The use of administrative data in PBO microsimulation models

M.R. Marino

EUROMOD WORKSHOP

Milano, 25-26 September 2019

upB ufficio parlamentare di bilancio

Outline

- PBO main tasks
- Access to data
- Instruments for the assessment of public policies: microsimulation models and the role of administrative data
- Lessons learned so far

PBO main tasks (L. 243/2012):

- **Endorsement** and **evaluation** of macroeconomic and public finance forecasts and **assessment of compliance** with national and European fiscal rules
- Ex-ante and ex-post **evaluation of macroeconomic and financial effects** of the most important legislative measures
- Analyses of the **long-term sustainability of public finances**
- Analyses and reports on its own initiative or at the request of the parliamentary committees responsible for public finance matters

Access to data

In order to carry out its functions, the (reinforced) law establishing the PBO guarantees **privileged access to information sources of administrative nature** (Law 243/2012, Article 18, paragraph 7):(l. 243/2012, art. 18, c. 7):

«Al fine di consentire all'Ufficio lo svolgimento dei propri compiti istituzionali, le amministrazioni e gli enti di cui al comma 6 [tutte le amministrazioni pubbliche, gli enti di diritto pubblico e gli enti partecipati da soggetti pubblici] assicurano all'Ufficio medesimo l'accesso a **tutte le banche di dati in materia di economia o di finanza** pubblica da loro costituite o alimentate»

To make this access effective, however, an amendment was approved unanimously by both Chambers of the Parliament:

«Ai fini dell'accesso ai dati raccolti per fini statistici ai sensi del decreto legislativo 6 settembre 1989, n. 322, l'Ufficio è **equiparato agli enti ed uffici facenti parte del Sistema statistico nazionale**» (L. 164/2016)

Concerning the **evaluation of public policies**, the PBO analyses mainly (not only) focus on the **quantification of financial effects**

- **Mainly first order effects** (static scoring): without accounting for behavioral reactions of involved economic subjects and markets
- **In some cases second-order and subsequent effects** (dynamic scoring): including the effects of the measures on the economy as a whole and the consequent feedback on public accounts

(i.e. consumption reactions to price and income shocks, changes in the take up rate of transfers policies, changes in the incentives to retire, companies reaction to tax rules, etc.)

The impact assessment of public policies also includes:

- Distributive analysis (revealing the effective nature of a tax reform)
- General equilibrium effects (interactions among markets)
- Analysis of the incentives produced on involved economic agents

Requirements for PBO microsimulation models

Given the PBO tasks, microsimulation models must meet **two important requirements**:

Accuracy ⇒ reliable and credible representation of actual tax liabilities and public transfers (e.g. fiscal variables net of tax evasion and erosion)

Completeness of information ⇒ wider set of information, beyond actual tax bases from administrative data

- «effective» economic conditions gross of tax evasion and erosion, to determine «effective» distributive effects
- need for several context variables to model complex phenomena, such as incentives, behavioral reactions, etc.

Key issue: accurate selection of data sources

A microsimulation model based exclusively on survey data to estimate **actual tax bases and tax liabilities** may suffer from several drawbacks:

- Small samples size
- «Measurement errors» of fiscal variables
  - Tax evasion
- Sampling distortions
  - Sample design not checking for variables which are relevant for tax calculations (non-response cases, underrepresented top incomes)
- Incomplete information (with respect to the calculation of relevant tax variables)
  - Erosion
  - Non-deductible costs
  - Individual choices
  - Income definitions
  - Other problems (cadastral values, imputed incomes)

Problems with administrative data

On the other hand, microsimulation models based only on administrative data (usually those used by government agencies) are **highly reliable to reproduce effective tax liabilities,** but:

- Limited demographic and socioeconomic information
- Underrepresent lower end of income distribution
- Administrative data reflect past legislation rules
  - Exclusion of potential new taxpayers or incomes relevant for policy assessment
- The assessment of distributive effects is difficult because of tax evasion and erosion

To overcome these problems a **direct linkage of survey data with administrative data** on the same population (or corporations obviously) is performed (using id codes):

- Administrative data are used to derive actual tax bases (overcoming the problems of measurement errors) and survey data are used to estimate the missing information in the administrative archives
- Identification of sampling distortions with respect to tax distribution and its correction through post-stratification techniques (calibration of survey weights, controlling their variability)
- Assessment of taxpayer economic conditions through information on net income collected in the survey to measure the redistributive effects of fiscal policies (hypotesis of no under reporting in answers to the survey questionnaire)

Similar linkage has been performed by the National Statistical Office and other similar institutions for Silc

However, **they have a different purpose**:

"The aim […] is to improve data quality on income components […] by means of imputation of item non-responses and reduction of measurement errors [of actual incomes]" (Consolini P. and Donatiello G. - 2015)

- The aim is to use administrative sources to get a better representation of "actual" incomes, rather than a precise measurement of fiscal aggregates:

("*when the net administrative incomes are higher than the survey incomes, the net and gross incomes completely arise from administrative data. On the opposite, […] the net incomes are those taken from the survey*")

- Since **original administrative data are not identified** in the publicly distributed data, we choose to perform a new matching procedure

PBO models and data sources

Microsimulation models

For policy assessment purposes, the PBO developped **a set of sectoral microsimulation models**:

- Tax-benefit model for households (C. Pollastri)
- Corporate tax model (F. Gastaldi)
- Short-term pension expenditure model (N. Salerno)
- Long-term pension expenditure dynamic microsimulation model (work in progress) (L. Toffoli)
- Short-term model on interest expenditure connected to public debt (C. Pollastri and C. Gabbriellini)

Static multi-period microsimulation model

Objectives

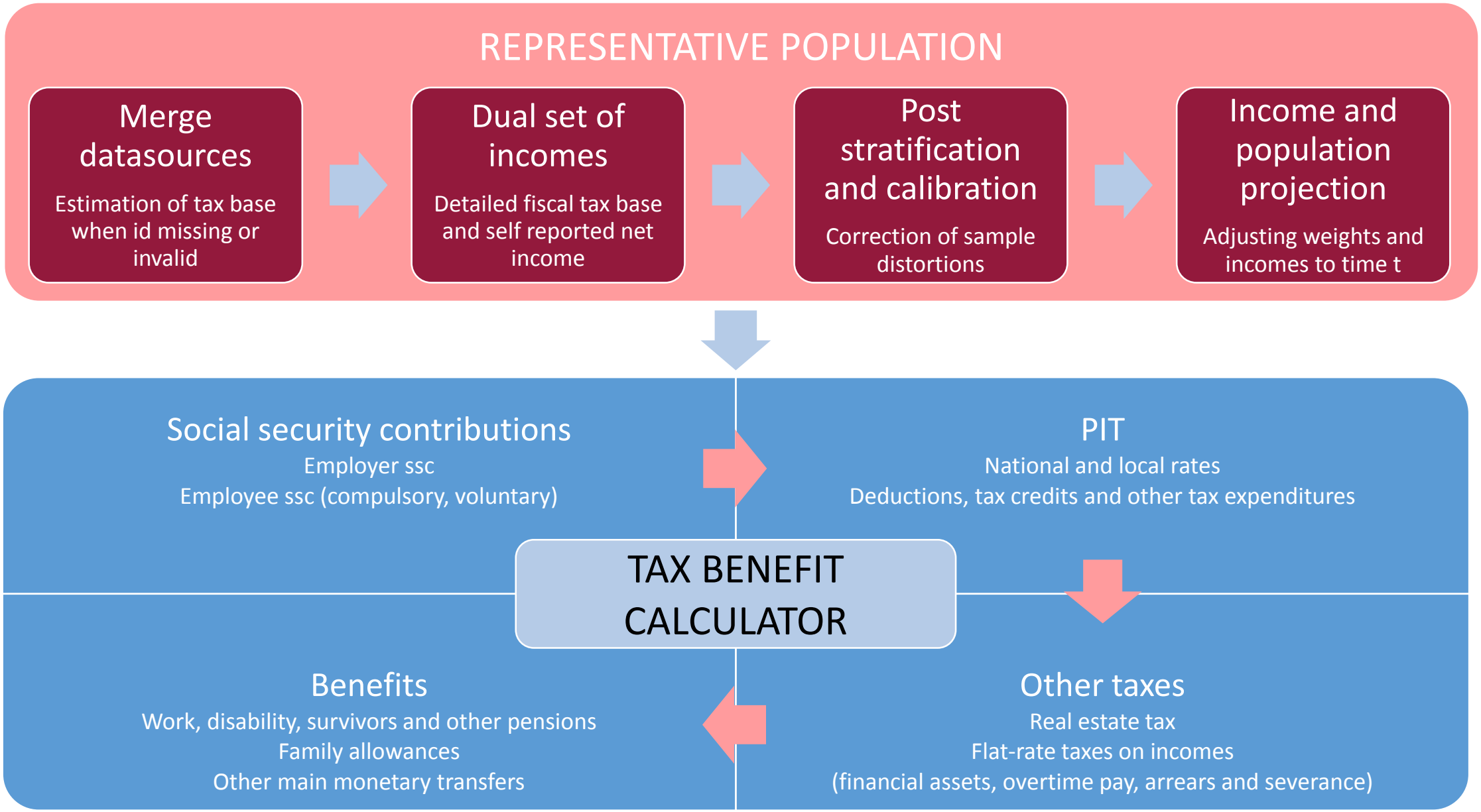
Allows accurate simulation on:

- Direct taxes
- Indirect taxes
- Benefits/Transfers
- Social security contributions

Models and data

- Direct taxes and transfers
  - It-Silc (S) + income tax returns (A) + social contributions and pensions (A)
- Direct and indirect taxes and transfers
  - Household expenditures (S) + income tax returns (A) + social contributions and pensions (A) + ISEE statements (A)
- Minimum income and other mean tested transfers
  - ISEE statements (A) + It-Silc (S) (in progress)
- Large sample of tax returns (A not linkable) for robustness analysis

Direct taxation and benefit model - Workflow



Developments of the model: behavioural aspects

- Short run behavioral reactions (costing purposes)
  - Modelling consumption reactions to price shocks (indirect taxation) and income shocks (direct tax + indirect tax + ssc effects on disposable income)
    - Estimation of demand side (substitution elasticities between groups of goods / services)
    - Estimation of MPCs to evaluate consumpion reactions to shocks on income
    - Integration with PBO macroeconomic model (memo.it) in order to estimate short term effects on the whole economy
- Other behavioral reactions (long run impact on the whole economy)
  - JRC assistance to develop an instrument for assessing the impact of reforms on labour supply
  - Linkage to a general equilibrium model (PBO version of Quest) to estimate long run effects

- Discrete choice model
  - Easier to run than other approaches (non linear budget constraint, household level utility optimization)
- Estimation steps
  - Heckman estimation of *w* (missing information for unemployed)
  - Building set of alternatives, $\{\overline{w}(h) - t(\overline{w}(h)), h \in H\}$
    - H: discrete set of hours worked (0,…,k);
    - $\overline{w}(h)$*:* estimated salary
    - $t(\overline{w}(h))$: microsimulated tax & benefit
  - Conditional logit estimation of the alternatives probabilities (given disposable income)
  - Estimation of labor supply (weighted average of hours in the set with estimated probabilities)
- Simulation
  - Recalculation of the work supply, given changes in disposable income due to changes in tax/benefit scheme

- Simulation of indirect and direct taxes on the same household
  - Exact match between household expenditure survey, tax returns and pensions of household members via administrative id
  - Vat and excises paid by households on the basis of HBS survey
    - Breakdown of VAT rates for Coicop classification of expenditure items
  - PIT estimated from tax returns, limited simulation capacity of other income taxes
    - Simplified tax calculator
    - Simplified income taxonomy (no income information from survey)
    - Some sources of income will be estimated to get a broad view of disposable income
- Past history of incomes (drawn from social security records) will give a life cycle perspective

- Simulations of welfare policies mean tested with ISEE
- Sample of 2% of ISEE statements submitted from 2016 to 2018 (1mln individuals)
  - Income (PIT taxable base and other exempted or separated taxation income), financial and housing wealth, household characteristics
- Simulations rely on the number of current recipients of benefits that could be eligible to the new policy criteria
  - On one side there is no information on “new” potential beneficiaries… but on the other it is possible to take into consideration current take-up
  - Robustness analysis performed on the whole population
  - Future development: model take-up on It-Silc linked to ISEE statements

Examples of PBO simulations 2015-2019

- Parliamentary hearing (Nov. 2015), PBO Budgetary Policy Report for 2016 and PBO Focus Paper No. 6/2015
  - Abolition of service tax
- Parliamentary hearing (Oct. 2016) and PBO Budgetary Policy Report for 2017
  - Reform of household related allowances and benefits
- Parliamentary hearing (Nov. 2017)
  - Social inclusion income (Reddito di inclusione)
- Parliamentary hearing (Nov. 2018) and PBO Budgetary Policy Report for 2019
  - Flat tax for self-employed and sole proprietors
- Parliamentary hearings (Nov. 2018, Feb., March and July 2019) and PBO Flash 1/2019
  - Citizenship income

Static multi-period microsimulation model

Objectives

Allows accurate simulations on:

- Corporation tax (IRES)
- Business tax (IRAP)
- Social security contributions (work in progress)

Data

- Censuary data at company level (S) (1 mln excluding financial sector) (BvD) linked (2015-2017) with administrative data (tax returns for IRES, IRAP, consolidated tax returns) (A)
- Working on the link with ISTAT survey data on active enterprises (ASIA)

Financial Accounting Data + aggregated tax return data (2011-17)

- Distributive analysis of tax revenue changes (full range of relevant differences between firms)
- Problems: tax codes contain complex and detailed rules for the determination of the tax base and tax due and necessary information details are not included in available firms data sets

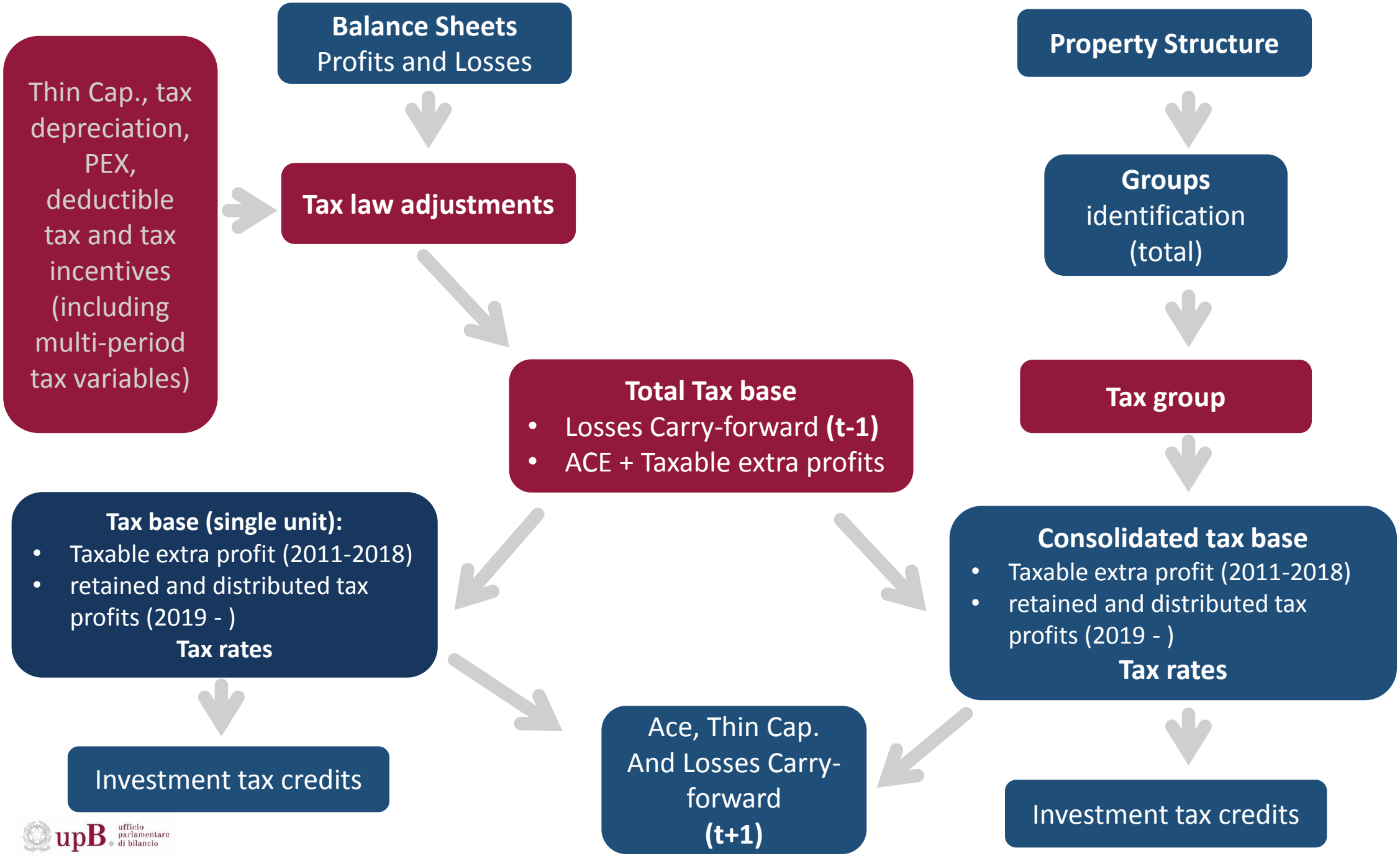
Integration with Micro Tax Return Data (Administrative register) (2015-17)

- Improves precision and allows more comprehensive analysis overcoming the previous problems
- Improves the assessment of budgetary impact
- Allows exogenous identification of tax groups for consolidated taxable base and sectoral matching
- A new module computes corporation tax for financial sector and allows simulations of some tax changes

Developments of the model: behavioural aspects

- Multiperiod extension
- Investment and financing decisions
- Possible assistance from European Commission Structural Reform Support Programme

Corporate tax model workflow



upB ufficio parlamentare di bilancio

Focus paper no. 1/2016

Investment incentives in the 2016 Stability Act

2016 Budgetary Policy Report

- *Corporate tax rate cut from 27.5% to 24%*
- *Extra depreciation (Superammortamento)*
- *Investment tax credit*

2017 Budgetary Policy Report

Changes in the ACE regime (2017 Budget Bill)

a) reduction of the notional return on equity

b) restriction of tax base relief

2018 Budgetary Policy Report

- *ACE - reduction of the notional return on equity*
- *Changes in thin capitalisation rules*

2019 Budgetary Policy Report and Focus Paper no. 4 / 10 June 2019

The redistributive effects of the main measures adopted with the budget package for 2019:

a) Abolition of the ACE,

b) introduction of the new reduced rate on retained profits,

c) increase in the deductibility of municipal property tax (IMU)

d) and extra depreciation (Superammortamento)

Static multi-period microsimulation model based on meso-data

Objectives

- Predict flows of new retirees in the short-term (EFD time horizon, t + 5, max t + 10) and related gross expenditure, with the possibility of separate evidence for pension schemes, type of pension (old-age, early retirement), calculation regime (wage, mixed, contributory)
- Simulate short-term effects of changes in the accessing rules and / or in the pension formula (ex ante and ex post analysis)

Data

- Sample of insured persons (i.e. active workers) provided on a yearly base by INPS. From 2018 (reference year 2015) the sample includes private employees (FPLD), public servants (ex INPDAP), ENPALS, self-employed, parasubordinate, separate INPS schemes
- The data are organized by cells (about 95,000), with individuals sharing a number of relevant characteristics. The universe is then obtained via a re-proportioning coefficient

How it works

1) Current legislation rules are applied to all active workers in year t in order to select the potential pensioners and calculate pension rights. Age and seniority of non retiring people are increased by one year (taking into account specific death probabilities of all ages)
2) At present the history of salaries is reconstructed starting from the labor income in year t under the assumption that the individual has worked continuously for a number of years equal to his contribution seniority and in the same job position (private or public employee, self employed, etc.)
3) In the forecasting horizon (up to t + 5) wages increase with GDP (or similar hypotheses) and pension rights are indexed to inflation
4) We are in the process of using instead data from INPS administrative archives with the effective wages and carrier histories for all workers

Examples of model application

Pension expenditure forecasts at least twice a year for public accounts validation purposes

Focus paper no. 6/2016

Evaluation of two proposals aiming at introducing more flexibility in retirement ("Damiano" and "Boeri")

Parliamentary Hearing on Bill no. 1018 on citizenship income and pensions (Feb. 2019)

Evaluation of:

a) *Quota 100 (early retirement possible if workers age + years of contributions ≥ 100) and Opzione donna*

b) *Temporary stop to the automatic link of pension requirements to life expectancy*

c) *Changes to the pension benefits indexation to consumer prices*

Microsimulation model with dynamic population (IT-Silc population is projected up to 2065)

Objectives

- Long-term projection of pension expenditure and long-term care (sustainability of public finances)
- Assessment of distributive effects of the social security and welfare systems

Data

- Integration of survey data (IT-Silc) with administrative data (INPS) on salaries and social contributions (A) of the interviewed people

Lessons from PBO experience

- Administrative data is increasingly used as a new source for **producing or complementing official statistics** (OECD, 2019)
- But also growing interest in the use of administrative data for the **evaluation of public policies** (forecasting, simulation and analysis of administrative phenomena):
  - by central and local public decision makers (strategic management)
  - by public units in charge of the analysis and the evaluation of public policies
  - from academics and researchers
- The availability of elementary microdata is crucial

Unlike operational management, the microdata needed for the evaluation of public policies on the one hand do not require the identification of individuals but on the other hand generally require an **enrichment of information** through the **precise and timely linkage of multiple archives**

Currently **limited access to administrative microdata**:

- limitations due to the legislation on personal data (even if the identifiers are not useful to analysts)
- slow affirmation of the open data culture between data managers
- further difficulties with the introduction of the General Data Protection Regulation (GDPR)

Outcomes:

- outside the SISTAN, only limited experiences of access to administrative microdata: VisitINPS, Laboratorio ADELE, but in controlled environments
- within the SISTAN, complex exchange when the use of linked microdata is at stake: data must be requested separately to the individual producing administrations (no exchange of linked sources)

  → costs and duplications of linkages

- A **central role of the National Institutes of Statistics** (i.e. ISTAT for Italy) for the distribution of administrative data necessary for the evaluation of public policies
  - Availability of a considerable amount of (already) linked administrative data
  - Possibility of further linkages of administrative data with its own survey data
  - Possible savings for entities in the SISTAN wishing to link databases
  - Wider diffusion outside the SISTAN of linked microdata (adequately anonymized)
- Higher acknowledgement of **the role of administrative information** in the field of statistical production. Acknowledgement of the information content and value of administrative data, even if they are “incomplete” and not “fully consistent” with national and international statistical standards
    - Example: ISTAT’s Archimede project for the dissemination of linked administrative microdata, not statistically treated and not necessarily consistent with the official statistics
    - Systematic integration in National statistical programmes of projects to link administrative data with the purpose of evaluating public policies

Thank you for the attention!

mariarosaria.marino@upbilancio.it

upB ufficio parlamentare di bilancio