BLS WORKING PAPERS

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

How large are revisions to estimates of quarterly labor productivity growth?

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

Working Paper 538

How large are revisions to estimates of quarterly labor productivity growth?

Last revised: April 6, 2021

Kendra Asher, Bureau of Labor Statistics
John Glaser, Bureau of Labor Statistics
Peter B. Meyer, Bureau of Labor Statistics
Jay Stewart, Bureau of Labor Statistics
Jerin Varghese, Bureau of Labor Statistics

Abstract

BLS's estimates of quarterly labor productivity, output per hour worked, are revised because of revisions to source data. Early estimates of hours worked and output are subject to substantial revisions for a variety of reasons. The BLS productivity program produces three regularly scheduled estimates of labor productivity growth: the preliminary estimate, the first revised estimate, and the second revised estimate. We consider revisions to the preliminary and first revised estimate relative to the second revised estimate. Our goal is to develop intervals to help data users better assess the size of these revisions.

Most of the revisions result from regularly scheduled updates of source data. We analyze these revisions to get a better understanding of their sources and to determine whether there are any systematic patterns that could be exploited to construct intervals. We find no evidence of trends or systematic patterns that we could exploit. Most notably, the largest revisions to current and prior quarter output coincide with the BEA's annual revision to GDP.

We then consider three alternative methodologies for constructing intervals: modified confidence intervals, model-based intervals, and percentile-based intervals. We argue that the percentile-based intervals are preferable, because they are less sensitive to outliers and therefore result in narrower intervals for a given level of statistical confidence.

Acknowledgements: We thank Barbara Fraumeni, Tara Sinclair, and Brady West for helpful comments.

1. Introduction and Background

The Bureau of Labor Statistics' (BLS) Labor Productivity and Costs (LPC) program produces quarterly estimates of labor productivity growth. For these estimates, BLS combines output data from the Bureau of Economic Analysis (BEA) with employment and hours data compiled from three BLS surveys: the Current Employment Statistics (CES) survey, the Current Population Survey (CPS), and the National Compensation Survey (NCS).[1] Of these data sources, two—the BEA output data and the CES employment and hours data—are revised multiple times after they are first released.[2] For each reference quarter, BLS releases three regularly scheduled estimates of labor productivity growth. The first, preliminary, estimate (**prelim**) is issued within 40 days of the end of the reference quarter. This initial estimate is revised as new data become available. The first revised estimate (**R1**) is released 30 days after prelim, and the second revised estimate (**R2**) is released 60 days after that. R2 is the last regularly scheduled release covering the reference period. The timing of the prelim and R1 estimates is dictated by BEA's release schedule for the Advance and Second GDP estimates. The R2 estimate is released at the same time as the prelim estimate of the following quarter, because the R1-to-R2 revisions are generally not large enough to warrant a separate news release. Subsequent revisions to the R2 estimates, such as BEA Comprehensive Revisions, can be large and can occur long after the reference quarter. Thus the estimates are never "final." In this paper, we focus on revisions to the prelim and R1 estimates relative to the R2 estimate.[3]

According to Fixler, et al, (2014) and Fixler, et al (2018), the early GDP estimates are subject to substantial revisions because: they are based on "…partial and preliminary source data as well as trend projections when data are not available." The source data for these early estimates come from "…a mixture of survey, tax, and other business and administrative data as well as various indicators, such as heating degree days…"

[1] A complete description of how BLS estimates total hours worked can be found in Eldridge, Sparks and Stewart (2018).
[2] The CPS data are almost never revised, but seasonal factors are subject to minor revisions. The NCS data are not revised because the LPC program uses NCS data for the fourth quarter and allocates changes to quarters using the Denton procedure. With this procedure, seasonal adjustment is not necessary.
[3] In future research, we will examine the long-run behavior of revisions. These later revisions are not as regular and often occur long after the end of the reference period.

In contrast, the main source of revisions to the hours data come from revisions to the CES data. There are three regularly scheduled releases, the first, second and third closing estimates. The CES's first closing estimates are usually released on the first Friday after the reference month, the second and third closing estimates are released in the following two months. These revisions are due to the collection of additional data. For example, establishments with monthly payrolls can never report in time for first-closing estimates. The largest revisions occur between the first and second closing estimates. As noted above, the preliminary labor productivity estimates are published about 40 days after the end of the reference quarter, which is too soon to incorporate CES data from the most recent Employment Situation report. Thus, the Q1 preliminary estimate uses first closing data for March, second closing data for February, and third closing data for January. The R1 estimate uses second closing data for March and third closing data for the other two months. And the R2 estimate uses third closing data for all three months. Thus, we would expect revisions to hours to be smaller than revisions to output.

The current news releases of the LPC program include intervals for the preliminary estimate of the business sector output-per-hour index relative to the second revision, based on data on revisions from 1995q4 through the current quarter.[4] They do not provide intervals for the productivity growth rate or for the first revised estimate (R1).

Confidence intervals are a fairly standard way of conveying uncertainty in survey-based statistics such as the unemployment rate and payroll employment. But as noted in Fixler, et al, (2018), it is not feasible to calculate conventional confidence intervals for GDP, because it is compiled from a number of survey, non-survey and administrative sources. The BEA does not publish intervals for its advance GDP estimate but Fixler, et al, (2014) and Fixler, et al (2018) discuss how intervals could be constructed. These calculations assume that each revised estimate is a better estimate of the true value, and inform data users about the likely size of these revisions. Some organizations publish "prediction intervals" that are based on subsequent revisions to preliminary estimates.

[4] The current language reads: RELIABILITY: Productivity and cost measures are regularly revised as more complete information becomes available. The measures are first published within 40 days of the close of the reference period; revisions appear 30 days later, and second revisions after an additional 60 days. In the business sector, the third publication (second revision) of a quarterly index of output per hour of all persons has differed from the initial value by between -1.5 and 1.4 index points approximately 95 percent of the time. This interval is based on the performance of this measure between the fourth quarter of 1995 and the second quarter of 2017.

For example, the Australian Treasury reports 70- and 90-percent intervals for its revenue projections that are based on historical forecasting errors. The Federal Reserve Board (hereafter The FRB) produces 70-percent intervals for its Industrial Production Index (IPI) that are based on historical revisions to preliminary estimates.

Galvao, Mitchell, and Runge (2019) consider how best to convey to the public information about the magnitudes of revisions. Based on surveys of both experts and general public readers of GDP growth statistics, they generally support of the approach of presenting intervals based on historical revisions. They conclude that it is more valuable to data users to quantify uncertainty than to provide qualitative descriptions. They also argue that quantifying uncertainty "…decreases the chance that the public misinterprets the uncertainty in information given to them, and does not reduce trust in the statistical office or encourage the view that data revisions are due to vested interests at the ONS [the United Kingdom's Office of National Statistics] or the Government."

For this study, we develop intervals based on revisions under the maintained assumption that each revision moves the estimate closer to the true value. Our focus is on prelim-to-R2 and R1-to-R2 revisions. We begin by summarizing the revisions and examining factors that might affect the size and direction of the revisions. We then consider alternative ways of constructing intervals for the quarterly estimates of aggregate U.S. nonfarm labor productivity growth. Our main focus is on estimates of growth from the previous quarter, since they receive the most attention. We report results for three alternative approaches to constructing intervals. The first is the modified confidence interval methodology discussed in Fixler, et al, (2014) and Fixler, et al (2018). The second is model-based, which allows us to control for differences across quarters. And the third approach constructs intervals based on percentiles of historical revisions, similar to the reliability estimates for the FRB's IPI. For each method, we generate 70-, 80- and 90-percent intervals.

The rest of the paper is organized as follows. Section 2 describes the data, presents summary statistics, and examines the sources of revisions. Section 3 presents the alternative methods for constructing intervals and Section 4 compares them. Section 5 discusses revisions to the 2020 estimates and the implications for generating intervals going forward. And Section 6 provides some concluding remarks.

2. Data and Summary of Revisions

Our data cover the period from 2000q1 through 2019q4 except 2018q4, which we dropped because the preliminary estimate was not produced due to the government shutdown, where the dates refer to the reference quarter. We experimented with extending the series back to mid-1990s, but our analysis indicated that nature and timing of revisions differed significantly in the 1990s compared to 2000 and later (see Appendix 2 for details).[5]

We begin our analysis by summarizing past revisions. We look at the magnitude and distribution of the revisions, and consider whether there are any systematic patterns that we could exploit in generating intervals or that could pose potential problems.

Table 1 shows summary statistics for prelim-to-R2 and R1-to-R2 revisions for quarter-to-quarter and year-over-year productivity growth. Looking at the quarter-to-quarter revisions in the left columns, we can see that both mean and median R1-to-R2 revisions are substantially smaller than the prelim-to-R2 revisions. The mean of 0.14 for the prelim-to-R2 revisions suggests a slight downward bias in the preliminary release, but it is not statistically significant. We also see that there is considerably more variability in the prelim-to-R2 revisions—the standard deviation is over 60 percent larger than that for R1-to-R2 revisions and the 90-10 difference is more than twice as large. The skewness statistics indicate that both distributions are left skewed, but the R1-to-R2 revisions are less so.[6] The kurtosis statistics indicate that most of the mass is concentrated in the center of the distribution, with the distribution of R1-to-R2 revisions being more concentrated as can also be seen from the 90-10 range. The revisions resulted in the estimate changing sign in a little under 10 percent of quarters. This is higher than Sinclair and Stekler (2013) found for revisions to GDP. However, about half of these sign changes were due to revisions that were less than one percentage point in absolute value. One reason for the larger number of sign changes compared to GDP is that the LPC data exclude the government sector output, which tends to have

[5] Our analysis is similar to that in Sinclair and Stekler (2013) in many ways. Their focus is on revisions to GDP and its components. Their study, which covers the period from 1977q1-2010q3, compares the BEA's advance estimates to their third estimates, which correspond to BLS's prelim and R2 estimates. They examine whether the advance estimates are biased and whether the estimates incorporate information about the state of the economy. For real GDP, they find a slight bias in the advance estimate and that the BEA estimates do incorporate business cycle information.

[6] Tests in Table A1 reject normality.

much smaller revisions. Together, these statistics indicate that there is no systematic bias in the quarter-to-quarter revisions and that R1 predicts R2 better than prelim does.

Chart 1 shows the distributions of prelim-to-R2 and R1-to-R2 revisions of quarter-to-quarter productivity growth. We can see that much of the deviation from normality is due to several rather large revisions. To see the impact of these outliers, we recalculated the statistics in Table 1 dropping revisions that were larger than 2 in absolute value. The mean revision is larger, 0.3, and is statistically significant. The skewness and kurtosis statistics, -0.08 and 3.15 respectively, are about the same as for a normal distribution. As a further check for bias, we estimated a series of OLS regressions to determine whether there is evidence of a trend and whether the prelim-to-R2 and R1-to-R2 revisions are correlated with prelim or R1. We found no evidence of a time trend in either the value or magnitude of either revision. Similarly, there is no evidence that the revisions are related to the value or magnitude of the preliminary estimates (prelim or R1). See Tables A4 and A5 in Appendix 1 for the regression results.

The right two columns of Table 1 show summary statistics for revisions to estimates of year-over-year productivity growth. Comparing the right two columns to the left two, we can see that revisions to year-over-year estimates are substantially smaller and more tightly distributed. The main reason for the smaller revisions to year-over-year estimates is that the estimates of quarter-to-quarter productivity growth are expressed as an annual rate. This means that the revisions are annualized as well, which results in those revisions being approximately four times as large. For this reason, and because quarter-to-quarter estimates receive the most attention from data users, the rest of the paper will focus on revisions to quarter-to-quarter estimates.

2.1 Quarterly Variation

Quarterly variation in the revisions can arise due to the timing of regularly scheduled revisions to source data. Table 2 shows the schedule of releases and regularly-scheduled revisions. There are two annual revisions that result in systematic differences in revisions across quarters. First, each July the BEA makes annual revisions to the GDP data for the previous calendar year. These are first used in productivity statistics in August's release of the preliminary estimate for Q2 and the R2 estimate for Q1, which is affected because it is a growth rate from the prior Q4.[7] Second, each March the CES is benchmarked to the QCEW and the seasonal adjustment factors are re-estimated. This revision is reflected in the R1 estimate for Q4.[8] Chart 3 gives a visual representation of the timing of news releases and revisions to the data for 2000q1.

Tables 3a and 3b show average revisions to estimates of nonfarm business sector labor productivity growth by reference quarter. They show that revisions of Q1 data are on average negative, while revisions to Q2-Q4 data tend to be positive with the largest revisions occurring for Q2 data. This is true for prelim-to-R2 and R1-to-R2 revisions, though R1-to-R2 revisions are smaller.

To investigate this further, we examined quarterly variation in a regression framework, and found differences across quarters to be similar to those in Table3a and 3b (see Appendix Tables A6 and A7). Only the prelim-to-R2 revisions for Q2 were statistically different from zero. More importantly, the coefficients on the quarterly dummies were not statistically different from each other with the exception of Q1 and Q2.

2.2 Decomposition of Revisions – Output vs. Hours

To better understand the magnitude and quarterly variability of revisions, it is useful to know what is driving them. Are revisions due mainly to revisions to output or hours? And are they due to revisions to current quarter or prior quarter data? And for which quarters do revisions to output and hours have the greatest impact on measured labor productivity?

To address these questions, we decompose 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 labor indexes, which makes it straightforward to decompose the revisions:

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

[7] Information on Updates to the National Income and Product Accounts | U.S. Bureau of Economic Analysis (BEA)
[8] CES National Benchmark Article (bls.gov)

where Q and L are indexes of real output and total hours worked, and the subscripts indicate the quarter. The four terms in this equation represent the four sources of revisions noted above. Under this specification of labor productivity growth, revisions are defined as:

$$Revision = \{[\ln(Q_t^{R2}) - \ln(Q_{t-1}^{R2})] - [\ln(L_t^{R2}) - \ln(L_{t-1}^{R2})]\} - \{[\ln(Q_t^{P}) - \ln(Q_{t-1}^{P})] - [\ln(L_t^{P}) - \ln(L_{t-1}^{P})]\},$$

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

$$(1)\ Revision = [\ln(Q_t^{R2}) - \ln(Q_t^{P})] - [\ln(Q_{t-1}^{R2}) - \ln(Q_{t-1}^{P})] - [\ln(L_t^{R2}) - \ln(L_t^{P})] + [\ln(L_{t-1}^{R2}) - \ln(L_{t-1}^{P})].$$

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. A similar expression can be written for the R1-to-R2 revisions. As before, we used data from 2000q1-2019q4 (excluding 2018q4).

Tables 4a and 4b show the average values for each term in equation (1) for prelim-to-R2 and R1-to-R2 revisions and for the average revision to quarterly LP growth. We multiplied each term by 4 so that quarterly changes and revisions are consistent with the annualized growth rates reported in the LPC news release. The average revisions calculated from our decompositions is about the same as the average revisions reported in Table 1.

The decompositions of prelim-to-R2 revisions in Table 4a reveal that the largest portion of the revision to labor productivity is due mainly to revisions to output, rather than revisions to hours. This finding, which is consistent with a study by Anderson and Kliesen (2006) that uses data through 2005, should not be too surprising given the sources of the revisions described in the introduction. Taking the difference between the current and previous quarter revisions to output yields a net revision of 0.12, which accounts for 88 percent of the total revision. Revisions to hours account for the remaining 12 percent (0.02). It is important to note that the revisions to current and prior quarter output tend to be in the same direction and therefore largely offset each other.

The same is true for hours. For prelim-to-R2 revisions, the correlation between current and prior quarter output revisions is 0.89. For hours, this correlation is even higher, 0.98.

Table 4a also highlights significant differences in revisions by quarter, which are mainly due to the revision schedules of the source data. The largest revisions to output occur for first quarter estimates. The revisions to first quarter output estimates tend to be large because they reflect BEA's annual GDP revisions (through Q4 of previous year) that are released in July and are reflected in the R2 estimates for Q1. These revisions directly affect the estimates of the previous year's Q4 output (see Table 2). They also indirectly affect the estimates of the current year's Q1 output, because the current year's Q1 output is calculated by projecting growth from the (revised) Q4 estimate. The revisions to hours are largest for Q3 and Q4.

Table 4b summarizes the R1-to-R2 revisions. As with the prelim-to-R2 revisions, the largest revisions to output occur for Q1 estimates. The revisions to prior quarter output (Q4 of the previous year) are nearly the same for prelim-to-R2 and R1-to-R2 revisions, because BEA's annual revision affects the R2 estimate as noted above (and in Table 2). The revisions to current quarter output are of a similar magnitude as revisions to previous quarter output, again because Q1 output is calculated as the growth in output relative to the revised Q4 output. Revisions to hours are largest for Q1 and Q3. Comparing R1-to-R2 revisions to prelim-to-R2 revisions, the correlation between current and prior quarter revisions are higher for output (0.96 vs. 0.89) and the same for hours (0.98).

The main difference between prelim-to-R2 and R1-to-R2 revisions is that the R1-to-R2 revisions to current quarter output are much smaller for Q2 and Q3. The smaller revisions to output growth in the R1-to-R2 revisions account for a substantial share of the overall difference between prelim-to-R2 and R1-to-R2 revisions. The smaller revisions to output in Q2 and Q3 result in output accounting for only about one-third of the average revision to labor productivity growth over all quarters, compared with almost 90 percent for prelim-to-R2 revision.

This analysis gives us a better understanding about the nature of the revisions. Although there are considerable differences in the revisions to output and hours, there does not seem to be any obvious way to use this information to help us construct intervals. The analysis does suggest that efforts to improve preliminary estimates of output would have a larger impact on reducing revisions than efforts to improve preliminary estimates of hours.

2.3 Business Cycle Influences

Given that our sample period includes two recessions, it is natural to wonder whether revisions for these recession quarters are representative and how they might affect the intervals that we construct below. Although we cannot use information on the current state of the economy to construct intervals,[9] it is useful know whether business-cycle effects might influence our conclusions about the variability of revisions across quarters.

According to the NBER, which dates recessions primarily based on GDP growth, the dates for the two recessions during our sample period were from March through November 2001 and from December 2007 through June 2009.[10] We also considered a second definition that extends the NBER definition to include months after the official end of the recession through the month during which the unemployment rate peaked. By this definition, the recessions extended from March 2001 through June 2003 and from December 2007 through October 2009. Since the start and end dates of the recessions do not exactly coincide with calendar quarters, we define a quarter as a recession quarter if it contains at least two recession months.

Tables 5a and 5b compare summary statistics for recession and non-recession quarters. Under both definitions, revisions are somewhat larger in recession quarters compared with non-recession quarters. However, this difference is largely driven by the large downward revision of 3.8 percent to the 2008q4 preliminary estimate.[11] If this observation is excluded, revisions in recession and non-recession quarters are not significantly different from each other. This is true regardless of how we define a recession quarter.

This analysis implies that there is no information in the revision history that can be used to improve preliminary estimates or to construct narrower intervals. This finding is consistent with Aruoba (2008), who finds that productivity revisions are partly predictable but not well-behaved,

---

[9] Business cycle dates are determined too late to use in any of the relevant news releases.
[10] Recession dates were obtained from http://www.nber.org/cycles.html.
[11] GDP from the fourth quarter of 2008 was revised more than any other quarter on record. (Croushore, 2011, p. 73)

and Jacobs and van Norden (2010, 2016), who find that revisions to productivity are large because revisions to output and labor inputs are not highly correlated.

3. Proposed Methods for Constructing Intervals

Our intervals differ from standard confidence intervals in that we are estimating bounds for a single observation. Our goal is to calculate intervals around prelim and R1 estimates such that the R2 estimates fall within those bounds a given percent of the time. We consider three alternative methods for estimating 70-, 80-, and 90-percent intervals.

The first method is a modified version of standard confidence intervals. The second method is a model-based approach. The third method constructs the intervals using a percentile approach. The modified confidence interval is discussed by Fixler, et al (2014) and Fixler, et al (2018) for BEA's preliminary GDP estimates, while the percentile method is used by the Fed for preliminary estimates of its Industrial Production Index (IPI). The model-based approach allows us to control for quarterly variation in revisions (and potentially other factors as well). We describe these methods for prelim-to-R2 revisions with the understanding that estimates for R1-to-R2 revisions are calculated the same way.

3.1 Modified Confidence Intervals

The first method we considered is a modified version of the standard confidence interval method based on historical revisions, and is the same formula that is presented in Fixler, et al. (2014). The modified confidence intervals are given by:

$$(3)\ Modified\ confidence\ interval =$$

$$\left(\overline{R2 - prelim}\right) \pm\ z_{\alpha/2} \times \sqrt{\frac{1}{n}\sum_{i=1}^{N}\left[(R2 - prelim) - \left(\overline{R2 - prelim}\right)\right]^2}$$

where *prelim* is the preliminary estimate, $\left(\overline{R2 - prelim}\right)$ is the average prelim-to-R2 revision and $z_{\alpha/2}$ is the critical value for α level of significance for a normal distribution. For a given value of α, we would expect the revision to prelim to be within these bounds (1–α) percent of the time. The upper and lower bounds of the interval are centered on the average value of the prelim-to-R2 revision. To adjust the interval so that it is relative to the value of prelim, we add the value of prelim to the upper and lower bounds.

3.2 Model-Based Intervals

Our model-based intervals are generated within a regression framework. One advantage of this approach is that it provides a convenient way to incorporate additional information, such as allowing intervals to vary by quarter. The general strategy is to estimate an OLS regression of R2 on prelim, and use the results of that regression to construct an interval. We first estimated the following regressions:

$$(4) \qquad R2 = \alpha + \beta \times prelim + \varepsilon$$

$$(4') \qquad R2 = \alpha + \beta \times R1 + \varepsilon$$

using data on prelim (R1) and R2 from prior quarters. For the current quarter, q, we generated a predicted value for R2, $\widehat{R2}$, and then constructed an interval around $\widehat{R2}$ using the following equation:

(5) *Model-based interval* =

$$\widehat{R2}_q \pm t_{N-2} \times \; s_{R2}\sqrt{1 + \frac{1}{n} + \frac{(prelim_q - \overline{prelim})^2}{(n-1)s^2_{prelim}}}$$

where $\widehat{R2}_q$ is the predicted value of R2 for quarter q from the regression, $s_{R2} = \sqrt{\frac{\sum_{i=1}^{N}(R2-\widehat{R2})^2}{N}}$ is the root mean squared error from the regression and s^2_{prelim} is the variance of the preliminary estimate. The interpretation of these intervals differs from the modified confidence intervals described above. Here the interval is centered on the predicted value of R2 and tells us the likely values of R2 given a value of prelim (or R1). To adjust the interval so that it is relative to the value of prelim, we subtract prelim from the upper and lower bounds.

Table 6 shows the estimated coefficients from equations (4) and (4') estimated over all 79 observations (2000q1-2019q4). The coefficients from equation (4) imply that preliminary and R2 estimates move together, but that the R2 estimate is on average about 0.15 of a percentage point larger than the prelim estimate. This is consistent with the mean revision reported in Table 1.

However, it is important to note that the constant is not statistically significantly different from zero. Thus, as noted above, we cannot conclude that the preliminary estimate is systematically biased. The results for Equation (4′), for R1-to-R2 revisions, are similar although the constant is somewhat larger than the mean R1-to-R2 revision in Table 1.

To account for the quarterly variation in revisions, we also estimated equations (4) and (4′) with quarterly dummies and recalculated intervals.

3.3 Percentile Intervals

There are two issues with the modified confidence interval and model-based approaches. First, they assume that the historical prediction errors are normally distributed, which Table 1 shows is not the case. And second, these intervals are calculated using squared deviations, which place greater importance on outliers. The percentile approach differs from these approaches in that it does not require any distributional assumptions and outliers are not over-weighted. As with the modified confidence intervals, we calculate intervals using revisions. Under this approach, the upper and lower bounds of, say, a 70-percent interval are calculated using the values of revisions at the 85th and 15th percentiles of the distribution.

One drawback to this approach is that it is potentially sensitive to the method used to calculate the percentiles. This is likely to be the case if there is a large difference between the two observations surrounding a given percentile. This is especially likely when working with small samples. To illustrate, if the sample has exactly 100 observations, each percentile is determined exactly. However, if the two adjacent percentiles are very different, then this method will be sensitive to the inclusion or exclusion of a single observation. When the number of observations is less than 100 (as is the case with our data), the percentile will lie between two observations. We consider three variations of the percentile approach: simple percentile, nearest percentile and weighted percentile.

Like the modified confidence intervals these intervals are for the revisions, but they are centered (approximately) around the median revision. To adjust the intervals so that they are relative to the value of prelim, we add prelim to the upper and lower bounds.

3.3.1 Simple and Nearest Percentile Intervals

The simple percentile approach uses the value of the first observation below the desired percentile for the lower bound and the value of the first observation above the desired percentile for the upper bound. This approach potentially results in intervals that are wider than necessary. The nearest percentile uses the value of the observation whose percentile rank is closest to the desired percentile. To illustrate, for the 95th percentile (the upper bound of a 90-percent interval) the two observations surrounding the 95th percentile are the 75th (94.94 percentile) and 76th (96.20 percentile) largest observations. The value of the 76th observation would be used for the simple percentile, while the value of the 75th observation would be used for the nearest percentile method. Note that given the compactness of the distribution, these observations often have the same value. The simple and nearest percentiles will differ only when the nearest percentile is below (above) the upper (lower) bound of the interval.

3.3.2 Weighted Percentile Prediction Intervals

Weighted percentile prediction intervals differ from the simple and nearest percentile prediction intervals in that they always account for the information contained in the surrounding observations. In the above example, the weighted percentile will be a weighted average of the simple and nearest percentile. Specifically, for each percentile, we used a weighted average of the values of the two surrounding observations, where the weights are equal to those numbers' distances from the stated percentile.

To calculate the lower weighted percentile we sorted the data by size of revision, and used the following formula:

(6) $Lower\ weighted\ percentile =$

$$Revision_1 \times \left(1 - \frac{(\mathrm{Percentile}_{lower} - \mathrm{Percentile}_1)}{(\mathrm{Percentile}_2 - \mathrm{Percentile}_1)}\right) +$$

$$Revision_2 \times \left(1 - \frac{(\mathrm{Percentile}_2 - \mathrm{Percentile}_{lower})}{(\mathrm{Percentile}_2 - \mathrm{Percentile}_1)}\right)$$

where subscript 1 denotes the first observation below the desired percentile and 2 denotes the first observation above the desired percentile. The *lower* subscript indicates the lower bound of the specified interval (5, 10, or 15 for 90-, 80-, or 70-percent intervals.) The formula for the upper weighted percentile is analogous:

$$(6') \quad Upper\ weighted\ percentile =$$

$$Revision_1 \times \left(1 - \frac{\left(\text{Percentile}_{upper} - \text{Percentile}_1\right)}{\left(\text{Percentile}_2 - \text{Percentile}_1\right)}\right) +$$

$$Revision_2 \times \left(1 - \frac{\left(\text{Percentile}_2 - \text{Percentile}_{upper}\right)}{\left(\text{Percentile}_2 - \text{Percentile}_1\right)}\right)$$

In cases where the two numbers above and below the desired percentile are the same, the weighted percentile confidence interval is the same as the simple (and nearest) percentile confidence interval. But, in general, we would expect the weighted confidence interval to be narrower than the simple percentile and wider than the nearest percentile.

4. Comparison of Methodologies and Recommendations

In this section we compare results from the different methods. For each methodology, we show the upper and lower bounds of the intervals, and the width of the interval for 90-percent, 80-percent and 70-percent intervals. Our goal is to compare the methods with respect to the frequency that R2 falls within the interval around prelim. To make this comparison, we express the upper and lower bounds of the intervals relative to prelim (or R1).

Our cross-validation methodology compares the methods using a leave-one-out approach to determine whether the intervals perform “as advertised.” That is, we wish to determine whether the percent of R2 values that fall within the specified interval is close to the stated confidence level for each methodology.

4.1 Cross validation Methodology

Typically, validation is done by generating the interval using the first N-1 observations, and validating the method using the omitted observation. To increase the number of test cases, we used the leave-one-out approach for cross validation. For each interval method, revision type (prelim-to-R2 or R1-to-R2), method, and confidence level (70-, 80-, or 90-percent), we:

(1) Drop the first observation and estimate intervals as described above over the remaining 78 observations.
(2) Repeat step (1), replacing the first observation and omitting the second observation. This process is repeated for each subsequent observation in sequence. Thus, for each observation, we have the following:
    a. A value for prelim
    b. A value for R2
    c. An interval around prelim
(3) For each observation, determine whether R2 lies within the interval around prelim.
(4) For each revision x method x confidence level cell, we calculate the "hit rate," which is the fraction of R2 values that lie within the interval constructed around prelim (R1).

4.2 Results

Table 7a shows our results for prelim-to-R2 revisions. The table shows the average value of the upper and lower bounds relative to the value of prelim, the interval width, and the fraction quarters in which the value of R2 falls within the interval.

As expected, the 90-percent intervals are the widest, followed by the 80- and 70-percent intervals. All of the methodologies generate fairly similar intervals. The modified confidence intervals and model-based intervals are narrower than the percentile intervals at the 90-percent confidence level, but are wider at the 80- and 70-percent levels. The hit rates at the 90-percent and 80-percent levels are about the same as the stated confidence level or a little higher. But at the 70-percent level, the hit rates for the modified confidence intervals and the model-based intervals are substantially higher than the stated confidence level, whereas the hit rates for the percentile intervals are about 70 percent. The larger difference in hit rates across methods for the 70-percent intervals occurs because there is a lot more bunching at the 15^{th} and 85^{th} percentiles than at the 5^{th} and 95^{th} percentiles.[12] Thus small differences in the 70-percent interval can have large

[12] Much of the bunching is due to rounding.

impact on the hit rate. This explains why the three percentile methods have different hit rates even though the intervals are nearly the same.

We noted above that the model-based approach allows us to control for quarterly differences in the size of the average revision. However, as can be seen in Table 7, controlling for quarterly variation in revisions did not make much difference in the width of the intervals, although there was a slight improvement in the hit rate.

The story is somewhat different for R1-to-R2 revisions (Table 7b). Here the biggest differences between the methods shows up in the 70-percent intervals. The model-based approach generates much wider intervals than the two percentile approaches because their formulas square prediction errors, which gives greater weight to outliers. Recall that most of the mass of the distribution of R1-to-R2 revisions is concentrated in the center of the distribution. The differences in the upper and lower bounds of the 90-percent intervals are fairly small because all observations at these ends of the distribution are, in some sense, outliers.

All of the intervals were calculated from growth rates that were calculated to a single decimal place of precision (like the published estimates). However, for later years (2006 onward), growth rates are available at a much greater level of precision (thousandths instead of tenths). To see if this mattered, we recalculated the intervals using the more-precise estimates. For the earlier years, we perturbed single-decimal estimates by adding a uniformly distributed random term, $U(-0.05,0.05)$, to each estimate, and then recalculated the intervals. We repeated this exercise 100 times. For each iteration, we calculated an average of the intervals. We then calculated an average of the averages. The results were virtually identical to those reported in Tables 7a and 7b.

4.3 Changes to the LPC News Release

Based on these results, the Reliability Note in the LPC News Release will be modified to include 70-, 80, and 90-percent intervals calculated using the weighted percentile method and the 20 most recent years of data. The percentile methods generate hit rates that are very close to the stated confidence level, and they are less sensitive to outliers than the other approaches. The weighted percentile method is relatively easy to calculate and explain to users. We believe that presenting multiple intervals at different levels of statistical significance will give data users a clearer picture regarding the expected size of revisions and make the estimates more useful. For example, forecasters could, in addition to using the point estimate, generate alternative predictions using the upper and lower bounds of the intervals.

The choice of the most recent 20 years of data balances the need for sufficient sample to generate intervals with the need for data that reflect how these estimates are currently being revised. We experimented with extending our data series back to 1994. However, our analysis (see Appendix 2) revealed that there were significant differences between the earlier period (1994-2006 and the later period 2000-2019—we used overlapping periods because there were not enough data).

5. Implications of Revisions in the COVID-19 Era

The COVID-19 pandemic severely disrupted economic activity and placed unprecedented demands on a statistical system that was not designed to measure the rapid changes that we saw in early 2020. First quarter estimates of labor productivity were particularly affected, because the sharp decline in economic activity occurred in the last two weeks of the quarter. On the output side, the BEA's use of projections for its advance estimate is particularly problematic 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 missed most of the declines in employment because the declines occurred largely after the reference periods of those surveys.

Both BEA and BLS quickly adapted to the new environment and modified their methods to provide a more accurate picture of output and productivity growth. For its Advance Estimate of GDP, BEA modified its procedures by incorporating high frequency data, and relying less on projections. The BLS Productivity Program modified its usual procedures for estimating hours worked using data on initial UI claims for its preliminary estimate.

For the preliminary Q1 estimate, the productivity program modified its methodology using weekly UI Initial Claims data. Employment was estimated week-by-week under the implicit assumption that the UI Initial Claims reflected actual job losses and that there were no transitions from non-employment to employment. These are admittedly 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. Although this adjustment was large, only wage and salary employment data were adjusted. Self-employed worker hours and average weekly hours of wage and salary workers were not adjusted, because there were no data on which to base an adjustment. However, once the April data were available, it became feasible to generate week-by-week estimates of hours by interpolating between the March and April hours estimates. This adjustment accounted for most of the 1.8 percentage point downward revision in hours worked. Output was revised downward by 0.3 of a percentage point.

Table 8 summarizes the revisions to labor productivity in 2020. The largest revision was the 3.3 percent prelim-to-R1 revision for Q2, which was entirely due to the revision to output. The next largest revision was the prelim-to-R2 revision for Q1, which was mostly due to revisions to hours. To put these revisions into perspective, the 3.3 percentage point revision for Q2 and the 2.2 percentage point revision for Q1 are among the largest revisions since 2000q1. It is worth pointing out that the large prelim-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.

However, given that percentile-based intervals are not sensitive to outliers, the 2020 revisions, though large, should not have a major impact on the width of percentile-based intervals.

6. Summary and Concluding Remarks

We have analyzed the regularly scheduled revisions to the BLS's quarterly labor productivity estimates with the goal of developing intervals that convey to data users the probable magnitude of future revisions. We found that there was no trend in revisions over time, no relationship between the magnitude of the estimate and the size of the revision, and that there were no business cycle effects. We did find some variation in the size of the revisions across quarters, but the differences were not statistically significant. Decomposing the revisions to labor productivity growth, we found that revisions to output accounted for the largest share of average prelim-to-R2 revisions, while the R1-to-R2 revisions were more evenly divided between revisions to output and revisions to hours.

This paper focused on the first three estimates of labor productivity growth. As we noted in the introduction, estimates can be revised long after the reference quarter. The next logical step is to examine these long-run revisions.

Appendix 1: Additional Summary Statistics and Diagnostics

In this Appendix, we present supplemental material on the prelim-to-R2 and the R1-to-R2 revisions. We begin by presenting some additional summary statistics on the two revisions, and then turn to the diagnostic regressions.

Normality of Revisions

The skewness statistics in Table 1 suggest that the historical errors are left skewed, and therefore do not have a normal distribution. We performed several statistical tests (see Table A1) and all of them reject normality.

We also tested the two components of labor productivity—output and hours—for normality (see Table A2). We can reject normality for all of these revisions except for prelim-to-R2 revisions to hours.

Correlation Between Revisions

We tested whether the prelim-to-R1 revisions are correlated with the R1-to-R2 revisions for each of our three variables. We have no reason to believe that they are correlated. But if they are, such information could potentially be exploited in constructing intervals.

The two tables below present correlations between prelim-to-R1 and R1-to-R2 revisions for reference quarters 1994q1 to 2019q4. We have 104 observations for hours, and 103 for output and productivity growth. The difference is due to the government shutdown in early 2019 which resulted in no prelim estimate for output (and therefore labor productivity) for 2018 Q4. There are 24 quarters in the sample in the 1990s, 40 in the 2000s, and 39 in the 2010s for output and productivity, and 40 in the 2010s for hours.

The first table shows that the correlations between prelim-to-R1 and prelim-to-R2 revisions are not consistently positive or negative, and are not far from zero for any of our variables. The second table shows same correlations for the absolute values of the prelim-to-R1 and R1-to-R2. Again, the correlations are small and not consistently positive or negative.

Correlation between prelim-to-R1 revision and R1-to-R2 revision amounts

Reference quarters	Hours	Output	Productivity
Overall 1994-2019	-0.012	-0.006	-0.041
1990s	0.148	-0.009	0.004
2000s	-0.140	-0.079	-0.116
2010s	-0.009	0.039	-0.032

Correlation between prelim-to-R1 revision and R1-to-R2 revision magnitudes

Reference quarters	Hours	Output	Productivity
Overall 1994-2019	-0.021	0.003	0.037
1990s	-0.020	-0.201	-0.171
2000s	-0.005	-0.043	0.101
2010s	-0.142	0.204	0.179

Other Summary Statistics

Table A3 presents some summary statistics for the magnitude (absolute value of) revisions over the 2000q1—2019q4 period. These statistics tell the same story as those in Table 1 in the text. R1-to-R2 revisions are on average smaller than prelim-to-R2 revisions, and the distribution of R1-to-R2 revisions is more compact. The only difference is that the distribution of the magnitude of R1-to-R2 revisions is more skewed than that for the prelim-to-R2 revisions.

Analysis of Trends and Bias

As noted above the mean and median prelim-to-R2 and R1-to-R2 revisions suggest a slight upward bias. We examine possible bias in a regression framework. Specifically, we are interested in whether there is a trend in the revisions and whether the revisions are related to the values of prelim and R1.

For the trend analysis, we estimated the following equations:

(A1) $$revision\_prelim\_R2 = \alpha + \beta \times time\_trend + \varepsilon$$

(A1′) $$revision\_R1\_R2 = \alpha + \beta \times time\_trend + \varepsilon$$

Table A4a shows the results. The coefficient on the time trend is essentially zero in both regressions indicating no upward or downward trend in revisions over time. The constant terms are about the same magnitude as the mean revisions shown in Table 1, but they are not statistically different from zero. It is also worth noting that the low R^2 indicates that the time trend explains none of the variation in the revisions.

To see if there is a trend in the absolute magnitude of the revisions, we re-estimated equations (A1) and (A2) but defined the right-hand-side variables are the absolute value of the revisions. Again, there is no evidence of a trend in the magnitude of the revisions (Table (A4b)).

We estimated similar equations to determine whether there is a relationship between the R2 estimate and the value of prelim or R1, because this information could potentially be used to improve our intervals. For this analysis, we ran the following regressions:

(A2) $$revision\_prelim\_R2 = \alpha + \beta \times prelim + \varepsilon$$

(A2′) $$revision\_R1\_R2 = \alpha + \beta \times R1 + \varepsilon$$

Tables A5a and A5b show the results for equations (A2) and (A2′) and the analogous absolute value versions. The small and not statistically coefficients on prelim and R1 indicate that there is no relationship between value of the estimate and the absolute magnitude of the revisions. Tables A5b and A5c show that there is no relationship between the absolute magnitude of the estimate and the size of the revision.

Quarterly Variation

The summary statistics in the text suggest that there may be some variation by quarter, but that the variation is not statistically significant. To further examine this issue, we estimated two sets of regressions with quarterly dummy variables.

(A2) $$revision\_prelim\_R2 = \beta_t \times time\_trend + \textstyle\sum_{q=1}^{4} \beta_q I_q + \varepsilon$$

(A2′) $$revision\_R1\_R2 = \beta_t \times time\_trend + \textstyle\sum_{q=1}^{4} \beta_q I_q + \varepsilon$$

(A3) $$revision\_prelim\_R2 = \beta_P \times prelim + \textstyle\sum_{q=1}^{4} \beta_q I_q + \varepsilon$$

(A3′) $$revision\_R1\_R2 = \beta_{R1} \times R1 + \sum_{q=1}^{4} \beta_q I_q + \varepsilon$$

The results are shown in Tables A6 and A7. Both sets of regressions indicate that quarterly differences in revisions are not statistically significant.

Appendix 2: Sample Period Considerations

Our sample period covers the period 2000q1 through 2019q4, but complete data starting in 1994q1 are available. Given the age of these earlier data, one concern is whether those data are representative of the current pattern of revisions. To examine this, we consider two time periods. Ideally we would have divided the 1994-2019 period into two mutually exclusive sub-periods, but there are not enough observations to do this. So we consider two overlapping time periods—1994-2006 and 2000-2019. Another consideration is whether the Great Recession had an impact on the revisions. So we also examined revisions excluding the Great Recession quarters.

Our first step is to compare summary statistics for the two time periods. Table A8a replicates Table 1, but adds columns for 1994-2006. The main takeaway from Table A8a is that the mean revisions are larger in the earlier period. The mean prelim-to-R2 revision is 0.40 vs. 0.14 for the latter period. And the mean R1-to-R2 revision is 0.18 vs. 0.04. Other summary statistics have similar values, although the prelim-to-R2 revisions are also less peaked in the earlier period. Table 8b shows the same summary statistics for the absolute value of revisions. These revisions are more similar in the two time periods. The reasons for this can be seen in Charts A1 and A2. For both prelim-to-R2 and R1-to-R2 revisions, there were more large-value negative revisions in the latter period. [13]

Tables A9 and A10 compare the revisions quarter by quarter for the 1994-2006 and 2000-2019 periods. Table A9 compares prelim-to-R2 revisions, and Table A10 compares R1-to-R2 revisions.

[13] Another consideration is whether the Great Recession had an impact on the revisions. When we omitted 2008 from the sample, there was only a minor impact on the summary statistics.

Looking at Table A9, in addition to differences in the size of the average revisions, there are also differences in how those revisions are distributed across quarters. The most notable difference is for Q1. In the earlier period, the revisions are large and positive whereas in the latter period the revisions are negative and about half the size in magnitude. In Table 10, we see a similar pattern for R1-to-R2 revisions, but the magnitudes are smaller.

Tables A11 and A12 compare the sources of revisions in the two time periods. For ease of comparison, we duplicated the decompositions in Tables 4a and 4b. For prelim-to-R2 revisions (Table A11), we can see that revisions to current quarter output were on average smaller. But the large revisions to current output (both current and previous quarter) occurred in Q2 and Q3, rather than Q1 as in the later periods. Revisions to hours were larger in the earlier period.

Comparisons of R1-to-R2 revisions in Table A12 also reveal differences between the two time periods. Revisions to output (both current and previous quarter) are larger in the latter period, but they completely offset one another. The large revisions to output occur in Q1 and Q2 in the earlier period, compared with Q1 in the latter period. Revisions to hours are larger in the earlier period, mainly due to larger revisions in Q1.

References

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

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

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

Fixler, Dennis J., Ryan Greenaway-McGrevy, and Bruce T. Grimm. 2011. "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*. https://www.bea.gov/scb/pdf/2018/01-January/0118-revisions-to-gdp-gdi-and-their-major-components.pdf

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

Galvao, Ana Beatriz, James Mitchell, and Johnny Runge. 2019. 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

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.

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.

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

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

Table 1: Summary Statistics of Revisions

	Quarter-to-Quarter		Year-Over-Year	
	Prelim-to-R2	R1-to-R2	Prelim-to-R2	R1-to-R2
Mean	0.14	0.04	0.03	-0.01
Median	0.3	0.1	0.1	0.0
10^{th} Percentile	-1.1	-0.5	-0.5	-0.2
90^{th} Percentile	1.3	0.6	0.4	0.2
Std. Dev.	1.04	0.64	0.32	0.24
Skewness	-1.17	-0.76	-0.99	-1.69
Kurtosis	5.41	7.10	3.53	7.31
Sign Changes	7	6	3	1
Observations	79	79	79	79

Note: The sample period is 2000q1-2019q4. 2018q4 was dropped due to the government shutdown. Table 1 shows simple percentiles. Skewness and kurtosis for a normal distribution would be 0.0 and 3.0.

Chart 1: Revisions – Quarter-to-Quarter (2000-2019)

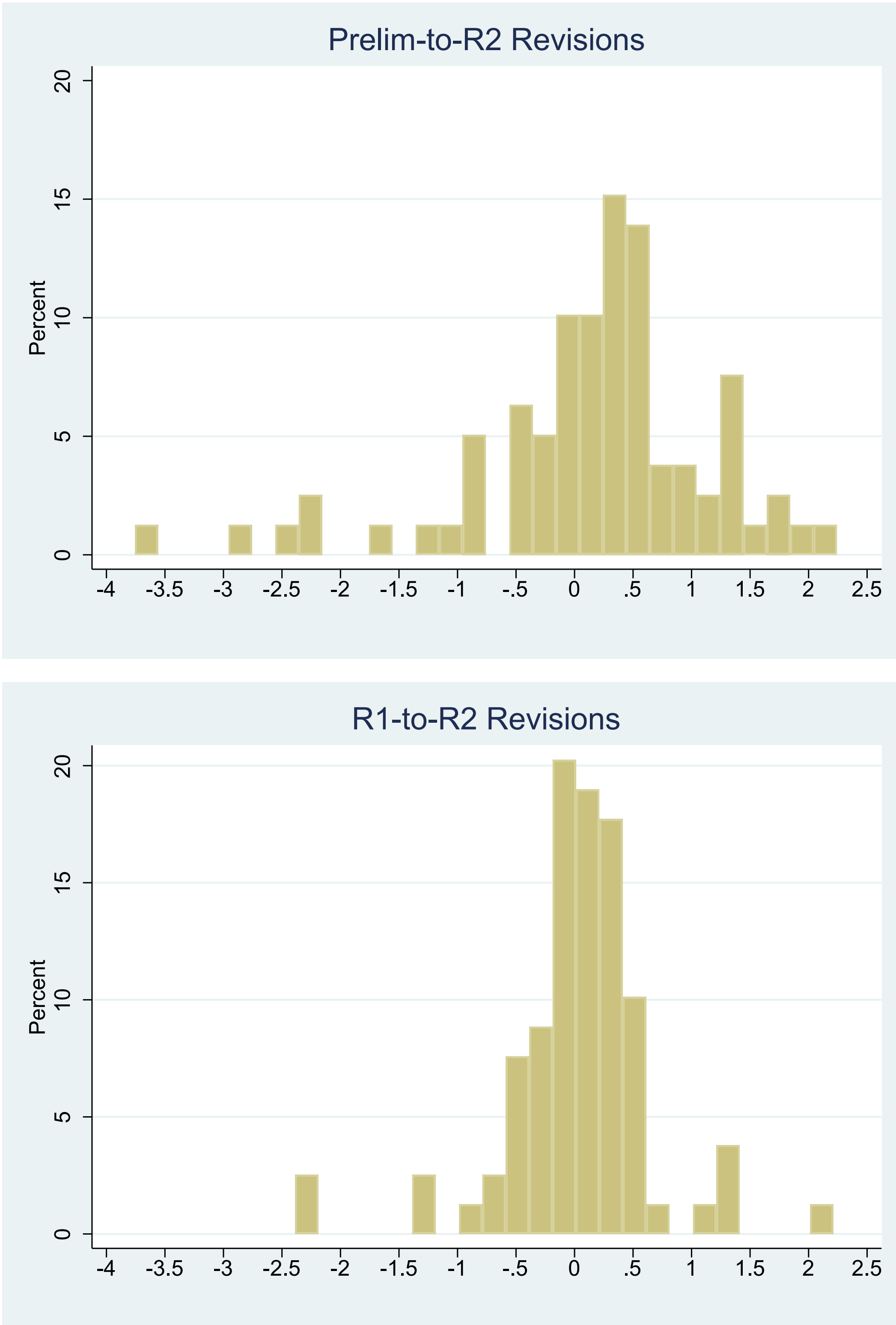
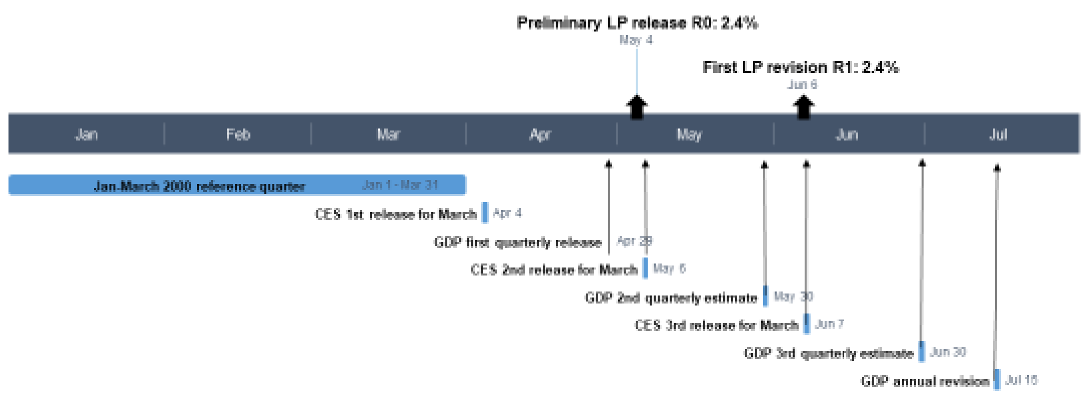


Table 2: Annual Data Calendar

Month	PFEI Releases	Revision notes
Feb.	**Prelim** for Q4, **R2** for Q3	
March	**R1** for Q4	Incorporates the annual CES benchmark revision through Q4 of the previous year. This affects both current and prior quarter hours.
May	**Prelim** for Q1, **R2** for Q4	
June	**R1** for Q1	
Aug.	**Prelim** for Q2, **R2** for Q1	Both prelim and R2 estimates incorporate the annual NIPA/GDP benchmark revision (current and prior quarter output).
Sept.	**R1** for Q2	
Nov.	**Prelim** for Q3, **R2** for Q2	
Dec.	**R1** for Q3	

Chart 3: Revisions to 2000Q1 labor productivity



Revisions to 2000Q1 labor productivity from histCV/prod_q_nf_updated.xlsx
Peter to add credits here for the web site that helps build timeline charts like this.

Table 3a: Prelim-to-R2 Revisions by Quarter

Period	N	Mean	Std. dev.	Min	Max	Mean absolute revision	Sign Changes
All Qs	79	0.14	1.04	-3.8	2.2	0.76	7
Q1	20	-0.24	1.09	-2.8	1.1	0.76	4
Q2	20	0.49	0.67	-0.9	2.2	0.62	0
Q3	20	0.24	1.10	-2.3	1.7	0.88	0
Q4	19	0.07	1.18	-3.8	2.0	0.77	3

Table 3b: R1-to-R2 Revisions by Quarter

Period	N	Mean	Std. dev.	Min	Max	Mean absolute revision	Sign Changes
All Qs	79	0.04	0.64	-2.4	2.0	0.43	6
Q1	20	-0.10	1.06	-2.4	2.0	0.72	4
Q2	20	0.12	0.47	-0.7	1.4	0.37	0
Q3	20	0.02	0.48	-1.0	1.3	0.35	1
Q4	19	0.13	0.30	-0.6	0.5	0.26	1

Table 4a: Decomposition of Prelim-to-R2 Revisions

	Average Revision to:				
	Output		Hours		
	Current Quarter	Previous Quarter	Current Quarter	Previous Quarter	Total
All Quarters	-0.16	-0.27	-0.06	-0.05	0.13
Q1	-1.29	-1.01	-0.15	-0.12	-0.25
Q2	0.25	-0.13	-0.11	0.00	0.49
Q3	0.38	0.02	0.32	0.18	0.22
Q4	0.04	0.04	-0.32	-0.25	0.07

Table 4b: Decomposition of R1-to-R2 Revisions

	Average Revision to:				
	Output		Hours		
	Current Quarter	Previous Quarter	Current Quarter	Previous Quarter	Total
All Quarters	-0.29	-0.30	0.00	0.02	0.03
Q1	-1.26	-1.07	-0.23	-0.16	-0.12
Q2	-0.04	-0.13	-0.05	0.00	0.14
Q3	0.04	0.01	0.29	0.26	-0.01
Q4	0.13	0.00	-0.03	-0.04	0.12

Note: The sample period is 2000q1-2019q4. 2018q4 was deleted due to the government shutdown.

Table 5a: Summary Statistics for Prelim-to-R2 Revision

Sample	N	Mean	Std. dev.	Min	Max	Sign Changes
Revisions	79	0.14	1.04	-3.8	2.2	7
Absolute value of revisions	79	0.76	0.72	0.0	3.8	
Statistics based on NBER-defined recession quarters						
Revisions in non-recession quarters	70	0.18	0.94	-2.8	2.2	6
Revisions in recession quarters	9	-0.16	1.71	-3.8	2.0	1
Absolute revisions in non-recession quarters	70	0.70	0.64	0.0	2.8	
Absolute revisions recession quarters	9	1.22	1.13	0.3	3.8	
Statistics based on labor-defined recession quarters						
Revisions in non-recession quarters	63	0.16	0.90	-2.8	2.2	5
Revisions in recession quarters	16	0.07	1.51	-3.8	2.0	2
Absolute revisions in non-recession quarters	63	0.66	0.63	0.0	2.8	
Absolute revisions recession quarters	16	1.13	0.96	0.0	3.8	

Table 5b: Summary Statistics for R1-to-R2 Revisions

Sample	N	Mean	Std. dev.	Min	Max	Sign Changes
Revisions	79	0.04	0.64	-2.4	2.0	6
Absolute value of revisions	79	0.43	0.48	0.0	2.4	
Statistics based on NBER-defined recession quarters						
Revisions in non-recession quarters	70	0.07	0.66	-2.4	2.0	6
Revisions in recession quarters	9	-0.17	0.52	-1.2	0.3	0
Absolute revisions in non-recession quarters	70	0.43	0.50	0.0	2.4	
Absolute revisions recession quarters	9	0.38	0.38	0.3	1.2	
Statistics based on labor-defined recession quarters						
Revisions in non-recession quarters	63	0.07	0.68	-2.4	2.0	6
Revisions in recession quarters	16	-0.09	0.49	-1.2	0.4	0
Absolute revisions in non-recession quarters	63	0.44	0.52	0.0	2.4	
Absolute revisions recession quarters	16	0.35	0.33	0.0	1.2	

Note: Sample period is 2000q1-2019q4. 2018q4 was dropped due to the government shutdown.

Table 6: Regression Results

	Dependent Variable: R2			
	Equation (4)		Equation (4’)	
	Coefficient	SE	Coefficient	SE
prelim	0.995	0.052		
R1			0.976	0.029
Constant	0.149	0.157	0.091	0.096
R-squared	0.828		0.935	

Table 7a: Results for prelim-to-R2 Revisions

90-percent Intervals

Method	Lower*	Upper*	Interval Width*	Percent in Interval
Modified CI	-1.57	1.84	3.41	89.9
Model-based	-1.63	1.91	3.54	89.9
w/Q dummies	-1.60	1.88	3.48	89.9
Percentile				
Simple	-2.40	1.69	4.09	91.1
Nearest	-2.29	1.49	3.79	88.6
Weighted	-2.31	1.51	3.82	88.6

80-percent

Method	Lower*	Upper*	Interval Width*	Percent in Interval
Modified CI	-1.19	1.47	2.66	84.8
Model-based	-1.23	1.51	2.75	83.5
w/Q dummies	-1.21	1.49	2.70	81.0
Percentile				
Simple	-1.19	1.30	2.49	82.3
Nearest	-1.08	1.30	2.38	81.0
Weighted	-1.10	1.30	2.40	81.0

70-percent

Method	Lower*	Upper*	Interval Width*	Percent in Interval
Modified CI	-0.94	1.21	2.15	76.0
Model-based	-0.97	1.25	2.22	76.0
w/Q dummies	-0.95	1.23	2.18	78.5
Percentile				
Simple	-0.89	1.08	1.97	70.9
Nearest	-0.75	1.00	1.75	69.6
Weighted	-0.79	1.03	1.82	69.6

* Average over all prediction intervals. Interval widths may not be consistent with upper and lower bounds due to rounding.

Table 7b: Results for R1-to-R2 Revisions

90-percent

Method	Lower*	Upper*	Interval Width*	Percent in Interval
Modified CI	-1.02	1.10	2.11	88.6
Model-based	-1.05	1.13	2.18	89.9
w/Q dummies	-1.06	1.14	2.20	88.6
Percentile				
Simple	-1.30	1.29	2.59	92.4
Nearest	-1.28	1.07	2.35	88.6
Weighted	-1.28	1.09	2.37	88.6

80-percent

Method	Lower*	Upper*	Interval Width*	Percent in Interval
Modified CI	-0.78	0.86	1.64	87.3
Model-based	-0.81	0.89	1.69	87.3
w/Q dummies	-0.81	0.89	1.71	87.3
Percentile				
Simple	-0.59	0.59	1.18	81.0
Nearest	-0.50	0.50	1.00	81.0
Weighted	-0.52	0.52	1.04	81.0

70-percent

Method	Lower*	Upper*	Interval Width*	Percent in Interval
Modified CI	-0.62	0.71	1.33	86.1
Model-based	-0.64	0.72	1.37	87.3
w/Q dummies	-0.63	0.73	1.38	83.5
Percentile				
Simple	-0.40	0.50	0.90	78.5
Nearest	-0.40	0.48	0.88	72.2
Weighted	-0.40	0.49	0.89	72.2

* Average over all prediction intervals. Interval widths may not be consistent with upper and lower bounds due to rounding.

Table 8: 2020 Revisions to Labor Productivity Growth

	Labor Productivity Estimates			Revisions	
	Prelim	**R1**	**R2**	**Prelim-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				

Table A1: Tests for Normality of Revisions to Labor Productivity

Prelim to R2 revision

Test	Statistic	P-Value	Normal
Shapiro - Wilk	0.915	0.0002	No
Kolmogorov - Smirnov	0.148	<0.0100	No
Cramer-von Mises	0.338	<0.0050	No
Anderson-Darling	1.872	<0.0050	No

R1 to R2 revision

Test	Statistic	P-Value	Normal
Shapiro - Wilk	0.905	0.0001	No
Kolmogorov - Smirnov	0.138	<0.0100	No
Cramer-von Mises	0.341	<0.0050	No
Anderson-Darling	2.073	<0.0050	No

Table A2: Tests for Normality of Revisions to Output and Hours (net revisions to current and prior quarter estimates)

Measure and Test	Statistic	P-Value	Normal
	Prelim to R2 revisions		
Output			
Shapiro - Wilk	0.949	0.0059	No
Kolmogorov - Smirnov	0.124	<0.0100	No
Cramer-von Mises	0.190	0.0071	No
Anderson-Darling	1.104	0.0067	No
Hours			
Shapiro - Wilk	0.971	0.1006	Yes
Kolmogorov - Smirnov	0.073	>0.1500	Yes
Cramer-von Mises	0.050	>0.2500	Yes
Anderson-Darling	0.414	>0.2500	Yes
	R1 to R2 revisions		
Measure and Test	**Statistic**	**P-Value**	**Normal**
Output			
Shapiro - Wilk	0.875	<0.0001	No
Kolmogorov - Smirnov	0.158	<0.0100	No
Cramer-von Mises	0.439	<0.0050	No
Anderson-Darling	2.608	<0.0050	No
Hours			
Shapiro - Wilk	0.816	<0.0001	No
Kolmogorov - Smirnov	0.199	<0.0100	No
Cramer-von Mises	0.824	<0.0050	No
Anderson-Darling	4.449	<0.0050	No

Table A3: Summary Statistics for Absolute Value of Revisions

	Quarter-to Quarter	
	Prelim-to-R2	R1-to-R2
Mean	0.76	0.43
Median	0.5	0.3
10th Percentile	0.1	0.0
90th Percentile	1.7	1.2
Std. Dev.	0.72	0.48
Skewness	1.67	2.29
Kurtosis	6.08	8.46

Note: Table A3 shows simple percentiles.

Table A4a: Time Trend Regressions: Prelim-to-R2 and R1-to-R2 Revisions

	(A1) **prelim-to-R2 Revisions**		(A1′) **R1-to-R2 Revisions**	
	Coefficient	SE	Coefficient	SE
Time trend	0.00	0.01	0.00	0.00
Constant	0.21	0.35	0.00	0.21
R-squared	0.001		0.0004	

Table A4b: Time Trend Regressions: Prelim-to-R2 and R1-to-R2 Revisions
Absolute Value of Prelim-to-R2 and R1-to-R2 Revisions

	(A1) **prelim-to-R2 Revisions**		(A1′) **R1-to-R2 Revisions**	
	Coefficient	SE	Coefficient	SE
Time trend	0.00	0.00	0.00	0.00
Constant	0.90	0.24	0.40	0.16
R-squared	0.0053		0.0005	

Table A5a: Regressions of Prelim-to-R2 and R1-to-R2 Revisions on the Value of the Preliminary Estimate

	(A2)		(A2′)	
	prelim-to-R2		**R1-to-R2**	
	Coefficient	SE	Coefficient	SE
Prelim	-0.01	0.05		
R1			-0.02	0.03
Constant	0.15	0.16	0.09	0.10
R-squared	0.000		0.009	

A5b: Regressions of Prelim-to-R2 and R1-to-R2 Revisions on the Absolute Value of the Preliminary and R1 Estimates

	(A2)		(A2′)	
	prelim-to-R2		**R1-to-R2**	
	Coefficient	SE	Coefficient	SE
abs(Prelim)	-0.01	0.06		
abs(R1)			0.00	0.04
Constant	0.17	0.19	0.05	0.12
R-squared	0.001		0.000	

Table A5c: Regressions of the Absolute Value of Prelim-to-R2 and R1-to-R2 Revisions on the Absolute Value of the Preliminary and R1 Estimates

	(A2)		(A2′)	
	abs(prelim-to-R2)		**abs(R1-to-R2)**	
	Coefficient	SE	Coefficient	SE
abs(Prelim)	0.07	0.04		
abs(R1)			0.00	0.03
Constant	0.60	0.13	0.43	0.09
R-squared	0.030		0.000	

Table A6: Time Trend Regressions with Quarterly Dummies

	Prelim-to-R2 Revisions		R1-to-R2 Revisions	
	Coefficient	SE	Coefficient	SE
Time trend	0.00	0.01	0.00	0.00
Q1	-0.16	0.39	-0.13	0.25
Q2	0.57	0.39	0.09	0.25
Q3	0.32	0.40	-0.02	0.25
Q4	0.15	0.40	0.09	0.03
R-squared	0.068		0.022	

Table A7: Regressions of Revisions on the Value of Preliminary Estimates with Quarterly Dummies

	Prelim-to-R2 Revisions		R1-to-R2 Revisions	
	Coefficient	SE	Coefficient	SE
Prelim	-0.02	0.05		
R1			-0.03	0.03
Q1	-0.21	0.25	-0.06	0.15
Q2	0.54	0.26	0.19	0.16
Q3	0.31	0.28	0.11	0.18
Q4	0.10	0.25	0.16	0.15
R-squared	0.070		0.032	

Table A8a: Comparison of Revisions

	Prelim-to-R2		R1-to-R2	
	1994-2006	2000-2019	1994-2006	2000-2019
Mean	0.40	0.14	0.18	0.04
Median	0.3	0.3	0.1	0.1
10^{th} Percentile	-0.8	-1.1	-0.4	-0.5
90^{th} Percentile	1.9	1.3	0.6	0.6
Std. Dev.	1.04	1.04	0.64	0.64
Skewness	0.54	-1.17	1.66	-0.76
Kurtosis	2.84	5.41	6.81	7.10
Observations	52	79	52	79

Note: Table A9a shows simple percentiles. Skewness and kurtosis for a normal distribution would be 0.0 and 3.0.

Table A8b: Comparison of Absolute Value of Revisions

	Prelim-to-R2		R1-to-R2	
	1994-2006	2000-2019	1994-2006	2000-2019
Mean	0.86	0.76	0.41	0.43
Median	0.6	0.5	0.3	0.3
10^{th} Percentile	0.2	0.1	0.0	0.0
90^{th} Percentile	1.9	1.7	1.2	1.2
Std. Dev.	0.70	0.72	0.52	0.48
Skewness	1.27	1.67	2.49	2.29
Kurtosis	4.49	6.08	8.93	8.46
Observations	52	79	52	79

Note: Table A9b shows simple percentiles. Skewness and kurtosis for a normal distribution would be 0.0 and 3.0.

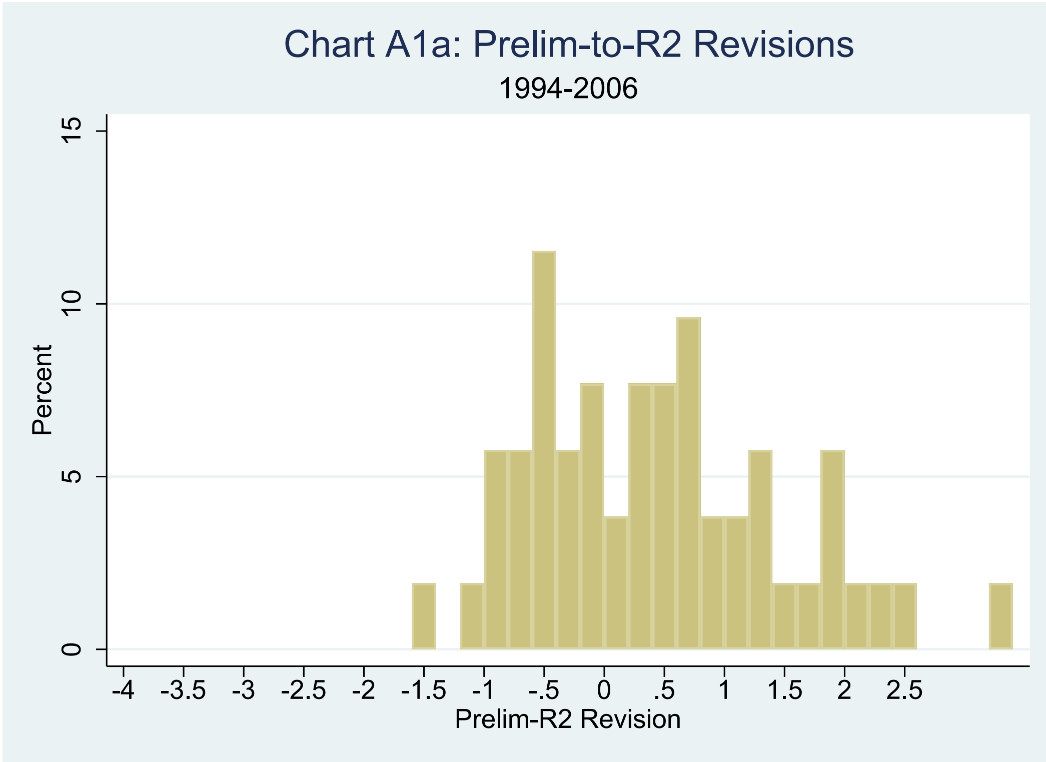


Chart A1a: Prelim-to-R2 Revisions
1994-2006
Percent
0
5
10
15
-4
-3.5
-3
-2.5
-2
-1.5
-1
-.5
0
.5
1
1.5
2
2.5
Prelim-R2 Revision

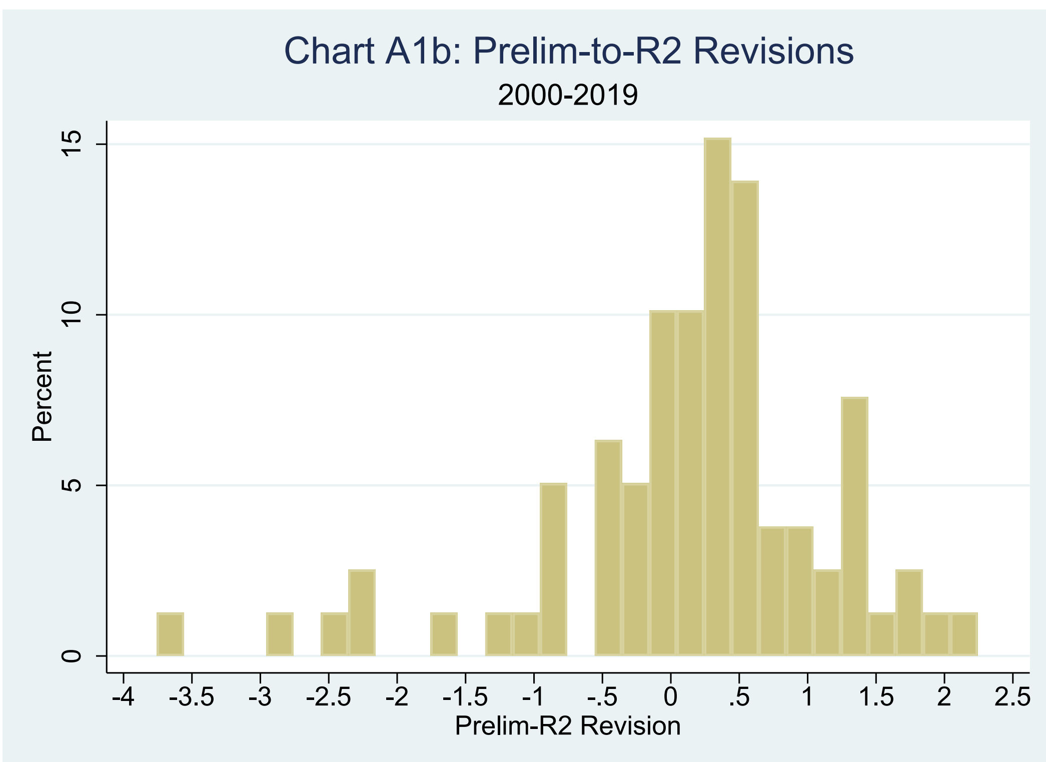


Chart A1b: Prelim-to-R2 Revisions
2000-2019
Percent
0
5
10
15
-4
-3.5
-3
-2.5
-2
-1.5
-1
-.5
0
.5
1
1.5
2
2.5
Prelim-R2 Revision

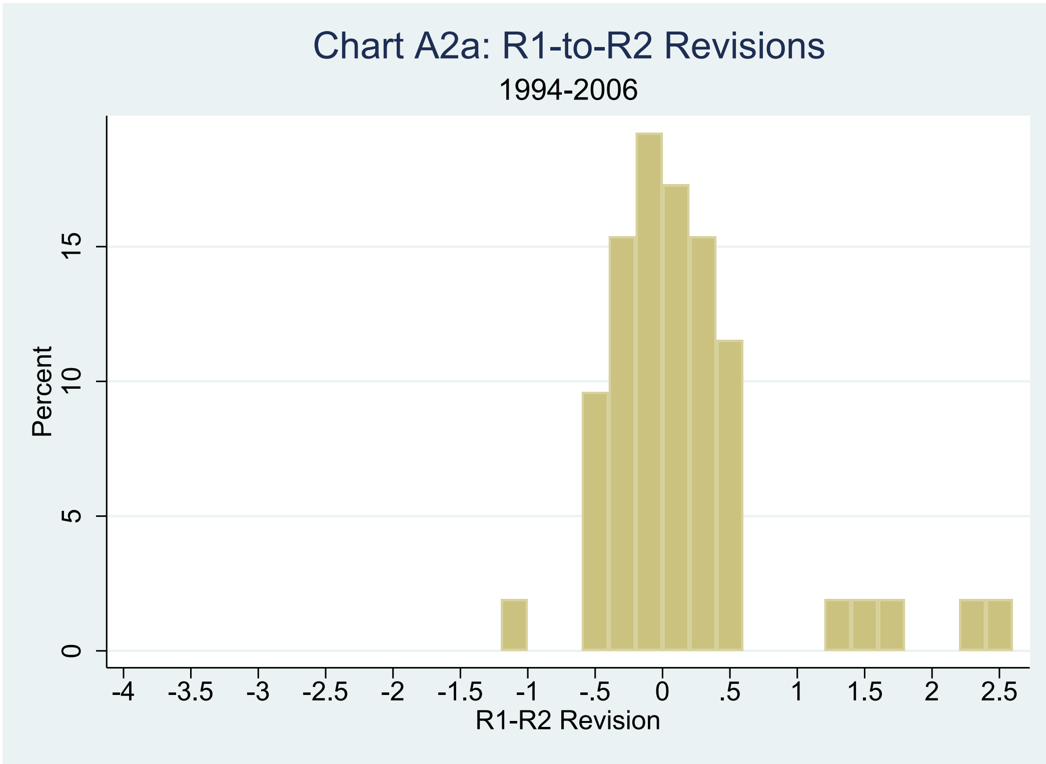


Chart A2a: R1-to-R2 Revisions
1994-2006
Percent
0
5
10
15
-4
-3.5
-3
-2.5
-2
-1.5
-1
-.5
0
.5
1
1.5
2
2.5
R1-R2 Revision

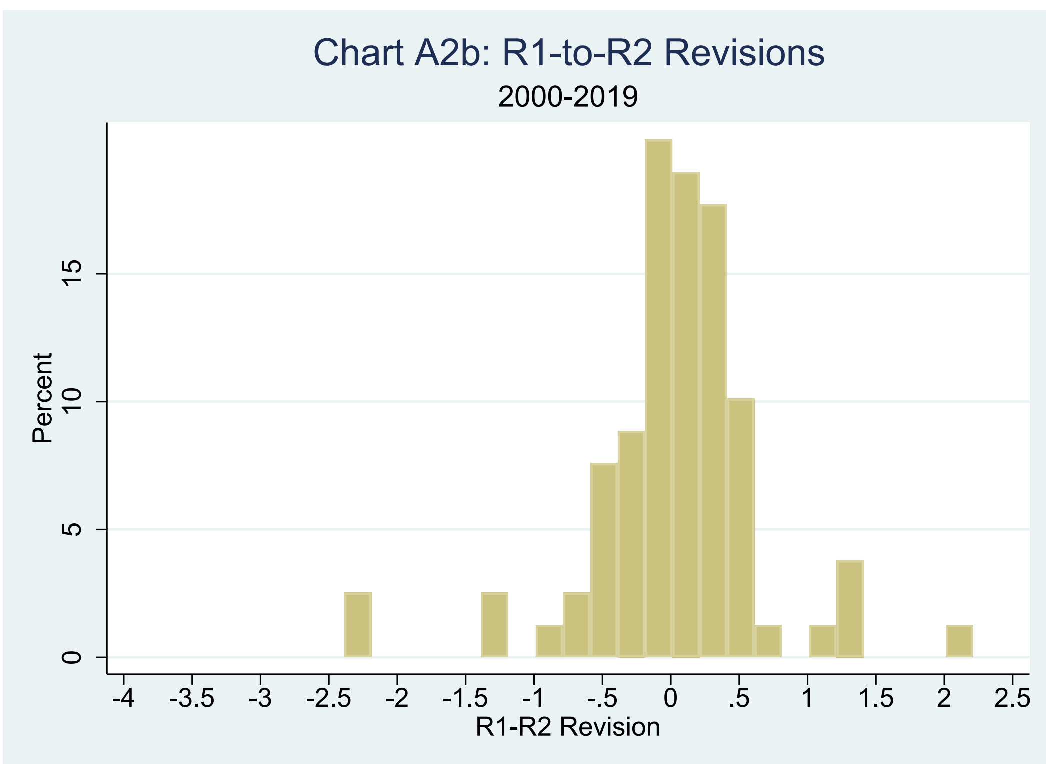


Chart A2b: R1-to-R2 Revisions
2000-2019
Percent
0
5
10
15
-4
-3.5
-3
-2.5
-2
-1.5
-1
-.5
0
.5
1
1.5
2
2.5
R1-R2 Revision

Table A9a: Prelim-to-R2 Revisions by Quarter - 1994-2006

Period	N	Mean	Std. dev.	Min	Max	Mean absolute revision
All Qs	52	0.40	1.04	-1.6	3.3	0.86
Q1	13	0.60	1.24	-0.8	3.3	0.95
Q2	13	0.51	0.88	-0.9	1.9	0.81
Q3	13	0.07	0.89	-1.6	1.5	0.68
Q4	13	0.42	1.15	-1.1	2.2	0.99

Table A9b: Prelim-to-R2 Revisions by Quarter - 2000-2019

Period	N	Mean	Std. dev.	Min	Max	Mean absolute revision
All Qs	79	0.14	1.04	-3.8	2.2	0.76
Q1	20	-0.24	1.09	-2.8	1.1	0.76
Q2	20	0.49	0.67	-0.9	2.2	0.62
Q3	20	0.24	1.10	-2.3	1.7	0.88
Q4	19	0.07	1.18	-3.8	2.0	0.77

Table A10a: R1-to-R2 Revisions by Quarter - 1994-2006

Period	N	Mean	Std. dev.	Min	Max	Mean absolute revision
All Qs	52	0.18	0.64	-1.2	2.4	0.41
Q1	13	0.33	0.95	-1.2	2.4	0.69
Q2	13	0.20	0.43	-0.4	1.4	0.31
Q3	13	-0.06	0.33	-0.5	0.4	0.29
Q4	13	0.23	0.68	-0.3	2.3	0.37

Table A10b: R1-to-R2 Revisions by Quarter - 2000-2019

Period	N	Mean	Std. dev.	Min	Max	Mean absolute revision
All Qs	79	0.04	0.64	-2.4	2.0	0.43
Q1	20	-0.10	1.06	-2.4	2.0	0.72
Q2	20	0.12	0.47	-0.7	1.4	0.37
Q3	20	0.02	0.48	-1.0	1.3	0.35
Q4	19	0.13	0.30	-0.6	0.5	0.26

Table A11a: Decomposition of Preliminary-to-R2 Revisions - 1994 - 2006

	Average Revision to:				
	Output		Hours		
	Current Quarter	Previous Quarter	Current Quarter	Previous Quarter	Total
All Quarters	0.03	-0.30	0.16	0.22	0.39
Q1	-0.45	-0.78	0.09	0.33	0.57
Q2	1.23	0.82	-0.05	0.10	0.56
Q3	-1.09	1.27	0.44	0.32	-2.49
Q4	0.42	0.04	0.16	0.13	0.36

Table A11b: Decomposition of Preliminary-to-R2 Revisions - 2000-2019

	Average Revision to:				
	Output		Hours		
	Current Quarter	Previous Quarter	Current Quarter	Previous Quarter	Total
All Quarters	-0.16	-0.27	-0.06	-0.05	0.13
Q1	-1.29	-1.01	-0.15	-0.12	-0.25
Q2	0.25	-0.13	-0.11	0.00	0.49
Q3	0.38	0.02	0.32	0.18	0.22
Q4	0.04	0.04	-0.32	-0.25	0.07

Note: 2018q4 was deleted due to the government shutdown.

Table A12a: Decomposition of R1-to-R2 Revisions - 1994-2006

	Average Revision to:				
	Output		Hours		
	Current Quarter	Previous Quarter	Current Quarter	Previous Quarter	Total
All Quarters	0.09	0.00	-0.05	0.02	0.15
Q1	-0.75	-0.81	-0.51	-0.30	0.26
Q2	0.97	0.82	-0.03	0.10	0.29
Q3	-0.07	0.02	0.31	0.29	-0.11
Q4	0.21	-0.02	0.03	-0.03	0.17

Table A12b: Decomposition of R1-to-R2 Revisions - 2000-2019

	Average Revision to:				
	Output		Hours		
	Current Quarter	Previous Quarter	Current Quarter	Previous Quarter	Total
All Quarters	-0.29	-0.30	0.00	0.02	0.03
Q1	-1.26	-1.07	-0.23	-0.16	-0.12
Q2	-0.04	-0.13	-0.05	0.00	0.14
Q3	0.04	0.01	0.29	0.26	-0.01
Q4	0.13	0.00	-0.03	-0.04	0.12

Note:2018q4 was deleted due to the government shutdown.