Miguel Almunia, Jarkko Harju, Kaisa Kotakorpi, Janne Tukiainen and Jouko Verho

Expanding access to administrative data: the case of tax authorities in Finland and the UK

Article (Published version)
(Refereed)

Original citation:
Almunia, Miguel and Harju, Jarkko and Kotakorpi, Kaisa and Tukiainen, Janne and Verho, Jouko (2018) Expanding access to administrative data: the case of tax authorities in Finland and the UK. International Tax and Public Finance. ISSN 0927-5940

DOI: https://doi.org/10.1007/s10797-018-9525-0

International Tax and Public Finance
https://doi.org/10.1007/s10797-018-9525-0

POLICY WATCH

Expanding access to administrative data: the case of tax authorities in Finland and the UK

Miguel Almunia[1,2,3,4] · **Jarkko Harju**[5,6] · **Kaisa Kotakorpi**[5,6,7] · **Janne Tukiainen**[5,8] · **Jouko Verho**[5]

Abstract
We discuss typical issues in getting access to and using high-quality administrative tax data for research purposes. We discuss research involving both quasi- and field experiments implemented together with the tax authority. We reflect on practical solutions that promote co-creation of knowledge and reduce information asymmetries between researchers and practitioners, based on our experiences of working with the tax authorities in Finland and the UK. We provide examples of how to improve the overall research environment focusing on two successful case studies: the HMRC Datalab in the UK and the remote access to data from Statistics Finland. We propose two key arguments to persuade policymakers elsewhere to follow similar practices: improved data security and equality of access across researchers.

Keywords Administrative data · Data access · Field experiments · Tax administration

JEL Classification C81 · C93 · H20

1 Introduction

The availability of administrative data has expanded in many countries in the last decades. The potential benefits of using these data for research are immense, but they have not yet been fully exploited in most countries. Administrative data typically cover entire populations rather than samples and therefore have fewer problems with attrition, non-response and measurement error than traditional survey data sources (Card et al. 2010). The large size and high quality of the data also help conduct more reliable causal inference, increasing the statistical power to identify causal effects and facilitating the study of heterogeneity in those effects, which is both policy relevant and academically interesting. At its best, administrative data allow combining information

✉ Kaisa Kotakorpi
kaisa.kotakorpi@vatt.fi

Extended author information available on the last page of the article

Published online: 16 November 2018

from many different registers or authorities, thus resulting in very rich information contents that would be impossible to achieve with traditional surveys.

While the advantages of having access to high-quality administrative data for research are well known, the actual process of how researchers can gain access to such data is less often discussed. Obtaining access to the right data is in many cases a key hurdle in conducting high-end policy-relevant empirical research. In this paper, we discuss the typical issues in accessing and using high-quality administrative data and provide some possible solutions to these issues. Our arguments are mainly based on our own experience of working with the tax authorities in Finland and the UK, and thus, we use the tax authority as a running example. We also provide two case studies from Finland and the UK as examples of how to achieve improvements in the research environment at an institutional level.

The use of administrative tax data also has some drawbacks, as pointed out recently by Slemrod (2016). The data are not always error free, as there can be data-entry errors or incomplete consistency checks that fail to correct mistakes made by taxpayers (see, e.g. Gillitzer and Skov 2018). Tax return data, in particular, tend to contain very little demographic information beyond age and gender of the taxpayer, preventing researchers from studying a wide range of questions if the data cannot be combined with information from other sources. However, in many countries such as in Finland and other Nordic countries, individual-level data from different government registers can be merged using unique personal identifiers. Finally, it is important to keep in mind that governments collect data for specific operational purposes and try to minimize the costs of compliance for individuals and firms. Therefore, the raw data may not always be suited to address the research questions of a particular study.

Our discussion relates to various different types of processes of getting access to administrative tax data for research. At one end of the spectrum, some countries lack any procedure for external researchers to access confidential administrative data sources. In other cases, data access may be granted after lengthy negotiations on a case-by-case basis, but researchers often face challenges to work with the data because there is no detailed documentation and it is stored in different databases within the tax administration. In more advanced settings, there are standardized practices for data access and project approval. This is the case of the HMRC Datalab in the UK, and Finnish individual tax return data held at the Statistics Finland, both discussed in more detail below. Finally, at the other end of the spectrum, we could place the data generated in the context of field experiments where researchers collaborate directly with the tax authority. Some of the issues that arise in those cases are discussed by List (2011) in an overview of field experiments. It is worth noting that while randomized field experiments are an extremely valuable recent development, the bulk of tax research still concerns the analysis of data that already exist for administrative purposes.

Data management practices vary a great deal across countries and research projects, but the obstacles to data access often appear to be similar everywhere. One key consideration is how to align the different interests of researchers and administrators. A second issue is how to ensure that data confidentiality is maintained while not making the access and research too difficult and burdensome. Finally, it is important to ensure the continuity of the collaboration between academics and the administration throughout the often-lengthy research process and beyond.

By reflecting on the experiences in Finland and the UK, we discuss how to make administrative data more easily accessible to researchers, for the benefit of both the academic community and policymakers. For the UK case, we describe the access to administrative tax records through the HMRC[1] Datalab, which welcomed applications from researchers from May 2011. For the case of Finland, we discuss a different model where the national statistics office (Statistics Finland) holds data on individual tax returns and also a variety of other types of register data such as the Finnish linked employer–employee data. Both models have already generated a large number of academic contributions. We discuss the benefits and limitations of these two models of data access, extracting general lessons that may apply to other countries.

2 Types of collaboration projects between tax practitioners and researchers

Access to administrative data is crucial for empirical research and policy evaluation. Conducting high-quality research that combines large data sets (often involving data on entire populations) and state-of the art methods can provide valuable advice for policymakers and offer information to the general public. There are many types of research that can be done with administrative data, but here we focus on two approaches: ex-post evaluations of policy reforms and field experiments.

First, ex-post evaluation of policy reforms utilizes quasi-experimental methods, where the researcher uses, for example, changes in tax law to study the responses of firms or individuals. A large literature has focused on estimating the effects of changes in tax systems and administrative rules on firm and individual behaviour.[2] This field of research has increased our knowledge of how and why individuals and firms respond to taxation and administrative burden, especially showing that there is a huge heterogeneity in responses across different types of individuals and firms. These results are also very policy relevant and can be used in policy preparation when new tax changes have been designed.

Second, a researcher can create variation in the data by designing a field experiment, involving a randomized controlled trial. This typically requires intense collaboration with practitioners. There are many good experiences from this type of collaboration, for example, in the context of tax research using data from Nordic countries. Kleven et al. (2011) randomize audit probabilities for taxpayers in Denmark and conclude that the availability of third-party information is an effective deterrent of income tax evasion. Another example of a field experiment is from Norway by Bott et al. (2017). They find that both offering moral reasons for tax payments (the importance of taxes for financing public goods and services) and increasing detection probabilities are effective tools for reducing tax evasion. Kosonen and Ropponen (2015) offer information about the

[1] Her Majesty's Revenue and Customs (HMRC) is the UK tax authority. HMRC was formed by the merger of the Inland Revenue and Her Majesty's Customs and Excise, which took effect on 18 April 2005.

[2] To name just a few recent examples studying Finland and the UK, see Devereux et al. (2014), Kosonen (2015), Harju and Matikka (2016), Almunia et al. (2017), Best and Kleven (2018), Guceri and Liu (2016), Harju et al. (2017), Matikka (2018), Liu et al. (2017), Harju et al. (2018) and Kotakorpi and Laamanen (2016).

tax code for firms in Finland. They show that tax rules and legislation are not salient for firms, while directly advising firms can increase their knowledge of complicated details of tax systems.

In addition to credible causal inference, running experiments with official authorities has several benefits over other types of experiments. First, most of the information that needs to be recorded on the experimental subjects would have been collected by the government in any case, which allows for both subject-blind and observer-blind research designs (see, e.g. List 2011; Levitt and List 2011). Second, it is relatively cheap to collect the data. In particular, while implementing the treatment may be costly, the control group can be as large as even the entire population at virtually no costs.

3 The role of co-creation and information in accessing administrative data

Despite the benefits of using administrative data, the process of getting access to data is not always straightforward. Intense cooperation between researchers and practitioners is often required both in the case of field experiments and research projects using existing data and quasi-experimental variation. Below, we discuss some of the challenges that may arise at different stages of such collaboration projects, as well as point to some potential ways forward.

We emphasize two themes in this section. The first is co-creation of knowledge, a process that requires the complementary expertise of both practitioners and researchers and produces knowledge that benefits both parties.[3] The second is managing the information flows between them. The latter is naturally required for the former. We divide the section into steps that can be taken before, during and after the project to facilitate these goals. We do not claim that this approach is the only way to pull through a successful project but merely discuss common challenges in cooperation that we have experienced in our projects and then offer potential solutions to them. In Sect. 4, we will turn to technical hurdles associated with data access.

3.1 Before the project

3.1.1 Information: acknowledging the differences in objectives and constraints between researchers and practitioners

The objectives of the tax administration and academic researchers are inevitably quite different. While the objective of research is to understand the mechanisms of firm or individual behaviour, the main task of the tax administration is to ensure the accurate and timely collection of taxes (although many tax authorities do undertake prioritized research and analysis activities themselves). Commitments to carry out research with academics clearly take time off these core activities. To promote fruitful dialogue, we

[3] Knowledge is a public good, which tends to lead to inefficiently low private contributions towards it. The complementary nature of the inputs to the public good in the case at hand may help reduce those inefficiencies (Cornes and Sandler 1984).

believe it is important for each party to acknowledge and communicate the differences in objectives and the legitimate concerns that arise from them. For example, the tax administration might have a long tradition of carrying out risk-based audits, and a researcher's request to carry out audits at random in association with a field experiment may appear as interfering with their own operations. Making changes to the routines is costly and possibly entails risks that an individual practitioner is likely reluctant to take.

Further, inaccurate beliefs about the other's objectives may lead to a perception of ignorance or arrogance, and such misunderstandings may hinder cooperation. For example, a researcher who is interested in the effects of tax enforcement on market outcomes may think that the tax administration should share this interest, in particular if market outcomes affect tax revenue. However, an individual practitioner whose key tasks are more narrowly defined rarely enters a project solely based on such broader considerations.

More generally, an important first step is simply to get to know each other in terms of methods, objectives and constraints. Understanding and accepting the differences in objectives is crucial for fruitful cooperation. One way to start closing the gap is to acknowledge that the long-term objectives of the administration and researchers may be much more similar than their short-term objectives. This is what we turn to next.

3.1.2 Co-creation: building common ground

Despite the differences in objectives and constraints between researchers and practitioners, we have found several opportunities for building common ground. Research collaboration between researchers and tax authorities can be seen as an investment: The lessons learned can be used to design more effective tax collection in future. In this context, we propose making a case for a carefully thought-out research design that produces reliable results, despite it being time-consuming, over a quicker study that may yield misleading advice. Good experiences from past projects and other countries can provide examples of the benefits of research collaboration.

Concrete examples of ways in which research results can be utilized are an important way of highlighting the potential benefits of the investment. That is, research findings can be tied closely to informing the core activities of the tax administration. For example, knowledge of the determinants of tax evasion may help in designing more effective audit rules; research can inform policymakers on how to increase the effectiveness of third-party information in tax collection. Further, it may be possible to utilize research results to develop new performance measures that, for example, take long-run gains of tax enforcement into account. Such measures would help tax administrators in quantifying the benefits of their investment. Another related development is increased interest in alternative means of persuading taxpayers to report correctly. Steering taxpayers to comply without audits is likely to be cost-effective, but requires new types of measures of administrative effectiveness. For example, audit hit rate is not a very useful measure of the quality of audit design if most taxpayers report correctly. On the other hand, this requires measuring the effects of different types of interventions, which is an interesting research question in itself.

Revenue authorities can also participate in the formulation of research questions, for example by regularly involving academics in the evaluation of their newly adopted practices. Policy relevance and academic interest often coincide, and thus cooperating in the process of defining research questions provides a great starting point for a fruitful collaboration.

3.2 During the project

3.2.1 Co-creation: project planning and timing

Correct timing is crucial for a successful research project. Here, we identify some of the risks related to timing. Again our key lesson is that the planning stage of every project also benefits from being a joint effort between practitioners and researchers.

First, field experiments are very time-consuming initiatives: They need to be planned carefully to make sure that the design uncovers the causal effects of interest. There can also be long lags in preparing data. This is the case in other types of research projects as well, but in many other situations delays are a nuisance rather than a critical threat to the entire project. In the context of tax research, timing is a first-order issue: for example, the tax administration has a fixed schedule for sending out tax forms, carrying out audits, etc. If data required at a given stage of a project miss a deadline, this may prevent the following steps from being implemented. Consider as an example a field experiment that involves studying the effect of treatment letters on income tax reporting. Treatment letters need to be sent out at about the same time as tax forms, for example in March, to avoid some people filing their taxes before receiving the treatment letter. If there is a delay in obtaining and merging background data required to form the target group of the experiment, for example tax information from the previous year that would be available say in February, the project might be delayed by an entire year. On the other hand, if audit resources have been allocated to the project for the current year, such a delay may not be feasible or would require a renewed commitment to alter auditing routines for the following year. Such changes would increase the risks associated with the project going through. Hence, timing is crucial and projects need to be planned well in advance to meet the deadlines. Moreover, the timing of an experiment should not interfere with normal tax collection practices. This is important for the tax authority, but also from the point of view of ensuring the integrity of the research design if the natural treatment of taxpayers is changed.

Second, while randomized field experiments are an extremely valuable recent development, the bulk of tax research still concerns the analysis of existing data and often takes the form of ex-post evaluation of reforms. In these instances, however, whether a policy lends itself to credible evaluation is often a matter of luck: the crucial questions are whether the policy has been designed in such a way that it can be evaluated (e.g. there exists a suitable comparison group to construct a counterfactual), and whether the necessary data are available (e.g. there are data available on outcomes prior to the reform). In particular, phased-out implementation of reforms is key, as the effects of a policy that is implemented on the entire population at once (e.g. as in the case of electronic tax filing in Finland) are difficult to evaluate. Even if there is a proper

comparison group to study a tax reform, we have found it useful to consult the tax administration to be certain that there were no other changes in the treatment of either group (treatment or control), for example in the tax reporting rules or special tax audit campaigns for certain industries, at the time period of interest. These types of changes could threaten the research design or at least affect the interpretation of the results obtained.

To address these issues, we emphasize regular and systematic dialogue between researchers and the tax administration. One way to assure frequent interaction is to set up a regular meeting schedule with short intervals, for example monthly or quarterly. If researchers learn about planned reforms ahead of time, researchers and policymakers can enter into a discussion on how to plan reforms in such a way that they can be credibly evaluated, and it can be ensured that necessary data gathering is carried out before implementing a policy change. Researchers can also provide expertise in planning pilot programs. On the other hand, practitioners have invaluable information on the behaviour of taxpayers, what are the most relevant policy questions, the details of the data available and the institutional and operational constraints.

3.2.2 Information: finding the right data and merging different sources

Typically, an ambitious research project requires different types of data from different sources that need to be linked together. The hurdles in getting access to and linking administrative data are still often non-negligible, and sometimes insurmountable.

A first challenge is to know what data actually exist. A lot of data are typically hidden in tax administration registers, without documentation that would be accessible to researchers. It is also often the case that different types of tax records are held at different departments within the authority, and thus, any single tax authority employee is unaware of all the potentially available data. This is mostly due to the different goals of various departments within the authority so there is no general interest to collect the data resources together.

The original purpose of administrative tax data is usually not to conduct empirical research. Understanding the content of these data in depth can be difficult for a researcher, and in our experience close cooperation with data specialists in the tax administration is often crucial. In addition, data recording practices can change over time, which has to be taken into account when constructing panel data sets. Also, it is important to be aware of the precise nature of the data, for example whether it is based on taxpayers' original reports or final (post-audit) tax information. The type of data required, of course, depends on the research question at hand.

Working towards a comprehensive documentation of tax administration data would be extremely helpful for researchers and most likely also quite useful for tax authorities themselves. For example in Finland, data on all individual income tax returns are held at the national statistical office (Statistics Finland), with appropriate documentation, while most other tax data are held at the tax administration and access is typically subject to negotiation on a case-by-case basis. A researcher should first contact the tax administration to find out whether the data required are available, and then apply for access to the data by submitting an application form together with a research proposal.

Some limited information on the Finnish tax authority data availability and content is published on their website.[4]

Regarding information on data availability, the HMRC Datalab is ahead of the Finnish tax authority as a complete list of the main data sets available in the HMRC Datalab is published on their website.[5] The website also includes a more detailed directory, the data catalogue,[6] which lists all data held by HMRC that could be made available to researchers. HMRC encourages researchers to contact the department if they wish to discuss these data sets, and similar to Finland, this is subject to negotiation on a case-by-case basis.

3.2.3 Co-creation: ensuring commitment throughout the project

Even after suitable data have been found, continued cooperation is key to make full and correct use of the data. Maintaining commitment on either side in a lengthy project and ensuring that necessary information flows in both directions can be challenging for a number of reasons.

First, tax administrations are large organizations, and their mere size and organizational structure pose challenges for interaction with researchers. According to our experience, it is important to have a clear commitment from the tax authority to the research project right from the beginning. For example, it may be easy to get a data analyst excited about a project in his/her field of expertise, but if commitment is not sought at a high enough level initially, risks related to the final project approval may be resolved only quite late in the planning process after a lot of resources have already been invested on each side. It is not plausible for an individual researcher to contact the general director of the tax administration, but the contact person should have the authority to make a decision on project procurement. For example, if the project studies corporate tax evasion, the commitment to carry out the project would likely come from the head of the corporate tax unit. One way forward would be for the tax administration to develop standardized practices for procuring research projects.

Second, the size and structure of the tax administration also implies challenges for communication during project execution. It might be hard to control what exactly goes on within the organization, whether the right instructions are passed on to the right people, so that the carefully thought-out research design is preserved at all stages. Information should flow in both directions: While researchers need to be committed to communicate the requirements of the research design to the tax authority, tax authority personnel on the other hand can share their expertise regarding data, the tax system and relevant reforms and what they have already learned about the behaviour of tax payers and goals of policymakers.

Third, frequent turnover of personnel at the tax administration is quite common. There may also be organizational changes that suddenly move a key person to a new role or to a different part of the organization. Such changes could prove disastrous especially for a project in its early stages unless a sufficient number of individuals in

[4] Available at: https://www.vero.fi/tietoa-verohallinnosta/tilastot/.

[5] Available at: https://www.gov.uk/guidance/hmrc-datalab-datasets-available.

[6] Available at: https://www.gov.uk/government/publications/hmrc-data-catalogue.

the tax administration are committed to the project initially. This also highlights the importance of clarifying the roles of different parties from the beginning of a project and sharing information about changes in those during the project.

A solution to these challenges is to have a project steering group that is sufficiently broad to ensure continuity in case of organizational changes and diverse to ensure that all necessary areas of expertise are covered. Ideally, it would involve individuals with detailed hands-on expertise on the precise data to be used; individuals with broader knowledge of developments within the specific branch of tax administration (e.g. changes to tax enforcement practices); and an individual who has the authority to make independent decisions regarding the execution of the project.

3.3 After the project

3.3.1 Co-creation: planning public relations

Public relation (PR) concerns are central to the debate related to joint projects, experiments and data access. When researchers and tax authorities carry out joint projects, they should agree on how to inform the public—not only about the results, but also about ongoing projects. This should ideally be coordinated already at the early stages of the project, for example before any experiments are implemented. For example, in the case of field experiments, there should be a common understanding of what type of information is given to the public at the time of the experiment so that the research design is not compromised. If an experiment with new enforcement measures is covered in the media, for example, the control group will receive some of the "treatment", and the results will underestimate the true effect of the policy under study.

As a public body, the tax administration is concerned about the fair and equal treatment of citizens. For example, sending audit threat letters associated with randomized audit studies may seem to interfere with this principle. To give a positive twist, researchers may argue that giving information on the audit probability in advance is a service to the taxpayer, and law-abiding citizens should not be affected by such letters (only tax evaders). Of course, the psychology of receiving an audit threat letter might be more complicated than this, and letters should be carefully crafted to avoid any unintended negative reactions.

Another consideration related to PR concerns has to do with data access. If any leakages of confidential data were to occur, the reputational costs both to the organization and the individuals involved could be enormous. Here, however, the incentives of the tax administration and the researcher are in fact aligned: the reputational cost to any individual researcher from a data leakage would be equally detrimental.

3.3.2 Information: disseminating the results

A key priority for researchers is to publish research results in peer-reviewed academic journals. However, other modes of disseminating research results are also important. A carefully planned dissemination strategy will make the benefits of investing in a

research project more salient for both parties—and therefore help in bridging the gap between the goals of researchers and tax practitioners.

Results of research projects that use administrative data are usually informative and helpful for the officials producing the data, as well as for policymakers. Even though writing policy-oriented reports is often regarded as unattractive by researchers, this type of work should not be underrated as new information can only be beneficial if legislators and practitioners are aware of it.

In this respect, one important reason for disseminating results in a form that is transparent to practitioners again arises from the different objectives of tax officials and researchers. Much of the work of tax administrators involves dealing with special cases, and therefore an individual administrator might not have an accurate view of the broader picture, such as the nature of compliance behaviour of taxpayers more generally. Research results based on statistical analysis of large data sets can be informative in this respect and increase knowledge overall within the administration. Further, simply presenting the research design and details of the econometric analysis (and perhaps also sharing programming codes that are used to obtain the results) to practitioners could be beneficial in their work.

To achieve still broader impact, reports and policy briefs targeted at policymakers and the general public can be drafted. Ideally, the results of tax research can inform policymakers of how to design a better tax system. This has potentially very large societal benefits. Many modern governments collect as much as about half of GDP as tax revenue, and doing this as effectively as possible is a key priority.

4 Accessing data: technical hurdles and solutions

Data access practices vary a lot across countries and have a crucial effect on the ability of researchers to effectively utilize the data. Data protection legislation governs the use of personal data in research. It dictates to what extent research data must be anonymous, i.e. allows the possibility that persons can be identified directly or indirectly using other publicly available data. In practice, individual privacy is also affected by the technical and physical environment where the data are analysed: appropriate technical solutions are required to ensure that only those with a research permit are able to access the data. These solutions are referred to as data security. It is important to note that researchers and practitioners have a very strong common interest in maintaining data security. Any data leakages would be equally detrimental to a person's career prospects on either side.

Merging data across registers, for example using the social security number as in Finland, is not possible in all countries. For example, the UK does not have a unique personal identifier that would be identical across administrative registers. HMRC uses a variety of sources to link information about the same taxpayers across different data sets, but most of that information is not available to external researchers. In some cases, the HMRC Datalab has linked multiple data sets and then provided researchers with the resulting de-identified data. One example is the linking of corporation tax records with the business financial statements (from FAME), which has been used by

Devereux et al. (2014) and others. We review experiences from the UK below in more detail in Sect. 4.1.

A very convenient solution that provides good data security, reasonably convenient access, ability to merge data across registers and information easily available in known format is to collect data from all administrative sources to the nation's statistical office and create a remote access system for researchers. Introducing remote access to administrative data also removes the travel costs associated with having to use the data at one, fixed location. This increases efficiency and is also important for equal treatment of researchers based at different universities and research institutions. Data security can be ensured through regulating the locations where remote access can be used, appropriate technical solutions and output controls that ensure that any results are only retrieved in a form that does not reveal any personal information.

Persuading the national statistics office to set up the remote access facilities may, however, be complicated. To highlight the practical hurdles, we describe in detail the process of introducing remote access to data at Statistics Finland in Sect. 4.2. We use first-hand information on this, as one the authors of this study, Jouko Verho, was the project manager on Statistic Finland introducing remote access. Because most countries still do not have this useful service in place, the Finnish lessons could be helpful for achieving a similar development in other countries (see, e.g. Card et al. 2010 for the US debate).

4.1 Case study I: the HMRC Datalab in the UK

The HMRC Datalab was set up in 2011 to provide researchers with controlled access to de-identified administrative tax records for the UK. The Datalab started as a pilot project to conduct a research study on the responsiveness of business profits to changes in the corporation tax rates (Devereux et al. 2014). In 2013, HMRC expanded the capacity of the Datalab to six computers connected to a dedicated Windows server and allowed a larger number of projects to take place. As of March 2017, the HMRC Datalab had approved 71 research projects by researchers belonging to more than ten different UK institutions (mainly universities and other academic institutions, but also some regional governments and think tanks).[7]

In order to obtain access to administrative data at the HMRC Datalab, researchers have to complete and submit a project proposal explaining the research design and how the project fulfils at least one of HMRC's core functions.[8] The proposals are reviewed in quarterly meetings by the Datalab committee, which includes stakeholders across HMRC. If approved, research can begin. Researchers new to the Datalab attend a one-day Safe User of Research Environments (SURE) training and are required to pass a test before being able to start working with the taxpayer-level data.

The Datalab is physically located at one of HMRC's offices in London. The secure room consists of an office with six computers that are connected to a dedicated server

[7] For details, see: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/600820/DatalabProposalsApproved_March2017.pdf.

[8] For information about the application process, see: https://www.gov.uk/government/organisations/hm-revenue-customs/about/research#the-hmrc-datalab.

that contains all the available data sets, which include all the major taxes, some surveys and import and export data.[9]

Each researcher has access only to the data sets that he/she requested in their proposal. At the moment, researchers do not have remote access to the data, and they are only allowed to work on the data during regular office hours at the HMRC offices in London.[10] This is due primarily to concerns about data confidentiality given the legislation covering taxpayer confidentiality in the UK, rather than technical limitations, but costs to authorities associated with providing a remote access system must also be considered. The lack of remote access implies that the cost of accessing the Datalab is unequal across researchers, depending on their distance from London. The restrictions on the location and opening hours of the Datalab contribute to slowing down the completion of research projects because the schedule overlaps with other activities performed by most researchers, such as teaching and academic seminars. As we discuss in the next section in the context of Finland, remote access can be a safe and efficient alternative (as also emphasized by Card et al. 2010).

Once researchers have obtained results from the analysis, they submit an output request to the Datalab team. One member of the team reviews the files to ensure that all statistical disclosure rules are respected, and then sends them for further review to a specialist on the specific tax being analysed within HMRC's central analytical directorate. The same procedure applies when researchers have completed a draft of their working papers or presentation slides for conferences. The processes of output release and working paper review usually take between 1 and 3 weeks, although in some cases it may take longer depending on the time pressure on HMRC's personnel.

It is worth noting that HMRC currently makes available an *anonymized* version of the Survey of Personal Incomes in the UK Data Archive.[11] This consolidates raw administrative data sets to provide comprehensive information for around 600,000 taxpayers that are representative of the population. Following registration and approved licensing, this data set can be downloaded by a researcher to work from a remote location and a researcher does not need to travel to the Datalab. HMRC makes available the Public User tape as well as the full Survey of Personal Incomes (de-identified) through the HMRC Datalab, to allow joining to other data in the Datalab.

4.2 Case study II: Statistics Finland and remote access

Finnish legislation on data protection has been fairly liberal in the sense that it has always allowed the use of high-quality administrative data for scientific research. The main data source for both empirical microeconomic and macroeconomic research has been Statistics Finland because it is responsible for the vast majority of official

[9] The full list of data sets available can be found at https://www.gov.uk/guidance/hmrc-datalab-datasets-available.

[10] The computers in the Datalab are not connected to the Internet, nor is it possible to extract any information through the USB ports. To access the Datalab's secure room, researchers are required to leave all of their electronic equipment (laptops, mobile phones) in outside lockers. This is to ensure that no results can be extracted without supervision, but has the cost of preventing researchers from using useful resources such as online help related to econometric software packages.

[11] See https://discover.ukdataservice.ac.uk/Catalogue/?sn=8044&type=Data%20catalogue.

statistics in Finland. High-quality data sets have been created as a by-product of official statistics when microdata have been collected from various administrative sources.

In early 2000s, Statistics Finland had two methods of releasing confidential microdata for research purposes. Samples of individual-level data were released directly to researchers after ensuring that data were anonymous to a sufficient degree. Access to firm-level data was more restricted due to Statistics Finland's concerns related to the survey response rates of Finnish firms. High-quality firm data could only be accessed in Statistics Finland's Datalab in Helsinki.

Running the Datalab was considered cumbersome for Statistics Finland because these types of services were not its core activities. The solution was also found unsatisfactory by researchers because of travel time costs and lacklustre work environment. With an aim to provide a better research environment, to address the geographical inequality between researchers and to improve data security for all microdata research, Statistics Finland started a project to develop a remote access system in 2008.

The remote access system has been operational since the beginning of 2009.[12] It consists of several Windows servers designed for statistical computing which can be accessed from Finnish universities and research institutes after they have been audited by Statistics Finland. A strong identification is required in the login process, and the researchers are unable to transfer any files in or out of the system themselves. All extracted files are first checked by Statistics Finland's personnel to ensure data protection.

At first, Statistics Finland decided to provide all firm-level data sets via remote access. Also individual-level data were available, but only after anonymity screening, which meant that the remote access benefit was not yet at its full potential. However, the secure analysis environment and the significantly improved control of data over their lifespan in the system contributed to a government proposal to allow also the use of rich population-wide individual-level data in research remotely. Although some statistical officials objected to this change on the grounds that it risks public confidence in their operation (Ministry of Finance 2012), the bill was accepted in 2013.

After the change in legislation, a linked employer–employee panel covering the Finnish workforce was provided in the remote access system. This became quickly the main data set used in economic research because it contains a rich set of socio-economic characteristics including details on personal taxation. The financial statements of firms or other firm data can be linked to individual data using firm and establishment identifiers.

Statistics Finland has also improved the application process for research permission. A principal investigator files a simple application form supplemented with a research proposal and the description of required data. The key improvement in the process was the provision of ready-made data sets with online documentation.[13] This documentation was not publicly available before, which made drafting research proposals difficult as researchers did not have detailed knowledge about the available data.

[12] See http://tilastokeskus.fi/tup/mikroaineistot/etakaytto_en.html.

[13] Description of data sets: http://tilastokeskus.fi/tup/mikroaineistot/aineistot_en.html and https://taika.stat.fi/en/.

The popularity of the remote access system has grown over time. The number of researchers using the system was around 70 in 2011 (Johnson 2017). In 2017, the number of users had reached 320, and the number of active research projects was 140. A total of 48 different research organizations have a contract with Statistics Finland to use the system.

Statistics Finland has been under constant pressure to increase the system's capacity. To provide sufficiently powerful servers for the service, Statistics Finland decided to outsource the system to CSC (Finnish IT centre for computing) in 2015.[14] CSC is able to provide scalable computing resources for research projects. After outsourcing the servers, Statistics Finland has focused on improving data descriptions, inclusion of new data sets and developing an electronic application process for data access.

Finally, researchers can also bring in their own data sets, possibly originating from other official registers, into Statistics Finland's system. All administrative data sets in Finland can be linked via personal identification codes. It would be extremely useful to work towards the inclusion of various types of register data in Statistics Finland's databases, with automatic updates. For example, other types of tax data besides data on individual tax returns, pension data and health data still need to be applied for on a case-by-case basis from the corresponding authorities.

5 Conclusions

Use of administrative data has great potential in empirical research, but getting access to suitable data is often a non-trivial hurdle. We have discussed the process of getting access to administrative data in the context of tax research in Finland and the UK and two types of research designs (field experiments and quasi-experiments). The hurdles relate to both technical aspects (e.g. ensuring data protection and linking different data sets) as well as to achieving smooth cooperation and information flow between researchers and practitioners. Even when technical issues are resolved, making proper use of the data (e.g. understanding the contexts of complex data sets built for administrative purposes) often requires intense cooperation between academics and practitioners. This relates to both quasi-experimental research designs, where cooperation ensures that researchers learn about policy reforms in time to be able to conduct a proper evaluation, as well as to field experiments where cooperation is obviously an integral part of the data generating process itself.

While we have proposed some intuitive practical guidelines on how to handle these issues when researchers work with the tax authority case-by-case, our main message is that it is possible to build institutions that help to solve the technical, operational and informational issues involved. We used the UK's HMRC Datalab and the Statistics Finland remote access as examples of substantial improvements in the research environment. Having administrative data stored in one place offers many benefits through the ability to link information from different registers, having a clear documentation of the available data and through reducing transaction costs related to data access. Stan-

[14] CSC remote access service (FIONA, Finnish Online Access): https://www.csc.fi/-/fiona-tilastokeskuksen-etakayttojarjestelma.

dardized practices and technical solutions for project approval and data protection also solve many potential challenges related to cooperation and information. There are two key arguments that may persuade policymakers to implement similar practices in countries where they do not yet exist. First, data protection and security actually increase with these arrangements, rather than decrease, as many practitioners seem to fear before the implementation. Second, remote access further reduces transaction costs associated with using administrative data for research purposes and ensures the equal treatment of researchers from different institutions.

Acknowledgements This paper summarizes and develops findings from a workshop "Bringing Together Tax Researchers and Tax Authorities: Learning about the Process" organized in Helsinki at VATT Institute for Economic Research in May 2017 in collaboration with the University of Warwick Department of Economics. The Workshop was funded by Public Economics UK, Academy of Finland (Grant No. 304807) and VATT. We thank the workshop participants for sharing their views and Essi Eerola, Aki Savolainen from the Finnish Tax Administration, Mark Brewin and Yee Wan Yau from HMRC and the editors for useful comments on the paper. The Finnish Tax Administration neither endorses nor disagrees with the views and opinions presented by the authors. While Her Majesty's Revenue and Customs (HMRC) has ensured factual information relating to the HMRC Datalab service and legislation underpinning the use of de-identified taxpayer-level data in the Datalab is correct at the time of publication, HMRC neither endorses nor disagrees with the views and opinions presented by the authors regarding the HMRC Datalab service or the paper's recommendations. Harju gratefully acknowledges financial support from the Foundation for Economic Education. Open access funding provided by University of Turku (UTU) including Turku University Central Hospital.

References

Almunia, M., Lockwood, B., & Scharf, K. (2017). *More giving or more givers? The effects of tax incentives on charitable donations in the UK*. Working paper, University of Warwick.

Best, M. C., & Kleven, H. J. (2018). Housing market responses to transaction taxes: Evidence from notches and stimulus in the UK. *Review of Economic Studies, 85*(1), 157–193.

Bott, K. M., Cappelen, A. W., Sorensen, E., & Tungodden, B. (2017). *You've got mail: A randomised field experiment on tax evasion*. NHH Department of Economics Discussion Paper No. 10.

Card, D., Chetty, R., Feldstein, M., & Saez, E. (2010). *Expanding access to administrative data for research in the United States*. Retrieved September, 2017 from https://eml.berkeley.edu/~saez/card-chetty-feldstein-saezNSF10dataaccess.pdf.

Cornes, R., & Sandler, T. (1984). Easy riders, joint production, and public goods. *Economic Journal, 94*(375), 580–598.

Devereux, M., Liu, L., & Loretz, S. (2014). The elasticity of corporate taxable income: New evidence from UK tax records. *American Economic Journal: Economic Policy, 6*(2), 19–53.

Gillitzer, C., & Skov, P. E. (2018). The use of third-party information reporting for tax deductions: Evidence and implications from charitable deductions in Denmark. *Oxford Economic Papers, 70,* 1–25.

Guceri, I., & Liu, L. (2016). *How effective are fiscal incentives in R&D-intensive sectors?* Working paper, Oxford University Centre for Business Taxation.

Harju, J., Kosonen, T., & Skans, O. N. (2018). Firm types, price-setting strategies, and consumption-tax incidence? *Journal of Public Economics, 165*, 48–72.

Harju, J., & Matikka, T. (2016). The elasticity of taxable income and income-shifting: What is "Real" and what is not? *International Tax and Public Finance, 23*(4), 640–669.

Harju, J., Matikka, T., & Rauhanen, T. (2017). *The effects of size-based regulation on small firms: Evidence from VAT threshold*. Working paper, VATT Institute for Economic Research.
Johnson, M. (2017). *Tutkimusaineistojen käyttö Fiona-etäkäyttöjärjestelmässä (Use of research data in fiona remote access system)*. Seminar presentation 29.5.2017, Statistics Finland.
Kleven, H., Knudsen, M., Kreiner, C., Pedersen, S., & Saez, E. (2011). Unwilling or unable to cheat? Evidence from a tax audit experiment in denmark. *Econometrica, 79,* 651–692.
Kosonen, T. (2015). More and cheaper haircuts after VAT cut? On the efficiency and incidence of service sector consumption taxes. *Journal of Public Economics, 131,* 87–100.
Kosonen, T., & Ropponen, O. (2015). The role of information in tax compliance: Evidence from a natural field experiment. *Economics Letters, 129*(C), 18–21.
Kotakorpi, K., & Laamanen, J.-P. (2016). *Prefilled income tax returns and tax compliance: Evidence from a natural experiment*. University of Tampere, Economics Working Papers, No. 104.
Levitt, S. D., & List, J. A. (2011). Was there really a hawthorne effect at the hawthorne plant? An analysis of the original illumination experiments. *American Economic Journal: Applied Economics, 3,* 224–238.
List, J. A. (2011). Why economists should conduct field experiments and 14 tips for pulling one off. *Journal of Economic Perspectives, 25*(3), 3–16.
Liu, L., Lockwood, B., & Almunia, M. (2017). *VAT notches, voluntary registration, and bunching: Theory and UK evidence*. Working paper, University of Warwick.
Matikka, T. (2018). The elasticity of taxable income: Evidence from changes in municipal income tax rates in Finland. *Scandinavian Journal of Economics, 120*, 943–973.
Ministry of Finance. (2012). *Tilastolain uudistamistyöryhmän ehdotus (Proposal of the working group on the Revision of the Statistics Act)*. Ministry of Finance Publications, July 2012.
Slemrod, J. (2016). Caveats to the research use of tax-return administrative data. *National Tax Journal, 69*(4), 1003–1020.

Affiliations

Miguel Almunia[1,2,3,4] · Jarkko Harju[5,6] · Kaisa Kotakorpi[5,6,7] · Janne Tukiainen[5,8] · Jouko Verho[5]

Miguel Almunia
miguel.almunia@cunef.edu

Jarkko Harju
jarkko.harju@vatt.fi

Janne Tukiainen
janne.tukiainen@vatt.fi

Jouko Verho
Jouko.verho@vatt.fi

1 Colegio Universitario de Estudios Financieros (CUNEF), Madrid, Spain

2 Centre for Economic Policy Research (CEPR), London, UK

3 Centre for Competitive Advantage in the Global Economy (CAGE), Coventry, UK

4 Oxford University Centre for Business Taxation, Oxford, UK

5 VATT Institute for Economic Research, Helsinki, Finland

6 CESifo, Munich, Germany

7 University of Turku, Turku, Finland

8 Department of Government, London School of Economics and Political Science, London, UK