Online Appendix

Appendix A- Choo-Siow Model Estimates

Table A1 reports estimates of matching by age-difference and education using the Choo-Siow (2006) framework to adjust for differences in the population supplies of different partner types. Following Choo and Siow (2006), let:

$$\pi_{ij} = \frac{\mu_{ij}}{\sqrt{\mu_{oi}}\sqrt{\mu_{oj}}}$$

where μ_{ij} is the number type *i* men married to type *j* women, μ_{i0} is the number of unmarried type *i* men and μ_{0j} is the number of unmarried type *j* women. In this analysis, couple types are categorized based on age and education. This standardization by μ_{i0} and μ_{0j} adjusts for supply side differences. In order to investigate matching by education and age-difference, the model specification is:

$$\pi_{ij} = \alpha_0 + \alpha_1 age_i + \alpha_2 age_i^2 + \alpha_3 age_i^3 + \beta_1 age_j + \beta_2 age_j^2 + \beta_3 age_j^3 + \sum_{k=1}^{15} \delta_k * EdCat_{kij} + \gamma(age_i - age_j)^2 + \sum_{k=1}^{15} \delta_k * EdCat_{kij} * (age_i - age_j)^2 + \varepsilon_{ij}$$

where *EdCat* is a vector of 15 indicator variables for all possible combinations of husband and wife education using four education categories: no high school, high school, college and advanced degree. The omitted 16th category is neither has a high school degree. We expect our estimate of γ to be negative, indicating that among couples in which neither spouse completed high school, pairings become less common as with-in couple age difference increases. Our interest is in the estimates of δ_k. If differently-aged couples are even less common in couples with higher education, we expect the δ_k estimates to be negative, and the magnitude to be larger for couples with higher educational attainment.

The estimates are reported below in Table A1. The couple type cells are calculated on the same age range, ages 25-60, as the rest of the analysis in the paper. Small cells, those with fewer than 20 couples in the 1960 and 1970 Census, and those with fewer than 100 couples in the 1980 Census, are dropped from the sample. This eliminates, in all three Census years, all cells in which an individual with a college or advance degree is partnered with a spouse without a high school degree, leaving at most 11 coefficient estimates.

The results are consistent with the educational attainment results in the paper, indicating that differently-aged couples are less prevalent among couples with higher education.

Table A1: Choo-Siow Matching Estimates for Educational Attainment and Age-Difference

	1960	1970	1980
$(age_i - age_j)^2$ *			
Education Category:			
H: HS W: no HS	-0.0008 (0.0001)	-0.0007 (0.0001)	-0.0003 (0.0001)
H: no HS W: HS	-0.0003 (0.0001)	-0.0004 (0.0001)	0.0000 (0.0001)
H: HS W: HS	-0.0019 (0.0001)	-0.0017 (0.0001)	-0.0003 (0.0001)
H: Coll W: HS	-0.0028 (0.0002)	-0.0030 (0.0002)	-0.0015 (0.0001)
H: HS W: Coll	-0.0053 (0.0010)	-0.0042 (0.0010)	-0.0022 (0.0003)
H: Coll W: Coll	-0.0059 (0.0011)	-0.0083 (0.0013)	-0.0059 (0.0005)
H: Adv W: HS	-0.0029 (0.0003)	-0.0027 (0.0001)	-0.0011 (0.0001
H: HS W: Adv	-	-	-0.0034 (0.0008)
H: Adv W: Coll	-0.0060 (0.0017)	-0.0074 (0.008)	-0.0050 (0.0003)
H: Coll W: Adv	-	-	-0.0062 (0.0007)
H: Adv W: Adv	-	-0.0061 (0.0015)	-0.0041 (0.0003)
N	2514	2720	3209

Notes: First row reports δ_k for couples in which husband has a high school degree and wife has less than a high school degree. Remaining rows report other δ_k estimates. Robust standard errors in parentheses. Missing coefficients in 1960 and 1970 due to small cells.

Appendix B- Decomposing women's earnings into hours and wages

Table A2 estimates equation (3) on the subsample of women in Column 4 of Table 3 with positive earnings using logged earnings, logged hours, and logged earnings per hour as dependent variables. This decomposes earnings into hours and earnings/hour:

Log(Earnings)=log((Earnings/Hours) *(Hours))=log(Earnings/Hours)+log(Hours).

Table A2: Earnings, Hours and Earnings per Hour for Women with Positive Earnings

	Log (Earnings)	Log(Hours)	Log(Earnings/Hr)
Age Difference:			
1980			
+8 or more	0.047 (0.006)	0.051 (0.005)	-0.004 (0.003)
+5 to 7	0.039 (0.004)	0.027 (0.004)	0.011 (0.003)
+2 to 4	0.014 (0.003)	0.009 (0.003)	0.010 (0.002)
-2 to -4	0.013 (0.007)	0.037 (0.006)	-0.010 (0.013)
-5 to –7	-0.011 (0.015)	0.054 (0.013)	-0.011 (0.004)
-8 or more	-0.097 (0.023)	-0.008 (0.019)	-0.031 (0.005)
N	600,883	600,883	600,883
1990			
+8 or more	0.083 (0.004)	0.064 (0.003)	0.019 (0.002)
+5 to 7	0.038 (0.004)	0.031 (0.003)	0.007 (0.002)
+2 to 4	0.008 (0.003)	0.007 (0.002)	0.002 (0.002)
-2 to -4	0.053 (0.005)	0.053 (0.004)	0.000 (0.003)
-5 to –7	0.087 (0.008)	0.079 (0.006)	0.009 (0.004)
-8 or more	0.071 (0.009)	0.080 (0.007)	-0.009 (0.005)
N	954,447	954,447	954,447
2000			
+8 or more	0.048 (0.004)	0.050 (0.003)	-0.001 (0.003)
+5 to 7	0.028 (0.004)	0.023 (0.003)	0.005 (0.002)
+2 to 4	0.008 (0.003)	0.008 (0.002)	0.001 (0.002)
-2 to -4	0.017 (0.004)	0.025 (0.003)	-0.008 (0.003)
-5 to –7	0.037 (0.006)	0.045 (0.005)	-0.008 (0.004)
-8 or more	0.029 (0.008)	0.060 (0.006)	-0.031 (0.005)
N	986,196	986,196	986,196

Notes: For 1990 and 2000, regressions are weighted using Census person weights. Robust standard errors are in parentheses.

Appendix C- 1990 and 2000 Census results

Table A3: Average Earnings in Occupation and Earnings by Age Difference with Spouse, 2000 and 1990 Census Data

	Men		Women	
	Avg Earnings/Hr in Occupation	Earnings	Avg Earnings/Hr in Occupation	Earnings
Age Difference:				
2000				
+8 or more	-0.609 (0.029)	-5355.5 (178.7)	0.002 (0.017)	1031.8 (112.8)
+5 to 7	-0.542 (0.024)	-4717.6 (149.5)	0.000 (0.015)	732.2 (97.2)
+2 to 4	-0.264 (0.018)	-2259.9 (114.2)	-0.020 (0.011)	232.2 (71.9)
-2 to -4	-0.321 (0.024)	-3082.9 (156.0)	-0.077 (0.018)	426.1 (111.3)
-5 to –7	-0.605 (0.036)	-5474.5 (222.8)	-0.079 (0.028)	625.1 (176.8)
-8 or more	-0.777 (0.039)	-7894.4 (231.4)	-0.182 (0.034)	273.1 (218.5)
N	1,486,287	1,486,287	1,283,405	1,283,405
1990				
+8 or more	-0.457 (0.021)	-5148.3 (129.1)	0.010 (0.012)	1532.8 (76.0)
+5 to 7	-0.410 (0.017)	-4098.9 (103.5)	-0.014 (0.010)	784.2 (65.1)
+2 to 4	-0.163 (0.012)	-1762.0 (77.6)	0.001 (0.008)	238.1 (47.7)
-2 to -4	-0.270 (0.018)	-2835.0 (116.7)	-0.046 (0.014)	866.8 (81.2)
-5 to –7	-0.478 (0.028)	-4718.4 (174.5)	-0.054 (0.023)	1381.1 (134.4)
-8 or more	-0.591 (0.032)	-6071.1 (196.8)	-0.125 (0.028)	1281.3 (177.0)
N	1,432,619	1,432,619	1,231,439	1,231,439

Notes: Sample is married couples with both spouses ages 25-60 in the 1990 and 2000 Decennial Censuses who report an occupation for most recent job in the past 5 years. Age difference categories measure the number of years the man is older than the woman, and omitted categories is 1 to -1. Columns 1 and 3 report coefficient estimates from equation (3). Columns 2 and 4 report coefficient estimates in which the dependent variable in equation (3) is replaced with earnings and estimated using a Tobit model. All regressions weighted using Census person weights. Robust standard errors are in parentheses.

Appendix D—Age Dispersion of Discussion Networks by Education

The 1985 and the 2004 General Social Survey (GSS) data contain a topical module in which the respondent is asked: "From time to time, most people discuss important matters with other people. Looking back over the last six months, who are the people with whom you discussed matters important to you?" Information on age, sex, education and family relationship are recorded for up to five members of the respondent's "discussion network."

Many members of the respondents' discussion networks are spouses, parents, siblings and children. We calculate age dispersion measures for non-family members of the discussion network.[1] Two measures are used. For the first measure, the first listed non-family member of the discussion network is used to calculate the age difference between the respondent and that network member. The second measure uses all respondents who have at least two non-family members of their discussion network to calculate the standard deviation of age of non-family discussion network members.

These age dispersion measures are regressed on years of education with controls for sex, race, age and age-squared, and the number of people listed in the discussion network. The results are reported in Table A4. In all cases, there is a negative relationship between education and age-dispersion of the network, indicating that individuals with higher levels of education have non-family networks that are less age diverse than those with lower levels of education. The coefficient estimates are, however, only statistically significant in the 2004 data.

[1] Marsden (1987) uses the 1985 GSS data to analyze the age, race and education heterogeneity of discussion networks, but does not calculate these measures separately for non-family members of the network.

Table A4: Age Dispersion of Discussion Network by Education, GSS 2004 and 1985 data

	Absolute Age Difference	Age Standard Deviation
2004		
Years of Education	-0.426 (0.002) ***	-0.233 (0.117)*
N	732	419
1985		
Years of Education	-0.158 (0.101)	-0.040 (0.075)
N	1060	755

Notes: First column samples contain those who report at least one non-family member of their discussion network. Second column samples contain those who report at least two non-family members of their discussion network. The dependent variable in column 1 is the absolute value of the age difference between the respondent and the first listed non-family member of the discussion network. The dependent variable in column 2 is the standard deviation of age for non-family members of the discussion network. All regressions control for sex, race, age, age-squared, and number of members of the discussion network. Robust standard errors are in parentheses.

Appendix E—Interactions with Age at first Marriage

If, at ages when marriages are most likely to occur, higher ability individuals typically interact with a wider age distribution than lower ability individuals, then we expect to observe a relationship between ability and marital age gap. Because it is likely that the networks of higher ability individuals become more diverse in age over the lifecycle, we would expect the relationship to be weaker among those who marry at later ages.

We therefore test whether the relationship between age difference and AFQT score is weaker for those who marry at older ages. In order to manage the number of coefficient estimates, the categorical age difference specification in equation (1) is replaced with two linear age-difference variables: *AgeDiffPos* is the number of years the man is older than the woman, and equals zero if the woman is older; *AgeDiffNeg* is the number of years the women is older than the man, and equals zero the man is older.

Because average age of first marriage is younger for those with lower educational attainment, this analysis is performed separately for individuals with and without a college degree. For college-educated individuals, the age difference variables are interacted with indicators for marriage by or after age 26.[2] For those with less than a college degree, the age difference variables are instead interacted with indicators for marriage by or after age 23.[3]

The results for AFQT scores are reported in Table A5. For men, the coefficient estimates on the age difference variables are, as predicted, weaker for men who married at later marriages.

[2] The main effects of the age of marriage indicators are included as well, while still retaining the linear and quadratic age of marriage controls.

[3] In the 1980 analysis sample, 21% of first marriages by college-educated individuals occur after age 26 and 29% of first-marriages for those with less than a college degree occur after age 29.

The differences in the coefficients are, however, not statistically significant. The coefficients for women are smaller and show no clear pattern.

Table A6 reports the same analysis for occupational wage. The results are consistent with our hypothesis that the relationship between occupational wage and age difference is weaker for those who marry at older ages. Coefficient pairs are in bold if there is a statistically significant difference between the coefficients for early marriages and late marriages, and among the statistically significant pairs the strong majority are also consistent with a weakening relationship.

Table A5: AFQT Scores by Age Difference with Spouse and Age of Marriage Interactions, NLSY79

	Men	Women
W/ College		
AgeDiffPos*AgeMarr ≤ 26	-1.736 (1.11)	-0.267 (0.289)
AgeDiffPos*AgeMarr>26	-0.837 (0.538)	-0.313 (0.252)
AgeDiffNeg*AgeMarr ≤ 26	-1.598 (1.071)	-0.081 (1.398)
AgeDiffNeg*AgeMarr>26	0.021 (0.883)	0.192 (0.694)
N	981	1141
W/O College		
AgeDiffPos*AgeMarr ≤ 23	-1.296 (0.487)**	-0.082 (0.186)
AgeDiffPos*AgeMarr>23	-0.452 (0.253)+	-0.213 (0.176)
AgeDiffNeg*AgeMarr ≤ 23	-1.055 (0.392)**	-0.362 (1.09)
AgeDiffNeg*AgeMarr>23	-0.406 (0.240)+	-0.254 (0.479)
N	3521	3744

Notes: Sample of first marriages in NLSY79. Dependent variable is AFQT score. AgeDiffPos is the number of years the man is older than the woman, and equals zero if the woman is older. AgeDiffNeg is the number of years the woman is older than the man, and equals zero if the man is older. 1979 Sampling weights are used, as are the same controls as in equation (1). Robust standard errors are in parentheses.

Table A6: Average Earnings per Hour in Occupation by Age Difference with Spouse Interacted with Age of Marriage Interactions, 1980 Census

	1980	1970	1960
Men			
W/ College			
AgeDiffPos*AgeMarr ≤ 26	**-0.120 (0.008)**	**-0.087 (0.021)**	**-0.072 (0.021)**
AgeDiffPos*AgeMarr>26	**-0.029 (0.007)**	**0.028 (0.019)**	**0.006 (0.016)**
AgeDiffNeg*AgeMarr ≤ 26	-0.106 (0.011)	-0.082 (0.031)	**-0.103 (0.027)**
AgeDiffNeg*AgeMarr>26	-0.083 (0.016)	-0.102 (0.045)	**-0.017 (0.036)**
N	275,659	40,299	27,756
W/o College			
AgeDiffPos*AgeMarr ≤ 23	**-0.067 (0.004)**	**-0.136 (0.008)**	**-0.119 (0.007)**
AgeDiffPos*AgeMarr>23	**-0.051 (0.003)**	**-0.076 (0.005)**	**-0.062 (0.009)**
AgeDiffNeg*AgeMarr ≤ 23	-0.052 (0.004)	**-0.043 (0.008)**	**-0.042 (0.008)**
AgeDiffNeg*AgeMarr>23	-0.062 (0.004)	**-0.083 (0.009)**	**-0.067 (0.008)**
N	756,381	188,039	200,560
Women			
/w College			
AgeDiffPos*AgeMarr ≤ 26	**-0.011 (0.003)**	-0.008 (0.010)	-0.018 (0.010)
AgeDiffPos*AgeMarr>26	**0.006 (0.006)**	0.015 (0.019)	0.005 (0.016)
AgeDiffNeg*AgeMarr ≤ 26	**-0.107 (0.017)**	-0.045 (0.046)	-0.039 (0.045)
AgeDiffNeg*AgeMarr>26	**-0.041 (0.014)**	-0.045 (0.036)	0.021 (0.028)
N	153,426	19,389	11,877
W/o College			
AgeDiffPos*AgeMarr ≤ 23	-0.009 (0.001)	**-0.009 (0.002)**	**-0.021 (0.002)**
AgeDiffPos*AgeMarr>23	-0.007 (0.002)	**-0.019 (0.004)**	**-0.029 (0.003)**
AgeDiffNeg*AgeMarr ≤ 23	**-0.074 (0.006)**	-0.046 (0.012)	**-0.095 (0.014)**
AgeDiffNeg*AgeMarr>23	**-0.045 (0.003)**	-0.057 (0.008)	**-0.046 (0.006)**
N	604,797	137,128	129,693

Notes: Samples are the same as used in Columns 3 and 4 of Table 3. Dependent variable is average earnings per hour in occupation. *AgeDiffPos* is the number of years the man is older than the woman, and equals zero if the woman is older; *AgeDiffNeg* is the number of years the women is older than the man, and equals zero the man is older. Interactions of these age difference measures with age of marriage indicators, as well as the age of marriage main effects, replace the age-difference categories in equation (3). Robust standard errors in parentheses. Coefficients in bold indicate a statistically significant difference (p-value<0.05) between the coefficient for those married early and those married late.

Appendix F- Age Dispersion within Occupations

We suggest that higher quality individuals are more likely to work in jobs with high upward mobility, and as a result, individuals who share their job description tend to be similarly-aged. On the other hand, lower quality individuals are likely to work in jobs with limited upward mobility. Therefore, there is greater age variation among co-workers who share their job description.

In this appendix, we use the 1980 Census to first test whether higher wage occupations have greater age segregation. We then test whether age heterogeneity in own job has a positive relationship with age difference with spouse.

First, within each occupation, we separate workers into categories based on deciles of the wage distribution in that occupation. Because we have no other information that allows us to separate workers into jobs within occupation, we define job categories based on wages. For each wage decile category in each occupation, we calculate the standard deviation of age of workers.

We first test whether high wage occupations have greater age segregation by "job" category. This is equivalent to testing for job ladders. For each occupation, we use the average wage of workers ages 35 to 50 as the occupational wage measure. The raw correlation between the occupational wage measure and the age heterogeneity measure is -0.46, suggesting that occupations with higher average wages are less age diverse by job category.

We test this relationship controlling for worker characteristics with the following regression specification:

(A1)

$$Job\_AgeStDev_{idj} = \beta_0 + \beta_1 Occ\_Earn\_Age3550_j + Race_i\alpha_1 + \alpha_2 AgeMarr_i + \alpha_3 AgeMarr_i^2 + \alpha_4 HS_i + \alpha_5 Coll_i + \alpha_6 Adv_i + + \sum_{a=1}^{A} \gamma_a * Age_{ia} + \sum_{s=1}^{S} \phi_s * State_{is} + \sum_{s=1}^{S} \psi_s * (State_{is} * Urban_i) + \varepsilon_i$$

where for person *i* in wage decile *d* in occupation *j*, *Job_AgeStDev* is the standard deviation of age of workers in the job-occupation category, *Occ_Earn_Age3550* is the average hourly earnings of workers ages 35-50 in that occupation, and the age, race, education, age of marriage and location controls are as described in the text of the paper.

To further test whether age heterogeneity in own job predicts age difference with spouse, we estimate:

(A2)

$$AbsAgeDiff_{ijklmn} = \beta_0 + \beta_1 Job\_AgeStDev\_W_{jk} + \beta_2 Job\_AgeStDev\_H_{mn} + \alpha_1 AgeMarr\_W_i + \alpha_2 AgeMarr\_W_i^2 + Race\_W_i\alpha_3 + Race\_H_l\alpha_4 + \alpha_5 HS\_W_i + \alpha_6 Coll\_W_i + \alpha_7 Adv\_W_i + \alpha_5 HS\_H_l + \alpha_6 Coll\_H_l + \alpha_7 Adv\_H_l + \sum_{a=1}^{A} \gamma_a * Age\_W_{ia} + \sum_{s=1}^{S} \phi_s * State_{is} + \sum_{s=1}^{S} \psi_s * (State_{is} * Urban_i) + \varepsilon_i$$

where for a couple with wife *i* in wage decile *j* in occupation *k* and husband *l* in wage decile *m* in occupation *n*, *AbsAgeDiff* is the absolute value of the age difference, *Job_AgeStDev_W* is the age standard deviation in wife's job and *Job_AgeStDEv_H* is age standard deviation in husband's job.

All analysis is conducted on the sample of married couples in which both husband and wife are in their first marriage. Panel A of Table A7 reports the results from equation (A1), which shows that higher wage occupations have lower age dispersion by wage category.

Panel B of Table A7 reports the results form equation (A2), showing that the age dispersion in husband's and wife's job predict couple age difference, but husband's age dispersion matters more.

Table A7: Age Dispersion in Job and Age Difference with Spouse, 1980 Census

Panel A: Age dispersion in job and average wage in occupation		
	Husbands	Wives
Average Hourly Earnings in Occupation, Workers Age 35-50	-0.204(0.000) ***	-0.274 (0.001)*
N	906,530	644,585
Panel B: Couple age difference and husband's and wife's age dispersion in job		
Age Standard Deviation in Wife's Job	0.009 (0.001)***	
Age Standard Deviation in Husband's Job	0.055 (0.001)***	
N	805,101	

Notes: The sample is married couples ages 25 to 60 in the 1980 Census where both spouses are in their first marriage. In Panel A, only the individual has to report an occupation in the past 5 years to be included in the sample. In Panel B, both spouses must report an occupation in the past 5 years to remain in the sample.

Online Appendix References

Choo, Eugene, and Aloysius Siow, "Who Marries Whom and Why" *Journal of Political Economy*. 114 (2006), 175-201.

Marsden, Peter, "Core discussion networks of Americans" *American Sociological Review*, 52 (1987), 122-31.