MPIDR-NES Training Programme

Moscow, New Economic School, 14th January - 1st February 2013

Population and Health

Лекция 12. Измерение межиндивидуального неравенства. Разброс по продолжительности жизни в таблице смертности

Lecture 12. Measurement of inter-individual inequality. Length-of-life disparity in the life table

MAX PLANCK INSTITUTE FOR DEMOGRAPHIC RESEARCH
MAX-PLANCK-INSTITUT FÜR DEMOGRAFISCHE FORSCHUNG

РЭШ
Российская экономическая школа

Outline of the lecture.

- ❖ An econometric introduction: the Lorenz curve.
- ❖ Desirable properties of inequality measures. Lorenz class measures.
- ❖ Life table and measures of inter-individual inequality in length of life: Gini/AID, IQR, STD, e-dagger.
- ❖ Inequalities in length of life: trends and cross-country comparisons.

Econometric introduction.
Shares of income and shares of population.

Suppose that there is a population of *n* individuals ordered in a non-descending order of their incomes: $y_1 \le y_2 \le y_3 \le .... \le y_n$. There is no inequality if for all individual incomes are the same. The inequality is at its maximum if one individual has the whole population's income, while all others have zero income.

Let μ be the mean income. Let the cumulative shares of population and income be:

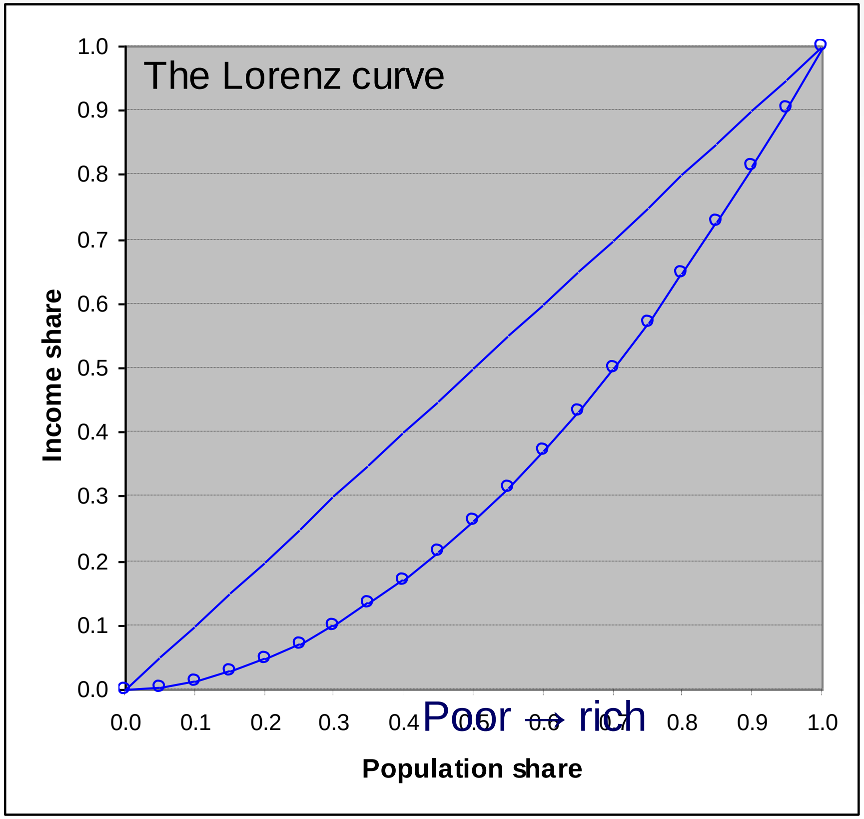
$$F_i = \frac{i}{n} \quad \text{and} \quad \Phi_i = \frac{1}{n\mu}\sum_{k=1}^{i} y_k$$

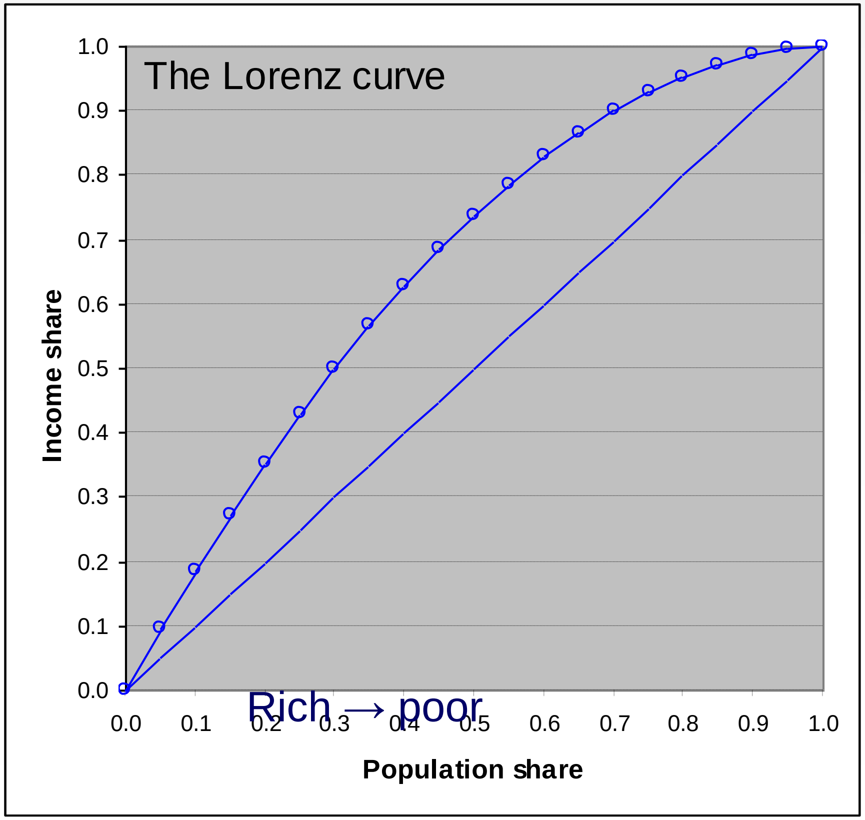
Then the closer the shares *F* and *FF* (*FF* denotes the Greek Φ), the lower the income inequality across the population. Situation of the perfect equality is observed then and only then *F* and *FF* are equal for every individual. That is to say that 10% of people have 10% of the total income, 50% of people have 50% of the total income, etc.

The inequality means that (for example) 10% of the richest people have 30% of the total nation's income while 10% of the poorest people have 5% of the total income.

Econometric introduction. Lorenz curve.

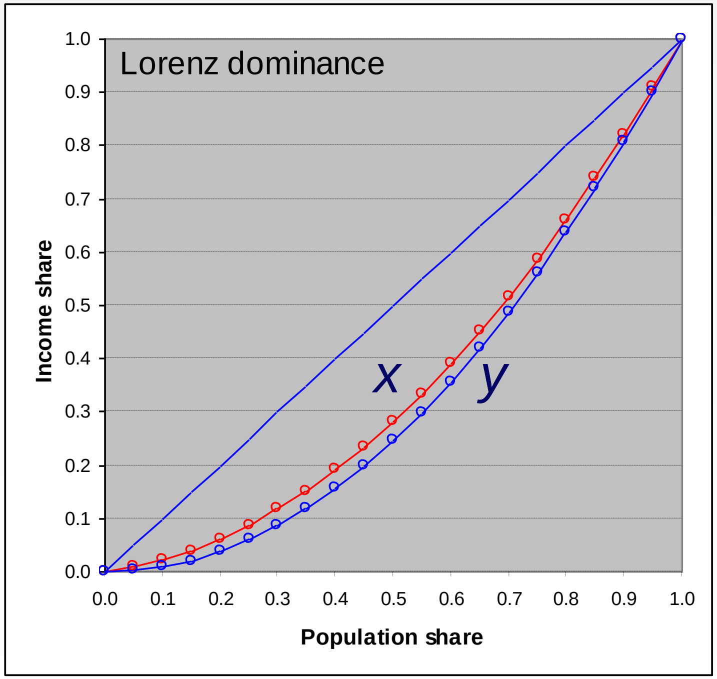
The Lorenz curve shows the whole distribution of income across population by displaying *FF* as a function of *F*.
Divergence between the Lorenz curve and a line of perfect equity (the diagonal) reflects a degree of income inequality. If the population shares on the horizontal axis are based on the ascending individual incomes, the L-curve lies below the diagonal. If the individual incomes are ordered in the descending order, the L-curve lies above the diagonal.





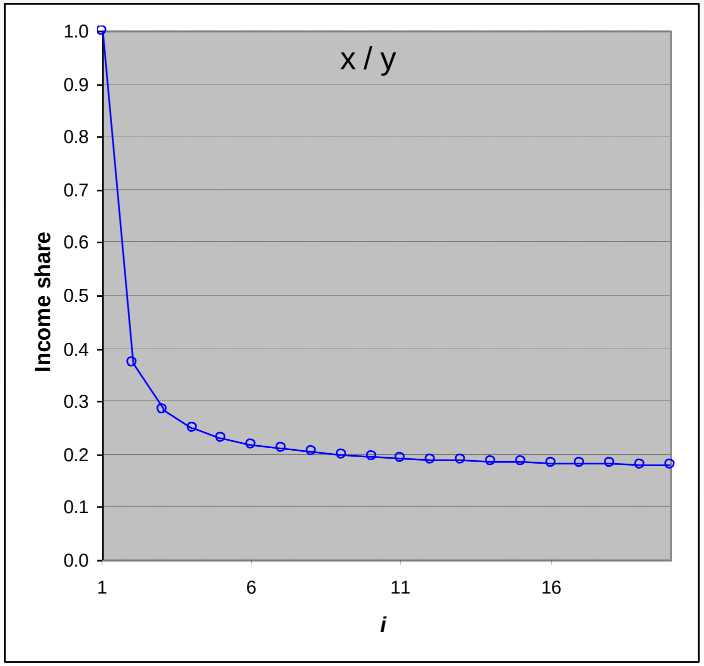
Lorenz-curves (Lorenz-curve-definitions-Gini1.xls)

Lorenz dominance.



Condition of the Lorenz dominance.

Suppose there are two ordered income distributions X ($0 \le x_1 \le x_2 \le x_3 \le .... \le x_n$) and Y ($0 \le y_1 \le y_2 \le y_3 \le .... \le y_n$) with the same numbers of individuals in them. X dominates Y if $x_1/y_1 \ge x_2/y_2 \ge .... \ge x_n/y_n$.



Lorenz-dominance (Lorenz-curve-definitions-Gini1.xls)

Econometric introduction. Lorenz-class measures.

Visual inspection of Lorenz curves helps to compare distributions with respect to amounts of inequality in them only if one distribution dominates another one. Correct ranking of distributions according to their amounts of inequality can be made by means of aggregated *Lorenz-class measures*.

Lorenz-class measure of inequality must satisfy the following conditions:

(1) mean (or scale) independence. It remains invariant if everyone's income is changed by the same amount;
(2) population-size independence. It remains invariant if the number of people at each level of income changes by the same proportion;
(3) Pigou-Dalton condition. It decreases after any transfer from a richer to a poorer person (or vice versa) that does not reverse their relative ranks.

Conditions (1) and (2) suggest that inequality can be measured without data on absolute income and population size.

The mean independence (1) is characteristic of *relative* measures of inequality that express the amount of diversity in income *relative* to its mean level. Almost all existing measures of diversity satisfy (2).
Condition (3) is a central one. It guarantees sensitivity of the inequality measure to any redistribution of income that influences shape of the Lorenz curve.

Econometric introduction.
Two most conventional dispersion measures once again.

In mathematical statistics the most common dispersion measures are the stadard deviation and the range of variation:

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \mu)^2} \qquad Range = \max_i(y_i) - \min_i(y_i)$$

Both indexes measure the absolute inequality and are mean-sensitive. Transition to relative measures can be achieved by rescaling:

$$CoefVar = \frac{\sigma}{\mu} \qquad RelRange = Range / \mu$$

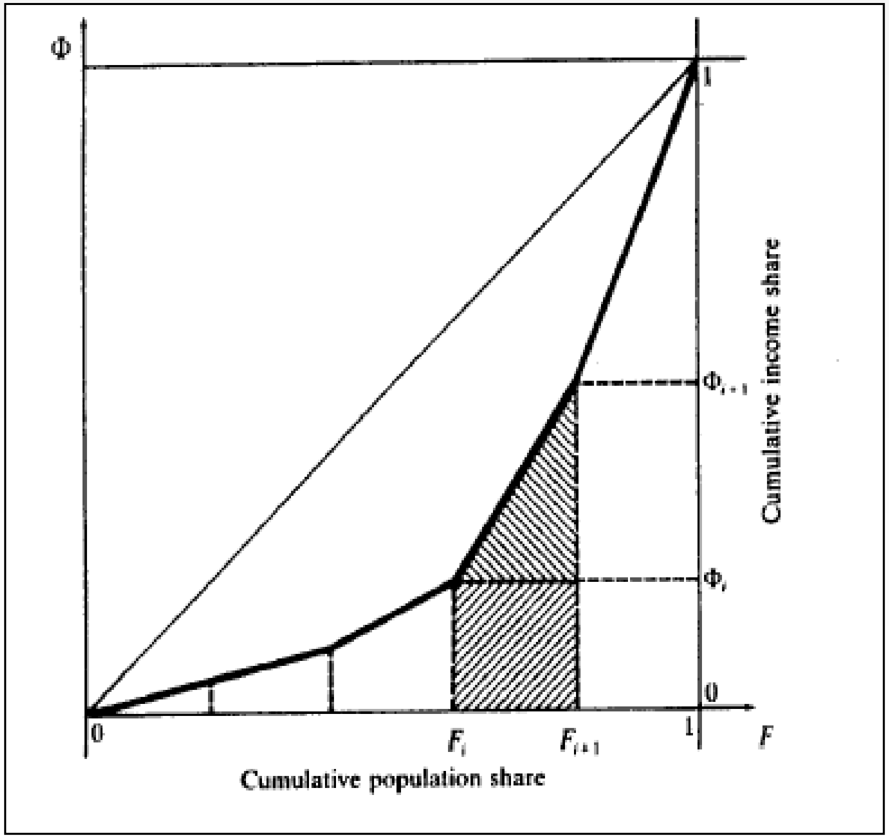
However, the range does not satisfy the P-D condition since it is insensitive to any redistribution that does not change the max and the min incomes. The same is true for the percentile-type indexes such as inter-quartile range: $IQR = Q_{25} - Q_{75}$. They are insensitive to any redistributions that do not change the quartiles. For example it would not change if the share of those getting the lowest income decreases from 10% to 8% and the share of those getting the second lowest income increases from 10% to 12%.

You can learn about other inequality measures such as Robin Hood index, VarLog (variance of logarithms), Theil‘s first and second measures related to the entropy and the information theory) from Anand (1983).

Econometric introduction.
Gini coefficient as a divergence from the diagonal.

There are several equivalent definitions of Gini coefficient:

1. Geometric definition. *G* is an area between the Lorenz curve and the diagonal relative to the whole area below (above) the diagonal.



Source: Anand, 1983

$$G = \frac{1}{1/2}\left[\frac{1}{2} - \frac{1}{2}\sum_{i=0}^{n-1}(F_{i+1} - F_i)(\Phi_{i+1} + \Phi_i)\right] =$$

$$= 1 - \sum_{i=0}^{n-1}(F_{i+1} - F_i)(\Phi_{i+1} + \Phi_i)$$

Econometric introduction. *G* as a mean inter-individual difference and as a covariance of income and its rank.

There are several equivalent definitions of Gini coefficient:

2. Kendall and Stuart (1963) definition. *G* is an average difference between individual incomes across population relative to mean income.

$$G = \frac{1}{2n^2\mu}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|y_i - y_j\right|$$

This definition suggests that $G_{abs} = G \cdot \mu$ is the average inter-individual difference (AID) in income across all pairs of individuals.

3. Definition via covariance. *G* is a covariance of individuals' rank and their incomes relative to the total population income. This definition permits to obtain *G* from regression.

$$G = \frac{2}{n\mu}\operatorname{cov}(i, y_i),$$

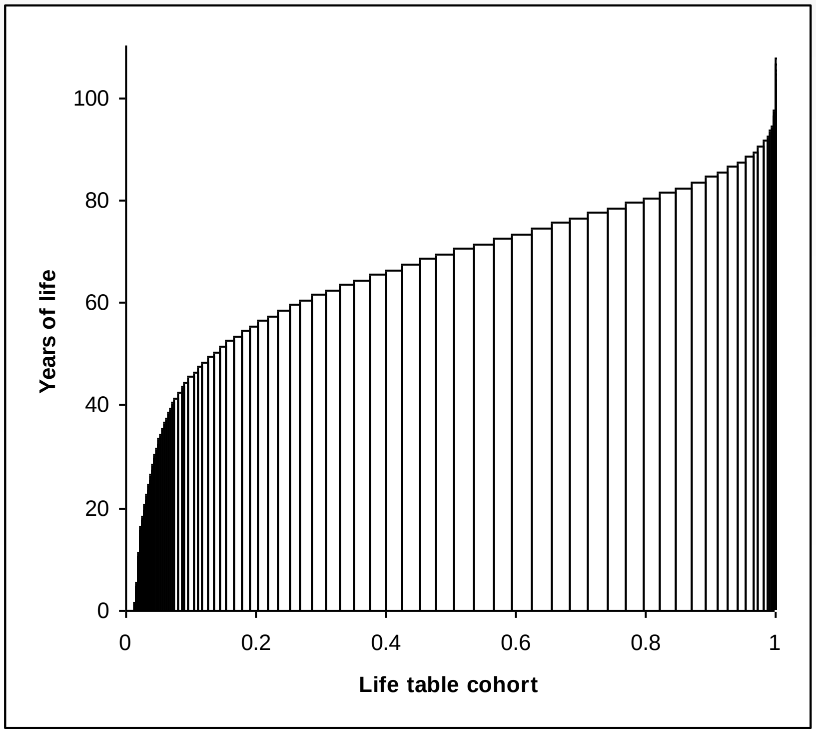
$$\operatorname{cov}(i, y_i) = \frac{1}{n}\sum_{i=1}^{n}(i-\bar{i})(y_i - \bar{y}) = \frac{1}{n}(\sum_{i=1}^{n} i y_i) - \bar{i}\bar{y}$$

Gini in Excel examples (Lorenz-curve-definitions-Gini1.xls)

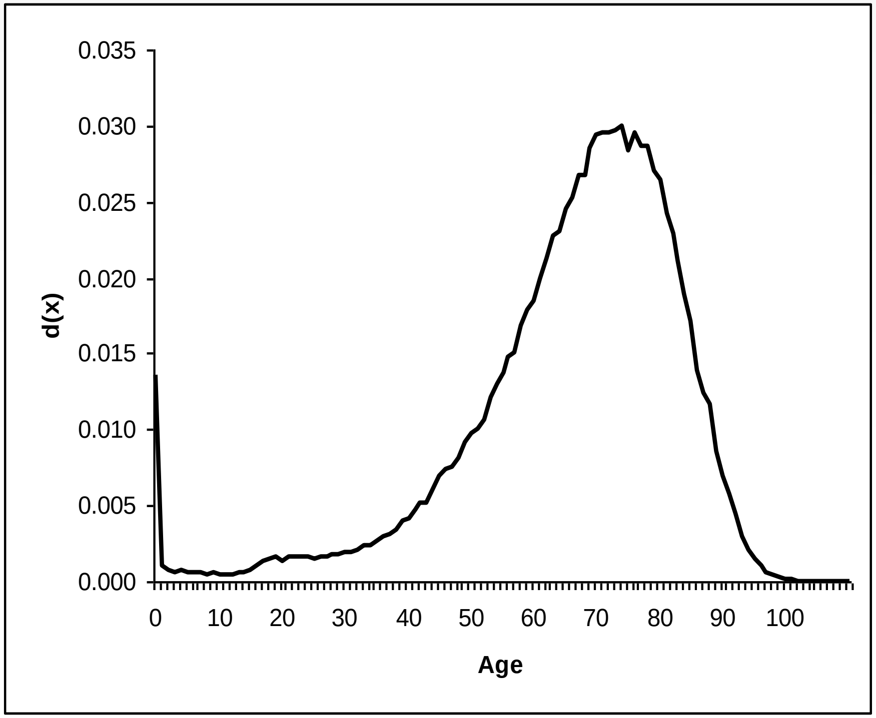
Inequality in the LT. Introduction -1.

Let us consider people's lengths of life (or ages at death) to be people's "incomes". All people will eventually die. But some people die at young age and are "poor" in terms of length of life. Other people die at advanced ages and are "rich" in terms of length of life. It is possible to learn how rich or poor was an individual only after death.

The life table can be presented as a distribution of the life table cohort by length of life.



=



Unusual distribution of the LT population by length of life

Familiar distribution of the LT deaths by age

Inequality in the LT. The Lorenz curve -2.

The life table is also an inequality arena.

The Lorenz curve can be constructed from the life table distribution by age at death. The *F(x)* and *FF(x)* functions can be defined in terms of the standard life table functions as

$$F(x) = 1 - l(x)/l(0) \qquad \Phi(x) = \frac{1}{e(0)l(0)} \int_0^x t d(t) dt$$

In discrete terms they can be expressed as

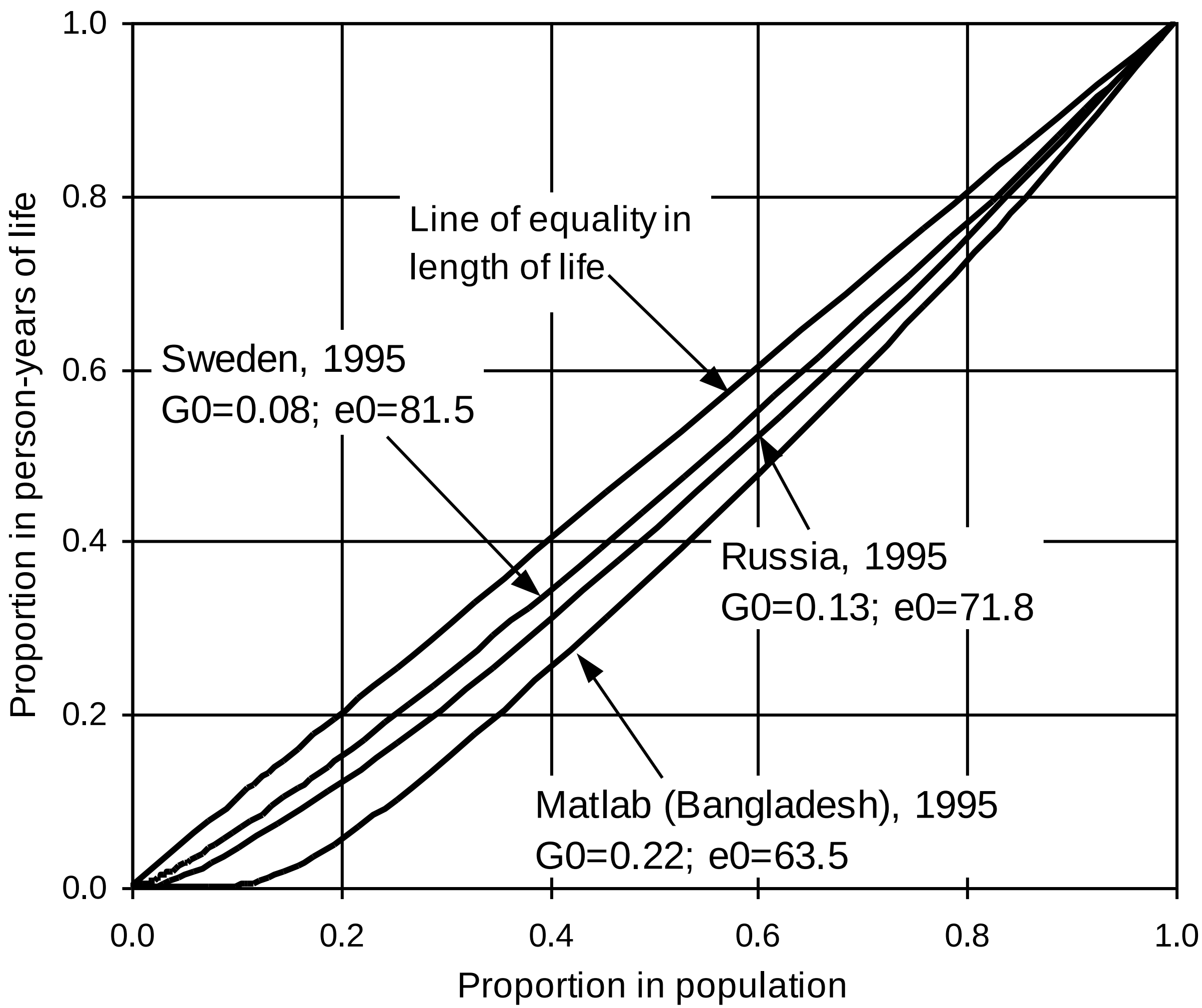
$$F_x = \frac{\sum_{t=0}^{x-1} d_t}{\sum_{t=0}^{\omega-1} d_t} = 1 - \frac{l_x}{l_0}, \quad x = 1,2,...,\omega; \qquad \Phi_x = \frac{\sum_{t=0}^{x-1} d_t \cdot \bar{t}}{\sum_{t=0}^{\omega-1} d_t \cdot \bar{t}} = \frac{T_0 - (T_x + x l_x)}{T_0}, \ x = 1,2,..,\omega - 1;$$

$$F_0 = 0 \qquad \Phi_0 = 0, \ \Phi_\omega = 1$$

Source: Shkolnikov, Andreev, Begun, 2003.

Inequality in the LT. The Lorenz curve -3.

Lorenz dominance: Sweden > Russia > Bangladesh.



Lorenz curves for three female populations with different levels and age distributions of mortality.

Source: Shkolnikov, Andreev, Begun, 2003.

Gini coefficient and AID -1.

According to the geometric definition, the Gini coefficient for ages above age *x* can be computed as

$$G_x = 1 - \sum_{y=x}^{\omega-1} (F_{y+1} - F_y)(\Phi_{y+1} + \Phi_y), \;\; x \le \omega - 1$$

According to the Kendall-Stuart definition, the Gini coefficient for the range of ages above age *x* can be computed as

$$G_x = \frac{1}{2(l_x)^2 e_x} \sum_{\substack{i=1, \\ x_i \ge x, x_j \ge x}}^{l_0} \sum_{\substack{j=1, \\ x_i \ge x, x_j \ge x}}^{l_0} \left| x_i - x_j \right| = \frac{1}{2(l_x)^2 e_x} \sum_{t=x}^{\omega} \sum_{\tau=x}^{\omega} d_t \cdot d_\tau \cdot \left| \bar{t} - \bar{\tau} \right|,$$

$$\text{for } (t, t+n] : \bar{t} = t + a_t n,$$

$$\text{for } t = \omega : \;\; \bar{t} = t + e_\omega$$

Absolute measure corresponding to G is named average inter-individual difference in age at death

$$AID_x = G_x \cdot e_x$$

Gini coefficient and AID -2.

Hanada (1983) developed a useful formula for Gini coefficient

$$G_x = 1 - \frac{1}{l(x)e(x)} \int_x^{\infty} [l(t)]^2 dt$$

For a complete LT with ages running up to around 100, numerical integration in the Hanada's formula can be done similarly to the numerical integration of *l(x)* in calculations of life expectancy (Shkolnikov et al., 2001)

$$\int_x^{\infty} [l(t)]dt \cong \sum_{y=x}^{\omega-1} L_y = \sum_{y=x}^{\omega-1} [l_{y+1} + a_y (l_y - l_{y+1})]$$

$$\int_x^{\infty} [l(t)]^2 dt \cong \sum_{y=x}^{\omega-1} [(l_{y+1})^2 + \hat{a}_y (l_y - l_{y+1})^2]$$

For a complete LT it is possible to assume $\hat{a}_y = a_y$

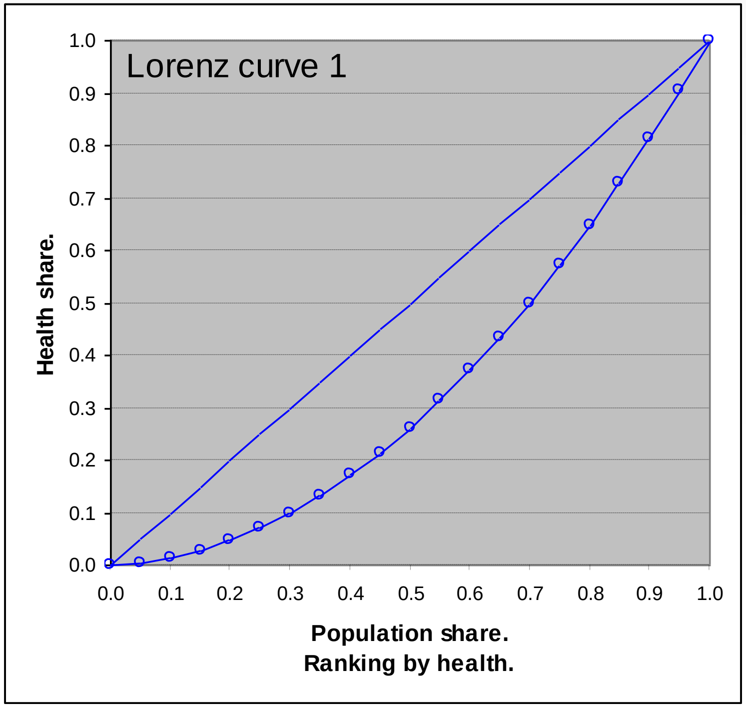
Inequality-measures-LT-complete (Inequality-measures-LT-complete1.xls)

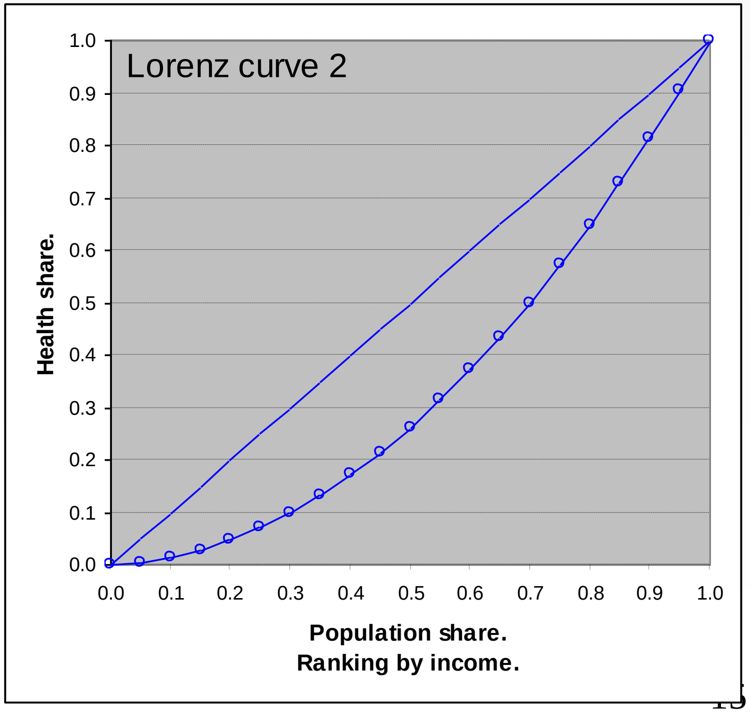
Inter-individual socioeconomic inequalities in health. Concentration index.

Lorenz curve 1: Based on distribution of health (length of life) across individuals. On the horizontal axis, individuals are ranked by health (length of life).

Lorenz curve 2: Based on distribution of health (length of life) across individuals‘ income (wealth). On the horizontal axis, individuals are ranked by income (wealth).

C = twice the area between the L-curve and the diagonal. + when L-curve is below the diagonal and – when the L-curve is above the diagonal. Varies from -1 to +1.





IQR and STD.

Other measures of inequality have been also used for assessing the LT disparities in length of life.

Goodwin and Vaupel (1985) proposed half-statistic as a measure of inequality. The Half-have statistic shows what proportion of the total years of life half of the LT cohort has. Let *x* be an age such that

$$l_x = 0.5 \quad \text{then} \quad HaveHalf_x = 1 - \frac{T_x}{T_0}$$

Wilmoth and Horiuchi (1999) proposed an inter-quartile range. *IQR* is a difference between age to which 25% of people survive and the age to which 75% of people survive. Let

$$Q_{25} = x \text{ such that } l_x = 0.25 \quad \text{and} \quad Q_{75} = x \text{ such that } l_x = 0.75$$

then $IQR = Q_{25} - Q_{75}$

Edwards and Tuljapurkar (2005) proposed to use standard deviation to measure the LT inequality

$$STD_x = \left\{ \frac{1}{l_x} \sum_{y=x}^{\omega} [d_y (\bar{y} - (x + e_x))^2] \right\}^{1/2} = \left\{ \frac{1}{l_x} \left(\sum_{y=x}^{\omega} d_y \bar{y}^2 \right) - (x + e_x)^2 \right\}^{1/2}$$

Inequality-measures-LT-complete (Inequality-measures-LT-complete1.xls)

e-dagger.

Vaupel and Canudas Romo (2003) proposed a measure called „e-dagger“. It counts person-years lost due to premature death and can be also considered as a measure of public health losses.

At every age, life table deaths are multiplied by the life expectancy lost by those dying at this age. The following two formulae can be used for calculations. The second formula is more precise than the first one.

$$e_x^{\dagger} = \frac{1}{l_x}\int_x^{\infty} d(t)e(t)dt = \frac{1}{l_x}\sum_{y=x}^{\omega-1}\left[d_y \frac{e_y + e_{y+1}}{2}\right] + \frac{1}{2l_\omega} d_\omega e_\omega$$

$$e_x^{\dagger} = \frac{1}{l_x}\int_x^{\infty} d(t)e(t)dt = \frac{1}{l_x}\sum_{y=x}^{\omega-1}\left[d_y (e_{y+1} + 1 - a_y)\right] + \frac{1}{2l_\omega} d_\omega e_\omega$$

The last term equals $\frac{1}{2}e_\omega$ since $d_\omega = l_\omega$.

If the last group is 100+ or higher, the formulae work well. The formulae are becoming somewhat problematic if the last age group is 85+. In this case one can first extrapolate the life table up to age 100+ or 105+.

Aversion to inequality and the Atkinson indices - 1.

Atkinson indices include a special parameter *epsilon* expressing degree of aversion to inequality.

$$A(\varepsilon) = \begin{cases} \left[\sum_x f_x (x + a_x)^{1/(1-\varepsilon)} \right]^{1/(1-\varepsilon)} & \text{for } \varepsilon \geq 0, \varepsilon \neq 1 \\ \exp\left(\sum_x f_x \ln(x + a_x) \right) & \text{for } \varepsilon = 1 \end{cases}$$

, where $f_x = d_x / l_0$

A(epsilon) **is equal to:**

-the arithmetic mean (life expectancy) if epsilion=0;

- the geometric mean when epsilon=1;

- the harmonic mean when epsilon=2.

Source: Anand et al., 2001

Aversion to inequality and the Atkinson indices - 2.

Inequality measures with different levels of aversion to inequality can differently evaluate temporal changes and inter-population differences in the amount of inequality.

Comparison of the length-of-life distributions between 1990 and 1995, Russian men.

Basic measures of central tendency and dispersion		
Arithmetic mean (life expectancy at birth)	63.80	58.43
Geometric mean	54.14	49.01
Harmonic mean	4.26	4.13
Median	67.02	60.52
Standard deviation	19.39	19.73
Coefficient of variation (= σ / *Xe* [0])	0.304	0.338
Relative mean deviation	0.232	0.266
Inequality measures		
Gini coefficient	0.163	0.187
Atkinson index		
Inequality aversion parameter 1.0	0.151	0.161
Inequality aversion parameter 1.5	0.581	0.579
Inequality aversion parameter 2.0	0.933	0.929

Note: For the definition of these measures see Appendix A.
Source: Anand 1998.

Explanation: In 1995 mortality was higher than that in 1990 with exception to only three ages: 0, 1, and 2 years.

Source: Anand et al., 2001

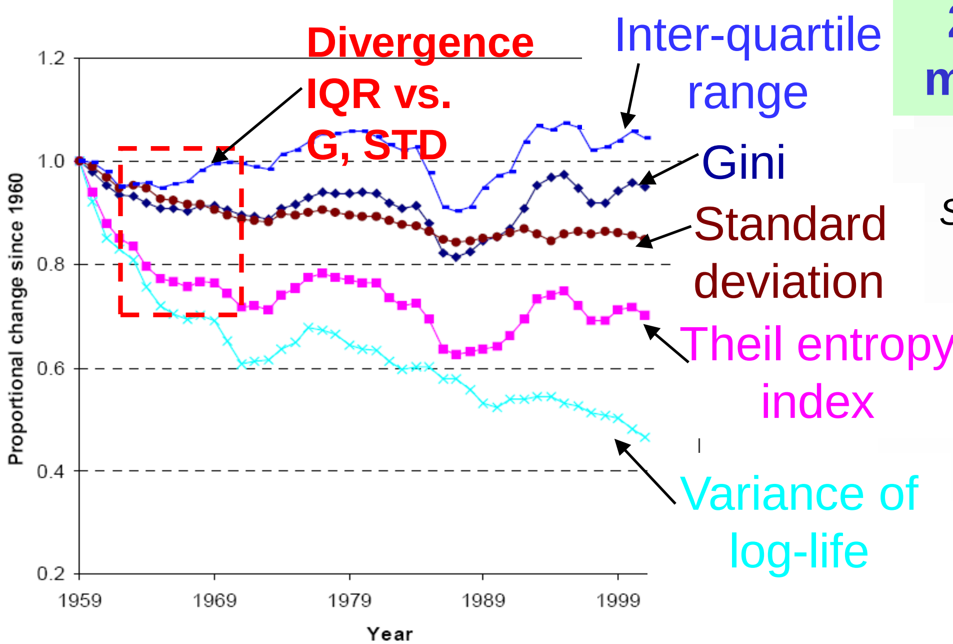
Trends and comparisons.Trends in various inequality measures: a high correlation, but also some differences.

1. Correlation coefficients for 10 measures of rectangularity and variability

	FR	MR	FD	SC	QP	PI	SD	IQR	G	H
Fixed Rectangle	1.000									
Moving Rectangle	0.997	1.000								
Fastest Decline	0.975	0.974	1.000							
Sharpest Corner	0.837	0.828	0.922	1.000						
Quickest Plateau	0.923	0.922	0.972	0.927	1.000					
Prolate Index	0.967	0.967	0.990	0.912	0.963	1.000				
Standard Deviation	–0.926	–0.918	–0.923	–0.835	–0.868	–0.906	1.000			
Interquartile Range	–0.956	–0.959	–0.926	–0.785	–0.856	–0.919	0.936	1.000		
Gini Coefficient	–0.993	–0.997	–0.959	–0.797	–0.895	–0.951	0.918	0.967	1.000	
Keyfitz's *H*	–0.987	–0.994	–0.948	–0.775	–0.886	–0.942	0.883	0.948	0.996	1.000

Calculated according to the data on Sweden, USA, and Japan. Source: Wilmoth & Horiuchi, 1999.

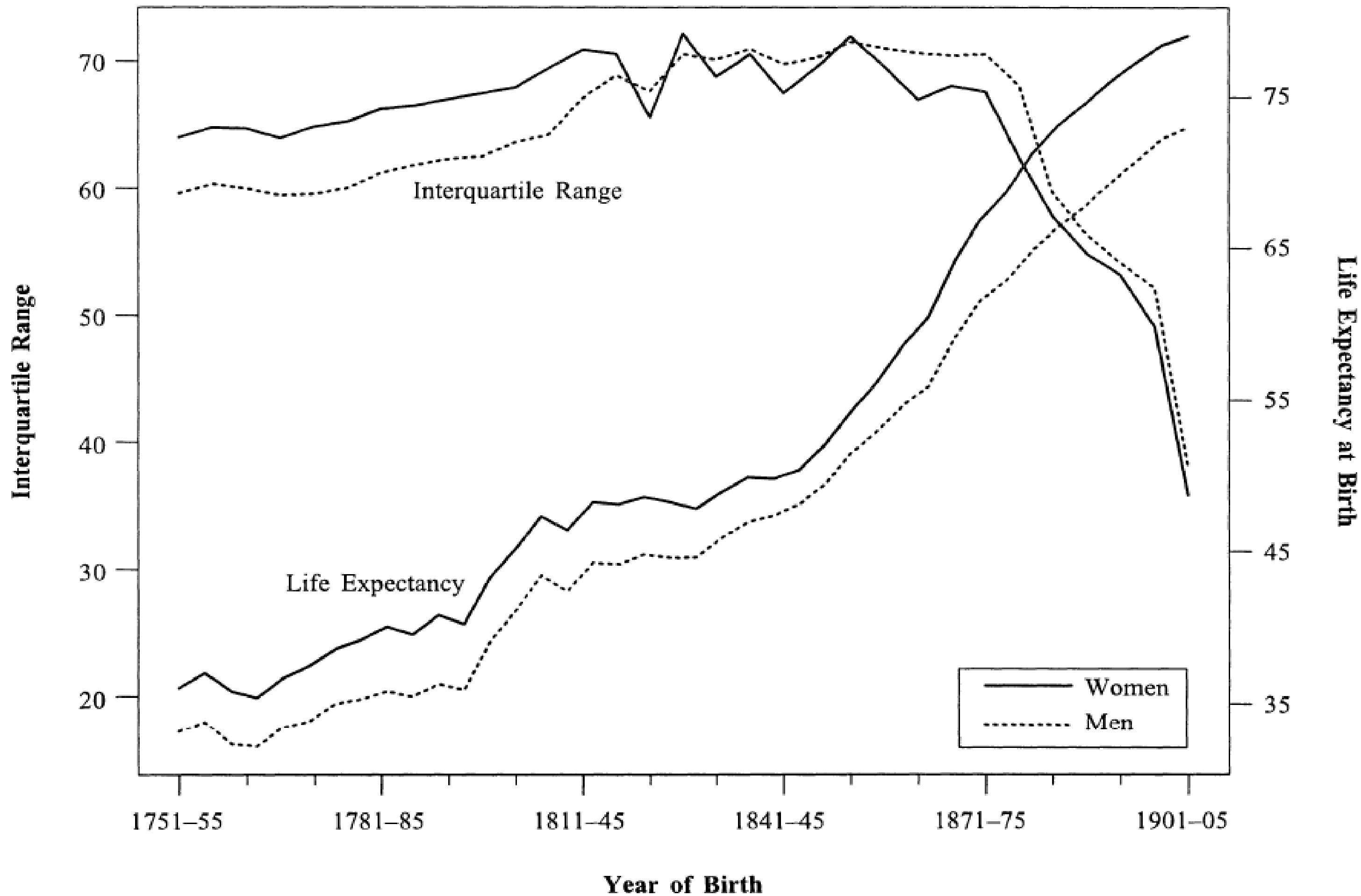
2. Proportional changes in inequality measures of Russian males, 1959-2000



Source: Shkolnikov, Andreev, Begun, 2003.

Trends and comparisons. Long term change: life expectancy increase is associated with reduction of inequality (rectangularization of the survival curve).

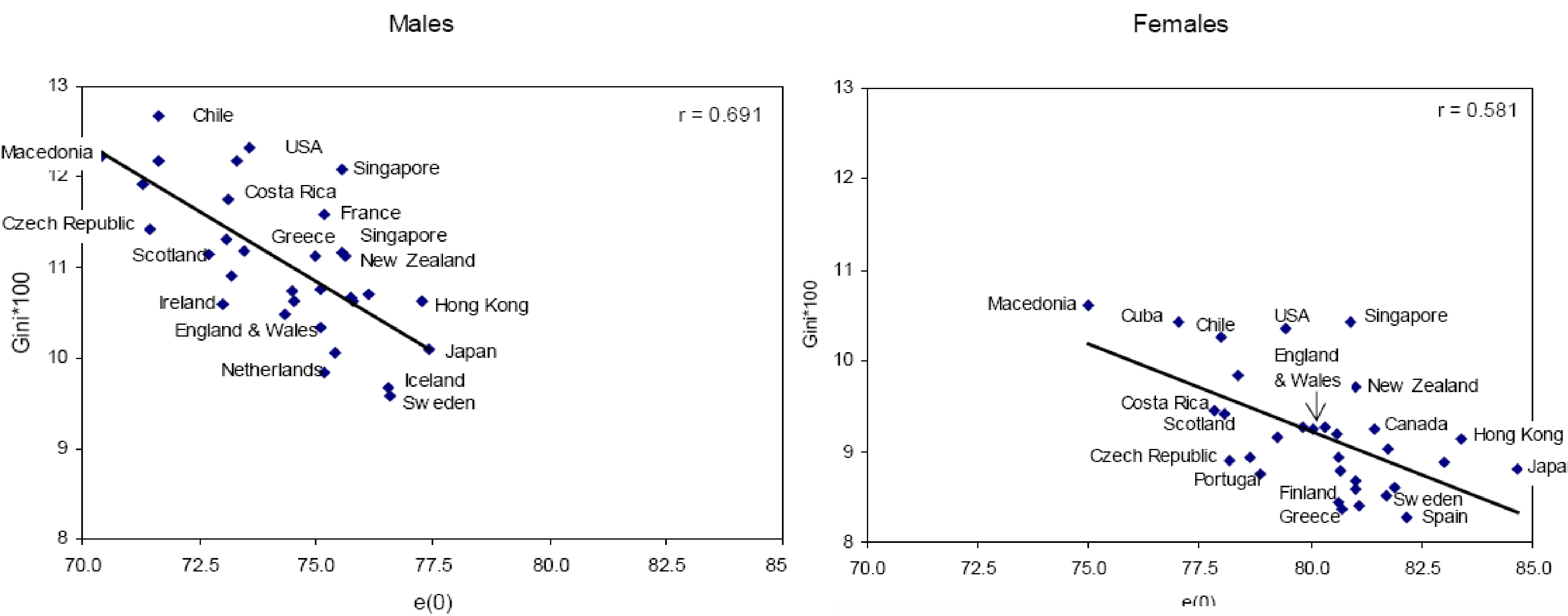
Inter-quartile range of life table ages at death, and life expectancy at birth. Swedish men and women, cohorts born in 1751-1905.



Source: Wilmoth & Horiuchi, 1999.

Trends and comparisons. A high cross-sectional correlation between life expectancy and inequality.

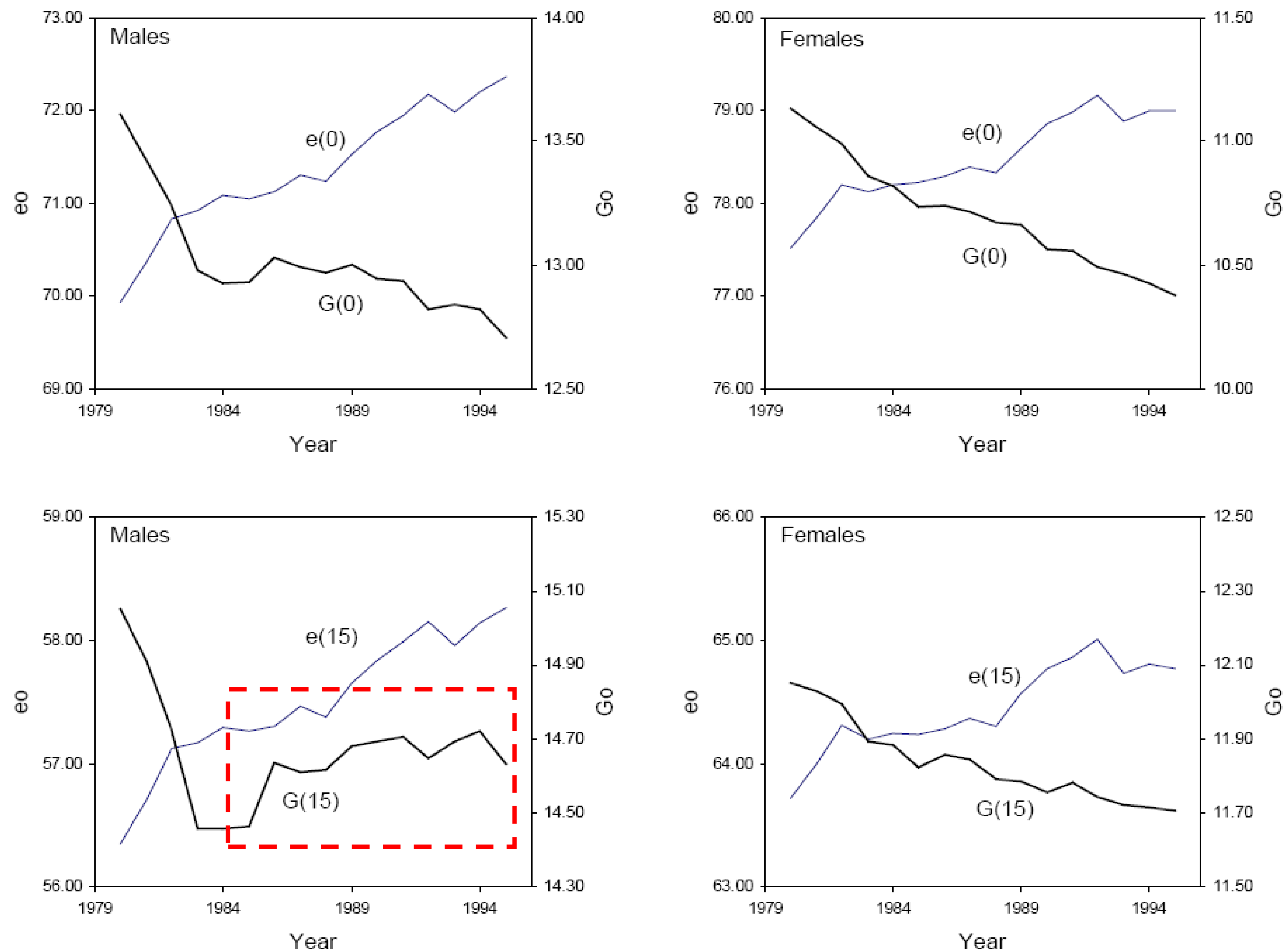
Relationships between life expectancy and Gini coefficient in 1996-99 for men and women among 31 countries with male life expectancies >= 70 years.



Source: Shkolnikov, Andreev, Begun, 2003.

Trends and comparisons. Peculiar changes in inequality and average length of life in the US.

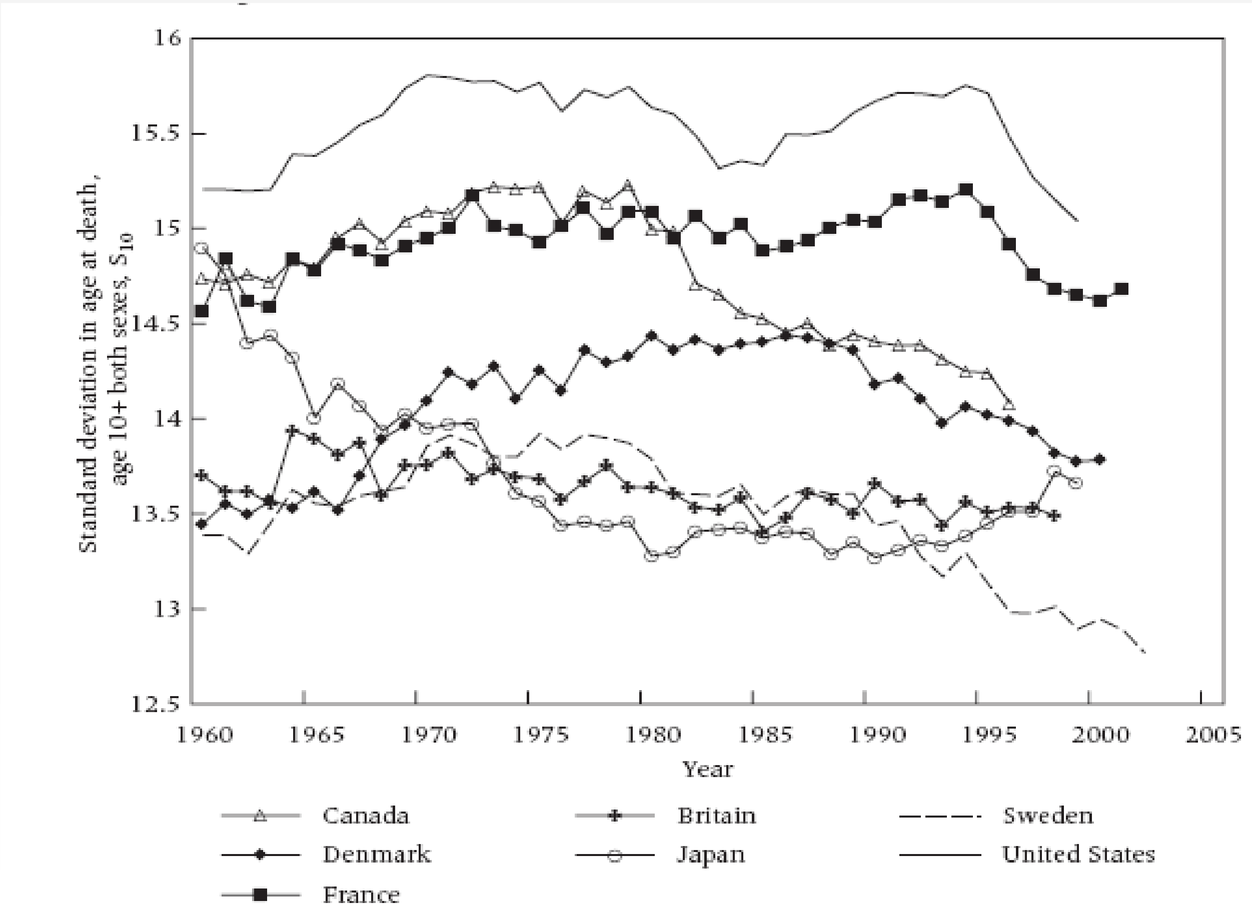
Trends in life expectancy and Gini coefficient (ages 0 and 15) for men and women in the USA in 1980-95.



Source: Shkolnikov, Andreev, Begun, 2003.

Trends and comparisons.
Particularly high inequality in the US.

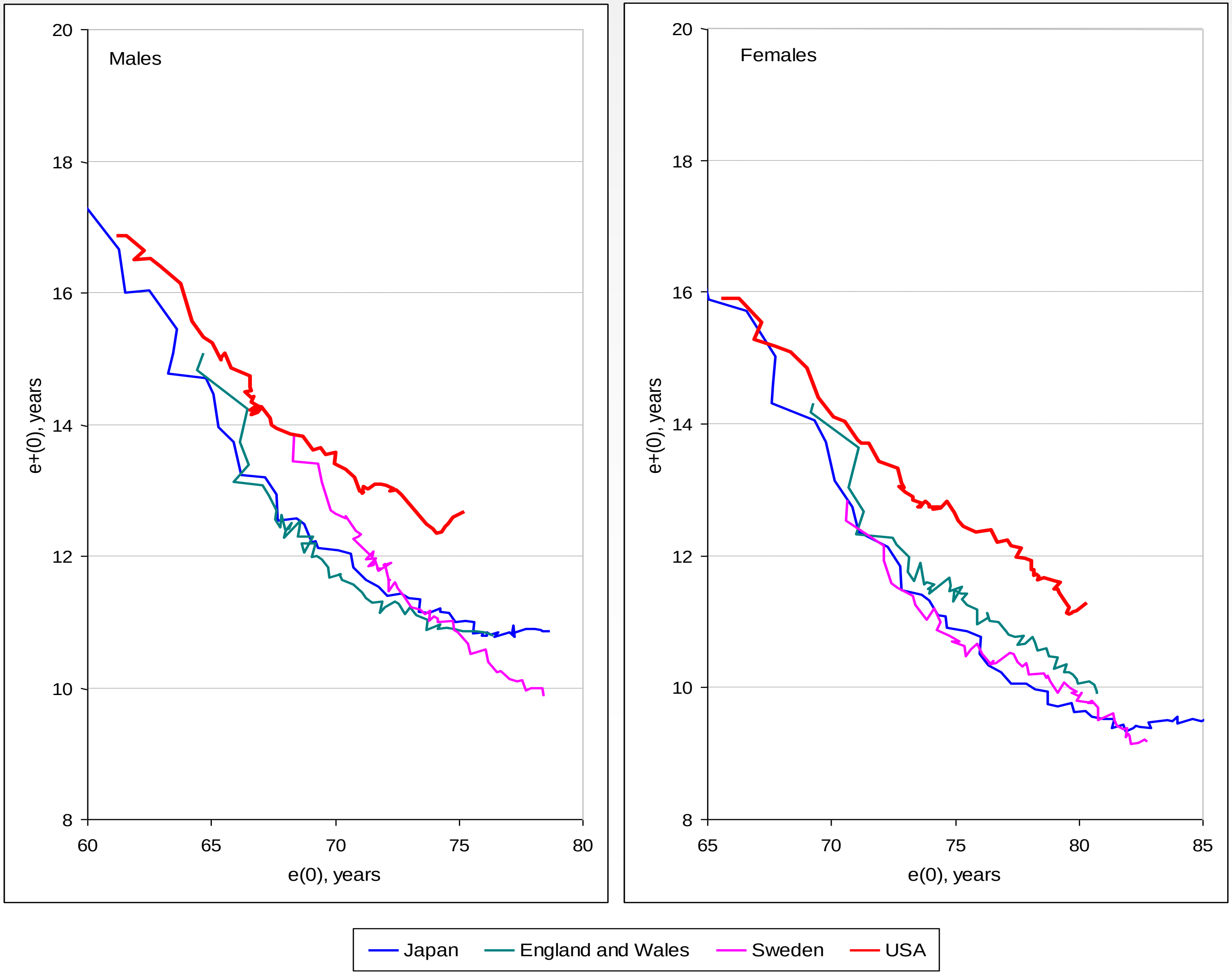
Conditional standard deviations in the age at death (STD_{10})
Seven high-income countries since 1960.



Source: Edwards & Tuljapurkar, 2005.

Trends and comparisons. Particularly high inequality in the US - 2.

The e†0-e0 trajectories in England and Wales, Japan, Sweden, and the US.



References

Anand, S. (1983). *Inequality and poverty in Malaysia. Measurement and decomposition.* New York: Oxford University Press.

Anand, S., Diderichsen, F., Evans, T., Shkolnikov, V., Wirth, M. (2001). Measuring disparities in health: methods and indicators. In: T. Evans, M. Whitehead, F. Diderichsen, A. Bhuiya, M. Wirth (Eds.), *Challenging inequities in health: from ethics to action*. New York: Oxford University Press.

Edwards and Tuljapurkar (2005). Inequality in life spans and a new perspective on mortality convergence across industrialized countries. *Population and Development Review*, Vol. 31 (4), pp. 645–674

Kendall, M., Stuart, A. (1963). *The advanced theory of statistics*. London: Charles Griffen and Company.

Le Grand, J., Rabin, M. (1986). Trends in British health inequality 1931-1983. In: A.J. Culyer, B. Jonsson (Eds.), *Public and Private Health Services*. Oxford: Blackwell.

Shkolnikov, V., Andreev, E., Begun, A. (2003). Gini coefficient as a life table function: computation from discrete data, decomposition of differences and empirical examples. Demographic Research, Vol. 8, pp. 306-357.

Vaupel and Canudas Romo (2003). Decomposing change in life expectancy : a bouquet of formulas in honor of Nathan Keyfitz's 90th birthday. Demography, Vol.40 (2), pp. 201-216.

Wilmoth, J., Horiuchi, Sh. (1999). Rectangularization revisited: variability of age at death within human populations. Demography, Vol.36, No.4, pp.475-495.