## SOSE EST

Workshop Economico, Statistico e Tecnologico

DATA SCIENCE
PER LE POLITICHE
PUBBLICHE

#### Roma - 30 OTTOBRE 2019

DESIGN AND USE OF MICRO-MACRO CONSISTENT LARGE ECONOMIC DATA BASES

Federico Perali

Dip.to di Scienze Economiche & ELL– Università di Verona



#### WHY IS IT IMPORTANT?

#### **SOSE SOSE**

"...FOR TOO LONG GOVERNING HAS BEEN CONSIDERED AN ART. NOW IT IS NECESSARY TO MAKE IT A SCIENCE" S. Cassese, 2013 "La qualità delle politiche pubbliche, ovvero del metodo nel governare"

#### • Evidence based policy making

- Ex Ante policy Analysis
  - DPEF, for example, should be based on an ex-ante estimation of both the macro and micro impact of the proposed policies that should be ranked in terms of their contribution to social welfare
- Ex Post policy analysis and program evaluation as normal practices of a good public administration
- Public debate should be based on certified facts and forecasts
  - Fakes and poorly argumented opinions are not satisfactory:
    - VAT Ex: What are the general equilibrium consequences of the VAT sterilization? Who looses, who wins in case of a VAT increase? By how much? How do consumers react? Labor supply?
- These dreams would not be possible if a reliable Micro-Macro Consistent Large Economic Data base were not implemented, efficiently maintained, and easily accessible and interpretable by both politicians and citizens!



#### **INSTITUTIONAL INVESTORS ON ECONOMICS BIG DATA**

#### **SOSE SOSE**

- BigNOMICS is a project within the JRC's Centre for Advanced Studies aiming at big data collection to provide better real-time **economic "fore and now casting"** through:
  - *Better models*: improve the accuracy of econometric models based on economic and financial variables observed at the daily, weekly and monthly frequencies also leveraging on artificial intelligence and machine learning techniques.
  - *Better data*: collect detailed economic datasets, newspaper articles, satellite images that can provide real-time and relevant indicators of economic activity.
- FAO's **RuLIS** (RUral Livelihoods Info System)
  - The FAO Statistics Division, the World Bank and IFAD are building **RuLIS**, a tool to support and monitor evidence-based policies for reducing poverty.
  - RuLIS brings together harmonized indicators and data for a large number of developing countries.



#### CHARACTERISTICS OF AN SOSE ST ECONOMY WIDE MICRO - MACRO DATA BASE

- Large
- Wholistic Micro-Macro Foundations of the Whole Economy
- Dynamic and spatial
- Integrated across surveys
- Consistent with National Accounts and quality certified
- Meaningfully Aggregated
- Apt for Macro and Micro Behavioral Simulations
- On-time access and data elaboration for ex-ante policy design and expost policy and program evaluation
- Understanding society: a socio-economic approach to causality



## FROM ECONOMIC THEORY TO DATA BASE DESIGN

- Demand for economic information
  - For direct access and interpretation
  - To feed micro and macro economic models
- Supply mainly from ISTAT
  - Not Ready for use
  - Data preparation and maintenance
  - Not designed to include instrumental variables
- Let us look at the "General" Equilibrium model to define the demand for statistical information necessary to feed the model and to design the associated data base maintaining an exact micro-macro consistency



#### A GENERAL REPRESENTATION OF AN ECONOMY

**SOSE SOSE** 

Flow of parameters and variables + space and time dimension



## REPRODUCING THE FUNCTIONING OF THE ECONOMY **SOSE SOSE SOSE**

#### MACRO

Dynamic Stochastic CGE Behavior of Hhs, Firms, Govt Data: National, regional, local SAMs + agg. time series



Tax-Benefit Simulator Non-Behavioral Models (Hhs, Firms) Data: Eusilc, HBS, Time Use - AIDA

**SO** 

#### MICRO

Consumption, Labor and Production Simulators Behavioral Models (Hhs, Firms) Data: Eusilc, HBS, Time Use - SOSE

#### MICRO-MACRO EXACT AGGREGATION



#### • MICRO

• Microsimulation analysis estimates the distributive effects of household or corporate taxation. If non-behavioural, then responses and adjustments in related markets to reform-driven changes in prices, wages and macro variables are not accounted for.

#### • MACRO

• Micro-funded CGE models assess the impact of exogenous variables and policy measures (i.e., tax rates) on economic equilibria (i.e., prices) that feed the micro level (e.g. EUROMOD). Because CGE models are based only on few households and firms' types, they do not properly capture agents' heterogeneity and related distributional effects.

Important to ensure accounting consistency between the disaggregate micro level and aggregate macro level using exact aggregation conditions in a presimulation phase and an exact micro-macro closure iterating until convergence (i.e. when two iterations give same wage rate) during simulation.



#### THE BIG (MICRO) DATA BASE FOR POLICY USE SOSE SET

- EUSILC matched with administrative data (Income, Savings, Poverty, Social Exclusion and Living Conditions)
  - Cross-sectional data pertaining to a given time or a certain time period
  - *Longitudinal* data pertaining to individual-level changes over time, observed periodically over a four-year period.
    - Social exclusion and housing condition information is collected at household level while labour, education and health information is for persons aged 16 and over. The core of the instrument, income at very detailed component level, is collected at personal level.
- EU-SILC based dataset for labor supply analysis
- Household Consumption Data
- EU national and regional level SAMs
  - Available at World Input-Output Database (WIOD), IRPET and ELL for Italian regions

www.sose.it

• Statistical Matching (Consumption + Income + Time Use + Quality of Life)





# Analytical

# Example



www.sose.it

#### EUROMOD ... IN SCRIBBLES

SOSE

#### **Household Gross Income**

(from EUSILC matched with admin. data)

$$y_h = rk_h + wl_h$$

**Household Net Income** 

( $\tau$  is the household specific tax-benefit rule)

$$\mathbf{y}_h^{net} = (1 - \tau_h)\mathbf{y}_h + transf_h =$$

$$= \underbrace{(1 - \tau_h)rk_h + transf_h}_{= y_h^{exo} + (1 - \tau_h)wl_h}$$



#### **SOSE SOSE**

#### THE MICRO LEVEL

$$\max_{\{c_h, l_h\}} \left\{ U_h^j(c_j, 1 - l_j) \,|\, pc_j = \mu_j \cdot \left( (1 - \tau_h) l_h w_h + y_h^{exo} \right) \right\}$$

whose solution gives optimal collective consumption (s=food, housing, education, health, transportation, recreation, others) and collective labor supply

$$\begin{cases} c_{h,s}^{j} = c_{h,s}^{j}(w, p, y_{h}^{exo}, \tau) \\ l_{h}^{j} = l_{h}^{j}(w, p, y_{h}^{exo}, \tau) \end{cases} \quad \forall j = male, female and \forall s sectors \end{cases}$$

that feed the macro CGE level.





#### THE MACRO CGE LEVEL



Solving Walrasian (to exemplify) equilibrium

$$\begin{split} & \sum_{h} l_{h} = L^{dem}(w/p, Q) \\ & \sum_{h} k_{h} = K^{dem}(w/p, Q) \\ & \sum_{h} c_{h} = Q \end{split} \implies (w^{*}, p^{*}, r_{numeraire}; Q^{*}) \end{split}$$





#### A GENERAL REPRESENTATION OF THE MICRO-MACRO LINK



•  $(w^*, p^*, r_{numeraire}; Q^*) \Longrightarrow$  to Euromod:

$$\mathbf{y}_h^{net} = \mathbf{y}_h^{exo} + (1 - \tau_h) \mathbf{w}^* l_h^*$$

then iterate until convergence when two iterations give the same wage rate.

• Incorporating labor supply and consumption schemes consists in replacing the FOCs in the CGE by estimated consumption and labor (collective) supply

$$\begin{cases} c_h^i = \hat{c}_h(w, p, y_h^{exo}, \tau) \\ l_h = \hat{l}_h(w, p, y_h^{exo}, \tau) \end{cases}$$



#### COMMENTS

#### **SOSE SOSE**

- Failing to close the model at the macro level would be "distortionary partial".
- Focusing on labor supply alone, would break the utility optimization:
  - $\hat{c}_h$  would be determined residually to satisfy the budget constraint

 $pc_h = (1 - \tau_h)l_h w_h + y_h^{exo}$ 

- An example:
  - consider an aggregate shock, such an economic crisis or ... a "Fr...exit", then use EUROMOD, and CGE iteratively, to generate  $\hat{y}_h$ ,  $\hat{G}$  and find the endogenous policy rule  $\hat{\tau}$  that neutralizes the shock.
- "Nowcasting" may gain from a close integration with the microbehavioral modelling of labour market changes accounting for major demographic changes such as migration, family re-formation and dissolutions.





## **Demand for**

## **Micro and Macro**

## Models



www.sose.it

#### THE MACRO DEMAND FOR BEHAVIORAL MODELS SOSE SEST

- Design and Estimation of a Spatial/Dynamic General Equilibrium Model that reproduces exactly:
  - The *benchmark economy* corresponding to the SAM's year using calibrated parameters (for simple preferences and technology representations) or parameters estimated by more flexible consumption, production and labor supply micro models
  - The endogenous mechanism determining fixed capital both for public and private investments.
  - The *actual fiscal system* in a fashion consistent with the non-behavioral models simulating both the direct and indirect taxation
  - *design and simulate the policy rules* related to the prospect reforms of public interest such as (FT, Citizens' income, IRAP elimination, VAT remodulation, Web tax, pet food tax, environmental taxes, etc.) in coherent fashion with micro-models related to households and firms.



#### THE MICRO DEMAND FOR BEHAVIORAL MODELS SOSE SET

- Micro-simulation models of household behaviour based on
  - Consumption C (Household Budgets HB)
  - Income I (SILC-ITA)
  - *Labor* supply L (SILC-ITA)
  - *Production P* (SOSE firm level data)
- The C,I,L models can be jointly estimated starting from an integrated database matching HB and SILC data bases
  - The equilibrium solution of CGE models by sector and income class are the inputs of micro-models.
  - The micro-macro accounting closure is ensured with a procedure with successive iterations
- CAVEAT for poverty-concerned policy makers
  - fragile families are missing from the sampling designs of household surveys



#### MICRO DATA PREPARATION – EXAMPLE FROM CONSUMPTION DATA

- Exact Aggregation
- From High to Low Level of Detail
  - Imputation of Missing Observations
  - Missing values
  - Corner solutions
  - Infrequency of purchases
  - Measurement errors
  - Endogeneity problems
  - Theory Consistent Aggregation
- Prices NOT AVAILABLE from ISTAT  $\rightarrow$  build pseudo prices
- Income Data not Available to the public
- Regional Representation





#### **CPI (NIC) – FOOD AND BEVERAGES**







#### EXPENDITURES AND SHARE TRENDS (HBS 1999-2013)







#### EXPENDITURES AND SHARE TRENDS (HBS 1999-2013)







#### NON-BEHAVIORAL MODELS - DEMAND SIDE SOSE SEST

- EUROMOD is a *static* tax-benefit microsimulation model for the European Union that enables researchers and policy analysts to calculate, in a comparable manner, the effects of taxes and benefits on household incomes and work incentives for the population of each country and for the EU as a whole. It is based on EU-SILC data
- However,
  - No indirect taxation
    - requires investing on household budgets data and stat matching techniques
  - No tax evasion
  - *Not applicable to local taxation* (but for Trento)
  - No Micro-Macro Integration





#### **NON-BEHAVIORAL MODELS - FIRM SIDE**



- ISTAT-MATIS is a corporate tax microsimulation model for Italy (Caiumi and Di Biagio 2015) simulating in a multiperiod framework corporate tax liabilities of given fiscal rules.
- Uses a large integrated database of corporate tax returns for all Italian corporations
  - reports results by firms category according to technological intensity, financing structure, profitability, size, age, location, export orientation, and ownership structure.
- The model predicts the distribution of the tax burden among taxpayers and the revenue impact of tax changes (e.g. IRES, IRAP, ACE).
  - Not behavioural, thus, limited to "dry runs" (first round effects).
- The multiperiod framework uses observations at the firm level for consecutive time periods (panel data). Currently, the integrated database spans the period 2005-2015.
- Data sources:
  - company accounts database,
  - ISTAT archive on national business groups,
  - the statistical register of Italian active enterprises (acronym ASIA),
  - information on spin-offs and mergers,
  - business structural surveys:
    - survey on foreign trade (COE); survey on Italian enterprises controlled by foreign firms (Fatsinward) and on resident firms with foreign subsidiaries (Fats-outward).



#### MAIN CHARACTERISTICS OF THE ISTAT-MATIS DATA BASE

#### **SOSE SOSE**





www.sose.it



## **Demand for**

## **Micro and Macro**

### Data



www.sose.it

#### MACRO DEMAND FOR DATA - I SOSE ST

- Based on a BLED (Big Local Economy Data Base) of Social Accounting Matrices (SAM) at the 20 Regions level (63 branches) and at the 107 Province (55 branches) level that aggregate exactly up to the national SAM available at the Economics Living Lab starting from 1999 up to present
- Interegional trade estimated using the CHARM method (Tobben, Kronenberg, Construction of Multi-Regional Input-Output Tables Using the Charm Method, Economic Systems Research, 2015) also useful to implement gravity models or using B2B (business-to-business) inter-firm trade data based on electronic invoicing.
- Extended SAMs using Satellite Accounts both regional and Province level
  - Environmental Impact Matrices
  - Industry/Occupation (highly detailed information on jobs by sector and workers'skills)
  - Commuting flows
  - Migratory flows also using GIS data
  - Public investments.



#### MACRO DEMAND FOR DATA - II SOSE ST

- This *in-house data warehouse* is obtained using a *top-down* methodology from publicly available data.
- Province and region level SAM data should be connected using a relational language SQL associated with modern data visualization tools ensuring a real time interaction with data in a single multiregional and multi-temporal "model array".
- Key Challenges
- *Challenge 1*: model and statistical coherency
  - SAM, micro and macro models should be coherent among themselves and statistically consistent mainly with national accounts.
- Challenge 2: robotization
  - The update of the *data warehouse* can be implemented every year at low cost but maintaining a high level of statistical precision.





#### MICRO DEMAND FOR DATA SOSE SEST FROM BEHAVIORAL AND NON-BEHAVIORAL MODELS

- Creation of a *data warehouse* starting at year 1990
  - EUSILC-ITA with further data preparation for the estimation of labor supply equations using Random Utility Models
  - ISTAT household budgets with
    - household specific pseudo unit-values (Menon, Perali and Tommasi 2018 Stata Journal) using ISTAT original *files*
    - registered income data from IRPEF records

facilitating the statistical *matching* with EUSILC-ITA and the *exact linkages* between macro and micro models also necessary for implementing non-behavioral tax-benefit models suitable for both IRPEF and VAT simulations.



#### DEMAND FOR INFORMATICS SERVICES

**SOSE SOSE** 

- Common platform with protected access
  - data warehouse (micro and macrodata)
  - software
  - micro and macromodels
- *Need for web-interfaced user-friendly* software (and associated apps working both on IOS and Android Environments) for the simulation analysis in real time of the SAMs that quantifies the impact of exogenous changes (environmental disasters included) on local economies and contiguous districts including direct and indirect effects
- ... more in detail



#### INTEGRATED MACRO DATA AND SOSE EST MODELING LABS where data warehouses of regional and national time-series social accounts are created

- Regional and national social accounts Lab:
  - Methods to assemble social accounts for regional and national economies, production functions and trade behavior models.
- Commodity Trade Modeling Lab:
  - Modeling environment to estimate shipments of commodities between foreign and domestic places of production and consumption to provide an economy-wide picture of interacting global supply chains. Create multiregional social accounts with unprecedented levels of spatial, temporal and sectoral granularity.
- Forecasting social accounts Lab:
  - Methods to assemble a time series of regional social accounts and project social accounts for future time periods.
- Satellite Accounts Lab: Applications
  - Adding Capital Accounts (financial, physical, human, natural, social, cultural, political);
  - Integrating current-year Quarterly Covered Employment and Wage data to social accounts;
  - Expanding government fiscal accounts in social accounts;
  - Integrating workforce Occupation / Knowledge / Skills / Abilities with social accounts; environmental; migration and mobility flows.





#### LOCAL AND REGIONAL DATA BASES SOSE ST

- Historical Data Warehouse of
  - Social Accounts for all Italian Provinces and Regions, 1990 2018.
  - Control Totals (Output, Employment, Value Added, and Final Demands) for all Italian Provinces, 1990 – 2018 used to assemble social accounts.
  - Production Functions used to assemble social accounts.
  - Commodity trade behavior for all Italian Provinces, 1990 2018 used to assemble social accounts.
- Methods to combine Control Totals, Production Functions and Commodity Trade to assemble regional and national social accounts for Italy.



### **SOCIAL ACCOUNTS**

#### **SOSE SOSE**

Because Leontief output models and CGE models are derived from the same regional social accounts, an opportunity exists to re-engineer and mass produce regional social accounts that are ready for use in CGE modeling equations.

- Social accounts:
  - double-entry statistical framework (bookkeeping) that measures all the sales and purchases in a region at a particular time.
- Based on and adds up to national input-output data.
- Expressed in row and column format.
- The rows shows the sales to meet demand.
- The columns shows the purchases to create supply.
- Each industry's sales equal their purchases.



#### **MULTIPLIER EFFECTS**





e.it

#### **MULTIPLIER EFFECTS**

#### **SOSE SOSE**

Indirect Spending







#### INTEGRATED INFORMATION ARCHITECTURE SOSE SEST



#### CONSTRUCT DATABASE PROTOCOLS SOSE SEST

- Protocols for developing the Integrated Database will be housing on a cloudbased, relational Structured Query Language (SQL) database server as a collection of SQL Stored Procedures for social accounting:
  - data acquisition processes as work processes for Extracting, Transforming and Loading (ETL) extant data into coherent relational databases that serve as SAM accounting identities or control totals.
  - database construction and maintenance. These are work processes for assembling and reconciling the data teams' control totals into relational databases that use the definitions and conventions of social accounts as a logical data framework.
- These processes may be performed in a Modelling Unit using support tools such as SAS Enterprise Modeler (SAS), Microsoft Business Intelligence (BI), and other custom BI tools. Data teams will be guided by the Integrated Data and Modelling Management Team.

![](_page_36_Picture_5.jpeg)

#### ANALYSIS DASHBOARDS SOSE ST FOR MACRO MODELS: SOFTWARE APPLICATIONS TO EXPLORE THE INTEGRATED DATA WAREHOUSE AND APPLY ADVANCED ECONOMIC ANALYSIS TOOLS

- Economic Structures Dashboard
  - Contribution Analysis; Structural path analysis; Income Distribution; Time Series / Comparative Advantage; Cross-Sectional / Economic Diversity.
- Economic Connections Dashboard
  - Supply chains & Value chains; Trade Modeling; Trade MRIO; Trade Mapping; Power Series; Labor force (Occupations & Skills).
- Analysis Models Dashboard
  - Traditional Input-Output Models;
  - Computable General Equilibrium Models
- Develop Graphical User Interfaces and Applications Program Interfaces

![](_page_37_Picture_9.jpeg)

#### MICROSIMULATION AND CAUSALITY SOSE ST

- The challenge: implementing causal analysis using observational data, not only experimental data!
  - Propensity scores methods, LATE techniques, partial identification via bound-analysis and relative-correlation restriction methods
  - Evaluation of counterfactual contrasts via statistical regression models informed by graphical causal models (directed acyclic graphs, or DAGs).
  - A supplementary method is agent-based modelling due to its potential for simulating counterfactuals.

![](_page_38_Picture_5.jpeg)

### CONCLUSIONS

![](_page_39_Picture_1.jpeg)

- Large and complex data bases are costly and require a large coordination effort despite investing in robotized data collection and organization processes
- A Dedicated Data and Modelling Lab is a Compelling Institutional Innovation to
  - Satisfy the demand for real-time, micro-macro consistent, data and independent analysis at the community, regional, national and super-national level
  - Avoid being "policy takers" at the European and global level
  - ... and to effectively make governing a science rather than an excessively "popular" art!

![](_page_39_Picture_7.jpeg)