ENTERPRISE MICROSIMULATION MODELS AND DATA CHALLENGES

PRELIMINARY RESULTS FROM THE DIECOFIS PROJECT•

Filippo Oropallo°

Abstract:
Microsimulation models and micro-founded indicators investigate the fiscal impact on enterprises and on their performance at a micro level, so they require a great deal of information, which is normally scattered in various statistical surveys and other (administrative) sources. Each different data source is conceived to serve different purposes and, in many instances, may refer to different units or different definitions may be used for the same unit. Any attempt to bring together data from different sources has to overcome complex problems in terms of sheer access, integration and systematisation. The paper shows the analysis of the integration problems that have been faced, the architecture of the integration process that has been adopted. Then first results bring to the construction of micro-founded indicators and a microsimulation study is carried out to analyze the tax burden on enterprise linked with a performance indicator.

Keywords: Data integration, micro-simulation, micro-founded indicators

Disclaimer: The views expressed in this paper are that of the author. They cannot and should not be attributed to his employer.

• The work is developed under the DIECOFIS (**D**evelopment of a system of **I**ndicators on **E**conomic **CO**mpetitiveness and **FIS**cal impact on enterprise performance) project, financed by the Information Society Technologies Programme (5th Research Programme of the European Union) (contract n. IST-2000-31125) coordinated by Paolo Roberti and Maria Grazia Calza, ISTAT, website www.istat.it/diecofis.
I am grateful to Alessandro Fazio, Lorenzo Lo Cascio and Danilo Orsini for the valuable contribution to the activities of the Diecofis Project and to the fulfilment of the paper.
I would like to express my gratitude to Roberto Monducci, Carlo De Gregorio, Giovanna Dabbicco, Alessandro Zeli, Giampiero Siesto, Cristina Casciano, Giulio Perani, Pasquale Papa (Business Statistics - ISTAT), Manlio Calzaroni, Giuseppe Garofalo, Caterina Viviano, Danila Filipponi, Monica Consalvi and Enrica Morganti (Business Census and Archives - ISTAT) for their availability in collaborating to Diecofis achievements.
° ISTAT (National Statistical Institute of Italy), email oropallo@istat.it.

1 Introduction

Understanding how policies affect economic performance and developing better indicators to gauge their effects is central to endow the policy makers with a set of efficient and fair policies. Existing knowledge on policy impact analysis may be considered approximate. Usually the "facts" on the impact of policies are charted only at the aggregate level and with a high degree of approximation. Macro indicators have well-known pitfalls and drawbacks. The gap in European knowledge and capacity for policy impact analysis is patent. The DIECOFIS EU-FP5 project has taken up the challenge of reducing this gap in the field of taxation. Results have been quite encouraging and have open new vistas for future work. Particularly notable has been the development of a system of micro-founded indicators, based on factuals and counterfactuals, estimated through micro-simulation models. Micro-founded indicators on enterprise performance and fiscal microsimulation models require a great effort in integrating data at a micro level. In the construction process of the datasets for microsimulation purposes, several problems were encountered in the combination of data from different sources and the following issues were handled: (i) the reconciliation of survey data with administrative data, (ii) the treatment of missing data and (iii) the sample weights adjustment.
This paper presents, in paragraph 2, the multi-source integrated and systematised data base of enterprises: the description of the sources, the architecture of the integration process and the solutions that has been adopted. In paragraph 3 the paper shows the preliminary results drawn from the microsimulation analysis on the distribution of the performance and of the tax burden.

2 DIECOFIS Database[1]

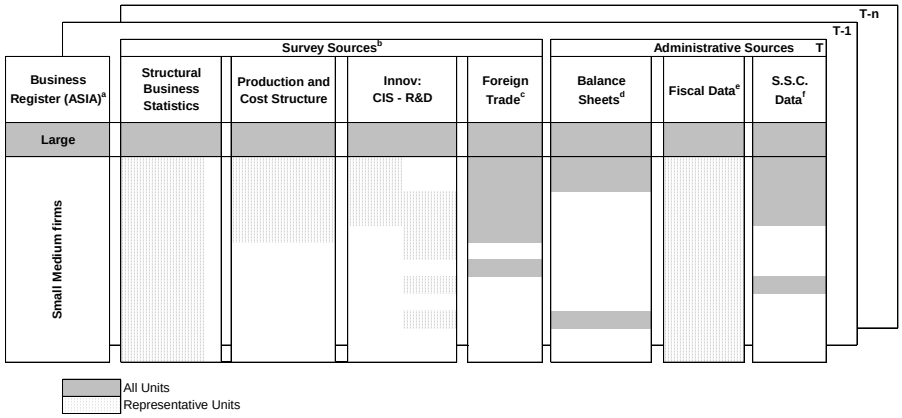
In order to build the Enterprise Integrated and Systematized Information System (EISIS) the first step is the selection of the "spine" information that will be used as a basis for the integration process. At ISTAT, the "spine" is constituted by the statistical register of Italian active enterprises (ASIA) [2].

[1] See Denk, Oropallo F. (2003) *Overview of the issues in longitudinal and cross-sectional multi-source databases* – www.istat.it/deicofis

[2] The ASIA project started in 1995, its goal is to improve and update the register of all Italian enterprises. It is the result of the integration of external sources with ISTAT Archives (old Sirio-nai archive, 7° Industry Census and survey SK). External sources are: VAT Register of

This is the result of an integration process of different administrative sources and represents the best "hanger" for data integration purposes. On this hanger, information from the following sources can be put. Large Enterprise Accounts (*SCI*); Small and Medium Enterprise Survey with less than 100 workers (*PMI*); Manufacturing Product Survey (*Prodcom*); Foreign Trade Archive (*COE*); Other surveys such as the Community Innovation Survey (*CIS*) and the *ICT* Survey. All of the above ISTAT surveys are based on common EUROSTAT standards and classifications (as shown in chart 1).

Chart 1 – General Framework



(a) The spine is the population of reference of the Business Register ASIA according to the Commission Regulation n. 2186/93 and n. 696/93 on statistical units

(b) These sources are governed by: (i) Council Regulations: no 3924/91 - survey of industrial production; (ii) n. 696/93 - statistical units for the analysis of the production system; (iii) n. 58/97 - structural business statistics; (iv) Commission Regulation: n. 1618/99 - evaluation of quality of structural business statistics

(c) Enterprises involved in foreign markets

(d) Incorporated enterprises only (Their account system is regulated by EU directives)

(e) Sample of Fiscal returns data - This database is shaped by the National Tax Legislation

(f) Enterprise with employees only - This database is shaped by the National Social Security Legislation

This implies that the DIECOFIS database can serve to simulate at a "micro level" the impact of public policies not only in Italy but also in other countries and that a path for the creation of an EU statistical information system has been traced. The main effort which it was necessary to undertake was the development of a methodology to allow data linkage between the information of the above surveys and the whole enterprise universe, represented by the data register on enterprises. In the ASIA archive, ISTAT files all active enterprises[3] except for those belonging to Agriculture, Forestry and Fishing

---

the Ministry of Finances; Chambers of Commerce; INAIL (National Institute of Insurance Against Accidents at Work); INPS (National Social Security Institute); Yellow Pages and other specific archives.

[3] cf. Eurostat 1999a.

(A, B sectors according to NACE classification) and to the Public Sector (L, O91, P and Q). This can be used as a starting point or common basis for the linkage of all survey data. In the ASIA archive the following information is included: identifier (internal code, name, fiscal code, vat number, telephone, address); localisation (geographical reference); typology (economic activity and legal form); demographic (status and transformations); size (turnover and employees).

The information coming from the administrative sources that have been integrated in the DIECOFIS database include: Commercial Accounts (*CA*) data from the Chamber of Commerce annual report that complement ISTAT business survey of account system (*SCI* and *PMI*) for all corporate, co-operatives and consortium enterprises only[4]; Fiscal data (*FISCAL*) from the Revenue Agency annual tax returns; Social Security data (*SSD*) from the Italian Social Security Institute (*INPS*). These two latter sources permit to obtain precise information on tax and social contribution revenues, and thus to calculate the actual tax burden on enterprises, which can be used to test the model's output (e.g. "counterfactuals"). Looking at the quality of the available information, enterprise size seems to be a "key" variable. In fact, exhaustive information (which covers the whole universe) is available for large enterprises that have at least 100 workers, while for small and medium ones only sample data is available. A second characteristic that appears to be very important is the legal form, as the type of tax that an enterprise is required to pay depends on it. The first problem is to identify the business unit. This means basically choosing a variable which can be a unique key and act as a natural bridge between the different sources. In almost all firms' databases the ID code is represented by the VAT code or the fiscal code. Another important question relates to possible changes to the business[5] during the enterprises' life. In fact, the same enterprise may appear as a different unit because of transformation events. Usually two types of changes are considered: changes involving a single unit (changes in kind of business classification, in size or localisation); changes in the number of units (death, birth, divestitures and splits, mergers and acquisitions). As a consequence of changes or in the presence of new-born firms, the business register doesn't contain all the units of a survey and it is necessary to distinguish between the case of new firms and that of transformed units. In the latter case, a problem of identifying the successor of the initial business can arise. In some cases, the VAT number of the new unit is different but the fiscal code is the same. A correspondence table containing old and new codes or a table containing the fiscal code and

[4] CA data contains a sample of corporate enterprises which have a precise a legal form. The variable legal form assumes the following values: Sole proprietorship (legal form =0), partnerships (legal form =1), corporate firms (legal form =2), co-operatives (legal form =3), consortium and other legal entities (legal form >3).

[5] cf. Black 2001.

the many VAT numbers used by the enterprise has been used in order to solve this kind of problem. This table has been constructed on the basis of VAT code and Fiscal code archives obtained from the Italian Revenue Agency. A further study on this kind of issue and in particular on the business demography is carried out at ISTAT and first results have been published in a Eurostat statistics in focus[6]. The study deals with the application of exact matching technique to distinguish new born firms and death firms excluding those firms involved in a transforming event.

2.1 Data Sources

At the start the integration process is concerned with the creation of the following datasets, as shown in chart 2: (1) *Business Register* (years: 1996–2000) includes basic information on the whole universe of Italian active enterprises, so that the contained characteristics can serve as auxiliary variables in the processes of imputation and estimation, these are: geographical reference, sector of economic activity, legal type, dimension (independent workers and employees) and the annual turnover (for this variable there is a small percentage of missing values).

Table 1 – Business Register

Year	*1996*	*1997*	*1998*	*1999*	*2000*
Large Ent.	8,091	8,684	8,924	9,240	9,741
SME	3,862,383	3,761,446	4,040,250	4,122,853	4,212,916
Total	3,870,474	3,770,130	4,049,174	4,132,093	4,222,657

(2) *Survey datasets* (Regional datasets, year 1999) include information on 55 thousand firms, coming from Structural Business surveys; information is exhaustive for large corporate firms (roughly 9 thousand firms with 100 workers or more), sample data are available for small and medium enterprises (roughly 46 thousand firms). The 1999 *SCI survey* (data are available also for year 1998) contains information on 8,734 enterprises. They refer to the universe of large enterprises with 100 or more workers (with the exclusion of the J division, financial sector). Among these, there are more than 7 thousand corporate enterprises. The 1999 PMI survey (data are available also for the year 1998) contains information on 45,867 enterprises, of which more than 15 thousand are corporate enterprises.

[6] cf. Eurostat 2003.

Table 2 – Survey data

	SCI (> 99 workers)		*PMI (<100 workers)*	
Year	*1998*	*1999*	*1998*	*1999*
Corporate Enterprises (legal form =2)	7,124	7,340	15,372	15,330
Non Corporate Enterprises (other legal forms)	1,330	1,394	32,112	30,537
Total	8,454	8,734	47,484	45,867

(3) *Administrative datasets* (Corporate datasets, year 1999), take account of 54 thousand corporate firms, coming from *CA*[7] data. Information is exhaustive for large corporate firms (7 thousand firms with 100 workers or more). Sample data is available for small medium sized corporations (47 thousand firms); this merging activity makes it possible to proceed with the second stage of integration: the missing data reconstruction, performed through the use of matching techniques. The different sources are now exposed in detail.
With respect to CA data (years '98-'00), there is, for the year 99, a sample of 53,532 enterprises: 6,911 with 100 or more workers and roughly 47 thousand with less than 100 workers. The *Fiscal* dataset contains a targeted sample of tax returns for the year 1999[8]. It contains all large corporate firms and a sample of small medium sized.

Table 3 – Administrative data

	CA (legal form =2)		*Fiscal (Corporate tax data)*
Year	*1998*	*1999*	*1999*
Large Enterprises (> 99 workers)	6,197	6,911	7,340
SME Enterprises (<100 workers)	48,261	46,621	4,535
Total	54,458	53,532	11,875

[7] A sample of Company Accounts (CA) is extracted from the whole archive of the Chamber Commerce, managed by Pitagora and available at ISTAT. In the 2000 the access to the database containing the balance sheets of all corporations (excluding the financial institutions) will be available.
[8] At this stage is only required a sample of fiscal data to test integration and harmonization procedures. The sample is coordinated with the PMI and SCI surveys.

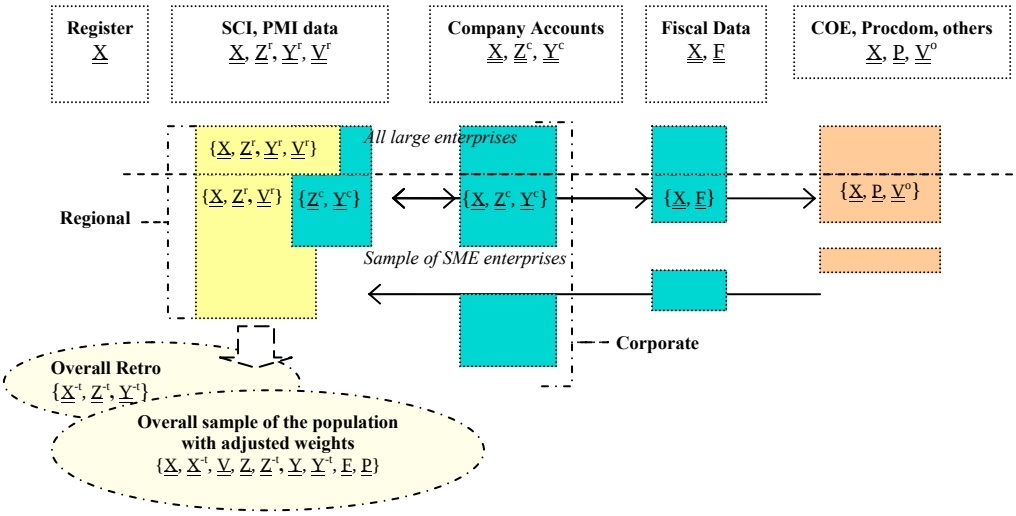
At this stage, other surveys have been linked with the business register. 1) *Prodcom survey*[9] is exhaustive for large enterprises and there is a sample of small and medium ones (approximately 35,000 units). It covers the manufacturing sector only. Other sectors, such as trade and services, remain uncovered. Moreover, for small and medium enterprises, there is the problem of a missing link between *PMI* sample units and units from the *Prodcom* sample. In this case the integrated information of *PMI* and *Prodcom* can refer simultaneously to the single stratum (domain) of the sample appropriately inflated to the population of the domain. 2) *Foreign trade archive (COE)* integrates information about foreign trade for the totality of enterprises[10]. It is derived from custom data and covers the population of enterprises engaged in foreign trade (approximately 260,000). It contains the value of every item exchanged (with a detail of 8 digits) for each country of destination and origin. 3) *Technological Innovation of enterprises* survey (CIS-Community Innovation Survey) collects information on expenses for innovation projects and on the type of innovation in question. The purpose is to estimate the input and output of the innovation process that takes place in enterprises. This survey is led on a representative sample of 5,256 enterprises that are part of the population of industrial enterprises with 20 workers or more. This is not an annual survey but is carried out every 4 years. 4) *Information and Communication Technologies (ICT)* survey tries to gather information on enterprises' use of information and communication technologies and electronic commerce, in order to highlight "new economy" activities. Enterprises with 10 or more workers in the manufacturing sector and in part of the services sector are the reference units.

2.2 Integration Issues

At this level three main issues can be distinguished: (1) the reconciliation of survey data with administrative data, (2) the treatment of missing data and (3) the problem of sample weights. *X* are auxiliary variables and represent the matching variables, *Z* are profit and loss variables which are present in survey and commercial account data, *Y* are asset variables which are not surveyed in *PMI.V* are other variables describing various features of the enterprises and *P* are the Prodcom variables. *F* are the fiscal variables.

[9] cf. ISTAT 2001a.
[10] cf. ISTAT 2002.

Chart 2 – Integration scheme: sources, units and variables



Legend:
←→ Exact matching (one to one)
← Statistical matching (similar to similar)
$\underline{X}$ = Matrix register (4 million of firms)
$\underline{Z}^r$ = Matrix profit & loss of SCI and PMI surveys that are combined in the regional dataset (54,000 units)
$\underline{Y}^r$ = Matrix assets & liabilities of SCI survey dataset (roughly 9,000 rows)
$\underline{V}^r$ = Matrix employment and other variables (SCI and PMI surveys are combined in the regional dataset)
$\underline{Z}^c$ = Matrix profit & loss of corporate dataset (sample of 54,000 corporations)
$\underline{Y}^c$ = Matrix assets & liabilities of corporate dataset (sample of 54,000 corporations)
$\underline{F}$ = Matrix of the sections RF RN RJ RU RS of fiscal returns declaration (tax receipts fiscal datasets)
$\underline{P}$ = Matrix of Prodcom datasets
$\underline{V}^o$ = Matrix of other datasets
$\{\underline{X}^{-t}, \underline{Z}^{-t}, \underline{Y}^{-t}\}$ = Matrices with retrospective information (t= 1996, 1997, 1998)

As it can be seen in the chart, information from administrative archives is appended to statistical information of the regional dataset (SCI and PMI). When information from different sources is combined, harmonization problems arise. For that a huge work in reconciling metadata information (record units, classifications, name and meaning of variables) is done.
In some cases the exact matching (←→) has not been possible, so the problem of lack of information is solved in two ways with statistical matching procedures (←):

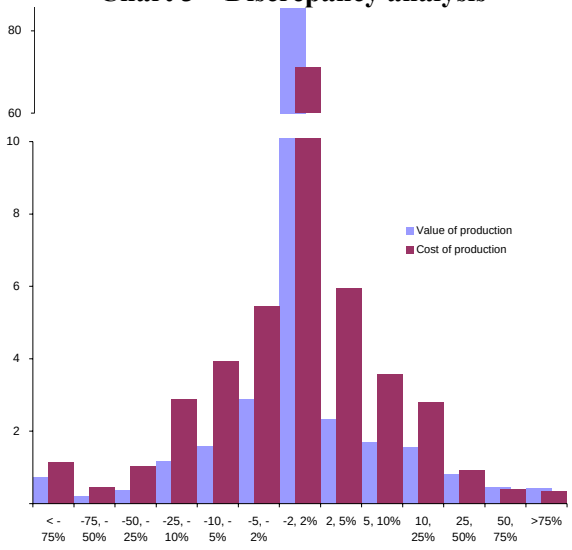
1) imputation through model based or nearest neighbour technique (the donor unit is chosen among the similar unit with complete data; similarity is calculated by a distance function)[11];
2) cell based imputation (where the cell is a specific domain of analysis and represent an aggregation of similar units)

After the harmonization and the reconstruction of $\underline{\boldsymbol{Y}}$ variables not present in surveys a discrepancy variable is calculated. It allows highlighting some units which are not coherent across the different sources and for this reason they are deleted.

2.2.1 Reconciling Survey and Administrative Data

The harmonization of variable definitions has been, preliminarily, required in order to produce metadata information. When it has been possible, the same variables from different sources have been compared. An example on two important variables is illustrated in the chart below. The comparison of the values of the variables included in both administrative and regional datasets shows a regular distribution of relative differences. The frequency in each class is equal to the number of units with the difference Z_{PMI}-Z_{CA} included in a percentage range of the value of Z_{PMI}.

Chart 3 – Discrepancy analysis



[11] cf. ISTAT 2002c.

With regards to the first variable roughly 86 percent of PMI under observation have a discrepancy range of ±2 percent, while with regards to the second variable roughly 71 percent of the enterprises fluctuate between ±2 percent.
Some information lacking in the surveys has been reconstructed by using administrative sources. Reconstruction is performed for 11,905 units exactly matched with administrative data (that is the 78% of PMI corporate firms). Starting from accounting variables in the surveys, and adding up new information from administrative sources, some variables has been recalculated according to compatibility rules of the balance sheet scheme. At the end the profit or loss of a single unit was recalculated. The relative difference between the initial value and the value after the reconstruction process is shown in the table below. Records of the extreme classes (1st and 13th classes) have excessively incoherent values. They represent 0.5%. These units have been considered as having missing data and they have been excluded from the analysis.

Table 4 – Discrepancy analysis after the reconstruction of information through administrative data

Cl.	Range	# records	Percentage value
1	*<-100%*	*44*	*0.4*
2	-100, -50%	55	0.5
3	-50, -25%	157	1.3
4	-25, -10%	392	3.3
5	-10, -5%	467	3.9
6	**-5, -2%**	**664**	**5.6**
7	**-2, 2%**	**8449**	**71.0**
8	**2, 5%**	**771**	**6.5**
9	5, 10%	473	4.0
10	10, 25%	323	2.7
11	25, 50%	76	0.6
12	50, 100%	26	0.2
13	*>100%*	*8*	*0.1*

2.2.2 Statistical matching

When exact matching is not possible the reconstruction of missing information at a micro level can be performed by linking the record with missing values with a similar record having complete information. This kind of problem has been treated as an imputation problem[12]. The table below shows on which basis missing data has to be imputed: being 1999 the year of reference and taking the corporate firms of the Regional dataset (15 thousand on a total of 46 thousand firms), 78% records are linked with the information contained in the administrative data sources (table 4 and table5). For the units for which no link has been possible, information has to be reconstructed and missing data has to be imputed (22% of PMI not linked with CA).

Table 5 – Results of merging activities

Cross Sectional integration (year 1999)				
Year 1999	*Regional (99)*	*of which corporates (legal form=2)*	*Corporates linked with CA*	*Corporates not linked with CA*
PMI	45,867	15,330	78%	22%
SCI	8,734	7,340	95%	5%
Total	54,601	22,670	83%	17%

		Longitudinal integration of which corporates		
Years 1999-1998	*Regional (99)*		*Corporates linked with CA*	*Corporates not linked with CA*
PMI	45,867	15,330	78%	22%
SCI	8,734	7,340	95%	5%
Total	54,601	22,670	83%	17%

			Linked with previous surveys datasets	*Not linked with previous survey datasets*
Years 99-97-96	*Regional (99)*	*of which corporates*		
PMI	45,867	15,330	26%	74%
SCI	8,734	7,340	91%	9%
Total	54,601	22,670	47%	53%

Regarding the reconstruction of the missing data, two approaches have been experimented[13]: (i) a non-parametric approach based on a distant function (ii) and a parametric approach based on a multiple regression model. In the two

[12] cf. D'Orazio, Di Zio and Scanu, 2001.
[13] cf. Oropallo, Inglese 2004.

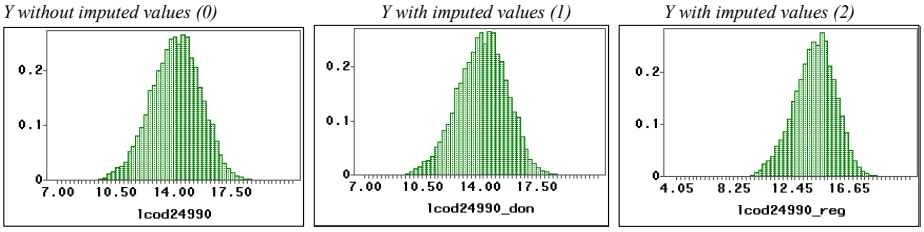
cases the analyzed variable is the total assets of the enterprise (*Y*). It has missing values in the case of the *PMI* records not linked with the *CA* dataset (dotted region in chart 2). In fact *PMI* has fewer variables than *SCI* and the assets variables (*Y variables* of the chart 2 = *fixed assets, current assets, net assets, debts and liabilities*) are considered missing data.
In the parametric approach has been utilised a distance based method: the *donor* technique[14]. In this approach units with incomplete data (*host* dataset) is reconstructed through units with complete data (*donor* dataset) by using a distance function calculated on matching variables. These are two quantitative variables: X1=value of production, X2=total employment and two qualitative variables: X3=localisation and X4=business sector.
In the second approach[15] the value of the missing datum has been estimated by applying a parametric method based on multiple regression model. Two explanatory variables selected after the analysis of the Pearson correlation coefficients and two dummies define the model. This model has been tested considering successively a model with all the variables in a linear form and a logarithmic model with two logarithmic variables and two dummies. Statistic tests identify the best models and sensitivity analysis methodologies calculate the data quality indicators for the two different approaches adopted to 'model incomplete data'[16]. The best multiple regression model are Y=f(X1,X2,X4) and Y=f(X1,X2,X3,X4), where Y=total asset, X1=value of production, X2=total employment, X3=localisation, and X4=business sector (the continuous variables are transformed in logarithmic values).

Comparison of the two approaches

Finally the distribution of the variable *Y* has been reconstructed in the two mentioned approaches: *(1) non-parametric*, *(2) parametric* and compared with the previous distribution without imputed values *(0)*.

[14] cf. Abbate 1997 and cf. Istat 2002c.
[15] cf. Nardo M. et al., EC-JRC 2003.
[16] cf. Rubin, 1987.

Chart 4 – Comparison of different approaches in the data imputation process (distributions of variable Y=total assets – log values)



Distribution of Y	*N*	*Mean*	*Variance*	*CV*	*Kurtosis*	*Skewness*
0	11,922	14.16	2.27	10.63	0.10	- 0.11
1	15,265	14.17	2.30	10.71	0.08	- 0.15
2	15,351	14.16	2.35	10.82	0.34	- 0.29

With donor imputation (1) the shape of the distribution has been preserved: the kurtosis and the skewness indexes are similar to the previous distribution with incomplete data. Thus with the imputation through multiple regression model (2) there is a higher kurtosis and negative skewness, but the mean value of the distribution has been maintained. Considering that the objective of the reconstruction process is first to keep up the distribution properties of a variable, the donor technique is preferred in this experimentation.

2.2.3 Weight Adjustment

To obtain reliable estimates from the microsimulation model it is necessary to employ weight adjustment techniques. For example, for the simulation of the Corporate Tax, a subset of the sample of small and medium enterprises has to be considered according to the legal form of the enterprise. At Istat the sampling and the estimation strategy does not consider the legal form of the units: auxiliary variables usually used for sample stratification are the business sector (NACE, four digits), geographical areas (Regions) and employment classes. In our case there is a need to add a new dummy variable, corporate/non corporate firm. The statistical methodology used is the Generalized Estimation Method[17], where calibrated estimators are applied. Weights are adjusted through the minimization of a distance function between the initial and final weights. The distance function is subject to these

[17] cf. Falorsi P.D., Pallara A. et al., 2000.

boundary conditions: 1) the sum of weighted firms in each stratum has to coincide with the sum of the population of firms in ASIA; 2) the sum of weighted employment in each stratum has to coincide with the sum of employment of the population in ASIA[18]. First results are produced by considering the dummy variable *corp* (0 = non corporate; 1 = corporate) and only one constraint, i.e. the total number of enterprises. At this stage, a first re-weighting procedure has been run, in order to correctly re-calculate sample weights, so that the sum of corporate weights is equal to the number of corporate enterprises[19].

Table 6 – Results of the calibration process

Legal type	Sum of observations	Sum of initial weights	Sum of final weights	Calibration Coefficient
Sole Proprietorship	16,696	2,804,793	2,787,100	0.9937
Partnership	11,761	762,813	758,004	0.9937
Corporate (corp=1)	19,158	354,036	468,867	1.3243
Co-operatives	2,833	40,460	40,231	0.9944
Public	333	5,153	5,122	0.9940
Others	269	5,952	5,919	0.9943
Overall dataset	51,050	3,973,206	4,065,243	1.0232

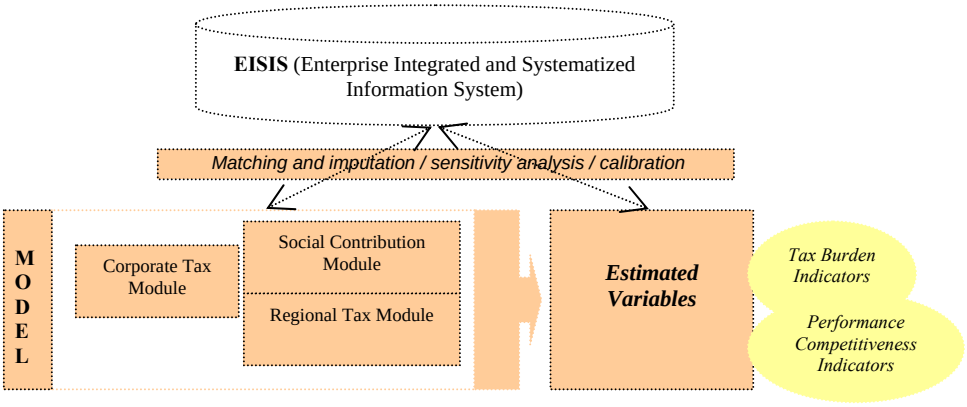
The table shows the result of the correction of the sample units to comply with the universe of corporations of the ASIA archive. The initial weight is multiplied by a coefficient of calibration to reach more reliable final weights, when classifications according the firm's legal type are produced. The last column gives the average calibration coefficient by firm's legal type.

[18] cf. Estevao, Hidiroglou, Sarndal, 1995.
[19] When there are a higher number of constraints, the optimization procedure has managed with the GENESEES software (cf. Istat 2002d).

3 Micro Analysis

The final result of the integration process is the overall dataset, which is representative of the universe of enterprises. Data marts are extracted from this database to serve fiscal microsimulation analysis and to produce systemic analyses.

Chart 5 – The Overall Structure

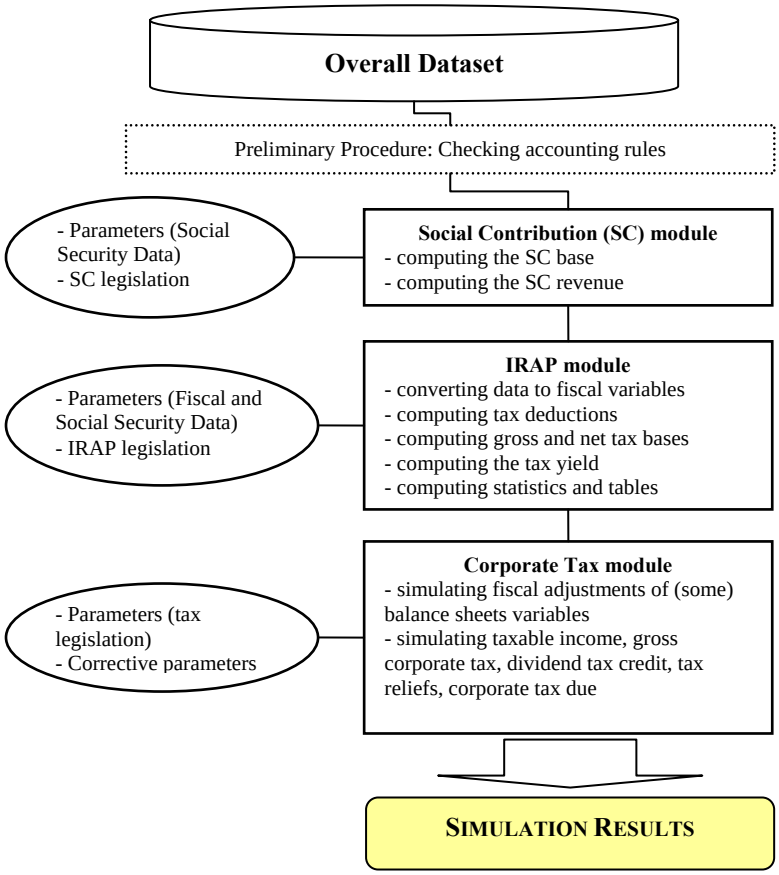


From the integrated data base the model should estimate the taxable yield for every type of tax. Even if ISTAT surveys do not cover all data needs for a precise calculation of taxable income or the voices of deductions from taxable income and of deduction of the tax, parameter data sets have been used to calibrate account data and to target them to aggregate results of the Italian revenue agency, available for the year 1998. Nonetheless it is possible to make hypothesis on the behaviour of enterprise on to select proxy variables.

3.1 The DIECOFIS Microsimulation Model[20]

There are three modules that calculate three types of taxes. It is structured to simulate hypothetic variation of tax legislation. The three modules are the Social Contribution Module, The Regional Tax Module and the Corporate Tax module.

Chart 6– The Microsimulation Model



[20] In short, "...the microsimulation approach to evaluating alternative legislative proposals involves modelling the impact of government programs at the level at which they are intended to operate…" cf. Orcutt G. 1957.

The Social Contributions module calculates the contribution for social security according to professional category (manager, employees, etc.) and type of contract (formation, collaborations, etc.).

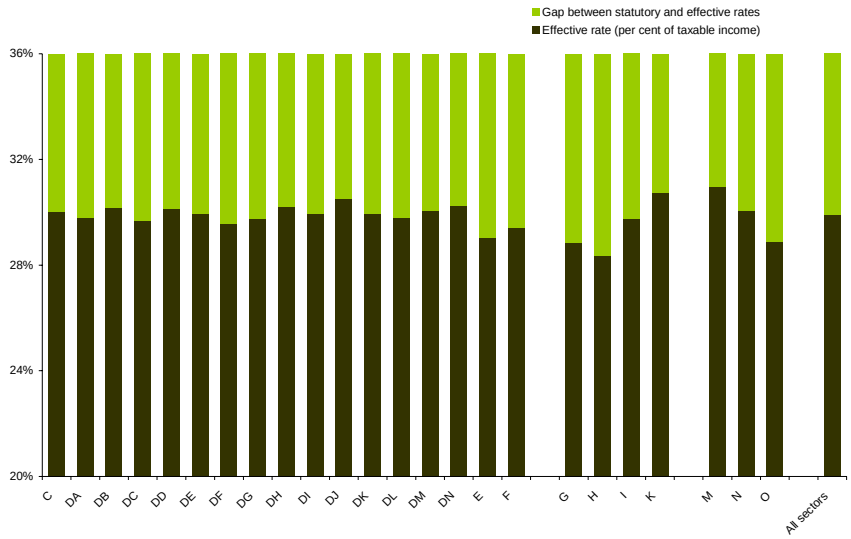
The calculation of the Regional Tax (IRAP) on Value Added is carried out in the following way:
$IRAP = tr\ (VP - CP - TO)$
Where: *tr* represents statutory rate of the regional tax. *VP* = Value of Production: Income from sales, variations of stocks, other income. *CP* = Costs of Production: Raw materials and consumables, other external charges, value adjustments, amortisation. *TO* = Other Deductions: INAIL (National Institute of Insurance Against Accidents at Work) Contributions, apprentices' costs, formation contract jobs and costs for disabled persons.

The Corporate Tax module calculates the tax burden on corporate enterprises. The equation for Italian corporate tax is the following:
$IRPEG = t_g(U + T_{IND.} + Cr + Crd - PP)$
Where: t_g is the legal rate, U is the profit, T_{IND} is the amount of non deductible taxes, *Cr* tax credits in profit and loss scheme, *Crd* Tax credit to share dividend and *PP* the amount of loss brought forward. Chart 6 shows the structure of the microsimulation model[21].

Results of the analysis demonstrate the heterogeneity and complexity of the business community and its taxation. For example, as chart 7 makes clear, statutory tax rates are not a very good proxy for tax incidence analysis, both within and across business sectors. The gap between statutory and effective tax rates can be not only large, but have a pattern characterized by widespread differences in tax incidence across groups and areas, even when rates are uniform and proportional. Investigating the importance and the factors behind these differences is essential, if analyses and comparisons are to be meaningful and revealing.

[21] Bardazzi, R., F., Pazienza, M.G., Parisi, V. (2003), The Effects of the Italian Tax Reform on Corporations: a Microsimulation Approach. http://www.istat.it/diecofis.

Chart 7 - Statutory and Effective Rates by business sectors[22]



The choice of tax indicators is important to bear out the real incidence of tax systems. Tax burdens may be calculated with different tools and for different aims. The first rather obvious measure to consider is the statutory tax rate, which gives a general idea of national tax policies. It is widely used in international comparisons, however it does not give a reliable measure of the actual tax burden, especially in international, dimensional and intersectoral contexts, as the actual tax burden closely depends on the definition of the tax base which is far from being standardized. For this reason *statutory tax rates* are often replaced by *effective tax rates* that take into account how the tax base is determined and whether tax laws provide for tax incentives. Within effective tax rates it is then possible to distinguish ex-post (backward looking) and ex-ante (forward looking). The backward looking approach includes all those indicators calculated as the ratio between taxes actually paid and a reference economic aggregate (e.g. profits, capital, value added, etc.). Using

[22] Te table shows the shares of enterprises and employment of the sample of data extracted from the Diecofis Database and utilized for the analysis.

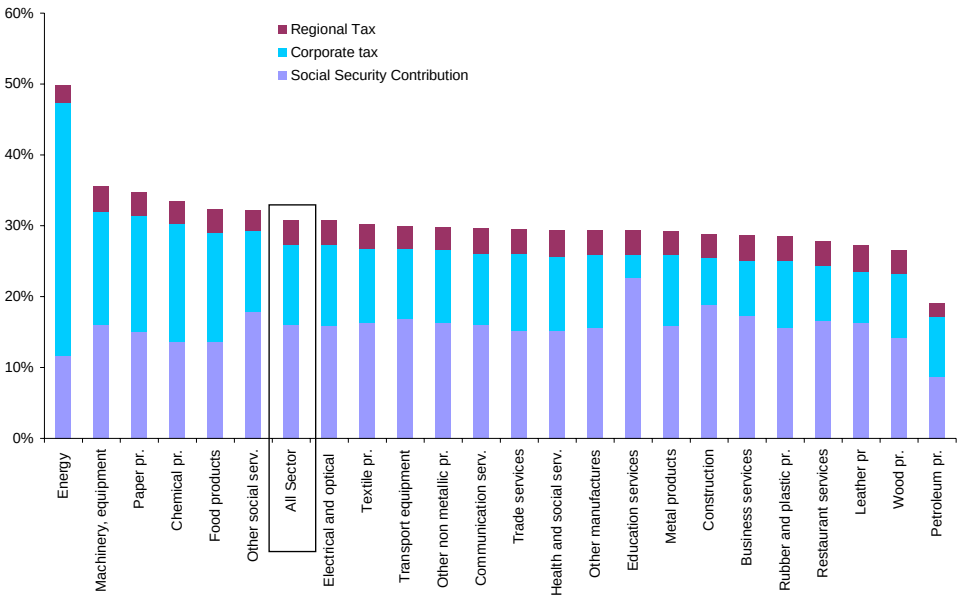
NACE	*Description*	*Firms %*	*Empl.. %*
C	Products from Mining and Quarrying	0.2	4.6
DA	Food Products, Beverages and Tobacco	4.5	2.9
DB	Textiles and Clothing Industry Products	8.6	2.1
DC	Leather and Leather Products	2.2	1.7
DD	Wood and Products of Wood and Cork (Except Furniture)	0.8	1.4
DE	Pulp, Paper and Paper Products; Recorded Media; Printing Services	3.5	2.4
DF	Coke, Refined Petroleum Products and Nuclear Fuel	0.4	6.3
DG	Chemicals, Chemical Products and Man-Made Fibres	4.9	3.6
DH	Rubber and Plastic Products	3.5	2.3
DI	Other non Metallic Mineral Products	3.6	2.5
DJ	Basic Metals and Fabricated Metal Products	9.1	2.3

NACE	*Description*	*Firms %*	*Empl.. %*
DK	Machinery and Equipment N.E.C.	10.7	2.7
DL	Electrical and Optical Equipment	6.6	3.6
DM	Transport Equipment	4.0	6.9
DN	Other Manufactured Goods N.E.C.	2.9	1.8
E	Electrical Energy, Gas, Steam and Water	0.7	20.4
F	Construction Work	3.9	2.1
G	Wholesale and Retail Trade Services	8.8	3.3
H	Hotel and Restaurant Services	2.4	3.6
I	Transport, Storage and Communication Services	4.7	13.2
K	Real Estate, Renting and Business Services	9.4	2.4
M-N-O	Education, Health and Social Services	4.6	6.0

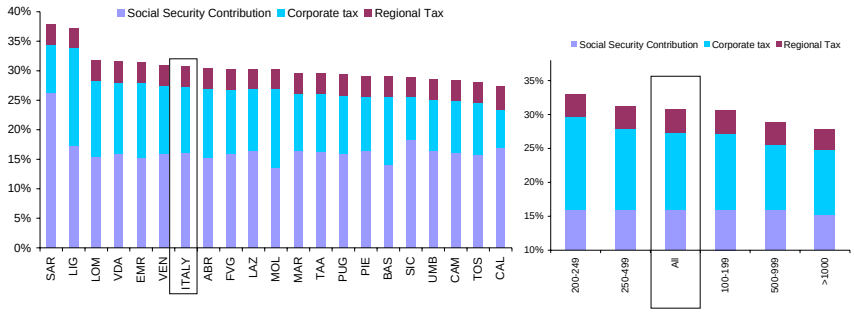
taxes actually paid allows the analyst to summarize the specific rules underlying the determination of the tax base and the specific choice of the firms. In this perspective, they are useful when the aim of the analysis is to examine the income effects of taxation and compare the taxation levels of different groups of firms. The second approach has a forward looking nature. It measures a theoretical incidence starting from tax laws applicable to a specific investment project at the margin, i.e. not producing extra-profits. This kind of measurement is useful to understand how tax systems affect firms' decisions and to understand also the biases (non-neutralities) of tax systems between different alternative investments.

In this work, *ex post implicit tax rates* are estimated in order to determine whether there are significant differences in the tax burden of companies with different characteristics, looking for evidence of discrimination between firms operating in different sectors, regions or having different size. Furthermore empirical results seem to suggest that the analysis of tax incidence can be of limited relevance, if based on summary indicators of tax burdens. What they show may not only be distorting but flawed.

Chart 8 - Total Tax Takes by Business Sectors, Regions and Size[23]
(as a percentage of Total Value Added)



[23] Include the Corporate income tax, the regional business tax and employers' social security contribution.



Another problem with traditional summary tax ratios is that they do not necessarily depict the full tax picture. Governments can make choices and use different instruments to collect their revenue. Whenever more than one tax instrument or different instruments are used to tax different types of enterprises (e.g. the Corporate tax, the Personal tax, the Regional tax and Employer's social security contributions), looking at one tax dimension only, leads to biased comparisons. Microsimulation models permit to overcome these shortcomings as they make it possible to add up different taxes and to estimate total *net* tax takes. The ensuing tax ratios, as shown in Chart 7 present a picture that is by no means similar to the one shown in Chart 8.

3.2 Enterprise Performance and Tax Burden

After a first excursus on the complexity of the tax burden analysis, a distributional or systemic analysis of enterprises performance is proposed. This kind of analysis starts from the distributional study proposed by the World Bank for poverty analysis. In this context it has been employed to study the performance of enterprises.

It is known that the GINI index[24] measures the concentration of a particular phenomenon (0=no concentration, 100=maximum concentration). The total measure of concentration can be divided into three elements[25]:

$$G = within + between + overlap\,.$$

---

[24] Gini Index formula:

$$G = \frac{(N+1)}{N}\left(1 - \frac{\mu_p}{\mu}\right)$$

con $\mu_p = \frac{\sum_i (N-r+1)y_r}{N + N(N-1)/2}$ e $\mu = \frac{\sum_r y_r}{N}$

μ_p *and* μ show respectively: the mean weighted with the position in the ranking (r = 1 to N) according to the "y" variable and the general simple mean value.

The index satisfies the axioms of: (1) Symmetry (anonymity); (2) Independence from the mean (scale); (3) Pigou Dalton's transfer principle.

[25] Decomposition is referred to Branko Milanovic *"True world income distribution, 1988 and 1993: First calculation based on household surveys alone"* – World Bank, Development Research Group.

$$G = \sum_{k=1}^{K} G_k p_k \pi_k + \frac{1}{\mu} \sum_{k}^{K} \sum_{k>i}^{K} (y_k - y_i) p_k p_i + L$$

p_k : Weight of the k^{th} class;

π_k : Share of the observed variable for the k^{th} class;

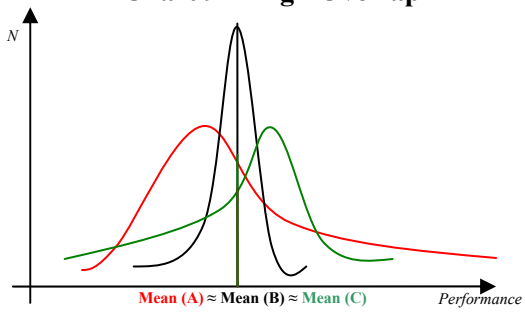
y_k : Mean value of the observed variable in the k^{th} class;

μ : Overall mean value of the observed variable.

The within element measures the concentration level into the group k. The between part measures the contribution to the total inequality of dissimilarities among groups. The overlap component represents the residual part and it is the result of the confusion of the two combined effects: within and between.

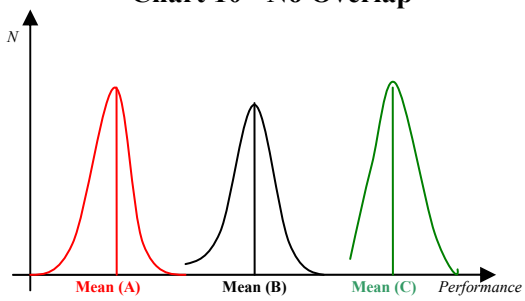
To explain the overlap effect, three classes can be supposed for the decomposition process. In a situation with a high overlap the queues of the distributions overlie:

Chart 9 – High Overlap



Thus when they are separated there is no overlap:

Chart 10 - No Overlap



Within and between components represent each one the sum of the single contribution of every class. Classes that have higher within part refer to

classes with higher internal concentration. The class with the higher between part refers to the class with a high-level mean value of performance.
The total Gini index and its decomposition are shown in the tables below. Three dimensions of enterprise performance have been considered (Export, Employment and Value added). The concentration level of each dimension is analysed through the most important classification variables (they represent the stratification variables and are included in the Business Register data): NACE sector, Region and size in term of workers.

Table 7 - Gini index and its components

Gini - Export				
Classes	*Total*	*Overlap*	*Within*	*Between*
NACE	83.2	33.6	4.9	44.7
Region	83.2	50.1	12.1	21.0
Size	83.2	14.9	11.9	56.4

Gini - Employment				
Classes	*Total*	*Overlap*	*Within*	*Between*
NACE	63.8	33.6	5.0	25.2
Region	63.8	27.7	7.7	28.4
Size	63.8	0.0	3.3	60.5

Gini - Value Added				
Classes	*Total*	*Overlap*	*Within*	*Between*
NACE	73.2	38.5	5.7	29.1
Region	73.2	28.8	9.4	35.0
Size	73.2	6.0	7.9	59.4

Other variables used to discriminate among enterprises are concerned with innovation features. It can be observed that *export* inequality is very high between enterprises with or without *innovation and R&D (26.8)*.

Table 8 - Gini index and Innovation features

Gini - Export					**Avg. Values (thous. of euro)**	
Classes	*Total*	*Overlap*	*Within*	*Between*	*Yes*	*No*
Product Innovation	83.2	22.1	39.1	22.1	35154	14266
Process Innovation	83.2	24.6	39.8	18.9	33689	15662
Innovation and R&D	83.2	9.8	46.6	26.8	70146	15939
Gini - Employment					**Avg. Values (thous. of euro)**	
Product Innovation	63.8	14.9	29.8	19.2	774	356
Process Innovation	63.8	14.2	29.4	20.2	794	350
Innovation and R&D	63.8	11.6	44.2	8.0	855	483
Gini – Value Added					**Avg. Values (thous. of euro)**	
Product Innovation	73.2	18.6	34.6	20.0	51386	22775
Process Innovation	73.2	18.5	34.4	20.4	52149	22782
Innovation and R&D	73.2	13.2	51.3	8.7	57848	31327

3.2.1 Performance Indicator

When more variables are considered, each one constitutes a particular aspect of enterprise performance. In the general case of m variables, several measures of enterprise performance can be obtained. Considering the Gini properties the between values, deriving from the m decomposition processes, are comparable to each other. A composite indicator can be calculated equal to the mean of all between values for each class.
For each variable or dimension it can be written:

$G^d = w^d + b^d + L^d$ d=1…m

The between value of the d^{th} variable can be decomposed in this way:

$$b^d = \sum_{k=1}^{K} b_k^d \qquad \text{d=1…m}$$

For each class k=1…K it can be written:

$$b_k = \sum_{d=1}^{m} b_k^d$$

It synthetically represents the sum of performance values for each class. Each performance value is a dimension of the overall performance of the enterprise.

The sum of b_k is equal to:

$$b = \sum_{k=1}^{K} b_k = \sum_{k=1}^{K} \sum_{d=1}^{m} b_k^d \qquad 0 < b < 100m$$

It represents the sum of all the between components across the classes and the dimensions. Its range is 0 (no classes inequalities for all "dimensions") and 100m (for all "dimensions" it is only present a inequality among classes with no diversities within the classes and with no overlap effects).

Finally each b_k^d, Gini's between of the d^{th} dimension and of the k^{th} class can be normalized by dividing each one with the maximum of each dimension:

$$bn_k^d = b_k^d \Big/ Max(b_k^d) \quad d = 1,..m$$

The final idea is to gather several performance rankings with a unique ranking. In this case is proposed the mean value of the three dimensions:

Composite Indicator = Mean(GINI's between)[26]

[26] This indicator has the following properties: (1) micro-founded (2) purpose oriented (3) consistent (4) decomposable.

In general weights for each dimension can be introduced to establish value judgment:

$$C = \frac{1}{m}\sum_{d=1}^{m} w_d bn_k^d$$

$$with \quad 0 \le w_d \le 1 \quad and \quad \sum_d w_d = 1$$

Several measures of concentration can be obtained to analyse multidimensional aspects of performance. The variables considered are: (i) the value added, which is the remuneration of production factors; (ii) the employment, which is the total number of persons employed; (iii) the total value of exports of enterprises. For each variable different concentration levels can be observed and through the "decomposition process", different performance rankings can be seen through the groups (between indicator) accordingly to the utilised variable or "dimension" of enterprise outcomes.

Table 9 - Different rankings and the Composite Indicator (by business sectors)

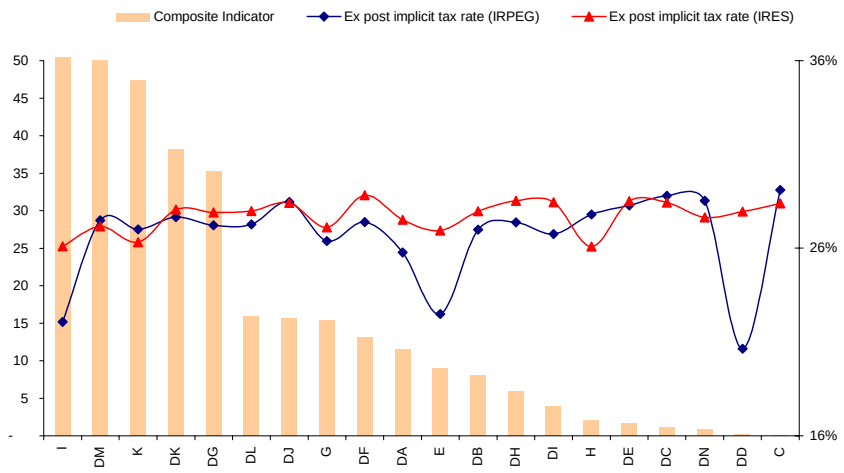
Sector	*Export*	*Sector*	*Employment*	*Sector*	*Value Added*	*Sector*	*Composite*
DK	100.0	I	100.0	K	100.0	I	51.4
DM	92.9	K	41.6	I	48.5	DM	50.1
DG	76.9	DM	31.0	G	36.4	K	47.3
DL	42.6	DK	7.4	DF	28.2	DK	38.2
DJ	36.4	DG	7.4	DM	26.4	DG	35.2
DB	22.2	G	6.5	E	24.1	DL	16.0
DH	15.6	DA	6.3	DG	21.5	DJ	15.7
DA	14.5	H	5.9	DA	13.6	G	15.5
DF	10.9	DJ	3.7	DK	7.3	DF	13.1
DI	8.0	E	2.7	DJ	7.0	DA	11.5
I	5.7	DL	2.7	DI	2.8	E	8.9
G	3.4	DB	1.2	DL	2.6	DB	8.0
DC	3.4	DH	0.8	DE	1.7	DH	6.0
DN	2.6	DI	0.8	DH	1.5	DI	3.9
DE	2.6	DE	0.4	DB	0.7	H	2.1
DD	0.5	DF	0.1	H	0.4	DE	1.6
K	0.5	DN	0.0	C	0.2	DC	1.1
C	0.0	DD	0.0	DD	0.0	DN	0.9
H	0.0	DC	0.0	DN	0.0	DD	0.2
E	0.0	C	0.0	DC	0.0	C	0.1

3.2.2 Taxation and Economic Performance

The view behind existing summary tax indicators is that they can be effectively paired with indicators of economic performance and interpreted on the basis of convictions, such as that a negative relationship exists between tax burdens and performance. The high level of aggregation and

heterogeneities of various data and often bar the inferences that can be drawn from them. Moreover, averages always hide the dispersion in the distributions from which they are calculated. Summary indicators are no different.

Chart 11 - Performance and Tax Burden Simulation by business sector



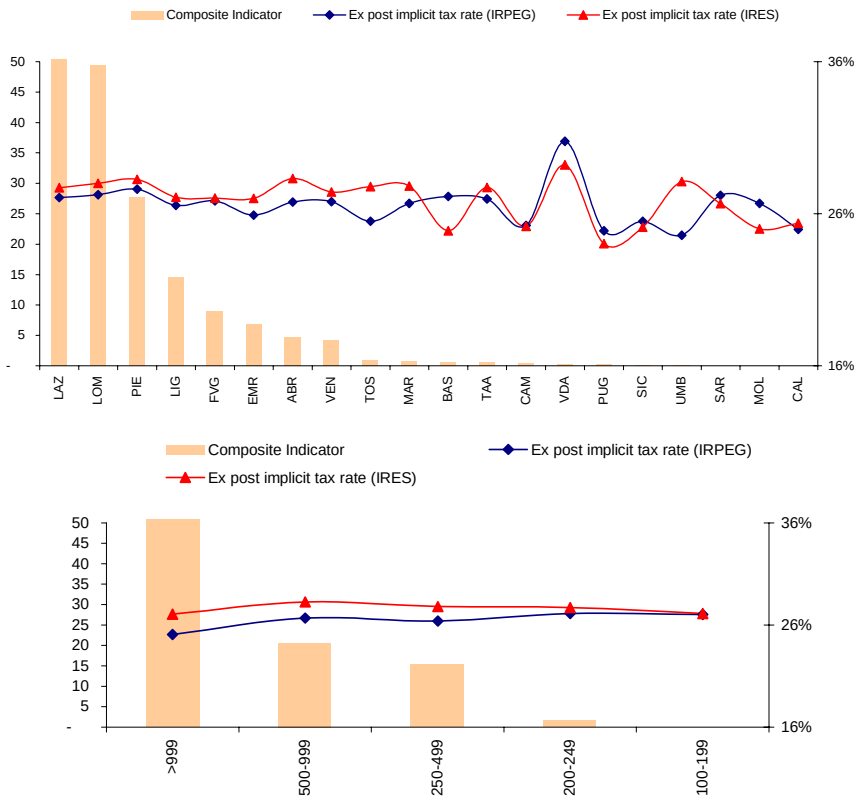
Diecofis has helped to progress on indicators and to shed new light throw new insight into existing relationships. It has made it possible to calculate sets of micro-founded indicators of impact and performance, characterized by high levels of homogeneity and precision (since they refer to the same populations). In turn, this has helped to finely map enterprises and their performance by, sectors, regions, size and so on, using sets of elementary and multidimensional (composite), indicators. As can be seen from Chart 11 and 12, Diecofis has paraded "dwarf" and "giant" enterprises (or groups of them), lined up from top to worst performers (with performance measured by means of a three-dimensional and composite indicator for NACE sector, and enterprises classified by region and size). Since the tax burden for each enterprise in the parade was known, the analysis plots rather precise maps and indications as to how the latter may relate to the former.

In addition Chart 11 and 12 contain two different scenarios of corporate taxation: The base scenario (IRPEG) and the IRES scenario[27]. The first one simulates the former corporate tax system; the second one describes the new tax reform starting from 2004. In the simulation analysis it is possible to identify which business sector, or which region or which group size lose or win following the simulation of the new tax system. One important

[27] cf. Bardazzi R., Pazienza M.G., Parisi V. (2003).

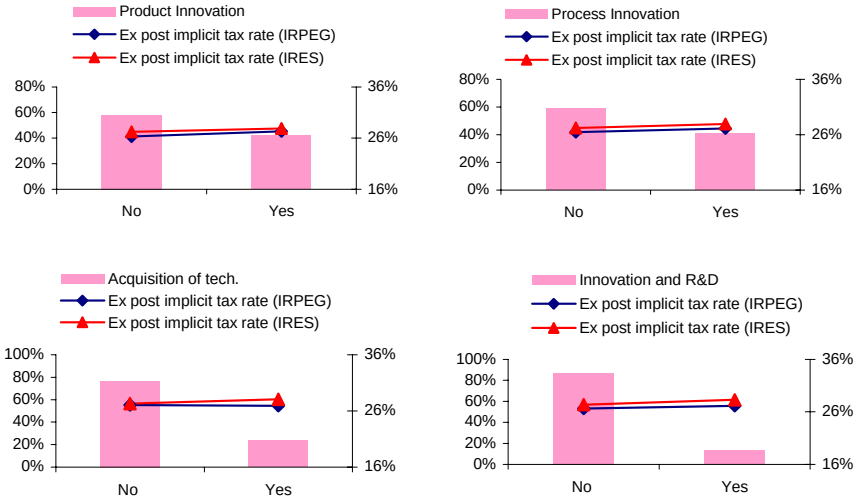
conclusion that can be drawn from these first results is that the reduction of statutory rates (from 36 to 33%) and the abolition of the dual income tax (lower rate 19% for reinvested income) should produce an increase of the tax burden in some cases. Firms operating in sectors which benefited most from the previous dual income taxation will bear most of the reform cost in terms of higher implicit rate (i.e. E="Electricity, gas and water supply").

Chart 12 - Performance and Tax Burden Simulation by Region and by Size



The next chart examines the response shares (yes/no) for each innovation feature of the enterprise: *product innovation, process innovation, acquisition of technology and research & development*, adding the effective tax burden in the two analyzed scenarios: *Irpeg (the former corporate tax system) and Ires (the reform scenario).* A reduced tax burden can be remarked for those enterprises which make process/product innovations, in other cases the situation should not change. Moreover it is very interesting to observe that a very small share of enterprise acquire new technologies and invest in R&D.

Chart 13 – Innovation Features and Tax Burden Simulation



4 Conclusions

The integration of all available information on enterprises into one multi-sources database performed in the framework of the Diecofis project gives the system high potential and opportunities in terms of economic analysis. It satisfies the data requirements of the microsimulation model for the estimation of the effects of fiscal policies on enterprises' performance and permits the investigation of more economic issues that can be analysed in a very detailed perspective through new demand driven methods. The use of a systematised and integrated system makes it possible to create new micro-founded indicators that are more appropriate to describe different economic systems and to understand their systemic strengths and weaknesses.

These first outcomes put out the potentiality of the microsimulation analysis, but many issues will be considered in future developments: adding up data sources of recent years, extending the microsimulation model to small enterprises, reconstructing reliable retrospective information for small enterprises by matching at a stratum aggregation level, simulation of consumption taxes (e.g. energy taxes) and cross country simulations. In fact one of the aims of the model is also to facilitate international comparisons of the business sector and its taxation. The Italian contribution has approached the issue from the perspective of a producer of data. Its research work has relied on both statistical surveys and administrative sources, including tax authority information. At the same time it has confirmed the difficulty of making direct comparisons across countries, since approaches are

heterogeneous and databases are far from common. In addition, the composition of the business sectors is quite different and tax systems are not easily comparable. Comparisons of different tax systems on a typical business basis approach are possible, but they are believed to be uninformative, since many tax heterogeneities may be hidden behind them. Further work is needed to develop the multi-sourced data and modelling described above in several countries so that, for example, detailed analysis of effective tax rates for large companies can be undertaken or, for small companies, the relationships between tax on directors remuneration and company profits can be compared. The analysis of the sources of heterogeneity across groups and countries is important and should be factored in developing integrated databases and models.

5 References

Abbate C. (1997), "Completeness of Information and Imputation from Donor with Minimum Mixed Distance", Quaderni di Ricerca ISTAT, n. 4/1997, pp. 68-102

Bardazzi, R., F., Pazienza, M.G., Parisi, V. (2003), "The Effects of the Italian Tax Reform on Corporations: a Microsimulation Approach" presented at the Public Finance and Financial Markets Conference organised by the International Institute of Public Finance (Prague August 2003) http://www.iipf.org/past.html

Black, J. (2001), Changes in Sampling Units in Surveys of Businesses, In: 2001 FCSM Research Conference Papers, US Census Bureau

Brick J.M., Kalton G. (1996), "Handling Missing Data in Survey Research", Statistical Methods in Medical Research, vol. 5, pp. 215-238

Brackstone, G. (1999), "Managing Data Quality in a Statistical Agency", Survey Methodology, December 1999, vol. 25, pp. 139-149

Denk M., Oropallo F. (2002), Overview of the Issues in Longitudinal and Cross-Sectional Multi-Source Databases, DIECOFIS Deliverable 1.1 www.istat.it/diecofis/deliverable_list.htm

Deville J. C., Särndal C. E. (1992), "Calibration Estimators in Survey Sampling", Journal of the American Statistical Association, vol. 87, pp. 376-382

D'Orazio, M., Di Zio, M., Scanu, M. (2001) Statistical Matching: a tool for integration data in National Statistical Institutes – paper ISTAT for NTTS 2001 – ETK 2001 – Crete Conference 18-22 June 2001 – Eurostat, JRC(ISIS)

Estevao, V., Hidiroglou, M.A., Sarnald, C.E. (1995) Methodological Principles for a Generalized Estimation System at Statistics Canada, Journal of Official Statistics 11 (2), 181-204

Eurostat (1999a) Use of Administrative Sources for Business Statistic Purposes: Handbook on Good Practices – Theme 4 (Industry, Trade and Services), Eurostat Edition

Eurostat (1999b), Model quality report in business statistics, Eurostat Working Group "Assessment of quality in business statistics"

Eurostat (2003), Business Demography in 9 member states, results for 1997-2000. Luxembourg: Eurostat, 2003. (Statistics in focus)

Falorsi, P.D., Pallara, S., Falorsi, S., Ballin, M. (2000) Il trattamento delle mancate risposte totali nelle indagini ISTAT sulle Famiglie e sulle Imprese (The Treatment of Total Non-Response in ISTAT) – ISTAT Methodological Studies

FCSM – Federal Committee on Statistical Methodology (1980) Report on Exact and Statistical Matching Techniques, Statistical Policy Working Paper 5, Washington, DC: U.S. Department of Commerce

Informer SA (2003), Concept of IT framework issues and development of software for the creation of a multi-source data base, DIECOFIS Work Package 2, http://www.istat.it/diecofis/deliverable_list.htm

ISTAT (1999) L'innovazione tecnologica nelle imprese (Firms' Technological Innovation) – Note Rapide – July 1999 (LE Survey) – http://www.istat.it/Imprese/Ricerca-e-/index.htm

ISTAT (2000) I risultati economici delle medio-grandi imprese Anni 1998-99 (Economic Performance of Medium-Large Size Enterprises) - Statistiche in breve - July 2000 (LE Survey) - http://www.istat.it/Imprese/Struttura-/index.htm.

ISTAT (2001a) Indagine Prodcom (Prodcom Survey) – Indagine sulla struttura dei Costi (Cost structure Survey) http://www.istat.it/Imprese-e-/index.htm

ISTAT (2001b) Struttura e competitività del sistema delle imprese industriali e dei servizi nel 1998 (Structure and competitiveness of industrial and service enterprise system in 1998). - Statistiche in breve - Luglio 2001 (LE & PMI Survey) http://www.istat.it/Imprese/Struttura-/index.htm.

ISTAT (2002a) Indagine sul Commercio Estero (Foreign Trade Survey). Current version available at http://www.coeweb.istat.it/

ISTAT (2002b) L'uso delle tecnologie dell'informazione e della comunicazione nelle imprese (The use of ICT in Italian firms) – Statistiche in breve http://www.istat.it/Imprese/Ricerca-e-/index.htm.

ISTAT (2002c) - CONCORD v1.0 - (Generalized Data Editing Software) CONtrollo e CORrezione dei Dati – ISTAT 2002 http://www.istat.it/Metodologi/index.htm

ISTAT (2002d) - GENESEES v1.0 - GENEralised software for Sampling Estimates and Errors in Surveys – ISTAT 2002 http://www.istat.it/Metodologi/index.htm

Jabine, T.B., Scheuren, F.J. (1986) Record Linkages for Statistical Purposes: Methodological Issues. Journal of Official Statistics 2 (3), 255-277

Kadane, J.B. (1978) Some Statistical Problems in Merging Data Files. In: 1978 Compendium of Tax Research, US Dept. of the Treasury, 159–171. (Reprinted in Journal of Official Statistics 17 (3), 423–433

Kalton G., Kasprzik D. (1986), "The treatment of missing survey data", Survey Methodology, vol. 12, n. 1, pp. 1-16

King M., Fullerton D. (1984), The taxation of income from capital, University of Chicago Press

Little R. J. A, Rubin D. B. (1987), "Statistical Analysis with Missing Data", Wiley & Sons, New York

Milanovic, B. (1999) “True world income distribution, 1988 and 1993: First calculation based on household surveys alone” – World Bank, Development Research Group

Nardo M., Kioutsioukis I. (EC-JRC 2003), Development of methodologies and of a software for the measurement of statistical quality, and for comparing the robustness of alternative multi-source, integrated databases, DIECOFIS Work Package 3, http://www.istat.it/diecofis/deliverable_list.htm

Orcutt G. “A New Type of Socio-economic System”, Review of Economics and Statistics, 80, 1957, pp.1081-1100.

Oropallo F., Inglese F., (2004), The Development of an Integrated and Systematized Information System for Economic and Policy Impact Analysis, “The Austrian Journal of Statistics” Vol 33/2004 N. 1&2 Special Issue, www.statistik.tuwien.ac.at/oezstat/.

Paass, G. (1986), "Statistical match: Evaluation of existing procedures and improvements by using additional information". G.H. Orcutt and H. Quinke (eds) Microanalytic Simulation Models to Support Social and Financial Policy. Amsterdam: Elsevier Science, pp. 401-422

Roberti P., Oropallo F., Inglese F., Lo Cascio L., de Martinis G. (2003) - Towards a Systemic Analysis of Italian Industrial Texture Review “Industria” 4/2002, Il Mulino

Renssen R. H. (1998), “Use of Statistical Matching Techniques in Calibration Estimation”, Survey Methodology, vol. 24, n. 2, pp. 171-183

Rodgers, W.L. (1984), An Evaluation of Statistical Matching. Journal of Business and Economic Statistics 2, 91-102

Rubin D.B. (1987), Multiple Imputation for Non-response in Surveys, Wiley & Sons, New York

Schafer J.L., Olsen M.K. (1998), “Multiple imputation for multivariate missing-data problems: a data analyst's perspective”, Multivariate Behavioural Research, vol. 33, pp. 545-571

Singh A. C., Mohl C. A. (1996), “Understanding Calibration Estimators in Survey Sampling”, Survey Methodology, vol. 22, n.2, pp. 107-115

Singh, A.C, Mantel, H.J., Kinack, M.D., Rowe, G. (1993) Statistical Matching: Use of Auxiliary Information as an Alternative to the Conditional Independence Assumption, Survey Methodology, June 1993- vol. 19, No.1 pp. 59-79 - Statistics Canada

Winkler W. E. (1995), “Matching and Record Linkage”, in B. G. Cox et al. (ed.), Business Survey Methods, Wiley & Sons, New York, pp. 920-935 (355-384)