Volume Title: Health and Labor Force Participation over the Life Cycle: Evidence from the Past

Volume Author/Editor: Dora L. Costa, editor

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-11618-2

Volume URL: http://www.nber.org/books/cost03-1

Conference Date: February 2-3, 2001

Publication Date: January 2003

Title: Internal Migration, Return Migration, and Mortality. Evidence from Panel Data on Union Army Veterans

Author: Mario Sanchez

URL: http://www.nber.org/chapters/c9633

8

Internal Migration, Return Migration, and Mortality
Evidence from Panel Data on Union Army Veterans

Mario A. Sánchez

8.1 Introduction

The United States has traditionally had high rates of internal migration. In 1860, 27 percent of native-born Americans were living outside their state of birth. This measure of mobility decreased throughout the second half of the nineteenth century, reaching a low of 20 percent in 1900. Since then, this figure has increased and in 1990 was 31 percent.[1] Although we know that Americans were mobile, there are still many unanswered questions about nineteenth century migration. Who moved in nineteenth century America and how often? Was this move temporary or permanent? What factors determined whether this move was temporary or permanent? How costly was this move?

This paper examines the characteristics of intercounty migrants and estimates the hazard rate of changing county of residence within a year. It investigates whether return migration was common and the characteristics of return migrants. Finally, it examines the costs of migration, in terms of mortality. The paper is novel on two grounds. First, it uses a large longitudinal data set of residential histories for Union Army veterans, allowing me to investigate not just the migration decision through a richer specification than

At the time this research was conducted, Mario A. Sánchez was a Ph.D. candidate in economics at the University of Chicago, and a graduate research assistant at the Center for Population Economics, University of Chicago. He is currently affiliated with the Inter-American Development Bank.

I thank Joseph Ferrie, Steven Levitt, Robert Margo, Richard Steckel, and Peter Viechnicki for valuable suggestions. I am especially indebted to Dora Costa and Robert Fogel for their comments and guidance. Suggestions from participants of the Center for Population Economics weekly meetings are gratefully acknowledged. I am fully responsible for all errors.

1. Figures computed from the 1 percent 1860–1990 Integrated Public Use Microdata Samples (Ruggles & Sobek 1997).

previous researchers have been able to use, but also the return migration decision, which may be workers' optimal reaction to temporary economic shocks. Second, while previous research has mainly used aggregated or cross-sectional data, this paper uses these longitudinal microdata to study the relationship between migration and life expectancy. Because migration, particularly to urban areas, may have decreased the life expectancy of workers, a "mortality wage premium" may partly account for wage differentials between cities and rural areas.

8.1.1 Previous Literature: Mobility

Data availability has stymied past research on migration in nineteenth-century America. Previous researchers relied on aggregate decennial census data or data for small communities. For example, Kuznets and Thomas (1957) limited themselves to using census survival techniques to produce net migration estimates at the state level for the 1870–1950 period.[2] Galloway and Vedder (1971) studied the effect of such state characteristics as per capita income, distance between states, number of jobs available, land availability, and similarity of climate and culture between states on net interstate migration flows. Because they had to define migrants as those whose state of residence at the time of the census differed from their state of birth, the time span over which migration could have occurred is very long, and death and undercounting bias their estimates.

Steckel (1989) was the first to use longitudinal microdata to present national estimates of mobility, geographical distribution, and distance traveled by migrants. He linked 1,600 families from the 1860 to the 1850 census, using the state of birth of children older than ten as a pointer to the residence of the family in 1850. Observing migration at the county level, and knowing that the movement occurred over a ten-year interval, he studied the effects of individual and household characteristics on the propensity to migrate, as well as on the distance and direction of the movement. Ferrie (1996) complemented Steckel's work by linking forward 4,938 families from the 1850 Public Use Sample of the Federal Census of Population to the 1860 federal census manuscripts. Using this data set, Ferrie (1999) analyzed the causes and consequences of migration to small towns and cities, and reexamined (Ferrie 1997) the theory that "the frontier" (90 degrees west longitude) served as a safety valve relieving pressure on urban labor markets in the east (Turner 1920).[3]

2. For a revision of Kuznets and Thomas's estimates using the state of birth–state of residence technique, see Lebergot (1970).

3. Steckel acknowledged that because of the "backwards" linkage technique, "the sample may not be representative of all families that had children, [and therefore] one must be cautious in interpreting the results. Inferences may apply only to that portion of the families" (191). Ferrie is in the process of linking approximately 7,000 families from the 1860 to the 1870 federal census manuscripts.

The outcome of these efforts to overcome data availability barriers is a more reliable picture of how mobile nineteenth-century Americans were, how far and where they decided to move, and the roles that national, regional, and individual characteristics played in shaping these decisions. However, the problems arising from making use of census manuscripts to assess how mobile Americans were still persist. In particular, mortality and census underenumeration, return migration, and the inability to observe more than one move within the ten-year interval potentially undermine the inferences made using samples linked across censuses.

This paper introduces the Union Army Migration Data set (UAMD), a panel data set consisting of postbellum residential histories for 17,017 Union Army veterans, with the intention of pushing forward this "data frontier." The Pension Board gathered the data from the recruit or his family, and they contain rich socioeconomic information on the recruit, his household, and his parents' household. Beyond serving as a basis to study the size of the bias of estimates of mobility for samples linked across decennial censuses of population, the use of UAMD should enhance our understanding of the circumstances under which these migration processes occurred.

8.1.2 Previous Literature: Migration and Life Expectancy

East-to-west migration and the movement from rural to urban areas were the two main sources that shaped geographical redistribution during the nineteenth century. As the century progressed, rural-to-urban migration became increasingly important (Haines 2000). Because urban areas had such high death rates from poor sanitation and overcrowding (Fogel 1986; Williamson 1990), it is reasonable to expect a decline in the life expectancy of movers to urban areas compared to individuals with similar characteristics who remained in or migrated to rural places. Curtin (1989, xiii) discussed this added cost to migration for the case of European soldiers migrating to other continents:

> From the beginning of European trade and conquest overseas, Europeans knew that strange "climates" could have fatal effects. Later, they came to understand that it was disease, not climate, that killed, but the fact remained that every trading voyage, every military expedition beyond Europe, had its price in European lives lost. For European soldiers in the tropics at the beginning of the nineteenth century, this added cost in deaths from disease—the "relocation cost"—meant a death rate at least twice that of soldiers who stayed home, and possibly much higher.

O'Rourke, Williamson, and Hatton (1994) presented evidence on the existence of an urban disamenities wage premium to migration. They studied the effect of such disamenities as population density, town size, and infant mortality on unskilled wage rates across British towns during 1834 and

1905. They concluded that these disamenities explained a good deal of the "bribe" that rural migrants had to be paid in order to move to urban places. They argued that high migration rates to cities indicate that workers were disposed to pay the disamenity costs of migration.

Fogel (1986) argued that both internal migration and immigration played important roles in explaining the sharp decline in life expectancy experienced by native-born Americans from the 1790s to the mid-nineteenth century, both in urban and rural places. Fogel referred to research that concluded that mortality rates were much higher in the immigrant wards than in the wards in which the native-born were preponderant, and that epidemics often began first in the foreign-born wards. Internal migration may have increased the spread of cholera, typhoid, typhus, malaria, dysentery, and other major killer diseases of the era in rural areas.[4]

The UAMD represents a unique opportunity to analyze the effect of the decision to migrate on the life expectancy of migrants. Its longitudinal structure allows me to estimate the impact of the migrant status on the length of the veteran's life span. This paper also investigates the effect of the characteristics of places of origin and destination on the number of years that recruits lived after the war and their causes of death. This will allow me to determine whether there was a mortality wage premium to migration.

8.2 The Data

The Center for Population Economics (CPE) at the University of Chicago collected military, socioeconomic, and medical information for 35,747 white males mustered into the Union Army during the Civil War.[5] This data set, collected as part of the Early Indicators of Later Work Levels, Disease, and Death (EI) project, contains information on the recruit, his household, and his parental household at numerous times during the recruit's life. They gathered the data by using the recruit's military-related information to link him to the 1850, 1860, 1900, and 1910 censuses of population[6] and to his pension records (PEN) and the medical examinations often included in them, and by using other historical documents as sources for ecological variables.

For those who survived the war and applied for a pension (either personally or through their legal heirs), EI is a rich source of residential information from the end of the war until the recruit's death. Postbellum residential information is abundant throughout PEN on pension claims, affidavits, correspondence with the Pension Board, envelopes retained in the pension

4. Fogel proposed an alternative approach to the bribery principle to compute the "mortality correction" to real wages due to a more rapid consumption of human capital in locations where disease was more prevalent.

5. For a discussion of sample design, data sources, and methodology see CPE (2000).

6. Linkage to the 1870 and 1880 censuses is in progress.

file, and vital statistics forms. These documents not only reveal the location of the recruit at the issuing time of the document, but often contain retrospective residential information provided by the recruit himself or, after his death, by his dependents. PEN also contains information on the date and location of a change in the recruit's marital status, of the birth of the recruit's children, and of the recruits' death.[7]

I combined the information in the pension documents to reconstruct the residential life history of the recruit after the Civil War; the final data set that I created is the UAMD. For a recruit linked to PEN, there are up to sixteen documents containing residential information, including death certificates or other communications to the Pension Board regarding the recruit's death. Although in some cases the city, county, and state of residence are recorded, it is more common to see the recruit at either the county or the city level. Because city and state of residence usually imply a unique county of residence, and not the other way around, UAMD contains residential histories recreated at the county level.

Of the 30,763 recruits in EI who survived the war, 20,674 provided information to the Pension Board. For 17,779 of these, information on residence and on dates of birth and death are provided. I omitted observations on recruits who were not born between 1820 and 1845 both to exclude unreasonable military ages and to exclude recruits in small cohorts. The final sample size of UAMD is 17,017.[8]

The next two subsections deal with the representativeness of UAMD and discuss the quality of the data.

8.2.1 Representativeness of the Sample

Fogel (1993) and Costa (1998) argued that white soldiers in EI constituted a representative cross section of their generation. A considerable proportion of white males of military age participated in the war[9] and soldiers' socioeconomic characteristics, life expectancies, and distribution of causes of death resembled those of the nonfighting population. As discussed in ap-

7. The following example of a pension document may clarify the type of information contained in PEN. On 16 April 1907 private James Dean declared in his first Declaration for Pension: "that his several places of residence since leaving the service have been as follows: Windsor, Conn., 1 year and 4 months; enlisted in regular service Nov. 21, 1867; returned to Windsor on December, 1870; reentered regular service March 21, 1871; discharged in 1876 and located in Bridgeport, Conn. until August, 1893; moved to Wallingford, Conn.; resided there until June 1885, returned to Bridgeport; have continued residing there until the present time." Thus, it is possible to reconstruct completely the residential life of a recruit up to the time when he or his heirs presented a pension application (or any other pension document), regardless of when and how many times they provided information.

8. See appendix A for a detailed discussion on the construction of UAMD.

9. From information provided by Dyer (1959), I computed that 40 percent of all Northern white men born between 1820 and 1845, 58 percent of those born between 1835 and 1845, and 80 percent of those who were born between 1840 and 1845 served in the Union Army during the Civil War.

pendix B in this chapter, being wounded in the war did not adversely affect soldiers' ability to move their place of residence. Mobility rates among veterans and nonveterans were similar. Sixty-three percent of Union veterans who survived to 1910 were living in a state other than their state of birth, compared to 67 percent of nonveterans who were either native-born or who, if foreign-born, migrated before the end of the war (figures computed from the 1 percent 1990 Integrated Public Use Microdata Sample).

Postbellum migration information is available only for those veterans in EI who applied directly for a pension or whose dependents applied for a pension.[10] However, while individual characteristics and military outcomes explained linkage to PEN, the most important factor explaining linkage failure was whether the recruit was dishonorably discharged (see appendix B). Because 90 percent of all soldiers were honorably discharged, the population in UAMD represents a large fraction of all soldiers.

8.2.2 Residential Information in UAMD

Figure 8.1 shows the number of recruits whose counties of residence are known from the end of the war until 1920[11] as a proportion of the total number of recruits alive at each year. Residential information is more abundant for those recruits who survived to later years, perhaps because as the number of widows' applications increased, so did the importance of collecting information to verify their claims. Because pretended widowhood was the most common way of filing a fraudulent claim (Glasson 1918), the recruit's residential history, as well as dates and places of birth of the veteran's children, were used to verify the validity of a widow's pension claim.

Figure 8.1 suggests a relationship between different pension regimes and the number of veterans with residential information for a particular year. I estimated a linear regression to explain the proportion of life after the war for which there is residential information for the recruit.[12] The variable with the most explanatory power was whether the veteran (or a dependent) submitted an application after 1907 (to have done so increases the proportion of life after the war with residential information by nearly 30 percent). Beginning in 1904, all pension forms (including widows' applications) explicitly asked for retrospective residential information. Figure 8.1 can therefore be explained by the evolution of the administrative procedures of the Pensions Board, rather than the behavior of the recruit seeking a pension.

10. Not all recruits linked to PEN entered the pension rolls. The recruits' legal heirs may have been the ones filing applications, and the Pension Board may have never accepted a claim as valid.

11. I chose 1920 as the last year of analysis, given that after that year residential information becomes increasingly scarce and the number of survivor veterans decreases rapidly.

12. The independent variables were as follows: number of applications recorded for the veteran, a set of dummy variables for pension regimes, age after the war, marital status, and an interaction term between admittance and year of entrance into the pension rolls.

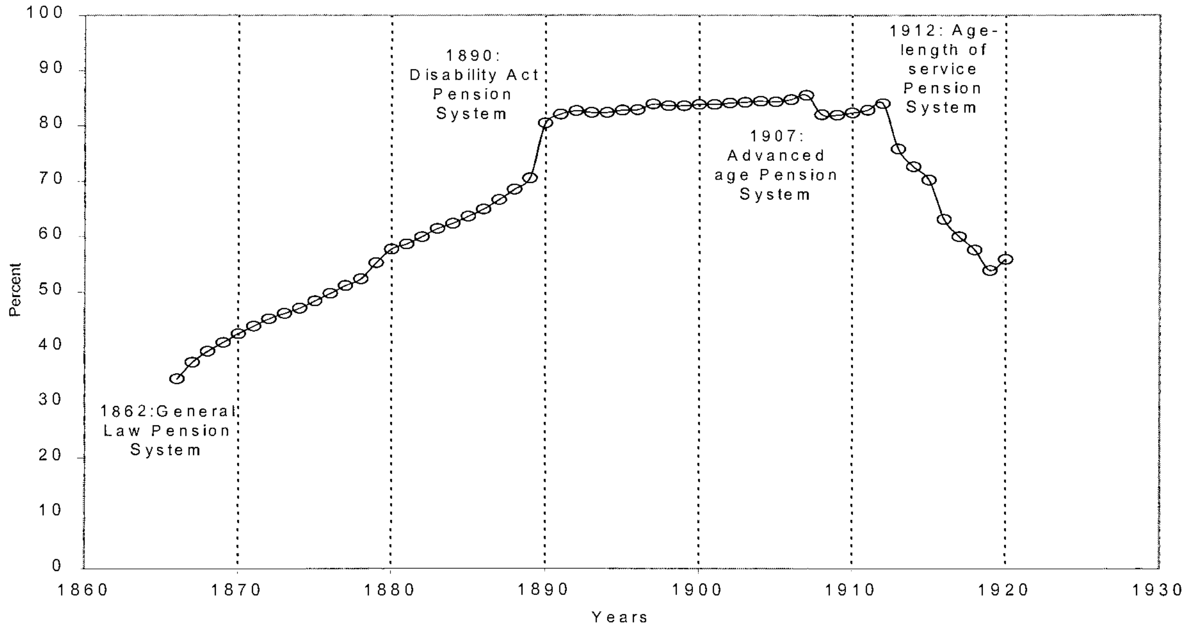

Fig. 8.1 Proportion of living recruits with residential information, 1866–1920

8.3 Empirical Framework and Methods

Models that serve as microfoundations to most of the research on migration can be classified in three broad categories: (a) models that assume that workers maximize their own expected discounted utility by choosing a geographical location and by investing in their capital, both human and physical (Sjaastad 1962; Todaro 1969); (b) models in which the decision units are not isolated individuals, but larger groups like couples, families, or households (Borjas 1994; Stark and Bloom 1985; Stark 1991); and (c) social networks models in which family and friends of the potential migrant, who have already established their residence at the location of destination, affect the migration decision by increasing the expected net returns of migration (Hugo 1981; Taylor 1986; Gurak and Caces 1992). The following is a summary of the implications of these three types of model that are relevant to this paper:

1. Migrants should be younger than the nonmigrant population because their longer life expectancy increases the discounted present value of their investment and because their physical costs of moving are lower. Migration of the young, however, may not be the best strategy for a household decisionmaker. For instance, a farm family faced with poor harvests may send its older, stronger son to work at another location, while the younger children stay to help with less strength-demanding tasks.
2. Married workers and those with children should move less often than single, childless ones because their moving costs are higher. However, at older ages, parents may move upon retirement to live with their children.
3. If there are differences in land availability or in location-specific economic conditions, people residing in less advantaged regions will be more likely to move to maximize their expected discounted wealth.
4. The sign of the correlation between wealth and the probability of changing locations is uncertain. Although relatively poorer individuals face a lower opportunity cost of migrating, financial constraints could prevent them from moving. On the other hand, a financially constrained household—or group—may decide to send one of its members to another location to accumulate assets.
5. Given the high correlation between wealth and type of job, it is not clear what the occupational distribution across the migrant population should be. After controlling for wealth, however, workers that depend more on their location-specific human or social capital should be less inclined to move out of their labor markets.
6. Healthier individuals endure better the physical and mental costs of migration. Thus, migrants could be positively self-selected in terms of their health. Migrants may also be healthier than nonmigrants because wage differentials across locations may be positively correlated with health. However, health itself may be a motive for migration. Workers and their families

may change their place of residence to escape from disease, and they may already be in poor health before moving. In this case, migrants would be negatively self-selected in terms of their health.

7. Individuals who have already experienced migration may be more prone to change locations, as they may belong to a social network that extends beyond their current place of residence. Lalonde and Topel (1997), for instance, showed that foreign-born immigrants tend to reside in ethnic enclaves, and that the locations that new waves of immigrants choose as their destination display a high concentration of ethnically similar individuals.

8.3.1 Estimation: Moving across Counties

Exploiting the yearly panel nature of the data, this paper studies the self-selection of migrants in terms of their observable characteristics—some of them time-varying—and their propensity to migrate over the life cycle, where the movements could have occurred within one year.

The probability that person i moves at the end of period t is modeled as a function of his possibly time-varying characteristics $X_{i,t}$, the characteristics of the location where i resides in period t, $Z_{i,t}$, and an error term $\mu_{i,t}$ that varies across individuals and time. Let $M_{i,t}$ be an indicator variable that takes a value of 1 if person i changes county of residence at t and a value of zero otherwise. Thus, this paper models the probability that at time t a person at risk of changing his place of residence will do so within the next year as

$$\text{(1)} \qquad \Pr(M_{i,t} = 1) = F(X_{i,t}, Z_{i,t}, \mu_{i,t}).$$

For empirical analysis, equation (1) takes the following form:

$$\text{(2)} \qquad P_{i,t} = \Pr(M_{i,t} = 1) = \Lambda(\alpha + \beta' X_{i,t} + \gamma' Z_{i,t} + \delta_t + \varepsilon_i)$$

where Λ represents the logistic cumulative distribution function and $\mu_{i,t}$ takes the form $\delta_t + \varepsilon_i$, where ε_i is a Gaussian term. The linearized version of equation (2) makes it easier to interpret the time coefficient δ_t:

$$\text{(3)} \qquad \log\left(\frac{P_{i,t}}{1 - P_{i,t}}\right) = \alpha + \beta' X_{i,t} + \gamma' Z_{i,t} + \delta_t + \varepsilon_i$$

The expected difference in the logarithm of the odds ratio of moving for two different persons i, j seen at times t and t' (respectively) is equal to

$$\text{(4)} \qquad E\left[\log\left(\frac{P_{i,t}}{1 - P_{i,t}}\right) - \log\left(\frac{P_{j,t'}}{1 - P_{j,t'}}\right)\right] = \beta'(X_{i,t} - X_{j,t'}) + \gamma'(Z_{i,t} - Z_{j,t'}) + (\delta_t - \delta_{t'}).$$

If i and j shared the same individual and location characteristics at t and t' (respectively); the expected difference in their odds ratio would simply be $\delta_t - \delta_{t'}$: the difference in the logarithm of the odds of moving for two individuals who decide to migrate at different points in time, but are otherwise identical.

Years with missing residential information represent a complication when estimating the mobility of the UAMD population. Individuals with longer periods without location information recorded are at lower risk *to be seen* moving. Thus, defining as migrant someone with a recorded movement would produce a downward-biased estimate of the mobility of the sample.[13] To overcome this problem, I estimate the probability that the county of residence differs within two consecutive years, using for the estimation only the subsample of recruits who have nonmissing information for those two years. In other words, I impute the migration decision of an unobserved individual through the decision taken by a synthetically identical individual. This procedure implicitly assumes that after controlling for observable characteristics, there is no correlation between the probability of having residential information and the probability of changing locations for every pair of contiguous years.

8.3.2 Estimation: Return Migration

I examine return migration by estimating a logistic regression, where the dependent variable is a dichotomous indicator, taking a value of 1 if the recruit is a return migrant and a value of zero otherwise. The subsample under study is all veterans in UAMD who changed locations at least once, and who never left the country.

Let t_i represent the ith year at which a new county of residence is observed for the veteran after the end of the war, and let l_i be the location at time t_i. Thus, the requirement for a veteran to be included in the analysis is that a location l_2 exists. Using this notation, a veteran is a return migrant if there exists a time t_i^*, with $i \geq 3$, such that l_i^* is equal to $l_{i'}$ for $i' < i$. Let R be an indicator function that takes a value of 1 if t_i^* exists, and a value of zero otherwise. Let us define the probability of becoming a return migrant as

$$(5) \qquad \Pr(R = 1) \equiv \Pr(\exists l_i^* \; s.t. \; l_i^* = l_{i'} \text{ for } i' < i) = \Lambda(\alpha + \beta' \boldsymbol{X}_2 + \mu),$$

where $\boldsymbol{X}_2$ is a row vector of individual characteristics of the recruit at the time of his first movement, and μ is a Gaussian term. Thus, in this paper, a return migrant is an individual with at least three different recorded locations and who returned to a location where he previously lived.[14]

Census-based estimators of mobility rates use the working assumption

13. Although 90 percent of the recruits in UAMD do not show "information gaps" larger than ten years, there is no available information on the literature that could give an idea on the size of the bias if I ignored this issue.

14. Some veterans returned to more than one location. However, this would be just a special case of equation (5), where the set of t_i^* such that $R = 1$ is not a singleton. Notice as well that this definition may underestimate the proportion of veterans who were returning migrants, as some of these return movements may not be observed.

that a person observed at the same location in two consecutive censuses did not move within the ten-year interval. Beyond acknowledging this caveat in their estimates, researchers have been unable to do much to measure the size of the return migration bias. Thanks to the panel nature of UAMD, it is possible, for the first time, to investigate the importance of return migration, and to describe the characteristics of the return migrants.[15]

8.3.3 Estimation: Migration and Life Expectancy

I study the effect of migrant status on life expectancy by estimating a time-varying covariates hazard model of mortality. Time until death is a function of a set $X_{i,t}$ of individual characteristics at time t, and a time-varying dichotomous variable $M_{i,t}$, which takes a value of zero if the recruit has not become a migrant by time t and a value of 1 otherwise. Let t^* be the year at which a veteran changes locations for the first time. Then, $M_{i,t} = 0$ for $t < t^*$, and $M_{i,t} = 1$ for $t \geq t^*$. For recruits who never acquire the status of migrant, $M_{i,t} = 0$ for all t.

It is important to model migrant status as a time-varying variable instead of just including a dummy variable M_i indicating whether the recruit ever changed locations. The coefficient on M_i is deceptive because recruits who died earlier were less likely to become migrants. A time-varying dummy, on the other hand, allows me to compare waiting times (the number of years that elapsed from the first year with information for the recruit to the year in which he actually moved) for recruits who were at risk of dying, with survival times for recruits who experienced events.

Following Cox (1984),[16] I model the hazard rate of dying as a function of time (represented by a fixed baseline function) and a function that depends on the covariates. Let T_i represent the year of i's death, and $h_i(t)$ be the probability i dies at $t + 1$ given that he has survived until t. Using this notation, the hazard rate of dying is

$$(6)\quad h_i(t) = \Pr\{T_i = \mathrm{t} + 1 | T_i \geq t\} \equiv h(t; X_{i,t}) = h_0(t)\exp\beta' X_{i,t} + \delta M_{i,t}).$$

Under this framework, this paper tests the hypothesis that once the recruit changed his county of residence at time t his hazard rate of dying increased with respect to other comparable recruits who had not moved at time t.

There is an additional complication in the estimation of equation (6) for the UAMD sample. I mentioned in section 8.1 that survival times are right-censored for those recruits who lived beyond 1920. The likelihood function produced by equation (6) must take censoring into account in order to get unbiased estimates. Fortunately, the partial-likelihood method developed

15. See appendix C for a discussion on the biases of census-based estimations of mobility.

16. The expression for and development of the maximum-likelihood function for a discrete-time Cox model with time-variant covariates are quite cumbersome. I refer the interested reader to read Cox (1984).

Table 8.1 Distribution of the Number of Times Recruits in UAMD Changed Locations after the War

Number of Times Recruit Changed Places of Residence	Number of Recruits	Percentage of Recruits
Never changed county of residence	10,015	58.9
Changed counties once	3,859	22.7
Changed counties twice	1,838	10.8
Changed counties three times	809	4.7
Changed counties four times	311	1.8
Changed counties five times or more	185	1.1
Total	17,017	

by Cox readily incorporates right-censoring, producing estimates that are consistent and asymptotically normal.[17]

I estimate a competing risks model to analyze how migration and the urban-rural status of both the location of origin and the hosting location affected the probability of dying of a particular type of disease. The procedure for estimating such models is simply to use time remaining to die from a particular type of disease as the dependent variable, while the life spans of people who died of other causes are treated as censored after the year of death. The only assumption behind this technique is that censoring is noninformative, that is, that conditional on the covariates, those who are at particularly high (or low) risk of dying of a particular type of disease are no more (or less) likely to die of any other type of disease.

8.4 Results

8.4.1 Results: Migration across Counties

Table 8.1 shows the distribution of the number of times recruits in UAMD changed county of residence. Approximately 41 percent of them migrated to another county at least once during the postbellum era.

Recall that I estimate mobility across counties using equation (3). Instead of running individual regressions (one for every pair of consecutive years), it is efficient to stack individual-time observations in a seemingly unrelated model (Greene 1997). Dummy variables for each year are included (omitting the dummy for 1867 to avoid perfect multicollinearity), and their slopes

17. Another issue in the Cox regression estimation is the existence of "tied data." The partial likelihood method assumes that it is possible to strictly order survival times. For the case of discrete data, it is common to find individuals who survived an equal number of years. In this case, it is not possible to strictly order the data. There are several methods proposed to handle such cases, and throughout this paper, I will be using a method called the *discrete method,* which was proposed by Cox (1984).

are interpreted as the effect of the state of external conditions during the year on the hazard rate of moving. Since residential information is missing for some individuals, I impute migration probabilities for synthetically identical individuals. This paper defines two individuals as synthetically identical if, for any given year, they were the same age, had the same marital status and number of children, lived in the same region, belonged to the same occupational group,[18] shared the same nativity, and had similar mobility and military pasts.

Table 8.2 shows the results for this stacked logistic regression. Younger, single, and rural recruits, who were foreign-born, had fewer children, resided in the Midwest, migrated before the war (as proxied by enlisting in a state different from state of birth), and had moved after the war[19] were at higher risk of moving next year. Surprisingly, occupational group does not significantly affect the propensity of veterans to move. Steckel (1989) found a similar result, and attributed it to the correlation between other variables in the regression and the occupational group.[20] Unexpectedly as well, people who were ill or wounded during the war were more mobile. It is possible that conditional on war survival, the average veteran's mobility was not seriously impeded.[21] Nevertheless, the reason the wounded were more mobile than the rest of the population remains as an interesting puzzle to be addressed in future research.

Figure 8.2 shows the hazard rate of movement across the life cycle, setting all time coefficients equal to zero (or equivalently assuming that veterans made all their moves—if any—in 1867) and all other variables at their means, and assuming that the mean veteran had not migrated during the postbellum period. Note that the propensity to move decreased continuously with age as the horizon over which individuals discounted the potential returns of moving decreased. Veterans were 62 percent more likely to migrate at age thirty compared to their propensity to move at age sixty.

How mobile were postbellum Americans? Let me rephrase this question as follows: How likely was a twenty-year-old recruit who lived until age seventy-five and who had the individual characteristics of the average veteran

18. I created the occupational-group dummy variables from the occupation of the recruit at enlistment time; therefore I am implicitly assuming that the veteran during his whole life kept the occupation he had at enlistment. It is possible to analyze the robustness of these results by linking UAMD to the several censuses of population available in EI. Moreover, for the segment of population linked to the censuses of population, wealth information is available and I could have incorporated it in the analysis for the linked sample. I did not do so, however, because that would have complicated the sample selection issues. See appendix A for a discussion on this matter.

19. As pointed out by Robert Margo in his referee report, the positive coefficients on migrant before and after the war could be evidence for state dependence or it could be evidence of a fixed effect.

20. However, he does not find a significant effect of family size.

21. In appendix B I make a case for this argument.

Table 8.2 Logistic Regression on the Probability of Moving Next Year, 1866–1920

Variable	Mean	Coefficient	Slope
Intercept		–2.53*	–0.0611
Age	52.51	–0.023*	
Age squared		0.00016**	
Married	0.8002	–0.3341*	–0.0082
Number of kids	2.7601	–0.0083**	–0.0002
Northeast region	0.2838	–0.4159*	–0.0100
South region	0.0557	–0.0869	–0.0021
West region	0.0445	–0.0823	–0.0022
Foreign-born	0.1679	0.1484*	0.0036
Migrant before the war	0.1699	0.0933***	0.0011
Migrant after the war	0.0477	0.5516*	0.0132
White-collar	0.0457	–0.0032	–0.0001
Manual worker	0.3227	–0.0121	–0.0003
Service worker	0.0˚193	0.1053	0.0027
Urban	0.2061	–0.1147*	–0.0028
Ill or wounded during the war	0.2910	0.1296*	0.0034
Prisoner of war	0.0709	0.0563	0.0014
Time enlisted	622.00	0.0001	0.0000
1868	0.0146	0.1786***	0.0044
1869	0.0152	0.0992	0.0024
1870	0.0157	0.0983	0.0024
1871	0.0162	0.0397	0.0010
1872	0.0166	0.0709	0.0017
1873	0.0169	0.0793	0.0019
1874	0.0173	–0.0020	0.0000
1875	0.0177	0.2021***	0.0049
1876	0.0182	0.0669	0.0016
1877	0.0187	0.1266***	0.0031
1878	0.0190	0.1648***	0.0040
1879	0.0199	0.2196**	0.0054
1880	0.0208	0.3046*	0.0074
1881	0.0211	0.1981***	0.0048
1882	0.0215	0.2048***	0.0050
1883	0.0219	0.2011***	0.0049
1884	0.0222	0.2233**	0.0054
1885	0.0225	0.2233**	0.0054
1886	0.0228	0.2364**	0.0057
1887	0.0232	0.1723**	0.0041
1888	0.0237	0.1529***	0.0038
1889	0.0241	0.1697***	0.0041
1890	0.0264	0.1308***	0.0032
1891	0.0270	0.0104	0.0003
1892	0.0271	0.0150	0.0004
1893	0.0268	0.0619	0.0015
1894	0.0265	0.0352	0.0009
1895	0.0260	–0.0512	–0.0012
1896	0.0257	0.1542***	0.0038
1897	0.0252	0.0190	0.0005
1898	0.0246	0.1050	0.0026

Table 8.2 (continued)

Variable	Mean	Coefficient	Slope
1899	0.0241	0.1399***	0.0034
1900	0.0235	0.3782*	0.0092
1901	0.0229	0.3148*	0.0077
1902	0.0223	0.2413*	0.0059
1903	0.0216	0.1235***	0.0030
1904	0.0209	0.2585*	0.0063
1905	0.0202	0.2645*	0.0064
1906	0.0195	0.1214***	0.0030
1907	0.0179	0.1111	0.0027
1908	0.0171	0.3167	0.0077
1909	0.0162	0.8597	0.0210
1910	0.0155	0.0454	0.0011
1911	0.0147	–0.1644	–0.0040
1912	0.0125	–0.0167	–0.0004
1913	0.0112	–0.5345*	0.0130
1914	0.0099	–0.1185*	–0.0029
1915	0.0082	–0.2367*	–0.0058
1916	0.0071	–0.5567*	–0.0136
1917	0.0062	–1.149*	–0.0280
1918	0.0052	–1.0477*	–0.0255
1919	0.0045	–1.3930*	–0.0340
Number of observations	371,577		
–2 log-likelihood	45,691		
Max-rescaled R^2	0.0186		

Notes: Omitted variables are single, Midwest region, farmer, native-born, enlisted in state of birth, has not moved after the war, not ill or wounded during the war, not being a POW, rural, and 1867.

***Significant at the 10 percent level.

**Significant at the 5 percent level.

*Significant at the 1 percent level.

to change his county of residence during 1867? Integrating under the curve depicted in figure 8.2, the answer is approximately 59 percent.[22] Previous researchers could not address this kind of question at that level of specificity.

8.4.2 Results: Return Migration

Why do we observe return migration? Workers may move temporarily if an economic shock leads to a wage gap across locations. Liquidity constraints are another explanation. For instance, an unemployed migrant may return to his previous residence once he exhausted the financial and social

22. This is the first time such a figure is computed for this time period, and thus there is no comparison point to address whether the figure is larger or lower than expected. However, this figure is informative if we compare it with the 30 percent ten-year-period mobility rates estimated by other researchers.

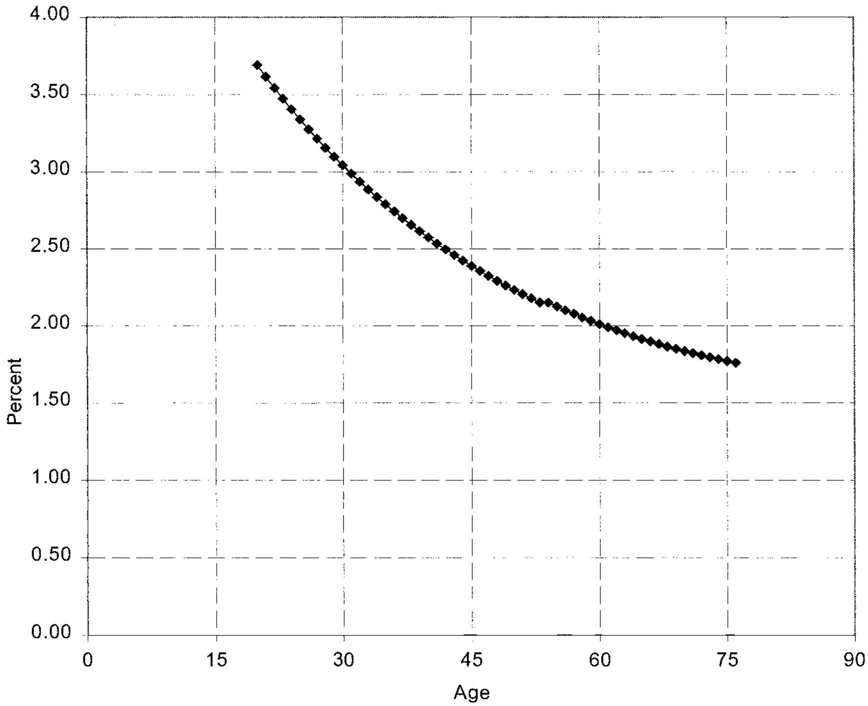

Fig. 8.2 Estimated probability of moving next year over the life cycle

Note: The graph is computed setting all time coefficients equal to zero and is computed at the sample mean for all other variables but age. The smoothness observed is the result of using a quadratic approximation.

capital that sustained him in the hosting location. Alternatively, he may accumulate assets while working at his new location and return home with these assets. Older migrants may retire to live with children who reside in a previous location. Finally, workers searching for the highest wages across locations may learn that they had already worked at the most attractive location available to them.

I study the characteristics of return migrants by examining recruits who moved at least once since the end of the Civil War until 1920 or their year of death, and who never left the country. Out of these 6,350 recruits, 992 (15.6 percent) returned to a previous county of residence, and the average time that elapsed for their return was 16.9 years.[23]

I estimate return migration probabilities using equation (5). The individual characteristics included as explanatory variables are age, marital status,

23. To provide an idea of the importance of return migration on census-based estimates of mobility, I computed the proportion of those return movements that occurred within a ten-year interval and found it to be equal to 3.5 percent (224 veterans), and significantly different from zero.

Table 8.3 **Logistic Regression on the Probability of Returning to a Previous Location**

Variable	Mean	Coefficient	Slope
Intercept	1.0000	0.8538*	0.1056
Age at first move	46.2814	–0.0348*	–0.0043
White-collar	0.0449	0.0368	0.0045
Manual worker	0.3041	0.1074	0.0133
Service worker	0.0236	0.1766	0.0218
Married at first move	0.7524	–0.2700*	–0.0334
Family size at first move	2.3387	0.0775*	0.0096
In Northeast after first move	0.1913	0.1421***	0.0176
In South after first move	0.0739	0.2898**	0.0358
In West after first move	0.0814	0.2396***	0.0296
In urban county after first move	0.1890	–0.0591	–0.0073
Logarithm of distance traveled	4.7421	–0.2309*	–0.0285
Number of observations	6,350		
–2 log-likelihood	2,646		
Max-rescaled R^2	0.0564		

Notes: Omitted variables are farmer and Midwest region.
***Significant at the 10 percent level.
**Significant at the 5 percent level.
*Significant at the 1 percent level.

and family size during the year the recruits first change county of residence. Additional explanatory variables are region and urban-rural status of the county in which they first lived as migrants, and the logarithm of the distance they traveled to their first new location.

Table 8.3 shows that return migrants were more likely to be younger, to be unmarried, to have bigger families, to have moved first to the South, and to have migrated only a short distance. As shown in the previous subsection, the young were more likely to migrate in the first place. The young are more financially constrained, face lower search costs, and have less information about various labor markets. If veterans left large families behind, then they may have been more likely to return to live with their children. Because migrating longer distances is more costly, those who traveled larger distances may have had characteristics that increased the likelihood of a successful first migration.

8.4.3 Results: Migration and Life Expectancy

Recall that I investigate the relationship between migration and life expectancy using the Cox regression model in equation (6). I restrict the sample to recruits with at least three years of residential information, with information on year of death, and who never left the country. The total sample size is 11,097.

Table 8.4 shows that once a recruit changed his county of residence, his

Table 8.4 Cox Regression on the Surviving Times of UAMD Veterans

Variable	Mean	Hazard Ratio	Pr > χ^2
Age at first observation	37.96	1.126	0.0001
Age at first observation squared		1.000	0.0001
White-collar	0.05	1.109	0.0357
Manual worker	0.34	1.106	0.0001
Service worker	0.02	1.118	0.1319
Northeast at first observation	0.32	1.077	0.0029
South at first observation	0.06	0.976	0.6094
West at first observation	0.03	0.957	0.5256
Urban at first observation	0.17	1.173	0.0001
Foreign-born	0.18	1.066	0.0215
Migrant status (time-varying)		1.083	0.0005
Number of observations	11,097		
Censored observations	23,223		
Model χ^2 (Wald)	6,103.7 (p-value = 0.0001)		

Notes: Omitted variables are farmer, Midwesterner at first observation, rural at first observation, and native-born.

probability of dying within the next year was 1.08 times higher than that of his counterpart who did not migrate. As expected, recruits first seen in an urban county had a higher risk of dying than those first seen in a rural county. Farmers, Midwesterners, and the native-born were other groups with lower hazard rates of dying.

Migrants' life expectancies could be shorter either because of the environmental changes to which they were exposed or because migrants were a self-selected group.[24] Although self-selection cannot be ruled out, examining mortality among migrants in detail will help us understand some of the mechanisms through which migrations may have shortened life expectancy. Table 8.5 shows the distribution of recruits' causes of death using five classification groups: chronic, infectious (tuberculosis, venereal diseases, typhoid, cholera, dysentery, and other bacterial and viral diseases), acute respiratory, other diseases, and missing cause of death. I use these causes of death to estimate a competing-risks model of mortality.

Table 8.6 shows that migrants were more likely to die of infectious diseases than nonmigrants. Are these results solely attributable to the fact that they changed locations? Did it matter whether they moved to a rural or an urban place? Including a set of dummy variables for the urban-rural status of both the origin and final destination of migrants shows that conditional on being a migrant, migrants to rural counties had higher life expectancies (see table 8.7). Migrants to rural areas had lower life expectancies com-

24. For instance, migrants may have been poorer, and lower income may have been negatively correlated with a longer life span. On the other hand, migrants may have been taller, stronger, and in general healthier than those who did not move, and therefore longer-lived.

Table 8.5 Distribution of Causes of Death

Type of Disease	Frequency	Percentage
Chronic, not infectious	4,645	68.39
Neoplasm	389	5.73
Diseases of the blood	539	7.94
Circulatory system	1,571	23.13
Genitourinary	605	8.91
Other chronic, not infectious	1,541	22.69
Infectious	736	10.84
Tuberculosis	380	5.50
Venereal diseases	218	3.21
Other infectious	138	2.03
Acute respiratory	636	9.37
Other diseases	774	11.40

Notes: Percentages are computed based on the 6,791 recruits with nonmissing cause of death. Other chronic, not infectious includes allergies, diabetes, diseases of metabolic origin, uremia, and other chronic diseases of not-infectious origin affecting the nervous, digestive, genitourinary, cardiovascular, and gastrointestinal systems. Other infectious includes typhoid, cholera, dysentery, and other bacterial and viral diseases. Acute respiratory includes influenza, pneumonia, bronchitis, and acute laryngitis and tracheitis. Other diseases includes mental diseases, other acute diseases, accidents and violence, and malaria.

Table 8.6 Competing-Risks Regression for the Time Remaining to Die from Different Types of Diseases

	Hazard ratio				
Variable	Chronic, not Infectious, Diseases	Infectious Diseases	Acute Respiratory Diseases	Other Diseases	Unknown Causes of Death
Age at first observation	1.122**	1.082**	1.156**	1.116**	1.152**
Age at first observation squared	1.000**	0.999**	0.999**	1.000**	1.000**
White-collar	1.275**	0.861	1.266	1.334**	0.894
Manual worker	1.045	1.235**	0.981	1.278**	1.141**
Service worker	1.094	1.239	0.755	0.971	1.234**
Northeast at first observation	1.238**	1.048	1.094	1.043	0.910**
South at first observation	0.819**	1.298**	1.284	1.023	1.023
West at first observation	0.868	0.769	1.006	0.785	1.131
Urban at first observation	1.256**	1.511**	1.385**	1.229**	0.974
Foreign-born	1.073**	1.063	1.438**	1.199**	0.967
Migrant status (time-varying)	1.082**	1.151**	1.032	1.187**	1.064
Number of observations	11,097				
Model χ^2 (Wald)	3,298				
p-value	0.0001				

Notes: Omitted variables are farmer, Midwesterner at first observation, rural at first observation, and native-born.

**Significant at the 5 percent level.

Table 8.7 **Competing-Risks Regression for the Time Remaining to Die from Different Types of Diseases, for Migrants**

	Hazard Rate					
Variable	Any Cause of Death	Chronic, not Infectious Diseases	Infectious Diseases	Acute Respiratory Diseases	Other Diseases	Unknown Causes
Age at first observation	1.124**	1.112**	1.082*	1.181*	1.128*	1.141*
Age at first observation squared	1.000**	1.000	1.000	0.999	1.000	1.000
White-collar	1.096	1.104	1.164	1.611*	1.697*	0.876
Manual worker	1.087*	0.981	1.288*	1.144*	1.292*	1.134*
Service worker	1.272*	1.255	1.011	1.117*	1.516	1.342
Northeast at first observation	1.087*	1.247*	1.118	1.095	0.955	0.932
South at first observation	1.022	0.859	1.461*	1.633	0.851	1.07
West at first observation	0.912	0.938	1.153	1.133	0.618	0.903
Moved from urban to rural	0.892*	0.828	1.16	0.803*	0.680	1.009
Moved from rural to urban	0.864*	0.853	0.686	0.859	0.917	0.902
Moved from rural to rural	0.849*	0.739*	1.024	0.769	0.820	0.985
Foreign-born	1.059	1.047	0.913	1.316	0.970	1.082
Number of observations	5,010					
Model χ^2 (Wald)	1,451.5					
p-value	0.0001					

Notes: Omitted variables are farmer, Midwesterner at first observation, urban-urban migrant, and native-born.

**Significant at the 5 percent level.

pared to nonmigrants. Excluding migrants to urban counties, table 8.8 shows that even after controlling for the urban-rural status of the original location, changing locations had a negative effect on life expectancy.[25] Even migrants to rural places had a 2 percent higher risk of dying within the next year compared to rural nonmigrants.[26] These findings imply that there was a "pure" migration effect, perhaps related to the stress of moving to another location and the physical costs that these moves required.

25. Recall that in subsection 8.3.3 I mentioned that because individuals with longer life spans are at higher risk of moving, including a time-fixed dummy variable to indicate migrant status produces downward-biased estimates of the migration effect on life expectancy. Therefore, I cannot simply include dummies for rural nonmigrant and urban nonmigrant in the previous regression to make comparisons between movers and nonmovers.

26. As table 8.2 shows, a considerable number of recruits moved more than once. I estimated this competitive risks model using not the final, but the first destination as the relevant arrival location for migrants. The qualitative results hold.

Table 8.8 **Cox-Regression on the Survival Times of UAMD Veterans, Excluding Migrants to Urban Counties**

Variable	Mean	Hazard Ratio	Pr > χ^2
Age at first observation	38.02	1.126	0.0001
Age at first observation squared		1.000	0.0001
White-collar	0.04	1.136	0.0177
Manual worker	0.34	1.107	0.0001
Service worker	0.02	1.093	0.2592
Northeast at first observation	0.31	1.063	0.0211
South at first observation	0.05	0.966	0.4931
West at first observation	0.02	0.969	0.6735
Urban at first observation	0.16	1.188	0.0001
Foreign-born	0.17	1.063	0.0384
Migrant status (time-varying)		1.026	0.0310
Number of observations	10,047		
Censored observations	2,076		
Model χ^2 (Wald)	5,474.5		
p-value	0.0001		

Notes: Omitted variables are farmer, Midwesterner at first observation, rural at first observation, and native-born.

8.5 Conclusions and Future Research

This paper used a new longitudinal data set to study migration in the postbellum United States and to examine the characteristics of migrants. The data allowed me not only to complement the work of previous researchers, but also to address questions that required longitudinal data. I estimated the hazard rate of changing county of residence within a year, arguing that this is a highly flexible specification for measuring mobility. This allowed me to investigate how the hazard rate of moving changed with age. I also showed that return migration was a common phenomenon during the second half of the nineteenth century. Younger, unmarried recruits who traveled shorter distances and who traveled to the South were more likely to return to a county where they previously resided. Migrants who left family behind were more likely to become return migrants. Migrant life expectancy was significantly shorter than that of counterparts who did not migrate because of migrants' higher probability of dying of infectious disease. Infectious diseases were particularly important in explaining the reduction in life expectancy for those who moved to urban counties. However, migrants across rural areas also suffered higher mortality rates relative to rural nonmovers.

The findings have implications for the extent of labor market integration in the postbellum United States. Price equalization across labor markets is achieved through migration. I found that workers were quite mobile, even at mature ages, and that many workers moved temporarily. These temporary moves suggest that workers responded to economic shocks even though migration reduced their life expectancies. Wage differentials between cities and rural areas (net of migration costs) may have been high because of reduced migrant life expectancy.

Appendix A
UAMD Variables

This section describes the variables in UAMD.

Year of birth. I imputed veterans' year of birth from date of and age at enlistment because these variables provide the most complete information on age.

Death year. This is generally known for all recruits linked to PEN. I assumed that all recruits without a wartime death date survived the war. This produced a wartime death rate of 12.8 percent, close to Dyer's (1959) estimate for the Union Army as a whole. Statewise comparisons of death rates yield equally consistent results.

Linkage to PEN. Any recruit with a pension application date, information on a Pension Board's resolution on claims made by him or his heirs, or with a pension certificate number was considered to be linked to PEN.

County of residence. After cleaning spelling errors, and inputting county from city of residence, I coded the county names using the ICPSR coding scheme created by Sechrist (1984). This coding scheme allows for changes in the names of counties through time, and I assigned a special code to the residences of recruits who reported having lived outside the United States during some part of their lives. Dates in which the veteran resided in a location accompany the location description. When two different sources reported conflicting information on the location of the recruit for a particular year (true in only 3 percent of all cases when two different documents are available), I coded the recruit's residence for that year as missing.

Occupation at enlistment. I coded the occupation of the recruit at enlistment using the 1950 census of population's four-digit classification scheme.

Urban-rural status. A county is considered urban if it contains a city with at least 25,000 inhabitants. I obtained the city population figures from the censuses of population. When two counties changed status from one census to another, I assumed exponential growth in population to impute the year in which the county's urban-rural status changed.

Number of children. Veterans and their dependents had to declare the number of children who were born alive and their dates of birth, as well as the dates of death of any of their children. If no death dates were given, I assumed that the veteran's children outlived their father.

Marital status. The Pension Board required veterans to inform them of any change in marital status. I assumed that if no death date was given for the wife that she outlived the veteran.

Distance traveled. Using Sechrist's (1984) data, I computed the distance (in miles) traveled by migrants from the county seats' longitude and latitude information.

Appendix B

Linkage Failure to PEN and the Effect of War-Related Wounds on the Mobility of Veterans

Veterans linked to PEN either applied claiming pensionable conditions or had legal heirs who claimed to be eligible pensioners. The definition of a pensionable condition changed constantly over time. Starting in 1862, as stated in the General Law of Pensions, pensions were granted to any honorably discharged veteran who served for more than ninety days and who had a war-related condition affecting his ability to perform manual labor. However, in 1890, with the Disability Act Pension Law System, veterans became eligible for pension if they suffered from any medical condition, war-related or not. By 1904, the federal government had equated older age and disability.[27]

Table 8B.1 presents the results of a logistic regression explaining linkage to PEN, where a value of 1 for the dependent variable indicates linkage and zero otherwise. The explanatory variables include individual characteristics, as well as variables for military outcomes, as these may have influenced the eligibility of the potential pensioner as well as the smoothness of the application process.[28]

Relatively older people, native-born veterans, farmers, people who enlisted in the Midwest, and people who were wounded or became prisoners during the war were more likely to be linked to PEN. Region of enlistment and occupation may predict linkage because pensions were an electoral weapon during the postbellum period (Costa 1998; Skocpol 1992). Political competition was highest in the Midwest and, as noted by Glasson (1918), farmers were overrepresented in the Grand Army of the Republic, the veterans' lobbying organization. Those who were honorably discharged (ap-

27. See Glasson (1918) for a detailed description of the main changes in the relevant pension legislation.

28. Only information recorded in military-time documents was used for this analysis. For some recruits, age, place of birth, occupation at enlistment, or time served during the Civil War is missing. When this is the case, a dummy variable indicating whether the information is missing for the recruit is included. Therefore, the coefficients for the nonmissing variables should be seen as interaction terms between having nonmissing information and the variable.

Table 8B.1 Logistic Regression on the Probability to be Linked to PEN

Variable	Mean	Logistic Coefficient	Slope
Intercept		0.0282*	0.0056
Age after the war	28.575	0.00938*	0.0529
Place of birth missing	0.03	–1.2942*	–0.0077
Foreign-born	0.286	–0.8129*	–0.0459
White-collar	0.061	–0.4854*	–0.0058
Service worker	0.008	–0.7414*	–0.0012
Manual worker	0.428	–0.3413*	–0.0288
Honorably discharged	0.82	1.1837*	0.1916
Prisoner of war	0.083	0.3473*	0.0057
Volunteer	0.246	0.0385	0.0019
Wounded during the war	0.265	1.3681*	0.0716
Time served	455.455	0.00021	0.0189
Enlisted in Northeast region	0.424	–0.2827*	–0.0237
Enlisted in South region	0.054	–0.1907*	–0.0020
Enlisted in West region	0.021	–1.6352*	–0.0068
N	30,763		
–2 log-likelihood	6593.127		
Max-rescaled R^2	0.2852		

Notes: Omitted variables are native-born, farmers and farm laborers, and enlisted in Midwest. The control variables include dummies indicating that age is missing, that place of birth is missing, that occupation is missing, that occupation is unclassifiable, and that time served is missing.
*Coefficient is significantly different from zero at least at the 1 percent level.

Table 8B.2 Distribution of Wounds Suffered during the War, by Severity

Wound Class	Frequency	Percent
Never had a wound examination	1,253	50.4
Examined for wounds, but given a zero rating	61	2.5
Wounds examined and granted a nonzero rating		
Equivalent to less than anchylosis of a wrist	566	22.8
Worse than above but less than third-degree	500	20.1
Worse than above but less than second-degree	82	3.3
Worse than above but less than first-degree	19	0.8
First-degree	5	0.2

Notes: Total number of recruits: 2,486. Third-degree was considered comparable to the loss of an arm below the elbow, second-degree to the loss of an arm above the elbow, and first-degree to the loss of an arm and a leg.

proximately 90 percent of all soldiers), had been prisoners of war, or were wounded during the war were also more likely to be pension recipients. Veterans who died at young ages were less likely to be linked, as they may not have developed any pensionable condition before dying.

Did being wounded in the war permanently affect recruits' capacity to relocate? Table 8B.2 shows the distribution of wounds suffered during the war by survivors who were wounded during the war and who were ever examined by a surgeon working for the Pension Bureau. Wounds are classified according to severity, where third degree is comparable to the loss of an arm below the elbow, second degree is equivalent to the loss of an arm above the elbow, and first degree is equivalent to the loss of an arm and a leg. This table shows that the majority of recruits in UAMD should not have been affected by their physical capacity to move. More than half did not suffer a permanent incapacity due to wounds suffered during the war. Nearly 30 percent of them were only mildly incapacitated. In fact, as shown in section 8.4.1, those recruits who were ill or wounded during the war were more mobile than those recruits who were neither ill nor wounded.

Appendix C

Size of Bias of Estimates of Mobility Relying on Samples Linked across Decennial Censuses of Population

Previous studies of nineteenth-century migration processes relied on estimates of mobility rates to answer the question "how mobile was this population?" Let P_t represent the number of persons who start period t at a different location with respect to the previous period, and who before t had not made a movement, and let N_{t1} be the total number of persons in the population of interest at $t1$. The mobility rate for the period ($t1$, $t2$) is defined as

(A1) $$M_{(t1,t2)} = \sum_{t1+1}^{t2} \frac{P_t}{N_{t1}}$$

Researchers relying on samples linked across censuses of population are confined to estimate the following version of equation (A1):

(A2) $$M_{(t1,t1+10)} = \sum_{t1+1}^{t1+10} \frac{P_t}{N_{t1}}$$

Rather than observing P_t, the number of persons with a different location at $t1$ and $t1 + 10$ is used to estimate $\sum_{t1+1}^{t1+10} P_t$. Because people may have moved between census years and returned to their previous locations (return migration) within the same ten-year period, estimation of equation (A2) is downwardly biased.

Although N_{t1} can be observed, inference about moves can be made only for those who were observed and survived to N_{t11+10} (by construction), and therefore this population is taken as the base population. Whether this sample of survivors is representative of the base sample depends on whether the variables explaining census undercount and mortality are correlated to the propensity of individuals to move.

The size of the return migration bias, and whether differences in life expectancy for migrants relative to nonmigrants exist, are issues that cannot be explored with decennial census data. My panel data enable me to explore more flexible specifications of the estimator for equation (A1) and to investigate return migration and the mortality of migrants.

References

Borjas, George J. 1994. The Economics of Immigration. *Journal of Economic Literature* 32 (4): 1667–1717.

Center for Population Economics. 2000. *Public use tape on the aging of veterans of the Union Army: Data user's manual.* Chicago: University of Chicago, Graduate School of Business, Center for Population Economics.

Costa, Dora L. 1998. *The evolution of retirement: An American economic history, 1880–1990.* Chicago: University of Chicago Press.

Cox, David R. 1984. *Analysis of survival data.* London: Chapman and Hall.

Curtin, Philip D. 1989. *Death by migration: Europe's encounter with the tropical world in the nineteenth century.* New York: Cambridge University Press.

Dyer, Frederick H. 1959. *A compendium of the War of the Rebellion.* Vol. 1. New York: Thomas Yoseloff.

Ferrie, Joseph P. 1996. A new sample of Americans linked from the 1850 public use sample of the federal census of population to the 1860 federal census manuscripts. *Historical Methods* 29 (4): 141–56.

———. 1997. Migration to the frontier in mid-nineteenth-century America: A reexamination of Turner's "safety valve." Northwestern University, Department of Economics. Unpublished manuscript.

———. 1999. How ya gonna keep 'em down on the farm (when they've seen Schenectady?): Rural-to-urban migration in the nineteenth-century America, 1850–70. Northwestern University, Department of Economics. Unpublished manuscript.

Fogel, Robert W. 1986. Nutrition and the decline in mortality since 1700: Some preliminary findings. In *Long-term factors in American economic growth,* ed. S. L. Engerman and R. E. Gallman, 439–555. Vol. 51 in Studies in Income and Wealth. Chicago: University of Chicago Press.

———. 1993. New sources and new techniques for the study of secular trends in nutritional status, health, mortality, and the process of aging. *Historical Methods* 26 (1): 5–43.

Gallaway, Lowell E., and Richard K. Vedder. 1971. Mobility of Native Americans. *Journal of Economic History* 31:3.

Glasson, William H. 1918. *Federal military pensions in the United States.* New York: Oxford University Press.

Greene, William H. 1997. *Econometric analysis.* 3rd ed. Upper Saddle River, N.J.: Prentice Hall.

Gurak, Douglas T., and Fe Caces. 1992. Migration networks and the shaping of migration systems. In *International migration systems: A global approach,* ed. Mary Kritz, Lin Lean Lim, and Hania Zlotnik, 150–76. Oxford, U.K.: Clarendon.

Haines, Michael R. 2000. The population of the United States, 1790–1920. In *The Cambridge economic history of the United States,* ed. Stanley L. Engerman and

Robert E. Gallman. Vol. 2, *The long nineteenth century,* 143–205. New York: Cambridge University Press.

Hugo, Graeme J. 1981. Village-community ties, village norms, and ethnic and social networks: A review of evidence from the third world. In *Migration decision making: Multidisciplinary approaches to microlevel studies in developed and developing countries,* ed. Gorden F. DeJong and Robert W. Gardner, 186–224. New York: Pergamon.

Kuznets, Simon and Dorothy Thomas. 1957. *Population redistribution and economic growth, United States, 1870–1950.* Philadelphia: American Philosophical Society.

Lalonde, Robert, and Robert Topel. 1997. Economic impact of international migration and the economic performance of migrants. In *Handbook of population and family economics,* eds. Mark R. Rosenzweig and Oded Stark, 799–850. Amsterdam: North–Holland.

Lebergott, Stanley. 1970. Migration within the U.S., 1800–1960: Some new estimates. *Journal of Economic History* 30:4.

O'Rourke, Kevin, Jeffrey Williamson, and Timothy Hatton. 1994. Mass migration, commodity market integration, and real wage convergence: The late-nineteenth-century Atlantic economy. In *Migration and the international labor market, 1850–1939,* ed. Timothy J. Hatton and Jeffrey G. Williamson, 203–20. London: Routledge.

Ruggles, Steven, and Matthew Sobek, et al. 1997. Integrated public use microdata series: Version 2.0. Minneapolis: Historical Census Projects, University of Minnesota, available at http://www.ipums.org.

Sechrist, Robert P. 1984. Basic geographic and historic data for interfacing ICPSR data sets, 1620–1983 (*United States*). ICPSR 8159. Baton Rouge, La.: Louisiana State University, and Ann Arbor, Mich.: Inter-university Consortium for Political and Social Research.

Sjaastad, Larry A. 1962. The costs and returns of human migration. *Journal of Political Economy* 70 (5): 80–93.

Skocpol, Theda. 1992. *Protecting mothers and soldiers: The political origins of social policy in the United States.* Cambridge: Harvard University Press.

Stark, Oded. 1991. *The migration of labor.* Cambridge, U.K.: Basil Backwell.

Stark, Oded and David Bloom. 1985. The new economics of labor migration. *American Economic Review* 75 (2): 173–78.

Steckel, Richard H. 1989. Household migration and rural settlement in the United States, 1850–1860. *Explorations in Economic History* 26 (2): 190–218.

Taylor, Edward J. 1986. Differential migration, networks, information and risk. In *Research in human capital and development,* vol. IV, ed. Oded Stark, 747–71. Greenwich, Conn.: JAI Press.

Todaro, Michael P. 1969. A model of labor migration and urban unemployment in less-developed countries. *American Economic Review* 59 (1): 138–48.

Turner, Frederick J. 1920. *The frontier in American history.* New York: Holt.

Williamson, Jeffrey G. 1990. *Coping with city growth during the British Industrial Revolution.* Cambridge: Cambridge University Press.

World Health Organization. 1949. *Manual of the international statistical classification of diseases, injuries, and causes of death.* 6th rev. of the *International lists of diseases and causes of death,* vol. 1. Geneva: World Health Organization.