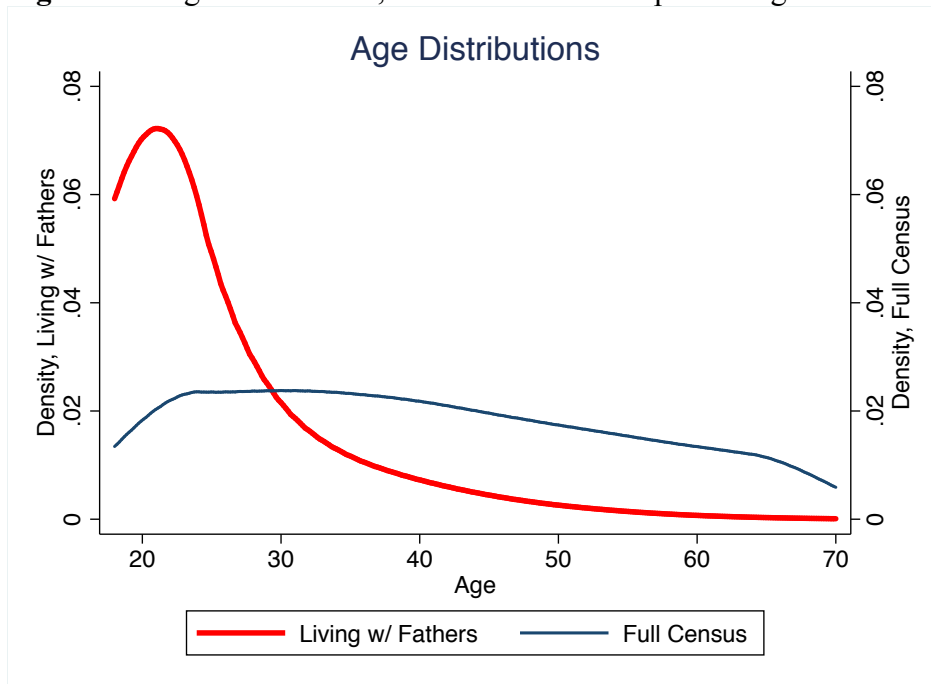
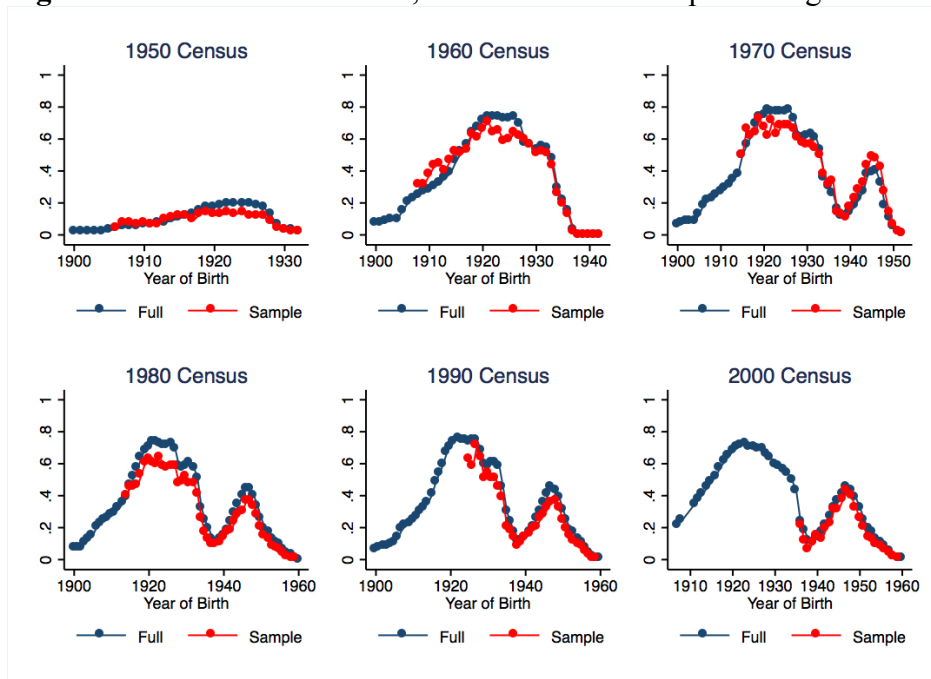


# Online Appendix

**Figure A1.** Age Distribution, Full Census vs. Sample Living with Father

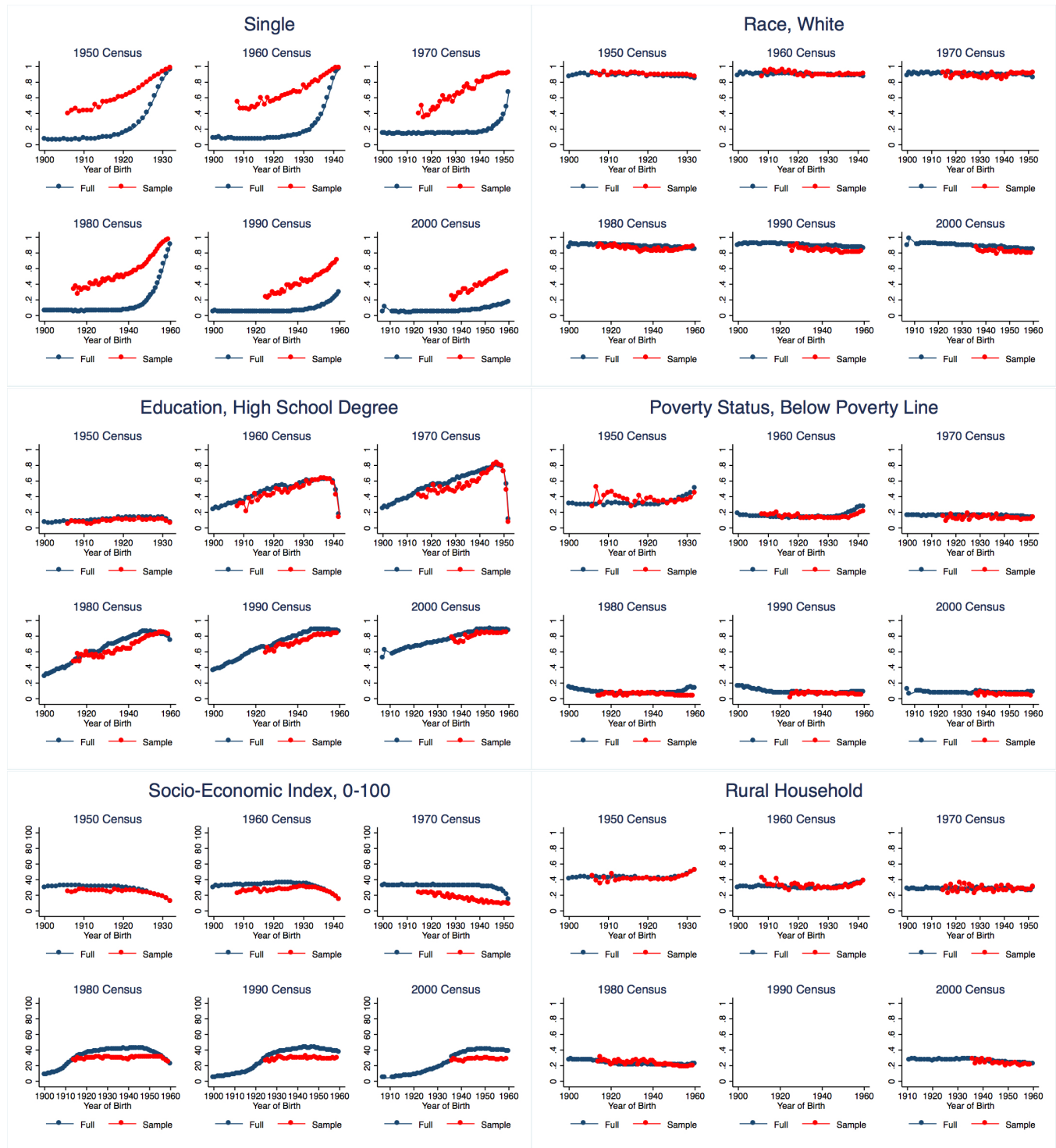


**Figure A2.** War Veteran Status, Full Census vs. Sample Living with Father



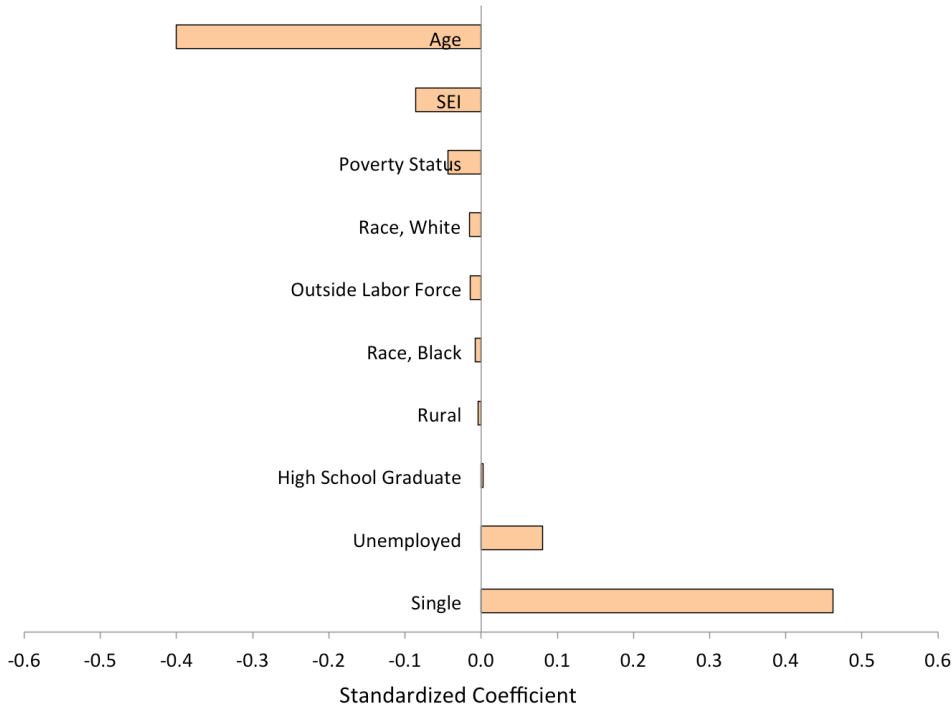
*Note:* The figure plots the mean of the war service dummy (WWII, Korea, Vietnam) in the full census data and our sample of sons living with their fathers. The blue dots refer to the full census means, and the red dots refer to the sample means. The data is restricted to cohorts with at least 100 observations. The figure shows the likelihood of serving in the population is approximately similar in the sample.

**Figure A3. Demographics and Socio-Economic Outcomes, Full Census vs. Sample, Means**



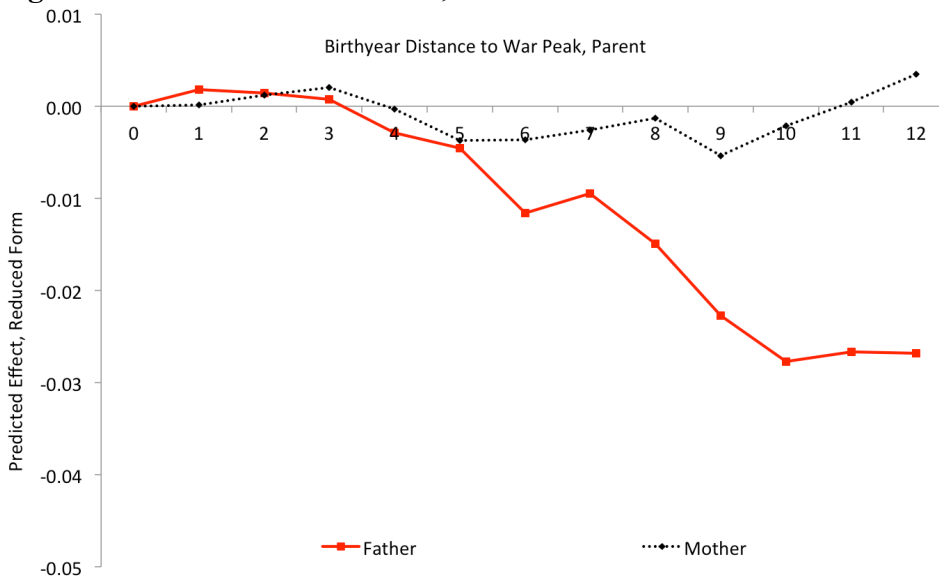
Note: The figure plots the mean of socioeconomic characteristics in the full census data of adult men, and in the sample of sons living with their fathers, across cohorts within the same census. The blue dots refer to the full census means, and the red dots refer to the sample means. The data is restricted to cohorts with at least 100 observations. The figure shows the characteristics in the population are approximately similar in the sample, with the exception of single status.

**Figure A4. Predictors of Living with Father, Standardized Coefficients**



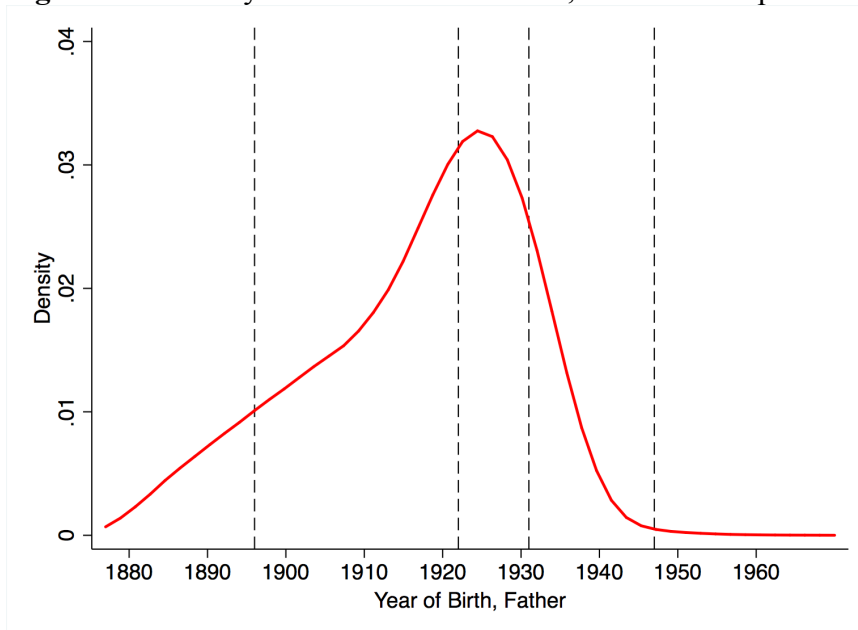
*Note:* Standardized coefficients from a bivariate regression predicting likelihood of living with the father, using the full census data from 1950-2000. The figure shows that that two strongest predictors are age at the time of the census and single status (never married).

**Figure A5. Reduced Form Effects, Fathers and Mothers**



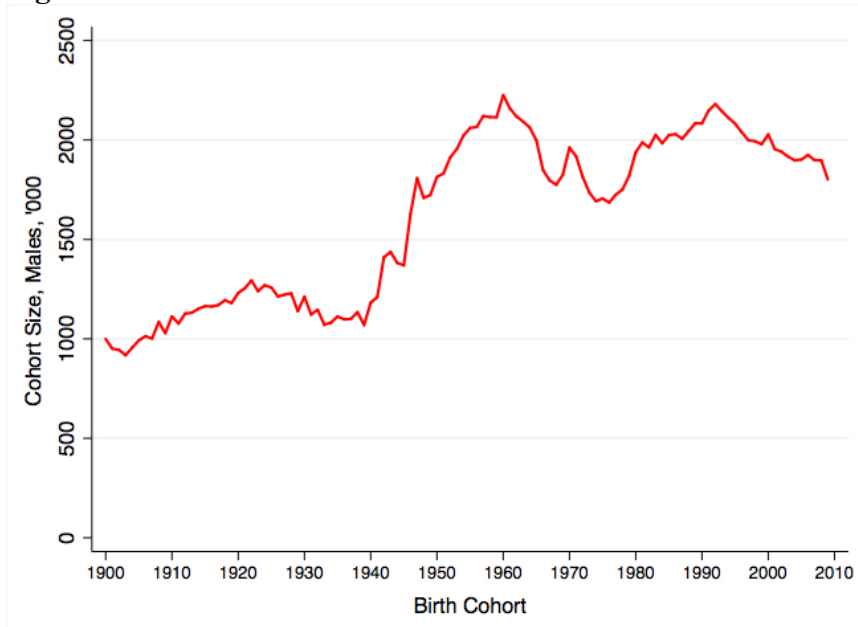
*Note:* The figure plots the coefficient estimates from the flexible specification using dummies to indicate the parent year of birth distance to the peak cohorts. The reduced form effect refers to the father, and the placebo to the mother. The figure shows that fathers - not mothers - drive the reduced form effects.

**Figure A6. Density of Father Birth Cohorts, Baseline Sample**



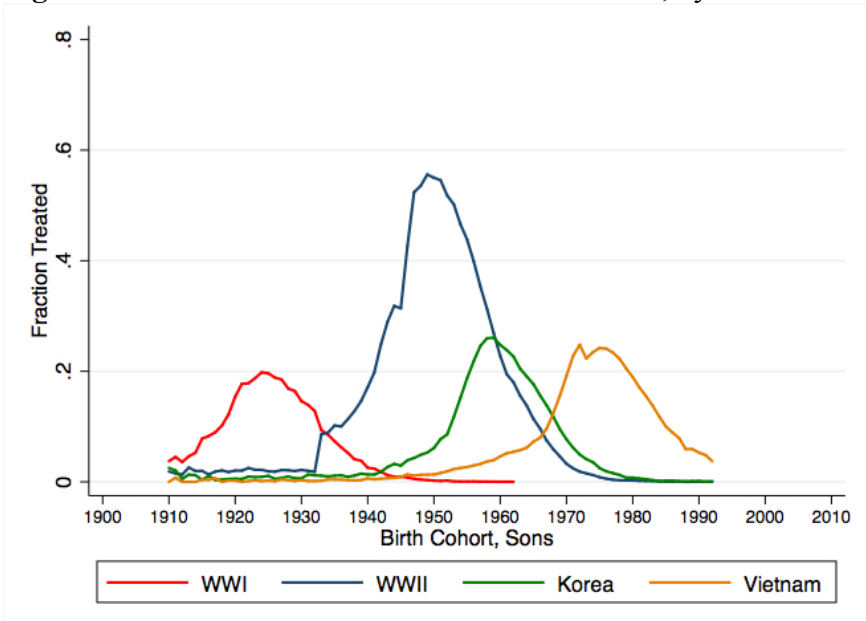
*Note:* The graph plots the distribution of father cohorts in the baseline sample used in Table 2 regressions. The dashed lines indicate the war peak cohorts.

**Figure A7. Cohort Size of Males**



*Note:* The graph plots the estimated total number of males born in each cohort, using the full U.S. Census data.

**Figure A8.** Share of Fathers who are War Veterans, by Cohort



*Note:* The figure shows the estimated share of fathers who served in each war, for each cohort among sons in the sample.

**Table A1. Sample Selection, Predictors of Likelihood of Living with Father**

Dependent Variable:	Living with Father, Dummy									
<b>SAMPLE:</b>	<b>Full Census</b>									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Age, Log	-0.278 (-0.400)									
Single		0.310 (0.462)								
Race, White			-0.013 (0.000)							
Race, Black				0.003 (0.000)						
High School Graduate					-0.005 (0.000)					
Poverty Status, Below Poverty Line						-0.042 (0.000)				
Socio-Economic Index							-0.092 (0.000)			
Unemployed								0.115 (0.001)		
Outside Labor Force									-0.010 (0.000)	
Rural										-0.002 (0.000)
Observations	14,901,322	14,901,322	14,901,322	14,901,322	14,901,322	13,940,463	14,901,322	14,901,322	14,901,322	9,527,601
R2	0.160	0.213	0.000	0.000	0.000	0.002	0.007	0.007	0.000	0.000
Age FE	No	No	No	No	No	No	No	No	No	No
Census FE	No	No	No	No	No	No	No	No	No	No
<b>Standardized Effects</b>	<b>-0.400</b>	<b>0.462</b>	<b>-0.016</b>	<b>-0.008</b>	<b>0.003</b>	<b>-0.043</b>	<b>-0.086</b>	<b>0.081</b>	<b>-0.014</b>	<b>-0.004</b>

Note: Sample using full Census data from 1950-2000, males age 18 and above. Bivariate regressions in all columns. The corresponding standardized effects are available in the bottom row, and plotted in Figure A4. Robust standard errors in parentheses.

**Table A2. Robustness, Dropping 1950 Census**

Sample	<i>Dep. Var.: War Veteran, Son</i>							External Validity: Inverse Probability Weighted Regressions			
	Baseline	Baseline	Baseline	Baseline	Son Age is 16-30 at Time of War	Son Age + Excl. Peak Cohorts	Baseline	Baseline, Excl. <0.5% Prob.	Baseline, Excl. <1% Prob.	Baseline, Excl. <1% Prob.	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
War Veteran, Father	0.075 (0.006)	0.081 (0.006)	0.079 (0.006)	0.066 (0.008)	0.071 (0.006)	0.083 (0.008)	0.078 (0.006)	0.187 (0.012)	0.111 (0.012)	0.065 (0.013)	0.095 (0.015)
High School Graduate, Father							-0.017 (0.002)				
Went to College, Father							0.002 (0.000)				
Years of Education, Father							-0.022 (0.001)				
Observations	398,295	398,295	398,295	398,295	397,297	300,406	398,295	398,295	383,230	365,239	339,016
Census FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Father Birthyear Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Son Birthyear FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Race Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
IV-variable	Linear	2-Poly.	3-Poly.	Dummy	Linear	Linear	Linear	Linear	Linear	Linear	Linear
Sample Predictors	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Baseline	Baseline	Baseline	Baseline+
Dependent Variable, Mean	0.121	0.121	0.121	0.121	0.121	0.127	0.121	0.121	0.108	0.101	0.094

Notes: The sample, specifications and variables definitions are the same as Table 3, except that the 1950 Census is dropped. Robust standard errors in parentheses.

**Table A3. Main Effects, Alternative Standard Errors**

Sample	War Veteran, Son						
	Baseline Sample	Baseline Sample	Baseline Sample	Baseline Sample	Son Age is 16-30 at Time of War	Son Age + Excl. Peak Cohorts	Baseline Sample
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
War Veteran, Father	0.098	0.101	0.099	0.092	0.100	0.108	0.100
Clustered SE: State	(0.006)	(0.006)	(0.006)	(0.008)	(0.008)	(0.007)	(0.007)
Clustered SE: Father YOB	(0.015)	(0.015)	(0.015)	(0.022)	(0.017)	(0.012)	(0.015)
Clustered SE: Father YOB + State	(0.015)	(0.015)	(0.015)	(0.021)	(0.017)	(0.012)	(0.015)
Clustered SE: Father YOB + Son YOB	(0.017)	(0.017)	(0.019)	(0.015)	(0.017)	(0.013)	(0.017)
Observations	458,181	458,181	458,181	458,181	456,877	350,320	458,181
Census FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Father Birthyear Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Son Birthyear FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Race Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Father Education Controls	No	No	No	No	No	No	Yes
IV-variable	Linear	2-Polynomial	3-Polynomial	Dummy	Linear	Linear	Linear

Note: This table is identical to columns 1-7 of Table 3, except it uses alternative ways to calculate the standard errors, clustered at different levels. State refers standard errors clustered at the state level, Father YOB refers to clustering at the level of the father year of birth. Father YOB + State refers to two-way clustered standard errors at both levels, as does Father YOB + Son YOB.



**Table A4. Main Effects By Region**

	Dep. Var.: War Veteran, Son							
	Northeast		Midwest		South		West	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
War Veteran, Father	0.119 (0.012)	0.120 (0.012)	0.090 (0.011)	0.090 (0.011)	0.092 (0.010)	0.091 (0.010)	0.087 (0.016)	0.088 (0.016)
Observations	127,721	127,721	116,281	116,281	141,666	141,666	72,513	72,513
Census FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Father Birthyear Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Son Birthyear FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	No	Yes	No	Yes	No	Yes	No	Yes
Race Controls	No	Yes	No	Yes	No	Yes	No	Yes
Dependent Variable, Mean	0.147	0.147	0.144	0.144	0.135	0.135	0.154	0.154

Note: IV/2SLS estimates, baseline instrument. Same sample and variable definitions as in Table 3. Baseline instrument. Robust standard errors in parentheses.

**Table A5. Extrapolation of Effects to the Population, IV/2SLS**

Dependent Variable:	War Veteran, Son											
<b>SAMPLE:</b>	<b>Baseline Sample</b>											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
War Veteran Father	0.099 (0.006)	0.112 (0.006)	0.100 (0.006)	0.098 (0.006)	0.107 (0.006)	0.091 (0.006)	0.110 (0.006)	0.110 (0.006)	0.111 (0.006)	0.107 (0.007)	0.096 (0.006)	0.129 (0.010)
War Veteran Father * Age, Log	0.240 (0.016)											
War Veteran Father * Single		-0.120 (0.011)										
War Veteran Father * Race, White			-0.113 (0.012)									
War Veteran Father * Race, Black				0.098 (0.013)								
War Veteran Father * High School					-0.137 (0.009)							
War Veteran Father * Poverty						0.098 (0.019)						
War Veteran Father * SEI							-0.123 (0.015)					
War Veteran Father * Unemployed								0.019 (0.011)				
War Veteran Father * Outside Labor Force									0.033 (0.008)			
War Veteran Father * Rural										0.015 (0.008)		
War Veteran Father * Predicted Prob(cohabit)											-0.143 (0.017)	-0.143 (0.017)
Observations	458,181	431,131	458,181	458,181	458,181	382,253	431,131	431,131	431,131	331,932	382,253	382,253
Census FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Father Birthyear Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Son Birthyear FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Race Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Predictor Demeaned at Mean of:	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Pop.
Std. Effect of Predictor on Prob(cohabit)	-0.400	0.462	-0.016	-0.008	0.003	-0.043	-0.086	0.081	-0.014	-0.004	N/A	N/A
<b>Predicted Effect at Population Mean</b>	<b>0.195</b> <b>(0.011)</b>	<b>0.177</b> <b>(0.010)</b>	<b>0.099</b> <b>(0.006)</b>	<b>0.097</b> <b>(0.006)</b>	<b>0.101</b> <b>(0.006)</b>	<b>0.094</b> <b>(0.006)</b>	<b>0.102</b> <b>(0.006)</b>	<b>0.109</b> <b>(0.006)</b>	<b>0.112</b> <b>(0.006)</b>	<b>0.106</b> <b>(0.007)</b>	<b>0.129</b> <b>(0.010)</b>	<b>0.129</b> <b>(0.010)</b>

Note: Baseline sample, US Census data from 1950-2000, men age 18 and above. IV/2SLS estimates using the baseline instrument and the interaction with the predictor. Each regression includes the predictor of the interaction term as a control variable. All other controls are the same as in Table 3. All predictors are demeaned at the sample mean in columns 1-10. The standardized effects of predictors on Prob(cohabit), i.e. the likelihood of living with the father, are from bivariate regressions in Table A1. The bottom row shows the predicted effect of War Veteran Father at the population mean (full Census) of the predictor. Column 11 shows the regressions when the interaction is the predicted likelihood of living with the father, estimated from the full census (see table A6, column 2), and demeaned at the sample mean. In column 12, it is demeaned at the population mean from the full census data 1950-2000 of men age 18 and above. The estimates are plotted in Figure 6. Robust standard errors in parentheses.

**Table A6. Estimating Prob(Living w/ Father), Census Data 1950-2000**

Dependent Variable:	Living with Father, Dummy	
<b>SAMPLE:</b>	<b>Full Census</b>	
	(1)	(2)
Age	-0.159*** (0.000180)	
Age, squared	0.00120*** (1.82e-06)	
Race, White	-0.177*** (0.00280)	-0.210** (0.00327)
Race, Black	-0.291*** (0.00330)	-0.408** (0.00388)
Single		1.173** (0.00155)
High School Graduate	-0.111*** (0.00139)	0.0134** (0.00175)
Socio-Economic Index		-0.00550** (2.94e-05)
Poverty Status, Below Poverty Line		-0.987** (0.00285)
Unemployed		0.260** (0.00263)
Outside Labor Force		0.285** (0.00194)
Observations	14,991,322	13,939,518
R-squared	0.160	0.213
Census FE	Yes	Yes
Age FE	No	Yes

Note: Sample using full Census data from 1950-2000, on males age $\geq$ 18. Probit regressions, from which predicted likelihood of living with father is estimated (Prob(Living w/ Father), used in inverse probability weighted regressions in Table 3 and A5. Column 1 includes pre-determined variables (before potentially enlisting in the armed forces), while column 2 add additional socio-economic characteristics. Robust standard errors in parentheses. \*\* p<0.01, \* p<0.05.

**Table A7. The Effects by Son Age at War Midpoint**

Sample: Son's age at War Midpoint	War Veteran, Son					
	Age: 14-16	Age: 17-19	Age: 20-22	Age: 23-25	Age: 26-28	Age: 29-31
	(1)	(2)	(3)	(4)	(5)	(6)
War Veteran, Father	0.0206 (0.0140)	0.0342 (0.0191)	0.0791 (0.0205)	0.0600 (0.0246)	0.0193 (0.0268)	-0.0371 (0.0205)
Observations	97,039	76,675	64,869	45,848	20,376	17,059
Census FE	Yes	Yes	Yes	Yes	Yes	Yes
Father Birthyear Controls	Yes	Yes	Yes	Yes	Yes	Yes
Son Birthyear FE	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Race Controls	Yes	Yes	Yes	Yes	Yes	Yes

Note: The variables definitions are the same as Table 3. Each sample is restricted to sons that were in the specified age range for at least one of the 20th Century wars, with the age at war midpoint specified at the top of each column. Robust standard errors in parentheses.

# The Intergenerational Transmission of War

## Data Appendix

Filipe Campante\* David Yanagizawa-Drott†

June 2015

In this appendix we describe in detail the data sets and procedures used in the paper.

### 1 US Census

We start from the US Census' Integrated Public-Use Microdata Series (IPUMS-USA), using the 5% sample whenever available (1980-2000), and the 1% sample otherwise. We merge together the data for each Census year between 1950 and 2000.

We obtain the information on fathers using the “*poploc*” variable, which for any individual indicates the number ascribed to that individual's father within the household. In other words, for each Census variable  $x$  described below, the father's information comes from “ $x\_pernum\_i$ ”, where  $poploc = i$ . This yields “ $x\_father$ .” Note that, as described by the Census, *poploc* identifies “social relationships (such as stepfather and adoptive father) as well as biological relationships.” The procedure is similar, using *momloc*, for information on mothers.

The key variables are: *birthyear* (for year of birth), *vetstat* (for veteran status), *vetwwi* (World War I veteran, from 1950-1980), *vetwwii* (World War II veteran, from 1950-2000), *vetkorea* (Korean War veteran, from 1960-2000), *vetvietn* (Vietnam War veteran, from 1970 to 2000), *vet90x95* (“Gulf War” veteran). We also use the variables *vet55x64*, *vet75x80*, *vet80x90* to look at service outside of wartime, and demographic variables – *educ* (years of education).

We define the key years (1896, 1922, 1931, 1947), subtract each of these numbers from *birthyear* (as long as the latter is within 20 years of the key year in question), and take the absolute value. The year-of-birth distance is the minimum over the four numbers thus generated. For our main regressions, we drop all observations for which we cannot define “*birthyear\\_father*,”

---

\*Harvard Kennedy School and NBER. Email: Filipe\_Campante@harvard.edu

†Harvard Kennedy School and NBER. Email: David\_Yanagizawa-Drott@harvard.edu

which leaves us with the sample of individuals who live with their fathers at the time of the Census, and keep adult male individuals at least aged 18.

The individual (son) is coded as a veteran of each war according to the aforementioned war-specific variables. He is coded as a “Big 3” veteran if he is a veteran of any of the three major wars (WWII, Korea, Vietnam). The father is coded as a “Big 4” veteran if he is a veteran of any of these wars, or WWI.

## **2 National Longitudinal Survey (NLS)**

We use three NLS datasets: The Young Men Survey, which includes 5225 men who were ages 14-24 when first interviewed in 1966; NLSY79, with 6403 males of age 14-22 when first interviewed in 1979, and; NLSY97, containing 4599 males of ages 12-17 when first interviewed in 1997.

In the Young Men Survey, the year of birth of the father is not a variable in the dataset. We first use the 1966 household roster to identify the father in the household, if present, and create a variable for his age. We then calculate his year of birth as the survey year (1966) minus his age. For those who have missing values, we search for the father by repeating the procedure in the surveys up to 1970. This procedure results in identified fathers for 66% of the observations. We use a similar procedure in NLSY97, resulting in identified year of birth of fathers for 55% of the male observations. In NLSY79, the age of the father is available as a variable for 75% of the males, enabling us to calculate father year of birth in those cases. In all three datasets, we then create the same instrumental variable as we did in the US Census data, described above.

The parenting style variables in NLS97 are available for the years 1997-2000. We create the authoritative parenting style dummy that equals one if the parenting is authoritative at any time in that time period, and zero otherwise. For the military aspirations data, we pool the Young Men of 1966 and the NLSY79 surveys. The former contains geographic identifiers at the census division level, while the latter only includes identifiers for region. We thus construct aggregated geographic identifiers at the region level in the Young Men survey, in order to have homogeneously defined geographical controls in the regressions.