# Sample-Selection Bias and Height Trends in the Nineteenth-Century United States


Ariell Zimran
ariell.zimran@vanderbilt.edu

Vanderbilt University and NBER


May 2018

Click Here for Most Recent Version


**Abstract**

I use census-linked military height data and a two-step semi-parametric sample-selection model to adjust trends in average stature in the nineteenth-century United States for selection on observables and on unobservables. This adjustment meaningfully and statistically significantly alters the trend in average stature between the birth cohorts of 1832 and 1860, validating concerns over bias in the historical heights literature. After adjustment, I find a net decline in average stature of 0.64 inches in the same cohort range, supporting the veracity of the Antebellum Puzzle—a deterioration of health during early modern economic growth in the United States.

**Acknowledgements** I am indebted to Joel Mokyr, Joseph Ferrie and Matthew Notowidigdo for encouragement and guidance, and to Ann Carlos (the editor), William Collins, and anonymous referees for detailed comments on several drafts of this paper. I also thank Ran Abramitzky, Hoyt Bleakley, Louis Cain, John Cawley, Stephanie Chapman, Carola Frydman, Seema Jayachandran, John Komlos, Peter Koudijs, Thomas Mroz, Aviv Nevo, Sangyoon Park, Pedro Sant'Anna, Yannay Spitzer, Richard Steckel, Benjamin Ukert, and Carlos Villareal for helpful suggestions and insightful comments. Thanks are also due to Roy Mill for access to the dEntry transcription system; to seminar participants at Northwestern University, the London School of Economics, Vanderbilt University, the University of Colorado Boulder, the University of Queensland, the Australian National University, the University of Melbourne, and Stanford University; to participants in the 2013 Asian Meetings of the Econometric Society, the 2015 Western Economic Association International Graduate Student Dissertation Workshop and Conference, the 2015 European Historical Economics Society Conference, the 2015 Illinois Economic Association Conference, the 2015 Social Science History Association Conference, and the 2015 H2D2 Research Day at the University of Michigan for helpful comments; and to Christie Jeung for excellent research assistance. This project was supported by the Northwestern University Center for Economic History, the Balzan Foundation, a Northwestern University Graduate Research Grant, and an Economic History Association Dissertation Fellowship. Computations were performed on the Social Sciences Computing Cluster at Northwestern University and on the Advanced Computing Cluster for Research and Education at Vanderbilt University. This project, by virtue of its use of the Union Army Data, was supported by Award Number P01 AG10120 from the National Institute on Aging. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institute on Aging or the National Institute of Health. All errors are my own.

**Notes** This is a revised version of chapter 2 of my dissertation. Previous versions of this paper were titled "New Perspectives on Historical Standards of Living: Evidence from US Military Enlistment in the Late Nineteenth Century," "Does Sample-Selection Bias Explain the Industrialization Puzzle? Evidence from Military Enlistment in the Nineteenth-Century United States," and "Does Sample-Selection Bias Explain the Antebellum Puzzle? Evidence from Military Enlistment in the Nineteenth-Century United States."

# 1 Introduction

The improvement of health through the elimination of chronic malnutrition is undoubtedly among the most important benefits from modern economic growth in developed countries (Fogel, 1994). Recent discoveries of declining or persistently poor health in rapidly growing developing countries such as India (Deaton, 2007; Jayachandran and Pande, 2017) and China (Trivedi, 2017) have therefore come as a surprise to some, who expected improving health to accompany rising incomes in these countries as well. But historical research finds that the transition from stagnation to growth was disruptive, even in developed countries. There is evidence that residents of both Britain (Floud, Wachter, and Gregory, 1990) and the United States (A'Hearn, 1998; Craig, 2016; Floud et al., 2011; Fogel, 1986; Haines, 2004; Komlos, 1987; Margo and Steckel, 1983; Zehetmayer, 2011) experienced declining health over several decades of the nineteenth century. Together, these patterns suggest that declining health may be a common aspect of the early development process.

In the American case, this phenomenon is known as the "Antebellum Puzzle." Despite rising income per capita in the 19th century, average height (a standard measure of health in historical contexts) appears to have declined precipitously in the birth cohorts of the 1830s to the 1850s and then to have stagnated for nearly 50 years.[1] This result has generated a large literature seeking to understand what mechanisms might have been responsible for the decline (see summary by Floud et al., 2011).

The implication that early modern economic growth in the United States came at the expense of health has been met with some skepticism. Instead of accepting the existence of this puzzle and seeking to explain it, some scholars have challenged its empirical basis, suggesting that the decline in height might be an artifact of the data rather than a true representation of living standards (e.g., Gallman, 1996). In particular, these scholars have argued that the data used to establish the existence of the Antebellum Puzzle and related phenomena may suffer from *sample-selection bias*, which arises when conclusions are drawn from a sample that is not representative of the population. Their main contention is that volunteer military records, from which the bulk of historical height data are drawn, represent only individuals who chose to enlist, and are therefore unlikely to have been representative of the whole population. This concern goes beyond the possible impacts of *selection on observables*, which would arise if the military and the whole population differed only on the basis of observable characteristics (e.g., if residents of urban areas were shorter and more likely to join the military, but all urbanites were equally likely to enlist)—concerns that the historical heights literature has long recognized and addressed (Fogel, 1986; Fogel et al., 1983). Instead, this concern is based on the possible presence of *selection on unobservables*, which would arise if the military and the population at large

[1] Further evidence of a deterioration in health is given by a decline in life expectancy during this period (Fogel, 1986).

differed in terms of characteristics unobservable to the researcher (e.g., if a childhood health shock made shorter urbanites more likely to enlist than taller ones).

Bodenhorn, Guinnane, and Mroz (2017) have recently argued that the existing literature has not satisfactorily addressed bias from selection on unobservables. Importantly, because time-invariant selection on unobservables would not bias the observed *trend* in heights, they argue that the improving economic conditions of the antebellum period may have made selection on unobservables more negative over time, leading to a spurious decline in observed stature among military enlisters.[2] That is, they argue that population height might have been rising (or at least not falling), and that only the height of enlisters declined because the composition of this group changed as successive cohorts faced better options in the civilian labor market.[3]

Although Bodenhorn, Guinnane, and Mroz (2017) have found evidence suggesting the presence of time-varying bias from selection on unobservables into American height data sets, it is still unknown whether sample-selection bias wholly, or even partially, accounts for the Antebellum Puzzle. More generally, it is not known whether correcting for bias from selection on unobservables would lead to any meaningful changes in conclusions drawn from historical samples of American height data.

To my knowledge, this paper is the first to adjust the trend in average stature for sample-selection bias stemming from selection on unobservables. I do this by estimating a two-step semi-parametric sample-selection model (Das, Newey, and Vella, 2003; Heckman, 1979; Klein and Spady, 1993; Newey, 2009; Vella, 1998) for height observed only among military enlisters, producing an estimated trend in average stature that is adjusted for selection on both observables and unobservables. The results enable me to shed light on two questions. First, does incorporating a correction for sample-selection bias meaningfully alter the conclusions drawn from stature data? I address this question by comparing my adjusted trend in average stature to the trend estimated using standard techniques of the anthropometric history literature, which do not correct for selection on unobservables. Second, is the Antebellum Puzzle an artifact of sample-selection bias? I address this question by determining whether the data exhibit an Antebellum Puzzle after incorporating the correction for sample-selection bias. To conclude that no puzzle is present, it must be the case that the estimated trend in average stature is increasing over time; simply eliminating the decline in average stature would not be sufficient to conclude that the Antebellum Puzzle is an artifact of sample-selection bias.

I draw my main height data from military records for US-born white males from the birth cohorts of 1832–

[2]Bodenhorn, Guinnane, and Mroz's (2017) critique extends beyond the Antebellum Puzzle in the United States to all anthropometric history and to the broader result known as the Industrialization Puzzle.

[3]As Mokyr and Ó Gráda (1996, p. 164) put it, data may have been "drawn increasingly from the left tail of a distribution which itself is shifting to the right." In Gallman's (1996, p. 194) words, military enlisters may not have "retained an unchanging character" over time.

1860. I collected data for the birth cohorts of 1832–1846 from Fogel et al. (2000), who provide information on individuals who served in the Union Army during the Civil War. Individuals born after 1846 would have been too young to serve in the Union Army (which was disbanded after the end of the war), so I collected data for the birth cohorts of 1847–1860 from the records of postbellum enlistments in the *Register of Enlistments in the U.S. Army, 1798–1914* (n.d., henceforth, *Register of Enlistments*). This source provides information on individuals enlisting in the professional Regular Army. This combination of sources has previously been used to establish the existence of the Antebellum Puzzle (e.g., Fogel, 1986, Table 9.6); but historical accounts of military enlistment (Bernardo and Bacon, 1955; Coffman, 1986; Foner, 1970; Weigley, 1967) indicate that the incentives for enlistment and the conditions of service were better in the Union Army than in the Regular Army, suggesting that there may have been more negative selection into the Regular Army after the Civil War (the 1847–1860 cohorts) as compared to the Union Army during the Civil War (the 1832–1846 cohorts).

Additional data are collected from the US censuses of 1850–1870. Individuals observed in the military data are linked to their census records in adolescence, thus adding socioeconomic characteristics of enlisters to the data set. I also collected a random sample of micro-level census data (Ruggles et al., 2015) from the complete population in these birth cohorts that was at risk for military service. These data make possible comparisons of the pre-enlistment socioeconomic characteristics of enlisters to those of the whole population, enabling me to characterize the determinants of military enlistment.

For identification of the sample-selection model, it is necessary to isolate variation in the probability of enlisting in the military that is unrelated to height, conditional on all covariates. To this end, I impose two restrictions on the model. First, I incorporate county-level vote shares for Abraham Lincoln in 1860 in a binary choice model of military enlistment (the first step of the two-step sample-selection model), but exclude this variable from the equation determining height. This variable, which is informative on a county's views on slavery and other central issues in the election, is a proxy for individuals' political ideology. By way of example, voting data for this election and others of the era have been shown to be important in the military desertion decision during the Civil War (Costa and Kahn, 2003, 2007) and in determining migration of Civil War veterans in the postbellum period (Eli, Salisbury, and Shertzer, 2018). Second, I allow the effects of covariates on the probability of enlisting in the military to vary based on whether an individual's birth year made him eligible to serve during the Civil War (i.e., by whether he was born in 1846 or earlier), whereas the equation determining height is assumed to be time-invariant. This restriction is based on the historical accounts of military enlistment cited above.

I find that failing to account for selection on unobservables can appreciably affect the conclusions drawn from historical height data. My estimated trend in average stature, which incorporates the correction for selection on unobservables, differs meaningfully and statistically from the trend estimated using the standard methodology of the historical heights literature, which includes no such correction. I thus validate concerns over the existence of sample-selection bias induced by selection on unobservables in historical height samples. In particular, the magnitude of the decline in average stature between 1832 and 1860 without correcting for selection on unobservables is 1.29 inches in my data, with 1.25 inches being the benchmark in the literature (e.g., Costa and Steckel, 1997; Craig, 2016). Adjusting for selection on unobservables results in a considerably smaller decline of only 0.64 inches, and it is possible to reject the null hypothesis of equality between this decline and the decline estimated according to the literature's standard techniques. The chief cause of changing selection on unobservables appears to be the changing composition of the military after the Civil War. Consistent with historical accounts, I find that the 1847–1860 cohorts, who enlisted in the Regular Army, were more negatively selected than were the 1832–1846 cohorts, who enlisted in the Union Army. Combining these sources without correcting for this concern, as is commonly done in the historical heights literature (e.g., Fogel, 1986), leads to bias.

Despite the presence of this cohort-varying sample-selection bias, my results do not support the suspicion that the Antebellum Puzzle is a statistical artifact. A decline in stature of 0.64 inches is evident in the trend incorporating the correction for selection on unobservables, and it is possible to reject the null hypothesis of no decline in heights over time (*a fortiori* ruling out the increase in average stature that would be required to solve the puzzle by sample selection alone). Thus, my results support the notion that early modern economic growth was disruptive to health in the United States.

This paper also addresses a broader challenge to research in economic history. Nearly all economic historical data are drawn from sources that are potentially vulnerable to sample-selection bias, and researchers often struggle to deal with it. Although I show that selection on unobservables may make drawing firm conclusions more difficult, I also find a path forward that allows the bias to be quantified and shows how researchers can learn from a selected sample without ignoring its potential pitfalls.

# 2 Empirical Framework

## 2.1 The Model

To explain how selection on observables and selection on unobservables generate sample-selection bias and how I adjust for it, I introduce a simple model of height and military enlistment. As is common in settings where researchers are concerned about selection on unobservables, I use a Tobit type-II model (Amemiya, 1985), which is an empirical application of the Roy model used by Bodenhorn, Guinnane, and Mroz (2014, 2017, p. 185) to illustrate their concerns regarding selection on unobservables.

Suppose that the height of individual $i$ from birth cohort $t$, $h_{it}$, is determined by

$$h_{it} = \gamma_t + \mathbf{x}'_{it}\theta + \varepsilon_{it}, \tag{1}$$

where $\gamma_t$ are cohort-specific intercepts, $\mathbf{x}_{it}$ is a vector of covariates affecting both height and military enlistment, and $\varepsilon_{it}$ are unobserved components in the determination of height. Let $y_{it}$ be an indicator variable equal to one if individual $i$ from birth cohort $t$ enlists and zero otherwise. Suppose that individuals enlist if and only if their latent utility of enlistment, $y^*_{it}$, is greater than zero. Let $y^*_{it}$ be determined by

$$y^*_{it} = \alpha_t + \mathbf{x}'_{it}\beta_k + \mathbf{z}'_{it}\delta_k + u_{it}, \tag{2}$$

where $\alpha_t$ are cohort-specific intercepts, $\mathbf{x}_{it}$ is (as above) a vector of covariates affecting both military enlistment and height, $\mathbf{z}_{it}$ is a vector of covariates affecting military enlistment but not height, and $u_{it}$ are unobserved components in the determination of military enlistment.[4] Note that the coefficients $\beta_k$ and $\delta_k$ in equation (2) are indexed by $k$ to indicate that they are permitted to vary by cohort group (i.e., $k$ could represent either the 1832–1846 birth cohorts, who were old enough to serve in the Civil War, or the 1847–1860 birth cohorts, who were not). I impose this structure because the nature of the enlistment decision likely differed between the cohorts that were eligible for Civil War service and those that were not. The variables $y_{it}$, $\mathbf{x}_{it}$, and $\mathbf{z}_{it}$ are observed regardless of enlistment status,[5] while $\varepsilon_{it}$, $u_{it}$, and $y^*_{it}$ are never observed. The main challenge is that $h_{it}$ is observed only if $y_{it} = 1$—that is, height is observed only if individual $i$ enlists.

Under standard assumptions,[6] it is possible to write the probability that individual $i$ enlists, given his

[4]In principle, one could allow equation (1) to contain variables that do not appear in equation (2); but because doing so would place a restriction on equation (2) that does not aid in identification, I follow the common approach of sample-selection models (e.g., Vella, 1998, p. 130) and let the data speak as to the forces that do and do not affect military enlistment.

[5]It will be necessary below to relax the observability of $y_{it}$ given the available data.

[6]I assume that $u_{it}$ satisfies the index assumption, so that it is possible to write

$$P(y_{it} = 1|\mathbf{x}_{it}, \mathbf{z}_{it}; t) = G(\alpha_t + \mathbf{x}'_{it}\beta_k + \mathbf{z}'_{it}\delta_k),$$

where $G(\cdot)$ is continuous and continuously differentiable. This assumption permits certain forms of heteroskedasticity and is sufficient for semi-parametric identification of equation (2) (Klein and Spady, 1993). Intuitively, this assumption requires a lack of correlation between the error $u_{it}$ and the regressors. I also assume that, conditional on the propensity score $G(\alpha_t + \mathbf{x}_{it}\beta_k + \mathbf{z}'_{it}\delta_k)$, $\varepsilon_{it}$ is uncorrelated with all functions of $(\mathbf{x}_{it}, \mathbf{z}_{it})$. Das, Newey, and Vella (2003) show that this assumption is sufficient for non-parametric identification of a sample-selection model for $h_{it}$. Das, Newey, and Vella (2003) point out that this assumption permits heteroskedasticity of unknown form.

observable characteristics (his *conditional probability of military enlistment*), as

$$P(y_{it} = 1|\mathbf{x}_{it}, \mathbf{z}_{it}; t) = G(\alpha_t + \mathbf{x}'_{it}\beta_k + \mathbf{z}'_{it}\delta_k). \tag{3}$$

These assumptions also permit height for enlisters to be written as

$$h_{it} = \gamma_t + \mathbf{x}'_{it}\theta + \Omega(\alpha_t + \mathbf{x}'_{it}\beta_k + \mathbf{z}_{it}{}'\delta_k) + \xi_{it}, \tag{4}$$

where $\Omega(\alpha_t + \mathbf{x}'_{it}\beta_k + \mathbf{z}'_{it}\delta_k) = E(\varepsilon_{it}|\mathbf{x}_{it}, \mathbf{z}_{it}, y_{it} = 1; t)$ and $\xi_{it}$ is an error term that is orthogonal to $u_{it}$. Thus, enlister $i$'s height is the sum of three components—the average height in the population for all individuals with the same observables ($\gamma_t + \mathbf{x}'_{it}\theta$), the difference in average height between enlisters and the whole population with the same observables ($\Omega(\alpha_t + \mathbf{x}'_{it}\beta_k + \mathbf{z}_{it}{}'\delta_k)$), and a "well behaved" error term ($\xi_{it}$). Note that if $\varepsilon_{it}$ and $u_{it}$ are uncorrelated (i.e., there is no selection on unobservables) then $\Omega(\cdot) = E(\varepsilon_{it}|\mathbf{x}_{it}, \mathbf{z}_{it}; t) = 0$ and the average height of enlisters is equal to the average height in the population for individuals with the same observables.

One notable difference between my model and the diagnostic test proposed by Bodenhorn, Guinnane, and Mroz (2017) is that I model the choice of whether or not to enlist in the military as a once-per-lifetime decision whereas they focus on the dynamic decision of military enlistment within a cohort. This is simply a matter of two different approaches to the same problem. Fundamentally, Bodenhorn, Guinnane, and Mroz (2017, p. 173) are concerned, as am I, with a situation "in which an individual enters the sample, in part, due to the unmeasured characteristics that are related to the outcome of interest." Their focus on the relationship of military enlistment with within-cohort changes in economic conditions is intended as a way to diagnose the presence of sample-selection bias using only the selected sample, which is not the fundamental concern of their paper. Thus, although I do not directly address Bodenhorn, Guinnane, and Mroz's (2017) specific concern regarding changing stature over time within a birth cohort, I do address the broader concern that unobservables determined both height and enlistment, generating time-varying sample-selection bias.

## 2.2 The Empirical Challenge

The goal of this paper is to learn the unconditional average height of birth cohort $t$, $E(h_{it}|t)$, for all $t \in \{1832, \dots, 1860\}$. If the heights of a random sample of the population were observed, it would be possible to estimate $E(h_{it}|t)$ simply by computing averages for each cohort or by regressing heights on a series of birth cohort indicators (with no controls). However, because height data are available only for military enlisters, and because non-random selection into military service generates sample-selection bias, it is impossible to accurately estimate $E(h_{it}|t)$ by this simple approach. Such selection comes in two forms—selection on observables and selection on unobservables.[7]

Selection on observables stems from the impact of $\mathbf{x}_{it}$ on both military enlistment and height. If observable characteristics impact the probability of military enlistment, then their distribution in the military will be different from their distribution in the population. If these characteristics also affect height, then the distribution of military heights will also differ from that of the population. For example, if residents of urban areas are both shorter than the population average and more likely to join the military, then they will be over-represented in the military, which will, in turn, be shorter than the whole population. If all bias is due to selection on observables (i.e., $\varepsilon_{it}$ and $u_{it}$ are uncorrelated, so $\Omega(\cdot) = 0$), then equation (4) shows that selection is random, conditional on observables. Thus, the average height of enlisters with a given set of observable characteristics is equal to the average height of the population with the same observables. Such selection can then be addressed by re-weighting the military data so that its distribution of observables matches that of the population. This is a standard approach in the historical heights literature (e.g., Fogel et al., 1983, p. 454) and can be achieved by computing weights from aggregate data or (as I do below) from estimates of conditional enlistment probabilities from equation (3).[8]

The formal basis for selection on unobservables is also illustrated in equation (4). Enlisters' average heights, conditional on all observables, differ from those of the whole population by $\Omega(\alpha_t + \mathbf{x}_{it}'\beta_k + \mathbf{z}_{it}'\delta_k)$, which is non-zero when $\varepsilon_{it}$ and $u_{it}$ are correlated with one another. Such a correlation might arise if, for instance, an unobserved adverse health shock in childhood harmed an individual's labor market prospects, making him more likely to enter the military, and also reduced his terminal height relative to others with his same observable characteristics. Re-weighting cannot address this bias because re-weighting requires selection into the military to be random, conditional on observables; in this case, the military will over-represent shorter individuals for a given set of covariates. Addressing selection on unobservables requires the estimation of $\Omega(\alpha_t + \mathbf{x}_{it}'\beta_k + \mathbf{z}_{it}'\delta_k)$ through the estimation of equation (4) for the selected sample. Once estimated, this term can be removed, leaving the adjusted average heights of enlisters equal to the average height of the population with the same observables. Any selection on observables can then be addressed by

[7] Online Appendix C shows formally how each type of selection affects naive estimates of average height and how such estimates can be corrected.

[8] Bodenhorn, Guinnane, and Mroz (2017) also discuss weighting approaches. Another common approach is to include these observable characteristics as controls in a regression of height on birth cohort indicators (e.g., Margo and Steckel, 1983; Zehetmayer, 2011). This approach also eliminates the problem of selection on observables, but estimates the conditional trend in heights $\gamma_t$ rather than the unconditional trend $E(h_{it}|t)$. My estimation strategy also adjusts the conditional trend for selection on unobservables.

re-weighting. I discuss the intuition of this approach below.

While sample-selection bias may cause the average height of military enlisters to differ from the average height of the population, it does not bias a naively estimated trend in average height unless its magnitude changes over time. For instance, if selection on unobservables caused military enlisters to always be one inch shorter than the population, then the trend in heights of military enlisters would be the same as the trend in heights of the population. Bodenhorn, Guinnane, and Mroz (2014) argue that selection on unobservables might have changed over time due to economic growth in the antebellum period: as economic growth improved the attractiveness of the civilian sector relative to the military sector, only those with increasingly poor civilian labor market opportunities would choose enlist; if success on the civilian labor market was positively correlated with height (for instance, if both were affected by childhood health shocks, as in the example above), then enlisters would be more negatively selected over time.

## 2.3 Intuition of the Correction for Selection on Unobservables

The historical heights literature typically corrects for selection on observables but not for selection on unobservables (see detailed discussion by Bodenhorn, Guinnane, and Mroz, 2017, pp. 173–174, 187–189). In so doing, the literature typically assumes (implicitly or explicitly) that $\varepsilon_{it}$ and $u_{it}$ are uncorrelated. In this paper, I impose no such assumption, allowing $\varepsilon_{it}$ and $u_{it}$ to be potentially correlated. I then adjust observed heights for sample-selection bias induced by selection on both observables and unobservables using the methods discussed above.

How is it possible to recognize and correct for sample-selection bias induced by selection on unobservables? The key insight is that the magnitude of the bias will vary with the probability of military enlistment conditional on observables, as defined in equation (3).[9] To see this, suppose that there is negative selection on unobservables (i.e., a negative correlation between $u_{it}$ and $\varepsilon_{it}$) that arises, for example, from unobserved adverse health shocks in childhood making the military relatively more attractive and reducing terminal height. Suppose that there is one group of potential enlisters whose probability of enlistment conditional on observables is close to zero (i.e., some combination of observable characteristics in the population that is almost never observed to enlist). For concreteness, this group might be thought of as the children of wealthy craftsmen.[10] These individuals enlist only if they experienced a particularly strong adverse health shock, and thus the members of this group observed enlisting are shorter than the whole population with the

[9] This is shown in equation (4), in which the bias $\Omega(\alpha_t + \mathbf{x}'_{it}\beta_k + \mathbf{z}'_{it}\delta_k)$ is a function of the same linear index as is the enlistment probability $G(\alpha_t + \mathbf{x}'_{it}\beta_k + \mathbf{z}'_{it}\delta_k)$ in equation (3).

[10] Of course, the true conditional enlistment probability in this group was not actually close to zero (nor was it close to one in the group to be discussed below), but it is helpful to think of extremes to develop the intuition.

same observable characteristics. On the other hand, suppose that there is a group of men whose conditional enlistment probability is close to one (i.e., some combination of observable characteristics in the population that is almost always observed to enlist). For concreteness, this group might be thought of as the children of poor unskilled laborers, for whom enlistment is more attractive than their civilian labor market options. They enlist regardless of their childhood health, and enlisters from this group are approximately the same height as the whole population with the same observable characteristics.

Thus, in the presence of negative selection on unobservables into the military height sample, there would be a positive correlation between the probability of enlistment implied by observables, and height, after conditioning on observables. The presence of such a correlation is how selection on unobservables is recognized, and precisely what is tested for by including $\Omega(\alpha_t + \mathbf{x}'_{it}\beta_k + \mathbf{z}'_{it}\delta_k)$ in equation (4). The precise nature of the relationship of height with the enlistment probability identifies (in the statistical sense) $\Omega(\cdot)$, enabling the determination of the magnitude of sample-selection bias.

Why are the exclusion restrictions required? That is, why is it necessary to include $\mathbf{z}_{it}$ in the military enlistment decision but not the height determination equation, or to allow $\beta$ and $\delta$ to differ by cohort group? The example above showed that selection on unobservables creates a correlation between the probability of enlistment and height, after conditioning on observables. But if enlistment probability is a function only of observables affecting height, then there is no variation in this probability after conditioning on observables, and thus no correlation can be found. The exclusion restrictions provide variation in enlistment probability while conditioning on all observables affecting height. In other words, they create different enlistment probabilities for individuals whose height-determining covariates are the same, implying that their average heights should be the same if there is no selection on unobservables.

To be concrete, let $\mathbf{z}_{it}$ represent (as it will below) political ideology. If ideology does not affect height, then all individuals with the same observables (other than ideology) should be of the same average height regardless of ideology. But negative selection on unobservables would cause individuals whose ideology renders them relatively likely to enlist to be observed to be taller than otherwise identical individuals whose ideology renders them relatively unlikely to enlist. As in the example above, individuals whose ideology renders them relatively unlikely to enlist do so only if they experienced a strong adverse health shock, whereas those whose ideology renders them relatively likely to enlist do so even without such shocks.

Permitting the effect of covariates on enlistment probability to differ by cohort group but ruling out such a difference in the effect on height gives a similar advantage. Whenever the correlation of height with a covariate differs between cohort groups, this is evidence of selection on unobservables. For instance, if the conditional urban height penalty differs between cohort groups, this indicates the presence of sample-selection bias because equation (1) permits no such variation; the only way that it could arise is if urban residence had a different impact on enlistment probability across the two cohort groups, and if this different enlistment probability led to different magnitudes of selection on unobservables.

In light of this intuition, it is instructive to consider how my model admits changing bias from selection on unobservables over birth cohorts. Because $\Omega(\cdot)$ is a function of the same single index $\alpha_t+\mathbf{x}_{it}'\beta_k+\mathbf{z}_{it}'\delta_k$ as is the conditional enlistment probability, selection changes whenever conditional enlistment probability changes. There are therefore three mechanisms by which selection can change over cohorts: changes in the distribution of $\mathbf{x}_{it}$ and $\mathbf{z}_{it}$ over cohorts, differences between cohorts in the cohort-specific intercept $\alpha_t$, and variation in $\beta_k$ and $\delta_k$ across cohort groups. As discussed above, only changes in enlistment probability generated by the exclusion restrictions are useful in identifying $\Omega(\cdot)$ (in a statistical sense);[11] but any differences in enlistment probability over birth cohorts (generated by any of these three sources of variation) are translated into differences in sample-selection bias induced by selection on unobservables.[12]

# 3 Data Sources

I estimate my model with a combination of three types of data. Military records are the foremost source of height data in the United States. Census data enable the estimation of the binary choice model for military enlistment (equation 3) through the comparison of the characteristics of individuals observed in the military to those of the population at risk for enlistment. Finally, voting data are useful for identification.

Researchers estimating sample-selection models typically collect data on a random sample of the population. In this context, such a data set would enable me to observe $\mathbf{x}_{it}$, $\mathbf{z}_{it}$, and $y_{it}$ for the whole sample and $h_{it}$ for those with $y_{it}=1$. Estimation of the binary choice model (equation 3) would be carried out by comparing the distribution of covariates for enlisters ($y_{it}=1$) to that of non-enlisters ($y_{it}=0$). The construction of such a data set is not possible in this context. While it is possible to learn the distribution of covariates for enlisters through the linkage of military records and census data (as discussed below), the fact that only a

[11] This is similar in interpretation to a local average treatment effect. It is therefore potentially a source of divergence from Bodenhorn, Guinnane, and Mroz's (2017) specific concern over changing labor market conditions. While these conditions would be captured by the $\alpha_t$ and would lead to changing sample-selection bias over time, they are not used to identify $\Omega(\cdot)$. It is possible that if (somehow) identification were based on these differences rather than on differences induced by political ideology, results would differ.

[12] The assumptions that allow the correction for selection on unobservables to be undertaken, intuitively require the standard uncorrelatedness of errors and covariates. It is therefore important to understand how violation of this assumption, due to the inability to observe variables that might belong in these equations, might jeopardize results. As long as it is possible to uncover some conditional-on-observables relationship between stature and enlistment probability, it is possible to learn about the magnitude of the sample-selection bias. Thus, as long as the most comprehensive set of covariates available is included in the the estimation of equation (4), the best correction possible is performed.

fraction of the military records for the period have been digitized makes it impossible to definitively identify individuals who did not enlist, and thus to learn the distribution of covariates for this group. Nonetheless, Cosslett (1981) has shown that my model can be estimated using a data set consisting of two types of samples that can be constructed using the available sources. The first is a *choice-restricted sample* of individuals located in military records. These individuals are known to have enlisted in the military, and census linkage provides information on their covariates.[13] The second is a *supplementary sample* of the population at risk for military enlistment, which uses census data to characterize the distribution of covariates in the whole population but has no information on military enlistment status. Equation (3) is estimated by comparing the distribution of covariates in the choice-restricted sample to that of the supplementary sample; the only remaining information necessary for estimation is the fraction of the population joining the military, which is computed from external data (see Online Appendix D). Equation (4) is estimated using the choice-restricted sample (and parameters estimated in equation 3) only.

## 3.1 Military Height Data

I collected military height data for the 1832–1860 birth cohorts from two sources that have previously been used to study the Antebellum Puzzle (e.g., Fogel, 1986, Table 9.6). The first is the Union Army Project (Fogel et al., 2000), which provides information collected at the time of entry into the Union Army (during the Civil War). I collected data from this source for individuals born between 1832 and 1846, for whom height, birth year, and age at height measurement are known.[14] It is not possible to use the Union Army data to extend the height series past the birth cohort of 1846 because younger birth cohorts would have been too young to serve during the Civil War (and the Union Army was disbanded after the war).[15] To extend the height series, I collected records of enlistments in the Regular US Army for the birth cohorts of 1847–1860 from the *Register of Enlistments*. This source contains the records of enlistments occurring between 1798 and 1914, though the birth cohorts of 1847–1860 largely enlisted after the Civil War in the 1860s, 1870s, and 1880s. To make my results comparable to those of most other studies of military stature in the United States, I restrict attention to native-born whites and exclude individuals born in the West region

[13]This is called a choice-restricted sample because it contains only individuals who chose to enlist.

[14]The type of enlistment is recorded for only about half of those in the Union Army sample. Among these about 86 percent are volunteers, 9 percent draftees, and 5 percent substitutes. In principle, studying only draftees should solve the selection problem. However, it is known that draftees were not a representative sample of the population because of the possibility of hiring a substitute. As a result, I do not distinguish between the different types of enlisters; this simply requires the interpretation of equation (3) as describing the binary event of being in the military.

[15]Some members of the 1847 birth cohort may have enlisted in the Union Army at age 18. However, given the end of the Civil War relatively early in 1865, I focus on enlisters from the 1846 cohort and earlier only from the Union Army data.

of the United States (e.g., Fogel, 1986; Zehetmayer, 2011).[16] I also exclude enlistments before age 18.[17]

Understanding the distinction between the Union and Regular Armies is crucial. The Union Army was a special temporary force raised during the Civil War (1860–1865). It was a citizen's army, comprising at some point nearly half of the eligible population, and reaching a peak strength of about one million men (Weigley, 1967, p. 267). Enlisters in the Union Army were drawn from all walks of life and were largely inspired to enlist by patriotism and wartime fervor (Weigley, 1967). The Regular Army is the same professional US Army that exists today. During the Civil War, enlistment in either the Regular or the Union Army was possible; but the vast majority of individuals serving during the war joined the Union Army.[18] Indeed, despite identical pecuniary incentives for enlistment, there was difficulty in maintaining the strength of the Regular Army during the Civil War. To my knowledge, no prior work exists on the composition of enlisters in the Regular Army during the Civil War (historians understandably focus on the Union Army). Between 1865 and 1898, all Army enlistment was in the Regular Army, which never exceeded a strength of 60,000 men (Department of Defense, 1997, Table 2-11) and included only about 2 percent of the eligible population in the 1847–1860 cohorts. Enlisters in the postbellum Regular Army were generally drawn from lower social strata and enlisted largely due to a lack of civilian alternatives (Coffman, 1986; Foner, 1970; Weigley, 1967).

The revealed preference for service in the Union Army when enlistment was possible in both it and the Regular Army is evidence of substantial differences between the two forces. Bernardo and Bacon (1955, pp. 201–206) hypothesize that the Union Army was preferred because, as compared to the Regular Army, it had a shorter term of service, less rigorous discipline, and the freedom to elect officers. Conditions of service in the postbellum Regular Army were also poor in comparison to those experienced in the Union Army (Coffman, 1986, pp. 328–329). Moreover, while volunteers in the Union Army were held in high esteem by the public, the Regular Army (outside of the war years) was more poorly perceived.

These differences underlie my decision to allow the effect of covariates on the military enlistment decision to differ between the 1832–1846 and the 1847–1860 cohort groups: the former group faced wartime patriotic motivations for enlistment and had the option to enlist in the Union Army; the younger cohorts only had the option to enlist in peacetime in the Regular Army. These differences also make concrete precisely the selection problems that concern Bodenhorn, Guinnane, and Mroz (2017) and that this paper addresses.

[16]This entails the removal of a small number of Californians, New Mexicans, and Oregonians.

[17]There is a real possibility that 18 year olds might not yet have reached terminal height (Frisancho, 1993). I address this in the empirical analysis by including measurement-age indicators in any specification in which height is the dependent variable.

[18]According to Department of Defense (1997, Table 2-11), the US Army in 1860 (i.e., the Regular Army) included 16,215 officers and men. In 1861, an additional 22,714 were authorized (Bernardo and Bacon, 1955, p. 202). Bernardo and Bacon (1955, pp. 201–203) discuss the relationship between the Union and Regular Armies during the Civil War, arguing that there was little interaction between the forces, which were largely kept separate.

This clear change in the incentive for military enlistment at the end of the Civil War makes it natural to suspect that the birth cohorts old enough to serve in the Civil War (the 1832–1846 cohorts) might have been differently selected than were those enlisting in the postbellum period (the 1847–1860 cohorts), even after conditioning on their observable characteristics. In particular, the worse conditions of service and contemporary reports (e.g., Coffman, 1986, p. 329) of the poor labor market attributes of enlisters suggest that selection into military service may have become more negative after (as compared to during) the war.

Selection into military service may also have differed between birth cohorts within each of these cohort groups. With the enlistments in the Union Army taking place over a short period and attracting individuals from many different birth cohorts (and thus ages), it is possible that enlistment may have appealed differently to individuals of different ages. Similarly, postbellum enlistment in the Regular Army may have appealed differently to individuals in different birth cohorts depending on the state of the civilian labor market in their prime years (as Bodenhorn, Guinnane, and Mroz, 2014 argue). Drawing a consistent trend in stature over time despite these changes is difficult. This is the heart of Bodenhorn, Guinnane, and Mroz's (2017) concern, and addressing it by permitting selectivity to differ is the contribution of the present paper.

## 3.2 Census Data

Data on the covariates $\mathbf{x}_{it}$ for military enlisters were collected through linkage of their military data with US census records from their childhood and adolescence. Enlisters in the Union Army sample have already been linked to the US Censuses of 1850 and 1860 by Fogel et al. (2000), but no previous linkage exists for the Regular Army enlisters whose information I extracted from the *Register of Enlistments*. I therefore linked the latter group to the US Censuses of 1860–1880 (using the procedure in Online Appendix E.1) and transcribed census information for the linked individuals and their households. For both the Union Army and the Regular Army I retain only individuals for whom census information could be located. Except for height and age of height measurement, all data pertain to the individual or his household as observed in the census in which he was aged 9–18, or to his county of residence in that census.[19] Census linkage provided information on the property ownership of the enlister's household, his place of residence, the size and composition of his household, the occupations of the members of his household, and his school attendance.

The use of census-linked data raises two concerns. One is that non-random failure to link enlisters to the census may introduce bias—that is, the linked data might not be representative of the military data as a whole. I study this possibility in Online Appendix E.2. Based on the fact that the trends in stature in the

[19] I make this limitation because it allows me to ensure that I observe individuals prior to enlistment.

linked data and in the complete collection of military data are nearly identical, I conclude that my results are unlikely to be affected by selection into the linked sample.[20] Another concern stems from the fact that, unlike the Union Army data, which are hand-linked by genealogists, my linkage of the Regular Army data is automated. Bailey et al. (2017) have recently shown that the rates of false links arising from automated linkage may be high.[21] To address this concern, I repeat my main analysis limiting the data to either hand matches (i.e., all of the Union Army data, where false positives are not a concern) or to exact automated matches, which are less likely to generate false positives. This exercise is discussed in Online Appendix G, where I show that limiting the sample in this way does not meaningfully affect the results.

The census-linked military data form the two choice-restricted samples—one for the 1832–1846 birth cohorts, generated by linking Union Army enlisters to the census, and one for the 1847–1860 birth cohorts, generated by linking the Regular Army enlisters to the census. For each of these two samples, a supplementary sample of the covariates of the population at risk for enlistment is required. To create such samples, I collected information on the covariates $\mathbf{x}_{it}$ for a random sample of the population at risk for enlistment for the 1832–1860 birth cohorts from the public use samples of the 1850, 1860, and 1870 censuses (Ruggles et al., 2015). I again restrict attention to native-born (outside of the West region) white males observed between ages 9 and 18. Dividing the random census sample along the same birth cohorts creates two supplementary samples.

For both the choice-restricted and supplementary samples I limit attention to individuals residing (at ages 9–18) in either non-seceding states or Virginia (i.e., I omit residents of Confederate states other than Virginia). I impose this restriction because residents of the excluded states are only rarely observed in Union Army, making it difficult to estimate their average heights. I impose the same restriction on the Regular Army data, which would have included individuals from these states, for comparability.

I also collected county-level data on agricultural production and population from Manson et al. (2017).

### 3.3 Voting Data

The Civil War was fought over the issues of slavery and preservation of the Union. Similarly, after the Civil War, one of the Regular Army's main duties was Reconstruction—the military occupation of the South. It is not hard to imagine that military service during these periods might have been attractive to those who had been opposed to slavery or supportive of preservation of the Union. Although it is not possible to observe

[20] I find selection into linkage on observables, and apply an inverse-probability weighting approach to correct for this.

[21] Part of this concern is mitigated by my avoidance of the Soundex algorithm to standardize names (one of the main criticisms of Bailey et al., 2017), my focus on individuals with unique characteristics (who would be less likely to be falsely matched), and my exclusion of any record with multiple possible matches (though I permit one census record to have multiple matches in the enlistments to reflect the potential for multiple enlistments).

political ideology, it is possible to observe a proxy—county-level voting patterns in the US Presidential Election of 1860, which also centered on the issues of slavery and preservation of the Union. Thus, it is likely that these voting patterns are informative regarding the military enlistment decision.

Election data have previously been used to predict military enlistment and desertion in the mid-nineteenth-century United States. Costa and Kahn (2003) relate voting data from the elections of 1856 and 1860 to the probability of desertion from the Union Army, finding that enlisters from counties with greater support for Republican candidates were less likely to desert. Similarly, Costa and Kahn (2007) use voting data from the 1864 election to measure a community's support for the Civil War, finding that deserters from communities with greater support for the war were more likely to migrate after the war, and were more likely to settle in more anti-war communities, as measured by voting. The relevance to desertion suggests relevance to enlistment. Eli, Salisbury, and Shertzer (2018) show that voting patterns in 1860 are predictive of the enlistment decisions of Kentuckians in the Civil War and of the subsequent migration of Civil War veterans.

Based on the likely impact of ideology, as measured by voting patterns, on military enlistment, I use county-level voting data for the Presidential Election of 1860 (ICPSR, 1999) for identification of the sample-selection model (that is, to act as the excluded variable $\mathbf{z}_{it}$). In particular, I focus on the fraction of the individual's county of residence (at the time that he is observed in the census between ages 9 and 18) voting for Abraham Lincoln in 1860. To be a valid exclusion restriction, this variable must satisfy two conditions. First, it must be related to military enlistment—that is, it must actually generate variation in the probability of enlistment. Given the discussion above, and the fact that Lincoln represented one extreme on the issue of slavery, it is plausible that voting patterns in this election should be related to military enlistment. The second requirement is that ideology (as proxied by voting) should be excludable from the determination of height; that is, conditional on all the observed covariates of height, voting patterns should be unrelated to height. Intuitively, the fact that I control for the available socioeconomic variables makes it likely that any factors that would cause voting and health to be correlated will be captured. Both relevance and excludability will be formally explored in the empirical analysis.

## 4 Summary Statistics

In Table 1, I summarize the structure of the sample, including information on the censuses from which each cohort's data are drawn. Because I draw each individual's census information from the census for which they are between ages 9 and 18 years old, individuals from the birth cohorts of 1832–1841 are observed in 1850, those from the birth cohorts of 1842–1851 are observed in 1860, and those from the cohorts of 1852–1860 are observed in 1870.[22] In this Table and throughout the paper I use the shorthand "UA" to refer to the Union Army and "RA" to refer to the Regular Army (1847–1860 cohorts). I abbreviate Cosslett's (1981) two terms by using "CR" to refer to a choice-restricted sample and "Supp." to refer to a supplementary sample.

Table 1: Distribution of observations by census and sample

| | | 1832–1846 | | 1847–1860 | |
|---|---|---|---|---|---|
| | | (1) | (2) | (3) | (4) |
| Census | Cohorts | UA (CR) | Supp. | RA (CR) | Supp. |
| 1850 | 1832–1841 | 3,347 | 5,879 | | |
| 1860 | 1842–1851 | 2,435 | 2,807 | 991 | 3,063 |
| 1870 | 1852–1860 | | | 1,477 | 6,208 |
| Total | | 5,782 | 8,686 | 2,468 | 9,271 |

*Notes*: Each cell reports the number of individuals in the sample indicated in the column header with data taken from the census indicated in the row. Samples are restricted to cover individuals with data on all individual-level variables. Abbreviations are as follows: UA is Union Army, RA is Regular Army, CR is choice-restricted sample, Supp. is supplementary sample.

In Figure 1 I present the distributions of heights in each military sample. The Union Army enlisters were statistically significantly taller than the Regular Army enlisters by 0.816 inches. More stringent enforcement of the minimum height requirement of 64 inches in the Regular Army than in the Union Army is also clear.

(a) Union Army, 1832–1846

(b) Regular Army, 1847–1860

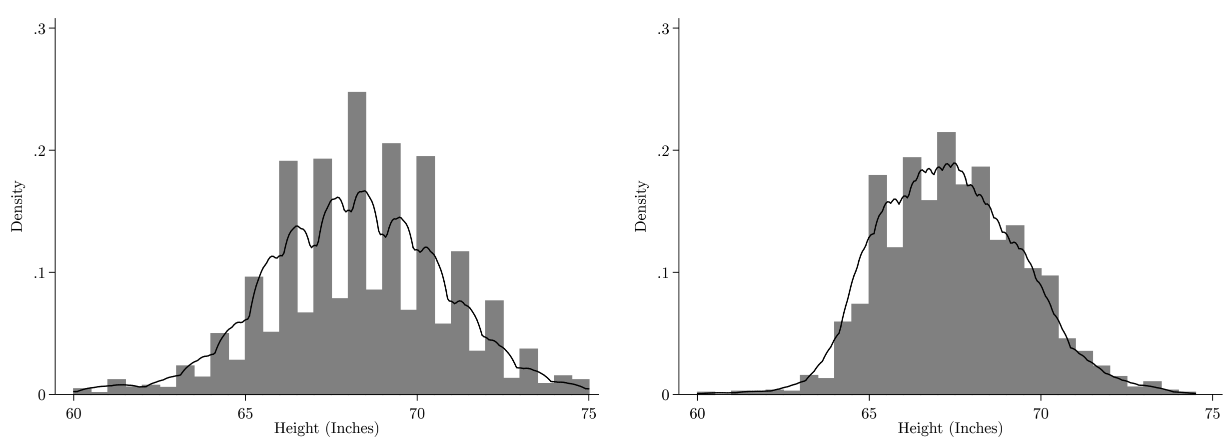



Figure 1: Height distributions

*Note:* These figures present histograms (with a bin width of 0.5 inches) and kernel density estimates of the height distributions for the two military height samples. Panel 1(a) covers the 1832–1846 birth cohorts (using Union Army data) while panel 1(b) covers the 1847–1860 cohorts (using Regular Army data).

[22]The Table should be read as follows, taking the second row as an example: there were 2,435 Union Army enlisters born 1842–1846 who were linked to the 1860 census, 991 Regular Army enlisters born 1847–1851 who were successfully linked to the 1860 census, 2,807 individuals born 1842–1846 drawn from the 1860 census without regard to their enlistment status, and 3,063 individuals born 1847–1851 drawn from the 1860 census without regard to enlistment status.

I collected data from the census on the property ownership of each individual's household (expressed in 1860 dollars, using deflators of Lindert and Margo, 2006), the composition of the household (including its size and whether the individual of interest, either the enlister or prospective enlister, was related to its head), the fraction of the household's county living in an urban area, and whether the individual of interest attended school in the year prior to observation. The occupations of each member of the individual's household were gathered and classified according to the system used by the Union Army Project (Fogel et al., 2000). The household is classified by the highest occupational status of any member; for example, if one member of the household is professional and the other is clerical, the household is categorized as professional. In addition, the birth region of the potential enlister is classified by region (Northeast, Midwest, and South). The sample is restricted to include only native-born whites, so individual nativity and race are not of interest.

Table 2: Summary statistics

| | 1832–1846 | | | 1847–1860 | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| *Variable* | UA (CR) | Supp. | Diff. | RA (CR) | Supp. | Diff. |
| *Individual or Household Variables* | | | | | | |
| Height (in) | 68.175 | | | 67.276 | | |
| Household Owns Property | 0.716 | 0.687 | 0.028* | 0.810 | 0.734 | 0.076*** |
| Household Real Property ($1,000) | 1.775 | 2.297 | −0.525*** | 2.092 | 2.436 | −0.369** |
| Related to Head of Household | 0.892 | 0.863 | 0.029*** | 0.875 | 0.896 | −0.019* |
| Household Size | 7.490 | 7.419 | 0.074 | 6.987 | 7.024 | −0.033 |
| Attended School | 0.573 | 0.648 | −0.076*** | 0.631 | 0.726 | −0.095*** |
| Household Occupation | | | | | | |
| Farmer | 0.484 | 0.520 | −0.036* | 0.282 | 0.428 | −0.142*** |
| Professional | 0.029 | 0.038 | −0.008** | 0.046 | 0.037 | 0.009* |
| Clerical | 0.030 | 0.066 | −0.036*** | 0.090 | 0.078 | 0.005 |
| Skilled and Artisan | 0.144 | 0.185 | −0.041*** | 0.246 | 0.169 | 0.078*** |
| Semi-Skilled and Operative | 0.050 | 0.058 | −0.008 | 0.105 | 0.099 | 0.006 |
| Unskilled | 0.048 | 0.065 | −0.017*** | 0.108 | 0.079 | 0.029*** |
| Farm Labor | 0.006 | 0.006 | 0.001 | 0.033 | 0.042 | −0.008 |
| Unproductive | 0.208 | 0.062 | 0.145*** | 0.090 | 0.068 | 0.022*** |
| Birth Region | | | | | | |
| Midwest | 0.461 | 0.287 | 0.173*** | 0.271 | 0.406 | −0.130*** |
| Northeast | 0.450 | 0.541 | −0.092*** | 0.559 | 0.439 | 0.131*** |
| South | 0.089 | 0.172 | −0.081*** | 0.169 | 0.154 | −0.001 |
| *County Variables* | | | | | | |
| Fraction Urban | 0.087 | 0.158 | −0.070*** | 0.344 | 0.237 | 0.095*** |
| Wheat Bushels per capita | 8.625 | 5.882 | 2.743*** | 6.400 | 8.570 | −2.171*** |
| Milk Cows per capita | 0.299 | 0.282 | 0.017** | 0.224 | 0.258 | −0.034*** |
| Swine per capita | 1.213 | 0.972 | 0.241*** | 0.510 | 0.716 | −0.206*** |
| Value of Agricultural Production per capita | 0.053 | 0.048 | 0.005*** | 0.050 | 0.061 | −0.011*** |
| Lincoln Vote Share (1860) | 0.508 | 0.455 | 0.053*** | 0.473 | 0.449 | 0.024 |
| Observations | 5,140 | 8,535 | | 2,174 | 8,828 | |

*Significance levels*: *** p<0.01, ** p<0.05, * p<0.1
*Notes*: Averages for the choice-restricted samples are weighted to correct for selection into linkage on the basis of observable characteristics. Standard deviations and standard errors are omitted for clarity. Sample sizes are the minimum of the column with observations for all variables. Abbreviations are as follows: UA is Union Army, RA is Regular Army, CR is a choice-restricted sample, and Supp. is a supplementary sample. Diff. is a difference.

Table 2 summarizes some of the individual- and household-level data taken from the census for each of the choice-restricted and supplementary samples, together with the voting data. Columns (1) and (4) present information for the choice-restricted samples of military enlisters. Columns (2) and (5) present information for the supplementary random samples of census information from the population as a whole. Columns (3) and (6) present $t$-tests of the difference between the supplementary and the choice-restricted samples for each of the military enlistment samples. Nearly all of the $t$-tests for the differences between the enlisters and the general population indicate statistically significant differences between enlisters and the general population at the one-percent level, and these differences are largely consistent with expectations. For instance, urban areas are over-represented in the Regular Army, as contemporary reports suggest that they should be. These differences also extend to the voting data, with both military samples drawn disproportionately from Lincoln-supporting areas (though the difference in the Regular Army is not statistically significant).

# 5 Estimation

The estimation of the trend in stature, incorporating the correction for selection on observables and unobservables, proceeds as follows. Steps 1 and 2 are essentially the two-step Heckman (1979) procedure. Step 3 computes the unconditional trend, which is smoothed in step 4.

1. Equation (3) is estimated semi-parametrically. The structure of the sample requires that I use an adapted Klein and Spady (1993) estimator, described in Online Appendix F. This yields estimates of the conditional enlistment probability $\hat{G}(\hat{\alpha}_t + \mathbf{x}_{it}'\hat{\beta}_k + \mathbf{z}_{it}'\hat{\delta}_k)$, and of its linear index $\hat{\alpha}_t + \mathbf{x}_{it}'\hat{\beta}_k + \mathbf{z}_{it}'\hat{\delta}_k$.

2. Equation (4) is estimated on the choice-restricted sample using Newey's (2009) method. To take into account the possibility that individuals in their late teens or early twenties might not yet have reached terminal height (Frisancho, 1993), I add to equation (4) a vector of measurement-age indicators $\mathbf{m}_{it}$ with coefficients $\pi$ to normalize heights to age 21. Estimation of equation (4) is weighted to account for the separate sampling of the two groups of birth cohorts. This yields an estimate $E(h_{it}|\mathbf{x}_{it}; t) = \hat{\gamma}_t + \mathbf{x}_{it}'\hat{\theta}$ and of the selection bias, $\hat{\Omega}(\hat{\alpha}_t + \mathbf{x}_{it}'\hat{\beta}_k + \mathbf{z}_{it}'\hat{\delta}_k)$.[23]

[23] Leaving the form of $\Omega(\cdot)$ free in equation (4) rather than assuming joint normality of $\varepsilon_{it}$ and $u_{it}$ implies that $\gamma_t$ and $\Omega(\cdot)$ are estimated only up to a constant. An intercept can be estimated by Andrews and Schafgans's (1998) method as

$$\hat{\mu} = \frac{\sum_{i=1}^{N} \Gamma(\hat{\alpha}_t + \mathbf{x}_{it}'\hat{\beta}_k + \mathbf{z}_{it}'\hat{\delta}_k)(h_{it} - \hat{\gamma}_t - \mathbf{x}_{it}'\hat{\theta} - \mathbf{m}_{it}'\hat{\pi})}{\sum_{i=1}^{N} \Gamma(\hat{\alpha}_t + \mathbf{x}_{it}'\hat{\beta}_k + \mathbf{z}_{it}'\hat{\delta}_k)},$$

where $\Gamma(\cdot)$ is a weighting function. Because this estimate is likely to be imprecise, because Online Appendix E.2 shows that it may be contaminated by selection into census linking, and because it does not play a role in comparisons across regions and sectors, I do not emphasize this estimation.

3. I estimate the selection-corrected average stature for cohort $t$ by computing

$$\hat{h}_t = \frac{\hat{k}_t}{N_t} \sum_{i \in t} \frac{h_{it} - \mathbf{m}'_{it}\hat{\pi} - \hat{\Omega}(\hat{\alpha}_t + \mathbf{x}'_{it}\hat{\beta}_k + \mathbf{z}'_{it}\hat{\delta}_k) + \hat{\mu}}{\hat{G}(\hat{\alpha}_t + \mathbf{x}'_{it}\hat{\beta}_k + \mathbf{z}'_{it}\hat{\delta}_k)}. \tag{5}$$

   This is accomplished by a regression of the selection- and measurement age-corrected height, $h_{it} - \mathbf{m}'_{it}\hat{\pi} - \hat{\Omega}(\hat{\alpha}_t + \mathbf{x}'_{it}\hat{\beta}_k + \mathbf{z}'_{it}\hat{\delta}_k) + \hat{\mu}$, for each member of the choice-restricted sample on birth-cohort indicators, weighting by inverse enlistment probability (i.e., by $\hat{G}(\hat{\alpha}_t + \mathbf{x}'_{it}\hat{\beta}_k + \mathbf{z}'_{it}\hat{\delta}_k)^{-1}$).[24]

4. I smooth the estimated average stature for each cohort using a kernel regression of $\hat{h}_t$ on birth year.[25]

For steps 1 and 2 standard errors can be computed analytically. For steps 3 and 4, bootstrapping is required.

To determine whether incorporating the correction for selection on unobservables affects results, I must estimate a trend in heights for comparison that does not adjust for selection on unobservables. To do this, I replace step 3 with estimation of a truncated regression of height on birth cohort indicators and measurement indicators with a truncation point of 64 inches (A'Hearn, 1998; Komlos, 1998), weighting by inverse enlistment probability. This approach matches the literature standard of correcting for truncation and for selection on observables, but not for selection on unobservables. I do not use a truncated regression in the main estimation procedure (to adjust for selection on both observables and unobservables) because the correction for selection on unobservables should also address truncation, which is a special case of positive selection on unobservables. In this case the binary choice model for military enlistment represents the compound event in which an individual meets the height requirement and chooses to join the military.[26]

[24]The term $\hat{k}_t$ in equation (5) is the normalizing constant to ensure that the inverse probability weights add to one. The estimate $\hat{\mu}$ in equation (5) is described in footnote 23.

[25]I do this because the anthropometric history literature generally focuses on average stature in bins of more than one cohort.

[26]To identify selection on unobservables from truncation, an exclusion restriction is needed that generates variation in the probability of being prevented from enlistment because of a height restriction, but does not affect height. Figure 1 shows that the minimum height requirement was not as strictly enforced for the Union Army as for the Regular Army. Thus, the exclusion restriction that allows military enlistment to vary by cohort group can give identification. Note that truncation and the suspected impact of health on enlistment and height imply different signs of bias. Just as thinking of binary choice model as describing the compound event of choosing to enlist and meeting the minimum height requirement would cause the coefficients to describe the average effect of covariates on the probability, the model should capture the net selection on unobservables. That is, it will capture whether a marginal change in enlistment probability is associated with taller or shorter average height, conditional on observables. The presence of stronger truncation in the Regular Army suggests that the model may be pre-disposed to finding more positive selection after the Civil War, which is the opposite of what I find.

# 6 Results

## 6.1 Selection into Military Service

The results of estimation of the binary choice model in equation (3) for military enlistment are presented in column (1) of Table 3, with the $\beta_k$ and $\delta_k$ presented in separate sub-columns for each cohort group $k \in \{1832\text{–}1846, 1847\text{–}1860\}$.[27] With the goal of correcting for sample-selection bias, two particular aspects of the results of column (1) of Table 3 are important. The first is the relevance of the vote share variables to the enlistment decision. In both cohort groups, the vote share for Lincoln in the county of residence enters with a statistically significant coefficient, indicating that ideology, as proxied by voting, was indeed relevant to military enlistment. The second aspect of interest in Table 3 is that a test of equality of the coefficients between cohort groups rejects the null hypothesis of equality at all levels of significance, indicating that the military enlistment did indeed vary by cohort group.

Although the coefficients themselves are important to demonstrating the relevance of the exclusion restrictions (because they generate the needed variation in the single index), they are not straightforward to interpret. I therefore present, in column (2) of Table 3, the average semi-elasticities associated with the estimates of column (1).[28] The semi-elasticity of the vote share variable indicates that it has an economically significant effect of the expected sign on the enlistment decision. The semi-elasticity of 1.853 for the 1832–1846 cohorts implies that a one-standard deviation increase in the vote share for Lincoln (approximately 23 percentage points) is associated with a roughly 43 percent increase in the probability of enlistment (e.g., a change of enlistment probability from a mean of 0.45 to about 0.69) in the Union Army. In the 1847–1860 cohorts, the semi-elasticity of 0.332 implies that the same change in Lincoln's vote share is associated with an approximately 7.6 percent increase in enlistment probability in the Regular Army (e.g., a change in enlistment probability from a mean of 2.2 percent to about 2.4 percent). Other semi-elasticities largely reflect the differences between enlisters and the whole population in the summary statistics (Table 2). For instance, school attendance is associated with a lower enlistment probability, as is the value of real property holdings. The fraction of the county of residence that is urban is associated with a higher enlistment probability for the Regular Army (consistent with reports that recruitment efforts were largely concentrated in urban areas), such that an individual from a fully urban county was about 13 percent more likely to enlist than an individual from a fully rural county. Conversely, the fraction of a potential enlister's county of residence that was urban was associated with a lower probability of enlistment in the Union Army.

[27] Household occupation indicators are excluded for clarity. They are included in Table B.1 in the Online Appendix.
[28] Computation of the semi-elasticities is discussed in Online Appendix F.4.

Table 3: Binary choice model estimation

| | (1) | | (2) | |
|---|---|---|---|---|
| | (a) | (b) | (a) | (b) |
| *Cohorts* | '32–'46 | '47–'60 | '32–'46 | '47–'60 |
| *Variables* | UA | RA | UA | RA |
| *Individual or Household Variables* | | | | |
| Household Owns Property | 0.081*** | 0.189*** | 0.232*** | 0.453*** |
| | (0.017) | (0.019) | (0.066) | (0.048) |
| Household Real Property (1,000) | −0.015*** | −0.046*** | −0.051*** | −0.120*** |
| | (0.002) | (0.004) | (0.006) | (0.012) |
| Related to Head of Household | 0.092*** | −0.062*** | 0.243*** | −0.189*** |
| | (0.019) | (0.018) | (0.052) | (0.070) |
| Household Size | 0.017*** | −0.000 | 0.059*** | −0.001 |
| | (0.003) | (0.003) | (0.009) | (0.007) |
| Attended School | −0.177*** | −0.148*** | −0.641*** | −0.501*** |
| | (0.016) | (0.015) | (0.049) | (0.065) |
| Birth Region (South excluded) | | | | |
| Midwest | 0.209*** | −0.054* | 0.947*** | −0.134* |
| | (0.033) | (0.031) | (0.158) | (0.078) |
| Northeast | 0.102*** | −0.046 | 0.343*** | −0.124 |
| | (0.033) | (0.030) | (0.112) | (0.090) |
| *County Variables* | | | | |
| Fraction Urban | −0.123*** | 0.049 | −0.420*** | 0.128* |
| | (0.039) | (0.031) | (0.121) | (0.072) |
| Wheat Bushels per capita | 0.013*** | −0.004*** | 0.043*** | −0.009*** |
| | (0.001) | (0.001) | (0.004) | (0.003) |
| Milk Cows per capita | 0.010 | −0.066 | 0.033 | −0.173 |
| | (0.038) | (0.070) | (0.130) | (0.190) |
| Swine per capita | 0.075*** | −0.076*** | 0.256*** | −0.198*** |
| | (0.011) | (0.017) | (0.032) | (0.043) |
| Value of Agricultural Production per capita | −0.046 | −0.042 | −0.157 | −0.109 |
| | (0.472) | (0.432) | (1.483) | (1.106) |
| Lincoln Vote Share (1860) | 0.544*** | 0.127** | 1.853*** | 0.332** |
| | (0.064) | (0.054) | (0.181) | (0.140) |
| Observations | 13,683 | 11,271 | 13,683 | 11,271 |

*Significance levels*: *** $p<0.01$, ** $p<0.05$, * $p<0.1$

*Notes*: Column (1) presents estimates of the coefficients $\beta$ and $\delta$ from the binary choice model. Column (2) presents the average semi-elasticity of the impact of each variable on enlistment probability as implied by the estimates of column (1). All specifications include cohort indicators and household occupation indicators. Standard errors are clustered at the county level. UA denotes Union Army. RA denotes Regular Army.

## 6.2 Selection-Corrected Height Regressions

The next step is to estimate equation (4), the second-stage selection-adjusted height regression. The results of this estimation are presented in column (1) of Table 4, alongside its unadjusted analog in column (2).[29] The results of the selection-adjusted regression of column (1) are similar to those of the unadjusted regression of column (2), though there are some exceptions. For example, the Northeast's conditional height disadvantage relative to the South decreases after the correction and becomes statistically insignificant. The conditional relationship between height and the fraction of the county's population that is urban is also smaller, though it remains strongly significant. The general similarity of the corrected and uncorrected coefficients would seem to indicate that the selection correction is not impactful; but it should be noted that this Table does not present the cohort-specific intercepts $\gamma_t$, which (as the analysis below will reveal) are impacted.

It is also possible to provide a direct test of the excludability of the vote share for Lincoln. This variable must satisfy two conditions to be used as an exclusion restriction for identification. The first is that it must be relevant to the enlistment decision. This was established in Table 3, in which it was shown that the vote share enters significantly into the enlistment equation (3). The second is that it is excludable from the height equation. That is, omitting the vote share from a regression of height on the covariates $\mathbf{x}_{it}$ in an unselected sample must not lead to omitted variables bias. Because allowing the coefficients of the binary choice model to differ by cohort group is sufficient for identification on its own, it is possible to include the vote share in the second stage to obtain selection-corrected estimates of its relationship with height, thus capitalizing on the over-identification of the model to directly test this assumption. Though this approach has validity as the null and assumes that it is appropriate not to include interactions in the second stage, it is informative to consider the results. Column (3) of Table 4 presents the result of this exercise, while column (4) presents the uncorrected analog. The first item of note is that the relationship of the vote share with height is not statistically significant in column (3). Moreover, the magnitude of the coefficient is small. Its interpretation is that a one standard deviation increase in the vote share for Lincoln (about 23 percentage points) is associated with a 0.07 inch (or less than 0.035 standard deviation) increase in stature. This result contrasts with the uncorrected regression of column (4), which shows a larger but still statistically insignificant coefficient for Lincoln's vote share. This supports the excludability of the vote share.

[29] Household occupation indicators are excluded for clarity. They are included in Table B.2 in the Online Appendix.

Table 4: Height regressions

| *Variables* | (1) Corr | (2) Not | (3) Corr | (4) Not |
|---|---|---|---|---|
| *Individual or Household Variables* | | | | |
| Household Owns Property | 0.092 | 0.104 | 0.100 | 0.104 |
| | (0.094) | (0.111) | (0.096) | (0.111) |
| Household Real Property (1,000) | −0.011 | −0.006 | −0.012 | −0.006 |
| | (0.012) | (0.008) | (0.012) | (0.008) |
| Related to Head of Household | 0.245** | 0.300** | 0.245** | 0.300** |
| | (0.104) | (0.131) | (0.104) | (0.131) |
| Household Size | 0.025* | 0.023 | 0.026* | 0.024 |
| | (0.014) | (0.015) | (0.014) | (0.015) |
| Attended School | 0.038 | 0.044 | 0.026 | 0.038 |
| | (0.077) | (0.084) | (0.079) | (0.084) |
| Birth Region (South excluded) | | | | |
| Midwest | 0.034 | 0.042 | −0.070 | −0.136 |
| | (0.163) | (0.169) | (0.173) | (0.198) |
| Northeast | −0.244 | −0.322* | −0.362** | −0.524** |
| | (0.162) | (0.181) | (0.181) | (0.216) |
| *County Variables* | | | | |
| Fraction Urban | −0.553*** | −0.613** | −0.556*** | −0.630*** |
| | (0.202) | (0.238) | (0.202) | (0.238) |
| Wheat Bushels per capita | −0.001 | 0.004 | −0.001 | 0.003 |
| | (0.007) | (0.006) | (0.007) | (0.006) |
| Milk Cows per capita | 0.008 | 0.080 | −0.016 | 0.048 |
| | (0.206) | (0.196) | (0.212) | (0.205) |
| Swine per capita | 0.077 | 0.064 | 0.094 | 0.087 |
| | (0.051) | (0.051) | (0.057) | (0.054) |
| Value of Agricultural production per capita | −3.964* | −3.758* | −4.305** | −4.284* |
| | (2.047) | (2.257) | (2.089) | (2.320) |
| Lincoln Vote Share (1860) | | | 0.292 | 0.469 |
| | | | (0.300) | (0.315) |
| Observations | 7,249 | 6,873 | 7,249 | 6,873 |

*Significance levels*: *** p<0.01, ** p<0.05, * p<0.1
*Notes*: Standard errors in parentheses. Dependent variable is height, measured in inches. All specifications include age-of-measurement, year-of-birth, and household occupation indicators. The selection-corrected specifications, indicated by the column header Corr, also include the selection-correction function $\Omega(\cdot)$. The uncorrected specifications, indicated by the column header Not, correct for truncation with a truncation point of 64 inches. Standard errors are clustered at the county level. The difference in sample sizes between columns is the result of the need to drop heights below 64 inches in the truncation-corrected regressions when not correcting for sample-selection bias.

### 6.3 Adjusted Trends in Height

I present the results of incorporating the correction for selection on both observables and unobservables in Figure 2. Panel 2(a) presents the smoothed and unsmoothed trends in average stature, either incorporating the correction for selection on both observables and unobservables ("Observables and Unobservables") or adjusting only for truncation and for selection on observables ("Observables Only"). The unsmoothed trends for the 1832–1846 cohorts are based on the Union Army height data, while the unsmoothed trends for 1847–1860 are based on the Regular Army data. For comparability to the existing literature, my main focus is on the smoothed trends. The trend adjusting only for selection on observables represents the state of the art of the historical heights literature, and shows a decline in average stature from 68.27 inches to 66.98 inches. The trend incorporating the correction for selection on both observables and unobservables is the contribution of this paper, and shows a decline in average stature from 68.83 inches to 68.19 inches. A 95 percent confidence interval for the decline in average height implied by the smoothed trend incorporating the correction for selection on both observables and unobservables is presented in panel 2(b).

(a) Estimated adjusted and unadjusted trends

(b) Confidence interval of adjusted trend

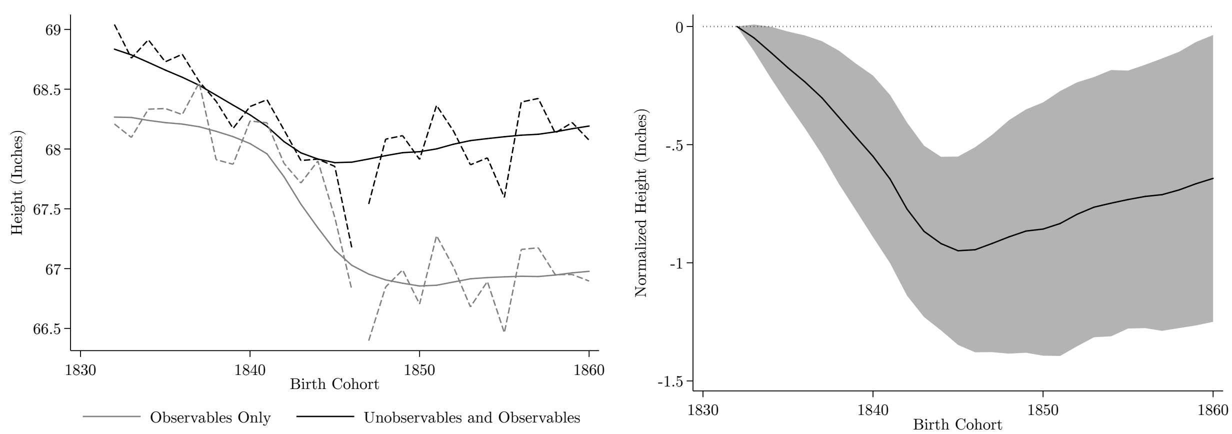



Figure 2: Trends in average stature

*Note:* Panel 2(a) plots four trends in average height by birth cohort. The first, in solid black (labeled "Unobservables and Observables"), incorporates the correction for selection on both observables and unobservables, and smoothed over birth cohorts; the second, in dashed black, is its unsmoothed analog. The third, in solid gray (labeled "Observables Only"), is corrected only for truncation and selection on observables, and is smoothed over birth cohorts; the fourth, in dashed gray, is its unsmoothed analog. The unsmoothed trends for the 1832–1846 cohorts are based on the Union Army data, while those for the 1847–1860 cohorts are based on the Regular Army data. Panel 2(b) presents bootstrap 95 percent pointwise confidence intervals clustered at the county level for the smoothed trend in average stature incorporating the correction for selection on both observables and unobservables (the solid black line in panel 2(a)).

Two key insights can be drawn from this Figure. The first is that my estimated trend incorporating the correction for selection on both observables and unobservables exhibits an Antebellum Puzzle. In the birth cohorts of 1832–1846, the estimated decline in average stature after adjusting for selection on both observables and unobservables and smoothing is 0.94 inches, and is statistically different from zero ($\chi^2_1 = 61.66$, $p < 0.01$). Moreover, the estimated smoothed and adjusted decline in average stature over the birth cohorts of 1832–1860 (i.e., the whole study period) is 0.64 inches, and is statistically different from zero ($\chi^2_1 = 4.43$, $p = 0.04$). I therefore conclude that the view, that the decline in average stature of the Antebellum Puzzle is an artifact of sample-selection bias, is not supported by the evidence. Of course, even if I had found no evidence of a decline in average heights, that would not constitute sufficient evidence to conclude that the Antebellum Puzzle was resolved. In this case, it would still be necessary to explain why stature did not increase in the presence of rapid economic growth. To resolve the puzzle based on selection alone, the trend incorporating the correction for selection on unobservables would have to show an increase in average stature.

The second key insight, evident in panel 2(a), is that incorporating the correction for selection on unobservables yields an estimated trend in average stature that is meaningfully different from the trend estimated according to the literature's state-of-the-art techniques (i.e., adjusting only for truncation and selection on observables).[30] In particular, when adjusting only for selection on observables, a decline of about 1.24 inches in the birth cohorts of 1832–1846 is evident in the smoothed trend, along with a net decline of about 1.29 inches in the birth cohorts of 1832–1860. Both of these estimates are larger than those reached when incorporating the correction for selection on unobservables (0.94 and 0.64 inches, respectively), though only the decline in average heights for the birth cohorts of 1832–1860 is statistically different between the two curves in panel 2(a) ($\chi^2_1 = 5.50$, $p = 0.02$). Thus, the general argument, that failing to properly account for sample-selection bias may lead to biased estimates of the trends in height over birth cohorts, is supported. Indeed, the difference between the two curves in panel 2(a) indicates that addressing selection on unobservables reduces by about half the estimated decline in average stature for the birth cohorts of 1832–1860.

Panel 2(a) further indicates that the difference between the estimated trends in stature is largely the product of distinctly different levels of sample-selection bias between the Union and the Regular Armies, and thus between the two portions of the sample.[31] This is evidenced in part by the fact that the decline in

[30] The difference between the "Observables Only" and "Unobservables and Observables" trends in panel 2(a) is not equal to the average of the function $\Omega(\cdot)$. This would be the difference if the trend adjusting only for selection on observables were calculated by OLS instead of a truncated regression. Figure A.1 presents the average estimated value of $\Omega(\cdot)$ for each birth cohort. If there were no changing bias from selection on unobservables over cohorts, this would simply be a horizontal line. The plot is not horizontal; indeed, it shows a decline in the average of $\Omega(\cdot)$ between the Union Army cohorts of 1832–1846 and the Regular Army cohorts of 1847–1860, which is consistent with the historical record's indication of more negative selection into the Regular Army as compared to the Union Army. Note that the standard truncation-correction approach implicitly assumes that the average $\Omega(\cdot)$ is greater (i.e., more positive, or less negative) in the 1847–1860 cohorts than in the 1832–1846 cohorts, because of more heavily enforced truncation after the Civil War (as shown in Figure 1).

[31] This view is also supported by Figure A.1, which shows a decline in the estimated $\Omega(\cdot)$ after 1846, indicating more negative selection on unobservables for the Regular Army than for the Union Army.

average stature for the birth cohorts 1832–1846 is not statistically different between the two smoothed trends ($\chi^2_1 = 2.00$, $p = 0.16$), whereas there is a statistically significant difference between the decline in average stature for the birth cohorts of 1832–1860 (as discussed above). It is also evident in the fact that in the unsmoothed trends of panel 2(a), the shift from the Union Army (1832–1846 cohorts) to the Regular Army (1847–1860 cohorts) entails a smaller fall in average stature after incorporating the correction for selection on unobservables.[32] The confounding effects of sample-selection bias thus arise when the two very differently selected samples are placed side by side and used to construct a trend.

When the rates of enlistment in each sample group are considered, this result is not surprising. The basic logic of selection models implies that the magnitude of sample-selection bias is decreasing in the fraction of the population that is represented in the selected sample. It is thus not surprising that the transition from the Union Army with its high rate of enlistment to the Regular Army with its lower rate of enlistment would distort the true trend in height to show a greater decline. This result also fits well with the historical accounts discussed above, which indicate that the postbellum Regular Army was likely to be composed of more negatively selected enlisters than the Union Army.

## 6.4 Cross-Sectional Patterns

Historical anthropometric studies of the United States suggest that the Northeast was the shortest region in the antebellum period (e.g., Komlos, 2012, p. 444). Whether this height disadvantage should be considered a cross-sectional analog of the temporal Antebellum Puzzle is debatable. On the one hand, Easterlin (1960) has shown that income per capita in the Midwest was only 51 percent of that of the Northeast in 1840, implying that there was a cross-sectional Antebellum Puzzle because the better economic well-being in the Northeast did not translate into better health. On the other hand, Margo (1999) has shown that real wages were higher in the Midwest than in the Northeast in the antebellum period, seemingly rationalizing the observed patterns in stature. Moreover, greater rates of urbanization and industrialization in the Northeast than in the Midwest can help to explain the Northeast's height penalty.

Regardless of whether the Northeast's height penalty can be rationalized by its relative prosperity, it is possible that it has been estimated incorrectly as a result of selection on unobservables that differs between regions. To determine whether sample-selection bias is responsible for the Northeast's height penalty I repeat the estimation above, averaging over regions (rather than cohorts) of birth. Results are presented in Table 5. Panel A shows the mean heights per region, adjusting for truncation and selection on observables only. The

[32] Approximating the trend with a piecewise function that admits different slopes and levels between the two armies also shows that the correction is driven largely by differences in the level of selection between them (results available on request).

Northeast's height penalty is evident, with Northeasterners 0.51 inches shorter than Midwesterners. Panel B also incorporates the correction for selection on unobservables. The difference between the Northeast and the Midwest is smaller after this adjustment, but is still present and statistically significant at 0.31 inches. Thus, sample-selection bias cannot wholly explain the Northeast's height disadvantage. But the researcher cannot disregard sample-selection bias, even in cross-sectional comparisons: as shown in Panel C, the difference in heights between Northeasterners and Midwesterners becomes smaller (though the change is only marginally statistically significant), decreasing the difference between the regions by roughly 48.5 percent.

A similar analysis is possible to investigate the urban height penalty—a robust finding that residents of urban areas were shorter than residents of rural areas. This penalty is usually attributed to the separation from food sources and poor sanitary conditions in cities. I define an urban county as one with any urban population and a rural county as one with no urban population. Averaging heights over sector makes it possible to determine to what extent the urban penalty is the result of sample-selection bias that differs by sector. The results of this procedure are presented in Table 6, in which an urban height penalty of 0.54 inches is present when adjusting for truncation and selection on observables (Panel A) and remains present and statistically significant at 0.29 inches when incorporating the correction for selection on unobservables (Panel B). Panel C shows a similar pattern to the regional case: addressing selection on unobservables statistically significantly and meaningfully changes the magnitude of the urban penalty, reducing it by 63.4 percent.

### 6.5 Results with Additional Data and Variables

In Online Appendix H, I present two additional sets of results. The first uses the same data as above, but instead of using Lincoln's vote share as the variable $\mathbf{z}_{it}$, it uses the vote share for Buchanan in 1856 and the vote share for Douglas in 1860. The results are qualitatively similar to those presented in the main text. The second alternative estimation uses the vote share for Lincoln for identification, but instead of using the Union Army to provide height data for the 1832–1846 cohorts, I collected an additional Regular Army sample, this time for the 1832–1846 cohorts, who largely (but not exclusively) enlisted during the Civil War. I combine this additional data set with data from the Regular Army on the 1847–1860 cohorts, used in the main text, and otherwise proceed similarly. In this case, there is little evidence of changing selection over birth cohorts, consistent with the fact that the data source does not change over time. There is no indication in either case that sample-selection bias can explain the decline in average stature.

Table 5: Tests for differences in levels, regional decomposition

| *Region* | (1) Northeast | (2) Midwest | (3) South |
|---|---|---|---|
| *Panel A: Observables Only* | | | |
| Northeast | 66.723$^{***}$ | | |
| | (0.230) | | |
| Midwest | $-0.510^{***}$ | 67.234$^{***}$ | |
| | (0.170) | (0.247) | |
| South | $-0.492^{*}$ | 0.018 | 67.215$^{***}$ |
| | (0.258) | (0.284) | (0.314) |
| *Panel B: Unobservables and Observables* | | | |
| Northeast | 67.915$^{***}$ | | |
| | (0.361) | | |
| Midwest | $-0.314^{**}$ | 68.229$^{***}$ | |
| | (0.144) | (0.362) | |
| South | $-0.304$ | 0.010 | 68.219$^{***}$ |
| | (0.197) | (0.222) | (0.437) |
| *Panel C: $B-A$* | | | |
| Northeast | 1.192$^{***}$ | | |
| | (0.370) | | |
| Midwest | 0.197$^{*}$ | 0.995$^{**}$ | |
| | (0.115) | (0.390) | |
| South | 0.188 | $-0.008$ | 1.003$^{***}$ |
| | (0.142) | (0.147) | (0.378) |
| Observations | 3,293 | 3,169 | 787 |

*Significance levels*: $^{***}$ p< 0.01, $^{**}$ p< 0.05, $^{*}$ p< 0.1
*Notes*: In Panels A and B, the diagonals present the estimated mean heights in each region, corrected for minimum height requirements with a truncation point of 64 inches, for the type of selection in the panel title, for measurement age, and for the separate sampling of the two groups of birth cohorts. The off-diagonals present the differences between the diagonal elements. Panel C presents differences between Panels A and B. In all cases, bootstrap standard errors clustered at the county level are in parentheses. Observation numbers are for the region in the column header for the estimates of Panel B.

Table 6: Tests for differences in levels, sectoral decomposition

| *Sector* | (1) Urban | (2) Rural |
|---|---|---|
| *Panel A: Observables Only* | | |
| Urban | 66.766$^{***}$ | |
| | (0.226) | |
| Rural | $-0.541^{***}$ | 67.307$^{***}$ |
| | (0.150) | (0.255) |
| *Panel B: Unobservables and Observables* | | |
| Urban | 67.958$^{***}$ | |
| | (0.359) | |
| Rural | $-0.287^{**}$ | 68.245$^{***}$ |
| | (0.122) | (0.378) |
| *Panel C: $B-A$* | | |
| Urban | 1.192$^{***}$ | |
| | (0.372) | |
| Rural | 0.254$^{***}$ | 0.938$^{**}$ |
| | (0.082) | (0.378) |
| Observations | 2,973 | 4,276 |

*Significance levels*: $^{***}$ p< 0.01, $^{**}$ p< 0.05, $^{*}$ p< 0.1
*Notes*: In Panels A and B, the diagonals present the estimated mean heights for each sector, corrected for minimum height requirements with a truncation point of 64 inches, for the type of selection in the panel title, for measurement age, and for the separate sampling of the two groups of birth cohorts. The off-diagonals present the differences between the diagonal elements. Panel C presents differences between Panels A and B. In all cases, bootstrap standard errors clustered at the county level are in parentheses. The urban sector is defined as a county with a non-zero urban population. Observation numbers are for the sector in the column header for the estimates of Panel B.

# 7 Conclusion

The Antebellum Puzzle is a major stylized fact of American economic history. Its surprising implication that living standards in the United States were not unambiguously improved by early modern economic growth has changed economists' understanding of economic development and led to a 40-year effort to document and explain the response of the human body to modern economic growth. This historical puzzle also has modern relevance. Deaton (2007) and Jayachandran and Pande (2017) report that economic growth in India has not been matched by improvements in height. They also report cross-sectional relationships in height that contradict monetary measures of welfare, with Africa poorer but taller than India. Similarly, Trivedi (2017) discusses a decline in life expectancy in China during the rapid growth of the 2000s. Thus, deteriorating health may be a symptom of the early stages of rapid economic growth.

It is possible, however, that that studies of historical heights have not sufficiently addressed sample-selection bias, which has the potential to undermine the veracity of the Antebellum Puzzle and of its analogs in countries other than the United States. In this paper, I supplement suggestive evidence from existing studies (Bodenhorn, Guinnane, and Mroz, 2017; Komlos and A'Hearn, 2016) with the first direct test of whether the Antebellum Puzzle is an artifact of sample-selection bias, and of whether failing to properly address such bias has affected conclusions drawn from historical height data. Based on the estimation of a two-step semi-parametric sample-selection model on a set of military-linked census data from the birth cohorts of 1832–1860 in the United States, I find that the trend in stature adjusted for sample-selection bias from selection on observables and on unobservables differs considerably from the baseline results in the literature and from the trend that I compute using standard techniques of the literature. The difference stems primarily from large changes in the degree of sample-selection bias across different sources of data. This result supports the argument that sample selection might have biased the conclusions of the anthropometric history literature. It also bolsters the general argument that future studies of historical heights must be cautious regarding the threat posed by sample-selection bias where it is likely to exist.

At the same time, I show that it is possible to learn from the selected sample without ignoring its potential pitfalls. I find evidence of an Antebellum Puzzle even after incorporating corrections for sample-selection bias, and thus conclude that the view, that the Antebellum Puzzle is merely an artifact of sample-selection bias, is not supported by the data. Moreover, there is no evidence of an increase in average stature, which would be necessary to resolve the puzzle by selection alone. The continuing effort to understand the causes of the Antebellum Puzzle must therefore focus on real explanations linking economic growth to health.

It should be noted that, precisely because I am studying the impacts of selection on unobservables, it is not possible to determine, by direct observation, whether I have purged the data of all of the bias induced by this kind of selection. The only way to be completely certain would be to use a sample of height data spanning this period in which selection on unobservables could be definitively ruled out.[33] Instead, I must rely on economic and statistical theory to indirectly infer the impact of selection on unobservables from the available data. Although I have employed the best available tools to address the possible presence of such selection, and have found that the evidence does not support the assertion that the Antebellum Puzzle is a statistical artifact, it is not possible to be completely certain that this was not the case. Moreover, whereas Bodenhorn, Guinnane, and Mroz's (2017) critique applies to the entirety of the anthropometric history literature and the more general finding of the Industrialization Puzzle, my assessment of its applicability is limited to the American Antebellum Puzzle. In settings where mass mobilizations such as that of the Civil War are not available to provide data, selection on unobservables may be more impactful.

# References


A'Hearn, Brian (1998). "The Antebellum Puzzle Revisited: A New Look at the Physical Stature of Union Army Recruits during the Civil War." In *The Biological Standard of Living in Comparative Perspective*. John Komlos and Jörg Baten (ed.). Stuttgart: Franz Steiner Verlag, pp. 250–267.

Amemiya, Takeshi (1985). *Advanced Econometrics*. Cambridge: Harvard University Press.

Andrews, Donald W. K. and Marcia M. A. Schafgans (1998). "Semiparametric Estimation of the Intercept of a Sample Selection Model." *Review of Economic Studies* 65:3, pp. 497–517.

Bailey, Martha, Connor Cole, Morgan Henderson, and Catherine Massey (2017). "How Well Do Automated Linking Methods Perform in Historical Samples? Evidence from New Ground Truth." NBER Working Paper 24019.

Bernardo, C. Joseph and Eugene H. Bacon (1955). *American Military Policy: Its Development Since 1775*. Harrisburg: The Telegraph Press.

Bodenhorn, Howard, Timothy W. Guinnane, and Thomas A. Mroz (2014). "Sample Selection Bias in the Historical Heights Literature." Cowles Foundation Working Paper, Yale University.

——— (2017). "Sample-Selection Biases and the Industrialization Puzzle." *Journal of Economic History* 77:1, pp. 171–207.

Coffman, Edward M. (1986). *The Old Army: A Portrait of the American Army in Peacetime, 1784–1898*. New York: Oxford University Press.

Cosslett, Stephen R. (1981). "Efficient Estimation of Discrete-Choice Models." In *Structural Analysis of Discrete Data with Econometric Applications*. Charles F. Manski and Daniel McFadden (ed.). Cambridge: MIT Press, 1990. Chap. 2, pp. 51–111.

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

——— (2007). "Deserters, Social Norms, and Migration." *Journal of Law and Economics* 50:2, pp. 323–353.

Costa, Dora L. and Richard H. Steckel (1997). "Long-Term Trends in Health, Welfare, and Economic Growth in the United States." In *Health and Welfare During Industrialization*. Richard H. Steckel and Roderick Floud (ed.). Chicago: University of Chicago Press. Chap. 2, pp. 47–90.

Craig, Lee A. (2016). "Antebellum Puzzle: The Decline in Heights at the Onset of Modern Economic Growth." In *Handbook of Economics and Human Biology*. John Komlos and Inas Kelly (ed.). Oxford: Oxford University Press.

Das, Mitali, Whitney K. Newey, and Francis Vella (2003). "Nonparametric Estimation of Sample Selection Models." *Review of Economic Studies* 70, pp. 33–58.

Deaton, Angus (2007). "Height, Health, and Development." *Proceedings of the National Academy of Sciences* 104:33, pp. 13232–13237.


[33] Steckel and Ziebarth (2016) are able to do this for a related puzzle in the literature on slave growth; but to my knowledge no such source exists for the stature of the native-born white male population in the antebellum United States.

Department of Defense (1997). *Selected Manpower Statistics: Fiscal Year 1997 (DIOR/M01-97)*. Washington: Government Printing Office.

Easterlin, Richard (1960). "Interregional Differences in Per Capita Income, Population, and Total Income, 1840–1950." In *Trends in the American Economy in the Nineteenth Century*. Conference on Research in Income and Wealth (ed.). Princeton: Princeton University Press, pp. 73–140.

Eli, Shari, Laura Salisbury, and Allison Shertzer (2018). "Migration Responses to Conflict: Evidence from the Border of the American Civil War." *Journal of Economic History* Forthcoming.

Floud, Roderick, Robert W. Fogel, Bernard Harris, and Sok Chul Hong (2011). *The Changing Body: Health, Nutrition, and Human Development in the Western World since 1700*. New York: Cambridge University Press.

Floud, Roderick, Kenneth W. Wachter, and Anabel S. Gregory (1990). *Height, Health and History: Nutritional Status in the United Kingdom, 1750–1980*. Cambridge: Cambridge University Press.

Fogel, Robert W. (1986). "Nutrition and the Decline in Mortality since 1700: Some Preliminary Findings." In *Long-Term Factors in American Economic Growth*. Stanley L. Engerman and Robert E. Gallman (ed.). Chicago: University of Chicago Press. Chap. 9, pp. 439–556.

——— (1994). "Economic Growth, Population Theory, and Physiology: The Bearing of Long-Term Processes on the Making of Economic Policy." *American Economic Review* 84:3, pp. 369–395.

Fogel, Robert W., Dora L. Costa, Michael R. Haines, Chulhee Lee, Louis Nguyen, Clayne Pope, Irvin Rosenberg, Nevin Scrimshaw, James Trussell, Sven Wilson, Larry T. Wimmer, John Kim, Julene Bassett, Joseph Burton, and Noelle Yetter (2000). *Aging of Veterans of the Union Army: Version M-5*. Chicago: Center For Population Economics, University of Chicago Graduate School of Business, Department of Economics, Brigham Young University, and the National Bureau of Economic Research.

Fogel, Robert W., Stanley L. Engerman, Roderick Floud, Gerald Friedman, Robert A. Margo, Kenneth Sokoloff, Richard H. Steckel, T. James Trussell, Georgia C. Villaflor, and Kenneth W. Wachter (1983). "Secular Changes in American and British Stature and Nutrition." *Journal of Interdisciplinary History* 14:2, pp. 445–481.

Foner, Jack D. (1970). *The United States Soldier Between Two Wars: Army Life and Reforms, 1865–1898*. New York: Humanities Press.

Frisancho, A. Roberto (1993). *Human Adaptation and Accommodation*. Ann Arbor: University of Michigan Press.

Gallman, Robert E. (1996). "Dietary Change in Antebellum America." *Journal of Economic History* 56:1, pp. 193–201.

Haines, Michael R. (2004). "Growing Incomes, Shrinking People—Can Economic Development Be Hazardous to Your Health?" *Social Science History* 28:2, pp. 249–270.

Heckman, James J. (1979). "Sample Selection Bias as a Specification Error." *Econometrica* 47:1, pp. 153–161.

ICPSR (1999). *United States Historical Election Returns, 1824–1968 (ICPSR 1)* [machine-readable database]. Ann Arbor.

Jayachandran, Seema and Rohini Pande (2017). "Why are Indian Children so Short? The Role of Birth Order and Son Preference." *American Economic Review* 107:9, pp. 2600–2629.

Klein, Roger W. and Richard H. Spady (1993). "An Efficient Semiparametric Estimator for Binary Response Models." *Econometrica* 61:2, pp. 387–421.

Komlos, John (1987). "The Height and Weight of West Point Cadets: Dietary Change in Antebellum America." *Journal of Economic History* 47:4, pp. 897–927.

——— (1998). "On the Biological Standard of Living of African-Americans: the Case of the Civil War Soldiers." In *The Biological Standard of Living in Comparative Perspective*. John Komlos and Jörg Baten (ed.). Stuttgart: Franz Steiner Verlag, pp. 236–249.

——— (2012). "A Three-Decade History of the Antebellum Puzzle: Explaining the Shrinking of the U.S. Population at the Onset of Modern Economic Growth." *Journal of the Historical Society* 12:4, pp. 395–445.

Komlos, John and Brian A'Hearn (2016). "The Decline in the Nutritional Status of the US Antebellum Population at the Onset of Modern Economic Growth." NBER Working Paper 21845.

Lindert, Peter H. and Robert A. Margo (2006). "Table Cc1–2: Consumer Price Indexes, for All Items, 1774–2003." In *Historical Statistics of the United States*. Susan B. Carter, Scott Sigmund Gartner, Michael R. Haines, Alan L. Olmstead, Richard Sutch, and Gavin Wright (ed.). Cambridge: Cambridge University Press, pp. 3.158–3.159.

Manson, Steven, Jonathan Schroeder, David Van Riper, and Steven Ruggles (2017). *IPUMS National Histortical Geographic Information System: Version 12.0* [Database]. Minneapolis: University of Minnesota.

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

Margo, Robert A. and Richard H. Steckel (1983). "Heights of Native-Born Whites During the Antebellum Period." *Journal of Economic History* 43:1, pp. 167–174.

Mokyr, Joel and Cormac Ó Gráda (1996). "Height and Health in the United Kingdom 1815–1860: Evidence from the East India Company Army." *Explorations in Economic History* 33, pp. 141–168.

Newey, Whitney K. (2009). "Two-Step Series Estimation of Sample Selection Models." *Econometrics Journal* 12:S1, S217–S229.

*Register of Enlistments in the U.S. Army, 1798–1914* (n.d.). (National Archives Microfilm Publication M233, 81 Rolls), Records of the Adjutant General's Office, 1780's–1917, RG 94. Washington, DC: National Archives.

Ruggles, Steven, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek (2015). *Integrated Public Use Microdata Series: Version 6.0* [machine-readable database]. Minneapolis: University of Minnesota.

Steckel, Richard H. and Nicolas Ziebarth (2016). "Selectivity and Measured Catch-up Growth of American Slaves." *Journal of Economic History* 76:1, pp. 109–138.

Trivedi, Anjani (2017). "Why Chinese Men Are Dying Despite Rising Income." *Wall Street Journal* February 25, 2017, B.10.

Vella, Francis (1998). "Estimating Models with Sample Selection Bias: A Survey." *Journal of Human Resources* 33:1, pp. 127–169.

Weigley, Russell F. (1967). *History of the United States Army.* New York: The Macmillan Company.

Zehetmayer, Matthias (2011). "The Continuation of the Antebellum Puzzle: Stature in the US, 1847–1894." *European Review of Economic History* 15, pp. 313–327.

# A Estimated $\Omega(\cdot)$ Function by Birth Cohort


Figure A.1: Estimated $\Omega(\cdot)$ function by birth cohort

*Note:* This graph plots the coefficients from a regression of the estimated function $\hat{\Omega}(\hat{\alpha}_t + \mathbf{x}_{it}'\hat{\beta}_k + \mathbf{z}_{it}'\hat{\delta}_k)$ on birth year indicators, weighting by inverse enlistment probability (in the dashed line), as well as these coefficients smoothed over birth cohorts (in the solid line).

# B Additional Tables (For Online Publication)

Table B.1: Binary choice model estimation

| | (1) | | (2) | |
|---|---|---|---|---|
| | (a) | (b) | (a) | (b) |
| *Cohorts* | '32–'46 | '47–'60 | '32–'46 | '47–'60 |
| *Variables* | UA | RA | UA | RA |
| *Individual or Household Variables* | | | | |
| Household Owns Property | 0.081*** | 0.189*** | 0.232*** | 0.453*** |
| | (0.017) | (0.019) | (0.066) | (0.048) |
| Household Real Property (1,000) | −0.015*** | −0.046*** | −0.051*** | −0.120*** |
| | (0.002) | (0.004) | (0.006) | (0.012) |
| Related to Head of Household | 0.092*** | −0.062*** | 0.243*** | −0.189*** |
| | (0.019) | (0.018) | (0.052) | (0.070) |
| Household Size | 0.017*** | −0.000 | 0.059*** | −0.001 |
| | (0.003) | (0.003) | (0.009) | (0.007) |
| Attended School | −0.177*** | −0.148*** | −0.641*** | −0.501*** |
| | (0.016) | (0.015) | (0.049) | (0.065) |
| Household Occupation (Unproductive excluded) | | | | |
| Farmer | −0.341*** | −0.223*** | −1.179*** | −0.565*** |
| | (0.038) | (0.024) | (0.109) | (0.085) |
| Professional | −0.286*** | −0.106*** | −0.810*** | −0.209*** |
| | (0.056) | (0.036) | (0.110) | (0.060) |
| Clerical | −0.361*** | −0.154*** | −1.013*** | −0.290*** |
| | (0.046) | (0.030) | (0.084) | (0.039) |
| Skilled and Artisan | −0.316*** | −0.097*** | −0.941*** | −0.224*** |
| | (0.038) | (0.021) | (0.075) | (0.045) |
| Semi-Skilled and Clerical | −0.322*** | −0.133*** | −0.908*** | −0.262*** |
| | (0.047) | (0.028) | (0.085) | (0.043) |
| Unskilled | −0.305*** | −0.101*** | −0.856*** | −0.212*** |
| | (0.042) | (0.022) | (0.073) | (0.042) |
| Farm Labor | −0.187* | −0.126*** | −0.557 | −0.236*** |
| | (0.102) | (0.033) | (0.342) | (0.040) |
| Birth Region (South excluded) | | | | |
| Midwest | 0.209*** | −0.054* | 0.947*** | −0.134* |
| | (0.033) | (0.031) | (0.158) | (0.078) |
| Northeast | 0.102*** | −0.046 | 0.343*** | −0.124 |
| | (0.033) | (0.030) | (0.112) | (0.090) |
| *County Variables* | | | | |
| Fraction Urban | −0.123*** | 0.049 | −0.420*** | 0.128* |
| | (0.039) | (0.031) | (0.121) | (0.072) |
| Wheat Bushels per capita | 0.013*** | −0.004*** | 0.043*** | −0.009*** |
| | (0.001) | (0.001) | (0.004) | (0.003) |
| Milk Cows per capita | 0.010 | −0.066 | 0.033 | −0.173 |
| | (0.038) | (0.070) | (0.130) | (0.190) |
| Swine per capita | 0.075*** | −0.076*** | 0.256*** | −0.198*** |
| | (0.011) | (0.017) | (0.032) | (0.043) |
| Value of Agricultural Production per capita | −0.046 | −0.042 | −0.157 | −0.109 |
| | (0.472) | (0.432) | (1.483) | (1.106) |
| Lincoln Vote Share (1860) | 0.544*** | 0.127** | 1.853*** | 0.332** |
| | (0.064) | (0.054) | (0.181) | (0.140) |
| Observations | 13,683 | 11,271 | 13,683 | 11,271 |

*Significance levels*: *** p<0.01, ** p<0.05, * p<0.1
*Notes*: Column (1) presents estimates of the coefficients $\beta$ and $\delta$ from the binary choice model. Column (2) presents the average semi-elasticity of the impact of each variable on enlistment probability as implied by the estimates of column (1). All specifications include cohort indicators. Standard errors are clustered at the county level. UA denotes Union Army. RA denotes Regular Army.

Table B.2: Height regressions

| Variables | (1) Corr | (2) Not | (3) Corr | (4) Not |
|---|---|---|---|---|
| *Individual or Household Variables* | | | | |
| Household Owns Property | 0.092 | 0.104 | 0.100 | 0.104 |
| | (0.094) | (0.111) | (0.096) | (0.111) |
| Household Real Property (1,000) | −0.011 | −0.006 | −0.012 | −0.006 |
| | (0.012) | (0.008) | (0.012) | (0.008) |
| Related to Head of Household | 0.245** | 0.300** | 0.245** | 0.300** |
| | (0.104) | (0.131) | (0.104) | (0.131) |
| Household Size | 0.025* | 0.023 | 0.026* | 0.024 |
| | (0.014) | (0.015) | (0.014) | (0.015) |
| Attended School | 0.038 | 0.044 | 0.026 | 0.038 |
| | (0.077) | (0.084) | (0.079) | (0.084) |
| Household Occupation (Unproductive excluded) | | | | |
| Farmer | 0.208 | 0.055 | 0.180 | 0.030 |
| | (0.134) | (0.123) | (0.140) | (0.125) |
| Professional | 0.234 | 0.039 | 0.214 | 0.019 |
| | (0.207) | (0.236) | (0.207) | (0.235) |
| Clerical | 0.011 | −0.260 | −0.012 | −0.279 |
| | (0.177) | (0.197) | (0.182) | (0.198) |
| Skilled and Artisan | −0.128 | −0.313** | −0.147 | −0.331** |
| | (0.144) | (0.148) | (0.147) | (0.149) |
| Semi-Skilled and Clerical | 0.080 | −0.125 | 0.061 | −0.140 |
| | (0.185) | (0.216) | (0.188) | (0.216) |
| Unskilled | 0.250 | 0.091 | 0.228 | 0.067 |
| | (0.173) | (0.187) | (0.178) | (0.189) |
| Farm Labor | −0.205 | −0.479* | −0.222 | −0.496* |
| | (0.230) | (0.283) | (0.232) | (0.285) |
| Birth Region (South excluded) | | | | |
| Midwest | 0.034 | 0.042 | −0.070 | −0.136 |
| | (0.163) | (0.169) | (0.173) | (0.198) |
| Northeast | −0.244 | −0.322* | −0.362** | −0.524** |
| | (0.162) | (0.181) | (0.181) | (0.216) |
| *County Variables* | | | | |
| Fraction Urban | −0.553*** | −0.613** | −0.556*** | −0.630*** |
| | (0.202) | (0.238) | (0.202) | (0.238) |
| Wheat Bushels per capita | −0.001 | 0.004 | −0.001 | 0.003 |
| | (0.007) | (0.006) | (0.007) | (0.006) |
| Milk Cows per capita | 0.008 | 0.080 | −0.016 | 0.048 |
| | (0.206) | (0.196) | (0.212) | (0.205) |
| Swine per capita | 0.077 | 0.064 | 0.094 | 0.087 |
| | (0.051) | (0.051) | (0.057) | (0.054) |
| Value of Agricultural production per capita | −3.964* | −3.758* | −4.305** | −4.284* |
| | (2.047) | (2.257) | (2.089) | (2.320) |
| Lincoln Vote Share (1860) | | | 0.292 | 0.469 |
| | | | (0.300) | (0.315) |
| Observations | 7,249 | 6,873 | 7,249 | 6,873 |

*Significance levels*: *** p<0.01, ** p<0.05, * p<0.1
*Notes*: Standard errors in parentheses. Dependent variable is height, measured in inches. All specifications include age-of-measurement, year-of-birth, and household occupation indicators. The selection-corrected specifications, indicated by the column header Corr, also include the selection-correction function $\Omega(\cdot)$. The uncorrected specifications, indicated by the column header Not, correct for truncation with a truncation point of 64 inches. Standard errors are clustered at the county level. The difference in sample sizes between columns is the result of the need to drop heights below 64 inches in the truncation-corrected regressions when not correcting for sample-selection bias.

# C Correcting for Selection: Formal Arguments (For Online Publication)

The model developed in section 2 points out the need to correct for selection into military service on the basis of both observable and unobservable characteristics in order to estimate $E(h_{it}|t)$ for each $t$—that is, the trend in average heights over birth cohorts. In this Appendix, I formally develop the weighting approach to correct for selection on observables and the control function approach to correct also for selection on unobservables.

## C.1 Selection on Observables

By the law of iterated expectations, the object of interest can be written as

$$\begin{aligned} E(h_{it}|t) &= \int_{\mathfrak{X}} E(h_{it}|\mathbf{x}_{it}, \mathbf{z}_{it}; t) f(\mathbf{x}_{it}, \mathbf{z}_{it}|t) \, \mathrm{d}\mathbf{w}_{it} \\ &= \int_{\mathfrak{X}} E(h_{it}|\mathbf{x}_{it}; t) f(\mathbf{x}_{it}, \mathbf{z}_{it}|t) \, \mathrm{d}\mathbf{w}_{it}, \end{aligned} \tag{C.1}$$

where $\mathbf{w}_{it} = [\mathbf{x}'_{it}, \mathbf{z}'_{it}]'$,[34] $\mathfrak{X}$ is the support of $\mathbf{w}_{it}$, $F(\cdot)$ is its distribution function, and $f(\cdot)$ is its density. Equation (C.1) follows from the assumption—implicit in equations (1) and (2)—that height is uncorrelated with $\mathbf{z}_{it}$ conditional on $\mathbf{x}_{it}$. If there were no self selection, either on the basis of observables or unobservables, the left hand side of equation (C.1) could be computed trivially from the data; however, the researcher observes $E(h_{it}|y_{it} = 1; t)$ and not $E(h_{it}|t)$ in a selected sample. Moreover, the components of the right-hand side of equation (C.1) cannot, in general, be directly computed from a sample consisting solely of military enlisters—the researcher observes $E(h_{it}|\mathbf{x}_{it}, y_{it} = 1; t)$ and not $E(h_{it}|\mathbf{x}_{it}; t)$; but if the selection is only on observables (i.e., $\varepsilon_{it}$ and $u_{it}$ are uncorrelated), the assumptions discussed above imply that

$$E(h_{it}|\mathbf{x}_{it}, y_{it} = 1; t) = E(h_{it}|\mathbf{x}_{it}; t) = \gamma_t + \mathbf{x}'_{it}\theta. \tag{C.2}$$

What remains as the main pitfall is that selection into military service on the basis of observables implies that $f(\mathbf{x}_{it}, \mathbf{z}_{it}|t) \neq f(\mathbf{x}_{it}, \mathbf{z}_{it}|y_{it} = 1; t)$. That is to say, simply averaging the observed heights within each birth cohort will not yield consistent estimates of the true heights because the weighting is based on the distribution of covariates in the selected sample, which differs from that in the population. However, Bayes's

[34] Although consideration of the exclusion-restriction variables $\mathbf{z}_{it}$ is unnecessary for the correction of this type of selection, I include them as they are required for identification in the correction for selection on unobservables.

Theorem implies that

$$f(\mathbf{x}_{it},\mathbf{z}_{it}|y_{it}=1;t)=\frac{P(y_{it}=1|\mathbf{x}_{it},\mathbf{z}_{it};t)f(\mathbf{x}_{it},\mathbf{z}_{it}|t)}{P(y_{it}=1|t)}, \tag{C.3}$$

so that

$$f(\mathbf{x}_{it},\mathbf{z}_{it}|t)=\frac{f(\mathbf{x}_{it},\mathbf{z}_{it}|y_{it}=1;t)P(y_{it}=1|t)}{P(y_{it}=1|\mathbf{x}_{it},\mathbf{z}_{it};t)}\propto\frac{f(\mathbf{x}_{it},\mathbf{z}_{it}|y_{it}=1;t)}{P(y_{it}=1|\mathbf{x}_{it},\mathbf{z}_{it};t)}. \tag{C.4}$$

Substituting expressions (C.2) and (C.4) into equation (C.1) gives

$$E(h_{it}|t)=\int_{\mathfrak{X}}E(h_{it}|\mathbf{x}_{it},y_{it}=1;t)f(\mathbf{x}_{it},\mathbf{z}_{it}|y_{it}=1;t)\frac{k_t}{P(y_{it}=1|\mathbf{x}_{it},\mathbf{z}_{it};t)}\,\mathrm{d}\mathbf{w}_{it}, \tag{C.5}$$

where $k_t$ is the normalizing constant for cohort $t$. Note that if the researcher were simply to take the (unweighted) average height for each birth cohort from a selected sample, he would estimate

$$E(h_{it}|y_{it}=1;t)=\int_{\mathfrak{X}}E(h_{it}|\mathbf{x}_{it},y_{it}=1;t)f(\mathbf{x}_{it},\mathbf{z}_{it}|y_{it}=1;t)\,\mathrm{d}\mathbf{w}_{it} \tag{C.6}$$

from its sample analog $\frac{1}{N_t}\sum_{i\in t}h_{it}$, where $N_t$ denotes the number of individuals in the sample belonging to birth cohort $t$ and $i\in t$ denotes the members of cohort $t$; because expression (C.6) is equivalent to the right-hand side of expression (C.5) save for the inclusion of the weights $\frac{k_t}{P(y_{it}=1|\mathbf{x}_{it},\mathbf{z}_{it};t)}$, it is natural to estimate expression (C.5) by its sample analog

$$\hat{h}_t=\frac{\hat{k}_t}{N_t}\sum_{i\in t}\frac{h_{it}}{P(y_{it}=1|\mathbf{x}_{it},\mathbf{z}_{it};t)}=\frac{\hat{k}_t}{N_t}\sum_{i\in t}\frac{h_{it}}{G(\alpha_t+\mathbf{x}_{it}'\beta_k+\mathbf{z}_{it}'\delta_k)}. \tag{C.7}$$

It can be shown that expression (C.7) is precisely the estimated coefficient on the year-of-birth indicator for cohort $t$ when observed heights are regressed on birth-cohort indicators (and no constant), weighting by inverse conditional enlistment probabilities, thus providing a method to perform this correction.

## C.2 Selection on Unobservables

When selection on unobservables is admitted alongside selection on observables (i.e., correlation is permitted between $\varepsilon_{it}$ and $u_{it}$), the arguments made in equations (C.3) and (C.4) continue to hold, as they did not rely on uncorrelatedness of $\varepsilon_{it}$ and $u_{it}$, but rather were simply an application of Bayes's theorem. However, expression (C.2) is no longer true. Instead,

$$E(h_{it}|\mathbf{x}_{it},\mathbf{z}_{it},y_{it}=1;t) = \alpha_t + \mathbf{x}'_{it}\theta + E(\varepsilon_{it}|\mathbf{x}_{it},\mathbf{z}_{it},y_{it}=1;t)$$
$$= E(h_{it}|\mathbf{x}_{it};t) + \Omega(\alpha_t + \mathbf{x}'_{it}\beta_k + \mathbf{z}'_{it}\delta_k), \tag{C.8}$$

where (C.8) follows from the assumptions regarding $u_{it}$ and $\varepsilon_{it}$ and equations (1) and (2). Thus, it is possible to write the missing piece of information in calculating expression (C.1) as a function of data and an unknown (but possible to estimate) object:

$$E(h_{it}|\mathbf{x}_{it};t) = E(h_{it}|\mathbf{x}_{it},\mathbf{z}_{it},y_{it}=1;t) - \Omega(\alpha_t + \mathbf{x}'_{it}\beta_k + \mathbf{z}'_{it}\delta_k).$$

The analog of equation (C.5) is then

$$E(h_{it}|t) = \int_{\mathfrak{X}} \left[E(h_{it}|\mathbf{x}_{it},\mathbf{z}_{it},y_{it}=1;t) - \Omega(\alpha_t + \mathbf{x}'_{it}\beta_k + \mathbf{z}'_{it}\delta_k)\right]$$
$$\times f(\mathbf{x}_{it},\mathbf{z}_{it}|y_{it}=1;t)\frac{k_t}{P(y_{it}=1|\mathbf{x}_{it},\mathbf{z}_{it};t)}\,\mathrm{d}\mathbf{w}_{it}, \tag{C.9}$$

which can also be estimated by its sample analog:

$$\hat{h}_t = \frac{\hat{k}_t}{N_t}\sum_{i\in t}\frac{h_{it} - \Omega(\alpha_t + \mathbf{x}'_{it}\beta_k + \mathbf{z}'_{it}\delta_k)}{G(\alpha_t + \mathbf{x}'_{it}\beta_k + \mathbf{z}'_{it}\delta_k)}. \tag{C.10}$$

## D Constructing Weights (For Online Publication)

This appendix describes the computation of the weights for estimation of the binary choice model. In order to compute the fraction of the relevant population serving in the Union Army, I consulted two sources. The first was Gould (1869, p. 28), who reports that 1,660,068 native-born men served in the Union Army. Next I consulted the 1860 census, finding that there were 3,720,008 native-born men aged 15–45 in the portions of the United States that did not secede. Thus, for the Union Army population

$$Q_1^{\mathrm{UA}} = \frac{1,660,068}{3,720,008} = 0.446.$$

To determine the value of $Q_1$ for the Regular Army, I again consulted two sources. First, I determined from the 1870 census that the total native-born white male population in non-seceding areas born between 1847 and 1860 was 3,467,695. Second, I collected the index of the *Register of Enlistments* for the native-born, and removed duplicate entries of name, birth year and state of birth.[35] This procedure yielded 77,836 distinct enlisters for the 1847–1860 cohorts, in each case restricting attention to those born in non-seceding areas. The estimate of $Q_1$ is thus

$$Q_1^{\mathrm{RA}} = \frac{77,836}{3,467,695} = 0.022.$$

[35] The removal of duplicates is necessary because individuals could enlist multiple times. Dropping duplicate appearances of name, birth year, and state of birth is an imperfect way of addressing this possibility. It is simultaneously too restrictive—there may have been two enlisters with the same name born in the same state in the same year—and too loose—slight deviations in the spelling or abbreviation of names, or misreporting of birth years would allow multiple enlistments by one individual to survive the removal of duplicates and be counted. Given that the actual value of $Q_1$ does not seem particularly important in practice, I do not investigate this potential problem further. An alternative method is to estimate the fraction of enlistments that are repeat enlistments using information from the *Register of Enlistments*; however, this information is not readily available for those in the earlier birth cohorts.

# E Linkage (For Online Publication)

## E.1 Procedure to Link Regular Army Enlisters to the US Censuses

I use standard census linking techniques developed by Ferrie (1996) to link the Regular Army samples to the US Censuses.

**Procedure E.1.** The procedure for linking the 1847–1860 cohorts to the censuses is as follows.

1. I obtained a 100 percent index of the 1880 census from Ruggles et al. (2015).

2. On the basis of first name, last name, state of birth, and year of birth ($\pm$ 4), I linked the 1880 census sample to itself. As above, no name standardization is made, but inexact matches are permitted. Any individual for whom another individual similar on these identifying characteristics existed was removed from the sample.

3. I obtained a 100 percent sample of individuals born 1847–1860 listed in the *Register of Enlistments* from Ancestry.com (2007).

4. Using the same identifying information and criteria described in step 2, I linked individuals in the *Register of Enlistments* to the remaining individuals from the 1880 census. Due to the possibility of multiple enlistments in the lifetime, I permit several individuals in the *Register of Enlistments* to match to one individual in the census. However, in the event that several individuals in the 1880 census are matched to a single individual in the *Register of Enlistments*, I drop all concerned individuals.

5. I match individuals in the *Register of Enlistments* who are matched to the 1880 census to the *Register of Enlistments* once more (using stricter criteria), thus bringing in additional enlistments. Multiple matches from the 1880 census to the *Register of Enlistments* are once again omitted.

6. I then link the 1880 individuals who were linked to the *Register of Enlistments* to the 1860 and 1870 censuses (gathered from Ancestry.com, 2009a,b) using the same information and criteria.

Table E.1 presents the numbers of individuals included in the sample at each stage. Note that I did not restrict attention during linkage to residents of non-seceding areas.

Table E.1: Sample sizes at each stage of linking for 1847–1860 cohorts

| (1) | (2) | (3) Sample Size | | (4) % of Previous | |
|---|---|---|---|---|---|
| | | (a) | (b) | (a) | (b) |
| Step No. | Description | Census | Enlistments | Census | Enlistments |
| 3 | Enlist Full | | 93,085 | | |
| 4 & 5 | 1880-Enlistments Link | 14,343 | 18,802 | | 20.20% |
| 6 | 1860 Link | 3,133 | 4,611 | 21.84% | 24.52% |
| 6 | 1870 Link | 3,129 | 4,632 | 21.82% | 24.64% |

## E.2 Representativeness of the Linked Samples

The military height data used in this paper differ from those typically used in the anthropometric history literature. While this literature generally takes random samples of the height data available in any particular source (leaving, in a military enlistment sample, the military enlistment decision as the only relevant point of sample selection), I limit my samples to the subset of these records that could also be linked to census records. The introduction of this second selection mechanism is necessary in order to gather covariates for comparison to the population as a whole and thus for estimation of the sample-selection model; but it may induce additional and potentially problematic bias.

Bias from non-representative linking can take two forms. First, I determine whether correcting for sample-selection bias has a meaningful effect on the trends in stature by comparing the trends corrected for selection on both observable and unobservable characteristics (which should represent the population trend in heights) to those corrected only for selection on observable characteristics (representing the conventional approach in the anthropometric history literature). If the sample-selection model corrects for selection both into military service and into census linking, then any difference in trends may be due to selection into census linking and not into military enlistment. Such bias would not be present in most studies of historical heights (because they do not use linked samples), but I would erroneously conclude that sample-selection bias existed in the height samples. Second, it is possible that the sample-selection correction would not properly correct for selection both into military enlistment and census linking. Mroz (2015) shows that studying two types of selection in a single index model has the potential to exacerbate any sample-selection bias, leading to estimates with even greater bias than those from a naive approach that ignores selection altogether.

Due to the possibly severe consequences of selection into census linkage, it is important to determine empirically whether such selection is likely to be present. As with the sample-selection issue that motivates this paper, selection into linkage is only problematic if it varies over cohorts. Fortunately, unlike the sample-selection problem, in which the outcome of interest is observed only for the selected sample, it is possible to directly test for selection of this type because the outcome variable (height) can be observed for both the selected (successfully census-linked) and unselected (failed to link to the census) samples. In order to test for sample-selection bias induced by selection into census-linking, I collected data on a random sample of Regular Army enlisters for the 1847–1860 cohorts without any attempt at linkage to the census. Similarly, I collected from the Union Army project information on enlisters without regard to linkage. Comparing the distributions and trends in heights of the linked sample and the unlinked sample (which represents the population of military enlisters as a whole rather than only those who could not be linked) makes it possible to determine whether problematic sample-selection bias is likely to exist. Throughout this analysis, I do not restrict attention to those living in non-seceding regions because this restriction was not imposed in linkage.

Table E.2 presents regressions comparing the trends in heights of the linked and unlinked (that is, representative of the whole enlisting population regardless of census linking) samples. Each column of this Table presents the results of two specifications. The first regresses heights on birth year indicators, measurement age indicators, and an indicator for being in the linked sample. The coefficient on the linked indicator is presented in Table E.2. This tests whether the linked and unlinked trends differ in level. The second regression adds interactions of the linkage indicator and the cohort indicators. The results of a $\chi^2$-test of joint significance of these trends—which is a test of whether the trends in height of the linked and unlinked differ from one another—are also presented in Table E.2. Results of these regressions show statistically significant evidence of positive selection into linkage on the basis of height. There is not, however, any indication of a statistically significant difference in trends. Figure E.1 replicates this analysis graphically by plotting the trends in average heights over birth cohorts in both the linked and the unlinked groups. While differences in levels are evident between the linked and unlinked trends, the trends themselves are visually quite similar.

It therefore appears that the only difference between the linked and unlinked trends is in level, with the linked taller than the population of military enlisters as a whole; that is, there exists selection into linkage, but it is cohort-invariant. The presence of positive selection into census linking on the basis of height is unsurprising, as census linkage is likely to favor those who provide accurate information in a number of sources, and who are therefore likely to be better educated (Ferrie, 1996), and thus also likely to be taller. This difference suggests that one should be cautious in interpreting the level of the trends corrected for selection on both observable and unobservable characteristics. However, I find no reason to believe that the trend itself is unrepresentative of the population of military enlisters, and therefore conclude that the

Table E.2: Regressions of selection into linkage

| | (1) | (2) | (3) |
|---|---|---|---|
| *Cohorts* | '32–'46 | '47–'60 | '32–'60 |
| *Variables* | UA | RA | UA & RA |
| Linked | 0.149*** | 0.192*** | 0.169*** |
| | (0.049) | (0.074) | (0.049) |
| Observations | 11,729 | 5,135 | 16,864 |
| $\chi^2$-Test of Birth Year FE × Linked | 10.99 | 12.72 | 23.11 |

*Significance levels*: *** p<0.01, ** p<0.05, * p<0.1
*Notes*: Dependent variable is height, measured in inches. Truncated regression is performed to account for minimum height requirements with a truncation point of 64 inches. All specifications include measurement-age and birth-year dummy variables. Standard errors are clustered by image for the unlinked sample. The sample includes linked and unlinked members of the Regular Army and Union Army. UA denotes the Union Army. RA denotes the Regular Army. The coefficients on linked are from a regression without interactions. The statistics on the interactions are from a separate regression with interactions.


Figure E.1: Height trends of linked and unlinked samples

*Note:* These trends compare enlisters who could be linked to census data (the "Linked") to a random sample collected without respect to linking (the "Unlinked"). Tests for the statistical significance of differences between the trends are presented in Tables E.2.

correction for sample-selection is informative regarding the trend in heights of the population. To put it briefly, any bias from census linking should be captured by the intercept.

To further explore differences between the linked and unlinked military samples, I gathered a number of covariates from the military enlistment records for both the linked and unlinked. As these are taken from military records rather than from census records located through linkage (as are the covariates used in the main analysis), these covariates are not necessarily comparable to those discussed in the main text. The covariates collected are region of birth, year of birth, year of enlistment, and occupation (categorized using the same categories as above). I also created measures of name complexity and length separately for first name and surname. Name complexity is measured by the scrabble score, which is increasing in the length and complexity of a name (Biavaschi, Giulietti, and Siddique, 2017). These measures are included in order to capture the fact that individuals with unique names are generally easier to match.

In Table E.3 I study whether the sample is balanced on the basis of these covariates. In particular, I present the results of a number of regressions of the form

$$x_i = \varsigma_0 + \varsigma_1 \ell_i + \nu_i,$$

where $x_i$ is some covariate and $\ell_i$ is an indicator for being in the linked sample. Cells of Table E.3 present estimates of $\varsigma_1$, which is the degree to which a particular covariate is overrepresented in the linked sample relative to the population of enlisters as a whole. For example, Northeasterners make up 3 percentage points less of the linked Union Army sample than the population of the Union Army. Overall, there are statistically significant differences between the linked and unlinked samples on the basis of name length and complexity, in terms of region of birth (generally under-representing the Northeast and over-representing the Midwest), and on the basis of occupation.

To correct for these imbalances, I compute weights in order to correct for selection into census linkage on the basis of these observable characteristics. In particular, I estimate a probit model for selection into linkage using Cosslett's (1981) likelihood, and use the results to compute inverse conditional linkage probabilities by which to weight.[36] I first reproduce the above analysis of the trends in heights of the linked and unlinked, weighting the linked samples by the inverse linkage probability. The results are presented in Table E.4 and Figure E.2. The differences in level between the heights of the linked and unlinked are smaller than in the unweighted equivalents, suggesting that some of the level differences are due to differences in these

[36] I omit the top one percent and bottom one percent of the sample, in terms of weights, in order to avoid having the results be driven by such outliers receiving too much weight.

Table E.3: Balancing tests for selection into linkage

| | (1) | (2) |
|---|---|---|
| *Cohorts* | 1832–1846 | 1847–1860 |
| *Dep. Variable* | Union Army | Regular Army |
| Name | | |
| Surname Scrabble Score | 0.023 | 0.036 |
| | (0.068) | (0.128) |
| First Name Scrabble Score | −0.052 | −1.922*** |
| | (0.073) | (0.107) |
| Surname Length | 0.008 | 0.124*** |
| | (0.029) | (0.047) |
| First Name Length | 0.023 | −0.806*** |
| | (0.032) | (0.047) |
| Region | | |
| Northeast | −0.030*** | −0.038** |
| | (0.009) | (0.015) |
| Midwest | 0.010 | 0.061*** |
| | (0.009) | (0.013) |
| South | 0.017*** | −0.024** |
| | (0.005) | (0.010) |
| Birth Year | −0.346*** | 0.623*** |
| | (0.068) | (0.161) |
| Enlistment Year | 0.003 | −0.000 |
| | (0.024) | (0.238) |
| Occupation | | |
| Farmer | 0.022*** | 0.031*** |
| | (0.008) | (0.012) |
| Professional | −0.001 | 0.010*** |
| | (0.003) | (0.003) |
| Clerical | 0.001 | 0.026*** |
| | (0.003) | (0.009) |
| Skilled and Artisan | −0.006 | 0.048*** |
| | (0.006) | (0.013) |
| Semi-Skilled and Operative | −0.002 | −0.089*** |
| | (0.003) | (0.010) |
| Unskilled | −0.011** | −0.035** |
| | (0.005) | (0.014) |
| Unproductive | −0.002 | 0.009*** |
| | (0.002) | (0.002) |
| Observations (Linked) | 5,412 | 2,340 |
| Observations (Unlinked) | 7,134 | 2,956 |

*Significance levels*: *** p<0.01, ** p<0.05, * p<0.1
*Notes*: Each cell represents the coefficient from a regression of the dependent variable in the first column on an indicator for linkage. The "unlinked" group is not composed only of the unlinked, but is a random sample of the population of enlisters, so the coefficients are to be interpreted as the difference in the dependent variable between the linked and the population of enlisters. All regressions cluster standard errors on image, and are weighted to account for stratification; for the Regular Army, weighting is also performed to make the enlistment years of the whole population similar to that of the linked.

observable characteristics. Importantly, any differences in trend between the linked and unlinked are small and statistically insignificant. In all analyses (except where indicated otherwise), I weight by this inverse linkage probability in order to correct for non-representative linkage.

Table E.4: Selection into linkage, weighted by inverse conditional linkage probability

| | (1) | (2) | (3) |
|---|---|---|---|
| *Cohorts* | '32–'46 | '47–'60 | '32–'60 |
| *Variables* | UA | RA | UA & RA |
| Linked | 0.120** | 0.080 | 0.089* |
| | (0.049) | (0.084) | (0.054) |
| Observations | 11,624 | 5,064 | 16,688 |
| $\chi^2$-Test of Birth Year FE × Linked | 11.17 | 12.56 | 22.50 |

*Significance levels*: *** p<0.01, ** p<0.05, * p<0.1
*Notes*: Dependent variable is height, measured in inches. Truncated regression is performed to account for minimum height requirements with a truncation point of 64 inches. All specifications include measurement-age and birth-year dummy variables. Standard errors are clustered by image for the unlinked sample. The sample includes linked and unlinked members of the Regular Army and Union Army. UA denotes the Union Army. RA denotes the Regular Army. The coefficients on linked are from a regression without interactions. The statistics on the interactions are from a separate regression with interactions.


Figure E.2: Height trends of linked and unlinked samples, weighting by inverse conditional linkage probability

*Note:* See Figure E.1. These graphs are weighted by inverse conditional linkage probability.

# F Estimation Details (For Online Publication)

## F.1 Adapting Klein and Spady's (1993) Estimator

Klein and Spady (1993) develop a method to estimate a model of the form of equation (3) in a simple random sample of a population. The sample used in the present research differs from such a sample in two ways. First, the sample is a choice-restricted sample with a supplementary sample (as discussed by Cosslett, 1981), rather than a simple random sample. Second, the sample is composed of two distinct subsamples that are sampled separately: the 1832–1846 cohorts and their supplementary sample, and the 1847–1860 cohorts and their supplementary sample.

Klein and Spady's (1993) estimator is a maximum likelihood estimator, where the likelihood function takes the usual form for binary choice models, and where the function $G(\cdot)$ is estimated using the leave-one-out Nadaraya (1964) and Watson (1964) (NW) estimator. To adapt this estimator to the sample available in the present context, it is useful to define the variable $s_{it}$ as

$$s_{it} = \begin{cases} 0 & \text{for members of the 1847–1860 cohorts} \\ 1 & \text{for members of the 1832–1846 cohorts} \end{cases}$$

and $\psi_{it} = \hat{\alpha}_t + \mathbf{x}'_{it}\hat{\beta}_k + \mathbf{z}_{it}{}'\hat{\delta}_k$. Let $\lambda(\cdot)$ denote the probability density function of $\psi_{it}$. I then use Bayes's Theorem and the law of total probability to write $P(y_{it} = 1|\mathbf{x}_{it}, \mathbf{z}_{it}; t)$ in terms of objects that can be learned from the available sample:

$$\begin{aligned}
P(y_{it} = 1|\psi_{it}) &= P(y_{it} = 1|\psi_{it}, s_{it} = 0)P(s_{it} = 0|\psi_{it}) + P(y_{it} = 1|\psi_{it}, s_{it} = 1)P(s_{it} = 1|\psi_{it}) \\
&= \frac{\lambda(\psi_{it}|y_{it} = 1, s_{it} = 0)P(y_{it} = 1|s_{it} = 0)}{\lambda(\psi_{it}|s_{it} = 0)} P(s_{it} = 0|\psi_{it}) \\
&\qquad + \frac{\lambda(\psi_{it}|y_{it} = 1, s_{it} = 1)P(y_{it} = 1|s_{it} = 1)}{\lambda(\psi_{it}|s_{it} = 1)} P(s_{it} = 1|\psi_{it}) \\
&= \frac{\lambda(\psi_{it}|y_{it} = 1, s_{it} = 0)P(y_{it} = 1|s_{it} = 0)}{\lambda(\psi_{it}|s_{it} = 0)} \frac{\lambda(\psi_{it}|s_{it} = 0)P(s_{it} = 0)}{\lambda(\psi_{it})} \\
&\qquad + \frac{\lambda(\psi_{it}|y_{it} = 1, s_{it} = 1)P(y_{it} = 1|s_{it} = 1)}{\lambda(\psi_{it}|s_{it} = 1)} \frac{\lambda(\psi_{it}|s_{it} = 1)P(s_{it} = 1)}{\lambda(\psi_{it})} \\
&= \frac{\lambda(\psi_{it}|y_{it} = 1, s_{it} = 0)P(y_{it} = 1|s_{it} = 0)P(s_{it} = 0)}{\lambda(\psi_{it}|s_{it} = 0)P(s_{it} = 0) + \lambda(\psi_{it}|s_{it} = 1)P(s_{it} = 1)} \\
&\qquad + \frac{\lambda(\psi_{it}|y_{it} = 1, s_{it} = 1)P(y_{it} = 1|s_{it} = 1)P(s_{it} = 1)}{\lambda(\psi_{it}|s_{it} = 0)P(s_{it} = 0) + \lambda(\psi_{it}|s_{it} = 1)P(s_{it} = 1)}.
\end{aligned} \tag{F.1}$$

Every portion of equation (F.1) can either be non-parametrically estimated from the available data, or can be deduced from aggregate statistics. The distribution of the linear index in the military service sample, $\lambda(\psi_{it}|y_{it}=1,\cdot)$, can be learned from each of the choice-restricted subsamples. The distribution of this same index in the population, $\lambda(\psi_{it}|\cdot)$ can be learned from each of the supplemental samples. The aggregate enlistment probabilities, $P(y_{it}=1|\cdot)$ are given in Online Appendix D. Finally, $P(s_{it}=0)$ and $P(s_{it}=1)$ can be learned from aggregate data.

In order to discuss the estimation procedure, it is convenient to define an indicator for whether individual $i$ is a member of the choice-restricted or supplementary sample. Define

$$\tilde{y}_{it} = \begin{cases} 1 & \text{for members of the choice-restricted sample} \\ 0 & \text{for members of the supplementary sample} \end{cases}.$$

Observations for which $\tilde{y}_{it}=1$ make it possible to learn the terms in equation (F.1) that are conditional on $y_{it}=1$, while those for which $\tilde{y}_{it}=0$ make it possible to learn the terms that do not condition on $y_{it}=1$ and which are not learned from aggregate data. I adapt the NW estimator and estimate equation (F.1) with the statistic

$$\hat{G}(\psi_{it}) = \widehat{P(y_{it}=1|\psi_{it})} =$$
$$\frac{\left\{\begin{array}{l} P(y_{it}=1|s_{it}=0)P(s_{it}=0)\left[\sum_j \tilde{y}_{j\tau}(1-s_{j\tau})\right]^{-1}\sum_{j\neq i}K\left(\frac{\psi_{it}-\psi_{j\tau}}{\omega}\right)\tilde{y}_{j\tau}(1-s_{j\tau}) \\ \qquad + P(y_{it}=1|s_{it}=1)P(s_{it}=1)\left[\sum_j \tilde{y}_{j\tau}s_{j\tau}\right]^{-1}\sum_{j\neq i}K\left(\frac{\psi_{it}-\psi_{j\tau}}{\omega}\right)\tilde{y}_{j\tau}s_{j\tau} \end{array}\right\}}{\left\{\begin{array}{l} P(s_{it}=0)\left[\sum_j (1-\tilde{y}_{j\tau})(1-s_{j\tau})\right]^{-1}\sum_{j\neq i}K\left(\frac{\psi_{it}-\psi_{j\tau}}{\omega}\right)(1-\tilde{y}_{j\tau})(1-s_{j\tau}) \\ \qquad + P(s_{it}=1)\left[\sum_j (1-\tilde{y}_{j\tau})s_{j\tau}\right]^{-1}\sum_{j\neq i}K\left(\frac{\psi_{it}-\psi_{j\tau}}{\omega}\right)(1-\tilde{y}_{j\tau})s_{j\tau} \end{array}\right\}}, \quad \text{(F.2)}$$

where $\omega$ is a bandwidth and $K(\cdot)$ is a kernel function. For estimation, I use a Gaussian kernel. I use this statistic rather than estimating the model separately for each of the two subpopulations in order to ensure that the estimates of $G(\cdot)$ and the linear index $\psi_{it}$ created by this procedure are comparable across subpopulations and can thus be used in the selection model. Thus, rather than simply maximizing the likelihood (F.5), defined below, separately for each sample, I maximize the sum of the likelihoods for each of the samples, estimating the choice probabilities by (F.2); to accommodate the separate sampling of the two groups of birth cohorts, I weight the likelihood, so that the final likelihood function becomes

$$\mathfrak{L} = L^{\{s_{it}=0\}}\left(\begin{bmatrix}\alpha' & \beta' & \delta'\end{bmatrix}'\right) \times \frac{P(s_{it}=0)}{\Xi(s_{it}=0)} + L^{\{s_{it}=1\}}\left(\begin{bmatrix}\alpha' & \beta' & \delta'\end{bmatrix}'\right) \times \frac{P(s_{it}=1)}{\Xi(s_{it}=1)}, \tag{F.3}$$

where $L^{\{s_{it}=j\}}(\cdot)$ is the likelihood function (F.5) for sample $s_{it}=j$, $j \in \{0,1\}$, $P(s_{it}=j)$ is the population proportion, and $\Xi(s_{it}=j)$ is the sample proportion.

Klein and Spady (1993) also suggest the use of a trimming function, though they report that the particular function is empirically unimportant. When estimating, I first assume that $G(\cdot)$ is normally distributed and estimate a probit model. I then compute a kernel density estimate of the estimated value of $\alpha_t + \mathbf{x}_{it}\beta_k + \mathbf{z}_{it}'\delta_k$, and trim from the estimation (that is, exclude from the likelihood function) individuals for whom the density falls below 0.005.

## F.2 Cosslett's (1981) Likelihood

The likelihood function also requires adaptation to the structure of the sample. Before proceeding further, it is useful to introduce some additional notation:

- $S = \{0,1\}$: the set of options for each individual in the sample, where 0 denotes never enlisting and 1 denotes enlisting in the military at some point in the lifetime
- $N$: the number of individuals in the choice-restricted (military enlister) sample
- $N_0$: the number of individuals in the supplementary (general population) sample
- $H_s = \frac{N_0}{N}$
- $Q_j$: the proportion of each choice $j \in S$ in the population
- $H_j$: the proportion of each choice $j \in S$ in the choice-restricted sample
- $\eta_j = \frac{H_j}{Q_j}$

I denote the choice-restricted sample by $i \in \{1, \ldots, N\}$ and the supplementary sample by $i \in \{N+1, \ldots, N+N_0\}$.

Models of this type are studied by Cosslett (1981), who provides a maximum-likelihood estimator.

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \\ \hat{\delta} \end{bmatrix} = \underset{[\,\alpha'\ \beta'\ \delta'\,]'}{\arg\max}\ L\left([\,\alpha'\ \beta'\ \delta'\,]'\right)$$

$$= \underset{[\,\alpha'\ \beta'\ \delta'\,]'}{\arg\max} \sum_{i=1}^{N} \log\left\{ \frac{\eta_1 P(y_{it}=1|\mathbf{x}_{it},\mathbf{z}_{it};t)}{\left[\sum_{j\in S}\eta_j P(y_{it}=j|\mathbf{x}_{it},\mathbf{z}_{it};t)\right]+H_s} \right\}$$
$$- \sum_{i=N+1}^{N+N_0} \log\left\{ \left[\sum_{j\in S}\eta_j P(y_{it}=j|\mathbf{x}_{it},\mathbf{z}_{it};t)\right]+H_s \right\}. \quad \text{(F.4)}$$

Since the choice-restricted sample contains only enlisters, $\eta_1 = \frac{1}{Q_1}$ and $\eta_0 = \frac{0}{Q_0} = 0$. Thus, the pseudo-log-likelihood function in expression (F.4) reduces to the following:[37]

$$L\left([\,\alpha'\ \beta'\ \delta'\,]'\right)$$
$$= \sum_{i=1}^{N} \log\left\{ \frac{\eta_1 P(y_{it}=1|\mathbf{x}_{it},\mathbf{z}_{it};t)}{\eta_1 P(y_{it}=1|\mathbf{x}_{it},\mathbf{z}_{it};t)+H_s} \right\} - \sum_{i=N+1}^{N+N_0} \log\left\{\eta_1 P(y_{it}=1|\mathbf{x}_{it},\mathbf{z}_{it};t)+H_s\right\}. \quad \text{(F.5)}$$

Finally, because $Q_1$ differs between the two groups of cohorts, I maximize the sum of equation (F.5) evaluated separately for the two subsamples, weighting the sums to account for the separate sampling of each group of birth cohorts, as described above.

### F.3 The Gradient Matrix

When the Gaussian kernel is used, the gradient matrix of the estimated probability with respect to the first-stage coefficients $\Theta_1 = [\alpha', \beta', \delta']$ is given by

$$\frac{\partial \widehat{P(y_{it}=1|\psi_{it})}}{\partial \Theta_{1\kappa}} = \frac{\widehat{P(y_{it}=1|\psi_{it})}}{E+F}(A+B-C-D)$$

[37]Location and scale normalizations are required. To this end, I omit a constant and require that

$$\begin{bmatrix}\alpha' & \beta' & \delta'\end{bmatrix}\begin{bmatrix}\alpha \\ \beta \\ \delta\end{bmatrix} = 1.$$

Note that this implies that the coefficients are identified only up to sign; however, the function $G(\cdot)$ adjusts to ensure that the marginal effect of each covariate is of the appropriate sign.

where

$$
\begin{aligned}
A =& P(s_{it}=0)\left[\sum_j (1-\tilde{y}_{j\tau})(1-s_{j\tau})\right]^{-1} \sum_{j\neq i}\left(\frac{\psi_{it}-\psi_{j\tau}}{\omega}\right)\left(\frac{x_{i\kappa}-x_{j\kappa}}{\omega}\right) \\
&\times K\left(\frac{\psi_{it}-\psi_{j\tau}}{\omega}\right)(1-\tilde{y}_{j\tau})(1-s_{j\tau}) \\
B =& P(s_{it}=1)\left[\sum_j (1-\tilde{y}_{j\tau})s_{j\tau}\right]^{-1} \sum_{j\neq i}\left(\frac{\psi_{it}-\psi_{j\tau}}{\omega}\right)\left(\frac{x_{i\kappa}-x_{j\kappa}}{\omega}\right) K\left(\frac{\psi_{it}-\psi_{j\tau}}{\omega}\right)(1-\tilde{y}_{j\tau})s_{j\tau} \\
C =& P(y_{it}=1|s_{it}=0)P(s_{it}=0)\left[\sum_j \tilde{y}_{j\tau}(1-s_{j\tau})\right]^{-1} \sum_{j\neq i}\left(\frac{\psi_{it}-\psi_{j\tau}}{\omega}\right)\left(\frac{x_{i\kappa}-x_{j\kappa}}{\omega}\right) \\
&\times K\left(\frac{\psi_{it}-\psi_{j\tau}}{\omega}\right)\tilde{y}_{j\tau}(1-s_{j\tau}) \\
D =& P(y_{it}=1|s_{it}=1)P(s_{it}=1)\left[\sum_j \tilde{y}_{j\tau}s_{j\tau}\right]^{-1} \sum_{j\neq i}\left(\frac{\psi_{it}-\psi_{j\tau}}{\omega}\right)\left(\frac{x_{i\kappa}-x_{j\kappa}}{\omega}\right) \\
&\times K\left(\frac{\psi_{it}-\psi_{j\tau}}{\omega}\right)\tilde{y}_{j\tau}s_{j\tau} \\
E =& P(s_{it}=0)\left[\sum_j (1-\tilde{y}_{j\tau})(1-s_{j\tau})\right]^{-1} \sum_{j\neq i} K\left(\frac{\psi_{it}-\psi_{j\tau}}{\omega}\right)(1-\tilde{y}_{j\tau})(1-s_{j\tau}) \\
F =& P(s_{it}=1)\left[\sum_j (1-\tilde{y}_{j\tau})s_{j\tau}\right]^{-1} \sum_{j\neq i} K\left(\frac{\psi_{it}-\psi_{j\tau}}{\omega}\right)(1-\tilde{y}_{j\tau})s_{j\tau}
\end{aligned}
$$

and $x_{it\kappa}$ may be substituted by any element of $\mathbf{x}_{it}$, $\mathbf{z}_{it}$, or a cohort fixed effect (for estimating $\alpha_t$) as appropriate.

## F.4 Computing Partial Effects

The semi-elasticities presented in Table 3 are computed as follows. When a Gaussian kernel is used, differentiating expression (F.2) with respect to $x_{it\kappa}$ yields the partial effect for individual $i$ for cohort $t$ for a continuous covariate $\kappa$:

$$\frac{\partial \log[\widehat{P(y_{it}=1|\psi_{it})}]}{\partial x_{i\kappa}} = \frac{1}{\widehat{P(y_{it}=1|\psi_{it})}}$$

$$\times \frac{\beta_\kappa}{\omega} \times \left[ \frac{-\left\{ \begin{array}{r} P(y_{it}=1|s_{it}=0)P(s_{it}=0)\left[\sum_j \tilde{y}_{j\tau}(1-s_{j\tau})\right]^{-1} \sum_{j\neq i} K\left(\frac{\psi_{it}-\psi_{j\tau}}{\omega}\right)\left(\frac{\psi_{it}-\psi_{j\tau}}{\omega}\right)\tilde{y}_{j\tau}(1-s_{j\tau}) \\ +P(y_{it}=1|s_{it}=1)P(s_{it}=1)\sum_{j\neq i}\left[\sum_j \tilde{y}_{j\tau}s_{j\tau}\right]^{-1} K\left(\frac{\psi_{it}-\psi_{j\tau}}{\omega}\right)\left(\frac{\psi_{it}-\psi_{j\tau}}{\omega}\right)\tilde{y}_{j\tau}s_{j\tau} \end{array}\right\}}{\left\{ \begin{array}{r} P(s_{it}=0)\left[\sum_j (1-\tilde{y}_{j\tau})(1-s_{j\tau})\right]^{-1} \sum_{j\neq i} K\left(\frac{\psi_{it}-\psi_{j\tau}}{\omega}\right)(1-\tilde{y}_{j\tau})(1-s_{j\tau}) \\ +P(s_{it}=1)\left[\sum_j (1-\tilde{y}_{j\tau})s_{j\tau}\right]^{-1} \sum_{j\neq i} K\left(\frac{\psi_{it}-\psi_{j\tau}}{\omega}\right)(1-\tilde{y}_{j\tau})s_{j\tau} \end{array}\right\}} \right.$$

$$-\widehat{P(y_{it}=1|\psi_{it})}$$

$$\left. \times \frac{-\left\{ \begin{array}{r} P(s_{it}=0)\left[\sum_j (1-\tilde{y}_{j\tau})(1-s_{j\tau})\right]^{-1} \sum_{j\neq i} K\left(\frac{\psi_{it}-\psi_{j\tau}}{\omega}\right)\left(\frac{\psi_{it}-\psi_{j\tau}}{\omega}\right)(1-\tilde{y}_{j\tau})(1-s_{j\tau}) \\ +P(s_{it}=1)\left[\sum_j (1-\tilde{y}_{j\tau})s_{j\tau}\right]^{-1} \sum_{j\neq i} K\left(\frac{\psi_{it}-\psi_{j\tau}}{\omega}\right)\left(\frac{\psi_{it}-\psi_{j\tau}}{\omega}\right)(1-\tilde{y}_{j\tau})s_{j\tau} \end{array}\right\}}{\left\{ \begin{array}{r} P(s_{it}=0)\left[\sum_j (1-\tilde{y}_{j\tau})(1-s_{j\tau})\right]^{-1} \sum_{j\neq i} K\left(\frac{\psi_{it}-\psi_{j\tau}}{\omega}\right)(1-\tilde{y}_{j\tau})(1-s_{j\tau}) \\ +P(s_{it}=1)\left[\sum_j (1-\tilde{y}_{j\tau})s_{j\tau}\right]^{-1} \sum_{j\neq i} K\left(\frac{\psi_{it}-\psi_{j\tau}}{\omega}\right)(1-\tilde{y}_{j\tau})s_{j\tau} \end{array}\right\}} \right] \quad \text{(F.6)}$$

For a discrete covariate, the partial effect for individual $i$ is calculated as

$$\frac{\Delta \log[\widehat{P(y_{it}=1|\psi_{it})}]}{\Delta x_{it\kappa}} = \frac{1}{\widehat{P(y_{it}=1|\psi_{it})}} \times \left[ \widehat{P\left(y_{it}=1|\psi_{it}\right)}\Big|_{x_{it\kappa}=1} - \widehat{P\left(y_{it}=1|\psi_{it}\right)}\Big|_{x_{it\kappa}=0} \right]; \quad \text{(F.7)}$$

that is, the difference between the estimated probability of enlistment for each of the two possible values of $x_{it\kappa}$.

The estimates presented in Table 3 are averages of expressions (F.6) and (F.7) across the supplementary samples, thus representing the average marginal effect of the covariate on enlistment in the whole population.

# G Results with Exact Matches Only (For Online Publication)

To address concerns over the role of false positives in linking, I repeat the main results limiting the sample to matches for which it is possible to be certain or nearly certain that there are no false positives. For the Union Army data, no data must be removed because the hand linking by genealogists removes the concern over false positives. For the Regular Army data, I limit the data to individuals whose census and enlistment characteristics (from the enlistment where height data are taken, which is generally the first enlistment) met the following criteria: absolute difference in age-implied birth years not more than one year, same place of birth, and no difference in name, except for double letters and abbreviations (e.g., "William" and "Wm" were allowed to match). These are much stricter requirements than those used for the main results, which allow up to a four-year difference in the age-implied birth year and allow for matching of similar but not identical first and last names. As might be expected from the more stringent requirements imposed on linking, the sample size for the Regular Army is reduced considerably by these refinements. In particular, the number of enlisters observed in the Regular Army data who are "exact" matches is 896, as opposed to 2,214 satisfying the standard criteria used in the main results.

As shown in Figures G.1 and G.2, the main results are largely unaffected by this restriction. Figure G.1 shows a similar pattern to Figure A.1, with a more negative estimated selection function in the later birth cohorts. Similarly, panel G.2(a) shows patterns similar to those of panel 2(a), with the existence of a decline in the corrected trend that is smaller than the decline in the uncorrected trend. However, the Regular Army data in this panel clearly show the much higher variability induced by the smaller samples. Finally, panel G.2(b) shows the confidence interval of the estimated decline. As in panel 2(b), it is possible to reject the null hypothesis of no decline in stature in the period in question ($\chi^2_{28} = 44.49$, $p = 0.025$). Unlike panel 2(b), it is not possible to reject the null hypothesis of no net decline in stature, as the confidence interval of the decline in stature to 1860 includes zero ($\chi^2_1 = 2.04$, $p = 0.153$); but this is not due to a very different estimate of the magnitude relative to the main results (the net decline is estimated as 0.591 inches in panel G.2(b) as opposed to 0.643 inches in panel 2(b)); the difference is largely due to the much higher standard error of this estimate (0.414 as opposed to the main result's standard error of 0.305). Similarly, although it was possible in the main results to reject the null hypothesis of equality of the estimated decline to 1860 between the fully corrected trends and the trends corrected only for selection on observables, it is not possible to do so in this case. Again, this is due to larger standard errors stemming from the loss of a considerable quantity of data rather than due to fundamental changes in the estimates: the difference in estimated declines was

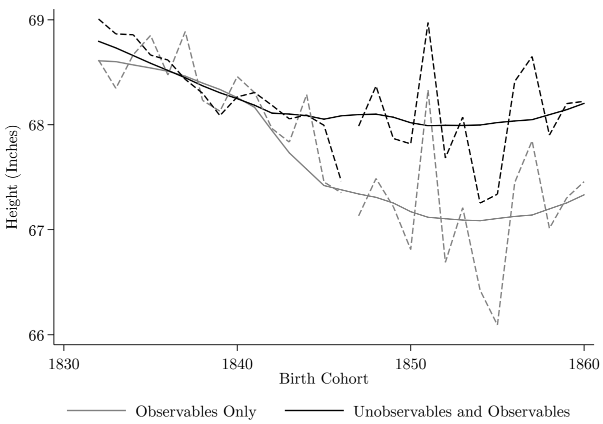
0.647 with a standard error of 0.276 in the main results; after the limitation to exact matches, it is 0.685 with a standard error of 0.398. Thus, I conclude that the main results are not driven by false positives in linkage.


Figure G.1: Estimated $\Omega(\cdot)$ function by birth cohort

*Note:* This graph plots the coefficients from a regression of the estimated function $\hat{\Omega}(\hat{\alpha}_t + \mathbf{x}_{it}'\hat{\beta}_k + \mathbf{z}_{it}'\hat{\delta}_k)$ on birth year indicators, weighting by inverse enlistment probability (in the dashed line), as well as these coefficients smoothed over birth cohorts (in the solid line).

That the main result that the height decline survives the correction for selection on unobservables is not surprising. This result is largely driven by the Union Army data, which are not in danger of false positives because they are hand matched. Thus, removing the inexact matches from the Regular Army data would be unlikely to affect the results in the Union Army data.

(a) Estimated adjusted and unadjusted trends




(b) Confidence interval of adjusted trend


Figure G.2: Trends in average stature

*Note:* Panel G.2(a) plots four trends in average height by birth cohort. The first, in solid black (labeled "Unobservables and Observables"), incorporates the correction for selection on both observables and unobservables, and smoothed over birth cohorts; the second, in dashed black, is its unsmoothed analog. The third, in solid gray (labeled "Observables Only"), is corrected only for truncation and selection on observables, and is smoothed over birth cohorts; the fourth, in dashed gray, is its unsmoothed analog. The unsmoothed trends for the 1832–1846 cohorts are based on the Union Army data, while those for the 1847–1860 cohorts are based on the Regular Army data. Panel G.2(b) presents bootstrap 95 percent pointwise confidence intervals clustered at the county level for the smoothed trend in average stature incorporating the correction for selection on both observables and unobservables (the solid black line in panel G.2(a)).

# H Results for Alternative Specification and Data Set (For Online Publication)

In addition to the Regular Army data for the 1847–1860 cohorts, used in the main text of the paper, I also collected Regular Army data for the birth cohorts of 1832–1846. In this Appendix, I present a specification that estimates the model using only Regular Army data (for the 1832–1860 cohorts) instead of the combination of the Union Army and Regular Army data, as used in the main text of the paper. That is, the data for the 1847–1860 cohorts are the same, but the data for the 1832–1846 cohorts are different.[38] These results are referred to throughout this appendix as those with *Different Data*. I also present results using the same data as in the main text, but replacing Lincoln's vote share with Douglas's vote share in 1860 and Buchanan's vote share in 1856. I refer to these results as those with *Different Variables*.

[38]Details of the census linkage used for construction of this sample are available on request.

## H.1 Summary Statistics for Regular Army, 1832–1846 Cohorts

Table H.1 presents the analog of Table 1 for the 1832–1846 Regular Army data, and Table H.2 presents the analog of Table 2. Note that the supplementary samples in these two tables are identical to those for the 1832–1846 cohorts in the main text because they represent the same birth cohorts. Figure H.1 presents the analog of Figure 1, exhibiting a similar pattern of heaping and left-censoring as the Regular Army sample for the 1847–1860 cohorts.

Table H.1: Distribution of observations by census

| | | 1832–1846 | |
|---|---|---|---|
| | | (1) | (2) |
| Census | Cohorts | RA (CR) | Supp. |
| 1850 | 1832–1841 | 984 | 5,879 |
| 1860 | 1842–1846 | 772 | 2,807 |
| Total | | 1,756 | 8,686 |

*Notes*: Each cell reports the number of individuals in the sample indicated in the column header with data taken from the census indicated in the row. Samples are restricted to cover individuals with data on all individual-level variables. Abbreviations are as follows: RA is Regular Army, CR is choice-restricted sample, Supp. is supplementary sample.

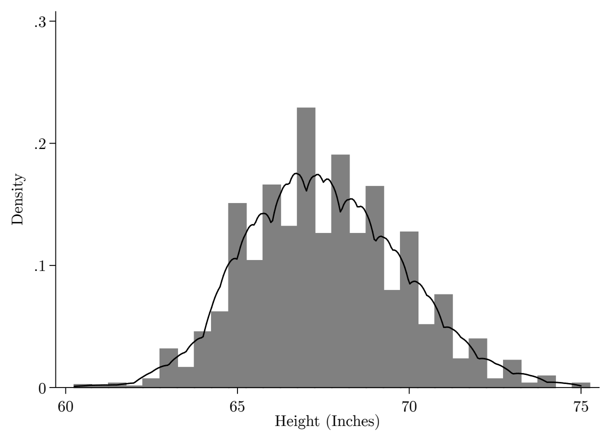



Figure H.1: Height distributions

*Note:* This figure presents a histogram (with a bin width of 0.5 inches) and a kernel density estimate of the height distribution for the height data for individuals from the 1832–1846 birth cohorts in the Regular Army.

## H.2 Selection into Military Service

Table H.3 presents the results of two alternative sets of estimation of the binary choice model (equation 3) in a manner analogous to Table 3. Column (1) presents the estimates of the coefficients $\beta_k$ and $\delta_k$ using

Table H.2: Summary statistics

| | 1832–1846 | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| *Variable* | RA (CR) | Supp. | Diff. |
| *Individual or Household Variables* | | | |
| Height (in) | 67.496 | | |
| Household Owns Property | 0.594 | 0.687 | −0.090*** |
| Household Real Property ($1,000) | 2.008 | 2.297 | −0.272 |
| Related to Head of Household | 0.777 | 0.863 | −0.085*** |
| Household Size | 7.242 | 7.419 | −0.177** |
| Attended School | 0.572 | 0.648 | −0.074*** |
| Household Occupation | | | |
| Farmer | 0.350 | 0.520 | −0.167*** |
| Professional | 0.052 | 0.038 | 0.015** |
| Clerical | 0.091 | 0.066 | 0.024* |
| Skilled and Artisan | 0.239 | 0.185 | 0.055*** |
| Semi-Skilled and Operative | 0.076 | 0.058 | 0.017* |
| Unskilled | 0.095 | 0.065 | 0.029*** |
| Farm Labor | 0.006 | 0.006 | 0.000 |
| Unproductive | 0.091 | 0.062 | 0.026*** |
| Birth Region | | | |
| Midwest | 0.213 | 0.287 | −0.073*** |
| Northeast | 0.697 | 0.541 | 0.162*** |
| South | 0.090 | 0.172 | −0.089*** |
| *County Variables* | | | |
| Fraction Urban | 0.242 | 0.158 | 0.077*** |
| Wheat Bushels per capita | 5.650 | 5.882 | −0.232 |
| Milk Cows per capita | 0.262 | 0.282 | −0.020** |
| Swine per capita | 0.606 | 0.972 | −0.367*** |
| Value of Agricultural Production per capita | 0.045 | 0.048 | −0.003** |
| Lincoln Vote Share (1860) | 0.515 | 0.455 | 0.060*** |
| Observations | 1,674 | 8,535 | |

*Significance levels*: *** p<0.01, ** p<0.05, * p<0.1
*Notes*: Averages for the choice-restricted samples are weighted to correct for selection into linkage on the basis of observable characteristics. Standard deviations and standard errors are omitted for space. Sample sizes are the minimum of the column with observations for all variables. Abbreviations are as follows: RA is Regular Army, CR is a choice-restricted sample, and Supp. is a supplementary sample. Diff. is a difference.

the same data as the benchmark results in the main text of the paper but with the Douglas and Buchanan vote shares instead of the Lincoln vote share for identification. Column (2) presents the associated semi-elasticities. This set of results exhibits magnitudes of the relationship of the vote shares for Douglas and Buchanan that are comparable in interpretation (though not in magnitude) to that of Lincoln's vote share in Table 3 for the Regular Army. Column (3) presents the estimates of the coefficients $\beta_k$ and $\delta_k$ using the same specification as the benchmark results, but replacing the Union Army data for the 1832–1846 cohorts with the alternative Regular Army data set for the same cohorts. Column (4) presents the associated semi-elasticities. These columns show a small and statistically insignificant role for the vote share in the military enlistment decision;[39] the model is still identified based on the difference in coefficients between the two cohort groups.

## H.3 Selection-Corrected Height Regressions

Results of estimation of equation (4) for alternative specifications are presented in Table H.4. Columns (1)–(4) use the Union Army data to represent the 1832–1846 cohorts, as in the benchmark specification, but base identification on the Buchanan and Douglas vote shares rather than on Lincoln's. Columns (5)–(8) base identification on Lincoln's vote share, and use the Regular Army to represent the 1832–1846 cohorts. Results are largely similar to those of the benchmark sample in Table 4. A key test provided in this Table is the overidentification test of columns (3) and (7). In the specification using Douglas's and Buchanan's vote shares, the vote share for Douglas enters marginally significantly before correction, but significance is lost and the coefficient decreases in magnitude after correction. The coefficient on Buchanan's vote share also decreases in magnitude after correction. In the specification using the Regular Army sample, excludability of the vote share is also supported by the lack of a statistically significantly coefficient.

## H.4 Adjusted Trends in Height

The main results for the alternative specifications are presented in Figures H.2–H.4. When the Douglas and Buchanan vote shares are used instead of Lincoln's, the results are qualitatively identical to those of the benchmark specification. As a result, I do not discuss them further. When the Regular Army enlisters are used to represent the 1832–1846 cohorts, however, differences are present. The first main result—that the Antebellum Puzzle is robust to the correction for sample-selection bias—is present in this specification as

[39] This estimation requires computation of $Q_1$ for this alternative Regular Army sample. Using methods similar to that for the 1847–1860 cohorts, I compute it as $Q_1 = \frac{90{,}201}{2{,}469{,}663} = 0.037$.

Table H.3: Binary choice model estimation

| | Different Variables | | | | Different Data | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | | (2) | | (3) | | (4) | |
| | (a) | (b) | (a) | (b) | (a) | (b) | (a) | (b) |
| *Cohorts* | '32–'46 | '47–'60 | '32–'46 | '47–'60 | '32–'46 | '47–'60 | '32–'46 | '47–'60 |
| *Variables* | UA | RA | UA | RA | RA | RA | RA | RA |
| *Individual or Household Variables* | | | | | | | | |
| Household Owns Property | 0.092*** | 0.195*** | 0.254*** | 0.455*** | 0.146*** | 0.272*** | −0.274*** | 0.451*** |
| | (0.017) | (0.020) | (0.054) | (0.065) | (0.039) | (0.030) | (0.052) | (0.039) |
| Household Real Property (1,000) | −0.016*** | −0.049*** | −0.050*** | −0.128*** | 0.006** | −0.019*** | −0.008** | −0.035*** |
| | (0.002) | (0.004) | (0.005) | (0.012) | (0.003) | (0.003) | (0.003) | (0.004) |
| Related to Head of Household | 0.103*** | −0.061*** | 0.263*** | −0.187*** | 0.209*** | −0.085*** | −0.332*** | −0.174*** |
| | (0.022) | (0.019) | (0.054) | (0.069) | (0.049) | (0.029) | (0.064) | (0.064) |
| Household Size | 0.018*** | 0.001 | 0.056*** | 0.002 | 0.014** | −0.001 | −0.020** | −0.002 |
| | (0.003) | (0.003) | (0.010) | (0.007) | (0.007) | (0.004) | (0.009) | (0.007) |
| Attended School | −0.187*** | −0.157*** | −0.634*** | −0.540*** | 0.073** | −0.195*** | −0.144*** | −0.420*** |
| | (0.017) | (0.016) | (0.048) | (0.067) | (0.036) | (0.022) | (0.050) | (0.052) |
| Household Occupation (Unproductive excluded) | | | | | | | | |
| Farmer | −0.310*** | −0.214*** | −0.994*** | −0.520*** | 0.155** | −0.176*** | −0.313*** | −0.321*** |
| | (0.038) | (0.024) | (0.100) | (0.094) | (0.065) | (0.042) | (0.093) | (0.072) |
| Professional | −0.256*** | −0.093** | −0.704*** | −0.189*** | −0.070 | −0.006 | 0.103 | −0.011 |
| | (0.062) | (0.036) | (0.139) | (0.052) | (0.093) | (0.062) | (0.126) | (0.119) |
| Clerical | −0.343*** | −0.147*** | −0.927*** | −0.279*** | 0.027 | −0.056 | −0.031 | −0.097 |
| | (0.046) | (0.028) | (0.081) | (0.041) | (0.082) | (0.048) | (0.104) | (0.075) |
| Skilled and Artisan | −0.287*** | −0.089*** | −0.819*** | −0.206*** | 0.048 | 0.020 | −0.069 | 0.038 |
| | (0.040) | (0.020) | (0.082) | (0.043) | (0.068) | (0.038) | (0.091) | (0.071) |
| Semi-Skilled and Clerical | −0.300*** | −0.127*** | −0.817*** | −0.249*** | 0.056 | −0.019 | −0.077 | −0.034 |
| | (0.050) | (0.025) | (0.099) | (0.039) | (0.088) | (0.045) | (0.109) | (0.079) |
| Unskilled | −0.282*** | −0.096*** | −0.767*** | −0.202*** | −0.075 | 0.042 | 0.119 | 0.081 |
| | (0.046) | (0.024) | (0.086) | (0.043) | (0.081) | (0.043) | (0.113) | (0.087) |
| Farm Labor | −0.184 | −0.126*** | −0.520 | −0.233*** | −0.009 | −0.041 | 0.015 | −0.072 |
| | (0.120) | (0.030) | (0.335) | (0.042) | (0.233) | (0.059) | (0.276) | (0.086) |
| Birth Region (South excluded) | | | | | | | | |
| Midwest | 0.391*** | 0.009 | 1.622*** | 0.024 | −0.121 | −0.141*** | 0.169 | −0.251*** |
| | (0.036) | (0.025) | (0.134) | (0.067) | (0.097) | (0.047) | (0.123) | (0.082) |
| Northeast | 0.319*** | 0.015 | 1.010*** | 0.039 | −0.081 | −0.087* | 0.121 | −0.162* |
| | (0.037) | (0.020) | (0.084) | (0.057) | (0.094) | (0.048) | (0.128) | (0.087) |
| *County Variables* | | | | | | | | |
| Fraction Urban | −0.166*** | 0.040 | −0.529*** | 0.105 | 0.056 | 0.133*** | −0.082 | 0.247*** |
| | (0.051) | (0.031) | (0.158) | (0.086) | (0.096) | (0.049) | (0.121) | (0.080) |
| Wheat Bushels per capita | 0.015*** | −0.003** | 0.048*** | −0.009** | −0.002 | −0.003 | 0.003 | −0.005 |
| | (0.002) | (0.001) | (0.004) | (0.004) | (0.003) | (0.002) | (0.004) | (0.003) |
| Milk Cows per capita | 0.060 | −0.062 | 0.190 | −0.162 | 0.092 | −0.029 | −0.133 | −0.054 |
| | (0.064) | (0.067) | (0.209) | (0.180) | (0.144) | (0.100) | (0.196) | (0.188) |
| Swine per capita | 0.051*** | −0.080*** | 0.162*** | −0.209*** | 0.209*** | −0.151*** | −0.303*** | −0.281*** |
| | (0.011) | (0.018) | (0.032) | (0.044) | (0.040) | (0.024) | (0.037) | (0.041) |
| Value of Agricultural Production per capita | −0.029 | −0.038 | −0.092 | −0.098 | −0.022 | −0.031 | 0.032 | −0.058 |
| | (0.545) | (0.451) | (1.628) | (1.246) | (1.172) | (0.679) | (1.478) | (1.148) |
| Lincoln Vote Share (1860) | | | | | 0.016 | −0.010 | −0.023 | −0.019 |
| | | | | | (0.145) | (0.081) | (0.190) | (0.144) |
| Buchanan Vote Share (1856) | −0.091 | −0.157*** | −0.289 | −0.408*** | | | | |
| | (0.064) | (0.055) | (0.203) | (0.143) | | | | |
| Douglas Vote Share (1860) | 0.113** | −0.026 | 0.362*** | −0.067 | | | | |
| | (0.045) | (0.037) | (0.131) | (0.097) | | | | |
| Observations | 13,570 | 11,000 | 13,570 | 11,000 | 10,249 | 11,271 | 10,249 | 11,271 |

*Significance levels*: *** p<0.01, ** p<0.05, * p<0.1
*Notes*: Columns (1) and (3) present estimates of the coefficients $\beta$ and $\delta$ from the binary choice model. Columns (2) and (4) present the average semi-elasticity of the impact of each variable on enlistment probability as implied by the estimates of columns (1) and (3), respectively. All specifications include cohort indicators. Standard errors are clustered at the county level. UA denotes Union Army. RA denotes Regular Army.

Table H.4: Height regressions

| | Different Variables | | | | Different Data | | | |
|---|---|---|---|---|---|---|---|---|
| *Variables* | (1) Corr | (2) Not | (3) Corr | (4) Not | (5) Corr | (6) Not | (7) Corr | (8) Not |
| *Individual or Household Variables* | | | | | | | | |
| Household Owns Property | 0.083 | 0.084 | 0.082 | 0.087 | −0.097 | −0.064 | −0.103 | −0.069 |
| | (0.098) | (0.113) | (0.098) | (0.113) | (0.110) | (0.133) | (0.110) | (0.133) |
| Household Real Property (1,000) | −0.021* | −0.006 | −0.020 | −0.006 | 0.003 | −0.002 | 0.004 | −0.002 |
| | (0.012) | (0.009) | (0.012) | (0.008) | (0.008) | (0.007) | (0.008) | (0.007) |
| Related to Head of Household | 0.250** | 0.305** | 0.248** | 0.299** | 0.283** | 0.279* | 0.286** | 0.278* |
| | (0.105) | (0.134) | (0.105) | (0.134) | (0.129) | (0.143) | (0.129) | (0.143) |
| Household Size | 0.026* | 0.023 | 0.027** | 0.025 | 0.025 | 0.016 | 0.025 | 0.017 |
| | (0.014) | (0.015) | (0.014) | (0.015) | (0.015) | (0.019) | (0.015) | (0.019) |
| Attended School | 0.040 | 0.051 | 0.046 | 0.058 | 0.204** | 0.230** | 0.195** | 0.217** |
| | (0.080) | (0.085) | (0.080) | (0.085) | (0.096) | (0.102) | (0.096) | (0.103) |
| Household Occupation (Unproductive excluded) | | | | | | | | |
| Farmer | 0.166 | 0.041 | 0.161 | 0.024 | 0.173 | 0.131 | 0.175 | 0.128 |
| | (0.140) | (0.125) | (0.139) | (0.124) | (0.176) | (0.200) | (0.176) | (0.200) |
| Professional | 0.189 | 0.010 | 0.183 | −0.007 | 0.081 | 0.093 | 0.086 | 0.100 |
| | (0.211) | (0.238) | (0.210) | (0.238) | (0.230) | (0.271) | (0.230) | (0.272) |
| Clerical | −0.023 | −0.279 | −0.034 | −0.303 | −0.099 | −0.286 | −0.093 | −0.279 |
| | (0.181) | (0.198) | (0.179) | (0.195) | (0.187) | (0.245) | (0.186) | (0.244) |
| Skilled and Artisan | −0.171 | −0.336** | −0.173 | −0.345** | −0.300* | −0.415** | −0.301* | −0.414** |
| | (0.148) | (0.150) | (0.147) | (0.149) | (0.164) | (0.203) | (0.164) | (0.202) |
| Semi-Skilled and Clerical | 0.058 | −0.127 | 0.051 | −0.144 | −0.093 | −0.214 | −0.082 | −0.202 |
| | (0.189) | (0.218) | (0.188) | (0.216) | (0.202) | (0.252) | (0.202) | (0.252) |
| Unskilled | 0.220 | 0.081 | 0.226 | 0.080 | 0.048 | −0.009 | 0.042 | −0.013 |
| | (0.179) | (0.191) | (0.179) | (0.191) | (0.194) | (0.230) | (0.194) | (0.229) |
| Farm Labor | −0.277 | −0.535* | −0.281 | −0.546* | −0.284 | −0.436 | −0.286 | −0.439 |
| | (0.231) | (0.287) | (0.232) | (0.288) | (0.254) | (0.315) | (0.254) | (0.316) |
| Birth Region (South excluded) | | | | | | | | |
| Midwest | 0.031 | 0.054 | −0.038 | −0.057 | 0.259 | 0.165 | 0.001 | −0.100 |
| | (0.169) | (0.174) | (0.171) | (0.181) | (0.167) | (0.196) | (0.212) | (0.251) |
| Northeast | −0.223 | −0.302 | −0.265 | −0.380** | −0.022 | −0.172 | −0.295 | −0.460* |
| | (0.171) | (0.186) | (0.166) | (0.187) | (0.155) | (0.195) | (0.212) | (0.256) |
| *County Variables* | | | | | | | | |
| Fraction Urban | −0.530** | −0.617** | −0.569*** | −0.691*** | −0.546** | −0.595** | −0.533** | −0.586** |
| | (0.206) | (0.247) | (0.207) | (0.253) | (0.226) | (0.269) | (0.231) | (0.273) |
| Wheat Bushels per capita | 0.003 | 0.007 | 0.003 | 0.007 | −0.001 | 0.008 | −0.002 | 0.007 |
| | (0.006) | (0.006) | (0.006) | (0.006) | (0.007) | (0.007) | (0.007) | (0.008) |
| Milk Cows per capita | 0.028 | 0.077 | 0.011 | 0.029 | 0.035 | −0.001 | −0.062 | −0.103 |
| | (0.209) | (0.206) | (0.212) | (0.215) | (0.352) | (0.415) | (0.361) | (0.431) |
| Swine per capita | 0.094* | 0.072 | 0.083 | 0.066 | 0.186* | 0.091 | 0.248** | 0.137 |
| | (0.051) | (0.052) | (0.053) | (0.053) | (0.108) | (0.090) | (0.112) | (0.092) |
| Value of Agricultural production per capita | −4.252** | −4.127* | −4.284** | −4.501* | −4.408* | −5.344* | −4.683* | −5.609* |
| | (2.099) | (2.320) | (2.144) | (2.374) | (2.556) | (3.126) | (2.547) | (3.119) |
| Lincoln Vote Share (1860) | | | | | | | 0.632* | 0.653 |
| | | | | | | | (0.341) | (0.399) |
| Buchanan Vote Share (1856) | | | −0.069 | −0.343 | | | | |
| | | | (0.297) | (0.360) | | | | |
| Douglas Vote Share (1860) | | | 0.305 | 0.450* | | | | |
| | | | (0.223) | (0.268) | | | | |
| Observations | 7,102 | 6,730 | 7,102 | 6,730 | 3,860 | 3,686 | 3,860 | 3,686 |

*Significance levels*: *** p<0.01, ** p<0.05, * p<0.1

*Notes*: Standard errors in parentheses. Dependent variable is height, measured in inches. All specifications include age-of-measurement, year-of-birth, and household occupation indicators. The selection-corrected specifications, indicated by the column header Corr, also include selection-correction function $\Omega(\cdot)$. The uncorrected specifications, indicated by the column header Not, correct for trunction with a truncation point of 64 inches. Standard errors are clustered at the county level. Specification (2) covers the 1832–1846 cohorts by the Union Army sample. Specification (3) covers the 1832–1846 cohorts by the Regular Army sample. the difference in sample sizes between columns is the result of the need to drop heights below 64 inches in the selection-corrected regressions when not correcting for sample-selection bias.

well. A decline in average stature is present in the fully corrected trend, and although it is not possible to reject the null of no net decline in stature over the study period, it is possible to reject the null of no decline at all ($\chi^2_{28} = 54.65$, $p < 0.01$). It might be objected that with the Lincoln vote share not entering significantly into the military enlistment decision for this sample, there is insufficient power to correct for sample-selection bias. However, the results for the benchmark specification showed that much of the change in the patterns in stature due to the correction came at the change in samples. With the other source of identification still present in this sample, such changes should have been present but are not.

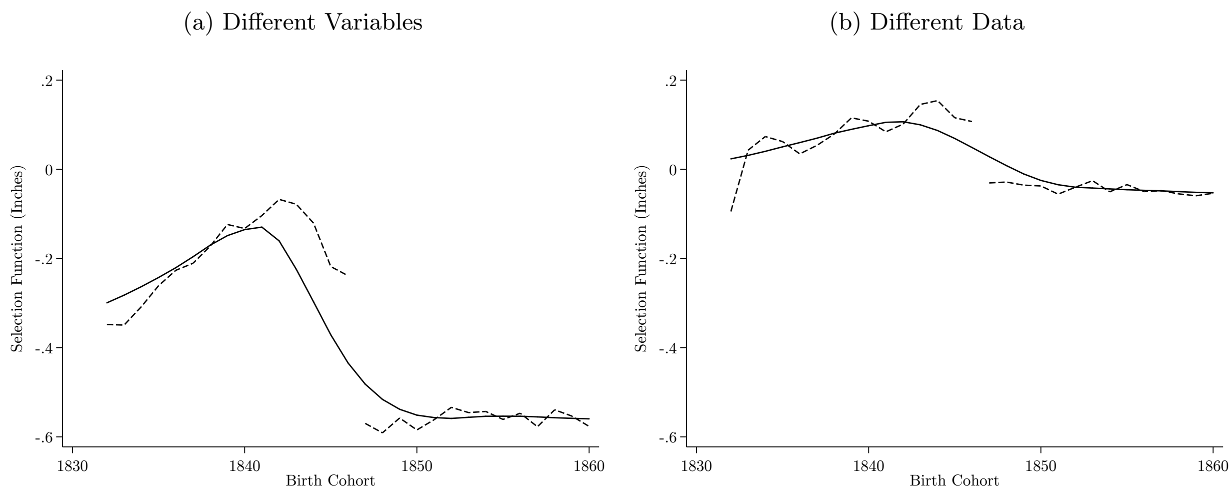



Figure H.2: Estimated $\Omega(\cdot)$ function by birth cohort

*Note:* Each graph plots the coefficients from a regression of the estimated function $\hat{\Omega}(\hat{\alpha}_t + \mathbf{x}_{it}'\hat{\beta}_k + \mathbf{z}_{it}'\hat{\delta}_k)$ on birth year indicators, weighting by inverse enlistment probability (in the dashed line), as well as these coefficients smoothed over birth cohorts (in the solid line). Panel H.2(a) presents the graphs using the Union Army sample to represent the 1832–1846 birth cohorts with identification based on the vote shares for Buchanan and Douglas, while Panel H.2(b) presents the graphs using the Regular Army sample to represent the 1832–1846 birth cohorts and use the vote share for Lincoln for identification; both panels use the Regular Army sample to represent the 1847–1860 cohorts.

The second main result—that correcting for sample-selection bias leads to meaningful and statistically significant changes in the patterns in average stature—is not replicated in the Regular Army-only sample. Indeed, visual inspection of Figures H.2(b) and H.3(b) shows only a small influence of the selection correction. This result is consistent with the different sources of data between this and the benchmark specification. In the benchmark specification, the change in institution with the end of the Civil War was responsible for the change in selection. When the institution remains the same over cohorts no such pattern is present.

The correction for sample-selection bias also contributes to solving another puzzle in the data. In particular, although they are in principle (if issues of selection are ignored) meant to represent the same populations, the sample using the Union Army data to represent the 1832–1846 cohorts and the sample using the Reg-

(a) Different Variables

(b) Different Data

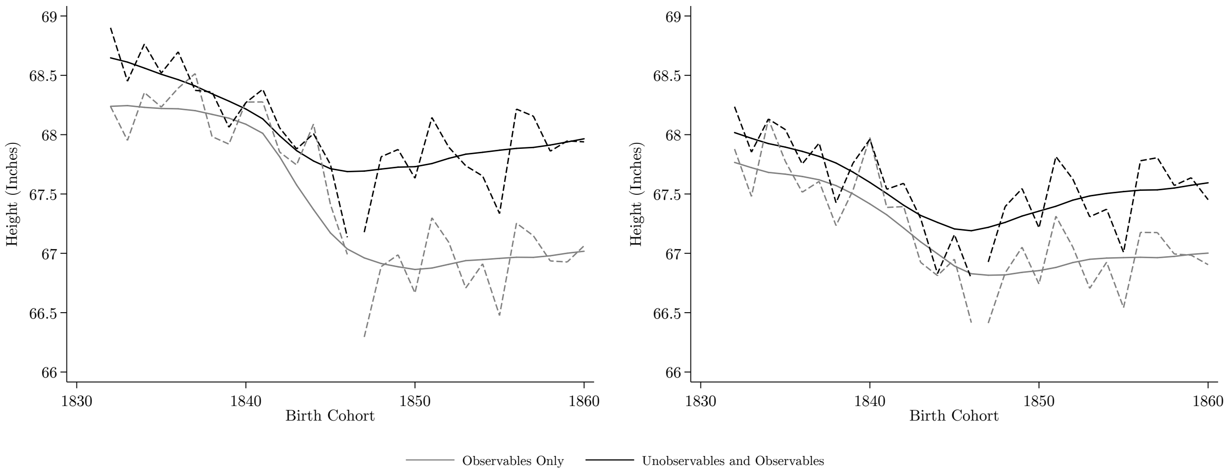



Figure H.3: Estimated adjusted trends by birth cohort

*Note:* Each graph plots four trends in average height by birth cohort. The first, in solid black (labeled "Unobservables and Observables"), incorporates the correction for selection on both observables and unobservables, and smoothed over birth cohorts; the second, in dashed black, is its unsmoothed analog. The third, in solid gray (labeled "Observables Only"), is corrected only for truncation and selection on observables, and is smoothed over birth cohorts; the fourth, in dashed gray, is its unsmoothed analog. The unsmoothed trends for the 1832–1846 cohorts are based on the Union Army data, while those for the 1847–1860 cohorts are based on the Regular Army data. Panel H.3(a) presents the graphs using the Union Army sample to represent the 1832–1846 birth cohorts with identification based on the vote shares for Buchanan and Douglas, while Panel H.3(b) presents the graphs using the Regular Army sample to represent the 1832–1846 birth cohorts and use the vote share for Lincoln for identification; both panels use the Regular Army sample to represent the 1847–1860 cohorts.

(a) Different Variables

(b) Different Data

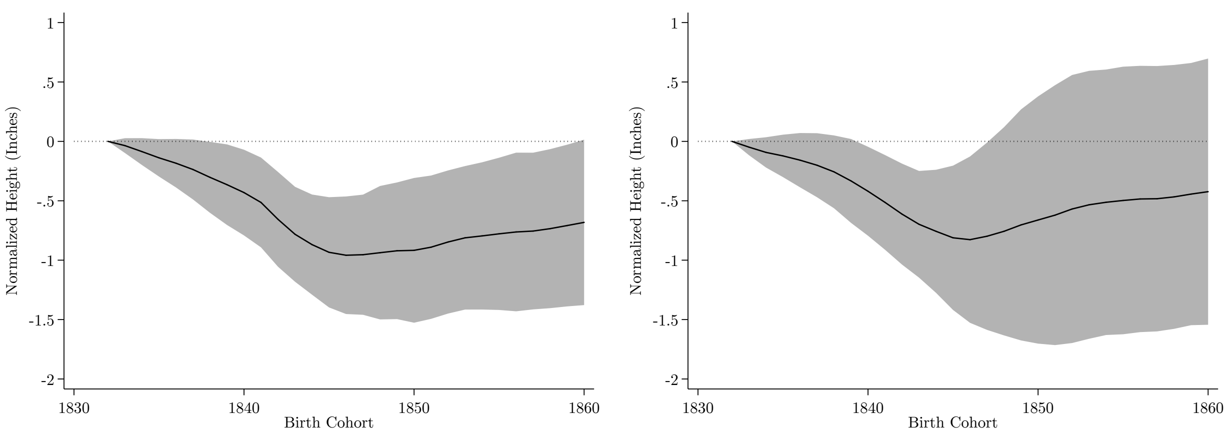



Figure H.4: Confidence intervals of adjusted decline

*Note:* Each graph presents bootstrap 95 percent pointwise confidence intervals clustered at the county level for the smoothed trend in average stature incorporating the correction for selection on both observables and unobservables (the solid black lines in Figure H.3). Panel H.4(a) presents the graphs using the Union Army sample to represent the 1832–1846 birth cohorts and bases identification on the Buchanan and Douglas vote shares, while Panel H.4(b) presents the graphs using the Regular Army sample to represent the 1832–1846 birth cohorts and bases identification on the vote share for Lincoln; both panels use the Regular Army sample to represent the 1847–1860 cohorts.

ular Army data to represent these cohorts give very different estimates for the decline in average stature over time before correction. In particular, before correction, the estimated net declines are 1.29 inches in the benchmark sample and 0.94 inches in the alternative sample. After the correction, the estimated net declines are 0.64 inches and 0.42 inches, respectively. The declines to 1846 are 1.24 inches in the benchmark sample and 0.94 inches in the alternative sample before correction, and 0.94 and 0.83 inches, respectively, after correction. Thus, I conclude that different sample-selection bias between the different armies is at least partially responsible for the different implications of the two samples.

## H.5 Cross-Sectional Patterns

Tables H.5 and H.6 present the results for the cross-sectional comparisons. The results here are similar to those of the temporal trends. Replacing the Lincoln vote share with the Douglas and Buchanan vote shares yields much the same results as does the benchmark specification, though the $p$-values for tests of significance of the effects of correcting for selection are slightly higher. Representing the 1832–1846 cohorts with Regular Army data shows a continued presence of the differences of interest, and only a small effect of correction.

Table H.5: Tests for differences in levels, regional decomposition

| | Different Variables | | | Different Data | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| *Region* | Northeast | Midwest | South | Northeast | Midwest | South |
| *Panel A: Observables Only* | | | | | | |
| Northeast | 66.714*** | | | 66.731*** | | |
| | (0.220) | | | (0.203) | | |
| Midwest | −0.563*** | 67.277*** | | −0.532*** | 67.263*** | |
| | (0.169) | (0.258) | | (0.148) | (0.235) | |
| South | −0.528** | 0.034 | 67.243*** | −0.572** | −0.040 | 67.303*** |
| | (0.241) | (0.290) | (0.306) | (0.250) | (0.271) | (0.325) |
| *Panel B: Unobservables and Observables* | | | | | | |
| Northeast | 67.633*** | | | 67.225*** | | |
| | (0.442) | | | (0.483) | | |
| Midwest | −0.386** | 68.019*** | | −0.502*** | 67.726*** | |
| | (0.175) | (0.454) | | (0.157) | (0.487) | |
| South | −0.376* | 0.010 | 68.009*** | −0.417** | 0.085 | 67.641*** |
| | (0.200) | (0.231) | (0.535) | (0.206) | (0.254) | (0.547) |
| *Panel C: $B - A$* | | | | | | |
| Northeast | 0.919* | | | 0.494 | | |
| | (0.471) | | | (0.504) | | |
| Midwest | 0.176 | 0.742 | | 0.030 | 0.464 | |
| | (0.156) | (0.520) | | (0.124) | (0.495) | |
| South | 0.152 | −0.024 | 0.767 | 0.156 | 0.125 | 0.338 |
| | (0.133) | (0.148) | (0.482) | (0.124) | (0.146) | (0.508) |
| Observations | 3,254 | 3,107 | 741 | 2,229 | 1,189 | 442 |

*Significance levels*: *** p< 0.01, ** p< 0.05, * p< 0.1

*Notes*: In Panels A and B, the diagonals present the estimated mean heights in each region, corrected for minimum height requirements with a truncation point of 64 inches, for the type of selection in the panel title, for measurement age, and for the separate sampling of the two groups of birth cohorts. The off-diagonals present the differences between the diagonal elements. Panel C presents differences between Panels A and B. In all cases, bootstrap standard errors clustered at the county level are in parentheses. Columns (1)–(3) present results using the Union Army to represent the 1832–1846 cohorts, with identification based on Lincoln's vote share. Columns (4)–(6) present results using the Regular Army to represent the 1832–1846 cohorts, with identification based on the Buchanan and Douglas vote shares. Observation numbers are for the region in the column header for the estimates of Panel B.

Table H.6: Tests for differences in levels, sectoral decomposition

| | Different Variables | | Different Data | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| *Sector* | Urban | Rural | Urban | Rural |
| *Panel A: Observables Only* | | | | |
| Urban | 66.764*** | | 66.802*** | |
| | (0.226) | | (0.214) | |
| Rural | −0.563*** | 67.327*** | −0.496*** | 67.297*** |
| | (0.152) | (0.257) | (0.160) | (0.248) |
| *Panel B: Unobservables and Observables* | | | | |
| Urban | 67.695*** | | 67.316*** | |
| | (0.446) | | (0.488) | |
| Rural | −0.331** | 68.026*** | −0.378*** | 67.695*** |
| | (0.142) | (0.465) | (0.138) | (0.488) |
| *Panel C: $B - A$* | | | | |
| Urban | 0.931** | | 0.515 | |
| | (0.469) | | (0.504) | |
| Rural | 0.233* | 0.699 | 0.117 | 0.398 |
| | (0.124) | (0.501) | (0.091) | (0.492) |
| Observations | 2,916 | 4,186 | 1,532 | 2,328 |

*Significance levels*: *** $p< 0.01$, ** $p< 0.05$, * $p< 0.1$
*Notes*: In Panels A and B, the diagonals present the estimated mean heights for each sector, corrected for minimum height requirements with a truncation point of 64 inches, for the type of selection in the panel title, for measurement age, and for the separate sampling of the two groups of birth cohorts. The off-diagonals present the differences between the diagonal elements. Panel C presents differences between Panels A and B. In all cases, bootstrap standard errors clustered at the county level are in parentheses. The urban sector is defined as a county with a non-zero urban population. Columns (1)–(2) present results using the Union Army to represent the 1832–1846 cohorts, with identification based on Lincoln's vote share. Columns (3)–(4) present results using the Regular Army to represent the 1832–1846 cohorts, with identification based on the Buchanan and Douglas vote shares. Observation numbers are for the sector in the column header for the estimates of Panel B.

# References (For Online Publication)


Ancestry.com (2007). *U.S. Army, Register of Enlistments, 1798–1914* [database on-line]. Provo: Ancestry.com Operations Inc.

——— (2009a). *1860 United States Federal Census* [database on-line]. Provo: Ancestry.com Operations Inc.

——— (2009b). *1870 United States Federal Census* [database on-line]. Provo: Ancestry.com Operations Inc.

Biavaschi, Costanza, Corrado Giulietti, and Zahra Siddique (2017). "The Economic Payoff of Name Americanization." *Journal of Labor Economics* 35:4, pp. 1089–1116.

Cosslett, Stephen R. (1981). "Efficient Estimation of Discrete-Choice Models." In *Structural Analysis of Discrete Data with Econometric Applications*. Charles F. Manski and Daniel McFadden (ed.). Cambridge: MIT Press, 1990. Chap. 2, pp. 51–111.

Ferrie, Joseph P. (1996). "A New Sample of Males Linked from the Public Use Microdata Sample of the 1850 US Federal Census to the 1860 US Federal Census Manuscript Schedules." *Historical Methods* 29:4, pp. 141–156.

Gould, Benjamin Apthorp (1869). *Investigations in the Military and Anthropological Statistics of American Soldiers*. Sanitary Memoirs of the War of the Rebellion. Collected and Published by the United States Sanitary Commission. New York: Hurd and Houghton.

Klein, Roger W. and Richard H. Spady (1993). "An Efficient Semiparametric Estimator for Binary Response Models." *Econometrica* 61:2, pp. 387–421.

Mroz, Thomas A. (2015). "Sample Selection with Multiple Selection Indicators in Two-Step Estimators." Mimeo., Georgia State University.

Nadaraya, Elizbar A. (1964). "On Estimating Regression." *Theory of Probability & Its Applications* 9:1, pp. 141–142.

*Register of Enlistments in the U.S. Army, 1798–1914* (n.d.). (National Archives Microfilm Publication M233, 81 Rolls), Records of the Adjutant General's Office, 1780's–1917, RG 94. Washington, DC: National Archives.

Ruggles, Steven, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek (2015). *Integrated Public Use Microdata Series: Version 6.0* [machine-readable database]. Minneapolis: University of Minnesota.

Watson, Geoffrey S. (1964). "Smooth Regression Analysis." *Sankhya: The Indian Journal of Statistics* 26:4, pp. 359–372.