BLS WORKING PAPERS

U.S. Department of Labor
U.S. Bureau of Labor Statistics
Office of Productivity and Technology

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

John Glaser, Peter B. Meyer, Jay Stewart, and Jerin Varghese,
U.S. Bureau of Labor Statistics

Working Paper 572
February 16, 2024

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

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

February 16, 2024

JEL codes: E24, O47, C82

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

* Corresponding author.

† We thank colleagues for their valuable advice: Victoria Battista, Barbara Fraumeni, Tara Sinclair, Brady West, Shelly Smith, Baoline Chen, Wendy Martinez, Brent Moulton, Terrance Savitsky, Shawn Sprague, John Stewart, Leo Sveikauskas, Shirley Tsai, and an anonymous reviewer from the Bureau of Economic Analysis. Kendra Asher did essential data work. Views expressed here are ours and do not necessarily reflect those of the BLS.

Abstract

This paper examines long-term revisions to official estimates of quarterly US labor productivity growth, and to its components output and hours growth, for 2000-2015. Estimates of output (GDP) and hours growth are revised substantially in the first months after the reference quarter. The data continue to be revised long after the end of the reference quarter, although the magnitudes of the revisions are negligible after 5 years. Revisions are due to the incorporation of additional microdata, benchmarking, adjustments to seasonal factors, and (for output) changes to definitions and methods—all of which are assumed to bring the estimates closer to "truth."

We find that revisions to output growth are substantially larger than revisions to hours growth and that the magnitude of revisions varies across reference quarters, with revisions being larger for Q1 and for recession quarters. Long-term revisions to growth rates tend to be smaller than revisions to levels because revisions to current quarter and prior quarter levels tend to be in the same direction and of approximately the same magnitude.

Following earlier research, we estimate Mincer-Zarnowitz regressions to examine whether these long-term revisions are "news" and whether they eliminate "noise." We find that the initial revisions to output and hours are news, while later revisions are not. Early revisions eliminate noise only for hours, while later revisions do not eliminate noise. These findings seem to contradict the assumption that revisions bring estimates closer to the "true" value. But further investigation resolves this apparent inconsistency. We also examine the ability of the early estimates to "predict" estimate values after five years of revisions.

1. Introduction

Labor productivity (output per hour) is an important source of economic growth, and the Bureau of Labor Statistics' estimates of quarterly labor productivity growth are critical inputs to the decision-making process of policymakers. For example, they are used by the Federal Reserve to assess the state of the economy for monetary policy and by the Congressional Budget Office when developing budget projections. However, because estimates of labor productivity growth are revised over time, the early estimates provide a noisy signal about true productivity growth.

Initial estimates of output growth are based on incomplete data and projections and are substantially revised as more and better data become available. Revisions to hours are typically much smaller than revisions to output and are mainly due to the receipt of additional survey data and to the benchmarking of employment data to the Quarterly Census of Employment and Wages (QCEW). Both output and hours are revised when seasonal factors are recalculated.

These revisions reflect "transitory uncertainty," in that the estimates improve in quality as more data become available (Manski, 2015). This differs from sampling error, conceptual uncertainty, or changes in definition, although there are occasional definitional changes in the form of periodic Comprehensive Revisions to GDP by the Bureau of Economic Analysis (BEA).

This paper examines long-run revisions to the official estimates of quarterly U.S. nonfarm business sector productivity growth for the 2000-2015 period. Previous research has examined short run revisions—those between the initial estimate and the second revised estimate, which is released 2 months later—to GDP (Sinclair and Steckler, 2013) or labor productivity (Asher, et al., 2022). Other papers have examined longer-run revisions (Fixler and Grimm, 2002, 2008; Aruoba, 2008; Fixler, Greenaway-McGrevy, and Grimm, 2011, 2014; Jacobs and van Norden, 2016; Fixler, Kanal, and Tien, 2018; and Jordan, et al., 2020). These papers use the latest vintage of the data and calculate the revisions as the difference between the latest estimate for each reference quarter and the initial estimate. Thus, estimates for more recent quarters have been revised fewer times than reference quarters that are further in the past.

Our paper differs from these by comparing estimates that have been revised the same number of times. In addition, because the nature of revisions changed between the late 1990s and the period after 1999, we examine estimates from more recent reference quarters rather than the entire series as other researchers have done. We focus our analysis on revisions to quarterly growth rates of labor productivity for the nonfarm business sector and its components: output and hours worked. We examine the sources, magnitudes, and distributions of revisions, as well as how the estimates change over time. We also follow earlier analysis by Jacobs and van Norden (2016) to determine whether the revisions are "news" and whether they eliminate "noise." Finally, we examine the extent to which early estimates of labor productivity predict later estimates.

We find that revisions to estimates of labor productivity growth can be large and are driven primarily by revisions to output growth. Revisions to estimates of hours growth become small after 2 years, while revisions to estimates of output growth (and labor productivity growth) become small after about 5 years. Interestingly, revisions to output and hours growth tend to be in the same direction, which mitigates the impact on productivity growth, although revisions to output and hours are only weakly correlated. Thus, there is substantial variability in the size of revisions to labor productivity growth. We find that estimates of output growth tend to be revised downward, with the revisions being largest for the Q1 and Q3 reference quarters. Estimates of hours growth tend to be revised downward slightly, with the largest revisions for Q1. Virtually all of the revisions to hours growth occur in the 2 years after the reference quarter.

Our findings are broadly consistent with the findings of Aruoba (2008) that productivity revisions are partly predictable, and of Jacobs and van Norden (2016) that revisions to productivity are large because revisions to output and labor inputs are not highly correlated. Following Jacobs and van Norden, we estimated Mincer-Zarnowitz regressions over our sample period to determine whether the revisions are "news" and whether they eliminate "noise." We found that the early revisions to output, hours and labor productivity are news, while early revisions eliminate noise only for hours. Like Jacobs and van Norden, we found that subsequent revisions are not news and do not eliminate noise, which seems to contradict the assumption that revisions bring the estimates closer to the true value. However, by taking a closer look at the later revisions, we resolve this apparent contradiction.

The rest of the paper is organized as follows. Section 2 describes the revisions to output and hours. Section 3 shows how revisions evolve over time, examines whether the revisions to labor productivity growth are due to revisions to current or prior quarter output or hours, and tests whether revisions are news or noise. Section 4 examines the role of BEA annual revisions to output. The next two sections (5 and 6) examine whether early estimates are good predictors of later values and generate "prediction" intervals. Sections 7 and 8 examine the effects of comprehensive revisions and revisions in the COVID-19 era. The final section concludes.

2. Why are Labor Productivity Estimates Revised?

Labor productivity (LP) is defined as:

$$Labor\ Productivity\ (LP)\ index\ = \frac{Output\ index}{Hours\ Worked\ index}$$

Throughout our analysis, we use seasonally adjusted annualized quarterly growth rates for all three variables. Labor productivity growth can be approximated as:[1]

$$Quarterly\ LP\ growth \approx Output\ growth\ -\ Hours\ Worked\ growth$$

For each reference quarter, BLS releases three regularly scheduled estimates of labor productivity growth that are released about one week after BEA releases the corresponding GDP estimates. The "preliminary" estimate (R0) is released within 40 days of the end of the reference quarter; the first revised estimate (R1) is released 30 days after R0; and the second revised estimate (R2) is released 60 days after R1. Subsequent revisions, including annual revisions and those due to BEA Comprehensive Revisions, can occur long after the reference quarter.

The output index is constructed from GDP data for the nonfarm business sector, which comprises about 75 percent of GDP. This output measure excludes general government, most non-profits, and private households because output for these sectors are not measured directly but rather derived mainly from data on inputs. The hours-worked data cover the same sectors and are compiled primarily from three BLS surveys: the Current Employment Statistics (CES) survey, the Current Population Survey (CPS), and the National Compensation Survey (NCS).[2]

Early estimates of output growth are based partly on preliminary source data and projections. These include survey, tax, and administrative data as well as indicators such as heating degree days.[3] These estimates are revised as actual output data become available and when seasonal factors are recalculated. Annual revisions to the previous 3 years of GDP

[1] The Labor Productivity and Costs program calculates labor productivity as the percentage change in the index of output divided by the index of hours worked, where the indexes have the same base year. This growth rate is then converted to an annual rate.
[2] In addition, several minor adjustments are based on other data sources. A description of the BLS methodology can be found in U.S. Bureau of Labor Statistics *Handbook of Methods*, and Eldridge, Sparks, and Stewart (2018).
[3] See Fixler, Greenaway-McGrevy, and Grimm (2014) and Fixler, Kanal, and Tien (2018).

estimates were published each July in our sample period.[4] However, the GDP data are never "final" because some of the data sources do not become available until long after the end of the reference period and because GDP estimates are subject to comprehensive revisions where the definition of output can change.

Revisions to the hours data are mainly due to revisions to estimates of employment, and to a lesser extent weekly hours, from the CES. There are three regularly scheduled releases for each reference month. The first estimates are usually released on the first Friday after the reference month, and the second and third estimates are released at the same time in the following 2 months.[5] Thus, a large fraction of the early revisions to the hours data are reflected in the preliminary (R0) estimate of labor productivity growth, and all of the early revisions are reflected in the second revised (R2) estimate. Subsequent revisions to hours-growth estimates are due to benchmark revisions and recalculation of seasonal adjustment factors. Each February the 21 months of (not-seasonally adjusted) CES employment data are benchmarked to data from the Quarterly Census of Employment and Wages (QCEW), and seasonal adjustment models are updated.[6] The NCS data are never revised, and the only revisions to CPS data are due to the recalculation of seasonal factors.[7]

---

[4] Starting in 2019, annual revisions cover the previous 5 years.

[5] More precisely, the Employment Situation report is typically released on the third Friday after the conclusion of the reference week, or the week which includes the 12th of the month.

[6] In a recent example, from the January 2022 Employment Situation news release, CES employment for March 2021 was benchmarked to QCEW totals, and employment estimates for April 2020 – February 2021 and for April 2021 – December 2021 were revised as follows:

(a) For employment estimates for April – December 2021: (1) the model used to forecast business births and deaths was re-estimated, (2) the monthly employment estimates were recalculated using the original sample-based growth rates from the new March 2021 base and the revised employer birth-death forecasts, and (3) seasonal factors were re-estimated. The employment estimates for November and December also incorporate additional sample collected into their sample-based growth rates as part of the regularly scheduled releases (the third release for November and the second release for December).

(b) For employment estimates for April 2020 – February 2021: (1) The adjustment to March 2021 employment was distributed linearly to the previous 11 months, and (2) seasonal factors were re-estimated.

Thus, except for March, each month's employment estimate is revised twice. After the second benchmark revision, the only revisions are due to the re-estimation of seasonal factors, which are recalculated for the previous five years (January 2017 through December 2021).

These revisions cover the 21-month period from April 2020 through December 2021. It is not necessary to revise the January-March 2022 data because employment for those months is estimated using sample-based growth rates from the revised December 2021 estimate. That is, the initial estimates for these months are calculated the same way as the revised estimates for April-December 2021 as in (a). More information on CES benchmarking can be found at the Benchmark section of the CES Handbook of Methods.

[7] The CES is the main source of hours data. The NCS data are used to convert the CES hours data from an hours-paid basis to an hours-worked basis. During our sample period, the CPS data were used to estimate hours worked for supervisory and nonproduction workers and for the self-employed. The LPC program uses NCS data for the

Revisions to labor productivity growth are often greater than revisions to GDP growth for two reasons. First, the 25 percent of output that is excluded (government, private households, and most non-profits) comes from administrative and compensation data, which are subject to much smaller revisions. Second, labor productivity estimates also incorporate revisions to labor hours.

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, 2022). We found that revisions had no significant trend over time, that there was no relationship between the magnitude of the initial estimate and the subsequent revisions, and that there were no significant business cycle effects. The magnitude of revisions varied across quarters, but not statistically significantly. 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.

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

Our first step is to examine the path of revisions and determine when revisions have become small enough for estimates to be considered final. Let reference quarters be numbered from $t = 1$ to T, k be the revision number (0-80, where 0 indicates the preliminary estimate) for that reference quarter, and X_t^k be the estimates for output, hours, or labor productivity for reference quarter t as of revision k. Then:

$$Absolute\ revision\ for\ release\ k = \left|X_t^k - X_t^0\right|$$

$$Mean\ absolute\ revision\ for\ release\ k = \frac{1}{T}\sum_{t=1}^{T}\left|X_t^k - X_t^0\right|$$

Thus, the mean absolute revision for release k is an average of the 64 reference quarters in our sample.

---

fourth quarter and allocates annual changes to quarters using the Denton procedure. Thus, seasonal adjustment is not necessary. See the U.S. Bureau of Labor Statistics *Handbook of Methods*, and Eldridge et al. (2018) for a more detailed description. BLS introduced a new method for estimating hours worked (see Eldridge et al., 2022). This change should not affect our results because virtually all of the revisions to total hours growth are due to revisions to estimates of employment from the CES, which has not changed.

Figure 1 shows the mean of the absolute value of the revisions between R0 and R*k* for *k* = 1,…,80 for each of our three variables, using the subsample of reference periods for which we have 10 years of data on revisions (2000q1-2010q4).[8]

Revisions to output, hours and productivity are the largest in the first few years after the initial R0 release, and revisions to hours are smaller than revisions to output and productivity. By R40, the absolute value of revisions to output and labor productivity have stabilized at around 1.7 percentage points, while the absolute value of revision to hours has stabilized at around 0.9 percentage points by R16. Because the estimates stabilize by R40 we treat these estimates as "final." For the rest of the paper, we use the sample for which we have 5 years of revisions for data on reference quarters from 2000q1 through 2015q4.

Table 1 and Figure 2 show summary statistics and distributions for the R0-to-R2, R2-to-R40 and R0-to-R40 revisions to estimates of labor productivity growth and its components and provide insight into how these distributions evolve over time.

Table 1 shows that R0-to-R2 revisions to LP tend to be positive (0.13 of a percentage point on average), whereas the R2-to-R40 revisions tend to be negative and somewhat larger in magnitude (−0.45) and net out to R0-to-R40 revisions that are equal to −0.32 on average. The pattern of these revisions is largely driven by the pattern of revisions to output, as the revisions to hours are very small on average. The larger mean revisions for R2-to-R40, compared to R0-to-R40 revisions, suggests that R0 estimates are better than R2 estimates in that the bias is smaller. However, variance matters too. To examine this further, we calculated the mean squared revision (MSR) relative to the R2 estimate (for R0-to-R2 revisions) and to the R40 estimate (for R2-to-R40 and R0-to-R40 revisions). The third column shows that the MSR for R2-to-R40 revisions is smaller than the MSR for R0-to-R40 revisions. Using this metric, the R2 estimate is a better predictor of the R40 value. We explore this further below.

[8] We considered extending our sample period back to 1995Q1. However, there were two important changes to the data between the late 1990s and the early 2000s. In 1999, BEA expanded the definition of output to include own-account software, which increased GDP levels significantly. The revisions to growth rates were small, but it seems likely that the large comprehensive revision distorted the revisions that we focus on here. In addition, revisions to hours levels were considerably larger in the late 1990s compared with 2000 forward. A possible reason for this is the CES conversion from a quota sample to a probability sample.

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

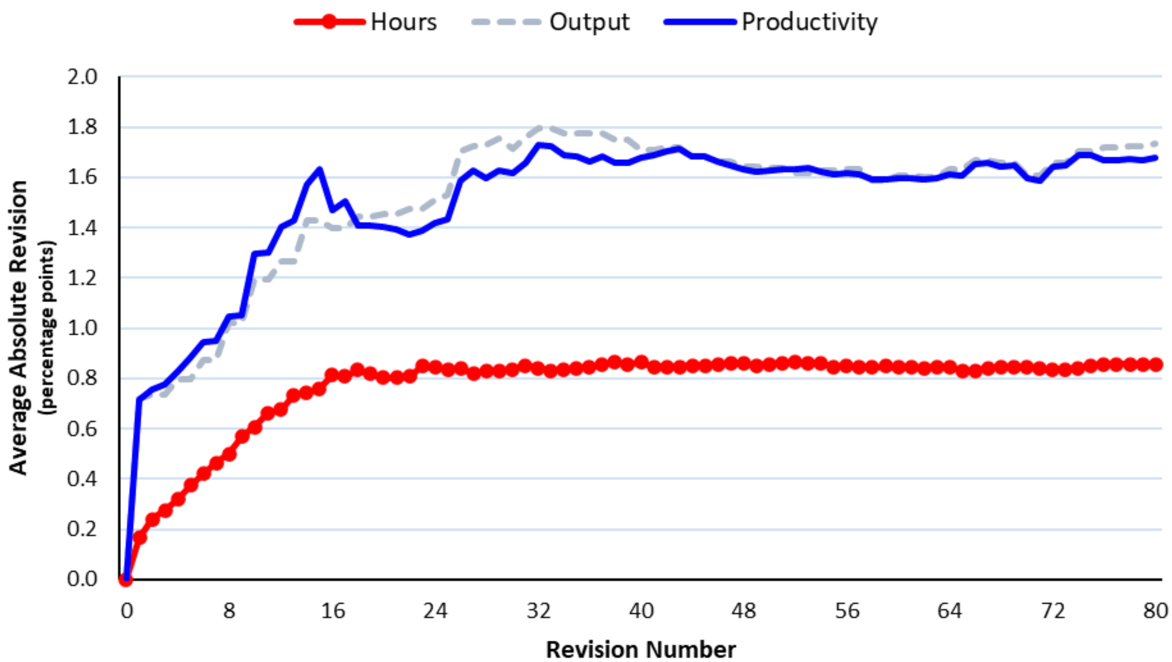


Table 1: Summary Statistics of Revisions to Growth Rates (in percentage points)

LP	Mean	Standard Deviation	Mean Squared Revision	Skewness	Kurtosis
R0 - R2	0.13	1.15	1.34	-1.05	4.50
R2 - R40	-0.45	1.90	3.80	0.06	2.81
R0 - R40	-0.32	2.17	4.79	-0.43	2.98
Output					
R0 - R2	0.08	1.13	1.29	-0.79	3.90
R2 - R40	-0.61	1.88	3.92	-0.12	2.66
R0 - R40	-0.53	2.13	4.81	-0.69	3.37
Hours					
R0 - R2	-0.05	0.27	0.08	-0.15	3.15
R2 - R40	-0.15	0.80	0.66	-0.09	2.46
R0 - R40	-0.20	0.90	0.84	-0.40	2.78

Figure 2: Distributions of Revisions to Labor Productivity, Output and Hours as of R0, R2, and R40

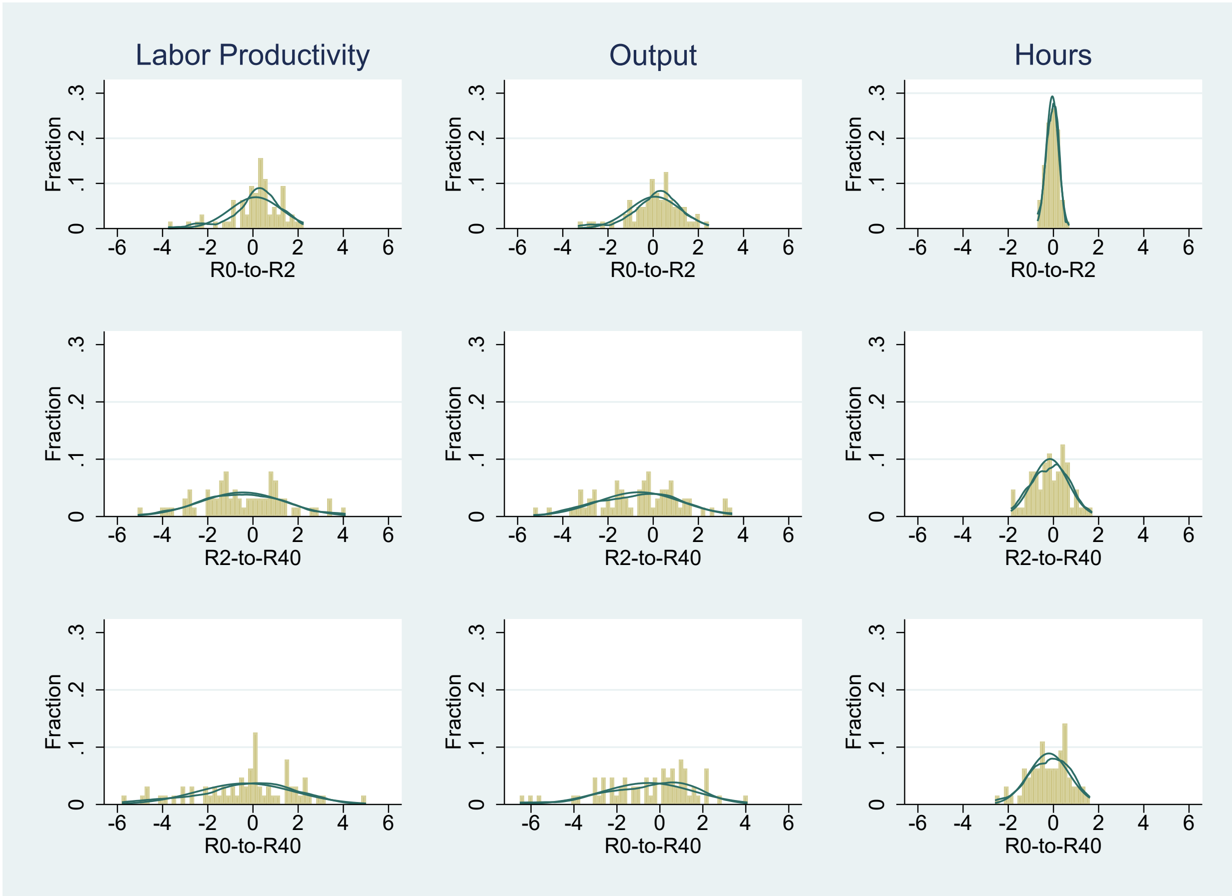
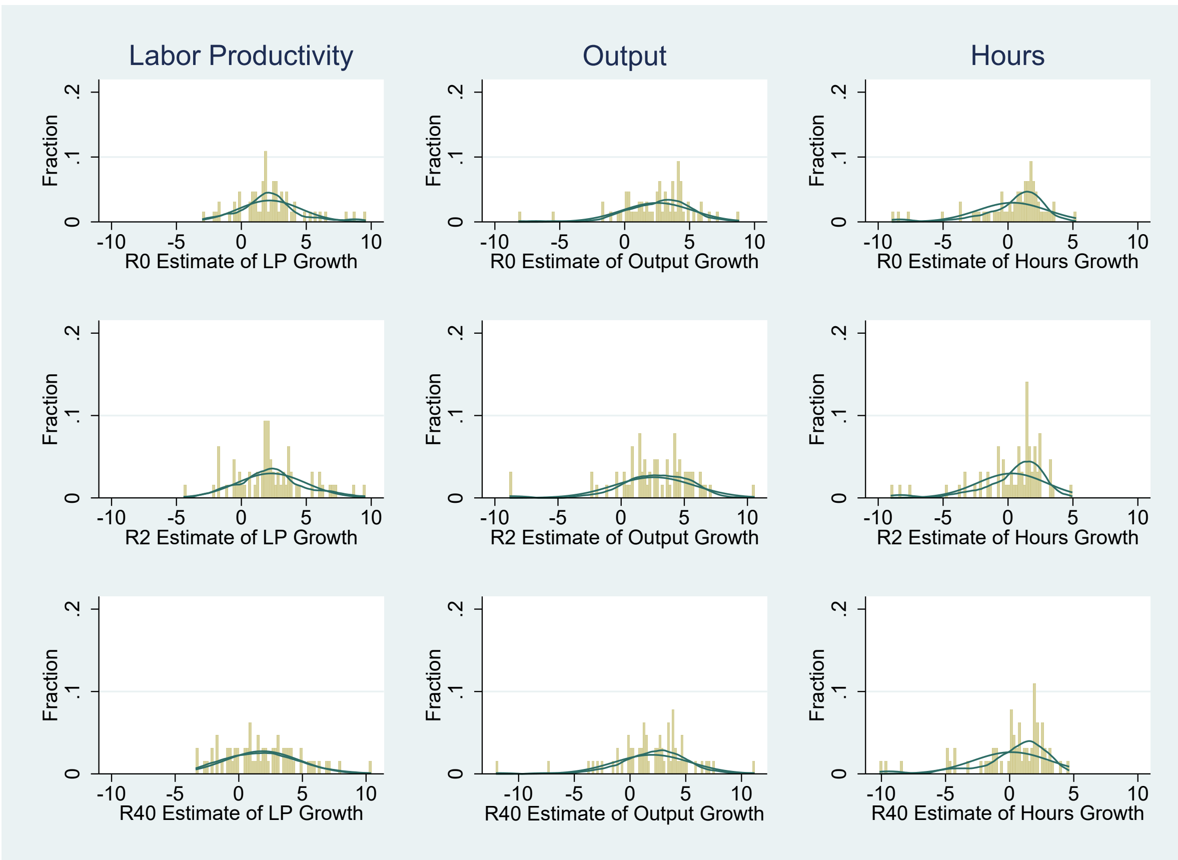


Figure 2 shows the histograms of the different revisions to LP, output, and hours, along with smoothed estimates of the distributions and a normal distribution for comparison. From the figure and the distribution statistics in Table 1, we can see that the distribution of R0-to-R2 revisions to output is slightly more peaked than a normal distribution and is left skewed. The tightness of the distribution of R0-to-R2 revisions to hours illustrates the fact that early revisions to hours are largely reflected in the R0 estimate. The combined effect is that the early revisions (R0-to-R2) LP growth are not normally distributed. The distribution is more peaked than a normal distribution and is left skewed. Appendix C shows statistical tests of normality of these distributions.

The distributions of R2-to-R40 revisions are more spread out and much closer to normal for all three measures. This could reflect the types of revisions. Early revisions to output substitute actual data for projections and proxies, whereas later revisions are due to changes in source data and methods, and the updating of seasonal factors. For hours data, the early revisions are due mainly to the collection of additional data, while the later revisions are due to benchmarking and revision of seasonal factors.

Figure 3: Distributions of Estimates of Labor Productivity, Output, and Hours as of R0, R2, and R40



The net effect of the early and later revisions is that the R0-to-R40 revisions to all three measures are more spread out and are approximately normally distributed, although the distributions of revisions to output follow an interesting pattern. The distribution of R0-to-R2 revisions is left skewed and is more peaked than a normal distribution. The R2-to-R40 revisions are much more spread out and are approximately normally distributed. The net result of these two sets of revisions is that the distribution of R0-to-R40 revisions to output is a spread-out version of the distribution of R0-to-R2 revisions.

The progression of the distribution of revisions to estimates of hours growth follow a similar pattern except that the distribution of R0-to-R40 revisions looks very similar to the distribution of R2-to-R40 revisions; most revisions to hours occur before the R2 estimate.

Figure 3 shows how the progression of revisions changes the distribution of estimated growth rates. The early estimates are more peaked around the mean compared to a normal distribution. But the distribution of R40 estimates is close to normal. For labor productivity growth, the distributions of R0 and R2 estimates look fairly similar in that both have spikes at around 2 percentage points. In contrast the R40 distribution has fewer and smaller spikes and more closely resembles a normal distribution. This is interesting, because the R40 distributions of output and hours are left skewed, with the distributions of output being less skewed. Thus, combining the left-skewed R40 distributions of output and hours results in a distribution of labor productivity that is close to normal.

Sources of Revisions

It is useful to know whether revisions to growth rates are due to revisions to current or prior quarter data and the extent to which these revisions vary by quarter. To shed additional light on these questions, we decomposed revisions to labor productivity growth into revisions to current and prior quarter output and hours.

To simplify the decompositions, we express labor productivity growth as the difference in the natural logs of the output and hours indexes:

$$Labor\ Productivity\ Growth \approx [\ln(Q_t) - \ln(Q_{t-1})] - [\ln(H_t) - \ln(H_{t-1})],$$

where Q and H are indexes of real output and total hours worked, and the subscripts indicate the reference quarter. With this specification, R0-to-R40 revisions to labor productivity can be written as:

$$Revision\ to\ LP\ Growth = \{[\ln(Q_t^{R40}) - \ln(Q_{t-1}^{R40})] - [\ln(H_t^{R40}) - \ln(H_{t-1}^{R40})]\} - \{[\ln(Q_t^{R0}) - \ln(Q_{t-1}^{R0})] - [\ln(H_t^{R0}) - \ln(H_{t-1}^{R0})]\}$$

where the superscripts indicate the release. This equation can be rewritten as:

$$(1)\quad Revision\ to\ LP\ Growth = [\ln(Q_t^{R40}) - \ln(Q_t^{R0})] - [\ln(Q_{t-1}^{R40}) - \ln(Q_{t-1}^{R0})] - [\ln(H_t^{R40}) - \ln(H_t^{R0})] + [\ln(H_{t-1}^{R40}) - \ln(L_{t-1}^{R0})].$$

Other revisions to LP (R0-to-R2 and R2-to-R40) can be decomposed similarly. This equation illustrates the four sources of revisions: the first term is the amount of the revision that can be attributed to revisions to current quarter output, the second is the contribution of revisions to prior quarter output, and the last two terms are the analogous measures for revisions to hours. As before, our data cover the period from 2000q1 through 2015q4.

Tables 2a, 2b, and 2c show the average values for each term in equation (1) for R0-to-R2, R2-to-R40, and R0-to-R40 revisions and for the average revision to quarterly LP growth. We multiplied the growth rates of output and hours by 4 so that quarterly changes are consistent with the annualized growth rates reported elsewhere. The mean revisions to the growth rates of output, hours, and labor productivity in the third, sixth, and seventh columns of the tables are the same (except for rounding) as mean revisions to growth rates reported in Table 1.

The first three columns of each table show the revisions to current quarter output, previous quarter output, and the growth rate (the difference between the first two columns multiplied by 4). The next three columns present the same information for hours revisions. The last two columns show the revisions to LP growth (the difference between the two Growth Rate columns) and the mean squared revision for LP growth.

As noted earlier, the R0-to-R2 revisions to hours are driven mainly by the collection of additional CES data for the current quarter. Revisions to previous quarter hours are due to the updating of seasonal factors, as data collection for the previous quarter was complete at the time of the R0 estimate. Table 2a shows that revisions to current quarter hours levels tend to be larger than revisions to previous quarter levels, although both are small.

As we saw in Table 1, the R0-to-R2 revisions to output growth are small on average, although there is also a fair amount of variation across quarters. Table 2a shows that this small revision is the net effect of relatively small negative revisions to current and previous quarter output levels that are of approximately the same magnitude. This table also sheds light on how the size and sign of revisions vary by reference quarter. In Table 2a, we see that the small 0.09 average revision to output growth masks variation across reference quarters. The R0-to-R2 revisions to output levels are significantly larger for Q1 than for the other quarters. We have more to say about this in the next section.

Table 2a: Decomposition of R0-to-R2 Revisions, 2000 - 2015

	Average Revision to:						Labor Productivity Growth	
	ln(Output)			ln(Hours)				
	Current Quarter	Previous Quarter	Growth Rate (x4)	Current Quarter	Previous Quarter	Growth Rate (x4)	Average Revision	Mean Squared Revision
All Quarters	-0.09	-0.11	0.09	-0.01	0.00	-0.03	0.12	1.31
Q1	-0.51	-0.41	-0.39	-0.05	-0.04	-0.04	-0.35	1.57
Q2	0.06	-0.04	0.42	-0.03	0.00	-0.13	0.54	0.82
Q3	0.10	0.01	0.39	0.09	0.06	0.13	0.26	1.43
Q4	0.00	0.01	-0.06	-0.06	-0.04	-0.10	0.04	1.65

Table 2b: Decomposition of R2-to-R40 Revisions, 2000-2015

	Average Revision to:						Labor Productivity Growth	
	ln(Output)			ln(Hours)				
	Current Quarter	Previous Quarter	Growth Rate (x4)	Current Quarter	Previous Quarter	Growth Rate (x4)	Average Revision	Mean Squared Revision
All Quarters	0.26	0.41	-0.61	0.13	0.17	-0.16	-0.45	3.80
Q1	0.19	0.41	-0.89	0.16	0.27	-0.45	-0.44	6.86
Q2	0.56	0.55	0.05	0.17	0.18	-0.04	0.09	3.69
Q3	0.19	0.51	-1.28	0.10	0.11	-0.05	-1.23	3.44
Q4	0.10	0.18	-0.34	0.11	0.13	-0.10	-0.24	1.88

Table 2c: Decomposition of R0-to-R40 Revisions, 2000 - 2015

	Average Revision to:						Labor Productivity Growth	
	ln(Output)			ln(Hours)				
	Current Quarter	Previous Quarter	Growth Rate (x4)	Current Quarter	Previous Quarter	Growth Rate (x4)	Average Revision	Mean Squared Revision
All Quarters	0.17	0.31	-0.52	0.12	0.17	-0.19	-0.33	4.78
Q1	-0.31	0.00	-1.28	0.11	0.24	-0.49	-0.79	8.10
Q2	0.62	0.50	0.47	0.13	0.18	-0.17	0.63	3.19
Q3	0.30	0.52	-0.89	0.19	0.17	0.09	-0.97	4.49
Q4	0.09	0.19	-0.40	0.04	0.10	-0.20	-0.20	4.18

The top lines of Tables 2b and 2c show that long-run revisions to current and previous quarter estimates of output and hours levels are positive and largely offset each other. But since revisions to the previous quarters' levels are larger, the net effect is a negative revision to the growth rates of output and hours. As in Table 2a, revisions to output growth are larger than revisions to hours growth, which results in downward revisions to LP growth, although these later revisions are larger than the early revisions.

Table 2b shows that there is considerably more variation across reference quarters in the size of later revisions. The revisions to output levels are smallest for Q4 estimates, although the revisions to Q2 growth rates are the smallest. Revisions to output growth are largest in magnitude for Q3 and Q1. As expected, revisions to hours and hours growth for each reference quarter are much smaller than the corresponding revisions to output. The R2-to-R40 revisions to hours levels are larger than the R0-to-R2 revisions because they include two benchmark revisions in addition to the recalculation of seasonal factors (see footnote 5 for a description). Like the BEA comprehensive revisions, the benchmarking process (both first and second benchmarks) generally results in small revisions to growth rates because both current and previous quarter estimates are revised.

The larger revisions to output levels are mainly due to annual revisions to GDP and the 2013 and 2018 comprehensive revisions to GDP. It seems likely that the 2013 comprehensive revision is the main culprit, because it added research and development and artistic originals as capital assets (and therefore as output). This addition increased the level of GDP in all periods but had relatively small effects on growth rates. We have more to say about comprehensive revisions in Section 7. Revisions to hours levels are much smaller than the revisions to output levels. But like output revisions, the revisions to current and previous quarter hours largely offset each other.

The net effect of the R0-to-R2 and R2-to-R40 revisions can be seen in Table 2c. The largest revisions to output growth are for Q1, with all of the revisions being due to revisions to current quarter output. In contrast, the large revisions to Q3 output growth are due to larger revisions to prior quarter output.

Comparing the MSR values for R2-to-R40 revisions by quarter yields some interesting insights. They suggest that estimates for Q4 are fairly reliable in that both the MSR and the average revision are small. The large MSR for Q1 implies that early estimates are considerably less reliable. The MSRs for Q2 and Q3 estimates are similar to each other. The Q2 estimates are low bias (smaller revisions) and high variance, while the reverse is true for Q3 estimates.

Are Revisions News or Noise?

So far, we have assumed that revisions add information and bring estimates closer to their later R40 values. Following the analysis in Jacobs and van Norden (2016), we consider two approaches to further examine this issue. We first look at the noise-to-signal ratio and then perform Mincer-Zarnowitz tests.

Reliability – Noise/signal ratios

Jacobs and van Norden (2016) calculate the noise-to-signal ratio for labor productivity and decompose it into the contributions of output, hours, and a cross-moment term. We calculated the same noise-signal ratios. The two main differences between our estimates and theirs are: (1) we restricted our analysis to the 2000-2015 period, rather than the entire series; and (2) our final values are the R40 estimates rather than the current estimates. We also present results for R2-to-R40 revisions in addition to those for R0-to-R40 revisions.

The equation for the Jacobs and van Norden noise-signal ratio (squared) for labor productivity (ϕ_{LP}^2) equation (using their notation) is given by:

$$\phi_{LP}^2 = \left[\frac{\sum R_Y^2}{\sum(LP-\overline{LP})^2} \times \frac{\sum(Y-\bar{Y})^2}{\sum(LP-\overline{LP})^2}\right] + \left[\frac{\sum R_H^2}{\sum(LP-\overline{LP})^2} \times \frac{\sum(H-\bar{H})^2}{\sum(LP-\overline{LP})^2}\right]$$
$$+ \left[-2 \cdot \frac{\sum R_H^2 \cdot R_Y^2}{\sum(H-\bar{H})(Y-\bar{Y})} \times \frac{\sum(H-\bar{H})(Y-\bar{Y})}{\sum(LP-\overline{LP})^2}\right]$$

where LP, Y, and H denote the R40 estimates of labor productivity, output, and hours; and R_x denotes revisions to variable x (=LP,Y,H). The bars indicate mean values. Our estimates are:

$0.56 = [0.40 \times 1.42] + [0.09 \times 1.09] + [(-0.15) \times 0.77]$ for R0-to-R40 revisions.
$0.45 = [0.33 \times 1.42] + [0.07 \times 1.09] + [(-0.12) \times 0.77]$ for R2-to-R40 revisions.
Output Hours Cross-moment

The main contributor to the noise-to-signal ratios for LP is output. The noise-to-signal ratios for output are larger than those for hours (0.40 and 0.33 for R0-to-R40 and R2-to-R40 revisions vs. 0.09 and 0.07 for revisions to hours). Both of these ratios are smaller than the noise-to-signal ratio for LP. But the final (R40) values of both output and hours growth are also relatively more variable than final values of LP growth, which results in "weights" that are greater than one. Our noise-signal ratios for LP are similar to the first-to-fifth year revision in Jacobs and van Norden—0.56 and 0.45 vs. 0.51—for early estimates for the nonfarm business sector. The lower value of the noise-signal ratio for R2-to-R40 revisions implies that the R2 estimate is an improvement over the R0 estimate, with most of the improvement coming from the output term.

Noise-versus-news regressions

We now turn to Mincer-Zarnowitz news vs. noise regressions. Following Jacobs and van Norden (and again using their notation), we estimate the following regressions:

"Noise" $$y_t^{Rn+} - y_t^{Rn} = \alpha_1 + \beta_1 y_t^{Rn+} + \nu_t$$

"News" $$y_t^{Rn+} - y_t^{Rn} = \alpha_2 + \beta_2 y_t^{Rn} + \varepsilon_t$$

where y_t^{Rn} and y_t^{Rn+} denote the initial estimate (R0 or R2) and the final estimate (R2 or R40) of variable *y* (LP, output or hours) for reference quarter *t*. In the "noise" regressions we test the hypothesis that the revisions eliminate noise, meaning that the revisions are independent of the value of the final estimate. In the "news" regressions we test the hypothesis that revisions are news, meaning that the revisions are not predictable by the earlier estimate of the same variable. In both sets of regressions, the null hypothesis holds if $\alpha_n = 0$ and $\beta_n = 0$, where n = 1,2.

Table 3 shows the results for R0-to-R2, R2-to-R40, and R0-to-R40 revisions. The top row of each panel shows the result for revisions to LP. Like Jacobs and van Norden, we reject the hypotheses that both short-term and long-term revisions to labor productivity eliminate noise, although we cannot reject the hypothesis that early revisions to hours eliminate noise.[9]

The results from our "noise" regressions are consistent with those in Jacobs and van Norden. We reject the null hypothesis for the three LP regressions. When we examine the components of LP, output and hours, we see a similar story except that we do not reject the hypothesis that the R0-to-R2 revisions to hours eliminate noise.

In contrast, the results of our "news" regressions for LP differ somewhat from those in Jacobs and van Norden and offer some interesting insights. We do not reject the hypotheses that the short-term (R0-to-R2) and long-term (R0-to-R40) revisions contain news, while the R2-to-R40 regression rejects that hypothesis. From this we conclude that revisions to labor productivity are news and that most of the news is contained in the R0-to-R2 revisions. The R0-to-R2 and R0-to-R40 revisions to output indicate that the revisions are news, whereas only the R0-to-R2 revisions to hours are news. The results suggest that the R2-to-R40 revisions to output and hours are not news and that most of the news is contained in the R0-to-R2 revisions.

[9] Jacobs and van Norden do not estimate these regressions for output and hours.

Table 3: "News" and "Noise" Regressions

	Noise (RHS var. = R2 estimate)			News (RHS var. = R0 estimate)		
R0-to-R2 Revisions	Constant	Slope	Test	Constant	Slope	Test
Labor Productivity	-0.30	0.19	[0.008]	0.12	0.00	[0.682]
	(0.20)	(0.06)		(0.20)	(0.07)	
Output	-0.43	0.20	[0.000]	-0.16	0.10	[0.161]
	(0.19)	(0.05)		(0.22)	(0.06)	
Hours	-0.05	0.00	[0.390]	-0.04	-0.01	[0.256]
	(0.03)	(0.01)		(0.03)	(0.01)	

	Noise (RHS var. = R40 estimate)			News (RHS var. = R2 estimate)		
R2-to-R40 Revisions	Constant	Slope	Test	Constant	Slope	Test
Labor Productivity	-0.98	0.28	[0.001]	-0.06	-0.17	[0.011]
	(0.27)	(0.07)		(0.31)	(0.08)	
Output	-1.05	0.22	[0.001]	-0.38	-0.09	[0.017]
	(0.28)	(0.07)		(0.35)	(0.09)	
Hours	-0.18	0.14	[0.000]	-0.18	0.09	[0.001]
	(0.08)	(0.02)		(0.09)	(0.03)	

	Noise (RHS var. = R40 estimate)			News (RHS var. = R0 estimate)		
R0-to-R40 Revisions	Constant	Slope	Test	Constant	Slope	Test
Labor Productivity	-1.12	0.43	[0.000]	0.08	-0.18	[0.149]
	(0.28)	(0.09)		(0.33)	(0.11)	
Output	-1.28	0.38	[0.000]	-0.52	0.00	[0.150]
	(0.30)	(0.07)		(0.46)	(0.14)	
Hours	-0.22	0.15	[0.000]	-0.23	0.07	[0.013]
	(0.10)	(0.03)		(0.11)	(0.03)	

The results in this subsection seem to suggest that the long run revisions do not add much to our understanding of productivity growth. They do not eliminate much noise, and only the early (R0-to-R2) revisions are news. But BEA's descriptions of their annual revisions make it clear that these revisions really do improve the estimates. After the R2 estimate, revisions are due to annual revisions and comprehensive revisions. The annual revisions, which revise data back 3 years, incorporate new data sources and minor changes to methodology.[10] Comprehensive revisions, which revise the entire series, are more significant and can include changes in concept such as the addition of new types of output.[11] In the next sections, we take a closer look at the revisions.

4. How Do Estimates Change with Revisions? The Role of Annual Revisions to GDP

We have shown that there can be substantial revisions to estimates of LP growth and that most of the revisions are due to revisions to output. Here we examine how estimates of growth rates for a given reference quarter change over time.

Table 4 shows summary statistics for estimates of output, hours, and labor productivity growth over the 2000-2015 sample period as of the R0, R2, and R40 estimates. In the first column we can see that early revisions to output and labor productivity growth increase estimated growth rates slightly, while subsequent revisions decrease estimated growth rates. In contrast revisions to hours slightly decrease estimated growth rates. As might be expected the standard deviations of the estimates are monotonically increasing. It is worth noting that the distributions of the output and hours are right-skewed and more peaked than a normal distribution, but that the distributions of LP estimates are approximately normal.

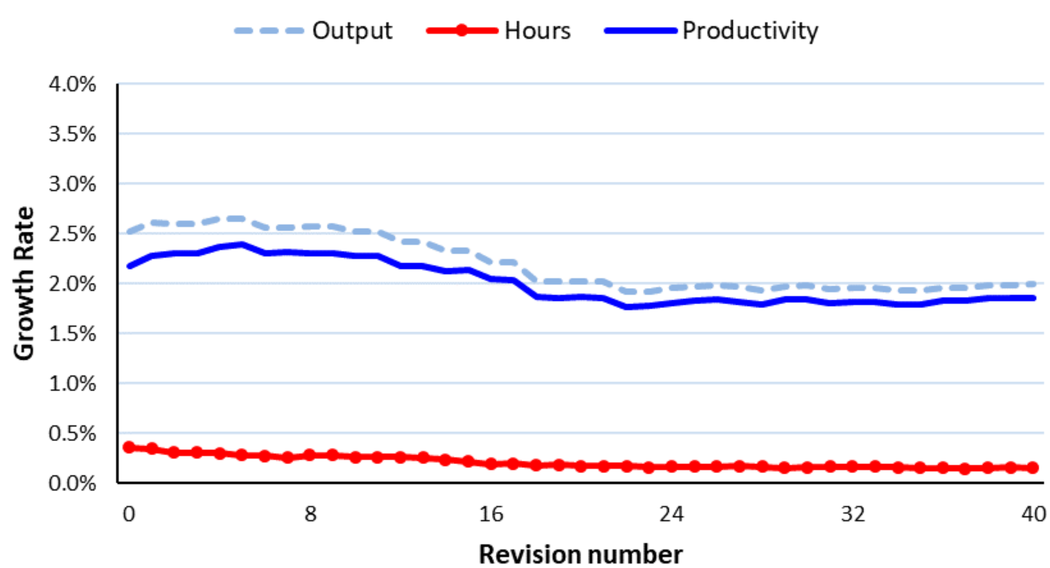
Figure 4 shows how the average estimated growth rates for these variables evolve as they are revised from the initial (R0) estimate to the R40 estimate. The horizontal axis graphs the revision number, while the vertical axis graphs the average percent growth rate from the previous quarter. Each point indicates the average growth rate—calculated for the 64 reference quarters from the 2000-2015 period—of output, hours, and LP as of the indicated release. Thus, at each point estimates for all 64 quarters have been revised the same number of times. The key things to note about Figure 4 is that revisions to labor productivity growth closely follow revisions to output growth because revisions to hours are relatively small. As we saw in Table 4, estimates of output growth, and therefore labor productivity growth, tend to be revised upward initially and then downward.

[10] Starting in 2019, the annual revisions go back 5 years.
[11] A description of the annual and comprehensive revisions can be found here: https://www.bea.gov/information-previous-updates-nipa-accounts.

Table 4: Summary Statistics of Estimates by Release Age

LP	Mean	Standard Deviation	Skewness	Kurtosis
R0 estimate	2.18	2.42	0.65	4.15
R2 estimate	2.31	2.69	0.29	3.32
R40 estimate	1.86	2.90	0.41	2.99
Output				
R0 estimate	2.52	2.73	1.13	6.22
R2 estimate	2.60	3.19	1.21	6.48
R40 estimate	1.99	3.45	1.08	6.74
Hours				
R0 estimate	0.36	2.70	1.65	6.10
R2 estimate	0.31	2.69	1.63	5.96
R40 estimate	0.15	3.03	1.64	5.75

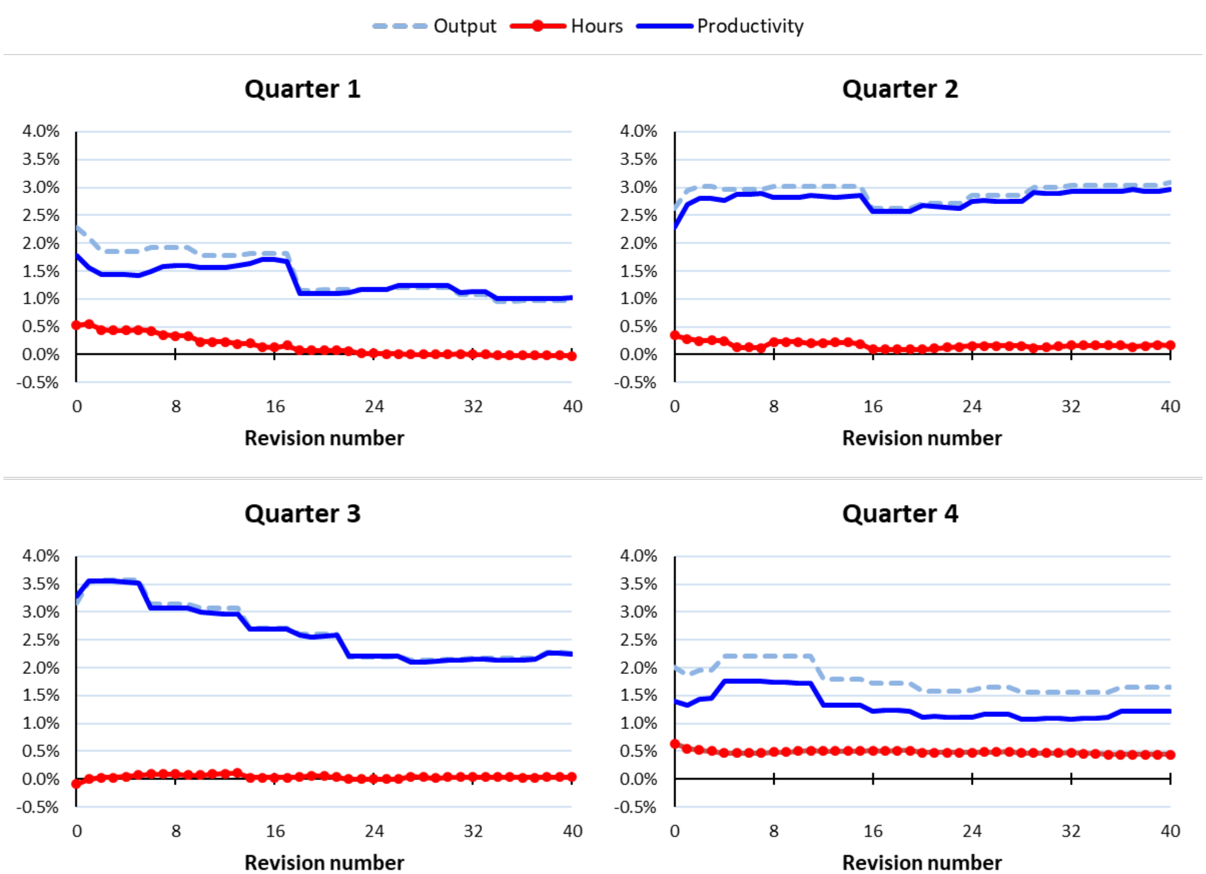
Figure 4: Average Growth Rates by Release, 2000-2015



Given the large variation in the size of revisions across reference quarters that we saw in Tables 2a, 2b, and 2c, one might suspect that the path of estimated growth rates also to vary by reference quarter. The variation in the paths of estimated growth rates can be seen in the graphs in Figure 5, which replicate Figure 4 by reference quarter. As we saw in Table 2c, estimates of output and LP are revised downward for all quarters except Q2 and the largest downward revisions are for Q1 and Q3. The Q1 revisions to output are larger than those for LP because hours growth is also revised downward. However, Figure 5's insights lie in the timing of those revisions. In contrast to Figure 4, which shows that estimated output (and LP) growth declines fairly smoothly with subsequent revisions, the quarterly graphs in Figure 5 show that most of this smooth decline can be traced to discrete drops (and occasional increases) at specific revision numbers that vary across quarters. These discrete changes to estimated output growth correspond to BEA annual revisions:[12]

- Q1: The drop between the R9 and R10 estimates and between the R17 and R18 estimates.
- Q2: The drop between the R15 and R16 estimates.
- Q3: The drops between the R5 and R6 estimates, between the R13 and R14 estimates, and between the R21 and R22 estimates.
- Q4: The jump between the R3 and R4 estimates and the drop between the R11 and R12 estimates.

Figure 5: Average Growth Rates by Revision for each Reference Quarter



[12] See Table A1 in the Appendix for the schedule of revisions.

For most of our sample period, the BEA annual revisions cover an approximately 3-year period and are first reflected in the August releases. For example, the August 2014 annual revision covers 2011q1 through 2014q1. Starting in 2019, the annual revisions cover a 5-year period.[13] There were no annual revisions in years when BEA did comprehensive revisions (2003, 2009, 2013, and 2018), which cover the entire series. Annual revisions are more expansive than other revisions (except comprehensive revisions) in that they include minor changes in methodology and the incorporation of new data sources. Thus, one might expect revisions to be larger when they coincide with the incorporation of annual revisions.

Given that the annual revisions cover a 3-year period, one might expect larger revisions to occur at the "seams". Using the example above, one might expect there to be a large revision to the 2011q1 growth rate to accompany the 2014 annual revision, because the estimate of 2011q1 output was revised while the 2010q4 estimate was not. For Q1, this "seam" effect would occur between the R25 and R26 estimates of output growth. But the largest revisions for Q1 are between the R17 and R18 estimates, which are in the middle year of the 3-year revision period. This same revision can be tied to the decline in the estimated output growth rates between the R15 and R16 estimates for Q2, the R13 and R14 estimates for Q3, and the R11 and R12 estimates for Q4. The average "middle-year" revisions were −0.65 for Q1, −0.40 for Q2, −0.35 for Q3, and −0.41 for Q4.

We can partly explain these sharp changes. Most of the large downward revisions to Q1-Q3 output were around the time of the Great Recession. The average middle-year revisions for the 2007-2013 period were: −1.09 percentage points for Q1, −0.76 for Q2, and −0.66 for Q3. In contrast, the middle-year revisions for Q4 were concentrated in the 2000-2006 period—mainly in the years following the 2001-2002 recession. In a breakout of the data, not shown here, we found that Q1 growth rates for reference years 2009 to 2012 were each revised sharply down, by more than 1.5% each, two years after the reference quarter, which correspond to the R17-R18 revisions for those quarters.

We do not know if there is a connection between the large downward revisions and residual seasonality that has been found in the GDP data. But a possible clue to these downward revisions to Q1 data can be found in an article by Moulton and Cowan (2016). In their discussion of the possible sources of residual seasonality, they note that revisions to estimates

[13] The 2021 annual revision was unusual in that it covered 1999q1 through 2021q1.

from the survey of the Value of Construction Put in Place cover a 2-year period, which creates a seam in the middle year of the BEA's annual revisions. Given that construction is especially sensitive to economic cycles, it seems likely that there could be large downward revisions in recession years. Further, given the seam between the Q4 and Q1 estimates, it also seems likely that revisions to this series could have a disproportionate impact on GDP revisions. It is worth noting that the BEA implemented changes to their seasonal adjustment procedures to address residual seasonality by mid-2018 (Mutikani, 2018). In our data, this distinctive set of downward revisions no longer occurred for post-Great Recession data.

Annual revisions likely play a role in the pattern of revisions that we saw in Tables 2a, 2b, and 2c. Starting with Table 2a, we see that the all-quarter average R0-to-R2 revisions to output growth are slightly positive. Average revisions are positive for Q2 and Q3, close to zero for Q4, but negative for Q1. A significant difference between Q1 and the other three quarters is that the R0-R2 revision for Q1 includes the August annual revision. Turning to Table 2b, we can see the impact of annual revisions on output levels (positive) and growth rates (negative). The all-quarter average revisions to output growth are large and negative. By quarter, the average revisions are negative except for Q2 revisions, which are (slightly) positive. This strongly suggests that, over our sample time period, the BEA annual revisions resulted in downward revisions to output growth. This would seem to explain the downward R0-to-R2 revision to Q1 output growth in Table 2a, because the R1-R2 revision includes an annual revision.

Putting the patterns in Tables 2a and 2b together suggests that early revisions to output growth rates tend to be positive and that later revisions, which are driven by BEA annual revisions, tend to be negative. This is significant because the Q1 estimates of LP are subject to one additional annual revision. The fact that the R2 estimates for Q1 coincide with annual revisions could explain why the negative R0-to-R40 revisions to Q1 output growth tend to be larger than for other quarters. Related to this, the revisions to Q2 output shown in Tables 2a, 2b, and 2c are consistently positive and are mainly due to R0-to-R2 revisions. It seems likely that the positive R0-to-R2 revisions to output growth for Q2 could be a collateral effect of the negative R0-to-R2 revisions to Q1 output growth.

We can see the impact of the additional annual revision on Q1 growth rates by performing a similar decomposition on the early revisions. Tables 5a and 5b show the decompositions for R0-R1 and R1-to-R2 revisions. Note that the R0-R1 revisions to output (Table 5a) are similar for all four quarters, with the largest revisions being the revisions to current quarter output for Q2 and Q3. Revisions to growth rates vary somewhat, with positive revisions for Q2 and Q3 and negative revisions for Q1 and Q4, but Q1 does not stand out as being significantly different. However, looking at the R1-to-R2 revisions (Table 5b), Q1 looks significantly different. The revisions to output levels are an order of magnitude larger than the revisions for the other three quarters. And the revisions to Q1 output growth are negative on average, whereas those revisions are positive and smaller in magnitude for the other three quarters. This provides evidence that the extra annual revision of Q1 output contributes to the large downward R0-to-R40 revisions for that quarter.

Table 5a: Decomposition of R0-to-R1 Revisions, 2000 - 2015

	Average Revision to:						Labor Productivity Growth	
	ln(Output)			ln(Hours)				
	Current Quarter	Previous Quarter	Growth Rate (x4)	Current Quarter	Previous Quarter	Growth Rate (x4)	Average Revision	Mean Squared Revision
All Quarters	0.03	0.01	0.10	-0.02	-0.01	-0.01	0.11	0.90
Q1	-0.02	0.02	-0.15	0.02	0.01	0.04	-0.19	0.54
Q2	0.08	0.00	0.33	0.02	0.00	0.09	0.24	0.80
Q3	0.10	0.00	0.38	0.00	0.03	-0.10	0.48	0.87
Q4	-0.03	0.01	-0.16	-0.06	-0.04	-0.10	-0.06	1.61

Table 5b: Decomposition of R1-to-R2 Revisions, 2000 - 2015

	Average Revision to:						Labor Productivity Growth	
	ln(Output)			ln(Hours)				
	Current Quarter	Previous Quarter	Growth Rate (x4)	Current Quarter	Previous Quarter	Growth Rate (x4)	Average Revision	Mean Squared Revision
All Quarters	-0.12	-0.12	-0.01	0.00	0.01	-0.02	0.01	0.48
Q1	-0.49	-0.43	-0.24	-0.07	-0.05	-0.08	-0.16	1.38
Q2	-0.02	-0.04	0.09	-0.01	0.00	-0.04	0.13	0.27
Q3	0.01	0.00	0.01	0.09	0.08	0.03	-0.02	0.27
Q4	0.03	0.00	0.10	0.00	0.00	0.00	0.10	0.10

The nature and timing of annual revisions sheds light on the Mincer-Zarnowitz results in Table 3. Nearly all of these revisions occur after the R2 estimate.[14] The R0-to-R2 revisions to output are "news" mainly because they replace projections and proxies with real data. In

[14] The only exception is the R1-to-R2 revision for Q1 in the year of the revision (see Table 5b).

contrast, the R2-to-R40 revisions are not "news" because they primarily incorporate changes in methods and data sources. Even though these annual revisions improve the accuracy of the estimates they can be viewed as random variation, which explains why these later revisions do not eliminate noise in the Mincer-Zarnowitz regressions.

An important caveat to this analysis is that, because the large downward revisions were for reference quarters in the Great Recession, this may not be an issue in more normal times.

5. How Well do Early Estimates Predict Later Estimates?

To further examine how well early estimates of LP growth predict the later estimates, we regress the R40 estimates on the R0 and R2 estimates. Table 6 shows the coefficients from these regressions. If the R0 and R2 estimates of the R40 value were unbiased, then the coefficients on these early estimates would be close to 1 and the constant would be close to 0.

We see in columns (1) and (2) of Table 5 that, as expected, the R0 and R2 estimates of labor productivity are strong predictors of estimates at R40. The coefficients on these early estimates are 0.82 and 0.83 and are not statistically different from 1. The constants are also not statistically significantly different from zero, although the joint hypothesis that the slope coefficient equals 1 and constant equals 0 is rejected in the R2 regression but not in the R0 regression.

The results in Table 6 are consistent with earlier findings. The coefficient estimates suggest that the R0 estimates are better than the R2 estimates but comparing the R-squared values in the first two columns of Table 6 we can see that the R2 explains more of the variation in the R40 estimate than does the R0 estimate. Column (3) shows that including the R0 estimate in the R2 regression does not add much information. The R-squared is about the same, although the coefficient on the R2 estimate is closer to 1 and the joint test no longer rejects the null hypothesis.

Given the variation in revisions by reference quarter that we see in Figure 5 and Table 4, we reran the regressions in Table 6 separately for each quarter. Table 7 presents these results along with the results from the all-quarter regressions from Table 6 for comparison. The first thing to note is that, as we might expect, the results vary quite a bit across reference quarters for both the R0 and R2 estimates.

The slope coefficients from the R0 regressions vary from 0.59 for Q1 to 1.06 for Q2 and none are statistically different from one. The constants are large, but none is statistically different from zero at conventional levels of significance. The R-squared values indicate that the R0 estimates for Q2 and Q3 perform the best and the estimates for Q1 perform the worst. The joint tests do not reject the hypotheses that the slope = 1 and the constant = 0.

Table 6: Early Measures as Predictors of LP_{40}

	Dependent variable is LP growth estimate as of R40		
	(1)	(2)	(3)*
R0 Estimate	0.816		-0.105
P-value (coef = 1)	0.103		0.647
R2 Estimate		0.831	0.916
P-value (coef = 1)		0.057	0.685
Constant	0.081	-0.058	-0.026
P-value (coef=0)	0.823	0.851	0.685
P-value (joint test)	0.133	0.029	0.915
R-squared	0.464	0.595	0.596
Observations	64	64	64

* P-value on R0 is for coefficient = 0. P-value for joint test is coefficient on R2 = 1 and constant = 0.

The results for the R2 regressions are similar in that there is a lot of variation in the coefficients across quarters, but the quarterly R2 estimates generally explain more of the variation in the R40 estimate than the corresponding R0 estimates. The R2 estimate explains more of the variation in the R40 estimate for quarters Q3 and Q4 compared with Q1 and Q2, although the joint hypothesis that the slope = 1 and the constant = 0 is rejected for Q3.

The low R-squared values in the Q1 regressions suggest that they are the main culprit for the low R-squared for the all-quarter regression. To verify this, we reran the equations in Table 6 excluding Q1 from the all-quarter regressions. In both of the restricted-sample regressions, the slope coefficients were closer to one, and we fail to reject the joint hypothesis (slope = 1 and constant = 0).

Table 7: Early Measures as Predictors of LP_{40}, by quarter

	Dependent variable is LP growth estimate as of R40				
	All Quarters	Q1	Q2	Q3	Q4
R0 Estimate	0.816	0.592	1.061	0.982	0.751
P-value (coef = 1)	0.103	0.150	0.786	0.931	0.260
Constant	0.081	-0.033	0.552	-0.977	0.169
P-value (coef=0)	0.823	0.968	0.417	0.264	0.776
P-value (joint test)	0.133	0.194	0.299	0.152	0.489
R-squared	0.464	0.258	0.625	0.619	0.471
Observations	64	16	16	16	16

	Dependent variable is LP growth estimate as of R40				
	All Quarters	Q1	Q2	Q3	Q4
R2 Estimate	0.831	0.601	0.824	1.136	0.964
P-value (coef = 1)	0.057	0.062	0.421	0.355	0.813
Constant	-0.058	0.157	0.649	-1.799	-0.164
P-value (coef=0)	0.542	0.811	0.415	0.011	0.693
P-value (joint test)	0.029	0.136	0.682	0.005	0.803
R-squared	0.595	0.399	0.517	0.820	0.749
Observations	64	16	16	16	16

Output

Table 8 replicates regressions in Table 7, but for output. There are some notable differences and similarities to the LP regressions. Comparing the all-quarter regressions in Table 8 to those in Table 7, the slope coefficients are much closer to one while the constants are somewhat larger. Only two coefficients are statistically different from their hypothesized values (the constants in the Q3 regressions), and the R-squared values are generally larger than those in the LP equation.

As expected, the R2 regression explains more of the variation in the R40 estimate of output than the R0 regression, with R-squared values of 0.709 vs. 0.619 for the all-quarter regressions. As with LP, we see large differences across quarters. First note that in both the R0 and R2 regressions, the R-squared is notably lower for the first and second quarters, which could be related to the patterns related to annual revisions discussed above. The R-squared values for the R2 regressions are greater than the corresponding R-squared values for the R0 regressions, with the exception of the Q3 regressions. The joint tests fail to reject the hypothesized values in all regressions except the R0 and R2 regressions for Q3, which is the quarter that has the largest downward revisions to output.

Table 8: Early Measures as Predictors of Output$_{40}$, by quarter

	Dependent variable is output growth estimate as of R40				
	All Quarters	Q1	Q2	Q3	Q4
R0 Estimate	0.996	0.728	0.942	1.300	1.198
P-value (coef = 1)	0.967	0.186	0.787	0.116	0.288
Constant	-0.515	-0.510	0.625	1.854	-0.754
P-value (coef=0)	0.165	0.398	0.372	0.019	0.249
P-value (joint test)	0.155	0.086	0.490	0.049	0.447
R-squared	0.619	0.497	0.588	0.790	0.762
Observations	64	16	16	16	16

	Dependent variable is output growth estimate as of R40				
	All Quarters	Q1	Q2	Q3	Q4
R2 Estimate	0.912	0.701	0.783	1.073	1.127
P-value (coef = 1)	0.240	0.087	0.333	0.487	0.280
Constant	-0.380	-0.339	0.727	1.569	-0.553
P-value (coef=0)	0.216	0.625	0.373	0.005	0.219
P-value (joint test)	0.022	0.097	0.608	0.002	0.402
R-squared	0.709	0.573	0.482	0.888	0.877
Observations	64	16	16	16	16

Hours worked

Table 9 replicates the regressions in Table 8 for hours worked. The first thing to note in Table 9 is that the early estimates do a much better job of predicting R40 values than is the case for output and LP. Looking at the R0 and R2 all-quarter regressions, we see that the early estimates of hours accurately predict the R40 values. The coefficients are not statistically different from their hypothesized values (slope = 1 and constant = 0), and the R-squared values are around 0.9. Except for the R0 all-quarter regression, we fail to reject the joint hypothesis. In contrast to Table 7, there is only a slight improvement in predictions due to revisions between the R0 and R2 estimates. This is not surprising because, as we noted earlier, most of the early revisions to CES data have already been incorporated into the R0 estimate and benchmark revisions typically do not have a large impact on month-to-month changes.[15]

Table 9: Early Measures as Predictors of Hours$_{40}$, by quarter

	Dependent variable is hours growth estimate as of R40				
	All Quarters	Q1	Q2	Q3	Q4
R0 Estimate	1.072	1.041	1.063	1.166	1.090
P-value (coef = 1)	0.084	0.689	0.478	0.100	0.096
Constant	-0.228	-0.567	-0.202	0.135	-0.259
P-value (coef=0)	0.045	0.073	0.393	0.490	0.119
P-value (joint test)	0.047	0.188	0.573	0.210	0.122
R-squared	0.917	0.885	0.915	0.916	0.971
Observations	64	16	16	16	16

	Dependent variable is hours growth estimate as of R40				
	All Quarters	Q1	Q2	Q3	Q4
R2 Estimate	1.090	1.060	1.091	1.138	1.104
P-value (coef = 1)	0.014	0.478	0.291	0.085	0.061
Constant	-0.182	-0.491	-0.106	0.006	-0.134
P-value (coef=0)	0.062	0.060	0.627	0.972	0.405
P-value (joint test)	0.015	0.150	0.523	0.214	0.147
R-squared	0.937	0.922	0.926	0.944	0.971
Observations	64	16	16	16	16

<u>Do Early Revisions Predict Later Revisions?</u>

A related question regarding predictability is whether R0-to-R2 revisions predict R2-to-R40 revisions. We examine this question by regressing R2-to-R40 revisions on R0-to-R2 revisions for LP, output, and hours. The results in Table 10 indicate that there is no relationship between the magnitude and direction of the early and later revisions. In the LP and output regressions,

[15] Extended regressions in Appendix C address other hypotheses, showing that for first quarter and recession quarters, hours tend to be revised significantly downward after R2. We do not find significant differences in these patterns across decades, or for reference quarters which had notable events. The magnitude of hours revisions has declined since the 1990s.

the slope coefficients are essentially zero, as are the R-squared values. The constants are approximately equal to the values of the R2-to-R40 revisions in Table 1. The results in the hours regression are slightly different. The slope coefficient is relatively large and is nearly statistically significant, but the R-squared value is still very close to zero. Thus, we conclude that the early revisions do not convey any information about future revisions.

Table 10: Regressions of Later Revisions on Early Revisions

	Dependent variable is R2-to-R40 Revison		
	LP	Output	Hours
R0-R2 Revision	-0.087	-0.114	0.649
P-value (coef=0)	0.678	0.592	0.078
Constant	-0.437	-0.600	-0.124
P-value (coef=0)	0.074	0.014	0.217
R-squared	0.003	0.005	0.049
Observations	64	64	64

Predictability Overall

Putting these results together, we come to several conclusions. First, we can do a better job of predicting the components of LP growth than we can of predicting LP growth. This suggests that LP can be better predicted by predicting the components and then constructing predicted LP growth from the predicted values of the components. This turned out not to be the case because revisions to output and hours are only weakly correlated, with correlation coefficient of 0.17.

Second, although the R0 estimates are closer on average to the corresponding R40 estimates, the R2 estimates are better in the sense that they explain more of the variation in the R40 estimates and have a smaller MSR. Third, there are significant differences in the predictive power of the early estimates by reference quarter, with early estimates of Q1 being the weakest predictor. And finally, early revisions to LP growth do not predict later revisions.

6. Prediction Intervals for Long-Run Revisions

In earlier research (Asher, et al., 2022), we developed a method for generating "prediction intervals" for the R0 and R1 estimates relative to the R2 estimate. The goal was to inform data users about the expected size of revisions. The two fan charts in Figures 6a and 6b show the 70-

percent, 80-percent, and 90-percent "prediction" intervals for the R0 and R2 estimates relative to each subsequent release for 2015q4. These intervals were calculated using revision data from 2000q1 through 2015q3 and the weighted percentile method described in Asher et al (2022).[16] The vertical axis shows the measured labor productivity growth rate. The horizontal axis graphs the release number. The dashed line shows the actual path of estimated 2015q4 productivity growth. To illustrate, in Figure 6a the R0 estimate is -3.0 percent and the R40 estimate is predicted to be between −1.4 and −5.3 about 70 percent of the time (as of the time of the R0 release). Figure 6b shows a similar fan starting at the R2 estimate of −1.7% for the same quarter.

There are several interesting things to note in these figures. First, the intervals are wide because, as we saw, revisions can be large. Second, the intervals for the R2 estimates are narrower than those for the R0 estimates, which is consistent with the assumption that revisions bring the estimates closer to "truth" and our finding that there is less variability in the R2-to-R40 revisions compared with R0-to-R40 revisions. However, it is worth noting that widths of the intervals are not monotonically increasing. This should not be too surprising given the path of average Q4 estimates that we saw in Figure 5. And third, the upper and lower bounds of the 90-percent intervals for the R0 estimate are not symmetric because, during this period, extreme downward revisions have tended to be larger than extreme upward revisions. The most extreme downward revisions were for reference quarters that were in the Great Recession—because early estimates of GDP use models and trends to fill in for missing data and these models tend to miss turning points.

Table 11 shows interval widths for the R0 and R2 estimates relative to the R40 estimate. As expected, the interval widths increase monotonically with the level of confidence, but the difference between the R0 and R2 interval widths does not. The smaller difference for the 90-percent interval indicates that most of the largest revisions occur between the R0 and R2 estimates.

[16] The weighted percentile method is less sensitive to other methods such as the modified confidence interval method described 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 a comparison to alternative methods.

Figure 6a: Fan chart for 2015Q4 labor productivity growth: 70%, 80%, and 90% intervals starting from R0

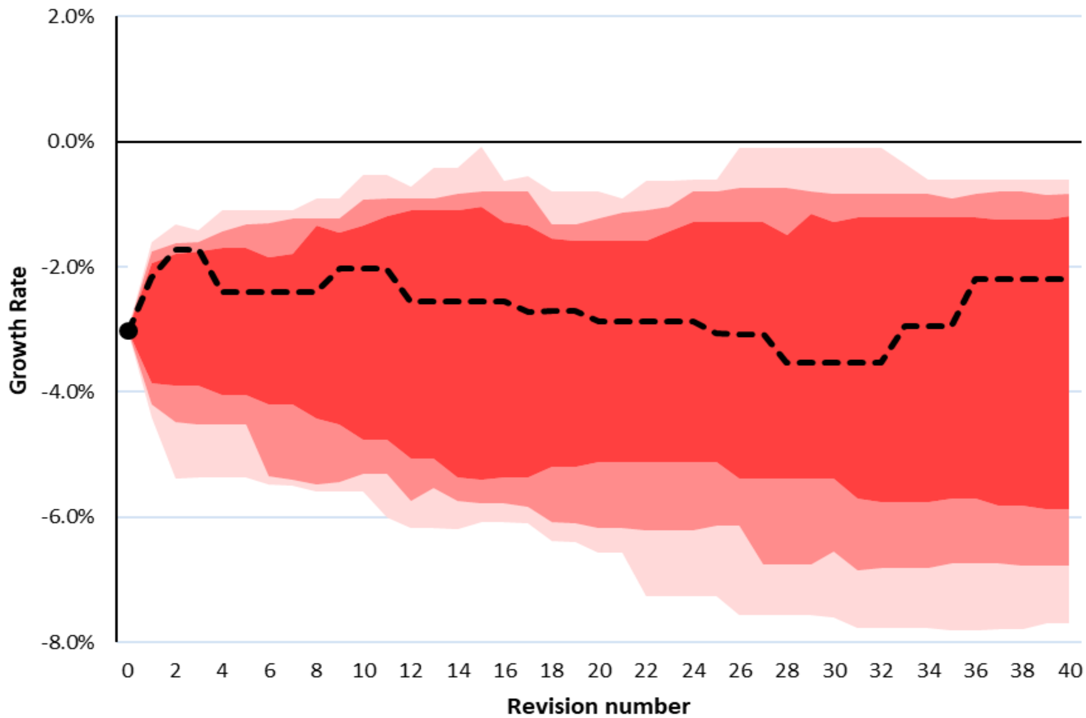


Figure 6b: Fan chart for 2015Q4 labor productivity growth: 70%, 80%, and 90% intervals starting from R2

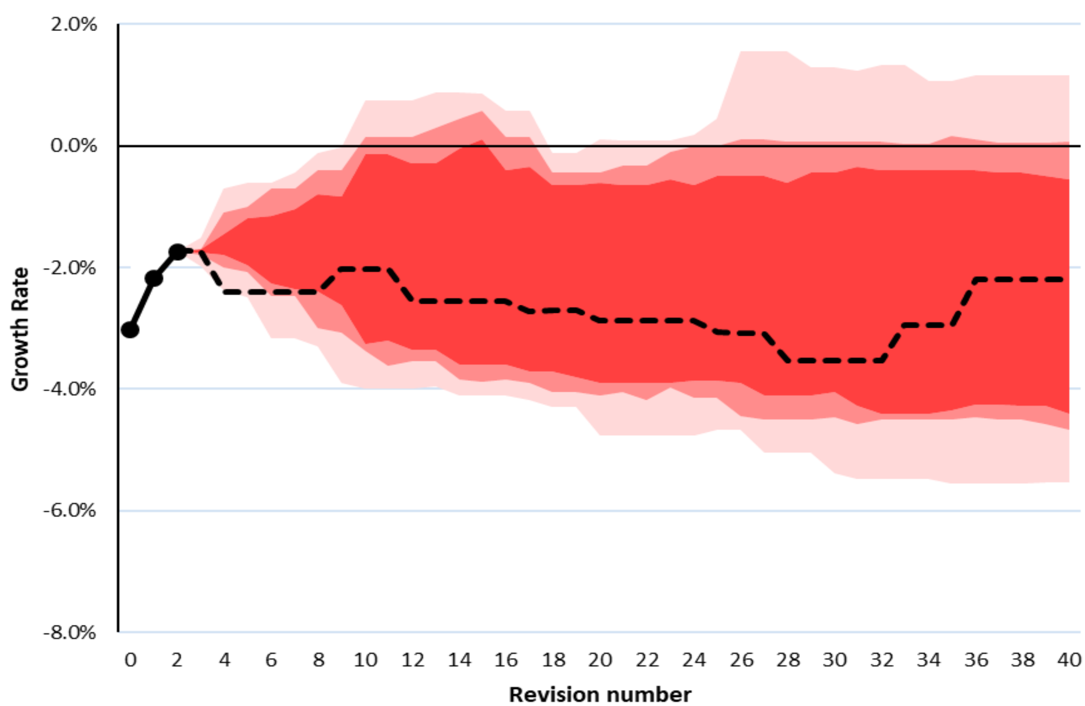


Table 11: Fan interval widths for R40 estimates

	Starting from…		
Confidence	**R0**	**R2**	**Difference**
70%	4.7	3.9	-0.8
80%	6.0	4.8	-1.2
90%	7.1	6.7	-0.4

7. Comprehensive Revisions to GDP

In addition to the regularly scheduled revisions to output and hours, the output data are updated by periodic comprehensive revisions. These revisions often include changes to the output concept—that is, what is included—but can also reflect new methods or new data sources. These revisions occur in years ending in 3 and 8 and are based largely on the Economic Censuses completed in the previous year. These revisions are truly comprehensive in that GDP levels and growth rates are revised for the entire time series. Comprehensive revisions tend to increase both the level and growth rate of measured GDP, because the revisions tend to expand the output concept to include goods and services that are growing faster than the rest of GDP. However, because the entire series is revised, changes to growth rates tend to be small, even when the changes to levels are large.

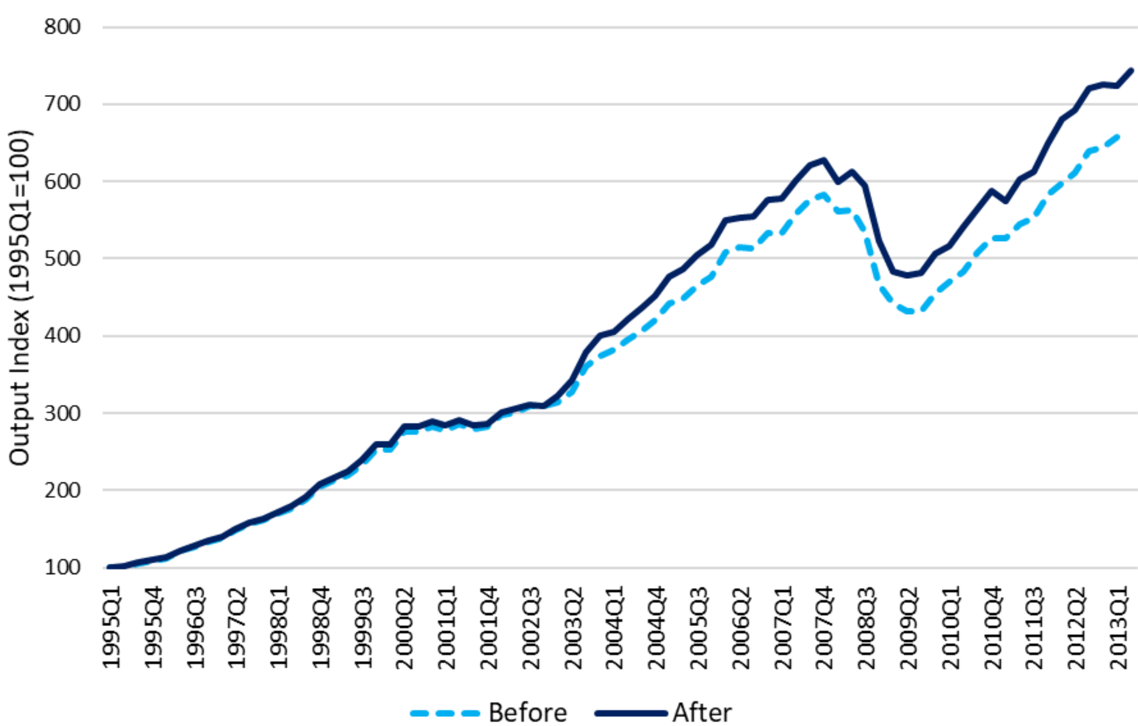
There were two comprehensive revisions to the output data in our sample period: 2013 and 2018.[17] The most significant of these was the 2013 revision, where BEA incorporated two important forms of intangible capital as investment: expenditures on research and development and on artistic originals. Because capital investments are included in output, in addition to being added to the capital stock, these changes resulted in an expanded definition of output. The result is that estimates of pre-2013 output levels and growth rates were revised upward. Because the entire series was revised, both current quarter and previous quarter estimates were revised upward by similar amounts, so that the net effect on GDP growth rates was relatively small. The 2013 revision increased the estimated level of GDP for 2011 by 3.1 percent but increased measured average annualized growth by only 0.174 percentage points per year over the 1995-2011 period. The impact of the comprehensive revisions can also be seen in Tables 2b and 2c

[17] The 2013 comprehensive revision affected 52 of the 64 reference quarters in our sample (2000q1-2012q4), while the 2018 revision affected all 64 quarters.

(along with the impact of annual revisions). In most quarters, the revisions to both the current and previous quarter's output are relatively large while the revisions to output growth are smaller (keeping in mind that the revisions to growth are annual rates).

Figure 7 shows the impact of the 2013 comprehensive revision on output levels. The comprehensive revision had a small impact on output until 2003. After 2003, we can see that the gap between pre- and post-revision output increases and becomes fairly substantial, but the growth rates are fairly similar. Because their effect on growth rates is small, comprehensive revisions do not have a significant impact on the width of the prediction intervals in the fan charts in Figures 6a and 6b.

Figure 7: Impact of 2013 Comprehensive Revision on Output



Notes: The 1995q1 level is indexed to 100.

Table 12: Net effects of Comprehensive Revisions in 2013 and 2018

Release month	Sample for estimated change	Average and range of revisions to output growth rates	Average and range of revisions to hours-worked growth rates	Average and range of revisions to LP growth rates
Aug 2013	1995-2011 (68 qtrs)	.174 (-2.43, 2.54)	.023 (-0.056, 0.159)	0.150 (-2.74, 2.55)
Aug 2018	1995-2016 (88 qtrs)	0.067 (-1.30, 1.43)	-.003 (-.071, 0.066)	0.071 (-1.26, 1.45)

Table 12 quantifies the effect of the 2013 and 2018 comprehensive revisions to the estimates of output and productivity. Revisions to hours, which are not part of the comprehensive revision, are shown for comparison. The average net effect was to increase measured average annualized output growth by 0.17 of a percentage point. The impact on average labor productivity growth over this period was slightly smaller (0.15 of a percentage point) because of concurrent revisions to hours. The 2018 comprehensive revision also increased average growth rates, but the increase was much smaller. It is worth noting that although the long-run growth rate was revised upward, growth rates for a given quarter could be revised upward or downward. And although the impact on average growth rates was small, the size of the revisions varied considerably from quarter to quarter with a range of revisions of −2.43 to 2.54 for the 2013 comprehensive revision and −1.30 to 1.43 for the 2018 revision. This variation can also be seen in Figure 7.

The timing of comprehensive revisions changed in 2022. Beginning with the release of the 2022 Annual Update of the NIPA, which revised data through 2022 Q1, rather than ending in Q4 of the previous year under the previous schedule. Annual and comprehensive revisions will first be reflected in the November (R0 for Q3) rather than the August (R0 for Q2) releases of Productivity and Costs. Therefore, the patterns associated with these longer-term revisions in our study may tend to shift forward by one reference quarter, or two releases.

8. Revisions in the COVID-19 Era

The COVID-19 pandemic severely disrupted economic activity in early 2020. 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 vulnerable, because the sharp decline in economic activity in the last 2 weeks of the quarter would not be reflected in source data on either output or hours using normal methods. 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.

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 Q1 GDP, BEA modified its procedures by incorporating high frequency data such as credit card transactions and relying less on projections. Because of these modifications, revisions to 2020 Q1 output were 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.

The BLS Productivity Program modified its usual procedures for estimating hours worked by incorporating data on initial Unemployment Insurance (UI) claims for its early estimates (R0, R1, and R2). The early revisions to hours were larger than usual and 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 assumption that the UI Initial Claims reflected actual job losses and that there were no transitions from non-employment to employment.[18] 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 because there were not enough data to adjust self-employed worker hours or average weekly hours of wage and salary workers. Once the April data became available, it was 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.

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

Table 13: Revisions to 2020 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 13 summarizes the R0-to-R2 and R1-to-R2 revisions to labor productivity in 2020. The largest revision was the 3.3 percent R0-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. The revisions for Q1 and Q2 are among the largest revisions since 2000Q1. 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.

BEA plans to continue using credit card transaction data and other high frequency data, and this change is likely to result in smaller revisions to output and labor productivity. We do not yet have sufficient data to measure the effect of these changes.

9. Conclusion

Estimates of quarterly labor productivity growth are revised long after the end of the reference quarter, with most of these revisions being due to revisions to estimated output growth. Revisions to estimated hours growth are smaller and are essentially zero after 2 years because the source data are subject to only minor revisions after they are benchmarked. In contrast, estimates of output growth stabilize after about 5 years but are revised non-trivially even 10 years after the end of the reference quarter. This is because GDP is estimated from multiple data sources, some of which do not become available until well after the end of the reference period, and because comprehensive revisions can change concepts and methods.

Given that estimates of labor productivity growth can be revised long after the reference quarter, it is important for policymakers to understand the properties of those revisions and to what extent early estimates of LP growth “predict” later estimates, which are assumed to be more accurate because the revisions bring the estimates closer to "truth." In this paper, we examined the behavior of long-term revisions to estimates of quarterly labor productivity growth, and its components—output and hours. Our study adds to earlier work along several dimensions.

First, we compare estimates that have been revised the same number of times rather than using the most current vintage of the data. The main disadvantage of using the current vintage is that more-recent reference quarters have been revised fewer times. We examine average revisions and the path of average growth rates as the estimates are revised from the preliminary (R0) estimate through the R40 estimate and find that estimates are initially revised upward, followed by downward revisions. We also restricted our sample to the 2000-2015 period, rather than using the entire series, because revisions to earlier reference quarters may not be representative of current procedures for revising data. For example, the nature and timing of revisions changed between the late 1990s and the early 2000s. One proviso is that the BEA changed its seasonal adjustment methodology to address residential seasonality—specifically, the lower seasonally adjusted growth rates for Q1 GDP. Unfortunately, we cannot look at the long-run implications of this change until more data are available.

Second, we examine the impact of BEA annual and comprehensive revisions to output. Annual revisions incorporate minor methodological changes and additional data sources, while comprehensive revisions also incorporate changes in concepts (for example, treating investment in intangible goods as an investment rather than a cost). Annual revisions cover the previous 3 years, while comprehensive revision cover the entire series. These revisions typically affect output levels by more than growth rates because revisions to the current and previous quarters' output are typically in the same direction and of approximately the same magnitude.

Third, we examined revisions by reference quarter. This yielded insights because the path of revisions varies by reference quarter. For example, we found that early revisions to Q1 output levels are large and can be traced to the annual revision that occurs between the R1 and R2 estimates. In contrast, early revisions to Q2-Q4 output levels are much smaller. By tracing the path of average productivity growth by reference quarter, we could identify the impact of annual revisions. Specifically, although the all-quarter path of average growth rates looks smooth, the paths by quarter reveal that this smoothness masks discrete changes that coincide with annual revisions.

The timing and nature of annual revisions help explain the results of our Mincer-Zarnowitz results in Table 3, which seem inconsistent with the assumption that revisions move estimates closer to “truth.” Early revisions to output are news because they are due to the replacement of projections and proxies with real data. However, subsequent revisions (annual and comprehensive) are due primarily to changes in methods, data sources, and concepts. These revisions, while improving the accuracy of the estimates, are not news in the Mincer-Zarnowitz sense because they are moving the target rather than bringing current estimates closer to a stationary target. For the same reason, these revisions do not eliminate noise.

Fourth, we take a closer look at how well the early estimates of labor productivity growth “predict” the corresponding R40 estimates. We took several approaches that told the same basic story. The preliminary (R0) estimate is generally a little closer to the R40 estimate than the R2 estimate. But the R2 estimates "explain” more of the variation in the R40 estimates and have a smaller mean squared revision. Our analysis also revealed that the early (R0 and R2) estimates for Q1 are significantly worse predictors of R40 values compared with the early estimates for the other quarters.

Appendix A. Timing of revisions

The schedule for labor productivity releases is driven by the major GDP releases from BEA of the first (Advanced) and second estimates of quarterly GDP growth. Revisions between the second and third releases of GDP are smaller and are incorporated with the next quarter's first release of labor productivity. Table A1 shows the release schedule for labor productivity, along with revision dates for estimates of output and hours.

Table A1: Annual data calendar for quarterly nonfarm labor productivity

		Source Data Subject to Revision		
Month	**LP Release**	**GDP**	**CES**	**Major Revisions**
January				
February	R0 for Q4 R2 for Q3	Advance for Q4 2nd revision for Q3	Oct. (3), Nov. (2), Dec. (1) July (3), Aug. (3), Sept. (3)	
March	R1 for Q4	1st revision for Q4	Oct. (3), Nov. (3), Dec. (2)	CES data incorporate annual benchmark revision (introduced in Jan. employment situation)
April				
May	R0 for Q1 R2 for Q4	Advance for Q1 2nd revision for Q4	Jan. (3), Feb. (2), Mar. (1) Oct. (3), Nov. (3), Dec. (3)	
June	R1 for Q1	1st revision for Q1	Jan. (3), Feb. (3), Mar. (2)	
July				
August	R0 for Q2 R2 for Q1	Advance for Q2 2nd revision for Q4	Apr. (3), May (2), June (1) Jan. (3), Feb. (3), Mar. (3)	GDP data incorporate annual revisions (introduced in July). Comprehensive revision (years ending in 3,8)
September	R1 for Q2	1st revision for Q2	Apr. (3), May (3), June (2)	
October				
November	R0 for Q3 R2 for Q2	Advance for Q3 2nd revision for Q2	July (3), Aug. (2), Sept. (1) Apr. (3), May (3), June (3)	
December	R1 for Q3	1st revision for Q3	July (3), Aug. (3), Sept. (2)	

Note: The numbers in parentheses under the CES revisions refer to the release number.
Note: The supervisory ratios are calculated each quarter for the R0 estimate and are never revised (except for seasonal factors)
Note: Beginning in 2022, the annual GDP revisions are incorporated into the second revision of Q2 estimates (introduced in September)

Tables A2a – A2c illustrate the impact of these regularly scheduled revisions on output, hours, and labor productivity. Each table covers reference quarters from 2000-2015, and all releases in the 2000-2019 period. There are 80 observations in each cell—16 reference years and 5 years of revisions.

As might be expected, the largest revisions to output are in August when the annual revisions or Comprehensive Revisions enter our data, and the largest changes to hours tend to occur in March when the CES annual update for benchmarking and seasonal adjustment enters the labor productivity series. Turning to Table A2c, we see that it is only revisions to output that have a material impact on estimates of labor productivity growth.

Table A2a: Average magnitudes of revision to Output growth by reference quarter and release month

	Reference quarter			
LPC Release	**Q1**	**Q2**	**Q3**	**Q4**
February	0.022	0.034	0.103	0.019
March	0.025	0.026	0.013	0.184
May	0.000	0.000	0.000	0.039
June	0.114	0.000	0.000	0.000
August	**0.903**	**0.670**	**0.574**	**0.556**
September	0.000	0.138	0.000	0.000
November	0.002	0.076	0.003	0.001
December	0.000	0.000	0.158	0.000

Table A2b: Average magnitudes of revision to Hours growth by reference quarter and release month

	Reference quarter			
LPC Release	**Q1**	**Q2**	**Q3**	**Q4**
February	0.007	0.004	0.022	0.005
March	**0.207**	**0.223**	**0.205**	**0.256**
May	0.011	0.011	0.011	0.021
June	0.032	0.014	0.014	0.011
August	0.120	0.085	0.079	0.079
September	0.001	0.026	0.000	0.001
November	0.000	0.022	0.000	0.000
December	0.009	0.006	0.037	0.000

Table A2c: Average magnitudes of revision to Labor Productivity growth by reference quarter and release month

	Reference quarter			
LPC Release	**Q1**	**Q2**	**Q3**	**Q4**
February	0.025	0.041	0.118	0.025
March	0.236	0.250	0.223	0.397
May	0.011	0.012	0.011	0.056
June	0.119	0.016	0.013	0.011
August	**0.936**	**0.685**	**0.594**	**0.611**
September	0.001	0.152	0.000	0.001
November	0.003	0.086	0.004	0.001
December	0.010	0.006	0.155	0.000

Appendix B. Revisions to concepts and methods

Since the mid-1990s, there have been significant changes to official productivity methods and source data over time. Revisions to output data include updated concepts, improved data, or changes in the kind of data that is available. The most important of these changes are the Comprehensive Revisions to GDP every five years, which are discussed in section 7. Revisions to hours data result from changes in methods used by the BLS productivity program and revisions to source data from the CES. The list below shows the significant methodological changes over our sample period. These changes typically changed levels but had little effect on growth rates.

Major Changes to Source Data

1996: Comprehensive revision of GDP. BEA introduced chained Fisher indexes to estimate GDP growth. These changes were introduced into the February 1996 release characterizing the preliminary estimate of 1995q4 LP.

1998 (August): CES benchmarking now occurs each August instead of June.

1999 (October): NIPAs were revised back to 1959 to incorporate own-account software and other changes. This was a significant revision but was not called a “comprehensive revision.” https://www.bea.gov/news/1999/gross-domestic-product-3rd-quarter-1999-advance-revised-estimates-1959-99

2000-2003: The CES transitioned from a quota sample to a probability sample and converted industry codes from SIC to NAICS.

2001: BLS discontinued its Hours at Work survey and started using hours-worked-to-hours-paid ratios from the NCS. NCS ratios were linked forward from the Hours at Work series.

2003 (February): CES benchmarking now occurs each February instead of August.

2003 (December): NIPA comprehensive revision. Reflected in LPCs February 2004 release for 2003q4.

2004: Industry codes were switched from SIC to NAICS in the NIPAs. The Census industry codes used in the CPS included NAICS equivalents.

2004 (July): OPT adopted a new method for estimating nonproduction worker hours using CPS data. Previously, it had been assumed that nonproduction workers worked the same average weekly hours as production workers.

2005 (June): The hours worked estimates started incorporating data on second jobs.

2009 (July): BEA comprehensive revision of GDP.

2013 (July): BEA comprehensive revision of GDP to incorporated two important forms of intangible capital as investment: expenditures on research & development and on artistic originals.

2017 (March): OPT modified the method for applying the HWHP ratios. OPT now uses three-year averages of fourth quarter estimates. Ratios from the Hours of Work survey are now adjusted to be consistent with the NCS estimates. NCS ratios before 2017 were adjusted to be consistent with estimates from the Hours of Work survey.

2018 (July): BEA comprehensive revision of GDP.

2022 (September): BEA annual revisions moved to September from August so that national and industry estimates are released at the same time. This was first reflected in the November 2022 LPC release.

2022 (November): Hours estimates start to be drawn from CES all-employee hours data instead of the method discussed in the text.

Impact of the 1999 Revision to GDP

As noted in the text, revisions to data for the late 1990s reference quarters were significantly different from those for the 2000s. This can be seen most readily by reproducing the decompositions in Tables 2a – 2c for the 1995-1999 period.

Table B1a: Decomposition of R0-to-R2 Revisions - 1995 - 1999

	Average Revision to:							
	ln(Output)			ln(Hours)			Labor Productivity Growth	Mean Squared Revision
	Current Quarter	Previous Quarter	Growth Rate	Current Quarter	Previous Quarter	Growth Rate		
All Quarters	0.22	-0.01	0.23	0.29	0.41	-0.12	0.35	1.48
Q1	1.35	1.26	0.10	0.97	1.51	-0.54	0.64	3.67
Q2	2.84	2.13	0.70	-0.08	0.07	-0.14	0.84	1.36
Q3	-3.30	-3.38	0.08	0.20	0.13	0.08	0.00	0.50
Q4	-0.01	-0.06	0.05	0.08	-0.07	0.14	-0.09	1.28

Table B1b: Decomposition of R2-to-R40 Revisions - 1995 - 1999

	Average Revision to:							
	ln(Output)			ln(Hours)			Labor Productivity Growth	Mean Squared Revision
	Current Quarter	Previous Quarter	Growth Rate	Current Quarter	Previous Quarter	Growth Rate		
All Quarters	8.67	8.30	0.37	2.44	2.14	0.30	0.07	3.10
Q1	6.42	6.83	-0.40	1.38	1.73	-0.35	-0.05	3.57
Q2	6.93	5.70	1.24	2.36	1.46	0.90	0.34	5.48
Q3	10.29	10.31	-0.02	3.07	2.23	0.84	-0.86	1.33
Q4	11.03	10.35	0.68	2.96	3.14	-0.18	0.86	3.96

Table B1c: Decomposition of R0-to-R40 Revisions - 1995 - 1999

	Average Revision to:							
	ln(Output)			ln(Hours)			Labor Productivity Growth	Mean Squared Revision
	Current Quarter	Previous Quarter	Growth Rate	Current Quarter	Previous Quarter	Growth Rate		
All Quarters	8.89	8.28	0.61	2.73	2.55	0.19	0.42	3.70
Q1	7.78	8.08	-0.31	2.34	3.23	-0.89	0.58	5.08
Q2	9.77	7.83	1.94	2.28	1.53	0.76	1.18	5.76
Q3	6.99	6.93	0.06	3.28	2.36	0.92	-0.86	1.60
Q4	11.03	10.29	0.74	3.04	3.07	-0.03	0.77	4.43

Comparisons of Table 2 to Table B1 show that the R2-to-R40 and R0-to-R40 revisions were much larger than those after 2000. BEA reported that the incorporation of own-account software increased GDP by about 5.8 percent. The larger revisions shown here are likely partly due to the fact that the 75 percent of output covered that comprises BLS's nonfarm business sector includes relatively more own-account software.

Revisions to hours—both levels and growth rates—are much larger in the 1995-1999 period compared to the 2000-2015 period. In the early 2000s, the CES moved from a quota sample to a probability sample. There were also improvements to the benchmarking procedures.

Appendix C. Distributions of each variable at R0, R2, and R40

Table C1 shows test statistics to evaluate whether estimates and revisions between them have normal (Gaussian) distributions. Each cell shows the p-value of the Shapiro-Wilk test. Figures below .05 indicate that a normal distribution was rejected at the 5% level of probability. Table cells for distributions that appear normal by that criterion are highlighted in **bold**.

For each variable, the R0 estimates do not have a normal distribution; those distributions are peaked and skewed. Revisions from R2 to R40 have approximately normal distributions.

Labor productivity estimates are closer to normal than the component variables. It is interesting that estimates of labor productivity at R2 and R40 are normally distributed even though the distributions of its components are not.

Table C1: P-values from Shapiro-Wilk tests for normality

	Hours	**Output**	**Labor productivity**
R0 estimates	0.00000	0.00102	0.01591
R0-to-R2 revisions	**0.59302**	0.02122	0.00133
R2 estimates	0.00000	0.00016	**0.31730**
R2-to-R40 revisions	**0.80970**	**0.81128**	**0.91290**
R40 estimates	0.00000	0.00077	**0.56141**

10. 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, and 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.

Eldridge, Lucy P., Sabrina Pabilonia, Drake Palmer, Jay Stewart, and Jerin Varghese (2022) "Improving Estimates of Hours Worked for U.S. Productivity Measurement." Monthly Labor Review, October 2022.

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. 2022. Can GDP measurement be further improved? Data revision and reconciliation. *Journal of Business & Economic Statistics*, 40:1, 423-431. https://doi.org/10.1080/07350015.2020.1831928

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

Jorda, Oscar, Noah Kouchekinia, Colton Merrill, and Tatevik Sekhposyan. 2020. "The Fog of Numbers" *FRBSF Economic Letter* 2020-20, July 15. https://www.frbsf.org/economic-research/publications/economic-letter/2020/july/fog-of-numbers-gdp-revisions/

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

Moulton, Brent R., and 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

Mutikani, Lucia. U.S. says has addressed residual seasonality in GDP data. Reuters, 27 July 2018.

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, and 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.

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.