Nikola Koepke and Joerg Baten, University of Tuebingen and CESifo

The Biological Standard of Living in Europe During the Last Two Millennia

JEL: N00, I12, O00

Nikola Koepke
University of Tuebingen
Department of Economic History
Mohlstraße 36
D- 72074 Tübingen
Germany
e-mail: nikola.koepke@uni-tuebingen.de

Joerg Baten
University of Tuebingen
Department of Economic History
Mohlstraße 36
D- 72074 Tübingen
Germany
e-mail: joerg.baten@uni-tuebingen.de
ABSTRACT

This study offers the first anthropometric estimates on the biological standard of living in Europe during the first millennium A.D., and extends the literature on the second millennium. The overall picture drawn from our data is one of stagnant heights. There was no larger-scale progress in European nutritional status over the period studied, not even for the period between 1000 and 1800, for which recent GDP per capita estimates diagnose increasing development. On a smaller level, however, we find that heights stagnated in Central, Western and Southern Europe during the Roman imperial period, while astonishingly increasing in the fifth and sixth centuries. Noteworthy is also the synchrony of height development in the three large regions of Europe. In an exploratory regression analysis of height determinants, population density turns out to have been economically (not statistically) significant and negative, indicating that the relevance of decreasing marginal product theories and Malthusian theory cannot be denied for the pre-1800 period. Of marginal significance, however, were climate (warmer temperatures being favourable for a good nutritional status), social inequality and gender inequality (both reducing average height). Lastly, we also discuss the limitations of our approach.
MAIN QUESTIONS

How did the standard of living in Europe develop in the very long run? In this study, we distinguish the biological standard of living (which includes important elements of the human utility function such as health, longevity, and quality of nutrition) from purchasing power-oriented standard of living components. For our analysis, which covers (mostly skeleton-based) height measurements from 9477 individuals, adult stature was used as a proxy for the biological components stated above. Out of these individual measurements, however, some were published in aggregated form by previous investigators and excavators, with each aggregate figure comprising between one and 360 individuals. Thus, the total number of height estimates available to us was 2974.

Our data were collected from 314 sites all over Europe. In this article, we find important deviations between the biological and purchasing power-related components of living standards over the past 2000 years. For instance, contrary to the widely accepted view that purchasing power should have increased during the period of the Roman Empire due to “Smithian Growth,” we find that mean stature stagnated. Another remarkable finding is that heights did in fact increase during the fifth and sixth centuries, after the breakdown of the Empire in the West.

A related study by Steckel (2004) found a substantial decline in Scandinavian heights from the Middle Ages until today. In a similar vein, Cohen and Armelagos (Armelagos, 1990; Cohen, 1989; Cohen and Armelagos, 1984) have stressed the detrimental influence of civilization on human health and nutritional status over the millennia (until the late 19th century). Maat (2003) has recently confirmed this view for the Netherlands. However, the number of observations covered by these studies is very limited, especially for the periods up to the Middle Ages. Our study, in contrast, will take into account a substantial number of sites and observations in order to trace the biological standard of living in Europe over the last two
millennia. Thus, the number of cases covered by our study compares favourably with earlier research (Table 1).

- Table 1 should appear approx. here -

After describing the sample characteristics and discussing the limitations of our sample and methods, we will - in a first step - estimate a general height trend as well as more disaggregated trends for three regions of Europe. This will be followed by a more detailed analysis of the second estimation’s time series, regressing it on population density, climate, social and gender inequality, and similar variables (the latter with substantial measurement error, however). Finally, we compare our height trends with purchasing power estimates.

DATA AND REGIONAL CHARACTERISTICS
Our height data stem mostly from archaeological excavations and represent the largest collection of observations on Central and Western Europe (see Table 2), covering contemporary Germany, Benelux, Austria, Northern France, Switzerland, and the UK.

- Table 2 -

The number of cases included in our sample is sufficient to subdivide the Central/Western region further into the regions along the Northern Rhine area (Benelux, Northern France, and Western Germany), the regions around the Southern Rhine (South-Western Germany, Switzerland, and Eastern France), Bavaria/Austria, and the UK. The category of “Northern and Eastern Europe” refers to those regions which had only limited contact with the Roman Empire and its provincial economy: Scandinavia, Poland, Northern and Eastern Germany, and Hungary, the latter being included here because the large region east of the Danube was not fully integrated into the Roman imperial economy. Northern and Eastern Europe constituted the least densely settled region in our sample, and it probably featured the highest per capita milk and beef consumption in Europe, since a high land ratio facilitates ceteris paribus a large number of cows and therefore a nutrition which is based on
high-quality proteins (on the role of high-quality proteins (such as milk) for pre-industrial nutrition and health, see Baten, 1999; Baten and Murray, 2000; and many others).

These regional characteristics require further substantiation. Recent research on archaeozoology sheds some additional light on our assumption that milk and beef consumption was high in Northern and Eastern Europe. King (1999) and others have analyzed the share of cattle bones, pig bones and other bones (sheep, goats) drawn from excavations of ancient towns, villages, and other civil and military settlements. A high share of cattle bones was typical of Northern and Eastern Europe during the Roman period, whereas a high share of pig bones was typical of the most urbanized part of the Empire, especially the Mediterranean region with its large cities (incl. Rome and the bay of Naples), until approximately the end of the fourth century A.D. The cattle bone share can be taken as approximating the cattle share – King reports no evidence on systematic bias from selective preservation or excavation. A high cattle share is plausible foremost in locations with low population density, since due to ancient technology, a large amount of land was already taken up by agriculture, while even more land was required for grazing. Pigs, in contrast, could be kept in locations with less land-intensive modes of production near large cities, and often be fed with side-products from other lines of agricultural production; in the case of Rome, for instance, even imported fodder was used (in addition to semi-spoiled grain and other foods). Hence, the cattle share might serve as a proxy for the consumption of beef and especially milk, which was of great importance for the protein-supply of low-income groups in pre-industrial societies. What is more, Peters (1998) points out that cow milk was not commonly used in the Mediterranean during Roman times, while bovine milk served in fact as the staple food of some populaces in Central Europe, like a number of tribes in Germania Magna. v.d. Driesch and Peters (2003) explain the missing depiction of udders on a Roman stone relief showing a cattle herd by the fact that dairy-farming was of no importance for Roman husbandry-farmers during the imperial period. In addition, they also mention that Germans preferred cow milk. Ancient
sources such as Tacitus and Plinius (Tac. Germ. 23; Plin. nat. VIII 179) emphasized this as well. Hence, we conclude that milk and beef consumption was considerably higher in Northern/Eastern Europe, and much lower in the Mediterranean already in Roman antiquity. We therefore need to control for those differences in the analysis. Even today, Scandinavians (and Dutch) rank highest in per capita milk and dairy product consumption, with East Asians ranking lowest. Western-Mediterranean dairy consumption was extremely low between the year 0 and ca. 1950 (in parallel to heights), but increased strongly thereafter (Quiroga Valle, 1998; Bisel, 1988). Milk consumption also increased after 1950 – thus, lactose intolerance was probably not a decisive limiting factor in Europe (on Asia, see Crotty 2001).

How many cases are available for the regions and periods? The “Mediterranean” region of our sample includes Italy, Spain, Portugal and the Balkans. Obviously, the small number of observations in Table 2 indicates that more work is to be conducted on the Mediterranean region in the future. For the early Middle Ages, the data available are quite abundant. After the 12th century, height data become more scarce, as bones in cemeteries were more often lost or mixed with bones from later epochs. From the 17th and 18th centuries onwards, archival (written) sources are available which provide much larger sample sizes, while at the same time posing additional selectivity and truncation problems (see Komlos, 2003). Lastly, because the period from the 18th to the 20th century is relatively well studied, we focus mainly on earlier centuries here.

Our sample consists of 2938 female and 6539 male height measurements which are rather equally distributed among all major periods. Only for the 17th and 18th centuries, an insufficient number of cases for women is available (12 and 0, respectively). A large proportion of the total 9477 height measurements were aggregated by the excavators and original investigators. Wherever possible, we collected disaggregated figures. Thus, our final database is comprised of 2972 different height measurements, after discarding extreme heights (<145 cm, > 200 cm). When the dating was imprecise, we used the average of the
earliest and latest date mentioned by the principal investigators, as the real date could have been both before and after the middle of a century. We experimented with estimation techniques granting smaller weight to imprecisely-dated observations or discarding them completely, but the main results remained robust. Because of these data limitations, our units of analysis are restricted to entire centuries. We organized all heights by century of birth and discarded such individuals who were still in the process of growing (< 23 years). Heaping and truncation did not play a large role, as is illustrated by the rather normal distribution of heights (Figures 1a and 1b). We also performed Jarque-Bera and Kolmogorov-Smirnov tests for normality (by century of birth), and found that the distributions of well-documented centuries were all distributed normally, except for the eighth century (details available from the authors).

- Figures 1a and b -

Our intention was to collect as much height data as possible, with the consequence of having to accommodate different types of height information. The majority of measurements were based on excavated long bones (see next section), but some information was also derived from complete skeletons; with such measurements, we relied on the original author’s judgement and adjustments (typically, for instance, 2 cm are added to cadaveric length in order to adjust for disappeared non-bone parts of the body, but none in the case of in situ measurements, as the post mortem stretch is compensated by missing skin: see Maat, 2003).

We also used heights that were estimated using “knight’s” armours from 16th and 17th century Central and Eastern Europe. One might assume that the armours did not fit those wearing them perfectly, but that they were in fact slightly larger in order to allow for some mobility. Fortunately, our data set contains a sufficient number of archaeological height measurements for those centuries, which can be compared to the armour data (the latter covering 12 height numbers or 198 individuals for the 16th century, and 4 height measurements or 105 individuals for the 17th century). The simple average difference between
armour height data and other height data was only about 0.3 cm for those periods, and thus insignificant. Once we controlled for social, regional, and inter-temporal influences in a multiple regression, we even found the difference to be only 0.17 cm (statistically insignificant, results not shown here). We therefore decided that any adjustment for armours should be omitted, since that might introduce an artificial measurement error.

HOW TO ESTIMATE HUMAN STATURE FROM LONG-BONES AND CREMATED BONES

Typically, archaeologists do not find complete skeletons in the (excavated) cemeteries of early settlements, but methods have been developed to estimate human stature from a little as a longbone or even a part thereof. So far, however, anthropologists have not reached full agreement on the optimal estimation procedure, since all available formulae naturally represent a mere approximation of real heights. Nevertheless, the length of a long bone can be regarded as a roughly constant proportion of full body height. The femur serves as the best approximation for reconstructing the height of inhumed skeletons, it luckily being the most frequently surviving bone.\textsuperscript{10} If only a part of the long bone is excavated, the length of the long bone has to be reconstructed first before full body height can be reconstructed. Since two regressions are needed in this case, the measurement error of the result is consequently larger. Fortunately, estimating long bone length with long bone heads (e.g. \textit{caput femoris}) results in a substantial $R^2$ of 0.6-0.8.\textsuperscript{11} Furthermore, completely preserved long bones provide more exact results than cremated bones.

A large number of regression models exist to reconstruct adult stature (see e.g. Herrmann et al., 1990; Rösing, 1988).\textsuperscript{12} In their search for the best model, anthropologists have evaluated various reconstruction models for singular long bones as well as for both genders, all leading to different results (e.g. Bach, Breitinger, Dupertius and Hadden, Manouvrier, Rösing, Trotter, Trotter and Gleser, Olivier et al., Pearson, Wolanski). As of yet,
however, no general consensus has been found, and anthropological analyses of various excavation sites continue to be based on these different reconstruction methods. Therefore, it is of little use to take over the stature estimates of excavation reports directly. What could we do to make our database homogenous? It was necessary to construct algorithms standardizing all height estimates according to the most appropriate method. Most anthropologists and archaeologists have used the following four methods to reconstruct the heights of inhumed individuals: out of 2712 height estimations, 47.9% were conducted using the model by Breitinger (for males) and Bach (for females); for 18.1% of cases, the Trotter/Gleser-model was used, while the Pearson (Wolanski respectively) model was applied in 11.3% of the cases covered by our study. Least frequently, i.e. for 5.9% of observations, Manouvrier’s reconstruction method was used. In contrast, we exclusively used Rösing’s model for cremation analyses, amounting to about 5.1% of observations.

In light of the above, which regression model is most adequate for our purposes and should thus serve as the basis for converting all remaining estimates? As the reconstruction formula introduced by Breitinger/Bach was the one most frequently used in our data, we adopted it as the ‘basic’ model for our estimates. Furthermore, the said formula leads to the most precise results (lowest mean of differences in stature) for the relevant spectrum of centimetres (Formicola, 1993): for the height range of 164 – 178.9 cm for males. Formicola arrives at the lowest estimation error for Breitinger, whereas the Trotter/Gleser and Pearson methods were proven to be less efficient (see also Søjvold, 1990). In our case, 89.3% of male observations fell precisely into the range of 164-178.9 cm. A third argument in favour of Breitinger/Bach’s method is that their height estimates lie in between the higher estimates of Trotter/Gleser and the lower height estimates of Pearson. Thus, the Breitinger/Bach estimates can be regarded as a justifiable “compromise” between the three most widely used methods. To illustrate this, the mean height of males in our sample is 169.7 cm when calculated with Breitinger’s method, compared to 166.1 cm with Pearson’s method and 170.7 cm with
Trotter/Gleser’s approach (for whites). Lastly, the Breitinger and Bach measurements correspond best with Rösing’s measurements of cremated bones, which cannot be converted since the underlying reconstruction formulae are based on different bone parts; in the case of cremated individuals, these are the diameters of long bone heads (e.g. femur F18; all measurements after Martin), while in the case of inhumed skeletons, the length of the long bones is decisive (e.g. femur F1 ‘largest length’).17

DETERMINANTS OF ADULT STATURE AND ESTIMATION OF HEIGHT TRENDS

Since our height data are not distributed perfectly equally over regions and over time, we run in a first step a number of regression analyses with temporal (century), regional, and other dummy variables. In a second step, we apply panel data analyses on the aggregate level, in order to explain the development of heights in the various regions of Europe.18

Quite clearly, studies based on archaeological data cannot take into account the same number of cases as studies which rely on written sources (those being available from the late 17th century onwards). Nevertheless, we are convinced that invaluable lessons can be learned from a study which brings together a variety of archaeological data and other information, albeit existing limitations. In order to avoid overrating the implications of our data, we should thus also address potential caveats. Measurement errors may occur for a number of reasons. Firstly, as is the case for archaeological findings in general, the dating of skeletons is not always precise. Secondly, excavators were not always able to determine the migrant status (ethnicity) or social stratification attributes of buried individuals. Even gender might be misclassified in some cases – although the latter measurement error should be randomly distributed, given the balance of gender frequencies for the various periods and regions; besides, we accounted for it by marking uncertain cases and performing regressions with and without these observations, obtaining no difference at all in the results. To a certain extent, such sources of measurement error can be controlled statistically, as will be shown below. In
addition, some of these measurement problems are less severe for the late Roman Empire and the early to high medieval periods, for which our sample includes a relatively large number of cases. For the late Middle Ages and the early modern period, however, the small number of cases available to us renders the resulting estimates substantially less reliable.

In what follows, we will discuss the variables used in our research, as well as their potential effects on height. First of all, social status is an important variable in this regard, since many studies on the 18th to 20th centuries have found height differences of typically 2-4 cm between adults from lower classes and adults belonging to the middle and upper classes (Table 3, studies see e.g. Baten, 2000).\(^19\) In our data set, we relied mostly on the original studies’ classification schemes. Thus, unless the skeletons found were of higher social rank, excavation reports did often not consider their social status worth mentioning.\(^20\) Therefore, we assigned dummy variables to the categories of middle and upper class origin only, leaving a “lower or unknown” group for the constant, which in turn implies that we should not over-interpret the coefficient of the social status variable. This variable is not only of importance in itself, however, but is also relevant in controlling for social composition and potential social selectivity when analyzing height trends. Although the bulk of our measurements stems from burial sites where all social strata are represented, we took great care of excluding the influence of potential social selectivity on height trends to the maximum extent possible. In the aggregate analysis, we find that overall middle and upper class heights exceeded the residual group by 0.6 cm (col. 2 and 3 in Table 3), a result which was at best marginally significant, however (p-value 0.11).

Another challenge we faced was to obviate any potential bias due to varying burial customs. In general, the excavation reports incorporated into our sample referred to entire populations as opposed to just a few noble men’s graves. When surveying the archaeological literature, we found no hints on burial customs which could have biased our results significantly, as rich and poor graves seem not to have been exposed to different preservation
conditions on average. As another strategy we pursued to examine this important aspect further was to compare different regions of Europe regarding their burial customs, expecting their height trends to be only imperfectly correlated. However, our investigations revealed similar trends for the different regions – except for a plausible decline in North-Eastern European heights (Little Ice Age). We concluded that time trends were not substantially influenced by local burial customs. Moreover, we took a closer look at some individual sites which were both characterized by a homogenous culture and burial tradition, and had been settled more than one century, finding the previously observed trend confirmed in most cases.

Another factor which is to be controlled for is migration. While some anthropologists are convinced that genetic height potentials play a substantial role in determining the average height of a populace, others have serious doubts whether genetic height potential can really explain any variation in average heights (Bogin, 1988, Mascie-Taylor and Bogin, 1995). Anthropometric historians have found environmental circumstances during bodily growth to be the most crucial factor for determining the variation in mean height. Two points are particularly important in this context, the first one being that most migrants were exposed to different environmental conditions than the autochthonous population during their first years of life. For instance, had they been born in a Northern or Eastern European agricultural environment and migrated to the Mediterranean region later on, we would expect them to be significantly taller than Mediterranean ‘natives’. Secondly, if the immigration rate was high enough, agricultural production techniques might have been transferred to the target region, provided that they turned out to be efficient in the new environment. If those techniques concentrated on cattle farming, there was a strong positive influence on height. We know that the most important migration streams over the period under study moved from the Mediterranean region into Central and Western Europe between the first and third century A.D., while there was also significant Germanic (and other) migration from Northern Europe.
to Eastern, Central and Southern Europe, as well as to the British Isles between the fourth and sixth centuries.

As our results indicate, migrants from the Mediterranean region to Central Europe (mainly Roman soldiers and officers, as well as administrative staff) were on average 4 cm shorter than the rest of the population (Table 3, col.4). Note that the only information available to us for the purpose of identifying this group were Roman burial objects such as balzamaria (flacons) and/or lamps (both necessary instruments for traditional Roman burial ceremonies), so that the respective cases in our sample are mainly representative of a core group of migrants who placed particularly strong emphasis on Roman customs. However, skeletons which could be identified as “Germanic migrants” were not significantly different from Eastern Europeans. Also not statistically significant, but economically relevant was their coefficient in the “Mediterranean” regression: Germanic migrants who died in the Mediterranean region were 1.63 cm taller than the autochtonic population.

As our intention was to provide a height trend estimation on the basis of all the height data available to us, we pooled male and female heights together and controlled for the gender difference by using a dummy variable, assuming that the secular height trends of both genders were moving more or less in the same direction (this assumption will be tested in further depth below). It is interesting to note that the largest difference between genders was found for the least densely populated Northern/Eastern regions of Europe, with the smallest difference prevailing in the Mediterranean region. This may be partially explained by a ceteris paribus increase of gender dimorphism (=gender differential) with average height, a hypothesis which requires further research (Koepke 2002).

- Table 3 -

Which regional differences can be observed when using the dummy variable? The tallest heights were to be found in Northern Europe; Eastern Europe ranked lower in comparison, on a similar level as the Northern Rhine region. Relatively short heights
dominated in the Bavarian/Austrian, Mediterranean and British regions, and in the latter especially during the Celtic-Roman period. Why these regional differences occurred is a question which will be addressed in more detail below. Finally, time dummies allow the formulation of a secular height trend, after controlling for the regional, social, age- and migratory composition of the sample.

- Figure 2 -

Thus, which overall height trend evolves when considering the century dummies of column 2 of the previous regression table (Table 3)? Over the period of the Roman Empire, heights did not increase at all (Figure 2), a finding which stands in contrast to the common view that living standards and especially purchasing power increased during the Roman Empire, due to economic growth and the protection of the pax Romana. What is similarly remarkable is that heights increased further in the fifth and sixth centuries, even after the breakdown of the Empire in the West.

As a result of population growth and adverse climate in the subsequent period, however, adult stature declined again until the 10th century A.D. Thereafter, the medieval warm period of the 11th and 12th centuries was characterized by a favourable height level. Heights collapsed again during the 13th century, but rose between the 14th and 16th centuries, possibly due to better nutritional conditions. The 17th century was a period of nutritional crisis. The estimates for the 16th to 18th centuries are less reliable, however, as their underlying number of cases is relatively small.

If we compare our skeletal measurements with other height estimates based on military archival sources, we find that skeletal male heights in the 18th century moved in the same range as, but were on average slightly higher than military heights: our values of 168 – 170 cm (17th/18th c.) correspond with the 164 and 172 cm estimates of the military samples, respectively. 17th century military heights were even lower. Thus, a height increase between the 17th and 18th century is supported by our skeleton height trend (see Komlos 1989, 2003).
How can we assess further whether this series reflects real height development? One counter-checking strategy would be to look at disaggregated data by region and gender. If the disaggregated series moved in a similar direction when expected to do so, while deviating only where this appeared justifiable from a theoretical perspective, this would consequently lend support to the validity of our overall height trend. When thus tested, the development of heights in the Mediterranean, Central/Western European and Northern/Eastern European regions turns out to be quite similar (Figure 3), as is the case with the decline in the fourth century in Central Western and North-Eastern Europe, the astonishing increase in the fifth and sixth centuries and the low marks in the 13th and 17th centuries. The reported increase in the 14th century and the high value during the 15th century can also be observed in more than one region. Deviations occur for the seventh and eighth centuries. While the Northern/Eastern height series remains at first at a constant level above all others, it loses its leading role from the 13th century onwards. In the 13th, 14th and 17th centuries, Scandinavians and Eastern Europeans became shorter than the British, Dutch and other Central/Western Europeans. It might have been the case that Northern and Eastern populations suffered exceptionally during the Little Ice Age (14th – 18th c.), whereas they benefited from the climatic maximum of the 11th and 12th centuries. In contrast, the maritime climate of the Netherlands, the British Isles and Western Germany allowed for a more favourable nutrition during the Ice Age.

- Figure 3 -

In a similar vein, male and female heights moved in an overall comparable direction (Figure 4). Their common maximum during the sixth century may have been caused predominantly by the female height series, but irrespective of that, both series reached a maximum at this point in time and another high value during the 11th and 12th centuries. During the high Middle Ages, females lost some ground in comparison to males. While the increase of female heights during the 15th century is supported by only 18 observations, the positive situation of the 16th century relies on 118 cases. Women might have benefitted from a
change in social roles during the Renaissance period, whereas gender discrimination was particularly severe during the “Dark Ages” when women's social position deteriorated, as archaeologists have demonstrated on the grounds of other sources (see Ulrich-Bochsler, 1996).

Are female heights in our sample characterized by a higher variance between birth centuries than male heights? In the following, we consider the average height's coefficient of variation by birth century (please note: this does not correspond to the CV of individual height distribution). The overall coefficient of variation is 0.68 for females and 0.49 for males. Biologists would expect female bodies to be more robust under adverse conditions (see e.g. Ortner, 1998), hence variability should be lower. As our results indicate, however, gender discrimination might have had a stronger impact on height than biological factors (compare also Moradi and Guntupalli, forthcoming). This finding correlates with the argument that the position of women deteriorates in relative terms when times are getting worse (Klasen, 1996). We have to admit, however, that the higher variability established here might at least in part be influenced by the lower number of female observations.

- Figure 4 -

POTENTIAL ECONOMIC DETERMINANTS OF HEIGHTS IN THE LONG RUN

In the following, we perform an exploratory analysis of the potential influence of a number of variables on stature by century and region.

(1) Land per capita and urbanization

Was Malthus right with his argument that land is the limiting factor of human development? Did population growth tend to outpace food production as a result of decreasing marginal product until culminating in a major demographic catastrophe? On the eve of a mortality crisis, Malthus would expect a decline in the nutritional status of the population under risk, a relationship which will be tested below. In the aftermath of mortality crises, in contrast,
nutritional quality might ameliorate as a result of the enhanced availability of agricultural land. Thus, to give just one example, cows and other forms of farm capital were not reduced to the same degree as the population during the major plague epidemics of the 6th, 14th and 17th centuries.

Apart from the effect of land availability, increasing urbanization might have separated urban dwellers from *de facto* untradeable goods such as milk. In addition, infectious diseases could spread more easily in urban centres. In this study, we measure population density as a weighted average of the country estimates by McEvedy and Jones (19802), while amending and improving them wherever possible with later and more detailed country studies (such as Wrigley and Schofield (1981); Dupâquier et al.1988; for further details, see appendix 3). We estimate urbanization levels as described in the appendix (based on Bairoch, 1976; Federico and Malanima, 2002; Allen, 2003).25

(2) Climate

One of the most fascinating topics in long-run economic history is the relationship between climate and human living standards. For example, did climatic change cause the demographic catastrophes of the 14th and 17th centuries, as Galloway (1986) has argued (see also Kelly, n.y., among others)? Colder winters tended to make food production (especially protein production) more difficult in Central Europe (on the 18th century climate-height effect, see Baten, 2002), with a consequential immense impact on human history. Grove (2002) demonstrated how the switch from the medieval warm period (lasting from 900 A.D. up until the early 13th century) to the Little Ice Age, starting in the late 13th century, decreased harvests and protein-production from cattle and sheep.26 Not only did temperatures decline, however, but since colder winters tended to be correlated with a higher frequency of climatic extremes, consecutive climatic problems created a deadly synergy effect. For instance, cattle epidemics spread rapidly in Northern and Western Europe as early as the 13th and early 14th centuries, killing a large share of the cattle stock. As Grove has argued, the resulting
agricultural production decline took place before and simultaneously with the Black Death of the mid-14th century.

Although plague is a highly infectious disease which is only mildly influenced by malnutrition, lower nutritional status might have weakened the immune systems of the European population, thus contributing strongly to the massive population loss of the 14th century. In addition, people often leave their households during famines and start moving around in search of other possibilities of subsistence (Mokyr and O’Grada, 2002). The cattle- and fishery-based economies of the northernmost region of Europe suffered most drastically. Thus, Iceland lost most of its population and the European population of Greenland disappeared completely.

The 15th and the first two thirds of the 16th century were warmer again, but the next climatic catastrophe came about in the 17th century. Pfister (1988) described how climatic changes reduced Swiss nutritional status in the last decades of the 16th and during most of the 17th century. While the major share of the population-decline in the 17th century is traditionally ascribed to the Thirty Years’ War and the hunger and infectious diseases following in its wake, rapid climatic deterioration could have contributed further to the large number of (at least partially) nutrition-related deaths during this devastating war. The interplay between protein malnutrition and disease-related deaths can also help to explain why population figures stagnated or declined even in countries which were not actively involved in the Thirty Years’ War. Lastly, milder episodes of climatic deterioration during the late 18th and mid-19th century coincided with milder average demographic effects (Grove 2002).

Recent research has brought forth new estimates on climatic change over the centuries by quantifying Alpine and Scandinavian glacier movements, Greenland ice kernels, oak tree rings and lake sediments. All of those series appear to be correlated in general. We mainly used glacier movements as explanatory variables, since the available data go back to the ancient period on the one hand, and might serve as more unequivocal evidence than, for
instance, oxygen isotope ratios from Greenland on the other hand (see Heide, 1997; Grove, 2002, p. 316). As is emphasized in the literature, however, glacier movements reflect temperature changes with a certain time lag. We therefore calculated the average of glacier movements in the previous and current century. In addition, we corroborated our glacier series with a tree-ring series from Northern Sweden which stretches back to the ancient period as well, and compared both with a shorter tree-ring series from the Alpine area - they moved in accordance (see Huntley et al., 2002, p. 278).  

Our comparison between the height and temperature series yielded some similarities and many differences (Figure 5). The well-documented climatic optimum of the 11th/12th centuries as well as the lower values before and afterwards are indeed discernible in the height series. Moreover, the low values of the seventh and eighth centuries and the crisis of the 17th century could have been caused by adverse climatic conditions. Important deviations, however, occur for the first to sixth and for the 13th centuries. Either only a weak relation exists, or a measurement error occurred, especially for the early period whose temperature estimates are known to be particularly imprecise. In our opinion, the most likely interpretation is that after the breakdown of the Roman Empire, several phenomena increased average height and nutritional status: (1) as a result of invasions and plague epidemics, population density and urbanization decreased and consumers moved back to the vicinity of nutrient production. Besides, infectious diseases might have occurred less frequently – although this factor is likely to be of secondary importance, as it would have been counter-acted by the disappearance of the famous Roman Public Health institutions, and contradicted by the second occurrence of the plague in the sixth century.  

(2) Germanic invaders introduced agricultural methods which emphasized protein production. Even if those methods were not efficient in a Mediterranean context, inhabitants might have adhered to them for a transitory period. In Central and Western Europe, they proved to be efficient as long as population density was low.
Both of these developments might explain why no climate-height relationship can be detected for the first six centuries. The low height-value of the 13th century is particularly interesting and deserves further research. Was it caused by the rapid urbanization of this period (along with more infectious diseases and less milk for rural-to-urban migrants)? Or was it rather a result of higher social or gender inequality? Lastly, is the height variable biased by a measurement error?

- Figure 5 -

One could argue that cold climate was not harmful, but rather beneficial to agriculture in the Mediterranean region, because colder climate might have brought about more frequent rainfall. We accounted for this hypothesis by testing whether our results would change if our small number of observations on the Mediterranean were excluded, yet found the climate coefficient unchanged (the coefficient was 2.74, compared with 2.97 when the Mediterranean was included).

(3) Income

Agricultural income was clearly the dominant source of subsistence over the major part of the last two millennia. To a large extent, it was a function of land per capita, land/soil quality and climate. Moreover, the level of income might have been augmented by industrial and service-sector production particularly during Roman times, the High Middle Ages, the 16th century and after the 18th century. We are unable to test this effect, however, because income estimates are completely unreliable for the first millennium, or based solely on urbanization rates and population growth (such as Maddison's estimates, 2001).

(4) Social Inequality

Inequality was identified by previous research as an important determinant of average height (Steckel, 1995). If growing income inequality and purchasing power disparity are not accompanied by changes in aggregate real GDP per capita, the rich might get richer and the poor get poorer to the same extent. Yet as the rich will spend a smaller proportion of their
extra income on additional food, while poor people’s already low nutrient level will decrease even further, average height will decline even if average purchasing power does not.

- Figure 6 -

Height data sets allow rough estimates of health inequality (see Baten, 2000). Wurm (1985) argued that inequality was particularly low in the early Middle Ages. In fact, social inequality increased dramatically between the early and the high Middle Ages (Figure 6), and again in the 15th – 18th century. This overall trend towards inequality corresponds well with other studies on income inequality. O’Rourke and Williamson (2002) have confirmed this for longer periods by using rent-wage-ratios, assuming quite reasonably that land-owners were relatively rich while wage earners were relatively poor (van Zanden, 1995). We controlled for social inequality by calculating the difference between individuals of middle/higher status and the lower/unknown category, by major periods.

(5) Public Health

How did Public Health develop over the last two millennia? Is our conception of the Romans’ impressive water-supply technology and especially of the Roman institution of public bathing facilities correct, and did a large share of the population benefit from these?29 To what extent did hygienic conditions deteriorate after the breakdown of the Roman bath-system? Were Public Health investments perhaps endogenous, being made primarily in times when urbanized and poorly nourished populations suffered more than usual from infectious diseases? Or did the Romans' contacts with distant populations (such as Chinese, Indians, Parthians) lead to the spread of new infectious diseases? In order to answer all these questions and to capture the potentially beneficial effects of the Roman Public Health system, we coded a “Roman Bath” dummy variable as 1 for the Mediterranean region over the centuries 1 to 4, and for Central/Western Europe for the centuries 2 to 4.30 This specification implies that apart from the Roman hygienic system (which may well have been of great importance for height and health levels), we might also use it to capture other aspects of Roman technology and the
imperial economic system. Therefore, we will name this variable “Roman bath or other technology”.

(6) Gender Inequality

In our estimation of height by (birth) century, we assumed that gender differentials were constant over time. In the following, we will relax this assumption and control explicitly for higher or lower gender inequality. Our expectation is that higher gender inequality *ceteris paribus* reduces average height, because Osmani and Sen (2003) have convincingly argued that female discrimination affects both girls' and boys' height via their mothers' nutritional status (see also Klasen, 2002). We measure the development of gender differentials over time using the dimorphism estimates graphed in Figure 4. However, here we calculate the percentage of height difference relative to the average male height, in order to adjust for possible level effects (Koepke 2002).

RESULTS

A large number of regression models were used in this study: fixed effects and random effects, weighted and unweighted least square, as well as generalized least square with adjustment for autocorrelation and heteroscedasticity (not shown in tables). The following results are robust across these various specifications. Unit root tests or cointegration might not have enough power based on 18 observations over time, but when nonetheless performing the ADF test, the height series for Northern/Eastern and Central/Western Europe turn out to be stationary. ADF North/East: even nesting the trend (which is insignificant) -5.12 (critical value 10%: -3.24); ADF Central-West: -3.43 (critical value 10%: -3.24) without nested trend, a unit root is rejected at the 5% level; Mediterranean: insufficient number of cases. Panel estimation techniques exploit the variation both over time and between cross-sectional units. We present two WLS estimates with regional dummies (equivalent to fixed effects), one of them with period dummies (for antiquity etc.) and the other without controls for unobserved
inter-temporal heterogeneity (Table 4a and b). The results are also robust when time trends are included (insignificant).

Given that we have considerable measurement error in our rough proxies for social and gender inequality, population density and urbanization, it is not astonishing that many coefficients are statistically insignificant. In addition, technological development cannot be appropriately captured over these two millennia, which means that a certain population density in 1800 might have resulted in a different average height than the same population density in 800 A.D. (simple time trends were insignificant, however). Following McCloskey and Ziliak (1996), we will not only focus on statistical, but also on economic significance, comparing the high and low height values predicted by our estimates with results from the 18th and 19th centuries, for which height differences between 1 and 3 cm are often interpreted as economically significant phenomena (Komlos and Baten, 1998, estimated that 1 cm in additional height corresponds to about 1.2 years of life expectancy).

Both the regional dummy variables and the period dummy representing antiquity are significant. In the regression without time dummies, the “Roman bath”-dummy becomes significant as well. However, Roman bath technology and especially other technology (e.g. in agricultural terms) were not able to improve height and health quality sufficiently enough to outweigh the negative effects of the Roman economic system, which explains why the coefficient of this variable is negative. We coded it as zero for North-Eastern Europe and Central-Western Europe in the first century, and this variable remains significant if we control the antiquity effect of low heights with another dummy variable (not shown). What was so different during Roman times, causing people under Roman reign to be shorter than others? One factor in this regard could have been income inequality. Jongman (2000) has argued that Roman wealth and income were distributed extremely unequally. A very small upper-class had an enormously large share of total income at their disposal. This upper class was too small to be captured by our estimate of height inequality, since this measure is only sensitive
when the high-welfare group represents a substantial proportion of the population. Inequality of heights is typically lower if rural areas are dominated by subsistence farming, including milk production. If subsistence farming is typical, even the poor can consume some perishable products such as milk and giblets, whereas the higher market integration during the Roman period might have reduced non-market entitlements to such products. In extreme cases, the poor became vegetarians. In addition, being a provincial subject in a large empire governed by a foreign elite speaking a foreign language typically leads to perceived or actual income losses, and a smaller supply of public goods such as Public Health (like quarantine measures against infected subjects etc.). Finally, due to contact with Persia, Asia etc., new infectious diseases could have been brought to Europe which were priorly unknown (although this should apply even more to the period of the crusades and the contact-period of the 16th century).

All other variables are statistically insignificant, but there is some indication of the expected outcome that warmer weather is favourable for harvests and protein production in the relevant range, which in turn increase height. The difference between the two standard deviations of our climatic series is 0.12, with the difference between minimum and maximum being 0.20 - values which can be interpreted as “good” and “bad” climate. The coefficient of the more appropriately specified model in Table 4b, column 2 is 2.97. The difference between “good” and “bad” climate amounted therefore to about 0.4 cm in height, with the difference between the extremes amounting to approximately 0.6 cm. Both values are at the margin of being economically significant (if period dummies are included). The tall stature of North-Eastern Europeans in the warm 11th/12th century and the subsequent dramatic decline in height lends further support to the importance of this variable.

- Table 4a and b -

Population density comes closest to statistical significance; in unweighted regressions, the p-value is even as low as 0.15 (not shown in table). This suggests that a lower population
density is advantageous for exactly that biological component of the standard of living which was reflected in stature during pre-industrial times, after controlling for large-regional effects and inequality. The analysis of the population density’s economic significance yields a height-effect of about 1.0 cm for the difference between typical “high” and “low” population densities of the time, and 2.2 cm between the most extreme observed values. In the other specification without time dummies, the economic significance of population density would even be enhanced by one third. Interestingly, the sign of the urbanization coefficient is positive once population density is controlled for. Without time dummies, it is even almost significant. The potentially large measurement errors prevent us from hypothesizing too much at this stage, but one could speculate that once the detrimental influence of high population density (which, because of the decreasing marginal product, implies: less protein per capita) is removed, the human capital-deepening effects of urban agglomerations on the entire country outweigh other negative effects (such as crowded cities and hygienic problems).

Both gender inequality and social inequality had negative signs. Given that these results are similar to those of many other studies on the 18th to 20th centuries, we tend to attribute fairly large credibility to them. In terms of economic significance, social inequality corresponded to a height-difference of 0.63 cm between high and low, and 0.74 cm between extreme values, whereas the effect of gender inequality was about half as strong. In sum, population density is definitely economically significant, albeit not of statistical significance. Climate, social inequality and perhaps gender inequality are thus all at the margin of being economically significant.

CONCLUSION: THE LARGER PICTURE, SOME SPECULATIONS AND PLANS
In sum, this study is the first to offer a European time-series of anthropometric estimates over the last two millennia (excluding the last two centuries, on which much research has been conducted already), although with some limitations. Height series are often correlated to
various biological aspects of the standard of living (such as longevity), but they do not necessarily capture other important aspects related to purchasing power. To illustrate this, a Northern Barbarian living in the sixth century was tall and certainly lived relatively long (as Herrmann, 1987, demonstrated), but in case he was amusement-loving and fond of consumer goods, he would most probably have preferred to live in second century Rome. Yet although we acknowledge this, we are unfortunately not able to measure such aspects of welfare. What we can and do capture, however, are other important aspects relating to height and development which have so far often been underrated.

The overall picture emerging from our study is one of stagnant heights over the past two millennia. We did not find much progress in European nutritional status, not even between 1000 and 1800, a period for which Maddison (2001) and others arrive at growing GDP per capita figures (Figure 7, but see Federico, 2002, for a critical view.).

Likely reasons for this divergence are (a) relatively favourable nutritional conditions during the Middle Ages, and especially the climatic optimum of the 11th/12th century; (b) the bias in pre-industrial and early GDP per capita estimates in favour of industrial goods consumed by middle- and upper class consumers, a possibility which is also supported by van Zanden's (1999) finding of a “negative link between economic development and the level of real wages” from the beginning of the 16th to the end of the 18th century. Van Zanden also described a decline of per capita meat and dairy product consumption. In a similar vein, Federico and Malanima (2002) estimated a downward trend of food consumption in Italy between 1300 and 1860.

- Figure 7 -

Our analysis stretches back to ancient times, measuring living standards during the Roman Empire and the following “Dark Ages”. We find that heights stagnated during the Roman imperial period in Central, Western and Southern Europe. In Northern and Eastern Europe, heights might have increased between the first and third century, but fell dramatically
in the fourth century. Whether this downward trend contributed to the onset of the migration of entire peoples in the fourth century awaits further exploration. One astonishing result is the height increase in the fifth and sixth centuries, the largest residual in our model. Declining population density in the former provinces after the breakdown of the *imperium Romanum*, and the plague of the sixth century might have played their role in this as well. Noteworthy is also the relative synchrony of the height development in the three regions.

Does the development of the cattle-bone rate as estimated by archaeozoologists (see data section above) support such a nutrition-based explanation for the fifth and sixth centuries? In their studies, we find our expectations to be fulfilled that during the imperial period of most extreme population density (but after the climax of power during the first century), large cities like Rome or Pompeii had a very small share of beef and milk consumption (grazing being too costly), and instead substituted meat and dairy products with grain and vegetables – and pork for the richer strata of society. In fact, the impressive cattle share of 28 % in Rome (Aqua Marcia excavation) between the first century B.C. and the first century A.D. fell to 7.9 % between the first and second century, reaching 0 % in the second and third century A.D. During the fourth century, the share was still negligible (0.6 % on the Palatine). Only in an excavation on the fifth century (Schola Praeconum), a substantial cattle share was found again, after the population density had decreased significantly and Germanic invaders had introduced a new agricultural system (and perhaps taste). Similarly in Naples, the cattle share was low throughout the first up until the third century A.D. (2-6 %), and only somewhat higher during the fifth to seventh centuries (6 – 9 %). Ostia and other excavation sites display a similar, but more mixed pattern. In general, the second to fourth centuries display low urban cattle rates in Italy.

We constructed potential explanatory variables of height development on the narrow basis of what we know about this early period. We have to emphasize the limitations of those estimates, but a first exploratory look might still generate informative insights. Population
density was clearly economically (but not statistically) significant. Thus, decreasing marginal product theories and Malthusian thought cannot be denied for the pre-1800 period. Of marginal significance were climate (warmer temperature being advantageous for nutritional status), social inequality and gender inequality (both reducing average height). When controlling for population density, urbanization was positive.

Questions about the social composition of the height samples over time as well as other potential biases can only be answered within certain confidence intervals (given the current state of research). However, we would argue that the error probability is smaller than for most other methods which can be applied to the first millennium (such as the urbanization-based GDP estimates by Maddison, 2001). If our intention is to study the economic history of Europe in the very long run, anthropometric techniques provide important insights into some (although naturally not all) central aspects of human life.
FOOTNOTES:

1 Komlos (1985) was the first to use this term.

2 Growth which is mostly stimulated by trade and comparative advantage, as opposed to Schumpeterian growth which is driven by technological change, see Mokyr (1992).

3 Richard Steckel and his co-operators have started a research project (Steckel, 2003) in which they will investigate height and a number of diseases (as far as they can be traced with bone material).

4 Crotty (2001) emphasized the importance of lactose intolerance in his bold attempt to explain the evolution of capitalism based on cattle-farming patterns. Crotty argued that lactose-intolerant people could not make sufficient use of cattle. Lactose intolerance means that people have digestive problems when drinking large quantities of milk after the age of 5–7, because at that age, genetically lactose-intolerant people lose their ability to digest fresh milk without encountering diarrhoea and similar problems. Especially East Asians (east of Tibet and Rajasthan), Native Americans and some African people suffer from lactose intolerance. For Southern Europe, the results are mixed – one study included Spain into the category with the highest degree of lactose tolerance (70 % and more) while the study put Greece in a middle position (30 – 70 % lactose tolerance). In Italy and Turkey, however, less than 30 % were classified as lactose tolerant (see Mace et al., 2003). Yet even lactose intolerant people can digest modified milk such as Kefir, Lassi and similar products. Moreover, all people can drink about one cup of milk per day if they train their intestinal bacteria to live in a milk environment. Even many South Koreans today consume some milk, using this method of permanent training. We thank Barry Bogin, Anthropology Department of the University of Michigan/ Dearborn, and S. Pak, Seoul National Univ., for sharing information on this.


6 The so-called primary deficit of females (smaller number of females in the case of patriarchally-structured societies) is typical for prehistoric and ancient populations: see e.g. Mays (1995).

7 The same applies to age estimates.

8 We also discarded old individuals (> 59 years) and included an “age 51-59 ”-variable in the beginning, because one earlier commentator pointed towards a possible selection mechanism. We
therefore used a dummy variable to control for any potential bias, but it turned out to be insignificant and is therefore not reported in the table. In contrast to data based on living heights, however, by using long bones we do not have to take into account the biologically determined shrinking process experienced by older people, since bodies shrink due to the compression of the disks between the vertebrae (as well as poor posture), while the femur does not change significantly with age. Some compression of joints occurs in the lower half of the body as well, but the length of the femur does not change enough to make a difference: Almost all of the age-related loss in height is derived from the collapse of the intervertebral disks, and the collapse of the vertebral bodies in the case of some individuals. Changes in femoral length terminate with the fusion of the epiphyses. The only way to achieve changes in length thereafter would be through a remodelling of the articular surfaces, or by bending the bone itself. Both of these changes would only be discernible under rare pathological conditions, such as severe osteoarthritis, femoral fractures, and perhaps osteomalacia.

We would like to thank Barry Bogin, Rick Steckel and Phil Walker for their friendly communication in this regard.

9 Note that most knight’s armours originated in fact from a time period when military technology had moved away from the horse-based knight armies which had proven so unsuccessful in the Hundred-Years’ War. Our armours probably stem from males from all social strata, who were hired and received salaries as soldiers.

10 In contrast to the older hypotheses that, apart from the femur, the humerus leads to a smaller measurement error than the radius and tibia, current researchers have come to the conclusion that normally, the long bones of the lower extremities (i.e. apart from the femur: the tibia) serve as the best approximation (see Herrmann et al. (1990) 93 f.).

11 We thank F. Rösing, Univ. of Ulm, for kind oral information.

12 Naturally, different regression formulae have to be used for females and males. Furthermore, the high number of models is the result of different reconstruction methods for inhumed and cremated bones, as the latter show signs of diminution due to exposure to heat; for example, for the reconstruction of inhumations alone, 43 models exist. On cremation, at least 22 researchers have worked using different models.
Stature reconstructions based on the also quite commonly used model by Olivier were sorted out, as they are not based on the same bone-measurements following Martin (1928) as the ones mentioned above (thus making it impossible to recalculate them).

If we know that the excavators and principal investigators used femurs, we can recalculate each stature estimate into the corresponding estimate of another reconstruction method. However, sometimes the authors do not mention whether their estimates are based on femurs only or, for example, on a combination of bones (which would render the estimates more exact), like femur and tibia, or on any other bone. In order to minimize recalculation error, we therefore prefer to use the calculation method employed by most principal investigators.

However, note that Pearson might be more efficient for the reconstruction of the height of particularly small individuals (males below 160 cm).

He did not include Manouvrier’s method.

This is the distance from the highest point of the caput to the lowest point of the condylus tibialis. According to J. Wahl, the stature estimates based on the Breitinger/Bach method come closest to the Rösing estimates. Since we want to use Rösing’s values for comparison, we are best-advised to use Breitinger and Bach measures. – We thank J. Wahl, Univ. of Tuebingen, for kind oral information.

In order to test the robustness of our first-step results, we ran the regressions with aggregate and disaggregated data (disaggregated by region and gender; gender-disaggregated regression not reported here).

The latter two are usually taken together, because the share of the upper class is very low, and even the share of the middle class is not very high in most historical populations.

In some cases, burial objects that would have indicated higher social status were certainly lost or robbed. This means that middle and upper class status is probably somewhat underreported.

To be sure, this refers to aggregated height. Individual height is clearly influenced by genetic factors. It is not fully clear whether some very isolated populations such as the Pygmies have a different height potential. In addition, Japanese people are sometimes outliers in regressions (although their strongest height increase after World War II correlates well with the introduction of dairy
products). Maya children who were brought to the U.S. and enjoyed good nutrition converged rapidly to North American growth paths – but never fully (Bogin, 1991).

Earlier views that North-Eastern French are genetically taller than other Europeans were recently rejected: once milk production and income are controlled for, the height difference disappears. The finding that Dutch people were particularly short during the nutritional crisis of the mid-19th century and that Indians (of Asian origin) and Central Asian nomads were particularly tall, also contradicts the explanatory power of genetic factors for mean height (Steckel and Prince, 2001).

22 The relatively high R-squares should be regarded with caution, as they stem mostly from the inclusion of gender dummies.

23 The congruence of bone evidence (168-170) and military evidence (164-172) is either good or poor, depending on how the military evidence is interpreted.

24 Allen (2000) finds a positive real wage development in the early modern period for the British urban case. The number of observations is (quite) comparable when combined into large regions (see Table 2: 13th (c.: 82 to 195 cases), 14th (c.: 553 to 120 cases), 17th (c.: 80 to 58 cases)).

25 We did not use the superior estimates of de Vries (1984), because they start after 1500 only.

26 Grain yields dropped between 1220 and 1320, see Grove (2002), figure 2.

27 We experimented with local temperature series for the three regions of Northern/Eastern, Central/Western and Southern Europe, but the differences between the series were extremely small, so that we abandoned this avenue of temperature measurement.

28 The so-called Antonine plague is regarded as the first one ever to occur.

29 In contrast to the generally positive view of Roman baths, Scobie (1986) argued that the bathes were quite unhygienic (e.g. water was rarely changed; it was a meeting point of ill and healthy people).

30 McKeown (1955) has argued that medical technology did not play an important role in societies prior to the 20th century.

31 For the case of the Roman provinces, some authors hypothesized that the urban population was better supplied with food than their rural counterparts.
The statistical insignificance is probably not caused by the fact that we controlled for social status in the regressions of table 4. We performed the same estimations without social status, and all results were almost identical.

In our data, Scandinavian and Eastern European heights even declined significantly. This supports Steckel's (2004) finding of a long-term height decline in Northern Europe. We took care not to use the data analyzed by him again, except for very few heights which we had recorded at an early stage of our project (6%).

Note also the potential urban bias of the real wage estimation (rural nutrition and living standard, in contrast, were influenced crucially by non-traded, high-quality proteins (such as milk), see Baten, 1999; van Zanden, 1995).

We also used plague dummies for the second, sixth, 14th, and 17th centuries and found them to be insignificant. In addition, we tested whether stature was higher in the second halves of those centuries, since the most violent plague waves occurred around the mid-centuries, just as the tall stature of the sixth and 14th centuries might have been caused by the lower population density afterwards. In fact, heights were half a centimeter taller in the second halves of those centuries, which is economically, but no statistically significant (not shown here).

ACKNOWLEDGEMENTS

We thank Stephen Broadberry, Hans de Beer, J.W. Drukker, James Foreman-Peck, Timothy Guinnane, Michael Haines, Willem Jongman, Alexander Moradi, Esteban Nicolini, Alan Olmstead, Sevket Pamuk, Richard Steckel, Joachim Wahl, and participants of the Fifth EHES Conference 2003, the Economic and Human Biology Conference 2004, and the Tuebingen Economic History Research Group, for important comments on earlier versions. Deborah Rice improved the language style.
REFERENCES


http://www.britarch.ac.uk/ba/ba2/ba2feat.html#mays


### Table 1

**STUDIES FOR COMPARISON: SKELETAL DATA**

<table>
<thead>
<tr>
<th>Author</th>
<th>Year of Publication</th>
<th>Dating (approx. cent.)</th>
<th>Region</th>
<th>N cases</th>
<th>(N) Underlying Heights</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Steckel/Rose</td>
<td>2002</td>
<td>-5500 – 0</td>
<td>America</td>
<td>33</td>
<td>1106</td>
<td></td>
</tr>
<tr>
<td>Steckel/Rose</td>
<td>2002</td>
<td>200 – 1800</td>
<td>America</td>
<td>14</td>
<td>5188</td>
<td></td>
</tr>
<tr>
<td>Steckel</td>
<td>2004</td>
<td>-900 – 1900</td>
<td>Scandinavia &amp; GB</td>
<td>23</td>
<td>6531</td>
<td>incl 2 sites also included by Maat</td>
</tr>
<tr>
<td>Maat and DeBeer</td>
<td>2003 and 2004</td>
<td>0 – 1800</td>
<td>Netherlands + Island</td>
<td>391</td>
<td>391</td>
<td>different height reconstruction models mixed (Maat)</td>
</tr>
<tr>
<td>Angel</td>
<td>1984</td>
<td>-3000 – 100</td>
<td>E- Mediterranean</td>
<td>974 ?</td>
<td>974</td>
<td></td>
</tr>
<tr>
<td>Angel</td>
<td>1984</td>
<td>100 – 1800</td>
<td>E- Mediterranean</td>
<td>254</td>
<td>254</td>
<td></td>
</tr>
<tr>
<td>Jäger/Zellner</td>
<td>1998</td>
<td>-5000 – 2000</td>
<td>NE-Germany</td>
<td>45</td>
<td>2130</td>
<td></td>
</tr>
<tr>
<td>Haidle</td>
<td>1997</td>
<td>-5000 – 1900</td>
<td>S-Germany &amp; N-Switzerland</td>
<td>488</td>
<td>488</td>
<td></td>
</tr>
<tr>
<td>Koepke/Baten</td>
<td>2000</td>
<td>0 – 1800</td>
<td>Europe</td>
<td>2974</td>
<td>9226</td>
<td>without harnesses (251 cases)</td>
</tr>
</tbody>
</table>

**Notes:** grey-shaded are studies which focus exclusively on the same period as our article (0 A.D. to 1800 A.D.), while other studies include earlier millennia. DeBeer uses 41 cases in addition to Maat. We distinguish between number of cases including averages and individual data, and the underlying number of height measurements. For example, Steckel (2004) considers 23 averages from the excavation literature, which averaged 6531 individual heights.
### Table 2

**AREAS COVERED BY THE DATA SET (NUMBER OF INDIVIDUALS)**

<table>
<thead>
<tr>
<th>Century</th>
<th>Central/Western Europe</th>
<th>Eastern/Northern Europe</th>
<th>Mediterranean region</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bavarian/Austrian region</td>
<td>Northern Rhine region</td>
<td>Southern Rhine region</td>
<td>Uk</td>
</tr>
<tr>
<td>1</td>
<td>21</td>
<td>21</td>
<td>14</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>56</td>
<td>6</td>
<td>210</td>
<td>711</td>
</tr>
<tr>
<td>3</td>
<td>16</td>
<td>95</td>
<td>24</td>
<td>184</td>
</tr>
<tr>
<td>4</td>
<td>361</td>
<td>1</td>
<td>26</td>
<td>227</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>113</td>
<td>9</td>
<td>68</td>
</tr>
<tr>
<td>6</td>
<td>338</td>
<td>208</td>
<td>380</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>146</td>
<td>35</td>
<td>456</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>225</td>
<td>20</td>
<td></td>
<td>179</td>
</tr>
<tr>
<td>9</td>
<td>78</td>
<td>5</td>
<td>36</td>
<td>164</td>
</tr>
<tr>
<td>10</td>
<td>49</td>
<td>24</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>14</td>
<td>4</td>
<td>229</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>18</td>
<td></td>
<td>20</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
<td>12</td>
<td>9</td>
<td>174</td>
</tr>
<tr>
<td>14</td>
<td>113</td>
<td>3</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>15</td>
<td>55</td>
<td></td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>16</td>
<td>73</td>
<td>55</td>
<td>314</td>
<td>17</td>
</tr>
<tr>
<td>17</td>
<td>21</td>
<td>19</td>
<td>39</td>
<td>39</td>
</tr>
<tr>
<td>18</td>
<td>103</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1343</td>
<td>777</td>
<td>1700</td>
<td>1483</td>
<td>1184</td>
</tr>
</tbody>
</table>

Sources: see [www.uni-tuebingen.de/uni/www/twomillennia.html](http://www.uni-tuebingen.de/uni/www/twomillennia.html)
Table 3
FOUR REGRESSIONS: DETERMINANTS OF MALE AND FEMALE HEIGHTS IN THE
THREE PARTS OF EUROPE

<table>
<thead>
<tr>
<th>Region</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>All</td>
<td></td>
<td>Central/Western</td>
<td></td>
<td>Mediterranean</td>
<td></td>
<td>Northern/Eastern</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>159.51</td>
<td>0.00</td>
<td>159.78</td>
<td>0.00</td>
<td>161.13</td>
<td>0.00</td>
<td>160.13</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Status mid/high</td>
<td>0.58</td>
<td>0.11</td>
<td>0.45</td>
<td>0.35</td>
<td>0.32</td>
<td>0.58</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Migr. Mediterr.</td>
<td>-3.87</td>
<td>0.01</td>
<td>-4.00</td>
<td>0.00</td>
<td>-0.20</td>
<td>0.89</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Migr. Gemanic</td>
<td>0.75</td>
<td>0.35</td>
<td>1.63</td>
<td>0.24</td>
<td>8.62</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>8.20</td>
<td>0.00</td>
<td>7.97</td>
<td>0.00</td>
<td>7.72</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Centuries:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 1.00</td>
<td>0.02</td>
<td>1.42</td>
<td>0.03</td>
<td>-0.55</td>
<td>0.74</td>
<td>0.59</td>
<td>0.60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 1.04</td>
<td>0.00</td>
<td>1.12</td>
<td>0.01</td>
<td>-1.20</td>
<td>0.39</td>
<td>1.51</td>
<td>0.19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 1.07</td>
<td>0.63</td>
<td>0.20</td>
<td>0.60</td>
<td>0.46</td>
<td>0.85</td>
<td>-0.95</td>
<td>0.44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 1.44</td>
<td>0.00</td>
<td>1.47</td>
<td>0.01</td>
<td>-0.38</td>
<td>0.83</td>
<td>1.76</td>
<td>0.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 2.80</td>
<td>0.00</td>
<td>2.92</td>
<td>0.00</td>
<td>1.30</td>
<td>0.47</td>
<td>2.88</td>
<td>0.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 1.99</td>
<td>0.00</td>
<td>2.24</td>
<td>0.00</td>
<td>0.31</td>
<td>0.88</td>
<td>-0.43</td>
<td>0.75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 1.05</td>
<td>0.01</td>
<td>0.30</td>
<td>0.51</td>
<td>0.78</td>
<td>0.72</td>
<td>1.06</td>
<td>0.35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 1.24</td>
<td>0.00</td>
<td>1.32</td>
<td>0.02</td>
<td>0.78</td>
<td>0.72</td>
<td>1.06</td>
<td>0.35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 0.86</td>
<td>0.17</td>
<td>1.21</td>
<td>0.15</td>
<td>0.37</td>
<td>0.78</td>
<td>1.44</td>
<td>0.19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 1.61</td>
<td>0.00</td>
<td>1.57</td>
<td>0.25</td>
<td>-0.49</td>
<td>0.75</td>
<td>1.86</td>
<td>0.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11 1.90</td>
<td>0.00</td>
<td>-0.94</td>
<td>0.59</td>
<td>-0.49</td>
<td>0.75</td>
<td>1.86</td>
<td>0.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12 -0.14</td>
<td>0.00</td>
<td>0.84</td>
<td>0.36</td>
<td>0.84</td>
<td>0.36</td>
<td>-0.59</td>
<td>0.61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13 1.22</td>
<td>0.02</td>
<td>2.68</td>
<td>0.01</td>
<td>4.54</td>
<td>0.06</td>
<td>0.34</td>
<td>0.77</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 0.90</td>
<td>0.32</td>
<td>3.61</td>
<td>0.02</td>
<td>0.61</td>
<td>0.75</td>
<td>-2.47</td>
<td>0.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 1.64</td>
<td>0.00</td>
<td>1.57</td>
<td>0.01</td>
<td>6.25</td>
<td>0.06</td>
<td>2.61</td>
<td>0.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16 -0.51</td>
<td>0.39</td>
<td>0.99</td>
<td>0.23</td>
<td>1.97</td>
<td>0.12</td>
<td>1.65</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17 1.62</td>
<td>0.19</td>
<td>2.09</td>
<td>0.10</td>
<td>1.62</td>
<td>0.19</td>
<td>2.09</td>
<td>0.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rhine, South</td>
<td>0.71</td>
<td>0.01</td>
<td>0.48</td>
<td>0.10</td>
<td>1.65</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rhine, North</td>
<td>1.14</td>
<td>0.01</td>
<td>0.62</td>
<td>0.16</td>
<td>1.65</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UK</td>
<td>-0.90</td>
<td>0.08</td>
<td>1.09</td>
<td>0.00</td>
<td>1.65</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Northern Eur.</td>
<td>2.21</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td>1.65</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eastern Eur.</td>
<td>0.78</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td>1.65</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mediterranean</td>
<td>-0.01</td>
<td>0.99</td>
<td></td>
<td></td>
<td></td>
<td>1.65</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adj. Rsq</td>
<td>0.54</td>
<td>0.51</td>
<td>0.63</td>
<td>0.61</td>
<td>0.61</td>
<td>0.61</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N (original)</td>
<td>9477</td>
<td>5303</td>
<td>542</td>
<td>3632</td>
<td>0.61</td>
<td>0.61</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

P-Values in columns 3, 5, 7, 9 in italics. The constants refer to a Bavarian/Austrian (col.2-5), a not further specified Mediterranean (col. 6/7), and an Eastern European one (col. 8/9). We also included a dummy for those aged 51-59 in order to control for a potential selection, but it was never significant. The weighted number of cases (adjusting for aggregated observations using square roots) is for the three regions 189 6, 86 and 990, respectively.

As we are working with grouped data, special estimation problems could arise. See, however, appendix 1, where we demonstrate that even the exclusion of observations with N > 1 does not change the results substantially.
Source: see Table 2

**Table 4a**

DESCRIPTIVE STATISTICS

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Climate warm</td>
<td>36</td>
<td>9.20</td>
<td>9.40</td>
<td>9.32</td>
<td>0.06</td>
</tr>
<tr>
<td>Gender inequality</td>
<td>36</td>
<td>4.14</td>
<td>6.11</td>
<td>5.37</td>
<td>0.5</td>
</tr>
<tr>
<td>Urban share</td>
<td>36</td>
<td>1.00</td>
<td>19.30</td>
<td>4.68</td>
<td>4.64</td>
</tr>
<tr>
<td>Social inequality</td>
<td>36</td>
<td>0.37</td>
<td>5.02</td>
<td>1.6</td>
<td>1.66</td>
</tr>
<tr>
<td>Population density</td>
<td>36</td>
<td>3.75</td>
<td>40.19</td>
<td>11.38</td>
<td>8.33</td>
</tr>
</tbody>
</table>

| Valid N (listwise)| 36 |

**Table 4b**

TWO REGRESSIONS: DETERMINANTS OF HEIGHT

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>144.24</td>
<td>0.00</td>
<td>164.37</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Climate warm</td>
<td>2.97</td>
<td>0.52</td>
<td>0.82</td>
<td>0.84</td>
<td></td>
</tr>
<tr>
<td>Gender inequality</td>
<td>-0.31</td>
<td>0.50</td>
<td>-0.29</td>
<td>0.46</td>
<td></td>
</tr>
<tr>
<td>Urban share</td>
<td>0.16</td>
<td>0.23</td>
<td>0.2</td>
<td>0.14</td>
<td></td>
</tr>
<tr>
<td>Population density</td>
<td>-0.06</td>
<td>0.37</td>
<td>-0.08</td>
<td>0.20</td>
<td></td>
</tr>
<tr>
<td>Roman Bath/ Roman</td>
<td>-2.05</td>
<td></td>
<td>0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technology</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social inequality</td>
<td></td>
<td></td>
<td>-0.17</td>
<td>0.58</td>
<td></td>
</tr>
<tr>
<td>Mediterranean</td>
<td>-1.66</td>
<td>0.05</td>
<td>-1.67</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>North-Eastern Europe</td>
<td>1.17</td>
<td>0.03</td>
<td>0.89</td>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td>Antiquity</td>
<td>-1.68</td>
<td>0.01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Late Medieval Period</td>
<td>-0.48</td>
<td>0.52</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Modern (15th to 18th c.)</td>
<td>-0.76</td>
<td>0.59</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adj. Rsq</td>
<td>0.33</td>
<td></td>
<td>0.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>36</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

P-Values in columns 3, 5 in italics.
Weighted Least Squares Regression: number of cases adjusted for aggregated observations using square roots.
Constant refers to a hypothetical height value for the Early Middle Ages, and Central/Western Europe.

Source: see Table 2 and text.
Figure 1a

MALE HEIGHT DISTRIBUTION, ALL CENTURIES

Source: see Table 2
**Figure 1b**

FEMALE HEIGHT DISTRIBUTION, ALL CENTURIES

Source: see Table 2
Figure 2

HEIGHT DEVELOPMENT, 1st TO 18th CENTURIES

(IN CM, MALE AND FEMALE)

Source: see Table 2. The level of heights was adjusted to average European male heights (using the regional coefficients and weighting them with sample weights).
Figure 3

HEIGHT DEVELOPMENT BY MAJOR REGIONS (IN CM)

Source: see Table 2
**Figure 4**

TWO-AXIS-DIAGRAM

HEIGHT DEVELOPMENT BY GENDER, 1st TO 18th CENTURIES (IN CM)

Source: see Table 2
Figure 5

HEIGHT AND TEMPERATURE DEVELOPMENT (1st TO 18th C),
BASED ON GLACIER MOVEMENTS AND TREE-RINGS

Source: see Table 2
Figure 6

DEVELOPMENT OF INEQUALITY

Source: see Table 2
**Figure 7**

HEIGHT DEVELOPMENT AND GDP PER CAPITA

Source: see Maddison (2001) and Table 2. Countries are excluded for which no height data is available.