Laboratory of Economics and Management

Sant'Anna School of Advanced Studies
Piazza Martiri della Libertà, 33 - 56127 PISA (Italy)
Tel. +39-050-883-343 Fax +39-050-883-344
Email: lem@sssup.it Web Page: http://www.lem.sssup.it/

LEM

Working Paper Series

A Dynamic Factor Analysis of Business Cycle on Firm-Level Data

Lucia Alessi*
Matteo Barigozzi*
Marco Capasso*

* Scuola Superiore Sant'Anna, Pisa, Italy

2006/27 October 2006

ISSN (online) 2284-0400

A Dynamic Factor Analysis of Business Cycle on Firm-Level Data

Lucia Alessi*
Matteo Barigozzi†
Marco Capasso‡

Laboratory of Economics and Management
Sant'Anna School of Advanced Studies, Pisa

October 2006

Abstract

We use the Generalized Dynamic Factor Model proposed by Forni et al. [2000] in order to study the dynamics of the rate of growth of output and investment and establish stylized facts of business cycles. By using quarterly firm level data relative to 660 US firms for 20 years, we investigate the number and the features of the underlying forces leading economic growth: evidence suggests the main shock to be the same across sectors and for the economy as a whole. Moreover, we disentangle the component of industrial dynamics which is due to economy-wide factors, the *common component*, from the component which relates to sectoral or firm-specific phenomena, the *idiosyncratic component*. We assess the relative importance of these two components at different frequencies and compare common components across sectors. Finally, we investigate the comovements of the common component of output and investment series both at firm level and at sectoral level.

Keywords: Dynamic Factor Analysis, Business Cycle, Comovements
JEL-classification: C51, E32, O30

* *E-mail*: lucia.alessi@sssup.it
† *E-mail*: matteo.barigozzi@gmail.com
‡ *E-mail*: marco.capasso@gmail.com

We thank Giovanni Dosi, Marc Giannoni, Marco Lippi and all the participants to the IV Ph.D. Conference "Research in Economics" 2006 in Volterra for helpful comments and suggestions. We are also grateful to Marc Hallin and Roman Liška for providing us with their codes. Matteo Barigozzi also gratefully acknowledges the financial support to the research by the European Union Project "CO3 - Common Complex Collective Phenomena in Statistical Mechanics, Society, Economics, and Biology". Of course, the responsibility for any error is entirely our own.

1 Introduction

A variety of important issues on industrial dynamics and business cycle analysis can be addressed by means of factor analysis. Because of their own nature, factor models focus on the few driving forces underlying the dynamics of the economy, sectors and firms. Indeed, each economic series of interest is represented as the sum of a common component, driven by the non observable economy-wide factors, and an idiosyncratic component, which is peculiar to each series and vanishes via aggregation. Moreover, this kind of model allows using large cross-sections of time series, thereby taking into account as much information as possible. Finally, the whole dataset can be broken up into smaller subsamples - e.g. sectoral or country-specific - in order to carry out a factor analysis on each of the subsamples and compare the results across different subsets. This work gives a contribution in this sense and, more generally, in the sense of investigating the strength of comovements in the economy and give some insights about the nature of the shocks at the root of these comovements.

This paper is in the spirit of the work by Forni and Reichlin [1998], who apply a dynamic factor model *à la* Sargent and Sims [1977] for studying the dynamics of output and productivity in US manufacturing. To our knowledge, no other works apply dynamic factor analysis to firm-level data, which have the advantage of being by and large less difficult to measure than macro variables. However, our study is not limited to the manufacturing sector, being our wider samples composed by 660 firms (when analyzing sales data) and 355 firms (when analyzing investments data) belonging to all sectors of the US economy. Moreover, we use about 20 years of quarterly data, from 1984 to 2005. Besides, the most important difference is in the model itself: we apply the Generalized Dynamic Factor Model (GDFM) by Forni et al. [2000], which allows for a limited amount of cross-sectional correlation across idiosyncratic components. Indeed, we believe that the assumption of mutual orthogonality of idiosyncratic components might be too restrictive when dealing with balance-sheet data relative to sales and investments. Finally, we improve on the heuristic procedure for determining the number of common factors by implementing the information criterion recently proposed by Hallin and Liška [2007], which indicates the minimum number of common dynamic factors to include in the model.

The main contribution of this work consists in sheding some light on the nature of cyclical economic fluctuations, which is accomplished by tackling a variety of issues all at once. Firstly, we want to assess the stochastic dimension of firm growth. In other words, we want to determine the number of deep forces leading this economic phenomenon, i.e. the number of non observable factors common to the whole economy which lead the process of firm growth and output fluctuations, from a microeconomic and macroeconomic point of view respectively. The Hallin-Liška criterion indicates the existence of just one dynamic common factor driving sales and one dynamic common factor driving investments at the economy-wide level. Secondly, we investigate the nature of these main forces. Then, we estimate the common component and the idiosyncratic component of growth and investments, and assess their relative importance by means of a straightforward variance decomposition. Finally, we investigate the patterns of the common component both at the economy-wide level and at the sectoral level by means of spectral analysis.
We address the problem of whether sectoral shocks may produce fluctuations at business cycle frequencies for the whole economy in two ways. In the first place, we compare the economy-wide dynamic common factor with the main sectoral factors coming out from the

factor decomposition run sector by sector. Secondly, we study the spectral densities of the common component of output and investment series retrieved at the economy level and at the sectoral level. Indeed, it seems there is just one main force governing the economy as a whole and each of the sectors, by the meaning that the many sectoral factors ultimately resemble the the dynamic common factor of the economy, although the issue regarding the nature of the latter deserves further investigation.
Finally, we obtain information with regard to the features of positive and negative comovements in the economy. We run this kind of analysis both at firm level, both on firms belonging to the same sector and on firms belonging to different sectors, and at sectoral level. In the first case, the evidence relative to output is consistent with the existence of some sort of competition mechanism and selection process which gets more stringent in the short run. In the second case, we check which sectors comove positively and which negatively, and at which frequencies. The spectral analysis of sectoral comovements, when applied to investments, provides information on the mapping of capital flows in the economy. On the other hand, the study of comovements in sales, assuming output as a crude proxy of employment, detects job flows.

Let us spend some words on the motivation underlying the choice of adopting a dynamic factor approach. In a nutshell, by reducing the dimension of the problem under study it combines the virtues of panel data techniques and VAR models. On this, Reichlin [2002] states:

> "Modern macroeconomic theory is based on the representative agent assumption, but macroeconomic empirics is mostly based on aggregate data. What is the cost of simplicity, i.e. are we losing valuable information by working with econometrics models containing few aggregate variables? How detailed do our models have to be to have a chance to provide the essential information on the macroeconomy? To try to answer these questions there is a need to develop econometric models which (a) are able to handle the analysis of many time series by reducing the number of the essential parameters to estimate; (b) can provide an answer on what is the relevant stochastic dimension of a large economy, i.e. on how many aggregate shocks are needed to study the macroeconomy which emerges from the behavior of many agents; (c) can help us identifying these (possibly few) shocks and studying the propagation mechanism through agents or through geographical space. This is what will help to bridge the gap between purely time series studies and the cross-sectional approach."

Indeed, the dynamic factor approach presents basically three big advantages with respect to the traditional VAR models (see Forni et al. [2007]):

- it allows dealing with much larger datasets, virtually infinite, i.e. a wider information set, virtually the whole economy;
- it does not need to impose any restriction in order to disentangle the common part and the idiosyncratic part of each series in the sample;
- it needs only a few identification restrictions for the structural shocks, being their number very small and equal to the number of underlying common factors, while in VAR models the number of structural shocks equals the number of series in the dataset.

Ultimately, dynamic factor models allow for an easier economic interpretation. Moreover, they do not suffer from two shortcomings of the VAR approach, i.e. the sensitivity to the choice of the variables to include in the analysis and the substantial arbitrarity of the identification restrictions. In this work we remain relatively agnostic and refrain from imposing any kind of theory-driven structure to achieve identification of the shocks, although a simple DSGE model would serve the purpose of offering a theoretical framework for our empirical analysis. However, we find the other arguments to be convincing enough for adopting the dynamic factor approach in order to empirically ground our economic research.
The novelty of our study with respect to the factor models literature stands in the use of the GDFM to investigate industrial dynamics issues and for business cycle analysis. The GDFM generalizes on the one hand the dynamic factor model proposed by Sargent and Sims [1977] and Geweke [1977] by allowing for mildly correlated idiosyncratic components; on the other hand the approximate factor model by Chamberlain [1983] and Chamberlain and Rothschild [1983] which is static. Indeed, static principal components along the lines of Stock and Watson [2002] are actually used in many of the works extracting common factors from large cross-sections (e.g. Marcellino et al. [2003], Beck et al. [2006], Eickmeier and Breitung [2006]), exceptions being for example Giannone et al. [2004], Giannone et al. [2002], and Sala [2002][1]. The key difference between the GDFM approach and the static principal component method is that the latter gives a static representation of the dynamic model, one shock and its lags being treated as many different shocks. In the GDFM, on the opposite, the time element is introduced in full into the analysis, the focus of the estimation being on the dynamic common factors rather than on their static counterparts.

The paper is structured as follows. In the following section we outline the GDFM and the estimation procedure. In section 3 we present the COMPUSTAT data and the deflators we use for building the dataset for the empirical analysis. For determining the number of factors to include in the model, in section 4 we apply on all our samples and subsamples both the criterion by Hallin and Liška and a heuristic procedure. In section 5 we retrieve the dynamic common factors and compare the main economy-wide factor to the sectoral ones. In section 6 we estimate the common and the idiosyncratic component of each series and study their spectral profile and their cospectra. Section 7 concludes and proposes complementary analyses and empirical applications.

2 The Model

We denote as $x_t = (x_{1t} \dots x_{Nt})'$ an N-dimensional vector process. Each of the series is stationary and second order moments $\gamma_{ik} = E[x_{it} x'_{it-k}]$ exist finite for all i and k. In the Generalized Dynamic Factor Model (GDFM), as proposed by Forni et al. [2000], it is assumed that each series x_{it} can be written as the sum of two mutually orthogonal unobservable components, the *common component* χ_{it} and the *idiosyncratic component* ξ_{it}. The common component is driven by a small number q of dynamic common factors or shocks u_{jt} with $j = 1, \dots, q$, which are loaded with possibly different coefficients and lags. Formally:

$$x_{it} = \chi_{it} + \xi_{it} = \sum b_{ij}(L) u_{jt} + \xi_{it} \qquad i = 1, \dots, N \quad j = 1' \dots, q \tag{1}$$

[1]Recent works on dynamic factor models within a Bayesian framework are, among others, Kose et al. [2003], Canova et al. [2007], Ciccarelli and Mojon [2005].

The q-dimensional vector process $u_t = (u_{1t} \dots u_{qt})'$ is an orthonormal white noise. The N-dimensional vector process $\xi_t = (\xi_{1t} \dots \xi_{Nt})'$ has zero mean and is stationary. Moreover, ξ_{it} is orthogonal to u_{jt-k} for all k, i and j. The polynomials in the lag operator $b_{i1}(L) \dots b_{iq}(L)$ are square-summable, two-sided filters in principle of infinte order.
We denote the spectral density matrices of the common part and the idiosyncratic part respectively as $\Sigma^{\chi}(\theta)$ and $\Sigma^{\xi}(\theta)$, with $\theta \in [-\pi, +\pi]$. Finally, we assume that the eigenvalues of $\Sigma^{\chi}(\theta)$ diverge almost everywhere while the largest eigenvalue of $\Sigma^{\xi}(\theta)$ is bounded as the number of series goes to infinity. This last condition, in other words, relaxes the assumption of mutual orthogonality of idiosyncratic components by allowing for a limited amount of cross-sectional correlation.

The estimation of the model follows the procedure proposed in Forni et al. [2000]. Firstly, the spectral density matrix of x_t, $\hat{\Sigma}^x(\theta)$, is estimated by applying the Fourier transform to the sample covariance matrices $\hat{\Gamma}_k$ (hereafter, all "hatted" symbols denote estimated values). Then the dynamic principal component decomposition is applied, thereby selecting the first q largest eigenvalues of $\hat{\Sigma}^x(\theta)$ and the corresponding eigenvectors $(p_1(\theta), \dots, p_q(\theta))$. From such eigenvectors we can build the corresponding filters using the inverse Fourier transform:

$$\tilde{p}_j(L) = \frac{1}{2\pi} \sum_{k=-\infty}^{+\infty} \left[\int_{-\pi}^{\pi} p_j(\theta) e^{i\theta k} \mathrm{d}\theta \right] L^k \tag{2}$$

The factor space is the minimal closed subspace that contains the q largest dynamic principal components: $U = \overline{\text{span}}\, \{\tilde{p}_j(L) x_t \ , \ j = 1 \dots q\}$; and the common part is just the projection of x_t onto the factor space. Given the orthonormality of the dynamic eigenvectors the following holds:

$$x_t = \left[\sum_{j=1}^{N} \tilde{p}_j^{\dagger}(L) \tilde{p}_j(L) \right] x_t$$

where with $p^{\dagger}$ we mean the transposed and complex conjugate of p. Finally, since by assumption the common and the idiosyncratic part are orthogonal, we have:

$$\chi_t = \left[\sum_{j=1}^{q} \tilde{p}_j^{\dagger}(L) \tilde{p}_j(L) \right] x_t \equiv K(L) x_t \tag{3}$$

We then obtain the idiosyncratic component simply as difference between the original series x_t and χ_t.
Note that the filter $K(L)$ is two-sided and in principle is also of infinite order, but for $t \leq 0$ and $t \geq T$ we do not have observations for x_t, hence we need to truncate the filter. Such operation will cause the loss of part of the variance of χ_t even for $N, T \to \infty$ so we must concentrate only on the central part of x_t. In practice this implies the choice of a truncation lag M_T such that consistency of all the estimates is preserved[2]; this is ensured provided that $M_T \to \infty$ and $M_T/T \to 0$ as $T \to \infty$ (we choose $M_T \sim [\sqrt{T}/2]$).
The estimated filter that in practice we use is:

$$\hat{K}(L) = \frac{1}{2M_T + 1} \sum_{k=-M_T}^{M_T} \left[\sum_{h=-M_F}^{M_F} \left(\sum_{j=1}^{q} p_j^{\dagger}(\theta_h) p_j(\theta_h) \right) e^{ik\theta_h} \right] L^k \tag{4}$$

[2]See Forni et al. [2000] for a detailed discussion of the problem. Moreover, estimations of the common component at the beginning and at the end of the sample are not reliable, thus such method is useless for prediction while it is good for structural analysis as it is in our case.

where $\theta_h = 2\pi h/(2M_F + 1)$ are the $2M_F + 1$ points for which the spectral density is estimated (we take $M_F = 10$).

A number of studies compare the one-sided estimation of the GDFM by Forni et al. [2005] with the dynamic factor model estimated by static principal components *à la* Stock and Watson [2002][3]: the evidence is mixed. We choose to carry out our analysis by estimating dynamic principal components rather than static principal components because the first exploit the information contained in lagged covariance matrices, while the static method makes use of contemporaneous covariances only. Indeed, while less sophisticated estimation procedures may turn helpful in those contexts where computational simplicity plays a role, when we come to structural analysis it becomes crucial to tell what the dynamic factors are rather than estimating the corresponding static factors, the dynamic factors being the primitive shocks which get then transmitted to the whole economy.

3 The Data

For the empirical analysis we use COMPUSTAT quarterly data relative to sales, by means of which we intend to proxy firm size, and investments of US firms. The whole COMPUSTAT database contains data on about 10,000 actively traded U.S. companies, standardized by specific data item definition and by financial statement in order to allow for intertemporal and inter-firm comparison. Given the large amount of data at our disposal, we can exclude from the analysis all those series presenting problems concerning missing values in the period under analysis. Moreover, we do not replace abnormal values because we cannot exclude their structural nature: for example, in the case of investment, they may be due to lumpiness. Therefore, the database we use is not affected by any arbitrary manipulation. What follows is the COMPUSTAT definition of sales and investments:

- Sales: this item represents gross sales (the amount of actual billings to customers for regular sales completed during the period) reduced by cash discounts, trade discounts, and returned sales and allowances for which credit is given to customers. The result is the amount of money received from the normal operations of the business (i.e. those expected to generate revenue for the life of the company).

- Capital expenditures: this item represents cash outflow or funds used for additions to the company's property, plant, and equipment, excluding amounts arising from acquisitions.

We do not consider those firms which have been affected at any point in the considered time span by any merger or acquisition or any kind of accounting changes, as well as those data which present discrepancies with respect to the standard definition (see Appendix A for a detailed description of how the data are built from balance-sheet items and a list of possible anomalies). The firms surviving the above selection have been grouped into two distinct datasets: the dataset for sales goes from the second quarter 1985 to the first quarter 2005 - 80 observations in total - and includes 57 firms; the dataset for investments goes from 1984 first quarter to 2005 first quarter - 85 observations in total - and includes 355 firms. The difference in the number of firms in the two samples is due to the fact that often, while the investment time series is long enough, the sales time series is too short for the same firm, so we had to drop those firms from the sales sample. However, since the cross dimension plays a crucial

[3] See for example Boivin and Ng [2005], Kapetanios and Marcellino [2004], D'Agostino and Giannone [2006].

role for the consistency results of the dynamic factor model we use, we run the analysis on a wider sales dataset as well, including 660 firms. In this latter dataset those firms experiencing significant mergers or acquisitions, whereby the effects on the prior year's sales constitute 50% or more of the reported sales for that year, are not included. Running the analysis on two sales datasets, i.e. the cleanest possible and the widest possible, ensures the accuracy of the results because both quality and quantity of the data are taken into account.
We are interested in inter-sectoral comparisons too. Therefore, we break up the two big samples into sectoral subsamples and apply the GDFM to the more numerous subsets. For the definition of sectors we follow the North American Industrial Classification System (NAICS), introduced in 1997, which is more disaggregated than the Standard Industrial Classification (SIC) and better designed. Moreover, it is more detailed on key sectors belonging to services and IT[4]. Table 1 summarizes the composition of the two bigger samples representing the whole economy at the 3-digit level for sales and investments.
We focus on the following four sectors, identified on the basis of the 3-digit NAICS code, which include a sufficient number of firms for applying the GDFM:

- Utilities (NAICS code: 221—): 109 firms in the sales subsample, investment subsample too small;
- Computer and electronic product manufacturing (NAICS code: 334—): 89 firms in the sales subsample, investments subsample too small;
- Chemical manufacturing (NAICS code: 325—): 80 firms in the sales subsample, 37 firms in the investments subsample;
- Machinery manufacturing (NAICS code: 333—): 61 firms in the sales subsample, investments subsample too small.

The raw data are deflated by means of the deflator series published by the U.S. Department of Labor - Bureau of Labor Statistics (BLS). For deflating capital expenditure series we use the Producer Price Index Finished Goods - Capital Equipment with base year 1982. For deflating output series, we use the Producer Price Index Revision - Current Series, which are NAICS-based sectoral deflators. According to the definition, they "reflect price movements for the net output of producers [...]. To the extent possible, prices used in constructing the indexes are the actual revenue or net transaction prices producers receive for sales of their outputs [...]. The PPI is meant to measure changes in prices received by domestic producers, import products are not priced in the survey". We are able to match firms and deflators with a good deal of accuracy, being the great majority of the codes at the 6-digit level. In some cases, for instance in the case of big enterprises active in more than one sector or when the 6-digit NAICS code deflator is not available, we take deflators relative to wider industry definitions, keeping the 3-digit level as the maximum limit for aggregation.
Series are seasonally adjusted by a simple moving average method, again aiming at manipulating the raw data as less as possible (see Appendix B for details). Finally, series are differenced in order to get stationarity and standardized, so that ultimately we work with rates of growth.

[4] Those firms belonging to some very recent niches, as the "Dot-com" enterprises in the computer manufacturing sector, are absent in our dataset.

3-digit NAICS code	Classification	Sales		Investments	
		No. of firms	Percentage	No. of firms	Percentage
111	Crop Production	0	0	2	0.56%
211	Oil and gas extraction	0	0	11	3.10%
212	Mining (except oil and gas)	8	1.21%	4	1.13%
213	Mining support activities	0	0	5	1.41%
221	Utilities	109	16.52%	9	2.54%
237	Heavy and Civil Engineering Construction	0	0	5	1.41%
311	Food manufacturing	21	3.18%	15	4.23%
312	Beverage and tobacco product manufacturing	7	1.06%	3	0.85%
313	Textile mills	3	0.45%	2	0.56%
314	Textile product mills	2	0.30%	2	0.56%
315	Apparel manufacturing	13	1.97%	1	0.28%
316	Leather and allied product manufacturing	10	1.52%	2	0.56%
321	Wood product manufacturing	6	0.91%	4	1.13%
322	Paper manufacturing	23	3.48%	17	4.79%
323	Printing and related support activities	9	1.36%	5	1.41%
324	Petroleum and coal products manufacturing	14	2.12%	7	1.97%
325	Chemical manufacturing	80	12.12%	37	10.42%
326	Plastics and rubber products mfg	16	2.42%	7	1.97%
327	Nonmetallic mineral product mfg	13	1.97%	6	1.69%
331	Primary metal manufacturing	15	2.27%	6	1.69%
332	Fabricated metal product mfg	35	5.30%	13	3.66%
333	Machinery manufacturing	61	9.24%	18	5.07%
334	Computer and electronic product mfg	89	13.48%	32	9.01%
335	Elec equip, appliance, and component mfg	32	4.85%	13	3.66%
336	Transportation equipment mfg	30	4.55%	14	3.94%
337	Furniture and related product mfg	11	1.67%	4	1.13%
339	Miscellaneous manufacturing	33	5.00%	8	2.25%
423	Merchant Wholesalers, Durable Goods	0	0	11	3.10%
424	Merchant Wholesalers, Nondurable Goods	0	0	8	2.25%
441	Motor vehicle and parts dealers	0	0	1	0.28%
442	Furniture and home furnishing stores	0	0	1	0.28%
444	Building material & garden equipment & supply dealers	0	0	1	0.28%
445	Food and beverage stores	0	0	7	1.97%
446	Health and personal care stores	0	0	3	0.85%
448	Clothing and clothing accessories stores	0	0	6	1.69%
452	General merchandise stores	0	0	5	1.41%
454	Nonstore retailers	0	0	1	0.28%
481	Air transportation	0	0	3	0.85%
482	Rail transportation	8	1.21%	5	1.41%
483	Water transportation	0	0	1	0.28%
484	Truck transportation	0	0	2	0.56%
492	Couriers and messengers	0	0	1	0.28%
511	Publishing industries (except internet)	12	1.82%	7	1.97%
515	Broadcasting (except internet)	0	0	1	0.28%
517	Telecommunications	0	0	3	0.85%
522	Credit Intermediation and Related Activities	0	0	3	0.85%
523	Finance	0	0	2	0.56%
524	Insurance carriers & related activities	0	0	3	0.85%
531	Real Estate	0	0	1	0.28%
532	Rental and leasing services	0	0	3	0.85%
533	Lessors of Nonfinancial Intangible Assets	0	0	1	0.28%
541	Professional and technical services	0	0	6	1.69%
561	Administrative and support services	0	0	3	0.85%
621	Ambulatory Health Care Services	0	0	1	0.28%
713	Amusement, Gambling, and Recreation Industries	0	0	1	0.28%
721	Accommodation	0	0	3	0.85%
722	Food Services and Drinking Places	0	0	6	1.69%
999	Unclassified	0	0	3	0.85%

Table 1: Firms included in the analysis disaggregated at the 3-digit level.

4 Determining the Number of Factors

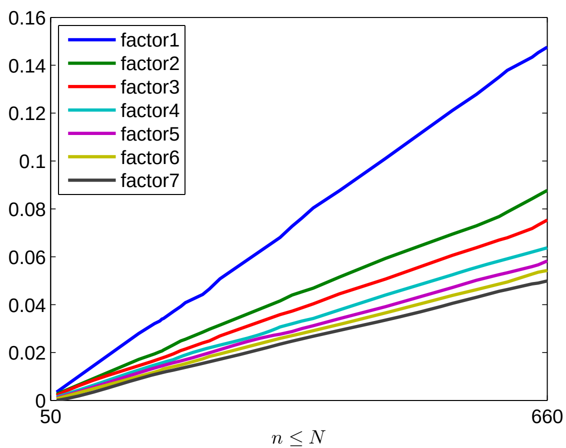
Firstly, we verify that our dataset does fulfill GDFM assumptions on the eigenvalues of the spectral density matrix of x_t. According to Brillinger [1981], we define the variance explained by the i^{th} dynamic factor, associated to the i^{th} largest eigenvalue $\lambda_i(\theta)$ of $\hat{\Sigma}^x(\theta)$, as:

$$EV_i = \frac{\int_{-\theta^*}^{\theta^*} \lambda_i(\theta) d\theta}{\sum_{j=1}^{N} \int_{-\theta^*}^{\theta^*} \lambda_j(\theta) d\theta} \qquad \theta^* = \pi, \pi/2, \pi/4, \pi/6 \tag{5}$$

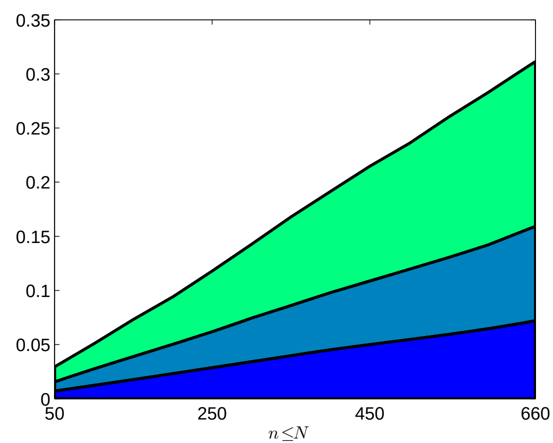
We require that, as $N \longrightarrow \infty$:

$$\begin{cases} EV_i \longrightarrow \infty & & \text{for} & i = 1, \dots, q \\ \exists\, M \in \mathbb{R}^+ & \text{s.t. } EV_i \leq M & \text{for} & i = q+1 \end{cases} \tag{6}$$

Indeed, this is the case for all samples. Figure 1 shows the explained variance and the cumulated explained variance computed in $[-\pi, +\pi]$ relative to the first eigenvalues for the 660 firms sales sample. These plots are built for $n = 50, \dots, 660$: starting from including just the firms belonging to the first sector listed in table 1 we compute the variance explained by the first eigenvectors and the cumulated explained variance. Then, at each step we add all the firms belonging to the following sector and re-compute the quantities.



(a) Explained variance

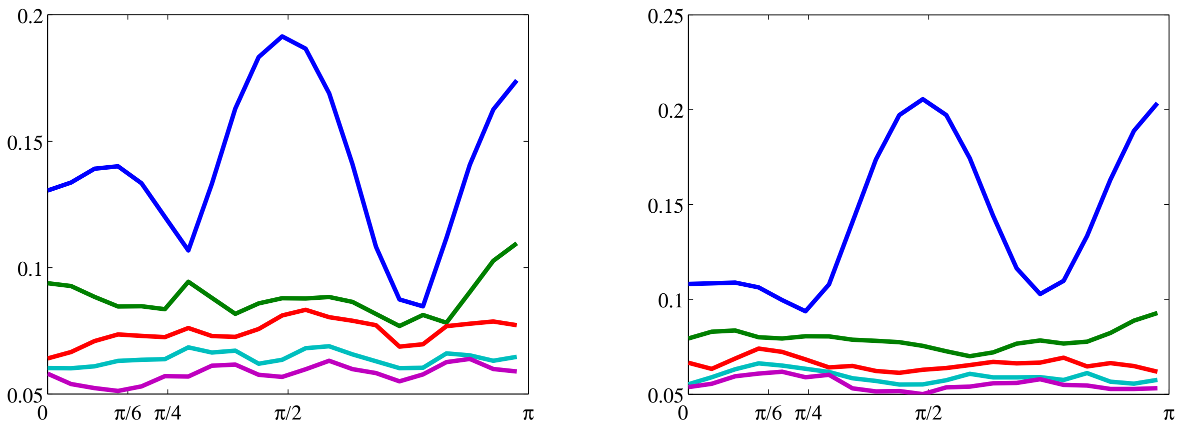


(b) Cumulated explained variance

Figure 1: Explained variance and cumulated explained variance by the first eigenvalues for the sales sample, 660 firms.

Moreover, we compute the explained variance and the cumulated explained variance relative to the first q eigenvectors also in the narrower frequency bands $[-\pi/2, +\pi/2]$ (horizon longer than one year), $[-\pi/4, +\pi/4]$ (horizon longer than two years) and $[-\pi/6, +\pi/6]$ (horizon longer than three years). Results are shown in tables 2 and 3: the bulk of the variance in sales data is generally concentrated in the $[-\pi/2, +\pi/2]$ frequency band except for the smaller economy-wide sample and the chemical industry, which are the only cases in which the variance explained by the first q eigenvalues decreases when we compute it only for periods longer than 1 year. On the opposite, the bulk of the variance in investments data is at higher frequencies: the amount of cumulated explained variance systematically decreases when we narrow the band keeping just lower frequencies.

Note that the first eigenvalue is actually much larger than all the others at all frequencies and presents no big jumps between adjacent frequencies. This suggests the presence of at least one common dynamic factor that explains most of the variance and has the same economic interpretation for all frequencies. Plots for sectors look as those in figure 2, except for the utilities sector where the first and the second eigenvalue are closer to each other. The peak at $\pi/2$ in the plots might be partly due to some residual seasonality that has not been washed away by the deseasonalization operated univariately on the original data.



(a) Sales, 660 firms

(b) Investments, 355 firms

Figure 2: Largest eigenvalues (frequencies on the horizontal axis).

However, we do not rely only on the intuition coming from the graphs for determining the number of factors to include in the model. We implement two complementary procedures by applying on the one hand the Hallin and Liška [2007] information criterion for determining the minimum number of common factors, and on the other hand a heuristic procedure still based on the variance explained by the eigenvalues.

The criterion by Hallin and Liška exploits the relation in the GDFM between the number of common factors and the number of diverging eigenvalues of the spectral density matrix of the observations. We choose the logarithmic form of the covariogram-smoothing version of the criterion, since Hallin and Liška maintain that this form has better finite sample performance than the non-logarithmic one. For given N and T, it consists in choosing the number of factors $\hat{q}_N^T$ so to minimize the following:

$$IC(q_N^T) = \log\left[\frac{1}{N}\sum_{i=q+1}^{N}\frac{1}{2M_T+1}\sum_{h=-M_F}^{M_F}\lambda_i(\theta_h)\right] + q_N^T c p(N,T), \qquad 0 \leq q_N^T \leq q_{\max} \leq N \quad (7)$$

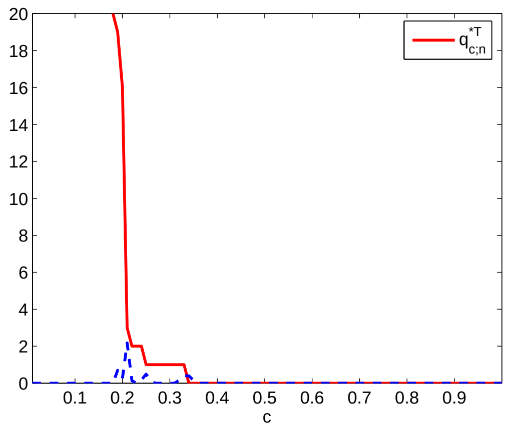
where θ_h, M_T and M_F are defined as in (4) and $p(N,T)$ is a penalty function satisfying

$$\lim_{N\to\infty} p(N,T) = 0 \quad \text{and} \quad \lim_{N\to\infty} Np(N,T) = \infty. \quad (8)$$

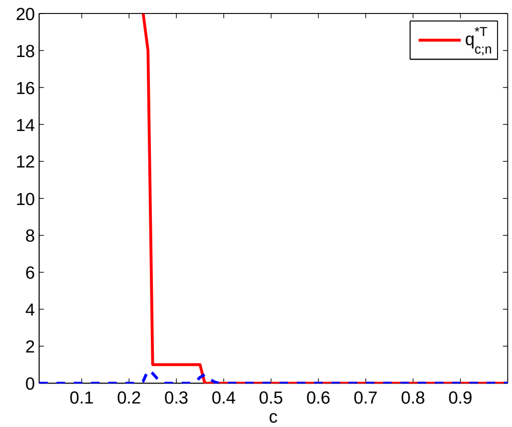
In principle, the maximum number of factors allowed $q_{\max}$ is the number of series in the dataset. Therefore, the penalty function should be large enough to avoid overestimation of

q, but at the same time it should not overpenalize. Multiplying the penalty function by a constant c is a way to tune the penalizing power of $p(N,T)$[5].
Hallin and Liška propose an automatic procedure for selecting $\hat{q}$ which basically explores the behavior of the variance of each q_N^T for the whole region of values of the constant c for N and T going to infinity. What we seek is the first stability region compatible with $\hat{q}_N^T < N$.
We build output plots (selected examples are in figure 3) by setting $q_{\max} = 20$, however results are robust to variations in $q_{\max}$. In these charts, the solid line indicates the value of q_N^T that minimizes $IC(q)$ for a given c and for the entire sample, while the dashed line measures (for the same c) the instability of q_N^T when considering different subsamples, i.e. different $n < N$. Roughly speaking, a dashed line approximating zero means that the value q suggested by the solid line is not biased by the sample dimension. In other words, the chosen $\hat{q}$ corresponds to the second zero-level interval of the dashed line corresponding to a plateau of the solid line, the first one being always associated to $q_{\max}$ and thus not indicative. We apply the the Hallin-Liška criterion on all the samples on different frequency bands, i.e. $[-\pi, +\pi]$, $[-\pi/2, +\pi/2]$, $[-\pi/4, +\pi/4]$, and $[-\pi/6, +\pi/6]$, in order to tell whether the estimated number of dynamic factors varies when taking into account different horizons: results are summarized in table 2. Indeed, as a general result sectoral common shocks are detected in the short run while in the long run the criterion indicates the existence of either less sector-specific factors or none at all. This is particularly evident when considering sales data in the chemicals sectors, whose plot is reported in figure 3 as an example. At longer horizons (i.e in the $[-\pi/6, +\pi/6]$ and $[-\pi/4, +\pi/4]$ frequency bands) the big economy-wide sales sample appears to be driven by one common dynamic factor, while the small sales sample and the sample for investments do not show any common factor. At higher frequencies, the Hallin-Liška criterion indicates the existence of one common factor for all the samples except the utilities industry, which is driven by two common dynamic factors.
These results point in the direction of the existence of (at most) one economy-wide shock at business cycle frequencies, which can be related to specific sectoral dynamics excluding the chemical industry, this latter possibly coming into play only at horizons shorter than 2 years. However, it could still be the case that the sectoral-specific factors and the economy-wide factor are ultimately the same: this point will be investigated in the next section. Finally, we compute the average variance of each series' common part over the total variance, which is one since data are standardized, giving also its standard deviation and its maximum and minimum values: this latter statistic in particular answers the question of why on average the variance of $\hat{\chi}_{it}$ is lower than the variance explained by the corresponding dynamic common factors. The heterogeneity of the series influences this result, since the variance of $\hat{\chi}_{it}$ is an average over the sample while the eigenvalues give a global and more reliable information. Moreover, we might interpret the variance explained by the first q dynamic common factors as a potential value reachable at the cost of including more and more lags to each dynamic common factor, trading off model parsimony and explanatory power.

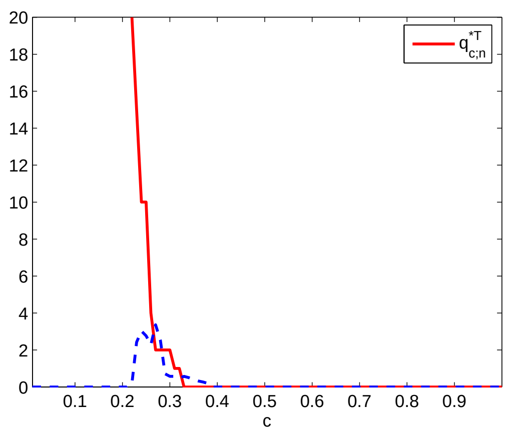
[5]We use $p(N,T) = \min(N, M_T^2, M_T^{-1/2} T^{1/2}) \log(\min[N, M_T^2, M_T^{-1/2} T^{1/2}])$.



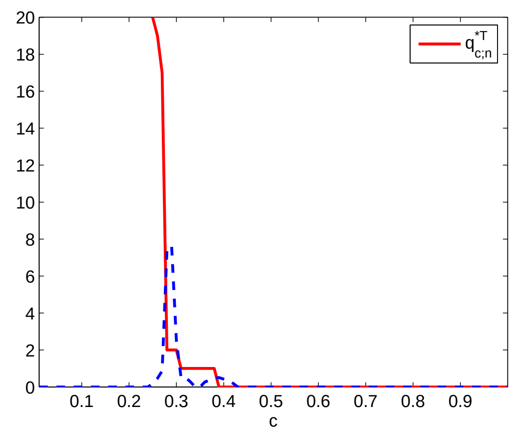
(a) Whole economy - long run



(b) Whole economy - short run



(c) Chemicals - long run



(d) Chemicals - short run

Figure 3: Selected Hallin-Liška criterion plots for sales data.

Sample	No. of series	Estimated q				Variance explained by the first q factors				Average variance of $\hat{\chi}_{it}$ over total							
										in $[-\pi,\pi]$				in $[-\pi/2,\pi/2]$			
		in $[-\pi,\pi]$	in $[-\frac{\pi}{2},\frac{\pi}{2}]$	in $[-\frac{\pi}{4},\frac{\pi}{4}]$	in $[-\frac{\pi}{6},\frac{\pi}{6}]$	in $[-\pi,\pi]$	in $[-\frac{\pi}{2},\frac{\pi}{2}]$	in $[-\frac{\pi}{4},\frac{\pi}{4}]$	in $[-\frac{\pi}{6},\frac{\pi}{6}]$	av	std	max	min	av	std	max	min
Sales	57	1	1	0	0	17%	17%	na	na	24%	17%	77%	3%	8%	10%	50%	0.5%
	660	1	1	1	1	14%	15%	13%	14%	21%	17%	98%	1%	7%	7%	51%	0.2%
Sales Chemicals	81	1	1	0	0	16%	15%	na	na	23%	17%	83%	3%	6%	4%	22%	0.4%
Sales Utilities	109	2	2	1	1	40%	43%	28%	27%	50%	21%	95%	6%	11%	7%	36%	0.4%
Sales Electronics	89	1	1	1	1	16%	19%	24%	24%	22%	16%	79%	1%	13%	16%	77%	0.3%
Sales Machinery	61	1	1	1	1	16%	17%	19%	20%	22%	14%	67%	5%	9%	8%	48%	0.5%
Investments	355	1	1	0	0	15%	13%	na	na	18%	13%	68%	1%	6%	4%	25%	0.1%
Investments Chemicals	37	1	1	0	0	23%	21%	na	na	26%	19%	71%	2%	8%	5%	20%	1%

Table 2: Factor decomposition according to the Hallin-Liška criterion.

Actually, the number of common dynamic factors suggested by the Hallin-Liška criterion is precisely the minimum number of factors required to satisfy the hypotheses of the model. As explained in Forni et al. [2000], including more dynamic common factors has asymptotically no consequences. Indeed, the criterion detects the number of dynamic factors q which satisfies the GDFM assumption requiring the variance explained by these q factors to diverge for N diverging. Adding more factors whose explained variance is bounded for N going to infinity yields then asymptotically the same result, the overall explained variance being diverging as well. Therefore, besides using the Hallin-Liška information criterion for determining the number of dynamic common factors, we are allowed to implement a complementary heuristic procedure which takes into account the amount of explained variance relative to the first q eigenvalues. Indeed, economic considerations might call for more than one common dynamic factor: different shocks could relate to the demand side, to the supply side (e.g. productivity or labor supply shocks), to monetary policy, to stock prices, to external developments (e.g. oil prices, exchange rate movements, fluctuations in world trade), etc.
In line with the literature, we aim at explaining at least 30% of the variance[6], hence we check how many dynamic common factors are necessary in order to fulfill the objective. As shown in table 3, from two to four dynamic factors are needed.

5 Retrieving the dynamic common factors

One might ask whether the first (or unique) dynamic common factor detected by the Hallin-Liška criterion is actually the same across all sectors and the economy as a whole, or we are ultimately dealing with sectoral factors, each one different from the economy-wide dynamic common factor. To answer this question, we first estimate the dynamic common factors and then compare them across samples.
We can easily recover the dynamic factors u_t once we obtain the dynamic eigenvectors of the spectral density matrix. Indeed the space spanned by the dynamic principal components coincides with the space spanned by the dynamic factors. The q largest eigenvectors, $p_1(\theta)\dots p_q(\theta)$, generate the two-sided filter:

$$\begin{aligned} \tilde{p}_j(L) &= \sum_{k=-M_T}^{M_T} c_{jk} L^k \qquad \text{for } j = 1, \dots, q \\ c_{jk} &= \frac{1}{2(2M_F+1)} \sum_{h=-M_F}^{M_F} p_j(\theta_h) e^{ik\theta_h} \qquad \theta \in [-\pi, +\pi] \end{aligned}$$

where θ_h is defined as in (4). Note that, as explained in section 2, in principle $\tilde{p}_j(L)$ is an infinite order filter, but in practice we truncate it at lag M_T. Moreover, given that we have only an estimation of the spectral density matrix, the Fourier coefficients c_{jk} are computed only for a finite number of frequencies. The dynamic principal components are then obtained just by applying these filters to the original series and we identify them with the dynamic factors:

$$u_{jt} = \tilde{p}_j(L) x_t \qquad \text{for } j = 1, \dots, q$$

In figure 4 we plot the first dynamic factor relative to the overall economy (solid line) against the first sectoral dynamic factor estimated by applying the GDFM to one sector at a time

[6] We do not consider a higher threshold because our data are highly disaggregated. See Forni and Reichlin [1998].

Sample	No. of series	q	Variance explained by the first q factors		Average variance of $\hat{\chi}_{it}$ over total							
			in $[-\pi, \pi]$	in $[-\pi/2, \pi/2]$	in $[-\pi, \pi]$				in $[-\pi/2, \pi/2]$			
					av	std	max	min	av	std	max	min
Sales	57	3	37%	37%	19%	12%	65%	2%	51%	16%	95%	21%
	660	3	31%	31%	15%	10%	82%	1%	44%	20%	99%	7%
Sales Chemicals	81	3	34%	35%	16%	10%	43%	3%	49%	18%	99%	15%
Sales Utilities	109	2	43%	40%	11%	7%	36%	0.4%	50%	21%	95%	6%
Sales Electronics	89	3	39%	35%	24%	19%	90%	2%	48%	18%	93%	8%
Sales Machinery	61	3	38%	37%	20%	14%	79%	4%	51%	17%	90%	21%
Investments	355	4	34%	35%	18%	9%	72%	4%	45%	18%	94%	10%
Investments Chemicals	37	2	31%	34%	13%	5%	20%	2%	39%	15%	69%	15%

Table 3: Factor decomposition according to the heuristic procedure.

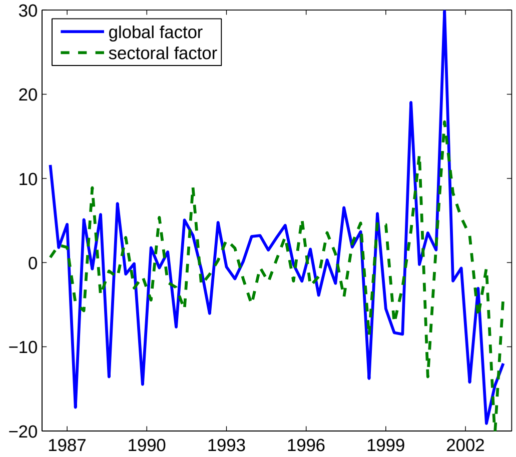
	$[-\pi, \pi]$			$[-\pi/6, \pi/6]$		
Sample	explanatory variable	t-value	R^2	explanatory variable	t-value	R^2
Sales Utilities	v_{t+6}	-6.7557	0.33288	v_{t+2}	10.379	0.61935
Sales Electronics	v_t	7.0106	0.39674	v_t	35.849	0.99286
Sales Chemicals	v_{t-3}	10.921	0.66121	no comm. factors	no comm. factors	no comm. factors
Sales Machinery	v_{t-1}	6.5346	0.48239	v_t	12.511	0.78364
Investments Chemicals	v_t	7.4695 7.6643	0.71711 0.72884	no comm. factors	no comm. factors	no comm. factors

Table 4: Regression results. Dependent variable u_t: economy-wide factor, explanatory variable v_{t+k}: sectoral factor.

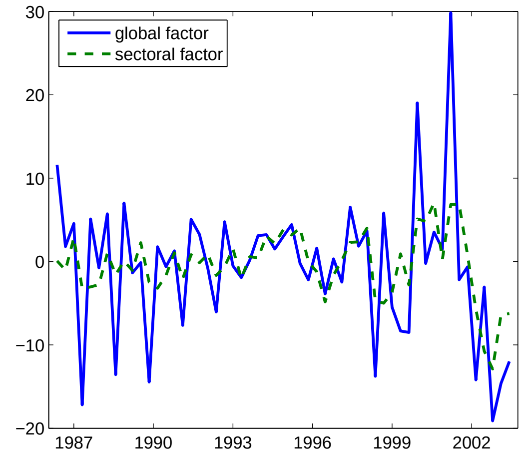
(dashed line), computed between $-\pi$ and π. In general, sectoral common factors look smoother than the economy-wide common factor, especially in the case of investments. Note that the peaks around 1996-1997 and 2001 in the macroeconomic common factor for sales are present also in the factors relative to the electronics industry and the utilities sector, respectively. However, since the dynamic factors are identified up to a dynamic rotation (or Blaschke matrix)[7], the relative timing of these peaks cannot be assessed precisely.

To investigate the stochastic dimension of the phenomenon, we run pairwise regressions of the economy-wide factor against leads and lags of sectoral factors, computed in both $[-\pi/6, +\pi/6]$ and $[-\pi, +\pi]$. In each regression we take as sole regressor the coincident value or one lead or lag of the first sectoral factor, up to the 8^{th}, so that we take into account correlations up to the correlation of the 4^{th} lead of one factor with the 4^{th} lag of the other. Table 4 summarizes the results for the regressions yielding the highest R^2 among the 17 regressions relative to the same sector. Note that the R^2 is generally fairly high, while the coefficients, although highly significant, are not reported since they have no informative power, the dynamic common factors being identified only up to a dynamic rotation.
This high correlation suggests the idea of the dynamic factors being actually just the same one for all sectors. Indeed, it seems there is one macroeconomic common factor which reflects into sectors, thereby inducing sectoral factors which closely resemble the economy-wide factor itself. The issue of the nature of the factor, however, deserves further investigation, although the 0.99 R^2 of the regression of the economy-wide factor on the electronics factor in the long run is a particularly striking insight about the nature of the economy-wide factor itself.

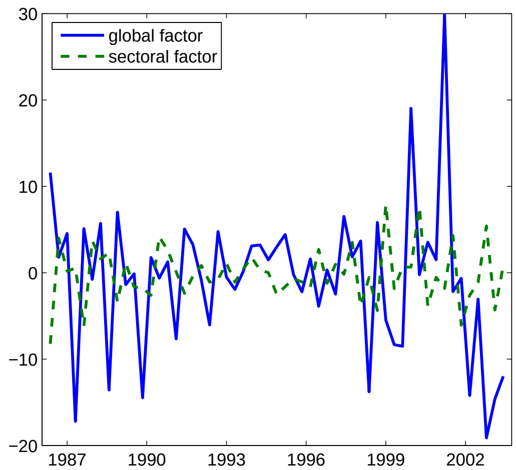
[7] See Forni et al. [2007].



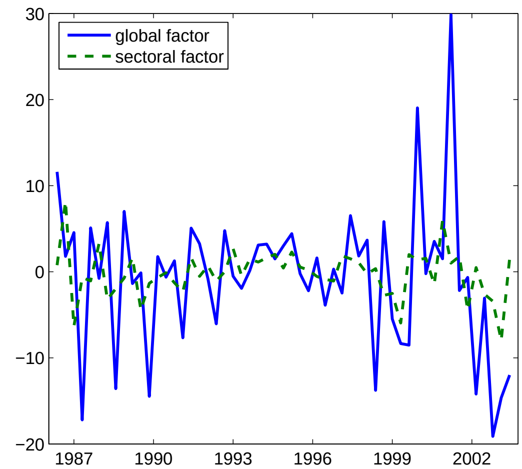
(a) Utilities - sales



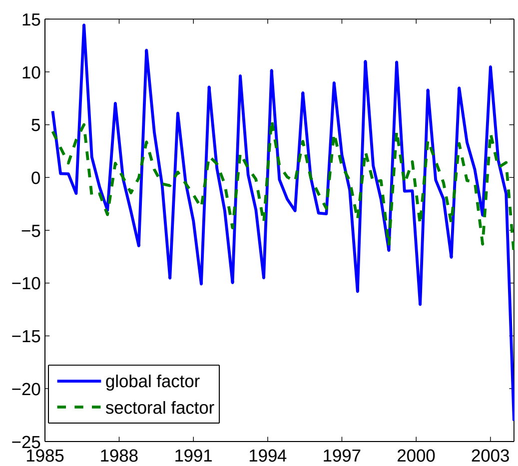
(b) Electronics - sales



(c) Chemicals - sales



(d) Machinery - sales



(e) Chemicals - investments

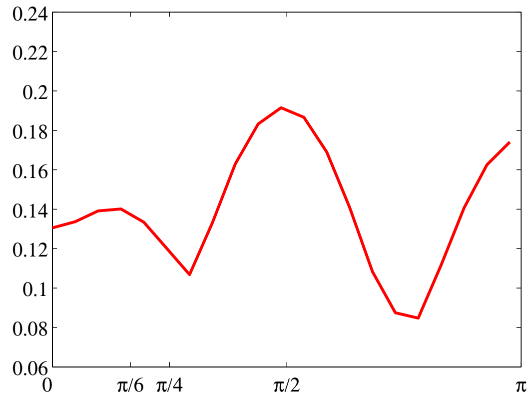
Figure 4: Economy-wide dynamic common factor (solid line) and sectoral common factor (dashed line).

6 Common and Idiosyncratic Components

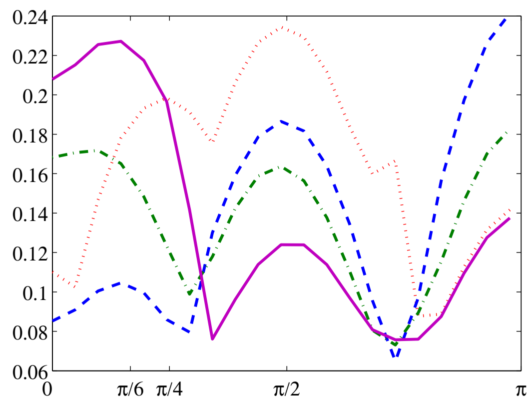
Spectral densities

Typically, the common component and the idiosyncratic component of the series present different spectral densities: low frequencies are more important in the common part, higher frequencies in the idiosyncratic part. Long run movements of the economy as a whole reflect into that part of firm growth which by definition is linked to the deep driving forces of the economy, while short run phenomena in the life of the firm are likely to be idiosyncratic to the firm itself. Finally, factors affecting firm growth may be classified at a third intermediate level, the sectoral level, when they influence more than one firm but their effects do not overcome the borders of a specific industry.
To check whether our data obey this rule, in figure 5 we have plotted the average spectral density of the common component obtained by including one dynamic common factor for frequencies belonging to $[-\pi, +\pi]$ (charts obtained by including more common factors and/or keeping narrower frequency bands do not provide additional information). We have grouped the common components estimated on all the firms of the economy according to the sector they belong to. Note that the obtained common components relate to only macroeconomic common factors since the influence of sectoral shocks vanishes via aggregation of sectors. Despite aggregation, however, different sectors still look different: indeed, two distinct patterns are detectable. In the electronics and machinery sectors the spectral density of the common component takes high values at business cycle frequencies and decreases more and more going towards one year; the chemicals and utilities spectral densities, on the opposite, take higher values the shorter the horizon. In other words, the bulk of sectoral comovements lies either at high or at low frequencies depending on the sector, the electronics and machinery sectors being more long-run driven and the chemicals and utilities sectors more short-run driven. As for the average spectral density of investments common parts, very high frequencies look more important than business cycle and long run frequencies. The spectral density for the chemical sector closely resembles the spectral density for the whole sample, meaning that comovements in investments are important in the short run also at a sectoral level.
In a nutshell, on sales data the results of the empirical analysis are consistent with the idea that macroeconomic common factors, leaving aside their specific nature, relative importance and interpretation, do affect firm growth at business cycle frequencies, but the relative importance of low frequencies versus high frequencies varies across sectors. Unfortunately we do not have a true sectoral disaggregation for investment data, where the bulk of comovements at least at the aggregate level and for the chemical sector is at short horizons, possibly shorter than one year.
The time domain counterpart of the average common component for sales data, computed with one dynamic common factor, is plotted in figure 6 (dashed line), together with the series of average sales (solid line): the average common component tracks the average sales very closely, this indicating that comovements play indeed a major role in sales dynamics.

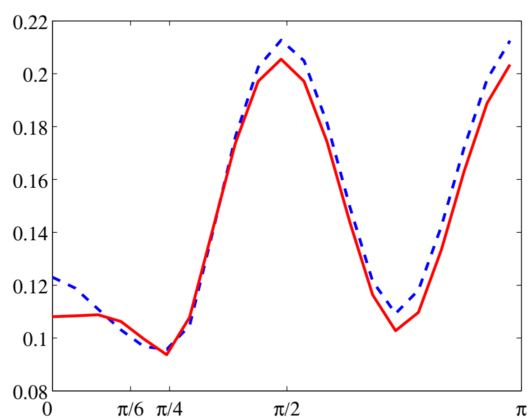
Finally, we estimated the common component of the sales and investment series only on the firms belonging to some sector. As expected, the common component in this case weights on average more than the common component estimated on all the firms of the economy since it captures not only what is common to the whole economy but also what is common only to the firms belonging to the same industry. Moreover, the common component estimated on the sectors is better estimated because homogeneity inside sectors is of course higher than in



(a) Sales.



(b) Sectoral sales. Solid line: electronics. Dashed line: chemicals. Dotted line: utilities. Dashed-dotted line: machinery.



(c) Investments. Solid line: all sectors. Dashed line: chemicals.

Figure 5: Average spectral densities of the common component by sector.

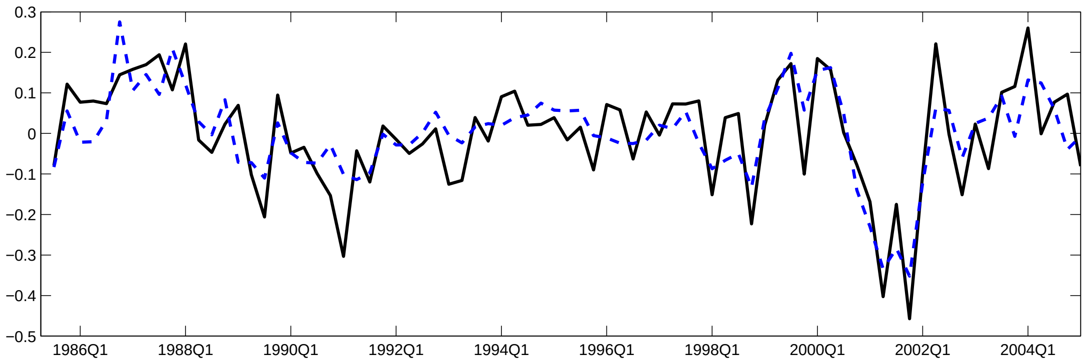


Figure 6: Solid line: average sales. Dashed line: χ_t with 1 factor.

the economy as a whole. However, the correspondent spectral densities have the same shape than those plotted in figure 5. This is consistent with the idea that nothing related to the sectoral level has ultimately a strong effect on the overall economy, at least at business cycle frequencies, in which case we should see some discrepancies between the spectral densities computed with the two approaches.

Cospectra

Besides the analysis of the spectral density of the common component, it is interesting to check how common components comove with each other. Focusing on comovements involving only the common component of the series adds value to the mere analysis of comovements among the series of output and investments as such. Indeed, by taking into account only the part of the series which is linked to the deep structure of the economy, we do not run the risk of interpreting as structural those comovements which are not.

The cospectrum $s_{ij}(\theta)$, relative to the common parts of series i and series j, is defined as the real part of the element of the spectral density matrix corresponding to series i and series j. We follow the approach proposed by Forni and Reichlin [1998] for measuring positive and negative comovements. Firstly, the cospectrum is decomposed into the sum of a positive cospectrum and a negative cospectrum, defined respectively as:

$$s_{ij}(\theta)_+ = [s_{ij}(\theta) + |s_{ij}(\theta)|]/2 \tag{9}$$

and

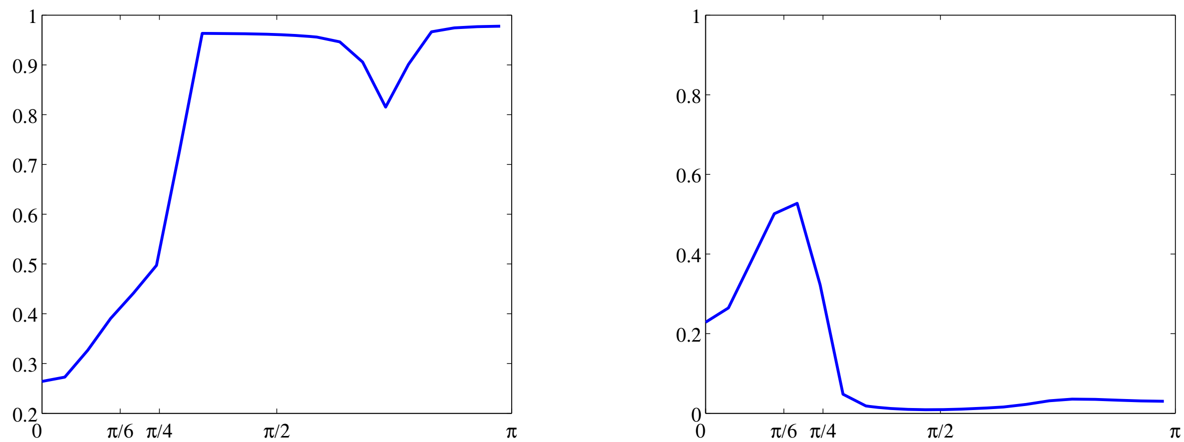
$$s_{ij}(\theta)_- = [s_{ij}(\theta) - |s_{ij}(\theta)|]/2 \tag{10}$$

with $\theta \in [-\pi, +\pi]$. Secondly, a synthetic measure is computed:

$$S(\theta) = -\frac{\sum_{ij} s_{ij}(\theta)_-}{\sum_{ij} s_{ij}(\theta)_+}. \tag{11}$$

A value of $S(\theta)$ close to 1 at a given frequency indicates strong negative comovements, while a value close to 0 indicates that positive comovements are important.

As shown in figure 7, where the quantity $S(\theta)$ has been plotted for the biggest samples of sales and investments, the pattern of comovements is pretty different between output and investment series. Firms' sales definitely comove more positively in the long run, i.e. for periods longer than three years. In the medium and short run, however, they comove negatively, and the shorter the period the more negatively they comove. This is reasonable: while in the long run every firm will be influenced by the economic conjuncture, thereby homogenizing its performance to the general boom or slowdown of the economy, in the medium run and especially in the short run the mechanism of competition and the process of selection play the major role. In the short run, the struggle for market shares by definition causes negative comovements on sales. The plot of $S(\theta)$ for investment series, instead, looks pretty much the opposite. Firms' investment behavior is extremely homogeneous in the short run, while it becomes more and more heterogeneous the longer the period. The peak of negative comovements in investment series corresponds to a 3 year period. A tentative interpretation of this finding relies in the monetary policy transmission mechanism, which is effective in bringing all the way down to firms each movement in interest rates, which ultimately affects investments. In the short run, every firm will be affected in the same way by monetary policy. On the other hand, investment decisions in the long run are inherent to the firm itself, depending on a number of factors as the attitude of the owner, in the case of a small firm, or the strategic decisions of the management in the case of a big company. This is consistent with the results of the analysis of the spectral density of the investments common component, which takes higher values in the short run.

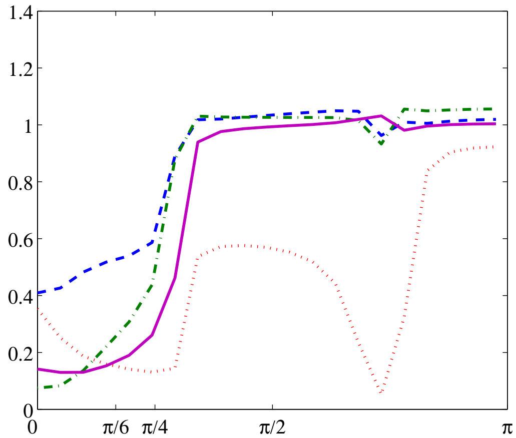


(a) All sectors, sales(660 firms)

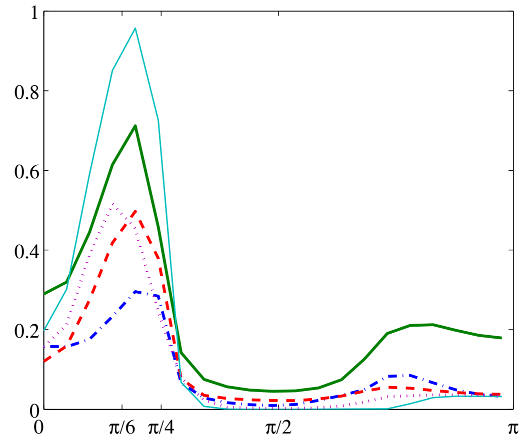
(b) All sectors, investments

Figure 7: Behavior of $S(\theta)$ for the whole economy.

Plots relative to sectoral subsamples are shown in figure 8 for both output and investments: they have been built by taking the correspondent off-diagonal elements of the spectral density matrix, which in turn has been computed including one dynamic common factor (plots for the three-factor case look alike). The plots for sales include the usual four sectors: utilities (solid line), chemicals (dashed line), electronics (dotted line) and machinery (dashed-dotted line). The plots for investments include the following 3-digit sectors, which are the most numerous in the investments sample: food manufacturing (NAICS 311), paper manufacturing (NAICS 322), chemicals (NAICS 325), machinery (NAICS 333), and electronics (NAICS 334). The pattern of comovements across firms belonging to the same industry does not diverge substantially from that of comovements across firms from all sectors, this supporting the idea of a unique deep force in the economy as a whole acting into sectors in a qualitatively similar manner. However note the presence of moderate negative comovements in investments also in the short run when taking into account only firms belonging to the same sector.



(a) Sales. Solid line: electronics. Dashed line: chemicals. Dotted line: utilities. Dashed-dotted line: machinery.



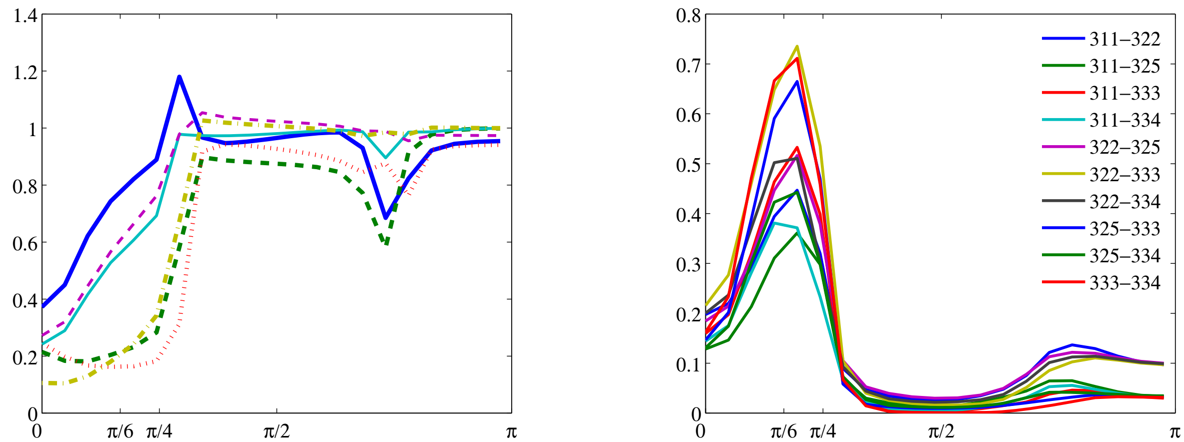
(b) Investments. Thin solid line: machinery. Thick solid line: paper manufacturing. Dashed line: chemicals. Dotted line: electronics. Dashed-dotted line: food manufacturing.

Figure 8: Behavior of $S(\theta)$ by sector.

Finally, we investigate reallocation effects and capital flows across different sectors. We have computed the quantity $S(\theta)$ by taking into account two sectors at a time, thereby considering the correspondent entries of the spectral density matrix. The first plot of figure 9 summarizes reallocation effects, which we measure by means of negative comovements of the common components of output series across sectors, while the seco shows capital flows, approximated by negative comovements of the common part of investment series across sectors. Again, we plot only the results obtained with one dynamic common factor. Reallocation effects are systematically more important in the short run, i.e. from one to two years, than in the long run. The relative importance of reallocation effects, however, varies depending on the two sectors we are considering, and on the frequency. Interestingly, job flows from and to the chemical industry are the more important at all frequencies, although lines tend to overlap each other in the very short run. However, this is likely to be a spurious result simply due to the fact that the chemical sector is the less prompt in reacting to the cycle, given the timing of plant building and R&D in chemicals and the magnitude of related investments. Finally, flows between the utility sector and the machinery industry are in general the most rare, especially in the short run.
The picture on capital flows is not as clear as on job flows. The pattern of comovement of inter-sectoral investments overall resembles the pattern of comovement of inter-firm investments, i.e. there are no preferential ways for inter-sectoral capital flows. Indeed, lines in the correspondent plot in figure 9 represent comovements across the five sectors considered for the intra-sector comovements and do not allow to detect any prominent direction.

7 Conclusions and further research

This work is an application of the Generalized Dynamic Factor Model to the study of industrial dynamics and the business cycle. In particular, we have analyzed output growth from a



(a) Reallocation effects. Thick solid line: utilities-chemicals. Thin solid line: chemicals-electronics. Thick dashed line: utilities-electronics. Thin dashed line: chemicals-machinery. Dotted line: utilities-machinery. Dashed-dotted line: electronics-machinery.

(b) Capital flows.

Figure 9: Behavior of $S(\theta)$ for sales (reallocation effects) and for investments (capital flows).

dynamic factor perspective, thereby aiming at disentangling that part of the phenomenon which is due to economy-wide shocks from that component which is idiosyncratic to each firm. We have used sales and investments quarterly series belonging to the COMPUSTAT database and covering the last 20 years. Thanks to the characteristics of the GDFM, we have been able to use extremely wide cross-sections, up to 660 series, and decompose them in sectoral subsamples, still about ten times wider than those used for traditional VAR analysis. As for the choice of the number of dynamic common factors to include in the model, consistently with the preliminary graphical analysis of the eigenvalues, the Hallin-Liška criterion indicates for all samples but the utilities sector just one dynamic common factor, explaining from 14% to 23% of the variance (considering all frequencies) depending on the sample. However, it is necessary to include up to three and four dynamic common factors if we aim at explaining at least 30% of the variance, for sales data and investments data respectively. A comparative analysis on the sectoral factors and the economy-wide factor suggests that we might ultimately deal with a unique driving force which could have no sectoral origin or just being linked only to some of the sectors of the economy, e.g. the electronics sector.
The results of the spectral analysis of the common component at the economy level are consistent with the idea that macroeconomic dynamic common factors do drive firm growth at business cycle frequencies. At the sectoral level, however, the bulk of comovements across firms may be at higher frequencies, as it is for the utilities and chemicals sectors.
Finally, we analyzed comovements across common components, detecting two opposite patterns in sales and investments. While sales comove strongly negatively in the short run and more positively in the long run, investments comove more negatively in the long run and strongly positively in the short run.

The main direction for further research is the identification and interpretation of the dynamic common factors, going beyond the distinction between common and idiosyncratic components and studying the building blocks of the common component itself. Note however that the model representation that we used in this work is not fundamental. As explained in Lippi and Reichlin [1994], nonfundamentalness poses a serious problem concerning the identification of economically sensible impulse-response functions associated to each shock u_t. Thus, before any identification strategy is proposed the problem of nonfundamentalness has to be solved, for example by adopting the alternative one-sided representation of the GDFM. Moreover, a challenging issue consists in extending the model in order to allow for structural breaks with respect to factor loadings or variance of the shocks.

In order to identify the dynamic shocks, further utilization of microdata is advisable, if not necessary, in order to take into account more different dimensions of the firm itself in addition to sales and investments, as R&D expenditure, inventories, and so on. At the same time, by exploiting the factor approach for the study of the firm, we would be able to disentangle the component of growth linked to external factors or related to the firm itself as a complex entity from the component concerning some specific inner function or division.

Finally, a natural extension of the present work would be an application of the GDFM on Euro-Area company data, e.g. the AMADEUS database. In this case one could investigate cross-country differences in shocks and shock transmission, assessing the issue of how much heterogeneity is due to asymmetric shocks and how important are country-specific shocks in the euro area. Indeed, investigating the strength of comovements within the monetary union, including the question of whether business cycle comovements have possibly increased over time, would help to understand to what extent “one policy fits all”.

References

G. W. Beck, K. Hubrich, and M. Marcellino. Regional inflation dynamics within and across euro area countries and a comparison with the US. Working Paper Series 681, European Central Bank, Oct. 2006.

J. Boivin and S. Ng. Understanding and comparing factor-based forecasts. NBER Working Papers 11285, National Bureau of Economic Research, Inc, May 2005.

D. Brillinger. *Time Series: Data Analysis and Theory.* Holden-Day, Inc., San Francisco, CA, 1981.

F. Canova, M. Ciccarelli, and E. Ortega. Similarities and convergence in G-7 cycles. *Journal of Monetary Economics*, 2007. Forthcoming.

G. Chamberlain. Funds, factors, and diversification in arbitrage pricing models. *Econometrica*, 51: 1281–1304, 1983.

G. Chamberlain and M. Rothschild. Arbitrage, factor structure and mean-variance analysis in large asset markets. *Econometrica*, 51:1305–1324, 1983.

M. Ciccarelli and B. Mojon. Global inflation. Working Paper Series 537, European Central Bank, Oct. 2005.

A. D'Agostino and D. Giannone. Comparing alternative predictors based on large-panel factor models. Working Paper Series 680, European Central Bank, Oct. 2006.

G. Dosi. Sources, procedures, and microeconomic effects of innovation. *Journal of Economic Literature*, 26(3):1120–71, September 1988.

G. Dosi. Statistical regularities in the evolution of industries. a guide through some evidence and challenges for the theory. LEM Papers Series 2005/17, Laboratory of Economics and Management (LEM), Sant'Anna School of Advanced Studies, Pisa, Italy, July 2005.

S. Eickmeier and J. Breitung. Business cycle transmission from the euro area to CEECs. Computing in Economics and Finance 2006 229, Society for Computational Economics, July 2006.

M. Forni and L. Reichlin. Let's get real: A factor analytical approach to disaggregated business cycle dynamics. *Review of Economic Studies*, 65(3):453–73, July 1998.

M. Forni, M. Hallin, M. Lippi, and L. Reichlin. The generalized dynamic factor model: identification and estimation. *The Review of Economics and Statistics*, 82(4):540–554, November 2000.

M. Forni, M. Hallin, M. Lippi, and L. Reichlin. The generalized dynamic factor model: one-sided estimation and forecasting. *Journal of the American Statistical Association*, 100(471):830–840, September 2005.

M. Forni, D. Giannone, M. Lippi, and L. Reichlin. Opening the black box - structural factor models with large gross-sections. Working Paper Series 712, European Central Bank, Jan. 2007.

J. Geweke. The dynamic factor analysis of economic time series. In D.J. Aigner and A.S. Golberger, eds., Latent Variables in Socio-Economic Models, 1977.

D. Giannone, L. Reichlin, and L. Sala. Tracking Greenspan: Systematic and unsystematic monetary policy revisited. CEPR Discussion Papers 3550, C.E.P.R. Discussion Papers, Sept. 2002.

D. Giannone, L. Reichlin, and L. Sala. Monetary policy in real time. Working Papers 284, IGIER (Innocenzo Gasparini Institute for Economic Research), Bocconi University, 2004.

M. Hallin and R. Liška. The generalized dynamic factor model: determining the number of factors. *Journal of the American Statistical Association*, 2007. Forthcoming.

G. Kapetanios and M. Marcellino. A comparison of estimation methods for dynamic factor models of large dimensions. Working Papers 489, Queen Mary, University of London, Department of Economics, Apr. 2004.

M. A. Kose, C. Otrok, and C. H. Whiteman. International business cycles: World, region, and country-specific factors. *American Economic Review*, 93(4):1216–1239, September 2003.

M. Lippi and L. Reichlin. VAR analysis, non-fundamental representation, Blaschke matrices. *Journal of Econometrics*, 63:307–325, 1994.

N. G. Mankiw. Real business cycles: A new keynesian perspective. *Journal of Economic Perspectives*, 3(3):79–90, Summer 1989.

M. Marcellino, J. H. Stock, and M. W. Watson. Macroeconomic forecasting in the euro area: Country specific versus area-wide information. *European Economic Review*, 47(1):1–18, February 2003.

L. Reichlin. Factor models in large cross-sections of time series. CEPR Discussion Papers 3285, C.E.P.R. Discussion Papers, Mar. 2002.

L. Sala. Monetary transmission in the euro area: A factor model approach. Macroeconomics 0205005, EconWPA, May 2002.

T. Sargent and C. Sims. Business cycle modeling without pretending to have too much a-priori economic theory. In C. Sims et al., eds., New Methods in Business Cycle Research, 1977.

J. H. Stock and M. W. Watson. Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics*, 20:147–162, 2002.

Appendix A

What follows is drawn from the COMPUSTAT User's Guide.

Sales (*data2*)
This item includes:

1. Advertising companies' net sales and commissions earned
2. Airline companies' transportation related revenues including passenger, cargo, mail, charter, and other
3. Any revenue source expected to continue for the life of the company
4. Banks and savings and loans' interest income and fee revenue
5. Banks' total current operating revenue and net pretax profit or loss on securities sold or redeemed
6. Broadcasting companies' net of agency commissions
7. Commissions
8. Equity in earnings/losses even if negative for real estate investment trusts and investors
9. Finance companies' earned insurance premiums and interest income. Finance companies' sales are counted only after net losses on factored receivables purchased are deducted from Sales - Net
10. Franchise operations' franchise and license fees and sales
11. Hospitals' sales net of provision for contractual allowances (sometimes includes doubtful accounts)
12. Income derived from equipment rental considered part of operating revenue
13. Installment sales
14. Leasing companies' rental or leased income
15. Life insurance and property and casualty companies' net sales in total income
16. Management fees
17. Net sales of tobacco, oil, rubber, and liquor companies' after deducting excise taxes
18. Oil and extractive companies' mineral royalty income
19. Operative builders' interest income from mortgage banking subsidiaries
20. Other operating revenue
21. Reimbursements for out of pocket expenses reported on the Income Statement
22. Rental income, if included by the company in Sales
23. Research and development companies' equity income from research and development joint ventures (when reported as operating income) and government grant income
24. Retail companies' finance charge revenues
25. Retail companies' sales of leased departments when corresponding costs are not available but are included in expenses
26. Royalty income and/or management fees when considered as part of operating income (such as, oil companies, extractive industries, publishing companies)
27. Security brokers' other income
28. Shipping companies' operating differential subsidies and income on reserve fund securities shown separately
29. Utilities' net sales total current operating revenue

This item excludes:

1. Broadcasting companies' agency commissions

2. Casinos' promotional allowances
3. Cost of delivery expenses for paper mills
4. Discontinued operations
5. Equity in earnings of unconsolidated subsidiaries
6. Excise taxes excluded from Sales (Net)
7. Gain on sale of securities or fixed assets
8. Interest income
9. Nonoperating income
10. Other income
11. Provision for contractual allowances for hospitals
12. Rental income

Capital Expenditures (*data90*)
This item includes:

1. Any items included in property, plant and equipment on the Balance Sheet
2. Expenditures for capital leases
3. Increase in funds for construction
4. Increase in Leaseback Transactions
5. Logging roads and timber
6. Reclassification of inventory to property, plant, and equipment

This item excludes:

1. Capital expenditures of discontinued operations
2. Changes in property, plant, and equipment resulting from foreign currency fluctuations when listed separately
3. Decreases in funds for construction presented as a use of funds
4. Deposits on property, plant and equipment
5. Net assets of businesses acquired
6. Property, plant, and equipment of acquired companies
7. Property, plant and equipment for real estate investment trusts
8. Software costs (unless included in property, plant and equipment on the Balance Sheet)

This item is not available for banks.
Data reflects year-to-date figures for each quarter.

What follows is the list of data anomalies causing a firm to be dropped from the databases used for the analysis:

- Data reflects an acquisition (purchase and/or pooling)
- Data reflects an accounting change
- Reflects fresh-start accounting upon emerging from Chapter 11 bankruptcy
- Excludes discontinued operations
- Includes excise taxes
- Includes other income/excludes some operating revenue
- Includes sales of leased departments

- Some or all data is not available because of a fiscal year change
- Some or all data is not available because the company has been in operation less than one year or presents more than or less than 12 months of data in their statements
- Includes six months of a merger or acquisition
- Includes nine months of a merger or acquisition
- Includes 12 months of a merger or acquisition
- Excludes six months of discontinued operations
- Excludes nine months of discontinued operations
- Excludes 12 months of discontinued operations

Appendix B

To seasonally adjust our dataset we used a multiplicative method based on the ratio between the original series y_t and its centered moving average, defined as:

$$z_t = \frac{1}{4}\left(\frac{1}{2}y_{t+2} + y_{t+1} + y_t + y_{t-1} + \frac{1}{2}y_{t-2}\right)$$

First of all we take the ratio $\tau_t = y_t/z_t$ from which we compute the seasonal indices i_q. Each of them is computed as the average of τ_t using observations only for the quarter q. We then adjust the indices so that they multiply to one; this is done by computing the seasonal factors as the ratio of the seasonal index to the geometric mean of the indices:

$$s_q = \frac{i_q}{(i_1 i_2 i_3 i_4)^{1/4}}$$

The interpretation is that the original series is s_j percent higher in quarter j relative to the adjusted series, which is finally computed as the ratio between y_t and the seasonal factors.