OXFORD JOURNAL
OF STUDENT
SCHOLARSHIP

A Regression-Based Composite Indicator for Economic Forecasting: Integrating Financial and Real-Sector Signals to Improve Macroeconomic Nowcasting

Shaurya Suresh
shauryasuresh9431@gmail.com

ABSTRACT

This paper develops a multivariate regression-based indicator constructed from non-traditional economic variables to improve real-time estimation of aggregate economic output and to identify potential economic downturns and recessions in future years by mimicking GDP at a higher frequency than the official value is released (i.e., quarterly). This indicator reduces economic reporting delays by blending real-activity data with forward-looking financial indicators through a structured lag framework, thereby capturing both general and specific economic dynamics in the global economy. The results of this analysis provide an accurate indicator (89.40% similarity to the industrial production index) and diagnostic insights from residual analysis. The results indicate that combining leading and contemporaneous indicators improves the robustness of macroeconomic nowcasting frameworks. The proposed indicator demonstrates the value of integrating multiple non-traditional economic channels into a unified framework for real-time macroeconomic monitoring, This paper argues that GDP is able to be accurately tracked using non-traditional, high frequency indicators, overcoming reporting delays associated with GDP quarterly release by using a multivariate regression model combining real-activity, financial stress, and labor market indicators and data.

INTRODUCTION

Gross Domestic Product, also known as “GDP,” is one of, if not the most important, indicators of a country’s overall economic health and status. It is used to help governments, large banks, companies, and consumers understand the overall financial situation and its implications for key metrics, thereby enabling institutions to determine the actions needed to address current and future challenges. Policymakers might use Gross Domestic Product as a basis for proposed laws on unemployment or tariffs. In contrast, large banks often rely on Gross Domestic Product to assess investment opportunities and to inform risk-management analyses across various asset classes two distinct objectives, connected by the same source.

However, like any indicator, Gross Domestic Product has major underlying limitations. Firstly, it has a low reporting frequency. Gross Domestic Product is recorded only four times per year, around a month after each fiscal quarter. This becomes problematic when GDP is used as primary data to substantiate or develop new theories in economics, finance, or law, because the sample size is too small to be used meaningfully. Additionally, the official GDP report is released with considerable delay, creating significant lags between actual economic activity and its measurement. GDP is also such an indicator of economic activity, that its movements are visible and can be tracked across many sectors and data sources not only in the fields of economics and finance, but also in commercial and industrial settings. This makes GDP versatile and indicates that its behavior can be reliably estimated using alternative, high-frequency data sources across a variety of different sectors.

LITERATURE REVIEW

Financial crises are recurring features of modern financial systems. This theory is not limited to the United States but can be extended to nearly every economy. Despite differences across countries in their financial institutions, a common crisis structure can be identified and this recurring pattern is documented across multiple historical episodes spanning both countries and centuries, as synthesized in the book, Manias, Panics and Crashes: A History of Financial Crises (Kindleberger, 1978). This book highlights the economic crisis structure as follows: displacement, boom, euphoria, crisis, and the crash (panic). A displacement occurs when a significant change in the economic environment changes profit expectations. This may come from shifts such as technological innovations, financial deregulation, monetary easing, or geopolitical changes in the world. This displacement alters investor expectations, creating the belief that previously returns that were not possible are now feasible. The displacement is initially justified, providing the base line for increased investing and risk-taking. Following the displacement, credit conditions tend to lose, and it leads to financial intermediaries expanding lending to their customers. Asset prices begin to rise as increased liquidity enters the system. Rising prices improve balance sheets, which further relaxes borrowing constraints, producing a self-reinforcing feedback loop between asset valuations and credit creation. At this stage, optimism is widespread but still partially grounded in fundamentals. During euphoria, speculative behavior dominates investment decisions. Asset purchases are increasingly being motivated by expectations of short-term capital gains rather than underlying productivity. Leverage rises and risk becomes systematically underpriced . From this, valuations become disconnected from fundamentals, and dissenting views are marginalized. The last stage, the crisis, begins when a triggering event occurs. This can include policy shifts, interest rate increases, defaults, or the revelation of fraud, all of which can cause asset prices to rise. As previously said, this framework can be utilized to analyze any economic downturn.

Prime examples of this can be drawn from cases concerning the United States and the Japanese economy. The U.S. dot-com stock market bubble occurred from 1995 to 2000, when the U.S. experienced a surge in technology and internet-related stocks, especially on the NASDAQ, leading to multiple new firms with little or zero profits receiving extremely high valuations from the now easy access to venture capital and equity financing, combined with strong macroeconomic growth. The economic crisis structure is clearly seen in this episode, with the displacement starting from the internet and information-technology revolution, which reshaped expectations about productivity and future growth. This was followed by a boom that can be characterized by favorable lending conditions and robust economic expansion, which fueled widespread stock purchases. As optimism intensified into euphoria, investors increasingly purchased technology stocks with the reasoning of future capital gains rather than current earnings.

The crisis phase began in 2000, when stock prices stopped rising, and culminated in a crash as the NASDAQ lost approximately 80 percent of its value over the subsequent years. Likewise, the Japanese asset price bubble occurred during the late 1980s, when Japan experienced a rapid increase in real estate and stock values after long economic expansion and financial liberalization. Asset prices rose dramatically as corporations and households were able to access never ending credit, while land and equity were increasingly viewed as guaranteed stores of value. The economic crisis's structural features are clearly reflected in this episode, with displacement stemming from financial deregulation and expansionary monetary policy following the Plaza Accord, which reshaped expectations regarding growth and asset appreciation. This was followed by a boom characterized by aggressive bank lending and rising leverage that fueled widespread investment in stocks and real estate. As optimism intensified into euphoria, investors and firms made speculative purchases based on the belief that asset prices would still rise indefinitely rather than on underlying economic principles. The crisis phase emerged in the early 1990s, when asset prices began to deteriorate, and combined in a crash, where both equity and property values had collapsed, leading to a prolonged period of economic deterioration known as Japan's Lost Decade. While this episode highlights how financial crises unfold over time, understanding their real-time economic impacts requires methods that can track macroeconomic conditions as they evolve.

"Nowcasting" is the process of estimating key macroeconomic variables at a higher frequency. This paper highlights the concept of "nowcasting" and applies it to produce a higher-frequency estimate of GDP, making the measure more useful for real-time applications. Because GDP influences economic decisions, policy development, and even independent research, improving the speed and accessibility of GDP estimates has important value across many fields. This definition was introduced in Giannone, Reichlin, and Small (2008), which provided the fundamental theory of nowcasting and laid the foundation for modern day nowcasting through introducing a formal framework that used multiple mixed-frequency indicators, producing real-time GDP estimates. Their approach to this framework is based on the Kalman filter and state-space representation. The Kalman filter continuously updates the estimate of the economy's underlying condition as new data become available (typically released weekly or monthly). Since GDP is recorded only quarterly, the Kalman filter can incorporate financial information (primarily news releases), filter out unwanted and irregular information, and then update the forecast index. The state-space representation condenses the nowcasting model into two core elements . One includes a transition equation that describes how the unobserved economic state evolves (i.e., the latent factor). This latent factor indicates the “true” level of economic activity that cannot be directly observed, but that drives many macroeconomic indicators.

$$F_t = AF_{t-1} + B_{ut} \tag{1}$$

In equation (1), F_t represents the hidden economic state, A represents how much of the previous days economy carries into the present day, F_{t-1} represents the hidden economic state of the economy of the previous day, it represents the new shocks or surprises that hit the economy(both positive and negative news), and B represents how impactful the shocks are. The second part of the state-space representation is the measurement equation, which explains how the hidden economic state is observed in the data. Although we cannot directly observe the underlying economic factor Ft-1, we observe indicators such as industrial production, employment data, surveys, and other macroeconomic releases. Each of these variables reflects part of the actual economic state, but they also contain noise unrelated to the underlying economy. The measurement equation captures this idea mathematically in equation (2).

$$y_t = \Lambda F_r + \xi_t \tag{2}$$

In equation (2), y_t denotes the observed data at time t, ΛF_t is the matrix of factor loadings that shows how strongly each indicator responds to the latent factor, and ξ_t is the individual error term representing noise specific to each data series. When combined with the Kalman filter, new data can be incorporated and noise filtered during nowcast updates, thereby integrating the two features and enabling successful nowcasting. These findings provide the basis for which the theoretical framework explains why the selected indicators are expected to improve GDP nowcasting.

THEORETICAL FRAMEWORK

This paper serves the argument that Gross Domestic Product (GDP) can be accurately measured and tracked through the use of non-traditional high frequency indicators, which allows for reporting delays and its low reporting frequency to be overcome. Many current existing indicators such the New York Fed's Staff Nowcast show how data taken at high frequencies can positively affect the timeliness of GDP estimation. However, these models heavily rely on traditional indicators (industrial production and retail sales). This paper tests whether non-traditional indicators contain sufficient information to accurately estimate GDP in real time, and that too at a higher frequency. Due to the fact that GDP reflects aggregate production, consumption, investment, and employment, these activities can show observable signals across many different sectors before they even appear in official GDP releases.

This indicator is made through a multivariate regression combining 3 categories of indicators, including real-activity, financial stress, and labor market indicators and data. Real activity variables indicate what the economy is producing and consuming in the present time and can be seen through variables such as freight volumes and electricity consumption. Financial variables capture future economic trends and activity, as financial markets are known to price in anticipated economic conditions before they appear in real output and can be seen in variables such as the yield curve (T10Y3M) and the National Financial Conditions Index (NFCI). Labor market indicators tend to provide an alternate perspective when economic downturns affect employment. When compared to financial markets and financial indicators which respond almost instantaneously to changes in expectations, employment adjusts over a period of time, making indicators such as unemployment insurance claims useful for confirming when a downturn has changed into a more broader and serious economic contraction. By combining non-traditional indicators across three dimensions, the proposed framework provides an accurate and timely representation of GDP.

DATA AND SUMMARY STATISTICS

Throughout this paper, a variety of economic data were used and discussed. This includes the following: Gross Domestic Product (GDP), Industrial Production Index (IPI), Freight Truck Volumes, Electricity Usage, Hotel and Motel Rentals, and Automobile Rentals. The primary measure of model fit used in this paper is the coefficient of determination (R^2), which identifies the proportion of the variance in the dependent variable explained by the independent variable(s) in a statistical regression model. In this case, the dependent variable is IPI, and the independent variables are the economic variables. Additionally, the data used in this paper are alternative and are not typically considered in traditional economic models. However, two traditional economic variables are used to refine this model: the 10-Year Treasury Constant Maturity Minus 3-Month Treasury Constant Maturity (T10Y3M) and the Chicago Fed National Financial Conditions Index (NFCI). These variables serve merely as an aid to improve the model's stability and explanatory power during periods of financial stress.

Index Type	**Industrial Production Index(IPI)**	**Freight Truck Volumes**	**Electricity Usage**	**Hotel and Motel Rentals**	**Automobile Rentals**	**T10Y3M**	**NFCI**	**INCA**
Mean	46.6	118.1	60.9	140.7	5905.2	1.5	0.0	361467.8
Minimum	3.9	94.9	0.6	100.0	1041.0	-1.7	-1.1	181200.0
Maximum	104.2	140.6	252.2	195.6	15309.0	4.2	5.1	4663250
Standard Deviation	35.4	13.6	70.9	22.7	3269.8	1.2	1.0	224636.2
First date recorded	January 1919	January 2000	January 1929	December 2003	January 1992	January 1982	January 1971	January 1967

Frequency	Monthly	Monthly	Yearly	Monthly	Monthly	Monthly	Monthly	Monthly

Table 1: Summary Statistics

Gross Domestic Product (GDP): Represents the total market value of all goods and services made and provided in a country during a specific period. Additionally, GDP is one of the most important economic indicators of a country's overall economic health and is the primary variable in this paper. Industrial Production Index (IPI): Measures the comparative output of the manufacturing, mining, electric, and gas industries over time. IPI is also occasionally used as a proxy for GDP, as it is reported more frequently (IPI is reported monthly, whereas GDP is reported quarterly) and is comparable in terms of the magnitude of GDP (Industrial Production | FRED). Freight Truck Volumes: The quantity of goods shipped or the space they occupy is used for logistics planning and cost calculations. A decline in freight volumes often indicates reduced demand for goods, signaling potential economic slowdowns (Freight Transportation Services | FRED). Freight volumes can be noticed as a direct proxy for manufacturing activity as industrial productions and GDP require the movement of raw materials and finished goods, leading to freight volumes naturally reflecting broader economic output. Electricity Usage: The Number of dollars of household electricity usage every month. A decrease in electricity use is associated with recessions, which, in turn, are associated with a drop in GDP. The same applies to increases in electricity use (Personal Consumption Expenditures | FRED). Electricity consumption also maintains a strong connection related to GDP and IPI as running factories with machines and equipment require electricity. When industrial production increases, it can be seen that there is an increase in electricity consumption/usage as more machines are operating at larger scales. The same effect is also seen vice versa when there is a decrease in electricity consumption, as less machines are operating during periods of decreased industrial production. Hotel and Motel Rentals: Number of Hotel and Motel rooms booked. A decline in hotel occupancy often reflects reduced travel and leisure activities, which typically occur during economic slowdowns (Producer Price: Hotels and Motels | FRED). Business travel and leisure spending contract during economic downturns and increase during economic "boom" periods, indicating that hotel and motel occupancy serves as a reflection of both consumer and corporate spending health, making it a reliable signal of broader economic conditions (GDP) such as industrial production. Automobile rentals: Number of automobiles booked. A rise in used-car sales often correlates with economic downturns, as consumers delay purchasing new vehicles to find more affordable prices (Retail Sales: Used Car Dealers | FRED). Automobile purchases are one of the largest discretionary purchases made by everyday consumers, and are noticed to delay/ downgrade vehicle purchases or vehicles in general in the event of an economic downturn. Similarly, consumers are also known to upgrade/speed up vehicles and vehicle purchase in an economic expansion. This directly reflects how consumer purchases, especially in the field of automobiles drive industrial production and GDP in general depending on current economic conditions. 10-Year Treasury Constant Maturity Minus 3-Month Treasury Constant Maturity (T10Y3M): The difference between long-term and short-term U.S. Treasury interest rates, commonly used as a leading indicator of economic activity, where a flattening or inversion of the yield curve often signals expectations of slower growth or an upcoming recession (T10Y3M | FRED). When short-term rates exceed long rates, it leads to banks becoming less willing to extend credit for both corporate and consumer use, as banks borrow short and lend long. With worsening credit conditions across the economy, investments and consumer spending leads to the contraction of industrial output and GDP. Chicago Federal National Financial Conditions Index(NFCI): Composite index measuring the overall tightness or looseness of U.S. financial conditions across money, debt, equity, and banking markets (Chicago Fed National Financial Conditions Index (NFCI) | FRED). When financial conditions tighten, households and firms lose wealth while credit becomes harder to access, which forces them to cut back on spending and investment, ultimately slowing industrial production. The opposite is also true, as looser financial conditions expand credit access and boost asset values, encouraging spending and investment that drives industrial output higher. Initial unemployment claims (ISCA): An index that captures the number of workers who file for unemployment insurance for the first time following job separation (Initial Claims (ICSA) | FRED | St. Louis Fed). When workers file for unemployment at increasing rates, it signals that businesses are cutting jobs, leaving households with less income to spend, which directly reduces the demand that drives industrial production. The opposite is also true, as declining claims signal a healthy labor market with strong consumer spending power To formally assess the predictive power of each economic variable, the analysis proceeds by applying a univariate regression with the Industrial Production Index (IPI) as the response variable.

METHODOLOGY

To assess the sole explanatory power of each variable, we run separate univariate regressions of IPI on each indicator. After estimating each regression, we record the coefficient of determination (R^2), which measures the proportion of IPI's variance accounted for by that single variable. The baseline univariate model is:

$$Y = \alpha + \boldsymbol{\beta} X \tag{3}$$

In equation (3), Y is the dependent variable, X is the independent variable, α(Alpha) is the intercept, and **β**(Beta) is the regression coefficient or slope. In Figure 1, a level regression model is used, with the variables in their original, untransformed form. The highest correlation, based on pure, unchanged data, is between electricity usage, and the least correlated data are freight truck volumes.

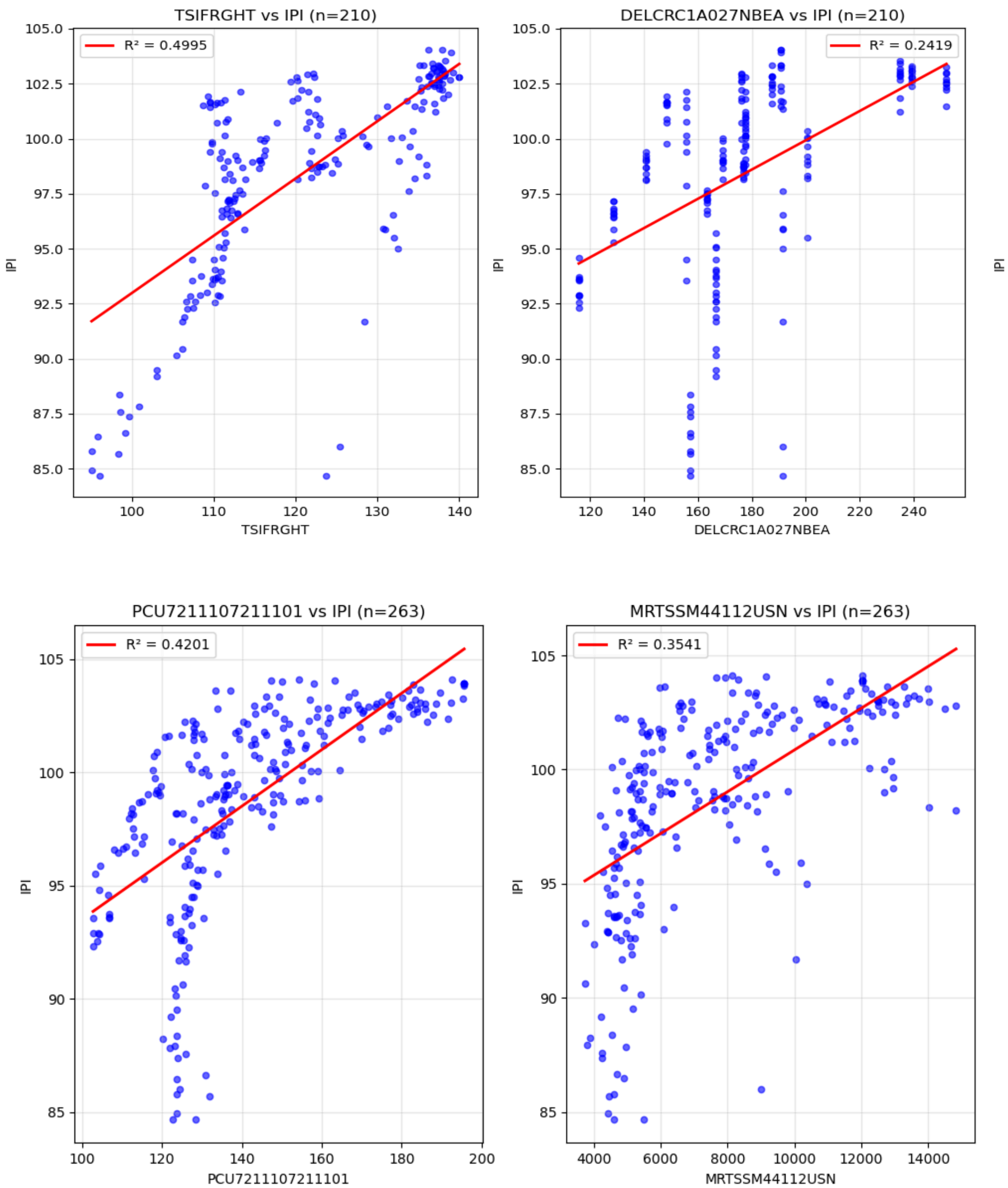
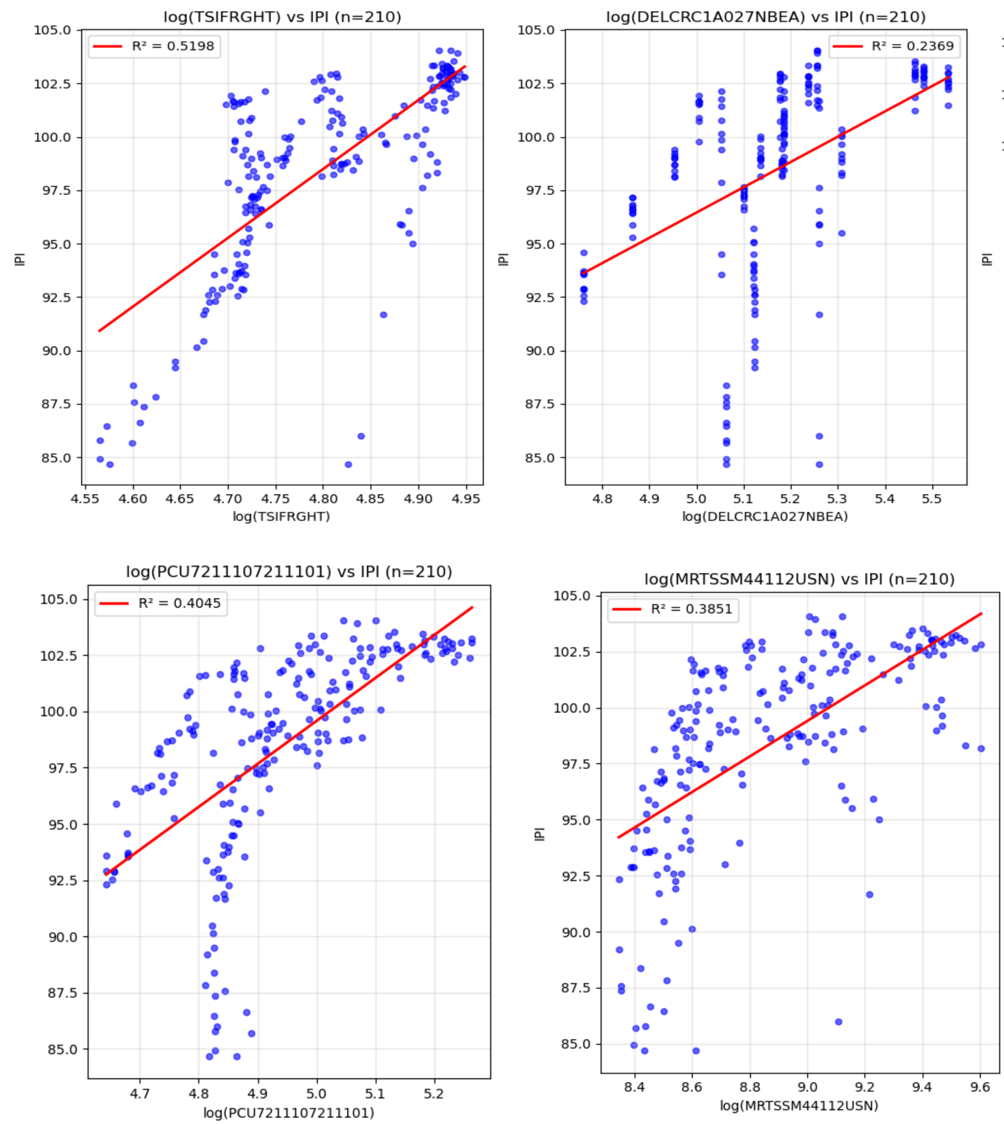


Figure 1: Preliminary Univariate Regression Models

A notable pattern in this data is that the data is not filtered to maximize correlation, reflecting the nature of working with raw datasets. In practice, datasets in general are not perfect, so there will always be outliers or unusual patterns. Outliers in a dataset can substantially alter statistics, including the mean, standard deviation, and, in this case, R^2, by making the data less connected and correlated with IPI. In addition, relationships in the real world are not always linear, suggesting the exploration of other nonlinear models. To address this, a second model is estimated in which the variables are transformed. A log-log regression model transforms both the dependent and independent variables by taking their natural logarithms, giving coefficients that represent elasticities. A 1% increase in the independent variable leads to a **β**(Beta)% change in the dependent variable. Log-log regression models are used to linearize relationships that are usually exponential or multiplicative. By applying logarithmic transformations to both variables, the model stabilizes the variance and reduces skewness, reducing the influence of extreme values and enabling a more precise assessment of the relationship between variables. Figure 2 highlights the log-log regression model on the variables.



Figure

Figure 2: Preliminary Univariate Log-Log Regression Models

From using the log-log regression model on the variables, it can be concluded that each of the variables went up in their coefficient of determination value(R^2), where visually all the data points achieve a more linear trend, with a greater number of data points aligning with the general shape of the line of best fit. However, it is worth noting that a log-log regression model does not eliminate all outliers; rather, it reduces their impact on the estimated coefficients. To evaluate whether the relationships in the model reflect logical effects rather than random variation, statistical significance is examined. Statistical significance measures likeliness of whether an estimated relationship is to occur by chance under the null or given hypothesis, which in this case is that the variables will have no effect. In this paper, coefficients are considered statistically significant if they differ significantly from zero at conventional significance levels, indicating a structured relationship between the explanatory variables and the industrial production index (IPI). Thus far, two models have been seen in this paper: Figure 1 and Figure 2, and across all economic variables examined in preliminary univariate analyses, statistical significance increased overall. While some variables changed more than others, all variables increased their independent contribution to IPI.

Multivariate regression analyses were also performed on these variables in combination with IPI, yielding less accurate results than univariate regression. Multivariate regression estimates and analyzes relationships among multiple independent and dependent variables. A multivariate regression equation can be written as follows

$$Y = \alpha_0 + \boldsymbol{\beta}_1 X_1 + \boldsymbol{\beta}_2 X_2 + \boldsymbol{\beta}_3 X_3 + \ldots + \boldsymbol{\beta}_p X_p \tag{4}$$

In equation (4), Y is the dependent variable, X is the independent variable, α_0 is the y-intercept, and $\boldsymbol{\beta}_1$ to $\boldsymbol{\beta}_p$ are the regression coefficients or slopes. Figure 3 presents the multivariate regression results for the economic variables relative to the actual Industrial Production Index, from 1980 to the most recent reported IPI at the start of this paper.

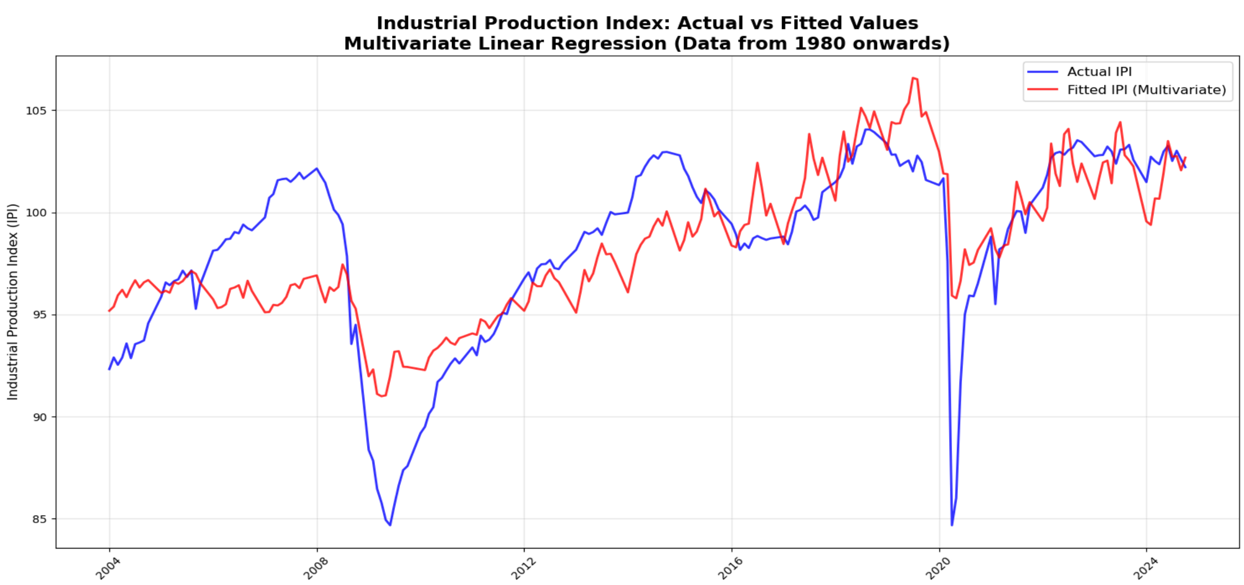


Figure 3: Preliminary Multivariate Regression Fitted IPI vs Actual IPI

The fitted IPI from multivariate regression closely resembles the actual IPI index. Both indices generally follow similar trends, indicating a positive correlation between IPI and the independent variables.

However, there are a few instances in which the fitted and actual IPIs do not match, resulting in discrepancies. These instances include 2008 and 2020 and are attributed to financial and economic downturns. Specifically, the 2008 housing market financial crisis and the COVID-19 pandemic. Both events disrupted the global economy and caused widespread inflation across financial and non-financial sectors. Moreover, while the fitted IPI correctly captures the downturn in the economy during these periods, it understates the effect of these crashes, as evidenced by the dramatic drop in the actual IPI relative to the slight drop in the fitted IPI. Fitted IPI captures normal economic relationships, leading to an incomplete accounting of sudden drops due to panics and crashes and to the averaging out of economic volatility. This builds on the idea shown in Figure 3 that the fitted IPI exhibits repeated cycles of peaks and troughs, whereas the actual IPI shows continuous, steady increases coupled with sudden drops during crashes.

As part of the preliminary multivariate regression calculations, additional residual plots and histograms were created. Figure 4 indicates both the residual plot and the histogram. The residual plot shows the difference between the predicted and actual IPI models. When the residuals are randomly scattered and centered around zero, it suggests that the model fits the data well and that the multivariate regression provides a reasonable representation of the actual IPI trends. The residual histogram shows the distribution of the residuals, which are the differences between the actual and predicted values from the multivariate regression model. Both show the systematic relationship between the missing and remaining data.

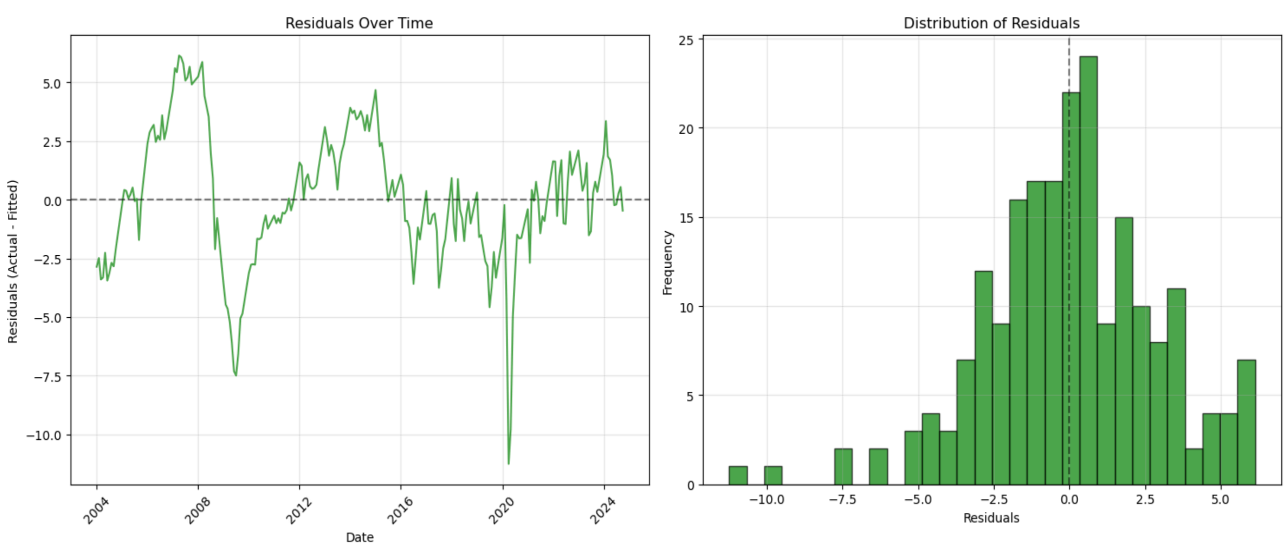


Figure 4: Residual Plot Distribution of Residuals

The first residual plot shows how the predicted IPI values differ from the actual values over time. In this case, the residuals generally stay close to zero, showing that the model fits reasonably well. However, there are downward spikes around major economic events, such as the 2008 financial crisis and the 2020 COVID-19 pandemic. These jumps, again, indicate that the model struggled to accurately capture sudden economic turns. The histogram shows the distribution of the residuals. Here, most residuals cluster near zero, however, a slight lean toward negative values can also be seen, indicating that at that point, the model tends to overpredict IPI more often than underpredict it. Even with that slight skew, the results still show that the model does a decent job at estimating the actual data overall.

FINDINGS AND NEXT STEPS

To better capture the IPI dynamics around economic downturns, as shown in our preliminary multivariate regression model, two additional economic variables can be introduced to further elucidate the existing trends. Specifically, these trends are found in 2008 and 2020 as previously described, due to the 2008 housing market crash and the COVID-19 pandemic, respectively. The first variable is the 10-Year Treasury Constant Maturity minus the 3-Month Treasury Constant Maturity (T10Y3M). This variable is a yield-curve spread that captures market expectations regarding future economic conditions. The yield curve, as a broader concept, illustrates the relationship between interest rates and the time to maturity of debt securities, typically government bonds. Under normal economic conditions, longer-term bonds offer higher yields than short-term bonds to compensate investors for the greater risk and uncertainty associated with longer maturities. Changes in the shape of the yield curve, particularly inversions where short-term rates exceed long-term rates, reflect market expectations of future economic slowdowns and shifts in monetary policy, making the yield curve a widely used leading indicator of economic activity. This measure reflects the spread between long- and short-term interest rates; inversions have historically signaled anticipated economic slowdowns. Additionally, T10Y3M is a leading indicator, an economic variable that changes before the overall economy, providing early signals about future economic conditions or turning points in the business cycle. As a leading indicator, T10Y3M allows for the signaling and prediction of economic slowdowns and downturns before they formally occur. When T10Y3M is included in our multivariate regression model, it improves the model's ability to predict trends observed in IPI, as shown in Figure 4.

In addition to the new yield curve variable, a fundamental step must be implemented to maximize the accuracy of the regression model and indicator. This step introduces a time lag, which is necessary because changes in economic conditions are not reflected immediately in GDP or in the economy as a whole. In this model, a 12-month time lag was used, and it is modeled in the multivariate regression formula as follows in equation (5).

$$Y_t = \alpha_{0t} + \boldsymbol{\beta}_{1t} X_{1t} + \boldsymbol{\beta}_2 X_{2t} + \ldots + \boldsymbol{\beta}_p X_{pt} + \boldsymbol{\beta}_q X_{q,t-12} + \varepsilon_t \tag{5}$$

In this equation, Y_t represents the dependent variable the predicted value, which in this case is GDP, α_{0t} represents the y-intercept or base level of output when all other variables are zero, X_{1t} - X_{pt} represents all other economic variables used in the model, $\boldsymbol{\beta}_{1,}$ - $\boldsymbol{\beta}_q$ represents the strength of how much each variable affects output, $X_{q,t-12}$ represents specifically the yield curve variable(T10Y3M), $\boldsymbol{\beta}_{t-12}$ represents the effect of the yield curve in 12 months, and ε_t represents the error term, which is everything affecting GDP that the regression model fails to capture. Additionally, t denotes time measured in months. For the yield curve variable, the time series is shifted back by 12 periods, corresponding to a 12-month lag. As a result, the model uses the 12-month lagged yield curve rather than a contemporaneous value. Equation (5) also poses a similarity to an equation already mentioned in this study, equation (2), where ΛF_t, the matrix of factor loadings, is nothing more than the notation, $\boldsymbol{\beta}_1 tX_{1t} + \boldsymbol{\beta}_2 X_{2t} + ... + \boldsymbol{\beta}_p X_{pt} + \boldsymbol{\beta}_q X_{q,\, qt-12}$ in a condensed form.

The second new variable used is the Chicago Fed National Financial Conditions Index (NFCI). This variable provides an overall weekly assessment of U.S. financial conditions across the money, debt, and equity markets, as well as the traditional and shadow banking systems. Positive values of the index indicate tighter-than-average financial conditions, while negative values reflect looser-than-average conditions. As a broad measure of financial stress, the NFCI captures changes in credit availability and risk that move closely with current economic conditions. NFCI aids in developing an indicator because it is a contemporaneous indicator, meaning it moves with current financial and economic conditions. Although it is not a leading indicator, it still contains attributes that are useful for predicting future economic crises and downturns.

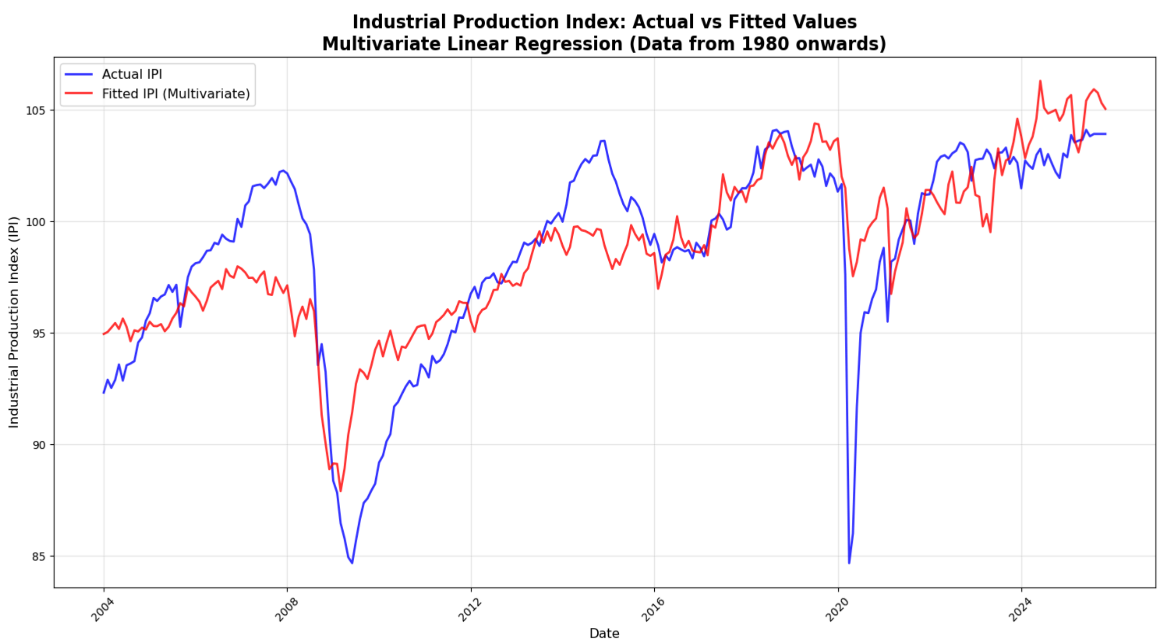


Figure 5: Updated Multivariate Regression model including T10Y3M

Figure 5 illustrates the relationship between actual and fitted Industrial Production Index values from the updated multivariate regression model that includes the yield curve spread. The model closely tracks long-run movements in industrial activity and captures major cyclical fluctuations. The inclusion of T10Y3M and NFCI improves the model's responsiveness to changing financial conditions, particularly ahead of economic downturns. This improvement is consistent with the theoretical framework, as financial markets typically include expectations regarding future economic conditions before its appearance in industrial production. By using this leading indicator, with an indicator of financial stress (NFCI), the updated model now captures predictions and real-time economic conditions that cannot be fully seen in real-activity variables alone. Although the model smooths extreme shocks and slightly underestimates the depth of sudden contractions, it demonstrates strong explanatory power, as reflected by an 80.14% correlation between fitted and actual IPI. Additionally, updated residual plot distributions were created to better illustrate trends in the presence versus absence of data, as shown in Figure 6.

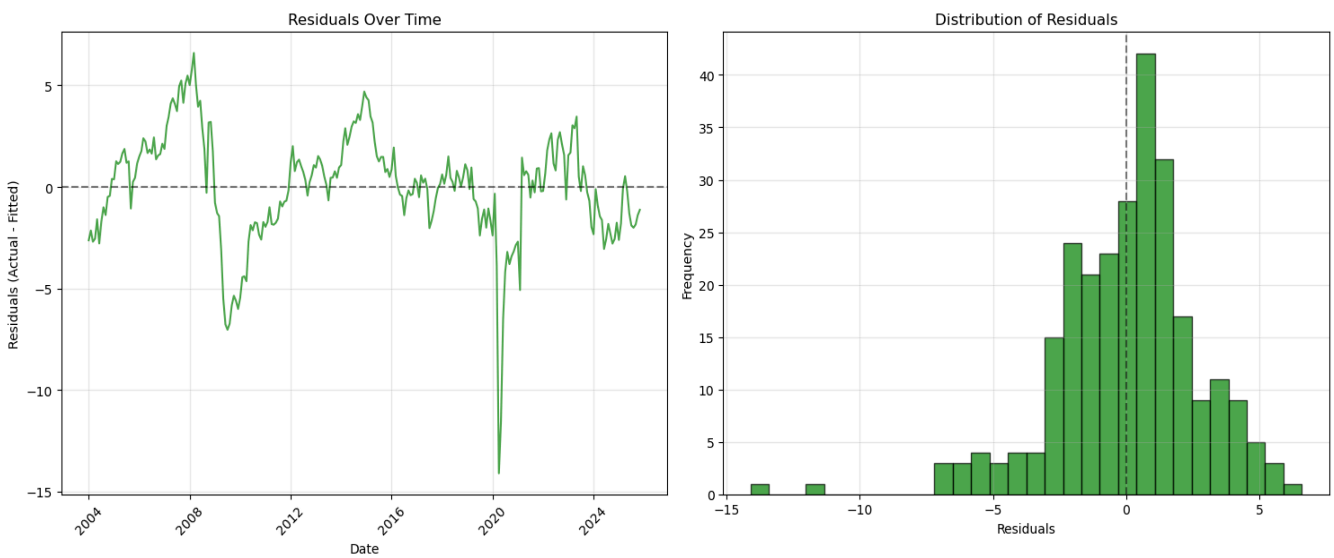


Figure 6: Updated Residual Plot Distribution of Residuals

It is evident that the new economic variables, T10Y3M and NFCI, have significantly affected the model's accuracy, compared with the preliminary residuals over time graph. In some areas, particularly during 2004 - 2010, there is a noticeable decrease in drop severity. Furthermore, the distribution of residuals is centered closer to zero, indicating that the updated model is less biased on average. Additionally, the concentration of residuals around the mean suggests fewer extreme prediction errors, reflecting improved model stability after the inclusion of T10Y3M. While some tail behavior persists due to major economic shocks, overall dispersion is lower than in the preliminary specification. Economically, this suggests that financial conditions have information related to future changes in production before they are observed in traditional economic variables, making the model able to better predict future turning points compared to only reacting to them. Compared with Figure 3, the accuracy from 2021 to the present has declined due to the addition of two new economic variables. However, this scenario, although unfavorable, indicates that the inverse functions of T10Y3M and NFCI are more accurate and predictable than the original IPI in the first half, whereas the second half is less accurate. To address this, one additional economic variable is included in the multivariate regression model. This variable is the initial unemployment claims (ISCA) and is reported when an individual who is unemployed files a claim after being released by a previous employer. A claim in this case refers to an individual's formal request to determine eligibility for Unemployment Insurance benefits after a loss of employment. Continued unemployment insurance claims are included as a measure of contemporaneous labor market stress, which captures real-time economic deterioration. Increases in jobless claims reflect rising unemployment and income disruptions, which directly constrain household consumption and reduce aggregate demand, making it an excellent variable for predicting recessions and, more specifically, declines in GDP. This variable will enable us to further examine and refine the composite indicator (Fitted IPI) and improve its accuracy relative to previously used economic variables. The results of adding this new economic variable to the composite indicator are shown in Figure 7, the final version of this multiple regression-based indicator.

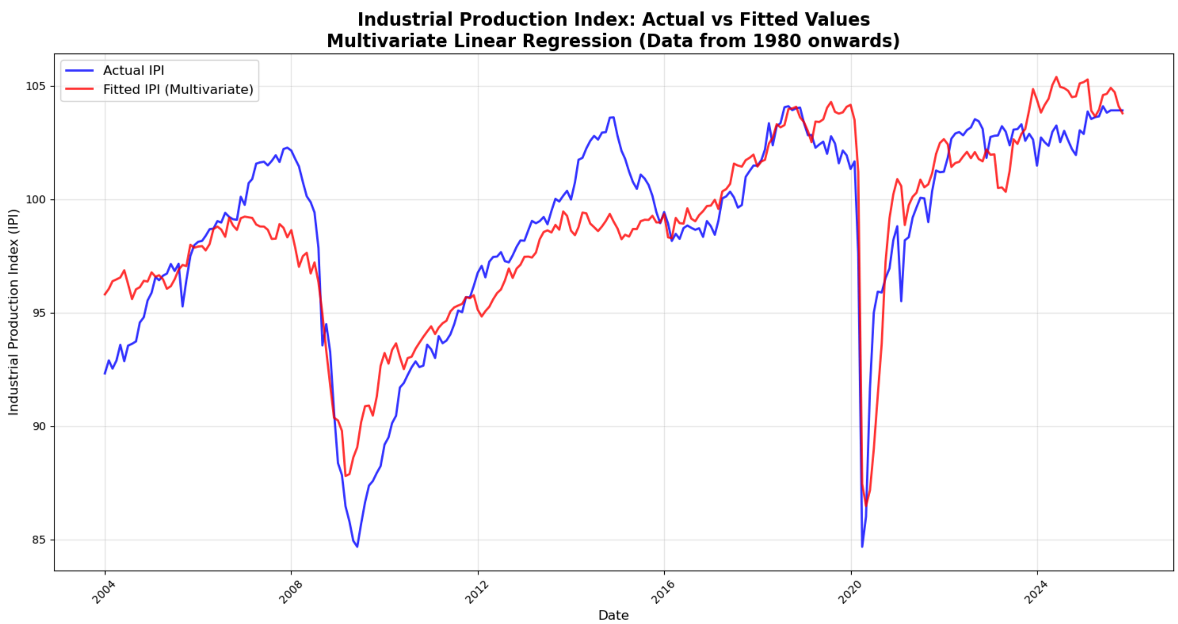


Figure 7: Final Multivariate Regression model including ICSA

Adding ISCA to the multivariate regression model increases the model's accuracy from 80.14% to 89.40%, a 9.26% improvement. This improvement can be seen due to one main reason. Labor market conditions adjust accordingly after changes occur in production and financial markets. When firms start to reduce employment, an increase/rise in unemployment claims indicate that there is an economic downturn that is large enough to influence household income and aggregate demand. Beyond statistics, the model's physical appearance strikingly imitates the indicator's appearance, capturing both general and specific trends. In addition, Figure 8 shows that both a residual plot and a distribution of residuals were created to support the statistical assumptions of this newly improved indicator.

Figure 8: Final Residual Plot Distribution of Residuals

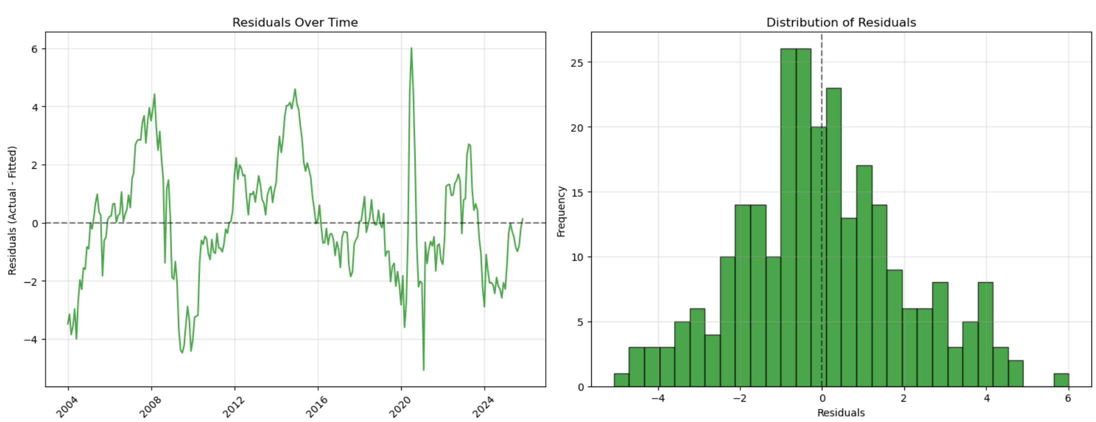
The inclusion of the labor market variable, jobless claims, has significantly improved the model's accuracy relative to the preliminary residuals-over-time specification. In particular, during periods of sharp economic contraction, the severity of residual deviations declines, indicating improved fit once labor market stress is incorporated. Furthermore, the distribution of residuals is centered closer to zero, suggesting that the updated model is less biased on average. The increased concentration of residuals around the mean implies fewer extreme prediction errors, reflecting enhanced model stability following the inclusion of jobless claims. While some tail behavior remains due to major economic shocks, overall residual dispersion is reduced relative to the preliminary specification. This improvement indicates how each category of variables, real-activity, financial stress, and labor market captures specific information of economic activity to the model that positively impacts it. Real-activity variables capture current industrial production, financial variables indicate early signals of changing economic conditions, and labor market variables double check when these changing conditions have spread to the broader economy through gradual rises or drops.

APPLYING THE NOWCAST MODEL

In addition to assessing fit, the model described above can be used to generate real-time nowcasts of GDP. For example, using the following model.

$$Y_t = \alpha_{0t} + \boldsymbol{\beta}_{1t}X_{1t} + \boldsymbol{\beta}_2X_{2t} + \ldots + \boldsymbol{\beta}_pX_{pt} + \boldsymbol{\beta}_qX_{q,t-12} + \varepsilon_t \quad (6)$$

Assume one is interested in knowing the approximate value of Industrial Production or GDP in the United States of America for the first quarter of 2026, but the index and data have yet to be released. However,



we have access to the following data and estimated parameters $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$ ...., $\boldsymbol{\beta}_t$

Variable	**Last Recorded Value**	**Type of Transformation**	**Estimated Coefficient**
Intercept	-	-	29.32
Freight truck volume	137.9	Levels	0.069
Hotel and motel rentals	186.4	Logs	10.51
Electricity Usage	252.24	Levels	-0.040
Automobile rentals	11570	Logs	2.04
Unemployment claims	1872000	Levels	0.000001
Yield curve	-0.62	Levels	-0.859
NFCI	-0.4415	Levels	2.33

Table 2: Application Regression Calculations

By applying the estimated coefficients to the most recent high-frequency indicators, the model yields an estimated Industrial Production Index (IPI) of 104.15 for March 2026. Given that the actual IPI value for March 2025 was 103.54, this corresponds to an implied annual growth rate of approximately 0.6%. The main usefulness of this methodology lies in its ability to nowcast economic activity before official data releases, which are often subject to publication lags and subsequent revisions. Individuals such as policymakers, investors, and researchers often need timely assessments of economic conditions, particularly during periods of heightened uncertainty. By using alternative indicators such as the ones used in this paper, the model gives a more immediate signal of underlying economic momentum than traditional quarterly or monthly aggregates already made. Additionally, this framework allows for scenario analysis. Scenario analysis is the process of evaluating changes in economic conditions that affect predicted output while holding the model structure constant and varying one or more variables. In the model, IPI is shown as a linear function and each variable's estimated coefficient represents how much of that indicator aids to IPI, conditional on all others. This structure allows researchers to simulate hypothetical economic environments and use it for many purposes such as recessions and crashes.

LIMITATIONS

Despite its encouraging results, this study is subject to several limitations that should be considered when interpreting the findings. Firstly, this composite indicator relies on a relatively small set of publicly available variables reported at monthly or annual frequencies. In contrast, many institutional nowcasting frameworks incorporate large-scale datasets, mixed-frequency inputs, and advanced machine-learning or state-space methods that can capture more complex, nonlinear economic dynamics at daily or intra-daily frequencies. Moreover, all data used in this study are sourced from publicly accessible Federal Reserve Bank of St. Louis databases, whereas more sophisticated models often integrate proprietary, high-frequency datasets obtained from paid or private sources. Additionally, the variables are restricted to the United States of America and its economy, which means that this indicator cannot be used to identify economic downturns in other countries. Finally, only linear or log specifications were applied to the economic variables; more complex machine learning models can incorporate more complex specifications (e.g., taking the sine or cosine of the variables).

Future extensions of this work could use the same variables from the dataset at higher frequencies, drawing on proprietary data sources, thereby enabling the integration of more advanced data into this indicator and improving its precision in forecasting future economic events. Additionally, more complex specifications could be applied to the variables to increase statistical significance and reduce outliers. Data from other countries could also be considered if the indicator is needed or used for other countries' economies. The proposed framework is not intended to compete with institutional nowcasting systems, but rather to demonstrate that a transparent and interpretable approach can still generate informative real-time estimates of industrial production using a limited set of economically meaningful indicators.

CONCLUSIONS

In this paper, univariate and multivariate regression models are applied to develop a new multivariate indicator of GDP (Gross Domestic Product) that can indicate future recessions in the economy, using a variety of non-traditional economic data. Using a unique dataset from the Federal Reserve Bank of St. Louis (FRED), a highly accurate indicator that showed 89.40% similarity to the Industrial Production Index (IPI), a substitute for GDP due to its frequency of reporting. The design behind this indicator can be traced back to a standard linear regression formula, $Y_t = \alpha_{0t} + \boldsymbol{\beta}_{1t}X_{1t} + \boldsymbol{\beta}_{2}X_{2t} + \ldots + \boldsymbol{\beta}_{p}X_{pt}$, which compiles a series of single univariate regressions on each variable to create a composite and final value. The findings of this paper stem and support the argument that GDP can be accurately nowcasted through the use of a combination of different non-traditional indicators. Through the use of real-activity, financial, and labor market non-traditional variables, this model accurately captures different dimensions of economic activity that is associated with GDP. This paper sets the groundwork for future investigations on the use of alternative data to predict both financial and economic trends, particularly in the context of recession detection and real-time macroeconomic nowcasting.

ACKNOWLEDGEMENTS

I would like to thank my mentor, Mr. Vicente Garcia, for his endless help, support, and guidance with my research and analysis into regression based analysis, macroeconomic forecasting and nowcasting, and the use of multidimensional economic indicators.

REFERENCES

Aliber, R. Z., & Kindleberger, C. P. (2015). *Manias, panics and crashes : a history of financial crises*. Palgrave Macmillan.

Arora, V., & Lieskovsky, J. (2014). *Electricity Use as an Indicator of U.S. Economic Activity*. 17. chrome-extension://dbkidnlfklnjanneifjjojofckpcogcl/pdf-viewer.html?file=https%3A%2F%2Fwww.eia.gov%2Fworkingpapers%2Fpdf%2Felectricity_indicator.pdf

Federal Reserve Bank of Chicago, & Federal Reserve Bank of St. Louis. (1971, January 8). *Chicago Fed National Financial Conditions Index*. FRED, Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/NFCI

Federal Reserve Bank of Cleveland. (2025). *Yield Curve and Predicted GDP Growth* [Review of *Yield Curve and Predicted GDP Growth*]. Clevelandfed.Org; Federal Reserve Bank of Cleveland. https://www.clevelandfed.org/indicators-and-data/yield-curve-and-predicted-gdp-growth

Federal Reserve Bank of St. Louis. (1982, January 4). *10-Year Treasury Constant Maturity Minus 3-Month Treasury Constant Maturity*. FRED, Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/T10Y3M

Federal Reserve Bank of St. Louis. (2020). *Initial Claims*. Stlouisfed.org. https://fred.stlouisfed.org/series/ICSA

Federal Reserve Bank of St. Louis. (2024). *Personal consumption expenditures: Services: Household utilities: Electricity*. Stlouisfed.org. https://fred.stlouisfed.org/series/DELCRC1A027NBEA

Federal Reserve Bank of St. Louis. (2025a). *Industrial Production Index*. Stlouisfed.org. https://fred.stlouisfed.org/series/INDPRO

Federal Reserve Bank of St. Louis. (2025b). *Producer Price Index by Industry: Hotels and Motels, Except Casino Hotels: Luxury and Resort Hotels Guestroom Rental*. Stlouisfed.org. https://fred.stlouisfed.org/series/PCU721110721110103

Giannone, D., Reichlin, L., & Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, *55*(4), 665–676. https://doi.org/10.1016/j.jmoneco.2008.05.010

Gillet, M., & Brave, S. A. (2024). What Does the NFCI Tell Us About Future Economic Growth? - Federal Reserve Bank of Chicago [Review of What Does the NFCI Tell Us About Future Economic Growth? - Federal Reserve Bank of Chicago]. ChicagoFed.Org. https://www.chicagofed.org/publications/chicago-fed-insights/2024/nfci-future-economic-growth

Hornstein, A. (2023). Unemployment Changes as Recession Indicators. Economic Brief, 23(13). https://www.richmondfed.org/publications/research/economic_brief/2023/eb_23-13

Mandelbaum, R., & Woodworth, R. M. (2024). U.S. Hotel Profitability During Recessions [Review of U.S. Hotel Profitability During Recessions]. Hotelexecutive.Com. https://www.hotelexecutive.com/newswire/19045/us-hotel-profitability-during-recessions

U.S. Bureau of Transportation Statistics, & Federal Reserve Bank of St. Louis. (2000, January 1). *Freight Transportation Services Index*. FRED, Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/TSIFRGHT

University of Michigan. (2013, February 19). GDP Predicts Auto Sales Worldwide [Review of GDP Predicts Auto Sales Worldwide]. University of Michigan News. https://news.umich.edu/gdp-predicts-auto-sales-worldwide/

U.S. Census Bureau, & Federal Reserve Bank of St. Louis. (1992, January 1). *Retail Sales: Used Car Dealers*. FRED, Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/MRTSSM44112USN

Yao, V. W., & Lahiri, K. (2006). Economic Indicators for the US Transportation Sector. https://doi.org/10.1016/j.tra.2006.02.006