Applied Economics, 2002, **34**, 295–304

Exploring 3D datasets: a factorial matrices analysis of the US industry in the 1980s

STEFANO FACHIN*, DENNIS G. SHEA‡ and MAURIZIO VICHI§

University of Rome "La Sapienza", DCNAPS-Faculty of Statistics, P.le A. Moro 5, 00185, Roma, Italy, ‡ *Pennsylvania State University, Dept. Of Health Policy and Administration, 116 Henderson Building, University Park PA 16802, USA* and § *University of Rome "La Sapienza", DSPSA-Faculty of Statistics, P.le A. Moro 5, 00185 Roma, Italy*

In this paper a Factorial Matrices technique suitable for exploratory analysis of multivariate, disaggregated time series is presented and applied to a data set covering 19 US manufacturing industries over the years 1979 to 1990. The empirical analysis confirms that the technique is a powerful tool, allowing otherwise difficult extraction of stylized facts from multidimensional datasets. In this case: (i) there are no signs of deindustrialization induced by growing import penetration, and, (ii), employment decline has generally not been associated to substitution of capital to labour.

I. INTRODUCTION

The question at the origin of this paper is very simple: what has occurred in the US manufacturing industry during the 1980s? The special interest in this decade stems from its peculiar features: it included a complete business cycle, both of Reagan's terms of office, and structural changes that have started a heated debate on the risk of deindustrialization of the US economy. Though the question is simple, providing a simple answer is not an easy matter: stylized facts are certainly one of the most powerful ways of conveying the results of empirical economic analysis, but grand simplifications causing no significant loss of information are often very hard to achieve. Indeed, the exploratory tools commonly employed provide little help when the data span over three dimensions, as is the case with multivariate time series covering several different industries – and we have to look at this sort of data in order to answer our question. There is clearly a need of a statistical technique allowing exploratory analysis of 3D datasets. The aim of this paper is thus twofold: on one hand, developing a Factorial Matrices statistical technique suitable for 3D exploratory analysis; on the other, understanding by means of this technique what has occurred in the US manufacturing industry during the 1980s.

The following section will discuss in more detail the empirical problem, the third will present the Factorial Matrices technique employed to solve it, in the fourth report the empirical analysis, and finally in the fifth draw some conclusions.

II. WHAT HAPPENED IN THE US INDUSTRY IN THE 1980S?

As long as the study confines itself to the aggregate level, a clearly naive but nevertheless common approach, the path followed by the US manufacturing industry over the 1980s appears pretty simple to describe. Comparing the data at the end and the beginning of the period, we find that employment fell, output and capital stock grew (thus suggesting widespread capital–labour substitution), both exports and imports grew very fast, however, imports largely outpaced exports (at current prices, respectively +200.3% and +131.0%). This raised the fears of deindustrialization mentioned above, well represented for instance by the essay collected in Bernstein and Adler (1994), a volume by the revealing title *Understanding American Economic Decline*. However, these fears are not universally shared, with the opposite view stressing the global tendency

* Corresponding author. E-mail: s.fachin@caspur.it; Web: www.uniroma1.it/dcnaps/tachin/

Applied Economics ISSN 0003–6846 print/ISSN 1466–4283 online © 2002 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/0003684011004374 9

towards specialization and the relative, rather than absolute, nature of the American decline. For instance, Dollar and Wolff (1993) state that 'Evidence for the more extreme versions of the America-in-decline thesis is scant' (p. 17).

If specialization is brought into the discussion, then the study clearly needs to consider industry-level data. Unfortunately, in this case things are less easy to summarize. Although employment did fall and output did grow in almost all industries (exceptions being furniture, printing, and rubber for the former where employment actually grew, and leather and primary metals for the latter where output fell), the individual rates of growth have been very variable. In most industries output grew by less than 10% (in other terms less than 1% a year), whereas for instance in the machinery industry it more than doubled (+117% between 1980 and 1990).

The situation gets worse if the study tries to shed some light on these differences by examining joint paths. First of all, just considering the figures at the beginning and the end of the period is now much less satisfactory, as it may obscure important comovements (this may be particularly serious in the case of the highly cyclical foreign trade variables). On the other hand, looking at the entire time series of several individual industries is unpractical even at the exploratory stage. The study shall then accept, for the moment, this simplification, and proceed: the result is a rather confused picture. For instance, although capital grew on the aggregate, it did fall in some industries. Obviously, this would be neither a surprise nor a problem, were it not for the fact that some of these have followed completely different paths as far as the other variables are concerned. Consider, for instance, the leather and lumber industries: in the former dramatic falls in both output and employment took place, suggesting indeed deindustrialization, whereas in the latter output grew and employment declined, suggesting capital *and* labour saving technical progress. Further, if the study looks, for instance at the Rubber Industry, it finds that a massive growth in output (62%) has been achieved with a small increase in capital stock (17%), while in other industries the reverse is true – the large increases in capital stock thus seemingly aimed more at cutting labour inputs than at increasing capacity (as, e.g., in the electrical machinery industry: capital +91%, output +50%, employment −7%).

In fact, any hopes that the labour productivity patterns may be obviously related to the growth of capital/labour ratios are short-lived. Taking the machinery and electrical machinery industries as examples it was discovered that these variables may even be inversely related, as can be appreciated by the data reported in Table 1.

Table 1. *Growth of capital per worker and labour productivity in the machinery and electrical machinery industries, 1979–1990*

	Capital per worker	Labour productivity
Machinery	+96.8%	+166.1%
Electrical Machinery	+104.6%	+62.9%

Source: authors' computations on data from *Survey of Current Business*, various issues.

Table 2. *Growth of labour productivity and international trade flows at current prices in the machinery and lumber industries, 1979–1990*

	Labour productivity	Exports	Imports
Machinery	+166.1%	+104.2%	+320.4%
Lumber	+29.5%	+85.3%	+11.6%

Source: authors' computations on data from *Survey of Current Business*, various issues.

Capital grew slightly more in the electrical machinery industry, but the increase in labour productivity in this industry has been less than half that achieved in the machinery industry.

Finally, no obvious relationships exist either between labour productivity, exports and imports, as shown very clearly by the examples reported in Table 2.

Despite a remarkable increase in labour productivity, machinery imports grew more than three times exports; on the other hand, a moderate productivity gain (about 2% a year) in the lumber industry has been accompanied by a large rise in exports, with imports hardly growing at all.

The cross-correlations are thus blurred enough to suggest that this study needs to take into account all the information available, and not just the averages: obviously, the point is how to do this, since computing annual growth rates, for instance, would clearly leave open the problem of synthesizing the resulting mass of information.

Summing up, two conclusions have been reached about how to proceed in order to answer this question:

(i) when looking at time trends, all the single industries should be examined at once.[1] Neither examining the aggregate (with an *a priori* unwarranted loss of information), nor all the individual industries separately (excessively fragmenting the analysis) really helps in understanding what happened.

[1] Incidentally, note that this objective is implicit in the exploratory analyses of small amount of disaggregate data routinely carried out for instance by looking at tables of average growth rates. The problem here is how to achieve this aim with a rather large data set (10 years, 19 industries).

(ii) cross-section cross-correlation analysis analogously must retain as much as possible of the time series information without excessively fragmenting the analysis, to understand what happened in the period of interest taken as a whole (rather than year by year, or limiting the analysis at the growth between the first and the last year).

At this point the question has become somehow more specific; more precisely, it has split into two:

(i) 'which are the time paths of the chosen set of variables best summarizing those followed by all the individual industries, taken as a whole?' Once these paths are identified, further analysis can be easily carried out exactly as could be done on aggregate or average data; the difference is that they are necessarily more representative than the latter.
(ii) the second question can be put in the following form: 'which are the cross-section cross-correlations best summarizing the joint movements of the chosen set of variables over the given industries?' Answering this question will help in assessing, e.g., if import penetration has been significantly linked to output and employment decline.

The natural objection is that our desire to have both synthesis and disaggregation (over either industries or time) resembles closely that of having the cake and eating it.[2] However, there is a solution, provided by the class of the so-called *multivay factorial* techniques (see, e.g., Kroonenberg, 1983; Rizzi and Vichi, 1995a, b). These powerful methods have been recently developed in order to analyse data that can be represented as matrices having three or more dimensions: considering individual industries as described by time series of some key variables (such as capital, employment, etc.), the objective can indeed be seen as that of analysing a three-dimensions (time $\times$ variables industries) data matrix. The next section will present the FAMA (FActorial MAtrices) method, a technique falling in this class.

III. THE FAMA METHOD

Structure and Objectives

The FAMA (FActorial Matrices Analysis) technique is aimed to analyse multivariate–multisituation phenomena, i.e. phenomena described by k variables measured on a given group of n units on r different situations (e.g. points in space or time, data sources, etc.). These data are collected in 3D, or three-way, matrix of size $n \times k \times r$, formed by two dimensional marginal matrices of sizes $n \times k$ (units $\times$ variables, one for each situation; in fact a sequence of multiple cross-section series), $n \times r$ (units $\times$ situations, one for each variable), and $r \times k$ (situations $\times$ variables, one for each unit; when the situations are time periods each of these is a multiple time series describing the path followed by a given unit). Taking for instance the point of view of the $n \times k$ (units $\times$ variables) matrices, the three-way matrix can be seen as a sequence of r two-way data matrices, one for each situation. In this case, for each year starting in 1979 and ending in 1990 there will be one matrix with 19 rows (one for each industry) and as many columns as the variables considered.

In the common case when the situations are points in time, analysing the complete three way matrix allows evaluating simultaneously the variability across units and the dynamics of the phenomenon of interest. The latter is instead absent if each observation in time is analysed separately, or lost if the average matrix is considered. It is indeed the relevance given to the time dimension that distinguishes all multiway dynamic teqhniques, pioneered by Tucker (1966, 1972) with the Three-mode factor analysis and Escoufier (1977, 1987) with the STATIS method, from the traditional static multivariate ones. We will later motivate the preference given to FAMA rather than STATIS or other available methods; let us however first describe its structure.

FAMA can be described as a two-steps factorial technique aimed (like all the other multiway techniques) at reducing the vast amount of information contained in a three-way matrix into more manageable terms. Most precisely, the objective of the first step is either:

(1) extracting a small number of unobservable *virtual units*, linear combinations of the original units that best summarize (in the usual least squares sense) the information given by the overall variability of the data across situations;
(2) extracting a small number of unobservable *virtual situations*, linear combinations of the original situations that best summarize (in the usual least squares sense) the information given by the overall variability of the data across units.

The two-way matrices describing respectively the virtual units and the virtual situations are called *factorial matrices*, and have the same dimensions of the matrices describing respectively the original units and situations. In case (1) the original $n \times k \times r$ dimensions are thus reduced to $n' \times k \times r$, with $n' \ll n$; in the second, to $n \times k \times r'$, with $r' \ll r$.

[2] Galbraith and Du Pin Calmon (1994) express the same concept stating that 'there is no simple answer to a question for which no simple metric can exist'.

It is useful for illustrative purposes to consider the extreme cases when $n' = r' = 1$, and the situations are points in time. In the first case the result is a two-way variables $\times$ situations matrix, i.e. a single virtual multiple time series; in the second, a two-way units $\times$ variables matrix, i.e. a pure cross-section series. In other terms, the results of the first step would be to marginalize the three-way matrix with respect to either the units or the time dimension, exactly as done when aggregating or taking averages. The crucial difference is that the weights of the linear combination chosen by FAMA minimize the loss of information according to the least squares optimality criterion, whereas aggregating or averaging impose using equal weights to all the two-way matrices and may thus entail an unnecessary loss of information. Further, when clusters of strongly correlated units or situations are present FAMA will extract more than one factorial matrix (i.e., $n' > 1$ or $r' > 1$), allowing separate analysis of these pieces of information.

In both cases 1 and 2 the objective of the second step is extracting independent *latent factors* summarizing the information provided by the k variables in each of the n' and r' factorial matrices, thus eventually reducing the dimension of the three-way matrix to respectively $n' \times k' \times r$ and $n \times k' \times r'$, with $n' \ll n, r' \ll r$ and k'.

Summing up, and retaining the identification of the situations with time periods, in case 1 the three-way matrix is seen as n multiple time series, and FAMA first summarizes its information with $n' \ll n$ latent independent multiple time series, including decreasing amounts of information common to highly correlated subsets of the observed multiple time series.[3] In the second stage of the method these latent multiple time series are further synthesized by independent latent factors. The final result is a three-way matrix with a smaller number of both units and variables (unobservable and interpreted on the basis of the correlations with the respective counterparts) and the same time periods of the original matrix.

On the other hand, in case (2) the three-way matrix is seen as r multiple cross section series, and FAMA first summarizes its content through $r' \ll r$ latent independent multiple cross-section series, each summarizing decreasing amounts of information common to highly correlated subsets of the observed cross-sections. In the second stage of the method these latent multiple cross-sections are further synthesized by independent latent factors. The final result is a three-way matrix with a smaller number of both time periods and variables (unobservable and interpreted on the basis of the correlations with the respective counterparts) and the same units of the original matrix.

FAMA can thus reach several objectives for both quantitative and qualitative data:

(1) examining the cross-correlations between all the pairs of multiple time series included in a three-way matrix, i.e. the links between the time series of the various variables measured on a given unit;
(2) examining the cross-correlations between all the pairs of multiple variables describing a multivariate-multisituation phenomenon, i.e. the links between the data observed on a given set of units in different situations, for all pairs of situations;
(3) identifying virtual situations, summarizing the information common to different observed situations;
(4) extracting factors summarizing the information common to different observed variables in each virtual situation;
(5) identifying virtual units, summarizing the information common to different observed units;
(6) extracting factors summarizing the information common to different observed variables on each virtual unit;
(7) tracing the paths followed by the observed units in the orthogonal basis formed by the factors defined in (4) above;
(8) tracing the paths followed by the observed units in the orthogonal basis formed by the factors defined in (6) above.

Some existing multiway techniques can be shown to be special cases of FAMA, that also deals with aspects not considered by any of them. STATIS (L'Hermier Des Plantes, 1976, Escoufier, 1977, 1987), deals with data matrices transformed in dissimilarity matrices and reaches the objectives (3), (4) and (7) under the constraint of a single factorial matrix ('compromise matrix'), while the *Multistep Principal Component Analysis* (MPCA; Rizzi, 1989; Bodo *et al.*, 1993), is aimed at the objectives (3) and (4) for quantitative data.

Methodology

This study will now examine more specifically the three operative phases of the method. Note that for simplicity the study will identify the situations with time periods, but the procedure outlined applies to any type of situations (for instance, points in space rather than time).

Dependence Analysis. In this first phase objectives (1) and (2) above are reached using the index of dependence (more precisely, association, for categorical data, or correlation, for quantitative variables) best suited to measure the overall links between (a) the variables observed on the units in period h and m, with $h, m = 1, \ldots, r$; or, (b) the pairs of multiple time series h ed m, with

[3] Note that we are referring here to the concept of correlation among matrices (cf. e.g. Rizzi and Vichi, 1989).

$h, m = 1, \ldots, n$. The choice of the index clearly depends on the nature of the data. Several different measures can be derived from the generalized relative index:

$$\text{dep}\,(X_h, X_m) = \frac{(\text{vec}\,X_h')'C\,\text{vec}\,X_m'}{\sqrt{(\text{vec}\,X_h')'C_h\,\text{vec}\,X_h'(\text{vec}X_m')'C_m\,\text{vec}\,X_m'}} \tag{1}$$

Defining $C = C_h = C_m = I_x \otimes W$, where I_x is the identity matrix of either order r, in case (a), or n, in case (b), and W is a square matrix of order k. If we want to take into account the correlation between pairs of different variables (say, X_i and X_j) in the different situations h and m we have $w_{ij} = 1$, otherwise $w_{ij} = 0$. More specifically, letting $W = I_k$ returns the *weak matrix correlation coefficient* (used e.g. in the MPCA mentioned above), that takes into account exclusively the correlations between the values of the same variables in different units or situations. Instead if $W = 1_k 1_k'$, with $1_k'$ the unit vector of size k, there is a *strong matrix correlation coefficient*, measuring the cross-correlations between all the variables over the different situations or units (and which is thus clearly more satisfactory, especially in economic applications).

Incidentally, note that STASIS is based on Escoufier's (1973) RV index, delivered by (1) when the data matrices are replaced by the covariance or scalar product matrices and $C = C_h = C_m = I_x \otimes W$, where I_x and W as in the case of the weak correlation coefficient. Now, under these conditions (1) is not a measure of correlation any more, as it does not satisfy the conditions introduced by Renyi (1959) and discussed by e.g. Ramsay *et al*. (1984), but a measure of similarity, as defined by Robert and Escoufier (1976).

Before proceeding further, note that an interesting byproduct of this stage of the analysis might be representing the X_h matrices as points in a space defined by the chosen measure of correlation between the pairs of matrices (X_h, X_m) $h, m = 1, \ldots, r$ or n, and then searching for clusters of correlated units or situations through techniques of cluster analysis.

Synthesis. In the second phase the aims are objectives 3 and 5, i.e. the extraction of the factorial matrices (virtual units or virtual situations, represented by matrices respectively of the same dimensions of the units and situations matrices) able to best summarize the information included in the three-way matrix. The g-th factorial matrix F_g is a normalized linear combination of the matrices X_h ($h = 1, \ldots, c$, with $c = r$ or $c = n$), summarizing the g-th largest amount of dependence (measured by 1); it is independent by the first g-1 factorial matrices. Formally:

$$F_g = \sum_{h=1}^{c} \text{a}_{hg} X_h \tag{2}$$

so that:

$$\sum_{h=1}^{c}\sum_{m=1}^{c} \frac{(\text{vec}\,X_h')'C\,\text{vec}\,X_m'}{\sqrt{(\text{vec}\,X_h')'C_h\,\text{vec}\,X_h'(\text{vec}\,X_m')'C_m\,\text{vec}\,X_m'}}\text{a}_{hg}\text{a}_{mg} = \max \tag{3}$$

subject to:

$$\sum_{h=1}^{c} \text{a}_{hg}^2 = 1$$

$$\sum_{h=1}^{c}\sum_{m=1}^{c} \frac{(\text{vec}\,X_h')'C\,\text{vec}\,X_m'}{\sqrt{(\text{vec}\,X_h')'C_h\,\text{vec}\,X_h'(\text{vec}\,X_m')'C_m\,\text{vev}\,X_m'}}\text{a}_{hf}\text{a}_{ml} = 0,$$

where $f \neq l; f, l = 1, \ldots, g$.

The virtual units and situations thus obtained are interpreted on the basis of the correlations with the original data (respectively, units and situations) measured by the generalized index 1.[4]

For a comparison, note that in the MPCA, based on the weak matrix correlation coefficient, (3) maximizes the sum of the linear combinations of the variance of the k variables in the r situations or n units and of the linear combination of the covariances between the same variable observed in pairs of situations or units.

Finally, because of the transformation of the data matrices in covariance matrices, STATIS is not able to extract more than one virtual unit or situation. It is thus not advised for data showing a high degree of heterogeneity, which can, on the contrary, be properly handled by FAMA.

Singular value decomposition. This third and final phase is aimed at objectives (4), (6), (7) and (8) through factor analyses applied on each factorial matrix (i.e. virtual unit or virtual situation). The factors extracted allow an appreciation of the covariance structure of the original variables over the whole set of original situations, while the paths followed by the units in the factor space an interpretation of the variability over the situations (in case of time series, of the actual time trend).

[4] Vichi (1993) shows that (3) is $\text{dep}(F_g, F_g)$, and thus explains part of the variability of the three-way matrix evaluated by $\Sigma(X_h, X_h)$, where $h = 1, \ldots, c$.

IV. FACTORIAL MATRICES ANALYSIS: ESTIMATION AND RESULTS

This study will now examine the main steps of the application of the FAMA method[5] to the three-way matrix including the series of fixed nonresidential private gross capital at 1982 dollars, employment, output, capital/labour ratio, labour productivity, exports and imports (the last two variables at current prices because of the difficulties involved with the construction of a foreign trade deflator at the chosen level of disaggregation), for the 19 two-digits SIC manufacturing industries[6] for the period 1979–1990. Given that this study is interested in the dynamics, all data have been normalized with respect to 1979, so that the period effectively analysed is 1980–1990. Covering both output, factor use, technology and foreign trade these series should provide a good description of the evolution of the structure of the US industry over the period examined.

As discussed in the introduction to the method given in Section II above, in this case the FAMA method can be applied following two approaches: (i) synthesizing the industries in one or more *virtual industries* representing them as a whole in the best possible way; (ii) synthesizing the observed years in one or more *virtual years*, representing them as a whole in the best possible way. 'Best' is taken in the least squares sense. Once these virtual years or industries have been computed further analysis can be carried out, e.g. examining correlations, etc.; when the set of variables used is rather large, as indeed in our case, a factor analysis is a rather natural solution, and as such it is part of the original structure of FAMA.

The correspondence of the two approaches with the questions defined in the Introduction is evident. In case (i) the factor analysis of the two-way matrices (one for each virtual industry) years $\times$ variables leads to factors summarizing the variance–covariance structure over time of the original variables, whereas in case (ii) the two-way matrices are as many as the virtual years, their dimensions are industries $\times$ variables, and the factors thus summarize the variance–covariance stucture across industries. In econometric terms the two approaches thus correspond respectively to time series and cross-section analysis.

Given that the data are normalized on 1979, the only transformation needed was centering on the means over time (in case (i)) or across industries (in case (ii)). This study then proceeded to the analysis of dependence, using as a measure the strong correlation coefficient. Measuring the correlation of every variable with all the others over different years or industries,[7] this is clearly the most suited for our aims. This study now examines the results in more detail.

Identifying the virtual industries

In this case the aim is identifying the virtual industries best representing the information given by the 19 multiple time series describing the paths followed by the various industries over the period 1979–1990, so to reduce the dimensions of the data space (industries $\times$ variables $\times$ years) from $19 \times 7 \times 11$ to $n' \times 7 \times 11$, with $n' \ll 19$. The strong correlation coefficients were thus computed between all the possible pairs of industries and extracted the factorial matrices describing the virtual industries. It turns out that 89% of the total variability is explained by the first virtual industry, highly correlated with all the industries except Tobacco and Petroleum and Coal. The paths followed by the individual industries over the period examined have thus been so closely correlated that only the Tobacco and Petroleum and Coal industries, a clearly casual cluster, stand apart. The most distinctive feature of both these industries are very large increases in capital stock with falling employment and stagnating or falling output; indeed, similar combinations are not found in any other industry (cf. Table 3). This is confirmed by the finding that these two industries are the only ones significantly correlated with the second virtual industry, which explains more than half of the residual variability but little enough of the original one (6% out of a residual 11%, with an eigenvalue of 1.1 only marginally higher than one) that this paper will not examine it further.

Having found that reducing the dimension of the industries portion of the space from 19 to 1 does not cause any substantial loss of information, this study proceeded to a factor analysis of the first virtual industry, so to achieve a further reduction of the dimension of the variables portion of the space to some $k' \ll 7$. The results confirm the expectations induced by the simple exploratory analysis carried out in the Introduction: there is only one significant factor (i.e., $k' = 1$), explaining 86% of the variability of the observations in that portion of the data space, positively correlated with all variables except employment (cf. Table 4).

The picture obtained thus coincides with those obtained by the simple observation of the aggregate data: growth of capital, output, exports, imports, capital per worker and labour productivity, fall in employment. The value added of the analysis is, first of all, knowing that no other patterns with any overall significance are present

[5] All the analysis has been carried out using a programme by M. Vichi running under Windows 3.1.
[6] The complete list is reported in Table 6. All data are from *Survey of Current Business*.
[7] Obvious examples of correlations between different variables in the same industry are accelerator effects from output to capital, and between the same variable in different industries input–output links.

Table 3. *The US manufacturing industry in 1990 (1979 = 100)*

	K	Q	N	Exp	Imp	Q/N	K/N
Food	120.62	130.27	96.25	152.19	189.27	135.34	125.32
Tobacco	188.89	96.67	70.00	522.82	184.31	138.10	269.84
Textiles	91.53	106.00	78.08	170.66	374.01	135.76	117.23
Apparel	95.04	111.01	79.98	201.41	417.55	138.79	118.82
Lumber	91.37	122.69	94.76	185.31	111.60	129.48	96.43
Furniture	141.67	132.54	102.41	447.61	518.83	129.42	138.33
Paper	140.11	125.63	100.29	268.88	240.00	125.27	139.71
Printing	165.35	160.23	127.45	325.75	337.00	125.72	129.74
Chemicals	117.14	120.59	98.56	216.45	268.46	122.36	118.86
Petroleum	122.39	101.03	75.24	322.30	130.50	134.28	162.67
Rubber	117.06	162.30	108.28	291.35	369.72	149.88	108.11
Leather	89.66	60.73	53.66	325.06	319.35	113.18	167.08
Stone	100.00	104.13	82.64	200.30	265.92	126.01	121.01
Primary Metal	100.19	81.35	60.29	136.20	148.70	134.94	166.19
Fabricated Metal	127.05	106.33	83.07	191.54	272.62	128.00	152.94
Machinery	164.42	222.32	83.53	204.16	420.39	266.15	196.83
Electrical equipment	190.93	151.98	93.31	294.08	415.21	162.88	204.63
Transport equipment	121.50	130.25	96.16	261.71	351.74	135.44	126.35
Instruments	181.64	152.02	99.80	303.31	469.91	152.32	182.00

Source: authors' calculations on data from *Survey of Current Business*, various years. K: Capital; Q: Output; N: Employment; Exp: Exports; Imp: Imports; Q/N: Labour Productivity; K/N: Capital per worker.

Table 4. *Factor analysis of the first virtual industry*

Factors	Eigenvalue			Variability explained			
F_1	6.00			85.7%			
Correlations between factors and variables							
	K	Q	N	Exp	Imp	Q/N	K/N
F_1	−0.99	−0.96	0.72	−0.80	−0.99	−0.99	−0.97

Note: Variability explained by the virtual year: 89%. All variables indices 1979 = 100.

in the data, and, second, being able to assess how much of the total variability of this data set is not covered by this description: 23.7% ($= 1 - 0.89 \times 0.86$, where the former is the variability explained by the first virtual industry and the latter that explained by the first factor).

Identifying the virtual years

In this case it was computed that the strong correlation coefficients computed between all the pairs of years, i.e. between all the pairs of cross-section series relative to the various years. Then identified were the virtual years best representing this information; only the first two turned out to have eigenvalues greater than 1 (more precisely, respectively 8.10 and 2.14), and thus to be significant. The first virtual year found explains 74% of the variability of the original data and is a linear combination of the individual years with weights growing very slightly in absolute value with time (from −0.27, for 1980, to −0.34, for 1990); the second one explains 16.0% of the original variance, and, not surprisingly, gives more weight to the initial years than to the last ones. A break is evident between 1983 and 1984, close to the turning point in the so-called 'Reagan Recession'.

An optimal synthesis of the time series information is thus given by two virtual years, linear combination with varying weights of all the individual years; in other terms, the time dimension of the data space can be reduced with no significant loss of information from 11 to 2. Note how this contrasts with the naive procedure of simply taking averages, which reduces the dimensions to 1 using constant weights. The successive step, as above, was to carry out separate factor analyses for each of the two virtual years, trying to reduce the dimension of the variables portion of the space to more manageable terms. This study will now examine the results (reported respectively in Table 5 and 6) in turn.

In the case of the first virtual year the variables can be replaced by three factors, jointly explaining 87% of the data variability in this portion of space, and all having interesting economic meanings.

To begin, the first factor found across all industries shows first of all that across industries the growth of output is positively correlated to that of both inputs (capital

Table 5. *Factor analysis of the first virtual year*

Factors	Eigenvalue			Variability explained			
F_1	2.97			42.2%			
F_2	1.73			25.5%			
F_3	1.35			19.3%			
Correlations between factors and variables							
	K	Q	N	Exp	Imp	Q/N	K/N
F_1	0.88	0.80	0.62	0.60	0.65	0.56	0.34
F_2	0.35	−0.54	−0.59	0.41	0.09	−0.19	0.87
F_3	−0.09	0.24	−0.49	−0.58	−0.02	0.78	0.32

Note: Variability explained by the virtual year: 74%. all variables indices 1979 = 100.

and labour) and labour productivity, depicting a rather traditional image of virtuous links between growth, productivity, and inputs demand. In other terms, comparing different industries, higher productivity growth is not synonymous of larger employment cuts. Cases like that of the electrical machinery industry cited above are the exception and not the rule: the industries that have experienced the fastest growth in capital accumulation are *not* those that have the most substituted capital to labour. Indeed, the largest employment fall (−46% between 1979 and 1990) took place in the leather industry, a mature industry with very small scope for productivity growth (over the same period only +13%, the smallest increase in all the 19 industries), and the Printing Industry, ranking third in terms of capital growth, is the industry with the largest employment increase. Second, the industries where output, capital, employment, labour productivity and exports grew most – in other terms, the healthy ones – are those whose imports grew most as well, refuting the view of import penetration leading to or being a symptom of deindustrialization. This finding lends strong support to Dollar and Wolff's free trade stance ('trade openness, particularly *import* openness, is a major contributor to the growth of the country productivity', Dollar and Wolff, 1993; 184; authors' italics) and is consistent with Krugman (1979), where increasing 'love of variety' is assumed to foster inter-industry trade.

A good example here is the primary metal industry, where capital stagnated and output and employment fell (the latter dramatically: −40% between 1979 and 1990). It is certainly a good candidate for the deindustrialization story, however, the growth in imports has been smaller than in any other industry except lumber and petroleum. At the other extreme, the industry where imports grew most, the furniture industry, is actually one of the few in which employment in 1990 was higher than a decade earlier.

Explaining about 31% of the original variability of the data matrix (42%, variability explained by the factor, times 74%, variability explained by the first matrix) this is the most important single story, but obviously not the only one. The second factor, positively related to output and employment and negatively to capital per worker (the remaining correlations are negligible), summarizes about 25% of the residual variance (18% overall). The interpretation here is that the faster growth in output and labour input (given the overall decining trends, for the latter it is often the case of slower fall) took place in the industries where that of capital per worker has been slower. Industries lagging behind in terms of capital deepening, and thus becoming backward according to the traditional view of embedded technological progress, have nevertheless been successful in terms of output growth and job defense. Indeed, the only industries where employment grew are printing (+27%), rubber (+8%), and furniture (+2%), ranking resepctively 11th, 8th and 10th according to the growth of capital per worker. It is interesting to note that this is largely the most significant correlation shown by the growth of capital per worker (a positive, but weak, link with output growth appears in the first factor), which thus explains only marginally those in output growth, and with a sign opposite to the traditional view.[8]

Finally, the third factor may appear quite obvious, but it is nevertheless interesting in that it lends further support to the free trade stance: higher export growth implies higher employment growth (or smaller decline).

This study now moves to the second virtual year, which, as mentioned above, is more correlated with the initial years of the decade, and explains 19% of the variability of the data. The significant factors are three again; the correlations are basically similar to those discussed above, thus suggesting that there are no substantial differences in the cross-correlation between the variables examined in the whole decade and its first part. Without

[8] Note that it is the link between the growth of capital per worker and that of output that ensures that the negative correlation between the former and employment growth it is more than a trivial tautology stating that less capital/labour substitution implies less employment cuts.

Table 6. *Factor analysis of the first virtual year*

Factors	Eigenvalue			Variability explained			
F_1	2.70			38.5%			
F_2	2.09			29.9%			
F_3	1.25			17.8%			
Correlations between factors and variables							
	K	Q	N	Exp	Imp	Q/N	K/N
F_1	−0.30	0.63	0.81	−0.51	0.71	0.05	−0.89
F_2	−0.81	−0.75	−0.08	−0.20	−0.14	−0.81	−0.38
F_3	−0.31	−0.07	−0.53	−0.74	0.21	0.49	0.18

Note: Variability explained by the virtual year: 16%. all variables indices 1979 = 100.

going into much detail: import growth is (again) not associated with decline; industries with lower growth of capital per worker exhibit the largest output increases; there are strong links between capital and labour productivity dynamics; export matter for employment growth.

V. CONCLUSIONS

This paper now draws some conclusions. The objective of this empirical analysis was to extract some stylized facts summarizing the evolution of the US manufacturing industry over the 1980s; this objective can be stated more precisely as, first, finding the best summary description of all the different time paths followed by the individual industries over the 1980s in terms of output, inputs (capital and employment), technology (represented by labour productivity and capital per worker), and foriegn trade (exports and imports), and, second, finding the cross-section cross-correlations between the same variables best summarizing the joint movements in the same decade.

In order to answer to these questions this study developed an exploratory technique based on Factorial Matrices (hence the label FAMA), which is essentially aimed at allowing an easier analysis of multidimensional data sets by reducing the number of dimensions according to the usual least squares optimality criterion. As seen below, on the basis of this empirical analysis it can be concluded that FAMA is a powerful exploratory technique to be recommended when the data set of interest spans over three dimensions-typically, variables, time, industries.

With respect to the first of the two empirical objectives, the conclusion suggested by the application of FAMA is that the aggregate story (growth of capital, output, foreign trade in both directions, capital per worker and labour productivity, fall in employment), explaining more than 75% of the total variability of this set of variables over the 19 manufacturing industries for the period of 1979–1990, is actually a good description of the individual trends at two-digits SIC disaggregation level taken as a whole. Note how by applying FAMA we have been able to measure the loss of information induced by considering aggregate rather than two-digits disaggregated data. As far as the second objective is concerned, first of all, their findings lend strong support to the free trade stance: the imports grew most in the healthiest industries and export growth is associated to employment growth (or rather, given the overall declining trends, smaller fall). Second, on the whole faster capital accumulation does not appear to be linked to larger employment cuts. Finally, relatively 'backward' industries, ranking lowest in terms of growth of capital per worker, have been very successful in terms of output growth and job creation (or, again, in limiting job losses). Here the value added delivered by FAMA lies in the ability, first, to extract these cross-correlations, and, second, to measure the amount of unexplained data variance (about 10%).

So, was America declining in the 1980s? On the basis of this analysis, the answer is 'no'. The deindustrialization picture – shrinking capacity, falling output and employment, growing net imports – has virtually no role in explaining the disaggregate data for the manufacturing industry in that decade. Rather, it was changing: foreign trade was becoming much more important than it ever used to be and jobs were becoming scarcer and more common than in the past in the industries using technologies with lower capital intensity (which, obviously, does not necessarily mean low technology industries).

REFERENCES

Bernstein, M. A. and Alder, D. E. (1994) *Understanding American Economic Decline*. Cambridge University Press, Cambridge.

Bodo, G., D'Alessio, G. and Signorini, L. F. (1993) Patterns of change in Italian regional labour markets: a multistep principal components analysis, *Applied Economics*, **25**, 303–13.

Dollar, D. and Wolff, E. D. (1993) *Competitiveness, Convergence, and International Specialization*, MIT Press, Cambridge (Mass.).

Escoufier, Y. (1977) Operators related to a data matrix, *in Recent Developments in Statistics* (eds) J. R. Barra *et al*., North Holland, Amsterdam, pp. 125–31.

Escoufier, Y. (1987) Three-mode data analysis: the STATIS method, *Methods for Multidimensional Data Analysis*, ECAS, 259–72.

Galbraith, J. K. and Du Pin Calmon, P. (1994) Industries, trades and wages, in *Understanding American Economic Decline* (eds) M. A. Bernstein and D. E. Adler, Cambridge University Press, Cambridge, pp. 161–98.

Kroonberg, P. M. (1983) *Three-mode principal component analysis: theory and applications*, DSWO Press, Leiden.

Krugman, P. (1979) Increasing returns, monopolistic competition and international trade, *Journal of International Economics*, **9**, 469–79.

L'Hermier Des Plantes, H. (1976) *Structuration des Tableaux a Trois Indices de la Statistique*, PhD Thesis, University of Montpelier.

Renyi, A., On Measures of depence *Acta Amthematica of the Academy of Science of Hungary*, **10**, 441–51.

Rizzi, A. (1989) *On the Synthesis of Three-way Data Matrices*, The International Meeting Multiway '88, pp. 28–30.

Rizzi, A. and Vichi, M. (1995a) Representation, synthesis, variability and data preprocessing of a three-way data set, *Computational Statistics & Data Analysis*, **17**, 1–20.

Rizzi, A. and Vichi, M. (1995b) Three-way data set analysis, in *Some Relations Between Matrices and Structures of Multidimensional Data Analysis* (ed.) A. Rizzi. CNR Applied Mathematics Monographs, Giardini, Pisa, pp. 93–166.

Ramsay, J. O., Ten, J., Berge and Styan, G. P. H. (1984) Matrix correlation, *Psychometrika*, **49**, 403–23.

Robert, P. and Escoufier, Y. (1976) A unifying tool for linear multivariate statistical methods for the *RV* coefficient, *Applied Statistics*, **25**, 257–65.

Tucker, L. R. (1966) Some mathematical notes on the three-mode factor analysis, *Psychometrika*, **31**, 279–311.

Tucker, L. R. (1972) Relations between multidimensional scaling and three-mode factor analysis, *Psychometrika*, **37**, 3–27.

Vichi, M. (1993) Variability of a three-way data set, *Proceedings of the International Statistical Institute Conference – Florence*.