JEREMY J. NALEWAIK
*Board of Governors of the Federal Reserve System*

# *The Income- and Expenditure-Side Estimates of U.S. Output Growth*

**ABSTRACT**    The two official measures of U.S. economic output, gross domestic product (GDP) and gross domestic income (GDI), have shown markedly different business cycle fluctuations over the past 25 years, with GDI showing a more pronounced cycle than GDP. This paper reports a broad range of results that indicate that GDI better reflects the business cycle fluctuations in true output growth. Results on revisions to the estimates, and correlations with numerous other cyclically sensitive variables, are particularly favorable to GDI. The most recent GDI data show the 2007–09 downturn to have been considerably worse than is reflected in GDP.

The Bureau of Economic Analysis (BEA) produces two conceptually identical official measures of U.S. economic output, currently called gross domestic product (GDP) and gross domestic income (GDI). These two measures have shown markedly different business cycle fluctuations over the past 25 years, with GDI showing a more pronounced cycle than GDP. These differences have become particularly glaring over the latest cyclical downturn, which appears considerably worse along several dimensions when measured by GDI. The aim of this paper is to determine which measure better represents the actual business cycle fluctuations in output growth. A wide variety of results suggest the answer is GDI.

Confusion about the information content of the two sets of estimates often starts with the nomenclature. "GDP" can mean either the true output variable of interest or an estimate of that output variable based on the expenditure approach—two very different things. Furthermore, since GDI has a different name than GDP, it may not be initially clear that GDI measures the same concept as GDP using the equally valid income approach. To keep things straight, this paper refers to the true variable of interest as true output, to the expenditure-side estimate of true output as GDP(E), and to the income-side estimate of true output as GDP(I).

The paper presents results for both the initial quarterly output growth estimates (those available three months after the end of the quarter) and the later estimates that have passed through more revisions. After presenting in section I some basic facts about the estimates, I discuss in section II the initial growth rate estimates and present numerous results favoring GDP(I) as the more accurate measure of output growth. First, there is some evidence that initial GDP(I) growth predicts revisions to GDP(E) growth, and no evidence of any tendency for GDP(E) growth to predict revisions to GDP(I) growth. Second, initial GDP(I) growth is the better predictor of a wide variety of business cycle indicators that should be correlated with true output growth. These include all measures of output growth in subsequent periods, the change in the unemployment rate in the current and subsequent periods, employment growth (measured using a household survey) in the current and subsequent periods, the Institute for Supply Management's Purchasing Managers' Index for manufacturing in the current and subsequent periods, changes in stock prices over previous periods, the slope of the Treasury yield curve in previous periods, and forecasts of GDP(E) growth itself from previous periods. Each of these results suggests that GDP(E) growth either is the noisier measure of true output growth or misses fluctuations in true output growth that appear in both GDP(I) growth and the other business cycle indicators. Third, initial GDP(I) growth has identified the onset of each of the last few cyclical downturns more quickly than initial GDP(E).

Section III discusses the latest revised growth rate estimates. I first establish some basic facts about the discrepancies between the fully revised estimates. On average, GDP(I) tends to grow faster than GDP(E) when the economy is expanding robustly, and to lag behind GDP(E) in recessions and in periods when the economy is sluggish. Because of this tendency, the statistical discrepancy between the two output measures is highly negatively correlated with the business cycle. Why is this the case? A thorough analysis of the nature of the source data suggests that GDP(E) misses part of the business cycle and that GDP(I) captures the business cycle better. Statistical analyses reach the same conclusion. First, the nature of the revisions suggests that they add cyclical variation to GDP(I) that is not added to GDP(E), implying that GDP(E) misses some cyclical variation. And second, the latest GDP(I) growth estimates are more highly correlated with a wide range of business cycle indicators, including changes in unemployment, the growth rate of employment, purchasing manager surveys (both manufacturing and nonmanufacturing), changes in stock prices over previous periods, the slope of the Treasury yield curve in previous periods, the spread between high-

yield bonds and Treasury securities from previous periods, and indicator variables for officially identified recessions.

Section IV discusses the behavior of the estimates over the most recent cyclical downturn. When measured by GDP(I), output decelerated sooner, fell at a faster rate at the height of the downturn, and recovered less quickly. Drawing on the results from the previous sections and the online appendices,[1] this section discusses how GDP(E) may have missed the severity of the downturn. Section V concludes with some thoughts about the implications of the results for both data users and the BEA.

## I.  Basic Facts about the Estimates

The BEA's first GDP(E) estimate for the most recent quarter, called the "advance" estimate, is released about a month after the quarter closes. Estimates of most components of GDP(I) for that quarter are included in the advance release, but not all of them; corporate profits and net income from the rest of the world are not released at that time. Those components and GDP(I) are first reported in the second release, about two months after the quarter ends, except for estimates for the fourth quarters, when GDP(I) first appears with the third release, about three months after the quarter ends. To work with a complete time series of the initial growth rates, I focus on these third-release estimates. However, in an online appendix, I repeat the regression results in section II using the second-release estimates for quarters where an estimate of GDP(I) is available and, alternatively, using my own advance GDP(I) estimates constructed using the available income-side components and forecasts of corporate profits and net income from the rest of the world.

Estimates of GDP(E) and GDP(I) growth continue to be revised after the third release. Table 1 shows the variances of the initial (third-release) estimates and the latest estimates, which have passed through more revisions, as well as the correlations between the two estimates; here and throughout the paper ΔGDP(E) and ΔGDP(I) stand for the annualized quarterly growth rates of output implied by the estimates. I focus on two samples here. The first starts in 1978Q3 and is dictated by the start date of the time series of third-release growth rates employed in the paper, which is based on a real-time dataset constructed by the BEA starting in 1978. When analyzing the latest revised estimates, I focus on a shorter sample starting in the

---

1. Online appendices for all papers in this issue may be found on the *Brookings Papers* webpage (www.brookings.edu/economics/bpea), under "Conferences and Papers."

**Table 1.** Correlations between and Variances of Initial and Latest Available Estimates of Growth in GDP(E) and GDP(I)[a]

| Measure | Initial $\Delta GDP(E)$ | Initial $\Delta GDP(I)$ | Latest $\Delta GDP(E)$ | Latest $\Delta GDP(I)$ |
|---|---|---|---|---|
| *Correlations, 1978Q3–2009Q3* | | | | |
| Initial $\Delta GDP(E)$ | 1.00 | | | |
| Initial $\Delta GDP(I)$ | 0.95 | 1.00 | | |
| Latest $\Delta GDP(E)$ | 0.85 | 0.81 | 1.00 | |
| Latest $\Delta GDP(I)$ | 0.77 | 0.82 | 0.79 | 1.00 |
| *Correlations, 1984Q3–2006Q4* | | | | |
| Initial $\Delta GDP(E)$ | 1.00 | | | |
| Initial $\Delta GDP(I)$ | 0.90 | 1.00 | | |
| Latest $\Delta GDP(E)$ | 0.68 | 0.61 | 1.00 | |
| Latest $\Delta GDP(I)$ | 0.63 | 0.66 | 0.60 | 1.00 |
| *Variances* | *1978Q3–2009Q3* | | *1984Q3–2006Q4* | |
| Initial estimates | 8.53 | 8.90 | 3.88 | 3.89 |
| Latest estimates | 9.44 | 10.29 | 4.23 | 4.96 |
| Revisions (difference between latest and initial estimates) | 2.78 | 3.60 | 2.57 | 3.05 |

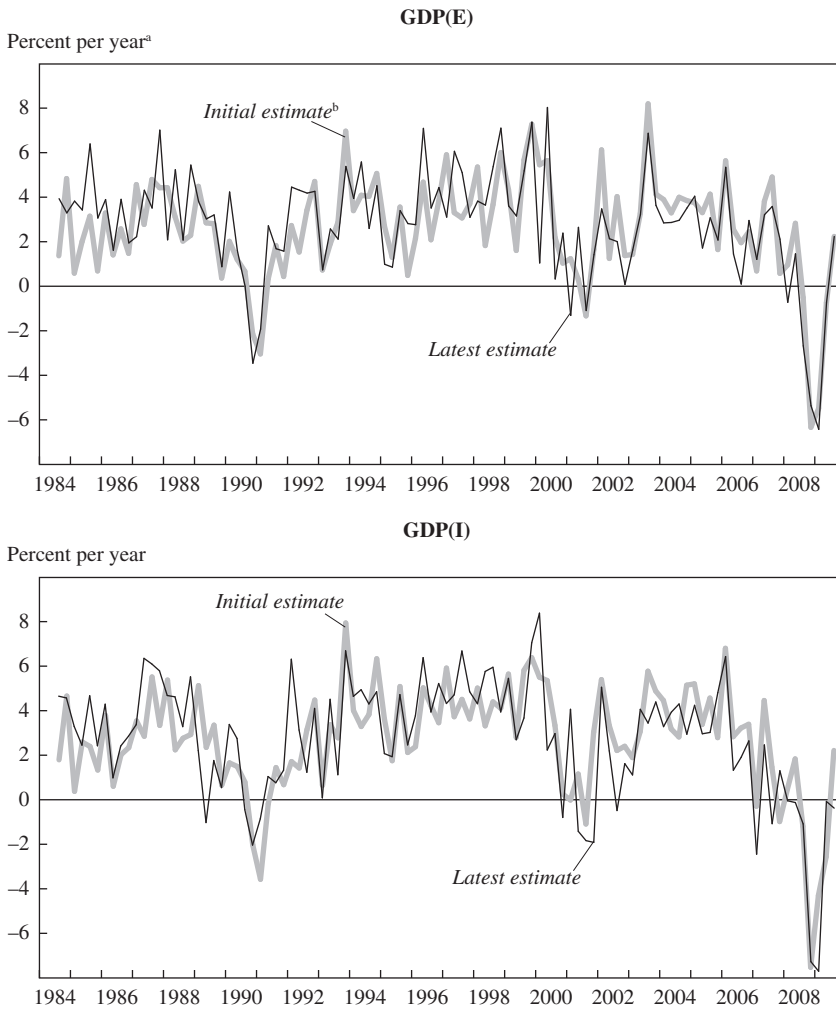Source: Author's calculations using BEA data.
a. "Initial" estimates are those in the third BEA release for each quarter.

mid-1980s, because the divergences between the estimates are particularly stark and highly cyclical over this period.[2] This second sample ends in 2006Q4 to ensure that the latest estimates have been revised to fully incorporate all their major annual source data. Figure 1 plots both $\Delta GDP(E)$ and $\Delta GDP(I)$ from the mid-1980s to the present.

The top two panels of table 1 show that the correlations of the initial estimates of $\Delta GDP(E)$ and $\Delta GDP(I)$ with the latest estimates of the same measures are fairly high: 0.85 and 0.82, respectively, for the first sample, and 0.68 and 0.66 for the second. Nonetheless, the revisions do change the estimates in important ways. First, the bottom panel shows that the variance of the revisions is somewhat larger in both samples for $\Delta GDP(I)$ than for $\Delta GDP(E)$. Moreover, the revisions tend to increase the variance of $\Delta GDP(I)$ more than the variance of $\Delta GDP(E)$. This suggests that the revisions add information to the latest $\Delta GDP(I)$ that is not added to the latest $\Delta GDP(E)$.

---

2. The start date chosen here is the econometric breakpoint marking the beginning of the once widely accepted phenomenon known as the Great Moderation. The precise start date is not particularly important, however; any start date around the mid-1980s gives similar results for the latest estimates.

**Figure 1.** Initial and Latest Available Estimates of Growth in Real GDP(E) and GDP(I), 1984Q3–2009Q3

**GDP(E)**

Percent per year[a]



*Initial estimate*[b]

*Latest estimate*

**GDP(I)**

Percent per year



*Initial estimate*

*Latest estimate*

Source: Author's calculations using BEA data.
a. Quarterly data, annualized.
b. Estimate from the third release for the indicated quarter.

Finally, and perhaps counterintuitively, the revisions tend to make the two measures less similar, reducing their correlation from 0.90 to 0.60 in the shorter sample.

Given the important differences between the latest estimates of GDP(E) and GDP(I), this paper investigates two questions. First, what is the relative information content of the initial estimates of ΔGDP(E) and ΔGDP(I)? Put differently, how much weight should one place on each of these initial estimates? Second, what is the relative information content of the estimated GDP(E) and GDP(I) growth rates after they have passed through all their revisions? In other words, how much weight should one place on each of these latest, revised growth rates?

Online appendix A provides more background information about GDP(E) and GDP(I). Appendix B discusses the source data used to construct the initial growth rates, and appendix C describes the source data incorporated at the annual and the benchmark revisions.

## II. The Information Content of the Initially Estimated Growth Rates

A detailed examination of the source data used to compute the initial (third-release) growth rates of GDP(E) and GDP(I) shows that both estimates suffer from similar types of measurement error problems (see online appendix B). These problems include missing data for a substantial portion of each estimate, sampling errors, and nonsampling errors such as incomplete coverage, survey nonresponse, and incomplete corrections for firm births and deaths. A compelling case for the superiority of either estimate cannot be made from such a detailed examination of source data, so this section proceeds immediately to the more informative statistical tests.[3]

---

3. In his comment on this paper, Steven Landefeld suggests that for the third-release estimates, a much greater fraction of GDP(I) than of GDP(E) is based on judgmental trends instead of early source data. Almost all of the source data used to compute the third-release estimates are flawed and unrepresentative in some way, and breaking down the data using such a binary classification scheme is a highly subjective exercise. The detailed discussion of the source data in online appendix B suggests that the evidence is less favorable to GDP(E) than this classification scheme suggests. Moreover, if a much greater fraction of GDP(I) were based on trends, one should expect third-release ΔGDP(I) to be much less variable than third-release ΔGDP(E), because trends should have less variance than the actual source data. The summary statistics in table 1 show that this is not the case: third-release ΔGDP(I) is actually slightly more variable than third-release ΔGDP(E).

   Table 2 reports the main regression results examining the information content of the initial growth rate estimates. A good place to start is by examining the power of the initial estimates to predict the latest estimates, which incorporate superior source data. Over the full sample, the initial ΔGDP(E) estimates predict well the latest estimates of ΔGDP(E), with initial ΔGDP(I) adding little after conditioning on initial ΔGDP(E). Similarly, initial ΔGDP(I) well predicts latest ΔGDP(I), with initial ΔGDP(E) adding little information after conditioning on initial ΔGDP(I). However, the final two sets of regressions in table 2 show results for a sample starting in 1994Q1; I stop this subsample in 2006Q4 to ensure that the latest estimates have passed through all their annual revisions, but extending the subsample to 2008Q4 produces similar results. The first specification constrains the coefficients on initial ΔGDP(E) and initial ΔGDP(I) to sum to 1, whereas the second does not; the results show that over this sample period, when initial ΔGDP(I) is 1 percentage point above initial ΔGDP(E), initial ΔGDP(E) has subsequently been revised upward about a third (0.28) to two-fifths (0.42) of a percentage point, on average. Dennis Fixler and Bruce Grimm (2006), using a broader set of conditioning variables, also find some tendency for initial ΔGDP(E) to be revised toward initial ΔGDP(I). The last set of results shows that, over this subsample, there remains no significant tendency for initial ΔGDP(E) to predict latest ΔGDP(I).

   Initial ΔGDP(I) may have predicted revisions to ΔGDP(E) in this sample because the initial ΔGDP(I) estimates are less noisy than the initial ΔGDP(E) estimates, or because they contain information about true output growth that is missed by the initial ΔGDP(E) estimates but is incorporated into latest ΔGDP(E) through revisions. Each of these explanations is likely part of the story. Averaging the data into year-over-year growth rates eliminates much of the noise in the quarterly data and shows the plausibility of the second explanation. The top panel of figure 2 plots the revisions in the year-over-year (fourth quarter to fourth quarter) growth rate of GDP(E) against the gap between GDP(I) and GDP(E) in the initial estimated year-over-year growth rates.[4] Broadly speaking, two periods drive the positive relationship that emerges.[5]

---

   4. This presentation was suggested to me by William Wascher.
   5. The line plots the predicted values from regressing the 13 Q4-over-Q4 growth rates of real GDP(E) on a constant and the gap between the initial estimates of Q4-over-Q4 GDP(I) and GDP(E) growth. The coefficient on the gap is 0.98, with a standard error of 0.44 and an adjusted $R^2$ of 0.25. I also experimented with corrections that removed the effects of major methodological changes from the revisions; this modification increased the $R^2$.

**Table 2.** Regressions Testing the Predictive Content of Initial Estimates of Growth in GDP(E) and GDP(I), 1978Q3–2009Q3 Sample[a]

| Dependent variable | Initial ΔGDP(E) | | | Initial ΔGDP(I) | | | Constant | Adjusted $R^2$ |
|---|---|---|---|---|---|---|---|---|
| | $t$ | $t-1$ | $t-2$ | $t$ | $t-1$ | $t-2$ | | |
| ΔGDP(E)$_t$, latest estimate | 0.77 | | | 0.12 | | | 0.40 | 0.71 |
| | (0.13) | | | (0.12) | | | (0.24) | |
| | 0.76 | −0.17 | | 0.07 | 0.27 | | 0.28 | 0.72 |
| | (0.12) | (0.15) | | (0.11) | (0.15) | | (0.21) | |
| | | −0.46 | | | 0.93 | | 1.42 | 0.25 |
| | | (0.28) | | | (0.29) | | (0.32) | |
| ΔGDP(E)$_t$, initial estimate | | −0.37 | | | 0.80 | | 1.38 | 0.22 |
| | | (0.26) | | | (0.25) | | (0.28) | |
| | | | −0.33 | | | 0.54 | 1.97 | 0.05 |
| | | | (0.22) | | | (0.26) | (0.33) | |
| ΔGDP(I)$_t$, latest estimate | 0.02 | −0.02 | | 0.86 | | | 0.34 | 0.66 |
| | (0.18) | (0.18) | | (0.18) | | | (0.25) | |
| | 0.03 | −0.17 | | 0.81 | 0.10 | | 0.24 | 0.66 |
| | (0.19) | (0.28) | | (0.20) | (0.17) | | (0.29) | |
| | | | | | 0.66 | | 1.34 | 0.22 |
| | | | | | (0.28) | | (0.45) | |
| ΔGDP(I)$_t$, initial estimate | | −0.17 | | | 0.67 | | 1.31 | 0.26 |
| | | (0.21) | | | (0.21) | | (0.32) | |
| | | | −0.43 | | | 0.69 | 1.91 | 0.09 |
| | | | (0.21) | | | (0.24) | (0.40) | |
| $(UR_t - UR_{t-1}) \times 4$ | 0.03 | | | −0.36 | | | 1.01 | 0.57 |
| | (0.08) | | | (0.08) | | | (0.18) | |
| | | 0.02 | | | −0.30 | | 0.87 | 0.40 |
| | | (0.07) | | | (0.08) | | (0.22) | |
| | | | 0.05 | | | −0.26 | 0.69 | 0.22 |
| | | | (0.10) | | | (0.12) | (0.24) | |

| Regressor | (1) | (2) | (3) | (4) | (5) | (6) | (7) | $R^2$ |
|---|---|---|---|---|---|---|---|---|
| $(E^{household}_t / E^{household}_{t-1})^4$ | 0.08 (0.12) | −0.03 (0.10) | −0.19 (0.13) | 0.35 (0.11) | 0.41 (0.13) | 0.43 (0.14) | 0.06 (0.27) | 0.47 |
| $ISM^{manuf.}_t$ | 0.15 (0.41) | 0.41 (0.38) | −0.03 (0.46) | 1.33 (0.41) | 0.93 (0.37) | 0.83 (0.44) | 0.17 (0.33) | 0.38 |
|  |  |  |  |  |  |  | 0.55 (0.39) | 0.16 |
|  |  |  |  |  |  |  | 47.23 (0.68) | 0.53 |
|  |  |  |  |  |  |  | 47.59 (0.66) | 0.43 |
|  |  |  |  |  |  |  | 49.00 (0.78) | 0.15 |
| $Log(S\&P500_t - S\&P500_{(t-4)})/4$ | −0.09 (0.30) |  |  | 0.73 (0.33) |  |  | 0.08 (1.08) | 0.19 |
| $r^{Treas(10)}_{t-8} - r^{Treas(2)}_{t-8}$ | −0.14 (0.07) |  |  | 0.21 (0.08) |  |  | 0.51 (0.15) | 0.08 |
| SPF forecast, $\Delta GDP(E)_{t,t}$ | 0.06 (0.11) |  |  | 0.43 (0.12) |  |  | 0.74 (0.28) | 0.51 |
| SPF forecast, $\Delta GDP(E)_{t,t-1}$ | 0.03 (0.07) |  |  | 0.22 (0.09) |  |  | 1.68 (0.30) | 0.19 |
| SPF forecast, $\Delta GDP(E)_{t,t-2}$ | 0.08 (0.07) |  |  | 0.08 (0.08) |  |  | 2.20 (0.33) | 0.11 |
|  |  |  |  |  |  |  | −0.29 (0.20) |  |
| $\Delta GDP(E)_t$, latest, | 0.58[b] (0.16) |  |  | 0.42 (0.16) |  |  | 0.31 (0.35) | 0.47 |
| 1994Q1–2006Q4 | 0.56 (0.16) |  |  | 0.28 (0.16) |  |  |  |  |
| $\Delta GDP(I)_t$, latest, | 0.15[b] (0.24) |  |  | 0.85 (0.24) |  |  | −0.16 (0.26) | 0.41 |
| 1994Q1–2006Q4 | 0.13 (0.23) |  |  | 0.74 (0.22) |  |  | 0.31 (0.59) |  |

Source: Author's regressions.

a. Each pair of rows reports results of a single regression. "Initial" estimates are those in the third BEA release for each quarter. Regressions are run on the full sample period unless stated otherwise. Numbers in parentheses are Newey-West standard errors using eight lags.

b. Regression coefficients are constrained to sum to 1.

**Figure 2.** Revisions to Output Growth Measures and the Statistical Discrepancy

**GDP(E)**

Revision to ΔGDP(E) (percentage points)[a]



Statistical discrepancy in initial estimated growth rates[b]
(percentage points)

**GDP(I)**

Revision to ΔGDP(I) (percentage points)[a]



Statistical discrepancy in initial estimated growth rates[c]
(percentage points)
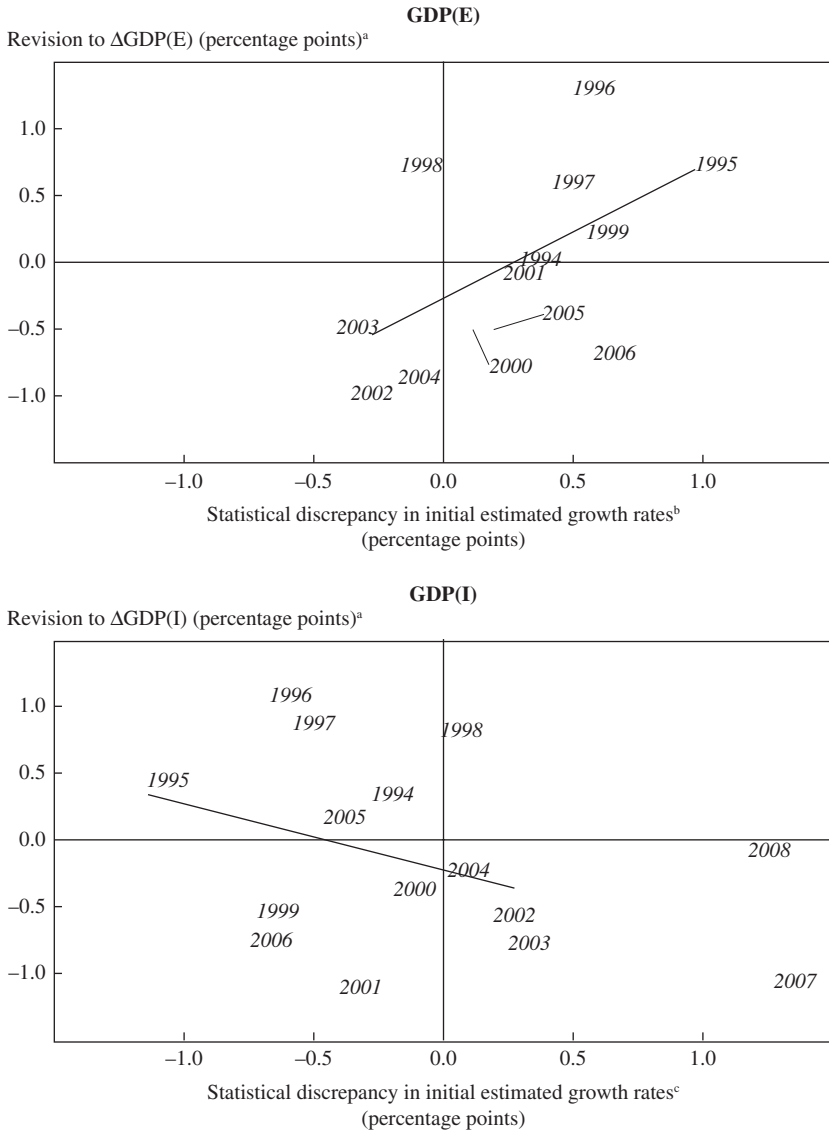
Source: Author's calculations using BEA data.
a. Difference between initial and latest available, fourth quarter to fourth quarter.
b. Difference between initial ΔGDP(I) and initial ΔGDP(E), fourth quarter to fourth quarter.
c. Difference between initial ΔGDP(E) and initial ΔGDP(I), fourth quarter to fourth quarter.

First, during the mid- to late 1990s, the gap between initial ΔGDP(I) and initial ΔGDP(E) was consistently positive, as the initial estimates showed GDP(I) growing faster than GDP(E). This phenomenon was discussed in real time by the Council of Economic Advisers (*Economic Report of the President 1997,* pp. 72–74), the Federal Reserve Board (Greenspan 2004), and the BEA itself (Moulton 2000), with conclusions generally favorable to GDP(I). Those conclusions were later vindicated, since ultimately ΔGDP(E) was revised upward toward initial ΔGDP(I). Initial ΔGDP(I) accurately captured information about the brisk pace of economic growth that the initial ΔGDP(E) estimates had missed and that was incorporated only later through revisions (and then probably only partially; see section III). Second, the initial estimates of ΔGDP(I) in 2002 and 2003, after the 2001 recession, showed a more sluggish recovery than the initial estimates of ΔGDP(E), so that the gap between the initial estimates was negative.[6] Ultimately, ΔGDP(E) was revised toward initial ΔGDP(I) again: the recovery was indeed quite sluggish, and this information was reflected in ΔGDP(I) before it appeared in ΔGDP(E). The bottom panel of figure 2, which plots revisions in the fourth quarter–to–fourth quarter growth rate of GDP(I) against the gap between GDP(I) and GDP(E) in the initial estimated year-over-year growth rates, shows no tendency for ΔGDP(I) to be revised toward ΔGDP(E); if anything, ΔGDP(I) tends to be revised in the opposite direction from the initial gap over this period.

This particular set of revision results uses a short sample and should therefore be taken with a grain of salt. As a robustness check, I used past issues of the *Survey of Current Business* to extend the sample back in time to 1966Q4; the results, reported in online appendix D, show a marginally statistically significant tendency for initial ΔGDP(E) to be revised toward initial ΔGDP(I) over this long sample. After the data have passed through their first annual revision, a statistically significant tendency for initial ΔGDP(E) to be revised toward initial ΔGDP(I) in subsequent revisions appears yet again, in regressions using the 1978Q3–2009Q3 sample. So it would probably be unwise to ignore these revision results entirely. However, if for some reason an analyst did decide to ignore the predictability of the revisions, then the weight that analyst should place on the initial

6. As online appendix C outlines, since 2002 the BEA has incorporated information from the Quarterly Census of Employment and Wages (QCEW) into its wage and salary estimates a couple of months after its third release. These QCEW revisions provided much of the information contained in ΔGDP(I) on the relative sluggishness of the recovery; see the discussion in Nalewaik (2007a). The year-over-year growth rates for fourth quarters available in real time reflect these QCEW revisions.

estimates is entirely determined by the weight he or she places on the latest, fully revised estimates. An analyst who believes that latest $\Delta GDP(E)$ is more accurate than latest $\Delta GDP(I)$ should believe that initial $\Delta GDP(E)$ is more accurate than initial $\Delta GDP(I)$, and vice versa. So the results outlined in the next section, addressing the paper's second question, are also critical to answering the first.

However, one can make considerable further progress on the paper's first question directly, by examining the predictive power of the initial estimates for other important cyclical indicators. Broadly speaking, these regressions help establish which estimate is more informative about the business cycle, but they also help answer the narrower question of which is the better estimate of true output growth. The inferior estimate of true output growth, containing relatively more noise or classic measurement error, should have a lower signal-to-noise ratio and be the inferior predictor of cyclical indicators correlated with true output growth, all else equal. (This assumes that the noise in the output growth estimates is uncorrelated with the measurement error in the other cyclical indicators, and I have chosen the other cyclical indicators carefully to avoid this problem.) An estimate may be inferior not only because it is noisier, but also because it misses more fluctuations in true output growth (that is, contains less news or signal about true output growth) than the other estimate. But, again, such an inferior estimate should have a lower signal-to-noise ratio and be the inferior predictor of cyclical indicators that reflect those missing fluctuations in true output growth.

Returning to table 2, the top panel shows that as a cyclical indicator of where output growth is headed, initial $\Delta GDP(I)$ is superior to initial $\Delta GDP(E)$. The initial estimates of $\Delta GDP(I)$ are positively related to output growth in the next quarter, whether the latter is measured by $\Delta GDP(E)$ or $\Delta GDP(I)$, initial or latest. Conditional on initial $\Delta GDP(I)$, initial $\Delta GDP(E)$ contains no information about output growth next quarter and may actually be negatively related to it. This result holds two quarters ahead as well, using either initial estimate of output growth. Following the logic outlined above, these results may obtain because initial $\Delta GDP(E)$ is noisier than initial $\Delta GDP(I)$, so that its signal about true output growth in subsequent periods is obscured, or because initial $\Delta GDP(E)$, but not initial $\Delta GDP(I)$, misses some of the shocks that produce serially correlated fluctuations in true output growth.

I examine next the relationship of the initial estimates to other cyclical variables that should be correlated with true output growth. To avoid correlated measurement errors and spurious correlation, I examine only vari-

ables that are not used in the construction of either GDP(E) or GDP(I). As outlined in the appendices, the GDP(E) and GDP(I) estimates make little use of the Current Population Survey (CPS), a monthly household survey used to produce estimates of the unemployment rate.[7] As one of the most important indicators of the business cycle, the unemployment rate is a good variable to use as a starting point for this analysis.

Table 2 shows that initial ΔGDP(I) has a strong negative relationship with the contemporaneous change in the unemployment rate and negatively predicts changes to the unemployment rate one and two quarters ahead, whereas the coefficients on initial ΔGDP(E) are insignificant and have the wrong sign when conditioning on initial ΔGDP(I). Again, this may be because initial ΔGDP(E) is noisier than initial ΔGDP(I), or because initial ΔGDP(E) misses fluctuations in true output that both appear in ΔGDP(I) and are reflected in the differenced unemployment rate.

The next set of regressions in table 2 shows results using quarterly annualized employment growth computed from the household survey data, adjusted for breaks introduced by Census updates to the population. Initial ΔGDP(I) is positively related to employment growth this quarter, as well as one and two quarters ahead, but initial ΔGDP(E) contains little additional information about employment growth beyond that contained in initial ΔGDP(I).

Broadening the results beyond labor market variables, the next set of regressions uses the PMI (formerly called the Purchasing Managers' Index) from the Institute for Supply Management (ISM) manufacturing survey. The ISM measure is computed quite differently from GDP(E) and GDP(I). It is an aggregation of several diffusion indexes, so that even though the companies participating in the ISM survey also participate in the surveys used to estimate GDP(E) and GDP(I), the measurement errors likely behave quite differently. Initial ΔGDP(I) explains the contemporaneous, one-quarter-ahead, and two-quarters-ahead movements in the ISM measure better than initial ΔGDP(E), which in fact provides no statistically significant information conditional on initial ΔGDP(I).

---

7. At first blush, some analysts might suspect that GDP(I) must be more correlated with the unemployment rate than GDP(E), because "income" is in the name and the unemployment rate is a labor market concept. However, this reasoning is incorrect. Of the various components of the two output measures, one may expect based on a priori considerations that compensation will have a higher-than-average correlation with unemployment, but the other components of GDP(I) should then have a lower-than-average correlation, since all the components of GDP(I) add up to the same conceptual measure of output as GDP(E). For example, stories in the press recently have suggested that some of the recent rebound in corporate profits was facilitated by weakness in the labor market, allowing firms to cut compensation costs.

Business cycle analysts use a host of other variables to predict ΔGDP(E), most notably different asset prices, and since these asset prices are not used in constructing the output growth estimates, they are prime candidate variables for testing the information content of the initial estimates. However, asset prices typically predict output growth in subsequent quarters, rather than being predicted by output growth, so to get the timing correct, I regress lagged values of these asset prices on the two initial output growth measures. This is a somewhat odd specification but still quite instructive. The results essentially reveal which initial estimate is more consistent with market expectations of the business cycle from earlier periods.

The first asset-price specification regresses the logarithm of the change in the S&P 500 stock price index from the end of quarter $t - 4$ to the end of quarter $t$ on the two initial output growth measures in quarter $t$. Initial ΔGDP(I) is strongly positively related to this measure of stock price changes, whereas the coefficient on initial ΔGDP(E) is insignificant and negative. The next specification examines the slope of the yield curve, measured as the difference in yields between 10- and 2-year Treasury notes. This variable is most closely related to the output growth measures about 2 years hence; a regression of this measure from quarter $t - 8$ on the two initial output growth measures in quarter $t$ yields a coefficient on initial ΔGDP(I) that is significant and has the correct (positive) sign, and a coefficient on initial ΔGDP(E) that is significant but has the wrong sign.

The final set of testing variables I employ consists of median forecasts of output growth from the Survey of Professional Forecasters (SPF). These forecasters are trying to predict initial ΔGDP(E), presumably inclusive of any measurement errors. However, if the forecasters do not yet have access to the source data used to compute ΔGDP(E) for the quarter they are trying to predict, their forecasts will likely reflect general information about the state of the economy, which may be better related to initial ΔGDP(I) than to initial ΔGDP(E). This may be the case even for the current-quarter forecasts, because the survey occurs relatively early in the quarter, before the analysts have much GDP(E) source data. The results show that those current-quarter forecasts are well explained by initial ΔGDP(I), with initial ΔGDP(E) providing no incremental explanatory power. The SPF forecasts for quarter $t$, made in the first half of quarter $t - 1$, are also better explained by initial ΔGDP(I) in period $t$ than by initial ΔGDP(E) in period $t$. Forecasters' expectations for how the economy will move in the current quarter and the next appear to play out more fully in the initial ΔGDP(I) estimates than in initial ΔGDP(E).

Given the tighter relationship of initial ΔGDP(I) to all these business cycle indicators, placing full weight on the initial ΔGDP(E) estimates and no weight on the initial ΔGDP(I) estimates necessarily implies several things: first, that one cares about true output growth in the current quarter only (and not about the business cycle more broadly or even where true output growth is headed next quarter); second, that one believes that the latest ΔGDP(E) estimates reflect all available information about true output growth, so that neither latest ΔGDP(I) nor any other variable provides any additional marginal information about true output growth; third, that one believes that the superior explanatory power of initial ΔGDP(I) for various other cyclical indicators reveals nothing about the relative accuracy of initial ΔGDP(I) and initial ΔGDP(E) as estimates of true output growth; and fourth, that one discounts entirely the evidence contained in revisions. Regarding the first point, this may be a reasonable position for the BEA to take, but for analysts it is less clear: true output growth may be the only variable of interest for some purposes, but not for others. The dismissal of initial ΔGDP(I)'s ability to predict other indicators could be justified only if initial ΔGDP(I) contains variation uncorrelated to true output growth but correlated with all the other dependent variables employed in table 2, including actual forecasts of output growth. This is clearly an extreme position. A much more plausible explanation is that initial ΔGDP(I) is more highly correlated with true output growth than initial ΔGDP(E), and that true output growth is correlated with all these other cyclical indicators. Similarly, the view that latest ΔGDP(E) reflects all available information about true output growth is quite an extreme position in favor of the accuracy of latest ΔGDP(E), and the results in the next section suggest that latest ΔGDP(I) does contain a considerable amount of information about true output growth that latest ΔGDP(E) misses.

The regression results in table 2 are broadly consistent with those in Nalewaik (2007a), where I use Markov switching models to show that ΔGDP(I) identifies cyclical turning points more quickly than ΔGDP(E) in real time. Specifically, at the start of the 1980, 1981–82, 1990–91, and 2001 recessions (as defined by the Business Cycle Dating Committee of the National Bureau of Economic Research, NBER), real-time estimates of a Markov switching model using ΔGDP(E) alone put the probability that the economy was in a low-growth state at 52 percent, 40 percent, 45 percent, and 23 percent, respectively. Substituting ΔGDP(I) in the model produced much more accurate probabilities: 78 percent, 44 percent, 72 percent, and 70 percent. Most of the research in Nalewaik (2007a) was carried out in 2005, and the subsequent cyclical downturn was the first out-of-sample

test of the main hypotheses of the paper. The model using $\Delta$GDP(I) again performed much better around the start of the downturn in real time; it also performed better than some popular models using monthly indicators (see section IV).

Although this section has focused on growth rates derived from the BEA's third releases, the information content of growth rates calculated from the two preceding releases is of critical importance for analysis in real time. Online appendix D reports results for these vintages, as well as results for the estimates once they have passed through their first annual revision. Briefly, when an official second-release estimate of $\Delta$GDP(I) is available, the results using second-release growth rates are very similar to those reported in this section using the third-release growth rates. And as discussed above, the results using growth rates based on the first annual revision are even more favorable to $\Delta$GDP(I) than the results using third-release growth rates, showing a statistically significant tendency for $\Delta$GDP(E) to be revised toward $\Delta$GDP(I) over the full sample.

For the advance estimates, when an official $\Delta$GDP(I) estimate is not available, the situation is quite different. I use profits data from previous quarters and NIPA components that are reported in the advance release to generate the profits forecast used in constructing the "advance $\Delta$GDP(I)" examined in the online appendix. However, some companies have already reported their quarterly profits for the latest complete quarter at the time of the advance release, and incorporating this information may produce a much-improved "advance $\Delta$GDP(I)" estimate. That said, these rather limited advance-release $\Delta$GDP(I) estimates perform poorly compared with the official advance-release $\Delta$GDP(E) estimates, which better predict most of the business cycle variables used in this section. In addition, when predicting latest $\Delta$GDP(I), about a two-thirds weight should be placed on the advance-release $\Delta$GDP(E), and only about a one-third weight on the constructed advance-release $\Delta$GDP(I) estimates. This suggests that the initial estimates of corporate profits produced by the BEA are highly informative and cannot be easily predicted based on lagged profits or other available NIPA variables. For fourth-quarter second-release estimates, when official profits numbers remain unavailable, this is presumably the case as well.

## III.  Information Content of the Latest Growth Rate Estimates

This section begins by showing that the latest estimates of GDP(I) and GDP(E) exhibit markedly different cyclical properties since the mid-1980s. I then discuss the evidence from revisions suggesting the superiority of the

**Figure 3.** Year-over-Year Growth Rates of GDP(E) and GDP(I), 1985Q1–2009Q3[a]

Percent
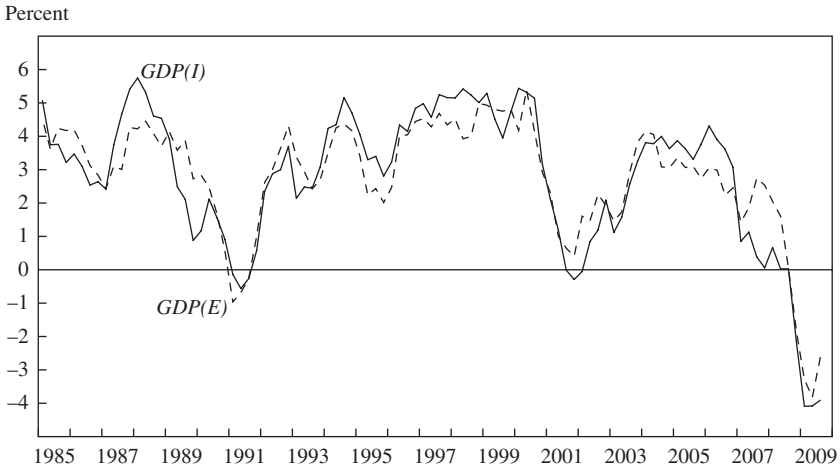


Source: Author's calculations using BEA data.
a. Calculations use latest available estimates of GDP(E) and GDP(I).

cyclical information in latest GDP(I), and I examine the relationship of the latest estimates to other cyclically sensitive variables.

### III.A. The Cyclicality of the Latest Estimates

Table 1 showed that the correlation between ΔGDP(E) and ΔGDP(I) dropped sometime around the mid-1980s, and the divergences between the estimates became highly cyclical around that time. Figure 3 shows this using year-over-year growth rates: GDP(I) rose faster than GDP(E) through most of the 1990s boom and the comparatively short boom period from 2004 to 2006; in contrast, GDP(I) growth fell below GDP(E) growth in the 2001 recession and in the latest cyclical downturn. (Of course, these data are subject to further annual and benchmark revisions.) Figure 4 plots the statistical discrepancy between GDP(E) minus GDP(I), as a percent of GDP(E), and the unemployment rate; work by Charles Fleischman first examined this relationship, to my knowledge. Fleischman and John Roberts (2010) have studied the relationships among GDP(E), GDP(I), the unemployment rate, and other variables in the context of a state space model of the business cycle. Their work points to the unemployment rate as an excellent measure of the state of the business cycle; it also suggests that GDP(E) is measured with more error than GDP(I). Figure 4 shows that the measurement errors in either GDP(I) or GDP(E) are clearly systematically

**Figure 4.** Statistical Discrepancy between GDP(E) and GDP(I) and the Unemployment Rate, 1984Q1–2009Q3[a]
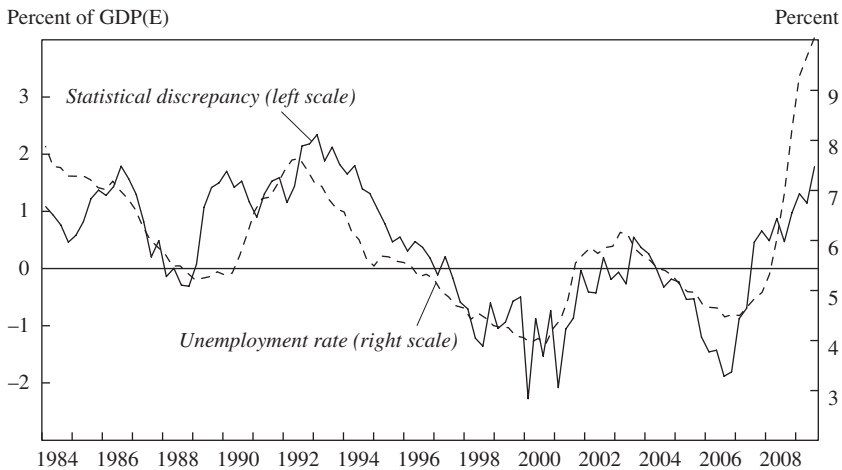


Source: Author's calculations using BEA and Bureau of Labor Statistics data.
a. Estimates use latest available data as of February 2010.

related to the business cycle, and the statistical discrepancy is not noise, as is commonly assumed.

To better understand the relationship depicted in figure 4, consider a very simple model. In Nalewaik (2008) I showed why the type of model outlined below is an incomplete characterization of the growth rates of GDP(E) or GDP(I), and I proposed models that fit the evidence better. However, the model is useful for the limited purpose of framing the subsequent discussion. Let true output be $Y_t^*$, and assume that it can be decomposed into a trend $\tau_t$ and a cycle $\psi_t$, so that $Y_t^* = \tau_t + \psi_t$. The unemployment rate $U_t$ is governed by an Okun's law relationship:

$$U_t - U_t^n = \gamma\left(Y_t^* - \tau_t\right) = \gamma\psi_t, \gamma < 0.$$

Now assume that GDP(I) and GDP(E) are systematically either too cyclical or not cyclical enough, so that:

$$GDP(E)_t = \tau_t + \alpha_E\psi_t$$
$$GDP(I)_t = \tau_t + \alpha_I\psi_t.$$

Then the statistical discrepancy $SD_t = $ GDP(E)$_t - $ GDP(I)$_t = (\alpha_E - \alpha_I)\psi_t$, and assuming that the systematic mismeasurement is not identical for the

two estimates, one should observe a relationship between the discrepancy and the unemployment rate:

$$U_t - U_t^n = \frac{\gamma}{\alpha_E - \alpha_I}(SD_t), \gamma < 0.$$

The strong positive relationship shown in figure 4 then implies that $\alpha_E < \alpha_I$, that is, that the magnitude of the cycle is smaller in GDP(E) than in GDP(I). Table 3 reports regressions investigating this relationship. The first three regressions show that the unemployment rate captures more than 60 percent of the variability of the statistical discrepancy from 1984Q3 through 2006Q4, and although the discrepancy is highly autocorrelated, the unemployment rate remains significant when an AR1 term is added. The last three columns of table 3 show specifications in first differences, to isolate the higher-frequency variation in the data. The first difference exhibits some negative autocorrelation, but the coefficient on the differenced unemployment rate remains positive, and the relationship is highly significant when the differenced unemployment rate is lagged one quarter. These regression results confirm that the statistical discrepancy is not noise, even in differences.

Having established that $\alpha_E < \alpha_I$ within the context of this very stylized model, three possibilities can be considered:

—Both GDP(I) and GDP(E) are more cyclical than true output, so that $\alpha_I > \alpha_E > 1$. In this case GDP(E) represents the cycle in true output better than GDP(I).

—Both GDP(I) and GDP(E) are less cyclical than true output, so that $\alpha_E < \alpha_I < 1$. In this case GDP(I) represents the cycle in true output better than GDP(E).

—GDP(E) is less cyclical and GDP(I) is more cyclical than true output, so that $\alpha_E < 1 < \alpha_I$. In this case GDP(I) represents the cycle in true output better than GDP(E) if $\alpha_I - 1 < 1 - \alpha_E$.

These possibilities frame the detailed discussion of the source data incorporated into the latest estimates in online appendix C. Plenty of evidence suggests that GDP(E) misses part of the business cycle, implying that the first possibility is unlikely. Some of the construction components of GDP(E) are smoothed; in particular, the additions and alterations ("adds and alts") component of residential structures is smoothed using a 3-year moving average. This is problematic, because smoothed estimates inherently understate the magnitude of business cycle accelerations and decelerations. Although "adds and alts" is a small component of GDP(E), it may have taken on outsized importance in the late-2000s downturn and

**Table 3. Regressions Explaining the Statistical Discrepancy between GDP(E) and GDP(I)**[a]

| | Levels specifications[b] | | | | | | Differences specifications[c] | | |
| | 1984Q3–2006Q4 | | | 1948Q1–1984Q2 | | | | | |
| Independent variable | 3-1 | 3-2 | 3-3 | 3-4 | 3-5 | 3-6 | 3-7 | 3-8 | 3-9 |
|---|---|---|---|---|---|---|---|---|---|
| Statistical discrepancy (SD) lagged one quarter $SD_{t-1}$ | 0.93 (0.05) | | 0.75 (0.12) | 0.74 (0.06) | | 0.72 (0.06) | | | |
| Change in SD lagged one quarter $\Delta SD_{t-1}$ | | | | | | | −0.33 (0.17) | −0.35 (0.18) | −0.36 (0.16) |
| Unemployment rate $U$ | | 0.88 (0.11) | 0.25 (0.11) | | 0.10 (0.05) | 0.03 (0.02) | | | |
| Current change in unemployment rate $\Delta U_t$ | | | | | | | | 0.28 (0.23) | |
| Change in unemployment rate lagged one quarter $\Delta U_{t-1}$ | | | | | | | | | 0.84 (0.25) |
| Constant $\alpha$ | −0.01 (0.05) | −4.69 (0.64) | −1.36 (0.62) | 0.15 (0.05) | 0.07 (0.28) | −0.02 (0.12) | −0.04 (0.05) | −0.03 (0.04) | −0.01 (0.04) |
| Adjusted $R^2$ | 0.83 | 0.63 | 0.84 | 0.55 | 0.06 | 0.55 | 0.10 | 0.10 | 0.22 |

Source: Author's regressions.

a. Each column represents a different regression specification. In the levels specifications, the dependent variable is the statistical discrepancy, the difference between GDP(E) and GDP(I), in the current quarter; in the differences specifications, it is the change in the statistical discrepancy from the previous quarter. Newey-West standard errors with eight lags are in parentheses.

b. The regression equation is $SD_t = \alpha + \beta SD_{t-1} + \theta U_t + \varepsilon$.

c. The regression equation is $\Delta SD_t = \alpha + \beta \Delta SD_{t-1} + \theta_0 \Delta U_t + \theta_1 \Delta U_{t-1} + \varepsilon$.

may have contributed to some of the fluctuations in the discrepancy around the 1990–91 recession.

Probably more important, over most of this sample the type of annual surveys used to compute the goods-producing sector of GDP(E) simply did not exist for most of the (enormous) services-producing sector. As a consequence the BEA was forced to cobble together estimates based on trade-source, administrative, and regulatory data, which may have missed part of the business cycle. For example, these data sources may miss fluctuations in the output of sole proprietors and some small businesses, both of which are highly cyclical. And the activities of many types of financial services companies or entities may have been missed by the regulatory data used by the BEA to compute personal consumption expenditures (PCE) for financial services. The magnitude of the booms and busts in financial services, then, may not be fully reflected in the PCE component of GDP(E) or exports of services. However, many of these firms and entities likely did file tax forms, so their activities would have been represented in the tax data used to compute GDP(I). This could explain part of the increase in the statistical discrepancy in 1989, 2001, and the latest episode.

Online appendix C also discusses potential reasons why GDP(I) might be too cyclical. It is possible that some capital gains, which should be excluded from the BEA's definition of output, were misreported to the Internal Revenue Service as ordinary income and thus included in the tax data used to compute GDP(I). Capital gains are likely to be highly procyclical, so failure to exclude them could have made GDP(I) more cyclical than true output. Although the evidence on this is thin, the third possibility above might be slightly more likely than the second. However, the evidence in favor of GDP(E) understating the cycle is stronger than the evidence in favor of GDP(I) overstating the cycle, so if the third possibility holds, it is probably the case that $\alpha_I - 1 < 1 - \alpha_E$.

Columns 3-4 through 3-6 of table 3 show that the statistical discrepancy is much less cyclical before the mid-1980s. Why might that be the case? Although PCE for services has always been a relatively large share of GDP(E), averaging 30 percent from 1947 to 1984, its share shot up to an average of 43 percent from 1985 to 2009 and reached 48 percent in 2009. As the share of services PCE has increased, the measurement problems in GDP(E) may have become more severe and more plainly visible. In addition, booms and busts in financial services may have accounted for a much larger share of the *variability* of the business cycle since the mid-1980s, a period that includes the junk bond boom and bust (as well as the savings and loan boom and bust) from the mid- to late 1980s, the day-trading boom in the

mid- to late 1990s and subsequent stock market crash from 2000 to 2002, and the mortgage securitization boom and bust from 2002 to 2008. GDP(E) may have missed much of this variation. But whatever the reason, since the really interesting divergences between the latest estimates occur in the post-1984Q3 period, the remainder of this section focuses on this sample.

### III.B. Information in the Revisions about the Latest Estimates

Consider the following hypothetical example. Two time series estimate the same unobserved variable of interest. The two time series happen to be identical, but they are known to be subject to considerable measurement error and may deviate widely from the true variable of interest. Suppose new information becomes available that leads to large revisions to one of the estimates, bringing it closer to the truth, while the other estimate remains unrevised. Which estimate is now better? Obviously, the estimate that was revised is better: it is now clear that it was far off initially and that the revisions corrected some or all of that measurement error, and that the estimate that was not revised remains far off. More generally, if two estimates start out identical, or reasonably close, and the revisions improve both estimates, then the estimate that is revised more will, on average, tend to be better than the estimate that is revised less. This is the underlying logic of Fixler and Nalewaik (2007). Table 1 shows that the initial estimates of $\Delta$GDP(I) and $\Delta$GDP(E) do start out with a very high correlation, but $\Delta$GDP(I) is revised more. Although the evidence in section II suggests that $\Delta$GDP(I) starts out as the better estimate, if one makes the relatively uncontroversial assumption that the revisions improve the estimates, then the larger revisions imply that $\Delta$GDP(I) has expanded its lead. Fixler and Nalewaik (2007) use this revisions evidence to establish bounds on the optimal weights to be placed on $\Delta$GDP(I) and $\Delta$GDP(E), and the bounds are favorable to $\Delta$GDP(I).

The revisions increase the variance of $\Delta$GDP(I) more than the variance of $\Delta$GDP(E), implying that they add some news, or actual variation in true output growth, to $\Delta$GDP(I) that is not added to $\Delta$GDP(E) (see Mankiw, Runkle, and Shapiro 1984, Mankiw and Shapiro 1986, and Fixler and Nalewaik 2007). This variation in true output growth missed by latest $\Delta$GDP(E) and captured by latest $\Delta$GDP(I) is closely related to the business cycle. In particular, Nalewaik (2007b) shows that the revisions tend to reduce $\Delta$GDP(I) by more than they reduce $\Delta$GDP(E) in low-growth states, so the extent of the weakness of true output growth in low-growth states appears to be part of the information missing from $\Delta$GDP(E) but appearing in $\Delta$GDP(I) through its more informative revisions. Since this weakness in low-growth states appears in neither initial estimate and remains missing in latest $\Delta$GDP(E),

if latest ΔGDP(E) is correct, the revisions showing this relative weakness in latest ΔGDP(I) must damage the estimates. More broadly, any suggestion that latest ΔGDP(E) is better than latest ΔGDP(I) would seem to imply that the variability added to ΔGDP(I) through the revisions moves it further away from true output growth. This seems hard to believe and, if carried to its logical conclusion, implies that the BEA should stop revising ΔGDP(I) and allocate its resources elsewhere. I do not think anyone at the BEA would seriously advocate taking that step. In contrast, the standard interpretation of the revisions is less problematic for the BEA: the revisions improve both ΔGDP(E) and ΔGDP(I), but the source data incorporated into ΔGDP(E) are simply not as informative as the source data incorporated in ΔGDP(I). But in that case, latest ΔGDP(I) is likely the better estimate.

### III.C.  Relationship to Other Business Cycle Variables

The logic behind the tests reported in table 4 is similar to the logic behind the regression results in table 2, but the table switches the regression order and reports results from pairs of regressions, one regressing latest ΔGDP(I) and one regressing latest ΔGDP(E) on each cyclical indicator. The $R^2$s in the second and third columns show that latest ΔGDP(I) is more highly correlated with every one of these cyclical indicators. Online appendix D repeats this exercise using annual instead of quarterly data, and the results are quite similar.[8] ΔGDP(I) is more highly correlated with lagged stock price changes, the lagged slope of the yield curve, and the lagged spread between high-yield corporate bonds and Treasury bonds (using a somewhat shortened sample).[9] It is more highly correlated with short and long differences of

---

8. Much of the source data incorporated at annual revisions are at an annual frequency, with no information on quarterly patterns, so the quarterly numbers are likely less reliable than the annual. For example, BEA analysts are confident that employee gains from exercising nonqualified stock options net out of the annual GDP(I) estimates (since profits fall by the same amount as the increase in compensation), but they are concerned that the quarterly pattern within years may be distorted (see Moylan 2008).

9. In his comment, Steven Landefeld suggests that stock market fluctuations may be more highly correlated with ΔGDP(I), because capital gains may be "leaking" into ΔGDP(I). If this is the case, the correlation between changes in stock prices and ΔGDP(I) should be contemporaneous, especially at an annual frequency, since a rising stock market translates immediately into a capital gain. Online appendix D shows that the evidence does not support this: using the annual output growth measures, ΔGDP(E) is slightly more highly correlated with the contemporaneous change in the stock market, whereas ΔGDP(I) is more highly correlated with the stock market change from one year earlier. The evidence is more suggestive of either the stock market anticipating changes in true output, or changes in the stock market affecting true output with a lag, with true output better represented by ΔGDP(I). See also Nalewaik (2008).

**Table 4.** Regressions of Growth in GDP(I) and GDP(E) on Selected
Business Cycle Indicators[a]

| | Adjusted $R^2$ | | Regression coefficients ($\beta$) | | Probability $p$ that the $\beta s$ are equal |
|---|---|---|---|---|---|
| Independent variable | GDP(I)$_t$ | GDP(E)$_t$ | GDP(I)$_t$ | GDP(E)$_t$ | |
| Log(S&P500$_t$/S&P500$_{t-7}$)/7 | 0.14 | 0.08 | 0.29 (0.06) | 0.20 (0.08) | 0.03 |
| $r_{t-1}^{\text{HYcorporate}} - r_{t-1}^{\text{Treasury (7yr)}}$ [b] | 0.28 | 0.19 | −0.67 (0.10) | −0.51 (0.14) | 0.06 |
| $r_{t-2}^{\text{HYcorporate}} - r_{t-2}^{\text{Treasury (7yr)}}$ [b] | 0.20 | 0.11 | −0.57 (0.13) | −0.41 (0.13) | 0.02 |
| $r_{t-3}^{\text{HYcorporate}} - r_{t-3}^{\text{Treasury (7yr)}}$ [b] | 0.18 | 0.06 | −0.54 (0.16) | −0.30 (0.15) | 0.00 |
| $r_{t-8}^{\text{Treasury (10yr)}} - r_{t-8}^{\text{Treasury (2yr)}}$ | 0.05 | 0.01 | 0.70 (0.37) | 0.36 (0.38) | 0.02 |
| (UR$_t$ − UR$_{t-1}$) × 4 | 0.26 | 0.24 | −1.47 (0.27) | −1.32 (0.30) | 0.40 |
| UR$_t$ − UR$_{t-4}$ | 0.25 | 0.10 | −1.74 (0.29) | −1.04 (0.32) | 0.00 |
| UR$_{t+2}$ − UR$_{t-2}$ | 0.35 | 0.21 | −2.19 (0.28) | −1.59 (0.34) | 0.02 |
| UR$_{t+4}$ − UR$_t$ | 0.24 | 0.18 | −1.81 (0.31) | −1.46 (0.31) | 0.08 |
| $(E_t^{\text{household}}/E_{t-1}^{\text{household}})^4$ | 0.30 | 0.20 | 1.00 (0.17) | 0.76 (0.21) | 0.11 |
| $E_t^{\text{household}}/E_{t-4}^{\text{household}}$ | 0.20 | 0.12 | 1.01 (0.19) | 0.74 (0.21) | 0.00 |
| $E_{t+2}^{\text{household}}/E_{t-2}^{\text{household}}$ | 0.34 | 0.24 | 1.36 (0.19) | 1.06 (0.22) | 0.04 |
| $E_{t+4}^{\text{household}}/E_t^{\text{household}}$ | 0.23 | 0.18 | 1.14 (0.23) | 0.93 (0.20) | 0.13 |
| ISM$_t^{\text{manuf.}}$ | 0.33 | 0.19 | 0.29 (0.05) | 0.20 (0.06) | 0.01 |
| ISM$_t^{\text{nonmanuf. c}}$ | 0.29 | 0.18 | 0.33 (0.08) | 0.24 (0.05) | 0.20 |
| Recession dummies | 0.29 | 0.24 | −5.05 (0.43) | −4.28 (0.77) | 0.46 |

Source: Author's regressions using BEA data.

a. The sample period is 1984Q3–2006Q4 except where noted otherwise. Newey-West standard errors with eight lags are in parentheses.

b. The sample period is 1988Q3–2006Q4.

c. The sample period is 1997Q3–2006Q4.

the unemployment rate, both contemporaneously and at leads and lags; the same holds true for the household survey measure of employment growth. Recall that there is no reason to suspect these measures to be spuriously correlated with $\Delta$GDP(I). $\Delta$GDP(I) is more highly correlated with the manufacturing ISM index and, when a shorter sample is used, with the nonmanufacturing ISM index as well. It is also more highly correlated with dummies for NBER recessions (see also Nalewaik 2007a).[10]

As in table 2, latest $\Delta$GDP(I) may be more highly correlated with all these variables because latest $\Delta$GDP(E) is contaminated with more noise. But the interpretation of the revisions offered in the previous subsection suggests a second type of measurement error, namely, that latest $\Delta$GDP(E) misses variation in true output growth that appears in all these cyclical indicators and is picked up by latest $\Delta$GDP(I). No matter which type of measurement error drives the results, latest $\Delta$GDP(I) is the better estimate of true output growth. For the more econometrically oriented reader, however, the table provides formal tests derived in Nalewaik (2008) that differentiate between the two types of measurement error. The tests reject the hypothesis that the second type of error, called lack of signal error (LoSE) in Nalewaik (2008), does not contaminate latest $\Delta$GDP(E). Regressions are run of each estimate on a testing variable, under the maintained assumption that the testing variable captures some of the variation missing from one estimate but included in the other. Nalewaik (2008) shows that the LoSE biases the regression coefficient on the testing variable toward zero, so the regression using the estimate that contains more LoSE yields a coefficient closer to zero. Note that it is measurement error of the LoSE form in the *dependent* variable that causes this attenuation bias, precisely the opposite of the conventional wisdom about classic measurement error, namely, that it is measurement error in the *explanatory* variable that causes attenuation bias. Testing the equality of the coefficients on the testing variable across the two regressions, Nalewaik

10. Note that these (often substantially higher) correlations are evidence against the crude model outlined in section III.A. In that model $\Delta$GDP(E) and $\Delta$GDP(I) contain rescaled versions of the same business cycle fluctuations, in which case the $R^2$ must be equal across the two regressions. That is clearly not the case: $\Delta$GDP(I) contains different business cycle fluctuations, which also show up in these other business cycle variables. Nalewaik (2008) uses essentially this same argument to reject a crude rescaling model in favor of the LoSE model; see below. Nevertheless, both the LoSE model and the rescaling model say the same thing, broadly speaking: GDP(E) growth misses some of the business cycle fluctuations in true output growth, which show up in GDP(I) growth as well as in other variables.
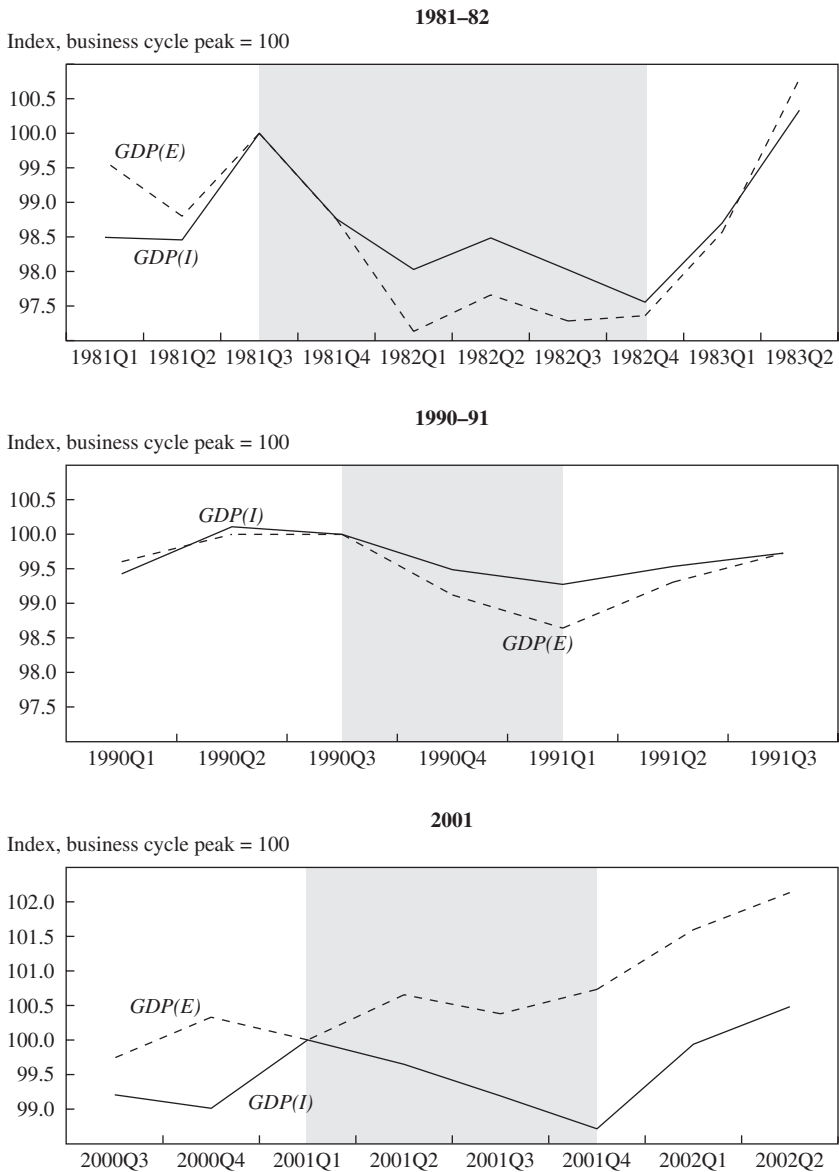
(2008) rejects using the asset price variables employed in the first five specifications of table 4, using a slightly different sample. Under the maintained assumption, ΔGDP(I) contains more signal about true output growth than ΔGDP(E)—signal that is reflected in stock and bond prices. Table 4 shows that this missing signal also appears in the differenced unemployment rate, household survey employment growth, and the ISM measures. The coefficients in table 4 are all larger, in absolute value, when ΔGDP(I) is the dependent variable; a relatively large amount of noise in ΔGDP(E) cannot explain these results, but a relatively large amount of LoSE can.

Less formal comparisons of GDP(E) and GDP(I) with other sources of information about the business cycle are also informative. In particular, one can compare the peaks and troughs in GDP(E) and GDP(I) with the NBER peak and trough dates. Grimm (2005) does this; figure 5 shows the results graphically for the three recessions before the most recent one. The one case where GDP(I) suggests a different dating than the NBER's is the 1990–91 recession: GDP(I) starts declining during the NBER peak quarter whereas GDP(E) is flat, but since the monthly peak was July 1990, the 1990Q3 GDP(I) decline seems consistent with the NBER dating. In the 1981–82 recession, the NBER called the trough in 1982Q4, the same quarter as the trough in GDP(I), whereas GDP(E) calls the trough three quarters earlier. In the 2001 recession it is difficult to discern any real cyclical downturn in GDP(E), whereas the NBER peak and trough dates line up perfectly with those of GDP(I). These peak and trough dates summarize the information in several other reliable indicators, and the fact that they line up better with GDP(I) is again suggestive that GDP(I) is the better estimate.

## IV. The Estimates over the 2007–09 Cyclical Downturn

The recent downturn looks considerably worse when output is measured using GDP(I) instead of GDP(E). Several differences can be cited. First, the effect on output appears sooner in GDP(I), which shows a sharp deceleration even before the NBER peak in late 2007. This deceleration was somewhat evident in the real-time estimates of GDP(I), but more important, the recession itself was much more evident in the real-time estimates of GDP(I) than in the real-time estimates of GDP(E). Second, the steepness of the plunge in output in late 2008 and early 2009 appears worse in the GDP(I) estimates. And third, with the BEA's February 2010 data release, the decline in output now appears more prolonged, extending into the summer of 2009.

**Figure 5.** GDP(E) and GDP(I) in Three Recessions[a]
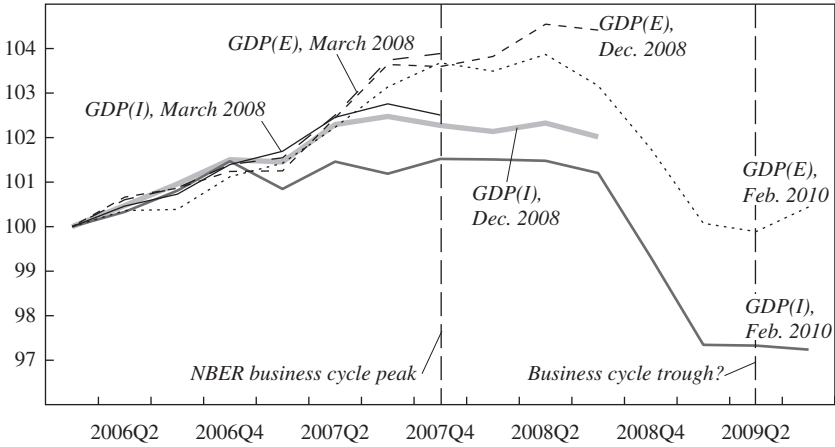
**1981–82**

Index, business cycle peak = 100



**1990–91**

Index, business cycle peak = 100



**2001**

Index, business cycle peak = 100



Source: Author's calculations using BEA data.
a. Shading indicates NBER recessions

**Figure 6.** GDP(E) and GDP(I) Estimates of Different Vintages, 2006Q1–2009Q3

Index, 2006Q1 = 100



Source: Author's calculations using BEA data.

Figure 6 shows GDP(E) and GDP(I) as they were measured at different dates, indexing them to 2006Q1 because the initial levels of the two series are different and have changed with revisions. The estimates released in March 2008, shortly after the start of the recession, both trend upward at a similar pace through 2006 and the first half of 2007 but then diverge considerably. GDP(I) shows the economy in a much more vulnerable state in late 2007, with output essentially flat over the second half of the year. GDP(E) shows little of this vulnerability in the second half; although growth was weak in 2007Q4, that weakness came on the heels of estimated annualized growth of almost 5 percent in 2007Q3.[11]

---

11. At the end of March 2008, the bivariate Markov switching model using ΔGDP(I) in Nalewaik (2007a) estimated a probability of around 90 percent that the economy had down-shifted to a low-growth state by 2007Q4, and this probability remained well above 50 percent throughout 2008. At the same time, a Markov switching model using ΔGDP(E) alone estimated a probability of less than 20 percent that the economy had downshifted to a low-growth state, a probability that remained low through most of 2008 (for example, it was 27 percent at the end of September), reaching 50 percent only after the BEA's advance 2008Q3 estimates released at the end of October. Models based on monthly indicators did no better: an implementation of the Diebold-Rudebusch (1996) monthly indicators model, based on Kim and Nelson (2000), did not rise above 50 percent until early November 2008, with the BEA's release of its initial personal income numbers for September 2008. The behavior of these models shows that real-time assessments of the state of the business cycle can be meaningfully improved by looking at GDP(I).

The GDP(E) and GDP(I) estimates released at the end of December 2008 followed the NBER's identification of December 2007 as a business cycle peak. All four of the monthly indicators that the NBER uses to date business cycles had peaked in late 2007 and early 2008, and GDP(I) was trending down slightly through the first three quarters of 2008 as well. GDP(E) was the only anomaly, showing continued growth at an annual rate of almost 2 percent in the first half of 2008.

The latest BEA estimates, those of February 2010, also depicted in figure 6, represent a downward revision of the initial ΔGDP(E) estimates for 2008 toward the initial ΔGDP(I) estimates, a continuation of the recent pattern in revisions discussed in section II. Revisions have also reduced GDP(I), but mainly in the first half of 2007. The latest estimates show that GDP(I) was essentially flat over the four quarters of 2007, declining in 2007Q1 and 2007Q3. These latest estimates suggest that the recent cyclical downturn caused a measurable deceleration in aggregate output much earlier than is commonly believed. Meanwhile, the latest estimate of GDP(E) shows no such early deceleration; instead GDP(E) grew 2.5 percent over the four quarters of 2007, about the same as in 2006. These differences over 2007 produce the bulk of the enormous swing in the statistical discrepancy observed in figure 4, from around −1.9 percent of GDP(E) in late 2006 to +1.8 percent in 2009Q3.

The current estimates show ΔGDP(E) actually slightly weaker than ΔGDP(I) in the first three quarters of 2008, but the current estimates of ΔGDP(I) then show a steeper downturn over the worst part of the recession. The current annualized ΔGDP(I) estimates for 2008Q4 and 2009Q1 are −7.3 and −7.7 percent, respectively, worse than the ΔGDP(E) estimates of −5.4 and −6.4 percent.

Finally, the latest ΔGDP(I) estimates for 2009Q3, released in late February 2010 and incorporating numbers from the Quarterly Census of Employment and Wages (see online appendix C), have called into question the timing of the trough of the recession. Before these numbers were released, a conventional wisdom was emerging that the recession had likely ended late in the second quarter of 2009, perhaps in June, with the economy resuming growth in 2009Q3. Figure 6 shows a modest rebound in GDP(E) in 2009Q3, but no rebound in GDP(I). Personal income less transfer payments and employment—two of the four indicators most emphasized by the NBER Business Cycle Dating Committee—continued to decline in 2009Q3.

What is one to make of these important differences between GDP(E) and GDP(I) over this cycle? All these estimates remain subject to considerable future revision, but the source data are most concrete for 2007, which

happens to be the period of greatest widening of the statistical discrepancy. Currently, two of the components of GDP(I), corporate profits and proprietors' income, incorporate IRS tax returns data through 2007, and declines in these two income categories account for the bulk of the deceleration in GDP(I) that year. Nonfarm proprietors' income (without inventory and capital consumption adjustments) increased by about $68 billion (nominal) in 2006 and fell $54 billion in 2007,[12] a deceleration of about $122 billion. The biggest declines in 2007 were in real estate, construction, and finance and insurance, as well as (less explicably) mining (see table 6.12D of the National Income and Product Accounts). As noted in online appendix C, some of the decline in proprietors' income may have represented a decline in capital gains from house flipping, which should not be included in the relevant concept of output. Real estate proprietors' income fell $24 billion in 2007, but it also fell $14 billion in 2006, suggesting that this type of mismeasurement cannot explain much of the widening of the statistical discrepancy in 2007. Construction proprietors' income decelerated from a $6 billion increase in 2006 to a $14 billion decline in 2007, with the current estimates showing a $46 billion decline in 2008. Part of this decline in proprietors' income should probably have shown up in lower spending on residential improvements, but as discussed earlier and in online appendix C, the BEA's averaging of its raw source data will tend to miss such a large deceleration. Currently, the raw estimates of improvements spending from the Census show declines of 4 and 14 percent in 2007 and 2008, respectively, steeper than the current BEA estimates of 1 and 4 percent. If the Census numbers are correct, GDP(E) should be $5 billion lower in 2007 and about $22 billion lower in 2008, so this also explains only a small portion of the widening of the statistical discrepancy.[13]

---

12. The 2007 decline in the raw IRS tax numbers was larger, about $66 billion (see table 7.14 of the National Income and Product Accounts), but the BEA reduced this figure with various adjustments.

13. Some other data sources suggest much larger declines in spending on residential improvements. For example, Greenspan and Kennedy (2005, 2007) use Flow of Funds data, and Mian and Sufi (2009) use data from credit rating agencies, to show that households extracted a very large amount of home equity in the mid-2000s, before banks cut credit lines in 2007 and 2008 and equity extraction dropped dramatically. Using survey evidence that households spend about a third of extracted home equity on home improvements—see Brady, Canner, and Maki (2000), Canner, Dynan and Passmore (2002), and Greenspan and Kennedy (2007) and the references therein—updated Greenspan-Kennedy estimates give declines in spending on home improvements of $66 billion in 2007 and $80 billion in 2008. Of course, this does not necessarily imply causality from equity extraction to spending, because households may have found other financing options when home equity lines of credit dried up.

Corporate profits increased by about $183 billion in 2006 and fell $48 billion in 2007—a deceleration of about $230 billion—and the current estimates for 2008 show a decline of $181 billion. The biggest decline in profits in 2007 was in the finance and insurance industry, where a $54 billion decline in 2007 followed an increase of $4 billion in 2006 and a massive increase of about $180 billion from 2000 to 2005. Looking more broadly, the sum of corporate profits, proprietors' income, and wages and salaries for the finance and insurance industry fell by close to 4 percent in 2007, while PCE for financial services increased by more than 12 percent. Although the categories are not strictly comparable, these numbers are difficult to reconcile unless there is severe measurement error in either the income measures or PCE.[14] In November 2007, in a technical note to its preliminary third-quarter GDP release, the BEA raised the issue of its ability to strip out capital losses (bad debt expenses and asset write-downs) from its initial estimates of financial companies' profits, but the availability of the tax data for 2007 likely made these subtractions much easier. The problems now appear more concentrated in the measurement of financial services PCE and services more generally on the expenditure side, as discussed in the previous section and in online appendix C. Given the advent of the financial crisis and the disappearance of many securitization markets in the second half of 2007, a 12 percent growth rate for financial services PCE seems implausibly high. To get a sense of the magnitudes involved, consider that a decline in financial services PCE of 4 percent would have lowered GDP(E) in 2007 by $76 billion from its current level, and by more if PCE missed the boom in financial services output over prior years. More recently, profits in the finance and insurance industry fell an additional $91 billion in 2008 (with proprietors' income and wage and salary income also falling), while financial services PCE increased once again. Since the tax data have not yet been incorporated for 2008, some risk remains that the income declines were too steep, but again it seems implausible that financial services PCE continued its uninterrupted growth.

Overall, this evidence suggests that although there may be problems on both sides of the accounts, they are likely more severe on the expenditure side. Given that, the latest downturn was likely substantially worse than the current GDP(E) estimates show. Output likely decelerated sooner, fell at a

---

14. The output of financial services could also have shown up in exports, or as an intermediate input into the production of other industries.

faster pace at the height of the downturn, and recovered less quickly than is reflected in GDP(E)—and in conventional wisdom.

## V. Concluding Thoughts

Considerable evidence suggests that the growth rate of GDP(I) better captures the business cycle fluctuations in true output growth than does the growth rate of GDP(E). For the initial growth rate estimates, the revisions evidence over the past 15 years, the correlations with other business cycle indicators, and the recent behavior of the estimates around cyclical turning points all point to this conclusion. For the latest estimates that have passed through their cycle of revisions, careful consideration of the nature of the source data, statistical analysis of the information added by the revisions, and statistical tests, as well as informal comparisons with other business cycle indicators, again all suggest that GDP(I) growth is better than GDP(E) growth at tracking fluctuations in true output growth.

These results strongly suggest that economists and statisticians interested in business cycle fluctuations in U.S. output should pay attention to the income-side estimates and consider using some sort of weighted average of the income- and expenditure-side estimates in their analyses. The evidence in this paper clearly suggests that the weights should be skewed toward GDP(I), but even a 50–50 average would be a marked improvement over an estimate that places all its weight on GDP(E). It would also follow the lead of the Council of Economic Advisers, who, after concluding in the 1997 *Economic Report of the President* that GDP(I) might be better than GDP(E), have subsequently given some weight to the income-side estimates in their productivity analyses: see *Economic Report of the President 2008,* p. 39, and *Economic Report of the President 2009,* pp. 47–48.

The results here also have implications for the BEA. When a new quarterly estimate of GDP(I) growth becomes available, the evidence here shows that it is likely to be a better estimate of output growth than the corresponding GDP(E) estimate. However, the first GDP(E) estimate for any given quarter, the advance estimate, is typically released about a month before the first GDP(I) estimate, and GDP(I) is delayed an additional month when the BEA is producing estimates for fourth quarters. As noted above, these delays occur because the BEA has incomplete information on corporate profits and is not comfortable releasing earlier estimates of profits. In general, the profits information released by the BEA appears tremendously useful, and the BEA does have some information on profits at these earlier release dates. An advance estimate of GDP(I) based on the available profits

information might be quite helpful for real-time assessment of the speed of economic growth. Earlier release of the fourth-quarter GDP(I) estimates, so that an estimate is available at least as early as the BEA's second release, might be similarly helpful; the BEA has still not released an estimate of GDP(I) growth for the fourth quarter of 2009 as of this writing in mid-March 2010. What the BEA decides will depend on how much information on profits is really available at these earlier dates, and a thorough assessment of this issue seems in order.

The BEA, the Census Bureau, and the Bureau of Labor Statistics doubtless will continue making improvements in their estimates where feasible, and the good news is that there have been substantial improvements over the past few years. The data on services have progressed by leaps and bounds, with the advent of the Census' Quarterly Services Survey in 2003, the recent expansions in the coverage of this survey, and the expansions in the coverage of the Census' Service Annual Survey. Further improvements are in train: in December 2010 the estimates from the Service Annual Survey will roughly double in coverage, expanding to mimic the sector coverage of the Economic Census. These data should improve the estimates of PCE and GDP(E).

Despite these improvements, however, problems with the output growth estimates will inevitably remain, and lack of coverage of services is only one of several important limitations of GDP(E). All the results in this paper suggest that the current reporting practice of the BEA, which puts nearly exclusive emphasis on GDP(E) over GDP(I), is suboptimal statistically. The BEA creates tremendous value by producing an income-based estimate of output growth, but current BEA reporting practice downplays that estimate so much that many analysts may not even be aware of its existence. The BEA's typical press release rarely discusses GDP(I), and it is reported only toward the back of the release, and then as a nominal level, requiring analysts to deflate and compute annualized quarterly growth rates themselves to arrive at a number comparable to headline real GDP(E) growth.

If the BEA finds the results here persuasive, there are several incremental steps it could take to increasing the prominence of GDP(I). Most obviously, the BEA could report real annualized growth rates of GDP(I) in its press releases, preferably in table 1 of the release so they can easily be compared with the annualized growth rates of real GDP(E). Second, it could give those annualized growth rates more prominence in the text of the press releases, discussing them at a level of detail similar to its current discussion of GDP(E). The BEA's discussion of the corporate profits estimates could be rolled into a more general discussion of GDP(I). Third, the BEA could bring

more balance to its statements about the reliability of GDP(E) and GDP(I). Steven Landefeld, Eugene Seskin, and Barbara Fraumeni (2008, p. 211) take a small step in this direction by stating, "what these studies remind users is that it is useful to look at growth in both GDP and gross domestic income in assessing the current state of the economy."

Featuring two measures of output growth in the same press release would raise communication challenges, and the BEA might fear that such a practice might prove too confusing for casual analysts.[15] Here the example of other countries is relevant: the United Kingdom and Australia, for example, report an average of the two sides of the accounts as their featured output growth measure. The BEA has considered taking this step in the past (see, for example, Moulton 2000), and it could report such an average of GDP(E) and GDP(I) as "GDP(A)."[16] The BEA could employ optimal weights guided by statistical analysis, as in Fixler and Nalewaik (2007), but the results here suggest that featuring even a simple 50–50 average would be a marked improvement over the current practice of featuring GDP(E) alone.

15. The BLS does report two estimates of the monthly change in employment in its employment report, one from the establishment survey and one from the household survey, but there are clear statistical reasons for favoring the establishment survey number at the monthly frequency. For the case of GDP(E) and GDP(I), making the case in favor of one measure over the other is more complicated.

16. Of course, the components of GDP(E) will not sum to the top-line GDP(A), nor will the components of GDP(I), and this may be confusing for some analysts. But if the evidence in this paper is convincing, the components of GDP(E) already do not sum to true output or even to the best estimate of true output; in fact, the sum of the components of GDP(E) misses important, systematic variation in true output. Reporting an average would simply make these facts explicit. Over the long run, allocating parts of the discrepancy to different components of GDP(E) and GDP(I) may be the right thing to do, but this would be an extremely complicated task, and much research would need to be done beforehand. But if the BEA does attempt to go down this path at some point, it should do so in a transparent and easily replicable fashion. The BEA is to be commended for its transparency in reporting the statistical discrepancy and should do nothing to compromise this transparency.

# References

Abel, Andrew B., and Ben S. Bernanke. 2001. *Macroeconomics,* 4th ed. Boston: Addison Wesley.

Brady, Peter J., Glenn B. Canner, and Dean M. Maki. 2000. "The Effects of Recent Mortgage Refinancing." *Federal Reserve Bulletin* (July), pp. 441–50.

Bureau of Economic Analysis. 2002. "Corporate Profits: Profits before Tax, Profits Tax Liability, and Dividends." Methodology Paper. Washington (September).

———. 2008a. "Updated Summary of NIPA Methodologies." *Survey of Current Business* (November), pp. 8–25.

———. 2008b. "Concepts and Methods of the U.S. National Income and Product Accounts (Introductory Chapters 1–4)." Washington (July).

Canner, Glenn, Karen Dynan, and Wayne Passmore. 2002. "Mortgage Refinancing in 2001 and Early 2002." *Federal Reserve Bulletin* (December): 469–81.

Diebold, Francis X., and Glenn D. Rudebusch. 1996. "Measuring Business Cycles: A Modern Perspective." *Review of Economics and Statistics* 78, no. 1: 67–77.

Fixler, Dennis, and Bruce Grimm. 2006. "GDP Estimates: Rationality Tests and Turning Point Performance." *Journal of Productivity Analysis* 25: 213–29.

Fixler, Dennis J., and Jeremy J. Nalewaik. 2007. "News, Noise, and Estimates of the 'True' Unobserved State of the Economy." FEDS Working Paper 2007-34. Washington: Board of Governors of the Federal Reserve System,

Fleischman, Charles, and John Roberts. 2010. "A Multivariate Estimate of Trends and Cycles." Washington: Board of Governors of the Federal Reserve.

Greenspan, Alan. 2004. "Risk and Uncertainty in Monetary Policy." *American Economic Review* 94, no. 2: 33–40.

Greenspan, Alan, and James Kennedy. 2005. "Estimates of Home Mortgage Originations, Repayments, and Debt on One-to-Four-Family Residences." FEDS Working Paper 2005-41. Washington: Board of Governors of the Federal Reserve System.

———. 2007. "Sources and Uses of Equity Extracted from Homes." FEDS Working Paper 2007-20. Washington: Board of Governors of the Federal Reserve System.

Grimm, Bruce T. 2005. "Alternative Measures of U.S. Economic Activity in Business Cycles and Business Cycle Dating." BEA Working Paper 2005-05. Washington: Bureau of Economic Analysis.

Landefeld, J. Steven, Eugene P. Seskin, and Barbara M. Fraumeni. 2008. "Taking the Pulse of the Economy: Measuring GDP." *Journal of Economic Perspectives* 22 (Spring): 193–216.

Mankiw, N. Gregory, and Matthew D. Shapiro. 1986. "News or Noise: An Analysis of GNP Revisions." *Survey of Current Business* (May): 20–25.

Mankiw, N. Gregory, David E. Runkle, and Matthew D. Shapiro. 1984. "Are Preliminary Announcements of the Money Stock Rational Forecasts?" *Journal of Monetary Economics* 14, no. 1: 15–27.

Mian, Atif R., and Amir Sufi. 2009. "House Prices, Home Equity-Based Borrowing, and the U.S. Household Leverage Crisis." Working Paper no. 15283. Cambridge, Mass.: National Bureau of Economic Research.

Moulton, Brent R. 2000. "Getting the 21st-Century GDP Right: What's Underway?" *American Economic Review* 90, no. 2: 253–58.

Moylan, Carol E. 2008. "Employee Stock Options and the National Economic Accounts." *Survey of Current Business* (February): 7–13.

Nalewaik, Jeremy J. 2007a. "Estimating Probabilities of Recession in Real Time Using GDP and GDI." FEDS Working Paper 2007-07. Washington: Board of Governors of the Federal Reserve System.

———. 2007b. "Incorporating Vintage Differences and Forecasts into Markov Switching Models." FEDS Working Paper 2007-23. Washington: Board of Governors of the Federal Reserve System.

———. 2008. "Lack of Signal Error (LoSE) and Implications for OLS Regression: Measurement Error for Macro Data." FEDS Working Paper 2008-15. Washington: Board of Governors of the Federal Reserve System.

# Comments and Discussion

COMMENT BY

**FRANCIS X. DIEBOLD**[1]   The topic of Jeremy Nalewaik's paper, the
measurement of aggregate output, is of central importance. His case for
preferring GDP(I) to GDP(E) is well argued and in certain key respects
persuasive. Henceforth it will be impossible for macroeconomic analy-
ses to proceed comfortably simply using GDP(E), as if the choice between
GDP(E) and GDP(I) were inconsequential. Exclusive focus on GDP(E)
will require justification and may have to be abandoned.

In my view, however, universal prescriptions (which Nalewaik does
not offer, but others might) are unlikely to emerge. Rather, the compara-
tive merits of GDP(E) and GDP(I) depend on the context. That is, use of
one measure or the other will likely produce different answers for some
questions and effectively indistinguishable answers for others. I will sub-
stantiate this claim in two contexts: aggregate output measurement and
business cycle measurement. I will emphasize, moreover, that the important
issue is not which of the two is "better," but rather how best to combine
them, and what is ultimately added by GDP(I). I will argue that GDP(I)
has much to add for aggregate output measurement, and little to add for
business cycle measurement.

Consider first the choice of measure for aggregate output. This is the
context in which Nalewaik primarily works, and in which, in my view,
his assertions are most persuasive. He argues from a variety of perspec-
tives that GDP(I) may be superior to GDP(E). That is initially surprising—

indeed, shocking—given the near-universal neglect of GDP(I). But one must not overinterpret the result. Even if one grants that several arguments favor GDP(I) over GDP(E), one must also recognize that there is no need to choose one or the other. Instead, there may be gains from combining the two.

Consider forming a combined GDP measure, GDP(C), by taking a convex combination of GDP(E) and GDP(I):

$$\text{GDP}(C) = \lambda \text{GDP}(E) + (1 - \lambda)\text{GDP}(I).$$

This is just a "portfolio" of the two measures. Under conditions from the forecast combination literature (see, for example, Diebold 2007), the optimal portfolio weight $\lambda^*$ is

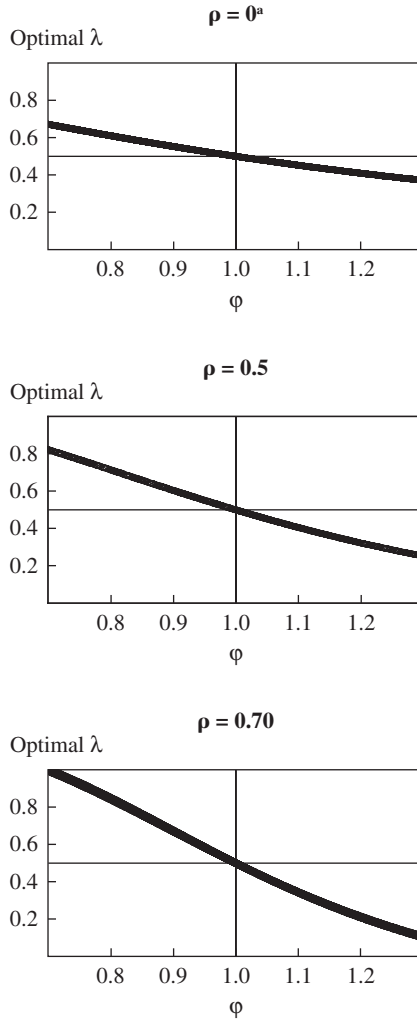$$\lambda^* = \frac{1 - \varphi\rho}{1 + \varphi^2 - 2\varphi\rho},$$

where $\varphi = \text{var}(e_{\text{GDP(E)}})/\text{var}(e_{\text{GDP(I)}})$, $\rho = \text{corr}(e_{\text{GDP(E)}}, e_{\text{GDP(I)}})$, $e_{\text{GDP(E)}} = \text{GDP} - \text{GDP(E)}$, and $e_{\text{GDP(I)}} = \text{GDP} - \text{GDP(I)}$. It is natural and desirable that $\lambda^*$ depend on the variance ratio $\varphi = \text{var}(e_{\text{GDP(E)}})/\text{var}(e_{\text{GDP(I)}})$. In particular, as $\text{var}(e_{\text{GDP(E)}})$ increases relative to $\text{var}(e_{\text{GDP(I)}})$, the optimal weight on GDP(E) drops, other things equal. It is similarly natural that $\lambda^*$ depend on $\rho$, which determines the benefits of portfolio diversification.

I illustrate the situation in figure 1, which plots $\lambda^*$ as a function of $\varphi$, for various values of $\rho$. For $\varphi = 1$, the optimal weight on GDP(E) is always ½, and the optimal weight drops toward zero as $\varphi$ increases. The speed with which it drops, moreover, increases as $\rho$ increases.[2]

The key observation is that, except for extreme values of $\varphi$ or $\rho$, or both, both GDP(E) and GDP(I) should receive significant weight in an informed assessment of aggregate output. Suppose, for example, that $\varphi = 1.1$, that is, that $\text{var}(e_{\text{GDP(E)}})$ is 10 percent greater than var $(e_{\text{GDP(E)}})$, and that $\rho = 0.5$, that is, that $e_{\text{GDP(E)}}$ and $e_{\text{GDP(I)}}$ are positively correlated, but not overwhelmingly so). Then the middle panel of the figure indicates an optimal GDP(E) weight of $\lambda^* = 0.4$. Weights near 0 or 1 would require extreme variance ratios, or extreme correlations, or both. Optimal weights may, however, be time

---

2. As $\rho$ increases, the gains from diversification decrease, and so one diversifies less, other things equal.

**Figure 1.** Optimal Portfolio Weights of GDP(E) and GDP(I) Given the Error Variance Ratio for Various Correlations of GDP(E) and GDP(I)

$\rho = 0^a$

Optimal $\lambda$

$\rho = 0.5$

Optimal $\lambda$

$\rho = 0.70$

Optimal $\lambda$

Source: Author's calculations.

a. $\rho = \text{corr}(e_{\text{GDP(E)}}, e_{\text{GDP(I)}})$; $\lambda^* = \dfrac{1 - \varphi\rho}{1 + \varphi^2 - 2\varphi\rho}$, $\varphi = \text{var}(e_{\text{GDP(E)}})/\text{var}(e_{\text{GDP(I)}})$, where $e_{\text{GDP(E)}} = \text{GDP} - \text{GDP(E)}$ and $e_{\text{GDP(I)}} = \text{GDP} - \text{GDP(I)}$.

varying, reflecting changes in measurement error variances and covariances (over the business cycle, for example).

Now consider measuring the business cycle, another task of central importance, as also emphasized in Nalewaik's paper. A key insight, emphasized by Arthur Burns and Wesley Mitchell (1946) and Robert Lucas (1977), and clearly reflected, for example, in the National Bureau of Economic Research's business cycle dating methodology, is that the business cycle is not about any single variable (including GDP). That is, many indicators of business conditions (including GDP) are *related* to the business cycle, but no single indicator *is* the business cycle.

The so-called dynamic factor model embodies the Burns-Mitchell-Lucas insight and has become a standard tool for empirical characterization of the business cycle (see, for example, Sargent and Sims 1977, Stock and Watson 1989, Diebold and Rudebusch 1996, and Aruoba and Diebold 2010). In the dynamic factor framework, one treats the state of the business cycle as latent, with observed indicators of business conditions providing noisy signals, and uses the Kalman filter to produce optimal estimates of the business cycle from the noisy signals.

Does the choice of GDP(E) or GDP(I) matter for business cycle measurement, which, as I have emphasized, involves monitoring not only GDP but also a variety of other business conditions indicators? I will address this question using a five-variable dynamic factor model nearly identical to that of the Federal Reserve Bank of Philadelphia based on payroll employment, industrial production, personal income less transfers, manufacturing and trade sales, and GDP (see Aruoba and Diebold 2010 for details).[3]
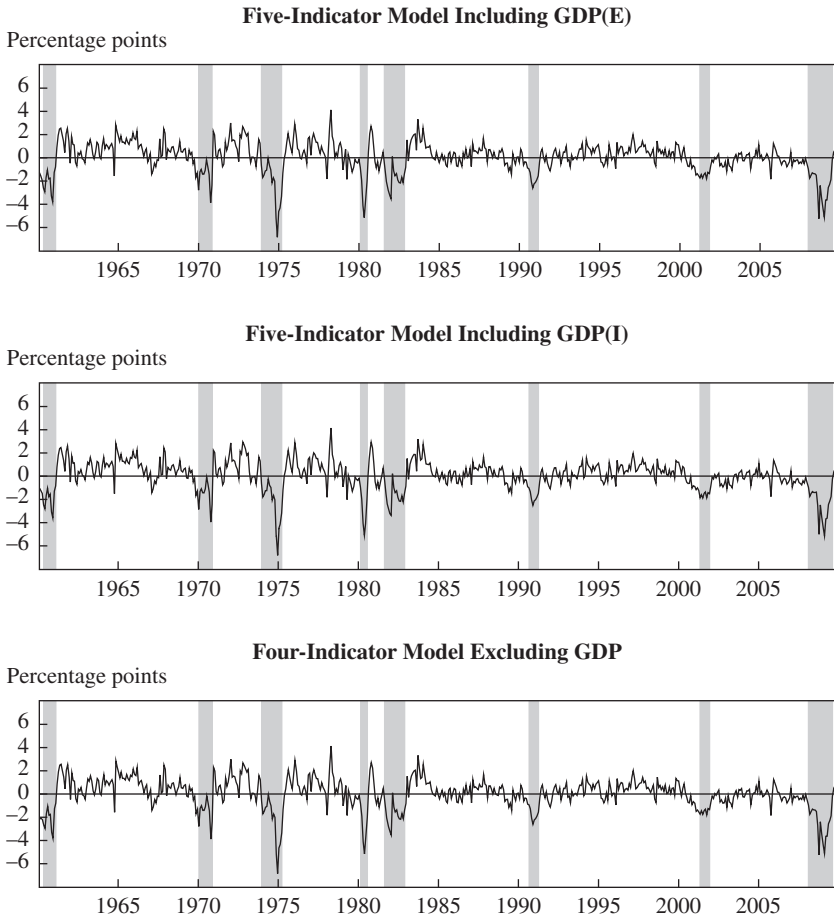
Figure 2 shows the business cycle factor extracted using several versions of the five-indicator dynamic factor model. The top panel uses GDP(E), and the middle panel uses GDP(I). The difference is negligible. Evidently, given the information in the other four indicators, it makes no difference which estimate of GDP is included as a fifth. Indeed, the bottom panel, based on a four-variable model that simply excludes GDP, produces a nearly identical business cycle factor.

To conclude, Nalewaik's insightful and eye-opening paper deserves significant attention. As I have emphasized, however, the relevant question is not likely to be, "Which of GDP(E) and GDP(I) is better?" or "Which of GDP(E) and GDP(I) should economists use?" Rather, it is how best to

---

3. The Philadelphia Fed model is described at www.philadelphiafed.org/research-and-data/real-time-center/business-conditions-index/.

**Figure 2.** Extracted Business Cycle Real Activity Factor Using Alternative Output Measures, 1960–2010

Five-Indicator Model Including GDP(E)

Percentage points



Five-Indicator Model Including GDP(I)

Percentage points



Four-Indicator Model Excluding GDP

Percentage points



Source: Author's calculations.

*blend* GDP(I) with GDP(E). GDP(I) has much to contribute in some contexts, and little in others.

REFERENCES FOR THE DIEBOLD COMMENT

Aruoba, S. Borağan, and Francis X. Diebold. 2010. "Real-Time Macroeconomic Monitoring: Real Activity, Inflation, and Interactions." *American Economic Review* 100, no. 2: 20–4.

Burns, Arthur F., and Wesley C. Mitchell. 1946. *Measuring Business Cycles.* New York: National Bureau of Economic Research.

Diebold, Francis X. 2007. *Elements of Forecasting,* 4th ed. Mason, O.: Thomson South-Western Publishing.

Diebold, Francis X., and Glenn D. Rudebusch. 1996. "Measuring Business Cycles: A Modern Perspective." *Review of Economics and Statistics* 78, no. 1: 67–77.

Lucas, Robert E. 1977. "Understanding Business Cycles." *Carnegie-Rochester Conference Series on Public Policy* 5: 7–29.

Sargent, Thomas J., and Christopher A. Sims. 1977. "Business Cycle Modeling without Pretending to Have too Much a Priori Economic Theory." In *New Methods in Business Cycle Research: Proceedings from a Conference.* Federal Reserve Bank of Minneapolis.

Stock, James H., and Mark W. Watson. 1989. "New Indexes of Coincident and Leading Economic Indicators." *NBER Macroeconomics Annual* 4: 351–94.

**COMMENT BY**

**J. STEVEN LANDEFELD**[1]    Jeremy Nalewaik's paper is an excellent piece of research. BEA appreciates work such as this on measurement issues related to its economic accounts, as well as the opportunity to discuss it in more detail.[1] External research, complemented by research at BEA, has long been the source of a wide range of statistical improvements, ranging from chain indexes to hedonic indexes. Although I have a number of questions about the conclusions outlined in this paper, it will certainly serve as the basis for several future research endeavors: first, for further research on the sources of apparent cyclical patterns in the statistical discrepancy between GDP and GDI—what Nalewaik refers to as GDP(E) and GDP(I), respectively; second, for reconciliation with related BEA work on revisions; third, for further work on the balancing of income, production, and expenditure now done in the industry accounts, with particular attention to their use in balancing annual GDP and GDI estimates; fourth, for exploration of the means by which BEA can better present data on GDI and the range of revisions in GDP and GDI estimates without unduly confusing the general community of users; and, finally, and most important, for continued work on improving the early source data for both GDP and GDI to address the measurement issues raised in this paper.[2]

---

1. These comments reflect the very helpful ideas and calculations of my colleagues, Brent Moulton, Dennis Fixler, Bruce Grimm, and Shaunda Villones.

2. For more information see Fixler and Nalewaik (2009) and Fixler and Grimm (2002, 2006).

In general, the conclusion that the gross domestic income measure of aggregate output is deserving of attention is noncontroversial. The national accounts have double-entry accounts for purposes not only of providing multiple estimates of the breakdown of GDP by expenditure and income, but also of providing a check on the consistency of the two sets of estimates and identifying and correcting sources of discrepancies. As far as I can determine, BEA has never suggested that GDP is the "true" estimate of output, or that GDI is not a meaningful and useful measure of economic activity.

However, the main conclusion of this paper is that GDI is a better indicator than GDP of economic activity over the business cycle. My own conclusions are as follows: First, the evidence suggests that GDP and GDI provide roughly the same picture of economic activity over the business cycle and that a review of the source data and performance of the two measures favors GDP rather than GDI, but both have their strengths and weaknesses. Second, any gain in accuracy from averaging the GDI and GDP estimates is likely to be small. And third, some of the measurement concerns raised in this paper about the ability of GDP and GDI to fully capture changes in the economy over the business cycle are in the process of being resolved, thanks to new quarterly source data on services from the Census Bureau and more comprehensive monthly data on wages and salaries from the Bureau of Labor Statistics. Other concerns, especially those related to the cyclicality of corporate profits and other variables on the income side, are less tractable and will require further research.

QUALITY OF SOURCE DATA FOR GDP AND GDI. In contrast to the paper's assessment, I would describe the source data for the early GDI estimates as considerably less complete, consistent, and timely than the source data for the early GDP estimates. As a result, a significantly smaller share of the early estimates for GDP is based on trend extrapolators rather than directly on source data. Moreover, BEA views the GDP source data as generally superior to the GDI source data, because they are collected for statistical purposes and based on a consistent set of survey definitions designed to be used with the national accounts. They are collected by the Census Bureau as part of a consistent set of business surveys using the same universe and samples to collect monthly, quarterly, annual, and comprehensive (once every five years) data. In contrast, the source data for the GDI estimates are mainly taken from financial statements or collected by a variety of regulatory and tax agencies for nonstatistical purposes. These "administrative" data utilize a wide range of concepts and definitions, many of which differ significantly from those used in the national accounts. They also differ in

scope and coverage. As a result, the income-side source data—especially for profits, proprietors' income, rental income, and interest income—differ significantly over time because of changes in business accounting and tax rules, changes in business practices, and changes in business conditions. The estimates from these sources also vary for the same time period, raising concerns about the consistency of estimates compiled from a combination of these sources. (For example, BEA's initial corporate profits estimates are based on companies' financial reports and financial accounting rules, whereas the latest estimates are based on companies' tax reports and IRS accounting rules.) Finally, significant tax incentives and corporate reporting requirements can bias information based on business, financial, and tax records. BEA takes great pains to adjust these administrative data to provide information consistent with the national accounts in terms of definition, scope, and timing, but such adjustments are challenging.

Whereas 86 percent of the early GDP estimates is based on some form of direct monthly or quarterly source data, only 37 percent of the early GDI estimates is based on such data; the rest is based on ratio adjustments, judgment, or trend estimators. The largest extrapolations are for the following: nonwage compensation, or supplements, which account for 18 percent of compensation and 10 percent of GDI; wages and salaries for nonproduction and supervisory workers, which include irregular payments and account for 45 percent of compensation and 21 percent of GDI; interest expense and rental income, which account for 8 percent of GDI; and proprietors' income, which includes large adjustments for misreporting and accounts for 8 percent of GDI (table 1).[3]

For the major components of GDI, revisions to later vintages of the estimates are sometimes significant. For example, the initial estimates of total wages and salaries have been subject to significant revision when the quarterly administrative (payroll tax) data become available in the Quarterly Census of Employment and Wages (QCEW). These revisions reflect the fact that although production and nonsupervisory workers account for roughly two-thirds of employment, they account for only a little more than

---

3. The 13 percent of the third estimates of GDP that is trend based is mainly in service components of personal consumption expenditures, including "other" services, "other" transportation, medical services, recreation, personal care, other personal business services, education and research, and religious and welfare services—as well as "other" state and local expenditures.

**Table 1. Source Data for Gross Domestic Income, Estimates for 2007[a]**

Percent of GDI

| Component of GDI | Judgmental trend | Early source data | Description of early source data or estimation method |
|---|---|---|---|
| Compensation of employees | | | |
| Wages and salaries | | | |
| Nonsupervisory and production workers | | 17.1 | BLS Current Employment Statistics: payroll survey |
| Supervisory/nonproduction workers | 20.6 | | Judgmental extrapolation based on payroll survey |
| Government | | 7.7 | Payroll survey employment and ECI |
| Supplements | 10.3 | | Judgmental trend extrapolation |
| Taxes on production and imports, less subsidies | | | |
| Property taxes | 2.8 | | Judgmental trend extrapolation |
| Other | | 4.1 | Federal Monthly Treasury data; Census data for sales taxes |
| Net interest and miscellaneous payments | 6.8 | | FDIC data for commercial banks; judgmental trend extrapolation based on interest rates for most of the remainder |
| Business current transfer payments | 0.7 | | Judgmental trend extrapolation |
| Proprietors' income | 7.8 | | Judgmental trend extrapolation based on BLS payroll survey, Census data, and other indicators |
| Rental income of persons | 1.0 | | Mixture of actual source data and judgmental extrapolation |
| Corporate profits | | 8.5 | Census Quarterly Financial Report, FDIC, Compustat data |
| Current surplus of government enterprises | 0.0 | | Judgmental trend extrapolation |
| Consumption of fixed capital | 12.5 | | Judgmental extrapolation based on BEA capital stocks |
| Total | 62.6 | 37.4 | |
| Percent based on early source data that are conceptually consistent with annual or benchmark data | | 11.8 | |

Source: Bureau of Economic Analysis.

a. BLS = Bureau of Labor Statistics; ECI = Employment Cost Index; FDIC = Federal Deposit Insurance Corporation.

half of wages and salaries, and the fact that the payroll survey does not capture stock options, bonuses, and other irregular payments.[4]

Further, although the QCEW data, which are available 4 months after the advance GDP report, cover virtually all workers, they are quite volatile and have proved to be extremely difficult to measure on a seasonally adjusted basis. Moreover, once annual QCEW data are received, there can be significant revisions in the quarterly data.

Corporate profits are even more difficult to measure, and early estimates based on corporate financial statements can differ significantly from both the economic accounting measure from BEA and the tax-based measure from the IRS. According to BEA's revision studies, corporate profits have the largest mean absolute revision of any component of GDP or GDI, except for farm proprietors' income.

The large revisions to profits reflect a number of factors, including the large differences between financial and tax accounting rules and BEA's economic accounting conventions; the use of financial corporate data for public companies to extrapolate profits for private or S corporations; and the possible effects of capital gains and losses or "unusual" losses—which should be excluded from GDI—in the source data for profits. And the final profits numbers differ widely depending on the source of the data. For 2005, profits as reported in the IRS Statistics of Income (SOI) increased by 43 percent, S&P operating profits by 9 percent, and Census Bureau Quarterly Financial Report profits by 15 percent. The mean absolute difference between the highest and the lowest estimate of growth in profits from 1999 to 2007 was 23 percentage points, with the largest differences recorded in 2001. Although many of these differences are relatively easily resolved, many others, such as those surrounding major changes in the economy, changes in accounting rules and practices, or changes in tax law, can be quite difficult.

The GDP estimates are, of course, not without their own limitations. As Nalewaik points out, one of the most important has been the absence of a timely, comprehensive data source for services in the early GDP estimates. Extrapolators for services may well have contributed to the tendency of the early GDP estimates to understate the decline in GDP during

4. Beginning with the first quarter of 2010, BEA estimates of wages and salaries reflect newly available monthly tabulations of hours and earnings for *all employees* on private non-farm payrolls from the BLS's expanded current employment statistics program. However, the new BLS monthly data do not include certain types of irregular pay, such as bonuses and stock options, which are included in the QCEW data.

contractions and the increase during the early stages of expansions (Fixler and Grimm 2002).[5]

One of the important advantages of the GDP estimates is that source data (mainly Census data) are quite timely: only about 25 percent of GDP is estimated using trend-based data for the first (or, as BEA calls it, the advance) estimate of GDP. That estimate is available approximately 1 month (25 days) after the end of the quarter, whereas sufficient source data for the first GDI estimate are not available until 2 months after the end of the quarter (3 months in the case of the fourth quarter). Also, the share of trend-based data for the first estimate of GDI is 63 percent, significantly higher than even the third GDI estimate.

However, the most important advantage of the GDP source data is the ability to develop an integrated benchmark for the GDP estimates once every 5 years based on detailed, high-quality data from the Economic Census. Equally important, the monthly and quarterly Census Bureau data are conceptually consistent with the definitions used by the Census Bureau for their every-5-year benchmark and annual data. Although this consistency does not provide clear evidence that GDP is closer to "true" production, for many users the conceptual consistency of the monthly, quarterly, annual, and every-5-year Economic Census data is a major source of comfort.

THE CYCLICALITY OF THE LATEST ESTIMATES.  Nalewaik's conclusions on the superiority of GDI rest mainly on his reading of the source data. As suggested above, a careful and detailed analysis suggests that the source data for the early GDI estimates are significantly weaker than those for the early GDP estimates. Nalewaik argues that the benchmarking procedures and the extrapolation of services make the GDP estimates too smooth, but he discounts the likelihood that the failure to fully remove capital gains and losses makes GDI too cyclical. Yet firms do seem to have the ability to time their receipts, expenses, and recognition of unusual losses in ways that would overstate the cyclicality of recorded profits relative to underlying economic activity. Firms may tend, for example, to recognize unusual losses when the overall economy and competing firms' sales and profits

5. As a result of a multiyear Census Bureau initiative to expand its services surveys, through new quarterly and expanded services, BEA is making substantial progress in improving the GDP source data for services. The Census Bureau's plans call for completing its program to provide complete coverage in the quarterly and annual services surveys by 2012.

are down and the losses are likely to have a smaller effect on investor perceptions and stock prices.

It is also known that quarterly wage data have included capital gains in the form of stock options, which were not taken out until the annual revisions, when profits data that excluded them were available. It also seems plausible that the misreporting adjustment that BEA applies to IRS data is countercyclical, yet BEA's misreporting adjustment is proportional and varies little over time, implying that measured income would be too cyclical.

Nonetheless, there is probably something to both sides of the argument over the cyclicality of GDP and GDI. Indeed, to address the issue, BEA is taking steps, outlined below, to improve the accuracy of services estimates through the incorporation of new quarterly services data in GDP, and to improve the early wage and salary estimates in GDI. Corporate profits will remain an issue, but this work by Nalewaik on the cyclicality of the statistical discrepancy suggests directions for future research.

THE RELATIONSHIP OF GDI AND GDP TO COINCIDENT MEASURES OF ECONOMIC ACTIVITY. Nalewaik shows that the cyclical measures that he examines move more closely in tandem with GDI than with GDP. Part of that closer correspondence relates to his use of income-type variables that either are used to measure GDI or can be seen as proxies for income-side measures of GDP. An examination of a broader set of cyclical measures, such as retail sales and manufacturing sales, shows that GDP has a closer correlation over 1984–2009 with these variables than does GDI. This partly reflects the fact that these variables either are used to measure GDP or can be seen as proxies for product-side measures of GDP. Moreover, an examination of nine business cycle measures—nonfarm employment, private services payroll, manufacturing employment, nonmanufacturing employment, personal income less transfers, industrial production, manufacturing sales, retail sales, and the nonmanufacturing ISM index—shows that their correlations with GDP and GDI are very similar, with only a slightly larger correlation for one or the other.[6] The average correlation of these variables

---

6. The correlations are as follows: nonfarm employment with GDP, 0.73, with GDI, 0.77; private services payroll with GDP, 0.69, with GDI, 0.73; manufacturing employment with GDP, 0.68, with GDI, 0.74; nonmanufacturing employment with GDP, 0.68, with GDI, 0.67; personal income less transfers with GDP, 0.65, with GDI, 0.76; industrial production with GDP, 0.50, with GDI, 0.60; manufacturing sales with GDP, 0.75, with GDI, 0.69; retail sales with GDP, 0.65, with GDI, 0.57; nonmanufacturing ISI index with GDP, 0.57, with GDI, 0.66.

with GDP was 0.66, compared with 0.69 for GDI, even though five of the nine variables are income-type variables.

I have not examined the leading index indicators discussed by Nalewaik because they do not seem to be a meaningful measure of the accuracy of either GDP or GDI. Stock prices, the yield curve, and high-yield bond spreads are not, as Nalewaik notes, measures of economic activity but rather leading indicators that are used to try to predict economic activity. As the former custodian of the leading indicators—which have been described as measurement without theory—I can report that they have a less-than-stellar history of predicting GDP *and* the business cycle, especially when examined in real time. The S&P 500 index, for example, may be a good financial indicator, but it has a checkered history as a leading economic indicator.

ACCURACY OF THE GDP AND GDI ESTIMATES IN 1984–2006. My review of most of the evidence marshaled by Nalewaik and a review of the current (rather than the revised, or latest) estimates from BEA suggest similar trend growth and cyclical patterns for the GDP and GDI estimates in 1984–2006. Both GDP and GDI provide very similar estimates of trend growth. Looking at revisions to the GDP and GDI estimates at the time of the comprehensive benchmarks—which are based on the Economic Censuses of 1982, 1987, 1992, 1997, and 2002—one can see that nominal GDP and GDI were revised by an average of 1 to 2 percent, and the growth rates for those 5-year periods were revised by about one-quarter of 1 percentage point.

A number of revision studies have shown that GDP and GDI estimates are both reliable indicators of general economic activity, as defined by whether growth is fast or slow relative to trend, whether growth is accelerating or decelerating, which of the major components are contributing to growth, and trends in saving and other major components of GDP. Dennis Fixler and Nalewaik (2009) have found that the revisions are larger around turning points, and given the degree of extrapolation in both the GDP and the GDI estimates, this makes sense. However, as I show in figures 1 and 2, the general patterns exhibited by the early estimates of both GDP and GDI are quite similar. In the last three business cycles, both early estimates show roughly the same peak, slowing pattern, trough, and recovery pattern. Most of Nalewaik's figures look at the differences in the revised GDP and GDI data, and except for 2007, the revised, or latest, data also show the same general cyclical patterns for GDP and GDI (figure 3).

ACCURACY OF THE GDP AND GDI ESTIMATES OVER 2007–09. The estimates of GDI and GDP for 2007–09 show a much larger cumulative drop in GDI than in GDP: GDI declined 4.1 percent from the fourth quarter of 2006 to

**Figure 1.** Changes in Real GDI and GDP, Latest-Release Estimates, 1984–2009[a]

Percentage points
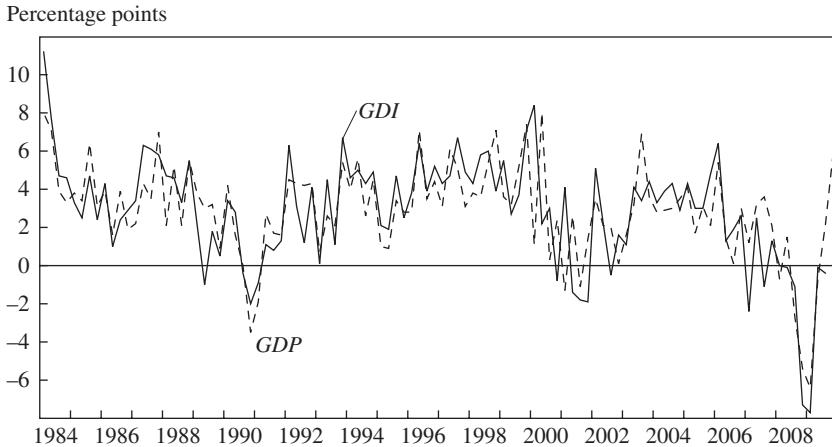


Source: Bureau of Economic Analysis.
a. Quarterly data, annualized.

the second quarter of 2009 (the trough in both GDP and GDI), while GDP declined 1.2 percent. Almost all of the cumulative difference occurs in 2007; between the relative peak in the fourth quarter of 2007 and the second quarter of 2009, the declines in GDP and GDI were much closer, with GDP declining 3.7 percent and GDI 4.1 percent.

**Figure 2.** Changes in Real GDI and GDP, Third-Release Estimates, 1999–2003[a]

Percentage points
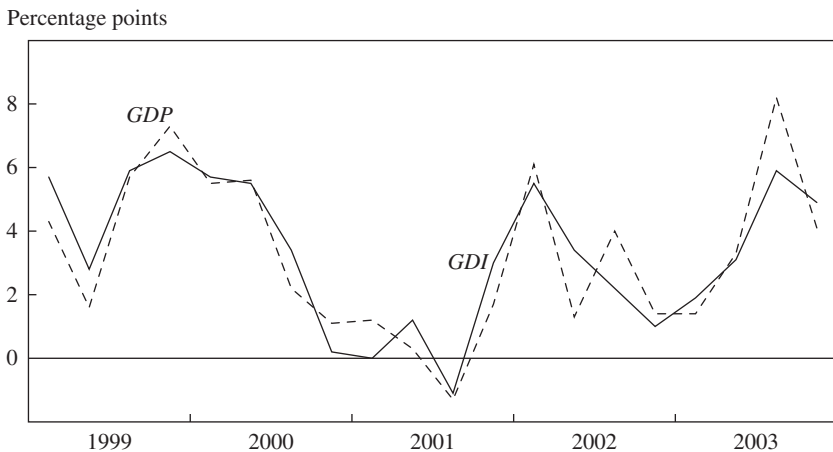


Source: Bureau of Economic Analysis.
a. Quarterly data, annualized.

**Figure 3.** Changes in Real GDI and GDP, Third-Release Estimates, 2006–09[a]

Percentage points
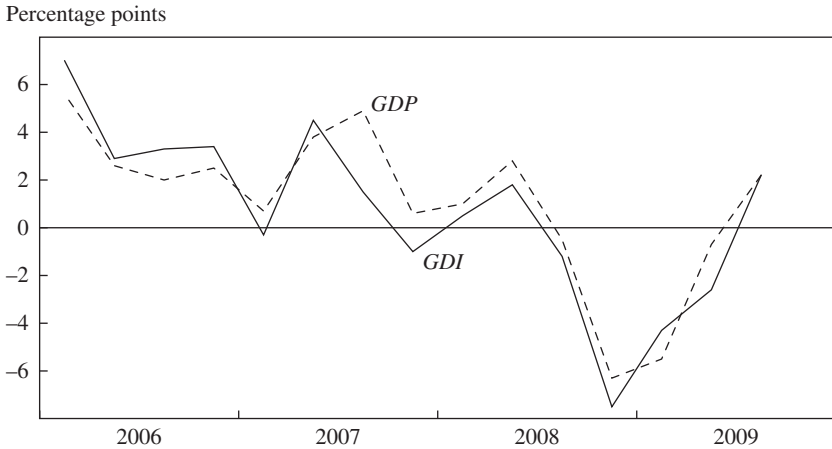


Source: Bureau of Economic Analysis.
a. Quarterly data, annualized.

Although both GDP and GDI growth began to show signs of weakness in 2007, with quarters of sharply lower and even negative growth, the latest estimates show GDI increasing only 0.1 percent over the four quarters of 2007, while GDP increased 2.5 percent. The main source of the slowdown in GDI comes from profits and proprietors' income, both of which declined in 2007. During 2007, compensation, including wages and salaries, continued to grow. This growth in compensation in GDI seems consistent with the 2.5 percent growth in GDP, which in turn seems consistent with the 0.8 percent growth in employment. Also, the residual growth in productivity (as measured by GDP per employee) of about 1.7 percent does not seem so high as to suggest that GDP growth was overestimated relative to employment (average productivity growth since 1995 averaged 2.5 percent).

The difference between the GDP and GDI estimates for 2007 (and early 2008) seems to turn on the accuracy of the profits and proprietors' income data. As discussed above, converting IRS data and financial report data to a national income and product accounts basis is extremely difficult, especially during periods of rapid change in markets. Firms have the ability to adjust the timing of their expenses and receipts and when they recognize unusual gains and losses. Tax law changes and changes in the economy can also affect the consistency of profits over time. The year 2007 marked the beginning of the financial crisis, and although BEA does its best to exclude unusual gains and losses, profit estimates for banks and other financial insti-

tutions were particularly challenging during and after the crisis. Estimating proprietors' income is challenging as well. IRS studies suggest that for each dollar reported to the IRS, another dollar is not reported. BEA therefore roughly doubles the annual estimate reported by the IRS. Unfortunately, only two comprehensive IRS estimates of underreporting have been published in the last 25 years: the 1988 IRS Taxpayer Compliance Measurement Program and the 2001 IRS National Research Program. So BEA's doubling may capture the long-run trend in compliance, but it may not be appropriate at times of significant change in the economy, and as noted above, it may cause measured income to be too cyclical.

It is also worth noting that the National Bureau of Economic Research, using data on employment, sales, and a number of other cyclical indicators including GDP, placed the cyclical peak in December 2007. This cyclical dating then counts most of 2007 as a period of expansion, which is consistent with growth in GDP, but not with the flat-to-declining pattern of GDI.

Overall, my reading of the behavior of the GDP and GDI data for 2007 and early 2008 suggests that GDP looks more consistent with the behavior of employment and unemployment than GDI. However, both sets of estimates will see further revisions, so interested readers should stay tuned.

SHOULD WE AVERAGE GDP AND GDI? Averaging GDP and GDI, using the two-to-one weighting recommended by Nalewaik, should produce an estimate of output that changes less in subsequent annual revisions than GDP has in recent years, but the statistical gain would not be large, on average, relative to the average revision. Moreover, the value of averaging must be weighed against two disadvantages: first, that of having larger revisions between the advance and the second estimate (when the income data are introduced), and second, the risk of having anomalous revisions that could reduce confidence in the overall accuracy of the national accounts.

Although the use of GDI and other real-time data may be able to reduce revisions to the early output estimates, the relative gain is likely to be small. Based on the estimates in the paper, the use of GDI over 1994–2006 would have reduced the mean absolute revision in the early GDP estimate by 0.2 percentage point, but the mean absolute revision to GDP over this period was 1.25 percentage points.[7] That is not insignificant, but the relative size, along with the fact that the early GDI and GDP estimates present a similar picture of the business cycle, needs to be considered in any proposal to produce a weighted average. Also, users of economic data may

---

7. Based on a replication of Nalewaik's equation, the mean absolute revision in 1994–2006 to GDP would be reduced from 1.25 to 1.18 percentage points.

perceive problems with an average GDP growth rate that is not consistent with the growth in the subcomponents for either GDP or GDI.

In the past, BEA has presented the idea of averaging the estimates to its key users. BEA's Advisory Committee, the Federal Reserve Board, and other users of the national accounts have consistently told us that if we want to balance GDP and GDI, we should continue to publish separate estimates of both, along with the statistical discrepancy, and then produce a balanced set of accounts that allocate the discrepancy using a replicable, statistically based method. BEA has been working on a methodology for balancing the input-output and industry accounts, but we do not think it is feasible to develop balanced quarterly GDP and GDI accounts. However, BEA will explore means of better presenting and highlighting the GDP and GDI estimates in ways that meet the differing needs of the various users.

BEA will also continue to work with the BLS, the Census Bureau, and the IRS to improve the source data for both GDI and GDP. Incorporation of the next steps in the expansion of the quarterly services survey should continue to bring significant improvement in the source data for GDP. Recent efforts by the BLS to collect data on all types of income—including bonuses, stock options, and other irregular payments—were unsuccessful, but the recent incorporation of the new BLS data on wages and salaries for all workers should significantly improve the estimates of wages in GDI. Finally, legislation now in prospect that would allow BEA, the BLS, and the Census Bureau more consistent access to tax data could be very helpful in reconciling the large differences between financial and tax accounting data.

### REFERENCES FOR THE LANDEFELD COMMENT

Fixler, Dennis J., and Bruce T. Grimm. 2002. "Reliability of GDP and Related NIPA Estimates." *Survey of Current Business* 82, no. 1 (January): 9–27.

———. 2006. "GDP Estimates: Rationality Tests and Turning Point Performance." *Journal of Productivity Analysis* 25, no. 3: 213–29.

Fixler, Dennis J., and Jeremy J. Nalewaik. 2009. "News, Noise, and Estimates of the 'True' Unobserved State of the Economy." Working paper. Washington: Bureau of Economic Analysis.

**GENERAL DISCUSSION** Christopher Sims asserted that the relevant issue is whether GDP(I) or GDP(E) is better by itself as an indicator, and not whether putting GDP(I) or GDP(E) into a factor model makes a difference in the outcome, as Francis Diebold had sought to demonstrate. If GDP(E) were simply GDP(I) plus noise, then the result would be the same, because Diebold is extracting the nonnoise part of GDP(I) and GDP(E). Sims

claimed that what is of real interest is which measure, GDP(I) or GDP(E), does most of the work in the factor model. Traditional aggregates, which are in a sense informal factors, can come close to extracting the main business cycle factor. It would be interesting to test a bivariate factor model to see whether an underlying factor in both GDP(I) and GDP(E) is closer to one or the other, and to ask how close GDP(I) or GDP(E) comes by itself. Francis Diebold responded that one could indeed treat GDP(E) and GDP(I) as indicators and extract a factor from them, but that that is a different exercise, in no way superior to or more appropriate than the results he had reported, which answer different but equally important questions.

Robert Hall noted that the press release from the NBER Business Cycle Dating Committee announcing its determination of the December 2007 business cycle peak said specifically that the committee consulted real gross domestic income. Clearly its usefulness is not news to the committee. The committee also rejected an approach to defining real activity that mixes a number of indicators together, because the mix always overweights manufacturing. As the manufacturing share of GDP has declined, it is important to avoid what would become substantial double counting. Hall further reported that the notional set of indicators that the committee listed does not reveal the weights the committee applies. The committee is focused on finding two things: the best possible measures of output and aggregate employment. The committee looks at the modern economy mainly with economy-wide, not sectoral, measures.

Hall found the paper persuasive on the point that the best early estimate of output ought to use a lot of highly relevant variables. The quality of the early estimates of both GDP(E) and GDP(I) could be improved by giving weight to private forecasts as well as the early data available to the BEA. Because the government might be squeamish about releasing an output estimate that relies mainly on forecasts and correlations, the BEA ought to consider leaving the close-to-real-time estimates—the nowcasting—to others. At a minimum, users of the BEA's early estimates should be consulting private nowcasts as well.

Phillip Swagel thought the paper raised a basic question about the very nature of a recession. Consider the debate over unemployment versus output measures: if real GDP growth were 1 percent for a considerable period, there would surely be net job losses. Would that be a recession, or not? The question is, What does one hinge the start and end dates on: measures of output or measures of the labor market?

Matthew Shapiro seconded Sims's comment that it would be useful to have the bivariate factor model calculate the optimal portfolio weights of

GDP(E) and GDP(I). He felt Diebold's illustrative model was misleading because it assumes that all the variance is error. If most of the variance were signal, the results might be quite different. The right weights will depend on the relative amounts of signal and noise in the two data series. Philip Howrey had done something similar 20 years ago, attempting to assign weights to the Bureau of Labor Statistics' household and establishment surveys. His analysis put about 80 percent of the weight on the establishment survey, which has now become the conventional wisdom.

Jan Hatzius noted that before the BEA reported estimates of GDP(E) and GDP(I) for the fourth quarter of 2009, the Federal Reserve's Flow of Funds tables already included an estimate of the fourth-quarter statistical discrepancy between the two, which shows an increase of about $130 billion over the previous quarter. That implies an estimate of real GDP(I) growth for the fourth quarter of just over 2 percent annualized. Hatzius was curious about how much, if any, weight should be put on that. He also observed that many analysts are concerned at the moment about the deviation between the performance of large firms and that of small firms, and what that might mean for preliminary estimates of GDP. Which of the two GDP measures is more vulnerable to that deviation? Finally, Hatzius wondered why other countries put more weight on income-based measures. Is it because they have different data sources, with different strengths and weaknesses than their U.S. counterparts, or do they simply reach a different conclusion about how important it is not to confuse the public, for example by taking an average of different measures?

Robert Gordon did not accept the characterization of the debate over output versus employment in the business cycle dating context as a tug-of-war between proponents of one or the other. Rather, he saw it as an econometric problem, one that involved studying the breakdown of changes in output and the output gap into their components, starting from the simple identity that output is equal to aggregate hours times output per aggregate hour, that is, aggregate productivity.

Gordon also acknowledged that the Okun's law relationship is quite different today from what it was in the mid-1980s, with unemployment becoming much more responsive to output than in the original formulation. Whereas in Okun's day aggregate hours responded by two-thirds of any movement in the output gap, today hours respond more than one for one. Further, the Great Recession witnessed departures from this relationship, with productivity growing faster and hours falling further than even the post-1986 equation would have predicted.

Gordon went on to point out that Nalewaik's figure 6, which shows the behavior of real GDP(I) and GDP(E) estimates over the most recent recession, reduces the residuals. In fact, the erroneous division of the equation between productivity and hours is almost eliminated. But this happened before the recession started; it is mainly a story about 2007. Gordon found it reassuring to have at least a partial explanation of why productivity had looked so good in the last 2 years.

Benjamin Friedman encouraged making a sharper distinction between two conceptual questions. The first is whether output or employment is the more meaningful concept for judging turns in business cycles. The second is which of the two statistics, GDP(E) or GDP(I), does a better job of measuring what we understand by output. The second question arises only because the statistical agencies use double-entry bookkeeping, which in a world of imperfect measurement necessarily leads to discrepancies. Even in a world of perfect measurement, where GDP(I) and GDP(E) are always identical, the first issue would still be a question, but the issue addressed by the paper would go away.

Friedman noted that the paper showed a very strong historical correlation between the statistical discrepancy between the two output measures and unemployment, which suggests that the question of what is going either unmeasured or mismeasured that gives rise to the difference is not just about, for example, the superiority of one or another source of data, but rather involves substantive questions of economics. He encouraged further analysis of what these measurement problems are. One might think the discrepancy is just noise, but the correlation he cited shows that it is not pure noise. Steven Davis added that the correlation suggests that the discrepancy is cyclically varying, which implies that determining the optimal weights for an average is more complicated.

Davis also remarked on Steven Landefeld's discussion of the administrative record inputs to the GDP(I) side, each of which is somewhat different in scope. Recognizing that making adjustments for these differences is challenging, he hoped that the BEA would drill down deeper and investigate the extent to which these administrative data sources line up when the discrepancies are more fully taken into account. There are potentially tremendous advantages to relying on administrative records when possible; their comprehensive nature creates much greater opportunities for disaggregation by type of activity, location, and other dimensions.

Steven Braun complimented the BEA for showing its dirty laundry, in the sense of making its best estimates of both GDP(E) and GDP(I) available so that economists could analyze the statistical discrepancy. Oral tradition

among analysts of BEA data has it that before 1980, the BEA managed the statistical discrepancy and did not allow it to change very much. Braun also said that what he would most like to see changed in the federal statistical system was not anything that the BEA does, but rather the way the BLS publishes its productivity data, preferring that the numerator for the productivity calculation be a weighted average of the two output measures.

Justin Wolfers expressed concern about Diebold's approach, on the grounds that the usefulness of any data is in what they reveal about the underlying real factor. When looking for the causes of the current recession, it matters for the diagnosis whether it appears that labor productivity was growing or falling. As Braun had argued, it is a question of getting the numerator right, and it has real economic significance.

Wolfers went on to make a plea for Landefeld to give the paper another chance. He thought the concern over users being confused by an average of GDP(E) and GDP(I) was overblown: sophisticated users are already taking an average. Moreover, when the choice is framed as one measure or the other, it is all too easy to regard whichever choice the administration's economists make as political. In any case, if the BEA is determined to rely on only one measure, all the metrics in the paper except one say it should be the income-based measure.

Alan Blinder observed that the paper shows that the GDP(E) revision, which previously had been thought not to be forecastable, is in fact forecastable by GDP(I). This alone, in his view, was a sufficient reason for the BEA to use it.

Steven Landefeld, responding to some of the concerns raised, noted that Europe uses GDP(I), probably for lack of the kind of sources available in the United States for expenditure-side data. Europe is now in the process of developing a set of monthly indicators for sales and output. He also commented that the gain in accuracy from "nowcasting" is relatively small. It does not dramatically change things. With respect to averaging, what some users, including many members of the BEA's advisory committee, have requested is, rather than a simple average, a statistical methodology that shares it out to components, creating a consistent picture. The BEA is working on something similar in its industry accounts.

Landefeld added that from his experience as the custodian of the index of leading indicators and previous research, he did not expect to find the use of cyclical indicators very helpful in improving the accuracy of GDP(E) or GDP(I), but that the BEA would continue to research and attempt to address the sources of the statistical discrepancy over the business cycle. He also noted that the BEA intended to better highlight GDP(I) in its reports.