

# Business cycle and Corporate Failure in France: is there a link?

Eric Bataille<sup>1</sup>, Catherine Bruneau<sup>1,2</sup>, Alexis Flageollet<sup>1,2</sup>, and Frédéric Michaud<sup>1,3</sup>

<sup>1</sup> Banque de France  
(e-mail: [eric.bataille@banque-france.fr](mailto:eric.bataille@banque-france.fr),  
[alexis.flageollet@banque-france.fr](mailto:alexis.flageollet@banque-france.fr))

<sup>2</sup> THEMA  
Université Paris X  
(e-mail: [cbruneau@u-paris10.fr](mailto:cbruneau@u-paris10.fr))

<sup>3</sup> Crédit Agricole du Morbihan  
(e-mail: [Frederic.michaud@ca-morbihan.fr](mailto:Frederic.michaud@ca-morbihan.fr))

**Abstract.** The aim of this paper is to extract cyclical factors, first from companies' data used to build the score functions estimated by the Bank de France and, second, from these functions themselves. The constraints are those of a database including a large number of variables and companies and a small number of time periods. The method chosen is the "principal components analysis" adapted by [Bai and Ng, 2000] and [Bai and Ng, 2004a] in the context of large N and limited T. We show that the factorial structure could be useful to immunize the score functions and the related decisions against the cyclical variations in the state of economy.

**Keywords:** Panel data, common factors, principal components analysis, scoring.

## 1 Introduction

Over the last 30 years, many research papers have focused on the early detection of corporate failures. The Banque de France has developed several scores for use by financial analysts in the branches and at head office. Indeed, for reasons of robustness and for the seek of simplicity, the Bank has chosen to implement the linear Fisher classification method to build its scores. The analysis is static: the classification is conducted with a cross-section estimation over one year and adjusted so as to be robust to changes over time, even if the score functions need to be regularly adapted. In certain cases, a complete reestimation is needed: this is the case, when the nature of corporate failure has significantly changed and the related structural change cannot be modelled *ex ante*. In the other cases, the score function remains valid but the discriminant threshold has to be adjusted. Thus, this one seems to be dependent on the position of the economy in the business cycle.

In that purpose, we choose to extract an endogeneous cyclical component directly from the corporate database, because it allows to answer directly the question about immunization of the scores against cyclical variations and

because the whole analysis remains at a microeconomic level, which avoids considering difficult questions about data agregation.

The business cycle can be represented as an unobservable component which can be identified with the principal components method presented in [Bai and Ng, 2000] and [Bai and Ng, 2004a].

After extracting cyclical component from corporate database and from the scores themselves, we look for interpreting these factors as cyclical co-movements, by comparing them to macroeconomic series which usually enter in the characterization of the French business cycle ([Bruneau *et al.*, 2002a]).

## 2 The dynamic factor analysis (DFA) in the lines of [Stock and Watson, 1998]

In this section, we recall the main contributions in Dynamic Factor Analysis starting with the work by [Stock and Watson, 1998] continuing with the papers by [Bai and Ng, 2000], [Bai and Ng, 2004a] and [Bai and Ng, 2004b].

### 2.1 The main assumptions

From now on,  $X_t$  will denote a  $N$ -dimensional multiple time series. The factor structure is as follows:

$$X_t = A_t F_t + u_t \quad (1)$$

where the dimensions are respectively :  $N \times 1$ ,  $N \times r$ ,  $r \times 1$  and  $N \times 1$ . The common part of  $X_t$  is  $A_t F_t$  and  $u_t$  denotes its idiosyncratic part. Note that, in the previous model, the dynamics is introduced in three ways:

- 1) the factors are assumed to evolve according to a time series (multivariate) process which is not observable;
- 2) the idiosyncratic error terms are serially correlated;
- 3) the factors can enter with lags (or even with leads).

Note also that the dynamic factor model can be rewritten such that  $A_t$  is constant by suitable redefinition of the factors and the idiosyncratic disturbances.

The factors as well as the loadings ( $A_t$ ) are considered as parameters that are estimated by solving a non-linear least squares problem which is decomposed into two successive ordinary least squares minimizations, which finally lead to solve an eigenvalue problem.

It is important to recall the assumptions:

- i)  $A_t = A_0$
- ii) the disturbances  $u_t$  are i.i.d. independent across series, normally distributed so that the covariance matrix  $\Sigma$  of the vector of residuals  $u = (u_1, \dots, u_T)$  is diagonal. (Its seems to be possible to allow a weak correlation structure between the  $u_{jt}$  for any date  $t$  ([Chamberlain and Rothschild, 1983]).

Thus the estimator of  $(A_0, F)$  solves the non-linear least squares problem with the objective function:

$$V_{NT}(A_0, F) = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T I_{it}(X_{it} - \lambda_{i0}F_t)^2 \tag{2}$$

where  $I_{it} = 1$  if the variable is observed at time  $t$  and equal to 0, otherwise.

The previous analysis is a standard principal component analysis with the only difference that dynamic features are taken into account.

Recently, [Bai and Ng, 2004a] have proposed a statistical procedure to extract factors without considering the degree of persistence in the series. It is the so-called PANIC approach (Panel Analysis of Non-stationary in Idiosyncratic and Common components).

### 2.2 PANIC analysis ([Bai and Ng, 2004a])

The model is the following:

$$\begin{aligned} X_{it} &= c_i + \beta_i t + \lambda'_i F_t + e_{it} \\ (1 - L)F_t &= C(L)u_t \\ (1 - \rho_i L)e_{it} &= D_i(L)\varepsilon_{it} \end{aligned}$$

with  $C(L) = \sum_{j=0}^{\infty} C_j L^j$  and  $D_i(L) = \sum_{j=0}^{\infty} D_{i,j} L^j$ . The idiosyncratic  $e_{it}$  is  $I(1)$  if  $\rho_i = 1$  and is stationary if  $|\rho_i| < 1$ .

When the residuals  $e_{it}$  are  $I(0)$  it is possible to get consistent estimates of the factors. When it is not the case,  $e_{it}$  are  $I(1)$ , one has to work with the first differences of the series. The model allows  $r_0$  stationary factors and  $r_1$  common trends with  $r = r_0 + r_1$ <sup>1</sup>. Equivalently, the rank of  $C(1)$  is equal to  $r_1$ .

Instead of testing for the presence of a unit root in  $X_{it}$ , the approach proposed here is to test the common factor and the idiosyncratic separately. PANIC has two objectives: first, to determine if non-stationarity comes from the common or from the idiosyncratic component. Second, to construct valid pooled tests for panel data when units are correlated; that is under the cross-sectional dependence (CSD) assumption.

More precisely, the objective of PANIC is to determine  $r_1$  and test if  $\rho_i = 1$  when neither  $F_t$  nor  $e_{it}$  is observed and is estimated by the method of principal components.

The large  $N$  permits consistent estimation of the factor and idiosyncratic components, whether or not they are  $I(1)$  or  $I(0)$ . A large  $T$  enables application of relevant central limit theorems so that limiting distributions of the tests can be obtained.

<sup>1</sup> The number of factors  $r$  is supposed to be given. Recently, [Bai and Ng, 2000] have proposed to use relevant information criteria to determine the number of factors in the S&P framework.

A important aspect of PANIC is that the idiosyncratic errors can be analysed (more specifically their stationarity) without knowing if the factors are stationary and vice-versa. More precisely, the tests on the factors are asymptotically (large  $N$  and  $T$ ) independent of the tests on the idiosyncratic terms.

When the idiosyncratic part is non stationary, [Bai and Ng, 2004a] recommend to deal with the firsts differenced series. The rank parameter  $r$  is identified by using an information criterium like the previous one, applied for the model in first differences.

Simulations show that the proposed tests have good finite sample properties even for panels with only 40 units.

It is worth noting that the factors are estimated more efficiently from the series in levels, if the idiosyncratic components are  $I(0)$ . The procedure we use can be find in [Bai and Ng, 2004a] and because computing individual  $p$ -values requires simulation, for that purpose we use the table computed by S. Ng for the DF distribution.

To give an economic interpretation of the factors extracted in the lines of [Bai and Ng, 2004a], we use the methodology presented in the paper by [Bai, 2003]. The point is to estimate a confidence interval around each of the (true) factors and check if an observed series lies or not in this interval.

### 3 Data and results

We first comment the main contents in the database before presenting the results.

#### 3.1 Corporate Database

The Banque de France built 8 scores to detect corporate failures. Estimates of score functions are based on data from company balance sheets in the Banque de France's Fiben banking database<sup>2</sup>. This database is used to construct a pool of ratios. Some of them enter the score functions. We work on the basis of ratio and on ones of the scores themselves. In all cases, we work on the averages of ratios or scores<sup>3</sup>.

The database of ratios covers 10 industries defined by the NES classification. It reports 91 ratios, that are usually employed in financial analysis and scoring decisions. They are estimated from the data characterizing firms over the 1989–2002 period. We formed their average on the 10 NES groups.

<sup>2</sup> For a complet description of the Banque de France scoring methodology, see [Bardos *et al.*, 2004].

<sup>3</sup> The restrictive choice of average measures can be *ex post* justified by the Fisher classification analysis employed at the Banque de France for the construction of the scores.

The statistical results can be summarized as follows, by focusing first on the stationarity properties of the series and, next, on the co-movements of these series.

### 3.2 The results

As some of the series exhibit a trend, we regress them on a linear function of time and replace them by the corresponding residuals when the trend is significant. In what follows, the residuals are designed as the detrended series<sup>4</sup>.

First, we focus on the idiosyncratic components, estimated from the first differences of the series. The test statistic  $P_{\hat{\epsilon}}^c$  takes the value 19.6, which leads to reject the stationarity of the idiosyncratic components  $e_i$ . However, as the time dimension is very low ( $T = 12$ ), we work with the level of the series to extract the factors. *A contrario*, the three extracted factors appear stationary.

According to the [Bai and Ng, 2000] criterion, 13 factors appear to be necessary to summarize the panel. Since 14 is the time dimension in the present analysis, this criterion does not appear to be relevant here. As the first three factors account for 74% of the variance (43.5% for the first one, 23.5% for the second and 7% for the third), we retain them to summarize the co-movements of the series at hand.

In addition, the contributions of the ten sectors appeared to be very homogenous<sup>5</sup>. It justifies the choice of implementing a global analysis and, more precisely, a business cycle analysis.

To make easier the interpretation of the contributions of the 91 ratios, we group them together in 10 financial-type ratios. The contribution of each synthetic ratio is just the sum of the contributions of the underlying ratios it summarizes. To take into account the number of contributing ratios to each synthetic ratio, we compare their effective contribution with the average contribution of all ratios. So, the first factor appears strongly associated with mark-up variables, the second one with solvability features and the third one with indebtedness characteristics.

Since the Principal Component Analysis is essentially static, the second and third factors may be lags or leads of the first factor. Correlation estimations accredit this hypothesis. The maximum correlations are achieved for a lag of 2 years for the second factor (0.89) and 4 years for the third one (0.84).

The use of the correlation procedure developed by [Bai, 2003] seems to confirm it. We should conclude that there is a unique factor, rather than

<sup>4</sup> The analysis was also conducted on the raw series to control the robustness of the method. It appeared that the first factor extracted in the later context looked like the trend of the series and the following factors like the factors extracted from the detrended series. Indeed, one finds high correlations between the factors of both analysis, around 0.8 (results available on request).

<sup>5</sup> The detailed results will be presented in the complet paper.

three, which really summarizes the co-movements of our series. But this result has to be considered with caution, because of the low time dimension of the series. Focusing on the contributions of each synthetic ratio to the factors supports this interpretation. Indeed, we can suppose that the cyclical co-movement, which we interpret as a business cycle effect (see the following paragraph), affects beforehand the profitability of the firms and, consequently, the balance sheet structure, which is weakened via the increase in debts or the degradation of stockholders' equity.

In what follows, we give a precise interpretation of the factors by comparing them to observable series that are usually considered as representative of the business cycle in France. Three macroeconomic series were chosen: the annual variation of French GDP in value, the output gap of the French GDP in volume obtained by the Hodrick-Prescott filter and the industrial production capacity utilization (TUC) as calculated by the Banque de France. We use confidence intervals in the lines of [Bai, 2003].

The results confirm the relationship between our three factors and the business cycle<sup>6</sup>. But at the last, we have to decide if using the factors to represent business cycle is more relevant, or efficient, than using particular macroeconomic series, as the output gap, for example. It would be also interesting to investigate the possibility of using variables to partially forecast the cyclical co-movements.

Indeed, such a forecasting power would allow setting up scenarii characterizing different states of economy.

The scoring method as it is implemented up to now at the Banque de France does not cover all activity sectors and we are not sure that this limitation does not influence the results of the factorial analysis. This question is taken into account in this study.

Moreover, as we aim at measuring the influence of the business cycle on the corporate ratios and especially on the failure risk, we have to take into account information on failure. However, the failure rate in our sample is very weak -around 2 and 5%. The weight of the failing firms is consequently very weak within the sample. Their influence on the detection of business cycle could be out of measure. To overcome this difficulty, we performe weighted averages for each sector/ratio.

To refine the diagnosis, we also use separately the samples of failing and non-failing firms. The PCA gives the results summarized in the following table. In spite of the rebalancing of sample, the contributions of the first three factors for the failing firms significantly differ from those obtained for the other firms.

In order to investigate if the restriction of the database or/and the distinction between failing and non-failing firms modified the results obtained over the whole sample, we projecte the first factors of every partial ACP on the space spanned by the first three factors stemming from the complete

---

<sup>6</sup> The detailed results and graphs will be presented in the complet paper.

sample. The results indicate a very strong similarity of the common factors<sup>7</sup>, indicating that the state of economy influences failing and non-failing firms in a similar way, whatever the sector considered.

After extracting common cyclical factors from corporate, we focus on the score functions themselves. Indeed, nothing says that the cyclical components are detectable from the scores, which have been precisely adapted to be immunized against cyclical variations.

The same plan of study as previously was thus applied to the averages of the score functions. To increase the cross-section dimension of the sample, the desagregation is made according to the NAF classification which is finer than the NES classification used before.

We had thus a sample of 8 averaged score functions over the same time period (1989 to 2003) for 49 sectors. So it includes 392 variables over 14 years. We then apply the principal components estimation on the 392x392 dimensional matrix. The first three factors account for 58.8% of the variance (33.6% for the first one, 17.5% for the second and 7.7% for the third). To compare these factors with the first three factors obtained before, we use again confidence intervals estimated in the lines of Bay [2003].

The results<sup>8</sup> show that the cyclical component is common to the ratios and the function scores. The projection of each height average score on the same space has come to the same conclusion.

The functions scores do not modify substantially the cyclical common component which is present in the original ratios.

So, we have to conclude that we should take into account the information about the cycle of activity to improve modeling of the corporate failure risk.

To finish, a question remains un-answered: how to implement scoring so as to account for cyclical variations? Have we just to adjust the decision thresholds or should we modify the score functions according to the state of economy? To give a first answer to these questions, we compute, for every year and for every score, the optimal decision threshold. Optimality equalizes both first and second type errors. Thus, we regress the different optimal thresholds onto the first three factors extracted from the scores database.

What do we observe? The decision thresholds are significantly correlated with the factors most of the time, as they generally belong to the confidence interval around each factor. But it is not always the case, so that we have to conclude that the analysis should be deepened in order to decide how to modify the scoring decision and more precisely the threshold, so as to include the cyclical variations observed in the score functions.

---

<sup>7</sup> The detailed results and graphs will be presented in the complet paper.

<sup>8</sup> The detailed results and graphs will be presented in the complet paper.

## 4 Conclusion

To summarize, we have to claim that the corporate ratios of the firms in our data base display significant cyclical comovements, which are quite similar when they are extracted from the whole corporate database and from the database just including the firms for which a score function has been estimated up to now in Banque de France. Moreover, the score functions themselves display the same cyclical comovements and, as a consequence, we have to conclude that these score functions are not immunized against cyclical variations in the state of economy.

Indeed, macroeconomic series, that are usually recognized as proxy variables to characterize the cyclical behavior of the French economy, appear to be significantly correlated with the estimated factors extracted from the database of the financial ratios as well as from the database of the scores.

Finally, we examine what kind of consequence this dependency of the score functions on the cyclical movements in economy may have on the scoring decision process itself.

We compared the decision thresholds to the estimated factors for each sector. We observed that the thresholds are correlated with the factors most of the time. Accordingly, the threshold should be vary over time like the underlying cyclical components in order to improve the scoring procedure. However, there are subperiods where the correlation disappears, indicating that changing the threshold is not always sufficient to account for changes in the economic environment. So, the analysis has to be deepened and this is left for further research.

## References

- [Bai and Ng, 2000]J. Bai and S. Ng. Determining the number of factors in approximate factor models. *Working Paper*, 2000.
- [Bai and Ng, 2004a]J. Bai and S. Ng. A PANIC attack on unit roots and cointegration. *Econometrica*, pages 1127–1177, 2004a.
- [Bai and Ng, 2004b]J. Bai and S. Ng. Evaluating latent and observed factors in macroeconomics and finance. *Working Paper*, 2004b.
- [Bai, 2003]J. Bai. Inferential theory for factor models of large dimensions. *Econometrica*, pages 135–172, 2003.
- [Bardos *et al.*, 2004]M. Bardos, Foulcher S., and E. Bataille. Banque de France scores: methode, results and applications. *Banque de France*, 2004.
- [Bruneau *et al.*, 2002a]C. Bruneau, O. de Bandt, A. Flageollet, and E. Michaux. Forecasting inflation using economic indicators: the case of france. *NER - Banque de France*, 2002a.
- [Chamberlain and Rothschild, 1983]G. Chamberlain and M. Rothschild. Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica*, pages 1281–1304, 1983.
- [Stock and Watson, 1998]J.H. Stock and M.W. Watson. Testing for common trends. *Journal of the American Statistical Association*, pages 1097–1107, 1998.



Part VII

**Data Analysis**

