

Joerg Baten

**Strategies to cope with potential labor market bias and other selectivities in height research**

*You can cite this version as “Working Papers in Economic History, University of Tuebingen (version August 2015) Number 91 b)”, but please contact the author about it before doing so.*

How can we measure the welfare development of historical populations? Evidence on early development is quite scarce, especially if we are not only interested in today’s rich countries in northwestern Europe and North-America, but want to take a global view on economic development. Heights and anthropometric research have offered a complement to conventional welfare indicators such as national income and real wages. The method is based on the influence which nutrition and the absence of disease have on average height. During childhood and adolescence the quality and quantity of nutrients supports bodily growth, while diseases absorb nutrients and lead to stunting. This height measurement project has produced a number of important insights, which have been documented in overviews such as Steckel (2009), Komlos and Baten (2004) and Floud (et al. 2011). Recently Bodenhorn, Guinnane and Mroz (2013, 2015) argued that the anthropometric history project of the last decades might have been the area of economic history research that received most resources (although this might be a question of definition).

There is also strong interest from outside of economic history in long-run developments, from growth and development economics for example (Deaton 2001; see the review in Moradi and Baten 2005). The advantages of height research relative to other indicators are for example that evidence is available for countries for which GDP estimates of reasonable quality might not be possible. In addition height often covers a large part of the

population<sup>1</sup>. In many historical societies for example self-employed peasants and farmers form the majority of the population. Their welfare development cannot be estimated with other methods. In addition, the intra-household allocation of nutrition and health related goods can be documented using height, for example the welfare of housewives can be estimated using anthropometric methods.

Bodenhorn, Guinnane and Mroz (2013, 2015) (referred to as BGM in the following) have recently argued that a substantial part of anthropometric research is based on “selective sources” and this did cast some doubt on the largest project of economic history over the last decades. They have suggested that selectivity issues are invalidating many anthropometric studies. In response, one recent study by Zimran (2015) analyzed whether changing selectivity of height samples exists for the US for the early and middle of the 19<sup>th</sup> century. He also assessed whether the “early industrial growth puzzle” (i.e., the puzzling observation that heights declined at a time when incomes increased) was only caused by BGM-type selectivity. The idea of Zimran was to merge the height sample of the volunteer army with census data. Thus he also has the observables on the non-soldiers, people who decided for the civilian sector. As another important source in this context, he linked this sample with an election sample of 1856 and 1860. This allows obtaining a variable which is relevant for the decision regarding joining the military or not: Since during the elections some candidates were either extremely pro-union or pro-secession and pro-slavery, also the voters were divided into two camps. Voters of the first political group could be seen as persons more likely to join the ranks of the northern Union Army, whereas voters of the secessionist political movement would be more likely to join the southern confederate army when the civil war broke out. So using this exogenous variation, which might influence the likelihood of joining one of the two military units, allows to study whether selectivity plays an important

---

<sup>1</sup> Other welfare indicators often concentrate on specific social groups, such as urban wage earners.

role. In the end Zimran finds that selectivity played a certain role -- the decline of height was not as large as estimated without adjusting for selectivity. Nevertheless, he finds that the height decline remains statistically significant. Hence the early industrial height decline remains as a stylized fact, namely as anthropometric values declined significantly between the 1830s and the 1840s.

In the following text I will suggest a number of strategies that might lead to more clarity about whether potentially selective height samples are indeed likely to suffer from changing selectivity. These strategies will help to mitigate potential selectivity problems. I will also present new evidence for countries in which the height of potentially selective samples can be compared to censuses of height.

### **The Issue**

Since scholars work with samples in order to learn about human behaviour in general, a standard issue is to which degree these samples are representative of the underlying population. Sometimes samples might not be completely representative in terms of levels, but even if this is the case, changes over time can shed light on important developments. In order to interpret these changes, it is important to study potential changes of the representativeness over time.

This applies also to research on the development of human stature, because some samples might be potentially plagued by potential labor market bias or other selectivities. In earlier studies height research has focused on various aspects of potential bias. Particularly prominent were considerations of minimum height requirements of military samples (Komlos and Kim 1990 give an overview), but this issue can be considered to be mostly solved (even standard software packages such as STATA incorporate truncated regression procedures such as “truncreg”). Also other selectivities were studied, such as social biases of samples relative

to the overall population as captured in censuses Baten (1999) compared to the occupational structure of prison samples and census samples, for example. Sometimes the parts of articles which discussed these issues were shortened or moved to online appendixes, because they tended to be quite technical in nature, and not always appealing a broader readership.<sup>2</sup> Some editors of journals preferred to have the most exciting and interesting parts of the analysis being published in the main text. In fact, much more work on selectivity was already done in previous height research than is might appear just from reading articles in well-known journals.

Selectivity is not only an issue of height studies but of many other samples that do not cover the whole population or a part of the population which can be considered to be representative. Hence the strategies suggested in the following apply to other research areas as well. Almost any evidence that we have can potentially be filtered, both by the intentions of contemporaries (any national income estimate is based on decisions of contemporary statisticians), administrative procedures that we cannot fully understand today, or often also by sheer luck – for example, some archives were destroyed by bombs in world wars, others were not.

If the height project would be dismissed because of the "doubt of selective samples" we would not have alternative indicators that do not suffer from potential selectivity. It is sometimes overlooked that some alternative indicators of economic development such as city growth rates or real wages of urban building laborers (and their cost-of-living baskets) only inform us about the urban population. I still find the information provided by studies based in these sources highly interesting and would never like to miss this work. I just mention this to put the harsh criticism of anthropometric work into perspective.

---

<sup>2</sup> Stolz et al. is an example: a lot of selectivity tests had to be moved to the Appendix. BGM (2015) criticize it based on reading only the main text. #check#

We need to distinguish between types of sources by “degree of selectivity risk”. Most at risk are height samples that consist of individuals having made a decision whether to be included in the sample. This is most typically the case in samples of volunteer armies and prison records. If a person joined a volunteer army, they typically made a decision about what they could expect from joining this volunteer army and what they could expect from alternative incomes and alternative career strategies. Similarly, if they thought about committing a crime, the economics of crime literature has shown that in particular property related crimes are following certain economic decision-making processes. For example, if a person feels that he would be able to steal some valuable item and that the likelihood of prosecution is low then he might commit the crime. However, the risk of being punished is usually difficult to estimate. He might still be caught by the police, in spite of low perceived likelihood. Individuals who are relatively poor, might be more tempted to risk to be imprisoned. In contrast, individuals who have a relatively good income, or optimistic individuals living in a growing economy, are weighing their opportunity costs and might not decide for the risk. It is important to note that this describes the theoretical possibility of changing labor market bias. It is an open empirical question whether this was the case in reality.<sup>3</sup>

These opportunity costs might change over time, and this is why volunteer armies and prison samples might be most at risk of being selective samples or underlying changing selectivity over time. So the obvious problem is the dependence on the opportunity costs as determined by the labor market: if the military sample was stable but other sector wages increased, groups of healthier and therefore taller, potential recruits may have decided for other sectors.

---

<sup>3</sup> Zimran finds that there is some evidence for the US in the mid 19<sup>th</sup> century, but that the extent was limited.

In addition, a second and potentially more severe issue arises: if the military paid a constant premium for additional health human capital (proxied by height), whereas the private sector did pay higher “health premia”, one could imagine the taller persons to move to the private sector – especially in a period of rising average income. However, is there any evidence for such a behavior in historical time periods? So far, there is little evidence, it is merely an assumption suggested by BGM (2014a, 2014b, 2015).<sup>4</sup> In fact, one could imagine that the military would be just as likely, if not more likely to treat taller soldiers better (for evidence on this, see the Appendix to BGM 2014a). Nobody knows whether any of these assumptions is true for a substantial number of countries for the period between, say, the 17<sup>th</sup> century and the 1970s (i.e., before systematic socioeconomic panels were collected in rich countries). It would require demanding additional research to find out about the existence of height premia. It is even unclear whether this research can be done with data that could be potentially mobilized for historical periods. We will discuss below some cases for which the height of selected and the overall population – including heights of persons selecting the civilian sector -- can be compared.

### **How large is the share of potentially biased samples used in global height research?**

In the first study on height selectivity, Bodenhorn et al. (2013) provided the impression that height research in general is problematic because of selective samples, in the second study they assessed the 19<sup>th</sup> century decline in the US and UK in a more focused way. Given the generalization that potential selectivity applies to “the height literature”, we will also discuss the question: how problematic is selectivity for the majority of height studies? How many height studies include potential labor market biases? To approach this question

---

<sup>4</sup> See Zimran (2015) as noted above, but he also does not have height values of the persons working in the civilian sector, but needs to refer to electoral proxies to identify the effect.

Baten and Blum 2012a presented global estimates of height trends (Figure 1). The work is based on thousands of different samples for more than 150 countries. This allows to calculate that some 90% of global height studies are based on samples which are not underlying strong selectivity. Among others, the relatively unbiased samples are military samples based on general conscription. Napoleon and the French revolutionary armies were among the first who advocated general conscription in the modern era, and many European states and states in other regions of the world followed the French example. They made it obligatory for a male of a certain age to be included in the military and usually there are very few exceptions to the rule. Often conscripts were first medically examined and their height was measured. Afterwards it was decided whether they would have been unfit for some medical reasons. If they were not, a lot was drawn whether they were really included in the active military service or not (Heyberger 2007, Weir 1997, Baten 2009). Interestingly general conscription was not only performed for European countries, but also for countries on other continents: Baten, Pelger, Twrdek (2009) study a census that includes height information in Argentina for the birth decades 1870 to 1910 (see below). Grajales-Porras and López-Alonso (2011) have census information on 18<sup>th</sup> century Mexican heights (for the place of Puebla), based on the Revillagigedo census.

In addition, there are also a lot of anthropological measurements that are not subject to labor market selectivities. These were most systematically executed for modern developing countries from the 1980s until today, covering the birth decades from the 1950s onwards. But also before, since the late 19<sup>th</sup> century, there were a lot of anthropological measurements taken which were not socially biased and did not underlie changing labor market bias. They measured, for instance, everybody in a certain regional unit as a representative sample. Here the regional composition or the estimation of birth decades can be more critical issues (Baten and Blum 2012a, 2014, Miller 2012).

Another source are school measurements (again without labor market biases) that were typically taken quite early in industrial countries and modern rich countries, although the heights have to be adjusted for the age at measurement.<sup>5</sup>

### **How can we obtain hints on volunteer army and prison sample selectivity change?**

The BGM studies claim that basically no remedy can be taken if samples might potentially contain selectivities, only diagnostics are suggested (such as estimating the joint significance of recruitment year dummy variables, or the changing population share of the army).

However, this is a relatively extreme position. It might lead to agnosticism about some interesting developments in economic history. Most people interested in early developments especially in countries which did not have conscript samples or anthropological measurements might decide to take a more pragmatic position, namely to study selectivity indicators (relative to underlying census populations characteristics or other samples that might not be biased). If several appropriate measures do not indicate changing selectivity, it seems reasonable to attribute informative value-added to a sample. A number of strategies can be applied to limit the risk of selectivities.

1. First of all, height samples can be constructed that were recorded in only one year (or for a short span of time, such as the U.S. Civil War 1861-65) and which contain different age groups. The height trends can be analyzed by organizing the sample by birth decades, keeping the labor market conditions at the time of recruitment more or less constant. For instance, in the U.S. there is still a height decline between the 1820s and 1840s if different age groups are included and only a small number of years with similar labor market conditions are

---

<sup>5</sup> A border case might be slave samples. In order to be included in those, the individual did not make a decision whether to join and become a part of slaved population, but somebody else was making a decision, such a slave trading person or in some cases local chiefs. So the potentially underlying bias is different from “free” labor market bias. These biases have been carefully studied and finally rejected by Eltis (1986) and by Baten and Blum (2012). See also the discussion about selectivity of slaves in the US, Komlos #, Steckel #, Pritchett #.



analyzed (A'Hearn 1998). We performed regressions using the Union army data set for the year 1862 only, and found that the cohort of the 1840s was significantly shorter than the 1820s/30s cohort. The coefficient is only 70% of the size compared to including all war years. An almost identical results emerges using only the recruitment year of 1864 (1862 and 1864 were the two years with the largest recruitment numbers of the Union Army dataset). This is actually a similar reduction of the coefficient compared to what Zimran (2015) finds using completely different methods. Clearly, other potential biases (such as potential military unit biases) also need to be studied before using this strategy, as always.

2. Further we can compare “side information” in height samples. For example, the occupational share of skilled persons in volunteer armies or prisons can be compared with the occupational share of the same occupational category in a census or in other samples not affected by changing labor market bias. There are a number of earlier studies who performed this comparison for the level, although not for the change (for example, Baten 1999; see also Steckel #, Johnson and Nicholas 1995). This comparison between observables of sample and census population can give hints about potential biases of the sample. It might be less revealing about selectivity driven relative civilian height premia assumed by BGM -- theoretically the taller unskilled daylaborers might still have preferred civilian occupations in rapidly developing economies. But in reality it is difficult to imagine that only height premia change, and nothing happens to the share of skilled occupations. Hence, the skilled share can give an important hint about whether this is likely. In sum, a substantial number of convincing indicators that do not indicate changing selectivity makes it more likely that changing labor market bias is not a problem (or otherwise that it is a problem).

It is necessary to note here that just controlling for occupations in a multiple regression is not necessarily sufficient. For instance, the share of skilled occupations in the underlying population as reported in different censuses might also have changed, especially in

industrializing societies. In addition, height and occupational status are not perfectly correlated. However, the signalling function of the skilled occupational share is what might be helpful here in identifying the direction of potential bias.

3. Other pieces of “side information” could be derived from the basic numeracy of sampled individuals in comparison with the numeracy based on censuses. Over the last decade, a vast literature has developed using age-heaping based numeracy estimates (A’Hearn et al. 2009, de Moor and van Zanden 2010, Humphries and Leunig 2009). The comparison of the numeracy of a height sample with the numeracy of a census sample was firstly applied by Stegl and Baten (2009) for a study of the Middle East (See Table 1 for the example). Unfortunately, this strategy cannot be applied to samples from the UK, the U.S. and similarly rich and relatively well-educated countries after the mid-19<sup>th</sup> century, because the basic numeracy problem did no longer exist at this development level, but for the critical period of the early 19<sup>th</sup> century it can yield important insights as well as for the rest of the world. In addition, soldiers might be upwardly biased -- but changes over time might still be informative.

New age-heaping based measurement concepts based on the 17-22 year-olds are particularly suitable for prisons and volunteer armies, because the standard numeracy measure for the age groups 23-72 requires the assumption of a relatively uniform distribution of end-digits (Prayon 2013). This is often violated for prison age distributions that often peaked in the early 20s. There were more 23-year-olds than 30-year-olds in most historical prisons, for example. The formula to calculate this ABCC2-Index is:

$$ABCC2 = \left(1 - \frac{(WI2 - 100)}{100}\right) \times 100 \text{ if } WI2 \geq 100; \text{ else } ABCC2 = 100$$

where WI2 is a transformation of the Whipple Index:

$$WI2 = \left( \frac{n_{18} + n_{20} + n_{22}}{\frac{1}{2} \times (n_{17} + n_{18} + n_{19} + n_{20} + n_{21} + n_{22})} \right) \times 100 \text{ if } WI2 \geq 100; \text{ else } WI2 = 100$$

4. Another strategy is to compare different height samples from different institutional contexts. Even if different samples might be affected by various potential selectivities, it is not always the case that different institutions generated identical directions of bias (Komlos #).

5. BGM suggest in addition a diagnostic of volunteer military samples to consider the changing share of the population that was recruited into the army, relative to the total labor force. If the share of the population recruited became substantially smaller compared to the underlying population of potential recruits, it might seem more likely that the selectivity changed. This was recently analyzed for example by Stolz et al. (2013) for the case of Portugal, where the share of the volunteer army remained quite similar in spite of a height decline (although income and height development moved in a similar direction in this country during the 18<sup>th</sup> century). This would be another hint that changing labor market bias might not have played a substantial role in this case.

6. Moreover, BGM suggest to include dummies for recruitment periods in addition to the dummy variables for birth decades, a technique that has been applied before.<sup>6</sup> This technique might lower the probability of labor market bias affecting the birth decade coefficient values.<sup>7</sup> BGM (2015) argued that including recruitment year dummies is not

---

<sup>6</sup> For example A'Hearn (1998) included recruitment year dummies, although he did not combine the pre- and post 1840-born union army soliders in his main regressions.

<sup>7</sup> Only if the sample is reduced to a small number by excluding all soldiers above 30, significance disappears (BGM 2015).

sufficient to remove the bias because there could be collinearity effects and other estimation problems<sup>8</sup>.

7. A systematic comparison between potentially biased and probably unbiased samples can perhaps reveal the most important insights, because also the selectivity on height can be assessed rather than the selectivity on observables. However, only final adult heights can be compared directly for the birth cohorts (see following section). Baten, Pelger and Twrdek presented evidence on potentially selective prison and military samples for Argentina and “height census sample” (Figure 3). This is one of the rare cases in which height of the total population can be compared with potentially selective samples. While the levels of height are different as expected, the stagnation between 1880 and 1900 is identical between potentially biased and unbiased (census) height samples. Hence we can summarize that a small number of unselective adult height samples exist and they allow to evaluate potentially selective samples.

8. A quite labor-intensive strategy could be to match the individuals who joined a volunteer army by name and place of birth and birth period to census samples that include the whole population. For example for the UK there are census samples for 1851 and 1881 censuses and it might be possible to identify the people who joined a volunteer army. Although matching by name, place of birth and birth period also poses some risk of measurement error. For South Africa a plan exists to match volunteer army soldiers with genealogical censuses of the population that include the South African population. However, this research plan depends currently on the access policies of the South African archives.

### **Studying other potential biases apart from labor market selectivity**

---

<sup>8</sup> It probably does not make sense to include individual year dummies as there might be random fluctuation, especially if the samples are not very large per year of recruitment or per year of birth. It is good practice in the literature to use decades or five-year-birth groups to average out random fluctuation.

Finally, studying histograms and normality tests can provide indications of sample distortions such as military height requirements. However, it is important to emphasize that these techniques are not a good indicator for labor market selectivity.

In addition, the regional composition of a sample compared to the underlying regional composition of the population we are interested in should be analyzed, because it might be that changing regional bias is often a stronger determinant of uncorrect average height change than labor market bias. In contrast to BGM (2015) who assume that proximity to protein effects cannot explain the Irish case, Baten (2009) has shown that a regional effect of proximity to dairy farming exists if non-selective samples are analyzed even controlling for income and other variables. It is a strong predictor of regional differences. Hence regional height differences need to be taken into account. The strong attention which labor market biases received recently should not lead to forget about other potential biases.

If volunteer armies are studied, it is also necessary to carefully study potential biases arising from military soldier categories, such as grenadiers and artillery army units. These types of biases have been studied intensively in the past, but if a new bias such as labor market selectivity bias is getting into the central stage of attention, some older biases might be ignored.

### **Final adult height and height of individuals that are still growing**

Some samples that are not biased by selectivities are slightly difficult to analyze because the individuals were still growing at the time of measurement. For example, BGM compare the height decline in the U.S. and UK volunteer armies with a stagnation or relatively small increase in general conscription samples such as in the Netherlands, Italy, France and Sweden. Perhaps the difference in trend was caused by the fact that the U.S. and UK volunteer soldiers were a selective sample, whereas the other countries had general

conscription? However, Baten (2000) argues that there is another issue that prevents us from comparing the short run movements of conscript sample time series and the military volunteer sample time series. There is an issue of timing of height measurement. If still-growing individuals are measured, the 1-2 years before measurement can be decisive for overall height level (Baten 2000). The body „waits“ for a better situation. For instance, if the individual is 20-year olds and the nutritional situation is favorable, then he will quickly finish his growth and we can interpret for example the height at age 21 is the final, et al. However, if the situation is adverse and might have been bad already in the previous years, then the body “waits” and does not grow any more. There is a biological logic behind this: If nutritional supply is not improving, it is better to stay relatively short, because a short individual has the opportunity to survive with less nutrients per capita (and eventually have children later-on). In contrast, an individual that opts for a very tall height level will require too many calories and proteins and might die from malnutrition or related risks. In 19<sup>th</sup> century Europe the growth of males continued often until the age of 23 or later. In this period, the good harvests of the 1820s and 1860s and the 1840s/early 1850s famines played a particularly important role for the short-run variation of height. This year-before-measurement effect often made birth decade effects invisible. For instance, 20.5-year-old South German conscripts measured in the late 1840s seemed short, and those measured in the late 1860s seemed tall (even if the latter were born in the hungry 1840s). Unfortunately, we have very few samples of generally conscripted males at age 23 and more for the mid- 19<sup>th</sup> century.<sup>9</sup> In Figure 1 we can observe a comparison of real wages, male heights (generally conscripted military recruits from lists of 20.5-year-olds) in Southern Germany (Bavaria), who were organized by the year-before-measurement). This is compared to – potentially selective – adult convicts in the same region, who are organized by birth decades. According to the graph we can notice that these two

---

<sup>9</sup> We have one sample for France for the late 19<sup>th</sup> century, however in this case we do not have the famine effects anymore. (Lion #Kestenbaum#, #)

series are correlated. Anthropometric values were high in the 1845-1849 period and then strongly declined up the 1855-1859 period. They recovered in the early 1860s and then fell again in the late 1860s. One of these two height developments is organized by birth decades, the other by year-before-measurement decades. Both mirror the real wage series for this region quite well. In Baten (2000) there are more examples of studies which confirm the observation that the year-before-measurement plays a large role for the individuals measured who are still growing. This does not invalidate conscript samples in general. They still provide a good indication of overall height levels especially as we can normally estimate the remaining growth after age 21 and so on. But interpreting the short-run variation during the 1820s-1860s period in Europe might be more complicated, if conscripts were still growing. On the other hand, we can note that the correlation between potentially biased prisoner heights and unbiased conscript statures (by decade of year before measurement) suggest that changing labor market bias did probably not play a large role in Southern Germany for prisoners.

## **Conclusion**

To sum up, in general the criticism of “height samples being based on selective samples” is of limited validity, because it refers to only few samples of globally available studies. Even for those it is mostly a theoretically plausible possibility that samples are biased. It might apply to a higher share of studies that are trying to trace the development of early 19<sup>th</sup> century U.S. and UK. If the BGM criticism would be taken too seriously, there might be a risk of agnosticism that people will shy away from analyzing the samples that could be potentially selective. Especially in case we want to trace the welfare history of developing countries and of modern rich countries as well, we cannot afford the ‘luxury’ of agnosticism, because height research is one of the crucial tools in the toolbox of the modern economic historian. Nevertheless,

paying attention to selectivities is crucial in all research dealing with potentially selective samples. Hence this article suggested a number of strategies that can help to diagnose changing labor market bias or its absence.



## References

- A'Hearn, Brian, 1998. "The Antebellum Puzzle Revisited: A New Look at the Physical Stature of Union Army Recruits during the Civil War", in *The Biological Standard of Living*. Edited by John Komlos and Joerg Baten. Stuttgart: Steiner, 1998.
- A'Hearn, Brian, Joerg Baten and Dorothee Crayen, 2009. "Quantifying Quantitative Literacy: Age Heaping and the History of Human Capital", *Journal of Economic History* 69-3, pp.783-808.
- Baten, Joerg, 2000. "Economic Development and the Distribution of Nutritional Resources in Bavaria, 1797-1839", in *Journal of Income Distribution* 9, pp. 89-106.
- Baten, Joerg, 1999. *Nutrition and Economic Development in Bavaria, 1730-1880*. Diss. University of München.
- Baten, Joerg, 2009. "Protein Supply and Nutritional Status in Nineteenth Century Bavaria, Prussia and France", *Economics and Human Biology* 7-2, pp. 165-180.
- Baten, Joerg, Ines Pelger and Linda Twrdek, 2009. „The Anthropometric History of Brazil, Lima (Peru), and Argentina during the 19th and early 20th Century", *Economics and Human Biology* 7-4, pp. 319-333.
- Baten, Joerg and Matthias Blum, 2012a. "An Anthropometric History of the World, 1810-1980: Did Migration and Globalization Influence Country Trends?", *Journal of Anthropological Sciences* 90, pp. 221-4.
- Baten, Joerg and Matthias Blum, 2014. "Why are you tall while others are short? Agricultural production and other proximate determinants of global heights", *European Review of Economic History* 18, pp. 144–165.
- Bodenhorn, Howard, Timothy Guinnane and Thomas Mroz, 2013. "Problems of Sample-selection Bias in the Historical Heights Literature: A Theoretical and Econometric Analysis", Economic Growth Center Discussion Paper No. 1023, Yale University.
- Bodenhorn, Howard, Timothy Guinnane and Thomas Mroz, 2015. "Sample-Selection Biases and the 'Industrialization Puzzle'", NBER Working Paper 21249, National Bureau of Economic Research.
- Deaton, 2001. "Health, Inequality, and Economic Development," Papers 200, Princeton, Woodrow Wilson School - Development Studies.
- De Moor, Tine and Jan Luiten van Zanden, 2010. "Girl power: the European marriage pattern and labour markets in the North Sea region in the late medieval and early modern period", *Economic History Review* 63, 1, pp. 1-33.

Floud, R., R. Fogel, B. Harris and S. Chul Hong, 2011. *The Changing Body Health, Nutrition, and Human Development in the Western World Since 1700*. Cambridge University Press.

Grajales-Porras and López-Alonso (2011). Physical stature of men in eighteenth century Mexico: Evidence from Puebla. *Economics and Human Biology* 9-3, 265-271.

Heyberger, Laurent, 2007. "Toward an Anthropometric History of Provincial France, 1780-1920", *Economics & Human Biology*, Vol. 5, No. 2, pp. 229-254.

Humphries, Jane and Timothy Leunig, 2009. Was Dick Whittington taller than those he left behind? Anthropometric measures, migration and the quality of life in early nineteenth century London?", *Explorations in Economic History* 46, pp. 120–131.

Johnson and Nicholas, 1995. "Male and Female Living Standards in England and Wales, 1812-1857: Evidence From Criminal Height Records", *The Economic History Review*, New Series, Vol. 48, No. 3, pp. 470-481.

Kesztenbaum, Lionel (2011). "The Health Cost of Living in a City: The Case of France at the End of the 19th Century," *Explorations in Economic History*, 48(2):207-225, 2011

#Komlos (add)

Komlos, John and Joerg Baten, 2004. "Looking Backward and Looking Forward: Anthropometric Research and the Development of Social Science History," *Social Science History* 28(2), pp. 191-210, also as working paper: Muenchner Wirtschaftswissenschaftliche Beitrage 2003-15.

Komlos, John and J. Kim, 1990. "Estimating trends in historical heights", *Historical Methods* 23, pp. 116-120.

López-Alonso, Moramay and Agustín Grajales-Porras, 2011. "Physical stature of men in eighteenth century Mexico: Evidence from Puebla", *Economics and Human Biology* 9, pp. 265–271.

Miller, Melinda, 2012. "Measuring the Measurers: Selection in the Boas Sample of American Indians".

Moradi, Alexander and Joerg Baten, 2005. "Inequality in Sub-Saharan Africa 1950-80: New Estimates and New Results", *World Development* Volume 33-8, pp. 1233-1265.

Prayon, Valeria, 2013. *Contributions to the discussion on the determinants of long-term human capital development in today's developing regions*. Diss. Universität Tübingen.

#Pritchett

#Steckel (add)

- Stegl, Mojgan and Joerg Baten, 2009. "Tall and Shrinking Muslims, Short and Growing Europeans: an Anthropometric History of the Middle East, 1840-2007", *Explorations in Economic History* 46, pp. 132-148.
- Stolz, Yvonne, Joerg Baten and Tarcisio Botelho, 2013. "Growth effects of 19th century mass migrations: "Fome Zero" for Brazil?", *European Review of Economic History* 17-1, pp. 95-121. Older version: Tübingen Working Papers in Economics and Finance 2. V.
- Weir, David R., 1997. "Economic Welfare and Physical Well-Being in France, 1750-1990." In *Health and welfare during Industrialization*. Edited by Richard H. Steckel and Roderick Floud. Chicago: University of Chicago Press.
- Zimran, Arielle, 2015. "Does Sample-Selection Bias Explain the Industrialization Puzzle? Evidence from Military Enlistment in the Nineteenth Century United States". Working Paper: Northwestern Univ.

Figure 1: Global height development. Source: Baten and Blum (2014)

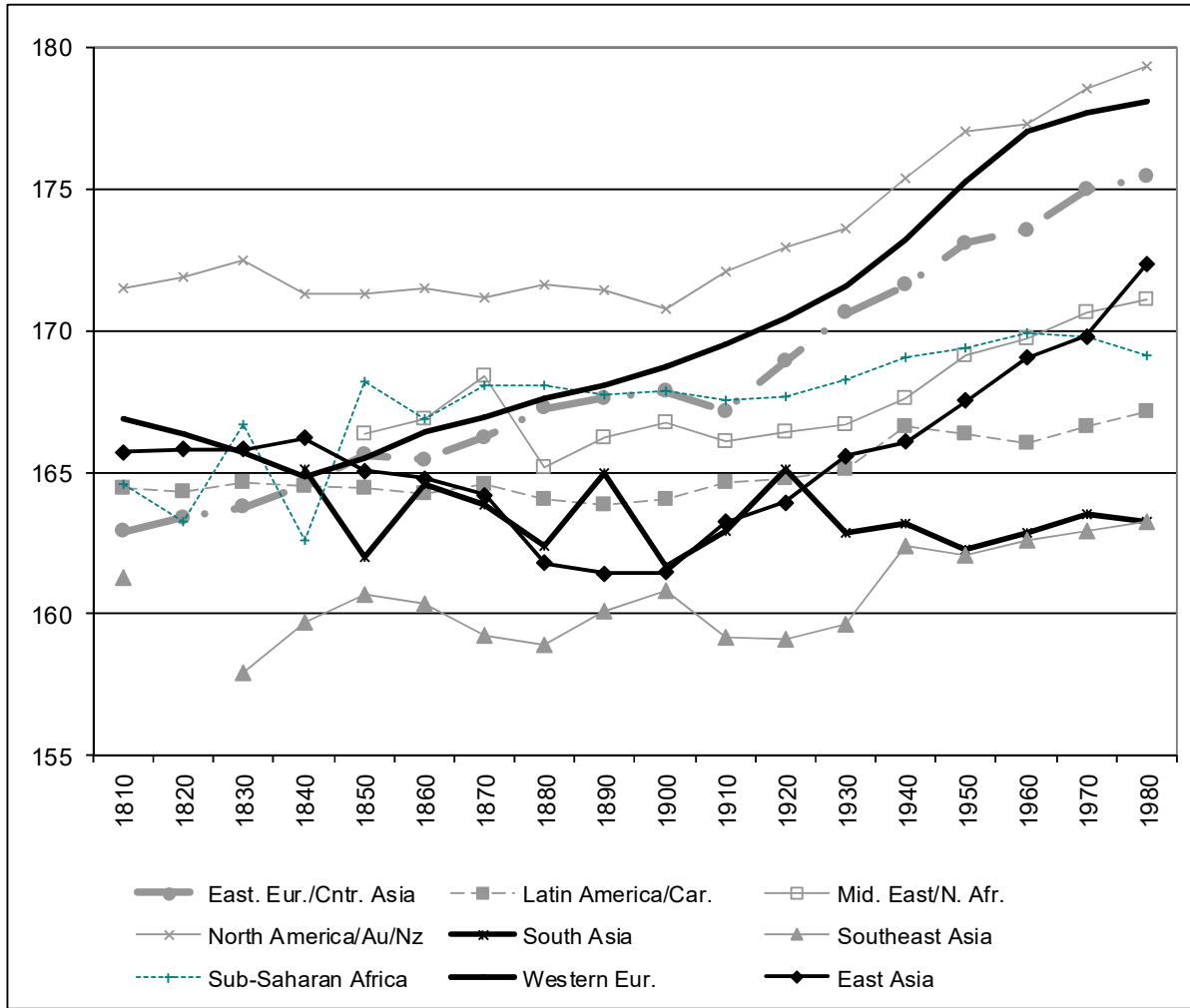
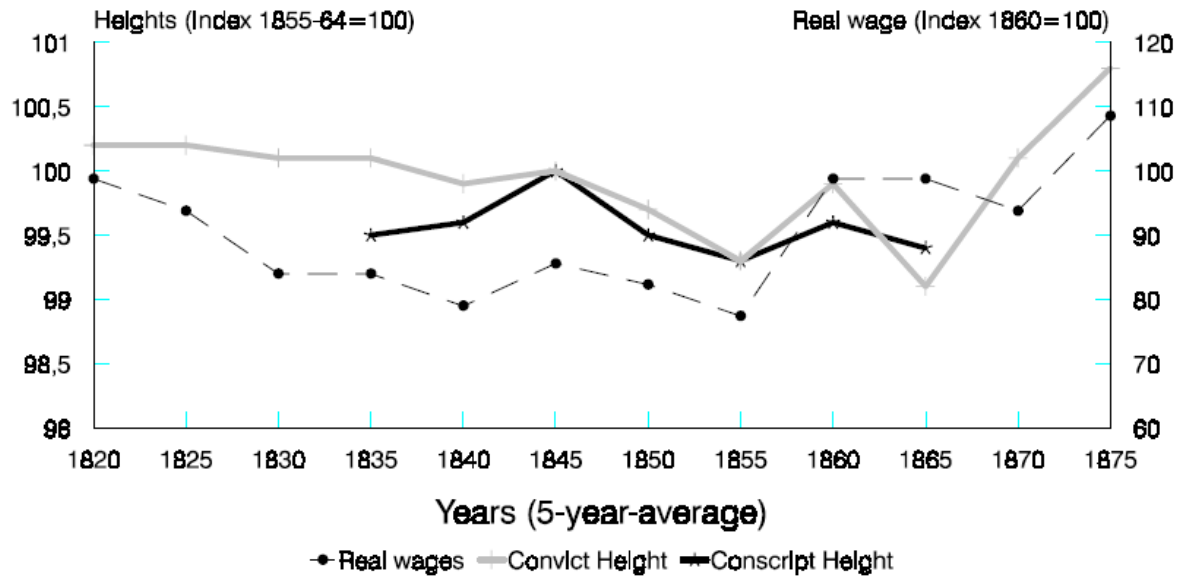
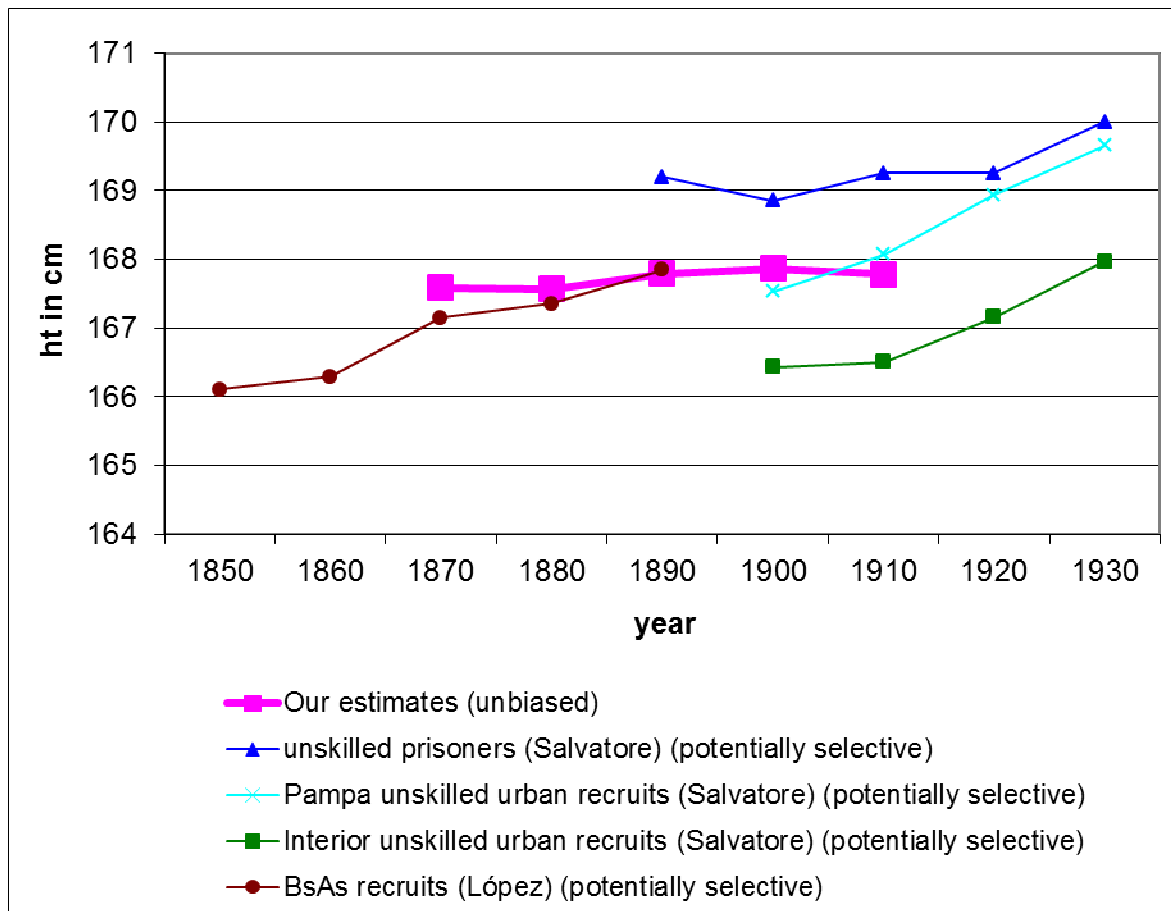


Figure 2: Real wages, male heights (general conscription military lists of 20.5-year-olds by year before measurement (YBM) and potentially selective adult convicts by birth year (BY)) in Bavaria



Note: YBM – heights aggregated by year before measurement, BY: by birth year. Source: Baten (2000), more country examples there

Figure 3: Comparison of various height estimates for Argentina (Source: Salvatore, 2007)



Source: Baten, Pelger and Twrdek (2009)

The last cohort of “Our estimates” contains 17 and 19- year olds and may be downward biased