RESEARCH BRIEF

Do State Indicators Trump National Ones for Predicting Economic Activity in the States? The Case of New York

Tosmai Puenpatom and Tom Stark
October 26, 2016

Tosmai Puenpatom is a senior regional economic analyst and Tom Stark is assistant director and manager of the Real-Time Data Research Center in the Research Department of the Federal Reserve Bank of Philadelphia.

Since 2010, the Federal Reserve Bank of Philadelphia has published a leading index of economic activity for each state in the U.S. These leading indexes project the six-month growth rate of our state coincident indexes and have drawn the interest of the private sector, state governments, and academics. Recently, at the request of the New York Department of Labor, we took a closer look at our leading index for New York to weigh the relative contributions of the state and national economic variables behind it.[1]

Specifically, we conducted an empirical exercise to isolate the predictive performance of each variable over time. Knowing the importance of the state versus the national variables in our indexes is of potential interest for understanding the results of our monthly updates, for serving as a guide for ways to improve our statistical model, and for understanding the differences among states.

The key findings of our empirical analysis are:

- New York state variables are more important than national variables in explaining variation in New York's coincident index, suggesting that state variables are more important than national ones for our leading index.
- New York state variables have been more important in the recent data than they were historically.

Our State Leading Indexes Model: A Primer

Our focus in this brief is on the variables, state and national, that we use in our statistical model for the leading indexes.The model produces each state's index as a six-month-ahead forecast for the coincident index. The forecast is based on a one-period near-vector autoregression (near-VAR) for the following variables:[2]

- That state's coincident index.
- State-specific building permits.
- State-specific unemployment claims.
- The interest rate spread between 10-year and three-month Treasury securities.
- The Institute for Supply Management's (ISM) delivery time diffusion index.[3]

The last two are national variables and the rest are state variables. Notably, the near-VAR model has five equations, one for each variable, and it produces projections over a six-month horizon for each variable. The six-month projection for the coincident index is the leading index. Importantly, as we describe more fully below, our interest in this brief is in only one of the five equations, the one for the coincident index itself.

Methodological Considerations

The relative contribution of each explanatory variable in the near-VAR used to produce the

New York state leading index is not an easy question to address, precisely because there is almost surely dynamic interaction among the variables in our statistical model, making it hard to attribute an independent contribution to each explanatory variable. Indeed, the near-VAR specification assumes the presence of such dynamic interactions. A precise answer can come only from a structural model, one that specifically addresses the linkages among variables, not from the reduced-form forecasting model we use to produce the leading index.

We take a nonstructural forecasting approach to assessing the contribution of the explanatory variables behind New York's leading index. One reasonable option would have been to conduct an out-of-sample experiment using our near-VAR framework to generate competing forecasts for the coincident index based on, alternately, including and excluding selected explanatory variables and comparing the relative accuracy of the corresponding projections. Instead, we take a somewhat more direct approach, looking at the in-sample predictive ability of the near-VAR.[4]

We limit our approach to assessing the in-sample importance of each right-hand-side variable in the near-VAR equation for the New York coincident indicator. The equation of interest from our near-VAR model is

$$DLindex_t^{NY} = \mu + \sum_{j=1}^{4}\phi_{1j}DLindex_{t-j}^{NY} + \sum_{j=1}^{4}\phi_{2j}DLpermits6_{t-j}^{NY} + \sum_{j=1}^{4}\phi_{3j}DLclaims3_{t-j}^{NY}$$
$$+\sum_{j=1}^{4}\phi_{4j}Spread_{t-j}^{US} + \sum_{j=1}^{4}\phi_{5j}ISMdelivery_{t-j}^{US} + \varepsilon_t$$

where $DLindex_t^{NY}$ is the month-over-month percent change in the New York coincident index, $DLpermits6_t^{NY}$ is the month-over-month percent change in the six-month average level of New York building permits, $Dlclaims3_t^{NY}$ is the month-over-month percent change in the three-month average level of New York initial unemployment claims, $Spread_t^{us}$ is the constant-maturity yield on 10-year Treasury securities minus the constant-maturity yield on three-month Treasury bills, and $ISMdelivery_t^{US}$ is the ISM's delivery time diffusion index. Notice that the first three variables are state-specific to New York and the last two are national. All percent changes are computed as the first-difference of the natural logarithm of the underlying level and expressed as nonannualized percents, not in percentage points. Interest rates are expressed in annualized percentage points.

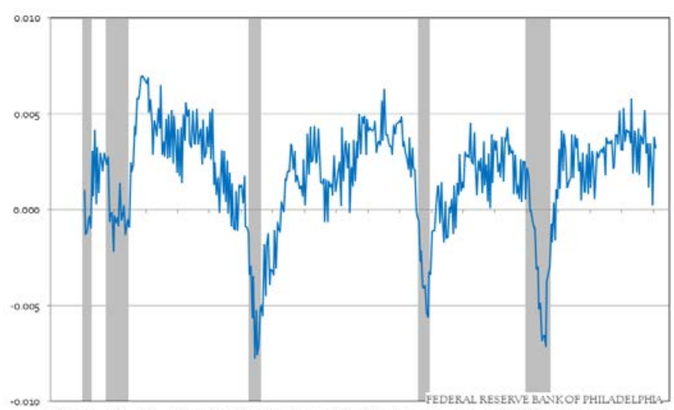
The underlying premise of our methodology is that in-sample evidence via standard T-tests and F-tests should give an indication of the importance of each right-hand-side variable for producing a forecast — the leading index — of the coincident index. Notably, we exclude an analysis of the importance of the coincident index itself because time series models always have substantial, almost always dominant, explanatory power from the lagged dependent variable. We concentrate instead on the in-sample contributions of New York state building permits and unemployment claims and the interest rate spread and ISM delivery time. We consider these variables individually but also as groups, specifically the state variables — building permits and claims — and the national ones — interest-rate rate spreads and ISM delivery times.

Full-Sample Results Suggest Importance of State Indicators

The data are shown in Figures 1–5, with A showing the level of the variable and B showing the transformation that enters into the near-VAR. The transformed data display no obvious econometric difficulties. The data appear stationary, have fairly regular variation, and, for the most part, appear cyclical. The New York coincident index estimates pronounced declines in each of the last three national recessions (Figures 1A–B).

FIGURE 1A
New York State Coincident Index: Level

FIGURE 1B
New York State Coincident Index: Growth



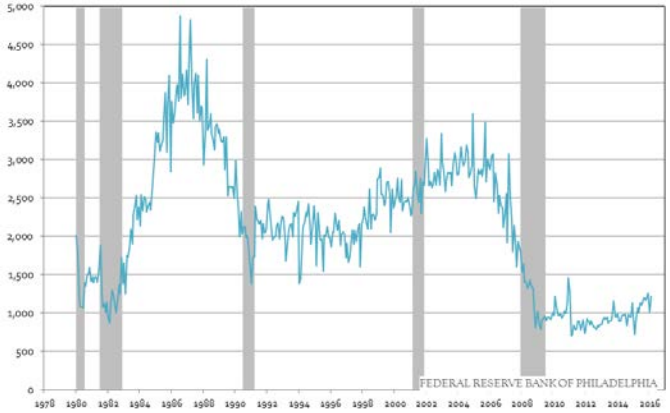
Note: Nonannualized percent using continuous compounding.

The most recent recession stands out for its estimated breadth and depth. It is also noteworthy that New York building permits generally fall in national recessions (Figures 2A–B), while New York unemployment claims rise (Figures 3A–B).

FIGURE 2A

New York State Building Permits: Level

Number of permits



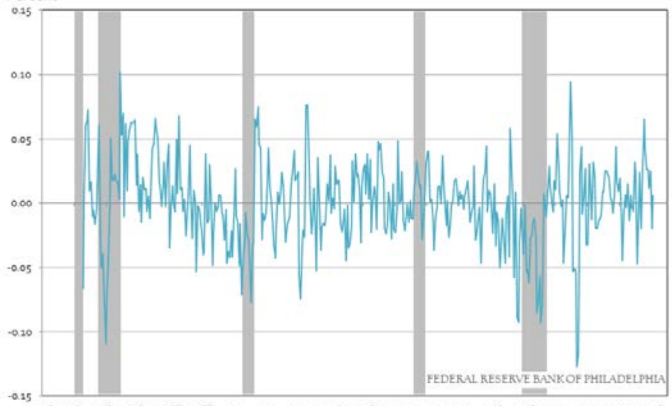
Source: Census Bureau.

FIGURE 2B

New York State Building Permits: Growth

Month-over-month growth of six-month moving average.

Percent



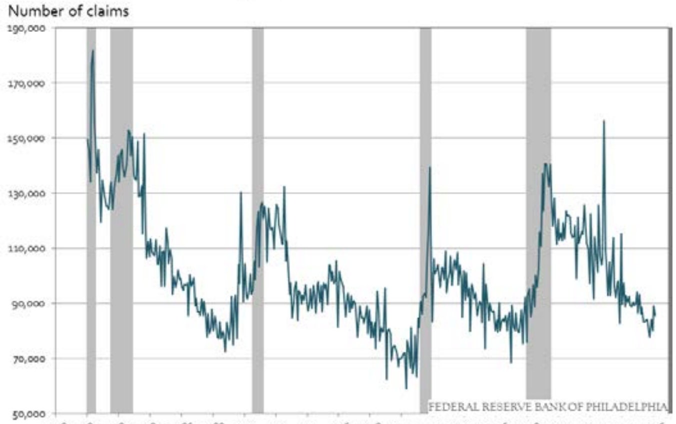
Source: Census Bureau.

Note: Nonannualized percent using continuous compounding.

FIGURE 3A

New York State Unemployment Claims: Level

Number of claims



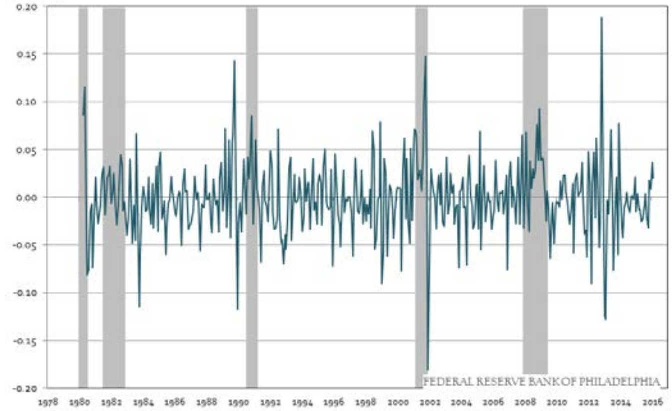
Source: Bureau of Labor Statistics.

FIGURE 3B

New York State Unemployment Claims: Growth

Month-over-month growth of three-month moving average.

Percent



Source: Bureau of Labor Statistics.

Note: Nonannualized percent using continuous compounding.

FIGURE 4

Interest Rate Spread: Level

Spread on yields of three-month and 10-year U.S. Treasury securities.

Percent

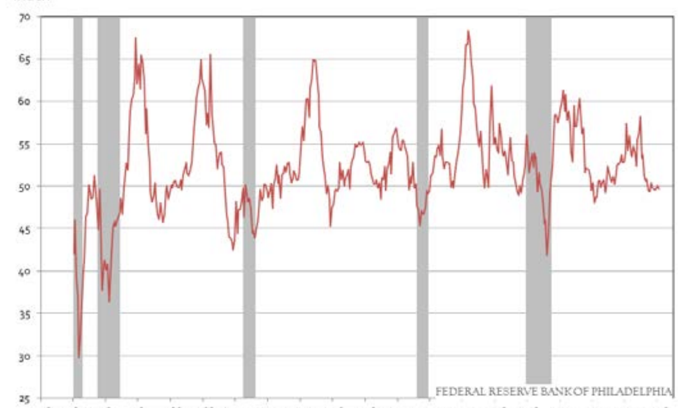
Source: Treasury Department.

Note: Nonannualized percent using continuous compounding.

FIGURE 5

Institute for Supply Management Delivery Time: Level

Index



Source: Institute for Supply Management via Haver Analytics.

Table 1 presents the full sample ordinary least squares estimation results for the above equation, whose dependent variable is the Philadelphia Fed's coincident index for New York. We show statistically significant coefficient estimates in red.

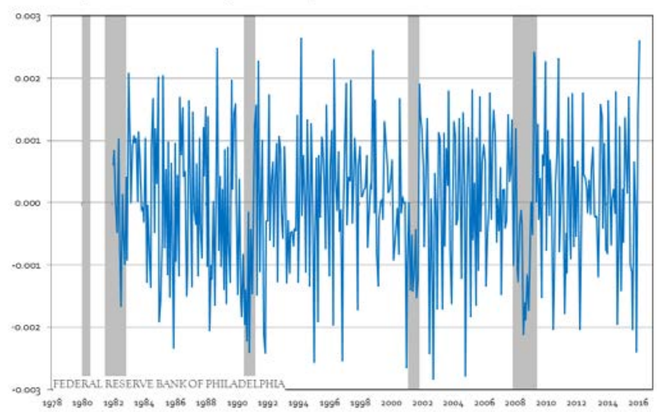
TABLE 1
Estimation Results for New York Coincident Index, January 1982–February 2016

Variable	Coefficient	T-Statistic	P-Value
Constant	0.001	1.062	0.289
Lagged values of NY coincident index			
1st lag	0.211	4.051	0.000
2nd lag	0.725	13.822	0.000
3rd lag	0.137	2.653	0.008
4th lag	-0.172	-3.339	0.001
Lagged values of NY building permits			
1st lag	0.005	2.108	0.036
2nd lag	-0.004	-1.517	0.130
3rd lag	0.001	0.547	0.584
4th lag	0.002	0.717	0.474
Lagged values of NY unemployment claims			
1st lag	-0.005	-2.912	0.004
2nd lag	0.001	0.720	0.472
3rd lag	-0.002	-1.348	0.179
4th lag	-0.000	-0.104	0.918
Lagged values of 10-year minus 3-month Treasury interest rate spread			
1st lag	-0.3×10^{-3}	-1.461	0.145
2nd lag	0.6×10^{-3}	1.493	0.136
3rd lag	-0.6×10^{-3}	-1.569	0.117
4th lag	0.4×10^{-3}	1.985	0.048
Lagged values of ISM delivery time			
1st lag	0.2×10^{-4}	0.672	0.502
2nd lag	0.3×10^{-4}	0.824	0.410
3rd lag	-0.7×10^{-4}	-1.718	0.087
4th lag	-0.1×10^{-5}	-0.033	0.974

FIGURE 6A
Regression Results for New York State Coincident Index
Realizations and in-sample predictions.

Sources: Bureau of Labor Statistics, Census Bureau, Treasury Department, Institute for Supply Management via Haver Analytics.
Note: Index is regressed on lags of itself and on lags of permits, claims, interest rate spread, and ISM delivery time.

FIGURE 6B
Regression Results for New York State Coincident Index
Residuals (realizations minus predictions).



Sources: Bureau of Labor Statistics, Census Bureau, Treasury Department, Institute for Supply Management via Haver Analytics.
Note: Index is regressed on lags of itself and on lags of permits, claims, interest rate spread, and ISM delivery time.

A reasonable initial impression is that all variables, state and national, have at least one lagged value whose coefficient estimate is statistically significant, suggesting that all variables are potentially useful in forecasting the New York state coincident index to produce the leading index.

Moreover, as shown in Figures 6A–B, a close correspondence exists between the in-sample predictions and realizations, and the residuals show no obvious signs of the unmodeled serial correlation or heteroscedasticity. These results suggest that the main equation of New York's near-VAR model, that for the New York coincident index, performs well in-sample.

Additional evidence on the relative importance of the right-hand-side variables in the equation for the New York coincident index appears in Table 2. The table presents the results of multiple statistical tests that certain variables can be excluded from the regression because their estimated coefficients take values of zero. The null hypothesis is that all coefficients on the lagged values of the indicated variable are zero. A low p-value means that we cannot accept the null hypothesis and we should therefore not exclude the variable from the regression. In other words, a low p-value — say, less than 0.10 — means the variable is an important predictor for the coincident index and we should expect that variable to contribute in a statistically significant sense to the New York leading index.

TABLE 2
P–Values for Exclusion Tests in Near-VAR Equation for New York State Coincident Index
Full Sample Estimates, January 1982 to February 2016. Statistically significant results in red.

Statistical Test	P–Value
Panel A. Excluding NY building permits, NY unemployment claims, interest rate spread, and ISM delivery time	
Null hypothesis: All coefficients are zero	0.025
Panel B. Excluding NY building permits	
Null hypothesis: All coefficients on NY permits are zero	0.198
Panel C. Excluding NY unemployment claims	
Null hypothesis: All coefficients on NY unemployment claims are zero	0.064
Panel D. Excluding Interest Rate Spread	
Null hypothesis: All coefficients on the interest rate spread are zero	0.160
Panel E. Excluding ISM Delivery Time	
Null hypothesis: All coefficients on ISM delivery time are zero	0.260
Panel F. Excluding state variables: NY permits and NY claims	
Null hypothesis: All coefficients on NY local variables are zero	0.073
Panel G. Excluding national variables: Interest rate spread and ISM delivery time	
Null hypothesis: All coefficients on national variables are zero	0.320

The results suggest that, as a group, New York building permits, New York unemployment claims, the interest-rate spread, and the ISM delivery time contribute significantly to explaining variations in the New York state coincident index (Panel A). However, individually, on a variable-by-variable basis, only New York unemployment claims are statistically significant for the New York state coincident index (Panel C). Notably, New York state variables are statistically significant as a group (Panel F), but national variables are not (Panel G).

Subsample Results Suggest Importance of State Indicators Since 1990

The full-sample results for the exclusion tests suggested that state variables carry more importance than national ones for predicting New York's coincident index. We wondered whether these findings hold over the entire sample period or only over various subsamples. Figure 7 shows the results of broadening the exclusion tests by recomputing the p-values on rolling fixed-window samples of 20 years. Rolling the sample tells us whether the full-sample results apply generally over the entire sample period or only over specific periods.

The results suggest that New York state variables have been particularly important for explaining, and hence forecasting, the New York coincident index since about 1990. Notice that the p-values for the exclusion of state variables (Figure 7E), plotted at the last period of the subsamples, become very small in 2010 and remain so thereafter, suggesting that the data 20 years prior reflect the importance of the state variables over the national ones.

It is noteworthy that we find no low p-values for the interest rate spread, suggesting that this variable is not particularly helpful in forecasting New York's coincident index. At the same time, the rolling results for ISM delivery times, the second national variable in the model, are somewhat mixed, but on balance suggest they have little predictive power for the coincident index.

Taken together, the rolling results suggest that national variables as a group provide little predictive power over any sample period for New York's coincident index. By contrast, New York's state variables appear to carry statistically significant predictive ability since 1990.

FIGURE 7A–F
P–Values for Exclusion Tests in Rolling Fixed-Window Regressions
P–values for null hypothesis that coefficients are zero on all lags of variable. Low p–values imply rejection of the null. Fixed-window is 240 months. Data plotted at sample endpoint.

FIGURE 7A
Excluding New York State Building Permits

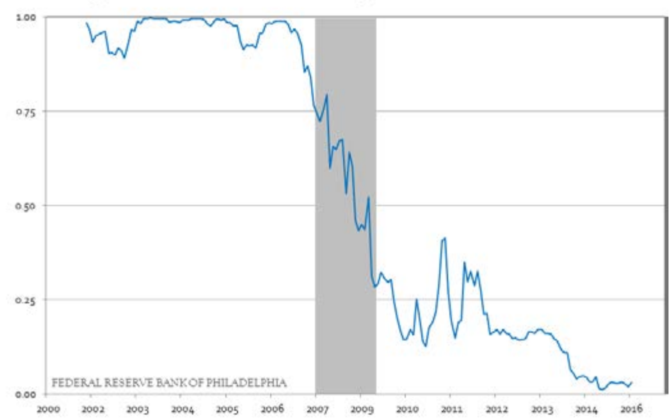


FIGURE 7B
Excluding New York State Unemployment Claims

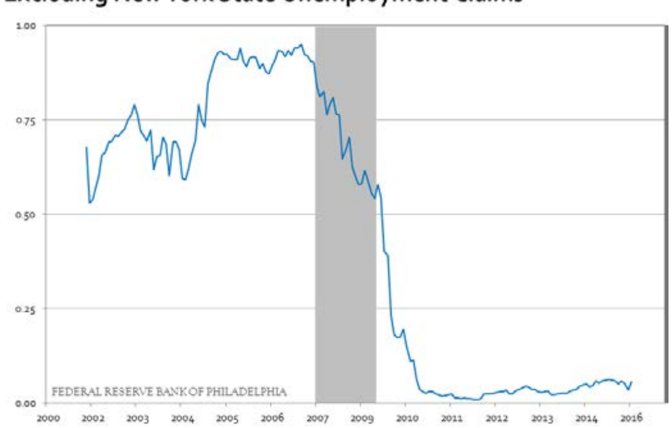


FIGURE 7C
Excluding Interest Rate Spread

FIGURE 7F
Excluding National Variables: Interest Rate Spread and ISM Delivery

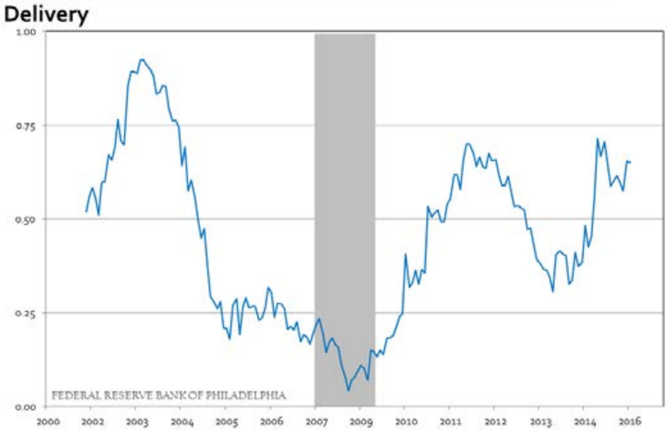
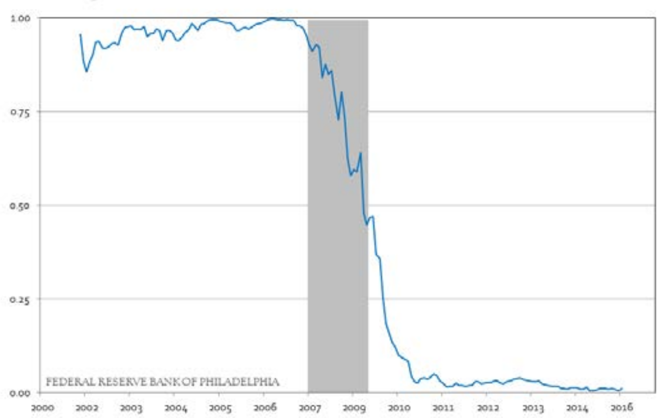


FIGURE 7D
Excluding ISM Delivery Time

FIGURE 7E
Excluding State Variables: Permits and Claims



Conclusions and Future Work

Our analysis of the relative contribution of the indicators included in the Philadelphia Fed's state leading index for New York points to the importance of state indicators over national ones in explaining the variation of our New York coincident index. We believe that our focus on the coincident index carries straightforward and obvious implications for the leading index because the latter is a six-month-ahead forecast of the coincident index, using a one-period near-VAR whose critical equation is the one we focused on.

Several avenues of future work are suggested by our findings for New York. First, we plan to extend the analysis to the remaining 49 states to see whether our New York findings can be generalized geographically. A second avenue for future work is to extend our approach to an out-of-sample environment, relying on the econometric literature previously cited. Third, we plan to check the robustness of our findings to data revisions, using real-time vintage data for the state coincident and leading indexes as well as the underlying indicator variables.

Notes

[1] We thank the state of New York Department of Labor for inquiring about the relative weights of the variables in our model, prompting this formal analysis and, we intend, further work in this area.

[2] Notice that we refer to the model as a near-vector autoregression and not a vector autoregression (VAR). The reason is that a VAR includes the same number of lagged values for each explanatory variable in each equation of the model. Our near-VAR, in contrast, uses four lagged values of the explanatory variables in the equation for the coincident index but only one lagged value in the remaining four equations.

[3] We construct the coincident index for each state using a mixed-observation frequency dynamic factor model estimated on (i) state-specific monthly nonfarm payroll employment, (ii) state-specific monthly average hours worked in manufacturing, (iii) the state-specific monthly unemployment rate, and (iv) state-specific quarterly wage and salary disbursements deflated by the consumer price index (U.S. city average). Each state's coincident index is retrended according to that state's real gross domestic product. Additional details can be found in Crone and Clayton-Mathews (2005) and on the Philadelphia Fed state coincident indexes web page, www.philadelphiafed.org/research-and-data/regional-economy/indexes/coincident. Additional information on the Philadelphia Fed's state leading indexes is available at www.philadelphiafed.org/research-and-data/regional-economy/indexes/leading. The background work on the state leading indexes can be found in Crone (2000) at https://www.philadelphiafed.org/-/media/research-and-data/publications/business-review/2000.

[4] See Clark and McCracken (2001) and Clark and West (2007) for leading examples of out-of-sample forecast evaluation. Diebold (2015) and Inoue and Kilian (2004) contain insightful discussions of the relative merits of in-sample and out-of-sample forecast evaluation.

References

Clark, Todd E., and Michael W. McCracken. "Tests of Equal Forecast Accuracy and Encompassing for Nested Models," *Journal of Econometrics,* 105 (2001), pp. 85–110.

Clark, Todd E., and Kenneth D. West. "Approximately Normal Tests for Equal Predictive Ability Accuracy in Nested Models," *Journal of Econometrics,* 138 (2007), pp. 291–311.

Crone, Theodore M. "A New Look at Economic Indexes for the States in the Third District," Federal Reserve Bank of Philadelphia *Business Review* November/December 2000, pp. 3–14.

Crone, Theodore M., and Alan Clayton-Matthews. "Consistent Economic Indexes for the 50 States," *Review of Economics and Statistics,* 87 (2005), pp. 593–603.

Diebold, F.X. "Comparing Predictive Accuracy Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold-Mariano Tests," *Journal of Business and Economic Statistics*, 33 (2015), pp. 1–24.

Inoue, A., and, L. Kilian. "In-Sample or Out-of-Sample Tests of Predictability: Which One Should We Use?" *Econometric Reviews,* 23 (2004), pp. 371–402.