Online Appendix for Changes in Assortative Matching and Educational Inequality: Evidence from Marriage and Birth Records in Mexico

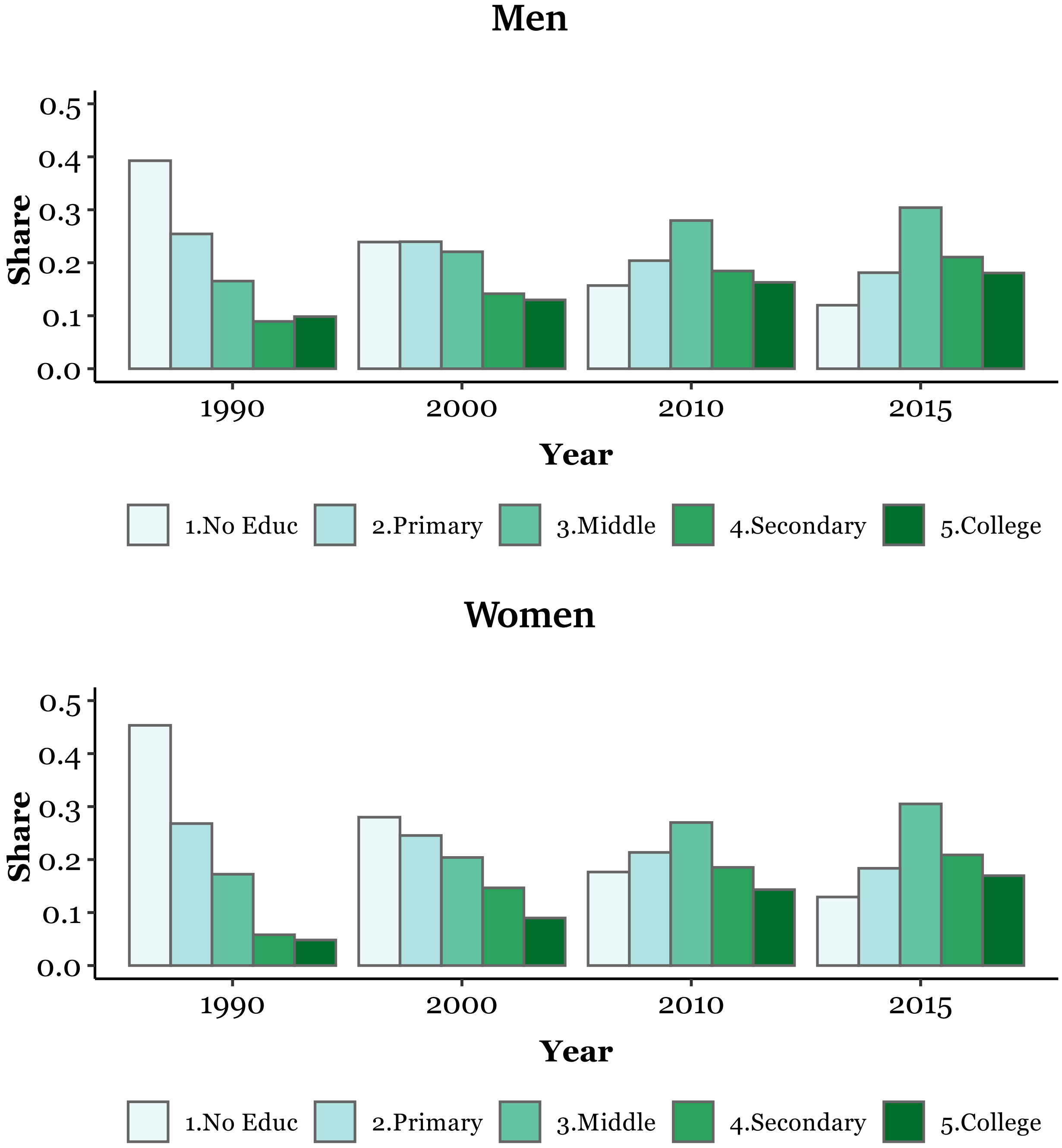
Lauren Hoehn-Velasco and Jacob Penglase

Contents

A Appendix

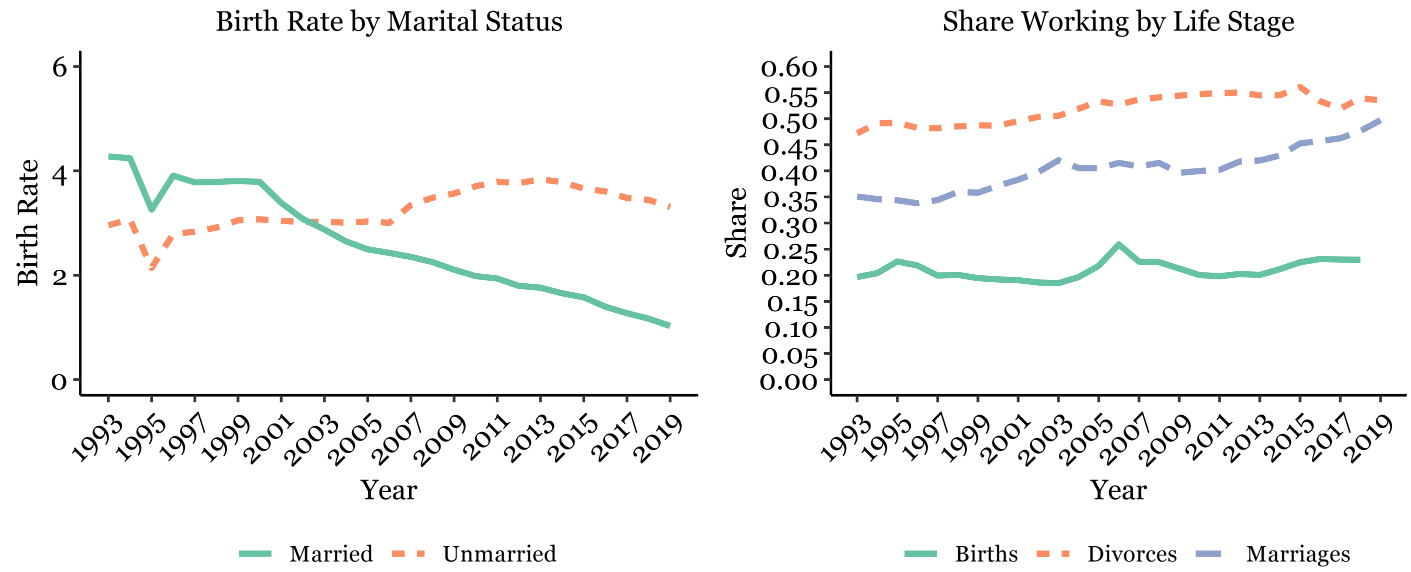
A.1 Additional Figures

Figure A1: Educational Attainment, Ages 25-54



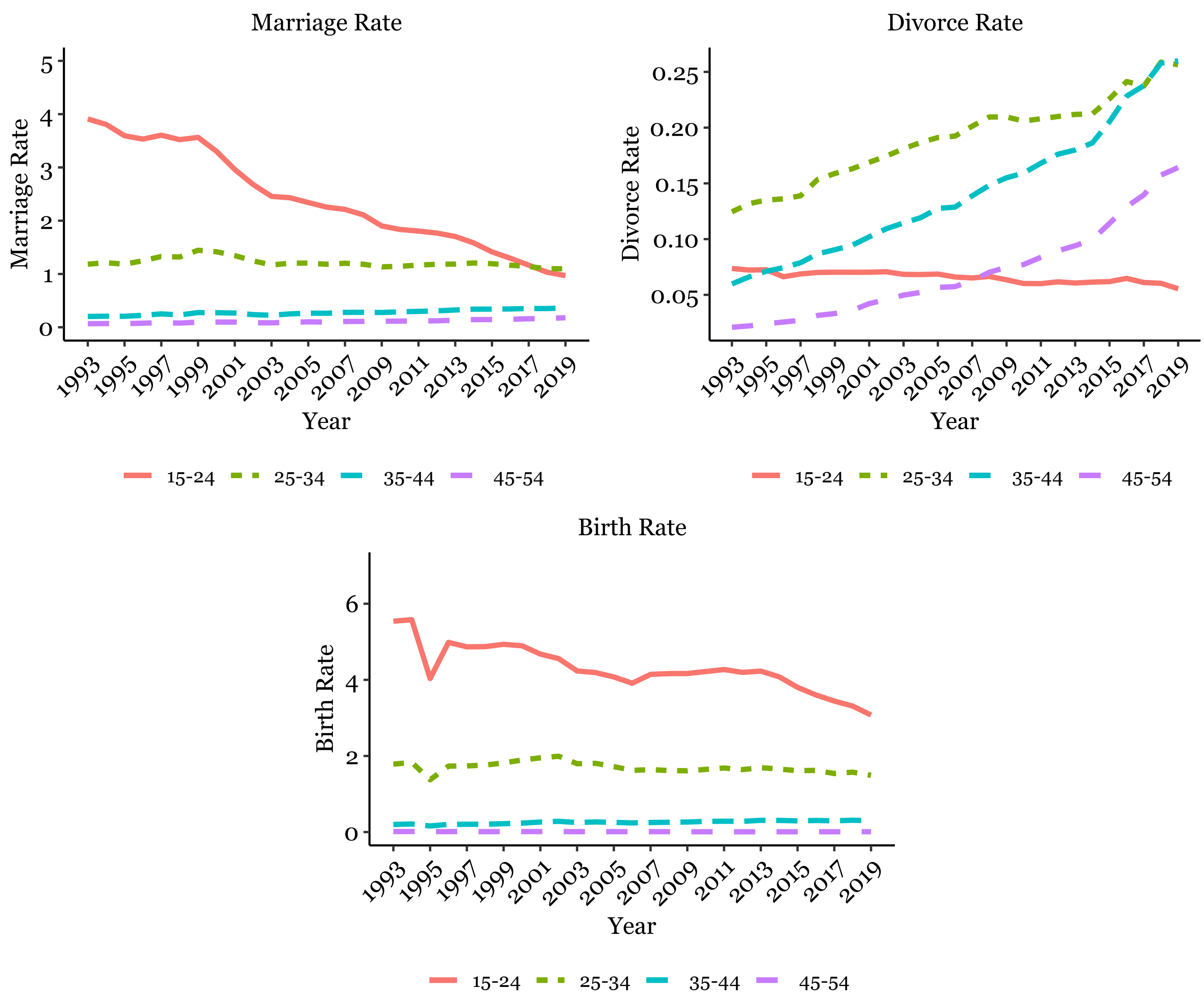
Source: Mexican IPUMS data.
Notes: Less than primary education is less than 6th grade. Primary education is between 6th and 8th grade. Middle school education is 9th grade to 11th grade. Secondary education is 12th grade to 15 years of education. College is greater than 15 years but not missing education. Population includes those between 25 and 54.

Figure A2: Births by Marital Status and Work Status during Life Stages



Sources: INEGI marriage, divorce, and birth statistics. Mexican IPUMS data. Notes: Rates are per 1,000 persons between 15 and 54.

Figure A3: Divorce, Marriage, and Birth Rates by Age

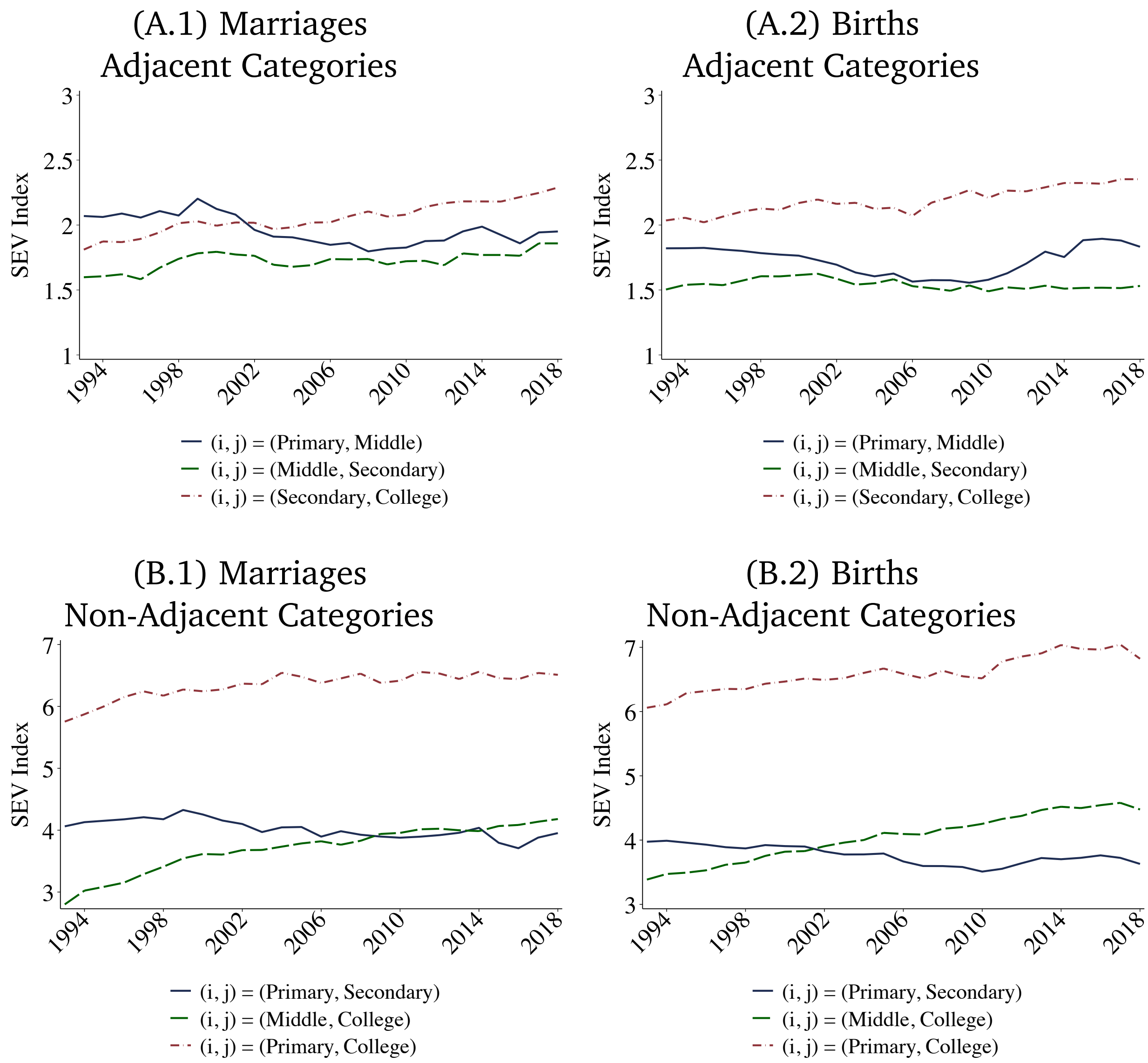


Sources: INEGI marriage, divorce, and birth statistics. Mexican IPUMS data. Notes: Rates are per 1,000 persons between 15 and 54. Less than primary education is either *sin escolaridad*, or education *1 a 3 años* and *4 a 5 años*. Primary education is *primaria completa*. Middle school education is *secundaria*. Secondary education is *preparatoria*. College is greater *professional*. Technical education is grouped with secondary.

Figure A4: Age of Matching Over Time (Marriages and First Births)

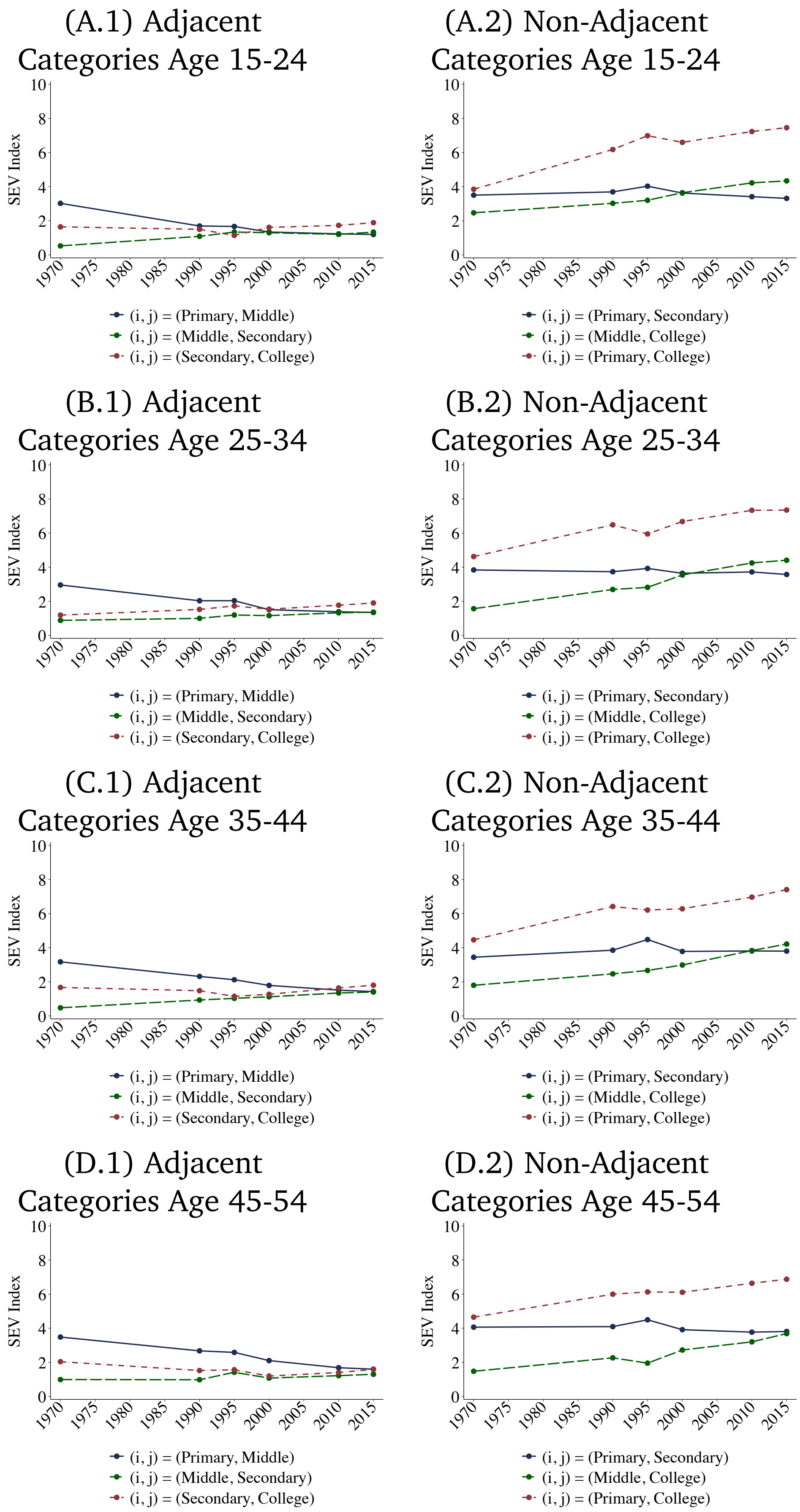
Notes: Vital Statistics Marriage and Birth Records. The sample includes marriages and first births where both spouses or parents are age 15 to 50.

Figure A5: Assortative Matching: Adults Age 25 to 54



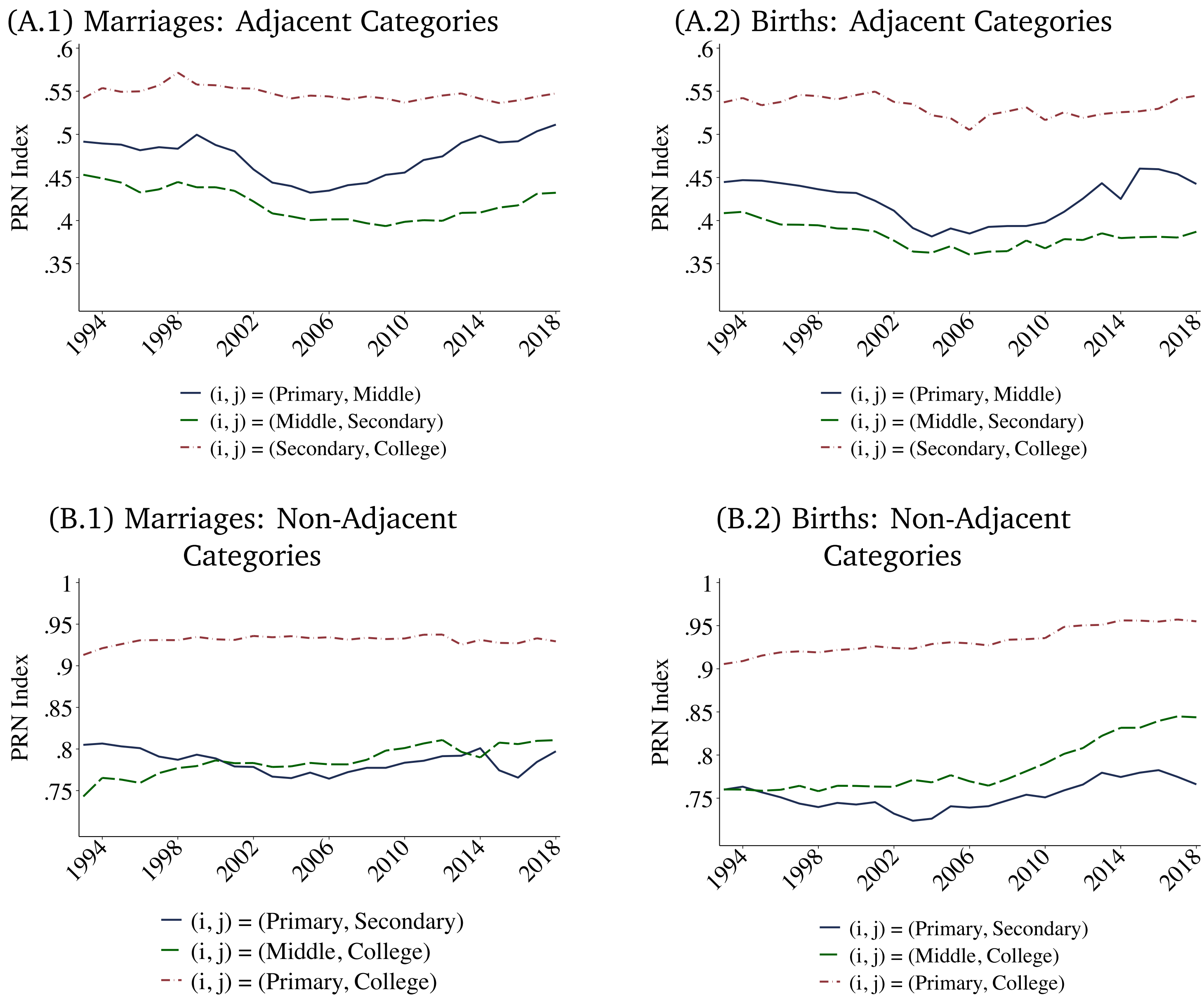
Notes: INEIGI Data. Men and women are divided into four mutually exclusive education categories: 1. Primary or Less, 2. Middle, 3. Secondary, 4. College. Each figure plots assortative matching for the diagonal 2×2 sub-matrices of the full sorting matrix using the Separable Extreme Value index. We restrict the sample in Figures A.2, B.2, and C.2 to marriages where at least one spouse or parent is between 25 and 54.

Figure A6: Assortative Marriage using Census Data



Notes: IPUMS Census Data. Men and women are divided into four mutually exclusive education categories: 1. Primary or Less, 2. Middle, 3. Secondary, 4. College. Each figure plots assortative matching for the diagonal 2×2 sub-matrices of the full sorting matrix using the Separable Extreme Value index. We divide couples into four age groups based on the age of the wife.

Figure A7: Assortative Marriage and Parental Matching (Perfect-Random Normalization)



Notes: Vital Statistics Marriage and Birth Records. Men and women are divided into four mutually exclusive education categories: 1. Primary or Less, 2. Middle, 3. Secondary, 4. College. In Panels A and B, each figure plots assortative matching for the diagonal 2×2 sub-matrices of the full sorting matrix using the Perfect-Random Normalization. Panel A plots adjacent education categories while Panel B plots non-adjacent categories. In Panel C, each line is a comparison of education i with the combined remaining three education categories. The weighted average curve is computed by averaging the assortative index across educational levels, where the weights are determined by diagonal value of the matching table given in Table A3.

A.2 Additional Tables

Table A1: Observed Marital Matching (1993 and 2018)

		Wife's Education			
		College	Secondary	Middle	Primary or Less
Panel A: 1993					
Husband's Education	College	7.2%	2.9%	2.9%	0.8%
	Secondary	1.9%	6.7%	6.1%	2.1%
	Middle	1.4%	3.5%	19.1%	9.8%
	Primary or Less	0.5%	1.2%	7.1%	26.8%
				Sum of Diagonal:	*59.8%*
Panel B: 2018					
Husband's Education	College	21.3%	5.3%	1.8%	0.4%
	Secondary	5.5%	16.0%	6.2%	1.1%
	Middle	2.5%	6.8%	16.7%	3.2%
	Primary or Less	0.5%	1.6%	4.4%	6.4%
				Sum of Diagonal:	*60.4%*

Source: Mexican INEGI data.
Notes: Primary or less is defined as less than 8 years of schooling or less. Middle is 9th to 11th grade. Secondary is 12 to 15 years of education. The sample includes marriages where at least one spouse is age 15 to 54.

Table A2: Observed Parental Matching (1993 and 2018)

		Mother's Education			
		College	Secondary	Middle	Primary or Less
Panel A: 1993					
Father's Education	College	5.8%	2.7%	2.8%	0.8%
	Secondary	1.5%	5.6%	6.2%	2.4%
	Middle	1.1%	3.3%	16.8%	10.3%
	Primary or Less	0.5%	1.3%	8.1%	30.5%
				Sum of Diagonal: 58.8%	
Panel B: 2018					
Father's Education	College	12.2%	3.9%	1.3%	0.2%
	Secondary	4.5%	15.0%	7.4%	1.3%
	Middle	2.3%	9.3%	21.1%	4.6%
	Primary or Less	0.5%	2.4%	6.1%	7.8%
				Sum of Diagonal: 56.2%	

Source: Mexican INEGI data.
Notes: Primary or less is defined as less than 8 years of schooling or less. Middle is 9th to 11th grade. Secondary is 12 to 15 years of education. The sample includes births where at least one parent is age 15 to 54. Births are limited to first births.

Table A3: Assortative Matching in a Four-Education Market

	College	Secondary	Middle School	Primary or Less
College	r_1	a	c	$n_1 - r_1 - a - c$
Secondary	b	r_2	d	$n_2 - b - r_2 - d$
Middle School	f	e	r_3	$n_3 - f - e - r_3$
Primary or Less	$m_1 - r_1 - b - f$	$m_2 - a - r_2 - e$	$m_3 - c - d - r_3$	$1 + r_1 + r_2 + r_3 + a + b + c + d + e + f - (m_1 + m_2 + m_3 + n_1 + n_2 + n_3)$

Notes: In the above tables, m_j are the shares of men who have graduated college, secondary school, and middle school, respectively. n_j gives the corresponding values for women. r_1 denotes the share of marriages where both spouses have a college degree, r_2 denotes the share of marriages where both spouses have a secondary school degree, and r_3 is the share where both have a middle school education. a, b, c, d, e, and f denote the shares of couples with different pairs of unequal education levels.

Table A4: Hypothetical SEV Index Values for Different Sorting Matrices

College-College Marriages (r) (1)	Share Men with a College Degree (m) (2)	Share Women with a College Degree (n) (3)	SEV Index $\ln(\frac{r(1+r-m-n)}{(m-r)(n-r)})$ (4)
0.00	0.40	0.40	$-\infty$
0.02	0.40	0.40	-3.49
0.04	0.40	0.40	-2.60
0.06	0.40	0.40	-2.00
0.08	0.40	0.40	-1.52
0.10	0.40	0.40	-1.10
0.12	0.40	0.40	-0.71
0.14	0.40	0.40	-0.35
0.16	0.40	0.40	0.00
0.18	0.40	0.40	0.35
0.20	0.40	0.40	0.69
0.22	0.40	0.40	1.05
0.24	0.40	0.40	1.42
0.26	0.40	0.40	1.81
0.28	0.40	0.40	2.23
0.30	0.40	0.40	2.71
0.32	0.40	0.40	3.26
0.34	0.40	0.40	3.93
0.36	0.40	0.40	4.84
0.38	0.40	0.40	6.31
0.40	0.40	0.40	∞

Notes: We illustrate how different values of r, m, and n correspond to different magnitudes of the SEV Index. All numbers are hypothetical. Column 1 provides the share of marriages where both spouses have a college degree. Column 2 and 3 give the share of men and women with a college degree. Column 4 gives the SEV index. The parameter r is bounded below by zero and above by the minimum of m and n.

Table A5: Changes in Assortativeness 1993-2018 (Perfect-Random Normalization)

	Marriages			Births		
	1993	2018	Difference	1993	2018	Difference
Panel A: 2 by 2 Comparisons	(1)	(2)	(3)	(4)	(5)	(6)
(i,j) = (Primary, Middle)	0.492	0.511	0.020	0.445	0.442	-0.002
(i,j) = (Middle, Secondary)	0.453	0.432	-0.021	0.409	0.387	-0.022
(i,j) = (Secondary, College)	0.542	0.547	0.006	0.537	0.545	0.008
(i,j) = (Primary, Secondary)	0.805	0.797	-0.008	0.760	0.766	0.006
(i,j) = (Lower Secondary, College)	0.743	0.811	0.068	0.760	0.844	0.084
(i,j) = (Primary,, College)	0.913	0.929	0.016	0.906	0.955	0.049
Panel B: Merged Categories						
Primary	0.596	0.548	-0.047	0.566	0.485	-0.081
Middle	0.320	0.390	0.070	0.290	0.346	0.056
Secondary	0.357	0.365	0.007	0.325	0.325	0.000
College	0.595	0.628	0.034	0.601	0.622	0.021
Weighted Average	0.486	0.488	0.002	0.470	0.423	-0.047
Observations	496,358	353,423		635,126	510,713	

Notes: Vital Statistics Marriage and Birth Records. Men and women are divided into four mutually exclusive education categories: 1. Primary or Less, 2. Middle, 3. Secondary, 4. College. In Panel A, each row provides the assortative matching measure for the diagonal 2×2 sub-matrices of the full sorting matrix using the Perfect-Random Normalization. In Panel B, each line is a comparison of education category i with the combined remaining three education categories. The weighted average measure is computed by averaging the assortative index across educational levels, where the weights are determined by diagonal value of the matching table given in Table A3.

B Data

We use national administrative records for births, marriages, and divorces from the Instituto Nacional de Estadística y Geografía (INEGI). The data includes an individual record for each record throughout Mexico over 1993-2018. These characteristics provided by INEGI for each record include the geographic location and the education levels of the couple.

We primarily rely on the birth records to measure assortativeness. The advantage of the birth records is that it details the information for married couples, cohabitating couples, as well as single women. We use records for first births to women in all 31 states as well as Mexico City. This allows us to look at assortativeness for all household arrangements, including those that are not formally cohabitating and would not appear in household surveys. We also utilize the INEGI marriage and divorce records, which collects similar information on education levels. These records include information over each marriage and divorce that occurred in Mexico, along with the characteristics of the couple.

There are several data limitations that require our attention. First, several states poorly reported education levels. Four states were particularly problematic for education reporting in the vital statistics records, as they defined education differently across years and contained excessive amounts of missing data.[1] To deal with this issue, we eliminate these four states from our analyses.[2] The results are In total, we focus on 27 states and Mexico City.

Even with omitting the four states with inconsistent and incomplete data, missing education levels is still a concern. The education level of one or both spouses is at times missing, and importantly this is not likely to be random; individuals with lower education may be more likely to leave certain categories of the marriage or birth certificate blank. A related problem in the birth records is that the father's information is often missing, and this is again likely to be correlated with age and education.

To examine the extent of this problem, we begin by plotting the percentage of missing education values in the marriage and birth records by year in Figure A8. Missing data in the marriage records is given in Panel A, while Panel B presents missing data in the birth records. We separately plot records where the husband/father's education is missing, the wife/mother's education is missing, and finally when either are missing. Several patterns emerge. First, missing data is considerably lower in the marriage records, especially among men. Roughly 10 percent of marriage records have a missing education value for either the husband or wife. In the birth records, this figure increases to more than 10-15 percent among mothers and more than 20 percent for fathers.[3] Second, Figure A8 shows that the number of missing values is higher in more recent years than it was in the 1990s. This is particularly concerning for us as we are interested in how assortative matching has evolved

[1] The four omitted states are México, Michoacán, Nayarit, and Querétaro.

[2] Several of these omitted states (e.g., Michoacán) have particularly high Mexico to U.S. Migration, which has been shown to affect assortativeness (Choi and Mare, 2012). As a sensitivity analysis, we adopt a less restrictive sample restriction where we only omit state-years where missing data is exceptionally high. This works out to omitting roughly one-third of state-years for each of the four problem states. Our results are not affected by this modification.

[3] In the United States, Shen (2020) finds that around 20 percent of birth records have missing information on the fathers education. Our data is therefore comparable in quality.

over time.

There is no obvious reason for the high number of missing values or for why the have grown over time. There are systematic differences across states, which suggests that state collection procedures may be driving the high number of missing values. Unfortunately, we can only speculate regarding what the actual education levels are for these observations. If they are random, then our estimates are robust. However, if they are correlated with education, i.e., if individuals with lower educational attainment are less likely to report their education level, this may affect our conclusions.

To address the missing data issue, we examine the sensitivity of the results to several different assumptions regarding the characteristics of the problem records. Consider a marriage where the wife's education is observed, but the husband's education is not. We first assume that every missing husband has an identical education level as the wife. This will result in the maximum level of assortativeness, or the "upper bound". We next assume that every missing husband has a different level of education compared to the wife. This will result in the minimum level of assortativeness possible, or the "lower bound". Finally, our preferred estimate, which is, in effect, what we do in our main analysis, is to assume that the missing data is random, and that the problem marriage record follows an identical pattern as the non-missing data. We make these assumptions only for when one of the two spouses has non-missing data.

We present the results in Table A6. In the interest of clarity, we limit our attention to the first and last years of the data. Panel A provides the results for marriages, while Panel B does the same for births. In columns (1) and (3), we report the main estimates that were calculated in the main text. In columns (2) and (4) we present the lower and upper bound of our assortativeness estimates based on whether couples with missing education levels are all homogamous, or instead non-homogamous. We find the bounds for the marriage records are not wide, which is perhaps unsurprising given that it is rare for only one spouse's education level to be observed. On the other hand, it is quite common to observe only the mother's education in the births records. As a result, the lower and upper bounds provided in Panel B are quite wide. Nonetheless, we wish to emphasize that the intervals are highly conservative, and the goal is only to place a bound on the assortative measures. For completeness, we repeat the analysis using our alternative measure of assortativeness in Table A7.

Table A6: Impact of Missing Data on Assortative Measures

	1993		2018	
	Estimate	[Lower bound, Upper bound]	Estimate	[Lower bound, Upper bound]
Panel A: Marriages	(1)	(2)	(3)	(4)
(i,j) = (Primary, Middle)	1.986	[1.950-2.022]	2.133	[2.116-2.159]
(i,j) = (Middle, Secondary)	1.718	[1.678-1.770]	1.807	[1.795-1.828]
(i,j) = (Secondary, College)	2.145	[2.093-2.211]	2.459	[2.358-2.517]
(i,j) = (Primary, Secondary)	4.302	[4.207-4.345]	4.132	[4.083-4.158]
(i,j) = (Lower Secondary, College)	3.363	[3.264-3.423]	4.298	[4.086-4.356]
(i,j) = (Primary, College)	6.138	[5.869-6.188]	6.613	[5.854-6.675]
Observations	496,398	502,609	358,423	363,783
Panel B: Births				
(i,j) = (Primary, Middle)	1.821	[0.903-2.269]	1.833	[0.731-2.354]
(i,j) = (Middle, Secondary)	1.504	[0.567-2.027]	1.531	[0.590-2.028]
(i,j) = (Secondary, College)	2.035	[1.084-2.481]	2.353	[1.309-2.747]
(i,j) = (Primary, Secondary)	3.975	[1.885-4.453]	3.630	[1.608-4.150]
(i,j) = (Lower Secondary, College)	3.386	[1.854-3.802]	4.477	[2.202-4.872]
(i,j) = (Primary, College)	6.060	[2.846-6.431]	6.823	[2.803-7.241]
Observations	635,126	726,159	510,713	585,381

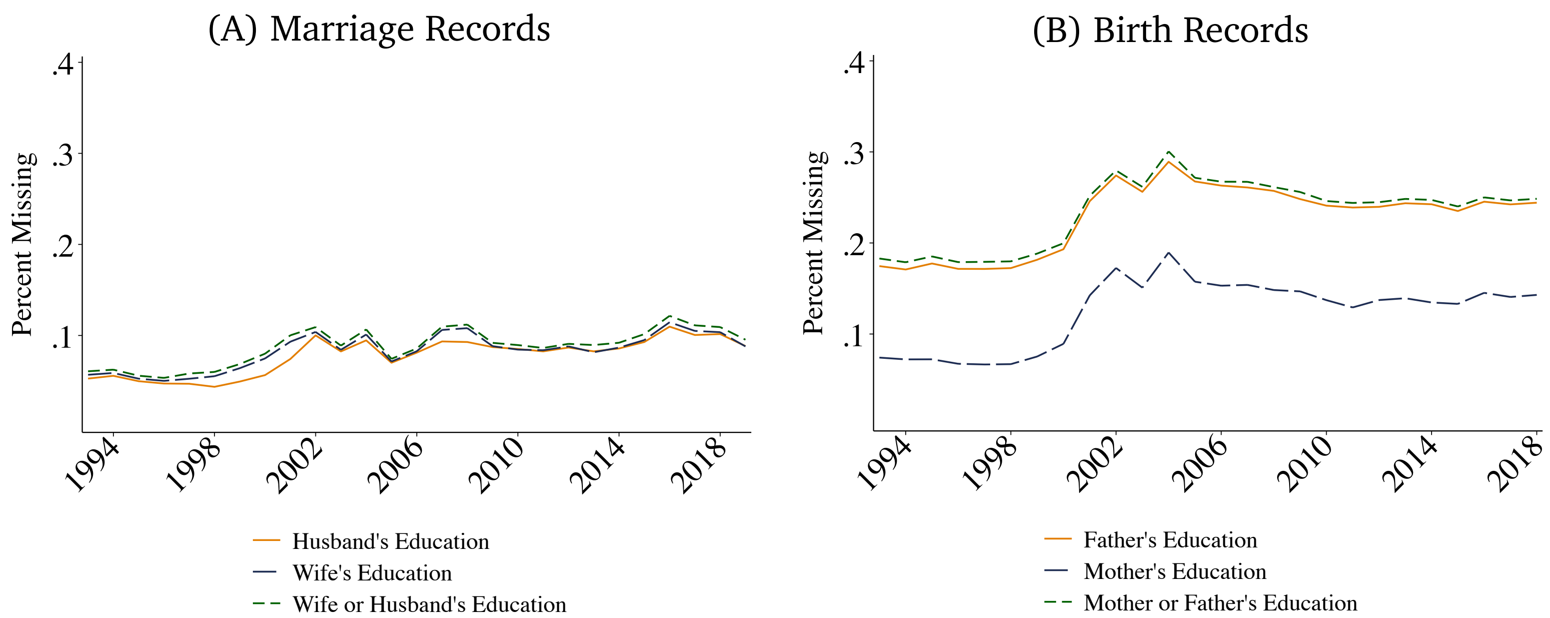
Notes: Vital Statistics Marriage and Birth Records. Men and women are divided into four mutually exclusive education categories: 1. Primary or Less, 2. Middle, 3. Secondary, 4. College. Each row provides the assortative matching measure for the diagonal 2×2 sub-matrices of the full sorting matrix using the Separable Extreme Value Index. The lower bound of assortativeness is calculated by assuming that all missing education levels for one partner is *different* than the observed education level for the other partner. The upper bound of assortativeness is calculated by assuming that all missing education levels for one partner is the *same* as the observed education level for the other partner. The bounds do not account for couples where both partners have missing educational attainment.

Table A7: Impact of Missing Data on Assortative Measures (Merged Categories)

	1993		2018	
	Estimate	[Lower bound, Upper bound]	Estimate	[Lower bound, Upper bound]
Panel A: Marriages	(1)	(2)	(3)	(4)
Primary	2.549	[2.499-2.580]	2.960	[2.770-2.992]
Middle	1.352	[1.311-1.391]	1.813	[1.810-1.845]
Secondary	1.860	[1.800-1.904]	1.657	[1.651-1.691]
College	3.101	[3.020-3.150]	3.052	[2.911-3.107]
Observations	496,398	502,609	358,423	363,783
Panel B: Births				
Primary	2.378	[1.967-2.729]	2.449	[2.019-2.867]
Middle	1.221	[0.657-1.640]	1.463	[1.095-1.878]
Secondary	1.737	[1.251-2.165]	1.417	[0.890-1.835]
College	3.164	[2.748-3.479]	3.159	[2.457-3.462]
Observations	635,126	726,159	510,713	585,381

Notes: Vital Statistics Marriage and Birth Records. Men and women are divided into four mutually exclusive education categories: 1. Primary or Less, 2. Middle, 3. Secondary, 4. College. Each row is a comparison of education category i with the combined remaining three education categories. The lower bound of assortativeness is calculated by assuming that all missing education levels for one partner is *different* than the observed education level for the other partner. The upper bound of assortativeness is calculated by assuming that all missing education levels for one partner is the *same* as the observed education level for the other partner. The bounds do not account for couples where both partners have missing educational attainment.

Figure A8: Missing Data by Year



Notes: Vital Statistics Marriage and Birth Records. The solid-orange line gives the percentage of marriage or birth records without the father's education by year. The blue, long-dashed line does the same for the mother's education. The blue-dashed line denotes the percentage of marriages or births by year where the education of either men and women are missing.

C Additional Results

C.1 Assortative Matching Across Merged Categories

As an alternative to the main findings, we select a single category k, merge the other three categories (i.e., category $\neg k$), and compare the single to the merged category. This second method results in four measures of assortativeness in total. The benefit of using the merged categories is that it allows us to compute a summary measure of assortativeness following Eika et al. (2019) and Shen (2020).

In Figure A9 and Table A8, we present the merged education categories, where we measure assortativeness between, e.g., individuals with a college degree, and those without. We also follow Eika et al. (2019) and Shen (2020) and compute weighted averages of assortativeness across these measures, where the weights are determined by the diagonal values given in Table A3. Again, the results are mostly flat, though there appears to be a slight increase.

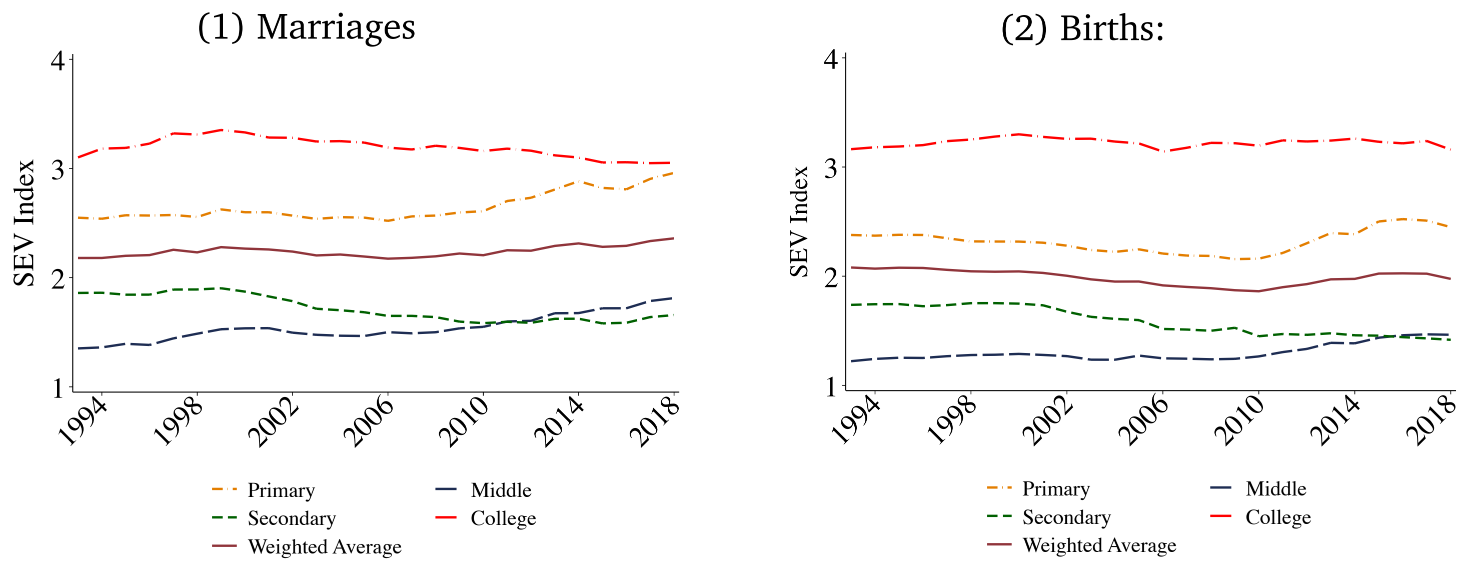
Interestingly, when we examine the merged categories, it appears assortativeness among college graduates has declined. However, what this suggests is that an increasing share of the population has a secondary degree, and we know from Panel A of Table 3 that assortativeness is relatively low when comparing secondary-college matches. The difference in results for comparisons involving college graduates highlights the local nature of assortativeness.

Table A8: Changes in Assortativeness 1993-2018 (Merged Categories)

	Marriages			Births		
	1993	2018	Difference	1993	2018	Difference
Merged Categories	(1)	(2)	(3)	(4)	(5)	(6)
Primary	2.549	2.960	0.411	2.378	2.449	0.072
Middle	1.352	1.813	0.461	1.221	1.463	0.242
Secondary	1.860	1.657	-0.203	1.737	1.417	-0.320
College	3.101	3.052	-0.049	3.164	3.159	-0.004
Weighted Average	2.180	2.360	0.180	2.080	1.975	-0.104
Observations	496,358	358,423		635,126	510,713	

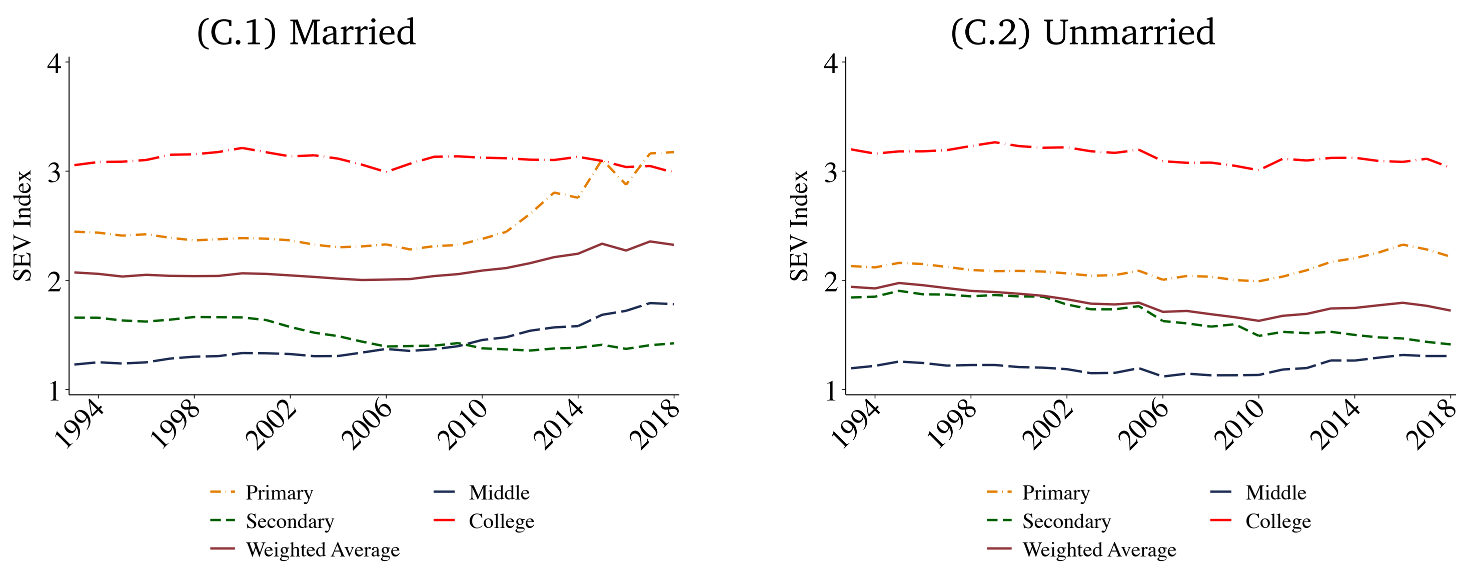
Notes: Vital Statistics Marriage and Birth Records. Men and women are divided into four mutually exclusive education categories: 1. Primary or Less, 2. Middle, 3. Secondary, 4. College. Each row is a comparison of education category i with the combined remaining three education categories. The weighted average measure is computed by averaging the assortative index across educational levels, where the weights are determined by diagonal value of the matching table given in Table A3.

Figure A9: Assortative Marriage and Parental Matching: Merged Categories



Notes: Vital Statistics Marriage and Birth Records. Men and women are divided into four mutually exclusive education categories: 1. Primary or Less, 2. Middle, 3. Secondary, 4. College. Each line is a comparison of education i with the combined remaining three education categories. The weighted average curve is computed by averaging the assortative index across educational levels, where the weights are determined by diagonal value of the matching table given in Table A3.

Figure A10: Assortative Parental Matching (Married vs. Non-Married): Merged Categories



Notes: Vital Statistics Birth Records. Men and women are divided into four mutually exclusive education categories: 1. Primary or Less, 2. Middle, 3. Secondary, 4. College. Each line is a comparison of education i with the combined remaining three education categories. The weighted average curve is computed by averaging the assortative index across educational levels, where the weights are determined by diagonal value of the matching table given in Table A3.

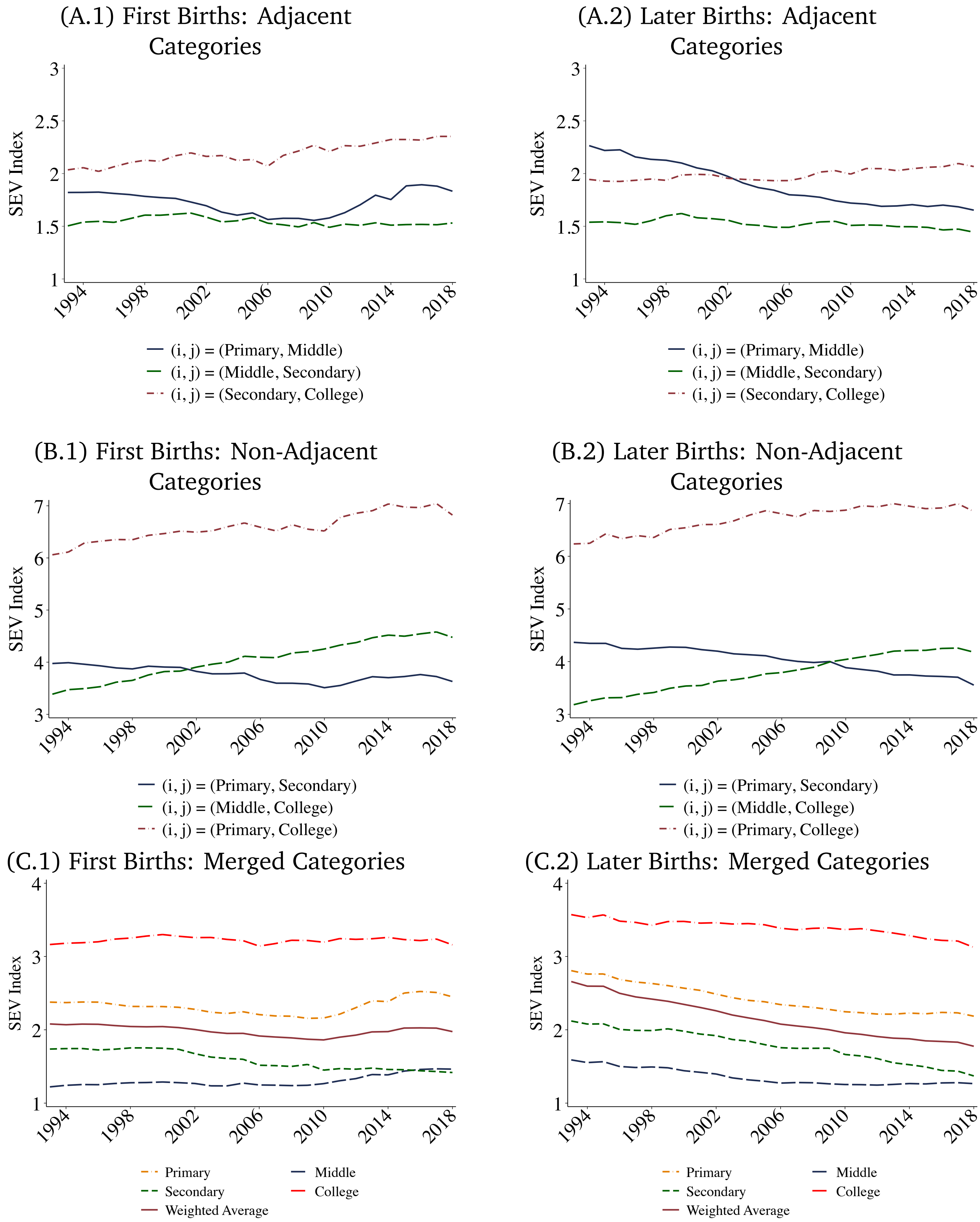
C.2 Additional Parental Matching Results

In Section 4.3, we focused our primary analysis on first births to avoid counting the same parental matches twice. This restriction prevents us from investigating the relationship between assortativeness and fertility. In order to shed light on this relationship, we compare our measures of assortativeness between first births (i.e., the main results) and non-first births using these two non-overlapping samples. If we observe lower assortativeness among non-first births, this *may* suggest that negative-assortative matches have more children. To see this, suppose that only negative assortative matches have multiple children, while all positive assortative matches have one child. In this case, we would find significantly more assortativeness among first births than non-first births. An important caveat of this implication arises from how we are forced to define first births. Specifically, we consider a child to be a first birth if it is the first child of the mother. It is possible that a couples first child is not the first child of the mother if she has had a previous birth with another partner. Hence, our results are only suggestive about the relationship between fertility and assortativeness.

The results are presented in Figure A11 and Table A9 in the Appendix. We first discuss the local measures presented in Panels A and B of Figure A11. Among comparisons involving college graduates, assortativeness is greater for first births, and has increased by more over time compared to later births. This suggests that negative-assortative matches involving college graduates have more children. This finding is consistent with the college-educated spouse specializing in market work, and the spouse with less than a college degree specializing in child care. The greatest difference in assortative measures is found in comparisons involving the primary-middle school $2x2$ sub-matrix. Among first births, there is little change in assortativeness over time. However, when looking at non-first births, we see a large decline.

Panel C of Figure A11 presents the merged categories. We hesitate to interpret these results too strongly as they are influenced by which parents are placed in the comparison group. Specifically, when we focus on first births, secondary and college educated individuals are a higher proportion of the merged category, and therefore there is more assortativeness (as can be seen in the local measures where those education levels tend to be more assortative). When we shift to focusing on later births, lower education couples comprise a larger share of the merged category, which will skew the results to be less assortative. For this reason, we prefer using the more local measures, consistent with the advice of Chiappori et al. (2020).

Figure A11: Assortative Parental Matching (First vs. Later Births)



Notes: Vital Statistics Marriage and Birth Records. Men and women are divided into four mutually exclusive education categories: 1. Primary or Less, 2. Middle, 3. Secondary, 4. College. In Panels A and B, each figure plots assortative matching for the diagonal 2×2 sub-matrices of the full sorting matrix using the Separable Extreme Value Index. We plot first births on the left, and non-first births on the right. Panel A plots adjacent education categories while Panel B plots non-adjacent categories. In Panel C, each line is a comparison of education i with the combined remaining three education categories. The weighted average curve is computed by averaging the assortative index across educational levels, where the weights are determined by diagonal value of the matching table given in Table A3.

Table A9: Changes in Assortativeness 1993-2018 (First vs. Later Births)

	First Births			Later Births		
	1993	2018	Difference	1993	2018	Difference
Panel A: 2 by 2 Comparisons	(1)	(2)	(3)	(4)	(5)	(6)
(i,j) = (Primary, Middle)	1.821	1.833	0.012	2.265	1.653	-0.612
(i,j) = (Middle, Secondary)	1.504	1.531	0.028	1.538	1.444	-0.094
(i,j) = (Secondary, College)	2.035	2.353	0.317	1.944	2.067	0.122
(i,j) = (Primary, Secondary)	3.975	3.630	-0.345	4.366	3.554	-0.811
(i,j) = (Lower Secondary, College)	3.386	4.477	1.091	3.183	4.179	0.997
(i,j) = (Primary, College)	6.060	6.823	0.763	6.234	6.850	0.616
Panel B: Merged Categories						
Primary	2.378	2.449	0.072	2.807	2.186	-0.621
Middle	1.221	1.463	0.242	1.590	1.266	-0.324
Secondary	1.737	1.417	-0.320	2.120	1.370	-0.750
College	3.164	3.159	-0.004	3.571	3.125	-0.447
Weighted Average	2.080	1.975	-0.104	2.658	1.775	-0.883
Observations	635,126	510,713		1,265,354	778,435	

Notes: Vital Statistics Birth Records. Men and women are divided into four mutually exclusive education categories: 1. Primary or Less, 2. Middle, 3. Secondary, 4. College. In Panel A, each row provides the assortative matching measure for the diagonal 2×2 sub-matrices of the full sorting matrix using the Separable Extreme Value Index. In Panel B, each line is a comparison of education category i with the combined remaining three education categories. The weighted average measure is computed by averaging the assortative index across educational levels, where the weights are determined by diagonal value of the matching table given in Table A3.

D Separable Extreme Value Model

The goal of this section is to provide additional details of the Separable Extreme Value (SEV) model, and to provide the derivation of the resulting measure of assortative matching. The content of this section follows closely with Chiappori et al. (2020). All terms are defined as in the main text.

This is a frictionless marriage matching model of heterosexual couples.[4] Denote men by the subscript i and women by the subscript j. Men and women each maximize their utility, where each potential marriage generates a surplus s_{ij} that is divided among the spouses. The model assumes Transferable Utility, so that the surplus is the sum of each spouses individual utility, with $s_{ij} = u_i + v_j$, where $u_i = U^{IJ} + \epsilon_i^J$, and $v_j = V^{IJ} + v_j^I$ represent the utility of men and women, respectively.

The SEV model relies upon a number of conditions. First, there must be a large number of men and women relative to the number of "types" of individuals (i.e., education categories), where the total number of types of men and women are given by I and J, respectively. Second, the surplus generated from a match must be composed of a deterministic component (Z^{IJ}) that does not vary across individuals, and a random term (γ_{ij}) that reflects unobserved individual preference heterogeneity, with $s_{ij} = Z^{IJ} + \gamma_{ij}$. Moreover, the utility for single individuals is normalized to zero and given by $s_{i0} = \epsilon_i^J$ and $s_{0j} = v_j^I$. The resulting matrix $Z = ([Z^{IJ}])$ then reflects individual preferences for different types of partners, and will be central to the SEV index measure of assortativeness.

For simplicity, assume there are two types of education categories for men and women; that is, $I = J = 2$. Then matrix Z will be a 2×2 matrix with a supermodular core (the sum of the diagonal elements minus the sum of the off-diagonal elements) of $S = Z_{11} + Z_{22} - Z_{12} - Z_{21}$. S is a measure of complementarity, and assortativeness will be increasing in S.

The third assumption is that the random term γ_{ij} is additively separable with $\gamma_{ij} = \epsilon_i^J + v_j^I$. These terms represent unobservable individual tastes for certain types of partners. Fourth, the SEV model assumes that these terms are Type 1 Extreme Value, which results in differences in utility across education partners following a logistic distribution. A more general model would allow unobservable preferences to follow a more flexible distribution (see e.g., Dupuy and Galichon (2014)), and would allow these terms to vary over time (see Ciscato et al. (2020) and Ciscato and Weber (2020)).

With this setup, Chiappori et al. (2020) construct a measure of assortativeness, stated below in the following proposition and subsequent proof:

Proposition: Let m, n, and r be defined as they are in the main text. Then table (m, n, r) can be generated by any SEV model such that its supermodular core satisfies

$$Z^{11} + Z^{22} - Z^{12} - Z^{21} = 2 \ln\Big(\frac{r(1 + r - m - n)}{(m - r)(n - r)}\Big)$$

[4] Ciscato et al. (2020) incorporate same-sex couples in their analysis of assortative matching in the United States. They find lesbian couples exhibit a higher degree of assortativeness relative to both gay and heterosexual couples.

One of the structural matrices that would generate Table (m, n, r) is:

$$Z = 2\begin{pmatrix} \ln r & \ln(m-r) \\ \ln(n-r) & \ln(1+r-m-n) \end{pmatrix}$$

Proof:

The probability P^{IJ} of any woman with education $i \in I$ matching with a man of education J is given by

$$\begin{aligned} P^{IJ} &= Pr(U^{IJ} + \epsilon_i^J \geq U^{IK} + \epsilon_i^K)\ \forall\ K \\ &= Pr(\epsilon_i^J - \epsilon_i^K \geq U^{IK} - U^{IJ})\ \forall\ K \end{aligned}$$

The above equation simply says that if we observe a woman marrying a particular type of man, it must be the case that she derives more utility from that specific match relative to any available alternative.

Similarly, for men, let Q^{IJ} be the probability of a man $j \in J$ being matched with a woman in category I:

$$\begin{aligned} Q^{IJ} &= Pr(V^{IJ} + \nu_i^I \geq V^{IK} + \nu_i^K)\ \forall\ K \\ &= Pr(\nu_i^I - \nu_i^K \geq V^{KJ} - V^{IJ})\ \forall\ K \end{aligned}$$

Since the error terms are assumed to be Type 1 Extreme Value, the probabilities are given by (Choo and Siow, 2006):

$$P^{IJ} = \frac{\exp U^{IJ}}{\sum_K \exp U^{IK}} \quad \text{and} \quad Q^{IJ} = \frac{\exp V^{IJ}}{\sum_K \exp V^{KJ}}$$

Then, in the two-education case, the matching probabilities for women are given by:

$$P^{11} = \frac{r}{m} = \frac{\exp U^{11}}{\exp U^{11} + \exp U^{12}} \quad \text{and} \quad P^{12} = 1 - P^{11}$$

$$P^{21} = \frac{n-r}{m} = \frac{\exp U^{21}}{\exp U^{21} + \exp U^{22}} \quad \text{and} \quad P^{22} = 1 - P^{21}$$

And for men,

$$Q^{11} = \frac{r}{n} = \frac{\exp V^{11}}{\exp V^{11} + \exp V^{21}} \quad \text{and} \quad Q^{21} = 1 - Q^{11}$$

$$Q^{21} = \frac{m-r}{n} = \frac{\exp V^{12}}{\exp V^{12} + \exp V^{22}} \quad \text{and} \quad Q^{22} = 1 - Q^{12}$$

Chiappori et al. (2020) implement the following normalizations: $U^{11} = U^{21} = V^{11} = V^{12} = 0$. Note that alternative normalizations would alter the resulting values of Z^{IJ} but would not affect the measure of assortativeness.

Then,

$$U^{12} = \ln\Big(\frac{m-r}{r}\Big)$$
$$U^{22} = \ln\Big(\frac{1+r-m-n}{n-r}\Big)$$
$$V^{21} = \ln\Big(\frac{n-r}{r}\Big)$$
$$V^{22} = \ln\Big(\frac{1+r-m-n}{m-r}\Big)$$

Recall that $Z^{IJ} = U^{IJ} + V^{IJ}$. Then,

$$Z^{11} = 0, \;\; Z^{22} = \ln\Big(\frac{(1+r-m-n)^2}{(m-r)(n-r)}\Big), \;\; Z^{12} = \ln\Big(\frac{m-r}{r}\Big), \;\; Z^{21} = \ln\Big(\frac{n-r}{r}\Big)$$

and the measure of assortativeness is:

$$S = Z^{11} + Z^{22} - Z^{12} - Z^{21} == 2\ln\Big(\frac{r(1+r-m-n)}{(m-r)(n-r)}\Big) \tag{A1}$$

The SEV Index can then be defined as:

$$I_{SEV} = \ln\Big(\frac{r(1+r-m-n)}{(m-r)(n-r)}\Big) \tag{A2}$$

References

Chiappori, Pierre-André, Monica Costa Dias, and Costas Meghir, "Changes in Assortative Matching: Theory and Evidence for the US," 2020. [20], [23], [24]

Choi, Kate H and Robert D Mare, "International Migration and Educational Assortative mating in Mexico and the United States," *Demography*, 2012, *49* (2), 449–476. [13]

Choo, Eugene and Aloysius Siow, "Who Marries Whom and Why," *Journal of Political Economy*, 2006, *114* (1), 175–201. [24]

Ciscato, Edoardo, Alfred Galichon, and Marion Goussé, "Like attract like? A structural comparison of homogamy across same-sex and different-sex households," *Journal of Political Economy*, 2020, *128* (2), 740–781. [23]

_ **and Simon Weber**, "The role of evolving marital preferences in growing income inequality," *Journal of Population Economics*, 2020, *33* (1), 307–347. [23]

Dupuy, Arnaud and Alfred Galichon, "Personality traits and the marriage market," *Journal of Political Economy*, 2014, *122* (6), 1271–1319. [23]

Eika, Lasse, Magne Mogstad, and Basit Zafar, "Educational Assortative Mating and Household Income Inequality," *Journal of Political Economy*, 2019, *127* (6), 2795–2835. [18]

Shen, Jenny, "Assortative Mating and Intergenerational Mobility," *Unpublished Manuscript*, 2020. [13], [18]