The Pitfalls of Macroeconomic Forecasting[1]

"What do you think is going to happen to interest rates?" Merton Miller upon being introduced to a stranger at a cocktail party

Nobel Prize winning economist Merton Miller was well-known for his keen sense of humor. He was aware that the public thought most economists spent their time forecasting macroeconomic indicators like interest rates (even though is hardly true and most economists spend no time doing so). Miller's way of avoiding having to explain all of this was to beat people to the punch by asking them first.

While it is true that economists' track records at some types of predictions are mediocre, that is in part because the task of forecasting events like recessions is difficult. Moreover, there are quite a few things that economists do know about the properties of good and bad forecasts. So although economic forecasting can be difficult, there is quite a bit of research that we can draw on in analyzing them. As we will see, there are several very different challenges depending on the type of prediction that is being made.

1. Basic Statistics

Before turning to the actual task of evaluating the quality of an economic forecast, it is helpful to review several basic statistical concepts that are relevant to forecasting. The first is the notion of randomness. It is useful to realize that economists recognize that the outcomes that they are speculating about are not certain and instead could turn out in many different ways. They usually assume that there is a "model" that can be used to describe a variable of interest, say Gross Domestic Product (GDP), but that model is an approximation in several respects, so that it will not perfectly describe GDP. This means that if you replayed history over and over, there would be a distribution of outcomes that would obtain. The forecaster is usually trying to determine an outcome that is relatively likely to occur, recognizing that an exact match is extremely unlikely.

---

[1] ©2022 by Anil Kashyap, March 2022. This case was prepared to serve as the basis for class discussion, rather than to demonstrate the effective or ineffective handling of an administrative situation. I thank Julien Weber for outstanding assistance in preparing the case.

Increasingly, economists, and others such as election forecasters who try to communicate their analysis with the public, have attempted to convey the uncertainty in their projections to provide the forecast users with context. One early, trend setter in this respect was the Bank of England which began publishing what it called "fan charts" based on the Bank's Monetary Policy Committee's forecast for output and inflation – see Exhibit 1. The charts show the Committee's assessment of the probability of different outcomes for the two variables over the coming quarters. The Committee has tried to use the charts to explain to the public that its judgments depend on much more than a point estimate for either of variable and further that missing the central projections is not by itself an indictment of a forecast.

For numeric variables, there is a large literature in statistics that analyses the question of how best to judge the quality of a forecast, based on the distance between the forecast and the realized value of the variable being forecast. There are many competing indicators for this purpose.[2] Rather than dwelling on such comparisons, we will instead focus on two properties of a forecast that are relevant for many of the applications in macroeconomics.

One property relates to how forecasts are updated when new information arrives. In many cases, we want a forecaster to give her best estimate for an outcome using all the information that is currently available.[3] In that case, when new information arrives, we should not be able to predict the direction of how a forecast will be revised. This is because when the forecast was made there should be no way to recombine what was already known to arrive at a better forecast. This property of expectations is called the "law of iterated expectations".

This law is often used to test forecasts to see if they are updated appropriately. The most common form of a test is to see whether a forecaster whose last revision was in a certain direction can be predicted to have the next revision in that same direction. For instance, suppose the government releases some information that leads a person to increase her estimate for next year's GDP growth. When the next piece of news becomes available, the fact that she had previously raised her GDP forecast should not make her any more or less likely to raise the forecast again. If the sign of the last revision did help you predict her upcoming revision, then that means that her original forecast did not fully account for everything that was known before the news arrived.

Another set of diagnostics is often used for evaluating forecasts of binary variables, such as whether the economy is in a recession or not. The variable of interest is either true or false, then kind of diagnostics that used for evaluating forecasts often differ from those used when the outcome variable is "continuous".[4]

---

[2] See for a discussion of these issues see Armstrong, J.S. Evaluating Forecasting Methods. In Principles of Forecasting: A Handbook for Researchers and Practitioners (Ed. J. Scott Armstrong). Kluwer, 2001.

[3] Whether this is optimal depends on what forecaster is being rewarded for. A common example comes from "The Price is Right" game show where the closest guess that does not exceed the true answer wins. In this case, the best forecast is not an expectation of the actual price because that would equally likely to be too high or too low.

[4] By continuous we mean that it can take on a large number of outcomes and where the relative rankings of the outcomes is clear, such as for percentage of correct answers on an exam where a higher percentage is better.

Exhibit 2 shows one way of summarizing outcomes for a binary variable. In this case, it is unambiguous that a good model maximizes the correct number of event and non-event predictions (A and D in the exhibit) while minimizing the number of incorrect predictions of events and non-events (B and C). Statisticians formalize this objective by reporting two statistics. The first is "sensitivity" which is defined as D/(C+D) so that it represents how many true positives are predicted to be so. The second is "specificity" which is defined as A/(A+B) so that it represents how many negatives are correctly predicted. With these definitions, a good model has high sensitivity and low false negatives (1-specificity).

These concepts can be extended to help judge different variables that might be used to make a prediction for a binary event. Exhibit 3 shows a hypothetical version of how a prediction of recession might vary with an indicator (that in this example is called the slope of the yield curve and is assumed to range between +8 and -8). The distribution for whether a recession did or did not occur is plotted on the y-axis. In this representation, specificity is the area of the graph that is denoted by α and sensitivity is in the area denoted by β where the cutoff for determining the regression is when the indicator is above or below zero. This representation is helpful because the indicator variable itself could be a complicated measure, such as the predicted value from a regression model or some other transformation of many variables.

Exhibit 4 then shows a way to summarize many models. In particular, suppose you made many versions of Exhibit 3 that each represented a different model that were associated with a different value of the indicator variable, as it ranged from -8 to +8. For each of these models, we can calculate the values of specificity and sensitivity. The curve labeled "Realized" traces out the values of 1- specificity and sensitivity are associated with various models. A model that was perfect, meaning every recession that was predicted occurred and every prediction of no recession also was correct, would lie in the upper left hand portion of the box (at the point labeled perfect).

Alternatively, a particularly bad model would be one that sits on 45 degree line (labeled as a useless model). In that case, the model is equally likely to make mistakes in either direction, incorrectly predicting a recession that does not materialize, or failing to identify one that does. In Exhibit 3 this would amount to having models for the recession and no recession sit on top of each other.

Combining these bits of intuition suggests one way to summarize the success of a set of models.[5] Consider summing up the area under the curve at each point on the realized curve. The higher is that area, the better is the model. If we had a number of different candidate indicator variables, this summary statistic would allow us to compare across indicators. The appeal of this approach is that it does not depend on knowing which cutoff for any particular indicator is estimated to be the best.

[5] See https://www.youtube.com/watch?v=4jRBRDbJemM for an explanation of this diagnostic.

2. Forecasting Incentives

Before beginning to compare forecasts, it is worth considering competing considerations that are likely to influence people who are making them. As a benchmark suppose that everyone was simply trying to be as accurate as possible. Then we would expect each person to report as their best guess the objectively correct number plus some error that comes from the imprecision in their own model.

If this were the right framing of the problem, it would make the problem for users of the forecast easy because averaging forecasts from many prognosticators would eliminate the mistakes by individuals so that the objectively best estimate would emerge. Indeed, the idea that averaging across many forecasts (or taking the median forecast from a set) is often referred to as harnessing the "wisdom of crowds". In some cases, this approach has been shown to work well.

There are, however, reasons to question whether forecasters are really seeking to be as accurate as possible. There are many people producing macro forecasts and to get paid to do it, a forecaster has develop a following. This is easiest to do if the forecaster is selling a narrative about a particular way to think about the economy or a scenario. In that case, there may be financial incentives to be an outlier and attract attention.

One of the most prominent collection of macroeconomic forecasts is the Survey of Professional Forecasters (SPF) that is currently conducted by the Federal Reserve Bank of Philadelphia. Research by Goldman Sachs analyzed the submissions by SPF respondents and looked at the percentage of submitters "with growth forecast errors of over 1pp over 85% of the time across all of their SPF submissions."[6] They dubbed these submitters "noise" forecasters since they routinely offered outlier forecasts (potentially either high or low). These submitters tend to be dropped relatively quickly from the SPF, and as Goldman notes the percentage of these "perma-bear" or "perma-bull" participants has declined over time (Exhibit 5).

Likewise, even if a forecaster is giving a best guess, there is a question of how much confidence clients have in someone who is constantly changing their mind. If a forecaster appears to have no conviction in the forecast she has produced, why would a client trust any particular forecast and use it for planning? This consideration might mean that forecasters hang on to their narratives and numbers for as long as they can to avoid being indecisive, aka a "wind sock". In that case, at any point in time some forecasts would be stale and some would be best guesses, so averaging them may not be advisable.

A second reason that reported forecasts may not be best guesses is if the forecast itself is the input into a process that the producer of the forecast is seeking to influence. The most notable example of this is when a government produces a forecast to accompany its budget submission. More optimistic growth forecasts in this situation allow the government to propose higher

[6] See "Can Fed Forecasters Really Beat the Wisdom of Crowds?" Goldman Sachs Economic Research April 20, 2109

spending or lower taxes without showing a deficit, or holding spending constant to report a lower budget deficit.

Research by Jeffrey Frankel and Jesse Schrager suggests that there are systematic biases on these situations[7]. Exhibit 6 shows their empirical analysis of forecasts for a set of advanced economies between 1997 and 2013 where they compared to governments official forecasts for GDP and budget deficits to the average of a set of private sector forecasts (obtained from Consensus Economics) as motivated by the wisdom of the crowds logic. The exhibit shows two striking findings. First, the private sector consensus forecast is more accurate than the official forecasts for both budgets and GDP. Second, the official forecasts are biased towards showing more favorable outcomes, with higher project GDP and higher projected budget surpluses (or lower deficits). They also present regression analysis that indicates that if the official government forecast is more optimistic than the consensus, then the realized outcome for the growth (or the surplus) is predicted to be below official error. In other words, knowing the private sector forecast can help correct biases in the government forecasts.

While this research suggests that consensus private sector views are valuable, there are some situations where that might not be the case, particularly for emerging markets economies during period stress. During these crises, governments are often under tremendous political pressure and facing difficult market conditions. The International Monetary Fund may also be involved working out a lending program. It is not uncommon for government officials to make it known that private sector analysts that offer unfavorable assessments may be cut off from briefings and denied access to off-the-record conversations with the officials. Thus, one often sees that in these situations, private sector forecasts will mirror public sector ones for some critical variables at least during early phases of a crisis.

Overall, these observations suggest that there are situations when it is particularly important to think about the motives and incentives of people making forecasts. Fortunately, for advanced economies in normal circumstances these considerations are not usually a problem for reputable private sector forecasts.

3. Forecasting Recessions

One of the most common consulted macroeconomic forecasts is for economic growth and more particularly whether the economy will be in recession. To forecast a recession one must start by defining one. The United States follows an unusual practice of having a private organization, the National Bureau of Economic Research (NBER), make the determination of whether the economy is in recession. That decision is taken by the Business Cycle Dating Committee of the NBER.[8] Exhibit 7 gives an overview of the NBER's description for how it determines recessions. In many countries there is not an official body that makes this determination and instead a more colloquial rule of thumb of two consecutive quarters of declining GDP is used.

---

[7] See Jeffrey Frankel and Jesse Schrager, "Bias in Official Fiscal Forecasts: Can Private Forecasts Help?" NBER Working Paper No. 22349, June 2016

[8] See https://www.nber.org/research/business-cycle-dating for more information.

Because the NBER rule involves much judgment and because the committee is not attempting to be timely in announcing the onset of a recession, many observers try to find ways to predict the committee's announcements. Of course, if the economy stalls but narrowly avoids a being labeled as being in recession by the NBER, anticipating this kind of slowdown is usually of interest to many people too. So exactly matching the committee's decisions is not just what forecasters are trying to do.

Staff at the IMF looked at the ability of different forecasters to predict recessions.[9] They studied 63 countries (29 advanced economies and 34 emerging economies) over the period 1992 to 2014 (though for some countries less data were available). They sought to assess whether in years where a country had negative growth for the year (that they use as their definition of a recession) economists were able to predict that. They considered forecasts made in April the year before and in October of the year in question. For example, the forecasts for the year 2009 would include predictions made as of April 2008 and October 2009.

For the representative private sector forecast they took the average growth estimate from the Consensus Economic collection of forecasts. Their results are shown in Exhibit 8. One of the challenges in forecasting a recession is that recessions are relatively rare, occurring only 11.7% of the time in this sample. These consensus forecasts rarely call for negative growth a year ahead, this happens less than 0.1% of the time, and when such forecasts are made they are more likely to be correct than incorrect. Only 3.3% of the actual recessions are correctly predicted.

By the time we reach October of the year in question, there is more willingness to forecast negative growth. The percentage of cases correctly predicted rises to 77.1%. Recall that given lags in data availability, forecasts made in October usually would have data through the middle of the year, nevertheless this evidence suggests that economists have limited success in foreseeing recessions defined in this way.

The IMF staff also looked at the forecast revisions that occurred as more information arrived. To do this they looked at how a revision between October of a given year (e.g. 2008) and April of the next year (2009), related to the next revision (between April 2009 and October 2009). These results are presented separately for advanced economies and emerging economies, and they also compare the consensus forecasts to those made by the IMF staff. The results are in Exhibit 9. Their findings suggest that revisions do have predictive power for future revisions, much more so for the consensus forecasts than the IMF ones. The persistence in the revisions for the emerging economies was clear for both forecasts, and only marginally significant for the IMF forecasts.

Another heavily researched area is the usefulness of individual variable in forecasting recessions. The largest body of this work pertains to the U.S. and much of the attention has focused on the information in interest rates, more particularly interest rate spreads that compare rates on long dated securities to shorter dated ones. These differences are often referred to as the slope of the

[9] See Zidong An, João Tovar Jalles, and Prakash Loungani "How Well Do Economists Forecast Recessions?" IMF Working Paper 18/39.

yield curve, although depending on which interest rates are under consideration the differences in yield could include factors besides just the maturity differences.

One reason for the interest in these kinds of variables is the fact that participants in financial markets are forward looking and very interested in the outlook for the economy. Because interest rates are available throughout the day it is very tempting to assign a macroeconomic interpretation to their movements. However, there is some historical basis for thinking that the slope of the yield curve has some predictive power for recessions in the United States.

Exhibit 10 graphs the difference in the yield for a 10 year constant maturity Treasury Note and a one year Treasury Bill since 1970. Recessions as defined by the NBER are shaded with grey. This yield spread has been negative at some point before each of the recessions in the last 50 years, although the lead time between the negative reading and the onset of the recession has varied. Because of this pattern, whenever the curve inverts, i.e. the spread becomes negative, the financial press pays great attention to that development.

Another indicator that has gained attention in recent years is a transformation of the unemployment rate that was proposed by Claudia Sahm, when she was working at the Board of Governors of the Federal Reserve System. Her measure is to take the three month average of the unemployment rate and compare that to the lowest recorded month in the last 12 months. Her indicator is graphed along with NBER recession dates in Exhibit 11. She noted that whenever this variable exceeded 0.5, a recession was underway. One way to interpret this observation is to note once the average unemployment is up by half a percent (relative to the recent trough) it keeps going. Put differently, the unemployment rate is sufficiently persistent that it does not have half point blips that are due to noise.

Of course, there is no theory that leads to the threshold of 0.5. In a 2019 blog post on the economy, Jay Shambaugh and colleagues at the Hamilton Project computed alternative versions of her rule with alternative thresholds.[10] The calculations are shown in Exhibit 12. Part of the interest in this rule comes from detecting a recession before the NBER has called one, so finding a lower threshold would be of interest. As their calculations show, surges in the average of 0.4 to 0.49 also appear to be highly predictive of impending recessions, though these calculations are based only on 17 months.

4. Medium Term Forecasts

A different set of issues arise when economists attempt to make forecasts about the medium terms, taken here to be 5 to 10 years ahead. At these horizons, the business cycle is nearly irrelevant in that knowledge of whether a recession or boom is currently underway gives very little information about growth prospects at that time.

Perhaps the most common application of this problem is by the U.S. Congressional Budget Office. The CBO is the non-partisan organization that advises the U.S. Congress on the budgetary impact of potential legislation, typically "scoring" that legislation. The CBO

[10] See https://www.brookings.edu/blog/up-front/2019/06/06/how-will-we-know-when-a-recession-is-coming/

calculates these impacts over 10 years, under the assumption that the current law will prevail over that period. In doing these calculations, the CBO does two pieces of analysis. The first is a baseline forecast for the economy over the next 10 years. This projection is made each January and then updated in August. These forecasts are intended to give Congress an assessment of the current trajectory of the economy. The most recent forecast, released in January 2021, is shown in Exhibit 13.

The first few years of the project take account of the current state of the economy and the CBOs judgment about how it will evolve. While the key input into the latter half of the projection is the CBO's estimate of potential GDP growth. The logic for doing so is well captured by this part of the CBO report that accompanied the January 2020 Economic Outlook:

> Typically, in CBO's forecasts, the growth of actual output and the growth of potential output converge in the second half of the 11-year projection period, and the level of actual output stays about 0.5 percent below that of potential output, which reflects the agency's consideration of the average effect of recessions from a historical perspective. However, that convergence is interrupted in the current forecast because the expiration of certain provisions of the 2017 tax act not only diminishes the growth of potential output by reducing the supply of labor but also temporarily slows the growth of overall demand. As a consequence, actual output temporarily falls relative to potential output. It then rises until the relationship between the levels of actual and potential output reaches its long-run average in the final years of the projection period. Correspondingly, the average growth of actual output during the 2025–2030 period is close to, but slightly slower than, that of potential output.

The second piece of the analysis is an adjustment to the baseline that would result from the legislation being evaluated. For tax legislation, the CBO typically takes estimates from the Joint Committee on Taxation, a non-partisan committee of Congress that works closely with the tax writing committees in both the House of Representatives and the Senate. The CBO staff attempt to incorporate all the direct and indirect effects of the programs that are affected. For instance, when the Affordable Care Act (aka Obamacare) was scored, the CBO estimated that it would be expected to *reduce* deficits over the coming years (despite the expansion of coverage that was an intended effect of the bill) – though the CBO noted there was substantial uncertainty surrounding the estimates. Exhibit 14 reproduces the summary table of the legislation that was sent to Congress that shows the myriad channels that are considered in the analysis.[11]

5. Long Term Projections

Some of the most consequential macroeconomic forecasts relate to very long-term projections that are made for the purposes of evaluating pension solvency. Even in rich countries like the United States, most people have very little savings and so they are highly dependent on promised government pensions. Hence, determining whether these pensions are adequately funded is a critical question.

In the U.S., social security represents a substantial portion of many people's pensions and the challenges in determining the funding situation for social security is a good case study of the issues associated with long-term projections. Social security is overseen by four government

[11] See https://www.cbo.gov/sites/default/files/111th-congress-2009-2010/costestimate/hr3962rangel0.pdf

officials and two Presidential-appointed outsiders (who must be confirmed by the Senate).[12] Each year the Trustees issue a report on the health of the system's finances, including an estimate of when the taxes that are due to be collected will fall short of the promised level of benefits.

Every four years, there is also a group of external technical experts that is convened to review the assumptions made by the Trustees and to evaluate the methods used by the Office of the Chief Actuary to project the future financial status of the program. The most recent external report was in submitted in 2019. Both the annual Trustees report and the quadrennial external assessment are long documents that contain extensive analysis.

One indicative, important decision that the Trustees must make each year is to forecast real wage growth over the next 75 years. Because benefits are tied to wages this one of the most important elements of their report. One peculiar aspect of this calculation is that in the near term, rising wages mean contributions to the program increase much more than payouts do because most people are not retiring right away. So the solvency of the program is improved for some time by increased real wages.

They make their projection by expressing real wages as the product of five terms as follows:

$$\frac{\dfrac{\text{Earnings}}{\text{CPI Deflator}}}{\text{Employment}} = \frac{\dfrac{\text{GDP}}{\text{GDP Deflator}}}{\textit{Hours}} \times \frac{\text{Compensation}}{\text{GDP}} \times \frac{\text{Earnings}}{\text{Compensation}} \times \frac{\text{Hours}}{\text{Employment}} \times \frac{\text{GDP Deflator}}{\text{CPI Deflator}}$$

Real earnings = Productivity x Labor's share x Benefit wedge x Hours worked x Relative price changes

There are several reasons for framing the projection in terms of these components. As we will see the different pieces have very different temporal patterns, so that handling them separately allows the modeling to account for that. In addition, this decomposition makes it easy to do sensitivity analysis to assess the importance of alternative assumptions. A third reason is that some of these same components will be relevant for other parts of the report, so this makes it easier to insure that there is internal consistency of the whole analysis.

The dominant determinant of real earnings is productivity (Exhibit 15). Unfortunately, as the technical experts noted this series is both volatile and hard to predict. Because it can swing by large amounts even when averaged over a decade, this poses a forecasting challenge. Many economists anchor their productivity forecasts on what happens in the non-farm business sector of the economy. This sector covers about 75% of the economy and up until recently the longer run average growth rates were relatively stable, however, since the 2008/9 global financial crisis productivity in that sector had been low.

The share of GDP that is accruing to workers is shown in Exhibit 16. That share had been relatively stable for the last decade, having trended down until then. Some of the fluctuations in

[12] There are actually two separate programs that are jointly administered, the Old-Age and Survivors Insurance and Disability Insurance Trust Funds. The first of these is what is commonly referred to as Social Security. The other program provides insurance for people who are deemed to be unable to work because of a disability.

this share is likely due to the changing nature of the economy and in particular to the type of capital used in production and to changes in housing markets. The shift toward capital that depreciates faster means that the base rate of investment will rise and the proceeds left for labor will shrink. Likewise, as housing becomes more expensive, the flow of services going to owners of housing will also rise. This again leaves less available to be paid to labor. The lower part of the Exhibit shows that once these compositional shifts are taken into account, the net share for labor in the remainder of GDP is much more stable. The Trustees were projecting a rebound in their forecast for the overall share.

The benefit wedge matters in these calculations because social security payments are tied to earnings and not total compensation (that includes any untaxed benefits that might nonetheless be valued by workers). So when firms shift to compensating employees with benefits such as health insurance or pension contributions, this reduces the fraction of compensation that is subject to the social security tax. Exhibit 17 shows that since about 2000 this share been relatively stable.

The average number of hours worked per employee is shown in Exhibit 18. The hour series trended down from 1960 until the early 1980s. The conventional explanation for that decline was the increase of women who were entering the labor market and often working part-time. Since the early 1980s the average has had a slightly negative trend and the Trustees project that pattern to continue.

Finally, Exhibit 19 show the ratio of the Consumer Price Index relative to the GDP deflator. These two different deflators matter because social security benefits are adjusted to account for changes in the CPI, while the level of employment depends on the size of the economy (and hence GDP). One way to think about this is to suppose that productivity is held fixed and consider only the impact from other additional changes in prices. This perspective explains why when prices change in ways to raise both the CPI and GDP, the effect essentially cancels out for the purposes of social security solvency. Hence, all that matters is when one price index moves more than the other. Relatively faster CPI inflation will raise benefits relative to contributions. Lately, the inflation measured by the GDP deflator has been higher than in the CPI. The Trustees project that the recent pattern will persist.

The study questions explore the implications of these assumptions.

6. Conclusion

Macroeconomists are asked to make many different kinds of forecasts. Because these forecasts are likely to be used in different ways, the methods that underlie them differ. Also the risks that could compromise a forecast change as we move from near-term to medium term to long run projections. This case has shown us some of these challenges.

Study Questions

Answer these questions and be sure to sign the honor code in accordance with the class rules. Your case write-ups are due in Canvas by 7:30 am on March 28. If you are not registered because you are taking the class via Take 3, please email your solutions to the TA.

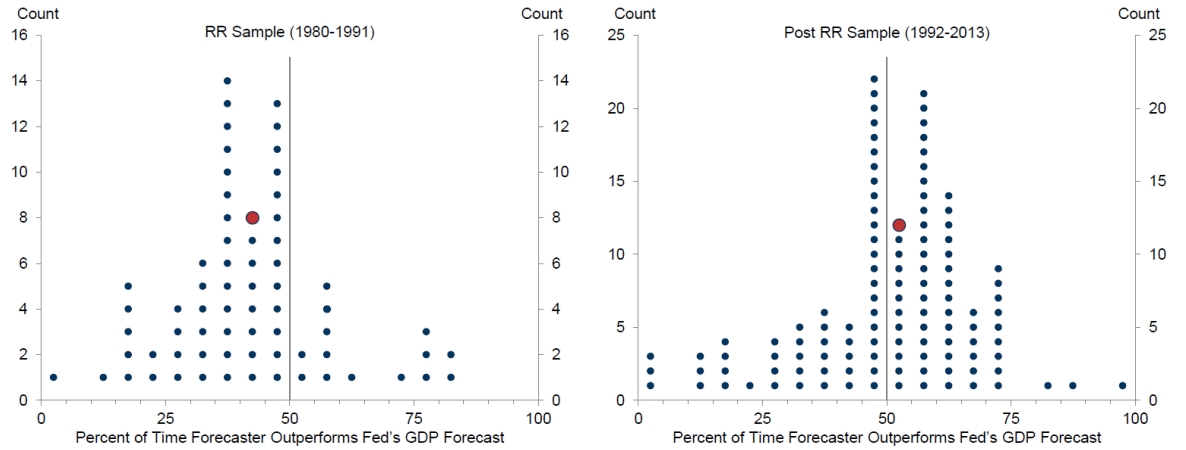
1) Give at least 3 reasons why a macroeconomist's current reported forecast might not be her best estimate of an outcome that is uncertain.

2) Looking at the data in Exhibit 9, what (if anything) does the law of iterated expectations say about these findings? Be sure to comment on both panels in the Exhibit.

3) What do the data in Exhibit 8 tell us about the ability of macroeconomists to predict slowdowns? What relation (if any) does this have to do with your answers to question 1?

4) The data shown below from the Goldman Sachs report mentioned in the case compare the forecasts of the submitters for the Survey of Professional Forecasters before and after 1992 with the Federal Reserve staff forecasts. (The split in the sample period is because a previous paper by Romer and Romer using data up until 1992 found that the Fed staff outperformed the private forecasts).

Exhibit 5: The Typical Individual Forecaster Has Improved Relative to the Fed

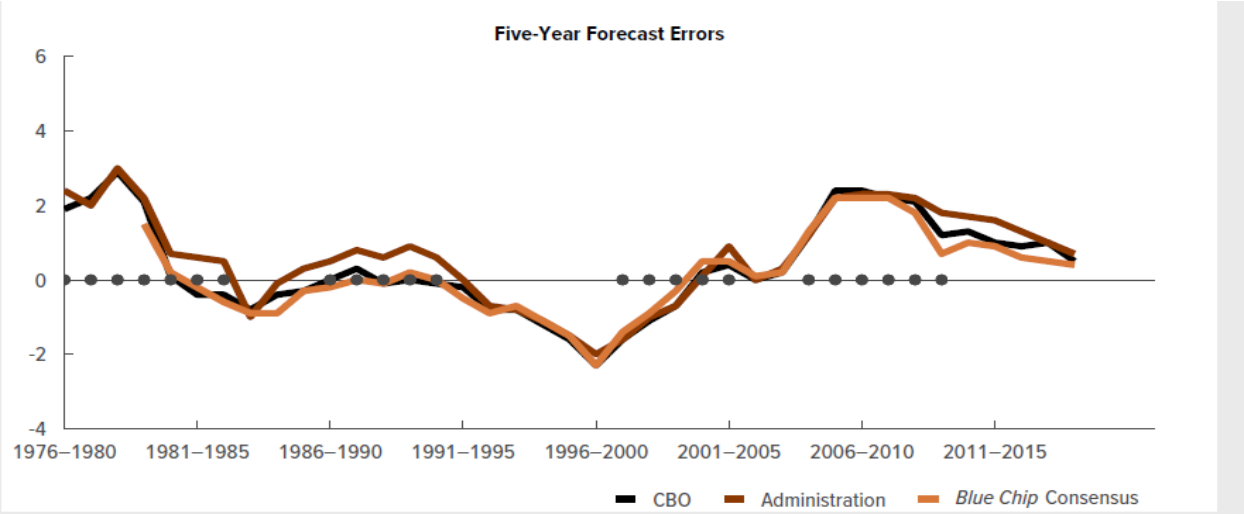


Note: Red dot indicates median forecaster.

Source: Federal Reserve Bank of Philadelphia, Goldman Sachs Global Investment Research

What might explain the differences in findings? Given the recent evidence what should the public conclude if they are told there is a difference between the private sector and the Fed staff's outlook for the economy? In particular, should the public conclude that the Fed has private information about the economy that is being revealed?

5) Below are data comparing forecast accuracy of forecasts made by the Office of Management and Budget (on behalf of the Presidential administration), the CBO and the Blue Chip consensus.



Sources: Congressional Budget Office; Office of Management and Budget; Wolters Kluwer, *Blue Chip Economic Indicators*; Bureau of Economic Analysis.

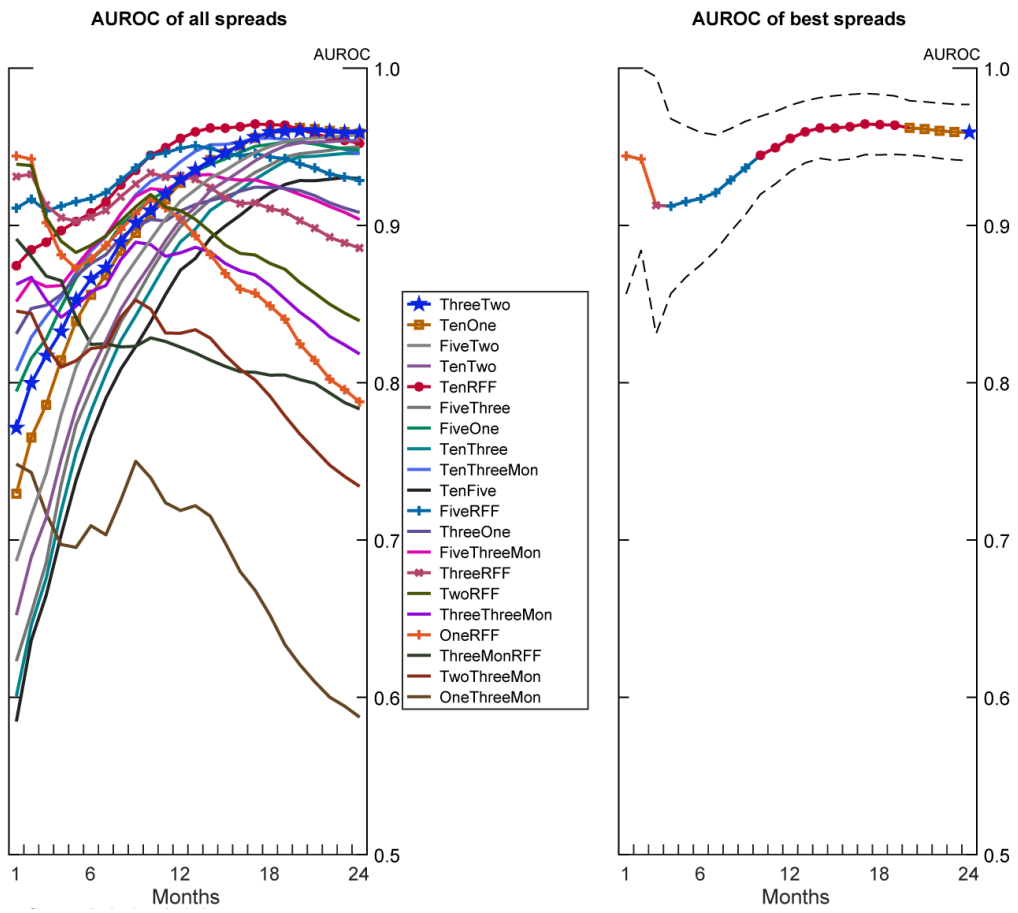
The measure of real output is gross national product in years before 1992 and gross domestic product in 1992 and later years. Positive errors represent overestimates. The dots shown on the horizontal axis indicate that the forecast period overlapped a recession by six months or more. The years indicate the time span covered by each of the forecast errors shown in the figure.

What conclusions do you draw from these data? Are Democratic or Republican administrations very different?

6) David Miller at the Federal Reserve Board of Governors analyzed the predictive power of different yield curve spreads for the purposes of forecasting recessions. More specifically, he estimated a model from 1984 to 2018 where the dependent variable is an indicator that, conditional on not being in a NBER dated recession this month, a recession begins within the next N months where the explanatory variable was alternative yield curve spreads. The area under the curve (analogous to what we described in Exhibit 3) is shown below.

The maturities of the yields being compared are listed in the names of the variables, where RFF is the federal funds rate, and the other rates considered are three months, 1 year, 2 years, 5 years, and 10 years.

Is there any economic explanation you can assign to why the spreads that are the best for forecasting change depending on the horizon of the forecast?

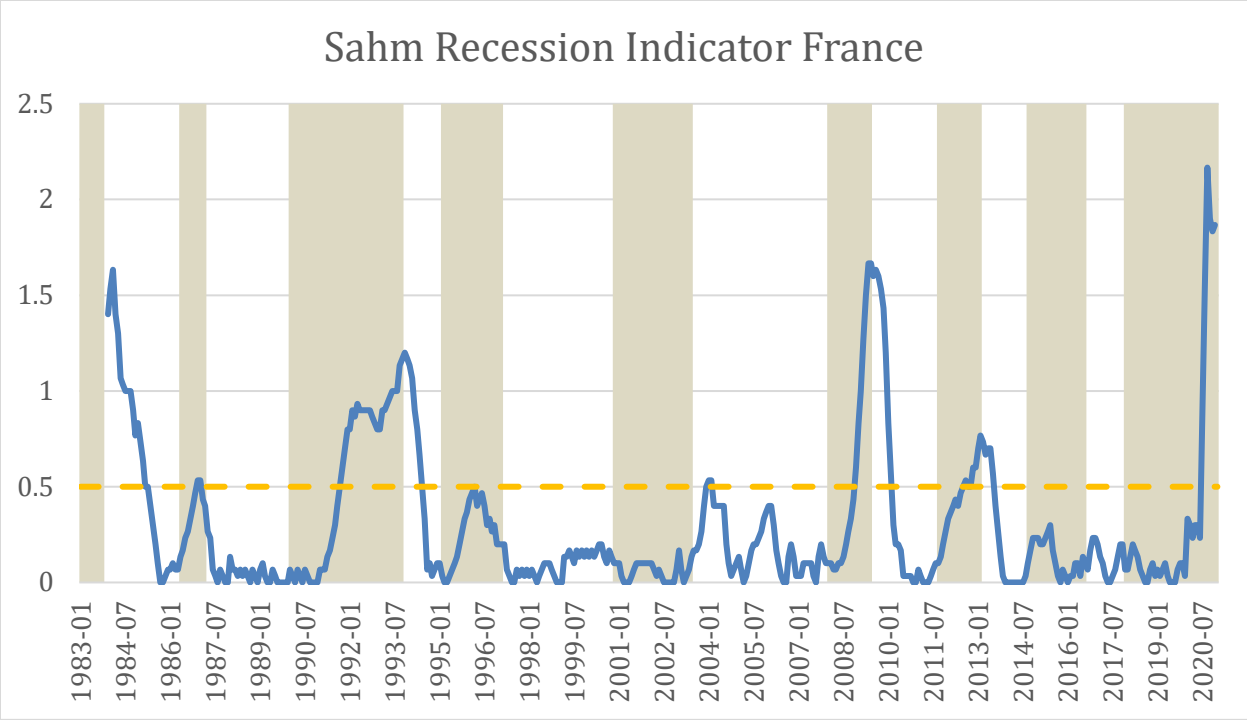


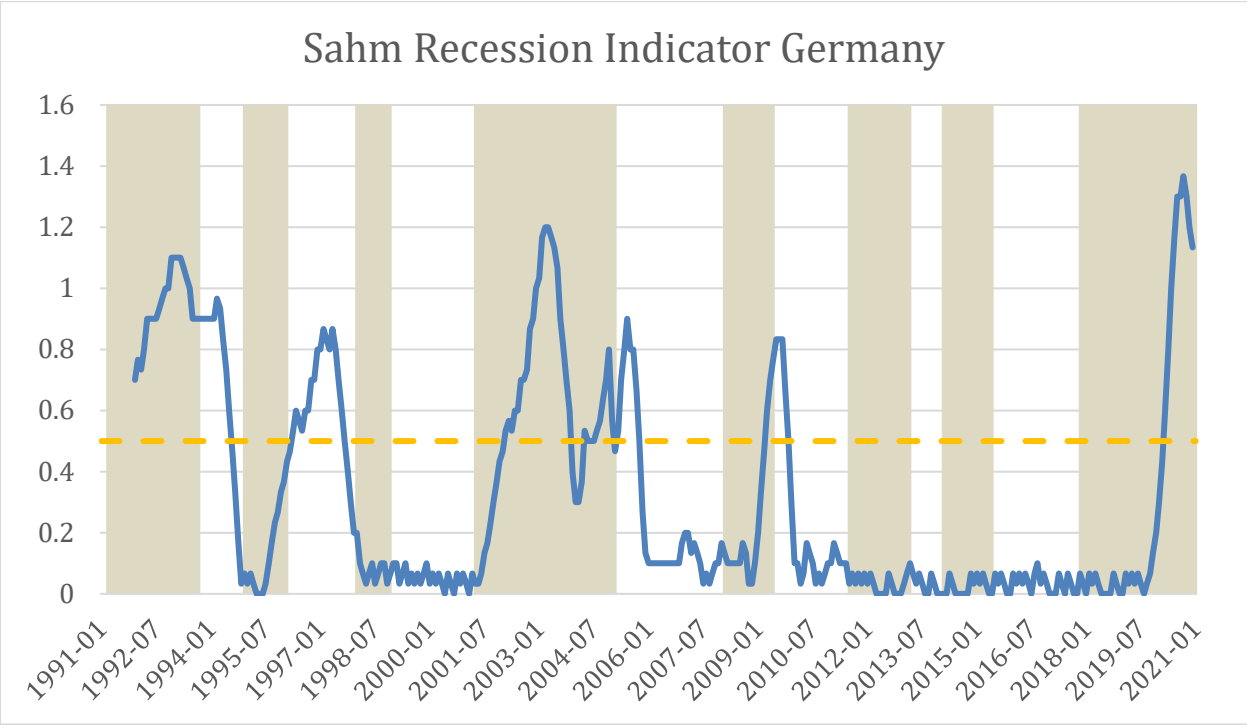
7) Below are data that replicate the Sahm indicator for several large European economies.

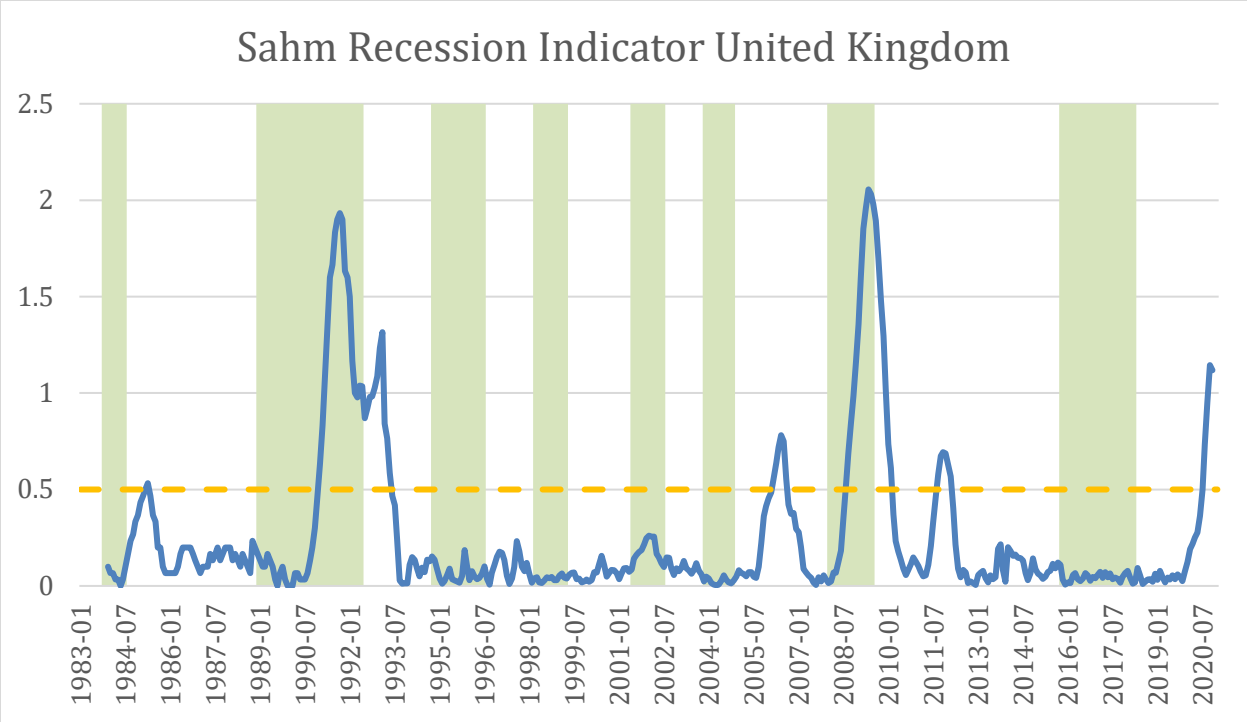
The recession indicators in the graphs come from the Organization for Cooperation and Development (OECD). The OECD dates its cycles using the deviation of GDP from its trend.

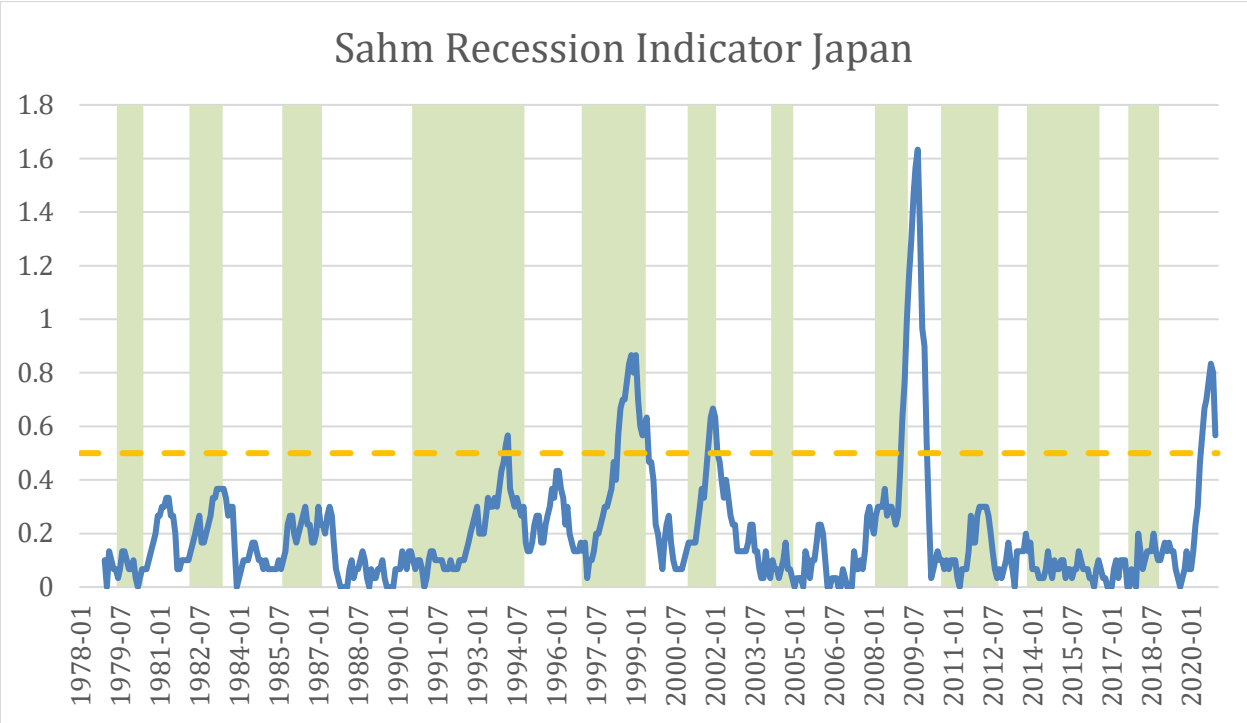
See http://www.oecd.org/sdd/leading-indicators/oecdcompositeleadingindicatorsreferenceturningpointsandcomponentseries.htm

How do you judge the success of the indicator in these countries? Can you explain why there might be differences across countries?


Sahm Recession Indicator France
2.5
2
1.5
1
0.5
0
1983-01
1984-07
1986-01
1987-07
1989-01
1990-07
1992-01
1993-07
1995-01
1996-07
1998-01
1999-07
2001-01
2002-07
2004-01
2005-07
2007-01
2008-07
2010-01
2011-07
2013-01
2014-07
2016-01
2017-07
2019-01
2020-07


Sahm Recession Indicator Germany
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
1991-01
1992-07
1994-01
1995-07
1997-01
1998-07
2000-01
2001-07
2003-01
2004-07
2006-01
2007-07
2009-01
2010-07
2012-01
2013-07
2015-01
2016-07
2018-01
2019-07
2021-01


Sahm Recession Indicator United Kingdom
2.5
2
1.5
1
0.5
0
1983-01
1984-07
1986-01
1987-07
1989-01
1990-07
1992-01
1993-07
1995-01
1996-07
1998-01
1999-07
2001-01
2002-07
2004-01
2005-07
2007-01
2008-07
2010-01
2011-07
2013-01
2014-07
2016-01
2017-07
2019-01
2020-07


Sahm Recession Indicator Japan
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
1978-01
1979-07
1981-01
1982-07
1984-01
1985-07
1987-01
1988-07
1990-01
1991-07
1993-01
1994-07
1996-01
1997-07
1999-01
2000-07
2002-01
2003-07
2005-01
2006-07
2008-01
2009-07
2011-01
2012-07
2014-01
2015-07
2017-01
2018-07
2020-01

8) The CBO scores legislation based on 10 year projections.

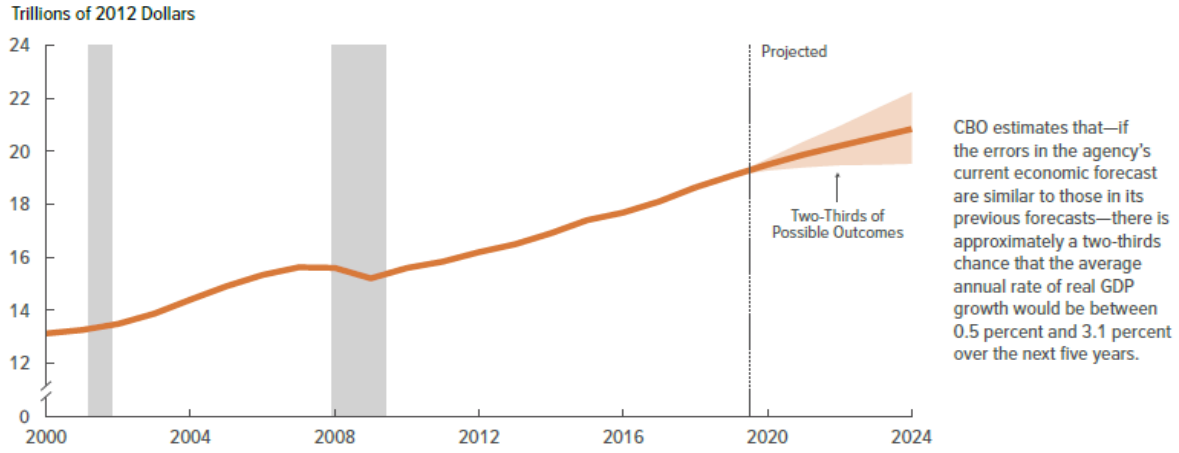
a) Is there any reasons why that would be a better horizon to use than say 3 to 4 years where the CBO might have a more informed view about the direction of the economy? What would be the downside of using the longer horizon?

b) In their projections, the CBO supposes that the economy grows slightly below potential. What is the rationale for doing that?

c) Below is an analysis from the CBO of the uncertainty regarding its estimates of potential GDP. Based on this analysis how reliable is their forecast of potential GDP?

Figure 2-8.

The Uncertainty of CBO's Projections of Real GDP



Sources: Congressional Budget Office; Bureau of Economic Analysis.

Real values are nominal values that have been adjusted to remove the effects of changes in prices.

The shaded area around CBO's baseline projection of real GDP illustrates the uncertainty of that projection. The area is based on the errors in CBO's one-, two-, three-, four-, and five-year projections of the average annual growth rate of real GDP for calendar years 1976 through 2018.

The value of real GDP for 2019 is CBO's estimate.

GDP = gross domestic product.

9) The Social Security Trustees make a baseline projection, as well as a high and low case projection. Their assumptions are shown below for the *annual* values of the five inputs that go into the real earnings calculation described in the case.

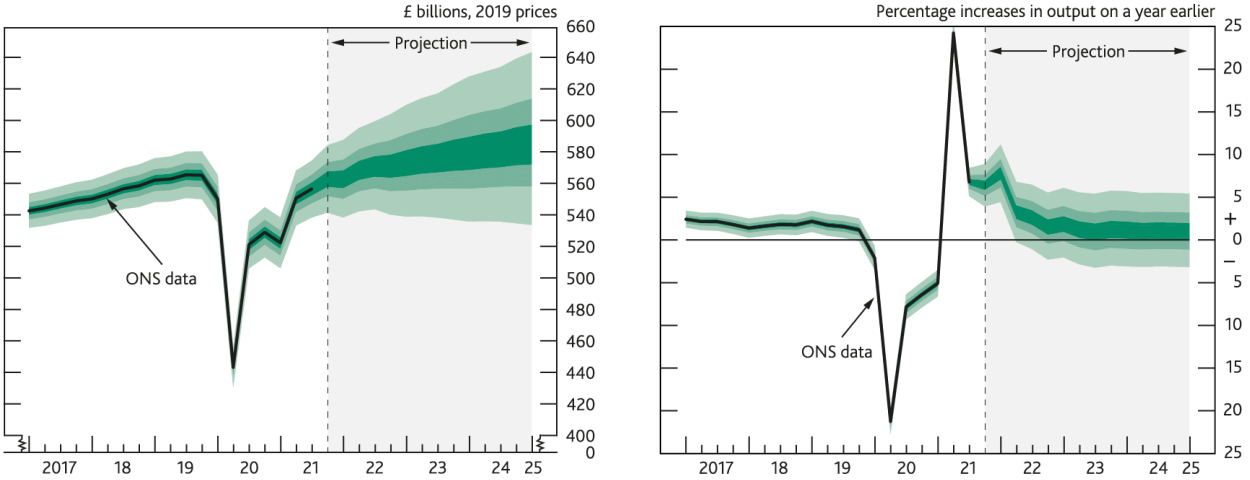
Longterm average annual percentage change (2028-2093)							
	Productivity	Benefit wedge	Hours worked	Labor's share	GDP deflator	CPI deflator	Real earnings
Baseline (trustees)	1.63	-0.05	-0.05	0	2.25	2.6	1.18
High (trustees)	1.93	0.03	0.05	0	2.95	3.2	1.76
Low (trustees)	1.33	-0.013	-0.15	0	1.55	2	0.72

Assume that the level of earnings was 100 at the start of the projection. Calculate the implied level 65 years in the future under in each of the scenarios. (Remember to compound the annual rates in this calculation.)

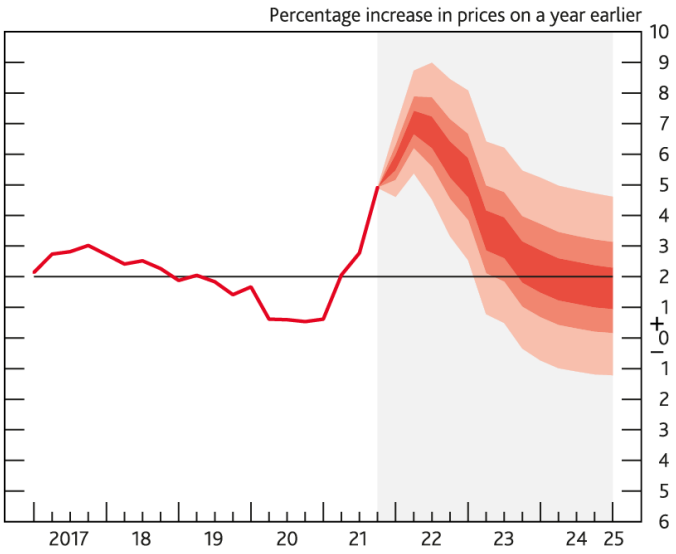
Recognizing the uncertainty over the inputs to the calculation, how confident can we be about the level of compensation in 65 years?

The case argues that the productivity assumption is the dominant driver of the results. What is the mathematical explanation for this claim and what is the economic explanation?

Exhibit 1: The Bank of England Fan Charts



The fan charts depict the probability of various outcomes for GDP and GDP growth. They have been conditioned on the assumptions in **Table 1.A** footnote (b). To the left of the vertical dashed line, the distribution reflects uncertainty around revisions to the data over the past. To aid comparability with the official data, it does not include the backcast for expected revisions, which is available from the 'Download the chart slides and data' link at *Monetary Policy Report* – February 2022. To the right of the vertical line, the distribution reflects uncertainty over the evolution of GDP and GDP growth in the future. If economic circumstances identical to today's were to prevail on 100 occasions, the MPC's best collective judgement is that the mature estimate of GDP (in **Chart 1.1**) or GDP growth (in **Chart 1.2**) would lie within the darkest central band on only 30 of those occasions. The fan chart is constructed so that outturns are also expected to lie within each pair of the lighter green areas on 30 occasions. In any particular quarter of the forecast period, GDP or GDP growth is therefore expected to lie somewhere within the fan on 90 out of 100 occasions. And on the remaining 10 out of 100 occasions GDP or GDP growth can fall anywhere outside the green area of the fan chart. Over the forecast period, this has been depicted by the light grey background. See the box on page 39 of the November 2007 *Inflation Report* for a fuller description of the fan chart and what it represents.



The fan chart depicts the probability of various outcomes for CPI inflation in the future. It has been conditioned on the assumptions in **Table 1.A** footnote (b). If economic circumstances identical to today's were to prevail on 100 occasions, the MPC's best collective judgement is that inflation in any particular quarter would lie within the darkest central band on only 30 of those occasions. The fan chart is constructed so that outturns of inflation are also expected to lie within each pair of the lighter red areas on 30 occasions. In any particular quarter of the forecast period, inflation is therefore expected to lie somewhere within the fans on 90 out of 100 occasions. And on the remaining 10 out of 100 occasions inflation can fall anywhere outside the red area of the fan chart. Over the forecast period, this has been depicted by the light grey background. See the box on pages 48–49 of the May 2002 *Inflation Report* for a fuller description of the fan chart and what it represents.

Source: Bank of England Monetary Policy Report February 2022

Exhibit 2 Specificity and Sensitivity in Forecasting Binary Outcomes

Actual Outcome	**Predict Event Does Not Occur**	**Predict Event Does Occur**
Event Does Not Occur	**A: True Negative**	**B: False Positive**
Event Occurs	**C: False Negative**	**D: True Positive**

A good model maximizes A and D while minimizing B and C

"Sensitivity" ≡ d/(c+d) how many true positives are predicted to be so

"Specificity" ≡ a/(a+b) how many negatives are correctly predicted

A good model has high sensitivity and low false negatives (1-specificity)

Exhibit 3 Hypothetical Illustration

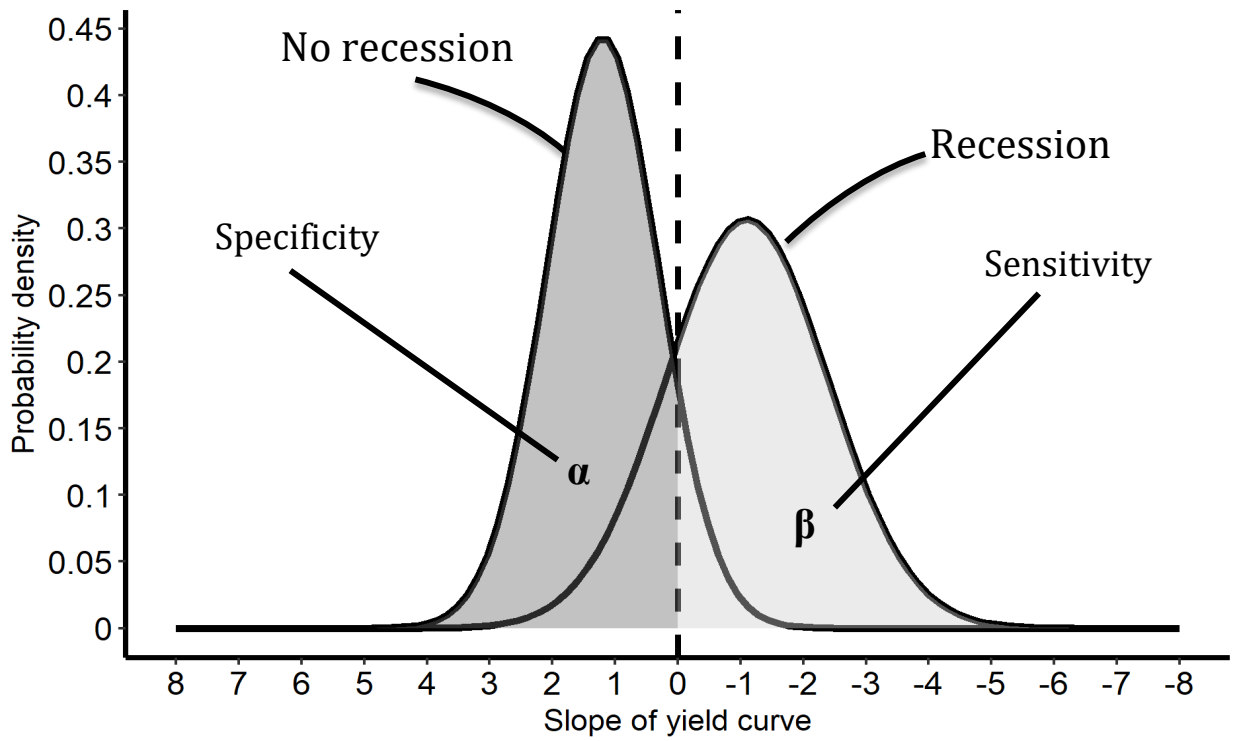


Exhibit 4 Model Comparisons

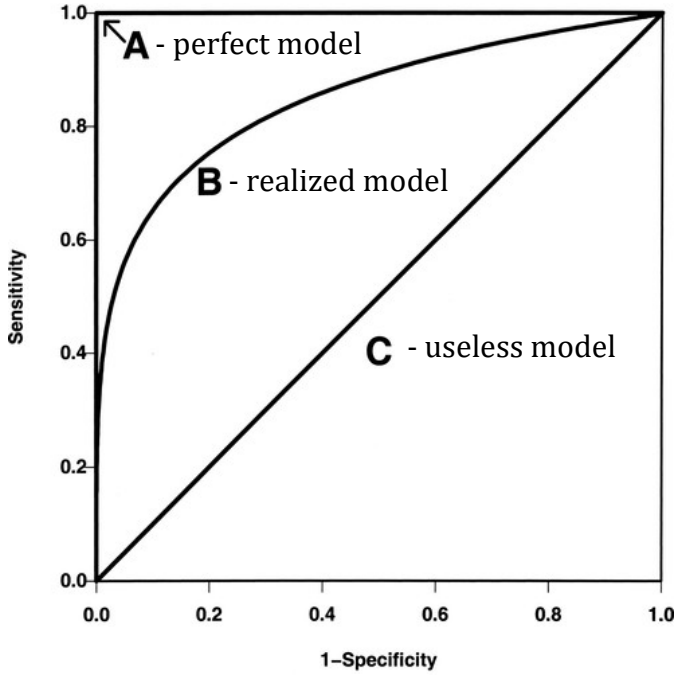
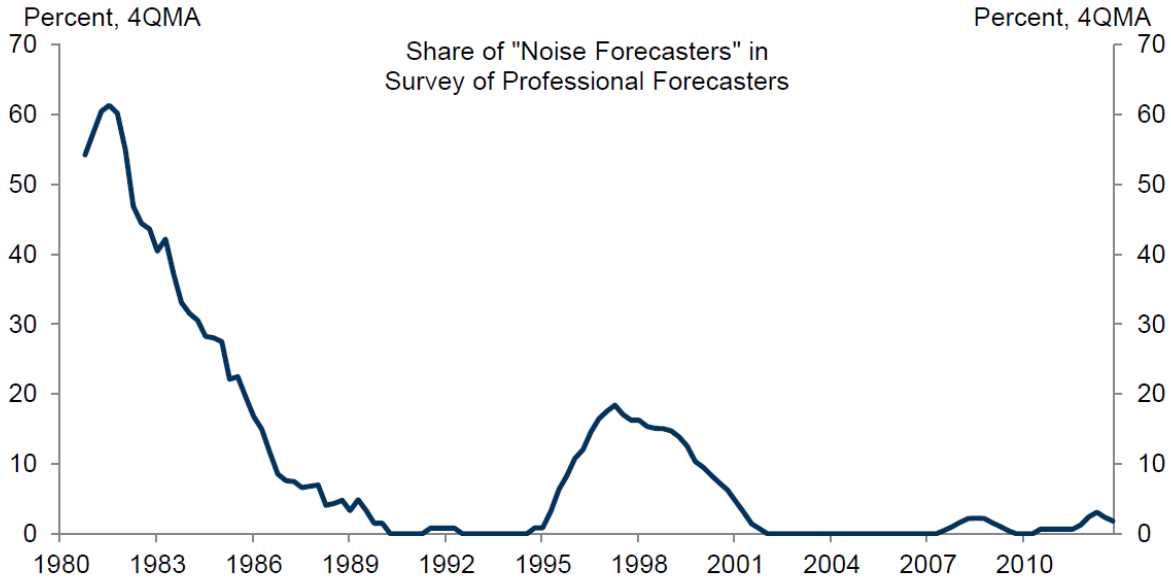


Exhibit 5: Percentage of Submitters for Survey of Professional Forecasters That are Consistently Overly Optimistic or Pessimistic About GDP Growth



Source: Federal Reserve Bank of Philadelphia, Goldman Sachs Global Investment Research

Exhibit 6: Comparison of the Accuracy of Official Sector and Private Sector Forecasts

Real GDP Forecast Errors

	Horizon	Mean	Standard Error	Observations	Countries
Official Minus Consensus	1Y	0.163***	(0.0276)	323	26
Official Forecast Error	1Y	0.465	(0.489)	323	26
Concensus Forecast Error	1Y	0.302	(0.477)	323	26
Official Minus Consensus	2Y	0.135**	(0.0484)	278	23
Official Forecast Error	2Y	1.244	(0.738)	278	23
Concensus Forecast Error	2Y	1.11	(0.736)	278	23

Driscoll-Kraay Standard Errors with 2 year lag. Only includes countries with at least 6 years of data.

Budget Balance Forecast Errors

	Horizon	Mean	Standard Error	Observations	Countries
Official Minus Consensus	1Y	0.261***	(0.0781)	123	12
Official Forecast Error	1Y	0.341	(0.335)	123	12
Concensus Forecast Error	1Y	0.0802	(0.340)	123	12
Official Minus Consensus	2Y	0.478***	(0.0863)	95	10
Official Forecast Error	2Y	1.060*	(0.541)	95	10
Concensus Forecast Error	2Y	0.582	(0.548)	95	10

Driscoll-Kraay Standard Errors with 2 year lag. Only includes countries with at least 6 years of data.

Note: source, Tables 3 and 4 from Jeffrey Frankel and Jesse Schrager, "Bias in Official Fiscal Forecasts: Can Private Forecasts Help?" NBER Working Paper No. 22349, June 2016

Exhibit 7: NBER Business Cycle Dating

The National Bureau's Business Cycle Dating Committee maintains a chronology of U.S. business cycles. The chronology identifies the dates of peaks and troughs that frame economic recessions and expansions. A recession is the period between a peak of economic activity and its subsequent trough, or lowest point. Between trough and peak, the economy is in an expansion. Expansion is the normal state of the economy; most recessions are brief. However, the time that it takes for the economy to return to its previous peak level of activity or its previous trend path may be quite extended. According to the NBER chronology, the most recent peak occurred in February 2020, ending a record-long expansion that began after the trough in June 2009.

The NBER's traditional definition emphasizes that a recession involves a significant decline in economic activity that is spread across the economy and lasts more than a few months. In our modern interpretation of this definition, we treat the three criteria—depth, diffusion, and duration—as at least somewhat interchangeable. That is, while each criterion needs to be met individually to some degree, extreme conditions revealed by one criterion may partially offset weaker indications from another. For example, in the case of the February 2020 peak in economic activity, the committee concluded that the subsequent drop in activity had been so great and so widely diffused throughout the economy that, even if it proved to be quite brief, the downturn should be classified as a recession.

In choosing the dates of business-cycle turning points, the committee follows standard procedures to assure continuity in the chronology. Because a recession must influence the economy broadly and not be confined to one sector, the committee emphasizes economy-wide measures of economic activity. It views real gross domestic product (GDP) as the single best measure of aggregate economic activity. This concept is measured two ways by the U.S. Bureau of Economic Analysis (BEA)—from the product side and from the income side. Because the two measures have strengths and weaknesses and differ by a statistical discrepancy, the committee considers real GDP and real gross domestic income (GDI) on an equal footing. It also considers carefully total payroll employment as measured by the Bureau of Labor Statistics (BLS).

The traditional role of the committee is to maintain a monthly chronology of business cycle turning points. Because the BEA figures for real GDP and real GDI are only available quarterly, the committee considers a variety of monthly indicators to determine the months of peaks and troughs. It places particular emphasis on two monthly measures of activity across the entire economy: (1) personal income less transfer payments, in real terms, which is a monthly measure that includes much of the income included in real GDI; and (2) payroll employment from the BLS. Although these indicators are the most important measures considered by the committee in developing its monthly business cycle chronology, it does not hesitate to consider other indicators, such as real personal consumption expenditures, industrial production, initial claims for unemployment insurance, wholesale-retail sales adjusted for price changes, and household employment, as it deems valuable. There is no fixed rule about which other measures contribute information to the process or how they are weighted in the committee's decisions.

The committee's approach to determining the dates of turning points is retrospective. It waits 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, it waits until the committee members are confident that a recession has occurred, even in the event that activity begins to rise again immediately. As a result, the committee tends to wait to identify a peak until a number of months after it has actually occurred.

Exhibit 8: Predicting Years where GDP Growth is Negative

		Consensus Forecasts: Apr [t-1]			Consensus Forecasts: Oct [t]		
		Non-recession	Recession	Total	Non-recession	Recession	Total
Actual	Non-recession	1145	8	1153	1120	33	1153
	Recession	148	5	153	35	118	153
	Total	1293	13	1306	1155	151	1306

Source: IMF World Economic Outlook and Consensus Forecasts

Reproduced from Zidong An, João Tovar Jalles, and Prakash Loungani "How Well Do Economists Forecast Recessions?" IMF Working Paper 18/39.

Exhibit 9: Predicting Forecast Revisions

Dependent Variable:	Consensus			IMF		
Revision	All	Advanced	Emerging	All	Advanced	Emerging
Lagged Revision	0.35***	0.29***	0.38***	0.21***	0.09*	0.27***
	(0.04)	(0.04)	(0.06)	(0.04)	(0.05)	(0.05)
Constant	0.06***	0.03*	0.07***	0.05**	0.02	0.05**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)
N. of Obs.	1306	639	667	1306	639	667
R-sq	0.18	0.12	0.21	0.08	0.01	0.13

Source: IMF World Economic Outlook, Consensus Forecasts, and authors' estimates.

Notes: The dependent variable is the forecast revision made between Oct[t] and Apr[t]. The independent variables are the forecast revision made between Apr[t] and Oct[t-1], dummy variable for recession. Country fixed effects are included but omitted for reasons of parsimony. Robust standard errors are reported in parentheses. *, **, *** denote statistical significance at the 10, 5 and 1 percent levels, respectively. Reproduced from Zidong An, João Tovar Jalles, and Prakash Loungani "How Well Do Economists Forecast Recessions?" IMF Working Paper 18/39.

Exhibit 10: Slope of the Yield Curve and the Business Cycle

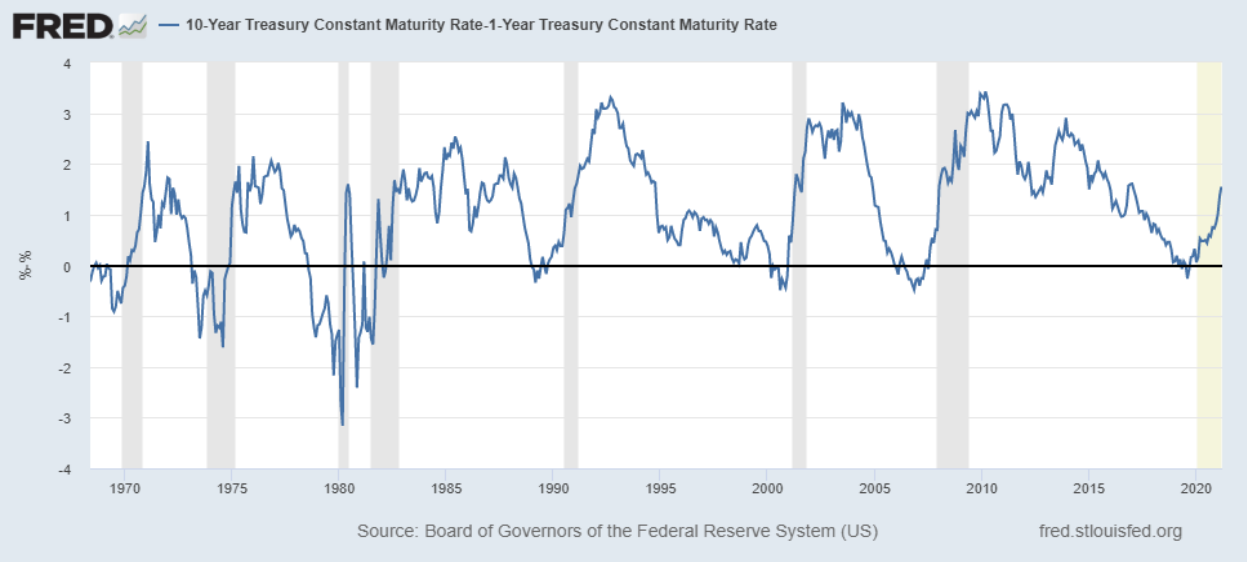


Exhibit 11: Sahm Recession Indicator

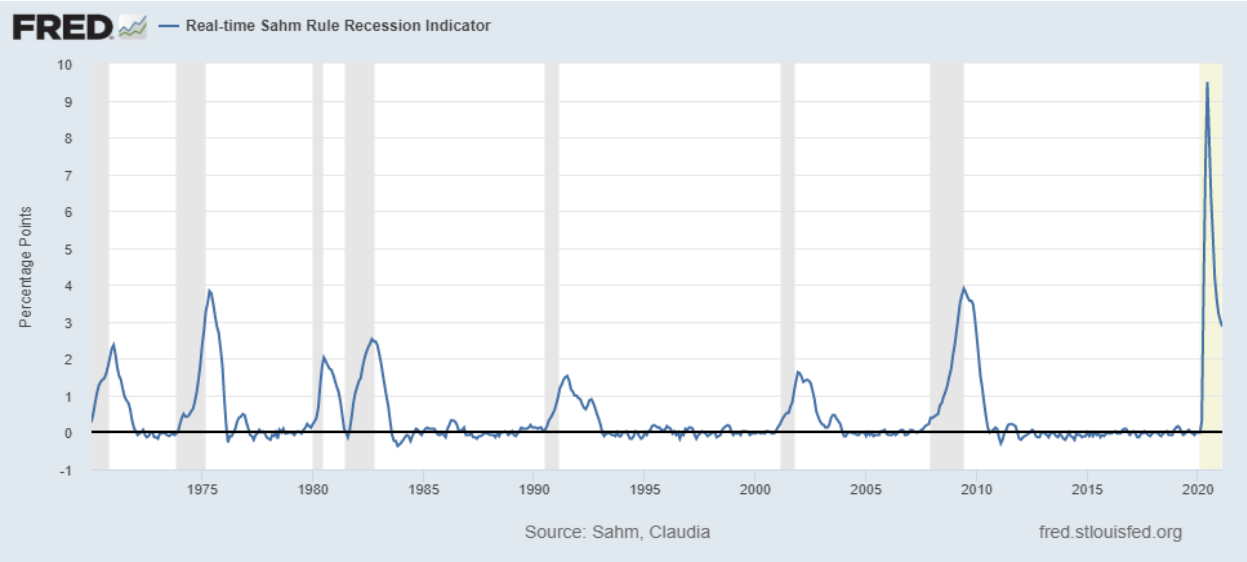


Exhibit 12: Sahm Indicator with Different Thresholds

Probability of a Recession by Sahm Recession Indicator, 1970–2019

Excluding Months Between NBER Announcements of Business Cycle Peaks and Troughs

Sahm recession indicator (percentage point increase in unemployment rate)	Recession now	Recession in 3 months	Recession in 6 months	Recession in 12 months	Recession in 24 months	Number of months
Less than 0	1%	2%	5%	10%	25%	203
0 to 0.09	2%	5%	8%	20%	39%	122
0.10 to 0.19	2%	13%	23%	33%	37%	52
0.20 to 0.29	11%	33%	33%	39%	39%	18
0.30 to 0.39	40%	40%	40%	40%	40%	10
0.40 to 0.49	76%	76%	76%	76%	76%	17
Greater than or equal to 0.50	**97%**	**97%**	**97%**	**97%**	**97%**	**30**
Probability of a recession at any unemployment rate	12%	15%	19%	25%	38%	452

Source: Current Population Survey, Bureau of Labor Statistics 1969–2019; Sahm 2019; authors' calculations.
Note: Calculations use real-time estimates of the three-month average unemployment rate. Values reflect data through May 2019. We exclude the months between when NBER announces the beginning and end of recessions (140 months out of 592 total in the sample period); at those times, it is already understood that a recession has occurred. For business cycles before NBER began formally announcing peak and trough dates in 1979, we use the average time between recession start and peak announcement, and recession end and trough announcement. When including all 592 months in the sample, the probability in any given month that there is a recession is 15 percent, compared to the 12 percent reported in the table. Future recession probabilities include the current month, plus the number of stated months in the future; e.g., "recession in three months" include the current month plus the following three months.

THE HAMILTON PROJECT
BROOKINGS

Exhibit 13: Congressional Budget Office January 2021 10 Year Projection

Table 2-1.

CBO's Economic Projections for Calendar Years 2021 to 2031

	2020	2021	2022	2023	Annual Average 2024–2025	Annual Average 2026–2031
	Percentage Change From Fourth Quarter to Fourth Quarter					
Gross Domestic Product						
Real[a]	-2.5	3.7	2.4	2.3	2.2	1.6
Nominal	-1.2	5.6	4.5	4.3	4.4	3.8
Inflation						
PCE price index	1.2	1.7	1.9	1.9	2.1	2.1
Core PCE price index[b]	1.4	1.5	1.9	1.9	2.1	2.1
Consumer price index[c]	1.2	1.9	2.2	2.3	2.4	2.4
Core consumer price index[b]	1.6	1.5	2.2	2.3	2.4	2.4
GDP price index	1.3	1.9	2.0	2.0	2.1	2.1
Employment Cost Index[d]	2.8	2.3	2.8	3.0	3.2	3.3
	Fourth-Quarter Level (Percent)					
Unemployment Rate	6.8	5.3	4.9	4.6	4.0[e]	4.3[f]
	Percentage Change From Year to Year					
Gross Domestic Product						
Real[a]	-3.5	4.6	2.9	2.2	2.3	1.7
Nominal	-2.3	6.3	4.9	4.2	4.4	3.8
Inflation						
PCE price index	1.2	1.6	1.8	1.9	2.0	2.1
Core PCE price index[b]	1.4	1.5	1.8	1.9	2.0	2.1
Consumer price index[c]	1.3	1.9	2.1	2.3	2.3	2.4
Core consumer price index[b]	1.7	1.6	2.1	2.3	2.4	2.4
GDP price index	1.2	1.6	1.9	2.0	2.1	2.1
Employment Cost Index[d]	2.9	2.1	2.6	2.9	3.1	3.3
	Annual Average					
Unemployment Rate (Percent)	8.1	5.7	5.0	4.7	4.2	4.1
Labor Force Participation Rate (Percent)[g]	61.7	61.9	62.1	62.0	61.9	61.2
Payroll Employment (Monthly change, in thousands)[h]	-765	521	145	145	135	40
Interest Rates (Percent)						
Three-month Treasury bills	0.4	0.1	0.1	0.2	0.4	1.7
Ten-year Treasury notes	0.9	1.1	1.3	1.5	2.0	3.0
Tax Bases (Percentage of GDP)						
Wages and salaries	44.8	44.0	43.9	43.9	43.9	43.6
Domestic corporate profits[i]	7.6[j]	7.9	7.5	7.7	8.2	8.0
Current Account Balance (Percentage of GDP)[k]	-2.8[j]	-2.9	-2.4	-2.0	-2.0	-2.2

Data sources: Congressional Budget Office; Bureau of Economic Analysis; Bureau of Labor Statistics; Federal Reserve. See www.cbo.gov/publication/56970#data.

Exhibit 14: Summary of the Budget Implications of the Affordable Care Act

TABLE 1. PRELIMINARY ESTIMATE OF THE EFFECTS ON THE DEFICIT OF H.R. 3962, THE AFFORDABLE HEALTH CARE FOR AMERICA ACT, AS INTRODUCED ON OCTOBER 29, 2009

	By Fiscal Year, in Billions of Dollars											
	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2010-2014	2010-2019
NET CHANGES IN THE DEFICIT FROM INSURANCE COVERAGE PROVISIONS [a]												
Effects on the Deficit	*	1	2	57	93	123	137	148	160	173	153	894
NET CHANGES IN THE DEFICIT FROM OTHER PROVISIONS AFFECTING DIRECT SPENDING [b]												
Effects on the Deficit of Changes in Outlays	7	17	-16	-25	-52	-51	-54	-72	-85	-96	-69	-426
NET CHANGES IN THE DEFICIT FROM OTHER PROVISIONS AFFECTING REVENUES [c]												
Effects on the Deficit of Changes in Revenues [d]	*	-33	-35	-57	-62	-67	-72	-77	-82	-86	-188	-572
NET CHANGES IN THE DEFICIT [a]												
Net Increase or Decrease (-) in the Budget Deficit	6	-15	-49	-25	-21	5	11	-1	-7	-9	-104	-104
On-Budget	6	-15	-49	-27	-23	4	10	-3	-8	-10	-108	-115
Off-Budget [e]	*	*	*	2	2	2	2	1	1	1	4	11

Sources: Congressional Budget Office and staff of the Joint Committee on Taxation (JCT).

Notes: Positive numbers indicate increases in the deficit, and negative numbers indicate reductions in the deficit.

Components may not sum to totals because of rounding; * = between $0.5 billion and -$0.5 billion.

a. Does not include effects on spending subject to future appropriations.

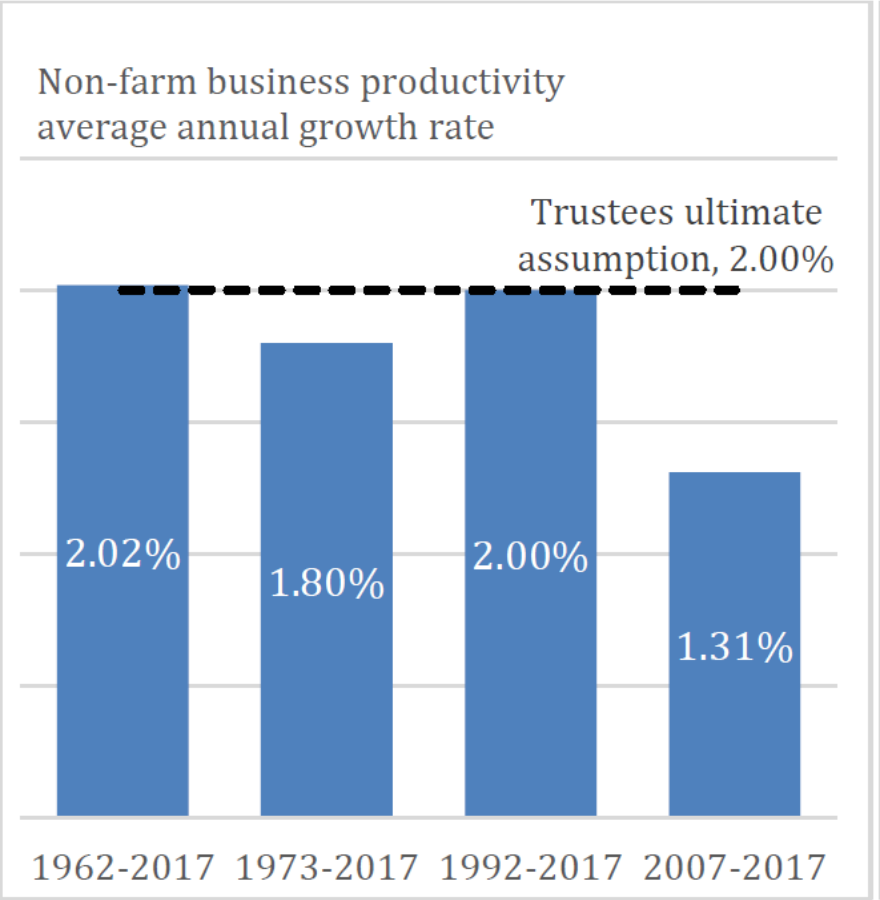
b. These estimates reflect the effects of interactions between insurance coverage provisions and other Medicare and Medicaid provisions. In addition, CBO has included $33 billion of spending over the 2010–2019 period for public health, prevention, and wellness provisions in these direct spending totals, as directed by the Committee on the Budget, even though that spending would be subject to future appropriation action.

c. The changes in revenues include effects on Social Security revenues, which are classified as off-budget.

d. The 10-year figure of $572 billion includes $558 billion in revenues from tax provisions (estimated by JCT) and $14 billion in additional revenues from certain provisions affecting Medicare, Medicaid, and other programs (estimated by JCT and CBO). (For JCT's estimates see JCX-43-09.)

e. Off-budget effects include changes in Social Security spending and revenues.

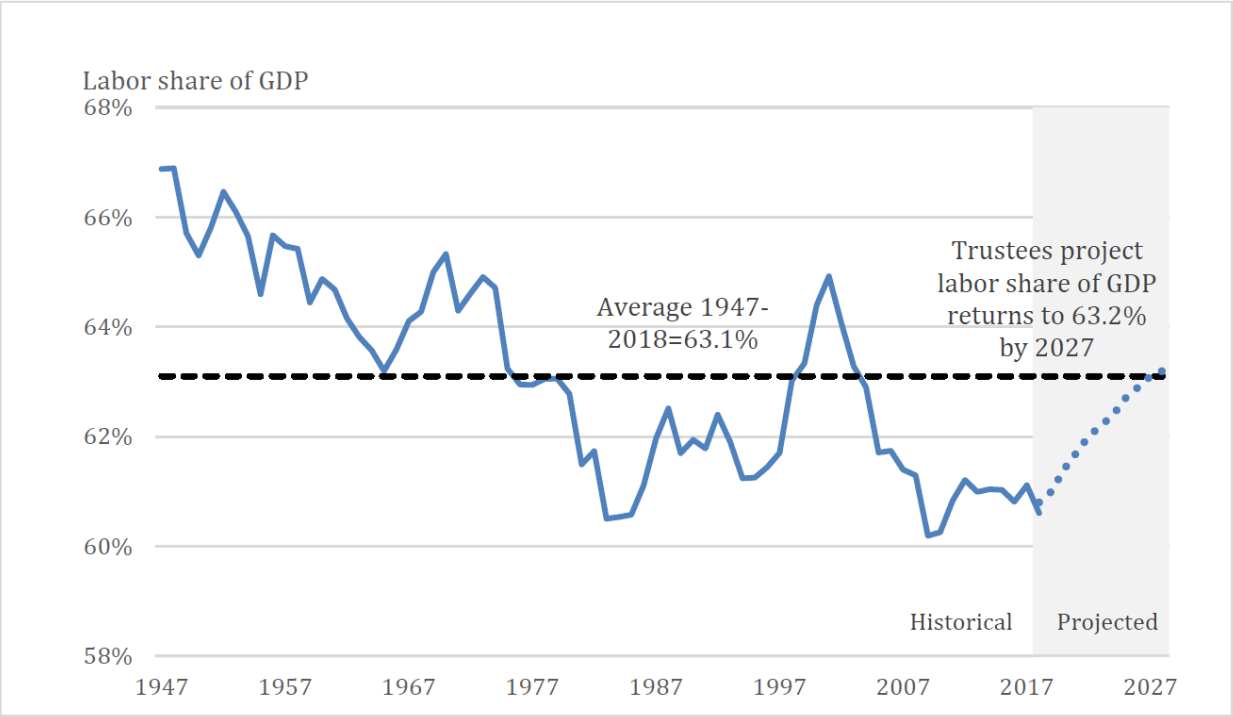
Exhibit 15: Productivity Forecasts



	Comparison of productivity growth assumptions with outside forecasts: average 2028-2048				
	CBO* (2019a,b)	IHS Markit	Moody's Analytics	Trustees 2019	Panel
Non-Farm Business	1.75	1.50	1.58	2.00	1.90
Economy-wide	1.49	--	--	1.63	1.55

*Non-farm business from CBO 2019a; Economy-wide from CBO 2019b; OCACT (2019)

Exhibit 16: Labor's Share Forecast



Historical interval	Labor share of GDP	Adjusted for depreciation	Adjusted for Housing
1958-2018	62.6%	61.4%	60.6%
1968-2018	62.3%	61.4%	60.6%
1978-2018	61.8%	61.2%	60.6%
1988-2018	61.9%	61.3%	61.0%
1998-2018	61.9%	61.6%	61.3%
2008-2018	60.9%	60.9%	60.8%

Exhibit 17: Benefit Wedge Forecast

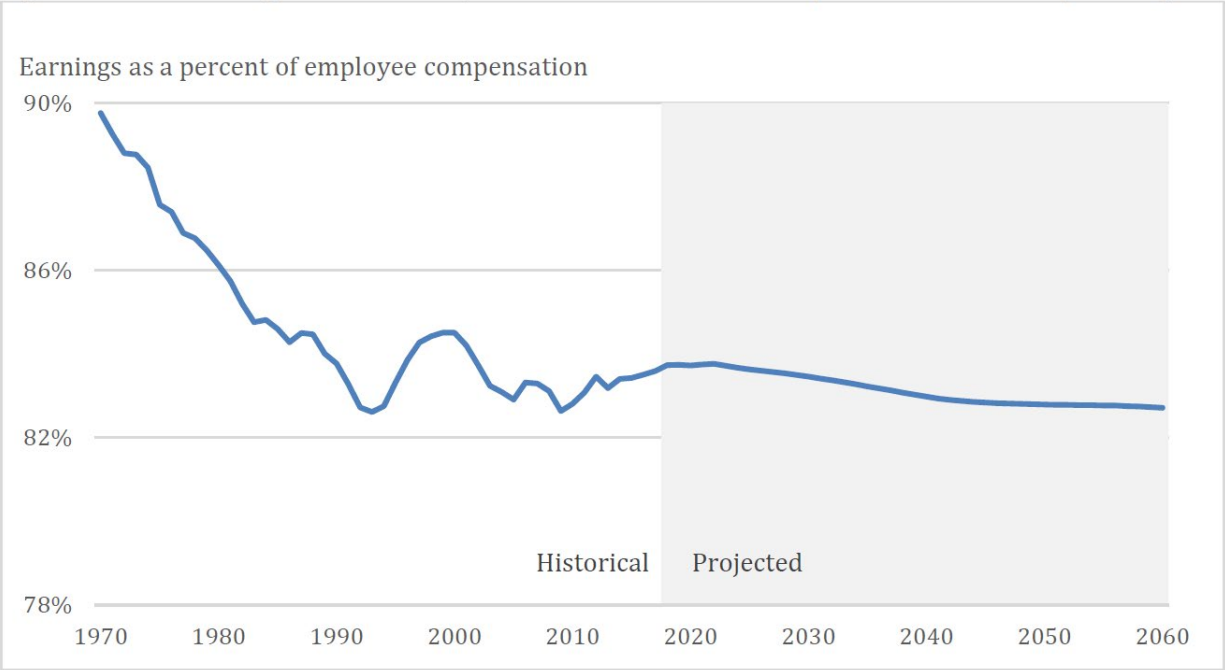


Exhibit 18: Hours Worked Forecast

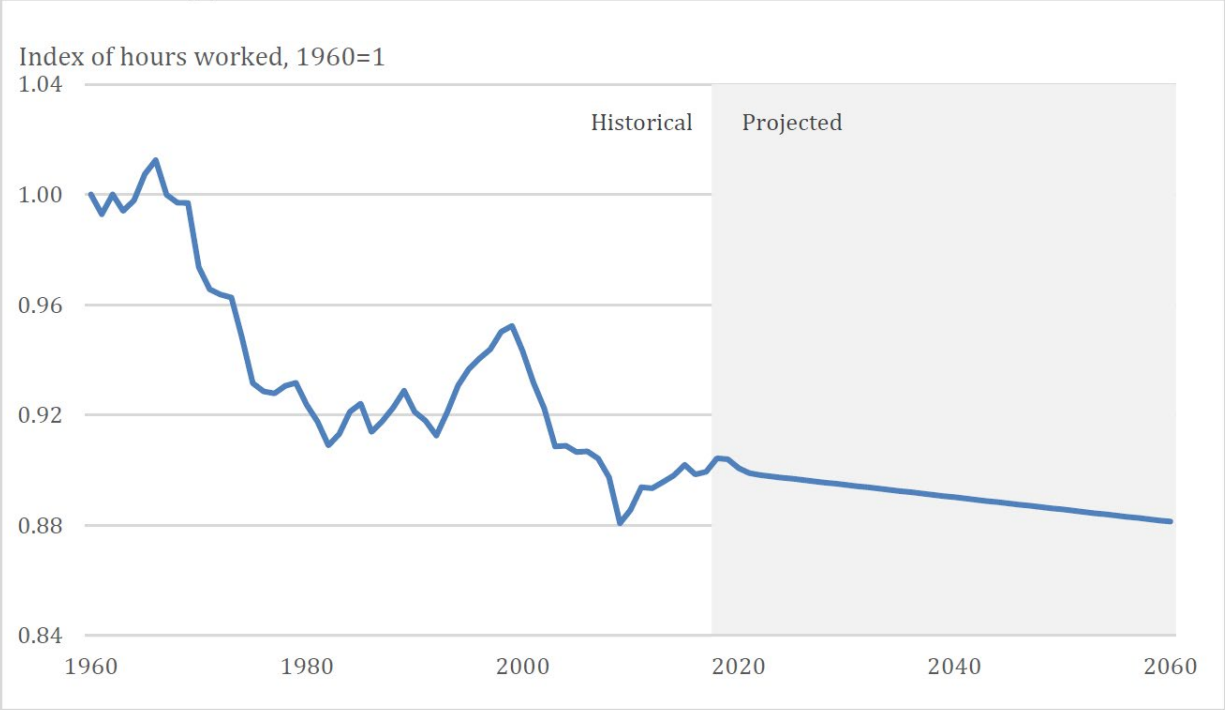
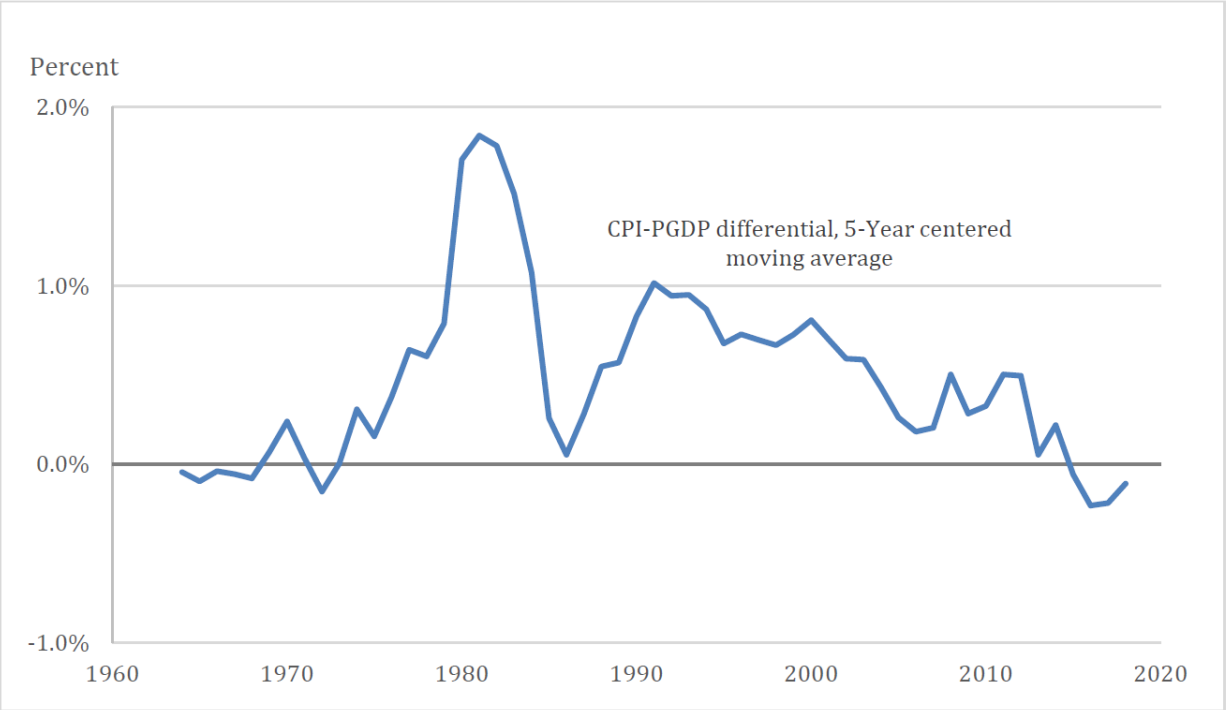



Exhibit 19: Forecast in Relative Price Changes