OECD Taxation Working Papers No. 45

The potential of tax microdata for tax policy

Seán Kennedy

<https://dx.doi.org/10.1787/d2283b8e-en>

OECD Taxation Working Papers

The Potential of Tax Microdata for Tax Policy

By Seán Kennedy

OECD Taxation Working Paper series

This series is designed to make available to a wider readership selected studies drawing on the work of the OECD Centre for Tax Policy and Administration. Authorship is usually collective, but principal writers are named. The papers are generally available only in their original language (English or French) with a short summary available in the other.

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 Centre for Tax Policy and Administration, OECD, 2 rue André-Pascal, 75775 Paris Cedex 16, France. This working paper has been authorised for release by the Director of the Centre for Tax Policy and Administration, Pascal Saint-Amans.

Comments on the series are welcome, and should be sent to either ctp.contact@oecd.org or the Centre for Tax Policy and Administration, 2, rue André Pascal, 75775 PARIS CEDEX 16, France.

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.

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.

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 acknowledgement of OECD as source and copyright owner is given. All requests for public or commercial use and translation rights should be submitted to rights@oecd.org. Comment on the series is welcome, and should be sent to ctp.contact@oecd.org.

Copyright © OECD 2019

Abstract

This paper explores one distinctive form of the ‘big data’ of economics – individual tax record microdata – and its value and potential for tax policy analysis. The paper draws on OECD collaborations with Slovenia and Ireland in 2018 where tax microdata was used.

Much of empirical economics is based on survey data. However, the current trend of low and falling response rates has placed a question mark over the future value of survey practice generally. By contrast, administrative microdata are increasingly used in some of the world’s most credible and influential economic research on public policy. Although tax microdata have limitations, they offer vastly greater scale and coverage, particularly among the highest earners, and have an inherent longitudinal structure with near perfect tracking rates. This paper argues that these characteristics produce analytical advantages in the depth and breadth of possible analysis. The paper also argues that the future of best-practice tax policy analysis will likely combine the unique advantages of tax microdata with survey and national account data. The complementary advantages of these combined data will be important for policymakers to address future policy challenges - including ensuring sustainable tax revenues in an era of population ageing, protecting citizens from rising income inequality and preserving fairness in a changing labour market.

Despite its potential for tax policy analysis, access to and use of tax microdata remains limited and sporadic. The primary reason for limited access is to protect privacy. Furthermore, new skills and technology will be needed to manage the unstructured and multidimensional nature of these data, which are not designed for research. Despite these challenges, tax administrations are increasingly adopting new technologies, statistical software and trained analysts to capture and process large volumes of data securely. This development could support new data access solutions where data are used more readily for tax policy research while limiting privacy risks.

Acknowledgments

This paper has benefited from support, comments and suggestions provided by David Bradbury and Bert Brys at the OECD and by Delegates of Working Party No. 2 on Tax Policy Analysis and Tax Statistics of the OECD Committee on Fiscal Affairs.

The author would like to thank Carlotta Balestra, Céline Colin, Marco Mira D'Ercole, Michael Forster, Alexander Hijzen, Herwig Immervoll, Anna Milanez, Bethany Millar-Powell, Pierce O'Reilly, David O'Sullivan, Sarah Perret, Hannah Simon and Kurt Van Dender for comments and input.

Thanks also to Anthony Bolton, Michael Sharratt, Violet Sochay for administrative support and Julien Dubuc, Karena Garnier, Hazel Healy and Natalie Lagorce for assistance with the publication.

Table of contents

OECD Taxation Working Paper series	2
Summary	3
Acknowledgments	4
1 Setting the scene	7
1.1. Introduction and structure	8
1.2. Tax microdata is increasingly a significant part of the 'big data' of economics	8
1.3. Some of the world's best economic research is increasingly using administrative microdata	9
1.4. Tax microdata is necessary but not sufficient for best-practice policy analysis	10
Notes	13
2 The Value and Potential of Tax microdata for Tax Policy: Case Studies with Slovenia and Ireland	14
2.1. Introduction	15
2.2. Structure	16
2.3. Measuring economic inequality in changing world	16
2.4. Measuring income mobility	21
2.5. Life-cycle analysis and preserving personal taxes in an era of ageing	23
2.6. Costing policy changes and distributional implications	26
2.7. Evaluating tax policies in an increasingly diverse labour market	28
2.8. Measuring the tax burden with complementary average tax rate statistics	32
Notes	35
3 Access Options	36
3.1. Tax microdata access remains limited but is slowly increasing	37
3.2. Five data access options	38
Notes	39
4 Conclusion	40
References	42

Tables

Table 1.1. The future of best-practice tax policy analysis will be based on combined data sources matched uniquely at the individual level	11
--	----

Table 2.1. Analysis of top incomes and concentration measures in Slovenia	18
Table 2.2. Tax rates developed based on microdata have key advantages	32
Table 2.3. Summary statistics using tax microdata can highlight key differences in tax burdens across cohorts and income levels	34

Figures

Figure 1.1. The rise of tax microdata in economic research	10
Figure 2.1. More microdata means more and better analysis for tax policymakers	15
Figure 2.2. Themes and analytical approaches	16
Figure 2.3. The income divide between the top 1% and the median is greatest in the professional, health and financial sectors in Ireland	17
Figure 2.4. Employee Income Distribution in Slovenia & Ireland	19
Figure 2.5. During the recent recovery in Ireland, incomes grew at the top but declined for the bottom 40%	20
Figure 2.6. Capital incomes are highly concentrated at the very top of the distribution	21
Figure 2.7. Over the past decade in Ireland, over half of taxpayers stayed in the top decile	22
Figure 2.8. Life-cycle earnings in Slovenia and Ireland	24
Figure 2.9. The vast majority of SSCs and PIT is paid by workers between 25 and 58	25
Figure 2.10. Ageing populations will reduce future taxpaying populations and tax revenues	26
Figure 2.11. The SSC loss associated with an employee SSC rate cut will be partly recovered through the PIT system	27
Figure 2.12. An employee SSC rate cut would increase disposable income across all deciles with greater relative (but not absolute) increases among the lowest deciles	28
Figure 2.13. Selected Cohort Analysis in Slovenia	30
Figure 2.14. Tax microdata can show where and how many taxpayers are responding to tax designs such as bunching before allowances are withdrawn	31
Figure 2.15. Backward-looking micro AETRs for employees at each percentile using tax microdata in Slovenia	34
Figure 3.1. Selected Access Options	38

Boxes

Box 2.1. How tax microdata expands the scope of possible analysis	28
---	----

1 Setting the scene

This chapter describes how tax and other microdata have become an increasingly important part of the ‘big data’ of economic research. It also discusses why the future of best-practice tax policy analysis will likely combine the complementary advantages of tax, survey and national account data.

1.1. Introduction and structure

This paper explores one distinctive form of the ‘big data’ of economics – administrative tax record microdata (tax microdata hereafter) – and its value and potential for tax policy analysis. The paper draws on OECD collaborations with Slovenia and Ireland in 2018 where tax microdata was used. For the current paper, no new microdata analysis was conducted. In April 2018, joint work between the OECD and the Irish tax administration was published on income dynamics and mobility (Kennedy, 2018^[1]). The analysis was based on the tax microdata of approximately 2.2 million tax units over the 10-year period from 2006 to 2015. In August 2018, the OECD published a *Tax Policy Review* of Slovenia on reshaping the personal income tax (OECD, 2018^[2]). The analysis was based on approximately 1.85 million taxpayers over the period 2015 and 2016. Analysis from these reports forms the basis of the current paper.

The purpose of the paper is to highlight the broad value of individual tax record microdata for many analytical methods and policy research areas. It is not intended to be a comprehensive assessment of any specific method or research area.

The scope of the paper is limited to individual taxation. This includes personal income taxes (PIT), social security contributions (SSC), capital taxes and property taxes. In the case of firm-level microdata, much progress has been made in recent years including at the OECD through the MultiProd and other projects (for a discussion, see (Berlingieri et al., 2017^[3]). The paper does not focus on private sector data, which also represents an important part of the ‘big data’ of economics. In addition, less attention is given to other administrative data (for example, on corporate tax, education, medical or business formation records).

The paper is structured in three parts. First, it sets the scene by discussing the role of tax microdata in economics and how it compares to survey and national account data. Second, it examines the value and potential of tax microdata for tax policy analysis across various topics and analytical methods. Finally, it reviews tax microdata access modes which countries could adopt including decentralised, extended decentralised, synthetic (see section 3.2), remote and direct.

1.2. Tax microdata is increasingly a significant part of the ‘big data’ of economics

With the exception of private sector data, ‘big data’ in economics often and increasingly means administrative microdata.¹ Despite the hyped ‘big data revolution’, there is no single definition of the term (‘big data’) or agreement on what constitutes ‘big’. For example, in the field of economics the average sample size of a dataset in microeconomics far exceeds that of its macroeconomic counterpart. The popular definition proposed by (Laney, 2001^[4]) is the three Vs characterisation: volume (i.e. the size), variety (i.e. the range of formats) and velocity (i.e. the speed of data generation). Tax microdata fit the first two Vs well: the data sample size is several orders of magnitude larger than traditional surveys and is increasingly available on new variables (for example, latitude and longitude information on property locations). The third V however, velocity, fits less well since tax returns are filed periodically. Overall, however, 2 out of 3 is not bad. Indeed, (Connelly et al., 2016^[5]) argue that administrative tax microdata are a distinctive form of big data. Another more concise description provided by (Schroeder, 2014^[6]) is that microdata represents ‘a step change in the scale and scope of knowledge about a given phenomenon’. Given these descriptions, tax microdata may be emerging as an increasingly significant part of the “big data” economics.

However, the tax microdata terrain is unpredictable and there is no metadata map. Most empirical economic analysis is based on neatly prepared survey and national account data². This can give an impression of agreed upon facts. In reality however, such data represent only one of many possible interpretations of the truth. This point is underlined by the fact that during the data preparation process statisticians are required to make a series of (largely unseen yet important) value judgements and

decisions about the data, often based on internationally agreed concepts, classifications and methodologies. By contrast, since administrative microdata are collected for a different purpose – in the case of tax, the assessment of tax liability – they are typically unstructured with little metadata. In this sense, the (tax microdata) terrain is more unpredictable and there is no (metadata) map. For example, income definitions can change with changes in tax rules over time. This distinction has led to the characterisation of tax microdata as ‘found’ rather than ‘made’ (Connelly et al., 2016^[5]). Given this data structure, the inductive data exploration phase to understand the data, such as descriptive statistics, becomes exponentially more challenging.

New skills are needed. (Einav and Levin, 2014^[7]) note that the skills required to work with such unstructured and multidimensional data are ‘not something that most empirical economists have been taught or have experience with’. These include the analytical skills to manage, process, shape and analyse large volumes of data in addition to the programming skills required to use the statistical software to do so.

Unlocking the value of tax microdata may require a shift in perspectives on the part of both tax administrations and users. A first step for tax administrations is to know the value of their tax microdata and its potential to help in shaping new policy insights. An OECD survey of tax administrations highlighted how tax microdata should not be thought of as the residue of an operational process but rather as an asset to be managed and developed (OECD, 2016^[8]). National Statistics Offices are increasingly combining administrative microdata sources with survey and national account data.

1.3. Some of the world’s best economic research is increasingly using administrative microdata

There is a rise in the use of microdata and the new avenues of analysis that it makes possible. The use of administrative microdata in economics, although still very limited, has risen sharply in recent years as shown in Figure 1.1. According to an analysis of Scopus, a database of peer review literature, use of the term ‘administrative data’ tripled in economic research over the past decade (from 76 articles in 2010 to 233 articles in 2017). Over the same period, use of the terms ‘administrative tax data’ grew even faster (albeit from a smaller base). Those strands of the economic literature which tend to use these data – income and wealth inequality, microsimulation and bunching analysis to take some examples - have expanded much faster than the economics field more broadly. The rise of such data also appears concentrated in the world’s best economic research. For example, (Einav and Levin, 2014^[9]) document the rise of papers which obtained an exemption³ for tax microdata in the American Economic Review (AER) – rising from 4% in 2006 to 26% in 2014. Similarly, using text analysis, (Kleven, 2018^[10]) documents a rise in the frequency of words such as administrative data and the top 1%.

Figure 1.1. The rise of tax microdata in economic research

Number of articles, 2007 – 2017 (Indexed: 2007 = 1.0)

Note: Based on the following terms appearing in the title, keyword or abstract: 'administrative data', 'wealth inequality', 'microsimulation' and 'bunching' respectively. While these research areas use tax administration data, they also use survey and other data sources. In 2017, these terms had 233, 104, 37 and 13 articles respectively. Economics includes all articles with the term 'economics' in any field. Only articles in the field of 'Economics, Econometrics and Finance' are included.

Source: Scopus article search September 2018.

1.4. Tax microdata is necessary but not sufficient for best-practice policy analysis

The future of best-practice tax policy analysis is likely to combine the unique advantages of tax, survey and national account data. Tax, survey and national account data have different advantages and limitations. This section briefly discusses a selection of these, with a focus on the advantages of tax microdata. For a comprehensive review of the pros and cons of data sources for measuring inequality see (Lustig, 2018^[11]). The definitions of each, and the distinctions between them, can however be unclear. For example, what constitutes a survey is not well defined. Moreover, national account data can be comprised of a mixture of survey and tax data. Nevertheless, some generalisations are useful on each of these three data sources (see Table 1.1). First, tax data have a high degree of accuracy, especially at the top of the distribution, but they often lack demographic information, income from social transfers and low-income individuals entirely. Second, survey data enjoy rich socioeconomic information on individuals and subgroups but suffer from high and rising non-response rates and poor coverage at the top. Third, national account data provide an international macroeconomic picture but they are aggregate and based on an economy-wide average. One way to bring these distinct advantages together is to start with tax microdata and match it tax microdata with survey and national account data. One challenge of linking administrative and survey data is respondent consent, which may be needed (for a discussion, see (Sakshaug and Kreuter, 2012^[12])).

Table 1.1. The future of best-practice tax policy analysis will be based on combined data sources matched uniquely at the individual level

Advantages and limitations for different data sources

Data	Advantages	Limitations	Tax Policy Themes
Tax data	<ul style="list-style-type: none"> - Universe of taxpaying population - Large sample size - Absent non-response and attrition - Near perfect longitudinal tracking - Relatively low cost 	<ul style="list-style-type: none"> - May not include untaxed transfers - Limited demographic information - Underrepresent lower end - Concepts not harmonised - Less quality control 	<ul style="list-style-type: none"> - Income inequality at the top - Capital & labour income dynamics - Income mobility over time - Natural & quasi-experiments (e.g. taxpayers bunching at kink points)
Survey data	<ul style="list-style-type: none"> - Rich demographics - Representative of total population - Represents lower end 	<ul style="list-style-type: none"> - High and rising non-response - Smaller sample size - Underrepresent top incomes 	<ul style="list-style-type: none"> - Income inequality at the bottom & social transfers - Changing nature of labour market (e.g. new forms of work)
National accounts data	<ul style="list-style-type: none"> - Comprehensiveness - Consistent macro picture - Harmonised concepts 	<ul style="list-style-type: none"> - Single economy-wide average - Too aggregate for subgroups 	<ul style="list-style-type: none"> - Compare international policies - Link to GDP, FDI, immigration - Link to productivity paradox

Note: Matching could occur at individual, family or household level.

Source: OECD analysis.

While the value of survey data for research is declining with high and rising non-response rates, the value of tax microdata is growing through its greater availability. Representative surveys, such as those compiled by National Statistics Offices (NSOs), have traditionally provided data for most empirical economic research. However, these processes are costly, complex and slow (Cavallo and Rigobon, 2016^[13]) and can involve large numbers of people collecting data over months followed by extensive data processing. Furthermore, in a world where ‘big data’ are increasingly captured organically, surveys appear intrusive by comparison. Consequently, survey participation has and is declining. As an example of declining survey participation rates, in a series of four SME surveys on tax conducted over the past decade in Ireland, response rates declined from 80% in 2006 to 46% in 2013 to 20% in 2017 (Kennedy et al., 2017^[14]).⁴ All of this has placed a question mark over the future value of traditional survey practice generally (Miller, 2017^[15]). By extension, this has implications for the empirical economic research that depends upon it. In sharp contrast, tax microdata have seen a marked revival. For example, according to (Slemrod, 2016^[16]) ‘the expanded availability of tax-return tax microdata has triggered an explosion of scholarly research addressing tax policy’.

Tax microdata have statistical advantages including in comprehensiveness and accurate measurement. (Groves et al., 2009^[17]) identifies four categories of error under which survey and tax microdata can be assessed; coverage, non-response, sampling and measurement. These are discussed in turn. First, tax microdata cover the entire taxpaying population (rather than the total population). This size advantage relative to surveys facilitates the study of more compelling research designs, natural experiments and exceptional event analysis. However, the tax data may underrepresent the lower part of the population income distribution. This can occur when government transfer payments and non-taxable employer provided benefits are not included (Auten and Splinter, 2018^[18]) or where low-income tax units are not required to file at all. Furthermore, the unit of analysis are tax units or taxpayers rather than individuals or households. Second, it is the law to file tax returns so non-response and attrition are mostly absent from the data. Third, when the taxpayer population is available sampling error is zero (albeit missing data is still a possibility where tax forms are incomplete). Finally, measurement and misreporting are less problematic with tax microdata. For example, the evidence shows that high-income survey respondents under-report incomes. By contrast, taxpayers have an incentive to accurately report incomes because not doing so is an offence. However, tax microdata also come with additional data challenges including data processing errors by the tax administration, unintentional and intentional mistakes (informality is a

significant issue) by the taxpayer (Slemrod, 2016^[16]). Since tax evasion differs significantly across countries and over time, this reduces comparability of tax data. Overall however, the advantages of tax data have caused leading economists in some countries to conclude that tax microdata may be 'highly preferable' to surveys for policy evaluation (Card et al., 2011^[19]).

Tax microdata naturally produce longitudinal data over time with near perfect tracking rates of the same taxpayers. Longitudinal data pave the way for more advanced analysis. This include survival and mobility analysis but also panel data econometrics that can statistically account for unseen taxpayer-specific differences such as skill levels (in the economic jargon, this is referred to as unobserved heterogeneity). In some countries, it is possible to link the tax microdata of taxpayers with their dependents over time. For example, (Chetty et al., 2014^[20]) use such data to examine intergenerational earnings mobility in the United States. However, longitudinal data also have drawbacks, if for example the definition of variables or tax units change over time. These data also may also increase privacy concerns. While survey data can also be longitudinal, they can suffer from smaller samples and attrition. A further statistical issue also can arise called time-in-sample bias where respondents' experience with a survey over time causes them to answer differently to those answering for the first time (Couper, 2000^[21]).

One of the great gaps of tax microdata is its lack of demographic and socioeconomic information on individual taxpayers, but these can be overcome by matching with survey data. Demographic and socioeconomic data could include educational attainment, years of experience, occupation, number of hours worked or health status. The potential of matching for analysis is greatest when done at the individual taxpayer level using unique identifiers (called 'record linking') although other levels of disaggregation are possible (for instance, incomes, age, sector, region). For a review of various approaches to combining information from various data sources, see (Lustig, 2018^[11]). As (Card et al., 2011^[19]) note 'there is tremendous value in carrying out research by merging data, for example educational data and earnings data'. Despite its potential, such matching is challenging - for example, given the need to reconcile potentially different definitions, concepts, timing and coverage.

Tax and macroeconomic data can be combined in various ways. An OECD Expert Group has developed a methodology that combines distributional measures of income, consumption and wealth across household groups using microdata (Zwijnenburg, Bournot and Giovannelli, 2017^[22]). The paper describes a method for matching components of the microdata to the national account items in different countries. (Piketty, Saez and Zucman, 2018^[23]) start with combine tax data and combine it with survey and national account data to build a new series on the distribution of national income in the United States (see also (Alvaredo, F., 2018^[24]). While the approach has limitations (as stressed by the authors), it produces a dataset that has consistency with both the macroeconomic picture (from national accounts data) and the microeconomic picture (from tax records and surveys data). Using a different approach, instead of attempting to systematically match national income, (Auten and Splinter, 2018^[18]) produce a United States top income share series by adding components to fiscal income.

Notes

¹ As discussed, private sector data, which is not the focus of this paper, also represent an increasingly significant part of the 'big data' of economics.

² National Statistics Offices (NSOs) often use administration data in preparing data including for example defining sampling frames and cross-validation of statistical results.

³ An exemption may be given where it is not possible to provide data to the journal, for example, confidential tax administration data.

⁴ This series of SME Surveys was conducted by the Irish Revenue Commissioners. The survey questionnaire differed in each survey and was administered by post in 2006 and 2010 and electronically in 2017.

2 The Value and Potential of Tax microdata for Tax Policy: Case Studies with Slovenia and Ireland

This chapter examines how tax microdata can expand the depth and breadth of analysis across many areas of tax policy research. For example, economic mobility, life-cycle analysis and microsimulation. The chapter draws on recent OECD collaborations with Slovenia and Ireland where tax microdata was used.

2.1. Introduction

This section describes how tax policy analysis can be enhanced using tax microdata. A number of areas of tax policy are considered, using examples from Slovenia and Ireland. These include income inequality and mobility, life-cycle analysis, the detection of earnings clustering through bunching analysis, microsimulation methods in addition to complementary measure of tax statistics¹. While many of these analyses can be conducted using survey and other data, this paper focuses on the value-added of using tax microdata.

Greater access to tax microdata will expand the range of analysis options for tax policy. Each tax administration has a different level of data disaggregation for internal tax policy analysis and for external researchers. To give a sense of this, Figure 2.1 shows a stylised spectrum with four different degrees of data disaggregation from low to high. Tax administrations fall somewhere along this spectrum. The spectrum seeks to highlight three points. First, greater data disaggregation will expand the range of possible analyses for tax policy. Second, greater data disaggregation will enhance existing analyses based on less disaggregate data. Third, a tension exists between greater analytical capability and protecting taxpayer confidentiality, which is represented by an inverse relationship on the vertical axis.

Figure 2.1. More microdata means more and better analysis for tax policymakers

Stylised trade-off between data access, capability & confidentiality

Source: OECD analysis.

On the horizontal axis are four stylised representations of data disaggregation:

- 1. Aggregate data** refer to broad summary statistics on personal income tax, perhaps placed on a website. They could also be more advanced statistics. From an analysis perspective, aggregate data are likely to restrict the options to summary income tax statistics for example across different taxpayer types, income ranges and regions. However, aggregate data can also be supplemented with tax records without access to individual data (Burkhauser et al., 2018^[25]).
- 2. Sample data** refer to some representative sample of the tax microdata, such as a public use-file, which could be made available to researchers. A representative sample of the tax microdata may allow for more advanced analysis such as income inequality measures and tax forecasting.
- 3. Taxpayer population data** refer to anonymised individual tax microdata for the full taxpaying population. It could be a synthetic version of the data and accessed remotely or directly. The data could be over one or many years. These data vastly expand the analysis options. For example, they have the sample size to facilitate microsimulation, life-cycle analysis, bunching analysis and other quasi-experimental research which would otherwise not be possible.

4. **Total population data** refer to longitudinal data on the taxpaying population, which is matched with survey and national accounts data so that it represents the total population. Compared to the taxpaying population, these data further enhance analysis and add some further options. For example, estimating the elasticity of taxable income with respect to the tax rate can be estimated with one year of the taxpaying population, but estimation is improved by using longitudinal data. Another example is mobility and survival analysis, which requires longitudinal data with low attrition rates to track the same taxpayers over time or a behavioural microsimulation requires rich survey data.

2.2. Structure

Figure 2.2 provides a summary of the themes and analytical approaches in the remainder of this section. The themes include economic inequality, economic mobility, life-cycle analysis, costing tax policy impacts and distributional implications, evaluating tax policies and complementary tax burden measures.

Figure 2.2. Themes and analytical approaches

Source: OECD analysis.

2.3. Measuring economic inequality in changing world

Technology and globalisation are transforming the global economy by providing new opportunities for growth. At the same time, they raise the risk of inequality (OECD, 2018^[26]). In most OECD countries, the gap between rich and poor is at its highest level in 30 years (OECD, 2015^[27]). Tax policy plays a role in supporting inclusive growth, that is, to ensure that the benefits of growth are shared broadly.

Tax microdata can help to build a comprehensive picture of the income and wealth distribution that is important for effective tax policy design. Income inequality is typically measured using survey data and the central concept of ‘equivalised household disposable income’. For example, income inequality across countries is monitored using the OECD’s *Income Distribution Database (IDD)*. Despite the significant definitional and conceptual differences that exist between survey and tax data (see section 1.4 for further discussion), tax microdata may be able to help. (Alvaredo et al., 2017^[28]) point out that administrative data are important to ‘properly understand the present as well as the forces which will dominate in the future, and to design potential policy responses’.

A compelling case for the value of tax microdata is how it can deliver unique policy insights with simple statistical descriptions, particularly when it is combined with other survey and national account data. A recent example using tax data is provided by (Piketty and Saez, 2014^[29]) who, by simply documenting basic facts on the long-run evolution of income and wealth, have reignited a global debate on inequality.

2.3.1. Measures of concentration

There are many options for measuring income inequality (Jenkins and Van Kerm, 2011^[30]). Perhaps the most widely used is the Gini coefficient, a synthetic point estimate of income across the entire population, which ranges from perfect equality to perfect inequality². Inter-decile ratios are intuitive and direct but, by construction, omit information on percentiles not selected and may be sensitive to outliers. (Piketty and Saez, 2014^[29]) argue that the 'simplest and most powerful measure' on inequality is the share of total pre-tax income going to the top decile. However, this measure is not informative about inequality at the bottom part of the distribution. Using tax microdata and survey data combined, the above measures can be developed, before and after taxes, and compared across cohorts (men and women; employees and self-employed).

Tax microdata allows for measuring the extent to which inequality is driven by specific sectors of the economy. Income inequality is typically measured at the aggregate countrywide level. How do the incomes of the highest earners compare to the median across sectors? Figure 2.3 shows relative income inequality across sectors in Ireland by displaying the P99/P50 ratio and its component parts – the median and the top 1% income threshold.

Figure 2.3. The income divide between the top 1% and the median is greatest in the professional, health and financial sectors in Ireland

Income divide between the median and top the 1% by sector, 2015

Note: 2015 prices. Gross income of tax units in 2006 deflated by the consumer price index. Tax units may represent one or two taxpayers and are not equivalised. For simplicity, the names of sectors are abbreviated. The data show relative rather than absolute mobility.

Source: OECD analysis based on Kennedy et al. (2018), Income Dynamics and Mobility in Ireland: Evidence from Tax Record Microdata, Irish Government Economic and Evaluation Service Publication Series.

In Slovenia, concentration measures suggest a relatively equal income distribution. Table 2.1 shows that the gross income earned by the top 20% of employees is 5.7 times greater than the gross income earned by the bottom 20%, as measured by the S80/S20 quintile share. Employees in the top 1% earn 6.5% of all income and contribute to 7% of all PIT and SSCs. These simple descriptive data are based on

about 740 000 employees who earn EUR 14.6 billion in gross income and pay EUR 6.8 billion in PIT and SSCs.

Table 2.1. Analysis of top incomes and concentration measures in Slovenia

Employee shares of income, PIT and SSCs, by decile, 2016

	Taxpayers (number)	Gross income	Disposable income	PIT	PIT & total SSCs
Total (EUR millions)	741 670	14 629	9 992	1 685	6 775
<i>Of which:</i>					
Bottom decile	74 167	2.4%	2.8%	0.1%	2.0%
Top decile	74 167	26.9%	24.5%	44.1%	30.0%
Top 1%	7 416	6.5%	5.8%	13.0%	7.2%
S80/S20 quintile share		5.7	4.6	83.1	7.5
S90/S10 inter-decile		11.3	8.8	703.0	15.3

Source: Reshaping the Personal Income Tax in Slovenia, OECD (2018).

2.3.2. Charting new distributional territory

Tax microdata can define and focus on almost any taxpayer cohort while retaining sufficient sample size for robust analysis. Perspectives on economic inequality can also be visualised graphically by plotting income distributions. A wide range of visual analyses are possible. These cohorts could include combinations across taxpayer types (employees, self-employed, pensioners), income sources (wages, self-employed, pensions, rental, capital income) and other demographic and socioeconomic characteristics (gender, household structure, age ranges, sectors of employment, geographic areas and income ranges). For a discussion on how tax microdata expand the scope of analysis see Box 2.1.

The Slovenian data reveal how wage levels are relatively low and equal, which has implications for the amount of tax that can be raised. Figure 2.4 shows the entire employee wage distribution at each percentile in Slovenia and Ireland. It may also shed light on the share of the population at risk of falling into poverty for whom targeted tax policies, such as allowances of earned income tax credits (EITCs), may be of benefit.

Figure 2.4. Employee Income Distribution in Slovenia & Ireland

Note: Slovenia relates to 2016 while Ireland relates to 2006 and 2015. For Slovenia, a cut-off is applied for gross income below EUR 500.

Source: Reshaping the Personal Income Tax in Slovenia, OECD (2018). ; Kennedy et al. (2018), Income Dynamics and Mobility in Ireland: Evidence from Tax Record Microdata, Irish Government Economic and Evaluation Service Publication Series.

During the recent economic recovery in Ireland, gross incomes grew at the top, but declined for the bottom 40%. Figure 2.5 shows three-year income growth by percentile in Ireland over the past decade. During the initial period of the economic boom and the onset of the recession (2006 – 2009), incomes continued to rise, most quickly in the bottom of the income distribution. However, the extent of this rise is likely inflated by the severity of unemployment among young workers during the period (Bergin, Kelly and McGuinness, 2015^[31]), which on the tax records gives the appearance of higher relative incomes, but may be more of a reflection of the profile of workers who remained in the workforce. Exceptionally, the top 1% were the only percentile to experience a significant decline, likely reflecting their higher concentration of income sources from capital (such as shares) and property. Using survey data on disposable incomes, (Savage et al., 2015^[32]) also find increased longer-range downward mobility during the worst of the recession. During the recovery period (2012 – 2015), the shape of this trend inverted: while overall wages growth was close to zero, incomes grew at the top, but declined for the bottom 40%. For a long-term view of household incomes in Ireland over the past 30 years, see (Callan, Maxime and Walsh, 2018^[33]), which shows that incomes have grown strongly and the distribution has been broadly stable.

Figure 2.5. During the recent recovery in Ireland, incomes grew at the top but declined for the bottom 40%

Three-year income growth in Ireland - Recession, Stabilisation and Recovery

Note: 2015 prices. Gross incomes of tax units. Previous years deflated by the consumer price index. Only incomes greater than 10th percentile are shown. Analysis compares percentiles at two points in time and does not track the same taxpayers.

Source: Kennedy et al. (2018), *Income Dynamics and Mobility in Ireland: Evidence from Tax Record Microdata*, Irish Government Economic and Evaluation Service Publication Series.

2.3.3. Measuring wealth and capital income inequality with tax data

In recent years, the composition of national income has shifted from labour to capital and wealth³, especially among top incomes (Piketty, 2014^[34]). While top income concentration has historically been driven by wages (since the 1970s), over the past 15 years the increase is attributable to capital (equity and bonds) at the top (Piketty, Saez and Zucman, 2018^[23]). At the same time, globalisation has increased the international mobility of capital income, creating greater opportunities for avoidance and evasion, which places further pressure on personal income tax (PIT) systems.

Wealth inequality is often measured using survey, tax and administrative records. For example, the OECD *Wealth Distribution Database (WDD)* includes estimates for 28 OECD countries based on national sources (Balestra and Tonkin, 2018^[35]). Using tax data, a good source to measure wealth concentration would be wealth tax declarations for the entire population, although no country in the world has such a perfect data source (OECD, 2018^[36]). **Tax microdata provides only a partial picture of wealth since not all wealth is subject to tax and recorded in the tax records.** However, its large-scale allows for examining specific components of income, such as capital income, across the income distribution while retaining sufficient sample size. In Slovenia, for the first half of the distribution, interest income from bank deposits typically comprises between half and three-quarters of capital income, with most of the remainder being dividends - income from capital gains does not start to become a significant component of capital income until after about the 85th percentile.

One alternative approach is to measure wealth indirectly. In a widely cited paper, (Saez et al., 2014^[37]) attempt this by bringing together capital assets in the tax microdata (for example, dividends, interest, rents, profits, mortgage payments) with a wealth dataset. The novelty of the paper is not its methods but rather its data - 'compared to previous attempts, our main advantage is that we have more data'. Using this method, they find that wealth inequality has followed a U-shape over the past 100 years and that 'the rise of wealth inequality is almost entirely due to the rise of the top 0.1% wealth share'.

Figure 2.6. Capital incomes are highly concentrated at the very top of the distribution

Capital income distribution by type, by capital income percentiles, in Slovenia 2016

Note: The total capital income decreases in some percentiles due to identical values for different taxpayers are assigned to the same percentile.
Source: Authors' calculations based on Ministry of Finance of Slovenia tax records microdata.

2.4. Measuring income mobility

While policymakers are rightly concerned about the well-documented evidence on rising income concentration at the top, it is often wrongly assumed that this means that the same rich individuals stay rich. In reality, the membership of this group is in constant flux - the people who are rich this year are not necessarily the people who are rich next year. The extent of this change is measured by income mobility – who moves up and down the income ladder over time? This question is crucial for tax policy because income inequality, and the extent to which countries choose to reduce it through the tax system, may be of less concern in a society with significant income mobility. Measuring upward mobility from the bottom of the distribution is also important. Base broadening tax policies targeted at specific cohorts, such as introducing or expanding an EITC, may also be an important factor for affecting the pattern of economic

mobility by acting as a ‘mobility lever’ to help move lower income workers up the income ladder over time (Mitnik et al., 2015^[38]).

The best mobility studies use longitudinal data, large sample sizes and cover long time horizons.

Social progress has for the most part been measured using average real income growth, which does not account for relative winners and losers. In economics, there is an excessive and damaging focus on averages over distributions. Ninety-five percent of applied econometrics is concerned with averages (Angrist and Pischke, 2009). Since high-quality longitudinal data are rare, few studies look at income mobility in and out of top income groups (Förster, Llena-Nozal and Nafilyan, 2014^[39]). Furthermore, existing survey evidence sometimes relies on small samples (Mitnik et al., 2015^[38]) and can suffer from attrition (people stopping participation over the course of the survey). For example, OECD has recently produced comprehensive income mobility analysis over the life course across countries using survey data (OECD, 2018^[40]). Increasingly however, income mobility studies appear to use tax microdata which can help to overcome some of these concerns (Mazumder, 2015^[41]), (Chetty et al., 2014^[20]); (Chetty et al., 2015^[42])).

One way to measure income mobility is to examine the positional change of individuals in the income distribution over time (Jäntti and Jenkins, 2013^[43]). For Slovenia and Ireland, transition matrices are calculated at two points in time for the same taxpayers using a balanced panel. Such analysis measures relative (but not absolute) income changes and does not capture those leaving the workforce. In addition, the time horizon selected for calculating a transition matrix matters – shorter periods will show less mobility for the simple reason that there is less time for it. Figure 2.7 shows income mobility of the taxpaying population in Ireland over the past decade between 2006 and 2015. Of those in the bottom decile in 2006, 1 in 5 (23 percent) remain entrenched in that decile over the ten year period while 4 in 5 (77 percent) move upwards. Of those in the top decile in 2006, over half (57 per cent) stay in that decile by 2015. This appears broadly similar to the United States, where around half of the top fifth of income earners remain in the top quintile after 11 years over the period 1970 to 1995 (Bradbury and Katz, 2009^[44]). However, the unit of analysis in this aforementioned paper is family incomes adjusted for family size rather than taxpayers. Similar mobility analysis in Slovenia for a shorter one-year period reveal that the self-employed have the greatest upward mobility compared to employees and then pensioners.

Figure 2.7. Over the past decade in Ireland, over half of taxpayers stayed in the top decile

Income Mobility in Ireland, 2006 - 2015

Source: Kennedy et al. (2018), Income Dynamics and Mobility in Ireland: Evidence from Tax Record Microdata, Irish Government Economic and Evaluation Service Publication Series.

2.5. Life-cycle analysis and preserving personal taxes in an era of ageing

2.5.1. Life-cycle incomes

Unlike aggregate age group data, combining the age of taxpayers with their income at the individual-level unlocks new analytical possibilities. This includes measuring life-cycle earnings, life-cycle tax contributions, the gap between official and effective retirement and the tax implications of an ageing society. For a comprehensive analysis of the development of inequality over the life cycle and how to prevent it see (OECD, 2017^[45]).

Such examinations of incomes by age for an entire population of taxpayers represent a first point of departure for enhancing visibility on three related but distinct effects: an age effect (the stages of the life cycle), a period effect (economic conditions in the current period) and a cohort effect (the group's initial level of inequality).⁴ Figure 2.8 compares incomes by age for the entire taxpaying populations of Slovenia and Ireland for the median, top and bottom decile taxpayers at each age from 15 to 90.

From a policy perspective, life-cycle analysis can identify poverty traps, for example among vulnerable older workers, which could be supported through targeted policy responses. The life cycle effect, that is, how people's income change over their lifetime, is apparent in both countries for young, middle-aged and older workers. The general trend is as follows. Incomes rise between 20 and 40 years of age, peak between 45 and 55 and decline thereafter. The sharp rise for young workers is in part driven by students, a group more likely to undertake part-time work, who also transition from study to employment. In Slovenia, more income is concentrated in the working-age population between 30 and 55 while the old-age incomes could be characterised as low, stable and relatively equal. The most striking comparison is the exceptionally steep income cliff observed in Slovenia as taxpayers' transition from work to retirement. At age 58 for instance, the top 10% of earners have gross incomes of EUR 31 063 but by 64, only a 6 year difference, incomes have almost halved. This trend, partly caused by low pensions, diverges sharply from other OECD countries such as Ireland and the United States (Auten, Rey Gee and Turner, 2013^[46]). From a policy perspective, this line of analysis can identify poverty traps among older workers. This could be alleviated through targeted policy responses, such as by encouraging part-time employment or self-employment through targeted age-specific tax incentives.

Figure 2.8. Life-cycle earnings in Slovenia and Ireland

Note: Slovenia relates to 2016 while Ireland relates to 2015. For Slovenia, a cut-off is applied for gross income below EUR 500.

Source: Reshaping the Personal Income Tax in Slovenia, OECD (2018); Kennedy et al. (2018), Income Dynamics and Mobility in Ireland: Evidence from Tax Record Microdata, Irish Government Economic and Evaluation Service Publication Series.

2.5.2. Life-cycle tax contributions

Life-cycle analysis can identify income, productivity and tax revenue risks to the economy. Figure 2.9 shows who contributes to PIT and SSCs in Slovenia across the age spectrum – the vast majority is contributed by the working age population, with the highest payments concentrated among workers aged 35 to 50. The sharp decline in incomes as workers age is also seen in the tax and contributions of that age-cohort. The figure highlights the dramatic economic loss through income, productivity and exchequer revenues to the Slovenian economy arising from comparatively early retirement.

Figure 2.9. The vast majority of SSCs and PIT is paid by workers between 25 and 58

Median PIT and SSCs by age, in Slovenia 2016

Source: Reshaping the Personal Income Tax in Slovenia, OECD (2018).

2.5.3. Estimating the statutory to effective retirement age gap

Measuring and then closing the gap between official and effective retirement can help to sustain tax revenues. A related analytical possibility is estimating the gap between the official retirement age (the statutory age set by government) and the effective retirement age (the age at which, in effect, most people retire from the workforce whether through early retirement or sick leave). An observable gap, which could for example be attributable to early retirement or sick leave, likely signals an economy-wide loss in productivity, income and tax which otherwise could be reversed through policy. In Slovenia, analysis of the tax microdata shows a gap of more than 4 years on average. This would place the effective retirement age at about 61, which is low by international standards. Reforms aimed at closing such a gap would do much to sustain future tax receipts and productivity more generally.

2.5.4. Preserving personal taxes in the era of ageing

The consequences of population ageing for tax revenues will be direct and significant. The global population is ageing rapidly. Virtually every country in the world is experiencing growth in the number and proportion of older persons (UN, 2017^[47]). The consequences for tax revenues will be direct and significant. First, pension systems often represent a transfer of income from those of working-age to the retired (under a pay-as-you-go system). As populations age, a growing retired population is supported by a contracting working population. This can widen intergenerational inequality. Second, older workers have lower participation rates in the labour market which means they contribute less in personal income taxes. Third, many governments provide generous pension deductions. As the population ages, the tax loss associated with those deductions rises. Tax policies can prepare countries for this looming demographic shift.

Tax policies can help prepare countries for a looming shift in demographics. Microdata can be applied to estimate tax base erosion and the extent to which it can be limited through tax policy. For example, in a research paper using individual data from a survey on national living incomes to estimate the impact of ageing on personal income tax (PIT) in Japan, (Yashio and Hachisuka, 2014^[48]) simulate changes in public pension deductions and the associated impact on individual incomes. The data allow them to estimate the effects of strengthening the tax on pension benefits for specific cohorts such as wealthy pensioners.

In Slovenia, the working age population currently pays the vast majority of PIT and SSCs. Therefore, the projected decline in the working age population, alongside a rise in older workers will have significant negative consequences for the revenues raised from PIT and SSCs in the coming decades. By applying the projected population changes by age group to the number of taxpayers in the same age groups in the tax microdata, it is possible to estimate the losses in tax associated with ageing over the period, which are likely to be significant.

Figure 2.10. Ageing populations will reduce future taxpaying populations and tax revenues

% change in PIT and SSCs arising from ageing in Slovenia

Source: Reshaping the Personal Income Tax in Slovenia, OECD (2018).

2.6. Costing policy changes and distributional implications

A number of policy options are available to countries to achieve policy objectives including enhancing tax revenues, economic growth or the redistributive impact of the tax system. At the heart of these policy choices, is a trade-off between equity and efficiency. On the one hand, the PIT and SSCs represent the main source of progressivity in most tax systems and changes in the tax rate schedule can encourage fairness by shifting the burden from lower to higher incomes. On the other hand, rate increases can reduce economic incentives to work, save and invest. Therefore, PIT and SSCs, as the most important sources of tax revenue in most OECD countries, have a crucial role to play in delivering both progressivity and tax revenues in society.

The growing availability of microdata has led to an expanded use of microsimulation methods to assess the impact of policy reforms. As the name suggests, microsimulation requires microdata, which are often large-scale and cross-sectional in nature. Most microsimulation models are static rather than behavioural or dynamic. Behavioural models can account for the behavioural response to taxation for different taxpayer segments. Dynamic models add a time dimension. The basic intuition is, for each taxpayer, to establish a baseline scenario (what would have happened with no policy change) and to compare it with a reform scenario (where a set of observed variables are replaced to reflect a policy change). Part of the appeal is practical – microsimulation can estimate the exchequer impacts of a new policy, or the impact on the income distribution, without the cost or risk of national rollout.

One key aspect of undertaking microsimulation analysis is the capacity to recalculate, or approximately recalculate, variables such as the tax paid by each taxpayer in the tax records. This is challenging in part due to the number of variables used in the calculation of tax (for example, gross income, allowances, credits) but also different taxpayer types for whom the calculation is different (for example, employees, self-employed, married, with and without children). For example, a microsimulation model in Germany, MIKMOD-Est, is capable of precisely recalculating current tax payments in part due to having over one thousand variables to calculate with (Flory and Stöwhase, 2012^[49]).

In Slovenia, a set of tax reform recommendations were costed using a simple static microsimulation approach. The data used were the entire employee taxpayer population in Slovenia of 741 670 in 2016.⁵ The recommended costings included measuring both the exchequer and distributional impacts of:

1. A reduced employee SSC rate (in combination with increased PIT rates); and
2. A reduced top PIT rate (including the abolition of the top rate).

2.6.1. Simulating a reduced employee SSC rate

One reform involved a cut in the employee SSC rate of one percentage point (Figure 2.11). The simulation analysis revealed that the reform was associated with a loss of approximately EUR 134 million in SSC revenues but that EUR 34 million (about one-quarter) could be recovered through the enhanced revenue raising capacity of the PIT system. That recovery occurs both directly, since reduced employee SSCs broadens the PIT base through higher taxable income, and indirectly as some taxpayers might be pushed into higher PIT rate brackets. In addition to the exchequer impact, microsimulation methods can examine distributional impacts by calculating disposable income before and after a reform. In Slovenia, this showed that reducing the employee SSC rate by 5 percentage points would increase disposable income but would benefit those on higher incomes more (Figure 2.12). One conclusion of the analysis is that the distribution of the gains of the reform could be shared more equally through a precise and simultaneous cut in employee SSCs with an increase in PIT rates.

Figure 2.11. The SSC loss associated with an employee SSC rate cut will be partly recovered through the PIT system

SSC loss and PIT gain from reducing the employee SSC rate from 22.1% to 16.86%, in Slovenia

Note: The analysis assumes no behavioural change and linearity from the employee SSC rate reductions.

Source: Reshaping the Personal Income Tax in Slovenia, OECD (2018).

Figure 2.12. An employee SSC rate cut would increase disposable income across all deciles with greater relative (but not absolute) increases among the lowest deciles

Mean disposable income from employment before and after a 5 percentage points SSC cut, by disposable income decile, in Slovenia

Notes: Based on disposable incomes of individual taxpayers in Slovenia.

Source: Reshaping the Personal Income Tax in Slovenia, OECD (2018).

2.7. Evaluating tax policies in an increasingly diverse labour market

2.7.1. Tax microdata expands the scope of possible analysis

The evaluation of existing or new tax policies, whether targeted tax rate or base changes, require the clearest possible picture of the current circumstances, and likely future behaviours, of the taxpayers that are affected by them. For instance, a redesigned PIT rate schedule along with a base broadening will produce a multitude of additional tax burdens and benefits across an income distribution with differing effects on different taxpayers. Tax microdata can help to clarify such complexity as it can be used to define and focus on a vast array of potential taxpayer cohorts while retaining sufficient sample size. When these cohorts are clearly defined, simple statistical descriptions can provide powerful insights for tax policy.

Box 2.1. How tax microdata expands the scope of possible analysis

1. **The data can be explored in boundless directions.** The basic building blocks of tax microdata analysis include the unit of analysis (e.g. taxpayer), the variable of interest (e.g. income) and the time period (e.g. year). For the former two, there are many types and therefore combinations of these can produce multiple groups for analysis. For example, taxpayers could be defined by an income source (for example, employee, self-employed and pensioner) and there are many sources of income (for example, wage, self-employment, pension, rental and capital). In addition, incomes can be examined at different stages before and after taxes and transfers (market, gross and net). Analysis can be further deepened by demographic and socioeconomic variables such as age, sector of employment, geographic region and income range.

2. **Existing taxpayer groups can be refined and better understood by applying rules to the data.** Is a taxpayer with a combination of salary, self-employment and pension income best characterised as an employee, self-employed or a pensioner? One approach to address this question would be to apply an income source cut-off rule. For example, one definition for self-employed is a taxpayer that derives *any* income from self-employment while another, aimed at capturing those with significant self-employment activity, is taxpayers with self-employment income which is at least 15% of all income. Robustness is best served by comparing the results from several definitions. In Slovenia for example, a total of six taxpayer groups were defined.
3. **New groups, such as those participating in the ‘gig economy’, could be explored.** The structure of the labour market is changing as a result of a rise in non-standard “gig economy” jobs (OECD, 2019^[50]), which poses challenges for tax revenues and tax fairness, particularly if such work is subject to different rates under tax and SSC systems (OECD, 2018^[51]). To examine this phenomenon, an approximate definition of a non-standard “gig economy” worker could be explored through the tax record microdata. This might be possible in some countries where there is a clear definition of self-employment in the tax law and text data is available on principal business activity. For example, (Jackson, Looney and Ramnath, 2017^[52]) use tax administration data to define a set of individuals who are likely to be “gig economy” workers by focusing on individuals who filed specific forms including those who report labour income outside of a formal employee-employer relationship. Furthermore, using two decades of tax return microdata in the United States, (Collins et al., 2019^[53]) document the share of the workforce with income from alternative non-employee work arrangements.
4. **There is the potential to evaluate the take-up and effectiveness of government programmes that offer tax incentives.** For example, in Slovenia, a flat-rate regime was introduced in 2013 to encourage entrepreneurship where for the purpose of deductions, costs are presumed to be 80% of business income (regardless of the actual costs incurred). An analysis over time could reveal both the take-up across sectors of the economy as well as the likely extent of the tax loss under different simulations up to and including the time of abolition.
5. **The large-scale of tax microdata supports more advanced research designs** including quasi-experimental research, natural experiments and rare event analysis that would be otherwise not be feasible with traditional data sources. Such analysis can be enhanced by borrowing the new econometric techniques from the so-called ‘credibility revolution’ in empirical economics (Angrist and Pischke, 2010^[54]).

Tax microdata allows for multiple definitions of the same taxpayer group. Figure 2.13 (Panel A) presents the distribution of income for pensioners in Slovenia using two definitions of pensioner. The first definition includes those taxpayers with *any* pension income and the second includes those with *only* pension income (referred to as ‘full pensioners’). For most of the distribution, the two groups have similar earnings but this changes at the higher end of the income distribution. After the 75th percentile, the first pensioner group start to become significantly better off due to their additional income sources from employment and self-employment. For the top 10%, full pensioners earn well below the gross earnings of pensioners with any pension income.

Income sources change further up the income ladder. Figure 2.13 (Panel B) shows an employee group analysis across the income distribution for different income sources. Salary income comprises the vast majority of gross income across most of the distribution with the exception of the top end where, among the top 1%, capital income has greater concentration, comprising 12% of all income. This is consistent with the literature. (Förster, Llana-Nozal and Nafilyan, 2014^[39]) show that wages comprise the majority of income for top earners and that the further up the income ladder, the smaller the share of wage income and the higher the share of capital gains, capital income and business income. One interpretation of the analysis is low wages in Slovenia may cause some pensioners to supplement income through multiple sources and jobs.⁶

Figure 2.13. Selected Cohort Analysis in Slovenia

Source: Reshaping the Personal Income Tax in Slovenia, OECD (2018).

2.7.2. Detecting earnings clustering using bunching analysis

The large sample size and small measurement error of tax microdata have provided a foundation for the development of new empirical techniques in economics, which were previously unworkable.⁷ One such technique is bunching analysis, which examines behavioural responses to

taxation. This can be done by measuring how many, and to what extent, taxpayers choose to set earnings before and after cut-offs in tax incentives.⁸

A simple plot of the raw tax record data can detect local earnings clustering and provide a starting point for understanding the responsiveness to a tax. Unlike with traditional survey data, the exact location of taxpayers can be identified on the tax schedule. Bunching can provide compelling evidence of a behavioural response, although there are challenges to predicting the effects of policy changes (Kleven, 2016^[55]). There are two bunching concepts in the literature relevant to taxation: kink points when the cut-off relates to marginal rates (Saez, 2010^[56]); and notches when it relates to average rates (Kleven and Waseem, 2013^[57]). Kink points are common, for example in progressive income tax systems or the phase-in and phase-out stages of an EITC. In a well-known study in the United States, (Saez, 2010^[56]) finds that bunching is concentrated among EITC recipients with self-employment income but not with wage earners. In that case, the evidence suggested that most of the bunching response may be due to reporting rather than labour supply effects.

Bunching analysis for Slovenia provides suggestive evidence that employees and employers respond to the tax allowance schedule. In Slovenia, a tax allowance was available for incomes up to EUR 10 866 in 2016⁹ and, as income increased beyond this point, it was phased-out in stages. This allowance structure could produce economic incentives for taxpayers, and potentially employers, to maintain incomes below these thresholds before the allowance was reduced. Exploring this empirically with tax administration data, Figure 2.14 shows that the highest number of taxpayers in any thousand euro band is EUR 10 000 and EUR 11 000; just before the most significant loss in the tax allowance. The number of taxpayers continues to fall as the allowance is further reduced in steps. Although this simple plot does not isolate a causal effect¹⁰, the analysis provides suggestive evidence that Slovenian employees and employers may be responding to the tax allowances schedule.

Figure 2.14. Tax microdata can show where and how many taxpayers are responding to tax designs such as bunching before allowances are withdrawn

Number of taxpayers in thousand EUR income bands, 2016

Source: Reshaping the Personal Income Tax in Slovenia, OECD (2018).

2.8. Measuring the tax burden with complementary average tax rate statistics

Backward-looking average effective tax rates (AETRs) that use tax microdata are complementary to more commonly adopted forward-looking measures. AETRs are the share of all income paid in taxes. This section summarises some of the advantages and limitations of backward-looking rates.

Table 2.2. Tax rates developed based on microdata have key advantages

Advantages and limitations of various income tax rates

Tax Rate	Definition	Usage & Adoption	Advantages	Limitations	Skill Needed & Data Availability
1. Statutory Rates	The headline rate in a given income range	Widely cited in media	Simple signal of tax burden	Inaccurate measure of true tax burden or economic incentives	None; publicly available
2. Marginal Effective Tax Rates	The tax paid on an additional unit change in income	Widely adopted in economics	Good for future investment decisions; can account for combined effects	Theoretical assumptions; not based on actual behaviour	Econometrics; publicly available
3. Macro Average Tax Rates	A single economy-wide average tax rate (e.g. tax/GDP)	Widely adopted in economics	Consistent; comparable; harmonised; readily available	Unrealistically assumes single economy-wide measure	Statistics; publicly available
4. Micro Discrete Change Tax Rates	The tax paid on additional discrete change in income	Rarely adopted or cited	Indication of marginal rates; measures impact of tax system for those already in work; benchmark for METR models	Not true marginal rates; do not hold other factors constant;	Data science; Access to tax microdata
5. Micro Average Effective Tax Rates	The actual tax paid relative to gross wage income	Rarely adopted or cited	Precisely measure tax burden across subgroups & incomes; Sharp evaluation of fairness both across & within groups	Not optimal for returns on prospective investment (where past differs from future)	Data science; Access to tax microdata

Note: The table draws from (Clark, 2003^[58]).

Source: OECD.

Backward-looking AETRs based on macro data use actual taxes paid but rely on a single economy-wide estimate. Backward-looking AETRs have the advantage of taking actual taxes paid into account. They are often referred to as ‘effective’ as they implicitly account for the combined effect of statutory rates, tax deductions, tax credits, tax planning, tax relief, avoidance and evasion (OECD, 2000^[59]).¹¹ Macro or micro data can be used. The macro approach can involve using tax-to-GDP ratios from the national accounts (or aggregate tax data). These data have the advantages of consistency, comparability and availability. However, this kind of rate analysis is unrealistic because it assumes a single economy-wide average tax rate on income for all taxpayers. It also does not necessarily reflect differences in tax policies, across countries or over time (OECD, 2000^[59]) because tax rate schedules, tax allowances and tax credits can differ across income levels, household structures and taxpayers.

For measuring the tax burden, backward-looking AETRs based on tax microdata (micro AETRs hereafter) help distinguish the actual tax burden on specific tax bases. Tax complexity generally arises through the definition of the tax base rather than the definition of the tax schedule. Backward-looking micro AETRs can examine the tax burden across groups including different taxpayer types, age ranges, household structures, sectors of employment and so on. Taking combinations of these groups, the tax burden can be calculated precisely across a vast range of potential tax bases while retaining sufficient sample size. Therefore, micro AETRs calculated on well-defined specific tax bases are uniquely placed to identify the actual tax burden. Compared to forward-looking marginal effective tax rate (METR) models, backward-looking micro AETRs have the drawback of being less suited for measuring the effective tax on prospective investments.

Backward-looking micro AETRs have greater value when group differences within societies are large, whether in characteristics or behaviours. For example, disaggregated data on both labour and capital could improve fairness by allowing for greater sharing of the tax burden across the two income categories.

Backward-looking Micro AETRs provide sharp evaluations of fairness in the tax system both across and within income groups (referred to as vertical and horizontal equity respectively¹²). In the more common former case, such micro AETRs provide a measure of the progressivity of the personal income tax system through an all-in measure of the tax burden across incomes. In the latter case, the large sample size of tax microdata is useful for measuring the similarity of tax paid by similar households or family types with similar incomes. For example, (Gravelle and Gravelle, 2006^[60]) show how an EITC can create effective tax rates favouring families with two children in the United States.

Tax microdata allows for the calculation of discrete tax rate changes (DTR) that measure the change in the tax burden for a given change in gross wage income over time. DTRs are different from METR models because they measure discrete rather than unit changes and do not hold other factors constant (Clark, 2003^[58]). They also provide a complementary benchmark statistic for comparing with METR model results.

The limited use of backward-looking micro AETRs in tax policy has more to do with data availability constraints than their usefulness. Arguably, statutory rates are the most commonly cited income tax rate measures by the media. METRs and macro AETRs are the most widely used in academia and by policymakers. The prominence of such measures in both academia and policy is in part because they represent the path of least resistance for those who develop them – the data and techniques to calculate them are readily available. Developing AETRs using tax microdata has been more challenging because of the need to access and use complex tax microdata.

In Slovenia, simple backward-looking micro AETRs highlighted differences in actual tax burdens across cohorts and income levels. Table 2.3 shows average PIT rates at each rate band for employees, pensioners and self-employed taxpayers in Slovenia. Personal average PIT rates, which include employee SSCs, are also provided. Most taxpayers are concentrated in the first tax bracket (87% of pensioners and 84% of self-employed) and pay small amounts of PIT as their taxable income base is narrowed by high employee SSCs and substantial allowances. At the other end of the income distribution, and among pensioners, half of all PIT is paid by only 1.4% of pensioners (7 973 individuals) in the top two brackets.

In Slovenia, the AETR for employees is examined across each percentile in the income distribution. Figure 2.15 shows the personal average tax rate (both as a percentage of gross income and total labour costs) for employees in each percentile. PIT as a percentage of labour costs is extremely low at the bottom of the distribution and increases slowly but progressively for higher incomes. For example, at the bottom 10th percentile, where average total labour costs are EUR 9 102 (average gross incomes are EUR 7 834) the average tax wedge is 32%. In the 90th percentile, the tax wedge is 43%.

Table 2.3. Summary statistics using tax microdata can highlight key differences in tax burdens across cohorts and income levels

Taxpayers are concentrated in the lower rate bands, EUR millions 2016

Taxpayer cohorts and tax brackets	Number of taxpayers	Gross income	PIT	Average PIT rate (%)	Personal average tax rate (%)
Employees	741 670	14 629	1 685	12%	32%
16%	372 162	4 221	184	4%	25%
27%	296 684	6 620	749	11%	32%
41%	69 803	3 308	613	19%	38%
50%	3 021	480	139	29%	47%
Pensioners	583 530	5 452	134	2%	5%
16%	505 910	3 820	12	0%	1%
27%	69 647	1 257	58	5%	9%
41%	7 702	334	53	16%	26%
50%	271	42	11	28%	41%
Self-employed	69 000	772	69	9%	14%
16%	57 944	424	8	2%	6%
27%	7 910	181	20	11%	19%
41%	2 884	131	28	21%	27%
50%	262	36	13	37%	39%

Note: The average PIT rate is defined as the amount of PIT divided by gross income.

Source: Reshaping the Personal Income Tax in Slovenia, OECD (2018).

Figure 2.15. Backward-looking micro AETRs for employees at each percentile using tax microdata in Slovenia

The labour tax burden is high in Slovenia, 2016

Source: Reshaping the Personal Income Tax in Slovenia, OECD (2018).

Notes

¹ Tax microdata is also particularly useful for studying taxpayer behaviour through large-scale randomised control trials (RCTs). It is also useful for analysing the geographic variation in policies, such as geographic dispersion in income mobility following a tax change, or the effects of tax rate changes on cross-border revenues.

² Other metrics include the S80/20 and the Palma ratio (S90/S40).

³ There is an important distinction between wealth and capital that has been identified in the literature. In general, wealth reflects control over resources while capital is an input to the production process. Consequently, net wealth taxes are imposed on total net wealth, even if they generate no income, whereas personal capital income taxes are levied on income flows from assets (for example, income and property).

⁴ Isolating the contribution of each effect is possible but requires sophisticated modelling as the three effects are linearly dependent since for any person (birth year = current year – age).

⁵ The approach taken was to reduce the reported employee SSC on the tax records and to substitute this into an estimate of taxable income. The PIT implications of a base broadening measure, such as incrementally reducing tax allowances, can be modelled in a similar way.

⁶ A causality issue arises here in that it may be that low wages drive supplementary incomes or that those with supplementary incomes tend not to have low wages.

⁷ With the exception of a very small number of studies, bunching analysis using survey data is often not possible or extremely limited due to sample size and measurement error.

⁸ A similar technique is the regression discontinuity design, which is also enhanced by administrative data and tax microdata, the difference being that the assignment variable is designed rather than chosen as in bunching.

⁹ This allowance structure relates to 2016 and has since been changed in 2018 in Slovenia.

¹⁰ For example, it may be that bunching occurs due to tax avoidance or labour supply responses.

¹¹ In practice, effective tax rates are often lower than the aforementioned statutory rates because the taxable income base is reduced by tax provisions before statutory rates are applied.

¹² Vertical equity means that those with greater ability to pay should pay proportionally more while horizontal equity means that those with similar incomes and circumstances should pay the same.

3 Access Options

Despite its potential for tax policy analysis, access to and use of tax microdata remains limited and sporadic. However, tax administrations are increasingly adopting new technologies, statistical software and trained analysts to capture and use large volumes of data more securely. This chapter discusses several data access options.

3.1. Tax microdata access remains limited but is slowly increasing

The extent of tax microdata access and use is a decision for tax administrations. This section explores a range of data access modes which could be adopted by tax administrations including decentralised, extended decentralised, synthetic, remote and direct (OECD, 2014^[61]). Despite its potential for tax policy analysis, access to tax microdata remains challenging and its use limited and sporadic. Restrictions are primarily enforced to protect taxpayer confidentiality. However, tax administrations are increasingly adopting new technologies and statistical software to capture and process large volumes of data securely. This development could support new data access solutions where data are used more readily for tax policy research while limiting privacy risks through enhanced platform security and analytical techniques for anonymising and synthesising data.

A small but increasing number of researchers are accessing tax microdata. Many countries allow anonymised access to government tax administrative data through statistical agencies¹. Indeed, in a report on the role of well-being metrics in policy, (Stiglitz, Fitoussi and Durand, 2018^[62]) recommend improving the quality of inequality measures ‘by allowing Statistical Offices to use tax records to capture developments at the top end of the distribution’ and that access to administrative records ‘should be facilitated, in ways that preserve confidentiality’. In Denmark, Statistics Denmark provides researchers access to de-identified data. In Ireland, the Central Statistics Office (CSO) provides access to microdata under certain conditions. In the UK, the ‘HMRC Datalab’ provides a secure environment for accessing anonymised taxpayer data. However, most microdata access stops at the first national boundary making international collaboration with microdata the exception rather than the rule (OECD, 2014^[61]).

Increasingly, tax administrations are adopting new technologies and statistical software to capture, process and analyse large volumes of data securely. For example, many tax administrations have now developed automated ‘data pipelines’ which can support data capture (often high volumes of diverse data types), processing (such as cleaning and transformations), analysis and visualisation (for details, see (OECD, 2016^[8]); (PWC, 2018^[63])). These developments are more likely to lead to an expansion in tax and other policy analysis in the future.

The importance of taxpayer confidentiality for governments should not be underestimated by the research community. Tax administration (and statistics agencies) have to reconcile two conflicting objectives. On the one hand, they must share data to support public transparency, debate and policy. On the other, they must protect taxpayer confidentiality (which is required by law).² In a conflict between these two objectives, it is certain that confidentiality will prevail. In the future, these government data are likely to be increasingly matched with other data sources which will widen this trade-off - the data will be improved from an analytical perspective but simultaneously disimproved from a privacy perspective (as the risk of identifying individuals rises) (Connelly et al., 2016^[5]).

Taking advantage of improved access options will require greater levels of technological security and capability to ensure confidentiality. Figure 3.1 summarises selected data access options based on current best-practice and OECD experience of working with statistical agencies direct (Bruno and Ahmad, 2006^[64]). This framework has been developed to place various data access modes on a spectrum with differing advantages and limitations. In general, the higher access options provide greater data disaggregation and therefore greater analytical flexibility and potential. However, they also raise the risk of taxpayer confidentiality disclosure and encountering legal restraints. The lower access options are more likely to place a greater work burden on national officials while the higher access options shift that burden to researchers.

Figure 3.1. Selected Access Options

Note: This framework draws significantly from the work of (Bruno and Ahmad, 2006^[64]) at the OECD.

Source: OECD analysis.

3.2. Five data access options

Each of the five access options are discussed below.

- 1. Decentralised approach where national officials work on national data to produce a range of standard indicators.** This reflects current practice in tax administrations. National officials may produce this data and place it online or send it securely to researchers on request by secure email.
- 2. A decentralised extension of this approach is to include more advanced data structures such as longitudinal data.** This could include other advanced summary statistics and can also be made available online or sent to researchers on request by secure email. A public use-file could be an example of this.
- 3. Synthetic microdata mimic actual data but are constructed in a way that protects privacy so that it can be publicly released.** While there is no universally agreed method for producing synthetic data, the general idea is that statistical processes are applied to tax microdata to create a new synthetic dataset that maintains the same underlying relationships across variables but that cleans and blurs variables in a way that protects confidentiality. The data may include age, number of children, region of work, sector of employment and may also focus on those on very high incomes. One or many synthetic subsamples could be drawn. The process could introduce random noise to some variables or predicting variables using statistical regressions. More recent methods include classification and regression trees which sort observations into similar groups and draw from the distribution of outcomes that occur in each group (Burman et al., 2017^[65]). In developing synthetic data, a balance must be struck between higher quality data and the risk of disclosure. (Card et al., 2011^[19]) argue that synthetic data is ‘much less attractive than providing direct access because it is ‘virtually impossible for researchers to fully specify the contents of the ideal synthetic dataset in advance’. In addition, they argue that synthetic data make it difficult to study subpopulations unless they are pre-specified. They also argue that it will require a large infrastructure of intermediaries whose job it is to construct such data, effectively ‘adding noise to the existing data’, which researchers will try to remove (but may not succeed).
- 4. Remote execution is where researchers submit statistical analysis that can be run on tax microdata.** (Bartelsman, 2004^[66]) pioneered the *distributed microdata approach*, which aims to conduct identical analysis across multiple countries and analyse combined outputs. In practice, this would involve the user sending code for certain statistical programs (such as R, STATA, SAS or SPSS) to officials who run the code and report results (back to the user).³ This approach has been adopted at OECD for measuring firm productivity by using a centrally written code that is flexible and automated enough to run across different data in different countries (Berlingieri et al., 2017^[3]). This remote execution approach provides users with analytical output without the confidential data being seen. Legal contracts are likely to be required to clarify responsibility and

ensure confidentiality. The advantages of this approach include cross-country comparability, a reduced burden on agencies and low cost reproducibility. However, it requires complex collaboration and relationship building in addition to challenges of sharing the costs of resources and risks of secondary disclosure (Bertelsman, 2004). It also limits analysis to relatively simple tabulations and may only work well for clean data. (Card et al., 2011^[19]) argue that such methods remain 'substantially inferior to direct access'. In the case of the analysis for Ireland presented in this paper, a partnership model was adopted where data analysts from the tax administration worked with guidance from OECD experts in the relevant field.

5. **Direct access to anonymised tax administration data is usually through a local research data centre or a secure remote connection for certain pre-defined users.** (Card et al., 2011^[19]) argue that 'direct access to micro-data is critical for success' and that 'researchers are almost always best served when starting from the raw data'. Direct access was possible in the case of our analysis for Slovenia.⁴ An essential advantage of direct access is that a competent researcher can learn from exploring complex data. This inductive process, which includes simple summary descriptive statistics of subpopulations for example, can provide a grasp of the data or reveal anomalies that might otherwise be missed by simply examining regression output run by others. For these reasons, (Card et al., 2011^[19]) argue that the above alternatives of synthetic data or remote execution will not allow for cutting edge policy relevant research.⁵ However, full access to such data is generally restricted to tax authorities, made available often on specific request for tax policy analysis, and in a few countries to outside researchers under strict confidentiality conditions. In 2011, the OECD Expert Group for International Collaboration on Micro data Access was formed to examine the challenges for cross-border collaboration with microdata.

Notes

¹ National Statistics Offices (NSOs) may not always have access to tax records.

² The newly introduced General Data Protection Regulation (GDPR) in the EU is an important part of the latter.

³ Web-based interface systems without human intervention are also possible however they can be inflexible.

⁴ In Slovenia, anonymised tax administration data was sent to the OECD server through a secured connection.

⁵ This argument is based on an assessment in the United States.

4 Conclusion

This chapter summarises the paper with concluding remarks.

This paper explores tax record microdata and its value and potential for tax policy analysis. The paper draws on OECD collaborations with Slovenia and Ireland in 2018 where tax microdata was used.

While new technology and skills will be required, these data offer important advantages for tax policy analysis. First, tax microdata can deliver unique policy insights with simple statistical descriptions. Second, tax microdata can expand the range and depth of analysis available to tax policymakers. Third, use of tax microdata can help to provide new answers to old tax policy questions including to what extent taxpayers move up and down the income ladder over time and to what degree will the income tax base be eroded in an era of ageing. Fourth, these data can define and focus on almost any taxpayer cohort while retaining sufficient sample size for analysis. However, they also have limitations including little to no demographic information on taxpayers, inconsistent alignment with macroeconomic data and that they cover the taxpaying rather than total population. These limitations may partly be overcome by combining them with the richness of individual demographic information (from survey data) and the consistency of distributed national income (from national accounts data).

The paper also explores a range of data access modes which could be adopted by tax administrations including decentralised, extended decentralised, synthetic, remote and direct (Bruno and Ahmad, 2006^[64]). Despite its potential for tax policy analysis, access to tax microdata remains challenging and its use limited and sporadic. Restrictions are primarily enforced to protect taxpayer confidentiality. However, tax administrations are increasingly adopting new technologies and statistical software to capture and process large volumes of data securely. This development could support new data access solutions where data are used more readily for tax policy research while limiting privacy risks.

References

- Alvaredo, F., E. (2018), *Distributional national accounts; in Stiglitz, J., J. Fitoussi and M. Durand (eds.), For Good Measure: Advancing Research on Well-being Metrics Beyond GDP*, OECD Publishing, Paris, <https://doi.org/10.1787/9789264307278-8-en>. [24]
- Alvaredo, F. et al. (2017), *Global Inequality Dynamics: New Findings from WID.world*. [28]
- Angrist, J. and J. Pischke (2010), "The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics", *Journal of Economic Perspectives*, Vol. 24/2, pp. 3-30. [54]
- Auten, G., G. Rey Gee and N. Turner (2013), "New Perspectives On Income Mobility And Inequality", *National Tax Journal*, Vol. 66/4, pp. 893-912. [46]
- Auten, G. and D. Splinter (2018), *Income Inequality in the United States: Using Tax Data to Measure Long-term Trends*, Draft version subject to change, <http://davidsplinter.com> (accessed on 21 November 2018). [18]
- Balestra, C. and R. Tonkin (2018), "Inequalities in household wealth across OECD countries: Evidence from the OECD Wealth Distribution Database", *OECD Statistics Working Papers*, No. 2018/01, OECD Publishing, Paris, <https://dx.doi.org/10.1787/7e1bf673-en>. [35]
- Bartelsman (2004), *Statistics Directorate Committee On Statistics The Analysis Of Microdata From An International Perspective To Be Held The Analysis Of Microdata From An International Perspective*. [66]
- Bergin, A., E. Kelly and S. McGuinness (2015), "Changes in labour market transitions in Ireland over the Great Recession: what role for policy?", *IZA Journal of European Labor Studies*, Vol. 4/1, p. 9, <http://dx.doi.org/10.1186/s40174-015-0035-0>. [31]
- Berlingieri, G. et al. (2017), "The Multiprod project: A comprehensive overview", *OECD Science, Technology and Industry Working Papers*, No. 2017/04, OECD Publishing, Paris, <https://dx.doi.org/10.1787/2069b6a3-en>. [3]
- Bradbury, K. and J. Katz (2009), "Trends in U.S. Family Income Mobility, 1967 - 2004", *Federal Reserve Bank of Boston Working Paper No. 09-7*. [44]
- Bruno, G. and N. Ahmad (2006), *Draft Assessing The Feasibility Of An Oecd Micro-Database For Research Across Contributing Countries Via Remote Access 2 Assessing The Feasibility Of An Oecd Micro-Database For Research Across Contributing Countries Via Remote Access*. [64]
- Burkhauser, R. et al. (2018), "Survey Under-Coverage of Top Incomes and Estimation of [25]

- Inequality: What is the Role of the UK's SPI Adjustment?", *Fiscal Studies*, Vol. 39/2, pp. 213-240, <http://dx.doi.org/10.1111/1475-5890.12158>.
- Burman, L. et al. (2017), *A Synthetic Income Tax Return Data File: Tentative Work Plan and Discussion Draft*, Tax Policy Center, Urban Institute and Brookings Institute. [65]
- Callan, T., B. Maxime and J. Walsh (2018), "Income growth and income distribution: a long-run view of Irish experience", <http://dx.doi.org/10.26504/bp201903>. [33]
- Card, D. et al. (2011), *Expanding Access to Administrative Data for Research in the United States*, SSRN Electronic Journal, <http://dx.doi.org/10.2139/ssrn.1888586> (accessed on 21 November 2018). [19]
- Cavallo, A. and R. Rigobon (2016), "The Billion Prices Project: Using Online Prices for Measurement and Research", <http://dx.doi.org/10.1257/jep.30.2.151>. [13]
- Chetty, R. et al. (2015), *The Impacts of Neighborhoods on Intergenerational Mobility: Childhood Exposure Effects and County-Level Estimates*, http://scholar.harvard.edu/files/hendren/files/nbhds_paper.pdf (accessed on 21 November 2018). [42]
- Chetty, R. et al. (2014), *Where is the land of opportunity? The Geography of Intergenerational Mobility in the United States*, NBER working paper series, <http://www.equality-of-opportunity.org>. (accessed on 21 November 2018). [20]
- Clark, W. (2003), *Using Micro-Data To Assess Average Tax Rates Presented At Cesifo Conference On Measuring The Tax Burden On Labour And Capital, Venice, July 2002*, <http://www.CESifo.de> (accessed on 14 August 2018). [58]
- Collins, B. et al. (2019), *Is Gig Work Replacing Traditional Employment? Evidence from Two Decades of Tax Returns **, <https://www.irs.gov/pub/irs-soi/19rpgigworkreplacingtraditionalemployment.pdf> (accessed on 8 July 2019). [53]
- Connelly, R. et al. (2016), "The role of administrative data in the big data revolution in social science research", *Social Science Research*, Vol. 59, pp. 1-12, <http://dx.doi.org/10.1016/J.SSRESEARCH.2016.04.015>. [5]
- Couper, M. (2000), *Review: Web Surveys: A Review of Issues and Approaches*, Oxford University Press/American Association for Public Opinion Research, <http://dx.doi.org/10.2307/3078739>. [21]
- Einav, L. and J. Levin (2014), "Economics in the age of big data.", *Science (New York, N. Y.)*, Vol. 346/6210, p. 1243089, <http://dx.doi.org/10.1126/science.1243089>. [9]
- Einav, L. and J. Levin (2014), *The data revolution and economic analysis*, <http://www.journals.uchicago.edu/t-and-c> (accessed on 14 August 2018). [7]
- Flory, J. and S. Stöwhase (2012), *A Static Microsimulation Model of Personal Income Taxation in Germany*, https://microsimulation.org/IJM/V5_2/5_IJM_5_2_note_Flory_Stowhase_clean.pdf (accessed on 21 November 2018). [49]
- Förster, M., A. Llana-Nozal and V. Nafilyan (2014), *Trends in Top Incomes and their Taxation in OECD Countries*, OECD. [39]
- Gravelle, J. and J. Gravelle (2006), *Horizontal Equity and Family Tax Treatment: The Orphan* [60]

- Child of Tax Policy*, <https://www.ntanet.org/NTJ/59/3/ntj-v59n03p631-49-horizontal-equity-family-tax.pdf> (accessed on 21 November 2018).
- Groves, R. et al. (2009), *Survey Methodology, 2nd Edition Sampling of Populations: Methods and Applications*. [17]
- Jackson, E., A. Looney and S. Ramnath (2017), *The Rise of Alternative Work Arrangements: Evidence and Implications for Tax Filing and Benefit Coverage*, <https://www.treasury.gov/resource-center/tax-policy/tax-analysis/Documents/WP-114.pdf> (accessed on 21 November 2018). [52]
- Jäntti, M. and S. Jenkins (2013), *Income Mobility*, <http://ftp.iza.org/dp7730.pdf> (accessed on 21 November 2018). [43]
- Jenkins, S. and P. Van Kerm (2011), *The Measurement of Economic Inequality*, Oxford University Press, <http://dx.doi.org/10.1093/oxfordhb/9780199606061.013.0003>. [30]
- Kennedy, S. (2018), *Income Dynamics and Mobility in Ireland: Evidence from Tax Records Microdata*, <https://igees.gov.ie/wp-content/uploads/2018/04/Income-Dynamics-Mobility-in-Ireland-Evidence-from-Tax-Records-Microdata.pdf> (accessed on 5 November 2018). [1]
- Kennedy, S. et al. (2017), *SME Survey 2017*, <https://igees.gov.ie/wp-content/uploads/2018/04/SME-Survey-2017.pdf> (accessed on 21 November 2018). [14]
- Kleven, H. (2018), *Language Trends in Public Economics*, https://www.henrikkleven.com/uploads/3/7/3/1/37310663/language_trends_slides_kleven.pdf (accessed on 21 November 2018). [10]
- Kleven, H. (2016), “Bunching, Annual Review of Economics, Vol. 8, pp. 435-464, 2016”, <http://dx.doi.org/10.1146/annurev-economics-080315-015234>. [55]
- Kleven, H. and M. Waseem (2013), “Using Notches To Uncover Optimization Frictions And Structural Elasticities: Theory And Evidence From Pakistan”, <http://dx.doi.org/10.1093/qje/qjt004>. [57]
- Laney, D. (2001), *Application Delivery Strategies*, <https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf> (accessed on 21 November 2018). [4]
- Lustig, N. (2018), *Measuring the distribution of household income, consumption and wealth, in Stiglitz, J., J. Fitoussi and M. Durand (eds.), For Good Measure: Advancing Research on Well-being Metrics Beyond GDP*, OECD Publishing, Paris,, <https://doi.org/10.1787/9789264307278-5-en>. [11]
- Mazumder, B. (2015), *Estimating the Intergenerational Elasticity and Rank Association in the U.S.: Overcoming the Current Limitations of Tax Data (Revised September 2015) - Federal Reserve Bank of Chicago*, <https://www.chicagofed.org/publications/working-papers/2015/wp2015-04> (accessed on 21 November 2018). [41]
- Miller, P. (2017), “Is There A Future For Surveys?”, *Public Opinion Quarterly*, Vol. 81, pp. 205-212, <http://dx.doi.org/10.1093/poq/nfx008>. [15]
- Mitnik, P. et al. (2015), *New Estimates of Intergenerational Mobility Using Administrative Data*, <https://www.irs.gov/pub/irs-soi/15rpintergenmobility.pdf> (accessed on 21 November 2018). [38]

- OECD (2019), *OECD Employment Outlook 2019: The Future of Work*, OECD Publishing, Paris, [50]
<https://dx.doi.org/10.1787/9ee00155-en>.
- OECD (2018), *A Broken Social Elevator? How to Promote Social Mobility*, OECD Publishing, [40]
 Paris, <https://dx.doi.org/10.1787/9789264301085-en>.
- OECD (2018), *OECD Tax Policy Studies The Role and Design of Net Wealth Taxes in the* [36]
OECD, <https://www.oecd-ilibrary.org/docserver/9789264290303-en.pdf?expires=1542811951&id=id&accname=ocid84004878&checksum=59FAC442B5C3FC71056819E64573F1F1> (accessed on 21 November 2018).
- OECD (2018), *Opportunities for All: A Framework for Policy Action on Inclusive Growth*, OECD [26]
 Publishing, Paris, <http://dx.doi.org/10.1787/9789264301665-en>.
- OECD (2018), *Reshaping the Personal Income Tax in Slovenia*, <http://www.oecd.org/tax/tax-policy/reshaping-the-personal-income-tax-in-Slovenia.pdf> (accessed on 5 November 2018). [2]
- OECD (2018), *Tax Policies For Inclusive Growth In A Changing World - Tax Policies For* [51]
Inclusive Growth In A Changing World, <http://www.oecd.org/g20/Tax-policies-for-inclusive-growth-in-a-changing-world-OECD.pdf> (accessed on 23 October 2018).
- OECD (2017), *Preventing Ageing Unequally*, OECD Publishing, Paris, [45]
<https://dx.doi.org/10.1787/9789264279087-en>.
- OECD (2016), *Advanced Analytics for Better Tax Administration: Putting Data to Work*, OECD [8]
 Publishing, Paris, <https://dx.doi.org/10.1787/9789264256453-en>.
- OECD (2015), *In It Together: Why Less Inequality Benefits All*, OECD Publishing, Paris, [27]
<https://dx.doi.org/10.1787/9789264235120-en>.
- OECD (2014), *Chair of the OECD Expert Group for International Collaboration on Microdata* [61]
Access, OECD, <http://www.oecd.org/sdd/microdata-access-final-report-OECD-2014.pdf>
 (accessed on 18 February 2019).
- OECD (2000), *Tax Policy Studies No. 2 – Tax Burdens: Alternative Measures*, [59]
<https://www.oecd.org/eco/public-finance/36986711.pdf> (accessed on 14 August 2018).
- Piketty, T. (2014), *Capital in the Twenty-First Century*, Cambridge Massachusetts: The Belknap [34]
 Press of Harvard University Press, 2014.
- Piketty, T. and E. Saez (2014), *Inequality in the long run*, Science, American Association for the [29]
 Advancement of Science, 2014, 344 (6186), pp.838-843.
- Piketty, T., E. Saez and G. Zucman (2018), “Distributional National Accounts: Methods And [23]
 Estimates For The United States”, *The Quarterly Journal of Economics*, pp. 553-609,
<http://dx.doi.org/10.1093/qje/qjx043>.
- PWC (2018), *The Data Intelligent Tax Administration Meeting the challenges of Big Tax Data and* [63]
Analytics 2, <https://www.pwc.nl/nl/assets/documents/the-data-intelligent-tax-administration-whitepaper.pdf> (accessed on 19 February 2019).
- Saez, E. (2010), “Do Taxpayers Bunch at Kink Points? American Economic Journal: Economic [56]
 Policy, Vol. 2, No. 3, August 2010”, <http://dx.doi.org/10.1257/pol.2.3.180>.
- Saez, E. et al. (2014), *Wealth Inequality In The United States Since 1913: Evidence From* [37]

Capitalized Income Tax Data, <http://eml.berkeley.edu/~saezandhttp://gabriel-zucman.eu/uswealth> (accessed on 21 November 2018).

- Sakshaug, J. and F. Kreuter (2012), “Assessing the Magnitude of Non-Consent Biases in Linked Survey and Administrative Data”, *Survey Research Methods*, Vol. 6/2, pp. 113-122, <http://dx.doi.org/10.18148/SRM/2012.V6I2.5094>. [12]
- Savage, M. et al. (2015), *The Great Recession, Austerity and Inequality: Evidence from Ireland Non-technical Summary The Great Recession, Austerity and Inequality: Evidence from Ireland*, <https://www.esri.ie/system/files/media/file-uploads/2015-07/WP499.pdf> (accessed on 8 July 2019). [32]
- Schroeder, R. (2014), “Big Data”, in *Society and the Internet*, Oxford University Press, <http://dx.doi.org/10.1093/acprof:oso/9780199661992.003.0011>. [6]
- Slemrod, J. (2016), “Caveats To The Research Use Of Tax-Return Administrative Data”, *National Tax Journal*, Vol. 69/4, pp. 1003-1020, <http://dx.doi.org/10.17310/ntj.2016.4.13>. [16]
- Stiglitz, J., J. Fitoussi and M. Durand (2018), *Beyond GDP - Measuring what counts for social and economic performance*, <https://www.oecd-ilibrary.org/docserver/9789264307292-en.pdf?expires=1562678950&id=id&accname=ocid84004878&checksum=386F94C335B71DD9FFA5FF53546BA919> (accessed on 9 July 2019). [62]
- UN (2017), *World Population Ageing*, Economic and Social Affairs, United Nations, http://www.un.org/en/development/desa/population/publications/pdf/ageing/WPA2017_Highlights.pdf (accessed on 21 November 2018). [47]
- Yashio, H. and K. Hachisuka (2014), *Impact of Population Aging on the Personal Income Tax Base in Japan: * Simulation Analysis of Taxation on Pension Benefits Using Micro Data*, https://www.mof.go.jp/english/pri/publication/pp_review/ppr026/ppr026f.pdf (accessed on 14 August 2018). [48]
- Zwijnenburg, J., S. Bournot and F. Giovannelli (2017), “Expert Group on Disparities in a National Accounts Framework: Results from the 2015 Exercise”, *OECD Statistics Working Papers*, No. 2016/10, OECD Publishing, Paris, <https://dx.doi.org/10.1787/2daa921e-en>. [22]