University of Innsbruck

# The logic of inequality extraction: An application to Gini and top incomes data

**Andreas Exenberger**

**Working Papers in Economics and Statistics**

2017-09

Research Platform
eeecon
Empirical and Experimental Economics
University of Innsbruck
http://eeecon.uibk.ac.at/

**University of Innsbruck**
**Working Papers in Economics and Statistics**

The series is jointly edited and published by

- Department of Banking and Finance
- Department of Economics
- Department of Public Finance
- Department of Statistics

Contact address of the editor:
Research platform “Empirical and Experimental Economics”
University of Innsbruck
Universitaetsstrasse 15
A-6020 Innsbruck
Austria
Tel: + 43 512 507 7171
Fax: + 43 512 507 2970
E-mail: eeecon@uibk.ac.at

The most recent version of all working papers can be downloaded at
http://eeecon.uibk.ac.at/wopec/

For a list of recent papers see the backpages of this paper.

# The Logic of Inequality Extraction: An Application to Gini and Top Incomes Data

Andreas Exenberger

*Associate Professor of Economic and Social History*
*at the Department of Economic Theory, Policy and History, University of Innsbruck,*
*Universitätsstraße 15, A-6020 Innsbruck, Austria; e-mail: andreas.exenberger@uibk.ac.at*

**Abstract**

As embodied in the concept of "inequality extraction" (Branko Milanovic), it is not possible to increase inequality (especially income inequality) in a society sustainably to levels beyond what is actually socially acceptable (and even less to levels endangering physical subsistence of parts of the population). Consequently, in international and intertemporal comparisons of inequality levels on the country (or the regional) level and for poorer societies in general, adaptations for the levels of inequality, which are actually feasible in physical and social terms, are necessary to arrive at meaningful conclusions. We extend this concept to cover also top-incomes ratios, which allows a broader database for measurement and analysis of cross-country historical inequality. A first look at the data uncovers different patterns and dynamics of inequality extraction when compared to untransformed inequality measures, sometimes to the point of trend reversal, which holds for extraction ratios based on Gini coefficients as well as top income shares. In the discussion, we also outline some directions of further research, especially about explanations of extraction patterns and dynamics, also with an application to global inequality data and in combination with institutional arrangements.

**Keywords:** Gini coefficients, top income shares, inequality extraction, empirical application

**JEL-Codes:** O15, N10, C82, D31, D63

# Introduction

The debate about the economic significance of inequality – and hence of the questions if and what should be done about it – is ages old. Opinions oscillate between the extreme positions that inequality is a main driving force of growth by providing incentives for those worse off to achieve more, and that inequality is a serious obstacle for growth and development by destabilizing social cohesion or market equilibria. Somewhere in the middle, positions about Kuznets-like dynamics can be found, arguing for some kind of a seemingly natural path of societies from low to high and back to low levels of inequality, while developing from poorer into richer economies. Recent empirical studies bring some evidence into this theoretical debate, when providing information that at least excessive inequality is negatively correlated with economic performance.[1] What the different voices in this debate all have in common is that they usually argue along a concept of inequality, which is everything but context-specific. However, especially when confronted with income inequality comparisons ("income" is understood as a proxy for the means to achieve a decent standard of living throughout this paper) between very poor and rather rich societies, in history as well as today, this is likely to be inappropriate. On the contrary, it may be more informative to look at the level of inequality from a perspective that is context-sensitive to understand better, how unequal societies actually are or have been, relative to what they could be or could have been. In a second step, this may provide substantially different answers to the question why these societies are unequal, what kind of problems may be associated with it and what can be done about it.

Branko Milanovic (together with various co-authors) provides a recent contribution to the literature, which is addressing this problem of context-insensitivity.[2] They focus on the amount of actual inequality in relation to an amount which is reasonably achievable in a society,[3] and hence on the level of inequality which is "extracted". This concept of "inequality extraction" rests on the simple insight that extreme income inequality (for example, a Gini coefficient close to or equal to 1) is practically impossible in any real existing society. If in a society one person earns all income and all others none, people will simply die and society cease to exist.[4] Hence, as long as the members of a society (or at least its elites) are interested in their physical survival (which of course is directly related to the actual production of any kind of income and of the continuance of society as such), they have to assure at least some kind of subsistence income for all (monetary or in kind). As a consequence, only the income exceeding this subsistence level (referred here as "surplus income") can reasonably be subject to unequal distribution.

---

[1] OECD (2015) provides a general discussion of this issue with an empirical investigation of OECD countries, focussing on human capital as central transmission channel (see also Cingano 2014 for a summary). Van der Weide and Milanovic (2014) show in-depth quantitative evidence for inequality harming growth of the poor in the United States. OECD (2015, p. 70) concludes on that issue: "lowering inequality by reducing income disparities at the bottom of the income distribution has a greater positive impact on economic performance than if the focus were on reducing top inequality". See also Ostry et al. (2014) for a discussion about the relation between inequality, growth and redistribution (with evidence in favour of the latter), as well as Halter et al. (2014) for a discussion of the timing of effects (with evidence of negative long-run effects of inequality on growth).

[2] See most notably Milanovic (2009, 2013) and Milanovic et al. (2007, 2010).

[3] While the actually most appropriate unit of analysis for the extraction concept is a "society", because of data availability the practical unit of analysis throughout this paper will be the "country". Hence, we use the two terms interchangeably and prefer for methodological reasons "society", while practically referring to countries.

[4] One could think of a situation in which all income in a group may formally be owned by a small elite or even one person (a king, high priest, or the like), but one cannot think of this situation without redistribution – at least in kind – to all members of this group.

In this paper, we apply different variants of this concept to actual inequality data and show, how this additional context changes the picture that we usually draw from inequality patterns in space and time. Further, we extend the concept to top incomes ratios, which has not been done so far. This will open up a much broader historical database for application of the extraction concept and resulting data transformation, which is necessary to come to meaningful comparisons in time and space as well as allowing the description of different and more relevant patterns and fresh explanations for the embodied dynamics. Overall, this paper is a step towards a general historical analysis of extraction patterns and trends. We show that the (negative) correlation between extraction and income levels is much more pronounced (and very robustly so) than between inequality and income levels (which is also mainly driven by a few rich countries). We further show that this logic also holds for top income shares, where changes in the patterns and dynamics are particularly obvious in the case of poorer societies. And we finally discuss some of the possible explanations for extraction, of the channels of transmission, global inequality trends and patterns, and of the consequences of extraction.

In the following section, we present the concept of inequality extraction and some extensions (already provided by Milanovic) and apply it to some recent and historical inequality data (especially Gini coefficients on the country level). Afterwards, we introduce the top incomes extraction ratio and as well apply it to historical inequality data (country time series of top-percentile and top-decile shares). Finally, we will discuss some conclusions from this analysis and directions of further research.

# Inequality extraction

Milanovic (2007) first introduced the concept of inequality extraction. It departs from the simple insight, that if a society produces on the aggregate not more than subsistence income[5] for all, no inequality at all is feasible – and would only be possible transitorily at the expense of the physical non-survival of its poorest members.[6] If it produces surplus income, inequality becomes feasible, but only up to a certain amount directly related to the ratio between mean and subsistence income (in per capita terms).[7] Approximately,[8] the relationship between maximum feasible inequality (*MFI*), measured here and in the following as Gini coefficient, the minimum subsistence income ($s$), and the mean income ($m$), is:

[5] Subsistence income is understood throughout the paper as an income (in money or kind) which allows an individual (or household) to achieve just as much resources as necessary to sustain its physical survival. To address the large differences in price levels across countries, we will work with incomes adjusted for purchasing power (actual methodologies for the provision of income data and the adaptations vary across databases).

[6] This situation is not sustainable, indeed. Cynically speaking: as long as there is no extra income generated by an external shock, in this setting a long-term equilibrium can only be reached at a level with not more than subsistence income for all who survive the adaptation process.

[7] For different levels of income, these amounts of "maximum feasible inequality" add up graphically to an increasing and concave "inequality possibility frontier" (Milanovic et al. 2010, p. 258).

[8] For a bit more maths see Milanovic (2013) and Milanovic et al. (2007). Mean income is used for mathematical reasons (different from median income, overall income can be directly calculated from it). The simplifying approximation is derived with the assumption that there are two income classes in society and that the share of people belonging to the upper one tends towards zero. Thus, it is decreasingly valid for large shares of rich people in a society. However, societies of that kind do hardly exist in the real world, especially not in history, and the mathematical influence of the magnitude of that share is rather small. A larger problem is posed by the assumption that there is no inequality within the two income classes, which is unrealistic in the real world, but this assumption is shown to be unproblematic at least for the upper class (Milanovic et al. 2007, pp. 11-12).

(1a) $$MFI = \left(\frac{m}{s} - 1\right)\left(\frac{s}{m}\right),$$

which, if we express the ratio of mean income ($m$) to subsistence income ($s$) as $\alpha$, simplifies to:

(1b) $$MFI = \frac{\alpha - 1}{\alpha}$$

This formula is used by Milanovic to derive everything that follows (implicitly already in Milanovic et al. 2007, p. 9, explicitly at least in Milanovic 2009, p. 15). *MFI* is hence dependent on $m$ and $s$, and it means numerically that in case that mean income in a society is double the level of subsistence income, the maximum level of feasible inequality is confined to a Gini coefficient of 0.5. Technically, this level is realized only if all persons but one in this society earn subsistence income while all surplus is distributed to this one person. This example also characterizes a situation in which the level of inequality extraction equals 100%, i.e. all surplus income is transformed into inequality. If this is not the case and some (at least two) people in society earn incomes exceeding subsistence level, while all others earn at least their subsistence, extraction is below 100%. Generally, the inequality extraction ratio (*IER*), in relation to maximum feasible inequality (*MFI*) and actual inequality ($G$) is given by:

(2) $$IER = \frac{G}{MFI}$$

In practice, an inequality extraction ratio exceeding 100% cannot be ruled out, at least in the short run. However, it is only possible if a considerable number of people earn less than subsistence income (at least for some time), which would accordingly increase the amount available for extraction. But if we find such ratios, this indicates unsustainable situations, high levels of extreme poverty, and ongoing physical destruction of parts of society – usually of marginalized people, as for example in extreme slave-holder societies or during genocidal episodes as well as in countries with considerable shares of undernourished people.

While the notion of "extraction" obviously bears some analytical meaning as well, it is at first primarily to be understood as a descriptive statement characterizing a distribution (of income). However, we will refer to extraction as something that an economic or political "elite", a small minority of all members of society, does to the (rest of) "population". This is done for reasons of simplification and does not necessarily imply a process by which this is actually done or any kind of intentionality, but is a descriptive statement about a quantitative pattern of distribution. It will nevertheless be necessary to clarify terminology and possible actual mechanisms at a later stage, because it is indispensable if we want to understand actual channels of transmission of extraction.

But why is it worth while to recalculate inequality levels according to these concepts? Mainly because it makes an important and substantial difference, not only in perception, if an absolute level of inequality as represented by a Gini coefficient of for example 0.5 refers to a situation in which only slightly more than 50% of feasible inequality is actually extracted by the elite, or a situation in which this level approximates 100%. While the former would be the case in societies with high levels of mean income, like Austria, where even poor people have access to considerably more than mere subsistence, the latter represents the case of a poorer

society, where poor people are much closer to endangered physical survival. Hence, a high level of extraction is also a hint for the incidence of widespread poverty. Further, dependent on differences in mean income, poor societies with comparably low levels of conservative inequality measures may actually turn out to be more unequal from the extraction perspective than rich societies with comparably high levels of these measures.[9] Because of this change of perspective, the questions asked and the answers received about inequality may change considerably, especially about the actual patterns, about potentially embodied injustice and about their reasons and justifications.

The idea of an absolute level of subsistence equal in all societies all over the world and at all times is attractive for reasons of mathematical simplicity, but economically not very reasonable. Even if the best of all possible adaptations for differences in purchasing power parity are applied, there is no question that anybody can survive in a developed society on anything close to a one-dollar-a-day basis. The respective goods (food and clothing, but even more housing and services) are simply not available in practice, definitely not on whatever kind of market. Further, these kinds of adaptations cannot at all account for non-material aspects of poverty, for relative status, no matter how low, in a more affluent society. In richer societies, the minimum standard of living to keep up with your peer group and to assure basic participation in activities regarded as common practice – in brief: to avoid social exclusion and hence deprivation –, is higher than in poor societies. Because of this and the perception of poverty in a society, poverty lines usually move upwards with mean income, and so should consequently a "social" level of subsistence.

Thus, subsistence levels should be allowed to differ between societies (in time and place). But this, of course, makes calculations more challenging. To keep things simple, we will follow Milanovic (2013) and work (at least for this paper) with a mean income elasticity of social subsistence ($b$), i.e. consider that the social subsistence level changes, when mean income is changing. This elasticity is reasonably non-negative but below 1, with a value of 0 representing the simple case presented before that social subsistence always equals physical (i.e., an absolute subsistence level totally independent of mean income) and a value of 1 representing the irrelevant case when social subsistence always equals mean income (and consequently, meaningful inequality would be mathematically impossible).[10] The level of social subsistence income ($\sigma$) is then formally given by:[11]

$$\sigma = s\left(\frac{m}{s}\right)^{b} = s\alpha^{b} \tag{3}$$

This means that the level of social subsistence is simply the level of physical subsistence multiplied by a term sensitive to different levels of mean income and calibrated to fulfil the range we discussed before. If we combine this with (1a) above (by replacing $s$ in the original formula with $\sigma$ as our new measure of relevant subsistence income), maximum feasible inequality with respect to social subsistence ($MFI^{*}$) results as:

$$MFI^{*} = 1 - \alpha^{b-1} \tag{4}$$

[9] A numerical example based on a subsistence income of 400$ per year: in this case the level of extraction in a 4,000$-society with a Gini coefficient of 0.6 is 67%, while the level of extraction in a 1,000$-society with a lower Gini coefficient of 0.5 is actually higher at 83%.

[10] Negative values are also mathematically possible, but only in the strange cases that poverty lines would fall in absolute terms in case of increasing income or rise in case of decreasing mean income levels.

[11] For a bit more maths, see again Milanovic (2013), pp. 7-9.

Because the ratio $\alpha$ is larger than or equal to 1 (or otherwise is unsustainable), and its exponent is reasonably somewhere between 0 and -1, $MFI^*$ will be between 0 and 1 as well as positively related to mean income und negatively to the level of physical subsistence.

The inequality extraction ratio based on social subsistence ($IER^*$) in relation to the respective maximum feasible inequality ($MFI^*$) and actual inequality ($G$) is given by:

$$IER^* = \frac{G}{MFI^*} \tag{5}$$

Following Milanovic (2013), it is reasonable to assume a mean income elasticity of social subsistence of 0.5, but higher or lower levels are of course possible and should be qualified with empirical data in follow-up work.[12] Under this assumption and if mean income in a society is double the level of physical subsistence income, the maximum level of feasible inequality would be confined to a Gini coefficient of only 0.29. Further, it would only be sustainably realized if all persons but one in this society earn the social level of subsistence income (which is about 1.4 times the level of physical subsistence, i.e. still not very much) while all the surplus is distributed to this one person. If elasticity is assumed to be lower (higher) than 0.5, the level of feasible inequality will turn out to be higher (lower) than 0.29. However, an extraction ratio of 100% is already achieved at this rather low value of the Gini coefficient.[13]

Generally, this recalibration of inequality data is of course very sensitive to assumptions about physical subsistence income (the higher it is set, the higher the level of extraction will turn out) and to assumptions about the mean income elasticity of social subsistence (the higher this elasticity is set, the higher will be the levels of extraction). Further, as a technical note, it should also be added that because – for mathematical reasons – $MFI$ as well as $MFI^*$ increase with an increase in mean income (although decreasingly so), it is likely that $IER$ and $IER^*$ will decrease with increasing income.[14] However, direction and magnitude of the effect is also dependent on the influence of this increase in mean income on the Gini coefficient. This means practically, that as long as the distribution of income growth is exactly mirroring the distribution of previous income or biased to favour the poor, $IER$ and $IER^*$ will decrease with increasing income level, but if growth favours the rich, $IER$ and $IER^*$ may also increase. Unfortunately, we cannot mathematically determine the respective threshold or the extent of the effect, because there is no mathematical relationship between the Gini coefficient and mean income ($dG/dm = 0$). Another serious caveat of all these calculations is certainly the poor quality of income data adjusted for purchasing power and the equally poor quality of inequality data, particularly for historical comparisons across time and space. A combination of these two will certainly multiply the estimation error. However, this is more a call for improvements in the databases than for abstaining from measure a meaningful phenomenon as good as we can with existing data.

[12] By using U.S. data in 2012 (mean income of 43,000$ per capita, a poverty line of 23,050$ for a 4-person-household and the assumption of physical subsistence at 300$ per capita), Milanovic (2013), p. 15, calculated a "real" mean income elasticity of social subsistence of 0.59 (for that country in that year).

[13] In case of a mean income 10 times the level of physical subsistence, the maximum feasible inequality is already at 0.68, in case of 100 times at 0.9 (under the assumption of $b = 0.5$), both levels considerably lower than without considering social subsistence.

[14] It is a mathematical property of the transformations applied in this paper, that at least the order of countries will necessarily be the same not matter if ordered by GDP per capita, $MFI$ or $MFI^*$.

## Some stylized facts about inequality and inequality extraction

Let us now take a first descriptive look into World Bank data (World Development Indicators, in brief WDI) about some inequality metrics. Table 1 provides an overview on GDP per capita and Gini coefficients as well as the resulting maximum feasible inequality and inequality extraction ratios of selected countries in years around 2010.

**TABLE 1: INEQUALITY METRICS IN AROUND 2010 (selected countries)**

| COUNTRY | YEAR | GDP p.c. | GINI | MFI | MFI* | IER | IER* |
|---|---|---|---|---|---|---|---|
| Switzerland | 2010 | 53,849 | 32.7 | 99.3 | 91.4 | 33.0 | 35.8 |
| United States | 2010 | 49,376 | 40.5 | 99.2 | 91.0 | 40.8 | 44.5 |
| Sweden | 2010 | 43,060 | 26.8 | 99.1 | 90.4 | 27.1 | 29.7 |
| Austria | 2010 | 43,006 | 30.3 | 99.1 | 90.4 | 30.5 | 33.5 |
| Germany | 2010 | 40,923 | 31.1 | 99.0 | 90.1 | 31.4 | 34.6 |
| Canada | 2010 | 40,713 | 33.7 | 99.0 | 90.1 | 34.0 | 37.4 |
| France | 2010 | 36,732 | 33.8 | 98.9 | 89.6 | 34.2 | 37.7 |
| United Kingdom | 2010 | 36,319 | 34.8 | 98.9 | 89.5 | 35.2 | 38.9 |
| Chile | 2011 | 20,154 | 50.8 | 98.0 | 85.9 | 51.9 | 59.2 |
| Turkey | 2010 | 16,758 | 38.8 | 97.6 | 84.6 | 39.7 | 45.9 |
| Brazil | 2011 | 14,301 | 53.1 | 97.2 | 83.3 | 54.6 | 63.8 |
| Thailand | 2010 | 12,822 | 39.4 | 96.9 | 82.3 | 40.7 | 47.9 |
| South Africa | 2011 | 11,910 | 63.4 | 96.6 | 81.7 | 65.6 | 77.6 |
| Colombia | 2010 | 10,777 | 55.5 | 96.3 | 80.7 | 57.6 | 68.7 |
| China | 2012 | 10,756 | 42.2 | 96.3 | 80.7 | 43.8 | 52.2 |
| India | 2011 | 4,883 | 35.2 | 91.8 | 71.4 | 38.3 | 49.2 |
| Pakistan | 2010 | 4,220 | 29.8 | 90.5 | 69.2 | 32.9 | 43.1 |
| Bangladesh | 2010 | 2,459 | 32.1 | 83.7 | 59.7 | 38.4 | 53.8 |
| Rwanda | 2010 | 1,262 | 51.3 | 68.3 | 43.7 | 75.2 | 117.5 |
| Ethiopia | 2010 | 1,081 | 33.2 | 63.0 | 39.2 | 52.6 | 84.7 |
| Congo, Dem. Rep. | 2012 | 742 | 42.1 | 46.1 | 26.6 | 91.4 | 158.4 |

Data Source: World Bank, World Development Indicators (online), data on inequality and income are from the year 2010 or the closest available (with respect to Gini coefficient) before or afterwards in a 5-year-interval.

Remarks: *MFI*, *MFI**, *IER* and *IER** are own calculations based on GDP per capita in PPP$ in constant values of 2011, a level of absolute subsistence of 400 PPP$ per capita and (for *MFI** and *IER**) the assumption of a mean income elasticity of social subsistence of 0.5.

Gini coefficients in the whole sample of 127 countries for which data in years between 2008 and 2012 is available vary between 24.8 and 63.4, GDP per capita between 737 and 89,979 PPP$ (in constant values of 2011). The data in table 1 vary between 26.8 (in the rather egalitarian case of Sweden) and 63.4 (in the very unequal case of South Africa, at the same time the sample maximum). Values for maximum feasible inequality, *MFI* and *MFI**, decrease with decreasing income levels from around 99 and 90 in rich countries to less than 50 in poor ones. This shows that while in rich countries almost all income could be transformed into inequality, the inequality possibility space is much smaller in poor countries.

Consequently, extraction levels are systematically larger in poorer countries: while all rich countries (with the notable exception of the United States) have extraction levels below 40% (no matter if using the concept of physical or social subsistence), all middle and lower income countries have extraction levels exceeding 40%, if calculated in consideration of social subsistence. Sometimes, the levels even exceed 100% as in the cases of Rwanda and the Democratic Republic of the Congo, with points to extreme forms of inequality and serious existential dangers for significant parts of the population in these countries directly related to distributional issues.[15]

The consequences for the interpretation of inequality metrics are important, indeed. Compare, for example, Switzerland, Bangladesh and Ethiopia. All three have almost identical Gini coefficients of 32.7, 32.1 and 33.2. But their extraction ratios differ significantly: while the level of extraction ($IER^*$) is still below 36% in Switzerland, but 54% in Bangladesh, and even exceeding 117% in Rwanda. This means that while in Switzerland only a third of surplus income is transformed into inequality, it is more than half in Bangladesh and in Rwanda, extraction even exceeds the level of maximum feasible inequality (at the obvious expense of serious deprivation). In the United States, the Gini coefficient is even much larger at 40.5, but the level of extraction is (at below 45%) still much lower than in Bangladesh.[16]

Let us now turn to a more analytical view at the data.[17] While there is no mathematical relationship between the Gini coefficient and mean income, there is at least a statistical one, as can be seen in table 2 and figures 1, 2 and 3. Generally, GDP per capita is negatively correlated with Gini coefficients and inequality extraction ratios, i.e. richer countries tend statistically to be less unequal than poorer ones, and figures 1a, 2a and 3a even could be read as (very weak) hand-sight evidence for Kuznets curves.[18] However, there are remarkable patterns in the data highly relevant for the argument of this paper, as shown in table 2, which displays correlation coefficients for income with inequality and extraction levels for years around 1990, 2000 and 2010. Further, robustness checks are provided with respect to high-income countries and outliers with high levels of extraction. What can be seen from these comparisons? First, correlation of extraction ratios with GDP per capita is systematically larger than that of raw Gini coefficients. The difference is smallest for the 2010-sample, but significant correlation of Gini coefficients with GDP per capita is even absent for the 2000-sample. All this can also be seen from the comparison of figure 1a, 2a and 3a with 1b, 2b and 3b. Secondly, if we exclude rich countries from the samples, correlation of extraction ratios with income changes considerably, but remains largely negative throughout, while correlation of Gini coefficients with income is totally disappearing. Hence, the general correlation of inequality with income is very much driven by a few rich countries, which on the other hand do hardly affect correlation of extraction with income. Thirdly, if we exclude countries with extraordinary large levels of extraction (which are particularly likely in very poor countries,

[15] World Bank data about poverty in these countries underlines that: The share of population living below the internationally comparable 1.90$-a-day poverty-line was 60.4 % in Rwanda 2013 (with a poverty gap of 24%) and 77.1 % in the DRC in 2012 (with a poverty gap of 39%), the latter being the highest level globally; but also the share of people below national poverty lines is large, with 45% in Rwanda 2010 and 64% in the DRC 2012.

[16] Even if we apply an elasticity of 0.59, which was calculated for the United Stated by Milanovic (2013), the level remains at 47% and hence considerably below the level in Bangladesh (and that of most other middle and lower income countries).

[17] However, keep in mind that the samples used are far from complete: they cover only about a third of all countries for years around 1990, more or less half for years around 2000, and still only about two thirds for years around 2010.

[18] Kuznets (1955) argued, based on data from the United States 1913–48, that there is an inverse U-shaped relationship between inequality and income: inequality is low in rather poor societies, increasing when they get wealthier, and finally decreasing again when they became rich.

because arrangements resulting in inequality extraction are certainly not growth promoting, and hence should drive correlation up), correlation of extraction on the basis of social subsistence is surprisingly even larger compared to the full sample, to the point of –0.67 in the 2010-sample. At the same time, correlation between income levels and Gini coefficients remains mainly insignificant (with the exception of the 2010-sample). Finally, correlation patters also differ slightly, when extraction ratios based on physical subsistence and extraction ratios based on social subsistence are compared, with the latter being systematically larger, although essentially because of mathematical reasons.

**TABLE 2: CORRELATION OF INCOME, INEQUALITY AND EXTRACTION LEVELS**

| **1990** | full sample | rich excluded | outliers excluded |
|---|---|---|---|
| no. of countries | 58 | 51 | 52 |
| GDP per capita / GINI | -0.288 (0.028) | -0.171 (0.231) | -0.220 (0.117) |
| GDP per capita / IER | -0.445 (0.000) | -0.477 (0.000) | -0.405 (0.003) |
| GDP per capita / IER* | -0.483 (0.000) | -0.551 (0.000) | -0.547 (0.000) |
| **2000** | full sample | rich excluded | outliers excluded |
| no. of countries | 94 | 86 | 84 |
| GDP per capita / GINI | -0.123 (0.236) | +0.111 (0.308) | -0.095 (0.390) |
| GDP per capita / IER | -0.406 (0.000) | -0.389 (0.000) | -0.331 (0.002) |
| GDP per capita / IER* | -0.479 (0.000) | -0.532 (0.000) | -0.523 (0.000) |
| **2010** | full sample | rich excluded | outliers excluded |
| no. of countries | 127 | 94 | 117 |
| GDP per capita / GINI | -0.423 (0.000) | -0.059 (0.571) | -0.393 (0.000) |
| GDP per capita / IER | -0.536 (0.000) | -0.445 (0.000) | -0.569 (0.000) |
| GDP per capita / IER* | -0.562 (0.000) | -0.561 (0.000) | -0.674 (0.000) |

Data Source: World Bank, World Development Indicators (online), data on inequality and income are from the reference year or the closest available (with respect to Gini coefficient) before or afterwards in a 5-year-interval.

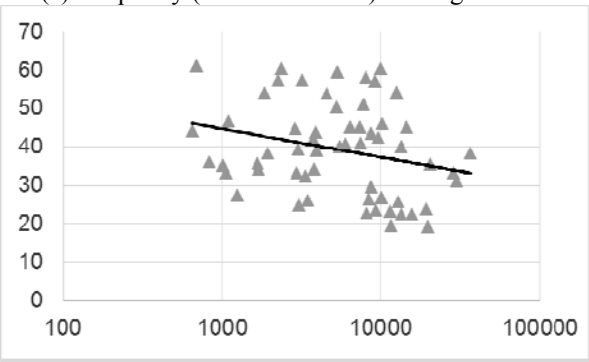
Remarks: 1) Data displayed are Pearson correlation coefficients of the respective data and two-sided error probabilities (p-values); 2) *IER* and *IER** are own calculations based on GDP per capita in PPP$ in constant values of 2011, a level of absolute subsistence of 400 PPP$ per capita and (for *IER**) the assumption of a mean income elasticity of social subsistence of 0.5; 3) "rich excluded" refers to countries with a GDP per capita lower than 20,000 PPP$ (2010) or lower than 15,000 PPP$ (1990 and 2000), "outliers excluded" refers to countries with $IER^* < 100$ (in all reference years).

All these results are also confirmed by a graphic comparison of scatter plots of Gini coefficients and extraction ratios based on social subsistence with (log) income, as can be seen in figures 1, 2 and 3. Trend lines are logarithmic and mirror the correlation coefficients displayed in table 2, full sample column (but keep in mind the different scales of the figures, which even plays differences down). It is obvious that the correlation of extraction with income levels is much more pronounced than with crude Gini coefficients. While especially among high income countries there are hardly any with high levels of inequality and consequently also hardly any with high levels of extraction, extraction levels are much larger in poorer countries. Low inequality there can be explained simply by a lack of opportunity:

Because there is hardly any surplus income available, there is only a small room for inequality in poor countries. Further, the extraction ratio also provides a more generalized argument for the seemingly Kuznets-like patterns, which are weakly embodied in figures 1a, 2a and 3a.

**FIGURE 1: INEQUALITY MEASURES IN AROUND 1990 (58 countries)**

(a) Inequality (Gini coefficient) and log income




(b) Inequality extraction ($IER^*$) and log income

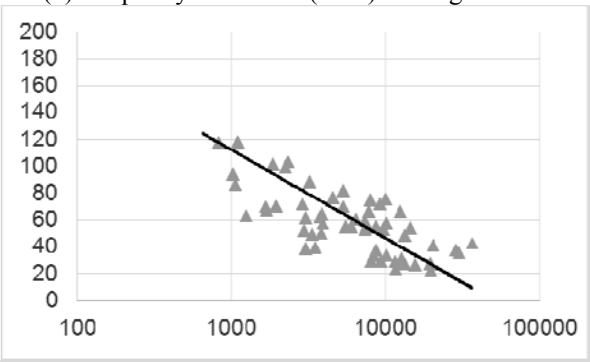
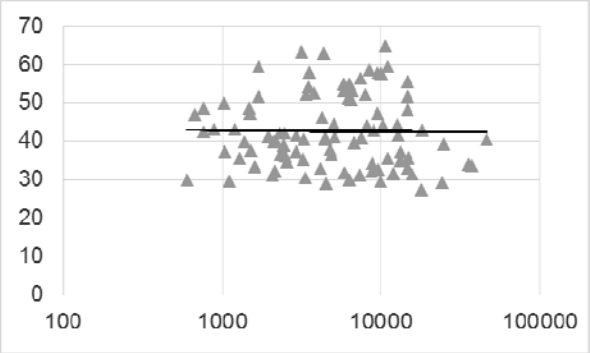



**FIGURE 2: INEQUALITY MEASURES IN AROUND 2000 (94 countries)**

(a) Inequality (Gini coefficient) and log income




(b) Inequality extraction ($IER^*$) and log income

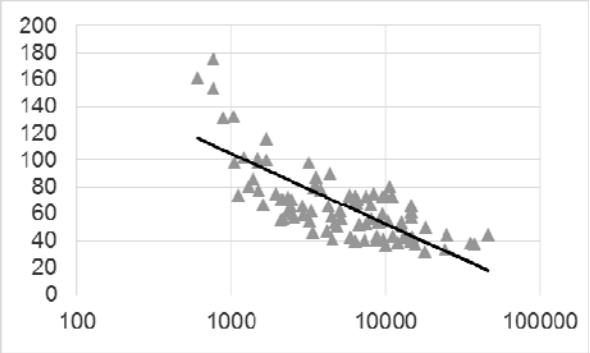
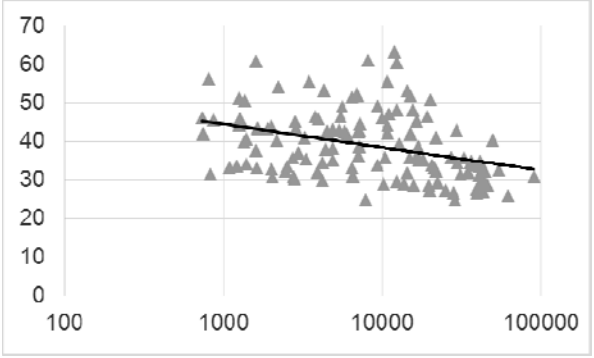
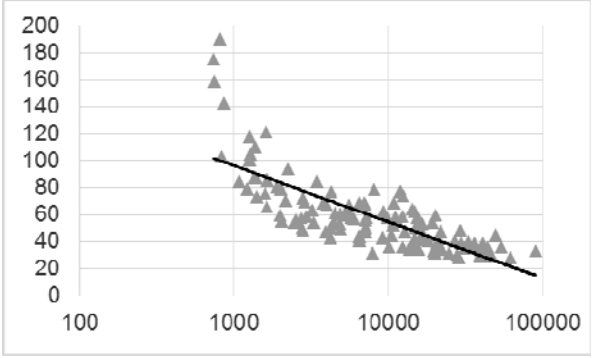



**FIGURE 3: INEQUALITY MEASURES IN AROUND 2010 (127 countries)**

(a) Inequality (Gini coefficient) and log income




(b) Inequality extraction ($IER^*$) and log income



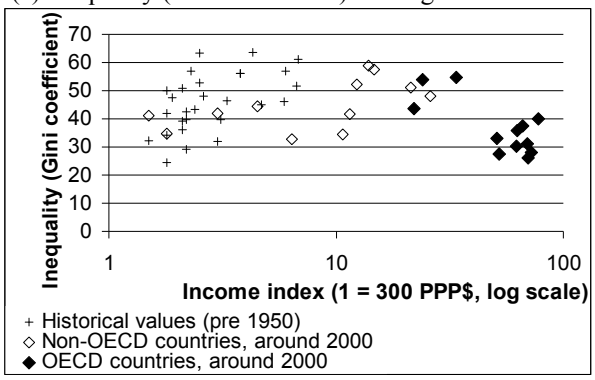

Data Source: World Bank, World Development Indicators (online), data on inequality and income are from the year 1990, 2000, or 2010 respectively, or the closest available (with respect to Gini coefficient) before or afterwards in a 5-year-interval.

Remarks: 1) Displayed trend lines are log-linear approximations; 2) The values of $IER^*$ are own calculations on the basis of GDP per capita in PPP$ in constant values of 2011, a level of absolute subsistence of 400 PPP$ per capita and (for $MFI^*$ and $IER^*$) the assumption of a mean income elasticity of social subsistence of 0.5; 3) outliers ($IER^* > 200$) are omitted from figures 1b (Central African Republic, Uganda) and 2b (Mozambique), but considered in trend calculation.

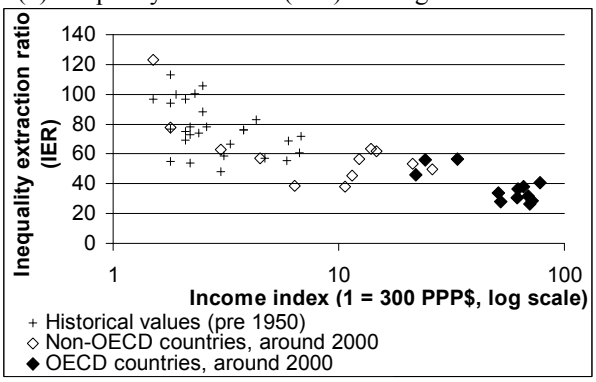
Over these 20 years, the overall pattern does not change fundamentally. Another comparison by using Milanovic's original pre-industrial and recent data (Milanovic et al. 2010), although calculated a bit different methodology, gives an even more pronounced picture of this general difference. In figure 4, three categories of countries are shown: 12 modern OECD countries, 12 modern non-OECD countries, and 28 pre-industrial societies.

**FIGURE 4: INEQUALITY MEASURES IN HISTORICAL COMPARISON**

(a) Inequality (Gini coefficient) and log income index




(b) Inequality extraction (*IER*) and log income index




Data Source: Milanovic et al. (2010), pp. 263-264; data calculated by the authors.

Remarks: 1) historical values include examples from all over the world: 5 from the (early) 20th century, 11 from the 19th, 6 from the 18th and 6 (all from Europe) earlier; 2) the values of *IER* are calculated by the authors on the basis of GDP per capita in PPP$ in constant values of 1990, a level of absolute subsistence of 300 PPP$ per capita and the assumption of a mean income elasticity of subsistence of 0; 3) a recalibration according to the methodology applied before (Tab. 1 and Fig. 1-3, own calculation) does not change any result substantially.

Together with figures 1, 2 and 3 this comparison shows how different the perspective becomes when we look at inequality from the angle of feasibility. While the data cloud in figure 4a embodies a statistical correlation of –0.397 (p=0.004) between inequality and income, the pattern does not look too much correlated and, if anything, Kuznets-U-like. This is obviously different for the data cloud in figure 4b: here we measure a statistical correlation between inequality extraction and income of –0.699 (p=0.000) and the pattern does also clearly look log-linear. Again, if we exclude the richest societies in the sample, the difference becomes more pronounced, correlation in figure 4a even turns to be positive.[19] Hence, this simple comparison of basic descriptive statistics already shows the considerable relevance of an extraction ratio perspective in historical comparisons of inequality data.

There are at least two further observations in need of better qualification. As a rule, very rich societies do rarely exceed a level of absolute extraction of 40% (the only close exception in Milanovic's data set is the United States in around 2000), while poorer societies do hardly fall below that level (the only close exceptions in Milanovic's data set are India and Indonesia in around 2000), which is the same statistical coincidence as already shown in table 1 with data from around 2010. However, there is also a temporal component to that argument: more recent data at the same income level tends to be lower in inequality and consequently also in extraction; further, at the country level (as can be seen in Milanovic et al. 2010), it is usually the case that extraction ratios decline with time, even if absolute inequality was rising.

[19] If the nine outliers with an income index larger than 50 (i.e. mean income of 15,000 PPP$ per capita or more) are excluded, correlation changes to +0.292 (p=0.057) in figure 4a and reduces to –0.503 (p=0.001) in figure 4b.

In the context of these arguments, a multivariate analysis of explanations of inequality and extraction and their dynamics is certainly needed.[20] Not only did most countries show positive growth performance at least over parts of the 20th century, resulting in rising average income, which means that – on the basis of statistical correlation – it is likely that they also became less unequal or at least less extractive. But the main reason for that is certainly not the passing of time, but the accumulation of wealth and rising demands for redistribution by the population. However, there are further factors not revealed by this simple descriptive comparison, especially institutional ones (which are certainly central when we come to explanations of extraction). Further, there is also a technical note about that relationship: as a matter of construction, *MFI* and $MFI^*$ decrease with increasing income levels and hence also *IER* and $IER^*$ tend to do the same.[21] We will come back on these issues in the discussion section of the paper.

# Top incomes and top income extraction

There is no particular reason why we should not recalibrate other inequality measures than the Gini coefficient following the logic of inequality extraction. Milanovic (2013, pp. 16-18 and annex) did that already for different Theill indices and the standard deviation of log incomes. We will apply the argument for the first time to also cover top income shares. Thanks to the efforts of scholars like Thomas Piketty and his colleagues, top income shares are today more widely available especially for historical time series as are Gini coefficients (which is partly due to the availability of historical tax data usually referring to top incomes only). While they by definition only give information about the upper tail of the distribution, there is also a strong statistical connection between top income shares and overall inequality. In an empirical study about top income shares and measures of inequality, like the Gini coefficient, the Atkinson Index, the 90:10-ratio, etc., Andrew Leigh finds a "strong positive relationship between the series, which is robust to the inclusion of country and year fixed effects" and concludes that "top income shares are far from perfect as a measure [of] the distribution of income across society. But where other data sources are limited, they may help to fill in some of the gaps."[22] Hence it is not only interesting to work with top income shares on their own sake, but also as a proxy for inequality as a whole, especially in historical times, which would allow long-run comparisons and analysis of dynamics in time.

In this context, the logic of the argument is the same as before, but the interpretation of extraction slightly differs. If we look at the share of the top-percentile, for example, then the hypothetically smallest possible share is of course 1% (in case of a totally equal distribution), while the largest income share of this richest 1% of the population, which could be sustainable, is not 100%, but a level that allows the other 99% at least the level of (social) subsistence. Hence the first step is again to consider feasibility and the second to relate actual income shares to feasible ones and hence to reinterpret the shares with respect to extraction.

In what follows, we will present the argument by use of social subsistence income, not physical, because we regard the further as the more realistic approach, while the latter is even

[20] Milanovic (2013), p. 19, already provided a first such analysis on the basis of his 28 historical data points. We will come back to that in the discussion section.
[21] They only tend to, because there is no direct mathematical influence of income on the Gini coefficient (again, $dG/dm = 0$), but – as shown – a statistical correlation, which may counterbalance the overall effect of a raise in income on *IER* and $IER^*$. It would be particularly welcome to isolate these more technical effects originating in the construction of the indicators, from the effects of actual interest, which are economic and socio-political.
[22] Leigh (2007), p. F630.

embodied in the former as a special case (of zero elasticity). Given a certain population share $t$ (for example the richest 10%, 1% or 0.1%), the maximum feasible top-income share of this group ($MFT_t^*$) is then simply given by total income minus social subsistence income for the rest of the population, expressed as a share of total income:

$$MFT_t^* = \frac{m - (1-t)s\alpha^{d_t}}{m} \tag{6a}$$

which contains a mean income elasticity of social subsistence, sensitive for the respective population share $t$ ($d_t$), because it is likely that reasonable values of this elasticity differ with increasing concentration (it is suggestive that they increase, but that issue needs empirical qualification). By introducing $\alpha$ as the ratio of mean to subsistence income, this can also be expressed as:

$$MFT_t^* = \frac{\alpha - 1}{\alpha} + t\alpha^{d_t - 1} \tag{6b}$$

Equation (6b) is the same as equation (1b) with an additional term. However, this term is rather small (its maximum is $t$, but only if $d_t = 0$), it is even decreasing with increasing $\alpha$ or $d_t$ and with decreasing $t$. Hence, especially for income shares of small elites in high-income countries, equations (1b) and (6b) are approximately identical.

The actual calculation of top income extraction is nevertheless slightly different. When we combine $MFT_t^*$ with the actual top-income share for the population share $t$ ($T_t$), we get the top-income extraction ratio with the application of a social subsistence level for the population share $t$ ($TER_t^*$), which is given by:

$$TER_t^* = \frac{T_t - t}{MFT_t^* - t} \tag{7}$$

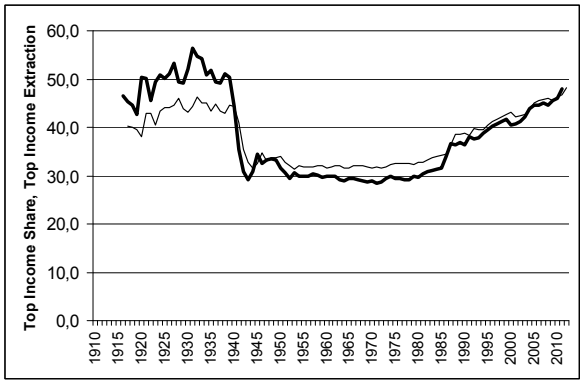
It is necessary to subtract $t$ from $T_t$ as well as $MFT_t^*$ because actual "extraction" is only possible at levels exceeding this lowest meaningful bound.[23] This also underlines that the interpretation of $TER_t^*$ is different from $IER^*$. Because the Gini coefficient ($G$) focuses on the whole distribution and is rather weak in addressing the margins (neither at the top nor at the bottom), $IER^*$ is also a rather indiscriminate measure of society-wide extraction. $TER_t^*$ is thus better suited to get an impression about actual extraction at the very top end of the distribution. Further, the lowest meaningful value for $G$ is 0 (in that case, also $IER^*$ should and does become 0), while the lowest meaningful value of $T_t$ is $t$ (and consequently, $TER_t^*$ should and does become 0 in that case). Additionally, because of that correction, $TER_t^*$ may even turn out to be smaller than $T_t$, especially for larger $t$'s (see for example figures 5a and 7). Further, to keep things simple for the moment, we assume the values of $d_t$ to be equal over all $t$'s and apply 0.5 as a reasonable value.[24] We apply that value for figures 5, 6 and 7 derived from data from the World Top Incomes Database and the New Maddison Project.

[23] Mathematically, it can only become 0 if $d = 1$ or if $m = s$, but it also allows for positive values in the case that $m < s$. Further, strictly mathematical it could even become negative, but only in the illogical case that the "richest" people earn less (!) income then poorer ones.

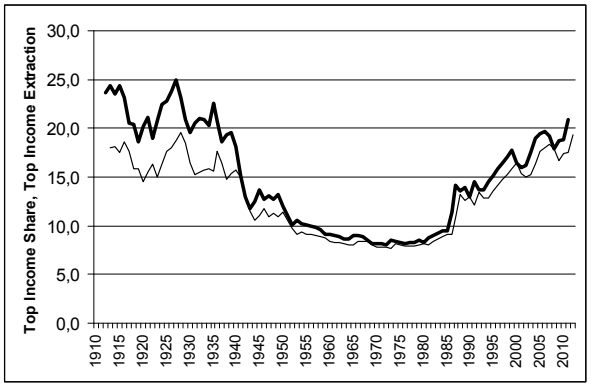
[24] Given actual distributions, $d_t = 0.5$ is especially reasonable for $t = 0.2$, because mean income is often close to actual income levels around the 80th percentile. Having said that, it is nevertheless likely that $d_t$ will be bigger for a smaller $t$, because the upper tail of the distribution is certainly more sensitive and able to avoid absolute poverty than the lower tail, hence incomes among these people will considerably exceed subsistence levels. It is hence very welcome to have empirically based assumptions or qualifications for $d_t$.

## FIGURE 5: TOP-INCOME EXTRACTION IN THE UNITED STATES

(a) Top-10% income share (thin line) and extraction ratio (solid line), U.S. 1917-2012

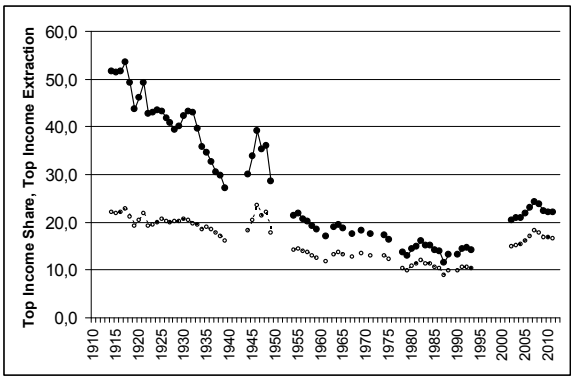



(b) Top-1% income share (thin line) and extraction ratio (solid line), U.S. 1913-2012

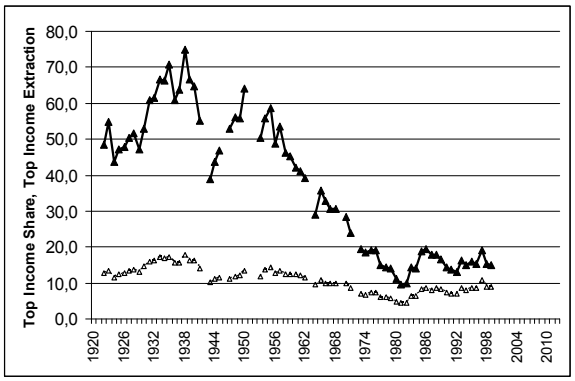



## FIGURE 6: TOP-INCOME EXTRACTION IN SOUTH AFRICA AND INDIA

(a) Top-1% income share (thin line/dots) and extraction ratio (bold line/dots), South Africa 1914-2011

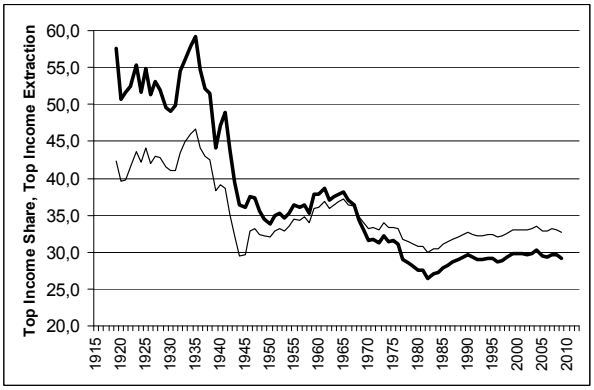



(b) Top-1% income share (thin line/dots) and extraction ratio (bold line/dots), India 1922-1999

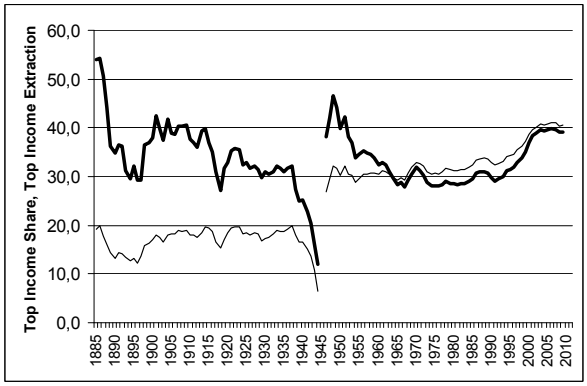



## FIGURE 7: TOP-INCOME EXTRACTION IN FRANCE AND JAPAN

(a) Top-10% income share (thin line) and extraction ratio (solid line), France 1919-2009




(b) Top income extraction (solid line), Japan 1886-2010 (Top-1% until 1945, Top-10% from 1947)




Data Sources: World Top Income Database (online) for data on top income shares, New Maddison Project Database (online) for data on income levels.

Remark: the values of $TER_t^*$ are own calculations on the basis of GDP per capita in PPP$ in constant values of 1990, a level of absolute subsistence of 400 PPP$ per capita and the assumption of a mean income elasticity of social subsistence of 0.5.

With this data from the World Top Incomes Database, which covers about 30 countries worldwide with time-series over much of the 20th century, partly even on an annual basis, we are able to get a clearer picture of developments during the last decades – which is impossible with respect to historical Gini coefficients, which are hardly available as such.[25] It turns out that extraction patterns are not so different from patterns of inequality and as diverse as these:[26] we see the typical U-shaped pattern in the United States (figure 5), which is obviously largely driven by developments of the top-1%-income share (figure 5b), or South Africa (figure 4a); we see the L-shaped pattern in countries like France (figure 5a) or India (figure 4b); and we see that these developments are likely to be more pronounced when looked at from the angle of extraction.

But the main message from the exercise of recalculating inequality data in the light of extraction is certainly about the differences it makes for poor countries. Not surprisingly, data clearly shows that top income extraction ratios are much larger than actual top income shares for example in South Africa and India, where the top-1%-income shares hardly exceed 20% (already a remarkable value, indeed), but extraction is exceeding 40% in early post-colonial South Africa and even 60% in colonial India (with the largest value of 75.0 in 1938), which actually means that almost or even more than half of all surplus income in these countries was extracted by the richest 1% of population. Hence, for example India and the United States show totally different patterns of top incomes extraction, while they look rather similar from the angle of crude top income shares. In this context, it is also worth to note that contrary to the level of inequality, the level of extraction in Japan over much of the second half of the 20th century was lower (sic!) on the top-10%-level than it had been on the top-1%-level over much of the first half of the 20th century (figure 7b).

# Discussion

There are two main contributions of applying an instrument like the extraction ratio to actual inequality data. The first one is that it reveals different patterns and developments, sometimes to the point of trend reversal. The second one is that it allows to ask (and hence answer) different questions. The main issue of this paper is to underline the first of these conclusions, while the main issue of this discussion part is to fuel the debate about the second.

This paper has especially provided a first look into top income shares from the World Top Incomes Database from the perspective of inequality extraction. Further, we also worked systematically into Gini coefficients from the World Development Indicators to show, how the data appears different after just a small change of the angle, from which to look at it. While there is significant and robust negative correlation between extraction and income levels, correlation between inequality and income levels is much smaller and even largely disappears, when we exclude rich countries. Hence, the descriptive evidence presented here shows that extraction is a much more pronounced issue than inequality as such, especially in historical contexts and in poorer societies. Historical levels of extraction are usually much larger than recent ones and anecdotal evidence confirms that in poorer societies the levels of extraction often decrease over time while absolute inequality is actually increasing. In Japan

[25] The World Income Inequality Database (online) contains several values, but they are rather diverse with respect to methodology applied, and hence do not combine to many consistent time series.
[26] See for an excellent overview on the diverse patterns of inequality development in the 20th century Atkinson et al. (2011), especially pp. 40-48.

between 1900 and 1940, for example, absolute inequality was increasing (the top-1%-income share was rising from about 15 to 20%), while inequality extraction was decreasing (from about 40 to 30%). This of course also mirrors a widening of the inequality possibility frontier with increasing income. Consequently, rich societies do not only show lower levels of inequality (especially compared to middle-income countries), but particularly also lower levels of extraction, a pattern remarkably stable over time. However, more recently we also observe that the levels of extraction follow the upward path of income inequality in these rich societies. Taken together, this takes the usual Kuznets-argument about inequality trends from its head back on the feet by re-inverting the inverse-U of inequality to again depicting a potentially "real" U of inequality extraction. While this mechanic picture is certainly too less nuanced to perfectly reflect reality and the upward part is still comparably small, it may provide a more promising workhorse hypotheses to analyse actual connections than Kuznets' original argument (which historically had its strong merits, anyway). A recent contribution to this kind of discussion is again provided by Milanovic (2016), where the main argument to explain long-run inequality dynamics are what he refers to as "Kuznets cycles" associated to innovation waves: inequality is rising at the beginning of a technological boom and falls again with increasing redistribution of its gains. However, the extraction perspective may also clarify the reasons for this and the mechanisms by which this is done as well as possibly reveal shifts of extraction levels from one innovation wave to the next.

**TABLE 3: GLOBAL INEQUALITY AND EXTRACTION**

| Year | Gini | income | MFI | IER | MFI* | IER* | MFI** | IER** |
|---|---|---|---|---|---|---|---|---|
| 1820 | 0.54 | 667 | 0.55 | 98 | 0.33 | 164 | 0.45 | 120 |
| 1850 | 0.56 | 791 | 0.62 | 90 | 0.38 | 146 | 0.52 | 108 |
| 1870 | 0.59 | 783 | 0.62 | 96 | 0.38 | 155 | 0.51 | 115 |
| 1890 | 0.61 | 1,133 | 0.74 | 83 | 0.49 | 126 | 0.63 | 97 |
| 1910 | 0.64 | 1,465 | 0.80 | 80 | 0.55 | 117 | 0.70 | 92 |
| 1929 | 0.67 | 1,784 | 0.83 | 81 | 0.59 | 114 | 0.74 | 91 |
| 1950 | 0.69 | 2,113 | 0.86 | 80 | 0.62 | 111 | 0.77 | 90 |
| 1960 | 0.68 | 2,775 | 0.89 | 76 | 0.67 | 101 | 0.81 | 84 |
| 1970 | 0.68 | 3,736 | 0.92 | 74 | 0.72 | 95 | 0.85 | 80 |
| 1975 | 0.72 | 4,095 | 0.93 | 78 | 0.73 | 99 | 0.86 | 84 |
| 1980 | 0.68 | 4,521 | 0.93 | 73 | 0.74 | 92 | 0.87 | 78 |
| 1985 | 0.68 | 4,763 | 0.94 | 73 | 0.75 | 91 | 0.87 | 78 |
| 1990 | 0.69 | 5,162 | 0.94 | 73 | 0.76 | 91 | 0.88 | 78 |
| 1995 | 0.69 | 5,452 | 0.94 | 73 | 0.77 | 90 | 0.89 | 78 |
| 2000 | 0.72 | 6,029 | 0.95 | 76 | 0.78 | 93 | 0.89 | 80 |

Data Source: van Zanden et al. (2011), p. 34. Global Gini coefficients are estimated by the authors, income data is average world GDP per capita in PPP$ (of 1990) is taken by them from the New Maddison Project.

Remarks: 1) *MFI*, *MFI**, *MFI***, *IER*, *IER** and *IER*** are own calculations; 2) *MFI* and *IER* are calculated on the basis of GDP per capita in PPP$ in constant values of 1990, a level of absolute subsistence of 300 PPP$ per capita and the assumption of a mean income elasticity of subsistence of 0; 3) the other measures are calculated with the same methodology, but additionally for *MFI** and *IER** a mean income elasticity of subsistence of 0.5 is assumed, and for *MFI*** and *IER*** a mean income elasticity of subsistence of 0.25.

Further, the extraction argument may also be applied to global inequality (understood as inequality among people). The most sophisticated attempt so far to calculate a meaningful estimation for these kinds of global Gini coefficients (on the basis of various indicators related to inequality across several countries for the period 1820–2000) is provided by van Zanden et al. (2014). If we apply the methodology used to produce figure 4 (which is appropriate given the nature of this data, which is comparable to the data displayed and processed there), the pattern revealed in table 4 is quite striking: although subject to various estimation errors, we may safely conclude that the level of extraction declined from values close to 100% over most of the 19th century to level around 80% over the first half of the 20th century and further to 73% by the end (the slight increase in 2000 is within the estimation error).[27] Further, if we consider social subsistence, then the level of extraction still exceeds 90% by the end of the 20th century, but having decreased continuously between 1870 and 1970. However, while this clearly shows that the global arena is still a very extractive place, certainly not the least because of the absence of any level of governance comparable to nation states, all these numbers have to be interpreted with great caution, because the quality of the underlying data is still rather weak. It also shows sensitivity to assumptions. If we apply a lower mean income elasticity of social substitution, which is reasonable precisely because of a lack of governance at the global level, which would be able to implement any kind of balancing mechanism, for example let $b = 0.25$, then the extraction ratios decline to 115% in 1870, less than 100% from 1890 on and finally a minimum of 78% in the 1980s and 1990s.

This case also shows that refinements in calibrating extraction data with respect to the level of social subsistence are certainly necessary. In this paper, the meaningful but overly simple assumption was applied to fix the respective mean income elasticity of subsistence at 0.5. It is certainly much more appropriate to let this elasticity vary in time and/or space (and as a consequence, the levels of social subsistence), which will definitively change the observed developments and patterns, because the level of extraction (and hence sometimes even the observed pattern) is very sensitive to the applied concept and level of subsistence. This also holds for a very likely share-sensitive elasticity with respect to top income shares. In this context, it would also be desirable to have more nuanced information about historical living standards to adapt social minima to better reflect actual conditions and to have information about the public acceptability of certain levels of inequality and poverty. All this would contribute to a better empirical foundation of the measurement concept applied, which is certainly a necessary step of improvement.

Another point to be addressed with consecutive research is the often large difference between inequality and extraction levels, especially in poor societies. While the extraction ratio is not a necessary instrument to understand the urgency of economic and political problems in countries where it exceeds 100%, it redirects attention to those where it is less excessive but still considerable. And this is a large group, indeed. While the Gini coefficient is only in 19% of all data points included in the WDI database since 1990 exceeding 0.5, the extraction ratio (on the basis of social subsistence) is exceeding 50% in no less than 580 of the 1,117 cases for which the necessary data is available (i.e. 52%, representing at least one point in time in 106 countries), as well as 66.7% in 286 (i.e. 26%, representing 73 countries), but 100% in only 49 (i.e. 4%, representing 21 countries).

[27] Milanovic (2009) already provided an analysis and a discussion of this issue for several inequality metrics, but he was confined to a far less sophisticated dataset. Hence, his result did show extraction ratios (with respect to his estimated global Gini coefficients) in the Seventies throughout.

Further research will hopefully contribute to stabilize the measurement of inequality extraction and hence to reveal more nuanced patterns than this rather descriptive and explorative exercise was able to. On this basis, a lot of potential research questions arise. The two in my view most promising questions are on the one hand, what does influence the level of extraction, and on the other how extraction does influence human development and its components. It is very tentative to expect that larger levels of extraction will have negative consequences for the dimensions of human development, including economic growth, and hence are harmful at least on the societal level. Besides the very rare cases of extremely benevolent (and very clear-sighted) dictators, large levels of extraction are reflecting a systematic lack of economic opportunity, which will depress individual activities and hence harm growth (for example via the obvious channel of insecure property rights). Hence, it is very likely that more refined empirical analysis will find more pronounced negative effects of extraction patterns (compared to inequality levels) on economic growth, and even more so to human development in a more general way. To our knowledge, no analysis of this kind has been carried out so far.

Secondly, the basic general explanation of the level of extraction is certainly about the political opportunity of the population to successfully demand a distribution (or at least re-distribution) of income from the elites which is perceived as "fair". This is obviously related to political and economic inclusion, insufficiently characterized maybe by terms like "democratization" (in the political sphere) and "liberalization" (in the economic sphere). Historically however, as for example the scattered empirical evidence from figure 4 shows, elites tended to extract as much as possible from their populations. Milanovic (2013, p. 19) even ran a regression analysis with the 28 historical observations and came to the conclusion that cases which were colonies tend to have much larger extraction ratios than those which were not, and that also smaller population density (but not urbanization in neither direction) did yield significant results. Both effects taken together even statistically offset the seemingly large connection of extraction ratios with income levels.[28]

More recently, a larger degree of redistribution seems to be necessary to achieve at least social appeasement, not to talk about stronger social cohesion. As long as this pattern of redistributive arrangements has the fragile nature of a "limited access order",[29] distribution is subject to institutions often directly associated with the threat or even actual use of violence. This reflects a serious development trap: as long as violence is an integral part of these arrangements, increased extraction is usually going hand in hand with decreased income, because the associated insecurity limits economic activities in general (or at least makes them more expensive because of higher costs of protection). This also partly explains, why we still observe some very large levels of extraction. These usually exist in political environments, which are either characterized by civil wars or strong political oppression with little capacity of political revolution.[30]

[28] While neither the geometrically decreasing effect of GDP per capita nor the increasing effect of urbanization proved to be statistically significant, the effect of colonization is large (extraction ratios are increased by 25 points in colonies) and the effect of population density considerable (extraction ratios are decreased by 2 points per 10 people more per square kilometer). Milanovic also shows that these conclusions hold for extraction calculated based on Gini coefficients as well as Theill coefficients.

[29] On that concept, see North et al. (2009).

[30] Milanovic (2013, pp. 20-23) analyses the influence of extraction on civil war dynamics and finds evidence that they are positively related, i.e. that higher levels of extraction contribute to a longer duration of conflicts. Generally, he finds stronger effects of extraction on conflicts than of inequality.

Only in the process of "maturing" and approaching the "doorstep conditions" to an "open access order",[31] societies are able to utilize more potential. This is usually associated with an inclusion process – at first at the level of elites, later also of a more general nature. This means "extending the franchise"[32] and hence allowing more room for (re-)distributive demands. The logical consequence is not only decreasing inequality, but especially – and even more pronounced – decreasing extraction. This may lead to the point when populations actually even accept larger levels of inequality in a development process as long as they are accompanied with credible commitments of lower and sustainably decreasing levels of extraction. Following the argument of Acemoglu and Robinson as well as North, Wallis and Weingast, this is only possible when it is in the interest of the elites to assure social peace by way of institutionalizing power-sharing arrangements (in more mature orders) or violence containment (in more fragile orders). Overall, in both kinds of societies, incentives faced by the elites and actual opportunities faced by the populations are crucial to understand the possibility space.

Hence, it is very likely that more refined empirical analysis will find that lower levels of extraction are related to more inclusive institutional arrangements. This would explain the general historical trend that extraction ratios tend to fall over time (which may be explained by an "extended franchise") and especially that they tend to be higher in colonial or otherwise repressive environments.[33] But it does not as such explain that countries with higher incomes tend to have lower levels of extraction. While it is clear that more income also leaves more to be redistributed, for extraction to fall it is not sufficient that additional income is distributed more equally (pro-poor growth) but it is necessary that income growth for the poor is also outpacing the increase of the social subsistence level. Hence, when we observe decreasing extraction ratios both conditions have to be fulfilled.

Finally, it is especially promising to extend the analysis to address global distribution and extraction more appropriately. It is very likely that the global arena is (and has been historically) a place of (very) limited political and economic access, and this is certainly not without consequences for extraction patterns globally, but also at more local levels. While we have addressed the characterization of the global order in the terminology of North et al. (2009) at least preliminarily elsewhere (Exenberger 2016), it is still totally open to further research to combine the levels and hence uncover the entanglement of global and local governance and extraction as another explanation for large extraction levels. In the end, the actual channels of extraction are not well understood so far and will very likely contain some kind of international cooperation of elites and some institutions by which this is organized (empire, financialization, citizenship rents, etc.). The case of imperial connections is just the most obvious tip of an iceberg and it is likely to be insufficient to analyse extraction ratios in colonies independently from their metropolises, but under the surface of this iceberg lies a large amount of research about the international and global entanglement of extraction patterns.

[31] See again North et al. (2009) on these concepts.

[32] This follows the arguments of Daron Acemoglu and James Robinson, who published a respective article already in 2000 and a book in 2006. North et al. (2009) apply a quite comparable argument, which is more context-sensitive but also far less formal and hence less operationalizable.

[33] Of course, also an alternative hypothesis deserves more close investigation: the extraction ratio could also be a useful proxy variable for the "extractiveness" (or otherwise inclusiveness) of an economy.

# Concluding remarks

Further research is clearly welcome on all the issues addressed in this paper and it is very likely to be a fruitful exercise. This paper has underlined the general necessity of refining, applying and explaining the extraction ratio as a relevant economic and political factor in historical as well as recent times, and it has proposed to extend the logic of the extraction ratio also on top income shares, which are much more widely available, especially in historical settings. Further research should also target patterns and trends specifically, i.e. be sensitive not only for context but especially also for dynamics. There is only one final technical caveat which should not be left unnoticed, although it is certainly well known to all working with developing country data. Besides all improvements of data quality with respect to recent as well as historical data, in many of the surveyed countries both, income nor inequality data, are very accurate and hence their combination is likely to multiply the measurement errors embodied in them. This has to be taken into account when we come to the level of actual conclusions and even more so policy recommendations. However, as long as we regard working with inequality metrics a meaningful endeavour, we shall not abstain from neither measuring nor concluding nor recommending only because the tasks are not trivial.

**Acknowledgements**

I want to thank Veronika Eberharter, Andrea Leiter-Scheiring, Andreas Pondorfer, and Engelbert Theurl, as well as the participants of the conference "Crisis and the Distribution" in Berlin (29-30 September 2014), an economic policy lecture in Vienna (21 April 2015), and the participants of the conference "Global Inequality: A Divided History" (Warwick, 19-21 April 2017) for valuable comments to preliminary work and early versions of this manuscript. All remaining errors and omissions are of course my own.

# Sources

**Cited Literature:**

Acemoglu D., Robinson J.A. (2000): "Why Did the West Extend the Franchise? Democracy, Inequality, and Growth in Historical Perspective", *Quarterly Journal of Economics* 115, pp. 1167-1199.

Acemoglu D., Robinson J.A. (2006): *Economic Origins of Dictatorship and Democracy*. Cambridge/UK: Cambridge University Press.

Atkinson A.B., Piketty T., Saez E. (2011): "Top Incomes in the Long Run of History", *Journal of Economic Literature* 49, pp. 3-71.

Cingano F. (2014): *Trends in Income Inequality and its Impact on Economic Growth*, OECD Social, Employment and Migration Working Papers 163, Paris: OECD Publishing.

Exenberger A. (2016): "African 'Limited Access Orders' and Post-colonial Conflicts", in: A. Exenberger and U. Pallua, eds.: *Africa Research in Austria: Approaches and Perspectives*. Innsbruck: innsbruck university press, pp. 153-170.

Halter D., Oechslin M., Zweimüller J. (2014): "Inequality and Growth: The Neglected Time Dimension", *Journal of Economic Growth* 19, pp. 81-104.

Kuznets S. (1955): "Economic Growth and Income Distribution", *American Economic Review* 45, pp. 1-28.

Leigh A. (2007): "How Closely Do Top Income Shares Track Other Measures of Inequality?", *The Economic Journal* 117, pp. F619-F633.

Milanovic B. (2009): *Global Inequality and the Global Inequality Extraction Ratio: The Story of the Past Two Centuries*, Policy Research Working Paper No. 5044, Washington/DC: The World Bank.

Milanovic B. (2013): *The Inequality Possibility Frontier Extensions and New Applications*, Policy Research Working Paper No. 6449, Washington/DC: The World Bank.

Milanovic B. (2016): *Global Inequality: A New Approach for the Age of Globalization*. Cambridge/MA, London: The Belknap Press of Harvard University Press.

Milanovic B., Lindert P.H., Williamson J.G. (2007): "Measuring pre-industrial inequality", NBER Working Paper No. 13550, Cambridge/MA: National Bureau of Economic Research.

Milanovic B., Lindert P.H., Williamson J.G. (2010): "Pre-Industrial Inequality", *The Economic Journal* 121, pp. 255-272.

North D.C., Wallis J.J., Weingast B.R. (2009): *Violence and Social Orders: A Conceptual Framework for Interpreting Recorded Human History*. Cambridge/UK: Cambridge University Press.

OECD (2015): *In It Together: Why Less Inequality Benefits All*, Paris: OECD Publishing.

Ostry J., Berg A., Tsangarides C. (2014): *Redistribution, Inequality, and Growth*, IMF Staff discussion note 14/02, Washington/DC: IMF Publishing.

van der Weide R., Milanovic B. (2014): *Inequality is bad for growth of the poor (but not for that of the rich)*, Policy Research Working Paper No. 6963, Washington/DC: The World Bank.

van Zanden J.L., Baten J., Földvari P., van Leeuwen B. (2014): "The Changing Shape of Global Inequality 1820–2000: Exploring a New Dataset", *Review of Income and Wealth* 60, pp. 279-297.

**Data sources:**

Milanovic B., Lindert P.H., Williamson J.G. (2010): "Pre-Industrial Inequality", *The Economic Journal* 121, pp. 255-272.

New Maddison Project Database: http://www.ggdc.net/maddison

World Development Indicators: http://databank.worldbank.org

World Top Income Database: http://topincomes.parisschoolofeconomics.eu

van Zanden J.L., Baten J., Földvari P., van Leeuwen B. (2011): *The Changing Shape of Global Inequality 1820–2000: Exploring a New Dataset*, CGEH Working Paper 1, Utrecht: Centre for Global Economic History.

**University of Innsbruck - Working Papers in Economics and Statistics**
**Recent Papers** can be accessed on the following webpage:

http://eeecon.uibk.ac.at/wopec/

2017-09 **Andreas Exenberger:** The Logic of Inequality Extraction: An Application to Gini and Top Incomes Data

2017-08 **Sibylle Puntscher, Duc Tran Huy, Janette Walde, Ulrike Tappeiner, Gottfried Tappeiner:** The acceptance of a protected area and the benefits of sustainable tourism: In search of the weak link in their relationship

2017-07 **Helena Fornwagner:** Incentives to lose revisited: The NHL and its tournament incentives

2017-06 **Loukas Balafoutas, Simon Czermak, Marc Eulerich, Helena Fornwagner:** Incentives for dishonesty: An experimental study with internal auditors

2017-05 **Nikolaus Umlauf, Nadja Klein, Achim Zeileis:** BAMLSS: Bayesian additive models for location, scale and shape (and beyond)

2017-04 **Martin Halla, Susanne Pech, Martina Zweimüller:** The effect of statutory sick-pay on workers' labor supply and subsequent health

2017-03 **Franz Buscha, Daniel Müller, Lionel Page:** Can a common currency foster a shared social identity across different nations? The case of the Euro.

2017-02 **Daniel Müller:** The anatomy of distributional preferences with group identity

2017-01 **Wolfgang Frimmel, Martin Halla, Jörg Paetzold:** The intergenerational causal effect of tax evasion: Evidence from the commuter tax allowance in Austria

2016-33 **Alexander Razen, Stefan Lang, Judith Santer:** Estimation of spatially correlated random scaling factors based on Markov random field priors

2016-32 **Meike Köhler, Nikolaus Umlauf, Andreas Beyerlein, Christiane Winkler, Anette-Gabriele Ziegler, Sonja Greven:** Flexible Bayesian additive joint models with an application to type 1 diabetes research

2016-31 **Markus Dabernig, Georg J. Mayr, Jakob W. Messner, Achim Zeileis:** Simultaneous ensemble post-processing for multiple lead times with standardized anomalies

2016-30 **Alexander Razen, Stefan Lang:** Random scaling factors in Bayesian distributional regression models with an application to real estate data

2016-29 **Glenn Dutcher, Daniela Glätzle-Rützler, Dmitry Ryvkin:** Don't hate the player, hate the game: Uncovering the foundations of cheating in contests

2016-28 **Manuel Gebetsberger, Jakob W. Messner, Georg J. Mayr, Achim Zeileis:** Tricks for improving non-homogeneous regression for probabilistic precipitation forecasts: Perfect predictions, heavy tails, and link functions

2016-27 **Michael Razen, Matthias Stefan:** Greed: Taking a deadly sin to the lab

2016-26 **Florian Wickelmaier, Achim Zeileis:** Using recursive partitioning to account for parameter heterogeneity in multinomial processing tree models

2016-25 **Michel Philipp, Carolin Strobl, Jimmy de la Torre, Achim Zeileis:** On the estimation of standard errors in cognitive diagnosis models

2016-24 **Florian Lindner, Julia Rose:** No need for more time: Intertemporal allocation decisions under time pressure

2016-23 **Christoph Eder, Martin Halla:** The long-lasting shadow of the allied occupation of Austria on its spatial equilibrium

2016-22 **Christoph Eder:** Missing men: World War II casualties and structural change

2016-21 **Reto Stauffer, Jakob Messner, Georg J. Mayr, Nikolaus Umlauf, Achim Zeileis:** Ensemble post-processing of daily precipitation sums over complex terrain using censored high-resolution standardized anomalies *published in Monthly Weather Review*

2016-20 **Christina Bannier, Eberhard Feess, Natalie Packham, Markus Walzl:** Incentive schemes, private information and the double-edged role of competition for agents

2016-19 **Martin Geiger, Richard Hule:** Correlation and coordination risk

2016-18 **Yola Engler, Rudolf Kerschbamer, Lionel Page:** Why did he do that? Using counterfactuals to study the effect of intentions in extensive form games

2016-17 **Yola Engler, Rudolf Kerschbamer, Lionel Page:** Guilt-averse or reciprocal? Looking at behavioural motivations in the trust game

2016-16 **Esther Blanco, Tobias Haller, James M. Walker:** Provision of public goods: Unconditional and conditional donations from outsiders

2016-15 **Achim Zeileis, Christoph Leitner, Kurt Hornik:** Predictive bookmaker consensus model for the UEFA Euro 2016

2016-14 **Martin Halla, Harald Mayr, Gerald J. Pruckner, Pilar García-Gómez:** Cutting fertility? The effect of Cesarean deliveries on subsequent fertility and maternal labor supply

2016-13 **Wolfgang Frimmel, Martin Halla, Rudolf Winter-Ebmer:** How does parental divorce affect children's long-term outcomes?

2016-12 **Michael Kirchler, Stefan Palan:** Immaterial and monetary gifts in economic transactions. Evidence from the field

2016-11 **Michel Philipp, Achim Zeileis, Carolin Strobl:** A toolkit for stability assessment of tree-based learners

2016-10 **Loukas Balafoutas, Brent J. Davis, Matthias Sutter:** Affirmative action or just discrimination? A study on the endogenous emergence of quotas *published in Journal of Economic Behavior and Organization*

2016-09 **Loukas Balafoutas, Helena Fornwagner:** The limits of guilt

2016-08 **Markus Dabernig, Georg J. Mayr, Jakob W. Messner, Achim Zeileis:** Spatial ensemble post-processing with standardized anomalies

2016-07 **Reto Stauffer, Jakob W. Messner, Georg J. Mayr, Nikolaus Umlauf, Achim Zeileis:** Spatio-temporal precipitation climatology over complex terrain using a censored additive regression model

2016-06 **Michael Razen, Jürgen Huber, Michael Kirchler:** Cash inflow and trading horizon in asset markets

2016-05 **Ting Wang, Carolin Strobl, Achim Zeileis, Edgar C. Merkle:** Score-based tests of differential item functioning in the two-parameter model

2016-04 **Jakob W. Messner, Georg J. Mayr, Achim Zeileis:** Non-homogeneous boosting for predictor selection in ensemble post-processing

2016-03 **Dietmar Fehr, Matthias Sutter:** Gossip and the efficiency of interactions

2016-02 **Michael Kirchler, Florian Lindner, Utz Weitzel:** Rankings and risk-taking in the finance industry

2016-01 **Sibylle Puntscher, Janette Walde, Gottfried Tappeiner:** Do methodical traps lead to wrong development strategies for welfare? A multilevel approach considering heterogeneity across industrialized and developing countries

University of Innsbruck

Working Papers in Economics and Statistics

2017-09

Andreas Exenberger

# The Logic of Inequality Extraction: An Application to Gini and Top Incomes Data

**Abstract**
As embodied in the concept of "inequality extraction" (Branko Milanovic), it is not possible to increase inequality (especially income inequality) in a society sustainably to levels beyond what is actually socially acceptable (and even less to levels endangering physical subsistence of parts of the population). Consequently, in international and intertemporal comparisons of inequality levels on the country (or the regional) level and for poorer societies in general, adaptations for the levels of inequality, which are actually feasible in physical and social terms, are necessary to arrive at meaningful conclusions. We extend this concept to cover also top-incomes ratios, which allows a broader database for measurement and analysis of cross-country historical inequality. A first look at the data uncovers different patterns and dynamics of inequality extraction when compared to untransformed inequality measures, sometimes to the point of trend reversal, which holds for extraction ratios based on Gini coefficients as well as top income shares. In the discussion, we also outline some directions of further research, especially about explanations of extraction patterns and dynamics, also with an application to global inequality data and in combination with institutional arrangements.

ISSN 1993-4378 (Print)
ISSN 1993-6885 (Online)