Federal Reserve Bank of Cleveland Working Paper Series

BLS Payroll Revisions: Forecasting Recessions

Rory Quinlan and Roberto Pinheiro

Working Paper No. 25-26

December 2025

Suggested citation: Quinlan, Rory and Roberto Pinheiro. 2025. "BLS Payroll Revisions: Forecasting Recessions." Working Paper No. 25-26. Federal Reserve Bank of Cleveland. https://doi.org/10.26509/frbc-wp-202526.

Federal Reserve Bank of Cleveland Working Paper Series
ISSN: 2573-7953

BLS Payroll Revisions: Forecasting Recessions*

Rory Quinlan and Roberto Pinheiro†

Abstract

We investigate the behavior of BLS monthly revisions to payroll growth at turning points. We find some evidence corroborating claims by former BLS commissioners and market analysts that revisions around turning points tend to be procyclical and more serially correlated. Furthermore, we do see large revisions before turning points. However, the ability to use revisions to forecast business cycles' turning points seems limited. First, we do see lots of false positives: large revisions occur without a subsequent recession. Second, even within-sample, other indicators, such as initial jobless claims, the Chicago Fed National Activity Index, and the Aruoba-Diebold-Scotti Index, do a better job at detecting recessions. Finally, out-of-sample forecasting performance of revisions is poor.

Keywords: BLS monthly revisions, Business cycles, Forecast

JEL E32, E37, C53

Introduction

Market analysts' statements suggest that BLS payroll revisions may have some predictive power to forecast turning points. As stated by former BLS commissioner Erica Groshen: *"Revisions tend to be bigger and procyclical at turning points."* The rationale is that the imputation procedure used by the BLS is usually a good approximation during normal times, but tends to be off during turning points. As stated by Groshen: *"(...) an implicit imputation for the non-reporters is that they are exactly like the ones that did report. When you are not at an inflection point, then that is usually pretty good. (...) But at the inflection point, then you got a lot of companies that are changing from adding jobs to cutting jobs (...) Then you get a serial correlation on the revisions, which does not happen normally."*[1]

*We appreciate the helpful comments of Bruce Fallick and James Mitchell. Any remaining errors are our responsibility. The views expressed here are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Cleveland or the Federal Reserve Board.

†Rory Quinlan, Federal Reserve Bank of Cleveland, Rory.Quinlan@clev.frb.org and Roberto Pinheiro (corresponding author), Federal Reserve Bank of Cleveland, roberto.pinheiro@clev.frb.org

[1]From: Moody's Talks Inside Economics - "In defense of the BLS." `https://www.moodys.com/web/en/us/about/insights/podcasts/moodys-talks-inside-economics.html`, August 7, 2025.

Unfortunately, there is no literature looking at the power of monthly BLS payroll revisions to predict business cycle turning points. Most of the literature focuses on the ability to predict revisions themselves.[2] In this memo, we apply the methods designed by Estrella and Mishkin (1998) as well as the methods presented by Berge and Jordà (2011) in order to evaluate the predictive power of revisions. There are a few benefits to using these approaches. First, as shown by Negro (2001) and Pelàez (2007), non-linear models that focus directly on predicting turning points have a better track record predicting recessions in out-of-sample tests than linear models built to predict future values of economic variables. Second, the rationale presented by market analysts and highlighted by Erica Groshen's comments implies that recessions and booms are quite distinct from "normal times." This view is in contrast with the maintained assumption behind most forecasting models of economic time series (VAR, dynamic factors, etc.) that the underlying structure of the economy does not change from a recession to an expansion.[3] The downside of the approaches presented here is that recessions are rare events. As a result, in the full sample of BLS payroll revisions that we have (starting in 1965 if we include the entire database, 1979 if we avoid large redesigns early in the time series), there are only 6 recessions, including the COVID episode.[4]

Revisions are larger at turning points

We start by investigating the claim that revisions are larger and procyclical at turning points.[5] Figure 1 shows the change in levels between the first and the third reading for total non-farm payroll growth. The figure also includes the recession indicators from the NBER. In order to make the graph more readable, we focus on a quarterly frequency in which quarterly revisions represent the average of the monthly values. While we usually see large negative revisions close to turning points as well as during recessions, we see quite a few false positives as well. Furthermore, revisions were actually small during the great recession.

To test the ability of BLS monthly revisions to predict recessions, we follow Estrella and Mishkin (1998) and consider a one-to twelve-month-ahead probit in which the NBER recession indicator is the dependent variable and consider different variables that represent revisions and revision changes. Results are presented in Tables 1 and 3. Table 1 shows that the magnitude of the revision has no statistical significance in predicting the NBER indicator one to twelve months ahead. Since Groshen's statement focuses on large revisions, we built a dummy variable that

[2]There is a small literature that discusses benchmark revisions and their ability to forecast labor market performance (see Haltom et al. (2005)).

[3]For more details in this discussion, see Negro (2001).

[4]Since there are only 6 recessions in our sample, there is a heavy class imbalance which could lead to an inflation of certain performance measures such as accuracy and Brier score, to avoid this we use the AUROC which is invariant to class incidence.

[5]We obtained data on BLS revisions from the BLS website `https://www.bls.gov/web/empsit/cesnaicsrev.htm`. For more information, see Bureau of Labor Statistics (1979).

equals to 1 if revisions are in the bottom 10 percent of revisions, indicating a large downward revision. Results are presented in Table 2. As we can see, large downward revisions are associated with a higher likelihood of recession within the next 8 months, at least in the in-sample case. However, this result is quite sensitive to the sample used in the regression. As shown in Table 3, the statistical significance for horizons 2 to 8 months ahead relied heavily on the 1980 recession.

We also consider an alternative approach suggested by Berge and Jordà (2011) based on the receiver operating characteristic (ROC) curve. The ROC curve's goal is to find an index's ability to correctly identify the true underlying state of the economy.[6] This ability is judged by comparing the index's true positive rate – the ability to rightly call the underlying unobserved state – versus the false positive rate. The summary of all the trade-offs contained in the ROC curve and a commonly used measure of overall classification ability is the area under the ROC curve (AUROC). The closer to 1 the area is, the better the performance.

Figure 2 presents the ROC curve in which we use the change in levels between the first and the third reading for total non-farm payroll growth (Δ_{13}) in order to classify a month as expansion (NBER recession indicator equals 0) or recession (NBER recession indicator equals 1). As we see in Figure 2, the AUROC is close to 0.5, implying that Δ_{13}'s ability to correctly classify expansions and recessions is quite close to random chance. In this figure, we are considering concurrent indexes, i.e., the ability to correctly classify the current conditions as a recession or an expansion. However, considering lags from 1 to 12 months of Δ_{13} only marginally affects the AUROC.

In summary, the ability of the change in levels between the first and the third reading for total non-farm payroll growth to predict business cycle movements seems quite limited.

Revisions' serial correlation and turning points

We move to the second claim that revisions tend to be serially correlated closer to turning points. To evaluate this claim, we run an AR(1) on the change in levels between the first and the third reading for total non-farm payroll growth in an 18-month rolling window. In particular, we run the following specification in an 18-month rolling window:

$$Revision_t = \alpha + \beta_{AR} Revision_{t-1} + \varepsilon_t \tag{1}$$

where *Revision* is the 3rd - 1st reading for non-seasonally adjusted payroll employment growth, and ε_t is a white noise. Evidence of serial correlation is found if β_{AR} is statistically different from zero.

[6] Since the NBER's Business Cycle Dating Committee's announcements are released with a lag of 12 to 18 months, indicators of business conditions that may provide a more timely signal, even at a concurrent date, may be quite valuable.

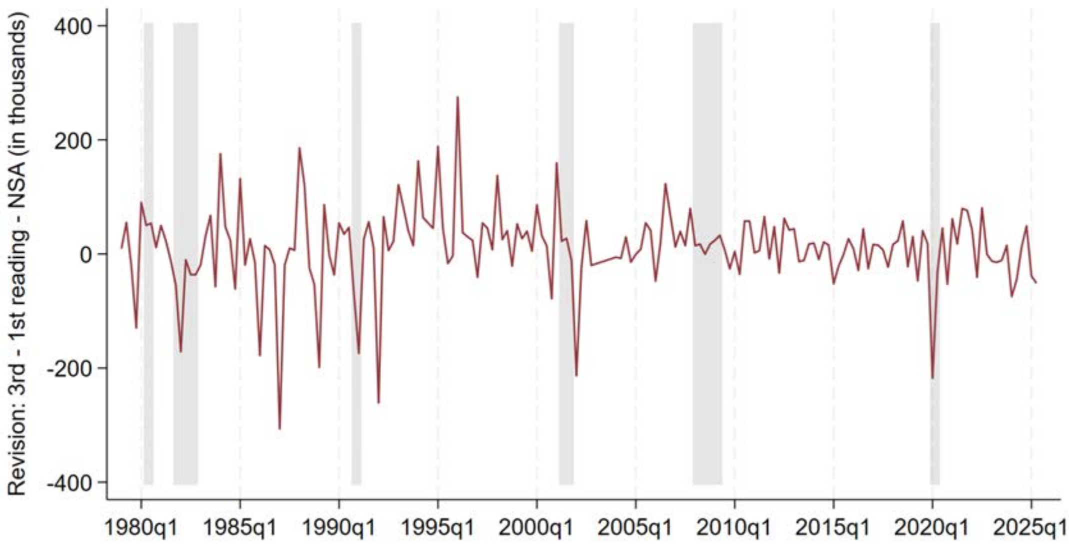


Figure 1: Revisions over the business cycle

Table 1: Probit for NBER recessions: Revisions as only explanatory variable – In-sample

	R_{t+1}	R_{t+2}	R_{t+3}	R_{t+4}	R_{t+5}	R_{t+6}
Δ_{13}	-0.0006	-0.0001	-0.0000	-0.0004	-0.0003	-0.0001
	(0.0007)	(0.0006)	(0.0007)	(0.0006)	(0.0006)	(0.0006)
N	555	555	554	553	552	551

	R_{t+7}	R_{t+8}	R_{t+9}	R_{t+10}	R_{t+11}	R_{t+12}
Δ_{13}	-0.0003	-0.0001	-0.0000	0.0000	-0.0000	0.0003
	(0.0006)	(0.0006)	(0.0004)	(0.0004)	(0.0004)	(0.0003)
N	550	549	548	547	546	545

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

We plot the coefficient of the regression β_{AR} against the NBER recession bars in Figure 3. The figure does corroborate Groshen's statement that around turning points revisions tend to be procyclical and we observe more serial correlation. In particular, near recessions, revisions tend to turn from positive to negative, as seen in Figure 1, delivering a negative β_{AR}. Furthermore, β_{AR} is more likely to be statistically different from zero during recessions, indicating serial correlation.

Table 2: Probit for NBER recessions: Large downward revision indicator as only explanatory variable – In-sample

	R_{t+1}	R_{t+2}	R_{t+3}	R_{t+4}	R_{t+5}	R_{t+6}
$D_{\text{Large Down}}$	0.5543***	0.3878*	0.3866*	0.5521**	0.5510**	0.4687*
	(0.2058)	(0.2091)	(0.2211)	(0.2268)	(0.2458)	(0.2564)
N	559	559	558	557	556	555

	R_{t+7}	R_{t+8}	R_{t+9}	R_{t+10}	R_{t+11}	R_{t+12}
$D_{\text{Large Down}}$	0. 0.4832*	0.3961*	0.2061	0.0986	-0.0208	-0.3180
	(0.2579)	(0.2280)	(0.2033)	(0.2159)	(0.2584)	(0.2525)
N	554	553	552	551	550	549

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table 3: Probit for NBER recessions: Large downward revision indicator as only explanatory variable – In-sample (1981m9-)

	R_{t+1}	R_{t+2}	R_{t+3}	R_{t+4}	R_{t+5}	R_{t+6}
$D_{\text{Large Down}}$	0.5474**	0.2500	0.2618	0.3859	0.2862	0.1744
	(0.2219)	(0.2473)	(0.2570)	(0.2516)	(0.2282)	(0.2251)
N	527	527	526	525	524	523

	R_{t+7}	R_{t+8}	R_{t+9}	R_{t+10}	R_{t+11}	R_{t+12}
$D_{\text{Large Down}}$	0.2018	0.2147	0.0850	0.0981	0.1116	-0.2614
	(0.2209)	(0.2270)	(0.2188)	(0.2639)	(0.2715)	(0.3086)
N	522	521	520	519	518	517

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

However, there are caveats. First, if we focus on the coefficient for β_{AR}, Figure 3 shows quite a few false positives: we do see large β_{AR} coefficients outside either turning points or recessions. Second, if we focus on the episodes in which β_{AR} becomes statistically different from zero, we do see that β_{AR} is more likely to be statistically significant during recessions. However, that usually happens after the recession has started. Moreover, we miss a few recessions altogether (2001 and 2008, for example), while identifying some false positives (2024m4 and 2003m8).[7]

We follow Estrella and Mishkin (1998) and consider a one-to twelve-month-ahead probit in

[7] However, 2003 is a particularly complicated year due to methodological changes in the data.

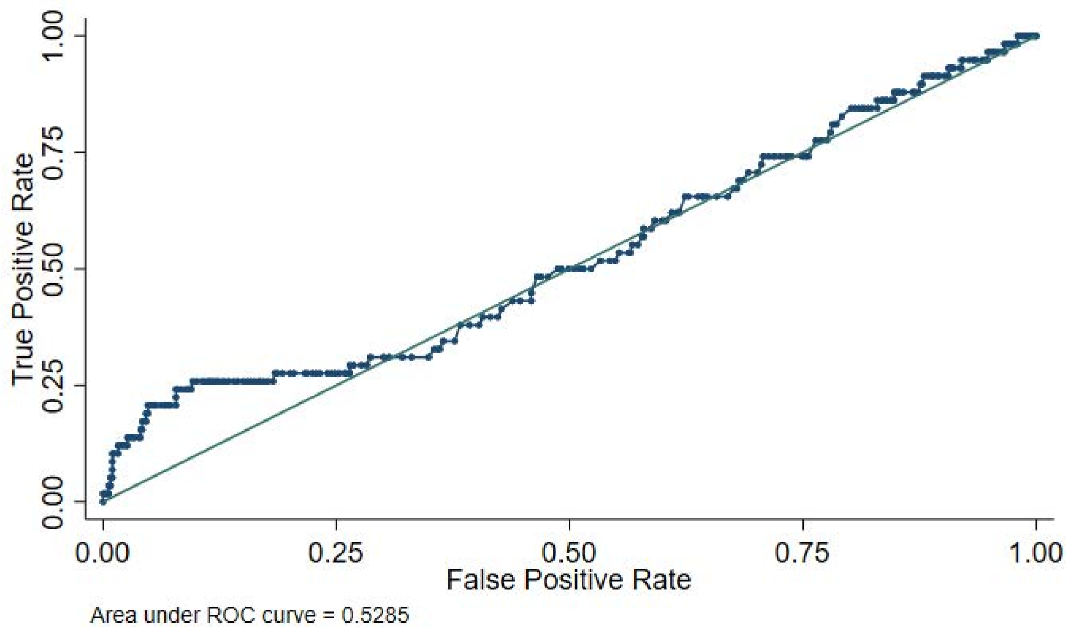


Figure 2: ROC curve – Revisions as classifier

which the NBER recession indicator is the dependent variable and β_{AR} is the only explanatory variable. We start with the in-sample case, following Mishkin's (1998) Table 1.[8] Our results are presented in Table 5. As we can see, β_{AR} is associated with a higher likelihood of recession within the next 6 months, at least in the in-sample case. The negative coefficient for β_{AR} indicates that negative β_{AR}s, i.e, β_{AR}s associated with significant swings from positive to negative revisions, are correlated with a higher probability of recessions.

We also consider an alternative approach suggested by Berge and Jordà (2011) based the ROC curve. Figure 4 presents the results for β_{AR} and other competing measures that are readily available. In this figure, we are considering concurrent indexes, i.e., the ability to correctly classify the current conditions as a recession or an expansion. As we can see, while β_{AR} does a good job at classifying the current economic conditions, it underperforms compared to other readily available indicators, such as the average initial claims in the period. Furthermore, other

[8] We consider an analysis in-sample when we use the entire database available to run the analysis and consider the within-sample forecasting performance. Out-of-sample analysis – also known as pseudo out-of-sample analysis – runs the model in a subsample, creates out-of-sample predictions, and compares forecasted results to data points not used in the model estimation phase.

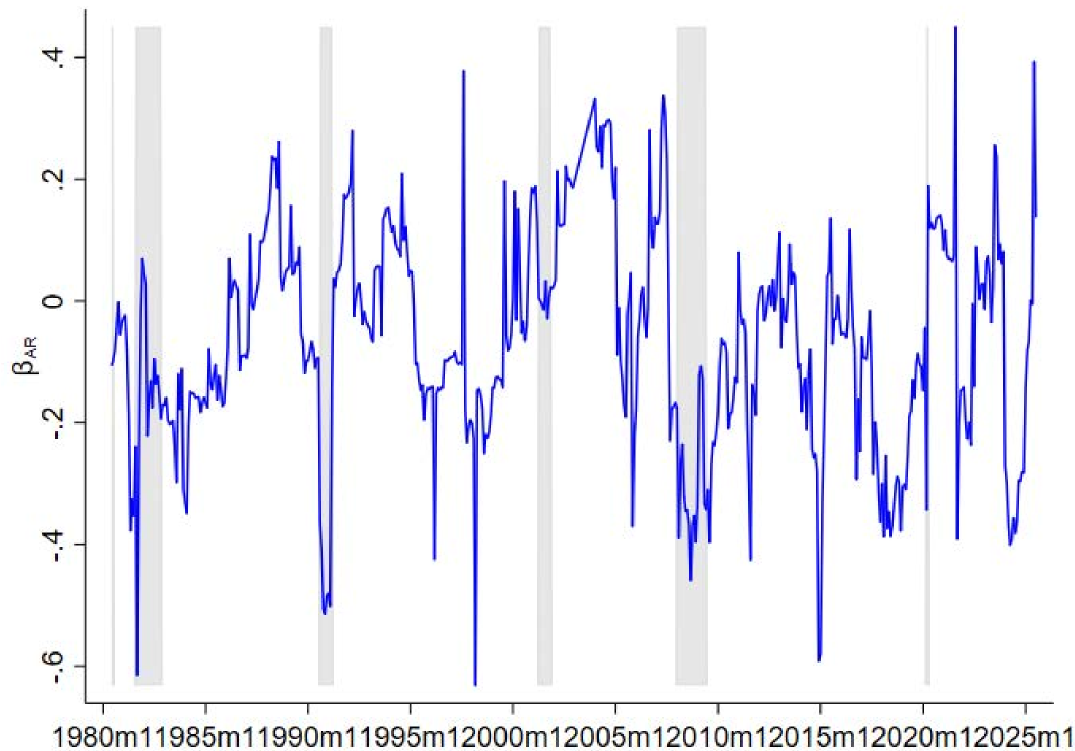


Figure 3: Autocorrelation of revisions over the business cycle

indexes constructed by the Federal Reserve System, such as the Chicago Fed National Activity Index[9] (CFNAI) and the Aruoba-Diebold-Scotti (ADS) Business Conditions Index from the Philadelphia Fed,[10] provide a much better classification ability, and are recommended by Berge and Jordà (2011). However, as seen in Table 4, classification ability of all these measures deteriorates rapidily when used to predict turning points into the future. Within a year ($h = -12$), they are just slightly better than a coin flip (AUROC=0.5).

[9]The Chicago Fed National Activity Index is a weighted average of 85 existing monthly indicators of growth in national economic activity drawn from four broad categories of data: 1) production and income, 2) employment and unemployment, and hours; 3) personal consumption and housing, and 4) sales, orders and inventories. For more information please visit `https://www.chicagofed.org/research/data/cfnai/about`. We retrieved monthly CFNAI data from Haver Analytics. However, the data is publicly available at `https://www.chicagofed.org/research/data/cfnai/current-data`.

[10]The Aruoba-Diebold-Scotti Business Conditions Index is designed to track real business conditions at high observation frequency. It has as underlying indicators the seasonally adjusted weekly initial jobless claims, monthly payroll employment, monthly industrial production, monthly personal income less transfer payments, monthly real manufacturing and trade sales; and quarterly real GDP. The average value of the ADS is zero. Progressively bigger positive values indicate progressively better-than-average conditions, whereas progressively negative values indicate progressively worse-than average conditions. For more information, see Aruoba et al. (2009). To access the data, visit `https://www.philadelphiafed.org/surveys-and-data/real-time-data-research/ads`.

Furthermore, when we evaluate the ability of β_{AR} out of sample, results are disappointing. We follow Berge and Jordà (2011) and consider an initial training sample with a window that spans June 1980 to December 1998. We then use a rolling window to estimate a model for each forecast horizon from 1 to 24 months ahead. Considering a threshold of 0.5 in order to consider that the model is predicting the beginning of a recession, a model with β_{AR} misses all 3 turning points in the out-of-sample period. Using the ROC curve methodology, Figure 5 shows that β_{AR}'s out-of-sample prediction is better than random chance only in the first 4 months. Furthermore, compared to the performance of the Conference Board's Index of Leading Indicators, β_{AR}'s performance is worse throughout the entire horizon (comparing Figure 5 against Berge and Jordà's (2011) Figure 9).

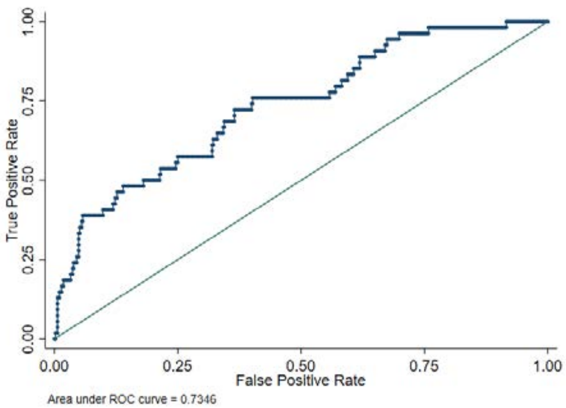
Conclusion

In summary, the ability to use BLS monthly revisions of payroll growth to forecast business cycles' turning points is limited. First, we see lots of false positives: large revisions occur without a subsequent recession. Second, even within-sample, other indicators, such as initial jobless claims, the Chicago Fed National Activity Index, and the Aruoba-Diebold-Scotti Index do a better job at detecting recessions. Finally, out-of-sample forecasting performance of revisions is poor.

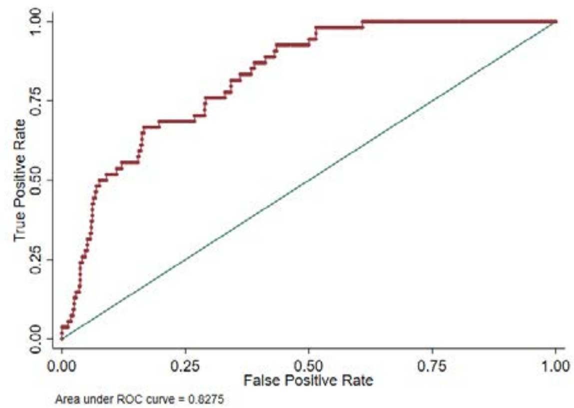
Table 4: Classification abilities of current business cycle

	β_{AR}	Initial Claims	CFNAI	ADS
$h = -12$	.543	.519	.620	.619
	(.043)	(.043)	(.040)	(.042)
$h = 0$	.735	.895	.957	.985
	(.036)	(.018)	(.019)	(.008)

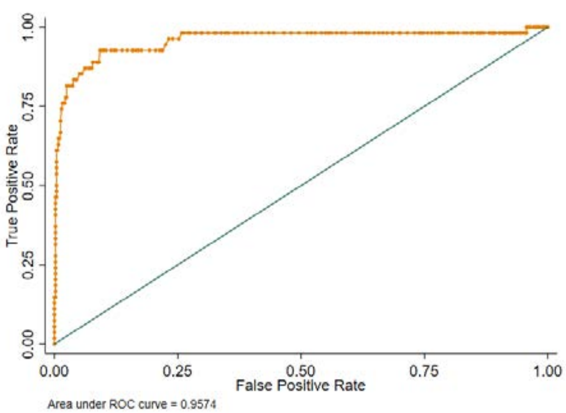
Note: This table reports area under the ROC curve (AUROC) for h periods out. Bootstrapped standard errors are reported in parenthesis.



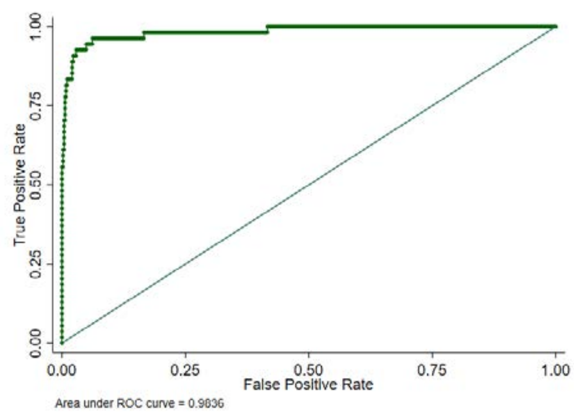
(a) β_{AR}



(b) Initial Claims



(c) Chicago Fed National Activity Index



(d) Aruoba-Diebold-Scotti Business Conditions Index

Figure 4: ROC curves – Alternative indexes

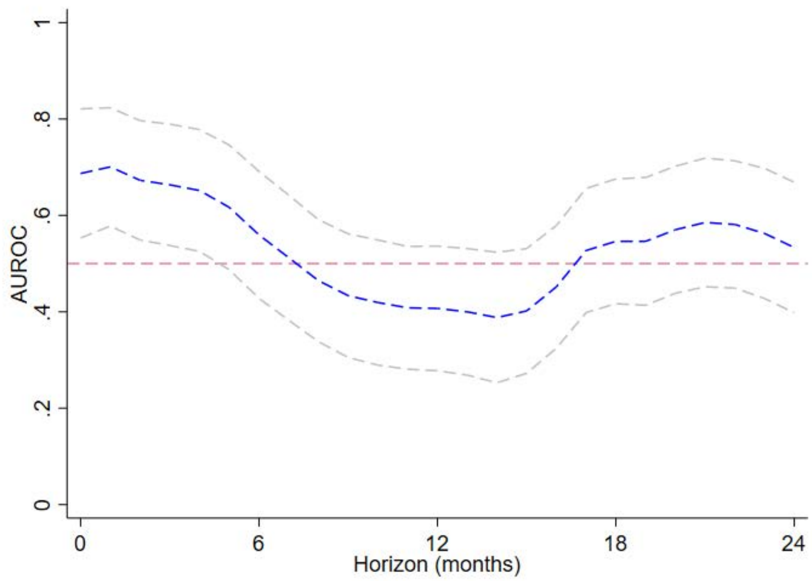


Figure 5: β_{AR}'s AUROC – Out of sample

References

Aruoba, S. Borağan, Francis X. Diebold, and Chiara Scotti (2009). "Real-Time Measurement of Business Conditions." *Journal of Business & Economic Statistics*, 27(4), pp. 417–427. doi:10.1198/jbes.2009.07205. URL https://doi.org/10.1198/jbes.2009.07205.

Berge, Travis J. and Òscar Jordà (2011). "Evaluating the Classification of Economic Activity into Recessions and Expansions." *American Economic Journal: Macroeconomics*, 3(2), pp. 246–277. doi:10.1257/mac.3.2.246. URL https://doi.org/10.1257/mac.3.2.246.

Bureau of Labor Statistics (1979). "Nonfarm Payroll Employment: Revisions between over-the-month estimates, 1979-present." URL https://www.bls.gov/web/empsit/cesnaicsrev.htm.

Estrella, Arturo and Frederic S. Mishkin (1998). "Predicting U.S. Recessions: Financial Variables as Leading Indicators." *The Review of Economics and Statistics*, 80(1), pp. 45–61. doi:10.1162/003465398557320. URL https://doi.org/10.1162/003465398557320.

Haltom, Nicholas L., Vanessa D. Mitchell, and Ellis W. Tallman (2005). "Payroll Employment Data: Measuring the Effects of Annual Benchmark Revisions." *Economic Review - Federal Reserve Bank of Atlanta*, 90(2), pp. 1–23. URL https://fraser.stlouisfed.org/title/economic-review-federal-reserve-bank-atlanta-884/second-quarter-2005-601624.

Negro, Marco del (2001). "Turn, turn, turn: Predicting turning points in economic activity." *Economic Review - Federal Reserve Bank of Atlanta*, 86(2), pp. 1–12. URL https://fraser.stlouisfed.org/title/economic-review-federal-reserve-bank-atlanta-884/second-quarter-2001-601608.

Pelàez, Rolando F. (2007). "Ex ante forecasts of business-cycle turning points." *Empirical Economics*, 32(1), pp. 239–246. doi:10.1007/s00181-006-0083-4. URL https://doi.org/10.1007/s00181-006-0083-4.

Table 5: Probit for NBER recessions: Revisions' serial correlation as only explanatory variable – In-sample

	R_{t+1}	R_{t+2}	R_{t+3}	R_{t+4}	R_{t+5}	R_{t+6}
β_{AR}	-2.86***	-2.55***	-2.44***	-2.26***	-1.97**	-1.53*
	(0.62)	(0.71)	(0.81)	(0.88)	(0.90)	(0.87)
N	541	540	539	538	537	536
	R_{t+7}	R_{t+8}	R_{t+9}	R_{t+10}	R_{t+11}	R_{t+12}
β_{AR}	-1.18	-0.86	-0.67	-0.59	-0.51	-0.42
	(0.86)	(0.87)	(0.87)	(0.83)	(0.81)	(0.82)
N	535	534	533	532	531	530

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table 6: Probit for NBER recessions: Revisions' serial correlation and 10-yr-3-month Treasury spread – In-sample

	R_{t+1}	R_{t+2}	R_{t+3}	R_{t+4}	R_{t+5}	R_{t+6}
β_{AR}	-2.91***	-2.43***	-2.20***	-1.99**	-1.73*	-1.32
	(0.66)	(0.74)	(0.83)	(0.90)	(0.95)	(0.96)
SPREAD	0.14*	0.05	-0.02	-0.10	-0.16	-0.21*
	(0.08)	(0.09)	(0.09)	(0.10)	(0.11)	(0.12)
N	526	525	524	523	522	521
	R_{t+7}	R_{t+8}	R_{t+9}	R_{t+10}	R_{t+11}	R_{t+12}
β_{AR}	-0.98	-0.65	-0.44	-0.36	-0.30	-0.23
	(0.99)	(1.04)	(1.05)	(0.98)	(0.94)	(0.93)
SPREAD	-0.25*	-0.30**	-0.34**	-0.37**	-0.41***	-0.47***
	(0.13)	(0.14)	(0.14)	(0.15)	(0.15)	(0.16)
N	520	519	518	517	516	515

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$