Stefano Fachin Giuseppe Venanzoni

IDEM: an Integrated Demographic and Economic Model of Italy

Abstract

The aim of this paper is to give an overview of IDEM, a multi-regional model of the Italian economy. IDEM includes two main multi-regional blocks, a demographic model, based on the so-called Spatial Cohort-Component approach, and an economic model, based on a Multi-Regional Input-Output approach. The two blocks are linked by several modules, devoted *e.g.*, to productivity growth, regional labour markets participation and internal migration flows. The exogenous inputs of IDEM is essentially given by a small set of national economic and demographic variables, including respectively variables such as investment and *per-capita* consumption growth and fertility rates trends. The model is expected to be used for a variety of purposes such as regionalising national macroeconomic forecasts made by the Italian Treasury, or evaluating the economic impact of regional development programs and localised investment plans.

Keywords and phrases: Italy, Input-Output Model, Multi-regional Models, Population.

1. Introduction[1]

It is quite obvious that regional economic analysis and policy making require availability of a geographically disaggregated model of the economy. However, geographical disaggregation alone is not enough, as regions react differently to the same macroeconomic shocks because the structures of their economies are different. Thus, a model aiming at supporting economic analysis at regional level ought to be sectorally disaggregated as well. Further, a point which is usually overlooked (although stressed in the authorative survey by Isserman, 1986) is that careful modelling at regional level should treat demographic variables as endogenous: for instance, migration flows, influenced by economic conditions, may deeply and rapidly affect regional labour markets.

Summing up, ideally a regional economic model should satisfy three fundamental requirements: high geographical disaggregation; high sectoral disaggregation; endogenous modelling of the main demographic variables affecting labour markets. Accordingly, the Italian Treasury sponsored the development of a multi-regional model of the Italian economy, IDEM (Italy Demographic Economic Model), including: a multi-regional demographic model, based on the so-called Spatial Cohort-Component approach developed by Willekens and Rogers (1978); an economic model based on a Multi-Regional Input-output (MRIO) table; a set of links between the two main blocks, modelling interregional migration flows as a function of regional labour markets conditions. The model, still under development, is expected to be used for a variety of purposes akin to RIMS II's, BEA's regional Input-Output model (U.S. Dept. of Commerce, Bureau of Economic Analysis, 1981, 1986): e.g., regionalising the macroeconomic forecasts of the government medium-term economic plan (*Documento di Programmazione Economica e Finanziaria*, DPEF), or evaluating the regional economic impact of localised investment plans (*e.g.*, those associated with Turin 2006 Winter Olympics).

It should be stressed that the single parts of the model are all entirely standard (and in some cases, because of data limitations, admittedly highly simplified at the minimum). The main point of interest of the model is indeed given by the way the various assembled parts interact. In this paper we shall essentially provide a general introduction to the structure and single blocks of the model (section 2). Some general comments and directions for future research are reported in section 3.

[1] This paper stems from a project under development at Consip S.p.A.; however, the views expressed here are the authors' sole responsibility. Among the many individuals that at various stages have worked at the model we should express our special thanks to Elena Forconi, Lorena Urbani and Paola Vieri. Correspondance to: s.fachin@caspur.it, giuseppe.venanzoni@uniroma1.it.

2. Structure of the Model

As mentioned in the Introduction, IDEM is based upon two main demographic and economic blocks, linked by a set of models devoted to modelling productivity, regional labour markets and interregional migration flows. The structure is summarised in the flow chart reported in Fig. 1. The geographical units considered are the 20 Italian NUTS 2 areas (*Regioni*, hereafter simply regions), while the economy is disaggregated into 17 sectors (approximately corresponding to the NACE rev. 1.1 (SEC79) classification, with disaggregation at one digit for Agricolture and Services and two digits for the manufacturing sector. Detailed lists of regions and industries will be given below, respectively in Tables 1 and 3.

The three main steps of a forecasting run may be summarised as follows.

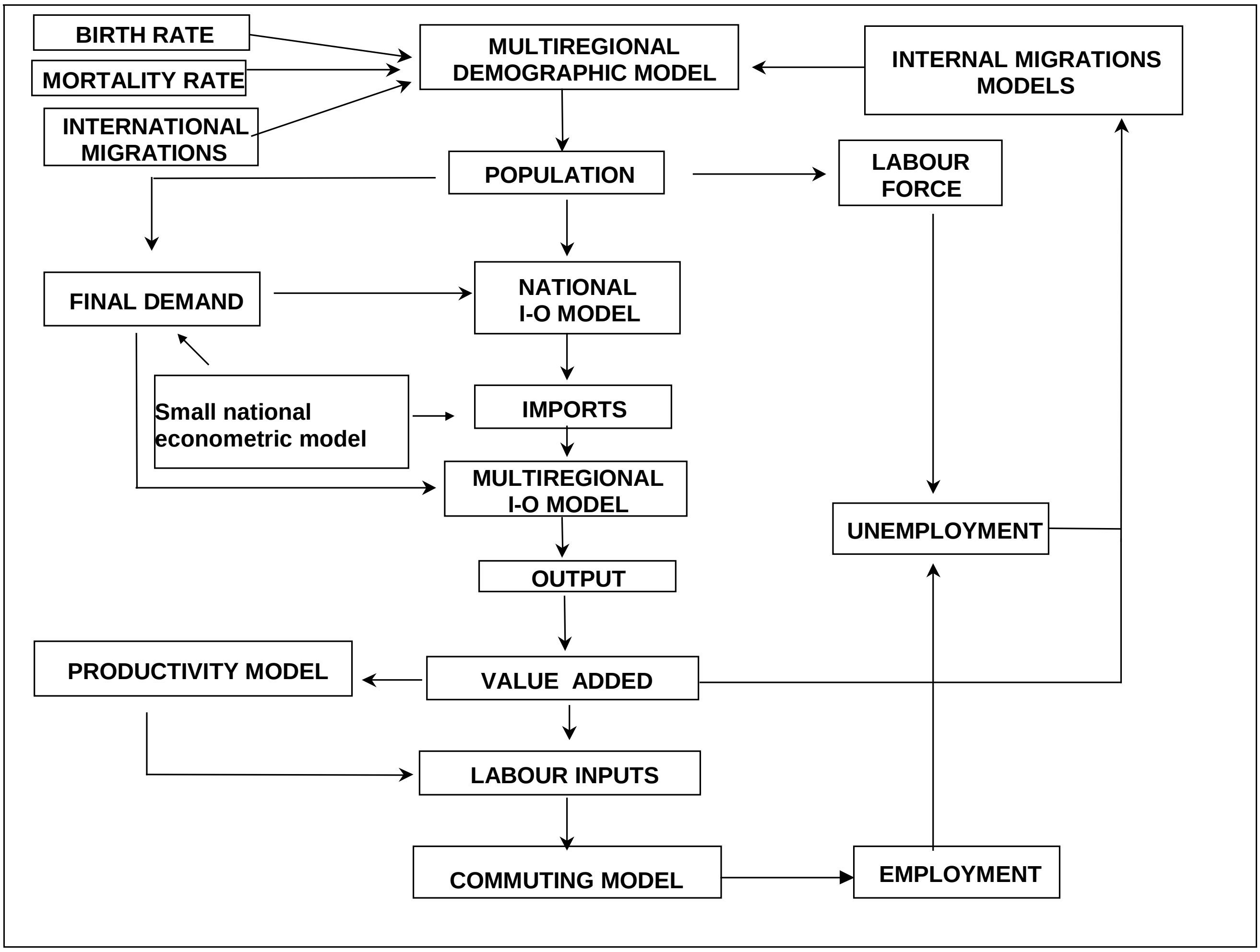
a) First of all, the demographic model computes yearly forecasts of population by region, yearly age group and gender. The starting point here are the last released official estimates of population by region of residence, age and gender and a set of hypotheses on fertility rates by region and age. These forecasts, given the regional participation rates by age and gender, yield forecasts of the labour force by region.

b) From a first set of exogenous hypotheses on national economic aggregates and a second set of regional allocation criteria, the economic model computes the net final demand by sector and region, and forecasts the corresponding value added. An interesting feature of the model is that the hypothesis on consumption may be stated in *per-capita* terms, with the regional values of aggregate consumption computed endogenously on the basis of the population forecasts. The latest version of the model employs a sketchy econometric model to compute import requirements. From the value added forecasts and the estimates of labour productivity delivered by the corresponding models regional labour inputs are obtained; the final step simply links the labour force and labour demand forecasts to obtain unemployment by region.

c) Finally, value added and unemployment are the main inputs of the interregional migration model, whose output is added to the population forecasts of the demographic model to obtain the final population forecasts.

From this outline one of the main shortcoming of the present form of the model should be apparent: it is of the “open Leontief” type, so that labour income multiplier effects are not taken into account. Developments in this direction are fully feasible and are in the current research agenda; estimates of the effects of changes in final demand on output and employment delivered by the current version of the model should be considered as somehow upward (demand fall) or downward (demand increase) biased. We shall now discuss the different parts in some detail, starting with the demographic model.

Fig. 1 Flow-chart of IDEM



2.1 The demographic model

2.1.1 *Multiregional population change*

IDEM's multiregional demographic model is able to produce yearly population forecasts by region, gender and single-year age group. It is based on the Spatial or Multiregional Population approach developed at IIASA: "[…] the dynamics of a multiregional population system are governed by its age-specific fertility, mortality and migration rates ... [these] determine not only the growth of the population, but also its age composition, spatial distribution, and crude rates" (Willekens and Rogers, 1978). This approach is an extension of traditional demographic models, which deal with single-region populations and are based only on the birth-death natural process, so considering only fertility and mortality age-specific schedules. Spatial models deal with multiregional populations, i.e. populations disaggregated simultaneously by age groups and regions. Each regional population is affected by a specific birth-death natural process, but also by an interregional migration process redistributing people across regions. It is possible to model such a process in a very similar way to the one followed in the traditional demographic model:

$$\{K_{t+1}\} = \mathrm{G}\{K_t\}$$

where $\{K_t\}$ is the age and regional distribution of the population at time t, and G is the generalized Leslie matrix (multiregional matrix growth operator) which contains the birth, death and migration schedules.

2.1.2 *International Migrations*

Although in principle international migration inflows might be endogenously modelled (se e.g. Venturini, 2001), in practice both the scarcity and low quality of the data[2] and the recent introduction of annual immigration quotas suggest that an *a priori* setting of total annual inflows and regional shares on the basis of recent experience might be a more effective option. Currently the total number of yearly foreign inflows is set at 160.000. Yearly foreign outflows are estimated by the demographic model, given the population forecasts, on the basis of the recent regional age- and gender-specific rates.

2.1.3 *Internal Migrations*

A key point here is that the age structure of migration rates is very stable (Fachin and Venanzoni, 1999): $m_{xijt} = \lambda_{xij} m_{ijt}$, where m_{xijt} is the migration rate at age x between region i and region j, m_{ijt} the mean migration rate between the same regions, and λ_{xij} is a time-independent coefficient mapping the former onto the latter. Thus, forecasts of age-specific migration rates, allowing disaggregate modelling of population changes induced by migrations, are easily recovered from those of the mean rates. These are obtained from a set of multinomial logit models for grouped data. One model is estimated for each region of origin and gender, with the regions of destination grouped in seven areas, in order to avoid modelling exceedingly small flows. Within each area the flows are allocated to each region on the basis of recent mean shares. The explanatory variables used in the models are rather standard (see for instance Daveri and Faini, 1999) and include per

[2] Foreign immigration was negligible as well as not regulated in Italy until the early 1990's; currently there are large legal immigration flows, but also significant and obviously largely unmeasured illegal ones. Thus, both time-length, coverage and overall quality of the data series are very limited.

capita valued added (as a proxy of income) and unemployment (as a proxy of the probability of obtaining a job) in the regions of origin and destination, employment shares of agricolture and construction in the regions of origin (with expected positive signs, on the grounds that jobs in these sectors are less stable and thus those holding them are more likely to migrate) and recent correlation between *per-capita* value added in the regions of origin and destination (with an expected negative sign, for instance because of risk-sharing within families). For obvious reasons of space here we will report only some summary results: RMSEs (Table 1) and the plots of the estimates of the male migration rates for two typical cases; in the first case the region of origin is Sicily (*Sicilia*), a typical southern low-income, high-unemployment region, and the area of destination the high-income North-West (Fig. 2); in the second case the region of origin is Piedmont (*Piemonte*), which is part of the North-West, and the destination area the depressed South (Fig. 3). Thus, the former are a typical example of migrations prompted by economic reasons, while the latter are bound to be return migrations.

Before commenting the table and plots some words of caution are of order: the task of modelling internal migrations in Italy is highly demanding. Some of the flows modelled are extremely small and thus erratic (as, e.g., those from Valle d'Aosta, a region with a population of just 100,000) while those originating from the wealthy regions are mostly short-distance movements, a statistical artifact due to the use of administrative boundaries, or, as in the example above, return migrations to poorer regions. Both are hard to explain on economic grounds. Finally, even long-distance migrations originating from depressed areas are somehow puzzling, as they all show a negative trend in spite of the growing unemployment differentials between North and South (Faini et al., 1997). In the light of these problems the generally not small RMSEs are hardly surprising. However, it should be pointed out that each of the RMSEs shown in table 1 is actually a mean of seven RMSEs; for each region of origin we generally find some good fits (such as those shown in Fig. 2) as well as very poor ones, even with no variable found to have any explanatory power.

Table 1
Migration Models: Root Mean Square Errors by region of origin
(x100 of mean migration rates)

Region of origin	Males	Females
	North-West	
Piemonte	14.3	13.4
Valle d'Aosta	26.4	29.4
Lombardia	9.6	10.3
Liguria	11.4	11.1
	North-East	
Trentino-Alto Adige	19.8	21.3
Veneto	11.5	12.4
Friuli	15.3	18.3
Emilia	9.7	9.6
Toscana	10.3	11.6
	Centre	
Umbria	13.3	16.1
Marche	14.9	16.0
Lazio	7.3	6.8
	South	
Abruzzo	13.4	17.1
Molise	17.8	16.6
Campania	8.1	8.5
Puglia	11.3	9.3
Basilicata	14.4	17.2
Calabria	11.4	11.4
Sicilia	9.9	8.5
Sardegna	15.1	15.4

Fig. 2 *Migration rates from Sicily (South Italy) to North-West Italy, 1975-1997*

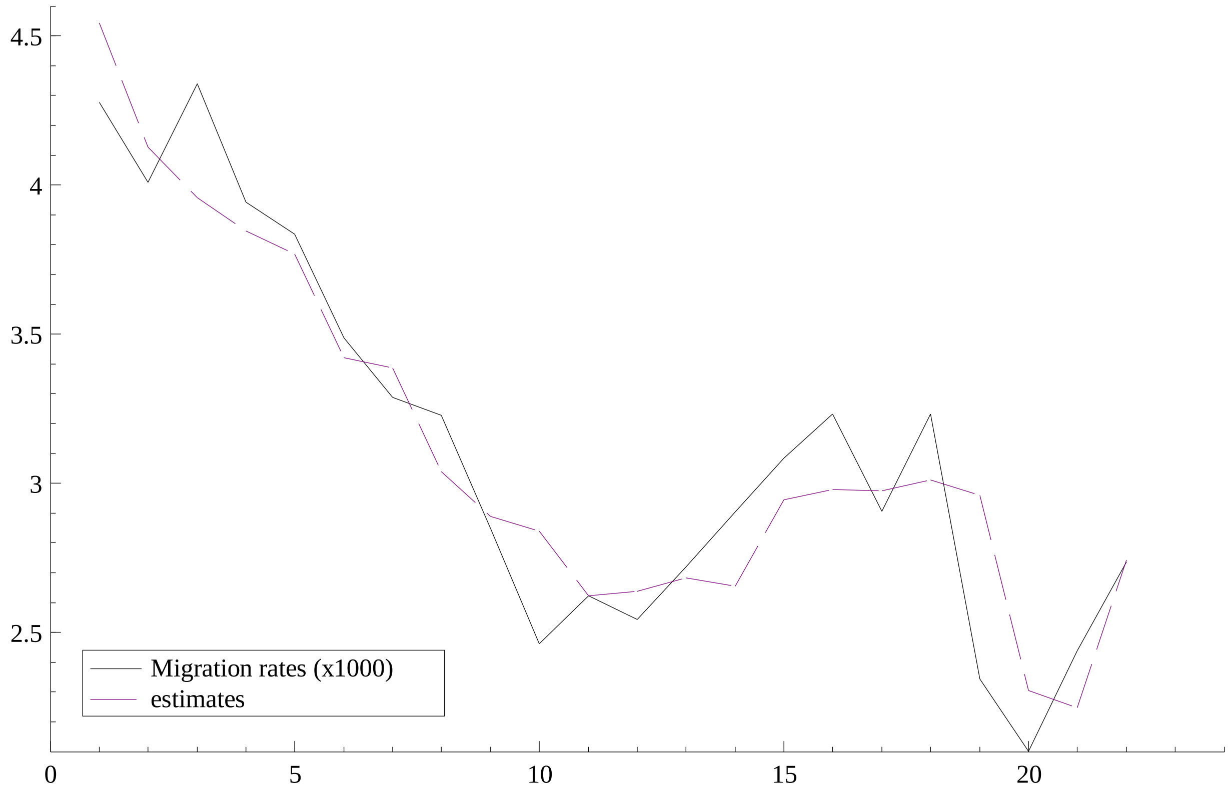
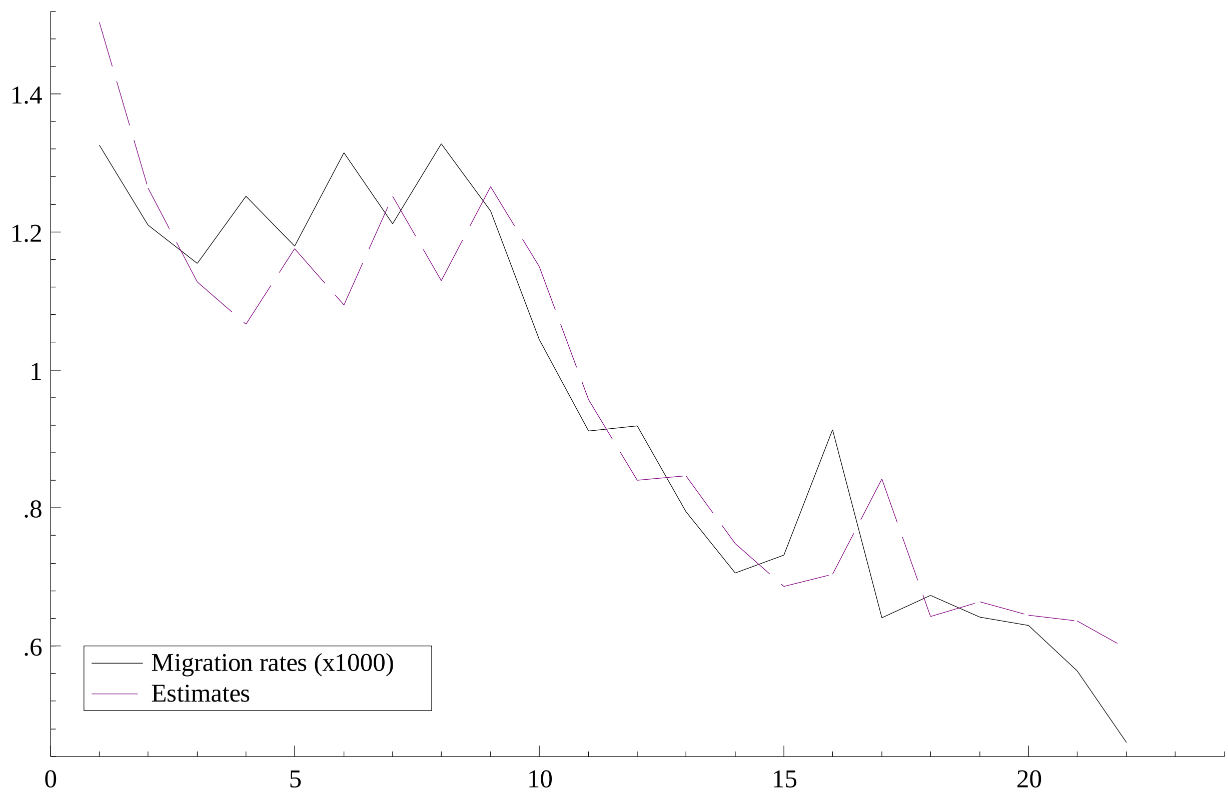


Fig. 3 Migration rates from Piemonte (North-West Italy) to South Italy, 1975-1997



2.1.3 *Labour Supply*

Labour force participation is much lower in Italy than in the other industrial economies. An exploratory analysis carried out in Fachin and Venanzoni (2000) concluded this low-participation model is a widespread and stable feature of the economy, found at disaggregate (by age a nd region) as well aggregate level. Regional participation rates may thus be modelled with no excessive risk of aggregation bias by a small number of equations resembling the β-regressions used in convergence analysis, with the difference D_i between the participation rate in region i (L_i) and the national participation rate (L) at time t modelled by an AR(1) process[3]:

$$D_{it} = (1+\gamma)D_{it-1} + u_{it}. \qquad [1]$$

Equation [1] may be rewritten as:

$$\Delta L_{it} = \Delta L_t + \gamma D_{it-1} + u_{it}, \qquad [2]$$

implying that the growth of the regional participation rate ΔL_{it} equals the national one ΔL_t corrected by a fraction γ of the lagged regional differential D_{it-1}. Three cases are possible:

i) $-1 \leq \gamma < 0$: *convergence to mean national level*;

ii) $\gamma = 0$: *constant regional differentials*;

iii) $\gamma > 0$: *divergence away from mean national level*.

Equation [2] is a particularly convenient way to model the regional dynamics of labour participation given national trends. In a baseline scenario the γ coefficient, governing the convergence processes, may be quite naturally fixed at the estimated level (thus adopting the neutral assumption that past regional differentials dynamics will continue in the future). Different scenarios (constant regional differentials, fast convergence, etc.) are easily defined by modifying γ as appropriate. For instance, with $\gamma = 0$ there is no convergence; setting $\gamma^* = k\gamma$ with $k > 1$ will increase the speed of convergence (if $\gamma < 0$) or divergence (if $\gamma > 0$).

Regional forecasts of labour participation rates may thus be constructed following two simple steps:

1. national mean participation rates for both genders are forecasted or fixed as scenario parameters;

[3] As participation Rates are defined in the (0,1) interval, we used a logistic transformation for the analysis.

2. regional participation rates for the same age groups are estimated by means of convergence equations, with coefficients possibly modified according to the convergence hypothesis adopted.

Currently, AR models (set in the first differences on the basis of standard ADF testing) are used to forecast national participation rates, with different scenarios set by modifying the equations constant term (measuring trend growth)[4].

Table 2
Regional Labour Participation: Convergence Equations

Age group	ϕ	σ_ϕ	R^2
		Males	
15-24	1.26	0.07	0.94
25-64	0.90	0.01	0.99
65-75	1.09	0.01	0.99
		Females	
15-24	1.09	0.06	0.93
25-64	0.76	0.10	0.77
65-75	1.06	0.02	0.99

estimated equation: $D_{iT} = \phi D_{i0} + u_{iT}$, with i =1,...,20, T=1999, 0=1993.

2.2 The economic model

2.2.1 *Structure*

The MRIO model is of the Chenery-Moses type. In order to establish notation it is convenient to recall its fundamental representation:

$$\mathbf{X} = (\mathbf{I} - \mathbf{TA})^{-1}[\mathbf{T}(\mathbf{C} + \mathbf{F} + \Delta\mathbf{S} + \mathbf{E}) - \mathbf{P}_{\mathrm{T}} - \mathbf{Imp}] \qquad [3]$$

Where: **X** is the output vector ($k \times s$) for the k sectors in the s regions; **A** is a block-diagonal matrix including s $k \times k$ blocks of regional input coefficients; **T** is the $k \cdot s \times k \cdot s$ matrix of interregional trade coefficients; **C, E, F,** Δ**S,** are the vectors of final demand (respectively, domestic consumption, exports, investment and inventory changes), while **Imp** is the vector of foreign imports and $\mathbf{P}_{\mathrm{T}}$ the

[4] There is a minor complication in the model. As long time series are available only for broad age groups (15-24, 25-64 and 65-75), the regional forecasts are computed for these groups, with more disaggregated forecasts (the demographic model uses 17 five-years age groups) obtained under the assumption of balanced growth within these broad groups.

vector of output transfers between sectors[5]. A base-year (1988) version of **A** and **T** matrices of regional input and interregional trade coefficients have been provided by courthesy of one the authors of the IRPET MRIO model (Casini Benvenuti et al., 1995); the two matrices are currently updated (the actual year is 1995), as soon as national and regional economic accounts are released by ISTAT, with indirect biproportional algorithms.

The starting point of a forecasting pattern is the trend of net national final demand; for instance this could be the DPEF forecast, i.e. the medium term (4-5 years) economic scenario assumed for fiscal and budgetary planning by the Italian Government. Should a longer time horizon be needed (10 to 20 years), exogenous trends of the different national aggregates have to be formulated. In both cases, national imports are then estimated with a National Input-Output Table (wich is simply the aggregation of the regional **A** matrix). The vectors of final demand and imports by regions and sectors are finally computed, depending on the scenario of interest[6] (but not entirely: as pointed out above, aggregate private consumption growth is partly endogenous *via* population growth). A research is currently done on the direct modelling of regional imports, *via* a reformulation of equation [3]:

$$\mathbf{X} = (\mathbf{I} - \mathbf{TA} + \mathbf{m\hat{}})^{-1}\,[\mathbf{T}(\mathbf{C} + \mathbf{F} + \Delta\mathbf{S} + \mathbf{E}) - \mathbf{P}_{\mathrm{T}}] \qquad [4]$$

$$\mathbf{M} = \mathbf{m\hat{}}\ \mathbf{X} \qquad [5]$$

where **m^** is a block-diagonal matrix including s $k{\times}k$ blocks of regional import input coefficients.

2.2.1 *Forecasting Value Added, Labour Inputs and Employment*

The MRIO model delivers forecasts of output by sectors and regions, which are the basis of value added and labour input forecasts. The output–value added link is simply provided by the value added input coefficients, which are embedded in the regional **A** matrix: these are kept constant for the the forecasting period. The step from value added to labour inputs is more troublesome, as, since Verdoorn (1949), labour productivity has been known to grow with value added (hence, unit labour inputs to fall): assuming constant labour input coefficients is clearly unsatisfactory. We thus need to estimate some sort of productivity equations. However, standard explanations of Verdoorn’s law, including increasing returns to scale and technical progress, point

[5] All these vectors have a dimension of $k{\times}s$; i.e. the components of net national demand are disaggregated by sector and region.
[6] For instance, a baseline scenario may assume regional growth of final demand as equal to the national mean values in all regions, implying that the regional distribution of aggregates is not changing over the forecasting period.

to a likely role of capital. This is a problem: IDEM produces no capital stocks forecasts. In principle, an auxiliary block may be added; however, any equation estimation is likely to be hampered by lack and uncertain quality of the data on regional capital stocks by sector. For the time being we had to leave this for future research, and settle for productivity equations including just value added and a time trend as explanatory variables. The modelling strategy is similar to the one adopted for participation rates, with regional labour productivity growth by sector estimated by means of a convergence-type equation expressing the current regional differential (regional productivity level minus the corresponding national one) as a function of that at the beginning of the period. Mean national productivity growth may be given by the forecasts of national models or fixed as a scenario parameter. In the current version of IDEM the former option has been preferred, with the forecasts derived from cointegrated VARs, or, if the cointegration hypothesis between labour productivity and valued added is rejected (a clear sign of the relevance of capital dynamics), Instrumental Variables equations in the first differences. The estimates of the convergence equations are reported in Table 3. The fit is generally acceptable, with some exceptions, which however have coefficients entirely in line with the rest.

As in the labour participation case, the baseline scenario simulations are based on the empirical estimates. Different forecasting options are easily introduced by replacing the estimated convergence coefficients γ_i as desired, with the most natural options being faster regional convergence or divergence. For instance, faster aggregate convergence is obtained by replacing the estimated coefficients γ_i for the various sectors with $k_i\gamma_i$, where $k_i > 1$ if $\gamma_i < 0$ and $k_i < 1$ else (the simplest way to do this is choosing a constant $\lambda > 1$ and setting respectively $k_i = \lambda$ and $k_i = 1/\lambda$): this has the effect of increasing the speed of convergence in the industries which have actually been converging and reducing that of divergence in those which have not, so that there is a faster aggregate convergence.

Table 3
Regional Labour Productivity: Convergence Equations

Industry	ϕ	σ_{ϕ}	R^2	$1+\gamma$
1	1.22	0.20	0.67	1.012
2	0.57	0.12	0.51	0.968
3.1	0.72	0.44	-	0.981
3.2	1.40	0.25	0.66	1.020
3.3	0.16	0.19	0.03	0.899
3.4	0.97	0.14	0.73	0.998
3.5	0.70	0.44	-	0.979
3.6	0.70	0.22	0.37	0.980
3.7	1.09	0.17	0.72	1.005
3.8	0.84	0.47	0.06	0.990
3.9	0.72	0.14	0.63	0.981
3.10	1.24	0.13	0.75	1.013
4	0.73	0.18	0.43	0.982
5	1.05	0.18	0.61	1.003
6	0.52	0.06	0.76	0.963
7	0.65	0.16	0.50	0.975
Mean	*0.83*	*0.22*	*0.46*	*0.98*
Median	*0.73*	*0.18*	*0.56*	*0.98*
Max	*1.40*	*0.47*	*0.76*	*1.02*
Min	*0.16*	*0.06*	-	*0.90*

NB

(a) *estimated equation:* $D_{iT} = \phi D_{i0} + u_{iT}$, with i =1,...,20, T=1996, 0=1980;

(b) $(1+\gamma) = \phi^{1/(1996-1980)}$ is the implicit AR(1) coefficient.

(c) Sectors as follows (productivity in the Public Administration, constant by convention, is not modelled):

1. Agriculture, hunting and forestry
2. Construction

3. Manufacturing

3.1 Food products, beverages and tobacco

3.2 Textiles and textile products, Leather and leather products

3.3 Wood, wood products, rubber and plastic products, Manufacturing n.e.c.

3.4 Pulp, paper and paper products; publishing and printing

3.5 Coke, refined petroleum products and nuclear fuel

3.6 Chemicals, chemical products and man-made fibres

3.7 Other non-metallic mineral products

3.8 Basic metals and fabricated metal products

3.9 Machinery, electrical and optical equipment

3.10 Transport equipment

4. Wholesale and retail trade, Hotels and restaurants
5. Transport, storage and communication
6. Financial intermediation
7. Services n.e.c.
8. Public Administration

3. Conclusions

The model presented certainly has many and serious limitations. However, it has several noteworthy features as well, such as the good integration of demographich and economic forecasts and the possibility of easily obtaining forecasts under different hypothesis on the future trends of regional differentials of some key variables. Among the much work to be done, we should mention the upgrading of the economic data-base to the new regional accounts based on the SEC95 (SNA93) standards, and the tranformation of the Input-Output model to an "open Leontief"-type one.

References

Casini Benvenuti, S., D. Martellato and C. Raffaelli (1995) INTEREG: A Twenty-region Input-Output Model for Italy, *Economic Systems Research*.

Daveri, F. and R. Faini (1999) "Where Do Migrants Go? Risk-Aversion, Mobility Costs and the Locational Choice of Migrants" *Oxford Economic Papers*, vol. 51, pp. 595-622.

Fachin, S. and G. Venanzoni (1999) "Patterns of Internal Migrations in Italy, 1970-1993", American Statistical Association, *1999 Proceedings of the Business and Economic Statistics Section*, pp. 76-81.

Fachin, S. and G. Venanzoni (2000) "Regional Trends in Labour Force Participation in Italy", American Statistical Association, *2000 Proceedings of the Business and Economic Statistics Section* (CD-ROM).

Faini, R., G. Galli, P. Gennari and F. Rossi (1997) "An Empirical Puzzle: Falling Migration and Growing Unemployment Differentials among Italian Regions" *European Economic Review*, vol. 41, pp. 571-579.

Isserman A. M. ed. (1986), *Population Change and the Economy: Social Science Theories and Models*, Kluwer-Nijhoff Publishing, Boston.

U.S. Dept. of Commerce, Bureau of Economic Analysis (1981) *Regional Input-Output Modelling System (RIMS II): Estimation, Evaluation, and Application of a Disaggregated Regional Impact Model* U.S. Government Printing Office Washington, D.C.

U.S. Dept. of Commerce, Bureau of Economic Analysis (1986) *Regional Multipliers: a User Handbook for the Regional Input-Output Modelling System* U.S. Government Printing Office, Washington, D.C. (3rd ed. 1997).

Venturini, A. (2001) *Le migrazioni e i paesi sudeuropei* UTET, Torino.

Verdoorn, P.J. (1949) "Fattori che regolano lo sviluppo della produttività del lavoro" *L'Industria*, vol. 3, pp. 45-53.

Willekens, F. and A. Rogers (1978) *Spatial Population Analysis: Methods and Computer Programs*, Research Report RR-78-18, International Institute for Applied Systems Analysis, Laxemburg, Austria.