Globalization, industrialization and welfare: 
The heights of 27 million Italian conscripts 1855–1910

Brian A’Hearn, Franco Peracchi and Giovanni Vecchi
Franklin & Marshall College and University of Rome “Tor Vergata”

Triangle Economic History Workshop, November 16, 2006
1 Introduction and background

- Random genetic variation of potential final height is typically described by a normal distribution.

- Its mean depends on net nutrition, especially in infancy: quality and quantity of nutrients (especially protein), disease burden and workload.

- Another, perhaps less important factor, is genetic differences between populations.

- Speed of growth is also affected by net nutrition. Body’s reaction to nutritional stress is to slow but also prolong growth. Historically, growth continued into the early 20s and onset of puberty was significantly later than today.

- Net nutrition is a broad measure of living conditions.

- It depends, directly or indirectly, on income: high income permits protein-rich diet, public health infrastructure, private access to health care, reduced workload, etc.

- Net nutrition may also depend on cultural and economic accidents, such as breastfeeding practices, type of diet and economic specialization (famous example: Irish skimmed milk after butter production for the English market, plus potatoes).

- Nonlinear impact of net nutrition on mean height due to genetically imposed upper bound on height. This implies that mean height may be affected by inequality. Redistribution will raise heights of favored group by less than it lowers them in impoverished groups.

- The dispersion and the shape of the distribution of heights may also be affected by inequality.

- Form of distribution under nutritional stress uncertain.

- Form of distribution before final height achieved is known to be affected by growth spurts. During puberty some individuals of a cohort rapidly reach their final height, or close to it, while others experience their growth acceleration only later. During this phase, negative skewness emerges.
To summarize:

- **Mean heights** may tell us about **average living conditions**, and (at least in the historical experience of today’s rich countries) are typically positively correlated with standard measures like per capita income.

- **Dispersion** and **shape** of height distribution may tell us about **inequality**, if we can control for the effects of **late growth spurt** in historical data.

- **Over-** or under-dispersion and **skewness** of heights are unambiguously associated with inequality, if growth spurt problem controlled.

- Mean may also be affected by inequality in living conditions (because of the nonlinearity between height and living conditions), but this effect is **hard to identify** separately from the impact of average conditions.

**Novelties of our paper**

- Fine geographical detail.

- Innovative age-adjustment procedure.

- Focus on the whole distribution of heights, not just its mean.
2 Sketch of Italian economic history: 1850–1910

Political unification

- **1848–49**: Revolutions of 1848–49 and first independence war.
- **1859**: Second independence war, incorporation of Lombardy (except Mantua), Parma–Piacenza and Modena–Reggio to the Kingdom of Sardinia.
- **1860**: Incorporation of Tuscany and Emilia-Romagna, cession of Nice and Savoy to France, Garibaldi’s expedition, and incorporation of Southern Italian regions, Marche and Umbria.
- **1861**: Creation of the new Kingdom of Italy.
- **1866**: Third independence war and annexation of Veneto and Friuli.
- **1870**: Annexation of Lazio.
- **1918**: Annexation of Trentino-Alto Adige and Venezia Giulia.
Economic unification

- **Backwardness**: Italy was a peripheral and very backward country.

- **Dualism**: North-South divide and the “Southern question”.

- **The 1870s**: Lack of growth.

- **The 1880s**: No consensus on this decade. The traditional story: agricultural crisis plus industrial flourishing. The “revisionist” view: Decade of growing prosperity.

- **Early 1890s**: The “darkest years”.

- **Mid 1890s–WWI**: Stability and relatively fast growth (discontinuity?). Catching up with Continental Europe begins.
Available empirical evidence

• All available aggregate welfare indicators (per-capita GDP, per-capita consumption, real wages, educational attainments, infant mortality, life expectancy at birth) show **improvements over time**.

• **Great geographic variation**: Improvements of indicators at the national level may mask very different patterns at the regional/provincial level.

• Data are **scarce** (especially consumption, wages, and income distribution data).
Issues in the economic history debate

- **Italian unification as globalization**: Removal of internal and external trade barriers, railroads and the creation of a common market, new tax system with a heavier burden, etc.

- **Gainers and losers after the unification**: Did Southern Italian regions gain or lose after the unification?

- **What do aggregate data hide?**: We only know something about mean levels, with some disaggregation at the regional level.

- **The late ’800: stagnation or growth**: Traditional estimates of consumption per-capita indicate falling real consumption or, at least, very slow growth between 1880 and 1900. A number of scholars (Federico, Fenoaltea) have recently challenged this “pessimistic” view arguing that it is all due to poor measurement of agricultural output.
How about heights?

- **Figure 1 in Arcaleni (2004):** Mean heights at age of visit and at age 20 of the cohorts born 1854–1980.

- Until the mid-1900, Italians were **short** in comparison to other countries. This is no longer true today.

- **Strong growth** for the cohorts born in **1945–65**. The growth seems to be **flattening out** for the cohorts born after 1970.
3 Our data

- Our height data are drawn from a series of reports published by the Ministero della Guerra from 1864 to 1932, covering men born from 1839 to 1910.

- We call these reports the Torre reports after Federico Torre (a mathematician/engineer and revolutionary of 1848 from Benevento) who headed the Direzione Generale delle Leve e della Trappa from its creation in 1861 until 1891.

- The Torre reports present summary tables with data on the height of all conscripts examined for fitness for service in the Italian army.

- The data consist of frequencies observed for particular height intervals. They are initially broken down by circondario then, from 1902, by province.

- For the cohorts born between 1839 and 1854, the height data are rather coarse (only 7 height classes) and do not cover the full country, as conscription was introduced in the South only in 1862 (cohort born in 1842), Veneto and Friuli were added in 1866 (cohort born in 1846), and Lazio in 1870 (cohort born in 1850).
Recruitment procedures

All men were to be entered in conscription lists during the calendar year they reached age 19, then examined for fitness for service, and potentially called to arms in the year they turned 20.

- Conscription operations were conducted by provincial draft councils, whose first task each year was to review the conscription rolls compiled by local officials. All men on the revised lists were required to appear at the local meeting of the provincial draft council to present any appeals, undergo a preliminary physical examination, and receive their draft lottery number.

- Up to the draft of 1874 (class of 1854), a system of exemptions was in place that aimed to ensure families were not unduly burdened by military service. At the draft council session, exempted people were sent home without being examined. The remaining conscripts were assigned a draft number by lottery and subjected to a physical examination to ascertain their fitness.

- A detailed list of physical and mental conditions disqualified from service. These included obvious defects (e.g. blindness, missing limbs) but also diseases (e.g. syphilis, malaria, pellagra). Conscripts were required to meet a minimum chest circumference and a minimum height set at 156 cm.

- Conscripts whose failure to pass the physical examination was due to a condition judged temporary were deferred to the next draft for renewed examination. Generally, poor health and insufficient muscular development were “deferrable” conditions. University students were allowed to delay participation in the draft until the completion of their studies.
• Conscripts passing the physical exam were drawn, in order of draft number, until the annual intake specified in the enabling legislation was filled. The total was divided into an active-duty contingent and a contingent destined for training and assignment to the reserves.

• It was permissible to provide a substitute, purchase relief from the obligation to serve, or trade draft numbers, but only a handful possessed the resources to avail themselves of such possibilities.

• Essentially this system remained in effect throughout the period under study. A significant change in 1874 affirmed the principle of universal service and suppressed family exemptions, but those formerly entitled to a family exemption were assigned to a new third category that could be called for duty only under the most extraordinary service. They were now examined for fitness to serve, however.

• Minor changes in the regulations were introduced over time, including reductions in the minimum height requirement, to 155 cm in 18xx and 150 cm in 19xx.
Sample selection

- **Choice of period**: For the period 1854–1910, we have counts of heights by centimeter (cm) from 125 cm up. For the earlier period 1839–1853, height frequencies are reported for seven height intervals. These intervals are fairly wide, and the two extreme intervals are open-ended. Because of this, and the insufficient geographical coverage, these data are not used. We also ignore the year 1854 because the old system of exemptions were still in force.

- **Choice of geographical disaggregation**: The raw data are initially reported at the sub-provincial level of *circondario*. The number of *circondari* remains about the same (slightly over 200) for the 1850–1892 cohorts. Starting with the 1893 cohort (year 1913), data are reported at the provincial level (about 70 provinces). We aggregate to the provincial level. At this level we are able to maintain stability of the boundaries and sufficient geographical detail.

- **Province definitions**: Province names and definitions change over time. In the early years, the names of the Southern provinces are those of the former Kingdom of the Two Sicilies. We use the 69 provinces of the year 1920, but merge Caserta with Napoli. For the years before 1913 (cohorts 1839–1892), we build the provinces up from the *circondari*.
Main data problems

The main problems with our data are:

- Possible selection due to migration and draft evasion.

- Effects of differences in the age at measurement.

- Effects of rounding.
Migration and draft evasion

- In addition to the observed frequencies at different heights, the Torre reports give the total number of young men on the conscription lists, and the residual who were not measured.

- Figure 1 plots the number of registered and measured people by year. The fraction of men skipping measurement increases from about 10% around 1870 to about 20% around 1890, and stays at that level afterwards.

- In some regions and some years, the fraction of men skipping measurement is as high as 30-35% due to massive emigration and draft evasion.

- Emigration was particularly intense from Veneto and Friuli during 1870–90, and from the Southern regions during 1890–1910.

- A relatively less important problem is the fact that our data do not include conscripts for the Navy. According to Arcaleni (2004) “they are about 2-3% of total conscripts and no specific elements are available to suppose a sensitive bias in stature due to their lack”. For people serving in the Navy, the draft preceded that of the Army.
Figure 1: Number of registered and measured people by year.
Changes in the age at measurement

- Age at measurement varied. In general, there was a downward trend as a transition from conscription at age 21 to age 20 was delayed and dragged out for many years. In addition:
  - The classes of 1839–42 were called up at different ages in a series of 8 extraordinary drafts from 1859 through 1862.
  - The classes of 1845 and 1846 were both called in the same year during the mobilization of 1866.
  - In connection with a reduction in the size of the Army, the classes of 1847–49 were called at age 22.
  - During WWI the classes of 1895–98 were called at age 19, and those of 1899–1900 at age 18.

- The law authorizing and regulating each year’s call-up specified dates for the beginning of conscription operations (the sessione di leva), completion of the revised conscription lists, assignment of draft lottery numbers, examination of conscripts, and conclusion of the draft session. These dates varied from one year to the next but were the same for all provincial councils.

- Figure 2: Following Costanzo (1948), average ages at measurement are computed by assuming
  - uniform spacing of births during the year (average birthday on June 30th),
  - average date of measurement halfway between the beginning of the examination period and the end of the draft session.
Figure 2: Age at measurement.
Rounding

- Individual measurements were rounded to the nearest full centimeter, a minority to the nearest half, and only a handful to the nearest millimeter.

- It is likely that, when compiling height frequencies at each centimeter measurement, officials did not round those measurements that were more accurate, but truncated them by dropping the fractional part.

- Following Boldrini (?), Costanzo (1948) interpreted all the data as the result of truncation, and therefore added .5 cm.

- We instead follow Livi (?), who suggested that most reported heights were rounded to the nearest integer, and therefore we let a reported height of 165 cm stand as 165 cm.
Other data problems

- **Heaping**: Heaping at heights ending with 0 or 5 (e.g. 160 and 165 cm.) are clearly seen in many provinces and years. Occasionally, minimum height requirements (which fell from 156 to 150 cm over time) appear to exert an effect.

- **Top coding**: Although there should be no top-coding in the data, there is evidence of top-coding for some province/year.

- **Bottom coding**: Data are bottom-coded at 125 cm through the class of 1903. After 1903, heights below 125 cm are grouped in a variety of idiosyncratic intervals that change from one year to the next.

- **Accuracy of measurements**: Accuracy is likely to be higher for stature about the minimum threshold (varying from 150 to 156 cm.).

- **Gross errors**: There is evidence of gross errors, due to copying errors, inversion of digits, or addition of extra digits. In some cases, we managed to correct after inspecting statistics based on the data.

- **Height distributions with unusual shapes**: Como 1900 and 1903, Napoli 1866, Palermo 1895 and 1897, Messina 1900, Reggio Calabria 1900-01 and 1904, Treviso 1885, Venice 1882-98.

- **Internal migrations**: Maybe important only for the period 1890–1910. It is also mostly short-range (except for Rome).
4 Adjusting for differences in age at measurement

Previous approaches

- Costanzo (1948) proposed a method for adjusting mean heights at the national level.
- His method was adopted by the Italian National Statistical Institute (Istat).
- From a study of Danish conscripts by Mackeprang, which he claimed had typical results, Costanzo took growth estimates for ages 18–19, 19–20, and so on, using linear interpolation to infer growth during fractions of a year. With these figures, he then adjusted heights measured at younger ages up, those at lower ages down, to a common standard of 20 years.
- This procedure assumes too much uniformity in the growth process. It is not obvious that growth increments should be the same between Denmark and Italy, between two time periods, or even between two Italian provinces.
- For example, there are large differences between Mackeprang’s growth estimates and those of Ridolfo Livi, based on a longitudinal study of Italian recruits over time. Both data show a gradual deceleration of growth with age, but Livi’s data show higher rates growth rates which are likely to reflect differences in net nutrition.
4 ADJUSTING FOR DIFFERENCES IN AGE AT MEASUREMENT

Our method

Our method is based on three principles:

- Age-adjust the whole distribution of heights rather than a few moments of it (but respect the non-negativity and adding-up properties of relative frequencies).
- Age-adjust at the finest possible level allowed by the data.
- Since changes in the age at measurement occur at different points in time, fit a model with both age at measurement and calendar time as predictors.

Let the continuous random variable $H_t$ represent the variability of heights in a given province at time $t$. Let $h_1 < \cdots < h_J$ be $J$ thresholds for height, and let

$$S_{jt} = \Pr\{H_t \geq h_j\}, \quad j = 1, \ldots, J.$$ 

Also let

$$\pi_{jt} = \begin{cases} \Pr\{H_t < h_1\} = 1 - S_{1t}, & \text{if } j = 1, \\ \Pr\{h_{j-1} \leq H_t < h_j\} = S_{j-1,t} - S_{jt}, & \text{if } j = 2, \ldots, J, \\ \Pr\{H_t \geq h_J\} = S_{Jt}, & \text{if } j = J + 1, \end{cases}$$

be the relative frequency of heights in the various intervals. Clearly

$$\sum_{j=1}^{J+1} \pi_{jt} = 1 - S_{1t} + (S_{1t} - S_{2t}) + \cdots + (S_{J-1,t} - S_{Jt}) + S_{Jt} = 1. \quad (1)$$

Instead of directly modeling the $\pi_{jt}$, we use a logit specification to model $\pi_{1t}$ and the “discrete hazards”

$$\lambda_{jt} = \Pr\{h_{j-1} \leq H_t < h_j \mid H_t \geq h_{j-1}\}$$

$$= \frac{\Pr\{h_{j-1} \leq H_t < h_j\}}{\Pr\{H_t \geq h_{j-1}\}}$$

$$= \frac{S_{j-1,t} - S_{jt}}{S_{j-1,t}}, \quad j = 2, \ldots, J.$$
Our logit models are of the form

\[ \eta_{1t} = \ln \frac{\pi_{1t}}{1 - \pi_{1t}} = \alpha_1 + \beta_1 \ln a_t + g_1(t), \]
\[ \eta_{jt} = \ln \frac{\lambda_{jt}}{1 - \lambda_{jt}} = \alpha_j + \beta_j \ln a_t + g_j(t), \quad j = 2, \ldots, J, \]

where \( \ln a_t \) is the log of age at measurement in year \( t \) and \( g_j(t) \) is cubic time trend.

Our method adjusts for variations due to different ages at measurement and eliminates short-term effects and random noise, leaving only the “trend component”.

Given estimates \( \hat{\eta}_{jt} \) of the \( \eta_{jt} \), estimates of \( \pi_{1t} \) and \( \lambda_{jt} \) are obtained as follows

\[ \hat{\pi}_{1t} = \frac{\exp \hat{\eta}_{1t}}{1 + \exp \hat{\eta}_{1t}} , \]
\[ \hat{\lambda}_{jt} = \frac{\exp \hat{\eta}_{jt}}{1 + \exp \hat{\eta}_{jt}}, \quad j = 2, \ldots, J. \]

Finally, given estimates \( \hat{\pi}_{1t} \) of \( \pi_{1t} \) and \( \hat{\lambda}_{jt} \) of \( \lambda_{jt} \), estimates of \( S_{jt} \) and \( \pi_{jt}, j > 1 \), are obtained from the recursions

\[ \hat{\pi}_{jt} = \hat{\lambda}_{jt} \hat{S}_{j-1,t}, \]
\[ \hat{S}_{jt} = \hat{S}_{j-1,t} - \hat{\pi}_{jt} = (1 - \hat{\lambda}_{jt}) \hat{S}_{j-1,t}, \quad j = 2, \ldots, J, \]

with \( \hat{\pi}_{J+1,t} = \hat{S}_{jt} \). Because of (1) and (2) the estimates \( \hat{\pi}_{jt} \) satisfy both the non-negativity and the adding-up constraints.

In fact, because of the parametric nature of our specification, we obtain estimates \( \hat{\pi}_{jt}(a) \) of the “counterfactual” relative frequencies \( \pi_{jt}(a) \) of heights at a fixed age \( a \). We can therefore estimate the “counterfactual” cumulative distribution function (cdf) of heights \( \hat{F}_t(h \mid a) = \Pr\{H_t \leq h \mid a\} \) as follows

\[ \hat{F}_t(h \mid a) = \sum_{j \leq h} \hat{\pi}_{jt}(a). \]
5 Mean heights and average living conditions

Stylized facts: national growth

- **Figure 3**: Mean height grew by some 3.1 cm over 55 years: good performance, though slow by comparison to *Miracolo economico*.

- The only real wobble in this steady growth is a deceleration among the birth cohorts of the late ’60s and early ’70s, followed by a (partial) recovery. Hence, fastest growth early and no crises.

- **Figure 4**: Compared with existing series of Costanzo/Istat/Arcaleni, our series is **lower** because of the assumption about rounding vs. truncation, **smoother** because we use long-run-trend values, and **more linear** because of differences in age-adjustment.
5 MEAN HEIGHTS AND AVERAGE LIVING CONDITIONS

Figure 3: Mean of the raw and the adjusted (at age 20) national distribution of heights.
Figure 4: Our estimates of national mean height at age 20, Costanzo’s original estimates, and Costanzo’s estimates recentered.
Age standardization

- Growth took two forms: (i) increase in mean final height, and (ii) increase in the speed (decrease in the age) it was reached. As living conditions improve, both effects contribute to the increase in height at a particular age. At ages where growth is completed (e.g. 25), only the effect on final height matters. At ages where individuals are still growing rapidly, growth anticipation effect can also be important.

- **Figure 5**: Comparing trends in mean height at ages 18, 20 and 22 shows that the increase at age 18 was faster, because both effects were at work. The trends at ages 20 and 22 are parallel to each other. Because the increase at age 22 must reflect almost entirely the second effect, the parallel course of the trend at age 20 suggests that the increase at this is also due mostly due to the second effect.
Figure 5: Mean of the adjusted national distribution of heights at various ages.
Stylized facts: provincial diversity in levels and growth

- **Figures 6–7**: Great diversity in provincial mean heights at both beginning and end of sample. Range of age-20 provincial means is 157.7 - 165.6 cm in 1855, 160.6 - 168.2 in 1910.

- North-South gradient evident at both beginning and end, corresponding to other measures of development. Provincial variation within-region is evident as well (for example in North in 1855, or within Sicily in 1910).

- **Figure 8**: Growth was universal, if different in timing and intensity.

- **Figure 9**: The trend in the national mean is **not** due to composition effects.
Figure 6: Map of mean heights at age 20. Year 1855.
Figure 7: Map of mean heights at age 20. Year 1910.
Figure 8: Map of the changes in mean heights at age 20 between 1855 and 1910.
Figure 9: Mean of the adjusted (at age 20) national distribution of heights under alternative sets of provincial weights.
Stylized facts: convergence and divergence

- **Figure 10**: Over the entire period, there was scarcely any convergence.

- **Figures 11–10**: Evidence of $\beta$-convergence of provincial mean heights for the period 1855–1885, not for the period 1885–1910. There is stronger evidence of $\beta$-convergence for the first period and some evidence for the second period if one allows for two groups: (i) the Northern provinces (NW and NE), and (ii) the Central and Southern provinces (C, S and I), with the possible exception of Tuscany.

- **Figure 13**: $\sigma$-convergence nonexistent over the period. Standard deviation of provincial means was 2.12 cm in 1855, 2.09 in 1910.
Figure 10: Scatterplot of provincial means of adjusted heights in 1885 versus differences in provincial means between 1855 and 1910.
Figure 11: Scatterplot of provincial means of adjusted heights in 1855 versus differences in provincial means between 1855 and 1885.
Figure 12: Scatterplot of provincial means of adjusted heights in 1885 versus differences in provincial means between 1885 and 1910.
Figure 13: Interquartile range (IQR) and interdecile range (IDR) of the provincial mean of adjusted heights.
Economic interpretations

- Economic and political unification were bad for no one in the medium run. Growth was most rapid for the cohorts surrounding unification, and convergence was fairly strong for the cohorts born in the 10–20 year after 1855.

- The end of convergence in 1870–90 coincides with the completion of the national market through railroad construction, the “grain invasion” from North America and Russia, and a commercial war with France. All this might have slowed the improvement of living conditions in the South and accelerated it in the North.

- No evidence that *industrialization* and *urbanization* were bad for your health. The industrializing North of the late 1800s and early 1900s begins to diverge from the South again. Provinces with major cities show heights above average or above-nearby-provinces: Torino, Genova, Milan, Venice, Bologna, Florence, Rome, Naples, Messina, Palermo. Height appears reasonably well correlated with conventional economic indicators (with some exceptions, like the North-East).

- The mass emigration of the period 1890–1910, which was particularly intense in the Southern provinces, is supposed to have pushed real wages up. There is no evidence that it pushed mean heights up more rapidly than elsewhere, possibly because emigration selected positively on height.

- Our data provide no basis for distinguishing between the role played by these economic forces and by factors such as improved systems of sewerage and clean water provision.
6 Distribution of heights: inequality

Higher moments at age 20

- We hope to learn something about inequality from the higher moments of the distribution of heights. Greater dispersion is almost certainly associated with greater inequality in living conditions. Greater skewness can also result from greater inequality.

- The national distribution of heights, in addition to shifting to the right as the mean grows, changes shape and more closely approximates a normal distribution over time (Figure 14).

- This is seen clearly in the trends over time in (adjusted at age 20) standard deviation, skewness (approaching 0) and kurtosis (approaching 3) (Figures 15–17).

- Was inequality really declining so decisively in Italy?
Figure 14: Density of the adjusted (at age 20) national distribution of heights in various years. The vertical lines correspond to the mean.
Figure 15: Standard deviation of the raw and the adjusted (age 20) national distribution of heights.
Figure 16: Skewness of the raw and the adjusted (at age 20) national distribution of heights.
Figure 17: Kurtosis of the raw and the adjusted (at age 20) national distribution of heights.
Importance of 22 as reference age

- Dispersion and negative skewness both *increase* during growth spurts. This is shown in Figure 18, which plots IQR and quantile symmetry, namely the ratio \((Q_{.90} - Q_{.50})/(Q_{.50} - Q_{.10})\), for US boys in year 2000.

- Historically, under conditions of net nutritional stress, the adolescent growth spurt happened later and was more gentle. Growth was prolonged into the early 20s. So our conscripts aged 20 and below would still have been in their growth phase when dispersion and skewness were high. Over time, as living conditions improved, the growth spurt would have occurred earlier, and the dispersion and skewness observed at age 20 would have fallen *even with no change in inequality*.

- Our age adjustment method allows us to standardize at any age. To avoid going too far out of sample, we standardize at age 22. We believe our conscripts would have been at, or close to, final height at this age.

- **Figures 19–20:** The effect of increasing the reference age to 22 is to shift the distribution to the right (increasing the mean), to narrow it (reducing the standard deviation), and to render it more symmetric (reducing negative skewness). These effects are greater in the early years of the sample, when living conditions were poorer on average and the growth spurt was later, and hence estimated age effects were larger.
Figure 18: IQR and quantile symmetry for US boys in year 2000. Source: NCHS.
Figure 19: Density of the adjusted national distribution of heights at ages 18, 20 and 22 in various years.
Figure 20: Cdf of the adjusted national distribution of heights at ages 18, 20 and 22 in various years.
National trends in higher moments at age 22

- **Figure 21**: Age-22 standard deviation of heights falls from 6.59 cm in 1855 to 6.19 in 1886, and the increase that begins by 1890 returns it to 6.37 in 1910.

- **Figure 22**: The age-22 skewness coefficient follows the same rising trend as age-20, but is so much higher as to be **positive** for most of the sample. Positive skewness could be consistent with inequality, in which case rising inequality would be indicated.

- **Figure 23**: Age-22 kurtosis follows a trend similar to age-20, converging to the Gaussian value of 3 from above.
Figure 21: Standard deviation of the adjusted national distribution of heights at various ages.
Figure 22: Skewness of the adjusted national distribution of heights at various ages.
Figure 23: Kurtosis of the adjusted national distribution of heights at various ages.
6 DISTRIBUTION OF HEIGHTS: INEQUALITY

Provincial trends in higher moments at age 22

- **Figure 24**: Variation at the national level is dominated by the within-province component which represents about 90% of total variance. Partly, this reflects random genetic variation within province, which is much greater than systematic economic variation between them.

- The decrease in the national variance of heights till 1885 is almost entirely driven by the decrease in average within-province variance.

- Between-province variation decreases till about 1885 (consistent with our earlier findings about convergence), and then starts rising. The increase in the variance of heights after 1890 is mainly due to the between-province component, although the within-province component also plays some role.

- Mean provincial standard deviation falls steadily between 1855 and 1885 (from 6.31 cm to 5.91 cm), it remains stable for about 15 years, and then rises slightly.

- Adjusting at age 22, there is little evidence of convergence (either $\beta$ or $\sigma$) of provincial standard deviations.

- Age-22 skewness at the provincial level follows the national trend. Mean provincial skewness goes from -0.014 in 1855 to 0.107 in 1877, and then back to 0.054 in 1910.

- Adjusting at age 22, there is little evidence of convergence of provincial skewness.

- Age-22 provincial kurtosis follows the national trend, showing a slightly more steady decrease toward 3.

- Adjusting at age 22, there is evidence of $\beta$-convergence (especially in the period 1885–1910) of provincial kurtosis but little evidence of $\sigma$-convergence.
Figure 24: Decomposition of the variance of the national distribution of adjusted (at age 22) heights into within- and between-province components.
Remarks

• Making inferences about inequality from the shape of the height distribution is difficult for two reasons.

1. The predicted effects of inequality on the distribution of final heights are not clear: (i) the effect on dispersion is pretty clear in direction, but not in magnitude, (ii) the effect on skewness could be in either direction, with the magnitude again uncertain, (iii) the effect on kurtosis is also not clear. We are not aware of good work on these issues.

2. A proper understanding of human biology – controlling for age is vital. The adolescent growth spurt induces an increase in variance and negative skewness during the teens. One would not want to confound inequality with puberty. We are not aware of any work at all on age-adjusting distributions.

• Once we standardize to age 22, our evidence offers a little (but only a little) evidence of a decrease in inequality of living conditions in Italy during the period 1855–1885 and no evidence for the later period 1885–1910.

1. Within-province dispersion of heights diminishes slowly but steadily up until 1885, then stagnates. The slight increase in the national standard deviation of heights in the final decades of the sample is driven by between-province divergence of means.

2. The trend in national and within-province skewness from slightly negative to slightly positive is not very informative. The convergence of national and within-province kurtosis towards the Gaussian benchmark is more evident but is hard to interpret.

• Decreasing inequality coincided with the period of unification of the national market and free trade. Stagnating (or slightly increasing) inequality coincided with the most rapid period of industrialization, and with mass emigration. Is this causal?

• As with trends in mean height, this evidence alone cannot tell us what role may have been played by other developments, such as public health measures.

• To study the effects of early industrialization and emigration on inequality would require a detailed analysis at provincial (or even circondario) level.
7 Next steps

- Further data checking/cleaning.

- More flexible (parametric or semi-parametric) age-adjustment procedures.

- Incorporating the data for the cohorts 1839–54. Issues:
  - Only a few (7), not equally spaced height classes.
  - Incomplete country coverage.

- Conditioning on additional provincial information. Need to collect province-specific information on:
  - Natality and mortality rates.
  - Emigration rates.
  - Weather conditions, crops, etc.
  - Schooling levels.
  - Occupational structure.