This PDF is a selection from a published volume from the National Bureau of Economic Research

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

Volume Author/Editor: Dora L. Costa, editor

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-11618-2

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

Conference Date: February 2-3, 2001

Publication Date: January 2003

Title: The Height of Union Army Recruits. Family and Community Influences

Author: Sven Wilson, Clayne L. Pope

URL: http://www.nber.org/chapters/c9630

5

The Height of Union Army Recruits
Family and Community Influences

Sven E. Wilson and Clayne L. Pope

5.1 Introduction

In the past two decades we have seen the birth and flourishing of a new line of inquiry among economic historians who have used physical stature (primarily height) as an indicator of the welfare of historical populations.[1] This line of research has highlighted the dramatic variability in the mean height of numerous populations across both time and place, including, quite surprisingly, some periods of moderate decline in the early stages of industrialization in both Europe and the United States.

Given the nature of the available data, there has been a tendency in most studies to make a strict dichotomy between genetic and environmental factors affecting height. We suggest that as better data become available, an understanding of the factors affecting human growth and development will necessitate a richer categorization and investigation of those factors typically lumped together in the environmental category. An obvious place to begin developing an expanded classification is by making a distinction between community-level and family-level influences on health. Family background may influence child health and welfare through a number of chan-

Sven E. Wilson is assistant professor of political science and adjunct assistant professor of economics at Brigham Young University. Clayne L. Pope is professor of economics and dean of the College of Family, Home, and Social Sciences at Brigham Young University, and a research associate of the National Bureau of Economic Research.

Financial support for this research was provided by the National Institute on Aging (AG10120) and Brigham Young University. Damon Cann, Lauren Cundick, Ben Howell, and Shawn Waddoups provided excellent research assistance.

1. Several excellent volumes and survey articles covering anthropometric history as it relates to both human biology and economic development have been written by leading scholars in recent years. See, for example, Steckel (1995), Floud, Wachter, and Gregory (1990), Komlos (1994), and Steckel and Floud (1997).

nels that are related to, but distinct from, community-level factors. In the case of socioeconomic variables, it is likely that the community level of economic development and the distribution of income are both important in explaining height. In modern populations, for instance, income distribution may matter less if social programs exist that augment the nutrition and health care of poor children in the community. Furthermore, parents make investments in the physical and intellectual human capital of their children that include nutrition, schooling, and labor requirements. Their investments will be a function of their own stock of human capital and the resource constraints that are present both within the household and at the community level. There are surely family-specific determinants of height that are neither genetic nor observable in aggregate statistics.

In this study we investigate family and community influences on adult height using newly collected data on Union Army recruits. Data on a sample of 35,570 Union Army recruits has recently been linked to almost 11,500 census records from the 1850 U.S. Census manuscripts. Furthermore, we have matched these census records to the county-level published data to investigate county-level economic and demographic data with the household-level variables present in the early lives of the recruits, including occupation, wealth, nativity, migration, school attendance, literacy, and family size.

Our analysis is an extension of important papers that used earlier versions of the same data source. Fogel (1986) drew attention to the similar patterns of decline in height and mortality for the antebellum period in the United States when gross domestic product (GDP) and GDP per capita were growing vigorously. He drew on the analysis of Margo and Steckel (1983), which used the same Union Army sample before it had been linked to the census manuscripts, and on unpublished work by Steckel using the National Guard.[2] From the enlistment records they were able to identify the recruit's height, occupation, and place of residence, but they lacked the early-life influences present in the census data. Costa (1993) was able to link the recruits to the census records, but her sample was much smaller in size and was limited almost exclusively to the states of New York and Ohio.

In addition to providing new estimates for family and community influences on adult height using ordinary least squares (OLS), we complement the analysis by employing a variety of alternative specifications. We explore potential differences between urban and rural populations and across occupational categories. We also estimate the model using county-level fixed effects and compare these results to the OLS equations. Finally, we estimate our main specification using quantile regression, a technique that has received increasing recognition over the past decade as a method of exploring the impact of covariates on the entire conditional distribution of the de-

2. Later published in Steckel and Haurin (1994).

pendent variable (in our case, height), rather than simply estimating the conditional mean, which is what OLS achieves.

This paper follows a somewhat different path than a similar analysis by Haines, Craig, and Weiss (2000). They use recruit information combined with previously unexploited county-level variables (such as aggregate calorie production and access to rail and waterways), but lack household-level data from the census. An important distinction, however, is that even though we use many of the same county variables, Haines, Craig, and Weiss use the 1850 aggregates from the county of birth, whereas we use the 1850 aggregates from the county in which the recruits were living in 1850. In short, we extend the county-level characteristics employed by Haines, Craig, and Weiss by emphasizing the early life household information from the 1850 census manuscripts.

5.2 Background

5.2.1 Height and History

Height is now widely accepted as a net measure of the cumulative nutritional history of an individual. The consensus is that while variation *within* a population is largely determined by genetics, variation *between* populations is seen as primarily determined by environmental factors. Height is a useful mirror of a society's well-being because it is thought to capture both the nutritional inputs available for growth and the energy demands upon those nutrients.

Heights were first used systematically by economic historians to gather evidence on the standard of living of slaves in the antebellum South (Steckel 1979; Margo and Steckel 1982; Fogel 1989). The use of heights gathered from coastwise manifests of ships transporting slaves and from records of ex-slave recruits into the Union Army were useful as evidence of the net nutritional status of slaves and, hence, their standard of living. This use of heights in the study of slavery quickly led to interest in using military records, which routinely recorded height to study trends in the standard of living of the broader population. When systematic trends in height were gathered for both Europe and the United States, the most striking findings were the declines in height during the early periods of industrialization. Margo and Steckel (1983) and Fogel (1986) find a downturn in heights for the United States in the antebellum period of industrialization, while Floud, Wachter, and Gregory (1990) find that height grew in early industrialization but declined in the mid-nineteenth century. Komlos (1994) found similar results on the continent of Europe. The decline in heights in the United States during the antebellum period is particularly puzzling because of the impressive growth in GDP and GDP per capita during that period (Gallman and Wallis 1992). Normally, one would expect decreases in height to be associated with increases in mortality and morbidity and with declines in productivity, as argued by Fogel (1994). Thus, this antebellum period of declining heights and increasing economic growth as conventionally measured is particularly anomalous.

Alternative explanations for the antebellum decline in heights have been suggested. Fogel (1986) emphasizes immigration and its effect on the infectious disease environment. Steckel (1995) suggests a number of possible explanations, including migration and urbanization. He also argues that an increase in the inequality of per capita income could offset the growth in income. In addition to citing changes in the distribution of income, Komlos (1998) suggests that sectoral shifts out of agriculture may have influenced height through a variety of mechanisms, including increases in the relative price of food, more cyclicality in income, and larger numbers of consumers per farmer. Costa and Steckel (1997) emphasize deterioration in height because of greater exposure to infectious disease due to international immigration, urbanization, and changes in workplace environment.[3] Finally, Margo (2000) suggests that short-term declines in real wages may have contributed to some of the decline in height, although this cannot be a dominant part of the explanation since real wages were generally rising, he notes, from 1820 to 1850.

5.2.2 The Impact of Environmental Variables

The dramatic secular increase in height over the past 150 years around the globe is evidence that environmental variables in general, and economic modernization in particular, are the driving forces behind improvements in health. Fogel (2000) has written extensively on what he terms the "technophysio evolution" that has occurred over recent centuries. This evolution is likely still proceeding, since socioeconomic differentials in height (as well as all other common measures of health) persist today, both across and within countries. Class and income differences are more pronounced within poorer countries than in the developed world, but they exist even in the United States and Western Europe (Eveleth and Tanner 1990).

In both modern and historical times, differentials in child health are driven both by the distribution of material resources available to individual families and by community-level factors. Many variables thought to affect health have both family and community counterparts. For instance, the occupation of individuals may matter because it affects wages, living conditions, and the extent of interaction with other people and, therefore, the family's nutritional intake and their infectious disease risk. But the same

3. Komlos argues that the height decline could not be exclusively caused by deterioration in the disease environment because there were periods where physical stature rose as population density, urbanization, and commercialization also increased (which would likely have worsened the disease environment), he concludes that the decline in heights could not be caused exclusively by a deterioration in the disease environment.

forces are at work at the community level. Growth in manufacturing, for instance, may raise wages, but it may also concentrate individuals in a fashion that facilitates the spread of disease—even to those who are not directly involved in manufacturing. Similarly, an individual's place of birth (particularly the nation of birth) may be a strong indicator of the childhood nutritional and disease environment, but living where there is a high concentration of foreign-born may have health consequences as well, even for the native-born. In short, a family's health risks, particularly in terms of infectious diseases, are a function of both the factors within the family and the larger disease environment.

5.3 Data

5.3.1 The Union Army Data Collection

The process of sample creation consisted of selection of Civil War companies, attachment of military, pension, and medical records to recruits in those companies, and finally, linkage of household entries from the census manuscripts for the recruit and his family in the censuses of 1850, 1860, 1900, and 1910. In 1850, and probably in 1860, the recruit was likely to be a son in the household of his parents. The sample consists of a random selection of 301 companies, which were part of the U.S. Army during the Civil War, excluding cavalry and companies composed of regular army professionals. The random sample of companies was chosen, in place of a random sample of individuals, to reduce the costs of collection to a manageable level. The collection of companies achieved good geographical balance across the Northern states. Analysis to date suggests that the base sample is a random sample of recruits into the U.S. Army during the Civil War. The most obvious differences between the Civil War recruit sample and the population of those same birth cohorts are the absence of women and African American men and severe underrepresentation of Southern-born men from the sample. Consequently, results should be interpreted as being confined to Northern, white males born between 1815 and 1847.

The analysis conducted here requires that the military and pension records of the recruit be linked to the census manuscript of his pre–Civil War family. Thus, biases may be introduced if the probability of linkage is correlated with the variables under consideration here. For example, it is possible that families who migrate will be more difficult to find in the census because of moves between 1850 and year of enlistment. Or it is possible that it is easier to link to households living in rural or farming counties than to households in larger urban areas. Such biases will imply that the rates or prevalence of various characteristics cannot be accurately measured from the sample. But the statistical purposes of this paper are served if we have sufficient observations to have the full range of characteristics, even if some characteristics, such as rural residence, are overrepresented in the sample relative to the population.

In general, linkage rates from the pension record to the antebellum censuses do exhibit some systematic variation, although rates do not vary widely by socioeconomic characteristics. Clearly, rural households are easier to link than urban households, with households in larger cities being the most difficult to link. However, census indices reduce the disparity between linkage rates for urban and rural households. Foreign-born recruits are less likely to be found in the census, especially the 1850 census. Rates of immigration were high between 1845 and 1857. Those migrating after 1850 are obviously not going to be linked to the census. Even the children of the foreign-born will have lower linkage rates because of the increased tendency for the foreign-born to migrate within the United States as they searched for economic opportunities. One would also expect a slightly lower rate of linkage for native households in migration near the census year since the process of movement was time consuming in the nineteenth century. Large families are easier to find in the census than small families because large families are less likely to be living in multiple-family households. There is no evidence that wealthier households are easier to link although the teenage children of poorer households may be more likely to have left home to work and supplement the income of their parents.

This variation in linkage rates will bias the mean height of the sample of recruits linked to the census above the mean height of the recruit sample as a whole. This bias is primarily due to the lower linkage rates for urban and foreign-born recruits. However, this variation in linkage rates does not necessarily bias the analysis of what follows. Here, the interest is in the effect of socioeconomic characteristics both at the household- and county-level on height. Hence, bias in estimates would require that linkage rates for households with a given characteristic were correlated with height (e.g., tall recruits of farm households are more likely to be linked to the census than short recruits, thereby overstating the effect of a farming occupation on height). Such biases could be conjectured, but they have not been investigated to date.

5.3.2 The Height Variable

An attractive feature of the Union Army data is that height is collected at the time of enlistment for 98.8 percent of the recruits in the sample.[4] Other heights are sometimes available in the data, most of which come either from a second enlistment date or from occasional comments in pension files. Although it would be possible to incorporate height from later observations to get a better measure of maximum adult height, this would introduce a

4. This figure is after a handful of extreme heights (those less than 48 or greater than 84 inches) were excluded. The range of heights in the sample actually used in the analysis, after linking to the Census records and applying age restrictions, is 52 to 81 inches. It is likely that almost all the extreme heights were inputting errors. Many of the shorter ones are even too small for young children who may have served.

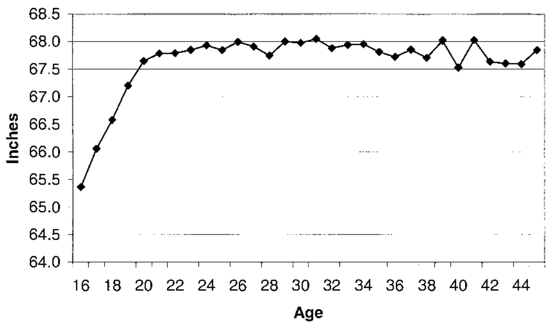

Fig. 5.1 Mean height, by age (16–45)

bias where those who enter into the pension system are systematically taller than those who have no pension files. We, therefore, restrict our attention to the first recorded enlistment height.

Enlistment height does, however, pose an important challenge to the analysis in that a significant portion of the recruits enlist at ages during which males are still growing. It is widely known that males can continue growing well into their twenties. Figure 5.1 shows mean height by age for the recruits in the sample aged sixteen to forty-nine. This graph shows relatively steep growth between ages sixteen to twenty-one in the Union Army sample. Note, however, that this figure incorporates the impact of both age and birth cohort.

Some studies have addressed this problem by restricting their analysis to older ages. For instance, Margo and Steckel (1983) drop cases below age twenty-five, and Costa (1993) looks only at those age twenty-three and older. We, on the other hand, restrict our sample to those over age sixteen and attempt to control for the significant growth between sixteen and twenty-one with age-specific dummy variables. A potential downfall of our approach is that the probability of growing significantly in late adolescence may not be random, but may be correlated with the other variables in the model, such as socioeconomic status. If this is the case, then a portion of the impact of other correlates will be captured by the age variables.

The primary reason for including the younger ages is that we are able to greatly increase sample size. Height is a variable that is primarily due to genetic variation. Furthermore, there is a likelihood of serious measurement error due to problems such as clumping and rounding. Since, even under the best of circumstances, environmental variables are going to explain only a small portion of the variance in height, if any reasonable degree of precision in estimating is desired, large sample size is a necessity.[5] Moreover, a higher

5. Increasing the minimum age restriction to twenty-three, for instance, lowers the sample size by over 70 percent. Exploratory analysis with this restricted data confirm the expectation of extremely large standard errors on almost all covariates and, for those few variables measured more precisely, little difference with the larger sample.

age restriction introduces another potential bias. Recruits who were, for instance, age twenty in 1861 and enlisted in the army at age twenty-three in 1864 would be included in the sample with a higher minimum age, but that same twenty year old who enlists in 1861 to 1863 would be excluded. Thus a higher minimum age restriction systematically excludes younger men who enlist early in the war, while including those who wait until the end of the war to enlist. Evidence presented later suggests that those who wait until the end of the war to enlist are significantly shorter than earlier entrants.

Linkage to the census, as discussed above, also creates a set of bias issues. Figure 5.2 provides a distribution of adult heights (twenty-one to forty-nine) for the full recruit sample and for those linked to the census. The full sample (aged twenty-one to forty-nine) has an average height of 67.9 in., while the census-linked group average 68.3 in. Note that the center of the distribution is the same for both samples (68 in.), but the census-linked cases are heavy on the right, and the full sample is heavier to the left. In the census-linked sample, 5.7 percent of recruits are 54 in. or shorter, while 8.4 percent of the full sample fall in this extreme. In the upper tail, 3.8 percent of recruits in the census-linked sample are taller than 72 in., compared to 5.1 percent of the full sample.

Clearly the census-linked sample is taller than the full recruit sample. This is likely because census matching is more successful among rural, native families than among urban, migrant families, categories which are also correlated with height. However, both distributions appear normally distributed, with no obvious truncation in either sample, and figure 5.2 suggests a more or less parallel shift in the distribution of heights as a result of the census linkage. Unfortunately, methods of controlling for selection bias

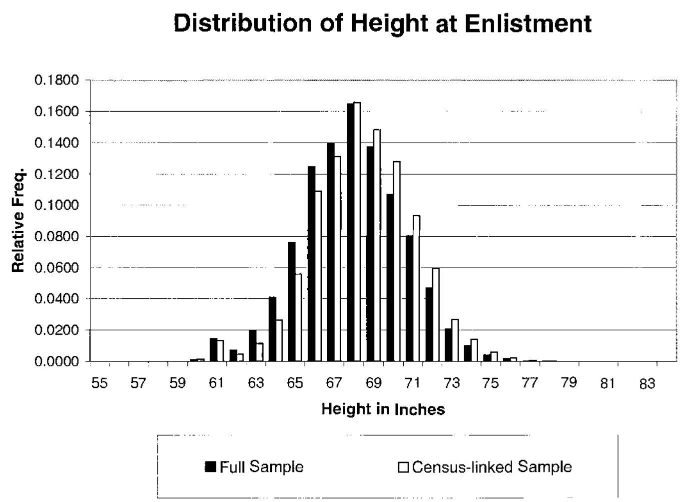

Fig. 5.2 Height distribution for full and linked samples

in this sample cannot be easily implemented, since variables that determine census linkage are also those that determine height.

5.3.3 Explanatory Variables

Linkage to the 1850 census allows the inclusion of a number of characteristics that affected the recruit's early life. These include the father's occupation, place of birth, literacy, and real estate wealth. Additionally, the place of birth of the recruit and the size of the recruit's family can be measured. Because county is known on all the census records, it is relatively simple to incorporate county-level variables from the published summary tables of the 1850 census.[6] Data available at the county level includes population, number of foreign-born, literacy rate, number of children in school, average family size, and number of deaths, as well as the detailed age distribution of the county.

At the family level, the most obvious candidates for variables that might influence height are the occupation of the family head and his real estate wealth. A problem with both these important variables is missing data. Roughly half of households report no real estate wealth. Since values of zero were not recorded on the census manuscripts, it is typically assumed that unreported cases are those where little or no wealth existed, but there is no known way to verify this assumption in individual cases. We create, therefore, five wealth classifications: (a) $1–100; (b) $101–500; (c) $501–1,000; (d) $1,001+; and (e) no reported wealth. Many occupations are, likewise, unknown. In fact, unknown is the most common occupation following farmer. We use occupation codes—provided in the data and found in the data documentation—to identify four classes of occupation: (a) farmers; (b) professionals and proprietors (codes 2 and 3); (c) artisans; and (d) laborers.

In addition to incomplete values, a limitation of this data set is that we only have real estate wealth and not personal wealth. An option would be to include the personal wealth variable in the 1860 census. While the 1860 values may prove fruitful in future analyses, we have not included them in this analysis for two reasons. First, linking the recruits to those who have nonmissing values in the 1860 census would further decrease the sample size, which is already relatively small given the nature of the question we are addressing. And second, by 1860 we are no longer talking about early life effects for most recruits, who are approaching adulthood by this point. While the 1860 wealth may be highly correlated with previous levels of wealth, a large number of recruits have already left their childhood homes by 1860, and information on their fathers is no longer available in the Union Army data.

6. We used the corrected and expanded version of ICPSR #3 created by Michael Haines (see Haines, Craig, and Weiss 2000), and we extend our appreciation to Professor Haines for graciously offering the use of the data.

Another important group of family-level variables are created from place of birth data. For each household member, the state or country of birth is almost always known, although city and county are often missing. Therefore, we classify both the recruit and his father as either native, an interstate migrant (meaning born in a different state than his 1850 residence), or foreign-born. We then interact the recruit and father's variables. This allows us to see the effect of migration holding constant the nativity of the father. It may be that interstate migration among the children of the foreign-born has different impacts on height than migration among children of native-born fathers. Of course, comparing 1850 residence to place of birth is a crude measure of internal migration, as is the absence of information on county-to-county and city-to-city moves within a state.

The final family variables we can get from the data are number of children and the literacy of parents. We expect that the effect of family size will vary by occupation. For farmers, children are an important labor input that can increase the economic well-being of the family. Fertility, consequently, is higher in farm families. Among other occupations, children are likely more of a drain on resources. Of course, with family size it is hard to determine the direction of causality, since family size (particularly in areas with high child mortality) is both a cause and a consequence of the family's economic condition. For literacy, we incorporate the mother's literacy, which is highly correlated with the father's literacy.

At the community level (and we use county as proxy for community), there are several variables that may influence height. We follow most other researchers in this area by hypothesizing that height is significantly affected by population and urbanization. There are potential positive aspects of urbanization—higher wages and closer access to agricultural markets—but we expect that the dominant effect is negative, likely because of increased exposure to infectious disease.

For our county-level variables we employ most of the variables used in Haines, Craig, and Weiss (2000). The most important is the percentage of the county that is urbanized, which is measured here as the percent of county residents living in towns of over 25,000 residents. We also employ the measure developed in Haines, Craig, and Weiss that indicates whether the county was served by rail or water transport (almost all the variation in this variable occurs, of course, in nonurbanized counties, since almost all urbanized communities had access to rail or water in 1850). Another variable strongly associated with urbanization is the percentage of the county that was foreign-born. Other important county-level variables reflect both urbanization and the economic vitality of the county, including the value of capital invested in manufacturing and agriculture; the "surplus" per capita calories produced in the county; the crude death rate; the school attendance rate; the illiteracy rate; and, finally, the average family size.

Mean height and 95 percent confidence intervals are given in table 5.1 for

Table 5.1 Mean Height, by Major Demographic Category

	N	Mean	Standard Deviation	95% Confidence Interval
All	5,758	67.74	2.66	(67.67, 67.81)
Region				
New England	858	67.67	2.66	(67.49, 67.84)
Mid-Atlantic	1,819	67.24	2.63	(67.11, 67.36)
South	530	68.26	2.79	(68.03, 68.50)
West	2,477	68.02	2.60	(67.92, 68.12)
	County-level variables			
Percent Urban				
0	3,955	67.92	2.68	(67.84, 68.01)
1–24%	1,085	67.38	2.55	(67.23, 67.53)
25–49%	429	67.54	2.60	(67.29, 67.79)
50–100%	315	66.94	2.61	(66.64, 67.23)
County population				
0–4.9K	330	68.39	2.85	(68.09, 68.70)
5–9K	608	68.05	2.65	(67.84, 68.26)
10–24K	2,155	67.93	2.63	(67.82, 68.04)
25–49K	1,620	67.68	2.61	(67.56, 67.81)
50K+	1,045	67.05	2.63	(66.89, 67.21)
Percent foreign-born				
0–10%	3,795	67.88	2.63	(67.79, 67.96)
10–24%	1,436	67.54	2.66	(67.40, 67.68)
25+%	527	67.29	2.66	(67.06, 67.52)
	Household-level variables			
Father and recruit native	2,630	67.65	2.66	(67.55, 67.75)
Father: interstate migrant				
Recruit: native	1,993	67.92	2.69	(67.80, 68.04)
Recruit: interstate migrant	522	68.02	2.63	(67.79, 68.24)
Father: foreign-born				
Recruit: native	406	67.22	2.55	(66.97, 67.46)
Recruit: interstate migrant	54	68.30	2.64	(67.57, 69.02)
Recruit: foreign-born	118	67.12	2.65	(66.64, 67.60)
Occupation				
Farmer	2,605	68.02	2.56	(67.92, 68.12)
Professional/proprietor	327	67.24	2.53	(66.96, 67.51)
Artisan	888	67.20	2.70	(67.02, 67.38)
Laborer	662	67.04	2.60	(66.84, 67.24)
Unknown	1,276	68.05	2.76	(67.90, 68.20)
Father's wealth				
$1–100	120	67.50	2.64	(67.02, 67.98)
$101–500	1,000	67.85	2.69	(67.68, 68.02)
$501–1000	931	67.96	2.65	(67.79, 68.13)
$1001+	1,516	67.94	2.60	(67.81, 68.07)
Unreported	2,191	67.47	2.68	(67.36, 67.58)
Number of children				
1–3	1,880	67.55	2.66	(67.43, 67.67)
4–6	3,185	67.77	2.64	(67.67, 67.86)
7+	693	68.13	2.76	(67.92, 68.33)

the major demographic variables discussed above. This table gives us our first look at how height varies according to the family-level and community-level variables we have identified. Although there are several significant differences between categories, we postpone a discussion of the patterns found in table 5.1 to the analysis of the OLS regression results. In brief, table 5.1 reveals the strong and significant differences for urbanization and population and for occupation, variables that prove to be important in the regression results of the next section.

Before proceeding further with the analysis, it is wise to point out two additional limitations of the variable definitions used above. First, county is a crude proxy for community. Any notion of community (and we shall not enter the quagmire of defining the term) is certainly smaller than a county. Neighborhood data would be much more desirable. Second, given the wide age range under study, the family-level and county-level variables in 1850 affect the sample individuals at very different points in the developmental process—some are in early childhood, and others are adolescents. What we know about the growth process indicates that certain time periods (particularly the first three years and the adolescent growth spurt) are particularly sensitive periods. Though unfortunate, this is an inescapable feature of using data where environmental variables were collected only at particular points in time.

An alternative to our approach would be to use information from the county of birth rather than the county of residence in 1850. Unfortunately, close to half the recruits in the Union Army sample do not have county of birth identified (although state is almost always present). We feel that linking the recruits to both county of birth and to the 1850 census reduces the sample size to such a degree that meaningful analysis is not possible and would hinder our primary objective of testing the effect of family variables on the determination of height.

5.4 Results

5.4.1 OLS Regression Results

Table 5.2 provides information relevant to the distribution of the independent variables used in the regression analysis, while table 5.3 presents our main results. In addition to estimates for the sample as a whole, we divide the sample into two parts, according to urbanization, and estimate the model for rural and urban counties separately (a county is defined as urban if it contains any communities with more than 25,000 residents, including towns that cross over county lines).[7] Table 5.4 further divides the sample

7. Margo and Steckel employed a similar technique, but they used population of the city or town, rather than county population. Their cutoff was at 2,500 persons.

Table 5.2 Means and Frequencies of Independent Variables

	Frequency (%)	Mean	Standard Deviation
Region			
New England	14.9		
Mid-Atlantic	31.6		
South	9.2		
West	44.3		
	County-level variables		
Percent urban		0.094	0.187
Agricultural capital per capita		191.434	83.080
Manufacturing capital per capita			
Surplus calories (in thousands)		2.212	1.478
Percent foreign-born		9.758	9.562
Rail or water connection		0.745	0.436
Crude death rate		12.831	5.693
School attendance rate		0.045	0.043
Illiteracy rate		0.628	0.164
Average family size		5.613	0.326
	Household-level variables		
Father and recruit native	45.7		
Father: migrant			
Recruit: native	34.6		
Recruit: interstate migrant	9.1		
Father: foreign-born			
Recruit: native	7.1		
Recruit: interstate migrant	0.9		
Recruit: foreign-born	2.0		
Occupation			
Farmer	45.2		
Professional/proprietor	5.7		
Artisan	15.4		
Laborer	11.5		
Unknown	22.2		
Father's wealth			
Excluding unreported		1759.84	3338.05
Assuming unreported = 0		1090.19	2762.63
$1–100	2.1		
$101–500	17.4		
$501–1000	16.2		
$1000+	26.3		
Unreported	38.1		
Mother literate	8.0		
Mother illiterate	92.0		
Number of children		4.39	1.78
1–3	32.7		
4–6	55.3		
7+	12.0		

(*continued*)

Table 5.2 (continued)

	Frequency (%)	Mean	Standard Deviation
	Enlistment variables		
Enlistment year			
1861	22.7		
1862	36.6		
1863	6.2		
1864	25.5		
1865	9.0		
Enlistment age		20.955	3.135
16	1.0		
17	2.5		
18	25.4		
19	13.3		
20	10.0		
21	12.0		
22	8.4		
23	6.8		
24	6.0		
25+	14.7		
N	5,758		

along a different dimension: occupation. Excluding those cases with occupation unknown, we divide the sample according to whether or not the recruit's father was a farmer in 1850. The unknown group are excluded since it is probably the case that a large share of them are farmers and inclusion of them in the nonfarmer category would mask the differences that might exist between farmers and nonfarmers. While we rely primarily on the full sample results, the estimation of the model across urban/rural and farm/nonfarm groups should be considered exploratory, since the reduced sample sizes significantly reduce the statistical power of our estimation.

We turn first to a discussion of the regional and county-level influences on height. Previously, table 5.1 showed noticeable differences in height across regions. The tallest recruits come from the South, followed in descending order by the West, New England, and the Mid-Atlantic States. More than an inch separates the South from the Mid-Atlantic.[8] The regression results, however, find smaller differences across regions, except for the Mid-Atlantic, which has significantly lower heights (by half an inch) than the other regions. The tallest recruits are from the West, which is particularly true in urban counties, indicating a substantial difference between Western and non-Western urban areas.

The regressions in table 5.3 and table 5.4 confirm that height is inversely

8. Recall that since the sample is drawn from the Union Army, the sample does not contain a representative sample of Southerners.

Table 5.3 **OLS Regressions**

	Full Sample			Rural Counties			Urban Counties		
	Mean or Frequency	Coefficient	*T*-statistic	Mean or Frequency	Coefficient	*T*-statistic	Mean or Frequency	Coefficient	*T*-statistic
Dependent variable = enlistment height (in.)	67.74			67.92			67.34		
Region									
Mid-Atlantic	.32			.26			.44		
New England	.15	.392	2.51	.10	.355	1.54	.26	.406	1.73
South	.09	.559	2.98	.12	.453	2.08	.03	.550	1.14
West	.44	.551	4.00	.52	.434	2.31	..27	.823	3.82
County-level variables (1850)									
Percent urban	.09	–.899	–2.83				.30	–.391	–.73
Rail or water connection	.75	–.281	–2.72	.63	–.295	–2.85	.99	–.313	–.48
Agricultural capital per capita	169.44	–.002	–2.57	163.87	–.002	–2.62	181.50	.000	–.18
Manufacturing capital per capita	21.99	.000	.06	13.80	–.002	–.48	39.73	.002	.40
Surplus calories (in thousands)	2.21	–.023	–.62	2.52	–.010	–.26	1.55	–.085	–1.12
Percent foreign-born (× 100)	9.76	–.009	–1.59	7.32	–.006	–.85	15.04	–.015	–1.28
Crude death rate (per 1,000)	12.83	0.15	1.99	11.87	.014	1.77	14.90	.013	.70
School attendance rate (× 100)	62.91	–.003	–.72	61.77	–.003	–.73	65.38	.001	.12
Illiteracy rate (× 100)	4.13	–.036	–3.00	4.78	–.042	–3.31	2.73	.014	.34
Average family size	5.61	.596	3.14	5.66	.559	2.61	5.52	.680	1.76

(*continued*)

Table 5.3 (continued)

	Full Sample			Rural Counties			Urban Counties		
	Mean or Frequency	Coefficient	*T*-statistic	Mean or Frequency	Coefficient	*T*-statistic	Mean or Frequency	Coefficient	*T*-statistic
				Household-level variables (1850)					
Father and son native	.46			.42			.55		
Father: interstate migrant									
Recruit: native	.35	–.101	–.96	.40	–.126	–.99	.23	–.109	–.54
Recruit: interstate migrant	.09	–.255	–1.81	.11	–.446	–2.61	.05	.422	1.61
Father: foreign-born									
Recruit: native	.07	–.195	–1.40	.05	–.193	–.98	.12	–.117	–.58
Recruit: interstate migrant	.01	.193	.55	.01	–.052	–.11	.01	.585	1.08
Recruit: foreign-born	.02	–.432	–1.51	.01	–.819	–2.06	.04	–.005	–.01
Occupation									
Farmer	.45			.51			.33		
Professional/proprietor	.06	–.391	–2.94	.04	–.341	–1.75	.10	–.475	–2.47
Artisan	.15	–.438	–3.94	.12	–.494	–3.35	.23	–.434	–2.39
Laborer	.11	–.433	–3.47	.09	–.335	–2.02	.17	–.582	–2.86
Unknown	.22	–.066	–.56	.25	.002	.02	.16	–.247	–1.26
Father's wealth									
$1–100	.02			.02			.01		
$101–500	.17	.191	.73	.20	.217	.72	.12	.066	.13
$501–1000	.16	.318	1.25	.17	.343	1.17	.14	.195	.37
$1000+	.26	.273	1.07	.26	.346	1.18	.28	.026	.05
Unknown	.38	.179	.70	.35	.228	.78	.45	–.027	–.05
Mother illiterate	.08	–.096	–.74	.10	–.003	–.02	.04	–.636	–2.24
Number of children									
1–3	.33			.30			.38		
4–6	.55	–.017	–.21	.57	.009	.09	.52	–.073	–.57
7+	.12	.190	1.46	.13	.136	.89	.10	.280	1.11

				Enlistment variables					
Enlistment year									
1861	.23			.22			.25		
1862	.37	.040	.37	.37	.039	.29	.35	.033	.20
1863	.06	–.336	–1.95	.06	–.371	–.171	.06	–.236	–.88
1864	.26	–.319	–2.91	.24	–.382	–2.73	.28	–.241	–1.45
1865	.09	–.642	–4.82	.11	–.703	–4.37	.06	–.453	–1.89
Enlistment age									
16	.01	–3.190	–9.75	.01	–3.062	–7.68	.01	–3.268	–5.83
17	.03	–1.960	–8.09	.03	–1.861	–6.33	.02	–2.322	–5.85
18	.25	–1.535	–12.74	.26	–1.555	–10.49	.24	–1.496	–7.06
19	.13	–.795	–6.61	.13	–.752	–4.86	.15	–.869	–4.57
20	.10	–.380	–2.83	.10	–.340	–2.06	.10	–.543	–2.30
21	.12	–.181	–1.35	.12	–.115	–.71	.13	–.330	–1.39
22	.08	–.286	–1.95	.09	–.270	–1.53	.08	–.331	–1.23
23	.07	–.277	–1.83	.07	–.162	–.90	.07	–.527	–1.84
24	.06	–.141	–.86	.06	–.110	–.54	.06	–.259	–.89
25+	.15			.14			.16		
Intercept		65.7465	54.89		66.10801	48.42		64.83274	26.46
Sample size		5,758			3,939			1,819	
R^2		.119			.115			.114	

Notes: Urban counties are those with at least one town of over 25,000 (including towns that cross over county lines). See text for other variable definitions. *T*-statistics are based on robust (heteroscedasticity-consistent) standard errors clustered at the county level.

Table 5.4 **OLS Regressions, by Farming Status**

	Full Sample			Farmers			Non-Farmers		
	Mean or Frequency	Coefficient	*T*-statistic	Mean or Frequency	Coefficient	*T*-statistic	Mean or Frequency	Coefficient	*T*-statistic
Dependent variable = enlistment height (in.)	67.74			68.02			67.15		
Region									
Mid-Atlantic	.32			.26			.46		
New England	.15	.392	2.51	.16	.560	2.63	.21	.143	.62
South	.09	.559	2.98	.03	.849	2.47	.02	.259	.47
West	.44	.551	4.00	.56	.406	2.15	.31	.766	3.60
County-level variables (1850)									
Percent urban	.09	–.899	–2.83	.05	–.705	–1.37	.17	–.175	–.39
Rail or water connection	.75	–.281	–2.72	.71	–.305	–2.22	.88	–.328	–1.42
Agricultural capital per capita	169.44	–.002	–2.57	167.54	–.002	–2.00	180.80	.000	.27
Manufacturing capital per capita	21.99	.000	.06	17.54	.001	.27	31.94	.000	.04
Surplus calories (in thousands)	2.21	–.023	–.62	2.40	.009	.22	1.78	–.153	–1.66
Percent foreign-born (× 100)	9.76	–.009	–1.59	9.31	–.009	–1.20	12.33	–.015	–1.70
Crude death rate (per 1,000)	12.83	0.15	1.99	12.62	.016	2.23	13.36	–.013	–.83
School attendance rate (× 100)	62.91	–.003	–.72	66.68	–.005	–.90	65.95	–.003	–.38
Illiteracy rate (× 100)	4.13	–.036	–3.00	3.78	–.044	–2.72	3.20	–.026	–.97
Average family size	5.61	.596	3.14	5.60	.482	1.72	5.54	.602	2.08
Household-level variables (1850)									
Father and son native	.46			.42			.52		
Father: interstate migrant									
Recruit: native	.35	–.101	–.96	.41	.025	.18	.25	–.232	–1.26
Recruit: interstate migrant	.09	–.255	–1.81	.10	–.139	–.72	.07	–.045	–.19
Father: foreign-born									
Recruit: native	.07	–.195	–1.40	.05	.126	.62	.11	–.525	–2.68
Recruit: interstate migrant	.01	.193	.55	.01	.368	.79	.01	1.144	1.73
Recruit: foreign-born	.02	–.432	–1.51	.01	–.832	–1.79	.04	–.283	–.75
Occupation									
Farmer	.45								
Professional/proprietor	.06	–.391	–2.94				.17	.026	.16

Artisan	.15	–.438	–3.94				.47	–.049	–.33
Laborer	.11	–.433	–3.47				.35		
Unknown	.22	–.066	–.56						
Father's wealth									
$1–100	.02			.02			.02		
$101–500	.17	.191	.73	.18	–.031	–.07	.16	.655	1.51
$501–1000	.16	.318	1.25	.18	.263	.58	.13	.497	1.13
$1000+	.26	.273	1.07	.37	.257	.59	.15	.393	.90
Unknown	.38	.179	.70	.25	.314	.70	.53	.287	.67
Mother illiterate	.08	–.096	–.74	.07	–.037	–.19	.06	–.154	–.59
Number of children									
1–3	.33			.31			.39		
4–6	.55	–.017	–.21	.56	–.033	–.28	.52	–.002	–.02
7+	.12	.190	1.46	.13	.279	1.64	.08	–.082	–.30
				Enlistment variables					
Enlistment year									
1861	.23			.19			.25		
1862	.37	.040	.37	.40	–.156	–.95	.34	.123	.71
1863	.06	–.336	–1.95	.06	–1.012	–4.18	.06	–.015	–.05
1864	.26	–.319	–2.91	.24	–.526	–3.18	.27	–.342	–1.83
1865	.09	–.642	–4.82	.11	–.703	–3.47	.09	–.737	–3.24
Enlistment age									
16	.01	–3.190	–9.75	.01	–2.453	–6.13	.01	–3.081	–5.82
17	.03	–1.960	–8.09	.02	–1.667	–3.84	.03	–1.841	–4.51
18	.25	–1.535	–12.74	.24	–1.296	–8.02	.28	–1.433	–7.20
19	.13	–.795	–6.61	.12	–.806	–4.22	.15	–.625	–2.93
20	.10	–.380	–2.83	.10	–.353	–1.79	.09	–.374	–1.37
21	.12	–.181	–1.35	.12	–.170	–.91	.12	–.327	–1.38
22	.08	–.286	–1.95	.09	–.444	–2.00	.08	–.032	–.12
23	.07	–.277	–1.83	.07	–.249	–1.26	.06	–.216	–.71
24	.06	–.141	–.86	.06	.003	.01	.06	–.306	–.93
25+	.15			.17			.12		
Intercept		65.831	55.95		66.5387	37.45		65.27195	34.32
Sample size		5,758			2,605			1,877	
R^2		.119			0.086			0.103	

Notes: Individuals with father's occupation unknown are excluded from both the farmer and nonfarmer groups. See text for other variable definitions. *T*-statistics are based on robust (heteroscedasticity-consistent) standard errors clustered at the county level.

related to urbanization. However, it appears that the greatest effect is between rural counties and counties with at least some urbanization. The regression in the urban group shows that little variation in height can be explained by the level of urbanization. Furthermore, table 5.4 shows that the urbanization effect is stronger among farmers than nonfarmers. All the results presented here are consistent with a farming advantage in terms of height, but closer examination reveals that farmers are further advantaged by living in rural locations.

Other county-level variables are also important. Access to transportation (primarily an issue in rural counties) has a significant negative effect. The aggregate level of capital is significant for agricultural capital, but negligible for manufacturing capital. Furthermore, the effect of capital in agriculture is concentrated almost solely in rural counties. Thus even though agricultural capital may raise farm output, it is negatively associated with height in rural areas. We speculate that this is because agricultural capital is indicative of access to centers of trade and commerce, which would also raise exposure to infectious disease, though we have not tested this argument directly.

The percent foreign-born is also a negative indicator (although not statistically significant), and, surprisingly, the illiteracy rate of the county has a statistically significant negative impact, though the magnitude of the effect is small and only exists in rural counties. Finally, average family size in the county turns out to be an important predictor of the variation in height (a 1 standard deviation increase in average family size raises height by 0.2 in.).

It should be noted that we find a few sharp differences between our results and those of Haines, Craig, and Weiss (2000). They find that surplus calories and the crude death rate are both negative, while we find essentially no effect of surplus calories and a positive correlation with the crude death rate. We also try their measure of protein and find similar results. Furthermore, we have not found that deleting other variables from the model changes these effects. The obvious difference between our two approaches is that we are looking at the county of residence in 1850 and they are looking at the 1850 values for the county of birth, but we have not reconciled these contrary findings.

Turning to family-level variables, clearly the most important and robust finding is the impact of occupation, almost all of which appears to be a farmer advantage. Professionals and proprietors do a little better than artisans and laborers, but the difference is not statistically significant. Furthermore, the farmer advantage exists in both rural and urban counties and, in fact, is slightly higher in the urban counties. This effect must be considered in association with the earlier result that population had greater effects among farmers than among nonfarmers. Taken together they imply that the farmer advantage is augmented by living in a rural setting, but significant advantages for farmers exist in urban counties as well.

The impact of father's wealth is somewhat ambiguous. The coefficient estimates in the full sample show that wealth in the $101 to $500 category increases height by 0.191 in. (relative to $1 to $100) and by 0.318 in. in the $501 to $1,000 range, with no further increases for the highest wealth category (0.273 in.). These magnitudes are nontrivial, but the estimates are not statistically significant. Interestingly, the estimates of wealth effects are much higher for nonfarmers than farmers, but they are virtually nonexistent for residents of urban counties in general. Because the effect of wealth may be nonlinear, occurring primarily at very low values, and because of our relatively small sample, we have found that it is infeasible to test for further interactions of wealth with other covariates, either through interacting wealth directly with other variables or by estimating the sample on selected subsets of the overall sample.

The migratory history of recruits and their parents is associated with very large differences in height, though individual coefficients tend to fall short of statistical significance at the 0.05 level. Overall, foreign-born recruits are in the worst situation. Foreign-born recruits in rural counties are 0.819 in. shorter than their native counterparts, and, similarly, the foreign-born who are farmers are 0.823 in. shorter than native-born farmers (note that these statistically significant effects are some of the largest presented here and are substantially larger than, for instance, the marginal effects of region, urbanization, or occupation). Another striking pattern is that effect of recruits who migrate across state lines. Movers who are the sons of native-born are somewhat shorter than nonmovers (though movers into urban counties are taller than natives of urban counties). However, movers whose fathers are foreign-born are taller than the nonmigrating recruits. This is especially true among the urban group and nonfarmers. In all cases there is a clear distinction between the sons of foreign-born who have moved and those who have not. The extreme case is among the nonfarming population: Among the sons of foreign-born fathers in this group, recruits who have moved are 1.67 in. taller than nonmovers. While moving can be traumatic, it is likely the case that movers are systematically more robust than nonmovers, especially among immigrants, since it was difficult and costly to move inland from the port cities to which they often settle upon first arriving in America.

Finally, the mother's literacy has little impact on height for the sample as a whole. A notable exception to this is the large and statistically significant impact of mother's illiteracy among those in urbanized counties. Indeed, among urban residents, the mother's literacy may be a more important marker of socioeconomic status than either the father's real estate wealth or occupation, since we do not have a measure in 1850 of personal wealth. Finally, recruits from households with seven or more children in 1850 were 0.19 in. taller than others, though the effect is predominantly concentrated within the farming group.

Finally, controls for age and year of enlistment are included. The age variables indicate that young men of this cohort tended to grow about three inches between age sixteen and age twenty-five, which, incidentally, is significantly higher than the rate of late growth in modern populations of males (U.S. Department of Health and Human Services 1987, 23), suggesting that "catch-up" growth is occurring for part of the recruit population. We also find that enlistment year is an important indicator of height, particularly in the rural counties. Recruits enlisting in 1863 to 1864 were about 0.33 in. shorter than earlier enlistees, whereas those enlisting in 1865 were 0.64 in. shorter. Apparently, early enlistees were healthier than later ones.[9]

5.4.2 Fixed Effects

An alternative way to control for county-level effects is to allow each county to have an unrestricted impact on height through the specification of a fixed effects model. If we index individuals by i and counties by j, we can specify the following model:

$$H_{ij} = a_j + B(\boldsymbol{X_{ij}}) + u_{ij}$$

Here H_{ij} is the height of the individual, $\boldsymbol{X_{ij}}$ is the vector of regressors, and a_j is the county-specific error term. The error terms are assumed to be independent across individuals and counties, as well as uncorrelated with the $\boldsymbol{X_{ij}}$. In this analysis there are 726 counties. The minimum number of persons per county is one, and the maximum is seventy-six. There are on average 7.9 recruits per county. Overall, the county-level fixed effects explain 32.5 percent of the variation in height, though not too much should be made of this result, given the low number of recruits per county.[10]

Table 5.5 repeats the earlier OLS results for the full model and compares them with a model with no county variables and one with county fixed effects. The primary differences between the regressions with no county variables and the "full model" are that the occupational differences are stronger without the county variables (more of a farmer advantage) as are the migration effects. These differences indicate that the occupation differences in height depend on the characteristics of the county. Likewise, moving across state lines would appear to have a strong positive effect on height if the county variables are excluded, but this is not found when the charac-

9. The enlistment year effects are remarkably similar in pattern and magnitude to those found by Margo and Steckel (1983); they even find that the year effects are stronger in rural areas than in urban ones, as do we.

10. In fixed effects models, as the number of individuals per groups falls, the variance of the dependent variable explained by variation across groups will automatically rise.

Table 5.5 Fixed Effects Estimates

	Mean or Frequency	Full Model		No County Variables		Fixed Effects	
		Coefficient	*T*-statistic	Coefficient	*T*-statistic	Coefficient	*T*-statistic
Dependent variable = enlistment height (in.)	67.74						
Region							
Mid-Atlantic	.32						
New England	.15	.392	2.51				
South	.09	.559	2.98				
West	.44	.551	4.00				
County-level variables (1850)							
Percent urban	.09	–.899	–2.83				
Rail or water connection	.75	–.281	–2.72				
Agricultural capital per capita	169.44	–.002	–2.57				
Manufacturing capital per capita	21.99	.000	.06				
Surplus calories (in thousands)	2212.23	.000	–.62				
Percent foreign-born (× 100)	9.76	–.009	–1.59				
Crude death rate (per 1,000)	12.83	0.15	1.99				
School attendance rate (× 100)	62.91	–.003	–.72				
Illiteracy rate (× 100)	4.13	–.036	–3.00				
Average family size	5.61	.596	3.14				

(*continued*)

Table 5.5 (continued)

	Mean or Frequency	Full Model		No County Variables		Fixed Effects	
		Coefficient	*T*-statistic	Coefficient	*T*-statistic	Coefficient	*T*-statistic
Household-level variables (1850)							
Father and son native	.46						
Father: interstate migrant							
Recruit: native	.35	–.101	–.96	.235	2.44	–.096	–.91
Recruit: interstate migrant	.09	–.255	–1.81	.160	1.24	–.241	–1.53
Father: foreign-born							
Recruit: native	.07	–.195	–1.40	–.281	–2.22	–.166	–1.09
Recruit: interstate migrant	.01	.193	.55	.450	1.29	.503	1.32
Recruit: foreign-born	.02	–.432	–1.51	–.478	–1.74	–.332	–1.26
Occupation							
Farmer	.45						
Professional/proprietor	.06	–.391	–2.94	–.605	–4.34	–.403	–2.52
Artisan	.15	–.438	–3.94	–.667	–6.21	–.453	–4.06
Laborer	.11	–.433	–3.47	–.672	–5.41	–.413	–3.17
Unknown	.22	–.066	–.56	.061	.54	.021	.15
Father's wealth							
$1–100	.02						
$101–500	.17	.191	.73	.182	.69	.043	.16
$501–1000	.16	.318	1.25	.248	.97	.204	.75
$1000+	.26	.273	1.07	.080	.32	.132	.49
Unknown	.38	.179	.70	–.005	–.02	.071	.27
Mother illiterate	.08	–.096	–.74	.010	.08	–.087	–.60
Number of children							
1–3	.33						
4–6	.55	–.017	–.21	.055	.67	–.133	–1.68
7+	.12	.190	1.46	.300	2.27	.115	.94

Enlistment variables							
Enlistment year							
1861	.23						
1862	.37	.040	.37	–.017	–.14	–.083	–.73
1863	.06	–.336	–1.95	–.253	–1.36	–.378	–2.08
1864	.26	–.319	–2.91	–.353	–3.25	–.452	–3.72
1865	.09	–.642	–4.82	–.598	–4.41	–.692	–4.23
Enlistment age							
16	.01	–3.190	–9.75	–2.992	–9.03	–3.378	–8.66
17	.03	–1.960	–8.09	–1.802	–7.52	–1.871	–7.56
18	.25	–1.535	–12.74	–1.440	–11.87	–1.469	–12.56
19	.13	–.795	–6.61	–.748	–6.18	–.734	–5.52
20	.10	–.380	–2.83	–.339	–2.50	–.348	–2.42
21	.12	–.181	–1.35	–.157	–1.16	–.083	–.61
22	.08	–.286	–1.95	–.234	–1.59	–.303	–2.01
23	.07	–.277	–1.83	–.247	–1.60	–.297	–1.82
24	.06	–.141	–.86	–.119	–.71	–.137	–.81
25+	.15						
Intercept		65.74645	54.89	65.746	54.89	68.789	232.07
Sample size		5,758		5,758		5,758	
R^2		.119		0.095		0.081	

Notes: See text for other variable definitions. *T*-statistics are based on robust (heteroscedasticity-consistent) standard errors clustered at the county level. The *R*-squared for the fixed effects model is the "within-groups" *R*-squared, which is the regression obtained from subtracting county-level means from the dependent and independent variables. The OLS regression values are taken from table 5.3.

teristics of the county of destination are controlled for. Both of these implications make intuitive sense.

In contrast, very little difference exists within the coefficients between the OLS and the fixed effects models. An important exception is that wealth is even less important in the fixed effects model than in the OLS framework. This is possibly attributed to a relationship between individual wealth and county, which might cause some of the wealth effect to be attributed to the county-specific error terms. Additionally, the effect of interstate migration is more pronounced in the fixed effects model than in the OLS case.

5.4.3 Quantile Regressions

The final alternative specification to be explored in this analysis is quantile regression. In estimating the effects of environmental variables on height, the principal assumption is that the body is deprived of the nutrients it needs during the growth process. Implicit in the OLS model is that contributing factors have a uniform effect across the distribution of heights. In other words, if factor *X* lowers height by 1 in., then a person who would be 72 in. tall will be 71 in., a person who would otherwise be 68 in. will end up at 67 in., and so on for each height. But what would it imply if the impact of a particular growth factor has nonuniform effects across the distribution of heights?[11] It is usually assumed that the distribution of heights is normal. But in conditions where growth is suppressed for some reason, such as due to disease or to nutritional deprivation, it is not clear a priori that the effect of such deprivation will be constant across the distribution of heights.

Quantile regression is a technique that has seen increased use in the past decade as a method of estimating the effects of covariates at different points in the conditional distribution of the dependent variable.[12] Estimates are derived by minimizing the sum of the absolute deviations around a designated point in the conditional distribution. Median regression is the simplest case, but any point can be estimated, such as the first quartile (the 25th percentile) or the first decile (10th percentile), and so on. In practical terms, this amounts to solving, as shown by Koenker and Bassett (1978), the following minimization problem

$$\min(B) \sum_{y \geq XB} \theta |h_i - X_i B| + \sum_{y < XB} (1 - \theta) |h_i - X_i B|$$

where θ is the quantile to be estimated, and the other notation is the same as used previously. As can be seen from the formula above, the values of the dependent variable to the left of θ are given more weight when $\theta < 0.5$ and

11. We should be clear that we are not talking here about the possible nonlinear effects of covariates; those can be accommodated in a straightforward fashion in the OLS framework or by nonlinear least squares.

12. Important examples in recent years include the study of wage distributions (Buchinsky 1994), intergenerational earnings transmissions (Eide and Showalter 1999), and wealth and economic mobility in the nineteenth century (Conley and Galenson 1998).

those to the right are given more weight when $\theta > 0.5$. Naturally, the weight given to extreme values of h_i increases as θ approaches the extremes of 0 and 1, making the estimation of very low or high values of θ potentially sensitive to some types of measurement error or other causes of outliers. In the case of heights, it is surely the case that a portion of those veterans with a height of 5'0", 5'1", and 5'2" are, in reality, 6'0", 6'1", and 6'2", but were entered incorrectly. In general, quantile regression is more robust in these cases since the absolute deviation is being modeled, rather than the squared deviation, as is the case with OLS.

It should be noted that, assuming an uncensored distribution of heights, quantile regression is not "right" and OLS "wrong." OLS has numerous desirable properties, not the least of which are small and large sample properties that are well known. The small sample properties of quantile regression, on the other hand, are not known. But quantile regression provides a method of investigating the complete conditional distribution of heights. At the very least, it is a useful comparison of the OLS results.

Table 5.6 presents quantile estimates for $\theta = 0.1, 0.5, 0.9$. Standard errors are calculated using the formula of Koenker and Bassett (1982) and Rogers (1992). The first column presents the OLS coefficients presented earlier. Though some of the variables in the model have relatively uniform effects across the quantiles estimated in table 5.3, there are several important exceptions. For instance, the effect of urbanization is much greater in the right tail of the distribution than in the center or at the first decile. It is in the left tail of the conditional distribution, however, that other county-level variables are the most important. Indeed, only the transportation and family size variables have any impact at the 0.9 quantile, whereas all the county-level variables other than urbanization and transportation have significant effects at the 0.1 quantile. The median effects are roughly equivalent to the OLS estimates.

The effects of other variables also differ across the conditional distribution of heights. Perhaps most important is the variation in occupational effects. In the left tail we find the first evidence that occupational variation among nonfarmers may be important, since the effect of being a laborer is nearly three times as large as the professional/proprietor effect (which is not statistically significant from farmers). The opposite occupational pattern holds, however, in the upper tail of the distribution. And again, the median results are qualitatively similar to the OLS results.

Estimating quantiles near zero is essentially a procedure that identifies those factors which best explain very low heights, while estimates of upper quantiles point to factors leading to great height. The key variable for identifying high heights is urbanization, while a variety of other county-level variables explain the low heights. We have not yet developed a theory to explain the pronounced nonuniformity of the effects at the extreme ends of the distribution, but the results suggest an important nonlinearity of effects,

Table 5.6 Quantile Regressions

	OLS		Quantile = .1		Quantile = .5		Quantile = .9	
	Coefficient	*T*-statistic	Coefficient	*T*-statistic	Coefficient	*T*-statistic	Coefficient	*T*–statistic
Dependent variable = enlistment height (in.)								
Region								
Mid-Atlantic								
New England	.392	2.51	.453	2.20	.499	2.61	.169	.87
South	.559	2.98	.336	1.35	.484	2.03	.953	4.01
West	.551	4.00	.589	3.63	.451	2.78	.665	3.80
County-level variables (1850)								
Percent urban	–.899	–2.83	.074	.19	.499	2.61	–1.906	–4.56
Rail or water connection	–.281	–2.72	–.186	–1.45	.484	2.03	–.320	–2.41
Agricultural capital per capita	–.002	–2.57	–.003	–2.99	.451	2.78	–.002	–2.35
Manufacturing capital per capita	.000	.06	–.007	–1.85	.000	–.02	.010	2.64
Surplus calories (in thousands)	–.023	–.62	–.088	–2.07	–.028	–.64	.007	.15
Percent foreign-born (× 100)	–.009	–1.59	–.025	–3.16	–.008	–1.13	–.004	–.52
Crude death rate (per 1,000)	.015	1.99	.022	2.35	.008	.82	.011	1.14
School attendance rate (× 100)	–.003	–.72	–.009	–1.99	.001	.26	.000	.03
Illiteracy rate (× 100)	–.036	–3.00	–.052	-3.23	–.034	–2.14	–.030	–1.83
Average family size	.596	3.14	.499	2.26	.718	3.76	.564	2.95
Household-level variables (1850)								
Father and son native								
Father: interstate migrant								
Recruit: native	–.101	–.96	–.240	–1.84	–.080	–.61	.055	.40
Recruit: interstate migrant	–.255	–1.81	–.262	–1.31	–.263	–1.36	–.236	–1.14
Father: foreign-born								
Recruit: native	–.195	–1.40	.004	.02	–.369	–1.82	–.560	–2.64
Recruit: interstate migrant	.193	.55	.356	.77	.087	.18	.369	.71
Recruit: foreign-born	–.432	–1.51	–.726	–1.98	–.534	–1.52	–.557	–1.63
Occupation								
Farmer								
Professional/proprietor	–.391	–2.94	–.296	–1.32	–.400	–1.86	–.251	–1.11
Artisan	–.438	–3.94	–.569	–3.76	–.500	–3.40	–.257	–1.69
Laborer	–.433	–3.47	–.635	–3.54	–.459	–2.68	–.201	–1.11
Unknown	–.066	–.56	–.372	–2.64	–.076	–.54	–.007	–.04

Father's wealth								
$1–100								
$101–500	.191	.73	.252	.69	.493	1.44	.423	1.16
$501–1000	.318	1.25	.425	1.16	.504	1.46	.566	1.54
$1000+	.273	1.07	.338	.93	.499	1.46	.245	.67
Unknown	.179	.70	.281	.79	.398	1.19	.363	1.02
Mother illiterate	–.096	–.74	.211	1.12	–.161	–.88	–.228	–1.18
Number of children								
1–3								
4–6	–.017	–.21	–.005	–.04	.024	.23	–.011	–.10
7+	.190	1.46	.005	.03	.355	2.20	.276	1.66
Enlistment variables								
Enlistment year								
1861								
1862	.040	.37	.232	1.77	–.112	–.88	.024	.18
1863	–.336	–1.95	–.753	–3.44	–.447	–2.08	–.411	–1.87
1864	–.319	–2.91	–.021	.15	–.427	–3.06	–.576	–3.96
1865	–.642	–4.82	–.520	–2.61	–.611	–3.22	–.858	–4.34
Enlistment age								
16	–3.190	–9.75	–3.544	–7.25	–3.152	–6.51	–3.067	–6.26
17	–1.960	–8.09	–2.018	–5.94	–2.072	–6.39	–1.522	–4.44
18	–1.535	–12.74	–1.380	–8.53	–1.648	–10.55	–1.631	–10.05
19	–.795	–6.61	–.688	–3.67	–.930	–5.20	–.833	–4.44
20	–.380	–2.83	–.114	–.56	–.455	–2.35	–.465	–2.31
21	–.181	–1.35	–.240	–1.27	–.265	–1.44	–.203	–1.05
22	–.286	–1.95	–.166	–.78	–.410	–2.00	–.308	–1.44
23	–.277	–1.83	–.010	–.04	–.214	–.98	–.371	–1.65
24	–.141	–.86	–.572	–2.43	0.35	.16	–.146	–.62
25+								
Intercept	65.831	55.95	63.914	45.27	64.800	52.24	68.527	55.22
Sample size	5,758							
R^2	.119		.075		.069		.059	

Notes: See text for other variable definitions. R-squared for the quantile regressions are "pseudo R-squared" values. The OLS regression values are taken from table 5.3.

likely associated with urbanization, that is not captured in the OLS framework. The quantile results for urbanization mirror the earlier results that variation in urbanization had a greater effect on height among farmers than upon nonfarmers, since farmers are likely to dominate in the upper tail of the conditional distribution. Further exploration of these nonlinearities and potential is certainly warranted.

5.5 Conclusions

Our results confirm and extend many of the previous findings related to height in antebellum America. In particular, we find significant effects for region and for various measures of urbanization. Although we do not know with certainty why urbanization matters, the most compelling explanation to date is the role of infectious disease. Though we have certainly not ruled out the importance of the quantity and quality of food in urban populations, it is notable that the effects of urbanization are strongest among farmers, where we would expect access to high-quality food to be less important. In addition to population density, which is proxied (though far from perfectly) by the percentage of the county living in urban areas, access to rail and water transport is important for rural counties. This fact further stresses the importance of infectious disease. We don't find, in contrast to Haines, Craig, and Weiss (2000), that calories produced within the county are important.

Our main intention here has been to explore the possibility that early life, family-level variation in socioeconomic status might have affected health and nutrition of the recruits in the sample. Previous studies have been limited to information available at the time of enlistment and to county-level aggregates. If there is a smoking gun here, it is occupation. The farming advantage exists even after controlling for urbanization, whether through the urbanization index or through estimating the equations separately for urban and rural counties. Furthermore, other investigations (not shown here) indicate that the farmers are at a distinct advantage in even sparsely populated counties. A lingering question in the field's current understanding of the health advantages accrued to the farming class is whether the story is one of access to food or remoteness from population centers and, hence, exposure to communicable disease. The analysis here points to the importance of *both* explanations, though the overall advantage of farming is likely due to several factors.

Other variables suggest the importance of individual characteristics. The descriptive statistics of table 5.1 reveal a positive relationship between wealth and height, but this occurs primarily at low levels of wealth. The regression results show a similar pattern, but the estimates are not statistically significant. The effects of the mother's literacy (in urban counties) and the large differences in effects of migratory history also suggest the importance of the family-specific variables. For instance, the sharp differences between the movers and nonmovers among the sons of the foreign-born point directly to family-specific influences on health and nutrition. The positive effect of geographic mobility (which is likely an indicator for a variety of unobserved family-specific characteristics) of the recruits with foreign-born parents is more than strong enough to offset the negative impact of foreign nativity on the height of the children.

We have attempted to confirm the robustness of the central results by using different estimation techniques, including a county-level fixed effects regression, which revealed no notable differences from the OLS results, and quantile regression. The quantile regressions demonstrate that even though many of the county and household variables simply shift the conditional distribution of height upward or downward, many other variables have sharply different effects at different points in the conditional distribution. The most important is that the effect of variation in urbanization is highly concentrated in the upper tail of the distribution. We conjecture that those in the lower end of the conditional distribution already have had so many insults to their health and nutrition that additional variation in urbanization does not matter significantly. In a sense, the quantile regressions can be interpreted as a diagnostic tool of the basic linear specification of the full OLS model. They reveal that the effects of urbanization are not constant across subgroups of the population—which is the same thing that is shown in making comparisons between urban and rural counties (table 5.3) and between farmers and nonfarmers (table 5.4).[13]

These results leave open many avenues for future research, including understanding why farmers (whether living in urban or rural counties) enjoyed a significant height advantage; what explains the regional variations in height; and what, in general, are the relative contributions of gross nutrition and infectious disease on height. We expect the progress will be made both in the exploitation of additional data and through implementing new empirical specifications and techniques of estimation. The results to date suggest that there is no single socioeconomic mechanism working in concert with genetics to determine height. Rather, variation in height is the result of an interplay of genetics, location specific characteristics and family characteristics.

References

Buchinsky, Moshe. 1994. Changes in the U.S. wage structure 1963–1987: Application of quantile regression. *Econometrica* 62 (March): 405–58.

13. In exploratory work, not presented here, we have explored a variety of interactive effects, where we interacted urbanization, wealth, and occupation. We found no compelling results in this exercise, and we note again that interacting variables often results in cell sizes that are much to small to estimate reliably given the size of our data set.

Conley, Timothy G., and David W. Galenson. 1998. Nativity and wealth in mid-nineteenth-century cities. *Journal of Economic History* 58 (June): 468–93.
Costa, Dora L. 1993. Height, wealth, and disease among the native born in the rural antebellum North. *Social Science History* 17 (Fall): 355–83.
Costa, Dora L., and Richard Steckel. 1997. Long-term trends in health, welfare, and economic growth in the United States. In *Health and welfare during industrialization,* ed. Richard H. Steckel and Roderick Floud, 47–90. Chicago: University of Chicago Press.
Eide, Eric, and Mark H. Showalter. 1999. Factors affecting the transmission of earnings across generations: A quantile regression approach. *Journal of Human Resources* 34:253–67.
Eveleth, P. B., and J. M. Tanner. 1990. *Worldwide variation in human growth.* 2nd ed. Cambridge: Cambridge University Press.
Floud, Roderick, Kenneth W. Wachter, and Annabel Gregory. 1990. *Height, health, and history: Nutritional status in the United Kingdom, 1750–1890.* Cambridge: Cambridge University Press.
Fogel, Robert W. 1986. Nutrition and the decline in mortality: Some preliminary findings. In *Long-term factors in American economic growth,* ed. Stanley L. Engerman and Robert E. Gallman, 439–555. Chicago: University of Chicago Press.
———. 1989. *Without consent or contract: The rise and fall of American slavery.* New York: Norton.
———. 1994. Economic growth, population theory, and physiology: The bearing of long-term processes on the making of economic policy. *American Economic Review* 84 (June): 369–95.
———. 2000. *The fourth great awakening and the future of egalitarianism.* Chicago: University of Chicago Press.
Gallman, Robert E., and John J. Wallis. 1992. *American economic growth and standards of living before the Civil War.* Chicago: University of Chicago Press.
Haines, Michael R., Lee A. Craig, and Thomas Weiss. 2000. Development, health, nutrition, and mortality: The case of the "antebellum puzzle" in the United States. NBER Working Paper no. H0130. Cambridge, Mass.: National Bureau of Economic Research, October.
Koenker, Roger, and Gilbert Bassett Jr. 1978. Regression quantiles. *Econometrica* 46:33–50.
Komlos, John. 1994. *Stature, living standards, and economic development: Essays in anthropometric history.* Chicago: University of Chicago Press.
———. 1998. Shrinking in a growing economy? The mystery of physical stature during the Industrial Revolution. *Journal of Economic History* 58 (September): 779–802.
Margo, Robert A. 2000. *Wages and labor markets in the United States: 1820–1850.* Chicago: University of Chicago Press.
Margo, Robert A., and Richard H. Steckel. 1982. The heights of American slaves: New evidence on slave nutrition and health. *Social Science History* 6 (Fall): 516–38.
———. 1983. Heights of native-born whites during the antebellum period. *Journal of Economic History* 43 (1): 167–74.
Rogers, W. H. 1992. sg11: Quantile regression standard errors. *Stata Technical Bulletin* 9:16–19.
Steckel, Richard H. 1979. Slave height profiles from coastwise manifests. *Explorations in Economic History* 16 (Fall): 363–80.
———. 1995. Stature and the standard of living. *Journal of Economic Literature* 33 (December): 1903–40.

Steckel, Richard H., and Roderick Floud, eds. 1997. *Health and welfare during industrialization.* Chicago: University of Chicago Press.
Steckel, Richard H., and Donald Haurin. 1994. Health and nutrition in the American Midwest: Evidence from the height of Ohio National Guardsmen, 1850–1990. In *Stature, living standards, and economic development,* ed. John Komlos, 117–28. Chicago: University of Chicago Press.
U.S. Department of Health and Human Services. 1987. Anthropometric reference data and prevalence of overweight, United States, 1976–1980. Publication no. 87-1688. Washington, D.C.: DHHS, October.