

NBER WORKING PAPER SERIES

CAVEAT LECTOR: SAMPLE SELECTION IN HISTORICAL HEIGHTS AND THE
INTERPRETATION OF EARLY INDUSTRIALIZING ECONOMIES

Howard Bodenhorn
Timothy Guinnane
Thomas Mroz

Working Paper 19955
<http://www.nber.org/papers/w19955>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2014

We thank John Murray for providing his data on Amherst College students. We also thank Ray Cohn and Tomas Cvrcek for comments on an earlier draft. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2014 by Howard Bodenhorn, Timothy Guinnane, and Thomas Mroz. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Caveat Lector: Sample Selection in Historical Heights and the Interpretation of Early Industrializing Economies

Howard Bodenhorn, Timothy Guinnane, and Thomas Mroz

NBER Working Paper No. 19955

March 2014

JEL No. N01,N31

ABSTRACT

Much of the research on height in historical populations relies on convenience samples. A crucial question with convenience samples is whether the sample accurately reflects the characteristics of the population; if not, then estimated parameters will be affected by sample selection bias. This paper applies a simple test for selection biased developed in Bodenhorn, Guinnane, and Mroz (2013) to several historical samples of prisoners, freed slaves, and college students. We reject the hypothesis of no selection bias in all cases. Using Roy's (1951) model of occupational choice, we interpret these findings as reflecting the economic forces that lead individuals to take the actions the led to inclusion in the sample. Our findings suggest that much of the evidence on the "industrialization puzzle" during the nineteenth century could reflect changing selection into the samples rather than changes in population heights.

Howard Bodenhorn
John E. Walker Department of Economics
College of Business and Behavioral Science
201-B Sarrine Hall
Clemson University
Clemson, SC 29634
and NBER
bodenhorn@gmail.com

Thomas Mroz
John E. Walker Department of Economics
222 Sarrine Hall
Clemson University
Clemson, SC 29624
tmroz@clemson.edu

Timothy Guinnane
Department of Economics
Yale University
PO Box 208269
New Haven CT 06520
timothy.guinnane@yale.edu

An online appendix is available at:
<http://www.nber.org/data-appendix/w19955>

1 Introduction

Introduced into the field just over three decades ago, anthropometric history is now one component of economic history's standard toolkit.¹ Anthropometrics' promise is founded on two bases, one conceptual and one practical. The conceptual basis originates in the recognition that humans are biological organisms that respond to their environments in predictable ways.² People grow taller the better their nutrition and the less harsh the disease and work environment. Changes in the average heights of populations reflect changes in well-being due to changes in the underlying environment, which is determined at least in part by socioeconomic factors, such as the absolute levels and the distributions of income, wealth, consumption, work effort and the incidence of chronic and acute infections. Average population height is a function of a society's level of net nutrition – the nutrients available for growth remaining after claims made on gross nutrition by labor and disease – so that increases in average height are typically associated with improvements in overall well-being and decreases in average height with a deteriorating environment in at least one dimension.³

The practical basis for anthropometrics follows from the wide availability of height data. Military recruiters measured prospective soldiers; school administrators measured their students; officials in the criminal justice system recorded various physical characteristics, including heights, of individuals entering into the system as a means of identification; slave traders did the same for slaves entering into the interregional slave trade; and county court clerks across the US South recorded the heights of free African Americans applying for their freedom papers. It is even possible to reconstruct heights from skeletal remains of ancient and aboriginal peoples.⁴ In periods when the data needed to construct the more familiar measures of economic well-being –

¹ Voth and Leunig, "Did Smallpox," 542, contend that, over the past three decades, more resources have been devoted to anthropometric history than to any other broad research agenda in economic history. Given the number of anthropometric papers appearing in the field's major journals, it is hard to argue with their assessment.

² Eveleth and Tanner, *Worldwide Variation*; Fogel et al, "Secular Changes;" Steckel, "Biological Measures."

³ Angus Deaton, *Great Escape*, 162, reminds us that while height data provides some information about well-being it is not a substitute for standard measures. Deaton points out several anomalies between incomes and heights in the modern developing world and argues that these "should help us resist the superficially attractive idea that population average heights can be used as some overall measure of wellbeing or of the standard of living.... Average height is not a sensible measure of wellbeing."

⁴ Steckel, "Health and Nutrition in Pre-Columbian America."

real per capita gross domestic product, real wages, and so forth – are unavailable, average height may provide insights into well-being; when familiar measures are available, average height may add nuance to our interpretations of those measures. Heights, to paraphrase the cliché, may help illuminate the statistical dark corners of economic history.

The product of the three-decade long anthropometric agenda can be usefully divided into either cross-sectional or time-series studies. Cross-sectional studies often find that relatively disadvantaged groups were shorter than relatively advantaged groups, though this need not be the case.⁵ Time-series studies trace average heights over time and the enterprise has developed into a cottage industry that documents the apparent anomaly that average heights in the nineteenth century appear to decline at about the moment modern economic growth emerged. People appear to have been shrinking in a growing economy. This so-called puzzle has garnered much attention and is increasingly accepted as one of the stylized facts emerging from cliometric history. John Komlos traces the puzzle from its introduction to the present and claims that the finding represents the sort of paradigm shift that demands a reinterpretation of the economic history of industrialization and modernization.⁶ Improvements in well-being did not necessarily follow from modern economic growth.⁷ Economic development, according to the “puzzle” interpretation, made a large enough proportion of the population worse off in one or more dimensions that average heights declined. The puzzle, thus, presents the era of early industrialization as one of decline in average well-being.

We are less ready to accept the puzzle as fact or the reinterpretation of early economic growth that follows from it. Our concern with the literature follows from its reliance on potentially unrepresentative samples, specifically its failure to account for the selection bias likely to arise from the use of these samples. The time-series and cross-sectional variation in sample heights may not correspond to the true underlying population values that are the statistics of genuine interest. Moreover, because the samples are subject to unknown selection bias, we doubt that the puzzle, as it is typically portrayed, was real. The failure to take selection seriously raises legitimate questions about the results.

⁵ Eveleth and Tanner, *Worldwide Variation*.

⁶ Komlos, “Three Decade ‘Kuhnian’ History .”

⁷ In some regard the puzzle is a generalization of the “pessimist” case made in the context of the British Industrial Revolution, though the heights literature emphasizes the existence of the puzzle in a much broader set of historical circumstances. See Feinstein, “Pessimism Perpetuated,” for a discussion of the real wages debate.

In an earlier paper (hereafter referred to as BGM), we adapt a standard Roy-type model of occupational choice (military versus civilian) to illustrate how standard military recruitment procedures may yield enlistments selected on height.⁸ If height is correlated with unobserved, individual-level productivity characteristics in the civilian and military sectors, or if height is differentially rewarded in the two sectors, then height samples drawn from either sector will provide a biased picture of the population distribution of heights. The problem is likely to be especially severe for military samples, when a small fraction of the population voluntarily enlists and enlistees are likely to be disproportionately drawn from a tail of the population height distribution, which would occur whenever there were time series variations in pay differentials by height or its correlates over time.

Assuming that the civilian sector offers a higher reward to abilities associated with height than does the military, BGM show that military samples typically underestimate the true population height, sometimes by biologically meaningful amounts. Moreover, the degree of bias in the military sample depends crucially on the rewards to heights and how its correlates differentially affect earnings in the two sectors. If, for example, real military wages are stable over time, but the returns to height in real civilian wages vary pro-cyclically over the business cycle, negative selection bias will be more severe during booms and less severe during busts. Heights observed in a military sample drawn across several cohorts and through long periods of time will confound true height changes over time with transitory shocks to military heights brought about by changes in enlistments over the business cycle. The existing literature has addressed the implications of *truncation*, the potential biases that result from the military practice of imposing minimum or maximum height requirements. It has failed to adequately address the potential for bias that arises from *self-selection* by height. Thus, our concern is not with the potential conflation of changes in heights and changes in well-being that may follow from changing height standards (or truncation). Rather, our concern is the conflation of changes in heights and changes in well-being that follow from changing height-based self-selection into the sample.

We apply the test for selection bias developed in BGM to three representative, non-military historical height data sets. Although the logic underlying our Roy-type model applies in the cross-section as well as the time-series, our diagnostic test is designed to reveal selection in

⁸ Bodenhorn, Guinnane and Mroz, “Problems of Sample-selection Bias.”

time-series data when there are multiple years of observation for the same birth cohort because it is these sorts of data that are used as evidence for the industrialization puzzle. In short, our diagnostic test is designed to reveal selection in a time-series, where year of enlistment, enrollment or other relevant observation-year variables are used to proxy for a range of macro forces that affect the decision to enlist or enroll. If heights at observation year for a given birth cohort vary significantly in the time series, provided that all measurements are made after the attainment of full adult height, the test provides evidence of height-based selection consistent with changes in aggregate or macroeconomic factors. It naturally follows that observed heights for birth cohorts need not well reflect the true population heights for these cohorts. We find compelling evidence of such height-based selection in each of the three samples we study here, which brings the puzzle into question.

2 Treatment of sample-selection bias in the analysis historical military heights

The scientific study of human physical growth – auxology – emerged in the 1830s, with studies by European scientists, including Louis-René Villermé, Adolphe Quetelet and Eduoard Mallet, who gathered information on the heights of army recruits in France, Belgium, and Switzerland respectively.⁹ Villermé drew a connection between height and health, Quetelet introduced the normal distribution to the study of practical scientific questions, including human growth; Mallet noticed a modern urban height advantage and attributed it to Geneva’s relative prosperity. A half-century later, J. T. Danson published his statistical study of English prisoners, which shows that males did not reach their terminal heights until after age 22.¹⁰ He concluded that armies should eschew 18-year old recruits because slightly older men who had already reached their terminal height would prove to be hardier soldiers. The thread that connects these studies to the modern literature, besides their concerns with human height and well-being, is that they relied on readily available convenience samples subject to unknown selection biases. Selection was an issue at the dawn of statistical anthropometrics and, as we argue below, remains an underappreciated issue in the literature. In this section, we review how selection issues have shaped the discussion of three principal sources of height data: military recruits, slaves and prisoners. With the exception of a fruitful debate about height-based selection into the slave

⁹ Staub et al, “Eduoard Mallet’s.”

¹⁰ Danson, “Statistical Observations.”

trade, concerns with selection and how it might influence the interpretation of results have not received sufficient attention.

2.1 Selection and military samples

Robert Fogel and his co-authors were the first to draw attention to the industrialization puzzle, namely that the positive correlation between height and per capita income observed in modern cross sectional studies did not hold in the time series for adult white male military enlistees in the late antebellum United States.¹¹ They postulated that the decline in heights was concentrated in the urban-born populations. Rapid industrialization, poor public sanitation, and more pronounced urban inequality meant that the conditions of city life deteriorated during industrialization. Urbanites were shorter as a consequence.

Several subsequent studies report similar patterns. Average heights of recruits into the Hapsburg military declined in the late eighteenth century.¹² The mid-eighteenth century peak was not attained again for nearly 150 years. A secular increase in English heights during the early stages of the Industrial Revolution reversed among cohorts born in the 1840s and 1850s, though recent analyses claim that the decline in heights may have begun as early as the 1770s, or even the 1750s.¹³ Between 1740 and 1800, the average height of northern Italian soldiers declined by 3cm.¹⁴ Existing interpretations of the available evidence from the Old World identify early industrialization as an era of shrinking men.

Studies of US soldiers report similar results.¹⁵ Mid-nineteenth-century military school students demonstrate trend changes in height inconsistent with existing evidence on trend

¹¹ Fogel et al, "Secular Changes."

¹² Komlos, *Nutrition and Economic Development*.

¹³ Floud, Wachter and Gregory, *Height, Health and History*.

¹⁴ A'Hearn, "Anthropometric Evidence."

¹⁵ Once the time series revealed the industrialization puzzle, it was quickly documented in several cross sections. Hapsburg army recruits from the most economically developed regions within the empire were the shortest, while recruits from the least developed regions were among the tallest. Similar patterns emerged elsewhere: poor Irish recruits into the English East India Company (EIC) were taller than less poor English recruits; mid-nineteenth-century Swedish soldiers from the less developed north and east were taller than recruits from the more developed west; urban Italians were shorter than their rural peers; Union Army troops from less-developed Kentucky and Tennessee were taller than troops from the Old Northwest who were taller yet than troops from industrializing New England; ex-slave recruits into the Union Army from less developed inland regions were taller than those from more commercialized coastal regions. These findings raise the same issues that our approach addresses, namely differential height-based selection based on occupational opportunities, but cross-section versions of the puzzle (why people in rural areas were taller than urban residents) are not as readily addressed as the time-series puzzle (why people grew shorter during early industrialization). See Komlos, *Nutrition and Economic Development*,

changes in economic growth. The mean height of 19-year old West Point cadets fell by a half-inch in the late antebellum era, but recovered by the end of the Civil War.¹⁶ The mean height of 19-year old Citadel students – a military academy in South Carolina – was stable up to 1900 and increased by about 2.5 inches by the 1930s. A widely reproduced diagram that combines average heights from U.S. soldiers and National Guardsmen demonstrates a decline in final adult heights for cohorts born between the 1830s and the 1880s, which reverses thereafter.¹⁷ This pattern of declining then rising height in the US and elsewhere, often referred to the literature as “height cycles,” is the centerpiece of the puzzle literature. Coclans and Komlos, in fact, insist that any residual controversy surrounding the puzzle centers on the “nature and causal connections of height cycles,” because the existence of the cycles “is no longer questioned.”¹⁸

But as Matthias Zehetmayer notes, the conventional wisdom, as expressed in the widely reproduced diagram, may be flawed because that diagram and others like it are developed from “peculiar” populations.¹⁹ Like Zehetmayer, we beg to differ with the Coclans-Komlos assertion. While we do not reject the possibility that average adult heights declined prior to the long secular increase in heights observed in the twentieth century in the developed world, we are skeptical about the existence and interpretation of eighteenth- and nineteenth-century cycles because the literature has not fully explored selection and the potential biases differential selection into the military over time may create in the data.

The failure to account for sample selection may follow, much like interpretations of the puzzle, from the reassurances offered by Fogel and his coauthors that selection bias in military samples was small and of little practical consequence.²⁰ As they note, selection bias is likely to be more significant in volunteer than conscript armies because volunteers may not be representative of the entire population. Their analysis of selection, however, focuses on three

Mokyr and O’Gráda, “Height and Health,” Sandberg at Steckel, “Overpopulation,” A’Hearn, “Anthropometric Evidence,” Johnson and Nicholas, “Health and Welfare,” and Margo and Steckel, “Heights of American Slaves.”

¹⁶ Komlos, “Height and Weight.”

¹⁷ See Steckel, “Stature and the Standard of Living,” p. .

¹⁸ Komlos and Coclans, “On the Puzzling Cycles,” 92. Explanations of the times-series puzzle build on those offered by Fogel and his coauthors (“Secular Changes”) and identify six likely causes: (1) increasing income inequality; (2) increasing income variability; (3) increases in the price of food relative to manufactured goods; (4) increasing distance between the production and the consumption of food, with the consequent loss of nutrients due to spoilage and waste; (5) increased work intensity; and, (6) increased infection rates and disease incidence. The difficulty lies in accounting for many or most of these effects, which, in the end, are more commonly asserted than shown to be the cause of height cycles or trends.

¹⁹ Zehetmayer, “Continuation of the Antebellum Puzzle,” 321.

²⁰ Fogel et al, “Secular Changes.”

features of the data: the representativeness of volunteers on observables; left-tail truncation; and normality of the observed heights distribution. It is clear that volunteer armies are not representative of the socioeconomic distribution of most populations in that recruits are disproportionately drawn from lower socio-economic classes, notably urban laborers and immigrants.

The issue is not whether there is selection on non-height observables. It is whether there is differential unobservable selection on height itself across the entirety of the height distribution. Knowing whether an army is made up of poor urban laborers does not provide much information about height-based incentives to enlist. Fogel and his coauthors argue against a sizeable selection effect based on the fact that the 41 percent of draftees rejected by the United States military in the Second World War averaged 0.22 inches less than inductees. The failure to account for rejected draftees would reduce average heights estimated from inductees alone by 0.09 inches, which is too small a bias to overturn any existing interpretation of height trends.

It is not clear how applicable the American experience in the Second World War is to the convenience samples of volunteer armies or to conscript armies that inducted much smaller fractions of the young adult male population or situations in which draftees could pay for substitutes or otherwise buy their way out. First, young men rejected from the WWII military were *shorter* than the general population, which suggests they were less healthy or suffered from maladies that may have been negatively correlated with height that disqualified them from service. Economic logic suggests, and the model in BGM predicts, an opposite effect for volunteer armies. Taller, healthier individuals whose civilian wages would likely exceed military wages may be increasingly less likely to volunteer as individual height increases, so that average military heights would be biased downward relative to the overall population, not upward as in WWII.

Second, solutions to the left-tail truncation problem, such as the quantile bend (QB) estimator, reduced sample maximum likelihood estimator (RSMLE), or other techniques designed to account for truncation, will not resolve any potential sample selection bias unless selection operates only through the minimum height standard. These empirical modeling approaches cannot control for any height-based selection above the truncation point. In a comment on the paper that introduced the RSMLE and QBE approaches, Meier notes that “recruitment effort,” and “disparagement of shorter individuals” might vary continuously over a

range of heights and could yield a selected military height distribution with a nearly normal distribution.²¹ In their rejoinder, Wachter and Trussell agreed that some scenarios for recruitment and volunteering could yield “spuriously normal observed distributions whose failure to represent the underlying population would be undetectable from internal evidence.”²²

Third, tests for normality alone do not identify possible selection on height. Despite an admonition that “careless application” of truncated sample correction procedures could lead to “erroneous conclusions,”²³ it is common practice in the literature to provide a graph of the distribution and be reassured from its apparent normality that selection bias is unlikely to compromise any of the results.²⁴ Fogel and his coauthors label the error of relying on normality to reveal potential selection bias the “basketball problem” in that a distribution (or QQ) plot of National Basketball Association (NBA) players is often not statistically different from a normal distribution, yet it is widely understood that professional basketball is an occupation for which participants are positively selected on height.²⁵ No reasonable researcher would claim that NBA heights, after correcting for selection on observables (race, socioeconomic status of parents, education, immigrant), reflect average population heights of young adult males.

The basketball-problem interpretation recognizes only one feature of potential selection bias; the complete selection problem is more complicated than recognizing that some subpopulations may be positively (NBA players) or negatively (jockeys) selected on height. BGM not only show that the power of standard normality tests is simply too low to uncover height-based selection bias, but also that selection on height – even in the NBA – might change over time. If competitive strategies change in basketball, it is unlikely that the returns to height would remain constant and that average NBA heights would be unresponsive to changes in the returns. The logic of a subtler conception of the “basketball problem” is applicable to military samples, as well. Evolving military strategy or recruitment efforts may lead to changes in

²¹ Meier, “Estimating Historical Heights: Comment,” 297.

²² Wachter and Trussell, “Estimating Historical Heights: Rejoinder,” 302.

²³ Fogel et al., “Secular Changes,” 460.

²⁴ See, for example, López-Alonso and Condey, “Ups and Downs,” whose discussion of sample-selection focuses exclusively on left-tail truncation; Morgan, “Economic Growth,” esp. note 8, whose discussion of sample-selection focuses on non-representative nature of government workers; Carson, “Biological Standard of Living,” p. 409 recognizes that Mexicans may be differently selected than US-born criminals, but does not consider the possibility that selection might change in response to economic conditions in the two countries; Tatarek, “Geographical Variation,” does not raise the issue of selection; Meisel and Vega, “Biological Standard,” do not provide a histogram, but report the Jacques-Bera test statistic for normality among Columbian passport holders, who the authors recognize as unrepresentative. Other examples are readily found in the literature.

²⁵ Fogel et al., “Secular Changes.”

average military heights that do not reflect changes in population heights. The military may have altered its minimum height standard over time, but that standard may or may not have reflected the totality of the military's (changing) preference for height. Komlos in fact, reports differential returns to heights for Hapsburg recruits.²⁶ Taller recruits were paid notably larger enlistment bonuses than recruits in the immediate neighborhood of the minimum height standard. If the Hapsburg recruiting officers changed the bonuses over time, cliometricians might observe changes in military height that are unrelated to changes in the mean heights of the underlying population of interest.

BGM explicitly model incentives to enlist in a volunteer army in response to changing states of the civilian and military economies in the spirit of Roy's occupational choice model.²⁷ Their approach generates a height-specific supply function for the military, which implies that the willingness of relatively tall individuals to enlist need not change in the same way as the willingness of relatively short people to enlist when the returns to employment in the two sectors do not change in unison. Simulations of the model reveal that modest differences in the returns to civilian and military employment lead to large changes in military heights even when mean population height does not change. Changes in observed military heights are driven by changes in the probability of individuals of different heights to appear in the military sample.

Our survey of the literature uncovered only a few instances in which the selection problem, as we characterize it, is recognized and carefully explored. Roderick Floud, for example, acknowledges that recruits into volunteer armies represented a self-selected sample that may not be representative of the population. But he argues that selection is "unlikely to be large enough to vitiate comparisons over time and between ... countries."²⁸ David Weir disagrees. He recognizes that if recruiter selection manifests as a strict exclusion only of those below the truncation point, methods such as RSMLE or QBE may correct for it, but if selection is continuous across the entire distribution of heights, these estimators will not generate accurate estimates of mean population height.²⁹ Farley Grubb, too, notes that "heroic assumptions about the randomness in height" of unmeasured potential recruits and about the nature of selection into

²⁶ Komlos, *Nutrition*, 237. The bonuses increased from 3 fl. For soldiers just 5'-0" to 45 fl. For soldiers 5'-5" and above.

²⁷ Roy, "Some Thoughts."

²⁸ Floud, "Heights of Europeans," p. 19.

²⁹ Weir, "Economic Welfare," 174-175.

the military are necessary to draw inferences about population heights from samples of recruits into volunteer armies.³⁰

In his critique of Komlos' analysis of West Point cadets, Robert Gallman not only refuted Komlos' claim of nutritional decline, but raised serious concerns about selection and representativeness. Even if average cadet height declined, that fact would be interesting, Gallman contended, only "if cadets can be taken to be a random sample of some larger, more interesting group – say all young white men in the United States. That, of course, cannot be."³¹ Not everyone was eligible for West Point and, of those who were, young men who attended were interested in either a military or engineering career. Despite the unrepresentative nature of the sample, it might still provide some insight but only if the characteristics of enrolling students were approximately constant over time. It seems unlikely that the pool of potential enrollees was unchanging because cohorts born after 1840 and interested in a military career faced the prospect of serving in an army with a large, permanent class of lieutenants, captains and majors with battlefield experience. This surely pushed many otherwise promising cadets into other pursuits. Gallman's critique is based on the notion that the "military economy" may evolve in ways that alter the height mix of people who choose to enlist.

Another discussion of selection in the heights literature comes from the finding that poor Irish recruits into the English East India Company (EIC) army were taller than similarly-situated English recruits.³² Several explanations have been offered for this counterintuitive insight, including the relatively nutritious diet of milk and potatoes and "epidemiological isolation."³³ But it is possible that the "poor-but-tall" Irish reflects differential self-selection into the EIC, rather than differential nutrition; because incomes were lower in Ireland than England, the relative quality of Irish recruits was higher. Irish soldiers in the EIC were relatively tall, even though they came from a poorer country, because the EIC drew a larger proportion of recruits from well-off Irish than well-off English families. Thus while the Irish EIC recruits were taller than their English counterparts, this does not imply that the populations from which they are drawn exhibit the same difference. Faced with less attractive civilian employment opportunities for a given height, taller Irish men were disproportionately more willing to enlist than were

³⁰ Grubb, "Lilliputians," 140.

³¹ Gallman, "Dietary Change," 194.

³² Mokyr and ÓGráda, "Heights of the British," and "Height and Health."

³³ Nicholas and Steckel, "Tall but Poor," pp.

equally tall English men, so that the relatively tall Irish recruits were a supply-driven rather than demand-driven selection phenomenon.³⁴

Our reading of the historical heights literature that uses military samples, which represents a plurality if not outright majority of historical studies, reveals that sample selection bias is underappreciated. When the issue is discussed, it is typically discussed in terms of left-tail truncation or selection on non-height observables, such as occupation, nativity, race and so on, rather than height itself.³⁵ The theory provided in BGM and the empirical analysis provided there and below shows that failure to account for sample selection bias in a meaningful way raises substantive concerns about the existence of and explanations for the industrialization puzzle.

3 Selection in nonmilitary samples

Military samples represent the earliest and the most commonly analyzed historical heights samples so it is not surprising that a great deal has been written about the statistical issues surrounding their use. How best to deal with truncation, height and age heaping, and selection are all concerns that have been discussed in the literature, though selection has not, in our opinion, received as thorough a discussion as it deserves. The historical heights literature does not rely solely on military heights, however. Several studies have used large samples of slaves, servants, students, and prisoners. Here again, some statistical issues about how best to work with the data and interpret the results have been discussed in the literature, including selection. This section provides a brief review of how the literature treats selection in the use of slave and prisoner data.

3.1 Selection and the slave trade

Given that one of the principal contributions of the modern heights literature is the demonstration that income and wealth inequality manifests itself in human height-at-age, it is not surprising that historians were intrigued by the anthropometric consequences of slavery. Relatively deprived children are consistently shorter at age than relatively well-off children, and it is hard to imagine a more potentially deprived population than slaves. Economic historians have produced several notable studies of slave heights, which have led to two general results.

³⁴ O'Gráda, "Anthropometric History," contends that height-based selection is likely to be responsive to current macroeconomic conditions.

³⁵ See, for example, Komlos, "Anthropometric History of Early-Modern France," 166-167.

First, slave children were “extraordinarily small.”³⁶ The mean height of slave children generally fell below the first percentile of modern stature until about their tenth birthday. Such low statures are sometimes observed in modern developing countries, but are unheard of in the developed world. Second, the growth of slaves in adolescence was so remarkably vigorous that adult slaves nearly attained the 20th percentile of modern stature. While they were shorter by about one inch than contemporary white Americans, *measured* adult slaves were taller than many contemporary European populations.

Economic historians have constructed plausible explanations for these two features of slave heights. Poor medical knowledge and poorer practice led to high infant mortality rates, which is probably indicative of high morbidity rates from both acute and chronic (gastrointestinal and diarrheal) infections.³⁷ Because diarrheal infection interferes with the nutrient-growth nexus, persistent endemic infection leads to height stunting. If nutrition is simultaneously low, the negative consequences on growth through infection are magnified. Slave children, according to this literature, survived on a low-quality diet of hominy and pork fat. Growth recovery began around age 10 because the typical slave child entered the plantation labor force around that time. Normally, the demands of heavy work expected of slaves would have further interfered with growth, but once children entered the labor force they received shoes, which reduced fecal-based gastrointestinal infections, as well as more and better food, perhaps as much as one-half pound of pork per day. The increased meat ration for working slaves was further supplemented by vegetables and legumes, which contributed to the apparent catch-up growth of North American slaves.

Although information on slave heights comes from a host of sources – “contraband” slaves that joined the Union Army in the 1860s,³⁸ notarized certificates of good behavior filed with New Orleans’ courts,³⁹ manumission and freedom papers,⁴⁰ and runaway slave advertisements,⁴¹ among others – the principal sources are the coastwise manifests filed by slave traders, which collected identifying information on slaves – including height – to limit the illegal

³⁶ Steckel, “Stature and the Standard of Living,” 1923; see also Steckel, “Peculiar Population,” and Steckel, “Growth Depression,” for discussions of the short stature of slave children.

³⁷ Kiple and Kiple, “Slave Child Mortality.”

³⁸ Margo and Steckel, “Heights of American Slaves.”

³⁹ Freudenberger and Pritchett, “Domestic United States Slave Trade.”

⁴⁰ Komlos, “Toward;” and Bodenhorn, “Manumission.”

⁴¹ Komlos, “Height of Runaway Slaves.”

international slave trade.⁴² The discussion of selection bias in these articles is limited to whether the slave trade was more subject to a lemons market or a “good apples” market. Heights would be downwardly biased if only slaves of below-average height entered the slave trade (a lemons market), or upwardly biased if the trade was characterized by an Alchian-Allen good-apples market, in which a fixed shipping cost represented a smaller fraction of the higher price received for taller and healthier slaves.⁴³ James Trussell and Richard Steckel doubt, but do not show, that either effect was substantial enough to introduce a selection bias into their analysis.⁴⁴

Jonathan Pritchett and his coauthors contend that the manifest sample is potentially subject to substantial selection on height of the “shipping the good apples out” variety.⁴⁵ That is, when a fixed transportation cost applies to similar goods, high-quality, high-priced goods (taller slaves in this instance) become relatively less expensive in the destination market. Four features are needed for the good-apples effect to hold: (1) transport costs must be non-negligible; (2) transport costs are not proportional to price in the source market; (3) the goods are close, but not perfect substitutes; and (4), the elasticity of substitution between each of the two goods in question and a composite third good (say, free or indentured labor) should not be substantially different.⁴⁶ Slave transport conditions satisfy conditions (1) and (2); conditions (3) and (4) are defensible. Tall and short slaves are imperfect substitutes in production; and free or indentured labor, as the historical record shows, were substitutable for slaves, whether tall or short.⁴⁷

Although the data have not been systematically analyzed with the selection-bias problem in mind, existing evidence suggests a positive selection on height bias among slave traders. Charles Calomiris and Jonathan Pritchett find that slave children shipped with their mothers were shorter than children of the same age shipped alone. This result is puzzling absent some selection into the slave trade based on height.⁴⁸ Custom frowned on and some states’ laws placed sharp limits on selling minor slave children, orphans excepted, without their mothers, and it defies the logic of the anthropometric and the economics-of-slavery literatures to believe that orphaned

⁴² Steckel, “Slave Height Profiles;” and Steckel, “Growth Depression and Recovery.”

⁴³ Alchian and Allen, *University Economics*.

⁴⁴ Trussell and Steckel, “Age of Slaves at Menarche,” 550-551.

⁴⁵ Freudenberger and Pritchett, “Domestic United States Slave Trade;” Pritchett and Freudenberger, “A Peculiar Sample;” and Pritchett and Chamberlain, “Selection.”

⁴⁶ Borcharding and Silberberg, “Shipping.”

⁴⁷ Grubb, “End of European Immigrant Servitude.”

⁴⁸ Calomiris and Pritchett, “Preserving Slave Families.”

slave children grew in more propitious environments than slave children residing with a mother.⁴⁹

Similarly, Barry Higman's study of Caribbean slaves provides evidence of height-based selection into the slave trade. The mean height of native-born Trinidadian adult males (25-40 years) measured in 1813 was 165.6cm (65.2 inches).⁵⁰ The mean height of creole male slaves, or those born in the New World, and imported into Trinidad from sugar islands was statistically significantly greater at 167.3cm (65.9 inches). Creole slaves imported from non-sugar islands were taller yet at 170.6cm (67.2 inches). Imported females were also taller than native-born slave women. The advantage of Higman's sample is that it is based on a census of Trinidadian slaves conducted by the British government in anticipation of general emancipation. Slave registers were open for public inspection, government officials visited plantations to confirm the initial returns, and corrections were made when necessary. The coastwise manifests, on the other hand, provide heights only of those entering into the trade and no account is made for static or dynamic selection into the manifest sample. Our reading of this literature leads us to conclude that slave children entered into the sample randomly, but that adults were likely positively selected on height. If this is so, the estimated catch-up is overstated, but we cannot know by how much because it is difficult to account for selection into the slave samples.

The available evidence, although not definitive, points toward height-based selection into the interregional slave trade. This type of selection need not be revealed by non-normality of height distributions, or heaping on round number heights or ages, or left-tail shortfall. That the selection process is not readily revealed does not, of course, imply that it is unimportant. Unrepresentative selected samples will yield incorrect inferences when selection is correlated with the variable of interest.

3.2 Criminals and prisoners

Anthropometricians' concerns with the condition of the working classes during industrialization, and the relatively wide availability of height data, have encouraged scholars to study the heights of incarcerated or transported criminals. Scholars making use of prison data

⁴⁹ Crawford, "Slave Family."

⁵⁰ Higman, *Slave Populations*. It is worth noting, too, that Higman ("Growth in Afro-Caribbean Slave") fails to find the remarkable catch-up growth discussed by Steckel. A not insubstantial fraction of the adult population of Caribbean slaves attain only the second centile of modern stature, which is far short of the 20th centile attained by the mean male in the coastwise manifests.

readily acknowledge that prison samples are not representative of the underlying populations in several dimensions: the samples are disproportionately made up of unskilled, low-skilled and low-wage workers, which some scholars cast as a positive feature of prison samples because they are interested primarily in the material conditions of the proletariat.⁵¹ Such studies could, in principle, re-weight these prison samples to reflect better the population distribution of occupations, race and other non-height observables, but this will not eliminate height-based selection unless all height-related selection were due to these few observable variables. What these studies cannot do, in practice, is correct the distribution of prisoner heights for selection on unobservable factors related to height. This is the crux of the problem.

The issue surrounding prison samples turns on whether there is a correlation between height and the propensity to engage in criminal activity and whether the strength of that correlation changes with changes in the macroeconomy. Negative selection on height into crime and prison seems likely. Modern studies find that height is positively associated with legitimate labor market outcomes (employment and wages), mediated through cognitive abilities and the accumulation of more human capital by taller youth and adolescents.⁵² Because taller individuals face relatively better legitimate labor market opportunities than shorter individuals, criminal activities are less attractive to taller individuals. Prisons are thus populated with short people because criminals are drawn disproportionately from the left-hand side of the height distribution.⁵³

The potential for compositional changes in the heights of individuals selected into prisons to drive estimated temporal changes in heights has not gone unnoticed. Several studies have considered how the relative under- or over-representation of certain groups, which may change over time, may influence estimated (versus true) population heights.⁵⁴ These assessments are irrelevant to the issue of height-based selection into their sample; they merely shows that there is no apparent differential selection into two sub-populations *within a given sample of prisoners*; it does not address how the subsets of the population selected into prison in the first place. The selection problem does not turn, for example, on the urban-rural mix of prisoners. Rather, the dynamic selection problem turns on whether men and women born into rural or urban areas,

⁵¹ Nicholas and Steckel, "Tall but Poor;" Riggs, "Standard of Living in Scotland;" and Carson, "Inequality."

⁵² Case and Paxson, "Stature and Status;" Persico, Postlewaite, and Silverman, "Effect of Adolescent Experience."

⁵³ Bodenhorn, Moehling and Price, "Short Criminals," table 2 and figure 1.

⁵⁴ See, for example, "Nicholas and Oxley, "Living Standards," and Johnson and Nicholas, "Health and Welfare."

conditional on height, were more or less likely to enter into crime, be apprehended, tried, and convicted in response to changes in the macroeconomy.

The Roy model of military recruitment developed in BGM can, with slight modifications, also be applied to prisoners. The model predicts that better economic conditions are consistent with shorter prisoners in both the cross-section and the time-series. If legitimate labor market opportunities are conditioned on height and economic expansions create more remunerative, legitimate opportunities, it is likely that taller individuals who may have selected into crime in bad times select into legitimate activities in good times. While criminals will exhibit a distribution of heights, they will be disproportionately drawn from the left-hand tail of the population distribution, more so as macroeconomic conditions improve. If the heights of prisoners serve as indicators of the biological standard of living, it appears that biological times are tough when economic times are good and vice versa, when, in fact, the negative correlation of heights and macroeconomic indicators reflects differential selection on height into the subset of the criminal class that gets caught and convicted.

We believe this effect is what underlies the finding that bad times produce tall criminals. In a study of Scottish prisoners, Paul Riggs claims that the sample is representative of Scottish working-class heights because “in a society of heavy drinkers ... many workers were at risk of being arrested and thus having their physical stature preserved in the historical record.”⁵⁵ Although his results generally jibe with the industrialization puzzle, he finds the “curious” result that those arrested in the 1840s were markedly taller than those arrested in other periods. Riggs considers this curious because the 1840s, known as the “hungry forties,” were years of hardship and hunger in Scotland. When considered in light of a Roy model of occupational choice, the result is not so curious. If the effect of food shortages and unemployment was that men moved from legitimate to criminal occupations, the deterioration in non-criminal opportunities would draw differentially more men into crime from the right-hand than the left-hand tail of the height distribution because the left-hand tail was already disproportionately represented in the criminal market prior to the downturn. Moreover, if right-tail entrants into the criminal market have relatively little criminal human capital on entry into the market, they were more likely to have been detected and arrested. Thus, heights apparently increase during a sharp economic downturn

⁵⁵ Riggs, “Standard of Living,” 64.

when the intuitive connection between the biological and economic standards of living suggest otherwise.

Studies of American prisoners, too, yield results seemingly at odds with the industrialization puzzle: heights of black men born into slavery did not decline in the 1830s and 1840s;⁵⁶ heights of Tennessee prisoners were stable in the late-antebellum era;⁵⁷ and the puzzle does not appear among Missouri's mid-nineteenth century prison population.⁵⁸ Ad hoc explanations were frequently used to reconcile these inconvenient facts with the puzzle. Slaves were insulated from the market and were provided with adequate nutrition; Tennesseans raised hogs and did not sell meat in the interregional market so that they had plentiful access to protein; and the tall-but-poor Ozark Missourians benefitted from access to dairy, unlike rich-and-short northern Missourians who mostly grew less protein-rich wheat (a variation of the tall-but-poor Irish explanation). In the last case, a Roy model predicts a south to north Missouri prisoner height gradient that is consistent with a north-to-south income gradient. The relatively less attractive non-criminal market opportunities in the Ozarks drew relatively taller men into criminal activity.

The BGM approach reflects the spirit of Occam's Razor, in that it provides internally consistent predictions that do not rely on ad hoc reconciliations of the evidence with the industrialization puzzle. Military and prisoner heights rise in bad economic times because alternative employments – military service and criminal activity – become relatively more attractive compared to prospects in the legitimate civilian labor market. The remainder of the paper sets out an intuitive approach to thinking about and empirically uncovering changes in height-based selection over time that might lead to biased results. We then take the test to the data and show evidence of selection in the data underlying three widely cited historical heights studies.

4. An empirical approach to uncovering selection bias

BGM develop a simple set of empirical tests for uncovering sample selection bias in nonrandom samples of heights. They recommend a regression of observed heights, for ages after the attainment of final adult height, on birth cohort dummy variables and variables capturing

⁵⁶ Komlos and Coclanis, "On the Puzzling Cycle."

⁵⁷ Sunder, "Height of Tennessee Convicts."

⁵⁸ Carson, "Inequality in the American South."

possible year of enlistment effects (e.g., enlistment year dummies, annual unemployment rates, or other relevant macroeconomic variables). If, after controlling for birth cohort-specific effects, heights are correlated with enlistment year variables, the data exhibits evidence of height-based sample selection. Such selection may result in biased estimates of cohort-specific height distributions (especially the estimated mean) over time. The basic intuition behind this approach is that different birth cohorts face different average macroeconomic conditions during the years when they are eligible to enlist. The *ceteris paribus* effect of birth cohort would confound this effect with any true variations in mean heights by birth cohort.

The BGM tests are all diagnostic: they can *detect* selection, but in general cannot *correct* for it. One reason for this is that most observed historical height distributions are self-selected as part of a dynamic decision making process. The anthropometric historian only observes men enlisting in the military at age 24 if they had not enlisted at age 23, or at age 22, at age 21, and so on. Similarly, a student enrolling in college at age 20 had not enrolled at 19 or 18 or 17. Samples of soldiers or students (or slaves or prisoners) are made of people who made a choice at one point that they had rejected at one or more prior periods. If that choice to enlist at time t is influenced by cohort effects *and* current conditions that may operate differently by height, we are faced with a dynamic selection process that may lead to incorrect inferences.

To highlight the essential issue of such a dynamic selection process consider a model of two successive cohorts, C_1 and C_2 . C_1 might be men born 1800, and C_2 men born in 1801. Members of each cohort can choose to enlist (or enroll or whatever) in at most one of two consecutive time periods. Overall, there are three time periods. In the first period (t_1) young members of cohort C_1 can enlist. If we are dealing with a military sample, t_1 might be 1820. In the second period (t_2) the now older members of cohort C_1 who had not enlisted in t_1 might enlist, as might young members of cohort C_2 . That is, enlistees in the period 1821 include both members of cohort C_1 (who are 21 years old) and members of cohort C_2 (who are 20). We assume there is a maximum age for enlistment, and members of cohort C_1 are too old to enlist in period t_3 . In period t_3 , only the now-older members of cohort C_2 who chose not to enlist in t_2 are eligible to enlist. This cohort enlistment implies four sets of cohort-enlistment pairs: (C_1, t_1) , (C_1, t_2) , (C_2, t_2) , and (C_2, t_3) . Denote the mean heights at enlistment for each of these sets of enlistees, respectively, be $M(1,1)$, $M(1,2)$, $M(2,2)$, and $M(2,3)$.

We assume that macroeconomic effects, if there are any, are independent of height and impact all cohorts in the same way. That is, if those in cohort C_1 who enlist at time t_2 are 1-inch shorter than their birth cohort's true mean height, then those in cohort C_2 who enlist at time t_2 are also 1-inch shorter than their birth cohort's true mean height. We could make alternative assumptions that would allow us separate macroeconomic time of enlistment effects from birth cohort effects, but we adopt a parsimonious approach here.

One regression model that can perfectly capture the four mean heights of enlistees, $M(1,1)$, $M(1,2)$, $M(2,2)$, $M(2,3)$, includes an intercept, a dummy variable for cohort C_2 , a dummy variable for period t_2 , and a dummy variable for period t_3 . Suppose we estimate this by OLS, where

$$h = \alpha + \beta_1 * t_2 + \beta_2 * C_2 + \beta_3 * t_3 + \varepsilon \quad (1)$$

where h is the individual enlistee's observed height; α is the estimated intercept, which is the estimated mean height of individuals from the first cohort enlisting in the first year of eligibility. β_1 , β_2 , β_3 are estimated cohort or period coefficients.

Then the estimated parameters correspond to:

$$\begin{aligned} \alpha &= M(1,1) \\ \beta_1 &= M(1,2) - M(1,1) \\ \beta_2 &= M(2,2) - M(1,2) \\ \beta_3 &= M(2,3) - M(1,1) - M(2,2) + M(1,2) \end{aligned}$$

That is, the cliometrician would fit exactly the mean heights for each of the four observed birth cohort-calendar year combinations in the data.

Consider, first, a situation where there is no self-selection related to height but there are time varying macroeconomic effects affecting enlistment "equally" at all heights. In this case, for each birth cohort the mean of heights for those eligible to enlist in their own cohort's second year is identical to that of the population at risk of enrolling in the birth cohort's first year. In this instance, the intercept would measure the population mean height for birth cohort C_1 plus the effect of t_1 macroeconomic conditions.

The coefficient on the C_2 dummy variable would measure how mean heights differ for birth cohort 1 and birth cohort 2. It is impossible to measure true mean height for either birth

cohort because we do not know whether the macroeconomic effect at any of the three time periods (t_1 , t_2 , t_3) is exactly zero. This appears to be the situation envisioned in many studies interpreting changes in observed mean heights over time; the claim is that samples might be biased, but the magnitude of bias does not change over time.⁵⁹ However, interpreting the C_2 coefficient as a true measure of cohort height differences relies crucially on the absence of dynamics in the enlistment process (or its observational equivalent).

Now suppose population height distributions do not vary across birth cohorts, the macroeconomic environment is unchanging over time, and heights are only recorded when the individual enlists or enrolls (dynamic enlistment). Suppose, further, there are individual level shocks at each time period that are independent across individuals and time periods and that the effects of these shocks on enlistment given height do not vary across time or birth cohorts. We also assume that these individual shocks equally affect the propensity of individuals of different heights to enlist. These assumptions generate a simple hypothetical world where all birth cohorts, all population level incentives, and all time periods are identical. In this situation there is an absence of height-based selection. The intercept measures the mean height, which is constant across birth cohorts. The coefficients on the three dummy variables (t_2 , t_3 , and C_2) would all equal zero because the height distribution for those who do enlist at any age and for every birth cohort would not change through time.

The interesting issue arises if we introduce negative selection on height for enlistment. We do this simply by specifying that for any particular value of the independent individual shock, shorter individuals are more likely to enlist in the military than taller individuals. Even with height-independent shocks, this introduction of negative selection on height implies that a larger proportion of shorter than taller individuals enter the military in their first year of eligibility. The height distribution for those individuals still eligible to enlist in the second year will first-order stochastically dominate that birth cohort's population height distribution. Given the stationarity, homogeneity, and monotonicity in height assumptions, the mean heights of enlistees in a cohort's second year of enlistment will be greater than the mean height of enlistees in the cohort's first year of enlistment.

This simple dynamic selection mechanism has a profound impact on the estimate of the coefficient on the dummy variable for the second birth cohort (C_2). First, the intercept will

⁵⁹ This is the assumption underlying Komlos and Kim, "Estimating Trends."

measure perfectly the mean height of those in birth cohort C_1 who chose to enlist in period t_1 . The coefficient on the dummy variable for second time period (t_2) will be determined so that predicted values perfectly match the mean height of enlistees for cohort C_1 in year t_2 . Given the simple dynamic selection process just described, this t_2 coefficient must be positive. Given the stationarity assumptions, the mean height for those from C_2 enlisting at time t_2 must be identical to the mean height of those from cohort C_1 enlisting at time t_1 . This implies that the coefficient on the dummy variable for the second birth cohort must be negative (and equal in magnitude to the coefficient on the t_2 dummy variable). *We have uncovered apparent evidence of declining “height” by birth cohort even though the hypothetical data generating-mechanism is perfectly homogenous across all birth cohorts and time periods.* It is because heights are recorded only at the time of first enlistment that we “uncover” such spurious effects.

The coefficient on the t_3 dummy variable will be estimated to fit exactly the mean height of those from C_2 who enlist at t_3 . Given the stationarity assumptions, it must equal twice the coefficient on the t_2 dummy variable. The time effects will appear as if the macroeconomic conditions are “improving” over time, in the sense that taller individuals are joining the military as time progresses.

Finally, consider adding additional birth cohorts (and time period effects) under the same set of stationarity, homogeneity, and monotonicity assumptions. With the large coefficient on the t_3 dummy variable obtained from the first 2 birth cohorts, the coefficient on the C_3 dummy variable must converge to twice the coefficient on the C_2 dummy variable. Cohort C_3 appears twice as bad off as cohort C_2 , relative to cohort C_1 , even though the populations are identical in all aspects. The time effects for t_4 and later would continue to be positive and increase linearly, and it would be necessary for the birth cohort effects to be negative and decline linearly in order to fit the data. Note that every time we add a new birth cohort we add two new “mean heights” but also two new parameters. Given the homogeneity, and monotonicity assumptions, we will fit all observed mean heights perfectly, but the pattern of cohort effects will appear to indicate a continuing decline in economic conditions over time for subsequent birth cohorts. That false interpretation/conclusion is due solely to the simple dynamics of the decision to enlist in the military.

5. Tests of selection bias in representative height samples

In this section we implement our selection diagnostic test described in §3 above in a study of three data sets representative of the types of data used in the literature, namely prisoners, free African Americans (including manumitted slaves), and students. Our diagnostic reveals some degree of selection in each data set, which raises concerns about inferences about trends or cycles in heights drawn from these and similar sources. Our objective is not to analyze every available dataset that includes heights. Rather, we focus on three relatively large datasets that encompass the types of studies in the literature.⁶⁰

5.1 *Pennsylvania prisoners*

Height data drawn from prison records are a widely used source of information on historical heights and many of the studies purport to provide evidence consistent with the antebellum puzzle.⁶¹ We ask whether there is evidence of self-selection in a typical sample of heights drawn from convicts incarcerated during the era of early industrialization drawing on data from the Pennsylvania penitentiary system between the late 1820s and the late 1870s. These data are similar to those used elsewhere in the literature. The Pennsylvania prisoner data are taken from ledgers maintained by clerks at the Eastern State Penitentiary in Philadelphia and at the Western State Penitentiary in Pittsburgh. At the prisoners' arrival at the prisons, clerks recorded basic information about the prisoners, including their names, ages, state of birth if native-born or country of birth if foreign-born, pre-incarceration occupation, the crime for which they were incarcerated, sentence length, prior convictions (if any), and the county of conviction. Two descriptive registers, one each from the Eastern and Western penitentiaries, included identifying information such as race (white, black, mulatto), eye color, brief descriptions of marks, scars, tattoos, or physical deformities. Most importantly for our purposes, the clerks recorded heights, typically to the nearest quarter-inch.

⁶⁰ BGM reports similar diagnostic tests for the British Army data that underlie Floud et al.

⁶¹ For studies of US prisons, see Komlos and Coelanis, "On the Puzzling Cycle;" Carson, "Inequality in the American South;" Maloney and Carson, "Living Standards;" Tatarek, "Geographical Height Variation;" Sunder, "Height of Tennessee Convicts." Nicholas and Steckel, "Heights and Living Standards;" Riggs, "Standard of Living;" Nicholas and Oxley, "Living Standards," investigate heights using prison records from Great Britain. Frank, "Stature" and Twrdek and Manzel, use heights from South American prisons.

The Pennsylvania prisoner data raises selection concerns because prisoners, especially those confined to state penitentiaries in the nineteenth century, were unlikely to represent random draws from the wider population. It is not even clear that they are representative of individuals engaged in criminal activity. Men incarcerated at the prison arrived after traversing a criminal process in which many participants in the law-enforcement sector made choices: individuals chose to (allegedly) commit a crime; the police chose whether to charge and arrest the suspect; the prosecutor chose whether to prosecute the case; a jury chose to convict and to impose a sentence of more than one year of incarceration. Ultimately, men committed to the state prisons were those who were convicted of relatively serious crimes. There is no guarantee that the choices made at each stage of the criminal process – commission, arrest, prosecution, conviction – was independent of the prisoners’ heights. Bodenhorn, Moehling and Price, in fact, show that criminals themselves were negatively selected on height.⁶² Criminals were short relative to their contemporaries and shorter men entered prison at younger ages. The potential for plausibly changing height-based selection into prisons raises concerns about inferences drawn from such data.

The mean age at admission into the Eastern and Western penitentiaries was 28.5 years, and ages ranged from 11 to 89 years. Criminologists identify the prime offending ages from the mid-teens to the mid-twenties, which is consistent with the historical data as well.⁶³ Because less-privileged individuals tend not to reach their terminal adult heights until around age 20 and because immigrants faced different childhood environments, we limit the sample to native-born, men 22 years and older.

We use the simple test for the absence of height-related selection in a particular type of sample as described above. Provided that all individuals in each birth cohort in the sample have reached terminal adult height and controlling for birth-year with a set of dummy variables, there should be no impact of any current (macroeconomic) variables on observed height. If heights, holding birth cohort fixed in such a sample, vary over time (measured either by age, observation year or specific current macroeconomic variable), the cliometrician can reject the null hypothesis of no height-based selection in favor of the alternative hypothesis of some form of height-based selection. Due to the exact relationship among age, time, and cohort variables, however, it is

⁶² Bodenhorn, Moehling and Price, “Short Criminals.”

⁶³ Moehling and Piehl, “Immigration, Crime and Incarceration .”

impossible to attribute any rejection of the null hypothesis to either a pure age effect or to a pure time period effect that influences the observed heights in the sample. All the cliometrician can do is reject the null that there is no height-based selection. It is important to note, as well, that implicit acceptance of such a null hypothesis underlies nearly all studies using variations in observed heights to draw inferences about changes in birth-cohort macroeconomic conditions, commonly interpreted as biological well-being, over time.

Our regressions for prisoners (and others) take the following general forms:

$$h_i = \alpha + \sum_c \beta_c * I_{ic} + \sum_t \beta_t * I_{it} + \gamma_i + \zeta_i + \rho_i + \varepsilon_i \quad (2)$$

I_{ic} is an indicator variable equal to one if the individual is a member of cohort c (i.e., born in year c) and zero otherwise, so the term $\sum_c \beta_c * I_{ic}$ captures a series of cohort or birth-year dummies. I_{it} is an indicator variable equal to one if the individual entered prison at time t and zero otherwise, so the term $\sum_t \beta_t * I_{it}$ captures a series of year-of-incarceration dummy variables, that proxy for year-specific macroeconomic effects. In some estimates we also include γ is a crime-specific dummy variable for crime (we restrict the sample to property crimes); ζ is an urban residence dummy variable; ρ represents a set of pre-incarceration occupation dummy variables; and ε is the error term.⁶⁴ Neither the results nor our interpretations are changed in a substantial way by the inclusion of additional correlates.

The excluded incarceration year is 1850 and the excluded birth year is 1825, so the estimated constant (reported in an online appendix) is the mean height of a 25-year old, born in 1825 and imprisoned for the first time in 1850. When we include the additional correlates, the excluded crime is larceny, the excluded region/urban category is rural, western Pennsylvania, and common laborer is the excluded occupation. We trim the sample by excluding individuals with reported height more than three standard deviations above or below the unconditional mean, which may reflect either coding errors or physical outliers (i.e., dwarfs). We also trim the sample to include only those individuals between 22 and 50 years, or those likely to have attained their

⁶⁴ The included crimes are arson, house breaking, burglary, counterfeiting, forgery, fraud, horse theft, larceny, and receiving stolen goods. Larceny is the most common crime. Urban places are Philadelphia and Pittsburgh. Occupations are divided into nine categories, including professionals, proprietors, service, operatives, craftsmen, farmers, clerks, farmers and no reported occupation.

terminal adult stature, but not so old as to have started to shrink. Finally, we trim the sample so that no cohort-year cell has fewer than five observations.

Table 1 reports the F-tests that result from the two OLS regression specifications described above (details of birth-year, age, and incarceration-year coefficient values are provided in an online appendix). The results summarized in Column 1, which pertain to all native-born white men, show that we can confidently reject the null that the incarceration year coefficients are jointly zero. Recall, that the hypothesis being tested is whether the heights of individuals in a given birth cohort differ based on the year in which they are observed. If entry into the prison was independent of macroeconomic factors, which we proxy with the incarceration years, then the estimated coefficients should be jointly zero. If, however, entry into prison was driven by some underlying macroeconomic feature such as employment or income that is plausibly correlated with height, the test should reject the null of no incarceration-year effects. Because the test statistic rejects the null at a $p\text{-value} = 0.005$, we can conclude that selection into Pennsylvania's prison was correlated with height.

Column 2 repeats the test on a smaller sample of native-born African-American men, who were overrepresented in the prison. We include a dummy variable to account for any systematic differences between black and mixed-race men. The results are consistent with the white-only sample. The F-statistics on the joint test of zero incarceration-year effects is rejected at standard levels. Column 3 reports test statistics for the combined white and African-American sample for which the null hypothesis of no selection is rejected. In all three samples, one (whites) relatively large by the standards of the historical heights literature and one (blacks) relatively small, we reject the null on the joint test of zero incarceration-year effects. We interpret the rejection to imply that some underlying dynamic selection process on height into Pennsylvania's prison operated in the mid-nineteenth century.

One question is the practical one of whether selection has any meaningful effect on the anthropometrician's interpretation of industrialization and the puzzle. Table 1 also includes F-tests of the null hypothesis that the birth-year coefficients are jointly zero. The results across the three columns report the curious case of a highly significant test statistic for blacks, and rejection at notably lower confidence levels for whites and for the combined black-white sample. Close inspection of coefficients reveals that the African-American birth-year coefficients are mostly positive with a slightly upward trend (excluded birth-year is 1825); the white birth-year

coefficients, by comparison, are mostly negative and demonstrate a modest downward trend. These results reveal one problem with drawing inference from selected samples. Because neither model corrects for (rather than simply identifies) selection, there is no way to determine which, if either, is the true effect.

That is, once we have identified statistically significant differences in height for different observations periods for a given birth cohort, then the interpretation issues arising from dynamic selection must be addressed. Real-world selection, however, would certainly be more complicated than that arising in the static macroeconomic environment offered in the discussion above. Selection is likely driven by time variations in macroeconomic forces, and this complicates the interpretation of the birth-year and observation-year effects in a non-trivial fashion. Thus, we do not discuss either birth-year or observation-year “effects” in our discussion, but we report the estimated coefficients in an online appendix.

5.2 Free-born and manumitted African Americans

A ten-fold increase in the free African-American population in Maryland and a five-fold increase in Virginia between the Revolution and the Civil War concerned contemporary whites who imposed a number of restrictions and regulations on manumission, the occupations African-Americans might pursue, and other features of free African-American life. One regulation imposed in both states in the post-Revolutionary era was that all free people of color were required to register with the county court and retain a notarized copy of the registration as proof of their freedom. County court clerks recorded information on the registrants, sometimes in special ledgers, sometimes in the regular court records. Most registrations included detailed descriptions of an individual registrant, including his or her name (including any known aliases), age, sex, height, complexion, any identifying scars or other notable physical attributes, and whether he or she were born free or manumitted. Some registrations included the county of birth; few reported an occupation. Komlos and Bodenhorn have used these data to investigate various features of the free African-American experience, including race- and complexion-based differences in height and whether African-American heights exhibited evidence of the antebellum puzzle.⁶⁵

⁶⁵ Komlos, “Toward;” Bodenhorn, “Troublesome Caste;” Bodenhorn, “Mulatto Advantage.”

One concern with the registration data is that, like in the prison sample, they may not be a true random draw from either the free-born or enslaved population. Virginia's 1793 "black code" required all free and manumitted African Americans to register with the court clerk of the county in which they resided.⁶⁶ Any free person who failed to do so was subject to arrest and liable for the jailor's fees incurred before they appeared in court and registered, which might be expected to have encouraged near universal registration because the law was enforced, even if unevenly.⁶⁷ But only a fraction of African Americans actually complied. In Campbell County, Virginia, for example, the clerk registered only 287 individuals between 1801 and 1850 even though the 1850 census enumerated 846 African-American residents. An effort to match registers (circa 1850 and 1860) to the 1850 and 1860 manuscript censuses for 17 Maryland and Virginia counties resulted in a 27 percent match rate, which may be only partly explained by the use of aliases, alternative spellings, and failure to match on common names. Registration was selected on something; we investigate whether it the selection was height-related.

A second feature of Virginia's 1793 act, namely that any employer who hired a free person of color without a proper registration was subject to a \$5 fine per violation, may have led to selective registration. Most free-born registrants appear in the records between the ages of 17 and 25, probably as a result of the \$5 fine-law. These are the ages at which young men and women left home and/or entered the wage labor market and employers required papers. If historical labor markets exhibit the same rewards to height observed in modern studies, African-American employment opportunities may have been correlated with height. Moreover, if the returns to height changed with changes in employment rates or wages (or other macroeconomic variables), dynamic selection may have operated in a way that improvements in labor market opportunities would have drawn differentially over cohorts and time from an otherwise stable height distribution. If this were the case, the selected data may reveal a (spurious) trend in height that would not appear in a representative random sample.

⁶⁶ Hening, *New Virginia Justice*, 546. Maryland lawmakers debated, but did not enact, a similar law in 1830 and 1831. See Wright, *Free Negro*, 269. Even without the employer law, there were strong incentives for Maryland's free blacks to register as they expected to be out and about between home and work. Having a copy of one's freedom papers would have reduced a free person being mistaken for and detained as a runaway slave.

⁶⁷ See, for example, Rothman, *Notorious in the Neighborhood*, ch. 3.

Further, because it includes only manumitted slaves, it is unlikely that the Maryland-Virginia registration sample includes a random draw of the slave population.⁶⁸ Historical studies into manumission practices across the New World point toward non-random, or selective manumission. Manumitted slaves were disproportionately young, female, mixed-race, skilled workers or domestic servants that resided in urban places.⁶⁹ Moreover, Shawn Cole found that slaves that achieved their freedom through self-purchase paid a 19 percent premium over market price adjusted for observable characteristics, which suggests that these slaves expected to realize productivities in freedom sufficient to compensate for the higher prices paid.⁷⁰ Arthur Budros, too, found a correlation between manumission rates in south-central Virginia and changes in slave and commodity prices.⁷¹ One feature of manumission that is widely accepted is that it was used selectively to reward favored slaves and to provide incentives for slaves to behave and work hard.⁷² If the selection process into manumission responded to either short-run cycles or long-run trends in the southern economy, the potential for dynamic selection is evident.

The selection-diagnostic regressions we estimate, which includes registration-year and birth-year dummies, takes the same general form as in Equation (2) above. We follow the same procedures in our estimations here as above. We trim the sample to include heights within three standard deviations of mean height (59 to 75 inches for men and 55 to 70 inches for women) and drop observation year- birth year cells with less than five observations. We include only men between 22 and 50 years and women 20 to 50 years, or those likely to have attained terminal adult height. After trimming, we are left with more than 4,000 observations in each sex-status subsample. The sample includes birth years between 1752 and 1843 and registration years between 1800 and 1864, so it is not clear whether it will have much to say about the puzzle, the onset of which is typically dated for birth years circa 1840. For all subsamples, the OLS constant is the estimated height of 25 year olds born in 1800 and measured in 1825.

Table 2 presents summaries of our tests for four subsamples taken from the Maryland-Virginia free black data. Three of the four F-tests on registration year, which is our proxy for unobserved macroeconomic factors, are statistically significant at p-values of 0.05 or less.

⁶⁸ Maryland data from Komlos, *Heights of African Americans*; Virginia data from Bodenhorn, “Mulatto Advantage,” and additional counties.

⁶⁹ Bodenhorn, “Manumission,” 146-147 and sources discussed therein.

⁷⁰ Cole, “Capitalism and Freedom.”

⁷¹ Budros, “Social Shocks.”

⁷² Whitman, *Price of Freedom*; Wolf, *Race and Liberty*.

Because these tests indicate some type of dynamic selection related to height, it is impossible to use these height data to draw valid inferences about trends in height (and, by implication, well-being) without an explicit model of height-based selection.⁷³

5.3 Amherst College students, 1861-1900

John Murray collected information on nearly 2500 Amherst College students who matriculated between 1861 and 1900.⁷⁴ Students' heights were often recorded by Edward Hitchcock, a professor of physical education and one of the leading American anthropometrists of the nineteenth century. Students were often measured more than once during their time at Amherst and Murray's dataset includes heights measured near the date at which the student left the college. Hitchcock recorded heights in meters to the third decimal place so that data heaping on round numbers or quarter-inches as is common in many American datasets is not an issue. We convert the metric values in the dataset into inches to make them comparable to the previously analyzed groups.

Murray readily acknowledges that the Amherst sample is unlikely to be representative of the general American population. The college was populated by well-to-do young men; only a small proportion was "poor boys" who worked to finance their educations.⁷⁵ Students were overwhelmingly from substantial middle-class families that had access to more economic resources than the typical American family. One advantage of the Amherst data is that over the period under study the demographic and socioeconomic characteristics of the student body remained fairly stable so long as age and a constant proportion of students receiving financial aid are good measures of similarity over time. But, again, enrollments may have been subject to a dynamic selection process on some feature correlated with height as the North American economy evolved for birth cohorts between 1830s and the 1870s.

Table 3 provides the results of two selection tests. Murray's sample includes approximately 1,000 young men, but we trim the sample to include men between 59 and 75 inches and exclude any birth year-measurement year cell with less than two observations. The

⁷³ Again, different subsamples present conflicting different interpretations of height trends and the puzzle. Graphs of birth-year coefficients for slave-born men and women reveal modest positive trends, though t-tests of individual coefficients and joint F-tests cannot reject the null of zero coefficients. Birth-year coefficients for free-born men and women reveal modest inverted U-shaped patterns that reach peaks in the 1820s.

⁷⁴ Murray, "Standards of the Present."

⁷⁵ Murray, "Standards of the Present," 591.

constant from the OLS regression is an estimate of the mean height of 23 year old Amherst students born in 1862 and measured in 1885. Column 1 trims the student sample to Amherst men between 22 and 25 years of age, or those who had likely achieved their adult stature and exited college at a fairly typical age. The F-test on the birth-year coefficients reveal strong evidence of selection on height. Column 2 adds 134 students who exited college between their 26th and 29th years. This slightly larger sample, too, shows evidence of height-based selection.

An interesting feature of Amherst student heights is that the birth-year coefficients on the 22-25 year subsample reveal a marked upward trend. Birth-year coefficients on the expanded 22-29 year sample reveal the U-shaped pattern commonly discussed in the puzzle literature. It is not clear how to interpret these results because it is likely that the selection process that resulted in graduates in their mid- to late-twenties differs from the process that generates students that graduate before age 26. It should come as no surprise that nineteenth-century college students, especially students willing to travel across the continent, were selected on some feature correlated with height.

BGM develop a simple test for dynamic selection in historical heights samples that can identify whether a sample is a nonrandomly selected on heights drawn from an underlying population of interest. In this section, we have implemented their test on three samples that have been discussed elsewhere in the historical heights literature, and are broadly representative of the types of samples studied in the literature, namely, prisoners, free African Americans and slaves, and students. In nearly every case, we find evidence of height-based selection into the sample. The one exception is manumitted males slaves. But research into manumission (discussed above) suggests that manumission was selective, but our tests do not show that the selection mechanism was correlated with height. Prisoners, free blacks and students did select into their respective samples based on height or with some characteristic correlated with height. Moreover, a model of dynamic selection suggests that selection can change over time, which makes valid inference problematic when the selection process cannot be modeled and estimated. Caveat lector seems an appropriate admonition for consumers of the historical heights literature. Trends in true heights may not have followed the time-series patterns that have emerged in the literature.

6. Conclusions

As we noted at the outset, anthropometrics is now a standard implement in the economic historian's toolkit. Books and articles that address the "industrialization puzzle" have been cited thousands of times. The puzzle, as Komlos characterizes it, is now a widely accepted stylized fact and one that has revolutionized how economists and economic historians think about the early stages of industrialization. We are less ready to accept the puzzle as a fact. Much of the evidence in support of the puzzle is drawn from sources where selection into the sample, on both observed and unobserved characteristics is likely – men who volunteered for military service, prisoners, manumitted slaves, and students – and it is possible that dynamic selection in response to changing macroeconomic conditions may be the driving force behind the observed decline in heights during early industrialization. Thus, what the cliometrician observes (declining heights) may not reflect what is actually happening in the economy. In fact, most historical heights studies that use conscripts (rather than volunteer soldiers) or other representative, random samples of individual height reveal little evidence of any notable height reversal.⁷⁶ Across Europe and the Americas and across two centuries, representative samples consistently reveal long-run increases in mean heights without much interruption.

Our review of the historical heights literature reveals an underappreciation for the potential consequences of dynamic, height-based selection on the interpretation of mean heights and their changes. Our paper offers a selection diagnostic that is easy to implement and, despite its simplicity, provides reliable evidence of selection when it is likely to be a concern. When the model is brought to bear against three sources of data common in the literature – prisoners, free African Americans, and students – it provides compelling evidence of some type of dynamic selection. We cannot, at this time, provide a solution to the selection problem. Doing so demands a clear understanding of the selection process, how it might change over time, the fraction of the relevant population that selects into the sample and other information that is not typically available in standard height sources. We cannot offer a solution to the selection problem. We can only offer a warning: interpreting the results without an adequate appreciation for the possibility that the results may reflect more an unaccounted for selection process than a real change in historical heights may lead to unwarranted inferences.

⁷⁶ Weir, "Economic Welfare;" Federico, "Heights, Calories and Welfare," 291; Arcaleni, "Secular Trend," 33; Jacobs and Tassenaar, "Height, Income and Nutrition," 186; and Meisel and Vega, "Biological Standard of Living," 104.

7. References

- A'Hearn, Brian. "Anthropometric Evidence on Living Standards in Northern Italy, 1730-1860." *Journal of Economic History* 63, no.2 (2003): 351-381.
- Arcalini, Emilia. "Secular Trend and Regional Differences in the Stature of Italians, 1854-1980." *Economics & Human Biology* 4, no. 1 (2006): 24-38.
- Bodenhorn, Howard. "Troublesome Caste ." *Journal of Economic History* (1999):
- Bodenhorn, Howard. "The Mulatto Advantage: The Biological Consequences of Complexion in Rural Antebellum Virginia." *Journal of Interdisciplinary History* 33, no. 1 (2002): 21-46.
- Bodenhorn, Howard. "Manumission in Nineteenth-Century Virginia." *Cliometrica* 5, no. 2 (2011): 145-164.
- Bodenhorn, Howard, Timothy Guinnane, and Thomas Mroz. "Problems of Sample-Selection Bias in the Historical Heights Literature: A Theoretical and Econometric Analysis." Yale University Economic Growth Center Working Paper No. 1023 (2013).
- Bodenhorn, Howard, Carolyn Moehling and Gregory N. Price. "Short Criminals: Stature and Crime in Early America." *Journal of Law & Economics*
- Borcherding, Thomas E and Eugene Silberberg. "Shipping the Good Apples Out: The Alchian and Allen Theorem Reconsidered." *Journal of Political Economy* 86, no. 1 (1978): 131-138.
- Brana-Shute, Rosemary and Randy J. Sparks. *Paths to Freedom: Manumission in the Atlantic World*. Columbia: University of South Carolina Press, 2009.
- Calomiris, Charles and Jonathan B. Pritchett. "Preserving Slave Families for Profit: Traders' Incentives and Pricing in the New Orleans Slave Market." *Journal of Economic History* 69, no. 4 (2009): 986-1011.
- Carson, Scott Alan. "The Biological Standard of Living in 19th Century Mexico and the American West." *Economics and Human Biology* 3, no 3 (2005): 405-419.
- Carson, Scott Alan. "Inequality in the American South: Evidence from the Nineteenth Century Missouri State Prison." *Journal of Biosocial Science* 40, no.4 (2008): 587-604.
- Case, Anne and Christina Paxson. "Stature and Status: Height, Ability, and Labor Market Outcomes." *Journal of Political Economy* 116, no. (2008): 499-532.

- Clark, Kevin. "The Hot New Thing in the NFL: Ancient Big Ten Offenses." *Wall Street Journal* (October 1, 2013): D6.
- Coclanis, Peter A. and John Komlos. "Nutrition and Economic Development in Post-Reconstruction South Carolina." *Social Science History* 19, no.1 (1995): 91-115.
- Crawford, Stephen. 1992. "The Slave Family: A View from the Slave Narratives." In *Strategic Factors in Nineteenth-century American Economic History: A Volume to Honor Robert W. Fogel*, pp. 331-350. Chicago: University of Chicago Press.
- Eveleth, Phyllis B. and James M. Tanner. *Worldwide Variation in Human Growth*. Cambridge: Cambridge University Press, 1976.
- Federico, Giovanni. "Heights, Calories and Welfare: A New Perspective on Italian Industrialization, 1854-1913." *Economics & Human Biology* 1, no. 3 (2003): 289-308.
- Feinstein, Charles. "Pessimism Perpetuated: Real Wages and the Standard of Living in the Britain during and after the Industrial Revolution." *Journal of Economic History* 58, no.3 (1998): 625-658.
- Floud, Roderick. "Wirtschaftliche und Soziale Einflüsse auf der Körpergrößen von Eurohaern seit 1750," *Jahrbuch für Wirtschaftsgeschichte* 2 (1985): 93-118. Translated by B. Feldmann. Available as "The Heights of Europeans since 1750: A New Source for European Economic History." NBER working paper 1318 (1984).
- Floud, Roderick, Kenneth Wachter and Annabel Gregory. *Height, Health and History: Nutritional Status in the United Kingdom, 1750-1980*. Cambridge: Cambridge University Press, 1990.
- Fogel, Robert W., Stanley L. Engerman, Roderick Floud, Gerald Friedman, Robert A. Margo, Kenneth Sokoloff, Richard H. Steckel, T. James Trussell, Georgia Villaflor, and Kenneth W. Wachter. "Secular Change in American and British Stature and Nutrition." *Journal of Interdisciplinary History* 14, no. 2 (1983): 445-481.
- Frank, Zephyr. "Stature in Nineteenth-Century Rio de Janeiro: Preliminary Evidence from Prison Records." *Revista de Historia Económica: Journal of Iberian and Latin American Economic History* 24, no. 3 (2006): 465-490.
- Freudenberger, Herman and Jonathan B. Pritchett. "The Domestic United States Slave Trade: New Evidence." *Journal of Interdisciplinary History* 21, no. 3 (1991): 447-477.

- Grubb, Farley. "The End of European Immigrant Servitude in British Colonial America: An Economic Analysis of Market Collapse." *Journal of Economic History* 54, no. 4 (1994): 794-824.
- Grubb, Farley. "Lilliputians and Brobdingnagians, Stature in British Colonial America: Evidence from Servants, Convicts, and Apprentices." *Research in Economic History* 19 (1999): 139-203.
- Hening, William Waller. *The New Virginia Justice, Comparing the Office and Authority of a Justice of the Peace, in the Commonwealth of Virginia*. Richmond: Johnson & Warner, 1810.
- Higman, Barry W. "Growth in Afro-Caribbean Slave Populations." *American Journal of Physical Anthropology* 50, no. 3 (1979): 373-385.
- Higman, Barry W. *Slave Populations of the British Caribbean, 1807-1834*. Baltimore: Johns Hopkins University Press, 1984.
- Jacobs, Jan and Vincent Tassenaar. "Height, Income, and Nutrition in the Netherlands; The Second Half of the Nineteenth Century." *Economics & Human Biology* 2, no. 2 (2004): 181-193.
- Johnson, Paul and Stephen Nicholas. "Health and Welfare of Women in the United Kingdom, 1785-1920." In *Health and Welfare during Industrialization*, 201-250. Edited by Richard H. Steckel and Roderick Floud. Chicago: University of Chicago Press, 1997.
- Kiple, Kenneth F. and Virginia H. Kiple. "Slave Child Mortality: Some Nutritional Answers to a Perennial Puzzle." *Journal of Southern History* 10, no. 3 (Spring 1977): 284-309.
- Komlos, John. "The Height and Weight of West Point cadets: Dietary Change in Antebellum America." *Journal of Economic History* 47, no.4 (1987): 987-927.
- Komlos, John. *Nutrition and Economic Development in the Eighteenth-Century Hapsburg Monarchy: An Anthropometric History*. Princeton: Princeton University Press, 1989.
- Komlos, John. "Toward an Anthropometric History of African Americans: The Case of the Manumitted Slaves of Maryland." In *Strategic Factors in Nineteenth-Century American Economic History: A Volume to Honor Robert W. Fogel*, 279-329. Edited by Claudia Goldin and Hugh Rockoff. Chicago: University of Chicago Press, 1990.
- Komlos, John. "The Secular Trend in the Biological Standard of Living in the United Kingdom, 1730-1860." *Economic History Review* 46, no.1 (1993): 115-144.

- Komlos, John. *Height of Free African Americans in Maryland, 1800-1864*. ICPSR 03422. Ann Arbor, MI: Inter-university Consortium for Political and Social Research (2002). doi:10.3886/ICPSR03422.v1
- Komlos, John. "A Three-Decade 'Kuhnian' History of the Antebellum Puzzle: Explaining the Shrinking of the US Population at the Onset of Modern Economic Growth." Unpublished working paper (2013).
- Komlos, John and Peter Coclanis. "On the Puzzling Cycle in the Biological Standard of Living: The Case of Antebellum Georgia." *Explorations in Economic History* 34, no. x (1997): 433-459.
- Komlos, John and Helmut Küchenhoff. "The Diminution of the Physical Stature of the English Male Population in the Eighteenth Century." *Cliometrica* 6, no. 1 (2012): 45-62.
- Lewis, Frank D. "The Transition from Slavery to Freedom through Manumission: A Life-Cycle Approach Applied to the United States and Guadeloupe." In *Slavery and the Development of the Americas, 150-177*. Edited by David Eltis, Frank D. Lewis, Kenneth L. Sokoloff. New York: Cambridge University Press (2004).
- López-Alonso, Moramay and Raúl Porrás Condey, "The Ups and Downs of Mexican Economic Growth: The Biological Standard of Living and Inequality, 1870-1950." *Economics and Human Biology* 1, no. 2 (2003): 169-186.
- Maloney, Thomas N. and Scott Alan Carson. "Living Standards in Black and White: Evidence from the Heights of Ohio Prison Inmates, 1829-1913." *Economics and Human Biology* x, no. x (2008):
- Margo, Robert A. and Richard H. Steckel. "The Heights of American Slaves." *Social Science History* 6, no. 4 (1982): 516-538.
- Meier, Paul. "Estimating Historical Heights: Comment." *Journal of the American Statistical Association* 77, no. 378 (1982): 296-207.
- Meisel, Adolfo and Margarita Vega. "The Biological Standard of Living (and Its Convergence) in Columbia, 1870-2003: A Tropical Success Story." *Economics and Human Biology* 5, no. 1: (2007): 100-122.
- Moehling, Carolyn M. and Anne Morrison Piehl. "Immigration, Crime, and Incarceration in Early 20th Century America." *Demography* 46, no. 4 (2009): 739-763.

- Mokyr, Joel and Cormac O'Gráda. "The Heights of the British and the Irish c. 1800-1815: Evidence from Recruits to the East India Company's Army." In *Stature, Living Standards, and Economic Development: Essays in Anthropometric History*, 39-59. Edited by John Komlos. Chicago and London: University of Chicago Press, 1994.
- Mokyr, Joel and Cormac O'Gráda. "Height and Health in the United Kingdom 1815-1860: Evidence from the East India Company Army." *Explorations in Economic History* 33, no. 2 (1996): 141-168.
- Morgan, Stephen L. "Economic Growth and the Biological Standard of Living in China, 1880-1930." *Economics and Human Biology* 2, no. 2 (2004): 197-218.
- Murray, John E. "Standards of the Present for People of the Past: Height, Weight, and Mortality among Men of Amherst College, 1834-1949." *Journal of Economic History* 57, no. 3 (1997): 585-606.
- Nicholas, Stephen and Deborah Oxley. "Living Standards of Women in England and Wales, 1785-1815: Evidence from Newgate Prison Records." *Economic History Review* 49, no. 3 (1996): 591-599.
- Nicholas, Stephen and Richard H. Steckel. "Heights and Living Standards of English Workers during the Early Years of Industrialization, 1770-1815." *Journal of Economic History* 51, no. 4 (1991), 937-957.
- Nicholas, Stephen and Richard H. Steckel. "Tall but Poor: Living Standards of Men and Women in Pre-famine Ireland." *Journal of European Economic History* 26, no. 1 (1997): 105-134.
- O'Gráda, Cormac. "Anthropometric History: What's in It for Ireland?" *Historie & Mesure* 11, no. 1-2 (1996): 139-166.
- Persico, Nicola, Andrew Postlewaite, and Dan Silverman. "The Effect of Adolescent Experience on Labor Market Outcomes: The Case of Height." *Journal of Political Economy* 112, no. (2005): 1019-1053.
- Pritchett, Jonathan B. "The Interregional Slave Trade and the Selection of Slaves for the New Orleans Market." *Journal of Interdisciplinary History* 28, no. 1 (1997): 57-85.
- Pritchett, Jonathan B and Richard M. Chamberlain. "Selection in the Market for Slaves: New Orleans, 1830-1860." *Quarterly Journal of Economics* 108, no. 2 (1993): 461-473.

- Pritchett, Jonathan B and Herman Freudenberger. "A Peculiar Sample: The Selection of Slaves for the New Orleans Market." *Journal of Economic History* 52, no. 1 (1992): 109-127.
- Riggs, Paul. "The Standard of Living in Scotland, 1800-1850." In *Stature, Living Standards, and Economic Development: Essays in Anthropometric History*, 60-75. Edited by John Komlos. Chicago and London: University of Chicago Press, 1994.
- Rothman, Joshua D. *Notorious in the Neighborhood: sex and Families across the Color Line in Virginia, 1787-1861*. Chapel Hill and London: University of North Carolina Press, 2007.
- Roy, A. D. "Some Thoughts on the Distribution of Earnings." *Oxford Economic Papers* 3, no. 2 (1951): 135-146.
- Steckel, Richard H. "Slave Height Profiles from Coastwise Manifests." *Explorations in Economic History* 16, no. 4 (1979): 363-380.
- Steckel, Richard H. "A Peculiar Population: The Nutrition, Health, and Mortality of American Slaves from Childhood to Maturity." *Journal of Economic History* 46, no. 3 (1986): 721-741.
- Steckel, Richard H. "Growth Depression and Recovery: The Remarkable Case of American Slaves." *Annals of Human Biology* 14, no. 1 (1987): 111-132.
- Steckel, Richard H. "Stature and the Standard of Living." *Journal of Economic Literature* 33, no. 4 (1995): 1903-1940.
- Steckel, Richard H. "Health and Nutrition in Pre-Columbian America: The Skeletal Evidence." *Journal of Interdisciplinary History* 36(1): 1-32.
- Steckel, Richard. "Biological Measures of Economic History." *Annual Review of Economics* 5 (2013): 15.1-15.23.
- Sunder, Marco. "The Height of Tennessee Convicts: Another Piece of the 'Antebellum Puzzle.'" *Economics and Human Biology* 2, no. 1 (2004): 75-86.
- Tatarek, Nancy E. "Geographical Height Variation among Ohio Caucasian Male Convicts Born 1780-1849." *Economics and Human Biology* 4, no. 2 (2006): 222-236.
- Twrdek, Linda and Kerstin Manzel. "The Seed of Abundance and Misery: Peruvian Living Standards from the Early Republican Period to the End of the Guano Era (1820-1880)." *Economics and Human Biology* 8, no. 2 (2010): 145-152.

- Voth, Hans-Joachim and Timothy Leunig. "Did Smallpox Reduce Height? Stature and the Standard of Living in London, 1770-1873." *Economic History Review* 49, no. 3 (1996): 541-560.
- Weir, David R. "Economic Welfare and Physical Well-Being in France, 1750-1990." In *Health and Human Welfare during Industrialization*. Edited by Richard H. Steckel and Roderick Floud. Chicago: University of Chicago Press, 1997.
- Whitman, T. Stephen. *The Price of Freedom: Slavery and Manumission in Baltimore and Early National Maryland*. Lexington: University Press of Kentucky, 1997.
- Wolf, Eva Sheppard. *Race and Liberty in the New Nation: Emancipation in Virginia from the Revolution to Nat Turner's Rebellion*. Baton Rouge: Louisiana State University Press, 2006.
- Wright, James Martin. *The Free Negro in Maryland, 1634-1860*. Columbia University Studies in History, Economics and Public Law, No. 222. New York: Longmans, Green & Co., 1921.

Table 1
 Summary of OLS estimates of height using Pennsylvania prisoners
 Selection diagnostic tests

	Native-born men		
	White	Black	All
	Ages 22 - 50		
	Heights: 59 - 75 inches		
	(1)	(2)	(3)
Test all incarceration years zero	1.58	1.47	1.58
p-value of F test	(0.005)	(0.020)	(0.006)
degrees of freedom of F test	[50, 4175]	[49, 857]	[50, 5151]
Test all birth years zero	1.3	2.93	1.23
p-value of F test	(0.052)	(0.000)	(0.093)
degrees of freedom of F test	[68, 4175]	[68, 857]	[69, 5151]
Additional correlates			
Black/Mixed race	na	Yes	Yes
Observations	4294	976	5273

Note: all regressions use robust standard errors

Table 2
 Summary of OLS estimates of height, using Maryland-Virginia African-American sample
 Selection diagnostic tests

	Men		Women	
	Free-born	Slave-born	Free-born	Slave-born
	Ages 22 - 50		Ages 20 - 50	
	Heights: 59 - 75 inches		Heights: 55 - 70 inches	
	(1)	(2)	(3)	(4)
Test all registration years zero	1.44	0.72	1.35	1.32
p-value of F test	(0.016)	(0.949)	(0.039)	(0.046)
degrees of freedom of F test	[59, 5680]	[62, 4322]	[59, 6425]	[63, 4494]
Test all birth years zero	1.28	0.94	1.58	1.12
p-value of F test	(0.059)	(0.618)	(0.002)	(0.228)
degrees of freedom of F test	[68, 5680]	[78, 4322]	[68, 6425]	[80, 4494]
Additional correlates	No	No	No	No
Observations	5808	4463	6553	4638

Note: all regressions use robust standard errors

Table 3
 Summary of OLS estimates of height, using Amherst College sample
 Selection diagnostic tests

	Amherst men	
	Ages; 22 - 25	Ages: 22 - 29
	Heights: 59 - 75 inches	
	(1)	(2)
Test all measurement years zero	2.19	1.86
p-value of F test	(0.000)	(0.003)
degrees of freedom of F test	[32, 739]	[32, 870]
Test all birth years zero	2.66	1.72
p-value of F test	(0.000)	(0.003)
degrees of freedom of F test	[41, 739]	[45, 870]
Additional correlates	No	No
Observations	814	948

Note: all regressions use robust standard errors