Intercity Interactions : Evidence from the US*

Andrea R. Lamorgese, Research Department - Banca d'Italia
Gianmarco I.P. Ottaviano, U. Bologna, FEEM and CEPR

February 2006

Abstract

Using national level input-output matrices, we propose a strategy to identify pecuniary externalities operating through the markets for intermediate goods at the local level. Then, controlling for common shocks in a spatial econometric framework, (i) we estimate the effect of pecuniary externalities on productivity growth; (ii) we disentangle such effect from the one of other local interactions (i.e. knowledge or other face-to-face spillovers) and that of local characteristics; (iii) we evaluate the scope of operating of all kind of externalities using different distance measures. Our estimates suggest that pecuniary externalities and other kinds of local interactions coexist, that their effect on productivity growth is decreasing with distance and that it depends on inter-city diversity and the pattern of local specialisation.

Preliminary and incomplete version
Do not cite without authors' permission

JEL classification: R12, R15

Keywords: customer- and supplier-driven externalities, knowledge spillovers, spatial autoregressive models, dynamic factors

---

*We thank Mario Forni, Sergio Paba, Daniele Terlizzese, and the participants to the workshop "Clusters, industrial districts and firms: The challenge of globalisation", Modena, September 2003, to the annual meeting of the EEA, Madrid, September 2004, and to the Department seminar at the University of Alicante for their comments. The views expressed herein are those of the authors and not necessarily those of the Bank of Italy.
Corresponding author: Andrea R. Lamorgese, email: andrea.lamorgese@bancaditalia.it

1 Introduction

When looking at the way economic activity spreads over the space, agglomeration stands out as a key feature; such pattern is apparent for instance in satellite night views of both United States and Europe. Spatial agglomeration supports the view that economic activities comove locally, possibly due to underlying economic interactions that are favoured or hampered by distance; in this sense geographical proximity matters for comovement at the local level. At the same time, however, similarities in their sectoral composition may generate comovements also between cities that are far apart (e.g., high tech poles in Silicon Valley, Boston Route 128, the North Carolina Research Triangle). This bears the question of the relative importance of geographical distance ('spatial hypothesis') and sectoral composition ('sectoral hypothesis') in explaining cities interactions and comovements among local business cycles. The nature and the channels of interaction among cities are relevant since they shape comovement among economic activities at the local level, i.e. since they determine local business cycles, which by aggregation may transmit into the aggregate business cycles.

This paper makes a contribution towards the empirical study of the spatial interrelations of economic activities. In particular, it aims at identifying the forces that drive the comovements of income and productivity across different locations. To avoid issues linked to 'border effects', which may distort the working of purely economic forces through administrative barriers, we focus on a sample of Metropolitan Statistical Areas (henceforth MSA or cities) within the US. On this sample we run a two-step procedure: first, we determine clusters of comoving cities; second, we investigate the economic determinants of clustering.

In order to assess the relative merits of the 'spatial hypothesis' and the 'sectoral hypothesis', initially we provide some description of the composition of the resulting clusters. It results apparent that geography is an important determinant of the ways cities endogenously group:[1] for the top half of clusters between 1/3 and 1/2 of the cities in each cluster belong to the same region. Nevertheless the remaining cities in each cluster show an high degree of correlation within the group which is not apparently due to a clear geographical pattern. Therefore in a second step of the analysis, we try to ascertain whether, conditionally to geographical proximity, we can identify further sources of correlation.

To uncover conditional correlations, we turn to spatial econometric analysis. Specifically, we use a spatial econometric model to assess the effect on the growth rates of each city of two measures of market interactions, after controlling for the effect of geographical proximity and other features. Such two measures — the average growth rates of per worker output of *customer* and *supplier* cities — are based on the loading of the I-O matrix.[2]

Beyond the economic meaning of such analysis, this paper provides a methodological contribution of more general interest, by setting a procedure

[1] As described in section 2 the algorithm of clustering we use does not impose ex-ente either the number of clusters or the number of cities in each cluster, thus allowing cities to divide into clusters only in order to maximise correlation within clusters and minimise it between clusters. In this simple sense the grouping can be considered endogenous.

[2] In their work on sectoral complementarities, Conley and Dupor (2003) construct different measures of distance among sectors, basing on the norm distance among the vectors of factor loadings of the I-O matrix. Preliminary analysis using their measure of distances within our framework confirms our results.

for identifying and estimating local interactions which keeps into account the possible action of common shocks and the considerable local heterogeneity, both very likely features in macro data with an important geographical dimension. The procedure consists in the joint use of three techniques: a spatial econometric model to jointly evaluate the interplay of local and global interactions, a dynamic factor model to condition out common shocks, and a dynamic clustering algorithm to account for fixed effects. Provided that the interaction under analysis is a local phenomenon, the paper shows that results would be biased if controls for common shocks and fixed effects were not included in the spatial regression.

Our overall findings reveal that the endogenous clustering structure does not seem to reflect solely geographical proximity. In particular, sectoral input-output linkages also matter in explaining local productivity growth. Nonetheless, the impacts of these linkages are crucially shaped by distance among locations. Therefore, both 'spatial' and 'sectoral' considerations are equally needed to explain the comovements among local business cycles.

The remaining of the paper is organised as follows: section 2 details the dynamic clustering exercise, along with the data on which the empirical exercise is run (section 2.1), and illustrates the features of the clustering structure obtained by means of some descriptive statistics (section 2.2). Section 3 describes the spatial econometric model, while section 4 explains why we need to nest this analysis within a dynamic factor model and a clustering exercise. Section 5 presents and discusses the estimates of the spatial econometric analysis. Section 6 concludes.

2 Dynamic clustering

In this section we perform a clustering exercise in order to uncover and describe the structure of local correlation in the data. In so doing, we only observe whether economic activities in different locations are linked, while leaving the analysis of the underlying economic reasons of such comovement to the next section.

Performing a clustering exercise amounts to grouping cities in different clusters depending on their affinities according to an pre-established measure. In this paper we address this question using non-parametric techniques based on dynamic cluster analysis (Rodrigues, 1998). The result will be a basic description of the structure of correlation of the data, which is informative of the forces driving local commonality.

Describing the data is not the only reason of interest towards the clustering structure of the data, though. A more technical reason is related to the spatial model we will estimate. Since the amount of idiosyncratic behaviour in city data is expectedly large, the spatial model needs to be estimated with fixed effects on the cluster the city belongs to. The choice of using fixed effect on the clusters rather than on the state or on the region reflects the consideration that the specific effects affecting interaction may cross states' or regions' borders and that they are better described by the grouping arising endogenously from the clustering exercise.

A further even more technical interest in the clustering structure is related the the estimation of the number of common factors in the dataset. Further details are provided in section 4.

The dynamic clustering exercise we perform amounts to dividing the series y_{it} in the dataset into groups ('clusters') such that a certain statistics is maximised within groups and minimised between groups.[3] The statistic we consider is the *coherence* spectrum, which has the analogous interpretation of an R^2 in the frequency domain.[4] Since the coherence spectrum exploits cross-sectional dependence among time series at all leads and lags and –as a frequency domain based approach– allows to focus on cycles of difference frequencies, the exercise is dynamic, unlike more traditional techniques which are based on contemporaneous correlation.[5] Another point of strength of this clustering algorithm, with respect to other more traditional techniques, is that it is non-parametric and it does not impose ex-ante any structure of clustering, that is, the classification method is model free and does not rely on prior beliefs on the clusters composition and number: they both arise endogenously from the data.[6]

[3] This is the sense of the joint minimisation of the *global cohesion criterion* in Rodrigues (1998), which we take as reference all along the dynamic clustering exercise.

[4] At any frequency λ, it can be shown that the coherence spectrum $h^2_{XY}(\lambda)$ between two series $\{X_t\}$ and $\{Y_t\}$ is the proportion of variance of $\{X_t\}$ captured by the best linear projection of $\{X_t\}$ on the leads and lags of $\{Y_t\}$, that is $R^2_{X|Y}(\lambda)$. By the symmetry of the coherence spectrum it can be shown that $h^2_{XY}(\lambda) = R^2_{X|Y}(\lambda) = R^2_{Y|X}(\lambda)$.

[5] A definition of the coherence spectrum which highlights its inherent dynamic nature is the following

$$h_{XY}(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{+\infty} e^{ih\lambda} \gamma_{XY}(h), \qquad \text{for} \quad -\pi < \lambda < +\pi$$

where $\gamma_{XY}(h)$ is the correlation coefficient between X and Y at lag h (Brockwell and Davis, 1991). The integral of $h_{XY}(\lambda)$ at (e.g.) the business frequencies represents therefore a measure of comovement between X and Y over those frequencies.

[6] Traditional clustering algorithms are either hierarchical or optimising: the former are free from initial conditions –that is invariant with respect to the initial partition– and do not impose the number of clusters, while the latter allow for comparison of different partitions with the same given number of clusters and therefore for reallocation of units among clusters. The algorithm used here nests the features of both at two different stages and therefore enjoys the advantage of both techniques.

2.1 Data description

Our dataset consists of 2 sets of observations concerning the growth rates of Total Personal Income (henceforth 'income') and the growth rates of Per Worker Personal Income or Labour Productivity (henceforth 'productivity') across 318 US cities in 32 years:[7]

$$y^i_{mt};\ m = 1, 2;\ i = 1, ..., 318;\ t = 1969, ..., 2000$$

where $m = 1$ refers to income and $m = 2$ refers to productivity. The latter variable is obtained dividing income by employment, worked hours being unavailable at this breakdown.[8] This is obviously just a crude measure of labour productivity. The data are contained in the REIS database by the Bureau of Economic Analysis, Economics and Statistic Administration of the U.S. Department of Commerce (see Appendix 1 for details). The series are stationary in log-differences.

Input-output (henceforth I-O) matrices used in this paper are the *use* tables taken from the 1985 and 1987 benchmark input-output accounts. They are provided at the national level;[9] in section 3 we explain our strategy to use information from the national I-O matrices to proxy intercity market

[7] Total personal income is the measure of output at the city level (See Appendix 1). Data on total personal income present a breakdown into 82 sectors.

[8] The metropolitan area definitions used by BEA for its personal income estimates are the county-based definitions issued by the Office of Management and Budget (OMB) for Federal statistical purposes. OMB's general concept of a metropolitan area is that of a geographic area consisting of a large population nucleus together with adjacent communities having a high degree of economic and social integration with the nucleus.

[9] The Bureau for Economic Analysis now provides also regional input-output multipliers under a program named Regional Input-Output Modeling System (RIMS II), but they are not available to us at the moment.

interactions.

The sector breakdown in the REIS database and in the I-O matrices, as well as the sector breakdown of the two I-O matrices for 1985 and 1987, do not match perfectly. After aggregating sectors opportunely in order to match the breakdown in the REIS database and in the I-O matrices, we obtain two distinct sector breakdowns, which are described in tables 4 and 5 in Appendix 1. As a robustness check we perform all the econometric exercises using weighting matrices based on either I-O matrix.

2.2 Results of the clustering exercise

We perform the dynamic clustering exercise on the two different set of variables, namely the growth rates of MSAs total personal income and the growth rates of MSAs productivity. Total personal income divides into five clusters, productivity into two.

We do not really want to comment these results comparatively, but we want to stress that the smaller the number of clusters arising, the more common the correlation structure. For instance, if all series in the panel were driven by only one shock which diffused quickly and dominated the interdependence of all series, we would get just one cluster. On the contrary, uncovering several clusters would mean that the interdependence among series is more local, an observation compatible, for instance, with slow diffusion of shocks originated locally, very idiosyncratic reactions to a single shock, several common shocks which hit the data with different idiosyncratic reactions.

If strong common shocks drive the dynamics of the panel, the clustering structure obviously tends to reflect the cross sectional dependence among those shocks. If we were interested in the underlying local interdependence structure, we should first control out the common shocks and then look at the clustering structure arising. By repeating the clustering exercise on the set of idiosyncratic components $y_{mt}^{i,\text{idio}}$ in equation (13), we discover that 39 different clusters arise for income and 50 for productivity. We will exploit this information in order to consider proper fixed effects in the estimation of the spatial model, as explained in section 4.

Looking at the composition of the first 24 clusters computed once the common shocks have been controlled out, geographical distance seems to be a very relevant force in explaining how cities endogenously group into clusters (see table 1): between 1/3 and 1/2 of the cities in each cluster belong to the same region. On the other hand between 1/2 and 2/3 of cities belong to the same cluster even if there is no clear dependence based on geography, hinting that some other force might be operating. In the remaining part of the paper we tests whether this further agglomeration force can be identified as the market interaction described by input-output linkages among sectors of U.S. cities.

group	region	state
1	6/14 Mid East 4/14 Great Lakes	5/14 PA
2	5/14 South East 3/14 South West 3/14 Far West	
3	4/14 Far West 3/14 Great Lakes 2/14 South West 2/14 New England	
4	4/13 South East 3/13 Great Lakes 3/13 South West	
5	5/13 Great Lakes 3/13 South West	
6	4/11 South East 3/11 New England	
7	6/11 South East 2/11 New England 2/11 Mid East	
8	6/11 South East 4/11 Mid East	
9	3/10 Plains 3/10 Far West	
10	4/9 Great Lakes 2/9 Mid East 2/9 South East	
11	3/9 Great Lakes 3/9 Mid East	
12	3/8 Rocky Mountains 2/8 Great Lakes	
13	3/8 Mid East 3/8 Far West	3/8 NY
14	4/8 South East 2/8 Grat Lakes	
15	4/8 South East 2/8 South West 2/8 Great Lakes	
16	4/8 Great Lakes	
17	3/7 Mid East	
18	3/7 South East	
19	2/6 South East	
20	3/6 Plains	
21	4/6 Plains	
22	4/6 South West	
23	3/6 South West	
24	3/6 Plains	

Notes: Composition of the largest 24 clusters of income per worker.

Table 1: Composition of the first 24 clusters on the idiosyncratic component of productivity

3 Spatial analysis

Suppose we want to estimate the effect on the growth rate of productivity in each city i (y^i_{2t}) of the growth rates of productivity of all cities $j \neq i$. All other cities cities may influence city i's productivity in two ways, namely through 'market interactions' (demand and supply linkages) and 'non-market interaction' (face-to-face contacts).

We suppose that market interactions between cities (i) and all other cities ($j \neq i$) can be approximated by the conditional correlations of local productivity growth rates (in city i) and two (input-output based) measures productivity growth in all other cities: an aggregate of the growth rates of *supplier* cities and the growth rates of *customer* cities.[10] Considering just one couple of cities i, j, we assume that all market interactions between these two cities can be approximated by the flows of inputs bought by each sector of city i from all supplier sectors in city j, and the flows of goods sold by each sector of city i to all customer sectors in city j. Market interactions of city i with all other cities are consequently described by the sum of all bilateral flows of inputs/goods purchased and sold by all sectors in city i and all sectors of all cities $j \neq i$, i.e. by the aggregates

$$\bar{y}^{i,\text{cust}}_{2t} = \sum_{j \neq i} \omega^c_{ij}\, y^j_{2t} \tag{1}$$

$$\bar{y}^{i,\text{suppl}}_{2t} = \sum_{j \neq i} \omega^s_{ij}\, y^j_{2t} \tag{2}$$

[10]This choice relies on Bartelsman, Caballero and Lyons (1994), who estimate such correlations for the U.S. manufacturing industry at the four digits sectoral level breakdown.

where ω^s_{ij} and ω^c_{ij} are weights based on the 'use of commodity-by-industry matrix' (U). Each column (k) of such matrix shows for a given industry the value of commodity k provided as input to the (m) industries shown on the rows. Therefore the kth-column sum of such matrix represents the value of all inputs that the kth industry provides to industries (its customers) producing the commodities shown on the rows. Analogously the mth row of the same matrix (or the mth column of its transposed) shows the value of each commodities that the mth sector requires as inputs from all industries; therefore, the mth-row sum of matrix U will show the value of all commodities that the sector producing the mth commodity obtains from all other industries (its suppliers).

Suppose that at the national level commodities from S sectors are represented in the U matrix, so that U is an SxS matrix. Also suppose that S_i sectors are active in each city i (for $i = 1 \dots N$); so that one can easily obtain an S_ixS_j submatrix $U^{m,k}$ by appropriately selecting the flows among the relevant sectors from the use matrix (U) at the national level. We argue that such flows represent the bilateral flows among all sectors of city i and city j.[11] To obtain the matrix of weights for customers cities, consider the

[11] Hence, for $i = j$, U^{ii} is the *use* matrix representing flows of inputs and goods among industries in city i. For $i \neq j$, U^{ij} is the use matrix representing flows of inputs and goods among industries in city i and j.

following matrix

$$\Omega^c_{i,j} = \begin{bmatrix} \sum\limits_{k\in S_1}\sum\limits_{m\in S_1} U^{m,k} & \sum\limits_{k\in S_2}\sum\limits_{m\in S_1} U^{m,k} & \cdots & \sum\limits_{k\in S_n}\sum\limits_{m\in S_1} U^{m,k} \\ \sum\limits_{k\in S_1}\sum\limits_{m\in S_2} U^{m,k} & \sum\limits_{k\in S_2}\sum\limits_{m\in S_2} U^{m,k} & \cdots & \sum\limits_{k\in S_n}\sum\limits_{m\in S_2} U^{m,k} \\ \vdots & & & \\ \sum\limits_{k\in S_1}\sum\limits_{m\in S_n} U^{m,k} & \sum\limits_{k\in S_2}\sum\limits_{m\in S_n} U^{m,k} & \cdots & \sum\limits_{k\in S_n}\sum\limits_{m\in S_n} U^{m,k} \end{bmatrix}. \tag{3}$$

Each element $\{i, j\}$ of matrix (3) represents the total flow of goods from all sectors of city i to all their customer sectors of city j. Hence, the sum of the ith row of matrix (3) represents the total flow of goods from city i to all her customer-cities. Now we drop the main diagonal and rescale matrix (3) so that rows add up to 1, so to obtain:[12]

$$\omega^c_{i,j} = \begin{bmatrix} 0 & \frac{\Omega^c_{12}}{\sum_{j\neq 1}\Omega^c_{1j}} & \cdots & \frac{\Omega^c_{1n}}{\sum_{j\neq 1}\Omega^c_{1j}} \\ \frac{\Omega^c_{21}}{\sum_{j\neq 2}\Omega^c_{2j}} & 0 & \cdots & \frac{\Omega^c_{2n}}{\sum_{j\neq 2}\Omega^c_{2j}} \\ \vdots & & & \\ \frac{\Omega^c_{n1}}{\sum_{j\neq n}\Omega^c_{nj}} & \frac{\Omega^c_{n2}}{\sum_{j\neq n}\Omega^c_{nj}} & \cdots & 0 \end{bmatrix} \tag{4}$$

The matrix of weights ω^s_{ij} is built analogously using the transpose of matrix

[12]Although having the main diagonal of the matrix of spatial weight equal to zero is a requirement of the spatial autoregressive model (LeSage, 1999), from a theoretical viewpoint, this is not necessarily a smart idea, since it means that we are neglecting the fact that i.e. richer cities might get their wealth from their backyard, which is the basic idea behind the *home market effect*. Empirically though, the effect captured from the weights on the main diagonal of the matrix of spatial weights proved not to be important in our model.

U, according to the reasoning described above.

Notice that we are supposing that sectors in close and distant cities have the same flow of goods and inputs. This is an extreme assumption which amounts to take to the limit the idea that face-to-face interactions are hampered by distance, while interactions through markets are not.[13]

With this two measures of input-output inter-relations among cities we can estimate the following spatial regression:

$$y_{2t}^{i} = \alpha + \rho \sum_{j \neq i} w_1^{ij}\, y_{2t}^{i} + \beta\, \bar{y}_{2t}^{i,\mathrm{suppl}} + \gamma\, \bar{y}_{2t}^{i,\mathrm{cust}} + \varepsilon_{it}, \qquad \forall i, t. \tag{5}$$

where y_{2t}^{i} is the annual growth rate of per worker income of city i at time t, and $w_1^{ij} = \left[\{w_d\}_{ij}\right]^{-2}$, $\{w_d\}_{ij}$ is the distance as crows fly between city i and city j . In a more compact matrix notation

$$(I - \rho W) y_2 = \alpha + \beta \bar{y}_2^{\mathrm{suppl}} + \gamma \bar{y}_2^{\mathrm{cust}} + \varepsilon. \tag{6}$$

with $W = I_T \otimes w_1$, and T is the time range of each series in the dataset. We use spatial econometrics because we believe that interactions among US cities are not exhausted by the sole input-output relation. Any other interaction which operates through space is therefore captured by the coefficient ρ of the spatial lag. Notice that neglecting this possible interaction would confine the term in the error, thus violating the standard hypothesis of uncorrelated errors.

As robustness check, we also estimate an analogous model with two

[13]When estimating the model, though, in some specifications we also consider spatial lags of productivity (income) growth of the customer and suppliers.

slightly different aggregates for customer cities and supplier cities. For each city i we choose its most representative sector (S_i^*), that is the sector which is most concentrated in that city, where for concentration of sector s in city i we mean the ratio between the relative weight of output of sector s in city i and the relative weight of sector s in the national GDP.[14] With this mapping from city i to sector S_i^* we construct the following weighting matrices for customer sectors

$$\Psi_{i,j}^c = \begin{bmatrix} U^{11}(S_1^*, S_1^*) & U^{12}(S_1^*, S_2^*) & \dots & U^{1n}(S_1^*, S_n^*) \\ U^{21}(S_2^*, S_1^*) & U^{22}(S_2^*, S_2^*) & \dots & U^{2n}(S_2^*, S_n^*) \\ \vdots & & & \\ U^{n1}(S_n^*, S_1^*) & U^{n2}(S_n^*, S_2^*) & \dots & U^{nn}(S_n^*, S_n^*) \end{bmatrix} \tag{7}$$

As before we drop the main diagonal and rescale so that rows add up to 1, so to obtain

$$\psi_{i,j}^c = \begin{bmatrix} 0 & \frac{U^{12}(S_1^*, S_2^*)}{\sum_{j\neq 1} U^{1j}(S_1^*, S_1^*)} & \dots & \frac{U^{1n}(S_1^*, S_n^*)}{\sum_{j\neq 1} U^{1j}(S_1^*, S_1^*)} \\ \frac{U^{21}(S_2^*, S_1^*)}{\sum_{j\neq 2} U^{2j}(S_1^*, S_1^*)} & 0 & \dots & \frac{U^{2n}(S_2^*, S_n^*)}{\sum_{j\neq 2} U^{2j}(S_1^*, S_1^*)} \\ \vdots & & & \\ \frac{U^{n1}(S_n^*, S_1^*)}{\sum_{j\neq n} U^{nj}(S_1^*, S_1^*)} & \frac{U^{n2}(S_n^*, S_2^*)}{\sum_{j\neq n} U^{nj}(S_1^*, S_1^*)} & \dots & 0 \end{bmatrix} \tag{8}$$

Analogously we obtain $\psi_{i,j}^s$ using the transpose of the use matrix U. We

[14] We have also tried mapping each city into its largest sector.

then replace the new weights in (1) to obtain

$$\bar{\bar{y}}\,{}^{i,\text{cust}}_{2t} = \sum_{j \neq i} \psi^{c}_{ij} y^{j}_{2t} \tag{9}$$

$$\bar{\bar{y}}\,{}^{i,\text{suppl}}_{2t} = \sum_{j \neq i} \psi^{s}_{ij} z^{j}_{2t}, \tag{10}$$

which we use to estimate

$$(I - \rho W)z = \alpha + \beta \bar{\bar{z}}\,^{\text{suppl}} + \gamma \bar{\bar{z}}\,^{\text{cust}} + \varepsilon. \tag{11}$$

4 Econometric issues

There are a number of econometric issues involved by the estimation of models (5) and (11). First, any spatial econometric model cannot be estimated by ordinary least squares, since the estimates would be biased and unconsistent; section 4.1 describe the two steps procedure which is customarily used to estimate such models. Second, if the series y^{i}_{mt} admit a dynamic factor representation, that is the dataset of city income is driven by a restricted number of common shocks, estimates are inconsistent if the common shocks are not taken into account in the estimation; section 4.2 describes the problem in details and shows how to obtain the controls for common shocks based on Forni and Reichlin (1998), Sala (2001) and Lamorgese and Ottaviano (2003).[15] Finally, there is a number of reasons to believe that

[15]Alternative methods of estimating the common component (z^{com}, according to the notation of section 5) consist in filtering the set of variables through a filter based on the eigenvectors associated with the largest q eigenvalues of the cross-spectra of the variables in the dataset (Forni, Hallin, Lippi and Reichlin, 2001), or through a filter based on the eigenvectors of the largest q eigenvalues of the variables in the dataset augmented by

the data used exhibit a certain heterogeneity across section. It is therefore a good idea to estimate equations (5) and (11) including controls for fixed effects. One possibility is to include fixed effects accounting for the state or the region each city belongs to. Both choices suffer from the problem that they attempt to capture systematic effects connected to unobserved economic mechanisms that are not necessarily bounded within the borders of a state or a region. We therefore prefer to consider fixed effects according to the clustering structure that arises endogenously from the dynamic clustering exercise in section 2.

4.1 Estimation of the spatial model

In this section we show why a spatial autoregressive model cannot be estimated by ordinary least square, and explain the customary two-step procedure of estimation. Consider model (6) and rewrite it as

$$y_2 = \rho\, W\, y_2 + \beta \bar{Z} + \varepsilon \tag{12}$$
$$\text{where } \bar{Z} = \left[\ \imath,\ \ \bar{y}_2^{\text{suppl}},\ \ \bar{y}_2^{\text{cust}}\right].$$

As suggested by Anselin (1988), the OLS estimate of the parameter ρ in this model is biased and not consistent due to the spatial interdependence. However, the OLS procedure can be described by the following steps:

1. OLS estimate of $y_2 = \bar{Z}\beta_0 + \varepsilon_0$, from which $\hat{\beta}_0 = (\bar{Z}'\bar{Z})^{-1}\bar{Z}'y_2$

an appropriate number of lags and leads of such variables (Stock and Watson, 1999). Although Forni and Lippi (1997) show that under a set of conditions averages of the variables (as we do in this paper) correctly approximates the common components, we plan to check the robustness of our results with respect to the estimation of common components *à la* Stock and Watson (1999) and Forni et al. (2001).

2. OLS estimate of $Wy_2 = \bar{Z}\beta_L + \varepsilon_L$, from which $\hat{\beta}_L = (\bar{Z}'\bar{Z})^{-1}\bar{Z}'Wy_2$

3. estimate ρ as a partial regression coefficient of (12), that is as $e_0 = \rho e_L + \varepsilon$, where e_L and e_L are the residuals of the OLS regression of y_2 and Wy_2 on $\bar{Z}$

4. given the estimate $\hat{\rho}$ for the autoregressive parameter, compute $\hat{\beta} = \hat{\beta}_0 - \rho\hat{\beta}_L$ and $\hat{\sigma}^2_\varepsilon = (1/n)(e_0 - \rho e_L)'(e_0 - \rho e_L)$

$$\hat{\rho} = (e_L' e_L)^{-1} \; e_L' \; e_0, \quad \text{therefore}$$

$$E\big[\hat{\rho}\big] = \rho + \underbrace{E\Big\{\big[(Wy_2 - \bar{Z}\beta_2)' \; (Wy_2 - \bar{Z}\beta_2)\big]^{-1}(Wy_2 - \bar{Z}\beta_2)'\varepsilon\Big\}}_{\text{bias}}.$$

Since Wy_2 is not fixed in repeated samples, one cannot pass the expectation operator over the term $\big[(Wy_2 - \bar{Z}\beta_2)' \; (Wy_2 - \bar{Z}\beta_2)\big]^{-1}(Wy_2 - \bar{Z}\beta_2)'$, which prevents the bias term from vanishing. This rules out also consistency since the $\text{plim}\{y_2'W'\varepsilon\}$ does not vanish either. The correct way of estimating ρ is to replace step 3 with

3 bis. given e_0 and e_L, find ρ which maximises the concentrated likelihood function: $L_C = -(n/2)\ln(\pi) - (n/2)\ln(1/n)(e_0 - \rho e_L)'(e_0 - \rho e_L)$ $+ \ln(\mid I - \rho W \mid)$.

4.2 Common shocks and spatial autoregressive models

In this section we describe a general problem arising for spatial autoregressive models in the presence of common shocks, and show how dynamic factor models provide a natural solution for such problems. In so doing we rely on the explanation provided by Giannone and Lenza (2002) on a related problem.

Let $\{y_{mt}^i\}_{i=1,\ldots,I}$ be a set of I time series of a certain variable y (e.g., growth rates of income or productivity in the I cities of the United States). Suppose that each y_{mt}^i admits a (possibly dynamic) factor representation, that is each y_{mt}^i series can be represented as the sum of a linear combination of common factors (that is, a set of shocks common to all i) and an idiosyncratic shock (that is, a shock specific to the series i), as

$$\begin{aligned} y_{mt}^i =& \underline{\lambda}(L)\ \underline{\mathbf{f}}_t + y_{mt}^{i,\text{idio}} = \lambda_1(L) f_{1t} + \lambda_2(L) f_{2t} + \ldots + \lambda_q(L) f_{qt} + y_{mt}^{i,\text{idio}} \\ =& y_{mt}^{i,\text{common}} + y_{mt}^{i,\text{idio}}. \end{aligned} \tag{13}$$

Suppose the following simple first-order spatial autoregressive relation holds between the idiosyncratic components of y_{mt}^i (the reasoning can be readily extended to a more general spatial autoregressive framework),

$$(I - \rho W) y_m^{\text{idio}} = \varepsilon_{it}, \tag{14}$$

using (13) we obtain

$$(I-\rho W)[y_m - \underline{\lambda}(L)\ \underline{\mathbf{f}}] = \varepsilon,$$

that is

$$y_m = \rho W y_m + (I-\rho W)\ \underline{\lambda}(L)\ \underline{\mathbf{f}} + \varepsilon. \tag{15}$$

Therefore, if we estimated (15) using

$$(I-\rho W) y_m = \varepsilon. \tag{16}$$

we would confine the term $(I-\rho W)\ \underline{\lambda}(L)\ \underline{\mathbf{f}}$ in the error, which would therefore be correlated with the spatial lag due to the presence of the common factors.

This problem arises as well in model (6) and (11). If we believe that interactions are local in nature, they occur (in the data generating process) among the idiosyncratic components of productivity,[16] as in

$$(I-\rho W) y_2^{\text{idio}} = \alpha_\nu + \gamma_1 \bar{y}_2^{s-\text{idio}} + \gamma_2 \bar{y}_2^{c-\text{idio}} + \varepsilon, \tag{17}$$

with $\bar{y}_{2t}^{i,s-\text{idio}} = \sum_{j\neq i} \omega_{ij}^s y_{2t}^{j,\text{idio}}$ and $\bar{y}_{2t}^{i,c-\text{idio}} = \sum_{j\neq i} \omega_{ij}^c y_{2t}^{j,\text{idio}}$. Plugging (13)

[16]Notice that we estimate (6) and (11) using fixed effects on the cluster structure, therefore we replace the constant term α with a set of fixed effects α_ν, which equals to estimate the using group-demeaned data.

into (17), we get, after some manipulation,

$$
\begin{aligned}
y_{2t}^i = & \alpha_\nu + \rho \sum_{j \neq i} w_1^{ij} y_{2t}^j + \gamma_1 \sum_{j \neq i} \omega_{ij}^s y_{2t}^j + \gamma_2 \sum_{j \neq i} \omega_{ij}^c y_{2t}^j \\
& + \underbrace{[1 - \rho \sum_{j \neq i} w_1^{ij} - \gamma_1 \sum_{j \neq i} \omega_{ij}^s + \gamma_2 \sum_{j \neq i} \omega_{ij}^c]}_{\delta} \ \underline{\lambda}(L) \ \underline{\mathbf{f}}_t + \varepsilon_{it},
\end{aligned}
$$

which shows that equation (17) can be properly estimated using (5) (or (6)) only if the common factors are included among the regressors.

4.2.1 The dynamic factor model

We consider the data generating model (13) where y_t^i is a zero mean covariance stationary vector stochastic process in $\Re^2$ (that is $y_t^i = (y_{1t}^i, y_{2t}^i)'$), $\underline{\mathbf{f}}_t = (\mathbf{f}_t^1, \mathbf{f}_t^2, \ldots \mathbf{f}_t^q)'$ is a column vector of q unit variance white noises (the 'common shocks'), $\underline{\lambda}(L)$ is a 2xq matrix of rational functions in the lag operator, and y_{it}^{idio} is a vector of 2 idiosyncratic shocks. In words, (13) poses that the realisations of the vector y for city i at time t can be written as the sum of the realisations of a common component ($\underline{\lambda}(L)$ $\underline{\mathbf{f}}_t$, which is a linear combination of q common shocks $\underline{\mathbf{f}}$'s) and those of an idiosyncratic component.

We can disentangle the common from the idiosyncratic component by assuming *(a)* that the city specific factors y_{it}^{idio} —which are possibly autocorrelated— are mildly correlated at all leads and lags and their variances are bounded from above; and *(b)* that the common shocks $\mathbf{f}_t$ are mutually orthogonal and orthogonal with respect to the idiosyncratic components. If that is the case, we can recover the common components by using the Law of Large Numbers. Indeed, since the idiosyncratic variances are bounded from above, by averaging y_{it} along the cross-section, the variance of the idiosyncratic component vanishes and the result converges in mean squares to the common component $\bar{y}_t = \underline{\lambda}(L)\ \underline{\mathbf{f}}_t$.

4.2.2 Number of common shocks

The first issue is to determine the rank q of the vector $\mathbf{f}_t$. To do this, we follow the 4-stage method proposed by Forni and Reichlin (1998), which consists in (i) choosing a partition of the set of cities i, (ii) averaging y_t^i within each subset of the partition, (iii) compute the spectral density of the resulting vector of averages and obtain its eigenvalues, (iv) choose q as the number of eigenvalues which explain at least 95% of the trace of the covariance matrix of the spectral density. The number of subsets of the partition has to be large enough to capture the number q of common shocks; at the same time it has to be small enough in order not to overstate q, since having a too fine partition tends to introduce too much noise. As to (i), we form subsets according to the clustering structure over y_{it} from section 2. This is different from Forni and Reichlin (1998), who instead adopt a random partition. Accordingly, our number of groups is endogenous, and based on an objective economic criterium.[17] As a result, we observe that

[17] We have as well tried to average over the groups resulting from the clustering exercise on income or productivity: the rank q does not change. When constructing averages on groups chosen randomly as in Forni and Reichlin (1998) we underestimate the number of common shocks.

two dynamic eigenvalues together represent 96% of the trace, and therefore we select $q = 2$ as the number of common shocks.

4.2.3 Estimation of components

Once the number of common shocks is known, we can estimate their observed counterparts, that is, the common components. Since the number of unobservable common shocks is the same as the number of the variables, then it is natural to assume that some dataset-wide linear aggregates of income ($\bar{y}_{1t}$) and productivity ($\bar{y}_{2t}$) are indeed the observable counterparts of the shocks.

In order to decompose each series of income and productivity into its common and idiosyncratic components, we estimate (equation by equation) the following disaggregated model:

$$y_{it}^{m} = \alpha_m^i + \beta_{m1}^i(L)\bar{y}_{1t} + \beta_{m2}^i(L)\bar{y}_{2t} + \eta_{mt}^i, \quad m = \{1, 2\} \tag{18}$$

where $\beta_{m1}^i(L)$ and $\beta_{m2}^i(L)$ are real valued polynomials in the lag operator.[18] To construct the aggregates, the weights are chosen such that the idiosyncratic component of variables vanishes through aggregation. This result is achieved by choosing the weights ω_m^i equal to $1/\mathrm{var}(\eta_{mt}^i)$. Since $\mathrm{var}(\eta_{mt}^i)$ is unknown, we start from $\omega_m^i = 1/\mathrm{var}(y_{mt}^i)$, then we get the estimated residuals and re-compute the weights. By iteration we converge to the required weights.

[18] Since we assume that the shocks are fundamental, the whole process can be written as a function of past innovations. Operationally, in the β-polynomials we include four lags and the contemporaneous value.

The aggregates are obtained as linear combination of the 311 series of income and productivity, using the vectors ω_1 and ω_2 as weights, as in

$$\overline{y}_{mt} = \sum_i \omega_m^i y_{mt}^i. \tag{19}$$

The common components ($y_{it}^{m,\text{common}}$) are represented by the fitted values of the above regressions:

$$y_{mt}^{i,\text{common}} = \hat{y}_{mt}^i = \hat{\alpha} + \hat{\beta}_{m1}(L)\bar{y}_{1t} + \hat{\beta}_{m2}(L)\bar{y}_{2t} = \underline{\lambda}_i^m(L)\underline{\mathbf{f}}_t$$

while the idiosyncratic components are recovered as the corresponding residuals:

$$y_{mt}^{i,\text{idio}} = y_{mt}^i - y_{mt}^{i,\text{common}}.$$

5 Results

Keeping into account all issues raised in section 4, we estimate the following four equations

$$(I - \rho W)y_2 = \alpha_\nu + \alpha_t + \underbrace{\delta_1 y_1^{\text{com}} + \delta_2 y_2^{\text{com}}}_{\delta z^{\text{com}}} + \gamma_1 \bar{y}_2^{\text{suppl}} + \gamma_2 \bar{y}_2^{\text{cust}} + \varepsilon \tag{20}$$

$$\begin{aligned}(I - \rho W)y_2 =& \alpha_\nu + \alpha_t + \delta z^{\text{com}} + \gamma_1 \bar{y}_2^{\text{suppl}} + \gamma_2 \bar{y}_2^{\text{cust}} + \zeta_1 W \bar{y}_2^{\text{suppl}} \\ &+ \zeta_2 W \bar{y}_2^{\text{cust}} + \varepsilon \end{aligned}\tag{21}$$

$$(I-\rho W)y_2 = \alpha_\nu + \alpha_t + \underbrace{\delta_1 y_1^{\text{com}} + \delta_2 y_2^{\text{com}}}_{\delta z^{\text{com}}} + \gamma_1 \bar{\bar{y}}\,_2^{\text{suppl}} + \gamma_2 \bar{\bar{y}}\,_2^{\text{cust}} + \varepsilon \tag{22}$$

$$(I-\rho W)y_2 = \alpha_\nu + \alpha_t + \delta z^{\text{com}} + \gamma_1 \bar{\bar{y}}\,_2^{\text{suppl}} + \gamma_2 \bar{\bar{y}}\,_2^{\text{cust}} + \zeta_1\, W\, \bar{\bar{y}}\,_2^{\text{suppl}} + \zeta_2\, W\, \bar{\bar{y}}\,_2^{\text{cust}} + \varepsilon \tag{23}$$

Equations (21) and (23) also estimate the effect of growth rates of suppliers and customers weighting their influence by the distance of the city where they are located. Since the weighting matrix we use here is based on the inverse of the squared bilateral distances, we end up giving a large weight to close suppliers and customers and a small weight to all others. In this sense coefficients ζ_1 and ζ_2 are estimates of the effect of the growth rate of productivity of close customers and suppliers on productivity of city i.

Results are rather consistent across all four specifications (see table 2). According to eq. (20) and (21), our most favourite specifications, once controls for specific time and clusters effect and for common shocks have been taken into account, market interactions with suppliers and customers weigh negatively in explaining cities' productivity growth (a 1% increase in customers' productivity growth subtracts around 0.7% to city productivity growth, while a 1% increase in suppliers' productivity growth subtract around 0.8% to city productivity growth).

At the same time market interactions with close suppliers and customers favour city i's productivity (a 1% increase in close customers productivity growth increases city i's productivity growth by 0.25%, while a 1% increase

Dependent variable: per worker personal income

	eq. (20)	eq. (21)	eq. (22)	eq. (23)
income common factor (δ_1)	-0.01 (-1.57)	-0.01 (-1.20)	-0.01 (-1.31)	-0.01 (-1.51)
per worker income common factor (δ_2)	0.47 (43.43)	0.46 (43.41)	0.91 (77.73)	0.92 (78.08)
externality from suppliers (γ_1)	-0.78 (-34.49)	-0.83 (-36.90)	-0.01 (-1.07)	0.00 (0.15)
externality from customers (γ_2)	-0.71 (-31.50)	-0.72 (-32.47)	-0.25 (-13.72)	-0.00 (-0.05)
externality from close suppliers (ζ_1)	-	0.55 (9.33)	-	0.02 (0.67)
externality from close customers (ζ_2)	-	0.25 (4.17)	-	-0.32 (-6.61)
spatial lag (ρ)	0.27 (21.30)	0.47 (26.17)	0.36 (23.32)	0.39 (24.74)
Adjusted R^2	0.78	0.79	0.87	0.87

Notes: t-statistics in parenthesis.

Table 2: Estimates of the spatial autoregressive model: Input-output matrix of 1985

in close suppliers productivity growth increases city i's productivity growth by 0.55%).[19] We take such estimates as evidence of the fact that increases in the productivity of city i's customers and suppliers located in city j have a positive effect on productivity of firms located in any city, decreasing with distance (direct effect); such increase of productivity in any city strengthens competition in the final goods markets (indirect effect), thereby possibly reducing some cities' market shares, which in turns lowers on city output

[19] All estimates are robust to the choice of the I-O matrix (see table 3).

and productivity. In other words, when considering market externalities from all customers and suppliers the indirect effect outweighs the direct one. The indirect effect is instead dominated when considering market externality only from close cities since close cities specialise in different productions to escape strong local competition. Far apart cities, which are sheltered by distance, may have similar specialisation, thus strengthening the indirect effect.

The negative overall effect of market-driven interactions should not surprise, since in a model with increasing returns to scale, imperfect competition and factor mobility, a single location might subtract market shares even to all other locations, thus inducing —even through relocation of human capital and the consequent knowledge spillovers— a negative overall effect on output and productivity. This latter is what is meant for core-periphery patterns of development, a concern which has not failed to raise policy debate. On the other hand, the effect captured by the parameters γ_1, γ_2 are idiosyncratic effects —and those captured by ζ_1, ζ_2 are local idiosyncratic effects, while all common effects are captured by the coefficient δ, therefore the overall positive effect on productivity (if any) of increased competition would be most likely captured by the latter parameter. Moreover, being a within estimate with fixed effect on clusters composition and time, the estimate should be intended as net of cluster and year averages.

All other non-market local interactions weight positively on the productivity growth of the city: all other things being equal, a 1% increase in all other (close) cities productivity growth increase city i's productivity growth between 0.27 and 0.47 percentage points.

When considering market interactions only with most concentrated sectors, only the market interaction with close customers and non-market interactions matter (eq. (23)).

Dependent variable: per worker personal income

	eq. (20)	eq. (21)	eq. (22)	eq. (23)
income common factor (δ_1)	0.00 (0.07)	-0.00 (0.10)	-0.01 (-1.30)	-0.01 (-2.13)
per worker income common factor (δ_2)	0.39 (39.10)	0.39 (39.01)	0.85 (74.29)	0.89 (75.02)
externality from suppliers (γ_1)	-0.83 (-21.22)	-0.86 (-25.23)	0.83 (13.37)	0.79 (12.75)
externality from customers (γ_2)	-0.97 (-24.54)	-0.99 (-25.23)	-0.91 (-14.96)	-0.78 (-12.69)
externality from close suppliers (ζ_1)	-	0.29 (2.65)	-	0.27 (1.76)
externality from close customers (ζ_2)	-	0.47 (4.21)	-	-0.49 (-3.34)
spatial lag (ρ)	0.27 (22.44)	0.44 (23.83)	0.25 (19.97)	0.35 (21.73)
Adjusted R^2	0.82	0.83	0.87	0.88

Notes: t-statistics in parenthesis.

Table 3: Estimates of the spatial autoregressive model: Input-output matrix of 1987

6 Concluding remarks

In this paper we have studied the origin of the comovements among the economic cycles of U.S. cities. In so doing, we have adopted a two-step procedure. First, we have analysed the clustering structure naturally arising in our dataset and observed that it does not seem to reflect just geographical proximity. Hence, using spatial econometrics, we have verified that sectoral input-output linkages also matter in explaining local productivity growth. Nevertheless, the importance of these linkages crucially depends on the distance among locations. In particular, once common shocks and specific time and space effects are taken into account, the effect of market interactions on the city's growth rate of per worker income is overall negative, but it is positive the effect of the interaction with close cities: this suggests that the direct positive effect of interaction with suppliers and customers is bounded in space, and —absent any barrier to trade— it is outweighed by a negative indirect effect of interaction, that is increased competition in the market for final goods. The inter-playing of direct and indirect effect might be related with the pattern of specialisation of each location, even if formal test of such conjecture is left for further analysis. Local business cycles are also transmitted by non market (face-to-face) interactions, whose effect is also bounded by geographical distance.

The policy relevance of the analysis performed here is apparent: since interactions both face-to-face and through markets are shown to be bounded by geographical distance, policies aiming at improving the access to peripheral regions might have important consequences on agglomeration and the decision of location.

References

Anselin, L. (1988). *Spatial econometrics: Models and methods*, Kluwer Academic Publishers, Dordrecht.

Bartelsman, E. J., Caballero, R. J. and Lyons, R. K. (1994). Customer- and supplier- driven externalities, *American Economic Review* **84**: 1075–84.

Brockwell, P. and Davis, R. (1991). *Time series: theory and methods*, Springer-Verlag, New York.

Conley, T. G. and Dupor, W. G. (2003). A spatial analysis of sectoral complementarity, *Journal of Political Economy* **111**(2): 311–352.

Forni, M. and Lippi, M. (1997). *Aggregation and The Microfoundations of Dynamic Macroeconomics*, Clarendon Press, Oxford.

Forni, M. and Reichlin, L. (1998). Let's get real: a dynamic factor analytical approach to disaggregated business cycle, *Review of Economic Studies* **65**: 453–473.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2001). The generalized factor model approach: identification and estimation, *Review of Economic and Statistics* **82**(4): 540–554.

Giannone, D. and Lenza, M. (2002). Explaining the Feldstein-Horioka puzzle, Mimeo, ECARES-ULB, Brussels.

Lamorgese, A. R. and Ottaviano, G. I. (2003). *Space, factors and spillovers*, Il commercio estero e la collocazione internazionale dell'economia italiana, Banca d'Italia and Università degli studi di Roma La Sapienza, Dipartimento di Scienze Economiche edn, pp. 175–196.

LeSage, J. P. (1999). The theory and practice of spatial econometrics, *Manuscript*, University of Toledo, OH. Available at www.spatial-econometrics.com.

Rodrigues, J. (1998). Clustering panels of inter-dependent time series in the frequency domain, Mimeo.

Sala, L. (2001). Monetary transmission in the Euro Area: a factor model approach, ECARES - Mimeo.

Stock, J. and Watson, M. (1999). Diffusion indexes, Unpublished manuscript.

Appendix 1 Data description

Employment is Total Full-time and Part-time Employment by Industry. Total Full-time and Part-time Employment by Industry (Table CA25) contains estimates of employment in Standard Industrial Classification (SIC) Division ("one-digit") detail. That's not a problem since we are disregarding sectors in this experiment. Employment is measured as the average annual number of jobs, full-time plus part-time; each job that a person holds is counted at full weight. The estimates are on a place-of-work basis. The estimates are organised both by type (wage and salary employment and self-employment) and by industry.

The source data for REMD's wage and salary employment estimates are from the Bureau of Labor Statistics (BLS) ES-202 series. The ES-202 series provides monthly employment and quarterly wages for each county in four-digit SIC detail. REMD releases local area employment estimates at the one-digit SIC level because self-employment is estimated– based mainly on data tabulated from individual and partnership income tax returns– at the one-digit level. (In the State annual series, however, the employment estimates are prepared and released at the SIC two-digit level.)

Personal income (Table CA05) is a measure of income received; therefore, estimates of State and local area personal income reflect the residence of the income recipients. The adjustment for residence is made to wages and salaries, other labor income, and personal contributions for social insurance, with minor exceptions, to place them on a place-of-residence (where-received) basis. The adjustment is necessary because these components of personal income are estimated from data that are reported by place of work (where earned). The estimates of proprietors' income, although presented on the table as part of place-of-work earnings, are largely by place of residence; no residence adjustment is made for this component. Net earnings by place of residence is calculated by subtracting personal contributions for social insurance from earnings by place of work and then adding the adjustment for residence, which is an estimate of the net inflow of the earnings of interarea commuters. The estimates of dividends, interest, and rent, and of transfer payments are prepared by place of residence only. **Total personal income** is the aggregate personal income received in the MSA.

Estimates of earnings by place of work are provided in CA05 at the two-digit Standard Industrial Classification (SIC) level. The principal source data for the wage and salary portion of REMD's earnings estimates are from the Bureau of Labor Statistics (BLS) ES-202 series. The ES-202 series provides monthly employment and quarterly wages for each county in four-digit SIC detail. REMD restricts its earnings estimates to the SIC Division ("one-digit") and two-digit levels and suppresses these estimates in many individual cases in order to preclude the disclosure of information about individual employers.

Wage and salary disbursements are defined as the monetary remuneration of employees. This remuneration includes the compensation of corporate officers (commissions, tips, and bonuses), voluntary employee contributions to certain deferred compensation plans (such as 401(k) plans), and receipts in kind, or pay-in-kind, that represent income. Wage and salary disbursements are measured before deductions, such as social security contributions and union dues, and they reflect the amount of wages and salaries disbursed, but not necessarily earned, during the year. The estimates are prepared, with a few exceptions, at the Standard Industrial Classification (SIC) two-digit level. Wage and salary disbursements accounted for about 57 percent of total personal income at the national level in 1993.

Other labour income consists of the payments by employers to privately administered benefit plans for their employees, the fees paid to corporate directors, and miscellaneous fees. The payments to private benefit plans account for more than 98 percent of other labour income. Other labour income excludes employer contributions for social insurance, which are paid to government-administered funds. Under the conventions of the national income and product accounts, the benefits paid from social insurance funds, not the employer contributions to the funds, are measured as part of personal income. These benefits are classified as transfer payments. Other labour income accounted for about 6.6 percent of total personal income at the national level in 1993.

Proprietors' income with inventory valuation and capital consumption adjustments is the current-production income (including the income in kind) of sole proprietorships and partnerships and of tax-exempt cooperatives.[20] Proprietors' income includes the imputed net rental income of owner-occupants of farm dwellings, but it excludes the dividends and the monetary interest that are received by nonfinancial business and the rental income received by persons not primarily engaged in the real estate business.[21] Proprietors' income accounted for approximately 8 percent of total personal income at the national level in 1993.

[20] A sole proprietorship is an unincorporated business owned by a person. A partnership is an unincorporated business association of two or more partners. A tax-exempt cooperative is a nonprofit business organisation that is collectively owned by its members.

[21] The dividends are included in personal dividend income, the monetary interest, in personal interest income, and the rental income, in rental income of persons.

Data are provided with a 2 digits SIC sectoral breakdown, which means that 82 sectors, disaggregated from 9 divisions (agricultural services forestry fisheries and other, mining, construction, manufacturing, transportation and public utilities, wholesale trade, retail trade, finance insurance and real estate services), are represented. Twenty two of them are manufactures.

Sector		Sector	
code	name	code	name
110	Agricultural services	380	Trans. equip. excl. motor vehicles
121	Forestry	390	Motor vehicles and equipment
122	Fisheries	410	Stone, clay, and glass products
140	Coal mining	420	Instruments and related products
150	Oil and gas extraction	430	Misc. manufacturing industries
160	Metal mining	450	Railroad transportation
170	Nonmetallic minerals, except fuels	470	Water transportation
181	General building contractors	481	Local & interurban passenger trans
182	Heavy construction contractors	482	Transportation by air
183	Special trade contractors	484	Transportation services
210	Food and kindred products	490	Communications
220	Textile mill products	500	Electric, gas, and sanitary service
230	Apparel and other textile products	510	Wholesale trade
240	Paper and allied products	520	Retail trade
250	Printing and publishing	540	Depository & non-dep. credit instit
260	Chemicals and allied products	553	Insurance carriers
270	Petroleum and coal products	570	Hotels and other lodging places
280	Tobacco products	580	Personal services
290	Rubber and misc. plastics products	601	Business services
300	Leather and leather products	602	Auto repair, services, and parking
320	Lumber and wood products	611	Amusement and recreation services
330	Furniture and fixtures	621	Health services
340	Primary metal industries	623	Educational services
350	Fabricated metal products	624	Social services
360	Machinery and computer equipment	627	Engineering and management services
370	Electric equipment, ex. computer e	660	State and local

Table 4: Sectoral breakdown in 1985

Sector		Sector	
code	name	code	name
81	Livestock and livestock products	390	Motor vehicles (passenger cars and trucks),
	Other agricultural products		Truck and bus bodies, trailers, and
120	Forestry and fishery products		motor vehicles parts
100	Agricultural, forestry, and	380	Aircraft and parts, Other transportation equipment
	fishery services	420	Scientific and controlling instruments, Ophthalmic
160	Metallic ores mining		and photographic equipment
140	Coal mining	430	Miscellaneous manufacturing
150	Crude petroleum and natural gas	450	Railroads and related services;
170	Nonmetallic minerals mining		passenger ground transportation
180	Construction	460	Motor freight transportation and warehousing
210	Food and kindred products	470	Water transportation
280	Tobacco products	482	Air transportation
220	Broad and narrow fabrics, yarn	483	Pipelines, freight forwarders, and related services
	and thread mills, Miscellaneous	490	Communications, except radio and TV,
	textile goods and floor coverings		Radio and TV broadcasting
230	Apparel, Miscellaneous	500	Electric services (utilities), Gas production and
	fabricated textile products		distribution (utilities), Water and
320	Lumber and wood products		sanitary services
330	Furniture and fixtures	510	Wholesale trade
240	Paper and allied products, except containers,	520	Retail trade
	Paperboard containers and boxes	530	Finance, Insurance, Owner-occupied
250	Newspapers and periodicals,		dwellings, real estate and royalties
	Other printing and publishing	570	Hotels and lodging places
260	Chemicals, Drugs, Paints and allied products	580	Personal and repair services (except auto)
	Cleaning and toilet preparations,	622	Legal, engineering, accounting, and related services
	Plastics and synthetic materials	601	Other business and professional services,
270	Petroleum refining and related products		except medical
290	Rubber and miscellaneous plastics products	527	Eating and drinking places
300	Footwear, leather, and leather products	524	Automotive repair and services
410	Glass and glass products, Stone and clay products	611	Amusements
340	Primary iron and steel manufacturing, Primary	621	Health services
	nonferrous metals manufacturing	623	Educational and social services,
350	Metal products, Screw machine		and membership organizations
	products and stampings, Engines and turbines,	640	Federal Government enterprises
	Other fabricated metal products	660	State and local government enterprises
360	Farm, construction, and mining machinery,	630	General government industry
	machinery and equipment	590	Household industry
370	Electrical industrial equipment and apparatus,		
	Household appliances		

Table 5: Sectoral breakdown in 1987