IARIW 2021

IARIW 2021

Monday 23 – Friday 27 August

Statistisk sentralbyrå
Statistics Norway

How Large are Long-run Revisions to U.S. Labor Productivity?

Peter B. Meyer

(U.S. Bureau of Labour Statistics)

John Glaser

(U.S. Bureau of Labour Statistics)

Kendra Asher

(U.S. Bureau of Labour Statistics)

Jay Stewart

(U.S. Bureau of Labour Statistics)

Jerin Varghese

(U.S. Bureau of Labour Statistics)

Paper prepared for the 36th IARIW Virtual General Conference
August 23-27, 2021
Session 24: Productivity
Time: Thursday, August 26, 2021 [16:30-18:30 CEST]

How large are long-run revisions to U.S. labor productivity?

Preliminary—Please do not cite without permission

Peter B. Meyer, John Glaser, Kendra Asher, Jay Stewart, and Jerin Varghese*

August, 2021

JEL codes: E24; O47; C82

Keywords: Labor productivity; real time data; data revisions; uncertainty communication; nonfarm business sector; official estimates

Abstract

This paper examines revisions to official estimates of quarterly US productivity growth, and to the underlying data on output and hours worked, for the period covering 1995-2015. These estimates are revised substantially in the first months after the reference quarter. Revisions are due to the incorporation of additional microdata, benchmarking, adjustments to seasonal factors, and changes to definitions and methods. The magnitudes of revisions decline to near zero within five years. Revisions to output are larger than revisions to labor hours. Revisions are larger to estimates for first quarter and recession periods. Later revisions are approximately normally distributed but early ones are not. Revisions are not very predictable.

---

* We thank colleagues for their valuable advice: Barbara Fraumeni, Tara Sinclair, Brady West, Shelly Smith, Baoline Chen, Brent Moulton, Shawn Sprague, Leo Sveikauskas, John Stewart, and Shirley Tsai. Any views expressed here are ours and do not necessarily reflect those of the BLS.

1. Introduction

The Bureau of Labor Statistics' (BLS) Labor Productivity and Costs (LPC) program produces quarterly estimates of labor productivity growth.[1] For these estimates, BLS combines output data from the Bureau of Economic Analysis (BEA) with employment and hours data compiled from three BLS surveys: the Current Employment Statistics (CES) survey, the Current Population Survey (CPS), and the National Compensation Survey (NCS).[2] Two of these data sources—the BEA output data and the CES employment and hours data—are revised multiple times after they are first released.[3]

For each reference quarter, BLS releases three regularly scheduled estimates of labor productivity growth. The first, preliminary, estimate (**R0**) is issued within 40 days of the end of the reference quarter. This initial estimate is revised as new data become available. The first revised estimate (**R1**) is released 30 days after R0, and the second revised estimate (**R2**) is released 60 days after that. R2 is the last regularly scheduled release covering the reference period. The timing of the R0 and R1 estimates is determined by BEA's release schedule for the Advance and Second GDP estimates. The R2 estimate is released along with the R0 estimate for the following quarter.[4] Subsequent revisions to the R2 estimates, such as those due to BEA Comprehensive Revisions, can be large and can occur long after the reference quarter. GDP is

---

[1] The BLS calculates the labor productivity index as the ratio of an index of output divided by an index of hours. Labor productivity growth is equal to the ratio of the current quarter index divided by the previous quarter's index (minus 1). All growth rates are annualized.

[2] A complete description of how BLS estimates total hours worked can be found in U.S. Bureau of Labor Statistics *Handbook of Methods*, and Lucy Eldridge, Chris Sparks, and Jay Stewart, "The BLS Productivity Program" in the *Oxford Handbook of Productivity Analysis*, Chapter 3. Oxford University Press, 2018.

[3] Seasonally adjusted GDP and employment data are used for the quarter-to-quarter measures. The CPS data are almost never revised, but seasonal factors are subject to minor revisions. The NCS data are not revised because the LPC program uses NCS data for the fourth quarter and allocates changes to quarters using the Denton procedure. With this procedure, seasonal adjustment is not necessary.

[4] This is because R1-to-R2 revisions are generally not large enough to warrant a separate news release.

estimated from multiple data sources, some of which do not become available until long after the end of the reference period. Thus, the estimates are never "final."

The output measure used by the BLS productivity program is a subset of GDP.[5] Early GDP estimates are subject to substantial revisions because they are based on "…partial and preliminary source data as well as trend projections when data are not available."[6] The source data for these early estimates come from "…a mixture of survey, tax, and other business and administrative data as well as various indicators, such as heating degree days…" Another source of revisions to GDP is the recalculation of seasonal factors.

The main source of revisions to the hours data is revisions to the CES data. There are three regularly scheduled releases for each reference month, the first, second and third closing estimates. The CES's first closing estimates are usually released on the first Friday after the reference month, and the second and third closing estimates are released in the following two months. These revisions are primarily due to the collection of additional data and to the recalculation of seasonal adjustment factors. In addition to the three regular closings, each February the CES employment data are benchmarked to data from the Quarterly Census of Employment and Wages (QCEW) and seasonal adjustment models are updated.[7] Revisions to hours growth are typically smaller than revisions to output growth.

Revisions to estimates of labor productivity growth are often larger than revisions to GDP growth for two reasons. First, the output measure used for labor productivity estimates excludes

---

[5] BLS excludes general government and non-profits from its output and hours measures, because output measures for these sectors are derived largely from measures of inputs, not output prices and quantities.

[6] See Dennis J. Fixler, Ryan Greenaway-McGrevy, and Bruce T. Grimm, "The Revisions to GDP, GDI, and Their Major Components." *Survey of Current Business,* August 2014; and Dennis J. Fixler, Danit Kanal, and Pao-Lin Tien, "The Revisions to GDP, GDI, and Their Major Components." *Survey of Current Business*, 2018. https://www.bea.gov/scb/pdf/2018/01-January/0118-revisions-to-gdp-gdi-and-their-major-components.pdf.

[7] Each February CES publishes benchmarks. The previous March's employment is set to match the QCEW total, and CES employment in the intermediate months are interpolated. Once a year they re-estimate the seasonal adjustment model.

government, for which data comes from administrative sources and is not revised much. And second, the revisions to labor hours are not perfectly correlated with revisions to output.

The revisions described above result from what Charles Manski has referred to as "transitory uncertainty," in that the estimates improve in quality as more data become available and estimates are revised.[8] This can be distinguished from conceptual uncertainty (such as ambiguity) or sampling error.

In an earlier paper, we focused on revisions to the R0 and R1 estimates relative to the R2 estimate and presented a methodology for constructing "prediction" intervals based on past revisions (Asher et al, 2021). In that paper, we found that there was no trend in revisions over time, no relationship between the magnitude of the estimate and the size of the revision, and that there were no significant business cycle effects. We did find some variation in the size of the revisions across quarters, but the differences were not statistically significant. Decomposing the revisions to labor productivity growth, we found that revisions to output accounted for the largest share of average R0-to-R2 revisions, while the R1-to-R2 revisions were more evenly divided between revisions to output and revisions to hours.

In this paper, we focus on longer run revisions to labor productivity growth. We find that estimates of output growth tend to be revised downward, with the revisions being largest for Q1. Estimates of hours growth tend to be revised downward slightly and the revisions are minimal after three years. The R2-to-R40 revisions are much larger than the R0-to-R2 revisions. And, using a simple regression framework, we show that the R2 estimate does a better job of explaining the variation in the R40 values of output and labor productivity than the earlier R0 estimate.

[8] Charles F. Manski, "Communicating Uncertainty in Official Economic Statistics: An Appraisal Fifty Years after Morgenstern." *Journal of Economic Literature* 53(3), 2015, pp. 631-653.

2. Long-Term Revisions to Output, Hours, and Labor Productivity

Estimates of labor productivity growth are revised long after the end of the reference quarter. For example, there are over 100 estimates for reference quarters in the 1990s. One might expect that these revisions become smaller over time and that the estimates converge to a value. We organize our data so that each observation of a value for our variables is associated with both a reference period and a release period.

Figure 1: The Absolute Value of Revisions to Output, Hours, and Labor Productivity

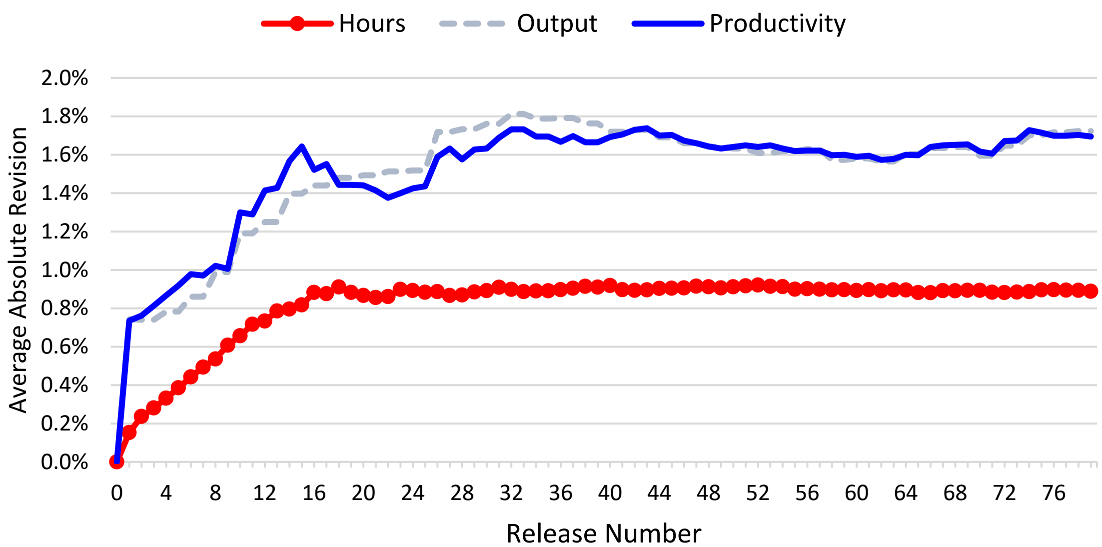


Figure 1 shows the average of the absolute value of the sum of revisions to output, hours, and labor productivity for ten years starting with the R0 estimate. For example, the value for release number 40 is equal to the absolute value of: (the R40 estimate minus the R0 estimate). The key things to note here are that revisions to hours are smaller and stabilize earlier than revisions to output and labor productivity. Second, the largest revisions to output and labor productivity are in the first two revised estimates, although there are further substantial revisions over the next year and a half. And third, the revisions to output and labor productivity have roughly stabilized by R40. For this reason, we make our comparisons relative to the R40 estimates, which we will treat as "truth."

Because we need 5 years of revisions for each reference quarter, we restrict our sample to the 1995-2015 period. We focus on U.S. nonfarm business labor productivity, which covers about 75% of GDP (general government and nonprofits are excluded). All data sources are seasonally adjusted either by the source agency or the BLS productivity program. Some of the details of how the BLS productivity program constructs its estimates can be found in the BLS *Handbook of Methods* and Eldridge et al (2018).

Table 1: Summary Statistics

(a)

	LP at R0	**LP at R2**	**LP at R40**
Mean	2.27	2.40	2.05
Std. dev.	2.19	2.51	2.72
Min	-2.05	-4.45	-3.39
Max	9.45	9.50	10.34

(b)

	Output at R0	**Output at R2**	**Output at R40**
Mean	2.98	3.05	2.66
Std. dev.	2.65	3.09	3.45
Min	-8.19	-8.84	-12.03
Max	8.80	10.40	11.01

(c)

	Hours at R0	**Hours at R2**	**Hours at R40**
Mean	.73	.66	.62
Std. dev.	2.65	2.60	2.90
Min	-8.95	-9.04	-10.15
Max	5.06	4.88	4.56

Tables 1(a)-(c) show averages of growth rates for three estimates of output, hours, and labor productivity. The R0 and R2 estimates are as described as above. The R40 estimate is the 40th revised estimate, which is about five years after the reference quarter. Each entry in Table 1 is the average of the estimated quarterly growth rates for all quarters in the sample period, where each observation in the entry has been revised the same number of times.

The average of the growth rates for output and labor productivity have an inverted U-shape, with standard deviations that are monotonically increasing. The estimates of hours growth decrease slightly over time, and there is not much change in the standard deviation.

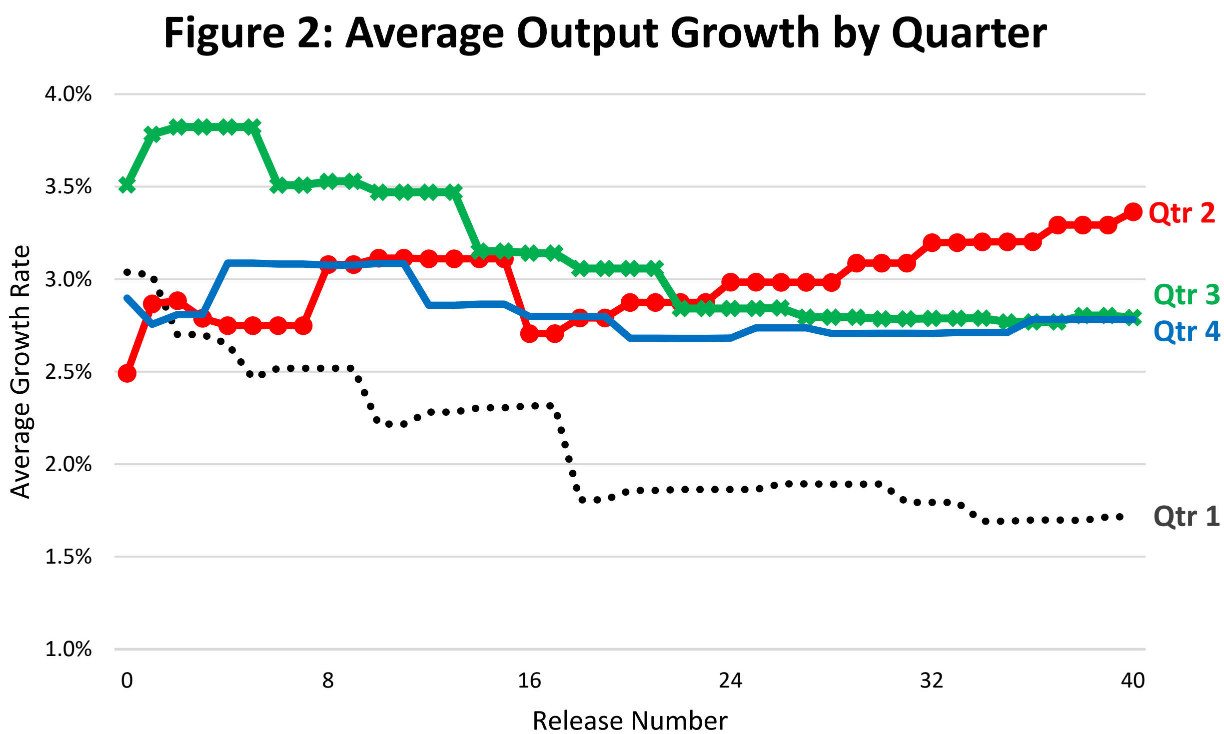
3. Revisions to Output, Hours, and Labor Productivity

In Figure 1, we saw how the magnitude of revisions changes over time. However, it is also useful to look at how estimated values of output and hours growth change with subsequent revisions. Given that there are regularly scheduled revisions that occur in the same quarter each year, we look at the patterns by reference quarter (Q1-Q4).

3.1 Revisions to output by quarter

Figure 2 summarizes how the estimated output growth rate changes with subsequent releases. The horizontal axis graphs the release number. The vertical axis graphs the average percent growth in output from the previous quarter, seasonally adjusted and annualized, from BEA. Each line represents a different quarter (Q1-Q4), and each point shows the average growth rate for all of the quarters between 1995Q1 and 2014Q4 for the indicated quarter as of the indicated release number. Thus, as of the 40th release, the average growth rate for all estimates of Q4 labor productivity over this period was about 1.75 percent.

The figure shows figure that estimates of output growth tend to be revised downward. The largest downward revisions are for Q1 and Q3. Estimates for Q2 estimates tend to be revised upward.

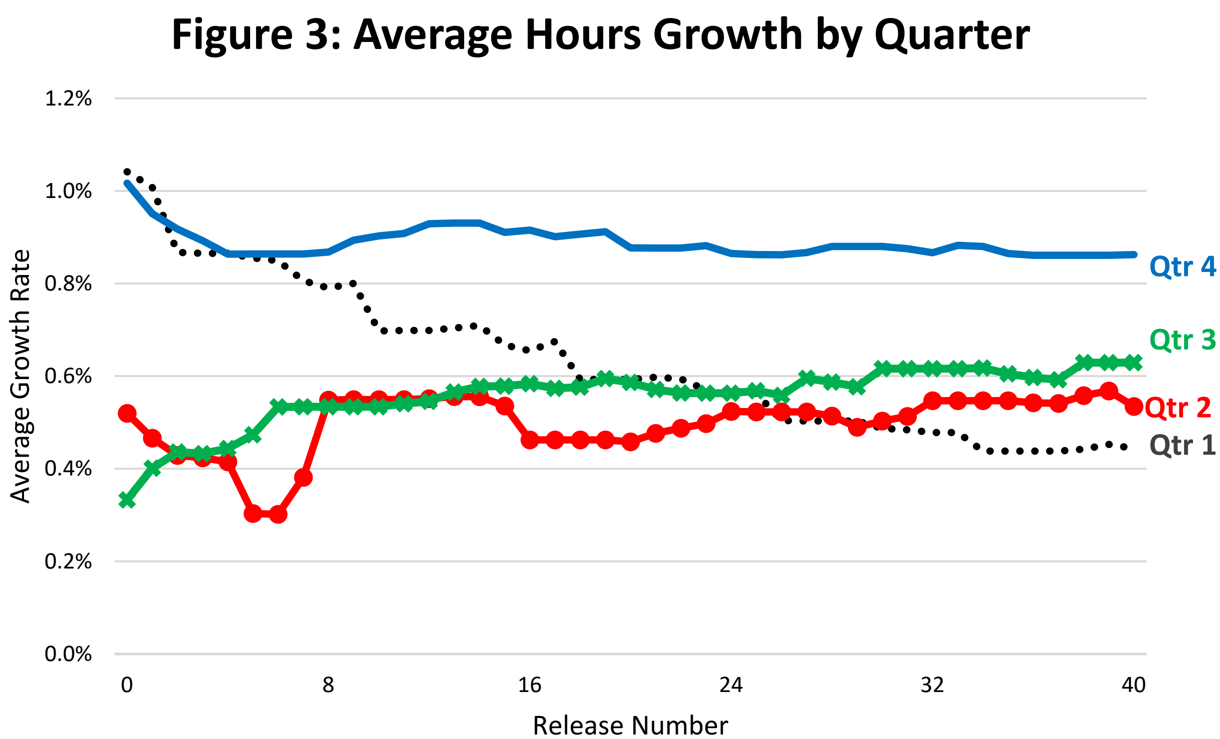


Some of the jumps can be traced to regularly scheduled revisions. The large downward jump at release 19 for Q1 corresponds to the August releases, when this output series incorporates annual revisions to GDP which are published each July.

3.2 Revisions to hours worked by quarter

The labor side of the picture is less volatile than the output side. The estimated growth in hours worked from the previous quarter, shown in Figure 3, is revised significantly less than the estimated growth in output. The period covered is 1995-2015. Note the difference in the vertical scale from Figure 2.

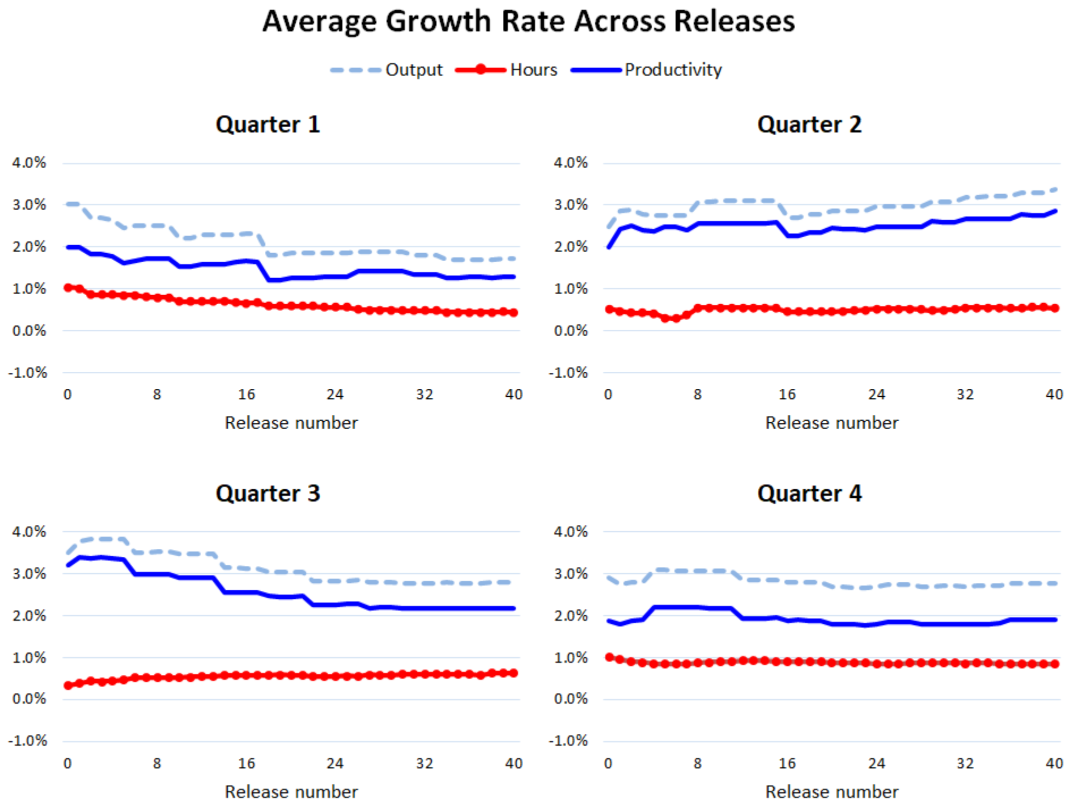
Estimated hours growth was revised downward in all quarters except Q4. Like estimated output growth, the largest downward revisions were for Q1. Revisions to hours were much larger in magnitude for the March and August releases than for other releases, which makes sense given the CES revision and benchmarking schedule.



<u>3.3 Putting the pieces together</u>

Figure 4 shows output, hours, and labor productivity by release number separately for each quarter. As we can see, the estimates of labor productivity growth closely follow estimates of output growth.

Figure 4:



4. The Distribution of Revisions

The preceding analysis suggests that revisions can be large. In this section, we examine the magnitude of revisions directly. Figure 5 shows the revisions to estimates of labor productivity between the R0 and R2 estimates, along with graphs of a kernel smoothing of the distribution and a normal distribution for comparison. Figure 6 shows the distribution for R2-to-R40 revisions.

The distribution of R0-to-R2 revisions in Figure 5 is more peaked and more skewed (left) than a normal distribution with most of the revisions being between −1 and 2. The mean is slightly positive (around 0.18) but would be higher were it not for several large downward revisions. If we were to exclude revisions that are greater than 2 in absolute value, the mean would be about 0.3.[9]

The R2-to-R40 revisions in Figure 6 appears to be approximately normally distributed based on the comparison of the kernel density function to the normal curve in the figure. Keeping in mind that the scales in Figures 5 and 6 are different, we see that there are many more R2-to-R40 revisions that are greater than 2 in absolute value. Further, a larger number of revisions are negative, resulting in a mean R2-to-R40 revision of −0.29.

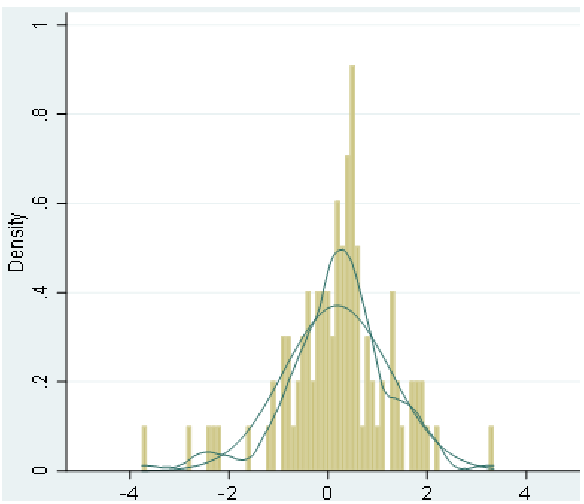


Figure 5. Revisions to LP from R0 to R2

Reference quarters 1995-2020, N=99, mean=0.18
A normal distribution is shown for comparison

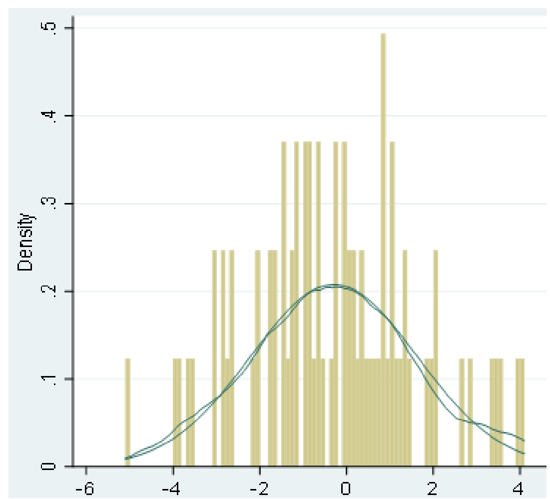


Figure 6. Revisions to LP from R2 to R40

Reference quarters 1995-2015, N = 81. Mean = -0.29

We tested for normality of the distribution of labor productivity, output, and hours revisions, using the Shapiro-Wilk and other tests. These tests reject normality of R0-to-R2 revisions to output and labor productivity, but the R2-to-R40 revisions of the variables are close to normal. Our simplified findings are shown in Table 2, using the term "compressed" to describe a distribution like the one in figure 5 that has a higher central peak than a normal distribution.[10]

---

[9] Our earlier paper (Asher et al., 2021) discusses R0-to-R2 revisions in detail. It shows that revisions of all three statistics from R0 to R2 were not normally distributed, especially for output. Historical revisions were left-skewed, with a long fat tail of sharp downward revisions. One quarter's R0 is missing because of a government shutdown.
[10] We can report test statistics or kurtosis measures in future versions of this paper.

Table 2: Normality tests for distributions

	Hours	Output	Labor productivity
R0 estimates	Compressed	Compressed	Compressed
R0 to R2 revisions	Normal	Compressed	Compressed
R2 to R40 revisions	Normal	Normal	Normal

5. Prediction Intervals for Long-Run Revisions

We have summarized the magnitude of the revisions and the path of estimated growth rates due to revisions. Now, we examine likely values of later estimates given R0 by constructing "prediction" intervals.

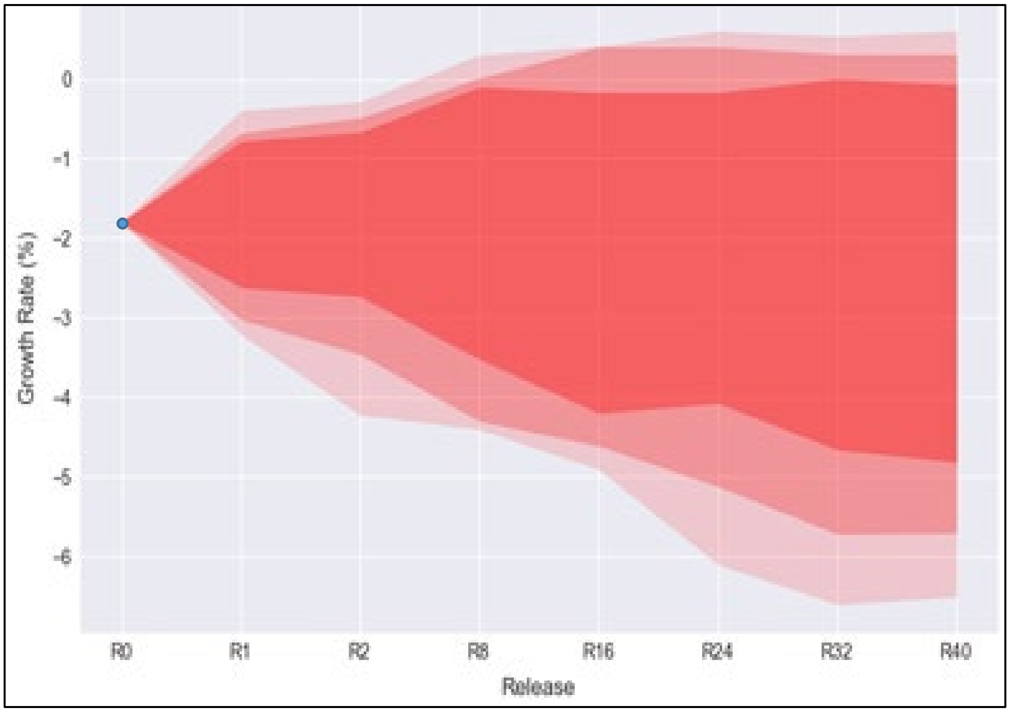
The fan chart in Figure 7 shows prediction intervals constructed as of the time of the R0 release. These intervals were calculated using 20 years of data on revisions and the weighted percentile method described in Asher, et al (2021).[11] The vertical axis shows the measured labor productivity growth rate. The horizontal axis graphs the release number. The first three tick marks indicate the three regularly scheduled news releases (R0, R1, and R2). The fourth and subsequent ticks represent release dates that are one year apart. There are 8 releases each year.

The R0 estimate of the labor productivity growth rate of -1.8% for 2014 Q4 is the starting point of the fan. The bounds of the intervals are estimated as of R0. The darkest zone is the 70% probability range, and the fainter one is the 80% range, and the widest and faintest one is the 90% interval. These do not represent changes in the economy over time; they are the range of predicted changes in the estimate for 2014 Q4 as data and methods are updated. The

[11] The weighted percentile method is less sensitive to other methods such as the modified confidence interval method in Fixler, et al (2014), and more accurately captures the fraction of revised estimates that fall within the intervals. See Asher, et al (2021) for a description of the weighted percentile method, and how we evaluated the different methods.

prediction intervals are wide because revisions can be large. One could construct a similar fan chart for each release period, illustrating tradeoffs between accuracy and timeliness.

Figure 7: Fan chart for 2014Q4 Labor productivity growth, with 70%, 80%, and 90% intervals



Upper and lower bounds are not symmetric around the R0 estimate because the extreme downward revisions have tended to be larger than the extreme upward revisions. The most extreme downward revisions tend to occur during recessions, because early estimates of GDP use models and trends to fill in for missing data and these models tend to miss turning points. So, in a recession, there is a tendency for the GDP growth to be revised downward.

6. How well do early estimates predict later estimates?

The fan chart provides one way to look at this question. In this section, we examine how well early estimates of labor productivity growth (the R0 and R2 estimates) predict the R40 estimate in a simple OLS framework. Table 3 shows the coefficients on regressions of the R40 estimate on early estimates (R0 and R2).[12] Ideally, the coefficients on early values in these regressions would be close to 1, the constant would be close to 0, and the R-squared would be high. The coefficients on early values of about 0.74 in columns (1) and (2) suggest that R0 and R2 predict R40 equally well. However, the constant in the R2 equation is closer to 0, although neither constant is statistically different from zero. The R-squared of the R2 regression is much higher than that of the R0 regression (0.50 vs. 0.37), indicating that R2 explains more of the variation in the R40 estimate than does the R0 estimate. This suggests R2 is better in a mean-squared-error sense. Column (3) shows that including the R0 estimate in the R2 regression does not meaningfully add information; the revision from R0 to R2 does not predict the later revision.

Table 3: Early Measures as Predictors of LP as of R40

	(1)	(2)	(3)
LP as of R0	0.743***		-0.089
	(0.106)		(0.202)
LP as of R2		0.756***	0.825***
		(0.083)	(0.177)
Constant	0.364	0.197	0.228
	(0.326)	(0.283)	(0.293)
R-squared	0.373	0.503	0.504

N=85 reference quarters, 1995-2015Q1. Standard errors in parentheses.
*** p<0.01, ** p<0.05

[12] Predictive accuracy is similar if we use R0 and R2 hours and output growth rates as predictors separately.

Table 4: Quarter Predictors of LP as of R40

VARIABLES	(1) LP40	(2) LP40
LP as of R0	0.558***	
	(0.205)	
LP as of R0 × Q2	0.313	
	(0.306)	
LP as of R0 × Q3	0.417	
	(0.289)	
LP as of R0 × Q4	0.233	
	(0.303)	
2nd quarter	0.800	0.891
	(0.839)	(0.726)
3rd quarter	-1.443	-1.836**
	(0.974)	(0.857)
4th quarter	0.079	0.010
	(0.842)	(0.702)
LP as of R2		0.612***
		(0.151)
LP as of R2 × Q2		0.089
		(0.224)
LP as of R2 × Q3		0.453*
		(0.231)
LP as of R2 × Q4		0.231
		(0.225)
Constant	0.325	0.224
	(0.570)	(0.472)
R-squared	0.450	0.572

N=85 reference quarters. Standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Given that our earlier research has shown that the size of the revisions varies by quarter, we extend the regressions in Table 3 by adding quarterly dummies and quarterly interactions (see Table 4). The interactions allow for the relationship between the early estimates (R0 and R2) to vary by quarter, while the dummies account for differences in average productivity growth across quarters. The results for R0 in column (1) are a little mixed. Adding the quarterly dummies and interactions increases the R-squared in both the R0 and R2 regressions. And it appears that the quarterly interactions are capturing some variation in the predictive power of the early estimates as evidenced by the decrease in the coefficients on R0 and R2. However, the coefficients on the quarterly dummies and interactions are not statistically different from zero (except the Q3 dummy and interaction in the R2 regression) or from each other.

Table 5: Predictors of Hours worked growth as of R40

	(1)	(2)	(6)
Hours as of R0	1.042***		
	(0.041)		
Hours as of R2		1.073***	0.937***
		(0.036)	(0.050)
1990s			0.250
			(0.228)
2000s			-0.286
			(0.230)
2nd quarter			0.638***
			(0.233)
3rd quarter			0.708***
			(0.234)
4th quarter			0.384
			(0.233)
Recession quarter			-1.195***
			(0.388)
Presidential election year			-0.091
			(0.201)
Constant	-0.120	-0.082	-0.190
	(0.112)	(0.097)	(0.238)
R-squared	0.889	0.915	0.938

N=85 reference quarters, 1994-2015Q1. Standard errors in parentheses.
*** p<0.01, * p<0.1

Given the results in Tables 3 and 4, it is natural to wonder if the R0 and R2 estimates of output and hours are better predictors of their R40 values, compared with the R0 and R2 estimates of labor productivity. Tables 5 and 6 show similar regressions for output and hours. The R-squared in both the output and hours regressions are higher than the R-squared in the labor productivity equation. This make sense, because output and hours are contributing independent information.

Table 5 shows that the early estimates of hours do a good job of predicting the R40 values. There is a slight improvement going from the R0 to the R2 estimate, as the R-squared in column 2 is slightly higher than that in column 1. The coefficients on the early estimates are approximately 1 in both regressions, and the constants are close to zero. This is as we would expect, because R0-to-R2 revisions are relatively small for hours.

Table 6: Predictors of Output growth as of R40

	(1)	(2)	(3)
Output as of R0	1.014***		
	(0.089)		
Output as of R2		0.912***	0.823***
		(0.069)	(0.080)
1990s			0.207
			(0.541)
2000s			-0.931*
			(0.504)
2nd quarter			1.632***
			(0.538)
3rd quarter			0.023
			(0.543)
4th quarter			0.840
			(0.537)
Recession quarter			-0.915
			(0.816)
Presidential election year			-0.503
			(0.458)
Constant	-0.312	-0.120	0.136
	(0.353)	(0.301)	(0.536)
R-squared	0.610	0.678	0.758

N=85 reference quarters. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

The results are somewhat different in the output regressions in Table 6. The R2 regression does a noticeably better job of explaining the variation in the R40 estimate of output, with an R-squared of 0.68 vs. 0.61. Although the coefficient is closer to 1 in the R0 regression, the constant is much larger. Putting these together, we conclude that the R2 estimate of output is an improvement over the R0 estimate.

Columns (3) in Tables 5 and 6, add reference quarter dummies, and dummies for recession quarters and quarters before presidential elections. Although some of the added variables in the hours equation are statistically significant, they do not improve the R-squared by very much. In contrast, these variables do improve the R-squared in the output regression—from 0.678 to 0.758.[13] The large and statistically significant coefficient on the Q2 dummy reflects the large difference in the average R40 values of output for Q1 and Q2 that can be seen in Figure 2.

7. Revisions in the COVID-19 era

The COVID-19 pandemic severely disrupted economic activity in early 2020 and the speed of this disruption placed unprecedented demands on a statistical system that was not designed to measure such rapid changes. First quarter estimates of labor productivity were particularly affected, because the sharp decline in economic activity occurred in the last two weeks of the quarter. On the output side, the BEA's use of projections for parts of its advance estimate would normally result in large revisions because projections cannot capture large changes that occur over a short period of time. On the input side (hours worked), the surveys that provide the employment and hours data did not capture most of the declines in employment because the declines occurred largely after the reference periods of those surveys.

Both BEA and BLS quickly adapted to the new environment and modified their methods to provide a more accurate picture of output and productivity growth. For its Advance Estimate of 2020 Q2 GDP, BEA modified its procedures by incorporating high frequency data such as credit

[13] In the next draft of the paper, we will use the predicted values from the output and hours regressions to calculate labor productivity growth, and compare it to the predicted value from the labor productivity equation.

card transactions, and relying less on projections. The BLS Productivity Program modified its usual procedures for estimating hours worked by incorporating data on initial Unemployment Insurance (UI) claims for its preliminary estimate.

Revisions to 2020 Q2 output have been relatively small. The R0-to-R1 revision was −0.3 of a percentage point (from −6.2 percent to −6.5 percent), and the R1-to-R2 revision was 0.1 of a percentage point to −6.4 percent. BEA plans to continue using high frequency data, and it is possible that this change will result in smaller revisions to output, and likely to labor productivity.

The revisions to hours were larger than revisions to output, mainly because of the one-time changes in methodology. For the preliminary Q1 estimate, employment was estimated week-by-week under the implicit assumption that the UI Initial Claims reflected actual job losses and that there were no transitions from non-employment to employment.[14] These are strong assumptions, but the adjustment significantly improved the estimate of total hours worked. The adjusted preliminary estimate of Q1 productivity growth was ─2.5 percent vs. the unadjusted estimate of ─5.2 percent. Only wage and salary employment data were adjusted; there was not enough data to adjust self-employed worker hours or average weekly hours of wage and salary. Once the April data were available, it became feasible to generate week-by-week estimates of hours by interpolating between the March and April estimates. This adjustment reduced the growth in hours worked by −1.8 percentage points, which more than offset the −0.3 revision to output and resulted in an upward revision to Q1 labor productivity growth of 1.6 percentage points to −0.9 percent. The R1-to-R2 revisions further increased Q1 labor productivity growth to −0.3 percent.

[14] The LPC program considered using changes in continued claims, but determined that initial claims more-accurately reflected actual job losses. A description of the methodological changes can be found here: https://www.bls.gov/covid19/effects-of-covid-19-pandemic-on-productivity-and-costs-statistics.htm#quarterly-LPC

Table 7: 2020 Revisions to labor productivity growth

	Labor productivity estimates			Revisions	
	R0	R1	R2	R0-to-R2	R1-to-R2
Q1	-2.5	-0.9	-0.3	2.2	0.6
Q2	7.3	10.1	10.6	3.3	0.5
Q3	4.9	4.6	5.1	0.2	0.5
Q4	-4.8	-4.2	-3.8	1.0	0.4

Table 7 summarizes the R0-to-R2 and R1-to-R2 revisions to labor productivity in 2020. The largest revision was the 3.3 percent prelim-to-R1 revision for Q2, which was entirely due to the revision to output. The next largest revision was the R0-to-R2 revision for Q1, which was mostly due to revisions to hours. To put these revisions into perspective, the 3.3 percentage point revision for Q2 and the 2.2 percentage point revision for Q1 are among the largest revisions since 2000q1. As noted above, the large R0-to-R1 revision to Q1 labor productivity growth was due mainly to the one-time modifications to the methodology for estimating hours. Had this modification not been made, the revision would have been smaller, but Q1 labor productivity growth would have been understated and Q2 growth would have been overstated.

7. Conclusions

In this paper, we have examined the behavior of long-term revisions to quarterly estimates of labor productivity, and its components—output and hours. We find that estimates of labor productivity are revised long after the end of the reference quarter, and that most of these revisions are due to revisions to output. Revisions to estimated hours growth are small after two years because the source data are subject to only minor revisions after they are benchmarked. In contrast, estimates of output growth stabilize after about five years, but are revised non-trivially even ten years after the end of the reference quarter. GDP is estimated from multiple data sources, some of which do not become available until well after the end of the reference period, and revisions may include changes to concepts and methods.

Revisions to both output and hours tend to be negative, with downward revisions being larger for Q1 estimates of output and hours. In comparing the distributions of R0-to-R2 revisions and R2-to-R40 revisions, we find several differences. As expected, the R2-to-R40 revisions are much larger than the R0-to-R2 revisions. The R2-to-R40 revisions are approximately normally distributed, whereas the R0-to-R2 revisions are left skewed.

Using a simple regression framework, we found that the R2 estimate does a better job of explaining the variation in the R40 values of output and labor productivity than the earlier R0 estimate. The R0 and R2 values of hours worked perform similarly, mainly because revisions to hours tend to be much smaller. Separate regressions on output and hours indicate that we can explain more of the variation in these variables, compared to labor productivity. This makes sense, because output and hours are not perfectly correlated.

One topic we have not addressed directly is the impact of BEA's comprehensive revisions. In analyses not presented here, we found that if estimates of their effect were taken out, there would be virtually no impact on the interval bounds of the fan chart in Figure 7. In future work we anticipate estimating the effect of these revisions on productivity measures, directly in our data and perhaps along the lines of Koh et al. (2020).

Appendix

Table A1: Annual data calendar for quarterly nonfarm labor productivity

Month	NIPA releases and revisions	CES releases and revisions	Nonfarm business labor productivity OPT publishes, with notes on source data
January	First (Advance) estimate of Q4 GDP	Dec employment release	
February	Second Q4 GDP release	Jan employment Annual benchmark to QCEW, and seasonal adjustment of recent years	Preliminary release for Q4, R2 for Q3 Uses Advanced Q4 NIPAs, CES data for December; Updates to SW and HWHP ratios
March	Third Q4 GDP	Feb employment, update to Jan employment ("second closing")	R1 for Q4 Incorporates the annual CES benchmark revision through Q4 of the previous year. This affects both current and prior quarter hours.
April	First Q1 GDP	March employment, and 3rd closing for Jan	
May	Second Q1 GDP	April employment	Preliminary for Q1, R2 for Q4; Updates to SW and HWHP ratios
June	Third Q1 GDP	May employment	R1 for Q1
July	First Q2 GDP and annual revision, usually of most recent three years	June employment	
August	Second Q2 GDP	July employment	Preliminary for Q2, and R2 for Q1 This release incorporates the annual NIPA/GDP benchmark revisions; Updates to SW and HWHP ratios
September	Third Q2 GDP	Aug employment	R1 for Q2
October	First Q3 GDP	Sept employment	
November	Second Q3 GDP	Oct employment	Preliminary for Q3, R2 for Q2; Updates to SW and HWHP ratios
December	Third Q3 GDP	Nov employment	R1 for Q3

Table A1 notes:

- OPT's timing of releases is designed to quickly incorporate the major releases from BEA of the First (Advanced) and second estimates of quarterly GDP growth. Revisions in the third releases are smaller, and OPT does not issue a regular productivity release immediately afterward, but incorporates them at the next opportunity.
- SW ratio = supervisory worker ratio, also called production/nonproduction ratio, computed from CPS microdata.
- HWHP ratio is hours worked/hours paid ratio, computed from NCS.

References

Aruoba, S. Borağan. 2008. Data Revisions Are Not Well Behaved. *Journal of Money, Credit and Banking*, 40:2/3. 319-340. https://www.jstor.org/stable/25096254

Asher, Kendra; John Glaser; Peter B. Meyer; Jay Stewart; and Jerin Varghese. How large are revisions to estimates of quarterly labor productivity growth? BLS working paper WP-538, April 2021

Bognanni, Mark; John Zito. 2016. New Normal or Real-Time Noise? Revisiting the Recent Data on Labor Productivity. Economic Commentary, Federal Research Bank of Cleveland.

CES program. Revisions, Technical Notes for the Current Employment Statistics Survey. bls.gov

CES National Vintage data. BLS public web page, in part discussing seasonal adjustment.

Chen, Baoline; Tucker McElroy; Osbert Pang. 2019. Assessing Residual Seasonality in the U.S. National Income and Product Accounts (NIPA) Aggregates. Presentation at Joint Statistical Meetings.

Clark, John, Caroline Gibbons, Susan Morrissey, Joshua Pooley, Emily Pye, Rhett Wilcox and Luke Willard. 2013. "Estimates of uncertainty around budget forecasts." Treasury Working Paper 2013–04. https://treasury.gov.au/publication/estimates-of-uncertainty-around-budget-forecasts/, summarized at http://www.budget.gov.au/2014-15/content/myefo/html/05_attachment_b.htm

Cowan, Benjamin; Shelly Smith; Sarahelen Thompson. 2018. Seasonal Adjustment in the National Income and Product Accounts. Results from the 2018 Comprehensive Update. *Survey of Current Business,* 98:8. https://apps.bea.gov/scb/2018/08-august/0818-gdp-seasonality.htm

Croushore, Dean. 2011. Frontiers of Real-Time Data Analysis. *Journal of Economic Literature* 49:1, 72-100

Dean, Edwin; Michael Harper; Phyllis Flohr Otto. Improvements to the quarterly productivity measures. *Monthly Labor Review*, October 1995. https://www.bls.gov/mlr/1995/10/art4full.pdf

Dey, Matthew; Mark A. Loewenstein. June 2017a. Quarterly Benchmarking for the Current Employment Survey. U.S. Bureau of Labor Statistics Working Paper 496.

Dey, Matthew; Mark A. Loewenstein. November 2017b. A quarterly benchmarking procedure for the Current Employment Statistics program. *Monthly Labor Review*.

Eldridge, Lucy P., Chris Sparks, and Jay Stewart. 2018. "The BLS Productivity Program," in the *Oxford Handbook of Productivity Analysis.* Oxford University Press.

Ferris, Erin E. Syron, Soo Jeong Kim, and Bernd Schlusche. 2017. "Confidence Interval Projections of the Federal Reserve Balance Sheet and Income." FEDS Notes, January 13,

2017. https://www.federalreserve.gov/econresdata/notes/feds-notes/2017/confidence-interval-projections-of-the-federal-reserve-balance-sheet-and-income-20170113.html

Fett, Nicholas A. 2015. Comparing with the original: a look at Current Employment Statistics vintage data. *Monthly Labor Review*.

Fixler, Dennis J. 2012. Revisions to BEA's Estimates of GDP and GDI. Presented to quarterly meeting of COPAFS. Council of Professional Association on Federal Statistics.

Fixler, Dennis J., Ryan Greenaway-McGrevy, and Bruce T. Grimm. The Revisions to GDP, GDI, and Their Major Components. *Survey of Current Business*. July 2011

Fixler, Dennis J., Ryan Greenaway-McGrevy, and Bruce T. Grimm. 2014. The Revisions to GDP, GDI, and Their Major Components. *Survey of Current Business*. Aug 2014.

Fixler, Dennis J., Danit Kanal, and Pao-Lin Tien. 2018. The Revisions to GDP, GDI, and Their Major Components. *Survey of Current Business.*

Fixler, Dennis J., and Bruce T. Grimm. 2002. Reliability of GDP and related NIPA estimates. *Survey of Current Business* 9-27.

Fixler, Dennis J., and Bruce T. Grimm. 2008. The reliability of the GDP and the GDI estimates. *Survey of Current Business.* (February), 16-32

Galvao, Ana Beatriz, James Mitchell, and Johnny Runge. 2020. Communicating Data Uncertainty: Experimental Evidence for U.K. GDP. Working paper, University of Warwick. https://docs.google.com/viewer?a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbnxhbmFiZ2FsdmFvfGd4OjEwNzBmMWY4N2E4N2RkM2E

Glaser, John, Kendra Asher, Peter Meyer, Jay Stewart, and Jerin Varghese. 2020. Reliability Statistics for Quarterly Labor Productivity Estimates]. Report to BLS Technical Advisory Committee, to be presented on November 20, 2020. https://www.bls.gov/advisory/tac/reliability-statistics-for-quarterly-labor-productivity-estimates-11-20-2020.pdf

Grimm, Bruce T. and Teresa L. Weadock. Gross Domestic Product: Revisions and Source Data. *Survey of Current Business.* 86 (February 2006):11–15.

Holdren, Alyssa E. 2014. Gross Domestic Product and Gross Domestic Income: revisions and source data. Survey of Current Business.

Holdren, Alyssa E. and Bruce T. Grimm. Dec 2008. Gross Domestic Income: Revisions and Source Data. *Survey of Current Business.* 88:14–20.

Jacobs, Jan P.A.M., Samad Sarferaz; Jan-Egbert Sturm; Simon van Norden. 2018. Can GDP measurement be further improved? Data revision and reconciliation. Working paper.

Jacobs, Jan P. A. M., Simon van Norden, 2010. Lessons from the latest data on U.S. productivity. Working Papers 11-1, Federal Reserve Bank of Philadelphia.. https://www.philadelphiafed.org/-/media/research-and-data/publications/working-papers/2011/wp11-1.pdf

Jacobs, Jan P.A.M. and Simon van Norden. 2016. "Why are initial estimates of productivity growth so unreliable?" *Journal of Macroeconomics* 47, pp. 200-213. https://doi.org/10.1016/j.jmacro.2015.11.004

Kim, Donald D., Erich H. Strassner, and David B. Wasshausen. 2014. Industry Economic Accounts Results of the Comprehensive Revision Revised Statistics for 1997–2012. https://apps.bea.gov/scb/pdf/2014/02%20February/0214_industry%20economic%20accounts.pdf

Koh, Dongya, Raul Santaeulalia-Llopis, and Yu Zheng. 2020. Labor Share Decline and Intellectual Property Products Capital. *Econometrica*, Volume 88, Issue 6, pp. 2609-2628

Kelly, Pamela A.; Stephanie H. McCulla; David B. Wasshausen. Sept 2018. Improved Estimates of the National Income and Product Accounts: Results of the 2018 Comprehensive Update. *Survey of Current Business*. https://apps.bea.gov/scb/2018/09-september/0918-nipa-update.htm

Manski, Charles F. 2015. "Communicating Uncertainty in Official Economic Statistics: An Appraisal Fifty Years after Morgenstern." *Journal of Economic Literature* 53(3), pp. 631-653.

Morgan, Ted; Dave Wasshausen. March 2014. Results of the 2014 Comprehensive Revision of the Industry Economic Accounts, 1997-2012. Industry Accounts Users' Conference.

Moulton, Brent R.; Benjamin D. Cowan. 2016. Residual Seasonality in GDP and GDI. *Survey of Current Business*. 1-6. https://apps.bea.gov/scb/pdf/2016/07%20July/0716_residual_seasonality_in_gdp_and_gdi.pdf

Robertson, Ken. 2017. Benchmarking the Current Employment Statistics survey: perspectives on current research. *Monthly Labor Review*. https://www.bls.gov/opub/mlr/2017/article/benchmarking-the-current-employment-statistics-survey-perspectives-on-current-research.htm

Rudebusch, Glenn D.; Daniel Wilson; Tim Mahedy. 2015 The Puzzle of weak first-quarter GDP growth. FRBSF Economic Letter. 2015-20165. (May 18, 2015)

Sinclair, Tara M. and H. O. Stekler. 2013. Examining the quality of early GDP component estimates. *International Journal of Forecasting* 29, 736-750.

Sprague, Shawn. 2017. Estimating hours worked for U.S. labor productivity measurement. Internal BLS/OPT memo.

U.S. Bureau of Labor Statistics. "Productivity Measures: Business Sector and Major Subsectors." *Handbook of Methods*. https://www.bls.gov/opub/hom/msp/pdf/msp.pdf

U.S. Bureau of Labor Statistics, Office of Productivity and Technology. 2017. https://www.bls.gov/news.release/prod2.htm as of Dec 2017. See spreadsheet named "DMSP PFEI current practice - BUS RN 2017.M11.xlsx" for data and computation.

United Kingdom, Office of National Statistics, Section on "Validation and quality assurance" at https://www.ons.gov.uk/economy/grossdomesticproductgdp/methodologies/grossdomesticproductgdpqmi#output-quality

Young, Alan H. 1993. Reliability and accuracy of the quarterly estimates of GDP. *Survey of Current Business* 73 (October): 29-43.

Young, Alan H. 1996. Reliability and accuracy of quarterly GDP estimates: a review. In *The New System of Economic Accounts*, edited by John W. Kendrick. pp 423-449.