



Munich Personal RePEc Archive

On the Construction of a Leading Indicator Based on News Headlines for Predicting Greek Deposit Outflows

Anastasiou, Dimitrios and Petralias, Athanassios

Athens University of Economics and Business, Alpha Bank

May 2021

Online at <https://mpra.ub.uni-muenchen.de/107602/>
MPRA Paper No. 107602, posted 10 May 2021 09:36 UTC

On the Construction of a Leading Indicator Based on News Headlines for Predicting Greek Deposit Outflows

Dimitrios Anastasiou^{†,1}

[†] Economic Research Division, Alpha Bank, Athens, Greece.

¹Department of Accounting and Finance, Athens University of Economics and Business

Email: anastasioud@aub.gr

Athanassios Petralias*

*Department of Statistics, Athens University of Economics and Business, Athens, Greece

Abstract

The purpose of this study is twofold. First, construct a leading indicator based on news headlines, and second, examine whether this novel indicator affects Greek bank deposit flows' trajectory. Employing alternative econometric methodologies, we find that this index proxies for depositors' crisis sentiment, and the higher this index becomes, the higher the depositors' negative sentiment becomes, leading them to withdraw their bank deposits. Monetary policy authorities or macroprudential regulators could adapt our model to assess the resilience of a bank or the whole banking sector.

Keywords: Bank deposit flows, news headlines; sentiment; uncertainty; Greece.

JEL classification: C22, C51, G10, E44

Disclaimer: The views and opinions expressed herein are those of the authors and do not reflect those of their respective institutions.

Funding: This research received no external funding.

Data Availability Statement: Data are not publicly available due to ethical reasons, though the data may be made available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

1. Introduction

Bank deposit (out) flows are clearly significant since any extreme fluctuations can disrupt aggregate consumption and aggregate investment, thus bringing about considerable adverse effects in the macroeconomic environment (Demirguç-Kunt and Detragiache, 1998; Anastasiou and Katsafados, 2020). Given that bank assets are usually illiquid assets, excessive deposit outflows can trigger banking insolvency, or even worse, a banking panic. As a consequence, this would disturb the credit flows both to households and enterprises, decreasing investment and consumption, and hence forcing even sustainable firms into bankruptcy (Demirguç-Kunt and Detragiache, 1998). Therefore, it becomes apparent that predicting bank deposit (out) flows is imperative for both policymakers and regulators.

Our work is part of a growing strand of the literature that explores how the sentiment affects bank deposits (Fecht *et al.*, 2019; Anastasiou and Katsafados, 2020; Anastasiou and Drakos, 2021a; Anastasiou and Drakos, 2021b). Our study builds on this prior work and presents a methodological framework for constructing a novel leading indicator for Greek bank deposit outflows, based on the news' headlines drawn from Bloomberg. We find that this leading indicator exerts a significant impact on bank deposit flows. Thus, it can be a valuable tool for macroprudential policymakers to forecast bank deposit outflows and hence predict potential bank runs.

2. Data and Methodology

The data for the Greek deposits refer to monthly changes in the outstanding amounts of deposits held by Greek banks, published on the website of the Bank of Greece (in millions €). Our data are in a monthly frequency, and the sample period spans from 2002 to 2018. The data for news refers to the number of times each month specific keywords appear in the economic press headlines, from over one hundred authoritative global sources, and were derived from Bloomberg.

In order to construct the leading indicator, we obtained from Bloomberg the number of specific keywords that appeared in the news headlines for the under-examination period, which we believe capture the so-called “negative sentiment”. By employing a Principal Component Analysis¹ (PCA hereafter) on these keywords, we isolated the common

¹ PCA originated with the work of Pearson (1901) and Hotelling (1933). PCA is a statistical technique used for data reduction. The leading eigenvectors from the eigen decomposition of the correlation or covariance matrix of the variables describe a series of uncorrelated linear combinations of the variables that contain most of the

component(s). The keywords we employed were: DEBT – BAILOUT – CRISIS – UNCERTAINTY – RECESSION – GREXIT – AUSTERITY – DEFAULT. All keywords were searched jointly with the word GREECE. The selection of these eight keywords was based on our identification assumption. Our prior beliefs are that since these eight news-based keywords denote the perceived crisis environment, then the common-isolated component (that is, the leading indicator) will be negatively (positively) associated with the Greek bank deposits (deposit outflows).

Table 1 shows the corresponding eigenvalue, proportion and cumulative proportion for each of the eight components. As we can depict, the first two components have the highest cumulative proportion in creating the common factor.

Table 1: PCA analysis results on the news headlines associated with the outflows of Greek deposits

Component	Eigenvalue	Difference	Proportion	Cumulative
NEWS_UNCERTAINTY	6.503	5.529	0.813	0.813
NEWS_AUSTERITY	0.974	0.781	0.121	0.934
NEWS_DEBT	0.193	0.031	0.024	0.958
NEWS_GREXIT	0.162	0.090	0.020	0.979
NEWS_DEFAULT	0.071	0.015	0.009	0.988
NEWS_RECESSION	0.056	0.028	0.007	0.995
NEWS_CRISIS	0.028	0.017	0.003	0.998
NEWS_BAILOUT	0.010	-	0.001	1.000

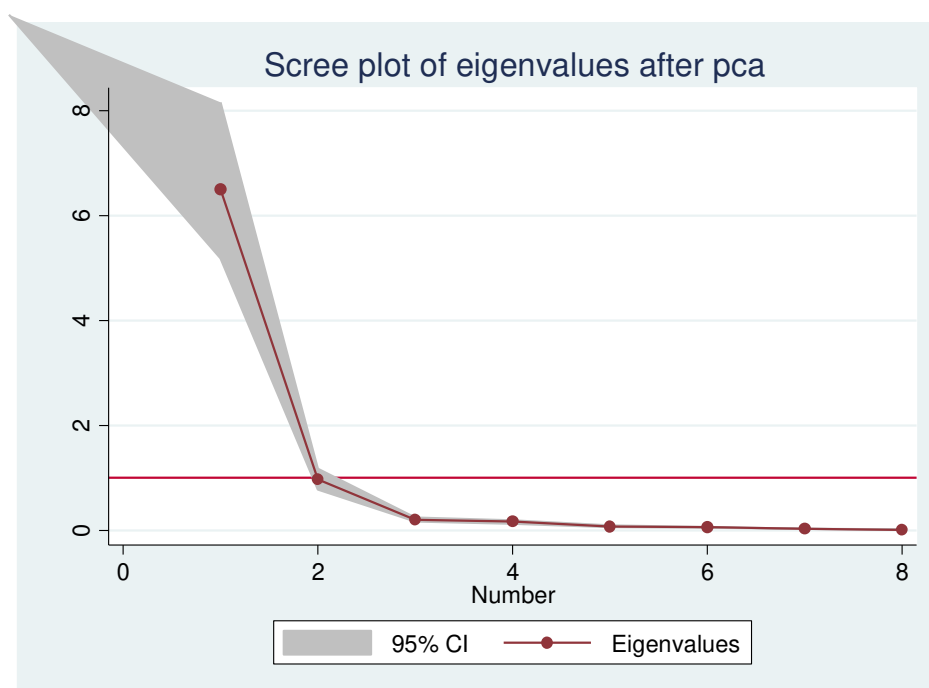
After computing the principal components, we wish to formally determine how many components to keep in order to construct the final indicator. In factor analysis, the question of the “true” number of factors is a complicated issue. With PCA, it is a little more straightforward. We may set a percentage of variance we wish to account for, say, >85%, and retain just enough components to account for at least that much of the variance. The relative magnitudes of the eigenvalues indicate the amount of variance they account for. According to Anastasiou *et al.*, (2020) an important benefit of the PCA is the fact that it produces the weights for each variable automatically, implying that the novel leading

variance. In addition to data reduction, the eigenvectors from a PCA are often inspected to learn more about the underlying structure of the data.

indicator that we constructed explains as much of the variance in the set of the different news headlines variables as possible.

A valuable tool for visualizing the eigenvalues relative to one another so that we can decide the number of components to retain in the scree plot proposed by Cattell (1966). As we can observe from Figure 1, the first two components have the highest proportion in creating the common factor. These two first components, which were found to be the main drivers of the PCA correspond to the keywords AUSTERITY and UNCERTAINTY, respectively used to construct the leading indicator (PC variable) based on the weights presented in Table 1.

Figure 1: Scree plot of eigenvalues for news headlines associated with Greek deposits outflows



After constructing our leading indicator, we examine its forecasting ability on bank deposit flows (DEPOSITS).² In particular, we first estimate a robust linear regression model (Model 1) in which we include the one month lag of the PC variable³, controlling for both autocorrelation and seasonality patterns via autoregressive (AR) and moving average (MA) terms and for some additional control variables which we believe that they also have an

² The dependent variable (DEPOSITS) which accounts for the monthly changes in the outstanding amounts of deposits, was found stationary (Dickey-Fuller test p-value=0.0155).

³ Including additional lags of the PC variable did not turn statistically significant.

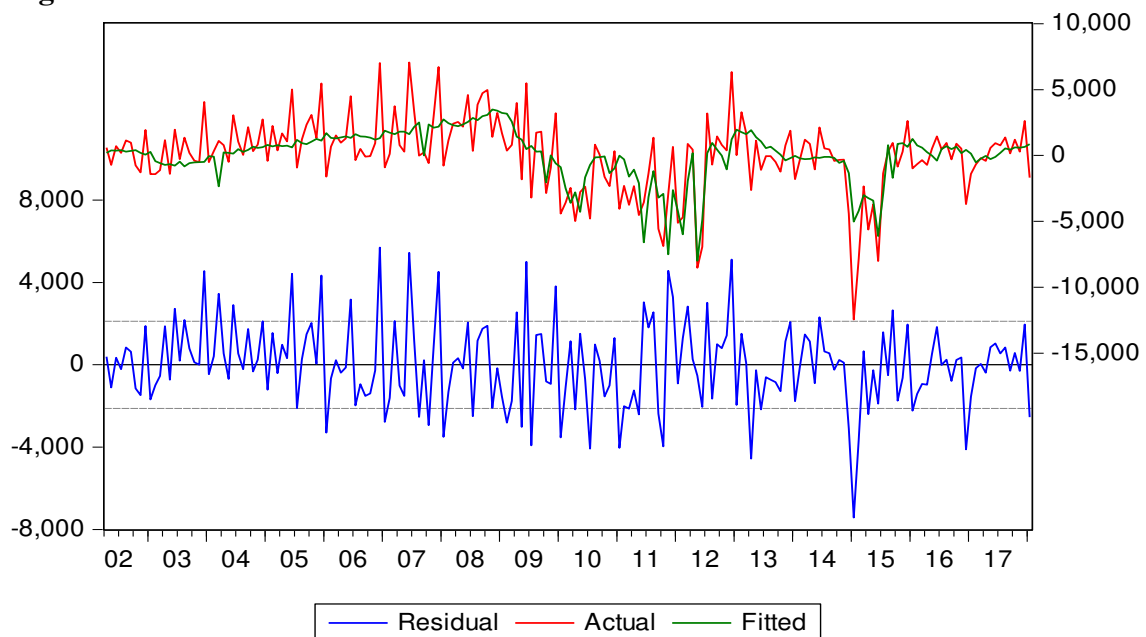
impact on bank deposit flows.⁴ Following among others Anastasiou and Drakos (2021b), the variables that we also consider as controls are defined as follows:

- INTEREST_RATE, which is the first time lag of real interest rate.
- ELECTIONS, which denotes a dummy variable which takes 1 in the years where we had elections and 0 otherwise, and
- CAPITAL_CONTROLS, which stands for a dummy variable that takes 0 (1) before (after) the imposition of the capital controls in the Greek banking system.

3. Discussion and Conclusion

According to the findings of the first model (Table 2), we found results that are compatible both with the economic theory and our priors. In particular, we found that PC is statistically significant at the 1% level and with the proper negative sign, denoting that as PC increases, more deposit outflows occur next month. In other words, as the PC increases, the higher the depositors' negative sentiment becomes, which leads them to withdraw their deposits from the banks. As far as the control variables are concerned, we found that the one-period lag of INTEREST_RATE and the CAPITAL_CONTROLS dummy variables have a positive impact on bank deposit flows. On the contrary, the ELECTIONS dummy variable was found to significantly negatively impact the Greeks' depositors' behavior. Finally, we also found evidence that the monthly outflow changes have persistence with strong autocorrelation.

Figure 2: Actual versus fitted values and residuals under Model



⁴ Note that we have tested to include additional control factors, such as spreads and stock exchange returns; nevertheless, these did not turn statistically significant and were excluded from the analysis.

The Root Mean Square Error (RMSE) of the one-step ahead forecasts under Model 1 was equal to 2079.09 and the Mean Absolute Error equal to 1594.9.⁵ As depicted in Figure 2, the actual values almost lie on the fitted values, having a common trajectory.

Table 2: Estimation results for Model 1

Dependent Variable: DEPOSITS	
Variable	Coefficient (std. error)
PC	-689.21*** (110.96)
INTEREST_RATE	143609.70*** (44718.89)
ELECTIONS	-2236.66*** (775.99)
CAPITAL_CONTROLS	4081.33*** (1162.41)
AR(1)	0.98*** (0.02)
MA(1)	-0.88*** (0.050)
Constant	-3357.35* (1727.71)
Number of Observations (Adjusted)	190
Diagnostics	
Adjusted R-squared	0.434
Hannan-Quinn criter.	18.239
Durbin-Watson stat	2.209
Probability Value (F-statistic)	0.000
Notes: (a) *, **, *** denote statistical significance at the 10, 5, and 1 percent level respectively, (b) numbers in parentheses denote robust standard errors, (c) variables INTEREST_RATE and PC are included in one period lag in the model.	

After estimating the first model, in which we observe the association between deposit flows and PC, we deem it appropriate to examine how our leading indicator influences the likelihood of going into a “high-outflows” regime. To do that, following Petropoulos *et al.*, (2018) we first employ a Markov Regime Switching Regression model on DEPOSITS with two regimes.

Table 3: Estimation results for a Markov Regime Switching Regression model with two regimes on the deposits outflows

⁵ Note that a different leading indicator tested, included performing PCA analysis directly in the lags (up to three lags) of the news headlines, as opposed to construct the indicator on basis of the same time instance, and then taking its lag (as in Model 1). This model had worse predicted ability with Root mean square error equal to 2236.5 and mean absolute error equal to 1700.6.

Dependent Variable: DEPOSITS				
Method: Switching Regression (Markov Switching)				
Included observations: 192				
Number of states: 2				
Ordinary standard errors & covariance using numeric Hessian				
Random search: 25 starting values with 10 iterations using 1 standard deviation				
Convergence achieved after 12 iterations				
Variable	Coefficient		Std. Error	z-Statistic
Regime 1				
C	896.96		180.62	4.97
Regime 2				
C	-3472.76		413.14	-8.41
Common				
LOG(SIGMA)	7.69		0.054	143.5
Transition Matrix Parameters				
P11-C	4.24	0.75	5.70	0.0000
P21-C	-2.66	0.73	-3.67	0.0002
Diagnostics				
Log likelihood			-1764.85	
S.E. of regression			2459.78	
Durbin-Watson stat			1.865	
Hannan-Quinn criter.			18.47	

In Table 3 are represented two states/regimes in the economy. In the first regime (good state) there are excess inflows (896.96 mil. € per month), while in the second regime (bad state) is represented a state of high outflows (-3472.76 mil. € per month). In Table 4, panel A are displayed the Markov transition probabilities, according to which the probability of moving from state 1 to state 2 is 0.014 while the probability of moving from state 2 to state 1 is 0.065. The probability of staying/remaining in the same regime is 0.985 and 0.934 for the first and the second regime, respectively. Furthermore, from panel B we observe that the expected duration for the first state of the economy is 70.72 months, whereas the expected duration of the bad state (increased outflows) equals 15.32 months. As depicted in Figure 3 the period of excessive outflows (state 2) spans in 2010-2012 and at the end of 2014 up to middle 2015, when the capital controls act was issued.

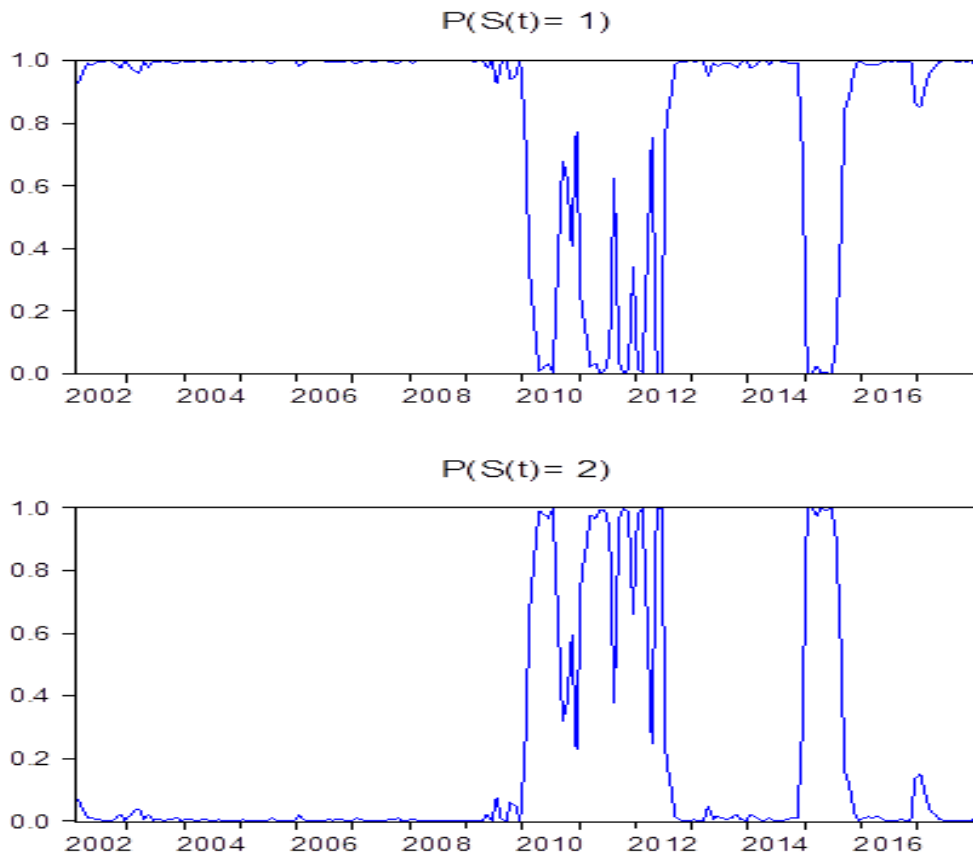
Based on the output of the Markov Switching model, we create a dummy variable (OUTFLOWS STATE) which takes 0 and 1 if we are in an inflow (state 1) or outflow (state 2) regime, respectively. Then, we estimate a Binary Logit model (model 2) by employing

the same control variables as before to investigate how the leading indicator constructed above is related to the probability of observing a regime with excessive outflows.

Table 4: Markov transition probabilities and expected duration

Panel A: Markov transition probabilities			
$P(i, k) = P(s(t) = k \mid s(t-1) = i)$ (row = i / column = j)			
Periods		1	2
	1	0.985	0.014
	2	0.065	0.934
Panel B: Expected duration			
Periods	1	70.720	
	2	15.315	

Figure 3: Filtered regime probabilities under the Markov Regime Switching Regression model with two regimes on the deposits outflows



The results from the estimation of the logit model are presented in Table 5. We observe that the one-month lag of the PC was found once again to be a significant predictor (p -value <0.0001) for observing a state of excessive outflows. Especially, we found that as the number of news increases, the greater the perceived fear becomes, we have increased probability the next month to observe deposit outflows. The real interest rate and the capital controls were found again to affect the probability of outflows. Although it was found to have the proper sign, the ELECTIONS variable was not significant in our second model.

Table 5: Estimation results for Model 2

Dependent Variable: OUTFLOWS STATE	
Variable	Coefficient (std. error)
PC	2.200*** (0.460)
INTEREST_RATE	-130.268*** (55.141)
ELECTIONS	-0.459 (0.764)
CAPITAL_CONTROLS	-2.596* (1.521)
Constant	0.305 (1.087)
Number of Observations (Adjusted)	191
Diagnostics	
McFadden R-squared	0.676
LR statistic	116.7938
Log likelihood	-27.92801
Probability Value (LR-statistic)	0.000
Notes: (a) *, **, *** denote statistical significance at the 10, 5, and 1 percent level respectively, (b) numbers in parentheses denote robust standard errors, (c) variables INTEREST_RATE and PC are included in one period lag in the model.	

A deposit outflow prediction model can be provided either from the first model, in which we tried to predict the expected deposit outflows directly, or from the second model, which we employed to predict the likelihood of going into a high outflow regime. These models can be expanded with additional covariates and be improved based on more advanced (i.e., non-linear) statistical methods while incorporating variance predictions. In this framework, the number of specific keywords appearing in news headlines seems to hold vital information for future deposit movements and can potentially serve as a leading indicator.

References

Anastasiou, D., Bragoudakis, Z., and Malandrakis, I., (2020), Non-performing loans, Governance Indicators and Systemic Liquidity Risk: Evidence from Greece, *Interdisciplinary Journal of Economics and Business Law*, 9(1), 91-114.

Anastasiou, D., and Drakos, K. (2021a). European depositors' behavior and crisis sentiment. *Journal of Economic Behavior & Organization*, 184, 117-136.

Anastasiou, D., and Drakos, K., (2021b), Nowcasting the Greek (Semi-) Deposit Run: hidden uncertainty about the future currency in a Google search?, *International Journal of Finance and Economics*, 26(1), 1133-1150.

Anastasiou, D., and Katsafados, A., (2020). Bank Deposits Flows and Textual Sentiment: When an ECB President's speech is not just a speech. MPRA Working Paper No. 99729, <https://mpra.ub.uni-muenchen.de/99729/>.

Cattell, R.B., (1966), The Scree test for the number of factors, *Multivariate Behavioral Research*, 1(2), 245-276.

Demirgüç-Kunt, A., and Detragiache, E., (1998). The Determinants of Banking Crises in Developing and Developed Countries. *International Monetary Fund Staff Papers*, 45(1), 81-109.

Fecht, F., Thum, S., and Weber, P., (2019). Fear, deposit insurance schemes, and deposit reallocation in the German banking system. *Journal of Banking and Finance*, 105, 151–165.

Hotelling, H., (1933), Analysis of a Complex of Statistical Variables into Principal Components, *Journal of Educational Psychology*, 24, 417-441.

Pearson, K., (1901), On lines and planes of closest fit to systems of points in space, *Philosophical Magazine*, 2(11), 559-572.

Petropoulos, A. , Vlachogiannakis, N.E. , Mylonas, D , (2018). Forecasting private sector bank deposits in Greece: determinants for trend and shock effects. *International Journal of Banking Accounting and Finance*, 9(2), 141–169.