SAMPLE-SELECTION BIAS AND HEIGHT TRENDS IN THE NINETEENTH-CENTURY
UNITED STATES

Ariell Zimran

Sample-Selection Bias and Height Trends in the Nineteenth-Century United States
Ariell Zimran
NBER Working Paper No. 24815
July 2018
JEL No. I15,N11,N31

ABSTRACT

After adjusting for sample-selection bias, I find a net decline in average stature of 0.64 inches in the birth cohorts of 1832--1860 in the US. This result supports the veracity of the Antebellum Puzzle—a deterioration of health during early modern economic growth in the US. However, this adjustment alters the trend in average stature, validating concerns over bias in the historical heights literature. The adjustment is based on census-linked military height data and uses a two-step semi-parametric sample-selection model to adjust for selection on observables and unobservables.

Ariell Zimran
Department of Economics
Vanderbilt University
2301 Vanderbilt Place
Nashville, TN 37235
and NBER
ariell.zimran@vanderbilt.edu

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 argument 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 view 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 provide 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

4

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 γ_t are cohort-specific intercepts, $\mathbf{x}_{it}$ is a vector of covariates affecting both height and military enlistment, and ε_{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 α_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 β_k and δ_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 ε_{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 $G(\alpha_t + \mathbf{x}_{it}\beta_k + \mathbf{z}'_{it}\delta_k)$, ε_{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 ξ_{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 (ξ_{it}). Note that if ε_{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 $t \in \{1832, \ldots, 1860\}$. 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., ε_{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 ε_{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 γ_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 ε_{it} and u_{it} are uncorrelated. In this paper, I impose no such assumption, allowing ε_{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 ε_{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 β and δ 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 α_t, and variation in β_k and δ_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 α_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 of the

---

[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.

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] Columns (1) and (3) present information for the choice-restricted samples of military enlisters. In this Table and throughout the paper I use the shorthand "UA" to refer to the Union Army (1832–1846 cohorts) and "RA" to refer to the Regular Army (1847–1860 cohorts). Columns (2) and (4) present information for the supplementary random sample of census information from the population as a whole. 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

Census	Cohorts	1832–1846		1847–1860	
		(1) UA (CR)	(2) Supp.	(3) RA (CR)	(4) 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.

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 relevant.

---

[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.

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 ($1,000)	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: All Individual or Household Variables are binary unless indicated otherwise. 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.
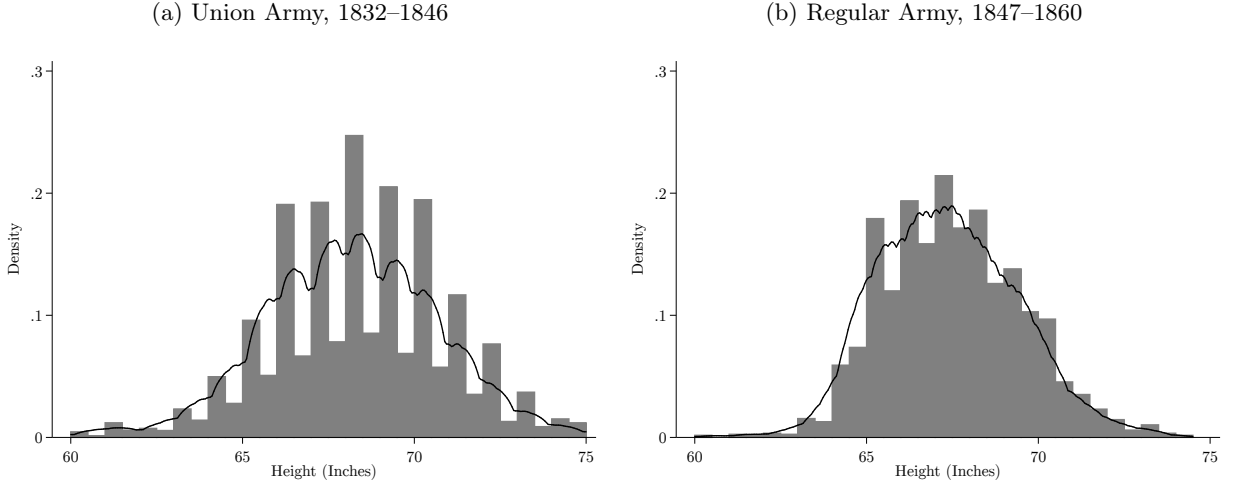
(a) Union Army, 1832–1846   (b) Regular Army, 1847–1860

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).

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, and under-represented in the Union Army. 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.

18

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 π 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]

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 so, 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

---

[23]Leaving the form of $\Omega(\cdot)$ free in equation (4) rather than assuming joint normality of ε_{it} and u_{it} implies that γ_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.

[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.

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]

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 β_k and δ_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

---

[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.

[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 (1,000)	−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 β and δ 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.

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.

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 γ_t, which (as the analysis below will reveal) are affected.

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

---

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

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 (1,000)	−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.

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 to 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.

6.3 Adjusted Trends in Height

I present the results of incorporating the correction for selection on both observables and unobservables in Figure 2.[30] 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 current methods 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. This 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).

Two key insights can be drawn from Figure 2. 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 ($\chi_1^2 = 61.66$, $p < 0.01$).

---

[30]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.
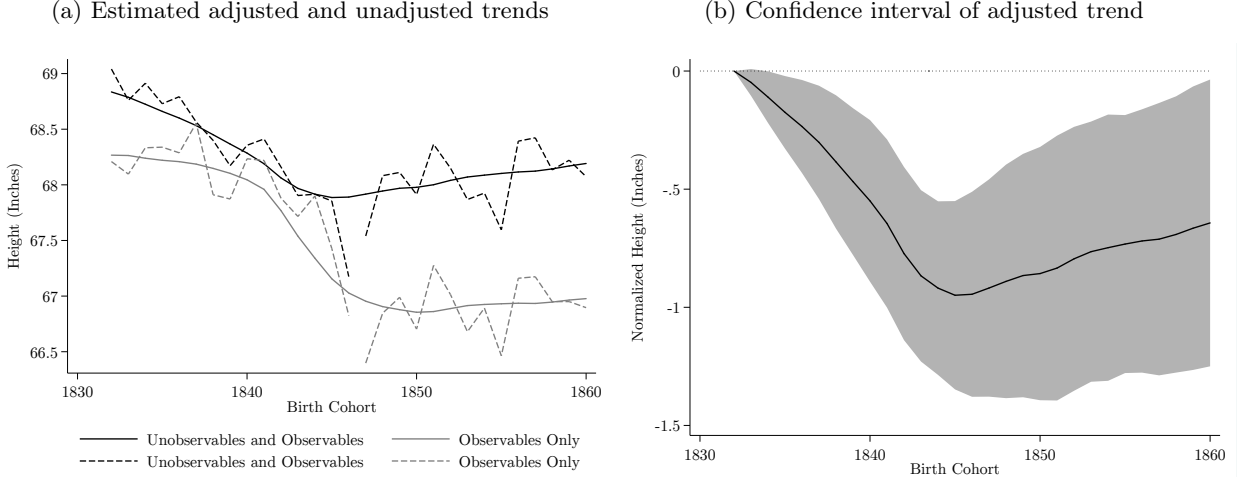
(a) Estimated adjusted and unadjusted trends

(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)).

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 ($\chi_1^2 = 4.43$, $p = 0.04$). 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 current literature's techniques (i.e., adjusting only for truncation and selection on observables).[31] In particular, when adjusting only for selection on observables, a decline of about 1.24 inches in the

---

[31]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

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) ($\chi_1^2 = 5.50$, $p = 0.02$). 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.[32] This is supported in part by the fact that the decline in average stature for the birth cohorts 1832–1846 is not statistically different between the two smoothed trends ($\chi_1^2 = 2.00$, $p = 0.16$), 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.[33] 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 that indicate that the postbellum Regular Army was likely to be composed of more negatively selected enlisters than the Union Army.

---

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).

[32] 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.

[33] 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).

6.4 Cross-Sectional Patterns

Historical anthropometric studies of the United States suggest that the Northeast was the region with the shortest average stature 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, another possibility is 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 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 (i.e., population living in places of at least 2,500 inhabitants) and a rural county as one with no urban population.[34] 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

---

[34]By this definition, there are 393 urban counties and 1,013 rural counties represented in the data.

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.

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 address the 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 ask 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 wherever 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

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*** (0.170)	67.234*** (0.247)	
South	−0.492* (0.258)	0.018 (0.284)	67.215*** (0.314)
Panel B: Unobservables and Observables			
Northeast	67.915*** (0.361)		
Midwest	−0.314** (0.144)	68.229*** (0.362)	
South	−0.304 (0.197)	0.010 (0.222)	68.219*** (0.437)
Panel C: B − A			
Northeast	1.192*** (0.370)		
Midwest	0.197* (0.115)	0.995** (0.390)	
South	0.188 (0.142)	−0.008 (0.147)	1.003*** (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*** (0.150)	67.307*** (0.255)
Panel B: Unobservables and Observables		
Urban	67.958*** (0.359)	
Rural	−0.287** (0.122)	68.245*** (0.378)
Panel C: B − A		
Urban	1.192*** (0.372)	
Rural	0.254*** (0.082)	0.938** (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.

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.[35] 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 exert a greater influence.

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.

---

[35]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.

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.

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.

Komlos, John (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.

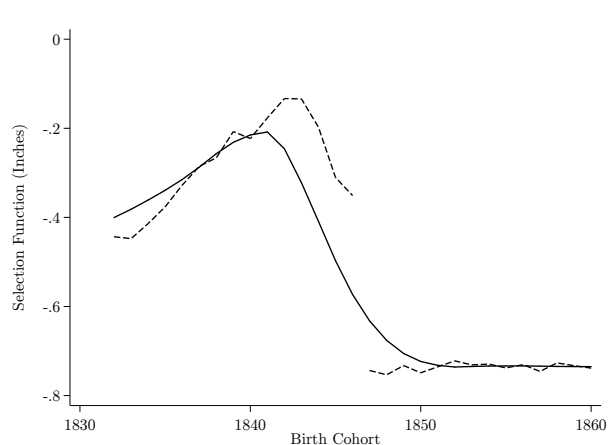# A Estimated $\Omega(\cdot)$ Function by Birth Cohort

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).