"Commonality, macroeconomic factors and banking profitability"

Orlando Joaqui-Barandica, Diego F. Manotas-Duque and Jorge M. Uribe-Gil

Abstract

We study banks' profitability in the US economy by means of dynamic factor models. Our results emphasize the importance of a few common cyclical market factors that greatly determine banking profitability. We conduct exhaustive regressions in a big data set of macroeconomic variables aiming to gain interpretability of our statistical factors. This allows us to identify three main macroeconomic factors underlying banking profitability: the financial burden of households and economic activity; household income and net worth and, in the case of ROA and ROE, corporate indebtedness. We also provide an integrated perspective to analyse banks' profitability dynamically and to inform policymakers concerned with financial stability issues, for which banks' profitability is fundamental. Our models allow us to provide several rankings of vulnerable financial institutions considering the common market forces that we estimate. We emphasize the usefulness of such an exercise as a market-monitoring tool.

JEL classification: E44, G20, G21.

Keywords: Banks' ROA, Indebtedness, Dynamic factors, Financial cycles.

Orlando Joaqui-Barandica: Universidad del Valle, Faculty of Engineering, School of Industrial Engineering, Colombia. Email: orlando.joaqui@correounivalle.edu.co

Diego F. Manotas-Duque: Universidad del Valle, Faculty of Engineering, School of Industrial Engineering, Colombia. Email: diego.manotas@correounivalle.edu.co

Jorge M. Uribe-Gil: Faculty of Economics and Business, Open University of Catalonia and Riskcenter, University of Barcelona, Spain. Email: juribeg@uoc.edu

1. Introduction

Understanding the sources of profitability in the banking sector is of prime importance to both policymakers and professionals in the banking industry. On the one hand, the profitability of banks is under intense and constant scrutiny by regulators and central banks because it is directly related to the soundness of financial institutions and, therefore, to financial stability and the probability of systemic risk materialization. On the other hand, profitability is a targeted indicator by managers and CEOs who determine the optimal combinations of operating assets and capital sources, hoping to increase a bank's value and its return on investment. It is not surprising that there is a lively debate in the literature about whether profitability in banks is mainly due to idiosyncratic characteristics of financial institutions, such as size, scope, capitalization, asset quality, efficiency, and the business model, or, in contrast, is more a reflection of underlying common market forces, over which bank managers lack any kind of influence, such as short-term policy rates, long-term rates, general financial conditions, or more broadly speaking, cycles in economic activity (see section 2). In practice, naturally, profitability answers to both sides of the narrative, but the key point is the degree to which one can rely on each side to explain the banking system performance.

In the first part of our study, we directly address this question. We answer how much of the profitability of the largest banks in the US banking system can be explained by common underlying market forces, which are not related to idiosyncratic characteristics. We adopt an eclectic perspective based on dynamic factor models, often used in machine learning asset pricing but novel in the context of our research question. Our approach allows us to directly quantify how much of the profitability of the largest US banks can be explained by some common factors to which all financial institutions are directly exposed. Three statistical factors[4] are sufficient to explain between 63% and 68% of the returns on assets (ROA), returns on equity (ROE) and EBITDA margin variations for the US banking industry from January 2000 to March 2021. This is a significant fact, considering that in a hypothetical case in which banks' profitability is fully explained by bank-specific characteristics, these three factors can be expected to account for up to 1.2 - 2.5% of the total variation in our panel of profitability indicators, which concern more than a hundred banks operating in the US market (in some cases up to 241). This first result is in line with the fact already documented in the literature and well known by central banks and practitioners that banking is a very cyclical industry and

[4] The optimal number of factors was decided following the criterion proposed by Bai and Ng (2002).

therefore dominated by underlying market macroforces, which determine most of the dynamics of a bank's performance. We offer a way to quantity to what extent this occurs.

In the second part, we go a step further by considering the interpretation of our statistical factors, estimated by regularized principal components analysis (PCA). This is actually a tough question, as PCA and machine learning in general are often criticized on the grounds of interpretability. In summary, to achieve reliable results when modelling a given economic (or other) phenomenon, it is often not enough to capture the statistical dynamics within the system, but we also need to understand what reasons underlie the reported empirical results if we aim to convincingly generalize our conclusions beyond the study sample. This problem brings us closer to the banking literature that has examined the macroeconomic determinants that impact bank profitability, such as GDP growth or market-wide liquidity. Unlike these studies, we follow a novel big data approach to address the problem. Rather than establishing in advance the macroeconomic factors responsible for the profitability of banks, which inevitably increases the risk of missing important confounding covariates unknown to the literature or the researcher, we let the data speak as freely as possible. Our approach consists of the following two steps. First, we collect and preprocess 248 quarterly series of US economic activity recently assembled by McCracken & Ng (2021), which together constitute a complete picture of real markets, financial markets and prices in the US. Then, we relate each of our three statistical factors to each series in the set of big data macro variables and select the heaviest loading series among them, using marginal R-squared statistics from exhaustive and separated regressions and thereby matching our statistical factors with well-known economic series that can be easily interpreted.

In this regard, we are inspired by McCracken & Ng (2021), who use statistical factors extracted from the same big data set of macrolevel variables as ours and then identify the strongest loading series using marginal R-squares among the original variables. This avenue confers interpretability to their results and ours. However, although similar in spirit to our approach, our statistical factors are estimated ***outside*** the macro system, i.e., from our banking profitability data. Therefore, there is no *a priori* reason to expect a high correlation between the factors and the individual series. Bearing this in mind, we are still able to identify what series drive the dynamics of each banking profitability factor, which makes our results solid and insightful. The first factor is related to the financial burden of households and economic activity, the second factor is associated with household income and, hence, employment, and the third factor is related to commercial loans (this factor involves changes in EBITDA, in which case it is closer to the dynamics of housing markets). Interestingly, our banking factors do not perfectly (or even closely) match three additional factors estimated using the macroeconomic series related to economic activity, prices and financial conditions. The largest correlation (in magnitude) among the banking and macroeconomic factors is -0.44, appearing between the second banking factor (household income) and the third macroeconomic factor (financial conditions). Otherwise, the two sets of factors depict largely independent trajectories, which highlights the advantages of our approach against the alternative path of identifying the market factors directly from the set of macroeconomic series.

In the third part of our results, we turn to the bank-specific side of the narrative. Having identified the main systemic forces behind the profitability of all banks and established their dominance to explain the cross-sectional profitability of banks, we investigate each bank's exposure to each factor and explore the explanatory power of our three-factor model on an individual basis. In this way, we can provide a ranking of banks according to their sensitivity to the three underlying market forces. The explanatory power of our factor model lies between a minimum of 6% and a maximum of 93%, with an interquartile range of 43%-77%. This highlights the heterogeneity of banks' exposure to the three market forces and the general adequacy of the proposed methodology to monitor banking profitability. We emphasize that ours is not necessarily an exercise in systemic risk, since we do not focus on particularly bad situations (for example, the lowest quantiles of profitability), but rather we analyse the average scenarios using average factors. However, we do gain novel insights in terms of systemic risk in this part of our results. We find that while the profitability dynamics of certain banks may be unrelated to certain systemic forces (e.g., the first and second banking factors), they may be highly linked to others (e.g., the third banking factor). The results here highlight the convenience of monitoring banks' profitability by using an integrated approach along the lines that we propose instead of by resorting to various indicators provided in the literature, such as the widely known approaches by Acharya et al., (2012) or Adrian & Brunnermeier (2016). Such indicators, despite reflecting different sides of systemic risk, are not mutually exclusive; therefore, they are highly correlated with each other. Our profitability indicators are directly identified as orthogonal combinations (i.e., PCA), such that they offer the integrated and complementary perspective that we emphasize.

Our results are robust to changing the method to estimate the factors from the baseline regularized PCA to traditional PCA and other methodological choices, to dividing our sample according to the market capitalization of the banks within the study sample (although the results seem more binding for the largest banks) and to using the three main indicators of profitability, namely, ROA, ROE and EBITDA (although the last factor has a different influence on EBITDA), which are additionally sampled at different frequencies, quarterly in the first two cases and annually in the last case.

The rest of this document is organized as follows. In the second part, we review in more detail the related literature and put our contribution into perspective. In the third part, we describe our methods. In the fourth section, we present our data and main results and provide some robustness for our claims, while in the fifth and last section, we conclude and offer some future research avenues.

2. Bank profitability literature

Bank profitability analysis has been one of the most important financial economics research challenges over time. There are different streams of the literature that analyse bank profitability. It is possible to identify in the literature two lines of analysis related to the determinants of the performance of financial institutions.

First, we find internal determinants, such as the structure and size of assets and the financial and capital structure of financial institutions. In this same group, internal determinants of operational character are analysed, such as the productivity of the workforce and the number of ATMs and customer service offices.

Le & Ngo (2020) investigate the determinants of bank profitability in 23 countries from 2002 to 2016. The main findings of this study point out operational determinants, such as the number of bank cards issued, the number of automated teller machines and the number of points of sale terminals, all of which can increase bank profitability. Additionally, these authors suggest that market power has a negative impact on bank profitability, and a more concentrated banking system is associated with lower profitability because nonprice competition may be more intense in more concentrated markets. Furthermore, managers can more easily engage in expense-preference behaviour so bank costs in such markets are higher, thus lowering profitability; in contrast, competition increases it. Other authors, such as Kumar et al., (2021), focus their analysis on bank profitability on topics related to financial inclusion. The main conclusions of this study suggest that banks with a wider scope of financial services are more profitable than their counterparts. Other key issues studied by these authors define cost management, credit risk management and bank size as key drivers of bank profitability.

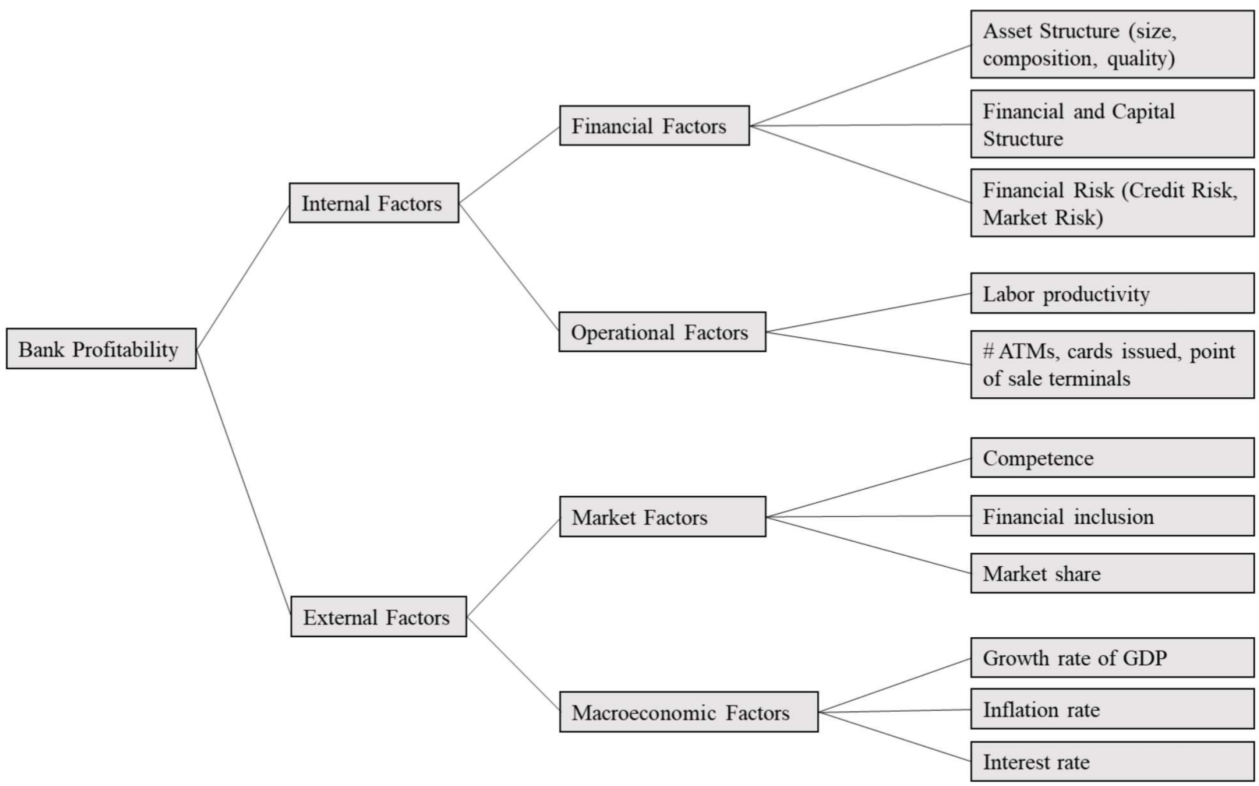

Figure 1. Lines of analysis of bank profitability. **Note:** This figure identifies in the literature on bank profitability two lines of analysis related with the determinants of the banks performance: 1). Internal determinants and 2) External determinants.

Duan & Niu (2020) propose another path of research focalized on the analysis of liquidity creation and bank profitability. These authors highlight that liquidity creation, related to the liability side, enhances bank profitability, while asset-side liquidity creation reduces bank profitability. Other authors who have studied the relationship between liquidity and bank profitability are Fernandes et al., (2021). They examine the effect of cash holdings on bank profitability using a worldwide database. Their results show that there is a nonmonotonic relationship between the cash conversion cycle and bank profitability. Other studies that have shown the relationship between liquidity levels and bank profitability are Davydov et al., (2021); Evans & Haq (2021).

The second stream of research focuses on the relationship of external factors and bank profitability. This group of factors includes fundamental macroeconomic determinants and market factors, such as competence and market share (Akhter & Daly, 2009; Bolt et al., 2012; de Mendonça & Silva, 2018; Kanas et al., 2012; Mohanty et al., 2021).

Regarding this macroeconomic environment, Kanas et al., (2012) show that the profitability of banks is affected by the economic cycle, short-term interest rates, inflation expectations, credit risk, and the effect of the structure of the loan portfolio on profits. Similar results are reported by Martins et al., (2019) for real estate banks in the United States, the United Kingdom, and Germany, where they point out that profitability depends on macroeconomic characteristics, such as volatility of interest rates and GDP. Along the same lines, Alessandri & Nelson (2015) show that in the long term, high interest rates present a positive relationship with profitability and bank margins; Molyneux et al., (2019), on their side, show that bank yields fell due to the rise of negative interest rates, although this effect also depended on the bank's structure and factors such as size and financing sources.

Studies have explored different relationships between the cyclical determinants of bank profitability and a variety of financial and monetary variables (e.g., Borio et al., 2017; Detragiache et al., 2018; Elekdag et al., 2020; Molyneux & Thornton, 1992). Some studies have focused their attention on the effects of macroeconomic dynamics on the profitability of banks in the European Union (Albertazzi & Gambacorta, 2009; Dietrich & Wanzenried, 2011; Djalilov & Piesse, 2016; Pasiouras & Kosmidou, 2007).

We fill an important gap in the literature by exploring all possible sources of macroeconomic profitability from a big data perspective. Our approach builds and expands on insights of past literature that link bank profitability with the macroeconomy.

3. Methodology

Our methodology consists of four sections. First, we present dynamic factor models as described by Bai & Ng (2002), which are used to explain the unobservable market-wide forces that determine banking profitability. Second, we present the statistical criteria used to determine the number of factors that suffice to construct our models. Third, we present our factor estimation method based on regularized PCA (Josse & Husson , 2012). Finally, in the fourth section, we show how we can gain insight into the factor's interpretability using the marginal coefficient of determination R^2_{Adj} from linear regressions of the estimated banking factors on a comprehensive set of macroeconomic series.

3.1. Dynamic Factor Models

A factor model is used to quantitatively measure the sensitivity of banks to unobservable systemic forces. We also establish how much of the dynamics of bank profitability are due to the idiosyncratic characteristics of banks and to common systemic forces.

We adapt our exposition form Bai & Ng (2002) and Bai & Ng (2008). Let y_{it} be the observed profitability data (either ROE, ROA, and EBITDA) for the i-th unit of the cross-section (i.e., bank) at time t, for $i = 1, \dots, N$ and t$= 1, \dots, T$. Consider the following model:

$$y_{it} = \lambda_i' \mathrm{F}_t + e_{it}$$

$$y_{it} = C_{it} + e_{it}, \tag{1}$$

where F_t is a vector of common factors, e_{it} is the idiosyncratic component of y_{it} and λ_i is a vector of factor loads associated with F_t. This is a vector of weights that unit i places on the corresponding r static common factors F_t. The term $C_{it} = \lambda_i F_t$ refers to the common component of the model. From Equation 1, letting $Y_t = (y_{1t}, y_{2t}, \dots, y_{Nt})'$, $F = (F_1, F_2, \dots, F_T)'$ and $\Lambda = (\lambda_1, \dots, \lambda_N)'$ in vector form, we have:

$$Y_t = \Lambda F_t + e_t, \tag{2}$$

Even though the model identifies a static relationship between y_{it} and F_t, F_t itself can be a dynamic vector process that evolves according to $A(L)F_t = u_t$,, where $A(L)$ represents a polynomial of the lag operator. The idiosyncratic error e_{it} can also be a dynamic process. The static model can be compared with the dynamic factor model defined as follows:

$$y_{it} = \lambda_i'(L) f_t + e_{it}, \tag{3}$$

where $\lambda_i(L) = (1 - \lambda_{i1}L - \cdots - \lambda_{is}L^s)$ is a vector of dynamic factor loadings of order s. When s is finite, we have a dynamic factor model. The law of motion governing the factor dynamics is given by:

$$f_t = C(L)\varepsilon_t, \tag{4}$$

where ε_t are *iid* errors. The dimension of f_t is the same dimension as ε_t, and it refers to the number of dynamic factors, denoted by q. Dynamic factor models with finite s are represented as static factor models with r finite; however, the dimension of F_t is often different from the dimension of f_t because F_t includes the lags and leads of f_t with $r \geq q$ (Bai & Ng, 2007). More generally, if we have q dynamic factors, we obtain $r = q(s+1) \geq q$ static factors.

Under this specification, it is necessary to determine the optimal number of estimated banking factors $\hat{\mathrm{F}}$. The factor loads associated with $\hat{\mathrm{F}}$ do not change over time, and they quantify the risk associated with each of the estimated profitability factors. $\hat{\lambda}_{0i}$ measures the part of the system's profitability, which is static and largely associated with a bank fixed effect. The larger the magnitude of this estimate, the less susceptible a bank is to fluctuating common variations. However, our main interest is to quantify how much of the system variation can be explained by common forces and how much by bank idiosyncrasies, so instead of analysing the intercepts of the model in Equation 3, we need to focus our attention on the slope coefficients known as the factor loadings $\hat{\lambda}_{1i}, \hat{\lambda}_{2i}$ ... and the residual variation e_{it}.

3.2. Selecting the number of factors

A strategy frequently used to estimate the optimal number of factors (r) conforming to the low-dimensional factor structure of a given system is the graphical representation of the system's eigenvalues. Specifically, we could use the point where the graph changes slope as an estimate of r. In contrast, a more transparent and quantitative way to determine the number of factors in the system is to balance the cost of adding an additional factor with increasing model complexity, and it was proposed by Bai & Ng (2020). The number of optimal factors $\hat{k}$ is obtained from the estimation of the corresponding factor loadings that accompany the observed factors, which can be consistently estimated from $\hat{k} = \arg min_{0 \leq k \leq kmax} PC(k)$ with $r \leq kmax$. Let:

$$PC(k) = V(k, \hat{F}^k) + kg(N, T), \tag{5}$$

where $PC(k)$ is the loss function. It is used to estimate $\hat{k}$, where $g(N, T)$ is an overfitting penalty, $\hat{F}^k$ is the matrix of k estimated factors and $V(k, \hat{F}^k)$ denotes the sum of squared residuals, as specified in Equation 6.

$$V(k, \hat{F}^k) = \min_{\Lambda} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (X_{it} - \lambda_i^{k'} \hat{F}_t^k)^2, \tag{6}$$

Another criterion by which r can be estimated consistently is

$$IC(k) = \ln \left(V(k, \hat{F}^k) \right) + kg(N, T), \tag{7}$$

The IC criterion thus resembles information criteria frequently used in time-series analysis, with the important difference that the penalty here depends on both N and T. It is often used in empirical work such as Bai & Ng (2019) and is obtained when

$$g(N, T) = \frac{(N + T)}{NT} \log \left(\frac{NT}{N + T} \right). \tag{8}$$

In such a way, r can be determine by

$$\hat{r} = \min_{k=0,\dots,rmax} IC(k), \tag{9}$$

We also verify that our estimated number of factors agrees with the empirical recommendation of using a relatively low number of factors, i.e., between three to seven, when working with panels of similar size to ours (Bai & Ng, 2007; Chudik & Pesaran, 2015; Chuliá et al., 2017; Gilchrist et al., 2009; Stock & Watson, 2005). We also estimate the number of dynamic factors, q, in our system following the path traced back to Bai & Ng (2007).

3.3. The (regularized) iterative PCA algorithm

In the estimation of the banking profitability factors $\hat{\mathrm{F}}$, we use the method of regularized principal components, due to Josse & Husson (2012). The objective is to determine a subspace that reduces the distances between individuals and their projections in this subspace. This is equivalent to finding two matrices $\mathrm{F}_{T\times S}$ and $\mathrm{U}_{N\times S}$ of rank S, with $S < T,$ that provide the best approximation of the matrix of the original dataset $\mathrm{X}_{T\times N}$, with *T: Time* and *N: Banks*, in the least-squares sense, that minimize the following criteria:

$$\vartheta = \left|\left|\mathrm{X} - \mathrm{M} - \mathrm{FU}'\right|\right|^2 = \sum_{t=1}^{T}\sum_{n=1}^{N}\left(X_{tn} - m_n - \sum_{s=1}^{S} F_{ts}U_{ns}\right)^2, \tag{10}$$

where M is a matrix of size $T \times K$ with each row equal to $(m_1, \dots, m_N)$, i.e., the vector with the mean of each variable. A common technique that deals with missing values in PCA is to ignore missing values by minimizing the least-squares criterion in Equation 10 overall non-missing entries. This can be achieved by introducing a weighted matrix W in the criterion, where $\mathrm{W}_{tn} = 0$ if X_{tn} is missing or $\mathrm{W}_{tn} = 1$ otherwise:

$$\vartheta = \sum_{t=1}^{T}\sum_{n=1}^{N} W_{tn}\left(X_{tn} - m_n - \sum_{s=1}^{S} F_{ts}U_{ns}\right)^2. \tag{11}$$

The iterative (regularized) PCA algorithm minimizes the criterion in Equation 11. It consists of the following steps:

1- Initial values such as the mean of each variable are used to replace missing values.

2- The second step of the iterative (regularized) algorithm is conducting PCA using the complete data set. Then, you impute the missing values with the reconstruction formulas (regularized) of order *ncp* (the fitted matrix calculated with *ncp* components for the scores and loads (regularized)). The number of components used (*ncp*) for the imputation of missing data is calculated by cross-validation in such a way that the mean square error of prediction is minimized.

   a) The optimal number of *ncp* components is estimated by cross-validation. The PCA is performed on the complete data set to estimate the parameters $\hat{\mathrm{M}}^{\ell}, \hat{\mathrm{F}}^{\ell}, \hat{\mathrm{U}}^{\ell}$.

b) Missing values are imputed with adjusted values $\widehat{\mathrm{X}}^{\ell} = \widehat{\mathrm{M}}^{\ell} + \widehat{\mathrm{F}}^{\ell}\,\widehat{\mathrm{U}}^{\ell}$; the new imputed data set is $X^{\ell} = W * X + (1 - W) * \widehat{\mathrm{X}}^{\ell}$; here the missing values are replaced by the fitted values.

3- Steps are repeated 2-a) and 2-b) until convergence is achieved.

The output of the algorithm is used to estimate the banking factors $\widehat{\mathrm{F}} \equiv \widehat{\mathrm{F}}_{Bank}$ that determine the profitability of banks, where the solution satisfies the following two Equations 12-13:

$$\widehat{\mathrm{U}}' = \left(\widehat{\mathrm{F}}'\widehat{\mathrm{F}}\right)^{-1}\widehat{\mathrm{F}}'\left(X - \widehat{\mathrm{M}}\right), \tag{12}$$

$$\widehat{\mathrm{F}}' = \left(X - \widehat{\mathrm{M}}\right)\widehat{\mathrm{U}}\left(\widehat{\mathrm{U}}'\widehat{\mathrm{U}}\right)^{-1}. \tag{13}$$

3.4. Interpreting the factors via marginal R-squared

We explore interpretability of the estimated factors, $\widehat{\mathrm{F}}_{Bank_p}$ with $p = 1, 2, 3$. To this end, we used FRED-QD database comprising of 248 macroeconomic series, we denote each series by MS_k, where $k = 1, \dots, 248$. The objective is to establish, based on the marginal determination coefficient of exahustive regressions, the top 5 macroeconomic variables that best explain each banking factor i. We then have 248 values of R^2_{Adj} extracted from each k model:

$$\widehat{\mathrm{F}}_{Bank_p} = \gamma_0 + \gamma_1 MS_k + u_k, \tag{14}$$

This procedure is carried out for the complete sample of banks for each financial indicator, as well as conditioning on bank size.

4. Data

Our data for US banking profitability and the macroeconomic environment come from two different sources. In the case of banking information, the data used were obtained from the Refinitiv database and correspond to profitability information measured through ROE (return on equity), ROA (return on assets), and EBITDA margin. ROA and ROE are sampled quarterly from 2000Q1 - 2020Q4. Hence, we have 111 banks for ROA and 118 for ROE. In the case of EBITDA margin, our data are sampled annually from 2001 to 2020 for a total of 241 banks. The number of banks differs by financial indicator since banks that presented more than 10 quarters with missing values were discarded. In addition to the profitability data, market capitalization data were retrieved from Refinitiv. This variable was used to divide the banks according to their size: small, medium, and large, corresponding to banks with market capitalization between the 1^{st} and 33^{rd} percentiles, 33^{rd} and 66^{th} percentiles, and 66^{th} to 99^{th} percentiles of the cross-sectional distribution as of March 2021.

In the case of macroeconomic information, the FRED-QD database assembled by McCracken & Ng (2021) was used. This big data set is publicly available and maintained by the Federal Reserve of St. Louis, through FRED service. This database is more ample than the traditional benchmark dataset provided by Stock & Watson (2012). In total, it consists of 248 series with quarterly frequency starting at 1959Q1. Many of these series have been aggregated from monthly series used in previous works by McCracken & Ng (2016), where they advance a collection of 128 macroeconomic series with monthly frequency. These macroeconomic series are classified into 14 groups following Stock & Watson (2012): NIPA (National Income and Product Accounts), Industrial Production, Employment and Unemployment, Housing, Inventories, Orders and Sales, Prices, Earnings and Productivity, Interest Rates Money and Credit, Household Balance Sheets, Exchange Rates, Other, Stock Markets and Non-Household Balance Sheets.

Using these two sources of information and after carrying out preprocessing and data cleaning, a single analysis period was consolidated from 2002Q1 to 2020Q4, consisting of 75 observations over time and per bank, in the case of ROE and ROA. In the case of EBITDA margin, the period 2001 to 2020 is covered, equivalent to 20 observations over time per bank. Table 1 shows summary statistics of profitability according to the bank's market capitalization. The table shows time averages of the cross-sectional mean, standard deviation, medians, skewness, and kurtosis. The size of the bank influences the different statistics; for instance, the larger the bank is, the larger the profitability. On their side, the variance is very similar for the three groups of banks, although the annual measures present a greater standard deviation. EBITDA presents a larger negative skewness and greater kurtosis than the quarterly indicators. Table 1 summary statistics indicate that although the three indicators of profitability are similar, they also seem to convey different information, thus emphasizing the convenience of considering the three of them in our calculations.

Table 1. Summary statistics of financial indicators

	Mean (%)	Sd (%)	Median (%)	Skewness	Kurtosis
ROE N=111 ; T = 75 Q					
All	9.67	2.98	9.63	0.03	0.96
Small	9.48	3.15	9.63	-0.42	0.54
Medium	9.70	3.01	9.61	0.35	2.40
Large	9.82	2.86	9.94	0.32	0.26
ROA N=118 ; T = 75 Q					
	Mean (%)	**Sd (%)**	**Median (%)**	**Skewness**	**Kurtosis**
All	0.82	0.23	0.81	0.22	0.87
Small	0.81	0.25	0.82	-0.15	0.34
Medium	0.81	0.22	0.80	0.14	1.19
Large	0.86	0.22	0.83	0.91	1.29
Margin EBITDA N=241 ; T = 20 Y					
	Mean (%)	**Sd (%)**	**Median (%)**	**Skewness**	**Kurtosis**
All	36.83	14.46	37.87	-1.89	11.16
Small	32.59	14.86	33.86	-0.54	1.85
Medium	38.36	14.52	39.61	-4.57	32.13
Large	39.86	12.88	39.84	-0.63	4.38

Note: Statistics estimated as mean, standard deviation (Sd), median, are presented as percentages (%). N denotes the number of banks used in each type of financial indicator. T denotes the number of periods in the sample, with Q: Quarter Y: Years.

5. Results:

In 5.1, we present the estimated factors of banking profitability and three factors of economic activity and prices for the US economy. In section 5.2, we estimate the marginal R-squared of the regression of each banking profitability factor on each of the series conforming to the big data macroeconomic set, aiming to gain interpretability. Finally, in section 5.3, we present the factor loads, and we rank financial institutions according to their exposure (either negative or positive) to the three sources of profitability dynamics.

5.1. Banking profitability and macroeconomic factors

We use the codes provided by the author of the original methodology, S. Ng[5], on her website to estimate the optimal number of factors. Once we determine the number of factors, we follow the approach proposed by Josse & Husson (2012) and describe the methodology used to estimate such factors.

Table 2 shows the percentage of the total variance in the panel of each of our three profitability indicators, ROE, ROA, EBITDA, explained by the three estimated banking profitability factors. In the table, small, medium and large banks correspond to those below the 3rd percentile, between the 33rd and 66th percentiles, and above the 66th percentile of market capitalization, respectively. The table also includes the sample size, N, used in each case, alongside the sampling frequency, which can be either quarterly, Q, or yearly, Y. In Table 2, we observe a high percentage of the total system variation explained by the three factors selected in the first step, namely, between 63% and 68% for the three profitability indicators. This result is robust to using ROE, ROA and even EBITDA margin (N=241), which is sampled at a different frequency and consists of a cross-sectional sample size that more than doubles that of ROE (N=111) and ROA (N=118). It is also robust to conditioning the results on the size of banks. Nevertheless, the percentage of explained variance increases for large banks with respect to small banks. The fact that the three factors explain similar percentages of the total variance for ROE and ROA, rather than for EBIDTA, confirms our number of selected factors. In short, it means that there are three main forces underling the profitability dynamics in the banking industry and that large banks are more susceptible to profitability commonality than small banks.

In the Appendix, we compare the results presented in Table 2 with alternatives in the literature to estimate the factors, such as the NIPALS algorithm (Andrecut, 2009; Wold, 1966), classic PCA (Mardia et al., 1979), and factor-based imputation for missing data (Bai & Ng, 2019, 2021; Cahan et al., 2021), and we further confirm the robustness of our claims.

---

[5] http://www.columbia.edu/~sn2294/

Table 2. Summary of variance explained by banking factors according to market cap and financial indicator.

		ROE	ROA	Margin EBITDA
		N=111 ; T = 75 Q	N=118 ; T = 75 Q	N=241 ; T = 20 Y
	Banking Factor	Exp. Variance (%)	Exp. Variance (%)	Exp. Variance (%)
All	F1	46.3	54.7	38.3
	F2	11.6	7.9	20.1
	F3	5.5	5.6	9.1
	Total	63.4	68.2	67.5
Small	F1	38.6	47.7	34.6
	F2	12.1	10.2	19.6
	F3	7.2	6.4	10.8
	Total	57.9	64.3	65.0
Medium	F1	52.4	58.0	39.6
	F2	12.2	8.2	17.8
	F3	5.4	5.4	10.9
	Total	70.0	71.6	68.3
Large	F1	50.8	61.7	48.7
	F2	11.5	7.5	19.7
	F3	6.7	5.2	6.6
	Total	69.0	74.4	75.0

Note: F1, F2, F3, specifically denote the estimated banking factors by the (regularized) iterative PCA algorithm. N denotes the number of banks used in each type of financial indicator. T denotes the number of periods in the sample, with Q: Quarter Y: Years.

In what follows, to ease the exposition and given the similarities of the results for our three profitability indicators, we focus the exposition on ROE, while the results regarding ROA and EBITDA can be found in the Appendix. Figure 2 shows the three banking factors for the total sample (solid line) and divides them according to bank size (dotted lines). Factor 1 presents a clear contraction in the wake of the Global Financial Crisis (GFC) in 2008-2009 and during the COVID-19 crisis in the first quarters of 2020. Factor 2 depicts smoother dynamics than factor 1, and it is depressed during the GFC. Factor 3 seems to recover from the crisis episodes around the GFC sooner than the other two factors. The three factors constructed with subsamples of small, medium and large banks present dynamics similar to those of the general factors, while the most noticeable divergence occurs during crisis episodes. In the case of ROA, Figure 3 shows that factor 1 has a downward trend long before the subprime crisis. Indeed, after 2005, the contraction becomes more pronounced. After the crisis, it presents a rapid recovery until reaching a maximum in 2019, and it suffers a great contraction again in the wake of COVID-19. In the case of factors 2 and 3, a sharp fall is evidenced in 2008, and then a gradual recovery is registered. These factors are far less sensitive to the COVID-19 crisis. The dynamics of the profitability factors, conditioning on bank size, are similar across categories.

The ROE indicator is obtained by dividing a bank's net profit by its total equity. ROE is the most widely used indicator to assess the profitability of a bank. The higher the ROE is, the greater the profitability that the bank can generate with the equity it uses to finance its operations. A measure greater than 10% is usually considered strong (Koch & MacDonald, 2015). On its side, ROA divides the profitability of the bank using total assets instead as the ratio denominator. A measure greater than 1% is considered strong (Koch & MacDonald, 2015). Differences observed in practice between the time evolution of ROA and ROE and, hence, between the estimated factors show that a shock to profitability affects the capital structure of a bank. In general, from Figures 2 and 3, we can observe that the dynamics of the estimated factors are very similar using either of the two indicators.

For the EBITDA margin, Figure 4 shows that factor 1 presents a large contraction during the GFC (2008-2009) and the COVID-19 crisis. This behaviour is similar to the dynamics reported with respect to ROE and ROA. Regarding factors 2 and 3, there is a slight contraction after the GFC. If we focus our analysis on the small and mid-sized banks in factor 2, we note that the dynamics are the opposite; here, the small and medium banks show a contraction during the crisis and subsequently recover. Regarding factor 3, if we analyse the medium-sized banks, we note that they present a contraction during the crisis and subsequent recovery, this behaviour is the opposite of that for factor 3.

Figure 5 shows three macroeconomic factors estimated from the FRED-QD database. As expected, extreme movements characterize the dynamics of the three macrofactors during crisis episodes, such as the GFC and the COVID-19 crisis. Interestingly, the factor's reaction to such episodes seems to occur before the depression recorded in banking factors 1 and 2.

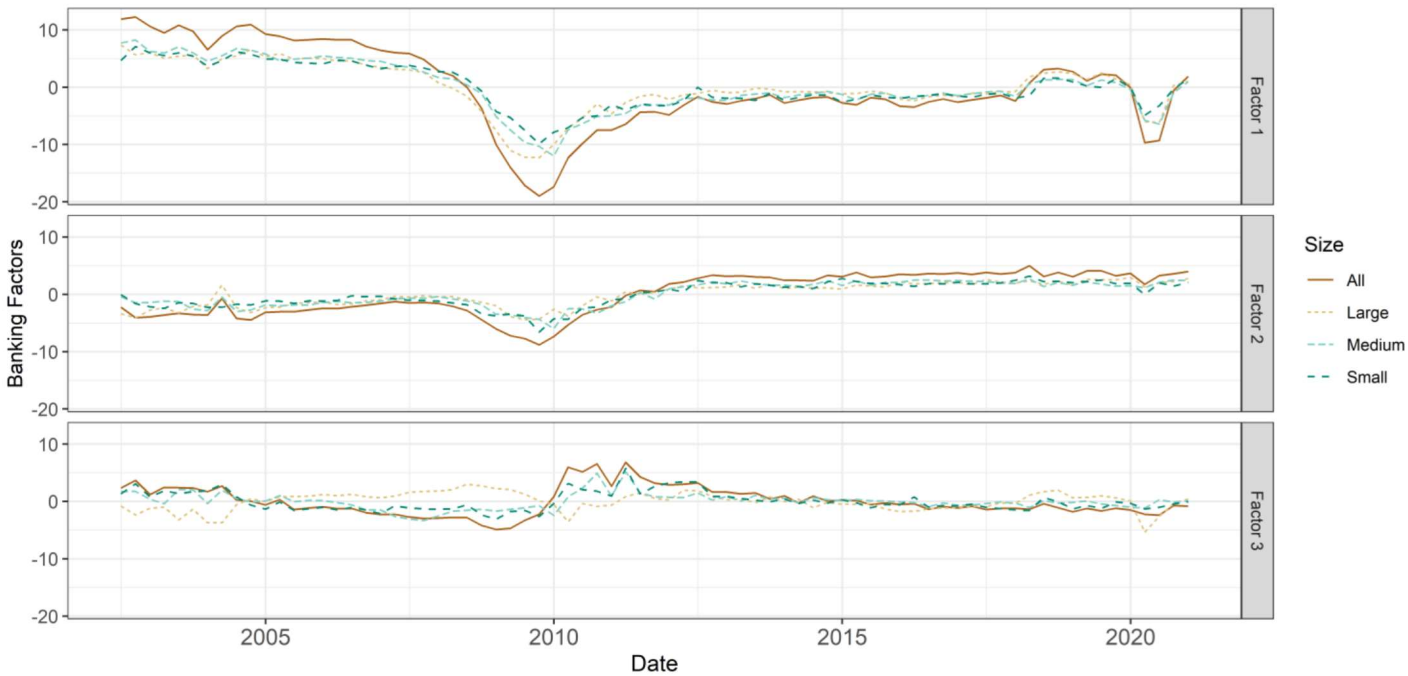

Figure 2. Banking factors from ROE. **Note:** This figure shows the three banking factors for the total sample (solid line) and divided according to the banks' size (dotted lines). These factors are estimated from the ROE information of 111 banks, using the iterative (regularized) PCA algorithm.

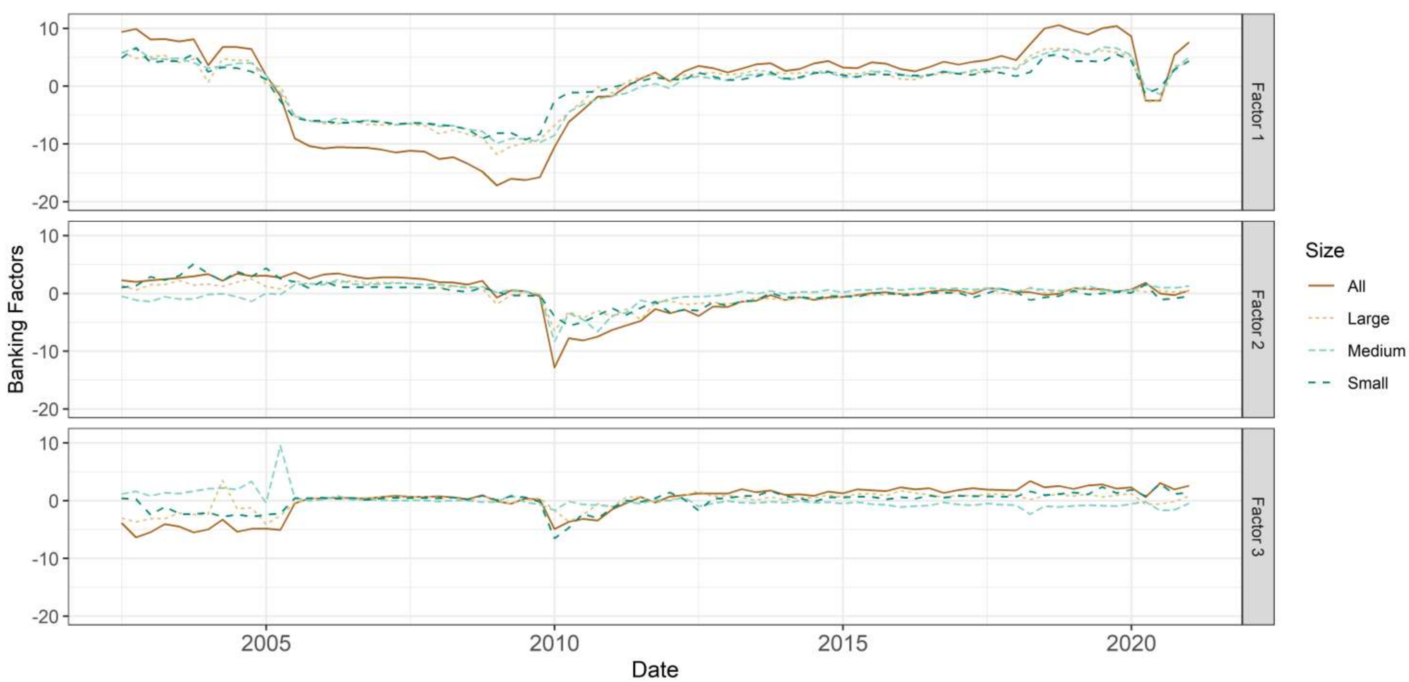

Figure 3. Banking factors from ROA. **Note:** This figure shows the three banking factors for the total sample (solid line) and divided according to the banks' size (dotted lines). These factors are estimated from the ROA information of 118 banks, using the iterative (regularized) PCA algorithm.

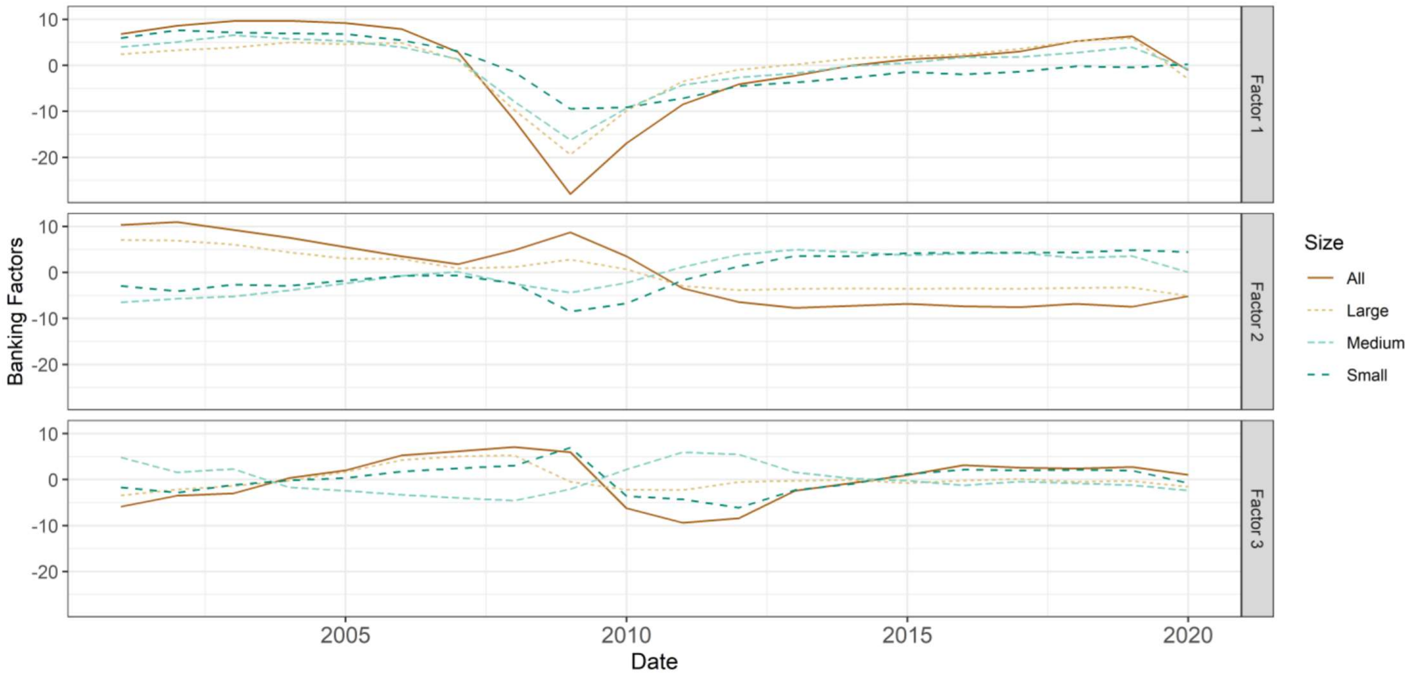

Figure 4. Banking factors from EBITDA. **Note:** This figure shows the three banking factors for the total sample (solid line) and divided according to the banks' size (dotted lines). These factors are estimated from the EBITDA information of 241 banks, using the iterative (regularized) PCA algorithm.

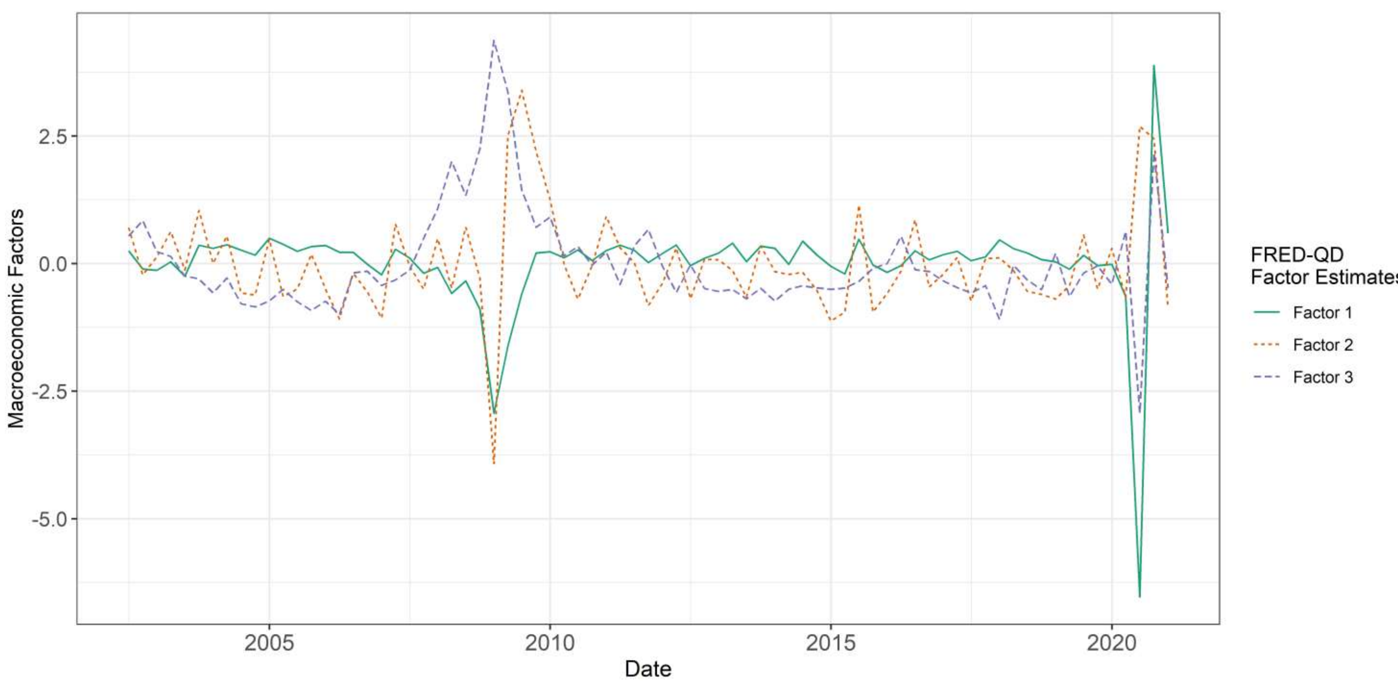

Figure 5. Macroeconomic factors from FRED-QD. **Note:** This figure shows the three macroeconomic factors for the total of 248 macroeconomic series from FRED-QD database, using the iterative (regularized) PCA algorithm.

5.2. Factor interpretation

Now, we turn our attention to the interpretation of our banking and macroeconomic factors. To do so, we estimate numerous, and exhaustive, bivariate OLS regressions in which the left-hand-side variable is the factor and the right-hand-side variable is each of the 248 macroeconomic series in the FRED-QD data set. From each regression, we keep the R-squared and construct a ranking, by factor, for the explanatory series, according to the predictive power of each variable on each factor. We interpret a factor according to the series it seems to be most closely related with. Table 3 shows the top 5 macroeconomic variables displaying great fit for the case of ROE factors. The variables are presented alongside their description and the group to which they belong, according to McCracken & Ng (2021). Note that we estimate three static factors, but only one dynamic or primitive factor. This means that even though each static factor is orthogonal to each other, by construction, there still exists only one primitive factor that determines the whole system dynamics. This primitive factor is likely related to the general and abstract concept of "economic activity", which includes characteristics of employability, income, financial burden, or industrial capacity utilization, etc.

The results show that factor 1 of banking profitability measured by ROE is mainly related to the level of household indebtedness (i.e., real estate loans and total real revolving credit owned and securitized) and with series that proxy for economic activity, particularly for industry capacity utilization. This first factor allows us to analyse how household indebtedness and industrial capacity, both of which are related to economic activity, are associated with bank profitability. In particular, the first static factor emphasizes the importance of mortgages and households' consumption for bank profitability (see for instance Jappelli et al. (2013)). Factor 2 is associated with household income, mainly with employment and the net worth of households and other economic actors. It could be rationalized as a forward-looking factor, according to which optimism on the side of households about income prospects, which can result in greater indebtedness and even greater interest payments. Factor 3 is clearly associated with companies' indebtedness and, to some extent, with government debt.

Regarding the identification of factors by bank size, we find a very similar dynamic for all factors; nevertheless, some differences are noteworthy. In the case of small banks, factor 1 is identical to factor 1 global (using the whole sample); in factor 2, some differences are perceived with respect to global factor 2, since for small entities, it is more closely associated with a series of liabilities of the nonfinancial business sector. Factor 3 is associated with the series with the net worth of households and nonprofit organizations, which are not as important for global factor 3. On the side of medium banks in comparison with the global factors, factor 1 is identical to global factor 1. Factor 2 includes a series of liabilities of the nonfinancial business sector, which are not included in the global factor, and factor 3 is more closely associated with a series of nonrevolving credits on property. Regarding large banks compared to global factors, factor 1 is more closely associated with the series of nonfinancial noncorporate real assets of the business sector; Factor 2 is more closely related to the series of liabilities of the nonfinancial business sector, while factor 3 makes a large difference since it is mainly associated with aspects of housing prices. The identification of the factors by bank size is presented in the Appendix.

Table 4 shows the variables that best explain the macroeconomic factors estimated from the FRED-QD database. In this case, the first factor is associated with income and production, the second factor is associated with prices, and the third factor is associated with real nonfinancial and household assets.

Figure 6 shows the correlation between the banking factors and the estimated macroeconomic factors. As expected, the correlation within factor groups (i.e., the profitability and macroeconomic sets) is zero because factors within groups are orthogonal by construction. On the other hand, we observe that the most correlated factors are household income (Banking 2) and financial conditions (Macro 3), with a negative correlation of -0.44.

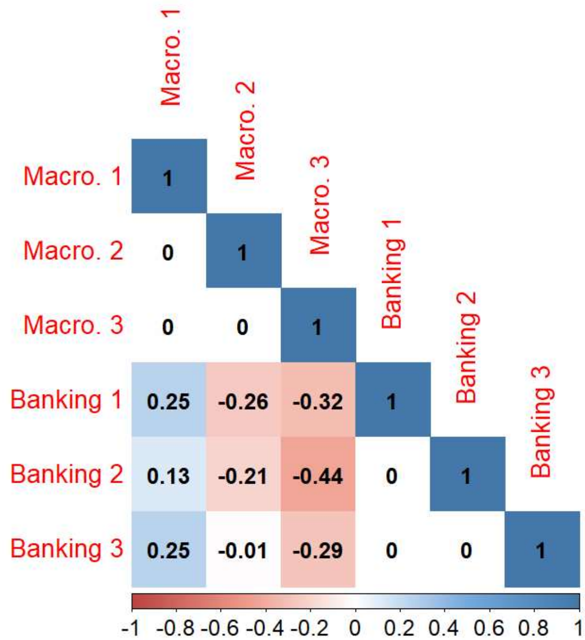

Figure 6. Banking and macroeconomic factor correlations. **Note:** This figure shows the correlation between the banking factors and the estimated macroeconomic factors. The intensity of the color shows the strength of association between the factors. The blue color denotes direct or positive association, while the red color denotes inverse or negative association.

Table 3. Top 5 of FRED's macroeconomic variables that best explain banking factors using ROE.

Banking Factors	Group	Description	R^2_{Adj}
		All	
	Household Balance Sheets	Real Total Liabilities of Households and Nonprofit Organizations (Billions of 2012 Dollars), deflated by Core PCE	0.66
	Industrial Production	Capacity Utilization: Manufacturing (SIC) (Percent of Capacity)	0.52
Factor 1	Industrial Production	Capacity Utilization: Total Industry (Percent of Capacity)	0.45
	Money and Credit	Real Real Estate Loans, All Commercial Banks (Billions of 2012 U.S. Dollars), deflated by Core PCE	0.39
	Money and Credit	Total Real Revolving Credit Owned and Securitized, Outstanding (Billions of 2012 Dollars), deflated by Core PCE	0.34
	Employment and Unemployment	Average Weekly Hours of Production and Nonsupervisory Employees: Manufacturing (Hours)	0.67
	Employment and Unemployment	Help-Wanted Index	0.59
Factor 2	Household Balance Sheets	Net Worth of Households and Nonprofit Organizations Relative to Disposable Personal Income (Percent)	0.34
	Employment and Unemployment	All Employees: Financial Activities (Thousands of Persons)	0.23
	Employment and Unemployment	All Employees: Construction (Thousands of Persons)	0.23
	Interest Rates	3-Month Commercial Paper Minus 3-Month Treasury Bill, secondary market (Percent)	0.33
	Non-Household Balance Sheets	Nonfinancial Noncorporate Business Sector Liabilities to Disposable Business Income (Percent)	0.31
Factor 3	Money and Credit	FRB Senior Loans Officer Opions. Net Percentage of Domestic Respondents Reporting Increased Willingness to Make Consumer Installment Loans	0.31
	Non-Household Balance Sheets	Real Nonfinancial Noncorporate Business Sector Liabilities (Billions of 2012 Dollars), Deflated by Implicit Price Deflator for Business Sector IPDBS	0.30
	NIPA	Real government state and local consumption expenditures (Billions of Chained 2012 Dollars), deflated using PCE	0.27

Note: This table presents the macroeconomic series of the FRED-QD most correlated with each of the banking factors estimated according to the $\boldsymbol{R^2_{Adj}}$ criterion. Each of the series is classified into a group determined by Stock & Watson (2012).

Table 4. Top 5 of FRED's macroeconomic variables that best explain macroeconomic factors.

Macroeconomic Factors	Group	Description	R^2_{Adj}
Factor 1	NIPA	Manufacturing Sector: Real Output (Index 2012=100)	0.91
	Industrial Production	Industrial Production: Manufacturing (SIC) (Index 2012=100)	0.91
	NIPA	Real Gross Domestic Product, 3 Decimal (Billions of Chained 2012 Dollars)	0.91
	NIPA	Business Sector: Real Output (Index 2012=100)	0.91
	Employment and Unemployment	Nonfarm Business Sector: Hours of All Persons (Index 2012=100)	0.90
Factor 2	Prices	Producer Price Index by Commodity Intermediate Materials: Supplies & Components (Index 1982=100)	0.57
	Prices	Personal Consumption Expenditures: Chain-type Price Index (Index 2012=100)	0.57
	Prices	Personal consumption expenditures: Goods (chain-type price index)	0.57
	Prices	Consumer Price Index for All Urban Consumers: Commodities (Index 1982-84=100)	0.57
	Prices	Consumer Price Index for All Urban Consumers: All items less shelter (Index 1982-84=100)	0.57
Factor 3	Stock Markets	CBOE S&P 100 Volatility Index: VXO	0.52
	Non-Household Balance Sheets	Real Nonfinancial Noncorporate Business Sector Assets (Billions of 2012 Dollars), Deflated by Implicit Price Deflator for Business Sector IPDBS	0.49
	Household Balance Sheets	Real Total Assets of Households and Nonprofit Organizations (Billions of 2012 Dollars), deflated by Core PCE	0.48
	Household Balance Sheets	Real Net Worth of Households and Nonprofit Organizations (Billions of 2012 Dollars), deflated by Core PCE	0.47
	Non-Household Balance Sheets	Real Nonfinancial Noncorporate Business Sector Net Worth (Billions of 2012 Dollars), Deflated by Implicit Price Deflator for Business Sector IPDBS	0.47

Note: This table presents the macroeconomic series of the FRED-QD most correlated with each of the macroeconomic factors estimated according to the R^2_{Adj} criterion. Each of the series is classified into a group determined by Stock & Watson (2012).

5.3. Sensitivity of banks to profitability factors

In this subsection, we estimate individual models for each bank in the study sample using the dynamic factor model presented in Equation 1. Figure 7 shows the kernel of the probability distribution of the adjusted R-squared, $\boldsymbol{R^2_{Adj}}$, alongside summary statistics of the distribution. In this way, we can evaluate the individual adjustment of the banking profitability dynamic factor model in the cross-section of banks. According to Figure 7, $\boldsymbol{R^2_{Adj}}$ **is** greater than 64% for more than half of the banks and above 77% for more than a quarter of the banks. This shows a high adjustment of the dynamic factor model for the panel of financial institutions. On average, there is an adjustment of 60%, and some banks present adjustments greater than 90%.[6]

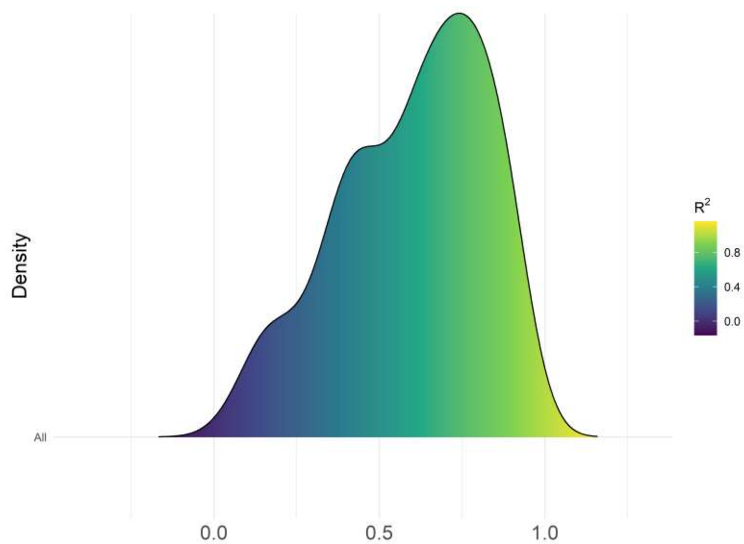

Figure 7. Distribution of R^2 of the model by factors. **Note:** This figure shows the density about $\boldsymbol{R^2_{Adj}}$ for the set of regressions of the dynamic factor models from ROE.

In Figure 8, we present the distribution of $\boldsymbol{\beta_i}$ for $i = 0, 1,2,3$, which corresponds to the intercepts $\boldsymbol{\beta_0}$ and the factor loads of our model. $\boldsymbol{\beta_0}$ shows an estimation of banking profitability, which is not time varying, that is, a fixed effect of banking profitability. $\boldsymbol{\beta_0}$ has a distribution concentrated at

[6] In the appendix, we show the ranking of the banks according to their goodness of fit, including the banks for which our model better explains profitability.

approximately 9.63%, with half of the sample between 8.14% and 11.03%. The right panel of the figure shows the distribution of $\boldsymbol{\beta_1}, \boldsymbol{\beta_2}, \boldsymbol{\beta_3}$. Analysing such a distribution allows us to identify the banks that are more vulnerable to shocks to each of the estimated factors. In the case of $\boldsymbol{\beta_1}$, it measures a bank's sensitivity to household burden and capacity utilization; for $\boldsymbol{\beta_2}$**,** it constitutes a risk indicator associated with household income; and $\boldsymbol{\beta_3}$ represents an indicator of risk according to corporate sector indebtedness. The exposure of each bank to each factor depends naturally on the business model and a bank's specialization and market target. Regarding $\boldsymbol{\beta_1}$, most banks present values greater than zero, between 0.26% and 0.63%. In the case of $\boldsymbol{\beta_2}$ and $\boldsymbol{\beta_3}$, there is greater variability, and approximately 50% of the banks present values greater than zero (see Table 5).

In Figure 9, we present the ranking of banks according to our three profitability factors. The spiral figure shows the magnitude of the estimated $\boldsymbol{\beta_i}$, which corresponds to the risk exposure to each factor. In Table 6, we show a summary of the top 10 banks according to the risk exposure to each systemic factor.

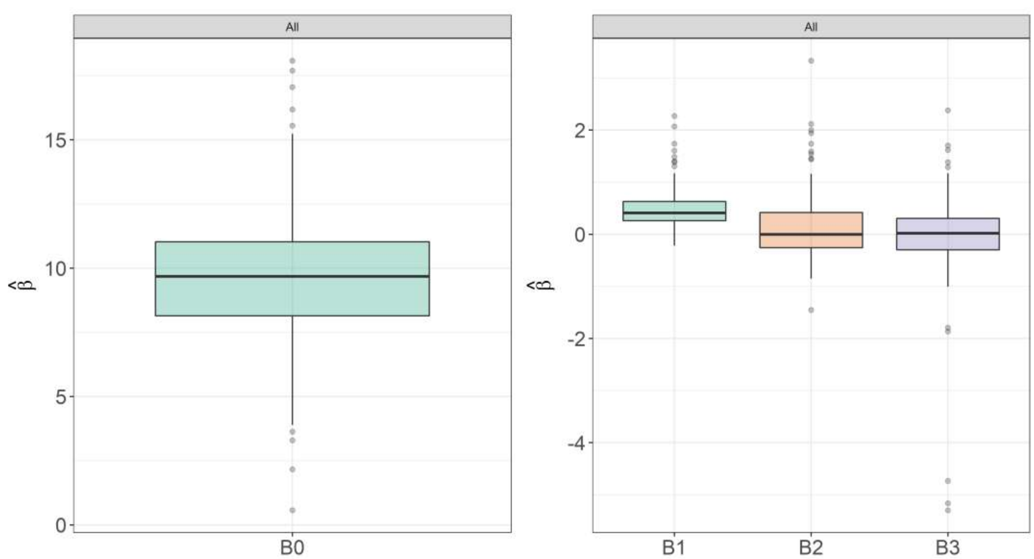

Figure 8. Distribution of model estimators by factors. **Note:** The left panel shows the distribution of the $\boldsymbol{\beta_0}$ estimator that represents the fixed effect. The right panel shows the distribution of the estimates of the effect $\boldsymbol{\beta_1}, \boldsymbol{\beta_2}, \boldsymbol{\beta_3}$ of the banking factors. Where $\boldsymbol{\beta_1}$ measures a bank's sensitivity to household burden and capacity utilization. $\boldsymbol{\beta_2}$ constitutes a risk indicator associated with household income and $\boldsymbol{\beta_3}$ represents an indicator of risk obeying corporate sector indebtedness.

Table 5. Summary statistics of estimators

Statistic	$\boldsymbol{\beta_0}$	$\boldsymbol{\beta_1}$	$\boldsymbol{\beta_2}$	$\boldsymbol{\beta_3}$
Min	0.57	-0.22	-1.45	-5.30
Q1	8.14	0.26	-0.26	0.30
Median	9.68	0.41	0.00	0.02
Q3	11.03	0.63	0.42	0.31
Mean	9.67	0.52	0.18	-0.07
St. Dev.	3.05	0.42	0.71	1.04
Max	18.08	2.27	3.33	2.38

Note: This table presents summary statistics for each one effects estimated. The statistics: minimum, maximum, mean, median, standard deviation, quantiles 25th and 75th are in percentage.

Figure 9. Total ranking of banks by risk indicators. **Note:** Each spiral shows the ranking of the banks with respect to the estimate of the effect of each banking factor. The ranking starts clockwise with the least vulnerable banks in each factor and ends with the most vulnerable banks in each factor. If the bar points towards the inside of the chart, it is because the estimate is negative.

Table 6. Top 10 banks according to the estimated parameters of the model by factors using ROE.

	Bank	R^2_{Adj}	Bank	β_0	Bank	β_1	Bank	β_2	Bank	β_3
					Top ranking of banks					
1	Westamerica Bancorp	0.927	Bank of Hawaii Corp	18.08	Central Pacific Financial Corp	2.27	Central Pacific Financial Corp	3.33	Popular Inc	2.38
2	Webster Financial Corp	0.921	Westamerica Bancorp	17.69	Synovus Financial Corp	2.07	Synovus Financial Corp	2.12	Huntington Bancshares Inc	1.71
3	Fulton Financial Corp	0.904	Bank Ozk	17.06	United Community Banks Inc	1.74	Huntington Bancshares Inc	2.00	Republic Bancorp Inc	1.62
4	Trustmark Corp	0.900	U.S. Bancorp	16.18	Seacoast Banking Corporation of Florida	1.60	Seacoast Banking Corporation of Florida	1.94	First Busey Corp	1.38
5	Synovus Financial Corp	0.898	City Holding Co	15.55	First Busey Corp	1.49	United Community Banks Inc	1.74	Bank Ozk	1.29
6	Hope Bancorp Inc	0.887	Stock Yards Bancorp Inc	15.24	Popular Inc	1.40	Popular Inc	1.59	Hanmi Financial Corp	1.17
7	East West Bancorp Inc	0.885	First Financial Bankshares Inc	14.24	Huntington Bancshares Inc	1.39	First Busey Corp	1.54	Financial Institutions Inc	1.15
8	Truist Financial Corp	0.884	Southside Bancshares Inc	13.73	Old Second Bancorp Inc	1.31	Regions Financial Corp	1.45	Citigroup Inc	1.04
9	Cullen/Frost Bankers Inc	0.884	Great Southern Bancorp Inc	13.62	First Horizon Corp (Tennessee)	1.17	Banner Corp	1.44	First Horizon Corp (Tennessee)	0.91
10	Zions Bancorporation NA	0.883	CVB Financial Corp	13.52	Regions Financial Corp	1.17	Financial Institutions Inc	1.16	Webster Financial Corp	0.85
					Lower ranking of banks					
1	South State Corp	0.065	Central Pacific Financial Corp	0.57	German American Bancorp Inc	-0.22	Westamerica Bancorp	-1.45	Central Pacific Financial Corp	-5.30
2	Republic Bancorp Inc	0.134	United Community Banks Inc	2.17	UMB Financial Corp	-0.11	Valley National Bancorp	-0.85	Old Second Bancorp Inc	-5.16
3	First Financial Corp	0.138	Seacoast Banking Corporation of Florida	3.29	Southside Bancshares Inc	-0.08	Southside Bancshares Inc	-0.82	United Community Banks Inc	-4.73
4	Bryn Mawr Bank Corp	0.144	Republic First Bancorp Inc	3.63	Brookline Bancorp Inc	-0.04	Republic Bancorp Inc	-0.74	Synovus Financial Corp	-1.87
5	UMB Financial Corp	0.158	Synovus Financial Corp	3.89	Republic Bancorp Inc	0.06	Sierra Bancorp	-0.72	Heritage Commerce Corp	-1.79
6	People's United Financial Inc	0.160	Banner Corp	4.46	Eagle Bancorp Inc	0.08	Bank Ozk	-0.67	Republic First Bancorp Inc	-1.01
7	Pacific Premier Bancorp Inc	0.180	Old Second Bancorp Inc	4.62	First Financial Corp	0.08	Bank of Hawaii Corp	-0.64	Sierra Bancorp	-0.91
8	First Financial Bancorp	0.185	Regions Financial Corp	5.40	First Financial Bankshares Inc	0.10	American River Bankshares	-0.60	American River Bankshares	-0.79
9	Berkshire Hills Bancorp Inc	0.231	Popular Inc	5.45	Community Bank System Inc	0.11	United Bankshares Inc	-0.60	Bank of Hawaii Corp	-0.79
10	Heritage Commerce Corp	0.239	Berkshire Hills Bancorp Inc	5.46	First Financial Bancorp	0.13	CVB Financial Corp	-0.56	Capital City Bank Group Inc	-0.74

Note: This table presents the consolidation of the top 10 banks according to the risk estimators associated with the estimated factors. Upper panel shows the banks most vulnerable. Lower panel shows the banks most robust respect to shocks macroeconomics.

Upper panel Table 6 shows the first 10 positions of the ranking with the highest $\boldsymbol{R^2_{Adj}}$ for $\boldsymbol{\beta_0}$, $\boldsymbol{\beta_1}$, $\boldsymbol{\beta_2}$, $\boldsymbol{\beta_3}$. The lower panel shows the 10 lowest positions according to the same criterion. The highest explanatory power of the general profitability factors that we identify here is shown by Westamerica Bancorp, Webster Financial Corp, and Fulton Financial Corp. Indeed, Central Pacific Financial Corp is the most vulnerable institution to the first and second factors associated with household indebtedness, economic activity and household income, respectively, followed by Synovus Financial Corp. Regarding the third factor, corporate debt, Popular Inc., is the most vulnerable financial institution. The following banks are jointly vulnerable to common factors according to the three factors: First Busey Corp, Popular Inc. and Huntington Bancshares Inc. In contrast, the following banks are only vulnerable to a single factor: Old Second Bancorp Inc, Banner Corp, Republic Bancorp Inc, Bank Ozk, Hanmi Financial Corp, Citigroup Inc. and Webster Financial Corp.

Focusing on the lower ranking, a lower β_0 is related to a greater vulnerability to the model common factors, such as Central Pacific Financial Corp, United Community Banks Inc., and Seacoast Banking Corporation of Florida. We note that the Central Pacific Financial Corp is vulnerable to factors one and two, but it is not vulnerable to factor three: commercial debt. The more countercyclical banks in the face of shocks to factors one and two are German American Bancorp Inc. and Westamerica Bancorp, respectively.

The heterogeneity of our ranking results demonstrates the convenience of the integrated approach that we propose to monitor bank profitability. The information conveyed by each factor is different and offers a new risk management perspective.

6. Conclusion

Using a sample of profitability indicators for the largest US banks according to market capitalization, we estimate the number of statistical factors that underlie profitability dynamics over time for the banking industry. Three factors are enough to describe 63.40%, 68% and 67.50% of ROE, ROA and EBITDA, respectively. The numbers increase according to bank size, which indicates that, as expected, the larger the bank is, the more cyclically it behaves. This provides a precise answer to the question of to what extent bank profitability is a matter of exposure to cyclical market forces related to the macroeconomy instead of bank-specific characteristics. Furthermore, we conduct an intensive search in a big data set comprising 248 macroeconomic variables for production, employment, housing, inventories, money and credit, and stock markets, among other groups, which are representative of the whole economy. Our results indicate that these three statistical factors are mainly related to households' financial burden and economic activity, households' income and employment and, in some cases, commercial loans (for ROA and ROE), while mortgage and housing markets are related in other cases (for EBITDA). Finally, we also provide a means to monitor profitability in the banking industry from an integrative perspective by establishing rankings of the financial institutions according to their exposure (either positive or negative) to the three market forces that we identify. The convenience of our approach is highlighted by a high adjustment of the factor models when explaining individual banks' profitability and the insights gained in terms of market monitoring after resorting to the integrative approach that we advance; i.e., while some banks are sensitive to specific banking factors, other banks are more sensitive to other market factors. Thus, ideally, we should keep track of the three factors simultaneously.

Our proposal is simple, yet intuitive and comprehensive; hence, it can be easily implemented by regulators and banking managers to keep track of market evolution and the most vulnerable financial institutions.

We focus on the largest 111-118 banks in the US system (241 for EBITDA) with more reliable information in our sample period. When we split the sample according to market capitalization, our three groups, namely, "large", "medium" and "small", are to be interpreted bearing in mind this caveat; hence, we do not truly consider the smallest financial institutions. Given that our results seem more relevant for large banks than for small banks, it could be that for banks outside of our sample, which are even smaller, the results lack the same validity. It would be interesting for future studies to explore this avenue by increasing the sample coverage.

References

Acharya, V., Engle, R., & Richardson, M. (2012). Capital shortfall: A new approach to ranking and regulating systemic risks. *American Economic Review*, *102*(3). https://doi.org/10.1257/aer.102.3.59

Adrian, T., & Brunnermeier, M. K. (2016). CoVaR. *American Economic Review*, *106*(7). https://doi.org/10.1257/aer.20120555

Akhter, S., & Daly, K. (2009). Bank health in varying macroeconomic conditions: A panel study. *International Review of Financial Analysis*, *18*(5), 285–293. https://doi.org/https://doi.org/10.1016/j.irfa.2009.09.002

Albertazzi, U., & Gambacorta, L. (2009). Bank profitability and the business cycle. *Journal of Financial Stability*, *5*(4), 393–409. https://doi.org/https://doi.org/10.1016/j.jfs.2008.10.002

Alessandri, P., & Nelson, B. D. (2015). Simple Banking : Profitability and the Yield Curve. *Journal of Money, Credit and Banking*, *47*(1). https://doi.org/10.1111/jmcb.12172

Andrecut, M. (2009). Parallel GPU implementation of iterative PCA algorithms. *Journal of Computational Biology : A Journal of Computational Molecular Cell Biology*, *16*(11). https://doi.org/10.1089/cmb.2008.0221

Bai, J., & Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, *70*(1), 191–221. https://doi.org/10.1111/1468-0262.00273

Bai, J., & Ng, S. (2007). Determining the number of primitive shocks in factor models. *Journal of Business and Economic Statistics*, *25*(1), 52–60. https://doi.org/10.1198/073500106000000413

Bai, J., & Ng, S. (2008). Large Dimensional Factor Analysis. *Foundations and Trends R in Econometrics*, *3*(2), 89–163. https://doi.org/10.1561/0800000002

Bai, J., & Ng, S. (2019). Rank regularized estimation of approximate factor models. *Journal of Econometrics*, *212*(1). https://doi.org/10.1016/j.jeconom.2019.04.021

Bai, J., & Ng, S. (2020). Simpler proofs for approximate factor models of large dimensions. *ArXiv*, *2018369*.

Bai, J., & Ng, S. (2021). Matrix Completion, Counterfactuals, and Factor Analysis of Missing Data. *ArXiv*, *1658770*.

Bolt, W., de Haan, L., Hoeberichts, M., van Oordt, M. R. C., & Swank, J. (2012). Bank profitability during recessions. *Journal of Banking & Finance*, *36*(9), 2552–2564. https://doi.org/https://doi.org/10.1016/j.jbankfin.2012.05.011

Borio, C., Gambacorta, L., & Hofmann, B. (2017). The influence of monetary policy on bank profitability. *International Finance*, *20*, 48–63. https://doi.org/10.1111/infi.12104

Cahan, E., Bai, J., & Ng, S. (2021). Factor-Based Imputation of Missing Values and Covariances in Panel Data of Large Dimensions. *ArXiv*. http://arxiv.org/abs/2103.03045

Chudik, A., & Pesaran, M. H. (2015). Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. *Journal of Econometrics*, *188*(2), 393–420. https://doi.org/10.1016/j.jeconom.2015.03.007

Chuliá, H., Guillén, M., & Uribe, J. M. (2017). Measuring uncertainty in the stock market. *International Review of Economics and Finance*, *48*(October 2016), 18–33. https://doi.org/10.1016/j.iref.2016.11.003

Davydov, D., Vähämaa, S., & Yasar, S. (2021). Bank liquidity creation and systemic risk. *Journal of Banking & Finance*, *123*, 106031. https://doi.org/https://doi.org/10.1016/j.jbankfin.2020.106031

de Mendonça, H. F., & Silva, R. B. da. (2018). Effect of banking and macroeconomic variables on systemic risk: An application of ΔCOVAR for an emerging economy. *The North American Journal of Economics and Finance*, *43*, 141–157. https://doi.org/https://doi.org/10.1016/j.najef.2017.10.011

Detragiache, B. E., Tressel, T., & Turk-ariss, R. (2018). Where Have All the Profits Gone? European Bank Profitability Over the Financial Cycle. *International Monetary Fund Working Paper 18/99*.

Dietrich, A., & Wanzenried, G. (2011). Determinants of bank profitability before and during the crisis: Evidence from Switzerland. *Journal of International Financial Markets, Institutions and Money*, *21*(3), 307–327. https://doi.org/https://doi.org/10.1016/j.intfin.2010.11.002

Djalilov, K., & Piesse, J. (2016). Determinants of bank profitability in transition countries: What matters most? *Research in International Business and Finance*, *38*, 69–82. https://doi.org/https://doi.org/10.1016/j.ribaf.2016.03.015

Duan, Y., & Niu, J. (2020). Liquidity creation and bank profitability. *The North American Journal of Economics and Finance*, *54*, 101250. https://doi.org/https://doi.org/10.1016/j.najef.2020.101250

Elekdag, S., Malik, S., & Mitra, S. (2020). Breaking the Bank? A Probabilistic Assessment of Euro Area Bank Profitability. *Journal of Banking & Finance*, *120*, 105949. https://doi.org/https://doi.org/10.1016/j.jbankfin.2020.105949

Evans, J. J., & Haq, M. (2021). Does bank capital reduce liquidity creation? *Global Finance Journal*, 100640. https://doi.org/https://doi.org/10.1016/j.gfj.2021.100640

Fernandes, G., Mendes, L. dos S., & Leite, R. de O. (2021). Cash holdings and profitability of banks in developed and emerging markets. *International Review of Economics & Finance*, *71*, 880–895. https://doi.org/https://doi.org/10.1016/j.iref.2020.10.018

Gilchrist, S., Yankov, V., & Zakrajšek, E. (2009). Credit market shocks and economic fluctuations: Evidence from corporate bond and stock markets. *Journal of Monetary Economics*, *56*(4), 471–493. https://doi.org/10.1016/j.jmoneco.2009.03.017

Jappelli, T., Pagano, M., Maggio, M. Di, Pagano, M., & Maggio, M. Di. (2013). Households' Indebtedness and Financial Fragility. *Journal of Financial Management, Markets and Institutions*, *I*(1), 23–46. https://doi.org/10.12831/73631

Josse, J., & Husson, F. (2012). Handling missing values in exploratory multivariate data analysis methods. *Journal de La Société Française de Statistique*, *153*(2), 79-99–99.

Kanas, A., Vasiliou, D., & Eriotis, N. (2012). Revisiting bank profitability: A semi-parametric approach. *Journal of International Financial Markets, Institutions and Money*, *22*(4), 990–1005. https://doi.org/https://doi.org/10.1016/j.intfin.2011.10.003

Koch, T., & MacDonald, S. (2015). *Bank Management* (8th Editio). Cengage Learning.

Kumar, V., Thrikawala, S., & Acharya, S. (2021). Financial inclusion and bank profitability: Evidence from a developed market. *Global Finance Journal*, 100609. https://doi.org/https://doi.org/10.1016/j.gfj.2021.100609

Le, T. D. Q., & Ngo, T. (2020). The determinants of bank profitability: A cross-country analysis. *Central Bank Review*, *20*(2), 65–73. https://doi.org/https://doi.org/10.1016/j.cbrev.2020.04.001

Mardia, K. V., Kent, J. T., & Bibby, M. (1979). *Multivariate Analysis*. London: Academic Press.

Martins, A. M., Serra, A. P., & Stevenson, S. (2019). Determinants of real estate bank profitability. *Research in International Business and Finance*, *49*, 282–300. https://doi.org/https://doi.org/10.1016/j.ribaf.2019.04.004

McCracken, M. W., & Ng, S. (2016). FRED-MD: A Monthly Database for Macroeconomic Research. *Journal of Business and Economic Statistics*, *34*(4), 574–589. https://doi.org/10.1080/07350015.2015.1086655

McCracken, M. W., & Ng, S. (2021). FRED-QD: A Quarterly Database for Macroeconomic Research. *Review*, *103*(1). https://doi.org/10.20955/r.103.1-44

Mohanty, D. K., Parida, A. K., & Khuntia, S. S. (2021). Financial market prediction under deep learning framework using auto encoder and kernel extreme learning machine. *Applied Soft Computing*, *99*, 106898. https://doi.org/https://doi.org/10.1016/j.asoc.2020.106898

Molyneux, P., Reghezza, A., & Xie, R. (2019). Bank margins and profits in a world of negative rates. *Journal of Banking & Finance*, *107*, 105613. https://doi.org/https://doi.org/10.1016/j.jbankfin.2019.105613

Molyneux, P., & Thornton, J. (1992). Determinants of European bank profitability: A note. *Journal of Banking & Finance*, *16*(6), 1173–1178. https://doi.org/https://doi.org/10.1016/0378-4266(92)90065-8

Pasiouras, F., & Kosmidou, K. (2007). Factors influencing the profitability of domestic and foreign commercial banks in the European Union. *Research in International Business and Finance*, *21*(2), 222–237. https://doi.org/https://doi.org/10.1016/j.ribaf.2006.03.007

Stock, J. H., & Watson, M. W. (2005). Implications of Dynamic Factor Models for VAR Analysis. *NBER Working Paper Series*, *11467*, 1–67.

Stock, J. H., & Watson, M. W. (2012). Disentangling the Channels of the 2007-2009 Recession. *Brookings Papers on Economic Activity*, 81–156.

Wold, H. (1966). Estimation of principal components and related models by iterative least squares. *Multivariate Analysis*, 391–420. http://ci.nii.ac.jp/naid/20001378860/en/

UBIREA
Institut de Recerca en Economia Aplicada Regional i Pública
Research Institute of Applied Economics
Universitat de Barcelona
Av. Diagonal, 690 • 08034 Barcelona

WEBSITE: www.ub.edu/irea/ • **CONTACT:** irea@ub.edu