Potential Output:
Little Explanation for a Big Number

Alex Williams
Research Analyst,
Employ America

Summary

Commentary about the Biden administration's proposed fiscal relief policies has relied heavily on estimates of the economy's potential output. However, few commentators or policymakers look under the hood to check how these estimates are calculated. Oftentimes, estimates of potential output — the maximum inflation-adjusted dollar value that the economy can sustain without runaway inflation — are produced by simply fitting a trend to past values of real GDP. More complex models like those used by the Congressional Budget Office rely on statistical residuals[1] like total factor productivity, and unsubstantiated assertions about the labor market.[2] In the face of the tremendous complexity of a modern industrial economy, assertions about potential output that rely on methodologies this fragile should be met with deeper skepticism.

That the economy faces constraints is obvious to everyone. However, the obviousness of this fact often masks the deeply suspect methodology used to calculate where these constraints may bind. When determining the necessary scale of fiscal relief, most estimates of potential output — and, accordingly, the scale of the output gap — are only useful for producing a very large number. The number itself tells us little about the economy or the needs of policy. To do this would require much more robust estimation methods, grounded in granular studies of individual production processes. Until these more robust methods are implemented, direct estimates of potential output will provide poor metrics for calibrating fiscal policy, and should be heavily discounted by commentators and policymakers alike. Instead of the output gap, fiscal relief and further stimulus should be calibrated with reference to achieving specific labor market outcomes, among other policy objectives.

Table of Contents

The Output Gap

The output gap has played a substantial role in economic[3] and policy discourse[4] around the Biden reconciliation bill. The worry expressed by most participants is that a fiscal bill that is substantially larger than the current output gap will lead to an overheated, inflationary economy. The logic is that — assuming the economy has some in-principle maximum capacity — if there is too much fiscal spending, then the economy will reach that maximum capacity and overheat. The problem is, the economy is an incredibly complex system with trillions of moving parts. Condensing this system to a single indicator always runs the risk of cutting out so much information that any argument based on that indicator is bound to be incomplete or misleading.

Ultimately, the output gap is useful as a ready-to-hand metric not because of how well it captures and communicates economic reality, but because it is incredibly effective at producing a dollar amount that is of a similar order of magnitude to most stimulus packages. That the dollar amount it produces tells us little about the economy is less important than the fact that it provides a "price anchor" for the discourse.

If the goal of economic and policy discourse is to create better policy, it would behoove us to have a quick discussion of the methodology used to produce output gap measurements. The idea that the economy has a maximum level of output is so obviously, intuitively true that most gloss over the methods by which that maximum level is estimated.

Before getting into the weeds of how the output gap should and shouldn't be used in policy conversations, it's important to understand how it is measured and calculated. To estimate the output gap, the first step is to estimate the inflation-adjusted dollar value of the economy's potential output. The Federal Reserve Bank of St. Louis has helpfully summarized the six most common approaches here.[5]

These six approaches follow two basic strategies: one a kind of technical trend analysis[6] that involves interpolating and extrapolating from data along a neat linear or exponential function, and the other a more "theory-laden" approach that centers on deviations from a "Natural Rate of Unemployment" and productivity growth. We will treat each approach described in the Fed's literature review above in turn, from simplest to most complex.

The GDP Trendline Approach

Most of the approaches described in the Fed's literature review above are simple trend-fitting strategies for GDP data — observe data, draw a line/curve, extrapolate said line/curve into the future.

Deterministic trend approaches are bafflingly simple: just take the last few years of real GDP and draw a line through them. That line then gives potential output in future years. If that feels too simplistic, some economists and agencies use different detrending and filtering methods to "see through" the volatility of business cycles and identify the "acyclical truth" within macroeconomic data. However, the core method of predicting potential output solely on the basis of past GDP values — and the cost of that dramatic oversimplification — are common to all. These trend fitting approaches are easy to use, but offer little insight about the level of potential output the economy can obtain without excessive inflation.

The heart of the problem is that GDP is not a simple or empirically observable quantity. Instead, it is calculated[7] using an estimate of dollars spent for all particular goods or services (nominal GDP), deflated by an estimate of the price of each particular good or service. Worse, much of nominal GDP is now composed of imputed measures;[8] housing imputations are better-known but imputed financial services are also growing in share and follow a highly complex methodology for estimating prices.[9]

To get a sense of just how sensitive the price estimates used to derive the GDP deflator are to methodological choices, consider the example of unlimited cellular data plans. As this article[10] explains, the adoption of unlimited data plans by a number of providers was incorporated as a decrease in the relative cost of a cell phone. Because of how the index was constructed, this small shift had an outsize effect, representing the difference between 1% and 2% annualized CPI inflation. While it makes sense to think of paying the same price to get more in terms of quality and output as a "real gain," deriving a quality-neutral price can introduce substantial fragility into the price deflator. Against annual real GDP growth in the low single digits, a variation this size in the price deflator matters.

However, this fragility extends far beyond cell phone plans. In part due to data availability constraints, hedonic adjustments are not applied consistently[11] across all components of output estimates. In certain segments, like information processing equipment, there has been a concerted effort to fold in the pace of improvement in the number of transistors that fit in integrated circuits (Moore's Law).[12] In other sectors like prescription drugs, healthcare services, or software, hedonic adjustments remain forthcoming.

This is not to say that real GDP is a useless indicator. It does a good job indicating the general direction of the economy; the evolution of the growth rate of real GDP is less sensitive to these issues than the outright level of real GDP. The problem is when it is taken as a basis for calculating precise dollar amounts beyond which inflation is inevitable. As a measurement, it is so distant from the actual production processes and labor market dynamics that cause inflation or increase capacity that purely real GDP-based estimates of potential output should be heavily discounted by policymakers and commentators.
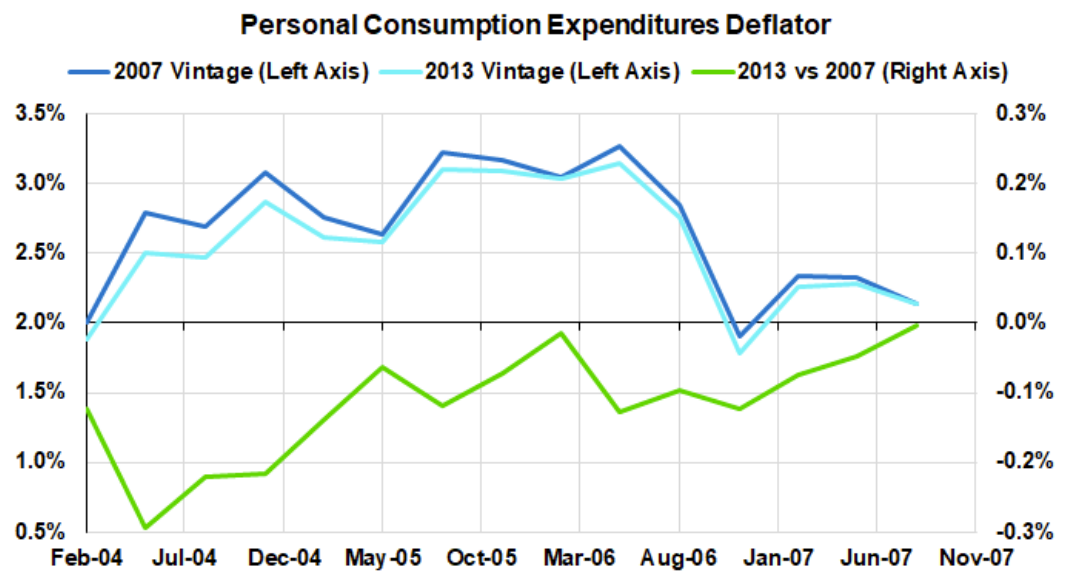
Beyond these measurement issues, relying on past levels of GDP to estimate present potential output embeds an additional assumption that the economy has spent as much time above as below its fully employed capacity in the recent past. Intuitively, the time spent above potential output should coincide with inflation above the then-symmetric 2% objective. The data on inflation, growth, and labor utilization over the past two decades show, to even an untrained observer, that this has not been the case.

This presumption also makes most estimates of potential look ridiculous in retrospect. This chart from a Fed piece[13] critical of the idea of potential output in 2014 makes the point quite succinctly:
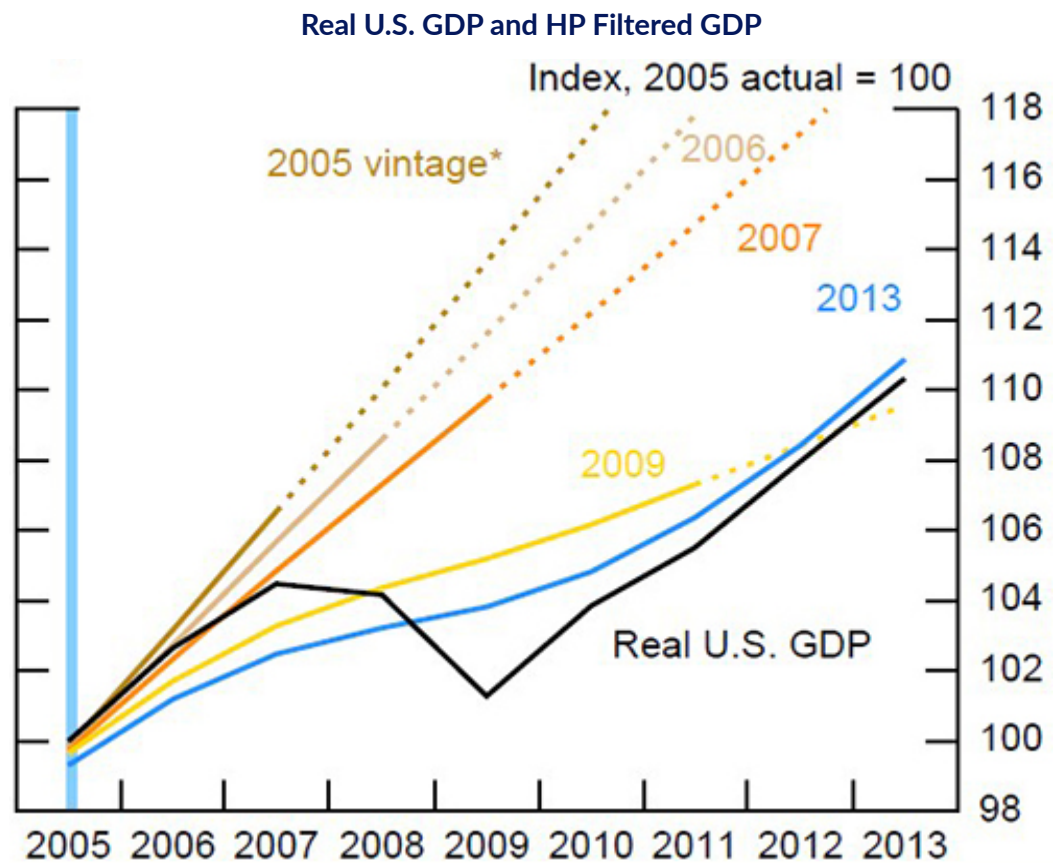
Source: OECD Economic Outlook, various vintages of actual and projected. Trends are calculated by applying an HP-filter to vintages of OECD GDP series. The trend is then extrapolated forward using the growth rate implied by the last year of HP-filtered data. The level of actual GDP is indexed to 100 in 2005.

To an observer in 2005, 2007 GDP was substantially below potential. However, to an observer in 2013, 2007 GDP was markedly above potential. One would expect — if exceeding potential output means higher inflation — that inflation readings would see upward revisions commensurate with potential output's downward revisions.

Personal Consumption Expenditures Deflator

— 2007 Vintage (Left Axis) — 2013 Vintage (Left Axis) — 2013 vs 2007 (Right Axis)

Source: BEA, ALFRED Archival Series

Clearly, this was not the case. In the 2013 revisions, core PCE was consistently revised downwards alongside potential output. Either output was not above potential in 2007, or there is no path from exceeding potential output — by two points, even! — to higher inflation.

Potential output is not some underlying fact about the economy that can be measured using only real GDP. At best, trendline-based estimates represent forecasts of likely future values, not upper limits on levels that can be attained without inflation. Worse still, treating measurements of potential output as hard limits, even as those measurements drift steadily downward, is tantamount to giving up on a strong economy in the long run. Deciding to abandon economic growth on the basis of a single filtered-trend indicator should be seen as an unforgivable abdication of political and economic responsibility.

The CBO Approach: Total Factor Productivity

The CBO's estimate of potential output is more widely used in policy circles than the approaches described above, and has its own distinct issues. Per the Fed's reconstruction[14] of their model, potential output is estimated by estimating deviations of unemployment from U* (the so-called Natural Rate of Unemployment), trend growth in total factor productivity, and trend growth in the capital stock across five sectors. Dealing with this model will take this section and the next, treating first the issues with productivity, then the issues with U*.

Total factor productivity (TFP) plays a key role in the CBO's estimate of potential output. It proxies increases in output that cannot be directly attributed to increases in the amount of capital and labor employed in production. Ideally, this is meant to stand in for technological, managerial, or process improvements, but in reality, it is calculated as a statistical residual.

Since it represents increases in output that don't come from increases in any other measured variable, TFP cannot be directly observed or measured, only derived from other estimates which each have their own measurement and specification issues. Estimates of TFP are derived from real GDP after adjusting for changes in labor and capital. As such, all of the problems described above with the precision of real GDP figures are embedded in estimates of TFP, while the methodology for proxying TFP introduces new, dynamic, problems.

The temporal relationship between investment — and thus capital deepening — and the employment of labor to effect that investment can play havoc with productivity estimates. To take one example, real GDP mainly measures market transactions and so has a hard time identifying productive non-market activity. Open-source software and own-account IT investment[15] are very tricky to account for within the intellectual property component of the national accounts. If investment is increasing in these sectors of the economy, we will see labor takeup rise but estimates of capital formation (a segment of real output) will lag… in which case published output per hour readings will actually decline!

This error propagation has serious consequences for the usefulness of productivity statistics for real-time dynamic policymaking. Errors embedded in the measurement of real GDP compound when dividing by hours worked, since productivity estimates respond much more quickly to changes in labor utilization than capital formation. This dynamic in turn drives the consistent surge in productivity when the economy enters recession.

Productivity Growth Booms In Recessions And Jobless Recoveries...
Nonfarm Productivity Growth: Year-over-Year % Change
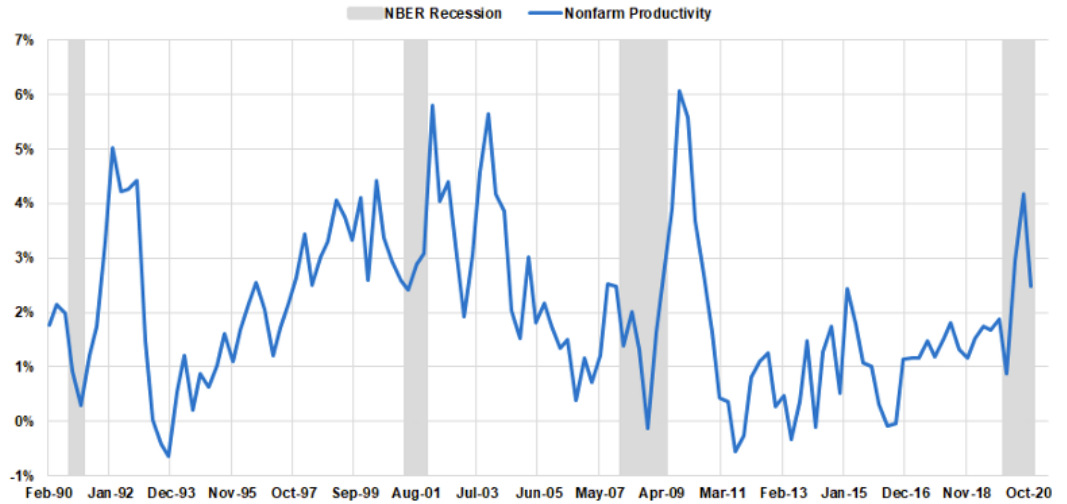
Source: Bureau of Economic Analysis (BEA)

The CBO's model for estimating potential output relies heavily on trend estimates of productivity growth. If positive trend productivity growth is associated with increases in potential output in the model they use. However, trend productivity growth has minimal predictive or explanatory power for forecasting long-run or short-run productivity growth. The following scatters show that short and long term productivity trends alike have minimal ability to forecast the path of future productivity growth:

Past Productivity Trends Do Not Predict Future Trends Annualized
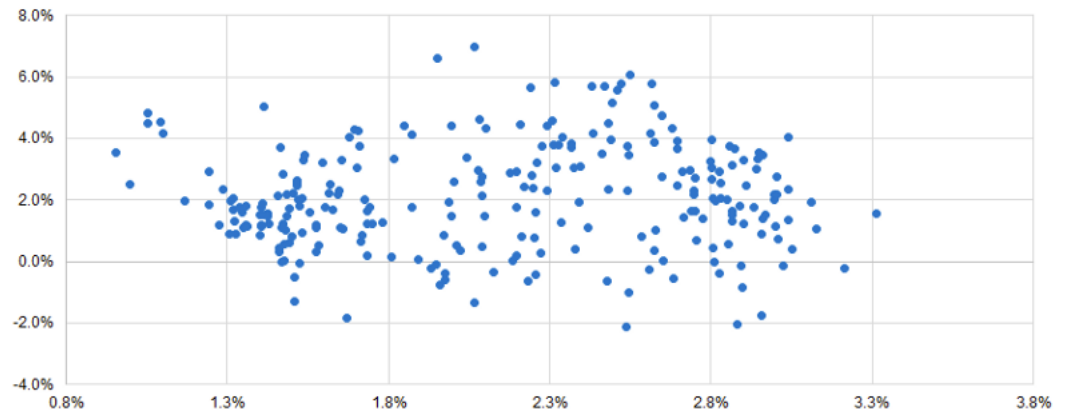Productivity Growth: Forward 5yr (y) vs Realized 10yr (x) (1957-2015)
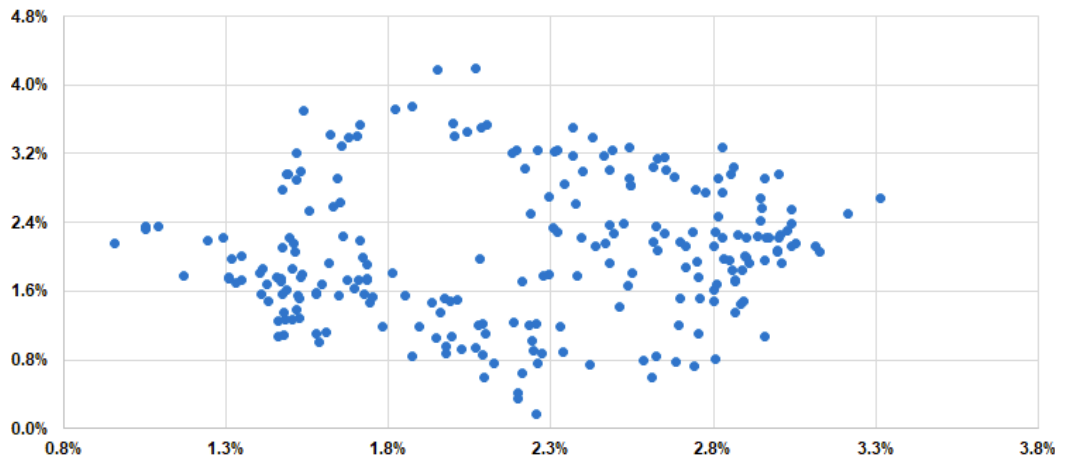
Source: BEA

The Trend Isn't Even Useful For Short-Term Forecasts
Annualized Productivity Growth: Forward 1yr (y) vs Realized 10yr (x) (1957-2019)

Source: BEA

Rather than a trend, the data shows us a cloud. Leaving aside the question that some raise as to whether productivity is a valid metric at all, the above should instill skepticism that such measurements are sufficiently robust to be taken into account when estimating potential output. One core part of the CBO's approach to making very precise measurements of the potential output of the economy as a whole amounts to little more than filtered noise.

The CBO Approach: U*

Alongside TFP, the CBO model relies on the deviation of unemployment from U* to calculate potential output. While the productivity data suffers from a wide variety of empirical issues, it at least attempts a grounding in real-world data. Most estimates of U*, commonly referred to as the Natural Rate of Unemployment, do not even go this far.

Like TFP, inertial U* is an unobservable, unmeasurable variable. Ultimately, it derives from a Phillips Curve dynamic, which assumes that there exists an optimal level of unemployment for the economy as a whole. We have talked about the flaws with this approach before,[16] but in a nutshell, employment above U* is claimed to lead to spiraling inflation, while employment below U* represents foregone output. The level of potential output not associated with accelerating inflation is limited by the level chosen for U*.

Unlike the CBO, the Federal Reserve is moving away from this type of model altogether. Both Chair Powell[17] and Vice Chair Clarida[18] have made speeches that emphasize the growing skepticism towards the use of U* and related measures in Fed policymaking.[19] We have advocated in past[20] that the Fed target, **at a minimum,** past peaks of labor utilization that were not associated with significant and persistent inflation at the time, most recently in February 2020. The CBO falls far short of this standard, choosing instead the average unemployment rate in 2005 as the natural rate of unemployment for the economy at all times.

While targeting a specific level of unemployment is not ipso facto bad policy, the level chosen matters tremendously. From the CBO:

> The Congressional Budget Office's approach to estimating the natural rate of unemployment — the rate that arises from all sources other than fluctuations in demand associated with business cycles — is based on the assumption that rates of unemployment for different demographic groups (by age, sex, education, and race) were approximately at their natural rates in 2005, a year in which employment in U.S. labor markets is thought to have been roughly at its maximum sustainable level. The agency uses estimates of the size of those groups as well as their time-varying rates of potential labor force participation as weights to calculate a weighted average of those 2005 natural rates over time, yielding an estimate of the natural rate over recent history. An implication of this approach is that all of the variation in the natural rate over time arises because of changes in the relative size and participation rates of different demographic groups rather than changes in the natural rates of those groups.[21]

This language may seem innocuous, but it embeds some truly galling assumptions. The first, narrowly technical, point is that the economy has enjoyed much tighter labor markets in the past thirty years than those achieved in 2005, without seeing significant inflation. The months leading up to the COVID pandemic provide a particularly relevant example.

What is much worse is the method the CBO uses to track its U* estimate against changing demographics. Rather than assuming the headline unemployment rate in 2005 is the natural rate, the CBO assumes that the unemployment rate by demographic group in 2005 represents the natural rate for that demographic. These demographics are then summed together to account for drift in the population. On an age or educational attainment-based measure, this could potentially make sense, however, the CBO breaks its estimate out by race. Obviously, it is important to take into account existing disparities when providing a descriptive account of the labor market. However, using those disparities to construct a prescriptive account of the labor market is hardly different from making those disparities into explicit policy. Deciding by fiat that the "natural" rate of unemployment for Black Americans is 10% reflects an intellectual laziness that becomes highly problematic in light of the pronounced cyclicality of the gap between Black and white unemployment rates.[22] This is in stark contrast with the Fed and Chair Powell, who has repeatedly affirmed[23] the responsibility of pursuing equitable fiscal and monetary policy.

The CBO's approach to measuring potential output, and thus the output gap, is founded on its estimates of trend productivity growth and U*. As we have seen, productivity growth is even more methodologically fragile than trend-based estimates, while U* amounts to a tendentious claim that labor markets can never be durably tighter than they were in 2005. Given that we have seen durably tighter labor markets since then, with minimal inflation, this assumption seems to ensure that the CBO's estimate of potential output will be biased sharply downward. Calibrating relief policies to the scale of the output gap requires tacit endorsement of the shaky measurements and specifications described above, and prevents policymakers from helping achieve economic growth and tight labor markets.

What Could The Output Gap Be?

None of this is to say that the goal of measuring the maximum productive potential of the economy should be abandoned. That there are limits to the productive power of the economy at any given time is obviously true. However, those limits have proven incredibly plastic throughout history. The example of WWII[24] is extreme, but instructive. Government economists developed and utilized Input-Output tables to ensure that any economic bottlenecks were widened or avoided. This process created durably tight labor markets and increased output without uncontrolled inflation as an incredible number of new entrants were drawn into the labor force.

However, this strategy requires a lot of research and effort, because it is founded on realistic assumptions about production processes and is informed by local business context. With the rise of services in the modern economy, the challenge of measuring capacity has only grown more complicated. Measuring capacity utilization is sufficiently complex even with respect to the production of discrete tradable goods (e.g. manufacturing, mining), but what would capacity estimates look like in transportation services? Healthcare? Information technology?

As COVID-19 has shown us, there really do exist capacity constraints in these areas, but constructing credible estimates of these constraints requires a strong appreciation for idiosyncrasy and meticulous empirical research. The current set of macroeconomic capacity estimates are neither informed by, nor say much about, the services that can be provisioned with currently available hospital beds, doctors, and nurses. Trucking capacity can be temporarily stretched to capacity constraints as delivery times get extended, only for new capacity to grow as fleets expand and drivers get hired and trained. Estimates of productivity and potential output generally lack the richness to appropriately incorporate such dynamics.

That current methods of estimating potential output fail to provide necessary detail does not mean we should abandon this line of research. In fact, creating durably tighter labor markets without excessive inflation should motivate us to develop measurements which provide a more nuanced read on the underlying productive capacity of the economy. However, the problems with measurements of potential output as they currently exist are so severe that commentators and policymakers should avoid relying on them when calibrating fiscal policy, and focus instead on labor market indicators.

Today, A Problem-Solving Approach To Fiscal Relief Calibration

None of what we have said resolves the challenge of calibrating the optimal size of a fiscal package, but there is no need to tie ourselves to flawed conceptions and estimates of potential output and the output gap.

The question of "how much money can be spent before the economy overheats" is a little bit like asking how much gasoline your car can use before the engine overheats. It depends first and foremost on where the gasoline is going: pouring it on the engine block risks overheating much faster than putting it in the gas tank. At the same time, a lot of things that aren't the amount of gasoline used matter as well: an engine will overheat much quicker if the radiator is broken and it's a hundred degrees outside. Focusing on the amount of gasoline alone won't give a final, or even particularly informative, answer.

It should also be acknowledged that the relief package is not solely for the purpose of smoothing the business cycle. Vaccine distribution funding and social safety net supports in the package are about direct attempts to solve problems of human suffering inflicted by this pandemic. We can make far more reliable claims about what these efforts will yield than we can regarding the location of persistent inflationary constraints.

To the extent we are interested in countercyclical policy, labor market indicators offer a much more robust starting point than real output. Unlike real GDP, employment data sees minimal revisions. The prime-age 25–54 employment-to-population ratio, a broad measure of labor market slack not subject to the survey distortions[25] that plague the headline unemployment rate, is still more than 4% away from its peak as of January 2021. The share of the working age population that is cyclically underemployed is also elevated by more than a percentage point since the pandemic. The pandemic has also led to the introduction of a unique survey distortion in which the unemployed are misclassified as "employed but not at work."[26] Taken together, the pandemic has led to several percentage points more of labor market damage than what the headline unemployment rate currently indicates.

A problem-solving approach to fiscal policy needs to be scaled to reflect the full scope of labor market pain that the pandemic has caused working Americans. In addition to addressing the current deficit of labor market opportunities, calibration efforts also need to account for how past fiscal efforts have helped to prevent worse employment outcomes from materializing in the present. The tremendous expansion of social insurance programs — aside from their plain social and moral necessity — has helped backstop labor markets as well. Workers who would have seen little or no income absent this expansion have used these transfer payments to support employment in areas tied to the production of household consumption goods. Supporting consumption and labor markets will always help topline GDP, while the reverse is not necessarily true. If fiscal policy is to solve the very real problems ordinary Americans face right now, it cannot afford to abandon its ambitions in the face of methodologically fragile estimates of potential output.

View Online

1. https://en.wikipedia.org/wiki/Solow_residual

2. https://fred.stlouisfed.org/graph/?g=Xtn

3. https://www.piie.com/blogs/realtime-economic-issues-watch/defense-concerns-over-19-trillion-relief-plan

4. https://www.washingtonpost.com/opinions/2021/02/04/larry-summers-biden-covid-stimulus/

5. https://research.stlouisfed.org/publications/review/2018/08/13/comparing-measures-of-potential-output

6. https://www.investopedia.com/terms/t/trendanalysis.asp

7. https://www.bea.gov/resources/methodologies/nipa-handbook

8. https://macromyths.blogspot.com/2017/10/about-that-great-moderation.html

9. https://www.ft.com/content/82232968-0433-357e-ace3-326fc7fc35f9

10. https://www.aier.org/article/cell-phone-plans-and-the-cpi/

11. https://www.bls.gov/cpi/quality-adjustment/home.htm

12. https://en.wikipedia.org/wiki/Moore%27s_law

13. https://www.federalreserve.gov/econresdata/notes/ifdp-notes/2014/potential-output-and-recessions-are-we-fooling-ourselves-20141112.html

14. https://research.stlouisfed.org/publications/review/2018/08/13/comparing-measures-of-potential-output

15. https://www.federalreserve.gov/econres/notes/feds-notes/own-account-it-equipment-investment-20171004.htm

16. https://employamerica.medium.com/inflation-the-good-the-bad-and-the-transitory-cc289f524c1c

17. https://www.federalreserve.gov/newsevents/speech/powell20180824a.htm

18. https://www.federalreserve.gov/newsevents/speech/clarida20190503a.htm

19. https://www.federalreserve.gov/newsevents/speech/powell20180824a.htm

20. https://employamerica.medium.com/the-feds-forward-guidance-prioritizing-labor-markets-to-ensure-a-robust-recovery-for-all-8a806a1b2bcf

21. https://www.cbo.gov/system/files/115th-congress-2017-2018/workingpaper/53558-cbosforecastinggrowthmodel-workingpaper.pdf

22. https://miro.medium.com/max/2400/0*RztdFM3IktbOf5PL

23. https://apnews.com/article/coronavirus-pandemic-jerome-powell-inflation-prices-economy-3cffc493578a7af0b0fda31b9010168f

24. https://rooseveltinstitute.org/publications/public-spending-as-an-engine-of-growth-and-equality-lessons-from-world-war-ii/

25. https://econweb.ucsd.edu/~jhamilto/AH2.pdf

26. https://www.nytimes.com/2021/02/22/business/economy/unemployment-rate-covid.html