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

Pascal Michaillat

June 2025

This paper develops a new method for detecting US recessions in real time. The method 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 training period without any false positives. By further selecting classifiers that lie on the high-precision segment of the anticipation-precision frontier, the method 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 1.9 months. Applied to May 2025 data, the classifier ensemble gives a 71% probability that the US economy is currently in recession. Backtesting to 2004, 1984, and 1964 confirms the algorithm's reliability. Algorithms trained on limited historical windows continue to detect all subsequent recessions without errors. Furthermore, they all detect the Great Recession by mid-2008—even when they are only trained on data up to 1984 or 1964. The classifier ensembles trained on 1929–2004, 1929–1984, and 1929–1964 data give a current recession probability of 58%, 83%, and 25%, respectively.

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

[1] There also exist numerous 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).

occurred since 1929 without false positives, whereas the Sahm rule breaks down 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, instead of extracting a better signal by looking at 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 times 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.

First, I construct millions of recession indicators from the unemployment and vacancy rates. The first step is to smooth the data using simple or exponentially weighted moving averages. The second step is to identify turning points by comparing smoothed data to their extrema over a defined period, yielding measures of increase for unemployment and decrease for vacancies. The increases and decreases are also scaled to cover all possibilities from absolute changes to relative changes. The individual unemployment and vacancy indicators are finally combined into single recession indicators through weighted averages or weighted min-max.

Having constructed the indicators, I build recession classifiers by applying a threshold. 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. Maybe surprisingly, 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, therefore, 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. They are unsuitable here given the presence of many perfect classifiers. Instead, I plot each classifier's mean detection delay against the standard deviation of detection delay in months. I then select classifiers on the low-delay, low-standard-deviation frontier—the anticipation-precision frontier. For a given mean detection delay, 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 lower 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 error 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-month wide or less. In all, I select 7 classifiers on the frontier.

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 delay. 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 delay. 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 delay'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.

Last, I backtest the detection method by shortening the training period and evaluating performance on subsequent recessions. For example, I train the system on data up to 2004 (13 recessions) and test its performance on later recessions. I repeat backtesting on data up to 1984 (11 recessions) to detect subsequent recessions, and on data up to 1964 (7 recessions) to detect more subsequent recessions. Backtesting confirms the robustness and reliability of the detection classifiers. Even 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 recession from the testing periods 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 breaks down only once I restrict training to 1929–1955 (5 recessions). Then, some of the classifiers in that ensemble do miss some of the recessions in the testing period, 1956–2021.

## 2. Data

In this section I present the data on US recessions, unemployment, and job vacancies that I analyze 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. Following the NBER's convention, I set the first month of a recession as the month following the peak and the last month of a recession as 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 before 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.

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

### 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.

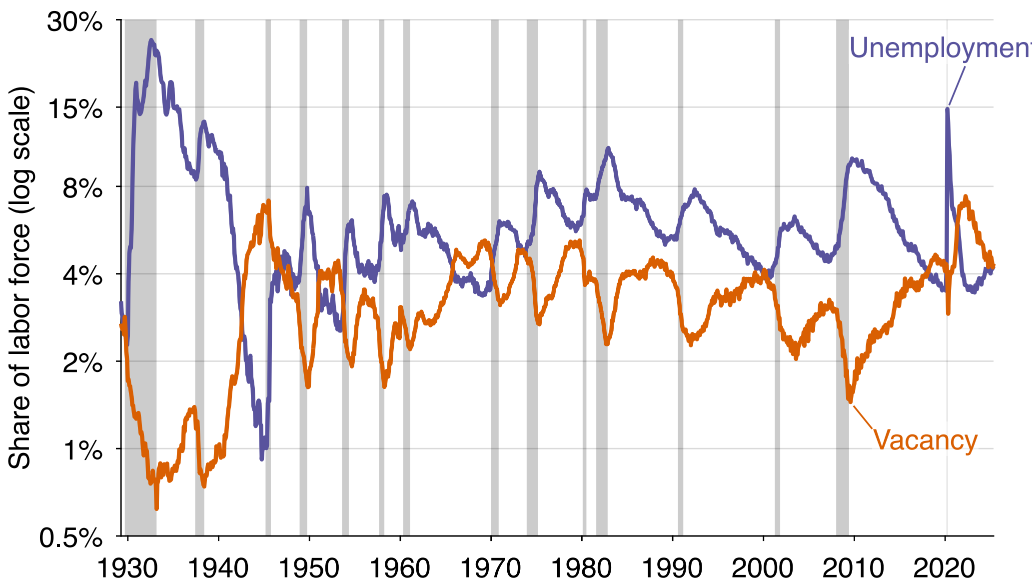



FIGURE 1. Monthly US unemployment and vacancy rates, April 1929–May 2025

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).

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.

### 2.3. Vacancy rate

Between April 1929 and December 1950, the vacancy rate is based on 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]

[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).

Between January 1951 and December 1959, 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 Survey (JOLTS), divided by the civilian 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). So 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).

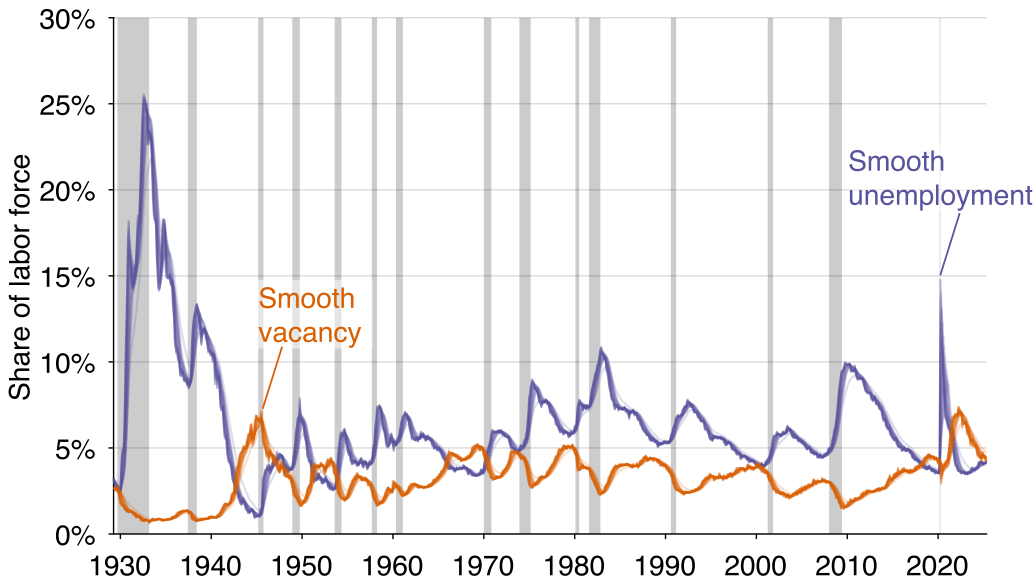



FIGURE 2. Smoothed US unemployment and vacancy rates, April 1929–May 2025

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. 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.

### 3.1. Smoothing the data

I smooth data by moving average. I use a simple trailing average. In that case the smoothed unemployment rate is given by

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

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

The vacancy rate is smoothed in the same way, similarly producing 12 series:

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

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. Here I compute all 10 series for $\alpha = 0.1, 0.2, 0.3, \ldots, 1$. $\alpha = 1$ corresponds to no smoothing.

The vacancy rate is smoothed in the same way, similarly producing 10 series.

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

### 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: $\beta = 1, 3, 6, 9, 12, \ldots, 36$ months. So the unemployment is compared to values ranging from last month to the minimum over the past 3 years:

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

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

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

The two steps of the computation are illustrated in figure 3.

I proceed analogously to determine turning points in the vacancy rate. I first take the maximum of the vacancy rate at various monthly horizons: $\beta = 1, 3, 6, 9, 12, \ldots, 36$ months. So the vacancy is compared to values ranging from last month to the maximum over the past 3 years:

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

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 two steps of the computation are illustrated in figure 4.

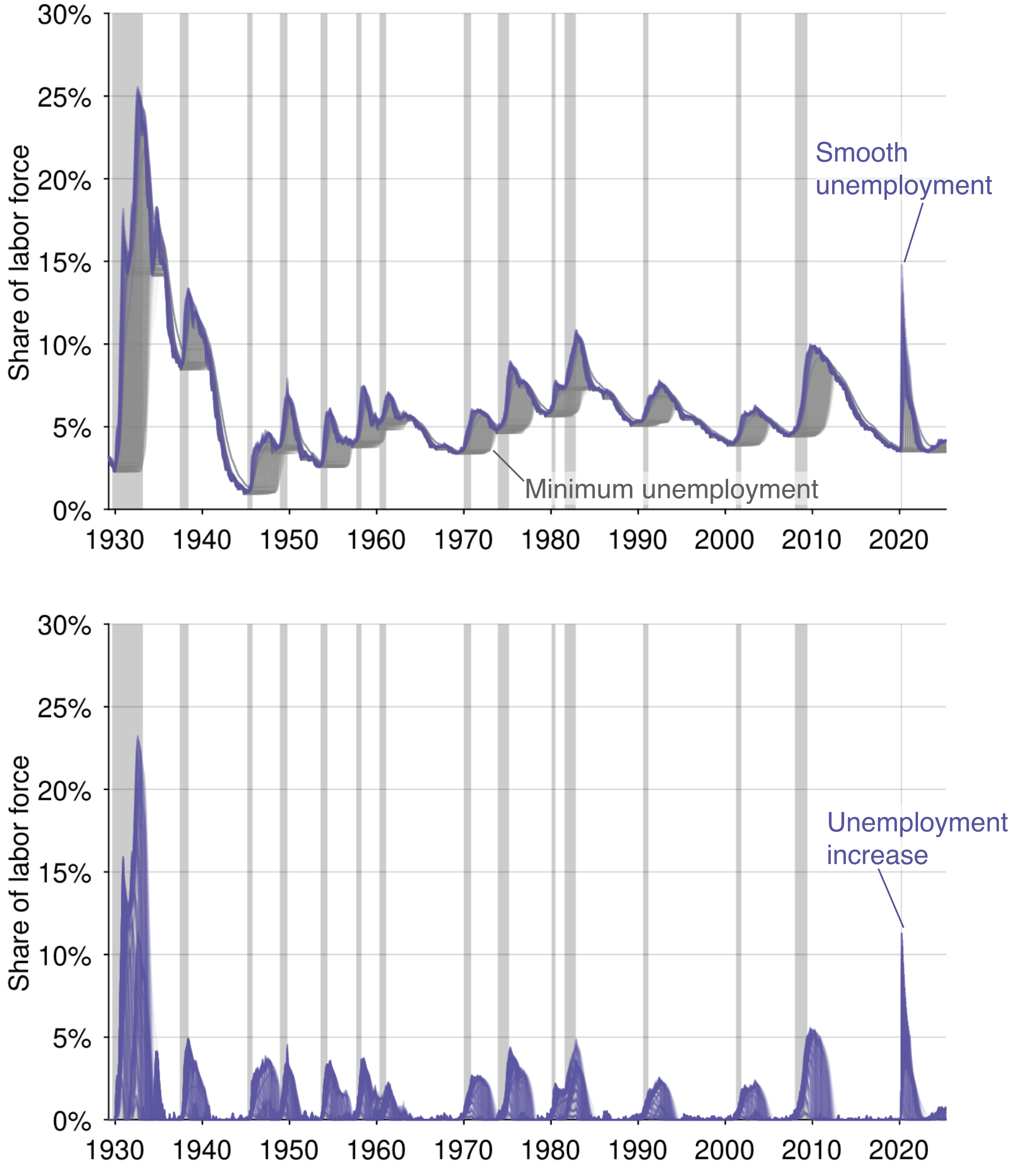



FIGURE 3. Increases in US unemployment rate, April 1929–May 2025

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 (6). Shaded areas indicate recessions dated by the NBER (2023).

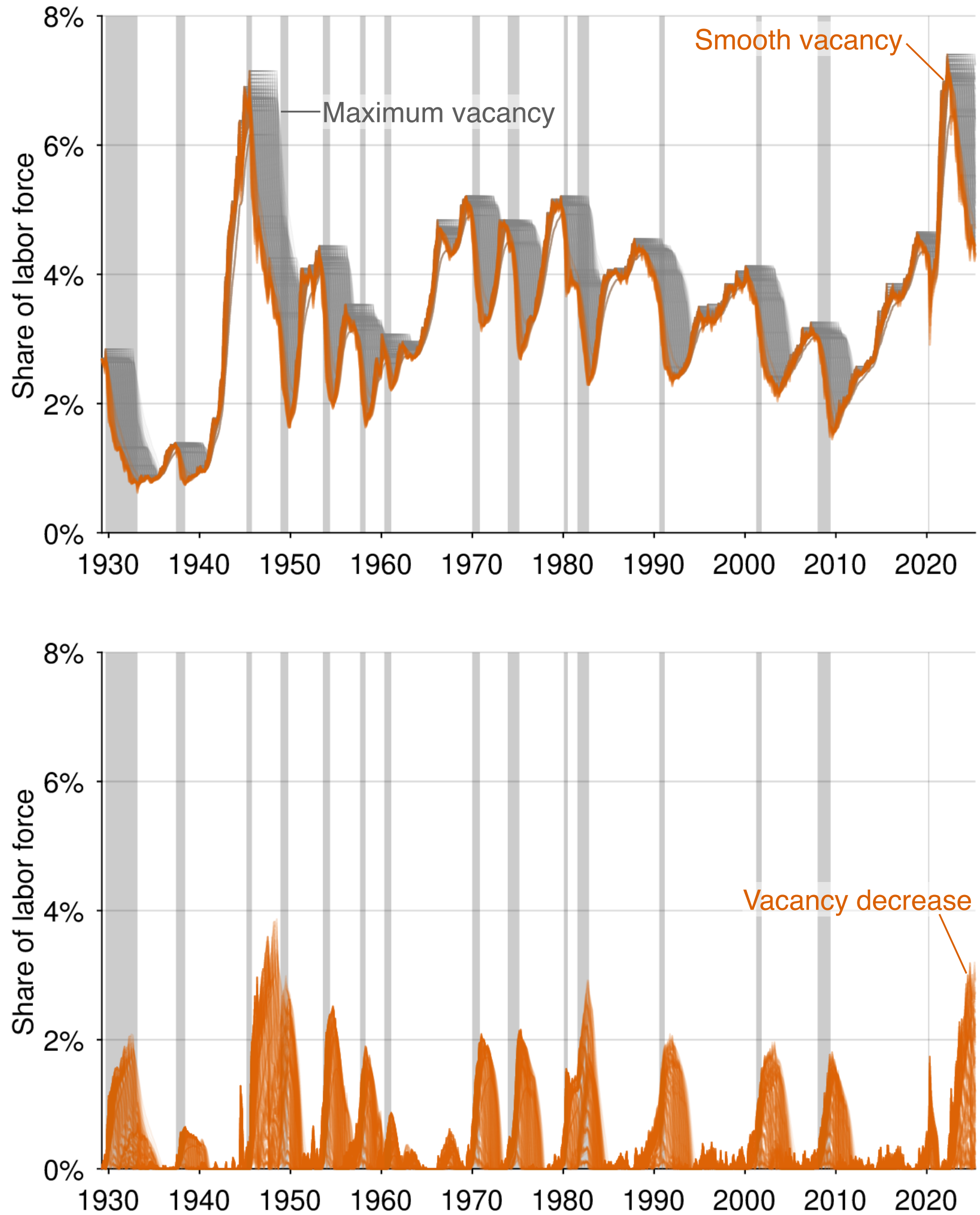



FIGURE 4. Decreases in US vacancy rate, April 1929–May 2025

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 (8). 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 changes in the rates of unemployment and job vacancy. It is not entirely clear, however, if what matters are level changes, such as $\tilde{u}(t)$ and $\tilde{v}(t)$, or relative 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 I consider level changes in unemployment and vacancy rates, as well as relative 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 $\gamma = 0, 0.1, 0.2, 0.3, \dots, 1$. The case $\gamma = 0$ reduces to percentage changes:

$$\hat{u}(t) = \log\left(\frac{\bar{u}(t)}{u^{\min}(t)}\right) \approx \frac{\tilde{u}(t)}{u^{\min}(t)}. \tag{10}$$

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

$$\hat{u}(t) = \tilde{u}(t). \tag{11}$$

Similarly, I consider all the vacancy indicators of the form

$$\hat{v}(t) = \frac{[v^{\max}(t)]^{\gamma} - [\bar{v}(t)]^{\gamma}}{\gamma} \approx \frac{\tilde{v}(t)}{v^{\max}(t)^{1-\gamma}} \tag{12}$$

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

### 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:

$$i(t) = \delta\hat{u}(t) + (1-\delta)\hat{v}(t), \tag{13}$$

where $\delta = 0, 0.1, 0.2, 0.3 \dots, 1$ is the weight on the unemployment indicator.

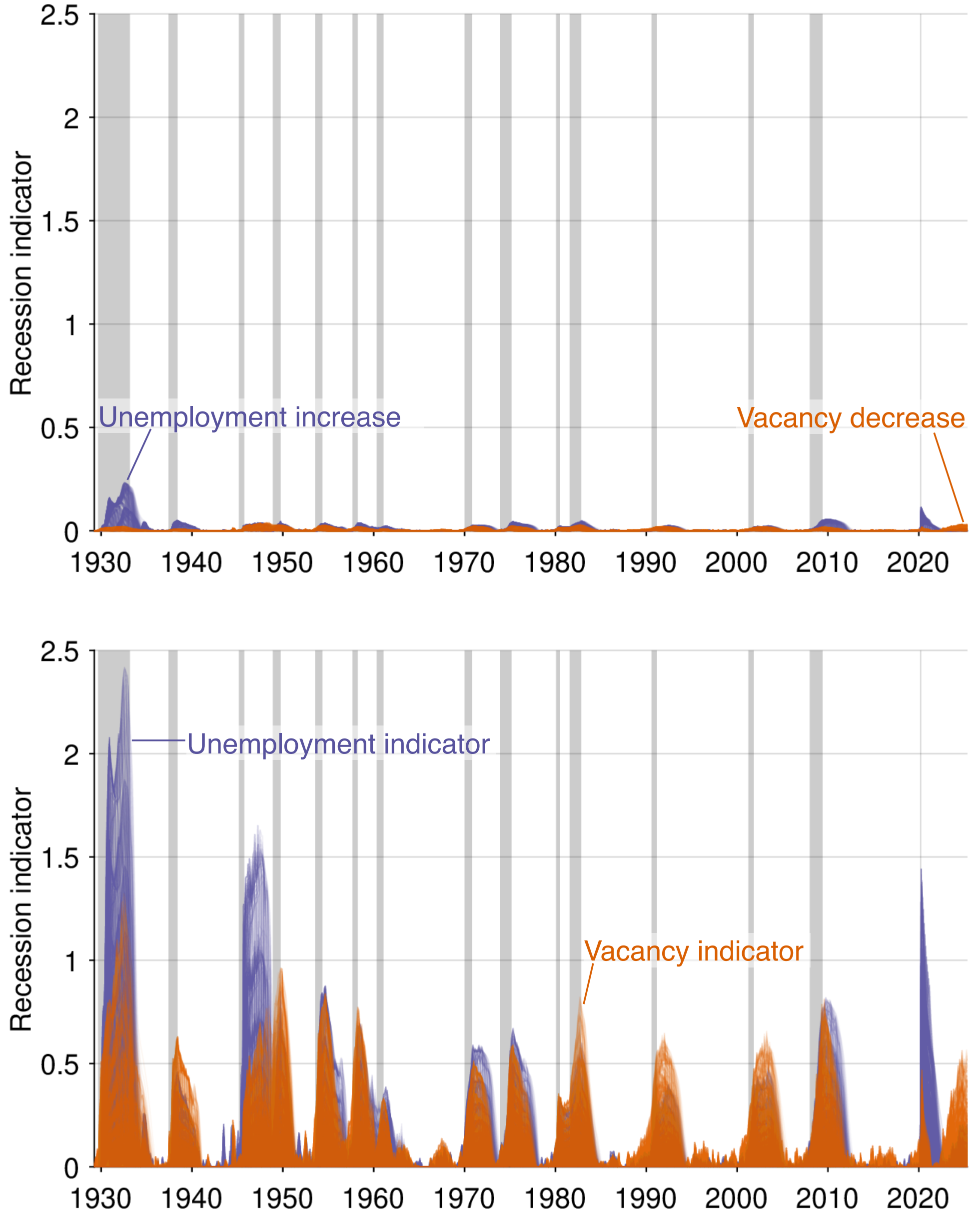



FIGURE 5. US unemployment and vacancy indicators, April 1929–May 2025

The unemployment increases and vacancy decreases come from figures 3 and 4. The unemployment indicators are computed from the unemployment increases and equation (9). The vacancy indicators are computed from the vacancy decreases and equation (12). Shaded areas indicate recessions dated by the NBER (2023).

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:

$$i(t) = \delta \min(\hat{u}(t), \hat{v}(t)) + (1 - \delta) \max(\hat{u}(t), \hat{v}(t)), \tag{14}$$

where $\delta = 0, 0.1, 0.2, 0.3 \dots, 1$ is the weight on the minimum indicator.

## 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 or recession that is updated dynamically as new recessions are triggered or as indicators signal a return to expansion—defined by the indicator value returning to zero.

Formally, a classifier $k$ is made up of an indicator $i(t) \geq 0$ and threshold $\tau > 0$. 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 ($i(0) < \tau$). The economy is in recession ($r(k, 0) = 1$) if the indicator is above the threshold ($i(0) \geq \tau$). In period $t$, we need to consider the state of the economy in period $t - 1$:

- If the economy was previously in recession ($r(k, t - 1) = 1$):
  - The economy remains in recession as long as the indicator remains positive: $r(k, t) = 1$ if $i(t) > 0$.
  - The economy enters an expansion if the indicator falls to zero: $r(k, t) = 0$ if $i(t) = 0$.
- If the economy was previously in expansion ($r(k, t - 1) = 0$):
  - The economy remains in expansion as long as the indicator remains below the threshold: $r(k, t) = 0$ if $i(t) < \tau$.

– The economy enters a recession if the indicator crosses the threshold: $r(k,t)=1$ if $i(t) \geq \tau$.

- If the classifier moves from expansion to recession in period $t$ for the $j$th time, we set the detection date of recession $j$ 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.

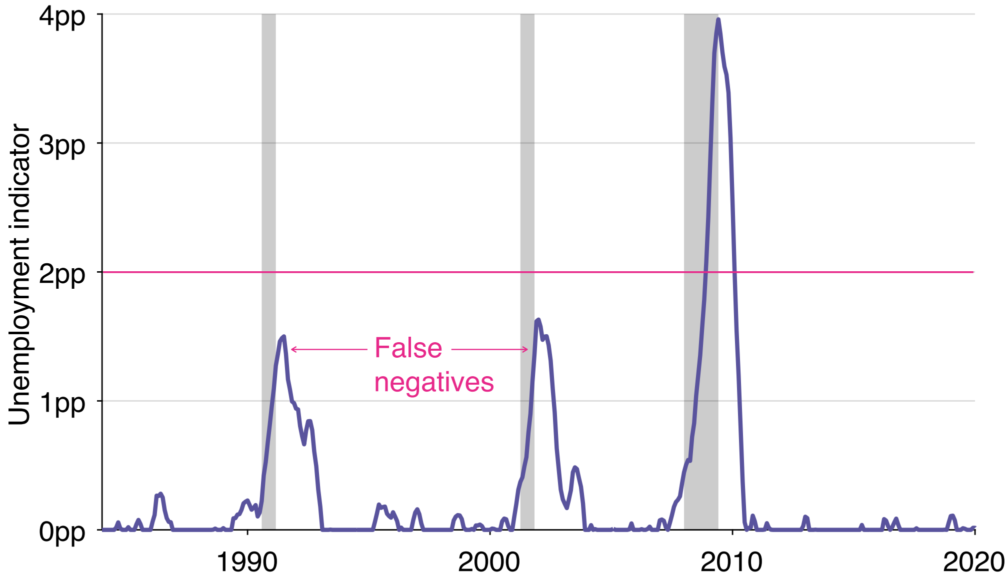
To eliminate such false positives, created by changes in the pace of economy slowdown around the threshold, I require that the economy is in expansion to be able to enter a new recession, and I require that the indicator falls to zero for the economy to be classified as in expansion. This classification procedure adds a bit of computational complexity, but it provides a more logical classification, which produces less noisy classifiers.

### 4.2. Selecting perfect classifiers

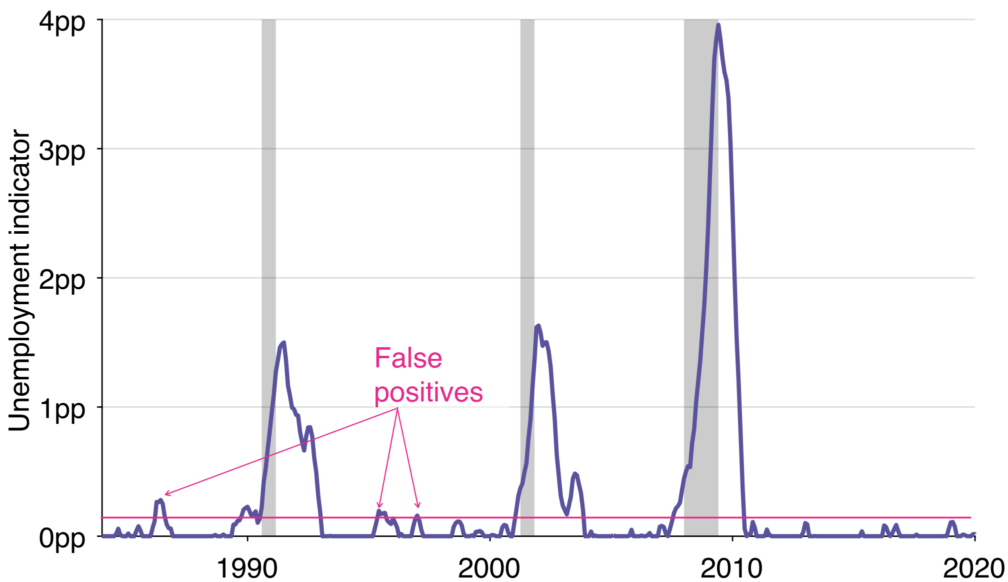
For all indicators and possible thresholds between 1 basis point and 1, $\zeta = 0.0001, 0.0002, 0.0003, \dots, 1$, I compute the number of recessions detected by each individual classifier. I evaluate each classifier between April 1929 and December 2021. The evaluation starts in April 1929 because this 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 throw away any classifier that makes any error: either false positives (non-recessions mistakenly detected by the classifier) or false negatives (actual recessions missed by the classifier). Figure 6 illustrates possible classification errors when the threshold is too high or too low.

I only keep perfect classifiers. Maybe surprisingly, there are many, many 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.




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).

### 4.3. Evaluating perfect classifiers

The challenge, therefore, 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 (Murphy 2022, section 5.1.3). They are unsuitable here given the presence of so many perfect classifiers.

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

$$\delta(k, j) = d(k, j) - s(j). \tag{15}$$

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

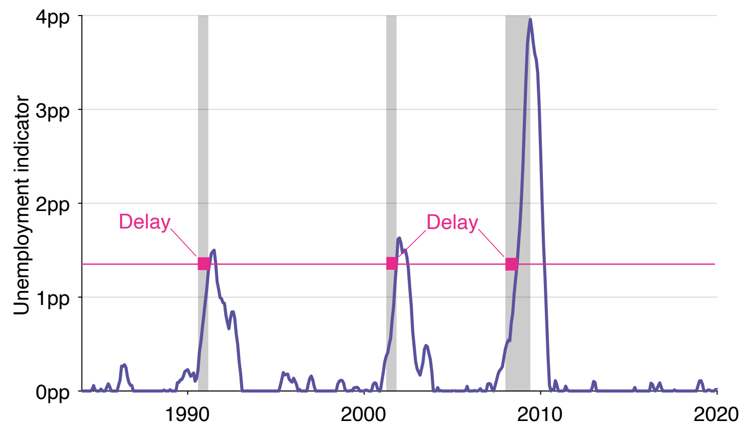
$$\mu(k) = \frac{1}{J} \cdot \sum_{j=1}^{J} \delta(k, j), \tag{16}$$

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

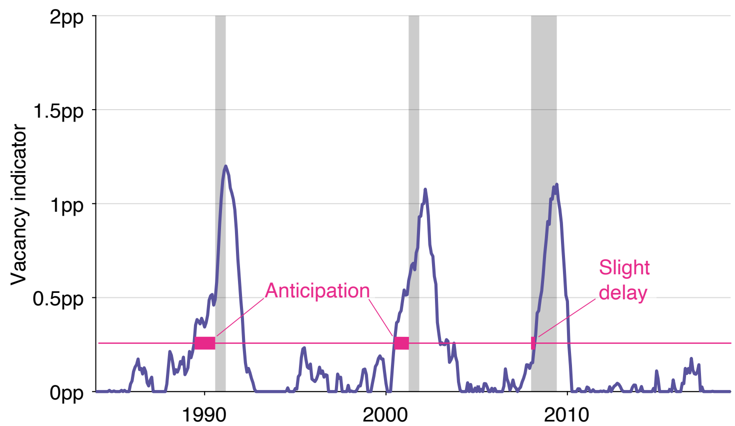
$$\sigma(k) = \sqrt{\frac{1}{J} \cdot \sum_{j=1}^{J} [\delta(k, j) - \mu(k)]^2}. \tag{17}$$

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.

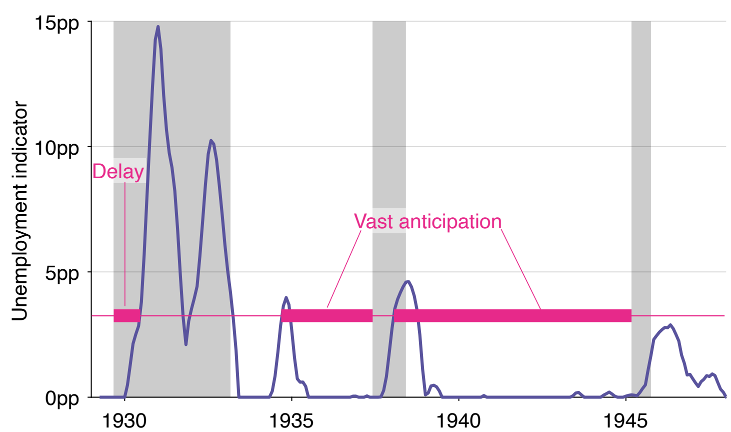
A classifier with higher precision (lower $\sigma(k)$) is also better, as it provides more precise, reliable information about the upcoming recession. It is clearly better to have a classifier that always detects a recession on its start date, than to have a classifier that is six months early half of the time and six months late half of the time—although both classifiers offer the same zero mean delay. For instance, the classifier in figure 7C is not appealing, although it anticipates the second and third recessions, because it actually anticipates




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).

these recessions too much. In that example, 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

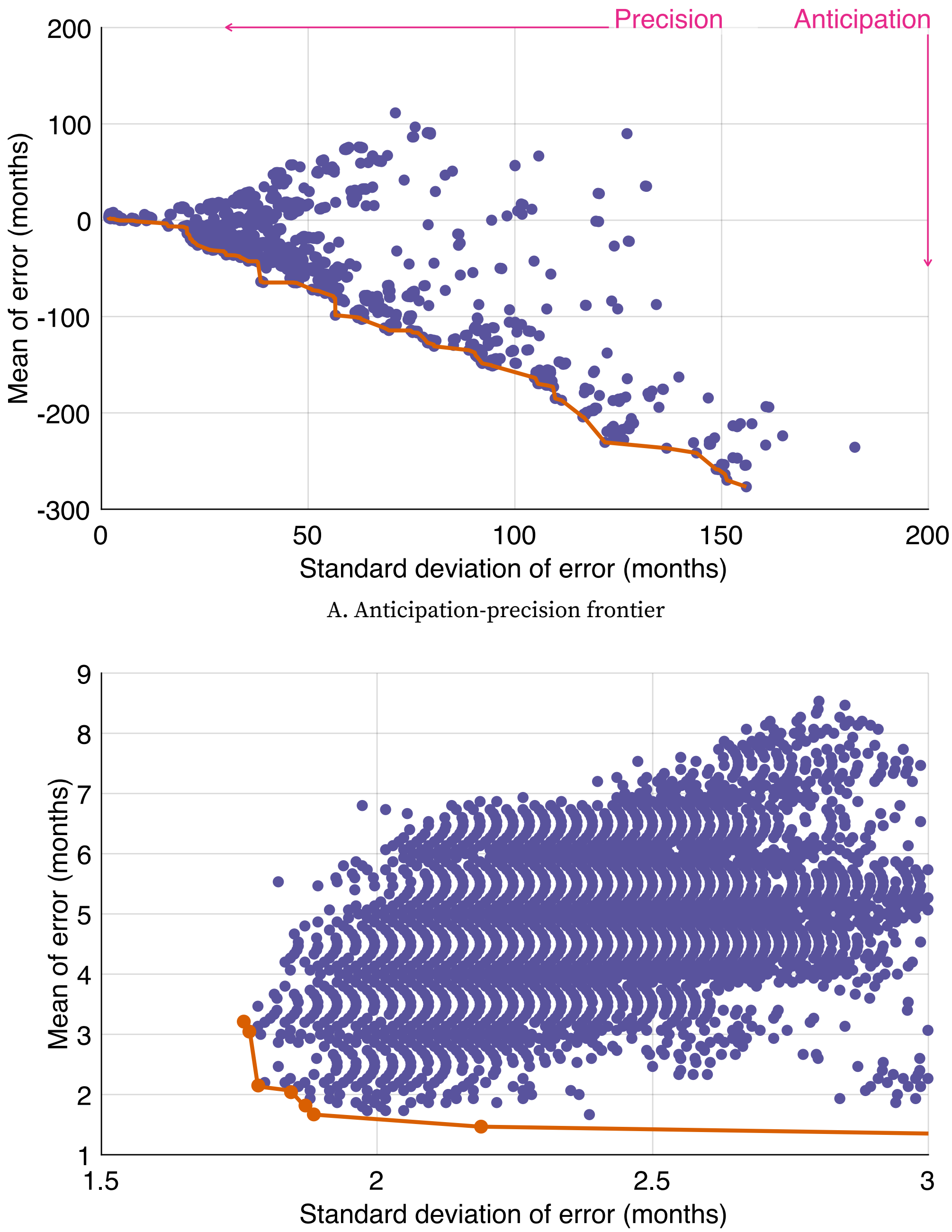
Next, I plot the mean and standard deviation of the detection delays for the 2,343,752 perfect classifiers (figure 8). I then select classifiers on the anticipation-precision frontier. The frontier comprises the classifiers with lowest mean delay (highest anticipation) and lowest standard deviation of delays (highest precision). For a given detection delay, no classifier is more precise than the classifier on the frontier; and for a given precision, no classifier detects recession as early as the classifier on the frontier. This frontier helps identify classifiers that balance early detection and accuracy.

The entire frontier includes 210 of the 2,343,752 perfect classifiers. The left-most classifier on the frontier has an anticipation of -3.5 months (mean delay of 3.5 months) and a precision of 1.6 months (standard deviation of delays of 1.6 months). The right-most classifier on the frontier has an anticipation of 273.4 months and a precision of 129.6 months.

### 4.5. Cropping the anticipation-precision frontier

How should we pick a classifier on the frontier to detect recessions out of sample? A policymaker with mean-variance preferences over detection delay would pick the classifier $k$ that minimize $\mu(k) + \lambda\sigma(k)$, 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 $\lambda$, unlike a policymaker who is more nimble, who might have a higher $\lambda$. 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, or of the user of the detection algorithm, it is not possible to pick the desirable classifier in that way. To circumvent this difficulty, 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 years. Indeed, assuming a normally distributed detection delay, 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




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 delays 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 delay is computed from (16) and the standard deviation of delays is computed from (17). The anticipation-precision frontier comprises the classifiers with lowest mean delay (highest anticipation) and lowest standard deviation of delays (highest precision).

confidence interval seems like a good choice to maximize the usefulness of the detection algorithm.[3]

Overall, the classifier ensemble comprises 7 classifiers on the frontier with a precision below 3 months. The 7 classifiers are perfect, so they detect the 15 recessions that occurred between 1929 and 2021 without producing false positives. The mean detection delay 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 delay 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 (the mean detection delay is positive) 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.

There are two clusters of classifiers on the anticipation-precision frontier (table 1). The first cluster consists of classifiers 1 and 2. They are broadly symmetric linear combinations of an unemployment 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 cluster consists of classifiers 3–7. They are the minimum of an unemployment indicator and a vacancy indicator, each representing broadly the change of an exponentially smoothed 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 a little away from the frontier. The Michez rule's mean detection delay over 1929–2021 is 1.9 months, while the standard deviation of its detection delay 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 delay 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

[3] The 3-month cutoff is an 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 $\alpha$ | 8m | 8m | 0.7 | 0.7 | 0.5 | 0.4 | 0.4 | |
| Turning horizon $\beta$ | 1m | 1m | 9m | 9m | 9m | 9m | 9m | |
| Curving weight $\gamma$ | 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 $\delta$ | 0.6 | 0.6 | 1 | 1 | 1 | 1 | 1 | |
| Threshold $\zeta$ | 1.24pp | 1.70pp | 0.78pp | 1.06pp | 0.70pp | 0.27pp | 0.19pp | |
| Detection delay (months) | | | | | | | | Average |
| Mean | 3.2 | 3.1 | 2.1 | 2.1 | 1.8 | 1.7 | 1.5 | 2.2 |
| Std 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 8. The ensemble comprises the classifiers on the frontier with a standard deviation of delays below 3 months—as depicted in figure 8B. The simple smoothing method with horizon $\alpha$ is given by (1) and (2); the exponential smoothing method with horizon $\alpha$ is given by (3) and (4). The turning horizon $\beta$ enters the construction of the classifiers through (5) and (7). The curving weight $\gamma$ enters the construction of the classifiers through (9) and (12). The u-v combination method with weight $\delta$ is given by (13); the min-max combination method with weight $\delta$ is given by (14). The detection delay is given by (15); its mean is given by (16); its standard deviation is given by (17).

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.

### 5.1. Recession probability from one classifier

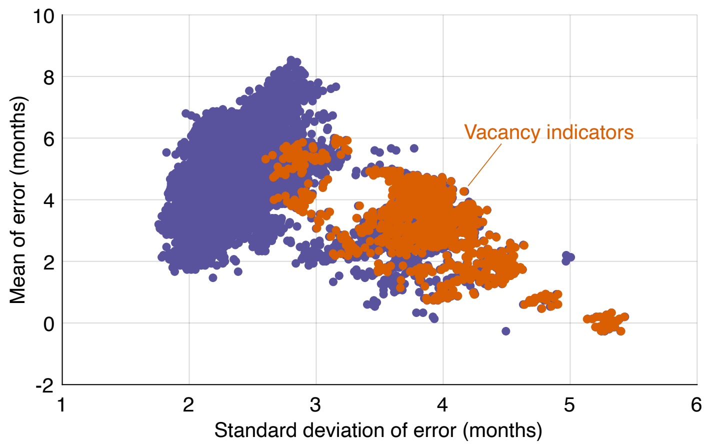
Whenever an individual recession indicator crosses its threshold, I infer the probability that the recession has already started based on the distribution of the detection error.

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 $k$ detect the $J$ recessions in the training period. Each detection $j$ generates a detection delay $d(k, j)$, which can be positive or negative if the detection anticipates the recession start. As defined in (16) and (17), the detection delays have mean $\mu(k)$ and standard deviation $\sigma(k)$.

Assuming that the detection delay $\delta(k)$ for classifier $k$ is normally distributed with mean $\mu(k)$ and standard deviation $\sigma(k)$—a convenient and neutral assumption—it is easy to compute the probability that a new recession has truly started at time $t_1$, given that the


A. Performance of unemployment indicators

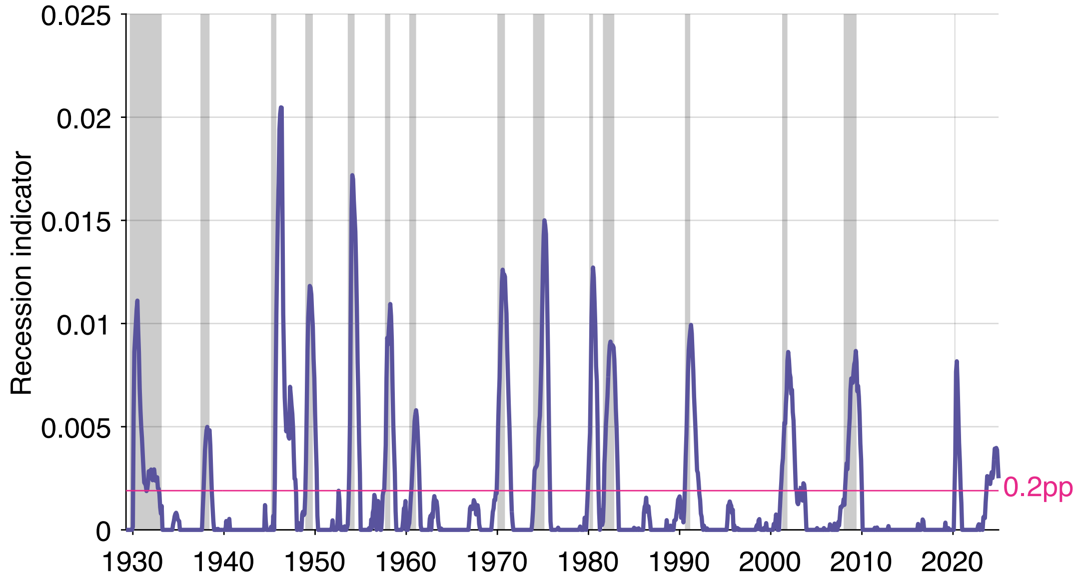



B. Performance of vacancy indicators

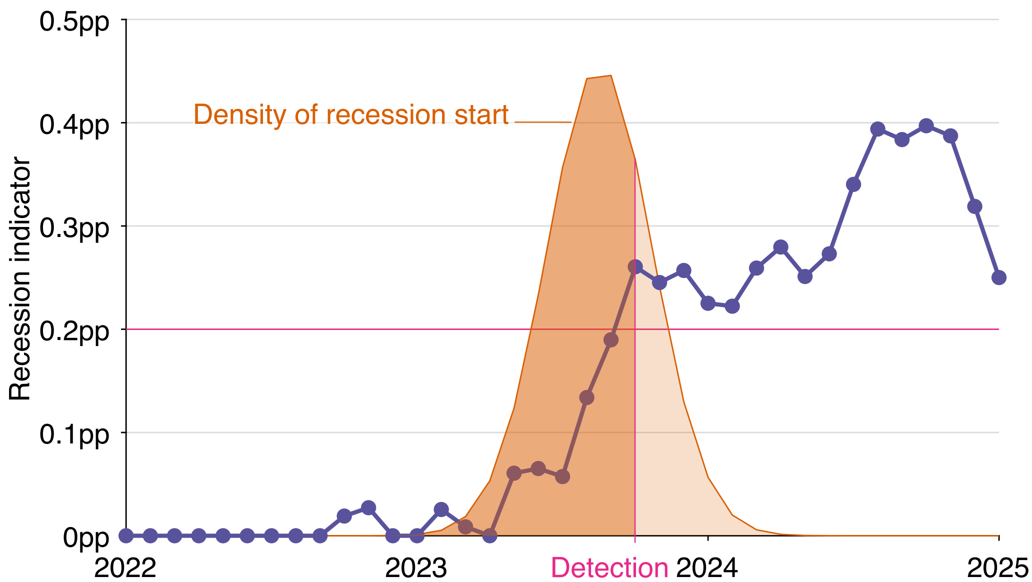

C. Performance of minimum indicators

FIGURE 9. Generalization of the Michez-rule results

The figure reproduces figure 8. In addition, it displays the mean and standard deviation of the detection delays for classifiers built from unemployment indicators (top panel), for classifiers built from vacancy indicators (middle panel), and for classifiers built from minimum indicators (bottom panel).




A. One selected recession classifier




B. Recession start probability density function from classifier

FIGURE 10. Computing the recession probability from one classifier

The figure displays as an illustration a specific recessions classifier and the associated probability density function of a recession start. Shaded areas indicate recessions dated by the NBER (2023).

classifier has detected a recession at time $t_0 \leq t_1$:

$$\begin{aligned} P(k, t_1) = \mathbb{P}(s < t_1 \mid d(k) = t_0, \mu(k), \sigma(k)) &= \mathbb{P}(d(k) - s > t_0 - t_1 \mid \mu(k), \sigma(k)) \\ &= \mathbb{P}(\delta(k) > t_0 - t_1 \mid \mu(k), \sigma(k)) \\ &= \mathbb{P}\left(\frac{\delta(k) - \mu(k)}{\sigma(k)} > \frac{t_0 - t_1 - \mu(k)}{\sigma(k)}\right) \\ &= 1 - \Phi\left(\frac{t_0 - t_1 - \mu(k)}{\sigma(k)}\right), \end{aligned}$$

where $\Phi$ 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 $d(k)$:

$$P(k, t) = \Phi\left(\frac{t + \mu(k) - d(k)}{\sigma(k)}\right), \tag{18}$$

as long as classifier $k$ remains in a recession state ($r(k, t) = 1$).

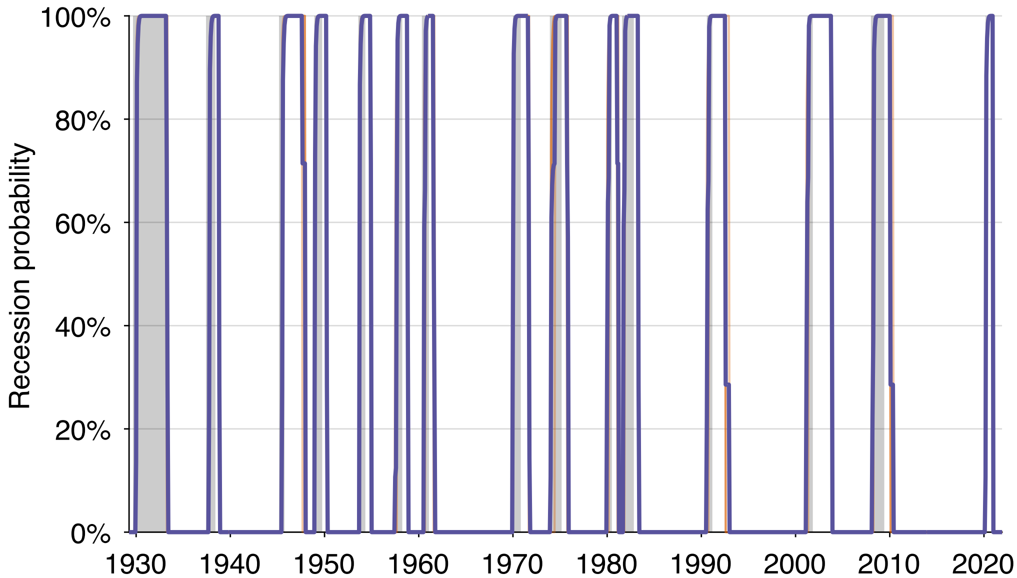
When the classifier $k$ signals that the economy is in expansion ($r(k, t) = 0$), I set the probability to $P(k, t) = 0$. The probability $P(k, t)$ is only positive when classifier $k$ signals that the economy is in recession ($r(k, t) = 1$). The value $d(k)$ used in the probability is then the most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

### 5.2. Recession probability from the classifier ensemble

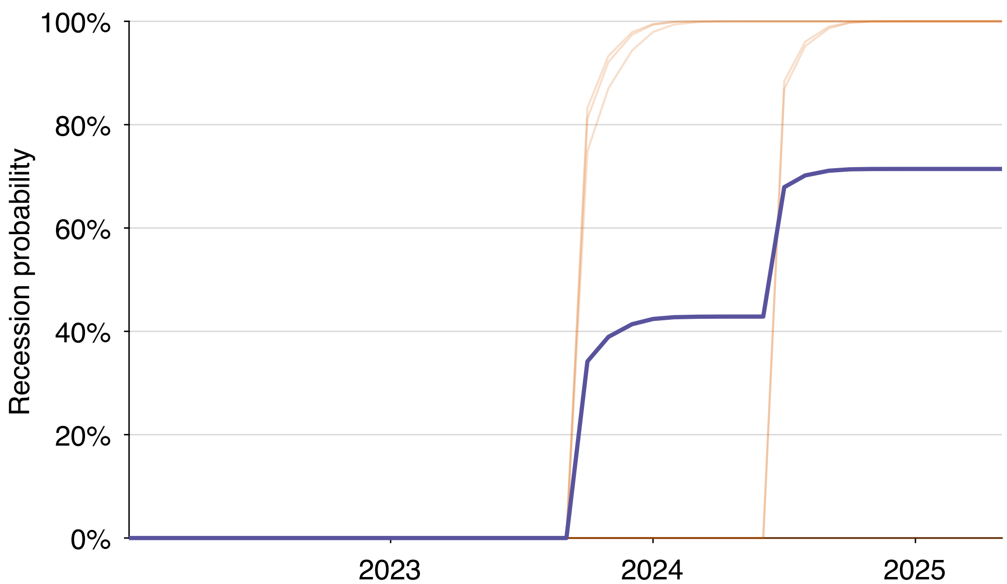
The final step is to average the probability that the recession has started across the $K = 7$ classifiers that I use for detection. Each classifier in the classifier ensemble generates a probability that the recession has started at time $t$, $P(k, t)$, given by (18). I compute the overall probability that the economy is in recession at time $t$ by averaging the probabilities produced by the individual classifiers in the ensemble:

$$P(t) = \frac{1}{K} \cdot \sum_{k=1}^{K} P(k, t). \tag{19}$$

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




A. Training over April 1929–December 2021




B. Testing over January 2022–May 2025

FIGURE 11. Recession probability from the classifier ensemble

The recession probability from the classifier ensemble is given by (19) (thick purple line). The probability is the average of the recession probabilities given by the individual classifiers in the ensemble, each given by (18) (thin orange lines). The classifiers in the ensemble are selected from the 1929–2021 anticipation-precision frontier, as illustrated by figure 8. Shaded areas indicate recessions dated by the NBER (2023).

TABLE 2. Performance of the algorithm on backtests

| | 2004 backtest | | 1984 backtest | | 1964 backtest | |
|---|---|---|---|---|---|---|
| # classifiers | 12 | | 6 | | 8 | |
| Delay (months) | Training | Testing | Training | Testing | Training | Testing |
| Mean | 2.4 | 2.7 | 2.3 | 1.2 | 3.4 | 3.9 |
| Std deviation | 1.9 | 1.9 | 2.1 | 1.4 | 1.5 | 2.6 |
| Min | −0.1 | 0.8 | −1.0 | −0.3 | 1.4 | 0.8 |
| Max | 5.7 | 4.6 | 5.2 | 3.5 | 5.3 | 9.3 |
| # recessions | 13 | 2 | 11 | 4 | 7 | 8 |

The table averages the mean, standard deviation, minimum, and maximum of the detection delays 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 delays 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.

that might mislead one classifier are less likely to mislead most of them.

### 5.3. Application of the recession detection algorithm to January 2022–May 2025

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 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 classifiers have been activated in the past 2 years (5/7 = 0.71). This is in turn due to the noticeable fall in the number of job vacancies and rise in the number of job seekers since the middle of 2022.

## 6. Backtesting the algorithm

Finally I backtest the detection method by shortening the training period and testing the algorithm on subsequent recessions. Backtesting confirms the robustness and reliability of the detection classifiers. Classifier ensembles produced from backtests also suggest a non-negligible probability that the US economy is in recession in 2025.

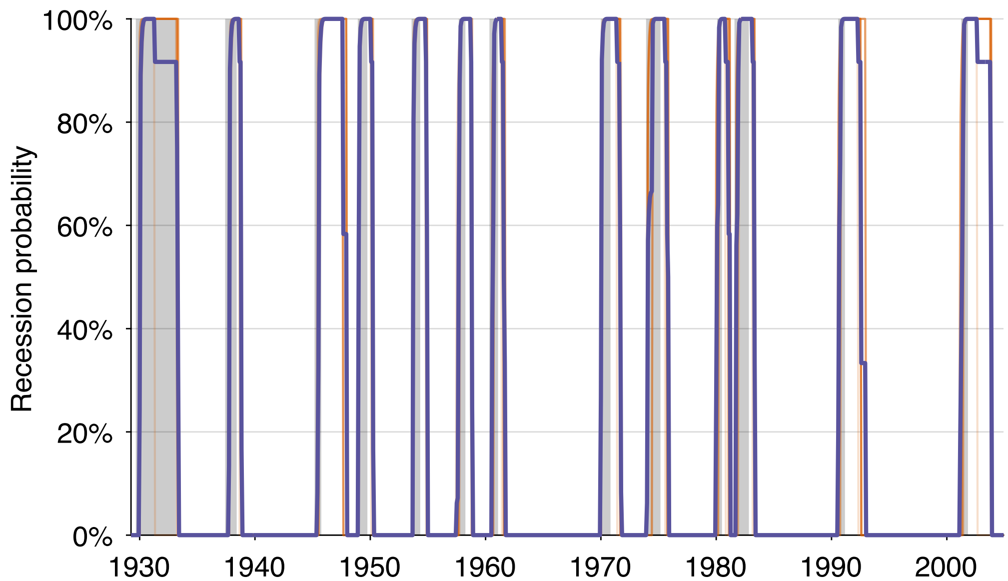
### 6.1. Backtesting to 2004

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 12). The training yields 12 classifiers on the anticipation-precision frontier with a

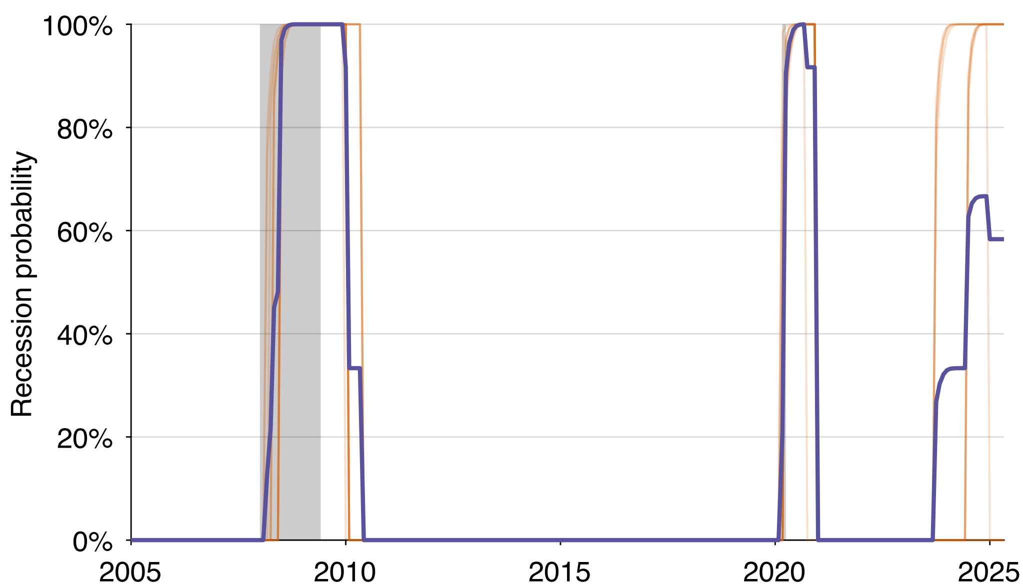
TABLE 3. Select simple classifiers from classifier ensembles obtained in backtests

| | 2004 backtest | | 1984 backtest | | 1964 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 $\alpha$ | 11m | 0.2 | 11m | 6m | 0.4 | 3m |
| Turning horizon $\beta$ | 1m | 1m | 1m | 1m | 9m | 9m |
| Curving weight $\gamma$ | 1 | 0 | 1 | 0 | 1 | 0.7 |
| Combination method | u-v | u-v | u-v | u-v | min-max | min-max |
| Combination weight $\delta$ | 1 | 0 | 1 | 0.4 | 1 | 1 |
| Threshold $\zeta$ | 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 delays 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 $\alpha$ is given by (1) and (2); the exponential smoothing method with horizon $\alpha$ is given by (3) and (4). The turning horizon $\beta$ enters the construction of the classifiers through (5) and (7). The curving weight $\gamma$ enters the construction of the classifiers through (9) and (12). The u-v combination method with weight $\delta$ is given by (13); the min-max combination method with weight $\delta$ is given by (14). The detection delay is given by (15); its mean is given by (16); its standard deviation is given by (17).




A. Training on 13 recessions between April 1929–December 2004




B. Testing on 2 recessions between January 2005–December 2021

FIGURE 12. Backtesting the algorithm to 2004

The recession probability from the classifier ensemble is given by (19) (thick purple line). The probability is the average of the recession probabilities given by the individual classifiers in the ensemble, each given by (18) (thin orange lines). The classifiers in the ensemble are selected from the 1929–2004 anticipation-precision frontier. Shaded areas indicate recessions dated by the NBER (2023).

precision below 3 months. I then use this 12 classifiers to detect the Great Recession and pandemic recession.

As an illustration, 2 of these classifiers are described in table 3. 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 is remarkably good. All 12 classifiers detect the 2 recessions without false positives. By June 2008 the recession probability was 1—although the algorithm had only been trained on data up to December 2004. By contrast, it was not entirely obvious that the recession had started in the summer of 2008. In August 2008, a notable macroeconometrician was circulating a working paper through the NBER arguing that there was no chance that the economy was currently in recession—and mocking Warren Buffet 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 2). In both cases the standard deviation of delays averages 1.9 months across classifiers. The mean delay averages 2.4 months over 1929–2004 instead of 2.2 months over 1929–2021.

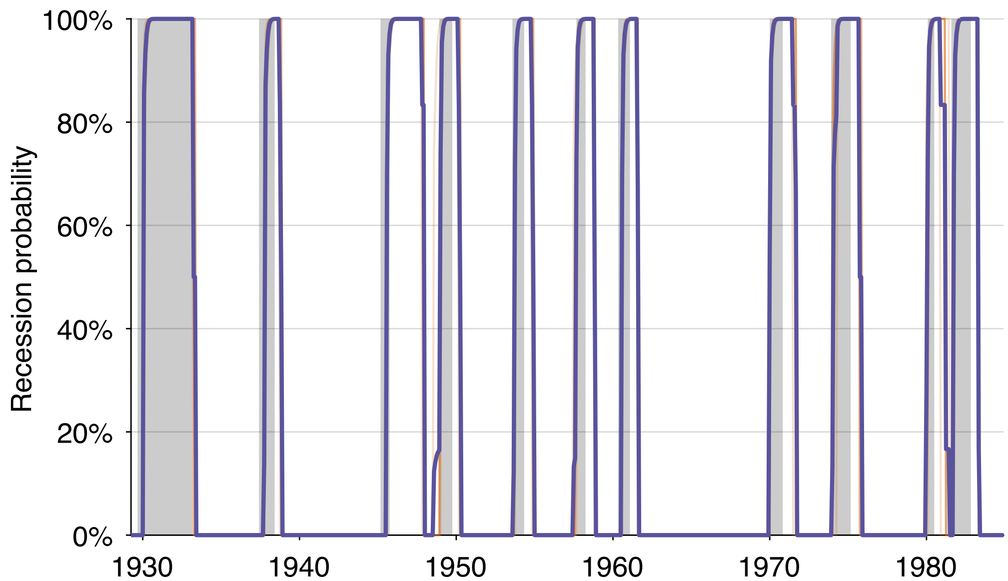
The performance of the algorithm over the testing period, 2005–2021, is comparable to that over the training period, 1929–2004. In both cases the standard deviation of delays averages 1.9 months across classifiers. The mean delay averages 2.7 months over the testing period, slightly longer than the 2.4 months over the training period.

### 6.2. Backtesting to 1984

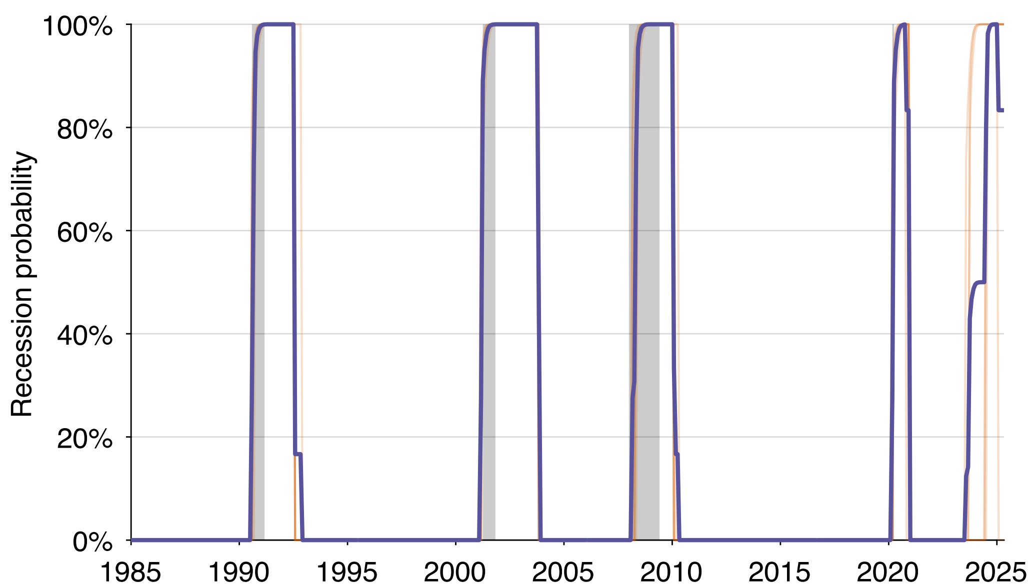
I then repeat backtesting by training the algorithm on data up to December 1984 and testing on subsequent data, from January 1985 to December 2021 (figure 14). The training period comprises 11 recessions, and the testing period features 4 recessions. The training yields 6 classifiers on the anticipation-precision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

As an illustration, 2 of these classifiers are described in table 3. 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 recessions without false positives. 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,




A. Training on 11 recessions between April 1929–December 1984




B. Testing on 4 recessions between January 1985–December 2021

FIGURE 13. Backtesting the algorithm to 1984

The recession probability from the classifier ensemble is given by (19) (thick purple line). The probability is the average of the recession probabilities given by the individual classifiers in the ensemble, each given by (18) (thin orange lines). The classifiers in the ensemble are selected from the 1929–1984 anticipation-precision frontier. Shaded areas indicate recessions dated by the NBER (2023).

before the dot-com recession had officially started.

The performance of the algorithm over the new training period, 1929–1984, is very slightly worse than that over the longer training period, 1929–2021 (table 2). The standard deviation of delays 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 delay 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 were not able to detect recessions as effectively.

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 delays averages only 1.4 months over the testing period, instead of 2.1 months over the training period. The mean delay 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.

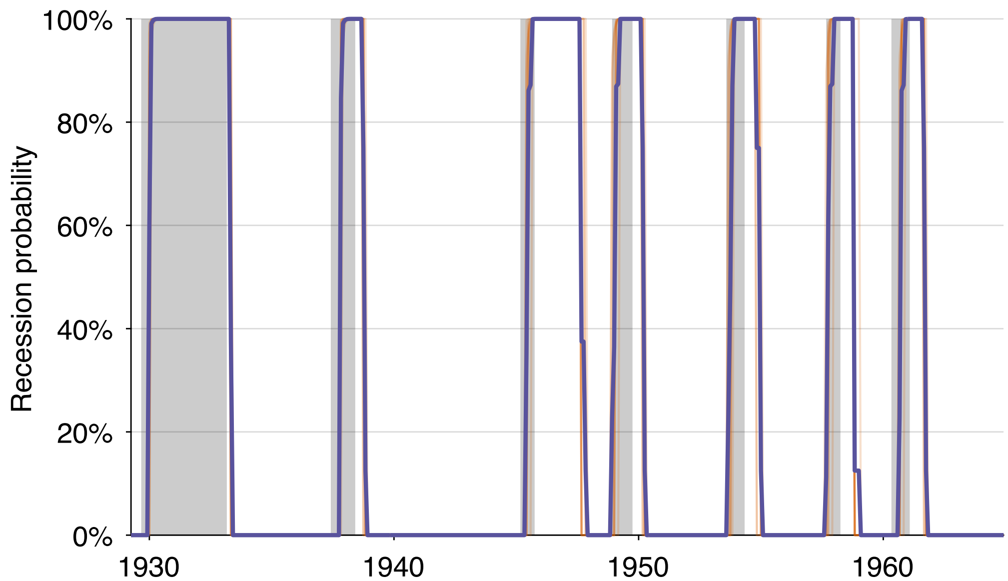
### 6.3. Backtesting to 1964

Last, I train the algorithm on data up to December 1964—which comprise 7 recessions—and test it on the subsequent 8 recessions (figure 14). 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.

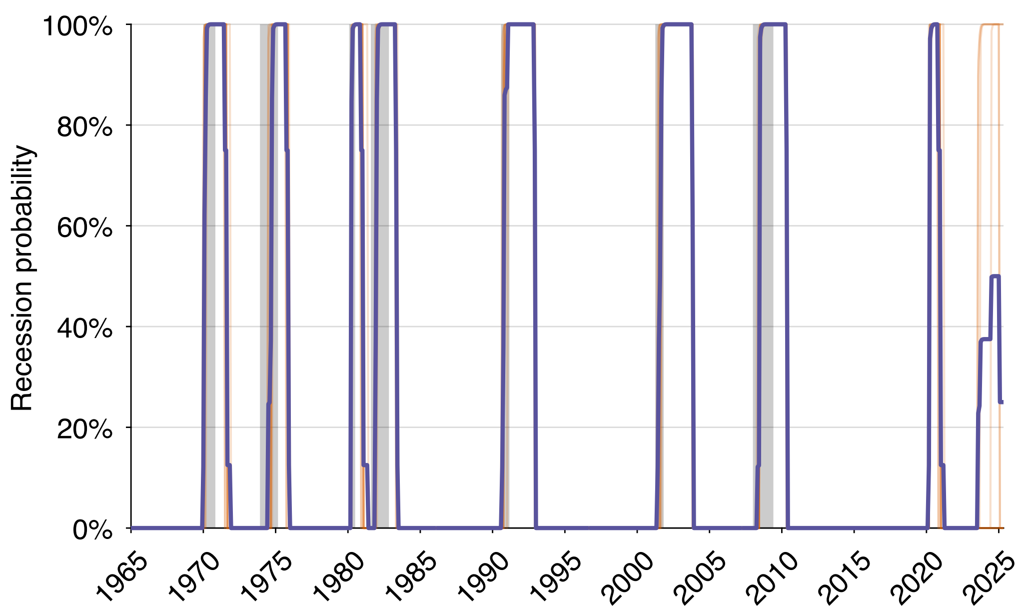
As an illustration, 2 of these classifiers are described in table 3. Both classifiers are based on minimum indicators. In fact, the indicator used to construct classifier 6 shares several similarities with the Michez-rule indicator: it is a minimum indicator; it smooths the data with 3-month trailing 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 that it needs to detect, the performance remains strikingly good. All 8 classifiers detect the 8 recessions without false positives. 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 that some of the state-of-the-art algorithms of the day that had seen real-time data (Leamer 2008).

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 2). The standard deviation of delays averages 1.5 months across classifiers for 1929–1964, which is less than the average of 1.9 months for 1929–2021. The mean delay 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



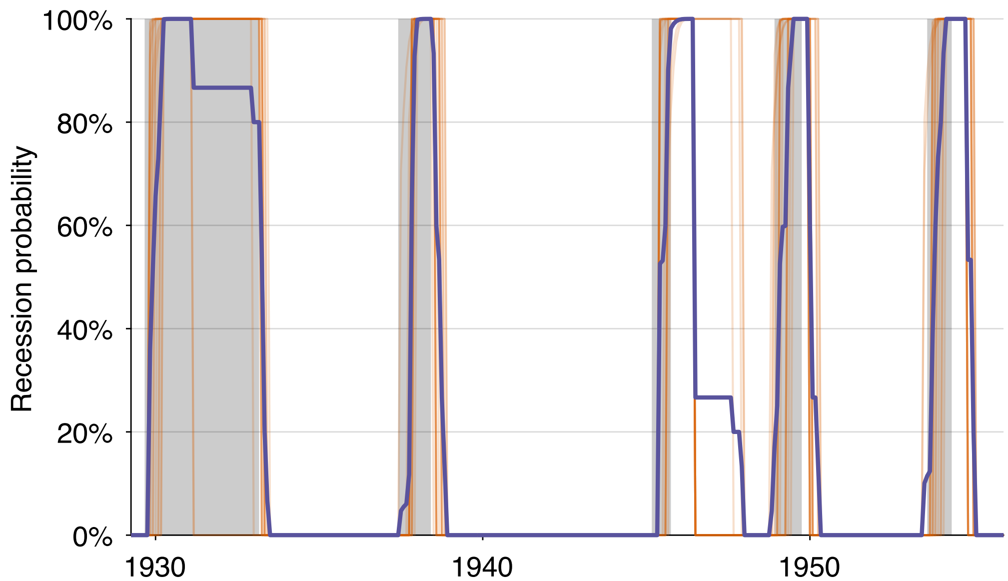

A. Training on 7 recessions between April 1929–December 1964



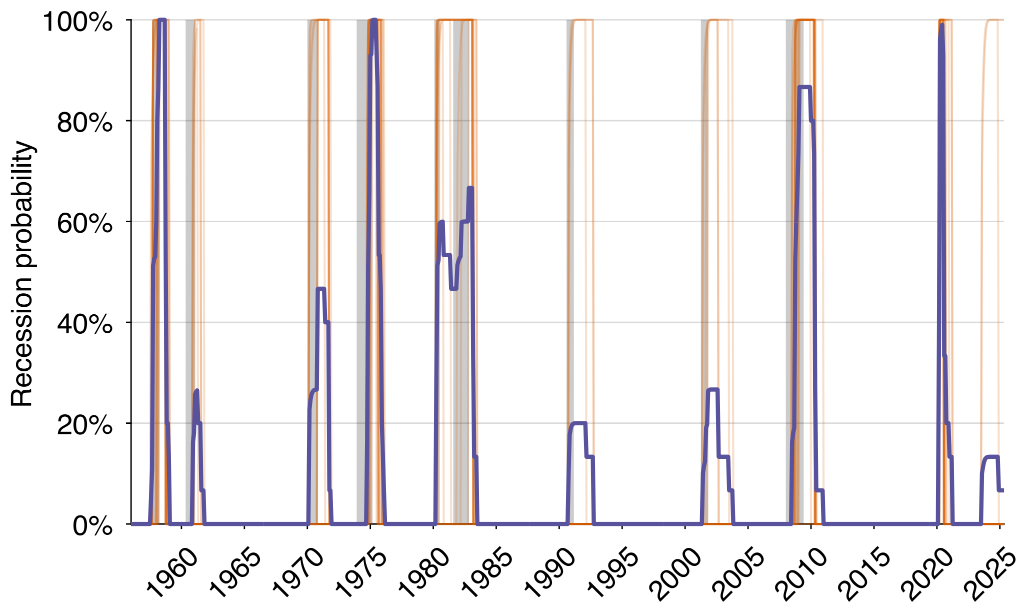

B. Testing on 8 recessions between January 1965–December 2021

FIGURE 14. Backtesting the algorithm to 1964

The recession probability from the classifier ensemble is given by (19) (thick purple line). The probability is the average of the recession probabilities given by the individual classifiers in the ensemble, each given by (18) (thin orange lines). The classifiers in the ensemble are selected from the 1929–1964 anticipation-precision frontier. Shaded areas indicate recessions dated by the NBER (2023).




A. Training on 5 recessions between April 1929–December 1955




B. Testing on 10 recessions between January 1956–December 2021

FIGURE 15. Backtesting the algorithm to 1955

The recession probability from the classifier ensemble is given by (19) (thick purple line). The probability is the average of the recession probabilities given by the individual classifiers in the ensemble, each given by (18) (thin orange lines). The classifiers in the ensemble are selected from the 1929–1955 anticipation-precision frontier. Shaded areas indicate recessions dated by the NBER (2023).

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 delays averages 2.6 months over the testing period, above the 1.5 months over the training period. The mean delay averages 3.9 months over the testing period, above the 3.4 months over the training period.

### 6.4. Backtesting to 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. When I train the algorithm until 1955 and test it over 1956–2021, the performance is noticeably worse (figure 15). The algorithm only has 5 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, 2 perfectly detect the 10 recessions over 1956–2021; 1 misses 1 recession; 1 misses 2 recessions; 1 misses 3 recessions; 2 miss 4 recessions; 3 miss 5 recessions; 3 miss 6 recessions; and the last 2 miss 7 recessions.

### 6.5. Application of the backtested classifier ensembles to January 2022–May 2025

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 11). The backtested classifier ensembles give different probabilities, but they all give positive recession probabilities in 2025—especially the more modern ensembles.

Indeed, 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 12). 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 13). 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 14). That probability peaked at 50% in 2024.

## 7. Conclusion

The paper develops a new method for detecting US recessions in real time from unemployment and vacancy data. The method uses an ensemble of recession classifiers selected from the anticipation-precision frontier. The method improves upon traditional approaches like the Sahm and Michez rules, which select arbitrarily how labor market data are filtered. 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 method 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 method optimizes jointly early detection and precision.

Using the ensemble of classifiers obtained from 1929–2021 data, I find that the probability that the US 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.

## 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. Data available at https://docs.google.com/spreadsheets/d/1fkMinSHkjTL99-bLZYFldQ8rHtgh8lxd.

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/books-and-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.

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. https://doi.org/10.1353/eca.2012.0024.

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* 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 Unemployment 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. 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.

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. Data available at https://drive.google.com/file/d/1NcuUMRR4_fOwJW6qHb-XxitVSXXONBHW.

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.

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.