

Comparing population distributions from bin-aggregated sample data: An application to historical height data from France

Jean-Yves Duclos*, Josée Leblanc[†], David Sahn[‡]

20th April 2009

Abstract

This paper develops a methodology to estimate the entire population distributions from bin-aggregated sample data. We do this through the estimation of the parameters of mixtures of distributions that allow for maximal parametric flexibility. The statistical approach we develop enables comparisons of the full distributions of height data from potential army conscripts across France's 88 departments for most of the nineteenth century. These comparisons are made by testing for differences-of-means stochastic dominance. Corrections for possible measurement errors are also devised by taking advantage of the richness of the data sets. Our methodology is of interest to researchers working on historical as well as contemporary bin-aggregated or histogram-type data, something that is still widely done since much of the information that is publicly available is in that form, often due to restrictions due to political sensitivity and/or confidentiality concerns.

Key words: Health, health inequality, aggregate data, 19th- century France, welfare

JEL Classification: C14, C81, D3, D63, I1, I3, N3.

* Institut d'Anàlisi Econòmica (CSIC), Barcelona, Spain, and Département d'économique and CIRPÉE, Université Laval, Canada; email: jyves@ecn.ulaval.ca

[†] Department of Finance, Ottawa, Canada; email: Leblanc.Josée@fin.gc.ca

[‡] Cornell University, Cornell, USA; email: david.sahn@cornell.edu

1 Introduction

There are many reasons to consider dimensions of well-being other than income or expenditure, both normative and practical. Following Sen (1985) and others, for example, one may wish to consider well-being as multidimensional, comprising characteristics such as , good health, nutrition, literacy, and freedom of association. Income may be instrumentally important to achieve these ends, but it is the capabilities themselves that are intrinsically important and merit recognition and measurement in their own right. Poverty can thus be defined as deprivation of basic capabilities or the failure of certain basic functionings, not just low levels of income. Deprivation of capabilities can also in turn contribute to low material standards of living.

This paper focuses on health, which is certainly important in a multidimensional understanding of well-being. In fact, even in a purely unidimensional welfarist framework, it can be argued that health contributes to welfare at least as much as income. Income may not even be a sound directional indicator of overall welfare in some environments. As Floud (1984) writes, “[t]here is little point in an improvement in real wages which is bought at the expense of a miserable life and an early death”. The example of the United States, for instance, shows that although health and economic growth have generally converged during the twentieth century, they diverged during the nineteenth century (“the antebellum period”) (Costa and Steckel 1997); strong economic growth during the nineteenth century coincided *inter alia* with a decrease in body-heights. Whether income is more indicative of welfare than health in such circumstances is then open to debate.¹

Beyond the conceptual arguments, there are many practical reasons to measure well-being in non-income dimensions. First, measurement difficulties may be less of a problem for some non-income variables. Collecting income (or expenditure) data is a complex procedure that contrasts, for instance, with the relatively straightforward procedure of collecting anthropometric data — data that also suffer typically less from misreporting. Measurement errors can of course still affect such data, but anthropometric data (unlike other self-reported and subjective measures of health) are more likely to be uncorrelated with important variables of interest — such as the welfare variable itself. Note also that health can be more

¹More generally, measures of health are often not highly correlated with incomes, either within a given country or across countries (Haddad and Ahmed 2003, Behrman and Deolalikar 1988, Behrman and Deolalikar 1990, Appleton and Song 1999), suggesting among other things that health variables may provide significant information on welfare that is not captured by income alone.

easily measured at the individual rather than at the household level, thus largely avoiding the need to make difficult assumptions on the well-being of individuals which requires an assessment of both the needs of individuals and how resources are allocated among household members relative to needs.

Perhaps most importantly, the choice of a welfare indicator obviously depends on the availability of data. Research on the distribution of well-being routinely takes advantage of the availability of large-scale micro-data sets that provide detailed information on money-metric and non-monetary indicators². The choice of welfare indicators is much more constrained for studies covering earlier historical periods where the available data is more limited and often of variable quality. It may for instance be the case that only part of the initially gathered welfare information was preserved, or that it comes from sources whose primary goal was not to capture data representative of the population³. Furthermore, historical household level data on incomes and expenditures are particularly difficult to collect since economies were less monetized, transactions were often in-kind, consumption was largely from home consumption rather than market purchases, and tax authorities did not have good records of gathering and verifying information on incomes.

Fortunately, quality historical data are widely available from many countries on one of the clearest manifestations of health and nutritional status, stature. It is also now well established that one of the best global indicators of living conditions is height, standardized for age and gender (de Onis, Frongillo, and Blossner 2000). More specifically, stature is the outcome of a combination of inputs that affect nutrition and disease, such as the local health environment, access to clean

²Some of the non-monetary indicators found in the literature include body-height (Fogel, Engerman, Floud, Steckel, Trussell, Wachter, Sokoloff, Villaflor, Margo, and Friedman 1982, Steckel and Floud 1997, Wagstaff 2002, Pradhan, Sahn, and Younger 2003, Sahn and Younger 2005), body mass index (Costa and Steckel 1997), amount of abdominal fat (Costa and Steckel 1997), birth weight (Costa 1999), life expectancy (Whitwell, Souza, and Nicholas 1997, Goesling and Firebaugh 2004), the overall mortality rate or mortality before a certain age (Costa and Steckel 1997, Floud and Harris 1997, Weir 1997, Whitwell, Souza, and Nicholas 1997, Wagstaff 2002), the prevalence of chronic or severe illness (Costa and Steckel 1997, Wagstaff 2002, Anson and Sun 2004), the individual's own assessment of his or her health (Deaton and Paxson 1998, Nolte and McKee 2004), disabilities (Anson and Sun 2004), difficulties accomplishing tasks (Anson and Sun 2004), and mental illness (Anson and Sun 2004).

³For example, Costa (1999)'s data on the birth weight of babies were obtained from hospital archives. These archives are not necessarily complete, since hospitals will not have kept every patient file since 1848; moreover, the hospital registries may represent a biased sample of the population, since wealthier women were over-represented among those who gave birth in hospitals.

water, nutrient intake, maternal health status, health technology, the organization of work, and so forth. In short, stature captures “multiple dimensions of the individual health and development and their socio-economic and environmental determinants” (Beaton, Kelly, Kevany, Martorell, and Mason 1990)⁴. And in particular, heights of young men entering adulthood is a cumulative indicator of their overall health and nutritional status during their formative years, particularly the period prior to the beginning of puberty.

Economic historians have thus expended considerable effort to examine changes in anthropometric outcomes of various populations over time (Fogel, Engerman, Floud, Steckel, Trussell, Wachter, Sokoloff, Villaflor, Margo, and Friedman 1982, Steckel and Floud 1997, Weir 1997, Deaton and Paxson 1998 and Goesling and Firebaugh 2004).⁵ An important concern that arises in how historians use anthropometric indicators is whether the summary health statistics they employ, particularly measures of central tendencies, can adequately capture the *distribution* of health. Indeed, it is increasingly recognized that looking at entire distributions of health, just as economists have long done with incomes, can provide valuable information that would otherwise be hidden by summary health statistics, such as means and the share of the population that falls before a normative standard, or cut-off point, that may define poor health⁶.

This paper gives prominence to the distributional analysis of health by examining both the evolution and the distribution of heights throughout France in the nineteenth century. More specifically, the paper uses a particularly rich data set collected on men who were called up for possible conscription into the French army during this period. The screening of all men at the age of 20 for manda-

⁴This is also supported by the existence of a significant correlation between body-heights and various indicators of health — see for instance Wagstaff (2002) — suggesting that the choice of an indicator other than body-height would yield similar results. Moreover, adult stature is not only a good indicator of prior episodes of, infection and chronic disease, but it has also been shown to be an important determinant of risk of morbidity and mortality.

⁵A number of other studies have also looked at health inequalities, using statistics such as covariances (such as Deaton and Paxson 1998 and Anson and Sun 2004), differences between various percentiles (Costa 1999), distances from the mean for different social classes (Anson and Sun 2004), or inequality indices (*e.g.*, Wagstaff 2002, Pradhan, Sahn, and Younger 2003, Goesling and Firebaugh 2004, Sahn and Younger 2005). Studies of health inequality have also attempted to explain whether inequality in some factors, such as income, can transmit itself into health inequality — see for instance Weir (1997), Anson and Sun (2004), and Nolte and McKee (2004)). Other studies have decomposed the evolution of health inequality across factors, such as Pradhan, Sahn, and Younger (2003), Goesling and Firebaugh (2004), Sahn and Younger (2005), and Wagstaff (2002).

⁶For instance, see Pradhan, Sahn, and Younger (2003) and Deaton (2003).

tory military service involved a physical examination, including measuring their height. Thus, we have a virtually complete census of heights for each year, disaggregated by administrative department. Data in such abundance for a whole century are a rare find. Consequently, socio-economic improvements as well as periods of adverse conditions in 19th-century France can be expected to have an observable effect measured by the stature achieved at 20 years of age - which is what we measure in our data.

While individual heights were recorded at the time of conscription, the data we have available are limited to the number of men that fall into classes, or bins of height intervals, for each year and department. This raises obvious challenges given our objective to compare the *entire* distributions of heights. These difficulties are not unique to our data, or the paper's historical period of interest. Even today, much of the information that is used and publicly available on distributions of income come from bin-aggregated, or histogram-type, data — important examples of this are the popular World Income Inequality and POVCAL databases⁷. In several countries, the unit data from early household surveys have not survived; in other cases, access to distant and/or more recent microdata is restricted by political sensitivity and confidentiality concerns. We therefore propose and implement a method that estimates the entire population distributions from the bin-aggregated sample data through the estimation of the parameters of mixtures of distributions that allow for maximal parametric flexibility⁸. While we only apply our method to the historical height data from France, this approach will be of general interest to both historians and other researchers working on contemporary bin level data on household incomes, heights, and other indicators of well-being.

Our data are abundant, consisting of more than 6000 different distributions of heights for the period 1819 to 1900 over 90-some departments. Using the sampling distribution of the estimators of the means and of the cumulative distributions of heights, we test for differences in means and also implement tests for robust comparisons across time and regions. We are also able to correct our inference procedures for the possible presence of measurement errors. This is rarely possible to do with the usual data used for comparing welfare; it is made feasible here by the richness of the data. We correct for measurement errors by measuring and taking into account the noise that is (possibly) introduced by measurement errors in the many year-to-year comparisons of the distributions that are

⁷See www.wider.unu.edu/wiid/wiid.htm and www.worldbank.org/LSMS/tools/povcal/

⁸See for instance Bandourian, McDonald, and Turvey (2002) for a review of some of the more restrictive functional forms that have been proposed in the literature.

made possible by our data. This renders the broader comparisons in which we are interested robust to both sampling and measurement errors. Again, this is done by estimating the parameters of mixtures of distributions with the maximum degree of parametric flexibility and by therefore exploiting all the statistical information that is present in the available height data.

In short, this paper develops a methodology that allows us to estimate the entire population distributions from the bin-aggregated sample data. We go on to illustrate how this methodology can be applied to a rich data set from France of 20-year-old army conscripts, and thereafter employ the generated distributions to address one of many possible questions on the evolution and the distribution of health and welfare in 19th-century France. These data are introduced in Section 2. The methodology for deriving from bin data the entire distribution as well as the average height of each department-year is described in Section 3.1 and extracts the maximum possible amount of information from the data. Statistical tests of differences in mean heights and in the entire distributions are performed using the test statistics and sampling distributions presented in Section 3.2. Further details in carrying out stochastic dominance tests appear in Section 3.3. The empirical results are presented in Section 4, where two main questions are more particularly considered to illustrate how our statistical procedures can be used. The first question considers the evolution of body-heights in France from the beginning to the end of the century; the second deals with the regional correlates of the distribution of body-heights. Section 5 concludes the paper. The Appendix in Section 6 provides more technical details on the data, the estimation procedures, and the adjustment for measurement errors.

2 Data

The data we use are from the registries of potential conscripts into the French army for the period 1819 to 1900, covering the 90-some departments of France⁹. More precisely, they are from the "Comptes rendus statistiques et sommaires." Depending on the year, they constitute either a "nearly" random sample, or a complete census of young French men aged 20 years old for each year of the century (this point is addressed in greater detail in the Appendix on page 22). They consist of 6369 different datasets, each representing the distribution of body-heights for a

⁹We are very grateful to Gilles Postel-Vinay for his generous assistance in making available to us, and helping us better understand the data. More details on the data set are also found in Sahn and Postel-Vinay (forthcoming).

given department and a given year. In total, we have measurements on close to 15 million young French men over the course of the century.

A particularity of these data is that they are not available in a totally disaggregated form. Rather, they are grouped into "classes," each of which contains individuals within a specific body-height range. The available data thus report the number of individuals in each of these classes. Furthermore, neither the number of classes, nor the boundaries between bins, were constant throughout the century. The data from 1819 to 1829 are divided into 15 classes, those from 1830 to 1871 into 16 classes, and those from 1872 to 1900 into 9 classes. The class boundaries for the various years are reproduced in Table 1. Note that the classes were based on the old imperial measurements until 1867 (1 French inch = 27.07 millimeters) so that 1.570 meters corresponds to 4 feet 10 inches, 1.597 meters to 4 feet 11 inches, etc. The metric measurements we employ are the public equivalent to the imperial measurements used during these years.

These registries contain a considerable amount of information. Indeed, the number of conscripts measured each year varies between one-quarter, and all, of the male population aged 20. Between 1819 and 1830, approximately 80,000 men were measured each year; from 1836 to 1885, approximately 150,000; and as of 1886, approximately 300,000. Only during this final period was the entire male population aged 20 measured each year.

3 Estimation and inference

3.1 Estimation of complete distributions

That the data available for this study are grouped into classes raises difficulties if the objective is to compare complete distributions of heights. To address this challenge, we first need to estimate the continuous population height distributions using the discontinuous sample histograms that are available. To do this, we solve a system of $(C - 1)$ equations in $(C - 1)$ unknowns, where C is the number of classes into which the heights are regrouped in the aggregated sample data. Each of these equations captures the probability of belonging to one of the C classes of height. Equation (1) defines such a probability for the class c of heights x , a class whose lower bound is \underline{x}_c and upper bound is \bar{x}_c :

$$F(\bar{x}_c; \hat{\Theta}) - F(\underline{x}_c; \hat{\Theta}) = \hat{H}_c, \quad (1)$$

where F stands for the height distribution function, $\hat{\Theta}$ is the vector of the estimated parameters for this function, and \hat{H}_c is the proportion of heights between \underline{x}_c and \bar{x}_c that is observed in our data.

F is specified as a mixture of normal distributions, namely, as a weighted sum of several normal distributions. We let the mixture use as many parameters as is statistically possible given the grouped form of our data. This mixture of normal distributions thus allows for the maximal possible amount of estimation flexibility. Note that since the normal distribution is smooth, this property will also be imposed on our estimated population height distributions.

Equation (2) provides an example of a “mixture” of three normal distributions with a set of 9 parameters:

$$F(x; \alpha_1, \alpha_2, \alpha_3, \mu_1, \mu_2, \mu_3, \sigma_1, \sigma_2, \sigma_3) = \alpha_1 \Phi\left(\frac{x - \mu_1}{\sigma_1}\right) + \alpha_2 \Phi\left(\frac{x - \mu_2}{\sigma_2}\right) + (1 - \alpha_1 - \alpha_2) \Phi\left(\frac{x - \mu_3}{\sigma_3}\right), \quad (2)$$

where Φ is the distribution function of the standard normal distribution, and where α_d , μ_d and σ_d correspond respectively to the weight, the mean, and the standard error of normal distribution d . Note that since $\alpha_1 + \alpha_2 + \alpha_3 = 1$, we can set $\alpha_3 = 1 - \alpha_1 - \alpha_2$. There are therefore 8 “free” parameters in (2). Thus, a mixture of $D \geq 1$ normal distributions contains $3D - 1$ free parameters. Similarly, there are only $(C - 1)$ “degrees of freedom” in data aggregated into C classes of heights, since the probabilities of belonging to one class is one minus the sum of the probabilities of belonging to the others. Hence, the problem is to solve a system of equations such as (1), with $c = 1, \dots, C - 1$, using the \hat{H}_c observed probabilities of belonging to C classes and choosing $D = C/3$.

For some years, however, $C/3$ is not an integer. For those years we use instead mixtures of $(C + 1)/3$ or $(C + 2)/3$ distributions, setting the last one (σ_{C+1}) or two (μ_{C+2} and σ_{C+2}) parameters in the mixture to some pre-specified values. These values are chosen as those that are estimated in the 1880 distribution of individual heights, which is the only year for which we have access to the entire set of individual-level data.

More technical details on the above estimation procedure can be found in the Appendix, Section 6.2.

3.2 Test statistics

Once the parameters $\hat{\Theta}$ in (1) are estimated for each year and each department, we can proceed to assess the evolution of the distributions of health in nineteenth-century France. We do this in two ways: first by comparing mean body-heights, which is one of the most common procedures in the literature, and second by comparing “health poverty rates”. Note that these poverty rates will be compared across ranges of possible “health poverty lines”, which will amount to testing for stochastic dominance of height distributions.

Mean height can be estimated as

$$\hat{\mu} = \int_{-\infty}^{\infty} x dF(x; \hat{\Theta}). \quad (3)$$

The height poverty rate (“the poverty headcount”) is the proportion of individuals below a height poverty line. Computing the poverty rate consists of evaluating the distribution function at the poverty line z :

$$F(z; \hat{\Theta}) = \int_{-\infty}^z dF(x; \hat{\Theta}). \quad (4)$$

Once estimated, $\hat{\mu}$ and $F(z; \hat{\Theta})$ can be compared across departments and years. For the comparisons of means, the null hypothesis is that the mean of distribution B does not exceed the mean of distribution A , and the alternative hypothesis is that it does. The test statistic that we use is then

$$\frac{\hat{\mu}_B - \hat{\mu}_A}{\sqrt{\widehat{var}(\hat{\mu}_B) + \widehat{var}(\hat{\mu}_A)}} \quad (5)$$

where

$$\widehat{var}(\hat{\mu}) = \frac{1}{n} \int_{-\infty}^{\infty} (x - \hat{\mu})^2 dF(x; \hat{\Theta}) \quad (6)$$

and n (which is always well above 500 in our data) stands for the number of soldiers over whom the aggregated bin data have been computed. Under the assumption that population heights follow the flexible form given by (2) and that the two means μ_B and μ_A are equal, the statistic (5) can be shown to follow asymptotically a normal distribution with mean zero and unit variance. At the conventional 5% level, the above null hypothesis will then be rejected if (5) is greater than 1.645.

For ordering poverty headcounts¹⁰, we use the test statistic

$$\frac{F(z; \hat{\Theta}_A) - F(z; \hat{\Theta}_B)}{\sqrt{\widehat{var}(F(z; \hat{\Theta}_B)) + \widehat{var}(F(z; \hat{\Theta}_A))}} \quad (7)$$

where

$$\widehat{var}(F(z; \hat{\Theta}_i)) = \frac{F(z; \hat{\Theta}_i) (1 - F(z; \hat{\Theta}_i))}{n}. \quad (8)$$

Under the null hypothesis of equality of the two distribution functions at z , and that population heights follow the flexible form given by (2), the distribution of $F(z; \hat{\Theta}_A) - F(z; \hat{\Theta}_B)$ is asymptotically normal with mean 0 and variance $\widehat{var}(F(z; \hat{\Theta}_B)) + \widehat{var}(F(z; \hat{\Theta}_A))$ (since the samples from A and B are independent).

Note that the expressions $\hat{\mu}$, $\widehat{var}(\hat{\mu})$, $F(z; \hat{\Theta}_A)$ and $\widehat{var}(F(z; \hat{\Theta}_i))$ are readily computed once the parameters $\hat{\Theta}$ in system (1) are estimated. Since $\hat{\Theta}$ contain the maximal number of parameters that can be estimated from our data, these statistics are also as distribution-free as they can be. Note furthermore that the asymptotic result for (7) is valid for the z located at the frontiers of the bins of the aggregated data even when population heights do not follow exactly the flexible form given by (2), since at such z we can estimate (7) and (8) directly from the \hat{H}_c in (1).

3.3 Dominance tests

A poverty comparison that uses (7) depends on the choice of the line z . It is also evidently dependent on the choice of the distribution function as a “poverty index”. To make the paper’s poverty comparisons more robust to such choices, stochastic dominance tests can be performed by comparing poverty rates over ranges of poverty lines. Pushing this approach farther, one can also compare cumulative height distributions over the entire range of possible heights.

To see what this implies in terms of poverty rankings, note that the poverty headcount F belongs to a general class of poverty indices, denoted as $\Pi^1(z^+)$, that can be defined with the help of two simple axioms and of a condition (see for instance Duclos and Araar 2006, Part III). The first axiom, a monotonicity axiom, says that an increase in the body-height of any one individual (provided that no one else’s body-height decreases) should (weakly) reduce the value of a poverty

¹⁰See for instance Davidson and Duclos (2000, Theorem 1).

index. The second axiom, a symmetry axiom, says that interchanging the body-height of any two individuals should not affect the poverty index. The condition is that the poverty index should use a poverty line that is below z^+ .

We then say that distribution B poverty dominates¹¹ distribution A if and only if the distribution function for B lies below that for A for all poverty lines in the interval $[0, z^+]$. Analytically, for generally-denoted poverty indices $P(z)$ and a distribution function $F(x)$, we have

$$\begin{aligned} P_A(z) &\geq P_B(z) && \forall P(z) \in \Pi^1(z^+) \\ \Leftrightarrow F_A(x) &\geq F_B(x) && \forall x \in [0, z^+]. \end{aligned} \quad (9)$$

This results says that ordering poverty headcounts over all lines in $[0, z^+]$ also ranks all poverty indices that meet the monotonicity and symmetry axioms, and for whatever choice of poverty lines below z^+ .

For statistical and normative reasons (see Davidson and Duclos 2006), these dominance tests are better implemented over ranges poverty lines ranging from of z^- to z^+ , rather than from 0 to z^+ (these tests are then denoted in the literature as restricted stochastic dominance tests). Empirically, this interval will correspond to $[1.53, 1.78]$, or from approximately the 3rd to the 97th percentile of the distributions of heights observed in nineteenth century France. The null and alternative hypotheses for the dominance tests that we conduct can then be written as:

$$\begin{aligned} H_0 : \quad & F(z_1; \hat{\Theta}_A) \leq F(z_1; \hat{\Theta}_B) \\ & \text{or} \\ & F(z_2; \hat{\Theta}_A) \leq F(z_2; \hat{\Theta}_B) \\ & \text{or} \\ & \dots \\ & \text{or} \\ & F(z_m; \hat{\Theta}_A) \leq F(z_m; \hat{\Theta}_B) \end{aligned} \quad (10)$$

versus

$$\begin{aligned} H_1 : \quad & F(z_1; \hat{\Theta}_A) > F(z_1; \hat{\Theta}_B) \\ & \text{and} \\ & F(z_2; \hat{\Theta}_A) > F(z_2; \hat{\Theta}_B) \\ & \text{and} \\ & \dots \\ & \text{and} \\ & F(z_m; \hat{\Theta}_A) > F(z_m; \hat{\Theta}_B), \end{aligned} \quad (11)$$

¹¹In the first order, since we could also test for higher-order dominance comparisons — see again Duclos and Araar (2006), Part III.

where the z_i 's represent m points in the interval $[z^-, z^+]$. The null hypothesis is an hypothesis of non-dominance of A by B . The alternative hypothesis is that B dominates A .

The decision rule differs from that for simpler test hypotheses, since H_0 and H_1 are sets of multiple hypotheses. The decision rule is to reject H_0 and conclude that B dominates A if and only if each of the inequalities in H_0 can be rejected at the 5% level. Since it involves testing separately over m hypothesis tests, this test procedure is generally conservative: the 5% nominal level for each inequality test in H_0 leads to a less than 5% probability of committing a Type I error of wrongly rejecting the joint hypothesis of non-dominance of A by B when non-dominance is true.

An illustration of a test procedure for stochastic dominance is provided by Figure 1, which uses data from the *Ain* department. We test here whether distribution B (1886) dominates distribution A (1819) (H_0 : 1886 does not dominate 1819). We see in the lower panel that the t test statistics (of equation (7)) exceed the critical value (the dotted line) across the entire interval $[1.53, 1.78]$, so we can reject H_0 and conclude that year 1886 dominates year 1819, that is, that year 1886 has less height for whatever choice of poverty measures in $\Pi^1(z^+)$.

4 Results

We now turn to the empirical results. Recall that the first step is to estimate an entire population distribution function for each of our 6369 datasets. This is done using the estimation techniques presented in Section 3. Once estimated, these distributions generally fit the empirical distributions very well, as is illustrated for instance in Figure 2. The histogram shows rectangles whose area \hat{H}_c (see (1)) comes from the aggregated data. The line is the estimated density function, a mixture of normal density functions (whose cumulative distributive functions appear in (2)).

Some estimated distributions are very close to the normal distribution, such as the one of Figure 3. This is more often the case of the distributions with eight parameters (a mixture of three normal distributions), *e.g.*, for many of the distributions from 1872 to 1900. Other distributions are less smooth, as Figure 4 shows. It proved impossible to find a satisfactory solution to the system (1) for 3 out of the 6369 distributions of our database. This may be due to numerical limitations in attempting to solve for systems of up to 15 equations in 15 unknowns, in which each equation is a sum of as many as 6 normal distribution functions —

also computed numerically. It can also be because of difficulties inherent in fitting empirical distributions that diverge (because, *e.g.*, of sampling variability) widely from the normal distribution. Representing less than 0.05% of our distributions, these three distributions are dropped from our analysis; further details on them can be found in Leblanc (2007).

4.1 National distributions of heights

We first consider the overall distribution of heights of the French during the 19th century. Figure 5 illustrates mean body-heights in each department for every year. Observe that there appears to be an upward trend in average body-heights. This phenomenon is most clearly seen in Figure 6, which reproduces the national mean of body-heights (*i.e.*, all departments aggregated — more information can be found in the Appendix) from 1819 to 1900. Note that the national mean seems to have increased from around 163.5 to 165.5 centimeters between 1819 and 1900. Figure 6 also suggests that this evolution of the mean was not constant. Finally, observe in Figure 5 that there is considerable variation across departments.

4.2 Year-to-year comparisons

Let us now turn to the year-to-year evolution of heights for each department. For each year, Tables 2 and 3 show the percentage of departments within which this year is better than (or dominates) the preceding year. Table 2 uses means and Table 3 compares distribution functions at a fixed z .

Notice that there appears to be a great deal of variation in the means from year to year. In Table 2, the percentages of departmental means dominating or dominated by the preceding year hover around 15% to 35%. These percentages therefore far exceed 5%, the level of the test (*i.e.* the proportion of times when sampling error would have erroneously led us to conclude that there was dominance when there was in fact none). The percentages of decline are also nearly as high as those of increase.

Similar results in Table 3 are obtained for comparisons of the proportion of individuals whose body-height was below 1.652 meters for the years 1819–1866 and 1.640 meters for 1867–1900.¹² The presumption that the variability in the

¹²These values were selected because they correspond to class boundaries (the proportion used thus corresponds to the sum of the number of individuals in each class below this boundary) and they differ because the class boundaries were changed between these two periods. They also

means actually springs from the raw data, and not from peculiarities in the estimation methods, is thus supported by the fact that a similar dominance rate exists in Table 3 as for means in Table 4.

There are two possible explanations to the relatively high rate of “acceptance” of dominance. The first is that the distributions within a single department truly do vary from one year to the next, and that the null hypothesis of non-dominance of the distributions must naturally be rejected more often than the nominal 5% level of the tests if we want our tests to have some power. The second, more cautious, explanation starts with the presumption that there should be little difference from one year to the next between the population distributions within a single department, and that we need to locate the source of the high rates of acceptance of dominance in measurement errors. This would then admonish caution in our interpretations of the dominance results.

The Appendix (Section 6.4) analyzes the effect of various possible sources of measurement errors on the validity of our inference results. A relatively conservative approach that emerges from that analysis is to consider the year-year acceptance of dominance rankings to stem from department-year-specific measurement errors that are identically, independently and normally distributed across the century. This would suggest a standard deviation of those measurement errors of the order of 1.7 times the size of the sampling error on the estimator of the mean — or roughly 0.23 cm. Note that a standard deviation of 0.23 cm in the distribution of (true) population average heights would also suffice to generate the high rates of dominance acceptance that we observe in Table 2.

Thus, rather than conclude that one year dominates another on the usual basis of rejecting non-dominance at the nominal level of 5%, we will only draw this conclusion if we can reject non-dominance of means at a nominal level of 20%, which is approximately the average rate of dominance rankings across departments from one year to the next over the century. In the case of stochastic (or distribution) dominance (Table 4), the corresponding average rate of dominance rankings is about 1.5%. Even though this is smaller than the 5% nominal level used in testing *each* of the null in the composite null H_0 in (10), recall from the discussion on page 11 that the decision rule is to reject that composite null and conclude that B dominates A *only if all* of the inequalities in H_0 can be rejected at the nominal level. The test procedure is therefore inherently conservative, lead-

correspond to points outside of the tails of these distributions. In fact, approximately 60% of individuals’ body-heights were less than 1.652 meters in 1819, and for 25% body-height was less than 1.640 meters in 1886. Because of this, it is worth noting that only the information present in the data (and not in the estimated distributions) were used for these tests.

ing to a probability of committing a Type I error that can be much lower than the nominal level. Allowing for the presence of measurement errors in our context is not enough to offset this. This explains why the 1.5% quoted above is below the nominal 5% level.

4.3 Did the body-height of the French increase during the 19th century?

We then turn to the following question: “Did the body-height of the French increase during the 19th century?” To answer this, we compare the distributions of the first ten years of available data (1819–1828) with those of the last ten years (1891–1900) on the basis of the elements discussed above — the means and the entire distributions. Comparisons of department-years were performed department by department, in order to account for “fixed effects” unique to each department (slight variations in genetic heritage, different geophysical conditions, *etc.*). If data are available for all years, this corresponds to a maximum of 100 comparisons per department. Overall, 7430 individual comparisons were performed. In addition, aggregate analyses were also performed: for these same years, the means and distribution for all of France were compared.

4.3.1 Comparisons of means

Overall, 93.8% of the distributions at the end of the century had a statistically greater mean than at the beginning. Thus, there was clearly a height progression over the course of the century, even in light of the aforementioned conservativeness of our inference procedures.

One of the reasons why this proportion is not 100% can be found in developments that are specific to certain departments, as Figure 7 illustrates for the Calvados department. We observe that, in this department, the evolution between the beginning and the end of the century is not very pronounced, and that furthermore the mean was already high at the start of the century. Figure 8 provides, however, a good illustration for the Ain department of the situation of a vast majority of the departments, namely a steady progression in body-height throughout the century. In addition, for purposes of verification, we identified the percentage of distributions at the beginning of the century that dominated the distributions at the end of the century. This proportion was only 2%, which adds to the strong statistical evidence of an increase in mean height over the course of the century.

Finally, the same test is performed for France as a whole, *i.e.* we tested for dominance of the aggregate means for the years in question. These comparisons reveal a 100% dominance rate of the end of the century over the beginning of the century (and 0% for the beginning over the end). This, again, confirms the strong evidence of a statistically significant increase in mean height.

4.3.2 Dominance tests

In order to test for a stronger normative ranking of the distributions, we test for dominance of end-of-century distributions over start-of-century distributions for values of z ranging from $z^- = 1.53$ to $z^+ = 1.78$ at intervals of 0.0025 (or 101 comparisons per pair of distributions). Using this approach, we find that 35.6% of distributions at the end of the century statistically and stochastically dominated those at the beginning of the century. Even in light of the conservativeness of the inference procedures outlined above, it therefore appears that a robust progression in the *distribution* of body-heights can be observed during the century. This evolution is confirmed by the percentage of dominance “in the other direction” — the dominance of distributions from the beginning of the century over those at the end of the century is near zero (more precisely, 0.02%). This is also confirmed by the evolution of aggregate distributions for each year. Indeed, 100% of distributions for France as a whole at the end of the century dominate those at the beginning of the century (and 0% of those at the beginning dominate those at the end).

Conversely, we see that in the case of poverty rates, the percentage of dominance of the end over the start of the century, by department, is substantially less than in the case of the means. This finding that we are less likely to be able to reject the null of non-dominance than the null of equality of means is expected and frequently reported in the literature for anthropometric data (see, for example, Sahn forthcoming and Sahn and Stifel 2002). The interpretation is that not all percentiles of the distribution benefited equally from improvements in living conditions. However, given that in the rejecting the null of no stochastic dominance rests on a very strong criterion, *i.e.* that no point of distribution A can lie beneath the corresponding point of B if this latter dominates the former, we limit the “rigidity” in comparisons employing bounds z_- and z_+ corresponding to 1.53 and 1.78.

The results are illustrated in Figure 9 (and partially in Table 5). We observe that, in general, the smaller the interval, the greater the proportion of dominance (which makes sense because it reduces the number of inequalities in (11) that must hold). For example, if we restrict the interval to $[1.55, 1.75]$ (approximately cor-

responding to percentiles 8.5 to 95) we obtain 53.1% dominance, or 17.5% more than in the case initially examined. At the limit, if we evaluate this dominance at a single point (the standard definition of a height poverty rate), we obtain a very high percentage of dominance (in the Figure, this corresponds to all points between the ordinate and the 45-degree line in the plane z^-z^+). For example, if we set the poverty line at 1.57 m, then 98.2% of the time the poverty rate is statistically lower at the end of the century than at the beginning. We also see that if we set the threshold at 1.62 m (as in ?), we obtain end-of-century dominance in 70.4% of cases.

In another vein, it is of some interest to examine Table 6, which shows the differences in the results for stochastic dominance in distribution according to the effective number of points compared in (11). In theory, we should set m to infinity, but in practice this is of course impossible. Instead, we use $m = 101$. Notice that the difference in the results obtained from between 6 and 101 points is small. Also observe that there is little advantage to using more than 101 points. It is therefore more the range, rather than the fineness of the intervals of thresholds, that matters for testing for height dominance.

Table 5 summarizes the results of the comparison between the beginning and the end of the century. Most decidedly, there is an increase in stature over the course of the century, whether this is assessed using means or distributional dominance. However, end-of-century dominance seems much greater when mean body-height is used than when more robust stochastic dominance techniques are applied.

4.4 Are there regional differences in heights?

Let us now turn our attention to showing how our techniques can be employed to address the spatial characteristic of the stature of conscripts. To illustrate, we first examine the question of whether the heights of conscripts from Paris differ from neighboring departments. Next we assess the extent of northeastern and southwestern differences in stature. In both instances, we are interested in the extent to which these differences are observed across the years.

4.4.1 Paris and neighboring departments

In order to assess the height correlate of living in Paris, we conduct a year-by-year comparison of department 75¹³ (which includes Paris) with the neighboring departments 60, 02, 51, 10, 89, 45, 28 and 27.¹⁴ These latter are the departments that immediately border the region Île-de-France (the geographical region including Paris), and they can be located on Figure 10 (where they are underlined).¹⁵

Aggregating across the years, the department of Paris is globally dominated by the bordering departments (Table 7). Indeed, 69.7% of the means of the neighboring departments exceed the Paris means (and only 17.0% of the latter dominate the former). Furthermore, 16.4% of the distributions of the adjacent departments stochastically dominate those of Paris (versus 2.5% for the converse). This proportion rises to 27.9% (versus 4.8%) if the dominance tests are restricted to the interval [1.55, 1.75]. This supports the results obtained previously by other researchers regarding the poorer quality of life in Paris in the nineteenth century.

Figure 11 contrasts the evolution of the means in the department of Paris with those of neighboring departments across the century. There appear to be three distinct periods: one that lasted until 1830, another that ended in approximately 1886 (or 1890), and a final one until 1900. Between 1819-30, 56.6% (versus 21.7%) of means and 8.4% (versus 3.6%) of distributions for Paris dominate those of the other departments. This last percentage increases to 16.9 (versus 6.0) if we compare dominance curves on the interval [1.55, 1.75]. This could be attributable to the fact that, at the beginning of the century, the positive economic effects of the population concentration dominated the negative effects of crowding, such as increased infectious disease, as Paris had yet to become nearly as highly densely populated as it would become later in the century.

Statistical tests for the years 1836-1886, like those for the entire century, indicate that the bordering regions dominate Paris at an even more convincing rate: 82.1% (versus 5.6%) of the means of the other departments dominate those of Paris and 20.7% (versus 0.0%) of the neighboring departments dominate Paris in distribution. In the last part of the century (1890-1900), however, we obtain more qualified results. Indeed, 58.0% (versus 28.4%) of the bordering departments dominate Paris in means. But as we have stressed throughout this paper,

¹³Seine.

¹⁴Oise, Aisne, Marne, Aube, Yonne, Loiret, Eure-et-Loir and Eure.

¹⁵Department 77 (Seine-et-Marne) was excluded from the comparison because its population density, while less than that of Paris, far exceeds that of the other bordering departments owing to its proximity to Paris (it is part of the Île-de-France region).

it is important to look beyond means and instead also focus on the changes in the distribution. Employing the usual interval, [1.53, 1.78], Paris stochastically dominates the other departments (9.1% versus 5.7%), but in the [1.55, 1.75] interval the bordering departments mostly appear to dominate (17.0% versus 12.5%). This suggests the interesting results that while mean height in Paris is lower, the poorest fared better in Paris than elsewhere.¹⁶

4.4.2 North-Eastern and South-Western regions

It has also been observed¹⁷ that the population of the north-east part of France are generally taller than those of the south-west. Some have even drawn a line from Saint-Malo to Geneva to divide these two regions (cf. Figure 10). We use this latter approach to determine whether this association is also suggested by the data and the rigors of our methodology, which once again relies on much more robust statistical testing of means and distributions than is found in the literature. More precisely, we compare all departments that are entirely to the north with those that are entirely to the south of this virtual demarcation (departments that are either bisected by the line or very close to it were omitted).¹⁸

Our results corroborate earlier work that suggests that the population in the northeast was, in fact, taller than that of the south-west. The means of the departments of the north dominate those of the southern departments in 86.4% of the cases versus 5.9% in the opposite direction. Moreover, 32.8% (versus 0.4%) of the distributions in the north stochastically dominate those of the south. These results are even stronger — 51.1% versus 0.7% — if we employ the interval [1.55, 1.75] in testing for dominance. It is also of interest to observe what happens when we examine north-south differences for two periods: 1819-1828 and 1891-1900. Our results indicate that 87.5% of the means of the northern distribution dominate those of the south, and the converse occurs in 4.2% of the cases at the beginning of the century. The comparable numbers for the latter period are 85.5% and 7.6%,

¹⁶It may also be of interest to compare these results with those of (?), who regress the number of individuals below a certain body-height (1.62 m) on several variables, including a dummy for residing in Paris. They perform this regression on data from the first half of the century (1830–50) as well as from the end of the century (1875–1900). They find a negative impact of living in Paris for both periods, although of greater magnitude for the latter period.

¹⁷See (?) for instance.

¹⁸Departments wholly in the north: 02, 08, 10, 14, 21, 25, 27, 28, 39, 45, 50, 51, 52, 54, 55, 57, 59, 60, 61, 62, 67, 68, 70, 75, 76, 77, 78, 80, 88, 89, and 90; departments wholly in the south: 04, 05, 06, 07, 09, 11, 12, 13, 15, 16, 17, 19, 20, 22, 23, 24, 26, 29, 30, 31, 32, 33, 34, 36, 38, 40, 42, 43, 44, 46, 47, 48, 49, 56, 63, 64, 65, 66, 69, 73, 79, 81, 82, 83, 84, 85, 86, and 87.

respectively. From the point of view of the mean, a few less distributions in the north thus dominate those of the south and a few more from the south dominate those from the north at the end of the century.

When we examine stochastic dominance, however, the results are not the same. Indeed, while 25.8% versus 0.01% of the distributions of the north stochastically dominate those of the south at the beginning of the century — 34.1% versus 0.4% dominate at the end of the century. Thus, there was not a decline, but rather an increase in the number of distributions in the north that dominate those of the south in poverty during the century. Residents of southern France thus appear to have progressed slightly more rapidly on average, but certain percentiles of the distribution have fared less well. Overall, and regardless of the point of comparison taken, inhabitants of the north-east of France broadly dominate those of the south-west in stature. Table 8 summarizes the results of all the north-south comparisons.

5 Conclusion

This study has focused on developing a methodology that enables us to generate complete population distributions from bin-aggregated or histogram-type sample data. By generating the entire population distributions from grouped data, we are able to make robust statistical and normative comparisons of the distributions. This is something that has previously not been done by economic historians that have instead relied on primarily non-statistical comparisons of measures of central tendency to examine changes in the stature of populations over time. Likewise, many researchers using contemporary data face similar challenges as they are often limited to bin-aggregated data, due to political constraints or concerns over confidentiality that prevent access to actual observations on households and individuals.

We go on to illustrate our methodology to height data from France where we first generate height distributions, and then use them to answer two main questions regarding the stature of the French during the nineteenth century: How did it evolve over the course of the century? And were there regional and geographical correlates with body-heights? The paper's methodological procedures and the richness of our data allows us to answer these questions using height distributions for the more than 6000 department-year distributions (representing around 15 millions young Frenchmen) and over close to a century.

The paper's first important empirical conclusion is that body-height increased

significantly in France over the course of the century (from 1819 to 1900). Thus, 93.8% of the means and 35.6% of the distributions of the end of the century statistically dominate those of the beginning of the century (as opposed to 2% and 0.02% in the reverse direction). This provides statistically and normatively strong evidence of a global improvement in living conditions over the course of this period.

The second main empirical finding is that the body-height of the French varied significantly with the geographical region in which they lived. First, Parisians were generally shorter than those residing in neighboring departments. 69.7% of means and 16.4% of distributions from bordering departments statistically dominated those of Paris across the century, as opposed to 17% and 2.5% in the opposite direction. This may be attributable to overpopulation of the city of Paris, which resulted in poor sanitation and drove up the cost of living, or to the migration of the poor (possibly shorter than the average) into the capital in search for work. However, this observed first-order stochastic dominance was not constant throughout the century. Indeed, Paris dominated its suburbs at the beginning of the century (1819–30). The situation was then inverted to make room for a strong dominance of the suburbs over Paris by the middle of the century — which persisted, though to a lesser extent, toward the end of the century.

The second “geographical” effect found in this study is that the French of the north-east were generally taller than those of the south-west. Thus, 86.4% of means and 32.8% of distributions of the north dominated those of the south throughout the century, as opposed to 5.9% and 0.4% in the opposite direction. One possible explanation is that some parts of the south featured more difficult geographic terrain (mountainous regions, etc.), and that industrialization and improvements in living conditions evolved less rapidly.

In short, robust statistical tests, both differences-in-means and stochastic dominance tests based on using data extracted from the conscription registries of the French army, allow us to ascertain the existence of clear differences across time and regions. The paper’s statistical methods are in particular designed to be more robust than is usually the case, in that they allow for the presence of both sampling and non-sampling errors while making maximal use of the available empirical information. The paper’s stochastic dominance tests allow ranking height distributions for almost any measure of poverty or social welfare based on body-height, which stands in contrast to the usual comparisons based on summary statistics such as median or mean heights.

Finally, the methodology developed and applied in this paper could be usefully applied to other types of grouped data, such as those on incomes and wealth,

in addition to answering other questions based on this paper's own data. In the case of the latter, for instance, future research could explore to what extent is the increase in body-height during the century attributable to a rise in the mean, and to what extent is it due to changes in the relative distribution. Likewise, it would be possible to use the generated distributions to assess how inequality in heights evolved over the course of the century as well as determine the contributions of years and departments to health poverty and health inequality across the century.

6 Appendix

6.1 Notes on the data

Though the data we use are rich, they also feature some drawbacks. First, because those called up to active service were drawn by lot, not all those selected were measured (or had their measurements retained). Weir (1997) finds that prior to 1872 approximately 19% of recruits were exempted because of physical deformities, 17% for legal reasons, 9% because of a weak constitution, and 7% for being too short. Among those exempted for legal reasons, there is no reason to presume that a bias is introduced into the data, since the exempted individuals then do not in principle differ from typical Frenchmen. We might expect, however, that intellectuals are richer, and thus taller, than the mean. This would imply that their exclusion from the sample might introduce a bias. However, since they represent a small proportion of the population, we can probably ignore this bias. Among those exempted for health reasons, those who were too short were entered into a category of their own (insufficient body-height), and thus are included in our data as a category for which we know the upper bound. It is highly unlikely that those rejected because of a weak constitution were too short, otherwise they would be classified as such.

The only category that could pose real problems for our analysis is physical deformities. In this case, we can only assume that their distribution was the same as that of the overall population in the contingent, *i.e.*, whose body-height exceeded a certain threshold.

Other concerns regarding the reliability of the data are that some individuals may have been measured with their shoes on, the data may have been rounded to the nearest unit, and that there was a switch from the imperial to the metric system in the 1867 registries may have caused significant rounding errors (Weir 1997)¹⁹.

¹⁹In fact, it is likely that during the transition between the two systems, body-height was first

And finally, there is some ambiguity about the class boundaries after 1867. More details on this and other data issues can be found in Leblanc (2007).

6.2 Classification of heights

The most direct way to use the information contained in our aggregated data is to apply an equation (of type (1)) to each available class. However, this will not necessarily yield estimated distributions that are very close to the true distributions, since it does not always impose sufficient constraints on the estimation. In particular, since the first class (corresponding to individuals categorized as too short) corresponds to $(-\infty, b_1]$ (where b_1 is the minimum body-height for being accepted into the army), direct application of this constraint would not impose any value on the distribution function for statures shorter than b_1 , which could lead to estimated distributions that are unrealistically far with any real distribution of body-heights. Thus, we created a supplementary interval by dividing this interval into two parts: $(-\infty, 1.47]^{20}$ and $[1.47, b_1]$. Since the proportion of the male population aged 20 whose body-height is below 1.47 meters should be negligible, we set the proportion of the population in that class to zero and grouped all those categorized as too short into the second class.

Also, for the years until 1871, we combined the two last classes provided by the bin data, since the numbers in each one were virtually nil. This allowed us to retain the same number of parameters as if we had not subdivided the first interval, and to do so without loss of information (since the omitted class did not contribute any additional information). Thus, the last class went from $[1.923, \infty)$ to $[1.896, \infty)$ for the years prior to 1896, and from $[1.96, \infty)$ to $[1.94, \infty)$ for the years 1867–71.

As to the years 1872 to 1900, we opted to treat the last class the same way as the first class. Indeed, the last class for these years contained as many as 20% of the sample, so using this class “as is” as a constraint on the estimation would not necessarily yield a distribution with only a handful of individuals taller than 2 meters (which should be the case in reality). Consequently, we divided this interval into two parts, i.e. $[1.73, 1.85]$ and $[1.85, \infty)$. We assumed that individuals categorized in the class $[1.73, \infty)$ were, in fact, all in the first segment of this interval, and that the proportion in the second segment was zero.

measured according to the imperial system, then the measure was rounded, and finally it was transformed into a metric measure. Heyberger (2005) also underscores that the switch between the two systems was performed in a transitional fashion prior to 1867.

²⁰In practice, we used -1000 rather than $-\infty$.

Since the number of constraints corresponds to the number of parameters, using the aforementioned constraints implies that the estimated distributions for the years 1819–1829 feature 14 parameters; those for 1830–1871, 15; and those for 1872–1900, 10.

6.3 Estimation of national distributions of heights

In order to aggregate the distributions (and the means) of the departments for purposes of generating the yearly distributions (or the means) for France as a whole, we used weights for each department that are drawn from the age pyramids present in the censuses of 1851 (for 1819 to 1859), 1876 (for 1860 to 1885), and 1896 (for 1886 to 1900). For each department, we used as a weight the number of men between 20 and 24 years old in the department divided by the total number of men in that cohort in France that year.²¹

If we let p_i be the number of men between 20 and 24 years old, and let $\hat{\Theta}_i$ represent the estimated parameters of the distribution of department i in a given year, then the estimator and the variance of the estimator of the distribution function for France are given by

$$\hat{F}_{FRANCE}(z^*) = \frac{\sum_i p_i F(z^*; \hat{\Theta}_i)}{\sum_i p_i} \quad (12)$$

$$\text{and } \widehat{\text{var}}(\hat{F}_{FRANCE}(z^*)) = \sum_i \left(\frac{p_i}{\sum_i p_i} \right)^2 \widehat{\text{var}}(F(z^*; \hat{\Theta}_i)). \quad (13)$$

Similarly, if we represent the estimated mean of department i by $\hat{\mu}_i$, the estimator of the overall mean for France and the variance of that estimator are given by

$$\hat{\mu}_{FRANCE} = \frac{\sum_i p_i \hat{\mu}_i}{\sum_i p_i} \quad (14)$$

$$\text{and } \widehat{\text{var}}(\hat{\mu}_{FRANCE}) = \sum_i \left(\frac{p_i}{\sum_i p_i} \right)^2 \widehat{\text{var}}(\hat{\mu}_i). \quad (15)$$

²¹There were some difficulties deriving the weights from 1860 to 1869 since departments 6, 73, and 74 were added in 1860. Additionally, departments 57 and 67 were eliminated in 1870. To deal with these problems we made a series of adjustments in weights of bordering departments.

6.4 Measurement errors

As mentioned in Section 4.2, there are several possible explanations to the relatively high rate of “acceptance” of dominance in Tables 2 and 3²². To examine some of them, let x_{ijk} be the true height x of observation i for department j and year k , and let an estimator of the true population mean μ_{jk} of height x for department j and year k be given by $\hat{\mu}_{jk}$,

$$\hat{\mu}_{jk} = n^{-1} \sum_{i=1}^n (x_{ijk} + \epsilon_{ijk}) + \eta_{jk}, \quad (16)$$

where ϵ_{ijk} is an observation-specific measurement error that we assume to have zero mean and to be uncorrelated with x_{ijk} ; η_{jk} is a department-year-specific disturbance term; and n denotes the sample size that is assumed for expositional simplicity to be the same across all distributions.

Firstly, let us set η_{jk} in (16) to zero. Assuming the existence of the appropriate population moments and using the central limit theorem, $\hat{\mu}_{jk}$ can be shown to be asymptotically normally distributed with an asymptotic variance that is given by $n^{-1} (\text{var}(x_{jk}) + \text{var}(\epsilon_{jk}))$. The presence of the ϵ_{ijk} measurement errors will therefore increase the variance of the estimator relative to the case of the absence of these errors, a variance that would equal $n^{-1} \text{var}(x_{jk})$. Confidence intervals around $\hat{\mu}_{jk}$ will therefore also be wider than in the absence of measurement errors. However, since $\text{var}(x_{jk} + \epsilon_{jk})$ can be estimated from the available sample information, and since $\hat{\mu}_{jk}$ is a consistent estimator of the true mean by the law of large number, the coverage probability of the confidence interval can be maintained to some desired level. Note also that the importance of the ϵ_{ijk} measurement errors fall with n .

Secondly, consider the case in which a measurement error η_{jk} is also allowed to vary unobservably across departments and years. Let $x_{ijk}^* = x_{jk} + \epsilon_{ijk}$, $\hat{\Delta} = \hat{\mu}_{jk} - \hat{\mu}_{lm}$, and also suppose that the η_{jk} are identically and independently normally distributed with variance $\text{var}(\eta)$. $\hat{\Delta}$ then follows an asymptotically normal distribution with variance

$$\text{var}(\hat{\Delta}) = n^{-1} (\text{var}(x_{jk}^*) + \text{var}(x_{lm}^*)) + 2 \text{var}(\eta). \quad (17)$$

Now consider the null that $\mu_{jk} = \mu_{lm}$, namely, that there is no difference in the true population means across the two distributions. Table 2 reports that, using a

²²We focus here on mean dominance, but similar insights apply to distribution dominance. Mean dominance is expositionally simpler to deal with because the estimator of mean height is linear in the observations of individual heights.

critical value of ± 1.65 — which delimits the usual 5% threshold on each side of the standard normal distribution — the null of non-dominance in means is rejected for about 20% of the year-to-year comparisons. Thus, under the above null and under the assumption that the η_{jk} follow a common normal distribution across the j and the k , we have that

$$1 - \Phi_{\hat{\Delta}}^* (1.65; n^{-1} (\text{var} (x_{jk}^*) + \text{var} (x_{lm}^*))) \cong 0.20, \quad (18)$$

where $\Phi_{\hat{\Delta}}^*(\cdot; v)$ is the cumulative normal distribution function of $\hat{\Delta}$ with mean zero and variance v . Transforming (18) into a standardized normal distribution, we find

$$1 - \Phi_{\hat{\Delta}/\text{std}(\hat{\Delta})}^* (0.84; 1) \cong 0.20, \quad (19)$$

where

$$\text{std}(\hat{\Delta}) = \sqrt{\text{var}(\hat{\Delta})} = \frac{1.65}{0.84} \sqrt{n^{-1} (\text{var} (x_{jk}^*) + \text{var} (x_{lm}^*))}. \quad (20)$$

Using (17) and (20), we then have

$$\frac{n^{-1} (\text{var} (x_{jk}^*) + \text{var} (x_{lm}^*)) + 2 \text{var}(\eta)}{n^{-1} (\text{var} (x_{jk}^*) + \text{var} (x_{lm}^*))} \cong 4. \quad (21)$$

Rearranging (21) and supposing for simplicity that $\text{var} (x_{jk}^*) \cong \text{var} (x_{lm}^*)$, we obtain that

$$\text{var}(\eta) \cong 3n^{-1} \text{var} (x_{jk}^*) = 3 \text{var} (\hat{\mu}_{jk}). \quad (22)$$

Expression (22) says that, under the above assumptions, the standard deviation of η is about $\sqrt{3} \cong 1.7$ times the standard error on the estimator of $\hat{\mu}_{jk}$, which is itself given by $n^{-1} \text{var} (x_{jk}^*)$. Since n is around 2000 and $\text{var} (x_{jk}^*)$ is about 6cm in our data, the standard deviation of η needs to be around 0.23cm (a bit more than 2mm) for the results of Table 2 to be accounted for.

Thirdly, note that the model in (16) could also be specified to allow for disturbance terms that are specific to a department j and a year k , such as v_j and ζ_k in

$$\eta_{jk} = v_j + \zeta_k. \quad (23)$$

For instance, the presence of a year-specific error term, ζ_k , may seem to be suggested by the variability of rejection rates across the rows of Table 2.

Fourthly, note that an important alternative interpretation of the term η_{jk} in (16), and of ζ_k in (23), is that they account for *true* variations in the distributions of heights within a single department from one year to the next. In such a case, if we want our tests to exhibit power, it is certainly desirable that the null hypothesis of non-dominance of the distributions be rejected more often than the nominal 5% level of the tests under the null of equality of mean. For this to happen around 20% of the time, as reported in Table 2 and as shown above in the context of measurement errors, would only require that the variability of the yearly changes in heights generate a standard error of about 2.3mm in η_{jk} .

All in all, for statistical prudence, we choose to follow a conservative inference procedure that assigns all year-to-year variations in distributions and means to measurement errors. Hence, rather than concluding that one year dominates another on the basis of rejecting non-dominance with a 5% nominal level, as the usual statistical inference procedure would suggest doing, we will only draw this conclusion if the proportion of rejections across departments exceeds that observed in year-to-year comparisons, namely, roughly 20% for comparisons of means and 1.5% for comparisons of distributions. A “super-conservative” procedure would set these last thresholds to the maximum proportion of rejections observed — 55% for means in Table 2 and 13% for distributions in Table 4. Most of the paper’s qualitative results are in fact robust to either of these procedures.

References

- ANSON, O. AND S. SUN (2004): "Health Inequalities in Rural China: Evidence from HeBei Province," *Health and Place*, 10, 75–84.
- APPLETON, S. AND L. SONG (1999): "Income and Human Development at the Household Level: Evidence from Six Countries," mimeo, Oxford University, Oxford.
- BANDOURIAN, R., J. MCDONALD, AND R. TURVEY (2002): "A comparison of parametric models of income distribution across countries and over time," Working Paper 305, Luxembourg Income Study.
- BEATON, G. H., A. KELLY, J. KEVANY, R. MARTORELL, AND J. MASON (1990): "Appropriate uses of anthropometric indices in children: a report based on an ACC/SCN workshop," Nutrition Policy Discussion Paper 7, United Nations Administrative Committee on Coordination/Subcommittee on Nutrition (ACC/SCN) State-of-the-Art Series, New York.
- BEHRMAN, J. R. AND A. DEOLALIKAR (1988): *Health and nutrition*, Amsterdam: North-Holland Press, vol. 1, 631–711.
- BEHRMAN, J. R. AND A. B. DEOLALIKAR (1990): "The Intrahousehold Demand for Nutrients in Rural South India: Individual Estimates, Fixed Effects, and Permanent Income," *Journal of Human Resources*, 25, 665–696.
- COSTA, D. L. (1999): "Unequal at Birth: A Long-Term Comparison of Income and Birth Weight," Working Paper 6313, National Bureau of Economic Research.
- COSTA, D. L. AND R. H. STECKEL (1997): "Long-Term Trends in Health, Welfare, and Economic Growth in the United States," in *Health and Welfare during Industrialization*, National Bureau of Economic Research, NBER Project Report.
- DAVIDSON, R. AND J.-Y. DUCLOS (2000): "Statistical Inference for Stochastic Dominance and for the Measurement of Poverty and Inequality," *Econometrica*, 68, 1435–64.
- (2006): "Testing for Restricted Stochastic Dominance," Working Paper 06-09, CIRPEE.

- DE ONIS, M., E. A. FRONGILLO, AND M. BLOSSNER (2000): "Is malnutrition declining? An analysis of changes in levels of child malnutrition since 1980," Tech. rep., World Health Organization.
- DEATON, A. (2003): "Health, Inequality, and Economic Development," *Journal of Economic Literature*, 41, 113–158.
- DEATON, A. AND C. PAXSON (1998): "Aging and Inequality in Income and Health," *The American Economic Review*, 88, 248–253.
- DUCLOS, J.-Y. AND A. ARAAR (2006): *Poverty and Equity Measurement, Policy, and Estimation with DAD*, Berlin and Ottawa: Springer and IDRC.
- FLOUD, R. (1984): "The Heights of Europeans since 1750 : a New Source for European Economic History," Working Paper 1318, National Bureau of Economic Research.
- FLOUD, R. AND B. HARRIS (1997): "Health, Height, and Welfare: Britain, 1700-1980," in *Health and Welfare during Industrialization*, National Bureau of Economic Research, NBER Project Report.
- FOGEL, R., S. L. ENGERMAN, R. FLOUD, R. H. STECKEL, J. TRUSSELL, K. WACHTER, K. L. SOKOLOFF, G. C. VILLAFLOR, R. A. MARGO, AND G. FRIEDMAN (1982): "Changes in American and British Stature since the Mid-Eighteenth Century : a Preliminary Report on the Usefulness of Data on Height for the Analysis of Secular Trends in Nutrition, Labor Productivity, and Labor Welfare," Working Paper 890, National Bureau of Economic Research.
- GOESLING, B. AND G. FIREBAUGH (2004): "The Trend in International Health Inequality," *Population and Development Review*, 30, 131–146.
- HADDAD, L. AND A. AHMED (2003): "Chronic and Transitory Poverty: Evidence from Egypt, 1997-99," *World Development*, 31, 71–85.
- HEYBERGER, L. (2005): *La révolution des corps*, Strasbourg: Presses Universitaires de Strasbourg.
- LEBLANC, J. (2007): "Santé et conditions de vie dans la France du dix-neuvième siècle," Master's thesis, Université Laval.
- NOLTE, E. AND M. MCKEE (2004): "Changing Health Inequalities in East and West Germany since Unification," *Social Science and Medicine*, 58, 119–136.

- PRADHAN, M., D. SAHN, AND S. D. YOUNGER (2003): “Decomposing World Health Inequality,” *Journal of Health Economics*, 22, 271–293.
- SAHN, D. AND D. STIFEL (2002): “Robust Comparisons of Malnutrition in Developing Countries,” *American Journal of Agricultural Economics*, 84, 716–35.
- SAHN, D. E. (forthcoming): “Weights on the Rise?” *Journal of Economic Inequality*.
- SAHN, D. E. AND G. POSTEL-VINAY (forthcoming): “Explaining Stunting in Nineteenth Century France,” *Economic History Review*.
- SAHN, D. E. AND S. D. YOUNGER (2005): “Improvements in Children’s Health: Does Inequality Matter?” *Journal of Economic Inequality*, 3, 125–143.
- SEN, A. (1985): *Commodities and Capabilities*, Amsterdam: North-Holland.
- STECKEL, R. H. AND R. FLOUD, eds. (1997): *Health and Welfare during Industrialization*, NBER Project Report, National Bureau of Economic Research.
- WAGSTAFF, A. (2002): “Inequalities in Health in Developing Countries: Swimming against the Tide?” Working Paper 2795, The World Bank.
- WEIR, D. R. (1997): “Economic Welfare and Physical Well-Being in France, 1750-1990,” in *Health and Welfare during Industrialization*, National Bureau of Economic Research, NBER Project Report.
- WHITWELL, G., C. D. SOUZA, AND S. NICHOLAS (1997): “Height, Health, and Economic Growth in Australia, 1860-1940,” in *Health and Welfare during Industrialization*, National Bureau of Economic Research, NBER Project Report.

Table 1: Bounds of the height classes for different years

Height (in meters)			
1819-1829 (15 classes)	1830 and 1836-1866 (16 classes)	1867, 1868 and 1871 (16 classes)	1872-1900 (9 classes)
less than 1.570	less than 1.560	less than 1.550	less than 1.540
1.570 - 1.597	1.560 - 1.569	1.550 - 1.579	1.540 - 1.629
1.598 - 1.624	1.570 - 1.597	1.580 - 1.609	1.630 - 1.639
1.625 - 1.651	1.598 - 1.624	1.610 - 1.639	1.640 - 1.649
1.652 - 1.678	1.625 - 1.651	1.640 - 1.669	1.650 - 1.659
1.679 - 1.705	1.652 - 1.678	1.670 - 1.699	1.660 - 1.669
1.706 - 1.732	1.679 - 1.705	1.700 - 1.729	1.670 - 1.699
1.733 - 1.760	1.706 - 1.732	1.730 - 1.759	1.700 - 1.729
1.761 - 1.787	1.733 - 1.760	1.760 - 1.789	1.730 and more
1.788 - 1.814	1.761 - 1.787	1.790 - 1.819	
1.815 - 1.841	1.788 - 1.814	1.820 - 1.849	
1.842 - 1.868	1.815 - 1.841	1.850 - 1.879	
1.869 - 1.895	1.842 - 1.868	1.880 - 1.909	
1.896 - 1.922	1.869 - 1.895	1.910 - 1.939	
1.923 and more	1.896 - 1.922	1.940 - 1.959	
	1.923 and more	1.960 and more	

Figure 1: Dominance test of 1886 distribution over that of 1819 for the Ain de-
partment

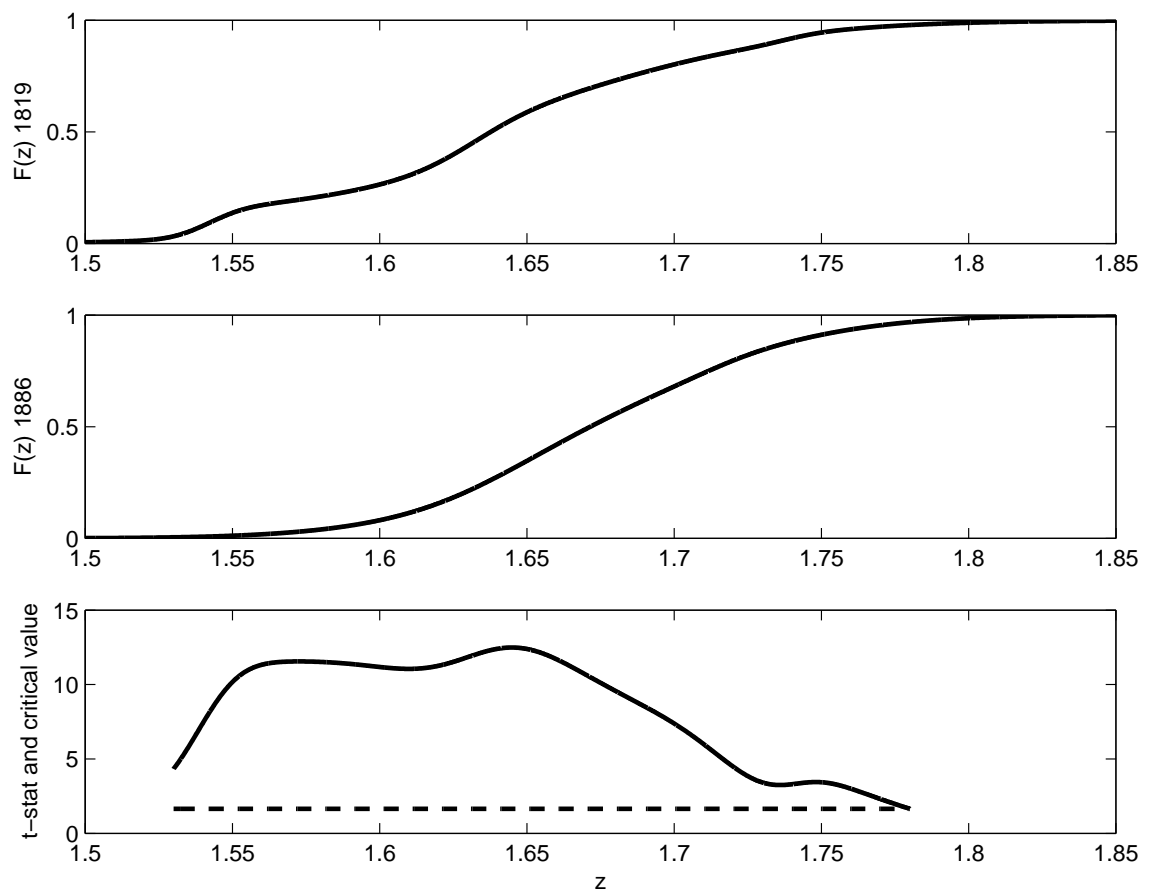


Table 2: Proportions of departments statistically dominating (or dominated by) the preceding year

Year	% of departmental means		(continued)		
	dominating	dominated by			
1820	31.3	7.5	1862	23.6	24.7
1821	11.8	26.3	1863	19.1	29.2
1822	9.7	22.2	1864	34.8	23.6
1823	11.1	27.8	1865	24.7	30.3
1824	50.6	10.1	1866	24.7	21.4
1825	32.5	11.7	1867	48.3	12.4
1826	55.7	8.6	1868	21.4	15.7
1827	30.0	27.1	1872	42.5	21.8
1828	12.9	32.9	1873	23.0	29.9
1829	18.5	35.4	1874	27.6	14.9
1830	44.6	14.3	1875	35.6	14.9
1837	22.1	24.4	1876	18.4	23.0
1838	34.9	10.5	1877	23.0	18.4
1839	27.9	18.6	1878	14.9	25.3
1840	15.1	32.6	1879	18.4	25.3
1841	22.1	24.4	1880	20.7	17.2
1842	23.3	24.4	1881	33.3	14.9
1843	24.4	22.1	1882	14.9	25.3
1844	37.2	18.6	1883	17.2	16.1
1845	29.1	24.4	1884	17.2	27.6
1846	23.3	22.1	1885	27.6	20.7
1847	12.8	44.2	1886	39.1	16.1
1848	39.5	15.1	1887	21.8	24.1
1849	36.5	15.3	1888	25.6	18.6
1850	20.0	24.7	1889	16.3	39.5
1851	30.2	19.8	1890	25.3	18.4
1852	12.8	23.3	1891	26.4	24.1
1853	31.4	19.8	1892	29.1	10.5
1854	12.8	46.5	1893	26.7	23.3
1855	30.2	25.6	1894	25.3	19.5
1856	31.4	26.7	1895	23.0	19.5
1857	30.2	22.1	1896	24.1	21.8
1858	25.6	26.7	1897	23.0	23.0
1859	29.1	22.1	1898	29.9	14.9
1860	19.8	19.8	1899	25.3	14.9
1861	21.4	25.8	1900	20.7	11.5

Table 3: Proportions of departments dominating (and dominated by) the preceding year at $z = 1.652$ (years 1819-1866) and $z = 1.640$ (1867-1900)

Year	% of department $F(z)$		(continued)		
	dominating	dominated by			
1820	27.5	10.0	1862	23.6	13.5
1821	14.5	21.1	1863	15.7	21.4
1822	16.7	23.6	1864	27.0	16.9
1823	9.7	26.4	1865	21.4	22.5
1824	35.4	7.6	1866	21.4	16.9
1824	35.4	7.6			
1826	44.3	10.0	1868	16.9	20.2
1827	28.6	32.9	1872	47.1	18.4
1828	27.1	30.0	1873	24.1	25.3
1829	18.5	29.2	1874	23.0	14.9
1830	17.9	33.9	1875	21.8	11.5
1837	16.3	22.1	1876	12.6	17.2
1838	31.4	10.5	1877	17.2	16.1
1839	25.6	23.3	1878	13.8	18.4
1840	15.1	30.2	1879	14.9	16.1
1841	24.4	22.1	1880	12.6	14.9
1842	24.4	17.4	1881	27.6	8.1
1843	17.4	22.1	1882	12.6	12.6
1844	23.3	18.6	1883	9.2	10.3
1845	24.4	24.4	1884	12.6	14.9
1846	29.1	24.4	1885	10.3	19.5
1847	12.8	31.4	1886	35.6	18.4
1848	26.7	23.3	1887	14.9	20.7
1849	29.1	17.4	1888	18.4	17.2
1850	22.1	23.3	1889	14.9	31.0
1851	30.2	24.4	1890	17.2	11.5
1852	15.1	19.8	1891	16.1	20.7
1853	16.3	17.4	1892	28.7	13.8
1854	19.8	36.0	1893	16.1	19.5
1855	29.1	26.7	1894	9.2	14.9
1856	27.9	20.9	1895	9.2	14.9
1857	19.8	19.8	1896	21.8	12.6
1858	25.6	29.1	1897	14.9	10.3
1859	23.3	16.3	1898	20.7	9.2
1860	18.6	13.9	1899	17.2	12.6
1861	18.0	18.0	1900	16.1	12.6

Table 4: Proportions of departments statistically dominating (or dominated by) the preceding year

Year	% of departmental distributions		(continued)		
	dominating	dominated by			
1820	3.8	0.0	1862	1.1	0.0
1821	0.0	1.3	1863	1.1	1.1
1822	0.0	1.4	1864	2.3	1.1
1823	0.0	0.0	1865	0.0	0.0
1824	2.5	0.0	1866	1.1	0.0
1825	3.9	0.0	1867	3.4	0.0
1826	1.4	0.0	1868	0.0	1.1
1827	1.4	1.4	1872	6.9	2.3
1828	0.0	4.3	1873	1.1	3.5
1829	0.0	1.5	1874	4.6	0.0
1830	3.6	0.0	1875	1.1	4.6
1837	1.2	0.0	1876	1.1	2.3
1838	3.5	1.2	1877	4.6	2.3
1839	0.0	0.0	1878	2.3	2.3
1840	0.0	0.0	1879	1.1	1.1
1841	1.2	1.2	1880	2.3	3.5
1842	3.5	2.3	1881	6.9	0.0
1843	1.2	0.0	1882	0.0	0.0
1844	2.3	0.0	1883	1.1	2.3
1845	0.0	2.3	1884	1.1	3.5
1846	2.3	1.2	1885	3.5	6.9
1847	0.0	12.8	1886	5.8	1.1
1848	7.0	1.2	1887	1.1	1.1
1849	2.4	1.2	1888	1.2	1.2
1850	1.2	0.0	1889	0.0	3.5
1851	1.2	0.0	1890	1.1	1.1
1852	1.2	0.0	1891	0.0	3.5
1853	0.0	0.0	1892	2.3	2.3
1854	0.0	4.7	1893	0.0	2.3
1855	0.0	2.3	1894	0.0	4.6
1856	1.2	0.0	1895	2.3	1.1
1857	0.0	0.0	1896	0.0	2.3
1858	0.0	2.3	1897	1.1	0.0
1859	4.7	0.0	1898	3.5	1.1
1860	0.0	2.3	1899	3.5	1.1
1861	0.0	0.0	1900	0.0	2.3

Figure 2: Aggregated data and estimated entire distribution for the Ain department in 1868

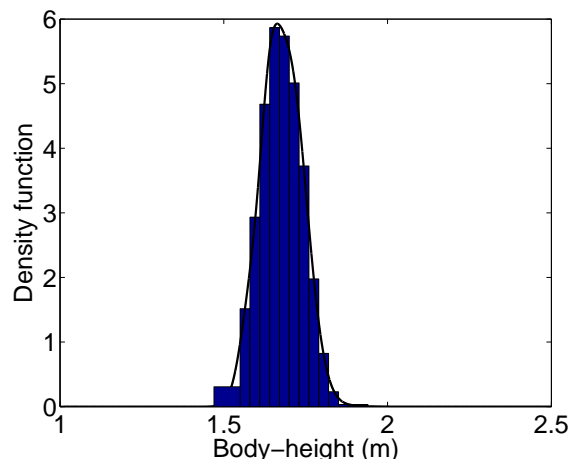


Table 5: % of dominance of end over beginning of century

Mean height dominance	
93.8%	
Stochastic dominance	
Ranges of poverty lines considered	
1.53 - 1.78	35.6%
1.55 - 1.78	40.3%
1.53 - 1.75	47.0%
1.55 - 1.75	53.1%
1.57 - 1.57	98.2%
1.62 - 1.62	70.4%

Table 6: Stochastic dominance of end over beginning of century

Number of points (<i>m</i>)	Distance between points	Percentage of dominance (%)
6	0.05	37.86
11	0.025	36.51
51	0.005	35.61
101	0.0025	35.57
2501	0.0001	35.57

Figure 3: Estimated distribution for the Ain department in 1873

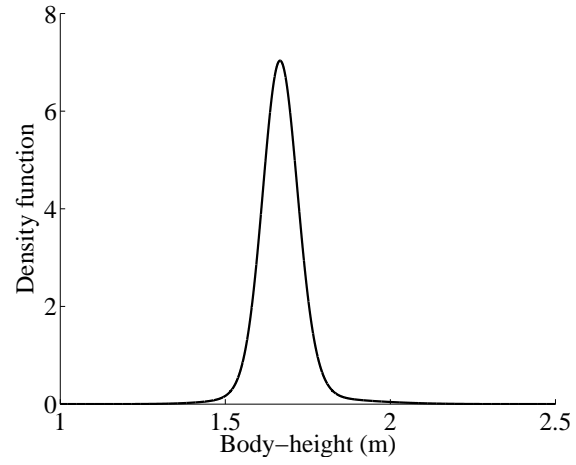


Figure 4: Estimated distribution for the Ain department in 1839

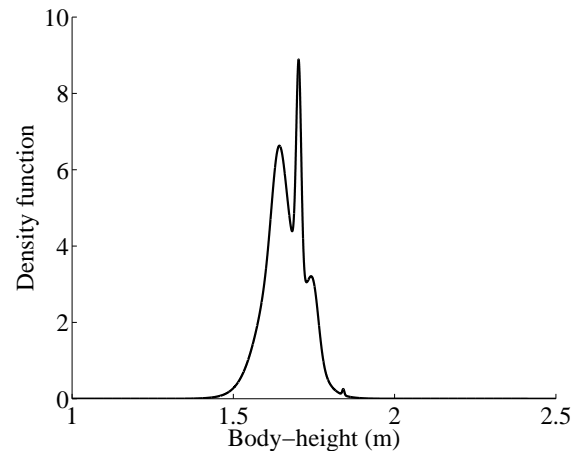


Figure 5: Mean height for each department from 1819 to 1900

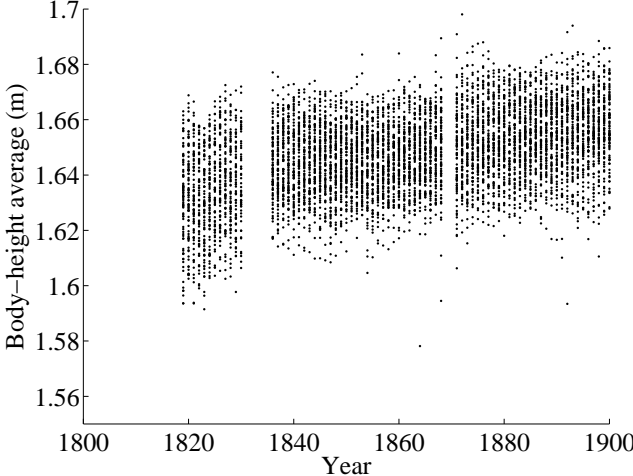


Figure 6: Mean height for all departments combined, from 1819 to 1900

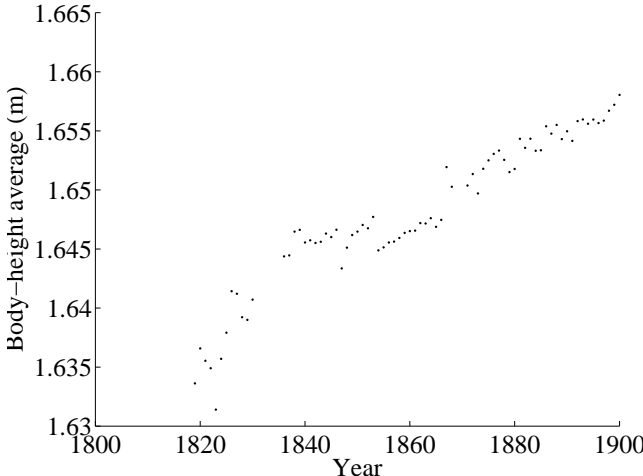


Figure 7: Mean height for the Calvados department

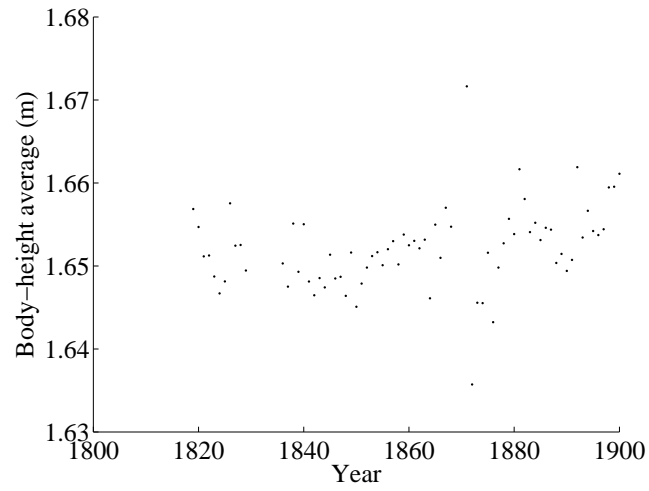


Figure 8: Mean height for the Ain department

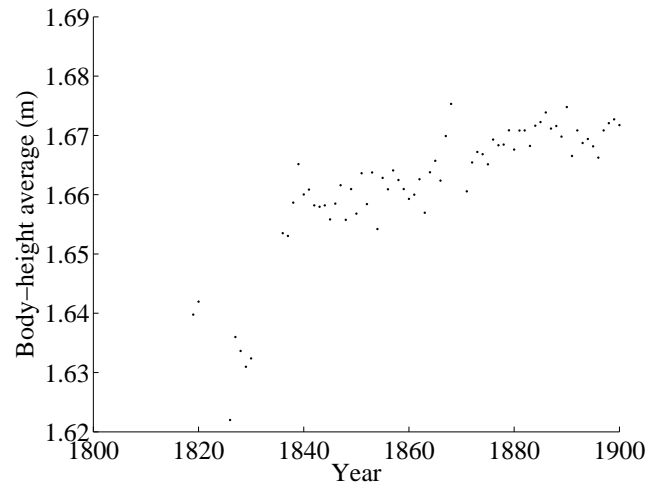


Figure 9: Stochastic dominance of the end over the start of the century with variable bounds for ranges of heights

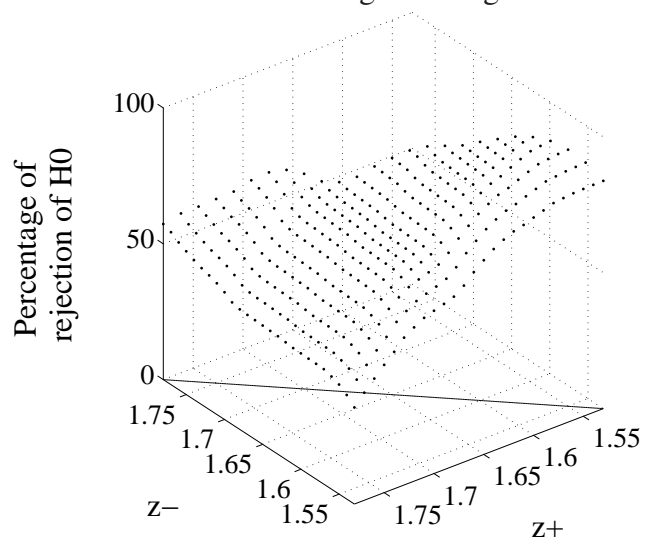


Table 7: Comparing Paris and neighboring departments

Period	Greater mean		Dominating distribution	
	Other dept.	Paris	Other dept.	Paris
1819 - 1900	69.7%	17.0%	16.4%	2.6%
1819 - 1830	21.7%	56.6%	3.6%	8.4%
1836 - 1886	82.1%	5.6%	20.7%	0.0%
1890 - 1900	58.0%	28.4%	5.7%	9.1%

Table 8: Comparing North and South

	Greater mean		Dominating distribution	
	North-east	South-west	North-east	South-west
1819 - 1900	86.4%	5.9%	32.8%	0.4%
1819 - 1900 without Paris	86.9%	5.6%	32.2%	0.3%
1819 - 1828	87.5%	4.2%	25.8%	0.01%
1891 - 1900	85.6%	7.6%	34.1%	0.4%

Figure 10: Map of French departments



Figure 11: Mean heights of department 75 (Paris) versus neighboring departments

