Early and Accurate Recession Detection Using Classifiers on the Anticipation-Precision Frontier

Pascal Michaillat

July 2025

This paper develops a new algorithm for detecting US recessions in real time. The algorithm constructs millions of recession classifiers by combining unemployment and vacancy data to reduce detection noise. Classifiers are then selected to avoid both false negatives (missed recessions) and false positives (nonexistent recessions). The selected classifiers are therefore perfect, in that 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 optimizes early detection without sacrificing precision. On average, over 1929–2021, the classifier ensemble signals recessions 2.2 months after their true onset, with a standard deviation of detection errors of 1.9 months. Applied to May 2025 data, the classifier ensemble gives a 71% probability that the US economy is currently in recession. A placebo test and backtests confirm the algorithm's reliability. The classifier ensembles trained on 1929–2004, 1929–1984, and 1929–1964 data in backtests give a current recession probability of 58%, 83%, and 25%, respectively.

University of California, Santa Cruz. I thank 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 (NBER 2021) occurs with a substantial delay—sometimes as late as 12 months after the recession's onset. As the NBER (2008) 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 (Stock and Watson 1989; Chauvet 1998; Chauvet and Piger 2008; Leamer 2008; Hamilton 2011; Giusto and Piger 2017; Huang and Startz 2020; Mertens 2022; Keil, Leamer, and Li 2023; Furno and Giannone 2024).[1]

Among the most effective approaches for real-time recession detection are threshold rules based on unemployment data (Schannep 2008; Hatzius and Stehn 2012; Sahm 2019; Sun, Feng, and Hu 2021; Philips 2024; O'Trakoun and Scavette 2025; Philips and Wang 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

[1]There also exist related algorithms that identify past US recessions retrospectively (Bry and Boschan 1971; Hamilton 1989; Harding and Pagan 2002, 2006; Chauvet and Hamilton 2006; Startz 2008; Stock and Watson 2010; Berge and Jorda 2011; Stock and Watson 2014), and that predict future US recessions at various time horizons (Stock and Watson 1993; Estrella and Mishkin 1998; Qi 2001; Chauvet and Potter 2005; Dueker 2005; Wright 2006; Kauppi and Saikkonen 2008; Nyberg 2010; Chen, Iqbal, and Lai 2011; Ng 2014; Christiansen, Eriksen, and Moller 2014; Berge 2015; Fornaro 2016; Bauer and Mertens 2018; Davig and Hall 2019; Vrontos, Galakis, and Vrontos 2021; Galvao and Owyang 2022; Leamer 2024; Diercks, Soques, and Wu 2024).

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 fails 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 time series to estimate US business cycle turning points. Ng (2014) considers 1,500 time series—132 distinct series and their lags—to forecast US recessions at various horizons. Fornaro (2016) considers 133 time 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 tens of thousands of 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 measures of increase for unemployment and decrease for vacancies. The increases and decreases are also scaled to cover the many possibilities from level changes to percentage changes. The individual unemployment and vacancy indicators are finally combined into single recession indicators through weighted averages or weighted min-max, using many different weights.

I then build millions of recession classifiers by applying many different recession thresholds to the previously constructed indicators. I train the algorithm such that only classifiers with no false negatives (actual recessions missed by the classifiers) or false positives (non-recessions mistakenly detected by the classifiers) between 1929 and 2021 are selected. There are many, many such perfect classifiers: more than 2 million classifiers detect the 15 recessions that occurred between 1929 and 2021 without false positives.

The challenge is to evaluate these perfect classifiers. 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 the presence of many perfect classifiers. Instead, I plot each classifier's mean detection error against the standard deviation of detection errors. I then select classifiers on the low-delay, 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. Hence, the frontier helps identify classifiers that balance 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 risk. 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, 7 classifiers from the high-precision segment of the frontier are selected.

From this ensemble of 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 already 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 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. Finally, I average the probability that the recession has started across the 7 classifiers from the classifier ensemble used for detection. Each classifier provides a recession-start probability; the ensemble averages these probabilities to yield a single recession risk.

I apply the trained classifiers to current data to obtain a real-time assessment of recession risk. As of May 2025, the recession probability given by the 7 classifiers is 71%, suggesting a high likelihood that a recession has begun at this point in time. The underlying reason is that 5 of the 7 classifiers have been activated in the past 2 years. 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 stark failure in the first-lady placebo test demonstrates that the model identifies genuine economic signals rather than spuriously fitting the data.

In addition, I backtest the detection algorithm by shortening the training period and evaluating performance on subsequent out-of-sample recessions. For example, I train the system on data up to 2004 (13 in-sample recessions) and test its performance on later recessions. I repeat backtesting on data up to 1984 (11 in-sample recessions) to detect subsequent recessions, and on data up to 1964 (7 in-sample recessions) to detect more subsequent recessions. Through backtesting, I find that the detection algorithm is reliable. Classifier ensembles trained on limited historical data (1929–2004, 1929–1984, and even 1929–1964) 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 trained on 1929–2004, 1929–1984, and 1929–1964 data give a current recession probability of 58%, 83%, and 25%, respectively. The algorithm only breaks down once I restrict training to 1929–1955 (5 in-sample recessions). Then, some of the classifiers in the selected ensemble miss some of the out-of-sample recessions.

Unlike the algorithm presented in this paper, the NBER's Dating Committee relies mostly on product market metrics (production, sales, consumption) to date recessions. The committee states that they do not look at the unemployment rate because it is lagging and noisy, and they have never mentioned looking at the vacancy rate (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 that the combination of unemployment and vacancy rates consistently outperforms industrial production in providing early and accurate recession signals; in fact the unemployment rate alone outperforms industrial production. Even when combined, unemployment and industrial production data do not achieve the same level of anticipation and precision as the unemployment-vacancy pairing, suggesting that current official dating practices might benefit from a greater emphasis on unemployment and vacancies.

2. Data

In this section I present the data on US recessions, unemployment, and job vacancies analyzed in the paper. I use data from April 1929 to May 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.

However, the official dates are published many months after recessions have actually started (NBER 2021). 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 7.3 months after a recession's onset.

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

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 May 2025, the unemployment rate is computed in the usual way: it is the number of jobseekers measured by the BLS (2020b) from the Current Population Survey (CPS), divided by the civilian labor force measured by the BLS (2024a) 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.

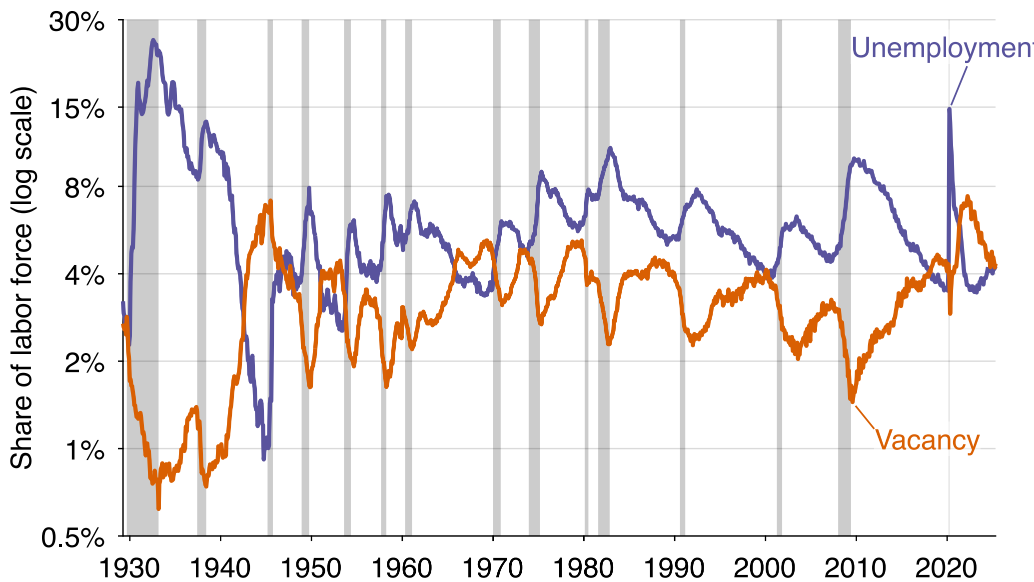


FIGURE 1. Monthly US unemployment and vacancy rates, April 1929–May 2025

The unemployment rate is computed from data produced by Petrosky-Nadeau and Zhang (2021) and the BLS (2020b, 2024a). The vacancy rate is computed from data produced by Petrosky-Nadeau and Zhang (2021), Barnichon (2010), and the BLS (2024a,c). Shaded areas indicate recessions dated by the NBER (2023).

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 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.[2]

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 May 2025, I compute the vacancy rate as the number of job openings measured by the BLS (2024c) from the Job Openings and Labor Turnover

[2] Petrosky-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 2024).

Survey (JOLTS), divided by the size of the labor force measured by the BLS (2024a) from the CPS. To best align labor force and vacancy data, I follow Michaillat and Saez (2024, 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 motivation for this shift is that the number of job openings from the JOLTS refers to the last business day of the month (Monday 31 March 2025), while the labor force from the CPS refers to the Sunday–Saturday week including the 12th of the month (Sunday 6 April 2025–Saturday 12 April 2025) (BLS 2020a, 2024b). So the number of job openings refers to a day that is closer to the next month's CPS reference week than to the current month's CPS reference week. In 2025, the CPS survey for April started in the same week as the JOLTS survey for March.

I then splice the three vacancy series to create a continuous vacancy rate covering April 1929–May 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 2024e). 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 (2024c) is preliminary and updated one month after its initial release, to incorporate additional survey responses received from businesses and government agencies (BLS 2024d). 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 2024d, 2025). 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

In this section I construct real-time recession indicators by combining unemployment and vacancy data for the United States, April 1929–May 2025. The indicators are created in 4 steps: smoothing the data, detecting turning points, computing variations, and combining single-variable indicators. Later I use these indicators together with appropriate thresholds to detect recessions in real time.

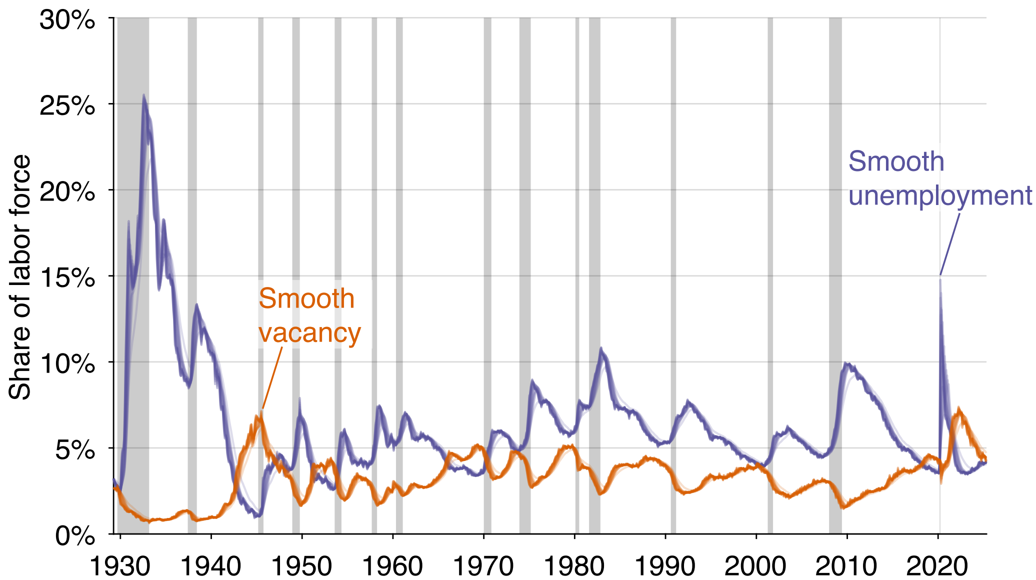


FIGURE 2. Smoothed US unemployment and vacancy rates, April 1929–May 2025

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 use simple trailing averages. In that case the smoothed unemployment rate is given by

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

where the parameter $0 \leq \alpha$ governs the amount of smoothing. The case $\alpha = 0$ corresponds to no smoothing; the case $\alpha = 11$ corresponds to a 12-month moving average. The Sahm and Michez rules have $\alpha = 3$. I compute 12 smoothed unemployment series by setting $\alpha = 0, 1, 2, \ldots, 11$ (figure 2).

The vacancy rate is smoothed in the same way:

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

I compute 12 smoothed vacancy series by setting $\alpha = 0, 1, 2, \ldots, 11$ (figure 2).

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

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

where the parameter $\alpha \in (0, 1]$ governs the amount of smoothing. The case $\alpha = 1$ corresponds to no smoothing. I compute 10 additional smoothed unemployment series by setting $\alpha = 0.1, 0.2, 0.3, \ldots, 1$ (figure 2).

The vacancy rate is smoothed in the same way:

$$\bar{v}(t) = \alpha v(t) + (1 - \alpha)\bar{v}(t - 1). \tag{4}$$

I produce 10 additional smoothed vacancy series by setting $\alpha = 0.1, 0.2, 0.3, \ldots, 1$ (figure 2).

3.2. Detecting turning points

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

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

where $\beta = 1, 3, 6, 9, 12, \ldots, 36$ months is the horizon of the trailing minimum (figure 3A). The Sahm and Michez rules use $\beta = 12$. Then, the increase in unemployment rate from the turning point is computed as

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

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

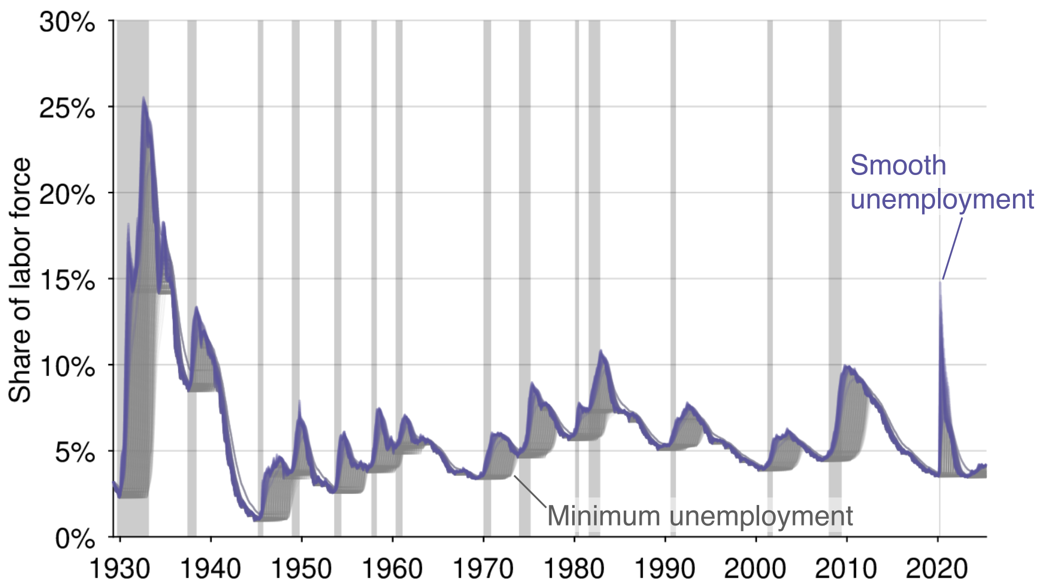
I proceed analogously to determine turning points in the vacancy rate. I first take the maximum of the vacancy rate at various monthly horizons:

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

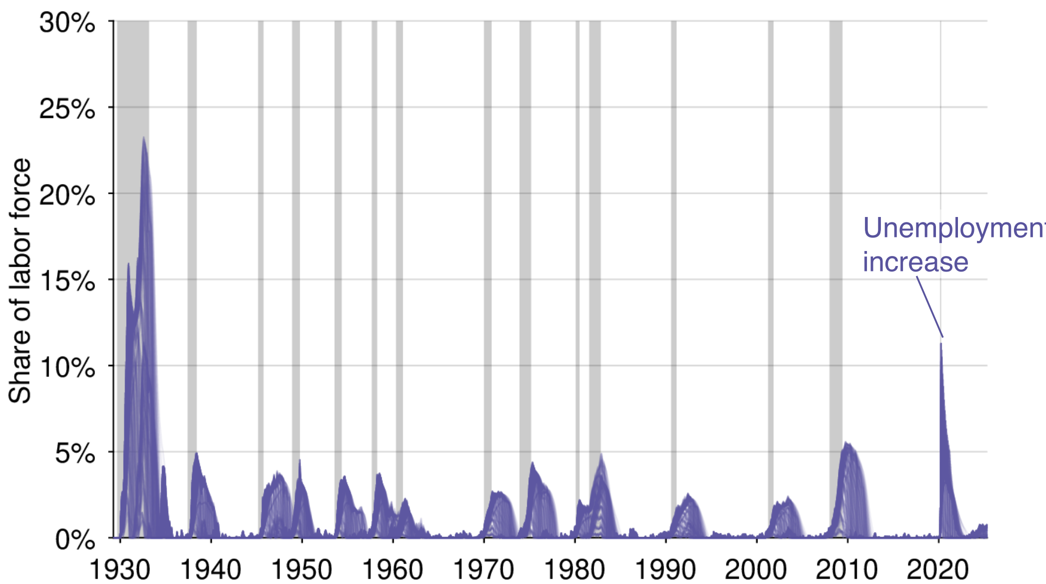
where $\beta = 1, 3, 6, 9, 12, \ldots, 36$ months (figure 4A). The Michez rule uses $\beta = 12$. Then, the decrease in vacancy rate from the turning point is computed as

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

The current vacancy rate is compared to values ranging from last month's vacancy rate to the maximum vacancy rate over the past 3 years (figure 4B).



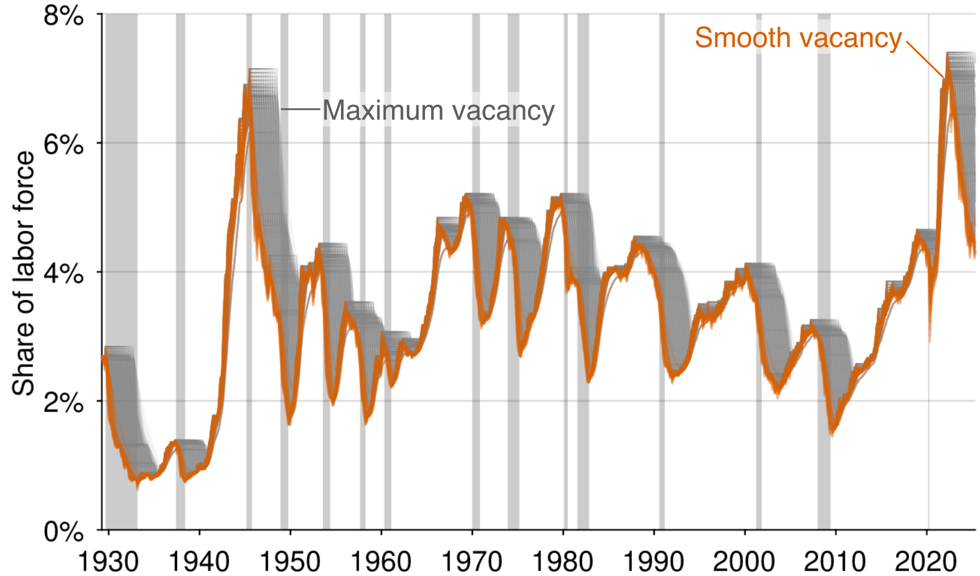
A. Trailing minimum of the unemployment rate



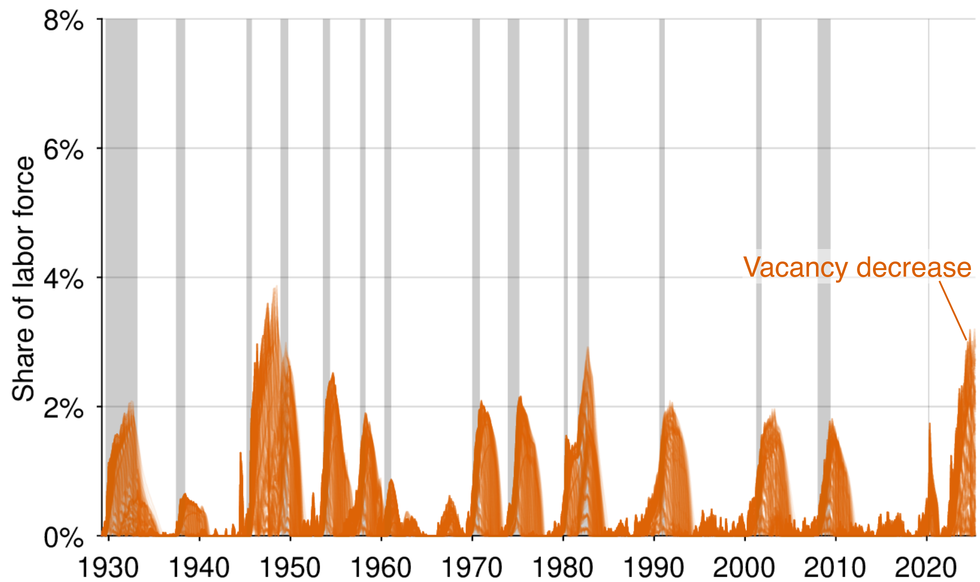
B. Increases in the unemployment rate

FIGURE 3. Increases in US unemployment rate, April 1929–May 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–May 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).

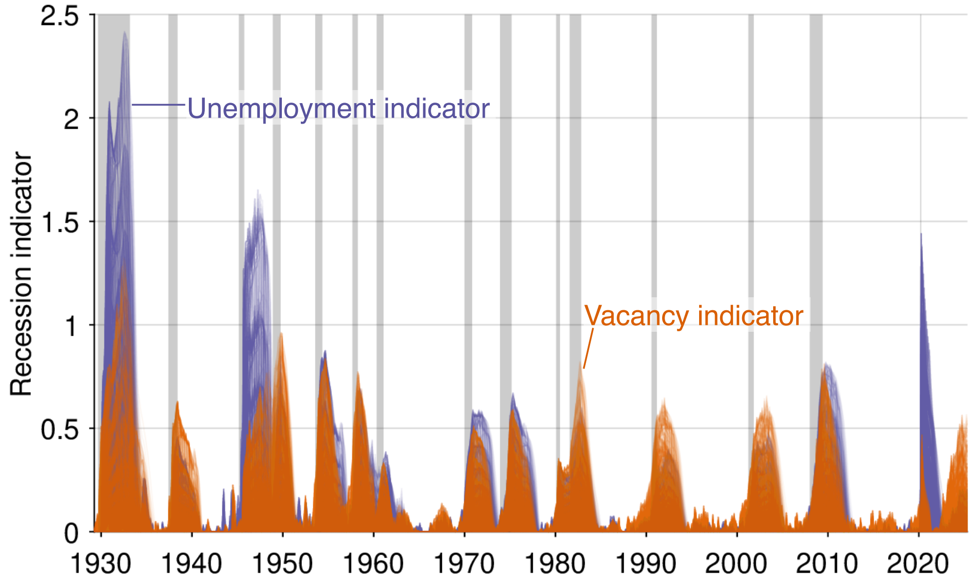


FIGURE 5. US unemployment and vacancy indicators, April 1929–May 2025

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).

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 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 artificially limit 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.

Formally, I consider all the unemployment indicators of the form

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

for values of the curvature parameter $\gamma = 0, 0.1, 0.2, 0.3, \ldots, 1$ (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 use $\gamma = 1$.

Similarly, I consider all the vacancy indicators of the form

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

for values of the curvature parameter $\gamma = 0, 0.1, 0.2, 0.3, \ldots, 1$ (figure 5). The Michez rule uses $\gamma = 1$.

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:

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

where $\delta = 0, 0.1, 0.2, 0.3 \ldots, 1$ is the weight on the unemployment indicator. The Sahm rule sets $\delta = 1$.

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. Motivated by this insight, I also consider the minimum and maximum of the unemployment and vacancy indicators, as well as all their linear combinations:

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

where $\delta = 0, 0.1, 0.2, 0.3 \ldots, 1$ is the weight on the minimum indicator. The Michez rule sets $\delta = 1$.

4. Constructing and evaluating recession classifiers

Having constructed the 69,212 indicators, I develop 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. The classification is recursive.

Initially, the economy is not in recession ($r(k, 0) = 0$) if the indicator is below the threshold ($i(0) < \zeta$). The economy is in recession ($r(k, 0) = 1$) if the indicator is above the threshold ($i(0) \geq \zeta$). In period t, we need to consider the state of the economy in period $t - 1$:

- 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 $i(t) > 0$.
  - 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 $i(t) < \zeta$.
  - 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 jth 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 zero for the economy to be classified as in expansion. This classification procedure adds a bit of computational complexity, but it provides a more logical classification, which produces less noisy classifiers.

4.2. Selecting perfect classifiers

For all indicators and possible thresholds between 1 basis point and 1, $\zeta = 0.0001, 0.0002, 0.0003, \dots, 1$, I compute the number of recessions detected by each individual classifier. I evaluate each classifier 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.

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).

I only keep perfect classifiers. Overall, the procedure produces 2,343,752 perfect classifiers, which detect the 15 recessions that occurred between 1929 and 2021 without false positives. The perfect classifiers are constructed from 65,952 different recession indicators, combined with 1,157 unique thresholds.

4.3. Evaluating perfect classifiers

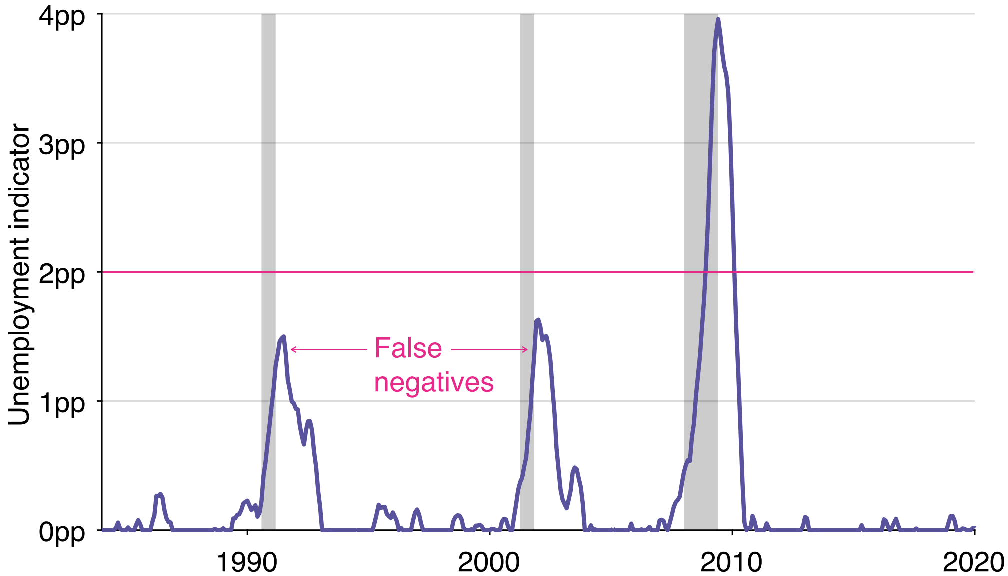
The challenge is to evaluate the many perfect classifiers. Standard evaluation methods, like ROC curves, rank classifiers based on the amount of false positives and false negatives that they generate.[3] They are unsuitable here given the number of perfect classifiers.

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, $d(k, j)$:

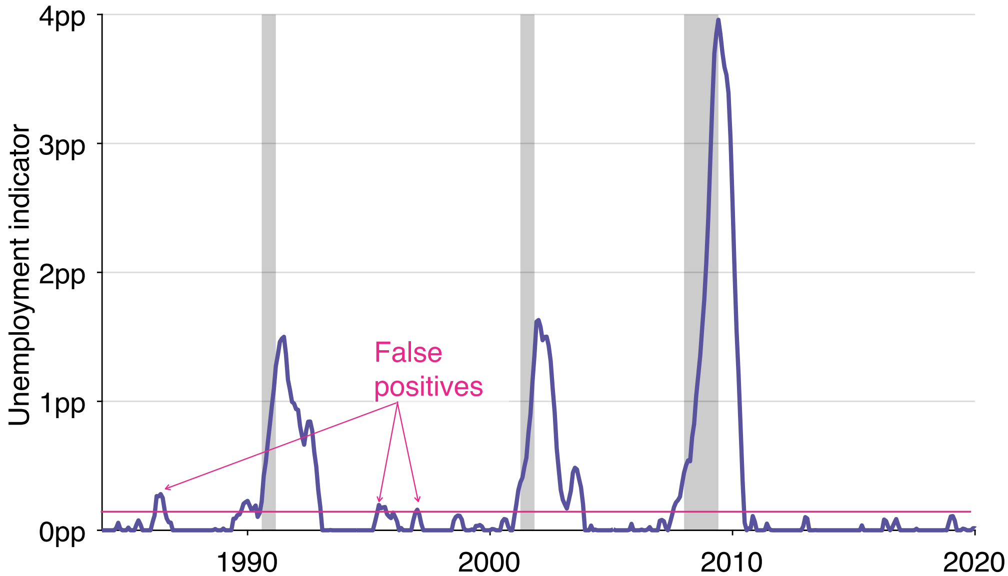
$$\epsilon(k, j) = d(k, j) - s(j). \tag{13}$$

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 some delay. If instead $\epsilon(k, j) < 0$, classifier k detects recession j with some anticipation.

[3] See for instance Murphy (2022, section 5.1.3) and Piger (2020, section 18.2.3).



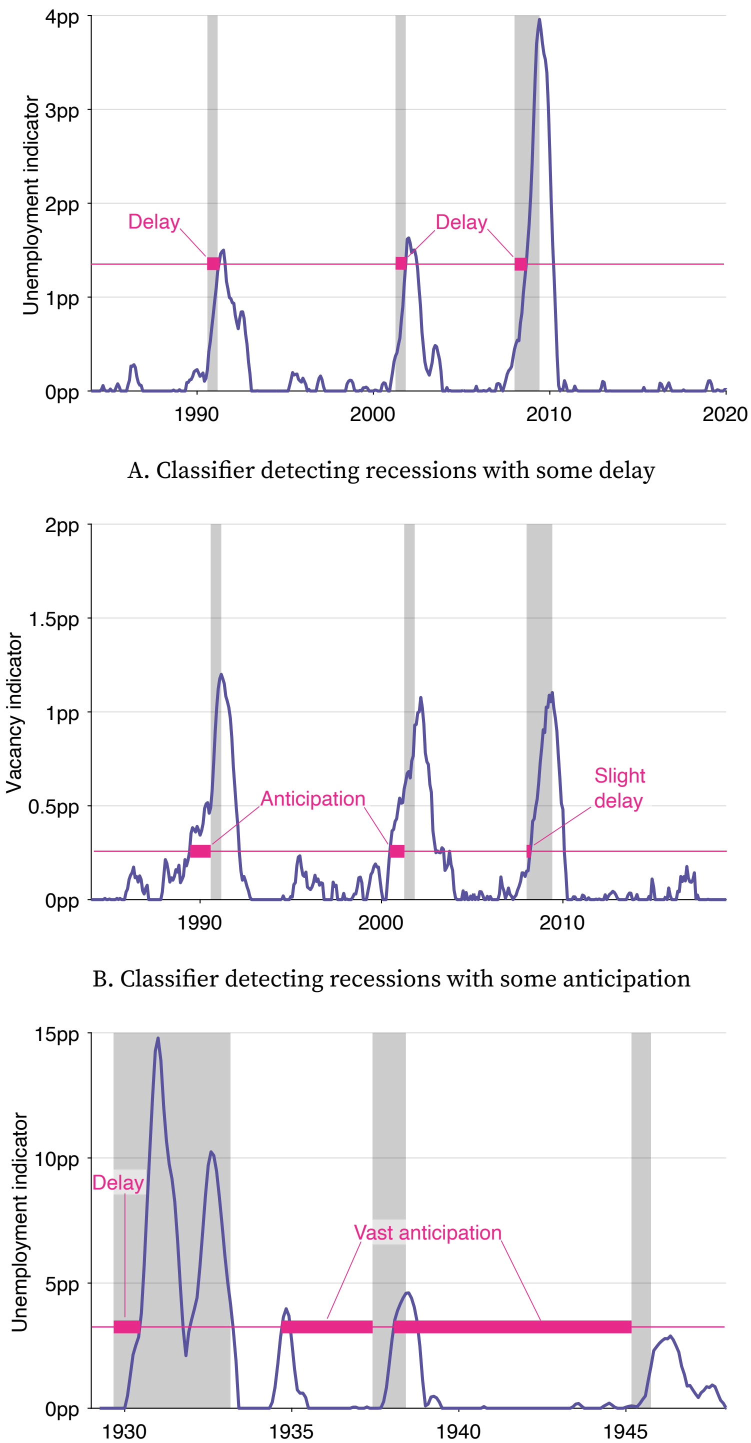
A. False negatives with high detection threshold



B. False positives with low detection threshold

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).



A. Classifier detecting recessions with some delay

B. Classifier detecting recessions with some anticipation

C. Classifier with imprecise recession detection

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).

Different classifiers, using different indicators and different thresholds, have different detection errors for each recession (figure 7).

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

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

where J is the number of recessions in the training period. Second, I compute the standard deviation of the detection errors:

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

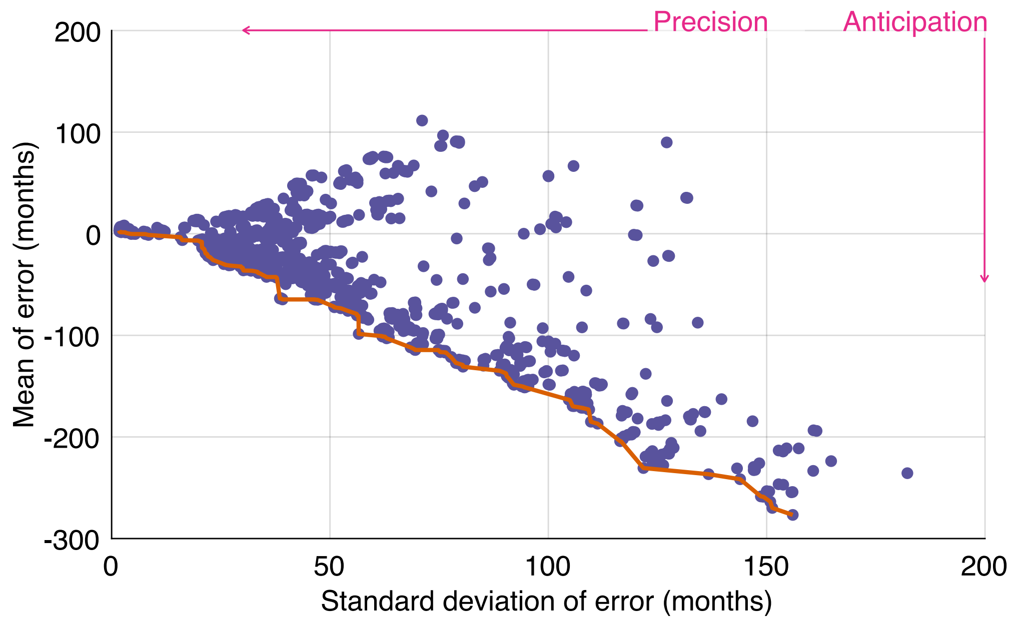
A classifier with higher anticipation (lower $\mu(k)$) is better, as it allows policymakers, firms, workers, investors, and other economic participants to foresee the recession and prepare for it. For instance, the classifier in figure 7B is superior to the classifier in figure 7A.

A classifier with better 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 actually 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.

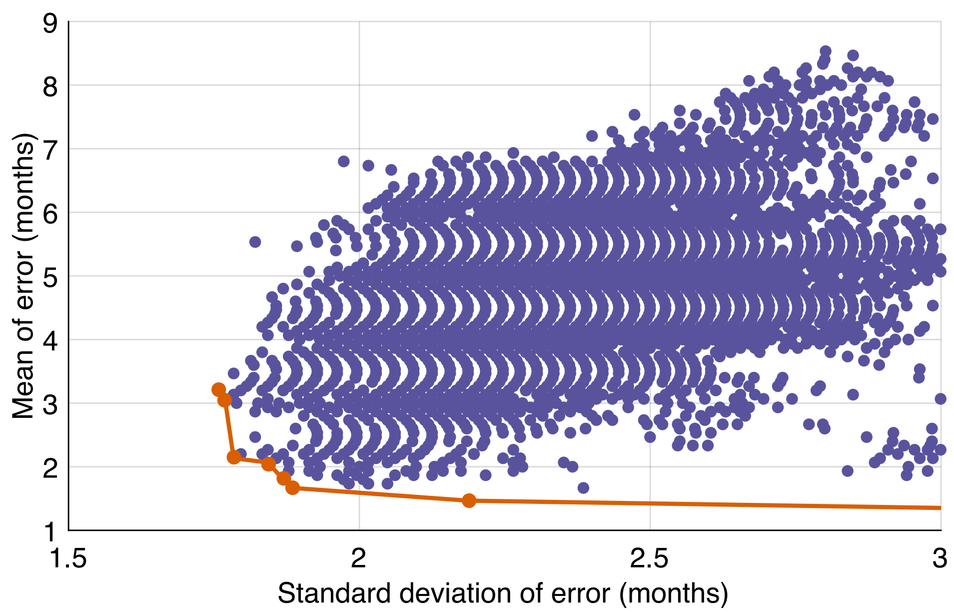
4.4. Finding the anticipation-precision frontier

Next, I plot the mean and standard deviation of the detection errors for the 2,343,752 perfect classifiers (figure 8A). I then select classifiers on the anticipation-precision frontier. 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 210 of the 2,343,752 perfect classifiers. The left-most classifier on the frontier has a delay (mean detection error) of 3.5 months and a precision (standard deviation of errors) of 1.6 months. The right-most classifier on the frontier has



A. Anticipation-precision frontier



B. Classifiers with the highest precision

FIGURE 8. Two million perfect recession classifiers for the United States, April 1929–December 2021

The figure displays the mean and standard deviation of the detection errors for 2,343,752 perfect classifiers, which detect the 15 recessions that occurred between 1929 and 2021 without false positives. The perfect classifiers are constructed from the recession indicators in figure 5 and recession thresholds between 1 basis point and 1. The mean error is computed from (14) and the standard deviation of errors is computed from (15). The anticipation-precision frontier comprises the classifiers with lowest mean error (highest anticipation) and lowest standard deviation of errors (highest precision).

an anticipation (– mean detection error) of 273 months and a precision of 129 months.

4.5. Cropping the anticipation-precision frontier

How should we pick a classifier on the frontier 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 $-\lambda$.

Not knowing the preferences of the policymaker, it is impossible to pick the desirable classifier that way. Instead, I pick all the classifiers on the frontier whose precision is below 3 months. These are classifiers for whom the 95% confidence interval for the recession's start date has a width below 1 year. Indeed, assuming a normally distributed detection error, the 95% confidence interval for the recession's start date has a width of 4 standard deviations, or $4 \times \sigma(k) < 12$ months. I pick this admittedly arbitrary threshold on the frontier because detecting a recession more than 6 months before its start does not seem extremely 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.[4]

The resulting classifier ensemble comprises 7 classifiers (table 1). These classifiers are perfect, so they detect the 15 recessions that occurred between 1929 and 2021 without producing false positives. The mean detection error of the classifiers ranges from 1.5 months to 3.2 months, with an average value of 2.2 months. The standard deviation of the detection errors ranges from 1.8 months to 2.2 months, with an average value of 1.9 months. 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 hindsight, not in real time, which allows them to place the start dates slightly earlier than the classifiers' detection dates.

The classifier ensemble can be organized in two clusters. The first consists of classifiers 1 and 2. They are broadly symmetric linear combinations of an unemployment indicator and a vacancy indicator, each representing the percentage change of the 8-month unemployment or vacancy average from the previous month's value. The second consists of classifiers 3–7. They are the minimum of an unemployment indicator and a vacancy indicator, each broadly representing the change of an exponentially smoothed

[4]The 3-month cutoff is a hyperparameter of the detection algorithm. In the future, it could be optimized via backtesting.

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

Classifier	1	2	3	4	5	6	7	
Smoothing method	simple	simple	exponential	exponential	exponential	exponential	exponential	
Smoothing horizon α	8m	8m	0.7	0.7	0.5	0.4	0.4	
Turning horizon β	1m	1m	9m	9m	9m	9m	9m	
Curving parameter γ	0.1	0	0.8	0.7	0.7	0.9	1	
Combination method	u-v	u-v	min-max	min-max	min-max	min-max	min-max	
Combination weight δ	0.6	0.6	1	1	1	1	1	
Threshold ζ	1.24pp	1.70pp	0.78pp	1.06pp	0.70pp	0.27pp	0.19pp	
Detection error (months)								Average
Mean	3.2	3.1	2.1	2.1	1.8	1.7	1.5	2.2
Standard deviation	1.8	1.8	1.8	1.8	1.9	1.9	2.2	1.9
Minimum	1	1	0	0	−1	−1	−2	−0.3
Maximum	7	7	5	5	5	5	5	5.6

The classifier ensemble is selected from the anticipation-precision frontier in figure 8A. The ensemble comprises the classifiers on the frontier with a standard deviation of errors below 3 months—as depicted in figure 8B. The simple smoothing method with horizon α is given by (1) and (2); the exponential smoothing method with horizon α is given by (3) and (4). The turning horizon β enters the construction of the classifiers through (5) and (7). The curving 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).

unemployment or vacancy rate from its 9-month extremum.

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 classifier 7 (whose detection error has mean of 1.5 months and standard deviation of 2.2 months). It is also a little less accurate and slower than classifiers 5 and 6.

The distance between the Michez-rule classifier and classifiers 5, 6, and 7 respectively are: $\sqrt{0.1^2 + 0.4^2} = 0.41$ months, $\sqrt{0.2^2 + 0.4^2} = 0.45$ 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. Its proximity to the frontier 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 and vacancy indicators alone (figure 9). The classifiers obtained from the minimum indicators tend to be more accurate and provide earlier recession signals than those obtained from the unemployment or vacancy indicators. Michaillat and Saez (2025) made this discovery by filtering the data in a specific way; but the result remains valid across a broad range of data 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 recessions are mostly 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, 2025). Therefore, a typical recession features both a decrease in vacancy rate and increase in unemployment rate, which are picked up by minimum indicators.

5. Computing recession probabilities

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.

A. Performance of unemployment classifiers

B. Performance of vacancy classifiers

C. Performance of minimum classifiers

FIGURE 9. Generalization of the Michez-rule results

The figure reproduces figure 8A. In addition, it displays the mean and standard deviation of the detection errors for classifiers built from unemployment indicators (panel A), for classifiers built from vacancy indicators (panel B), and for classifiers built from minimum indicators (panel C).

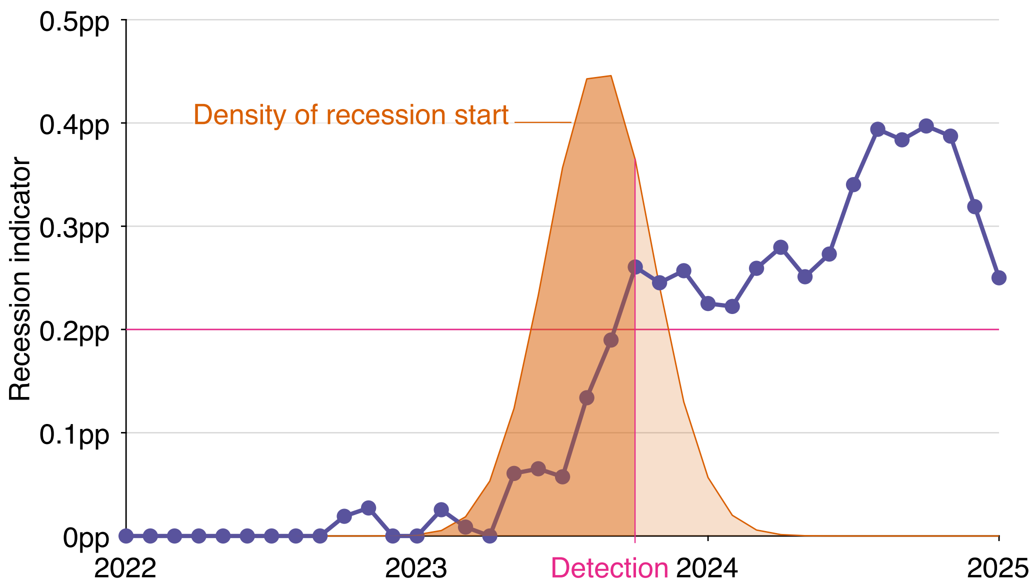


FIGURE 10. Computing the recession probability from a classifier

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

5.1. Recession probability from one 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 $\epsilon(k, j)$. I assume that classifier k's detection error $\epsilon(k)$ is normally distributed with mean $\mu(k)$ and standard deviation $\sigma(k)$—a convenient and neutral assumption. Then, it is easy to compute the probability that a new recession's start time, s, truly occurred before time t_1, given that the classifier detected a recession at time $d(k) = t_0 \leq t_1$:

$$
\begin{aligned}
P(k, t_1) = \mathbb{P}(s < t_1 \mid d(k) = t_0, \mu(k), \sigma(k)) &= \mathbb{P}(d(k) - s > t_0 - t_1 \mid \mu(k), \sigma(k)) \\
&= \mathbb{P}(\epsilon(k) > t_0 - t_1 \mid \mu(k), \sigma(k)) \\
&= \mathbb{P}\left(\frac{\epsilon(k) - \mu(k)}{\sigma(k)} > \frac{t_0 - t_1 - \mu(k)}{\sigma(k)}\right)
\end{aligned}
$$

$$= 1 - \Phi\left(\frac{t_0 - t_1 - \mu(k)}{\sigma(k)}\right),$$

where Φ is the cumulative distribution function of the standard normal distribution. Hence, the probability that a recession has started at time t if classifier k detected a recession at $d(k)$:

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

as long as classifier k remains in a recession state (figure 10).

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.

5.2. Recession probability from the classifier ensemble

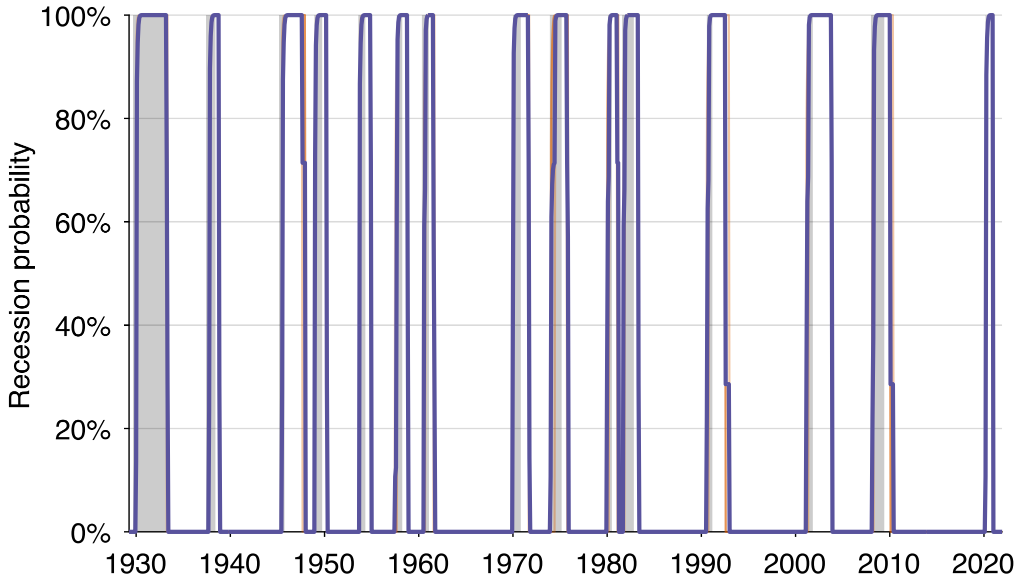
The final step is to average the probability that the recession has started across the $K = 7$ classifiers that I use for detection. Each classifier in the classifier ensemble generates a probability that the recession has started at time t, $P(k,t)$, given by (16). I compute the overall probability that the economy is in recession at time t by averaging the probabilities produced by the individual classifiers in the ensemble:

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

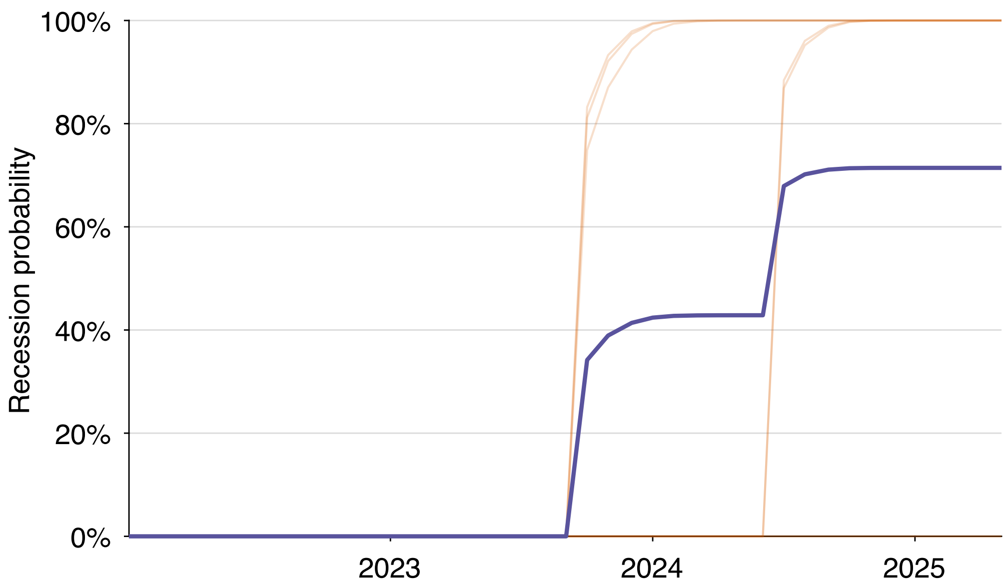
Each classifier provides a recession-start probability. The classifier ensemble averages these probabilities to yield a single risk score. The ensemble was built by collecting all classifiers with high accuracy. The reason was that it is unknown how the user of the detection algorithm trades off anticipation versus precision. 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.

Figure 11A shows the aggregate recession probability given by the classifier ensemble on the training period, 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 right after or somewhat after the recession has ended.[5]

[5]This algorithm does not aim to determine the end of recessions. But it would be possible to build another



A. Recession probability on training data, April 1929–December 2021



B. Recession probability on testing data, January 2022–May 2025

FIGURE 11. Recession probability from the classifier ensemble

The recession probability from the classifier ensemble is given by (17) (thick purple line). The probability is the average of the recession probabilities given by the individual classifiers in the ensemble, each 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, as illustrated by figure 8A. Shaded areas indicate recessions dated by the NBER (2023).

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`.

5.3. Application of the recession detection algorithm to current data

Finally, I apply the ensemble classifiers obtained from 1929–2021 data to current data so as to assess recession risk in real time (figure 11B). As of May 2025, the recession probability given by the classifier ensemble is 71%, suggesting a high likelihood that a recession has begun at this point in time. The underlying reason is that 5 of the 7 selected classifiers have been activated in the past 2 years (5/7 = 0.71). This is in turn due to the sharp decrease in the number of job vacancies and increase in the number of job seekers since the middle of 2022 (figure 1).

6. Addressing concerns about overfitting via placebo test

A challenge for the algorithm is the risk of overfitting: the algorithm uses a vast set of potential indicators to detect only 15 recessions. To address this concern, I conduct a placebo test. The principle 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 sample period. If the algorithm, which is based on labor market data, is robust and not

algorithm on the same principles to determine recession ends as early and accurately as possible.

overfitting, it should be unable to detect these economically unrelated events. My findings show that the algorithm indeed fails to detect the placebo events, which alleviates concerns about overfitting.

The challenge 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 US presidents. By happenstance, there have been exactly 15 such deaths between April 1929 and December 2021 (table 2). Obviously such deaths are not caused by the US economy, nor do they affect it in any noticeable way.

A first manifestation of the algorithm's inability to detect the deaths of first ladies appears in the anticipation-precision plane (figure 12A). To conduct the test, I reuse the set of 2,343,752 perfect classifiers from the recession analysis (figure 8A)—that is, all classifiers that generate exactly 15 signals between 1929 and 2021. Since the number of placebo events is also 15, this set of classifiers is, by definition, also the perfect set for the placebo test. I then evaluate their performance by plotting the mean and standard deviation of their detection errors, this time relative to the dates of the first ladies' deaths.

I assess the algorithm's performance based on how early and, more importantly, how precisely it can detect the first-lady deaths. The mean error is computed using (14) and the standard deviation of errors using (15), with start dates of recessions substituted by the death dates of first ladies. The results show a stark contrast with the recession detection task. For the placebo test, every classifier yields a standard deviation of detection errors exceeding 500 months (figure 12A). This is dramatically different from the recession analysis, where the best classifiers achieved a standard deviation below 2 months (figure 8A). A standard deviation of over 500 months—more than 41 years—indicates that the detection dates are effectively random with respect to the death dates, confirming that the algorithm cannot meaningfully detect the deaths of first ladies.

Overall, the anticipation-precision frontier comprises only 7 classifiers (figure 12A), with an average standard deviation along the frontier of 543 months. By contrast, the 7-classifier frontier obtained from recession detection on 1929–2021 data has an average standard deviation of 1.9 months (table 1).

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 12B). The overall probability is the average of the death probabilities given by the 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

TABLE 3. Performance of the algorithm on backtests

	2005 backtest		1985 backtest		1965 backtest	
	Ensemble of 12 classifiers		Ensemble of 6 classifiers		Ensemble of 8 classifiers	
Detection error (months)	Training	Testing	Training	Testing	Training	Testing
Mean	2.4	2.7	2.3	1.2	3.4	3.9
Standard deviation	1.9	1.9	2.1	1.4	1.5	2.6
Minimum	−0.1	0.8	−1.0	−0.3	1.4	0.8
Maximum	5.7	4.6	5.2	3.5	5.3	9.3
Recession count	13	2	11	4	7	8

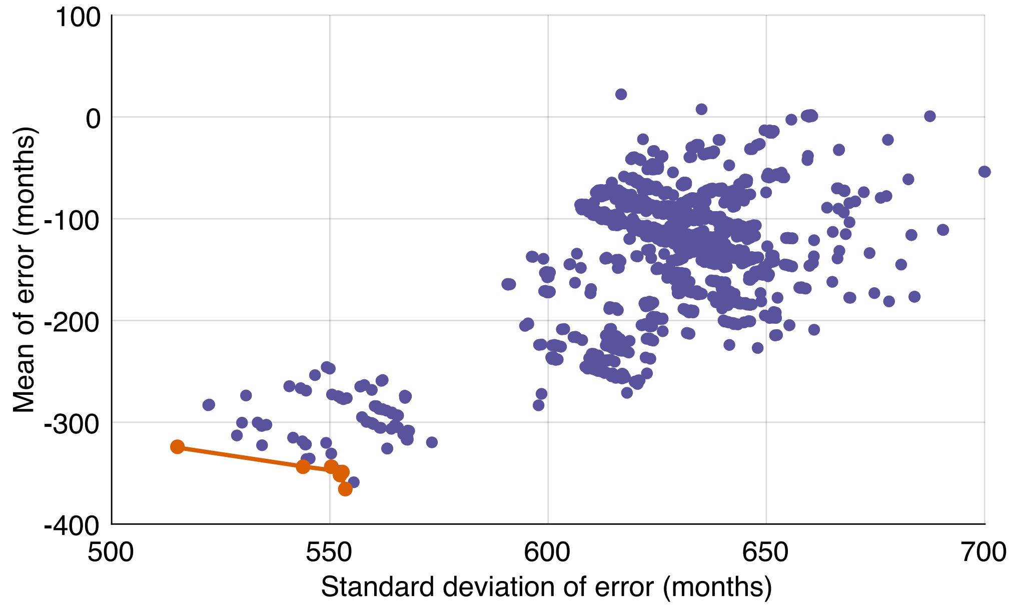
The table averages the mean, standard deviation, minimum, and maximum of the detection errors across all the classifiers in the selected classifier ensemble. In each backtest, the classifier ensemble comprises all the perfect classifiers that present a standard deviation of detection errors below 3 months on the training period. The training periods considered in the backtests are April 1929–December 2004, April 1929–December 1984, and April 1929–December 1964. The testing periods considered are January 2005–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).

7 deaths occurred. Another striking 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 11A).

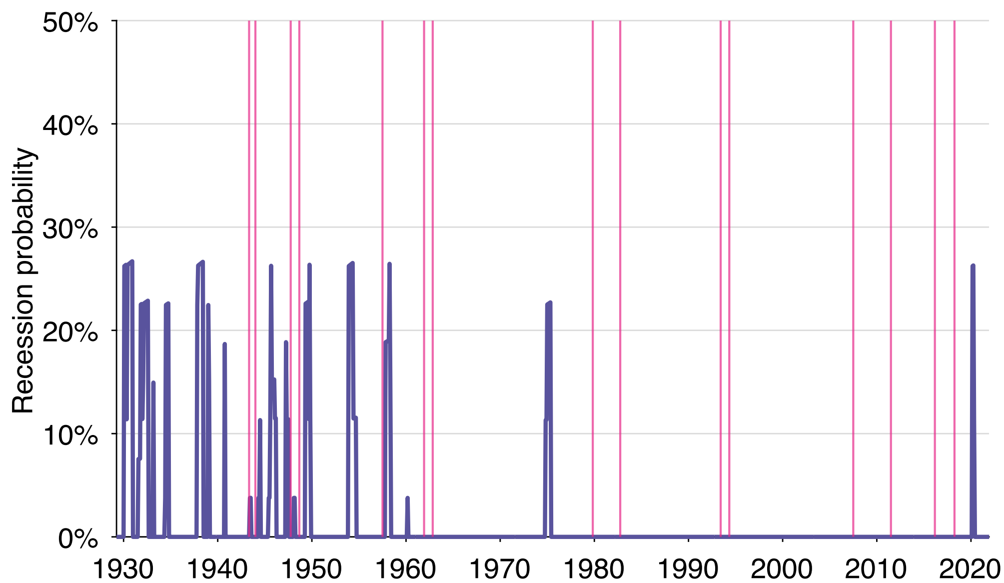
Overall, the huge standard deviation of detection errors and the randomly scattered death probability spikes indicate that the algorithm is unable to detect the placebo events. This inability implies that the algorithm does not overfit the data. The reason is that, although the algorithm builds tens of thousands of indicators, these indicators are all constructed from only two 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 streams 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. As a result, the algorithm cannot spuriously fit 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.



A. Anticipation-precision frontier of classifiers in placebo test



B. Death probability obtained from classifiers on the frontier

FIGURE 12. 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 same 2,343,752 perfect classifiers from figure 8A, which detect exactly 15 first-lady deaths between 1929 and 2021. The mean error 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 only 7 classifiers. Panel B gives the death probability from the 7 frontier classifiers, computed from (17). The probability is the average of the death probabilities given by the individual classifiers, each given by (16). Pink vertical lines are the death dates of the first ladies, as described in table 2.

Classifier ensembles produced from backtests all suggest a non-negligible probability that the US economy is in recession in 2025.

7.1. Backtesting from 2005

First, I train the system on data up to December 2004, which feature 13 recessions, and test the algorithm on January 2005–December 2021, during which 2 recessions occurred (figure 13). The training yields 12 classifiers on the anticipation-precision frontier with a precision below 3 months. I use these 12 classifiers to detect the Great Recession and pandemic recession. As an illustration, 2 of these classifiers are described in table 4. I pick classifiers that are particularly simple and representative of clusters in the ensemble. Interestingly, one of the classifiers in the ensemble is based on an unemployment indicator (classifier 1) while another one is based on a vacancy indicator (classifier 2). Both of these indicators are constructed from heavily smoothed unemployment and vacancy rates and by comparing the current data value to last month's value.

The performance of the algorithm in the backtest is remarkably good. All 12 classifiers detect the 2 out-of-sample recessions without false positives (figure 13B). By June 2008 the recession probability reaches 1—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 entirely 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).

The performance of the algorithm over the new training period, 1929–2004, is similar to that over the longer training period, 1929–2021 (table 3). In both cases the standard deviation of errors averages 1.9 months across classifiers. The mean error averages 2.4 months over 1929–2004 instead of 2.2 months over 1929–2021. The performance of the algorithm over the testing period, 2005–2021, is comparable to that over the training period, 1929–2004. In both cases the standard deviation of errors averages 1.9 months across classifiers. The mean error averages 2.7 months over the testing period, slightly longer than the 2.4 months over the training period.

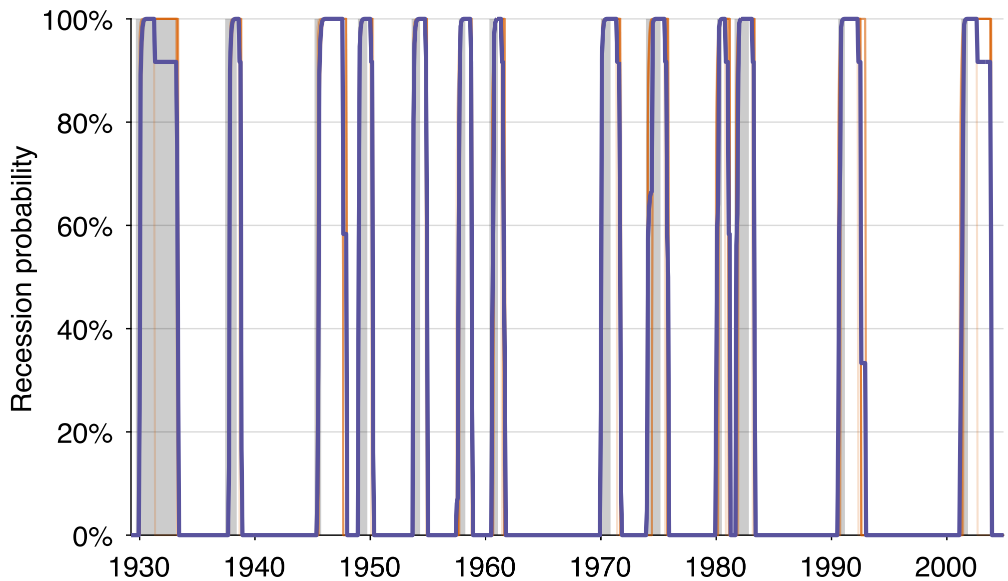
7.2. Backtesting from 1985

I then repeat backtesting by training the algorithm on data up to December 1984 and testing on subsequent data, from January 1985 to December 2021 (figure 14). The training period comprises 11 recessions, and the testing period features 4 recessions. The training yields 6 classifiers on the anticipation-precision frontier with a precision below 3 months. I then use these 6 classifiers to detect recessions that occurred after 1984. As an illustration,

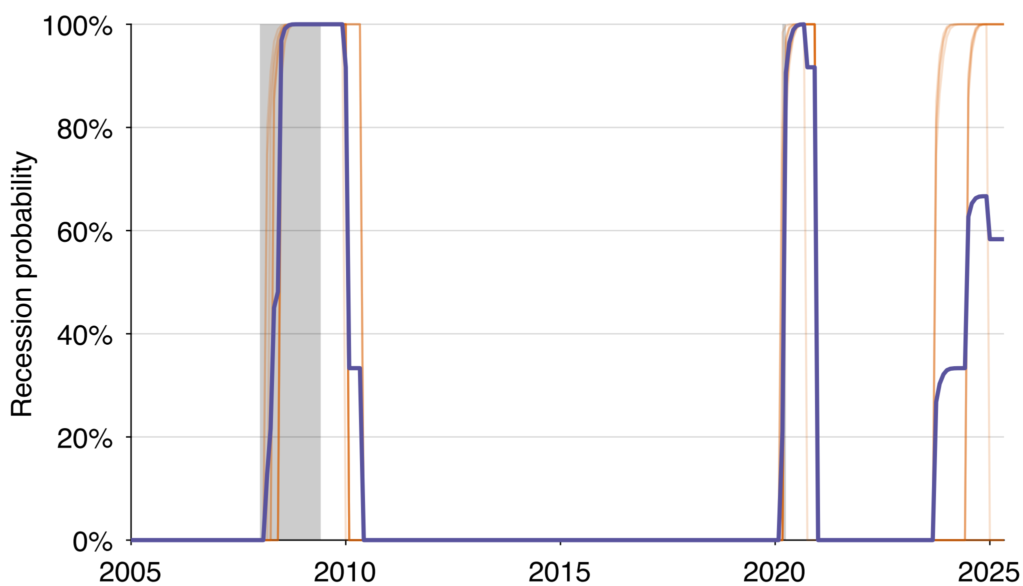
TABLE 4. Select classifiers from the classifier ensembles obtained in backtests

	2005 backtest		1985 backtest		1965 backtest	
	Ensemble of 12 classifiers		Ensemble of 6 classifiers		Ensemble of 8 classifiers	
Classifier	1	2	3	4	5	6
Smoothing method	simple	exponential	simple	simple	exponential	simple
Smoothing horizon α	11m	0.2	11m	6m	0.4	3m
Turning horizon β	1m	1m	1m	1m	9m	9m
Curving parameter γ	1	0	1	0	1	0.7
Combination method	u-v	u-v	u-v	u-v	min-max	min-max
Combination weight δ	1	0	1	0.4	1	1
Threshold ζ	0.16pp	2.36pp	0.09pp	2.18pp	0.19pp	0.84pp

The classifier ensembles comprise the classifiers on the anticipation-precision frontier with a standard deviation of errors below 3 months. The ensembles are built from training data. The training periods considered in the backtests are April 1929–December 2004, April 1929–December 1984, and April 1929–December 1964. The ensembles are then evaluated over testing data. The testing periods considered are January 2005–December 2021, January 1985–December 2021, and January 1965–December 2021. The simple smoothing method with horizon α is given by (1) and (2); the exponential smoothing method with horizon α is given by (3) and (4). The turning horizon β enters the construction of the classifiers through (5) and (7). The curving 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).



A. Training on 13 in-sample recessions between April 1929 and December 2004



B. Testing on 2 out-of-sample recessions between January 2005 and December 2021

FIGURE 13. Backtesting the algorithm from 2005

The recession probability from the classifier ensemble is given by (17) (thick purple line). The probability is the average of the recession probabilities given by the individual classifiers in the ensemble, each given by (16) (thin orange lines). The classifiers in the ensemble are selected from the high-precision segment of the 1929–2004 anticipation-precision frontier. Shaded areas indicate recessions dated by the NBER (2023).

2 of these classifiers are described in table 4. Interestingly, the unemployment indicator used to build classifier 1 in the previous backtest is also used to build classifier 3 in this backtest. So certain indicators remain useful across backtests.

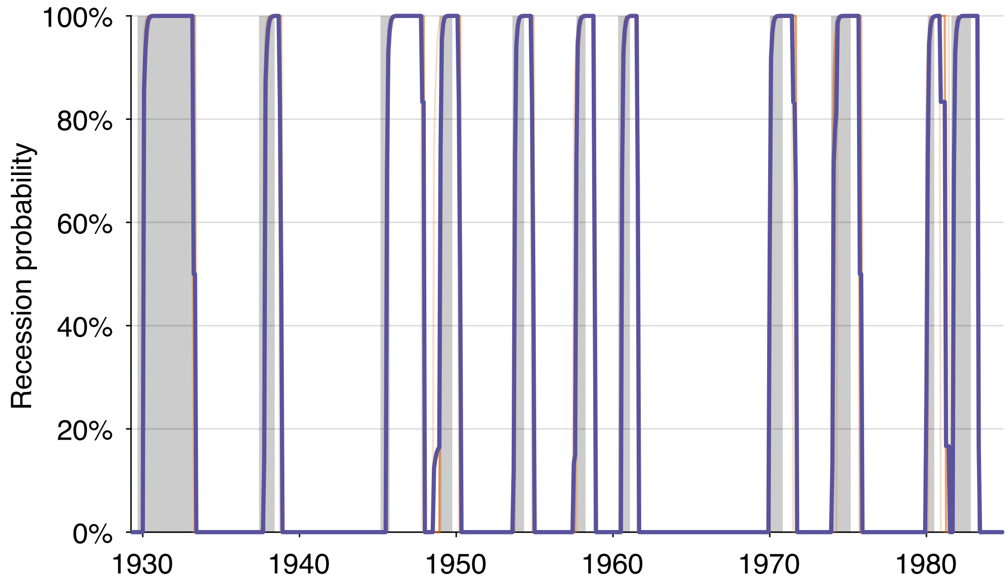
The performance is again remarkably good. All 6 classifiers detect the 4 out-of-sample recessions without false positives (figure 14B). Once again, the recession probability reached 1 by the summer of 2008. So an algorithm that had only been trained on data up to December 1984 proficiently detects the Great Recession. The recession probability also became positive in early 2001, before the dot-com recession had officially started.

The performance of the algorithm over the new training period, 1929–1984, is slightly worse than that over the longer training period, 1929–2021 (table 3). The standard deviation of errors averages 2.1 months across classifiers for the new training period, slightly higher than the 1.9 months over the 1929–2021 training period. The mean error averages 2.3 months over 1929–1984, slightly higher than 2.2 months over 1929–2021. The slightly worse performance reflects the fact that the data at the beginning of the period were noisier so less informative. The performance of the algorithm over the testing period, 1985–2021, is actually much better than that over the training period, 1929–1984. The standard deviation of errors averages only 1.4 months over the testing period, instead of 2.1 months over the training period. The mean error averages only 1.2 months over the testing period, much less than the 2.1 months over the training period. My interpretation is that unemployment and especially vacancy data are of higher quality at the end of the period so they better delineate the last 4 US recessions.

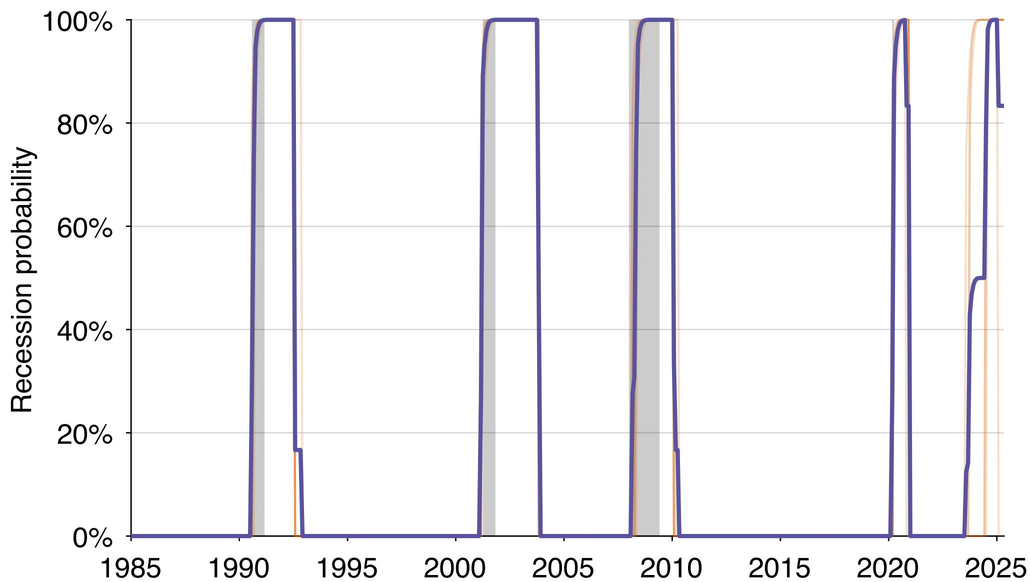
7.3. Backtesting from 1965

Last, I train the algorithm on data up to December 1964—which comprise 7 recessions—and test it on the subsequent 8 recessions (figure 15). The training yields 8 classifiers on the anticipation-precision frontier with a precision below 3 months. I use these 8 classifiers to detect the 8 recessions that occurred after 1964. As an illustration, 2 of these classifiers are described in table 4. Both classifiers are based on minimum indicators. In fact, the indicator used to construct classifier 6 shares several similarities with the Michez-rule indicator: it is a minimum indicator; it smooths the data with 3-month trailing averages; and it detects turning point based on a fairly long horizon (9 months here and 12 months in the Michez rule).

Given that the algorithm is trained on fewer recessions than it must detect, the performance remains strikingly good. All 8 classifiers detect the 8 out-of-sample recessions without false positives (figure 15B). Looking at the Great Recession, the recession probability reaches 1 by the summer of 2008. So an algorithm that had only seen data up to December 1964 detects the Great Recession in better time than some of the state-of-the-art algorithms of the day that had seen real-time data (Leamer 2008).



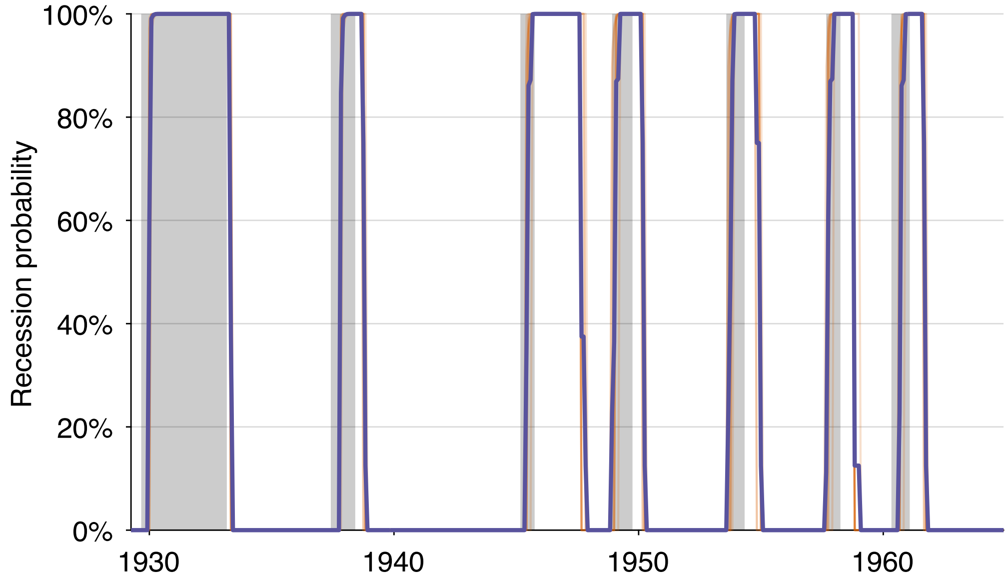
A. Training on 11 in-sample recessions between April 1929 and December 1984



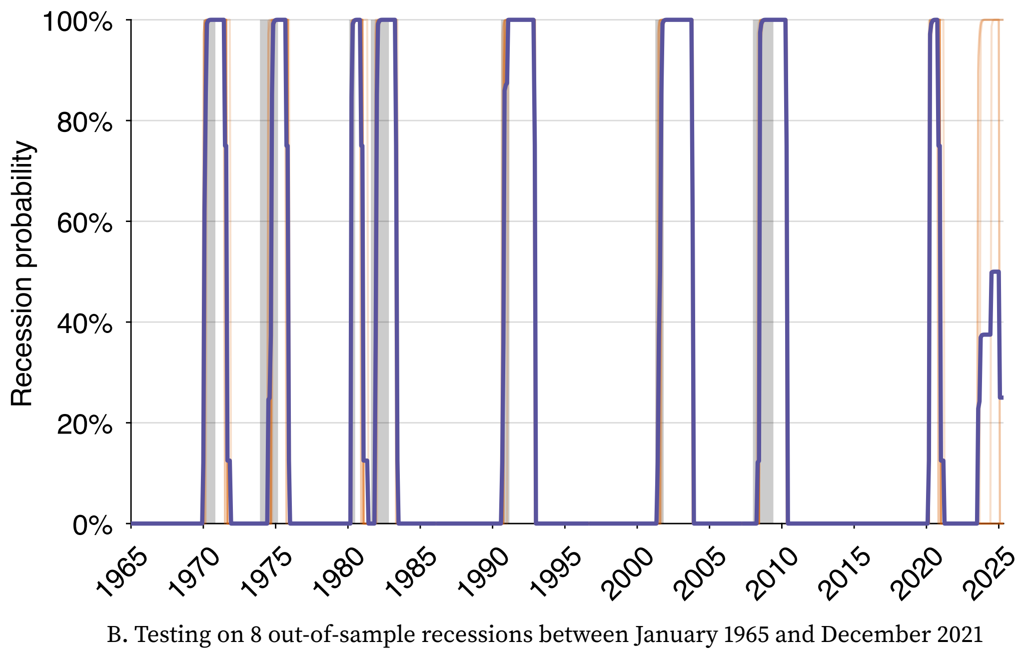
B. Testing on 4 out-of-sample recessions between January 1985 and December 2021

FIGURE 14. Backtesting the algorithm from 1985

The recession probability from the classifier ensemble is given by (17) (thick purple line). The probability is the average of the recession probabilities given by the individual classifiers in the ensemble, each given by (16) (thin orange lines). The classifiers in the ensemble are selected from the high-precision segment of the 1929–1984 anticipation-precision frontier. Shaded areas indicate recessions dated by the NBER (2023).



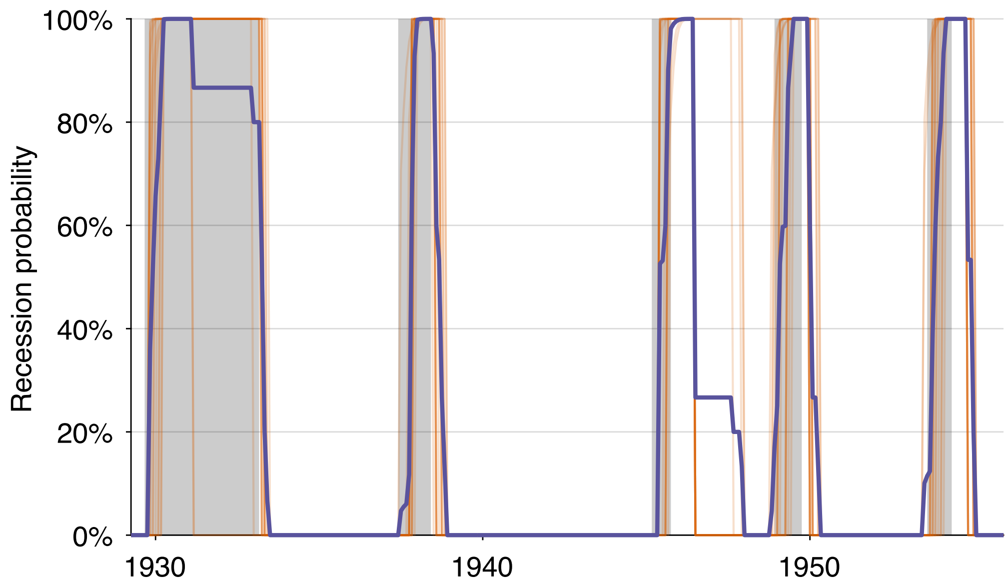
A. Training on 7 in-sample recessions between April 1929 and December 1964



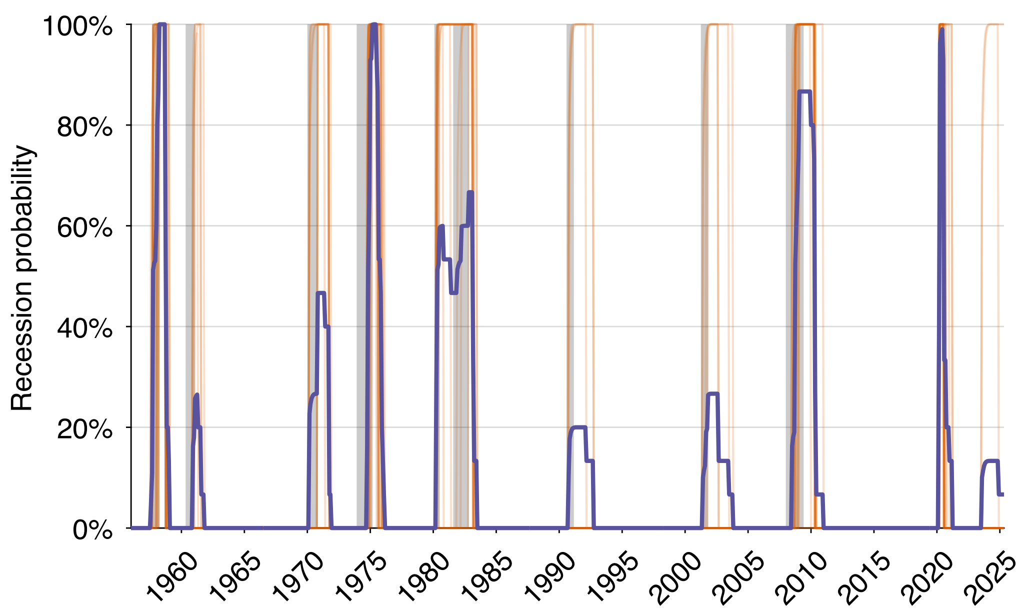
B. Testing on 8 out-of-sample recessions between January 1965 and December 2021

FIGURE 15. Backtesting the algorithm from 1965

The recession probability from the classifier ensemble is given by (17) (thick purple line). The probability is the average of the recession probabilities given by the individual classifiers in the ensemble, each given by (16) (thin orange lines). The classifiers in the ensemble are selected from the high-precision segment of the 1929–1964 anticipation-precision frontier. Shaded areas indicate recessions dated by the NBER (2023).



A. Training on 5 in-sample recessions between April 1929 and December 1955



B. Testing on 10 out-of-sample recessions between January 1956 and December 2021

FIGURE 16. Backtesting the algorithm from 1956

The recession probability from the classifier ensemble is given by (17) (thick purple line). The probability is the average of the recession probabilities given by the individual classifiers in the ensemble, each given by (16) (thin orange lines). The classifiers in the ensemble are selected from the high-precision segment of the 1929–1955 anticipation-precision frontier. Shaded areas indicate recessions dated by the NBER (2023).

The performance of the algorithm over the new training period, 1929–1964, is different but not strictly worse than that over the longer training period, 1929–2021 (table 3). The standard deviation of errors averages 1.5 months across classifiers for 1929–1964, which is less than the average of 1.9 months for 1929–2021. The mean error averages 3.4 months over 1929–1964, which is more than 2.2 months over 1929–2021. So the algorithm is more precise but slower over the 1929–1964 training period than over the 1929–2021 training period. The performance of the algorithm over the testing period, 1965–2021, is somewhat worse than that over the training period, 1929–1964. This shouldn't be unexpected given that the algorithm is trained on fewer recessions than are present in the testing period. The standard deviation of errors averages 2.6 months over the testing period, above the 1.5 months over the training period. The mean error averages 3.9 months over the testing period, above the 3.4 months over the training period.

7.4. Backtesting from 1956

How far can we go in backtesting and continue to obtain a perfect performance in testing—without false positives or false negatives from the classifiers selected in training? I find that the algorithm breaks down by going back a further decade in time. The performance of the algorithm trained until 1955 and tested over 1956–2021 is noticeably worse (figure 16). The algorithm only has 5 in-sample recessions to learn from. Through training, it selects 15 classifiers that perfectly detect these 5 recessions with a precision of less than 3 months. Of these 15 classifiers, however, only 2 perfectly detect the 10 subsequent out-of-sample recessions. The remainder miss at least 1 out-of-sample recession, with some missing as many as 7 recessions (figure 16B).

7.5. Application of the backtested classifier ensembles to current data

Finally, I use the classifier ensembles produced by the backtests to evaluate the current recession risk. The classifier ensemble created by training the algorithm on 1929–2021 data gives a 71% probability that the US economy is in recession in May 2025 (figure 11B).

The backtested classifier ensembles give different probabilities, but they all give positive recession probabilities in 2025—especially the more modern ensembles. The classifier ensemble created by training the algorithm on 1929–2004 data gives a 58% probability that the US economy is in recession in May 2025 (figure 13B). That probability peaked at 67% in 2024. The classifier ensemble created by training the algorithm on 1929–1984 data produces the highest recession probability. It gives an 83% probability that the US economy is in recession in May 2025 (figure 14B). That probability peaked at 100% in 2024. Finally, the classifier ensemble created by training the algorithm on 1929–1964 data gives a 25% probability that the US economy is in recession in May 2025 (figure 15B). That probability peaked at 50% in 2024.

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.[6] 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:

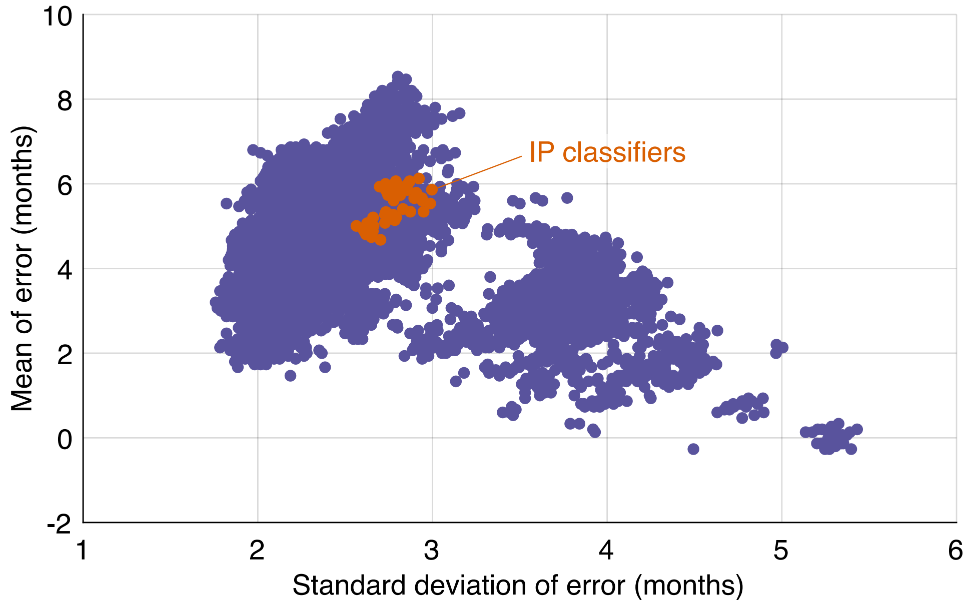
> The determination of the months of peaks and troughs is based on a range of monthly measures of aggregate real economic activity published by the federal statistical agencies. These include real personal income less transfers, nonfarm payroll employment, employment as measured by the household survey, real personal consumption expenditures, wholesale-retail sales adjusted for price 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" variable and "noisy", so 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 2008, 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:

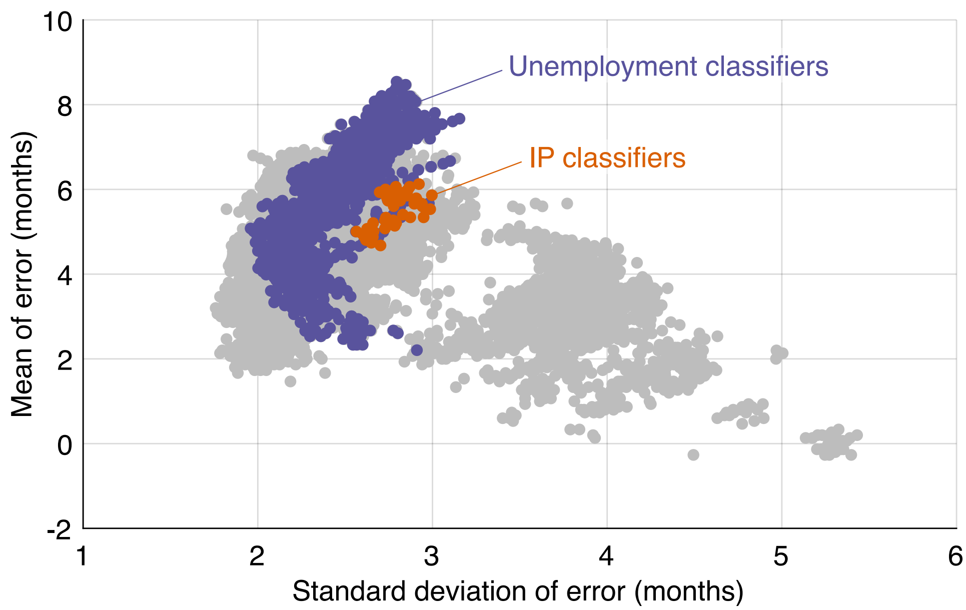
> The unemployment rate is a trendless indicator that moves in the opposite direction from most other cyclical indicators.... The NBER business-cycle chronology considers economic activity, which grows along an upward trend. As a result, the unemployment rate sometimes rises before the peak of economic 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

[6] Strangely, 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.



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



B. Performance of industrial-production classifiers versus unemployment classifiers

FIGURE 17. Comparing labor-market classifiers to product-market classifiers

The figure reproduces figure 8A. In addition, it displays 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).

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, both industrial production and the unemployment rate featured prominently in recession announcements (NBER 1979, 1980, 1982). A third advantage is that of all the main series examined by the Dating Committee, industrial production is probably the least connected to the unemployment and vacancy rates, which renders the exercise more interesting.

I filter the industrial production index just like the unemployment and vacancy rates and produce 3,146 indicators. I then find thresholds such that the resulting classifier perfectly detects the 15 recessions between April 1929 and December 2021. There are 385,627 such classifiers. I finally compute, for each perfect classifier, the mean and standard deviation of the detection error. A good classifier should be early and accurate, so it should have a low (or better, negative) mean error and a low standard deviation of 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 17). I first compare the industrial-production classifiers to all the classifiers constructed with unemployment and vacancy data (figure 17A). 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 2.6 months, while the most precise labor-market classifier has a standard deviation of detection error of 1.8 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 5.0 months while the most precise labor-market classifier detects recessions with an average delay of 3.2 months (table 1). So the most precise industrial-production classifier is about $\sqrt{0.8^2 + 1.8^2} = 2.0$ 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.

Of course, the previous task was quite challenging for industrial-production classifiers because they rely on one data source but competed against a larger number of classifiers constructed from two data sources (unemployment and vacancies). To make the task easier, I compare the industrial-production classifiers to classifiers constructed only with unemployment data. Even then, I find that industrial-production classifiers are less performant than unemployment classifiers (figure 17B). For a given precision, there are unemployment classifiers with much lower delay than the industrial-production classifiers. And for a given delay, there are unemployment classifiers with better precision than the industrial-production classifiers.

Given that industrial production was often used in tandem with the unemployment rate to determine recession starts in the early days of the NBER Dating Committee, I also construct classifiers by combining industrial production and unemployment rate, and compare their performance to that of classifiers constructed by combining unemployment and vacancy rates (figure 18). The unemployment-industrial production classifiers do clearly better than the industrial-production classifiers, so including the unemployment rate in the analysis improves the detection power of industrial production. This result suggests that the Dating Committee might benefit from incorporating the unemployment rate into their considerations. Even if the unemployment rate is "lagging" and "noisy," it significantly improves the performance of industrial production for recession detection.

Overall, however, classifiers based on labor market data continue to outperform those based on industrial production. There is no unemployment-industrial production classifier that enters the segment of the labor-market anticipation-precision frontier with a standard deviation of detection error below 6 months (figure 18A). In fact, the unemployment-industrial production anticipation-precision frontier remains above the unemployment-vacancy anticipation-precision frontier (figure 18B). The average mean error along the unemployment-industrial production frontier is 2.9 months; the average standard deviation of errors is 2.0 months. Both are higher than the corresponding statistics for the unemployment-vacancy anticipation-precision frontier: average mean error of 2.2 months and average standard deviation of errors of 1.9 months (table 1).

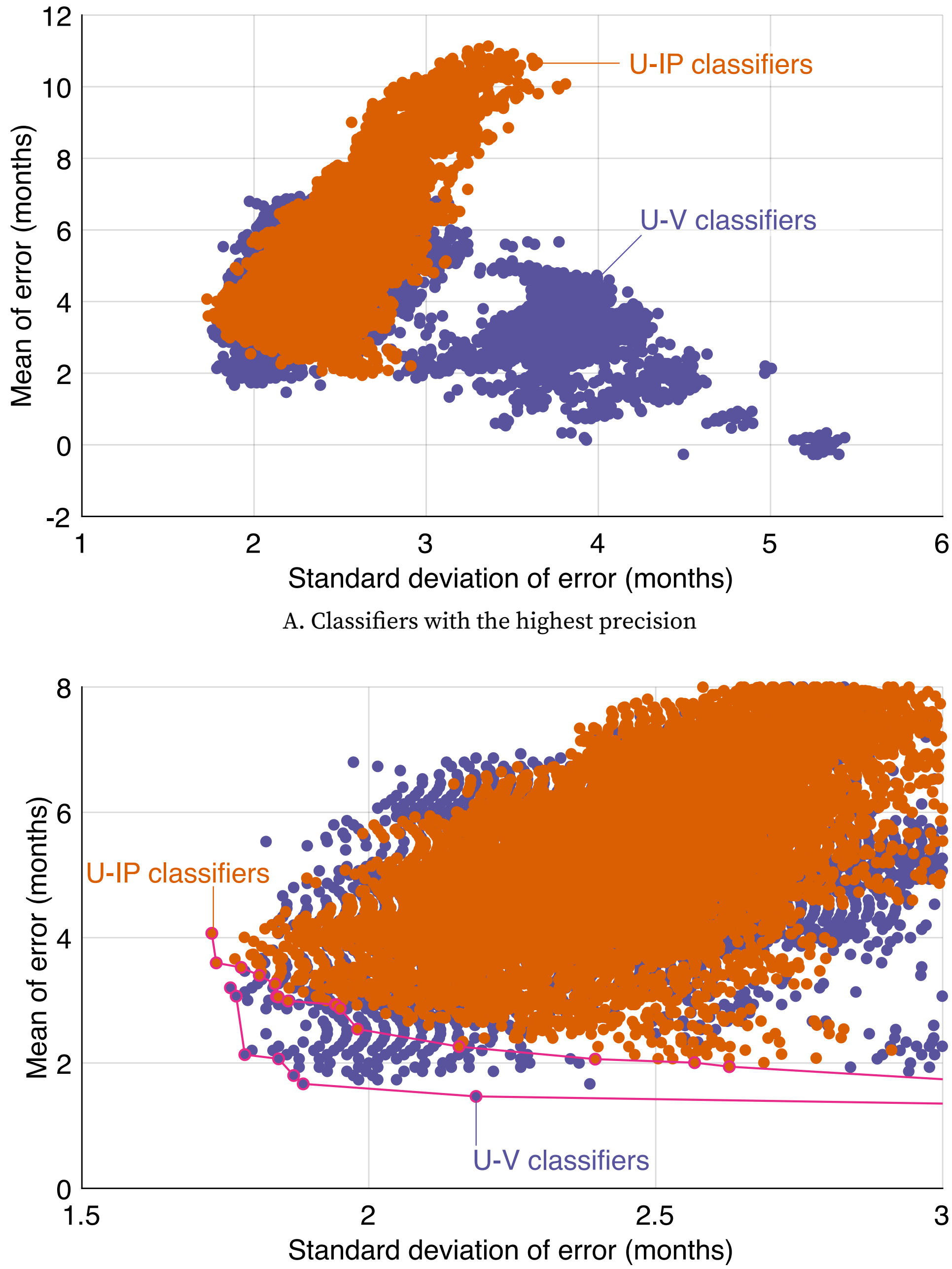
9. Conclusion

The paper develops a new algorithm for detecting US recessions in real time 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 systematically constructs millions of recession classifiers by combining unemployment and vacancy data. The classifiers are selected to avoid both false negatives (undetected true recessions) and false positives (falsely detected recessions). Then, by further selecting classifiers that lie on the anticipation-precision frontier, the algorithm optimizes jointly early detection and precision.

Using the ensemble of classifiers obtained from 1929–2021 data, I find that the probability that the United States is in recession in May 2025 is 71%. The classifier ensembles obtained through backtests on 1929–2004, 1929–1984, and 1929–1964 data also indicate an elevated recession risk: current recession probability of 58%, 83%, and 25%, 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



A. Classifiers with the highest precision

B. Anticipation-precision frontiers

FIGURE 18. Comparing unemployment-vacancy classifiers to unemployment-industrial production classifiers

The figure reproduces figure 8A. In addition, it displays the mean and standard deviation of the detection errors for classifiers built from unemployment and industrial production indicators (panel A). It also displays the anticipation-precision frontier of unemployment-vacancy classifiers and the anticipation-precision frontier of unemployment-industrial production classifiers (panel B).

the implication if, in retrospect, no recession occurred? In that case, the algorithm would need to be retrained on the longer 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 the current period would be eliminated. Given that many of the classifiers on the frontier do signal a recession, and that these would be eliminated, the anticipation-precision frontier would shift out after retraining. We would therefore learn that detecting recessions with labor market data is harder than the 1929–2021 data suggested. In other words, if a recession does not materialize, we would 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.

Bauer, Michael D. and Thomas M. Mertens. 2018. "Information in the Yield Curve about Future Recessions." FRBSF Economic Letter 2018-20. https://perma.cc/RP53-M5RQ.

Berge, Travis J. 2015. "Predicting Recessions with Leading Indicators: Model Averaging and Selection over the Business Cycle." *Journal of Forecasting* 34 (6): 455–471. https://doi.org/10.1002/for.2345.

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

BLS. 2020a. "Labor Force Statistics from the Current Population Survey: Overview." https://perma.cc/RN3P-S4SL.

BLS. 2020b. "Unemployment Level." FRED, Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/UNEMPLOY.

BLS. 2023. "Labor Force Statistics from the Current Population Survey: Concepts and Definitions." https://perma.cc/6LGU-AEU6.

BLS. 2024a. "Civilian Labor Force Level." FRED, Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/CLF16OV.

BLS. 2024b. "Job Openings and Labor Turnover Survey Overview Page." https://perma.cc/Y6EQ-WBXF.

BLS. 2024c. "Job Openings: Total Nonfarm." FRED, Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/JTSJOL.

BLS. 2024d. "JOLTS Frequently Asked Questions." https://perma.cc/22QW-ASAL.

BLS. 2024e. "Release Calendar." https://perma.cc/ZC78-SDPN.

BLS. 2025. "Technical Documentation (CPS)." https://perma.cc/DZ9K-AJP8.

Bry, Gerhard and Charlotte Boschan. 1971. *Cyclical Analysis of Time Series: Selected Procedures and Computer Programs*. Cambridge, MA: NBER. https://www.nber.org/booksand-chapters/cyclical-analysis-time-series-selected-procedures-and-computer-programs.

Chauvet, Marcelle. 1998. "An Econometric Characterization of Business Cycle Dynamics with Factor Structure and Regime Switching." *International Economic Review* 39 (4): 969–996. https://doi.org/10.2307/2527348.

Chauvet, Marcelle and James D. Hamilton. 2006. "Dating Business Cycle Turning Points." In *Nonlinear Time Series Analysis of Business Cycles,* edited by Costas Milas, Philip Rothman, and Dick van Dijk, chap. 1. Amsterdam: Elsevier. https://doi.org/10.1016/S0573-8555(05)76001-6.

Chauvet, Marcelle and Jeremy Piger. 2008. "A Comparison of the Real-Time Performance of Business Cycle Dating Methods." *Journal of Business & Economic Statistics* 26 (1): 42–49. https://doi.org/10.1198/073500107000000296.

Chauvet, Marcelle and Simon Potter. 2005. "Forecasting Recessions Using the Yield Curve." *Journal of Forecasting* 24 (2): 77–103. https://doi.org/10.1002/for.932.

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.

Christiansen, Charlotte, Jonas Nygaard Eriksen, and Stig Vinther Moller. 2014. "Forecasting US Recessions: The Role of Sentiment." *Journal of Banking & Finance* 49: 459–468. https://doi.org/10.1016/j.jbankfin.2014.06.017.

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.

Davig, Troy and Aaron Smalter Hall. 2019. "Recession Forecasting Using Bayesian Classification." *International Journal of Forecasting* 35 (3): 848–867. https://doi.org/10.1016/j.ijforecast.2018.08.005.

Diercks, Anthony M., Daniel Soques, and Jing Cynthia Wu. 2024. "Perfect Recession Predictors." https://ssrn.com/abstract=5022493.

Dueker, Michael. 2005. "Dynamic Forecasts of Qualitative Variables: A Qual VAR Model of US Recessions." *Journal of Business & Economic Statistics* 23 (1): 96–104. https://doi.org/10.1198/073500104000000613.

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

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.

Furno, Francesco and Domenico Giannone. 2024. "Nowcasting Recession Risk." In *Research Methods and Applications on Macroeconomic Forecasting,* edited by Mike Clements and Ana

Galvao. Cheltenham: Edward Elgar Publishing.
Galvao, Ana Beatriz and Michael T. Owyang. 2022. "Forecasting Low Frequency Macroeconomic Events with High Frequency Data." *Journal of Applied Econometrics* 37 (7): 1314–1333. https://doi.org/10.1002/jae.2931.
Giusto, Andrea and Jeremy Piger. 2017. "Identifying Business Cycle Turning Points in Real Time with Vector Quantization." *International Journal of Forecasting* 33 (1): 174–184. https://doi.org/10.1016/j.ijforecast.2016.04.006.
Hamilton, James D. 1989. "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle." *Econometrica* 57 (2): 357–384. https://doi.org/10.2307/1912559.
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.
Harding, Don and Adrian Pagan. 2002. "Dissecting the Cycle: A Methodological Investigation." *Journal of Monetary Economics* 49 (2): 365–381. https://doi.org/10.1016/S0304-3932(01)00108-8.
Harding, Don and Adrian Pagan. 2006. "Synchronization of Cycles." *Journal of Econometrics* 132: 59–79. https://doi.org/10.1016/j.jeconom.2005.01.023.
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.
Huang, Yu-Fan and Richard Startz. 2020. "Improved Recession Dating Using Stock Market Volatility." *International Journal of Forecasting* 36 (2): 507–514. https://doi.org/10.1016/j.ijforecast.2019.07.004.
Kauppi, Heikki and Pentti Saikkonen. 2008. "Predicting US Recessions with Dynamic Binary Response Models." *Review of Economics and Statistics* 90 (4): 777–791. https://doi.org/10.1162/rest.90.4.777.
Keil, Manfred, Edward Leamer, and Yao Li. 2023. "An Investigation into the Probability That This Is the Last Year of the Economic Expansion." *Journal of Forecasting* 42 (5): 1228–1244. https://doi.org/10.1002/for.2939.
Leamer, Edward E. 2008. "What's a Recession, Anyway?" NBER Working Paper 14221. https://doi.org/10.3386/w14221.
Leamer, Edward E. 2024. "Data Patterns That Reliably Precede US Recessions." *Journal of Forecasting* 43 (7): 2522–2539. https://doi.org/10.1002/for.3140.
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.
Mertens, Thomas M. 2022. "Recession Prediction on the Clock." FRBSF Economic Letter 2022-36. https://perma.cc/3QES-APLB.
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. 2024. "$u^* = \sqrt{uv}$: The Full-Employment Rate of Unem-

ployment in the United States." *Brookings Papers on Economic Activity* 55 (2).

Michaillat, Pascal and Emmanuel Saez. 2025. "Has the Recession Started?" *Oxford Bulletin of Economics and Statistics*. 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. 2008. "Business Cycle Dating Committee Announcement December 1, 2008." https://perma.cc/488B-ZCGY.

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/87A8-J3DS.

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.

Nyberg, Henri. 2010. "Dynamic Probit Models and Financial Variables in Recession Forecasting." *Journal of Forecasting* 29 (1–2): 215–230. https://doi.org/10.1002/for.1161.

O'Trakoun, John and Adam Scavette. 2025. "A Better Sahm Rule? Introducing the SOS Recession Indicator." *Economic Letters* 247: 1121175. 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.

Philips, Thomas K. and Tianqi Wang. 2025. "Two Simple Real-Time Algorithms to Identify Turning Points in US Business Cycles." https://ssrn.com/abstract=5243579.

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.

Qi, Min. 2001. "Predicting US Recessions with Leading Indicators via Neural Network Models." *International Journal of Forecasting* 17 (3): 383–401. https://doi.org/10.1016/S0169-2070(01)00092-9.

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.

Startz, Richard. 2008. "Binomial Autoregressive Moving Average Models With an Application to US Recessions." *Journal of Business & Economic Statistics* 26 (1): 1–8. https://doi.org/10.1198/073500107000000151.

Stock, James H. and Mark W. Watson. 1989. "New Indexes of Coincident and Leading Economic Indicators." *NBER Macroeconomics Annual* 4: 351–394. https://doi.org/10.1086/654119.

Stock, James H. and Mark W. Watson. 1993. "A Procedure for Predicting Recessions with Leading Indicators: Econometric Issues and Recent Experience." In *Business Cycles, Indicators and Forecasting*, edited by James H. Stock and Mark W. Watson, chap. 2. Chicago: University of Chicago Press. http://www.nber.org/chapters/c7190.

Stock, James H. and Mark W. Watson. 2010. "Indicators for Dating Business Cycles: Cross-History Selection and Comparisons." *American Economic Review* 100 (2): 16–19. https://doi.org/10.1257/aer.100.2.16.

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.

Vrontos, Spyridon D., John Galakis, and Ioannis D. Vrontos. 2021. "Modeling and Predicting US Recessions Using Machine Learning Techniques." *International Journal of Forecasting* 37 (2): 647–671. https://doi.org/10.1016/j.ijforecast.2020.08.005.

Weir, David R. 1992. "A Century of US Unemployment : 1890–1990. Revised Estimates and Evidence for Stabilization." *Research in Economic History* 14: 301–346.

Wright, Jonathan H. 2006. "The Yield Curve and Predicting Recessions." FEDS Working Paper 2006-07. https://ssrn.com/abstract=899538.

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.