How to (and How Not to) Analyze Deficient Height Samples: an Introduction

John Komlos
Department of Economics
University of Munich
Ludwigstraße 33/IV
D-80539 Munich, Germany
Telephone: +49-89-2180-5824
+ 49-89-8983-9700
Voice Mail: +49-89-2180-3169
Fax: +49-89-33-92-33
email: jk@econhist.de

Many thanks to Brian A'Hearn, George Alter, Bernard Harris, Guido Heineck, and Ricardo Salvatore for helpful comments on an earlier version of the manuscript.

Abstract:

There are many challenges associated with the analysis of historical height data from military records on account of the existence of minimum height requirement prior to universal conscription. This essay provides an overview of two decades of experience working with such deficient height distributions, and outlines various strategies to use and to avoid in order to obtain robust estimates. The conclusion emerges that the QBE procedure is to be avoided, because the other procedures available, the K&K method, or truncated OLS regression or truncated maximum likelihood regression are more robust and are the appropriate methods to analyze such samples.

How to (and How Not to) Analyze Deficient Heights Samples

The systematic study of the physical characteristics of human beings reach back well into the eighteenth century (Tanner, 1981). By the 1830s Adolphe Quetelet and Louis R. Villermé recognized that biological outcomes, such as physical stature were influenced both by the natural, as well as the socio-economic environment (Villermé, 1829; Quetelet, 1831). However, until the late 1960s, when French historians of the Annales tradition began to explore the socio-economic correlates of human height, the topic was of interest primarily to scholars in sister disciplines such as anthropology, biology, medicine, or military history (Le Roy Ladurie, Bernageau, and Pasquet, 1969). The true expansion of anthropometric history began in the mid-1970s when American cliometricians (quantitative economic historians), in search of a new measure of well-being, set out in earnest to study the history of human physical stature during the course of the last two centuries (Harris, 1994; Komlos, 1985, 1995; Steckel, 1995; Komlos and Cuff, 1998).

Anthropometric historians are generally interested in estimating the changes in mean height of a population over time, insofar as that provides an indication of how well the biological organism was able to thrive in its socio-economic and epidemiological environment. From these trends one can infer changes in the level, variability, and the distribution of income, or how well self-sufficient peasants were able to care for their children. They provide an overall indication of the access to nutrients, influenced by their relative prices. In addition, measures of social, gender, and spatial differences in nutritional status are obtained. This is a considerable advantage, insofar as height data are plentiful for segments of the population at the regional level at a time when conventional indicators of economic welfare – such as income - are sparse or completely nonexistent.

This innovative perspective opened up new windows to the past. We now know that during the pre- and early industrial period the biological standard of living depended on such socio-economic and demographic factors that influenced the availability of nutrients, particularly of protein (Komlos, 1998). Population density, population growth, urbanization and the degree of commercialization of the economy all had important effects on the human growth process. We also learned that prior to the onset of modern economic growth, people who were self-sufficient in food production, were living on productive land and in regions with low population density (i.e. relatively removed from urban markets and their disease pools) tended to be relatively tall even if they were poor in conventional terms (Komlos, 1985). Propinquity to nutrients invariably conferred considerable nutritional advantages in the early-industrial period vis-à-vis urban populations, and those engaged in industrial activity prior to the emergence of refrigerated trucks and railroad cars (Cuff, 1998; Craig and Weiss, 1998; Haines, 1998; Komlos, 1989). Physical stature declined at the onset of modern economic growth in the then developed world, and did not begin to improve substantially until well into the second half of the 19th century. This is an indication that there were hitherto unrealized hidden costs to industrialization even in such resource-abundant country as the United States. In marked contrast, the secular trend in stature in OECD countries in the 20th century was less subject to cyclical downturns, except during wars. The reason is that markets in food products became better integrated, so that local shortages in Europe and North America were quickly alleviated. Government expenditures on public health and medical care increased sufficiently to make a major impact on the biological well being of the population. Children's work declined or was entirely eliminated, freeing up calories for the growth process. Welfare programs increased, so that the effect of short term fluctuations in income had a negligible effect on children's heights (Komlos, 1996).

Yet, the aim of this paper is not to provide an overview of the accomplishments of anthropometric history. Rather, the goal is to offer some basic guidelines on how to analyze deficient height samples which are rather common particularly in European data sets. With two decades of experience behind us, it should be useful to share our understanding of current practice. To be sure, in many cases one can apply standard procedures to estimate trends and cross-sectional patterns using ordinary least squares (OLS) linear regression analysis. Such institutions as prisons, passport agencies, and armies with universal conscription did not impose a height requirement (HR) on entrants, and, therefore, the samples drawn from their records are representative, in the main, of the underlying population from which the institution drew its members.[1] In such cases one can obtain a random sample of heights representative of the universe of observations upon which the archival information is based.[2] Those samples are in principal unbiased, and standard procedures of analysis apply, even if the universe might be limited by gender or confined to certain social classes.[3]

Deficient Height Samples

Height data often stem from institutions that imposed a HR as a precondition of entrance. Samples drawn from such records are perforce deficient i.e., incomplete, inasmuch as a substantial portion of the underlying population's height distribution is unavailable for analysis. Obviously, samples drawn from such records are not representative of the universe of observations: people below or above a certain threshold are either missing completely, or are underrepresented. There might be a minimum HR on the left side of the distribution (HR_{min}, at $?_m$,), or on the right side, due to a maximum height requirement (HR_{max}, at $?_x$) (or both). HR_{min} are most common, but HR_{max} were also imposed, if being much taller than average was a disadvantage, as aboard ships, because of the high center of gravity. Estimating procedures devised to correct for a bias caused by the HRs are based on the biological fact that the height distribution of adult (homogeneous) populations are (approximately) normally distributed N (? ,?) (Tanner, 1978; Bogin, 1999). Yet, one does not need to test for normality in such samples because only a part of the distribution is available for analysis and statistical tests of normality have not been devised for distributions with a HR.[4]

A complicating factor is that the HRs varied over time, and were enforced with varying stringency (but *never* perfectly): in effect, the records went through a (time-variant) filter: people shorter than ?$_{m}$, or taller than ?$_{x}$, had a lower probability of entering the sample. Such distributions have a "shortfall" (of unknown magnitude) beyond the HRs, and are thus inherently biased (Wachter, 1981). Shortfall implies that the sample is not random, but that it is not perfectly truncated either. [5] There are some observations missing below ?$_{m}$, or above ?$_{x}$. The amount of shortfall is the (unknown) share of the observations missing in the sample beyond the HRs, i.e., the percent of people who were not accepted into the institution because of the HRs. Without proper statistical correction any results obtained by such deficient samples, such as means, correlations, or regression coefficients are unreliable, misleading and mostly wrong[6] (Komlos, 1993; Heintel and Baten, 1998). (See Appendix A for definitions and abbreviations.) In the presence of shortfall the sample is normally distributed only within the range (?$_{m}$, ?$_{x}$).[7] As an example, consider the height distribution (histogram) of 11,000 adult French soldiers recruited prior to 1740 (Figure 1). An erosion of the distribution below 62 French inches (F.i.) is quite apparent, implying the existence of shortfall. Note, however, that in practice the truncation is not perfect, implying that the HR_{min} was not enforced consistently.[8] This is the typical pattern in armies that did not have universal conscription. An example of the effect of HR_{max} is found in a sample of poor boys admitted into the Marine Society of London, a charitable institution, between 1792 and 1798 (N=559) (Figure 2). The institution prepared boys for life at sea, did not have an official HR_{max}, but an informal one.[9]

Figures 1 and 2 about here

In order to discover such deficiencies in the sample, the very first step in any analysis of height should be a visual inspection of the sample histograms. This is mandatory before proceeding further, even if one has reason to think that HRs did not exist, because informal ones were often enforced, and by examining such evidence one gains a sense of the degree of shortfall, of rounding, and of other possible sample deficiencies. Note that heights were not always recorded precisely, particularly in military samples. Observations were often rounded to a nearest unit (e.g., inch or cm). Heaping on whole, or on even numbers, or numbers divided by 5 is often evident.[10] However, symmetric rounding (heaping) does not introduce appreciable systematic bias into the analysis.[11] The distribution within the range ($?_m$, $?_x$) should be approximately normally distributed, but heaping on favorite numbers might well produce samples which deviate from this general rule.[12] As a consequence, testing for normality is not very helpful. In addition, although the mean and the mode are identical in normal distributions, this is not the case in the presence of heaping (rounding), and consequently the mode is not a robust estimator of the mean in such samples.[13]

Height Distributions

The aim of the visual examination of the histograms is to identify the actual HRs for the period under consideration, and if they changed over time, because one needs the de facto, and not the prescribed HRs for further analysis.[14] In creating the histograms for initial inspection, the height measurements must be left in the original units, because conversion into another system can introduce distortions.[15] Samples from different military units, i,e. cavalry, artillery, navy, infantry, and ancillary troops should not be conflated, because different units generally had different HRs, and the combined sample would become a mixture of truncated normal distributions for which estimating procedures are much more complicated.[16] In historical populations those who were older than 23-years old have reached their final height and can therefore be considered adults. The height distributions of younger soldiers should be analyzed separately, because the time during which their growth could be influenced by environmental circumstances is not perfectly coincident with those of adults.[17] Insofar as height begins to diminish after age 50 – those above this age should not be considered in the analysis.

In samples covering an extended time period, several histograms should be produced. Histograms by decades are desirable if there is sufficient number of observations, but longer intervals are appropriate otherwise. It is usually necessary to separate peace-time from war-time recruiting, because HRs were frequently overlooked during times of need.[18] The dates at which legal changes were made in the conscription laws are also obvious choices for periods.[19] For this part of the analysis the relevant date is that of recruitment, and not that of the time of birth, because HRs changed for recruitment years and not for birth years. Ideally, there should be about 500 records for each histogram in order to minimize the confounding effect of small sample size.[20]

Independent Variables

Only after the HRs are determined, can the actual process of analysis commence. The aim is to estimate: the mean height of the population from which the soldiers were drawn, ? ?and the effect of covariates (independent variables) on height including time, i.e., the trend.[21] Usual explanatory variables available for military samples include age, date of birth, region of birth and of residence, urban/rural provenance, date of enlistment, and occupation. Physical stature is affected by many socio-economic variables. These include, but are not limited to the epidemiological environment – hence the mortality rates in the region of birth or of residence is a legitimate explanatory variable. Social stratification is also a crucial determinant of height insofar as income determines the budget constraint. Education (literacy) has an effect, because better educated parents have superior consumption skills, are better informed about long-range health effects of consumption patterns, and, thus, are usually able to take better care of their off-springs. Education also correlates positively with income. Height is a function of income inasmuch as the consumption of nutrients, particularly of proteins, vitamins, and minerals, and the regularity with which those nutrients are consumed, all influence height at a particular age until adulthood. Urban/rural differences (population density) are also useful predictors of height insofar as the disease environment and medical services vary spatially. Population density also determines the speed with which disease vectors are transmitted throughout the population. From the region of birth many ecological variables might be inferred, depending on collateral information, such as the distribution of income. Occasionally some data might be available on the parents as well.[22] All of these variables can be useful, though usually a small percentage of the variation in height can be explained at the individual level, because most of the individual variation in height is genetic.[23]

Estimation Procedures

Four estimating procedures, available for analyzing normal distributions with shortfall, are discussed below: the Quantile Bend Esimator (QBE), the Komlos-Kim method (K&K), Truncated Ordinary Least Squares (TOLS), and Truncated Regression (TR) using maximum likelihood procedure (Appendix B).[24]

The QBE was suggested by Wachter and Trussell (1981, 1982) to estimate ? ?the true mean height of the underlying normally distributed population from a sample with shortfall: theoretically E(?$_{QBE}$) = ? ??? owever, empirical work has found that ? ?has such a high variance that QBE is not reliable (Table 1).[25] The algorithm estimates the amount of shortfall by filling in the missing observations below ?$_{m}$ until the sample distribution becomes normal (Gaussian).[26] In contrast, the first step for the other three methods is to discard *all* observations outside of the range (?$_{m}$,??$_{x}$) in order to equalize the bias across the whole sample. The K&K method is the easiest of the four procedures and can be done most conveniently with all statistical programs: one calculates the mean of the part of the sample remaining after truncation, i.e., after discarding the portion outside of the range (?$_{m}$,??$_{x}$).[27] While ?$_{K\&K}$? ? ??the direction of the trend of the truncated mean using the K&K method is also the direction of the trend in ? ??the mean height of the population from which the sample originates.[28]

Table 1 about here

The usual approach using ordinary least squares regression (OLS) is unsuitable (biased) in case of samples with shortfall, and should not be used on such data in any circumstances. Two regression methods are appropriate to analyze the correlates of height in case of truncated normal distributions: TOLS and TR.[29] TOLS, first suggested by Komlos (1989), is the easier approach, being the an extension of the K&K method for estimating the effect of covariates of height. TOLS has the advantage that it can be performed with almost all statistical computer programs, including SPSS. This is not the case with TR, which is confined to a more limited set of statistical packages such as STATA (2001, p. 209) and EViews (2000, p. 438). TOLS is OLS regression analysis is after the data outside of the range $(?_m,?_x)$ have been eliminated. Clearly, the coefficients obtained with TOLS are biased, but their relative sizes, as well as their signs are correct.[30] Note, however, that accurate statistical inference (confidence intervals, hypothesis testing) is not appropriate with TOLS. For this purpose the fourth method, the TR, using maximum likelihood estimation, is required.[31]

For the three methods TR, TOLS, and K&K, observations outside of the range $(?_m , ?_x)$ are discarded. As a consequence, the height distributions have to be examined carefully and if $(?_m , ?_x)$ changed over time, then one has to consider which of the $(?_m , ?_x)$ to use. For the TR procedure this is not a problem, insofar as it can be used with several sets of $(?_m , ?_x)$ simultaneously; for the K&K and the TOLS methods, however, the largest $?_m$ ($?_{\mathbf{m}}^{max}$), and the smallest $?_x$ ($?_{\mathbf{x}}^{min}$) are the relevant ones to use for truncation, if $(?_m , ?_x)$ changed over time. These then become the effective $(?_m , ?_x)$ for further analysis. (In subsequent notation we drop the superscripts on the ?'s and assume that they are the appropriate ones for the particular analysis.) For instance, in the case of the French army, $?_m$ was lowered from 62 to 60 French inches (F.i.) between 1740 and 1762.[32] Hence, with K&K and TOLS, one would choose the larger of the two, i.e. 62 F.i., as the binding $?_m$ for the whole period 1716-1789.[33] Thus, the distribution of heights is conditioned to have "identical biases", within the range ($?_m$, $?_x$), and can be used for further analysis. In other words, we make certain that the biases are the same throughout the period under consideration. The observations outside of the range ($?_m$, $?_x$) are then not used for further analysis.[34]

The QBE is the only one of the four methods that does not need information on the HRs, and as is the case with the K&K method, it cannot estimate the impact of covariates on height (Table 1). Moreover, it cannot be used on doubly truncated samples, whereas the other three methods can. However, it was clear from the very beginning of the anthropometric research program that $?_{QBE}$ was inaccurate (Komlos 1985, 1989, p. 52).[35] Simulation exercises confirm that $?_{QBE}$ is an inefficient estimator of ? (Table 2). Experiments using 28 different specifications (each repeated 1,000 times) with various sample sizes, varying amounts of shortfall, and varying truncation points confirm that the QBE procedure is unambiguously inferior to all three other ones (Heintel, 1996a,b). These simulations indicate, that the average bias = $?_{QBE} - ?$ is about twice as large as that of the TR procedure (Table 2). In addition, the variance of $?_{QBE}$ is 1.7 - 3.5 times as large that of $?_{TR}$, and its mean square error ($?^2$ + $bias^2$) is 3.3-3.5 times as large as that of $?_{TR}$ (Table 2). In other words, $?_{QBE}$ is inefficient as the estimates have a large variance compared to the TR procedure.

Table 2 about here

Ascertaining accurately the direction of trend in physical stature is an important issue in anthropometric history: consequently we proceed to estimate the probability of correctly estimating whether the difference between the estimated means of two samples has the correct sign. Consider an experiment in which two samples are drawn from two populations N($?_1$, $?_1$) and N($?_2$, $?_2$). Discard

20% of the sample to the left of $?_{m1}$, $?_{m2}$, and then calculate the probability that the three methods, QBE, TR, and K&K estimate the direction of the trend correctly. That is to say, in how many cases out of 1,000 is it true that if $?_1 < ?_2$ then $?_{1,QBE} < ?_{2,QBE}$ or $?_{1,TR} < ?_{2,TR}$ or $?_{1,K\&K} < ?_{2,K\&K}$. The results of some 20,000 simulations indicate that K&K method is the most robust followed by TR procedure. With a one cm change in the mean, the probability that the K&K method estimated the sign of the change correctly is 0.87; the TR method estimated it correctly in 71 percent of the cases, while the QBE did so in 65 percent of the cases (Table 3).[36] In other words, the QBE procedure is not able to estimate even the sign of the trend correctly in about one-third of the cases.[37] The relative accuracy of the K&K procedure increases until the difference between the two means ($?_1 - ?_2$) reaches 0.5 cm and remains at being about 40% more accurate until $?_1 - ?_2 = 1.25$ cm (Figure 4).[38] Hence, by all measures, the QBE is unacceptably unreliable, and should not be used as an estimator of either the true population means, or of trends, insofar as more accurate procedures are available.

Figure 4 and Table 3 about here

In contrast, the K&K method is quite accurate. While it is suitable primarily for estimating the direction of the trend, it can also be used to convert $?_{TR}$ into ?? as described in fn. 42, although that is a bit complicated. It is primarily useful in obtaining a quick and accurate impression of trends over time. Insofar as it cannot be used to estimate correlates of height, the sample first has to be subdivided in order to ascertain the trends in height for particular subgroups, by geographic provenance, for example.[39] For the K&K method to be more accurate than the TR, one of the following three conditions need to hold: a) ? remain constant; or b) if ?? is not constant, it changes in the same direction as the mean height, ?;[40] or c) if ?? and ?? move in opposite direction, ? should not change by more than about six percent (about 4 mm) (Table 5). If the true heights increase and at the same time as ? declines by about 4 mm (from say 6.5 to 6.1 cm) the TR becomes the superior method. These are weak restrictions, however, insofar as ? tends to remain stable over time, even over centuries, and even as mean heights change considerably. TR is superior to QBE with any change in ? ??Figure 5?. In sum, the K&K method is a very useful first step in any analysis of height trends. It is not foolproof, though, and it is therefore important to supplement it with the two regression methods, TOLS or TR, that are able to explore the effect of covariates (i.e., age, time, birth place, socioeconomic environment, mortality rate, population density) on heights, provided such data are available.

Figure 5 about here

The advantage of TOLS is its ease of application – almost as convenient as the K&K method, but in contrast to K&K, it also estimates the relative effect of the covariates on heights. The estimates $\hat{?}_{TOLS}$ are first converted into (truncated) height estimates. For example, in order to estimate the height of French soldiers who have not yet reached adult height, dummy variables were included for ages ? 22. The constant (171.62 cm) then pertains to the height of 22-year-old soldiers: $?_{TOLS(age=22)} = \hat{?}_{TOLS(age=22)}$ =171.62 cm, which becomes the reference category. If the estimated coefficient of the 20-year-old soldiers, for example, is $\hat{?}_{TOLS(age=20)}$ = –0.44 cm, then this is the amount by which 20-year old soldiers were shorter than 22-year-old soldiers. One can calculate the height of 20-year-old soldiers (above $?_{min}$) as follows: $?_{TOLS(age=20)} = \hat{?}_{TOLS(age=22)} - \hat{?}_{TOLS(age=20)}$ = 171.62–0.44= 171.18 cm. Note that this is the estimated height of 20-year-old soldiers and not that of 20-year-olds in the population. That is to say, $E(?_{TOLS})$? ? . However, these heights can be converted into estimates of the mean height of the sub-populations from which the soldiers were selected, even if the conversion is relatively complicated. The 171.18 implies that 20-year-olds were about 162.3 cm in France in the first half of the 18th

century.[41] These estimates are the same as the maximum likelihood estimates obtained with TR (A'Hearn, 2004).

Though TR is often a bit more difficult to implement in practice, insofar as the procedure is available on fewer statistical computer programs, it has the considerable advantage of immediately providing consistent and unbiased estimates of the coefficients of the independent variables, as well as their standard errors, thereby allowing for further statistical inference, such as the calculation of the t-values of the estimates.[42] Just as importantly, with TR one can use several truncation points for different sub-periods, in contrast to TOLS. One only has to discard the part of the sample outside of the range ($?_{m}$,$??_{x}$) pertaining to the various sub-samples, and specify these upper and lower limits within the command before running the program.[43] A'Hearn (2004) has shown that the accuracy of the TR method might be improved by constraining the historical height distribution to have a ? ?similar to those obtained in modern populations.

Conclusion

The analysis of height data can be challenging if a part of the underlying height distribution is missing from the sample, as is the case for records stemming from institutions which imposed a minimum or maximum height requirement on entrants, as many military establishments, in fact, did prior to the establishment of universal conscription.[44] Several methods that have been used on such deficient height samples over the course of the last two decades are reviewed in this article. The conclusion emerges that it is imperative to begin the analysis with a visual inspection of the sample height distributions, in order to recognize possible anomalies and to determine the actual height requirements in practice, which could deviate from the legally prescribed ones. We conclude that if the estimates obtained fluctuate wildly over time, the presumption is that they are probably incorrect, and need further investigation. During the course of the last century and a half mean height of European populations increased at a rate of between 1-1.5 cm per decade (Cole, 2003), and this should be considered a normal level of change from decade to decade. If the estimates diverge from this order of magnitude substantially then the chances are something is wrong.[45]

Four methods (QBE, K&K, TOLS, TR) for estimating means, trends and covariates of heights from a deficient sample with shortfall have been discussed. We infer on the basis of simulation experiments that the QBE procedure is unambigously inefficient, and should not be used to estimate the mean of a deficient height sample. An example of incorrect inferences that have been drawn with the use of the QBE procedure can be illustrated on the basis of the height of the British population based on military samples. While the original estimates fluctuate implausibly (Figure 6), Floud et al. nonetheless concluded on the basis of the pattern obtained that heights "rose from the middle of the eighteenth century and into the late 1820s…" (1990, pp 28, 275). The incorrect inference is then made pertaining to the very old controversy about the secular trend in living standards of English workers during the Industrial Revolution that, "the era of the early Industrial Revolution seems on this evidence to have led to an improving standard of living for the working population…" (1990, pp 151-152). Yet, the revised estimates, based on the K&K method, yield an entirely different pattern of generally falling physical stature in the late 18th century (Figure 6). The two estimates for the 1810s are as far as 7 cm apart. Moreover, the revised estimates indicate that the peak in the early 19th century is well below the 18th-century peak, so that the proper inference is that living standards declined rather than increased[46] (Komlos 1998).

Figures 6 and 7 about here

The fact that the Floud et al. estimates are inaccurate can be corroborated on the basis of four independent data sets which consistently show a very different pattern than that reported by Floud et al., on the basis of which it becomes clear that heights did decline during the course of the Industrial

Revolution (Figure 7). These estimates are also lower than that of Floud et al. by of about 6 cm for the 1810s, but are greater than that of Floud et al. by 5.5 cm in the 1740s. So that the increase of 7.5 cm increase in height suggested by Floud at al. (1750-1820) is turned into a 4 cm decline!

Another example of the wide margin of error of the QBE procedure is provided by the Floud et al. (1982, 1990, p. 167) estimates of the height of poor English boys. These estimates also fluctuate widely, the estimates frequently cross one another for different ages, although one would expect older boys to be taller (Figure 8). Just as importantly, a general pattern is difficult to recognize. In contrast, the estimates made by Stata's "*trunc reg*" program shows a much more consistent pattern among the different ages. The fluctuations are much more plausible, and heights decline unambiguously at all ages at the end of the 18th century, corresponding thereby with the revised estimates of the military data (Figure 9). Hence, the height estimates of the poor English boys corroborate the revised estimates of the height of soldiers. The QBE procedure obfuscated the patterns considerably, thereby leading to incorrect historical inferences in significant ways.

Figures 8 and 9 about here

In sum, the inescapable conclusion is that the QBE program should not be used. Of the other three methods, The K&K method is the quickest way to obtain an overall impression of trends. Its major limitation is its inability to estimate the covariates of height. In contrast, the TOLS is able to estimate at least the relative size of the covariates, while the TR method is able to estimate both the covariates and their standard error (Table 1). Because all three methods, TR, K&K and the TOLS, have some shortcomings, it is useful to use at least two of the methods to validate the results obtained. Only to the extent that the conclusions obtained with two methods support each other, should the results be accepted as valid. Deficient height samples are rather difficult to analyze, and require a lot of careful consideration of the procedures used. One cannot work with them as easily as one does when one has the full distribution of heights.

Appendix A. Abbreviations – Definitions

amount of shortfall - the percent of missing observations from the sample beyond the HRs.

E – expectation operator.

F.i. – French inches

HR - height requirement(s) (minimum and/or maximum) to be accepted into an institution.

K&K – Method to estimate trends in truncated height distributions. Uses means of the sample after observations outside of the range ($?_m$, $?_x$) have been eliminated.

? ?? ?The true mean height of the population from which the sample is drawn.

$?_{K\&K}$ – mean height of the sample using the K&K estimator: E($?_{K\&K}$) ? ?? ?

$?_{TR}$ – Estimate of ? ?using the maximum likelihood estimator: E($?_{TR}$) = ? ?

$?_{QBE}$ – Estimate of ? ?using the QBE estimator, E($?_{TR}$) = ? ??but $?_{QBE}$ is not an efficient estimator of ? .

$?_{TOLS}$ – Estimate of ? ?using TOLS method: E($?_{TOLS}$) ? ? ?

HR_{min} - minimum height requirement.

HR_{max} - maximum height requirement.

? ???,? ??? ?Normal distribution with mean ? ?and standard deviation ? .

OLS - ordinary least squares regression analysis.

QBE – Quantile Bend Estimator – a method of estimating ? ?and ? ?by estimating the amount of shortfall. It is inefficient.

RSML – reduced sample maximum likelihood estimator – see TR.

? ?? ? the true standard deviation of the population from which the sample is drawn.

shortfall – missing observations from the sample on account of the HR.

$?_m$ – minimum truncation point.

$?_x$ - maximum truncation point.

TOLS – truncated ordinary least squares; linear regression analysis after the values of the sample outside of the range ($?_m$, $?_x$) have been eliminated.

TR – truncated regression: uses maximum likelihood procedure to estimate ? ??its standard error, as well as ? ; previously referred to as RSML, reduced sample maximum likelihood or as TML, truncated maximum likelihood.

truncated sample – the sample obtained after the observations outside of the range ($?_m$, $?_x$) are eliminated.

truncation point – the value at which shortfall begins and where the sample should be truncated.

Appendix B. The Truncated Maximum Likelihood Estimate

Let x_i, $i = 1,..., n$, the n observations in a (sub)-sample fully truncated below $?_m$. The density of this sample is the product of the individual densities $f(x_i)$, which are densities of a normal distribution with mean ? and a standard deviation ?, normalized by the factor $1-?\ [\frac{?_m ? ?}{?}]$, so that the density integrates to one over the range above the truncation point. ? is the standardized normal distribution function. The estimates can be found numerically by maximizing the sample density subject to the unknown parameters ? and ? (maximum likelihood estimation).

With rounding, the likelihood function to be maximized is given by:

$$L = \underset{i?1}{\overset{n}{?}} \frac{\frac{1}{?\sqrt{2?}} e^{[?\frac{1}{2?^2}(x_i??)^2]}}{1??\ [\frac{?_m ????}{?}]}$$

, where ? the linear measure (inch, cm) of the rounding interval below the truncation point (Mokyr and O´Gráda. 1996). That is, in case of the 18th century French military $?_m$ was 62 French inches (F.i.), and we assumed that ? = 0.25, i.e., that men whose height equaled 61.75 F.i. were allowed to pass muster, by having their height rounded up to 62 F.i. This is a crucial assumption, insofar as it increases the height estimates by about 1 cm. Even if the assumption is arbitrary – it is plausible, and should be implemented as the best guess estimate.

Appendix C. Density Estimation (Smoothing)

To estimate the continuous density of a sample X = (x_i, $i = 1,..., n$), perform the following steps Scott (1992, ch. 6): Calculate the optimal bandwidth $h = 2.78\ ?_e n^{-1/5}$, where $?_e$ is the estimated standard deviation.[47] Define a stepwidth s (for heights in cm $s = 0.1$, for heights in inches $s = 0.05$ is a reasonable choice). With this stepwidth generate points P_j, $j = 1,...,$ the smallest element being min(X) - h, the largest about max(X) + h. Calculate the estimated density on every point P_j by weighting the observations X on each of these points: $f_e(P_j) = \frac{1}{\mathbf{nh}} \sum_{\mathbf{i}=1}^{\mathbf{n}} \mathbf{K}[\frac{\mathbf{P_j} - \mathbf{x_i}}{\mathbf{h}}]$. The function $K(u)$ is $\frac{15}{16}(1 - u^2)^2$ if $|u| \leq 1$. Otherwise $K(u)$ is zero.

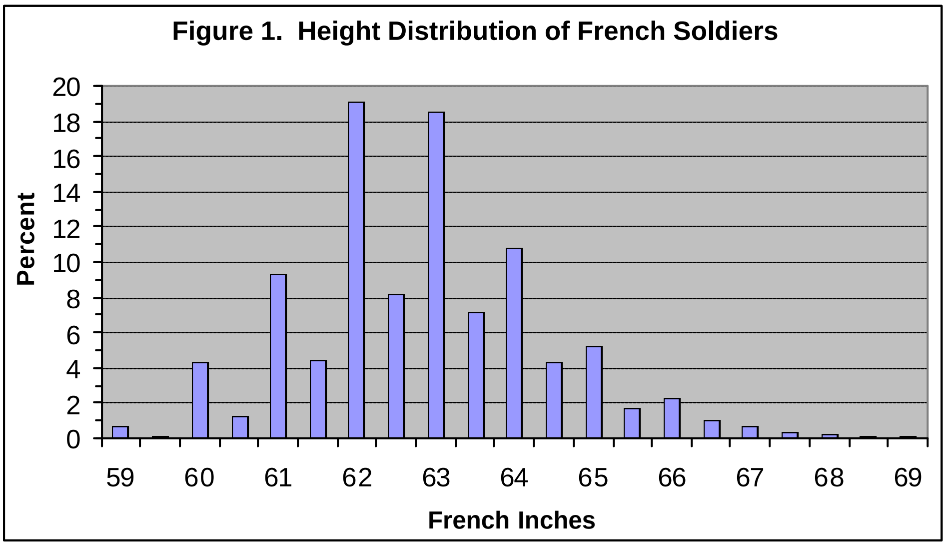


Figure 1. Height Distribution of French Soldiers
20
18
16
14
12
10
8
6
4
2
0
Percent
59
60
61
62
63
64
65
66
67
68
69
French Inches

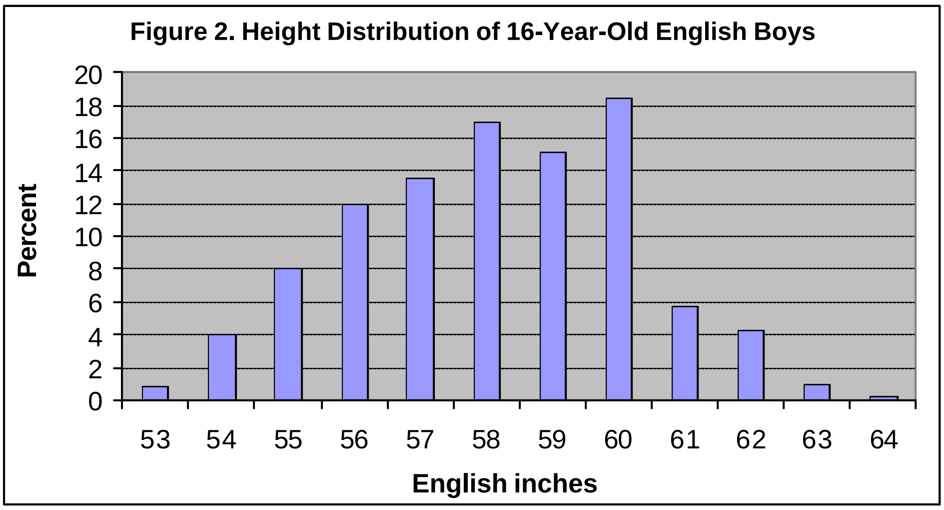


Figure 2. Height Distribution of 16-Year-Old English Boys
20
18
16
14
12
10
8
6
4
2
0
Percent
53
54
55
56
57
58
59
60
61
62
63
64
English inches

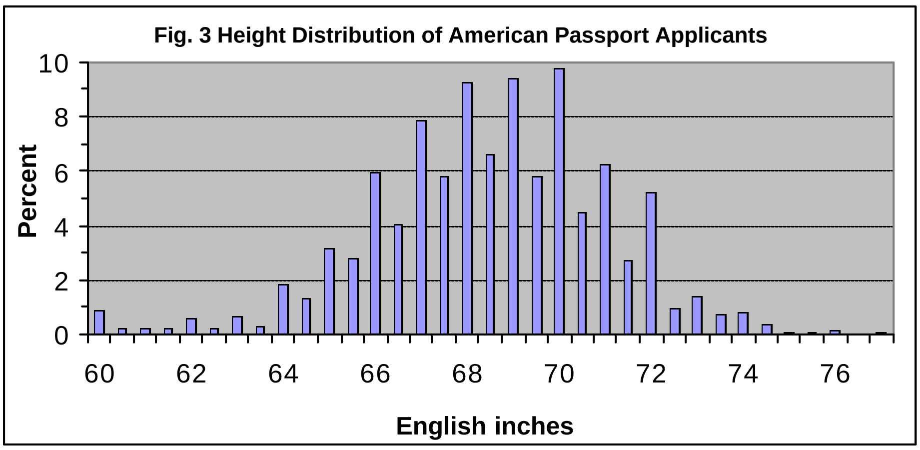
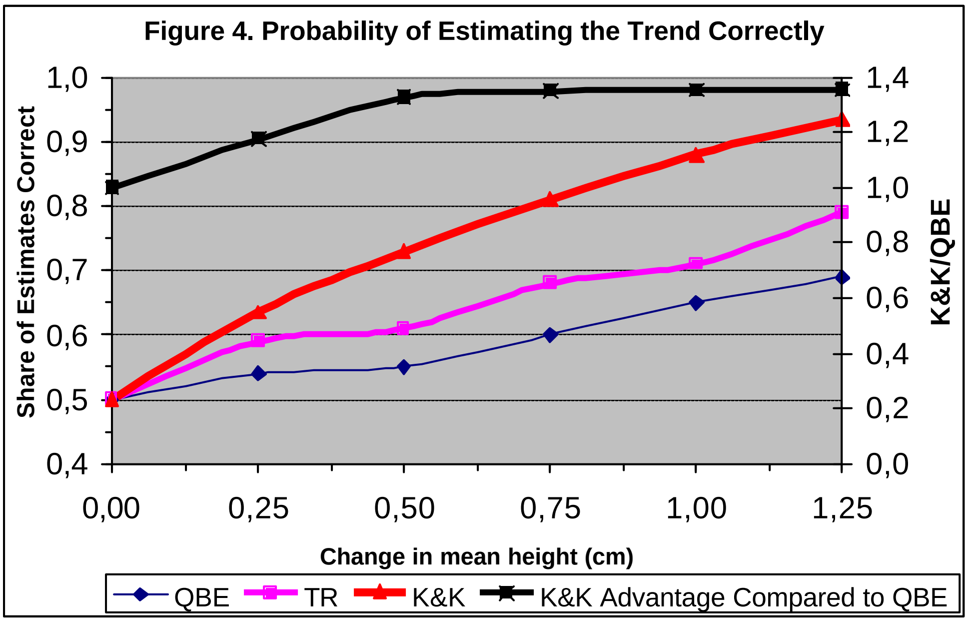
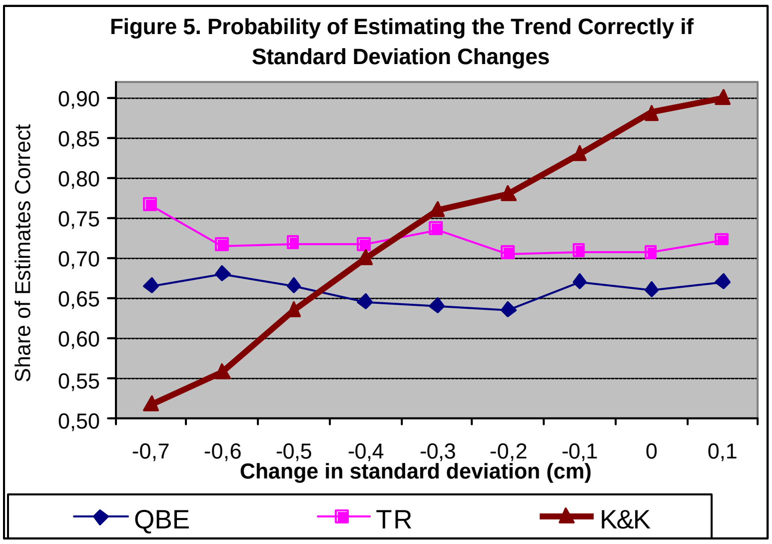


Fig. 3 Height Distribution of American Passport Applicants
10
8
6
4
2
0
Percent
60
62
64
66
68
70
72
74
76
English inches

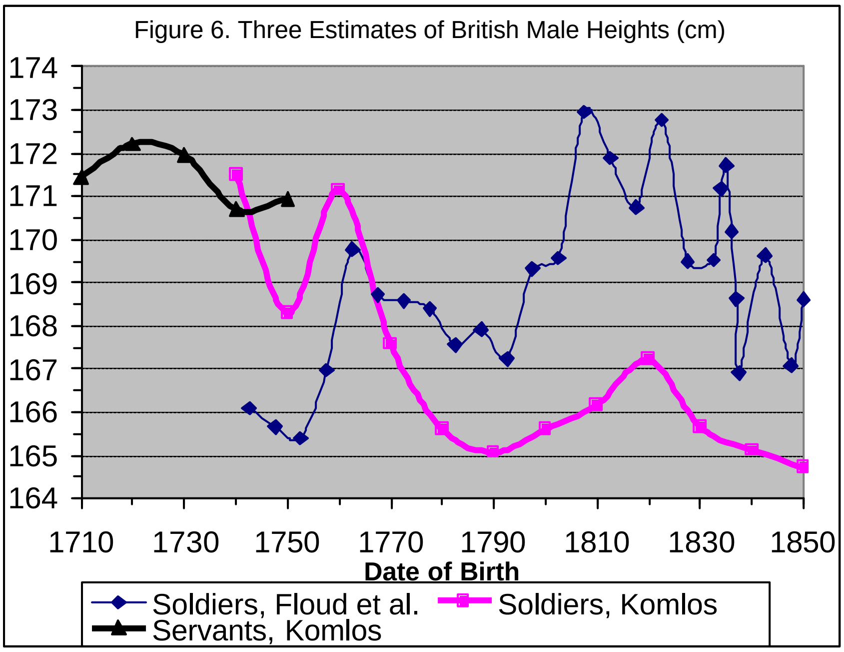
- -



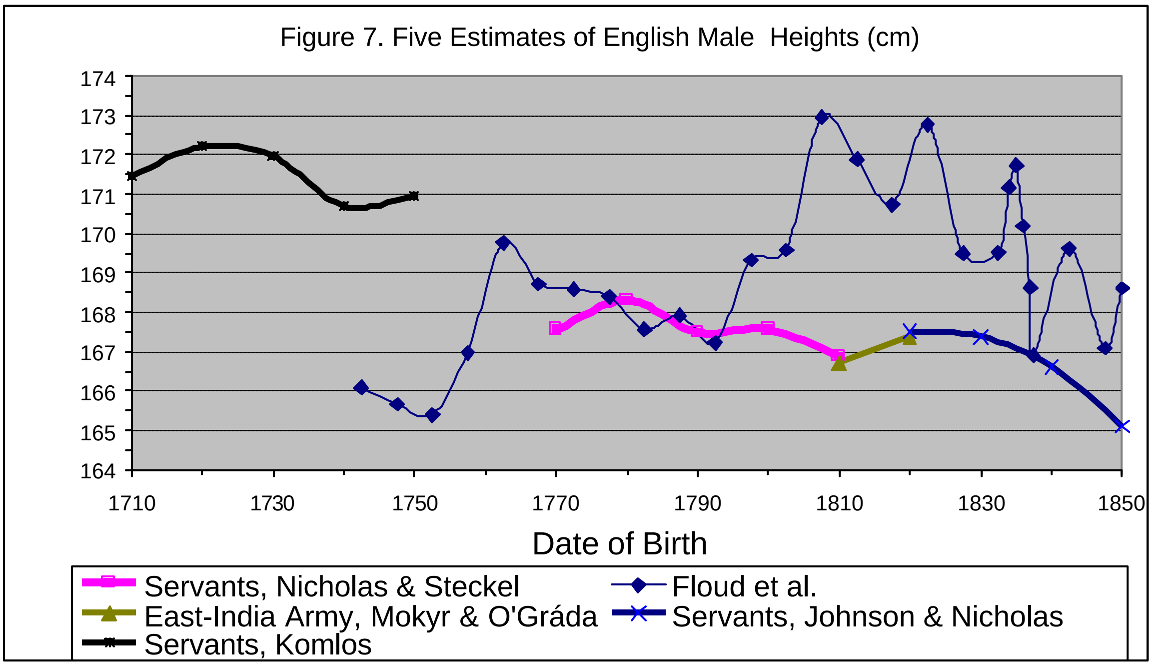
Source: Heintel, 1996b



Source: Heintel, 1996b



Sources: Floud, Wachter and Gregory, 1990, p. 148; Komlos, 1993, 1999.



Sources: Floud, Wachter and Gregory, 1990, p. 148; Johnson and Nicholas, 1995; Komlos, 1993, 1999; Mokyr and O´Gráda, 1994, 1996; Nicholas and Steckel, 1991.

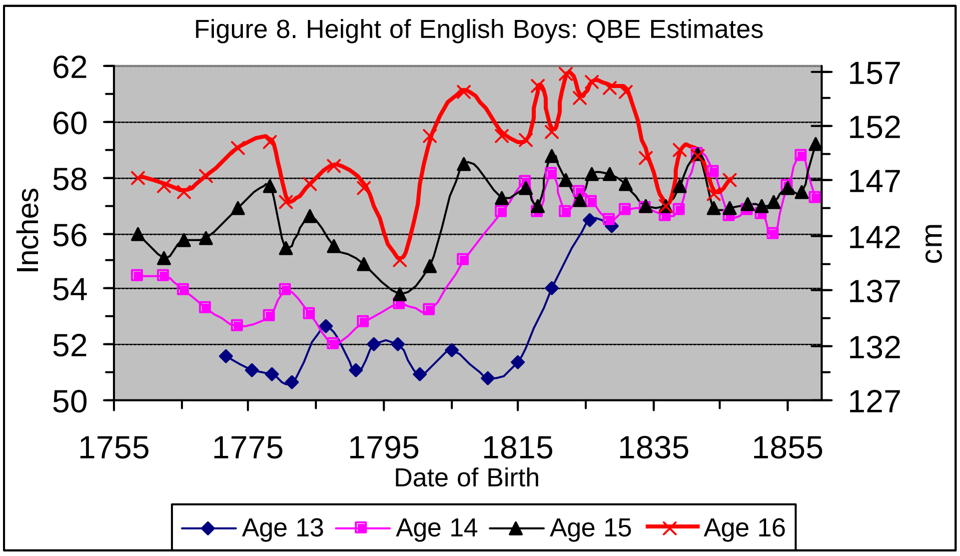


Figure 8. Height of English Boys: QBE Estimates

Source: Floud et al., 1990, p. 167.

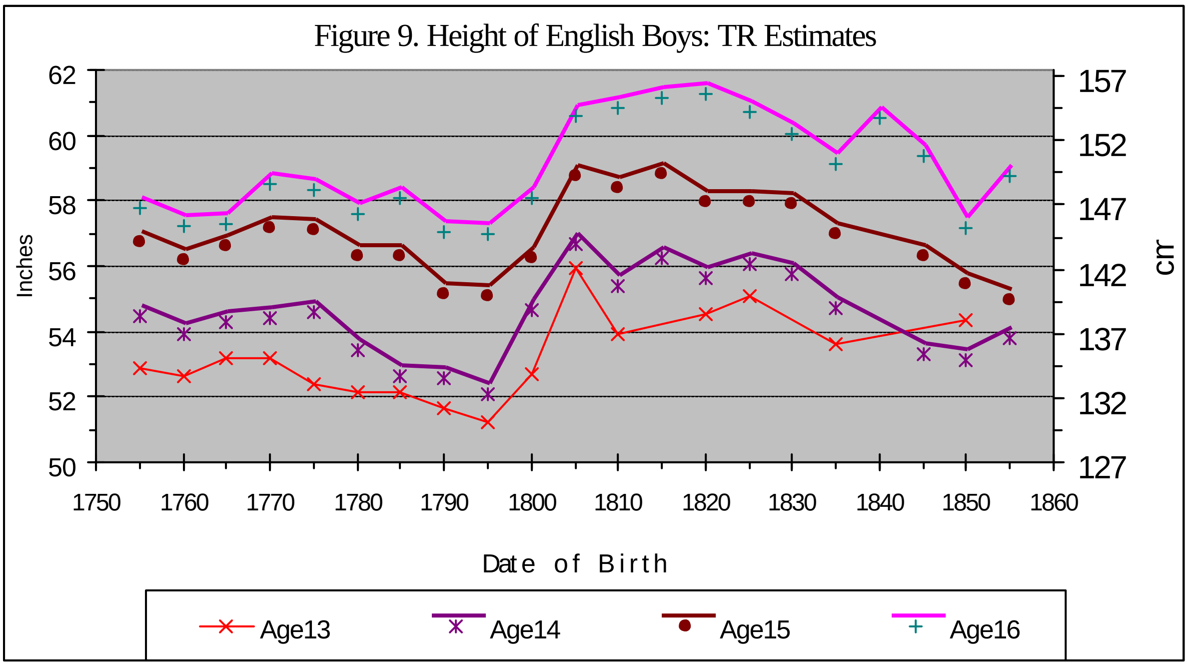


Figure 9. Height of English Boys: TR Estimates

University of Essex. ESRC Data Archive, Marine Society Data Set.

Table 1. Deficient Height Samples: Comparison of the Four Estimators

	QBE	K&K	TR	TOLS
Reliability	poor	excellent	very good	excellent
? unbiased in theory	yes	no	yes	no
? unbiased in practice	no	no	yes	no
estimates ?	yes	no	yes	no
needs ? for ? ?	no	yes	no	yes
Needs HR	no	yes	yes	yes
Double truncation ok	no	yes	yes	yes
Program needed	special	any	Stata, Eviews	any
Easy to use	no	yes	medium	yes
Estimates covariates	no	no	yes	yes
Recommended	no	yes	yes	yes

Table 2. Bias and Variance of ? (cm): Comparison of the QBE and TR Procedures

Continuous Distributions[a]

		Means of						
n	N	Bias of ?		Variance of ?		Mean Square Error		
		QBE	TR	QBE	TR	QBE	TR	Ratio
1000	12	-0.40	+0.21	2.89	0.83	3.05	0.87	3.52

Discrete Distributions[b]

n	N	Bias of ?		Variance of ?		Mean Square Error		
		QBE	TR	QBE	TR	QBE	TR	Ratio
1,000	16	-0.25	+0.13	1.66	0.96	3.24	0.99	3.26

Note: n = number of times the simulation was run with each specifications;
N = number of various specifications that were estimated with different sample sizes and amounts of shortfall (shortfall: 10-30 percent; sample sizes: 250-500);
[a] True mean of the population from which the samples were drawn was 165 cm with a ? of 6.5 cm.
[b] The amount of shortfall was between 5 and 45 percent. The true mean of the population from which the samples were drawn was 65 inches with a ? of 2.6 inches. The samples were rounded to the nearest inch.
Source: Heintel, 1996a, 1998.

Table 3. The Accuracy of Trend Estimates. The Probability of Obtaining Correct Direction of Change with Three Methods

	Probability				K&K
N	QBE	TR	K&K	n	Advantage
10,000	0.64	0.69	0.84	250	0.15 - 0.20
10,000	0.67	0.74	0.91	500	0.17 - 0.24
Average	0.65	0.71	0.87		

Note: N = number of pairs of samples for which direction of change was estimated with all three methods; n = sample size of each simulation. Simulations were made with 20 different specifications, with 1,000 simulations each. The true mean of the population from which samples were drawn increased in 9 specifications from 165 cm to 166 cm, and from 165 to 165.5 in one specification. The ? was 6.5 cm throughout. Shortfall varied from 10 to 50 percent, and the truncation points from 160 to 165 cm.
Source: Heintel, 1996b.

References

A'Hearn, B. 2004. A Restricted Maximum Likelihood Estimator for Truncated Height Samples. *Economics and Human Biology* 2, 1, forthcoming.

Baten, J., and M. Heintel. 1995. Zum Problem der Verteilung mit Shortfall bei der Nutzung des Indikators 'Durchschnittliche Körpergröße'. *Historical Social Research* 20 4:135-154.

Bogin, B.. 1999. *Patterns of Human Growth.* Cambridge: Cambridge University Press, 2nd edition.

Cheung C., and A. Goldberger. 1984. Proportional Projections in limited Dependent Variable Models. *Econometrica* 52:531-34.

Chay, K.Y., and J.L. Powell. 2001. Semiparamertic Censored Regression Models. *Journal of Economic Perspectives* 15, 4: 29-42.

Cole, T.J. 2003. The secular trend in human physical growth: a biological view. Economics and Human Biology 1, no. 2: 161-168.

Craig, L., and T. Weiss. 1998. Nutritional Status and Agricultural Surpluses in the Antebellum United States, in John Komlos and Jörg Baten, eds., *The Biological Standard of Living in Comparative Perspective*, Stuttgart: Franz Steiner, pp. 190-207.

Cuff, T. 1998. Variation and Trends in the Stature of Pennsylvanians, 1820-1860, in John Komlos and Jörg Baten, eds., *The Biological Standard of Living in Comparative Perspective*, Stuttgart: Franz Steiner, pp. 208-235.

DiNardo, J., and J.L. Tobias. 2001. Nonparametric Density and Regression Estimation. *Journal of Economic Perspectives* 15, 4: 11-28.

EViews 4 User's Guide. 2000. Irvine CA: Micro Software.

Floud, R., and K. Wachter. 1982. Poverty and Physical Stature: Evidence on the Standard of Living of London Boys 1770-1870. *Social Science History* 6: 422-52.

Floud, R., K. Wachter, and A. Gregory. 1990. *Height, Health and History*. Cambridge: Cambridge University Press.

Freiman, J.A., T.C. Chalmers, H. Smith, and R.R. Kuebler. 1978. The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial. *New England Journal of Medicine*, 299 690-4; cited in C.G. Nicholas Mascie-Taylor, Statistical Issues in Anthropometry. in S.J. Ulijaszek and C.G.N. Mascie-Taylor eds., Anthropometry: the Individual and the Population. Cambridge University Press, 1994, 56-77.

Frisancho, A.R. 1990. Anthropometric Standards for the Assessment of Growth and Nutritional Status. Ann Arbor: University of Michigan Press.

Greene, W.H. 1993. *Econometric Analysis*. New York: MacMillan.

Haines, M.R. 1998. Health, Height, Nutrition, and Mortality: Evidence on the ‚Antebellum Puzzle‘ from Union Army Recruits for New York State and the United States. in John Komlos and Jörg

Baten, eds., *The Biological Standard of Living in Comparative Perspective*, Stuttgart: Franz Steiner, pp. 155-180.

Harris, B. 1994. Health, Height, and History: An Overview of Recent Developments in Anthropometric History. *Social History of Medicine* 7: 297-320.

Heintel, M. 1996a. Historical Height Samples with Shortfall: A Computational Approach. *History and Computing* 8: 24-37.

Heintel, M. 1996b. Estimating Means and Trends of Normal Samples with Shortfall. Working Paper no. 96-20, Department of Economics, University of Munich.

Heintel, M. 1997. Über einige statistische Probleme und Möglichkeiten der anthropometrishen Geschichte. in Jörg Baten and Markus Denzel (eds). *Wirtschaftsstruktur und Ernährungslage, 1770-1870* (St. Katharinen: Scripta Mercaturae Verlag, 1997), pp. 108-130.

Heintel, M., and J. Baten. 1998. Smallpox and Nutritional Status in England, 1770-1873: On the Difficulties of Estimating Historical Heights. *Economic History Review* 51: 360-371.

Heintel, M., L. Sandberg, and R. Steckel. 1998. Swedish Historical Heights Revisited. New Estimation Techniques and Results. in John Komlos and Jörg Baten, eds., *The Biological Standard of Living in Comparative Perspective*, Stuttgart: Franz Steiner, 1998, pp. 449-458.

Johnson, P., and S. Nicholas. 1995. Male and Female Living Standards in England and Wales, 1812-1857: Evidence from Criminal Height Records. *Economic History Review* 48, no. 3: 470-81.

Komlos, J. 1985. Stature and Nutrition in the Habsburg Monarchy: The Standard of Living and Economic Development in the Eighteenth Century. *American Historical Review* 90: 1149-61.

Komlos, J. 1989. *Nutrition and Economic Development in the Eighteenth Century Habsburg Monarchy* (Princeton: Princeton University Press).

Komlos, J. 1993. The Secular Trend in the Biological Standard of Living in the United Kingdom. *Economic History Review* 46: 115-44.

Komlos, J. 1995. *The Biological Standard of Living in Europe and America 1700-1900*. Aldershot: Variorum Press.

Komlos, J. 1996. Anomalies in Economic History: Toward a Resolution of the 'Antebellum Puzzle'. *Journal of Economic History* 56: 202-214.

Komlos, J. 1998. Shrinking in a Growing Economy. The Mystery of Physical Stature during the Industrial Revolution. *Journal of Economic History* 58 (1998) 3: 779-802.

Komlos, J. 1999. On the Nature of the Malthusian Threat in the Eighteenth Century. *Economic History Review* 52, no. 2: 730-48.

Komlos, J., in collaboration with M. Hau and N. Bourguinat. 2003. The Anthropometric History of Early-Modern France, 1666-1766. *European Review of Economic History* 7: 159-189.

Komlos, J., and J.H. Kim. 1990. Estimating Trends in Historical Heights. *Historical Methods* 23: 116-120.

Komlos, J., and J. Baten. 1998. *The Biological Standard of Living in Comparative Perspectives: Proceedings of a Conference Held in Munich January 18-23, 1997*. Stuttgart: Franz Steiner Verlag.

Komlos, J., and T. Cuff. 1998. *Classics of Anthropometric History*. St. Katharinen: Scripta Mercaturae.

Le Roy Ladurie, E., N. Bernageau, and Y. Pasquet. 1969. Le Conscrit et l'ordinateur: Perspectives de recherches sur les archives militaires du XIXe siècle francais. *Studi Storici* 10: 260-308.

Livi, R. 1883. Sulla Statura degli Italiani," *Archivio per l'antropologia e la etnologia* 13: 243-90, 317-379.

Livi, R. 1896. *Antropometria militare*. Rome: Presso il Giornale medico del Regio Esercito.

Mokyr, J., and C. O´Gráda. 1994. The Heights of the British and the Irish c. 1800-1815. Evidence from Recruits to the East India Company's Army. In *Stature, Living Standards, and Economic Development: Essays in Anthropometric History*, edited by J. Komlos, Chicago: The University of Chicago Press, pp. 117-28.

Mokyr, J., and C. O´Gráda. 1996. Height and Health in the United Kingdom 1815-1860: Evidence from the East India Company Army. *Explorations in Economic History* 33: 141-68.

Nicholas, S., and R.H. Steckel. 1991. Heights and Living Standards of English Workers during the Early Years of Industrialization, 1770-1815. *Journal of Economic History* 51, no. 4: 937-57.

Preece, M.A., and M.J. Baines. 1978. A New Family of Mathematical Models describing the Human Growth Curve. *Annals of Human Biology* 5: 1-24.

Pindyck R.S., and D.L. Rubinfeld, *Econometric Models and Economic Forecasts*, 4th edition, Boston: Irwin McGraw Hill, 1998.

Quetelet, A. 1831. Recherches sur la loi de croissance de l'homme. *Annales d'Hygiène Publique et de Médicine Légale* 6: 89-113.

Salvatore, R., and J. Baten. 1998. A Most Difficult Case of Estimation: Argentinian Heights, 1770-1840. in John Komlos and Jörg Baten, eds., *The Biological Standard of Living in Comparative Perspective*, Stuttgart: Franz Steiner, pp. 90-96.

Sandberg, L., and R. Steckel. 1987. Heights and Economic History: The Swedish Case. *Annals of Human Biology* 14: 101-10.

Scott, D.W. 1992. *Multivariate Density Estimation*. New York: Wiley.

Stata Reference Manual. Release 7. 2001. College Station, TX: Stata Press, vol. 4.

Steckel, R. 1979. Slave Height Profiles from Coastwise Manifests. *Explorations in Economic History* 16: 363-380.

Steckel, R. 1995. Stature and the Standard of Living. *Journal of Economic Literature* 33: 1903-40.

Tanner, J.M. 1978. *Foetus into Man: Physical Growth from Conception to Maturity.* Cambridge: Cambridge University Press.

Tanner, J.M. 1981. *A History of the Study of Human Growth*. Cambridge: Cambridge University Press.

Trussell, J., and K. Wachter. 1984. Estimating the Covariates of Historical Heights. NBER Working Paper No. 1455.

Twarog, S. 1997. Heights and Living Standards in Germany, 1850-1939: The Case of Württemberg. In *Health and Welfare during Industrialization*, edited by R Steckel and R. Floud. Chicago: The University of Chicago Press, 285-330.

Villermé, L.R. 1829. Mémoire sur la taille de l'homme en France. *Annales d'Hygiène Publique et de Médicine Légale* 1: 551-59.

Wachter, K. 1981. Graphical Estimation of Military Heights. *Historical Methods* 14: 31-42.

Wachter, K., and J. Trussell. 1982. Estimating Historical Heights. *Journal of the American Statistical Association* 77: 279-293.

Endnotes

[1] For example, prison samples pertain essentially to the lower classes of the society, while passport samples in the 19th century are likely to contain a disproportionate number of people from the middle class.

[2] Because of regulations and financial constraints, archival sampling can be very challenging, insofar as there are limits on the number of documents one can obtain in a given time, due to the shortage of archivists. Under such circumstances it can be very difficult to ascertain at the outset even how many data are available, and it can be often impractical to obtain a truly random sample under conditions that prevail in many archives.

[3] Yet, problems may arise even in case of universal conscription, if the individual data are unavailable and one is working with published sources. There are several such sources of summary results of heights distributions in categories. George Alter suggests that some of the Italian height distributions found in Livi (1883, 1896) are quite distorted, but not because of the minimum height requirement. Apparently recruits who were initially shorter than the minimum height requirement at mustering were asked to return the following year. Hence, some people were measured twice, and there may be double counting near the MHR of 155 cm. In addition, one cannot hold age constant in these published height distributions, as it contains height at different ages. Moreover, there might be heterogeneity in such summary statistics that cannot be controlled for, since the occupation of the conscripts are unknown, and the obtained height distribution might be a mixture of normal distributions. Thus, there are many more problems to be worked on from a statistical point of view. We do not claim that we have solved all problems pertaining to deficient height samples.

[4] In addition, in the presence of rounding, the statistical tests of normality (such as the Jarque-Bera test) of even complete height distributions lose power, that is, are ineffective.

[5] The fact that the HRs were not enforced with stringency implies that the samples themselves are not truncated. In practice, the researcher has to truncate the sample after identifying the HRs. Truncation is different from the problem of censoring. "… Suppose for example, that we are studying the wages of women. We know the actual wages of those women who are working, but we do not know the 'reservation wage' (the minimum wage at which an individual will work) for those who are not. The latter group is simply recorded as not working. Or suppose that we are studying automobile purchasing behavior using a random survey of the population. For those who happened to buy a car, we can record their expenditure, but for those who did not we have no measure of the maximum amount they would have been willing to pay at the time of the survey. In both of the examples just described, the dependent variable is censored: information is missing for the dependent variable, but the corresponding information for the independent variables is present. (If both kinds of data are missing, we describe the dependent variable as truncated)..." (Pindyck and Rubinfeld, 1998, p. 325). However, in anthropometric research, in most cases we do not have information on the number of people who had been rejected from military service.

[6] For example, the French military had a HR_{min} of 62 French inches (F.i.) (or 167.81 cm), implying that men shorter than this height had a lower probability of being accepted into the infantry and, therefore, into the sample. The mean height of the soldiers was about 171.5 cm, but this obviously does not pertain to the height of the population of able-bodied men living in France at the time: this mean is upwardly biased. However, one can use this information to estimate the mean height of the population of all men by assuming that 18th century heights were also normally distributed with a standard deviation (?) equal to that of modern adult populations (about 6.86 cm). On the basis of this assumption one can calculate that the mean height of the male population must have been about

161.7 cm. That, in turn implies, that 82.5 percent of the population must have been shorter than the HR_{min} of 167.8 cm. However, that is far from being the case in the sample: only 20.5 percent of the sample is to the left of 167.8 cm, and the difference between the two values (62 %) is the amount of shortfall, i.e., 62 % of the observations below 167.8 cm is missing from the data set. The ? of height distributions is remarkably similar across populations and across time. That of black and white males between the ages of 18 and 25 is 7.0 in the contemporary US, and that of females is 6.5 among blacks and 6.4 among whites (Frisancho, 1990, pp. 144, 164).

[7] Deviations from the normal distribution might be due to insufficient sample size. Thus, in case of archival sampling, one should examine the histograms during the sampling process, in case the sample needs to be enlarged. However, one cannot afford to disregard observations outside of the range ($?_m$, $?_x$) already at the archival stage. Though that practice might save some time in collecting data, one does need to collect a complete sample insofar as in the early stages the researcher only knows the legal HRs, and does not know the extent to which practice deviated from them.

[8] The value at which shortfall begins is referred to as the truncation point, even if the HR is never perfectly enforced.

[9] Heintel and Baten (1998) describe some statistical difficulties with these data. Boys of this age who were extraordinarily tall might have been deemed unsuitable for life at sea or might have had the opportunity to go directly into the labor force, and might not have required the charitable support of the Marine Society.

[10] There might be slight deviations from normality due to rounding. Americans who traveled abroad applied for passports as proof of citizenship. There were obviously no HR for traveling abroad, so one would not expect such a height distribution to be truncated. The distribution of height of a sample

of 2040 male adult (older than 22 years) applicants, extracted from archival records in Washington D.C. confirms that the heights were not truncated (Figure 3). Thus, the sample can be analyzed as usual. Nonetheless, it is evident that there are departures from a normal distribution due to rounding on 70 inches. Heaping on favorite numbers is observed in almost all height samples, even in modern ones. Thus, slight departures from normality, even in the absence of height requirements is possible, usually due to rounding on "attractive" numbers. However, such biases tend to have only a marginal effect on results, and can be ignored generally (Komlos 1999).

Figure 3 about here

[11] It might be the case, say, that 67 inches has fewer observations than warranted, but the rounding to 68 and 66 inches would tend to cancel each other. However, there might be some situations when rounding does have a significant impact on the analysis. For instance, in an Argentinean data set a large number of the height of recruits were recorded as being exactly equal to the HR_{min} (Baten and Salvatore, 1998). Obviously, the height of recruits was rounded up to the HR_{min} in order to enable them to enter the military. In order to obtain accurate estimates, the sample had to be truncated one unit above the HR_{min}. This example illustrates again the utmost importance of a visual inspection of the height distributions, in order to ascertain data anomalies.

[12] Heaping might also affect information on age. Ex-slaves of the union army, for instance, did not know their age exactly. Hence, one can observe heaping on certain ages, and the age-by-height profile cannot be very accurate.

[13] This is also the case because the truncation point can coincide with or be to the right of the mean.

[14] A visual inspection of the distributions suffices in most cases. However, Heintel (1996a, 1998) developed procedures to estimate truncation points, based on the fact that the increase in sample

density is steepest at ?, because this point is where the distribution changes discontinuously from a complete normal distribution to one with a shortfall (Appendix C). However, this procedure does not tend to produce superior estimates to visual inspection.

[15] Youth and adult height distributions should also be inspected separately. This is the case, because the HR_{min} were at times not as stringently applied to youth, inasmuch as they were expected to grow subsequently.

[16] Floud, Wachter and Gregory (1990) merged data from the British Army and from the Royal Marines even though they had different HRs, which also were enforced with different degrees of stringency. Additionally, the Royal Marines had a maximum height requirement. Therefore, their results fluctuated unreasonably and are inaccurate (see Figures 6 and 7 below) (Komlos, 1993).

[17] An adult born in 1740, for instance, was influenced by the nutritional circumstances between 1740-63, while an 18-year-old born in the same year was influenced by those of 1740-1758. Hence, they do not overlap for the years 1759-63.

[18] In the French example, cited above, the HR_{min} was lowered to 60 French inches after 1740.

[19] For 18th-century samples, one might also consider examining the histograms for different regiments in order to see if there were some anomalies in the recruitment procedures used in the field, because there was still considerable local autonomy in recruiting.

[20] The sample size (n) needed, for a given degree of accuracy, can be calculated from the following equation: $\mathbf{n} ? \frac{2?^{2}(\mathbf{Z}_{1??} ? \mathbf{Z}_{1??})^{2}}{\mathbf{x}^{2}}$, where ? is the standard deviation, ? the is desired level of significance, ? ?is the power of the test (defined as the probability of correctly rejecting the null hypothesis, if it is false), and x is the difference between two means (Freiman et al, 1978, 58).

Hence, the sample size needed to ascertain a 1 cm difference at a 0.05 significance level, for ? ?= 0.7, would be: $\mathbf{n} ? \frac{2(6.8)^2(1.96 ? 0.5244)^2}{1^2}$ = 571 observations in each cell. Admittedly, this sample size is often impractical in real world situations given financial and archival constraints, but at least it gives one a sense of the ideal sample size to strive for.

[21] The appropriate time interval (annual, quinquennial, decadal, quarter century, etc.) is determined by the number of observations available for analysis. The number of observations per interval should be several hundred for robust estimates.

[22] Although the occupation of an adult soldier would not have had an effect on his height during his growing years, this variable has been found to be significant insofar as social mobility was limited, and therefore, the soldier's occupation can serve as a proxy of his family's social status during childhood and adolescence. In some cases, height becomes a determinant of occupation – such as lumberjacks, for instance. In that case height is no longer a valid dependent variable, since it determines occupation, and not the other way around. In that case lumberjacks might be subsumed under a larger category, such as lower class blue collar occupations.

[23] One has to make sure that the explanatory variables entered into the analysis are legitimate in the sense that they are variables that could have affected the height of the soldiers prior to adulthood. The year of recruitment during the American Civil War has been used as an explanatory variable, although it is not actually pertinent to the determination of the soldiers' height. The height of the soldiers was not determined by the year of recruitment, in the same sense as the income of the soldier's parents was a determining factor in the height of the soldiers. In contrast, different unobserved recruitment procedures (or self-selection) meant that a particular birth cohort was sorted into different recruitment years by

height. In other words, heights could have determined enlistment year, and not the other way around. Hence, recruitment year is not a useful right hand side variable in that case.

[24] The discrete histograms can be turned into continuous smooth distributions using a kernel density estimator (which incorporates a smoothing function), as height is a continuous variable (see Appendix C). The continuous distribution so obtained can also be used to estimate ??? as well as the mode, which in turn is an estimator of ?, as the mode and the mean are identical in a normal distribution. Successive such estimates provide trend estimates of the mean height. The accuracy of the kernel density estimator of ? ? has not been explored. The method also does not allow for estimation of the effect of covariates on heights (Heintel 1996a, 1998; DiNardo and Tobias 2001).

[25] It also provides estimates of the variance of the height of the population. For the QBE procedure the observations below the HR_{min} do not have to be discarded.

[26] The algorithm estimates the amount of shortfall by minimizing the bending in a quantile-quantile plot. The method completes the sample in such a way that the part of the plot above the truncation point forms a straight line. This line is estimated by means robust regression. If *a* is the amount of shortfall, ? is the standard normal distribution function and F(y) is the empirical distribution function of the sample at point y, then the aim is to estimate *a* such that the line: $\mathbf{x} ? ?^{?1}[(1? \mathbf{a})(1? \mathbf{F}(\mathbf{y}))]$ is as straight as possible. Yet, because we usually have only a handful of points, the estimation of the straight line has a relatively large variance, and a few additional observations can make a large difference in the estimate (Komlos, 1989).

[27] If one would like to estimate urban heights distinctly from rural heights, one has to first divide the sample into urban and rural provenance before proceeding with the K&K analysis. One cannot do it in one step as with TR or with TOLS by including a dummy variable for urban heights.

[28] This is the case, because ? ?is a monotonic function of ? $_{K\&K}$. In other words, if ? $_{1,K\&K}$< ? $_{2,K\&K}$ then it follows that ? $_1$<? $_2$.

[29] The truncated regression is essentially the model presented in Appendix B replacing the overall mean ? by the individual mean in a particular category.

[30] By eliminating the sample outside of the range (? $_m$,? $_x$) the artificial impact of potentially different shortfall patterns in the sub-samples (caused by different truncation points and/or different amounts of shortfall) is avoided by "equalizing" the bias over the complete sample (Heintel and Baten, 1998, footnote 17, is a example of artificial correlations if one fails to equalize the bias). If ? $_i$ is the true coefficient vector, *i* = 1,... n, and ? $_i^*$ the TOLS coefficient vector, then ? $_i^*$ = ? $_i$ (?) with 0 < ? < 1, and the bias is given by ? $_i^*$ - ? $_i$ = ? $_i$ (?-1) (Cheung and Goldberger, 1984). This implies that the sign, and the relative ordering of the ? $_i$'s are unaffected by the bias caused by truncation. As a consequence, the coefficients of the time dummy variables indicate the true direction of height trend over time. Furthermore, as ? is the same for all the coefficients, one can infer the ordering of the covariates, because the values of the coefficients reflect their importance relative to one another.

[31] The coefficient vector is found numerically by maximizing the likelihood function subject to the unknown parameters (Chay and Powell, 2001). The TR method can estimate ? , and therefore one does not have to assume that ? remain constant as one does if one converts ? $_{TOLS}$ into ? . This is a

disadvantage, however, if $?_m$ is close to, or to the right of ? (A'Hearn, 2004). If that is the case, it is better to constrain the TR regression to a given ? , as in fn. 42.

[32] The distribution for 1740-1762 is not reported here (Komlos 2003).

[33] The recording of height data began in 1716, and the recruiting system remained in effect until the revolution. Only between 1740 and 1762 was the $?_m$ lowered to 60 F.i.

[34] However, the data within a particular HR regime can be used for further analysis. For the French example, the soldiers recruited during the war years 1740-1762 can be also analyzed separately. These results could corroborate the findings obtained with the combined data set. In effect, one can create two samples for analysis. A sample which encompasses the whole period 1716-1786, with an effective $?_m$ of 62 French inches (F.i.), and another sample for the period 1740-62 with $?_m$ of 60 F.i.. Using TOLS soldiers recruited with different HRs should not be conflated, unless the HRs has been equalized between them. They can be analyzed together using TOLS only if the largest of the two $?_m$ is applied to both of them – in this case 62 F.i.. With TR, however, one can use both 60 and 62 F.i. as the appropriate $?_m$ by specifying them in the program as the lower limit for the appropriate time interval.

[35] See, for example, the unstable QBE estimates in Floud and Wachter (1982), Floud Wachter and Gregory (1990), Sandberg and Steckel (1987), Twarog (1997). Heintel, Sandberg & Steckel (1998) confirm the unreliability of the QBE procedure.

[36] The known HR was used in the estimations. 20,000 simulations were run with 10 different specifications: n= 250, 500; ? ?= 6.5 cm and unchanged, ? H = 1 cm in 9 specifications and 0.5 cm in one specification.

[37] None of the methods shows substantial improvements in accuracy with increased sample size (Table 3).

[38] In contrast, the relative accuracy of the TR v.s. the QBE does not change substantially in this range of height changes.

[39] This is also the case with the QBE, but not with TR or TOLS.

[40] That is to say, if height decreases, then ? ?can also decrease, and the advantage of the K&K is still obtained.

[41] Assume that ? = 6.86 cm, the same as for modern populations, and that the effective $?_m$ was not actually equal to 62 F.i. (167.81 cm), as indicated by the histograms, but slightly lower, at 61.75 F.i., (167.14 cm), because it is plausible to suppose that those recruits who were slightly shorter than the $?_m$ were probably allowed to slip through by having their height measurement rounded up to the nearest whole F.i.. (This assumption increases the estimated population means by about 1 cm.) Next take a normal distribution with ? = 170 cm and ? =6.86 cm and discard all observations below 167.14 cm. Then calculate the mean of the truncated distribution, to obtain 173.8 cm. Thus, we can reverse the calculation and assert that if $?_{TOLS}$=173.8 cm then ? =170 cm. In this manner we obtain a conversion schedule for $?_{TOLS(converted)}$. This estimator is unbiased, it is the same as the maximum likelihood estimator, i.e., it has the property that E($?_{TOLS(converted)}$)= ? . In this manner we obtained the following schedule (millimeters):

$?_{TOLS}$	$?_{TOLS(converted)}$
1738	1700
1733	1690
1729	1680
1726	1670
1722	1660

1719	1650
1716	1640
1713	1630
1711	1620
1709	1610
1707	1600

This schedule can be used to obtain (by interpolation) estimates of the height of the French male population from which the soldiers were drawn. 1716.2 mm converts approximately to 1640.1 mm while 1711.8 mm converts to 1623.2 mm. One can also run a linear regression with $?_{TOLS(converted)}$ as the dependent variable and $?_{TOLS}$ as the independent variable in order to obtain a conversion formula. This procedure can also be used to convert the estimates obtained by the K&K method into population height estimates.

[42] In addition, it estimates ? ?of the height distribution (provided that the truncation point is to the left of the mean), rather than assume it as with TOLS.

[43] For instance, in the French example above, one would discard all observations smaller than 60 French inches (F.i.) during the period 1740-1762, and those smaller than 62 F.i. for the remainder of the sample. In this case, we can use two different truncation points.

[44] Anthropometric results should be treated cautiously until corroborated with an independent data set, or with collateral evidence. After all, the socio-economic composition of the institution studied might have varied over time, even in the absence of explicit changes in the admission criteria. This might be due both to supply and demand considerations. The willingness of individuals to enter the military, for instance, might have varied over time. Similarly, the size of the military might have expanded sufficiently so that individuals were accepted who would have been rejected at an earlier time. In either case, the underlying population which provided members of the organization might

have changed over time. This problem is quite intractable. How the supply of, and demand for, potential entrants into an institution fluctuated over time might not be ascertainable at all. Yet, anthropometric history is frequently concerned with long-run changes, so that even social processes which move at a glacial pace might have an impact on the institutions in question. It is advisable to look for other data sets from other types of institutions to corroborate or refute the findings.

[45] Such as insufficient sample size, wrong HRs, not controlling for officers or grenadiers within infantry units, etc.

[46] This very wide margin of error was obtained not only because of the inaccuracy of the QBE procedure, but also because of the fact that Floud et al. combined samples from different units of the military with different height requirements before estimating mean heights. This led to a mixture of normal distributions which further exacerbated the inaccuracy of the QBE program.

[47] For extremely rounded data a modification of the bandwidth might be useful (Heintel and Baten, 1998, footnote 6).