2013

Marriage, migration and work: three essays on mobility in the United States, 1850-1930

BOSTON UNIVERSITY

GRADUATE SCHOOL OF ARTS & SCIENCES

Dissertation

MARRIAGE, MIGRATION, AND WORK: THREE ESSAYS ON MOBILITY IN THE UNITED STATES, 1850-1930

by

LAURA SALISBURY

B.A., McGill University, 2006

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2013

Approved by

First Reader

Robert A. Margo, Ph.D
Professor of Economics

Second Reader

Claudia Olivetti, Ph.D
Associate Professor of Economics

Third Reader

Daniele Paserman, Ph.D
Associate Professor of Economics

Acknowledgments

I thank Robert Margo, Claudia Olivetti, Daniele Paserman, and Carola Frydman for invaluable advice. Comments from Timothy Bond, Joseph Burton, Louis Cain, Dora Costa, Shari Eli, Joseph Ferrie, Michael Haines, Frank Lewis, Aloysius Siow, Richard Steckel, and seminar participants at Boston University, Michigan, Harvard, Northwestern, CPE-University of Chicago, Toronto, Guelph, Antwerp, National Bureau of Economic Research, and the annual meetings of the Economic History Association are also gratefully acknowledged. I also thank Noelle Yetter at the CPE, and the helpful and knowledgeable staff at the National Archives in Washington, DC. I acknowledge financial support from the National Science Foundation (SES-1227471), the Economic History Association, and the Institute for Economic Development at Boston University. Any errors are mine.

MARRIAGE, MIGRATION, AND WORK: THREE ESSAYS ON MOBILITY IN THE UNITED STATES, 1850-1930

(Order No.)

LAURA SALISBURY

Boston University, Graduate School of Arts & Sciences, 2013

Major Professor: Robert A. Margo, Professor of Economics

ABSTRACT

This dissertation studies three forms of mobility in the United States during the late nineteenth and early twentieth centuries. The first chapter uses newly collected data from Union Army widows pension files to isolate the causal effect of womens income on their decisions about marriage. Making use of exogenous variation in the processing time of pension applications, I show that receiving a pension caused widows to remarry at a significantly slower rate. This suggests that womens income directly influenced marital outcomes, largely by making women more selective in the marriage market. The second chapter explores the extent to which nineteenth century internal migrants in the United States were motivated by the possibility of upward occupational mobility. Drawing on the literature on contemporary migrant selection and sorting, I argue that workers with greater potential for occupational upgrading should have selected themselves out of counties with low skill premiums and sorted themselves into counties with high skill premiums. Using linked data from the U.S. Census and county-level wage data, I present results consistent with this argument.

The third chapter of the dissertation (co-authored with Claudia Olivetti and Daniele Paserman) examines intergenerational income mobility across three generations between 1850 and 1930. Making use of the socioeconomic content of names, pseudo-panels of three generations are created by grouping samples of individuals by first name. Using G1, G2, and G3 to index generations one two and three, respectively, we find a significant correlation between G1 and G3, controlling for G2. We also find differences in this correlation by gender, suggesting that the process by which income was transferred from fathers to daughters was not the same as the process by which it was transferred from fathers to sons.

Contents

List of Tables

List of Figures

List of Abbreviations

2SLS	....	Two Stage Least Squares
AFDC	....	Aid to Families with Dependent Children
AR	....	Autoregressive
CPE	....	Center for Population Economics
FEBRL	....	Freely Extensible Biomedical Record Linkage
G1	....	Generation 1
G2	....	Generation 2
G3	....	Generation 3
ICPSR	....	Inter-university Consortium for Political and Social Research
IPUMS	....	Integrated Public Use Microdata Series
NBER	....	National Bureau of Economic Research
NPB	....	Nam-Powers-Boyd
NYIIS	....	New York Identification and Intelligence System
OLS	....	Ordinary Least Squares
RAND	....	Research and Development
SSDI	....	Social Security Disability Insurance
UA	....	Union Army

Chapter 1

Women's Income and Marriage Markets in the United States: Evidence from the Civil War Pension

1.1 Introduction

Marriage markets in the United States changed substantially over the course of the 19th century. The average female age at first marriage rose from roughly 20 during the colonial period to a peak of 23.6 in 1890 (Haines 1996). While aggregate trends in marriage market outcomes are well documented for this period, a virtual absence of micro-level data following women through marriage makes it difficult to account for these patterns. In particular, the factors affecting women's decisions about when and whom to marry are not well understood. A number of explanations for the observed patterns have been proposed, including declining land availability, which increased the the cost of establishing new households,[1] and falling male-to-female ratios,[2] most notably in the aftermath of the Civil War. Much less attention has been paid to the role of women's economic opportunities in altering the desirability of marriage to women.[3] This paper fills this gap in the literature by presenting new evidence exploiting shocks to the income of Union Army widows. Through the compilation

[1] See for example Easterlin (1971; 1976) and Haines (1996).

[2] For example, see Haines (1996) and Hacker (2008).

[3] One recent study (Hacker 2008) includes this in a set of possible correlates of first marriage and documents a correlation between the age at first marriage and labor force participation among unmarried women in the 1860 census. However, this paper does not address the potential endogeneity of female labor force participation to norms of marriage or marriage market conditions.

of a novel database, this paper also helps to rectify the scarcity of data tracing 19th century women through marriage.

The desirability of marriage to women is largely ignored in accounts of 19th century marriage patterns. Economists model marital outcomes as the result of a balancing of costs and benefits; any factor affecting these costs and benefits may influence women's choices. If outside economic opportunities for women lower the net gains from marriage, we would expect them to substitute away from marriage and toward these alternatives.[4] This channel is considered very important for the later decades of the 20th century, which saw a simultaneous drop in marriage rates and explosion of female labor market opportunities. Between 1970 and 1995, the fraction of women ages 20-24 who had ever been marriage dropped from 64 to 34 percent (Blau Kahn and Waldfogel 2000); at the same time, female labor force participation increased from 49 to 72 percent, and the average female-male wage ratio rose from 0.56 to 0.72 (Blau 1998). While the 19th century did not see such a radical increase in opportunities for women, industrialization in the later part of the century facilitated women's work (Wanamaker 2012), as did the rise of the clerical sector beginning around 1890 (Goldin 1984).

In this paper, I offer evidence that women's income had a causal effect on their choices about marriage during the years immediately following the American Civil War. Using data newly collected for this project, I assess the effect of Civil War pension income on the behavior of Union Army widows in the marriage market. The Civil War pension provides a rare setting for studying this behavior. Under the General Law, passed on July 14, 1862, a woman was eligible to receive a pension if her husband served honorably in the Union Army and died as a consequence of this service; however, she lost her right to the pension if she remarried. As such, a pension increased the value of remaining single, but it was not correlated with individual

[4] See Becker (1973; 1991) for a theoretical development of this argument.

characteristics that affect marriage market outcomes, nor should it have rendered women more attractive to potential mates. In other words the effect of pension income on marital outcomes should work solely through women's preferences.[5]

Such a natural experiment is especially useful because establishing that women's income has a causal effect marital outcomes is difficult. Simply documenting a correlation between economic opportunities for women and delayed marriage is insufficient because of the interrelatedness of decisions regarding career and family. For example, both of the following explanations for such a correlation are plausible: women marry later because they have better labor market opportunities; or, women invest more in improving their labor market outcomes because norms of marriage have changed.[6] Moreover, labor market opportunities for women can affect their behavior through multiple channels: women may prefer market work to home production; at the same time, the increased income these opportunities afford may render them more selective. A social assistance program that carries a marriage penalty directly isolates this latter channel, which is akin to an income effect. Most recent examples of such programs are age-based or means tested.[7] During the period of focus of this study, the Civil War pension is neither of these.

Providing evidence for any mechanism driving 19th century marriage patterns is challenging because of data limitations. The first Census tabulations of marital status by age and sex were not published until 1890 (Hacker 2008). Moreover, samples that follow women through marriage during this period are all but impossible to construct from census data: a primary tool for creating linked census samples is last names, and all women changed their last names upon marriage. The creation of a novel database

[5]The argument is somewhat more subtle than this because a soldier's children could also receive a pension. I explain this in fuller detail in a later section.

[6]The latter story is consistent with the work of Goldin and Katz (2002) and Bailey (2006) on the relationship between contraception and marriage and female labor supply.

[7]Rosensweig (1999); Baker, Hanna and Kantarevic (2004); Brien, Dickert-Conlin and Weaver (2004).

following women through marriage is an important contribution of this paper. This database has the potential to provide insight into any number of questions about women's behavior in marriage markets during this period.

This paper seeks to determine the extent to which exogenous income shocks altered the relative costs and benefits associated with marriage. To illustrate how such shocks translate into observable outcomes, I use a theoretical model of search in the marriage market. I show that, by subsidizing the search for mates, pensions allow women to be more selective in their search process. Thus, pensions raise both the average time to remarriage and average match quality, conditional on remarrying at all. I show that the same predictions hold true in a comparison between women with accepted versus pending pension claims; this is due to uncertainty about if and when a pending claim will be approved. To assess the extent to which pensions caused women to delay remarriage, I make use of variation in the timing of pension decisions, or pension processing times. Because pension amounts were standardized, I argue that this is the most appropriate source of variation to use. I estimate a proportional hazards model of remarriage in which the rate of remarriage is allowed to shift at the moment a pension is granted. As such, I estimate a treatment effect of transitioning from having a pending claim to having an accepted claim. To evaluate the effect pensions had on match quality, I use links to the 1870 and 1880 censuses, which allow me to observe the characteristics of women's second husbands. I compare women who remarry with and without pensions along several plausible dimensions of match quality, including the second husband's occupational status and literacy.

One concern this paper addresses is the possible endogeneity of pension processing times to marital outcomes. This is largely due to sample selection, which is generated by the decision to apply for a pension. Women whose pensions take a long time to process tend to be those with ambiguous claims, and those who choose to incur the cost of applying for a pension even though their claims are ambiguous may be systematically different from those who apply with straightforward claims. To address this concern in my analysis of the relationship between pensions and the timing of remarriage, I exploit the fact that my treatment variable is a duration variable, which provides more information than is available in a standard cross-sectional setting. As I explain in a later section, variation in observables and the relationship between the hazard rates of pension receipt and remarriage provide sufficient information to correct for correlated unobserved heterogeneity in these two risks (Abbring and Van den Berg 2003; 2005). As an additional test, I estimate a linear version of this model using two stage least squares. My instrument for pension processing time is a measure of surname spelling homogeneity, calculated as the dispersion of unique spellings within phonetic surname groups in the censuses of 1860, 1870 and 1880. This generates variation in the difficulty of proving a soldier's identity, which altered the amount of time it took for a claim to be granted.

While I do not find conclusive evidence that pensions affected match quality, I do find a significant effect of pensions on the timing of remarriage. Specifically, I find that receiving a pension caused the rate of remarriage to drop by 40 percent, implying an increase in the median time to remarriage of approximately three years. This is especially striking because of the size of the pension. At eight dollars per month, the pension was less than half the monthly income of a typical farm laborer in 1870, so it was hardly enough to comfortably support a family. This finding lends credence to the idea that the incremental changes in female labor market opportunities seen in the 19th century may have contributed to the aggregate changes in marriage patterns that occurred during this period. In addition to offering new information about the way 19th century marriage markets worked, these results shed light on the behavior we observe during the 20th century. In particular, they suggest that the substitution of economic opportunities for marriage is not an entirely new behavior brought about by changing social norms.

1.2 Marriage and Women's Income in Historical Context

While the literature on marriage patterns in the United States before 1890 is small (Hacker 2008), it provides a broad picture of trends since the Colonial period. Haines (1996) shows an increase in the female age at first marriage up to about 1890. Fitch and Ruggles (2000) also find an increase in the female age at first marriage between 1850 and 1880; however, this increase is quite small, and seems to be concentrated in the years following the Civil War. It is well established that, during the last years of the 19th century, the age at first marriage began to fall, for men but more substantially for women, and it continued to decline until the middle of the 20th century.[8] Since the 1970s, the age at first marriage for women has been steadily increasing (Blau, Kahn and Waldfogel 2000).

Most explanations for 19th century trends in marriage focus on opportunities rather than preferences for marriage. In contrast to Western Europe, where "couples often delayed marriage until the prospective bridegroom inherited the family farm" (Fitch and Ruggles 2000, p. 62), land in the United States was cheap and abundant and did not pose an impediment to early marriage. However, as land became increasingly settled, marriage patterns started to more closely resemble those in Europe. As farmland grew scarcer and more expensive, "men were forced either to postpone marriage, working as farmhands or manual laborers until they had saved up enough money to set up their own farms, or to migrate to the western frontier" (Hacker 2008, p. 312). Easterlin (1976) also links the closing of the frontier to fertility control within marriage.[9] As international and internal migration patterns changed over the

[8]See Fitch and Ruggles (2000) and Haines (1996), for example.

[9]For further elaborations of this argument, see Haines (1996), Easterlin (1971), Haines and

course of the 19th century, declining male-to-female ratios likely contributed to the rising age at first marriage among women (Haines 1996; Hacker 2008). This would have been especially true in the years immediately following the Civil War.[10]

A small number of studies link women's economic opportunities to delayed marriage before the 20th century. Hacker (2008) offers evidence from the 1860 census that women tended to marry later in areas in which economic opportunities for women were greater; this is measured by local unmarried female labor force participation. In a somewhat related study, Wanamaker (2012) links industrialization to declining fertility in the 19th century, with a focus on fertility within marriage. Goldin (1995) indirectly links economic opportunities to delayed marriage by noting a tendency for women's education and marriage to be mutually exclusive. She describes a "stark set of alternatives between career and family" (p. 1) for women born at the end of the 19th century, noting that approximately half of college-educated women graduating in 1910 were childless. While this references a somewhat later period, women's colleges in the late 19th century were similarly labeled "spinster factories" (Monahan 1951, p. 242). Some historical writing notes a tendency for women to delay or forgo marriage in the presence of favorable alternatives. Paraphrasing a critical 1871 account of this behavior, Calhoun (1919) writes that "the opening sphere for women's talents is rendering marriage less popular for women; they are reluctant to marry a poor man; education inclines toward celibacy rather than marriage with poverty" (p. 205). Overall, the economic literature on 19th century marriage patterns is quite small. Moreover, data limitations severely limit its ability to provide evidence in support of the various drivers of these patterns.

Investigations into modern marriage markets place much more stock in the role of women's income in altering their behavior. There is a well developed theoretical

Hacker (2006).

[10]See Abramitzky, Delavande and Vasconcelos (2011) for an analysis of the effect of sex ratios on assortative matching in post-WWI France.

literature about this mechanism. In Becker's transferable utility model (1973, 1991), marriage generates utility by allowing couples to exploit increasing returns through division of labor, or by allowing both parties to consume collective goods such as children. A marriage will occur if marital output exceeds the sum of the output that both partners produce while single. As the gains from marriage arise from division of labor, married women will tend to specialize in home production as long as their market wages are lower than those of men, which has typically been the case. Thus, "an increase in the wage rate of women relative to men would tend to decrease the incentive to marry" (Becker 1973, p 822). Weiss (1997) notes that if labor market returns are higher for men than for women, high-earning women will experience relatively smaller gains from marriage than low-earning women.

Another class of model that generates this relationship between female income and marriage rates comes from search theory. If women's labor income functions as an alternative to marriage, it should raise the value of being single relative to the value of being matched. If being single increases in value, women will require more valuable matches in order to marry. Under random matching, such an increase in reservation match quality will lower the probability that any given proposal of marriage will be deemed suitable; thus, it will cause women to remain single longer. It will also raise average match qualities conditional on marrying at all.[11]

Most of the empirical literature on the effect of female income on marriage rates is descriptive, largely demonstrating a negative correlation between opportunities for women and marriage rates.[12] This type of exercise is subject to several biases. For

[11]See Rogerson, Shimer and Wright (2005) for a survey of basic search models. See Weiss (1997) for a review of search models applied to marriage markets. Gould and Paserman (2003) and Loughran (2002) use a search framework to investigate the effect of wage inequality on marriage rates.

[12]Keeley (1977) finds that women with high wages tend to marry later, although men with high wages tend to marry earlier. Ruggles (1997) argues that increasing female labor market opportunities contributed to the rise in divorce rates during the twentieth century. Weiss and Willis (1997) find that women with high earnings are more likely to divorce, while the opposite is true of men with high earnings. Price-Bonham and Balswick (1980) argue that widows are less likely to remarry than

one thing, income depends on human capital investment, which may be endogenous to preferences for marriage. A paper that deals explicitly with this causality issue is Blau, Kahn and Waldfogel (2000), who look at the effect of city-wide marriage and labor market conditions on marriage rates. They find that better female labor markets tend to decrease marriage rates, while better male labor markets tend to increase them. Still, it is not clear from this analysis that female labor market opportunities cause women's choices about marriage to change: areas in which these opportunities are greater may have different norms of marriage. A different approach is due to Choo and Siow (2006), who propose a statistic to directly measure the net gain from marriage for a given pair of male and female "types."[13] They attempt to quantify the net benefit from marriage for men and women using data from the 1970 U.S. Census and Vital Statistics. They find that the net benefit of marriage declined between 1970 and 1980 for both men and women, but more so for women. This is suggestive, as opportunities in the labor market for women grew significantly during this decade.

Other work takes a similar approach to this paper, looking at the effect of marriage penalties on the behavior of social assistance recipients. Rosensweig (1999) studies the effect of the AFDC program on marriage and out-of-wedlock childbearing for young women, and he finds that AFDC benefits tend to encourage fertility outside marriage. Baker, Hanna and Kantarevic (2004) find a significant negative effect of marriage penalties on remarriage, which they identify through the removal of marriage penalties from the public pension system in Canada during the 1980s. Brien, Dickert-Conlin and Weaver (2004) find that American widows and widowers delayed remarriage until after the age of 60 in response to the marriage penalty built into Social Security before

divorced women, as are older and more educated women with fewer children. Bahr (1979) finds that more affluent women are less likely to remarry after divorce. See also Waite and Spitze (1981) for an investigation into determinants of female age at first marriage.

[13]This statistic is the ratio of the number of matches formed by these types to the geometric mean of the number men and women of these types that remain single.

1979.

1.3 Institutional Background: Widows and the Civil War Pension Law

The original Civil War pension law, called the General Law, was passed on July 14, 1862. This act provided compensation for soldiers and the dependents of soldiers who had fought honorably for the Union and who had been wounded in such a way that they were unable to work. Over time, this pension system expanded into a form of old-age security for Union Army veterans and their families. Pension expenditures grew from $29 million in 1870 to $160 million by 1910, covering almost one million veterans and their dependents (Linares 2001). It is generally considered America's first large-scale social assistance program (Skocpol 1993; 1995).

Eligibility for a widow's pension under the General Law depended three main criteria. A widow was entitled to a pension if she did not remarry, and if her husband had served honorably in the Union army and died of a disease or injury sustained in the service. The qualifying widow of a private in the Union Army was entitled to eight dollars per month plus two dollars per minor child (under the age of 16) beginning on July 25, 1866.[14] To give a sense of the size of this income, a typical daily wage for a common laborer in the north was approximately one dollar in 1860 and two dollars in 1870; a farm worker would typically make 11to 15 dollars per month in 1860 and 18 to 20 dollars per month in 1870, which included room and board (Margo 2000).

The pension law was amended at various times. The most significant amendment was the act of June 27, 1890, which changed the eligibility requirements for both veterans and widows. Under this law, a widow could claim a pension if her husband had served honorably for at least 90 days in the Union Army, regardless of how he

[14]Glasson (1900; 1918); Song (2000). Officers' widows were entitled to a larger pension than widows, but the UA data contains only privates.

died. However, she had to demonstrate that she was "dependent upon her daily labor for support" (Linares 2001). Under the act of July 14, 1862, widows permanently lost their right to a pension if they remarried. However, later changes to the General Law altered this somewhat. As of June 7, 1888, a widow who had remarried could apply for a General Law pension in arrears, commencing on the date of her first husband's death and terminating on the date of her remarriage.[15] On March 3, 1901, a widow who was eligible under the General Law but had remarried was allowed to be restored to the pension rolls after her new husband died, provided she had never divorced this second husband, and she was needy. It became progressively easier for remarried widows to be restored to the rolls through the 1920s (Glasson 1900).

1.3.1 Procedures for Filing for and Collecting Pensions

The process of applying for pensions was costly and time consuming. In contrast to soldiers who filed pension claims, widows did not need to be examined by a surgeon; however, they were required to provide a great deal of evidence in support their claims. A widow had to appear before a court of record. If she lived more than 25 miles from a court of record, she could appear before a pension notary stationed in her locality (Oliver 1917). Here, she would make her declaration, which involved filling out a form in the presence of witnesses. The instructions attached to this form outline the information and documents she was required to furnish:

> She must prove the legality of her marriage, the death of her husband, and that she is still a widow. She must also furnish the names and ages of her children under sixteen years of age, at her husband's decease, and the place of their residence... The legality of the marriage may be ascertained by the certificate of the clergyman who joined them in wedlock, or by the testimony of respectable persons having knowledge of the fact, in default of Record evidence. (Widow's Certificate No. 8,336).

[15] ibid

This evidence was mailed to the pension bureau in Washington, DC, where claims were adjudicated. This adjudication process involved obtaining the soldier's military record from the war department. If a widow could not prove that she was legally married to the soldier or that his death was a direct result of his military service, her claim would be rejected.

In many instances, claimants hired attorneys to prosecute their claims. The quality of the attorney could have a dramatic effect on the speed with which a claim was processed; there are ample instances of claims pending for years because of attorney neglect, a problem well known to the pension board. The 1883 annual report of the pension commissioner condemns the behavior of these pension lawyers:

> There are certain ignorant, unscrupulous, and useless persons, whose only object seems to be, first, to procure applications from soldiers, regardless of merit, to be filed through them, and then, while acting simply as transmitters of the papers, assiduously dun the claimant until the ten-dollar fee is secured, and thereafter practically abandon the case (United States Pension Bureau 1883, p. 16).

Pensions were disbursed from agencies, located in cities and towns across the country. There were 33 such agencies in operation in 1863; by 1872, this had expanded to 57.[16] These agencies grew out of an existing infrastructure for distributing military pensions, inherited from the much smaller pension system already in place.[17] Payments were initially made semiannually, but this was increased to quarterly in 1870. Vouchers were drawn up and mailed from the pensioner's local agency. Upon receiving this voucher, the pensioner would fill it out and return it to the agency, which would mail back a check drawn on the U.S. treasury (Oliver 1917, p. 30).

[16] United States Pension Bureau 1864 and 1873. The agencies were generally considered inefficient and expensive (Oliver 1917; United States Pension Bureau 1883), and were reduced in number by the end of the 1870s (United States Pension Bureau 1883).

[17] See Glasson 1900 and 1918 for details. These pensions were for veterans (and dependents) of the Revolutionary War, the War of 1812, and the Mexican-American War.

1.3.2 Minors' Pensions

If a widow remarried, she lost her right to a pension. Entitlement to the pension then passed to the soldier's minor children, who were allowed to receive it until the youngest turned sixteen.[18] I have argued that pensions should only affect marital outcomes through widows' preferences: because they terminated upon remarriage, pensions should not make widows more desirable in the marriage market. However, if the widow's children were entitled to the pension when she remarried, she would, in a sense, continue to receive it. This means that she would be bringing an additional income stream into her new marriage, which might change the profile of matches available to her.

While the soldier's children were collectively entitled to the same monthly pension as the widow, there is variation in minors' pensions that is distinct from widows' pensions. This means that the effect minors' pensions had on widows' outcomes can be controlled for in the empirical analysis. This independent variation is due to several features of the pension law. First of all, minors' pensions terminated when the youngest child reached the age of sixteen; therefore, the lifetime value of these pensions was significantly lower than that of a widow's pension. There was an additional cost to obtaining a minor's pension: the children (or their guardian) needed to file a separate application, which took time to process. They also needed to obtain proof of their ages and legitimacy, as well proof that their mother was no longer eligible for the pension due to remarriage or death.

Finally, there were restrictions on the consumption of a minor's pension. These pensions were intended to be spent only on children's maintenance and schooling. Funds were paid directly to guardians, not to the children themselves; proof of guardianship had to be provided "under seal of the Court from which their appoint-

[18]One pension could be issued to all the soldier's children, which they would share.

ment is obtained" (Widow's Certificate No. 8,336). In some cases, the guardian was the widow or her new husband; in others, it was a third party. Even if the guardian was the widow or her second husband, there were steps taken to ensure that the pension was spent on the children and not on the guardian's consumption. In particular, the guardian had to account for the expenditure of the children's property at court. This requirement was laid out explicitly in many guardianship documents and in some cases codified in law. For example, a case from Michigan requires "a true account if the property of said ward in your hands" to be provided to the Probate Office "within one year from this date [December 10, 1867]" (Widow's Certificate No. 73,022). The law pertaining to guardianship in New York state required such an "inventory and account" annually (Legislature of New York 1837). In order to secure the guardian's obligations to his wards, he would post a bond with the county probate court. The pension cited above notes that the guardian rendered "a Bond with good and sufficient security to be approved by our said County Judge... in the penal sum of nine hundred" (Widow's Certificate No. 73,022). Another pension file includes proof of guardianship that describes a bond "in the penalty of fifteen hundred dollars conditional that the said [guardian] should faithfully, in all things, discharge the duties of a guardian" (Widow's Certificate No. 35,292).

Certainly, minors' pensions would have affected widows' outcomes, largely by rendering children less of a detriment in the marriage market. Potential husbands may have been more likely to propose to a woman with young children if these children were self-supporting. However, variation in minors' pensions that is independent of widows' pensions allows me to control for this effect in the empirical analysis. Specifically, I can control for potential minors' pensions using information on the number and ages of each widow's children.

1.3.3 Fraud

An obvious concern with using information about marital status from pension records is accuracy. Widows had a clear incentive to hide remarriages from the pension board, since disclosing this information would result in loss of pension. The incentive to fabricate marriages to veterans also existed. As the 1872 annual report of the pension commissioner remarks, "So long as pensions are to be granted upon evidence which (except record evidence) is purely *ex parte*, so long frauds will continue to exist" (United States Pension Bureau 1872, p. 13). The pension bureau was especially concerned about widows' claims: "The evidence to sustain a widow's or dependent's case is purely *ex parte*. As a result of this, a very considerable percentage of those cases are wrongfully established" (United States Pension Bureau 1872, p. 13).

If the pension authorities suspected a fraud, they would send a special examiner to the widow's place of residence to conduct an investigation. If found guilty of fraud, the widow lost her pension. Fraud was usually reported by either the postmaster who oversaw the delivery of pension vouchers and checks, or by members of the pensioner's community. There are a handful of examples in my sample of both sources reporting frauds[19]. However, notwithstanding the pension bureau's concerns about

[19] A letter of instruction to a special examiner in the case of Catherine Matthews describes allegations of remarriage by the postmaster of Malone, New York. The examiner is instructed to ascertain "whether the pensioner, by regular ceremony, by cohabitation, or by any other manner has performed such an act as will constitute marriage (re-marriage) under the laws of New York" (Widow's Certificate No. 6,916). Another example of fraud is the case of Maria van Buren, whose remarriage to Frank Stoffer is reported to the pension board by a close acquaintance. An excerpt from the examiner's report reads, "Stoffer had in his possession several letters, written in the same chirography, with the one hereto attached, none having a signature, all about equally dirty, but differing vastly in tone and purpose. The first a threatening message, demanding that she return to him by 7 o'clock and at least bid him farewell 'like a lady,' or he would have her in the penitentiary immediately. The next, breathing undying attachment of enormous dimensions, and asking her forgiveness for having 'told on her'. The third a sarcastic letter to Stoffer, and the fourth a letter of farewell and filled with threats of vengeance for her rejection of his 'ardent heart.' Mrs Van Buren acknowledged that she was living wtih Stoffer, and had done so 'off and on when she felt like it', but denied that she had married him, denied that he is Van Buren, who is now, she remarked, if not in heaven, certainly not on earth; denied that she intended to run away and professed several times an unusually strong desire to be arrested. I was, of course, satisfied that the case was not one

fraud, there is little evidence that hidden remarriages were a frequent occurrence. Women receiving pensions regularly interacted with the pension board throughout their lives; yet, in only about 15 out of the 500 cases analyzed in this study is there any indication of investigation into pension fraud. Moreover, only a few of these cases resulted in the widow being stripped of her pension. Still, to address concerns about fraud, I check marital status using links to the federal censuses of 1870 and 1880. Unless a large number of women were engaged in an elaborate fraud involving hiding second husbands from census enumerators, hidden remarriages or cohabitation do not appear to pose a significant problem.

1.4 Theoretical model

The aim of this paper is to assess the effect of an independent income source on women's choices about marriage. In this section, I describe a theoretical model to characterize the way in which such an income source should affect observable outcomes by altering the relative gains women perceived from marriage. A Civil War pension is income that a woman earns while she remains single but loses upon remarriage. As such, it is analogous to unemployment benefits in a search model of the labor market.[20] Thus, a simple search model of the marriage market is a natural framework for analyzing this question. I restrict the analysis to the female side of the market, which implicitly assumes that these pensions do not have general equilibrium effects on marriage market conditions. This is justified if the number of pensioners is relatively small.[21]

which I was authorized to further investigate without direct instruction" (Widow's Certificate No. 23,529). She was ultimately removed from the pension rolls because of remarriage, demonstrated by "cohabitation and recognition" (Widow's Certificate No. 23,529).

[20]See Rogerson, Shimer and Wright 2005 for a review.

[21]In fact, the number of pensioners was relatively small. There were just over 100,000 widows and other dependents on the pension rolls in 1872 (United States Pension Bureau 1872), and the number of dependents on the General Law pension rolls peaked in the early 1870s (Linares 2001). In 1870, the number of single women over the age of 17 was on the order of four million (Ruggles

The model is set up as follows. Unmarried women periodically receive proposals of marriage. A match generates value for the woman, and she must determine whether or not this exceeds the value she derives from remaining single. The value of staying unattached incorporates whatever flow utility she gets, as well as an "option value" of waiting for a potentially better match. If pension income raises the value of remaining single, this will raise the minimum match quality a woman will require in order to accept a proposal of marriage. An increase in this reservation match quality lowers the probability that a given match will be accepted, which will tend to increase the time spent searching. And, it will raise the average quality of a match, conditional on being matched at all. Simply put, if a woman is able to better support herself while single, she will be willing to wait longer for a better match.

In this model, I endogenize the frequency of marriage proposals. The effort women spend on search in the marriage market affects their outcomes: the more effort women allocate to search, the more frequently they receive proposals of marriage. However, search is costly. Because pensions raise the value of being single, women with pensions will tend to allocate less effort to the search process. This can be interpreted as an income effect: women "spend" a portion of this additional value on mitigating search costs.

Suppose there are two types of single women: those who receive a pension (indexed by P) and those who do not (indexed by N).[22] Married women are indexed by M. Assume for simplicity that there is no divorce.[23] A marriage generates flow utility θ,

et all 2010), putting the fraction of unmarried women on the pension rolls at no more than two or three percent.

[22]For the purposes of the model, I am assuming that women with and without pensions are otherwise identical.

[23]Allowing divorce does not qualitatively change the implications of the model. In any case, divorce was relatively uncommon. Preston and McDonald (1979) estimate that around six percent of marriages ended in divorce during the 1870s, compared to more than twenty percent in the 1950s. Work by Cvrcek (2009) demonstrates that this underestimates the true extent of marital separation: he estimates that ten to fifteen percent of marriages contracted during this period were disrupted, which is still a clear minority of marriages.

which is drawn from a distribution $F(\theta)$, and discounting occurs at a rate r. Each state, married or single, is associated with a lifetime expected value, V. For all women, the value of being in a marriage with match quality θ is given by:

$$rV^M = \theta$$

In words, this is the present discounted value of receiving utility θ forever. The value of being single is different for pensioned and unpensioned women. Suppose remaining single generates a flow utility s, and women with pensions receive additional utility p. Marriage proposals have a poisson arrival rate α, which depends on search effort. Specifically, it costs a widow $c(\alpha)$ in utility to obtain a rate of proposals α. I assume that costs are increasing and convex in α, so $c'(\alpha) > 0$ and $c''(\alpha) > 0$.[24] Then, the value to a pensioned woman of remaining single with proposal rate α_P^* can be written

$$rV^P = s + p - c(\alpha_P^*) + \alpha_P^* E[\max\{V^M - V^P, 0\}] \tag{1.1}$$

This is composed of two elements: the instantaneous utility a woman receives $(s + p - c(\alpha_P^*))$ and a term that reflects additional value, over and above the value of remaining single, from anticipated future proposals of marriage. It is a standard result that these unmarried women will have a reservation match quality, θ_P, which means they will accept any match carrying quality $\theta \geq \theta_P$. This has the property that $V^M(\theta_P) = V^P = \theta_P/r$. In other words, the reservation match quality is such that the woman is indifferent between remaining single and accepting the match. Substituting this into (1), and re-writing the expectation as an integral, we get the

[24] This standard assumption follows Mortenson (1986). It merely means that the marginal cost of search is increasing.

following equation that implicitly defines this reservation match quality:

$$\theta_P = s + p - c(\alpha_P^*) + \frac{\alpha_P^*}{r} \int_{\theta_P}^{\infty} (\theta - \theta_P) dF(\theta)$$

Women will choose α_P^* that maximizes the value of being unmarried. The maximizing level α_P^* will solve the following first order condition:[25]

$$rc'(\alpha_P^*) = \int_{\theta_P}^{\infty} (\theta - \theta_P) dF(\theta)$$

Similarly, for women who do not receive pensions, the reservation match quality is

$$\theta_N = s - c(\alpha_N^*) + \frac{\alpha_N^*}{r} \int_{\theta_N}^{\infty} (\theta - \theta_N) dF(\theta)$$

The optimal α_N^* is defined similarly to α_P^*. Notice that α_i^*, $i \in \{P, N\}$, does not depend directly on p. Instead, it depends on θ_i, which in turn depends on p. It is straightforward to show that θ_P is increasing and α_P^* is decreasing in p;[26] therefore, $\theta_P > \theta_N$ and $\alpha_P^* < \alpha_N^*$. In other words, women with pensions should be more selective and should spend less effort on search in the marriage market.

How are these differences manifested in observable outcomes? First, we can derive the rate of remarriage, which depends on both reservation match qualities and search effort. For a woman of type $i \in \{P, N\}$, the rate of exit from widowhood into marriage (H_i), or probability of remarrying at a given point in time conditional on staying single until then, can be written as

$$H_i = \alpha_i^*(1 - F(\theta_i))$$

This can be interpreted as the probability of both receiving a marriage proposal and

[25] See Mortenson (1986).

[26] See Mortenson (1986) or Rogerson Shimer and Wright (2005)

accepting it. Then, because $\theta_P > \theta_N$ and $\alpha_P^* < \alpha_N^*$, it follows that $H_P < H_N$. This means that the average time spent as a widow will be greater for women with pensions than without. Additionally, we have

$$E[\theta|\theta \geq \theta_P] > E[\theta|\theta \geq \theta_N]$$

Women receiving a pension have higher expected match qualities, conditional on being matched. This is simply because the minimum θ for women with pensions is higher.

In the empirical section of this paper, I will find it useful to specify a third group of women: those with a pending pension application. Suppose that, during an interval Δ, the (endogenous) probability of a woman with a pending claim receiving a proposal of marriage is $\Delta\tilde{\alpha}^*$, and the probability of having a claim decided is $\Delta\lambda$.[27] The probability that the decision will be favorable is given by π. Then, the value of being a widow with a pending pension claim can be written:

$$r\tilde{V} = s - c(\tilde{\alpha}^*) + \tilde{\alpha}^*\Big(E[\max(V^M - \tilde{V}, 0)]\Big) + \lambda\Big(\pi V^P + (1-\pi)V^N - \tilde{V}\Big) \qquad (1.2)$$

See appendix A for proof. Again, this is composed of three parts: the flow utility while single, additional value from future marriage proposals, and additional value from future pension rulings. Because V^M is strictly increasing in θ, the right hand side of this equation is also strictly increasing in θ. This implies that there exists a reservation match quality $\tilde{\theta}$ for women with pending pension applications. Then, we have the following equation that implicitly defines this reservation match quality:

$$\tilde{\theta} = s - c(\tilde{\alpha}^*) + \frac{\tilde{\alpha}^*}{r}\int_{\tilde{\theta}}^{\infty}(\theta - \tilde{\theta})dF(\theta) + \frac{\lambda}{r}\Big(\pi\theta_P + (1-\pi)\theta_N - \tilde{\theta}\Big) \qquad (1.3)$$

The optimal $\tilde{\alpha}^*$ will resemble that of the other two groups.

[27]This set-up follows Rogerson Shimer and Wright (2005).

Proposition 1: For $\pi \in (0, 1]$, $\theta_N < \tilde{\theta} < \theta_P$ *and* $\alpha^*_N > \tilde{\alpha}^* > \alpha^*_P$.

Proof: See appendix A. The intuition behind this is simple. Women with pending claims should have higher reservation match qualities than women receiving no pension with certainty because of the possibility of future pension income. However, they should have lower reservation match qualities than women whose claims have already been approved because of discounting and the possibility that the pending claim will be rejected. Again, the "income effect" coming from differences in the value of singlehood for these three types will generate differences in optimal search effort.

1.5 Data

1.5.1 Pension and Military Records

The data used in this paper comes from three main sources, two of which are newly collected from primary sources. The first data source is the Union Army (UA) database created by the Center for Population Economics (CPE) at the University of Chicago.[28] I have chosen a random sample of 500 women who were married to soldiers in the UA database. Useful for this study, this database provides information about soldiers' families, including when, where, and to whom they were married, as well as the birth dates and names of their children. I use this information to identify women that meet two important conditions. First, I restrict my attention to women wid-

[28]These data were collected as part of the project Early Indicators of Later Work Levels, Disease, and Death, sponsored by the National Institutes of Health and the National Science Foundation (Federal grant number P01 AG10120; see Fogel 2000). The data are drawn from three principal sources: the military, pension and medical records are compiled from sources at the National Archives including military service records and Civil War pension records; data from the Surgeons Certificates contain detailed information about veterans health status, which was used to determine pension eligibility; further socioeconomic information is gathered by linking veterans to the Federal Censuses of 1850, 1860, 1900 and 1910. These data have primarily been used to study health and aging in the late 19th and early 20th centuries. See for example Costa 1997, 1995, 1993; Fogel 2004; Eli 2010. They have also been used to analyze group dynamics in military settings (Costa and Kahn 2003, 2008). The data contain information about every soldier who enlisted in 303 randomly sampled companies of white volunteer infantry regiments. The database contains 39,341 observations and 3,230 variables (Fogel et al. 2000).

owed by 1880. This is because I expect such women to be most representative of the unmarried female population; they will be relatively young and thus more plausible marriage candidates.[29] I choose 1880 as a cutoff because it facilitates the linking of my observations to the 1880 census.[30]

The second restriction is that the widow had to apply for a pension within five years of her first husband's death. This restriction is intended to minimize sample selection bias due to limited data availability. Ideally, one would observe the widows of all soldiers in the UA database. However, because of the nature of this data source, the availability of spousal information depends on actions taken by subjects. For soldiers who died before 1880, all such information comes from dependents' pension applications, the vast majority of which are widows' applications. As such, it is extremely rare to know about widows who do not file for a pension at some point in their lives.[31] Women who first apply for a pension, say, ten years after widowhood will be those who had not applied earlier *and* had not remarried during those ten years. This will be a highly selected sample of *all* widows who did not file for a pension before ten years had elapsed. Given that my sample is necessarily restricted to applicants, there is a certain amount of selection that is unavoidable; however, I expect including late applicants to exacerbate this problem.

[29] Another consideration has to do with later amendments to the pension law. Under the General Law, the only requirement for pension eligibility was that a woman's husband served honorably in the Union Army and died from an injury or disease contracted in the service. However, following the act of June 27, 1890, a widow could receive a pension regardless of how her husband died, provided she could prove financial need. I expect financial need to be correlated with marital outcomes, more so than the details of a widow's first husbands death. So, it is beneficial to restrict the sample to women who could only have applied for a pension under the General Law, at least during the years immediately after widowhood.

[30] I cannot link widows to the 1890 census, because these manuscripts were lost in a fire. Linking to the 1900 census is less useful, as most Civil War widows were well past the age at which they could reasonably expect to remarry by 1900. The importance of census links is described later in this section.

[31] Soldiers on the pension in 1898 were required to inform the pension bureau of the name of their spouse and children. Before 1898, it is possible to have spousal information about a soldier if his widow never filed a claim but his mother or children did; however, this is quite rare.

The majority of the information I use in this paper comes from data that I collected from the Civil War pension files at the National Archives in Washington, DC. The CPE project focuses on soldiers' outcomes, so the UA database does not contain information about widows and children after the soldier died. After drawing my sample, I collect information about widows' pensions and marriage histories from their pension files. See appendix B for details of the data collection process. Because these data are compiled from historical records and not from surveys designed to avoid selection bias, the source of every piece of information is important. With this in mind, I will explain in detail where my most important variables come from.

The pension information is largely straightforward to collect, as any action a widow took with respect to pensions is recorded in her correspondence with the pension bureau. The case files contain all materials in a widow's pension application, which includes her application form and supporting evidence. If the widow was granted a pension, her file will contain both a pension brief and a pension certificate, indicating the amount of the pension, the effective start date, the date at which the pension was granted, the agency she was to be paid from, and the name of her attorney.[32] If the widow did not receive a pension, it is more difficult to determine why. In later years, rejected claims contain a brief indicating the date of and reason for rejection; however, during the years immediately following the Civil War, information about rejection merely consists of a stamp somewhere in the file that reads "rejected." In such cases, it is impossible to determine the reason for or date of rejection. Similarly, if a widow abandons her claim, we cannot be certain why or when.

Information about a widow's remarriage is slightly more complicated. Figure 1.1

[32]This information can be independently verified using the index to the pension files, which indicates the number attached to the widow's application and pension certificate. As these numbers are issued chronologically, the approximate date of application and issuance of the certificate can be inferred from these numbers.

illustrates the possible pension and marital outcomes for women in my sample. The first thing that occurs is the widow's pension application. After applying, the widow may remarry or die before her claim is adjudicated. Otherwise, she will receive a decision from the pension board, which may be favorable or not. After receiving this decision, the widow may or may not remarry. The outcome of a pension application is always certain; however, in 20 percent of cases it is impossible to determine whether or not the widow ever remarried.[33]

Table 1.1 lists possible sources of information about marital status and their frequency by pension status. A widow's remarriage is observable if her children file a pension claim or she applies to be restored to the pension rolls under the act of March 3, 1901.[34] A widow's failure to remarry is observable if her death date is known, and there is no indication of remarriage. If she is receiving a pension when she dies, her file will often contain a card indicating that she has been dropped from the pension rolls due to death. If not, this information may come from minors' pension applications or other correspondence with the pension board. Marital status is not observable if the widow stops communicating with the pension board some time before her death. Notice that the frequency of sources of information differs by pension status; this will be important to the sensitivity analysis I do later on.

Table 1.2 presents summary statistics from the pension file data I have collected (498 records in total). All women in this sample applied for a pension within five years of widowhood and had not remarried before doing so. The average age when widowed is 32; however, this ranges from 15 to 73. There are 397 women for whom remarriage status is certain, meaning that I observe them either remarrying or dying

[33]After around 1880, the pension bureau started including records of pensioners being dropped from the rolls for any reason. Women whose marital status is unknown are missing these records; thus, if they were on the pension, it is likely that they died, remarried, or stopped collecting their pensions some time before 1880.

[34]In some cases, a widow may have filed a claim for a pension she was not entitled to, or there may have been some other correspondence with the pension board indicating that she had remarried.

while single. There is no evidence that the other 101 women either remarried or died. Of these 397 women, 55 percent remarried at some point in their lives, which implies that the true fraction of women who ever remarried is between 44 and 64 percent. Of the 425 women for whom this information is available, 18 percent remarried before receiving a pension.[35] On average, a woman who remarried did so 4.3 years after her first husband's death. This average is much lower among women who remarried before getting a pension (2.4 years), which is unsurprising. It is, however, suggestive that the average time that elapsed between receiving a pension and remarriage is 3.9 years, which is much greater than 2.4 years.

The average amount of time that elapsed between the soldier's death and his widow filing for a pension was eight months, and the median was less than four months. The probability of ever having a General Law claim accepted was 0.86; however, fewer than 80 percent of women were receiving a General Law pension within five years of applying. The average processing time for a pension was more than two years, although this is highly skewed: the median processing time is slightly less than one year. Most women in my sample were first married during the 1850s and were widowed during the war. These women tended to come from the Mid Atlantic region (30 percent) or the East North Central region (41 percent). Very few come from Southern or Western regions.

1.5.2 Census Links

I use information from the pension file data to link my observations to the federal censuses of 1870 and 1880. These links are important because they provide information about widows' second marriages. In the pension file data, such information is available in a minority of cases, which makes it difficult to evaluate the effect of pen-

[35]Even if I do not know whether or not a widow *ever* remarried, I may know that she did not remarry with a pending claim if she communicated with the pension board subsequent to her claim being granted.

sion income on match quality. Another reason for linking widows to the census is that it provides a check on the marriage information available in the pension data. For one thing, these links allow independent verification of widows' marital status, which alleviates concerns about inaccuracies due to fraud. These links also help mitigate concerns about missing data.

As explained above, although marital status is known in most instances, it is unknown for 20 percent of my sample. A concern is that the availability of information about marital status is not random, and this might bias my results. A remarried widow must do one of two things to be identified in the pension data: she must have young children who apply for a minor's pension after she remarries; or, she must survive long enough to apply for a pension under the act of March 3, 1901. Women who do not remarry do not need to meet these restrictions in order to be observed. Therefore, my sample of remarried widows may be younger and healthier than my sample of widows who do not remarry, simply by virtue of the way the data are collected. If the effect of the pension on marriage behavior depends on age or health, this sample selection might bias my results.

Identifying widows with uncertain marital status through census links is a challenge: if a widow did remarry, her last name would have changed. I use an alternative method for linking these ambiguous cases. The names and ages of children from the widow's first marriage are available in the pension data, so I can link these children to the census; in principle, a child's surname would not change if his or her mother remarried.[36] If I locate a child who is living with a married mother with a different last name (but whose birth year and first name match the widow in my sample), I assume that I have identified a remarried widow. See appendix B for further details. Data collected this way will still favor women with young children; however, this will

[36]The availability of information about children does not impart additional bias, as all widows were required to list minor children in their pension applications; thus, this information is available for every widow who made a pension application.

apply equally widows who have remarried and those who have not. These data may generate other biases. For example, a remarried woman may be less likely to keep her children living at home, so I might underestimate the fraction of widows who remarry. Still, because the availability of these data does not depend on details of the pension application process, they will be a useful complement to the pension data.

Table 1.3 presents statistics on the success rate of this procedure. The top panel lists the fraction of widows who were linked to the 1870 and 1880 census, overall and by marital status. The linkage rate is quite high overall, close to 60 percent in both years. In 1870, the linkage rate is higher among widows who are known to have remarried (69 percent) than it is among women who are known to have remained unmarried (63 percent); in 1880, the linkage rate is higher among women known not to have remarried (76 versus 68 percent). The fraction of widows with uncertain marital status who were successfully linked through children from their first marriage is much lower (18 to 27 percent); however, this partly reflects the fact that some of these women are childless. Among women who may theoretically be linked this way, 26 to 37 percent were located successfully.[37] The vast majority of widows with unknown marital status turned out to be unmarried: only one had remarried by 1880.

The bottom panel of table 1.3 contains the fraction of widows who were theoretically "linkable" through children from their first marriage. This is to get a sense of the effectiveness of my strategy for linking widows with unknown marital status. In fact, a large number of widows, both married and unmarried, reside with children who have kept their deceased father's surname. In 1870, 88 percent of unmarried widows and 52 percent of married widows live with such children. In 1880, these fractions

[37]One reason for the linkage rate for these women to fall below the linkage rate for women with known marital status is that, for women with unknown marital status, I have little information on place of residence in 1870 or 1880; these women have largely disappeared from the sample by this time. Note that these linkage rates still compare favorably to other projects that create samples of linked census data. See Ruggles et al (2010) and Ferrie (1996).

are 80 and 44 percent, respectively. This decline in the fraction of women who are linkable through children is likely caused by the increasing tendency for children to leave home as they age. While at appears that linking widows through their children does underrepresent those who have remarried, a significant fraction of such widows can still be linked.

1.5.3 Representativeness

In order for a widow to appear in my sample, she must satisfy two conditions. First, she must have been married to a Union Army soldier who died before 1880; second, she must have filed an application for a pension. In this section, I investigate the extent of the bias introduced by the decision to apply for a pension, which will be important when considering what these results imply about all women, or even all Civil War widows. A natural starting point is to establish the fraction of women widowed before 1880 who ever made pension applications. Recall that spousal information comes almost exclusively from widows' pension applications, so I will treat making an application and appearing in the pension data as interchangeable[38].

To know for certain the fraction of women widowed by 1880 who made pension applications, we need both a numerator and a denominator. More precisely, we need three pieces of information: (i) the number of women widowed by 1880 who made pension applications; (ii) the number of soldiers who died before 1880; and (iii), how many of these soldiers were married. We know (i) but not (ii) or (iii). In order to establish a lower bound estimate of the application rate among women widowed by 1880, it is necessary to make assumptions about missing data. Table 1.4 contains some of these estimates. Out of a sample of 39,341, we know for certain that 7,953 soldiers died before 1880. Of these, we know that 3,102 were married because there

[38]As described earlier, spousal information before the early 1900s was collected through dependents' pension applications, so it was very unusual to have this information if no pension application was submitted.

is spousal information in the UA data; we also know that 714 were not married. If the 7,953 soldiers whose death dates are known to be prior to 1880 constitute a fully representative sample of all soldiers who died before 1880, it would be reasonable to infer the application rate among women widowed by 1880 was at least 45 percent.[39]

However, these 7,953 soldiers are almost certainly not a representative sample of soldiers who died by 1880, because knowledge of a soldier's death date is highly correlated with his widow making a pension application. To see this, notice that 95 percent of soldiers with missing death dates also have missing spousal information. This is because information on death dates for soldiers who died prior to 1880 often comes from widows' pension applications. So, in order to establish a lower bound on the fraction of widows who appear in the data, we must allow for the possibility that some soldiers with missing death dates died before 1880. Depending on the reference group and assumptions about the fraction of soldiers who were married, I derive reasonable lower bounds that range from 17-46 percent.[40]

Using only soldiers who died during the war as a reference group provides a potentially more reliable lower bound estimate of the true application rate. Information about death dates of soldiers who died *in the service* can be obtained from sources other than widows' pension applications, such as military or hospital records. Thus, it is more reasonable to treat these soldiers as a random sample of casualties, with respect to widows' pension applications. If we assume that the overall Union Army casualty rate of 16 percent (Costa and Kahn 2008) prevailed in this sample, the lower

[39]This lower bound assumes that every soldier with missing spousal information was married.

[40]In calculations using soldiers dead by 1880 as the reference group, I assume that all soldiers with missing death dates died before 1880, which is quite conservative. In the most conservative calculation, I assume that all soldiers with missing marital status were married; in another, I use an imputed marriage rate for these soldiers. This imputation is based on a regression of marital status on age, state, and occupational class dummies using the 1860 one percent IPUMS sample. The imputed marriage rate is the predicted fraction of UA soldiers who would have been married in 1880 (the most conservative death date assumption for individuals with unknown death dates), using the coefficients from the above regression.

bound ranges from 28-46 percent, depending on assumptions about the marital status of men with missing spousal information. Based on this calculation, a lower bound application rate of about one half is reasonable.

While a large fraction of women widowed by 1880 made pension applications, it seems likely that not every widow did so. The next question is: how did women who made pension applications differ from those who did not? Establishing this is complicated by the fact that women who never made pension applications do not appear in the pension file data. However, the UA data contains links to the 1860 federal census.[41] Using these links, I infer the soldier's marital status from the composition of the household in which he resides.[42] I compare soldiers who were married in 1860 and whose wives appear in the pension data with those whose wives do not appear. I restrict the sample to men who died during the war, for reasons explained above.

Table 1.5 contains these results. Column (1) contains the mean of each variable among wives who appear in the pension data, and column (2) contains the mean among wives who do not. Column (3) presents the difference in means between these two groups. Column (4) contains an OLS regression of an indicator for appearing in the pension data on all of the variables in the table. These results provide strong evidence for selection on the basis of marriage prospects or affluence. Women who file pension claims tend to be older and to come from less wealthy households. Their husbands are more likely to be illiterate. These husbands are more likely to hold skilled blue collar occupations, such as craftsmen and skilled factory operatives, and

[41]These data strongly favor men whose wives appear in the pension data, as this information was used to make the links. However, this is the only information I can provide here.

[42]I call household occupants "potential wives" if they are female, less than 15 years older or 30 years younger, and have the same last name as the soldier. This is somewhat more conservative than the IPUMS procedure for imputing spousal relationships; this procedure uses 10 and 25 year cutoffs, respectively (Ruggles et al 2010). If the soldier is a household head and the second household member is a potential wife, I assume he is married. If he is not a household head, I infer marital status from the relative position of potential wives and potential children in the household using standard rules for imputing family interrelationships (see Ruggles et al 2010).

are are less likely to be skilled professionals or proprietors. Notice that the regression coefficient on the wife's age is negative, while the coefficient on the soldier's age is positive and larger in magnitude. This reflects the high correlation between the ages of husbands and wives, and can be interpreted to mean that women who were married to older men were more likely to apply for a pension.[43]

These apparent differences between pension applicants and non-applicants have no bearing on the internal validity of this study. However, they are important to keep in mind when extrapolating the results to the general population. I will discuss this further after presenting my empirical findings.

1.6 Pensions and the Timing of Remarriage

1.6.1 Empirical Framework

In this section, I describe my approach to evaluating the extent to which pension income slowed the rate of remarriage among Civil War widows. This is a challenge because pension amounts are standardized, so there is no variation in pension income among pensioners. Moreover, it is not straightforward to compare women who had pensions to those who did not, as I do not observe women who never make pension applications. The are two possible sources of variation in pension income: the pension board's decision and the timing of this decision.

The pension board's decision is not an ideal source of variation for a few reasons. First, this variable is only defined for women who complete their claims. Recall from figure 1.1 that at least twelve percent of my sample remarried while their claims were pending. A simple comparison between women with accepted and rejected claims will discard this potentially valuable information. Another issue is that rejections take significantly longer to process than acceptances. It takes approximately five

[43] If husband's age is omitted from the regression, the coefficient on wife's age becomes positive and highly significant.

years longer to reach the “rejected” node in figure 1.1 than the “accepted” node. Thus, my sample of rejected widows ought to look very different from the universe of potentially rejected widows, as many of these are likely to have remarried before the board’s decision was rendered. A final technical issue has to do with accuracy: it is often unclear when or why a claim was rejected.

Because of these issues, I use variation in the timing of the pension board’s decision, rather than the outcome, to estimate the effect of pensions on the timing of remarriage. Specifically, I look for a treatment effect of having a pension claim granted, or of transitioning from having a pending claim to an accepted claim. Recall from section 4 that women with pending claims should behave differently from women who have their pensions in hand, due to discounting and the possibility of rejection. I estimate a proportional hazard model of both pensions and marriage, allowing the rate of remarriage to shift at the moment a pension is granted. Variation in processing times allows me to observe women with and without pensions at every point in time, which allows me to estimate a hazard rate of remarriage that differs by pension status.

Some of this variation is plausibly exogenous. For example, idiosyncrasies in the postal service, clerical errors, or unexpectedly capricious behavior on the part of pension attorneys certainly affected processing times in a random fashion. However, a portion of the variation in processing times is likely endogenous to marital outcomes. For example, women with poor marriage prospects may have been more invested in getting a pension because they knew their alternatives were poor. So, those who got pensions quickly may have tended to remarry slowly because of poor marriage prospects, not because of a causal effect of the pension. Another concern is that processing times are highly correlated with the quality of a pension claim: rejections take significantly longer to process than acceptances.

Why is this a threat to identification? If we accept that pension eligibility is random, then the ambiguity of a claim should be similarly exogenous. However, bias may be introduced by the decision to apply. Applying for a pension is costly: a widow will choose to incur this cost if the benefit is great enough. The expected benefit from applying is lower for a widow with an ambiguous claim, as the probability of ever receiving a pension is low. Thus, women who apply with ambiguous claims may be systematically different from women who apply with straightforward claims. In particular, they may have worse alternatives, either financially or in the marriage market. The direction of this bias on the timing of remarriage is unclear: women with poor alternatives might receive fewer proposals per unit of search effort; however, they might also be less selective.

To overcome these endogeneity problems, I use a method developed by Abbring and Van den Berg (2003a). This is a novel approach to identifying treatment effects in the presence of an endogenous treatment when both the treatment and outcome are duration variables. The approach involves jointly estimating the hazard rates of pensions and remarriage, allowing for correlation between the unobserved heterogeneity in these two risks. The hazard rate at time t refers to the probability of realizing an outcome (pension or marriage) at t, conditional on not having realized it earlier. The hazard rate of pension income is given by

$$\theta_p(t|X, v_p) = \lambda_p(t)\exp(X\beta_p + v_p) \tag{1.4}$$

and the hazard rate of marriage is given by

$$\theta_m(t|X, v_m, t_p) = \begin{cases} \lambda_m(t)\exp(X\beta_m + v_m) & \text{if} \quad t \leq t_p \\ \lambda_m(t)\exp(X\beta_m + \delta + v_m) & \text{if} \quad t > t_p \end{cases} \tag{1.5}$$

For each $i \in \{p, m\}$, λ_i is the baseline hazard function, which characterizes duration dependence, and X is a matrix of explanatory variables that may shift the hazard rate. The term t_p represents the time at which a pension is granted, and v_i reflects unobserved heterogeneity.

Allowing for duration dependence ($\lambda_i(t)$) and the effect of covariates ($X\beta_i$) is crucial to the identification of δ. Duration dependence refers to the way in which the hazard rate changes over time; for instance, whether marriage becomes more or less likely as time passes. Failing to account for duration dependence will bias the estimate of δ. For example, suppose there is negative duration dependence in the rate of remarriage, so the probability of remarrying declines with time in the marriage market. Then, women will appear to remarry at a slower rate upon receiving a pension, simply because these women will have been in the marriage market longer. Thus, we will overestimate δ. Failure to account for observables will bias the estimate of δ to the extent that these are correlated with pension status. For example, suppose the hazard rate of pension receipt increases with age, and the hazard rate of marriage declines with age. If we do not control for age when estimating δ, the estimate will be biased away from zero, as women who receive pensions quickly will tend to be older, and these women will tend to remarry slowly.

Every concern I have just described applies to a standard proportional hazards model. An additional issue that arises in this particular setting is the possibility that v_m and v_p are correlated. For example, if v_m and v_p are negatively correlated, the estimate of δ may be negative even if the true δ is zero. Correlated unobserved heterogeneity generates bias in a similar fashion to omitted observable controls. If women who get pensions quickly tend to have large v_p, they will also tend to have small v_m, which means they are likely to take longer to remarry.

Abbring and Van den Berg (2003a; 2003b) show that this model is identified even if v_m and v_p are correlated. Moreover, it is identified without exclusion restrictions or assumptions about the functional form of either the baseline hazard or the joint distribution of the unobserved heterogeneity terms. The unobserved heterogeneity directly affects the *rate* of treatment but not the precise timing of treatment. Put another way, a high v_p raises the probability of receiving a pension at time t; however, there remains a stochastic element to which event, pension or no pension, actually occurs at time t. The problem is disentangling this random assignment from the non-random assignment.

To understand how this is possible, first notice that, in a simple proportional hazards setting, the distribution of unobserved heterogeneity is identified from variation in observables. To see this, consider the rate of pension receipt. Suppose one woman has a very good pension attorney (high $X\beta_p$), and a second woman has a poor pension attorney (low $X\beta_p$). Now, suppose these two women both take a long time to receive a pension (large t_p). We can infer from this that the probability that the first woman has an ambiguous pension claim (low v_p) is higher than it is for the second woman. In general, the distribution of v_p, conditional on t, depends on observables, which allows its distribution to be pinned down.

How does this help us identify correlated unobserved heterogeneity in the rates of remarriage and pension receipt? Using the same example, suppose that the quality of pension attorney has no direct effect on the rate of remarriage, so women with good and bad pension attorneys have the same $X\beta_m$.[44] This means that we should not expect to see systematically different marital outcomes by the quality of pension lawyer. However, recall that, conditional on t, the distribution of v_p is not independent of the quality of pension lawyer. So, if v_m and v_p are correlated, the distribution of v_m will similarly be dependent on pension lawyer quality. Say v_m and v_p are negatively

[44]This example is used for clarity and does not imply the necessity of an exclusion restriction for identification. In general, as long as $\beta_m \neq \beta_p$ and there is sufficient variation in the data, there exists some X, X' such that $X\beta_m = X'\beta_m$ but $X\beta_p \neq X'\beta_p$. This is all that is required. Also notice that the values of β_m, β_p are identified using "early" parts of the sample, when v_m and v_p are independent of observables. This dependency arises "later" in sample, due to selective sample attrition.

correlated, and recall that, fixing t, $E(v_p)$ is higher for women with bad lawyers than it is for women with good lawyers. This means that, among women who are in the sample at time t, those with good lawyers will tend to remarry fastest, because these women tend to have higher v_m. Similarly, if v_m and v_p are positively correlated, women with bad lawyers will tend to remarry more quickly. In other words, different joint distributions of v_m and v_p will be observationally distinct. Once the correlation between v_m and v_p has been corrected for, the remaining difference between the marriage rate before and after a pension is granted can be interpreted as a causal effect of the pension.

I estimate this model by maximum likelihood. To explain the estimation process, I define a series of functions that are elements of the likelihood function. The survival function, or the probability of remaining a widow (m) or not having a pension (p) at time t, is denoted $S_i(t)$, and it has the following form:[45]

$$S_i(t) = \exp\left(- \int_{t_0}^{t} \theta_i(s) ds \right), \qquad i \in \{m, p\}$$

If t is a random variables denoting time an event occurs, its density is given by

$$f_i(t) = \theta_i(t) S_i(t)$$

So, the likelihood of an event occurring at t depends on both the hazard function and the survival function. For pensions, the survival function is straightforward to define:[46]

$$S_p(t|X, v_p) = \exp\left(- \int_{t_0}^{t} \lambda_p(t) \exp(X\beta_p + v_p) \right)$$

The survival function for marriage is somewhat more complicated, because it shifts

[45] See Lancaster (1990).

[46] This construction follows Abbring and van den Berg (2005), who apply this model to evaluating the effect of unemployment insurance sanctions on the rate of transition to employment.

at a point in time. The survival function before and after receiving a pension are given by the following two equations, respectively:

$$S_{m,1}(t|X,v_m) = \exp\left(-\int_{t_0}^{t} \lambda_m(t)\exp(X\beta_m + v_m)\right)$$

$$S_{m,2}(t|X,v_m,t_p) = S_{m,1}(t_p|X,v_m) \times \exp\left(-\int_{t_p}^{t} \lambda_m(t)\exp(X\beta_m + \delta + v_m)\right)$$

To understand the definition of $S_{m,2}$, consider the meaning of its two parts separately. Suppressing X and v_m, the first term reflects $Pr(t_m \geq t_p)$, and the second term reflects $Pr(t_m \geq t | t_m \geq t_p)$.

There are four possible outcomes for women in the sample, which I index by $k \in \{1,2,3,4\}$. A woman can remarry before she gets her pension ($k = 1$); she can remarry after her claim is granted ($k = 2$); she can be censored before her claim is granted, meaning that she dies or disappears from the sample ($k = 3$); or she can be censored after her claim is granted ($k = 4$). Each of these events is associated with a different likelihood. Conditional on her unobserved heterogeneity terms, the likelihood contribution of woman i can be written as

$$L_i(t) = \begin{cases} \theta_m(t|X,v_m,t_p)S_{m,1}(t|X,v_m)S_p(t|X,v_p) & \text{if} \quad k=1 \\ \theta_m(t|X,v_m,t_p)S_{m,2}(t|X,v_m,t_p)\theta_p(t_p|X,v_p)S_p(t|X,v_p) & \text{if} \quad k=2 \\ S_{m,1}(t|X,v_m)S_p(t|X,v_p) & \text{if} \quad k=3 \\ S_{m,2}(t|X,v_m,t_p)\theta_p(t_p|X,v_p)S_p(t_p|X,v_p) & \text{if} \quad k=4 \end{cases}$$

To estimate this model, I make certain parametric assumptions about the baseline hazard rate and the joint distribution of the unobserved heterogeneity terms, v_m and v_p. I attempt to make the least restrictive parametric assumptions possible. For the baseline hazard, I use a piecewise constant function, where time is divided into discrete "bins," and $\lambda(t) = \lambda_t$ takes on some unrestricted value for each of these bins. I use bins of one year, with a single bin for the tail of the time distribution, extending from $t = 8$ until the last observation leaves the sample. Following eight years after widowhood, first marriages and pensions occur with insufficient frequency to identify hazard rates at finer intervals.

For the unobserved heterogeneity terms, I assume a discrete distribution in which both v_m and v_p have two unrestricted mass points:[47] $v_m \in \{v_m^{low}, v_m^{high}\}$ and $v_p \in \{v_p^{low}, v_p^{high}\}$. Thus, there are four possible combinations of v_m and v_p, each of which is associated with a certain probability. The location of each of these mass points and the probability of each combination of the two are estimated in the model. A discrete distribution is considered the most flexible parametric assumption that can be made about the joint distribution of unobserved heterogeneity terms, as it allows any correlation between the two variables to be achieved; other assumptions, like allowing unobserved heterogeneity terms to take on infinite values that follow a set distribution, restrict these correlations.[48] A discrete distribution with more than two mass points is not feasible with the sample size I am working with.[49]

Intuitively, this particular about the distribution for v_m and v_p means that women may be one of two "pension types" and one of two "marriage types." Meaning, a woman can be likely or unlikely to get a pension quickly, and she can be likely or unlikely to remarry quickly. The main threat to identification is that "high" pension types may tend to be "low" marriage types, and vice versa. If this is the case, then even if pensions have no true effect on marriage rates, I might estimate such an effect simply because women who remarry quickly also take longer to get their pensions.

Estimating a model that accounts for unobserved heterogeneity is complicated because the heterogeneity is unobserved, which means that I cannot calculate the correct likelihood contribution of each observation. To estimate the model, I use the

[47] This follows an application of this model by Abbring and Van den Berg (2005).

[48] Heckman and Singer (1984); Abbring and Van den Berg (2005); Van den Berg (1996).

[49] Notice that the number of parameters increases exponentially with each additional mass point in the distribution of v_m and v_p, as any combination of these two variables must be allowed to occur.

EM algorithm.[50] This procedure does the following. I start with a vector of parameters, ϕ_0, which includes $\delta, \alpha_m, \alpha_p, \beta_m, \beta_p$, $v = (v_m^{low}, v_m^{high}, v_p^{low}, v_p^{high})$, and probability weights, $\pi = (\pi_1, \pi_2, \pi_3, \pi_4)$, associated with each of the four unobserved heterogeneity "groups" my observations may fall into. Using these values, I construct a set of weights for each observation:

$$\delta_{i,j}^0 = \frac{\pi_i^0 L_{ij}^0}{\sum_{k=1}^4 \pi_k^0 L_{kj}^0}$$

The letter j indexes the individual, and i indexes the unobserved heterogeneity group. Given the data and parameter choices, this reflects the probability that individual j falls into group i. I fix these weights, and then construct an expected log likelihood function, which I maximize over ϕ to obtain ϕ_1. Based on ϕ_1, I construct a new set of weights, δ^1, and repeat the process to convergence.

1.6.2 Results

Before presenting estimates of the model described above, it is useful to get a sense of what the hazard rates of remarriage and pension receipt look like. Figure 1.2 plots the empirical hazard rate of both pensions and remarriage, estimated non-parametrically using a kernel method.[51] The top panel illustrates the rate of remarriage measured before and after a pension is granted; the bottom panel illustrates the hazard rate of pension decisions. Time is measured in years since widowhood; however, individuals do not enter the sample until they apply for a pension. Notice that, for the first five years, the rate of remarriage for women who have not yet received a pension lies uniformly above that of women who have pensions. After five years, the two

[50]This is frequently used procedure, which was developed to deal with missing data. See Heckman and Singer (1984) and Lancaster (1990).

[51]This is done using the STS package in STATA. For ease of comparison, I truncate this graph at $t = 10$. This is because it becomes impossible to estimate the rate of remarriage for women without pensions for later periods, as there are insufficient observations.

lines are very close together. This may indicate that the pension only lowers the rate of remarriage in the short run; however, it may also reflect differences in the characteristics of pensioned and unpensioned women in later years. Women who are still in the sample without pensions, say, ten years after widowhood are those who are still trying, unsuccessfully, to get a pension after ten years. These women may have very different characteristics, either observable or unobservable, than women who are in the sample without pensions only a year or two after widowhood. It is also worth mentioning that the sample of women without pensions becomes very small as time passes. For instance, there are only 27 such women in the sample more than five years after widowhood.

Table 1.6 contains parameter estimates for the model described above, with the estimated effect of covariates on the rate of pension receipt listed next to their estimated effect on the rate of remarriage. In column (1), I estimate the model with no covariates or correction for correlated unobserved heterogeneity. In this specification, the estimated effect of the pension is negative, but it is not significantly different from zero. In column (2), I add covariates to the hazard rate of both risks, which significantly increases the magnitude of the estimate, to -0.49 (0.19). This suggests that selection on observables biases this effect toward zero. Recall that this bias could go either way. Women who experience long processing times are likely to have ambiguous claims, and women who apply with ambiguous claims may be different from those who apply with straightforward claims. If these women are less wealthy, for example, it may be more difficult for them to receive marriage propsals; however, they may also be less selective. These results suggest that observable characteristics of women with ambiguous claims tend to slow the rate of remarriage, leading to an underestimate of the effect of the pension when these controls are omitted.

In column (3), I introduce the possibility of correlated unobserved heterogeneity in the rates of pension receipt and remarriage. At -0.54 (0.22), the estimated effect of the pension changes little from the previous specification, suggesting that much of the selection problem is captured by the controls for covariates. The estimate from the full model can be interpreted to mean that receiving a pension lowered the hazard rate of remarriage by approximately 40%.[52] This estimate implies that, for a woman with median characteristics, immediately granting her a pension would raise her median time to remarriage from 4.7 to 7.8 years, an increase of more than three years.[53] This timing increase is consistent with the summary statistics from table 1.2, although the implied medians are substantially higher than they are in this table, as they should be. These summary statistics are calculated using women who actually remarry. The medians implied by the model estimates incorporate information from women who never remarry, which will tend to raise them substantially.

Other variables affect the rate of remarriage in plausible ways. Older women tend to remarry more slowly, as do women with more children. The year of widowhood has a negative effect on the rate of remarriage, which may reflect sample selection, as claims become more ambiguous the farther removed is the soldier's death from the war. Characteristics of the widow's first husband have some effect on marriage rates: women who are married to older and shorter men tend to remarry more quickly. This latter finding could reflect women's reservation match qualities, especially if height is positively correlated with socioeconomic status. The county male to female ratio speeds up remarriage quite significantly, which is to be expected. The only variable that significantly affects the hazard rate of pension income is year of widowhood, which presumably reflects the fact that claims become more ambiguous with distance

[52] This comes from the fact that $\theta^{PEN}/\theta^{NOPEN} = \exp(-0.54) = 0.58$, so $\frac{\theta^{PEN} - \theta^{NOPEN}}{\theta^{NOPEN}} = -0.42$.

[53] For women with pensions, this calculation is done by solving the following for t_{med}:

$$0.5 = Pr(t \geq t_{med}) = S_2(t_{med}|X, v_m)$$

For women without pensions, I do the same calculation, replacing S_2 with S_1. For X, I use median characteristics and mean regions; I integrate over v_m and v_p using estimates from the model.

from the war. There are also regional differences: claims from the New England seem to be processed significantly faster than claims from the Mid-Atlantic, the Midwest or the South.

The parameters of $\lambda_m(t)$ and $\lambda_p(t)$ are also listed in table 1.6, with λ_m and λ_p on the interval [0,1) both normalized to 1. These estimates suggest non-monotonic duration dependence in both risks. In both cases, the hazard rate initially increases and then falls. One can imagine plausible explanations for this pattern in the hazard rate of marriage. The rate of remarriage may rise in the short run if women lower their reservation match qualities as time passes, either due to revised expectations or changing preferences for matching. However, this rate is likely to fall eventually if part of what makes women desirable in the marriage market is fertility. In the case of pensions, this pattern may reflect changes in the composition of claims as time passes. Among very straightforward claims, the probability of receiving a pension is likely to increase with processing time. However, at some point, all straightforward claims will have been processed, leaving only ambiguous ones. The probability of ever getting a pension with an ambiguous claim is low.

The unobserved heterogeneity terms are quite imprecisely estimated. Notice that the two estimated values of v_p are very close to one another, and the probability weights attached to each unobserved heterogeneity group have very large standard errors. This may indicate that unobserved heterogeneity in the rate of pension receipt is well controlled for by covariates and the duration dependence function, leaving few systematic unobserved differences.

1.7 Sensitivity Analysis

1.7.1 Instrumental Variables Analysis

The hazard model described in section 6 is the most exact representation of the relationship between the receipt of pensions and the rate of remarriage. However, a concern is that the estimates may be sensitive to some of the parametric assumptions made in estimation. So, as a complement to the analysis in section 6, I include a linear analysis of the relationship between pensions and the timing of remarriage.

Using a series of time frames ranging from one to five years ($\tau \in \{1, 2, 3, 4, 5\}$), I create an indicator variable equal to one if a widow had received a pension within the time frame ($I(t_p \leq \tau)$) and an indicator equal to one if she had remarried within the time frame ($I(t_m \leq \tau)$). I estimate the following by OLS:

$$I(t_m \leq \tau) = \alpha + \beta I(t_p \leq \tau) + X\gamma + u$$

The matrix X includes all controls used in section 6. I expect to find $\beta < 0$. Here, the endogeneity problem is quite severe: many women who were not receiving pensions within, say, three years of applying had been denied pensions *because* they had remarried. I use instrumental variables to circumvent this problem.

Details of the application and review process provide potentially valid instruments for pension income.[54] The instrument that I use is based on the spelling of last names.

[54] This approach is similar in spirit to Maestas, Mullen and Strand (2011) who use spending allowances of the examiners assigned to individual cases as an instrument for disability insurance to identify a causal effect of disability insurance on labor supply. An alternative possibility follows Eli (2010), who uses political variables as instruments for pension income. This approach uses the observation that Union Army pensions were used to secure votes for the Republican party (Eli 2010; and Skocpol 1993), so pension amounts would be inflated in contested congressional districts. It is conceivable that pensions also would have been processed more quickly in politically expedient areas, so political variables may be valid instruments in this case. I do not make use of these variables for several reasons. For one thing, women could not vote, so expediting widows' pensions would have been less politically beneficial for the Republican party. Still, one could make the argument that generosity with widows' pensions may have generated good will among male veterans. However, the period during which pensions were widely used as political patronage occurred later, largely in the

As described earlier, to receive a pension a widow had to prove that she was married to a soldier, that he served honorably in the military, and that his death was connected to the service. This involved locating military service records, hospital records, and marriage certificates. If there were discrepancies in the spelling of his name in these records, additional steps were required to demonstrate that the records referred to the same individual. In the pension files, there are examples of secondary affidavits explaining name spelling discrepancies.

I construct an indicator of name spelling homogeneity from the one percent IPUMS samples from 1860, 1870, and 1880. I compile a list of all household heads in each of these years, and I group last names by codes generated using the NYIIS algorithm (Atack and Batemen 1992). Frequently used to create linked census samples,[55] this algorithm collects names into phonetically similar groups. I construct a Herfindahl index of the dispersion of unique name spellings within these phonetic groups. Greater values indicate that there is little variation in name spelling; smaller values indicate that names in this group are spelled in many different ways. I perform two tests of the validity of this measure. First, I check whether or not a low name homogeneity index predicts multiple spellings of the veteran's last name in the pension data. I find that a one standard deviation increase in this index raises the probability of observing multiple surname spellings in the pension data by 8.5 percentage points; this is highly significant. Second, I check whether or not a name with a high homogeneity index is more likely to exactly match the most common spelling in its phonetic group in the census. Again, I find that a one standard deviation increase in the index raises the

1870s and 1880s. The majority of my sample was widowed during the war and applied for a pension before 1870. Thus, political variables ought to explain little of the variation in their pension outcomes. I have experimented with using county-level election variables as instruments in this context, and they are unable to explain a satisfactory amount of variation in pension outcomes. Granted, county-level variables are an approximation: the appropriate unit of analysis is the congressional district. Still, my sample is predominantly rural, so the approximation should be a good one.

[55] Ferrie 1996; Abramitzky, Boustan and Eriksson 2010.

probability of such a match by 25 percentage points, which is also highly significant.

A concern is that this measure may not be exogenous to marital outcomes. Names that belong largely to immigrants may be spelled in multiple ways; immigrant status is likely endogenous to marital outcomes. Names that belong to lower socioeconomic status families may be frequently misspelled if the literacy rate is low among these families. Because there is no information on nativity or literacy in the pension data, I cannot control for these variables without restricting my sample to individuals linked to the census. However, I can control for average literacy, immigrant status and socioeconomic status, measured as the occupational income of the household head,[56] by phonetic name group in the IPUMS data. I include these controls to preserve the validity of the instrument.

Table 1.7 contains first stage results. For all possible values of τ, name homogeneity strongly predicts receiving a pension, even conditional on the immigration, occupational income, and literacy controls added from the census. The first stage F statistics are not quite as high as one would like, ranging from 3.18 to 7.16; however, they are substantially higher than the F statistics for any other potential instrument. The relationship between pension status and other explanatory variables is broadly consistent with results on the rate of pension receipt from section 6.

Table 1.8 contains both OLS and 2SLS results. The OLS estimate is negative for all values of τ, but only significant at the five percent level when $\tau \geq 3$. The 2SLS estimates are also everywhere negative, but they are close to one in magnitude, and the standard errors are quite large. The estimates are only significantly different from zero when $\tau \geq 4$. Because the first stage F statistics point to the possibility that the instrument is weak, I also present 95 percent Anderson-Rubin confidence intervals for the effect of the pension, which are robust to weak instruments.[57] In

[56] See section 7 for an explanation of this variable.

[57] To calculate this confidence region, I use the condivreg command in Stata.

most cases, these confidence regions do not include zero. Given their imprecision, it is difficult to attach significance to the size of the 2SLS estimates. However, this analysis provides some corroborating evidence that the causal effect of pensions on the timing of remarriage is negative.

1.7.2 Alternative Sample Restrictions

An additional concern is that the results may be sensitive to the source of information on remarriage. Recall that knowledge of a widow's remarriage is contingent on her communicating in some way with the pension board. Specifically, I observe a widow's remarriage if her children file a minors' claim, or if she files a new claim under the act of March 3, 1901. If the source of information is distributed differently among women who remarry before and after obtaining a pension, and if the source of this information is correlated with marital outcomes, this might bias my results. As an example, recall from table 1.1 that minors' pension applications are the source of evidence for remarriage in 64 percent of cases that occur before a pension is granted and 84 percent of cases that occur after a pension is granted. This means that my sample of women who remarry before receiving a pension may be disproportionately composed of childless women who lived to 1901. These women may be younger and healthier by construction, and thus better marriage prospects.

I use two alternative sample restrictions to address this concern. First, I restrict the sample to women who have children under the age of 16 when they are widowed, and I stop following these women once their youngest child turns 16. So, the sample is restricted to women whose marital status *might* be known through a minor's pension application. Second, I discard any information that comes from a source other than a General Law pension claim, either widow or minor. Thus, any woman whose marital status is known only from a pension application under the law of March 3, 1901 becomes an observation with missing marital status.

Panels A and B of figure 1.3 plot the empirical hazard rate of remarriage by pension status, in similar fashion to figure 1.2, under these two sample restrictions. While the overall picture looks similar, as time passes the rate of remarriage for women without pensions starts to lie solidly below that of women with pensions. This could reflect the fact that the sample size is substantially reduced by these restrictions. It may also indicate that the effect of the pension on women's behavior is simply smaller for those with small children, so differences by pension status shrink when the sample is restricted to these women. However, we cannot rule out the possibility that differences in the source of information on remarriage are biasing the estimated effect of the pension away from zero.

The model described in section 6 is estimated under these sample restrictions, and the results appear in table 1.9. The baseline results, with and without a correction for correlated unobserved heterogeneity, are repeated in panel A. Panels B and C contain results from the sample restrictions outlined above. As seen in panel C, the results are not sensitive to the omission of information from pension applications under the act of March 3, 1901. When the sample period is restricted to years in which the widow has a minor child, the estimate remains negative; however, it decreases in magnitude relative to the baseline, and the standard errors increase. In panel B, we can only say with about 80 percent certainty that the coefficient is different from zero. Still, these results broadly support the finding of a negative effect of the pension, even if the estimate becomes noisier under one of the sample restrictions.

Panel C of figure 1.3 and panel D of of table 1.9 impose a different sample restriction. These use only women who are successfully linked to the census of 1870 and/or 1880. These data provide independent verification of the information on marital status in the pension files. Women have an incentive to lie to the pension board about marital status; however, there should be no such incentive to lie to census enumerators. By including only women whose marital status can be verified in the census, I mitigate accuracy issues that stem from pension fraud. Another benefit of the linked data is that it allows me to observe potentially important demographic variables such as birthplace and literacy. As seen in figure 1.3 and in table 1.9, the results are not sensitive to restricting the sample to women linked to the census, or to including controls for immigration and literacy. Panel D of figure 1.3 and panel G of table 1.9 restrict the sample to women widowed during the war years. Dying during the war is arguably more random than failing to recover from a non-life-threatening injury or disease contracted during the war, so it is worth verifying that the results are robust to this sample restriction. The restriction has little effect on the estimate.

Finally, I estimate OLS and 2SLS models that are similar to those in the previous subsection, restricting the sample to women who are linked to the census of 1870 or 1880 through children from their first marriages. As explained earlier, it is desirable to use an alternative way of identifying remarried widows, as the source of marriage information in the pension data may generate artificial differences between widows who remarry and those who do not. Table 1.10 contains results from regressions of an indicator for being remarried in the census on an indicator for having received a pension within five years of applying.[58] These are similar to the regressions presented in table 1.8. In column 1, I use links to the 1870 census; in column 2, I use links to the 1880 census; and in column 3, I pool both years and cluster standard errors by widow. Columns 4, 5, and 6 repeat these specifications using two stage least squares, where the instrument is the name homogeneity index used earlier. This instrument explains a reasonable amount of variation in pension status for the sample linked to the 1880 census, but it performs very badly for the sample linked to the 1870 census. This suggests that much of the variation being explained by the instrument is coming from

[58] I also try this with different time frames, and the results are similar.

women widowed in the later part of my sample.[59] Still, while the number of women linked in this fashion is small, and the estimates are often noisy, these results broadly support the basic findings. The coefficient on pension income is always negative, and the 2SLS estimate is significant at the ten percent level when the 1880 census is used.

1.8 Pensions and Match Quality

1.8.1 Empirical Framework

If pension income raises the minimum match quality women are willing to accept, it should increase the average quality of the matches they make, conditional on being matched at all. I use links to the federal censuses of 1870 and 1880 to evaluate this empirically. In principle, I would like to measure match-specific quality; however, this is not observable. Instead, I attempt to measure the "quality" of the second husband, controlling for the "quality" of the widow. I use four plausible measures of quality available in the linked census data. The first is the occupational income of the second husband, measured using the 1900 occupational wage distribution, with an imputed wage for farmers, assigned to 1950 occupational codes.[60] Another measure is literacy of the second husband. I also use the squared difference between the age of the husband and wife, the idea being that people of closer age may be better matched. Finally, I use an indicator equal to one if the second husband is present in the household.

Using my sample of remarried widows who have been linked to the census, and I

[59] The backlog of claims at the pension office grew over time, so it is possible that variation in processing time stemming from name spelling ambiguity was amplified in later years. See Oliver (1917). In fact, when I re-do the analysis in table 1.8 using only war dead, the first stage F statistic declines substantially.

[60] Occupational wages are taken from Preston and Haines (1991) and the farmer's wage is imputed from the 1900 census of agriculture using a procedure from Abramitzky Boustan and Eriksson (2010) and Olivetti and Paserman (2012).

estimate the following by OLS:

$$Q_{husb} = \beta_0 + \beta_1 PEN + \gamma X + u$$

The variable PEN is an indicator for the marriage having taken place after the widow received a pension, and Q_{husb} is a measure of match quality. The matrix X contains explanatory variables including the widow's age, literacy, immigrant status, the woman's age at widowhood, age at remarriage, characteristics of the woman's first husband from enlistment records, and county-level and region controls. I also include the number of children from the widow's first marriage and the potential amount of pension income these children could receive on the date of remarriage; these are both interacted with pension status. I do this to control for the role minors' pensions may have played in making these women more attractive to potential mates. What remains should capture the effect of the pension on women's selectivity in choosing a husband.[61]

I pool all married women linked to 1870 and 1880 in order to maximize the sample size. In some cases, women are linked to both the 1870 and 1880 census, so these individuals appear twice in the sample. With this in mind, I cluster standard errors by widow.

1.8.2 Results

Table 1.11 contains results from the regression model describe above. These results offer little evidence that marriages that occur after a pension is granted look different from marriages that occur before a pension is granted. With the exception of hus-

[61]Notice that I do not do this in the previous section. This is because I expect the effect of minors' pensions on the timing of remarriage to go in the opposite direction. If minors' pensions make women more desirable in the marriage market, they should receive more proposals per unit of search effort, tending to increase the rate of remarriage. Experimenting with interacting the effect of the pension with the number of children or potential minor's pension suggests that including these does not change the results. For ease of exposition, I do not include these in table 1.6.

band's literacy, the relationship between pension status and each measure of match quality has the anticipated sign; however, these estimates are very noisy. Because the OLS estimates do not suggest a relationship between pensions and match quality, I do not present additional results that correct for potential endogeneity of pensions to marital outcomes, as I do in the previous section.[62]

Is this conclusive evidence that pensions had no effect on match quality? Not necessarily. One possibility is that the measures of match quality I am using are very rough approximations, and a much larger sample size would be required to say anything conclusive about the effect of pensions on these measures. It is also possible that the achievable range of variation in match quality was quite small. If marriage markets are segmented by socioeconomic class, it may be difficult for a low socioeconomic status woman to marry a high socioeconomic status man, no matter how selective she is. Moreover, my sample is largely rural, which means that marriage markets are quite small. This would also tend to decrease the range of variation in match quality that is possible for individuals.

These measures also fail to capture other aspects of match quality that we expect to be important. For example, it matters how much a woman and her potential husband like each other, or how this potential husband relates to her children. In a small number of cases, the case files include descriptions of physical and emotional abuse of the part of second husbands, which is likely underreported. Differences in match quality along dimensions such as these will fail to be included in this analysis.

On the other hand, these results may be informative about the precise mechanism behind the results from section 6. I have described two channels through which an increase in the value of being unmarried may affect the rate of remarriage: it is expected to increase a woman's reservation match quality and to lower her optimal

[62] Attempting to use an instrumental variables approach, with the instrument described in the next section, yielded similarly noisy results. These are omitted here for the sake of brevity.

level of search effort. These results might indicate that the latter channel is more important. In fact, this is consistent with recent empirical investigations into the effect of unemployment insurance on post-employment outcomes.[63] In any case, the evidence presented in this section is inconclusive, and further investigation is required here.

1.9 Implications and Discussion

This paper's most robust and important finding is that receiving a pension had a causal effect on the timing of remarriage for Union Army widows who filed for a pension. Having a claim granted lowered the rate of remarriage by 40 percent. This implies that a typical widow who immediately received a pension would tend to remarry three years later than an identical widow with no pension. This is a striking result, for which I provide context and interpretation in this section.

The main issue is generalizability. Namely, is it reasonable to infer that allocating this modest amount of income to all single women in the late 19th century U.S. would have raised the median age at first marriage by three years? Perhaps not. Interpreting the results in the context of the model, it seems that pensions did cause women to raise reservations match qualities or lower search effort. However, the size of this shift may very well depend on other parameters of the model: the distribution of match qualities, a woman's flow utility while single, her discount rate, etc. This effect is estimated using a sample of women who apply for pensions, and I have shown evidence that these women are different from those who do not make pension applications. I also find a significant interaction effect with age at widowhood (not shown): with each additional year of age, the effect of the pension increases in magnitude by 0.07 (0.02). My sample of pension applicants seems to be older on average than non-applicants,

[63] See Card, Chetty and Weber (2007) and van Ours and Vodopivec (2008).

and they are certainly older that the average unmarried woman by virtue of the fact that they have been married before. It is possible that the response among women in the general population would be more muted.

On the other hand, my estimates are generated by a comparison between women who have been granted a pension and women who are *still waiting for a pension.* The rationale behind this approach is that there is uncertainty about if and when the pension claim will be granted, so discounting and the possibility of rejection should generate differences in behavior. However, if the data allowed a comparison between women with a pension and women *with no possibility* of a pension, the differences may be starker. Another interesting point to note is that the probability of rejection, at about 14 percent, is quite low. So, the results likely reflect a high discount rate, which suggests liquidity constraints. Again, this may point to different effects in a more representative sample: women in my sample appear to be less wealthy on average, so liquidity constraints may have been more binding.

It is probably also important that my sample consists of widows and not never-married women. Preferences for marriage may have been different for these two types of women. Historical literature suggests that widows, if financially viable, may have been less constrained by social norms in their ability to take part in public organizations, such as charities or other benevolent associations.[64] If women value this kind of freedom, preferences for marriage may have been lower for widows. However, widows may have been more financially constrained than never-married women, especially if they had small children; this may augment their preferences for marriage. The fact that they had been previously married may also signal greater a preference for marriage.

While the composition of my sample makes it difficult to extrapolate the magni-

[64] See Reinhart, Tacardon and Hardy (1998) and Boylan (1986) for discussions of the different roles for widows, as opposed to never-married women, in these organizations.

tude of this effect to the population of unmarried women, it does offer evidence that women responded to economic alternatives when making decisions about marriage. This is informative about changes in first marriage that occurred over the course of the 19th century, and also across regions. While this pales in comparison to the revolution in female labor market opportunities that occurred more recently, there were increases in women's work opportunities during this century, largely due to industrialization. There were also regional differences in opportunities for women, which have been shown to be correlated with delayed marriage (Hacker 2008). A major contribution of this paper is to demonstrate that economic opportunities had a causal effect on women's behavior. This is especially important because of the multitude of potential drivers of the patterns we observe in the 19th century.

1.10 Conclusion

This paper documents the effect of pension income on the marital outcomes of Union Army widows during the late 19th century. While there is little evidence that women receiving pensions married systematically "better" husbands, my results suggest that receiving a pension significantly lowered the rate of remarriage. I argue that this effect can be presumed to work through widows' preferences for mates, suggesting that American women during this period did respond to outside economic opportunities when making decisions about marriage. This gives new insight into the functioning of marriage markets during this period. It also provides an early example of the kind of behavior we observe on a greater scale at the end of the 20th century.

The results of this paper demonstrate that women's economic incentives mattered for marriage market outcomes in the 19th century. As such, my findings suggest that factors affecting the gains from marriage for women are important to understanding differences in behavior over time and in different regions. This paper has focused on the role of economic alternatives in reducing preferences for marriage; however, future work might look into how women responded to events that raised the gains from marriage. Over the course of the 19th century, marriage became a significantly better 'deal' for women. Divorce laws were gradually liberalized, allowing women to escape from bad marriages if necessary (Doepke and Tertlit 2008). Laws allowing married women to independently engage in business and to hold property were enacted, with almost all states adopting such laws by 1895 (Doepke and Tertlit 2008; Fernandez 2009). Women's inheritance laws were also amended to allow widows greater ownership and control of their spouses' assets (Hirsch 2009). Understanding the way women responded to these and other developments will be key to understanding patterns of marriage in the 19th century.

1.11 Appendix

1.11.1 Proofs

Proof of equation (2).

Suppose the arrival rate of pension decisions is λ, the arrival rate of marriage proposals if α, and the probability of an acceptance is π. Take Δ to be an arbitrarily small period of time, and note that, for search effort $c(\alpha)$, the probability of receiving a marriage proposal during this interval is $\alpha\Delta$; similarly, the probability of receiving a decision from the pension bureau is $\lambda\Delta$. Call V^S the expected value of being single, which will be a weighted average of the value of being single in each potential state of “singlehood”. Then, it must be that

$$\begin{aligned}
\tilde{V} &= \Delta(s - c(\alpha)) + \frac{\Delta\alpha}{1+\Delta r}\Big(E[\max(V^M, V^S)]\Big) + \frac{1-\Delta\alpha}{1+\Delta r}E[V^S] \\
&= \Delta(s - c(\alpha)) + \frac{\Delta\alpha}{1+\Delta r}\Bigg(\Delta\lambda\Big(\pi E[\max(V^M, V^P)] + (1-\pi)E[\max(V^M, V^N)]\Big)+ \\
&\quad + (1-\Delta\lambda)E[\max(V^M, \tilde{V})]\Bigg) + \frac{1-\Delta\alpha}{1+\Delta r}\Bigg(\Delta\lambda\Big(\pi V^P + (1-\pi)V^N\Big) + (1-\Delta\lambda)\tilde{V}\Bigg) \\
&= \Delta(s - c(\alpha)) + \frac{\Delta\alpha}{1+\Delta r}\Bigg(\Delta\lambda\Big(\pi E[\max(V^M - V^P, 0)] + (1-\pi)E[\max(V^M - V^N, 0)]\Big)+ \\
&\quad + (1-\Delta\lambda)E[\max(V^M - \tilde{V}, 0)]\Bigg) + \frac{\Delta\lambda}{1+\Delta r}\Big(\pi V^M + (1-\pi)V^N - \tilde{V}\Big) + \frac{1}{1+\Delta r}\tilde{V}
\end{aligned}$$

Re-arranging, dividing by Δ, and taking the limit as $\Delta \to 0$, we get (2).

Proposition 1. For $\pi \in (0,1]$, $\theta_N < \tilde{\theta} < \theta_P$ and $\alpha^*_N > \tilde{\alpha}^* > \alpha^*_P$.

Proof. Throughout, I use the well known result that $\int_{\theta_i}^{\infty}(\theta - \theta_i)dF(\theta) = \int_{\theta_i}^{\infty}(1 - F(\theta))d(\theta)$ First notice that $\tilde{\theta}$ is strictly increasing in π:

$$\begin{aligned}
\frac{\partial\tilde{\theta}}{\partial\pi} &= -\frac{\tilde{\alpha}^*}{r}(1 - F(\tilde{\theta}))\frac{\partial\tilde{\theta}}{\partial\pi} + \frac{\lambda}{r}(\theta_P - \theta_N) \Rightarrow \\
\frac{\partial\tilde{\theta}}{\partial\pi} &= \frac{\lambda(\theta_P - \theta_N)}{r + \tilde{\alpha}^*(1 - F(\tilde{\theta}))} > 0
\end{aligned}$$

Now, suppose $\pi = 0$. Call $\tilde{\theta}^0$ the reservation match quality for those with pending claims when $\pi = 0$. Then, $\tilde{\theta} \geq \tilde{\theta}^0$. So, if $\tilde{\theta}^0 \geq \theta_N$, then $\tilde{\theta} > \theta_N$ for $\pi > 0$.

If $\pi = 0$, then the reservation match quality for women with pending claims becomes

$$\tilde{\theta} = s - c(\tilde{\alpha}^*) + \frac{\tilde{\alpha}^*}{r}\int_{\tilde{\theta}}(1 - F(\theta))d(\theta) + \frac{\lambda}{r}(\theta_N - \tilde{\theta})$$

The left hand side of this equation is strictly increasing in $\tilde{\theta}$ and the right hand side is strictly decreasing in $\tilde{\theta}$, so it has a unique solution. I will show that $\tilde{\theta} = \theta_N$ and

$\tilde{\alpha^*} = \alpha_N^*$ solve both this equation and the first order condition:

$$\begin{aligned}\theta_N = \tilde{\theta} &= s - c(\alpha_N^*) + \frac{\alpha_N^*}{r}\int_{\theta_N}(1 - F(\theta))d(\theta) + \frac{\lambda}{r}(\theta_N - \theta_N) \\ &= s - c(\alpha_N^*) + \frac{\alpha_N^*}{r}\int_{\theta_N}(1 - F(\theta))d(\theta) \\ &= \theta_N\end{aligned}$$

The first order condition defining $\tilde{\alpha}^*$ is $rc'(\tilde{\alpha}^*) = \int_{\tilde{\theta}}^{\infty}(1 - F(\theta))d(\theta)$, which is set up the same way as the condition defining α_N^*. Thus, $\tilde{\theta} = \theta_N$ and $\tilde{\alpha^*} = \alpha_N^*$ satisfy this condition as well. So, when $\pi = 0$, $\tilde{\theta} = \theta_N$. Therefore, for $\pi > 0$, $\tilde{\theta} > \theta_N$.

Now, define $\tilde{\theta}^1 = \tilde{\theta}$ when $\pi = 1$. If $\theta_P > \tilde{\theta}^1$, then $\theta_P > \tilde{\theta}$ for every $\pi \leq 1$. When $\pi = 1$:

$$\tilde{\theta} = s - c(\tilde{\alpha}^*) + \frac{\tilde{\alpha}^*}{r}\int_{\tilde{\theta}}(1 - F(\theta))d(\theta) + \frac{\lambda}{r}(\theta_P - \tilde{\theta})$$

Suppose $\tilde{\theta} \geq \theta_P$. Because the optimal α^* is decreasing in reservation θ_i (see below), it follows that $\alpha_P^* \geq \tilde{\alpha}^*$. Two inequalities follow from this: First,

$$\frac{1}{r}\int_{\tilde{\theta}}(1 - F(\theta))d(\theta) \leq \frac{1}{r}\int_{\theta_P}(1 - F(\theta))d(\theta)$$

And, from convexity of $c(\alpha)$, we get the following inequality:

$$-c(\tilde{\alpha}^*) \leq -c(\alpha_P^*) + c'(\alpha_P)(\alpha_P^* - \tilde{\alpha}^*)$$

This implies the following:

$$\begin{aligned}
\tilde{\theta} &= s - c(\tilde{\alpha}^*) + \frac{\tilde{\alpha}^*}{r}\int_{\tilde{\theta}}(1-F(\theta))d(\theta) + \frac{\lambda}{r}(\theta_P - \tilde{\theta}) \\
&\leq s - c(\tilde{\alpha}^*) + \frac{\tilde{\alpha}^*}{r}\int_{\tilde{\theta}}(1-F(\theta))d(\theta) + \frac{\lambda}{r}(\theta_P - \theta_P) \\
&\leq s - c(\tilde{\alpha}^*) + \frac{\tilde{\alpha}^*}{r}\int_{\theta_P}(1-F(\theta))d(\theta) \\
&\leq s - c(\alpha_P^*) + c'(\alpha_P^*)(\alpha_P^* - \tilde{\alpha}^*) + \frac{\tilde{\alpha}^*}{r}\int_{\theta_P}(1-F(\theta))d(\theta) \\
&= s - c(\alpha_P^*) + \frac{1}{r}\int_{\theta_P}^{\infty}(1-F(\theta))d\theta(\alpha_P^* - \tilde{\alpha}^*) + \frac{\tilde{\alpha}^*}{r}\int_{\theta_P}(1-F(\theta))d(\theta) \\
&= s - c(\alpha_P^*) + \frac{\alpha_P^*}{r}\int_{\theta_P}(1-F(\theta))d(\theta) \\
&= \theta_P - p < \theta_P
\end{aligned}$$

This is a contradiction. So, it must be that, when $\pi = 1$, $\theta_P > \tilde{\theta}$, which further implies that $\theta_P > \tilde{\theta}$ for all $\pi \leq 1$. Therefore, for all $\pi \in (0,1]$, $\theta_N < \tilde{\theta} < \theta_P$.

The result that $\alpha_P^* < \tilde{\alpha}^* < \alpha_N^*$ follows from the fact that α^* is decreasing in reservation match quality. Recall that, for reservation match quality θ_i, α^* is defined by the following condition:

$$rc'(\alpha^*) = \int_{\theta_i}^{\infty}(1-F(\theta))d(\theta)$$

Then, $\partial\alpha^*/\partial\theta_i$ is given by:

$$\frac{\partial\alpha^*}{\partial\theta_i} = \frac{-(1-F(\theta_i))}{rc''(\alpha^*)} < 0$$

This follows from the convexity of search costs.

1.11.2 Data

Data Collection

In this section, I describe the process by which I collected the data for this project. The most important effort is the collection of pension records from the National Archives in Washington, DC. Using the indices to the Civil War pension files available on ancestry.com and fold3.com, I compile a list of all pension applications made and certificates issued on behalf of soldiers married to the women in my sample. Then, I request these files from the National Archives. In approximately 90 percent of cases, these files are successfully located, and I am able to collect digital images of them. Files that could not be located had either been taken out by another use (37% of cases), or the file number was incorrectly recorded, and the record puller was unable to find it (63% of cases). Where possible, I make use of digital images of widows' pensions from the website fold3.com. This website is in the process of uploading images of accepted widows' pensions, which they are doing chronologically. It is not possible to make exclusive use of this resource for several reasons. First, this project is expected to take several years to complete. Second, they do not include rejected pension applications. Third, they do not include minors' pensions. If a widow remarried and her children applied for the pension, her file would be consolidated with theirs, and the entire file would be classified as a minor's pension. So, it would be excluded from the fold3.com project. In total, 30 percent of my sample can be collected from this resource.

Because of the importance of these variables to the paper, I describe the source of information on pension outcomes and marriages in the body of the text. However, there are other important variables collected from the pension files. Other available information includes the widow's age and place of residence, as she had to furnish this information in her pension application. If a remarried widow applied to be restored to the pension rolls under the act of March 3, 1901, her file will contain further information about her second husband. For example, she had to provide proof of her husbands death, which usually meant furnishing a death certificate. In some cases, these death certificates contain the age, birthplace, and occupation of the husband.

The second source of information consists of links to the census of 1870 and 1880. I perform these links manually using the genealogy website ancestry.com. When marital status is certain, I search for the widow using the appropriate surname. If I am unable to find her, I search for the children from her first marriage. If her marital status is uncertain, I search only for her children. Whenever there is insufficient information to distinguish between two candidate links, I discard the observation. However, because of the detailed information available in the widow's pension application, including place of residence, this is a rare occurrence. I am able to make very high quality links in most cases.

A concern is that being linked to the federal census may not be random. Table 1.12 contains OLS regressions of an indicator for a widow being linked to the census on explanatory variables from the pension data. For each census year, the sample is comprised of women who are widowed by that year and who are not known to have died. The only significant determinant of linkage to the 1870 census is the number of children from the widow's first marriage; this is unsurprising, as information about family members is used to create these links. Age and time since widowhood have no significant effect on linkage to the 1870 census. Neither do pension status, measured by an indicator equal to one if the widow had received a pension within five years of applying, or the region in which the widow's first husband enlisted. The omitted category is the northeast.

The number of children from the widow's first marriage also significantly increases the probability of her being linked to the 1880 census. Women whose husbands died more recently are also more likely to be linked, as are women whose husbands enlisted in the midwest or the south (relative to those who enlisted in the northeast). The former result can likely be explained by the fact that information about women whose husbands died closer to 1880 is more current; women widowed earlier are more likely to have died, which might not have been recorded in the pension data: death records for pensioners were not consistently kept before the 1880s. Linking women from the midwest and the south may be more successful because I am using information about place of residence from the pension file data. Women residing in smaller towns or counties are less likely to have multiple positive matches, so these women may be less likely to go unlinked. This may be more of a problem in 1880 than 1870 because fewer women are residing with linkable children in 1880, so residential information is more important in this census year.

Tables 1.13 and 1.14 present further descriptive information about the linked data. In table 1.13, widows linked to the census are compared with nationally representative samples of women from IPUMS by marital status. Mean characteristics from the IPUMS data are presented unadjusted and re-weighted to obey the same distribution of five-year birth cohorts as the analogous sample from the linked widows data. Table 1.14 conveys information about the household composition in the linked widows data by year and marital status.

Variables

Variable	Source	Notes
Date of first husband's death	Union Army database (UA; Fogel et al. 2000)	Based on dependents' pension applications or military death records
Date of pension application	Widows' pension (WP; Salisbury)	Date at which widow filled out pension declaration form; if missing, date at which pension application received by pension bureau
Date of pension receipt	WP	Date of issuance on pension certificate; if missing, date of pension approval on pension brief
Date of remarriage	WP	Based on marriage certificates or affidavits rendered in support of minors' pension application or application for widow to be restored to the pension rolls under a later act
Date of death	WP	Base on pension drop cards, or death records filed in support of minors' pension application
Age at widowhood	WP	Deduced from widow's first pension declaration, in which age and date of application are both provided.
Number of children	UA	Equal to number of children under the age of 16 when widow first filed for pension
Potential minor pension	UA	Calculated as \$8/mo until youngest child turns 16, or \$8/mo plus \$2/mo for each child under 16 if widowed after July 25, 1866
No pension attorney	WP	Equal to one if the widow did not hire an attorney at the time of filing her first claim
Washington pension attorney	WP	Equal to one if the widow first hired an attorney from a Washington firm at the time of filing her first claim
First husband: height	UA	Soldier's height at enlistment
First husband: log occupational wage	UA; Preston and Haines (1991); United States Census of Agriculture (1900)	Based on soldier's occupation at enlistment
First husband: age at death	UA	Based on implied birth year from age at enlistment

County of residence	WP	County listed on first pension application form
County male-to-female ratio	Haines and ICPSR (2010)	Weighted mean of male-to-female ratio in 1860, 1870 and/or 1880, depending on date of application
County percent urban	Haines and ICPSR (2010)	See above
County population density	Haines and ICPSR (2010)	See above
Name homogeneity index	Ruggles et al. (2010); Atack and Bateman (1992)	Herfindahl index of concentration of unique spellings within phonetic surname groups among household heads in 1 percent IPUMS sample from 1860–1880. Phonetic groups created using NYIIS algorithm.
Last name: mean occupational income	Ruggles et al. (2010); Preston and Haines (1991); United States Census of Agriculture (1900)	Mean occupation status of household head, calculated using 1900 wage distribution, by phonetic name group in IPUMS 1 percent sample from 1860–1880.
Last name: mean immigrant status	Ruggles et al. (2010);	Mean immigrant status of household head by phonetic name group in IPUMS 1 percent sample from 1860–1880.
Last name: mean literacy	Ruggles et al. (2010);	Mean literacy of household head by phonetic name group in IPUMS 1 percent sample from 1860–1880.
Last name: mean farm residence	Ruggles et al. (2010);	Mean farm status of household head by phonetic name group in IPUMS 1 percent sample from 1860–1880.
Literacy	Linked widow sample (Salisbury); Ancestry.com	Literate in census of 1870 or 1880
Immigrant status	Linked widow sample; Ancestry.com	Immigrant in census of 1870 or 1880

1.12 References

Abbring and Van den Berg (2005). "The Effect of Unemployment Insurance Sanctions on the Transition Rate from Unemployment to Employment." *Economic Journal.* 115(505): 602-630.

—. (2003a). "The Nonparametric Identification of Treatment Effects in Duration Models." *Econometrica.* 71(5): 1491-1517.

—. (2003b). "The Identifiability of the Mixed Proportional Hazards Competing Risks Model." *Journal of the Royal Statistical Society, Series B.* 65(3): 701-710.

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson (2010). "Europes Tired, Poor, Huddles Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration." NBER Working Paper no. 15684.

Abramitzky, Ran, Adeline Delavande, and Luis Vasconcelos (2011). "Marrying Up: The Role of Sex Ratio in Assortative Matching." *American Economic Journal: Applied Economics.* 2: 124-157.

Atack, Jeremay and Fred Bateman (1992). "Matchmaker, Matchmaker, Make Me a Match: A General Personal Computer-Based Matching Program for Historical Research" *Historical Methods.* 25(2):53-65.

Bahr, Stephen J. (1979). "The Effects of Welfare on Marital Stability and Remarriage." *Journal of Marriage and the Family.* 41(3): 553-560.

Bailey, Martha J. (2006). "More Power to the Pill: The Impact of Contraceptive Freedom on Women's Life-Cycle Labor Supply." *Quarterly Journal of Economics.* 121(1): 289-320.

Baker, Michael, Jasmin Kantarevic and Emily Hanna (2004). "The Married Widow: Marriage Penalties Matter!" *Journal of the European Economic Association.* 2(4): 634-664.

Becker, Gary S. (1973). "A Theory of Marriage: Part I." *Journal of Political Economy.* 8(4): 813-846.

Becker, Gary S. (1991). *A Treatise on the Family.* Cambridge: Harvard University Press.

Blau, Francine D. (1998). "Trends in the Economic Well-Being of Women." *Journal* *of Economic Literature.* 36(1): 112-165.

Blau, Francine D., Lawrence M. Kahn and Jane Waldfogel (2000). "Understanding Young Womens Marriage Decisions: The Role of Labor and Marriage Market Conditions." *Industrial and Labor Relations Review.* 53(4): 624-647.

Brien, Michael J., Stacy Dickert-Conlin and David A. Weaver (2004). "Widows Waiting to Wed?" *Journal of Human Resources.* 39(3): 585-623.

Calhoun, Arthur W. (1919). *Social History of the American Family.* Vol. 3. Cleveland: The Arthur H. Clark Company.

Card, D., R. Chetty and A. Weber (2007). "Cash-On-Hand and Competing Models of Intertemporal Behavior: New Evidence from the Labor Market." *Quarterly Journal of Economics.* 122: 1511-1560.

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

Costa, Dora L. (1993). "Height, Wealth, and Disease Among the Native-Born in the Rural, Antebellum North." *Social Science History.* 17(3): 355-383.

—. (1995). "Pensions and Retirement: Evidence from Union Army Veterans." *Quarterly Journal of Economics.* 110(2): 297-319.

—. (1997). "Displacing the Family: Union Army Pensions and Elderly Living Arrangement." *Journal of Political Economy.* 105(6): 1269-1292.

Costa, Dora L. and Matthew E. Kahn (2003). "Cowards and Heroes: Group Loyalty in the American Civil War." *Quarterly Journal of Economics.* 118(2): 519-548.

—. (2008). *Heroes and Cowards: The Social Face of War.* Princeton, NJ: Princeton University Press.

Cvercek, Tomas (2009). "When Harry Left Sally: A New Estimate of Marital Disruption in the U.S., 1860-1948." *Demographic Research.* 21: 719-758.

Doepke, Matthias and Michele Tertlit (2008). "Women's Liberation: What's in it for Men?" NBER Working Paper, no. 13919.

Easterlin, Richard A. (1976). "Change and Farm Settlement in the Northern United States." *Journal of Economic History.* 36(1): 45-75.

—. (1971). “Does Human Fertility Adjust to the Environment” *American Economic Review.* 61(2): 399-407.

Eli, Shari (2010). “Wealth is Health: Pensions and Disease Onset in the Gilded Age.” Unpublished paper, Berkeley University.

Fernandez, Raquel (2009). “Women’s Rights and Development.” NBER Working Paper, no. 15355.

Ferrie, Joseph P. (1996). “A New Sample of Americans Linked from the 1850 Public Use Micro Sample of the Federal Census of Population to the 1860 Federal Census Manuscript Schedules.” *Historical Methods.* 29: 141- 156.

Fitch, Catherine A. and Steven Ruggles (2000). “Historical Trends in Marriage Formation: The United States, 1850-1990.” *The TIes that Bind: Perspectives on Marriage and Cohabitation.* Ed. Linda J. Waite, Christine Bachrach, Michelle Hindin, Elizabeth Thomson, and Arland Thornton. New York: de Gruyter: 59-88.

Fogel, Robert W. (2000). *Public Use Tape on the Aging of Veterans of the Union Army: Military, Pension, and Medical Records, 1860-1940, Version M-5.* Center for Population Economics, University of Chicago Graduate School of Business, and Department of Economics, Brigham Young University.

Fogel, Robert W. (2004). “Changes in the Process of Aging During the Twentieth Century: Findings and Procedures of the Early Indicators Project.” *Population and Development Review.* 30: 19-47.

Glasson, William H. (1900). *History of Military Pension Legislation in the United States.* Doctoral Dissertation. New York: Columbia University.

Glasson, William H. (1918). *Federal Military Pensions in the United States.* New York: Oxford University Press.

Goldin, Claudia (1995). “Career and Family: College Women Look to the Past.” NBER Working Paper, no. 5188.

—-. (1984). “The Historical Evolution of Female Earnings Functions and Occupations.” *Explorations in Economic History.* 21: 1-27.

—-. (1977). “Female Labor Force Participation: The Origin of Black and White Differences, 1870 and 1880.” *Journal of Economic History.* 37(1): 87-108.

Goldin, Claudia and Lawrence Katz (2002). "The Power of the Pill: Oral Contraceptives and Women's Career and Marriage Decisions." *Journal of Political Economy.* 110(4):730-770.

Gould, Eric D. and M. Daniele Paserman (2003). "Waiting for Mr. Right: Rising Inequality and Declining Marriage Rates." *Journal of Urban Economics.* 53(2): 257-281.

Hacker, J.David (2008). "Economic, Demographic, and Anthropometric Correlates of First Marriage in the Mid-Nineteenth-Century United States." *Social Science History.* 32(3): 307-345.

Haines, Michael R. (1996). "Long-Term Marriage Patterns in the United States from Colonial Times to the Present." *History of the Family.* 1(1): 15-39.

Haines, Michael R., and Inter-university Consortium for Political and Social Research (2010). Historical, Demographic, Economic, and Social Data: The United States, 1790-2002 [Computer file]. ICPSR02896-v3. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2010-05-21. doi:10.3886/ICPSR02896

Haines, Michael R. and J. David Hacker (2006). "The Puzzle of the Antebellum Fertility Decline in the United States: New Evidence and Reconsideration." NBER Working Paper, no. 12571.

Heckman, James J. and Burton Singer (1984). "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data." *Econometrica.* 52(2): 271-320.

Hirsch, Adam J. (2009). "Inheritance: United States Law." *Oxford International Encyclopedia of Legal History.* Ed. Stanley N. Katz. New York: Oxford University Press, 235-240.

Keeley, Michael C. (1977). "The Economics of Family Formation." *Economic Inquiry.* 15(2): 238-250.

Lancaster, Tony (1990). *The Econometric Analysis of Transition Data.* Cambridge: Cambridge University Press.

Legislature of New York (1837). "Excerpt from an Act Concerning Executors, Administrators, Guardians, Wards &c., Passed May 16, 1837." *Case Files of Approved Pension Applications of Widows and Other Dependents of Civil War Veterans, ca.*

1861 - ca. 1910. Records of the Department of Veterans' Affairs, Record Group 15. National Archives Building, Washington, D.C. Digital Images, "Civil War Widows' Pensions," fold3.com.

Linares, Claudia (2001). "The Civil War Pension Law." Chicago: Center for Population Economics Working Paper no. 2001-6.

Loughran, David (2002). "The Effect of Male Wage Inequality on Female Age at First Marriage." *Review of Economics and Statistics.* 84(2): 237-250.

Maestas, Nicole, Kathleen J. Mullen and Alexander Strand (2011). "Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt." RAND Working Paper WR-853-2.

Margo, Robert A. (2000). *Wages and Labor Markets in the United States, 1820-1860.* Chicago: University of Chicago Press.

Monohan, Thomas P. (1951). *The Pattern of Age at Marriage in the United States.* Vol. 2. Philadelphia: Stephenson Brothers.

Mortenson, Dale T. (1986). "Job Search and Labor Market Analysis." *Handbook of Labor Economics.* Ed. O. Ashenfelter and R. Layard. Amsterdam: North Holland, 849-920.

Oliver, John William (1917). "History of the Civil War Military Pensions." *Bulletin of the University of Wisconsin Historical Series.* 4(844): 1-120.

Olivetti, Claudia and M. Daniele Paserman (2012). "In the Name of the Father: Marriage and Intergenerational Mobility in the United States: 1850-1930." Boston University.

Preston, Samuel H. and Michael Haines (1991). *Fatal Years: Child Mortality in Late Nineteenth Century America.* Princeton: Princeton University Press.

Preston, Samuel H. and John McDonald (1979). "The Incidence of Divorce Within Cohorts of American Marriages Contracted Since the Civil War." *Demography.* 16(1): 1-25.

Price-Bonham, Sharon and Jack O. Balswick (1980). "The Noninstitutions: Divorce, Desertion, and Remarriage." *Journal of Marriage and the Family.* 42(4): 959-972.

Rogerson, Richard, Robert Shimer and Randall Wright (2005). "Search-Theoretic

Models of the Labor Market: A Survey." *Journal of Economic Literature.* 43(4): 959-988.

Rosenzweig, Mark R. (1999). "Welfare, Marital Prospects, and Nonmarital Childbearing." *Journal of Political Economy.* 107(56): 3-32.

Ruggles, Steven (1997). "The Rise of Divorce and Separation in the United States.

Demography. 34(4): 455-466.

Ruggles, Steven J., Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek. *Integrated Public Use Microdata Series: Version 5.0* [Machine-readable database]. Minneapolis: University of Minnesota, 2010.

Skocpol, Theda (1993). "Americas First Social Security System: The Expansion of Benefits for Civil War Veterans." *Political Science Quarterly.* 108(1): 85-116.

Skocpol, Theda (1995). *Protecting Soldiers and Mothers: The Political Origins of Social Policy in the United States.* Cambridge: Harvard University Press.

Song, Chen (2000). "Filing for the Union Army Pension: A Summary from Historical Evidence." Unpublished Paper, Center for Population Economics, University of Chicago.

United States Pension Bureau (1883). *Annual Report of the Commissioner of Pensions to the Secretary of the Interior for the Year Ending June 30, 1883.* Washington: Government Printing Office.

—. (1872). *Annual Report of the Commissioner of Pensions to the Secretary of the Interior for the Year Ending June 30, 1872.* Washington: Government Printing Office.

—. (1864). *Annual Report of the Commissioner of Pensions to the Secretary of the Interior for the Year Ending June 30, 1864.* Washington: Government Printing Office.

—. (1997). "Association Measures for Durations in Bivariate Hazard Rate Models." *Journal of Econometrics.* 79(2): 221-245.

Van Ours, Jan C. and Milan Vodopivec (2008). "Does Reducing Unemployment Insurance Generosity Reduce Job Match Quality?" *Journal of Public Economics.* 92: 684-695.

Waite, Linda J. and Glenna D. Spitze (1981). "Young Women's Transition into Marriage." *Demography.* 18(4): 681-694.

Wanamaker, Marianne (2012). "Industrialization and Fertility in the Nineteenth Century: Evidence from South Carolina." *Journal of Economic History.* 72(1): 168-196.

Weiss, Yoram (1997). "The Formation and Dissolution of Families: Why Marry? Who Marries Whom? And What Happens Upon Divorce?" *Handbook of Population and Family Economics.* 1: 81-123.

Weiss, Yoram and Robert J. Willis (1997). "Match Quality, New Information and Marital Dissolution." *Journal of Labor Economics.* 15(1): S293-329.

Widow's Certificate No. 8,336. Jonathan Timberman, Private, Compay I, 12th New York Infantry. *Case Files of Approved Pension Applications of Widows and Other Dependents of Civil War Veterans, ca. 1861 - ca. 1910.* Records of the Department of Veterans' Affairs, Record Group 15. National Archives Building, Washington, D.C. Digital Images, "Civil War Widows' Pensions," fold3.com.

Widow's Certificate No. 73,022. Henry Havens, Private, Company A, 141st New York Infantry. —.

Widow's Certificate No. 35,292. Henry Detrich, Private, Company K, 111th New York Infantry. —-.

Widow's Certificate No. 35,292. Henry Detrich, Private, Company K, 111th New York Infantry. *Case Files of Approved Pension Applications of Widows and Other Dependents of Civil War Veterans, ca. 1861 - ca. 1910.* Records of the Department of Veterans' Affairs, Record Group 15. National Archives Building, Washington, D.C.

Widow's Certificate No. 6,916. William Matthews, Private, Company C, 98th New York Infantry. —-.

Widow's Certificate No. 23,539. Andrew Van Buren, Private, Company E, 100th New York Infantry. —-.

1.13 Tables and Figures

Table 1.1: Sources of Information on Marital Status in Pension File Data

Group: Source of Information:	Without General Law Pension	With General Law Pension
	Remarried	
Minor's application after remarriage	64%	84%
Widow's application under late law	30%	11%
Other communication	6%	5%
N, remarried	77	141
	Never Remarried	
Dropped from pension rolls	0%	81%
Minor's application after death	52%	13%
Widow's application under law of 1890	48%	0%
Other communication	0%	6%
N, never remarried	26	153
	Unknown	
N, unknown	14	87
N, total	117	381

This table summarizes the sources of information about widows' remarriage status, separately by pension status. Sample includes women widowed by 1880 and who applied for a pension within five years of widowhood.

Table 1.2: Summary Statistics from Pension File Data

Variable:	Mean	Median	SD	Min	Max	N
Pension Variables						
Applied within 1 year	0.817	1.000	0.387	0.000	1.000	498
Time to first application	0.674	0.285	0.958	0.014	5.767	498
General law claim accepted	0.865	1.000	0.342	0.000	1.000	498
Processing time of accepted gen law claim	2.280	0.906	4.583	0.112	50.500	431
Age/Marriage Variables						
Age widowed	31.867	30.000	9.410	15.000	73.000	487
Age at first marriage	20.838	20.000	5.025	9.000	48.000	474
Age at remarriage	32.080	31.000	7.641	18.000	65.000	213
Number of children (first marriage)	2.566	2.000	2.240	0.000	13.000	498
Husband died during war years	0.721	1.000	0.449	0.000	1.000	498
Remarried	0.549	1.000	0.498	0.000	1.000	397
Remarried without pension	0.181	0.000	0.386	0.000	1.000	425
Time to Remarriage:						
All	4.305	3.348	3.708	0.230	26.036	215
Remarried with pending claim	2.392	1.838	1.866	0.230	8.778	76
Remarried after pension	5.351	4.096	4.038	0.915	26.036	139
Time to remarriage following pension	3.911	2.573	3.956	0.047	25.463	134
Calendar Years						
First marriage	1854.6	1856	7.820	1822	1879	489
Widowhood	1865.6	1864	4.526	1862	1879	497
Remarriage	1868.9	1867	5.124	1863	1889	215
Pension application	1866.2	1864	5.044	1862	1883	498
Pension certificate	1869.0	1866	9.145	1862	1928	451
Region of Residence						
New England	0.129	0.000	0.336	0.000	1.000	488
Mid Atlantic	0.303	0.000	0.460	0.000	1.000	488
East North Central	0.410	0.000	0.492	0.000	1.000	488
West North Central	0.090	0.000	0.287	0.000	1.000	488
South Atlantic	0.023	0.000	0.149	0.000	1.000	488
East South Central	0.041	0.000	0.198	0.000	1.000	488
West South Central	0.002	0.000	0.045	0.000	1.000	488
Mountain	0.000	0.000	0.000	0.000	0.000	488
Pacific	0.002	0.000	0.045	0.000	1.000	488

Sample includes women who were widowed before 1880 and who applied for a pension within five years of widowhood. Sample drawn from Union Army Database (Fogel et al 2000). Data collected from Civil War pension files at the National Archives in Washington, DC.

Table 1.3: Linkage Rates to 1870 and 1880 Census

Group:	All	Married	Unmarried	Marital status unknown	
				all	has children from 1st marriage
Linkage Rate:					
1870 census	0.576	0.686	0.628	0.271	0.377
N	408	159	164	85	53
1880 census	0.606	0.681	0.764	0.184	0.262
N	467	204	165	98	61
Fraction linkable through children					
1870 census	0.711	0.523	0.879	1.000	1.000
N	235	111	124	23	20
1880 census	0.629	0.435	0.805	1.000	1.000
N	283	131	149	18	16

Sample includes all women widowed by relevant census year and who are not known to have died by this year. A woman is considered linkable through children if she is living with a child from her first marriage who has the same last name as her first husband.

Table 1.4: Estimated Fraction of Widows Observed in Union Army Data

Panel A: Distribution of Observations by Date of Death and Marital Status

Category:	Total	Died before 1880 (N=7,953)						Date of death unknown (N=11,552)		
		Total			Dead during war (N=5,777)					
		married	unmarried	unknown	married	unmarried	unknown	married	unmarried	unknown
N:	39,341	3,102	714	3,777	1,755	654	3,446	572	46	10,934

Panel B: Lower-Bound Estimates of the Fraction of Widows Observed in UA Data

	Assumption	Implied fraction of widows that appear in the UA data
Reference group:		
Died before 1880	A1: Everyone with missing death date died before 1880, and everyone with missing marital status was married.	16.9%
	A2: Everyone with missing death date died before 1880, and 63.4% of men with missing marital status were married	23.9%
Died during war	A3: 16% casualty rate, and everyone with missing marital status was married.	27.9%
	A4: 16% casualy rate, and 52.6% of soldiers with missing marital status were married.	45.7%

This table provides lower-boumd estimates of the fraction Union Army widows who filed pension applications, using different assumptions about missing data. Marriage rates in A2 and A4 are imputed marriage rates for the full sample and the sample killed in the war, respectively. These are based on marriage probabilities imputed from a regression of marital status on age, state, and occupational class using the 1860 1 percent IPUMS sample.

Table 1.5: Characteristics of Wives Identified in 1860 Census Links: Soldiers who died during war

	t test for equality of means			*OLS regression*
	Mean: Wife observed in pension data	Mean: Wife not observed in pension data	(1) - (2)	Dependent variable=1 if wife observed in pension data
Wife's age	29.4890	26.8571	**2.6320*** **	**-0.0096*** **
	(8.5129)	(8.1493)		(0.002)
Soldier's age	33.0864	26.5940	**6.4920*** **	**0.0193*** **
	(8.4058)	(8.2747)		(0.002)
Wife literate	0.9109	0.9323	-0.0200	0.0064
	(0.285)	(0.2521)		(0.048)
Soldier literate	0.9153	0.9699	**-0.0560** **	**-0.0845* **
	(0.2786)	(0.1714)		(0.050)
Wife immigrant	0.1359	0.2045	**-0.0680** **	**-0.0880* **
	(0.3429)	(0.4049)		(0.051)
Soldier immigrant	0.1542	0.1805	-0.0280	0.0107
	(0.3614)	(0.386)		(0.048)
HH head personal property	0.1708	0.6727	**-0.5000*** **	-0.0226
($1,000)	(0.2903)	(2.6348)		(0.014)
HH head real estate	0.4900	1.4727	**-0.9840*** **	**-0.0205*** **
($1,000)	(1.0522)	(4.3341)		(0.008)
Soldier farmer	0.3051	0.3083	-0.0040	
	(0.4608)	(0.4635)		
Soldier professional or proprieter	0.0336	0.0677	**-0.0360* **	-0.0518
	(0.1803)	(0.2521)		(0.067)
Solder skilled worker	0.2088	0.0752	**0.1320*** **	**0.1106*** **
	(0.4067)	(0.2647)		(0.035)
Soldier laborer	0.2146	0.2632	-0.0480	0.0064
	(0.4108)	(0.442)		(0.032)
Solder no occupation	0.0073	0.0075	-0.0002	0.1013
	(0.0852)	(0.0867)		(0.142)
Urban county	0.1441	0.1870	**-0.0440* **	**-0.1707*** **
	(0.2455)	(0.2938)		(0.057)
NE	0.4131	0.3534	0.0600	0.0000
	(0.4928)	(0.4798)		(0.000)
MW	0.5153	0.6165	**-0.1000** **	**-0.0506* **
	(0.5001)	(0.4881)		(0.028)
SO	0.0657	0.0301	0.0360	0.0649
	(0.2479)	(0.1714)		(0.057)
N	685	133		801

Sample of soldiers in UA data who died during the war, are linked to the 1860 census, and who appear to be married based on the composition of their household in 1860. Regression model includes a constant, and R2=0.175

Table 1.6: Determinants of the Hazard Rate of Remarriage and Pension Receipt

	(1)		(2)		(3)	
Outcome:	Remarriage	Pension	Remarriage	Pension	Remarriage	Pension
Effect of pension	-0.1923		-0.4867**		-0.5361**	
	(0.1602)		(0.1935)		(0.2213)	
Age at widowhood			-0.0967***	0.0065	-0.1147***	0.0065
			(0.0157)	(0.0103)	(0.0213)	(0.0105)
Number of Children			-0.1327**	-0.0359	-0.1600**	-0.0359
			(0.0605)	(0.0364)	(0.0784)	(0.0371)
Year of widowhood			-0.0463*	-0.0694***	-0.0679**	-0.0697***
			(0.0255)	(0.0169)	(0.0312)	(0.0170)
Time to pension application			-0.0346	-0.0998	-0.0504	-0.1032
			(0.1206)	(0.0967)	(0.1460)	(0.0968)
Potential minor pension at widowhood			0.1149	0.1402	0.1184	0.1407
			(0.1613)	(0.1174)	(0.1902)	(0.1188)
No pension attorney			0.1884	0.2669	0.3276	0.2680
			(0.3196)	(0.2484)	(0.3516)	(0.2491)
Washington pension attorney			0.0358	0.0859	-0.0642	0.0851
			(0.2125)	(0.1700)	(0.2986)	(0.1730)
First husband: age at death			0.0278*	-0.0132	0.0248	-0.0133
			(0.0153)	(0.0126)	(0.0173)	(0.0128)
First husband: log occupational wage			0.1280	-0.1646	0.2955	-0.1629
			(0.4251)	(0.2881)	(0.4562)	(0.2951)
First husband: height (feet)			-0.7049**	0.1418	-0.9165**	0.1417
			(0.3543)	(0.2696)	(0.4221)	(0.3077)
County male-to-female ratio			2.1664**	-0.0439	2.2800*	-0.0377
			(1.0528)	(1.1809)	(1.1672)	(0.9914)
County percent urban			0.2710	0.3678	0.3360	0.3681
			(0.3611)	(0.2828)	(0.3927)	(0.2723)
County population density			-0.0310	-0.0177	-0.0322	-0.0177
			(0.0208)	(0.0112)	(0.0229)	(0.0113)
Mid Atlantic			0.2804	-0.7319***	0.4838	-0.7345***
			(0.2674)	(0.2078)	(0.3866)	(0.2088)
East North Central			0.2525	-0.6505***	0.4748	-0.6539***
			(0.2608)	(0.2149)	(0.3659)	(0.2127)
West North Central			0.5323	-0.3906	0.9626*	-0.3948
			(0.3622)	(0.2943)	(0.5185)	(0.2990)
South			-0.4368	-0.7574**	-0.3212	-0.7615**
			(0.4183)	(0.2971)	(0.5355)	(0.3008)
λ for years:						
[1,2)	1.5122	1.0519***	1.8779	1.1925***	3.4782	1.1982***
	(0.4158)	(0.1409)	(0.5721)	(0.1748)	(1.8574)	(0.1767)
[2,3)	1.5575	0.6498***	2.3504*	0.8273***	5.9411	0.8340***
	(0.4432)	(0.1191)	(0.7447)	(0.1708)	(4.3221)	(0.1736)
[3,4)	1.5361***	0.4793***	2.7242***	0.7398***	7.7541	0.7470***
	(0.4538)	(0.1154)	(0.9048)	(0.1996)	(6.2593)	(0.2030)
[4,5)	1.2324***	0.3610***	2.5961***	0.7981***	7.8301	0.8094***
	(0.3906)	(0.1091)	(0.9380)	(0.2683)	(6.6760)	(0.2740)
[5,6)	0.9929***	0.2130**	1.9935**	0.3428*	6.2107	0.3480*
	(0.3388)	(0.0890)	(0.8000)	(0.1837)	(5.5455)	(0.1871)
[6,7)	0.5071**	0.1742**	1.3197**	0.3158	4.2473	0.3211
	(0.2172)	(0.0885)	(0.6135)	(0.1923)	(4.0310)	(0.1970)
[7,8)	0.6036**	0.1464*	1.2622**	0.3824	4.1337	0.3885
	(0.2489)	(0.0855)	(0.6110)	(0.2329)	(3.9533)	(0.2385)
[8. ∞)	0.0808***	0.2443***	0.2095***	0.5012***	0.7081	0.5099***
	(0.0256)	(0.0526)	(0.0781)	(0.1412)	(0.6612)	(0.1456)
v_{low} *(constant in columns 1-2)*	-2.4050***	-0.4650***	0.3372	0.6750	-0.2168	0.6105
	(0.2192)	(0.0891)	(3.4815)	(2.6406)	(3.7550)	(2.6362)
v_{high}					3.3428	0.7443
					(3.4592)	(2.7011)
π_1					0.5817	
					(0.7635)	
π_2					0.3364	
					(0.7946)	
π_3					0.0589	
					(0.1364)	
π_4					0.0229	
					(0.1043)	
Log Likelihood	-1377.486		-1141.420		-1140.0844	
Observations	482		429		429	

Hazard coefficients are reported. Sample: women who applied for a pension within five years of husband's death. Column (3) includes a correction for correlated unobserved heterogeneity, and does not include a constant as this is not identified separately from one of the mass points in the distribution of the unobserved heterogeneity terms; columns (1) and (2) make no such adjustment, and include a constant. Age at widowhood and all widows' pension variables (including county of residence) are taken from the pension file data collected by the author. First husband characteristics come from the UA data and are based on enlistment variables; occupational wages measured using 1900 occupational wage distribution assigned to 1950 occupational codes, with an imputed wage for farmers (Preston and Haines 1992; Abramitzky Boustan and Eriksson 2010; Olivetti and Paserman 2012). County-level variables are taken at the time of pension application; they are the weighted average of these variables at the decadal censuses preceding and following the date of pension application (Haines and ICPSR 2010). On the time interval [0,1), the hazard rate of both risks is normalized to one (this is necessary because I include a constant in the model). The variables v_{low} and v_{high} are the two mass points in the distributions of v_m and v_p. The variables π1-π4 are the estimated probability of each unobserved heterogeneity event.

Table 1.7: First Stage Results: Effect of Name Homogeneity Index on Pension

VARIABLES	(3) pen1	(5) pen2	(7) pen3	(9) pen4	(11) pen5
Name homogeneity index	0.1660*	0.2014**	0.1670**	0.1933**	0.2061***
	(0.091)	(0.087)	(0.082)	(0.077)	(0.074)
Age at widowhood	-0.0081*	0.0029	0.0033	0.0009	0.0030
	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)
Number of children	0.0025	-0.0107	-0.0095	-0.0098	-0.0145
	(0.015)	(0.014)	(0.013)	(0.012)	(0.012)
Year of widowhood	-0.0297***	-0.0345***	-0.0393***	-0.0393***	-0.0387***
	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)
Time to pension application	-0.0777**	-0.1107***	-0.0882***	-0.0653**	-0.0737***
	(0.031)	(0.029)	(0.027)	(0.025)	(0.025)
Potential minor pension	-0.0000	0.0001	0.0001*	0.0000	0.0000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
No pension attorney	0.0628	0.0791	-0.0060	-0.0056	-0.0229
	(0.103)	(0.097)	(0.090)	(0.084)	(0.082)
Washignton pension attorney	-0.1599**	-0.0647	-0.0504	0.0092	0.0372
	(0.069)	(0.066)	(0.062)	(0.059)	(0.057)
First husband: height	-0.1080	0.0285	0.0704	0.0238	-0.0288
	(0.111)	(0.108)	(0.103)	(0.099)	(0.096)
First husband: log occupational wage	0.0792	0.2034	0.2714**	0.0027	-0.0079
	(0.141)	(0.135)	(0.127)	(0.120)	(0.116)
First husband: age at death	0.0063	0.0003	-0.0002	0.0016	0.0010
	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)
County male-to-female ratio	0.4369	0.0476	-0.1043	-0.1993	-0.1620
	(0.455)	(0.431)	(0.401)	(0.374)	(0.363)
County percent urban	0.1287	0.1933*	0.1307	-0.0260	-0.0529
	(0.122)	(0.117)	(0.109)	(0.104)	(0.101)
County population density	-0.0000**	-0.0000	-0.0000	0.0000	0.0000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Last name: mean occupational income	0.2845	0.3510	0.3417	0.3826	0.3658
	(0.440)	(0.421)	(0.391)	(0.366)	(0.361)
Last mean: mean immigrant status	-0.0097	-0.1688	-0.0866	-0.0180	-0.0476
	(0.150)	(0.143)	(0.134)	(0.127)	(0.123)
Last name: mean literacy	-0.3253	-0.3456	-0.4106	-0.1187	-0.1211
	(0.422)	(0.404)	(0.380)	(0.358)	(0.347)
Constant	54.2775***	61.6002***	70.0858***	71.6802***	71.0150***
	(13.329)	(12.627)	(11.748)	(11.155)	(10.834)
Observations	368	362	356	348	347
R-squared	0.219	0.245	0.270	0.269	0.281
First stage F statistic	3.18	5.26	4.35	6.73	7.16

Name homogeneity index is Herfindahl index of unique surname spellings within phonetic name groups, calculated using household heads from IPUMS 1 percent samples, 1860-1880. See notes to table 6 for description of sample and remaining variables.

Table 1.8: OLS and 2SLS Estimates of Relationship between Pension and Remarriage

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Dependent variable: married w/in time frame (OLS)					*Dependent variable: married w/in time frame (2SLS)*				
Time Frame	1 year	2 years	3 years	4 years	5 years	1 year	2 years	3 years	4 years	5 years
Pension granted w/in time frame	-0.0473	-0.0810*	-0.2182***	-0.2736***	-0.2564***	-0.3180	-0.8348	-1.0244	-1.0627*	-1.1018**
	(0.032)	(0.049)	(0.057)	(0.063)	(0.066)	(0.365)	(0.518)	(0.656)	(0.557)	(0.534)
Age at widowhood	-0.0039	-0.0115***	-0.0156***	-0.0186***	-0.0211***	-0.0068	-0.0099*	-0.0139**	-0.0202***	-0.0204***
	(0.003)	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)	(0.005)	(0.006)	(0.005)	(0.006)
Number of children	-0.0131	-0.0175	-0.0287**	-0.0287**	-0.0313**	-0.0119	-0.0256	-0.0354*	-0.0353*	-0.0433**
	(0.009)	(0.013)	(0.014)	(0.014)	(0.015)	(0.010)	(0.018)	(0.019)	(0.018)	(0.019)
Year of widowhood	0.0040	0.0032	-0.0082	-0.0123*	-0.0137*	-0.0028	-0.0208	-0.0409	-0.0439*	-0.0466**
	(0.004)	(0.006)	(0.007)	(0.007)	(0.007)	(0.011)	(0.019)	(0.026)	(0.022)	(0.021)
Time to pension application	-0.0228	-0.0315	-0.0233	-0.0248	-0.0372	-0.0467	-0.1218*	-0.0892	-0.0711	-0.0963*
	(0.018)	(0.027)	(0.029)	(0.030)	(0.030)	(0.036)	(0.069)	(0.070)	(0.052)	(0.055)
Potential minor pension	-0.0000	-0.0000	0.0000	-0.0000	-0.0000	-0.0000	0.0000	0.0001	-0.0000	-0.0000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
No pension attorney	0.0665	0.0242	-0.0015	-0.0403	-0.0472	0.0857	0.0954	-0.0041	-0.0459	-0.0571
	(0.061)	(0.088)	(0.095)	(0.097)	(0.098)	(0.073)	(0.125)	(0.121)	(0.118)	(0.121)
Washignton pension attorney	0.0073	0.0346	-0.0060	0.0029	0.0105	-0.0382	-0.0131	-0.0686	-0.0210	0.0184
	(0.041)	(0.060)	(0.065)	(0.068)	(0.069)	(0.074)	(0.086)	(0.089)	(0.083)	(0.087)
First husband: height	-0.0666	-0.0666	-0.0886	-0.1981*	-0.2849**	-0.0868	-0.0377	-0.0389	-0.1797	-0.2991**
	(0.066)	(0.098)	(0.110)	(0.116)	(0.116)	(0.082)	(0.131)	(0.150)	(0.141)	(0.142)
First husband: log occupational wage	0.0090	-0.0151	0.0678	-0.1829	0.0091	0.0247	0.1148	0.2681	-0.1999	-0.0329
	(0.084)	(0.124)	(0.136)	(0.140)	(0.140)	(0.096)	(0.188)	(0.237)	(0.168)	(0.173)
First husband: age at death	0.0023	0.0050	0.0071	0.0058	0.0076	0.0043	0.0059	0.0079	0.0086	0.0093
	(0.003)	(0.004)	(0.005)	(0.005)	(0.005)	(0.004)	(0.006)	(0.006)	(0.006)	(0.006)
County male-to-female ratio	0.6539**	0.2972	0.4341	0.6191	0.9918**	0.8111**	0.3744	0.3307	0.4141	0.8311
	(0.267)	(0.385)	(0.420)	(0.429)	(0.431)	(0.340)	(0.517)	(0.543)	(0.539)	(0.545)
County percent urban	0.0453	-0.0520	0.0151	0.0245	0.0529	0.0813	0.0940	0.1253	-0.0021	0.0185
	(0.073)	(0.107)	(0.117)	(0.121)	(0.122)	(0.094)	(0.174)	(0.171)	(0.146)	(0.151)
County population density	-0.0000	0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Last name: mean occupational income	0.0823	-0.1783	-0.4119	0.0008	0.1302	0.2760	0.1466	-0.0387	0.6612	0.8199
	(0.228)	(0.331)	(0.361)	(0.370)	(0.377)	(0.307)	(0.532)	(0.566)	(0.552)	(0.563)
Last mean: mean immigrant status	0.0279	-0.1311	-0.1325	-0.2075	-0.1505	0.0110	-0.2973	-0.2040	-0.2099	-0.2446
	(0.085)	(0.123)	(0.135)	(0.140)	(0.140)	(0.100)	(0.208)	(0.201)	(0.180)	(0.189)
Last name: mean literacy	0.0665	0.2034	0.3486	0.2987	0.0060	0.0083	-0.0237	-0.0970	0.0570	-0.0104
	(0.165)	(0.239)	(0.261)	(0.267)	(0.268)	(0.304)	(0.516)	(0.571)	(0.505)	(0.516)
Constant	-8.2092	-4.2195	18.1355	25.5147*	26.4345**	3.5142	38.0825	76.3273	81.3946**	84.6270**
	(7.861)	(11.471)	(12.635)	(13.207)	(13.331)	(20.909)	(33.982)	(47.071)	(41.250)	(39.485)
Observations	375	369	363	355	354	368	362	356	348	347
AR 95% Confidence Region for pension effect						(-∞, +∞)	[-5.57, -0.057]	[-20.61, 0 .031]	[-4.390, -0.148]	[-3.686, -0.247]
R-squared	0.065	0.094	0.173	0.235	0.256					

Instrument used in 2SLS specification is name homogeneity index. "Last name" variables are means by phonetic name group among household heads in IPUMS 1 percent sample from 1860-1880. AR 95% confidence region for pension effect is 95 percent confidence interval for the effect of the pension based on the Anderson-Rubin statistic, which is robust to weak instruments. See notes to table 6 for description of sample and other variables.

Table 1.9: Sensitivity of Estimates to Sample Restrictions

Model:	Simple	Full
	Panel A. Baseline	
Effect of Pension on Marriage Rate	-0.4867**	-0.5361**
	(0.1935)	(0.2213)
Log-Likelihood	-1141.420	-1140.0844
Observations	429	429
	Panel B. Limit to time with minor children	
Effect of Pension on Marriage Rate	-0.2782	-0.2967
	(0.2111)	(0.2406)
Log-Likelihood	-943.556	-941.6477
Observations	339	339
	Panel C. Limit to Information from General Law Pension Applications	
Effect of Pension on Marriage Rate	-0.3980*	-0.5024**
	(0.2100)	(0.2427)
Log-Likelihood	-1023.468	-1021.626
Observations	429	429
	Panel D. Linked Only	
Effect of Pension on Marriage Rate	-0.4777**	-0.4915**
	(0.2181)	(0.2428)
Log-Likelihood	-896.135	-893.9608
Observations	302	302
	Panel E. Linked Only: immigrant and literacy controls	
Effect of Pension on Marriage Rate	-0.4675**	-0.4597*
	(0.2206)	(0.2449)
Log-Likelihood	-892.845	-890.5271
Observations	302	302
	Panel G. Husband died during war	
Effect of Pension on Marriage Rate	-0.4635**	-0.4751**
	(0.2191)	(0.2320)
Log-Likelihood	-902.990	903.314
Observations	348	348

All specifications include the full set of controls from table 6; see notes to this table for explanation. The top panel replicates the baseline results. Panel B restricts the analysis to years in which the widow has a child under the age of 16. Panel C discards information that comes from applications under the law of March 3, 1901. Panel D restricts the sample to women who are successfully linked to the census of 1870 or 1880. Panel E poses a similar restriction, but includes imigrant and literacy controls available in the census data.

Table 1.10: Pensions and the Timing of Remarriage: Widows Linked to Census through Children

	(1)	(2)	(3)	(4)	(5)	(6)
			Dependent variable: Remarried in Census			
Model:		*OLS*			*2SLS*	
Year:	1870	1880	Pooled	1870	1880	Pooled
=1 if pensioned w/in 5 years	-0.2304**	-0.1561	-0.1511*	-3.2024	-0.7379*	-1.1183
	(0.112)	(0.098)	(0.079)	(5.137)	(0.433)	(0.776)
Age at widowhood	-0.0226***	-0.0255***	-0.0244***	-0.0156	-0.0219***	-0.0203***
	(0.007)	(0.007)	(0.004)	(0.021)	(0.008)	(0.007)
Number of children	-0.0743***	-0.0356*	-0.0516***	-0.1451	-0.0377*	-0.0646***
	(0.024)	(0.019)	(0.016)	(0.134)	(0.020)	(0.021)
Year of widowhood	-0.0130	-0.0097	-0.0087	-0.0422	-0.0338*	-0.0469
	(0.030)	(0.009)	(0.008)	(0.088)	(0.019)	(0.031)
Time to pension application	-0.0574	-0.0690*	-0.0693**	-0.3429	-0.1072**	-0.1444*
	(0.061)	(0.040)	(0.028)	(0.524)	(0.052)	(0.082)
Potential minor's pension	0.0000	0.0000	0.0000	-0.0005	0.0000	-0.0000
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
No pension attorney	0.0465	-0.1756	-0.0592	0.2369	-0.2592	-0.0664
	(0.133)	(0.153)	(0.113)	(0.460)	(0.164)	(0.177)
Washington pension attorney	-0.0014	0.0672	0.0329	0.2398	0.1187	0.1132
	(0.108)	(0.108)	(0.078)	(0.492)	(0.130)	(0.123)
First husband: height	-0.0745	0.0061	-0.0114	-0.4418	0.0254	-0.0158
	(0.191)	(0.184)	(0.155)	(0.777)	(0.195)	(0.179)
First husband: log occupational income	0.2998	0.2280	0.2371	0.7901	0.2782	0.3190
	(0.215)	(0.195)	(0.179)	(1.020)	(0.211)	(0.330)
First husband: age at death	0.0075	0.0111	0.0089	0.0128	0.0116	0.0105
	(0.009)	(0.008)	(0.006)	(0.022)	(0.008)	(0.008)
Coutny male-to-female ratio	0.5028	0.8729	0.8098*	-4.0906	1.0195*	0.5259
	(1.016)	(0.549)	(0.452)	(8.294)	(0.596)	(0.634)
County percent urban	0.0597	0.1430	0.1198	-1.2462	0.1867	-0.0013
	(0.228)	(0.196)	(0.179)	(2.319)	(0.203)	(0.243)
County population density	-0.0000	-0.0000	-0.0000*	0.0000	-0.0000	-0.0000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Literate	-0.1888*	0.1791	-0.0370	-0.3085	0.1809	-0.0697
	(0.102)	(0.124)	(0.080)	(0.321)	(0.139)	(0.089)
Immigrant	0.0499	-0.0900	-0.0522	0.5253	-0.1118	-0.0137
	(0.134)	(0.117)	(0.063)	(0.879)	(0.127)	(0.102)
Year=1880			0.0209			0.0457
			(0.041)			(0.053)
Constant	23.7010	16.5803	15.0851	85.0484	61.4020*	86.9513
	(56.548)	(16.361)	(14.425)	(170.098)	(36.525)	(58.893)
Observations	147	158	305	144	153	297
R-squared	0.328	0.352	0.306			
First stage F statistic				0.333	7.917	4.008

Sample includes women linked to the census of 1870 or 1880 who are residing with a child from their first marriage who has kept his or her deceased father's last name. In columns (3) and (6), both census years are pooled and standard erorrs clustered by widow. See notes to table 6 for a detailed description of the variables and the notes to table 8 for a description of the instrument used in columns 4-6.

Table 1.11: Pensions and Match Quality

VARIABLES	(1) Husband log occupational wage	(2) Husband literate	(3) Husband-wife age difference^2	(4) Husband present in household
Remarried after pension	0.0581	-0.0048	-36.8906	0.0010
	(0.091)	(0.058)	(77.837)	(0.094)
Number of children	-0.0432	0.0290	-23.6364	0.0234
	(0.033)	(0.033)	(16.468)	(0.029)
Number of children X pension	0.0097	-0.0468	7.4168	-0.0485
	(0.042)	(0.038)	(19.893)	(0.037)
Potential minor pension	0.0700	-0.0935	11.3229	-0.0477
	(0.091)	(0.094)	(54.726)	(0.089)
Potential minor pension X pension	-0.0930	0.0657	4.7589	0.1079
	(0.110)	(0.103)	(62.245)	(0.107)
Age in census	0.0095	0.0079	4.5264	-0.0059
	(0.010)	(0.006)	(3.590)	(0.009)
Literate	0.1302*	0.4953***	-5.2923	0.0731
	(0.072)	(0.109)	(44.771)	(0.065)
Immigrant	0.1496	0.0190	-33.0041	0.1550**
	(0.109)	(0.056)	(38.211)	(0.066)
Age at widowhood	-0.0053	-0.0072	-1.2926	-0.0014
	(0.010)	(0.008)	(5.888)	(0.011)
Age at remarriage	0.0031	-0.0013	-0.2236	-0.0003
	(0.011)	(0.005)	(4.296)	(0.009)
First husband: age at death	-0.0054	0.0055	-0.0419	-0.0005
	(0.005)	(0.003)	(3.319)	(0.005)
First husband: height	0.0720	0.0581	-57.2034	-0.0020
	(0.099)	(0.068)	(53.081)	(0.081)
First husband: log occupational wage	0.0323	0.0557	280.3327**	0.2091
	(0.207)	(0.119)	(116.120)	(0.176)
County male-to-female ratio	-0.0199	-0.0053	-16.5068	-0.0033
	(0.023)	(0.011)	(11.165)	(0.033)
County population density	-0.0156	0.0090	-4.9571	0.0313
	(0.023)	(0.011)	(11.063)	(0.026)
County percent urban	0.0737	0.0335	-89.6366	-0.1637
	(0.142)	(0.066)	(56.799)	(0.173)
Census year = 1870	0.1009	0.0905*	57.5048	0.0452
	(0.094)	(0.050)	(42.502)	(0.095)
Constant	5.4848***	-0.4433	-1,434.0604**	-0.1901
	(1.376)	(0.720)	(679.590)	(1.187)
Observations	191	197	197	221
R-squared	0.145	0.485	0.111	0.097

Sample consists of remarried widows who linked to the census of 1870 or 1880. These census years are pooled, and stanard errors are clustered by widow. County-level variables from Haines and ICPSR (2010). See table 6 for description of other variables.

Table 1.12: Determinants of Linkage to 1870 and 1880 Census among Widows

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent variable:	*Linked to 1870 census*					*Linked to 1880 census*				
Age at Widowhood	-0.0026	-0.0027	-0.0027	-0.0001	-0.0000	-0.0014	-0.0042	-0.0042	-0.0021	-0.0018
	(0.003)	(0.003)	(0.003)	(0.005)	(0.005)	(0.002)	(0.003)	(0.003)	(0.004)	(0.004)
Number of Children	0.0497***	0.0493***	0.0493***	0.0610***	0.0591***	0.0613***	0.0601***	0.0601***	0.0627***	0.0615***
	(0.012)	(0.012)	(0.012)	(0.013)	(0.014)	(0.010)	(0.010)	(0.010)	(0.011)	(0.011)
Year of Widowhood		-0.0051	-0.0051	-0.0019	-0.0004		0.0157**	0.0157**	0.0131**	0.0137**
		(0.018)	(0.018)	(0.019)	(0.019)		(0.006)	(0.006)	(0.007)	(0.007)
Time to first application		-0.0248	-0.0248	-0.0034	-0.0122		0.0169	0.0169	0.0309	0.0282
		(0.040)	(0.040)	(0.042)	(0.042)		(0.028)	(0.028)	(0.029)	(0.029)
No pension attorney		0.0832	0.0832	0.0738	0.0748		0.0325	0.0325	0.0272	0.0242
		(0.102)	(0.102)	(0.108)	(0.108)		(0.093)	(0.093)	(0.098)	(0.098)
Washington pension attorney		0.0995	0.0995	0.0845	0.0864		0.0047	0.0047	0.0065	0.0125
		(0.070)	(0.070)	(0.072)	(0.073)		(0.064)	(0.064)	(0.067)	(0.067)
Pensioned w/in 5 yrs		-0.0021	-0.0021	-0.0144	-0.0071		0.0675	0.0675	0.0735	0.0779
		(0.075)	(0.075)	(0.077)	(0.077)		(0.064)	(0.064)	(0.066)	(0.066)
First husband: log occupational wage				-0.1218	-0.1332				-0.0202	-0.0328
				(0.138)	(0.139)				(0.123)	(0.124)
First husband: age at death				-0.0045	-0.0043				-0.0032	-0.0032
				(0.005)	(0.005)				(0.005)	(0.005)
First husband: height				0.0823	0.0518				0.1745*	0.1420
				(0.108)	(0.110)				(0.099)	(0.100)
Region of enlistment: Midwest					0.0533					0.0639
					(0.055)					(0.049)
Region of enlistment: South					0.1918					0.2060*
					(0.120)					(0.111)
Region of enlistment: West					0.0000					0.0000
					(0.000)					(0.000)
Constant	0.5346***	9.9560	9.9560	4.4928	1.7777	0.5014***	-28.7767**	-28.7767**	-24.8121**	-25.7085**
	(0.091)	(34.445)	(34.445)	(35.956)	(35.989)	(0.077)	(11.379)	(11.379)	(12.310)	(12.288)
Observations	401	401	401	376	376	458	458	458	426	426
R-squared	0.042	0.049	0.049	0.067	0.074	0.074	0.093	0.093	0.102	0.111

Age at widowhod, time to first application, and all widows' pension variables are taken from the pension file data collected by the author. The remaining variables are from the Union Army database. First husband characteristics based on enlistment variables; occupational wages measured using 1900 occupational wage distribution assigned to 1950 occupational codes, with an imputed wage for farmers (Preston and Haines 1992; Abramitzky Boustan and Eriksson 2010; Olivetti and Paserman 2012). Sample includes women widowed by the relative census year and who are not known to have died by this year.

Table 1.13: Sample Means: Linked Widows Data versus Nationally Representative IPUMS Data

	Married Women			Unmarried Women (16+)			Female Household Heads		
	Linked widows data	IPUMS	IPUMS age adjusted	Linked widows data	IPUMS	IPUMS age adjusted	Linked widows data	IPUMS	IPUMS age adjusted
					Panel A. 1870				
Age	32.41	36.67	32.82	39.71	32.26	40.96	40.76	48.39	40.55
Family size	5.58	5.47	5.90	5.18	6.19	5.51	4.75	4.01	4.44
Works	0.00	0.01	0.01	0.16	0.26	0.27	0.14	0.15	0.19
Literate	0.84	0.86	0.86	0.88	0.86	0.81	0.84	0.77	0.78
Immigrant	0.08	0.26	0.29	0.11	0.19	0.26	0.12	0.24	0.25
Reas estate	75.96	61.73	60.43	540.58	342.49	717.47	739.80	1593.00	1430.58
Personal property	8.17	26.64	19.79	166.08	178.48	330.22	220.66	710.22	534.71
Husband's age	38.75	41.45	37.93						
Husband's occupational income (1900)	578.84	595.54	603.37						
Husband literate	0.88	0.90	0.90						
Husband's real estate	1204.18	2325.95	2096.40						
Husband's personal property	517.04	1048.54	1016.49						
County population density (1000/mile^2)	0.18	1.44	1.61	0.78	1.90	2.37	0.96	2.10	2.25
County percent urban	0.06	0.27	0.29	0.06	0.31	0.34	0.06	0.30	0.31
County male-to-female ratio	1.11	1.03	1.03	1.02	1.01	1.00	1.02	1.01	1.01
Northeast	0.36	0.38	0.37	0.40	0.42	0.44	0.40	0.39	0.36
Midwest	0.62	0.38	0.39	0.50	0.31	0.26	0.51	0.28	0.28
South	0.02	0.22	0.21	0.10	0.26	0.27	0.09	0.31	0.33
West	0.01	0.02	0.02	0.00	0.02	0.02	0.00	0.02	0.03
Observations	104	57,217	34,247	139	40,544	16,669	99	6,951	5,018
					Panel B. 1880				
Age	41.22	37.41	40.95	48.41	33.13	48.96	47.90	50.81	48.25
Family size	6.12	5.31	6.08	4.30	5.91	4.87	3.88	3.81	4.08
Works	0.02	0.02	0.02	0.11	0.29	0.24	0.11	0.20	0.22
Literate	0.96	0.89	0.89	0.85	0.90	0.85	0.82	0.83	0.83
Immigrant	0.04	0.24	0.30	0.14	0.18	0.28	0.16	0.30	0.33
Husband's age	46.64	42.21	45.75						
Husband's occupational income (1900)	590.60	610.90	620.69						
Husband literate	0.94	0.91	0.91						
County population density (1000/mile^2)	0.33	1.44	1.56	0.82	2.09	2.30	0.63	2.10	2.21
County percent urban	0.07	0.30	0.32	0.09	0.36	0.38	0.07	0.37	0.37
County male-to-female ratio	1.07	1.05	1.04	1.03	1.02	1.01	1.02	1.02	1.02
Northeast	0.29	0.34	0.36	0.38	0.40	0.42	0.35	0.40	0.37
Midwest	0.65	0.39	0.39	0.51	0.34	0.29	0.55	0.31	0.32
South	0.04	0.23	0.21	0.09	0.24	0.27	0.09	0.26	0.27
West	0.02	0.03	0.03	0.01	0.03	0.03	0.01	0.04	0.04
Observations	112	74,698	45,068	171	54,697	25,264	128	9,869	7,428

See appendix B for description of variables. The IPUMS age adjusted sample is weighted to have the same distribution of five-year birth cohorts as the analogous sample in the linked widows data.

Table 1.14: Composition of Households in Linked Widows Data

	1870		1880	
	Married	Unmarried	Married	Unmarried
Resides with child having first husband's surname	0.51	0.87	0.44	0.75
Resides with chlld of first husband (imputed)	0.77	0.87	0.51	0.75
Resides with second husband's child from previous marriage	0.32		0.22	
Has children from second marriage	0.58		0.74	
# Children from previous marriage (widow)	1.49	2.78	0.94	2.10
# Children from previous marriage (husband)	0.63		0.33	
# Children from second marriage	1.05		2.33	
Widow is household head	0.00	0.71	0.00	0.75
Observations	104	139	112	171

Figure 1·1: Possible Outcomes for Widows in Sample

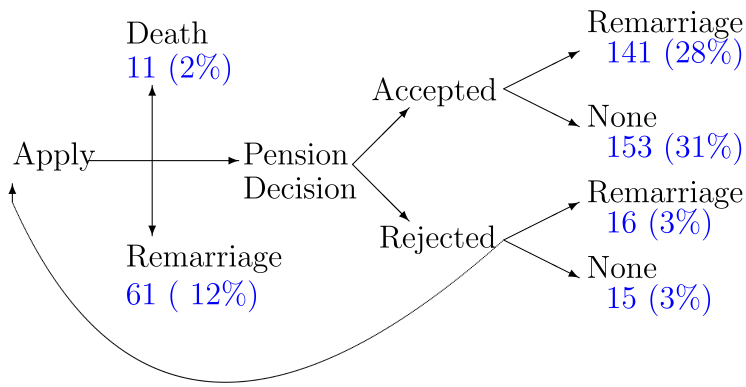
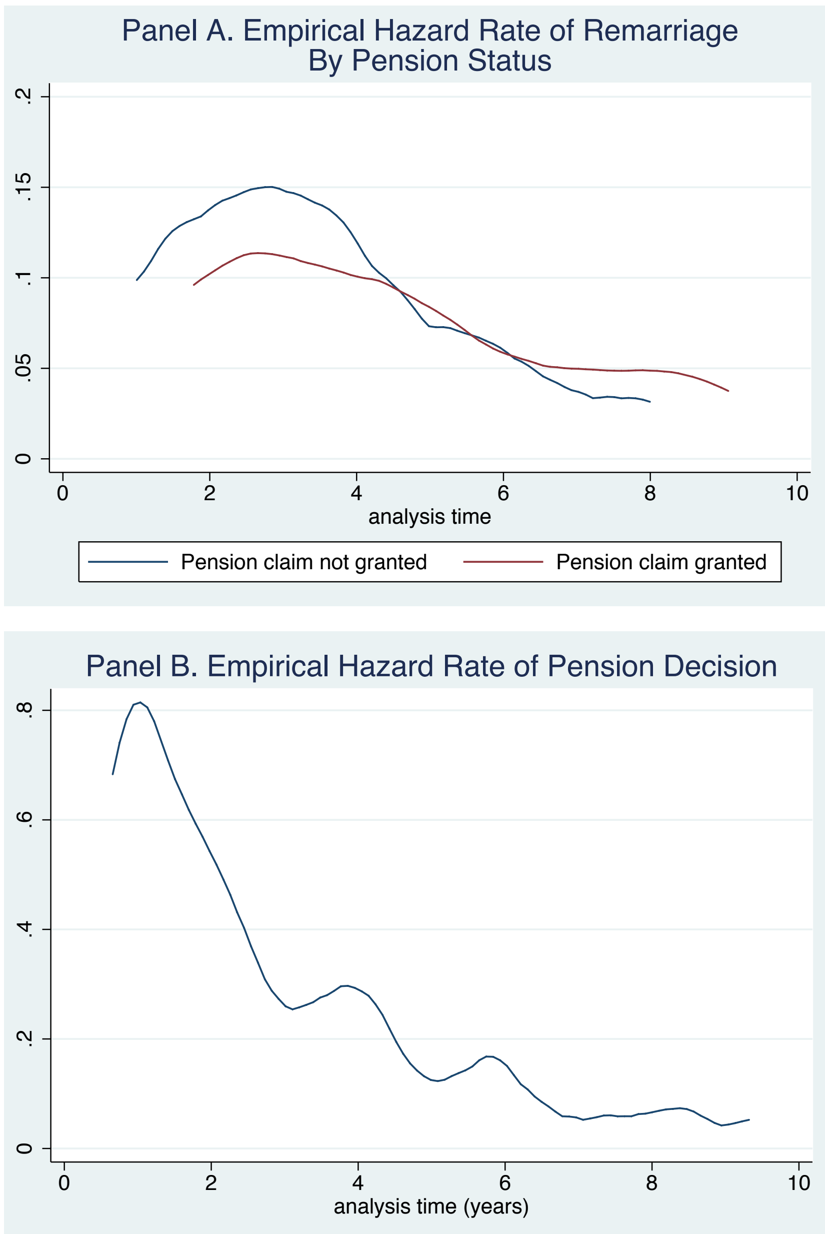
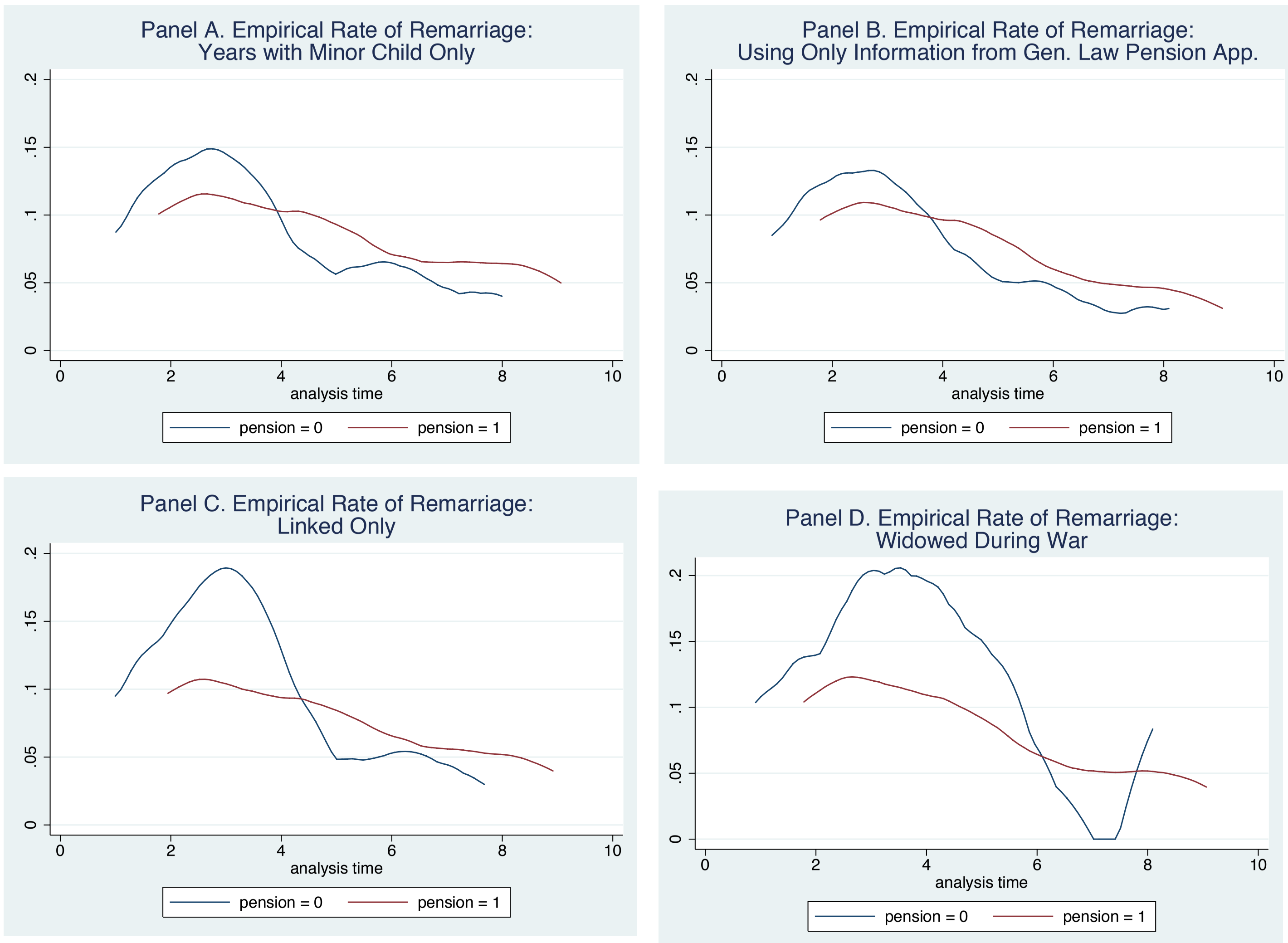


Figure 1·2: Empirical Hazard Rate of Remarriage and Pension Decision



Panel A plots the nonparametric empirical hazard rate of remarriage, separated by pension status, and etimated using kernal method (STS package in STATA). Panel B does the same for the hazard rate of pension receipt. The picture is cut off at t=10 because the rate of remarriage cannot be estimated for women without pensions at t>10.

Figure 1·3: Empirical Hazard Rate of Remarriage under Different Sample Restrictions



Nonparametric empirical hazard rate of remarriage, separated by pension status, and etimated using kernal method (STS package in STATA). Pictures are truncated at t=10 for ease of comparison; the rate of remarriage among women without pensions cannot be estimated at t>10. See notes to table 9.

Chapter 2

Selective Migration, Wages, and Occupational Mobility in Nineteenth Century America

2.1 Introduction

This paper analyzes the connection between occupational and geographic mobility in the United States between 1850 and 1880. During this period, internal migration was more common than it was at the turn of the 20th century (Rosenbloom and Sundstrom 2003). Between 20 and 25 percent of white native-born Americans were residing outside their state of birth (Hall and Ruggles 2004), a figure that masks considerable within-state migration. Internal migration in the nineteenth century was also fundamentally different from such migration in the United States today. In particular, many migrants moved from settled to unsettled parts of the county. For this reason, the opportunities available to them in their destinations would have looked different from those at home.

The concept of opportunity on the frontier has informed much of the work on internal migration during this period.[1] For example, Ferrie (1997) finds evidence that frontier migrants between 1850 and 1870 were negatively selected, providing some empirical support for the hypothesis that migrants to the American West were largely unskilled laborers seeking to escape less attractive working conditions in the more

[1] See the following section for a review.

urban East. An implication of this hypothesis is that migration was associated with career mobility. The iconic example is the urban wage worker who became a farmer; however, differences in the industrial composition of regions, or in the availability of skilled workers, may have affected opportunities in the non-agricultural sector as well. Kim (1998) shows that overall economic activity in the United States was more specialized by region in 1870 and 1880 than it was by the late 1980s, which implies that job prospects varied geographically. This idea adds a level of complexity to a simple migration model, which predicts a positive relationship between average wages and in-migration. Migrants may have maximized income in a way that we cannot observe through data on average wages if the possibility of upward occupational mobility factored into their decisions.

In this paper, I explore the extent to which the prospect of occupational upgrading motivated internal migration in the United States in the nineteenth century. In particular, I consider the capacity for upward mobility as a "skill" by which migrants may have selected and sorted themselves, and I analyze the effect this had on migration patterns. Accounts of the settlement of the United States are littered with references to the qualities that are thought to have enabled this settlement: resourcefulness, adaptability, or perseverance, for example. Lebergott quotes an 1818 description of the American work force, which lauds the "versatility of [the American laborer's] talent (1964, p. 8)." In reference to the particular qualities that enabled success on the frontier, Goodrich and Davison quote an 1853 description of a new colony in Minnesota, which reports that "none but the persevering ones [are] staying... Most of those who left were mechanics and artisans, principally from New York City, who knew nothing about farming or the wants of a new settlement (1936, p 98)." While these characteristics are difficult to quantify, one might expect them to be associated with a propensity for upward occupational mobility.

Drawing on the modern migration literature, I argue that if migrants selected and sorted themselves on the basis of upgrading potential, there should be a systematic relationship between wage profiles in home and destination counties and migrants' propensity for occupational mobility. Workers who were likely to transition into skilled occupations should have selected themselves out of places in which the return to skill was low and sorted themselves into places in which the return to skill was high. One reason we should expect to see this is that the return to occupational upgrading was higher in areas with greater returns to skill. Another possibility is that large skill premiums signal opportunities for upgrading: these may reflect relatives shortage of skilled workers. By and large, counties with large skill premiums were less urban, which could mean that skilled labor was in shorter supply. It could also mean that demand for skilled labor was higher, if these places lagged behind urban areas in the adoption of new production technologies that favored unskilled labor over skilled artisan labor.[2] These counties should have been most attractive to workers who were best able to take advantage of the opportunities that existed there.

To evaluate this hypothesis, I take a sample of linked census data and combine it with county-level wage data. I measure occupational mobility by an individual's change in log occupational income over the course of a decade, and I measure the skill premium in a county as the log ratio of carpenters' to laborers' wages. I show that the return to migration, in the sense of occupational upgrading, was increasing in the destination county skill premium and decreasing in the home county skill premium. This suggests that migrants from counties with small skill premiums were positively selected in terms of upgrading potential. Conditioning on migrating, I find that workers who moved to places with large skill premiums were more likely to upgrade than those who moved to places with small skill premiums. Again, this suggests that migrants sorted themselves based on upgrading potential. I close by offering

[2] See Atack Bateman and Margo (2000)

suggestive evidence for my interpretation of this result.

2.2 Related Literature

2.2.1 Migration, Selection, and Sorting

Work on internal migration in the United States in the nineteenth century focuses on two questions. One is the relationship between wage levels and migration patterns. The well-known Easterlin paradox (Margo 1999) posits that internal migration moved in the "wrong" direction, in the sense that it did not tend to follow high per capita incomes; migration tended to flow East to West, even though average incomes were higher in the East. Lebergott (1964) finds a weak relationship between average wages and state population growth during the nineteenth century. As long as cross-state differences in population growth are driven by migration, this suggests a weak migration response to regional wage differentials. He claims that "migrants suboptimized" (Lebergott 1964, p 76), in that sense that they failed to take advantage of potential wage gains in different parts of the country. These puzzles can be explained in part by regional price differences: real wages tended to be higher in the West than in the East (Coelho and Shephard 1976; Margo 1999).

Another strain of research on internal migration in the United States focuses on characterizing frontier migrants.[3] This is broadly motivated by the Turner "safety valve" hypothesis: the presence of the frontier relieved pressure on eastern labor markets, as unskilled urban wage workers could go west as working conditions in the urban East deteriorated. An implication of this hypothesis is that frontier migrants were disproportionately drawn from the lower tail of the occupational distribution, a proposition that was impossible to test before the emergence of large samples of linked census data. Using a sample of males linked between the 1850, 1860 and

[3] See Steckel (1983) and Shaefer (1985), for example.

1870 federal censuses, Ferrie (1997) finds some evidence that frontier migrants were negatively selected; in other words, they tended to be unskilled workers with lower potential for wealth acquisition than stayers in the East. Stewart (2005) uses linked census data to show that households who moved to the frontier between 1860 and 1870 tended to be poorer, less literate, and less likely to own land than households who did not migrate. Galenson and Pope (1989) find evidence of significant economic opportunity on the farming frontier, looking specifically at Appanoose county, Iowa. They find that the return to frontier migration, measured by wealth acquisition, was highest for early settlers. Abramitzky Boustan and Eriksson (2010) find evidence of negative selection of Norwegian immigrants to the United States around the turn of the twentieth century.

There is a well developed literature on the relationship between home and destination wage profiles and the characteristics of migrants, particularly during recent periods. This literature addresses differences in the skills of movers and stayers from a given home area (selection) and differences among migrants to various destination areas (sorting). Borjas (1987) argues that migrant selection accounts for part of the observed wage differential between immigrants and native born Americans, as immigrants from countries with higher levels of wage dispersion than the United States tend to be negatively selected. Borjas, Bronars and Treo (1992) focus on migrant sorting within the United States. They argue that skilled individuals sort themselves into states with high earnings variance, as the return to skill is highest in these places. Grogger and Hanson (2008) address both selection and sorting of migrants. They find that skilled migrants are selected out of countries with a small gap between skilled and unskilled wages and sorted into countries where this gap is large. They also find that international migrants tend to be positively selected in general, so immigrants on average are more skilled than stayers.

The concepts of migrant selection and sorting based on wage profiles may be relevant to internal migration in nineteenth century America. For one thing, it suggests a more nuanced approach than coarsely distinguishing "East" from "West". As Lebergott notes, "there have been many 'Wests' (1964, p. 13)." Selection implies that a person who migrates from Massachusetts to Minnesota will differ from one who migrates from Michigan to Minnesota, for example. Similarly, sorting implies that a person from Massachusetts who migrates to Michigan will be differently skilled than one who migrates to Minnesota. The modern literature on migrant selection measures skill as educational attainment, a function of wages, or an aptitude test score. As described earlier, other skills reflecting potential for occupational upgrading ought to be more relevant during the period of focus of this study.

2.2.2 Occupational Mobility

Other work explores the nature of occupational mobility in nineteenth century America. Today, a sizable human capital investment is needed to ascend most job ladders. However, there is (anecdotal) evidence that this was not true of 19th century America. Lebergott quotes several contemporary accounts of the ease with which unskilled workers could perform tasks which, in Europe, were restricted to those with particular training: "Every man can use an axe, a saw and a hammer. Scarcely one who cannot do any job of rough carpentering, and mind a plough or a wagon.[4]" A Swedish immigrant in the 1850s describes "a person I have seen going about working as a mason [who] served for a couple of months as an assistant in a drug store in Milwaukee, whereupon he laid aside the trowel, got himself some medical books, and assumed the title of doctor.[5]" He remarks that "the speed with which people here change their life-calling and the slight preparation generally needed to leave one

[4]Lebergott page 6, quoting Cobbett 1818
[5]p 120, quoting Olsson 1950

calling for another are really surprising, especially to one that has been accustomed to our Swedish guild-ordinances." More recent work by Ferrie (2005) and Long and Ferrie (2005) has shown that intergenerational occupational mobility was higher in the United States than in Britain through the early 20th century.

Other work has assessed the determinants of occupational mobility in different sectors of the American economy during the mid to late 19th century. Ferrie (1996b) looks at occupational mobility among European immigrants before the Civil War. He traces migrants from passenger ship lists between 1840 and 1850 to federal censuses in 1850 and 1860. He finds that many immigrants who held skilled occupations in their home countries worked as unskilled laborers immediately after emmigrating; however, these immigrants were likely to transition out of unskilled jobs by 1860. Immigrants' labor market performance also varied by country of origin and literacy, for example.

Other work has focused explicitly on mobility in agriculture. The agricultural ladder consists of wage workers, sharecroppers (in the south), tenants, and owners.[6] The prevalence of different tenancy arrangements differed by region and over time; however, farming hierarchies existed even in the antebellum north where land was supposedly freely available[7]Atack (1988) finds that 16 percent of northern farms were operated by tenants in 1860. Many contemporaries considered wage labor and tenant arrangements stepping stones into self-employed farming, and there is some evidence that this was the case.[8] And, there is evidence that individuals engaged in these various arrangements differed by age, immigrant status, schooling, and wealth, among other dimensions. Alston and Ferrie (2005), using data from Jefferson County Arkansas between 1890 and 1938, find that age and education tended to increase mobility among tenant farmers (but not farm laborers), and inheritance tended to

[6]One can also distinguish between share tenants and cash tenants, and owners and part owners. See Atack (1988, 1989), Atack and Bateman, Alston and Ferrie (2005), Alston and Kauffman (1997).

[7].

[8]See Atack (1988, 1989).

increase upward mobility across the board. Overall, they find that the rate of upward mobility among farm laborers was comparable to that in the nonfarm population. To summarize, mobility was common during this period, and it varied by individual. This supports the characterization of aptitude for upward mobility as an important individual skill.

2.2.3 Wages

There is a literature on the evolution of wages in the United States, beginning in the colonial period. While real wage growth during the antebellum period kept pace with productivity growth on average, it was uneven across time and space (Margo 2000, p 4). The South consistently lagged behind the North (Margo 2000), and the Midwest had higher real wages, though lower nominal wages, than the Northeast for much of the nineteenth century (Coelho and Shepherd 1976; Margo 1999). There is not much evidence of regional convergence until after the Civil War (Rosenbloom 1996).

There is also a large literature on patterns of wage inequality in the United States. Lindert (1976) summarizes this overall pattern as follows:

> Throughout the antebellum period, starting around 1820, wide earnings gaps opened up, skill premia were on the rise, and wealth concentration accelerated. In short, skilled labor, professional groups, and urban wealth holders prospered much faster than farm hands and the urban unskilled. A dramatic change in northeastern America's income distribution was largely complete by 1860 or 1880. After the Civil War, earnings and total income inequality fluctuated around historically high levels with one last secular inequality surge, at least in urban America, appearing from the 1890s to World War I (Lindert 1976, pp 5-6).

The trend in wage inequality can be attributed largely to changes in America's industrial structure: Atack, Bateman and Margo (2000) attribute a growth in wage inequality between 1850 and 1880 to in increase in the number of unskilled workers in large firms, which tended to pay lower wages during this period.

This paper will rely on geographic variation in wage inequality within the United States. There is less evidence about this in the existing literature. In the following section, I will present very casual evidence that skill premiums tended to be higher in less urban counties. In theory, this would be consistent with the relative supply of skilled workers being greater in more urban areas. It would also be consistent with rural areas lagging behind urban areas in the adoption of new industrial technologies, thus causing the demand for skilled artisan labor in these areas to exceed that of urban areas.

2.3 Data

This paper uses data from three major sources. The first two are nationally representative samples of individuals in two consecutive decades, constructed from census data; the 1850-60 sample is from Joseph Ferrie,[9] and the 1870-80 sample is from IPUMS.[10] The third is a sample of county-level wage data from Robert Margo[11] covering 1850, 1860, and 1870.

The IPUMS sample links observations in the 1880 complete-count database to observations in the 1870 one percent sample. The Church of Jesus Christ of Latter-Day Saints recently digitized 100 percent of the 1880 census manuscripts; IPUMS finalized its version of these data in January of 2010. Links to the 1870 census are based on six variables: birth year, place of birth, last name, first name, race, and age. These variables were chosen because, in theory, they should be identical across years

[9]See Ferrie (1996a) for details.
[10]Ruggles et al (2010)
[11]See Margo (2000).

(or they should differ predictably, as does age). However, the linking procedure allows some variability in name and age, as these variables may be reported with error. For example, names are often misspelled or abbreviated; people may also use nicknames or diminutives in one year but not another. The procedure involved searching within a seven-year age window for possible matches based on sex, race and birthplace, then assessing the quality of the match based on age and name similarity.[12] This procedure yielded 20,782 links between 1870 and 1880. Of course, the links are only probable links which adds a dimension of measurement error, but an attempt is made to minimize this error.

The 1850-60 sample is from Joseph Ferrie. It is constructed by matching individuals in the 1850 one percent IPUMS sample with individuals listed in the index to the 1860 census. Once individuals are located in both censuses, original manuscripts from the 1860 census were obtained. Names are matched using the NYIIS algorithm,[13] and possible matches are discarded if they violate certain criteria. For example, any individual with more than ten potential matches is discarded. This procedure generated 4,938 matches.

These individual-level samples are used to identify migrants and to determine demographic characteristics (age, nativity, literacy), geographic characteristics (urban and farm status, etc), and occupation. Both samples contain an occupation variable that uses 1950 occupational codes. Because the census did not collect information on income before 1940 or information on wealth after 1870, occupation is the only available measure of success on the labor market. I use occupational wages from 1900 to measure upgrades and downgrades. These are compiled by Preston and Haines (1991), based on the 1901 cost of living survey. Following Olivetti and Paserman

[12]Software used was Freely Extensible Biomedical Record Linkage (FEBRL) created by Peter Christen and Tim Churches. Jaro-Winkler distance algorithm was used to assess the similarity of name variables. See Ruggles et al (2010).

[13]See Atack and Bateman (1992) for a description of this procedure.

(2012), I assign a 1900 income value to each 1950 occupational code. The income of farmers is imputed from the 1900 Census of Agriculture, using a method that follows Abramitzky, Boustan and Eriksson (2011) and is used in Olivetti and Paserman (2011). I use alternative occupational rankings, most of which are based on the 1950 wage distribution, as a robustness check.[14]

There are several challenges associated with measuring occupational mobility in this way. First, the wage distribution from 1850-1880 was not identical to that in 1900, so some job changes that were not actually upgrades will be coded as such. This procedure will be especially sensitive to the placement of farmers in the distribution, as many people moved to and from the farm sector during the period in question. Moreover, it fails to capture the difference in status between farm owners and tenants, as these two types are given the same 1950 occupational code. For the 1850-60 sample, I use information about real estate wealth to distinguish farm owners from tenants, but I cannot do this for the 1870-80 sample because this information was not collected in 1880. This procedure is also subject to error in the original coding of occupations. For example, suppose a farm laborer does not change occupations between 1850 and 1860, but he is coded as a farm laborer in 1850 and a common laborer in 1860. He will be treated as an upgrader even though his occupation has not changed. I attempt to limit this error by re-assigning "laborers" who live on farms the status of farm laborers.

The county-level wage data is from Robert Margo. These data come from the 1850, 1860 and 1870 Censuses of Social Statistics. Data on average wages paid to common laborers, farm laborers, carpenters and domestics, as well as the cost of board, were collected for each minor civil division in these census years. Day wages to laborers are reported with and without board; farm laborers' wages are reported by the month, while domestics' wages and board are reported by the week. These

[14] These variables are provided by the Census, and are described further below.

data are aggregated to the county level. They are available for 1265 counties in 20 states in 1870, and 1022 counties in 20 states in 1850.[15]

I use day wages to laborers (excluding board) as the primary measure of unskilled wages, and I use carpenters' wages as the sole measure of skilled wages. For the 1850-60 sample, the wage in 1850 is taken to be the migrant's pre-move expectation of the wage in the county to which he migrates. For the 1870-80 sample, I use the 1870 wage. It is important to use a measure of wages that pre-dates migration because migration should have an independent effect on wages. To approximate migration costs, I use two variables: distance migrated and a measure of railroad access. Railroad access is measured as the fraction of a given county located within fifteen miles of a railroad.[16] County population is taken from Haines and ICPSR (2010).

The sample is restricted to white males age 15-60 who are unskilled blue collar workers or farm laborers in the first year I observe them. I include only those who reported an occupation and resided in a county with available wage data in both periods. There are 2,036 such people. The first two columns of table 1 compare all unskilled workers to unskilled workers for whom I have wage data. The two samples are largely similar; however, the restriction that wage data be available in both years seems to bias the sample against interstate migrants. While 22 percent of the full sample moved states over the course of a decade, only 10 percent of my sample with available wage data did so. To address this issue, I use a second sample of migrants alone, and I only impose the restriction that they have available wage data in the second year I observe them. Columns 3 and 4 of table 1 show summary statistics for

[15]Wage data is available for the following states: Massachusetts, New York, Pennsylvania, Illinois (1870), Indiana, Michigan, Iowa, Nebraska, Virginia, Arkansas, Florida, Georgia, Louisiana, Mississippi (1850), North Carolina (1850), South Carolina (1850), Texas, Kentucky, Tennessee, Washington, and DC. I have added Kansas and South Carolina in 1870 to the sample using manuscripts published by ancestry.com

[16]Data provided by Robert Margo. See Atack, Bateman, Haines, and Margo (2010) for further information.

this group and a full national sample of migrants. These compare more favorably.

Table 2 contains summary statistics for wages in 1850 and 1870. As other literature has found, nominal laborers' wages tended to be higher in the Northeast than in the Midwest or the South; however, real laborer's wages were higher in the Midwest than in the Northeast. Interestingly, both real and nominal carpenters' wages were highest in the South. The skill premium was highest in the South in both years, and it grew uniformly across the U.S. between 1850 and 1870. Table 3 compares rural and urban wages across the United States and within regions. In general, counties with a nonzero urban population had higher nominal wages than entirely rural counties. However, real wages tended to be higher in rural counties. Skill premiums were uniformly higher in rural counties, especially in 1850.

2.4 A Motivating Result

An important hypothesis of this paper is that internal migrants in the United States during the nineteenth century were motivated by the prospect of upward job mobility. Before examining this idea more closely, I evaluate whether, on average, migration and occupational mobility were associated. Table 4 contains results from regressions of my measure of occupational upgrading on an indicator equal to one if the individual is a migrant, using a full sample of men and a sample of men who are unskilled workers in the first year I observe them. While the coefficient on migrant status is positive and significant when no controls are included, it becomes insignificant for all workers when demographic controls are added, and insignificant for unskilled workers alone when geographic controls (including region fixed effects) are added. Notice also that the R squared in the regression is never more than 0.12, suggesting that there is substantial unexplained variation in occupational upgrading. In the rest of the paper, I explore the possibility that this heterogeneity reflects unobserved propensity in occupational upgrading, and that selection and sorting on the basis of this skill can explain some of the migration patterns we observe among unskilled workers.

2.5 Empirical Approach

If internal migrants selected and sorted themselves on the basis of upgrading potential, there should be a relationship between "skill" (occupational mobility) and wages that is similar to what we observe in the modern migration literature. Namely, those most likely to upgrade occupations should be selected out of counties with small skill premiums sorted into counties with large skill premiums. Empirically, we should see a migration premium (in terms of occupational income) that is increasing in the destination skill premium and decreasing in the home skill premium. And, when we compare migrants, the incidence of upgrading should be higher in places with larger skill premiums.

To evaluate this, I estimate the following equation:

$$\Delta \log Y_{occ} = \beta_0 + \beta_1 MIG + \beta_2 \log W_{home} + \beta_3 MIG \times \log W_{home} + \beta_4 MIG \times \log W_{dest} + \gamma X + u \tag{2.1}$$

where Y_{occ} is occupational income, W is the county-level skill premium, and X is a matrix of individual and county-level variables, including region fixed effects. To understand the expected signs of these coefficients, consider three counties, A B and C, with $W_C > W_B > W_A$. Selection implies that a mover from A to B is more likely to upgrade than a stayer in A, and a stayer in B is more likely to upgrade that a stayer in A. Selection also implies that a mover from A to C is more likely to upgrade that a mover from B to C. Sorting implies that a mover from A to C is more likely to upgrade than a mover from A to B. The following table summarizes the differences in expected occupational upgrades for these groups:

Group 1	Group 2	$E[\Delta \log Y_{occ}^1] - E[\Delta \log Y_{occ}^2]$
Stayer B	Stayer A	$\beta_2(\log W_B - \log W_A)$
Mover A to B	Stayer A	$\beta_1 + \beta_3 \log W_A + \beta_4 \log W_B$
Mover A to C	Mover B to C	$-(\beta_2 + \beta_3)(W_B - W_A)$
Mover A to C	Mover A to B	$\beta_4(W_C - W_B)$

The relationships described above imply the following about the signs of these coefficients. The fourth row implies $\beta_4 > 0$, and the first row implies $\beta_2 > 0$. The third row implies that $\beta_3 + \beta_2 < 0$, which means $\beta_3 < 0$. Then, as predicted, the second row implies a migration premium that is decreasing in the home county skill premium and increasing in the destination county skill premium.

Restricting the sample to individuals with wage data in their homes and destinations biases the sample against inter-state migrants. To address this, I use a sample of migrants alone, and I only require that wage data be available in their destination counties. Grogger and Hanson (2008) argue that, among migrants, sorting should be independent of wage profiles at home. With this in mind, I estimate

$$\Delta \log Y_{occ} = \alpha_0 + \alpha_1 \log W_{dest} + \gamma X + u \tag{2.2}$$

Here, I expect to find $\alpha_1 > 0$.

One concern is that, while migrant sorting implies $\alpha_1 > 0$, the reverse is not true. Because we do not observe upgrading potential, we cannot be sure that the observed relationship between wages and upgrading is due to sorting and not solely to geographic differences in opportunity reflected in wage profiles. Moreover, this relationship could be generated by a different process. For instance, say we can write the actual change in occupational income as the expectation plus an error:

$$\Delta \log Y = E\Big[\Delta \log Y\Big] + e$$

Suppose workers select counties completely at random. If large skill premiums induce effort toward upgrading, we may observe more upgrading among migrants in counties with large skill premiums even though counties are chosen randomly. Put another way, the correlation between $\log W$ and $\Delta \log Y$ might reflect a relationship between $\log W$ and e, and not between $\log W$ and $E[\Delta \log Y]$. If this is the case, it would be difficult to argue that migrants chose destinations with occupational mobility in mind.

I will address this by using the gap between an older relative's occupational income and the migrant's occupational income in the first year I observe him as a proxy for the migrant's expected occupational upgrade. The idea is that a person's expected career trajectory is related to that of the other members of his family. So, the occupation of his father (or grandfather, or older brother) is a reasonable measure of what he expects his occupation to be in the future. This reflects expectations alone because it will not be induced by destination county wages. If individuals with large gaps between their relative's and their own occupations tend to choose counties with high skill premiums, this hints at the presence of sorting. To evaluate this, I regress the skill premium in a migrant's destination county (measured before the move) on the gap between his older male relative's log occupational income and his own. I focus on migrants because I do not want the results to be confounded by children inheriting land or businesses from their fathers; for example, a young man who works on his father's farm might report an upgrade from "farm laborer" to "farmer" because he has inherited his father's land, not because he is a particularly skilled agricultural worker. The problem with this approach is that, because so many of my observations are household heads in both decades, it reduces my sample size to 248 observations. So, these results will be presented as "suggestive evidence."

2.6 Results

2.6.1 Basic Results

I estimate equation (1) by OLS using a sample of workers who were unskilled blue collar workers or farm laborers in 1850 or 1870. Table 5 shows the results from these regressions, adding a series of demographic, county-level, and state controls. If migrants who were able to upgrade occupations were selected out of counties with low skill premiums and into counties with high skill premiums, we should see home county skill premiums exerting a positive effect on upgrading for stayers and a negative effect on upgrading for migrants; and, we should see destination county skill premiums positively affecting upgrading for migrants. By and large, this appears to be the case. However, notice that skill premiums only have a positive effect on upgrading for stayers when geographic and state controls are excluded.

These results can be summarized in the same way as the theoretical coefficients above, taking hypothetical counties A, B, and C with $W_C > W_B > W_A$. Using column (4) of table 5, in which geographic controls are included in the regression, the coefficients imply the following differences in expected occupational upgrade by home and destination county:

Group 1	Group 2	$E[\Delta \log Y_{occ}^1] - E[\Delta \log Y_{occ}^2]$
Stayer B	Stayer A	$0.0579(\log W_B - \log W_A)$
Mover A to B	Stayer A	$0.0031 - 0.1960^{(*)} \log W_A + 0.1823^{(**)} \log W_B$
Mover A to C	Mover B to C	$0.1381(W_B - W_A)$
Mover A to C	Mover A to B	$0.1823^{(**)}(W_C - W_B)$

These predicted differences all have the anticipated sign. However, notice that sorting appears more significant than selection. In the first row, skill premiums do not significantly affect differences in upgrading among stayers. And, the negative coefficient on log W_A in the second row of the table is only significant at the 10 percent level.

As mentioned earlier, restricting the sample to individuals with both home and destination wage data biases the sample against interstate migrants. In table 6, I include individuals with wage data in their destination counties alone, and I restrict the analysis to migrants. Notice that the strongly positive and significant coefficient on the destination skill premium is robust to including all covariates. So, the correlation between the skill premium and occupational upgrading does not seem to be driven by differences in the composition of migrants who choose different counties, or by county or state-level factors that may have jointly affected wages and job opportunities. For example, because so many upgraders are farm laborers who become farmers, one might worry that the observed relationship is driven by some correlation between land values or availability and wages. However, the significant coefficient remains even when controlling for these things.

These results mask large differences across time. Table 7 contains results estimated separately by year. The results are more robust for the 1870-80 sample than they are for the pooled sample, and the results for the 1850-60 sample are not at all robust. While differences between these two decades are not the focus of this paper, one story that may explain this discrepancy is increasing national labor market integration over the course of this period. Rational responses to wage differentials are symptomatic of a labor market that operates well across space. There is evidence that such a labor market did not emerge in the United States until the 1870s.[17] Differences in the linking procedure might also explain some of these differences, although it is not clear how.

One major concern is that farm laborers who become farmers comprise an overwhelming majority of the upgraders in my sample. This raises several concerns: one is that the imputed farmer's income that I use is intended for owner-occupier farmers, which is not an appropriate level of income to assign to tenant farmers. While I can infer tenancy in the 1850-60 sample by the absence of real estate wealth, I am

[17]Rosenbloom 1996.

unable to do this for the 1870-80 sample because this information was not collected in 1880.[18] And, this coarse distinction between tenant and owner classifies all "tenants" the same way, which neglects variation in income caused by different tenancy arrangements. Sharecroppers, for example, were worse off than share tenants, who were worse off than cash tenants.[19] Moreover, one worries that moves into farming are not directly comparable to moves into other occupations. Becoming a farmer may have been considered a good in and of itself, in part because it was associated with land acquisition and not just labor income. Table 8 excludes individuals who were farmers in either period. The results are robust to this exclusion.

I test the sensitivity of my results to my particular measure of the skill premium. The Census of Social Statistics provides alternatives to day laborer's wages as a measure of unskilled wages. The data also contain information on common laborer's wages including board and monthly farm wages (which included board). Table 9 repeats the analysis using these other two wage measures, which broadly confirms the baseline results.

Another concern is the use of the 1900 occupational wage distribution to measure occupational upgrades. This imposes two characteristics on job changes: whether or not it is an upgrade (ordinal) and the size of the upgrade or downgrade (cardinal). Both of these restrictions may not be valid for this sample. I use coarse indicators of upgrading and downgrading to verify that the results hold when the size of the occupational income gain is not imposed by the 1900 occupational wage distribution. Table 10 contains the results from regressions that use an indicator for upgrading instead of the actual change in log occupational income. The indicator used in the

[18]I assign farm tenants in 1850 and 1860 the status of farm laborers. Atack (1988) argues that, "to the extent that the rental market functioned well, landlords were able to capture much of the surplus over and above labor returns that were generated by the tenant. As a result, tenants probably did little better than farm laborers and may have even been worse off to the extent that they bore increased income risk." (p. 23-24).

[19]See Atack (1988, 1989), Atack and Bateman, Alston and Ferrie (2005).

first four columns is equal to one if the person upgrades; the indicator used in the last four columns is equal to one if the person moves up an occupational class.[20] The standard errors become larger in the last four columns of the table, but the estimate remains significant in all but one specification.

The results may also be sensitive to the ranking of occupations. To address this possibility, I define $\Delta \log Y$ using different occupational rankings available in the IPUMS data. These are the median 1950 occupational wage, the Duncan socioeconomic index, the Siegel occupational prestige score, and the Nam-Powers-Boyd occupational status score. The Duncan index is a composite index based on average occupational earnings and educational attainment.[21] The relative weights given to these two variables are based on a regression of occupational prestige rankings from the 1947 National Opinion Research Center survey on earnings and education for a small number of occupations. The NPB score is also a composite of earnings and education, but these two measures are given equal weight.[22] The Siegel prestige score is based on a series of National Opinion Research Center surveys from the 1960s; it uses no information on average earnings or education.[23]

Figure 1 plots the distribution of each occupational ranking using the one percent IPUMS sample from 1870. The baseline measure and the Duncan index rank farmers above both farm and common laborers but below carpenters. The Nam-Powers-Boyd index and the 1950 occupational wage both rank farmers below common laborers. The Siegel prestige score ranks farmers above carpenters and farm laborers above common laborers. Table 11 includes results that assign changes in occupational income according to these different rankings. This supports the main finding for every ranking except the Siegel prestige score, where the coefficient is still positive but not

[20]i.e. If he becomes a skilled craftsman, a farmer, or a white collar worker.
[21]See Duncan (1961) for details
[22]See Nam and Boyd (2004).
[23]See Siegel (1971).

robust to including all controls.

In general, it seems that unskilled blue collar workers who were able to ascend the occupational ladder chose counties in which carpenter's wages were high relative to laborer's wages. This results is robust to including various controls and altering the definition of the skill premium and occupational upgrading.

2.6.2 Suggestive Evidence for Sorting

In the preceding analysis, I demonstrated the robustness of the positive correlation between county skill premiums and occupational upgrading among migrants. This relationship remains intact when I control for observable variables that might jointly affect average wages and the opportunity for occupational advancement; the results are also robust to changing the definition of upgrading and skill premiums. Still, this does not necessarily mean that occupational mobility was something migrants explicitly considered when deciding where to move.

For example, a positive estimate of α_1 could be generated by the following. Suppose migrants choose counties at random. Then, if a migrant lands in a county with high skilled relative to unskilled wages, he may expend greater effort toward upgrading, thus generating the relationship observed in the data.

To address this problem, I take a subsample of migrants who were not household heads in the first period (but were related to the household head), and I use the difference between the head's log occupational income and the migrant's as a proxy for the migrant's expected occupational income growth. I regress the destination skill premium on this proxy, the idea being that migrants who are more likely to upgrade should be drawn to counties with high skill premiums. Unfortunately, this procedure completely cuts out the 1850-60 sample, as virtually none of these resided with an older working relative in 1850.[24] And, it reduces the sample of migrants with wage

[24]This is likely due to the way this sample is constructed. Because observations are located in

data to 248. Still, I will present these results as suggestive that induced effort toward upgrading cannot explain all of my results.

Table 12 contains results a regression of the change in log occupational income on the difference between the household head's log occupational income and the migrant's. The first four columns use only the 248 migrants with wage data, and the last four columns use everyone. There is a very strong and positive relationship between the change in occupational income and the proxy variable; in fact, these covariates explain approximately one third of the variation in occupational upgrading.

Table 13 contains results from a regression of destination skill premiums on my proxy variable. While the coefficient is not significant in all specifications, the results suggest that at least some selection and sorting can explain the relationship between upgrading and skill premiums. While the results I have presented in this section are only suggestive, there does not seem to be overwhelming evidence that the basic results are driven by reverse causality.

2.7 Conclusion

This paper explores the relationship between geographic and occupational mobility in nineteenth century America. One might expect these to be closely connected; however, a straightforward analysis of the data suggests no relationship between migration and occupational upgrading. I offer evidence that this masks heterogeneity among migrants: those who were better able to upgrade occupations seem to have sorted themselves into counties with high skill premiums. I also present suggestive evidence to argue that reverse causality cannot explain this result.

This selection and sorting of migrants might shed light on puzzles identified in previous work. Lebergott (1964) finds a surprisingly weak relationship between pop-

the 1860 census using the index to the census manuscript, all men in this sample had to have been household heads in 1860.

ulation growth and average wages at the state level, arguing that migrants failed to take full advantage of potential wage gains. Others have found frontier migration puzzling because of the "Easterlin paradox" that Midwestern per capita incomes were lower than those in the Northeast,[25] although this can be explained in part by regional price differences.[26] In fact, these observations seem to mask a tendency for some workers to seek upward occupational mobility instead of higher average wages; workers who migrated from counties with low skill premiums to counties with high skill premiums tended also to end up in counties with lower real laborers' wages. Conversely, those who opted for counties with lower skill premiums tended to end up in counties with higher real laborers' wages.

2.8 References

Abramitzky, Ram, Leah Platt Boustan, and Katherine Eriksson (2010). "Europes Tired, Poor, Huddles Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration." NBER Working Paper no. 15684.

Alston, Lee J. and Joseph P. Ferrie (2005). "Time on the Ladder: Career Mobility in Agriculture, 1890-1938." *Journal of Economic History.* 65(4): 1058-1081.

Alston, Lee J. and Kyle D. Kauffman (1997). "Agricultural Chutes and Ladders: New Estimates of Sharecroppers and "True Tenants" in the South, 1900-1920." *Journal of Economic History.* 57(2): 464-475.

Atack, Jeremy (1989). "The Agricultural Ladder Revisited: A New Look at an Old Problem with some Data from 1860." *Agricultural History.* 63(1): 1-25.

—-. (1988). "Tenants and Yeomen in the Nineteenth Century." *Agricultural History.* 62(3): 6-32.

Atack, Jeremy and Fred Bateman (1992). "Matchmaker Matchmaker Make Me a Match." *Historical Methods.* 25(2): 1-18.

[25]Easterlin estimates that per capita incomes in the Midwest were roughly half what they were in the northeast in 1840 (Margo 1999 p. 130).

[26]See Coelho and Shephard (1976); Margo (1999).

Atack, Jeremy and Fred Bateman (1987). *To Their Own Soil: American Agriculture in the Antebellum North.* Iowa State University Press.

Atack, Jeremy, Fred Bateman, and Robert A. Margo (2004). "Skill Intensity and Rising Wage Dispersion in Nineteenth-Century American Manufacturing." *Journal of Economic History* 64: 172-192.

Atack, Jeremy, Fred Bateman, Michael Haines, and Robert A. Margo (2010). "Did Railroads Induce or Follow Economic Growth? Urbanization and Population Growth in the American Midwest, 1850-60." Forthcoming in *Social Science History.*

Borjas, George J., Stephen G. Bronars, and Stephen J. Trejo (1992). "Self-Selection and Internal Migration in the United States." *Journal of Urban Economics.* 32(2): 159-85.

Coelho, Philip R. and James F. Shepherd (1976). "Regional Differences in Real Wages: The United States, 1851-1880." *Explorations in Economic History.* 13: 203-230.

Duncan, O.D. (1961). "A Socioeconomic Index for All Occupations." in A. Reiss et al., *Occupations and Social Status.* Free Press.

Ferrie, Joseph P. (2005). "The End of American Exceptionalism: Mobility in the U.S. since 1850." NBER Working Paper no. 11324.

—-. (1997). "Migration to the Frontier in Mid-Nineteenth Century America: A Re-Examination of Turner's 'Safety Valve.' " Department of Economics, Northwestern University.

—-. (1996a). "A New Sample of Americans Linked from the 1850 Public Use Micro Sample of the Federal Census of Population to the 1860 Federal Census Manuscript Schedules." *Historical Methods.* 29: 141- 156.

—-. (1996b). "The Entry into the U.S. Labor Market of Antebellum European Immigrants, 1840-60." *Explorations in Economic History.* 34(3): 295-330.

Forestall, Richard L. (1996). *Population of States and Counties of the United States: 1790-1990* Washington, DC: US Department of Commerce, Bureau of the Census.

Galenson, David W. and Clayne L. Pope (1989). "Economic and Geographic Mobility on the Farming Frontier: Evidence from Appanoose County, Iowa, 1850-1870."

Journal of Economic History. 49: 635-656.

Gallaway, Lowell and Richard K. Vedder (1971). "Mobility of Native Americans." *Journal of Economic History* 31(3): 613-649.

Grogger, Jeffrey and Gordon H. Hansen (2008). "Income Maximization and the Selection and Sorting of International Migrants." NBER Working Paper no. 13821.

Goodrich, Carter and Sol Davison (1935). "The Wage-Earner in the Westward Movement I." *Political Science Quarterly.* 50(2): 161-185.

Goodrich, Carter and Sol Davison. (1936). "The Wage-Earner in the Westward Movement II." *Political Science Quarterly.* 51(1): 61-116.

Haines, Michael R., and Inter-university Consortium for Political and Social Research (2010). Historical, Demographic, Economic, and Social Data: The United States, 1790-2002 [Computer file]. ICPSR02896-v3. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2010-05-21. doi:10.3886/ICPSR02896

Hall, Patricia K. and Steven Ruggles (2004). " 'Restless in the Midst of their Prosperity': New Evidence on the Internal Migration of Americans, 1850-2000." *Journal of American History.* 91(3): 829-846.

Kim, Sukkoo (1998). "Economic Integration and Convergence: U.S. Regions, 1840-1987." *Journal of Economic History.* 58(3): 659-683.

Lebergott, Stanley (1964). *Manpower and Economic Growth: The American Record since 1800.* New York: McGraw-Hill.

Long, Jason, and Joseph Ferrie (2005). "A Tale of Two Labor Markets: Intergenerational Occupational Mobility in Britain and the U.S. Since 1850." NBER Working Paper no. 11253.

Nam, C.B. and M. Boyd (2004). "Occupational Status in 2000: Over a Century of Census-based Measurement." *Population Research and Policy Review.* 23: 327-358.

Margo, Robert A. (2000). *Wages and Labor Markets in the United States, 1820-1860.* Chicago: University of Chicago Press.

Margo, Robert A. (1999). "Regional Wage Gaps and the Settlement of the Midwest." *Explorations in Economic History.* 36: 128-143.

Olivetti, Claudia and Daniele Paserman (2012). "In the Name of the Father: Marriage and Intergenerational Mobility in the United States, 1850-1930." Boston University.

Preston, Samuel H. and Michael Haines (1991). *Fatal Years: Child Mortality in Late Nineteenth Century America.* Princeton: Princeton University Press.

Rosenbloom, Joshua (2002). *Looking for Work, Searching for Workers: American Labor Market during Industrialization.* Cambridge: Cambridge University Press.

—. (1996). "Was There a National Labor Market at the End of the Nineteenth Century? *Journal of Economic History.* 56(3): 626-656.

Rosenbloom, Joshua and William Sundstrom (2003). "The Decline and Rise of Interstate Migration in the United States: Evidence From the IPUMS, 1850-1990." NBER Working Paper no. 9857.

Rothenberg, Winifred B. (1988). "The Emergence of Farm Labor Markets and the Transformation of the Rural Economy: Massachussetts 1750-1855." *Journal of Economic History.* 48(3): 537-566.

Ruggles, Steven J. et. al. (2010). *Integrated Public Use Microdata Series: Version 5.0.* Minneapolis: University of Minnesota.

Schaefer, Donald F. (1985). "A Statistical Profile of Frontier and New South Migration: 1850-1860," *Agricultural History* 59(4): 563-578.

Siegel, P.M. (1971). "Prestige in the American Occupational Structure." Ph.D. dissertation, University of Chicago.

Steckel, Richard H. (1983). "The Economic Foundations of East-West Migration during the 19th Century." *Explorations in Economic History* 20: 14-36.

Stewart, James I. (2003). "Migration to the Agricultural Frontier and Economic Mobility, 1860-1880." Paper presented at 42nd Cliometrics Conference, North Carolina State University.

2.9 Tables and Figures

Table 2.1: Summary Statistics, 1850-1860 and 1870-1880

Sample:	All Unskilled	Unskilled workers with home and destination wage data	Unskilled Migrants	Unskilled migrants with destination wage data
Age	27.273	27.413	25.695	25.664
Literate	0.912	0.9	0.911	0.915
Immigrant	0.129	0.118	0.208	0.215
Unskilled laborer 2	0.465	0.469	0.455	0.472
Farmer 2	0.369	0.375	0.344	0.315
Occupational upgrade	0.638	0.631	0.662	0.651
Occupational downgrade	0.115	0.108	0.124	0.117
Moved counties	0.394	0.319	1	1
Moved states	0.224	0.104	0.568	0.508
Number of county-years, t	1546	824	933	594
Number of county-years, t+10	1680	893	982	529
Obs per county-year, t	2.761	2.471	1.805	1.54
Obs per county-year, t+10	2.541	2.28	1.715	1.73
Northeast, t	0.475	0.497	0.444	0.466
Midwest, t	0.273	0.216	0.318	0.292
South, t	0.235	0.287	0.211	0.223
West, t	0.017	0	0.027	0.02
Northeast, t+10	0.447	0.477	0.371	0.462
Midwest, t+10	0.301	0.242	0.39	0.318
South, t+10	0.228	0.28	0.193	0.216
West, t+10	0.025	0.001	0.046	0.003
t=1870	0.599	0.558	0.479	0.438
N	4269	2036	1684	915

Notes: Data from IPUMS (2010), Ferrie (1996), Margo (2000). Sample consists of men between the ages of 15 and 60 in year t (1850 or 1870) who report an occupation in both years and who work in unskilled jobs in year t. Unskilled workers are defined as farm laborers or unskilled blue collar workers. In the 1850-60 sample, farmers who report zero real estate wealth are defined as farm laborers. Occupational upgrades and downgrades are coded using the 1900 occupational wage distribution with an imputed wage for farmers (Preston and Haines 1991; Abramitzky Boustan and Eriksson 2010).

Table 2.2: Summary Statistics for Wages, 1850 and 1870

Wage Concept	**Region**	**1850**					**1870**				
		Mean	SD	Min	Max	N	Mean	SD	Min	Max	N
Nominal Carpeneter	USA	1.513	0.465	0.500	5.000	1004	2.684	0.605	1.000	6.000	1258
	Northeast	1.331	0.172	0.964	1.750	136	2.631	0.444	1.500	4.000	110
	Midwest	1.412	0.213	1.000	2.500	157	2.850	0.349	2.000	4.500	426
	South	1.570	0.527	0.500	5.000	711	2.550	0.667	1.000	5.000	701
Nominal Laborer	USA	0.766	0.227	0.250	2.500	993	1.347	0.451	0.500	3.000	1256
	Northeast	0.880	0.125	0.627	1.250	136	1.754	0.302	1.000	2.750	110
	Midwest	0.825	0.164	0.500	2.000	157	1.654	0.273	0.500	2.860	426
	South	0.731	0.244	0.250	2.500	700	1.069	0.344	0.500	3.000	700
Real Carpenter	USA	6.787	2.363	1.750	28.000	981	6.243	2.003	2.625	20.000	1254
	Northeast	5.358	0.920	3.500	8.167	136	4.709	0.920	2.625	8.167	109
	Midwest	6.673	1.581	3.908	17.500	157	5.525	1.005	2.800	9.625	426
	South	7.096	2.594	1.750	28.000	688	6.935	2.295	2.625	20.000	698
Real Laborer	USA	3.394	1.003	0.875	14.000	979	3.009	0.849	0.875	10.818	1253
	Northeast	3.528	0.554	2.032	5.273	136	3.133	0.561	2.000	5.000	109
	Midwest	3.892	1.136	2.170	14.000	157	3.196	0.633	0.875	6.125	426
	South	3.253	1.000	0.875	11.200	686	2.868	0.971	0.875	10.818	698
Skilll Premium	USA	2.058	0.665	0.500	6.667	986	2.160	0.716	0.588	8.333	1254
	Northeast	1.524	0.161	1.200	1.993	136	1.533	0.313	0.708	2.750	110
	Midwest	1.746	0.283	1.000	2.667	157	1.765	0.405	1.000	7.700	426
	South	2.233	0.705	0.500	6.667	693	2.511	0.720	0.588	8.333	698

Table 2.3: Urban versus Rural Wages by Region, 1850 and 1870

	1850						
Wage	Region	Mean (urban)	Mean (rural)	Urban - rural	P value (urban=rural)	N (urban)	N (rural)
Carpenter (nominal)	USA	1.472	1.518	-0.046	0.345	100	904
	northeast	1.314	1.340	-0.025	0.414	47	89
	midwest	1.395	1.413	-0.018	0.763	14	143
	south	1.688	1.564	0.125	0.151	39	672
Laborer (nominal)	USA	0.894	0.752	0.142	**0.000**	100	893
	northeast	0.895	0.872	0.023	0.311	47	89
	midwest	0.880	0.819	0.061	0.184	14	143
	south	0.897	0.721	0.176	**0.000**	39	661
Carpetner (real)	USA	5.440	6.940	-1.500	**0.000**	100	881
	northeast	4.883	5.609	-0.727	**0.000**	47	89
	midwest	6.212	6.718	-0.506	0.254	14	143
	south	5.835	7.172	-1.337	**0.002**	39	649
Laborer (real)	USA	3.310	3.403	-0.093	0.379	100	879
	northeast	3.312	3.642	-0.330	**0.001**	47	89
	midwest	3.895	3.892	0.003	0.992	14	143
	south	3.098	3.262	-0.165	0.318	39	647
Skill premium	USA	1.679	2.100	-0.421	**0.000**	100	886
	northeast	1.477	1.548	-0.072	**0.013**	47	89
	midwest	1.602	1.760	-0.158	**0.046**	14	143
	south	1.951	2.250	-0.299	**0.010**	39	654
	1870						
Wage	Region	Mean (urban)	Mean (rural)	Urban - rural	P value (urban=rural)	N (urban)	N (rural)
Carpenter (nominal)	USA	2.749	2.668	0.081	**0.056**	251	1007
	northeast	2.650	2.587	0.063	0.501	77	33
	midwest	2.814	2.863	-0.050	0.195	113	313
	south	2.755	2.530	0.225	**0.012**	61	640
Laborer (nominal)	USA	1.582	1.288	0.294	**0.000**	251	1005
	northeast	1.781	1.691	0.090	0.155	77	33
	midwest	1.658	1.653	0.005	0.856	113	313
	south	1.192	1.058	0.134	**0.003**	61	639
Carpetner (real)	USA	5.407	6.452	-1.045	**0.000**	251	1003
	northeast	4.653	4.844	-0.191	0.326	77	32
	midwest	5.400	5.571	-0.171	0.120	113	313
	south	6.372	6.988	-0.616	**0.045**	61	637
Laborer (real)	USA	3.047	3.000	0.047	0.434	251	1002
	northeast	3.121	3.161	-0.041	0.732	77	32
	midwest	3.175	3.203	-0.028	0.685	113	313
	south	2.716	2.883	-0.167	0.199	61	637
Skill premium	USA	1.827	2.244	-0.417	**0.000**	251	1003
	northeast	1.523	1.557	-0.034	0.604	77	33
	midwest	1.724	1.780	-0.056	0.212	113	313
	south	2.400	2.521	-0.121	0.210	61	637

Table 2.4: Effect of Migration on Occupational Mobility, 1850-60 and 1870-80

	Dependent variable: change in log occupational income							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Group	*All*				*Unskilled Workers*			
Migrant	0.0400***	-0.0082	-0.0102	-0.0066	0.0303**	0.0251*	0.0167	0.0064
	(0.009)	(0.009)	(0.009)	(0.011)	(0.015)	(0.015)	(0.015)	(0.017)
Age		-0.0053***	-0.0052***	-0.0051***		-0.0056***	-0.0044***	-0.0043***
		(0.000)	(0.000)	(0.000)		(0.001)	(0.001)	(0.001)
Immigrant		-0.0559***	-0.0202*	-0.0266**		-0.1821***	-0.1212***	-0.1297***
		(0.012)	(0.012)	(0.012)		(0.020)	(0.021)	(0.021)
Literate		-0.0641***	-0.0431**	-0.0436**		0.0002	0.0575**	0.0586**
		(0.017)	(0.017)	(0.017)		(0.026)	(0.025)	(0.026)
Married		-0.1485***	-0.1587***	-0.1588***		-0.0968***	-0.0908***	-0.0902***
		(0.011)	(0.010)	(0.011)		(0.019)	(0.019)	(0.019)
Constant	0.1254***	0.4621***	0.5559***	0.5446***	0.3526***	0.5738***	0.6332***	0.7228***
	(0.009)	(0.022)	(0.045)	(0.077)	(0.014)	(0.034)	(0.087)	(0.128)
Home county controls			X	X			X	X
Destination county controls				X				X
Observations	11,197	10,924	10,856	10,756	4,112	4,064	4,035	3,989
Adjusted R-squared	0.00343	0.0975	0.111	0.113	0.00131	0.0728	0.112	0.113

Notes: Standard errors in parentheses. County-level population and agricultural variables taken from Haines and ICPSR (2010); railroad access variable from Atack, Bateman, Haines, and Margo (2010). Land availability is the fraction of agricultural land that is unimproved. Railroad access is the percent of a county within 15 miles of a railroad. See Table 1 for further details.

Table 2.5: Effect of Migration and Wages on Occupational Mobility, 1850-60 and 1870-80

Dependent variable: change in log occupational income

	(1)	(2)	(3)	(4)
Migrant	-0.0240	-0.0122	0.0163	0.0031
	(0.053)	(0.052)	(0.052)	(0.053)
Skill premium (home)	0.2994***	0.2452***	0.0762	0.0579
	(0.053)	(0.052)	(0.058)	(0.059)
Migrant X Skill premium (home)	-0.2899***	-0.2710***	-0.2288**	-0.1960*
	(0.097)	(0.097)	(0.098)	(0.100)
Migrant X Skill premium (dest)	0.3799***	0.3241***	0.2128***	0.1823**
	(0.076)	(0.077)	(0.080)	(0.084)
Age		-0.0042***	-0.0035***	-0.0035***
		(0.001)	(0.001)	(0.001)
Immigrant		-0.1763***	-0.1242***	-0.1243***
		(0.027)	(0.028)	(0.028)
Literate		0.0478	0.0675**	0.0692**
		(0.033)	(0.033)	(0.033)
Married		-0.0970***	-0.0881***	-0.0850***
		(0.025)	(0.025)	(0.025)
Population density (home)			0.0066*	0.0022
			(0.003)	(0.004)
Percent urban (home)			-0.2949***	-0.2350***
			(0.045)	(0.076)
Railroad access (home)			0.0212	-0.0056
			(0.033)	(0.050)
Land availability (home)			-0.0938	-0.1183
			(0.071)	(0.112)
Log farm value per acre (home)			-0.0475	-0.0006
			(0.080)	(0.125)
Midwest (home)			0.0925***	0.0325
			(0.028)	(0.060)
South (home)			0.1265***	0.0422
			(0.036)	(0.079)
West (home)			-0.1274**	0.2656
			(0.052)	(0.182)
Population density (dest)				0.0064**
				(0.003)
Percent urban (dest)				-0.0712
				(0.079)
Railroad access (dest)				0.0393
				(0.053)
Land availability (dest)				0.0365
				(0.109)
Log farm value per acre (dest)				-0.0646
				(0.130)
Northeast (dest)				0.3932**
				(0.181)
Midwest (dest)				0.4597***
				(0.175)
South (dest)				0.4884**
				(0.199)
West (dest)				0.0000
				(0.000)
Distance Migrated				0.0058
				(0.008)
1870-1880	-0.0076	-0.0021	0.0170	0.0168
	(0.025)	(0.023)	(0.023)	(0.024)
Constant	0.2066***	0.3682***	0.4870***	0.1066
	(0.032)	(0.056)	(0.128)	(0.223)
Observations	2,035	2,018	2,018	2,018
Adjusted R-squared	0.0347	0.0920	0.124	0.124

Note: Standard errors in parentheses. Skill premium is the log ratio of carpenter's to laborer's wage at the county level. Standard errors clustered at the destination county level. See Tables 1 and 2 for details.

Table 2.6: Effect of Destination Skill Premium on Occupational Mobility among Migrants, 1850-60 and 1870-80

Dependent variable: change in log occupational income

	(1)	(2)	(3)	(4)	(5)
Destination skill premium	0.3081***	0.2359***	0.2135***	0.1976**	0.2735***
	(0.061)	(0.064)	(0.074)	(0.083)	(0.094)
Home skill premium					-0.0563
					(0.096)
Age		-0.0047**	-0.0052**	-0.0054**	-0.0040
		(0.002)	(0.002)	(0.002)	(0.003)
Immigrant		-0.1999***	-0.1775***	-0.1767***	-0.2323***
		(0.035)	(0.040)	(0.039)	(0.044)
Literate		0.0630	0.0687	0.0621	0.0176
		(0.046)	(0.048)	(0.048)	(0.052)
Married		-0.0464	-0.0298	-0.0275	-0.0665
		(0.043)	(0.043)	(0.043)	(0.051)
Distance migrated			0.0074	0.0059	0.0104
			(0.005)	(0.007)	(0.008)
Population density (home)			-0.0028	-0.0021	-0.0015
			(0.004)	(0.005)	(0.005)
Population density (dest)			0.0049	0.0053	0.0053
			(0.003)	(0.003)	(0.005)
Percent urban (home)			-0.2288***	-0.2077***	-0.1738**
			(0.074)	(0.075)	(0.088)
Percent urban (dest)			0.0712	0.0785	0.0103
			(0.081)	(0.082)	(0.098)
Railroad access (home)			-0.0382	-0.0266	-0.0019
			(0.049)	(0.049)	(0.057)
Railroad access (dest)			0.0212	0.0261	0.0227
			(0.053)	(0.054)	(0.065)
Land availability (home)			-0.2200*	-0.2408**	-0.1800
			(0.115)	(0.119)	(0.137)
Land availability (dest)			0.0669	0.0553	0.0437
			(0.111)	(0.118)	(0.126)
Log farm value per acre (home)			0.0272	-0.0142	0.0155
			(0.116)	(0.122)	(0.152)
Log farm value per acre (dest)			-0.1245	-0.1408	-0.1014
			(0.126)	(0.130)	(0.154)
Midwest (home)				0.0689	0.0138
				(0.048)	(0.065)
South (home)				-0.0524	0.0265
				(0.060)	(0.080)
West (home)				0.1389	0.0000
				(0.205)	(0.000)
Northeast (dest)				-0.0647	0.4890**
				(0.079)	(0.201)
Midwest (dest)				-0.0629	0.5072***
				(0.065)	(0.186)
South (dest)				0.0000	0.4945**
				(0.000)	(0.212)
West (dest)				-0.4563***	0.0000
				(0.170)	(0.000)
1870-1880	0.0159	0.0089	0.0301	0.0124	-0.0110
	(0.031)	(0.031)	(0.037)	(0.038)	(0.049)
Constant	0.2136***	0.3780***	0.5907***	0.7333***	0.0806
	(0.035)	(0.079)	(0.215)	(0.226)	(0.348)
Observations	914	908	897	897	645
Adjusted R-Squared	0.0261	0.0813	0.0899	0.0914	0.0961

Note: Standard errors in parentheses. Skill premium is the log ratio of carpenter's to laborer's wage at the county level. Standard errors clustered at the destination county level. See Tables 1 and 2 for details.

Table 2.7: Effect of Destination Skill Premium on Occupational Mobility among Migrants, By Year

Dependent variable: change in log occupational income

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Year		*1850-1860*				*1870-1880*		
Destination skill premium	0.0930	0.0493	0.0315	0.1164	0.2945***	0.2795***	0.2955**	0.3654***
	(0.093)	(0.117)	(0.121)	(0.140)	(0.085)	(0.091)	(0.115)	(0.118)
Constant	0.4581***	0.4880	0.4561	0.0885	0.3390***	0.6353**	0.6992*	0.7895
	(0.097)	(0.341)	(0.354)	(0.485)	(0.115)	(0.299)	(0.402)	(0.494)
Demographic controls	X	X	X	X	X	X	X	X
Geographic controls		X	X	X		X	X	X
Region controls			X	X			X	X
Home county wage				X				X
Observations	513	506	506	376	395	391	391	269
Adjusted R-Squared	0.0749	0.0727	0.0688	0.0675	0.0989	0.102	0.117	0.143

Note: Standard errors in parentheses. Skill premium is the log ratio of carpenter's to laborer's wage at the county level. Standard errors clustered at the destination county level. See Tables 1 and 2 for details.

Table 2.8: Effect of Destination County Skill Premium on Occupational Mobility, 1850-60 and 1870-80: Excluding Farmers

Dependent variable: change in log occupational income

	(1)	(2)	(3)	(4)
Destination skill premium	0.1857**	0.2751***	0.2519**	0.2898**
	(0.089)	(0.104)	(0.117)	(0.136)
Constant	0.2530**	-0.1580	0.0528	-0.1763
	(0.100)	(0.285)	(0.305)	(0.390)
Demographic controls	X	X	X	X
Geographic controls		X	X	X
Region controls			X	X
Home county wage				X
Observations	624	616	616	443
Adjusted R-squared	0.0566	0.0829	0.0827	0.0734

Note: Standard errors in parentheses. Skill premium is the log ratio of carpenter's to laborer's wage at the county level. Sample excludes workers who become farmers in year t+10 (1860 or 1880) Standard errors clustered at the destination county level. See Tables 1 and 2 for details.

Table 2.9: Effect of Destination County Skill Premium on Occupational Mobility, 1850-60 and 1870-80: Different Wage Measures

	Dependent variable: change in log occupational income							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Wage concept	*Laborer including board*				*Farm laborer*			
Destination skill premium	0.1107***	0.1164**	0.1006*	0.1225**	0.1168***	0.1119**	0.0999*	0.1669***
	(0.037)	(0.049)	(0.051)	(0.060)	(0.042)	(0.050)	(0.053)	(0.062)
Constant	0.2427**	0.3571	0.0960	0.4873	0.1948	0.3601	0.5495**	0.4242
	(0.119)	(0.263)	(0.306)	(0.358)	(0.144)	(0.260)	(0.278)	(0.375)
Demographic controls	X	X	X	X	X	X	X	X
Geographic controls		X	X	X		X	X	X
Region controls			X	X			X	X
Home county wage				X				X
Observations	906	895	895	643	890	880	880	625
Adjusted R-squared	0.0791	0.0868	0.0894	0.0889	0.0807	0.0892	0.0915	0.0956

Note: Standard errors in parentheses. Skill premium is defined two ways: first, as the log ratio of "real" carpenter's wage (deflated by the cost of board) to laborer's wage including board; second, as the log ratio of the real carpenter's wage to the farm wage (which includes board). Standard errors clustered at the destination county level. See Tables 1 and 2 for details.

Table 2.10: Effect of Destination County Skill Premium on Occupational Mobility, 1850-60 and 1870-80: Coarse Indicators

	Dependent variable: Indicator for occupational upgrade							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Upgrade indicator:	=1 if upgrades				=1 if moves up occupational class			
Destination skill premium	0.1626**	0.1810**	0.1852**	0.2400***	0.2163***	0.1407*	0.1270	0.1842*
	(0.067)	(0.076)	(0.083)	(0.092)	(0.072)	(0.085)	(0.090)	(0.103)
Constant	0.6817***	0.4533**	0.6233**	-0.2060	0.3665***	0.3797	0.5505**	-0.3485
	(0.085)	(0.229)	(0.244)	(0.385)	(0.088)	(0.242)	(0.262)	(0.402)
Demographic controls	X	X	X	X	X	X	X	X
Geographic controls		X	X	X		X	X	X
Region controls			X	X			X	X
Home county wage				X				X
Observations	908	897	897	645	908	897	897	645
Adjusted R-squared	0.0649	0.0661	0.0755	0.0735	0.0485	0.0630	0.0690	0.0727

Note: Standard errors in parentheses. Skill premium is the log ratio of carpenter's to laborer's wage at the county level. Standard errors clustered at the destination county level. See Tables 1 and 2 for details.

Table 2.11: Effect of Destination County Skill Premium on Occupational Mobility, 1850-60 and 1870-80: Different Occupational Status Measures

Dependent variable: change in log occupational status measure

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Occupational Status Measure:	*1950 occupational wage*				*Nam-Powers-Boyd Index*			
Destination skill premium	0.1664**	0.2330***	0.2543***	0.3791***	0.4176***	0.5000***	0.5017***	0.7488***
	(0.075)	(0.082)	(0.094)	(0.107)	(0.149)	(0.167)	(0.190)	(0.218)
Constant	0.2967***	0.4391*	0.5898**	-0.2853	0.6589***	1.1939**	1.5202***	-0.2652
	(0.095)	(0.251)	(0.264)	(0.401)	(0.192)	(0.514)	(0.543)	(0.775)
Demographic controls	X	X	X	X	X	X	X	X
Geographic controls		X	X	X		X	X	X
Region controls			X	X			X	X
Home county wage				X				X
Observations	908	897	897	645	908	897	897	645
Adjusted R-squared	0.0434	0.0506	0.0490	0.0520	0.0634	0.0803	0.0799	0.0784
Occupational Status Measure:	*Siegel Prestige Score*				*Duncan Index*			
Destination skill premium	0.1496**	0.1051	0.0553	0.1027	0.2177**	0.2953**	0.2350*	0.3700***
	(0.063)	(0.071)	(0.076)	(0.088)	(0.109)	(0.116)	(0.126)	(0.143)
Constant	0.3767***	0.3989*	0.3570	0.1224	0.5235***	0.5296	1.0037***	-0.8864
	(0.072)	(0.205)	(0.272)	(0.325)	(0.136)	(0.357)	(0.388)	(0.567)
Demographic controls	X	X	X	X	X	X	X	X
Geographic controls		X	X	X		X	X	X
Region controls			X	X			X	X
Home county wage				X				X
Observations	907	896	896	645	908	897	897	645
Adjusted R-squared	0.0343	0.0571	0.0697	0.0590	0.0385	0.0583	0.0684	0.0614

Note: Standard errors in parentheses. Skill premium is the log ratio of carpenter's to laborer's wage at the county level. The 1950 occupational wage is the median wage paid to workers in a given occupation in 1950; the Nam-Powers-Boyd index and the Duncan index are composites of wages and educational attainment from 1950, weighted differently; the Siegel score is based on opinion surveys from the 1960s. Standard errors clustered at the destination county level. See Tables 1 and 2 for details.

Table 2.12: Relationship between Occupational Mobility and Older Male Relative's Occupation, 1870-1880.

	Dependent variable: change in log occupational income							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sample:	*Migrants with wage data*				*All Migrants*			
Log relative's/own occupational income	0.6384***	0.6358***	0.6414***	0.6410***	0.7212***	0.7193***	0.7007***	0.6985***
	(0.059)	(0.062)	(0.067)	(0.067)	(0.042)	(0.044)	(0.048)	(0.048)
Constant	0.1104***	0.0194	0.4225	0.5268	0.0969***	0.0502	0.1742	0.3169
	(0.034)	(0.147)	(0.354)	(0.413)	(0.025)	(0.103)	(0.231)	(0.310)
Demographic controls		X	X	X		X	X	X
Geographic controls			X	X			X	X
Region controls				X				X
Observations	250	250	248	248	480	480	440	440
Adjusted R-squared	0.315	0.320	0.308	0.324	0.378	0.384	0.378	0.380

Note: Standard errors in parentheses. Sample is restricted to men who live with an older, employed male relative in year 1870 (none meets this restriction in the 1850-60 sample). Sample expanded to include workers who begin in skilled jobs to increase sample size. See Tables 1 and 2 for details.

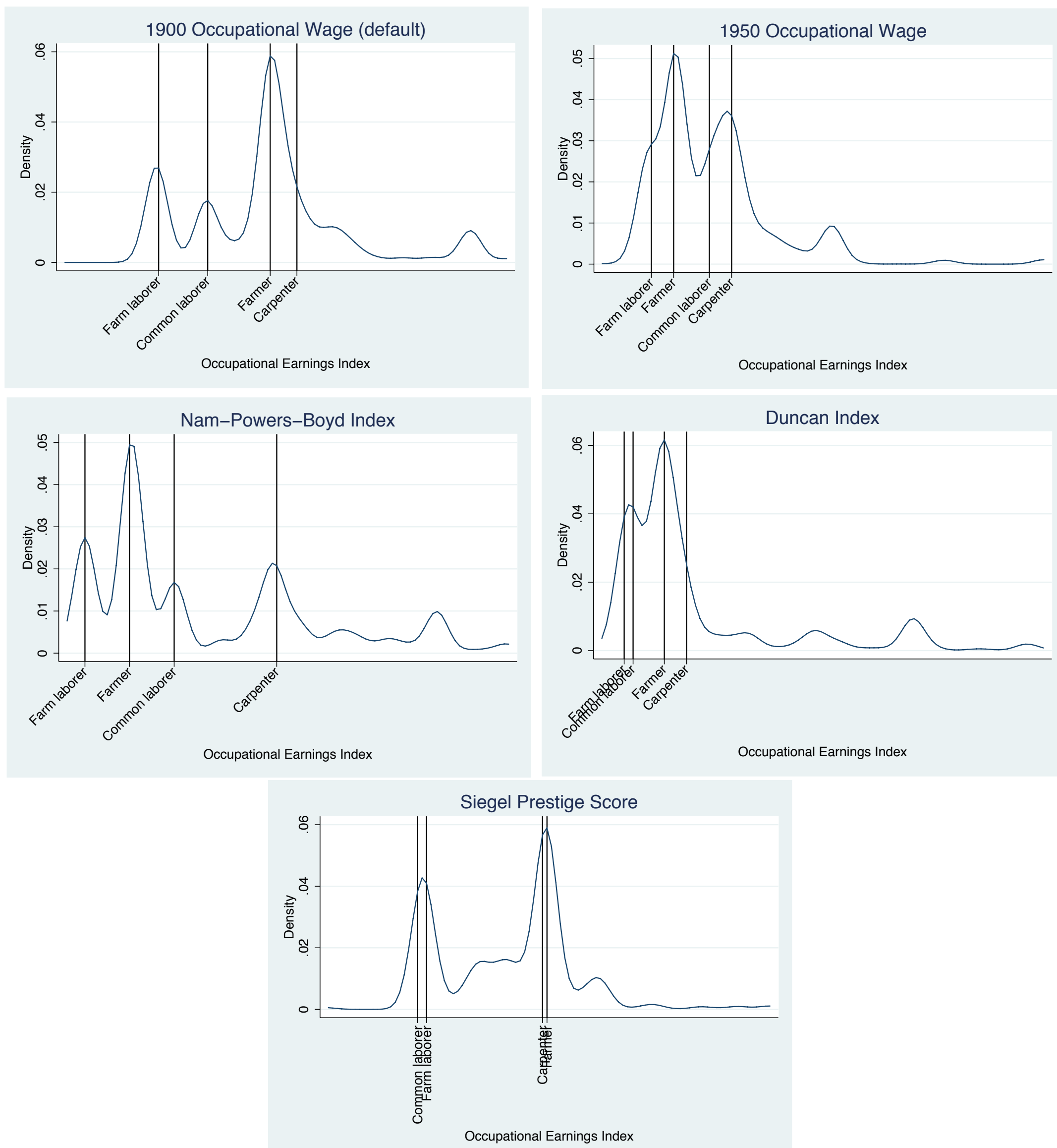
Table 2.13: Effect of Older Male Relative's Occupation on Choice of Destination County, 1870-1880

Dependent variable: destination county skill premium

	(1)	(2)	(3)	(4)	(5)
Log relative's/own occupational income	0.1352***	0.1021**	0.0400	0.0271	0.0801*
	(0.040)	(0.044)	(0.042)	(0.037)	(0.041)
Constant	0.5460***	0.8089***	1.2933***	0.6110**	0.9451***
	(0.024)	(0.097)	(0.233)	(0.268)	(0.353)
Demographic controls		X	X	X	X
Geographic controls			X	X	X
Region controls				X	X
Home county wage					X
Observations	250	250	248	248	172
Adjusted R-Squared	0.0456	0.100	0.286	0.420	0.440

Note: Standard errors in parentheses. Skill premium is the log ratio of carpenter's to laborer's wage at the county level. Sample is restricted to men who live with an older, employed male relative in year 1870 (none meets this restriction in the 1850-60 sample). Sample expanded to include workers who begin in skilled jobs to increase sample size. Standard errors clustered at the destination county level. See Tables 1 and 2 for details.

Figure 2·1: Placement of Select Occupations in Different Occupational Rankings in 1% 1870 IPUMS sample



Chapter 3

Intergenerational Mobility across Three Generations in the Nineteenth Century: Evidence from the U.S. Census (with Claudia Olivetti and Daniele Paserman)

3.1 Introduction

Recent remarks made by Alan Krueger, President of the Council of Economic Advisers, have reignited economists' interest in intergenerational mobility. How is economic status transmitted across generations? How much does one's childhood environment matter for economic outcomes? The research to date (see Solon, 1999 and Black and Devereux, 2011 for extensive surveys) has mostly focused on the estimation of the father-son intergenerational elasticity, the parameter that measures the effect of a proportional increase of income in the father's generation on the income of the son's generation. These studies have often assumed that the intergenerational transmission of income follows an AR(1) process, an assumption driven primarily by lack of data linking more than two generations. However, this assumption may not hold in reality, for a number of reasons. As modeled in Solon (2013), grandparents may make independent human capital investments in grandchildren. Even if grandparents do not have a direct effect on children's outcomes, the inclusion of multigenerational effects may serve to rectify attenuation bias stemming from the mis-measurement of single generation effects. For instance, in the presence of discrimination or "ethnic capital," a parent's individual income may provide an incomplete picture of the transfer made from one generation to the next. This may also be true if socioeconomic status is measured with error.[1]

Identifying these multigenerational effects has important implications about the persistence of income inequality over time. Most existing estimates place the intergenerational elasticity for the US at between 0.4 and 0.5. If the process of intergenerational income transmission is AR(1), this implies that a given shock to income will fade out relatively quickly: the third order autocorrelation coefficient is between 0.064 and 0.125. Clearly, a higher order autoregressive process would imply a much slower regression to the mean. In other words, the degree of persistence of socioeconomic status across generations could be a lot higher than what we believe.

In this paper, we attempt to overcome the limitations of existing studies by estimating intergenerational elasticities across three generations for the United States during the late 19$^{\text{th}}$ and early 20$^{\text{th}}$ centuries. We apply and extend the methodology originally developed by Olivetti and Paserman (2012) in order to measure intergenerational elasticities between fathers (G1), children (G2) and grandchildren (G3). The key to the methodology is the construction of synthetic cohorts, or pseudo-panels, where individuals are grouped by their first names. Specifically, intergenerational correlations between fathers and sons are calculated as the correlation between the average outcomes of individuals with a specific first name in a given year and the average outcomes of the fathers of individuals with that name 20 or 30 years earlier. Extending this idea, the correlation between fathers and *grandsons* (i.e., the second-order autocorrelation) can be calculated as the correlation between the average income of fathers of children with a given first name, and an estimate of the average income of sons of children with that first name.

The intuition for why this methodology works can be explained using a simple ex-

[1]See Solon (2013) for a theoretical treatment of each of these possibilities.

ample. Assume that the only possible names for boys in generation G2 are Adam and Zachary. Moreover, assume that in generation G1, high socioeconomic status parents are more likely to name their child Adam, while Zachary is more common among low socioeconomic status parents. In a society with a high degree of intergenerational mobility, we would not expect the adult Adams to have much of an advantage on the adult Zacharys. Moreover, in the next generation (G3) the sons of Adam should be almost indistinguishable from the sons of Zachary. Therefore, the correlation in average income of people with a given name, their fathers and their sons will be a good measure of intergenerational mobility.

One important advantage of this methodology is that it applies equally well to women: just replace Adam and Zachary in the previous example with Abigail and Zoe, and use husband's income as the measure of women's socioeconomic status. Olivetti and Paserman (2012) used this methodology to provide the first estimate of intergenerational mobility between fathers and daughters in the late 19th and early 20th Centuries. In the case of three generations, the methodology will allow us to estimate four different channels of intergenerational transmission of socioeconomic status: fathers-sons-grandsons, fathers-sons-granddaughters, fathers-daughters-grandsons and fathers-daughters-granddaughters. Thus our analysis has the potential to uncover different mechanisms through which gender differentials in intergenerational mobility may arise.

Using data from the 1850 to 1930 US Census 1% sample, we find evidence of a strong second-order autoregressive coefficient for the process of intergenerational transmission of income. That is, even after controlling for the income in generation G2 ("father's income"), the income of generation G1 ("grandfather's income") has a large and positive effect on the income of generation G3 ("grandson's income"). This finding suggests that traditional estimates of intergenerational mobility that assume a first-order autoregressive process for income may substantially understate the true extent of intergenerational persistence in economic status.[2] In addition we find evidence of a gender differential in the strength of the correlation between the three generations. Our results indicate that the intergenerational elasticity between grandfathers and grandchildren is stronger if G2 is a male. We discuss a three generation dynastic model in which there is tension between G1's and G2's preferences over G2's (and G3's) consumption. This framework can rationalize our findings if the timing of intergenerational transfers is gender specific; for example, if parents assign dowries to their daughters and leave bequests to their sons.

The rest of the paper is organized as follows. The next section discusses the methodology as well as the data used for the analysis and some measurement issues. The main results and some robustness checks are presented in Section 3.3. Finally, Section 3.4 presents a simple dynastic model, which we use to provide a possible interpretation for our findings.

3.2 Methodology and Data

We use data from the 1850 to 1930 US Census 1% samples from IPUMS (Ruggles et al., 2010). Even though IPUMS has recently released a set of Linked Representative Samples that link records from 1880 complete count database to each one of the 1850-1930 on percent Census samples, these data do not allow individuals to be linked across more than two generations. Moreover, these data cannot be used to connect married women to their fathers. Therefore, it is necessary to use our synthetic-cohort-based methodology to create a data set with information on economic status across three generations. The core idea, as described in the previous section, is to use first

[2]Another possible interpretation is that this finding reflects measurement error, as noted by Gary Solon in remarks made at the PSID Conference on Inequality across Multiple Generations, Ann Arbor, September 2012.

names to construct the synthetic cohorts.

Linking generation G1 to generation G2

To link individuals from generations G1 and G2, we follow the same approach used in Olivetti and Paserman (2012). Assume that we observe G1 and G2 in two separate cross-sections. The first is at time t, that is, the Census year in which G1 individuals are adults and G2 individuals are children, and the second is at time $t + 20$, that is, the census year in which G2 individuals are adults (20 years later in this example). Our strategy is to base our intergenerational links on individuals' first names, which are available for both adults and children in each cross-section.

Define $\tilde{y}_{j,t}$ as the average log earnings of G1 fathers of children named j and $\tilde{y}_{j,t+20}$ as the average log earnings of G2 adults named j. One can then merge the two cross sections by first names, and then estimate the income elasticity across generations G1 and G2 by a weighted least squares regression of $\tilde{y}_{j,t+20}$ on $\tilde{y}_{j,t}$. The weights are equal to the frequency counts of first names in the sample of G2 adults.[3]

Olivetti and Paserman (2012) show that if names carry information about economic status, this estimator will be informative of the underlying parameters governing the process of intergenerational mobility. In addition, they conduct a numerical exercise to show that the estimated intergenerational elasticity is not sensitive to the characteristics of the name distribution (the degree of concentration of names, and the extent to which names carry economic content). Therefore, they argue that the observed trends in estimated intergenerational mobility are unlikely to be driven by changes in the parameters governing the name distribution, but rather reflect more fundamental changes in the parameters that govern the income transmission process across generations (the returns to human capital, the degree of inheritability of traits,

[3]This "means-on-means" regression is similar to the synthetic cohort approach pioneered by Browning et al. (1985) and Attanasio and Weber (1995). In our case, the synthetic cohorts are defined on the basis of both first names and age.

and the degree of assortative mating in the marriage market).

Linking generation G2 to generation G3

Linking generations G2 and G3 is slightly more complicated. We would like to know the average log earnings of G3 children born to G2 fathers named j. However, as the earnings of G3 are observed when these individuals are adults, we can no longer link them to their fathers. Therefore, we proceed as follows. First, we calculate $q_{j,k}$ as the fraction of children (G3) named k of fathers (G2) named j. This value is taken from Census year $t+20$, in which G2 individuals are adults, and G3 children still live at home with their parents. Second, we calculate $\tilde{y}_{k,t+40}$, the average log earnings of G3 adults named k (this average is calculated from Census year $t+40$). Finally, we calculate $\tilde{y}_{j,t+40}$ as:

$$\tilde{y}_{j,t+40} = \sum_k q_{j,k} \tilde{y}_{k,t+40}$$

In other words, the average log earnings of the children of G2 parents named j are a weighted average of the name-specific average log earnings of G3 adults, with the weights equal to the fraction of G3 individuals with that name among all the children of G2 parents named j. For example, suppose that adults named Adam in year $t+20$ have children named David, Edward and Fred. The income assigned to G3 for the group of adults named "Adam" is the weighted average, with weights $\frac{1}{3}$, $\frac{1}{3}$, $\frac{1}{3}$, of the average income at time $t+40$ of all the G3 individuals named David, Edward and Fred, respectively.

One can then merge this cross section to the one linking G1 and G2 by first names, then obtain an estimate of the income elasticity across the three generations by running a weighted least squares regression of $\tilde{y}_{jt+40}$ on $\tilde{y}_{j,t+20}$ *and* $\tilde{y}_{j,t}$, where the weights are equal to the frequency counts of first names in the sample of G2 adults.

The description above was presented in terms of the father-son-grandson relationship. It is easy to see, however, that the methodology can be applied to fathers-son-granddaughters, fathers-daughters-grandsons, and fathers-daughters-granddaughters. Therefore, we will be able to analyze gender differentials in the transmission of economic status across multiple generations.

Data and Measurement Issues

Data with individual names is available from IPUMS for every decadal Census from 1850 to 1930, with the exception of 1890. This means that we can calculate our three-generation measures of intergenerational mobility for two triplets observed at a distance of 20 years from one another (1860-1880-1900, and 1880-1900-1920); and for two triplets of observations observed at a distance of 30 years from one another (1850-1880-1910 and 1870-1900-1930). This gives us a unique long-run perspective on the transmission of economic status across generations.

A challenge that applies to all computations of historical intergenerational elasticities is to obtain appropriate quantitative measures of socioeconomic status. Because income and earnings at the individual level are not available before the 1940 Census, we are constrained to use measures of socioeconomic status that are based on individuals' occupational status. While this contrasts with the current practice among economists, who prefer to use direct measures of income or earnings if available, there is a long tradition in sociology to focus on occupational categories (Erikson and Goldthorpe, 1992). One of the advantages of the IPUMS data set is that it contains a harmonized classification of occupations, and several measures of occupational status that are comparable across years. For our benchmark analysis, we choose the OCCSCORE measure of occupational standing. This variable indicates the median total income (in hundreds of dollars) of persons in each occupation in 1950.

A second challenge arises from our methodology for measuring generation G3 occupational income. As explained above, the income of children of generation G2 is computed as a weighted mean of mean incomes by first names. This implies that the distribution of income for G3 is more compressed than that of G1 and G2. As we demonstrate below, this produces an even stronger attenuation bias in the OLS estimate. Therefore, in most of our analysis we use percentile rank of mean log occupational income as a way to get around this problem.

3.3 Results

In this section, we present our estimates of intergenerational income elasticities across three generations. Panel A of table 1 presents these estimates when we use the log occupational income score as our dependent variable. The first column shows that the intergenerational elasticity between the 1860 and 1880 cohorts is about 0.37, a result in line with that obtained in Olivetti and Paserman (2012). Column 3 regresses the log occupational scores of G3 males on those of their grandfathers (G1). The coefficient is substantially smaller, but still highly statistically significant. Column 4 includes the income of both G1 and G2 males on the right hand side. Both coefficients are statistically significant, with the coefficient on G2 income about three times as large. The statistically significant coefficient on G1 income implies that the intergenerational income transmission process is better characterized as an AR(2), and ignoring the second order autoregressive term will lead to overstating the extent of long-run mobility across generations.

One concern with the results above is that the estimated elasticities appear to be substantially smaller when we use grandson's income as the dependent variable. This problem is illustrated in column 2, where we estimate the simple one-generation intergenerational elasticity using the 1880 and 1900 cohorts. Here, the generations are linked by the first names of the older generation, meaning that average log earnings of the younger generation must be calculated as the weighted average of name-specific log earnings, as described previously. This procedure may introduce substantial measurement error, potentially biasing the estimated elasticity toward zero. In addition, the distribution of G3 earnings, being calculated as an average of averages, tends to be very compressed. This may in itself lead to a downward bias in the estimates.

To overcome these problems, we re-estimate the regressions, but now using the log earnings percentile rank as the dependent variable. This should at least alleviate the problem of excessive compression of the distribution of the dependent variable. The results are presented in panel B of table 1. Columns 1 and 2 show the one-generation estimates of intergenerational mobility. The coefficient in column 1 indicates that going from the bottom to the top percentile of earnings in generation G1 is associated with an increase of about 50 percentiles in the earnings of generation G2. The coefficient is somewhat smaller for the 1880-1900 generations, but at least has the same order of magnitude. Column 3 links G3 to G2, and Column 4 adds the link to G1. As in panel A, we find that G1's earnings rank has a large and significant effect on the earnings rank of G3, even after controlling for the earnings rank of G2.

Table 2 presents percentile rank regressions of G3 earnings on the earnings of G1 and G2, using all possible gender combinations and both decade triplets in which the distance between generations is 20 years (1860-1880-1900; 1880-1900-1920). In the top panel, we show results using both decade triplets and genders for G3 when G2 is male; the bottom panel repeats this analysis using G2 females. Several patterns emerge from this exercise. First, there is a slight upward time trend in the coefficient on G1 earnings, indicating that the effect of grandparents' earnings on grandchildren's earnings is increasing over time.

Differences in the G1-G3 intergenerational elasticity by gender can be seen in this table. First of all, holding the decade and the gender of G3 constant, the effect of G1 earnings is greatest when G2 is male. To see this, compare the top and bottom panels of each column in table 2. The coefficient on G1 earnings is always substantially larger in the top panel. What this means is that earnings are more strongly related to paternal grandfathers than maternal grandfathers. Second, holding year and the gender of G2 constant, the coefficient on G1 earnings is larger when G3 is female. This can be seen from a comparison between columns (1) and (3) and between columns (2) and (4) in the top panel, and between these same columns in the bottom panel. In other words, grandfathers and granddaughters have more closely related earnings than grandfathers and grandsons.

Table 3 repeats the analysis in table 2 using decade triplets separated by 30-year intervals (1850-1880-1910; 1870-1900-1930). These coefficients do not systematically differ in magnitude from those in table 2. However, this table offers little evidence of an upward time trend. The gender differences seen in table 2 are less pronounced here; however, by and large, they are still present.

One concern we have is that comparisons by gender may be sensitive to the way our samples are constructed. For example, we measure a woman's socioeconomic status by the earnings of her husband. This means that all women in our sample are married, whereas men need not be married to be included. Then, we may be measuring differences in intergenerational income transmission by marital status rather than gender. Another concern is that G1 individuals need to have children ages 0-15 in order to appear in the sample; the same is not true of later generations.

To ensure that our results are not being driven by these details of our sample construction, we redo the analysis imposing different restrictions on G2 and G3. The restrictions we impose on G2 are (i) Men must be married; (ii) Men and women must be married with children ages 0-15; (iii) Men and women must be married with children ages 0-15, and their children must have names that are linked to the sample of G3 adults. We also impose restrictions (i) and (ii) on G3. We calculate the G1-G3 intergenerational elasticity for each of 12 combinations of these sample restrictions (including the baseline of no restrictions). The results, using the 1860-1880-1900 sample, are reported in appendix table 1.

To clearly summarize these results, we compile all G1-G3 intergenerational elasticities estimated under different sample restrictions in each decade triplet. There are 192 such estimates. We regress these on indicators for the decade in which G2's earnings are measured (1880 or 1900), the interval that separates generations (20 or 30 years), the gender of G2, the gender of G3, and categorical variables indicating which sample restrictions are imposed. Standard errors are clustered at the specification level. We report these results in table 4. The positive time trend does not stand up to the above sample restrictions, nor does the finding that the relationship between grandfathers and granddaughters is stronger than the relationship between grandfathers and grandsons. However, the finding that the correlation between G1 and G3 is greater when G2 is male persists. We consider this our main finding so far.

3.4 Model

In order to interpret our findings, we present a simple three-generation dynastic model of consumption and human capital investment. The key ingredient of the model is that there is a tension between the desired allocation of consumption across the three generations between the decision-makers in generations 1 and 2. This tension derives from the fact that each generation discounts heavily the utility of future generations relative to its own utility, but the discount factor between any two future generations is relatively low. In other words, each generation is characterized by quasi-hyperbolic, or $\beta-\delta$ preferences.[4] The investment in the third generation's human capital depends on whether the second generation is able to reoptimize over the allocation of resources between G2 and G3, or whether it must follow the allocation chosen by G1. We conjecture that, because of the structure of marriage markets in the 19$^{\text{th}}$ Century and the timing of transfers across generations, second generation daughters (and their husbands) may have been more likely to reoptimize, thus inducing the lower elasticity between first and third generation's income when the second generation is female.

Formally, we consider a three-generation dynasty. Each generation derives utility from its own consumption and from that of the following generations. Therefore:

$$
\begin{aligned}
U_1(c_1, c_2, c_3) &= \ln(c_1) + \beta\delta \ln(c_2) + \beta\delta^2 \ln c_3 \\
U_2(c_1, c_2) &= \ln(c_2) + \beta\delta \ln(c_3) \\
U_3(c_3) &= \ln(c_3)
\end{aligned}
$$

We assume throughout that $\beta < 1$, reflecting the fact that each generation puts more weight on its own utility relative to future generations' utility; and $\delta < 1$, reflecting the fact that the weight placed on more distant generations' utility also declines. Notice that for G1, the discount factor between its own utility and that of G2 is $\beta\delta$, while the discount factor between G2 and G3's utility is only δ. This captures the fact that the discount rate between the present and any period in the future is higher than the discount rate between any two periods in the distant future.

Each generation can allocate its income Y_t between its own consumption c_t and

[4] Quasi-hyperbolic preferences have been made popular in recent years to model the intra-personal self-control problems in consumption and savings decisions and other contexts (Laibson, 1997; O'Donoghue and Rabin, 1999; DellaVigna and Paserman, 2005). However, one of the first applications of $\beta-\delta$ preferences (Phelps and Pollak, 1968) was to an intergenerational growth model very much like the one considered here.

investment in the following generation's human capital, I_{t+1}. Generation $t+1$'s income is a function of generation t's investment:

$$Y_{t+1} = RI_{t+1}.$$

To solve for the optimal allocation of resources across generations, we consider two alternative possibilities. In the first case, G1 decides on how to allocate resources for all three generations. In the second case, G2 can reoptimize and decide on the allocation of resources from that point onwards. G1's decision takes into account G2's decision, and decides how much to consume and how much to invest in the next generation as a best response to G2's actions. In the language if the quasi-hyperbolic discounting literature, the first case corresponds to that of an agent who can perfectly commit to the full sequence of decisions made in period 1, while the second case corresponds to that of a *sophisticated* agent. We label the optimal consumption choices made by the two agents $\left\{c_t^{COMM}\right\}_{t=1}^3$ and $\left\{c_t^{SOPH}\right\}_{t=1}^3$, respectively; and the resulting income levels $\left\{Y_t^{COMM}\right\}_{t=1}^3$ and $\left\{Y_t^{SOPH}\right\}_{t=1}^3$.

We are interested primarily in the *incomes* of G2 and G3, and how they correlate with G1's income. The following proposition holds:

Proposition 1. *(a) If G1 can commit to all future decisions, the incomes of G2 and G3 will be, respectively*

$$Y_2^{COMM} = \frac{R\beta\delta\left(1+\delta\right)}{1+\beta\delta+\beta\delta^2}Y_1$$

and

$$Y_3^{COMM} = \frac{R^2\beta\delta^2}{1+\beta\delta+\beta\delta^2}Y_1;$$

(b) If G1 cannot commit to all future decisions, the incomes of G2 and G3 will be, resepctively

$$Y_2^{SOPH} = \frac{R\beta\delta\left(1+\delta\right)}{1+\beta\delta+\beta\delta^2}Y_1$$

and

$$Y_3^{SOPH} = \frac{R^2\beta^2\delta^2\,(1+\delta)}{(1+\beta\delta+\beta\delta^2)\,(1+\beta\delta)}Y_1.$$

Proof. When G1 can commit to all future resource allocations, he will simply solve the following straightforward maximization problem:

$$\max_{c_1,c_2,c3} \ln(c_1) + \beta\delta\ln(c_2) + \beta\delta^2\ln(c_3) \quad \text{s.t.} \quad Y_1 = c_1 + \frac{c_2}{R} + \frac{c_3}{R^2}$$

This generates the following optimal choices of c_1, c_2, and c_3:

$$c_1 = \frac{1}{1+\beta\delta+\beta\delta^2}Y_1$$
$$c_2 = \frac{\beta\delta R}{1+\beta\delta+\beta\delta^2}Y_1$$
$$c_3 = \frac{\beta\delta^2 R^2}{1+\beta\delta+\beta\delta^2}Y_1$$

Part (a) follows from the fact that $Y_3 = c_3$ and $Y_2 = c_2 + \frac{Y_3}{R}$.

When G1 cannot commit to future resource allocations, he will anticipate G2's resource allocation decision and make his decisions accordingly. Taking Y_2 as given, G2 will solve the following:

$$\max_{c_2,c_3} \ln(c_2) + \beta\delta\ln(c_3) \quad \text{s.t.} \quad Y_2 = c_2 + \frac{c_3}{R}$$

The solution to this problem yields the following optimal choices of c_2 and c_3:

$$c_2 = \frac{1}{1+\beta\delta}Y_2$$
$$c_3 = \frac{\beta\delta R}{1+\beta\delta}Y_2$$

Then, G1's optimization problem can be written:

$$\max_{c_1,Y_2} \ln(c_1) + \beta\delta\ln\left(\frac{Y_2}{1+\beta\delta}\right) + \beta\delta^2\ln\left(\frac{\beta\delta R Y_2}{1+\beta\delta}\right) \quad \text{s.t.} \quad Y_1 = c_1 + \frac{Y_2}{R}$$

The solution to this problem for Y_2 is

$$Y_2 = \frac{R\beta\delta(1+\delta)}{1+\beta\delta+\beta\delta^2}Y_1$$

The value of Y_3 follows from the solution for c_3 given above, and from the fact that $Y_3 = c_3$. $\square$

Proposition 1 also allows one to calculate the relationship between the incomes of the different generations. Let $\eta_{2,1}$ and $\eta_{3,1}$ be, respectively, the slope coefficients in regressions of Y_2 and Y_3 on Y_1. It follows directly from the proposition that $\eta_{2,1}^{SOPH}$=$\eta_{2,1}^{COMM}$ and $\eta_{3,1}^{SOPH} < \eta_{3,1}^{COMM}$.

The intuition for the second result is straightforward. If G1 can commit to a given consumption path for all three generations, it will allocate resources between G2 and G3 in a relatively egalitarian way: from its perspective, G3's utility is discounted only by a factor δ relative to G2's utility. On the other hand, if G2 can reoptimize given its allocation, it will put much more weight on its own consumption, as the discount factor that it applies between its own utility and G3's utility is $\beta\delta$.[5]

How do these results relate to our findings on how the strength of the G1-G3 intergenerational elasticity depends on the gender of the middle generation? There is a literature that examines the relationship between marriage institutions, postmarital location rules and property rights. Botticini and Siow (2003) argue that in *virilocal* societies, where married daughters leave their parental nest and married sons do not, altruistic parents will leave dowries to their daughters and bequests to their sons to mitigate a free-rider problem. Other papers focus on the role of marital arrangements, with men remaining immobile and specializing in farm production and women moving to new households, for consumption smoothing and agency problems (see for example, Rosenzweig and Stark, 1989, based on data on rural India, and Fafchamps and Quisumbing, 2005, on rural Ethiopia). While dowries were relatively uncommon in North America in the 19$^{\text{th}}$ Century,[6] a quick examination of census

[5]The result that the second generation's income is identical under both allocation rules is less interesting, and depends on the specific functional form of the utility function (logarithmic utility).

[6]Botticini and Siow (2003) document that in late 18$^{\text{th}}$ Century Connecticut, between 46 and 67 percent of married daughters were assigned, inter vivos transfers from their family of origin, likely

data reveals a tendency for young married couples to reside with husbands' rather than wives' parents.[7] During the period of focus of this study, only 10-12 percent of married couples under 35 resided in the same household a parent; however, that parent was significantly more likely to belong to the husband. This is especially true of agricultural families: young couples residing with a parent were twice as likely to be living with the husband's parents rather than with the wife's parents. This may also mask a tendency for families to reside in the same locality as the husband's parents, even if they do not reside in the same household. As such, it is possible that transfers from parents to daughters were more likely to occur at marriage than were transfers from parents to sons, even if formal dowries were unusual.

We conjecture that the timing of intergenerational transfers may affect the ability of G2 to decide on the allocation of consumption between itself and G3. Specifically, assume that G2 daughters receive transfers from their parents upon marriage. Since 19$^{\text{th}}$ Century women relinquished control of their assets to their husbands, it seems likely that the G1 patriarch would have little say over the allocation of resources between G2 and G3. On the other hand, G2 sons were more likely to receive a bequest, upon G1's death. The fact that G1 could withhold the transfer of resources to his male offspring implies that it was also easier for G1 to monitor the allocation of resources between G2 and G3, and therefore guarantee that the investment in the grandchild's human capital would be sufficiently high.[8] If families were more likely to locate near paternal than maternal grandparents, it may also be that the direct human capital transfer from parental grandparents to grandchildren was greater.

at the time of their marriage. However, by 1820's, only 40 percent received such transfers.

[7] These tabulations are based on the 1 percent IPUMS samples from 1880 and 1900.

[8] For more on the timing of transfers across generations, see Botticini and Siow (2003), Rosenzweig (1988), Rosenzweig and Stark (1989).

3.5 Conclusion

In this paper, we have estimated intergenerational elasticities across three generations for the US spanning the late 19th and early 20th Century. We find that the intergenerational income process exhibits a strong second-order autoregressive coefficient. We also find that the grandfather-grandchild intergenerational elasticity is larger when the middle generation is male, and we rationalize these findings using a simple three-generation dynastic model where there is a tension between G1's and G2's preferences over G2's consumption, and the timing of transfers is gender specific. These results can have important implications for our understanding of the persistence of socioeconomic status over the long run.

3.6 References

Attanasio, Orazio P. and Guglielmo Weber (1995). "Is Consumption Growth Consistent with Intertemporal Optimization? Evidence from the Consumer Expenditure Survey." *Journal of Political Economy*, 103(6), pp. 1121-1157.

Browning, Martin; A. Deaton and M. Irish (1985). "A Profitable Approach to Labor Supply and Commodity Demands over the Life-Cycle." *Econometrica,* 53(3), pp. 503-544.

Black, Sandra E. and Paul J. Devereux (2010). "Recent Developments in Intergenerational Mobility." NBER Working Paper 15889.

Botticini, Maristella and Aloysius Siow (2003). "Why Dowries?" *American Economic Review*, 93(4), pp. 1385-1398.

Della Vigna, Stefano and M. Daniele Paserman (2005). "Job Search and Impatience." *Journal of Labor Economics*, 23(3), pp. 527-588.

Erikson, Robert and John H. Goldthorpe (1992). *The Constant Flux: A Study of Class Mobility in Industrial Societies.* Oxford: Clarendon Press.

Fafchamps, Marcel and Agnes R. Quisumbing (2005). "Marriage, Bequest, and Assortative Matching in Rural Ethiopia." *Economic Development and Cultural Change.*

53(2): 347-380.

Laibson, David I. (1997). "Golden Eggs and Hyperbolic Discounting." *Quarterly Journal of Economics*, 112(2), pp. 443-477.

O'Donoghue, Ted and Matthew Rabin (1999). "Doing It Now or Later." *American Economic Review*, 89(1), pp. 103-124.

Olivetti, Claudia and M. Daniele Paserman (2012). "In the Name of the Father: Marriage and Intergenerational Mobility in the United States, 1850-1930." Manuscript, Boston University.

Phelps, Edmund S. and Robert A. Pollak (1968). "On Second-Best National Saving and Game-Equilibrium Growth." *Review of Economic Studies*, 35(2), pp. 185-199.

Rosenzweig, Mark R. (1988). "Risk, Private Information, and the Family." *American Economic Review.* 78(2): 245-250.

Rosenzweig, Mark R. and Oded Stark (1989). "Consumption Smoothing, Migration, and Marriage: Evidence from Rural India." *Journal of Political Economy.* 97: 905-26.

Ruggles, Steven J., Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek (2010) Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database]. Minneapolis: University of Minnesota.

Solon, Gary R. (2013). "Theoretical Models of Inequality Transmission Across Multiple Generations." NBER Working Paper no. 18790.

—. (1999). "Intergenerational Mobility in the Labour Market," in (O. Ashenfelter and D. Card, eds.), *Handbook of Labor Economics*, Volume 3A, pp. 1761-800, Amsterdam: North-Holland.

3.7 Tables

Table 3.1: Intergenerational Income Elasticities for Three Generations: Levels and Percentile Rank

	(1)	(2)	(3)	(4)
Dependent variable:	1880 G2 male	1900 G3 male	1900 G3 male	1900 G3 male
	Panel A: Log occupational income			
1860 G1	0.369		0.0535	0.0298
	(0.0295)		(0.0105)	(0.0112)
1880 G2 male		0.0764		0.0642
		(0.0109)		(0.0118)
Constant	1.837	2.754	2.821	2.703
	(0.0857)	(0.0316)	(0.0304)	(0.0370)
Observations	874	882	875	874
R-squared	0.152	0.053	0.029	0.061
	Panel B. Percentile rank of log occupational income			
1860 G1	0.456		0.243	0.120
	(0.0288)		(0.0318)	(0.0350)
1880 G2 male		0.328		0.270
		(0.0321)		(0.0363)
	32.53	28.89	33.91	25.14
Constant	(1.784)	(1.976)	(1.972)	(2.251)
Observations	874	882	875	874
R-squared	0.224	0.106	0.063	0.118

Table 3.2: Intergenerational Elasticities Across Three Generations: Percentile Rank Regressions at 20 Year Intervals

VARIABLES	(1) G3 Male 1900	(2) G3 Male 1920	(3) G3 Female 1900	(4) G3 Female 1920
G1	0.120	0.191	0.218	0.227
	(0.0350)	(0.0324)	(0.0367)	(0.0310)
G2 Male	0.270	0.296	0.231	0.183
	(0.0363)	(0.0342)	(0.0378)	(0.0327)
Constant	25.14	19.38	23.65	25.06
	(2.251)	(1.938)	(2.348)	(1.850)
Observations	874	1,161	848	1,171
R-squared	0.118	0.170	0.140	0.137
G1	-0.0218	0.0479	0.151	0.176
	(0.0292)	(0.0237)	(0.0283)	(0.0239)
G2 Female	0.298	0.379	0.397	0.401
	(0.0324)	(0.0253)	(0.0311)	(0.0256)
Constant	31.14	22.60	19.55	16.20
	(1.729)	(1.477)	(1.658)	(1.474)
Observations	1,146	1,675	1,132	1,627
R-squared	0.090	0.176	0.259	0.266

Table 3.3: Intergenerational Elasticities Across Three Generations: Percentile Rank Regressions at 30 Year Intervals

VARIABLES	(1) G3 Male 1910	(2) G3 Male 1930	(3) G3 Female 1910	(4) G3 Female 1930
G1	0.0902	0.195	0.106	0.215
	(0.0308)	(0.0288)	(0.0302)	(0.0276)
G2 Male	0.321	0.305	0.186	0.282
	(0.0357)	(0.0326)	(0.0360)	(0.0313)
Constant	24.66	19.07	32.76	20.72
	(2.150)	(1.842)	(2.168)	(1.772)
Observations	996	1,250	970	1,265
R-squared	0.123	0.184	0.066	0.193
G1	0.0845	0.0875	0.156	0.0489
	(0.0286)	(0.0240)	(0.0284)	(0.0240)
G2 Female	0.258	0.354	0.359	0.472
	(0.0341)	(0.0264)	(0.0343)	(0.0262)
Constant	29.09	23.25	21.20	20.01
	(1.962)	(1.544)	(1.961)	(1.540)
Observations	958	1,414	935	1,388
R-squared	0.101	0.180	0.204	0.255

Table 3.4: Summary of G1-G3 Intergenerational Income Elasticities under Different Sample Restrictions: Percentile Rank Regressions

Dependent variable:	Intergenerational income elasticity: G1-G3
Year (G2)=1900	0.0198
	(0.0203)
Interval = 30 years	-0.0278
	(0.0203)
G2 Male	0.0458**
	(0.0203)
G3 Male	-0.0262
	(0.0203)
Sample restrictions:	
G2 married	-0.0185**
	(0.00639)
G2 married w children 0-15	-0.0168**
	(0.00723)
G2 married w children, linked to G3 sample	-0.00426
	(0.00886)
G3 married	0.0153**
	(0.00549)
G3 married w children 0-15	0.0174*
	(0.00897)
Constant	0.129***
	(0.0243)
Observations	192
R-squared	0.369

Table 3.5: Intergenerational Income Elasticities Across Three Generations, 1860-1880-1900: Different Sample Restrictions for G2 and G3

		G3: Baseline				G3: Married				G3: Married with children ages 0-15			
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	VARIABLES	G3 Male	G3 Female	G3 Male	G3 Female	G3 Male	G3 Female	G3 Male	G3 Female	G3 Male	G3 Female	G3 Male	G3 Female
G2: Baseline	G1	0.120	0.218	-0.0185	0.151	0.206	0.218	0.0492	0.151	0.240	0.250	0.0747	0.142
		(0.0350)	(0.0367)	(0.0292)	(0.0283)	(0.0372)	(0.0367)	(0.0270)	(0.0283)	(0.0379)	(0.0364)	(0.0268)	(0.0279)
	G2 Male	0.270	0.231			0.278	0.231			0.242	0.206		
		(0.0363)	(0.0378)			(0.0385)	(0.0378)			(0.0390)	(0.0375)		
	G2 Female			0.298	0.397			0.361	0.397			0.361	0.415
				(0.0324)	(0.0311)			(0.0301)	(0.0311)			(0.0298)	(0.0307)
	Constant	25.14	23.65	31.16	19.57	21.45	23.65	25.64	19.57	24.00	23.69	25.49	19.62
		(2.251)	(2.348)	(1.732)	(1.657)	(2.388)	(2.348)	(1.598)	(1.657)	(2.411)	(2.317)	(1.583)	(1.636)
	Observations	874	848	1,146	1,132	865	848	1,137	1,132	861	846	1,137	1,131
	R-squared	0.118	0.140	0.091	0.259	0.153	0.140	0.182	0.259	0.150	0.148	0.200	0.271
G2: Married	G1	0.112	0.160	-0.0185	0.151	0.149	0.160	0.0492	0.151	0.178	0.176	0.0747	0.142
		(0.0345)	(0.0356)	(0.0292)	(0.0283)	(0.0357)	(0.0356)	(0.0270)	(0.0283)	(0.0363)	(0.0354)	(0.0268)	(0.0279)
	G2 Male	0.266	0.334			0.370	0.334			0.355	0.325		
		(0.0364)	(0.0372)			(0.0376)	(0.0372)			(0.0381)	(0.0369)		
	G2 Female			0.298	0.397			0.361	0.397			0.361	0.415
				(0.0324)	(0.0311)			(0.0301)	(0.0311)			(0.0298)	(0.0307)
	Constant	26.12	21.31	31.16	19.57	19.58	21.31	25.64	19.57	21.21	21.15	25.49	19.62
		(2.234)	(2.277)	(1.732)	(1.657)	(2.307)	(2.277)	(1.598)	(1.657)	(2.338)	(2.261)	(1.583)	(1.636)
	Observations	871	845	1,146	1,132	862	845	1,137	1,132	858	843	1,137	1,131
	R-squared	0.109	0.170	0.091	0.259	0.181	0.170	0.182	0.259	0.183	0.176	0.200	0.271
G2: Married with children ages 0-15	G1	0.122	0.162	0.00253	0.167	0.162	0.162	0.0667	0.167	0.180	0.176	0.0911	0.156
		(0.0347)	(0.0359)	(0.0289)	(0.0278)	(0.0361)	(0.0359)	(0.0267)	(0.0278)	(0.0365)	(0.0355)	(0.0266)	(0.0273)
	G2 Male	0.243	0.307			0.337	0.307			0.344	0.302		
		(0.0367)	(0.0375)			(0.0381)	(0.0375)			(0.0384)	(0.0371)		
	G2 Female			0.272	0.387			0.337	0.387			0.335	0.411
				(0.0318)	(0.0303)			(0.0294)	(0.0303)			(0.0292)	(0.0299)
	Constant	26.82	22.72	31.69	19.33	20.69	22.72	26.15	19.33	21.70	22.49	26.20	19.22
		(2.234)	(2.278)	(1.727)	(1.648)	(2.316)	(2.278)	(1.594)	(1.648)	(2.336)	(2.251)	(1.585)	(1.623)
	Observations	867	843	1,146	1,132	858	843	1,137	1,132	854	841	1,137	1,131
	R-squared	0.101	0.155	0.084	0.261	0.165	0.155	0.174	0.261	0.177	0.162	0.188	0.278
G2: Married with children ages 0-15 and linked to G3 sample	G1	0.127	0.181	0.0414	0.193	0.160	0.181	0.103	0.193	0.168	0.191	0.128	0.185
		(0.0336)	(0.0352)	(0.0283)	(0.0273)	(0.0347)	(0.0352)	(0.0261)	(0.0273)	(0.0348)	(0.0348)	(0.0258)	(0.0268)
	G2 Male	0.221	0.266			0.304	0.266			0.344	0.253		
		(0.0347)	(0.0363)			(0.0355)	(0.0363)			(0.0357)	(0.0357)		
	G2 Female			0.234	0.351			0.297	0.351			0.303	0.378
				(0.0312)	(0.0302)			(0.0288)	(0.0302)			(0.0286)	(0.0297)
	Constant	27.96	24.16	31.61	19.91	22.74	24.16	26.44	19.91	22.45	24.65	26.00	19.53
		(2.226)	(2.312)	(1.752)	(1.687)	(2.286)	(2.312)	(1.613)	(1.687)	(2.298)	(2.272)	(1.605)	(1.657)
	Observations	864	832	1,144	1,127	855	832	1,135	1,127	851	830	1,135	1,126
	R-squared	0.093	0.139	0.075	0.242	0.150	0.139	0.161	0.242	0.177	0.141	0.181	0.263

Bibliography

Abbring and Van den Berg (2005). "The Effect of Unemployment Insurance Sanctions on the Transition Rate from Unemployment to Employment." *Economic Journal.* 115(505): 602-630.

—. (2003a). "The Nonparametric Identification of Treatment Effects in Duration Models." *Econometrica.* 71(5): 1491-1517.

—. (2003b). "The Identifiability of the Mixed Proportional Hazards Competing Risks Model." *Journal of the Royal Statistical Society, Series B.* 65(3): 701-710.

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson (2010). "Europes Tired, Poor, Huddles Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration." NBER Working Paper no. 15684.

Abramitzky, Ran, Adeline Delavande, and Luis Vasconcelos (2011). "Marrying Up: The Role of Sex Ratio in Assortative Matching." *American Economic Journal: Applied Economics.* 2: 124-157.

Alston, Lee J. and Joseph P. Ferrie (2005). "Time on the Ladder: Career Mobility in Agriculture, 1890-1938." *Journal of Economic History.* 65(4): 1058-1081.

Alston, Lee J. and Kyle D. Kauffman (1997). "Agricultural Chutes and Ladders: New Estimates of Sharecroppers and "True Tenants" in the South, 1900-1920." *Journal of Economic History.* 57(2): 464-475.

Atack, Jeremy (1989). "The Agricultural Ladder Revisited: A New Look at an Old Problem with some Data from 1860." *Agricultural History.* 63(1): 1-25.

—-. (1988). "Tenants and Yeomen in the Nineteenth Century." *Agricultural History.* 62(3): 6-32.

Atack, Jeremay and Fred Bateman (1992). "Matchmaker, Matchmaker, Make Me a Match: A General Personal Computer-Based Matching Program for Historical Research" *Historical Methods.* 25(2):53-65.

—. (1987). *To Their Own Soil: American Agriculture in the Antebellum North.* Iowa State University Press.

Atack, Jeremy, Fred Bateman, and Robert A. Margo (2004). "Skill Intensity and Rising Wage Dispersion in Nineteenth-Century American Manufacturing." *Journal of Economic History* 64: 172-192.

Atack, Jeremy, Fred Bateman, Michael Haines, and Robert A. Margo (2010). "Did Railroads Induce or Follow Economic Growth? Urbanization and Population Growth in the American Midwest, 1850-60." Forthcoming in *Social Science History.*

Attanasio, Orazio P. and Guglielmo Weber (1995). "Is Consumption Growth Consistent with Intertemporal Optimization? Evidence from the Consumer Expenditure Survey." *Journal of Political Economy*, 103(6), pp. 1121-1157.

Bahr, Stephen J. (1979). "The Effects of Welfare on Marital Stability and Remarriage." *Journal of Marriage and the Family.* 41(3): 553-560.

Bailey, Martha J. (2006). "More Power to the Pill: The Impact of Contraceptive Freedom on Women's Life-Cycle Labor Supply." *Quarterly Journal of Economics.* 121(1): 289-320.

Baker, Michael, Jasmin Kantarevic and Emily Hanna (2004). "The Married Widow: Marriage Penalties Matter!" *Journal of the European Economic Association.* 2(4): 634-664.

Becker, Gary S. (1973). "A Theory of Marriage: Part I." *Journal of Political Economy.* 8(4): 813-846.

Becker, Gary S. (1991). *A Treatise on the Family.* Cambridge: Harvard University Press.

Black, Sandra E. and Paul J. Devereux (2010). "Recent Developments in Intergenerational Mobility." NBER Working Paper 15889.

Blau, Francine D. (1998). "Trends in the Economic Well-Being of Women." *Journal of Economic Literature.* 36(1): 112-165.

Blau, Francine D., Lawrence M. Kahn and Jane Waldfogel (2000). "Understanding Young Womens Marriage Decisions: The Role of Labor and Marriage Market Conditions." *Industrial and Labor Relations Review.* 53(4): 624-647.

Borjas, George J., Stephen G. Bronars, and Stephen J. Trejo (1992). "Self-Selection and Internal Migration in the United States." *Journal of Urban Economics.* 32(2): 159-85.

Botticini, Maristella and Aloysius Siow (2003). "Why Dowries?" *American Economic Review*, 93(4), pp. 1385-1398.

Brien, Michael J., Stacy Dickert-Conlin and David A. Weaver (2004). "Widows Waiting to Wed?" *Journal of Human Resources.* 39(3): 585-623.

Browning, Martin; A. Deaton and M. Irish (1985). "A Profitable Approach to Labor Supply and Commodity Demands over the Life-Cycle." *Econometrica,* 53(3), pp. 503-544.

Calhoun, Arthur W. (1919). *Social History of the American Family.* Vol. 3. Cleveland: The Arthur H. Clark Company.

Card, D., R. Chetty and A. Weber (2007). "Cash-On-Hand and Competing Models of Intertemporal Behavior: New Evidence from the Labor Market." *Quarterly Journal of Economics.* 122: 1511-1560.

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

Coelho, Philip R. and James F. Shepherd (1976). "Regional Differences in Real Wages: The United States, 1851-1880." *Explorations in Economic History.* 13: 203-230.

Costa, Dora L. (1993). "Height, Wealth, and Disease Among the Native-Born in the Rural, Antebellum North." *Social Science History.* 17(3): 355-383.

—. (1995). "Pensions and Retirement: Evidence from Union Army Veterans." *Quarterly Journal of Economics.* 110(2): 297-319.

—. (1997). "Displacing the Family: Union Army Pensions and Elderly Living Arrangement." *Journal of Political Economy.* 105(6): 1269-1292.

Costa, Dora L. and Matthew E. Kahn (2003). "Cowards and Heroes: Group Loyalty in the American Civil War." *Quarterly Journal of Economics.* 118(2): 519-548.

—. (2008). *Heroes and Cowards: The Social Face of War.* Princeton, NJ: Princeton University Press.

Cvercek, Tomas (2009). "When Harry Left Sally: A New Estimate of Marital Disruption in the U.S., 1860-1948." *Demographic Research.* 21: 719-758.

Della Vigna, Stefano and M. Daniele Paserman (2005). "Job Search and Impatience." *Journal of Labor Economics*, 23(3), pp. 527-588.

Doepke, Matthias and Michele Tertlit (2008). "Women's Liberation: What's in it for Men?" NBER Working Paper, no. 13919.

Duncan, O.D. (1961). "A Socioeconomic Index for All Occupations." in A. Reiss et al., *Occupations and Social Status.* Free Press.

Easterlin, Richard A. (1976). "Change and Farm Settlement in the Northern United States." *Journal of Economic History.* 36(1): 45-75.

—. (1971). "Does Human Fertility Adjust to the Environment" *American Economic Review.* 61(2): 399-407.

Eli, Shari (2010). "Wealth is Health: Pensions and Disease Onset in the Gilded Age." Unpublished paper, Berkeley University.

Erikson, Robert and John H. Goldthorpe (1992). *The Constant Flux: A Study of Class Mobility in Industrial Societies.* Oxford: Clarendon Press.

Fafchamps, Marcel and Agnes R. Quisumbing (2005). "Marriage, Bequest, and Assortative Matching in Rural Ethiopia." *Economic Development and Cultural Change.* 53(2): 347-380.

Fernandez, Raquel (2009). "Women's Rights and Development." NBER Working Paper, no. 15355.

Ferrie, Joseph P. (2005). "The End of American Exceptionalism: Mobility in the U.S. since 1850." NBER Working Paper no. 11324.

—-. (1997). "Migration to the Frontier in Mid-Nineteenth Century America: A Re-Examination of Turner's 'Safety Valve.' " Department of Economics, Northwestern University.

—-. (1996a). "A New Sample of Americans Linked from the 1850 Public Use Micro Sample of the Federal Census of Population to the 1860 Federal Census Manuscript Schedules." *Historical Methods.* 29: 141- 156.

—-. (1996b). "The Entry into the U.S. Labor Market of Antebellum European Immigrants, 1840-60." *Explorations in Economic History.* 34(3): 295-330.

Fitch, Catherine A. and Steven Ruggles (2000). "Historical Trends in Marriage Formation: The United States, 1850-1990." *The TIes that Bind: Perspectives on Marriage and Cohabitation.* Ed. Linda J. Waite, Christine Bachrach, Michelle Hindin, Elizabeth Thomson, and Arland Thornton. New York: de Gruyter: 59-88.

Fogel, Robert W. (2000). *Public Use Tape on the Aging of Veterans of the Union Army: Military, Pension, and Medical Records, 1860-1940, Version M-5.* Center for Population Economics, University of Chicago Graduate School of Business, and Department of Economics, Brigham Young University.

Fogel, Robert W. (2004). "Changes in the Process of Aging During the Twentieth Century: Findings and Procedures of the Early Indicators Project." *Population and Development Review.* 30: 19-47.

Forestall, Richard L. (1996). *Population of States and Counties of the United States: 1790-1990* Washington, DC: US Department of Commerce, Bureau of the Census.

Galenson, David W. and Clayne L. Pope (1989). "Economic and Geographic Mobility on the Farming Frontier: Evidence from Appanoose County, Iowa, 1850-1870." *Journal of Economic History.* 49: 635-656.

Gallaway, Lowell and Richard K. Vedder (1971). "Mobility of Native Americans." *Journal of Economic History* 31(3): 613-649.

Glasson, William H. (1900). *History of Military Pension Legislation in the United States.* Doctoral Dissertation. New York: Columbia University.

Glasson, William H. (1918). *Federal Military Pensions in the United States.* New York: Oxford University Press.

Goldin, Claudia (1995). "Career and Family: College Women Look to the Past." NBER Working Paper, no. 5188.

—-. (1984). "The Historical Evolution of Female Earnings Functions and Occupations." *Explorations in Economic History.* 21: 1-27.

—-. (1977). "Female Labor Force Participation: The Origin of Black and White Differences, 1870 and 1880." *Journal of Economic History.* 37(1): 87-108.

Goldin, Claudia and Lawrence Katz (2002). "The Power of the Pill: Oral Contraceptives and Women's Career and Marriage Decisions." *Journal of Political Economy.* 110(4):730-770.

Gould, Eric D. and M. Daniele Paserman (2003). "Waiting for Mr. Right: Rising Inequality and Declining Marriage Rates." *Journal of Urban Economics.* 53(2): 257-281.

Grogger, Jeffrey and Gordon H. Hansen (2008). "Income Maximization and the Selection and Sorting of International Migrants." NBER Working Paper no. 13821.

Goodrich, Carter and Sol Davison (1935). "The Wage-Earner in the Westward Movement I." *Political Science Quarterly.* 50(2): 161-185.

Goodrich, Carter and Sol Davison. (1936). "The Wage-Earner in the Westward Movement II." *Political Science Quarterly.* 51(1): 61-116.

Hacker, J.David (2008). "Economic, Demographic, and Anthropometric Correlates of First Marriage in the Mid-Nineteenth-Century United States." *Social Science History.* 32(3): 307-345.

Haines, Michael R. (1996). "Long-Term Marriage Patterns in the United States from Colonial Times to the Present." *History of the Family.* 1(1): 15-39.

Haines, Michael R., and Inter-university Consortium for Political and Social Research (2010). Historical, Demographic, Economic, and Social Data: The United States, 1790-2002 [Computer file]. ICPSR02896-v3. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2010-05-21. doi:10.3886/ICPSR02896

Haines, Michael R. and J. David Hacker (2006). "The Puzzle of the Antebellum Fertility Decline in the United States: New Evidence and Reconsideration." NBER Working Paper, no. 12571.

Hall, Patricia K. and Steven Ruggles (2004). " 'Restless in the Midst of their Prosperity': New Evidence on the Internal Migration of Americans, 1850-2000." *Journal of American History.* 91(3): 829-846.

Heckman, James J. and Burton Singer (1984). "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data." *Econometrica.* 52(2): 271-320.

Hirsch, Adam J. (2009). "Inheritance: United States Law." *Oxford International Encyclopedia of Legal History.* Ed. Stanley N. Katz. New York: Oxford University Press, 235-240.

Keeley, Michael C. (1977). "The Economics of Family Formation." *Economic Inquiry.* 15(2): 238-250.

Kim, Sukkoo (1998). "Economic Integration and Convergence: U.S. Regions, 1840-1987." *Journal of Economic History.* 58(3): 659-683.

Laibson, David I. (1997). "Golden Eggs and Hyperbolic Discounting." *Quarterly Journal of Economics*, 112(2), pp. 443-477.

Lancaster, Tony (1990). *The Econometric Analysis of Transition Data.* Cambridge: Cambridge University Press.

Lebergott, Stanley (1964). *Manpower and Economic Growth: The American Record since 1800.* New York: McGraw-Hill.

Legislature of New York (1837). "Excerpt from an Act Concerning Executors, Administrators, Guardians, Wards &c., Passed May 16, 1837." *Case Files of Approved Pension Applications of Widows and Other Dependents of Civil War Veterans, ca. 1861 - ca. 1910.* Records of the Department of Veterans' Affairs, Record Group 15. National Archives Building, Washington, D.C. Digital Images, "Civil War Widows' Pensions," fold3.com.

Linares, Claudia (2001). "The Civil War Pension Law." Chicago: Center for Population Economics Working Paper no. 2001-6.

Long, Jason, and Joseph Ferrie (2005). "A Tale of Two Labor Markets: Intergenerational Occupational Mobility in Britain and the U.S. Since 1850." NBER Working Paper no. 11253.

Loughran, David (2002). "The Effect of Male Wage Inequality on Female Age at First Marriage." *Review of Economics and Statistics.* 84(2): 237-250.

Maestas, Nicole, Kathleen J. Mullen and Alexander Strand (2011). "Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt." RAND Working Paper WR-853-2.

Margo, Robert A. (2000). *Wages and Labor Markets in the United States, 1820-1860.* Chicago: University of Chicago Press.

—. (1999). “Regional Wage Gaps and the Settlement of the Midwest.” *Explorations in Economic History.* 36: 128-143.

Monohan, Thomas P. (1951). *The Pattern of Age at Marriage in the United States.* Vol. 2. Philadelphia: Stephenson Brothers.

Mortenson, Dale T. (1986). “Job Search and Labor Market Analysis.” *Handbook of Labor Economics.* Ed. O. Ashenfelter and R. Layard. Amsterdam: North Holland, 849-920.

Nam, C.B. and M. Boyd (2004). “Occupational Status in 2000: Over a Century of Census-based Measurement.” *Population Research and Policy Review.* 23: 327-358.

O’Donoghue, Ted and Matthew Rabin (1999). “Doing It Now or Later.” *American Economic Review*, 89(1), pp. 103-124.

Oliver, John William (1917). “History of the Civil War Military Pensions.” *Bulletin of the University of Wisconsin Historical Series.* 4(844): 1-120.

Olivetti, Claudia and M. Daniele Paserman (2012). “In the Name of the Father: Marriage and Intergenerational Mobility in the United States: 1850-1930.” Boston University.

Phelps, Edmund S. and Robert A. Pollak (1968). “On Second-Best National Saving and Game-Equilibrium Growth.” *Review of Economic Studies*, 35(2), pp. 185-199.

Preston, Samuel H. and Michael Haines (1991). *Fatal Years: Child Mortality in Late Nineteenth Century America.* Princeton: Princeton University Press.

Preston, Samuel H. and John McDonald (1979). “The Incidence of Divorce Within Cohorts of American Marriages Contracted Since the Civil War.” *Demography.* 16(1): 1-25.

Price-Bonham, Sharon and Jack O. Balswick (1980). “The Noninstitutions: Divorce, Desertion, and Remarriage.” *Journal of Marriage and the Family.* 42(4): 959-972.

Rogerson, Richard, Robert Shimer and Randall Wright (2005). “Search-Theoretic Models of the Labor Market: A Survey.” *Journal of Economic Literature.* 43(4): 959-988.

Rosenbloom, Joshua (2002). *Looking for Work, Searching for Workers: American Labor Market during Industrialization.* Cambridge: Cambridge University Press.

—. (1996). "Was There a National Labor Market at the End of the Nineteenth Century? *Journal of Economic History.* 56(3): 626-656.

Rosenbloom, Joshua and William Sundstrom (2003). "The Decline and Rise of Interstate Migration in the United States: Evidence From the IPUMS, 1850-1990." NBER Working Paper no. 9857.

Rosenzweig, Mark R. (1988). "Risk, Private Information, and the Family." *American Economic Review.* 78(2): 245-250.

—. (1999). "Welfare, Marital Prospects, and Nonmarital Childbearing." *Journal of Political Economy.* 107(56): 3-32.

Rosenzweig, Mark R. and Oded Stark (1989). "Consumption Smoothing, Migration, and Marriage: Evidence from Rural India." *Journal of Political Economy.* 97: 905-26.

Rothenberg, Winifred B. (1988). "The Emergence of Farm Labor Markets and the Transformation of the Rural Economy: Massachussetts 1750-1855." *Journal of Economic History.* 48(3): 537-566.

Ruggles, Steven (1997). "The Rise of Divorce and Separation in the United States.

Demography. 34(4): 455-466.

Ruggles, Steven J., Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek. *Integrated Public Use Microdata Series: Version 5.0* [Machine-readable database]. Minneapolis: University of Minnesota, 2010.

Schaefer, Donald F. (1985). "A Statistical Profile of Frontier and New South Migration: 1850-1860," *Agricultural History* 59(4): 563-578.

Siegel, P.M. (1971). "Prestige in the American Occupational Structure." Ph.D. dissertation, University of Chicago.

Skocpol, Theda (1993). "Americas First Social Security System: The Expansion of Benefits for Civil War Veterans." *Political Science Quarterly.* 108(1): 85-116.

Skocpol, Theda (1995). *Protecting Soldiers and Mothers: The Political Origins of*

Social Policy in the United States. Cambridge: Harvard University Press.

Solon, Gary R. (2013). "Theoretical Models of Inequality Transmission Across Multiple Generations." NBER Working Paper no. 18790.

—. (1999). "Intergenerational Mobility in the Labour Market," in (O. Ashenfelter and D. Card, eds.), *Handbook of Labor Economics*, Volume 3A, pp. 1761-800, Amsterdam: North-Holland.

Song, Chen (2000). "Filing for the Union Army Pension: A Summary from Historical Evidence." Unpublished Paper, Center for Population Economics, University of Chicago.

Steckel, Richard H. (1983). "The Economic Foundations of East-West Migration during the 19th Century." *Explorations in Economic History* 20: 14-36.

Stewart, James I. (2003). "Migration to the Agricultural Frontier and Economic Mobility, 1860-1880." Paper presented at 42nd Cliometrics Conference, North Carolina State University.

United States Pension Bureau (1883). *Annual Report of the Commissioner of Pensions to the Secretary of the Interior for the Year Ending June 30, 1883.* Washington: Government Printing Office.

—. (1872). *Annual Report of the Commissioner of Pensions to the Secretary of the Interior for the Year Ending June 30, 1872.* Washington: Government Printing Office.

—. (1864). *Annual Report of the Commissioner of Pensions to the Secretary of the Interior for the Year Ending June 30, 1864.* Washington: Government Printing Office.

—. (1997). "Association Measures for Durations in Bivariate Hazard Rate Models." *Journal of Econometrics.* 79(2): 221-245.

Van Ours, Jan C. and Milan Vodopivec (2008). "Does Reducing Unemployment Insurance Generosity Reduce Job Match Quality?" *Journal of Public Economics.* 92: 684-695.

Waite, Linda J. and Glenna D. Spitze (1981). "Young Women's Transition into Marriage." *Demography.* 18(4): 681-694.

Wanamaker, Marianne (2012). "Industrialization and Fertility in the Nineteenth Century: Evidence from South Carolina." *Journal of Economic History.* 72(1): 168-196.

Weiss, Yoram (1997). "The Formation and Dissolution of Families: Why Marry? Who Marries Whom? And What Happens Upon Divorce?" *Handbook of Population and Family Economics.* 1: 81-123.

Weiss, Yoram and Robert J. Willis (1997). "Match Quality, New Information and Marital Dissolution." *Journal of Labor Economics.* 15(1): S293-329.

Widow's Certificate No. 8,336. Jonathan Timberman, Private, Compay I, 12th New York Infantry. *Case Files of Approved Pension Applications of Widows and Other Dependents of Civil War Veterans, ca. 1861 - ca. 1910.* Records of the Department of Veterans' Affairs, Record Group 15. National Archives Building, Washington, D.C. Digital Images, "Civil War Widows' Pensions," fold3.com.

Widow's Certificate No. 73,022. Henry Havens, Private, Company A, 141st New York Infantry. —.

Widow's Certificate No. 35,292. Henry Detrich, Private, Company K, 111th New York Infantry. —-.

Widow's Certificate No. 35,292. Henry Detrich, Private, Company K, 111th New York Infantry. *Case Files of Approved Pension Applications of Widows and Other Dependents of Civil War Veterans, ca. 1861 - ca. 1910.* Records of the Department of Veterans' Affairs, Record Group 15. National Archives Building, Washington, D.C.

Widow's Certificate No. 6,916. William Matthews, Private, Company C, 98th New York Infantry. —-.

Widow's Certificate No. 23,539. Andrew Van Buren, Private, Company E, 100th New York Infantry. —-.

Curriculum Vitae

[redacted]

[redacted]