Tax after coronavirus Inquiry – Submission: HMRC's Datalab
28 August 2020

Executive Summary

- The HMRC Datalab enables academic researchers to undertake cutting-edge analysis in the public interest using UK tax data.
- The Datalab is perfectly positioned to advance the government's 'big data' agenda, but it is currently critically under-resourced. This limits its utility for academic research and constrains the scope for productive collaboration between HMRC and academic researchers.
- Additional resourcing for the Datalab is a highly cost-effective means of enhancing the government's capabilities in data analysis, by leveraging input from the UK's world-leading academic research institutions.

About us

Arun Advani is Assistant Professor of Economics and Impact Director of the CAGE Research Centre at the University of Warwick. He is also a Research Fellow at the Institute for Fiscal Studies, and a Visiting Fellow at the LSE International Inequalities Institute. He studies issues of tax compliance and tax design, with a particular focus on those with high incomes or wealth.

Andy Summers is an Associate Professor of Law at the London School of Economics and an Associate of the International Inequalities Institute at LSE. His teaching and research focuses on tax law and policy, particularly the taxation of wealth. His work also investigates the measurement of inequality using tax data.

Our research

We are collaborating on a series of academic research projects relevant to this review. Most of our work has taken place in HMRC's Datalab facility, giving us access to anonymised data from the personal tax records of every UK resident in recent years. Having done a significant amount of exploration of these data, we are in a unique position to submit evidence on these issues.

A key focus of our work has been capital gains, including the characteristics of those who earn the highest taxable gains, the tax rates they typically pay, and the impact of this on inequality. Our work documenting empirical evidence on the sources and distribution of taxable capital gains has been made public in a series of papers and briefings, to which our submission below refers.

About the Datalab

1. The HMRC Datalab was established in 2012 to enable accredited academic researchers to access de-identified tax data for projects approved by HMRC that benefits both the department and wider public. Its purpose is to facilitate cutting-edge independent analysis of tax policy by providing access to information that is not available in the public domain, via a secure setting that maintains strict protection of taxpayer confidentiality. The facility is physically housed within HMRC's Canary Wharf offices and is staffed by a small team from HMRC's Knowledge, Analysis and Intelligence division ('KAI').

2. Before a Datalab project can be approved, it must be shown that the research will assist HMRC's functions of tax design, collection and administration. A full list of approved projects is available via the HMRC Datalab webpage. Academic researchers work closely with HMRC at each stage of the research process, but the results are published independently and are subject to academic peer-review. The Datalab accordingly provides a crucial channel for transparent tax policy analysis conducted in the public interest; it also represents excellent value-for-money for government since the time of the academic researchers is provided for free.

Data security

3. Only trained and accredited researchers, whose project proposal has been approved by HMRC in advance, can access the Datalab. The Datalab follows HMRC's strict data protection policies and ensures safe working practices to safeguard taxpayer confidentiality. In particular, the Datalab follows 'five safes' framework which is used by other secure centers such as the UK Data Service:

   • **Safe projects**: proposals must make appropriate and ethical use of the data, must deliver clear benefits, and must be published to enable scrutiny and further research.

   • **Safe people**: researchers must have relevant technical skills, complete training courses and assessments, and sign an agreement to protect data confidentiality.

   • **Safe setting**: while accessing data, researchers have no access to email, the internet, printers, or any other way of taking out protected data. There are also additional security controls including CCTV as part of the IT infrastructure.

   • **Safe data**: before being made available to researchers, data are de-identified by removing disclosive information such as names, addresses and any other details that would directly identify the data subjects.

   • **Safe outputs**: raw data cannot leave the Datalab environment; any outputs (including tables and charts) are checked for statistical disclosure by Datalab staff to ensure they cannot be used to identify information about individual data subjects.

Datalab research informs government policy

4. Recent Datalab-based research has informed tax policymaking and resulted in important reforms introduced by the government at the March 2020 Budget:

   a. **Advani (2017):** Who does and doesn't pay taxes?
   This research identified key characteristics of individuals who are most likely to underpay taxes, and showed how their behaviour changes after a tax audit.
   Advani, Elming and Shaw (2019): The Dynamic Effects of Tax Audits
   This research showed that tax audits are highly cost-effective: HMRC not only receives repayment of the taxes owed in the audited year, but also 1.5-1.8 times as much additional revenue over subsequent years as audited individuals continue to declare more income.
   This work recommended an increase HMRC's budget to facilitate more audits, which was implemented in the March 2020 Budget.

   b. **Miller, Pope and Smith (2019):** Intertemporal income shifting and the taxation of owner-managed businesses
   This research by the Institute for Fiscal Studies showed that company owner-managers respond to differential tax rates on income and capital gains by adjusting how and whether they take money out of their company, rather than by increasing investment. The research documented economic inefficiencies created by Entrepreneurs' Relief for Capital Gains Tax and improved understanding of behaviours within the UK's small business population.
   The lifetime limit to Entrepreneurs' Relief was reduced from £10 million to £1 million in the March 2020 Budget.

5. Current work by Datalab researchers will answer important policy questions relevant to the government's key priorities:

   a. **Advani, Koenig, Pessina & Summers: Tax and International Mobility**
   This research investigates the characteristics of high-income migrants to the UK, the effects of the tax system on their decisions to move to and remain in the UK, and the impact of that migration on the UK economy. The researchers have recently applied to link HMRC tax data with Home Office visa data to evaluate the effectiveness of the UK's visa system.
   In January the Migration Advisory Committee recommended better linking of government data 'to allow a better understanding of the employment outcomes of migrants' and 'for the evaluation of all visas'.

   b. **Advani & Summers (2020a, 2020b):** Capital Gains and UK Inequality & How much tax do the rich really pay? New evidence from tax microdata in the UK
   Initial findings from this research investigated the characteristics of individuals who reported large capital gains and documented the impact of capital gains on effective average tax rates, taking into account all sources of income and gains. This ongoing research will estimate behavioural responses to Capital Gains Tax to evaluate its effectiveness in incentivising entrepreneurship and investment.
   In July the Chancellor asked the Office of Tax Simplification to carry out a review of Capital Gains Tax, with specific reference to how gains are taxed compared to other types of income.

The Datalab is crucial to the government's evaluative agenda

The importance of policy evaluation

6. [Cabinet Office](): In his Ditchley Lecture, Michael Gove called for improvements to the way in which government evaluates the effectiveness and value-for-money of public policies: "Government needs to evaluate data more rigorously and that means opening up data so others can judge the effectiveness of programmes as well" – *June 2020*

7. [National Audit Office](): A report on management of tax expenditures noted that the impact of such measures is 'not guaranteed' and recommended that both HMRC and Treasury should "formally establish their accountabilities for tax expenditures and enable greater transparency" – *February 2020*

8. [Public Accounts Committee](): Following the NAO report, the PAC launched an inquiry into how government evaluates tax reliefs, questioning officials on "management of tax reliefs, the number of reliefs and the Government's understanding of whether they represent value for money" – *June 2020*

9. [The Treasury Committee](): As part of its 'Tax after coronavirus' enquiry, the Committee will seek evidence on "the role of tax reliefs in rebuilding the economy" – *July 2020*

10. [Office for Budget Responsibility](): The 2019 Fiscal Risks Report highlighted that government does not know the overall cost of tax reliefs & expenditures, but that the cost of identified expenditures is almost 8% of GDP: "It is not clear that the Government gives tax reliefs and expenditures adequate scrutiny to control their cost" – *July 2019*

The role of the Datalab

11. Academic researchers working within HMRC Datalab use cutting-edge computational and statistical methods for the analysis of big data and have world-leading expertise in tax policy. They also have the freedom and incentive-structure to pursue bigger-picture and longer-term research projects than are typically feasible for HMRC analysts: a typical Datalab project may last 2-3 years. Consequently, Datalab projects can supplement and enhance HMRC's internal analytic capacity, for example by creating new links between data sources and applying new econometric tools to investigate behavioural responses to the tax system.

Datalab value-for-money

12. The time of academic researchers working in HMRC Datalab is wholly funded by their universities or from external research grants. Consequently, these skills and expertise are provided by the researchers at zero direct cost to government. The researchers routinely share results of their analysis with HMRC KAI and policy teams as well as HM Treasury. After undergoing appropriate checks, all research findings produced by Datalab researchers are made publicly available and can be used to inform government policy. In accordance with UK Research and Innovation's framework for Knowledge Exchange and Impact, researchers have strong incentives to communicate their findings to policymakers and the public.

Lack of resources

13. In its recent report 'Overcoming the Barriers to Tax Reform', the Institute for Government made the following observations about HMRC Datalab:

    "The process once [researchers] have been granted access to the Datalab – including getting access to the right data – is often slow, which means that the research is produced in a less timely manner. The main reasons for this are that some of the documentation is poor … and that the Datalab is under-resourced."

14. The under-resourcing of the Datalab manifests in two main ways:
    a. **The Datalab is under-staffed**

    The Datalab staff team comprises the Datalab Manager, two Datalab Assistants, and two IT staff. Between them they support over thirty active research projects. For each project, this involves:

    - assessing the project proposal;
    - liaising with data owners regarding new data requests;
    - sourcing and compiling documentation on datasets;
    - validating and de-identifying datasets for input to Datalab;
    - investigating data queries from researchers;
    - undertaking data-matching where not possible using pseudonymised identifiers
    - checking outputs for statistical disclosure control;
    - setting up and attending meetings with KAI and policy teams

    In addition, Datalab staff maintain the IT infrastructure, manage bookings and other queries from researchers, run training courses for new researchers and implement data protection policies, amongst other roles.

    The Datalab staff are highly skilled and it is evident that they work very hard (for example, sending emails outside regular office hours) but there are not enough hours for them to support researchers effectively. This results in delays when processing data requests and queries, which hold up the progress of research projects.

    b. **KAI are not adequately resourced to support Datalab requests**

    Datalab projects crucially rely on regular engagement with data owners and analysts within HMRC's Knowledge Analysis and Intelligence division (KAI). For example:

    - providing metadata to inform new data requests;
    - assessing data requests;
    - answering queries about data construction;
    - providing feedback on data analysis;
    - setting up and attending meetings with researchers

KAI staff are in general supportive of Datalab projects but it evident that requests for assistance from Datalab researchers must be delivered on top of routine workloads. For example, delays in answering data requests and queries are often due to concurrent pressures in delivering official statistical publications and analysis for ministers.

Delays in responding to new data requests and queries about data validation frequently result in 'roadblocks' that can halt progress on a project entirely for several weeks at a time, or alternatively require researchers to adopt inefficient workarounds.

Recommendations

The Institute for Government made the following recommendation in its 'Overcoming the Barriers to Tax Reform' report:

> *An increase in Datalab resources – which are a modest part of the HMRC budget – would speed up the research process, allowing for more timely outputs and higher-quality analysis. This, in turn, would lead to more researchers using the data, answering policy-relevant questions and building up a richer academic tax community more generally.*

As researchers currently working on projects in HMRC Datalab, we strongly agree with this assessment. Although at the margins we feel there could be room for improved working practices, by far the main obstacle to undertaking research within the Datalab is that HMRC staff are overstretched and unable to give sufficient time to our projects.

Consequently, we make two concrete recommendations on resourcing:

(1) **HMRC should employ at least two additional staff working full-time at the Datalab** (this is in addition to replacement of one recent departure). At least one of these additional staff members should be devoted to processing requests for new datasets and data matching and assisting with data validation and queries; this role requires a high level of skill in data analysis.

(2) **HMRC should directly allocate KAI staff time to supporting Datalab projects.** At minimum, the staff member within KAI who is assigned as project 'lead' should have time allocated to support the project, in particular acting as a proactive go-between the Datalab researchers and other parts of HMRC. Other staff in KAI should also be appropriately resourced to respond promptly to queries and new data requests.

We believe that these two steps – which could be achieved with very minimal additional spend in the context of HMRC's total budget for data analysis – would radically transform the capacity of academic researchers to deliver impactful and policy-relevant projects in the public interest.

Better resourcing of the Datalab would also make this facility more attractive to academic researchers in future, thereby enhancing the capacity of government to leverage external expertise to advance its agenda to improve policy evaluation using big data.

References

Advani, A. (2017): "Who does and doesn't pay taxes?," IFS Briefing Note BN218.

Advani, A., W. Elming and J. Shaw (2017): "The Dynamic Effects of Tax Audits," CAGE Working Paper no. 414.

Advani, A. and A. Summers (2020a): "Capital Gains and UK Inequality: New evidence from tax microdata," CAGE Policy Briefing no. 19.

Advani, A. and A. Summers (2020b): "How much tax do the rich really pay? New evidence from tax microdata in the UK," CAGE Policy Briefing no. 27.

Gove, M. (2020): "The privilege of public service," Ditchley Annual Lecture, 27 June 2020

Migration Advisory Committee (2020): "A Points-Based System and Salary Thresholds for Immigration," report published January 2020.

Miller, H., T. Pope and K. Smith (2019): "Intertemporal income shifting and the taxation of owner-managed businesses," IFS Working Paper W19/25.

National Audit Office (2020): "The management of tax expenditures," report published February 14 2020.

Office for Budget Responsibility (2019): "Fiscal Risks Report", July 2019.

Office of Tax Simplification (2020): "Chancellor requests OTS review of Capital Gains Tax," correspondence published 14 July 2020.

Public Accounts Committee (2020): "Management of tax reliefs: Inquiry," website accessed 29 July 2020.

The Treasury Committee (2020): "Tax after coronavirus: Inquiry," website accessed 29 July 2020.

Tetlow, G., et al (2020): "Overcoming the Barriers to Tax Reform", report published by Institute for Government, April 2020