

Composite Leading Indicator for GVA-Manufacturing for India

by Anirban Sanyal, Shivangee Misra and Sanjay Singh[^]

This paper develops a quarterly Composite Leading Indicator (CLI) for GVA-Manufacturing using a two-stage procedure that combines systematic variable selection with subsequent aggregation. The indicator set—comprising commodity prices, survey-based expectations, industrial credit flows, and global variables—is identified through multiple validation techniques and then incorporated into machine-learning models, notably Random Forest and XGBoost. The resulting CLI exhibits a stronger leading property, yielding a cross-correlation of 0.86 at a one-quarter lead, compared with 0.72 contemporaneously. Its turning points consistently precede those of manufacturing GVA by one quarter, highlighting its usefulness for short-term monitoring and forecasting.

Introduction

Business cycle leading indicators are a vital component of macroeconomic surveillance, offering timely insights into emerging shifts in economic momentum. In particular, the leading business cycle indicators enable the identification of prospective turning points in the business cycle of the reference series, thereby providing early signals on evolving economic conditions. Against this backdrop, this article introduces a new leading indicator designed to track the real gross value added (GVA) growth of the manufacturing sector in India.

Business cycle analysis has a long intellectual lineage, tracing its roots to the seminal contributions of Burns and Mitchell (1946) and the later empirical

refinements of Stock and Watson (1989). Subsequent research by Moore (1982) and Zarnowitz & Boschan (1975) underscored the importance of cyclical synchronisation across economic indicators, setting the stage for systematic development of leading business cycle indicators across advanced and emerging economies. International experience highlights considerable heterogeneity: in Italy, monetary and financial variables have been found to lead domestic cycles by 12–16 months, with international cycles exhibiting a high degree of co-movement (Altissimo et al., 2000); for Turkey, a leading indicator index was constructed from nine key economic series spanning imports, monetary aggregates, and fiscal expenditures (Murutoglu, 1999). Many such indicators draw on business and consumer survey data, with evidence—such as Finland's industry survey—demonstrating strong correlations between forward-looking expectations and subsequent industrial production (Penna Urrila, 2001). Collectively, composite leading indicators have proven useful in anticipating turning points in reference series, thereby strengthening short-term forecasting and policy assessment (Altissimo et al., 2000; Murutoglu, 1999).

In the Indian context, Roy and Biswas (2012) developed a composite leading indicator (CLI) for the Index of Industrial Production (IIP), employing both growth-cycle and growth-rate-cycle approaches to track turning points in overall industrial activity. That indicator, constructed for the 2004–05 base, served as a timely gauge of cyclical dynamics of industrial growth in India at the time. Since then, however, the role of IIP in national accounts has diminished, and the index itself has undergone a base revision to 2011–12. Given the centrality of the manufacturing sector in gross value added (GVA) and the need for more robust high-frequency cyclical assessment, this article proposes a composite leading indicator for GVA-manufacturing at a quarterly frequency.

[^] The authors are from the Department of Statistics and Information Management. The views expressed in the article are of the authors and do not represent the views of the Reserve Bank of India..

The construction of the CLI for GVA-manufacturing in this study follows a structured two-step framework, supported by a range of cross-validation techniques. In the first stage, potential leading variables are identified on the basis of their signal strength vis-à-vis the reference series, complemented by a machine-learning-based kitchen-sink approach to further refine the selection. The cyclical characteristics of the shortlisted indicators are subsequently examined through wavelet analysis to ensure robustness of their leading properties. In the second stage, these selected variables are combined using multiple aggregation methods to derive the composite indicator. The resulting CLI encompasses indicators reflecting cost pressures, external demand, policy uncertainty, and credit flows to industry, and the proposed CLI exhibits a lead of one quarter over the growth rate cycle of GVA manufacturing.

Rest of the article is organised as follows – Section II provides background and current practice of CLI. The empirical framework is noted in Section III. Section IV documents the data sources and frequency. Empirical findings are listed in Section V. Within empirical findings, the variables selected through various approaches are mentioned, followed by the findings of the wavelet analysis. Later, CLI is constructed using the selected variables through various aggregation approaches. The validation of the leading property of each CLI is assessed through turning point analysis. Lastly, the findings are summarised in the concluding remarks in Section VI.

II. Literature Review and Current Practice of Composite Leading Indicator

The construction of the leading indicator is guided by the requirement that it attain its turning points—both peaks and troughs—ahead of the coincident index, which captures contemporaneous economic conditions. This property underpins its usefulness in forecasting near-term fluctuations and

supports forward-looking decision-making. As such, the leading indicator provides policymakers, financial analysts, investors, and firms with a systematic means of anticipating shifts in macroeconomic momentum. When used in conjunction with the coincident index, the leading economic indicator enhances the monitoring architecture for the Indian economy and delivers early warning signals of prospective expansions or contractions in activity (Dua and Banerji, 1999).

The analytical foundations of this approach trace back to Mintz's (1969, 1972 and 1974) development of the growth-cycle methodology for identifying cyclical turning points. Earlier work by Burns and Mitchell (1967) established the empirical significance of coincident and leading indicators for business-cycle analysis. Building on these contributions, Klein and Moore (1985) advanced the study of economic cycles at the National Bureau of Economic Research, setting the precedent for modern indicator-based monitoring systems.

Among the existing studies for India, Chitre (1982) documents substantial synchronicity among a wide range of Indian macroeconomic indicators around their long-run trends using a growth-cycle framework. His analysis spans non-agricultural net national product, industrial output, capital formation, monetary aggregates, and bank credit, among others, from which fifteen variables were ultimately selected to construct a composite index intended to proxy aggregate economic activity. The study identifies five distinct growth cycles for the Indian economy between 1951 and 1975. Using annual data from 1950 to 1985, Hatekar (1993) similarly identifies turning points in major macroeconomic aggregates and examines their comovement patterns within a growth-cycle perspective. Dua and Banerji (1999), applying the traditional NBER methodology, derived both classical business cycles and growth-rate

cycles for India and develop a composite leading index drawing on indicators from the monetary, construction, and corporate sectors. Expanding the empirical base, Chitre (2001) analyzes 94 monthly series for 1951–1982, and derived a reference cycle from eleven indicators using diffusion indexes, composite indexes, and principal components methods.

This paper is closely related to Roy and Biswas (2012) which proposes a CLI for the Index of Industrial Production using eight high-frequency indicators (HFIs) (Table 1). To account for differences in scale across the HFIs and the IIP, the authors apply a cumulative density function transformation prior to aggregation. Lead–lag relationships are assessed through cross-correlation analysis between each candidate series and the reference series. Following indicator selection, the target series (IIP) is regressed on lagged values of the individual indicators; the resulting adjusted R^2 values—interpreted as the share of variation in the target explained by each indicator and its lags—serve as a metric of leading performance. These adjusted R^2 values are subsequently employed as weights in the construction of the composite index.

III. Empirical Framework

The CLI for GVA-Manufacturing, proposed in this article has been constructed in two phases. In the first phase, the HFIs are selected using various variable selection methods. The selected indicators are

aggregated to derive CLI in the phase 2. The variable selection is carried out through signal extractions where the leading property of each variable is tested individually and in a collective manner. The signal strength of each HFI is validated using cross correlation analysis, regression estimates using ordinary least squares (OLS), quantile regression, mutual information criteria and dynamic time warping (DTW). Using the kitchen sink approach the information on the leading properties of multiple indicators is assessed through recursive feature eliminations (RFE) with k-nearest neighbour (k-NN), random forest and XGBoost approach. Further, different sets of HFIs are generated from the common variables selected by various methods. Lastly, HFIs having high wavelet coherence are selected as additional group of the selected variables. Additionally, indicators appearing in at least one criteria are taken together as separate sets of variables through various combinations to improve the information contents. A brief discussion of the methods used for variable selection is provided in Annex I.

The aggregation of the selected indicators is conducted through various methods, namely, simple average, weighted average with various choices of weights and dynamic factor models (DFMs). The weighted average approach is developed using inverse standard deviations and correlation as weights. DFM is employed to extract the common signal strength from the selected variables. Lastly, the CLI is smoothened using HP filter to remove the irregular variations in the data. The lambda value in the HP filter ($\lambda=4$) is selected based on the cross-validation technique used by Grehmann and Yetman (2018). Lastly, the performance of the proposed CLI is carried out using cross-correlation, coherence and turn-around point analysis proposed by Bry and Boschan (1971) and later, modified for quarterly data by Harding & Pagan (2012).

Table 1: Indicators Used in CLI for Industry

Sr. No.	Indicator	Weight
1	Commercial Motor Vehicle Production	11.4
2	Dollar/Rupee Exchange Rate (Monthly Average)	13.3
3	Monetary Aggregate M1	19.7
4	Non-Oil Imports	10.8
5	Railway Freight	6.3
6	BSE SENSEX Index	15.8
7	Steel Production	18.0
8	CP Spread	4.8

Source: Roy and Biswas (2012).

IV. Data Used

The high frequency economic indicators which are likely to influence the economic activities with a lead period, are selected from major dimensions, namely, i) Domestic demand condition; ii) Domestic industrial production; iii) Domestic price conditions; iv) Foreign trade; v) Employment condition; (vi) Trade, transport and other services indicators; (vii) Public finance and payment Indicators; (viii) Exchange rate; (ix) Global commodity price; (x) Policy uncertainty; (xi) Forward looking survey – Industrial Outlook Survey and PMI Manufacturing; (xiii) Cost of borrowing proxy; and (xiv) Global economic indicators. The variables are transformed into year-on-year (y-o-y) growth except for the borrowing cost (i.e. interest rate) proxy, policy uncertainty and exchange rate. The interest rate proxies, and policy uncertainties are used in level values. Exchange rates are transformed into quarter-on-quarter (q-o-q) annualized growth rate. The detailed list of the HFIs considered within each segment is provided in Annex II.

The data used for this analysis spans from April 2013 to December 2024. As the reference series (*i.e.*, GVA-manufacturing) is available at quarterly interval, the variable selection is carried out at quarterly frequency and the CLI is also calculated at quarterly frequency.

V. Empirical Findings

V.1 Variable Selection

The correlation analysis of the HFIs show very high negative correlation of global commodity prices with GVA manufacturing. Within commodity prices, crude oil prices affect the manufacturing growth with lag of 1 – 2 quarters. IMF all commodity prices (excluding gold) also drags manufacturing growth with a lag of 1 - 2 quarters. Merchandise imports moderate manufacturing growth in 1-2 quarter lag, while non-oil exports improve growth in the manufacturing

Table 2: Cross Correlation Estimates of GVA-Manufacturing Growth with Top 10 HFIs

Variable	Pearson	Kendall	Spearman
IMF Crude Oil Price (-2)	-0.48	-0.40	-0.53
IMF industrial Input (-2)	-0.50	-0.39	-0.56
IMF All Commodity Price (-2)	-0.48	-0.38	-0.53
IMF All (Excl. Gold) Commodity Price (-2)	-0.48	-0.38	-0.54
Non-oil exports (-1)	0.49	0.38	0.54
Real Credit to Industry (-2)	0.47	0.37	0.51
IMF Metal Prices (-2)	-0.41	-0.37	-0.49
Merchandise Imports (-2)	-0.44	-0.35	-0.52
US Non-farm Payroll Employment SA (-1)	0.34	0.43	0.57
WPI Industrial Raw Material (-2)	-0.42	-0.33	-0.45

Note: The number indicated in the parentheses indicate the lags of the variables, measured in quarters.

Source: Authors' calculation.

sector with a lag of one quarter. Among the global variables, US non-farm payroll employment has positive correlation with GVA manufacturing with one quarter lead (Table 2).

The variable selection using regression estimates is carried out using OLS regression and quantile regression (for median). The regression includes the lagged value of GVA - manufacturing as additional regressor to knock out any time persistent effects in the data generating process of GVA - manufacturing. The regression coefficient attached to the HFI is extracted if the coefficient estimate is statistically significant at 10 per cent level of significance.¹

The regression estimates show similar effects of commodity prices on GVA – manufacturing growth. Railway freight and petroleum consumption appear to have significant positive relation with GVA - manufacturing. Non-food credit and real credit to industry also improves the manufacturing growth with a lead of 2 quarters. The borrowing cost proxy, namely, G-Sec 10-year yield and treasury bill rate increases the borrowing burden and

¹ The standard errors of the quantile regression are asymptotic standard error.

Table 3: Regression Coefficients of GVA - Manufacturing with HFIs

Variable	OLS Regression	Quantile Regression (Median)
Indian Basket Crude Oil Price (-1)	-1.40	-0.54
WPI Manufacturing (-1)	-1.03	-0.45
Railway Freight Traffic (-2)	1.03	0.31
Petroleum Consumption (-1)	0.86	0.26
Real Credit to Industry (-2)	0.67	0.23
Real Non-food Credit (-2)	0.62	0.22
G-Sec 10 Yrs Yield (-2)	-0.98	-1.10
T-Bill 91 Days Rate (-1)	-1.05	-0.94
WPI Headline (-1)	0.64	-0.29
Merchandise Imports (-2)	-0.43	-0.48

Note: The number indicated in the parentheses indicate the lags of the variables, measured in quarters.

Source: Authors' calculation.

thereby, moderates the manufacturing growth. Merchandise imports also moderate the manufacturing growth (Table 3).

Lastly, the variable selection using mutual information, cosine similarity and (L1 and L2) distance measure² filters capacity utilization, outlook of raw material cost, PMI manufacturing index and its components, trade - transport indicators, borrowing cost proxy and commodity prices (Table 4).

The variables selected from the signal strength of the individual HFIs ignores the interactions among HFIs. For that, all variables are used simultaneously in a single framework (with different lags) to identify the suitable variables. This kitchen-sink approach uses three broad methods to identify the important variables—recursive feature elimination (RFE), random forest (RF) and XGBoost. Apart from the commodity prices, global indicators and policy uncertainties are selected as leading variables of GVA - manufacturing growth. Real non-food credit and credit to industry are selected in RFE and Random Forest. The domestic

² L1 (or Manhattan) distance is the absolute deviation between two vectors whereas L2 (or Mahalanobis) distance is the square root of the sum of square deviation between two vectors.

Table 4: Variables Selected in Other Criteria

Variable	Cosine	DTW	MI
Backlogs of Work(-2)	1	0	0
IOS Cost of Raw Material Expectation (-1)	0	1	0
IOS Cost of Raw Material Expectation (-2)	0	1	0
Indian Basket Crude Oil Price (-2)	0	0	1
CU (-1)	1	1	0
CU (-2)	0	1	0
PMI Employment (-1)	0	0	1
PMI Employment(-2)	1	0	0
PMI Future Output_(-1)	0	1	0
PMI Future Output(-2)	1	1	0
G-sec 10Yrs Yield (-1)	1	0	0
G-sec 10Yrs Yield (-2)	1	0	0
IMF Industrial Input (-1)	0	0	1
IMF Industrial Input (-2)	0	0	1
PMI Input Prices (-1)	0	1	0
PMI Input Prices (-2)	0	1	0
PMI New Export Orders (-2)	1	1	0
PMI New Orders (-1)	1	1	0
PMI New Orders (-2)	0	1	0
PMI Output Prices (-2)	1	0	0
PMI Output (-1)	1	1	0
PMI Output (-2)	0	1	0
Petroleum Consumption (-1)	0	0	1
PMI Index (-1)	1	1	0
PMI Index (-2)	1	1	0
Railway Freight Traffic (-2)	0	0	1
Real Non-food Credit (-1)	0	0	1
Real Non-food Credit(-2)	0	0	1
UPI Payments (-1)	0	1	0
UPI Payments (-2)	0	1	0
WPI Headline (-1)	0	0	1
WPI Headline (-2)	0	0	1
WPI Manufactured Products (-1)	0	0	1
WPI Manufactured Products (-2)	0	0	1

Note: 1. '0' indicates not selected and 1 stands for selected.

2. The number indicated in the parentheses indicate the lags of the variables.

Source: Authors' calculations.

economic indicators and borrowing cost proxies are also filtered (Table 5).

V.2 Cross Validation of the Business Cycle Properties of Selected HFI

Combining the variables selected through various methods provides a comprehensive list of HFIs having some leading information about GVA-manufacturing growth. The information content of

Table 5: Variables Selected in RFE, RF and XGBoost

Variables	Lead in Quarters
Method = RFE	
US Non-farm Payroll Employment (SA)	1
International Air Passenger Traffic	2
European EPU	2
Real Non-food Credit	1
IOS Cost of Raw Material (Expectation)	1
T-Bill 91 Days Yield	2
IMF All (Excl. Gold) Commodity Prices	1
Method = Random Forest	
Real Credit to Industry	1
Global EPU	1
Cement Production	1
G-Sec 10 Years Yield	2
USA EPU	1
WPI Headline	2
Commercial Motor Vehicle Sales	1
India EPU	1
Method = XGBoost	
International Cargo Traffic	1
Global EPU	1
WCMR	1
Exports to Emerging and Developing Asia	2
Domestic Air Passenger Traffic	1
WPI Manufacturing	1
High Speed Diesel	1
US Non-farm Payroll Data	1
IIP Consumer Durables	1
IIP Consumer Non-durables	1

Source: Authors' calculations

the selected HFIs is verified using wavelet analysis. The cross-wavelet analysis provides the degree of coherence between the HFIs and the benchmark series (*i.e.*, GVA-manufacturing).³

The wavelet coherence estimates shows that GVA manufacturing growth shows high coherence with eight core industries, merchandise exports, IIP use-based classifications, employment from US Non-

³ Here, the leading property is not directly sought from the cross-spectrum due to limitation of the length of time series data. The data used for this analysis, spans from Q1: 2013-14 till Q3: 2024-25 which includes disruptions due to COVID pandemic. The lack of business cycle coverage in the selected data, may lead to overfitting of wavelets and may lead to biased estimate of phase difference. Hence, the coherence measure is used in this context.

Table 6: Coherence from Cross Wavelet Analysis

Series	Coherence Period (in quarters)	Coherence
Eight Core: Overall	2	0.78
Merchandise Export	2	0.75
IIP Intermediate goods	2	0.73
IIP Primary goods	2	0.73
US Non-farm Payroll (SA)	2	0.68
IIP Infrastructure/ construction goods	2	0.64
Indian Basket Crude Oil Price	2	0.59
Cement Production	2	0.59
WPI IRM	2	0.59
US IIP	2	0.59

Source: Authors' calculations

farm payroll, Indian basket crude oil price, cement production and US IIP (Table 6).

The wavelet coherence plot also confirms the leading property of these indicators with GVA-manufacturing in 1-2 quarter lead. This common set of indicators having high coherence, are also selected as separate set of variables (along with various other subsets of variables from the previous selections) in the variable selection set (Annex III). A short description of interpretation of the wavelet charts is provided in Annex IV.

a. Composite Leading Indicator

The CLI is constructed using the selected HFIs from different methods. For each selection of HFI, CLI is constructed using simple average, weighted average and DFM. Following this approach, 48 different CLI are constructed and the performance of each CLI is validated using cross-correlation and turnaround point analysis for quarterly data. The cross-correlation is checked with lead of 1 quarter and 2 quarters. The cross-correlation results shows that variables selected from Random Forest - XGBoost and combined using inverse standard deviation weights, provide the highest tracking at 1-2 quarter lead. CLI from random forest also provide a better tracking than others (Table 7). The detailed list of cross-correlation estimates is provided in Annex V.

Table 7: Cross Correlation of Selected CLIs with GVA Manufacturing

Variable Selection	CLI Construction	Lag = 1		Lag = 2	
		Full Sample	Excluding COVID	Full Sample	Excluding COVID
RF-XG Boost	Simple Average	0.39	0.29	0.16	0.37
RF-XG Boost	Weighted Average - SD	0.86	0.85	0.59	0.71
RF-XG Boost	Weighted Average - Correlation	0.40	0.23	0.17	0.31
RF-XG Boost	Dynamic Factor Model (2) factors)	0.27	0.39	0.00	0.42
RF-XG Boost	Dynamic Factor Model (1) factor)	0.46	0.05	0.16	0.09
RF	Simple Average	0.62	0.64	0.34	0.61
RF	Weighted Average - SD	0.74	0.78	0.61	0.73
RF	Weighted Average - Correlation	0.59	0.66	0.28	0.61
RF	Dynamic Factor Model (2) factors)	0.07	0.34	0.29	0.32
RF	Dynamic Factor Model (1) factor)	0.07	0.51	0.26	0.50
RF-Spearman	Simple Average	0.64	0.77	0.54	0.74
RF-Spearman	Weighted Average - SD	0.56	0.73	0.64	0.68
RF-Spearman	Weighted Average - Correlation	0.58	0.76	0.50	0.75
RF-Spearman	Dynamic Factor Model (2) factors)	0.06	0.52	0.24	0.50
RF-Spearman	Dynamic Factor Model (1) factor)	0.05	0.48	0.28	0.49
RF-Quantile	Simple Average	0.66	0.69	0.39	0.65
RF-Quantile	Weighted Average - SD	0.66	0.75	0.69	0.67
RF-Quantile	Weighted Average - Correlation	0.70	0.72	0.44	0.69
RF-Quantile	Dynamic Factor Model (2) factors)	0.07	0.32	0.24	0.24
RF-Quantile	Dynamic Factor Model (1) factor)	0.08	0.54	0.23	0.47
Coherence	Simple Average	0.46	0.68	0.51	0.65
Coherence	Weighted Average - SD	0.49	0.68	0.55	0.61
Coherence	Weighted Average - Correlation	0.39	0.66	0.48	0.63
Coherence	Dynamic Factor Model (2) factors)	0.12	0.51	0.32	0.54
Coherence	Dynamic Factor Model (1) factor)	0.08	0.47	0.32	0.50

Note: • Highlighted cells have correlation higher than 50 per cent.
 • The numbers within parentheses indicate number of factors used in DFM.

Source: Authors' calculations

Next, the turnaround point analysis was carried out on the proposed CLI and the reference series using Harding & Pagan (2002) approach. The turnaround points of CLI are mapped with the reference series to track the leading property. CLI based on Random Forest with XGBoost has one quarter lead, whereas the Random Forest has lead of average one to two quarters (Table 8).

Lastly, the time series plot of the proposed CLI and reference series establish the leading properties of the CLI visually except for the COVID period. The

economic disruption during COVID pandemic led to broad based slowdown in Indian economy followed by a gradual recovery. The extent of recovery varied across segments which weakened the leading property of CLI during the recent pandemic period (Chart 1).

Following the derivation, CLI is proposed using variables selected from Random Forest - XGBoost criteria and aggregating those using inverse standard deviation as weights. The selected variables are listed in Table 9.

Table 8: Leading property of CLI

Variable Selection	CLI Construction	Peak	Trough
Random Forest - XGBoost	Weighted Average - SD	2015.50	2016.75
Random Forest - XGBoost	Weighted Average - SD	2017.75	2020.00
Random Forest - XGBoost	Weighted Average - SD	2021.25	2022.25
Random Forest - XGBoost	Weighted Average - SD	2023.50	
Random Forest	Weighted Average - SD	2015.25	2017.00
Random Forest	Weighted Average - SD	2018.00	2019.00
Random Forest	Weighted Average - SD	2021.25	2022.25
Random Forest	Weighted Average - SD	2023.50	
Reference Series			
Reference Series	GVA Manufacturing	2015.75	2017.25
	GVA Manufacturing	2018.00	2020.00
	GVA Manufacturing	2021.25	2022.50
	GVA Manufacturing	2023.75	

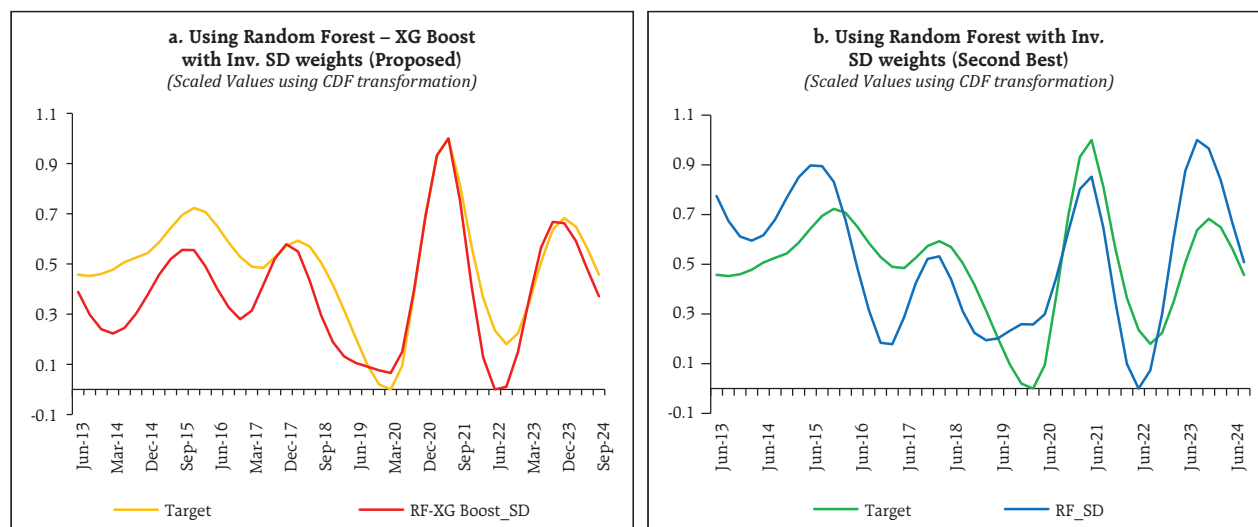
Note: 1. Green shaded CLI is the proposed one and white shaded CLI is the next best one.

2. The timeline of peaks and troughs are provided in fractions. YYYY.00 represents Q1 of year YYYY, YY.25 is Q2 of YYYY, YY.50 is Q3 of YYYY and YY.75 is for Q4 of YYYY. All years are financial years.

Source: Authors' calculations.

It may be mentioned that the business cycle exhibits long-term pattern which evolves over time. For business cycle analysis using quarterly data, it is generally recommended to have a time series length of at least 30 to 50 years (*i.e.*, 120 to 200 quarterly observations). This duration allows for the identification of multiple complete business cycles,

which typically last between 5 to 10 years (OECD, 2008; Canova, 1998). A longer time series also helps in improving the reliability of statistical methods used in trend-cycle decomposition, filtering techniques (*e.g.*, HP filter), and econometric modelling. In this analysis, 12 years data is used for deriving the leading indicator of GVA-manufacturing which is

Chart 1: CLI and Reference Series

Sources: : Authors' calculations.

Table 9: Variables selected in Random Forest and Mutual Information

Cement Production	WPI Headline
Commercial Motor Vehicle Sales	WPI Manufacturing
IIP Consumer Durable	G-Sec 10 Years Yield
IIP Consumer Non-durable	WCMR
High Speed Diesel	IMF All (Excl. Gold) Commodity Prices
International Air Cargo	Exports to Emerging and Developing Asia
Domestic Air Passenger Traffic	US Non-farm Payroll Employment
Real Credit to Industry	Global EPU
India EPU	

reasonable but may not be sufficient to study longer cycles, particularly due to the presence of COVID-19 led disruptions. The post pandemic data spans for two years which is insufficient to understand any changes in the data generating process of the economic variables. Following the data limitations and disruptions due to the COVID pandemic, it is recommended to revisit the construction of CLI at regular interval with better data availability.

VI. Concluding Remarks

This paper proposed a composite leading indicator for tracking the growth rate cycle of GVA Manufacturing using various high frequency indicators. Among the high frequency indicators, commodity prices, use-based classification of IIP, forward looking survey-based indicators, credit disbursed to industry, policy uncertainty and global commodity prices appeared to possess leading property on GVA manufacturing growth.

The CLI constructed using random forest and XGBoost exhibits the highest tracking power with cross correlation of 86 per cent at lag of one quarter (contrary to 72 per cent contemporaneous correlation). The turnaround points of the constructed CLI leads the GVA-manufacturing turnaround points by one quarter. The leading property of the proposed CLI shows robustness in the pre-COVID period and post-pandemic recovery period.

Reference

- Altissimo, Filippo, Alessandro Bassanetti, Roberto Cristadoro, Mario Forni, Marco Hallin, Marco Lippi, and Giovanni Veronese (2000). A Real Time Coincident Indicator of the Euro Area Business Cycle. European Central Bank Working Paper Series No. 42. Frankfurt: European Central Bank.
- Berndt, D. J., and Clifford, J. (1994). Using Dynamic Time Warping to find patterns in time series. Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining (AAAI Workshop), 359-370.
- Boschan, Charlotte (1975). Cyclical Analysis of Time Series: Selected Procedures and Computer Programs. National Bureau of Economic Research Technical Paper No. 20. New York: NBER.
- Bry, G., & Boschan, C. (1971). Cyclical analysis of time series: Selected procedures and computer programs. New York: National Bureau of Economic Research.
- Burns, A. F., and Mitchell, W. C. (1946). Measuring Business Cycles. National Bureau of Economic Research.
- Canova, F. (1998). Detrending and business cycle facts. *Journal of Monetary Economics*, 41(3), 475-512.
- Chitre, V. S. (1982). Growth Cycles in the Indian Economy. Gokhale Institute of Politics & Economics / Artha Vijnana, 158 pp.
- Chitre, Vikas (2001). Indicators of Business Recessions and Revivals in India: 1951-1982. *Indian Economic Review*, 36(1), pp. 79-105.
- Cover, T. M., and Thomas, J. A. (2006). Elements of Information Theory (2nd ed.). Wiley.
- Diebold, Francis X., & Rudebusch, Glenn D. (1991). Forecasting Output with the Composite Leading Index: A Real-Time Analysis. *Journal of the American Statistical Association*, 86(415), 603-610.
- Drehmann, Mathias & Yetman, James (2018). Why You Should Use the Hodrick-Prescott Filter — At Least to Generate Credit Gaps. BIS Working Papers No. 744, Bank for International Settlements.

- Dua, Pami & Banerji, Anirvan (1999). An Index of Coincident Economic Indicators for the Indian Economy. *Journal of Quantitative Economics*, 15, 177-201.
- Greene, W. H. (2012). *Econometric Analysis* (7th ed.). Pearson.
- Grehmann, K., & Yetman, J. (2018). How monetary policy affects economic conditions: Evidence from cross-country data. BIS Working Paper No. 729.
- Harding, Don & Pagan, Adrian (2002). Dissecting the Cycle: A Methodological Investigation. *Journal of Monetary Economics*, 49(2), 365–381.
- Hatekar, N. H. (1993). The short-term dynamics of money and output in India. Mimeograph, University of Bombay, Bombay.
- Huang, A. (2008). Similarity measures for text document clustering. *Proceedings of the Sixth New Zealand Computer Science Research Student Conference (NZCSRSC)*, 49-56.
- Klein, Philip A. & Moore, Geoffrey H. (1985). *Monitoring Growth Cycles in Market-Oriented Countries: Developing and Using International Economic Indicators*. Cambridge, MA: Ballinger / NBER Studies in Business Cycles.
- Koenker, R., and Bassett, G. (1978). Regression quantiles. *Econometrica*, 46(1), 33-50.
- Mintz, Ilse (1961). World Imports and United States Business Cycles. *American Exports during Business Cycles, 1879–1958*, pp. 35–41. National Bureau of Economic Research.
- Mintz, I. (1969). *Dating postwar business cycles: Methods and their application to Western Germany, 1950–1967*. New York: National Bureau of Economic Research.
- Mintz, I. (1972). *Dating postwar business cycles: Methods and their application to France, 1949–1967*. New York: National Bureau of Economic Research.
- Mintz, I. (1974). *Dating postwar business cycles: Methods and their application to Italy, 1950–1967*. New York: National Bureau of Economic Research.
- Moore, Geoffrey H. (1982). *Business Cycles, Inflation, and Forecasting*. 2nd Edition. National Bureau of Economic Research (NBER) Studies in Business Cycles, No. 24. Chicago: University of Chicago Press.
- Penna Urrila, M. (2001). Business cycle asymmetry: An application of nonlinear models to the Finnish economy. Bank of Finland Discussion Paper No. 19/2001.
- Murutoglu, G. (1999). Business cycles: Persistence, asymmetries and international transmission. Unpublished manuscript, Department of Economics, University of Illinois at Urbana–Champaign.
- OECD (2008). *OECD System of Composite Leading Indicators: A Revised Framework*. Organisation for Economic Co-operation and Development.
- Roesch, A., & Schmidbauer, H. (2018). WaveletComp: computational wavelet analysis. R package version, 1(1).
- Roy, I., and Biswas, D. (2012). Construction of a composite index of leading indicators for the index of industrial production in India. *Journal of Quantitative Economics*, 10(1), 17-31.
- Stock, J. H., and Watson, M. W. (1989). New indexes of coincident and leading economic indicators. *NBER Macroeconomics Annual*, 4, 351-393.
- Stock, J. H., and Watson, M. W. (1999). Business cycle fluctuations in US macroeconomic time series. *Handbook of Macroeconomics*, 1, 3-64.
- Urrila, Penna (2001). *Suhdanneindikaattoreiden käyttö talouskehityksen seurannassa*. (Working Paper No. 780). ETLA — Elinkeinoelämän tutkimuslaitos, Helsinki.
- Zarnowitz, Victor & Moore, Geoffrey H. (1986). Major Changes in Cyclical Behavior, *The American Business Cycle: Continuity and Change*, edited by Robert J. Gordon, 519–582. Chicago: University of Chicago Press / National Bureau of Economic Research

Annex I : Methods used for Variable Selection

Cross-Correlation:

Cross-correlation analysis is a statistical technique used to measure the relationship between two time series at different lags, helping to identify lead-lag relationships and synchronicity between variables. In business cycle analysis, it is often employed to examine how economic indicators move in relation to the overall cycle, determining whether they are leading, coincident, or lagging indicators (Stock and Watson, 1999). A high cross-correlation at a positive lag suggests that one variable tends to lead the other, while a strong correlation at zero lag indicates simultaneous movement. This method is widely used in macroeconomic research to assess the predictive power of economic indicators and understand transmission mechanisms across sectors.

Regression Analysis

Regression analysis using Ordinary Least Squares (OLS) is a fundamental statistical method for estimating relationships between dependent and independent variables in economic and financial research. OLS minimizes the sum of squared residuals to derive the best-fitting linear equation, making it widely used for business cycle analysis, forecasting, and policy evaluation (Greene, 2012).

Quantile Regression

Quantile regression is an econometric technique that extends traditional Ordinary Least Squares (OLS) regression by estimating the conditional relationship between variables at different points of the outcome distribution (Koenker and Bassett, 1978). Unlike OLS, which models the mean effect, quantile regression provides a more comprehensive view of the data by capturing heterogeneous effects across quantiles. This is particularly useful in business cycle analysis, where economic relationships may vary during recessions and expansions.

Cosine Similarity

Cosine similarity is a metric used to measure the similarity between two non-zero vectors by computing the cosine of the angle between them. It is widely applied in text analysis, machine learning, and economic research to compare patterns in high-dimensional data. Unlike Euclidean distance, cosine similarity focuses on the direction rather than the magnitude of vectors, making it useful for comparing time series with different scales. In business cycle analysis, it can be employed to assess the similarity of economic indicators or compare the cyclical patterns of different countries over time. A value close to 1 indicates high similarity, while a value near 0 suggests no correlation (Huang, 2008).

Mutual Information

Mutual information (MI) is an information-theoretic measure that quantifies the dependency between two random variables by capturing both linear and nonlinear relationships (Cover and Thomas, 2006). Unlike correlation, which only detects linear dependencies, MI assesses the reduction in uncertainty about one variable given knowledge of another. In business cycle analysis, MI can be used to evaluate the strength of associations between macroeconomic indicators, such as GDP growth and inflation, across different economic conditions.

Dynamic Time Warping

Dynamic Time Warping (DTW) is an algorithm used to measure the similarity between two time series by allowing non-linear distortions in the time axis (Berndt and Clifford, 1994). Unlike traditional distance metrics, such as Euclidean distance, DTW aligns sequences of different lengths or with temporal shifts by finding an optimal warping path that minimizes the cumulative distance between corresponding points. This makes it particularly useful in business cycle analysis, where economic indicators may exhibit phase shifts or different speeds of fluctuation across countries or industries.

Annex II : List of HFI Considered for CLI

IIP Data	Exchange Rate
IIP Manufacturing	REER
IIP Headline Index	NEER
IIP Primary Goods	INR - USD Exchange Rate
IIP Capital Goods	Global Commodity Price
IIP Intermediate Goods	IMF All Commodity Prices
IIP Infrastructure Goods	IMF Commodity prices excluding Gold
IIP Consumer Durables	IMF Commodity Price of Industrial Raw Material
IIP Consumer Non-durables	IMF Commodity Price of Metals
Global Trade	IMF Commodity Price of Base Metals
Exports to Emerging and Developing Asia	IMF Commodity Price of Fuel
Exports to Europe	IMF Commodity Price of Crude Oil
US Non-farm payroll Data	IMF Commodity Price of Coal
US IIP	World Bank Price - Aluminium
China IIP	World Bank Price - Iron
External Trade	World Bank Price - Copper
Merchandise Exports	World Bank Price - Lead
Merchandise Imports	World Bank Price - Tin
Non-oil non-gold imports	World Bank Price - Nickel
Non-oil exports	World Bank Price - Zinc
Export of services	Indian Basket Crude Oil Price
Import of services	WTI Crude Oil Price
Import of Capital Goods	Brent Crude Oil Price
Employment Condition	Dubai Crude Oil Price
CMIE Labour Force Participation - All India	PMI Data
CMIE Labour Force Participation - Urban	PMI Index
CMIE Labour Force Participation - Rural	PMI Output
CMIE Unemployment Rate - All India	PMI New Orders
CMIE Unemployment Rate - Urban	PMI Employments
CMIE Unemployment Rate - Rural	PMI Supplier Delivery Time
CMIE Employment Rate - All India	PMI Stock of Purchase
CMIE Employment Rate - Urban	PMI Input Prices
CMIE Employment Rate - Rural	PMI Quantity of Purchase
NAUKRI Job Speak Index	PMI Stocks of Finished Goods
MGNREGA Work Demand	PMI New Export Orders
Payment and Public Finance	PMI Output Prices
RTGS Payments	PMI Backlog of Work
UPI Payments	PMI Future Output
E-Way Bills	Eight Core Headline
GST Collection	Coal Production
Revenue Expenditure (less interest payments and subsidy) of Central Government	Crude Oil Production
Fertiliser Sales	Natural Gas Production
	Petroleum Products production
	Fertilizers production
	Steel Production
	Cement Production
	Electricity Production

Domestic Price Condition

WPI Headline Index

WPI Food Prices

WPI Primary Articles

WPI Fuel and Power

WPI Manufactured Items

WPI Industrial Raw Material

CPI Headline

CPI Index excluding Food, Fuel and Beverages

Consumption Indicators

Finished Steel Consumption

Petroleum Consumption

High Speed Diesel Consumption

Motor Spirit Consumption

Aviation Turbine Fuel Consumption

Trade, Transport and Demand Indicators

Domestic Air Passenger Traffic

International Air Passenger Traffic

Domestic Air Cargo

International Air Cargo

Railway Freight

Port Traffic

Passenger Vehicle Sales (Wholesale)

Passenger Vehicle Sales (Wholesale) - LMV

Two Wheeler Sales (Wholesale)

Two Wheeler Sales (Retail)

Three Wheeler Sales (Domestic)

Tractor Sales

Electricity Demand

Policy Uncertainty

Global Policy Uncertainty

India Policy Uncertainty

USA Policy Uncertainty - Three factor model

USA Policy Uncertainty - Newspaper Based

China Policy Uncertainty

European Union Policy Uncertainty - Newspaper

Germany Policy Uncertainty -

Italy Policy Uncertainty -

UK Policy Uncertainty -

France Policy Uncertainty -

Spain Policy Uncertainty -

PE Ratio of listed companies

Realized volatility of BSE Companies

IOS and OBICUS Data

Industrial Outlook Survey - Production (Expectation)

Industrial Outlook Survey - Order Book (Expectation)

Industrial Outlook Survey - Capacity Utilization (Expectation)

Industrial Outlook Survey - Exports (Expectation)

Industrial Outlook Survey - Imports (Expectation)

Industrial Outlook Survey - Inventory of Raw Material (Expectation)

Industrial Outlook Survey - Inventory of Finished Goods (Expectation)

Industrial Outlook Survey - Employment (Expectation)

Industrial Outlook Survey - Financial Condition (Expectation)

Industrial Outlook Survey - Cost of Finance (Expectation)

Industrial Outlook Survey - Cost of Raw Material (Expectation)

Industrial Outlook Survey - Selling Price (Expectation)

Industrial Outlook Survey - Profit Margin (Expectation)

Industrial Outlook Survey - Overall Business Condition

Capacity Utilization

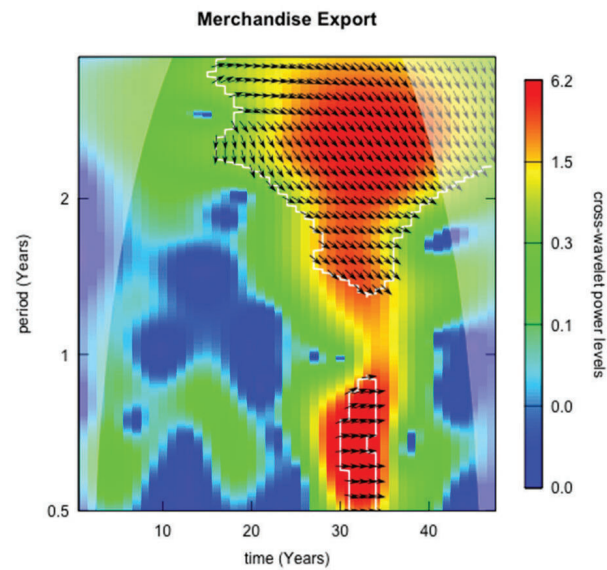
Cost of Borrowing Proxy

Weighted Average Call Money Rate

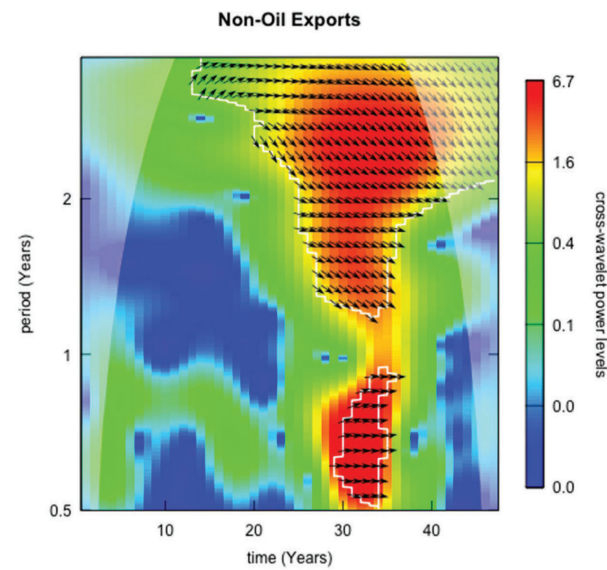
91 days T-Bill Rate

G-Sec 10 Years Yield

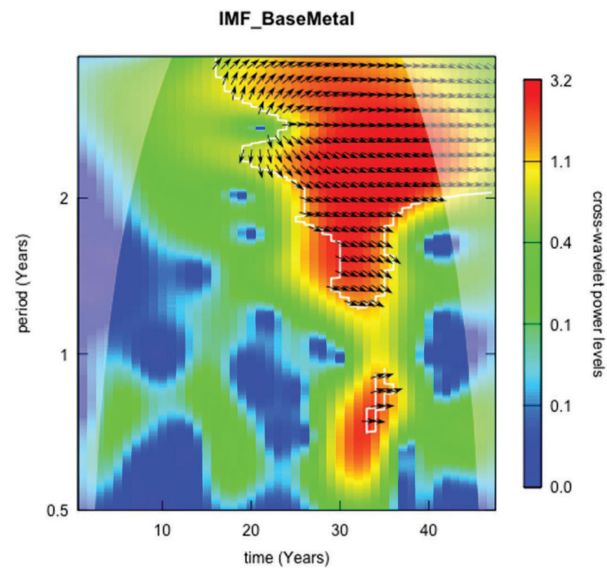
Annex III: Wavelet Coherence



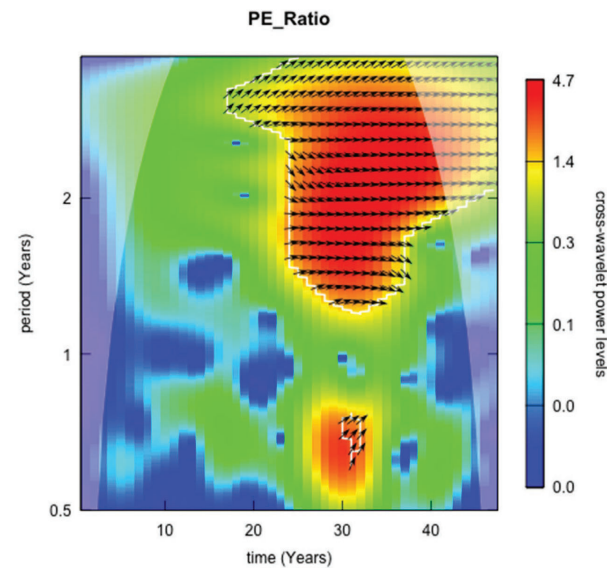
Source: Authors' calculations



Source: Authors' calculations



Source: Authors' calculations

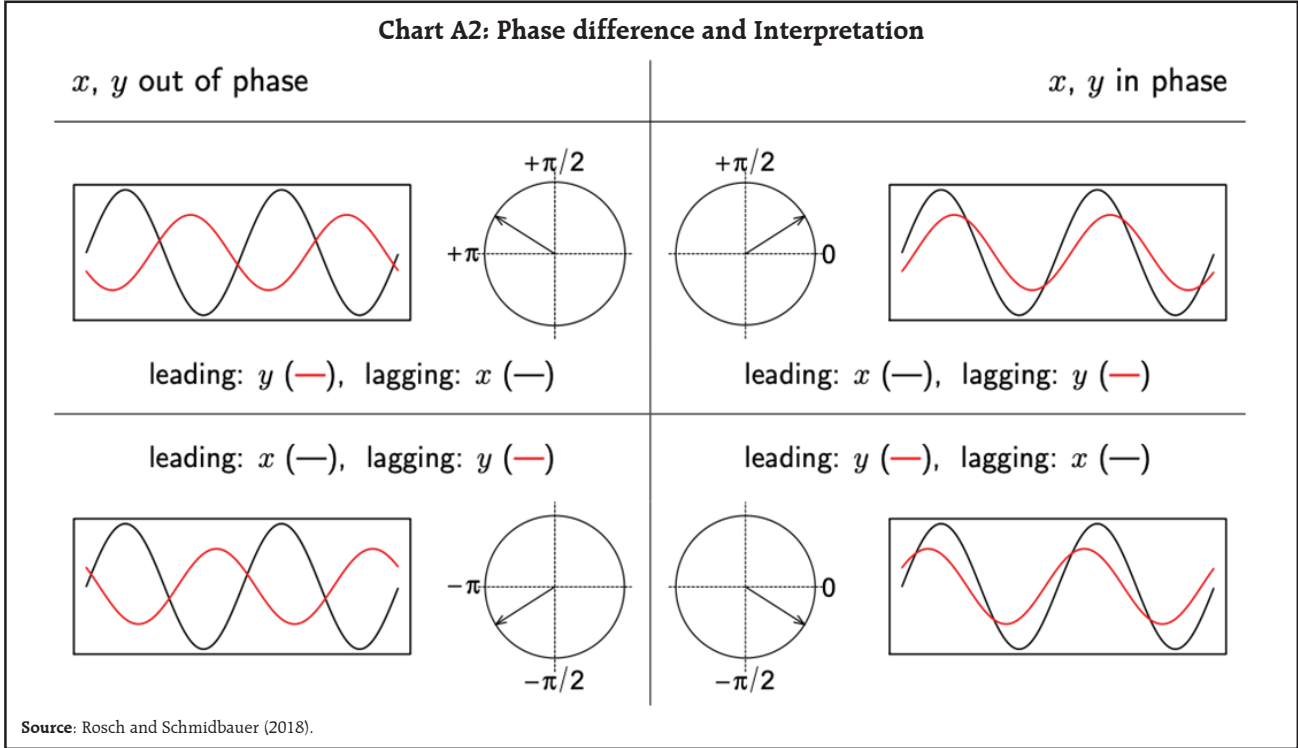


Source: Authors' calculations

Annex IV : Interpretation of Wavelet Charts

Wavelet charts, or wavelet power spectra, are used to analyze signals in both time and frequency domains simultaneously. They help in detecting transient features, periodicities, and localized frequency variations in data. The x-axis of the wavelet charts plots the time whereas the y-axis shows the

different periodicities. The wavelet charts plot the phase differences between two series which help to identify the business cycle leading – lagging properties between two economic indicators. The direction of arrows in the phase diagram indicates the relationship between the economic indicators (Chart A2).



Annex V : Cross Correlation Estimates

Variable Selection	CLI Construction	Lag = 1		Lag = 2	
		Full Sample	Excluding COVID	Full Sample	Excluding COVID
RFE	Simple Average	0.08	0.43	0.24	0.31
RFE	Weighted Average - SD	0.20	0.42	0.37	0.30
RFE	Weighted Average - Correlation	0.10	0.43	0.26	0.32
RFE	Dynamic Factor Model (2 factors)	0.06	0.34	0.20	0.24
RFE	Dynamic Factor Model (1 factor)	0.06	0.51	0.24	0.48
RF	Simple Average	0.62	0.64	0.34	0.61
RF	Weighted Average - SD	0.74	0.78	0.61	0.73
RF	Weighted Average - Correlation	0.59	0.66	0.28	0.61
RF	Dynamic Factor Model (2 factors)	0.07	0.34	0.29	0.32
RF	Dynamic Factor Model (1 factor)	0.07	0.51	0.26	0.50
XG Boost	Simple Average	0.41	0.20	0.17	0.29
XG Boost	Weighted Average - SD	0.63	0.44	0.35	0.30
XG Boost	Weighted Average - Correlation	0.41	0.18	0.17	0.27
XG Boost	Dynamic Factor Model (2 factors)	0.42	0.15	0.15	0.27
XG Boost	Dynamic Factor Model (1 factor)	0.46	0.07	0.17	0.08
Pearson	Simple Average	0.50	0.53	0.54	0.43
Pearson	Weighted Average - SD	0.53	0.43	0.49	0.26
Pearson	Weighted Average - Correlation	0.44	0.52	0.51	0.44
Pearson	Dynamic Factor Model (2 factors)	0.02	0.44	0.19	0.49
Pearson	Dynamic Factor Model (1 factor)	0.08	0.47	0.32	0.50
Kendall	Simple Average	0.15	0.53	0.35	0.51
Kendall	Weighted Average - SD	0.25	0.55	0.45	0.51
Kendall	Weighted Average - Correlation	0.16	0.53	0.35	0.50
Kendall	Dynamic Factor Model (2 factors)	0.23	0.57	0.38	0.50
Kendall	Dynamic Factor Model (1 factor)	0.06	0.51	0.26	0.50
Spearman	Simple Average	0.12	0.53	0.34	0.54
Spearman	Weighted Average - SD	0.26	0.60	0.45	0.56
Spearman	Weighted Average - Correlation	0.28	0.62	0.46	0.60
Spearman	Dynamic Factor Model (2 factors)	0.07	0.44	0.33	0.48
Spearman	Dynamic Factor Model (1 factor)	0.06	0.48	0.29	0.49
Quantile	Simple Average	0.22	0.39	0.40	0.30
Quantile	Weighted Average - SD	0.20	0.12	0.32	0.06
Quantile	Weighted Average - Correlation	0.21	0.48	0.40	0.38
Quantile	Dynamic Factor Model (2 factors)	0.02	0.14	0.11	0.19
Quantile	Dynamic Factor Model (1 factor)	0.03	0.52	0.04	0.53
MI	Simple Average	0.66	0.59	0.54	0.55
MI	Weighted Average - SD	0.29	0.53	0.30	0.50
MI	Weighted Average - Correlation	0.29	0.52	0.41	0.53

Variable Selection	CLI Construction	Lag = 1		Lag = 2	
		Full Sample	Excluding COVID	Full Sample	Excluding COVID
MI	Dynamic Factor Model (2 factors)	0.15	0.40	0.20	0.38
MI	Dynamic Factor Model (1 factor)	0.09	0.56	0.24	0.49
Cosine	Simple Average	0.22	0.00	0.14	0.05
Cosine	Weighted Average - SD	0.20	0.06	0.15	0.03
Cosine	Weighted Average - Correlation	0.18	0.01	0.07	0.06
Cosine	Dynamic Factor Model (2 factors)	0.07	0.07	0.09	0.12
Cosine	Dynamic Factor Model (1 factor)	0.08	0.06	0.08	0.03
DTW	Simple Average	0.15	0.26	0.17	0.30
DTW	Weighted Average - SD	0.38	0.21	0.38	0.33
DTW	Weighted Average - Correlation	0.15	0.26	0.17	0.30
DTW	Dynamic Factor Model (2 factors)	0.15	0.02	0.01	0.07
DTW	Dynamic Factor Model (1 factor)	0.10	0.06	0.05	0.02
Union	Simple Average	0.14	0.26	0.17	0.30
Union	Weighted Average - SD	0.67	0.74	0.67	0.65
Union	Weighted Average - Correlation	0.14	0.26	0.17	0.30
Union	Dynamic Factor Model (2 factors)	0.07	0.55	0.21	0.55
Union	Dynamic Factor Model (1 factor)	0.08	0.47	0.32	0.50
RFE-RF	Simple Average	0.43	0.62	0.42	0.50
RFE-RF	Weighted Average - SD	0.58	0.70	0.61	0.57
RFE-RF	Weighted Average - Correlation	0.67	0.67	0.50	0.58
RFE-RF	Dynamic Factor Model (2 factors)	0.05	0.51	0.19	0.46
RFE-RF	Dynamic Factor Model (1 factor)	0.07	0.51	0.25	0.49
RFE-XG Boost	Simple Average	0.38	0.29	0.15	0.32
RFE-XG Boost	Weighted Average - SD	0.53	0.54	0.53	0.38
RFE-XG Boost	Weighted Average - Correlation	0.39	0.26	0.16	0.31
RFE-XG Boost	Dynamic Factor Model (2 factors)	0.01	0.47	0.20	0.41
RFE-XG Boost	Dynamic Factor Model (1 factor)	0.12	0.20	0.17	0.12
RFE-Spearman	Simple Average	0.10	0.51	0.29	0.42
RFE-Spearman	Weighted Average - SD	0.25	0.58	0.45	0.50
RFE-Spearman	Weighted Average - Correlation	0.11	0.54	0.31	0.49
RFE-Spearman	Dynamic Factor Model (2 factors)	0.14	0.43	0.40	0.36
RFE-Spearman	Dynamic Factor Model (1 factor)	0.06	0.51	0.25	0.49
RFE-Quantile	Simple Average	0.08	0.44	0.26	0.32
RFE-Quantile	Weighted Average - SD	0.26	0.46	0.45	0.35
RFE-Quantile	Weighted Average - Correlation	0.10	0.43	0.28	0.32
RFE-Quantile	Dynamic Factor Model (2 factors)	0.07	0.31	0.24	0.24
RFE-Quantile	Dynamic Factor Model (1 factor)	0.08	0.54	0.23	0.47
RFE-MI	Simple Average	0.23	0.52	0.36	0.41
RFE-MI	Weighted Average - SD	0.30	0.58	0.41	0.50
RFE-MI	Weighted Average - Correlation	0.33	0.55	0.44	0.46

Variable Selection	CLI Construction	Lag = 1		Lag = 2	
		Full Sample	Excluding COVID	Full Sample	Excluding COVID
RFE-MI	Dynamic Factor Model (2 factors)	0.09	0.48	0.15	0.42
RFE-MI	Dynamic Factor Model (1 factor)	0.08	0.54	0.24	0.49
RFE-Cosine	Simple Average	0.10	0.44	0.27	0.33
RFE-Cosine	Weighted Average - SD	0.34	0.29	0.46	0.25
RFE-Cosine	Weighted Average - Correlation	0.14	0.46	0.30	0.36
RFE-Cosine	Dynamic Factor Model (2 factors)	0.08	0.05	0.08	0.01
RFE-Cosine	Dynamic Factor Model (1 factor)	0.08	0.06	0.08	0.03
RF-XG Boost	Simple Average	0.39	0.29	0.16	0.37
RF-XG Boost	Weighted Average - SD	0.86	0.85	0.59	0.71
RF-XG Boost	Weighted Average - Correlation	0.40	0.23	0.17	0.31
RF-XG Boost	Dynamic Factor Model (2 factors)	0.27	0.39	0.00	0.42
RF-XG Boost	Dynamic Factor Model (1 factor)	0.46	0.05	0.16	0.09
RF-Spearman	Simple Average	0.64	0.77	0.54	0.74
RF-Spearman	Weighted Average - SD	0.56	0.73	0.64	0.68
RF-Spearman	Weighted Average - Correlation	0.58	0.76	0.50	0.75
RF-Spearman	Dynamic Factor Model (2 factors)	0.06	0.52	0.24	0.50
RF-Spearman	Dynamic Factor Model (1 factor)	0.05	0.48	0.28	0.49
RF-Quantile	Simple Average	0.66	0.69	0.39	0.65
RF-Quantile	Weighted Average - SD	0.66	0.75	0.69	0.67
RF-Quantile	Weighted Average - Correlation	0.70	0.72	0.44	0.69
RF-Quantile	Dynamic Factor Model (2 factors)	0.07	0.32	0.24	0.24
RF-Quantile	Dynamic Factor Model (1 factor)	0.08	0.54	0.23	0.47
RF-MI	Simple Average	0.73	0.74	0.45	0.69
RF-MI	Weighted Average - SD	0.55	0.65	0.44	0.60
RF-MI	Weighted Average - Correlation	0.68	0.68	0.36	0.64
RF-MI	Dynamic Factor Model (2 factors)	0.10	0.47	0.24	0.47
RF-MI	Dynamic Factor Model (1 factor)	0.07	0.52	0.26	0.50
RF-Cosine	Simple Average	0.62	0.62	0.34	0.59
RF-Cosine	Weighted Average - SD	0.55	0.43	0.44	0.45
RF-Cosine	Weighted Average - Correlation	0.52	0.48	0.21	0.47
RF-Cosine	Dynamic Factor Model (2 factors)	0.09	0.00	0.04	0.04
RF-Cosine	Dynamic Factor Model (1 factor)	0.08	0.06	0.08	0.03
XG Boost-Spearman	Simple Average	0.39	0.26	0.16	0.34
XG Boost-Spearman	Weighted Average - SD	0.63	0.76	0.66	0.70
XG Boost-Spearman	Weighted Average - Correlation	0.38	0.32	0.14	0.39
XG Boost-Spearman	Dynamic Factor Model (2 factors)	0.10	0.52	0.30	0.52

Variable Selection	CLI Construction	Lag = 1		Lag = 2	
		Full Sample	Excluding COVID	Full Sample	Excluding COVID
XG Boost-Spearman	Dynamic Factor Model (1 factor)	0.04	0.47	0.28	0.49
XG Boost-Quantile	Simple Average	0.40	0.20	0.17	0.29
XG Boost-Quantile	Weighted Average - SD	0.49	0.33	0.41	0.22
XG Boost-Quantile	Weighted Average - Correlation	0.41	0.19	0.17	0.27
XG Boost-Quantile	Dynamic Factor Model (2 factors)	0.29	0.19	0.03	0.26
XG Boost-Quantile	Dynamic Factor Model (1 factor)	0.46	0.06	0.17	0.08
XG Boost-MI	Simple Average	0.40	0.24	0.16	0.32
XG Boost-MI	Weighted Average - SD	0.56	0.58	0.47	0.54
XG Boost-MI	Weighted Average - Correlation	0.40	0.24	0.16	0.32
XG Boost-MI	Dynamic Factor Model (2 factors)	0.30	0.07	0.04	0.13
XG Boost-MI	Dynamic Factor Model (1 factor)	0.03	0.49	0.24	0.48
XG Boost-Cosine	Simple Average	0.40	0.20	0.17	0.29
XG Boost-Cosine	Weighted Average - SD	0.42	0.12	0.26	0.10
XG Boost-Cosine	Weighted Average - Correlation	0.41	0.17	0.17	0.26
XG Boost-Cosine	Dynamic Factor Model (2 factors)	0.38	0.04	0.09	0.05
XG Boost-Cosine	Dynamic Factor Model (1 factor)	0.08	0.06	0.08	0.03
Spearman-Quantile	Simple Average	0.13	0.54	0.36	0.53
Spearman-Quantile	Weighted Average - SD	0.30	0.50	0.50	0.44
Spearman-Quantile	Weighted Average - Correlation	0.28	0.61	0.48	0.57
Spearman-Quantile	Dynamic Factor Model (2 factors)	0.08	0.39	0.25	0.31
Spearman-Quantile	Dynamic Factor Model (1 factor)	0.08	0.54	0.23	0.47
Spearman-MI	Simple Average	0.56	0.59	0.57	0.57
Spearman-MI	Weighted Average - SD	0.43	0.56	0.49	0.54
Spearman-MI	Weighted Average - Correlation	0.66	0.57	0.57	0.55
Spearman-MI	Dynamic Factor Model (2 factors)	0.10	0.43	0.28	0.47
Spearman-MI	Dynamic Factor Model (1 factor)	0.06	0.48	0.29	0.49
Spearman-Cosine	Simple Average	0.16	0.52	0.37	0.54
Spearman-Cosine	Weighted Average - SD	0.34	0.44	0.51	0.45

Variable Selection	CLI Construction	Lag = 1		Lag = 2	
		Full Sample	Excluding COVID	Full Sample	Excluding COVID
Spearman-Cosine	Weighted Average - Correlation	0.15	0.54	0.38	0.54
Spearman-Cosine	Dynamic Factor Model (2 factors)	0.09	0.42	0.27	0.46
Spearman-Cosine	Dynamic Factor Model (1 factor)	0.06	0.48	0.29	0.49
Quantile-MI	Simple Average	0.66	0.60	0.57	0.55
Quantile-MI	Weighted Average - SD	0.40	0.44	0.47	0.40
Quantile-MI	Weighted Average - Correlation	0.40	0.47	0.43	0.45
Quantile-MI	Dynamic Factor Model (2 factors)	0.07	0.30	0.24	0.22
Quantile-MI	Dynamic Factor Model (1 factor)	0.08	0.54	0.23	0.47
Quantile-Cosine	Simple Average	0.31	0.20	0.31	0.24
Quantile-Cosine	Weighted Average - SD	0.31	0.13	0.34	0.15
Quantile-Cosine	Weighted Average - Correlation	0.32	0.12	0.31	0.19
Quantile-Cosine	Dynamic Factor Model (2 factors)	0.08	0.21	0.04	0.26
Quantile-Cosine	Dynamic Factor Model (1 factor)	0.03	0.52	0.04	0.53
MI-Cosine	Simple Average	0.66	0.56	0.55	0.54
MI-Cosine	Weighted Average - SD	0.36	0.40	0.39	0.40
MI-Cosine	Weighted Average - Correlation	0.59	0.41	0.44	0.44
MI-Cosine	Dynamic Factor Model (2 factors)	0.12	0.05	0.01	0.07
MI-Cosine	Dynamic Factor Model (1 factor)	0.09	0.06	0.06	0.02
Coherence	Simple Average	0.46	0.68	0.51	0.65
Coherence	Weighted Average - SD	0.49	0.68	0.55	0.61
Coherence	Weighted Average - Correlation	0.39	0.66	0.48	0.63
Coherence	Dynamic Factor Model (2 factors)	0.12	0.51	0.32	0.54
Coherence	Dynamic Factor Model (1 factor)	0.08	0.47	0.32	0.50

Note: • Highlighted cells have correlation higher than 50 per cent.
 • The numbers within parentheses indicate number of factors used in DFM.

Source: Authors' calculations.