OECD Economics Department Working Papers No. 1265

What do household surveys suggest about the top 1% incomes and inequality in OECD countries?

Nicolas Ruiz,

Nicolas Woloszko

https://dx.doi.org/10.1787/5jrs556f36zt-en

OECD

Unclassified **ECO/WKP(2015)83**

Organisation de Coopération et de Développement Économiques
Organisation for Economic Co-operation and Development **21-Jan-2016**

English - Or. English

ECONOMICS DEPARTMENT

ECO/WKP(2015)83
Unclassified

WHAT DO HOUSEHOLD SURVEYS SUGGEST ABOUT THE TOP 1% INCOMES AND INEQUALITY IN OECD COUNTRIES?

ECONOMICS DEPARTMENT WORKING PAPERS No. 1265

By Nicolas Ruiz and Nicolas Woloszko

OECD Working Papers should not be reported as representing the official views of the OECD or of its member countries. The opinions expressed and arguments employed are those of the author(s).

Authorised for publication by Christian Kastrop, Director, Policy Studies Branch, Economics Department.

Document available in pdf format only.

All Economics Department Working Papers are available at www.oecd.org/eco/workingpapers

JT03389206

Complete document available on OLIS in its original format
This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

English - Or. English

OECD Working Papers should not be reported as representing the official views of the OECD or of its member countries. The opinions expressed and arguments employed are those of the author(s).

Working Papers describe preliminary results or research in progress by the author(s) and are published to stimulate discussion on a broad range of issues on which the OECD works.

Comments on Working Papers are welcomed, and may be sent to the Economics Department, OECD, 2 rue André-Pascal, 75775 Paris Cedex 16, France, or by e-mail to eco.contact@oecd.org.

All Economics Department Working Papers are available at www.oecd.org/eco/workingpapers.

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

© OECD (2016)

You can copy, download or print OECD content for your own use, and you can include excerpts from OECD publications, databases and multimedia products in your own documents, presentations, blogs, websites and teaching materials, provided that suitable acknowledgment of OECD as source and copyright owner is given. All requests for commercial use and translation rights should be submitted to **rights@oecd.org**

ABSTRACT/RÉSUMÉ

What do household surveys suggest about the top 1% incomes and inequality in OECD countries?

Standard income inequality figures, based on official household survey statistics covering most of the population, report a steady rise of inequality across a majority of advanced countries. The usefulness of these data sources in providing a timely and internationally comparable picture of inequality is undisputed, but one well-known limitation is their under-reporting of top incomes. This matters insofar as separate data sources devoted specifically to top incomes evolution report substantially faster inequality growth in recent years compared to conventional statistics. This paper proposes a methodology to adjust household survey data for the under-reporting of top incomes. More specifically, the analysis delivers a set of top incomes-adjusted income distribution series that bring together the bottom 99% and the top 1%. Unsurprisingly, the results point to a significant increase of the level of inequality measured by standard statistics based on official figures: the Gini coefficient adjusted for top incomes was in 2011 on average 6 percentage points higher, moving from 0.31 to 0.37 for the average OECD country; similarly, the gap between the mean income of the richest and the poorest 10% rises from 10 to 15 as a result from the adjustment. Inequality trends are also significantly altered, albeit in ways that differ across countries.

Classification JEL: D31, D63, O15

Keywords: Income, inequality, top incomes, household surveys

****

Que peut-on apprendre des hauts revenus à partir de données d'enquêtes dans les pays de l'OCDE ?

Les chiffres sur les inégalités de revenus, basés sur les enquêtes auprès des ménages, couvrent la plupart de la population et font état d'une augmentation constante de l'inégalité dans la majorité des pays avancés. L'utilité de ces sources afin de fournir une image satisfaisante et comparables à l'échelle internationale de l'inégalité est reconnue, mais une limite bien connue est leur sous-déclaration des hauts revenus. Cela importe dans la mesure où des sources de données distinctes consacrées spécifiquement aux hauts revenus décrivent une évolution de l'inégalité sensiblement plus rapide au cours des dernières années par rapport aux statistiques classiques. Cet article propose une méthodologie pour ajuster les données d'enquêtes auprès des ménages pour la sous-déclaration des hauts revenus. Plus précisément, l'analyse fournit un ensemble de série sur la répartition des revenus ajustés pour les hauts revenus et qui rassemblent ainsi les 99% et les 1% des ménages les plus riches. Les résultats montrent une augmentation significative du niveau de l'inégalité par rapport aux statistiques standard basées sur des chiffres officiels: le coefficient de Gini corrigé des hauts revenus a été en 2011 en moyenne de 6 points de pourcentage supérieur, passant de 0,31 à 0,37 pour la moyenne des pays de l'OCDE ; De même, l'écart entre le revenu moyen des plus riches et les plus pauvres passe de 10 à 15 suite à l'ajustement. L'évolution des inégalités est également modifiée de façon significative, quoique de manière hétérogène selon les pays.

Classification JEL: D31, D63, O15

Keywords: Revenue, inégalité, hauts-revenus, enquêtes auprès des ménages

TABLE OF CONTENTS

WHAT DO HOUSEHOLD SURVEYS SUGGEST ABOUT THE TOP 1% INCOMES AND INEQUALITY IN OECD COUNTRIES? 6
Introduction and main findings 6
Data sources for measuring inequality 8
Household surveys 8
Secondary cross-national datasets 10
Tabulations from tax data 10
Estimation of top income shares from household surveys 13
Methodology 13
Data sources 15
Estimation and comparison with the WTID 15
Top income shares in all OECD countries 21
Income inequality in OECD countries based on adjusted income distribution measures 21
Effects of the adjustment on mean incomes 22
Inequality across the 100% 24
Conclusion 28

BIBLIOGRAPHY 29

Annex 1: Conditionnal estimate of an income distribution upper tail from household survey 32

Annex 2: Adjusted vs. non-adjusted ratios of mean income of the richest 10% to the median in OECD countries 35

Annex 3: Adjusted vs. non-adjusted generalised means curves in OECD countries (from mid-1990's to late-2000's) 40

Tables

1. Example of tax data tabulation for the United States (2007) 11
2. Summary of the differences between estimated top incomes series and the World Top Incomes Database 20

Figures

1. Reversed Pareto Coefficient : comparison between IDD and WTID data for selected OECD countries 16
2. Top 1% income shares : comparison between IDD and WTID data for selected OECD countries 18
3. Top 1% income shares in OECD countries (2005 and 2010) 21
4. Adjusted vs. non-adjusted mean income (as a multiple of median income) 23
5. Adjusted vs. unadjusted top decile mean income for selected OECD countries (as a multiple of median income) 24

6.. Impact of the adjustment on standard indexes of income inequality..... 26
7. Growth in household income across the distribution between the mid-1990s and late 2000s: selected countries..... 27

Boxes

Box 1. Pareto distribution..... 12
Box 2. Conditional likelihood estimator of top incomes..... 13

WHAT DO HOUSEHOLD SURVEYS SUGGEST ABOUT THE TOP 1% INCOMES AND INEQUALITY IN OECD COUNTRIES?

By Nicolas Ruiz and Nicolas Woloszko[1]

Introduction and main findings

1. Concerns about the distribution of income weigh heavily in the public policy debate. Indeed, in most OECD countries and many emerging economies, the gap between the rich and the poor has widened over the past decades. The economic crisis has added urgency to deal with the policy issues related to inequality. Greater inequality raises economic, political and ethical challenges as it risks leaving behind a growing number of individual in an ever changing economic landscape. It can even undermine future growth prospect in the long run, should income inequality further translate into inequality of opportunity, decrease in social mobility and the weakening of incentives to invest in knowledge.

2. If the current challenges are clear, there is less common understanding about the definition and measurement of income inequality. Going back to the 1970s and the more widespread availability of micro data sets, income inequality at national levels has been mostly gauged on the basis of household surveys. International initiatives such as the OECD Income Distribution database or the United Nations World Income Distribution database have since made possible the use of such data for international comparisons. Data from these sources have been used to form the backbone of official statistics to guide policy reflexion on inequality and to feed most of the research works related to the topic.

3. A consensual finding from these data is that income inequality has been steadily rising between the mid-1980s and the late 2000s, but at a much slower pace after the mid-1990s: on average among OECD countries, the Gini coefficient on post-redistribution income has increased by 2.6 percentage points, with most of the increase (2.1 percentage points) occurring between the mid-1980s and the mid-1990s (see OECD, 2011). During the crisis years, recent evidence point also to stable inequality levels on average (OECD, 2015).

4. However, a wave of research based on new data has challenged this conclusion. Using tabulations from tax return data for the United States – a methodology later applied to a wide range of countries -- more recent studies have shown that the share of total income held by the very richest households actually rose faster in the 1990s than in the 1980s (Piketty and Saez; 2003; Atkinson *et al.*, 2011). These findings indicate that income inequality has actually grown more rapidly over the last fifteen years than previously thought, particularly in English-speaking countries, and that following the crisis there has been a significant upsurge in top incomes (Piketty and Saez, 2013). These new findings sparked a debate, particularly in the United States, about the real trends on inequality and the validity of the numbers used in underlying studies, leading to consider official statistics as not fully suited for casting light on the relevant policy issues.

5. These findings altered the popular and expert views on inequality. In particular, the notion of inequality is now often narrowed down to the gap between the very rich and the rest of the population (or

---

1. Nicolas Ruiz is an Economist in the OECD Economics Department. At the time of conducting this work, Nicolas Woloszko was a trainee in the OECD Economics Department. The authors would like to thank Orsetta Causa, Alain De Serres, Christian Kastrop, Jean-Luc Schneider and various colleagues from the Statistics and Employment, Labour and Social Affaires Directorate for their valuable comments and suggestions. Special thanks go to Agnès Cavaciuti for statistical assistance and to Caroline Abettan for technical assistance.

in other terms into the divide between the top 1% and the bottom 99%, as popularised notably by the Occupy Wall Street Movement). This view is undoubtedly relevant but in itself too narrow. A growing share of national income accumulated by the top 1% can indeed have a deleterious effect on society and shape important issues in terms of the political economy. The fact that between 1975 and 2007 around 47% of total GDP growth per capita in the United States went to the top 1% (see OECD, 2014a) justifies indeed a specific focus on top incomes. But the divide between the 99% and the 1% constitutes only one aspect of inequality.

6. In fact, casting light mainly on top incomes and how they tend to move apart from the rest of the population occult what is happening within the bulk of a society. And as argued recently by Acemoglu and Robinson (2014), an exclusive focus on the top 1% "may not be the most fruitful way of thinking about the societal and welfare consequences of inequality". Other elements, such as the gap between the bottom and the middle class, are equally valuable to understand inequality, social mobility and thus for relevant policy discussions.[2] As such, a complete treatment of inequality requires a picture of the 100%, i.e. the need to bring together a small group of the population with a large share of income, that is the top 1%, and a group comprising most of the population but where income concentration is less severe, that is the 99%.[3]

7. This requirement seems quite natural and at first glance pretty straightforward. But in practice deriving a complete picture of inequality is a task fraught with difficulties, mainly because income inequality figures derived from household surveys are weakly comparable to the sources used to measure top shares, i.e. tax return-based tabulations. As a result, the information upon which evidence-based policy recommendations are made is currently fragmented. On the one hand, inequality figures drawn from household surveys measure income dispersion on a comprehensive and representative portion of the population, but are not able to capture top incomes. Yet it is in this portion of the distribution that most of the changes in inequality seem to have occurred over the last fifteen years (Piketty and Saez; 2003; Atkinson *et al.*, 2011). This implies in particular a systematic underestimation of inequality levels based on surveys. On the other hand, figures derived from tax return data offer an accurate picture for top incomes, but remain mute about how top incomes fit in the overall distribution and what are the global inequality trends: in particular, the complete evolution of inequality in a country where top incomes shares increase but where inequality across the rest of the population decreases cannot be determined. In fact, no overall measure is available for quantifying what happens to inequality when the situation of the very rich is evolving differently from the situation of the rest of the distribution.

8. Against this background, this paper proposes a methodology to derive top incomes estimations from household surveys-based data in order to reconstitute a comprehensive picture of inequality in OECD countries. Currently, the official data on income inequality from the OECD[4] do not cover developments in the very top of the distribution. In a nutshell, the analysis delivers a set of top incomes-adjusted income distribution series that bring together the bottom 99% and the top 1%, based on a technical exploration of official and comparable OECD data[5] . This is done by correcting household survey data for underreporting in the upper-tail of income distributions. The resulting new top incomes estimates are compared with those

2. At the theoretical level, an exclusive focus on the top 1% also entails the postulate of a specific social welfare function which could be rather limited to represent a society's preference *vis-à-vis* inequality.

3. There are also well-known weaknesses from income distribution data sources to capture accurately the bottom of the distribution, in particular the very poor. However, this requires different data sources and methodologies that are beyond the scope of this paper.

4 . See http://www.oecd.org/social/income-distribution-database.htm

5 . The estimates generated in this paper are the responsibility of the authors and should not be treated as official OECD data. The Income Distribution Database remains the official source.

of the World Top Incomes Database[6] (hereafter WTID), the natural, tax-based benchmark for comparison, which gathers all the series estimated by Atkinson *et al.* (2011). In order to allow for a systematic and up to-date assessment of inequality, the approach presented in this paper requires minimal data requirements and can be easily extended across countries and over time. More specifically, the estimates are derived from secondary datasets, i.e. using semi-aggregated income distribution figures, rather than direct micro data, which are generally not made available in international databases. Using harmonised secondary datasets allows for a cross-country comparison.

9. The main insights emerging from this paper are:

- Levels of top incomes shares estimated from OECD data are in line with those of the WTID: on average across countries, the difference between the two sources in the share of income going to the top 1% is less than 2 percentage points in 2010.

- Estimated trend evolutions in top incomes shares are also in line with those of the WTID: on average between the mid-1980s and 2010, estimated top incomes shares are trending in the same direction as those in the WTID in 70% of the time.

- Accounting for top incomes results in a significant increase of the level of inequality measured by standard statistics based on household surveys: across OECD countries, the Gini coefficient corrected for top incomes is in 2011 on average 6 percentage points higher, moving to 0.37 from 0.31. Similarly, the gap between the mean income of the richest and the poorest 10% rises from 10 to 15.

- Accounting for top incomes also significantly alters the assessment of inequality trends delivered by standard statistics based on household surveys. This adjustment can either reinforce (e.g. Chile, Japan, and the United States) or moderate (e.g. Australia, Austria and France) the unadjusted trend evolution.

10. The rest of the paper is structured as follows. Section 2 provides a summary of inequality measurement through respectively household surveys, secondary cross-national datasets and tax return sources. Section 3 presents the analytical framework used in this paper to correct household surveys for top incomes and delivers new estimates of top income shares from household surveys for OECD countries, while providing a systematic comparison with the WTID. This allows in Section 4 for a comprehensive assessment of the income distribution and therefore of the level and evolution of inequality in all OECD countries, combining top incomes with the rest of the population. The last section concludes.

Data sources for measuring inequality

Household surveys

11. In OECD countries, the bulk of official income distribution statistics is derived from household surveys, generally conducted on a yearly basis. In theory, the ideal income survey would cover all households in the population, with an equal selection probability for each of those and a comprehensive response by those actually surveyed at the implementation stage. Under such unbiased survey design, very rich households would be surveyed and represented.

12. In practice, for numerous reasons, the design of a household survey is far more complex. Non-compliance from selected households can be high, leading to various measurement errors and associated

6. See http://topincomes.parisschoolofeconomics.eu/

issues for income distribution analysis (see Deaton, 1997 for a comprehensive review of household survey methodology and underlying issues). In the context of this paper, it is important to highlight that in recent years concerns have been raised about the capacity of household surveys to appropriately capture developments in the upper end of the income distribution. Factors that may disproportionately affect the quality of estimates for the highest income group from income surveys are:

- Higher than average non-response rates (due to respondent's high opportunity costs, concerns about sensitivities of complex personal affairs and difficulties in gaining access to this group).

- Unwillingness to provide accurate data due to respondent concerns about confidentiality and possible disclosure of sensitive information.

- Survey collection and processing constraints such as any limit on the number of digits that can be recorded in the survey instrument or dataset.

- Top coding processes applied to the public use datasets, i.e. censuring of income amount above a certain threshold in order to limit the risk of disclosing confidential information.

13. At first glance, those issues may be considered as fairly minor and the underestimation of top incomes in household surveys as having all in all little impact on overall inequality. In fact, simple calculations show that this is not the case. Assuming that the top income group is an infinitesimal share of the total population but with a finite share of total income, one can show that if between two periods the Gini coefficient on the bottom 99% of a distribution remains stable at (say) 40 percentage points, then a rise of 14 percentage points in the share of income held by the top 1% (as happened in the United States between 1976 and 2006) implies a rise of 8.4 percentage points in the Gini coefficient on the overall population[7]. As a result, missing top incomes in household surveys can lead to large bias in the measurement of inequality through, for example, the Gini index.

14. Acknowledging this issue, recent studies have tried to improve the upper tail information of household surveys, in particular by attempting to reconcile for the United States the top income estimates from Piketty and Saez (2003) with those from household surveys (see Burkhauser *et al.*, 2012). While the results point to promising similarities between the two sources, the methodology used is unfortunately not of general interest in practice (particularly in a cross-national comparisons perspective) for two reasons:

- The comparison uses data internal to the United States Census Bureau, data generally not available to external users. The extension of these results to other countries is thus a challenging task. But clearly, cross-country comparisons of inequality need to rely to the largest extent possible on statistical capacity available to most users.

- These studies focus specifically on the issue of top coding: to protect the confidentiality of respondents, income received by households above a certain threshold is replaced by the threshold value. Since the coverage of top-coded data is generally better in publicly available database than in data internal to National Statistical Offices, the use of the former instead of the latter leads mechanically to an improved appraisal of top incomes. However, while the partial or complete removal of top coding is a step in the right direction, it is only one of the sources of underestimation of top incomes in household surveys. Under-coverage and underreporting are also major and probably more prominent drivers. Thus relaxing top

7. See Atkinson *et al.*, 2011 for the corresponding decomposition formula of the Gini coefficient.

coding through the use of less censored data only solves part of the problem: the fundamental issue is that very rich households are simply hard to survey.

- In fact, even full access to non-top coded data does not prevent the necessity of using imputation techniques for the top 1%: while Burkhauser *et al.*, (2012) replicate quite accurately the income share held by the top 10%, they nonetheless derive estimates for the top 1% using power law-based distributions.

15. Evidence also suggests that the under-coverage of top income has been on the rise. In particular, the richer a country becomes, the higher is the probability that the response rate to a household survey will be negatively related to income.[8] This is also true between countries: despite generally better statistical capacity, developed countries tend to suffer more from under-coverage of top incomes than developing countries. Why it occurs is an issue still debated. However, a direct consequence is that the estimation of top incomes has to be achieved on alternative data sources to be deemed reliable.

Secondary cross-national datasets

16. In a cross-national perspective, the study of inequality generally relies on the use of secondary cross-national datasets, i.e. ready-made data representing income distributions in different countries at different dates. Such datasets summarize household surveys with a set of tabulations and key statistics that provide sufficient information on inequality but without the need for users to delve into the original micro-data. Most popular secondary datasets are currently the Deininger and Squire (1996) database compiled by the World Bank and the World Income Inequality Database compiled by the United Nations (derived also from World Bank sources). Both sources have been widely used and have allowed the introduction of distributional variables into a broad range of studies, e.g. Barro, 2000; Dollar and Kraay, 2002; Dollar *et al.*, 2013. In this paper, the secondary dataset used is the OECD Income Distribution Database (hereafter IDD), which gathers a number of inequality summary indicators and decile points for income distributions over the 34 OECD countries since the mid-70s.

17. Because they are mostly based on household surveys[9], secondary datasets suffer from the same issues, failing in particular to give a proper image of top incomes. However, one additional problem is the departure of the original survey due to data grouping, which can potentially imply a loss of information. But following Davies and Shorrocks (1989), just five income groups are needed to produce a Gini coefficient equal to 95% of the true value computed on the original micro data, while only twelve groups are required to get 99% of the true value. Secondary datasets are generally between the two, with ten income groups for each country and at each date in the case of the IDD. As a result, the errors involved due to data grouping appear as negligible.

Tabulations from tax data

18. The recent literature on top incomes generally uses aggregate tax data grouped according to the thresholds of the income taxes prevailing in each country, coming under the form of Table 1. These data are published by the tax administration, almost every year since the setting of progressive taxation in most advanced countries. More generally, all countries with a well-developed income tax system compile and publish tabulations based on the exhaustive set of income tax returns.

---

8. A direct consequence of this phenomenon is that the ratio between aggregate income amounts from household surveys and National Accounts measures of income tends to decrease over time (see Bourguignon, 2005; Deaton, 2005).

9. The IDD is also based for some countries on tax data or some combinations of tax and survey data (see Gasparini and Tornarolli, 2015 for a detailed presentation of the IDD).

19. These data have the clear advantage of exhaustiveness as all the incomes of tax-filers, thus in particular top incomes, are included and the data are available across countries over a long-time span (almost one century for Western countries). But a clear disadvantage is that incomes of the non-tax filers are not included. As a result these data cover only part of the population. In order to match these data with the entire population, taxable income must first be linked to the total income of the economy. To ensure consistency over time and across countries, National Accounts income data are used to estimate total income. After adjusting for differences in tax systems, the income of non tax-filers is then derived as a residual.[10]

Table 1. Example of tax data tabulation for the United States (2007)

Tax rate	Number of returns	Total income	Total tax paid
	'000	$m	$m
5 percent	13,218	76,924	3,846
10 percent	108,976	1,101,418	110,142
15 percent	81,501	1,955,871	293,381
25 percent	30,354	844,825	211,206
28 percent	5,904	293,691	82,217
33 percent	2,730	260,213	85,870
35 percent	1,061	586,067	240,123

20. Thus, a central feature of the literature on top incomes is that it is not based on household or individual micro records *e.g.* administrative, census files. Only grouped tabulations retrieved from fiscal administration are used. These tabulations, because they are structured according to income tax brackets, do not generally single out the group of interest such as the top 1%; the groups depicted by these data are only driven by the tax legislation of each country. A model is therefore required to estimate the summary statistics for the group of interest. The preferred estimation methodology is the fit of a Pareto distribution (Box 1). If the number of tax brackets reported is large, measurement errors with this methodology are generally deemed to be small (see Atkinson *et al.*, 2011).

21. Despite the exhaustiveness of tax data and the confidence in the methodology applied to them, several factors can reduce the validity of tax data-based top incomes estimation (see Leigh, 2007 for a complete discussion). In particular:

- Income shifting induced by changes in tax legislation, e.g. lowering of personal tax rates might lead individuals to shift their business income from corporate tax returns to individual tax returns, which may inflate the share of (personal) taxable income accruing to the very

10. It has to be noted that the unit of observation varies in tax data, contrary to household survey, where equivalised disposable household income is normally used for income distribution analysis (see Deaton, 1997).

rich. That is, people with very high incomes may not earn more income but simply declare more of their income as part of their personal tax declaration.

Box 1. Pareto distribution

A Pareto law is given by the following cumulative distribution function F(y) for income y:

$$F(y) = P(Y < y) = 1 - \left(\frac{y}{k}\right)^{-\alpha} \ \forall y > k$$

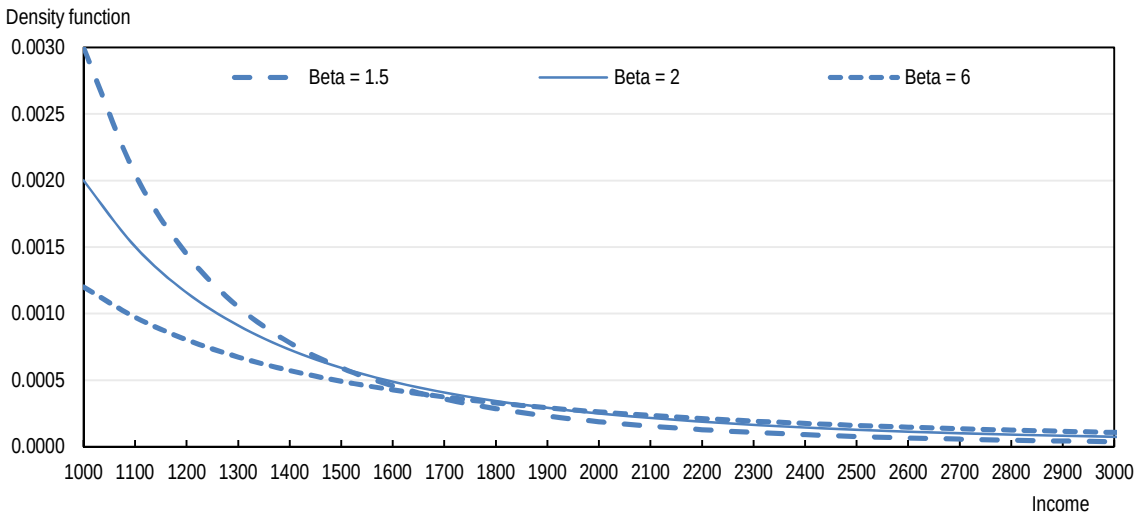
$$= 0 \ \forall y \leq k$$

where k is a positive scaling parameter and α is called the Pareto coefficient. The corresponding density function f(y) is given by:

$$f(y) = \alpha \frac{k^{\alpha}}{y^{1+\alpha}}$$

The key property of a Pareto distribution is that the ratio of the average income y* of individuals or households with income above y to y does not depend on the income threshold y and is equal to the inverted Pareto coefficient β, i.e. ${}^{y\,*}/_{y} = \beta \ with \ \beta = {}^{\alpha}/_{\alpha - 1}$. For instance, if $\beta = 2$, the average income of individuals with income superior to $100,000 is $200,000. The inverted Pareto coefficient has thus an intuitive economic interpretation: the higher the coefficient, the fatter the upper tail of the distribution (see Figure below). In its own an inverted Pareto coefficient can be interpreted as a measure of inequality.

Pareto density functions



Note: The higher is the reversed Pareto coefficient beta, the thicker is the upper-tail of the distribution, which implies a more unequal distribution of income and a greater number of very high incomes.

Once the computation of the inverted Pareto coefficient has been derived from data such as in Table 1, the share of income for the top 5%, top 1%, top 0.1%... can be computed from local approximations of f(y). While estimations at different levels of brackets can lead to different inverted Pareto coefficients (and thus different top income shares), in practice this is rarely the case: for a given country and year, inverted Pareto coefficients are generally invariant with y (Atkinson *et al.*, 2011).

The Pareto distribution appears to be a first-choice option to explain income distribution and has benefited from an increasing popularity, culminating recently with the recent literature on top incomes (Pareto 1896; Macaulay, 1922; Champernowne 1953; Tinbergen 1956; Mandelbrot 1960; Piketty 2001; Klass *et al.*, 2006). Pareto laws belong to the class of Power Laws, well known for the numerous regularities they can explain: size of human settlement, the occurrence of words in any language, CEO pays… Such empirical regularities also triggered interest in the use of power laws in economics beyond income distribution (Schumpeter 1949; Gabaix, 2009), and are still the objects of broad investigations.

- Similarly, changes in marginal tax rates could lead to changes in the income that is reported on tax records. Most studies on the subject suggest that the elasticity of taxable income with respect to changes in tax rates is positive and is typically between 0.5 and 1% (Gruber and Saez, 2002). This implies for example that, when the upper marginal income tax rate drops from (say) 40 to 30% the amount of income reported would increase by between 8% and 16%. Again, the share of income going to the top income bracket may not mean that the rich are getting richer, only that more of their income is reported on their tax declaration.
- A significant and shifting incidence of tax evasion and tax avoidance schemes.

22. Bearing in mind these limitations, tabulation-based top income figures point nonetheless to patterns that are large enough to change the picture of income distribution. They also highlight the doubts that surround the assessment of trends in the upper end of the distribution based on survey data alone, and on the overall level of inequality in countries. But they also create a divide in the measurement of income inequality, with on the one hand the (say) 99%, measured from household surveys and on the other hand the 1% measured from tax data. This situation is not satisfactory in the sense that it does not provide measures of the full extent of inequality that are ultimately required in various areas of economic analysis.

Estimation of top income shares from household surveys

Methodology

23. This paper delivers a new assessment of the income distribution, covering at the same time the top 1% and the bottom 99% in a consistent manner, across countries and over time. The methodology relies only on household survey data, which are adjusted for top incomes under-coverage by applying the West (1985) conditional likelihood estimator (Box 2).

Box 2. Conditional likelihood estimator of top incomes

The West's conditional likelihood estimator builds on two assumptions which are at the core of the top incomes literature and therefore hardly binding in this context (see Annex 1 for additional details on the derivation). The first is that the upper tail of the income distribution conforms to a Pareto law. This is the central assumption adopted for deriving top incomes shares from tax data (see Box 1). The second is that statistical inference of the Pareto parameter can rely on the observed income distribution which is censored beyond a specific upper-tail threshold, after which incomes are unobserved.

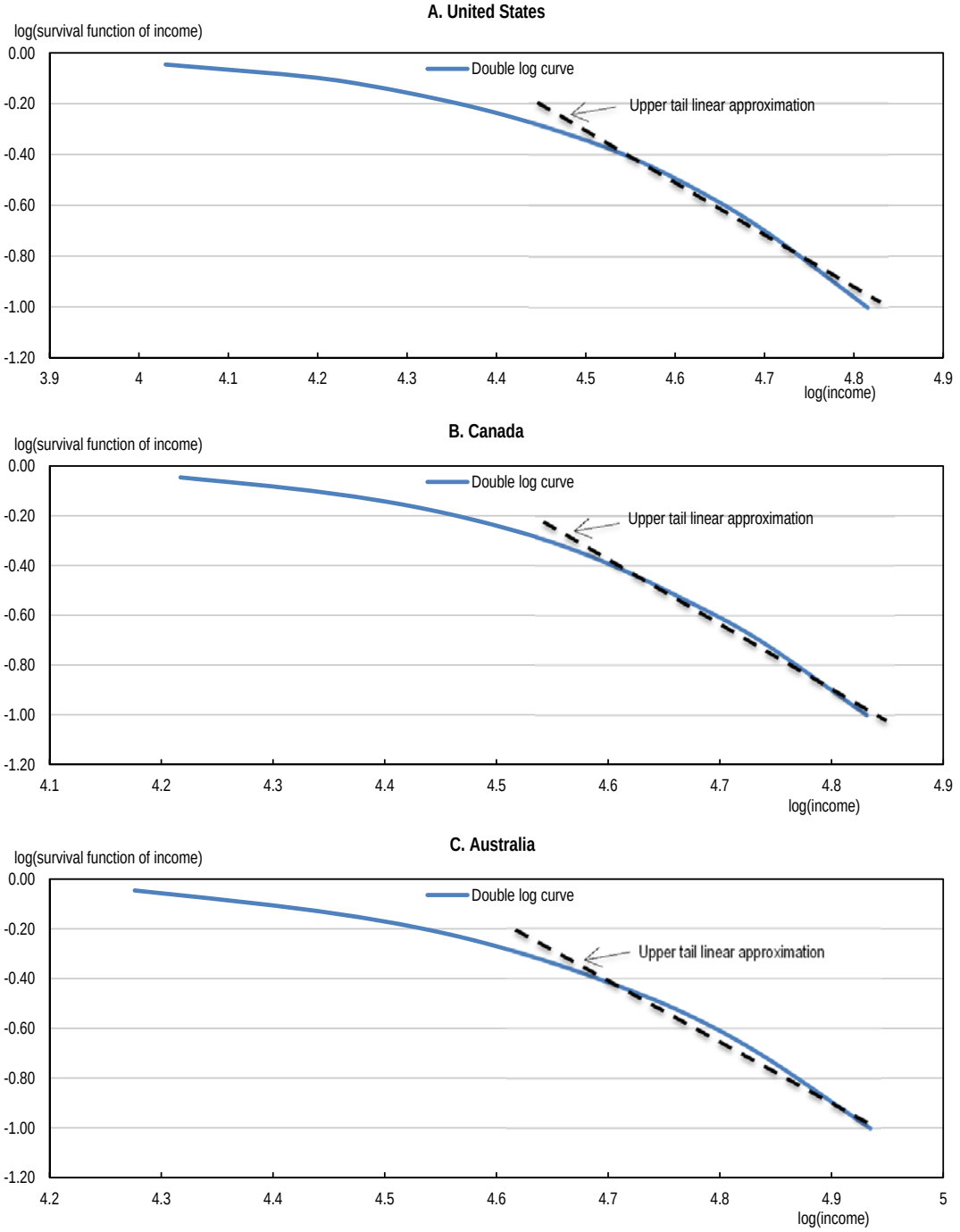
In practice, this second condition means that at least part of an observed income distribution, that is, below the upper-tail threshold above which top incomes are not observed (or partially) in household surveys, also displays some degrees of similarity with a Pareto law, so as to minimize measurement errors. Such conformity can be checked at least visually by plotting the logarithm of the survival function (the function that describes the probability that income takes on a value greater than specified thresholds) of the income distribution against the logarithm of the associated thresholds' values. If the income distribution follows a Pareto distribution, then the resulting plot is a straight line with its slope equal to the law's parameter (see Mandelbrot, 1960). The Figure below suggests that this requirement can be met in practice, with illustrative examples from Australia, Canada and the United States in 2010*.

The visual analysis of survey-based income distribution data suggests that the plot above the median can be fairly approximated by a straight line (the dashed lines being some representation of such approximations in the Figure). This is a fundamental rationale behind the application of the West estimator: some parts of observed income distributions can be used to infer Pareto parameters without any prior observation of top incomes. Pareto parameters can then be used to derive top incomes series, similarly to the approach initiated by Piketty and Saez (2003).

However, regarding the selection of the truncation threshold M_s, no specific rule applies (see Hill, 1975). It has been found in West (1985) that truncation at the interval containing the average of the original distribution gives the most satisfactory results (when comparing the estimated mean in the upper part with the true mean). However, across countries and time, this paper finds that truncation at the median gives the closest estimates of the Pareto parameters available in the WTID. Since one of the main objectives of this paper, is to conform to the WTID, this threshold has been retained.

Box 2. Conditional likelihood estimator of top incomes (*continued*)

Pareto laws double-log curves for selected OECD countries (around 2010



Note: The double log curve is the log of the survival function of an income distribution expressed in function of log(income). In the case of a Pareto distribution, the curve is linear. For the three countries considered, the curves are not linear globally, nor in their right-ends. However, as indicated by the dashed lines, the right-ends can be reasonnably approximated by a straight line.

---

* Figures for the 34 OECD countries over the period 1985-2011 are available upon request to the authors.

24. In the literature, the aim of adjusting household surveys for upper-tail mismeasurement is not new.[11] However, the previous studies (essentially produced before the rise of the recent top incomes literature) proceed in specific ways that cannot be easily generalised. In fact, in most cases the quality of the adjustment is assessed from the comparison with data unavailable to a general audience, e.g. micro data internal to National Statistical Offices (Jenkins *et al.*, 2011; Burkhauser *et al.*, 2012), or administrative records that can be accessed only by government officials (Fichtenbaum and Shahidi, 1988). Thus, the generalisation across countries and time appears problematic, as this approach to gauge the validity of the estimations cannot be easily conducted. . Other studies make adjustments by assuming an ad-hoc Pareto parameter applied beyond a certain threshold (such as in Autor *et al.*, 2008), e.g. incomes above a specific threshold are assumed to be equal to 1.5 or 2 times the threshold. Such an approach is also obviously limited given the arbitrary setting of parameters involved.

25. Surprisingly, none but one paper (Burkhauser *et al.*, 2012) undertook the task of estimating top incomes from the US income survey and then comparing the resulting adjustment with the WTID. There is little contention that the WTID is now the standard against which any top income estimates have to be compared. The WTID has also the advantage of being fully accessible to any users. But as stated above, Burkhauser *et al.* (2012) based their estimates on internal information from the Census Bureau, which precludes the possibility of extending their adjustment framework to other countries. In this paper, the methodology used is applicable in more general settings: in what follows figures from household surveys, generally easily available, are used.

Data sources

26. As highlighted above, the IDD gathers income distributions statistics over the 34 OECD countries since the mid-70s. It is based on the central concept of "equivalized household disposable income", i.e. the total market income received by households (gross earnings, self-employment and capital income) plus transfers less the current taxes they pay, adjusted for household size (with an equivalence scale where all incomes are divided by the square root of the household size). The method of data collection used for the IDD aims to maximize international comparability as well as inter-temporal consistency of data. This is achieved by a common set of protocols and statistical conventions, e.g. on income concepts and components, used to derive estimates. As a result, the IDD is less subject to the usual limitations of secondary datasets. It fails however to measure properly top incomes and such issue underlies the adjustment procedure presented in this paper.

27. In this section, the income concept used to produce top income share estimates from the IDD is pre-taxes and post-transfers. This concept is what comes closest to the income definition used in the WTID. This is however not entirely satisfactory in the sense that numerous methodological differences remain between household survey-based and tax return-based definitions of income – as well as differences between underlying units of observation. However, the room for methodological alignment is very narrow in practice and differences in concepts are unavoidable (see Leigh, 2007; Atkinson *et al.*, 2011; Burkhauser *et al.*, 2012; Armour *et al.*, 2013; OECD, 2014a).

Estimation and comparison with the WTID

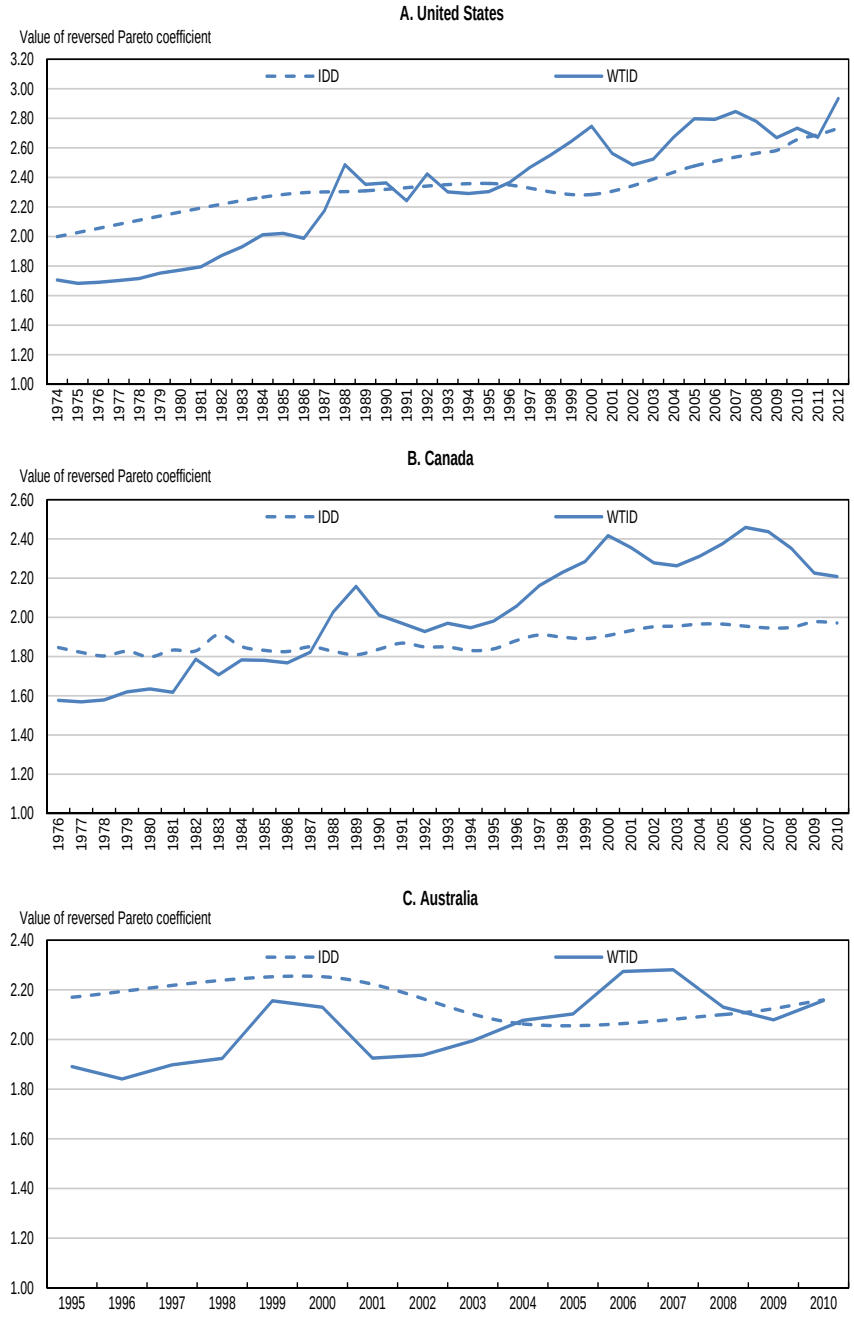
28. The estimation builds on the assumption that income distributions reported in the IDD fail to capture the entirety of the top decile, i.e. households with incomes greater than the last decile threshold are censured to that threshold[12]. The inverted Pareto coefficients are derived using the West estimator and the

11. See among other references West, 1985; Fichtenbaum and Shahidi, 1988; Katz and Murphy, 1992; Juhn *et al.*, 1993; Bishop *et al.*, 1994; Autor *et al.*, 2008; Jenkins *et al.*, 2011; Burkhauser *et al.*, 2012.

12. Despite the fact that the IDD reports in practice the average income value for the last decile, it remains a reasonable assumption in the sense that at least the top 1% are not covered properly by the data while they

resulting Pareto laws are applied subsequently on the last decile of each distribution. Incomes of the remaining 90% of the population are assumed to be captured without errors by household surveys: only the top 10% and above are subject to estimation in this paper.

Figure 1. Reversed Pareto Coefficient : comparison between IDD and WTID data for selected OECD countries



account for most of the income amount of the last decile (and of the income amount of a distribution in general).

29. Pareto coefficients derived by applying the West estimator to household survey data are in line with those reported in the WTID (Figure 1). This is particularly the case for the United States, the most widely and comprehensively studied country in the top incomes literature. Some discrepancies nonetheless occur, but by no means does this imply any systematic and significant bias – be it across countries and over time.

30. Household survey-based top income shares can be computed from estimated Pareto coefficients[13]. Figure 2 presents the top 1% series estimated from the OECD data over the set of countries common with the WTID. In most cases, survey-based series are consistent with tax-based series, particularly so for Australia, France, Italy, Japan, the Netherlands, Portugal, the United States, Spain and Sweden. Norway is a notable exception, because WTID data for this country display an erratic pattern after the early-1990s; before however, estimates are again very close. Survey-based series are also generally smoother than the WTID series. This is expected in the sense that survey methodologies in OECD countries are to a large extent stable over time while, as highlighted above, tax data are subject to factors such as tax shifting, tax avoidance and changes in tax legislation that are likely to make the top income series erratic.

13. The formulas for the derivation of top incomes summary statistics based on Pareto distributions can be found in Atkinson and Piketty (2007).

Figure 2. Top 1% income shares : comparison between IDD and WTID data for selected OECD countries

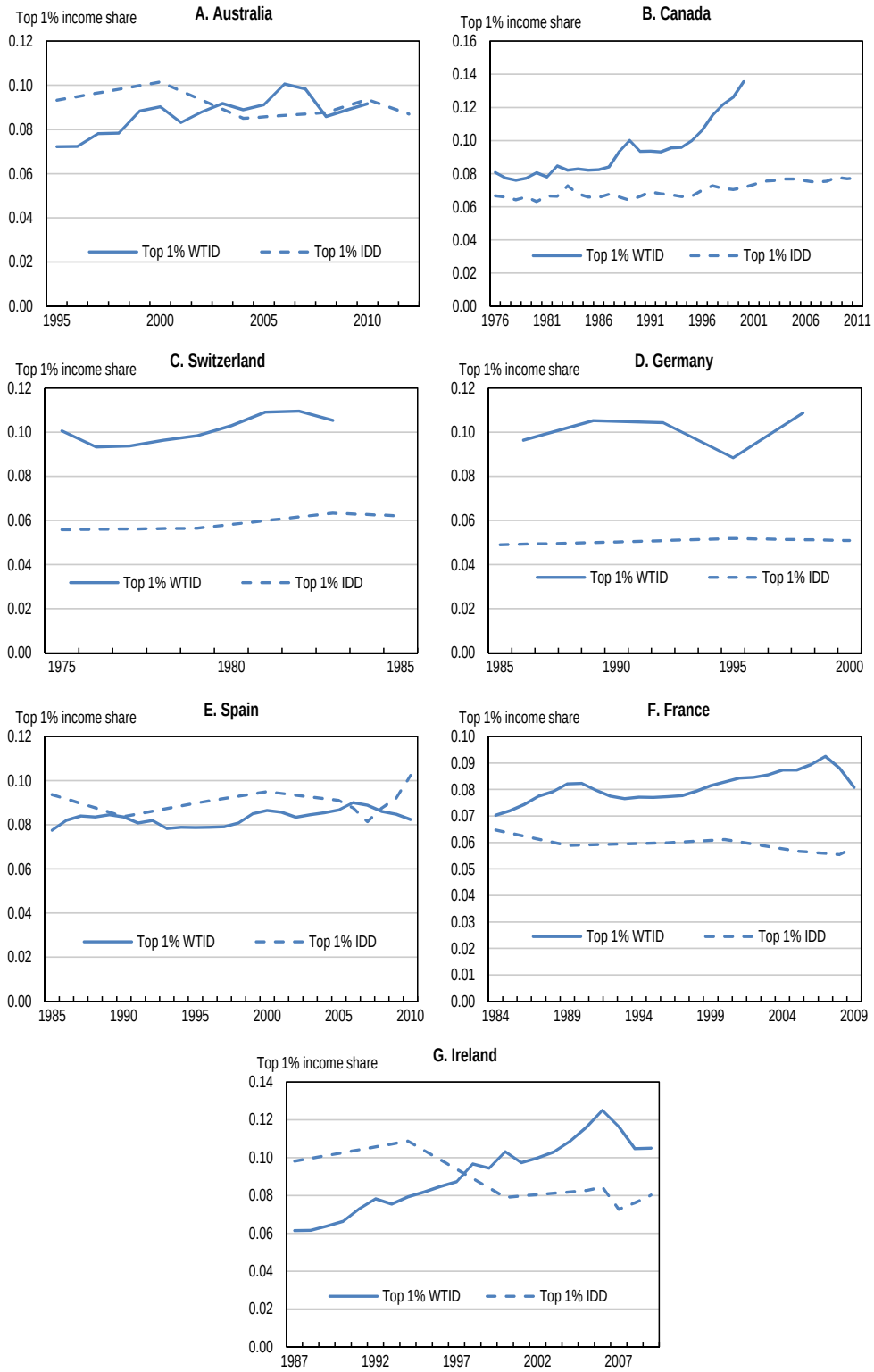
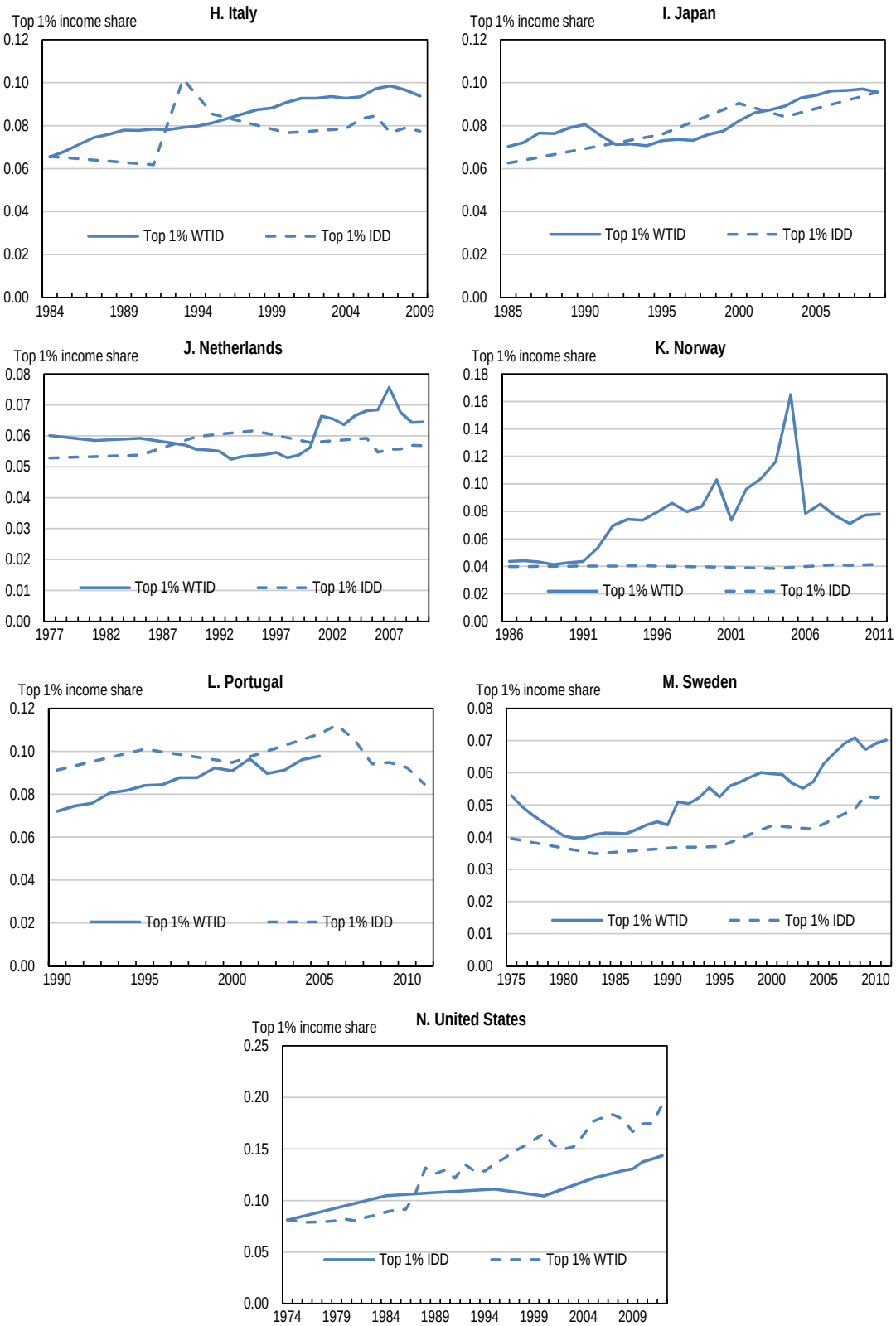


Figure 2. Top 1% income shares : comparison between IDD and WTID data for selected OECD countries



31. Table 2 delivers a synthetic comparison of top income shares estimated from OECD data and the WTID, covering both the cross-section and time-series dimensions. For countries common to both sources in 2010, the average difference in the share of income accruing to the top 1% is less than 2 percentage points; on average between the mid-1980s and 2010, the two top-income share series exhibit the same trend direction 70% of the time[14].

Table 2. Summary of the differences between estimated top incomes series and the World Top Incomes Database (for OECD countries common to both sources in 2010)

Differences in levels (2010)	
Countries	Differences in top 1% income shares (in percentage points)
AUS	0.01%
ESP	2.00%
FRA	2.10%
IRL	2.10%
ITA	1.80%
JPN	0.10%
NLD	0.83%
NOR	3.71%
SWE	0.35%
USA	1.80%
Average	1.48%
Trend comparison (between mid-180s and 2010)	
Proportion of common trends between estimated series and WTID (between mid-1980s and 2011)	70.22%

14. Trend directions are measured by computing moving averages of order 6 of first-differenced series. In other words, the idea is to compare the structural evolutions of top 1% incomes shares series, which is done by filtering-out short term variations. This is done for every OECD country available in the WTID but the time horizon differs depending on data availability.

Top income shares in all OECD countries

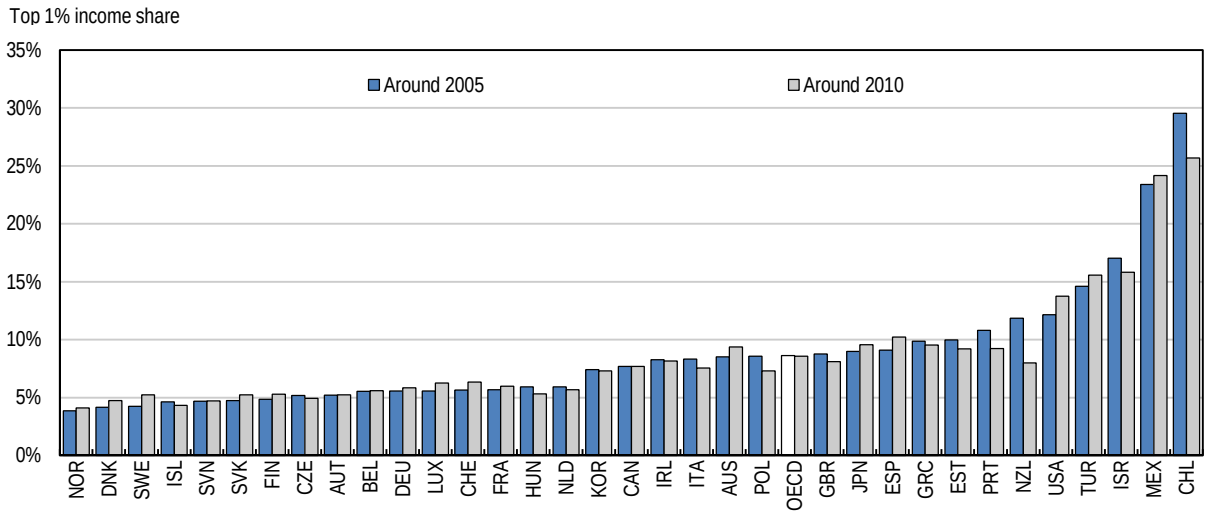
32. The methodology developed above allows for estimating top incomes across the 34 OECD countries, which expands the country coverage of the WTID. This can be used to shed light on recent trends in top incomes, comparing the pre-crisis period with the latest data point, generally 2010:

- The share of income held by the top 1% has been on the rise in about half of OECD countries between 2005 and 2010, but fairly modestly (Figure 3). The most significant increase occurred in the United States, where top incomes have been rebounding after a temporary decline over the crisis.[15]

- Some OECD countries have been experiencing significant declines in top income inequality: this is notably the case in Chile, where the share of income held by the top 1% dropped by 4 percentage points. Still, it remains, at more than 25%, the highest across the OECD.

- Overall, this new evidence delivers a more nuanced view on the evolution of top incomes in the aftermath of the crisis. This is due to a wider country coverage compared to the one that can be achieved by tax-based tabulation (see Piketty and Saez, 2013 for the estimation of top incomes based on the latest vintage of the WTID).

Figure 3. Top 1% income shares in OECD countries (2005 and 2010)



Income inequality in OECD countries based on adjusted income distribution measures

33. The estimations of top income shares can now be used to derive new, adjusted income distributions and inequality measures, where all the population is taken on board, i.e. the 99% and the 1%. Such assessment is now based on household disposable income (i.e. income post-transfers and taxes) as opposed to taxable income in the last section. Household disposable income is conceptually most representative of the material conditions experienced by household (see Stiglitz et al., 2009).

15. See Piketty and Saez (2013) for a similar characterisation.

34. The adjustment proceeds as in the last section: IDD data are directly used to measure incomes for the bottom 90% of the population. Statistics for the richest 10% are retrieved from Pareto estimations and adjusted for the missing top incomes, for each year and country, applying the above-presented West estimator. This results in a new quantification of the levels and evolutions of inequality in OECD countries.

35. It should be noted that using secondary datasets rule out the possibility to consider alternate cut-offs, since the latter would require working with the original micro data for probing more deeply into the Pareto property. In principle, this could result in a loss of precision in the computation of the Pareto parameter[16]. In practice, assuming that the top 10% of an income distribution conforms to a Pareto law is a rather weak assumption given the empirical regularities of such laws identified in the literature as well as in the income distributions used in this paper over 34 countries and almost 30 years.

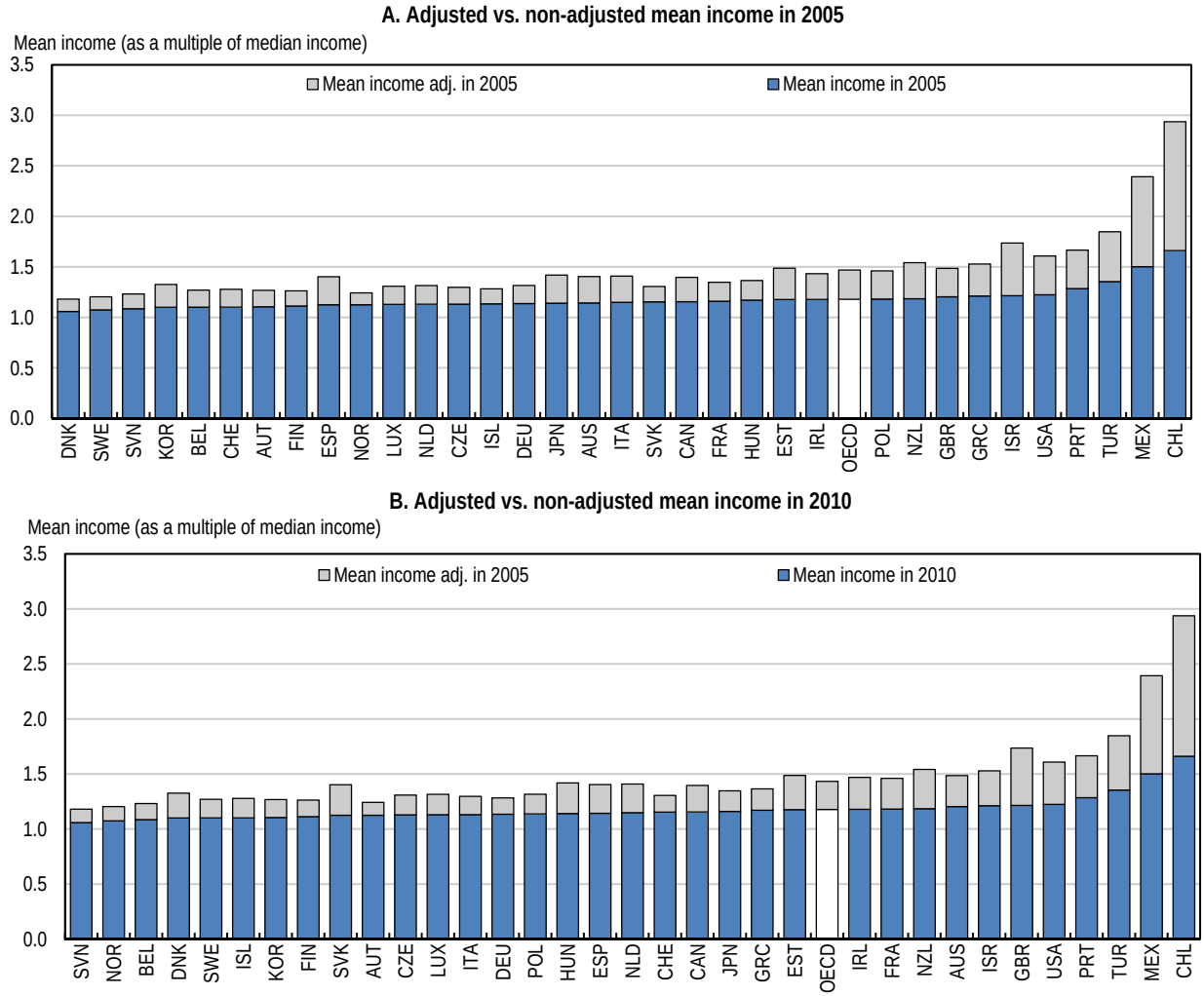
Effects of the adjustment on mean incomes

36. Adjusting income distributions for top incomes results in an up-scaling of mean income, which makes income distributions significantly more skewed (Figure 4). For instance, in 2010, the gap between mean and median incomes was 1.2 on average across the OECD; when adjusted for top incomes, this gap is re-evaluated at 1.5[17] (from 1.1 to 1.4 in 2010). The most notable changes occur in Chile and Mexico and to a lesser extent in Israel, Turkey and the United States. Generally, the more unequal a country is within the 99%, the larger is the adjustment for the top 1%.

16. Resolving such imprecisions is a path for future research (see for example Fairfield and Jorrat (2014) for an attempt), notably when a broader access to micro data among several countries will be made possible. But as stated in the introduction the primary perspective of this paper is cross-national comparisons, which necessitates an almost exclusive use of semi-aggregated inequality figures.

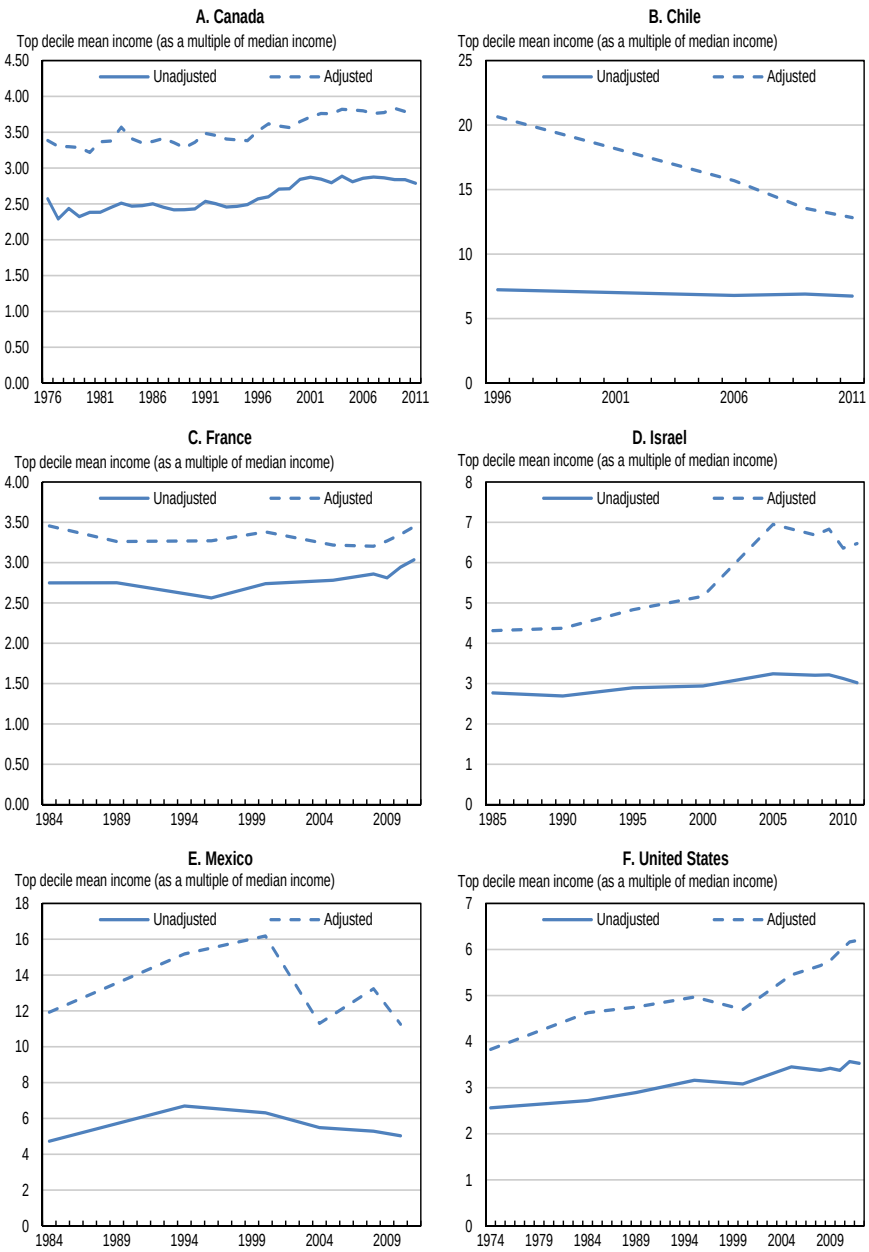
17. A result of up-scaling is that micro-based mean incomes figures align better with macro-based National Accounts mean incomes figures. This allows for achieving a more accurate micro-based estimate of mean household income. See Deaton (2005) and Causa et al., (2014), for a short summary of micro versus macro-based mean household income measures

Figure 4. Adjusted vs. non-adjusted mean income (as a multiple of median income)



37. By construction the adjustment occurs in the last decile of each income distribution. The adjusted mean income of the richest 10% thus provides an alternative way to gauge the magnitudes implied by the top incomes correction. Over time, there is no clear pattern in the magnitude of the adjustment (Figure 5 and Annex 2): it can continuously increase (Israel), decrease (Chile), remain broadly stable (Canada) or having a more erratic pattern (Mexico). The adjusted series disclose in fact some varying trends across countries, which were masked by the unadjusted series. For instance, in Chile the gap between the mean income of the richest 10% and the median income remains constant over time, at around 7 between the mid-1990s and 2011 for the unadjusted series, while the adjusted series reveals a declining trend, from 20 to 12. The reverse occurs for Israel: while the unadjusted ratio remains close to 3 between the mid-1980s and 2011, the adjusted ratio displays a marked increase, from 4.1 to 6.5.

Figure 5. Adjusted vs. unadjusted top decile mean income for selected OECD countries (as a multiple of median income)



Inequality across the 100%

38. Adjusting income distributions for top incomes allows for a reassessment of the official levels of inequality across countries. Summary measures of income dispersions, such as the Gini coefficient, risk under-estimating the real size of inequality when applied to unadjusted series, as the latter basically fail to cover top incomes. Indeed, the main findings from this adjustment can be summarised as follows:

- The Gini coefficient increases by 6 percentage points, from 0.31 to 0.37, on average across OECD countries when the data are adjusted for top incomes (Figure 6). Stated otherwise, the top incomes-adjusted level of inequality in OECD countries is on average equivalent to the

level of inequality prevailing in the United States before the adjustment. This magnitude is in line with earlier, specific adjustments of the Gini coefficient for top incomes (Atkinson and Piketty, 2007; Alvaredo, 2011).

- The ratio of mean incomes of the richest to the poorest 10% increases from 10 to 15, on average across OECD countries. This inequality measure is most sensitive to top incomes and the average adjustment masks substantive heterogeneity across countries: the incomes of the richest 10% are 6 times higher than those of the poorest 10% instead of 5 times higher in Denmark – a relatively negligible under-estimation; at the opposite end of the spectrum, the incomes of the richest 10% are 64 times higher than those of the poorest 10% instead of 29 times higher in Mexico – a dramatic under-estimation (Figure 6).

- Country rankings remain however broadly unaltered following the adjustment, whatever the indicator considered.

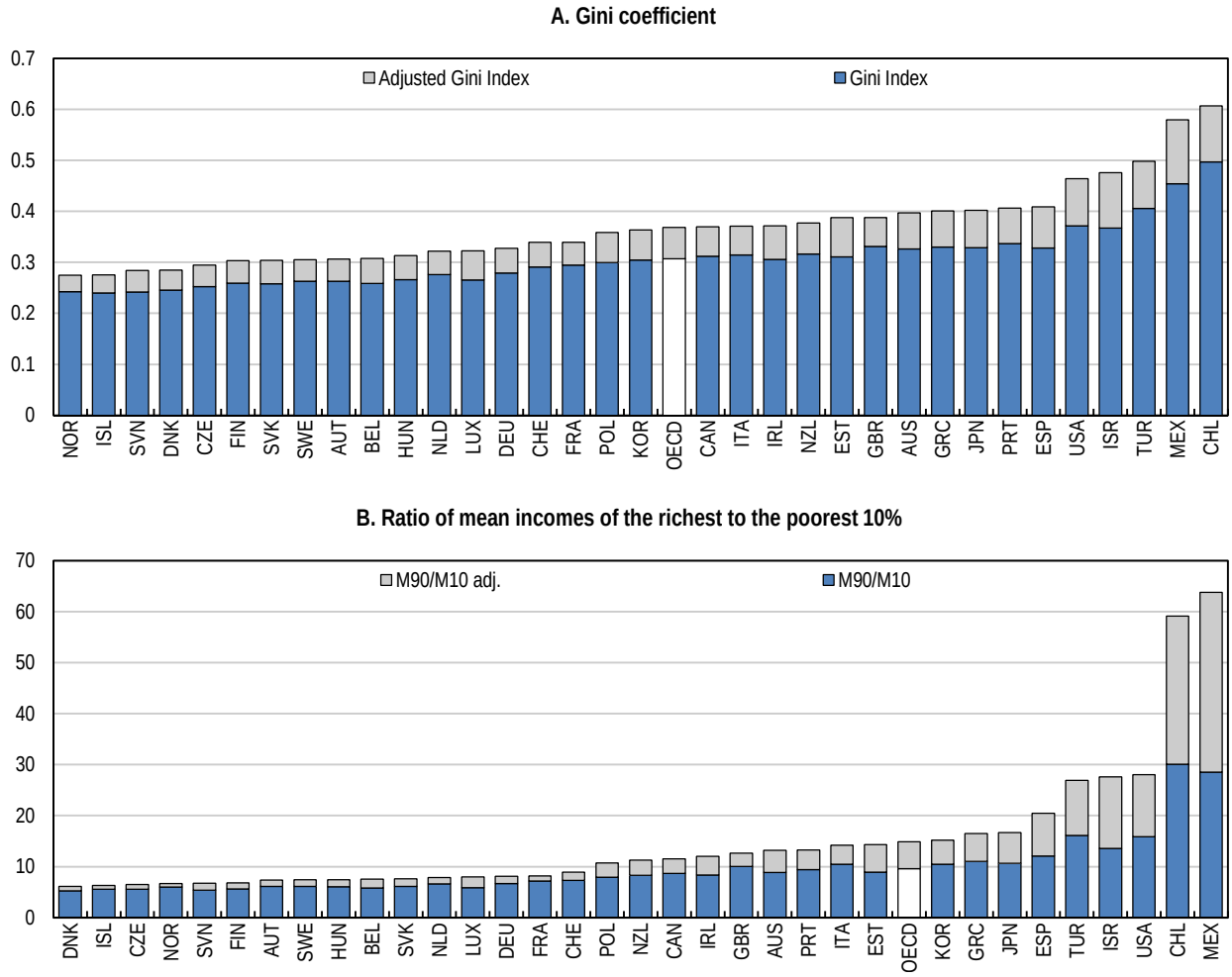
39. Adjusting income distributions for top incomes allows also for a reassessment of the evolution of inequality. Such assessment is performed here on the basis of a more comprehensive and granular approach for measuring inequality than that based on summary indexes, i.e. through general mean curves[18]. This allows for visualizing the broad evolution of inequality but also associated profiles, i.e. the sources and patterns of changing inequality within each country. Main findings from this adjustment can be summarised as follows, focusing on the period between mid-1990s to late 2000s (Figure 7 and Annex 3):

- While the adjustment delivers automatically higher levels of inequality, evolutions will depends on how the mean income of the top 1% evolves compared to the mean incomes of the bottom 99%. Over time, no common pattern emerges. In some countries, the adjustment for the top 1% leads to a smaller rise in inequality than on the 99%. This is the case for Australia, Austria, Denmark, France, Ireland and the United-Kingdom.

- In other countries such as the Czech Republic, Hungary, Italy, the Netherlands and Portugal, the original fall in inequality is reinforced by the inclusion of the top 1%. And in countries where inequality was rising before the adjustment, the rise is also reinforced (e.g. Israel, Luxembourg, the United States, Spain and Sweden).

18. These curves follow from the use of specific income standards which summarise the income distribution by a single "representative" level of income without relying on the setting arbitrary cut-offs along the distribution. See Foster and Szekely, 2008; Causa *et al.*, 2014.

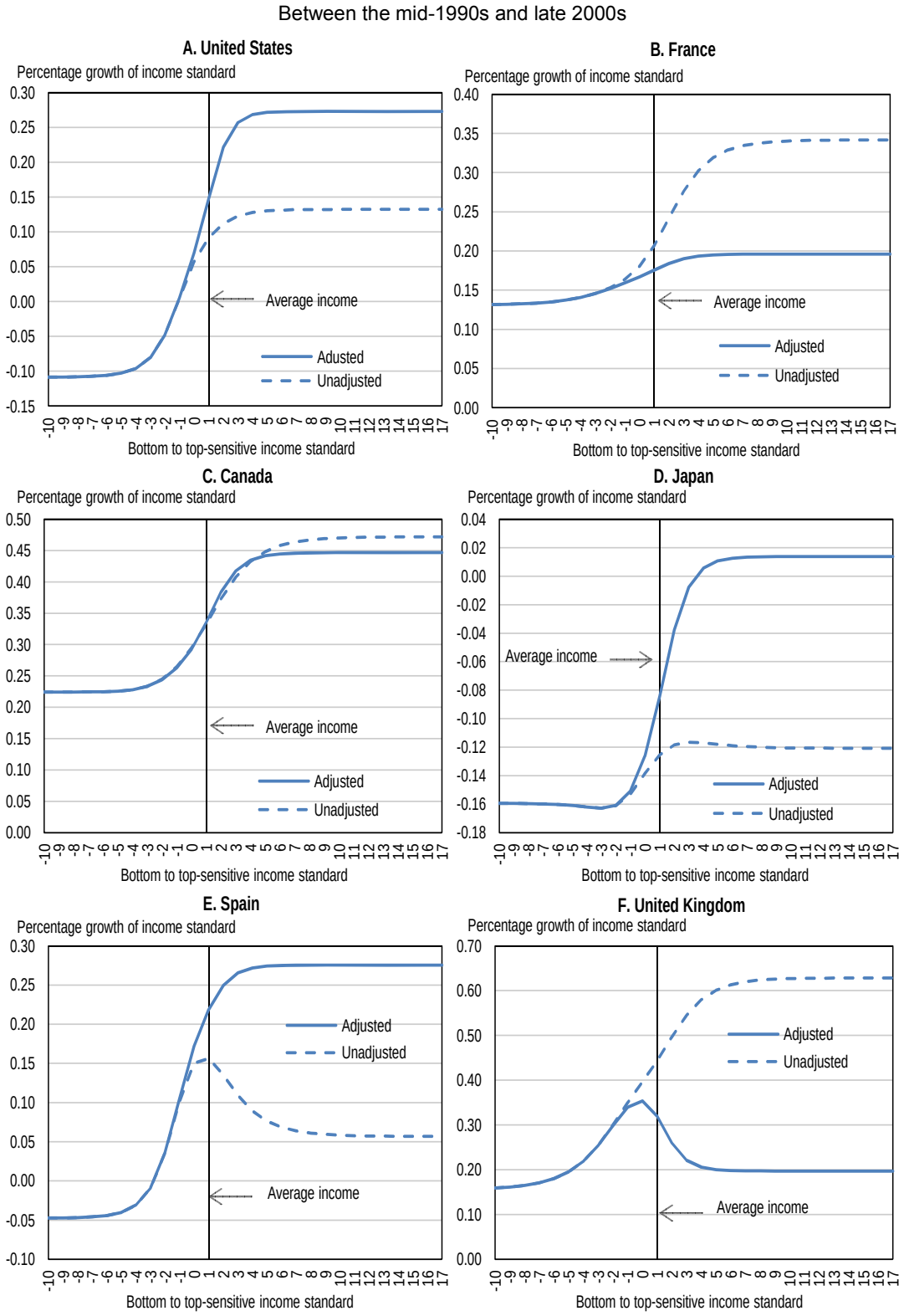
Figure 6. Impact of the adjustment on standard indexes of income inequality
Around 2010



- However, in a majority of countries trends in income inequality across the 100% remain *qualitatively* unchanged compared to those derived from the 99%. The upper portion of general means curves is tilted upward in countries where top incomes have been experiencing stronger relative income gains and downward in countries where top incomes have been experiencing weaker relative income gains. The overall profiles of the curves remain the same, indicating the same qualitative evolution of income dispersion. Nevertheless, the magnitudes involved are quite different when accounting for top income developments: the rise in inequality is much more pronounced in Japan, Spain and the United States and much less in Australia and Austria; the decline in inequality is more pronounced in Chile and Italy.

- Several factors can explain these differential developments. In particular, top incomes are generally more sensitive to the business cycle than other income groups, moving up and down faster than the rest of the population when the economy expands or contracts. When combined to relatively large welfare state, the over-sensitivity of top incomes leads them to increase slower than the rest of the population (or decrease faster than the rest), during a downturn. This appears to be the case in France, where top incomes grow slower than the mean incomes of the 99%, leading to an overall increase in inequality lower than when assessed on the 99%.

Figure 7. Growth in household income across the distribution between the mid-1990s and late 2000s : selected countries



- Other policy changes directed specifically toward the top incomes can also explain why the evolution of inequality on the whole population delivers new conclusions. For example in the United-Kingdom, the top statutory personal income tax rates has been increased from 40% to 50% in 2010 (albeit reversed partially in 2013), introducing then an additional wedge between the growth rates of the 99% mean income and the top 1%. This assuredly contributes to a more nuanced increase in overall inequality.

Conclusion

40. The evidence gathered on top incomes over the last fifteen years has changed the general perception about inequality both among specialists and the wider public and in fact this has raised awareness and attention devoted to the topic. At the same time the top income debate resulted in a more fragmented approach towards measuring and assessing inequality. The under-estimation of the level of inequality associated with the use of household surveys to characterise the income distribution is now well recognised, leading to separate analyses of top incomes. At the same time, the analysis of top income through tax return data puts the focus on a small group of the population and ignores the majority, e.g. the middle class and the poor. A comprehensive approach is needed for assessing inequality, especially at a time when international organisations are treating the reduction in income inequality as an important policy objective along with boosting incomes: such is the case of the OECD Inclusive Growth Initiative (OECD, 2014b) and the World Bank Shared Prosperity Initiative (World Bank, 2014).

41. This paper develops a new approach towards achieving a comprehensive and unified assessment of inequality, trying to put together the 99% and top 1%. The approach is applied to estimate levels and trends in inequality across OECD countries. The analysis uses information from household surveys to deliver top income estimates. Such estimates are compared and found in line with those available in the benchmark WTID database This approach is still preliminary and subject to various limitations. Nevertheless it is potentially a promising step to raise the accuracy of the data and analytical material underlying policy discussions on inclusive growth.

BIBLIOGRAPHY

Acemoglu, D., and J. A. Robinson (2014), "The Rise and Fall of General Laws of Capitalism", *Working Paper, Harvard University*.

Alvaredo, F. (2011), "A Note on the Relationship between Top Incomes and the Gini Coefficient", Economics Letters, Vol. 110, pp. 374-277.

Armour, P., R. V. Burkhauser, and J. Larrimore (2013), "Deconstructing Income and Income Inequality Measures: A Crosswalk from Market Income to Comprehensive Income", *American Economic Review*, vol. 103(3), pp. 173-177.Atkinson, A. B., and T. Piketty (2007), *Top Incomes over the Twentieth Century: a Contrast Between Continental European and English-Speaking Countries*, Oxford : Oxford university press.

Atkinson, A. B., T. Piketty, and E. Saez (2011), "Top incomes in the long run of history", *Journal of Economic Litterature*, vol. 49 (1), pp. 3-71.

Autor, D., L. Katz, and M. Kearney (2008), "Trends in U.S. Wage Inequality: Revising the Revisionists", *The Review of Economics and Statistics*, vol. 90, pp. 300–323.

Brandolini, A., and A. B. Atkinson (2001), "Promise and Pitfalls in the Use of "Secondary" Data-Sets: Income Inequality in OECD Countries as a Case Study", *Journal of Economic Literature*, vol. 39(3), pp. 771-799.

Barro, R. J. (2000), "Inequality and Growth in a Panel of Countries", *Journal of Economic Growth*, vol. 5, pp. 5-32.

Bishop, J. A., J-R. Chiou, and J. P. Formby (1994), "Truncation Bias and the Ordinal Evaluation of Income Inequality", *Journal of Business and Economic Statistics*, vol. 12, pp. 123-127.

Bourguignon, F. (2005), "Comments on: Measuring Poverty in a Growing World (or Measuring Growth in a Poor World)", *The Review of Economics and Statistics*, vol. LXXXVII, pp. 20-22.

Burkhauser, R., S. Feng, S. Jenkins, and J. Larrimore (2012), "Recent Trends in Top Incomes Shares in the United States: Reconciling Estimates from March CPS and IRS Tax Return Data", *The Review of Economics and Statistics*, vol. XCIV, pp. 371-388.

Causa, O., A. De Serres, and N. Ruiz (2014), "Can Pro-Growth Policies Lift All Boats? An Analysis Based on Household Disposable Income", *OECD Economics Department working paper,* forthcoming.

Champernowne, D. G. (1953), "A Model of Income Distribution", *The Economic Journal*, vol. 63, pp. 318-51.

Davies, J. B., and A. F. Shorrocks (1989), "Optimal Grouping of Income and Wealth Data", *Journal of Econometrics*, vol. 42(1), pp. 97-108.

Deaton, A. (1997), *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*, Baltimore, Johns Hopkins University Press for the World Bank.

Deaton, A. (2005), "Measuring Poverty in a Growing World (or Measuring Growth in a Poor World)", *The Review of Economics and Statistics*, vol. LXXXVII, pp. 1-19.

Deininger, K., and L. Squire (1996), "New Data Set Measuring Income Inequality", *World Bank Economic Review*, vol. 10, pp. 565-591.

Dollar, D., T. Kleineberg, and A. Kraay (2013), "Growth is Still Good for the Poor", *Policy Research Working Paper 6568*, The World Bank.

Dollar, D., and A. Kraay (2002), "Growth is Good for the Poor", *Journal of Economic Growth*, vol. 7, pp. 195-225.

Fairfield, T., and M. Jorratt (2014), "Top Incomes Shares, Business Profits and Effective Tax Rates in Contemporary Chile", International Center for Tax and Development Working Paper, 17.

Fichtenbaum, R., and H. Shahidi (1988), "Truncation Bias and the Measurement of Income Inequality", *Journal of Business and Economic Statistics*, vol. 6, pp. 335-337.

Foster, J. E., and M. Székely (2008), "Is Economic Growth Good for the Poor? Tracking Low Incomes Using General Means", *International Economic Review*, vol. 49 (4), pp. 1143–1172.

Gabaix, X. (2009), "Power Laws in Economics and Finance", *Annual Review of Economics*, vol. 1, pp. 255-293.

Gasparini, L. and L. Tornarolli (2015), "A Review of the OECD Income Distribution Database", *Journal of Economic Inequality*, forthcoming.

Gruber, J., and E. Saez (2002), "The Elasticity of Taxable Income: Evidence and Implications", *Journal of Public Economics*, vol. 84, pp. 1-32.

Hill, B. M. (1975), "A Simple General Approach to Inference about the Tail of a Distribution." *The Annals of Statistics,* vol. 3(5), pp. 1163-1174.

Jenkins, S. P., R. V. Burkhauser, S. Feng, and J. Larrimore (2011), "Measuring Inequality Using Censored Data: a Multiple-Imputation Approach to Estimation and Inference", *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 174(1), pp. 63-81.

Juhn, C., K. M. Murphy, and B. Pierce (1993), "Wage Inequality and the Rise in Returns to Skill", *Journal of Political Economy*, vol. 101(3), pp. 410-442.

Katz, L. F., and K. M. Murphy (1992), "Changes in Relative Wages, 1963-87: Supply and Demand Factors", *Quarterly Journal of Economics*, vol. 107, pp. 35-78.

Klass, O. S., O. Biham, M. Levy, O. Malcai, and S. Solomon (2006), "The Forbes 400 and the Pareto Wealth Distribution" *Economics Letters* vol. 90(2), pp. 290-295.

Leigh, A. (2007), "How Closely Do Top Incomes Shares Track Other Measures of Inequality?", *The Economic Journal*, vol. 117, pp. F589-F603.

Macaulay, F. Robertson (1922), *The Personal Distribution of Income in the United States*, Harcourt Publishing, New York.

Mandelbrot, B. (1960), "The Pareto-Levy Law and the Distribution of Income", *International Economic Review*, vol. 1(2), pp. 79-106.

OECD (2011), *Divided we Stand: Why Inequality Keeps Rising*, OECD Publishing, Paris.

OECD (2014a), *Focus on Top Incomes and Taxation in OECD Countries: Was the crisis a game changer?*, OECD Publishing, Paris.

OECD (2014b), *All on Board: Making Inclusive Growth Happen*, OECD Publishing, Paris.

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

Pareto, V. (1896), *La Courbe de la Répartition de la Richesse*, writings by Vilfredo Pareto collected by G. Busino, Librairie Droz, 1965, pp. 1–15.

Piketty, T. (2001), *Les Hauts Revenus En France Au XXème Siècle*, Éditions Grasset, Paris.

Piketty, T., and E. Saez (2003), "Income Inequality in the United States, 1913-1998", *Quarterly Journal of Economics*, vol. 118(1), pp. 1-39.

Piketty, T., and E. Saez (2013), "Top incomes and the Great Recession", *IMF Economic Review*, vol. 61, pp. 456-478.

Schumpeter J. (1949), "Vilfredo Pareto (1848-1923)", *Quarterly Journal of Economics*, vol. 63, pp. 147-172.

Stiglitz, J.E., A. Sen and J.-P. Fitoussi (2009), *Report by the Commission on the Measurement of Economic Performance and Social Progress*.

Tinbergen, J. (1956), "On the Theory of Income Distribution", *Weltwirtschaftliches Archiv*, vol. LXXVII, pp. 155-75.

World Bank (2014), *Ending Poverty and Sharing Prosperity*, World Bank Group Edition.

West, S. (1985), "Estimation of the Mean from Censored Earnings Data", *Proceedings of the Section on Survey Research Methods*, American Statistical Association, pp. 665-670.

ANNEX 1: CONDITIONNAL ESTIMATE OF AN INCOME DISTRIBUTION UPPER TAIL FROM HOUSEHOLD SURVEY

Formulation of the problem

42. The distribution of income across N households is denoted by:

$$X_1, X_2, \dots, X_N$$

and $X_{(1)}, X_{(2)}, \dots, X_{(N)}$ denote the same distribution ranked in increasing order of magnitude. Data are observed under the form of r mutually exclusive (and exhaustive of the income scale) intervals $I_1, I_2, \dots, I_r$, with associated frequencies $f_1, f_2, \dots, f_r$ and average-points $m_1, m_2, \dots, m_r$. As incomes reported are generally positive in household surveys, I_1 has a natural lower bound of zero and contains all the $X_{(j)}$ less than a fixed threshold U_1. I_r contains all the $X_{(j)}$ greater than or equal to a fixed threshold $U_{(r-1)}$ but has no upper bound: incomes can go up to very large values. For i=2,3,…,(r-1), I_i are all bounded intervals containing all the $X_{(j)}$ greater than or equal to $U_{(i-1)}$ but less than $U_{(i)}$.

43. A usual estimator of the mean of the income distribution is the grouped mean:

$$\overline{X}_G = \frac{\sum_{i=1}^{r} f_i m_i}{N}$$

In practice, $\overline{X}_G$ is biased downward in household surveys given the fact that m_r is underestimated: very rich household are not observed. m_r can be estimated through the fitting of a law in the upper tail of an income distribution truncated on the right.

44. The choice of a Pareto law as the parametric assumption to infer upper tail behaviour is quite natural and popular when dealing with income distribution. It is however unlikely that this parametric form holds globally. As a result, in practice the fitting can be performed on at most (but generally less) r-1 average points and an estimate of m_r can be derived as the mean of the conditional distribution $P(X \leq x | X > U_{(r-1)})$. $\widehat{m}_r$, an estimate of m_r, is thus given by:

$$\widehat{m}_r = \int_{U_{(r-1)}}^{\infty} x dP\left(X \leq x \middle| X > U_{(r-1)}\right) = \int_{U_{(r-1)}}^{\infty} x\, f(x|U_{(r-1)})$$

where $f(x|U_{(r-1)})$ denotes the conditional density of X given that X is greater than or equal to $U_{(r-1)}$.

45. By considering a Pareto distribution given by the following cumulative distribution function:

$$F(x) = P(X \leq x) \begin{cases} = 1 - \left(\frac{K}{x}\right)^{\alpha} & \forall x \geq K > 0 \; and \; \alpha > 0 \\ = 0 \; \forall x < K & \end{cases}$$

and by noting that:

$$P(X \geq x | X \geq U_{(r-1)}) = {U_{(r-1)}}^{\alpha} x^{-\alpha}$$

one obtains the following expression for the conditional density of X given that X is greater than or equal to $U_{(r-1)}$:

$$f\left(x\middle|U_{(r-1)}\right) = \alpha {U_{(r-1)}}^{\alpha} x^{-\alpha-1} \; \forall x \geq U_{(r-1)}$$

It results the following expression for $\widehat{m}_r$:

$$\widehat{m}_r = U_{(r-1)} \frac{\alpha}{\alpha - 1}$$

The parameter α of the Pareto law remains to be estimated.

The West estimator

46. Many methods exist to estimate the parameter α, mainly but not exclusively regression techniques or quantile estimation methods (see Gabaix, 2009 for a review). In this paper, it has been found however that the use of a conditional likelihood estimator developed by West (1985) gives superior results compared to regressions and quantile estimations, in the sense that the estimates produced by the estimator display strong similarities with the estimates of the Pareto coefficients available in the WTID[19]. To the best of an extensive literature review, it has been used only in the original paper that presents the estimator.

47. The West estimator is derived from two specific contribution patterns to a likelihood function:

- Some households with income contained in one of the bounded intervals $\mathrm{I}_1, \ldots, \mathrm{I}_{\mathrm{r}-1}$ contribute to the likelihood function with the value reported in the household survey.

- Each household with an income above the threshold $\mathrm{U}_{(\mathrm{r}-1)}$ contributes to the likelihood function with an observation indicating that its income, unobserved in this case, is at least equal to $\mathrm{U}_{(\mathrm{r}-1)}$.

Analytically this gives:

$$\begin{cases} \forall\, 0 < K < X_j < U_{(r-1)} \; f\left(X_j\right) = \alpha K^{\alpha} {X_j}^{-\alpha-1} \; with \; \alpha > 1 \\ \forall\, X_j \geq U_{(r-1)} \; f\left(X_j\right) = 1 - F\left(U_{(r-1)}\right) = \left(\dfrac{K}{U_{(r-1)}}\right)^{\alpha} \end{cases}$$

The likelihood function L is thus:

$$L = (N^N - f_r) \prod_{i=1}^{r-1} \left[\frac{\alpha K^{\alpha}}{X_i^{\alpha+1}}\right]^{f_i} \left[\frac{K}{U_{(r-1)}}^{\alpha}\right]^{f_r}$$

which leads to the following maximum likelihood estimator $\hat{\alpha}$ of α:

19. Estimated Pareto coefficients using regressions and quantile estimators over the 34 OECD countries are available upon request to the authors.

$$\hat{\alpha} = \frac{N - f_r}{\left[\sum_{i=1}^{r-1} f_i \log(X_i)\right] - [N - f_r]\log(\widehat{K}) - f_r \log\left(\frac{\widehat{K}}{U_{(r-1)}}\right))}$$

with $\widehat{K}$ equal to the minimum income in the distribution.

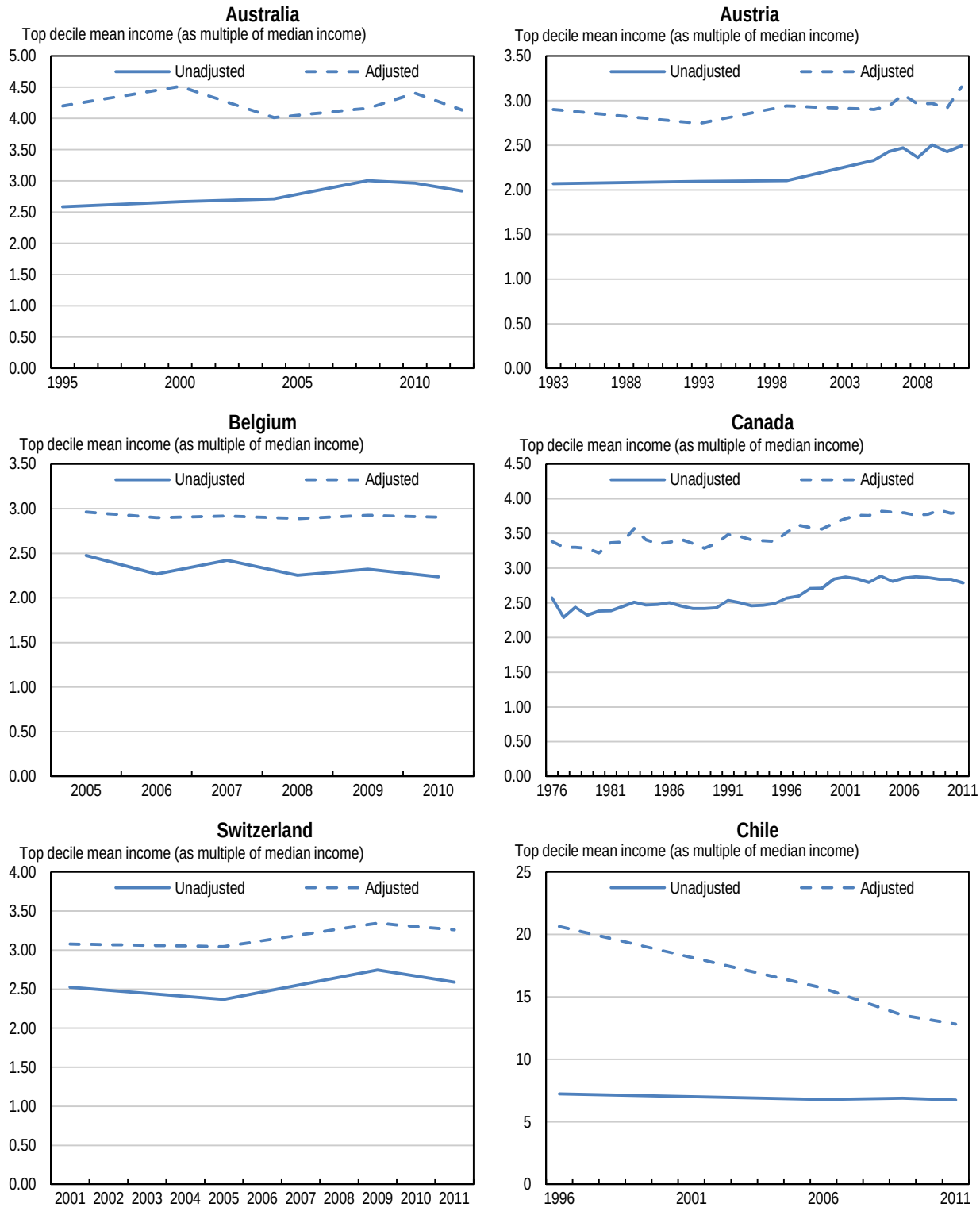
48. Given that generally only the upper-end of an income distribution is Pareto-distributed, in practice the estimation has to be performed on a left-truncated version of the distribution. Assuming the income distribution is truncated on the left at M_s, the modified maximum likelihood estimator thus becomes:

$$\hat{\alpha} = \frac{\sum_{i=s}^{r-1} f_i}{\left[\sum_{i=s}^{r-1} f_i \log(X_i)\right] - \sum_{i=s}^{r-1} f_i * \log(X_s) - f_r \log\left(\frac{X_s}{U_{(r-1)}}\right)}$$

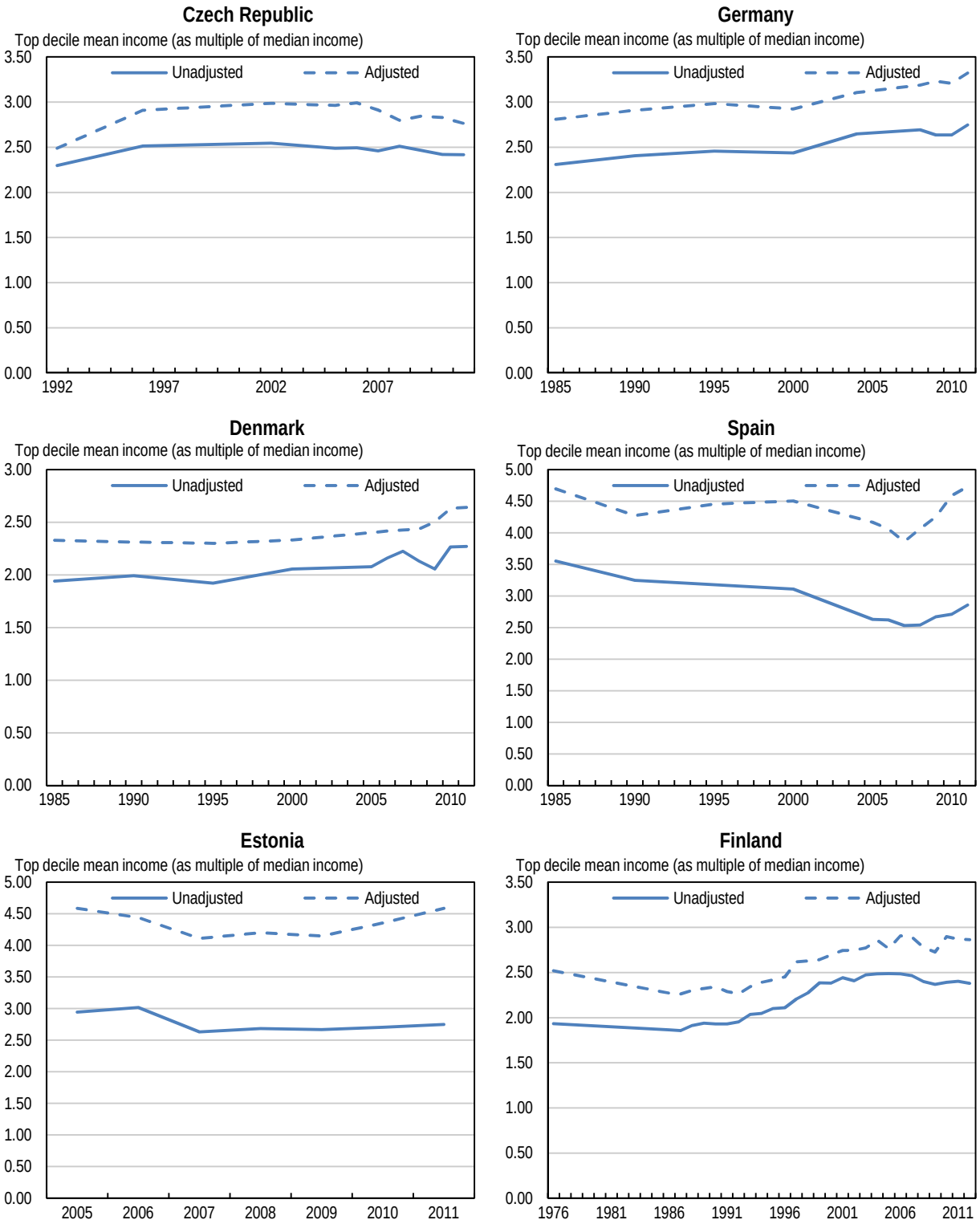
Regarding the selection of the truncation threshold M_s, there is no specific rule to apply (see Hill, 1975). It is found in West (1985) that truncation at the interval containing the average of the original distribution gives the most satisfactory results (when comparing the estimated mean in the upper part with the true mean). However, across countries and time, this paper finds that truncation at the median gives the closest estimates of the Pareto parameters available in the WTID.

ANNEX 2: ADJUSTED VS. NON-ADJUSTED RATIOS OF MEAN INCOME OF THE RICHEST 10% TO THE MEDIAN IN OECD COUNTRIES

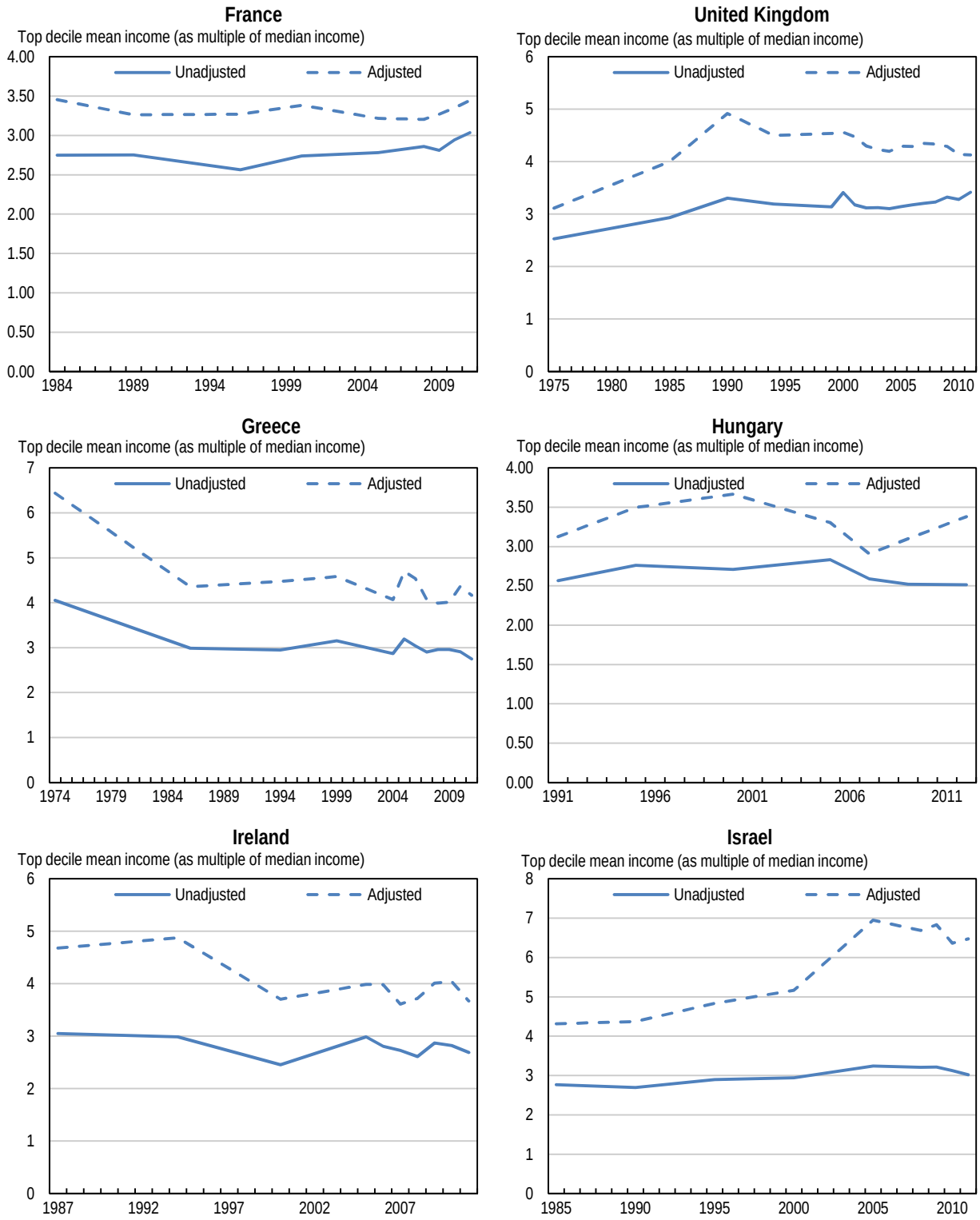
Annex 2. Part A



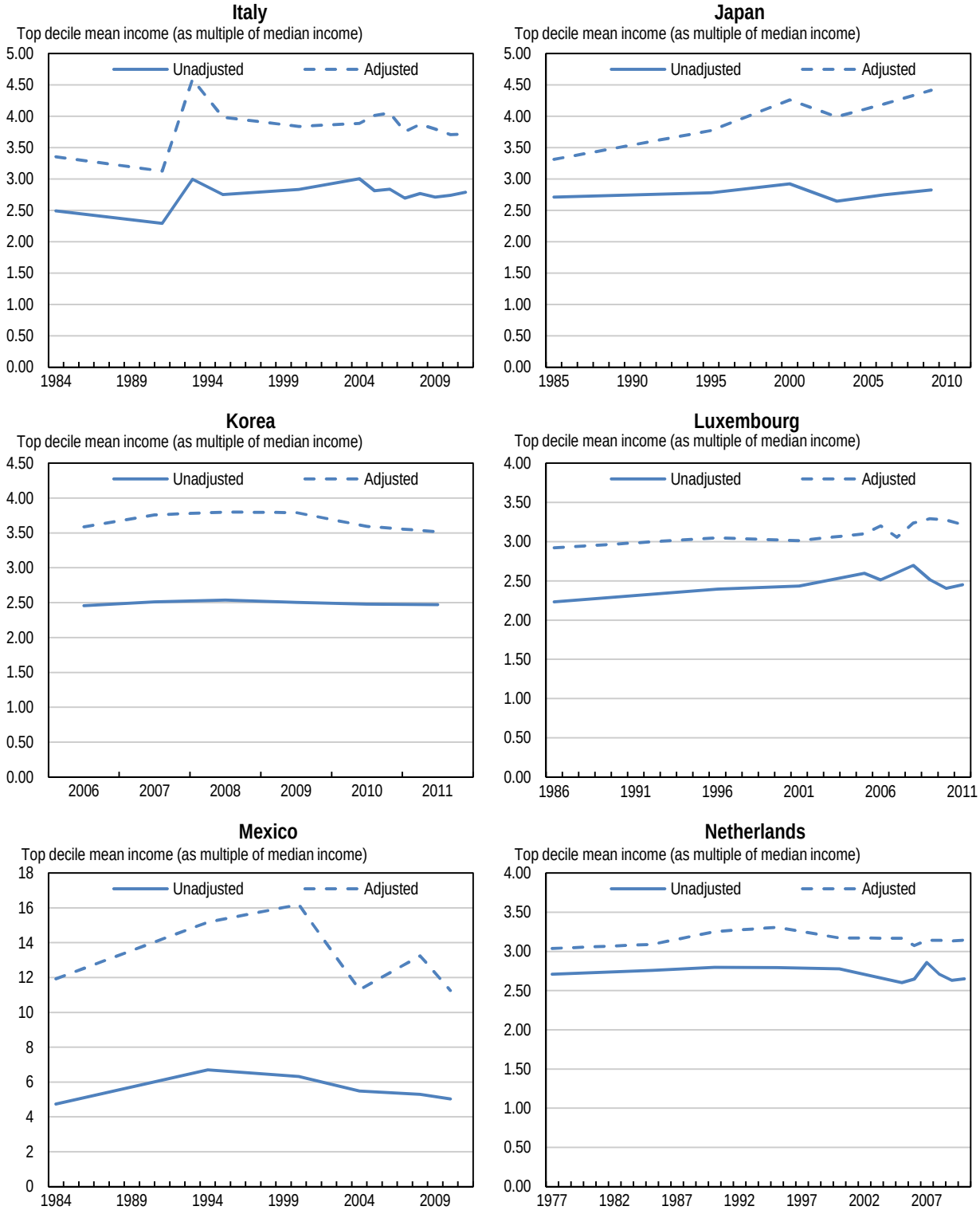
Annex 2. Part B



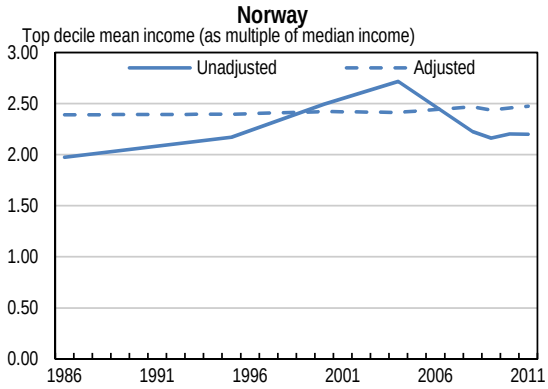
Annex 2. Part C

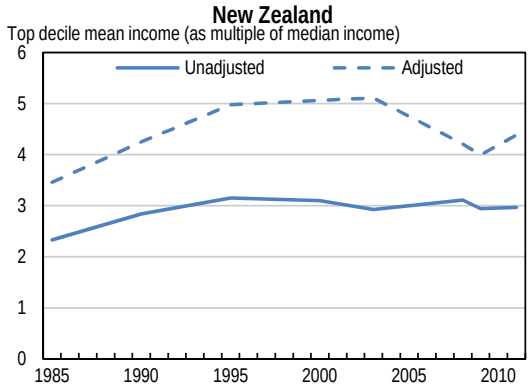


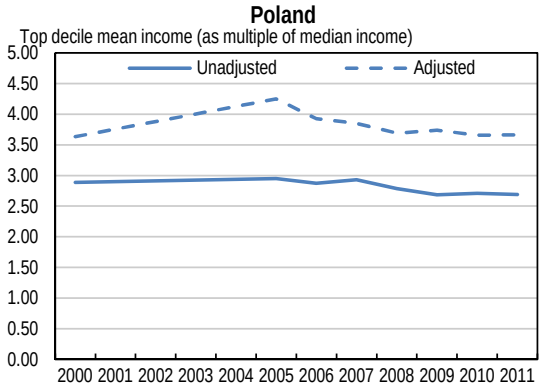
Annex 2. Part D

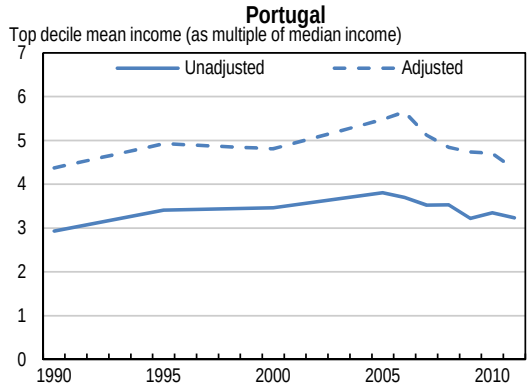


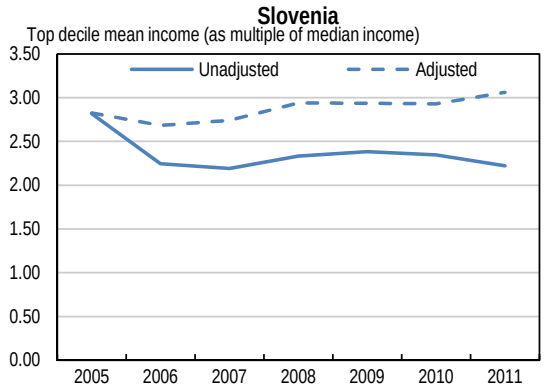
Annex 2. Part E

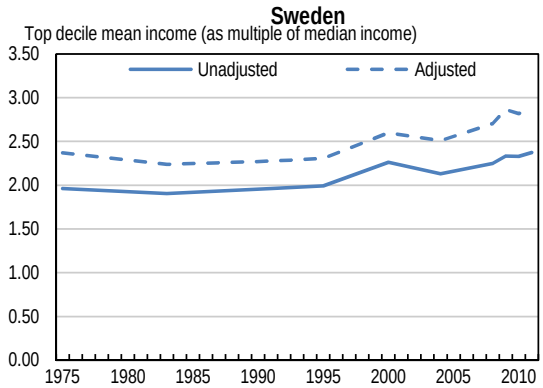


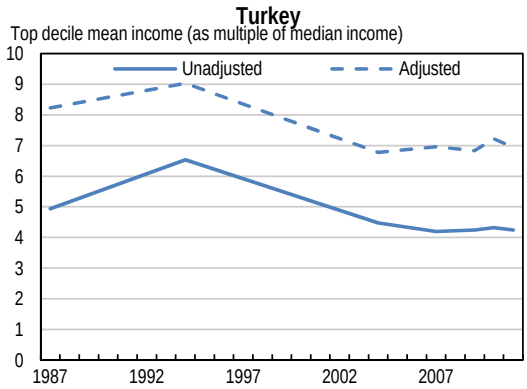


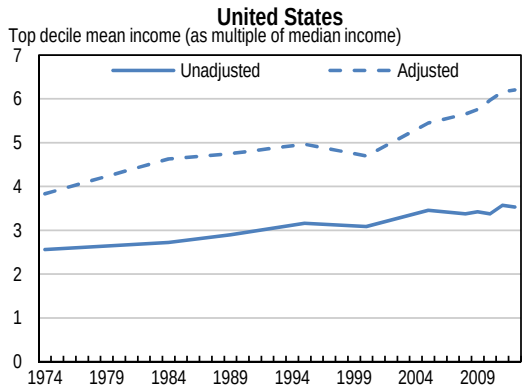






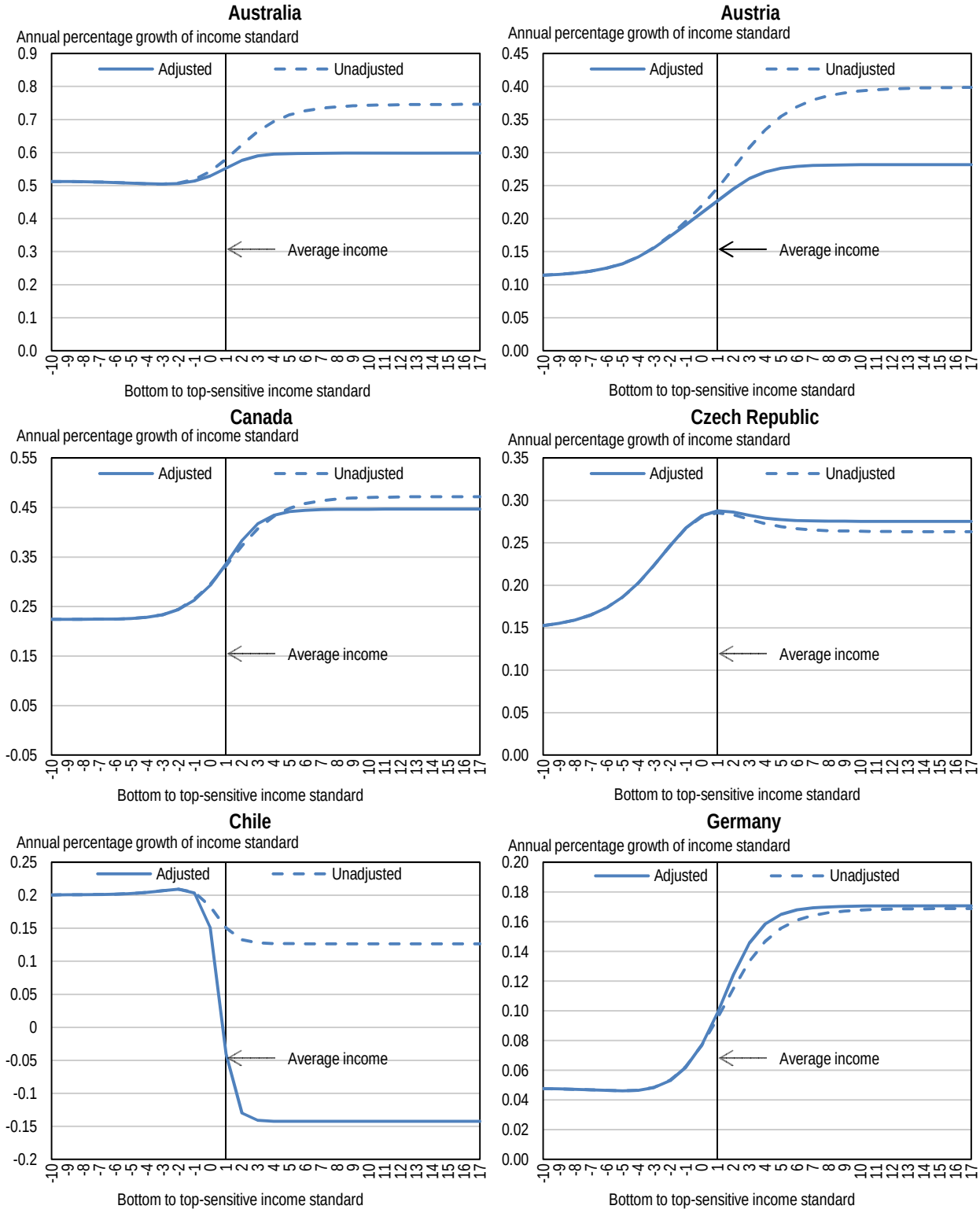




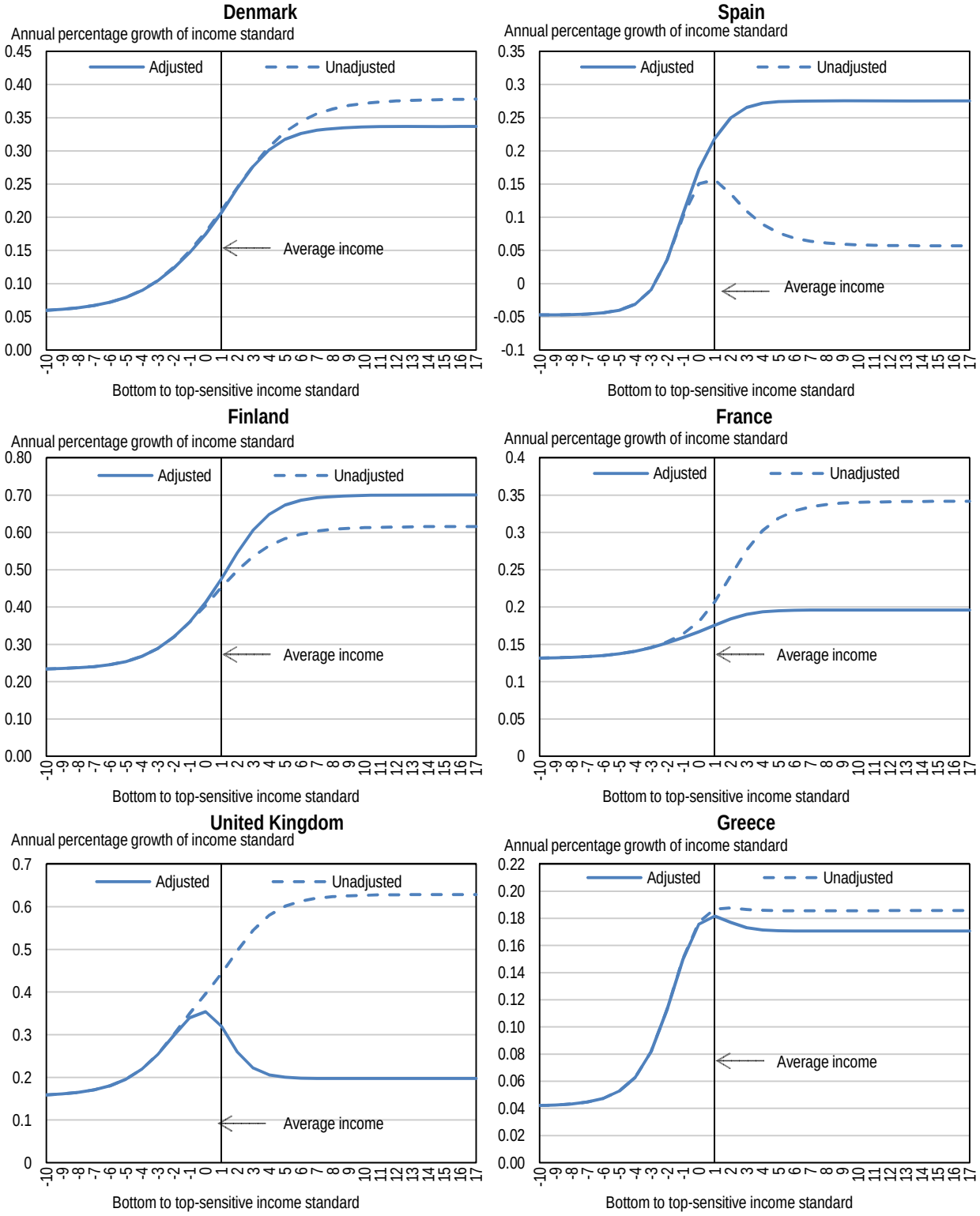


ANNEX 3: ADJUSTED VS. NON-ADJUSTED GENERALISED MEANS CURVES IN OECD COUNTRIES (FROM MID-1990'S TO LATE-2000'S)

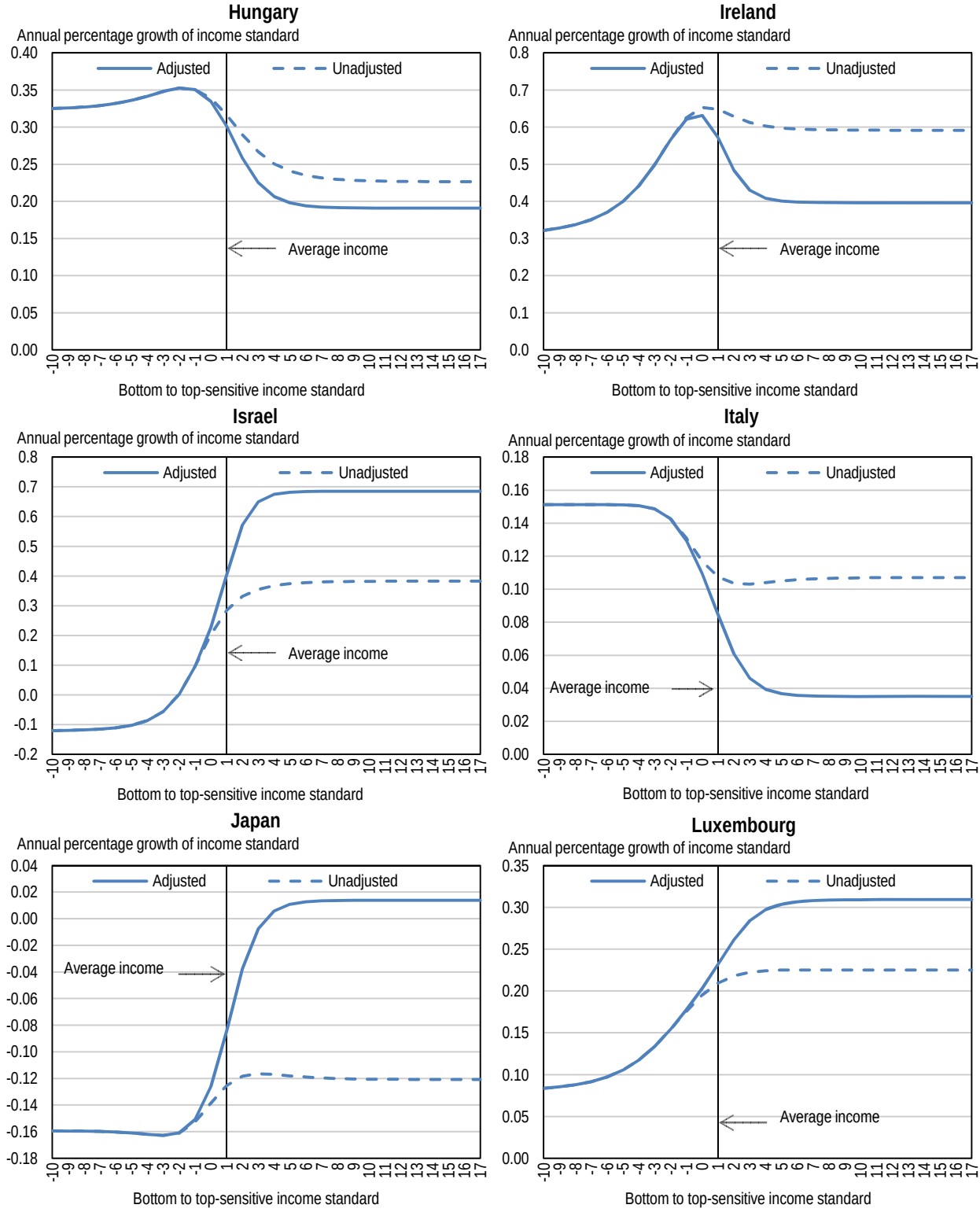
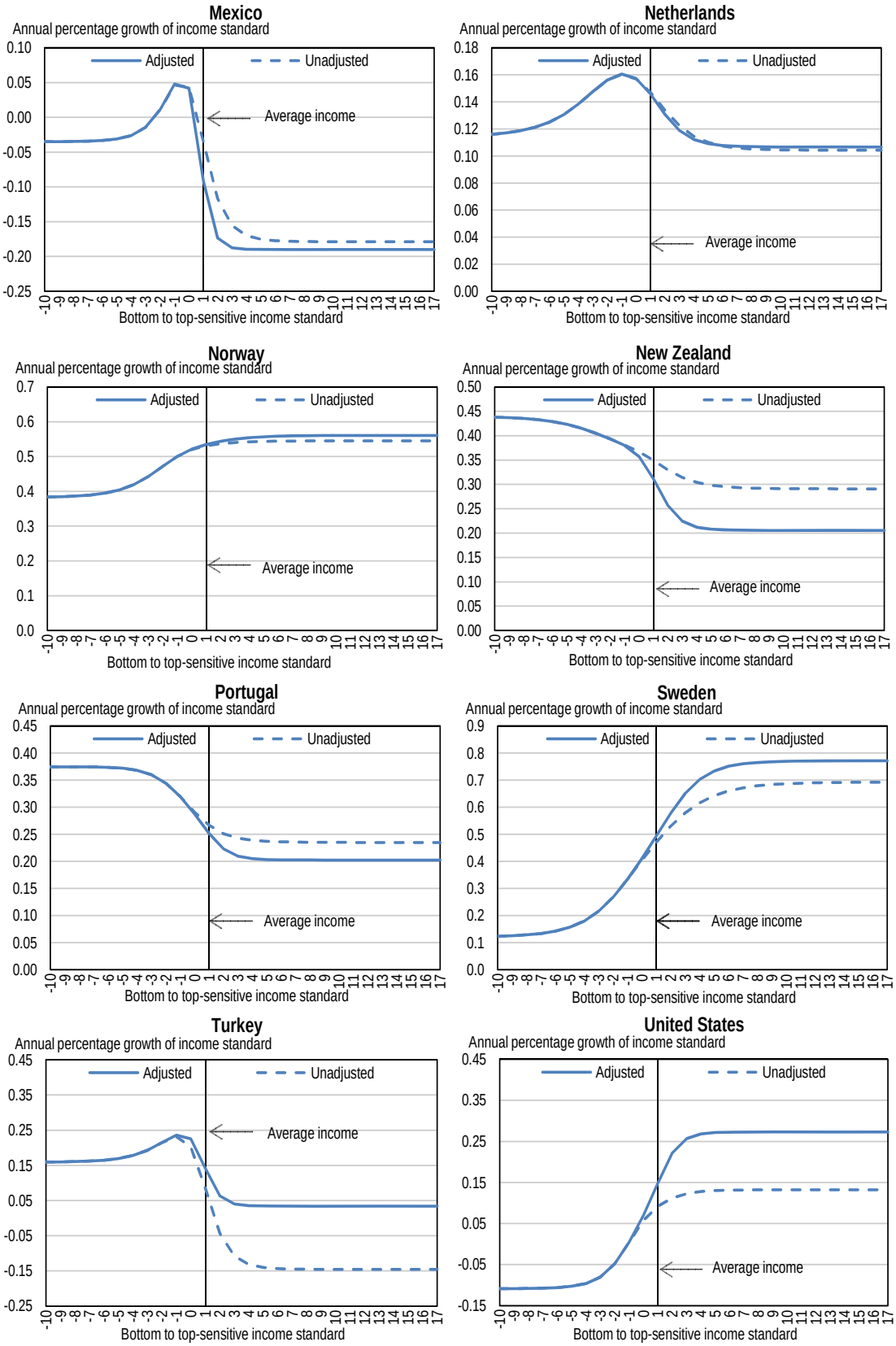
Annex 3. Part A



Annex 3. Part B



Annex 3. Part C



Annex 3. Part D