

TAXIPP, TAXIPP-LIFE, TAXIPP-FIRM: Exploiting French administrative data for policy evaluation

Antoine Bozio

25th September 2019 – Milano



- **A research centre dedicated to policy evaluation**
 - Based in Paris
 - Created in 2012
 - Scientific partnership between the Paris School of Economics (PSE) and the CREST
 - Funded by research projects

- **A research centre dedicated to policy evaluation**
 - Based in Paris
 - Created in 2012
 - Scientific partnership between the Paris School of Economics (PSE) and the CREST
 - Funded by research projects

- **Methods for policy evaluation**
 - Evaluation ex ante
 - ▶ Microsimulation techniques
 - Evaluation ex post
 - ▶ Randomized controlled trials (RCT)
 - ▶ Quasi-experimental designs (DiD, RDD, event studies, etc.)
 - ▶ Structural econometrics

- **TAXIPP**
 - Arithmetical/static model with behavioural responses
 - Model of the French tax and benefit system
 - Based on administrative data at household level
 - Current version TAXIPP 1.0

- **TAXIPP**

- Arithmetical/static model with behavioural responses
- Model of the French tax and benefit system
- Based on administrative data at household level
- Current version TAXIPP 1.0

- **TAXIPP-LIFE**

- Dynamic model of French population until 2050
- Model of the French pension system and elderly care provisions
- Based on administrative panel data at household/ind. level
- Current version TAXIPP-LIFE 0.1 still under test

- **TAXIPP**

- Arithmetical/static model with behavioural responses
- Model of the French tax and benefit system
- Based on administrative data at household level
- Current version TAXIPP 1.0

- **TAXIPP-LIFE**

- Dynamic model of French population until 2050
- Model of the French pension system and elderly care provisions
- Based on administrative panel data at household/ind. level
- Current version TAXIPP-LIFE 0.1 still under test

- **TAXIPP-FIRM**

- Model of firm taxation
- Based on corporate income tax data
- Still in development

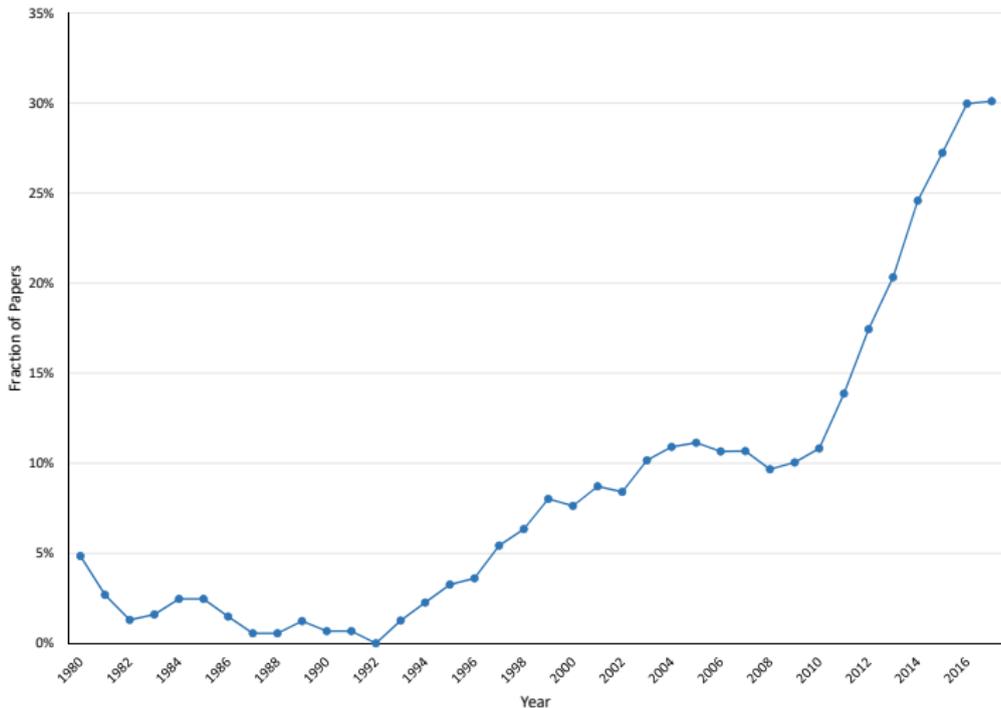
Use of administrative data at IPP

- **A widespread use**
 - For quasi-experimental evaluations
 - For microsimulation models
 - Survey data are still used, but as complement

Use of administrative data at IPP

- **A widespread use**
 - For quasi-experimental evaluations
 - For microsimulation models
 - Survey data are still used, but as complement
- **Why?**
 - Trend in applied public economics
 - Peer-reviewed journals are leading the way
 - Recent change in access to administrative data in France
 - Much improved quality/precision of microsimulation models
 - e.g., heterogeneity within top income decile
 - e.g., spatial heterogeneity
 - Easier link between ex ante and ex post evaluations
 - Building on the expertise of the admin data sources

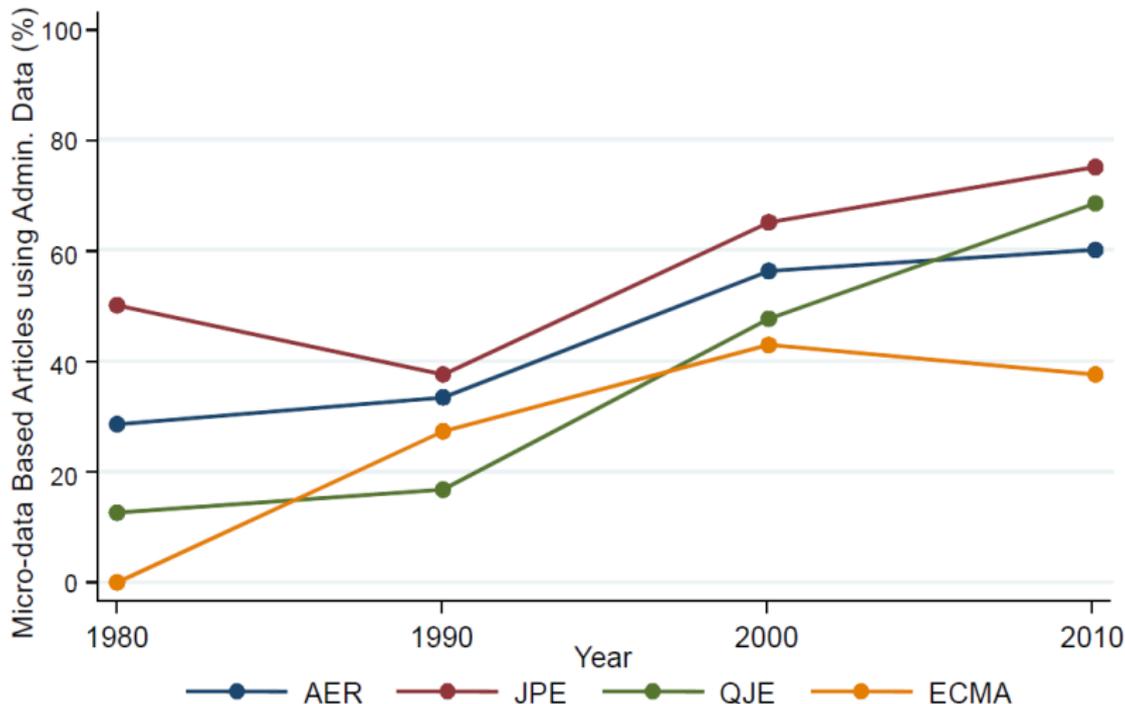
The Rise of Administrative Data in Public Economics



SOURCE: Kleven (2018), "Language trends in Public Economics".

NOTES: NBER working papers 1975–2018 tagged "public economics" (4676 papers).

Use of Admin Data in top Economics Journals, 1980–2010



SOURCE: Chetty and Bruich (2012), Public Economics Lectures.

NOTES: "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.

- **A major change in access to admin data**
 - Law in 2010 for allowing access to health data
 - Law in 2013 for tax data
 - Law in 2016 for all other administrative data (or even data collected for general interest by non public bodies)

- **A major change in access to admin data**
 - Law in 2010 for allowing access to health data
 - Law in 2013 for tax data
 - Law in 2016 for all other administrative data (or even data collected for general interest by non public bodies)
- **Principles of access**
 - Public administrations are mandated to keep confidentiality of individual data collected for administrative purposes
 - New laws provide legal grounds for exception to these strict confidentiality rules
 - Administrations have the right (but not the duty) to provide access to researchers for scientific research only
 - See report by Bozio and Geoffard (2017)

- **Accreditation process**

- Researchers need to obtain agreement from data producers
- Project submitted to *Comité du secret statistique*
- Validation of scientific purpose
- Confidentiality agreement signed by researchers

- **Accreditation process**

- Researchers need to obtain agreement from data producers
- Project submitted to *Comité du secret statistique*
- Validation of scientific purpose
- Confidentiality agreement signed by researchers

- **Remote access technology CASD**

- Centre for Access to Secured data (CASD)
- Technology through a box providing remote access
- Identification by finger prints and digital cards
- Control ex post of all results exported out of the secure environment
- Cost supported by researchers

1. Introduction: administrative data in France
2. TAXIPP, modelling tax and benefits in France
3. TAXIPP-LIFE, modelling pensions and elderly care

TAXIPP, modelling tax and benefits in France

- **Language**
 - Originally written in Stata (versions 0.x)
 - Transcription into Python (from versions 1.0)

- **Language**

- Originally written in Stata (versions 0.x)
- Transcription into Python (from versions 1.0)

- **Version control**

- TAXIPP 0.5 last Stata version
- TAXIPP 1.0 python version
- TAXIPP 2.0 forthcoming Dec. 2019

- **Language**

- Originally written in Stata (versions 0.x)
- Transcription into Python (from versions 1.0)

- **Version control**

- TAXIPP 0.5 last Stata version
- TAXIPP 1.0 python version
- TAXIPP 2.0 forthcoming Dec. 2019

- **Documentation**

- IPP Methodological Guides (Ben Jelloul et al. 2019 for TAXIPP 1.0)

- **IPP tax and benefit tables**

- Legal parameters from 1945 onwards
- Linked to the legal source at *Journal officiel*
- Available online

www.ipp.eu/en/ipp-tax-and-benefit-tables/

- **IPP tax and benefit tables**

- Legal parameters from 1945 onwards
- Linked to the legal source at *Journal officiel*
- Available online

www.ipp.eu/en/ipp-tax-and-benefit-tables/

- **Opensource calculator OpenFisca**

- Collaborative endeavour by Etalab, IPP and other institutions
- Written in python, available on github.com
- Versatile calculator aiming at exact computations
- Used by government agencies to provide information

e.g., <https://mes-aides.gouv.fr>

- **Adaptation of the data to the calculator**

- Different data sources can be used

- **Data from local taxation**

- Database Fideli (DGFIP-Insee)
- Original source is *Taxe habitation* merged with other admin data (income tax returns)
- Universe of all housing units in France
- Precise localisation
- Informations about households, tax units and housing

- **Data from local taxation**

- Database Fideli (DGFIP-Insee)
- Original source is *Taxe habitation* merged with other admin data (income tax returns)
- Universe of all housing units in France
- Precise localisation
- Informations about households, tax units and housing

- **Social Security earnings data**

- Database DADS (Insee)
- Employer-employee data with all earnings and income from self-employed
- Universe of all earners and self-employed in France
- Details about labour contract, employer, occupation, hours of work, etc.
- Base to compute Social Security contributions

- **Income tax return data**
 - Database POTE (DGFIP)
 - Universe of tax returns at tax unit level
 - Details about all taxable incomes, tax credits, etc.

- **Income tax return data**

- Database POTE (DGFIP)
- Universe of tax returns at tax unit level
- Details about all taxable incomes, tax credits, etc.

- **Wealth tax returns**

- Database ISF-IFI (DGFIP)
- Information about taxable wealth

- **Income tax return data**
 - Database POTE (DGFIP)
 - Universe of tax returns at tax unit level
 - Details about all taxable incomes, tax credits, etc.
- **Wealth tax returns**
 - Database ISF-IFI (DGFIP)
 - Information about taxable wealth
- **Consumption data**
 - Survey *Budget des familles* (Insee)
 - Information about consumption basket at household level

- **Income tax return data**
 - Database POTE (DGFIP)
 - Universe of tax returns at tax unit level
 - Details about all taxable incomes, tax credits, etc.
- **Wealth tax returns**
 - Database ISF-IFI (DGFIP)
 - Information about taxable wealth
- **Consumption data**
 - Survey *Budget des familles* (Insee)
 - Information about consumption basket at household level
- **Rents and house prices**
 - Information collected from rent/sale websites (Chapelle, 2018)
 - Precise information about housing units, localisation, etc.

- **Statistical matching**
 - DADS, POTE, FIDELI are matched statistically
 - Exact matching with common variables (age, sex, number of children, housing localisation, taxable income)
 - Matching with Mahalanobis distance for households which have multiple candidates for the match (e.g., households with zero earnings) with additional variables
 - Common identifier between wealth tax returns and income tax returns

- **Statistical matching**
 - DADS, POTE, FIDELI are matched statistically
 - Exact matching with common variables (age, sex, number of children, housing localisation, taxable income)
 - Matching with Mahalanobis distance for households which have multiple candidates for the match (e.g., households with zero earnings) with additional variables
 - Common identifier between wealth tax returns and income tax returns
- **Adding other sources by statistical matching**
 - Consumption patterns matched at household level with propensity score matching on fewer variables and less precision (e.g., income by decile)
 - Housing information added by statistical matching using information on property characteristics and localisation

- **Tax and benefits simulated**
 - Most taxes except firm taxation
 - Most benefits except disability benefits

- **Tax and benefits simulated**
 - Most taxes except firm taxation
 - Most benefits except disability benefits
- **Behavioural responses**
 - Indirect taxation : responses to prices modelled by demand systems
 - Labour supply: responses through extensive and intensive elasticities
 - Capital income: responses through specific capital income elasticities

- **Survey vs admin data**

- Comparing TAXIPP with admin data with same model using ERFs survey data
- ERFs data is aligned to match tax aggregates
- TAXIPP has no alignment : aggregates come from admin data + model
 - ▶ Close result for aggregates
 - ▶ Increased inequality using TAXIPP

- **Survey vs admin data**

- Comparing TAXIPP with admin data with same model using ERFs survey data
- ERFs data is aligned to match tax aggregates
- TAXIPP has no alignment : aggregates come from admin data + model
 - ▶ Close result for aggregates
 - ▶ Increased inequality using TAXIPP

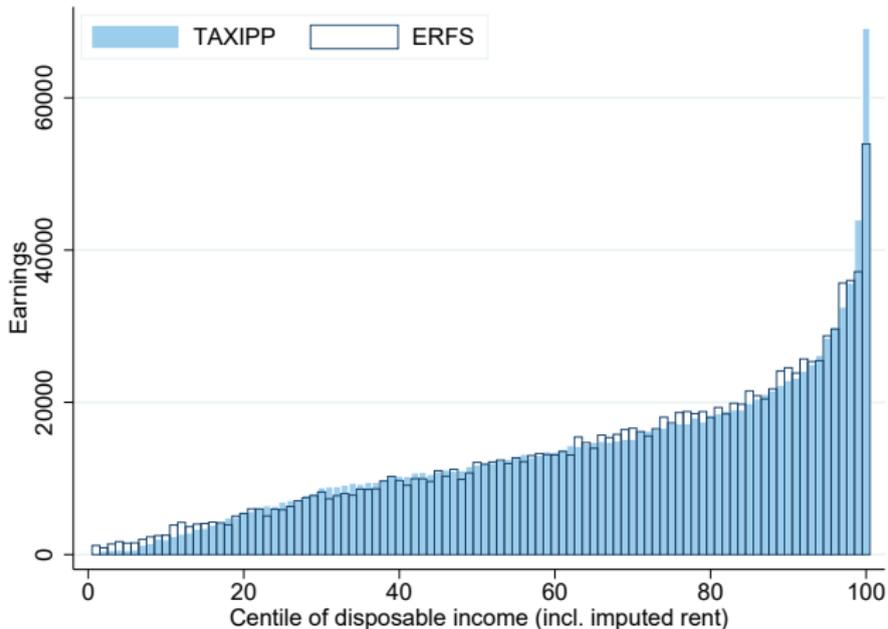
- **Understanding causes of divergence**

- TAXIPP has more poor young households
- TAXIPP has higher incomes for top 1 percent

- **Major differences with 2.0**
 - Tax dataset FELIN is a sample of 500 K tax units
 - FIDELI tax is used for local taxation without statistical matching
 - Labour force data is used for computing SSCs
 - Discrepancy between household and tax unit sources
- **Example of studies with 1.0**
 - Analysis of budget measures
 - Behavioural responses to capital income taxation

Comparison TAXIPP vs survey ERFS

Earnings by centile (Average/consumption unit)

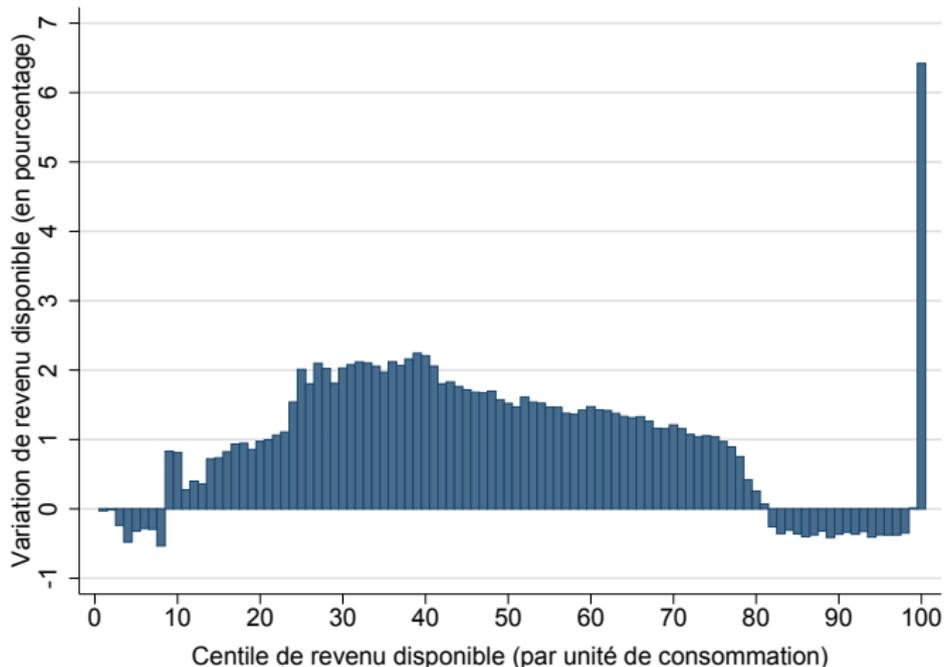


Source: TAXIPP 1.0, ERFS.

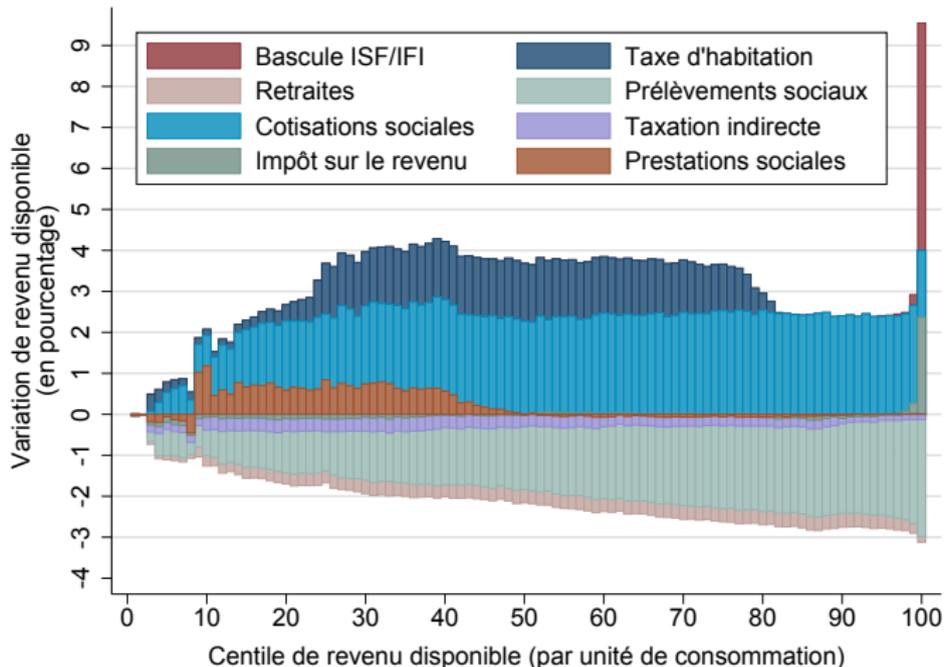
- **Estimating impact of budget measures**
 - Estimation of impact on public finances and redistribution
 - Static analysis (“the day after the budget”)
 - Using macroeconomic estimates from the government
 - Counterfactual of inflation uprating of tax/benefit schedules
 - Household ranked according to disposable income by consumption unit

- **Estimating impact of budget measures**
 - Estimation of impact on public finances and redistribution
 - Static analysis (“the day after the budget”)
 - Using macroeconomic estimates from the government
 - Counterfactual of inflation uprating of tax/benefit schedules
 - Household ranked according to disposable income by consumption unit
- **Example: 2018-19 French budget**
 - Shift from SSCs to flat-rate income tax CSG
 - Cut in local taxation
 - Increases in in-work tax credits
 - Introduction of flat-rate tax for capital incomes
 - Removal of wealth tax

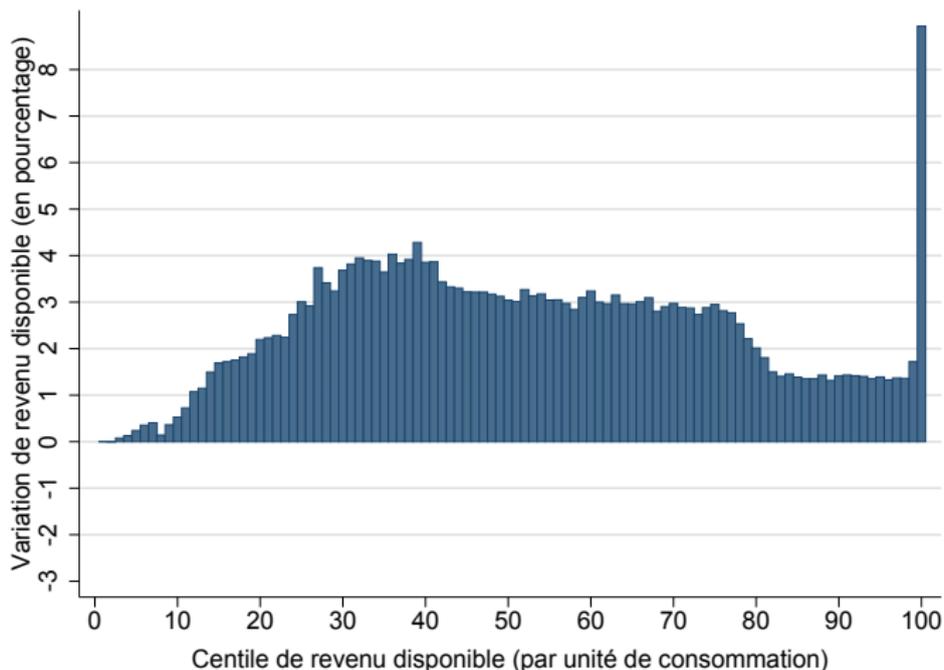
Gain/loss in disposable income



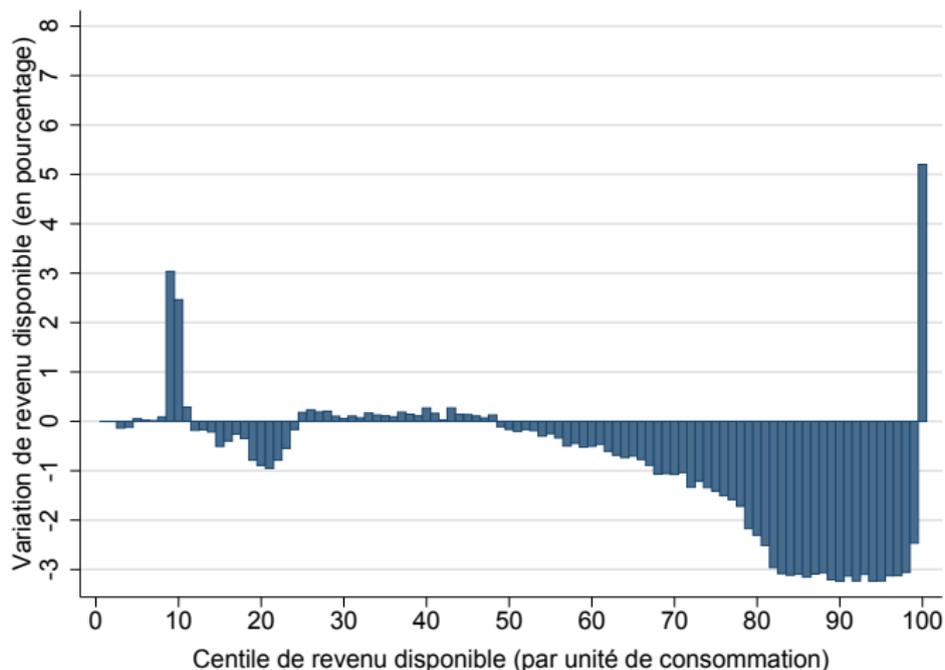
Gain/loss in disposable income



Gain/loss in disposable income



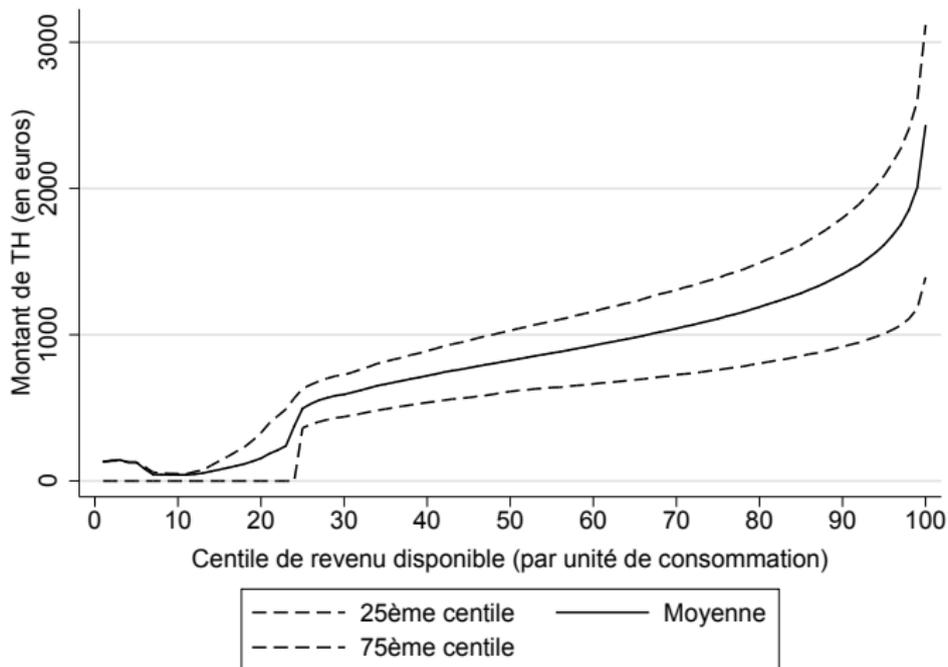
Gain/loss in disposable income



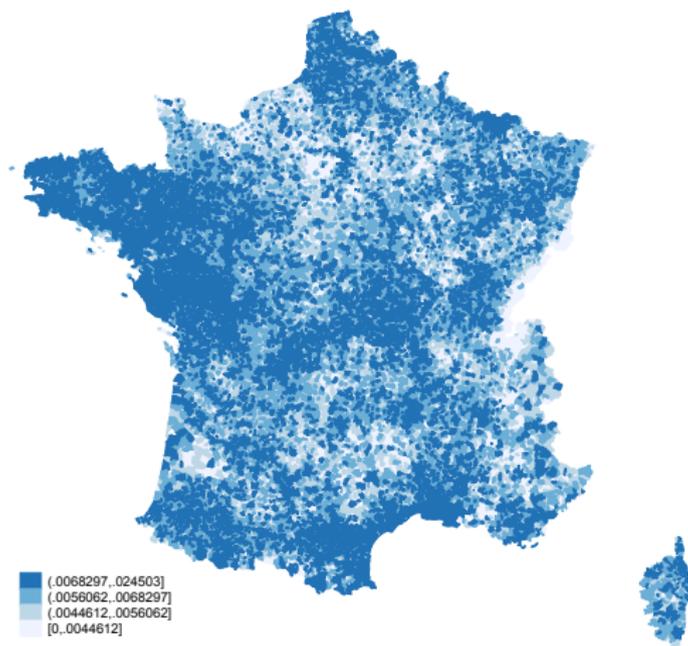
- **Cut in local tax “taxe d’habitation”**
 - Compensated by central gov. transfer
 - Creates significant heterogeneity by local authority

- **Cut in local tax “taxe d’habitation”**
 - Compensated by central gov. transfer
 - Creates significant heterogeneity by local authority
- **Exploiting TAXIPP local granularity**
 - Gains to tax cut depends from previous rate and previous tax base
 - Tax base based on 1970 rents
 - Significant heterogeneity conditioning on household income

Pre-reform local tax rates by household incomes



Average gain per local authority
(in percent of average local income)



- **Ex post evaluation**
 - Use TAXIPP to identify affected households : first stage
 - Estimate elasticities using administrative data
 - Use elasticities in TAXIPP for counterfactual analysis

- **Ex post evaluation**

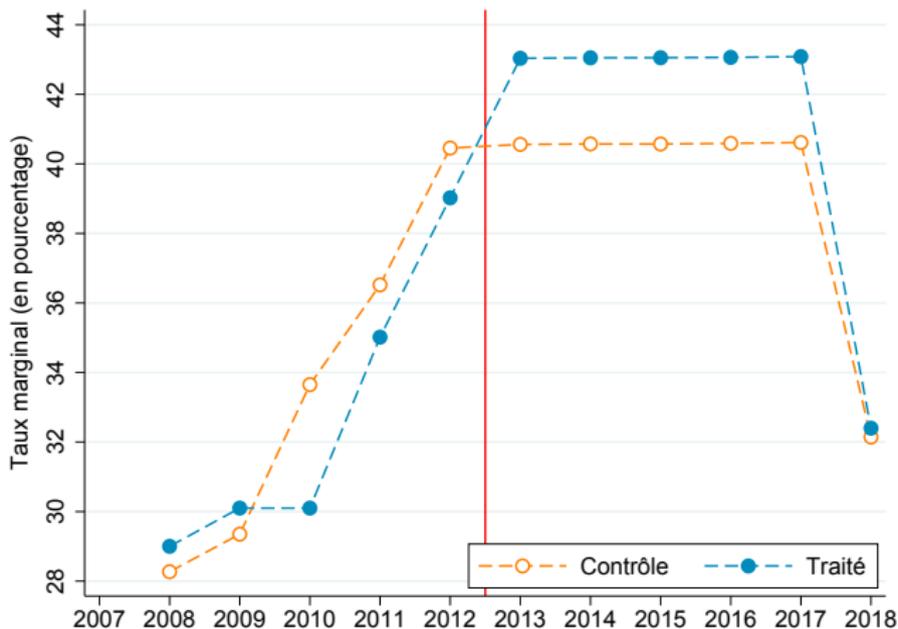
- Use TAXIPP to identify affected households : first stage
- Estimate elasticities using administrative data
- Use elasticities in TAXIPP for counterfactual analysis

- **Example: capital income taxation**

- 2013: capital incomes incorporated into regular income tax schedule
- Pre-2012 households could opt for flat rate tax
- 2018: flat-rate tax becomes default option for capital income taxation
- Bach et al. (forthcoming 2019) conduct impact evaluation using household and firm data

Ex post evaluation of capital income tax changes

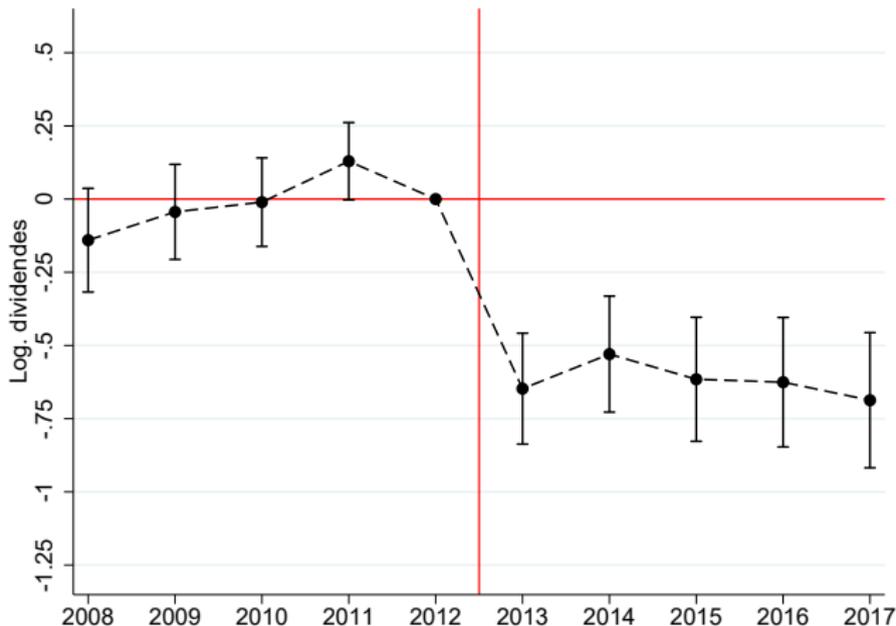
First stage – change in marginal tax rates (treated vs controls)



Sources: TAXIPP 1.0, POTE (DGFIP), Bach et al. (2019).

Ex post evaluation of capital income tax changes

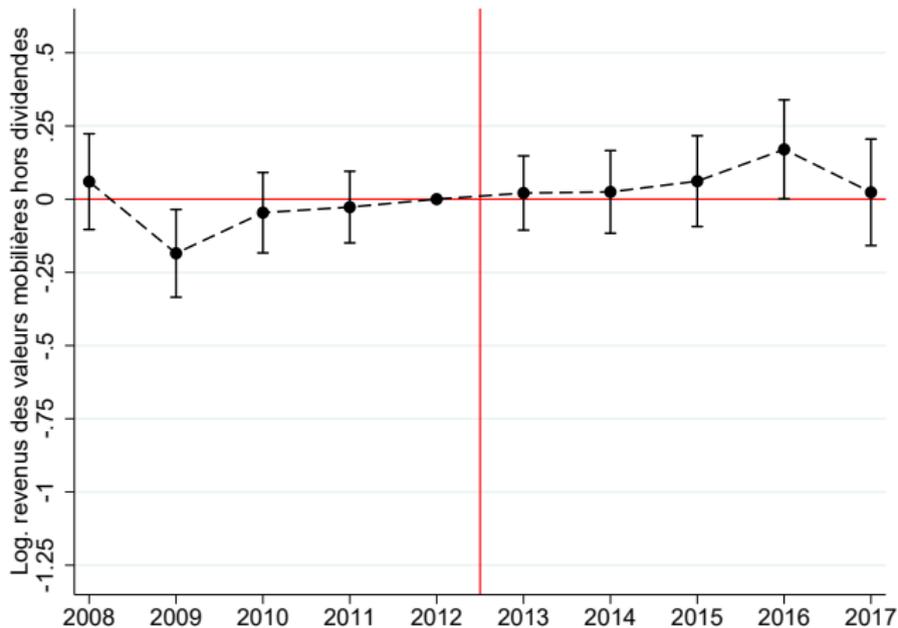
Difference in difference estimates – dividends



Sources: TAXIPP 1.0, POTE (DGFIP), Bach et al. (2019).

Ex post evaluation of capital income tax changes

Difference in difference estimates – other capital incomes



Sources: TAXIPP 1.0, POTE (DGFIP), Bach et al. (2019).

- **Static estimates were off the mark**
 - Government's estimates of revenues were overestimated in 2013 and underestimated in 2018
 - Redistributive analysis is also biased by static approach
 - But real impact depends of the mechanism of behavioural responses

- **Static estimates were off the mark**
 - Government's estimates of revenues were overestimated in 2013 and underestimated in 2018
 - Redistributive analysis is also biased by static approach
 - But real impact depends of the mechanism of behavioural responses
- **Mechanisms: exploiting firm data**
 - Bach et al. (forthcoming 2019) show that responses driven by firms controlled directly by affected individuals
 - Drops in dividends matched by increased in retained earnings
 - Importance of taking into account relationship between firms and households

TAXIPP-LIFE

- **Dynamic model based on administrative data**
 - Many different pension schemes in France
 - Good administrative panel data on earnings (DADS)
 - Issue to combine individual level data with household level
 - Destinie model (Insee) on Wealth Survey
 - Prisme model (CNAV) on pension scheme admin data

- **Dynamic model based on administrative data**
 - Many different pension schemes in France
 - Good administrative panel data on earnings (DADS)
 - Issue to combine individual level data with household level
 - Destinie model (Insee) on Wealth Survey
 - Prisme model (CNAV) on pension scheme admin data
- **TAXIPP-LIFE: combining survey with admin data**
 - Insee has merged individual data with household information
 - Survey data on elderly care and health status

- **Panel data on earnings matched with Census**
 - Database EDP++ (Insee)
 - Original source is Census, matched with DADS data, and tax returns
 - Sample of French born in first week of October
 - Earnings since 1967
 - Household composition, children, marital status from Census and State registry

- **Panel data on earnings matched with Census**
 - Database EDP++ (Insee)
 - Original source is Census, matched with DADS data, and tax returns
 - Sample of French born in first week of October
 - Earnings since 1967
 - Household composition, children, marital status from Census and State registry

- **Administrative data from pension system**
 - Database EIC and EIR (Drees)
 - Information necessary to compute pension rights in all mandatory schemes
 - Matched with DADS data and unemployment records

- **Survey data on elderly care**
 - Survey CARE (Drees)
 - French population aged 60 and above
 - Information about health conditions, ADL, IADL, informal care, etc.

- **Survey data on elderly care**
 - Survey CARE (Drees)
 - French population aged 60 and above
 - Information about health conditions, ADL, IADL, informal care, etc.

- **SHARE data**
 - European population, ADL, IADL, in panel dimension

- **Simulation of French pension system**
 - Most pension schemes simulated (public and private)
 - Information at household level allows to assess poverty/disposable income issues
 - Large sample size reduces variance of results

- **Simulation of French pension system**
 - Most pension schemes simulated (public and private)
 - Information at household level allows to assess poverty/disposable income issues
 - Large sample size reduces variance of results
- **Simulation of elderly care projections**
 - Estimation of transition matrices from Survey data
 - Imputation of health condition in 50+
 - Projection of population in need of care according to various scenarios

Conclusion

- **Access to administrative data is transformative**
 - Ex post evaluation of many policies become possible
 - Microsimulation models can be enriched by exploitation of large sample, precised information about schemes/benefits/tax credits, etc.
 - Zoom on specific population (by incomes, by localisation, etc.) or specific scheme

- **Access to administrative data is transformative**
 - Ex post evaluation of many policies become possible
 - Microsimulation models can be enriched by exploitation of large sample, precised information about schemes/benefits/tax credits, etc.
 - Zoom on specific population (by incomes, by localisation, etc.) or specific scheme
- **Looking forward: plenty of work!**
 - Potential of current data sources is much bigger than currently exploited
 - TAXIPP, TAXIPP-LIFE, TAXIPP-FIRM could be combined
 - Behavioural responses should be improved in many dimensions

- Bozio, A. et Geoffard, P.-Y. (2017) *L'accès des chercheurs aux données administratives – État des lieux et propositions d'actions*, Rapport du groupe de travail du Cnis.
- Ben Jelloul, M., Bozio, Douenne, T., Fabre, B., and Leroy, C. (2019) "The 2019 French budget: impacts on households", IPP Briefing notes, No. 37, January.
- Ben Jelloul, M., Bozio, Douenne, T., Fabre, B., and Leroy, C. (2019) "Le modèle de microsimulation TAXIPP – version 1.0", *Guide méthodologique IPP*, September.
- Bach, L., Bozio, A., Fabre, B., Guillouzouic, A., Leroy, C. and Malgouyres, C. (2019) "Évaluation d'impact de la fiscalité des dividendes", *Rapport IPP* No. 26, October 2019.
- Bach, L., Bozio, A., Fabre, B., Guillouzouic, A., Leroy, C. and Malgouyres, C. (2019) "Follow the money! Combining household and firm-level evidence to untangle the tax elasticity of dividends", manuscript, PSE.