Turing-HSBC-ONS Economic Data Science Awards 2018

Guidance for Applicants

The Economic Data Science research programme at The Alan Turing Institute (the Turing) invites proposals for research in economics and data science to improve understanding of how the economy works or contributes tools that can be used for that purpose. Proposals may cost up to £150,000 and last 6-24 months. Part of the aim of the call is to generate results quickly, so proposals for projects lasting longer than 12 months should include a description of what can be produced within a 6-9-month time frame.

This call for proposals aims to catalyse research in economics and data science to advance the Economic Data Science research programme. It is funded through the five-year strategic partnership between The Alan Turing Institute and HSBC. For the purpose of this call, The Alan Turing Institute is also collaborating with the Data Science Campus of the Office for National Statistics (ONS) to make ONS data and resources available to researchers, subject to access and appropriate use requirements.

Contents

Background

The aim of the economic data science research programme at the Turing is to make major advances understanding how the economy works in its complexity and interconnectedness, and to use that insight to address economic challenges of national and international importance. This will be possible through using econometric and data science techniques to harness the scale and scope of economic data now available to analyse the economy with greater accuracy and granularity.

ONS is the UK's largest independent producer of official statistics and the recognised national statistical institute of the UK. It is responsible for collecting data and publishing statistics related to the economy, population and society at national, regional and local levels. ONS uses data from surveys, the census, administrative and commercial sources for statistical and research purposes only. ONS produces official and research statistics to inform important decisions that affect us all and for the wider benefit of society and the economy. The Data Science Campus at the Office for National Statistics is the UK Government hub for data science. It brings together data, digital expertise and leadership. Its research goals are to explore new data sources and perform cutting-edge research using a new generation of tools and technology, to support the goals of ONS.

All applications will be shared with an HSBC Steering Committee made up of stakeholders from HSBC with interests and expertise in economic data science and data analytics. This Steering Committee is not part of the review panel that will assess applications, but will instead look for projects or research teams that HSBC might be interested in collaborating with in the future.

Scope

This call aims to catalyse research in economics and data science under the three themes listed below to improve understanding of how the economy works or to contribute tools that can be used for that purpose.

Proposals are invited from research teams for small (≤£50,000) or medium scale (£50,000-£150,000) projects that lie at the boundary between economics and data science and combine expertise from those two fields. Proposals must be innovative in nature, fit under one or of the three research themes outlined below, have potential for substantial impact and present good value for money. Note that proposals will not be assessed under a single theme so may address issues under multiple themes. Projects funded through this call may be candidates for future funding at a larger scale but this will not affect their assessment at this funding round.

The three research themes under which proposals are invited are as follows.

Theme 1: Economic activity over the business cycle

New data, whether from the government or private sector, opens up the possibility of measuring economic activity – output, employment, wages, spending, regional or international trade etc – faster, more precisely and at a more disaggregated level than has previously been possible. This will mean that individual, business and government decisions can be based on a much more complete and up-to-date picture of what is going on in the economy. It will also allow a better understanding of economic networks and propagation mechanisms and the drivers of productivity differences, and enable researchers to address issues relating to household consumption/saving decisions, firm financing over the business cycle and the economic impact of Brexit.

Theme 2: The changing nature of work

Recent decades have seen important changes in the labour market across four dimensions. First, there have been changes in how long is spent at work. This includes the growth in part-time work as well as changes in participation across life for men and for women. Second, there have been changes in how work is organised. The size and structure of employers has changed and outsourcing has grown in importance. Alongside this, a growing number of individuals are on zero-hours contracts or are self-employed (some in the so-called “gig economy”) or in new forms of informal employment. The third change is in what work involves. This includes both the types of occupation available and the tasks performed while at work. And fourth, there have been changes in the return to work. Wage growth since the financial crisis has been subdued. Those lower down the distribution have tended to fare less badly than those higher up, but there have also been important developments among the very rich – with implications for inequality. There has also been substantial variation across occupations and skill levels. These changes raise challenges for measurement, and create a need to understand what is driving the changes, analysing the implications for welfare and establishing what role policy has to play.

Theme 3: Machine learning for economic data

Modern econometrics emphasizes causal inference and counterfactual analysis to assess the impact of policy interventions, but typical models are built for quantitative array data of limited size. Data of a different form is increasingly available: for example, observations now regularly appear with hundreds or even thousands of recorded variables; and text, satellite images, and web search profiles have non-standard data structures but contain vast amounts of information. Machine learning methods have developed in the last decade in response to the growth of such data, although these usually prioritize accurate out-of-sample prediction with computationally efficient algorithms. The goal of this theme is to develop an understanding of the role of machine learning techniques in empirical economics. The first component is methodological. For example, when and how do economists need to adapt machine learning algorithms to address their empirical challenges? What are the statistical properties of estimates that algorithms produce? The second component is to apply machine learning methods within specific fields to answer economic questions that traditional data cannot.

Data

Existing datasets

Researchers are free to use whatever existing datasets they choose but they must have agreement to access to those datasets when they submit their proposal (i.e. awards will not be made if data access is not already guaranteed). Proposals should not include data collection.

ONS datasets

As part of this call, ONS are offering inward secondments to work on ONS datasets. These datasets are in addition to the rich microdata available from ONS's surveys via the Virtual Microdata Laboratory, and may be particularly relevant for addressing questions under the three research themes. Any potential secondee will require the correct visa and work checks, and will need to be security cleared at either Baseline Personnel Security Standard (BPSS) or Security Check (SC) level, depending on the data to be accessed and the access environment. Secondments will allow remote access to the data (i.e. researchers will not need to be at ONS full time to use the datasets). Data held by ONS may only be used for approved purposes, i.e. for statistical research and functions aligned with ONS's strategic objectives. No data in which individuals or individual companies can be identified may be removed from ONS' secure environments, or shared with individuals not employed by ONS.

Some of the datasets that ONS hold are as follows.

VAT turnover and expenditure

Turnover and expenditure data are collected by HMRC through VAT returns. This information is provided by businesses on a monthly, quarterly or annual basis dependent on their VAT reporting pattern. The information is a census of all businesses which are registered to pay VAT. The current VAT threshold for businesses for the financial year ending 2016 is a turnover of more than £82,000 per year. Turnover data are available from 2007, and expenditure from 2012 onwards, both to the current period. Data are from HMRC's daily feed, and are sent to ONS monthly. The data are mixed frequency, having either monthly, quarterly or annual reference. Data are reported at the VAT unit level, and ONS apportion this to enterprise level. The key variables are turnover, expenditure, VAT unit reference number, reference period and stagger (flag for monthly, quarterly or annual).

Debt securities, commercial and consumer lending

ONS hold a range of financial data from commercial sources, including:

- Intelligence and data relating to the ownership of debt securities and equity (e.g. shares, bonds, gilts) by and of UK companies. Data will be available from mid-February 2018, and will cover the last 20 years up until the present period. Data are reported quarterly, and are at the firm level or transaction level, with 100s of variables.
- Information on borrowing by UK households and companies - specifically who is borrowing from whom and for what purpose. Data are available from 2015 Q1 to 2017 Q3, and will be extended to cover periods up to 2018. Data are reported quarterly, and are at firm level, or (consumer borrowing) low level geographies. rich information about the value and type of lending is included.

Shipping movements

ONS holds Automatic Identification System (AIS) and Consolidated European Reporting System (CERS) data on shipping movements from the Marine and Coastal Agency. These data show the GPS location of ships, reported with a frequency between 2 seconds and 3 minutes, depending on the speed of the ship. All ships over 300 tonnes are legally required to have AIS on board, and many ships below this weight limit also have AIS installed. The dataset covers UK waters, for the period 1 August 2016 to 31 July 2017. The CERS data also captures UK ship arrival and departure times, dangerous and polluting goods notifications, and information about the vessel, voyage and ports.

ONS would also be happy to discuss the availability of other datasets they hold.

Researchers can only use ONS data for the purposes permitted by the legal framework governing ONS, and researchers would also need to agree to disseminate results across ONS and more widely.

Researchers considering using one or more of these ONS datasets should discuss this with Louisa Nolan (louisa.nolan@ons.gov.uk) at ONS prior to submitting an application to ensure that suitable access arrangements can be made.

A statement of commitment from ONS should be included with the application.

Collaborations

HSBC

All applications have the option of being shared with an HSBC Steering Committee made up of stakeholders from HSBC with interests and expertise in economic data science and/or data analytics. This Steering Committee is not part of the review panel that will assess applications, but it will look for research teams that HSBC might be interested in collaborating with in the future.

Applicants are also invited to propose how they could run a second stage of their project (within the project's lifetime) in collaboration with HSBC, should they wish. Please note that this is not is mandatory part of an application; any project proposal that relies on collaboration with HSBC to be successful will not be considered by the Panel.

ONS expertise

In addition to providing data, ONS are offering the opportunity to partner with ONS staff. This may be particularly useful where researchers are seeking to use ONS datasets and could therefore benefit from dataset-specific expertise. Such collaboration would be provided free of charge but would be subject to capacity constraints and would need to be negotiated on a case-by-case basis in advance of the application being submitted. A statement of commitment from ONS should be included with the application. Proposing to partner with ONS researchers is no guarantee of funding. Please contact Louisa Nolan (louisa.nolan@ons.gov.uk) at ONS to discuss this further.

Turing Research Engineer expertise

There is also an opportunity to work with the Research Engineering department at The Alan Turing Institute (please see: https://www.turing.ac.uk/research-engineering/).

The members of The Alan Turing Institute's Research Engineering Group (informally known as 'Hut23') are experienced data science researchers who are committed to professional delivery of impactful research, rather than to personal research interests. They support the mission of the institute by connecting pure research to applications. Hut23 are a permanent pool of staff with postdoctoral level skills in data science, software engineering, and modelling. They are available to provide collaborative support to research proposals as a specialist resource, charging a day rate.

For projects that need support and the skills and background to create readable, reliable and efficient compute- and data-intensive software, embodying in code sophisticated mathematical and statistical concepts, please contact Martin O'Reilly - moreilly@turing.ac.uk prior to submitting an application.

For projects that require support and expertise in computational statistics, inference, and machine learning, and also in mathematical and computational modelling of complex systems, knowledge representation, and operations research, please contact James Geddes – jgeddes@turing.ac.uk prior to submitting an application.

Leveraging other resources

Applicants are invited to draw on other resources they might have access to such as already-funded PhD/PDRA posts aligned with this work, or already-funded PI time. Applications should include a "Leverage Statement" to confirm this resource commitment.

Eligibility

All proposals must be led by one senior academic researcher with a strong track record of grant support and research outputs. Projects must start before 01 October 2018.

Lead applicants must be either Turing Fellows, Turing Research Fellows, Turing Visiting Researchers or situated at one of the Turing's university partners (please note that funding is subject to the University-Turing partnership agreement being in place by 30 June 2018). Additional researchers who are not directly involved with the Turing or its university partners may participate but may not lead. These additional researchers should be based at a Higher Education Institution (HEI) or Independent Research Organisation (IRO).

The Turing is actively committed to promoting equality and diversity.

Funding available

Applicants may submit proposals for up to £150,000 costs for projects of 6-24 months in duration. Part of the aim of the call is to generate results quickly, so proposals for projects lasting longer than 12 months should include a description of what can be produced within a 6-9-month time frame.

Awards may include:

- Staff salaries, whether directly allocated or directly incurred (e.g. existing academic staff, research assistants or other research staff, technicians and other support staff). *Please note that PhD studentships cannot be supported through this call.*
- Research consumables directly attributable to the project
- Research equipment essential for the project
- Travel and subsistence, and other meeting costs where relevant
- Dissemination costs such as conference costs

Awards may not include indirect costs such as overheads.

How to apply

All applications must be made online via the Turing's Flexigrant portal: https://ati.flexigrant.com.

Applicants will be required to complete the following sections on the Flexigrant portal directly or as a PDF upload to the portal, as indicated:

- Project title and summary
- Lead applicant contact details
- Scientific abstract and lay summary (max. 200 words each)
- Research proposal (max. 4 pages, plus an additional page for references if needed, as a single PDF file) that includes:
    - Case for support (including background, aims and objectives, tools and methods, relevance and beneficiaries, dissemination and impact)
    - Workplan (including timeline, milestones and deliverables)
- Confirmation of access to data
- A description of how the applicant(s) could run a second stage of the project (within the project's lifetime) in collaboration with HSBC, should they wish (max. 400 words)
- Financial details:
    - Directly Allocated costs (e.g. staff salaries)
    - Directly Incurred costs (e.g. consumables, equipment)
    - Other costs (e.g. travel & subsistence)
    - Other resources being leveraged
    - Ineligible costs
- Leverage statement (max. 400 words)
- Impact statement (max. 400 words)
- Ethics statement (max. 200 words) and status of Ethics Approval
- CVs and publication list
    - CVs as a single pdf file: Lead applicant (max. 2 pages) Other investigators (max. 1 page each)
    - Combined publication list (max. 2 pages, pdf)
- Letters of support
    - From applicants' employer to show support to cover indirect costs and any additional provision of resources as stated in the application (max. 1 page .pdf)
    - From ONS if the proposal seeks to make use of ONS data or resources (max. 1 page .pdf)

Applications must be submitted no later than **09:00 BST, Monday 09 April 2018**.

Only applications submitted through the Turing's Flexigrant system will be accepted for processing. The application submitted through Flexigrant will be taken to be the final version, and will be the version used for assessment.

Assessment

All applications will be assessed by a specially-constituted panel, against the following criteria.

Research excellence

What is the novelty and added value of the proposed work in relation to current knowledge or ongoing work by the applicants or elsewhere? Does it combine techniques from economics and data science? How likely is the proposed research to be successful? Is the necessary data access guaranteed?

Fit to programme themes

How closely does the proposed research align to one or more of the programme's three research themes?

Potential for impact

Will the proposed research improve understanding of how the economy works, or contribute tools that can be used for that purpose? How will insights from the research will address economic challenges of national and international importance? What plans been made for engagement within and outside academia? What results will be generated within a 6-9-month time frame?

Value for money

Overall value for money: Are the proposed research outputs and associated potential for impact good value for money given the costs? How well does the proposal leverage other sources of funding? Assessment will be made relative to the costs requested (under £50,000 or £50,000-£150,000).

Individual elements of resourcing the proposal: Are all the specific cost components essential and sufficient for the proper conduct and exploitation of the research proposed?

What we will do with your information

In accordance with the Data Protection Act 1998, the personal information that you provide within the application will specifically be used for the purpose of administering this call. The information will be viewed by selection panel members and the Turing, ONS and HSBC staff, including an HSBC Steering Group to identify shared interests and explore possible collaborations. Your information will not be used for any other purpose without your specific consent.

Application and award timetable

Opening date for applications	09:00, Thursday 08 February 2018
Closing date for applications	09:00, Monday 09 April 2018
Review panel meeting	w/c Monday 28 May 2018
Notification of awards	Friday 29 June 2018
Ratification of awards by the Turing-HSBC Programme Board	w/c Monday 16 July 2018
Project start date	before Monday 01 October 2018

Contact

For questions regarding the application process or other elements of the call, please contact Dr Mahlet Zimeta (mzimeta@turing.ac.uk).