

Recession Detection Using Classifiers on the Anticipation-Precision Frontier

Pascal Michaillat

December 2025

This paper develops an algorithm for detecting US recessions in real time. The algorithm constructs hundreds of millions of recession classifiers by combining unemployment and vacancy data. Classifiers are then selected to avoid both false negatives (missed recessions) and false positives (nonexistent recessions). The selected classifiers are perfect in a statistical sense: they identify all 15 historical recessions in the 1929–2021 training period without any false positives. By further selecting classifiers that lie on the high-precision segment of the anticipation-precision frontier, the algorithm delivers early detection without sacrificing accuracy. On average between 1929 and 2021, the selected classifier ensemble signals recessions 2.1 months after their true onset, with a standard deviation of detection errors of 1.8 months. The classifier ensemble is much faster than the NBER Business Cycle Dating Committee: between 1979 and 2021, the committee takes on average 6.3 months to determine recession starts, while the classifier ensemble only takes 1.2 months. Applied to September 2025 data, the classifier ensemble gives a 64% probability that the US economy has entered a recession. A placebo test and backtests confirm the algorithm's reliability.

University of California, Santa Cruz. I thank Clement Bohr, Yuriy Gorodnichenko, Manfred Keil, Michael Leung, Adam McCloskey, Jeremy Piger, Mahdi Rastad, Emmanuel Saez, Dick Startz, and Eduardo Zambrano for helpful comments.

1. Introduction

Detecting recessions in a timely manner is critical for devising an effective policy response. Yet the official declaration of recessions by the Business Cycle Dating Committee of the National Bureau of Economic Research occurs with a substantial delay: sometimes as late as 11 months after the recession's onset (table A1). As the NBER (2008b) explains:

The committee's approach to determining the dates of turning points is retrospective. We wait until sufficient data are available to avoid the need for major revisions. In particular, in determining the date of a peak in activity, and thus the onset of recession, we wait until we are confident that, even in the event that activity begins to rise again immediately, it has declined enough to meet the criterion of depth. As a result, we tend to wait to identify a peak until many months after it actually occurs.

Because of this delay, waiting for the NBER to announce a recession before acting is not practical for policymakers—or for businesses and households.

To remedy this issue, several algorithms have been designed to detect US recessions in real time, using a variety of data and methods (Hamilton 2011). Among the most effective approaches for real-time recession detection are threshold rules based on unemployment data (Schanep 2008; Hatzius and Stehn 2012; Sahm 2019; Sun, Feng, and Hu 2021; Philips 2024; O'Trakoun and Scavette 2025; Garimella, Jorda, and Singh 2025). Indeed, Crump, Giannone, and Lucca (2020a,b) show that unemployment data, combined with simple threshold rules, provide a more reliable signal of US recessions than other detection methods. The logic behind unemployment-threshold rules is simple: unemployment always goes up in recessions, so a recession can be detected when the unemployment rate increases sharply. In that way, the unemployment rate measures well the latent state of the economy. A famous example is the Sahm (2019) rule, which takes the difference between the 3-month average of the unemployment rate and its 12-month minimum and detects a recession whenever the difference crosses a threshold of 0.50pp.

However, the unemployment rate is only a noisy measure of that latent state. Another measure is the job vacancy rate. Indeed, recessions feature not only an increase in the unemployment rate but also a decline in the vacancy rate as the economy moves along the Beveridge curve (see figure 1). Therefore, by combining data on unemployment and job vacancies, Michaillat and Saez (2025) obtain a recession indicator that is less noisy than unemployment-based indicators. Thanks to the reduced noisiness, the detection threshold can be lowered once both unemployment and vacancy data are used. The rule that they construct (Michez rule) detects recessions faster than the Sahm rule: with an average delay of 1.2 months instead of 2.7 months, and a maximum delay of 3 months instead of 7 months. It is also more robust than the Sahm rule: it identifies the 15 recessions that

occurred since 1929 without false positives, while the Sahm rule produces several false positives before 1960.

While the Sahm rule, Michez rule, and other threshold rules choose their thresholds optimally, they filter the data in an arbitrary way. As such, they may not extract the most information possible from unemployment and vacancy data. This paper aims to develop optimized recession classifiers, which extract the most information possible from labor market data. By optimally filtering the data and selecting the threshold, recessions can be detected more rapidly and accurately than with the Michez or Sahm rule.

This paper aims to extract a better recessionary signal by considering a larger number of ways to process two data series, rather than by processing a larger set of data. For example, Chen, Iqbal, and Lai (2011) consider 141 time series to forecast US recessions. Stock and Watson (2014) look at 270 series to estimate US business cycle turning points. Ng (2014) considers 1,500 series (132 distinct series and their lags) to forecast US recessions at various horizons. Fornaro (2016) considers 133 series from the dataset constructed by McCracken and Ng (2016) to forecast US recessions. My motivation is that unemployment and vacancy data capture recessions almost perfectly (as is apparent in figure 1), so instead of looking at other, less informative data sources, I try to make progress by determining the best possible lens through which to look at the data. The focus on a small number of predictive variables fits with the assessment by Piger (2020) that a narrow set of variables has the best classification performance for US business cycles.

First, I construct 95,832 recession indicators from the unemployment and vacancy rates. These indicators are constructed by processing unemployment and vacancy data in many different ways. The first step is to smooth the data using simple or exponentially weighted moving averages, with many different smoothing parameters. The second step is to identify turning points by comparing smoothed data to their extrema over periods of many different lengths, yielding many increases for unemployment and decreases for vacancies. The increases and decreases are also subject to Box-Cox transformations to cover the many possibilities from level changes to percentage changes. The 4,356 unemployment and vacancy indicators are finally combined into single recession indicators through weighted averages or weighted min-max, using many different weights.

Next, I build hundreds of millions of recession classifiers by combining the 95,832 recession indicators with many different recession thresholds. Among this vast number of recession classifiers, I then select those that produce no false negatives (actual recessions missed by the classifiers) and no false positives (non-recessions mistakenly detected by the classifiers) between 1929 and 2021—the longest period for which unemployment data, vacancy data, and recession dates are available. The selected classifiers are perfect in a statistical sense: they detect the 15 recessions that occurred between 1929 and 2021 without any false positives. There are many perfect classifiers: 4,481,622 in total.

The next task is to evaluate these perfect classifiers and select the best ones to detect recessions. Standard evaluation methods, like ROC curves, rank classifiers based on the amount of false positives and false negatives that they generate. These methods are unsuitable here given that so many classifiers are perfect, in the sense that they produce neither false positives nor false negatives. Instead, I compute the mean and standard deviation of detection errors for each classifier. I then select classifiers on the low-mean, low-standard-deviation frontier—the anticipation-precision frontier. For a given mean detection error, no classifier is more precise than a classifier on the frontier; for a given precision, no classifier detects recessions earlier. The anticipation-precision frontier allows me to identify 248 classifiers that optimize anticipation and precision.

A policymaker could pick any classifier on the anticipation-precision frontier, based on their need for early recession notice and tolerance for errors. A policymaker who can implement stabilization policies rapidly would pick a classifier with little anticipation but high precision. A policymaker who requires advance notice to enact stabilization policies would settle for a classifier with higher anticipation but worse precision.

Since policymakers' preferences are not available, I pick a collection of classifiers on the frontier and average their predictions. I select all classifiers with a standard deviation of detection errors below 3 months. For the selected classifiers, the probability that the actual recession starts 6 months before or after detection is less than 5% (assuming normal errors). In other words, the 95% confidence interval for the recession's start date is 12 months wide or less.

In all, I select 11 classifiers from the high-precision segment of the frontier. On average, over 1929–2021, the 11 classifiers detect recessions 2.1 months after their true onset, with a standard deviation of detection errors of 1.8 months. The performance is quite remarkable given that unemployment and vacancy data are constructed from a patchwork of sources before 1948 and 2001, respectively (Petrosky-Nadeau and Zhang 2021). The 11 classifiers are much faster than the NBER Dating Committee: while the committee takes on average 6.3 months to determine recession starts between 1979 and 2021, the 11 classifiers only take 1.2 months over the same period.

From this ensemble of 11 classifiers, I compute a recession probability that indicates the likelihood that a recession has begun. Whenever an individual recession indicator crosses its threshold, I infer the probability that the recession has started based on the distribution of its detection error. If for instance the classifier is on time on average, the recession has started with 50% probability when the classifier is activated—assuming a symmetric detection error. If the classifier is early on average, the probability is less than 50%; if it is late on average, the probability is more than 50%. In the months following detection, the probability converges to 1 along the detection error's cumulative distribution function. I then average the probability that the recession has started across the 11 classifiers and

obtain a single recession risk.

I apply the method to current data to obtain a real-time assessment of recession risk. As of September 2025, the recession probability given by the 11 classifiers is 64%, suggesting a high likelihood that a recession has begun at this point in time. This is in turn due to the noticeable decrease in the number of job vacancies and increase in the number of job seekers since the middle of 2022.

To ensure the algorithm's success is not due to overfitting, I run a placebo test. I ask the algorithm to detect a series of events that is random, as frequent as recessions, but not connected in any way to the US economy: the 15 deaths of US first ladies between 1929 and 2021. The algorithm is unable to detect these placebo events. While it identifies recessions with a precision of under 2 months, the standard deviation of its detection errors for the first-lady deaths exceeds 40 years. This failure in the first-lady placebo test confirms that the model identifies genuine economic signals rather than spuriously fitting the data.

In addition, I backtest the algorithm by shortening the training period and evaluating the algorithm's performance on subsequent out-of-sample recessions. Through backtesting, I find that the algorithm is reliable. Classifier ensembles selected on limited historical data (1929–2014, 1929–2004, 1929–1994, 1929–1984, and 1929–1974) all detect the Great Recession by mid-2008. Furthermore, all classifiers from these ensembles detect all out-of-sample recessions without any false positives. The classifier ensembles selected in these backtests give a current recession probability between 50% and 67%—confirming the recession risk detected by the classifier ensemble selected on the 1929–2021 sample. The algorithm only shows some cracks once I restrict training to 1929–1964 data (7 in-sample recessions): all the selected classifiers detect the 8 out-of-sample recessions, but for the first time, one of the classifiers produces a false positive.

Unlike the present algorithm, the NBER Dating Committee relies mostly on product market metrics (production, sales, consumption) to date recessions. The committee states that they do not look at unemployment because it is a lagging and noisy variable, and they have never mentioned looking at job vacancies (NBER 2001). Accordingly, I examine whether it is possible to detect recessions earlier and more accurately by applying the algorithm to some of the product market data used by the NBER (2025). I find, however, that the combination of unemployment and vacancy rates outperforms industrial production in providing early and accurate recession signals; in fact the unemployment rate alone outperforms industrial production. This finding suggests that official recession dating practices might benefit from a greater emphasis on unemployment and vacancies.

2. Data

This section presents the data on US recessions, unemployment, and job vacancies used in the paper. The data cover April 1929 to September 2025, which is the longest period for

which the data are available (Michaillat and Saez 2025).

2.1. Recession dates

This paper aims to develop an algorithm to detect recessions that is timely and fully automated. To build the algorithm, I compare the number of recessions that various classifiers detect and the detection dates to the number of official US recessions and their start dates.

US recessions are officially identified by the Business Cycle Dating Committee of the NBER (2023). The NBER (2024) identifies the peaks and troughs of US business cycles by looking holistically at numerous macroeconomic variables. By convention, the first month of a recession is the month following the peak and the last month of a recession is the month of the trough.

Between April 1929 and September 2025, the NBER (2023) identifies 15 recessions. The NBER-dated recessions are displayed in Figure 1.

Critically, recession start dates are announced many months after recessions have actually started (table A1). For instance, the NBER did not announce until December 2008 that the previous business cycle peak had occurred in December 2007 and therefore that the Great Recession had started in January 2008. On average, between 1979 and 2021, the NBER announces recession starts 6.3 months after a recession's onset.

2.2. Unemployment rate

Between April 1929 and December 1947, I use the monthly unemployment rate constructed by Petrosky-Nadeau and Zhang (2021). They extrapolate Weir (1992)'s annual unemployment series to a monthly series using monthly unemployment rates compiled by the NBER.

Between January 1948 and September 2025, the unemployment rate is computed in the usual way: it is the number of jobseekers measured by the BLS (2025d) from the Current Population Survey (CPS), divided by the civilian labor force measured by the BLS (2025a) from the CPS. This is the standard, official measure of unemployment, labelled U3 by the BLS (2023).

The unemployment rate used in the analysis is plotted in Figure 1. It is countercyclical, rising sharply at the onset of all recessions.

2.3. Vacancy rate

Between April 1929 and December 1950, the vacancy rate is based on the help-wanted index created by the Metropolitan Life Insurance Company (MetLife). This index aggregates help-wanted advertisements from newspapers across major US cities. It is considered a

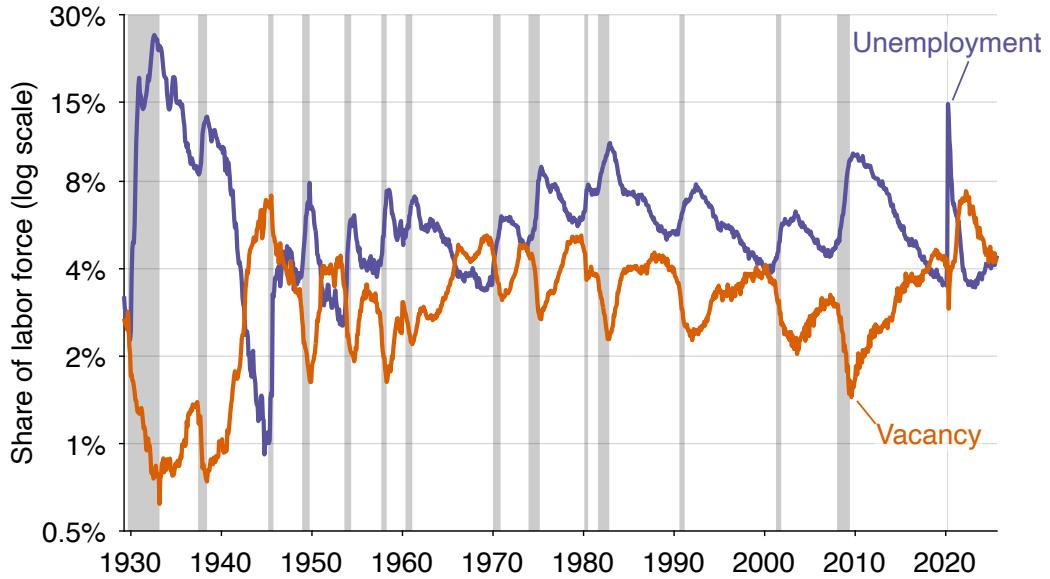


FIGURE 1. Monthly US unemployment and vacancy rates, April 1929–September 2025

The unemployment rate is computed from data produced by Petrosky-Nadeau and Zhang (2021) and the BLS (2025d,a). The vacancy rate is computed from data produced by Petrosky-Nadeau and Zhang (2021), Barnichon (2010), and the BLS (2025a,b). Shaded areas indicate recessions dated by the NBER (2023).

reliable proxy for job vacancies (Zagorsky 1998). The MetLife index is scaled to align with Barnichon (2010)'s vacancy rate at the end of 1950, effectively translating the index into a vacancy rate.¹

Between January 1951 and December 2000, I use the vacancy rate produced by Barnichon (2010). This series is based on the Conference Board's help-wanted advertising index, adjusted to account for the shift from print advertising to online advertising in the 1990s. The Conference Board index aggregates help-wanted advertising in major metropolitan newspapers in the United States. It serves as a reliable proxy for job vacancies (Abraham 1987; Shimer 2005). The Conference Board index is scaled to align with the JOLTS vacancy rate in 2001, effectively translating the index into a vacancy rate.

Between January 2001 and September 2025, I compute the vacancy rate as the number of job openings measured by the BLS (2025b) from the Job Openings and Labor Turnover Survey (JOLTS), divided by the size of the labor force measured by the BLS (2025a) from the CPS. To best align labor force and vacancy data, I follow Michaillat and Saez (2024b, 2025) and shift forward by one month the number of job openings from JOLTS. For instance, I assign to April 2025 the number of job openings that the BLS assigns to March 2025. The

¹Petrosky-Nadeau and Zhang (2021) produce a vacancy series that starts in 1919 and an unemployment series that starts in 1890. I only begin the analysis in April 1929, however, because there are some limitations with the prior data (Michaillat and Saez 2024b).

motivation for this shift is that the number of job openings from the JOLTS applies to the last business day of the month (Monday 31 March 2025), while the labor force from the CPS applies to the Sunday–Saturday week including the 12th of the month (Sunday 6 April 2025–Saturday 12 April 2025) (BLS 2020, 2024a). So the number of job openings applies to a day that is closer to the next month’s CPS reference week than to the current month’s CPS reference week.

I then splice the three vacancy series to create a continuous vacancy rate covering April 1929–September 2025. The vacancy rate is plotted in Figure 1. It is procyclical, dropping sharply at the onset of all recessions.

2.4. Availability and revisions of labor market data

The unemployment and vacancy data required to apply the algorithm in any given month are released in the first week of the following month, usually on a Tuesday for the JOLTS data and on a Friday for the CPS data (BLS 2024c). Thus, the algorithm can be applied in real time.

The recession probability constructed in real time might not be its final value because the unemployment and vacancy data are revised after their initial release. The number of job openings released by the BLS (2025b) is preliminary and updated one month after its initial release, to incorporate additional survey responses received from businesses and government agencies (BLS 2024b). Additionally, the BLS revises the prior five years of CPS and JOLTS data each year at the beginning of January, to account for revisions to seasonal factors, population estimates, and employment estimates (BLS 2024b, 2025c). Yet, revisions to labor market data are generally minimal, especially compared to GDP revisions, so the information provided in real time should be “almost indistinguishable” from the information provided in the final version (Crump, Giannone, and Lucca 2020a).

3. Construction of the recession indicators

This section constructs 95,832 recession indicators by combining the unemployment and vacancy data. The indicators are created in 4 steps: smoothing the data, detecting turning points, scaling variations, and combining single-variable indicators. Later I will use these indicators together with appropriate thresholds to detect recessions.

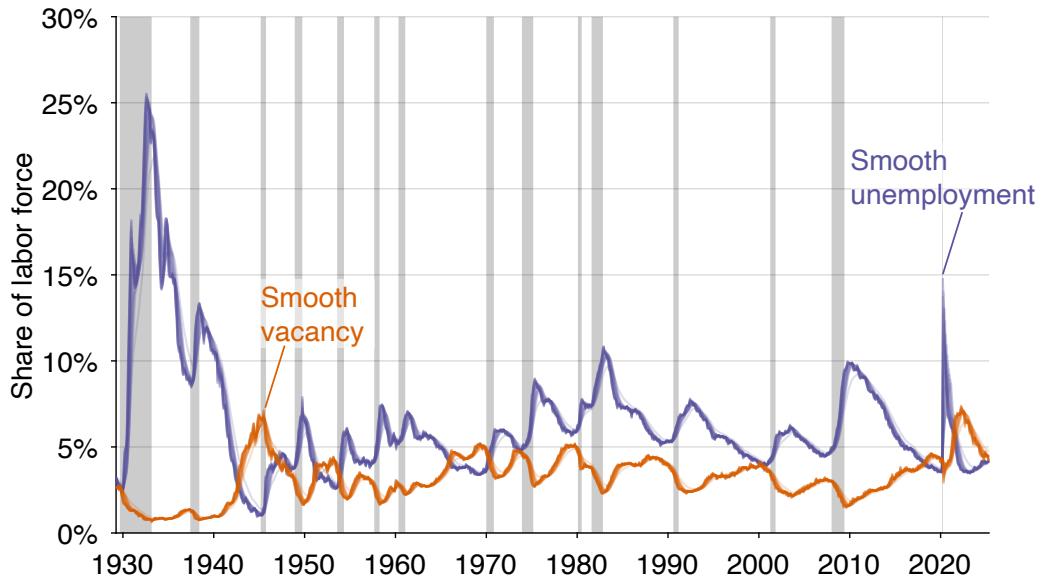


FIGURE 2. Smoothed US unemployment and vacancy rates, April 1929–September 2025

The smoothed unemployment rate is computed from the unemployment rate in figure 1 and the smoothing algorithms in equations (1) and (3). The smoothed vacancy rate is computed from the vacancy rate in figure 1 and the smoothing algorithms in equations (2) and (4). Shaded areas indicate recessions dated by the NBER (2023).

3.1. Smoothing the data

I smooth data by moving average. I first compute simple trailing averages of the unemployment rate:

$$(1) \quad \bar{u}(t) = \frac{\sum_{k=0}^{\alpha} u(t-k)}{\alpha + 1},$$

where $\alpha = 0, 1, 2, \dots, 11$ determines the amount of smoothing (figure 2). The case $\alpha = 0$ corresponds to no smoothing; the case $\alpha = 11$ corresponds to a 12-month trailing average. The Sahm and Michez rules set $\alpha = 2$, which corresponds to a 3-month trailing average.

The vacancy rate is smoothed in the same way:

$$(2) \quad \bar{v}(t) = \frac{\sum_{k=0}^{\alpha} v(t-k)}{\alpha + 1},$$

where $\alpha = 0, 1, 2, \dots, 11$ (figure 2).

As an alternative, I also use an exponentially weighted moving average, which is defined recursively by

$$(3) \quad \bar{u}(t) = \alpha u(t) + (1 - \alpha) \bar{u}(t-1),$$

where $\alpha = 0.1, 0.2, 0.3, \dots, 1$ determines the amount of smoothing (figure 2). The case $\alpha = 1$ corresponds to no smoothing, and lower values of α generate more smoothing.

The vacancy rate is smoothed in the same way:

$$(4) \quad \bar{v}(t) = \alpha v(t) + (1 - \alpha)\bar{v}(t - 1),$$

where $\alpha = 0.1, 0.2, 0.3, \dots, 1$ (figure 2).

3.2. Detecting turning points

To detect turning points in the unemployment rate, I isolate the minimum of the unemployment rate at various horizons:

$$(5) \quad u^{\min}(t) = \min_{0 \leq k \leq \beta} \bar{u}(t - k),$$

where $\beta = 1, 2, 3, \dots, 18$ months is the horizon of the trailing minimum (figure 3A). The Sahm and Michez rules use $\beta = 12$, which corresponds to a 12-month trailing minimum.

Then, the increase in unemployment rate from the turning point is computed as

$$(6) \quad \tilde{u}(t) = \bar{u}(t) - u^{\min}(t).$$

Thus the current unemployment rate is compared to values ranging from last month's unemployment rate to the minimum unemployment rate over the past 18 months (figure 3B).

I proceed analogously to determine turning points in the vacancy rate. I first take the maximum of the vacancy rate at various monthly horizons:

$$(7) \quad v^{\max}(t) = \max_{0 \leq k \leq \beta} \bar{v}(t - k),$$

where $\beta = 1, 2, 3, \dots, 18$ months (figure 4A). Then, the decrease in vacancy rate from the turning point is computed as

$$(8) \quad \tilde{v}(t) = v^{\max}(t) - \bar{v}(t)$$

The current vacancy rate is compared to values ranging from last month's vacancy rate to the maximum vacancy rate over the past 18 months (figure 4B).

3.3. Scaling variations

The unemployment and vacancy variations $\tilde{u}(t)$ and $\tilde{v}(t)$ measure level changes in the unemployment and vacancy rates. It is not entirely clear, however, whether what matters are

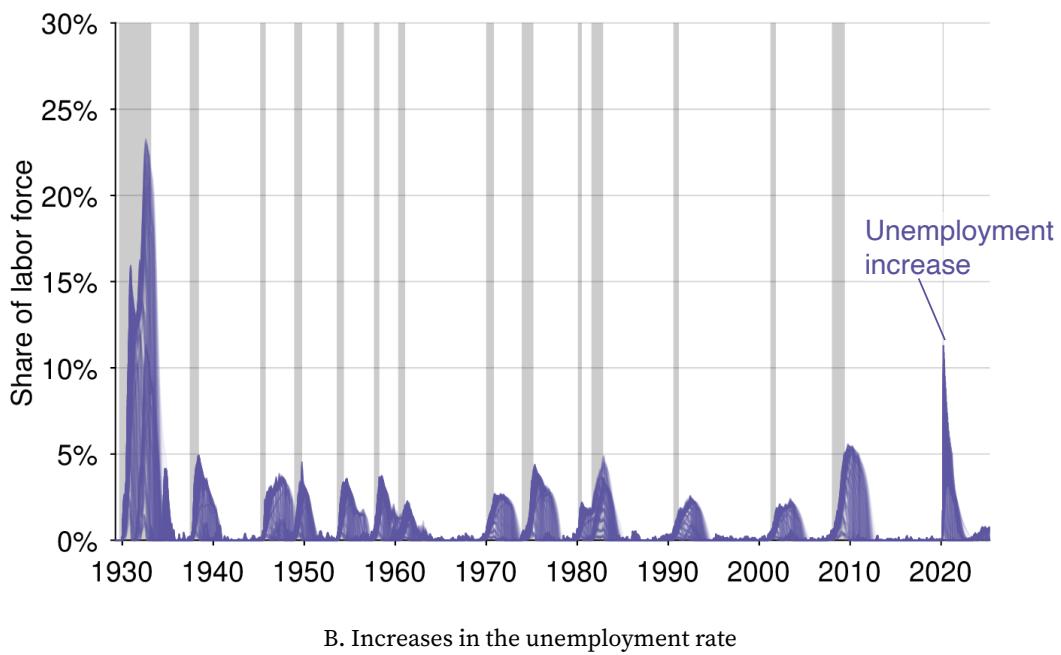
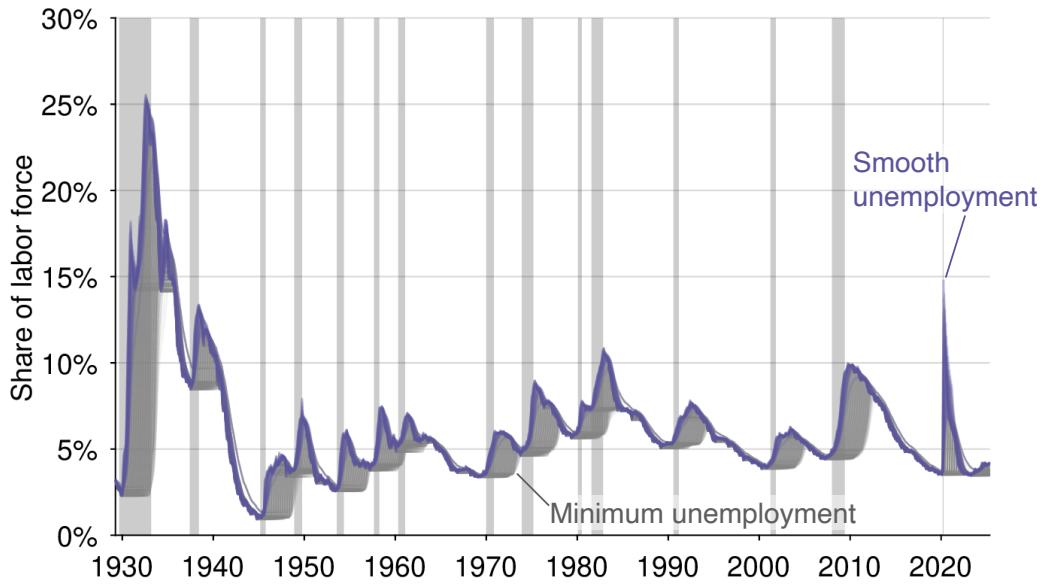
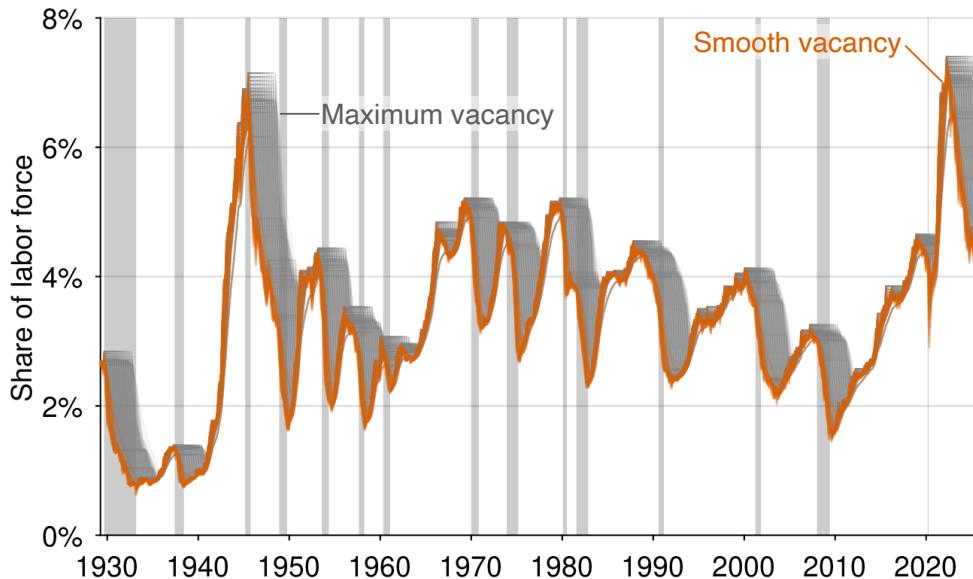
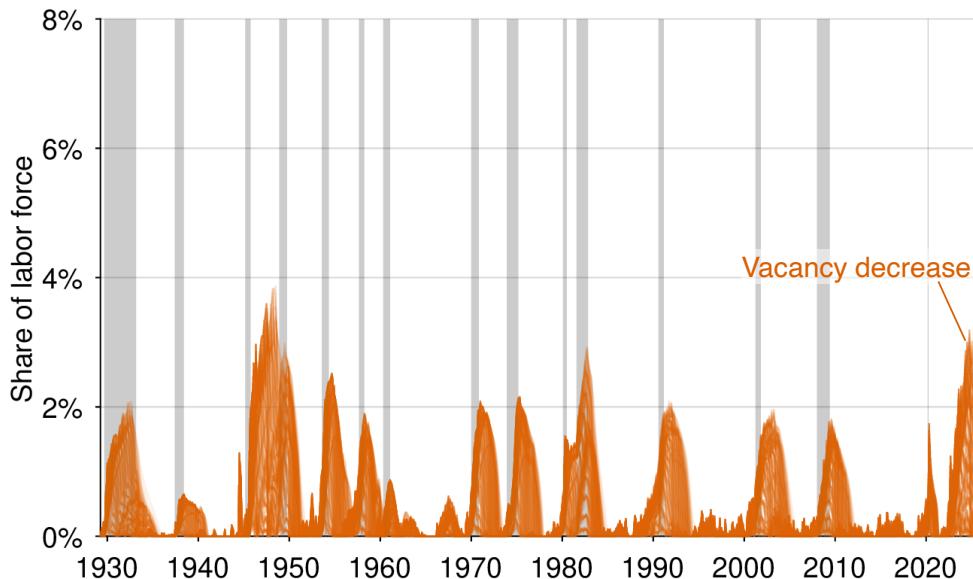


FIGURE 3. Increases in US unemployment rate, April 1929–September 2025

The smoothed unemployment rate comes from figure 2. The minimum unemployment rate is computed from equation (5). The unemployment increase is the difference between the smoothed unemployment rate and its minimum, as showed by equation (6). Shaded areas indicate recessions dated by the NBER (2023).



A. Trailing maximum of the vacancy rate



B. Decreases in the vacancy rate

FIGURE 4. Decreases in US vacancy rate, April 1929–September 2025

The smoothed vacancy rate comes from figure 2. The maximum vacancy rate is computed from equation (7). The vacancy decrease is the difference between the smoothed vacancy rate and its maximum, as showed by equation (8). Shaded areas indicate recessions dated by the NBER (2023).

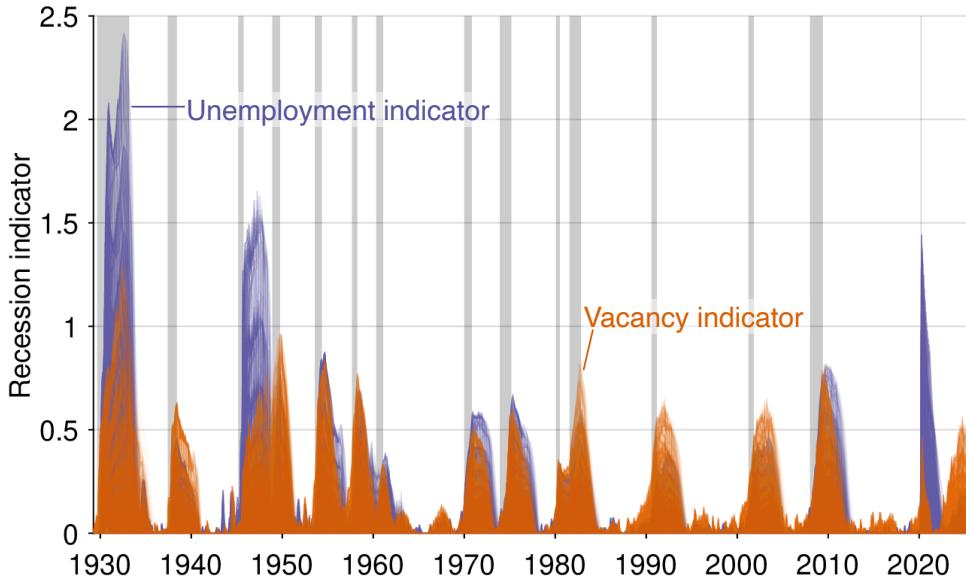


FIGURE 5. 4,356 unemployment indicators and 4,356 vacancy indicators

The unemployment indicators are computed by scaling the unemployment increases from figure 3B according to equation (9). The vacancy indicators are computed by scaling the vacancy decreases from figure 4B according to equation (10). Shaded areas indicate recessions dated by the NBER (2023).

level changes or percentage changes. Maybe recessions occur when the unemployment rate increases by 1pp, but it is just as possible that recessions occur when the unemployment rate increases by 10%. Since I am looking for the best recession indicator, I do not want to restrict the type of indicators that I consider.

So in addition to level changes in unemployment and vacancy rates, I also consider percentage changes and intermediate changes. To do that, I consider Box-Cox transformations of the unemployment indicator. Formally, I consider all the indicators of the form

$$(9) \quad \hat{u}(t) = \frac{[\bar{u}(t)]^\gamma - [u^{\min}(t)]^\gamma}{\gamma} \approx \frac{\tilde{u}(t)}{u^{\min}(t)^{1-\gamma}},$$

where the Box-Cox parameter takes values $\gamma = 0, 0.1, 0.2, \dots, 1$ (figure 5). The case $\gamma = 0$ reduces to percentage changes:

$$\hat{u}(t) = \log\left(\frac{\bar{u}(t)}{u^{\min}(t)}\right) \approx \frac{\tilde{u}(t)}{u^{\min}(t)}.$$

The case $\gamma = 1$ reduces to level changes:

$$\hat{u}(t) = \tilde{u}(t).$$

The Sahm and Michez rules both consider level changes, so $\gamma = 1$.

Similarly, I consider Box-Cox transformations of the vacancy indicator. That is, I construct all the vacancy indicators of the form

$$(10) \quad \hat{v}(t) = \frac{[v^{\max}(t)]^\gamma - [\bar{v}(t)]^\gamma}{\gamma} \approx \frac{\tilde{v}(t)}{v^{\max}(t)^{1-\gamma}},$$

where $\gamma = 0, 0.1, 0.2, \dots, 1$ (figure 5).

3.4. Combining single-variable indicators

The last step is to combine the unemployment and vacancy indicators constructed previously.

Besides using all the individual unemployment and vacancy indicators, I also construct new indicators that are linear combinations of the unemployment and vacancy indicators:

$$(11) \quad i(t) = \delta \hat{u}(t) + (1 - \delta) \hat{v}(t),$$

where $\delta = 0, 0.1, 0.2, \dots, 1$ is the weight on the unemployment indicator. The Sahm rule only uses the unemployment indicator, which corresponds to $\delta = 1$.

Moreover, Michaillat and Saez (2025) showed that the minimum of the unemployment and vacancy indicators performs very well to detect recessions because it is a less noisy signal of recessions than either the unemployment indicator or the vacancy indicator alone. Motivated by this insight, I also consider the minimum and maximum of the unemployment and vacancy indicators, as well as their linear combinations:

$$(12) \quad i(t) = \delta \min(\hat{u}(t), \hat{v}(t)) + (1 - \delta) \max(\hat{u}(t), \hat{v}(t)),$$

where $\delta = 0, 0.1, 0.2, \dots, 1$ is the weight on the minimum indicator. The Michez rule only uses the minimum indicator, which corresponds to $\delta = 1$.

3.5. Summary

Overall, from the unemployment rate, I construct $[12 + 10] \times 18 \times 11 = 4,356$ recession indicators. Similarly, from the vacancy rate, I construct another 4,356 recession indicators. By combining these indicators linearly, I produce a total of $4,356 \times 11 = 47,916$ indicators. By combining the minimum and maximum of these indicators, I produce another batch

of 47,916 indicators. So in total, I have $47,916 + 47,916 = 95,832$ recession indicators, from which I will build recession classifiers.

4. Constructing and evaluating recession classifiers

Having constructed the 95,832 recession indicators, I construct hundreds of millions of recession classifiers by applying thresholds to the indicators.

4.1. Detection methodology

A specific classifier is a specific indicator with a specific threshold. The classifier delineates periods as recessions by identifying points where the indicator crosses the threshold from below, contingent upon the economy being in expansion prior to the crossing. The classifier maintains a state variable for expansion and recession that is updated dynamically as new recessions are detected or the economy returns to expansion—defined by the indicator value returning to 0.

Formally, a classifier k is made up of an indicator $i(t) \geq 0$ and threshold $\zeta > 0$. The classifier keeps track of the state of the economy: $r(k, t) = 0$ if the economy is in expansion at time t and $r(k, t) = 1$ if the economy is in recession at time t . Initially, the economy is not in recession ($r(k, 0) = 0$). In period t , if the economy was previously in recession ($r(k, t-1) = 1$), it remains in recession if the indicator is positive ($r(k, t) = 1$ if $i(t) > 0$), but it enters an expansion if the indicator falls to 0 ($r(k, t) = 0$ if $i(t) = 0$). If the economy was previously in expansion ($r(k, t-1) = 0$), it remains in expansion if the indicator is below the threshold ($r(k, t) = 0$ if $i(t) < \zeta$), but it enters a recession if the indicator crosses the threshold ($r(k, t) = 1$ if $i(t) \geq \zeta$). If the classifier moves from expansion to recession in period t for the j th time, the detection date of recession j is set to t : $d(k, j) = t$.

A simpler way to proceed would have been to delineate recessions simply as periods when the indicator is above the threshold. This is for example the approach taken by Sahm (2019) and Michaillat and Saez (2025). The simple approach is not entirely desirable because it creates noise when the economy is exiting recessions. As the recession ends, the unemployment rate increases less rapidly and the vacancy rate declines less rapidly, so indicators start falling toward the threshold. During their fall, indicators sometimes drop below the threshold, then temporarily climb above the threshold, before falling down below it again. (Such blips appear on the indicator in figure 6 in the aftermath of the 1990 and 2001 recessions.) This secondary period above the threshold is not a new recession, however, and it should not be counted as such.

To eliminate such false positives, created by changes in the pace of economic slowdown around the threshold, I require that the economy is in expansion before entering a new recession, and I require that the indicator falls to 0 for the economy to be in expansion.

This classification procedure adds a bit of computational complexity, but it provides a more logical and more robust classification.

4.2. Selecting perfect classifiers

I construct recession classifiers by combining all the recession indicators with all thresholds $\zeta = 0.0001, 0.0002, 0.0003, \dots, 0.25$. In total, I therefore evaluate $95,832 \times 2,500 = 239,580,000$ recession classifiers. The threshold ζ is in the units of the recession indicator. When $\gamma = 1$ the indicator measures a level change in percentage points, so the threshold is in percentage points, but otherwise it is a transformed unit.

For each classifier, I compute the number of recessions detected between April 1929 and December 2021. The evaluation starts in April 1929 because this is when the data become available. It stops in December 2021 because it is too early to say if and when a recession occurred after that. Between April 1929 and December 2021, the NBER identifies 15 recessions. I therefore only select classifiers that detect 15 recessions during that time. In other words, I only keep classifiers that are perfect in a statistical sense, and I discard any classifier that makes any error—either false positive or false negative. A false negative is an actual recession missed by the classifier; it occurs when the threshold is too high (figure 6A). A false positive is a non-recession mistakenly detected by the classifier; it occurs when the recession threshold is too low (figure 6B).

Overall, the procedure produces 4,481,622 perfect classifiers, which detect all the recessions that occurred between 1929 and 2021 without false positives.

4.3. Evaluating perfect classifiers

Given that there are so many perfect classifiers, I need to rank them in some way to isolate the best one. Standard evaluation methods, like ROC curves, rank classifiers based on the amount of false positives and false negatives that they generate.² They are unsuitable here because the perfect classifiers generate no errors.

Instead, I evaluate each classifier k based on how quickly and accurately it detects recessions. For each recession j , I compute the detection error, which is the difference between the date when the recession officially started, $s(j)$, and the date when the classifier first detected the recession, $d(k, j)$:

$$(13) \quad \epsilon(k, j) = d(k, j) - s(j).$$

This is doable because each classifier detects the correct number of recessions, so there are as many start dates as detection dates. If $\epsilon(k, j) > 0$, classifier k detects recession j with

²See for instance Murphy (2022, section 5.1.3) and Piger (2020, section 18.2.3).

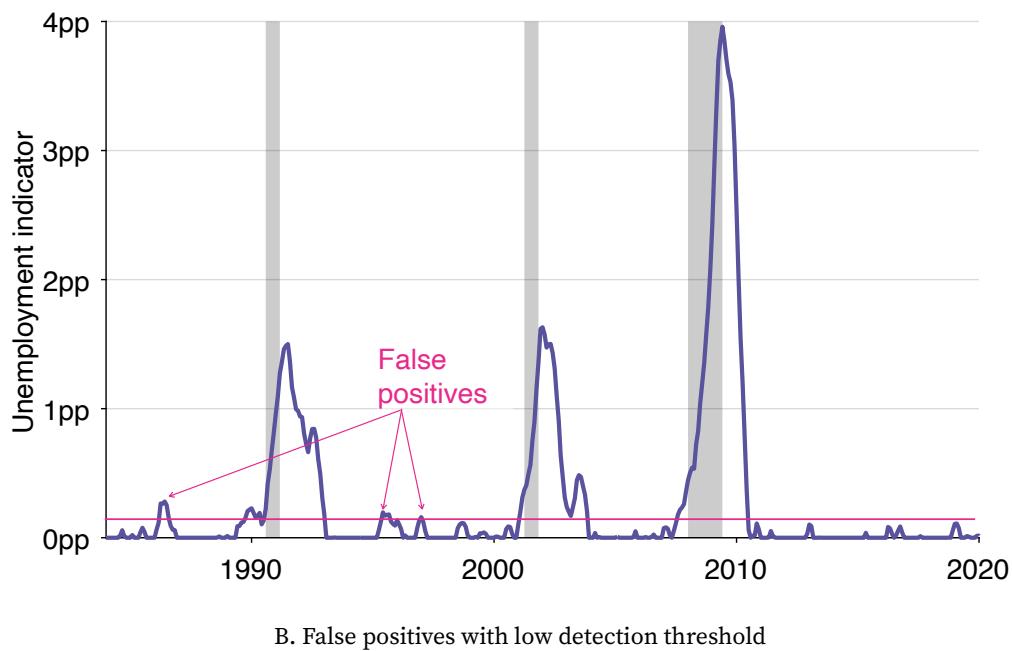
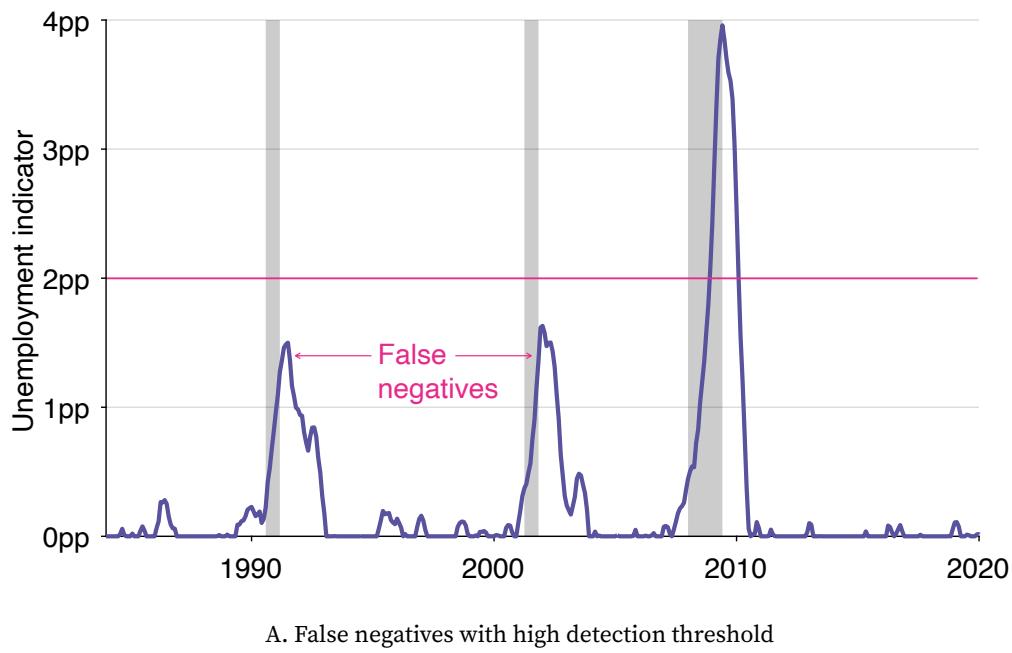


FIGURE 6. Possible classification errors

The figure displays one specific unemployment classifier and two possible recession thresholds as an illustration. Shaded areas indicate recessions dated by the NBER (2023).

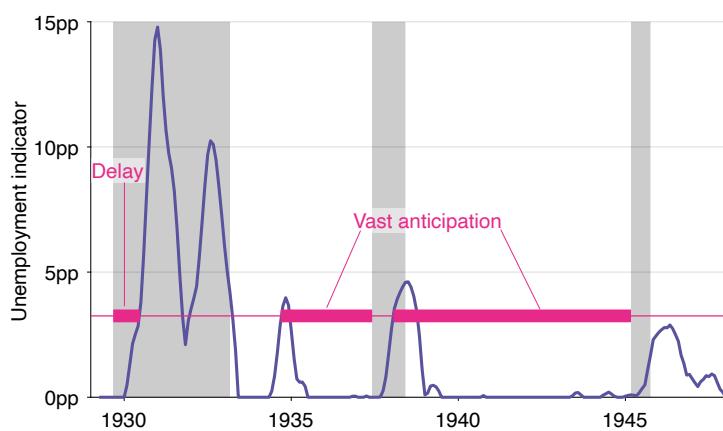
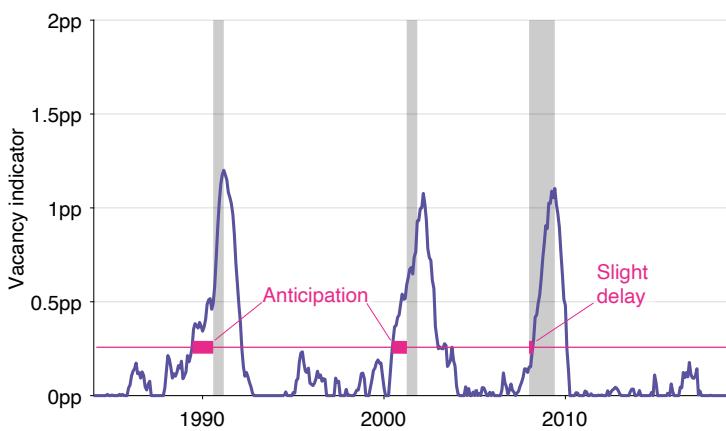
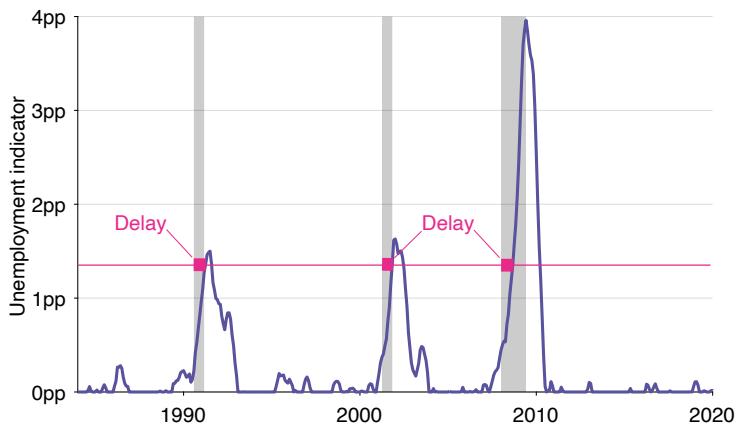


FIGURE 7. Evaluating perfect recession classifiers

The figure displays specific recessions classifiers and possible recession thresholds as an illustration. Shaded areas indicate recessions dated by the NBER (2023).

some delay. If instead $\epsilon(k, j) < 0$, classifier k detects recession j with some anticipation. Different classifiers have different detection errors for each recession (figure 7).

For each classifier k , I then compute two performance measures. First, I compute the mean of the detection errors on the training sample:

$$(14) \quad \mu(k) = \frac{1}{J} \cdot \sum_{j=1}^J \epsilon(k, j),$$

where J is the number of recessions in the training sample. For instance, $J = 15$ in the April 1929–December 2021 sample. Second, I compute the standard deviation of the detection errors:

$$(15) \quad \sigma(k) = \sqrt{\frac{1}{J} \cdot \sum_{j=1}^J [\epsilon(k, j) - \mu(k)]^2}.$$

A classifier with higher anticipation (lower $\mu(k)$) is better, as it allows policymakers, firms, workers, investors, and consumers to foresee the recession and prepare for it. For instance, the classifier in figure 7B is superior to the classifier in figure 7A.

A classifier with higher precision (lower $\sigma(k)$) is also better, as it provides more accurate information about the upcoming recession. It is clearly better to have a classifier that always detects a recession on its start date, than to have a classifier that is six months early half of the time and six months late half of the time—although both classifiers offer the same zero mean error. For instance, the classifier in figure 7C is not appealing, although it anticipates the second and third recessions, because it anticipates these recessions too much. The high and unusual anticipations reflect the fact that the classifier picks up a misleading increase in unemployment in 1935 and then misses an informative increase in unemployment in 1946.

4.4. Finding the anticipation-precision frontier

Having computed the mean and standard deviation of the detection errors for the 4,481,622 perfect classifiers, I restrict my attention to the classifiers on the anticipation-precision frontier (figure 8A). The frontier comprises the classifiers with lowest mean error (highest anticipation) and lowest standard deviation of errors (highest precision). For a given anticipation or delay, no classifier is more precise than the classifier on the frontier. And for a given precision, no classifier anticipates recession as much as the classifier on the frontier. This frontier helps identify classifiers that optimize early detection and accuracy.

The entire frontier includes only 248 of the 4,481,622 perfect classifiers. The left-most classifier on the frontier has a mean detection error of 3.1 months (slight delay) and a standard deviation of errors of 1.6 months (very precise). The right-most classifier on the

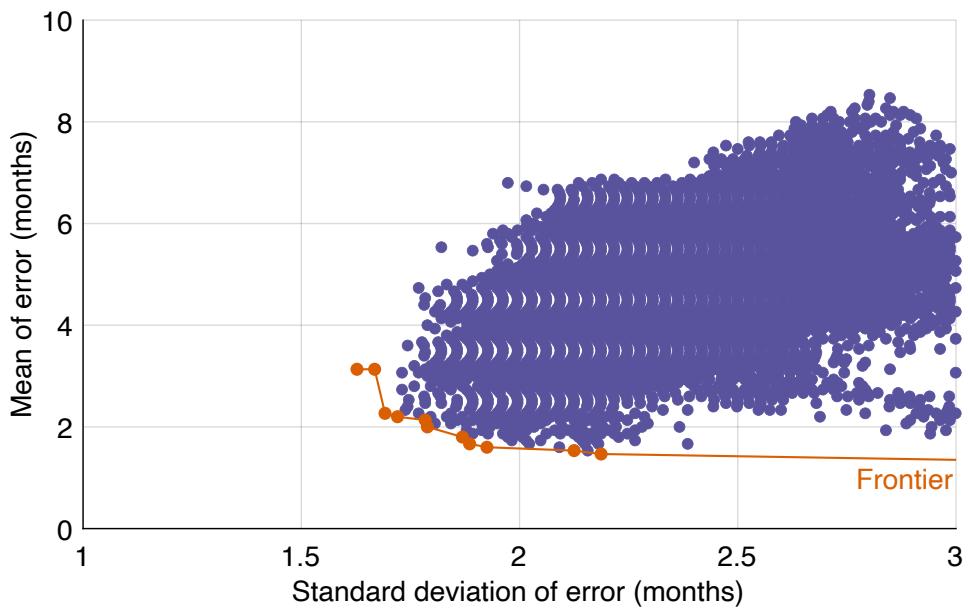
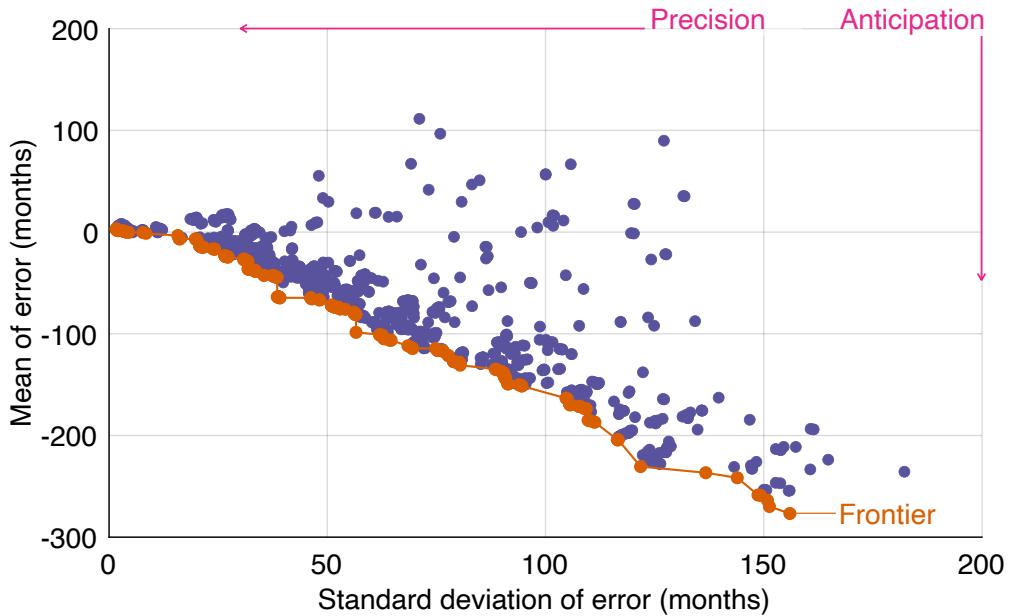


FIGURE 8. Anticipation and precision of 4,481,622 perfect recession classifiers for the United States, April 1929–December 2021

The figure displays the mean and standard deviation of the detection errors for 4,481,622 perfect classifiers, which detect the 15 recessions that occurred between 1929 and 2021 without false positives. The perfect classifiers are constructed from the single-variable indicators in figure 5, which are mixed using (11) and (12), and combined with recession thresholds between 0.0001 and 0.25. The mean error is computed from (14) and the standard deviation of errors is computed from (15). The anticipation-precision frontier is the set of perfect classifiers with lowest mean error (highest anticipation) and lowest standard deviation of errors (highest precision). The high-precision segment of the anticipation-precision frontier is the set of frontier classifiers with standard deviation of errors below 3 months.

frontier has a mean detection error of -277 months (huge anticipation) and a standard deviation of errors of 156 months (huge imprecision).

4.5. Selecting the frontier classifiers with highest precision

Which one of the 248 frontier classifiers should be picked to detect future recessions? A policymaker with mean-variance preferences over detection error would pick the classifier k that minimizes $\mu(k) + \lambda\sigma(k)$, where the parameter $\lambda > 0$ captures how much the policymaker values precision relative to anticipation. For example, a policymaker who requires a lot of time to implement stabilization policies might have a low λ , unlike a policymaker who is more nimble, who might have a higher λ . The policymaker would find the desirable classifier by finding the point on the frontier that is tangent to a line with slope $-\lambda$.

Without knowing the preferences of policymakers, it is impossible to pick the desirable classifier. Instead, I pick all the frontier classifiers with a standard deviation of errors $\sigma(k)$ below 3 months (figure 8B). For these classifiers, the 95% confidence interval for recession start dates has a width below 1 year. Indeed, assuming a normally distributed detection error, the 95% confidence interval for recession start dates has a width of 4 standard deviations, or $4 \times \sigma(k) < 12$ months. I pick this admittedly arbitrary threshold because detecting a recession more than 6 months before its start does not seem helpful, and detecting it more than 6 months late is not helpful either, as the NBER officially announces recessions on average 7 months after their starts. So imposing a 12-month confidence interval seems like a good choice to maximize the usefulness of the detection algorithm.

There might be added benefits from using an ensemble of classifiers, as shown by other ensemble methods such as bagging and boosting (Murphy 2022, section 18). By pooling information across classifiers, the ensemble might detect recessions more reliably than any single classifier. The ensemble might also be more stable: random data fluctuations or measurement errors that might mislead one classifier are less likely to mislead most of them.

By focusing on the high-precision segment of the anticipation-precision frontier, I select an ensemble of 11 classifiers (table 1). These classifiers detect the 15 recessions that occurred between April 1929 and December 2021 without producing false positives. The mean detection delay of the classifiers ranges from 1.5 months to 3.1 months, with an average value of 2.1 months. The standard deviation of the detection errors ranges from 1.6 months to 2.2 months, with an average value of 1.8 months.³

On average the classifiers detect recessions after their official start dates because the official start dates are backdated (NBER 2021). The NBER identifies recessions with

³For completeness, the 11 recession indicators and thresholds behind the 11 classifiers in table 1 are plotted in appendix B.

TABLE 1. Classifier ensemble selected from the anticipation-precision frontier

Smoothing method	Smoothing parameter α	Turning horizon β	Box-Cox parameter γ	Combination method	Combination weight δ	Threshold ζ	Detection error (months)			
							1929-2021		1979-2021	
							Mean	Standard deviation	Mean	Standard deviation
(1)	simple	4m	4m	1	min-max	1	0.23	3.1	1.6	2.5
(2)	simple	3m	8m	0.7	min-max	1	0.84	3.1	1.7	2.5
(3)	exponential	0.5	5m	0.9	min-max	1	0.38	2.3	1.7	1.2
(4)	exponential	0.5	5m	1	min-max	1	0.27	2.2	1.7	1.5
(5)	exponential	0.5	5m	1	min-max	1	0.25	2.1	1.8	1.3
(6)	exponential	0.7	10m	0.6	min-max	1	1.42	2.0	1.8	1.4
(7)	exponential	0.5	8m	0.7	min-max	1	0.70	1.8	1.9	1.5
(8)	exponential	0.4	8m	0.9	min-max	1	0.27	1.7	1.9	0.5
(9)	exponential	0.4	9m	0.9	min-max	1	0.27	1.6	1.9	0.3
(10)	exponential	0.3	8m	1	min-max	1	0.14	1.5	2.1	0.3
(11)	exponential	0.4	9m	1	min-max	1	0.19	1.5	2.2	0.3
Average:							2.1	1.8	1.2	1.3

The classifier ensemble is composed of the classifiers on the anticipation-precision frontier with a standard deviation of errors below 3 months (figure 8B). The simple smoothing method with parameter α is given by (1) and (2); the exponential smoothing method with parameter α is given by (3) and (4). The turning horizon β enters the construction of the classifiers through (5) and (7). The Box-Cox parameter γ enters the construction of the classifiers through (9) and (10). The u-v combination method with weight δ is given by (11); the min-max combination method with weight δ is given by (12). The detection error is given by (13); its mean is given by (14); its standard deviation is given by (15). The detection errors are computed between April 1929 and December 2021, and between January 1979 and December 2021.

hindsight, not in real time, which allows them to place the start dates slightly earlier than the classifiers' detection dates.

However, the classifiers detect recessions significantly earlier than the NBER. The NBER Dating Committee has been operating since 1979. On average, since its inception, the committee has taken 6.3 months to announce recession starts (table A1). By comparison, over the same period, the 11-classifier ensemble only takes 1.2 months to detect recession starts. The classifier ensemble is also quite precise over that period: the standard deviation of the detection errors averages only 1.3 months. So the classifier ensemble outperforms the NBER Dating Committee by 5.1 months.

The classifier ensemble clearly performs better over the modern, 1979–2021 period than over the entire, 1929–2021 sample—no doubt reflecting the higher quality of the data in more recent decades. Classifiers (9)–(11) in particular perform exceedingly well over the modern period. They detect the 6 recessions in the period with an average delay of only 0.3 month, so they are almost able to detect recessions in real time, just as they start. The standard deviation of detection errors is only 1.2 months.

4.6. Distance of the Michez rule from the frontier

The Michez rule also detects the 15 recessions that occurred between 1929 and 2021 without producing false positives, but it is slightly off the frontier. The Michez rule's mean detection error over 1929–2021 is 1.9 months, while the standard deviation of its detection error is 2.3 months (Michaillat and Saez 2025, tables 1 and 2). So it is a little less accurate and a little slower than classifiers (7)–(11), whose detection errors have means below 1.8 months and standard deviations below 2.2 months.

The distance between the Michez-rule classifier and classifiers (7)–(11) respectively are: $\sqrt{0.1^2 + 0.4^2} = 0.41$ months, $\sqrt{0.2^2 + 0.4^2} = 0.45$ months, $\sqrt{0.3^2 + 0.4^2} = 0.5$ months, $\sqrt{0.4^2 + 0.2^2} = 0.45$ months, and $\sqrt{0.4^2 + 0.1^2} = 0.41$ months. So overall, the Michez rule is roughly 0.4 months away from the anticipation-precision frontier. Given the simplicity of the Michez rule, it is quite striking that it is not further away from the frontier. Such proximity confirms that the Michez rule is a good option for users looking for a simple yet performant recession detection rule.

In fact, the key insight from the analysis by Michaillat and Saez (2025) remains valid here: taking the minimum of unemployment and vacancy indicators provides earlier and more accurate recession signals than relying on unemployment or vacancy indicators alone. Indeed, all the 11 classifiers in the ensemble selected to detect recession are obtained by taking the minimum of unemployment and vacancy indicators (table 1). (The combination method is "min-max" and the combination weight is 1, which means that these are minimum indicators, as (12) shows.) Hence, minimum indicators tend to be more accurate and provide earlier recession signals than those obtained from the

unemployment or vacancy indicators alone, and in fact than those obtained from other combinations of the unemployment and vacancy indicators. Michaillat and Saez (2025) made this discovery by filtering labor-market data in a specific way; but the result remains valid across a wide range of filtering.

The general insight is that combining data on unemployment and job vacancies—two noisy but independent measures of the state of the economy—provides a clearer signal of the latent state than looking at unemployment or job vacancies in isolation. The reason is that most recessions are caused by drops in aggregate demand, which produce negative comovements between unemployment rate and vacancy rate as the economy moves along the Beveridge curve (Michaillat and Saez 2015, 2022, 2024a). Therefore, a typical recession features both a decrease in vacancy rate and increase in unemployment rate. Such joint movements are picked up by minimum indicators.

5. Detecting recessions

The final step to building the recession detection algorithm is to aggregate the detection signals produced by the classifier ensemble. Once the signals are aggregated, I use them to compute the current recession risk.

5.1. Overview of the empirical construction of recession probabilities

For each classifier k and each recession j in the training period, the detection date $d(k, j)$ can be compared to the NBER start date $s(j)$, which yields the empirical detection error $\epsilon(k, j) = d(k, j) - s(j)$. These historical errors summarize how classifier k has behaved in past episodes: their mean $\mu(k)$ measures systematic anticipation or delay, and their standard deviation $\sigma(k)$ measures timing precision. No model structure is imposed beyond these summaries; they are computed directly from the training data.

When classifier k issues a detection at a new date $d(k)$ in real time, the unknown quantity is the start date s of the current recession. The key step is to map this new detection into a probability that the recession has begun by a candidate date t . To do so, I treat the historical detection errors $\epsilon(k, j)$ as an empirical forecast-error sample for classifier k and approximate their distribution by a normal distribution with the same mean $\mu(k)$ and variance $\sigma^2(k)$. This approximation is the only distributional assumption in this step, and it uses information obtained solely from the training sample. Given this approximation, the event $\{s \leq t\}$ is equivalent to $\{\epsilon(k) \geq d(k) - t\}$, where $\epsilon(k) = d(k) - s$ is the detection error associated with the new episode. The probability that the recession has begun by t , conditional on a detection at date $d(k)$, then follows directly from the fitted error distribution.

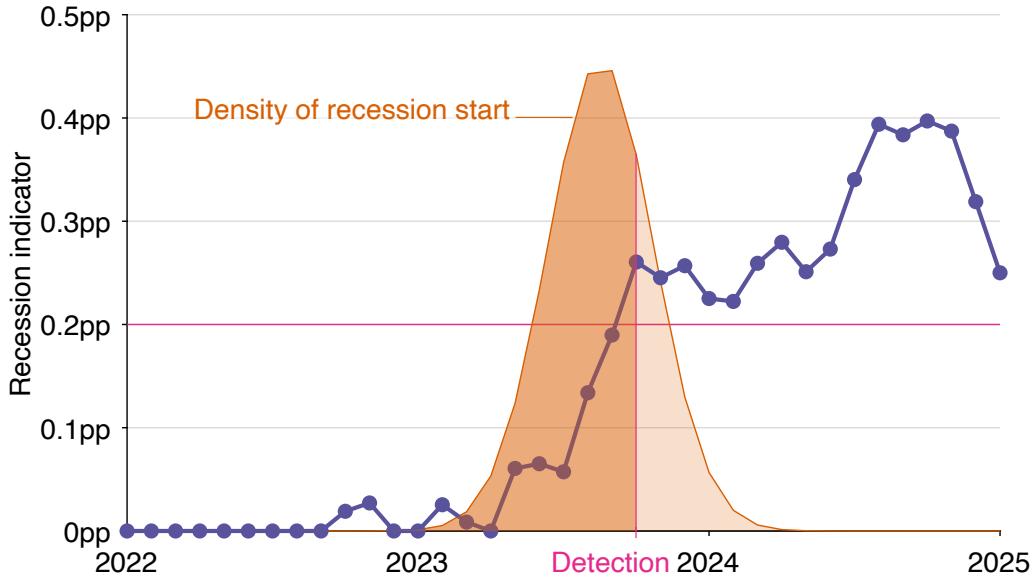


FIGURE 9. Computing the recession probability from an individual classifier

The figure displays as an illustration the probability density function of a recession start obtained from one specific recession classifier.

5.2. Recession probability from an individual classifier

Whenever an individual recession indicator crosses its threshold, I infer the probability that the recession has already started based on the distribution of the detection error.

If for instance the classifier is on average exactly on time, the recession has started with 50% probability when the classifier is activated—assuming a symmetric detection error. If the classifier is early on average, the probability is less than 50%. If the classifier is late on average, the probability is more than 50%, and so on. In the months following detection, the probability converges to 1 along the detection error's cumulative distribution function.

More formally, a perfect classifier k detects the J recessions in the training period. Each detection j generates a detection date $d(k, j)$ and detection error $\epsilon(k, j)$. For convenience, I assume that classifier k 's detection error $\epsilon(k)$ is normally distributed with mean $\mu(k)$ and standard deviation $\sigma(k)$. Then, it is easy to compute the probability that a new recession's start date, s , truly occurred before date t , given that the classifier detected a recession at date $d(k) \leq t$:

$$P(k, t) = \mathbb{P}(s < t | d(k), \mu(k), \sigma(k)) = \mathbb{P}(d(k) - s > d(k) - t | \mu(k), \sigma(k)).$$

Introducing the detection error $\epsilon(k) = d(k) - s$, I rewrite the probability as

$$P(k, t) = \mathbb{P}(\epsilon(k) > d(k) - t \mid \mu(k), \sigma(k)) = \mathbb{P}\left(\frac{\epsilon(k) - \mu(k)}{\sigma(k)} > \frac{d(k) - t - \mu(k)}{\sigma(k)}\right).$$

Finally, introducing the cumulative distribution function Φ of the standard normal distribution—the distribution followed by $[\epsilon(k) - \mu(k)]/\sigma(k)$ —I find that the probability that a recession has started at time t if classifier k detected a recession at $d(k)$ is:

$$(16) \quad P(k, t) = 1 - \Phi\left(\frac{d(k) - t - \mu(k)}{\sigma(k)}\right) = \Phi\left(\frac{t + \mu(k) - d(k)}{\sigma(k)}\right),$$

as long as classifier k remains in a recession state (figure 9).

When the classifier k signals that the economy is in expansion ($r(k, t) = 0$), I set the probability to $P(k, t) = 0$. The probability $P(k, t)$ is only positive when classifier k signals that the economy is in recession ($r(k, t) = 1$). The value $d(k)$ used in the probability is then the most recent recession detection date.

5.3. Recession probability from the classifier ensemble

The final step is to average the probability that the recession has started across the $K = 11$ classifiers used for detection. Each classifier in the ensemble generates a probability that the recession has started at time t , $P(k, t)$, given by (16). To compute the overall probability that the economy has entered a recession at time t , I average the probabilities produced by the individual classifiers in the ensemble:

$$(17) \quad P(t) = \frac{1}{K} \cdot \sum_{k=1}^K P(k, t).$$

Each classifier provides a recession-start probability. I then average these probabilities to yield a single recession-risk score. I could in principle aggregate the individual probabilities using generalized least-squares weights to correct for correlation across recession indicators. In practice, this approach is unreliable because the 11 classifiers are all constructed from transformations of only two data series (the unemployment rate and the vacancy rate). The covariance matrix of the underlying indicators is nearly singular, so the generalized least-squares weights are numerically unstable: small changes in the regularization of the covariance matrix generate large shifts in the weights. This instability reflects the collinearity of the indicators rather than useful information. For robustness and transparency, I therefore simply aggregate the individual probabilities using equal weights.

Figure 10A shows the aggregate recession probability given by the classifier ensemble

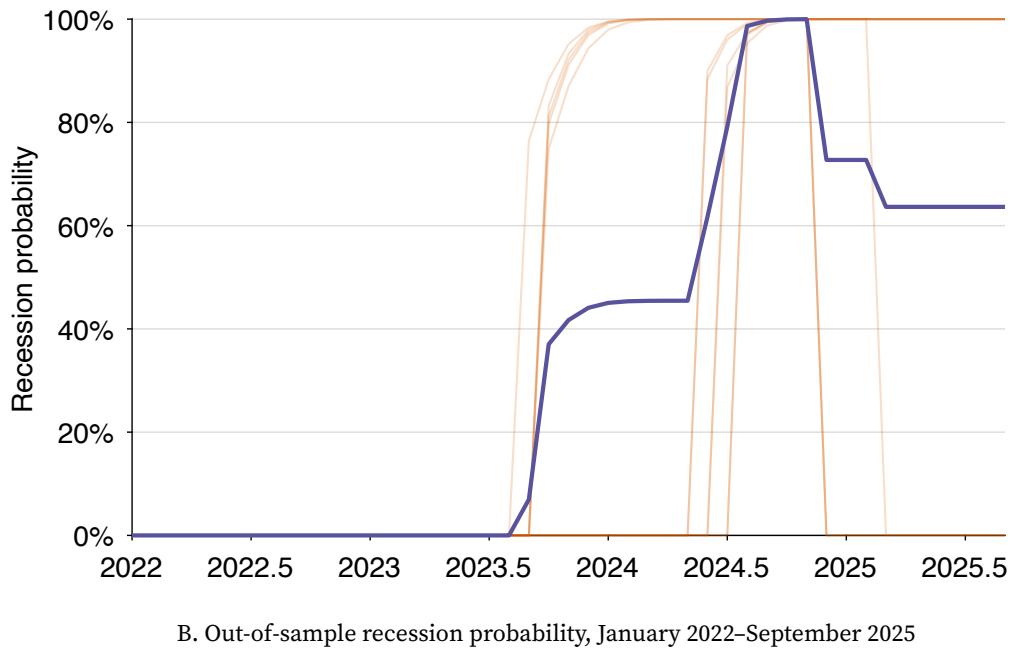
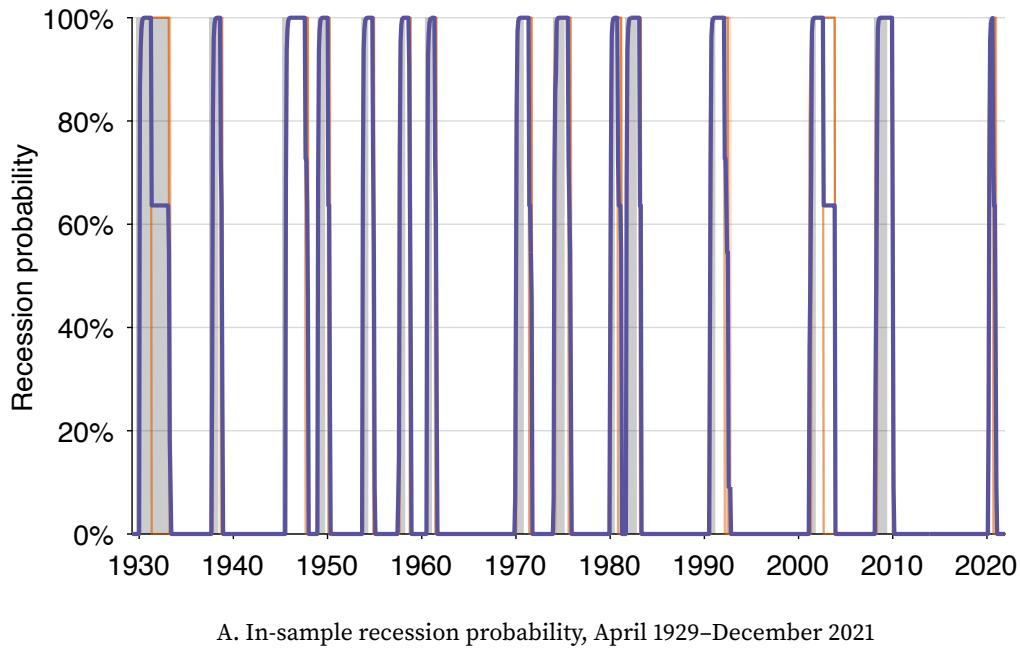


FIGURE 10. US recession probability given by the algorithm trained on April 1929–December 2021 data

The recession probability is computed from (17) (thick purple line). It is the average of the recession probabilities given by the 11 classifiers in the selected ensemble (table 1). Each individual probability is given by (16) (thin orange lines). The classifiers in the ensemble are selected from the high-precision segment of the 1929–2021 anticipation-precision frontier (figure 8B). Shaded areas indicate recessions dated by the NBER (2023).

on the training sample, 1929–2021. In each of the 15 in-sample recessions, the recession probability rapidly rises at the onset of the recession and quickly reaches 1. The probability then starts declining after the recession has ended.⁴

5.4. Application of the recession detection algorithm to current data

Finally, I apply the classifier ensemble obtained by training the algorithm on April 1929–December 2021 data to the current period: January 2022–September 2025. The aim is to assess the recession risk currently faced by the US economy (figure 10B).

A recession risk first emerged in September 2023, when the first classifier in the ensemble was activated. The recession probability given by the classifier ensemble reached 45% by January 2024. The recession probability increased significantly in May–August 2024, when all the remaining, unactivated classifiers of the ensemble were in turn activated. The probability that the US economy had entered a recession reached 99% in August 2024 and remained at that level until November 2024.

This negative assessment of the health of the US economy is in turn due to the sharp decrease in the vacancy rate and increase in the unemployment rate between the summer of 2022 and the summer of 2024 (figure 1). In the historical record, such cooling of the labor market coincides with a recession, which is why all the classifiers in the ensemble signalled a recession.

The recession probability then fell to 64% in March 2025. Indeed, in the winter of 2025, several classifiers in the ensemble became deactivated. This reduction in the recession probability reflects the fact that the labor market stopped cooling in the fall of 2024 (figure 1). As a result, some of the recession indicators returned to 0, and the corresponding classifiers signalled an expansion instead of a recession.

As of September 2025, 7 of the 11 classifiers in the ensemble trained on 1929–2021 data still indicate a recession. The aggregate recession probability given by the classifier ensemble is 64%. This probability suggests a high likelihood that a recession has begun at this point in time.

6. Addressing concerns about overfitting via placebo test

Given that the algorithm uses a vast set of potential classifiers to detect only 15 recessions, there is a risk of overfitting the data. To address this concern, I conduct a placebo test. The idea is to apply the same detection algorithm to a series of non-economic events that, by chance, occur with the same frequency as recessions in the training period. If the algorithm is not overfitting, it should be unable to detect these non-economic events.

⁴This algorithm does not aim to determine the ends of recessions. But it would be possible to build another algorithm on the same principles to determine recession ends as early and accurately as possible.

TABLE 2. First ladies in placebo test

First lady	In office	Husband	Date of death	
			Month	Year
Helen Taft	1909–1913	William Howard Taft	May	1943
Lou Hoover	1929–1933	Herbert Hoover	January	1944
Frances Cleveland	1886–1889, 1893–1897	Grover Cleveland	October	1947
Edith Roosevelt	1901–1909	Theodore Roosevelt	September	1948
Grace Coolidge	1923–1929	Calvin Coolidge	July	1957
Edith Wilson	1915–1921	Woodrow Wilson	December	1961
Eleanor Roosevelt	1933–1945	Franklin D. Roosevelt	November	1962
Mamie Eisenhower	1953–1961	Dwight D. Eisenhower	November	1979
Bess Truman	1945–1953	Harry S. Truman	October	1982
Pat Nixon	1969–1974	Richard Nixon	June	1993
Jacqueline Kennedy	1961–1963	John F. Kennedy	May	1994
Lady Bird Johnson	1963–1969	Lyndon B. Johnson	July	2007
Betty Ford	1974–1977	Gerald Ford	July	2011
Nancy Reagan	1981–1989	Ronald Reagan	March	2016
Barbara Bush	1989–1993	George H. W. Bush	April	2018

The table presents the 15 first ladies whose death occurred between April 1929 and December 2021. The dates of death are used as placebo to test whether the recession detection algorithm overfits the data or not. Source: White House Historical Association, available at <https://www.whitehousehistory.org/collections/first-lady-biographies>, and Wikipedia, available at https://en.wikipedia.org/wiki/List_of_first_ladies_of_the_United_States.

We will see that the algorithm indeed fails to detect the placebo events, which alleviates concerns about overfitting.

The challenge here is to find an event that is random, occurred 15 times between April 1929 and December 2021, and is not related in any way to the US economy. An event that fits these three criteria is the death of US first ladies. This event is random, and in fact more random than the death of US presidents because the ages of first ladies are much more variable than those of presidents. Obviously such deaths are not caused by the US economy, nor do they affect it in any way. By happenstance, there have been exactly 15 such deaths between April 1929 and December 2021 (table 2).

The algorithm is entirely unable to detect the deaths of first ladies. In the anticipation-precision plane, every perfect classifier has a standard deviation of detection errors above 500 months—more than 41 years (figure 11A). Such large standard deviation indicates that the detection dates are effectively random with respect to the death dates, confirming that the algorithm cannot detect the deaths of first ladies.

Another way to see that the algorithm is unable to detect first-lady deaths is to use the frontier classifiers to compute the probability of a first-lady death at any point in time (figure 11B). The overall probability is the average of the death probabilities given by the

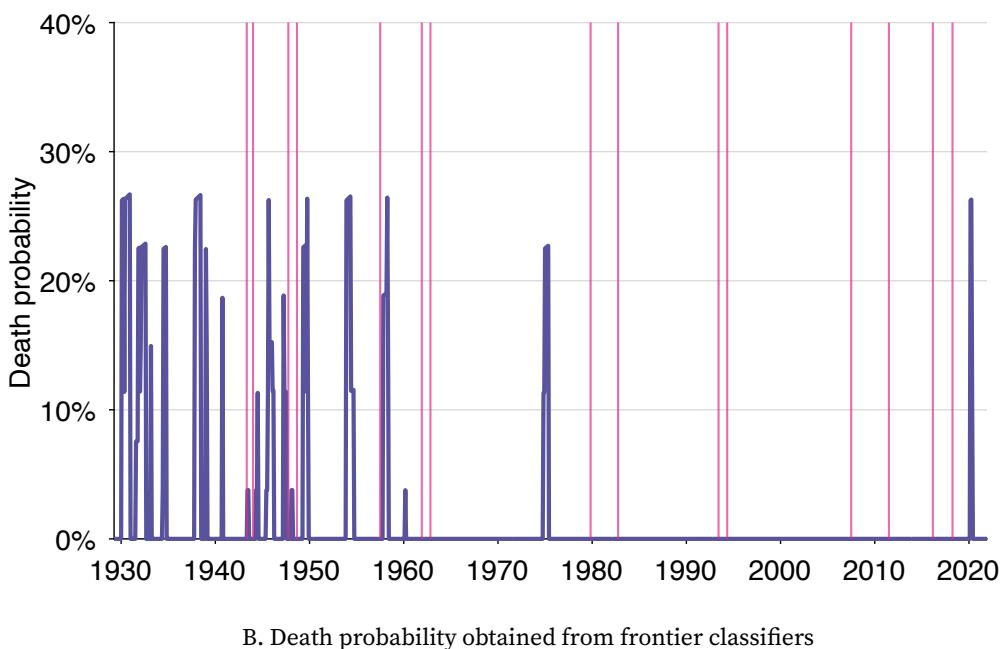
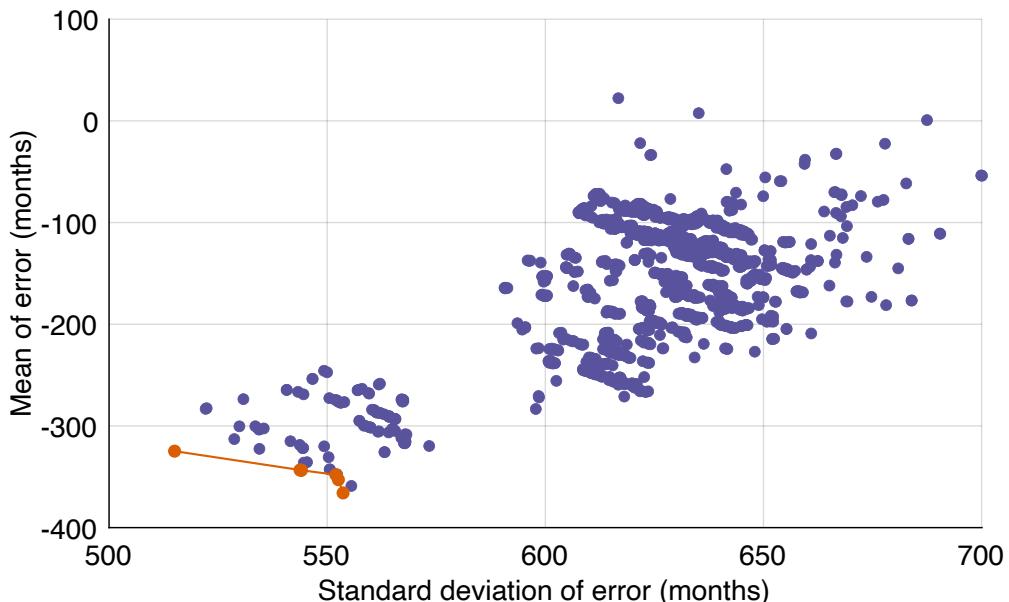


FIGURE 11. Placebo test of the algorithm: detecting the deaths of US first ladies

Panel A displays the mean and standard deviation of the detection errors for the perfect classifiers in the placebo test. These classifiers detect exactly 15 first-lady deaths between 1929 and 2021. The mean of the errors is computed from (14) and the standard deviation of errors is computed from (15), where the recession start dates are replaced by the first-lady death dates. The anticipation-precision frontier is composed of 7 classifiers. Panel B gives the death probability obtained from the 7 frontier classifiers. The probability is the average (17) of the death probabilities (16) given by the individual classifiers. Pink vertical lines are the death dates of the first ladies, as described in table 2.

individual classifiers, each given by (16). Vertical lines are the death dates of the first ladies, as described in table 2. The classifiers detect the placebo events at different times, resulting in many spikes of the event probability. Furthermore, many spikes occur in the 1930s, when no death occurred, and there is no spike between 1980 and 2020, when 7 deaths occurred. Another feature is that the death probabilities tend to spike between 10% and 30% and return to 0% rapidly. That is very different from the recession probabilities over the training period, which reached 100% and stayed there until the end of the recessions (figure 10A).

In sum, the huge standard deviation of detection errors and the randomly scattered death probability spikes indicate that the algorithm is unable to detect the placebo events. This inability implies that the algorithm does not overfit the data. The reason is that, although the algorithm considers hundreds of millions of classifiers, these classifiers are all constructed from only 2 independent time series: the unemployment rate and the vacancy rate. These 2 independent series must then detect 15 independent events. This structure inherently limits the potential for overfitting, as the number of independent data series is much smaller than the number of events to detect. While the various data transformations introduce additional degrees of freedom, they primarily refine the existing information rather than injecting new, independent information. Hence, the algorithm cannot spuriously detect a random series of events, which is why it fails the first-lady placebo test.

7. Assessing the reliability of the recession detection algorithm via backtests

To further assess the reliability of the algorithm, I run several backtests with out-of-sample detection. I shorten the training period and test the detection algorithm on subsequent out-of-sample recessions. Backtesting shows that the detection algorithm is robust and reliable. Classifier ensembles produced from backtests all give a non-negligible probability that the US economy has entered a recession in 2025.

7.1. Backtesting from 2015

First, I train the algorithm on data up to December 2014, which feature 14 recessions (figure A12). The training yields 11 classifiers on the high-precision segment of the anticipation-precision frontier (table A2). The insight from the Michez rule continues to hold in this backtest, as 10 of the 11 classifiers are based on minima of unemployment and vacancy indicators.

I then test the algorithm on January 2015–December 2021 data. The performance of the algorithm in the backtest is good. All 11 classifiers detect the out-of-sample, pandemic recession without false positives (figure 12). Furthermore, the detection is rapid: the

TABLE 3. Performance of the algorithm on backtests

Detection error (months)	2015 backtest		2005 backtest		1995 backtest		1985 backtest		1975 backtest		1965 backtest	
	11 classifiers		11 classifiers		10 classifiers		9 classifiers		8 classifiers		10 classifiers	
	Training	Testing	Training	Testing	Training	Testing	Training	Testing	Training	Testing	Training	Testing
Mean	2.3	1.1	2.1	2.5	2.3	1.8	2.4	1.6	2.8	1.6	3.6	-
Standard deviation	1.9	0.5	1.9	1.3	2.0	1.5	2.0	1.4	2.0	1.4	1.5	-
Minimum	-0.2	0.0	-0.5	1.2	-0.7	0.4	-0.7	0.3	-0.5	-0.1	1.7	-
Maximum	5.4	2.0	5.2	3.7	5.3	3.9	5.3	4.0	5.4	4.3	5.5	-
Recession count	14	1	13	2	12	3	11	4	9	6	7	8

The table averages the mean, standard deviation, minimum, and maximum of the detection errors across all the classifiers in the selected ensembles. In each backtest, the classifier ensemble comprises all the frontier classifiers that achieve a standard deviation of detection errors below 3 months on the training sample. The training samples considered are April 1929–December 2014, April 1929–December 1994, April 1929–December 1984, April 1929–December 1974, and April 1929–December 1964. The testing periods considered are January 2015–December 2021, January 2005–December 2021, January 1995–December 2021, January 1985–December 2021, and January 1965–December 2021. The detection error is given by (13); its mean is given by (14); its standard deviation is given by (15). The classifier ensembles are fully described in appendix C. Results cannot be computed on the testing sample for the 1965 backtest because one of the classifiers produces a false positive.

average detection delay across the 11 classifiers is just 1.1 months. The recession probability becomes positive in March 2020, when the recession officially started, and reaches 95% in May 2020.

7.2. Backtesting from 2005

In the next backtest, I train the algorithm on data up to December 2004, which feature 13 recessions (figure A13). The training yields 11 classifiers on the high-precision segment of the anticipation-precision frontier (table A3). Here too the insight from the Michez rule applies, as all 11 classifiers are based on minima of unemployment and vacancy indicators.

I test the algorithm on data for January 2005–December 2021, during which 2 recessions occurred: the Great Recession and pandemic recession. The performance of the algorithm in the backtest is again remarkably good. All 11 classifiers detect the 2 out-of-sample recessions without false positives (figure 13). The recession probability becomes positive in March 2008 and by June 2008 the recession probability reaches 95%—although the algorithm is only trained on data up to December 2004. This is quite good because in the summer of 2008, it was not obvious that the recession had already started. In August 2008, a notable macroeconometrician circulated a working paper through the NBER arguing that there was no chance that the economy was currently in recession—and mocking Warren Buffett and others who were arguing that the economy had indeed entered a recession (Leamer 2008).

7.3. Backtesting from 1995

I then repeat backtesting by training the algorithm on data up to December 1994 (figure A14). The training period presents 12 recessions. The training yields 10 classifiers on the high-precision segment of the anticipation-precision frontier (table A4). The insight from the Michez rule continues to hold: 9 of the 10 classifiers are based on minima of unemployment and vacancy indicators.

I test the algorithm on subsequent data, from January 1995 to December 2021. The testing period features 3 recessions: dot-com recession, Great Recession, and pandemic recession. The performance is again remarkably good. All 10 classifiers detect the 3 out-of-sample recessions without false positives (figure 14). The dot-com recession, which started in April 2001, is detected quite early: the recession probability becomes positive in March 2001 and by June 2001 the recession probability reaches 97%.

The algorithm continues to detect well the Great Recession and pandemic recession. For instance, once again, the recession probability becomes positive in March 2008 and by June 2008 the recession probability reaches 95%—so the algorithm does as well as in the 2005 backtest, although here it is only trained on data up to December 1994.

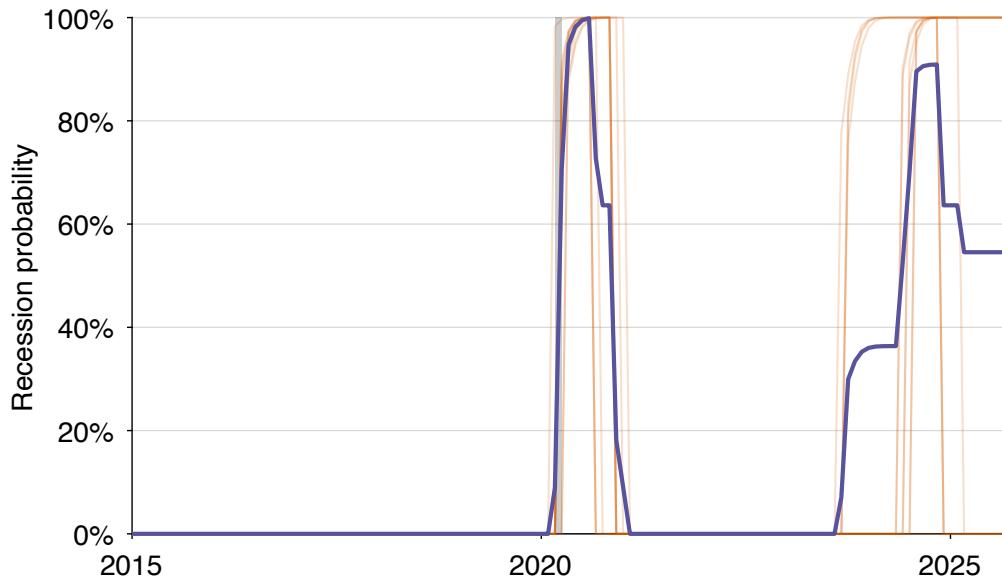


FIGURE 12. Out-of-sample US recession probability, January 2015–September 2025

The recession probability from the classifier ensemble is given by (17) (thick purple line). It is the average of the recession probabilities given by the individual classifiers in the ensemble, each given by (16) (thin orange lines). Shaded areas indicate recessions dated by the NBER (2023).

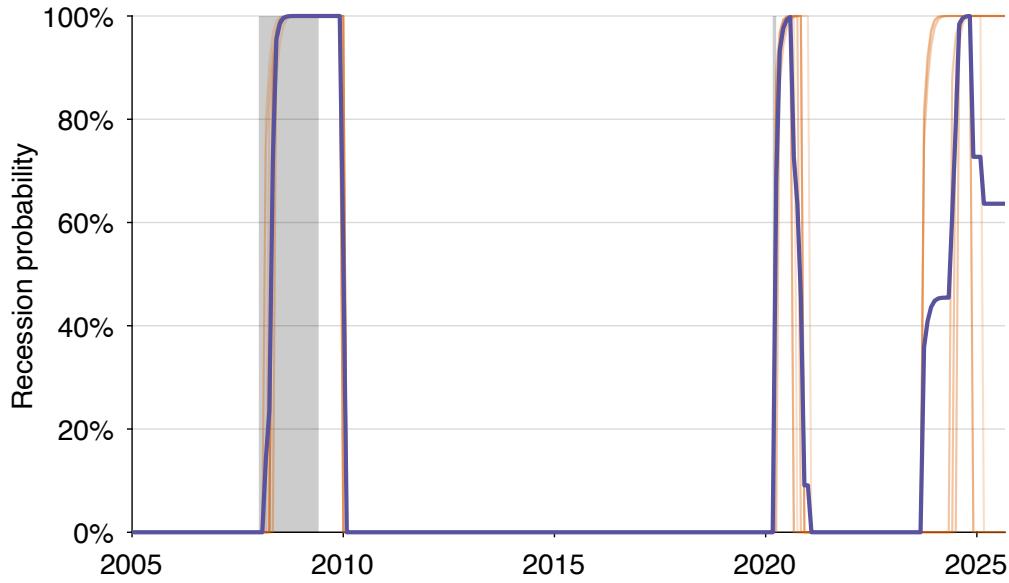


FIGURE 13. Out-of-sample US recession probability, January 2005–September 2025

The recession probability from the classifier ensemble is given by (17) (thick purple line). It is the average of the recession probabilities given by the individual classifiers in the ensemble, each given by (16) (thin orange lines). Shaded areas indicate recessions dated by the NBER (2023).

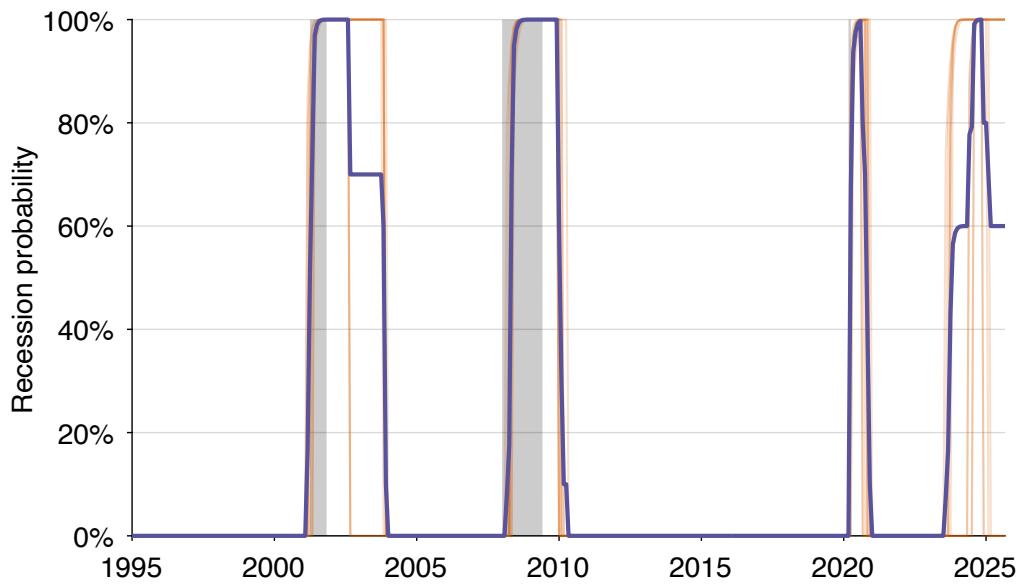


FIGURE 14. Out-of-sample US recession probability, January 1995–September 2025

The recession probability from the classifier ensemble is given by (17) (thick purple line). It is the average of the recession probabilities given by the individual classifiers in the ensemble, each given by (16) (thin orange lines). Shaded areas indicate recessions dated by the NBER (2023).

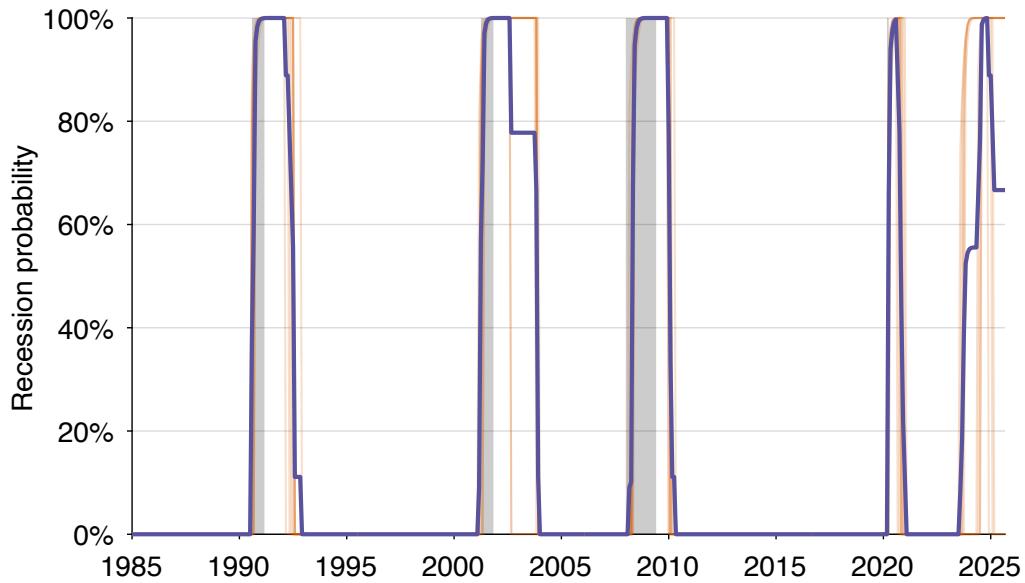


FIGURE 15. Out-of-sample US recession probability, January 1985–September 2025

The recession probability from the classifier ensemble is given by (17) (thick purple line). It is the average of the recession probabilities given by the individual classifiers in the ensemble, each given by (16) (thin orange lines). Shaded areas indicate recessions dated by the NBER (2023).

7.4. Backtesting from 1985

To continue exploring the robustness of the algorithm in backtests, I next train the algorithm on data up to December 1984 (figure A15). The training period presents 11 recessions. The training yields 9 classifiers on the high-precision segment of the anticipation-precision frontier; all but one are based on minima of unemployment and vacancy indicators (table A5).

I then test the algorithm on the subsequent period, from January 1985 to December 2021, which comprises 4 recessions. The performance of the algorithm is again remarkably good. All 9 classifiers detect the 4 out-of-sample recessions without false positives (figure 15). The recession that started in August 1990 is detected early: the recession probability becomes positive in August 1990 and it reaches 95% two months later, in October 1990.

The algorithm continues to detect well the subsequent recessions, including the Great Recession. Once again, the recession probability becomes positive in March 2008 and it reaches 95% by June 2008—so the algorithm does as well as in the 2005 backtest, although here it is only trained on data up to December 1984.

7.5. Backtesting from 1975

Next, I train the algorithm on data up to December 1974, which feature 9 recessions (figure A16). The training yields 8 classifiers on the high-precision segment of the anticipation-precision frontier (table A6). Again, all but one classifier are based on minima of unemployment and vacancy indicators.

I test the algorithm on the subsequent period, January 1975–December 2021, which features 6 recessions. All 8 classifiers detect the 6 out-of-sample recessions without any false positives (figure 16). The algorithm detects the twin Volcker recessions—which started in February 1980 and August 1981—without much delay: the recession probability turns positive in January 1980 and October 1981, and it reaches 95% in March 1980 and again in November 1981.

The algorithm also continues to detect the subsequent recessions quite early. Over the entire testing period, the mean detection error is 1.6 months, and the standard deviation of errors is 1.4 months, which are both less than the values obtained by training the algorithm over 1929–2021 (table 3).

7.6. Backtesting from 1965

Last, I train the algorithm on data up to December 1964—which features 7 recessions (figure A17). The training yields 10 classifiers on the high-precision segment of the anticipation-precision frontier (table A7). Here, none of the selected classifiers are of the minimum

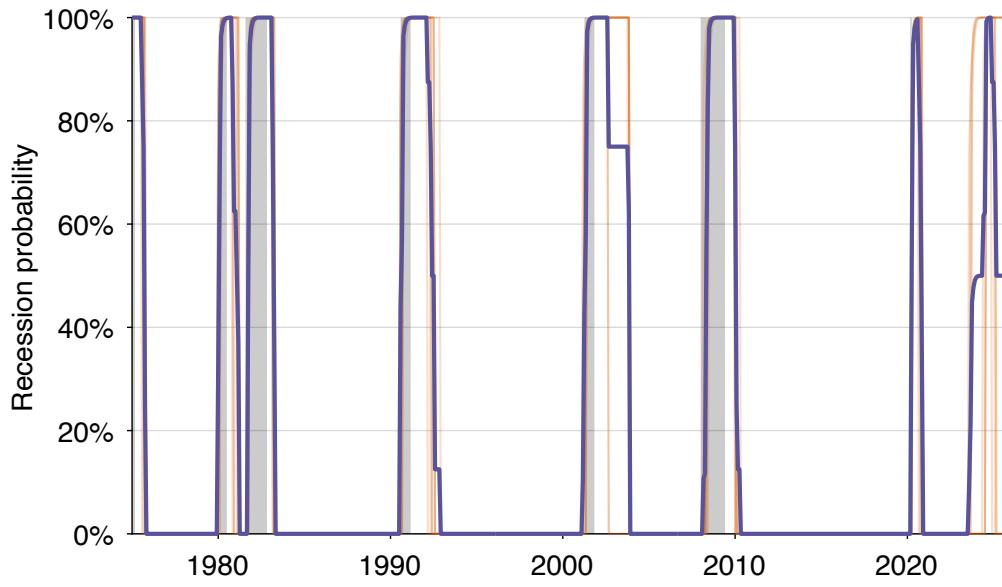


FIGURE 16. Out-of-sample US recession probability, January 1975–September 2025

The recession probability from the classifier ensemble is given by (17) (thick purple line). It is the average of the recession probabilities given by the individual classifiers in the ensemble, each given by (16) (thin orange lines). Shaded areas indicate recessions dated by the NBER (2023).

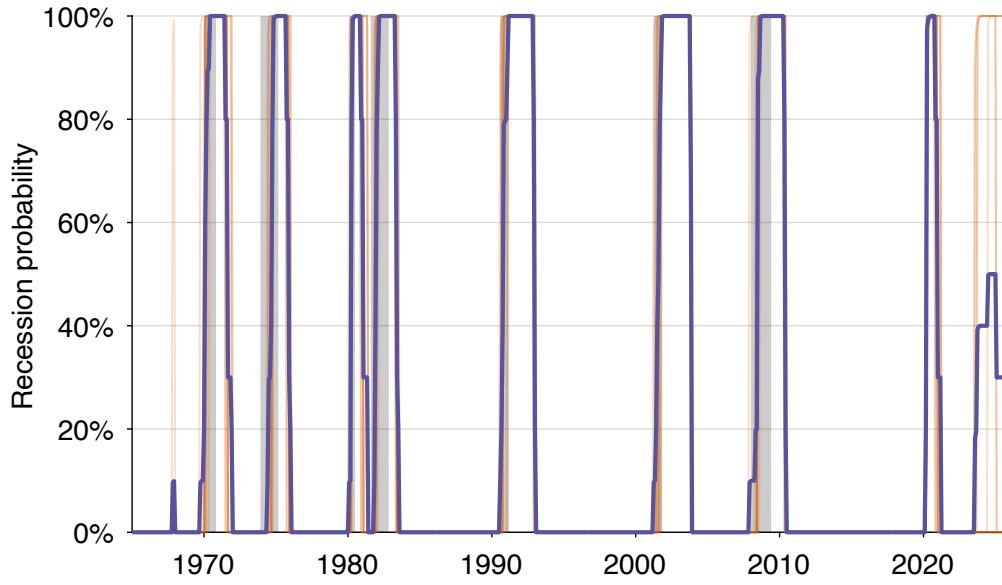


FIGURE 17. Out-of-sample US recession probability, January 1965–September 2025

The recession probability from the classifier ensemble is given by (17) (thick purple line). It is the average of the recession probabilities given by the individual classifiers in the ensemble, each given by (16) (thin orange lines). Shaded areas indicate recessions dated by the NBER (2023).

form; most of them are linear combinations of unemployment and vacancy indicators.

I test the algorithm on the 8 recessions of the subsequent period, January 1965–December 2021. Given that the algorithm is trained on fewer recessions than it must detect, the performance remains strikingly good. All 10 classifiers detect the 8 out-of-sample recessions (figure 17). However, for the first time in the backtests, 1 of the 10 classifiers produces a false positive. This error occurs in 1967: the false positive is the small blip in the recession probability in figure 17. Given that the trained algorithm produces its first error, I stop the backtests at this point.

7.7. Application of the backtested classifier ensembles to current data

Finally, I use the classifier ensembles produced by the 1975–2015 backtests to evaluate the current recession risk.

The classifier ensembles constructed in backtests give recession probabilities that are slightly different but all positive in 2025—confirming that the recession risk is elevated in 2025. The classifier ensemble created by training the algorithm on 1929–2014 data gives a 55% probability that the US economy has entered a recession in September 2025 (figure 12). The classifier ensemble created by training the algorithm on 1929–2004 data gives a 64% probability that the US economy has entered a recession in September 2025 (figure 13). The classifier ensemble created by training the algorithm on 1929–1994 data gives a 60% recession probability in September 2025 (figure 14). The classifier ensemble created by training the algorithm on 1929–1984 data produces the highest recession probability for September 2025: 67% (figure 15). Finally, the classifier ensembles obtained by training the algorithm on 1929–1974 data give a recession probability of 50% in September 2025 (figures 16).

8. Detecting recessions with product market data

The NBER’s Business Cycle Dating Committee does not rely on unemployment or vacancy data to date recessions.⁵ As the NBER (2025) explains, the committee relies mostly on product market data (payroll employment, production, sales, and consumption) to date the turning points of business cycles:

The determination of the months of peaks and troughs is based on a range of monthly measures of aggregate real economic activity published by the federal statistical agencies. These include real personal income less transfers, nonfarm payroll employment, employment as measured by the household survey, real personal consumption expenditures, wholesale-retail sales adjusted for price

⁵Strangely, the NBER (2025) prominently displays the US unemployment rate at the top of the webpage explaining that business-cycle dating does not rely on the unemployment rate.

changes, and industrial production.... In recent decades, the two measures we have put the most weight on are real personal income less transfers and nonfarm payroll employment.

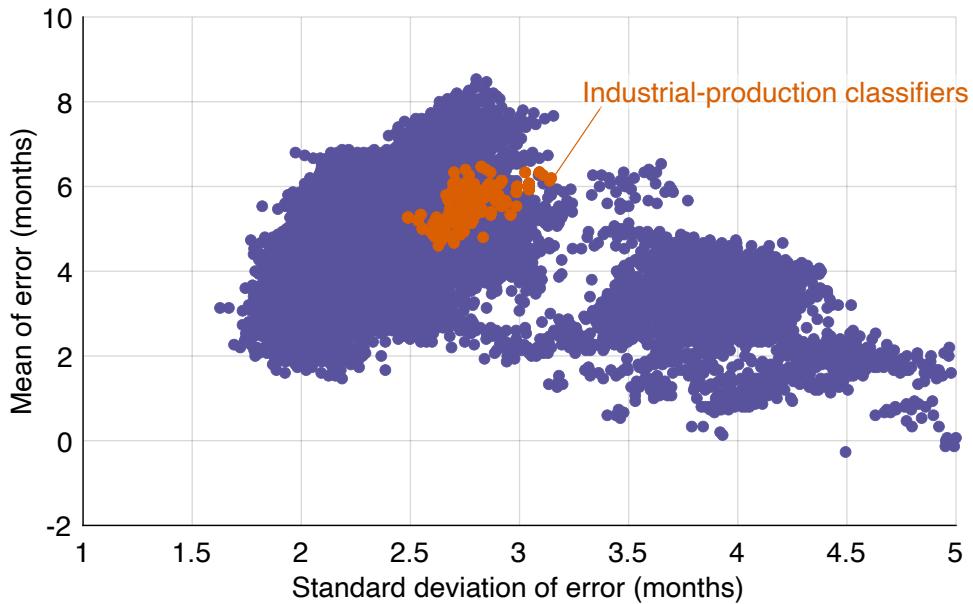
When the Dating Committee announced the start of the dot-com recession, it explained that the unemployment rate was a lagging and noisy variable, so it was not appropriate to detect recessions (NBER 2001). When it announced the starts of the Great Recession and pandemic recession, it did not mention unemployment at all (NBER 2008a, 2020). Among the frequently asked questions answered by the Dating Committee, one concerns unemployment: “How do the cyclical fluctuations in the unemployment rate relate to the NBER business-cycle chronology?” The NBER (2024) answers that because the unemployment rate is trendless while the variables considered are growing, the unemployment rate is not a reliable variable to date business cycles:

The unemployment rate is a trendless indicator that moves in the opposite direction from most other cyclical indicators.... The NBER business-cycle chronology considers economic activity, which grows along an upward trend. As a result, the unemployment rate sometimes rises before the peak of economic activity, when activity is still rising but below its normal trend rate of increase.... On the other hand, the unemployment rate often continues to rise after activity has reached its trough.

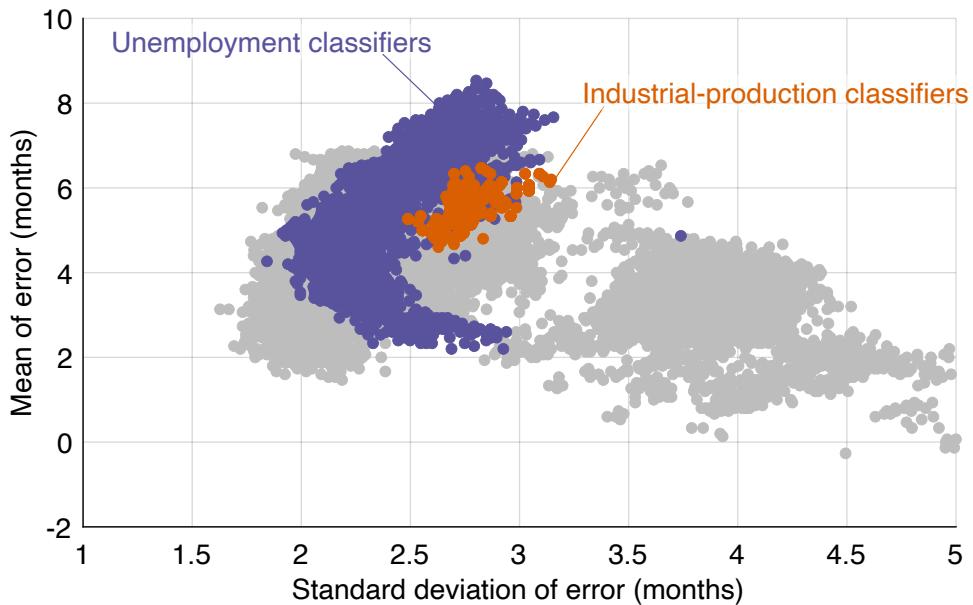
In light of all the purported limitations of the unemployment rate, one naturally wonders whether it would be possible to detect recessions earlier and more accurately by applying the algorithm developed in this paper to some of the product market data used by the Dating Committee. In this section, I apply the algorithm to the industrial production index constructed by the Federal Reserve Board (2025). Of all the variables examined by the Dating Committee, this is the only variable available since 1929. Another advantage of industrial production is that it has a long history as a marker of business cycles. In the early days of the NBER’s Dating Committee, industrial production featured prominently in recession announcements (NBER 1979, 1980, 1982). A third advantage is that of all the series examined by the Dating Committee, industrial production is probably the least connected to the unemployment and vacancy rates, which makes the exercise especially interesting.

I filter the industrial production index just like the unemployment and vacancy rates and produce 4,356 recession indicators. I then find thresholds such that the resulting classifiers perfectly detect the 15 recessions that occurred between April 1929 and December 2021. There are 404,590 such perfect classifiers. Finally, for each perfect classifier, I compute the mean and standard deviation of the detection errors.

To compare the performance of the industrial-production and labor-market classifiers, I display all the classifiers in the same anticipation-precision plane (figure 18). I



A. Performance of industrial-production classifiers versus unemployment-vacancy classifiers



B. Performance of industrial-production classifiers versus unemployment classifiers

FIGURE 18. Comparing labor-market classifiers to product-market classifiers

The figure reproduces figure 8A. In addition, it highlights the mean and standard deviation of the detection errors for classifiers built from industrial-production indicators (panels A and B) and for classifiers built from unemployment indicators (panel B).

first compare the industrial-production classifiers to all the classifiers constructed with unemployment and vacancy data (figure 18A). Clearly, the industrial-production classifiers are some distance away from the anticipation-precision frontier constructed from labor-market data. For instance, the most precise industrial-production classifier has a standard deviation of detection error of 2.5 months, while the most precise labor-market classifier has a standard deviation of detection error of 1.6 months (table 1). The same industrial-production classifier is not only less precise, it has less anticipation than the labor-market classifier: it detects recessions with an average delay of 5.3 months while the most precise labor-market classifier detects recessions with an average delay of 3.1 months (table 1). So the most precise industrial-production classifier is about $\sqrt{0.9^2 + 2.2^2} = 2.4$ months away from the labor-market anticipation-precision frontier. Thus, the unemployment-vacancy combination is better than industrial production to detect recession starts early and accurately.

The previous task was quite challenging for industrial-production classifiers because they rely on one data source but competed against classifiers constructed from two data sources (unemployment and vacancies). To make the task easier, I compare the industrial-production classifiers to classifiers constructed only from unemployment data. Even then, industrial-production classifiers are less performant than unemployment classifiers (figure 18B). For a given precision, there are unemployment classifiers with much lower detection delay than the industrial-production classifiers. And for a given delay, there are unemployment classifiers with much better precision than the industrial-production classifiers.

This result suggests that at the minimum the NBER Dating Committee might benefit from incorporating the unemployment rate into their considerations. Despite being labelled lagging and noisy, the unemployment rate performs significantly better than industrial production for recession detection. Of course, it would be even better to include both unemployment and vacancy rates into their considerations.

9. Conclusion

The paper develops a new algorithm for detecting US recessions from unemployment and vacancy data. The algorithm improves upon traditional approaches like the Sahm and Michez rules, which arbitrarily determine how labor market data are filtered (Sahm 2019; Michaillat and Saez 2025). It is possible to construct many other recession classifiers by filtering the data differently. Then, by optimizing the filtering process, it is possible to detect recessions earlier and more accurately.

The algorithm starts by constructing hundreds of millions of recession classifiers from unemployment and vacancy data. The algorithm then selects perfect classifiers, which avoid both false negatives (undetected true recessions) and false positives (falsely

detected recessions). By further selecting classifiers that lie on the anticipation-precision frontier, the algorithm optimizes jointly the earliness and accuracy of detection. Although they detect recessions very early, many classifiers on the frontier are too imprecise to be helpful. To ensure that the classifiers used to detect recessions are precise enough, the algorithm finally selects classifiers that are on the high-precision segment of the frontier.

The algorithm is trained on April 1929–December 2021 data—a period that experienced 15 recessions. On the training sample, the algorithm performs quite well. On average, the selected classifier ensemble signals recessions 2.1 months after their true onset, with a standard deviation of detection errors of 1.8 months. The classifier ensemble is much faster than the NBER Business Cycle Dating Committee: on average between 1979 and 2021, the committee takes 6.3 months to determine recession starts, while the classifier ensemble only takes 1.2 months.

Finally, I apply the trained algorithm to current data. The classifier ensemble obtained from 1929–2021 data gives a 64% probability that the United States has entered a recession in September 2025. The classifier ensembles obtained in various backtests also indicate an elevated recession risk. The classifier ensembles selected on the 1929–2014, 1929–2004, 1929–1994, 1929–1984, and 1929–1974 subsamples give a current recession probability of 55%, 64%, 60%, 67%, and 50%, respectively.

Overall, the algorithm developed in the paper shows that labor market conditions characteristic of a recession are not on the horizon—they are already here. What would be the implication if, in retrospect, no recession started in 2025? In that case, the algorithm would need to be retrained on the extended period that includes 2022–2025. Only classifiers that do not detect a recession in 2022–2025 would be selected by the algorithm. The classifiers that misdetected a recession in 2022–2025 would be eliminated. Many of the classifiers on the anticipation-precision frontier currently signal a recession, so they would be eliminated, and the anticipation-precision frontier would shift out after retraining. If a recession does not materialize, we would therefore learn that using labor market data to detect recessions is inherently more challenging than suggested by the 1929–2021 historical record.

References

- Abraham, Katharine G. 1987. “Help-Wanted Advertising, Job Vacancies, and Unemployment.” *Brookings Papers on Economic Activity* 18 (1): 207–248. <https://doi.org/10.2307/2534516>.
- Barnichon, Regis. 2010. “Building a Composite Help-Wanted Index.” *Economics Letters* 109 (3): 175–178. <https://doi.org/10.1016/j.econlet.2010.08.029>.
- BLS. 2020. “Labor Force Statistics from the Current Population Survey: Overview.” <https://perma.cc/RN3P-S4SL>.
- BLS. 2023. “Labor Force Statistics from the Current Population Survey: Concepts and Definitions.” <https://perma.cc/6LGU-AEU6>.

- BLS. 2024a. "Job Openings and Labor Turnover Survey Overview Page." <https://perma.cc/Y6EQ-WBXF>.
- BLS. 2024b. "JOLTS Frequently Asked Questions." <https://perma.cc/22QW-ASAL>.
- BLS. 2024c. "Release Calendar." <https://perma.cc/ZC78-SDPN>.
- BLS. 2025a. "Civilian Labor Force Level." FRED, Federal Reserve Bank of St. Louis. <https://fred.stlouisfed.org/series/CLF160V>.
- BLS. 2025b. "Job Openings: Total Nonfarm." FRED, Federal Reserve Bank of St. Louis. <https://fred.stlouisfed.org/series/JTSJOL>.
- BLS. 2025c. "Technical Documentation (CPS)." <https://perma.cc/DZ9K-AJP8>.
- BLS. 2025d. "Unemployment Level." FRED, Federal Reserve Bank of St. Louis. <https://fred.stlouisfed.org/series/UNEMPLOY>.
- Chen, Zhihong, Azhar Iqbal, and Huiwen Lai. 2011. "Forecasting the Probability of US Recessions: A Probit and Dynamic Factor Modelling Approach." *Canadian Journal of Economics* 44 (2): 651–672. <https://doi.org/10.1111/j.1540-5982.2011.01648.x>.
- Crump, Richard, Domenico Giannone, and David Lucca. 2020a. "Reading the Tea Leaves of the US Business Cycle – Part One." Liberty Street Economics, Federal Reserve Bank of New York. <https://perma.cc/W79A-EFPF>.
- Crump, Richard, Domenico Giannone, and David Lucca. 2020b. "Reading the Tea Leaves of the US Business Cycle – Part Two." Liberty Street Economics, Federal Reserve Bank of New York. <https://perma.cc/8CEL-TK37>.
- Federal Reserve Board. 2025. "Industrial Production: Total Index." FRED, Federal Reserve Bank of St. Louis. <https://fred.stlouisfed.org/series/INDPRO>.
- Fornaro, Paolo. 2016. "Forecasting US Recessions with a Large Set of Predictors." *Journal of Forecasting* 35 (6): 477–492. <https://doi.org/10.1002/for.2388>.
- Garimella, Rohit, Oscar Jorda, and Sanjay R. Singh. 2025. "Tracking Labor Market Stress." FRBSF Economic Letter 2025-19. <https://perma.cc/LXQ5-HSE4>.
- Hamilton, James D. 2011. "Calling Recessions in Real Time." *International Journal of Forecasting* 27 (4): 1006–1026. <https://doi.org/10.1016/j.ijforecast.2010.09.001>.
- Hatzius, Jan and Sven Jari Stehn. 2012. "Comment on 'The Ins and Outs of Forecasting Unemployment: Using Labor Force Flows to Forecast the Labor Market'." *Brookings Papers on Economic Activity* 43 (2): 118–131. <https://doi.org/10.1353/eca.2012.0024>.
- Leamer, Edward E. 2008. "What's a Recession, Anyway?" NBER Working Paper 14221. <https://doi.org/10.3386/w14221>.
- McCracken, Michael W. and Serena Ng. 2016. "FRED-MD: A Monthly Database for Macroeconomic Research." *Journal of Business & Economic Statistics* 34 (4): 574–589. <https://doi.org/10.1080/07350015.2015.1086655>.
- Michaillat, Pascal and Emmanuel Saez. 2015. "Aggregate Demand, Idle Time, and Unemployment." *Quarterly Journal of Economics* 130 (2): 507–569. <http://doi.org/10.1093/qje/qjv006>.
- Michaillat, Pascal and Emmanuel Saez. 2022. "An Economical Business-Cycle Model." *Oxford Economic Papers* 74 (2): 382–411. <https://doi.org/10.1093/oep/gpab021>.
- Michaillat, Pascal and Emmanuel Saez. 2024a. "Beveridgean Phillips Curve." arXiv:2401.12475. <https://doi.org/10.48550/arXiv.2401.12475>.
- Michaillat, Pascal and Emmanuel Saez. 2024b. " $u^* = \sqrt{uv}$: The Full-Employment Rate of Un-

- employment in the United States.” *Brookings Papers on Economic Activity* 55 (2). <https://doi.org/10.1353/eca.2024.a964373>.
- Michaillat, Pascal and Emmanuel Saez. 2025. “Has the Recession Started?” *Oxford Bulletin of Economics and Statistics* 87 (6): 1047–1058. <https://doi.org/10.1111/obes.12685>.
- Murphy, Kevin P. 2022. *Probabilistic Machine Learning: An Introduction*. Cambridge, MA: MIT Press. <http://probml.github.io/book1>.
- NBER. 1979. “Business Cycle Dating Committee Announcement December 31, 1979.” <https://perma.cc/7XKA-526Z>.
- NBER. 1980. “Business Cycle Dating Committee Announcement June 3, 1980.” <https://perma.cc/8MNK-GM2S>.
- NBER. 1982. “Business Cycle Dating Committee Announcement January 6, 1982.” <https://perma.cc/TL5X-9M8D>.
- NBER. 2001. “Business Cycle Dating Committee Announcement November 26, 2001.” <https://perma.cc/T6T8-46KQ>.
- NBER. 2008a. “Business Cycle Dating Committee Announcement December 1, 2008.” <https://perma.cc/488B-ZCGY>.
- NBER. 2008b. “Business Cycle Dating Committee Announcement January 7, 2008.” <https://perma.cc/U737-K42A>.
- NBER. 2020. “Business Cycle Dating Committee Announcement June 8, 2020.” <https://perma.cc/9HDG-A7QL>.
- NBER. 2021. “Business Cycle Dating Committee Announcements.” <https://perma.cc/JL65-XDWH>.
- NBER. 2023. “US Business Cycle Expansions and Contractions.” <https://perma.cc/B4LH-BVT8>.
- NBER. 2024. “Business Cycle Dating Procedure: Frequently Asked Questions.” <https://perma.cc/32QQ-D5NW>.
- NBER. 2025. “Business Cycle Dating.” <https://perma.cc/2DQB-AUAX>.
- Ng, Serena. 2014. “Viewpoint: Boosting Recessions.” *Canadian Journal of Economics* 47 (1): 1–34. <https://doi.org/10.1111/caje.12070>.
- O’Trakoun, John and Adam Scavette. 2025. “A Better Sahm Rule? Introducing the SOS Recession Indicator.” *Economics Letters* 247: 112175. <https://doi.org/10.1016/j.econlet.2025.112175>.
- Petrosky-Nadeau, Nicolas and Lu Zhang. 2021. “Unemployment Crises.” *Journal of Monetary Economics* 117: 335–353. <https://doi.org/10.1016/j.jmoneco.2020.01.009>.
- Philips, Thomas K. 2024. “A Simple Real-Time Algorithm to Identify Turning Points in US Business Cycles.” <https://ssrn.com/abstract=4897450>.
- Piger, Jeremy. 2020. “Turning Points and Classification.” In *Macroeconomic Forecasting in the Era of Big Data*, edited by Peter Fuleky, chap. 18. New York: Springer.
- Sahm, Claudia. 2019. “Direct Stimulus Payments to Individuals.” In *Recession Ready: Fiscal Policies to Stabilize the American Economy*, edited by Heather Boushey, Ryan Nunn, and Jay Shambaugh, chap. 3. Washington: Brookings Institution. <https://www.brookings.edu/articles/direct-stimulus-payments-to-individuals/>.
- Schannep, Jack. 2008. *Dow Theory for the 21st Century*. Hoboken, NJ: John Wiley & Sons.

- Shimer, Robert. 2005. "The Cyclical Behavior of Equilibrium Unemployment and Vacancies." *American Economic Review* 95 (1): 25–49. <https://doi.org/10.1257/0002828053828572>.
- Stock, James H. and Mark W. Watson. 2014. "Estimating Turning Points Using Large Data Sets." *Journal of Econometrics* 178: 368–381. <https://doi.org/10.1016/j.jeconom.2013.08.034>.
- Sun, Jiandong, Shuaizhang Feng, and Yingyao Hu. 2021. "Misclassification Errors in Labor Force Statuses and the Early Identification of Economic Recessions." *Journal of Asian Economics* 75: 101319. <https://doi.org/10.1016/j.asieco.2021.101319>.
- Weir, David R. 1992. "A Century of US Unemployment : 1890–1990. Revised Estimates and Evidence for Stabilization." *Research in Economic History* 14: 301–346.
- Zagorsky, Jay L. 1998. "Job Vacancies in the United States: 1923 to 1994." *Review of Economics and Statistics* 80 (2): 338–345. <https://doi.org/10.1162/003465398557438>.

Appendix A. Announcements by the NBER Business Cycle Dating Committee

Table A1 lists the announcement dates of recession starts and ends by the NBER Business Cycle Dating Committee since its inception. As of September 2025, the committee has announced 6 recession starts. On average, the announcements come 6.3 months after the starts; the standard deviation of announcement delays is 2.7 months. For completeness, the table also lists announcement dates for recession ends and related statistics.

Appendix B. Classifier ensemble selected from 1929–2021 data

This appendix displays the recession classifiers included in the classifier ensemble built in section 4 and used to detect recessions in section 5. The classifiers are selected using 1929–2021 data and are described in table 1. The classifiers are plotted in figures A1–A11. Each classifier is based on a recession indicator and a recession threshold. The classifier is activated when the indicator crosses the threshold; the classifier is deactivated once the indicator returns to 0.

Appendix C. Additional results for 1965–2015 backtests

This appendix describes the recession classifiers included in the classifier ensembles built in several backtests in section 7. The classifier ensemble used in the 2015 backtest is described in table A2; the in-sample performance of the ensemble is depicted in figure A12. The classifier ensemble used in the 2005 backtest is described in table A3; the in-sample performance of the ensemble is depicted in figure A13. The classifier ensemble used in the 1995 backtest is described in table A4; the in-sample performance of the ensemble is depicted in figure A14. The classifier ensemble used in the 1985 backtest is described in table A5; the in-sample performance of the ensemble is depicted in figure A15. The classifier ensemble used in the 1975 backtest is described in table A6; the in-sample performance of the ensemble is depicted in figure A16. The classifier ensemble used in the 1965 backtest is described in table A7; the in-sample performance of the ensemble is depicted in figure A17.

TABLE A1. NBER Business Cycle Dating Committee announcements

Start date	Recession start		Recession end	
	Announcement date	Announcement delay (months)	End date	Announcement date
February 1980	June 1980	4	July 1980	July 1981
August 1981	January 1982	5	November 1982	July 1983
August 1990	April 1991	8	March 1991	December 1992
April 2001	November 2001	7	November 2001	July 2003
January 2008	December 2008	11	June 2009	September 2010
March 2020	June 2020	3	April 2020	July 2021
Mean:	6.3			15.2
Standard deviation:	2.7			4.4

Source: NBER (2021, 2023).

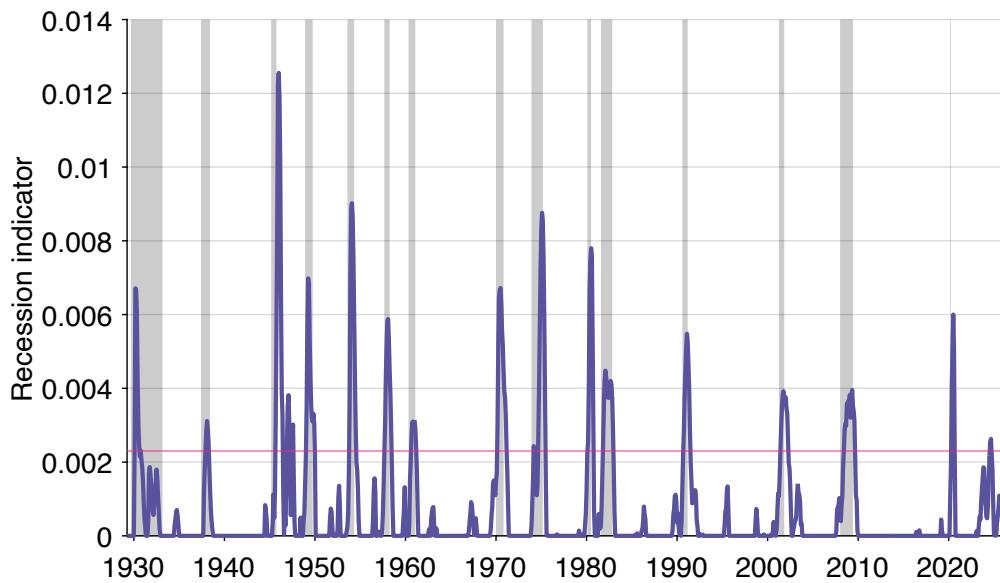


FIGURE A1. Classifier (1) in the classifier ensemble selected from 1929–2021 data

The classifier is described in table 1. The thick, purple line is the recession indicator on which the classifier is based. The thin, pink line is the recession threshold on which the classifier is based. Shaded areas indicate recessions dated by the NBER (2023).

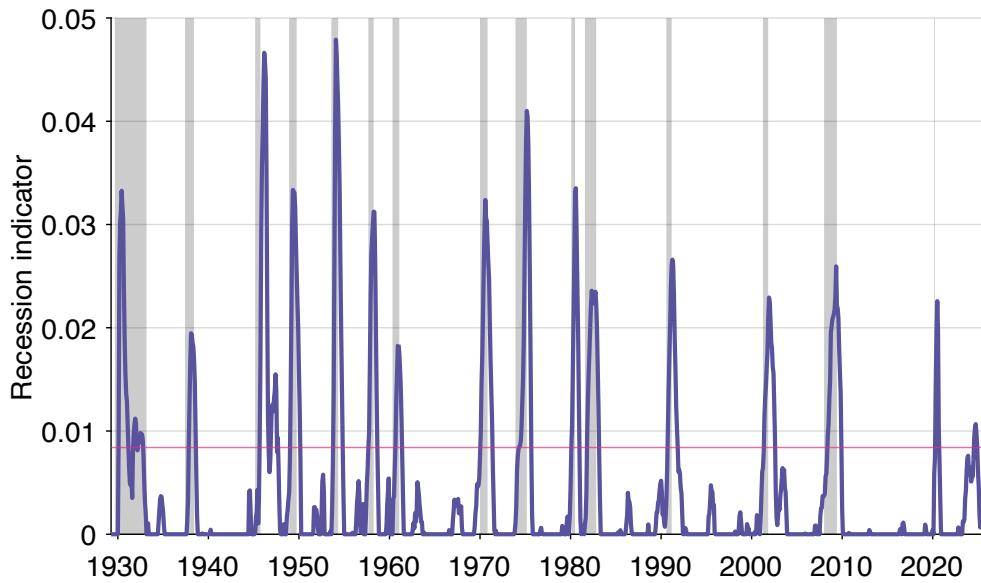


FIGURE A2. Classifier (2) in the classifier ensemble selected from 1929–2021 data

The classifier is described in table 1. The thick, purple line is the recession indicator on which the classifier is based. The thin, pink line is the recession threshold on which the classifier is based. Shaded areas indicate recessions dated by the NBER (2023).

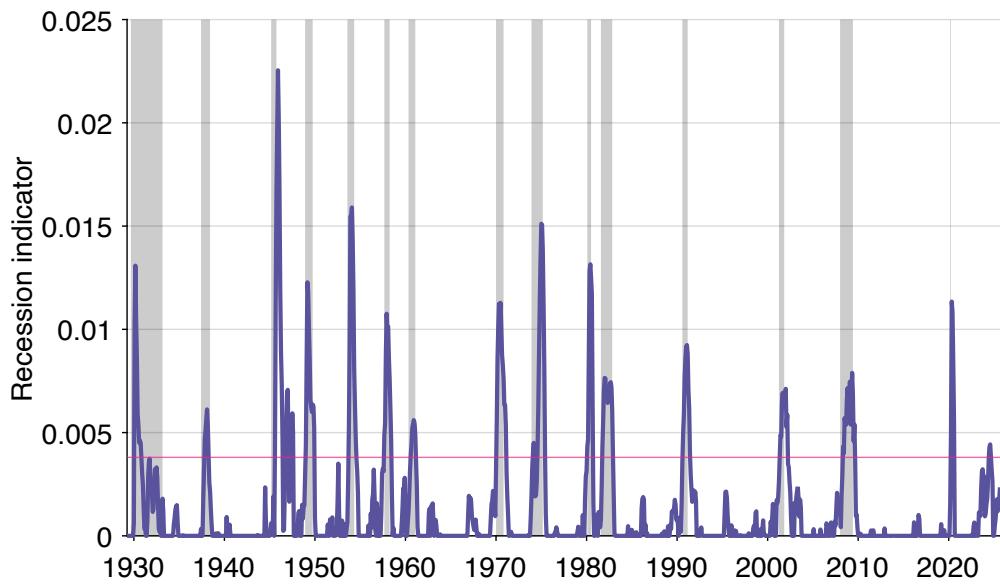


FIGURE A3. Classifier (3) in the classifier ensemble selected from 1929–2021 data

The classifier is described in table 1. The thick, purple line is the recession indicator on which the classifier is based. The thin, pink line is the recession threshold on which the classifier is based. Shaded areas indicate recessions dated by the NBER (2023).

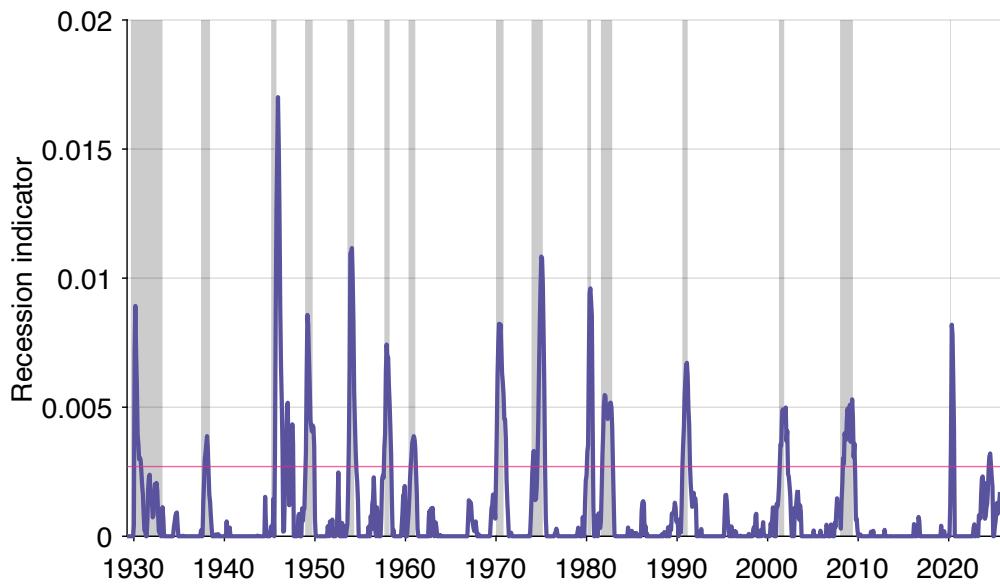


FIGURE A4. Classifier (4) in the classifier ensemble selected from 1929–2021 data

The classifier is described in table 1. The thick, purple line is the recession indicator on which the classifier is based. The thin, pink line is the recession threshold on which the classifier is based. Shaded areas indicate recessions dated by the NBER (2023).

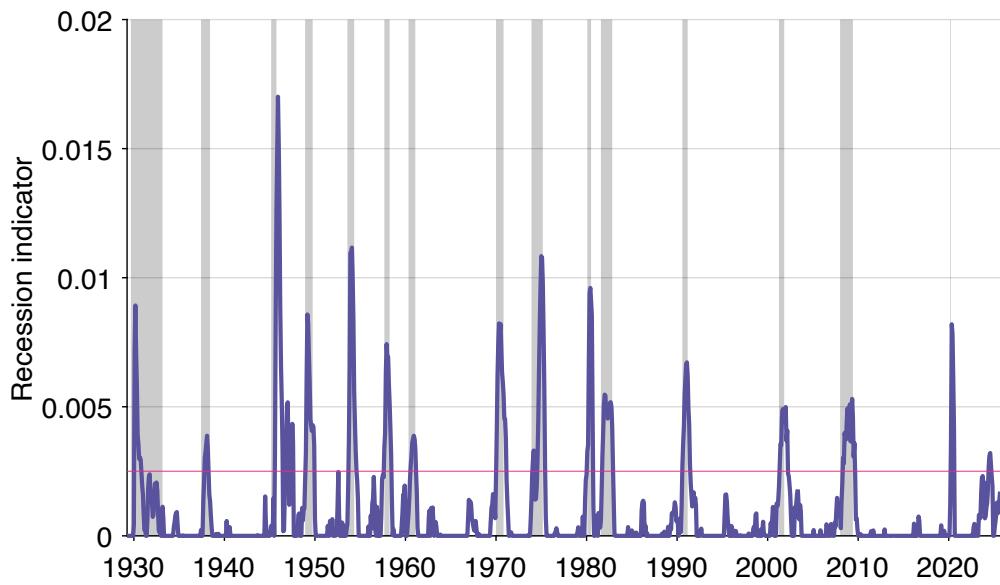


FIGURE A5. Classifier (5) in the classifier ensemble selected from 1929–2021 data

The classifier is described in table 1. The thick, purple line is the recession indicator on which the classifier is based. The thin, pink line is the recession threshold on which the classifier is based. Shaded areas indicate recessions dated by the NBER (2023).

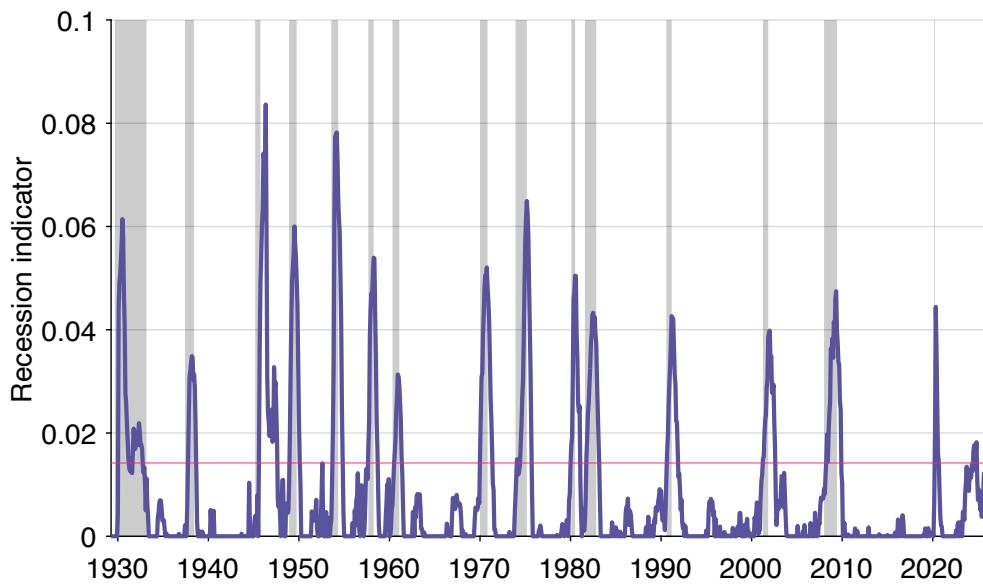


FIGURE A6. Classifier (6) in the classifier ensemble selected from 1929–2021 data

The classifier is described in table 1. The thick, purple line is the recession indicator on which the classifier is based. The thin, pink line is the recession threshold on which the classifier is based. Shaded areas indicate recessions dated by the NBER (2023).

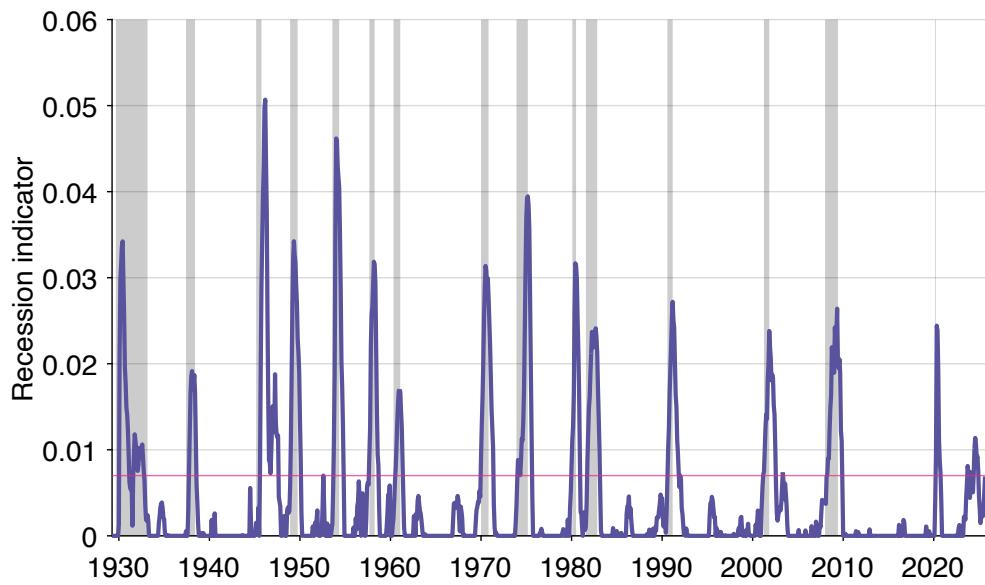


FIGURE A7. Classifier (7) in the classifier ensemble selected from 1929–2021 data

The classifier is described in table 1. The thick, purple line is the recession indicator on which the classifier is based. The thin, pink line is the recession threshold on which the classifier is based. Shaded areas indicate recessions dated by the NBER (2023).

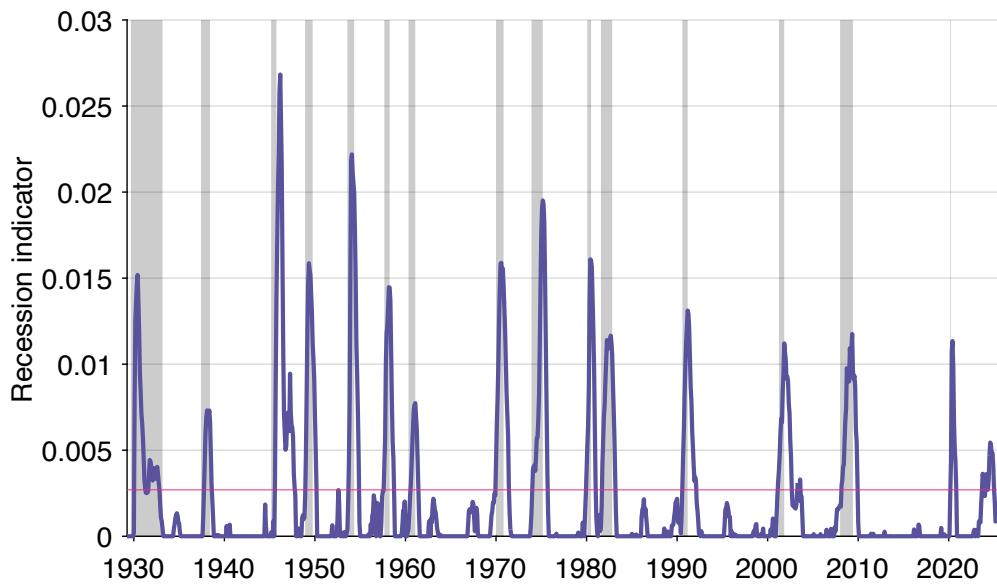


FIGURE A8. Classifier (8) in the classifier ensemble selected from 1929–2021 data

The classifier is described in table 1. The thick, purple line is the recession indicator on which the classifier is based. The thin, pink line is the recession threshold on which the classifier is based. Shaded areas indicate recessions dated by the NBER (2023).

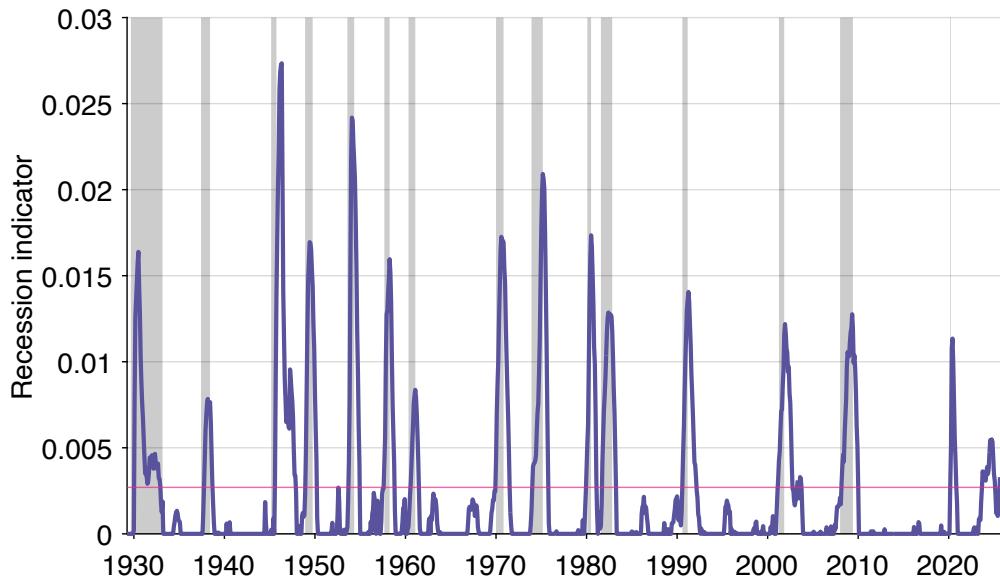


FIGURE A9. Classifier (9) in the classifier ensemble selected from 1929–2021 data

The classifier is described in table 1. The thick, purple line is the recession indicator on which the classifier is based. The thin, pink line is the recession threshold on which the classifier is based. Shaded areas indicate recessions dated by the NBER (2023).

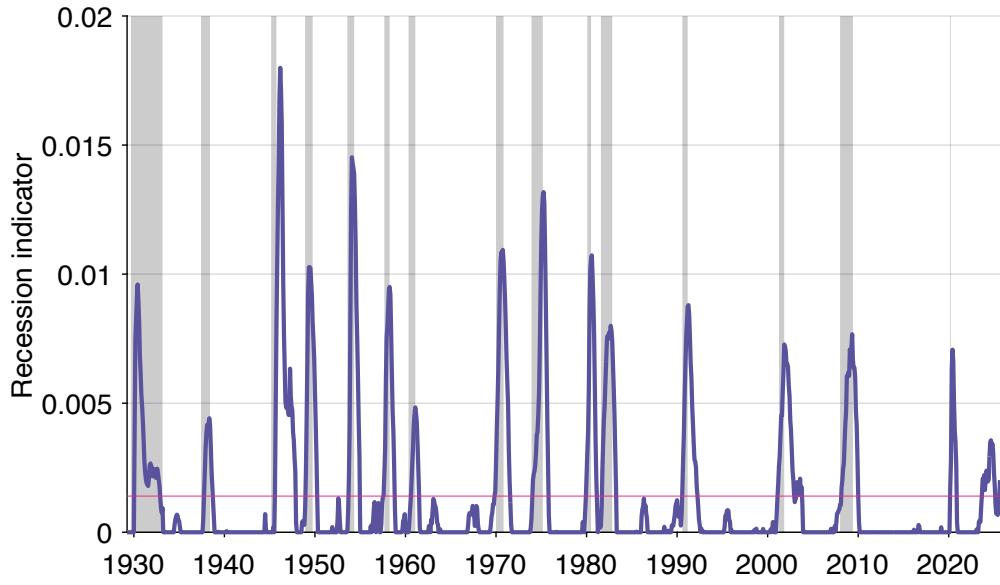


FIGURE A10. Classifier (10) in the classifier ensemble selected from 1929–2021 data

The classifier is described in table 1. The thick, purple line is the recession indicator on which the classifier is based. The thin, pink line is the recession threshold on which the classifier is based. Shaded areas indicate recessions dated by the NBER (2023).

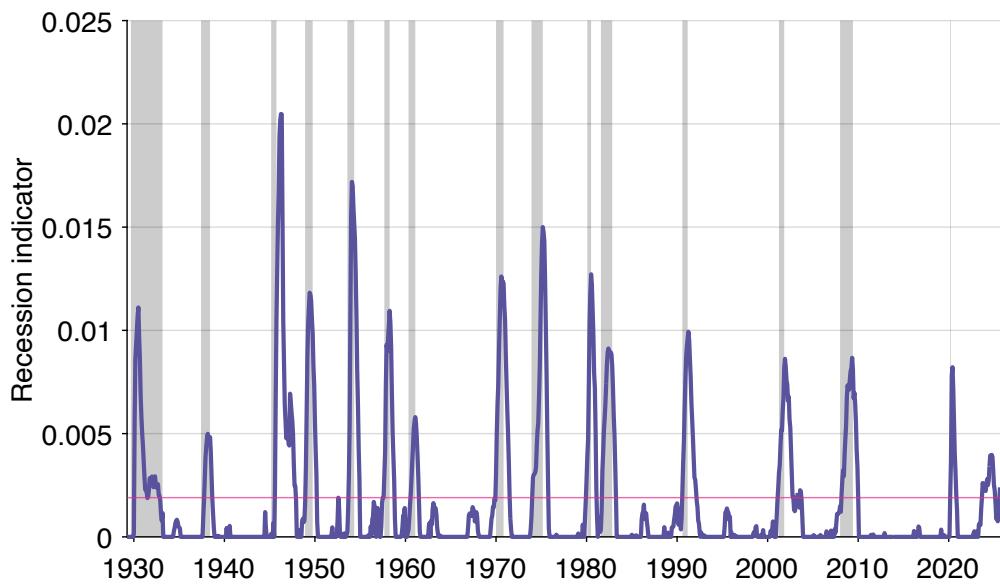


FIGURE A11. Classifier (11) in the classifier ensemble selected from 1929–2021 data

The classifier is described in table 1. The thick, purple line is the recession indicator on which the classifier is based. The thin, pink line is the recession threshold on which the classifier is based. Shaded areas indicate recessions dated by the NBER (2023).

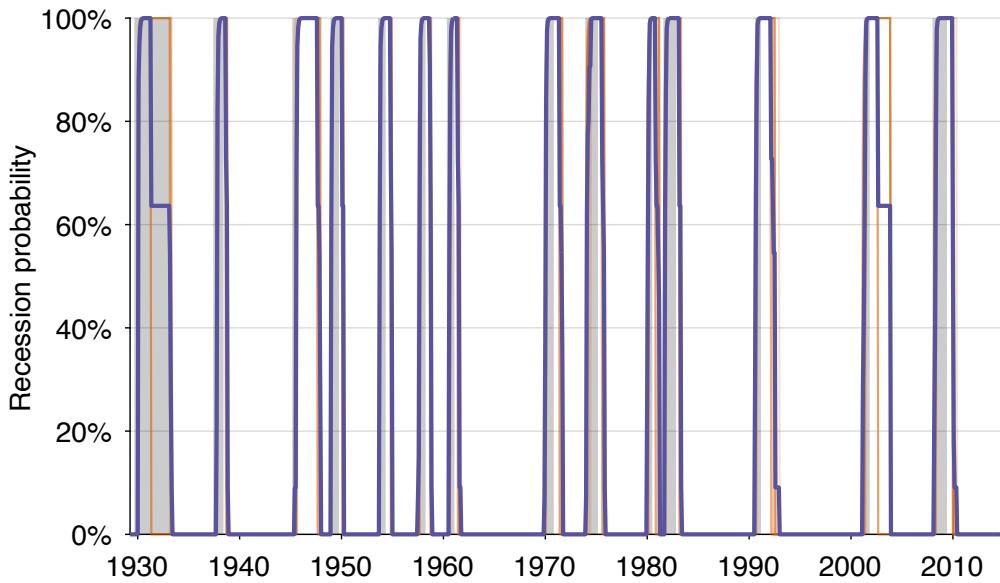


FIGURE A12. 2015 backtest: In-sample recession probability, April 1929–December 2014

The classifier ensemble used in the backtest is described in table A2. Shaded areas indicate recessions dated by the NBER (2023).

TABLE A2. Classifier ensemble selected in the 2015 backtest

Smoothing method	Smoothing parameter α	Turning horizon β	Box-Cox parameter γ	Combination method	Combination weight δ	Threshold ζ	Training error	
							Mean	Standard deviation
(1)	simple	8m	1m	0	u-v	0.7	1.70	3.4
(2)	simple	4m	4m	1	min-max	1	0.23	3.2
(3)	simple	3m	8m	0.7	min-max	1	0.84	3.2
(4)	exponential	0.5	5m	0.9	min-max	1	0.38	2.4
(5)	exponential	0.5	5m	1	min-max	1	0.27	2.3
(6)	exponential	0.5	5m	1	min-max	1	0.25	2.2
(7)	exponential	0.7	10m	0.6	min-max	1	1.42	2.1
(8)	exponential	0.5	8m	0.7	min-max	1	0.70	1.9
(9)	exponential	0.4	8m	0.9	min-max	1	0.27	1.7
(10)	exponential	0.3	7m	1	min-max	1	0.14	1.6
(11)	exponential	0.4	8m	1	min-max	1	0.19	2.2

The ensemble comprises the classifiers on the anticipation-precision frontier with a standard deviation of errors below 3 months. The ensemble is selected from training data covering April 1929–December 2014. The ensemble is then evaluated over testing data covering January 2015–December 2021. The simple smoothing method with parameter α is given by (1) and (2); the exponential smoothing method with parameter α is given by (3) and (4). The turning horizon β enters the construction of the classifiers through (5) and (7). The Box-Cox parameter γ enters the construction of the classifiers through (9) and (10). The u-v combination method with weight δ is given by (11); the min-max combination method with weight δ is given by (12).

TABLE A3. Classifier ensemble selected in the 2005 backtest

Smoothing method	Smoothing parameter α	Turning horizon β	Box-Cox parameter γ	Combination method	Combination weight δ	Threshold ζ	Training error	
							Mean	Standard deviation
(1)	simple	4m	1	min-max	1	0.23	3.1	1.6
(2)	simple	3m	0.7	min-max	1	0.84	3.1	1.7
(3)	exponential	0.5	5m	min-max	1	0.38	2.2	1.7
(4)	exponential	0.5	5m	min-max	1	0.27	2.2	1.8
(5)	exponential	0.5	5m	min-max	1	0.25	2.1	1.8
(6)	exponential	0.7	10m	0.8	min-max	1	0.78	2.0
(7)	exponential	0.4	7m	0.9	min-max	1	0.27	1.8
(8)	exponential	0.5	8m	0.7	min-max	1	0.70	1.8
(9)	exponential	0.4	8m	0.9	min-max	1	0.27	1.7
(10)	exponential	0.4	7m	1	min-max	1	0.19	1.6
(11)	exponential	0.4	8m	1	min-max	1	0.19	1.5

The ensemble comprises the classifiers on the anticipation-precision frontier with a standard deviation of errors below 3 months. The ensemble is selected from training data covering April 1929–December 2004. The simple smoothing method with parameter α is given by (1) and (2); the exponential smoothing method with parameter α is given by (3) and (4). The turning horizon β enters the construction of the classifiers through (5) and (7). The Box-Cox parameter γ enters the construction of the classifiers through (9) and (10). The u-v combination method with weight δ is given by (11); the min-max combination method with weight δ is given by (12).

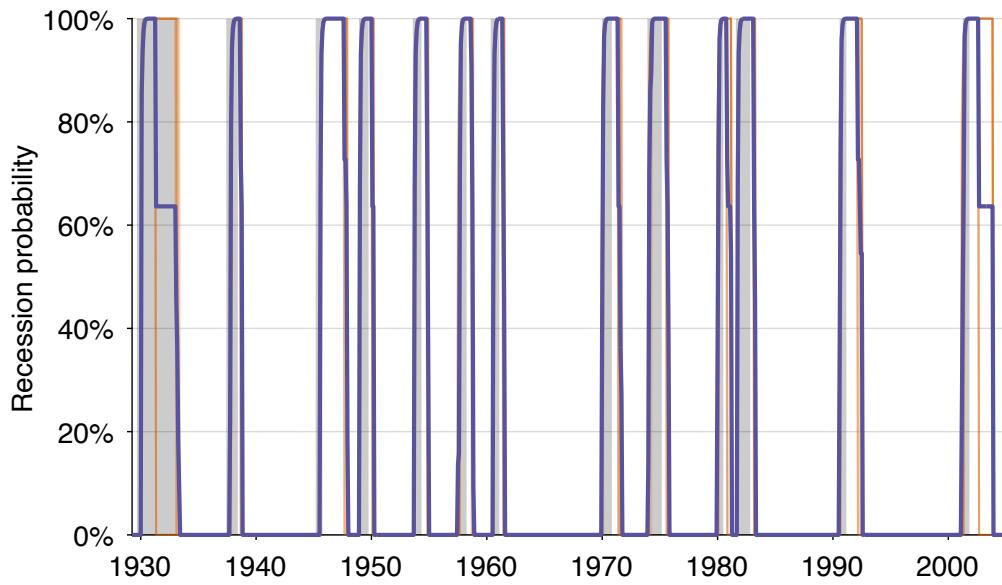


FIGURE A13. 2005 backtest: In-sample recession probability, April 1929–December 2004

The classifier ensemble used in the backtest is described in table A3. Shaded areas indicate recessions dated by the NBER (2023).

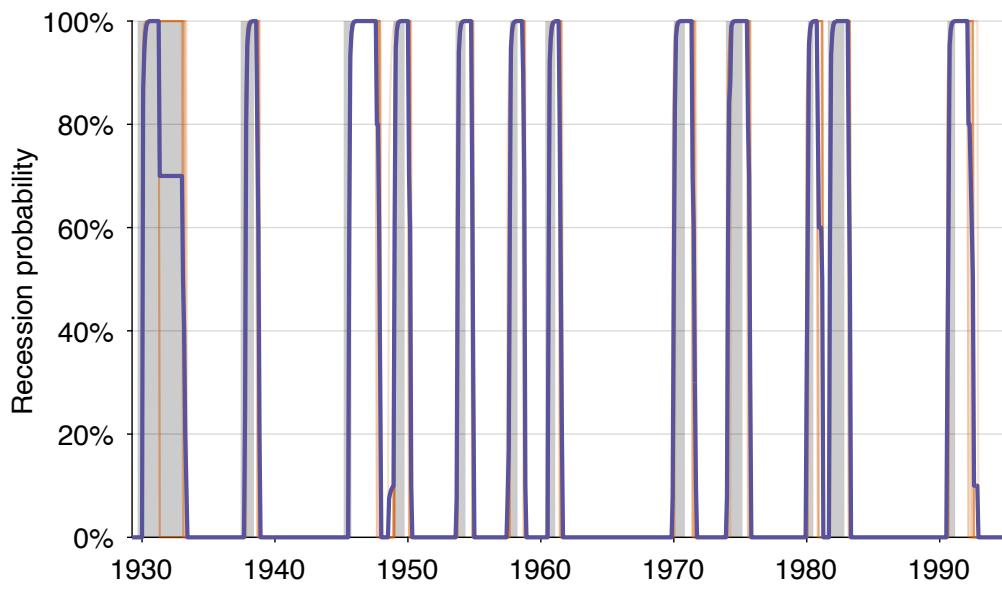


FIGURE A14. 1995 backtest: In-sample recession probability, April 1929–December 1994

The classifier ensemble used in the backtest is described in table A4. Shaded areas indicate recessions dated by the NBER (2023).

TABLE A4. Classifier ensemble selected in the 1995 backtest

	Smoothing method	Smoothing parameter α	Turning horizon β	Box-Cox parameter γ	Combination method	Combination weight δ	Threshold ζ	Training error	
								Mean	Standard deviation
(1)	simple	4m	4m	1	min-max	1	0.23	3.2	1.7
(2)	simple	3m	8m	0.7	min-max	1	0.84	3.2	1.7
(3)	simple	4m	8m	0.9	min-max	1	0.32	2.9	1.8
(4)	exponential	0.5	5m	1	min-max	1	0.27	2.3	1.8
(5)	exponential	0.5	5m	1	min-max	1	0.25	2.2	1.9
(6)	exponential	0.5	8m	0.7	min-max	1	0.70	2.0	1.9
(7)	exponential	0.4	7m	0.9	min-max	1	0.27	1.9	1.9
(8)	exponential	0.3	7m	1	min-max	1	0.14	1.8	2.2
(9)	exponential	0.4	7m	1	min-max	1	0.19	1.8	2.3
(10)	simple	6m	1m	0.6	min-max	0.9	0.18	1.7	2.5

The ensemble comprises the classifiers on the anticipation-precision frontier with a standard deviation of errors below 3 months. The ensemble is selected from training data covering April 1929–December 1994. The simple smoothing method with parameter α is given by (1) and (2); the exponential smoothing method with parameter α is given by (3) and (4). The turning horizon β enters the construction of the classifiers through (5) and (7). The Box-Cox parameter γ enters the construction of the classifiers through (9) and (10). The u-v combination method with weight δ is given by (11); the min-max combination method with weight δ is given by (12).

TABLE A5. Classifier ensemble selected in the 1985 backtest

	Smoothing method	Smoothing parameter α	Turning horizon β	Box-Cox parameter γ	Combination method	Combination weight δ	Threshold ζ	Training error	
								Mean	Standard deviation
(1)	simple	4m	4m	1	min-max	1	0.23	3.3	1.7
(2)	simple	3m	8m	0.7	min-max	1	0.84	3.3	1.8
(3)	simple	4m	8m	0.9	min-max	1	0.32	3.0	1.8
(4)	exponential	0.5	5m	1	min-max	1	0.25	2.4	1.8
(5)	exponential	0.7	10m	0.8	min-max	1	0.78	2.3	1.9
(6)	exponential	0.4	7m	0.9	min-max	1	0.27	2.1	1.9
(7)	exponential	0.3	7m	1	min-max	1	0.14	2.0	2.2
(8)	exponential	0.4	7m	1	min-max	1	0.19	1.9	2.3
(9)	simple	6m	1m	0.6	min-max	0.9	0.18	1.7	2.6

The ensemble comprises the classifiers on the anticipation-precision frontier with a standard deviation of errors below 3 months. The ensemble is selected from training data covering April 1929–December 1984. The simple smoothing method with parameter α is given by (1) and (2); the exponential smoothing method with parameter α is given by (3) and (4). The turning horizon β enters the construction of the classifiers through (5) and (7). The Box-Cox parameter γ enters the construction of the classifiers through (9) and (10). The u-v combination method with weight δ is given by (11); the min-max combination method with weight δ is given by (12).

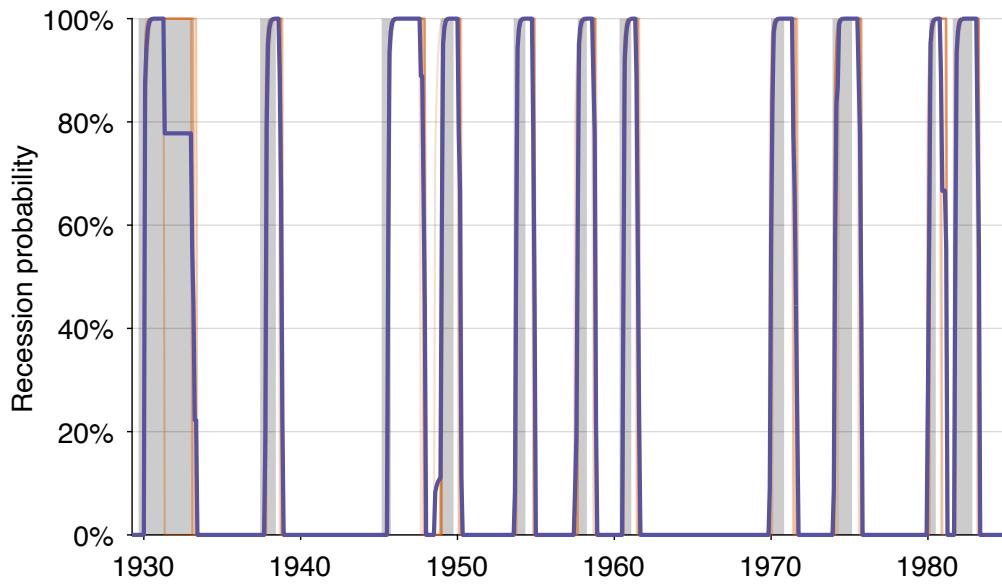


FIGURE A15. 1985 backtest: In-sample recession probability, April 1929–December 1984

The classifier ensemble used in the backtest is described in table A5. Shaded areas indicate recessions dated by the NBER (2023).

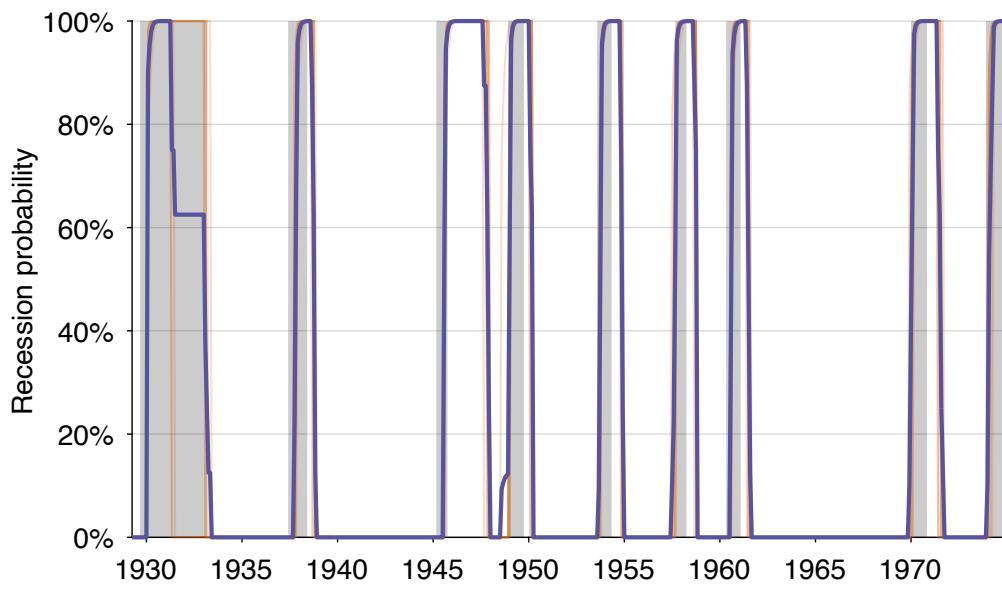


FIGURE A16. 1975 backtest: In-sample recession probability, April 1929–December 1974

The classifier ensemble used in the backtest is described in table A6. Shaded areas indicate recessions dated by the NBER (2023).

TABLE A6. Classifier ensemble selected in the 1975 backtest

	Smoothing method	Smoothing parameter α	Turning horizon β	Box-Cox parameter γ	Combination method	Combination weight δ	Threshold ζ	Training error	
								Mean	Standard deviation
(1)	simple	3m	7m	0.9	min-max	1	0.45	3.7	1.7
(2)	simple	4m	4m	1	min-max	1	0.23	3.6	1.7
(3)	simple	3m	8m	0.7	min-max	1	0.84	3.6	1.8
(4)	exponential	0.5	5m	1	min-max	1	0.25	2.7	1.8
(5)	exponential	0.4	7m	0.9	min-max	1	0.27	2.4	1.8
(6)	exponential	0.3	7m	1	min-max	1	0.14	2.3	2.2
(7)	exponential	0.4	7m	1	min-max	1	0.19	2.2	2.4
(8)	simple	6m	1m	0.6	min-max	0.9	0.18	1.9	2.9

The ensemble comprises the classifiers on the anticipation-precision frontier with a standard deviation of errors below 3 months. The ensemble is selected from training data covering April 1929–December 1974. The simple smoothing method with parameter α is given by (1) and (2); the exponential smoothing method with parameter α is given by (3) and (4). The turning horizon β enters the construction of the classifiers through (5) and (7). The Box-Cox parameter γ enters the construction of the classifiers through (9) and (10). The u-v combination method with weight δ is given by (11); the min-max combination method with weight δ is given by (12).

TABLE A7. Classifier ensemble selected in the 1965 backtest

	Smoothing method	Smoothing parameter α	Turning horizon β	Box-Cox parameter γ	Combination method	Combination weight δ	Threshold ζ	Training error	
								Mean	Standard deviation
(1)	simple	11m	2m	1	u-v	1	0.22	6.0	1.1
(2)	simple	11m	1m	1	u-v	1	0.09	5.1	1.1
(3)	simple	8m	1m	0.3	u-v	0.6	0.80	3.7	1.3
(4)	simple	8m	1m	0	u-v	0.8	1.82	3.6	1.3
(5)	simple	8m	1m	0.1	u-v	0.5	1.52	3.4	1.4
(6)	simple	8m	1m	0.3	u-v	0.5	0.70	3.3	1.5
(7)	simple	8m	1m	0.2	min-max	0.6	0.77	3.1	1.6
(8)	simple	11m	2m	0.3	u-v	1	0.40	3.1	1.7
(9)	simple	6m	1m	0	u-v	0.4	2.36	2.6	1.8
(10)	simple	6m	1m	0	u-v	0.4	2.18	2.4	1.9

The ensemble comprises the classifiers on the anticipation-precision frontier with a standard deviation of errors below 3 months. The ensemble is selected from training data covering April 1929–December 1964. The simple smoothing method with parameter α is given by (1) and (2); the exponential smoothing method with parameter α is given by (3) and (4). The turning horizon β enters the construction of the classifiers through (5) and (7). The Box-Cox parameter γ enters the construction of the classifiers through (9) and (10). The u-v combination method with weight δ is given by (11); the min-max combination method with weight δ is given by (12).

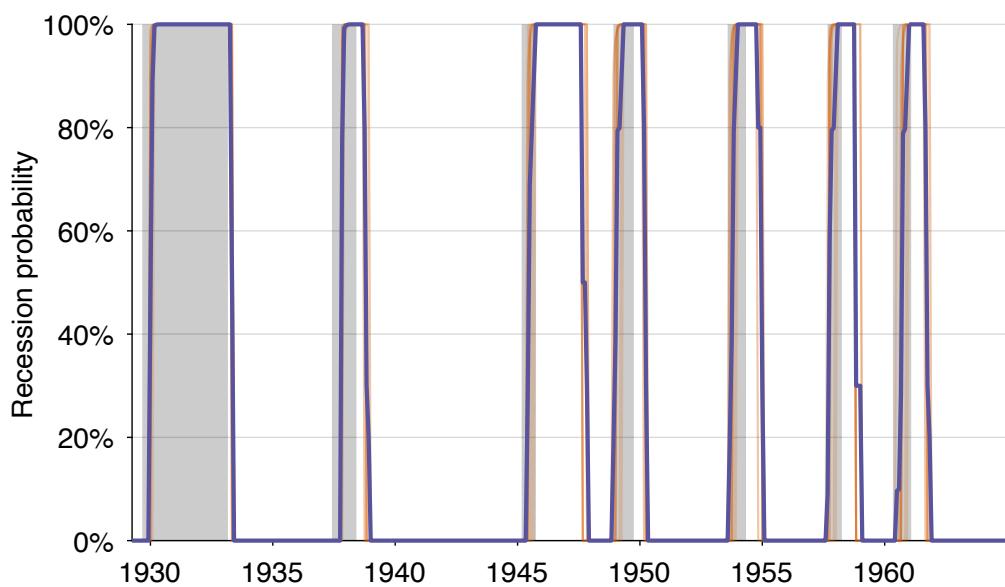


FIGURE A17. 1965 backtest: In-sample recession probability, April 1929–December 1964

The classifier ensemble used in the backtest is described in table A7. Shaded areas indicate recessions dated by the NBER (2023).