A Primer for Doing Tax Research with Administrative Data

Dario Tortarolo
University of Nottingham

December 13, 2022

Tax Research Workshop

- **Today:** empirical tax research + my (Berkeley's) view on what/how we do it + some applications + best practices
- **Learning outcomes:** learn about the process of how faculty do research + hear about the work itself + get you inspired for your own research
- What this class is not about: teaching you econometric methods or coding; it's not a public finance crash course (PF, henceforth)

Outline

Overview of the field: the big picture, topics, methods

Ingredients of a tax project

- Research questions
- Tax variation
- Administrative data
- First stage
- Salience
- Main analysis
- Structural model

Administrative Data: the nuts & bolts

Outline

Overview of the field: the big picture, topics, methods

Ingredients of a tax project
- Research questions
- Tax variation
- Administrative data
- First stage
- Salience
- Main analysis
- Structural model

Administrative Data: the nuts & bolts

Big Picture: What is Public Economics?

Government is instrumental in most aspects of economic life:

- (1) Taxes (people and firms), (2) Regulation (min wages, zoning, environment), (3) Spending (public goods, infrastructure, defense, education), (4) Macro stabilization

Big Picture: What is Public Economics?

Government is instrumental in most aspects of economic life:

- (1) Taxes (people and firms), (2) Regulation (min wages, zoning, environment), (3) Spending (public goods, infrastructure, defense, education), (4) Macro stabilization

Public Economics (PE) studies the role of the government in the economy:

1) **How do government policies affect the economy?** (positive)
   E.g., the # of people working, distribution of income, etc.
2) **How should we optimally design policies to maximize welfare?** (normative)

Big Picture: What is Public Economics?

Government is instrumental in most aspects of economic life:

- (1) Taxes (people and firms), (2) Regulation (min wages, zoning, environment), (3) Spending (public goods, infrastructure, defense, education), (4) Macro stabilization

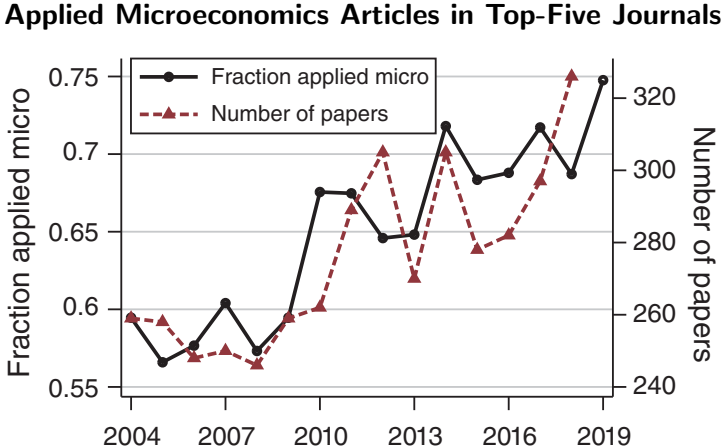
Public Economics (PE) studies the role of the government in the economy:

1) **How do government policies affect the economy?** (positive)
   E.g., the # of people working, distribution of income, etc.
2) **How should we optimally design policies to maximize welfare?** (normative)

Some motivations for studying PE:

- Contentious debate on govt appropriate role: injecting science has practical value
- Is the end point for many subfields. So, it's a natural combination
- At the frontier of data-driven approaches to answering key policy questions

Plus: Applied Micro is on the rise!



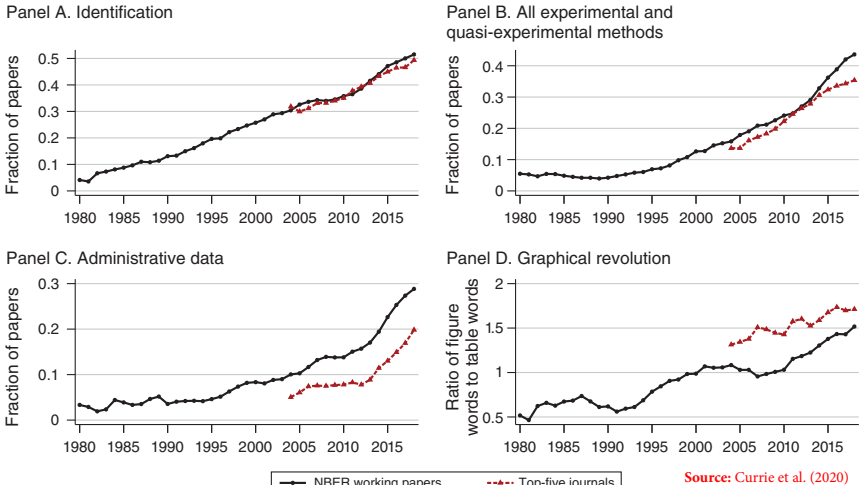
Source: Currie et al. (2020)

Overview of PE: How do government policies affect the economy?

Modern PE tightly integrates **theory** with **empirical evidence** to derive quantitative predictions about policy (**direct** and **unintended** effects) to inform and improve the debate

- **Quasi-experiments:** Research in PE exploits a variety of quasi-experimental research designs to identify parameters of interest (a "collage" approach)
- **Figures:** Emphasis on non-parametric graphical techniques rather than parametric regression models (the "graphical revolution")
- **Data:** Compelling implementation of quasi-experimental methods requires a lot of data ("Big Data")

PE is a big contributor of the "Credibility Revolution"



Source: Currie et al. (2020)

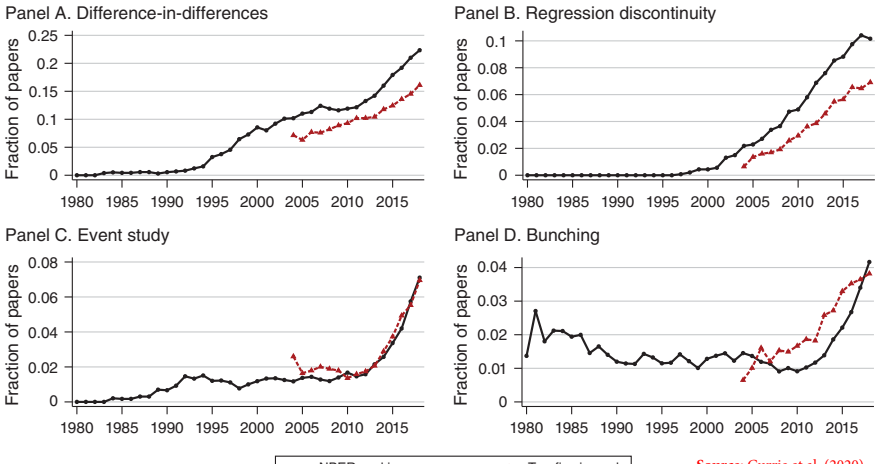
Some topics studied in graduate PF

1. Overview of the tax system (usually country-specific)
2. Optimal labor income taxes and transfers
3. Responses to income taxation: Labor supply, Taxable income, Migration, Innovation
4. Capital income taxation: canonical models + empirics
5. Wealth and property taxation
6. Corporate taxation
7. Consumption/commodity taxation (excise, sales, VAT)
8. Tax compliance and enforcement
9. Tax incidence
10. Social insurance

Methods used in PF

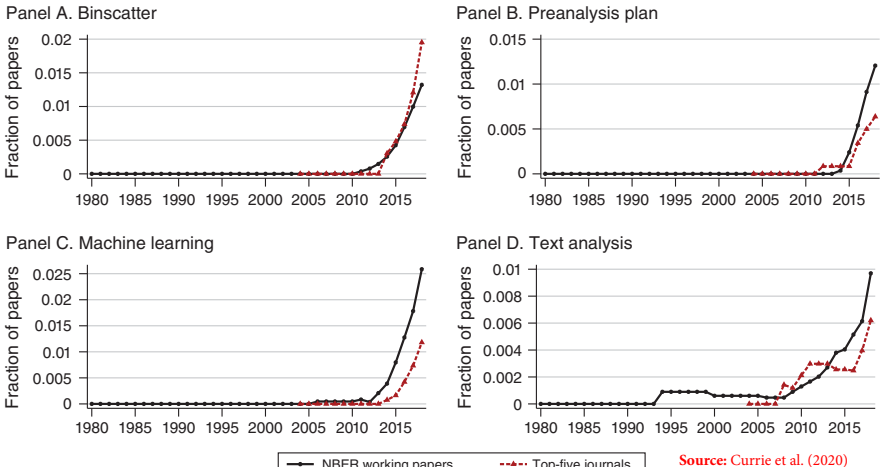
- Linear Regressions
- Instrumental Variables (IV)
- Difference-in-Differences (DiD)
- Event Studies
- Synthetic Control
- Regression Discontinuity Design (RDD)
- Randomized Experiments (RCT)
- Bunching to *kinks* and *notches*

The rise of quasi-experimental methods



Source: Currie et al. (2020)

What's new?



Source: Currie et al. (2020)

Outline

Ingredients of a tax project

1. **Research questions:** think BIG!
2. **Tax variation:** what does the law say? what changed?
3. **Administrative data:** the basics
4. **First stage:** is the variation large enough?
5. **Salience:** do people/businesses understand the variation?
6. **Main analysis:** graphical evidence first, then tables with coefficients
7. **Structural model:** for policy counterfactuals and welfare analysis

Outline

1. Research questions

- Be ambitious, think BIG!
- "But how do I know?" Talk to people (classmates, professors, etc)
- Some rules of thumb: Is this a hot topic? Are people/media talking about this? Would your findings change the way we think about XYZ?
- Regardless, choose a question you are excited about and willing to spend a significant amount of time researching/answering. Preferably, two-sided projects!
- Often-times we come across sharp policy changes that allow us to answer important questions. Stay alert! (and try to connect shocks with theory)

1. Research questions

A non-exhaustive list of some recent *broad* topics (e.g., no convincing evidence on a classic PF topic, issues discussed in newspapers)

- Inequality and intergenerational mobility (by diversity, gender, income)
- Environmental taxes
- Taxation of the digital economy
- International taxation (e.g., global minimum corporate tax of 15%) [link]
- Offshore tax evasion
- Sin taxes (alcohol, cigarettes, sugary drinks, gambling)
- Wealth taxation (inequality in general)
- Gender taxation (e.g., pink tax, tampon tax)

Outline

2. Tax variation

- Essentially, you need to find settings or shocks that differentially affect otherwise comparable people/firms (cross-sectional or longitudinal variation)
  - The more unprecedented/unique the better
- A good tax researcher must carefully read the **legislation**
  - Understand the context: how the tax works, the schedule, how people/businesses file the tax (or receive a subsidy)
  - Understand the nature: what changes and what doesn't, who is affected, is it permanent or temporary, is it anticipated or unanticipated, how is the policy funded, what was the official (and unofficial) purpose of the policy
- It's not boring, it's crucial! The key to figure out the **research design**
- Plus: if possible, always talk to tax accountants

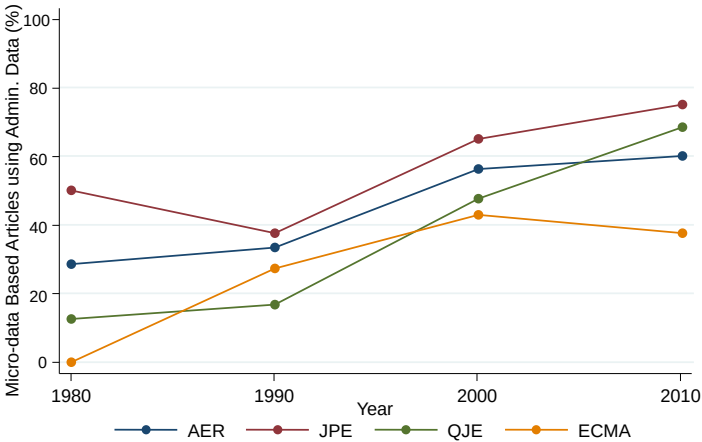
Outline

3. Administrative Data

- Compelling implementation of quasi-experimental methods requires a lot of data
- Myriad of socio-economic databases are created by govts for the administration of taxes and programs (so-called "admin data")
- Great resource for studying people and businesses responses to policy changes
- Ultimately, can be used for **tax** and **non-tax** research (e.g., intergenerational mobility)
- Let me postpone the "nuts and bolts" until the end

Rising use of admin data in leading economic journals

Source: Chetty (2012)



Note: "Administrative" datasets refer to any dataset that was collected without directly surveying individuals (e.g., scanner data, stock prices, school district records, social security records). Sample excludes studies whose primary data source is from developing countries.

Outline

4. First Stage

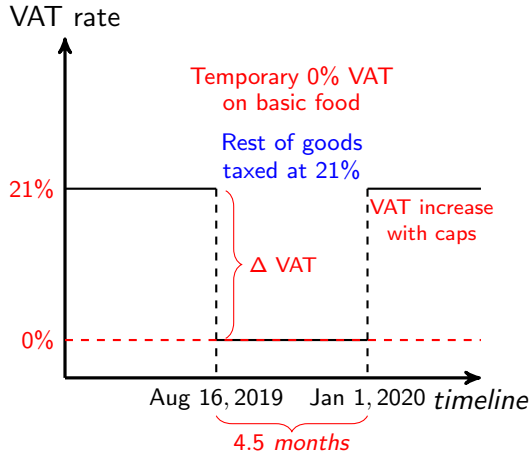
- Is your variation large enough to detect/estimate something meaningful?
  Some rules of thumb:
  - Does it change any macro time series? (e.g., tax revenue collected or number of taxpayers)
  - Does it target key players in the economy? (e.g., the ultra rich, the poor, women, etc.)

- This is the #1 thing you should check before moving forward

- Let's see an example from the tax incidence lecture

Example: VAT holiday

- VAT holiday in Argentina
- VAT cut: unanticipated, large, salient, and temporary
- VAT increase: anticipated, but capped price increase
- Great setting to study VAT incidence (pass-through to prices) and intertemporal consumption responses



Source: Benzarti, Garriga, Tortarolo (2021)

Example: VAT holiday

- Huge/unprecedented tax cut + capped tax increase
  - VAT cuts are usually economy-wide
  - VAT rate on some products goes from 21% to 0% (e.g., tea 0% but coffee 21%)

- The setting triggers interesting and policy-relevant questions
  1. Do supermarkets pass it on to prices or they pocket the money?
  2. Is there anything govt can do to mandate the pass-through? (e.g., real-time monitoring)
  3. Do consumers bring consumption forward to take advantage of cheaper goods?

- Novel data: high-frequency retail scanner data & a panel of shoppers

Outline

5. Salience

- Are people/businesses aware of the variation/policy change? Also...
  - How the tax/program works in practice
  - How the tax change affects their incentives

- This section is usually the least rigorous but it matters
  - It speaks well of your skills (a true researcher)
  - Can help you with the narrative and to convey the practical relevance
  - It makes the presentation/paper livelier

- Some strategies/resources are:
  - Media: newspapers, magazines, TV screenshots
  - Google Trends
  - Anecdotal evidence from practitioners, accountants, etc. (e.g., forums)
  - Other: pay slips, brochures, photos

Example: VAT holiday

Media coverage: a very salient VAT cut


Clarín

Viernes 16.8.2019

Spot
En la piel de Carlos Tevez

Soldano, la nueva apuesta de gol en Boca

Tema del día » Nuevas medidas económicas

Eliminan IVA en alimentos, congelan cuotas UVA y habría cambios de Gabinete

Para Alberto F. el dólar a $ 60 está bien y no hay razón para que aumente

El diálogo o un atiborro de tregua

En busca de la mística perdida

Nuevo ataque del hacker que filtró datos de la Federal

Lavagna suspendió la campaña y pidió que lo imiten


LA NACION

Viernes 16 de agosto de 2019

Macri eliminó el IVA de 14 alimentos y congeló las cuotas de los créditos UVA

Dólar: cuyó por primera vez desde las PASO

Presión militar china a Hong Kong

Permitirán estacionar junto a algunas bicisendas

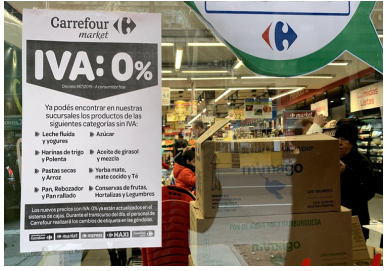
Filtraron archivos de la Federal sobre Chocobar

Ejecutaron a tres hombres en un búnker de drogas

Devaluación. Cómo se protegen los clubes

Example: VAT holiday

Mandatory banners and tags



Outline

6. Main analysis (or second stage)

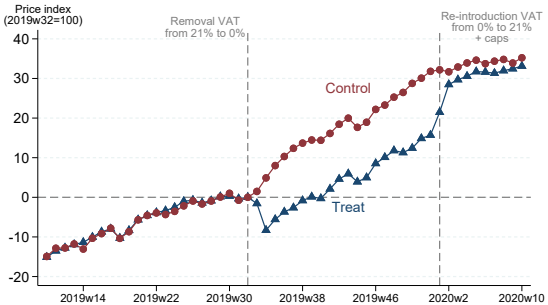
- Figures first, tables with numbers second
  - Both should be self-contained (i.e., VERY detailed footnotes)
  - Take the reader by the hand and walk them through each figure/table

- Effects should be visible in graphs!
  - Figures should be crystal clear, not too cluttered (it's almost an art)
  - The goal is to obtain 5 to 10 convincing figures
  - Bury the rest in an online appendix

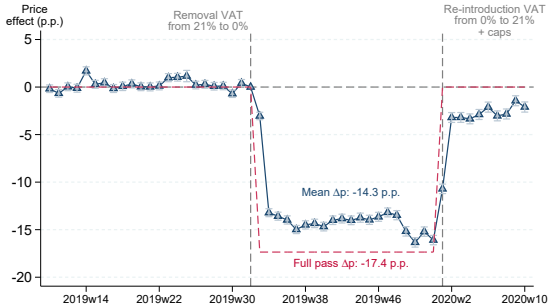
- Once you are happy with the figures, summarize the estimates, S.E., etc in tables
  - Tables should also be crystal clear, not too cluttered (max 7 columns)
  - **Trick:** often possible to compute elasticities (and s.e.) with a 2SLS regression. It scales the reduced-form by the first-stage change in taxes. Alternative: by hand + Delta Method

Example: VAT holiday

Price levels



Price pass-through



Outline

7. Structural model

Do I need a model? Maybe. Combining convincing reduced-form evidence with structural approaches is becoming common in successful PF papers

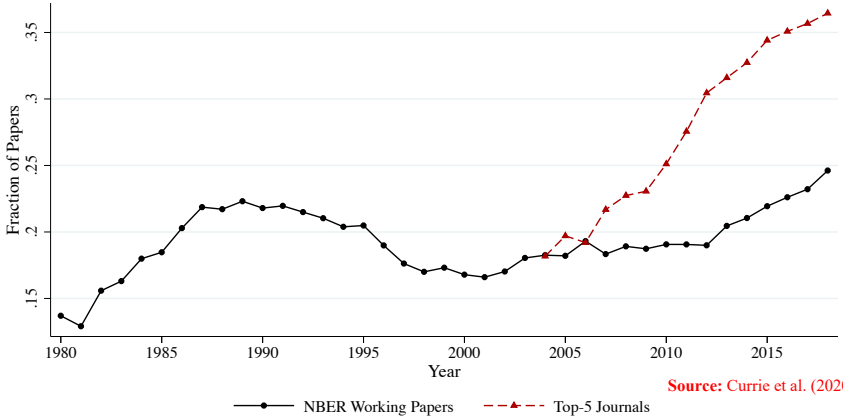
- **Structural approach:** specifies complete models of economic behavior and estimates/calibrates "deep" primitives [Blundell, 2017; Low & Meghir, 2017]
  - Useful to analyze policy counterfactuals and welfare, quantifying impacts on specific outcomes, short- and long-run effects, identify mechanisms

- **Sufficient-statistics approach:** formulas for welfare analysis that depend on reduced-form elasticities rather than structural primitives [Chetty, 2009; Kleven, 2021]

- Some recent nice PF work combining models with credible identification:
  Suarez Serrato & Zidar (2016); Brockmeyer et al. (2021); Gadenne et al. (2021)

Fraction of papers referring to **structural methods**

Currie et al. (2020)

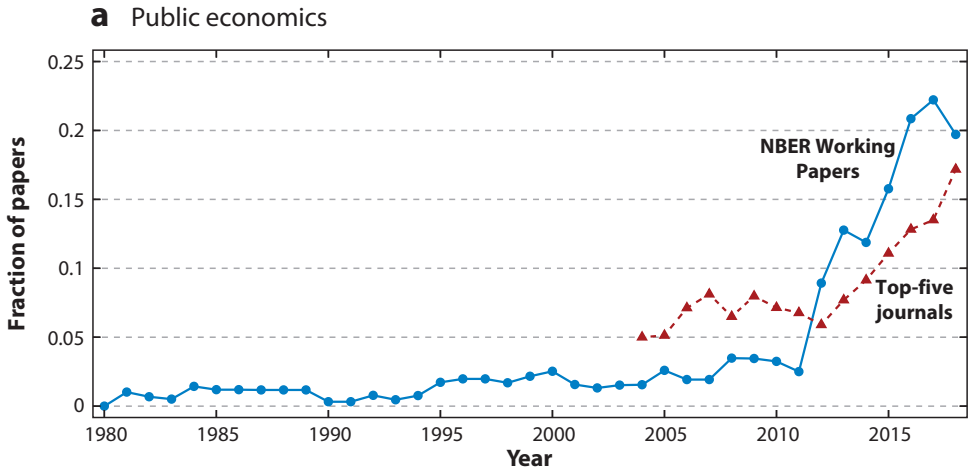


Source: Currie et al. (2020)

Notes: This figure shows the fraction of all papers in applied microeconomics referring to structural methods.

Fraction of papers referring to the **sufficient statistics** approach

Kleven (2021)



Outline

Overview of the field: the big picture, topics, methods

Ingredients of a tax project
- Research questions
- Tax variation
- Administrative data
- First stage
- Salience
- Main analysis
- Structural model

Administrative Data: the nuts & bolts

Administrative Data

- High quality data are key to empirical research in social sciences
- Recent availability of very large datasets has transformed research in applied microeconomics and the analysis of economic policy
- Admin databases are ubiquitous: central, subnational, and municipal agencies; ministries; tax administrations; central banks; credit bureaus; etc.
- **Key research priority:** develop **direct** and **secure** access to administrative data for research. Mixed experience across countries. Key for frontier empirical research

Administrative Data

What are the advantages?

- Excellent coverage
  - Longitudinal structure over long periods of time
  - Ability to match wide variety of data (tax, earnings, family, health, educational records)
- Higher quality information
  - Virtually no missing data or attrition
  - Less prone to measurement error
- Very large sample sizes
  - Can develop new non-parametric, quasi-experimental research designs
  - Explore heterogeneity in its own right

Fundamental issue: it's hard to merge data from different agencies (e.g., subnational taxes and SSA records). Mainly due to confidentiality and "data jealousy"

Admin Data: nuts & bolts

Types of tax data

- Tax registers (e.g., property cadasters)
- Self-assessment declarations (e.g., VAT, CIT, PIT tax returns)
- Third-party/information declarations (e.g., VAT annexes, withholding by employers, credit/debit card companies, and banks)
- Customs data (e.g., import and export records)
- Internal data (e.g., audits, tax amnesties)

Admin Data: nuts & bolts

Modes of accessing tax data

- Online (e.g., Mexico, Argentina, the UK)
  - Usually very limited and incomplete
- Handed over to PIs (de-identified data in a secure computer/room)
  - This is the first best
- Remote access (e.g., VPN access to a secure server)
  - This is the second best
- On site (e.g., in person, via RAs or staff)
  - This is one works, but it's time consuming and will require blind coding and a lot of travelling to make real progress

Admin Data: nuts & bolts

How to access tax data: a matching problem

- Diplomacy and entrepreneurial spirit is necessary to make the project a reality
- **Option 1:** Countries where you have connections are usually the starting point
  - Flag a key contact person (a high-level senior, a professor/friend connected to them)
  - Pitch your topic and methodological approach in plain language (see next slide)
  - If successful, discuss the data needs and associated logistics
  - Be ready to offer co-authorship
  - Some positions are political/temporary. Get close to permanent staff once your are in
- **Option 2:** Get involved in projects with seniors currently using admin data. Propose your own idea and request access to data
- **Option 3:** Apply online to access de-identified open data
  - Example 1: UK HMRC Datalab (check out their Data Catalogue)
  - Example 2: household & retail scanner data (price/quantity data from 54+ countries)

Admin Data: nuts & bolts

Pitching your project

- Pitch the project to your government counterpart in plain language
  - This is a key step. It might be your only shot, so be strategic
  - Usually a one pager: be clear, succinct, and emphasize the benefits to them (e.g., policy impact, more revenue, less evasion, etc.)
  - Don't be technical. It's not about you, it's about bewitching them
  - Make sure they actually have the data you need
  - Ultimately, the goal is to trigger interest and get a meeting
  - Failure is likely. It happened to me many times. Don't get discouraged. People are busy

- Extra tips
  - Ask your friends/classmates/advisor to read the proposal
  - Ask your contact what they currently worry about (policy needs)
  - Be ready to offer technical support, even if it's unrelated to your project
  - Play up your own experience in working with tax admin data and policy impact (if any)

Admin Data: nuts & bolts

Requesting data

Data request: a key step once you have green light

- Usually data are not organized and ready to be used.
  Agencies need to pull it out from servers or storage clouds
- Be as precise as possible: variables needed, frequency, period of time, identifiers anonymized consistently over time and across databases
- Limit the need for follow-up requests as much as possible: there's a delicate balance of asking as much as you can, without jeopardizing your relation
- Specify the mode of access (provide examples of how colleagues accessed data)

Admin Data: nuts & bolts

Confidentiality agreements

Some projects may require signing confidentiality/collaboration agreements

- It's time-consuming, but often times necessary
  - It provides a legal framework and reassurance to the provider
  - It prevents some headaches for the publication stage (academic independence)
- Agencies might be used to doing this or you might be the first one
  - Be prepared to extend the HR team a draft
  - Click here and here for an example of my own
- If required, get legal advise and support from your institution
  - If possible, avoid the red tape and solve this independently

Admin Data

Some (unsolicited) advise: adopt the "mindset of a plumber"

A true tax champion would also

- Inspect the tax returns where the data are pulled out from (e.g., 1040 and 1099 in the US, 931 and 649 in Argentina, etc.)
- Read the manuals on how to file tax returns (walk in the shoes of tax payers/preparers)
- Explore the tax filing software: how does it look like? how do people fill in the fields
- Request/search codebooks of the raw data (description of variables, label for categories and special codes)
- Maintain regular contact with officers during the analysis, share intermediate results & methods, show interest in their work, listen to them and build that into your agenda

Final thoughts

Research takes time, especially in the early stages of your career. See, for instance, the first and last presentation of my JMP

- Document every step of your project. Put the key pieces into slides early on
- Presentations: opinions are divided. My view: present, present, present. Go to conferences, talk to people, sign up to office hours and meet speakers
- Present your project to practitioners. It could be beneficial to their work and yours too

I hope you found this material useful.

Contact me if you have suggestions.

E-mail: dario.tortarolo@nottingham.ac.uk

Web: http://dtortarolo.com.ar/

References

Almunia, M., Harju, J., Kotakorpi, K. et al. (2019) "Expanding access to administrative data: the case of tax authorities in Finland and the UK." *International Tax and Public Finance* 26, 661-676.

Brockmeyer, A. (2019). "Working with Administrative Tax Data: A How-to-Get-Started Guide", MTI Practice Notes, Number 7, World Bank Group.

Connelly, R., Playford, C. J., Gayle, V., and Dibben, C. (2016). "The role of administrative data in the big data revolution in social science research." *Social science research*, 59, 1-12.

Currie, Janet, Henrik Kleven, and Esmée Zwiers. (2020). "Technology and Big Data Are Changing Economics: Mining Text to Track Methods." *AEA Papers and Proceedings*, 110: 42-48.

Duflo, Esther. (2017). "The Economist as Plumber." *American Economic Review*, 107 (5): 1-26.

Pomeranz, D. & Vila-Belda, J. (2019). "Taking State-Capacity Research to the Field: Insights from Collaborations with Tax Authorities." *Annual Review of Economics*, 11:1, 755-781

Saez, E., Slemrod, J., and Giertz, S (2012). "The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review." *Journal of Economic Literature*, 50(1): 3-50.

Slemrod, J. (2019). "Caveats to the Research Use of Tax-Return Administrative Data." *National Tax Journal*, 69 (4), 1003–1020.

Slemrod, J. (2019). "Tax Compliance and Enforcement." *Journal of Economic Literature*, 57 (4): 904-54.