Building the Census Bureau Index of Economic Activity (IDEA)[*]

Jose Asturias[1]
William R. Bell[2]
Rebecca Hutchinson[3]
Tucker McElroy[2]
Katherine J. Thompson[3]

March 10, 2023

[*] Any opinions and conclusions expressed herein are those of the authors and do not represent the views of the U.S. Census Bureau. All results in this paper use publicly available data from Census Bureau websites. The Census Bureau has reviewed the public domain products for unauthorized disclosure of confidential information and has approved the disclosure avoidance practices applied to this release (Approval ID: CBDRB-FY23-ESMD-010-012).

[1] Center for Economic Studies, Research and Methodology Directorate, U.S. Census Bureau
[2] Research and Methodology Directorate, U.S. Census Bureau
[3] Economic Directorate, U.S. Census Bureau

Abstract

The Census Bureau Index of Economic Activity (IDEA) is constructed from 15 of the Census Bureau's primary monthly economic time series. The index is intended to provide a single time series reflecting, to the extent possible, the variation over time in the whole set of component series. The component series provide monthly measures of activity in retail and wholesale trade, manufacturing, construction, international trade, and business formations. Most of the input series are Principal Federal Economic Indicators. The index is constructed by applying the method of principal components analysis (PCA) to the time series of monthly growth rates of the seasonally adjusted component series, after standardizing the growth rates to series with mean zero and variance 1. Similar PCA approaches have been used for the construction of other economic indices, including the Chicago Fed National Activity Index issued by the Federal Reserve Bank of Chicago, and the Weekly Economic Index issued by the Federal Reserve Bank of New York. While the IDEA is constructed from time series of monthly data, it is calculated and published every business day, and so is updated whenever a new monthly value is released for any of its component series. Since release dates of data values for a given month vary across the component series, with slight variations in the monthly release date for any one component series, updates to the index are frequent. It is unavoidably the case that, at almost all updates, some of the component series lack observations for the current (most recent) data month. To address this situation, component series that are one month behind are predicted (nowcast) for the current index month, using a multivariate autoregressive time series model.

This report discusses the input series to the index, the construction of the index by PCA, and the nowcasting procedure used. The report then examines some properties of the index and its relation to quarterly U.S. Gross Domestic Product and to some monthly non-Census Bureau economic indicators.

1 Introduction

The Census Bureau Index of Economic Activity (hereafter "the index") is constructed from 15 of the Census Bureau's primary economic indicators intended to provide, to the extent possible, a single time series reflecting the variation over time in the whole set of component series. Thus, the index is designed to provide an "at-a-glance" snapshot of the combined movement of these different series. The component series, listed and discussed in Section 2, provide monthly measures of activity in retail and wholesale trade, manufacturing, construction, international trade, and business formations.

The index is constructed by applying the method of principal components analysis (PCA) to the time series of monthly growth rates of the seasonally adjusted component series (standardized month-to-month percentage changes as described in Section 2.1). The PCA approach has been used for the construction of other economic indices, including the Chicago Fed National Activity Index (see Brave, 2008), which is issued by the Federal Reserve Bank of Chicago, and the Weekly Economic Index (see Lewis, Mertens, Stock, and Trivedi, 2022), which is issued by the Federal Reserve Bank of New York.[1] The PCA approach gives weights to be applied to each of the component series. PCA determines these weights (subject to a scaling constraint) so the resulting index is the particular linear combination that reflects the maximum amount of variation over time of the set of component series compared to any other linear combination. To aid interpretation of index movements, the index is then rescaled by subtracting its mean and dividing by its standard deviation to produce the final version of the index. Note that we standardize the growth rates of the inputs to put the component series on a comparable scale to prevent components whose growth rates inherently fluctuate the most from dominating the contributions to the index. Details of these calculations are discussed in Section 4.

We refer to the index as a monthly index since it is constructed from time series of monthly growth rates. However, the index is calculated every business day using the latest values of the components drawn from the Census Bureau's Application Programming Interface (API).[2] The index is thus updated whenever a new value is released for any of its component series. Since release dates of data values for a given month vary across the component series, with slight variations in the monthly release date for any one component series, updates to the index are frequent. It is unavoidably the case that, at almost all updates, some of the component series lack observations for the current (most recent) data month. To address this situation, component series that are one month behind are predicted (nowcast) for the current

[1] More detailed background information about the Weekly Economic Index (WEI) can be found at https://www.newyorkfed.org/research/policy/weekly-economic-index#/, and more information about the Chicago Fed National Activity Index (CFNAI) can be found at www.chicagofed.org/cfnai.

[2] More information on the Census Bureau's API is available at https://www.census.gov/data/developers/data-sets/economic-indicators.html.

index month, using a multivariate autoregressive time series model. The update schedule and the nowcasting procedure are also discussed in Section 3.

Section 4 presents the index for the period from August of 2004 through September of 2022, the period for which all component series provided data as of this writing. Examining the graph of the index reveals large movements due to the Great Recession and the COVID-19 pandemic, as would be expected. Further analyses compare the index to the individual component series and examine the weights which reflect the contributions to the index of the various component series. Section 4 also compares the index to economic indicators produced by other federal agencies: the industrial production index of the Federal Reserve Board, non-farm employment and the unemployment rate from the Bureau of Labor Statistics, and quarterly gross domestic product from the Bureau of Economic Analysis.

Section 5 examines sensitivity of the index to some alternative choices of the time frame over which PCA is applied to determine the weights on the component series. This analysis was motivated by the very large deviations of the input series growth rates, and hence the index, during the periods of the Great Recession and the COVID-19 pandemic. We find broadly similar results whether the Great Recession and the COVID-19 pandemic are included in the time frame of the data used to determine the weights.

Section 6 provides conclusions and discusses some possible directions for future work. Appendices to the report provide technical details on PCA and on the nowcasting approach, as well as graphs displaying the index jointly with each of the component series.

2 Data

We now describe the data series that we used to construct the index and discuss their release dates.

2.1 Monthly Census Economic Indicators Used

We apply PCA to monthly economic indicators produced in the Economic Directorate of the Census Bureau. Economic indicator surveys are a hallmark of official statistics programs in National Statistical Institutes (NSI). High frequency collections, monthly or quarterly, along with the publication of aggregated estimates for a small number of variables, characterize these programs. Their measures are inputs to National Accounts, to Price Indices, and to Gross Domestic Products. Policymakers, business communities, and economists monitor these measures to assess the state of the economy. Consequently, timeliness is paramount. It is a common practice to produce a preliminary estimate early in the collection cycle ("early indicators"), with revised estimates for the same time-period published in succeeding months (two to three revisions, depending on the program). In many cases, differences between the corresponding indicators' estimates are primarily due to changes in the underlying data (e.g., amended reports, late reports). Consequently, revisions between corresponding preliminary and final estimates are unavoidable. In many, but not all, cases, differences between the first and second revised estimates are trivial. That said, the media reports on changes between these preliminary indicators and their prior period estimates, and data users monitor the direction of any change and act accordingly.

Table 1 lists the monthly economic indicators that we used in our analysis. Criteria for inclusion in the initial version of the index were twofold: (1) the series is widely used and frequently monitored by public data users and (2) the series must be produced in-house. The first requirement ensures relevance. An advantage to the second criterion is that for movements of the index driven by movements in particular input series, Census Bureau analysts who are subject matter experts would be available who could potentially explain the movements in the input series. As the index matures, it may be possible to expand the component series to include measures from other Census Bureau programs.

A byproduct of this purposive series selection is that most of the series included in the index are Principal Federal Economic Indicators (PFEIs). An advantage of PFEIs is that they are widely watched economic indicators that tend to be correlated with other measures of economic activity. Asturias et al. (2021), for example, show that these PFEIs are positively correlated with nonfarm employment.

We focused on the primary data series for each economic data program, e.g., Retail Trade and Food Services Monthly Sales from the Advance Monthly Retail Trade Survey, Wholesale Trade Inventories from the Advance Economic Indicator Report. Of the series that are not PFEIs, the Business Formation Statistics (BFS) is included because, as documented by Asturias et al. (2021), its Business Application series tends to be a leading economic indicator.

As inputs to our index, we use growth rates of the seasonally adjusted series listed in Table 1. More specifically, if Y_t is the seasonally adjusted value for month t of one of the series listed, then the corresponding growth rate X_t is defined as $X_t = \log Y_t - \log Y_{t-1}$. With this definition, $100 \times X_t$ is a close approximation to the month-to-month percentage change, $100 \times (Y_t - Y_{t-1})/Y_{t-1}$, for percentage changes smaller in magnitude than about 20 percent. The approximation deteriorates for percentage changes larger in magnitude than 20 percent. Graphs of the input growth rates given in Appendix D show that changes exceeding 20 percent in magnitude are rare, occurring in our data only for pandemic period outliers for a few of the input series.

Table 1. Data Series Included in Index

Series	Source
Business Applications	Business Formation Statistics*
Exports of Goods and Services	International Trade: Goods & Services
Housing Units Authorized in Permit-Issuing Places	New Residential Construction
Housing Units Completed	New Residential Construction
Housing Units Started	New Residential Construction
Imports of Goods and Services	International Trade: Goods & Services
Manufacturing Inventories	Manufacturers' Shipments, Inventories, and Orders
Manufacturing Value of New Orders	Manufacturers' Shipments, Inventories, and Orders
New Orders for Durable Goods	Advance Report Durable Goods
New Single-Family Houses for Sale	New Residential Sales
New Single-Family Houses Sold	New Residential Sales
Retail Inventories	Advance Economic Indicator Report*
Retail Trade and Food Services Monthly Sales	Advance Monthly Retail Trade
Total Construction Spending	Construction Spending
Wholesale Trade Inventories	Advance Economic Indicator Report*

* Not a Principal Federal Economic Indicator (PFEI). Revisions to Retail Inventories and Wholesale Trade Inventories are included in the Manufacturing and Trade Inventories and Sales and Monthly Wholesale Trade PFEI releases, respectively.

The earliest date for which data on all the time series listed in Table 1 are available is July 2004. Since we use growth rates of the series as inputs to the index, the earliest date for which we can construct the index is August 2004. As noted in the Introduction, as of this writing, the latest date for which data on all the input series was available is September 2022, so that the longest available time span for which we could compute the index is August 2004 through September 2022. Appendix D provides plots of the input series growth rates and the index over this time span. As can be seen in these plots, almost all of the input time series, and hence the index, show dramatic effects from the Great Recession (around 2008 through 2009) and the COVID-19 pandemic (starting in March or April of 2020). We decided to apply PCA to compute the index weights to the growth rate data from August 2004 through February 2020, thus avoiding the effects of the COVID-19 pandemic but not the Great Recession. The weights are then applied to the full data series from August 2004 to September 2022. Some analyses presented in this report, including the application of PCA to compute the index weights, compare results obtained using restricted time spans of the data to examine possible sensitivity of the results to these effects.

2.2 Varying Release Dates of Input Data

One challenge in creating a monthly index is that the input series data are released at different points in time. Table 2 lists the release dates for the index input series for the data month January 2022 (i.e., data that cover January 2022). For example, Table 2 indicates that the January 2022 value for Business Applications, the first series released covering January 2022, was released on February 14, and that the data for the other input series continued to be released until March 8. The PCA methodology requires the growth rates of all 15 series to compute the index for the month of January. In Section 3.2 we describe the methodology that we use to nowcast (i.e., predict current values of economic data) the missing values for a particular month.

Another important point is that we want the index to incorporate the most recently available data for each given input series. For example, in the case of January 2022 data, an advance estimate of retail inventories was published on February 28, 2022, in the Advance Economic Indicator Report, and included in the computation of the index. On March 16, 2022, a revised retail inventories estimate for January 2022 was published as part of the Monthly Retail Trade Survey. As of March 16, this revised series replaced the advance series in the computation of the index.

Table 2. Initial Release Dates January 2022 Data

Series	Release Dates	Data Month
Business Applications	February 14	January 2022
Retail Trade and Food Services Sales	February 16	January 2022
Housing Units Authorized in Permit-Issuing Places	February 17	January 2022
Housing Units Completed	February 17	January 2022
Housing Units Started	February 17	January 2022
New Single-Family House Sold	February 24	January 2022
New Single-Family Houses for Sales	February 24	January 2022
New Orders for Durable Goods	February 25	January 2022
Retail Trade Inventories	February 28	January 2022
Wholesale Inventories	February 28	January 2022
Total Construction Spending	March 1	January 2022
Manufacturing Value of New Orders	March 3	January 2022
Manufacturing Inventories	March 3	January 2022
Imports of Goods and Services	March 8	January 2022
Exports of Goods and Services	March 8	January 2022
Business Applications	March 9	February 2022

3 Methodology

3.1 Using Principal Components Analysis (PCA) to Construct Weights and the Index

PCA is a dimension reduction technique that, as noted in the Introduction, has been previously applied for the construction of economic indices. The first principal component weights from PCA are applied to each economic indicator to construct the index. Note that with p time series the PCA calculations provide weights for up to p principal components, with each successive component explaining less of the variation in the original set of p time series. While additional principal component series beyond the first could be examined, our objective was to compute a single index series capturing as much of the variation of the input series as possible. For this purpose, the first principal component is the clear choice. See Appendix A for more details about PCA.

We now describe in detail how we prepare the data, calculate the weights, and construct the index. For data preparation, we take the following steps:

- Download the series described in Table 1 from the Census Bureau API using data starting in July 2004.
- Calculate the growth rate of each series by finding the log difference. If Y_t is the value of the seasonally adjusted series in month t, then the growth rate, X_t, is calculated as $X_t = \log Y_t - \log Y_{t-1}$.
- Standardize the growth rates of all the series. The standardized growth rate, x_t^S, is calculated as $x_t^S = \frac{X_t - \mu}{\sigma}$, where μ and σ are the mean and standard deviation, respectively, of the non-standardized growth rates from August 2004 to February 2020.

To determine the index weights for each series, we apply PCA to all 15 of the standardized series, x_t^S, over the period August 2004 to February 2020, and use the weights for the first principal component to construct the index. As noted in Section 2.1, we left out the data after February 2020 from the application of PCA to calculate the index weights out of concerns about COVID-19 pandemic effects possibly distorting the results. We do apply the resulting weights to the input data to calculate the index over the full range of data available. In Section 4, we make some comparisons that involve computing the index weights from alternative time frames of the data to examine the sensitivity of the calculations to the effects of the Great Recession and the COVID-19 pandemic.

Finally, to compute the index, we take the following steps:

- Apply the weights from the first principal component to the standardized growth rates, x_t^S of each series.
- Standardize the resulting index using its mean and standard deviation over the period August 2004 to February 2020. Thus, the final index has mean 0 and standard deviation 1 over that period.

3.2 Nowcasting

As described in Section 2.2, at a given point in time input data for the most recent month may not be available for all the input series. For example, on February 14, 2022, Business Applications was the first input series released for the month of January 2022. Thus, at that point in time we had the values for January for just 1 of the 15 series in the index. Rather than waiting for the release of January values for all the remaining input series to compute the index for January, we estimate the January index value by imputing the missing values. Our methodology to do this employs a nowcasting framework based upon a fitted Vector AutoRegressive (VAR) model. Specifically, the growth rates of the input time series, denoted by $\mathbf{X}_t$, are modeled over the time span of August 2004 through February 2020 (avoiding COVID-19 pandemic effects but not the Great Recession as noted in Section 2.1) with a first order VAR model (VAR(1)), being fitted to enforce a degree of sparsity in the coefficient matrix as described in McElroy and Trimbur (2022). A sparsity threshold of 50% is used for the *p*-value of whether the likelihood changes significantly when a new zero is enforced in the estimation, as discussed in Section 2.2 of McElroy and Trimbur (2022); also see McElroy and Findley (2015).

Over the span of data used for estimating the model, all the data is fully observed, and so the VAR(1) model can be fitted without imputation. But we now wish to impute the missing values in the most recent $\mathbf{X}_t$ by a linear projection (i.e., a conditional expectation presuming the data process is Gaussian) that utilizes the observed past values of $\mathbf{X}$ as well as the observed portion of $\mathbf{X}_t$. (For the VAR(1) model, the observed past of $\boldsymbol{X}$ as of time t can be condensed to just $\boldsymbol{X}_{t-1}$.) We refer to this type of projection as a nowcast, since it uses available present data in addition to past data; the projection combines regular one-step ahead forecasts with the partial contemporaneous information available in $\mathbf{X}_t$. The linear projection formula, which is given in Appendix B, is applied using estimated means, variances, and covariances obtained from the fitted VAR(1) model. The nowcast formula given in Appendix B is the particular case for the VAR(1) model of more general results given in McElroy et al. (2021), which also describes computational details for a general VAR process. The R function “mvar_forecast.r” used in the code base is a direct implementation of these more general results.

Applying these formulas, we calculate the imputations of the missing values of $\mathbf{X}_t$, thereby obtaining a completed data set. Now the index can be calculated at time t in the usual manner – the input time series of growth rates are standardized, the linear combination using PCA weights yields the initial index, which is then standardized to have mean zero and variance one over the period August 2004 to February 2020.

4 Results

In this section, we present the results of applying the PCA methodology to the monthly Census economic indicators discussed in Section 2. We also explore the properties of the index, e.g., we study the correlations with other widely watched economic indicators.

4.1 Weights

The index weights assigned to each economic indicator can be found in the second column of Table 3. We see that there is a wide range of values for the weights, from 0.37 for manufacturing new orders to 0.02 for business applications. The series with the highest PCA weights are related to manufacturing (new orders, new orders for durable goods, and manufacturing inventories), international trade (exports and imports of goods and services), retail trade (sales and inventories), and wholesale trade (inventories). Table 4 reports the weights assigned to the series belonging to each activity (column 2 of Table 3 indicates the activity corresponding to each series). We find that the series related to the manufacturing sector received 28 percent of the weights, followed by the series relating to construction receiving 27 percent of the total weight. International trade, retail, and wholesale receive 20 percent, 17 percent, and 8 percent of the total weights.

Note that we could have dropped the series in Table 3 that were given a low weight by the PCA without significantly changing the index. We chose not to drop any of these inputs so that weights for these inputs could be recomputed at a future date. Thus, the series with small weights added little information at present but might in the future as the economy evolves.

Table 3. PCA Weights Assigned to each Data Series

Series	Activity	Abbreviation	PCA Weight
Manufacturing Value of New Orders	Manufacturing	manu orders	0.37
Exports of Goods and Services	International trade	exports	0.36
Manufacturing Inventories	Manufacturing	manu inv	0.34
Imports of Goods and Services	International trade	imports	0.34
Retail Trade and Food Services Inventories	Retail	retail inv	0.30
Retail Trade and Food Services Monthly Sales	Retail	retail sales	0.28
New Orders for Durable Goods	Manufacturing	durable orders	0.28
Wholesale Trade Inventories	Wholesale trade	wholesale inv	0.28
Housing Units Authorized in Permit-Issuing Places	Construction	permits	0.24
Housing Units Started	Construction	starts	0.22
Total Construction Spending	Construction	constr spend	0.19
New Single-Family Houses for Sale	Construction	houses for sale	0.17
New Single-Family Houses Sold	Construction	houses sold	0.07
Housing Units Completed	Construction	completions	0.05
Business Applications	Business formations	ba	0.02

Table 4. Percentage Contribution by Activity to Total PCA Weights

Activity	Percent of Total Weights
Manufacturing	28
Construction	27
International trade	20
Retail	17
Wholesale trade	8
Business formations	1

To gain a better understanding of what could account for the differences in weights across the economic indicators, Figure 1 shows a scatterplot of the PCA weight on the y-axis, and the mean correlation of the series with all the other input series on the x-axis. To find the mean, we calculate the pairwise correlations between a given indicator and the 14 other indicators that we use for the analysis; we then take the mean over these 14 pairwise correlations. Because the weights are calculated using the growth rates between July 2004 to February 2020, we use the same period to calculate the pairwise correlations. Column 3 of Table 3 lists the labels used for each series in Figure 1.

The figure shows a striking positive correlation between the weight and the mean correlation. For example, if we estimate a linear regression, we find an R^2 of 0.96, indicating a high degree of correlation. Thus, in this application, the PCA finds common patterns among the inputs (reflected by a high mean correlation) and determines weights that reflect these patterns in the index.

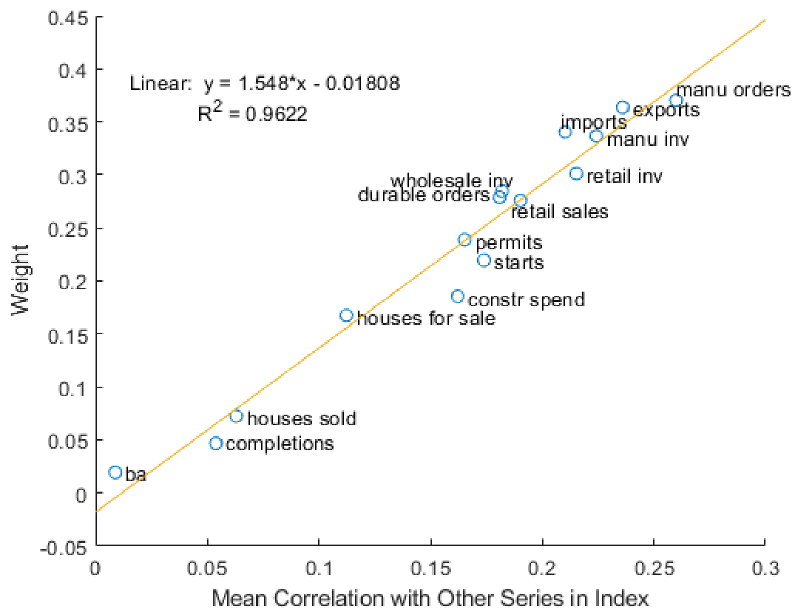


Figure 1. The PCA Weights (in Index) are Plotted Against the Mean Correlation with Other Series. The line is from a least squares regression of the weight against the mean correlation with other series in the index; the results from this regression are reported in the top left side of the figure.

Figure 2 shows the plot of the index through time. As discussed in Section 3.1, we standardize the index using the mean and standard deviation calculated over the time span August 2004 to February 2020 so that the index in Figure 2 has mean 0 and standard deviation 1 over that period. Figure 2 also shows bands indicating one, two and three standard deviations away from the mean of zero. Two features that immediately stand out are the large declines during the Great Recession and the COVID-19 pandemic. For example, during the Great Recession the index declined to –5.3 (December 2008) and at the start of the pandemic (April 2020) the index declined to –11.0.[3] Furthermore, there are large periods of time in

[3] In May 2020 the index rebounded to 1.3, but this did not indicate a full recovery of the index; it was an artifice of the May growth rate coming off a low level due to the enormous drop in April. Similar rebounds can be seen in the growth rates of many economic series that experienced large drops in April 2020.

which we do not see large deviations in the index. For example, during the period January 2010 to February 2020, the index was at least two standard deviations from the mean on only two occasions. These two occasions occurred in March 2011 and July 2014. On both occasions, the index exceeded two for only one month.

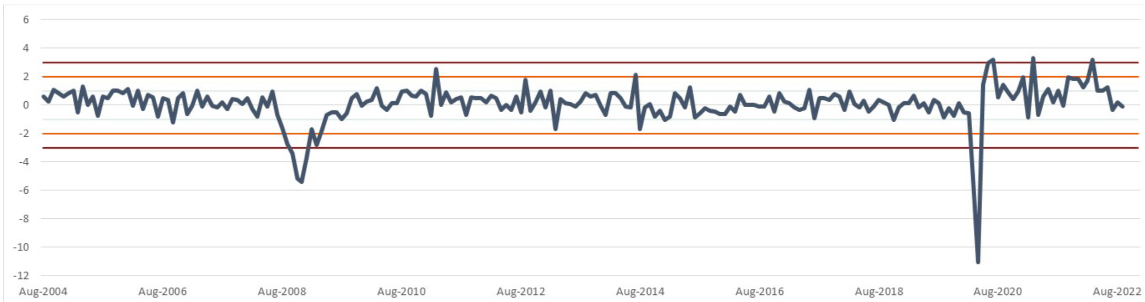


Figure 2. Index Plot (August 2004 to November 2022).

4.2 Comparison of the Index with Other Economic Indicators

To better understand the properties of the index (i.e., how does the index relate to measures that are not in the index?), we compare it to three widely watched monthly economic indicators that are not in the index: the industrial production index (IPI) issued by the Federal Reserve Board; the unemployment rate issued by the Bureau of Labor Statistics (BLS); and nonfarm employment issued by BLS. We also compare it to quarterly Gross Domestic Product (GDP) issued by the Bureau of Economic Analysis (BEA).

Panel (a) of Figure 3 plots the IPI and the index plotted in Figure 2 up to December 2019, which is the period leading up to the COVID-19 pandemic. We first focus on the period before the pandemic because the pandemic saw very large changes, so graphing the data through the pandemic makes it difficult to visually see the patterns before the pandemic. The plots show that both the IPI and the index have large declines during the Great Recession and that the timing of the turning point was similar.[4] Panel (b) shows a similar figure except that we plot the standardized growth rates of nonfarm employment and the index; Panel (c) plots the standardized growth of the unemployment rate and the index. Note that for Panel (c), we use minus the growth of the unemployment rate to make it easier to visually compare with the index (i.e., both the index and minus the unemployment rate fall during recessions). In these figures, we find similar patterns during the Great Recession except that the index has a small lead over both series.[5]

[4] During the Great Recession, the lowest point for the index was December 2008. The lowest growth rate for IPI took place in September 2008 and then the IPI growth rates had another sharp dip in December 2008. This pattern in the IPI can be seen in panel (a) of Figure 3.

[5] During the Great Recession, the lowest point for the index was December 2008 whereas the lowest growth rate for

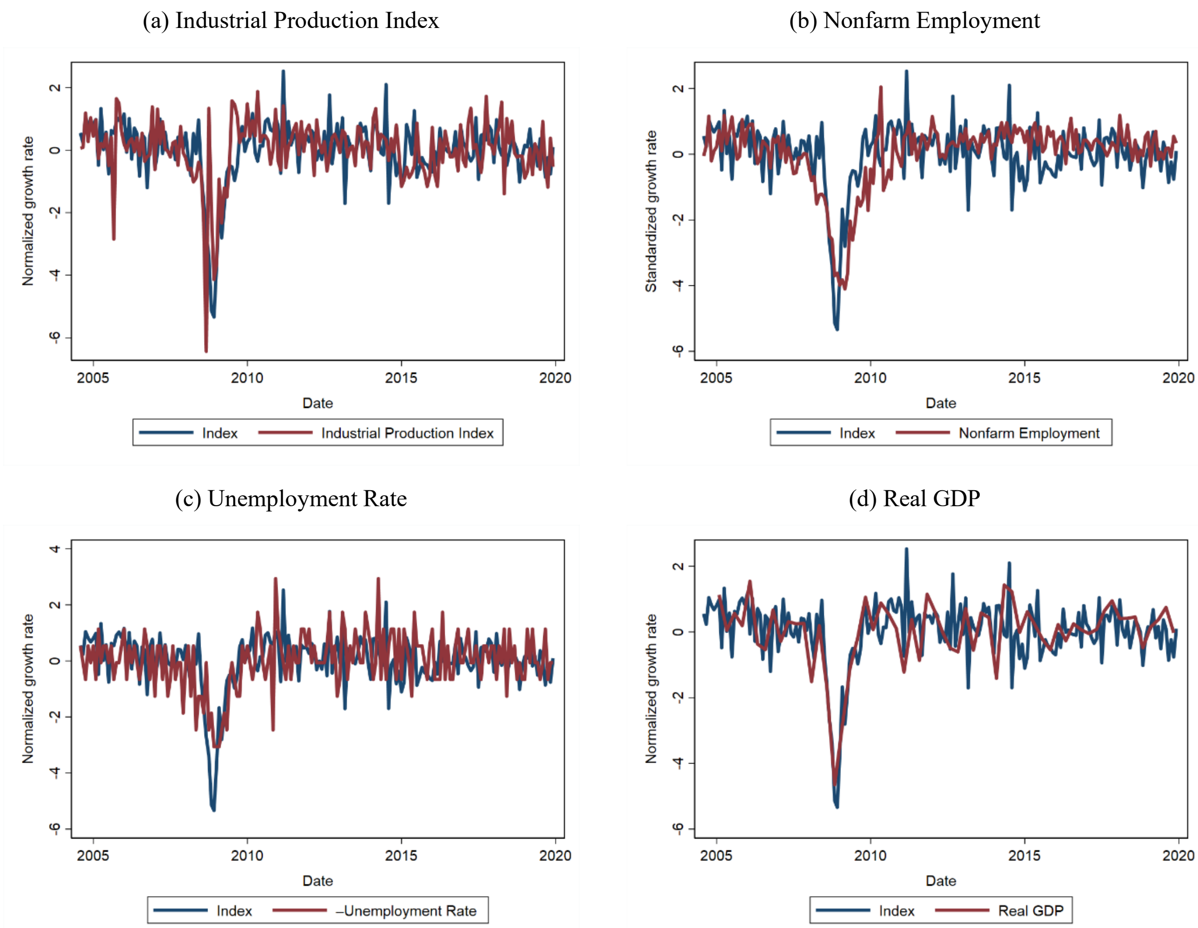


Figure 3: Index Compared to Other Economic Indicators (August 2004 to December 2019).

Panels (a)-(c) of Figure 4 report similar plots as Panels (a)-(c) of Figure 3 except that the starting date is January 2020. We find that in all three cases, both the index and the other time series experienced declines until April 2020 followed by a sharp reversal. One thing to note is that, although the index is consistent with the timing of the decline and subsequent reversal of economic activity, the other indicators experienced much larger declines in April 2020. For example, nonfarm employment saw declines of more than 80 standard deviations away from its long-term mean (as mentioned before, the mean and standard deviation were calculated using data between August 2004 and February 2020).

nonfarm employment took place on March 2009 and the highest growth rates of the unemployment rate took place on December 2008 through February 2009.

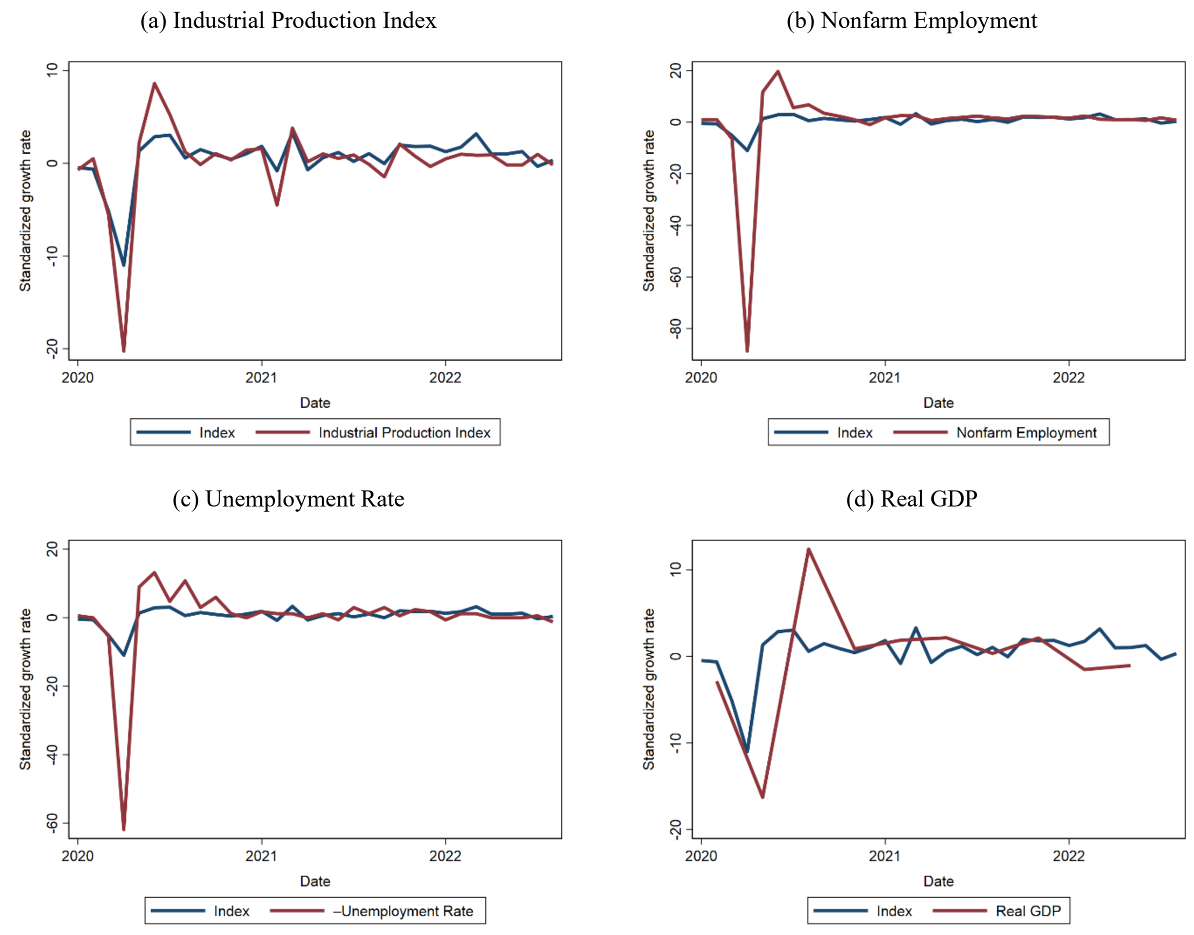


Figure 4. Index Compared to Other Economic Indicators Starting from January 2020 to November 2022.

To study how the index relates to these other economic indicators in a more systematic manner, we calculated the following pairwise correlations:

$$corr(\Delta y_t, x_t),$$

where Δy_t is the growth rate of a target economic indicator at time t, and x_t is the index at time t. Note that the index itself can be thought of as like a growth rate since it is computed as a linear combination of the growth rates of the input series. The results of these calculations can be found in Table 5. Column 2 shows the correlations using data for the whole period. We find that the correlation between the index and IPI, the unemployment rate, and nonfarm employment is 0.75, –0.66, and 0.67, respectively.

Table 5. Correlations between Index and Growth Rates of Other Economic Indicators

Monthly indicator	Entire time span (8/2004 – 8/2022)	Remove pandemic (8/2004 – 2/2020)	Between GR & pandemic (1/2010 – 2/2020)	Post-2010 period (1/2010 – 8/2022)
Industrial production index	0.75	0.52	0.25	0.81
Unemployment rate	–0.66	–0.42	–0.01	–0.74
Nonfarm employment	0.67	0.59	–0.08	0.75

Notes: All results have significance levels of less than 0.01, except for the correlation between the index and the unemployment rate between the Great Recession and the pandemic, and the correlation between the index and nonfarm employment between the Great Recession and the pandemic, which have a significance level of greater than 0.10.

We are also interested in correlations between the index and these other monthly indicators during additional different time periods also reported in Table 5. Column 3 reports the results when we use growth rates up to February 2020, which excludes the COVID-19 pandemic from the analysis. We find moderate declines in the magnitude of the correlations. For example, the correlation between the index and IPI declines to 0.52. Column 4 reports results when we use growth rates for the period between the Great Recession and the COVID-19 pandemic (i.e., January 2010 to February 2019). We find that the correlation between the index and IPI over this time frame is much lower, declining to 0.25, and the index does not have statistically significant correlation with either nonfarm employment or the unemployment rate. Column 5 shows the correlations for the time frame after the Great Recession and including the pandemic are a bit larger in magnitude than those for the full time-span in Column 2. Note that the Great Recession and COVID-19 pandemic periods contribute a substantial share of the overall variation of the index and other series. Thus, the correlations for the series computed from data omitting the Great Recession and the COVID-19 pandemic periods are much lower because they lack these important contributions to variation.

We also conduct a similar set of comparisons between the index and the growth rates of real GDP.[6] Panel (d) of Figures 3 and 4 plot real GDP growth rates and the index for the different time periods that we discussed before. Note that panel (d) in both figures plots the growth rates of real GDP, which is quarterly, and the index, which is monthly; we thus plot the real GDP growth rate in the middle month of each quarter (e.g., the first quarter real GDP growth rate is plotted in February). We find similar patterns as we did when studying the monthly indicators. For example, during the Great Recession, the index and real GDP exhibited their lowest growth rates in the same quarter (i.e., the index bottomed in December of 2008 and GDP saw its bottom in quarter 4 of 2008). Similarly, during the pandemic, the index saw its lowest point in April 2020, which coincides with real GDP having its lowest growth rates in the second quarter of 2020.

[6] For real GDP, we use the series "Real Gross Domestic Product, Chained Dollars" found in Table 1.1.6 of BEA's National Income and Product Accounts (NIPA).

Table 6. Correlations between real GDP growth rates and quarterly index growth rates

	Entire time span (Q1 2005 – Q2 2022)	Remove pandemic (Q1 2005 – Q4 2019)	Between GR & pandemic (Q1 2010 – Q4 2019)	Post 2010 period (Q1 2010 – Q2 2022)
Real GDP	0.58	0.74	–0.03	0.59

Notes: All results have significance levels of less than 0.01, except for the correlation between the Great Recession and the pandemic, which have a significance level of greater than 0.10.

Table 6 reports the correlations between real GDP growth rates and the index over the different time spans used in Table 5. To compute these correlations, it is necessary to convert the monthly index to a quarterly index. Thus, we sum the index over all the months in a particular quarter. For example, the growth in quarter one of a given year would be the sum of the index for the months of January, February, and March of that year. We find that for the entire time span, the correlation is 0.58 and if we remove the pandemic the correlation rises to 0.74. If we focus on the period between the Great Recession and the pandemic, the correlation becomes statistically insignificant. The correlation using the time period after the Great Recession is very close to that for the full timespan.

In Appendix C, we report the results of the cross-correlations of the growth rates of the four series considered in this subsection and lagged values of the index. This analysis provides systematic evidence of whether the index leads or lags these series. We find the results of the cross-correlations are consistent with the analysis in this subsection in which we focus on the turning points during the Great Recession and COVID-19 pandemic.

5 Sensitivity Analysis of the Index to Computing the PCA Weights from Alternative Time Frames of the Data

When we applied PCA to our 15 input time series described in Section 2 to compute index weights we used data for August 2004 to February 2020 to avoid possibly distorting effects of the COVID-19 pandemic. We also could have chosen a time frame for the PCA to avoid the Great Recession by starting the data in January 2010 or even later. Figure 5 shows overlay plots comparing the index as calculated using weights from PCA applied to data from August 2004 to February 2020 (our baseline index), with alternatives obtained with weights determined by PCA applied over three alternative time frames: August 2004 to September 2022 (the most recent available data as of this writing), January 2010 to February 2020 (the time frame between the Great Recession and the pandemic), and January 2010 to September 2022 (avoiding the Great Recession but not the pandemic). The figure also shows the difference between the baseline index and the indexes constructed using alternative weights.

We see that the alternatives produce results that are generally similar to the baseline index. In particular, all the alternatives reflect the largest movements, which come from the Great Recession period and the COVID-19 pandemic. The most similar alternative is displayed in panel (b), which computed

PCA weights using data from January 2010 to February 2020. Panels (a) and (c) show more differences, including differences in the depth of the dips from the Great Recession and the start of the pandemic. The differences in the depths of these dips are not terribly important. While most economic time series were severely impacted by the Great Recession and the pandemic, the impacts did differ across series, so while it is important that the index reflect large downturns for these periods, there is no necessarily correct depth for the dips. In any case, taken together the graphs show that whether the pandemic time period data was included in computing the index weights mattered some to the resulting index, but whether or not the Great Recession time period data was included in computing the index weights mattered very little.

Table 7 shows correlations measuring the strength of the relation between the alternative versions of the index obtained by computing the PCA weights over the different time frames and the baseline index obtained by computing the weights using data from August 2004 to February 2020. For each alternative, correlations were computed over the four alternative time frames as shown in the table. The first thing to note about these results is that all the correlations are very high – above 0.9 in all cases. Another thing to note is that the correlations are highest when computed over the time frame covering the most recent data including the pandemic period (the second and fourth columns of correlations). This is because the large movement starting with the pandemic has a strong positive effect on the correlations. Omitting the Great Recession period from calculation of the correlations has a somewhat lesser effect on the results as can be seen by comparing the first two columns with the last two columns.

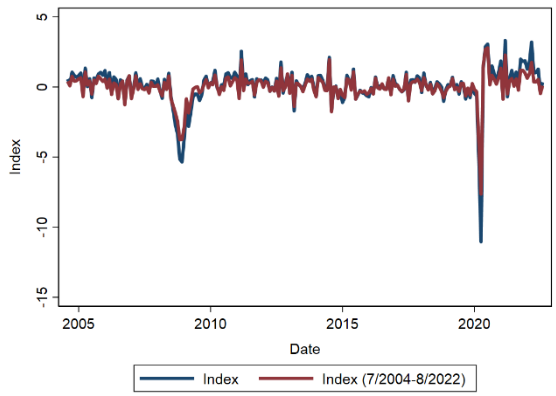
Table 7: Correlations between Baseline Index and Indices Constructed Using Alternative Weights

Alternative version of index (time frame used to compute PCA weights)	Time frame over which the correlations were computed			
	Remove pandemic (8/2004 – 2/2020)	Entire time span (8/2004 – 9/2022)	Between GR & pandemic (1/2010 – 2/2020)	Post-GR period (1/2010 – 9/2022)
Entire time span (8/2004 – 9/2022)	0.98	0.98	0.97	0.98
Between GR & pandemic (1/2010 – 2/2020)	0.97	0.94	0.91	0.97
Post-GR period (1/2010 – 9/2022)	0.95	0.94	0.95	0.96

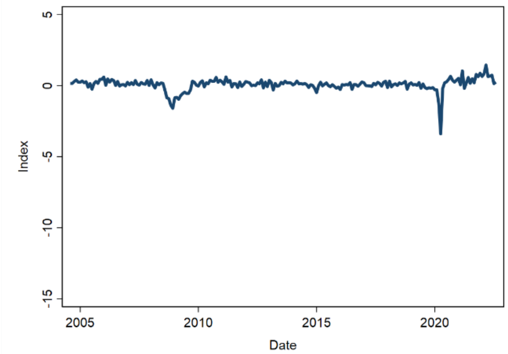
Notes: All results have significance levels of less than 0.01.

(a) Index if Weights Constructed using the Entire Time Span (8/2004 – 9/2022)

Baseline and Alternative Index

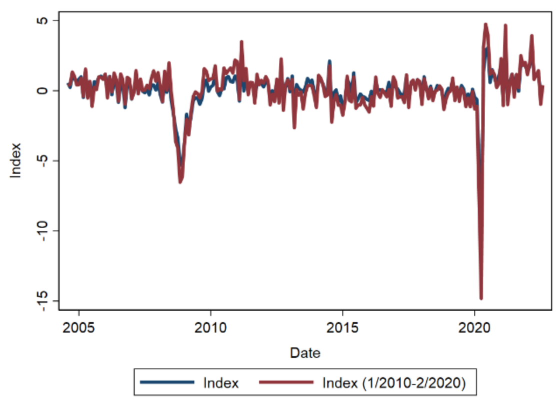


Difference (Baseline Index – Alternative Index)

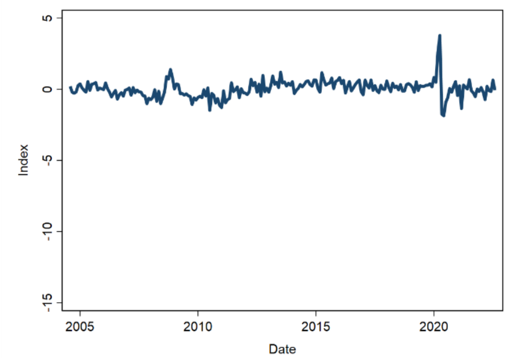


(b) Index if Weights Constructed using the Period Between the GR & Pandemic (1/2010 – 2/2020)

Baseline and Alternative Index

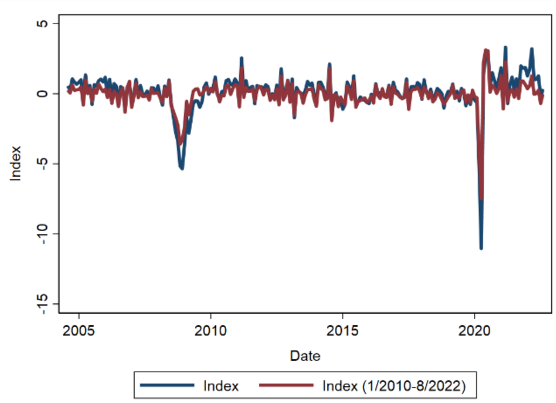


Difference (Baseline Index – Alternative Index)



(c) Index if Weights Constructed Using the Post-2010 Period (1/2010 – 9/2022)

Baseline and Alternative Index



Difference (Baseline Index – Alternative Index)

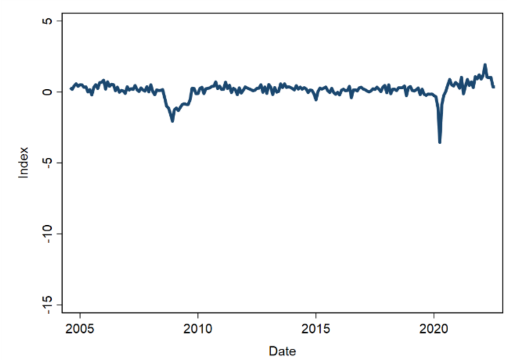


Figure 5. Index with Weights Constructed Using Alternative Time Periods (August 2004 to November 2022).

6 Conclusion and Directions for Future Work

In this paper, we used the method of principal components analysis to construct an economic index from 15 monthly Census Bureau economic indicators covering a broad range of economic activity. We then examined its properties and its relation to other economic indicators. One direction for future work is to somehow extend the index to reflect the incorporation of data related to the service sector, an important part of the economy not covered by the 15 monthly indicators in the current index. The constraint we face is that the Census Bureau's estimates of activity from the service sector are released only quarterly, from the Quarterly Services Survey (QSS), whereas the index constructed in this paper is monthly. One option would be to develop a corresponding quarterly index, which would allow us to incorporate QSS information into the index easily. Another option, which poses more technical challenges, would be to combine quarterly services data with the 15 monthly indicators of the current index to produce an enhanced monthly index.

Another area for future work will involve determining how often to update calculation of the PCA weights. As was mentioned in Section 3, we ended the time frame of the data for applying PCA to calculate the weights in February of 2020 to avoid potential distortions of the calculation from the large movements in the data due to the COVID-19 pandemic. In the future it clearly makes sense to decide that, beyond some point, potential serious distortions of the calculations from the pandemic have ceased, so that more recent data can be added into the PCA weight calculations. It is possible that we could decide that this point has already occurred. Continuing further onward, and apart from any possible large future economic disruptions, we will need to decide what schedule will be regularly used for incorporating new data to update the PCA weight calculations.

One topic not fully addressed here concerns possible measures of statistical uncertainty for the index. Results given in Appendix B show how variances of the nowcasting errors can be obtained from the fitted nowcasting model, but this uncertainty shows up only in the last month of the index time series. In addition to nowcasting errors, errors in the historical values of the input time series will be inherited by the index, with contributions influenced by the PCA weights shown in Table 3. All the input series are subject to various nonsampling errors, some of which are reflected in revisions made to their historical estimates, particularly for recent months. Information about the input series, possibly including information about their nonsampling errors, can be found at the links for the series given in Table 1. Nonsampling errors are, unfortunately, difficult to quantify. Some of the input series derive from Census Bureau sample surveys, and these series are subject to sampling error. The amount of sampling error in the input series thus ranges from none (for series not obtained from sample surveys, such as Exports and Imports), to a minor amount (from very aggregate series such as Retail Sales), to substantial (for some of the construction series). Since sampling variance estimates are available for those series subject to sampling error, and the index is a linear function of the input series growth rates, there is the possibility of deriving sampling variances for the index. Complications include possible dependence of sampling errors across series obtained from the same or related surveys (for construction series), and the fact that the index weights can be regarded as estimates of true weights with these estimated weights subject to sampling and other errors of the input series. It must also be kept in mind that this would not account for the various nonsampling errors in the input series. Future research will investigate the possibilities for quantifying error in the index and providing uncertainty measures.

Appendix A: Principal Components Analysis (PCA)

PCA is a multivariate statistical method of dimension reduction. If one has data on p characteristics $X_1, \dots, X_p$, PCA determines a reduced set of $m < p$ linear combinations that are uncorrelated[7] with each other and that provide, in some sense, as much of the information in the original set of X_i as possible, where i indexes the characteristic. These linear *combinations* are the "principal components." The simplest case has $m = 1$ and determines the single linear combination $\sum_i w_i X_i$ with maximum variance subject to the constraint $w_1^2 + \dots + w_p^2 = 1$. This is known as the first principal component.

The solution to the PCA task is to compute the eigenvalues and eigenvectors of the covariance matrix Σ of $(X_1, \dots, X_p)'$. The weights w_i for the first principal component are the elements of the eigenvector corresponding to the first (largest) eigenvalue of Σ. In practice, Σ is estimated using the available data, which can be written as vectors of observations $\boldsymbol{X}_t = (X_{1t}, \dots, X_{pt})'$ for $t = 1, \dots, n$. This is generally done using the usual formula:

$$\hat{\Sigma} = \frac{\mathbf{1}}{\boldsymbol{n-1}} \sum_{t=1}^{n} (\boldsymbol{X}_t - \bar{\boldsymbol{X}})(\boldsymbol{X}_t - \bar{\boldsymbol{X}})' \quad \text{where} \quad \bar{\boldsymbol{X}} = \left(\bar{X}_1, \dots, \bar{X}_p\right)' = \frac{\mathbf{1}}{\boldsymbol{n}} \sum_{t=1}^{n} \boldsymbol{X}_t.$$

Suppose the first eigenvalue of $\hat{\Sigma}$ is denoted λ_1 and the corresponding eigenvector by $\boldsymbol{w_1} = (w_{11}, \dots, w_{1p})'$. Then the first principal component is $\sum_i w_{1i} X_i$, and its variance is $var(\sum_i w_{1i} X_i) = \lambda_1$.[8] The value of the first principal component for any observation $\boldsymbol{X}_t$ can be computed as $\boldsymbol{w}_1{}'\boldsymbol{X}_t = \sum_i w_{1i} X_{it}$. The contribution of the first principal component to the overall variation in $(X_1, \dots, X_p)'$ is the ratio of its variance λ_1 to the "total variance" defined as the trace of the estimated covariance matrix $\hat{\Sigma}$, which is also the sum of all its eigenvalues λ_j. That is, the contribution is defined as $\lambda_1 / tr(\hat{\Sigma}) = \lambda_1 / (\lambda_1 + \dots + \lambda_p)$.

A question arises as to whether the PCA should be carried out using the original variables X_i or with standardized versions of these variables, $(X_{it} - \bar{X}_i)/\sqrt{var(X_i)}$. If the latter are used, then the eigenvalues and eigenvectors computed are those of the correlation matrix $R = D^{-1/2} \hat{\Sigma} D^{-1/2}$, where D is the

---

[7] If x_{it} are the variables for the i^{th} principal component, then the uncorrelatedness result says that $corr(x_{it}, x_{jt}) = 0$ for $i \neq j$. If the X_{it} for given i are independent random variables, not correlated over time, then $corr(x_{it}, x_{jk}) = 0$ for all t and k. If, however, the X_{it} are correlated over time, then generally $corr(x_{it}, x_{jk}) \neq 0$ for $t \neq k$.

[8] The p eigenvalues $\lambda_1, \dots, \lambda_p$ and corresponding eigenvectors $\boldsymbol{w}_1, \dots, \boldsymbol{w}_p$ of the $p \times p$ matrix Σ are the p solutions to the equation $\Sigma \boldsymbol{w} = \lambda \boldsymbol{w}$. Assuming the eigenvalues are distinct, they are ordered by decreasing magnitude, i.e., $\lambda_1 > \dots > \lambda_p$. Because any solution $\boldsymbol{w_j}$ multiplied by a nonzero scalar yields another solution with the same λ_j, the constraint $w_{j1}^2 + \dots + w_{jp}^2 = 1$ is imposed to determine a particular solution for a given eigenvalue λ_j (though this still leaves $-\boldsymbol{w_j}$ as an alternative solution). Note that $var(\boldsymbol{w_j'X}) = \boldsymbol{w_j'}\Sigma \boldsymbol{w}_j = \boldsymbol{w_j'}(\lambda_j \boldsymbol{w}_j) = \lambda_j$. When PCA is done with the covariance matrix $\hat{\Sigma}$, the principal components (also called "principal component scores") may be defined in terms of centered variables as $\sum_i w_{ji}(X_{it} - \bar{X}_i)$. When PCA is done with the correlation matrix R, the principal components are defined in terms of the standardized variables as $\sum_i w_{ji} (X_{it} - \bar{X}_i)/\sqrt{var(X_i)})$.

diagonal matrix with elements $\sqrt{var(X_i)}$. Also, since the diagonal elements of R are all ones, its trace is p, and thus the contribution of the first principal component from R to the overall variation in R is λ_1/p. One reason for using the correlation matrix R rather than the covariance matrix $\hat{\Sigma}$ to define the principal components is if the original variables X_i are measured in different units, in which case the determination of the first principal component could be dominated by the variable(s) whose measurements are the largest numbers simply due to the scale of their measurement units. While that is not the case for our application of PCA to growth rates of a set of Census Bureau economic indicators, we nevertheless decided to use the correlation matrix of the growth rates rather than the covariance matrix out of concern that component series with the most volatile growth rates could contribute more to the index than would be desirable. High volatility in the growth rate of a given indicator could stem from inherent noisiness that did not reflect meaningful economic movements, as could occur for an indicator that was survey estimates that contained substantial sampling error.

More information on PCA can be found in texts on multivariate statistics, in a review article by Jolliffe and Cadima (2016), and in the econometrics book by Stock and Watson (2019).

Appendix B: Nowcasting with a VAR(1) model

Let $\boldsymbol{X}_t$ be a $p \times 1$ vector time series following a VAR(1) model: $\boldsymbol{X}_t = \boldsymbol{\mu} + \Phi(\boldsymbol{X}_{t-1} - \boldsymbol{\mu}) + a_t$ where $\boldsymbol{\mu} = E(\boldsymbol{X}_t)$, $a_t \sim i.i.d.\, N(0, \Sigma)$, and Φ is the $p \times p$ AR(1) parameter matrix. Break $\boldsymbol{X}_t$ into two subvectors, $\boldsymbol{X}_t = (\boldsymbol{U}_t', \boldsymbol{V}_t')'$, and partition the VAR(1) model conformably as:

$$\boldsymbol{X}_t \equiv \begin{bmatrix} \boldsymbol{U}_t \\ \boldsymbol{V}_t \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} + \begin{bmatrix} \Phi_1 \\ \Phi_2 \end{bmatrix} (\boldsymbol{X}_{t-1} - \boldsymbol{\mu}) + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} \quad \text{with} \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{21}' \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}.$$

Suppose that, at time t, $\boldsymbol{U}_t$ has been observed, as has the full $\boldsymbol{X}_j$ for $j < t$, but $\boldsymbol{V}_t$ is unobserved. The following are the formulas for predicting (nowcasting) $\boldsymbol{V}_t$ from $(\boldsymbol{U}_t, \boldsymbol{X}_{t-1})$, and for the prediction error covariance matrix:

$$\widehat{\boldsymbol{V}}_t \equiv E(\boldsymbol{V}_t | \boldsymbol{U}_t, \boldsymbol{X}_{t-1}) = \boldsymbol{\mu}_2 + \Phi_2(\boldsymbol{X}_{t-1} - \boldsymbol{\mu}) + \Sigma_{21}\Sigma_{11}^{-1}[\boldsymbol{U}_t - \boldsymbol{\mu}_1 - \Phi_1(\boldsymbol{X}_{t-1} - \boldsymbol{\mu})]$$

$$var\left(\boldsymbol{V}_t - \widehat{\boldsymbol{V}}_t\right) = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{21}'.$$

Derivation: For simplicity, we first derive the results for the case where the mean vector, $\boldsymbol{\mu}$, of $\boldsymbol{X}_t$, is zero.

$$\begin{aligned} \widehat{\boldsymbol{V}}_t \equiv E(\boldsymbol{V}_t | \boldsymbol{U}_t, \boldsymbol{X}_{t-1}) &= E(\Phi_2 \boldsymbol{X}_{t-1} + a_{2t} | a_{1t}, \boldsymbol{X}_{t-1}) && \text{since } a_{1t} = \boldsymbol{U}_t - \Phi_1 \boldsymbol{X}_{t-1} \\ &= \Phi_2 \boldsymbol{X}_{t-1} + E(a_{2t}|a_{1t}) + E(a_{2t}|\boldsymbol{X}_{t-1}) && \text{since } a_{1t} \perp \boldsymbol{X}_{t-1}{}^{9} \\ &= \Phi_2 \boldsymbol{X}_{t-1} + \Sigma_{21}\Sigma_{11}^{-1}(\boldsymbol{U}_t - \Phi_1 \boldsymbol{X}_{t-1}) && \text{since } E(a_{2t}|\boldsymbol{X}_{t-1}) = 0. \end{aligned}$$

The covariance matrix of the nowcast error, $\varepsilon_{2t} = \boldsymbol{V}_t - \widehat{\boldsymbol{V}}_t$, is obtained as follows:

$$\begin{aligned} \boldsymbol{\varepsilon}_{2t} &= \boldsymbol{V}_t - \Phi_2 \boldsymbol{X}_{t-1} - E(a_{2t}|a_{1t}) \\ &= a_{2t} - E(a_{2t}|a_{1t}) \\ \Rightarrow var(\boldsymbol{\varepsilon}_{2t}) &= \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{21}'. \end{aligned}$$

For the case where $\boldsymbol{\mu} \neq \boldsymbol{0}$, we can replace $\boldsymbol{X}_t$, $\boldsymbol{U}_t$, and $\boldsymbol{V}_t$ in the expression from the derivation for $\widehat{\boldsymbol{V}}_t$ when $\boldsymbol{\mu} = \boldsymbol{0}$ by $\boldsymbol{X}_t - \boldsymbol{\mu}$, $\boldsymbol{U}_t - \boldsymbol{\mu}_1$, and $\boldsymbol{V}_t - \boldsymbol{\mu}_2$. Slightly rearranging terms in the resulting expression for $\widehat{\boldsymbol{V}}_t$ gives the nowcast for the case of $\boldsymbol{\mu} \neq \boldsymbol{0}$. The expression for $var(\boldsymbol{\varepsilon}_{2t})$ remains the same because the derivation assumes that $\boldsymbol{\mu}$ is known. These results for the VAR(1) model are a special case of more general results derived in McElroy, et al. (2021).

The expression for $var(\boldsymbol{\varepsilon}_{2t})$ can be augmented with an additional term to account for error in

[9] A general result on linear projections is that if X, Y, and Z are zero mean jointly Gaussian random variables, with X and Y orthogonal ($X \perp Y$, meaning (cov(X,Y) = 0), then $E(Z|X,Y) = E(Z|X) + E(Z|Y)$.

estimating $\boldsymbol{\mu}$ from the observed data. Further augmentation to account for error in estimating the AR and covariance matrix parameters, Φ and Σ, is also possible but more complicated.

For the case where an arbitrary subset of the elements of $\boldsymbol{X}_t$ is observed, we define $\boldsymbol{U}_t$ as the observed elements of $\boldsymbol{X}_t$, $\boldsymbol{V}_t$ as the unobserved elements of $\boldsymbol{X}_t$, select the corresponding elements of $\boldsymbol{\mu}$ to define $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, and the corresponding rows of Φ and rows and columns of Σ to define the blocks of Φ and Σ, and then apply the above formulas. For example, suppose $\boldsymbol{X}_t$ has 5 elements (5 input variables) of which the first, third, and fourth are observed at time t and the second and fifth are not. Let $[k_1] \sim [1,3,4]$ denote selection of the first, third, and fourth elements of a vector, or of the rows and/or columns of a matrix, with $[k_2] \sim [2,5]$ analogously defined. Then $\boldsymbol{U}_t = \boldsymbol{X}_t[k_1]$, $\boldsymbol{V}_t = \boldsymbol{X}_t[k_2]$, $\boldsymbol{\mu}_1 = \boldsymbol{\mu}[k_1]$, $\boldsymbol{\mu}_2 = \boldsymbol{\mu}[k_2]$, $\Phi_1 = \Phi[k_1,]$, $\Phi_2 = \Phi[k_2,]$, $\Sigma_{11} = \Sigma[k_1, k_1]$, $\Sigma_{21} = \Sigma[k_2, k_1]$, and $\Sigma_{22} = \Sigma[k_2, k_2]$.

Appendix C: Cross-correlation Analysis

In Section 4.2, we compared the growth rates through time of the index and a few widely watches series, such as nonfarm employment, that are not in the index. For example, we analyzed how the index and these series behaved during the Great Recession and the COVID-19 pandemic. We find that the cross-correlation analyses give similar patterns as we found in Section 4.2 regarding these turning points in the economy.

In order to analyze the relationship between these series in a more systematic manner, we use cross-correlations. We define the cross-correlation between series x and y to be

$$\rho_{xy}(k) = corr(y_t, x_{t+k}),$$

where y_t is the growth rate of one of the series considered in Section 4.2 at time t and x_{t+k} is the index at time $t + k$. Note that k is the number of periods that we use to lag the growth rates of the index. For example, if $k = 0$ then the cross-correlation would be the same as those reported in Section 4.2 (i.e., the growth rates of both series); if $k = -1$ then cross-correlation would report the correlation between the growth rate of the series such as nonfarm employment at time t and the growth rate of the index at time $t - 1$. In our exercises, we consider k from –12 to 12.

Panels (a)-(c) of Figure 6 report the results of these cross-correlations between the 3 monthly series that we consider and the index when we use the entire time span for the calculations, where k or the number of periods that we use to lag the growth rate of the index is plotted on the x-axis. Intuitively, if the index leads a particular series then we would expect stronger correlations to the left of $k = 0$; if the index lags a particular series then we would expect stronger correlations to the right of $k = 0$. We find that for all 3 series, the strongest correlation is for the case of $k = 0$, suggesting that these three series are coincident with the index when we use the entire time span. Note that these results are consistent with the timing of the turning points of these series during the COVID-19 pandemic, as discussed in Section 4.2.

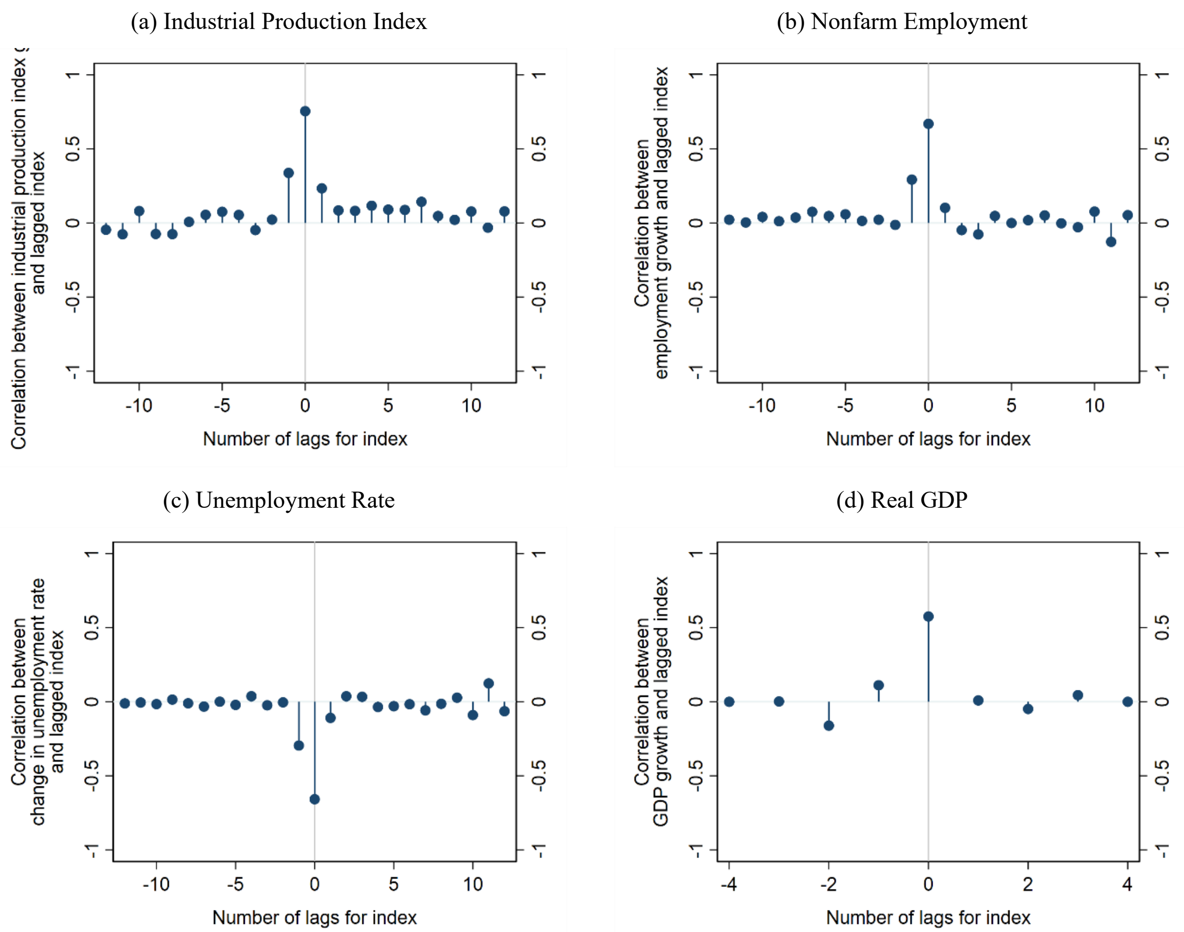


Figure 6. Cross-correlations between the Growth Rates of Other Economic Indicators and the Index using Growth Rates from August 2004 to August 2022 (Entire Time Span).

As mentioned before, the inclusion of the COVID-19 pandemic and Great Recession can have important effects on these correlations. For that reason, Figure 7 reports the same results except that we remove the pandemic period, including only the growth rates until February 2020. We find in panel (a) that, in the case of industrial production, the strongest correlations tend to be centered around 0, which reflects that the index is coincident with industrial production. In panels B and C, we find that the strongest correlations are centered around $k = -2$, which suggests that the index leads nonfarm employment and the unemployment rate by 2 months. Note that these results are consistent with the timing of the turning points of these series during the Great Recession, as discussed in Section 4.2.

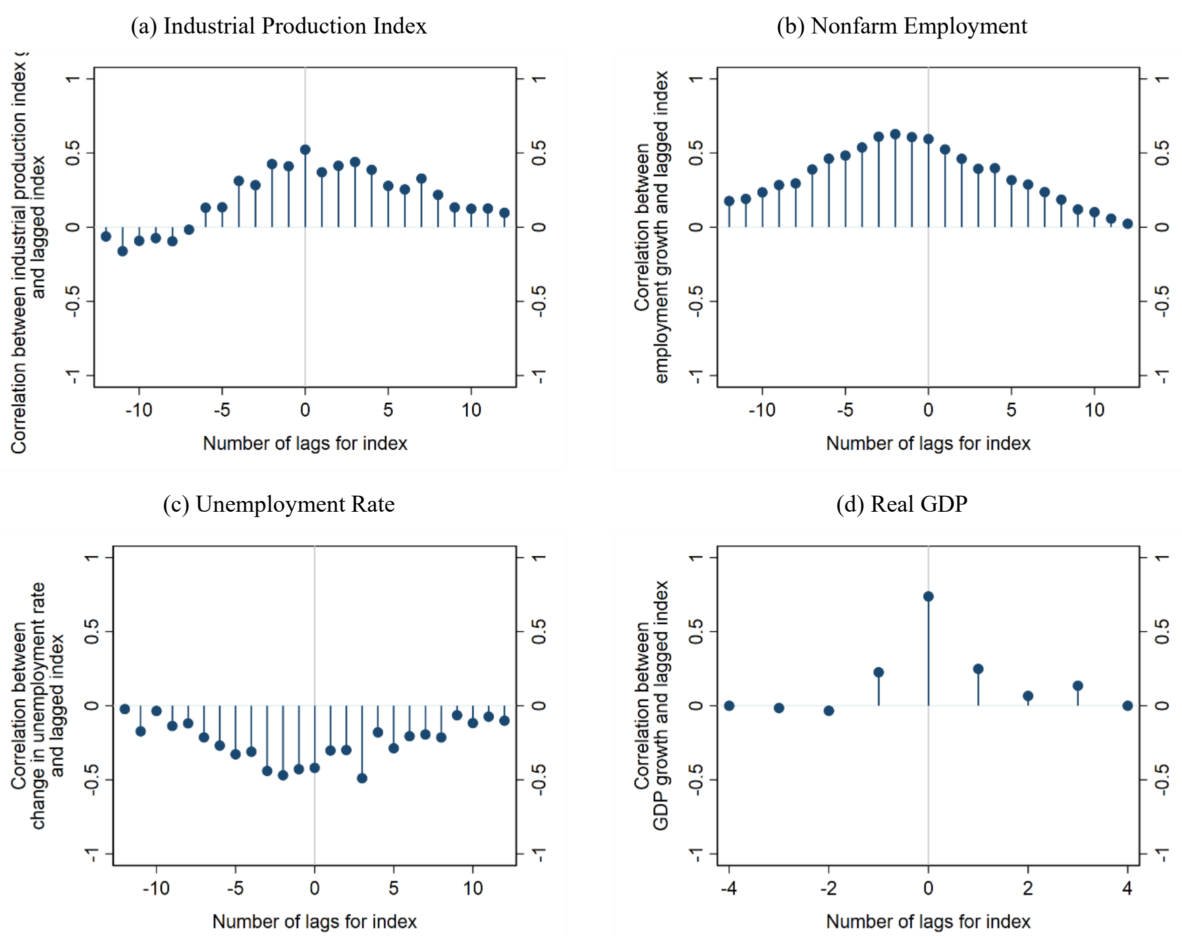


Figure 7. Cross-correlations between the Growth Rates of Other Economic Indicators and the Index using Growth Rates from August 2004 to February 2020 (Remove Pandemic).

Finally, we redo the exercise, except that now we only include the growth rates between January 2010 and February 2020, which excludes the Great Recession. Panels (a)-(c) of Figure 8 report these results. Panel (a) shows that the strongest correlations between the index and industrial production occur between values of k ranging from –4 to 0, suggesting that the index leads industrial production by a few months over this time period, although the correlations tend to be much smaller relative to other periods considered. Panel (b) reports the cross-correlations between nonfarm employment and the index and panel (c) reports the cross-correlations between the unemployment rate and the index. In Section 4.2, we found that, over this time period, the correlations are not statistically different from 0 when $k = 0$. We find in panels (b) and (c) that the correlations do not increase appreciably in magnitude when lagging the growth rates of the index using various values of k.

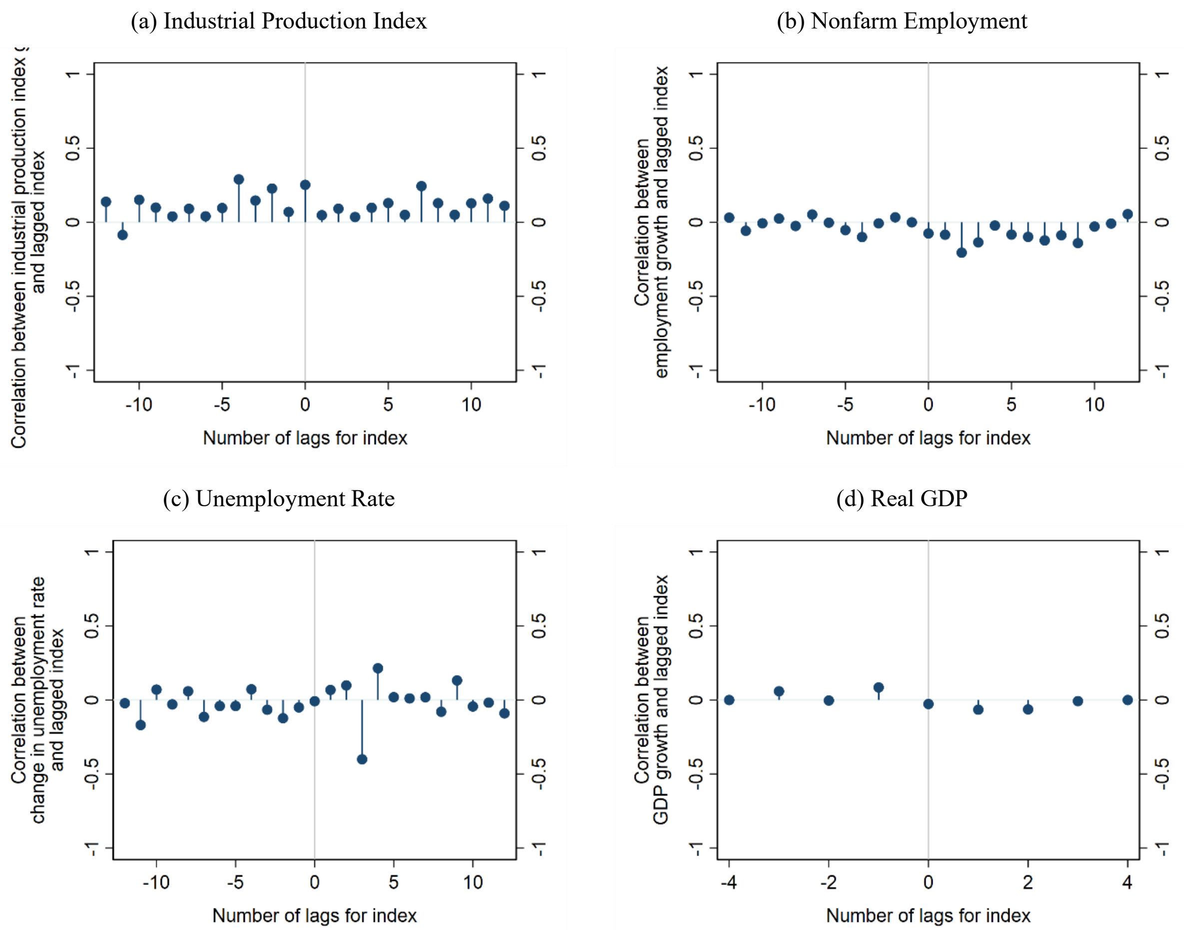
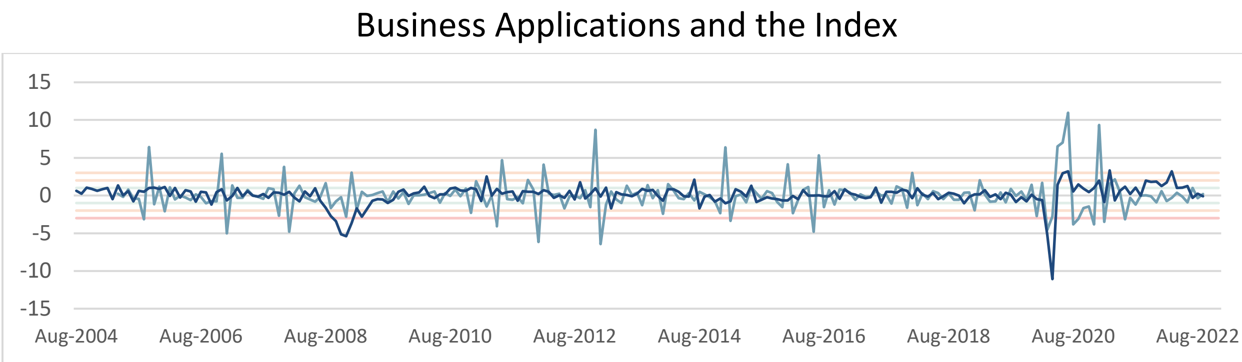
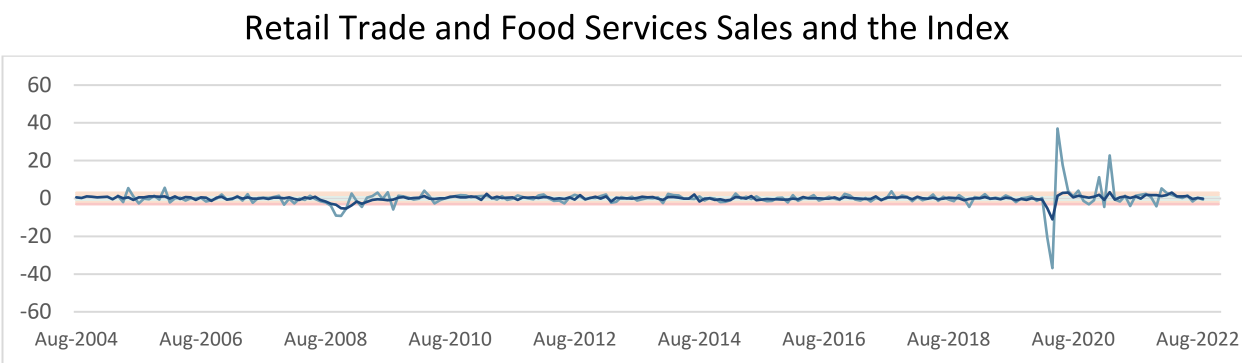


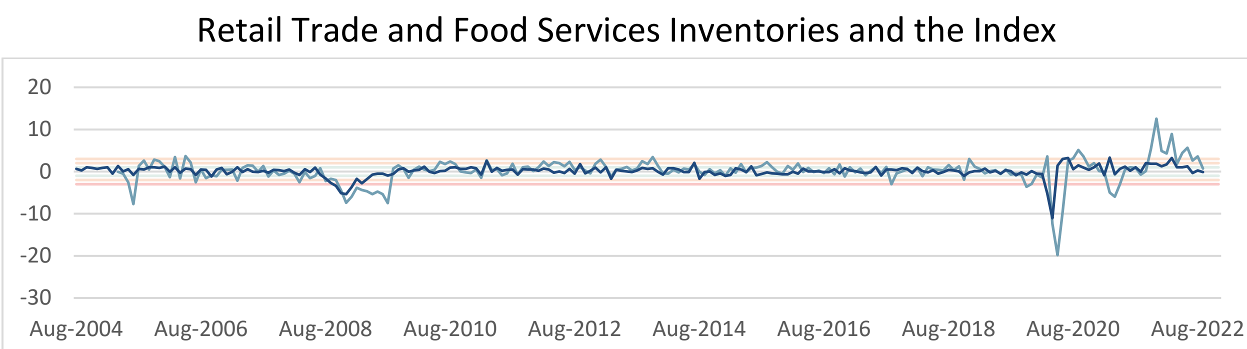
Figure 8. Cross-correlations between the Growth Rates of Other Economic Indicators and the Index using Growth Rates from January 2010 to February 2020 (Between Great Recession and Pandemic).

Panel D of Figures 6-8 reports the cross-correlations between the index and GDP for the various time periods considered. Note that, as in Section 4.2, we create a quarterly index by summing the index over all months in a quarter. Furthermore, we use values of k ranging from –4 to 4, which is consistent with the lags used to analyze the monthly series. We find that, when we include either the COVID-19 pandemic or the Great Recession in the analysis, the strongest correlations are found when $k = 0$, which suggests that the index is coincident with GDP. Furthermore, we find low correlations between the index and GDP during the time period January 2010 to February 2020, even when we consider various lags for the index. Note that in Section 4.2 we found that the correlation between the index and the growth rate of GDP was not statistically significant over this period.

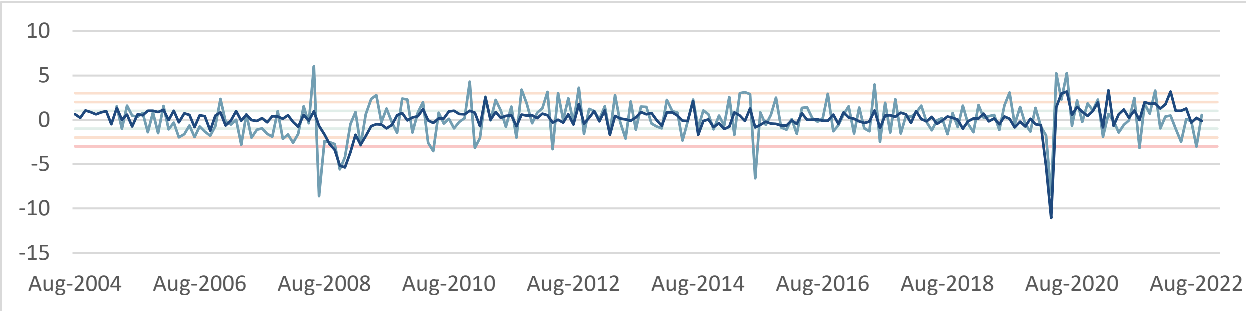
Appendix D: Plots of Input Time Series Growth Rates (in blue) with the Index (in black)



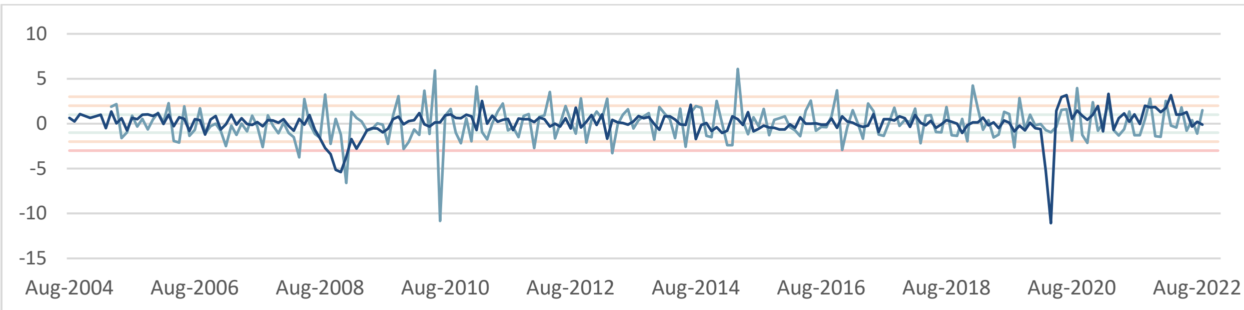




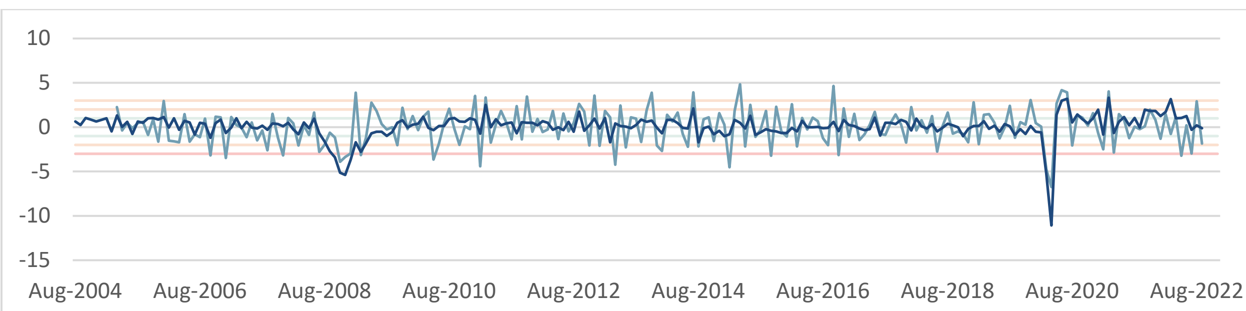
Housing Units Authorized in Permit-Issuing Places and the Index



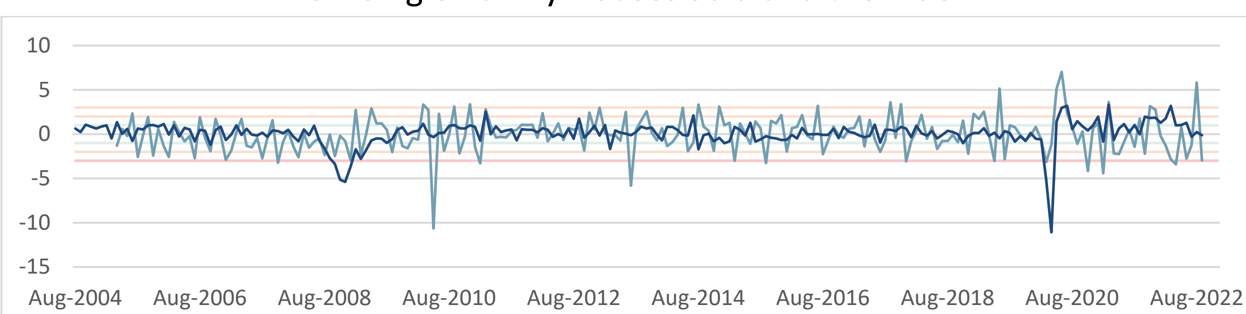
Housing Units Completed and the Index



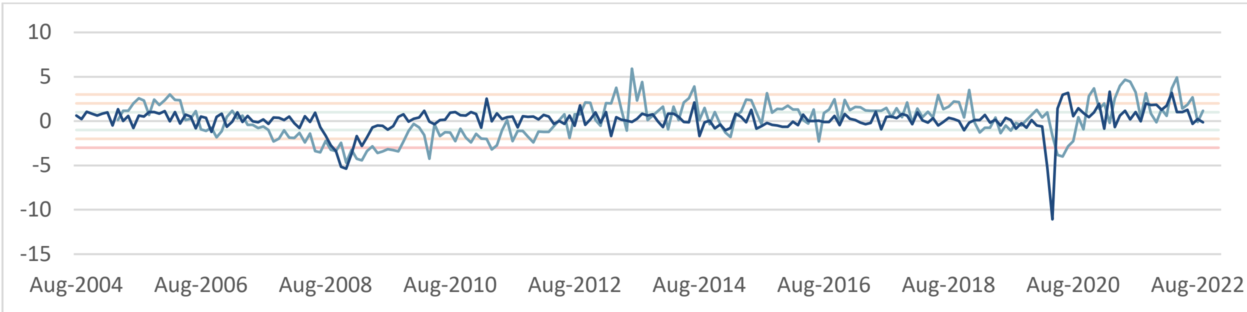
Housing Units Started and the Index



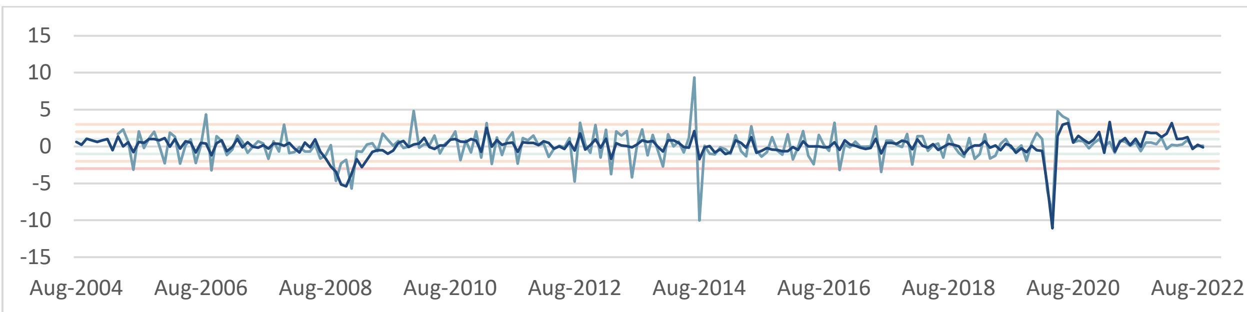
New Single-Family Houses Sold and the Index



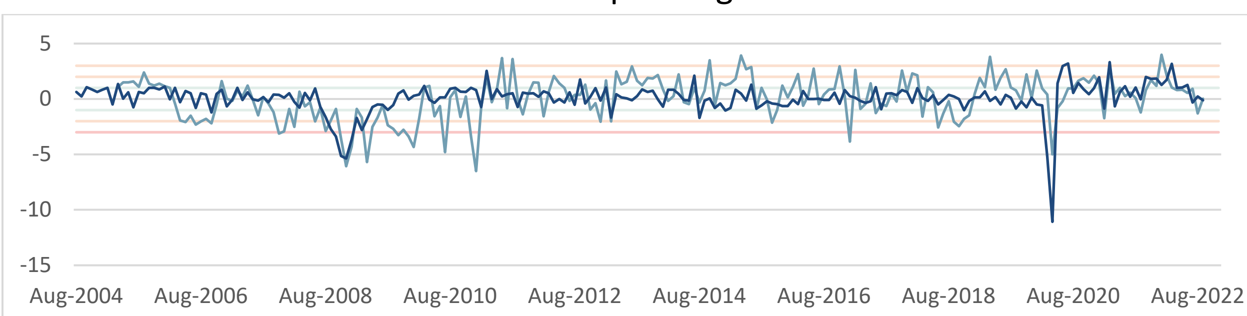
New Single-Family Houses for Sale and the Index



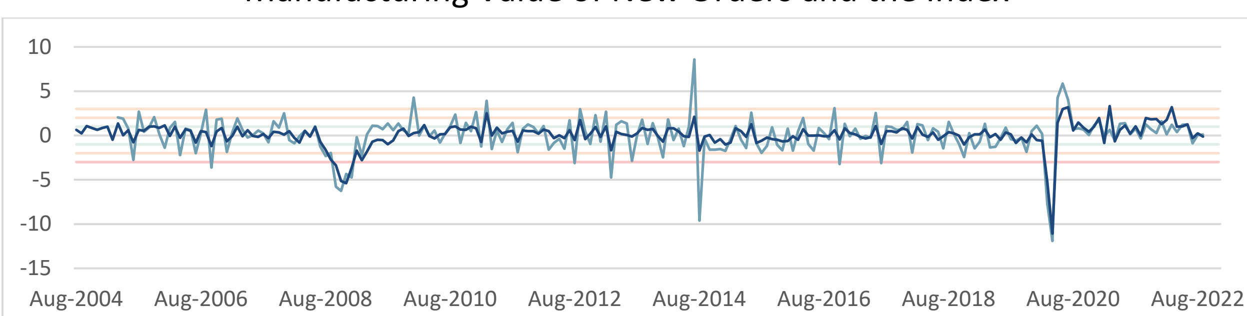
New Orders for Durable Goods and the Index



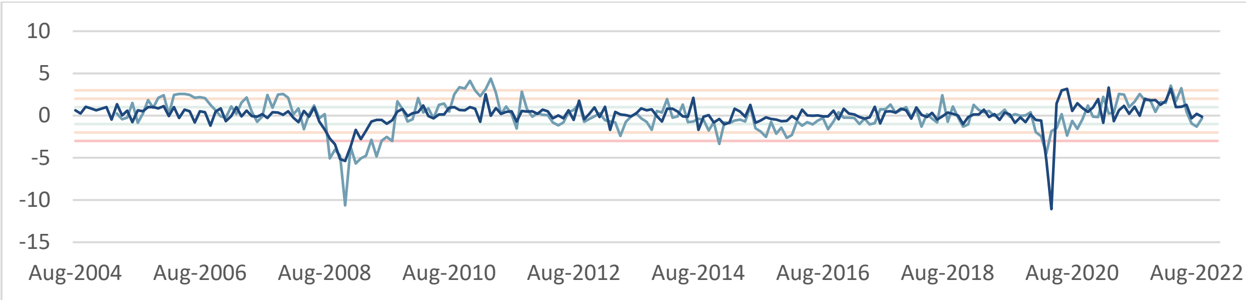
Total Construction Spending and the Index



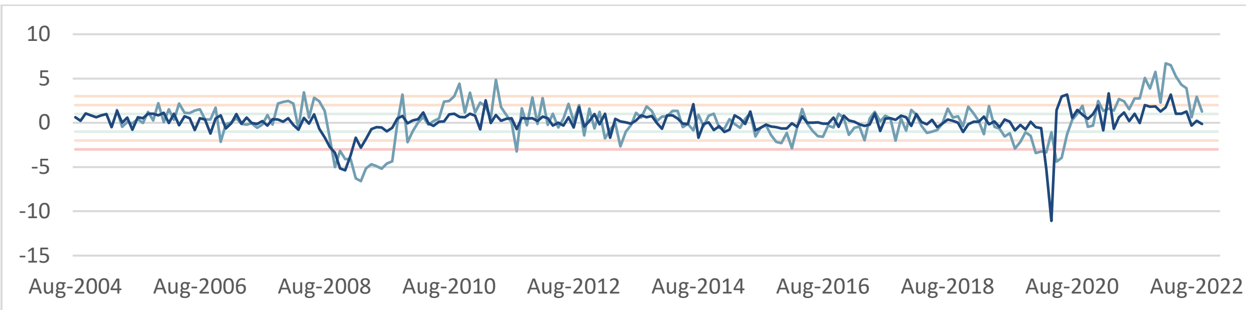
Manufacturing Value of New Orders and the Index



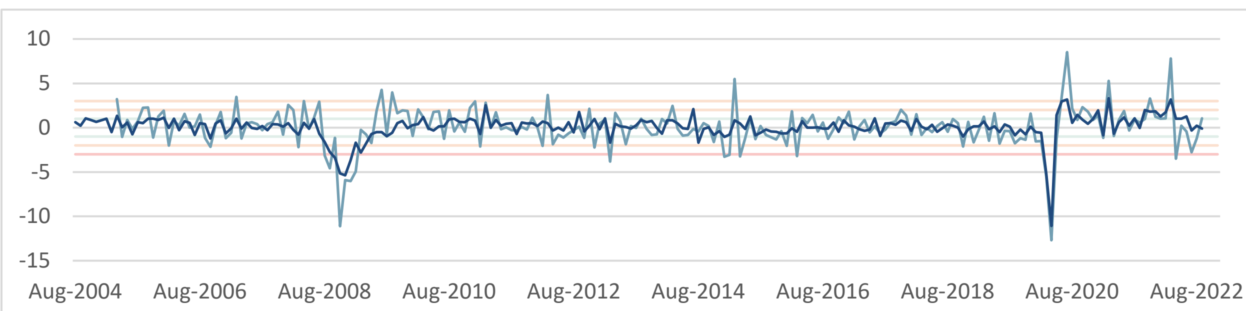
Manufacturing Inventories and the Index



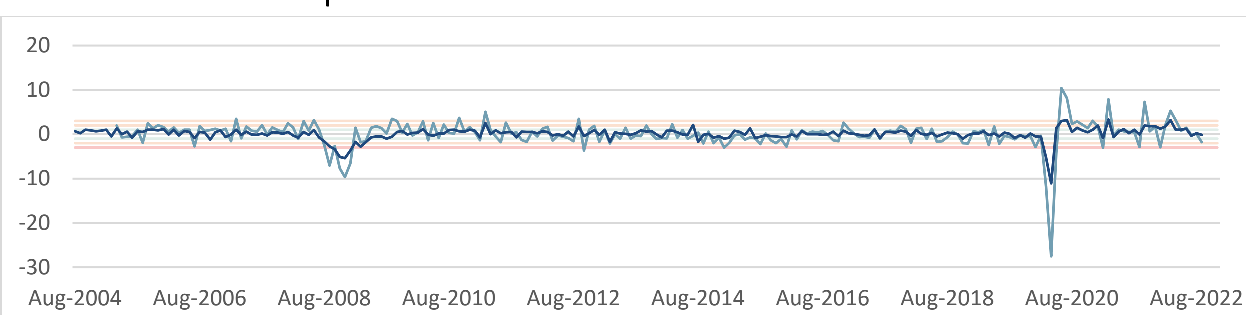
Wholesale Trade Inventories and the Index



Imports of Goods and Services and the Index



Exports of Goods and Services and the Index



References

Asturias, Jose, Emin Dinlersoz, John Haltiwanger, and Rebecca Hutchinson. 2021. "Business Applications as Economic Indicators." *Center for Economic Studies Working Paper 20-09*.

Baumeister, Christiane, Danilo Leiva-Leon, and Eric R. Sims. 2021. "Tracking Weekly State-level Economic Conditions." *NBER Working Paper 29003*.

Brave, Scott. 2008. "Economic trends and the Chicago Fed National Activity Index." *Chicago Fed Letter*, no. 250. https://www.chicagofed.org/publications/chicago-fed-letter/2008/may-280.

Jolliffe, Ian T., and Jorge Cadima. 2016. "Principal Component Analysis: a Review and Recent Developments." *Philosophical Transactions of the Royal Society A* 374, no. 2065. http://dx.doi.org/10.1098/rsta.2015.0202.

Lewis, Daniel J., Karel Mertens, James H. Stock, and Mihir Trivedi. 2022. "Measuring Real Activity Using a Weekly Economic Index." *Journal of Applied Econometrics* 37, no. 4: 667–687.

McElroy, Tucker, and David F. Findley. 2015. "Fitting Constrained Vector Autoregression Models." In *Empirical Economic and Financial Research – Theory, Methods, and Practice*, edited by Jan Beran, Yuanhua Feng, and Hartmut Hebbel, 451–470. New York: Springer.

McElroy, Tucker, Anindya Roy, James Livsey, Theresa Firestine, and Ken Notis. 2021. "Anticipating Revisions in the Transportation Services Index." *Journal of the International Association of Official Statistics* 37, no. 2: 641–653.

McElroy, T. S., and Thomas Trimbur. 2022. "Variable Targeting and Reduction in Large Vector Autoregressions with Applications to Workforce Indicators." *Journal of Applied Statistics*: 1-23.

Stock, James H., and Mark W. Watson. 2019. *Introduction to Econometrics*. New York: Pearson.