# Intuition to Recognize and Address Sample-Selection Bias in Historical Sources

With Illustrations from the Historical Heights Literature

Ariell Zimran
ariell.zimran@vanderbilt.edu

Vanderbilt University

December 2017

Prepared for Consideration for the *Social Science History* Special Issue on Selection Bias
Click Here for Most Recent Version

**Abstract**

The idiosyncratic availability and survival of historical records often forces quantitative social science historians to rely on sources describing only a subset of the population that they wish to study. Whether and how such sources can be used to study this population is a key concern. In this paper, I illustrate the intuition of an approach that enables researchers to recognize the presence of sample-selection bias—a systematic difference between the sample and the population of interest that can render conclusions drawn from the sample uninformative about the population of interest—and to draw conclusions about the broader population without ignoring this potential pitfall. A key part of this discussion concerns the role of a variable that affects whether an individual is observed but not the outcome of interest. I focus specifically on differences in average stature between the Northeast and the Midwest in the antebellum United States as observed in historical military data. Using voting patterns in the Presidential Election of 1860 as this variable, I show what simple patterns in the military data can indicate the presence of sample-selection bias, and outline how the variable enables the researcher to understand the potential impact of this bias.

**Acknowledgements** I am grateful to Ran Abramitzky and Richard Steckel, whose advice on another paper led to the creation of the exercises in this one; to William Collins and Marlous van Waijenburg for helpful discussions; and to Kris Inwood, Hamish Maxwell-Stewart, and Ewout Depauw for organizing this special issue. Thanks are also due to Timothy Cuff for sharing his data on Pennsylvania recruits to the Union Army. This project, by virtue of its use of the Union Army Project data, was supported by Award Number P01 AG10120 from the National Institute on Aging. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institute on Aging or the National Institute of Health. All errors are my own.

# 1 Introduction

To social science historians, the value of studying history is self evident. Social scientists with a modern focus, however, must sometimes be persuaded that important insights of modern interest can be gained by studying the past. In describing to economists the contribution of the study of history to advancing knowledge, economic historians tend to give pride of place to the advantage of working with historical data (e.g., Abramitzky, 2015; McCloskey, 1976). By enabling the study of complete life spans and lineages, of economic development in the long term, and of unique institutions, among other benefits, historical data have enriched our understanding of economics and opened new frontiers of knowledge in this and other social sciences. Combined with the advent of new digitization and record linkage technologies that have reduced the cost of collecting and analyzing historical data (Collins, 2015), these benefits have helped to generate a surge in research intended for an audience of general economists that uses historical data (Abramitzky, 2015, Figure 1, p. 1243).

But historical data are not a panacea. Like any other data source, they carry a potential for unique errors and biases. One challenge is that researchers are limited to the records that historical circumstances caused to be collected and that the filter of time has permitted to survive to the present. A key danger is therefore that the conclusions drawn from historical data are generated by the availability of data as much as by the content thereof. This is a sample-selection problem: if data are available only for a non-random subset of the population of interest, and if the subset is systematically different from the population, then the conclusions drawn from studying the subset alone might not generalize to the whole population. The error that results from drawing conclusions for a population from an unrepresentative sample is called *sample-selection bias*. Bias of this type can arise under two circumstances. First, it occurs if the sample and the population of interest differ in terms of observable characteristics related to the outcome of interest (*selection on observables*). For example, a sample might over-represent certain occupations or regions. Such bias can also occur even if the sample and the population are identical in their observables, if characteristics unobservable to the researcher affect both entrance into the sample and the outcome of interest (*selection on unobservables*), making the sample unrepresentative of the population. For instance, a sample might over-represent more productive laborers, but the researcher would only be aware of having sampled laborers generally.[1] Bias generated by this channel is much more challenging to recognize and address.

Recent scholarship in economic history has ignited a debate over the extent to which sample-selection bias has led scholars to draw spurious conclusions and over how to best work with data that are suspected to

[1] This point is clearly made by Kosack and Ward (2014).

suffer from sample-selection bias (Bodenhorn, Guinnane, and Mroz, 2017). The debate is salient throughout the field: selection issues are present in many contexts because historical data were often recorded for purposes other than those for which modern researchers use them, and therefore (either unintentionally or by design) encompass parts of the population that were systematically different from the populations as a whole. Despite this wide applicability, however, the debate has centered on the historical heights literature, which takes advantage of the wealth of height data recorded for identification purposes in historical settings to study secular patterns in health, of which average height is a standard measure. These data form the basis for the important finding that health declined during early industrialization in the United States and England. The data collected for this purpose, however, are largely drawn from the records of prisoners and military volunteers. Naturally, there is reason to believe that prisoners and soldiers were not randomly drawn from the population, raising the question of whether the patterns in stature among these groups are informative of those in the complete population of interest. This literature has addressed the danger posed by selection on observables; but the recent debate has focused on whether changing unobservable characteristics of soldiers and prisoners induced by changing incentives and alternatives for military service or crime might have generated sample-selection bias.

One extreme position taken in this debate is that, because sample-selection bias might generate spurious conclusions, it is almost impossible to learn from data sources unless they are known to be representative of the population of interest. Such a restriction on the scope of scholars' analysis, by ignoring precious data, would render impossible the study of many important questions and contexts. But in many cases, the challenges posed to the use of historical data by such selection issues are not insoluble, and if appropriate precautions are taken, it is still possible to learn from data suspected to be unrepresentative without ignoring potential pitfalls. In the historical heights literature, Zimran (2017) takes this position and demonstrates a solution to the problem of sample-selection bias generated by selection on unobservables. By studying the determinants of military enlistment, he quantifies the sample-selection bias in a sample of historical stature data drawn from military records and formally tests whether existing conclusions of the historical heights literature can be purely artifacts of such bias. He finds evidence that sample-selection bias can affect the conclusions drawn from historical data, but that the historical heights literature's key findings are not artifacts of sample-selection bias. While the data and application studied by Zimran (2017) are specific to the historical heights literature, his analysis encompasses an approach that can be used to address and quantify bias induced by selection on unobservables in historical data from a variety of sources and with a variety of outcomes of interest.

Zimran's (2017) method is a simple extension of a standard statistical method employed by economists when dealing with sample-selection bias (e.g., Heckman, 1979; Vella, 1998). Given the existence of a number of textbook treatments of this approach, as well as Zimran's (2017) description of his application, I do not provide a guide on estimation in the presence of sample-selection concerns. Instead, in this paper, I provide a simple discussion of the intuition of this method without dwelling on its technical aspects. A key part of this discussion focuses on the role of a variable that affects only whether an individual enters the sample, but which has no impact on the outcome of interest.

I begin by providing additional background information on the historical heights literature (and the greater standards-of-living debate of which it is a part) and on approaches to selection problems in economics in section 2. I then lay out several examples in section 3 that illustrate the most basic intuition of how selection on observables and selection on unobservables can generate sample-selection bias, and how the researcher can use a combination of data from the available sample (including information on the outcome of interest) and from the population as a whole (giving information on some characteristics for the whole population, but not for the outcome of interest) to recognize this bias,[2] to understand how it might affect conclusions drawn from the data, and to draw conclusions from potentially problematic sources without ignoring the pitfalls of sample-selection bias. Four pieces of intuition are developed. Intuition 1 is that sample-selection bias generated by selection on observables occurs whenever observable characteristics affecting the outcome are over or under represented in the sample relative to the population—that is, when these characteristics also affect the likelihood of entering the sample; but bias generated in this way is simple to correct (and such corrections are routinely performed). Intuition 2 is that sample-selection bias generated by selection on unobservables can be detected when a variable that affects whether an individual enters the sample, but is known to have no effect on the outcome in the population, has a relationship with the outcome in the sample; or, more generally, when the probability of individuals to be observed is related to their outcomes in the sample, controlling for observable characteristics. Intuition 3 is that if individuals whose observable characteristics make them more likely to be observed in the sample have a greater (smaller) value of the outcome, then the sample-selection bias generated by selection on unobservables is negative (positive). Intuition 4 is that the magnitude of the sample-selection bias generated by selection on unobservables is decreasing as individuals are more likely to enter the sample, and if individuals can be identified whose observable characteristics make them almost certain to be observed, their data can be used to learn about the outcome of interest without concern over sample-selection bias.

[2] If the outcome of interest were observed in the population as a whole, there would be no need to rely on the selected sample, and therefore no problem of sample-selection bias.

I then move to my empirical analysis. In section 4, I introduce my data source—the Union Army data (collected by Cuff, 2005; Fogel et al., 2000), which is one of the chief sources in the historical heights literature—and provide summary statistics. In section 5, I apply the intuition developed in section 3 to the study of the difference in average stature between the Northeast and the Midwest, which has been cited as being possibly affected by sample-selection bias (Bodenhorn, Guinnane, and Mroz, 2017; Mokyr and Ó Gráda, 1996; Zimran, 2017). I compare individuals observed in the Union Army to the population in the 1860 census to estimate the determinants of military enlistment in the Union Army. By showing that region and occupation were related both to height and to the probability of military enlistment, I show the presence of selection on observables in the data. I also include, in my estimation of the determinants of military enlistment, voting patterns at the county level in the Presidential Elections of 1860 under the assumption that they have no effect on height in the population. Finding that this variable is positively related to the probability of enlistment and to height in this sample, and that enlistment probability is positively related to height after conditioning on other observables, I conclude that sample-selection bias from selection on unobservables was present and negative. Finally, I show that the greater probability of enlistment in the Midwest than in the Northeast is consistent with part of the Midwest's observable height advantage being the result of sample-selection bias, but that the presence of a Midwestern height premium at all levels of enlistment probability implies that it is not an artifact of sample-selection bias.

Although this paper focuses on the historical heights literature, the intuition that it provides can be applied in a variety of settings, as long as enough information is available to characterize the process by which individuals come to be observed. In section 6, I conclude by discussing general strategies for application of this intuition to contexts outside of the historical heights literature in the antebellum United States.

## 2 Background

Bodenhorn, Guinnane, and Mroz's (2017) discussion of sample-selection bias, though methodologically relevant to a broad range of subjects in quantitative social science history, is empirically situated in the "standards-of-living debate," which focuses on the welfare effects of early industrialization and modern economic growth, primarily for the working classes. The primary focus of this debate is England during the Industrial Revolution, although considerable attention has also been devoted to the United States and other Western European countries in the nineteenth century. Contributors to this debate have divided into three camps. "Optimists" argue that the bulk of the population experienced an appreciable short- to medium-term increase in living standards. "Pessimists" argue that living standards deteriorated ("strong pessimists"), or at best improved only slowly ("weak pessimists") during this period.

Four classes of indicators are generally called upon to assess living standards in this debate. The first three classes are conventional measures of the economic standard of living—GDP per capita, real wages, and consumption. In the English case, these indicators tend to support the weak pessimist perspective (Brown, 1990; Clark, Huberman, and Lindert, 1995; Crafts, 1985, 1997; Crafts and Harley, 1992; Feinstein, 1998; Mokyr, 1988; Schwarz, 1985; Voth, 2003): although industrialization was not a boon to the English working classes in the short- to medium-term, it did not make them worse off. The situation in the United States, on the other hand, was more positive, with rapid improvements evident in these conventional indicators of welfare indicating rapid welfare gains (Costa and Steckel, 1997; Goldin and Margo, 1992; Margo and Villaflor, 1987).

The fourth class of indicators—measures of health—are a boon to the strong pessimist case. Both in the United States and in the United Kingdom, the "historical heights literature" indicates that the biological standard of living as measured by average stature declined in the mid-nineteenth century (A'Hearn, 1998; Costa and Steckel, 1997; Floud et al., 2011; Floud, Wachter, and Gregory, 1990; Fogel, 1986; Fogel et al., 1983; Komlos, 1987, 1992; Margo and Steckel, 1983; Mokyr and Ó Gráda, 1996; Voth and Leunig, 1996; Zimran, 2017), indicating deterioration of the biological standard of living. While the decline in the United Kingdom was slight, a precipitous decline in the United States contrasts with the strong improvements in the Economic Standard of Living. In the United States, declining life expectancy in the nineteenth century provides additional supporting evidence (Fogel, 1986; Pope, 1992).

This divergence between the economic and biological measures of living standards is known as the "Industrialization Puzzle" in the English context, as the "Antebellum Puzzle" in the American context, and generally as the "Industrial Growth Puzzle" to emphasize that scholars who viewed rapid economic growth as unambiguously welfare-improving were surprised to observe declining health during early industrialization (Haines, 2004). In the American context this pattern is illustrated in Figure 1. These puzzles also have analogs in the cross section. In the United Kingdom, Mokyr and Ó Gráda (1996) show that the Irish, despite being poorer than the English, held a height advantage over the latter. Similarly, in the United States, the poorer Midwest is generally found to be taller than the Northeast (Haines, Craig, and Weiss, 2003; Margo and Steckel, 1983; Zimran, 2017) despite higher per capita income in the latter. These patterns also raise the question of why better economic standards of living did not translate to greater health in the early 19th century.


Figure 1: The Antebellum Puzzle

*Note:* GDP data from Bolt and van Zanden (2013) are smoothed and presented in log scale. Height data are of native-born white males and are based on birth cohorts from data presented by Costa and Steckel (1997), Craig (2015), Floud et al. (2011), and Zehetmayer (2011).

Despite the wealth of evidence establishing their existence, these puzzles have always been contentious. Because the trend in health constitutes the only evidence in favor of the pessimist perspective on the welfare effects of early economic growth in the United States, scholars skeptical of this pessimist perspecgive have criticized the research establishing its existence. Although some criticism focuses on how to interpret these conflicting patterns in economic and biological indicators, the greatest criticism has focused on the sources from which the height data were drawn. The main concern is that, even though the historical heights literature has always recognized the danger posed by selection on observables and has taken steps to address it, sample-selection bias generated by selection on unobservables has not been properly addressed and may have created these patterns spuriously (Bodenhorn, Guinnane, and Mroz, 2014, 2017; Gallman, 1996; Mokyr and Ó Gráda, 1996). In the time series, the crux of the argument formalized by Bodenhorn, Guinnane, and Mroz (2014, 2017) is that the economic growth of the nineteenth century made civilian labor market options more attractive over time. The alternative to civilian employment, military enlistment, thus became less attractive, leading to a change in the composition of the part of the population for which height is observed. The military was generally composed of individuals whose unobservable characteristics made them shorter than the population—for instance, because unobservable health shocks in childhood would reduce terminal height and harm future labor market outcomes, making the military alternative more attractive—and the growing attractiveness of the civilian labor market may therefore have caused this negative selection to become more severe. Such a pattern could generate a decline in observed average stature even if population average stature did not decline. In the cross section the argument is that Northeasterners may have had better opportunities in the labor market than Midwesterners, leaving only the shortest Northeasterners to enlist, while the Midwest appeared taller because worse opportunities outside of the military that led a more representative portion of the population to enlist. It is argued that these forces led to an exaggeration of any height premium of the Midwest over the Northeast, or even its spurious creation.

The potential patterns of sample-selection bias generated by selection on unobservables threaten the conclusions of a large and important literature. As a result, a variety of empirical tests have been performed to determine whether a given data source is affected by sample-selection bias induced by selection on unobservables. Among these, two are notable. Bodenhorn, Guinnane, and Mroz (2017) propose a test based on the logic that if the composition of military enlisters responded to changes in the state of the economy over the life time, then long-term improvements in living standards must have affected the composition of the height data over birth cohorts, and thus inference from such data. This test, however, is unable to detect the presence of all sample-selection bias and can also incorrectly indicate the presence of such bias when it is not actually present. It also does not enable the quantification of or correction for sample-selection bias if it is detected. Bodenhorn, Guinnane, and Mroz's (2017) application of this test suggests the presence of sample-selection bias in a number of historical height sources, leading them to take the view that "we cannot draw firm conclusions about long-run trends in US or British heights" (p. 201) from these sources.

Zimran (2017), on the other hand, addresses concerns of sample-selection bias in historical heights with an approach that formally tests for the presence of sample-selection bias and quantifies it providing a correction for it where it exists. The different scope of Zimran's (2017) approach is enabled by his study of the process by which individuals came to enter the sample through the comparison of the sample to the complete population of interest on the basis of observable characteristics, rather than Bodenhorn, Guinnane, and Mroz's (2017) restriction of attention to the sample only. Using voting patterns in the Presidential Election of 1860 to isolate variation in military enlistment probability that is unrelated to height in the population, and exploiting changes in the incentives for military enlistment with the end of the Civil War, he finds that the data used to establish the Antebellum Puzzle did in fact suffer from sample-selection bias that varied over time and geography. But he also finds that the magnitude of the bias was not sufficient to be solely responsible for the puzzling patterns discovered. This test, unlike that of Bodenhorn, Guinnane, and Mroz (2017), is applicable beyond concerns over changing sample-selection bias in historical heights.

Zimran's (2017) test is an application of formal selection correction methods to the historical heights literature. Although this application is novel, these methods are not themselves novel. The logic and mechanism of the correction date to Heckman (1979). The canonical application is to the study of the determinants of women's wages—a setting in which it is important to study the participation decision as a first stage. A number of survey pieces review this literature in depth (e.g., Vella, 1998), and many of the advances in the literature on selection models have focused on relaxing the assumptions required to perform the correction. These assumptions are largely technical. In applications of this method, the focus is on the exclusion restriction—the variable that affects whether or not an individual enters the sample in which outcomes are observed but does not affect the outcome of interest. It is this variable that allows the researcher to make a distinction between changes in sample-selection bias and actual changes in the outcome.

## 3 Intuition

The intuition of how selection on observables and selection on unobservables can create sample-selection bias, and of how this bias can be recognized and addressed, can be illustrated with some simple examples. To fit the motivation of this paper, the empirical setting of sections 4 and 5, and the focus of the bulk of the debate over sample-selection bias in economic history, I situate these examples in the context of trying to determine the true difference in unconditional average stature between Midwesterners and Northeasterners from a sample of military data.[3]

The most basic intuition of the sample-selection problem is that, absent information on how the sample was formed, it is impossible to determine whether any difference in average stature between regions as observed in the sample reflects a true difference in the heights of the populations of each region, sample-selection bias induced by differences in selection into the sample across regions, or some combination of these two forces. That is, absent information about how individuals came to enter the military, it is impossible to make conclusions regarding population average stature from the average stature of the sample. In some countries, this challenge is overcome by conscription: if everyone (or a randomly selected group) were required to serve in the military (or at least to appear to be examined and have their heights recorded, as in Italy), then observed heights can be taken as representative of those of the population.

If, on the other hand, military enlistment is the product of individual choice (as in Britain and the United States in the nineteenth century), then the translation of the average stature observed in the sample to the average stature of the population is less straightforward. The first example exhibits this. Suppose that each region is divided into an urban and a rural sector, and that ruralists are, on average, taller than urbanites. Suppose further that there is no difference in the average heights of individuals of the same sector across

[3]The unconditional difference in average stature is the difference between the average height of all Northeasterners and the average height of all Midwesterners, not taking into account differences in urbanization, occupation, or any other characteristics.

regions, and that the only other determinant of height is genetic variation which is the same in each region-sector and averages away in random samples. Finally, suppose that the fraction of the population that is rural is greater in the Midwest than in the Northeast, implying greater average stature in the Midwest than in the Northeast. Panel A of Table 1 presents an example of average heights satisfying these restrictions. But these are the true average heights: what the researcher wishes to learn but does not observe.

Table 1: Example height distributions

| | Average Heights | | Fractions | | Average Height |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| *Region* | Urban | Rural | Urban | Rural | All |
| *Panel A: Population (Actual)* | | | | | |
| Northeast | 67.00 | 69.00 | 0.75 | 0.25 | 67.50 |
| Midwest | 67.00 | 69.00 | 0.25 | 0.75 | 68.50 |
| *Panel B: Military (Observed)* | | | | | |
| Northeast | 67.00 | 69.00 | 0.90 | 0.10 | 67.20 |
| Midwest | 67.00 | 69.00 | 0.50 | 0.50 | 68.00 |

*Notes*: Panel A describes the population of interest. Columns (1) and (2) describe the average heights of each region-sector, and columns (3) and (4) describe the distribution of each region's population across these sectors (so that each row sums to one). Column (5) of Panel A shows the true average height of each region, and thus the true difference in average heights between regions. Columns (3) and (4) of Panel A are observed, but the other columns are not. Panel B describes the observed population—the military enlisters. The contents of all five columns are observed, but because the greater tendency of urbanites to enlist causes columns (3) and (4) to differ from Panel A, the observed average height of each region and the difference between them does not match the true difference in Panel A.

Consider first the case in which only urbanites enlist in the military. In this case, the observed heights of Northeasterners and Midwesterners would be the same, despite the true Midwestern premium. A less extreme case allows both urbanites and ruralists to enlist, but retains the greater tendency for urbanites to enlist relative to ruralists. As the observed data would over-represent urbanites relative to the population, the average stature of each region as observed in the enlistments would differ from the true stature, and the regional differences in the sample would not reflect regional differences in the population. Panel B of Table 1 depicts such an example, and shows that these patterns cause the observed difference in the heights of Midwesterners and Northeasterners (0.80 inches in Panel B) to differ from the actual difference (1.00 inch in Panel A).[4] If the urban status of enlisters is observed, and if the fraction of each region that is

[4]The contents of Table 1 can be generated by a model of the form

$$h_i = 67.00 + 2.00R_i + 0.00N_i + \varepsilon_i$$
$$P(y_i = 1|R_i, N_i) = 0.75 + 0.00N_i - 0.50R_i,$$

where $h_i$ denotes the height of individual $i$, $R_i$ is an indicator equal to one if the individual lives in the rural sector, $N_i$ is an indicator equal to one if the individual lives in the Northeast, $y_i$ is an indicator equal to one if an individual enters the military, and $\varepsilon_i$ is a mean-zero stochastic error term that is uncorrelated with $R_i$ and $N_i$ (and thus is unrelated to military enlistment). Height is observed only if $y_i = 1$. The $N_i$ is included in the model despite its zero coefficients in both equations to emphasize that the researcher is trying to learn the difference between the average height of each region, and so must allow for region-specific differences in height and enlistment probability.

urban is also known, then this is an example of *selection on observables* because the variable driving the non-representativeness of the data (sector) is observed in both the sample and in the population. Once this variable is controlled for (by looking at heights within each region-sector), however, the sample is random and observed heights represent actual heights; the only problem is that the fractions of each sector in the sample differ from those in the population. But the population fractions are known, making possible computation of true average stature by simply combining observed stature for each region-sector with its population fraction. That is, in Table 1, the researcher can compute population average heights using Panel B's height data in columns (1) and (2), and Panel A's fractions of the population in columns (3) and (4). The same logic applies when more than one variable affects selection or if the variable or variables affecting selection are continuous. If all variables affecting both height and entrance into the sample are observed, then selection conditional on these variables is random and the average stature of individuals with any given set of observables is known. Again, information on the distribution of these observables in the population enables the calculation of true average heights by providing the correct weights.

The first piece of intuition to carry forward is generated by this example.

**Intuition 1.** Recognizing selection on observables is straightforward. It occurs whenever an observable characteristic that affects the outcome of interest is over- or under-represented in the observed sample relative to the population of interest (that is, whenever an observable characteristic that affects the outcome also affects entrance into the sample). The correction for the bias that it induces is also straightforward. Because selection is random conditional on observables, the average height for every combination of observables is known. The researcher must simply compute a weighted average of these observed outcomes using population weights.

Given the relative ease by which bias induced by sample selection of this type can be recognized and addressed, most analyses of historical data using potentially unrepresentative routinely check for it and address it, either by studying the determinants of entry into the sample, or by including observable controls in regression analysis.

If, on the other hand, enlisters' sector were not observed, this would be a case of *selection on unobservables* because an unobserved factor (in this case, sector) affects both height and military enlistment.[5] Sample-selection bias induced by selection on unobservables can be present even if the sample and the population are identical on all observable characteristics; it is thus more difficult to recognize and address and sometimes it

[5]In Table 1, if sector is not observed, or if the fraction of the population that is urban is not known, then it is impossible to compute the true average height of each region. All that would be observed would be the contents of column (5) of Panel B, which is not enough information to compute the true average height of each region.

is impossible to do so. This bias can be better illustrated with another example. For this example, remove the urban-rural distinction so that all individuals are in the same sector and the distribution of heights is the same in each region. Suppose instead that individuals differ in their wages, and that only those with wages below a particular threshold enlist in the military. Suppose also that lower wages imply lower stature, that wages are higher in the Northeast, and that the relationship between height and wages is the same in each region once differences between regions in wages are accounted for.[6]

Figure 2 illustrates this example. Higher wages in the Northeast are evident from the rightward shift of its wage-height relationship relative to that of the Midwest. The same distribution of heights in each region is illustrated by the same range of each line (and an implicit assumption of a uniform distribution along the line). Three implications of this graph are notable. First, the relationship between wages and stature implies that only the shorter members of each region tend to join the military. Second, Midwesterners are more likely to enlist, as evidenced by the greater share of the Midwest's line that is below the cutoff for enlistment, $\bar{w}$. Finally, higher wages in the Northeast imply that there is a range of heights such that the wages are low enough to enlist in the Midwest but not in the Northeast (the range $B$–$C$). This would make the Midwest appear taller in the enlistments data—Midwesterners would have an observed average height of $\frac{1}{2}(C-A)$ while Northeasterners would have an observed average height of $\frac{1}{2}(B-A)$—even though it was assumed above that the regions are in fact of the same height. If wages are not observed, then this is a case of selection on unobservables.[7]

Because wages are not observed in the enlistments data, there is seemingly no way of knowing whether the difference in stature observed between regions is because of different incentives to enlist (the true reason) or because of differences in health between regions (as the literature on historical heights has usually interpreted the results). All that the researcher observes is that Northeasterners in the data are shorter on average than

[6]This example can be generated by a model of the form

$$\begin{aligned} h_i &= \beta_0 + \beta_1 N_i + \varepsilon_i \\ w_i &= \alpha_0 + \alpha_1 N_i + u_i \\ y_i &= \mathbb{1}\{w_i < \bar{w}\}, \end{aligned}$$

where $\mathbb{1}\{\cdot\}$ is the indicator function, $h_i$ is the height of individual $i$, $N_i$ is an indicator equal to one for Northeasterners, $w_i$ is the wage of individual $i$, $y_i$ is an indicator equal to one for military enlisters, $\varepsilon_i$ and $u_i$ have means of zero, are uncorrelated with $N_i$, and $\text{corr}(\varepsilon_i, u_i) = 1$. Height is observed only if $y_i = 1$. The perfect correlation between the errors is helpful for illustration, but is not necessary. If there is a non-zero positive correlation between them, the lines in Figure 2 can be taken to represent regression lines, and the intuition is the same. The equivalence of average heights in the two regions implies that $\beta_1 = 0$, but it is included to emphasize that the researcher is looking to learn this difference and cannot do so if region is excluded from the height equation.

[7]If wages were observed in both the population and the sample, then this would be another example of selection on observables. Since there are a range of wages whose heights are not observed (because they do not enlist), it is necessary to exploit the linear structure of the model to learn the true average heights. In a more realistic case in which the lines of Figure 2 represent regression lines rather than true data (in the language of footnote 6, if the correlation of $\varepsilon_i$ and $u_i$ is positive but not equal to one), then this is simply a generalization of the example in Table 1 with a continuum of values of the observables.

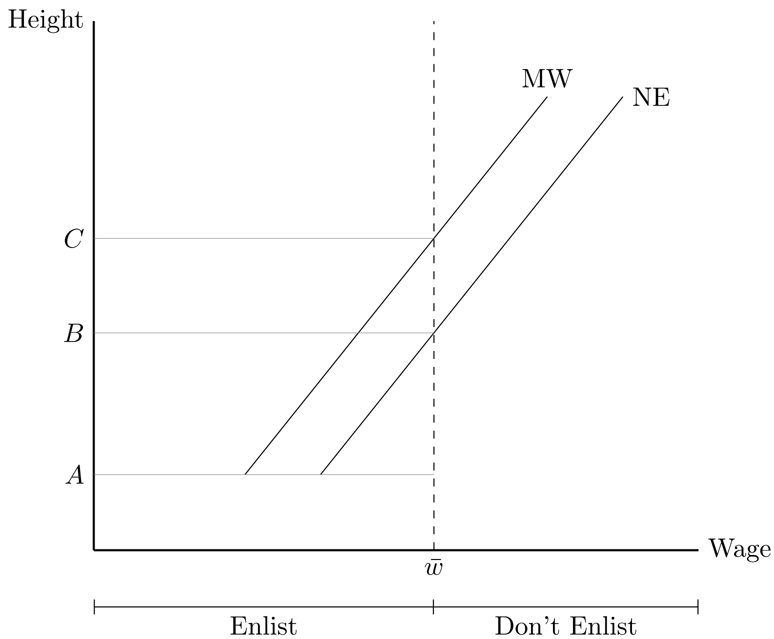



Figure 2: Hypothetical relationship of wages, heights, and military enlistment

Midwesterners in the data. Indeed, in the absence of further information the selection problem cannot be solved—any height difference observed between regions can be explained equally well by differences in true heights or by differences in selection between the regions.

Can the selection on unobservables problem be solved with additional information? If the additional information satisfies certain conditions, the answer is yes. Continuing the example depicted in Figure 2, suppose that the population is divided between hawks and doves, that hawk-dove status in the population and in the military is observable, and that the division between hawks and doves is independent of height and wage so that the distribution of heights and wages in each region is the same between hawks and doves. Finally, suppose that the threshold wage for hawks' enlistment is higher than that of doves.[8]

The utility of the hawk-dove division in solving this problem stems from the following insight, illustrated

[8]This is a simplification of the idea that patriotism played a role in driving military enlistment in the Civil War (Komlos and A'Hearn, 2016; Zimran, 2017). It can be represented by a model of the form

$$\begin{aligned} h_i &= \beta_0 + \beta_1 N_i + \varepsilon_i \\ w_i &= \alpha_0 + \alpha_1 N_i + (\bar{w}^D - \bar{w}^H) H_i + u_i \\ y_i &= \mathbb{1}\{w_i < \bar{w}^D\}, \end{aligned}$$

where $\mathbb{1}\{\cdot\}$ is the indicator function, $h_i$ is the height of individual $i$, $N_i$ is an indicator equal to one for Northeasterners, $w_i$ is the wage of individual $i$, $y_i$ is an indicator equal to one for military enlisters, $H_i$ is an indicator equal to one for hawks, $\varepsilon_i$ and $u_i$ have means of zero, are uncorrelated with $N_i$ and $H_i$, and $\text{corr}(\varepsilon_i, u_i) = 1$. The perfect correlation between these errors is helpful for illustration, but is not necessary. If there is a non-zero positive correlation between them, the lines in Figure 3 can be taken to represent regression lines, and the intuition is the same. The equivalence of average heights in the two regions implies that $\beta_1 = 0$, but it is included to emphasize that the researcher is looking to learn this difference and cannot do so if region is excluded from the height equation.

in Figure 3. While doves enlist only if their wages are below $\bar{w}^D$, hawks with wages in the range $[\bar{w}^D, \bar{w}^H]$ also enlist (as well as hawks with wages below $\bar{w}^D$). This implies that hawks have a higher probability of enlistment. It also implies that observed hawks (in the military) would be taller than observed doves (again, in the military) in each region despite there being no relationship of hawk-dove status with height in the population: hawks in the Northeast who are observed in the military include individuals of heights $A$ to $D$, while observed doves in that region include only those of heights $A$ to $B$; similarly, observed hawks in the Midwest include individuals of heights $A$ to $E$, while observed doves in that region include only those of heights $A$ to $C$. That is, hawk-dove status has no relationship to height in the population; but because it affects the military enlistment decision, bringing individuals with wages between $\bar{w}^D$ and $\bar{w}^H$ into the sample, the observed heights of hawks are greater than those of doves.

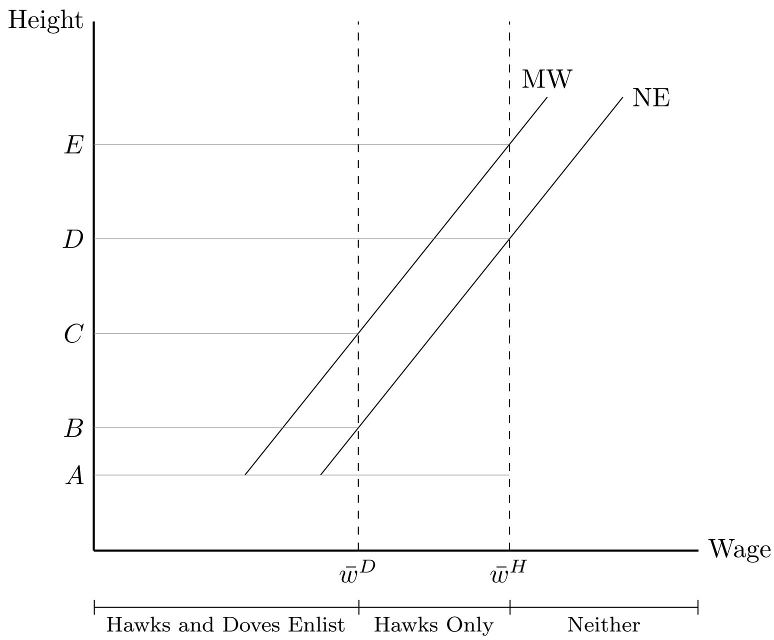



Figure 3: Hypothetical relationship of wages, heights, and military enlistment with hawks and doves

This observed difference in height between hawks and doves generates the second piece of intuition.

**Intuition 2.** Consider a variable that affects selection into the sample but is unrelated to outcome of interest in the population. If this variable is related to the outcome in the sample, then there is selection on unobservables. In a multivariate setting, if, after conditioning on variables affecting both the outcome and the probability of entering the sample, this probability is related to the outcome, then there is selection on unobservables.

Four clarifications of this intuition are important. First, in a multivariate setting, it is the correlation between the probability of being in the sample and the outcome that is important. Correlations of individual variables that affect this probability but not the outcome in the population, with the outcome in the sample are suggestive, but finding or failing to find such a correlation is not definitive evidence when there are other drivers of the probability of being observed. Second, in the multivariate setting, it is crucial that selection on unobservables is present if there is a relationship between height and enlistment probability after controlling for other observables; an unconditional relationship is not informative, as the failure to control for observables affecting height and enlistment probability would trivially create a relationship.[9] This discussion also shows why it is not possible to use the differences in enlistment probability between regions to find sample-selection bias, and why hawk-dove status is necessary. From a technical perspective, this is because there would be no variation in enlistment probability or height to compare after conditioning on this observable. Intuitively, it is because there is no way to know what difference observed between the Northeasterners and Midwesterners is due to the different probability of enlistment and what is due to true differences in height. Third, this example makes clear why hawk-dove status must be unrelated to height in the population. If the two were related in the population, then the hawks' height premium in the sample could reflect the role of ideology in driving enlistment and thus selection on unobservables, but could also reflect an actual height premium for hawks in the population. The available data would make it impossible to distinguish between these two explanations. Finally, it is tempting to conclude from this discussion that it is possible to detect selection in the selected sample alone; indeed, the correlation between height and hawk-dove status is from the selected sample only. While such an analysis can provide suggestive evidence (and is certainly better than nothing), data on the population at risk for enlistment are needed to establish that hawk-dove status actually affects military enlistment. In a multivariate setting, these data are needed to compute each individual's enlistment probability.

The nature of the relationship of height with hawk-dove status in the sample reveals the nature of the bias induced by selection on unobservables (i.e., whether it is positive or negative). As illustrated in Figure 3, military enlisters are negatively selected on their unobservables—that is, only the shortest enlist. But the fact that hawks are more likely to enlist than are doves, and that hawks are taller than are doves within each region, indicates that this selection must be negative. Essentially, hawks draw in a greater fraction of their respective populations to enlistment, and the fact that doing so brings in taller individuals implies that enlistment must be from the bottom of the height distribution. Had hawks instead been observed to

[9]For instance, if farmers are taller and less likely to enlist, then there would be an unconditional relationship, but it would not be indicative of selection on unobservables.

be shorter than doves, that would indicate that selection into military service was positive.[10] This logic provides the third piece of intuition.

**Intuition 3.** If, after conditioning on variables affecting both the outcome and the probability of entering the sample, individuals whose observable characteristics make them more likely to be observed (in the example, hawks) are taller, then selection on unobservables is negative. If they are shorter, then selection on unobservables is positive.

This intuition forms the basis of Heckman's (1979) formal correction: the precise nature of the relationship of the outcome and the probability of entering the sample reveals the relationship between the probability of entering the sample and the sample-selection bias. Indeed, it is this variation that would enable the formal correction to find, in this example, that there is no actual difference in average heights between regions.

Finally, suppose that there is a third group ("zealots") that enlists regardless of its wage, but continue to assume that the membership in the hawks, doves, or zealots is observed and unrelated to height. In this case, it is possible to learn the true heights of each region simply from the zealots. More generally, the bias is decreasing as the probability of entering the military increases from doves to hawks to zealots (where there is no selection on unobservables). This underlies the fourth intuition.

**Intuition 4.** The more predisposed individuals are to be observed on the basis of their observable characteristics, the less is the sample-selection bias induced by selection on unobservables among these individuals. This is intuitive from the fact that this implies that more of the population will be observed. At the limit, if there are individuals who enlist regardless of their unobservable characteristics (because their observables so strongly predispose them to enlist that their unobservables are unimportant), then there is no selection problem among these individuals.

# 4 Data

## 4.1 Sources

The main data used in this paper are collected from four sources. The first, which provides the stature data and covariates for military enlisters (that is, the sample for which height is observed), is collected from Records of the Adjutant General's Office (1861–1865). Data from this source are the products of two

[10] A key implication of selection models that allows this conclusion to be reached is that selection is from the tails of the distribution. In cases of negative selection, the shortest are more likely to enlist than are taller individuals. This assumption would be violated if the selection came from the middle of the distribution.

projects, each of which which provides data on a random sample of enlisters in the Union Army, including data on stature, age at enlistment, date of enlistment, place of birth, place of enlistment and occupation at the time of enlistment. The first is Fogel et al.'s (2000) Union Army Project, which provides information on a random sample of 16,483 enlisters. The second is the data set of Cuff (2005), which adds information on an additional 10,304 enlisters from the state of Pennsylvania. The total number of observations is thus 26,787. An obvious concern is the over-sampling of Pennsylvania, which is addressed by weighting all analyses so that the distribution of states of enlistment in the data matches that of the Union Army (Gould, 1869). I limit the data to native-born white males born and living in the Northeast and Midwest, restricting the sample to the birth cohorts of 1820 to 1846, and allowing only enlistments after age 18. Finally, because the place of enlistment will be treated as the place of residence in the analysis to follow, I also exclude individuals who enlisted in a state other than the state of their regiment.[11]

The second source provides data on the covariates of the population at risk for military enlistment. Such data are taken from the one-percent sample of the 1860 United States Census, as provided by Ruggles et al. (2015). When applying the same filtering criteria as applied to the military data, this data set provides information on 29,844 individuals. It provides information on age, place of residence, and occupation. This source also provides information on other economic measures (such as school attendance and household characteristics), but these are not used because they are not observed in the enlisting population and so it is not possible to determine their impact on the military enlistment decision.

The third source is a collection of county-level data from the Census of 1860, provided by the Manson et al. (2017, henceforth NHGIS). This source provides information on county-level agricultural and manufacturing production and capital stocks, wealth, and population density.

The final main source provides data on voting patterns in the Presidential Election of 1860. Collected from ICPSR (1999), this source provides the number of votes cast for each candidate in each county in the sample. The main variable of interest calculated in this case is the share of each county's vote cast for Abraham Lincoln, the Republican candidate. This is to fill the role of the variable that affects military enlistment but not height. Relevance to the enlistment decision is straightforward to show, as will be done below (and has been shown by Zimran, 2017). Moreover, the fact that the main issue at hand in the election was the same issue over which the Civil War was fought suggests that political attitudes should be relevant to enlistment. There is also evidence that such voting patterns are relevant to the decision of whether to enlist in the Confederate Army (Eli, Salisbury, and Shertzer, 2016), to desert from the Union Army (Costa and

[11] For instance, if an individual enlists in an Ohio regiment while it was in the field in Virginia, I do not take the place of enlistment in Virginia as the place of residence, as this might give indications of an incorrect place of residence.

Kahn, 2003), and to migrate after the Civil War (Costa and Kahn, 2007; Eli, Salisbury, and Shertzer, 2016). It is necessary to assume that this variable is unrelated to height in the population once all other observable factors are taken into account. This assumption is not testable, but as I control for all observed economic variables, a channel linking economic conditions to voting and to the enlistment decision is addressed to the extent that the data allow.

These data will be used to study two patterns. The first is the determinants of military enlistment. To perform this analysis, individuals are first linked to the characteristics of their county of residence—the vote patterns and the county data from NHGIS. For members of the 1860 census sample, the county of residence is directly observed. For members of the military sample, the county of enlistment is taken as the county of residence. This assignment is unlikely to be completely accurate, but is the best available indicator of actual place of residence without performing census linkage. A main difficulty in analyzing the military enlistment decision with these data is that the military enlistment decision is unobserved in the census data, and does not vary (by construction) in the military data. Thus, it is possible (and, given the rate of enlistment in the Union Army, very likely) that the census data include military enlisters (rather than just being a sample of non-enlisters), and the ratio of observations in the Union Army data to the number of observations in the census data is based solely on sampling decisions and not on actual enlistment probabilities. It is still possible, however, to estimate the determinants of military enlistment using a slightly different estimation procedure (Cosslett, 1981). The intuition of this approach is that the census data are informative of the distribution of characteristics in the population, whereas the Union Army data are informative of the distribution of characteristics in the military data. If these data reveal a difference in the distribution of characteristics between the two groups, then they reveal correlates of the enlistment decision.

The second pattern studied is the determinants of height, which is done—putting aside selection issues—by studying the relationship of stature data in the Union Army with the characteristics observed therein or through the linkage to county characteristics based on the county of enlistment.

## 4.2 Summary Statistics

Figures 4 and 5 summarize the stature data. Figure 4 presents histograms of raw observed heights separately for Fogel et al.'s (2000) Union Army sample and for Cuff's (2005) Pennsylvania sample together with nonparametric estimates of the density in each sample, weighting Fogel et al.'s (2000) sample to match Gould's (1869) distribution of states. The Pennsylvania enlisters in panel (b) are 0.15 inches taller than Pennsylvania enlisters from the Union Army sample in panel (a), and the difference is statistically significant. In both cases, the distributions show the characteristic tendency to heap on whole inches, and the slight under-representation of individuals below the official minimum height requirement of 64 inches. Figure 5 presents the distributions of observed heights of Midwesterners and Northeasterners, combining the Fogel et al. (2000) and the Cuff (2005) data with the use of the Gould (1869) weights. A height premium for the Midwest is evident.

(a) Union Army (b) Pennsylvania Sample

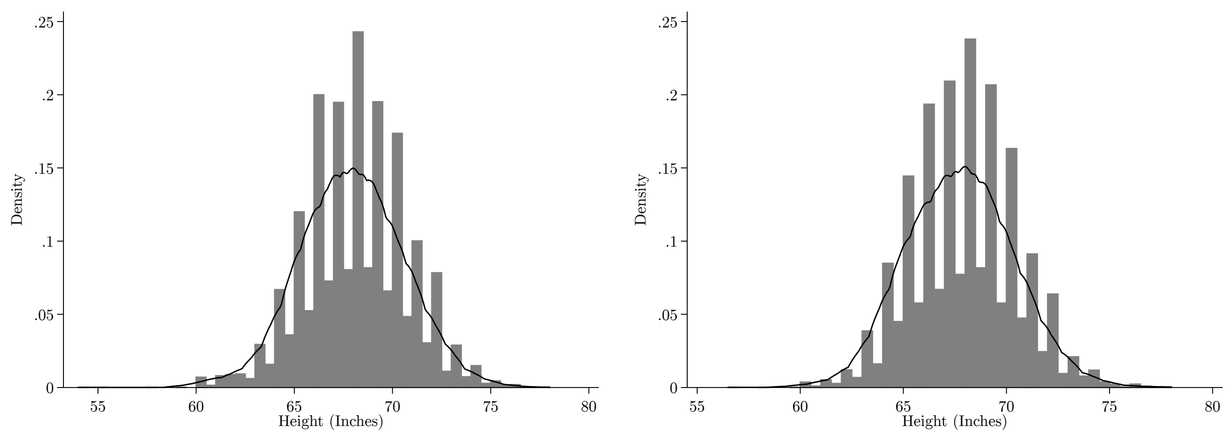



Figure 4: Height histograms and distributions by sample

*Note:* Panel (a) presents a histogram (with a bin width of 0.5 inches) for observations collected from Fogel et al.'s (2000) Union Army sample, along with an estimate of the underlying distribution, weighting the sample so that the distribution of states of enlistment matches the distribution provided by Gould (1869). Panel (b) provides a similar figure for data collected from Cuff's (2005) sample of Pennsylvanians in the Union Army.

Table 2 presents summary statistics for the variables described above. Columns (1) and (2) present summary statistics for enlisters in the Midwest and the Northeast; column (3) presents difference-in-means tests comparing columns (1) and (2); column (4) presents summary statistics for all individuals in the military data; columns (5) and (6) present summary statistics for the population of the Midwest and the Northeast; column (7) presents difference-in-means tests comparing columns (5) and (6); column (8) presents summary statistics for all census data; and column (9) presents difference-in-means tests comparing columns (4) and (8). In all cases, the enlistment data are weighted so that the distribution of states of enlistment matches the distribution presented by Gould (1869).

The two variables of greatest interest in this Table are height and the vote share for Lincoln. The first row of the Table confirms the insight given by Figure 5, that Midwesterners were taller than Northeasterners. In particular, Midwesterners exhibit a height advantage over Northeasterners of 0.73 inches, and the average of all observed heights is 67.83 inches, both after adjusting for the over-representation of Pennsylvanians.

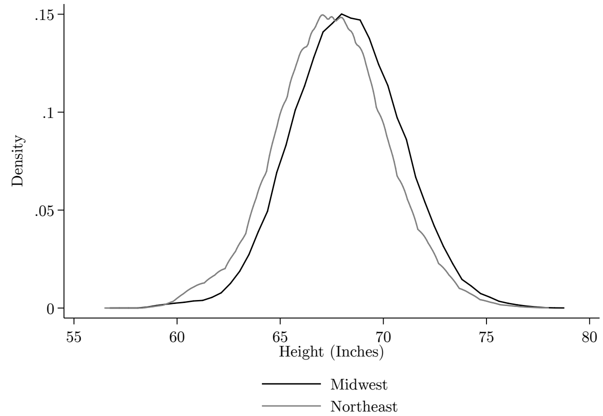



Figure 5: Height distributions by region

*Note:* This Figure presents the estimated height distributions for each region, combining both sources of military data and weighting observations to match Gould's (1869) distribution of states.

Table 2: Summary statistics

| | Military Only | | | | Census Only | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| *Variable* | MW | NE | MW-NE | All | MW | NE | MW-NE | All | Difference |
| Height | 68.209 | 67.481 | $0.728^a$ | 67.830 | | | | | |
| Inches | (2.515) | (2.613) | [0.089] | (2.592) | | | | | |
| Lincoln Vote Share | 0.525 | 0.591 | $-0.066^a$ | 0.559 | 0.531 | 0.572 | $-0.041^a$ | 0.555 | 0.004 |
| | (0.109) | (0.107) | [0.018] | (0.113) | (0.113) | (0.104) | [0.014] | (0.110) | [0.006] |
| Midwestern | | | | 0.479 | | | | 0.412 | $0.068^b$ |
| | | | | (0.500) | | | | (0.492) | (0.028) |
| Birthyear | 1838.324 | 1837.949 | $0.375^b$ | 1838.129 | 1835.963 | 1835.192 | $0.771^a$ | 1835.510 | $2.619^a$ |
| | (6.322) | (6.495) | [0.179] | (6.415) | (7.461) | (7.766) | [0.114] | (7.651) | [0.106] |
| log(Population Density) | 3.829 | 4.820 | $-0.991^a$ | 4.345 | 3.648 | 4.810 | $-1.162^a$ | 4.332 | 0.013 |
| | (0.804) | (1.769) | [0.324] | (1.478) | (0.700) | (1.792) | [0.311] | (1.555) | [0.068] |
| log(Agricultural Value per capita) | 3.885 | 3.302 | $0.582^b$ | 3.581 | 3.976 | 3.308 | $0.668^a$ | 3.583 | −0.001 |
| | (0.527) | (1.292) | [0.229] | (1.043) | (0.428) | (1.307) | [0.219] | (1.090) | [0.047] |
| log(Manufacturing Value per capita) | 3.469 | 4.390 | $-0.921^a$ | 3.949 | 3.175 | 4.334 | $-1.159^a$ | 3.860 | 0.089 |
| | (0.824) | (0.756) | [0.142] | (0.914) | (0.841) | (0.806) | [0.104] | (0.999) | [0.057] |
| log(Manufacturing Capital per capita) | 2.736 | 3.801 | $-1.064^a$ | 3.291 | 2.448 | 3.778 | $-1.330^a$ | 3.234 | 0.057 |
| | (0.856) | (0.702) | [0.134] | (0.944) | (0.850) | (0.765) | [0.095] | (1.034) | [0.059] |
| log(Agricultural Capital per capita) | 5.562 | 5.222 | $0.340^b$ | 5.385 | 5.644 | 5.229 | $0.415^a$ | 5.400 | −0.015 |
| | (0.451) | (0.997) | [0.168] | (0.803) | (0.432) | (1.004) | [0.157] | (0.843) | [0.035] |
| log(Real and Personal Estate per capia) | 6.174 | 6.330 | $-0.156^a$ | 6.255 | 6.141 | 6.327 | $-0.185^a$ | 6.251 | 0.004 |
| | (0.314) | (0.303) | [0.046] | (0.318) | (0.322) | (0.304) | [0.030] | (0.324) | [0.017] |
| White Collar | 0.053 | 0.071 | −0.018 | 0.062 | 0.096 | 0.170 | $-0.074^a$ | 0.140 | $-0.078^a$ |
| | (0.224) | (0.256) | [0.011] | (0.241) | (0.294) | (0.376) | [0.021] | (0.347) | [0.010] |
| Skilled | 0.164 | 0.318 | $-0.154^a$ | 0.244 | 0.153 | 0.283 | $-0.131^a$ | 0.231 | 0.013 |
| | (0.370) | (0.466) | [0.026] | (0.430) | (0.360) | (0.451) | [0.016] | (0.421) | [0.013] |
| Unskilled | 0.065 | 0.208 | $-0.142^a$ | 0.139 | 0.087 | 0.147 | $-0.060^a$ | 0.123 | $0.017^c$ |
| | (0.247) | (0.406) | [0.016] | (0.346) | (0.282) | (0.354) | [0.008] | (0.328) | [0.009] |
| Farmer | 0.718 | 0.404 | $0.314^a$ | 0.555 | 0.665 | 0.400 | $0.265^a$ | 0.507 | $0.048^a$ |
| | (0.450) | (0.491) | [0.032] | (0.497) | (0.472) | (0.490) | [0.031] | (0.500) | [0.016] |
| Observations | 9873 | 16716 | | 26589 | 11611 | 16594 | | 28205 | |

*Notes*: All figures in the enlistments are weighted to match Gould's (1869) distribution of states of enlistment. Standard deviations in parentheses. Standard errors, clustered by county, in square brackets.

The absence of stature data in the census (the source of the selection problem) makes a direct comparison of the heights of enlisters and the population as a whole impossible. The second row of the Table compares the vote shares for Lincoln. Northeasterners' counties of residence, both in the military and in the population, had a greater vote share for Lincoln than those of Midwesterners. This difference is about 7 percentage points in the military sample and about 4 percentage points in the population. Comparing the enlisting and non-enlisting groups reveals virtually no difference between them on the basis of the voting variables. Although this similarity between the vote shares suggests, on its face, that voting patterns did not enter into the enlistment decision, it must be remembered that this Table presents only the unconditional difference; the difference between columns (2) and (6) (enlisters and the population of the Northeast) suggest a conditional role for the vote share in determining military enlistment. Figure 6 presents the distribution of the vote share by region. While the support of the distribution is wider in the Midwest than in the Northeast, the Northeast has a higher average vote share for Lincoln.

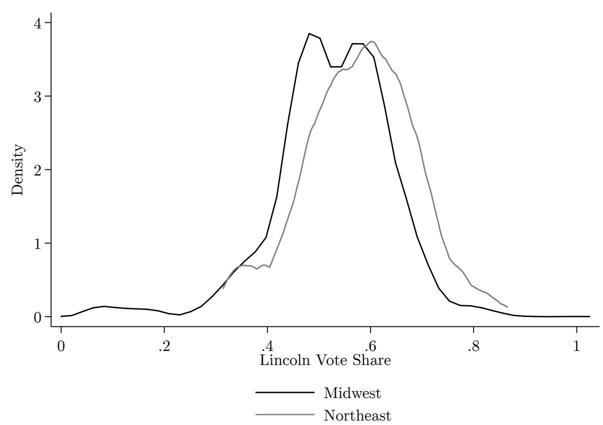



Figure 6: Distribution of vote shares by region

Indeed, except for differences in the regional representation (the enlistment sample statistically significantly over-represents the Midwest by about 7 percentage points) and in terms of birth year (enlisters are, on average, about 2.6 years younger) none of the other county-specific variables exhibit a large or statistically significant difference between the enlisting population and the census. For the only individual level variables that are observed, the occupational indicators, there are differences between enlisters and the complete population. In particular, the enlisted sample over-represents farmers and the unskilled, and under-represents those with white collar occupations. These patterns are typical of military enlistment in the 19th century (e.g., Zimran, 2017). But the comparison between the occupations of the military and census samples is complicated by the fact that they are observed up to five years apart and thus may not be directly comparable. For the birth cohort of 1847, for instance, the occupations in the census are of 13 year olds, while the occupations at enlistment are from 18 year olds in 1865.

# 5 Empirical Demonstrations of the Intuition

## 5.1 Selection on Observables

Table 3 presents two sets of regressions describing the relationship between the observable characteristics described in Table 2 on the one hand, and military enlistment and observed height on the other.[12] Columns (1)–(4) present the results of probit regressions for the probability of military enlistment, with columns (1) and (2) including a Midwest indicator, and columns (3) and (4) including state-specific fixed effects.[13] One key result of columns (1)–(4) is that the vote share for Lincoln enters with a positive and statistically significant coefficient, indicating that individuals from counties that were more supportive of Lincoln were more likely to enlist. The coefficients are not directly interpretable (because these are probit coefficients), but it can be shown that the coefficient 2.21 in column (3) can be interpreted as indicating that a 10 percentage point increase in Lincoln's vote share increases the probability of enlistment by 7.2 percentage points, relative to a base probability of 44.6 percent. These results show that, despite the preliminary indicators of Table 2, the vote share for Lincoln was related to the military enlistment decision.

Three other features of the results of columns (1)–(4) are notable. The first is that, all else equal, Midwesterners were more likely to enlist than Northeasterners. The second is that individuals from counties with greater manufacturing production per capita were more likely to enlist. The last is that individuals with skilled or unskilled occupations, or who were farmers, were more likely to enlist than were individuals with white collar occupations (the excluded group). The latter result, however, is potentially the result of different ages at which occupations were recorded in the military data and in the census, as discussed above.

Columns (5)–(8) present OLS regressions for the correlates of height in the military data without any correction for potential sample-selection bias. Columns (5) and (6) include a Midwest indicator and columns (7) and (8) include state-specific fixed effects. A conditional height advantage for the Midwest of about a

[12]Due to small sample sizes, I omit enlisters from Minnesota, Missouri, and Rhode Island, who would otherwise drive results because of the strong weight given to them by the Gould (1869) weights.

[13]Due to the unusual structure of the sample, columns (1)–(4) are not estimated by an ordinary probit regression (though the interpretation is the same). In the standard setting, the researcher observes a random sample of the population with the military enlistment status of all individuals. In this setting, I observe a sample of military enlisters and their covariates and a sample of the complete population with their covariates but without information on the military enlistment decision. Following Zimran (2017), I use Cosslett's (1981) method to estimate the model; I also use Zimran's (2017) weights that are necessary for this estimation.

Table 3: Relationship of covariates to enlistment probability and observed heights

| *Variables* | (1) Enl | (2) Enl | (3) Enl | (4) Enl | (5) Ht | (6) Ht | (7) Ht | (8) Ht | (9) Ht |
|---|---|---|---|---|---|---|---|---|---|
| Lincoln Vote Share | $1.112^b$ | $0.903^c$ | $2.212^a$ | $1.852^a$ | | | | | $0.720^b$ |
| | (0.530) | (0.489) | (0.696) | (0.586) | | | | | (0.334) |
| Midwest | $0.624^a$ | $0.517^a$ | | | $0.538^a$ | $0.465^a$ | | | |
| | (0.137) | (0.118) | | | (0.090) | (0.091) | | | |
| log(Population Density) | 0.035 | 0.069 | −0.106 | −0.066 | $-0.140^b$ | $-0.138^b$ | $-0.217^a$ | $-0.211^a$ | $-0.218^a$ |
| | (0.091) | (0.077) | (0.110) | (0.089) | (0.058) | (0.056) | (0.055) | (0.054) | (0.054) |
| log(Agricultural Value per capita) | 0.241 | 0.283 | 0.039 | 0.126 | $-0.239^c$ | $-0.301^b$ | $-0.262^c$ | $-0.340^b$ | $-0.304^b$ |
| | (0.217) | (0.186) | (0.260) | (0.220) | (0.143) | (0.133) | (0.149) | (0.136) | (0.143) |
| log(Manufacturing Value per capita) | $0.363^b$ | $0.385^a$ | $0.679^a$ | $0.648^a$ | −0.083 | −0.064 | 0.012 | 0.039 | 0.037 |
| | (0.165) | (0.148) | (0.209) | (0.179) | (0.096) | (0.094) | (0.102) | (0.101) | (0.104) |
| log(Manufacturing Capital per capita) | −0.049 | −0.141 | −0.284 | $-0.318^c$ | 0.028 | 0.029 | −0.064 | −0.079 | −0.084 |
| | (0.160) | (0.145) | (0.191) | (0.168) | (0.090) | (0.089) | (0.090) | (0.089) | (0.092) |
| log(Agricultural Capital per capita) | −0.200 | −0.227 | −0.224 | −0.291 | $0.321^c$ | $0.368^b$ | 0.253 | $0.322^c$ | 0.263 |
| | (0.204) | (0.173) | (0.258) | (0.223) | (0.181) | (0.168) | (0.206) | (0.181) | (0.201) |
| log(Real and Personal Estate per capita) | −0.043 | −0.062 | 0.083 | 0.025 | $-0.431^b$ | $-0.426^b$ | −0.255 | −0.247 | −0.286 |
| | (0.209) | (0.186) | (0.251) | (0.223) | (0.206) | (0.191) | (0.184) | (0.169) | (0.185) |
| Skilled | | $0.804^a$ | | $0.837^a$ | | 0.005 | | 0.005 | |
| | | (0.066) | | (0.065) | | (0.081) | | (0.080) | |
| Unskilled | | $0.797^a$ | | $0.805^a$ | | −0.049 | | −0.078 | |
| | | (0.088) | | (0.093) | | (0.098) | | (0.097) | |
| Farmer | | $0.638^a$ | | $0.667^a$ | | $0.317^a$ | | $0.324^a$ | |
| | | (0.061) | | (0.060) | | (0.086) | | (0.084) | |
| Constant | $-2.768^a$ | $-3.209^a$ | $-3.346^b$ | $-3.314^a$ | $69.905^a$ | $69.664^a$ | $69.624^a$ | $69.365^a$ | $69.460^a$ |
| | (1.065) | (0.976) | (1.507) | (1.271) | (0.812) | (0.801) | (0.796) | (0.772) | (0.797) |
| Observations | 54,660 | 45,651 | 54,660 | 45,651 | 26,585 | 25,470 | 26,585 | 25,470 | 26,585 |
| R-squared | | | | | 0.088 | 0.091 | 0.093 | 0.097 | 0.094 |
| State FE | No | No | Yes | Yes | No | No | Yes | Yes | Yes |
| Pseudo R-squared | 0.047 | 0.139 | 0.052 | 0.146 | | | | | |

*Significance levels*: $^a$ p<0.01, $^b$ p<0.05, $^c$ p<0.1
*Notes*: Dependent variable is an indicator for military enlistment in columns with the header Enl and height in inches in columns with the header Ht. The sample includes all individuals with height data or in the census sample, excluding residents of Missouri, Minnesota, and Rhode Island. All specifications include birth year fixed effects and all specifications with height as the outcome also include age-of-measurement fixed effects to standardize age of measurement to age 21. All specifications are weighted to match the distribution of states of enlistment. Standard errors in parentheses, clustered at the county level.

half inch is evident in columns (5) and (6), and in all four of these specifications individuals from counties with greater population density and greater agricultural output per capita tended to be shorter. Moreover, individuals reporting an occupation of farmer at enlistment were, on average, about 0.30 to 0.35 inches taller than individuals reporting other occupations.

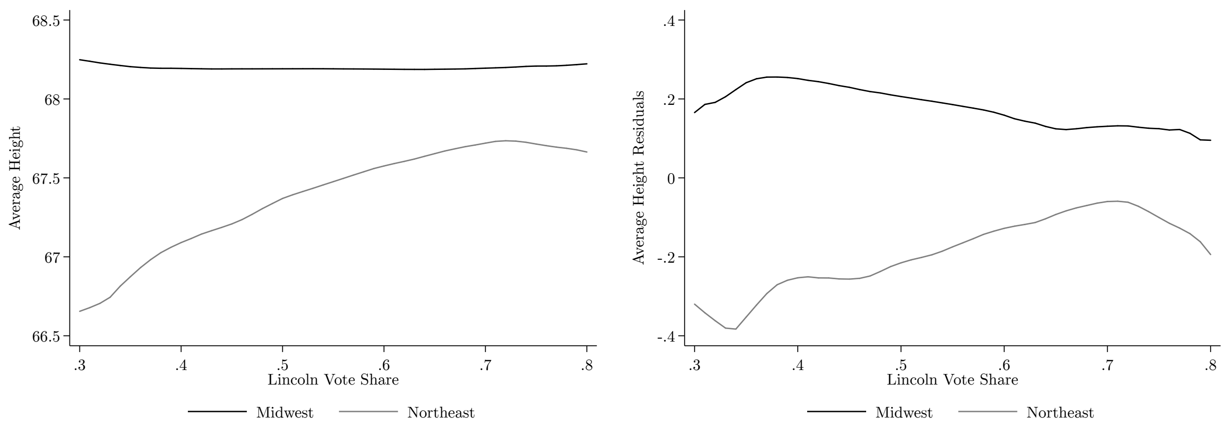
The results of columns (1)–(8) of Table 3 relate to Intuition 1, that if a variable that affects height is over- or under-represented in the sample relative to the population, selection on observables has occurred. If the goal were to compute the average height of the population using the observed military data, a simple calculation would seemingly suffer from bias due to selection on observables: all else equal, Midwesterners appear to have been taller and were more likely to enlist in the military. However, this evidence is only suggestive, because the regressions of columns (5)–(8) do not correct for potential selection on unobservables. The correlation of occupation with height and with the probability of military enlistment also provides suggestive evidence of the presence of selection on observables that would affect the comparison of Northeastern and Midwestern heights. This evidence is only suggestive, however, both because of selection and because of the difficulties in comparing occupations between the sources, as discussed above. The lack of another variable that affects military enlistment and that (appears to) affect height (not taking into account selection on unobservables) suggests that sample-selection bias from selection generated by other observable characteristics is unlikely to be severe (but again, no firm conclusion can be drawn from these regressions because no correction for selection on unobservables is made).

## 5.2 Selection on Unobservables

Column (9) of Table 3 repeats the regression of column (8) of the same Table, but adds Lincoln's vote share as a regressor. Figure 7 presents a similar exercise graphically. Panel (a) plots a non-parametric relationship between average stature and the vote share for Lincoln, divided by region, while panel (b) does the same, replacing average height with the average of the residuals of height from the regression of column (5) of Table 3, excluding the Midwest indicator.[14] All three of these exercises relate to Intuition 2, that if a variable that is unrelated to the outcome in the population but influences entrance into the sample is correlated with the outcome in the sample, then selection on unobservables is present. The vote share for Lincoln is taken as such a variable. Its relationship to the enlistment decision is clear from columns (1)–(4) of Table 3. That it does not affect height in the population must be assumed, as testing this directly would require observation of heights of the entire population, or at least of a representative sample. Under this assumption, however,

[14] If the Midwest indicator were retained, both sets of residuals would have zero means by construction.

these exercises suggest the presence of sample-selection bias induced by selection on unobservables, as a positive relationship between height and the vote share is evident. As this is a multivariate setting, however, no firm conclusion can be drawn without studying enlistment probability as I will do below. Similarly, it is important to note that if the vote share were not related to height in the sample, this would not indicate the absence of selection on unobservables, as it is ultimately the probability of entering the sample that must be related or unrelated to the outcome.[15]




(a) Height and vote share

(b) Height residuals and vote share

Figure 7: Height and residuals by region and Lincoln's vote share

*Note:* The residuals are from the regression of height on all variables in column (5) of Table 3 except the Midwest indicator. Each graph presents the results of a non-parametric regression for each region. The upper and lower extremes of the support of the vote share are omitted for clarity.

Figure 7 also enables an informal application of Intuition 4, that the magnitude of the sample-selection bias induced by selection on unobservables is decreasing in the probability of entering the sample and is zero at the upper extreme of enlistment probability. In this Figure, this intuition can be informally applied by realizing that the results of columns (1)–(4) imply that the probability of entering the sample is increasing in the vote share. The application is only informal because the vote share is only one determinant of many of the probability of enlistment; a formal application of that intuition requires study of the probability of enlistment. The key pattern in Figure 7 is that the Midwest's height premium over the Northeast is present over all values of the vote share—and in particular at its upper extreme—suggesting that it is not simply the product of sample-selection bias.

These exercises can be formalized by studying the estimated probability of enlistment for each individual

[15]It is also possible that using a different exclusion restriction would show a relationship. I do not delve deeper into this issue, as I have assumed that the models for height and enlistment are correctly specified.

in the data. Figure 8 presents the distribution of estimated conditional enlistment probabilities for each region based on the estimates of column (3) of Table 3. Panel (a) of this Figure presents these estimates for the census sample, while panel (b) covers the enlisting population. Two features of these plots are notable. The first is that, unsurprisingly, the enlisters have a greater enlistment probability than the population as a whole. Second, the enlistment probabilities in the Midwest are, on average, greater than in the Northeast. This indicates that the greater tendency of Midwesterners to enlist outweighs the combination of the greater tendency to enlist of Lincoln-supporting counties and the greater Lincoln vote share in the Northeast (Figure 6). It should be noted that the fact that the vote share is overpowered in determining enlistment by the Midwest indicator is not problematic—the vote share need not be the chief determinant of military enlistment; it simply must generate variation in military enlistment without generating variation in actual stature.

(a) Census (b) Enlisters

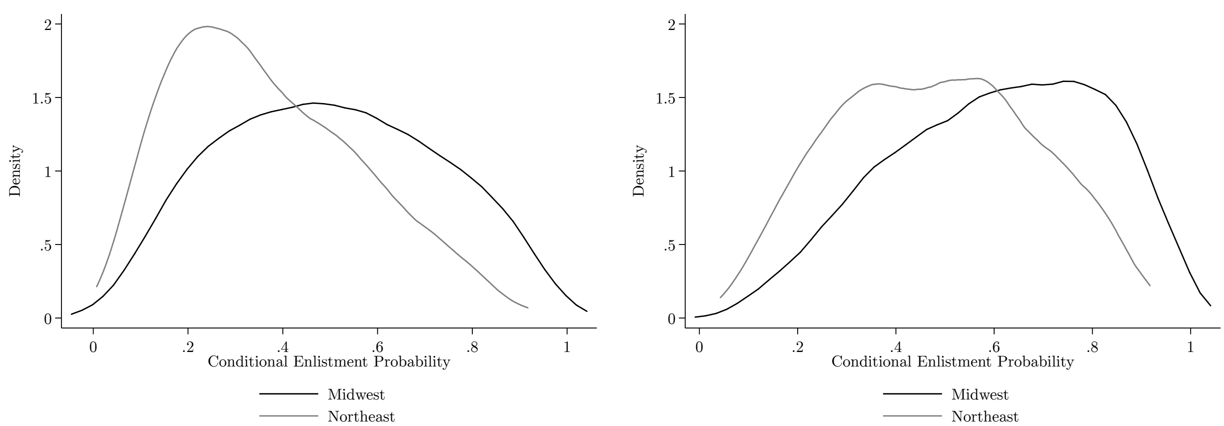



Figure 8: Distribution of estimated enlistment probabilities by region

*Note:* Enlistment probability is estimated from the results of the probit regression of column (3) of Table 3.

Figure 9 repeats the plots of Figure 7, but replaces the vote share on the $x$-axis with the estimated conditional enlistment probabilities whose distributions are plotted in Figure 8. Two patterns in this Figure are notable. The first is the difference in regional average stature among individuals with a high conditional enlistment probability. Intuition 4 implies that individuals with a high conditional enlistment probability do not suffer from sample-selection bias induced by selection on unobservables as much as do individuals with a lower enlistment probability. The persistence of the Midwest's height premium in the case of individuals with high enlistment probability supports the notion that it is a true difference and not a statistical artifact. Indeed, a Midwestern premium is present throughout the range of enlistment probabilities, further supporting the notion that it is not an artifact. The second notable pattern in Figure 9 is that panel (b) shows that, after taking into account the influence of observable characteristics, height is increasing in the probability of enlistment. This bears on Intuition 2, that the presence of such a relationship indicates the presence of sample-selection bias generated by selection on unobservables. It also bears on Intuition 3, that if the outcome is increasing in the probability of entering the sample, then selection on unobservables is negative. Thus, the increase of height residuals with with enlistment probability indicates that selection into the military is negative. Note that the positive slopes of Panel (b) conflict with the negative slopes of Panel (a). This drives home the importance of conditioning on observables in making the comparison of the outcome to the probability of entering the sample. When this conditioning is not done, as in Panel (a), the relationship is negative because the observable factors that make enlistment more likely are also associated with lower stature. Only when the role of these observables is addressed, as in Panel (b), the role of unobservables in driving negative selection is revealed.

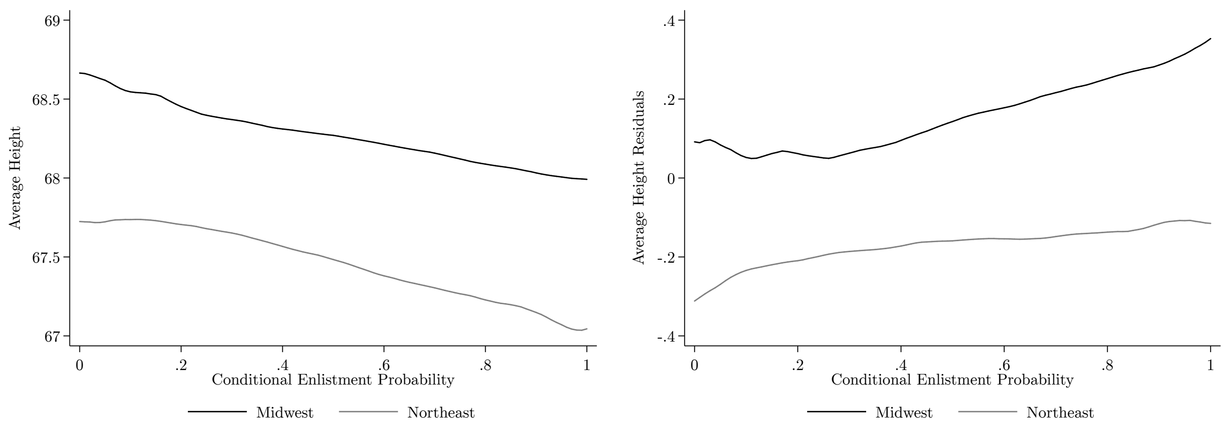



Figure 9: Height and residuals by region and estimated enlistment probability

*Note:* The $x$-axis in both graphs is the enlistment probability estimated in column (3) of Table 3. The residuals of height are residuals from a regression of heights on all variables in column (5) of Table 3 except for state fixed effects. Each graph presents results of a non-parametric regression for each region.

How would the presence of sample-selection bias induced by selection on unobservables affect the comparison of average heights of the Northeast and the Midwest? Figure 10 repeats panel (b) of Figure 9, but indicates approximately the points at which the bulk of Northeasterners and Midwesterners are located in the distribution of enlistment probability, as shown in Panel (b) of Figure 8—points $A$ and $C$, respectively. The effect of selection on unobservables on estimation of the height difference between the regions can be illustrated by comparing these points. Point $A$ is (loosely) the average observed height of Northeasterners, while point $C$ is (again loosely) the average observed height of Midwesterners. A comparison of these two points to one another yields the Midwest's observed height advantage. But this comparison conflates two differences—the true Northeast-Midwest difference and the difference in sample-selection bias between the regions, which is greater for the Northeast at point $A$ than for the Midwest at point $C$ (based on Intuition 4). A better comparison would be of points $A$ and $B$, which compares individuals with the same enlistment probability, and thus the same degree of sample-selection bias. More generally, rather than computing the difference in heights between the Midwest and the Northeast using the distribution of enlistment probabilities in the data (Figure 8), a correct comparison would be a weighted average of differences between individuals across regions with the same enlistment probability. This is essentially the intuition of the formal correction originally proposed by Heckman (1979) and applied to the historical heights literature by Zimran (2017).

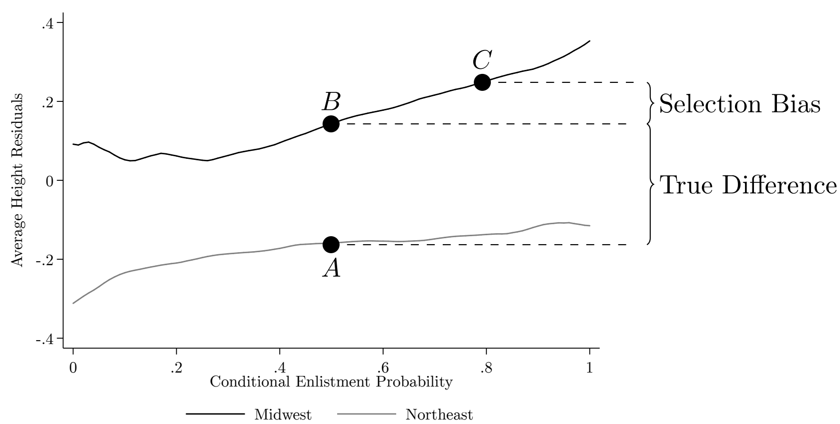



Figure 10: Residuals of height and enlistment probability, annotated

*Note:* This Figure repeats panel (b) of Figure 9, but is annotated to show the danger of comparing individuals with different enlistment probabilities instead of making comparisons only for individuals with the same enlistment probabilities.

# 6 Conclusion

Sample-selection bias generated by selection on observables and selection on unobservables poses a central challenge to the use of historical data to draw conclusions about broader populations of interest. While the presence of sample-selection bias generated by selection on observables is commonly recognized and addressed through simple methods, bias generated when characteristics unobservable to the researcher affect both the probability of entering the sample and the outcome of interest is less commonly acknowledged and more difficult to address. Bodenhorn, Guinnane, and Mroz (2017) have recently pointed out that this shortcoming in the historical heights literature has the potential to have generated spuriously some important conclusions in economic history, and their critique is applicable to a variety of historical sources.

In this paper, I illustrate the intuition of a method that is commonly used by economists to address sample-selection bias generated by selection on unobservables, and which has recently been applied to the historical heights literature by Zimran (2017). I develop four basic pieces of intuition to guide quantitative social science historians in their engagement with sources whose use in drawing conclusions about the broader population might be prevented by the presence of sample-selection bias. Of these, three relate to selection on unobservables. After demonstrating them in simple stylized examples, I apply them to the case of using the Union Army data to draw conclusions about the difference in average stature between the Midwest and the Northeast, showing that there is evidence of negative selection into the military that was stronger in the Northeast than in the Midwest, but that the Midwest's height premium is likely not a statistical artifact.

Although the focus of this paper is on the historical heights literature, the intuition explained and exhibited is applicable to a broad range of applications, and can help researchers to recognize when sample-selection bias may be present in their data and to understand the role that this bias might play in affecting conclusions drawn from this data. The most obvious application is to cases like the historical heights literature in which the researcher wishes to learn the average of some outcome for a population from a possibly selected sample. But sample-selection bias need not be thought of as simply an obstacle to overcome. In some cases, selection on unobservables may be an outcome of interest, such as in Ferrie's (1997) and Stewart's (2006) studies of migration to the frontier in the nineteenth-century United States. In such cases, although the role played by selection is different, the intuition to recognize its presence and to understand its role is the same as the example discussed throughout the present paper.[16]

The key to this intuition is the recognition of a variable that affects only whether an individual is observed, but has no impact in the population of interest. In the case of enlistment in the Union Army and the observation of height therein, that variable is taken to be the county-level vote share for Abraham Lincoln in the Presidential Election of 1860. Scholars seeking to apply this intuition to other settings will likely find that choosing this variable is the greatest challenge. The key is to ensure that this variable is what economists call an "exclusion restriction" and not an "omitted variable." That is, the researcher must be careful that the variable selected does not belong in the model as a variable affecting the outcome.

It is important to note that the intuition highlighted here is not a substitute for the direct and formal correction proposed by Heckman (1979) and applied to the historical heights literature by Zimran (2017). The goal of the present paper is simply to develop a better understanding of what it is that this method

[16]In the case of these studies, the outcome of interest is wealth accumulation, and the selection issue is the choice of where to live (on the frontier or not). The selection problem is that, for example, frontier wealth accumulation is observed only for those who chose to move to the frontier and not for the whole population.

actually does, and to provide scholars with a simple (but incomplete and informal) method to check for the presence and impact of sample-selection bias generated by selection on unobservables. Regardless of the way in which researchers confront problems of bias in their data, it must be kept in mind that no statistical exercise is a substitute for serious consideration of the limitations of a data source.

# References


Abramitzky, Ran (2015). "Economics and the Modern Economic Historian." *Journal of Economic History* 75:4, pp. 1240–1251.

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.

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.

Bolt, J. and J. L. van Zanden (2013). "The First Update of the Maddison Project: Re-Estimating Growth Before 1820." Maddison Project Working Paper 4.

Brown, John C. (1990). "The Condition of England and the Standard of Living: Cotton Textiles in the Northwest, 1806–1850." *Journal of Economic History* 50:3, pp. 591–614.

Clark, Gregory, Michael Huberman, and Peter H. Lindert (1995). "A British Food Puzzle, 1770–1850." *Economic History Review* 48:2, pp. 215–237.

Collins, William J. (2015). "Looking Forward: Positive and Normative Views of Economic History's Future." *Journal of Economic History* 75:4, pp. 1228–1233.

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

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

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

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

Crafts, N. F. R. (1985). "English Workers' Real Wages During the Industrial Revolution: Some Remaining Problems." *Journal of Economic History* 45:1, pp. 139–144.

——— (1997). "Some Dimensions of the 'Quality of Life' during the British Industrial Revolution." *Economic History Review* 50:4, pp. 617–639.

Crafts, N. F. R. and C. Knick Harley (1992). "Output Growth and the British Industrial Revolution: A Restatement of the Crafts-Harley View." *Economic History Review* 45:4, pp. 703–730.

Craig, Lee A. (2015). "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.

Cuff, Timothy (2005). *The Hidden Cost of Economic Development: The Biological Standard of Living in Antebellum Pennsylvania*. Burlington, Vermont: Ashgate.

Eli, Shari, Laura Salisbury, and Allison Shertzer (2016). "Migration Responses to Conflict: Evidence from the Border of the American Civil War." NBER Working Paper 22591.

Feinstein, Charles H. (1998). "Pessimism Perpetuated: Real Wages and the Standard of Living in Britain during and after the Industrial Revolution." *Journal of Economic History* 58:3, pp. 625–658.

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

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.

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.

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

Goldin, Claudia and Robert A. Margo (1992). "Wages, Prices, and Labor Markets before the Civil War." In *Strategic Factors in Nineteenth Century American Economic History*. Claudia Goldin and Hugh Rockoff (ed.). Chicago: University of Chicago Press. Chap. 2, pp. 67–104.

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

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

Haines, Michael R., Lee A. Craig, and Thomas Weiss (2003). "The Short and the Dead: Nutrition, Mortality, and the 'Antebellum Puzzle' in the United States." *Journal of Economic History* 53:2, pp. 382–413.

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.

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.

——— (1992). "Toward an Anthropometric History of African-Americans: The Case of the Free Blacks in Antebellum Maryland." In *Strategic Factors in Nineteenth Century American Economic History: A Volume to Honor Robert W. Fogel*. Claudia Goldin and Hugh Rockoff (ed.). Chicago: University of Chicago Press. Chap. 10, pp. 297–329.

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.

Kosack, Edward and Zachary Ward (2014). "Who Crossed the Border? Self-Selection of Mexican Migrants in the Early 20th Century." *Journal of Economic History* 74:4, pp. 1015–1044.

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

Margo, Robert A. and Georgia C. Villaflor (1987). "The Growth of Wages in Antebellum America: New Evidence." *Journal of Economic History* 47:4, pp. 873–895.

McCloskey, D. N. (1976). "Does the Past Have Useful Economics?" *Journal of Economic Literature* 14:2, pp. 434–461.

Mokyr, Joel (1988). "Is There Still Life in the Pessimist Case? Consumption during the Industrial Revolution, 1790–1850." *Journal of Economic History* 48:1, pp. 69–92.

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.

Pope, Clayne (1992). "Adult Mortality in America before 1900: A View from Family Histories." In *Strategic Factors in Nineteenth Century American Economic History: A Volume to Honor Robert W. Fogel*. Claudia Goldin and Hugh Rockoff (ed.). Chicago: University of Chicago Press. Chap. 9, pp. 267–296.

Records of the Adjutant General's Office (1861–1865). Regimental records, including descriptive rolls, order, and letter books, and morning reports, of volunteer organizations, Civil War, 1861–65. Records relating to volunteers and volunteer organizations. Record Group 94.2.4. Washington, DC: National Archives Building.

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.

Schwarz, L. D. (1985). "The Standard of Living in the Long Run: London 1700–1860." *Economic History Review* 38, pp. 24–41.

Stewart, James I. (2006). "Migration to the Agricultural Frontier and Wealth Accumulation." *Explorations in Economic History* 43, pp. 547–577.
Vella, Francis (1998). "Estimating Models with Sample Selection Bias: A Survey." *Journal of Human Resources* 33:1, pp. 127–169.
Voth, Hans-Joachim (2003). "Living Standards during the Industrial Revolution: An Economist's Guide." *American Economic Review, Papers and Proceedings* 93:2, pp. 221–226.
Voth, Hans-Joachim and Timothy Leunig (1996). "Did Smallpox Reduce Height? Stature and the Standard of Living in London, 1770–1873." *Economic History Review* 49:3, pp. 541–560.
Zehetmayer, Matthias (2011). "The Continuation of the Antebellum Puzzle: Stature in the US, 1847–1894." *European Review of Economic History* 15, pp. 313–327.
Zimran, Ariell (2017). "Does Sample-Selection Bias Explain the Antebellum Puzzle? Evidence from Military Enlistment in the Nineteenth-Century United States." Mimeo., Vanderbilt University.