

Appendix A for online publication: Theoretical Framework

Ingvild Almås,^{1,2} Andreas Kotsadam,³ Espen R. Moen,⁴
Knut Røed,³

¹IIES, Stockholm University,

²Norwegian School of Economics,

Department of Economics, N-5045 Bergen, Norway

³Frisch Center, University of Oslo, N-0317 Oslo, Norway

⁴Norwegian Business School

The logic of hypergamy

There exists a substantial literature on two-sided matching in the marriage market, as well as in other markets such as the labor market; see, e.g., Browning et al. (2014) and Chade et al. (2017). Within the framework of a standard matching model of the marriage market, we analyze how asymmetries between the genders may lead to asymmetric matching outcomes, in which females tend to marry males with earnings potential that exceeds their own. We discuss two potential mechanisms. The first mechanism is related to the biological fact that men are reproductive for a longer period than women, and hence may stay in the marriage market longer. As previously highlighted by, e.g., Siow (1998) and Polachek et al. (2015), this implies that fertile women are relatively scarce and can be choosy with respect to a male partner's attributes such as education or earning. The second mechanism occurs through gender differences in preferences over potential mating partners. This mechanism has similarities with that in Chen and Perroni (2013), in which the females income potential is discounted as they are systematically discriminated in the labor market.

The fertility model

Time is discrete. In the beginning of each period, a unit measure of males and of females enter the market. The type of each individual (both males and females)

is stochastic and drawn from a cumulative distribution $F(y)$. All agents are born fertile. At the end of a period, a fertile agent may be exposed to a fertility shock and lose his/her fertility. The probability of this event is τ^f for females and τ^m for males. We assume that $\tau^f \geq \tau^m$. Infertility is an absorbing state.

All agents are born single. In each period, all single, fertile individuals join a fully competitive and frictionless marriage market. Matching is assortative, all individuals marry the person with the highest y that is willing to marry them. At the end of each period, a married couple may be hit by a divorce shock, which happens with probability k . For simplicity we assume that the divorce shock is independent of the fertility status of the spouses. After divorce, those agents who are still fertile join the matching market, while the infertile agents stay single (or exits).

The model is stationary. Due to the Markov nature of the fertility process, any fertile individual of a particular gender has the same expected future lifetime as fertile (perpetual youth). Since individuals have the same type y over their lifetime, they will end up marrying a spouse of the same type y each time they go to the marriage market.

Since females are always on the "short side" of the market, they will always marry. The stock of fertile females is $1/\tau^f$, and they are all married. Hence at the end of each period, k/τ^f fertile females divorce. A fraction $1 - \tau^f$ of these females enter the marriage market next period. Hence the measure of females that enter the marriage market in each period is given by

$$M^f = 1 + k \frac{1 - \tau^f}{\tau^f} \quad (1)$$

The first term is the new entrants (which are all fertile). The second term is the inflow of divorced females which is equal to the stock of fertile females times the divorce rate times the probability that they "survive" the fertility shock in that period.

The least attractive males may stay single. Let y^c be the cut-off point; a male with a type $y < y^c$ never marries, while those above always marry. The stock of fertile males in the economy is $1/\tau^m$. Of these, a fraction $(1 - F(y^c))$ is married. The flow of fertile males that enter the market after divorce is thus $k(1 - F(y^c))(1 - \tau^m)/\tau^m$. The flow of entering males with type above y^c is $(1 - F(y^c))$. Hence the stock of males of type above y^c in the marriage market is

$$M^m(y^c) = (1 - F(y^c)) \left(1 + k \frac{1 - \tau^m}{\tau^m} \right) \quad (2)$$

The equilibrium condition is that $M^m(y^c) = M^f$. It is convenient to write $\hat{\tau}^j =$

$\frac{1-\tau^j}{\tau^j}$, $j \in \{m, f\}$. Then we can write

$$F(y^c) = \frac{\hat{\tau}^m - \hat{\tau}^f}{1 + \hat{\tau}^m k} k \quad (3)$$

which uniquely determines y^c .¹ Note that if $k = 0$ or $\tau^f = \tau^m$, $F(y^c) = 0$, that is, all men marry. If $\tau^f > \tau^m$, then $F(y^c) > 0$. The higher is k , the higher is y^c *cet. par.* The mechanism behind this is that as the divorce rates increase, the ratio of fertile unmatched men to fertile unmatched women increases.

The distribution of married men is thus given by $\tilde{F}(y) = \frac{F(y) - F(y^c)}{1 - F(y^c)}$ for $y \geq y^c$. A woman of type y^f , with relative position $F(y^f)$, is married to a man of type y^m with the same relative position in the \tilde{F} -distribution. Hence the types (y^f, y^m) in any couple satisfies $\tilde{F}(y^m) = F(y^f)$, or

$$F(y^m) = F(y^f) + (1 - F(y^f))F(y^c) \quad (5)$$

Asymmetric preferences

Suppose individuals have two characteristics x (appearance) and y (income) that are independent and uniformly distributed on $[0, 1]$. The underlying assumption is that the ranking of a females' appearance and her income in the population of females count equally much for the males' overall ranking of females as potential spouses. With different distributional assumptions, the results will change quantitatively but not qualitatively. Suppose males rank females according to their average score of x and y , $z = (x + y)/2$, while females rank males according to y only. The cumulative distribution function is given by $F^z(z) = 2z^2$ for $0 \leq z < 1/2$ and $F^z(z) = 1 - 2(1 - z)^2$ for $1/2 \leq z \leq 1$. For any z , the expected income potential is equal to $E[y|z] = z$.² The cumulative distribution function of y is simply $F^y(y) = y$. Note that $F^z < F^y$ for values below $1/2$ while $F^z > F^y$ for values above $1/2$. The cumulative distribution functions are shown in Figure 1.

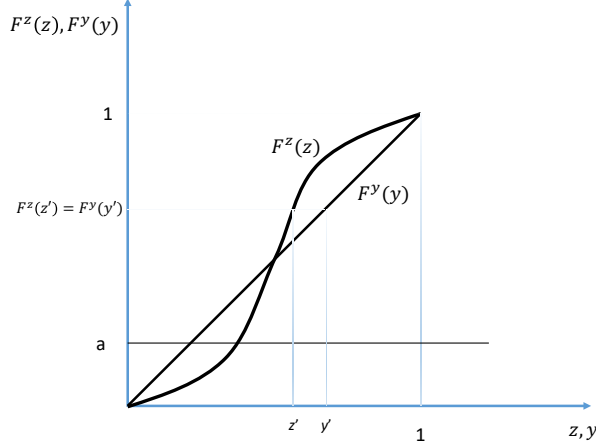
Matching is assortative, so that the most attractive female mates with the most attractive male and so on. There are equally many males and females. Assortative matching then implies that for any pair (y', z') , $F^y(y') = F^z(z')$. Hence it follows

¹Written out, it follows that

$$F(y^c) = k \frac{\tau^f - \tau^m}{\tau^f \tau^m - k(1 - \tau^m) \tau^f} \quad (4)$$

²Due to symmetry, $E[y|z] = E[x|z]$. Furthermore, $E[(x + y)/2|z] = z$. It follows that $E[y|z] = z$.

Figure 1: Gender differences in mating preferences and hypergamy



Note: The figure illustrates matching in the mating market in the model where women have preferences for potential partners' earnings potential y only, whereas men have preferences over $z = z = (x + y)/2$, where x is a factor that comes in addition to income y . The figure illustrates one match between a woman with index z' and a man with index y' . As discussed in the main text, a is a fraction of lower-ranked individuals that do not mate.

that $z' < y'$ for $y' \in (1/2, 1)$ and that $z' > y'$ for $y' \in (a, 1/2)$, see Figure 2 for an illustration.³ In the figure, agents of type less than a don't marry. In that case it follows that on average women mates up for $z > 1/2$ and down for $z < 1/2$.

Suppose now that a woman of type y^f never accepts to marry a male if his productivity is below $\kappa y^f \in (1/2, 1)$, where κ is a constant. We assume that $\kappa \in (1/2, 1)$, while a is set to zero. As above, matching is assortative, in the sense that a male marries the female of the highest type z that accepts him. Females marry the male with the highest type y that accepts him, with the additional requirement that his productivity exceeds a fraction κ of her own productivity.

Among those who marry, write the couples as $(z, y^m(z))$. At the top of the distribution, the marriage pattern is as without the constraint. However, at some point $z = \bar{z}$, the participation constraint of females starts to bind. Hence for $z \geq \bar{z}$, the marriage pattern is given by $1 - F(z) = 1 - y^m(z)$, or $2(1 - z)^2 = 1 - y^m(z)$ which gives $y^m(z) = 1 - 2(1 - z)^2$. The first female that rejects a man has productivity $y^f = 1$, and thus rejects a man with productivity κ . It follows that \bar{z} is given by

$$\bar{z} = 1 - \sqrt{(1 - \kappa)/2}$$

³More precisely, for $y' < 1/2$, $z' = \sqrt{y'}/2$. For $y' > 1/2$, $z' = 1 - \sqrt{(1 - y')}/2$.

On an interval below \bar{z} (stretching at least to $z = 1/2$), the most productive females choose not to marry. For $z \leq \bar{z}$, the probability that a woman does not marry, $\pi(z)$, is equal to

$$\pi(z) = P[y^f \geq \kappa y^m(z) | \frac{x^f + y^f}{2} = z]$$

We have that for $z \geq 1/2$, $y^f|z$ is uniformly distributed on $[2z - 1, 1]$. It follows that at this interval,

$$\pi(z) = \frac{1 - \kappa y^m(z)}{2(1 - z)}$$

For $z \in \{1/2, \bar{z}\}$, it follows that

$$4(1 - z)(1 - \pi(z))dz = y^m \iota(z) dz$$

On a small interval dz , the left-hand side shows the number of women in that interval that marry (proportional to the probability density times the propensity to marry at this interval), while the right-hand side shows the number of males that marry in a corresponding interval. Written out, the equation reads

$$y^m \iota(z) = 2 - 2\kappa y^m(z) \quad (6)$$

This is an ordinary first order differential equation with a well defined solution. The solution is given by

$$y^m(z) = C_1 e^{-2\kappa z} + 1/\kappa \quad (7)$$

To find C , we use the initial condition that $y(\bar{z}) = \kappa$, which gives

$$C_1 = -\left(\frac{1}{\kappa} - \kappa\right) e^{2\kappa\left(1 - \sqrt{\frac{1-\kappa}{2}}\right)} \quad (8)$$

Consider then the situation with $z < 1/2$. In this case, $y|z$ is uniformly distributed on $[0, 2z]$. Since some females don't marry, there must exist a cut-off y^c below which men don't marry because they don't find a spouse. It follows that for $z < y^c \kappa$, all females marry. Above y^c , $\pi(z) = \frac{2z - \kappa y^m(z)}{2z}$. Recall that $F(z) = 2z^2$ on this interval. It follows that

$$4z(1 - \pi(z))dz = y^m \iota(z) dz$$

Or, written out, at the interval were the lower bound on π does not bind;

$$y^m(z) = 2y^m(z) \quad (9)$$

Which also has a closed form solution. It follows that

$$y^m(z) = C_2 e^{2\kappa z} \quad (10)$$

To find C_2 , we utilize that $y^m(z)$ is continuous at $1/2$. C is determined so that $C_2 e^\kappa = C_1 e^{-\kappa} + 1/\kappa$.

Finally, define z^0 implicitly by the equation $y^m(z^0) = 2z^0$. At z^0 and below, all females marry. Define $y^0 = y^m(z^0)$. Below z^0 , the marriage pattern is defined by the equation

$$F(z^0) - F(z) = y^0 - y \quad (11)$$

To summarize, the equilibrium marriage pattern $\{z, y^m(z)\}$ has the following properties:

1. For $z \geq \bar{z} = 1 - \sqrt{(1 - \kappa)/2}$, all females marry, and $y^m(z) = 1 - 2(1 - z)^2$.
2. For $z \in [1/2, \bar{z}]$, $y^m(z)$ is given by (??) and the end condition that $y(\bar{z}) = \kappa\bar{z}$.
3. For $z \in [z^0, 1/2]$, $y^m(z)$ is given by (??) and the end condition that $\lim_{z \rightarrow 1/2^-} y^m(z) = y^m(1/2)$.
4. For $z \in [0, z^0]$, $y^m(z)$ is given by (??)

Appendix B: First stage results

In Tables 2 and 3 of the main paper, own earnings rank is instrumented with parental earnings rank. We here present the first stages of these regressions. When we use the quadratic model we use both parental rank and parental rank squared as instruments for own rank and own rank squared. For the quadratic model we therefore present several first stages. In Table B1 we present the first stages for the results in Table 2. As the sample in tables 2 and 3 are different we also present the corresponding first stages for Table 3 in Table B2

Table B1: First stages for the IV results in Table 2

	Linear model		Quadratic model			
	(1) Own rank men	(2) Women	(3) Own rank men	(4) Women	(5) Own rank sq. Men	(6) Women
Parental rank	19.3*** (0.11)	15.2*** (0.12)	23.7*** (0.46)	24.8*** (0.47)	1565.8*** (46.2)	1910.8*** (47.7)
Parental rank sq			-0.044*** (0.0044)	-0.095*** (0.0045)	4.81*** (0.46)	-2.98*** (0.47)
F-value	28499.25	16571.34	14412.48	8635.84	15081.27	8955.99
N	757868	723317	757868	723317	757868	723317

Notes: Estimates and standard errors are multiplied by 100, such that they are measured in percentage points. Robust standard errors in parentheses. ***/*** indicates statistical significant at the 10/5/1 percent level.

Table B2: First stages for the IV results in Table 3

	Linear model		Quadratic model			
	(1) Own rank men	(2) Women	(3) Own rank men	(4) Women	(5) Own rank sq. Men	(6) Women
Parental rank	16.4*** (0.22)	11.5*** (0.22)	18.9*** (0.46)	16.9*** (0.86)	1481.8*** (92.0)	1346.2*** (87.0)
Parental rank sq			-0.024*** (0.0083)	-0.005*** (0.0083)	3.17*** (0.90)	-1.26*** (0.86)
F-value	5852.06	2880.04	2934.47	1475.83	3062.21	1549.70
N	200074	202449	200074	202449	200074	202449

Notes: Estimates and standard errors are multiplied by 100, such that they are measured in percentage points. Robust standard errors in parentheses. **** indicates statistical significant at the 10/5/1 percent level.

Appendix C: Robustness

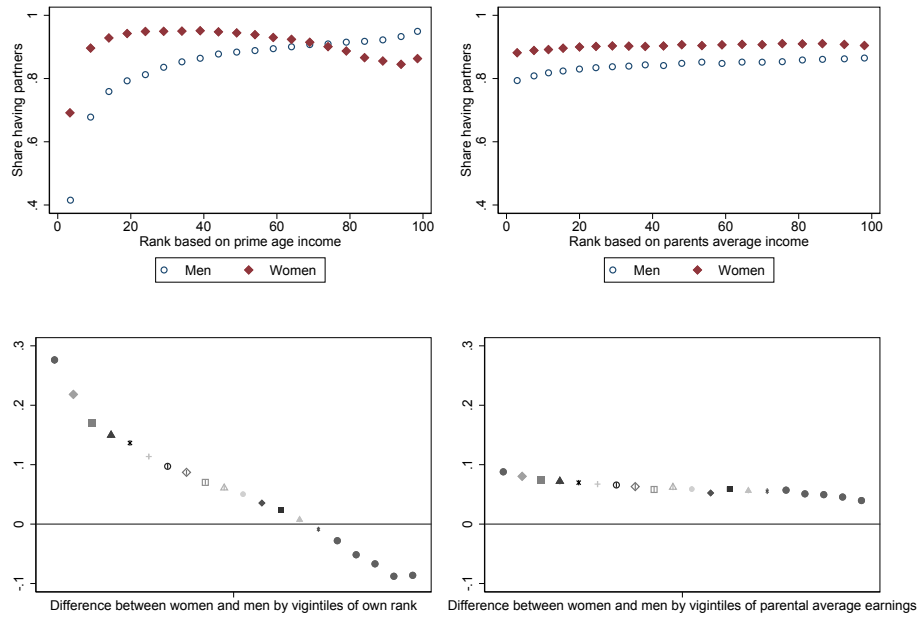


Figure C1: Probability of having found a partner by 2015. By own or parental average earnings

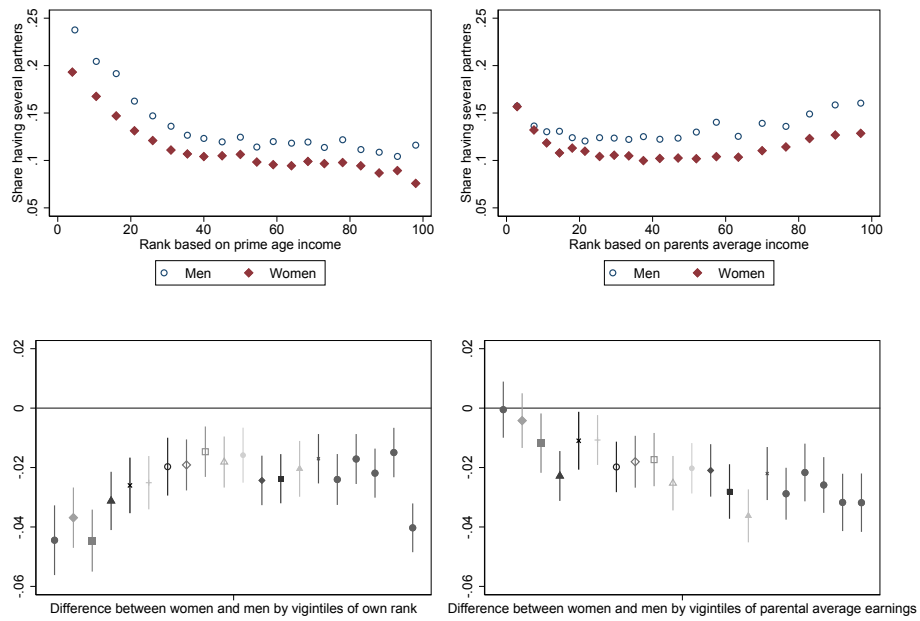


Figure C2: Probability of having had multiple partners by 2015. By own and parental average earnings

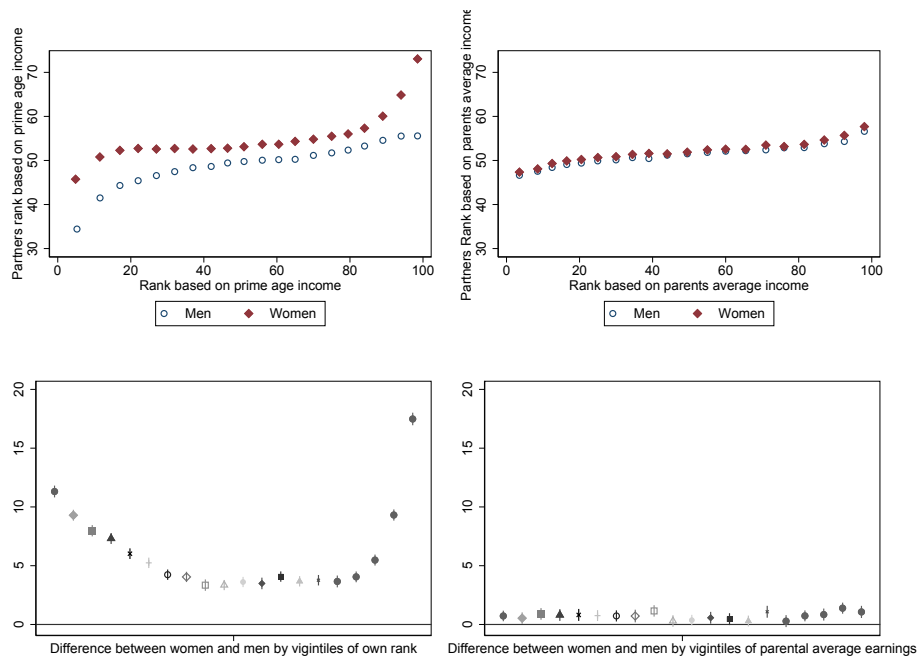


Figure C3: Average partner rank by own rank. By own or parental average earnings

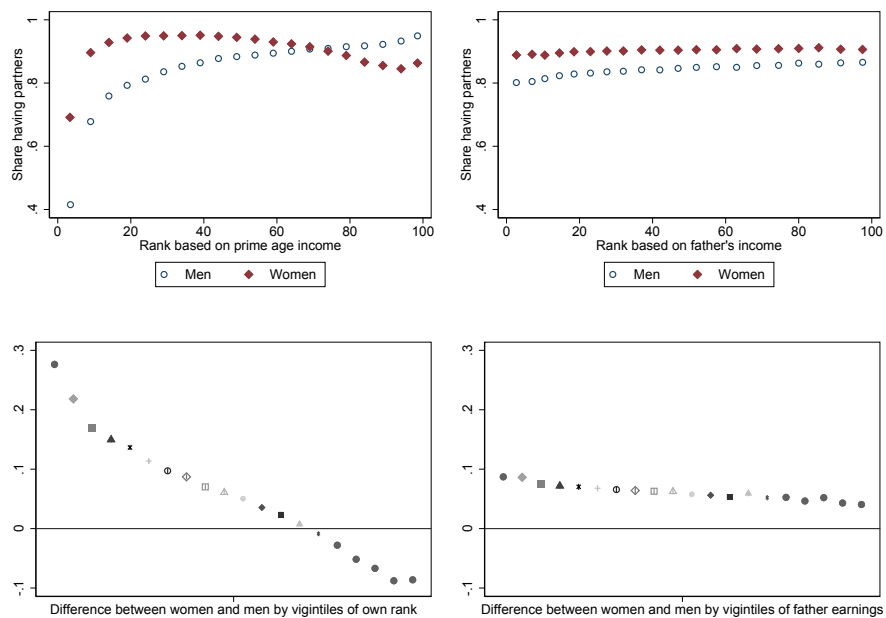


Figure C4: Probability of having found a partner by 2015. By own or father earnings

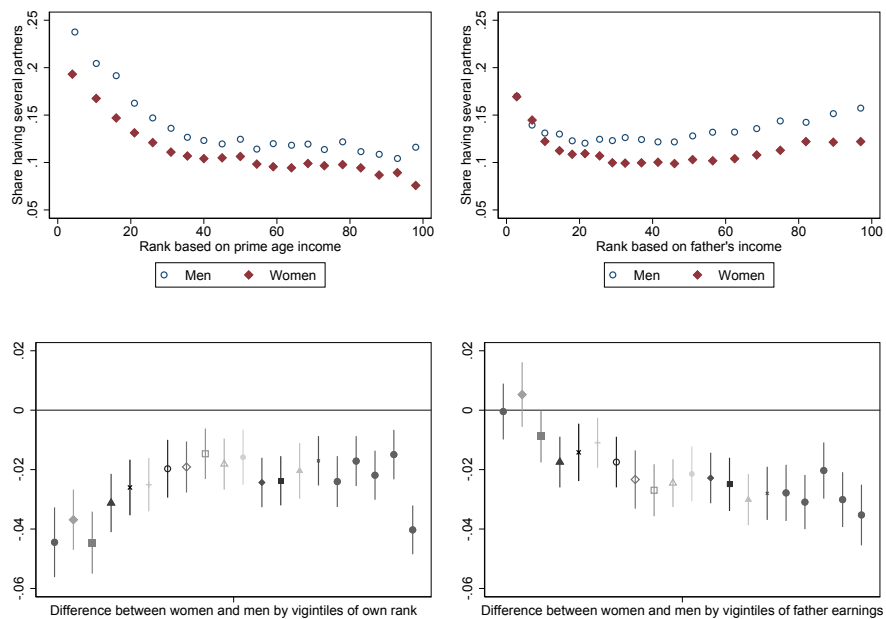


Figure C5: Probability of having had multiple partners by 2015. By own and father earnings

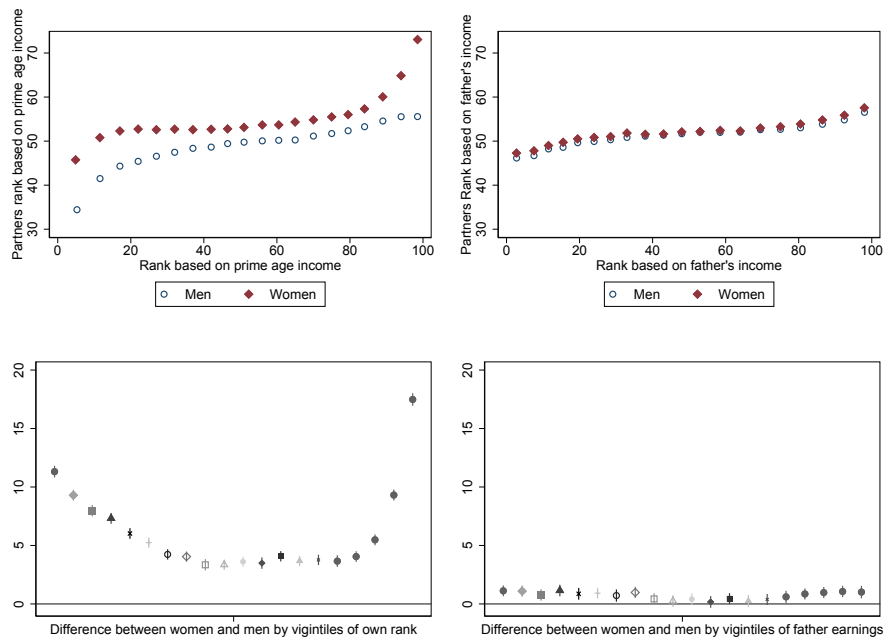


Figure C6: Average partner rank by own rank. By own or father earnings

Table C1: Gender difference in partnering. Instrumental variables (IV) estimates with ranks based on average parental incomes.

	(1) Men	(2) Women	(3) Gender diff.	(4) Men	(5) Women	(6) Gender diff.
Own rank	0.46*** (0.0092)	0.23*** (0.0099)	0.20*** (0.014)	0.89*** (0.084)	1.00*** (0.10)	-0.087 (0.013)
Own rank squared				-0.0040*** (0.00078)	-0.0072*** (0.00097)	0.0027** (0.0013)
Mean of outcome	0.84	0.90	0.87	0.84	0.90	0.87
N	757,868	723,317		757,868	723,317	

Notes: Own earnings rank is instrumented with average parental earnings rank. Estimates and standard errors are multiplied by 100, such that they are measured in percentage points. The gender differences in columns (3) and (6) are evaluated within a joint model with gender interactions on all variables. Robust standard errors in parentheses. */**/***/ indicates statistical significant at the 10/5/1 percent level.

Table C2: Gender difference in partnering. Instrumental variables (IV) estimates with ranks based on fathers' incomes.

	(1) Men	(2) Women	(3) Gender diff.	(4) Men	(5) Women	(6) Gender diff.
Own rank	0.40*** (0.010)	0.19*** (0.013)	0.21*** (0.016)	0.58*** (0.095)	0.50*** (0.13)	0.077 (0.16)
Own rank squared				-0.0016* (0.00086)	-0.0028** (0.0012)	0.0012 (0.0014)
Mean of outcome	0.84	0.90	0.87	0.84	0.90	0.87
N	600456	572503		600456	572503	

Notes: Own earnings rank is instrumented with father earnings rank. Estimates and standard errors are multiplied by 100, such that they are measured in percentage points. The gender differences in columns (3) and (6) are evaluated within a joint model with gender interactions on all variables. Robust standard errors in parentheses. */**/***/ indicates statistical significant at the 10/5/1 percent level.

Table C3: Gender difference in multiple partnerships. Instrumental variables (IV) estimates with ranks based on average parental incomes.

	(1) Men	(2) Women	(3) Gender diff.	(4) Men	(5) Women	(6) Gender diff.
Own rank	0.16*** (0.022)	-0.080*** (0.029)	0.23*** (0.036)	-3.7*** (0.0033)	-5.9*** (0.0059)	2.3*** (0.68)
Own rank squared				0.034*** (0.0030)	0.053*** (0.0055)	-0.0020*** (0.0063)
Mean of outcome	0.13	0.11	0.12	0.13	0.11	0.12
N	200,074	202,449		200,074	202,449	

Notes: Own earnings rank is instrumented with ranks based on average parental incomes, the table is otherwise identical to table 3 in the main paper. Estimates and standard errors are multiplied by 100, such that they are measured in percentage points. The gender differences in columns (3) and (6) are evaluated within a joint model with gender interactions on all variables. Robust standard errors in parentheses. */**/***/ indicates statistical significant at the 10/5/1 percent level.

Table C4: Gender difference in multiple partnerships. Instrumental variables (IV) estimates with ranks based on fathers' incomes.

	(1)	(2)	(3)	(4)	(5)	(6)
	Men	Women	Gender diff.	Men	Women	Gender diff.
Own rank	0.066*** (0.025)	-0.12*** (0.038)	0.18*** (0.046)	-3.1*** (0.37)	-5.5*** (0.86)	2.4** (0.94)
Own rank squared				0.027*** (0.0033)	0.048*** (0.0079)	-0.021** (0.0085)
Mean of outcome	0.14	0.12	0.13	0.14	0.12	0.13
N	116,639	117,127		116,639	117,127	

Notes: Own earnings rank is instrumented ranks based on fathers' incomes, the table is otherwise identical to table 3 in the main paper. Estimates and standard errors are multiplied by 100, such that they are measured in percentage points. The gender differences in columns (3) and (6) are evaluated within a joint model with gender interactions on all variables. Robust standard errors in parentheses. */**/** indicates statistical significant at the 10/5/1 percent level.

Table C5: Gender difference in partner's parental ranks. Ordinary least squares (OLS) estimates with ranks based on average parental incomes.

	(1)	(2)	(3)	(4)
	Partner rank	Partner with higher rank	Partner rank	Partner with higher rank
Own rank	0.16*** (0.0014)	-0.50*** (0.0019)	0.079*** (0.0014)	-0.54*** (0.0019)
Female	6.18*** (0.12)	25*** (0.17)	0.66*** (0.11)	7.4*** (0.16)
Female*Own rank	-0.0061*** (0.0020)	-0.31*** (0.0026)	0.0014 (0.0019)	-0.053*** (0.0027)
N	1,065,534	1,242,148	1,058,692	1,237,577

Notes: Columns 3 and 4 use average parental income as basis for the parental ranks, the table is otherwise identical as Table 4 in the main paper. For the dichotomous outcome in columns (2) and (4), the estimates and standard errors are multiplied by 100, such that they are measured in percentage points. The regressions are based on the 1952-1975 birth cohorts. All regressions control for year of birth fixed effects. Robust standard errors in parentheses. */**/** indicates statistical significant at the 10/5/1 percent level.

Table C6: Gender difference in partner's parental ranks. Ordinary least squares (OLS) estimates with ranks based on fathers' incomes.

	(1)	(2)	(3)	(4)
	Partner rank	Partner with higher rank	Partner rank	Partner with higher rank
Own rank	0.16*** (0.0014)	-0.50*** (0.0019)	0.078*** (0.0015)	-0.46*** (0.0021)
Female	6.18*** (0.12)	25*** (0.17)	0.59*** (0.13)	9.4*** (0.19)
Female*Own rank	-0.0061*** (0.0020)	-0.31*** (0.0026)	0.0012 (0.0022)	-0.083*** (0.0031)
N	1,065,534	1,242,148	855,899	996,840

Notes: Columns 3 and 4 use father income as basis for the parental ranks, the table is otherwise identical as Table 4 in the main paper. For the dichotomous outcome in columns (2) and (4), the estimates and standard errors are multiplied by 100, such that they are measured in percentage points. The regressions are based on the 1952-1975 birth cohorts. All regressions control for year of birth fixed effects. Robust standard errors in parentheses. */**/** indicates statistical significant at the 10/5/1 percent level.

Appendix D: Vignette Survey Experiment

We here provide descriptive statistics of our vignette survey experiment sample as well as screenshots for our survey experiment.

	Age	Share female	Level of education
Mean	47.84	0.58	3.45
Standard dev	16.30	0.49	1.21
p25	34		2
p50	48		4
p75	61		4
p90	70		5
p95	74		5
p99	79		5

Note: The table shows the descriptive statistics for the sample of our vignette survey experiment. Level of is given in level 1-5 where 1 is the lowest level (compulsory schooling or lower), 2 is high school, 3 is vocational training, 4 is college degree whereas 5 is higher university education (master's or PhD level).

Screenshots from survey experiment

For kvinner (two treatments, a) or b)):

Enten:

- a) Tenk deg at (Emma/Anne/Inger/Anna)¹ er singel og leter ett fast langvarig forhold. Hun møter en mann som er snill og hensynsfull, som tjener godt, og som hun synes er pen og attraktiv.

Hvor sannsynlig tror du at det er at hun er interessert i et langvarig forhold med kvinnen?

1 2 3 4 5 6 7 8 9 10
|-----|-----|-----|-----|-----|-----|-----|-----|-----|

Helt usannsynlig

Helt sannsynlig

Eller:

- b) Tenk deg at (Emma/Anne/Inger/Anna) er singel og leter ett fast langvarig forhold. Hun møter en mann som er snill og hensynsfull, som ikke tjener så godt, men som hun synes er pen og attraktiv.

Hvor sannsynlig tror du at det er at hun er interessert i et langvarig forhold med kvinnen?

1 2 3 4 5 6 7 8 9 10
|-----|-----|-----|-----|-----|-----|-----|-----|-----|

Helt usannsynlig

Helt sannsynlig

For menn (to treatments a) og b)):

Enten:

- a) Tenk deg at (Markus/Jan/Arne/Per)² er singel og leter ett fast langvarig forhold. Han møter en kvinne som er snill og hensynsfull, som tjener godt, og som han synes er pen og attraktiv.

Hvor sannsynlig tror du at det er at han er interessert i et langvarig forhold med mannen?

1 2 3 4 5 6 7 8 9 10
|-----|-----|-----|-----|-----|-----|-----|-----|-----|

Helt usannsynlig

Helt sannsynlig

¹ Emma: mest populære jentenavn i 2005, Anne, Inger og Anna er de tre mest populære jentenavnene fra 1900 til 1999. Forslaget er å randomisere hvilket av dem vi bruker.

² Markus: mest populære guttenavn i 2005, Jan, Arne og Per er de tre mest populære guttenavnene fra 1900 til 1999.

- 1: Min partner tjente mye mer
- 2: Min partner tjente mer
- 3: Vi tjente like mye
- 4: Jeg tjente mer
- 5: Jeg tjente mye mer

Totalt sett, hvem gjør mest av husholdsarbeidet i hjemmet?

- 1: Min partner gjør mye mer
- 2: Min partner gjør mer
- 3: Vi gjør like mye
- 4: Jeg gjør mer
- 5: Jeg gjør mye mer

Hvis du tenker tilbake til tiden da dere begynte å bo sammen, hvordan fordelte dere hussarbeidet da?

- 1: Min partner gjorde mye mer
- 2: Min partner gjorde mer
- 3: Vi delte det likt
- 4: Jeg gjorde mer
- 5: Jeg gjorde mye mer

Hvem jobber flest timer utenfor hjemmet?

- 1: Min partner jobber mye mer utenfor hjemmet
- 2: Min partner jobber mer utenfor hjemmet
- 3: Vi jobber like mye utenfor hjemmet
- 4: Jeg jobber mer utenfor hjemmet
- 5: Jeg jobber mye mer utenfor hjemmet

Hvis du tenker tilbake til tiden da dere begynte å bo sammen, hvem jobbet flest timer da?

- 1: Min partner jobbet mye mer utenfor hjemmet

2: Min partner jobbet mer utenfor hjemmet

3: Vi jobbet like mye utenfor hjemmet

4: Jeg jobbet mer utenfor hjemmet

5: Jeg jobbet mye mer utenfor hjemmet

Har du barn?

Hvis ja:

Bor du sammen med den du fikk ditt første barn med?