Is GDP or GDI a better measure of output? A statistical approach.*

Ryan Greenaway-McGrevy
The Bureau of Economic Analysis

February 7, 2011

Abstract

Gross domestic product (GDP) and gross domestic income (GDI) are in theory estimates of the same concept, namely economic production over a defined span of time and space. Yet the two measures are compiled using different source data, and the two measures often give different indications of the direction of the economy. This raises the issue of which of the two measures is a more accurate estimate of economic production. In this paper we present a time-series statistical framework for addressing this issue. Our findings indicate that the latest vintage of GDP has been a better measure of true output over the 1983-2009 period than the latest vintage of GDI. Our model also implies an optimal weighting of GDP and GDI can yield a more accurate estimate of economic output than either GDP or GDI alone. Our empirical findings indicate that a weighting of approximately 60% to GDP yields the best estimate for the 1983-2009 period. When we consider vintages of estimated output, we find that GDI often contains additional information to GDP regarding true output.

Keywords: GDP, GDI, statistical discrepancy, signal-to-noise ratios, optimal combination of estimates, business cycles

JEL Classification: C1, C82.

*The views expressed herein are those of the author and not necessarily those of the Bureau of Economic Analysis or the Department of Commerce. I wish to thank Parker Sheppard and Mikhael Lal for their dedicated research assistance, and Dennis Fixler and Bruce Grimm for their valuable thoughts and comments.

1 Introduction

Is Gross Domestic Product (GDP) a better measure of economic output than Gross Domestic Income (GDI)? GDP and GDI are constructed using different source data but are in theory estimates of the same concept, namely the output of the economy over a given period of time. Yet the growth rate in GDP often gives a different picture of the growth in the economy to the growth in GDI (see, e.g., Grimm, 2007; Nalewaik, 2010). While GDI growth has exhibited more volatility than GDP growth (see table 1 below), the average growth rate in GDI has been slightly less than GDP growth over the 1983-2009 period.[1] As noted by Fixler and Nalewaik (2007) and Nalewaik (2010), lower variation in GDP growth does not in itself indicate that GDP is a better measure of true output.

Table 1: Output growth summary statistics; 1983-2009

	mean	variance	correlation with GDI
GDI	5.410	10.144	-
GDP	5.456	7.984	0.810

Figure 1 below depicts GDI and GDP growth over the 1983-2009 period. Throughout the relatively mild 1991 recession GDI growth remains positive, whereas GDP growth remains positive throughout the similarly mild 2000 recession. The more recent 2008 financial crisis and associated recession saw a much more profound collapse in GDI growth than GDP growth.

Given that both GDP and GDI can give substantially different pictures of the direction of the macroeconomy, whether GDP or GDI is a better measure of output growth is of interest to policymakers and analysts alike. Yet it is difficult to assess which measure is more accurate because true output is inherently unobservable. While we may have a sound basis for thinking that later vintages of a given measure of output are more accurate than earlier vintages (for example, as time passes the source data for the given measure become more complete), there is little obvious justification for thinking that the latest vintage of either GDP, GDI or another observable variable, is true output. Thus our question cannot be answered by comparing the distance the given output measures and a given observable. Instead, in order to answer this question the researcher must rely on other methods of inference for the determining the accuracy of a given estimate. Several inferential frameworks have been employed in the past.

[1] Throughout, growth rates are the log-difference current dollar quarterly GDP and GDI, multiplied by 400.

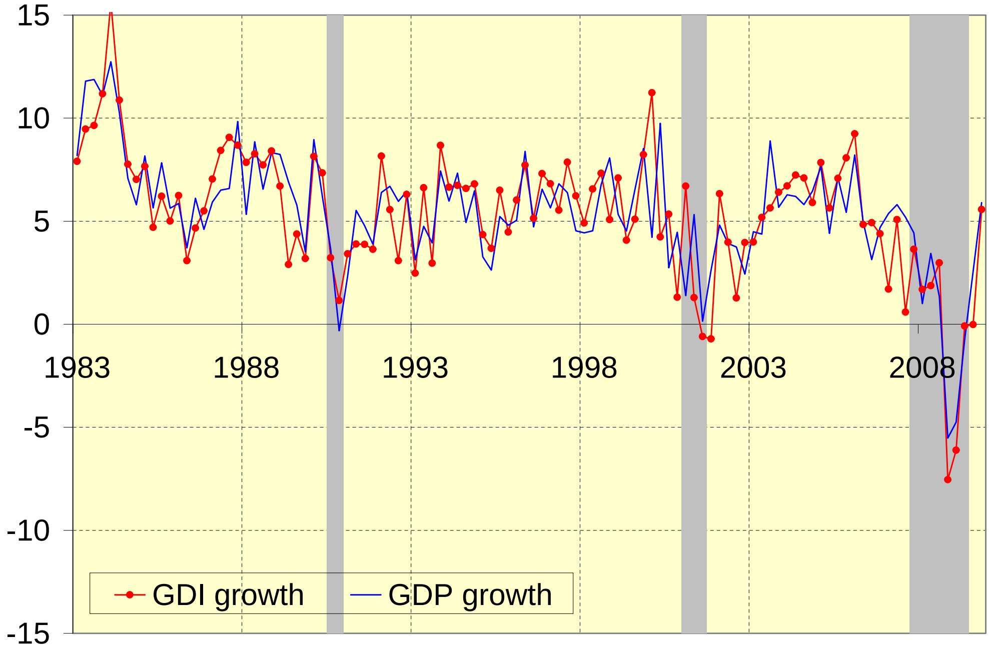

Figure 1: Quarterly GDP and GDI growth 1983-2009. Darker areas indicate NBER recessions.

First, we may evaluate the quality of the source data used to construct the estimates. In this regard, the source data for the early GDI estimates are considerably less complete and consistent than the source data for the early GDI estimates (Landefeld, 2010). About 86 percent of the early GDP estimates are based on some form of direct monthly or quarterly source data, while only 37 percent of the early GDI estimates are based on monthly or quarterly source data. However, these criticisms do not apply to the later vintages of GDP and GDI, as the source data for both measures become more complete with the passage of time. However, Landefeld (2010) also notes that the source data for GDP is gathered from statistical surveys with consistent definitions tailored to national accounting purposes, whereas GDI estimates are based in part on "administrative data" published by firms for the purposes of regulatory compliance. We refer the reader to Landefeld (2010) for a more detailed discussion of these and related issues.

Alternatively, we may compare the measures of output with other economic indicators, such as the unemployment rate, or other macro-economic events, such as the beginning and end of recessions. Nalewaik (2010) shows that GDI is typically more highly correlated with unemployment, stock returns and interest rates than GDP. Nalewaik (2010) also shows that the last few downturns in economic activity have shown up in GDI sooner than in GDP, while Grimm (2005) shows that the sign of GDI growth lines up better with NBER recession dates than the sign of the GDP growth rate. Yet one weakness of these approaches is that they require either a modelling assumption or a stylized fact to describe the relationship between the indicators and unobserved output. For example, Nalewaik (2010) uses a linear regression model to describe the relationship between output growth and various economic indicators. This specification is justified under a given economic model. For example, Okun's law states that the growth rate of output is linearly related to the growth rate in unemployment. Indeed, Nordhaus (2010) makes this assumption explicit in his analysis. Yet our point is that the researcher must make a stand on the specificity of economic model adopted, without being able to test empirically whether the model is correctly specified. (In order to test the model, we would have to observe true output.) The analysis then rests on the validity of a modelling assumption that can never be verified. To return to our example, if Okun's law does not hold, then the unemployment rate cannot be used as an instrument to infer true output. Similarly, if we compare GDP and GDI to recession dates, we presuppose that recessions are dated using true output as the sole indicator, which is not the case. For example, the NBER business cycle committee "examines and compares the behavior of various measures of broad activity: real GDP measured on the product and income sides, economy-wide employment, and real income" (see http://www.nber.org/cycles/recessions.html).

In this paper we take a different approach to evaluating whether GDP or GDI is more accurate. Although the focus of policy-makers, analysts and the media is typically on the *change* in output, we begin by considering GDP and GDI in *levels*. We show that although the difference between the latest vintage of GDP and GDI in levels - the "statistical discrepancy" - can be large and persistent over time, the difference between the GDP and GDI is a mean-zero stationary series. In statistical terms the level of GDP and GDI follow the same common trend or equivalently, GDP and GDI are cointegrated. In other words, as GDP and GDI have grown over time, they have never drifted too far from each other. Figure 2 below exhibits the difference between log GDP and log GDI over the 1983Q1-2009Q4 period.[2]

[2]The log transformation converts the exponential trend in real GDP and GDI to a linear trend, thereby circumventing statistical problems associated with linear transformations of exponentnial series, such as upward-trending variance in the difference of GDP and GDI. Throughout the paper, references to GDP and GDI will be in reference to the log level of current dollars estimates of US GDP and GDI.

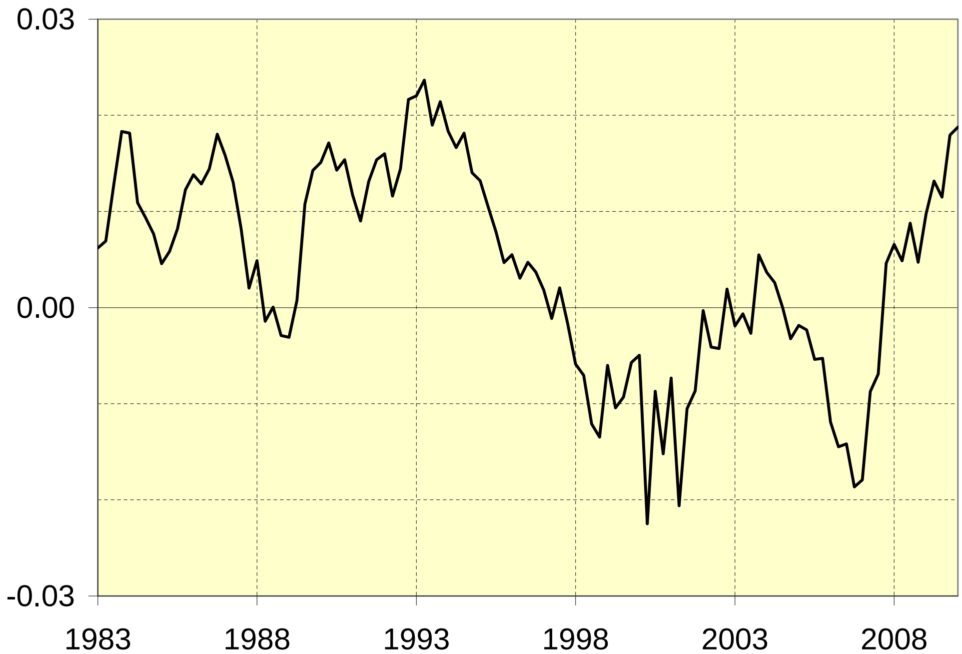

Figure 2: The statistical discrepancy 1983Q1-2009Q4. Latest Vinatges.

We interpret this finding as a robust justification of the assumption that the level of GDP and GDI are measures of the same latent variable, and that this latent variable is true output. Although this interpretation has been used in much of the related literature (see, e.g., Fixler and Nalewaik, 2007; Weale, 1992), the interpretation has typically rested on a conceptual argument rather than on statistical evidence. While GDP and GDI should *in theory* be measures of the same concept, this is not in itself sufficient for GDP and GDI to have been measures of the same concept *in practice*. By showing that GDI and GDP follow the same common trend, we provide statistical evidence that GDP and GDI are in practice measures of the same latent variable. A similar approach is adopted in Patterson and Heravi (2004), who document that different vintages of GDP follow a single common trend. They then interpret this trend as “true” economic production, and use this interpretation as the basis for revisiting the “news” versus “noise” question posed by Mankiw and Shapiro (1984).

Given that GDP and GDI in *levels* are measures of the same latent variable, we can use time-series statistical techniques to decompose *changes* in GDP and GDI into two components: A common component, which we interpret as the change in “true” unobserved economic output;

and measurement error terms, capturing the difference between the change in GDP (or GDI) and the common component. This decomposition can then be used to answer the main question of the paper. Specifically, the output measure that has historically exhibited smaller measurement error is the better measure of output.

Our framework is not beholden to whether the "news" or the "noise" hypotheses of Mankiw and Shapiro (1986) hold for GDP or GDI estimates. In particular, we do not preclude correlation between the measurement error and true output growth, which has been characterized as reflecting the "noise" hypothesis (see, e.g., Fixler and Nalewaik, 2007).[3] The interpretation of our results thus remains agnostic with respect to the validity of the "news" or the "noise" hypothesis.

To estimate the model we use standard Kalman filter methods (Kalman, 1960, 1963). The Kalman filter has been commonly used to estimate a time series of latent variables that are not directly observed by the econometrician. For example, Fama and Gibbons (1982) use the Kalman filter to extract the anticipated inflation rate from the bond market, whereas Stock and Watson (1991) use the filter to extract a business cycle index from a low dimensional vector of economic indicators. In the measurement literature, Chen and Zadroni (2007) and Patterson (1994) apply the Kalman filter to different vintages of GDP in order to produce real-time measurement of output.

We focus on the period spanning 1983Q1-2009Q4. This period coincides roughly with the break in output volatility in the U.S. known as the "Great Moderation" (McConnell and Perez-Quiros, 2000). By focussing on the great-moderation period exclusively we circumvent any structural breaks that would otherwise confound our results.[4] In addition, the more recent accuracy of GDP and GDI is likely to be of more interest to policy-makers and analysts.

We find that over the 1983Q1-2009Q4 period, GDP has been a stronger signal of latent output growth than GDI: The measurement error of GDP is smaller - in terms of both bias and variance - than the measurement error of GDI. Specifically, the average squared GDI measurement error is approximately 1.7 times greater than the average squared GDP measurement error. This is strong evidence in favor of GDP being a more accurate measure than GDI, and it is verified by more rigorous statistical tests of accuracy.

Having established our main result, we pursue some extensions that may be of interest to analysts and policy-makers.

[3]Our approach thus differs to earlier decompositions of GDP and GDI into measurement error and true growth, such as Weale (1992), in that moment conditions do not impose orthogonality betweeen the estimates of the measurement error and true growth. Indeed, in the results to follow we see that measurment error is correlated with true growth.

[4]For example, Stock and Watson (2002) estimate that the break in output (GDP) volatility occurred between 1983Q2 and 1984Q3.

1.1 The accuracy of earlier vintages of GDP and GDI

Measures of output are subject to periodic revision, and each successive revision or "vintage" should in theory represent a more accurate estimate of output. Our main result is obtained using the "latest" estimate of output. Yet policy-makers are often concerned with more timely estimates of economic indicators. It is therefore of interest to investigate the extent to which the earlier vintages of GDP and GDI, which are available much sooner in terms of calender time than the "latest"estimate, can "predict" unobserved output. We find that in this respect GDI does a marginally better job than GDP in predicting true output growth.

1.2 Combining GDP and GDI

Other research has suggested that some weighting of earlier releases of GDP and GDI may be closer to latent output (see, e.g., Fixler and Nalewaik, 2007, Nordhaus, 2010). The state space decomposition can be used to derive an optimal weighting for such an average. The weights are optimal in the sense that the squared measurement error of the weighted average is minimized with respect to the weights. We show that for the 1983-2009 period, the optimal weight of GDP is 0.60, a result consistent with our finding the GDP was a better measure of latent output over this time period.

The derivation of these weights depends on the covariance of the measurement error in GDP and GDI rather than the covariance of GDP and GDI growth themselves. Thus our weighting methodology is not beholden to either the "news" or "noise" view of output estimate accuracy (see Mankiw and Shapiro, 1986). For example, Fixler and Nalewaik (2007) derive optimal weights based on the directly observable variance and correlation between GDP and GDI growth.

The remainder of the paper is organized as follows. In the following section we document the time series properties of the statistical discrepancy, in particular we show that it is a stationary series. In section 3 we give our state space representation of GDP and GDI, and decompose the two series into measurement error and true output. In particular we show that GDP has less measurement error than GDI. In section four we pursue some extensions of our findings from section three. Section five concludes. Throughout, growth rates will be the log-difference in the level series, multiplied by 400, since we are using quarterly data.

2 Properties of the statistical discrepancy

We define the statistical discrepancy at time t as

$$x_t := y_{\text{GDP},t} - y_{\text{GDI},t}$$

where $y_{\text{GDP},t}$ is the log level of the latest estimate of quarterly GDP at time t, and $y_{\text{GDI},t}$ is the log level of the latest estimate of quarterly GDI. Figure 2 in the introduction exhibits the statistical discrepancy over the 1983-2009 period. Notably the statistical discrepancy appears rather persistent over time, and does not often change signs in the period. To statistically test whether x_t reverts towards zero over the long run, we run the standard Augmented Dickey-Fuller regression model of the form

$$x_t = \rho x_{t-1} + \textstyle\sum_{j=1}^{p} \theta_j \Delta x_{t-j} + u_t, \tag{1}$$

where u_t is a random variable or "shock" that is independent and identical distributed over time, "Δ" denotes the first difference operator, i.e., $\Delta x_t := x_t - x_{t-1}$, and the number of lags p is estimated using information criteria. Note that if $\rho < 1$, then the effect of a given shock u_t dies out over time, so that, in response to the shock, x_t returns to zero in the long run. That is, if $|\rho| < 1$, x_t is a stationary time series. By empirically estimating ρ we can determine whether the statistical discrepancy returns to zero over the long run.

The least squares (LS) point estimate of is 0.926, which indicates a high degree of persistency, and the Schwartz criterion selects the number of lags p to be 2. The associated t-statistic on the null hypothesis that $\rho = 1$ is -1.854, which has an associated p-value of 0.061 under the Dickey-Fuller distribution. Thus we can reject the null hypothesis of a unit root at the 10% significance level. The lack of ability to reject the null at a conventional significance level such as 5% may be due to insufficient power (we have 108 observations). This leads us to consider a unit root test with more power under the alternative hypothesis. We therefore adopt the more powerful Elliot, Rothenberg and Stock (1996) point-optimal test. The test yields a t-statistic of 2.472, which is below the 5% critical value of 3.127. Thus we reject the null of a unit root in the (log-difference) statistical discrepancy at the 5% significance level. Our finding that the statistical discrepancy is stationary implies that the (log level) of GDP and GDI follow a common trend. That is, while both GDP and GDI tend to trend linearly upwards over time, they do not move too far away from each other.[5]

[5] Even if $|\rho| < 1$, the possibility remains that x_t is stationary with a non-zero mean. In this case, GDP and GDI continue to follow the same common trend, it is just that log GDP is higher (lower) on average than log GDI if the mean is positive (negative). However, we fail to reject the null that the mean of x_t is zero at conventional signifi-

The natural interpretation of this common trend is that it is "true" output, since the common trend drives all the low frequency variation in GDP and GDI. The question at hand then becomes: which output measure - GDP or GDI - is more closely associated with that common trend? We now turn to answer this question.

3 A statistical framework for assessing output measure accuracy

Although the level of output is important, it is more often the growth rate in output that receives attention. Because both GDP and GDI are signals of the same common trend, GDP and GDI growth rates are measures of the growth rate of that common trend. We may impose a statistical model on GDP and GDI growth in order to extract an estimate of the growth rate in the common trend.

3.1 State space decomposition

Consider the following "observation equations" describing the growth rates of GDP and GDI.

$$\begin{pmatrix} \Delta y_{\text{GDP},t} \\ \Delta y_{\text{GDI},t} \end{pmatrix} = \Delta y_{\text{TRUE},t} + \begin{pmatrix} \varepsilon_{\text{GDP},t} \\ \varepsilon_{\text{GDI},t} \end{pmatrix}, \tag{2}$$

where $y_{\text{TRUE},t}$ is the common component, or the change in the common trend in GDP and GDI in log levels. In light of the results from the previous section we interpret $\Delta y_{\text{TRUE},t}$ as true output growth. Also $\varepsilon_{\text{GDP},t}$ and $\varepsilon_{\text{GDI},t}$ are measurement errors that capture the deviation of the measure of output growth from true output growth.

The "noise" view of estimate accuracy rests on the proposition that greater accuracy is associated with lower variance of a given output measure (Mankiw and Shaprio, 1986). Yet this view has recently received much criticism (see, e.g., the discussion in Fixler and Nalewaik, 2007, and Nalewaik, 2010). Notably our decomposition given in (2) remains agnostic to the news versus noise proposition. Only if we impose zero correlation between the measurement errors and true output growth would there be a clear negative relationship between the accuracy of a given output measure and its variance, as posited under the "noise" hypothesis. As we will see below, when we estimate the model there is correlation between the estimated measurement errors and estimated true output.

cance levels. We use the Newey-West HAC estimator of the standard error to account for the serial dependence in x_t.

To estimate the model we use the Kalman filter. To operationalize the Kalman filter we must specify a linear filter for the series $\Delta y_{\text{TRUE},t}$, such that the filtered series are independent over time. This filter is referred to as the "state equation" as it describes the evolution of the state variable $\Delta y_{\text{TRUE},t}$ over time. We model $\Delta y_{\text{TRUE},t}$ as an AR(1) process, namely

$$(3) \quad \Delta y_{\text{TRUE},t} = \delta_1 + \delta_2 \Delta y_{\text{TRUE},t-1} + \upsilon_t, \quad |\delta_2| < 1$$

so that true growth $\Delta y_{\text{TRUE},t}$ is dependent on the previous period growth. Under the assumption that the vector $(\varepsilon_{\text{GDP},t}, \varepsilon_{\text{GDI},t})'$ is identically and independently Gaussian, the state variable $\Delta y_{\text{TRUE},t-1}$ can be recursively predicted based on the observed GDP and GDI growth rates. The Kalman filter is a recursion that estimates true output growth in each period based on observed GDP and GDI growth and equation (3). The filter works predicting true output growth in the current period using true output growth in the previous period. It then constructs the mean square error of observed GDP and GDI output growth in the current period. The filter updates the estimate of true growth in the current period by minimizing the mean square error of GDP and GDI growth. The filter recursively estimates the whole time-series of true output growth by recursively updating the estimates until no further reductions in mean square error can be made. We refer the reader to chapter 13 of Hamilton (1994) for a detailed explanation on the implementation of the Kalman filter.

Figure 3 exhibits the estimated output growth, $\Delta y_{\text{TRUE},t}$, using the Kalman filter, alongside the GDI and GDP measures of output growth.

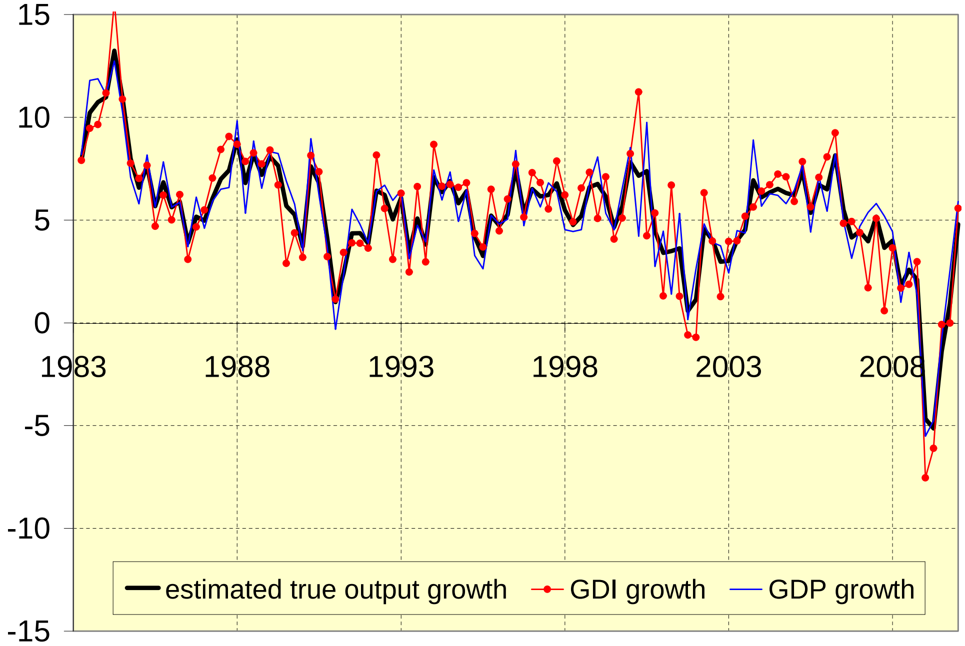

Figure 3: Estimated true output growth 1983-2009.

Estimated output growth closely tracks both GDP and GDI growth over this time period. It appears however that GDI growth is more volatile than GDP growth, which has been noted elsewhere in the extant literature (Fixler and Nalewaik, 2007), as well as above in table 1. In figure 4 we exhibit the past decade only, so that we may observe the small differences between the three series.

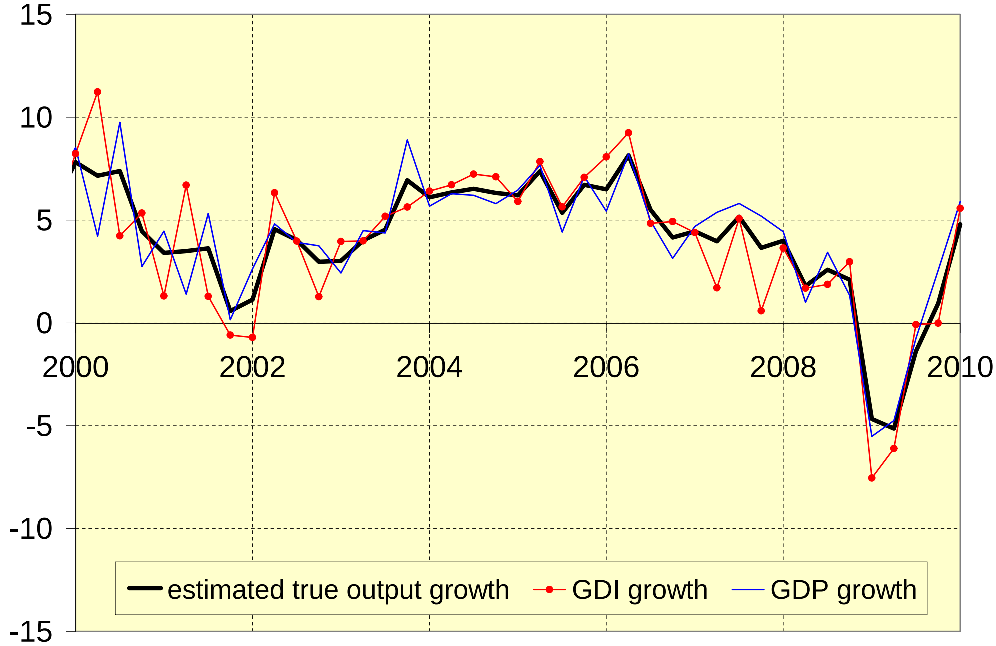

Figure 4: Estimated true output growth 2000-2009.

In general we can see that estimate output lies between GDP and GDI, although this need not be the case for each observation in the entire sample (for example, 2008Q1).

The estimated measurement error terms, namely $\hat{\varepsilon}_{\text{GDP},t}$ and $\hat{\varepsilon}_{\text{GDI},t}$, can be used to tell us which output measure is more accurate. The table below exhibits the summary statistics.

Table 2: Output growth decomposition summary statistics

Kalman filter estimates of equation (2)

	GDI measurement error	GDP measurement error	True output growth
mean	0.017	0.063	5.393
variance	1.443	0.758	6.994
correlation with:			
GDP measurement error	-0.649	1	-
True output growth	0.269	0.051	1

GDI measurement error is the estimate of $\varepsilon_{\text{GDI},t}$, GDP measurement error is the estimate of $\varepsilon_{\text{GDP},t}$, and true output growth is the estimate of $\Delta y_{\text{TRUE},t}$ in (2).

Note that both the absolute value of the mean and the variance of the measurement error in GDI growth is greater than that of the measurement error in GDP growth. Thus the overall squared measurement error is greater for GDI than for GDP, indicating that GDI was a noisier estimate of true output for the 1983-2009 period. However, both measures are strong signals of the underlying true growth component. Comparing the variance of the measurement errors given in table 2 to the total variance given in table 1, we can see that the proportion of the total variation in GDP growth that is attributable to variation in true output growth is about 0.88. The corresponding figure for GDI growth is 0.69. It is also interesting that while GDP measurement error appears largely uncorrelated with estimated output growth (with a correlation coefficient of 0.051), GDI measurement error is substantially positively correlated with estimated output growth (with a correlation coefficient of 0.269).

Another way to measure the distance between measured output growth and the growth rate in the common trend is to run regressions of the form

$$\text{(4)} \quad \widehat{\Delta y}_{\text{TRUE},t} = \beta \Delta y_{\text{GDX},t} + error, \ x = \text{P,I}$$

The R-squared of the linear regression tells us the degree of correlation between the series. The table below exhibits the results.

Table 3: Regression Results

OLS estimation of equation (4)

sample:		GDP growth	GDI growth
1983Q1-2009Q4	$\hat{\beta}$	**0.965**	**0.940**
	R-squared	0.880	0.796
1983Q1-1992Q4	$\hat{\beta}$	**0.970**	**0.976**
	R-squared	0.912	0.841
1993Q1-2009Q4	$\hat{\beta}$	**0.961**	**0.907**
	R-squared	0.835	0.738

Bold face type indicates significance at the 5% level. Standard Errors were computed using the Newey-West method.

Evidently, GDP growth explains more of the variation in the estimated signal than GDI growth. The accuracy of both measures appears to have decreased over the recent 1993Q1-2009Q4 period.

However, in order to formally test whether GDP has less measurement error than GDI, we run a multivariate regression of the form

$$\widehat{\Delta y}_{\mathrm{TRUE},t} = \beta_1 \Delta y_{\mathrm{GDP},t} + \beta_2 \Delta y_{\mathrm{GDI},t} + error, \tag{5}$$

and test whether β_1 is larger than β_2. The table below exhibits the results.

Table 4: Multivariate Regression Results

OLS estimation of equation (5)

sample:	$\hat{\beta}_1$	$\hat{\beta}_2$	R^2	t-stat $(\beta_1 = \beta_2)$
1983Q1-2009Q4	**0.58**	**0.40**	0.96	3.34
1983Q1-1992Q4	**0.58**	**0.40**	0.97	2.18
1993Q1-2009Q4	**0.57**	**0.40**	0.94	2.50

Bold face type indicates significance at the 5% level. Standard Errors were computed using the Newey-West method.

It is clear that in each regression, the coefficient of GDP growth is statistically larger than that on GDI growth, indicating that GDP growth is has more information than GDI.

3.2 Decomposing the statistical discrepancy into GDP and GDI measurement error

Much extant research has focussed on what causes GDP and GDI to be different from each other. For example, Klein and Makino(2000) found that mismeasurement of corporate profits, proprietors' income, exports, and government expenditures explain the statistical discrepancy for the 1947-1997 period. However, when Grimm (2007) replicates the study for the 1984-2004 period, he finds that the statistical discrepancy cannot be attributed to a small set of product- or income- side components.

Our state space decomposition offers another framework for determining what causes that statistical discrepancy. The model allows one the decompose the statistical discrepancy into GDP measurement error and GDI measurement error. To see this, note that by construction the statistical discrepancy is equal to the difference between the measurement error of GDP and the measurement error of GDI. That is for $\tilde{x}_t := x_t - x_0$, we have

$$\tilde{x}_t = y_{\text{GDP},t} - y_{\text{GDP},0} - y_{\text{GDI},t} + y_{\text{GDP},0} = \sum_{s=1}^{t} \hat{\varepsilon}_{\text{GDP},s} - \sum_{s=1}^{t} \hat{\varepsilon}_{\text{GDP},s} =: \hat{e}_{\text{GDP},t} - \hat{e}_{\text{GDI},t},$$

say. The above equation shows that the statistical discrepancy - normalized to be zero at some base period - can be decomposed into cumulated GDP measurement error and cumulated GDI measurement error. Hence, using this decomposition we can attribute the normalized discrepancy $\tilde{x}_t$ to either GDP or GDI. Figure 5 plots $\tilde{x}_t$ against cumulated GDP measurement error ($\hat{e}_{\text{GDP},t}$) and cumulated GDI measurement error ($-\hat{e}_{\text{GDI},t}$).

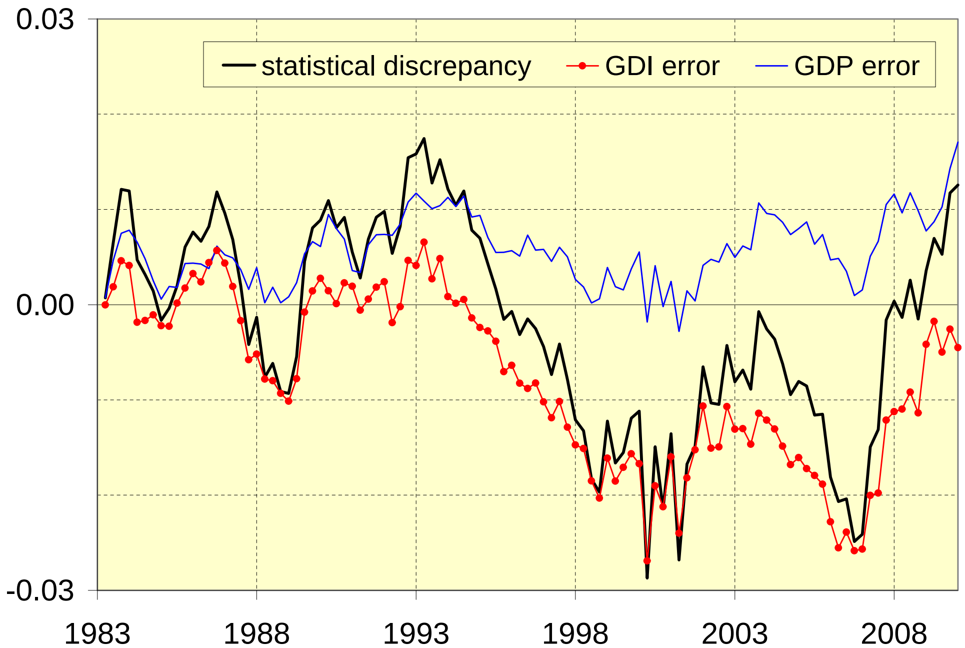

Figure 5: Decomposition of the statistical discrepancy into GDP and GDI measurement error.

It is evident that GDI measurement error more closely follows the statistical discrepancy than GDP measurement error, meaning that most of the variation in the statistical discrepancy is due to the GDI measurement error over the 1983-2009 period. The table below confirms this observation.

Table 5: Summary statistics for cumulated GDI and GDP measurement error

	GDI error	GDP error
mean	-3.144	2.290
variance	12.592	2.070
correlation with statistical discrepancy	0.952	0.652
correlation with cumulated GDI measurement error	1	0.387

cumulated GDI measurement error is $\sum_{s=1}^{t} \hat{\varepsilon}_{\mathrm{GDI},s}$; cumulated GDP measurement error is $\sum_{s=1}^{t} \hat{\varepsilon}_{\mathrm{GDP},s}$; where $\hat{\varepsilon}_{\mathrm{GDI},s}$ and $\hat{\varepsilon}_{\mathrm{GDP},s}$ are the Kalman filter estimates of $\varepsilon_{\mathrm{GDI},s}$ and $\varepsilon_{\mathrm{GDP},s}$ in (2).

Note that the cumulated GDI measurement error is much more highly correlated with the statistical discrepancy than the cumulated GDP measurement error. Moreover, the variance of the cumulated GDI measurement error is about 6 times greater than that of GDP. From this we infer that the majority of the variation in the statistical discrepancy is due to measurement error in GDI.

4 Extensions

Our results from the previous section suggest that GDP is a better indicator of GDI growth. In this section we pursue some extensions of our main result.

4.1 Performance during expansions and recessions

In this section we examine the performance of GDP and GDI at different stages of the business cycle. We split the sample into NBER dated recessions and expansions, and repeat estimation of the regression (4). For the purposes of these regressions, both the quarter of the reported trough and the quarter of the reported peak are included in the recession sub-sample. We do this to increase the number of observations included in the recession sample. The table below exhibits the results below.

Table 6: Recession Regression Results

OLS estimation of equation (4)

sample:		regressor: GDP growth	GDI growth
NBER Recessions	$\hat{\beta}$	**0.79**	**0.75**
(14 observations)	R-squared	0.81	0.77
NBER Expansions	$\hat{\beta}$	**0.97**	**0.95**
(94 observations)	R-squared	0.92	0.85

NBER recession dates: 1990Q3:1991Q1; 2001Q1:2001Q4; 2007Q4:2009Q2.

GDP growth continues to be a better measure of true output growth in both recessions and expansions. These results corroborate Fixler and Grimm (2002), who found that GDP is a more accurate gauge of turning points in the business cycle than GDI. Interestingly, both measures performs worse in recessions relative to expansions. However, this may be due to the fact that there are far less data-points with which to estimate model parameters.

4.2 Assessing the accuracy of earlier vintages

In the previous section we argued that the latest vintage of GDP is a better measure of economic output than the latest vintage of GDI. Yet the latest vintage is itself subject to many revisions, and it can be published by the BEA long after the reference quarter. The BEA's first GDP estimate for the most recent quarter, called the "advance current quarterly" estimate, is released about a month after the quarter ends. The first estimate of GDI appears two months after the reference quarter ends, with the "second current quarterly" release of GDP, except the estimates for fourth quarters, in which GDI first appears with the "third current quarterly" release about three months after the quarter ends. To work with a complete time series of the initial growth rates, we use the "third current quarterly" estimates of GDI and GDP. We also consider the "first annual" vintages, which are released in July of the year following the reference quarter. We also consider the "second annual" and "third annual" vintages, released in the July following two and three years, respectively, after the year of the reference quarter.

In this section we explore how well the earlier releases of GDP and GDI predict our estimate of true unobserved output, $\widehat{\Delta y}_{\text{TRUE},t}$. To achieve this we consider regressions of the form

$$\widehat{\Delta y}_{\text{TRUE},t} = \beta_0 + \beta_1 \Delta y^{v}_{\text{GDP},t} + \beta_2 \Delta y^{v}_{\text{GDI},t} + error \quad (6)$$

where *v* is used to index the various vintages of GDP and GDI. The predictive ability of a given vintage of output is measured using the R-squared of the regression.

Table 7: Predicting true output using GDI and GDP vintages

OLS estimation of equation (6)

vintage	3rd current qtrly	1st annual	2nd annual	3rd annual
$\hat{\beta}_1$	0.111	-0.008	0.035	**0.173**
$\hat{\beta}_2$	**0.284**	**0.457**	**0.407**	**0.338**
R-squared	0.411	0.570	0.441	0.568

Bold face type indicates significance at the 5% level. Standard Errors were computed using the Newey-West method. 1983Q1-2007Q4

Several results are of note. First, the coefficients on GDP growth are not statistically different from zero at conventional significance levels for the 3rd current quarterly, first annual and second annual vintages. In contrast, the coefficient on GDI growth are significant for all vintages. Thus the earlier vintages of GDI growth have been a better predictor of true output growth than the earlier vintages of GDP growth. This result in itself is rather interesting as it suggests that early vintages of GDI contain more information about true output than early vintages of GDP. It supports earlier research that suggests that early vintages of GDI have a significant information content. For example, Fixler and Grimm (2006) find evidence that GDI predicts revisions to GDP**.** Second, later vintages are not always better predictors of true output growth, as we would expect if later vintages were more accurate estimates. The first annual vintage is a more accurate predictor of true output growth than the second annual vintage, and it is marginally better than the third annual vintages. Note however that the second and third annual vintages are better predictors than the third current quarterly vintage.

4.3 Optimal weighting of GDP and GDI

The estimated model can provide optimal weightings for a mean of GDP and GDI growth rates[6]. Define

$$\Delta\tilde{y}_t(\lambda) := \Delta y_{\text{GDP},t}\lambda + \Delta y_{\text{GDI},t}(1-\lambda) = \Delta y_{\text{TRUE},t} + \lambda e_{\text{GDP},t} + (1-\lambda)\, e_{\text{GDI},t}$$

as the average with a weight of $\lambda \in (0, 1)$ on GDP growth. Note that the weightings sum to one by definition (c.f., Fixler and Nalewaik, 2010, who propose a weights that do not necessarily sum to one). Suppose we wish to find

$$\lambda^* = \arg\min_{\lambda} E\left(\Delta\tilde{y}_t(\lambda) - \Delta y_{\text{TRUE},t}\right)^2$$

So-defined λ^* is the GDP weight that minimizes the squared error of the weighted average. Because $Ee_{\text{GDP},t} = Ee_{\text{GDI},t} = 0$ by construction, minimization of the squared measurement error of the weighted average is equivalent to minimization of the variance of the measurement error of the weighted average. Then since

$$E\left(\Delta\tilde{y}_t(\lambda) - \Delta y_{\text{TRUE},t}\right)^2 = \lambda^2\sigma^2_{\text{MEGDP}} + (1-\lambda)^2\,\sigma^2_{\text{MEGDI}} + 2\lambda\,(1-\lambda)\,\rho\sigma_{\text{MEGDP}}\sigma_{\text{MEGDI}}$$

where $\sigma^2_{\text{MEGDP}} := E(e^2_{\text{GDP},t})$, $\sigma^2_{\text{MEGDI}} := E(e^2_{\text{GDI},t})$, denote the variance of the measurement error in GDP and GDI, respectively, and ρ denotes the correlation between $e_{\text{GDP},t}$ and $e_{\text{GDI},t}$, we can obtain

$$\lambda^* = \frac{\sigma^2_{\text{MEGDI}} - \rho\sigma_{\text{MEGDP}}\sigma_{\text{MEGDI}}}{\sigma^2_{\text{MEGDP}} + \sigma^2_{\text{MEGDI}} - 2\rho\sigma_{\text{MEGDP}}\sigma_{\text{MEGDI}}}$$

Intuitively, as σ_{MEGDI} grows large with all else is held constant, $\lambda^* \to 1$, meaning that GDP receives a heavier weighting as the variance in the GDI measurement error increases. Conversely as σ_{MEGDI} approaches zero with all else held constant, we have $\lambda^* \to 0$, meaning that GDP receives no weighting. Intuitively, as $\sigma_{\text{MEGDI}} \to 0$, GDI has no measurement error and thus GDI becomes a perfect measure of true output. Even if σ_{MEGDI} is greater than σ_{MEGDP}, $\lambda^* < 1$ provided that the measurement errors are not perfectly correlated (i.e. $\rho \neq 0$). Intuitively, if $\rho \neq 0$ some of the measurement error can be attenuated by taking an average.

Using the above formula and the summary statistics given in table 1, we obtain $\lambda^* = 0.6$

[6]Results may differ if we take the growth rate of a weighted average of the level and GDP and GDI. We choose to take w weighted average of the growth rates because of the focus on growth rates in output rather than the levels. A weighted average of growth rates is thereby less intensive for users to calculate based on estimate releases than a growth rate of a weighted level of output.

for the 1983-2009 sample. This value of λ^* by definition provides us with the weighted average with the lowest mean square error. Figure 6 depicts the optimal weighted average for the 1983-2009 sample period for the time frame spanning 1995-2009.

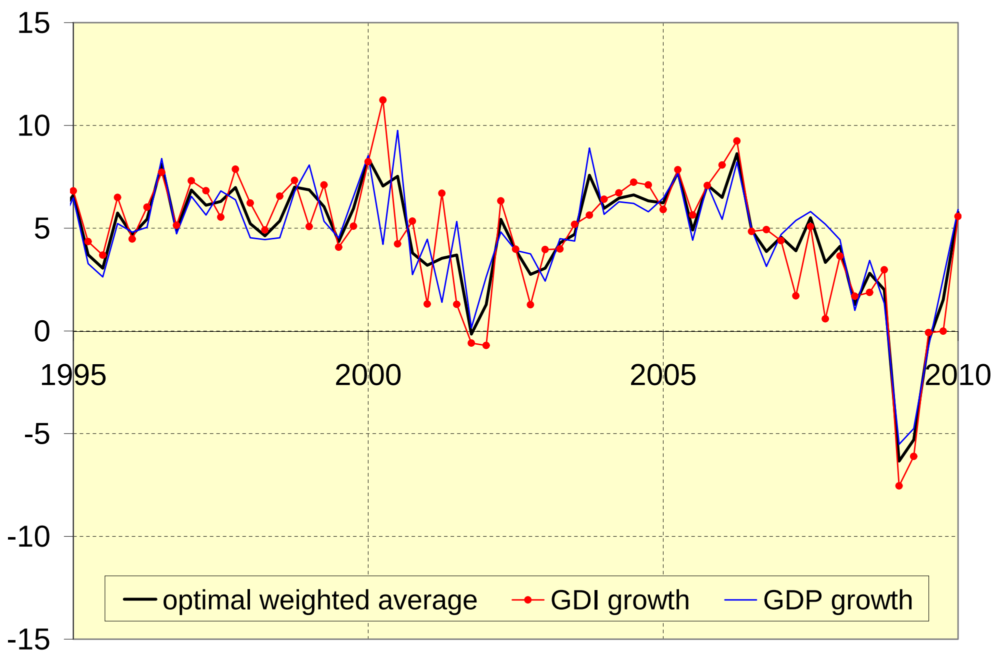

Figure 6: Optimal weighted average of GDP and GDI: 1995-2009

The table below gives the summary statistics for these various measures of output. Notably the optimal weighted average has a variance only slightly above that of the growth rate in GDP.

Table 8: Summary Statistics of output growth measures: 1983Q1-2009Q4

	True output growth	GDP growth	GDI growth	optimal weighting
mean:	5.393	5.410	5.456	5.438
variance:	6.994	10.144	7.984	8.000

The optimal weighted average gives GDP a weight of 60%

Notably the optimal weight here is derived based on the variance and correlation of the measurement error of GDP and GDI, and not on the variance and correlation between GDP and GDI growth rates themselves. Our weighting scheme thus differs to those employed elsewhere in the extant literature, where the directly observable variance and correlation between GDP and GDI growth are used to derive weights (e.g., Fixler and Nalewaik, 2007).

To cross-validate whether the optimal weighted average is indeed a more accurate measure of output we can consider how revisions to earlier vintages of the weighted average behave. Obviously, the smaller the revision, the more accurate the initial estimate. In the table below, we report summary statistics for the revision from the "third" to the "latest" vintages for various weights $\lambda = (0, 0.1, 0, \ldots, 0.9, 1)$. The third vintage is released approximately 3 months after the close of the quarter. While this represents the third release for GDP, it is the second GDP release in the first second and third quarters, and the first release of GDI in the fourth quarters.

Table 9: Revisions to weighted GDP and GDI

	GDP weight λ						
	1	0.8	0.6	0.5	0.4	0.2	0
mean	-0.34	-0.33	-0.33	-0.32	-0.32	-0.31	-0.31
variance	7.01	6.50	6.37	6.44	6.60	7.21	8.18
mean square error	7.12	6.61	6.47	6.54	6.71	7.31	8.27

Third current quarterly to latest revision

For the considered values of λ, a weight of 0.6 towards GDP performs best in terms of mean square error. (Mean square error takes into account both the mean and variance of the revision.) This is also the optimal weight implied by the common component decomposition. We interpret these results as cross-validating the idea that a weighted average of GDP and GDI, with GDP weighted at around 0.6, provides a more accurate measure of output than either GDP or GDI alone.

Note that as the weight increases incrementally from 0 towards 1 the mean square error of the revisions to the weighted average initially decreases, reaching a minimum at 0.6, before increasing again. This is consistent with the fact that revisions to GDP and GDI are not perfectly correlated, so that smaller revisions can be achieved by taking a weighted average of GDP and GDI growth.

5 Concluding Remarks

In this paper we use time series statistical techniques to evaluate whether GDP or GDI is a better indicator of true output. We demonstrate that the statistical discrepancy - the difference between the log-level of GDP and the log-level of GDI - is a stationary series. This implies that

- Over the long term, the log-level of GDP and GDI follow the same common trend.

While GDP and GDI are *in theory* of the same concept, this finding indicates that the two are also estimates of the same latent variable in *practice*. A natural interpreatation of this latent variable is that it is "true" output. Based on this interpretation, we then turn our attention to growth rates, using a state-space framework to decompose GDP and GDI growth into true output growth and measurement error. We estimate the model using the Kalman filter. We find that

- In terms of growth rates, the latest vintage of GDP is a better indicator of true output growth than the latest vintage of GDI. However, a weighted average of the GDP and GDI growth rates, with a weight of 60% to GDP, is a better indicator of true output growth than either GDP or GDI alone.

- When we consider earlier vintages of GDP and GDI, it is apparent that earlier vintages of GDI can contain more information than earlier vintages of GDP regarding the true state of the economy.

References

[1] Dickey, D. A., and W. A. Fuller (1979) "Distribution of the Estimators for Autoregressive Time Series with a Unit Root". Journal of the American Statistical Association 74, 427-431.

[2] Elliott, G., Rothenberg, T. J., and J. H. Stock (1996) "Efficient Tests for an Autoregressive Unit Root," Econometrica 64, 813-836.

[3] Engle, R. F., and M. Watson (1981) "A One-Factor Multivariate Time Series Model of Metropolitan Wage Rates," Journal of the American Statistical Association, 76, 774-781.

[4] Chen, B, and P. Zadroni (2007). "Real-Time State-Space Method for Com- puting Filtered Estimates of Future Revisions of U.S. Quarterly GDP," Proceedings of American Statistical Association, Business and Economic Statistics Section.

[5] Fama, E. F., and M. R. Gibbons (1982) "Inflation, Real Returns, and Capital Investment"Jounral of Monetary Economics, 9, 297-323.

[6] Fixler, D. and J. Nalewaik (2007) "News, Noise, and Estimates of the True Unobserved State of the Economy," Board of Governors of the Federal Reserve System, FEDS working paper 2007-34.

[7] Fixler, D. and B. Grimm (2002) "Reliability of GDP and Related NIPA Estimates," Survey of Current Business 82, 9-27.

[8] Fixler, D. and B. Grimm (2006) "GDP estimates: Rationality tests and turning point performance," Journal of Productivity Analysis, 25, 213-229.

[9] Grimm, B. (2005) "Alternative Measures of U.S. Economic Activity in Business Cycles and Dating," BEA working paper 2005-05.

[10] ——— (2007) "The Statistical Discrepancy," BEA working paper 2007-01.

[11] Kalman, R. E. (1960) "A New Approach to Linear Filtering and Prediction Problems." Journal of Basic Engineering, Transactions of the AMSE Series D, 82, 35-45.

[12] Kalman, R. E. (1963) "New Methods in Wiener Filtering Theory." Pro- ceedings of the First Symposium of Engineering Applications of Random Function Theory and Probability, 270-388. New York: Wiley.

[13] Klein, Lawrence R. and J. Makino (2000) "Economic interpretations of the statistical discrepancy" Journal of Economic and Social Measurement, 26, 11-29.

[14] Landefeld, S. (2010), "Comments on 'The Income- and Product Side- Estimates of U.S. Output Growth'", Presented at the Brookings Pane on Economic Activity, Spring 2010.

[15] Mankiw, N., Runkle, D., and Shapiro, M. (1984) "Are Preliminary Announcements of the Money Stock Rational Forecasts?"Journal of Monetary Economics, 14, 15-27.

[16] Mankiw, N. and Shapiro, M. (1986) "News or Noise: An Analysis of GNP Revisions"Survey of Current Business, 20-25.

[17] McConnell, M., and G. Perez-Quiros (2000) "Output Fluctuations in the United States: What Has Changed since the Early 1980s?" American Economic Review, 90, pp. 1464-76.

[18] Nordhaus, W. D. (2010) "The Statistical Discrepancy" Presentation to the BEA Advisory Committee, May 7 2010.

[19] Patterson, K. D., and S. M. Heravi (2004) "Revisions to Official Data on U.S. GNP: A Multivariate Assessment of Di¤erent Vintages" Journal Offficial Statistics, 4, 573–602.

[20] Patterson, K.D. (1994), "A State-Space Model for Reducing the Uncertainty Associated with Preliminary Vintages of Data with An Application to Aggregate Consumption," Economic Letters, 46, 215-222.

[21] Stock, J., and M. Watson (1991). "A Probability Model of the Coincident Economic Indicators," in Leading Economic Indicators: New Approaches and Forecasting Records, edited by K. Lahiri and G. Moore, Cambridge University Press, 1991.

[22] ——— (2002) "Has the Business Cycle Changed and Why?" NBER Macroeconomics Annual 2002, Mark Gertler and Ken Rogoff (eds), MIT Press.

[23] Weale, M. (1992) "Estimation of Data Measured With Error and Subject to Linear Restrictions," Journal of Applied Econometrics 7, 167-174.