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 analyses 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 comparative 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 substantiate 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 perspectives that GDP(I) may be superior to GDP(E). That is initially surprising—

1. For helpful comments I thank the participants at the Brookings Panel conference, especially Robert Hall, Christopher Sims, and Justin Wolfers. For research support I thank the National Science Foundation and the Real-Time Data Research Center at the Federal Reserve Bank of Philadelphia.

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):

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

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 λ^* is

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

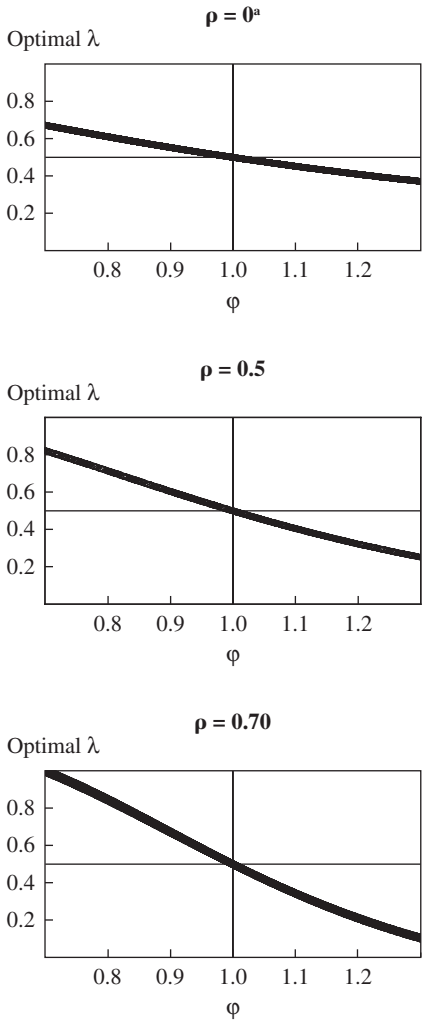
where $\varphi = \mathrm{var}(e_{\mathrm{GDP(E)}})/\mathrm{var}(e_{\mathrm{GDP(I)}})$, $\rho = \mathrm{corr}(e_{\mathrm{GDP(E)}}, e_{\mathrm{GDP(I)}})$, $e_{\mathrm{GDP(E)}} = \mathrm{GDP} - \mathrm{GDP(E)}$, and $e_{\mathrm{GDP(I)}} = \mathrm{GDP} - \mathrm{GDP(I)}$. It is natural and desirable that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

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

The key observation is that, except for extreme values of φ or ρ, 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 $\mathrm{var}(e_{\mathrm{GDP(E)}})$ is 10 percent greater than var $(\mathrm{e}_{\mathrm{GDP(E)}})$, and that $\rho = 0.5$, that is, that $e_{\mathrm{GDP(E)}}$ and $e_{\mathrm{GDP(I)}}$ 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 ρ 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)



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.

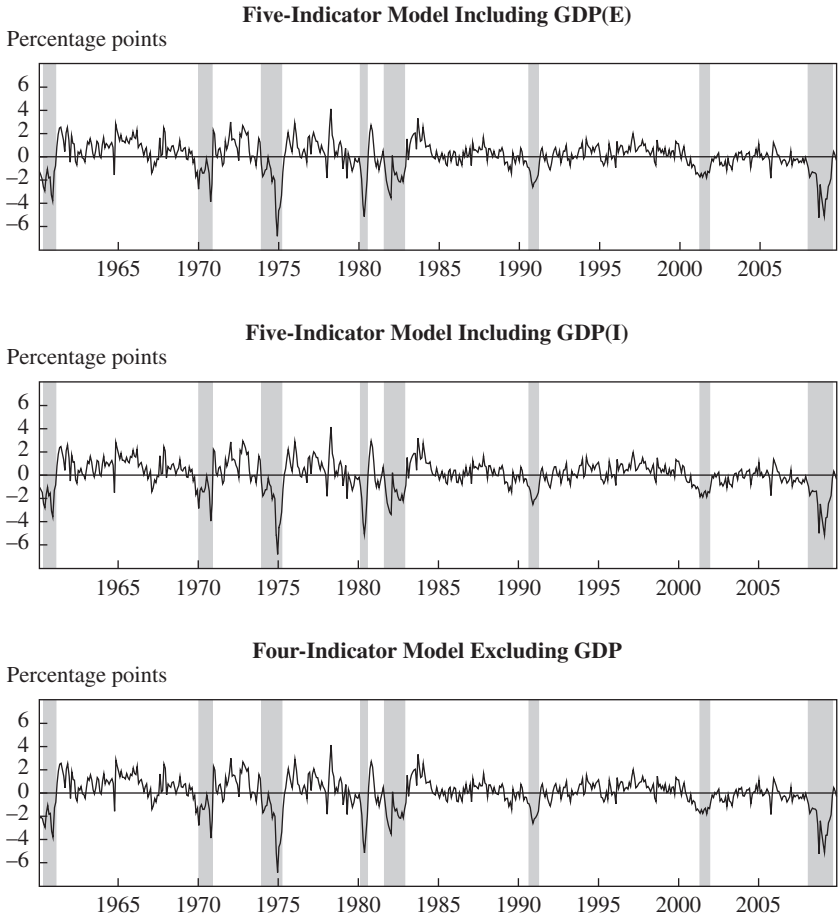
Does the choice of GDP(E) or GDP(I) matter for business cycle measurement, which, as I have emphasized, involves monitoring not only GDP but also a variety of other business conditions indicators? I will address this question using a five-variable dynamic factor model nearly identical to that of the 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



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 nonfarm 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.

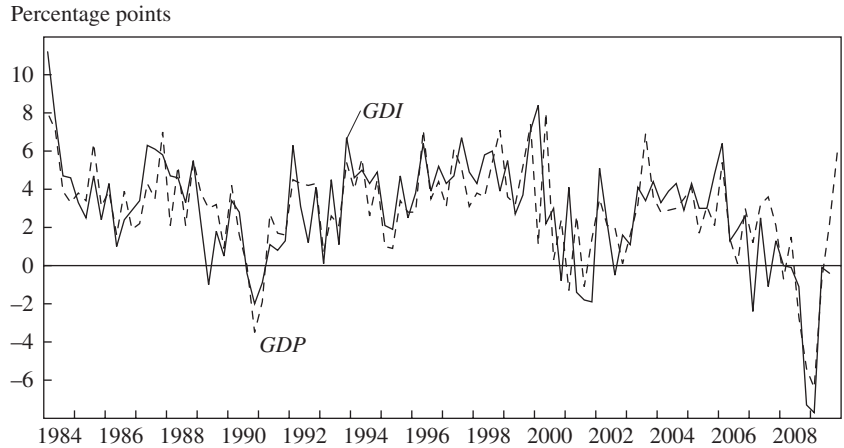
I have not examined the leading index indicators discussed by Nalewaik because they do not seem to be a meaningful measure of the accuracy of either GDP or GDI. Stock prices, the yield curve, and high-yield bond spreads are not, as Nalewaik notes, measures of economic activity but rather leading indicators 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

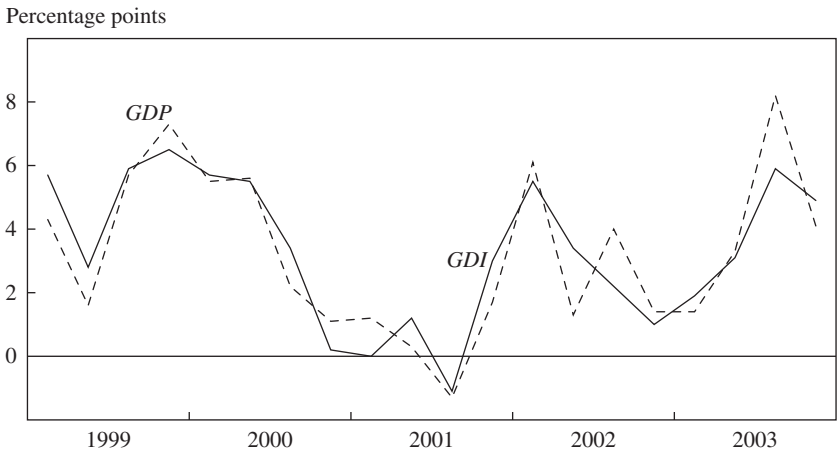
Figure 1. Changes in Real GDI and GDP, Latest-Release Estimates, 1984–2009[a]



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

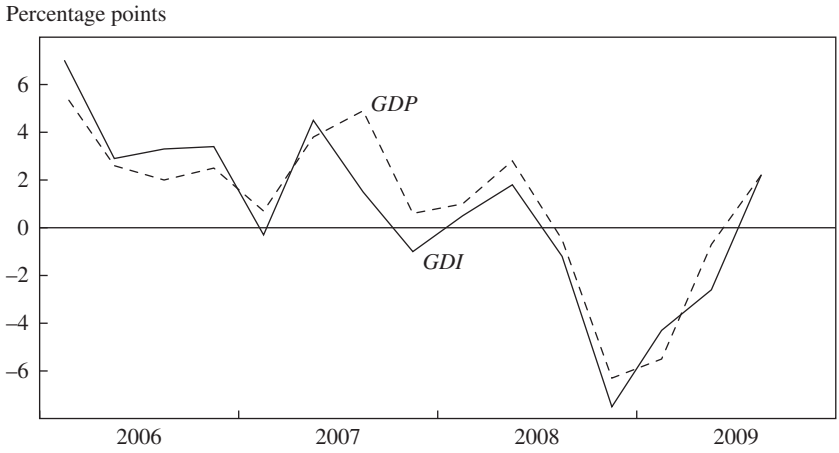
the second quarter of 2009 (the trough in both GDP and GDI), while GDP declined 1.2 percent. Almost all of the cumulative difference occurs in 2007; between the relative peak in the fourth quarter of 2007 and the second quarter of 2009, the declines in GDP and GDI were much closer, with GDP declining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

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



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

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



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.