{s_m^2}{2}$ (the safe career and the risky career have the same expected income), $\rho_m = 0$, and $\alpha_m = 1$ (the marriage surplus is linear in income), then $\delta_m = \alpha_w \frac{\sigma_{y_w}}{\sigma_{y_m}} \frac{s_m^2}{2} > 0$. If $t_m - c = -\frac{s_m^2}{2} - e_m$, then $\delta_m = -e_m + ((1-\rho_m)(\alpha_m + \alpha_w \frac{\sigma_{y_w}}{\sigma_{y_m}}) - 1)\frac{s_m^2}{2}$. As long as $e_m < ((1-\rho_m)(\alpha_m + \alpha_w \frac{\sigma_{y_w}}{\sigma_{y_m}}) - 1)\frac{s_m^2}{2}$, a risk-averse man would strictly prefer the risky career that yields a lower expected income. Similarly, if $t_w - c = -\frac{s_w^2}{2} - e_w$, then $\delta_w = -k - e_w + ((1-\rho_w)(\alpha_w + \alpha_m \frac{\sigma_{y_m}}{\sigma_{y_w}}))\frac{s_w^2}{2}$. As long as $e_w < ((1-\rho_w)(\alpha_w + \alpha_m \frac{\sigma_{y_m}}{\sigma_{y_w}}))\frac{s_w^2}{2}$, a risk-averse woman would strictly prefer the risky career with a lower expected income.

A.8. Proof of Proposition 2

A greater income variance for men than for women, $\sigma_{y_m}^{*2} = \sigma_x^2 + p_m^* s^2 > \sigma_x^2 + p_w^* s^2 = \sigma_{y_w}^{*2}$, implies

$$p_m^* = \frac{\sigma_{y_m}^2 - \sigma_x^2}{s^2} > \frac{\sigma_{y_w}^2 - \sigma_x^2}{s^2} = p_w^*.$$

A.9. Proof of Proposition 3

Since $0 < p_m^* < 1$ and $0 < p_w^* < 1$, the two equilibrium premiums are

$$\delta_m{}^* = t - c^* + (1-R)\left(\alpha_m + \alpha_w \frac{1}{r^*}\right)\frac{s^2}{2} = 0$$

and

$$\begin{aligned}
\delta_w{}^* =& t - c^* - k + (1-R)(\alpha_w + \alpha_m r^*)\frac{s^2}{2}\\
=& t - c^* - k + r^*(1-R)\left(\alpha_m + \alpha_w \frac{1}{r^*}\right)\frac{s^2}{2} = 0,
\end{aligned}$$

where $r^* = \sigma_{y_m}^*/\sigma_{y_w}^*$. Subtract the two equations to get

$$(r^* - 1)(1-R)\left(\alpha_m + \alpha_w \frac{1}{r^*}\right)\frac{s^2}{2} = k.$$

Since $1 - R$ is assumed to be positive, in order for the left-hand side of the equation to be positive, $r^* > 1$, so $\sigma_{y_m}^* > \sigma_{y_w}^*$.

A.10. Proof of Remark 1

Rearranging the equilibrium condition

$$(r-1)(1-R)(\alpha_m + \alpha_w \frac{1}{r})\frac{s^2}{2} = k,$$

we get

$$(r-1)(\alpha_m r + \alpha_w) = \frac{2k}{(1-R)s^2},$$

$$\alpha_m r^2 - \alpha_m r + \alpha_w r - \alpha_w = \frac{2k}{(1-R)s^2} r,$$

and

$$r^2 - \left(1 - \frac{\alpha_w}{\alpha_m} + \frac{1}{\alpha_m}\frac{2k}{(1-R)s^2}\right) - \frac{\alpha_w}{\alpha_m} = 0.$$

Therefore,

$$r^* = \frac{1}{2}\left[\left(1 - \frac{\alpha_w}{\alpha_m} + \frac{1}{\alpha_m}\frac{2k}{(1-R)s^2}\right) + \sqrt{\left(1 - \frac{\alpha_w}{\alpha_m} + \frac{1}{\alpha_m}\frac{2k}{(1-R)s^2}\right)^2 + 4\frac{\alpha_w}{\alpha_m}}\right].$$

Obviously, r^* increases in k, decreases in $(1-R)$, and decreases in s^2. It is less obvious to derive $\partial r^*/\partial \alpha_w$:

$$\begin{aligned}\frac{\partial r^*}{\partial \alpha_w} =& \frac{1}{2}\left[-\frac{1}{\alpha_m} + \frac{1}{2}\frac{\left(1 - \frac{\alpha_w}{\alpha_m} + \frac{1}{\alpha_m}\frac{2k}{(1-R)s^2}\right) + \frac{4}{\alpha_m}}{\sqrt{\left(1 - \frac{\alpha_w}{\alpha_m} + \frac{1}{\alpha_m}\frac{2k}{(1-R)s^2}\right)^2 + 4\frac{\alpha_w}{\alpha_m}}}\right] \\ =& \frac{1}{2}\frac{1}{\alpha_m}\frac{2-r^*}{\sqrt{\left(1 - \frac{\alpha_w}{\alpha_m} + \frac{1}{\alpha_m}\frac{2k}{(1-R)s^2}\right)^2 + 4\frac{\alpha_w}{\alpha_m}}}.\end{aligned}$$

$\frac{\partial r^*}{\partial \alpha_w} > 0$ if $r^* < 2$, and $\frac{\partial r^*}{\partial \alpha_w} < 0$ if $r^* > 2$.

A.11. Proof of Lemma 5

Marriage timing is related to the career choice as follows.

1. A man who chooses the safe career gets a marriage payoff of $v_m(x_m)$ regardless of his marriage timing decision, so he is indifferent between marrying in the first period and marrying in the second period.
2. A woman who chooses the safe career gets a marriage payoff of $v_w(x_w)$ if she marries in the first period and gets $v_w(x_w - k) < v_w(x_w)$ if she marries in the second period, so she chooses to marry in the first period if she chooses the safe career.
3. An ability-x_m man who chooses the risky career gets a marriage payoff of

   $$s(x_m - c + \varepsilon_m, \tilde{z}_w) - v_w(\tilde{z}_w)$$

   if $\tilde{z}_w$ is his wife's marriage type, and he realizes income $x_m - c + \varepsilon_m$. However, if he waits until the second period to get married, then he gets

   $$v_m(x_m - c + \varepsilon_m) = s(x_m - c + \varepsilon_m, z_w(x_m - c + \varepsilon_m)) - v_w(z_w(x_m - c + \varepsilon_m)).$$

   By the stability condition, for any $\tilde{z}_w$,

   $$v_m(x_m - c + \varepsilon_m) \geq s(x_m - c + \varepsilon_m, \tilde{z}_w) - v_w(\tilde{z}_w).$$

   Therefore, for any realization of ε_m, the man is weakly better off to wait until the second period to marry. Since for different ε_m, the man marries a different partner in the second period, he is almost always strictly better off to wait until the second period to marry. The expected marriage payoff is strictly higher when a man who chooses the risky career marries in the second period.
4. An ability-x_w woman who chooses the risky career gets a marriage payoff of

   $$s(\tilde{z}_m, x_w - c - k + \varepsilon_w) - v_m(\tilde{z}_m),$$

   if $\tilde{z}_m$ is her husband's marriage type and she realizes income $x_w - c + \varepsilon_w$. However, if she waits until the second period to get married when she realizes income $x_w - c + \varepsilon_w$, then she gets

   $$v_w(x_w - c + \varepsilon_w) = s(z_m(x_w - c + \varepsilon_w), x_w - c + \varepsilon_w) - v_m(z_m(x_w - c + \varepsilon_w)).$$

By the stability condition, for any $\widetilde{z}_w$,

$$v_w(x_w - c + \varepsilon_w) \geq s(\widetilde{z}_m, x_w - c + \varepsilon_w) - v_m(\widetilde{z}_m).$$

Therefore, for any realization of ε_w, the woman is weakly better off to wait until the second period to marry. Since for different ε_w the woman marries a different husband in the second period, it is almost always strictly better off to wait until the second period to marry. Hence, the expected marriage payoff is strictly higher when a woman who chooses the risky career marries in the second period.

A.12. Gender differences in postmarital career choices

While our model is mostly about premarital career choices, it also generates some interesting predictions for the postmarital period. If a man marries in the first period after choosing the risky career, the expected surplus he gets from marrying an income-y_w woman is $\mathbb{E}[s(x_m + \varepsilon_m, y_w)|x_m]$. Hence, a man who chooses the risky career and marries early is treated as if he chooses the safe career and marries early. As a result, a male risk-taker is better off waiting to marry in the second period. This result again highlights the fact that the marital benefits from the risky career come from the possibility of switching partners. Remember that the wife a man marries is the wife that maximizes his personal marriage payoff, so choosing the risky career while unmarried is better than choosing the risky career while married. In contrast, an unmarried woman faces the additional reproductive cost, so she is more inclined to choose a safe career while unmarried. However, after she is married and has had children, she no longer worries about the reproductive decline. If she can choose a different career, she may opt to choose a riskier career. This gender-differential effect yields the second part of the proposition below.

Proposition 6. *A married person is less likely than an unmarried person to choose a risky career, and a married woman is more likely to switch to a risky career than a married man.*

Proof of Proposition 6. The effects of marital status on the risky career choice differ by gender in the following ways. If a type-z_m man who is married to a type $\widetilde{z}_w$ woman chooses the risky career and realizes an income $z_m - c + \varepsilon_m$, he gets a marriage payoff of $s(z_m - c + \varepsilon_m, \widetilde{z}_w) - v_w(\widetilde{z}_w)$. In contrast, if the same type z_m man who is unmarried chooses the risky career and realizes an income $z_m - c + \varepsilon_m$, he gets a marriage payoff of $s(z_m - c + \varepsilon_m, z_w(z_m - c + \varepsilon_m)) - v_w(z_w(z_m - c + \varepsilon_m)) \geq s(z_m - c + \varepsilon_m, \widetilde{z}_w) - v_w(\widetilde{z}_w)$, and the inequality holds strictly as long as $-c + \varepsilon_m \neq 0$.

If a type-z_w woman who is married to a type $\widetilde{z}_m$ man chooses the risky career and realizes an income $z_w - c + \varepsilon_w$, she gets a marriage payoff of

$$s(\widetilde{z}_m, z_w - c + \varepsilon_w) - v_m(\widetilde{z}_m).$$

In contrast, if the same type z_w woman who is unmarried chooses the risky career and realizes an income $z_w - c + \varepsilon_w$, she gets marriage payoff of

$$s(z_m(z_w - c + \varepsilon_w - k), z_w - c + \varepsilon_w - k) - v_m(z_m(z_w - c + \varepsilon_w - k)).$$

The difference between the marriage payoffs can be written as

$$\begin{aligned}&[s(z_m(z_w - c + \varepsilon_w - k), z_w - c + \varepsilon_w - k) - v_m(z_m(z_w - c + \varepsilon_w - k))]\\ &- [s(\widetilde{z}_m, z_w - c + \varepsilon_w - k) - v_m(\widetilde{z}_m)]\\ &+ [s(\widetilde{z}_m, z_w - c + \varepsilon_w - k) - v_m(\widetilde{z}_m)] - [s(\widetilde{z}_m, z_w - c + \varepsilon_w) - v_m(\widetilde{z}_m)].\end{aligned}$$

The first two terms together are positive, but the third term is negative. The first two terms show that by deciding whom to marry after the income is realized, an unmarried woman has a higher incentive to choose the risky career. The last term shows that the absence of a reproductive decline gives a married woman a higher incentive to choose the risky career. □

While this is an interesting theoretical prediction, there are other considerations with regard to switching one's careers after marriage, many of which are not included in the model. The accumulation of specific human capital makes it costly to switch to another career. Once a person is married and has children, household production demands an increasing amount of time, and the burden mostly falls on the wife. *Prima facie* evidence does not fully support Proposition 6. From the 2016 American Community Survey, the share of men in risky careers gradually increases with age, from around 57% between ages 25 and 34 to about 65% between 45 and 54. In contrast, the share of women in risky careers is rather stable across different age groups, at around 50%. Using a balanced panel from the National Longitudinal Survey of Young 1979 and exploiting within-person variation, we find that being married is associated with a 1.2 pp reduction in the probability of being in the risky occupation for men, and a 4.2 pp reduction for women. This is consistent with voluminous studies that show wives are more likely to sacrifice career for family (e.g., Goldin, 1990, 2015; Bertrand et al., 2010).

Table B.1
Most and least risky popular occupations.

	(1)	(2)	(3)	(4)
	% in skilled			
Occupation	workers	% male	$OccRisk_o$	$\bar{\hat{\varepsilon}}_o$
Panel A: Most risky occupations				
Real estate sales occupations	0.93	49.77	0.86	-0.14
Insurance sales occupations	0.49	61.37	0.80	0.03
Lawyers	2.06	59.27	0.78	0.5
Retail sales clerks	1.17	52.93	0.78	-0.36
Supervisors and proprietors of sales jobs	2.73	62.62	0.77	-0.08
Chief executives and public administrators	1.91	74.18	0.75	0.48
Other financial specialists	1.4	61.89	0.73	0.27
Physicians	1.73	60.41	0.73	0.87
Nursing aides, orderlies, and attendants	0.74	23.27	0.71	-0.64
Management analysts	1.23	56.25	0.71	0.24
Panel B: Least risky occupations				
Managers in education and related fields	1.69	38.5	0.48	0.01
Police, detectives, and private investigators	1.13	80.58	0.47	-0.01
Physical therapists	0.47	32.17	0.46	0.12
Computer software developers	2.44	78.01	0.46	0.25
Registered nurses	3.96	12.96	0.46	0.17
Primary school teachers	7.52	20.14	0.44	-0.25
Special education teachers	0.48	13.44	0.44	-0.2
Secondary school teachers	1.51	40.84	0.43	-0.25
Social workers	1.31	18.29	0.43	-0.23
Pharmacists	0.5	37.11	0.42	0.49

Note: Among the 50 largest occupations for college-educated workers, the table reports the 10 most risky occupations and the 10 least risky occupations according to the measure of occupational riskiness in Eq. (10). The underlying data is the 5-year ACS ending in 2016. The sample includes workers with a college degree, report an occupation, worked at least part-time in a typical week in the previous year, and are between 45 and 49 years old. For each occupation, the table reports its share in total college graduates employment (Column 1), share of male workers among college graduates in this occupation (Column 2), occupational riskiness (Column 3), and the average residual log wage (Column 4).

Table B.2
Include both college and non-college graduates.

	(1)	(2)	(3)	(4)	(5)
	= 1 if	= 1 if never married		= 1 if have children	
	risky occ	women	men	women	men
= 1 if male	0.132***				
	(0.001)				
= 1 if risky occupation		0.013***	-0.003	-0.009	0.010***
		(0.001)	(0.002)	(0.002)	(0.002)
N	2234207	414531	433812	426546	451366
mean dep var	0.560	0.118	0.135	0.658	0.643

Note: The data is from the 5-year ACS ending in 2016. The table reports robustness results to gender gap in risky occupational choices as well as marital and fertility outcomes. Here the sample includes both college and non-college graduates. Log residual wage is obtained from Eq. (9) separately for college and non-college graduates. Risky occupations are defined, separately for each skill level, according to whether the standard deviation of the log residual wage is above or below the median. The table investigates the gender gap in risky occupational choice (Column 1), as well as marriage and fertility outcomes (Columns 2 through 5). Individual characteristics and other occupational characteristics are controlled for. The specification in Column 1 follows that of Table 3 Column 6. The specifications in Columns 2 to 5 follow those of Columns 2 to 5 in Table 4. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Appendix B. Additional empirical results

B.1. Most and least risky popular occupations

See Table B.1.

B.2. Occupation wage riskiness and female ratio by detailed education levels

See Fig. B.1 and Table B.2.

B.3. Alternative definitions of occupational riskiness

See Tables B.3–B.5.

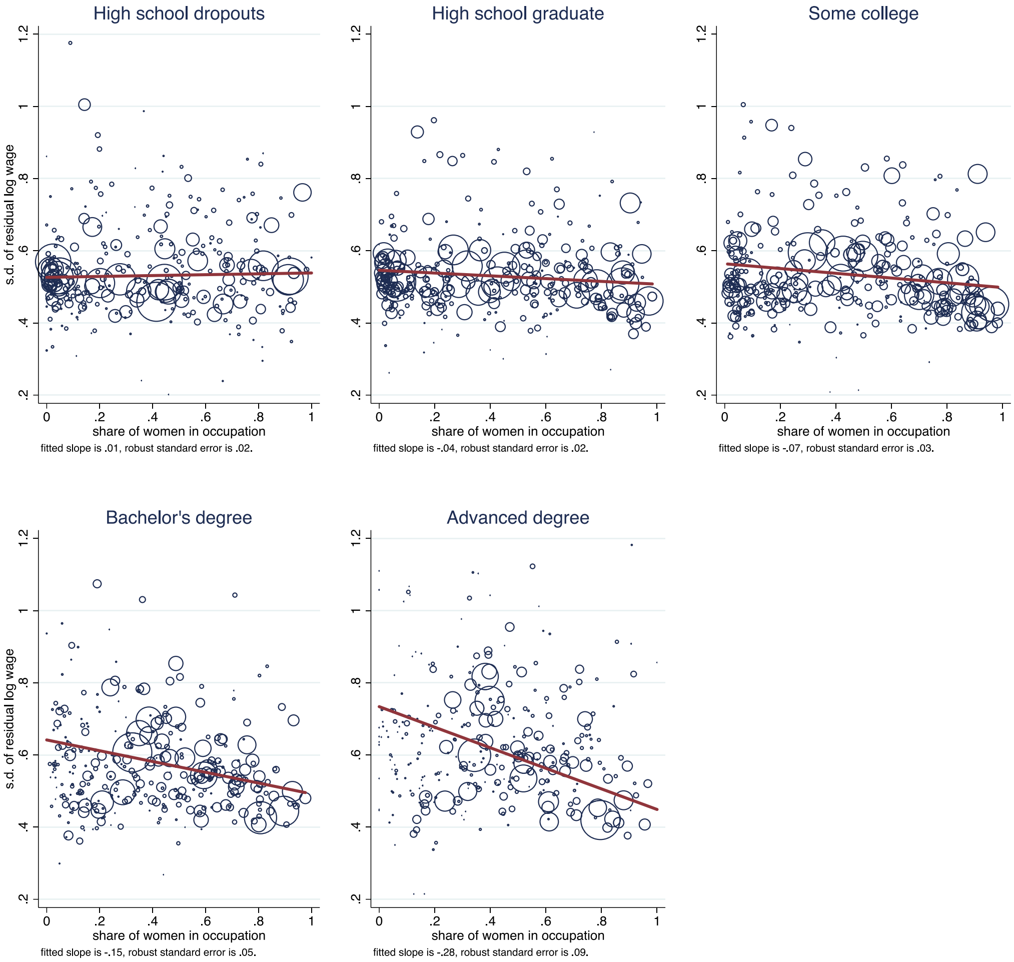


Fig. B.1. Occupation Wage Riskiness and Female Ratio — by Detailed Education Levels. Note: The data is from the 5-year American Community Surveys (ACS) ending in 2016, which represents 5% of the US population. The sample includes those between ages 16 and 64 who report an occupation and are not currently enrolled in school. Hourly wage is imputed as the division between last year's earnings by annual working hours. Separately for each detailed occupation level, log wage is first regressed against age and age squared, an indicator for gender, a set of race indicators, as well as the exhaustive interactive terms of the three sets of variables. Occupation wage riskiness is measured as the standard deviation of the residual log wage within the occupation. Each bubble in the graph represents an occupation defined by the IPUMS. The size of the bubble is proportional to the number of workers in the occupation for the specific skill group, which is also used as the weight in the linear fit regression shown in the red line.

B.4. Other confounding factors

Although the results presented in this section are only intended to be correlational, rich information in the NLSY79 helps to rule out several confounding factors. One such factor is "effort". It is conceivable that the earnings variation within an occupation does not unravel on its own; instead, it is the result of early-career investment. Such investment could take the form of on-the-job training or climbing the corporate ladder. Workers who go through more training and accumulate a higher human capital or those who get promoted will make a lot more than those who do not. In this view, women are less likely to embark on a career that requires a long period of apprenticeship because they are less willing or more constrained to commit an exorbitant amount of time to work. It is worth noting that women's unwillingness to choose demanding careers may be precisely because those careers demand a lot of time and women cannot afford waiting too long before starting a family and have children. We have also controlled for the prevalence of working overtime, and a measure of mom-friendliness, which at least partly account for this alternative explanation.Using the longitudinal nature of the NLSY79, we measure the actual cumulative experience at the time when risky occupation is measured (at age 30).[16] We also control for ability (AFQT score adjusted for age) as one may suspect that workers with a higher ability may sort into more competitive occupations.[17]

[16] The variable is log average annual hours of work between the time of finishing formal schooling and age 30.

[17] Although the dispersion of residual log wage within an occupation is largely uncorrelated with the average residual log wage of that occupation.

Table B.3
Alternative ways of calculating residual wage.

dep var	= 1 if	= 1 if never married		= 1 if have children	
	risky occ (1)	women (2)	men (3)	women (4)	men (5)
Panel A: Control for location FE; corr. coeff. with baseline = 0.99					
= 1 if male	0.057*** (0.001)				
= 1 if risky occupation		0.009** (0.003)	−0.002 (0.003)	−0.012** (0.004)	0.013*** (0.004)
Panel B: Wage from age group 30–35; corr. coeff. with baseline = 0.85					
= 1 if male	0.073*** (0.001)				
= 1 if risky occupation		0.013*** (0.002)	−0.008** (0.003)	−0.021*** (0.003)	0.016*** (0.003)
Panel C: Wage from age group 35–40; corr. coeff. with baseline = 0.92					
= 1 if male	0.055*** (0.001)				
= 1 if risky occupation		0.012*** (0.003)	0.003 (0.003)	−0.020*** (0.004)	0.004 (0.004)
Panel D: Wage from full-time workers; corr. coeff. with baseline = 0.99					
= 1 if male	0.041*** (0.001)				
= 1 if risky occupation		0.006** (0.003)	0.003 (0.003)	−0.012** (0.004)	0.005 (0.004)

Note: The table reports robustness results where occupational riskiness is defined using log residual wage calculated in alternative ways. Panel A adds a set of commuting zone fixed effects in Eq. (9). Panel B calculates the residual log wage using workers between 30 and 35 years old (instead of those between 40 and 49 years old). Panel C uses workers between 35 and 40 years old. Panel D uses full-time workers only (instead of at least part-time). Risky occupations are then defined as those with residual log wage standard deviation above the median. For each alternative definition of a risky occupation, the table investigates the gender gap in risky occupational choice (Column 1), as well as marriage and fertility outcomes (Columns 2 through 5). Individual characteristics and other occupational characteristics are controlled for. The specification in Column 1 follows Table 3 Column 6. Columns 2 through 5 correspond to Table 4. Robustness standard errors are in parentheses. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table B.4
Choice of risky occupation: alternative measures of occupational riskiness.

	(1)	(2)	(3)	(4)
	= 1 if risky occupation measured at			continuous
	p50	p75	p90	s.d.
= 1 if male	0.042*** (0.001)	0.040*** (0.001)	0.001** (0.000)	0.015*** (0.000)
mean dep var	0.536	0.242	0.032	0.588

Note: The data is from the 5-year ACS ending in 2016. The sample includes those between ages 25 and 39 who have a college degree and report an occupation. In Columns 1 through 3, the dependent variable is an indicator for whether the person is in a risky occupation, with risky occupation defined as those above the median (Column 1, baseline), 75th percentile (Column 2), and 90th percentile (Column 3) of the standard deviation of the log residual wage, respectively. In Column 4, the dependent variable is the standard deviation of residual log wage, a continuous variable. Individual and other occupational characteristics are controlled for as in Table 3 Column 6. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Regression results with those additional controls (on top of what are already controlled for in Table 5) are reported in Column 2 of Appendix Table B.6 (for the gender gap in risky occupational choice) and Appendix Table B.8 (for marriage and fertility outcomes). The estimated coefficients are largely comparable to those in Table 5.

Other confounding factors are preferences and expectations. It is possible that the preference for job riskiness exhibits a gender difference and is correlated with attitudes toward marriage, children, and other life outcomes. Again, the gender gap in risk preference could reflect that women rationally expect marriage-market disadvantages if they embark on a risky occupation and wait too long before they enter the marriage market, which is consistent with our hypothesis.

Typically, preferences and expectations are difficult to observe. One unique feature of the NLSY79 is that it asked the respondents about their expected age of marriage and the number of children when the respondents were between teens and early 20 s. After the respondents were in their late 20 s and early 30 s, the survey asked a series of hypothetical questions on whether the respondent would take jobs that exhibit varying degrees of income uncertainty, thus eliciting their risk tolerance specifically regarding labor market outcomes.

Appendix Table B.7 shows the correlation coefficients between stated expectations or preferences and actual choices later in life (Panel A), and between expectations/preferences and the choices of risky occupations (Panel B). Expectations and actual life

Table B.5
Marriage and fertility outcomes: alternative measures of occupational riskiness.

	= 1 if never married		= 1 if have children	
	women	men	women	men
Panel A: Cutoff at median (baseline)	(1)	(2)	(3)	(4)
= 1 if risky occupation	0.008**	0.003	-0.013**	0.005
	(0.003)	(0.003)	(0.004)	(0.004)
Panel B: Cutoff at 75th percentile	(1)	(2)	(3)	(4)
= 1 if risky occupation	0.006**	0.003	0.001	0.014***
	(0.003)	(0.002)	(0.004)	(0.003)
Panel C: Cutoff at 90th percentile	(1)	(2)	(3)	(4)
= 1 if risky occupation	0.038***	0.001	-0.040**	0.053***
	(0.011)	(0.010)	(0.014)	(0.013)
Panel D: s.d. of res. log wage	(1)	(2)	(3)	(4)
1st quintile (lowest s.d.)	-	-	-	-
2nd quintile	0.031***	0.017***	-0.048***	-0.032***
	(0.004)	(0.004)	(0.005)	(0.006)
3rd quintile	0.028***	0.020***	-0.052***	-0.029***
	(0.004)	(0.005)	(0.005)	(0.007)
4th quintile	0.035***	0.015***	-0.060***	-0.023***
	(0.004)	(0.005)	(0.005)	(0.006)
5th quintile	0.032***	0.014**	-0.044***	-0.008
	(0.004)	(0.005)	(0.005)	(0.006)
# of obs.	149764	143789	159484	150616

Note: The table investigates the correlation between occupational riskiness and marriage and fertility outcomes. The data is from the 5-year ACS ending in 2016. The sample includes those between ages 45 and 49 who have a college degree and report an occupation. Individual and other occupational characteristics are controlled for as in Table 4. In addition, to match the flexible definition of risky occupation, which is based on the standard deviation of log residual wage, flexible polynomials of average log residual wage are also controlled for. In Panels A through C, risky occupations are defined as those with the standard deviation of the residual log wage above the median (Panel A, the baseline), above the 75th percentile (Panel B), and above the 90th percentile (Panel C). In Panel D, occupations are divided into quintiles according to the standard deviation of the residual log wage. $*p < 0.1$, $**p < 0.05$, $***p < 0.01$.

Table B.6
Gender gap in risky occupational choices: NLSY79 robustness checks.

dep var: = 1 if in a risky occ by age 30	(1) baseline	(2) + AFQT, actual exp.	(3) + risk attitude	(4) + marriage/fertility expectations
= 1 if male	0.122***	0.122***	0.104***	0.099***
	(0.020)	(0.021)	(0.023)	(0.023)
adj. AFQT		-0.001	-0.001	-0.000
		(0.001)	(0.001)	(0.001)
log actual exp. (25–29)		-0.305**	-0.306**	-0.183
		(0.108)	(0.115)	(0.120)
risk tolerance (from low to high), leave out category: 1st quartile				
2nd quartile			0.030	0.015
			(0.033)	(0.032)
3rd quartile			0.037	0.045
			(0.029)	(0.030)
4th quartile			0.030	0.020
			(0.028)	(0.029)
preferred # of children, leave out category: 0				
1				-0.038
				(0.030)
2-3				-0.004
				(0.027)
≥4				0.073
				(0.082)
expected age of marriage, leave out category: less than 25				
25–29				0.037*
				(0.023)
above 30				0.032
				(0.049)
never				0.042
				(0.069)
N	3021	2925	2244	2036

Note: The dependent variable is an indicator for whether the person is in a risky occupation by age 30. The sample includes men and women in NLSY79 with a college degree. All regressions also include log wage at age 30, year of birth fixed effects, individual demographic characteristics and other occupational characteristics, and are weighted by sample weights. Column 1 is the baseline, which corresponds to Column 1 of Table 5. Column 2 adds the AFQT score adjusted for age, as well as actual work experience. Column 3 adds quartiles of risk tolerance, derived from hypothetical questions in the NLSY79 on whether the respondent is willing to take jobs that exhibit different levels of riskiness. Column 4 adds indicators for the expected number of children and the expected age of first marriage. Robust standard errors are in parentheses. $*p < 0.1$, $**p < 0.05$, $***p < 0.01$.

Table B.7
Expectations and preferences: correlations with actual choices.

Panel A: expectations vs. actual choices	all	male	female
# of children in the household	.083	.073	.101
# of biological children (women only)	.	.	.079
age at marriage	.25	.241	.238
Panel B: expectations & pref. vs. risky occ. choice	all	male	female
risk attitude	.057	.034	.045
expected number of children	.02	.076	-.027
expected age of marriage	.047	.016	.034

Note: The table shows the correlation coefficients between stated expectations/preferences and observed choices. The sample includes men and women in the NLSY79 with a college degree. Panel A reports the correlation coefficients between stated expectations regarding marriage and fertility and actual life outcomes in later years. The total number of children the respondent expected to have is from a question in the 1979 wave (when the respondents were between 14 and 22 years old). The actual outcomes are (1) the observed number of children in the household when the respondent was 40 years old, and (2) the total number of biological children in the last wave of survey when the respondent is surveyed (for female respondents only, conditional on the last wave of survey was in or after 2000. The number of biological children is adjusted for age at the last survey). The age expected to marry (for the first time) is from a question in the 1979 wave, in which the respondent was asked to choose among (1) less than 20, (2) 20–24, (3) 25–29, (4) above 30, and (5) never. Actual marital history by age 40 is grouped into the same set of bins to match the answers to the question on the expected marriage age. Those who had never been married by 40 are classified as the never-marry category. Panel B reports the correlation coefficients between expectations and whether the person was in a risky occupation by age 30. Risky attitude is a four-category variable derived from a set of hypothetical questions in the 1993 wave (when the respondents were between 28 and 36 years old). The respondent were asked whether they are willing to take a job that "can either double the family income or reduce it by X percent," where X takes value of 50, 30, and 25. We call a respondent has the highest risk tolerance if they are willing to take a job that could reduce family income by half (the risk attitude variable takes value 4) and one that has the lowest risk tolerance if they answer no to the question regardless of the value of X (the risk attitude variable takes value 1).

Table B.8
Marriage and fertility outcomes, controlling for ability and actual experiences.

	(1) = 1 if ever married by age 40	(2)	(3) # of children in hhds by 40	(4)	(5) biological children #	(6) any
	female	male	female	male	female	female
= 1 if in risky occ by age 30	-0.022	-0.029	-0.183*	-0.090	-0.159	-0.074**
	(0.030)	(0.035)	(0.095)	(0.113)	(0.101)	(0.032)
ln inc. by age 30	-0.018**	0.022**	-0.094***	0.084***	-0.077**	-0.015**
	(0.006)	(0.010)	(0.020)	(0.025)	(0.021)	(0.006)
adj. AFQT	0.001	0.002**	0.006**	0.002	0.003	-0.000
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)
log actual exp. (25–29)	0.351**	0.015	-0.416	-0.038	-0.558	-0.032
	(0.136)	(0.147)	(0.443)	(0.557)	(0.477)	(0.132)
year of birth FE	X	X	X	X	X	X
occ. chars.	X	X	X	X	X	X
age when last surveyed					X	X
ind. demo.	X	X	X	X	X	X
# of individuals	1244	988	1133	903	1171	1171
mean dep var	0.758	0.772	1.564	1.339	1.705	0.759

Note: The sample includes men and women with a college degree in the NLSY79. The dependent variable is whether the respondent has ever been married by age 40 in Columns 1 and 2, the number of children in the household when the respondent was 40 in Columns 3 and 4, the number of biological children for women who were observed for the last time after year 2000 in Column 5, and whether the women had any biological children in Column 6. All regressions include a set of year of birth fixed effects, individual demographic characteristics (Black, Hispanic, or non-Black and non-Hispanic), other occupational characteristics, and are weighted by sample weights, as in Table 5 Columns 2 through 7. Other occupational characteristics include the mean residual log wage, the spatial dissimilarity index, long-hour wage premium, and an index measuring the occupation's overall friendliness toward women with young children. In addition, all regressions include a set of additional controls as in Appendix Table B.6 Column 2. They include the AFQT score adjusted for age and the actual work experience between ages 25 and 29. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

outcomes are positively correlated, but the correlation is not particularly strong. The correlations between expectations and choices of risky occupations are also rather weak. Elicited risk tolerance with regard to earnings uncertainty is positively correlated with the actual choice of a risky occupation. Those who report a later marriage (or would never marry) are more likely to choose a risky career. Interestingly, the expected number of children is positively correlated with the choice of a risky occupation for men, but negatively correlated for women.

Table B.9
Marriage and fertility outcomes controlling for preferences and expectations.

	(1) = 1 if ever married by age 40	(2)	(3) # of children in hhds by 40	(4)	(5) biological children #	(6) any
	female	male	female	male	female	female
= 1 if in risky occ by age 30	-0.016	-0.032	-0.213**	-0.137	-0.191*	-0.049
	(0.031)	(0.037)	(0.095)	(0.109)	(0.102)	(0.035)
ln inc. by age 30	-0.011**	0.026***	-0.061***	0.076***	-0.060**	-0.010*
	(0.006)	(0.010)	(0.020)	(0.021)	(0.021)	(0.006)
adj. AFQT	0.001	0.002**	0.007***	0.003	0.004**	0.000
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)
log actual exp. (25–29)	0.312**	-0.016	-0.388	-0.347	-0.459	-0.000
	(0.153)	(0.150)	(0.428)	(0.560)	(0.471)	(0.146)
year of birth FE	X	X	X	X	X	X
occ. chars.	X	X	X	X	X	X
age when last surveyed					X	X
ind. demo.	X	X	X	X	X	X
indicators for						
risk attitude	X	X	X	X	X	X
marriage expectations	X	X	X	X	X	X
fertility expectations	X	X	X	X	X	X
# of individuals	1089	892	980	812	1033	1033
mean dep var	0.765	0.775	1.508	1.342	1.643	0.746

Note: The sample includes men and women in the NLSY79 with a college degree. In Columns 1 and 2, the dependent variable is whether the respondent has ever been married by age 40. In Columns 3 and 4, the dependent variable is the number of children in the household when the respondent was 40. In Columns 5 the dependent variable is the number of biological children for women who were observed for the last time after year 2000. In Column 6, the dependent variable is whether the women has any biological children. All regressions include a set of year of birth fixed effects, individual demographic characteristics (Black, Hispanic, or non-Black and non-Hispanic), other occupational characteristics, and are weighted by sample weights, as in Table 5 Columns 2 through 7. Other occupational characteristics include mean residual log wage, the spatial dissimilarity index, long-hour wage premium, and an index measuring the occupation's overall friendliness toward women with young children. In addition, all regressions include a set of additional controls as in Appendix Table B.6 Column 4. They include the AFQT score adjusted for age, the actual work experience between ages 25 and 29, sets of indicators capturing risk tolerance, preferred number of children, and preferred age of marriage. Robust standard errors in parentheses. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

We add to measures capturing preferences and expectations on top of the specifications in Table 5 (as well as controls for aptitude and actual experiences). Columns 3 and 4 of Appendix Table B.6 report results regarding the gender gap in risky occupational choices; Appendix Table B.9 reports results regarding family and fertility outcomes. Results are largely comparable to those in Table 5.

Appendix C. Supplementary data

Supplementary material related to this article can be found online at https://www.dropbox.com/sh/nl5oaouoycgcfxs/AADzhda2isOzFZR-PH-SuS2ja?dl=0.

References

Adda, J., Dustmann, C., Stevens, K., 2017. The career costs of children. J. Polit. Econ. 125 (2), 293–337.
Altonji, J.G., Blank, R.M., 1999. Race and gender in the labor market. In: Ashenfelter, O.C., Card, D. (Eds.), Handbook of Labor Economics. Vol. 3, Elsevier Science B. V., pp. 3143–3259.
Angrist, J., 2002. How do sex ratios affect marriage and labor markets? Evidence from america's second generation. Q. J. Econ. 3 (117), 997–1038.
Bailey, M., 2006. More power to the pill: The impact of contraceptive freedom on women's life cycle labor supply. Q. J. Econ. 121 (1), 289–320.
Barbulescu, R., Bidwell, M., 2012. Do women choose different jobs from men? Mechanisms of application segregation in the market for managerial workers. Organ. Sci. 1–20.
Becker, G.S., 1964. Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education. The University of Chicago Press, Chicago.
Becker, G.S., 1973. A theory of marriage: Part I. J. Polit. Econ. 81 (4), 813–846.
Becker, G.S., 1974. A theory of marriage: Part II. J. Polit. Econ. 82 (2), S11–S26.
Becker, G.S., Murphy, K.M., Werning, I., 2005. The equilibrium distribution of income and the market for status. J. Polit. Econ. 113 (2), 282–310.
Ben-Porath, Y., 1967. The production of human capital and the life cycle of earnings. J. Polit. Econ. 75 (4, Part 1), 352–365.
Benson, A., 2014. Rethinking the two-body problem: The segregation of women into geographically dispersed occupations. Demography 51 (5), 1619–1639.
Bergstrom, T.C., Bagnoli, M., 1993. Courtship as a waiting game. J. Polit. Econ. 101 (1), 185–202.
Bertrand, M., Goldin, C., Katz, L., 2010. Dynamics of the gender gap for Young professionals in the financial and corporate sectors. Am. Econ. J.: Appl. Econ. 2, 228–255.
Bhaskar, V., Hopkins, E., 2016. Marriage as a rat race: Noisy premarital investments with assortative matching. J. Polit. Econ. 124 (4), 992–1045.

Bronson, M.A., 2019. Degrees are forever: Marriage, educational investment, and lifecycle labor decisions of men and women. Mimeo.
Bursztyn, L., Fujiwara, T., Pallais, A., 2017. 'Acting wife': Marriage market incentives and labor market investments. Amer. Econ. Rev. 107 (11), 3288–3319.
Buser, T., Niederle, M., Oosterbeek, H., 2014. Gender, competitiveness, and career choices. Q. J. Econ. 1–39.
Byrnes, J.P., Miller, D.C., Schafer, W.D., 1999. Gender differences in risk taking: a meta-analysis. Psychol. Bull. 125 (3), 367.
Cárdenas, J.-C., Dreber, A., Von Essen, E., Ranehill, E., 2012. Gender differences in competitiveness and risk taking: Comparing children in Colombia and Sweden. J. Econ. Behav. Organ. 83 (1), 11–23.
Charness, G., Gneezy, U., 2012. Strong evidence for gender differences in risk taking. J. Econ. Behav. Organ. 83 (1), 50–58.
Chiappori, P.-A., Iyigun, M., Weiss, Y., 2009. Investment in schooling and the marriage market. Amer. Econ. Rev. 99 (5), 1689–1713.
Corinne Low, 2021. The human capital - reproductive capital tradeoff in marriage market matching. Mimeo.
Deming, D.J., 2017. The growing importance of social skills in the labor market. Q. J. Econ. 132 (4), 1593–1640.
Dreber, A., Hoffman, M., 2007. 2D:4D and risk aversion: Evidence that the gender gap in preferences is partly biological. Mimeo.
Eckel, C.C., Grossman, P.J., 2002. Sex differences and statistical stereotyping in attitudes toward financial risk. Evol. Hum. Behav. 23 (4), 281–295.
Fan, J., Zou, B., 2021. The dual local markets: Family, jobs, and the spatial distribution of skills. Mimeo.
Friedman, M., 1953. Choice, chance, and the personal distribution of income. J. Polit. Econ. 61 (4), 277–290.
Friedman, M., Savage, L.J., 1948. Utility analysis of choices involving risk. J. Polit. Econ. 56 (4), 279–304.
Gershoni, N., Low, C., 2021a. Older yet fairer: How extended reproductive time horizons reshaped marriage patterns in Israel. Am. Econ. J.: Appl. Econ. 13 (1), 198–234.
Gershoni, N., Low, C., 2021b. The power of time: The impact of free IVF on women's human capital investments. Eur. Econ. Rev. 133, 103645.
Gill, D., Prowse, V., 2014. Gender differences and dynamics in competition: The role of luck. Quantitative Economics 5, 351–376.
Goldin, C., 1990. Understanding the Gender Gap: An Economic History of American Women. Oxford University Press, New York.
Goldin, C., 2014. A grand gender convergence: Its last chapter. Amer. Econ. Rev. 104 (4), 1–30.
Goldin, C., 2015. Career and family: Collision or confluence. The Eighth Annual Kenneth Arrow Lecture.
Goldin, C., Katz, L., 2002. The power of the pill: Oral contraceptives and women's career and marriage decisions. J. Polit. Econ. 110 (4), 730–770.
Goldin, C., Katz, L.F., 2016. A most egalitarian profession: Pharmacy and the evolution of a family-friendly occupation. J. Labor Econ. 34 (3), 705–746.
Gretsky, N.E., Ostroy, J.M., Zame, W.R., 1992. The nonatomic assignment model. Econom. Theory 2, 103–127.
Iyigun, M., Walsh, R.P., 2007. Building the family nest: Premarital investments, marriage markets, and spousal allocations. Rev. Econom. Stud. 74, 507–535.
Kleinjans, K., 2009. Do gender differences in preferences for competition matter for occupational expectations? J. Econ. Psychol. 30 (5), 701–710.
Koellinger, P., Minniti, M., Schade, C., 2013. Gender differences in entrepreneurial propensity. Oxf. Bull. Econ. Stat. 75 (2), 213–234.
Kroeger, S., La Mattina, G., 2017. Assisted reproductive technology and women's choice to pursue professional careers. J. Popul. Econ. 30 (3), 723–769.
Lafortune, J., 2013. Making yourself attractive: Pre-marital investments and the returns to education in the marriage market. Am. Econ. J.: Appl. Econ. 5 (2), 151–178.
Lundberg, S., 1985. The added worker effect. J. Labor Econ. 3 (1, Part 1), 11–37.
Mincer, J., 1978. Family migration decisions. J. Polit. Econ. 86 (5), 749–773.
Niederle, M., Vesterlund, L., 2007. Do women shy away from competition? Do men compete too much? Q. J. Econ. 122 (3), 1067–1101.
Patnaik, A., Venator, J., Wiswall, M., Zafar, B., 2020. The role of heterogeneous risk preferences, discount rates, and earnings expectations in college major choice. J. Econometrics.
Robson, A.J., 1992. Status, the distribution of wealth, private and social attitudes to risk. Econometrica 60 (4), 837–857.
Robson, A.J., 1996. The evolution of attitudes to risk: Lottery tickets and relative wealth. Games Econom. Behav. 14, 190–207.
Rodham, K., Hawton, K., Evans, E., Weatherall, R., 2005. Ethnic and gender differences in drinking, smoking and drug taking among adolescents in England: a self-report school-based survey of 15 and 16 year-olds. J. Adolescence 28 (1), 63–73.
Rosen, S., 1997. Manufactured inequality. J. Labor Econ. 15 (2), 189–196.
Rubin, P.H., Paul, C.W., 1979. An evolutionary model of taste for risk. Econ. Inq. 17 (4), 585–596.
Ruggles, S., Genadek, K., Goeken, R., Grover, J., Sobek, M., 2017. Integrated Public Use Microdata Series: Version 7.0 [dataset]. Minneapolis: University of Minnesota.
Siow, A., 1998. Differential fecundity, markets, and gender roles. J. Polit. Econ. 106 (2), 334–354.
Smith, A., 1776. An Inquiry Into the Nature and Causes of the Wealth of Nations. W. Strahan and T. Cadell, London.
Stephens, Jr., M., 2002. Worker displacement and the added worker effect. J. Labor Econ. 20 (3), 504–537.
United Nations, 1990. Patterns of first marriage: Timing and prevalence. Tech. rep., United Nations. Department of International Economic and Social Affairs.
United Nations, 2000. World marriage patterns 2000. Tech. rep., United Nations. Department of International Economic and Social Affairs.
Wang, H., 2019. Intra-household risk sharing and job search over the business cycle. Rev. Econ. Dyn. 34, 165–182.
Wiswall, M., Zafar, B., 2021. Human capital investments and expectations about career and family. J. Polit. Econ. 129 (5), 1361–1424.
Wozniak, D., Harbaugh, W.T., Mayr, U., 2014. The menstrual cycle and performance feedback alter gender differences in competitive choices. J. Labor Econ. 32 (1), 161–198.
Wu, J., Zhang, H., 2021. Preference evolution in different matching markets. Eur. Econ. Rev. 137, 103804.
Xiao, P., 2021. Wage and employment discrimination by gender in labor market equilibrium. Mimeo.
Zafar, B., 2013. College major choice and the gender gap. J. Hum. Resour. 48 (3), 545–595.
Zhang, H., 2015. Essays in matching, auctions, and evolution (Ph.D. thesis). University of Chicago.
Zhang, H., 2020. Pre-matching gambles. Games Econom. Behav. 121, 76–89.
Zhang, H., 2021. An investment-and-marriage model with differential fecundity: On the college gender gap. J. Polit. Econ. 129 (5), 1464–1486.