TI 2008-095/3
Tinbergen Institute Discussion Paper

The Bias of the Gini Coefficient due to Grouping: Revisiting First-order Corrections

Tom Van Ourti[1]
Philip Clarke[2]

[1] *Erasmus School of Economics, Erasmus University Rotterdam, and Tinbergen Institute;*
[2] *University of Sydney.*

Tinbergen Institute

The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam
Roetersstraat 31
1018 WB Amsterdam
The Netherlands
Tel.: +31(0)20 551 3500
Fax: +31(0)20 551 3555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Most TI discussion papers can be downloaded at http://www.tinbergen.nl.

THE BIAS OF THE GINI COEFFICIENT DUE TO GROUPING: REVISITING FIRST ORDER CORRECTIONS

Tom Van Ourti and Philip Clarke[1]

October 2008

Abstract—We propose a first order bias correction term for the Gini index to reduce the bias due to grouping. The first order correction term is obtained from studying the estimator of the Gini index within a measurement error framework. In addition, it reveals an intuitive formula for the remaining second order bias which is useful in empirical analyses. We analyze the empirical performance of our first order correction term using income data for 15 European countries and the US, and show that it reduces a considerable share of the bias due to grouping.

JEL-classification: C19, D31, I30

Keywords: Gini index, grouped data, measurement error, first-order correction

---

[1] Erasmus University Rotterdam, Tinbergen Institute, and University of Sydney respectively.
This research was partially supported from the NETSPAR project "Income, health and work across the life cycle". We thank EUROSTAT for access to the ECHP, and the Netherlands Central Bureau of Statistics for access to the linked datasets used for this research ("Own calculations of Erasmus University Rotterdam using data files made available by the Netherlands Central Bureau of Statistics on the regional income distribution of persons and households derived from the Tax Administration"). Tom Van Ourti is a Postdoctoral Fellow of the Netherlands Organisation for Scientific Research – Innovational Research Incentives Scheme – Veni. Dr. Clarke receives support from a Senior Research Fellowship from the University of Sydney. Part of this work was undertaken while he was a visitor of economics RSSS at Australian National University. We would like to thank Hans van Kippersluis for helpful comments and excellent research assistance. The study has benefited from the comments and suggestions of Teresa Bago d'Uva, Bob Breunig, George Deltas, John Einmahl, Andrew Leigh as well as participants at seminars given at Australian National University, Tilburg University and Erasmus University Rotterdam. The usual caveats apply and all remaining errors are our responsibility.

I. Introduction

The Gini index is the most commonly applied inequality measure in the literature, probably because of its link with Lorenz curves which give an intuitive and graphical representation of inequality. Its main application has been in the measurement of inequalities in income and wealth, but it has also a long history in other areas. For example, it has appeared as an inequality measure of health indicators (among others Le Grand, 1987, Pradhan *et al.*, 2003), educational attainment (among others Sheret, 1988, Lin, 2007), business concentration (among others Hart, 1971, Buzzacchi and Valletti, 2006), scientific publications and citations (among others Allison and Stewart, 1974), legislative malapportionment (Alker, 1965), astronomy (Abraham *et al.*, 2003), and many others.

The most frequently cited shortcomings of the Gini index are its violation of subgroup decomposability (see e.g. Bourguignon, 1979), and the bias due to data that is grouped by categories or into ranges (see e.g. Gastwirth, 1972).[1] The latter issue commonly arises with income or tax statistics that are grouped into deciles or quintiles for confidentiality reasons. Grouped data is also the main source of information on income distributions provided through the POVCALNET interactive computational tool of the World Bank (World Bank, 2008), and recent publications on regional and global inequality have also used grouped data (among others Sala-i-Martin, 2006). Previous empirical research suggests the grouping of income into relatively small number of categories imparts a non-negligible bias. For example, using the 1984 US Current Population Survey and the 1979-1980 Israeli Family Expenditure Survey, Lerman and

[1] Lambert and Aronson (1993) and Aronson *et al.* (1994) have successfully rationalized the violation of subgroup decomposability as a desirable property of the Gini index as a tool for analyzing the redistributive effect of an income tax.

Yitzhaki (1989) show that the bias from using grouped data with 10 and 5 income categories is about 2.5 and 7 percent of the Gini as calculated from micro data.

Several solutions have been proposed to cope with the dependence of the Gini index on the number of groups. First, a common approach is to reduce the bias due to grouping by estimating parametric functions that satisfy the properties of a theoretical Lorenz curve.[2] The estimated parameters are then used to estimate the Gini coefficient (among others Kakwani, 1980, Kakwani, 1986, Villaseñor and Arnold, 1989, Ryu and Slottje, 1996). This approach is popular among applied researchers (among others Datt and Ravallion, 1992, Bigsten and Shimeles, 2007) and has been implemented in the POVCAL software of the World Bank (2008). Despite its popularity, empirical uncertainty is the major disadvantage of this approach. Schader and Schmid (1994) show that most parametric functions give unreliable estimates of the Gini coefficient.

A second approach is to define non-parametric bounds on the Gini index (Gastwirth, 1972, Mehran, 1975, Murray, 1978, Fuller, 1979, Ogwang, 2003, Ogwang, 2006) which has the advantage that – compared to parametric functions – it does not make any assumption on the shape of the underlying Lorenz curve. These non-parametric bounds have been shown to outperform the approach using parametric functions (Schader and Schmid, 1994), but do require information on the lower and upper limit of each group.[3] The intuition is that the lower bound of the Gini corresponds to a situation where all individuals within a group have the same amount, while the upper bound reflects a situation where inequality is maximal in each of the groups.

[2] These are: twice differentiable, convex, monotonically increasing and passing through $(0,0)$ and $(1,1)$.

[3] The various methods mainly differ with respect to the information requirements of the overall and group-specific means.

In a recent study Deltas (2003) has attempted to address the related issue of small-sample bias, particularly in the context of business concentration. Here the bias arises not because of grouping, but is due to only having a few observations such as firms in an industry (Spiezia, 2003, Blyde, 2006, Li, 2006, Reynolds-Feighan, 2007). Deltas (2003) approach involves dividing the estimated Gini by its potential maximum to reduce the bias due to small samples which he denotes as a first-order correction term. The main advantage of this procedure is its relative simplicity and transparency in application. However, as the bias of the Gini is distribution specific, there might be a remaining (second-order) bias after applying this procedure. Despite the latter bias, his Monte Carlo simulations show that the procedure manages to reduce the bias in small samples.

In this paper, we develop a simple first-order correction term to deal with the bias of the Gini due to grouping by treating grouping as a form of measurement error. We first revisit the first-order bias correction term of Deltas (2003) that addresses small-sample bias, and show it worsens matters if applied to grouping. Our first-order correction term reduces the bias due to grouping considerably when applied to the income distributions of the 15 European countries and the US. It also provides an exact expression for the remaining second-order bias with an intuitive interpretation, and thus allows assessing the bias reduction of the first-order correction term for various shapes of the underlying distribution functions. An additional advantage is that it allows for groupings of unequal size.

The remainder of this paper contains four sections. We start by revisiting Deltas' (2003) first-order correction for the Gini. The next section derives and discusses our first-order correction. We then illustrate the bias due to income groupings of the Gini and the performance in terms of reducing the bias of our first-order correction in the third section. The final section contains the conclusions.

II. The first-order correction term of Deltas (2003) revisited

The Gini can be estimated using several equivalent formulas. For our purposes the following one is the most useful (Kakwani *et al*., 1997), i.e.

$$G_n = \frac{2\sum_{i=1}^{n} y_i R_i}{n\overline{y}} - 1 \qquad (1)$$

where y_i is the income of individual $i = 1,\dots,n$ with individuals ranked from poor to rich, i.e. $y_1 \le y_2 \le \cdots \le y_n$, $R_i = n^{-1}\left(i - 1/2\right)$ is the fractional income rank (Lerman and Yitzhaki, 1989), and $\overline{y} = n^{-1}\sum_{i=1}^{n} y_i$ denotes average income.[4] Equation (1) is a consistent, but downwardly biased estimator of the Gini index due to the convexity of the underlying Lorenz curves, but the magnitude of the bias is distribution specific (Lerman and Yitzhaki, 1989).

Deltas (2003) addresses this small-sample bias of the Gini by considering the inverse of the Gini in equation (1) that applies if all income would be concentrated among the richest person. It is well known that in finite samples the upper bound of the Gini is $n^{-1}\left(n-1\right)$. One way to correct this bias is to multiply the Gini by the inverse of

[4] We discuss the Gini of income, but obviously everything also holds for any variable which distribution is analyzed.

this maximum, i.e. $n(n-1)^{-1}$, so one obtains an upper bound of the Gini index equalling +1 which is independent of n.[5]

Deltas (2003) shows that his first-order correction term reduces the absolute value of the small-sample bias using a Monte Carlo simulation, but as the correction term only depends on n, a second-order bias remains, except for the case where the distribution of income is exponential. The magnitude of the remaining bias is increasing in the variance and kurtosis of the underlying distribution and is reducing in the skewness. The latter is in line with the Gini being mostly sensitive to transfers close to the mode as there are a high number of individuals between the transferring parties (Borooah, 1991).

III.A first-order correction term for grouping

In this section, we present a first-order correction term for the bias of the Gini due to grouping and discuss its properties. It derives from three steps. First, we compare equation (1) for n observations and for a situation where one constructs K groups from these n observations.[6] Second, we draw a parallel with the econometric literature on measurement error models (for example Cameron and Trivedi, 2005, chapter 26). Third, we let the number of observations approach infinity, while keeping the number of groups (and their relative size) fixed.

[5] The first-order correction term removes the small-sample bias of the absolute mean difference. Deltas (2003) shows that this unbiasedness of the absolute mean difference does not translate to the Gini since the Gini is a non-linear combination of the absolute mean difference and the mean.

[6] Note the similarity with the difference between the OLS and between estimator for panel models (Cameron and Trivedi, 2005, chapter 21).

Kakwani *et al.* (1997) have shown that the Gini index in equation (1) can also be calculated as the OLS estimate of β , i.e.

$$2\sigma_R^2 \frac{y_i}{\bar{y}} = \alpha + \beta R_i + \varepsilon_i \qquad (2)$$

where $\sigma_R^2 = \left(n^2 - 1\right)\left(12n^2\right)^{-1}$ is the variance of R_i, ε_i is an error term with zero mean and α, β are parameters.[7] It is important to note that the equality between equation (1) and (2) holds under the properties of OLS as arithmetic tool, and that no additional assumptions must be made.[8]

A. Groups of equal size

In order to understand the bias of the Gini that results from grouping n observations into K groups of equal size[9], it is helpful to see that equation (2) reduces to

$$2\sigma_{R^K}^2 \frac{y_g}{\bar{y}} = \alpha^K + \beta^K R_g + \varepsilon_g \qquad (3)$$

where we have added 'K'-superscripts to refer to the grouped data case, $R_g = K\left(n\right)^{-1}\left(g - 1/2\right)$ is the fractional income rank of group $g = 1,\ldots,K$, $\sigma_{R^K}^2 = \left(K^2 - 1\right)\left(12K^2\right)^{-1}$ is the variance of R_g, and y_g is the average income within group g . The OLS estimate of β^K equals the Gini index calculated from the K groups *and* is smaller (if there is income variation within at least one of the K groups) or equal

[7] Consult appendix A for a derivation of the variance of the fractional rank.
[8] Consult appendix B for a derivation of the equality between equation (1) and (2).
[9] For ease of exposition, we first derive the first-order correction term for equally sized groups, i.e. n/K observations per group. The intuition and derivation is similar for unequally sized groups and is shortly discussed in the next section.

(if there is no income variation in each of the K groups) to the estimate of the Gini based on n observations, i.e.

$$\beta^K = G_n^K = \frac{2\sum_{g=1}^{K} y_g R_g}{K\bar{y}} - 1 \leq G_n = \beta \tag{4}$$

The goal of the remainder of this section is to establish an exact relationship between G_n and G_n^K using equations (2) and (3). Comparing the latter equations reveals that both RHS and LHS differ. The difference in the RHS can be interpreted as a measurement error problem, i.e. we observe the rank of income at the level of the groups rather than one at the level of the n observations. More exactly, let's start from equation (2) and add an equation that describes the measurement error

$$R_i^g = R_i + \delta_i^g \tag{5}$$

where δ_i^g is the measurement error with zero mean and R_i^g is the fractional income rank of group g *defined at the individual level*, i.e. every individual in group g gets the fractional income rank of group g, i.e. R_g. Assuming that we do observe the actual income level y_i but not the actual fractional income rank R_i, i.e. substituting equation (5) into equation (2), gives

$$2\sigma_R^2 \frac{y_i}{\bar{y}} = \alpha + \beta R_i^g + \left(\varepsilon_i - \beta\delta_i^g\right) \tag{6}$$

It is impossible to estimate β from equation (6) using OLS (as an arithmetic tool) since we do not observe $\left(\varepsilon_i - \beta\delta_i^g\right)$. Instead, we can only estimate

$$2\sigma_R^2 \frac{y_i}{\bar{y}} = \alpha^{MER} + \beta^{MER} R_i^g + \eta_i \tag{7}$$

where η_i is a zero mean error term, and the superscript '*MER*' refers to measurement error. Using some algebra, exploiting the fact δ_i^g and R_i^g are uncorrelated[10], the fact that ε_i and R_i are uncorrelated (which holds due the using OLS as an arithmetic tool only), it is easy to show that the OLS estimate of β^{MER} in equation (7) and the OLS estimate of $\beta = G_n$ in equation (2) are related[11]

$$\beta^{MER} = G_n + \frac{\frac{1}{n}\sum_{i=1}^{n}\delta_i^g \varepsilon_i}{\sigma_{R^K}^2} \tag{8}$$

In order to derive an expression relating G_n and G_n^K, we need to establish one additional relationship that addresses the difference between the LHS of equations (2) and (3). After some algebra, one can establish that[12]

$$\beta^{MER} = G_n^K \left(\frac{\sigma_R^2}{\sigma_{R^K}^2}\right) = G_n^K \left[\frac{K^2\left(n^2-1\right)}{n^2\left(K^2-1\right)}\right] \tag{9}$$

which shows that β^{MER} is related to G_n^K by the ratio of the variances of the actual fractional income rank and that of the fractional income rank of group g – $K^2\left(n^2-1\right)\left[n^2\left(K^2-1\right)\right]^{-1}$.

Combining equation (8) and (9), allows us to come up with a useful equation that expresses the Gini estimated from n observations as a function of – among others – the Gini estimated from a grouping of these n observations, i.e.

[10] δ_i^g and R_i^g are uncorrelated since R_i^g equals the average R_i of group g, i.e. $\sum_{i\in g}\delta_i^g R_i^g = \sum_{i\in g}\left(R_i^g - R_i\right)R_i^g = 0$, and hence $\sum_{i=1}^{n}\delta_i^g R_i^g = \sum_{g=1}^{K}\left(\sum_{i\in g}\delta_i^g R_i^g\right) = 0$.

[11] Consult appendix C for a full derivation.

[12] Consult appendix D for a full derivation.

$$G_n = G_n^K\left(\frac{\sigma_R^2}{\sigma_{R^K}^2}\right) - \frac{\frac{1}{n}\sum_{i=1}^{n}\delta_i^g\varepsilon_i}{\sigma_{R^K}^2} = G_n^K\left[\frac{K^2\left(n^2-1\right)}{n^2\left(K^2-1\right)}\right] - \left[\frac{12K^2}{K^2-1}\right]\left[\frac{1}{n}\sum_{i=1}^{n}\delta_i^g\varepsilon_i\right] \qquad (10)$$

Assuming that $n \to +\infty$ and $K < +\infty$ (i.e. the number of groups in the population and their relative size is fixed) results in $G_\infty = \frac{K^2}{K^2-1}\left[G_\infty^K - 12\operatorname{cov}\left(\delta_i^g, \varepsilon_i\right)\right]$. Equation (10) reveals some interesting insights. First, we have only used the properties of OLS as an *arithmetic* tool and the properties of the fractional rank to come up with equation (10). Second, equation (10) provides a first-order correction term and an expression for the remaining second-order bias. The first-order correction $\left(K^2-1\right)^{-1}K^2$ resembles Deltas' term, but is smaller and has two intuitive interpretations, i.e. it equals a "grouped data" adjustment of the variance of the fractional rank, and it is also related to the inverse of the covariance between the actual fractional rank at the individual level and that of group g, i.e. $\left(K^2-1\right)^{-1}K^2 = \left[12\operatorname{cov}\left(R_i^g, R_i\right)\right]^{-1}$.[13] The latter interpretation is intuitive as a high covariance between the grouped and actual fractional ranks implies a low first-order correction term. The second order bias $-12\operatorname{cov}\left(\delta_i^g, \varepsilon_i\right)K^2\left(K^2-1\right)^{-1}$ also has an intuitive interpretation as it is a function of the covariance between the measurement error and the error term from equation (2). A few things can be said about this covariance. First, although one can always observe δ_i^g, ε_i is unobservable and thus the value and sign of this covariance is always unknown. Nevertheless, if one has an idea on the shape of the distribution function of y_i, it is straightforward to get an idea on its sign and magnitude. For example the second order bias is zero for a uniform distribution as the variance of ε_i is equal to zero. Second, the covariance will be smaller the higher

[13] More information on the derivation can be found in appendix E.

the number of groups K, which is easily inferred from the equality $\operatorname{cov}\left(\delta_i^g,\varepsilon_i\right)=\operatorname{cov}\left(R_i^g,\varepsilon_i\right)$ (see also appendix E). Third, although the sign of the covariance cannot be predicted a priori, it is likely to be negative (i.e. implying an undercorrection after applying the first-order correction term) for an asymmetric unimodal distribution (i.e. left or right skewed). For example, an extreme long right tail is likely to result in a negative covariance (see equation (2) and (5)).

B. Groups of unequal size

Until now we have assumed that the K groups are equally sized. Equation (10) is however easily generalised to groups of unequal size. Assume that n_u is the number of observations in group $u=1,\ldots,K$ (with u referring to 'unequal group size'), that $R_u=(n)^{-1}\left(1/2\,n_u+\sum_{j=1}^{u-1}n_j\right)$ equals the fractional income rank of group u, and that the variance of the latter is defined as $\sigma^2_{R_u^K}=(n)^{-1}\sum_{u=1}^{K}n_u\left(R_u-1/2\right)^2$. We have now sufficient information to derive the equivalent expressions of equation (3) and (4):

$$2\sigma^2_{R_u^K}\frac{y_u}{\overline{y}}\sqrt{n_u}=\alpha^u\sqrt{n_u}+\beta^u R_u\sqrt{n_u}+\varepsilon_u\sqrt{n_u} \tag{11}$$

$$G_n^{K,u}=\beta^u=\frac{2\sum_{u=1}^{K}n_u y_u R_u}{n\overline{y}}-1=\frac{2\sum_{u=1}^{K}\left[\left(\frac{K}{n}n_u\right)y_u R_u\right]}{K\overline{y}}-1\leq G_n \tag{12}$$

Equation (11) is a Weighted Least Squares (WLS) generalisation of equation (3), and equation (12) reduces to equation (4) if all groups have equal size. The relationship between $G_n^{K,u}$ and G_n is established by combining equation (2) with an 'unequal size' generalization of equation (5)

$$R_i^u = R_i + \delta_i^u \quad (13)$$

where δ_i^u is the measurement error with zero mean and R_i^u is the fractional income rank of group u *defined at the individual level*. This results in[14]

$$G_n = \frac{\sigma_R^2}{\sigma_{R_u^K}^2} G_n^{K,u} - \frac{\frac{1}{n}\sum_{i=1}^{n} \delta_i^u \varepsilon_i}{\sigma_{R_u^K}^2} \quad (14)$$

It is straightforward to see that equation (11) and (14) are identical, except for the unequal group sizes. The first order correction term still measures the ratio of the variance of the actual fractional rank and that of the fractional rank of group u *and* is easy to calculate, and we still obtain an expression of the second-order bias with the covariance interpretation.

IV. Empirical illustration

A. Data

In this section, we illustrate the dependence of the Gini index of *income* on the number of groups, and show the performance of our first-order correction term in reducing the bias if applied to *income* distributions. First, we analyzed this bias for the Netherlands using administrative data on more than five million individual income tax files for 2004. The advantage of administrative data is that it allows us to compare the Gini indices obtained from income groupings with the one obtained from this *population*. Unfortunately, it was not possible to obtain administrative data for other countries. Instead, we used European microdata from the European Community

[14] Consult appendix F (and appendix C) for a full derivation.

Household Panel (ECHP) and US microdata from the Medical Expenditure Panel Survey (MEPS). As we report below, the findings based on these microdata are very much in line with those resulting from the Dutch administrative data.

We have not resorted to Monte Carlo simulations since one might draw empirically irrelevant inferences from these. As discussed in the introduction, the approach using parametric functions to reduce the bias from grouping suffers from empirical uncertainty. This suggests that Monte Carlo simulations using parametric cumulative distribution functions will be of limited value in understanding the performance of our first-order correction term if applied to actual *income* distributions. While the results from the ECHP and the MEPS do not allow us to compare with the Gini in the *population* in the respective countries, it nevertheless gives useful information on the variability of the first order correction as it compares the reduction of the underestimation across a set of 16 countries.

The ECHP was designed and coordinated by EUROSTAT. It contains socioeconomic information for individuals aged 16 or older, uses a standardised questionnaire, and covers 15 EU member states: Austria (AT), Belgium (BE), Denmark (DK), Finland (FI), France (FR), Germany (DE), Greece (GR), Ireland (IRL), Italy (IT), Luxembourg (LU), Netherlands (NL), Portugal (PT), Spain (ES), Sweden (SE) and the United Kingdom (UK). We use the first wave for all countries, i.e. the 1994 wave, except for Austria that joined the survey in 1995, Finland that joined in 1996, and Sweden that joined in 1997. We supplement this with US income microdata from the 2001 wave of MEPS. We use the first wave of the ECHP as it does not suffer from attrition, and thus has more observations which is useful for illustrating the first-order correction term and the dependence of the Gini upon the number of income groups.

Note that all calculations in this section only serve the purpose of illustrating the methods explained in the previous sections, and not to deliver any hard evidence on income inequality in the EU and US.

The key variable for this study is income. The Dutch income tax files provide annual equivalent disposable household income (where the equivalence-factor gives weight 1 to the head of the household, each following household member over 18 receives weight 0.38, while household members under 18 receive a weight – depending on their age and birth order – between 0.19 and 0.30). The ECHP income measure is annual disposable (i.e. after-tax) household income, which is all net monetary income received by the household members during the previous year. It includes income from work (employment and self-employment), private income (from investments and property and private transfers to the household), pensions and other direct social transfers received. No account has been taken of indirect social transfers (e.g. reimbursement of medical expenses), receipts in kind and imputed rent from owner-occupied accommodation. The MEPS income measure is similarly defined.[15] We measure all incomes in national currencies.[16] The income variable was further divided by the OECD modified equivalence scale in order to account for household size and composition (giving a weight of 1.0 to the first adult, 0.5 to the second and each subsequent person aged 14 and over, and 0.3 to each child aged under 4 in the household). Table A.1 in appendix G reports descriptive statistics of equivalent income

[15] Note that two individuals in the MEPS data report negative incomes. We have recoded these negative into zero income values (see also Chen *et al.*, 1982).

[16] We did not take the trouble to convert the national currencies into a common currency and did neither deflate to correct for inflation as the Gini index is a relative inequality measure that is invariant to proportional income changes. The national exchange rates (national currency=1 euro) for 1994 were 39.66 Belgian francs, 7.54 Danish krone, 6.58 French francs, 1.92 German mark, 288.03 Greek drachma, 0.79 Irish pound, 1915.06 Italian lire, 39.66 Luxembourgian francs, 2.16 Dutch guilders, 196.90 Portuguese escudo, 158.92 Spanish peseta, 0.78 British pounds. The 1995 exchange rate was 13.18 Austrian schilling, the 1996 one was 5.83 Finnish markka, and the 1997 one was 8.65 Swedish krona (EUROSTAT, 2003). The 2001 US exchange rate was 0.90 US dollars (OECD, 2008).

in each of the countries. As we are analyzing the behaviour of estimates of the Gini index for varying grouping sizes, it is reassuring to note that all samples are sufficiently large (at least 5500 observations, except for Luxembourg that has about 2000 observations).

The analysis takes three steps. First, we calculate the Gini index based on the Dutch administrative data and the ECHP and MEPS datasets. Second, we create income categories from the full samples; and analyze the effect that follows from these groupings. Third, we illustrate the performance of our first-order correction term in terms of reducing the underestimation.[17]

B. Full sample Gini indices

Figure 1 presents the estimates of the Gini indices based on the Dutch administrative data and the full samples of the ECHP and the MEPS where we have ranked countries from low to high relative income inequality. The point estimates are obtained using equation (2) and the confidence intervals for the microdata result from a bootstrap procedure (see e.g. Mills and Zandvakili, 1997).[18,19] Figure 1 illustrates that the 15 European countries and the US differ widely in terms of relative income inequality. The Gini index of the three Scandinavian countries – that have the lowest

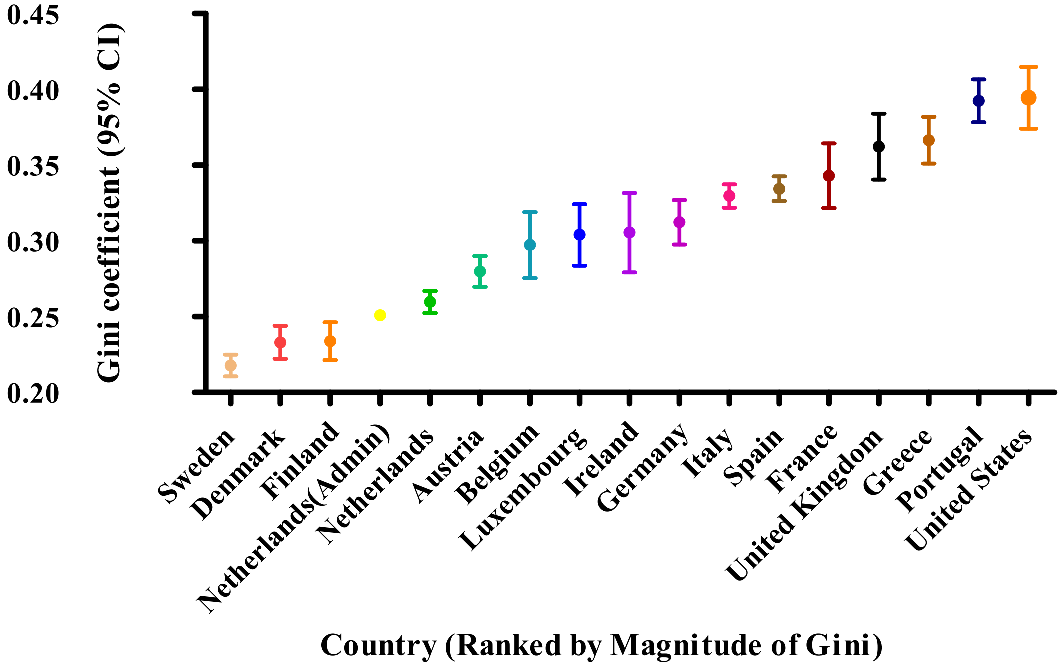
[17] We have replicated each of the three steps for random subsamples of the MEPS 2001 (i.e. 90, 80, 70,…, 10 percent of the sample size) to check whether sample size might affect our findings. As expected, we confirmed all conclusions based on the full sample. The only difference was plausible, i.e. a reduction of the statistical precision.

[18] We use a fractional rank that accommodates individuals with identical equivalent incomes, i.e. $R_i = n^{-1}\left\{q(y_i) + 0.5\left[q'(y_i) - q(y_i)\right]\right\}$ where $q(y_i) = \sum_{k=1}^{n} 1(y_k < y_i)$ and $q'(y_i) = \sum_{k=1}^{n} 1(y_k \le y_i)$.

[19] We draw 1000 bootstrap samples on the level of the fractional income rank – rather than on the level of the individual – to account for individuals with identical equivalent income levels. We adjust the standard bootstrap sampling procedure such that the probability to draw a fractional income rank is inversely related to the number of individuals with the corresponding equivalent income level. From the resulting 1000 bootstrap samples, we compute standard errors and confidence intervals.

inequality – is almost half of that in Portugal and the US. It is also the case that the sampling variability differs widely across countries, but this is only partially explained by differences in sample size (compare e.g. France and Spain).

FIGURE 1. – GINI INDICES IN THE EU AND US IN SELECTED YEARS: POINT ESTIMATES AND 95 PERCENT CONFIDENCE INTERVALS



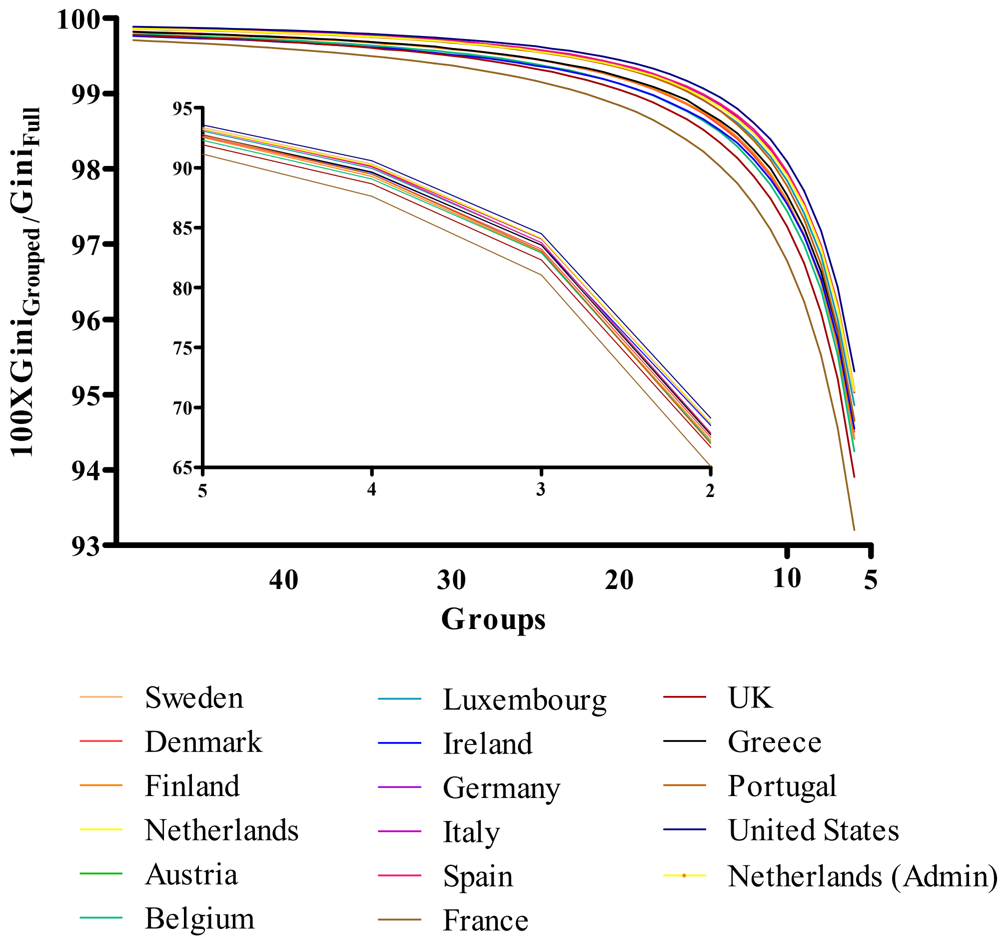
Source: Netherlands (admin) refers to authors' calculations from linked Dutch administrative data 2004

C. Gini index and the number of income groupings

The estimates presented in Figure 1 are in this study considered as the benchmark estimates against which the effect of grouping the data and the performance of our first-order correction in terms of reducing the underestimation are evaluated where we have explicitly included the Dutch administrative data to have also results based on a very high number of observations, i.e. more than 5 million observations. First, we subdivide the full sample into 50 equally sized (equivalent) income categories. Second, we calculate average equivalent income for each income category. Finally, we calculate the Gini index from these average equivalent incomes using equation (3). This three step procedure is repeated for 49 to 2 income categories, and the resulting Gini indices are expressed as a proportion of the benchmark Gini's estimated from the full sample, i.e. $100\times\left(G_n^K/G_n\right)$ with $K=50,\ldots,2$. The resulting proportions are presented in figure 2. In table A.2 in appendix G, we also present the point estimates and the standard errors of the Gini indices calculated from the income groupings in the ECHP and MEPS.[20]

[20] The standard errors are obtained using the same bootstrap procedure, with the difference that we have drawn income groups, rather than fractional income ranks.

FIGURE 2. – THE GINI AND ITS DEPENDENCE ON THE NUMBER OF INCOME GROUPINGS IN THE EU AND US



Source: Netherlands (admin) refers to authors' calculations from linked Dutch administrative data 2004

Figure 2 reveals several interesting insights. First, due to the convexity of Lorenz curves, the Gini index based on grouped data always underestimates the one in the full sample, i.e. all lines lie underneath 100. Second, the underestimation – expressed as $100\times\left[1-\left(G_n^K/G_n\right)\right]$ – is similar across countries. The largest horizontal difference between the lines in figure 2 is observed at 2 income groups, i.e. US has the lowest underestimation of 30.88 percent and France has the highest underestimation of

34.91 percent. The range of the underestimation (about 4% percent) seems low given the much higher value of the underestimation itself. The cross-country similarity of the underestimation suggests that the shape of the underlying distribution functions is similar across countries, but that the spread differs (otherwise the Gini would take a similar value in all countries). In addition, it shows that there is scope for improving cross-country inequality comparisons using the first-order correction terms if there cross-country differences in the number of income categories. Third, the underestimation of the Gini index from grouping the data increases at an increasing pace when lowering the number of income categories. It seems that most of the action is taking place for 20 or less income groups. In the extreme case of 2 income groups, the Gini index based on grouped income data is only between 65 and 70 percent of the one based on the full sample. For 5 income groups, the underestimation is between 9 and 6 percent, and for 10 income groups, the underestimation still amounts to about 2 to 3 percent. These percentages do represent important underestimations. In order to get an intuitive feeling for their magnitude, it is worthwhile to make some comparisons. Consulting table A.2 in appendix G to compare with the sampling variability of the Gini index in the full sample shows that the underestimation is not negligible. Comparing the evolution of the Gini over time in the full sample is a second benchmark. For all countries in the ECHP, we have calculated the proportional change in the Gini between the first available and last wave using a balanced panel, and calculated the underestimation that results from grouping the data in the first wave of the balanced panel. We find that in *all* countries, the proportional change in the Gini over time (8 years for most countries) is smaller than the underestimation resulting from 5 income groups. A final comparison to grasp the importance of the underestimation from income groupings, is to consider the impact of income grouping in one country on the income inequality ranking of countries in Figure 1. This is illustrated in Table 2 in section F, and again confirms the importance of the underestimation (see below for additional discussion). Finally, given the similarity of the bias in the Dutch administrative data and the underestimation in the ECHP and the MEPS, and given that the first-order-correction does not depend upon the income distribution, we stick to the latter microdatasets in the remainder of the analysis.

D. Determinants of the underestimation from income groupings

We analyze in more detail some potential determinants of the magnitude of the underestimation using pooled regression on 784 observations, i.e. 49 income groupings for 16 countries. Our baseline regression model results from rearranging and dividing equation (10) by G_n and assuming that $n \to +\infty$ and $K < +\infty$. The resulting model, i.e. $\frac{G_\infty^K}{G_\infty} = \frac{K^2-1}{K^2} + \eta_K \frac{12}{G_\infty}$ with $\operatorname{cov}\left(\delta_i^g, \varepsilon_i\right) = \eta_K$ and $K = 2, \ldots, 50$, can be estimated with OLS by excluding a constant. In other words, this model estimates the 49 covariance terms using between-country variation. We find that the latter regression fits the data well (i.e. the uncentered and standard R^2 equal 1.0006 and 0.9817 respectively). All 49 covariance terms are negative, their value increases monotonically with the number of income groupings, and they are precisely estimated.[21] This shows two things. First, the combination of the good fit of the model with the low negative covariance terms shows that our first-order correction term is likely to reduce the underestimation considerably

[21] For example, the covariance equals -0.00183 for two income groupings, and -0.00003 for 50 groups. The Huber-White standard errors are 0.00013 and 3.32e-06 respectively.

for income distributions. Second, applying our first order correction term will – for the income data of these 16 countries – underestimate the Gini resulting from the full sample, and the lower the number of income categories the higher the underestimation.

TABLE 1. – DETERMINANTS OF THE UNDERESTIMATION IN THE ECHP AND MEPS

	G_n^K/G_n	G_n^K/G_n	G_n^K/G_n	G_n^K/G_n	$G_n - G_n^K$	$G_n - G_n^K$	$G_n - G_n^K$	$G_n - G_n^K$
G_n		-0,0556*		0,0342*		0,7316*		-0,0041
var			-0,0057*	-0,0114*			0,0024*	0,0031*
skew			0,0003	0,0012+			0,0001	0,0000
kurt			-0,0002	-0,0010			-0,0007	-0,0006
AT	0,0017*	0,0003			-0,0011+	0,0178*		
BE	-0,0007+	-0,0012*			0,0000	0,0061*		
DK	0,0002	-0,0038*			-0,0017+	0,0513*		
FI	0,0000	-0,0032*			0,0120*	0,0545*		
FR	-0,0047*	-0,0026+			0,0025*	-0,0251*		
DE	-0,0011+	-0,0007			0,0005+	-0,0045*		
GR	0,0006*	0,0040*			0,0011*	-0,0435*		
IT	0,0024*	0,0038*			-0,0003	-0,0180*		
LU	0,0017*	0,0017*			-0,0006§	0,0005*		
NL	0,0023*	-0,0003			-0,0016+	0,0319*		
PT	0,0011*	0,0060*			0,0015+	-0,0621*		
ES	0,0022*	0,0038*			-0,0001	-0,0213*		
SE	0,0024*	-0,0025			-0,0025*	0,0617*		
UK	-0,0018*	0,0013			0,0019*	-0,0396*		
US	0,0033*	0,0083*			0,0007	-0,0644*		
K=2	-0,3223*	-0,3220*	-0,3223*	-0,3225*	0,1052*	0,1015*	0,1052*	0,1052*
K=3	-0,1654*	-0,1652*	-0,1654*	-0,1656*	0,0550*	0,0524*	0,0550*	0,0550*
K=4	-0,1029*	-0,1028*	-0,1029*	-0,1030*	0,0348*	0,0328*	0,0348*	0,0348*
K=5	-0,0711*	-0,0710*	-0,0711*	-0,0712*	0,0244*	0,0227*	0,0244*	0,0244*
K=6	-0,0524*	-0,0523*	-0,0524*	-0,0525*	0,0183*	0,0168*	0,0183*	0,0183*
K=7	-0,0404*	-0,0403*	-0,0404*	-0,0405*	0,0143*	0,0130*	0,0143*	0,0143*
K=8	-0,0321*	-0,0320*	-0,0321*	-0,0322*	0,0115*	0,0104*	0,0115*	0,0115*
K=9	-0,0262*	-0,0261*	-0,0262*	-0,0263*	0,0095*	0,0085*	0,0095*	0,0095*
K=10	-0,0218*	-0,0217*	-0,0218*	-0,0218*	0,0080*	0,0071*	0,0080*	0,0080*
K=11	-0,0184*	-0,0183*	-0,0184*	-0,0184*	0,0069*	0,0060*	0,0069*	0,0069*
K=12	-0,0158*	-0,0157*	-0,0158	-0,0158*	0,0059*	0,0052*	0,0059*	0,0060*
K=13	-0,0136*	-0,0136*	-0,0136*	-0,0137*	0,0052*	0,0045*	0,0052*	0,0052*
K=14	-0,0119*	-0,0118*	-0,0119*	-0,0119*	0,0046*	0,0039*	0,0046*	0,0046*
K=15	-0,0104*	-0,0104*	-0,0104*	-0,0105*	0,0041*	0,0035*	0,0041*	0,0041*
K=16	-0,0092*	-0,0092*	-0,0092*	-0,0092*	0,0037*	0,0031*	0,0037*	0,0037*
K=17	-0,0082*	-0,0081*	-0,0082*	-0,0082*	0,0033*	0,0028*	0,0033*	0,0033*
K=18	-0,0073*	-0,0073*	-0,0073*	-0,0073*	0,0030*	0,0025*	0,0030*	0,0030*
K=19	-0,0065*	-0,0065*	-0,0065*	-0,0066*	0,0027*	0,0022*	0,0027*	0,0027*
K=20	-0,0059*	-0,0058*	-0,0059*	-0,0059*	0,0025*	0,0020*	0,0025*	0,0025*
K=21	-0,0053*	-0,0053*	-0,0053*	-0,0053*	0,0023*	0,0018*	0,0023*	0,0023*
K=22	-0,0048*	-0,0048*	-0,0048*	-0,0048*	0,0021+	0,0017*	0,0021*	0,0021*
K=23	-0,0043*	-0,0043*	-0,0043*	-0,0043*	0,0019+	0,0015*	0,0019*	0,0019*
K=24	-0,0039*	-0,0039*	-0,0039*	-0,0039*	0,0018+	0,0014*	0,0018*	0,0018*
K=25	-0,0036*	-0,0035*	-0,0036*	-0,0036*	0,0016§	0,0013*	0,0016*	0,0016*
K=26	-0,0032*	-0,0032*	-0,0032*	-0,0033*	0,0015§	0,0012*	0,0015*	0,0015*
K=27	-0,0029*	-0,0029*	-0,0029*	-0,0030*	0,0014	0,0011+	0,0014*	0,0014*
K=28	-0,0027*	-0,0027*	-0,0027*	-0,0027*	0,0013	0,0010+	0,0013+	0,0013*
K=29	-0,0024*	-0,0024*	-0,0024*	-0,0025*	0,0012	0,0009+	0,0012+	0,0012+
K=30	-0,0022*	-0,0022*	-0,0022*	-0,0022*	0,0011	0,0008§	0,0011+	0,0011+
K=31	-0,0020*	-0,0020*	-0,0020*	-0,0020*	0,0011	0,0007§	0,0011+	0,0011+
K=32	-0,0018*	-0,0018*	-0,0018*	-0,0018*	0,0010	0,0007	0,0010+	0,0010+
K=33	-0,0016*	-0,0016*	-0,0016*	-0,0017*	0,0009	0,0006	0,0009+	0,0009+
K=34	-0,0015*	-0,0015*	-0,0015*	-0,0015*	0,0009	0,0006	0,0009§	0,0009+
K=35	-0,0013*	-0,0013*	-0,0013*	-0,0014*	0,0008	0,0005	0,0008§	0,0008§
K=36	-0,0012+	-0,0012+	-0,0012*	-0,0012*	0,0007	0,0005	0,0007§	0,0008§
K=37	-0,0011+	-0,0011+	-0,0011*	-0,0011+	0,0007	0,0004	0,0007§	0,0007§
K=38	-0,0010§	-0,0009§	-0,0010+	-0,0010+	0,0006	0,0004	0,0006	0,0006
K=39	-0,0008	-0,0008	-0,0008+	-0,0009§	0,0006	0,0004	0,0006	0,0006
K=40	-0,0007	-0,0007	-0,0007§	-0,0007	0,0006	0,0003	0,0006	0,0006
K=41	-0,0006	-0,0006	-0,0006	-0,0006	0,0005	0,0003	0,0005	0,0005
K=42	-0,0006	-0,0005	-0,0006	-0,0006	0,0005	0,0003	0,0005	0,0005
K=43	-0,0005	-0,0004	-0,0005	-0,0005	0,0005	0,0002	0,0005	0,0005
K=44	-0,0004	-0,0004	-0,0004	-0,0004	0,0004	0,0002	0,0004	0,0004
K=45	-0,0003	-0,0003	-0,0003	-0,0003	0,0004	0,0002	0,0004	0,0004
K=46	-0,0002	-0,0002	-0,0002	-0,0003	0,0004	0,0002	0,0004	0,0004
K=47	-0,0002	-0,0002	-0,0002	-0,0002	0,0003	0,0001	0,0003	0,0003
K=48	-0,0001	-0,0001	-0,0001	-0,0001	0,0003	0,0001	0,0003	0,0003
K=49	-0,0001	0,0000	-0,0001	-0,0001	0,0003	0,0001	0,0003	0,0003
cste	0,9977*	1,0147*	1,0015*	0,9939*	-0,0003	-0,2232*	-0,0007	0,0002
n	784	784	784	784	784	784	784	784
R^2	0,9983	0,9983	0,9976	0,9981	0,9544	0,9752	0,9241	0,9242
country	0,0000	0,0000			0,0000	0,0000		
skew/kurt			0,0142	0,0000			0,0000	0,0000

Note: G_n^K : Gini index estimated from K income groupings, G_n : Gini index estimated from the full sample, var: variance divided by the squared mean of equivalent income in the full sample, skew: skewness divided by the cube of the mean of equivalent income in the full sample, kurt: kurtosis divided by the fourth power of the mean of equivalent income in the full sample, country: p-value of a test on joint significance of the country dummies (excluded category is Ireland), skew/kurt: p-value of test on joint significance of skew and kurt, *: significant at 1% level, +: significant at 5% level; §: significant at 10% level, in all occasions we used the Huber-White covariance matrix

Besides the above regression model that naturally results from section III, we also report results from other regression models to analyse some potential determinants of the underestimation in table 1. The models in the 4 left columns use a relative indicator of the underestimation whereas the 4 right columns use an absolute indicator.[22] We compute Huber-White standard errors and consider four sets of explanatory variables. First, we include dummies for the number of income groups (50 income groups is the excluded category). Second, we include country dummies (Ireland is the excluded category).[23] Third, we include the value of the Gini index calculated from the full sample. Finally, we included scale-free summary measures of the shape of the underlying income distribution *in the full sample*, much along the lines of Deltas (2003). We included the normalized variance (var) – i.e. divided by the square of mean equivalent income –, the normalized skewness (skew) – i.e. divided by the cube of the mean –, and the normalized kurtosis (kurt) – i.e. divided by the fourth power of the mean. Mean equivalent income was *not* included as (a) the mean was used to normalize the other summary measures of the income distribution, (b) as it is expressed in different currencies, and (c) since the Gini index is a relative inequality measure.

We draw 4 lessons from the estimates in table 1. First, we prefer the regressions in the left columns since the treatment in section III naturally leads to a proportional presentation of the underestimation, but also since the R^2's show that it is more difficult to explain the underestimation expressed as an absolute difference. Second, the dummies for the number of income groups explain the majority of the underestimation. This is easily seen from a comparison between figure 2 and the estimates in the first

[22] We use both relative and absolute indicators to provide a more complete understanding of the underestimation, but also since the relative indicators seem more appropriate for a within country analysis and the absolute differences for between country analyses.

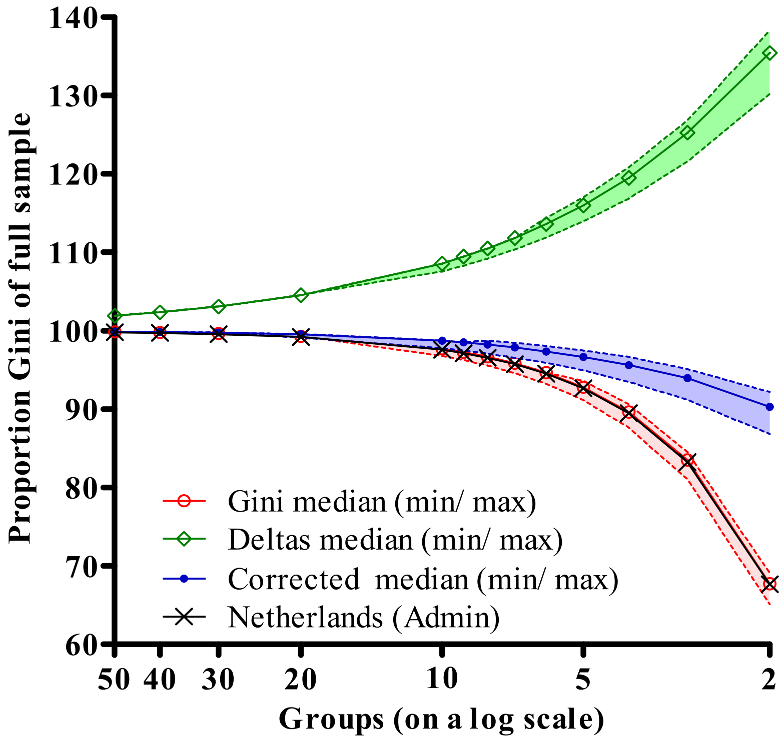
[23] There are insufficient degrees of freedom to check the relevance of interactions between the country dummies and the number of income groups.

column of table 1, but also from observing that the estimates are hardly influenced by the inclusion of other explanatory variables. Third, the Gini calculated from the full sample tends to increase the underestimation (see 2nd and 6th column). Fourth, the 'average' differences between countries are small, but nevertheless jointly significant as can be seen from the 'country' row. Therefore, we try to explain what features of the income distribution might be driving these country differences. We exclude the country dummies and include the three summary measures of the underlying (full sample) income distribution. We find that all three measures are jointly significant (see row skew/kurt), but only the estimate of the variance is individually significant showing that it has a similar effect as the Gini calculated from the full sample, which seems plausible as both are dispersion measures. This is also in line with our earlier observation that the cross-country similarity of the underestimation suggests that the shape of the underlying distribution functions is similar across countries, but that the spread differs.

E. Reduction of underestimation after first order correction

This section discusses the performance of our first-order correction term as applied to income distributions. The results are presented in figure 3 – which has a similar setup as figure 2 – and some more detailed results are available in table A.3 in appendix G. The lines with unfilled circles represent the median value for the Gini as a proportion of the Gini based on the full sample and the shaded region is the area between the minimum and maximum values across all countries presented in figure 2. The lines with filled circles give the remaining underestimation after applying our first-order correction term based on equation (10). The figure also contains Deltas (2003) first-order correction to illustrate the consequences of applying a small-sample bias correction method as a method of adjusting the bias that arises from grouping. The remaining underestimation is represented with unfilled diamonds.

FIGURE 3. – GINI AND ITS DEPENDENCE ON THE NUMBER OF INCOME GROUPINGS: FIRST ORDER CORRECTION TERMS



Note: Gini min, median, max: the minimum, median and maximum value (across countries) of the Gini estimated from grouped income data as a proportion of the Gini index calculated from the full sample; Deltas min, median, max: the minimum, median and maximum value (across countries) of the Gini index estimated from grouped data after applying Deltas' first-order correction term, i.e. $K^{-1}(K-1)$, as a proportion of the Gini index calculated from the full sample; Corr min, median, max: the minimum, median and maximum value (across countries) of the Gini index estimated from grouped data after applying our first-order correction term resulting from equation (10), i.e. $(K^2-1)^{-1}K^2$; as a proportion of the Gini index calculated from the full sample.
Source: Netherlands (admin) refers to authors' calculations from linked Dutch administrative data 2004.

A first thing to note is that our first-order correction term reduces the underestimation in each of the 16 countries. This is evident in figure 3, but can in more detail be inferred from table A.3 in appendix G. Second observation is that application of our first-order correction term never results in an overestimation of the Gini index. Although the sign of the remaining underestimation cannot be signed a priori, its magnitude is similar across the 16 countries, as can be inferred from the estimates of the covariance term in equation (10) (consult column 'covar' in table A.3). Third observation is that our first-order correction term removes more than half of the underestimation, but this obviously implies a higher remaining percentage point underestimation at a low number of income groups (consult the increasing value of 'covar' in table A.3). Fourth observation is that applying Deltas' first-order correction term always results in an overestimation that is larger (in absolute value) compared to the original underestimation resulting from applying the Gini to grouped income data. This finding shows that Deltas' correction should not generally be used to correcting bias that arises due to groupings of income. The same advice applies to our correction term, if it is applied to small-sample bias.[24] A final interesting observation is that for $n \leq 6$ the maximum underestimation after applying our first-order correction term is always smaller than the minimum original underestimation. The latter suggests – and we tend to believe that the comparison between minimum and maximum is an extremely conservative test – that cross-country comparative research with different number of income groupings per country is almost guaranteed to improve after applying our first-order correction term.

[24] A Monte Carlo experiment using MEPS income data showed that the correction proposed by Deltas almost completely removed the small sample bias for Gini indices calculated using between two and 50 individuals, while the first order correction proposed here mitigated rather than removed the bias.

F. Case study: income inequality rankings and first-order corrections

Although figure 3 shows that our first-order correction term removes a substantial part of the underestimation for each country separately, we believe it is worthwhile to present a case study on the potential of our first-order correction term to reduce the effect of income groupings on the income inequality ranking of the 16 countries. More exactly, we have analyzed how the income inequality ranking of the countries is affected if one were to use the Gini indices based on grouped income data reported in figure 2 for one country and the benchmark indices in figure 1 for all other countries, and to what extent our first-order correction term manages to restore the ranking in figure 1.[25] We prefer a case study where only the Gini for one country is affected by income groupings – as compared to a case where the Gini's of all countries are based on a different number of income groupings –as it is more likely to lead to a conservative assessment of the performance of our correction term.

[25] An alternative case study could focus on the effect of income groupings on longitudinal variation, and would reach similar conclusions. This would for example refer to the case where the number of income categories used in a questionnaire changes over time.

TABLE 2. – OUR FIRST-ORDER CORRECTION TERM: A CASE STUDY ON CROSS-COUNTRY COMPARISONS IN THE ECHP AND MEPS

	Sweden		Denmark		Finland		Netherlands		Austria		Belgium		Luxembourg		Ireland		Germany		Italy		Spain		France		UK		Greece		Portugal		US		Total	
	G	C	G	C	G	C	G	C	G	C	G	C	G	C	G	C	G	C	G	C	G	C	G	C	G	C	G	C	G	C	G	C	G	C
full	1		2		3		4		5		6		7		8		9		10		11		12		13		14		15		16			
50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
40	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0
20	0	0	0	0	-1	-1	0	0	0	0	0	0	0	0	-1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	-3	-2
15	0	0	0	0	-1	-1	0	0	0	0	0	0	0	0	-1	-1	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	-1	0	-4	-2
14	0	0	0	0	-1	-1	0	0	0	0	0	0	0	0	-1	-1	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	-1	-1	-4	-3
13	0	0	0	0	-1	-1	0	0	0	0	0	0	0	0	-1	-1	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	-1	-1	-4	-3
12	0	0	0	0	-1	-1	0	0	0	0	0	0	0	0	-1	-1	0	0	0	0	-1	0	0	0	0	0	-1	0	0	0	-1	-1	-5	-3
11	0	0	0	0	-1	-1	0	0	0	0	0	0	0	0	-1	-1	-1	0	0	0	-1	0	-1	0	0	0	-1	-1	0	0	-1	-1	-7	-4
10	0	0	0	0	-1	-1	0	0	0	0	0	0	0	0	-1	-1	-1	0	0	0	-1	0	-1	0	0	0	-1	-1	0	0	-1	-1	-7	-4
9	0	0	0	0	-1	-1	0	0	0	0	0	0	-1	0	-2	-1	-2	0	0	0	-1	0	-1	-1	0	0	-1	-1	0	0	-1	-1	-10	-5
8	0	0	0	0	-1	-1	0	0	0	0	0	0	-1	0	-2	-1	-2	-1	0	0	-1	-1	-2	-1	0	0	-1	-1	0	0	-1	-1	-11	-7
7	0	0	0	0	-1	-1	0	0	0	0	0	0	-1	0	-2	-1	-2	-1	0	0	-1	-1	-2	-1	0	0	-1	-1	0	0	-1	-1	-11	-7
6	0	0	0	0	-1	-1	0	0	0	0	0	0	-1	-1	-2	-2	-3	-2	0	0	-1	-1	-2	-2	-1	0	-1	-1	0	0	-1	-1	-13	-11
5	0	0	-1	0	-2	-1	0	0	0	0	-1	0	-1	-1	-2	-2	-3	-2	-1	0	-2	-1	-2	-2	-2	0	-2	-1	-1	0	-1	-1	-21	-11
4	0	0	-1	0	-2	-1	0	0	-1	0	-1	0	-2	-1	-3	-2	-4	-3	-3	0	-4	-1	-5	-2	-3	-1	-4	-1	-2	0	-3	-1	-38	-13
3	0	0	-1	0	-2	-1	-2	0	-2	0	-2	-1	-3	-1	-4	-2	-5	-3	-5	-1	-5	-1	-7	-2	-6	-1	-5	-1	-5	0	-5	-1	-59	-15
2	0	0	-1	-1	-2	-2	-3	0	-4	-1	-5	-1	-6	-2	-7	-3	-8	-4	-8	-3	-9	-4	-10	-5	-9	-3	-10	-3	-10	-2	-11	-2	-103	-36

Note: full: rank in full sample, 50-2: change in rank from income grouping/our correction term in the respective country while using the Gini from the full sample for all other countries, total: sum of rank changes over all countries, G: Gini index, C: Gini after our first order correction.
Countries are ranked according to full; light grey implies an improvement over the ranking using the Gini based on grouped data.

Table 2 presents the results of our case study. The row “full” shows the income inequality ranking using the full samples. The column “G” shows the change in the ranking from grouping the data for the country under study (and using the full sample Gini indices for the other countries). For example, Germany drops 4 places (from rank 9 to rank 5) for 4 income groups. The column “C” shows the change in the country ranking after applying our correction term to the country under study. Comparing columns “C” and “G” reveals the potential of our correction in restoring the income inequality ranking in row “full”. Cells in light grey imply that applying the correction term comes closer to the “full” country ranking. In the final column “Total”, we sum the change in the country rankings over all countries (i.e. the sum over the separate columns) giving an overall indicator of the performance of our correction term. A final issue to note is that our case study has a few built-in tendencies. Since we use the change in the country ranking, it is obvious that one is more likely to observe changes for countries that are ranked in the middle and at the top. In addition, since income groupings always lead to an underestimation of the Gini index, the “full sample” country ranking of the lowest ranked country (i.e. Sweden) will never change.

We find that changes in the income inequality ranking occur frequently, especially in case of a low number of income groups (see “G” columns). We also find that our correction term never worsens the income inequality ranking based on the grouped data, and often improves upon the latter. In other words, although it does not always restore the full sample country ranking, it never harms to use it in our case study.

V. Discussion and conclusion

This paper analyses the bias of the Gini index due to grouped data complicating comparisons of Gini indices calculated from such data. We develop a first-order correction term that results from studying the Gini in a measurement error framework, and show that it is inversely related to the covariance between the fractional rank at the individual and group level. Besides its simplicity and transparency, our procedure provides an exact and intuitive expression for the remaining and distribution-specific second-order bias allowing assessing a priori the performance of the first-order correction term for various shapes of the underlying distribution functions. We show that it exactly removes the bias due to income groupings for a uniform distribution, and is likely to remove a substantial share of the bias for an asymmetric unimodal distribution.

Using Dutch administrative data with more than 5 million observations and microdata from the ECHP and MEPS on income distributions of 15 European countries and the US, we illustrate that the underestimation from income groupings is similar across the 16 countries. Despite the wide variability in the Gini indices in the full samples, the value of the Gini has only a small effect on the magnitude of the underestimation. We further illustrate that the underestimation increases at an increasing pace when lowering the number of income categories, and that the underestimation is substantial relative to the sampling variability of the Gini index, its evolution over time, and cross-country differences in the value of the Gini.

Next, we illustrate the performance of our first-order correction term, and show that it reduces the underestimation of the Gini due to grouping considerably in all countries. We reached similar conclusions from a case study on the performance of our correction term in restoring the income inequality ranking if one were to use the Gini indices based on grouped income data for one country and the Gini's in the full samples for all other countries. In addition, our results suggest that the bias resulting from income groupings is fundamentally different from small-sample bias although both entail a small number of data points in practice. The latter bias is generally better addressed using the first-order correction term of Deltas (2003), but his correction should not be used to correcting bias that arises due to groupings of income.

A final issue concerns the terminology we have used throughout this paper. We have deliberately used 'income groupings' to abstract from a situation where the individuals in each income group have the same income. In the latter case, the Gini index estimated from grouped data is not biased, and thus application of our correction term would introduce an upward bias. 'Income groupings' instead point to a situation where microdata/official income statistics/etc. are grouped into a limited number of income groups, and thus neglecting within income group income variation leads to an underestimation.

Although this paper deals with the bias due to income groupings of the Gini index, we believe it is also useful for the widely used concentration index. For example, Wagstaff *et al.* (1991), Wagstaff and van Doorslaer (2000), and Burström *et al.* (2005) present applications to bivariate distributions in the health domain (inequalities in health/health care use/health care expenditures by income and occupational categories, etc.), Lambert (2001) gives an overview of applications to taxation (progressivity, redistributive effect, etc.), and many other applications have been reported in the economics literature. Its main difference with the Gini is that the fractional rank and the cumulative shares refer to different variables (for example cumulative shares of health over occupational groups), and thus the bias of the concentration index can be both down- and upward as the underlying concentration curves need not be convex and may have inflection points.

An important assumption in the theoretical and empirical part of this paper is that income groupings result in measurement error within income groups only, i.e. we assume that measurement error and the fractional group ranks are not correlated, and that the income ranking in the full sample is measured without error. This assumption allows studying the bias due to income groupings of the Gini in isolation, but neglects other types of measurement error. When answering a survey for example, a respondent may round off his/her reported income instead of reporting an exact amount or more generally income might be misreported. In combination with income groupings, the latter might introduce a misclassification bias to estimates of the Gini, i.e. an individual might be classified into the wrong income group based on his reported income. It is clear that misclassification and bias due to income groupings might be offsetting each other, and these issues have been analyzed for the variance of log incomes, the Theil and Atkinson inequality index by van Praag *et al.* (1983). Although we believe future research should analyze the relative importance of both biases in the Gini index, our results show that the bias from income groupings can be considerable.

REFERENCES

Abraham, Roberto, van den Bergh, Sidney, and Preethi Nair, "A new approach to galaxy morphology. I. Analysis of the Sloan digital sky survey early data release," *The Astrophysical Journal* 588 (2003), 218–229.

Alker, Hayward, *Mathematics and politics* (New York: the Macmillan company, 1965).

Allison, Paul, and John Stewart, "Productivity differences among scientists: evidence for accumulative advantage," *American Sociological Review* 39 (1947), 596-606.

Aronson, J. Richard, Johnson, Paul, and Peter Lambert, "Redistributive effect and unequal tax treatment," *Economic Journal* 104 (1994), 262-270.

Bigsten, Arne, and Abebe Shimeles, "Can Africa reduce poverty by half by 2015?," *Development policy review* 25 (2007), 147-166.

Blyde, Juan, "Convergence dynamics in Mercosur," *Journal of Economic Integration* 21 (2006), 784-815.

Borooah, Vani, "Problems in the measurement of inequality and poverty: a survey," *Indian economic journal* 38 (1991), 12-38.

Bourguignon, Francois, "Decomposable income inequality measures," *Econometrica* 47 (1979), 901-920.

Burström, Kristina, Johannesson, Magnus, and Finn Diderichsen, "Increasing socio-economic inequalities in life expectancy and QUALYs in Sweden 1980-1997," *Health Economics*14 (2005), 831-850.

Buzzacchi, Luigi, and Tommaso Valletti, "Firm size distribution: testing the "independent submarkets model" in the Italian motor insurance industry," *International journal of industrial organization* 24 (2006), 809-834.

Cameron, Colin, and Pravin Trivedi, *Microeconometrics: methods and applications* (Cambridge: Cambridge University Press, 2005).

Chen, Chau-Nan, Tsaur, Tien-Wang, and Tong-Shieng Rhai, "The Gini coefficient and negative income," *Oxford Economic Papers* 34 (1982), 473-478.

Datt, Gaurav, and Martin Ravallion, "Growth and redistribution components of changes in poverty measures: A decomposition with applications to Brazil and India in the 1980s," *Journal of Development Economics* 38 (1992), 275-295.

Deltas, George, "The small-sample bias of the Gini coefficient: results and implications for empirical research," *Review of economics and statistics* 85 (2003), 226-234.

EUROSTAT, *ECHP UDB Description of variables: Data Dictionary, Codebook and Differences between Countries and Waves* (Luxembourg: European Commission, 2003).

Fuller, Mike, "The Estimation of Gini Coefficients from Grouped Data: Upper and Lower Bounds," *Economics Letters* 3 (1979), 187-192.

Gastwirth, Joseph, "Robust Estimation of the Lorenz Curve and Gini Index," *Review of economics and statistics* 54 (1972), 306-316.

Hart, Peter, "Entropy and other measures of concentration," *Journal of the Royal Statistical Society, series A (general)* 134 (1971), 73-85.

Kakwani, Nanak, *Income inequality and poverty: Methods of estimation and policy applications* (London: Oxford University Press, 1980).

Kakwani, Nanak, *Analyzing redistribution policies* (London: Cambridge University Press, 1986).

Kakwani, Nanak, Wagstaff, Adam, and Eddy van Doorslaer, "Socioeconomic inequalities in health: measurement, computation, and statistical inference," *Journal of Econometrics* 77 (1997), 87-103.

Lambert, Peter, *The distribution and redistribution of income: third edition* (Manchester: Manchester University Press, 2001).

Lambert, Peter, and J. Richard Aronson, "Inequality decomposition analysis and the Gini coefficient revisited," *Economic Journal* 103 (1993), 1221-1227.

Le Grand, Julian, "Inequalities in health: some international comparisons," *European Economic Review* 31 (1987), 182-191.

Lerman, Robert, and Shlomo Yitzhaki, "Improving the accuracy of estimates of Gini coefficients," *Journal of Econometrics* 42 (1989), 43-47.

Li, Suxi, "The effect of demand uncertainty on price dispersion: the case of the airline industry," University of Miami, Coral Gables, mimeograph (2006).

Lin, Chun-Hung, "Education expansion, educational inequality, and income inequality: evidence from Taiwan, 1976-2003," *Social indicators research* 80 (2007), 601-615.

Mehran, Farhad, "Bounds on the Gini Index Based on Observed Points of the Lorenz Curve," *Journal of the American Statistical Association* 70 (1975), 64-66.

Mills, Jeffrey, and Sourushe Zandvakili, "Statistical Inference via Bootstrapping for measures of Inequality,“ *Journal of Applied Econometrics* 12 (1997), 133-150.

Murray, David, "Extreme Values for Gini Coefficients Calculated from Grouped Data," *Economics Letters* 1 (1978), 389-393.

OECD, *Main Economic Indicators* (Paris: OECD, 2008).

Ogwang, Tomson, "Bounds of the Gini index using sparse information on mean incomes," *Review of Income and Wealth* 49 (2003), 415-423.

Ogwang, Tomson, "An upper bound of the Gini index in the absence of mean income information," *Review of Income and Wealth* 52 (2006), 643-652.

Pradhan, Menno, Sahn, David, and Stephen Younger, "Decomposing world health inequality," *Journal of health economics* 22 (2003), 271-293.

Pyatt, Graham, Chen, Chau-nan, and John Fei, "The distribution of income by factor components," *The quarterly journal of economics* 95 (1980), 451-473.

Ryu, Hang, and Daniel Slottje, "Two flexible functional form approaches for approximating the Lorenz curve," *Journal of Econometrics* 72 (1996), 251-274.

Reynolds-Feighan, Aisling, "Competing networks, spatial and industrial concentration in the US airline industry," *Spatial Economic Analysis* 2 (2007), 237-257.

Sala-i-Martin, Xavier, "The world distribution of income: Falling poverty and…convergence, period," *Quarterly journal of economics* 121 (2006), 351−397.

Schader, Martin, and Friedrich Schmid, "Fitting Parametric Lorenz Curves to Grouped Income Distributions – A Critical Note," *Empirical Economics* 19 (1994), 361-370.

Sheret, Michael, "Evaluation studies equality trends and comparisons for the education system of Papua New Guinea," *Studies in Educational Evaluation* 14 (1988), 91-112.

Spiezia, Vincenzo, "Measuring regional economies," *Statistics Brief* 6, (Paris: OECD, 2003).

van Praag, Bernard, Hagenaars, Aldi, and Wim van Eck, "The influence of classification and observation errors on the measurement of income inequality," *Econometrica* 51 (1983), 1093-1108.

Villaseñor, José, and Barry Arnold, "Elliptical Lorenz curves," *Journal of econometrics* 40 (1989), 327-338.

Wagstaff, Adam, Paci, Pierella, and Eddy van Doorslaer, "On the measurement of inequalities in health," *Social Science and Medicine* 33 (1991), 545-557.

Wagstaff, Adam, and Eddy van Doorslaer, “Equity in health care finance and delivery” (pp. 1803-1862), in Anthony Culyer and Joseph Newhouse (Eds.), *Handbook of Health Economics* (Amsterdam: Elsevier Science, 2000).

World Bank, *PovcalNet* (Washington: World Bank, 2008). Available at http://iresearch.worldbank.org/PovcalNet/jsp/index.jsp (accessed 11 July 2008).

APPENDIX

A. Derivation of the variance of the fractional rank

Let σ_R^2 be the variance of the fractional rank

$$\sigma_R^2 = \frac{1}{n}\sum_{i=1}^{n}\left(R_i - \overline{R}\right)^2 \qquad \text{(A.1)}$$

Note also that the average fractional rank equals $1/2$ since

$$\overline{R} = \frac{1}{n}\sum_{i=1}^{n} R_i = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{i-0.5}{n}\right) = \frac{1}{n}\left[\frac{1}{n}\left(\frac{1+n}{2}n\right) - 0.5\right] = \frac{1}{2} \qquad \text{(A.2)}$$

Substituting equation (A.2) into equation (A.1) shows that

$$\sigma_R^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{2i-1-n}{2n}\right)^2 \qquad \text{(A.3)}$$

$$= \frac{1}{4n^3}\sum_{i=1}^{n}\left(2i-1-n\right)^2 \qquad \text{(A.4)}$$

$$= \frac{1}{4n^3}\left[\sum_{i=1}^{n}4i^2 - \sum_{i=1}^{n}4i - \sum_{i=1}^{n}4in + n + n^3 + 2n^2\right] \qquad \text{(A.5)}$$

$$= \frac{1}{4n^3}\left[\frac{2n(n+1)(2n+1)}{3} - 2n(1+n) - 2n^2(1+n) + n(n+1)^2\right] \qquad \text{(A.6)}$$

$$= \frac{1}{4n^3}\left[2n(n+1)\left(\frac{-2-n}{3}\right) + n(n+1)^2\right] \qquad \text{(A.7)}$$

$$= \frac{n^2-1}{12n^2} \qquad \text{(A.8)}$$

B. Equality between Gini estimated from equation (1) and (2)

Let us use OLS as an arithmetic tool to calculate β in equation (2), i.e.

$$\beta = \frac{\sum_{i=1}^{n}\left(R_i - \frac{1}{2}\right)\left(2\sigma_R^2 \frac{y_i}{\bar{y}} - 2\sigma_R^2\right)}{\sum_{i=1}^{n}\left(R_i - \frac{1}{2}\right)^2} \tag{B.1}$$

$$\beta = \left[\frac{2}{n}\sum_{i=1}^{n}\left(R_i - \frac{1}{2}\right)^2\right]\frac{\sum_{i=1}^{n} R_i \frac{y_i}{\bar{y}} - \sum_{i=1}^{n} R_i - \sum_{i=1}^{n}\frac{1}{2}\frac{y_i}{\bar{y}} + \sum_{i=1}^{n}\frac{1}{2}}{\sum_{i=1}^{n}\left(R_i - \frac{1}{2}\right)^2} \tag{B.2}$$

$$\beta = \frac{2}{n}\left[\sum_{i=1}^{n} R_i \frac{y_i}{\bar{y}} - \frac{n}{2} - \frac{n}{2} + \frac{n}{2}\right] = \frac{2}{n}\left[\sum_{i=1}^{n} R_i \frac{y_i}{\bar{y}} - \frac{n}{2}\right] = \frac{2\sum_{i=1}^{n} y_i R_i}{n\bar{y}} - 1 \tag{B.3}$$

C. Derivation of equation (8)

Let us start from equation (7). The OLS point estimate of β^{MER} equals

$$\beta^{MER} = \frac{\sum_{i=1}^{n}\left(R_i^g - \frac{1}{2}\right)\left(2\sigma_R^2 \frac{y_i}{\bar{y}} - 2\sigma_R^2\right)}{\sum_{i=1}^{n}\left(R_i^g - \frac{1}{2}\right)^2} \tag{C.1}$$

Next we use fact that the LHS of equation (6) and (7) are similar, such that

$$\beta^{MER} = \beta + \frac{\sum_{i=1}^{n}\left(R_i^g - \frac{1}{2}\right)\left(\varepsilon_i - \beta\delta_i^g - \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i + \beta\frac{1}{n}\sum_{i=1}^{n}\delta_i^g\right)}{\sum_{i=1}^{n}\left(R_i^g - \frac{1}{2}\right)^2} \tag{C.2}$$

Note that $n^{-1}\sum_{i=1}^{n}\varepsilon_i = 0$ due the properties of OLS as an arithmetic tool. It also holds that $n^{-1}\sum_{i=1}^{n}\delta_i^g = n^{-1}\left(\sum_{i=1}^{n}R_i^g - \sum_{i=1}^{n}R_i\right) = 0$ since the average of R_i^g and R_i equals $1/2$. Note that we have *not* relied upon $n \to +\infty$ to derive both properties.

$$\beta^{MER} = \beta + \frac{\sum_{i=1}^{n} R_i^g \varepsilon_i - \frac{1}{2}\sum_{i=1}^{n}\varepsilon_i - \beta\sum_{i=1}^{n}\delta_i^g R_i^g + \frac{\beta}{2}\sum_{i=1}^{n}\delta_i^g}{\sum_{i=1}^{n}\left(R_i^g - \frac{1}{2}\right)^2} \tag{C.3}$$

which similarly reduces to

$$\beta^{MER} = \beta + \frac{\sum_{i=1}^{n} R_i^g \varepsilon_i - \beta\sum_{i=1}^{n}\delta_i^g R_i^g}{\sum_{i=1}^{n}\left(R_i^g - \frac{1}{2}\right)^2} \tag{C.4}$$

It is important to note that the properties of the fractional rank make the measurement error δ_i^g uncorrelated with the fractional rank of group g defined at the individual level R_i^g (see footnote 11). Combining this information with the definition of the measurement error in equation (5), equation (C.4) reduces to

$$\beta^{MER} = \beta + \frac{\sum_{i=1}^{n} R_i^g \varepsilon_i}{\sum_{i=1}^{n}\left(R_i^g - \frac{1}{2}\right)^2} = \beta + \frac{\sum_{i=1}^{n} R_i \varepsilon_i + \sum_{i=1}^{n}\delta_i^g \varepsilon_i}{\sum_{i=1}^{n}\left(R_i^g - \frac{1}{2}\right)^2} \tag{C.5}$$

Finally, we should remember that OLS as an arithmetic device imposes on equation (2) that $n^{-1}\sum_{i=1}^{n}R_i\varepsilon_i = 0$. In combination with $\sum_{i=1}^{n}\left(R_i^g - 1/2\right)^2 = n\left(K\right)^{-1}\sum_{g=1}^{K}\left(R_g - 1/2\right)^2$ and the fact that equation (2) learns that $\beta = G_n$, equation (C.5) reduces to

$$\beta^{MER} = \beta + \frac{\sum_{i=1}^{n}\delta_i^g \varepsilon_i}{\frac{n}{K}\sum_{g=1}^{K}\left(R_g - \frac{1}{2}\right)^2} = G_n + \frac{\frac{1}{n}\sum_{i=1}^{n}\delta_i^g \varepsilon_i}{\sigma_{R^K}^2} \tag{C.6}$$

D. Derivation of equation (9)

In order to derive equation (9) it is worthwhile to notice that equation (7) – here (D.1) – and (D.2) give the same point estimate of β^{MER} since[1] equation (D.2) is basically a 'grouped data average' of equation (D.1)

$$2\sigma_R^2 \frac{y_i}{\bar{y}} = \alpha^{MER} + \beta^{MER} R_i^g + \eta_i \tag{D.1}$$

$$2\sigma_R^2 \frac{y_g}{\bar{y}} = \alpha^{MER} + \beta^{MER} R_g + \eta_g \tag{D.2}$$

Note that the LHS of (D.2) includes the variance of the fractional rank at the individual σ_R^2 and not the variance at the grouped level. The OLS estimate of β^{MER} in equation (D.2) equals

$$\beta^{MER} = \frac{\sum_{g=1}^{K}\left(R_g - \frac{1}{2}\right)\left(2\sigma_R^2 \frac{y_g}{\bar{y}} - 2\sigma_R^2\right)}{\sum_{g=1}^{K}\left(R_g - \frac{1}{2}\right)^2} \tag{D.3}$$

$$\beta^{MER} = \frac{\frac{2\sigma_R^2}{\bar{y}}\sum_{g=1}^{K} R_g y_g - 2\sigma_R^2 \sum_{g=1}^{K} R_g - \frac{\sigma_R^2}{\bar{y}}\sum_{g=1}^{K} y_g + \sum_{g=1}^{K}\sigma_R^2}{\sum_{g=1}^{K}\left(R_g - \frac{1}{2}\right)^2} \tag{D.4}$$

$$\beta^{MER} = \sigma_R^2 \left[\frac{\frac{2}{\bar{y}}\sum_{g=1}^{K} R_g y_g - K - K + K}{\sum_{g=1}^{K}\left(R_g - \frac{1}{2}\right)^2}\right] \tag{D.5}$$

[1] The point estimate is identical, while the standard error will differ. This is unimportant since we only use OLS as an arithmetic tool, not as a statistical device. Consult Pyatt *et al.* (1980) for a similar issue related to the covariance.

$$\beta^{MER} = \frac{\sigma_R^2}{\sigma_{R^K}^2}\left[\frac{\frac{2}{\bar{y}}\sum_{g=1}^{K} R_g y_g - K}{K}\right] = \frac{\sigma_R^2}{\sigma_{R^K}^2} G_n^K = \left[\frac{K^2\left(n^2-1\right)}{n^2\left(K^2-1\right)}\right] G_n^K \qquad \text{(D.6)}$$

E. Covariance between fractional rank at individual and grouped level

Let's start from equation (10)

$$G_n = \frac{K^2}{K^2-1}\left[\left(\frac{n^2-1}{n^2}\right)G_n^K - 12\frac{1}{n}\sum_{i=1}^{n}\delta_i^g \varepsilon_i\right] \qquad \text{(E.1)}$$

Next, we focus on the second term between brackets of this equation

$$\frac{12}{n}\sum_{i=1}^{n}\delta_i^g \varepsilon_i = \frac{12}{n}\sum_{i=1}^{n}\left(R_i^g - R_i\right)\varepsilon_i = \frac{12}{n}\sum_{i=1}^{n} R_i^g \varepsilon_i = \frac{12}{n}\sum_{i=1}^{n} R_i^g\left(2\sigma_R^2\frac{y_i}{\bar{y}} - \alpha - \beta R_i\right) \qquad \text{(E.2)}$$

Combine equation (E.2) with $\alpha = 2\sigma_R^2 - \left(2\right)^{-1}\beta$ (cf. 'averaging' of equation (2)), gives

$$\frac{12}{n}\sum_{i=1}^{n}\delta_i^g \varepsilon_i = \frac{12}{n}\left[\frac{2\sigma_R^2}{\bar{y}}\sum_{i=1}^{n} R_i^g y_i - \frac{n}{2}\left(2\sigma_R^2 - \frac{\beta}{2}\right) - \beta\sum_{i=1}^{n} R_i^g R_i\right] \qquad \text{(E.3)}$$

After some algebra, and noting that $\beta = G_n$, we get:

$$\frac{12}{n}\sum_{i=1}^{n}\delta_i^g \varepsilon_i = \frac{12}{n}\left[\sigma_R^2\left(\frac{2\sum_{i=1}^{n} R_i^g y_i}{\bar{y}} - n\right) - G_n\left(\sum_{i=1}^{n} R_i^g R_i - \frac{n}{4}\right)\right] \qquad \text{(E.4)}$$

$$\frac{12}{n}\sum_{i=1}^{n}\delta_i^g \varepsilon_i = 12\left\{\sigma_R^2 G_n^K - \frac{1}{n}\left[\sum_{i=1}^{n}\left(R_i^g - \frac{1}{2}\right)\left(R_i - \frac{1}{2}\right)\right] G_n\right\}$$

(E.5)

Combining (E.1) and (E.5), shows that

$$G_n = \frac{K^2}{K^2-1}\left[G_n^K\left(\frac{n^2-1}{n^2} - 12\frac{n^2-1}{12n^2}\right) + \frac{12}{n}\left[\sum_{i=1}^{n}\left(R_i^g - \frac{1}{2}\right)\left(R_i - \frac{1}{2}\right)\right] G_n\right] \qquad \text{(E.6)}$$

Equation (E.6) can only hold if

$$\frac{1}{n}\left[\sum_{i=1}^{n}\left(R_i^g-\frac{1}{2}\right)\left(R_i-\frac{1}{2}\right)\right]=\frac{K^2-1}{12K^2} \tag{E.7}$$

F. The case of unequal group sizes

Similar to equation (D.2), we start from

$$2\sigma_R^2\frac{y_u}{\overline{y}}\sqrt{n_u}=\alpha^{MER,u}\sqrt{n_u}+\beta^{MER,u}R_u\sqrt{n_u}+\eta_u \tag{F.1}$$

Next, we solve for $\beta^{MER,u}$

$$\beta^{MER,u}=\frac{\left(\sum_{u=1}^{K}n_u\right)\left(\frac{2\sigma_R^2}{\overline{y}}\sum_{u=1}^{K}n_uR_uy_u\right)-\left(\frac{2\sigma_R^2}{\overline{y}}\sum_{u=1}^{K}n_uy_u\right)\left(\sum_{u=1}^{K}n_uR_u\right)}{\left(\sum_{u=1}^{K}n_u\right)\left(\sum_{u=1}^{K}n_uR_u^2\right)-\left(\sum_{u=1}^{K}n_uR_u\right)^2} \tag{F.2}$$

$$\beta^{MER,u}=\frac{2\sigma_R^2}{\overline{y}}\frac{\left(\sum_{u=1}^{K}n_u\right)\left[\left(\sum_{u=1}^{K}n_uR_uy_u\right)-\left(\frac{1}{2}\sum_{u=1}^{K}n_uy_u\right)\right]}{\left(\sum_{u=1}^{K}n_u\right)\left[\left(\sum_{u=1}^{K}n_uR_u^2\right)-\left(\frac{1}{4}\sum_{u=1}^{K}n_u\right)\right]} \tag{F.3}$$

$$\beta^{MER,u}=\frac{2\sigma_R^2\left[\left(\sum_{u=1}^{K}n_uR_uy_u\right)-\left(\frac{1}{2}\sum_{u=1}^{K}n_uy_u\right)\right]}{\left(\sum_{u=1}^{K}n_uy_u\right)\sigma_{R_u^K}^2}=\frac{\sigma_R^2}{\sigma_{R_u^K}^2}G_n^{K,u} \tag{F.4}$$

G. Additional tables

TABLE A1. – DESCRIPTIVE STATISTICS OF EQUIVALENT INCOME

	obs	mean	stdev
Sweden	8889	137.947	63.268
Denmark	5899	131.497	69.759
Finland	8171	86.900	50.580
Netherlands (admin)	5.104.844	22.673	34.882
Netherlands	9351	28.788	15.363
Austria	7382	214.317	123.594
Belgium	6664	609.200	507.861
Luxembourg	2044	866.215	563.721
Ireland	9890	7.715	7.081
Germany	9390	31.414	24.164
Italy	17323	15.943	10.558
Spain	17757	1.107.543	763.037
France	13794	94.265	98.806
UK	10484	9.431	9.664
Greece	12423	1.562.758	1.347.131
Portugal	11445	887.748	750.996
US	17399	30.011	23.662

Source: Netherlands (admin) refers to authors' calculations from linked Dutch administrative data 2004

TABLE A2. – POINT ESTIMATES AND STANDARD ERRORS OF GINI INDICES ESTIMATED FROM FULL SAMPLE AND INCOME GROUPINGS IN ECHP AND MEPS

	Sweden		Denmark		Finland		Netherlands		Austria		Belgium		Luxembourg		Ireland	
	gini	s.e	gini	s.e	gini	s.e	gini	s.e	gini	s.e	gini	s.e	gini	s.e	gini	s.e
full	0,218	0,004	0,233	0,006	0,234	0,006	0,260	0,004	0,280	0,005	0,297	0,011	0,304	0,010	0,306	0,013
50	0,218	0,022	0,233	0,035	0,233	0,035	0,259	0,029	0,279	0,034	0,297	0,049	0,304	0,038	0,305	0,051
40	0,217	0,024	0,233	0,037	0,233	0,039	0,259	0,032	0,279	0,038	0,296	0,052	0,304	0,042	0,305	0,051
30	0,217	0,025	0,232	0,040	0,233	0,041	0,259	0,036	0,279	0,042	0,296	0,055	0,303	0,046	0,304	0,054
20	0,216	0,028	0,231	0,041	0,232	0,044	0,258	0,040	0,278	0,047	0,295	0,058	0,302	0,054	0,303	0,060
10	0,213	0,032	0,227	0,045	0,228	0,051	0,254	0,047	0,274	0,055	0,290	0,063	0,298	0,062	0,298	0,067
9	0,212	0,034	0,226	0,045	0,227	0,050	0,253	0,048	0,273	0,057	0,288	0,064	0,296	0,067	0,297	0,072
8	0,211	0,032	0,225	0,045	0,226	0,051	0,252	0,048	0,271	0,055	0,287	0,066	0,295	0,066	0,295	0,068
7	0,210	0,036	0,223	0,046	0,224	0,053	0,250	0,052	0,269	0,058	0,284	0,065	0,292	0,067	0,293	0,069
6	0,207	0,037	0,220	0,044	0,221	0,053	0,247	0,052	0,265	0,061	0,280	0,070	0,288	0,074	0,289	0,075
5	0,203	0,042	0,216	0,050	0,216	0,055	0,242	0,060	0,260	0,067	0,274	0,066	0,283	0,073	0,283	0,077
4	0,197	0,045	0,209	0,053	0,209	0,060	0,234	0,065	0,251	0,071	0,265	0,076	0,274	0,090	0,274	0,087
3	0,183	0,058	0,194	0,070	0,194	0,073	0,219	0,081	0,234	0,089	0,246	0,087	0,255	0,100	0,255	0,099
2	0,149	0,075	0,157	0,079	0,157	0,079	0,178	0,089	0,189	0,095	0,200	0,100	0,206	0,103	0,209	0,105

TABLE A2. – CONTINUED

	Germany		Italy		Spain		France		UK		Greece		Portugal		US	
	gini	s,e,	gini	s,e,	gini	s,e,	gini	s,e,	gini	s,e,	gini	s,e,	gini	s,e,	gini	s,e,
full	0,312	0,008	0,330	0,004	0,335	0,004	0,343	0,011	0,362	0,011	0,367	0,008	0,393	0,007	0,395	0,004
50	0,312	0,051	0,329	0,038	0,334	0,041	0,342	0,083	0,361	0,074	0,366	0,058	0,392	0,058	0,394	0,047
40	0,311	0,054	0,329	0,042	0,334	0,046	0,342	0,085	0,361	0,073	0,366	0,060	0,392	0,066	0,394	0,050
30	0,311	0,057	0,329	0,046	0,334	0,050	0,341	0,089	0,360	0,083	0,365	0,066	0,391	0,070	0,393	0,056
20	0,310	0,064	0,328	0,052	0,333	0,056	0,339	0,094	0,359	0,087	0,364	0,070	0,390	0,081	0,392	0,067
10	0,304	0,072	0,323	0,057	0,328	0,072	0,332	0,093	0,352	0,097	0,358	0,078	0,384	0,097	0,387	0,080
9	0,303	0,071	0,322	0,059	0,326	0,073	0,330	0,097	0,351	0,096	0,356	0,079	0,382	0,102	0,386	0,080
8	0,301	0,078	0,320	0,064	0,324	0,074	0,328	0,093	0,348	0,094	0,354	0,079	0,380	0,103	0,383	0,081
7	0,298	0,075	0,317	0,061	0,322	0,080	0,325	0,099	0,345	0,104	0,351	0,085	0,376	0,111	0,380	0,087
6	0,294	0,077	0,313	0,065	0,318	0,080	0,320	0,100	0,340	0,103	0,347	0,089	0,372	0,111	0,376	0,098
5	0,288	0,083	0,307	0,065	0,312	0,094	0,313	0,100	0,333	0,111	0,340	0,087	0,364	0,128	0,369	0,098
4	0,277	0,085	0,297	0,076	0,301	0,099	0,301	0,108	0,321	0,112	0,329	0,093	0,351	0,123	0,357	0,113
3	0,257	0,098	0,277	0,103	0,280	0,115	0,278	0,115	0,298	0,123	0,306	0,116	0,326	0,134	0,333	0,130
2	0,207	0,104	0,227	0,113	0,227	0,113	0,223	0,112	0,242	0,121	0,248	0,124	0,263	0,131	0,273	0,136

TABLE A3. – THE PERFORMANCE OF DELTAS' AND OUR FIRST-ORDER CORRECTION TERM TO ADDRESS THE UNDERESTIMATION OF THE GINI INDEX CALCULATED FROM GROUPED INCOME DATA IN THE ECHP AND MEPS

	Sweden				Denmark				Finland				Netherlands				Austria				Belgium				Luxembourg				Ireland			
	Gini	Deltas	Corr	covar	Gini	Deltas	Corr	covar	Gini	Deltas	Corr	covar	Gini	Deltas	Corr	covar	Gini	Deltas	Corr	covar	Gini	Deltas	Corr	covar	Gini	Deltas	Corr	covar	Gini	Deltas	Corr	covar
full	0,2178				0,2331				0,2339				0,2597				0,2798				0,2973				0,3041				0,3056			
50	0,2175	0,2220	0,2176	0,0000	0,2327	0,2374	0,2328	0,0000	0,2334	0,2382	0,2335	0,0000	0,2594	0,2647	0,2595	0,0000	0,2795	0,2852	0,2796	0,0000	0,2966	0,3027	0,2968	0,0000	0,3037	0,3099	0,3038	0,0000	0,3049	0,3111	0,3050	0,0000
40	0,2174	0,2230	0,2175	0,0000	0,2325	0,2385	0,2327	0,0000	0,2333	0,2392	0,2334	0,0000	0,2592	0,2659	0,2594	0,0000	0,2793	0,2865	0,2795	0,0000	0,2964	0,3040	0,2966	-0,0001	0,3035	0,3113	0,3037	0,0000	0,3046	0,3124	0,3048	-0,0001
30	0,2171	0,2246	0,2174	0,0000	0,2321	0,2402	0,2324	-0,0001	0,2329	0,2409	0,2332	-0,0001	0,2589	0,2678	0,2592	0,0000	0,2790	0,2886	0,2793	0,0000	0,2959	0,3061	0,2962	-0,0001	0,3031	0,3136	0,3034	-0,0001	0,3041	0,3146	0,3044	-0,0001
20	0,2164	0,2278	0,2170	-0,0001	0,2312	0,2434	0,2318	-0,0001	0,2320	0,2442	0,2326	-0,0001	0,2581	0,2717	0,2587	-0,0001	0,2780	0,2927	0,2787	-0,0001	0,2947	0,3102	0,2954	-0,0002	0,3021	0,3180	0,3029	-0,0001	0,3029	0,3188	0,3037	-0,0002
15	0,2155	0,2309	0,2165	-0,0001	0,2301	0,2466	0,2312	-0,0002	0,2309	0,2474	0,2320	-0,0002	0,2570	0,2754	0,2582	-0,0001	0,2769	0,2966	0,2781	-0,0001	0,2932	0,3142	0,2945	-0,0002	0,3008	0,3223	0,3021	-0,0002	0,3015	0,3230	0,3028	-0,0002
14	0,2153	0,2318	0,2164	-0,0001	0,2298	0,2474	0,2310	-0,0002	0,2306	0,2483	0,2318	-0,0002	0,2567	0,2764	0,2580	-0,0001	0,2765	0,2977	0,2779	-0,0002	0,2928	0,3153	0,2943	-0,0002	0,3004	0,3235	0,3019	-0,0002	0,3010	0,3242	0,3026	-0,0002
13	0,2149	0,2328	0,2162	-0,0001	0,2294	0,2485	0,2307	-0,0002	0,2302	0,2494	0,2315	-0,0002	0,2563	0,2776	0,2578	-0,0002	0,2760	0,2990	0,2776	-0,0002	0,2922	0,3165	0,2939	-0,0003	0,2999	0,3249	0,3017	-0,0002	0,3005	0,3256	0,3023	-0,0003
12	0,2145	0,2340	0,2160	-0,0002	0,2288	0,2496	0,2304	-0,0002	0,2297	0,2506	0,2313	-0,0002	0,2558	0,2790	0,2576	-0,0002	0,2754	0,3005	0,2774	-0,0002	0,2915	0,3180	0,2936	-0,0003	0,2993	0,3265	0,3014	-0,0002	0,2998	0,3271	0,3019	-0,0003
11	0,2140	0,2354	0,2158	-0,0002	0,2282	0,2510	0,2301	-0,0002	0,2290	0,2519	0,2309	-0,0002	0,2551	0,2806	0,2573	-0,0002	0,2748	0,3022	0,2771	-0,0002	0,2907	0,3198	0,2931	-0,0003	0,2985	0,3284	0,3010	-0,0003	0,2990	0,3289	0,3015	-0,0003
10	0,2133	0,2370	0,2155	-0,0002	0,2274	0,2527	0,2297	-0,0003	0,2282	0,2536	0,2305	-0,0003	0,2543	0,2826	0,2569	-0,0002	0,2738	0,3043	0,2766	-0,0003	0,2896	0,3218	0,2926	-0,0004	0,2975	0,3306	0,3005	-0,0003	0,2980	0,3311	0,3010	-0,0004
9	0,2125	0,2390	0,2151	-0,0002	0,2264	0,2547	0,2292	-0,0003	0,2271	0,2555	0,2300	-0,0003	0,2533	0,2849	0,2564	-0,0003	0,2726	0,3067	0,2760	-0,0003	0,2883	0,3244	0,2919	-0,0004	0,2962	0,3332	0,2999	-0,0003	0,2967	0,3338	0,3004	-0,0004
8	0,2112	0,2414	0,2146	-0,0003	0,2250	0,2571	0,2286	-0,0004	0,2257	0,2579	0,2293	-0,0004	0,2518	0,2878	0,2558	-0,0003	0,2711	0,3098	0,2754	-0,0004	0,2865	0,3275	0,2911	-0,0005	0,2945	0,3366	0,2992	-0,0004	0,2950	0,3371	0,2997	-0,0005
7	0,2096	0,2445	0,2140	-0,0003	0,2230	0,2602	0,2277	-0,0004	0,2237	0,2610	0,2283	-0,0005	0,2498	0,2914	0,2550	-0,0004	0,2688	0,3136	0,2744	-0,0004	0,2840	0,3313	0,2899	-0,0006	0,2919	0,3406	0,2980	-0,0005	0,2925	0,3413	0,2986	-0,0006
6	0,2071	0,2486	0,2130	-0,0004	0,2203	0,2643	0,2266	-0,0005	0,2208	0,2650	0,2271	-0,0005	0,2469	0,2962	0,2539	-0,0005	0,2655	0,3185	0,2730	-0,0006	0,2802	0,3362	0,2882	-0,0007	0,2885	0,3461	0,2967	-0,0006	0,2889	0,3467	0,2972	-0,0007
5	0,2033	0,2542	0,2118	-0,0005	0,2159	0,2699	0,2249	-0,0007	0,2164	0,2705	0,2254	-0,0007	0,2422	0,3027	0,2523	-0,0006	0,2603	0,3254	0,2712	-0,0007	0,2744	0,3431	0,2859	-0,0009	0,2830	0,3537	0,2948	-0,0007	0,2834	0,3543	0,2952	-0,0008
4	0,1967	0,2623	0,2099	-0,0006	0,2086	0,2781	0,2225	-0,0008	0,2088	0,2784	0,2227	-0,0009	0,2343	0,3125	0,2500	-0,0008	0,2514	0,3352	0,2681	-0,0009	0,2648	0,3530	0,2824	-0,0012	0,2735	0,3647	0,2918	-0,0010	0,2738	0,3651	0,2920	-0,0011
3	0,1833	0,2750	0,2062	-0,0009	0,1940	0,2910	0,2183	-0,0011	0,1940	0,2910	0,2182	-0,0012	0,2185	0,3278	0,2459	-0,0010	0,2335	0,3503	0,2627	-0,0013	0,2463	0,3695	0,2771	-0,0015	0,2547	0,3820	0,2865	-0,0013	0,2553	0,3830	0,2872	-0,0014
2	0,1495	0,2990	0,1993	-0,0012	0,1573	0,3147	0,2098	-0,0015	0,1573	0,3146	0,2097	-0,0015	0,1783	0,3567	0,2378	-0,0014	0,1894	0,3789	0,2526	-0,0017	0,1995	0,3990	0,2660	-0,0020	0,2061	0,4122	0,2748	-0,0018	0,2093	0,4186	0,2790	-0,0017

Note: Gini: point estimate of the Gini index; Deltas: point estimate of the Gini index using Deltas' first-order correction term, i.e. $K^{-1}(K-1)$; Corr: point estimate of the Gini index using our first-order correction term resulting from equation (10), i.e. $(K^2-1)^{-1}K^2$; covar: the covariance term in equation (10).

TABLE A3. – CONTINUED

	Germany				Italy				Spain				France			
	Gini	Deltas	Corr	covar	Gini	Deltas	Corr	covar	Gini	Deltas	Corr	covar	Gini	Deltas	Corr	covar
full	0,3123				0,3298				0,3346				0,3432			
50	0,3117	0,3181	0,3119	0,0000	0,3294	0,3361	0,3295	0,0000	0,3342	0,3410	0,3343	0,0000	0,3422	0,3492	0,3424	-0,0001
40	0,3115	0,3195	0,3117	-0,0001	0,3292	0,3377	0,3294	0,0000	0,3340	0,3425	0,3342	0,0000	0,3418	0,3506	0,3420	-0,0001
30	0,3110	0,3217	0,3113	-0,0001	0,3288	0,3402	0,3292	-0,0001	0,3336	0,3451	0,3339	-0,0001	0,3410	0,3528	0,3414	-0,0001
20	0,3097	0,3260	0,3105	-0,0002	0,3277	0,3450	0,3286	-0,0001	0,3325	0,3500	0,3333	-0,0001	0,3392	0,3571	0,3401	-0,0003
15	0,3081	0,3301	0,3095	-0,0002	0,3264	0,3497	0,3279	-0,0002	0,3312	0,3548	0,3327	-0,0002	0,3371	0,3612	0,3386	-0,0004
14	0,3076	0,3313	0,3092	-0,0003	0,3260	0,3510	0,3276	-0,0002	0,3308	0,3562	0,3324	-0,0002	0,3364	0,3623	0,3382	-0,0004
13	0,3070	0,3326	0,3089	-0,0003	0,3254	0,3526	0,3274	-0,0002	0,3302	0,3578	0,3322	-0,0002	0,3357	0,3636	0,3377	-0,0005
12	0,3063	0,3342	0,3085	-0,0003	0,3248	0,3543	0,3271	-0,0002	0,3296	0,3596	0,3319	-0,0002	0,3347	0,3652	0,3371	-0,0005
11	0,3054	0,3360	0,3080	-0,0004	0,3240	0,3564	0,3267	-0,0003	0,3288	0,3617	0,3315	-0,0002	0,3336	0,3669	0,3363	-0,0006
10	0,3042	0,3380	0,3073	-0,0004	0,3230	0,3588	0,3262	-0,0003	0,3278	0,3642	0,3311	-0,0003	0,3321	0,3690	0,3355	-0,0006
9	0,3027	0,3405	0,3065	-0,0005	0,3216	0,3618	0,3256	-0,0003	0,3263	0,3671	0,3304	-0,0003	0,3303	0,3716	0,3344	-0,0007
8	0,3007	0,3436	0,3054	-0,0006	0,3198	0,3655	0,3249	-0,0004	0,3245	0,3708	0,3296	-0,0004	0,3278	0,3747	0,3330	-0,0008
7	0,2979	0,3476	0,3041	-0,0007	0,3172	0,3701	0,3238	-0,0005	0,3218	0,3754	0,3285	-0,0005	0,3245	0,3786	0,3313	-0,0010
6	0,2939	0,3527	0,3023	-0,0008	0,3134	0,3761	0,3224	-0,0006	0,3179	0,3815	0,3270	-0,0006	0,3199	0,3838	0,3290	-0,0011
5	0,2877	0,3596	0,2997	-0,0010	0,3074	0,3843	0,3202	-0,0008	0,3119	0,3899	0,3249	-0,0008	0,3127	0,3909	0,3258	-0,0014
4	0,2772	0,3696	0,2957	-0,0013	0,2973	0,3964	0,3171	-0,0010	0,3013	0,4018	0,3214	-0,0010	0,3007	0,4009	0,3207	-0,0018
3	0,2570	0,3855	0,2891	-0,0017	0,2773	0,4159	0,3119	-0,0013	0,2802	0,4203	0,3152	-0,0014	0,2781	0,4172	0,3129	-0,0022
2	0,2074	0,4148	0,2765	-0,0022	0,2265	0,4530	0,3020	-0,0017	0,2271	0,4542	0,3028	-0,0020	0,2234	0,4468	0,2979	-0,0028

	UK				Greece				Portugal				US			
	Gini	Deltas	Corr	covar	Gini	Deltas	Corr	covar	Gini	Deltas	Corr	covar	Gini	Deltas	Corr	covar
full	0,3623				0,3666				0,3925				0,3946			
50	0,3615	0,3689	0,3616	-0,0001	0,3660	0,3734	0,3661	0,0000	0,3920	0,4000	0,3922	0,0000	0,3941	0,4022	0,3943	0,0000
40	0,3612	0,3704	0,3614	-0,0001	0,3657	0,3751	0,3659	-0,0001	0,3918	0,4018	0,3920	0,0000	0,3939	0,4040	0,3942	0,0000
30	0,3605	0,3729	0,3609	-0,0001	0,3651	0,3777	0,3655	-0,0001	0,3913	0,4048	0,3917	-0,0001	0,3935	0,4071	0,3939	-0,0001
20	0,3588	0,3777	0,3597	-0,0002	0,3638	0,3829	0,3647	-0,0002	0,3900	0,4105	0,3910	-0,0001	0,3924	0,4130	0,3934	-0,0001
15	0,3569	0,3824	0,3585	-0,0003	0,3621	0,3880	0,3637	-0,0002	0,3883	0,4160	0,3900	-0,0002	0,3909	0,4188	0,3926	-0,0002
14	0,3563	0,3837	0,3581	-0,0003	0,3616	0,3894	0,3635	-0,0003	0,3877	0,4175	0,3897	-0,0002	0,3904	0,4204	0,3924	-0,0002
13	0,3556	0,3852	0,3577	-0,0004	0,3610	0,3911	0,3632	-0,0003	0,3870	0,4192	0,3893	-0,0003	0,3898	0,4223	0,3921	-0,0002
12	0,3547	0,3869	0,3572	-0,0004	0,3602	0,3930	0,3627	-0,0003	0,3862	0,4213	0,3889	-0,0003	0,3891	0,4244	0,3918	-0,0002
11	0,3536	0,3890	0,3566	-0,0005	0,3593	0,3952	0,3623	-0,0004	0,3851	0,4236	0,3883	-0,0003	0,3882	0,4270	0,3914	-0,0003
10	0,3522	0,3914	0,3558	-0,0005	0,3580	0,3978	0,3616	-0,0004	0,3838	0,4264	0,3877	-0,0004	0,3870	0,4300	0,3909	-0,0003
9	0,3505	0,3943	0,3549	-0,0006	0,3564	0,4010	0,3609	-0,0005	0,3820	0,4298	0,3868	-0,0005	0,3855	0,4337	0,3903	-0,0003
8	0,3481	0,3978	0,3536	-0,0007	0,3543	0,4049	0,3599	-0,0006	0,3797	0,4339	0,3857	-0,0006	0,3834	0,4382	0,3895	-0,0004
7	0,3449	0,4024	0,3521	-0,0008	0,3514	0,4099	0,3587	-0,0006	0,3763	0,4390	0,3842	-0,0007	0,3805	0,4439	0,3884	-0,0005
6	0,3402	0,4083	0,3499	-0,0010	0,3470	0,4164	0,3569	-0,0008	0,3715	0,4458	0,3821	-0,0008	0,3761	0,4513	0,3868	-0,0006
5	0,3330	0,4163	0,3469	-0,0012	0,3400	0,4250	0,3542	-0,0010	0,3642	0,4553	0,3794	-0,0011	0,3692	0,4615	0,3846	-0,0008
4	0,3212	0,4282	0,3426	-0,0015	0,3287	0,4382	0,3506	-0,0013	0,3512	0,4682	0,3746	-0,0014	0,3575	0,4766	0,3813	-0,0010
3	0,2982	0,4473	0,3354	-0,0020	0,3062	0,4594	0,3445	-0,0016	0,3260	0,4890	0,3668	-0,0019	0,3335	0,5002	0,3751	-0,0014
2	0,2416	0,4832	0,3221	-0,0025	0,2484	0,4967	0,3311	-0,0022	0,2628	0,5257	0,3505	-0,0026	0,2727	0,5454	0,3636	-0,0019

Note: Gini: point estimate of the Gini index; Deltas: point estimate of the Gini index using Deltas' first-order correction term, i.e. $K^{-1}(K-1)$; Corr: point estimate of the Gini index using our first-order correction term resulting from equation (10), i.e. $(K^2-1)^{-1}K^2$; covar: the covariance term in equation (10).