Revisiting the Economic Crisis after a Decade: Statistical and Machine Learning Perspectives

Jiayan YU⋆, Jingqian ZHANG⋆⋆, Hee Eun SHIN⋆⋆⋆, Jooan KONG⋆⋆⋆⋆

A R T I C L E   I N F O

A B S T R A C T

This paper examines the 2008-09 crisis experienced by the world economy as one of the largest of its kind in four decades. Based on ample research published in the decade since, a simple statistical correlation coefficient is applied to selected variables, and a principal component analysis is conducted for significant indicators that an economic crisis sends as signals of its imminence. Recovery from the crisis is also analyzed not only for the reinstatement of original relationships but also for new signs that falling into a crisis and recovering from it may take quite different paths.

1. Introduction

Economic memory is either short or it is a nightmare to be forgotten, and this is perhaps why financial crises tend to recur over time. Identifying factors that might have led or may lead to such events is difficult to analyze or forecast (Brian, 2010). In terms of magnitude, the global crisis of 2008-09 was more than just an economic turbulence. The U.S. unemployment rate doubled from 4.6% to 9.3% during the 2007-2009 period (BLS, 2012). International trade plummeted by 12% in 2009 (WTO, 2010), and by 2011, government debt across OECD countries jumped to 100% of their combined GDP.

Research shows that the 2008 economic crisis was based on the global economy's structural problems. This suggests a need for a thorough analysis of underlying factors to ensure sustainable future growth. Adomnicăi (2019) reviewed theories and models of economic growth for economic sustainability, and Demyanyk and Van Hemert (2009) showed the 2008 crisis to generally follow the boom-and-bust cycle noted in Dell'Ariccia et al. (2008). Several studies suggested that economic crises do not indicate a free market failure but market participants' distorted incentives (Demirguc-Kunt and Serven, 2009).

Demyank and Hasan (2010) reviewed bank failure and crisis prediction models, and Celik and Karatepe (2007) considered artificial neural network models for the prediction of non-performing loans. Alam et al. (2000) found the efficacy of self-organizing neural networks and fuzzy clustering for bank failure prediction.

This paper presents an extended analysis of factors underlying the 2008 financial and economic crisis, which began with the U.S. banking sector and quickly spread across the world. Intervariable correlations are explored, and an in-depth principal component analysis reveals significant observations where the crisis sends prior signals and reveals withdrawal symptoms through various economic indicators for a perspective through which essential behaviors of a crisis can be analyzed and its recovery may be traced.

2. Literature review

Demirguc-Kunt and Detragiache (2005) review two crisis-warning methods including a signal approach and a multivariate probability model, and identifies the influence of macroeconomic shocks, structure of the banking market, broad credit-market institutions and political-economic variables as the major driving factors responsible for bank fragility. Bell and Pain (2000) review the application of popular indicator models to financial crisis prediction, stressing that the success of an empirical macro-prudential model depends largely on the accuracy of the explanatory variables so to be able to forecast economic calamities with a short frequency.

⋆Johns Hopkins University, 3400 N. Charles St, Baltimore, MD 21218, United States, ⋆⋆ University of California, Irvine, 405 Aldrich Hall, UC Irvine, CA, 92697, United States, ⋆⋆⋆ Daewon Foreign Language High School, 26 Yongmasan-ro, Gwangjin-gu, Seoul, 04939, Korea, ⋆⋆⋆⋆ Gangnam International School, Saemmaeul 2-gil 33, Seoul, Korea, 06794, Korea.
E-mail address: aryansaif908@gmail.com   (Jiayan YU – Corresponding author).

Korotayev and Tsirel (2010) explored Kondratieff waves (or K-waves) in world GDP dynamics using GDP growth rates for the 1870-2007 period and attributed Kondratieff wave patterns to price indices, interest rates, international trade, and inflation shocks from war. The oscillation period is found to be about 50 years, but this loses significance after World War II. On the other hand, Kitchin cycles (with a 40-59-month period) are found in inventory fluctuations, and Juglar (or business) cycles (with a 7-11-year period) are related to investment in fixed capital. Kuznets swings (with a 15-25-year period) are considered to be related to demographic processes such as immigrant flow and resulting construction activity and the infrastructure investment cycle.

Rios-Morales et al. (2009) identified some good governance indicators to be the key factors underlying a country's political and economic stability and identified corruption management as a key indicator. In the second level, political stability, voice and accountability seem to be the major factors discriminating political instability of countries.

Recent studies of predictive models based on artificial intelligence and statistical learning have demonstrated their wide applicability. Imandoust and Bolandraftar (2013) employed the k Nearest Neighbour algorithm and stressed the usefulness of Support Vector Machines, Artificial Neural Network, Naïve Bayesian Classifier and Genetic Algorithm in classifying and predicting financial distress in order to speculate the onset of bankruptcy. Vishwakarma (1994) used a four-layer neural network to analyze the growth rate of four U.S. business cycles indicators, viz. manufacturing and trade sales, industrial production, non-agricultural employment and personal income, for the 1965-1989 period. On the other hand Berge (2015) notes that many academically well-known economic indicators (such as interest rates, stock index, inflation, industrial production, personal income and unemployment rate) are effective for prediction only at specific time ranges, whereas many indicators have non-linear effects on forecasting business cycle turning points. The author employs equally weighted model average, BMA model, and linear and nonlinear boosting to predict recessions 0 to 24 months ahead. Giusto and Piger (2017) applied learning vector quantization, improving on the dynamic factor Markov switching model for real-time recession prediction. They have used data for growths of non-farm payroll employment, industrial production index, real personal income and real manufacturing and trade sales to identify US business cycle turning points including the 5 recessions in the last 35 years. Their method is reported to correctly identify the dates of each of the five recessions accurately, with no false positives, and that too in a shorter time-span compared to traditional statistical methods.

3. Materials and Methods

This section discusses the dataset and its source, features and missing data attributes and provides the mathematical background for the two major analyses conducted in this research.

3.1. Dataset

The data from the World Bank Global Economic Monitor (GEM, 2019) had about 40 indicators for around 200 countries for the 1990-2019 period. However, most were replications by seasonal adjustment or reduction to a "constant 2010" index. In addition, more than 50% of the data were missing overall. Given this, a slice of data in which most of the data were present and relevant to this study was taken, and this slice is described below.

We considered 9 variables as follows:
1. Core CPI, seasonally adjusted (CPI)
2. GDP at market prices, constant 2010 US$, seasonally adjusted (GDP)
3. Industrial Production, constant 2010 US$, seasonally adjusted (InP)
4. Imports Merchandise, Customs, constant 2010 US$, seasonally adjusted (Imp)
5. Exports Merchandise, Customs, constant 2010 US$, seasonally adjusted (Exp)
6. Official Exchange Rate (XRa)
7. Foreign Reserves, Months Import Cover, Goods (FRs)
8. Stock Markets, US$ (Sto)
9. Unemployment Rate, seasonally adjusted (Upl)

Seasonally adjusted values and figures with a constant base of 2010 instead of absolute values were considered wherever possible. For missing values for any variable or year, only those countries with more than 50% of values were considered, and missing values were imputed using the `cart` method in the R package `mice`.

3.2. Correlation between variables

Pairwise correlations between 9 variables were initially studied for insights into the degree of dependence or redundancy. The Pearson correlation coefficient between two variables x and y is given by

$$r_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \,, \tag{1}$$

where the summation is across all countries (denoted by subscripts i). The resulting correlograms are presented in Figure 1, and their implications are discussed in Section 4.1.

3.3. Principal Component Analysis

Data with many variables are difficult to visualize or analyze, and redundancy naturally occurs in many datasets across several variables. Addressing these issues often necessitates the reduction of dataset dimensionality, with the known limitation of losing some detailed information in the process.

Data dimensionality can be reduced by considering the first few principal components with most of the relevant information. Here principal components (PCs) are linear combinations of variables as eigen vectors of the covariance matrix.

A dataset may viewed as a collection of m observations where each is characterized by n variables $v_1, v_2, \ldots, v_n$. Here let $v_{i(k)}$ denote the value of the ith variable for the kth observation. The covariance matrix for the dataset is then

$$S = \begin{pmatrix} \mathrm{cov}(v_1, v_1) & \mathrm{cov}(v_1, v_2) & \cdots & \mathrm{cov}(v_1, v_n) \\ \mathrm{cov}(v_2, v_1) & \mathrm{cov}(v_2, v_2) & \cdots & \mathrm{cov}(v_2, v_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{cov}(v_n, v_1) & \mathrm{cov}(v_n, v_2) & \cdots & \mathrm{cov}(v_n, v_n) \end{pmatrix} \tag{2}$$

where $\mathrm{cov}(v_i, v_j)$ is the covariance of two variables expressed as

$$\mathrm{cov}(v_i, v_j) = \frac{1}{n} \sum_{k=1}^{n} \left(v_{i(k)} - \overline{v_i} \right) \left(v_{j(k)} - \overline{v_j} \right) \tag{3}$$

The first principal component (PC1) can be calculated as the eigen vector corresponding to the largest eigen value of S, the second PC (PC2), as the eigen vector corresponding to the second largest eigen value, and so on. The amount of variance explained by each PC is proportional to the magnitude of the corresponding eigen value (Shlens, 2014).

Loading plots were used to visualize variable interrelation. The 9 variables appeared as different points in the loading plot, where their respective locations reveal insights into their relationships. Variables that are close together or in the same line of PC1-PC2 plots were considered to have the same type of effect. For example, to identify GDP factors, those variables located along the GDP vector (blue arrow from the origin in Figure 2) need to be examined. Here a quantitative measure of variable dependence on GDP is the *projection* of that variable (considered as another vector) on the GDP vector. Therefore, variables perpendicular to the indicator vector are considered to have no effect on the indicator variable. In addition, those on the side of the indicator vector are considered to have positive effects, whereas those on the opposite side are seen to have negative effects.

4. Results and Discussion

This section discusses the main results from two statistical and machine learning techniques. Plots are first analyzed from a mathematical perspective for connotative insights, and then their econometric interpretations are discussed.

3.1. Analyzing inter-variable correlations

Pairwise correlations between each variable pair in section 3.1 were analyzed for every three-year interval from 2000 to 2018. However, since the focus of this study is the 2008-09 crisis, 2008 and 2010 are additionally included for analysis. Figures 1 (a) – (i) show the resulting correlograms.
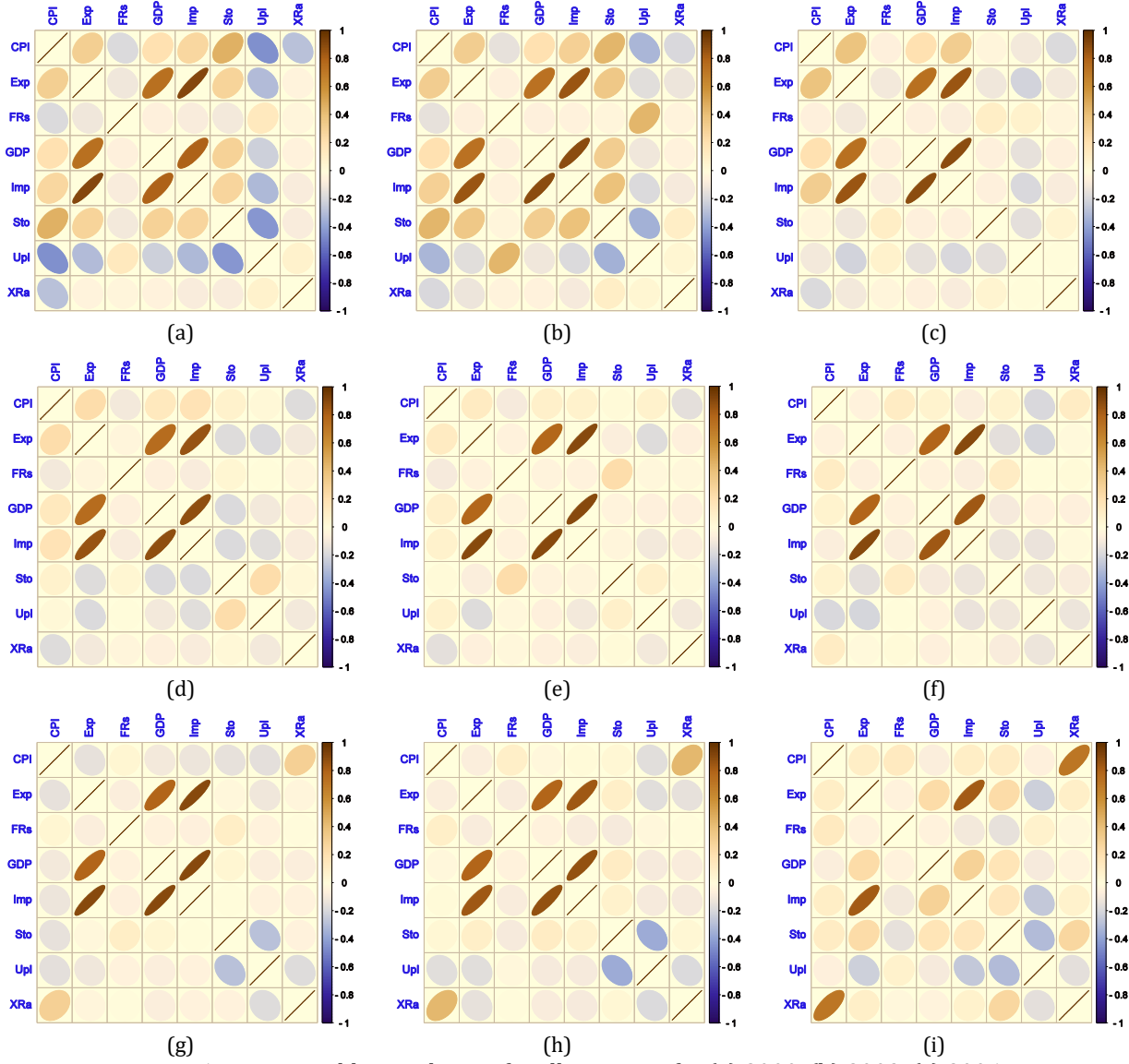
Figure 1. Inter-variable correlations for all countries for (a) 2000, (b) 2003, (c) 2006;
(d) 2008, (e) 2009, (f) 2010, (g) 2012, (h) 2015, and (i) 2018

The top row (Figures 1 (a), (b) and (c)) indicates the pre-crisis period, and the bottom row (Figures 1 (g), (h) and (i)) shows the post-crisis period, both with a three-year period. The middle row (Figures 1 (d), (e) and (f)) contrastingly show correlations for three consecutive years of the crisis period (2007-2009). One noteworthy observation in inter-variable correlations is that in the pre-crisis period, most inter-variable correlations (except GDP, Export and Import) gradually fade away, almost vanishing in 2006. This can be interpreted as the world economy going out of control, where a controlled situation can be viewed as a good inter-relation between well-behaved features. This hypothesis is reinforced by the post-crisis correlograms, where the inter-correlations again show signs of building up, and in 2018, most features including unemployment rate (Upl) and stock markets (Sto) regain strong relationships with other variables such as import (Imp), export (Exp) or exchange rate (XRa), although not to the 2000-03 level.

Another characteristic in the post-crisis period typically not present in the pre-crisis years is the strengthening positive correlation between the consumer price index (CPI) and the official exchange rate (XRa), which peaks in 2018 (Figure 1 (i)). Recovery from a crisis is of course qualitatively different from falling into it, but this is a striking feature in the post-crisis period, which may indicate that the pass-through from exchange rate movements to consumer prices primarily through exports takes place more in this period (Ulvedal and Vonen, 2016) when stronger economies more intensely affected by the crisis tend to pass their burdens on weaker economies.

4.2. Exploring results from principal component analysis

PCA loadings for 2000-2018 are plotted in Figures 2 (a) – (i). The two vectors of GDP and unemployment rate (Upl) are in blue and red, respectively, and epitomize the manifestation and effect of the crisis.
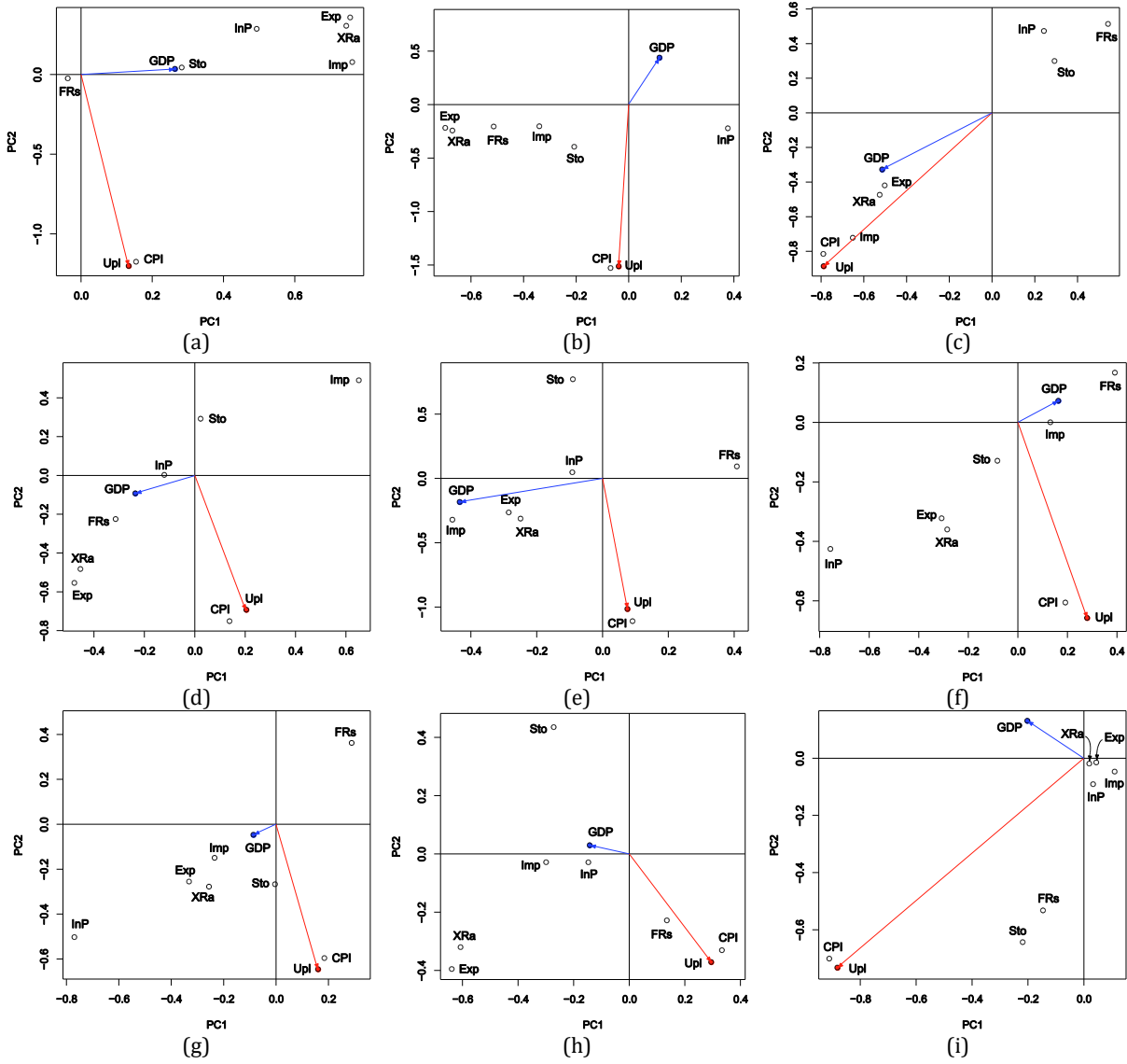
Figure 3. The PCA loadings of the nine variables for the years (a) 2000, (b) 2003, (c) 2006; (d) 2008, (e) 2009, (f) 2010, (g) 2012, (h) 2015, and (i) 2018

Principal component (i.e., PC1 vs. PC2) plots reveal some important insights. First, the CPI and unemployment rate (Upl) are together in almost every year, as in the case of the official exchange rate (XRa) and exports (Exp). This appears a *general feature* of any country in all periods and is not particular to either the pre-crisis or post-crisis period. They may be interpreted as dual indicators of the internal market and foreign trade dynamics of a country. Second, exports and imports (Exp, Imp) appear to be close together (or at least on the same side of the plot from the origin) in the pre-crisis period (this is particularly notable in Figures 2 (a), (b) and (c)), whereas they first take diametrically opposite positions in two opposite quadrants in 2007 (Figure 2 (d)). This indicates that economies are *in turmoil* or *out of balance*. Antagonistic positions are also seen in 2009 (Figure 2 (f)) but regain closeness from 2012 (Figure 2 (g), (h) and (i)). The crisis-impending year 2008 is not consistent with this hypothesis, plausibly from the effect of the tumultuous nature of economic twists and turns experienced by most countries in that period, when any phenomena may be overshadowed by other more imminent current events, masking intended behaviors.

Another noteworthy characteristic of the unemployment rate (Upl) and consumer price index (CPI) is that they are placed in a lone region of almost all PCA loading plots except in 2006. The distinct nature of these two indicators is likely due to their more *demographic nature*. Upl and CPI reflect the balance of the labor and consumer goods with the financial market, but the relative distance of economic-demographic features from remaining factors falls off prior to the crisis. It may be premature to infer this phenomenon as an indicator of a global crisis, but their exceptional behaviors are clearly noteworthy. Mügge (2015) notes that CPI and unemployment are among the most important indicators of economic health. In addition, foreign reserves (FRs) and stock markets (Sto) appear to be a buffer for economies recovering from the crisis, and this is evident from three plots in the bottom row (Figures 2 (g), (h) and particularly in 2 (i)), where they

occupy an intermediate position between the export-import-GDP-exchange rate cluster and the CPI-unemployment rate pair. Frankel and Saravelos (2012) suggest that a fall in reserves and movements in the exchange rate are two leading indicators of crises, using a drop in GDP or industrial production, currency depreciation, participation in IMF programs and stock market performance to analyze the 2008-09 crisis.

5. Conclusions

The 2008-09 economic and financial crisis is considered one of the largest of its kind in modern history. Several studies have examined the factors underlying such crises using econometric theory as well as statistical and machine learning methodologies. This paper employs a statistical pair-wise correlation coefficient for an overview of indicators of the crisis and its impending and recovery periods and uses a more sophisticated principal component analysis for deeper insights into interrelationships between major economic indicators of an approaching crisis and into cues for its recovery. According to the results, GDP remains closely related to a country's export and import, and these three factors significantly lose inter-relationships with other variables including exchange rate and stock market or unemployment rate a couple of years before the crisis but gradually regain the inter-relationship a few years after the crisis, when markets begin recovering. PCA loading plots indicate that the exchange rate has the same kind of effect as exports, inflation and unemployment. However, exports and imports in the same quadrant of the PCA plot indicate a healthy economy and move to opposite quadrants just before or after a crisis, when economies undergo a fundamental breakdown or recovery. Here foreign reserves and stock markets may have a buffering effect on the impact of a crisis and facilitate economies' recovery to their pre-crisis states.

References

1. Adomnicăi, V. I. (2019). The Sustainability of economic growth based on consumption. The case of Romania. Ecoforum Journal, 8(1).
2. Alam, P., Booth, D., Lee, K. and Thordarson, T. (2000), The use of fuzzy clustering algorithm and self-organizing neural networks for identifying potentially failing banks: an experimental study. Expert Systems with Applications, 18, pp. 185-199.
3. Bell, J., & Pain, D. (2000), Leading indicator models of banking crises–a critical review. Financial Stability Review, 9, 113-129.
4. Berge, T. J. (2015), Predicting recessions with leading indicators: Model averaging and selection over the business cycle. Journal of Forecasting, 34(6), 455-471.
5. BLS, U.S. Bureau of Labor Statistics (2012), The Recession of 2007–2009.
6. Brian, K., & Patrick, L. (2010). OECD insights from crisis to recovery the causes, course and consequences of the great recession: The causes, course and consequences of the great recession. OECD Publishing.
7. Celik, A. E. and Karatepe, Y. (2007) Evaluating and forecasting banking crises through neural network models: An application for Turkish banking sector. Expert Systems with Applications, 33, pp. 809-815.
8. Dell'Ariccia, G., Igan, D. and Laeven, L. (2008), Credit booms and lending standards: Evidence from the subprime mortgage market. Report No. 2008-2106, International Monetary Fund.
9. Demirgüç-Kunt, A., & Detragiache, E. (2005). Cross-country empirical studies of systemic bank distress: a survey. National Institute Economic Review, 192(1), 68-83.
10. Demirguc-Kunt, A. and Serven, L. (2009), Are all the sacred cows dead? Implications of the financial crisis for macro and financial policies. Working paper 4807, World Bank
11. Demyanyk, Y. and Hasan, I. (2010), Financial crises and bank failures: A review of prediction methods. Omega, 38(5), pp.315-324.
12. Demyanyk, Y. and Van Hemert, O. (2009), Understanding the subprime mortgage crisis. The Review of Financial Studies, 24(6), pp.1848-1880.
13. Frankel, J., & Saravelos, G. (2012), Can leading indicators assess country vulnerability? Evidence from the 2008–09 global financial crisis. Journal of International Economics, 87(2), 216-231.
14. Giusto, A., & Piger, J. (2017), Identifying business cycle turning points in real time with vector quantization. International Journal of Forecasting, 33(1), 174-184.
15. Imandoust, S. B., & Bolandraftar, M. (2013), Application of k-nearest neighbor (knn) approach for predicting economic events: Theoretical background. International Journal of Engineering Research and Applications, 3(5), 605-610.
16. Korotayev, A. V., and Tsirel, S. V. (2010). A spectral analysis of world GDP dynamics: Kondratieff waves, Kuznets swings, Juglar and Kitchin cycles in global economic development, and the 2008–2009 economic crisis. Structure and Dynamics, 4(1).
17. Mügge, D. (2016), Studying macroeconomic indicators as powerful ideas. Journal of European Public Policy, 23(3), 410-427.
18. Pål Bergset Ulvedal and Nikka Husom Vonen (2016), Pass-through from exchange rate movements to consumer prices. Monetary Policy, Norges Bank Staff Memo Report No. 3, ISSN: 1504-2596.
19. Rios-Morales, R., Gamberger, D., Šmuc, T., & Azuaje, F. (2009). Innovative methods in assessing political risk for business internationalization. Research in International Business and Finance, 23(2), 144-156.
20. Shlens, J. (2014), A tutorial on principal component analysis. arXiv preprint arXiv:1404.1100.
21. Ulvedal, P. B., & Vonen, N. H. (2016). Pass-through from Exchange Rate Movements to Consumer Prices. Norwegian Bank Report.
22. Vishwakarma, K. P. (1994), Recognizing business cycle turning points by means of a neural network. Computational Economics, 7(3), 175-185.
23. WB, The World Bank (2019), Global Economic Monitor Databank. https://databank.worldbank.org/data/reports.aspx?source=global-economic-monitor Accessed on: 14th February, 2019.
24. WTO, World Trade Organization (2010), The trade situation in 2009-10, World Trade Report 2010.