w o r k i n g
p a p e r

17 08

Forecasting GDP Growth with NIPA Aggregates

Christian Garciga and Edward S. Knotek II

FEDERAL RESERVE BANK OF CLEVELAND

ISSN: 2573-7953

Working Paper 17-08 May 2017

Forecasting GDP Growth with NIPA Aggregates

Christian Garciga and Edward S. Knotek II

Beyond GDP, which is measured using expenditure data, the U.S. national income and product accounts (NIPAs) provide an income-based measure of the economy (gross domestic income, or GDI), a measure that averages GDP and GDI, and various aggregates that include combinations of GDP components. This paper compiles real-time data on a variety of NIPA aggregates and uses these in simple time-series models to construct out-of-sample forecasts for GDP growth. Over short forecast horizons, NIPA aggregates—particularly consumption and GDP less inventories and trade—together with these simple time-series models have historically generated more accurate forecasts than a canonical AR(2) benchmark. This has been especially true during recessions, although we document modest gains during expansions as well.

Keywords: forecasting, GDP, GDI, real-time data, consumption.

JEL classifications: C53, C32, E01.

Suggested citation: Garciga, Christian, and Edward S. Knotek II, 2017. "Forecasting GDP Growth with NIPA Aggregates," Federal Reserve Bank of Cleveland, Working Paper no. 17-08. https://doi.org/10.26509/frbc-wp-201708.

Christian Garciga is at the Federal Reserve Bank of Cleveland. Edward S. Knotek II (corresponding author) is also at the Federal Reserve Bank of Cleveland (edward.knotek@clev.frb.org).

I. Introduction

Real gross domestic product (GDP) is a comprehensive measure of an economy's output. In the United States, the Bureau of Economic Analysis (BEA) publishes estimates of GDP in the National Income and Product Accounts (NIPAs). These estimates are derived using expenditure data: real GDP is the chain-weighted sum of the components consumption, investment, government, and net exports; see Landefeld et al. (2008). In turn, these components can be separated into subcomponents; e.g., investment comprises nonresidential fixed investment, residential investment, and private inventory investment. While the expenditure-based GDP measure receives considerable attention, the NIPAs contain estimates of the theoretically equivalent income-based measure of the economy's output—gross domestic income (GDI); see Nalewaik (2010)—and the average of GDP and GDI. In addition, the BEA publishes information on a number of other NIPA aggregates, which are combinations that typically omit one or more of the (sub)components of GDP.

Akin to the way in which economists have attempted to use core inflation—which excludes volatile food and energy prices—to predict headline inflation, the omission of (sub)components of GDP in these NIPA aggregates may be useful in extracting the signal for where GDP is going.[1] For example, because inventory investment can be volatile from one quarter to the next, a measure of GDP that excludes the change in inventories may help forecast GDP. In practice, economists have drawn conclusions about which aggregates most closely resemble overall GDP and therefore may be helpful for forecasting on the basis of in-sample fit using the most revised data; see, e.g., Council of Economic Advisers (2015) and Kawa (2017).

[1] Aruoba et al. (2012, 2016) posit that GDP is a noisy measure of true output growth; by extension, NIPA aggregates may be useful in extracting the true underlying state of the economy.

This paper examines the ability of NIPA aggregates to forecast GDP growth. Instead of focusing on in-sample fit with the most revised vintage data, the first contribution of the paper is to compile or, in some cases, to reconstruct real-time vintages for NIPA aggregates. While real-time vintages for real GDP and real personal consumption expenditures (PCE) are readily available from real-time data repositories, we also collect real-time vintage data on the growth rates of: GDI; the average of GDP and GDI; real final sales of domestic product; real final sales to domestic purchasers, which we refer to as domestic final purchases (DFP); and real final sales to private domestic purchasers, which we refer to as private domestic final purchases (PDFP). We format these vintages of real-time data series to match the formatting of other real-time data repositories for other researchers to use in their research.

Using these real-time data vintages, the second contribution is to assess the ability of NIPA aggregates to forecast GDP growth out-of-sample. Our consideration of a range of NIPA aggregates in a real-time, out-of-sample forecasting exercise for GDP growth is novel. Despite a great interest in forecasting GDP growth and a large body of research, the survey article by Chauvet and Potter (2013) shows that, in general, a univariate AR(2) process is a difficult forecasting benchmark to beat. Using simple forecasting techniques, we show that a variety of forecasts generated with NIPA aggregates have historically been more accurate than the forecasts from the AR(2) benchmark. These results are strongest for DFP—which is the sum of PCE, nonresidential fixed investment, residential investment, and government purchases—and for PCE, over horizons of 1 to 3 quarters; by the 4-quarter horizon, it is difficult to improve on the AR(2) benchmark. The forecasting accuracy gains are especially visible during recessions, although we document modest gains during expansions as well. Our results also illustrate a key limitation of GDI. Nalewaik (2010) advocates for GDI as a more compelling measure of economic output than GDP and documents that it also has explanatory power for GDP in real time; Nalewaik (2012) and Aruoba et al. (2012, 2016) provide further support for GDI. While asking and answering a different question, our results indicate that the ability of GDI to forecast GDP is hampered by data release lags.

II. Real-Time Data Vintages

We compile or construct real-time vintage data for seven quarterly series in the NIPAs; see Bureau of Economic Analysis (2016) for more details on the series. To do so, we use data from the Federal Reserve Bank of St. Louis' Archival Economic Data (ALFRED) online database; the Federal Reserve Bank of Philadelphia's online Real-Time Data Set for Macroeconomists (RTDSM), as documented in Croushore and Stark (2001); the BEA Data Archive: National Accounts (NIPA); historical print issues of the BEA's monthly *Survey of Current Business*, available on the Federal Reserve Bank of St. Louis' FRASER website; Haver Analytics; and Nalewaik (2012). The seven series are:

1. Real gross domestic product (GDP).
2. Real gross domestic income (GDI).
3. The average of GDP and GDI, which we denote by GDA (for "average"). The BEA defines GDA as the arithmetic average of nominal GDP and nominal GDI, divided by the implicit GDP deflator. The BEA began officially publishing this measure in July 2015 with the release of the first or "advance" estimate of second quarter GDP for 2015, although in practice economists could have been calculating it themselves at any point

before then; see the discussion in Nalewaik (2010, p. 127) for evidence that this was being done in practice.[2]

4. Real final sales of domestic product (FSDP). The BEA defines this series as GDP minus the change in private inventories; alternatively, it can be built up as the sum of PCE, nonresidential fixed investment, residential investment, government spending, and net exports.[3]
5. Real domestic final purchases (DFP). Formally called real final sales to domestic purchasers by the BEA, this aggregate is GDP minus the change in private inventories and net exports; alternatively, it can be built up as the sum of PCE, nonresidential fixed investment, residential investment, and government spending.
6. Real private domestic final purchases (PDFP). Formally called real final sales to private domestic purchasers by the BEA, but also referred to as PDFP (see Council of Economic Advisers 2015) or private final domestic purchases (PFDP), this aggregate is GDP minus the change in private inventories, trade, and government spending; alternatively, it is the sum of PCE, nonresidential fixed investment, and residential investment. As with GDA, the BEA only began officially publishing this series in July 2015, although economists could have been calculating it themselves at any point before then, in an effort to remove volatility in GDP coming from inventories, trade, and government spending.
7. Real personal consumption expenditures (PCE).

The data vintages we construct are similar to those in ALFRED. The first row in our dataset contains vintage dates, which identify when NIPA data releases occurred. This convention allows us to generate forecasts using the data that would have been available at any

[2] See Fixler et al. (2014) for analysis of combinations of GDP and GDI as well as the research cited below.
[3] For the sake of exposition in the text, we abstract from chain-weighting when summing series to form aggregates, but we follow BEA practice as necessary when constructing measures.

given point in the past. The first available vintage date for real GDP in ALFRED is December 4, 1991, while most other series have earlier vintage dates.[4] The last data vintage comes from the NIPA release on April 28, 2017. Most of the data vintages have their first growth rate observation in 1947:Q2.[5] At each point in our forecasting exercises, data availability is determined by the vintage date. In all cases, we compute quarterly annualized growth rates using the standard BEA formula as $g(X_t)=100\times[(X_t/X_{t-1})^4-1]$, where X_t is the level of the series in quarter t. Appendix A provides a detailed accounting of how we compiled or constructed vintages of each aggregate. While GDI, GDA, and PCE differ in their construction from the other aggregate measures we consider, we include them under the rubric of "NIPA aggregates" for the sake of expositional simplicity.

III. Methodology

We evaluate the ability of NIPA aggregates, in combination with simple forecasting methods, to forecast out-of-sample quarterly annualized real GDP growth using real-time data. We do so by recursively generating forecasts using the same information sets that professional forecasters participating in either the Survey of Professional Forecasters (SPF) or the Blue Chip Economic Indicators (BC) survey would have had available. Because the information sets differ for these two surveys, we conduct two separate forecasting exercises.

---

[4] The earliest GDI and GDA vintages are from January 29, 1992. The earliest vintages for FSDP, DFP, and PDFP are from August 19, 1965, while the earliest vintage for PCE is from December 21, 1958.

[5] Because the levels of most series begin in 1947:Q1, growth rates begin in Q2. In some cases, especially around comprehensive revisions, complete time series data extending all the way back to 1947:Q1 were not available, often temporarily. In such cases, we make the reasonable assumption that forecasters would have temporarily backfilled the missing growth rate data using the data that had been available prior to the revision.

The first exercise uses SPF information sets. The SPF survey is conducted once per quarter, in the first half of the middle month of the quarter.[6] We take the forecast origin in each quarter to be the closing date for submissions to that quarter's SPF release. At this point in quarter *t*, we generally have the "advance" (i.e., first) release of GDP for quarter $t-1$ along with many NIPA aggregates. However, because of publication delays for GDI and thus GDA, the last observation for these variables is usually from quarter $t-2$.[7] In addition, the BEA makes the advance estimate using incomplete source data. Hence, this exercise implicitly allows us to examine the effect of measurement errors in the initial estimates of GDP and many NIPA series compared with publication lags in GDI and GDA on the ability of these aggregates to forecast real GDP growth.

The second exercise uses BC information sets. The BC survey is released three times each quarter on the 10th of each month, with surveys conducted in the prior week.[8] Following Chauvet and Potter (2013), we make out-of-sample forecasts in the first month of each quarter using the data that would have been available to Blue Chip survey contributors when submitting their forecasts. At this point in quarter *t*, we have available the BEA's third NIPA release with GDP *and* GDI for quarter $t-2$. Thus, this timing allows us to assess the forecasting ability of NIPA aggregates conditional on using the BEA's third estimates—which include more complete source data than is included in the advance release—through the same final quarter for all series.

[6] SPF survey dates and release dates are available from the Federal Reserve Bank of Philadelphia's website at https://www.phil.frb.org/-/media/research-and-data/real-time-center/survey-of-professional-forecasters/spf-release-dates.txt?la=en.
[7] While this delay is the normal pattern, there are a few exceptions in our dataset when GDI was available at this point in the quarter.
[8] The BC survey is conducted over a two-day period. If it is reported in the release, we use the closing date for submissions. In cases where the BC survey date is not available, we follow Knotek and Zaman (forthcoming) and assume the survey date was the first Thursday of the month, unless that is the first day of the month, in which case we assume the survey date was the first Tuesday of the month.

We look at the ability of each NIPA aggregate to generate out-of-sample forecasts using real-time data for 1- through 3-step-ahead GDP growth.[9] We focus on relatively simple forecasting methods and time series models in order to let the information in the NIPA aggregates speak. With more complicated forecasting models and methodologies, the contribution of the data versus the model becomes less clear. While a potential limitation of our analysis, the principle of parsimony tends to be successful in a variety of forecasting contexts.[10] For the sake of exposition, assume that when making a forecast in quarter t, the information set includes the most recently available observation on GDP growth from quarter $t-j$, $g(GDP_{t-j})$, while the most recently available observed growth rate for the aggregate X is from quarter $t-k$, $g(X_{t-k})$, where $k \geq j$. We consider four basic forecasting methods:

1. RW: Our first method takes the flavor of a random walk: the most recent observation on the growth of aggregate X is the forecast for quarterly GDP growth going forward, which we denote with a hat, "^".

$$\hat{g}(GDP_{t-j+i}) = g(X_{t-k}),\ i = 1,2,3,4 \tag{1}$$

   In this case, a forecaster would simply predict future GDP growth by reading the most recently available growth rate of aggregate X from the NIPA release.

2. RW-AR(p): The second method estimates an AR(p) model using growth in the aggregate X up through time $\tau = t-k$.

$$g(X_\tau) = a + \sum_{h=1}^{p} b_h g(X_{\tau-h}) + e_\tau \tag{2}$$

[9] While not reported, and consistent with Chauvet and Potter (2013) and the literature such as Bańbura et al. (2013), we found essentially no ability of our measures to outperform univariate GDP growth forecasts from the 4-step horizon onward.
[10] See, e.g., Chauvet and Potter (2013) and Edge et al. (2010) for GDP, and Atkeson and Ohanian (2001) and Faust and Wright (2013) for inflation.

We estimate and then iteratively use equation (2) to forecast the growth of the aggregate X at time $t-j+i$, $\hat{g}(X_{t-j+i})$, for i=1,2,3,4, and, as in equation (1), we treat that forecast as the forecast for GDP growth: $\hat{g}(GDP_{t-j+i}) = \hat{g}(X_{t-j+i})$.

3. Direct(p): The third method makes direct i-step-ahead forecasts for GDP growth using p lags of the aggregate X. We estimate regressions of the form:

$$g(GDP_{\tau+i}) = a + \sum_{h=1}^{p} b_h g(X_{\tau+1-h}) + e_{\tau+i},\ i = 1,2,3,4 \tag{3}$$

and use the estimated coefficients from equation (3) to generate forecasts for $\hat{g}(GDP_{t-j+i})$.

4. VAR(p): The fourth method uses bivariate vector autoregressions in GDP growth and the growth of the aggregate X with p lags. We use equation-by-equation OLS to estimate VARs of the form:

$$\begin{bmatrix} g(GDP_{\tau}) \\ g(X_{\tau}) \end{bmatrix} = \mathbf{A} + \sum_{h=1}^{p} \mathbf{B}_h \begin{bmatrix} g(GDP_{\tau-h}) \\ g(X_{\tau-h}) \end{bmatrix} + \mathbf{e}_{\tau} \tag{4}$$

and use the estimated coefficients from equation (4) to iteratively generate forecasts of $\hat{g}(GDP_{t-j+i})$ for i=1,2,3,4.

Implicitly, the RW and RW-AR(p) methods assume that the growth rate of the aggregate X is similar to the underlying growth rate of GDP and that forecasts can take advantage of this relationship, while the direct(p) and VAR(p) methods are able to better capture persistent differences in growth rates. For the RW-AR(p), direct(p), and VAR(p) models, we use expanding windows with the first data point in 1964:Q2 to estimate the parameters of the respective model.[11] We consider p lags between 1 and 4. For our purposes, we will refer to a combination of forecasting method, choice of lag length p, and NIPA aggregate as a "model."

[11] In Appendix C, we estimate the parameters of the models using rolling windows of 40 quarters.

To assess our models' forecasting performance, we examine root mean squared forecast errors (RMSFEs) relative to the benchmark used in Chauvet and Potter (2013): an AR(2) model with an expanding estimation window starting in 1964:Q2. Edge et al. (2010) also employ an AR(2) benchmark for GDP growth; D'Agostino et al. (2006) document that it is difficult to improve upon the GDP growth forecasts from simple benchmarks in the post-1985 period.

As in all studies using real-time data, we need to take a stand on what constitutes the "true" GDP growth rate for measuring forecasting accuracy, and we use the third NIPA release for a given quarter from the RTDSM; Tulip (2009) provides a justification for using this vintage as the "truth." The evaluation period is 100 quarters long, beginning in 1992:Q1 and ending in 2016:Q4.[12] Further, as in Chauvet and Potter (2013), we look at forecasting performance based on the entire sample, the subsample including only expansions as defined by the NBER dating committee, and the subsample including only recessions. While NBER recessions and expansions are not known in real time, examining the results conditional on the business cycle augments our understanding of what drives the full-sample, true real-time results. Our entire evaluation sample includes two recessions which lasted a combined eleven quarters: the 2001:Q1-2001:Q4 recession and the 2007:Q4-2009:Q2 recession. To examine predictive ability conditional on being in an economic expansion, we remove all forecasting errors corresponding to quarters in which the U.S. economy was in recession; we follow a similar procedure when we condition on being in a recession. For models with relative RMSFEs less than 1, we assess statistical significance using the Diebold and Mariano (1995) (DM) test for equal predictive accuracy between a given forecast and the forecast from the AR(2) benchmark with the small-

[12] Following the BEA, we use fixed-weight real GDP at the beginning of the evaluation period and chain-weighted real GDP from 1996 onward.

sample correction of Harvey et al. (1997), using MSE as the metric and two-sided t-statistic tests.[13]

IV. Survey of Professional Forecasters Exercise Results

SPF surveys are released in the middle of the second month of each quarter. At that point, the first estimate of GDP growth is available for the previous ($t-1$) quarter. The first forecast for quarterly annualized real GDP growth in each survey is for the quarter t in which the survey is released (i.e., the nowcast). Our exercise considers forecasts for quarters t, $t+1$, and $t+2$.[14]

Using matched SPF information sets, Table 1 presents relative RMSFEs, with the AR(2) benchmark in the denominator, for the full sample results across all model specifications and 1- through 3-step-ahead forecast horizons. We shade entries that outperform the benchmark in green, with darker shading implying lower relative RMSFEs. We note three findings from this exercise.

First, NIPA aggregates are useful in generating forecasts for GDP growth that outperform the AR(2) benchmark over short horizons. In general, the gains are modest. At the 1-step horizon, using DFP, PDFP, or PCE produces lower RMSFEs than the benchmark model across nearly all methods, with RMSFE gains in the range of 3% to 8%. At the 2-step horizon,

[13] Clark and McCracken (2009) discuss issues with tests of equal forecast accuracy when using real-time data. Hence, we view these DM test statistics as approximations. In Appendix B, we present results from the directional forecast accuracy approach of Pesaran and Timmermann (2009) and show success ratios that measure the frequency with which, for a given forecast horizon i, the forecast $\hat{g}(GDP_{t-j+i})$ and the third release $g(GDP_{t-j+i})$ are both either greater than or less than the most recently available real-time GDP growth rate at the time of the forecast, $g(GDP_{t-j})$.

[14] As an example, the 2016:Q1 SPF survey was conducted in February 2016. At that time, the advance estimate of 2015:Q4 GDP was available, as were estimates for most other NIPA aggregates we consider; however, GDI and GDA were only available through 2015:Q3. The 1- through 3-step-ahead forecasts were for 2016:Q1, 2016:Q2, and 2016:Q3.

forecasting gains remain widespread among these aggregates. In most cases, the size of the forecasting gains over the AR(2) benchmark shrinks compared with the 1-step horizon results. At the 3-step horizon, there continue to be improvements over the AR(2) benchmark from PCE, PDFP, and DFP, along with a few improvements from using FSDP, but the magnitude of the gains has decreased to between 1% and 2%. At longer forecast horizons, meaningful forecast improvements disappear, and the forecasts from using NIPA aggregates are about as good (or bad; see Edge and Gürkaynak 2010) as those from the benchmark. Despite some gains in RMSFEs approaching 10%, relatively few of the improvements in forecast accuracy are statistically significant at the 10% level based on the DM test; they come from using DFP in RW-AR(p) models at the 1-step horizon, PDFP in the VAR(1) model at the 2-step horizon, and FSDP in the RW-AR(2) model at the 3-step horizon.

Second, on the basis of the information sets available in this exercise, the publication lag in GDI compared with GDP and other NIPA aggregates renders the forecasting accuracy of GDI- and GDA-based models nearly uniformly worse than that of the AR(2) benchmark for forecasting future GDP growth. This is true even though the AR(2) benchmark is using the imprecise advance release for quarter $t-1$ to iteratively inform future forecasts.

Third, the RMSFEs from the SPF median forecasts have been smaller than those from both the models using NIPA aggregates and the AR(2) benchmark. However, the improvement in SPF forecast accuracy compared with the AR(2) benchmark is only statistically significant at the 10% level at the 1-step (nowcast) horizon based on the DM test. For the sake of comparison, nowcasting models that use mixed-frequency data or higher-than-quarterly frequency data and forecast evaluation periods similar to ours—such as the dynamic factor models in Bańbura et al. (2013) and Liebermann (2014), and the mixed-frequency models estimated with Bayesian methods and featuring stochastic volatility in Carriero et al. (2015)—produce nowcast RMSFEs that are comparable to, but greater than, those from SPF. Thus, even though we are only using quarterly NIPA aggregates for near-term forecasting, the combination of these aggregates with simple forecasting tools helps close the gap between the quarterly AR(2) benchmark and the nowcasts from SPF that incorporate judgment.

Chauvet and Potter (2013) show that the relative forecasting ability of competing models varies over the business cycle. To examine this issue, Table 2 presents the RMSFEs conditional on a given forecast quarter being part of an expansion as defined by the NBER. During expansions, improvements in forecast accuracy over the AR(2) benchmark are more model-specific than in the full sample. The RW-AR(1) model outperforms the benchmark for all measures except for PDFP at the 1-step horizon and for all measures at the 2-step horizon, with small gains from 3 of the 6 measures at the 3-step horizon. The VAR(1) performs well across all measures at the 2-step horizon, with small gains at the 3-step horizon. In looking at NIPA aggregate measures, the PCE aggregate is the most useful during expansions; together with a RW-AR(*p*), it offers the largest improvements in forecast accuracy over the benchmark—on the order of 4% to 6% based on RMSFEs at the 1-step horizon and 3% to 4% at the 2-step horizon. As was the case with the full sample, however, there are few cases in which the improvements are statistically significant based on the DM test. Whereas GDI and GDA had suffered from release lags in the full-sample results, there are some cases in which they outperform the benchmark when conditioning on the economy being in an expansion. This finding essentially says that old information can still be relevant—provided the economy does not suddenly turn down. Finally, while SPF outperforms the benchmark in the nowcast quarter during expansions, it underperforms the benchmark at other horizons. Thus, we document a number of cases in which NIPA aggregates and simple forecasting models improve upon the AR(2) benchmark conditional on the economy being in an expansion; using PCE with the RW-AR methods produces forecasts that essentially match SPF at the 1-step horizon and outperform SPF forecasts at the 2- and 3-step horizons.

Table 3 displays RMSFEs conditional on a given forecast quarter being part of a recession. With only eleven recession quarters, inference is based on a very small sample. Nevertheless, the table illustrates that DFP, PDFP, and PCE all outperform the AR(2) benchmark across most methods at all horizons conditional on the economy being in a recession at that point. In a number of cases, the gains are nearly as large as 40% in terms of RMSFE at the 1-step horizon and about 25% at the 2- and 3-step horizons. Interestingly, the largest improvements in forecast accuracy come from the simplest RW method for using PCE, PDFP, and DFP—in which the forecaster would simply read off growth rates for the aggregate from the NIPA release and use those values as the forecast for GDP growth. Conditioning on recession quarters, the SPF forecast is superior to both the AR(2) benchmark and our model-based forecasts 1- and 2-steps ahead, but by the 3-step horizon a random walk in either DFP or PDFP is substantially more accurate.

Comparing the split sample and full sample results, it is clear that the relative strength of DFP, PCE, and PDFP in the full sample is driven in large part by the relative forecasting performance of these aggregates during recessions. Conversely, GDI and GDA do poorly in the full sample because their lagged release relative to the survey dates and to the other series causes them to miss turning points. The relatively strong forecasting performance of SPF comes from the fact that the professionals handily outperform the benchmark in and around recessions and are about on a par with it—outside the nowcast quarter—during expansions.

V. Blue Chip Exercise Results

To highlight the role of different information sets, we run a second exercise in which we use the closing date of the first BC survey in each quarter to generate our forecasts. At the start of the first month of quarter t, all of our series—including GDI and GDA—would be available up through quarter $t-2$.[15] Furthermore, the observations would come from the third NIPA release, which incorporates more comprehensive information than is available in the advance release. This exercise replicates the design of Chauvet and Potter (2013), who compare the real-time forecasting ability of an AR(2) in real GDP growth with twelve competing models, including the BC survey.

Table 4 presents the relative RMSFEs for the entire sample. The results at all three forecast horizons for FSDP, DFP, PDFP, and PCE are broadly similar to those obtained when forecasting at the SPF date: there are widespread gains in forecasting accuracy from using DFP, PCE, and PDFP at the 1-step horizon that diminish at the 2-step horizon and are smaller still at the 3-step horizon; meanwhile, there are few forecasting gains from FSDP. The main difference with the SPF exercise comes from the results for GDI and GDA: because information from these series no longer lags the other measures, simple models utilizing GDI or GDA now also outperform the AR(2) benchmark. In a number of cases, these gains are statistically significant based on the DM test, even though they are usually smaller than those from the equivalent forecasting methods using DFP, PDFP, or PCE; e.g., at the 1-step horizon, a direct(p) or RW-AR(p) model using GDA is more accurate than the AR(2) benchmark but, in most cases, is less

[15] For example, for the BC survey released in January 2016, the end of our estimation sample across models is 2015:Q3. The 1-step forecast across the models would be the forecast made for 2015:Q4, the 2-step forecast would be for 2016:Q1, and the 3-step forecast would be for 2016:Q2.

accurate than the same model in DFP, PDFP, or PCE across all lags *p*, yet the gains are statistically significant at the 10% or 5% levels.

Results conditional on only expansionary quarters, shown in Table 5, are similar to those in Table 2. Across forecasting methods, the RW-AR(1) and VAR(1) again produce notable gains in forecast accuracy relative to the AR(2) benchmark, a number of which are statistically significant based on the DM test. Across measures, PCE and DFP generally perform better than the others.

During recessions, as shown in Table 6, GDI- and GDA-based models now show broad forecasting gains over the AR(2) benchmark across forecasting horizons. Across all aggregate measures and all forecasting methods, the smallest relative RMSFEs come from the simplest models—the RW specification using PCE growth for the 1-step and 2-step horizons, and either DFP or PDFP at the 3-step horizon—although the statistically significant forecasting gains based on the DM test are concentrated in models using GDA and GDI. Clearly, the improvements in forecast accuracy that come from using GDA and GDI in the full sample are due almost exclusively to their usefulness during recessionary quarters.

Our results in this section are directly comparable to those of Chauvet and Potter (2013). During expansions, Chauvet and Potter (2013) document that a range of models are unable to produce more accurate forecasts than an AR(2) benchmark at the 1- and 2-step horizons.[16] We show that NIPA aggregates can help in this regard, with some models producing lower RMSFEs since the early 1990s. During recessions, Chauvet and Potter (2013) find that an autoregressive model augmented with additional regressors coming from a dynamic factor model with Markov switching (AR-DFMS) produces RMSFEs relative to the AR(2) benchmark of 0.575 at the 1-step

[16] As noted earlier, nowcasting approaches that take advantage of higher-frequency data have enjoyed success in outperforming univariate benchmarks; see, e.g., Giannone et al. (2008), Bańbura et al. (2013), Carriero et al. (2015), and Higgins (2014).

horizon and 0.583 at the 2-step horizon.[17] In the AR-DFMS model, the Markov switching is driven by monthly data on four coincident indicators that the NBER Business Cycle Dating Committee uses to date recessions, one of which is real manufacturing and trade sales, which is released with a 2-month lag. Thus, when making a forecast at, e.g., the second quarter's BC survey date which is in early April, the model is taking signal from monthly data available through January. In our approach, the relative RMSFE from our best measure during recessions, the RW in PCE, is 0.585 at the 1-step horizon and 0.760 at the 2-step horizon. The former reading essentially matches the performance of the AR-DFMS model, while the AR-DFMS model clearly improves upon the latter. But it is worth noting that the PCE reading we use is from two quarters earlier; e.g., in early April, we would be using the fourth quarter PCE reading when making this forecast. Based on the timing of data releases, *monthly* PCE readings would be available through February at that point. With PCE being a useful near-term predictor of growth during recessions, fully utilizing the monthly PCE data could yield further forecasting gains over those we present.

VI. Discussion and Related Literature

This paper takes a novel approach to forecasting GDP growth out-of-sample by using real-time data on NIPA aggregates. This exercise replicates the forecasting process and often generates different results compared with in-sample fit using the most revised data. Notably, PDFP is intuitively appealing because it omits idiosyncratic quarterly fluctuations in inventories, trade, and government spending (especially those coming from defense spending), and it appears

[17] This approach builds on work by Chauvet (1998), Chauvet and Hamilton (2006), Chauvet and Piger (2008), and Chauvet et al. (2013).

to have the highest correlation with GDP growth at the 1-step horizon based on revised data (Council of Economic Advisors 2015). But we find little evidence that this measure is generally superior to others in our exercises: forecast accuracy from the PCE- and DFP-based measures is essentially the same as PDFP over the entire sample and during recessions, and PCE-based forecasts made with the RW-AR(*p*) approach estimated over an expanding window are more accurate during expansions.

Because some of our aggregates omit relatively volatile components or subcomponents from GDP, our approach has parallels to parts of the inflation forecasting literature that have examined the ability of core inflation measures—which can be interpreted broadly to be inflation measures that exclude food and energy prices, focus on price changes in the center of the monthly distribution by using medians or trimmed means, or omit other components than food and energy—to predict headline inflation. The results from that literature have been mixed, with some studies reporting that various core measures produce relatively more accurate forecasts for headline inflation than does headline inflation itself, while others find little forecasting benefit from core measures vis-à-vis headline inflation; see Bryan and Cecchetti (1994), Smith (2004), Meyer and Pasaogullari (2010), Crone et al. (2013), and Meyer and Venkatu (2014). The ability of core inflation to forecast headline inflation can be related to a broader literature on the usefulness of disaggregates in forecasting an aggregate; see, e.g., Lütkepohl (2006) or Hendry and Hubrich (2011). While much of this literature has focused on inflation, recent work on nowcasting suggests potential gains from aggregating across component nowcasts to generate both inflation nowcasts (see Knotek and Zaman forthcoming) and GDP nowcasts (see Foroni and Marcellino 2014 and Higgins 2014).

Overall, our results suggest that NIPA aggregates that exclude inventories and trade data—which includes the measures DFP and PDFP, as well as PCE itself—are useful predictors for future GDP growth and can outforecast a typical univariate benchmark. This is true whether we have the advance NIPA release (the SPF exercise) or the third release (the BC exercise). By contrast, GDI and GDA show predictive ability at the beginning of the first month of the quarter in the BC exercise, but they are outperformed by other NIPA aggregates by the start of the second month of the quarter. These data release lags diminish the usefulness of GDI and GDA for forecasting purposes.

While forecasting the start of a recession is challenging, the ability of our models to forecast during recessions is important in determining where they stand relative to the benchmark: much of the improvement in forecastability of GDP growth relative to the benchmark comes during recessions, where gains are widespread across measures. During expansions, by contrast, PCE and DFP show the largest forecasting gains. Because the best simple models to forecast GDP growth differ depending on the state of the business cycle, a reliable real-time indicator of the state of the business cycle, such as the dynamic factor model with regime switching of Chauvet (1998), could be useful and would allow for switching between, e.g., the expanding-window RW-AR(2) in PCE growth during expansions and the RW in PCE growth during recessions. Given its predictive ability at the quarterly frequency, monthly frequency PCE readings could be particularly useful such real-time indicators.

The results in this paper at SPF survey dates find little benefit from looking at GDI growth compared with a univariate benchmark when forecasting GDP growth, whereas GDI is competitive with other measures and often outperforms GDP growth at the BC survey dates we examine.[18] On the surface, these mixed findings contrast with other recent research that looks favorably at GDI, although details matter. In one study, Nalewaik (2012) finds that real-time GDI growth has historically been a better indicator than real-time GDP growth in determining that the economy has fallen into a recession. In working with the real-time data, Nalewaik (2012) uses the BEA's third NIPA release—at which point both GDP and GDI data are available—to compute the recession probability for quarter *t*. Due to data release lags, the third NIPA release occurs at the end of the third month of quarter t+1; hence, the study has an important backcasting component to it. When making forecasts based on BC survey dates, we also are using the data from the BEA's third NIPA release, and at that point we find more positive results for GDI and GDA.[19] Nevertheless, our performance metric—forecasting GDP growth per se—differs from determining whether the economy is in a recession or not.

Looking across a range of indicators, Nalewaik (2010) finds that GDI tends to provide a better broad summary of the state of the economy than GDP, which is constructed using an expenditure approach. Using the real-time history of third NIPA releases, Nalewaik (2010) shows that lagged GDI growth readings have more explanatory content for 1- and 2-step ahead GDP growth than lagged GDP growth. Running a formal out-of-sample forecasting exercise using third-release data, our results at the BC survey dates are similarly favorable for GDI-based forecasts. But we also document that the predictive content of GDI for forecasting GDP growth out-of-sample is greatly diminished at the SPF survey dates due to data release lags; in attempting to generate an "advance" estimate of GDI that would match up with the advance

[18] Indeed, GDI- and GDA-based forecasts enjoyed more statistically significant improvements than was the case for the other measures.

[19] E.g., for estimating the probability that the economy was in recession in Q1, Nalewaik (2012) uses the third NIPA release for Q1 data which would have become available at the end of June, which is in Q2. When making forecasts using BC data, our forecasts made in Q3 (July) use the data available at that time, which would have been the estimates released in June for Q1 data. As noted in footnote 15, the 1-step forecast would have been for Q2, etc.

estimate of GDP, Nalewaik (2010) finds that the explanatory power of GDI declines substantially. In related research, the object of underlying interest in Aruoba et al. (2016) is the unobserved true state of output growth in the economy, whereas both GDP growth and GDI growth are assumed to contain measurement error; this unobserved true output growth loads more heavily on GDI growth than on GDP growth.[20] However, Aruoba et al. (2016) note that delays in the arrival of GDI data would pose challenges in real time, and they do not examine the ability of their approach to forecast (reported) GDP growth.

VII. Conclusion

This paper collects and, in some cases, reconstructs real-time data vintages for multiple aggregate series in the BEA's national income and product accounts. We format these vintages of real-time data to match the formatting conventions of other real-time data repositories for others to use in their research. Using these real-time vintages in out-of-sample GDP growth forecasting exercises, we document gains in forecasting accuracy from NIPA aggregates combined with relatively simple time series models compared with a canonical autoregressive benchmark model.

Our results are strongest for domestic final purchases—also known as final sales to domestic purchasers, which is the sum of PCE, nonresidential fixed investment, residential investment, and government purchases—and for PCE itself, over horizons of 1 to 3 quarters. We find little ability of private domestic final purchases to outperform these other measures in forecasting GDP growth. The measure of GDP that excludes only inventories, final sales of

[20] The same results are true in Aruoba et al. (2012) using a forecast-error based approach to recovering the unobserved true state of output growth.

domestic product, appears to offer little forecasting ability over the AR(2) benchmark. By contrast, the ability of gross domestic income to compete with these other measures as a forecasting tool depends on its availability; it suffers in cases in which it has not yet been released with the other advance NIPA data.

VIII. Literature Cited

Aruoba, S. Boragan, Francis X. Diebold, Jeremy Nalewaik, Frank Schorfheide, and Dongho Song (2016) "Improving *GDP* Measurement: A Measurement-Error Perspective" *Journal of Econometrics* 191(2): 384-397.

Aruoba, S. Boragan, Francis X. Diebold, Jeremy Nalewaik, Frank Schorfheide, and Dongho Song (2012) "Improving *GDP* Measurement: A Forecast Combination Perspective" in X. Chen and N. Swanson, eds., *Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis: Essays in Honor of Halbert L. White Jr.*, Springer.

Atkeson, Andrew, and Lee E. Ohanian (2001) "Are Phillips Curves Useful for Forecasting Inflation?" Federal Reserve Bank of Minneapolis *Quarterly Review* (Winter): 2-11.

Bańbura, Marta, Domenico Giannone, Michele Modugno, and Lucrezia Reichlin (2013) "Now-casting and the Real-Time Data Flow" in Graham Elliott and Allan Timmermann, eds., *Handbook of Economic Forecasting*, vol. 2, Elsevier: North Holland.

Bryan, Michael F., and Stephen G. Cecchetti (1994) "Measuring Core Inflation" in N. Gregory Mankiw, ed., *Monetary Policy*, University of Chicago Press: Chicago.

Bureau of Economic Analysis, U.S. Department of Commerce (2016) *Concepts and Methods of the U.S. National Income and Product Accounts* https://www.bea.gov/national/pdf/all-chapters.pdf, accessed May 5, 2016.

Carriero, Andrea, Todd E. Clark, and Massimiliano Marcellino (2015) "Realtime Nowcasting with a Bayesian Mixed Frequency Model with Stochastic Volatility" *Journal of the Royal Statistical Society: Statistics in Society: Series A* 178(4): 837-862.

Chauvet, Marcelle (1998) "An Econometric Characterization of Business Cycle Dynamics with Factor Structure and Regime Switches" *International Economic Review* 39(4): 969-996.

Chauvet, Marcelle, and James D. Hamilton (2006) "Dating Business Cycle Turning Points" in Costas Milas, Philip Rothman, and Dick van Dijk, eds. *Nonlinear Time Series Analysis of Business Cycles*, Elsevier: North Holland.

Chauvet, Marcelle, and Jeremy Piger (2008) "A Comparison of the Real-Time Performance of Business Cycle Dating Methods" *Journal of Business and Economic Statistics* 26(1): 42-49.

Chauvet, Marcelle, Mi Emmy Lu, and Simon Potter (2012) "Forecasting Output Growth during the Great Recession" unpublished manuscript.

Chauvet, Marcelle, and Simon Potter (2013) "Forecasting Output" in Graham Elliott and Allan Timmermann, eds., *Handbook of Economic Forecasting*, vol. 2, Elsevier: North Holland.

Clark, Todd E., and Michael W. McCracken (2009) "Tests of Equal Predictive Ability with Real-Time Data" *Journal of Business and Economic Statistics* 27: 441-454.

Crone, Theodore M., N. Neil K. Khettry, Loretta J. Mester, and Jason A. Novak (2013) "Core Measures of Inflation as Predictors of Total Inflation" *Journal of Money, Credit and Banking* 45(2-3): 505-519.

Croushore, Dean, and Tom Stark (2001) "A Real-Time Data Set for Macroeconomists" *Journal of Econometrics* 105(1): 111-130.

Council of Economic Advisers (2015) *Economic Report of the President* Washington, DC.

D'Agostino, Antonello, Domenico Giannone, and Paolo Surico (2006) "(Un)Predictability and Macroeconomic Stability" European Central Bank working paper series no 605.

Diebold, Francis X., and Roberto S. Mariano (1995) "Comparing Predictive Accuracy" *Journal of Business and Economic Statistics* 13: 253-263.

Edge, Rochelle M., and Refet S. Gürkaynak (2010) "How Useful Are Estimated DSGE Model Forecasts for Central Bankers?" *Brookings Papers on Economic Activity* 41(2): 209-259.

Edge, Rochelle M., Michael Kiley, and Jean-Philippe Laforte (2010) "A Comparison of Forecast Performance Between Federal Reserve Staff Forecasts, Simple Reduced-Form Models, and a DSGE Model" *Journal of Applied Econometrics* 25: 720-754.

Faust, Jon, and Jonathan H. Wright (2013) "Forecasting Inflation" in Graham Elliott and Allan Timmermann, eds., *Handbook of Economic Forecasting*, vol. 2, Elsevier: North Holland.

Fixler, Dennis J., Ryan Greenaway-McGrevy, and Bruce T. Grimm (2014) "The Revisions to GDP, GDI, and Their Major Components" *Survey of Current Business* 94 (August): 1-23.

Foroni, Claudia, and Massimiliano Marcellino (2014) "A Comparison of Mixed Frequency Approaches for Nowcasting Euro Area Macroeconomic Aggregates" *International Journal of Forecasting* 30: 554-568.

Giannone, Domenico, Lucrezia Reichlin, and David Small (2008) "Nowcasting: The Real Time Information al Content of Macroeconomic Data Releases" *Journal of Monetary Economics* 55(4): 665-676.

Harvey, David, Stephen Leybourne, and Paul Newbold (1997) "Testing the Equality of Prediction Mean Squared Errors" *International Journal of Forecasting* 13: 281-291.

Hendry, David F., and Kirstin Hubrich (2011) "Combining Disaggregate Forecasts or Combining Disaggregate Information to Forecast an Aggregate" *Journal of Business and Economic Statistics* 29: 216-227.

Higgins, Patrick (2014) "GDPNow: A Model for GDP 'Nowcasting'" Federal Reserve Bank of Atlanta working paper 2014-7.

Kawa, Luke (2017) "Here Comes the Economic Growth that Confidence Data's Predicting" https://www.bloomberg.com/news/articles/2017-02-06/here-comes-the-economic-growth-that-confidence-data-s-predicting, accessed February 6, 2017.

Knotek, Edward S., II, and Saeed Zaman (forthcoming) "Nowcasting U.S. Headline and Core Inflation" *Journal of Money, Credit and Banking*.

Landefeld, J. Steven, Brent R. Moulton, and Cindy M. Vojtech (2003) "Chained-Dollar Indexes: Issues, Tips on Their Use, and Upcoming Changes" *Survey of Current Business* 83 (November): 8-16.

Landefeld, J. Steven, Eugene P. Seskin, and Barbara M. Fraumeni (2008) "Taking the Pulse of the Economy: Measuring GDP" *Journal of Economic Perspectives* 22(2): 193-216.

Liebermann, Joelle (2014) "Real-Time Nowcasting of GDP: A Factor Model vs. Professional Forecasters" *Oxford Bulletin of Economics and Statistics* 76(6): 783-811.

Lütkepohl, Helmut (2006) "Forecasting with VARMA Models" in Graham Elliott, Clive W.J. Granger, and Allan Timmermann, eds., *Handbook of Economic Forecasting*, vol. 1, Elsevier: North Holland.

Meyer, Brent H., and Mehmet Pasaogullari (2010) "Simple Ways to Forecast Inflation: What Works Best?" Federal Reserve Bank of Cleveland *Economic Commentary* 2010-17.

Meyer, Brent, and Guhan Venkatu (2014) "Trimmed-Mean Inflation Statistics: Just Hit the One in the Middle" Federal Reserve Bank of Cleveland working paper no. 12-17R.

Nalewaik, Jeremy J. (2010) "The Income- and Expenditure-Side Estimates of U.S. Output Growth" *Brookings Papers on Economic Activity* (Spring): 71-106.

Nalewaik, Jeremy J. (2012) "Estimating Probabilities of Recession in Real Time Using GDP and GDI" *Journal of Money, Credit and Banking* 44(1): 235-253.

Pesaran, M. Hashem, and Allan Timmermann (2009) “Testing Dependence Among Serially Correlated Multicategory Variables” *Journal of the American Statistical Association* 104: 325-337.

Smith, Julie K. (2004) “Weighted Median Inflation: Is This Core Inflation?” *Journal of Money, Credit and Banking* 36: 253-263.

Tulip, Peter (2009) “Has the Economy Become More Predictable? Changes in Greenbook Forecast Accuracy” *Journal of Money, Credit and Banking* 41(6): 1217–31.

Appendix A: Real-Time Data

We detail the sources for the seven real-time data series we use in the paper.

Real gross domestic product (GDP)

Real-time vintages for real GDP come from the Federal Reserve Bank of St. Louis' ALFRED online database.

Real gross domestic income (GDI)

Our real-time real GDI vintages begin on January 29, 1992, and end on April 28, 2017. For vintages dated September 27, 2002, through April 28, 2017, we construct real-time vintages of nominal GDP, nominal GDI, and real GDP using the BEA Data Archive: National Accounts (NIPA).[21] Real-time vintage dates are assigned based on the dates of NIPA releases as reported by the BEA. From these, we deflate nominal GDI using the implicit GDP deflator computed from nominal GDP and real GDP.[22]

For vintages dated January 29, 1992, through August 29, 2002, we hand collect real GDI data as published in the BEA's monthly *Survey of Current Business* (*SCB*) series. Most *SCB* publications include the most recent estimate of real GDI and a handful of earlier observations, while those corresponding to BEA annual revision releases include observations from the previous three years. To fill in the remaining time-series of these vintages, we use the real GDI

[21] Data are available at: https://www.bea.gov/histdata/histChildLevels.cfm?HMI=7.

[22] The time series history prior to 1995:Q1 is missing for vintage dated July 31, 2009, for nominal GDP, nominal GDI, and real GDP. ALFRED data indicate that, for this vintage, the deep-history is the same as that for the following vintage, dated August 27, 2009, for nominal GDP and real GDP. Hence, we fill in the history of vintage dated July 31, 2009, using the August 27, 2009 vintages of nominal GDP, nominal GDI, and real GDP. The real GDP vintage dated December 23, 2003, is also missing. ALFRED data indicate that data for this vintage is the same as for the January 30, 2004, vintage, except for the last (i.e., the 2003:Q4) observation. Hence, we fill in the vintage dated December 23, 2003, of real GDP using all but the last value of vintage dated January 30, 2004.

vintage dataset from Nalewaik (2012).[23] The columns (which represent different vintages) in this dataset are unlabeled, which necessitated assigning a vintage date to each column. We do this by finding the intersection of the hand collected *SCB* data and the Nalewaik (2012) data.[24,25] On the basis of this match, we fill in the observations for the 1992M1-2002M8 vintages going back to 1959Q4. Finally, we correct the following three vintage labels in the final real GDI dataset:

- The May 1997 *SCB* had the GDP release date as April 30, 1997. However, the *SCB* data actually come from the release on May 7, 1997, because the deep-history of real GDP in the May 7, 1997 vintage of ALFRED matches the deep-history of real GDP in this *SCB* (i.e., this *SCB* coincided with a comprehensive revision). Also, after matching the Nalewaik (2012) data to the data collected by hand in the *SCB*, the deep-history data changes from vintage March 28, 1997, to April 30, 1997. Hence, the date of this vintage in the collected-by-hand file was changed from April 30, 1997, to May 7, 1997.
- The November 1999 *SCB* had the GDP release date as October 28, 1999. However, judging by the real GDP and GDP deflator ALFRED vintages, there was a release on October 29, 1999, with the exact same data from 1994 onwards and additional observations before that which differed from the previous vintage with a deep-history. That is, this release coincided with a comprehensive revision. Also, after matching the Nalewaik (2012) data to the data collected by hand from the *SCB*, the deep history data changes from vintage dated September

[23] Data are available at: https://jmcb.osu.edu/archive/volume-44.
[24] As a robustness check, we also assigned vintage labels to the Nalewaik (2012) data by using ALFRED vintages of nominal GDP, real GDP, and the GDP deflator, as well as Haver vintages of nominal GDI going back to 1999M9. The resulting real-time real GDI set is identical to the one derived by assigning vintage labels to Nalewaik (2012) data on the basis of the hand collected *SCB* data.
[25] In cases where a column in the Nalewaik (2012) data clearly corresponds to a particular *SCB* vintage but the data differ between the two, we assign the vintage label to the Nalewaik (2012) column but take the available *SCB* data to be the truth. This occurs for five SCB vintages: February 23, 1996; May 2, 1996; August 1, 1996; October 30, 1996; and October 28, 1999. In each case, there is one additional observation in the *SCB* vintage beyond the final observation available in the Nalewaik (2012) column, or the final observation differs between the two.

30, 1999, to October 28, 1999. Hence, we changed the date of this vintage from October 28, 1999, to October 29, 1999.

- The April 2000 *SCB* had the GDP release date as March 30, 2000. However, the *SCB* data actually comes from the release on April 3, 2000, which is when, according to this *SCB*: "On March 30, 2000, as part of the comprehensive revision of the NIPA's, BEA released revised NIPA estimates for 1929-58 that incorporated the definitional and statistical changes that had been incorporated earlier into the estimates beginning with 1959. In addition, BEA released revised estimates beginning with 1959 that incorporated corrections and a previously announced methodological improvement. The revisions were not sizable enough to affect the average annual growth rate in real GDP for 1929-58 or for 1959-98, but the growth rates for individual years were revised by as much as 0.5 percentage point." (page *i*) This *SCB* had a longer history than most *SCB*s. In the real GDP ALFRED data, the deep history for the vintage dated April 3, 2000, is slightly different from that for the vintage dated March 30, 2000. Finally, after matching the Nalewaik (2012) data to the *SCB* data that were collected by hand, the deep history data changes slightly from vintage dated February 25, 2000, to March 30, 2000. Hence, we changed the date of this vintage in the collected by hand file from March 30, 2000, to April 3, 2000.

While these changes have no impact on our results for GDI, we make the changes in order to more accurately match vintages of real GDI with real GDP in order to construct the average of GDP and GDI.

The average of GDP and GDI, which we denote by GDA

Given real-time vintages of quarterly annualized growth rates of real GDP and real GDI, we calculate real GDA growth by first matching each vintage of GDI with a vintage of GDP. All vintages of GDI have an exact match, except for a handful. For these, we take the nearest GDP vintage which occurs within five days of the GDI vintage and in the same month, and label the GDA vintage the later of the GDI and GDP vintages. We are able to match each vintage of GDI in this way.

Due to the identities (real GDP)=(nominal GDP)/deflator and (real GDI)=(nominal GDI)/deflator, real GDA may be calculated as the arithmetic average of real GDP and real GDI. Therefore, we convert the growth rates of real GDP and real GDI to level indices, take the arithmetic average, and compute the annualized quarter-over-quarter growth rates.

Real final sales of domestic product (FSDP)

Real-time vintages for real final sales of domestic product (FSDP) come from ALFRED.

Personal Consumption Expenditures (PCE)

Real-time vintages for real PCE come from ALFRED.

Real private domestic final purchases (PDFP)

Formally called real final sales to private domestic purchasers by the BEA, PDFP is GDP minus the change in private inventories, trade, and government spending; alternatively, it is the sum of PCE, nonresidential fixed investment, and residential investment. As with GDA, the BEA only began officially publishing this series in July 2015, although economists could have been calculating it themselves at any point before then. The real-time vintage data for PDFP are available in ALFRED starting with the July 30, 2015, NIPA release date. Prior to that time, we reconstruct the growth rates of PDFP that economists would have been able to construct for themselves in real time. To do so, we sum components rather than starting from GDP and subtracting components.

In addition to real-time vintage data on real PCE, we also collect real-time vintage data on real nonresidential (or business) fixed investment, BFI, and real residential investment, RES. In both cases, we require data from ALFRED and the RTDSM.[26] Prior to the changeover to chain-weighting with the January 19, 1996, NIPA release, real PDFP was simply the sum of real PCE, real BFI, and real RES. Starting with the January 19, 1996, NIPA release and running up through the July 30, 2015, NIPA release, we construct PDFP using the chain-weighting formula in, e.g, Landefeld et al. (2003):

$$I_t = I_{t-1}\sqrt{\frac{\mathbf{P}'_{t-1}\mathbf{Q}_t}{\mathbf{P}'_{t-1}\mathbf{Q}_{t-1}} \times \frac{\mathbf{P}'_t\mathbf{Q}_t}{\mathbf{P}'_t\mathbf{Q}_{t-1}}} \tag{5}$$

where I_t is the chain-type quantity index at time t for PDFP, $\mathbf{Q}_t$=[PCE,BFI,RES]′ is the vector of real quantities at time t for the components included in PDFP, and $\mathbf{P}_t$ is the vector with their associated chain-type price indexes. The PCE price index is available in ALFRED starting with the January 19, 1996, NIPA release. We compiled the BFI and RES price indexes from multiple sources. For vintages dated January 19, 1996, through August 26, 1999, we hand collected BFI and RES price indexes from monthly issues of the *SCB*. For vintages dated September 30, 1999, through January 30, 2013, we collected real-time data on BFI and RES price indexes from the

[26] Following the BEA, the ALFRED data on BFI and RES from December 2003 onward do not have the entire time series history; instead, they stop in the 1990s. The RTDSM data go back to 1947 by using the chain-weight quantity indexes to backfill the missing observations. We splice the series together to fill in the missing data.

BEA via Haver Analytics.[27] For vintages dated February 28, 2013, onward, the BFI and RES price indexes are available via ALFRED.[28]

As a check, for the July 30, 2015, NIPA vintage, we compared the time series history of annualized growth rates from our procedure to reconstruct the series with the annualized growth rates of the as-reported BEA series from ALFRED. The root mean squared error was 0.0151 percentage point, and the maximum absolute error was 0.0692 percentage point.

Real domestic final purchases (DFP)

Formally called real final sales to domestic purchasers by the BEA, DFP is GDP minus the change in private inventories and net exports; alternatively, it can be built up as the sum of PCE, nonresidential fixed investment, residential investment, and government spending. While the BEA has a long history of reporting DFP, real-time vintage data begin in ALFRED with the February 28, 2013 vintage.

For vintages prior to January 19, 1996, which do not use chain-weighting, we collect real-time vintage data on real government spending, GOVT, from ALFRED, and we compute real DFP as the sum of real PCE, real BFI, real RES, and real GOVT. For the January 19, 1996, vintage, we collect real DFP by hand from the *SCB* for the period 1992:Q1 through 1995:Q3. For observations prior to 1992:Q1, we collect the chain-type price index for government spending from the *SCB* and use equation (5) to extend the real DFP series back to 1959:Q3. For vintages up through August 26, 1999, we hand collect real DFP from issues of the *SCB.* For

[27] We date the price indexes vintages with the corresponding real series release date as recorded in ALFRED. In two cases, there were two NIPA releases within several weeks of each other which were not correctly captured in the Haver Analytics database—on October 28, 1999, and December 10, 2003. In these cases, we manually corrected the price index observations for these NIPA release dates.

[28] In both cases, the ALFRED data contain errors. The January 10, 2014, vintage is erroneously labeled, so we drop it. In addition, the vintages from August 29, 2013, through December 20, 2013, are missing. We insert these data using data collected in real time from the BEA.

vintages dated September 30, 1999, through January 30, 2013, we collected real-time data on real DFP from the BEA via Haver Analytics.[29]

Appendix B: Pesaran-Timmermann Success Ratio Results

Given the limitations of the DM test identified by Clark and McCracken (2009) when using real-time data, we present results from the directional forecast accuracy approach of Pesaran and Timmermann (PT) (2009) and show success ratios that measure the frequency with which, for a given forecast horizon i, the forecast $\hat{g}(GDP_{t-j+i})$ and the third release $g(GDP_{t-j+i})$ are both either greater than or less than the most recently available real-time GDP growth rate at the time of the forecast, $g(GDP_{t-j})$. Tables B1 through B3 show the results using the SPF survey dates, while Tables B4 through B6 show the results using the BC survey dates, estimating the models described in Section III with expanding windows of data. Entries highlighted in green in the tables show that a model's success ratio at a given forecast horizon is greater than the success ratio from the AR(2) benchmark.

Based on the SPF survey dates, we find a large number of cases in which the success ratios from our NIPA aggregate models are greater than those from the AR(2) benchmark during the full sample at the 1- and 2-step horizons. In some cases, especially when using the RW-AR(p) approach with DFP, we also document success ratios that are greater than those from the SPF survey. We find fewer such results at the 3-step horizon, when the models' success ratios are generally comparable with those from the AR(2) benchmark. Similar results hold for the sample considering only expansions. For the sample including only recessions, beyond the 1-

[29] We date the DFP vintages based on the corresponding real series release dates as recorded in ALFRED. In three cases (October 28, 1999; April 3, 2000; and December 10, 2003), we manually corrected observations that were not correctly captured in the Haver Analytics database.

step horizon, most success ratios that are greater than the success ratios from the AR(2) benchmark come from the RW models.[30]

Based on the BC survey dates, for the entire sample we again find a large number of success ratios from our NIPA aggregate models that are greater than those from the AR(2) benchmark, especially at the 2- and 3-step horizon. Interestingly, most of these improvements on the AR(2) model based on success ratios come from using DFP, PDFP, and PCE, not from using GDI or GDA—even though the DM test results in Table 4 documented more statistically significant improvements over the AR(2) benchmark from GDI and GDA than from the other NIPA aggregates. The results from the sample considering only expansions are again similar to those from the entire sample. For the sample considering only recessions, we document a considerable number of cases at the 2- and 3-step horizon in which the success ratios are greater than those from the AR(2) benchmark, with the largest improvements again coming from the RW model.[31]

Appendix C: Results Using Rolling Windows

Our baseline results use expanding windows with the first observation in 1964:Q2 to estimate the parameters in equations (2), (3), and (4). For the sake of comparison, we show results if we instead estimated the parameters using 40-quarter rolling windows. Tables C1

[30] We omit stars denoting statistical significance based on the PT test because some of the tables would have stars for nearly every entry. For the full sample, the PT *p*-values are less than 0.1 for 69 of the 78 model combinations at the 1-step horizon, for 76 combinations at the 2-step horizon, and for 73 combinations at the 3-step horizon. Results are similar for the sample considering only expansions. By contrast, when we consider only recessions, the PT *p*-values are less than 0.1 for 23 of the 78 combinations at the 1-step horizon, for 2 combinations at the 2-step horizon, and for 0 combinations at the 3-step horizon.

[31] Similar to the SPF survey results, statistically significant results based on the PT test are widespread. For the entire sample, the number of combinations (out of 78) having a *p*-value less than 0.1 is 62 at the 1-step horizon, 72 at the 2-step horizon, and 72 at the 3-step horizon. Results are similar during expansions. During recessions, the numbers fall to 11 at the 1-step horizon, 2 at the 2-step horizon, and 0 at the 3-step horizon.

through C3 show the results using the SPF survey dates. Tables C4 through C6 show the results using the BC survey dates. For comparison with Tables 1-6, note that we show RMSFEs relative to the same benchmark model as above—i.e., the AR(2) for GDP growth is estimated using an expanding window. In addition, we show the results from the RW models, which involve no estimation and are identical to the values reported in Tables 1-6.

Table 1: Relative RMSFEs for Real GDP Forecasts at SPF Survey Dates, Full Sample

	1-Step Horizon						2-Step Horizon						3-Step Horizon					
Method	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**
RW	1.32	1.32	1.13	1.05	1.16	1.07	1.30	1.33	1.26	1.17	1.25	1.10	1.28	1.34	1.21	1.21	1.30	1.12
RW-AR(1)	1.04	1.02	1.00	0.93*	0.95	1.00	1.01	1.01	1.01	0.99	1.01	1.03	1.01	1.01	1.01	1.01	1.03	1.03
RW-AR(2)	1.04	1.02	1.01	0.94*	0.96	0.97	1.01	1.01	1.01	0.97	0.99	0.98	1.00	1.00	0.99*	0.99	1.01	1.01
RW-AR(3)	1.04	1.03	1.01	0.93*	0.97	0.95	1.01	1.01	1.00	0.97	1.00	0.96	1.00	1.00	0.99	0.99	1.02	0.99
RW-AR(4)	1.03	1.03	1.01	0.93*	0.96	0.95	1.00	1.00	1.01	0.97	1.00	0.96	1.00	1.00	0.99	0.99	1.02	1.00
Direct(1)	1.03	1.02	0.99	0.92	0.93	0.97	1.01	1.01	1.02	0.97	0.96	0.97	1.00	1.00	1.00	1.00	0.99	0.98
Direct(2)	1.03	1.02	1.02	0.93	0.93	0.94	1.00	1.00	1.03	0.98	0.96	0.95	1.00	1.02	1.00	1.00	0.99	0.99
Direct(3)	1.03	1.02	1.05	0.94	0.93	0.93	1.01	1.03	1.02	0.98	0.97	0.96	1.01	1.02	1.00	1.00	0.99	0.99
Direct(4)	1.04	1.05	1.05	0.95	0.94	0.94	1.02	1.03	1.01	0.97	0.96	0.95	1.02	1.02	1.01	1.00	0.99	0.98
VAR(1)	1.42	1.10	1.00	0.94	0.94	0.97	1.03	1.01	1.01	0.99	0.98*	0.99	1.02	1.01	1.01	1.00	0.99	1.00
VAR(2)	1.43	1.13	1.07	0.94	0.94	0.95	1.01	1.01	1.02	0.97	0.96	0.96	1.00	1.00	1.00	0.99	0.98	0.98
VAR(3)	1.44	1.12	1.08	0.94	0.94	0.94	1.00	1.00	1.02	0.97	0.96	0.94	1.05	1.01	1.01	1.00	1.00	0.99
VAR(4)	1.45	1.11	1.08	0.95	0.94	0.95	1.00	0.99	1.02	0.98	0.95	0.95	1.04	1.00	1.01	1.00	0.99	1.00
SPF	0.82*						0.91						0.97					
AR(2) Benchmark	2.16						2.26						2.29					

Notes: The AR(2) benchmark reports root mean squared forecast errors (RMSFEs) for quarterly annualized real GDP growth; all other statistics are relative to the AR(2) benchmark. For model relative RMSFEs less than 1 (shaded in green), *, **, *** denote rejection of the null hypothesis of equal predictive accuracy of the model and the benchmark at the 10%, 5%, and 1% level, respectively, based on the Diebold and Mariano (1995) test with the small-sample correction of Harvey et al. (1997). GDI is gross domestic income. GDA is the average of GDP and GDI. FSDP is final sales of domestic product, or GDP minus the change in inventories. DFP is domestic final purchases, or FSDP minus net exports. PDFP is private domestic final purchases, or DFP minus government. PCE is personal consumption expenditures. Forecasting methods using expanding windows are described in Section III.

Table 2: Relative RMSFEs for Real GDP Forecasts at SPF Survey Dates, Expansion-Only Sample

	One-Step Horizon						Two-Step Horizon						Three-Step Horizon					
Method	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**
RW	1.45	1.46	1.27	1.20	1.35	1.24	1.54	1.58	1.41	1.38	1.50	1.28	1.56	1.66	1.44	1.49	1.63	1.31
RW-AR(1)	0.97	0.96	0.96	0.95	1.02	0.96	0.97	0.97	0.97	0.97	0.99	0.97	0.99	1.00	0.99	0.99	1.01	1.00
RW-AR(2)	1.00	0.99	1.00	0.97	1.04	0.95	0.98	0.98	1.00	1.01	1.04	0.97	0.99	1.00	0.98**	1.01	1.02	0.98
RW-AR(3)	1.01	1.03	0.99	0.97	1.05	0.94	1.00	1.03	1.01	1.01	1.07	0.96	1.00	1.03	0.99	1.02	1.07	1.00
RW-AR(4)	1.01	1.02	0.99	0.97	1.04	0.95	0.99	1.01	1.01	1.01	1.05	0.96	1.00	1.01	1.00	1.03	1.06	1.00
Direct(1)	0.99	0.97	1.03	0.98	1.02	1.03	0.98	0.98	1.04	1.02	1.02	1.02	0.99	1.01	0.99	0.99	0.99	0.99
Direct(2)	0.99	0.98	1.07	1.00	1.03	1.04	0.98	0.99	1.05	1.02	1.02	1.00	1.01	1.02	0.99	1.00	1.01	1.01
Direct(3)	0.99	0.98	1.10	1.00	1.03	1.04	1.00	1.02	1.04	1.02	1.02	1.01	1.02	1.04	1.00	1.01	1.01	1.01
Direct(4)	1.02	1.04	1.11	1.02	1.04	1.05	1.01	1.04	1.04	1.02	1.02	1.01	1.05	1.06	1.02	1.01	1.01	1.01
VAR(1)	1.29	1.00	1.04	1.01	1.05	1.04	0.97	0.97	0.97	0.98*	0.98	0.97	1.00	1.00	0.99	0.99	0.99	0.99
VAR(2)	1.33	1.06	1.12	1.03	1.07	1.06	1.02	1.01	1.05	1.03	1.02	1.01	1.00	0.99	1.00	1.01	1.01	0.99
VAR(3)	1.36	1.05	1.14	1.02	1.06	1.04	1.01	1.00	1.04	1.02	1.02	0.99	1.06	1.01	1.01	1.01	1.03	1.01
VAR(4)	1.43	1.09	1.14	1.03	1.06	1.06	1.03	1.02	1.04	1.03	1.02	1.00	1.07	1.02	1.01	1.03	1.04	1.03
SPF	0.94						1.03						1.06					
AR(2) Benchmark	1.86						1.85						1.80					

Notes: The AR(2) benchmark reports root mean squared forecast errors (RMSFEs) for quarterly annualized real GDP growth; all other statistics are relative to the AR(2) benchmark. For model relative RMSFEs less than 1 (shaded in green), *, **, *** denote rejection of the null hypothesis of equal predictive accuracy of the model and the benchmark at the 10%, 5%, and 1% level, respectively, based on the Diebold and Mariano (1995) test with the small-sample correction of Harvey et al. (1997). GDI is gross domestic income. GDA is the average of GDP and GDI. FSDP is final sales of domestic product, or GDP minus the change in inventories. DFP is domestic final purchases, or FSDP minus net exports. PDFP is private domestic final purchases, or DFP minus government. PCE is personal consumption expenditures. Forecasting methods using expanding windows are described in Section III.

Table 3: Relative RMSFEs for Real GDP Forecasts at SPF Survey Dates, Recession-Only Sample

	One-Step Horizon						Two-Step Horizon						Three-Step Horizon					
Method	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**
RW	1.03	1.02	0.79	0.68	0.63	0.61	0.84	0.82	0.98	0.77	0.78	0.76	0.84	0.79	0.85	0.74	0.74	0.84
RW-AR(1)	1.16	1.12	1.06	0.89	0.82	1.09	1.08	1.06	1.07	1.01	1.03	1.11	1.03	1.02	1.03	1.02	1.07	1.07
RW-AR(2)	1.10	1.09	1.04	0.87	0.80	1.00	1.04	1.04	1.01	0.93	0.92	1.01	1.01	1.01	1.00	0.97	1.00	1.04
RW-AR(3)	1.08	1.03	1.03	0.86	0.79	0.97	1.02	0.97	1.00	0.92	0.91	0.96	1.01	0.97	0.99	0.96	0.97	0.98
RW-AR(4)	1.08	1.05	1.03	0.85	0.78	0.96	1.02	0.99	1.00	0.91	0.91	0.97	1.00	0.99	0.98	0.95	0.96	0.99
Direct(1)	1.11	1.11	0.91	0.80	0.72	0.82	1.05	1.04	0.99	0.91	0.88	0.89	1.00	0.99	1.02	1.00	0.98	0.97
Direct(2)	1.11	1.10	0.91	0.79	0.71	0.70	1.04	1.02	1.00	0.92	0.88	0.87	1.00	1.01	1.01	0.98	0.97	0.96
Direct(3)	1.10	1.08	0.93	0.80	0.72	0.69	1.04	1.04	0.99	0.91	0.88	0.87	1.00	1.00	1.01	0.98***	0.97	0.96
Direct(4)	1.08	1.08	0.92	0.78	0.71	0.69	1.02	1.03	0.97	0.90	0.87	0.86	0.98	0.99	0.99	0.97**	0.96	0.95
VAR(1)	1.63	1.26	0.92	0.78	0.70	0.83	1.11	1.07	1.05	1.00	0.98	1.02	1.04	1.03	1.02	1.00	0.99	1.02
VAR(2)	1.61	1.26	0.96	0.75	0.64	0.71	1.01	1.01	0.99	0.89	0.86	0.87	1.01	1.01	0.99	0.96	0.95	0.97
VAR(3)	1.59	1.23	0.96	0.77	0.65	0.69	0.99	0.99	0.99	0.91	0.86	0.87	1.04	1.01	1.01	0.99	0.95	0.97
VAR(4)	1.50	1.17	0.96	0.77	0.66	0.72	0.96	0.95	0.99	0.90	0.85	0.87	0.99	0.97	1.01	0.97	0.92	0.96
SPF	0.52						0.71						0.86					
AR(2) Benchmark	3.82						4.33						4.60					

Notes: The AR(2) benchmark reports root mean squared forecast errors (RMSFEs) for quarterly annualized real GDP growth; all other statistics are relative to the AR(2) benchmark. For model relative RMSFEs less than 1 (shaded in green), *, **, *** denote rejection of the null hypothesis of equal predictive accuracy of the model and the benchmark at the 10%, 5%, and 1% level, respectively, based on the Diebold and Mariano (1995) test with the small-sample correction of Harvey et al. (1997). GDI is gross domestic income. GDA is the average of GDP and GDI. FSDP is final sales of domestic product, or GDP minus the change in inventories. DFP is domestic final purchases, or FSDP minus net exports. PDFP is private domestic final purchases, or DFP minus government. PCE is personal consumption expenditures. Forecasting methods using expanding windows are described in Section III.

Table 4: Relative RMSFEs for Real GDP Forecasts at Blue Chip Survey Dates, Full Sample

	1-Step Horizon						2-Step Horizon						3-Step Horizon					
Method	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**
RW	1.11	1.15	1.16	1.04	1.18	1.05	1.26	1.25	1.30	1.17	1.27	1.09	1.28	1.29	1.30	1.23	1.34	1.11
RW-AR(1)	0.96*	0.95	1.00	0.92**	0.96	1.00	0.99	0.97***	1.00	0.97*	0.99	1.01	1.00	1.00	1.01	1.00	1.03	1.03
RW-AR(2)	0.96*	0.95*	1.01	0.93*	0.97	0.97	0.98	0.97**	1.00	0.96	0.98	0.96	1.00	1.00	1.00	0.99	1.02	1.01
RW-AR(3)	0.96*	0.96	1.00	0.93*	0.98	0.95	0.98	0.98	0.99	0.96	0.99	0.94	1.00	1.00	1.00	0.99	1.03	0.98
RW-AR(4)	0.96*	0.95	1.00	0.92**	0.97	0.95	0.98	0.97	1.00	0.96	0.98	0.94*	1.00	1.00	1.00	0.99	1.02	0.99
Direct(1)	0.96*	0.96	0.99	0.91	0.92	0.95	0.99**	0.98***	1.02	0.96	0.95	0.95	1.00	1.00	1.01	0.99	0.99	0.97
Direct(2)	0.96**	0.95	1.01	0.93	0.93	0.93	0.98**	0.97**	1.03	0.96	0.95	0.93	1.00	1.00	1.00	0.99	0.99	0.98
Direct(3)	0.96*	0.95	1.04	0.93	0.93	0.92	0.98*	0.97	1.02	0.96	0.95	0.94	1.01	1.02	1.00	1.00	0.99	0.98
Direct(4)	0.96**	0.95	1.05	0.94	0.93	0.93	0.99	1.01	1.01	0.96	0.95	0.93	1.01	1.03	1.01	0.99	0.99	0.98
VAR(1)	0.97	0.97	1.00	0.92	0.94	0.95	0.97***	0.97***	1.00	0.97**	0.97**	0.97*	1.00	1.00	1.01	1.00	0.99	1.00
VAR(2)	1.00	1.01	1.06	0.93	0.95	0.94	1.01	1.01	1.02	0.95	0.93	0.94	1.00	1.00	1.00	0.99	0.98	0.98
VAR(3)	1.01	1.01	1.08	0.94	0.95	0.93	1.00	1.00	1.02	0.95	0.94	0.92	1.00	1.00	1.01	1.00	1.00	0.98
VAR(4)	1.01	1.02	1.08	0.94	0.95	0.95	0.99	0.99	1.02	0.95	0.93	0.92	0.99	0.99	1.01	1.00	0.99	0.99
AR(2) Benchmark	2.15						2.30						2.29					

Notes: The AR(2) benchmark reports root mean squared forecast errors (RMSFEs) for quarterly annualized real GDP growth; all other statistics are relative to the AR(2) benchmark. For model relative RMSFEs less than 1 (shaded in green), *, **, *** denote rejection of the null hypothesis of equal predictive accuracy of the model and the benchmark at the 10%, 5%, and 1% level, respectively, based on the Diebold and Mariano (1995) test with the small-sample correction of Harvey et al. (1997). GDI is gross domestic income. GDA is the average of GDP and GDI. FSDP is final sales of domestic product, or GDP minus the change in inventories. DFP is domestic final purchases, or FSDP minus net exports. PDFP is private domestic final purchases, or DFP minus government. PCE is personal consumption expenditures. Forecasting methods using expanding windows are described in Section III.

Table 5: Relative RMSFEs for Real GDP Forecasts at Blue Chip Survey Dates, Expansion-Only Sample

	1-Step Horizon						2-Step Horizon						3-Step Horizon					
Method	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**
RW	1.26	1.31	1.30	1.18	1.37	1.21	1.44	1.43	1.45	1.36	1.49	1.25	1.57	1.59	1.54	1.51	1.68	1.29
RW-AR(1)	0.98	1.00	0.96	0.94*	1.02	0.95	0.96***	0.96***	0.95**	0.94***	0.97	0.94*	1.00	1.00	1.00	0.99	1.01	1.00
RW-AR(2)	1.00	1.00	0.99	0.96	1.04	0.95*	0.99	0.98	0.99	0.98	1.02	0.94*	1.01	1.01	0.99	1.00	1.03	0.98
RW-AR(3)	1.00	1.03	0.98	0.96	1.05	0.94*	1.00	1.02	0.99	0.98	1.05	0.93*	1.03	1.06	1.00	1.02	1.07	0.99
RW-AR(4)	1.00	1.02	0.98	0.96	1.05	0.94*	1.00	1.00	0.99	0.98	1.03	0.93*	1.03	1.04	1.01	1.02	1.06	0.99
Direct(1)	0.98	1.00	1.03	0.97	1.01	1.01	0.99	0.97**	1.03	0.99	0.99	0.99	1.00	1.00	0.99	0.99	0.99	0.98
Direct(2)	0.99	1.00	1.06	0.99	1.02	1.02	0.99	0.97*	1.04	0.99	0.99	0.98	1.01	1.02	0.99	1.00	1.01	1.00
Direct(3)	0.99	1.01	1.09	0.99	1.03	1.02	0.99	0.98	1.03	1.00	1.00	0.99	1.03	1.06	1.00	1.01	1.01	1.00
Direct(4)	0.99	1.00	1.10	1.00	1.03	1.04	1.01	1.03	1.03	0.99	0.99	0.98	1.05	1.07	1.02	1.01	1.01	1.00
VAR(1)	1.02	1.02	1.04	0.98	1.04	1.02	0.95***	0.96***	0.96**	0.96***	0.96**	0.95**	1.00	1.00	0.99	0.99	0.99	0.99
VAR(2)	1.07	1.07	1.10	1.00	1.07	1.04	1.01	1.01	1.04	0.98	0.98	0.98	1.00	1.00	1.00	1.01	1.01	0.98*
VAR(3)	1.07	1.08	1.12	1.00	1.06	1.02	1.01	1.01	1.04	0.97	0.98	0.95	1.02	1.02	1.00	1.01	1.03	1.00
VAR(4)	1.09	1.10	1.13	1.00	1.06	1.04	1.03	1.03	1.03	0.98	0.98	0.96	1.05	1.04	1.01	1.03	1.04	1.02
AR(2) Benchmark	1.87						1.90						1.80					

Notes: The AR(2) benchmark reports root mean squared forecast errors (RMSFEs) for quarterly annualized real GDP growth; all other statistics are relative to the AR(2) benchmark. For model relative RMSFEs less than 1 (shaded in green), *, **, *** denote rejection of the null hypothesis of equal predictive accuracy of the model and the benchmark at the 10%, 5%, and 1% level, respectively, based on the Diebold and Mariano (1995) test with the small-sample correction of Harvey et al. (1997). GDI is gross domestic income. GDA is the average of GDP and GDI. FSDP is final sales of domestic product, or GDP minus the change in inventories. DFP is domestic final purchases, or FSDP minus net exports. PDFP is private domestic final purchases, or DFP minus government. PCE is personal consumption expenditures. Forecasting methods using expanding windows are described in Section III.

Table 6: Relative RMSFEs for Real GDP Forecasts at Blue Chip Survey Dates, Recession-Only Sample

	1-Step Horizon						2-Step Horizon						3-Step Horizon					
Method	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**
RW	0.74	0.70	0.80	0.69	0.65	0.58	0.91	0.90	1.03	0.80	0.82	0.76	0.80	0.79	0.92***	0.76	0.76	0.86
RW-AR(1)	0.90	0.83*	1.07	0.89	0.82	1.09	1.02	0.99	1.07	1.01	1.03	1.10	1.01	1.00	1.03	1.02	1.06	1.07
RW-AR(2)	0.89*	0.82*	1.05	0.87	0.80	1.01	0.97*	0.96	1.01	0.93	0.92	1.00	0.98	0.98	1.01	0.97	1.00	1.04
RW-AR(3)	0.87*	0.79*	1.05	0.87	0.80	0.98	0.95	0.91	1.00	0.92	0.91	0.96	0.96	0.92	0.99	0.96	0.97	0.98
RW-AR(4)	0.87**	0.79**	1.05	0.85	0.79	0.97	0.95**	0.92*	1.01	0.91	0.90	0.96	0.96*	0.93**	0.99	0.95	0.96	0.99
Direct(1)	0.92	0.85*	0.91	0.79	0.71	0.80	0.99*	0.99*	1.00	0.91	0.88	0.87	1.00	0.99**	1.02	1.00	0.98	0.97
Direct(2)	0.91*	0.84*	0.92	0.79	0.71	0.69	0.98*	0.98	1.01	0.92	0.88	0.85	0.98	0.97	1.01	0.98	0.97	0.96
Direct(3)	0.91*	0.84*	0.94	0.80	0.71	0.68	0.97	0.96	1.00	0.91	0.88	0.86	0.98*	0.98	1.01	0.98*	0.97	0.96
Direct(4)	0.89*	0.83*	0.92	0.78	0.70	0.68	0.97	0.97	0.98	0.90	0.87	0.85	0.97	0.97	0.99	0.97	0.96	0.95
VAR(1)	0.84*	0.84*	0.92	0.79	0.70	0.81	1.00	1.00	1.05	1.00	0.98	1.01	1.00	1.00	1.02	1.00	0.99	1.02
VAR(2)	0.87*	0.87*	0.97	0.78	0.67	0.71	1.01	1.01	0.99	0.89	0.85	0.86	0.99***	0.99***	0.99	0.96	0.95	0.97
VAR(3)	0.85	0.85	0.98	0.79	0.67	0.70	0.98	0.98	1.00	0.91	0.86	0.86	0.97	0.97	1.01	0.99	0.95	0.96
VAR(4)	0.83*	0.83*	0.97	0.79	0.68	0.72	0.93	0.94	1.00	0.91	0.84	0.86	0.92	0.92	1.00	0.96	0.92	0.95
AR(2) Benchmark			3.72						4.34						4.61			

Notes: The AR(2) benchmark reports root mean squared forecast errors (RMSFEs) for quarterly annualized real GDP growth; all other statistics are relative to the AR(2) benchmark. For model relative RMSFEs less than 1 (shaded in green), *, **, *** denote rejection of the null hypothesis of equal predictive accuracy of the model and the benchmark at the 10%, 5%, and 1% level, respectively, based on the Diebold and Mariano (1995) test with the small-sample correction of Harvey et al. (1997). GDI is gross domestic income. GDA is the average of GDP and GDI. FSDP is final sales of domestic product, or GDP minus the change in inventories. DFP is domestic final purchases, or FSDP minus net exports. PDFP is private domestic final purchases, or DFP minus government. PCE is personal consumption expenditures. Forecasting methods using expanding windows are described in Section III.

Table B1: Pesaran-Timmermann Success Ratios at SPF Survey Dates, Full Sample

	1-Step Horizon						2-Step Horizon						3-Step Horizon					
Method	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**
RW	0.60	0.61	0.56	0.61	0.60	0.53	0.64	0.65	0.54	0.57	0.62	0.65	0.54	0.53	0.59	0.54	0.51	0.54
RW-AR(1)	0.66	0.67	0.68	0.70	0.66	0.68	0.69	0.71	0.68	0.73	0.73	0.70	0.67	0.67	0.68	0.69	0.65	0.67
RW-AR(2)	0.67	0.68	0.65	0.75	0.67	0.66	0.71	0.71	0.69	0.73	0.72	0.71	0.66	0.66	0.66	0.67	0.64	0.67
RW-AR(3)	0.67	0.69	0.67	0.75	0.67	0.68	0.70	0.70	0.72	0.76	0.70	0.75	0.66	0.68	0.66	0.65	0.63	0.64
RW-AR(4)	0.66	0.64	0.65	0.76	0.66	0.65	0.73	0.72	0.72	0.76	0.72	0.77	0.67	0.66	0.67	0.66	0.62	0.65
Direct(1)	0.67	0.67	0.63	0.70	0.69	0.63	0.71	0.72	0.65	0.73	0.70	0.70	0.66	0.66	0.66	0.66	0.65	0.64
Direct(2)	0.67	0.66	0.65	0.74	0.69	0.68	0.73	0.71	0.66	0.70	0.70	0.74	0.69	0.67	0.63	0.65	0.66	0.65
Direct(3)	0.69	0.65	0.60	0.73	0.70	0.66	0.70	0.70	0.65	0.69	0.74	0.70	0.65	0.66	0.66	0.66	0.67	0.65
Direct(4)	0.64	0.61	0.59	0.70	0.71	0.63	0.70	0.69	0.64	0.70	0.71	0.70	0.65	0.66	0.66	0.67	0.67	0.63
VAR(1)	0.68	0.67	0.61	0.69	0.71	0.62	0.72	0.71	0.69	0.73	0.72	0.69	0.67	0.68	0.68	0.66	0.67	0.68
VAR(2)	0.68	0.64	0.61	0.68	0.69	0.66	0.67	0.69	0.65	0.73	0.73	0.73	0.66	0.67	0.67	0.65	0.66	0.67
VAR(3)	0.68	0.70	0.59	0.67	0.69	0.66	0.72	0.72	0.63	0.73	0.73	0.75	0.68	0.68	0.63	0.64	0.66	0.66
VAR(4)	0.66	0.63	0.61	0.70	0.68	0.63	0.67	0.68	0.62	0.72	0.73	0.74	0.66	0.66	0.65	0.64	0.66	0.66
SPF	0.69						0.75						0.66					
AR(2) Benchmark	0.64						0.70						0.67					

Notes: Model success ratios whose values are greater than the AR(2) benchmarks at a particular horizon are shaded in green. GDI is gross domestic income. GDA is the average of GDP and GDI. FSDP is final sales of domestic product, or GDP minus the change in inventories. DFP is domestic final purchases, or FSDP minus net exports. PDFP is private domestic final purchases, or DFP minus government. PCE is personal consumption expenditures. Forecasting methods using expanding windows are described in Section III. Pesaran-Timmermann success ratios are described in Appendix B.

Table B2: Pesaran-Timmermann Success Ratios at SPF Survey Dates, Expansion-Only Sample

	1-Step Horizon						2-Step Horizon						3-Step Horizon					
Method	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**
RW	0.62	0.63	0.56	0.61	0.60	0.52	0.65	0.66	0.52	0.56	0.63	0.65	0.55	0.55	0.60	0.55	0.52	0.55
RW-AR(1)	0.69	0.70	0.71	0.73	0.69	0.71	0.72	0.74	0.71	0.76	0.76	0.74	0.72	0.72	0.74	0.75	0.71	0.74
RW-AR(2)	0.70	0.71	0.67	0.79	0.69	0.69	0.74	0.74	0.72	0.76	0.75	0.74	0.71	0.71	0.71	0.72	0.69	0.72
RW-AR(3)	0.70	0.71	0.70	0.79	0.69	0.71	0.73	0.73	0.75	0.80	0.73	0.78	0.71	0.74	0.71	0.70	0.68	0.69
RW-AR(4)	0.69	0.66	0.67	0.80	0.67	0.67	0.76	0.75	0.75	0.80	0.75	0.81	0.72	0.71	0.72	0.71	0.67	0.70
Direct(1)	0.70	0.70	0.65	0.73	0.72	0.64	0.74	0.75	0.67	0.76	0.73	0.73	0.71	0.71	0.71	0.71	0.70	0.69
Direct(2)	0.70	0.69	0.67	0.76	0.71	0.69	0.76	0.74	0.68	0.73	0.73	0.77	0.75	0.72	0.68	0.70	0.71	0.70
Direct(3)	0.72	0.67	0.62	0.76	0.72	0.66	0.73	0.73	0.67	0.72	0.77	0.73	0.70	0.71	0.71	0.71	0.71	0.70
Direct(4)	0.66	0.64	0.61	0.73	0.73	0.63	0.73	0.72	0.67	0.73	0.74	0.73	0.70	0.72	0.70	0.72	0.71	0.68
VAR(1)	0.71	0.70	0.63	0.72	0.72	0.63	0.75	0.74	0.72	0.76	0.75	0.72	0.72	0.74	0.74	0.71	0.72	0.74
VAR(2)	0.71	0.66	0.63	0.70	0.70	0.66	0.69	0.72	0.67	0.76	0.76	0.76	0.71	0.72	0.72	0.70	0.71	0.72
VAR(3)	0.71	0.73	0.61	0.69	0.70	0.66	0.75	0.75	0.65	0.76	0.76	0.76	0.74	0.74	0.68	0.69	0.71	0.71
VAR(4)	0.67	0.64	0.63	0.72	0.70	0.63	0.71	0.72	0.65	0.75	0.75	0.76	0.71	0.71	0.70	0.69	0.71	0.71
SPF	0.69						0.75						0.70					
AR(2) Benchmark	0.66						0.73						0.72					

Notes: Model success ratios whose values are greater than the AR(2) benchmarks at a particular horizon are shaded in green. GDI is gross domestic income. GDA is the average of GDP and GDI. FSDP is final sales of domestic product, or GDP minus the change in inventories. DFP is domestic final purchases, or FSDP minus net exports. PDFP is private domestic final purchases, or DFP minus government. PCE is personal consumption expenditures. Forecasting methods using expanding windows are described in Section III. Pesaran-Timmermann success ratios are described in Appendix B.

Table B3: Pesaran-Timmermann Success Ratios at SPF Survey Dates, Recession-Only Sample

	1-Step Horizon						2-Step Horizon						3-Step Horizon					
Method	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**
RW	0.46	0.46	0.55	0.64	0.64	0.64	0.55	0.55	0.64	0.64	0.55	0.64	0.46	0.36	0.55	0.46	0.46	0.46
RW-AR(1)	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.36	0.27	0.27	0.27	0.27	0.18	0.18
RW-AR(2)	0.46	0.46	0.46	0.46	0.55	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.27	0.27	0.27	0.27	0.27	0.27
RW-AR(3)	0.46	0.55	0.46	0.46	0.55	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.27	0.27	0.27	0.27	0.27	0.27
RW-AR(4)	0.46	0.46	0.46	0.46	0.55	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.27	0.27	0.27	0.27	0.27	0.27
Direct(1)	0.46	0.46	0.46	0.46	0.46	0.55	0.46	0.46	0.46	0.46	0.46	0.46	0.27	0.27	0.27	0.27	0.27	0.27
Direct(2)	0.46	0.46	0.46	0.55	0.55	0.64	0.46	0.46	0.46	0.46	0.46	0.46	0.27	0.27	0.27	0.27	0.27	0.27
Direct(3)	0.46	0.46	0.46	0.46	0.55	0.64	0.46	0.46	0.46	0.46	0.46	0.46	0.27	0.27	0.27	0.27	0.36	0.27
Direct(4)	0.46	0.36	0.46	0.46	0.55	0.64	0.46	0.46	0.36	0.46	0.46	0.46	0.27	0.18	0.36	0.27	0.36	0.27
VAR(1)	0.46	0.46	0.46	0.46	0.64	0.55	0.46	0.46	0.46	0.46	0.46	0.46	0.27	0.27	0.27	0.27	0.27	0.27
VAR(2)	0.46	0.46	0.46	0.55	0.64	0.64	0.46	0.46	0.46	0.46	0.46	0.46	0.27	0.27	0.27	0.27	0.27	0.27
VAR(3)	0.46	0.46	0.46	0.55	0.64	0.64	0.46	0.46	0.46	0.46	0.46	0.64	0.27	0.27	0.27	0.27	0.27	0.27
VAR(4)	0.55	0.55	0.46	0.55	0.55	0.64	0.36	0.36	0.36	0.46	0.55	0.55	0.27	0.27	0.27	0.27	0.27	0.27
SPF	0.73						0.73						0.36					
AR(2) Benchmark	0.46						0.46						0.27					

Notes: Model success ratios whose values are greater than the AR(2) benchmarks at a particular horizon are shaded in green. GDI is gross domestic income. GDA is the average of GDP and GDI. FSDP is final sales of domestic product, or GDP minus the change in inventories. DFP is domestic final purchases, or FSDP minus net exports. PDFP is private domestic final purchases, or DFP minus government. PCE is personal consumption expenditures. Forecasting methods using expanding windows are described in Section III. Pesaran-Timmermann success ratios are described in Appendix B.

Table B4: Pesaran-Timmermann Success Ratios at Blue Chip Survey Dates, Full Sample

	1-Step Horizon						2-Step Horizon						3-Step Horizon					
Method	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**
RW	0.53	0.53	0.56	0.67	0.61	0.63	0.61	0.61	0.57	0.62	0.64	0.62	0.61	0.61	0.56	0.63	0.57	0.69
RW-AR(1)	0.68	0.65	0.72	0.74	0.70	0.73	0.70	0.69	0.71	0.75	0.73	0.73	0.69	0.67	0.69	0.72	0.75	0.71
RW-AR(2)	0.70	0.66	0.70	0.75	0.67	0.74	0.70	0.73	0.75	0.77	0.74	0.74	0.68	0.67	0.67	0.72	0.72	0.72
RW-AR(3)	0.67	0.65	0.70	0.73	0.68	0.71	0.73	0.73	0.72	0.75	0.74	0.75	0.69	0.68	0.68	0.72	0.74	0.72
RW-AR(4)	0.66	0.65	0.70	0.75	0.69	0.71	0.74	0.75	0.74	0.76	0.74	0.71	0.67	0.69	0.69	0.71	0.75	0.71
Direct(1)	0.69	0.66	0.68	0.77	0.75	0.70	0.70	0.69	0.75	0.77	0.75	0.73	0.69	0.69	0.69	0.72	0.71	0.71
Direct(2)	0.70	0.66	0.68	0.77	0.71	0.69	0.70	0.72	0.72	0.77	0.75	0.77	0.68	0.69	0.70	0.71	0.74	0.69
Direct(3)	0.70	0.65	0.67	0.75	0.72	0.68	0.70	0.70	0.73	0.77	0.77	0.74	0.71	0.69	0.71	0.69	0.70	0.71
Direct(4)	0.70	0.65	0.63	0.74	0.72	0.68	0.73	0.71	0.72	0.73	0.74	0.74	0.72	0.69	0.70	0.69	0.71	0.68
VAR(1)	0.67	0.67	0.68	0.76	0.75	0.70	0.69	0.69	0.72	0.75	0.75	0.72	0.68	0.68	0.69	0.72	0.71	0.68
VAR(2)	0.65	0.65	0.67	0.76	0.73	0.69	0.69	0.69	0.73	0.77	0.74	0.73	0.70	0.70	0.69	0.72	0.71	0.70
VAR(3)	0.65	0.65	0.61	0.77	0.70	0.69	0.73	0.73	0.71	0.77	0.77	0.76	0.71	0.71	0.74	0.70	0.70	0.70
VAR(4)	0.64	0.63	0.63	0.76	0.68	0.70	0.71	0.71	0.74	0.76	0.76	0.75	0.68	0.68	0.71	0.68	0.68	0.68
AR(2) Benchmark	0.72						0.71						0.69					

Notes: Model success ratios whose values are greater than the AR(2) benchmarks at a particular horizon are shaded in green. GDI is gross domestic income. GDA is the average of GDP and GDI. FSDP is final sales of domestic product, or GDP minus the change in inventories. DFP is domestic final purchases, or FSDP minus net exports. PDFP is private domestic final purchases, or DFP minus government. PCE is personal consumption expenditures. Forecasting methods using expanding windows are described in Section III. Pesaran-Timmermann success ratios are described in Appendix B.

Table B5: Pesaran-Timmermann Success Ratios at Blue Chip Survey Dates, Expansion-Only Sample

	1-Step Horizon						2-Step Horizon						3-Step Horizon					
Method	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**
RW	0.53	0.53	0.55	0.66	0.61	0.61	0.60	0.60	0.56	0.60	0.65	0.61	0.60	0.60	0.56	0.63	0.56	0.71
RW-AR(1)	0.69	0.64	0.73	0.74	0.70	0.74	0.73	0.72	0.74	0.78	0.76	0.76	0.74	0.71	0.74	0.77	0.79	0.76
RW-AR(2)	0.70	0.65	0.71	0.75	0.66	0.74	0.73	0.75	0.78	0.80	0.76	0.77	0.72	0.71	0.71	0.76	0.77	0.77
RW-AR(3)	0.66	0.64	0.70	0.73	0.67	0.71	0.75	0.75	0.75	0.77	0.76	0.77	0.74	0.71	0.72	0.76	0.77	0.76
RW-AR(4)	0.65	0.64	0.70	0.75	0.69	0.71	0.76	0.77	0.77	0.78	0.76	0.73	0.71	0.74	0.74	0.75	0.78	0.76
Direct(1)	0.70	0.65	0.69	0.78	0.75	0.69	0.73	0.72	0.78	0.80	0.77	0.75	0.74	0.74	0.74	0.77	0.76	0.75
Direct(2)	0.70	0.65	0.69	0.78	0.71	0.67	0.73	0.74	0.75	0.80	0.77	0.80	0.72	0.72	0.74	0.75	0.77	0.72
Direct(3)	0.70	0.64	0.67	0.75	0.72	0.66	0.72	0.72	0.76	0.80	0.80	0.76	0.76	0.74	0.76	0.74	0.74	0.75
Direct(4)	0.70	0.64	0.63	0.74	0.72	0.66	0.75	0.74	0.76	0.75	0.76	0.76	0.77	0.74	0.75	0.74	0.75	0.71
VAR(1)	0.66	0.66	0.69	0.76	0.74	0.69	0.72	0.72	0.75	0.78	0.77	0.75	0.72	0.72	0.74	0.77	0.76	0.72
VAR(2)	0.64	0.64	0.67	0.76	0.72	0.67	0.72	0.72	0.76	0.80	0.76	0.75	0.75	0.75	0.74	0.76	0.75	0.74
VAR(3)	0.64	0.64	0.61	0.78	0.69	0.67	0.75	0.75	0.74	0.80	0.80	0.77	0.75	0.75	0.77	0.74	0.74	0.74
VAR(4)	0.63	0.62	0.63	0.76	0.67	0.69	0.73	0.73	0.77	0.78	0.78	0.77	0.71	0.71	0.75	0.71	0.71	0.71
AR(2) Benchmark	0.72						0.74						0.74					

Notes: Model success ratios whose values are greater than the AR(2) benchmarks at a particular horizon are shaded in green. GDI is gross domestic income. GDA is the average of GDP and GDI. FSDP is final sales of domestic product, or GDP minus the change in inventories. DFP is domestic final purchases, or FSDP minus net exports. PDFP is private domestic final purchases, or DFP minus government. PCE is personal consumption expenditures. Forecasting methods using expanding windows are described in Section III. Pesaran-Timmermann success ratios are described in Appendix B.

Table B6: Pesaran-Timmermann Success Ratios at Blue Chip Survey Dates, Recession-Only Sample

	1-Step Horizon						2-Step Horizon						3-Step Horizon					
Method	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**
RW	0.55	0.55	0.64	0.73	0.64	0.82	0.64	0.64	0.64	0.73	0.55	0.64	0.73	0.73	0.55	0.64	0.64	0.55
RW-AR(1)	0.64	0.73	0.64	0.73	0.73	0.64	0.46	0.46	0.46	0.46	0.46	0.46	0.36	0.36	0.36	0.36	0.36	0.36
RW-AR(2)	0.73	0.73	0.64	0.73	0.73	0.73	0.46	0.55	0.46	0.55	0.55	0.46	0.36	0.36	0.36	0.46	0.36	0.36
RW-AR(3)	0.73	0.73	0.73	0.73	0.73	0.73	0.55	0.55	0.46	0.55	0.55	0.55	0.36	0.46	0.36	0.46	0.46	0.46
RW-AR(4)	0.73	0.73	0.73	0.73	0.73	0.73	0.55	0.55	0.46	0.55	0.55	0.55	0.36	0.36	0.36	0.46	0.46	0.36
Direct(1)	0.64	0.73	0.64	0.73	0.73	0.82	0.46	0.46	0.46	0.55	0.55	0.55	0.36	0.36	0.36	0.36	0.36	0.46
Direct(2)	0.73	0.73	0.64	0.73	0.73	0.82	0.46	0.55	0.46	0.55	0.55	0.55	0.36	0.46	0.46	0.46	0.46	0.46
Direct(3)	0.73	0.73	0.64	0.73	0.73	0.82	0.55	0.55	0.46	0.55	0.55	0.55	0.36	0.36	0.36	0.36	0.46	0.46
Direct(4)	0.73	0.73	0.64	0.73	0.73	0.82	0.55	0.46	0.36	0.55	0.55	0.55	0.36	0.36	0.36	0.36	0.46	0.46
VAR(1)	0.73	0.73	0.64	0.73	0.82	0.82	0.46	0.46	0.46	0.46	0.55	0.46	0.36	0.36	0.36	0.36	0.36	0.36
VAR(2)	0.73	0.73	0.64	0.73	0.82	0.82	0.46	0.46	0.46	0.55	0.55	0.55	0.36	0.36	0.36	0.46	0.46	0.46
VAR(3)	0.73	0.73	0.64	0.73	0.82	0.82	0.55	0.55	0.46	0.55	0.55	0.64	0.46	0.46	0.46	0.46	0.46	0.46
VAR(4)	0.73	0.73	0.64	0.73	0.73	0.82	0.55	0.55	0.46	0.55	0.55	0.55	0.46	0.46	0.46	0.46	0.46	0.46
AR(2) Benchmark	0.73						0.46						0.36					

Notes: Model success ratios whose values are greater than the AR(2) benchmarks at a particular horizon are shaded in green. GDI is gross domestic income. GDA is the average of GDP and GDI. FSDP is final sales of domestic product, or GDP minus the change in inventories. DFP is domestic final purchases, or FSDP minus net exports. PDFP is private domestic final purchases, or DFP minus government. PCE is personal consumption expenditures. Forecasting methods using expanding windows are described in Section III. Pesaran-Timmermann success ratios are described in Appendix B.

Table C1: Relative RMSFEs for Real GDP Forecasts at SPF Survey Dates, Full Sample; Estimation Using Rolling Windows

	1-Step Horizon						2-Step Horizon						3-Step Horizon					
Method	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**
RW	1.32	1.32	1.13	1.05	1.16	1.07	1.30	1.33	1.26	1.17	1.25	1.10	1.28	1.34	1.21	1.21	1.30	1.12
RW-AR(1)	1.09	1.04	1.02	0.94	0.96	0.98	1.06	1.02	1.02	1.03	1.04	1.01	1.04	1.02	1.02	1.07	1.08	1.03
RW-AR(2)	1.11	1.09	1.04	0.94	0.98	0.95	1.09	1.08	1.02	1.03	1.06	0.97	1.05	1.06	1.01	1.09	1.11	1.03
RW-AR(3)	1.13	1.10	1.05	0.96	1.00	0.96	1.11	1.09	1.06	1.08	1.13	0.98	1.09	1.07	1.04	1.14	1.19	1.04
RW-AR(4)	1.15	1.16	1.06	0.95	1.01	0.96	1.13	1.15	1.07	1.08	1.14	0.99	1.12	1.13	1.07	1.16	1.20	1.06
Direct(1)	1.07	1.03	1.01	0.95	0.94	0.98	1.03	1.02	1.04	1.01	1.00	0.98	1.06	1.04	1.00	1.03	1.01	1.00
Direct(2)	1.08	1.04	1.05	0.94	0.93	0.97	1.07	1.05	1.06	1.04	1.02	0.98	1.07	1.05	1.06	1.07	1.05	1.02
Direct(3)	1.09	1.05	1.06	0.95	0.93	0.97	1.08	1.07	1.09	1.06	1.04	0.99	1.10	1.09	1.06	1.08	1.06	1.04
Direct(4)	1.10	1.06	1.07	0.95	0.95	0.97	1.11	1.09	1.09	1.06	1.05	1.00	1.10	1.09	1.08	1.09	1.07	1.04
VAR(1)	1.37	1.12	1.01	0.95	0.94	0.98	1.05	1.03	1.03	1.02	1.01	1.00	1.05	1.04	1.01	1.03	1.03	1.00
VAR(2)	1.44	1.15	1.08	0.94	0.91	0.99	1.09	1.05	1.06	1.03	1.01	0.99	1.08	1.04	1.03	1.05	1.03	1.02
VAR(3)	1.45	1.17	1.09	0.94	0.92	0.98	1.11	1.07	1.09	1.05	1.01	0.99	1.12	1.08	1.08	1.11	1.09	1.06
VAR(4)	1.53	1.22	1.12	0.96	0.93	1.00	1.17	1.12	1.12	1.08	1.01	1.00	1.18	1.13	1.14	1.15	1.10	1.10
SPF	0.82*						0.91						0.97					
AR(2) Benchmark	2.16						2.26						2.29					

Notes: The AR(2) benchmark reports root mean squared forecast errors (RMSFEs) for quarterly annualized real GDP growth; all other statistics are relative to the AR(2) benchmark. For model relative RMSFEs less than 1 (shaded in green), *, **, *** denote rejection of the null hypothesis of equal predictive accuracy of the model and the benchmark at the 10%, 5%, and 1% level, respectively, based on the Diebold and Mariano (1995) test with the small-sample correction of Harvey et al. (1997). GDI is gross domestic income. GDA is the average of GDP and GDI. FSDP is final sales of domestic product, or GDP minus the change in inventories. DFP is domestic final purchases, or FSDP minus net exports. PDFP is private domestic final purchases, or DFP minus government. PCE is personal consumption expenditures. Forecasting methods are described in Section III.

Tables C2: Relative RMSFEs for Real GDP Forecasts at SPF Survey Dates, Expansion-Only Sample; Estimation Using Rolling Windows

	One-Step Horizon						Two-Step Horizon						Three-Step Horizon					
Method	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**
RW	1.45	1.46	1.27	1.20	1.35	1.24	1.54	1.58	1.41	1.38	1.50	1.28	1.56	1.66	1.44	1.49	1.63	1.31
RW-AR(1)	1.10	1.03	1.01	0.99	1.04	0.98	1.11	1.05	1.02	1.08	1.09	0.97	1.10	1.08	1.04	1.13	1.15	1.01
RW-AR(2)	1.14	1.12	1.02	0.98	1.06	0.98	1.15	1.16	1.04	1.13	1.17	0.99	1.11	1.14	1.03	1.22	1.23	1.03
RW-AR(3)	1.19	1.15	1.03	0.99	1.07	1.00	1.20	1.18	1.11	1.21	1.28	1.03	1.19	1.16	1.11	1.33	1.40	1.11
RW-AR(4)	1.23	1.25	1.05	1.00	1.09	1.00	1.25	1.29	1.14	1.23	1.30	1.04	1.25	1.28	1.17	1.37	1.42	1.16
Direct(1)	1.08	1.02	1.04	1.00	1.03	1.07	1.06	1.05	1.08	1.09	1.09	1.06	1.13	1.07	1.03	1.11	1.09	1.04
Direct(2)	1.09	1.05	1.10	1.01	1.03	1.09	1.13	1.09	1.11	1.15	1.12	1.06	1.15	1.12	1.14	1.20	1.16	1.09
Direct(3)	1.12	1.06	1.10	1.01	1.02	1.09	1.15	1.13	1.17	1.18	1.15	1.08	1.18	1.14	1.13	1.21	1.17	1.11
Direct(4)	1.13	1.08	1.13	1.02	1.05	1.10	1.17	1.15	1.17	1.18	1.17	1.09	1.17	1.13	1.16	1.21	1.18	1.11
VAR(1)	1.33	1.09	1.03	1.02	1.04	1.07	1.11	1.09	1.04	1.08	1.08	1.03	1.12	1.10	1.03	1.11	1.12	1.04
VAR(2)	1.38	1.13	1.12	1.01	1.01	1.11	1.14	1.10	1.12	1.13	1.10	1.08	1.15	1.11	1.08	1.18	1.15	1.10
VAR(3)	1.41	1.16	1.12	1.01	1.02	1.09	1.17	1.14	1.15	1.18	1.13	1.08	1.20	1.15	1.15	1.27	1.24	1.17
VAR(4)	1.45	1.20	1.19	1.04	1.04	1.13	1.24	1.18	1.22	1.21	1.11	1.09	1.27	1.19	1.27	1.32	1.24	1.21
SPF	0.94						1.03						1.06					
AR(2) Benchmark	1.86						1.85						1.80					

Notes: The AR(2) benchmark reports root mean squared forecast errors (RMSFEs) for quarterly annualized real GDP growth; all other statistics are relative to the AR(2) benchmark. For model relative RMSFEs less than 1 (shaded in green), *, **, *** denote rejection of the null hypothesis of equal predictive accuracy of the model and the benchmark at the 10%, 5%, and 1% level, respectively, based on the Diebold and Mariano (1995) test with the small-sample correction of Harvey et al. (1997). GDI is gross domestic income. GDA is the average of GDP and GDI. FSDP is final sales of domestic product, or GDP minus the change in inventories. DFP is domestic final purchases, or FSDP minus net exports. PDFP is private domestic final purchases, or DFP minus government. PCE is personal consumption expenditures. Forecasting methods are described in Section III.

Table C3: Relative RMSFEs for Real GDP Forecasts at SPF Survey Dates, Recession-Only Sample; Estimation Using Rolling Windows

	One-Step Horizon						Two-Step Horizon						Three-Step Horizon					
Method	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**
RW	1.03	1.02	0.79	0.68	0.63	0.61	0.84	0.82	0.98	0.77	0.78	0.76	0.84	0.79	0.85	0.74	0.74	0.84
RW-AR(1)	1.08	1.06	1.06	0.85	0.79	1.00	0.98	0.98	1.03	0.96	0.96	1.05	0.95	0.95	0.99	1.00	1.00	1.06
RW-AR(2)	1.05	1.01	1.07	0.87	0.81	0.89	1.00	0.95	0.99	0.88	0.88	0.93	0.97	0.94	0.98	0.93	0.94	1.02
RW-AR(3)	1.00	1.00	1.09	0.89	0.85	0.89	0.97	0.95	0.97	0.85	0.86	0.90	0.95	0.95	0.95	0.87	0.88	0.95
RW-AR(4)	0.99	0.96	1.07	0.85	0.82	0.87	0.94	0.90	0.96	0.82	0.84	0.90	0.94	0.91	0.94	0.85	0.87	0.94
Direct(1)	1.04	1.04	0.96	0.82	0.75	0.78	0.99	0.98	0.99	0.88	0.85	0.86	0.96	0.99	0.95	0.91	0.89	0.93
Direct(2)	1.05	1.03	0.96	0.81	0.72	0.70	0.98	1.00	0.98	0.87	0.85	0.85	0.96	0.98	0.95	0.90	0.89	0.94
Direct(3)	1.03	1.04	0.98	0.82	0.73	0.70	0.98	0.97	0.97	0.87	0.85	0.86	0.99	1.01	0.96	0.90	0.89	0.94
Direct(4)	1.03	1.01	0.96	0.82	0.72	0.66	1.01	0.99	0.97	0.86	0.85	0.84	1.01	1.03	0.97	0.91	0.90	0.94
VAR(1)	1.46	1.19	0.97	0.79	0.72	0.79	0.96	0.94	1.02	0.93	0.89	0.95	0.97	0.95	0.98	0.94	0.90	0.96
VAR(2)	1.53	1.19	1.00	0.79	0.69	0.71	1.01	0.98	0.97	0.86	0.85	0.84	0.98	0.96	0.96	0.87	0.87	0.90
VAR(3)	1.53	1.17	1.02	0.80	0.70	0.72	1.01	0.97	0.98	0.83	0.81	0.85	1.00	0.98	0.99	0.88	0.87	0.92
VAR(4)	1.67	1.27	0.98	0.80	0.68	0.69	1.07	1.03	0.96	0.86	0.85	0.87	1.06	1.04	0.96	0.89	0.89	0.96
SPF	0.52						0.71						0.86					
AR(2) Benchmark	3.82						4.33						4.60					

Notes: The AR(2) benchmark reports root mean squared forecast errors (RMSFEs) for quarterly annualized real GDP growth; all other statistics are relative to the AR(2) benchmark. For model relative RMSFEs less than 1 (shaded in green), *, **, *** denote rejection of the null hypothesis of equal predictive accuracy of the model and the benchmark at the 10%, 5%, and 1% level, respectively, based on the Diebold and Mariano (1995) test with the small-sample correction of Harvey et al. (1997). GDI is gross domestic income. GDA is the average of GDP and GDI. FSDP is final sales of domestic product, or GDP minus the change in inventories. DFP is domestic final purchases, or FSDP minus net exports. PDFP is private domestic final purchases, or DFP minus government. PCE is personal consumption expenditures. Forecasting methods are described in Section III.

Table C4: Relative RMSFEs for Real GDP Forecasts at Blue Chip Survey Dates, Full Sample; Estimation Using Rolling Windows

	1-Step Horizon						2-Step Horizon						3-Step Horizon					
Method	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**
RW	1.11	1.15	1.16	1.04	1.18	1.05	1.26	1.25	1.30	1.17	1.27	1.09	1.28	1.29	1.30	1.23	1.34	1.11
RW-AR(1)	0.98	0.96	1.03	0.93	0.97	0.98	1.02	0.99	1.00	1.01	1.03	0.99	1.04	1.03	1.02	1.07	1.10	1.03
RW-AR(2)	0.98	0.98	1.04	0.94	0.99	0.95	1.02	1.01	1.01	1.02	1.06	0.95	1.03	1.05	1.01	1.10	1.13	1.03
RW-AR(3)	0.99	1.01	1.06	0.96	1.02	0.96	1.03	1.02	1.03	1.06	1.13	0.97	1.06	1.06	1.06	1.15	1.22	1.05
RW-AR(4)	1.01	1.05	1.06	0.95	1.02	0.96	1.06	1.07	1.05	1.06	1.13	0.98	1.08	1.12	1.08	1.17	1.23	1.07
Direct(1)	0.98	0.96	1.01	0.93	0.94	0.96	1.02	0.99	1.04	0.99	0.98	0.97	1.01	1.01	1.00	1.03	1.01	1.00
Direct(2)	0.99	0.97	1.04	0.94	0.93	0.96	1.04	1.01	1.05	1.02	1.00	0.96	1.05	1.03	1.06	1.07	1.05	1.03
Direct(3)	1.00	0.98	1.05	0.94	0.93	0.96	1.06	1.02	1.08	1.04	1.02	0.98	1.06	1.05	1.07	1.08	1.05	1.04
Direct(4)	1.02	0.99	1.06	0.95	0.95	0.96	1.07	1.04	1.08	1.04	1.03	0.98	1.09	1.08	1.09	1.09	1.07	1.03
VAR(1)	0.98	0.98	1.02	0.94	0.94	0.98	1.02	1.02	1.02	1.02	1.01	0.99	1.03	1.03	1.01	1.04	1.03	1.00
VAR(2)	0.99	0.99	1.08	0.94	0.91	0.98	1.01	1.01	1.06	1.01	0.99	0.98	1.01	1.01	1.03	1.06	1.04	1.03
VAR(3)	1.00	1.00	1.09	0.95	0.92	0.98	1.03	1.03	1.08	1.04	1.00	0.99	1.05	1.05	1.08	1.11	1.09	1.07
VAR(4)	1.03	1.04	1.13	0.96	0.93	0.99	1.09	1.09	1.12	1.07	1.00	1.01	1.09	1.09	1.13	1.15	1.10	1.10
AR(2) Benchmark	2.15						2.30						2.29					

Notes: The AR(2) benchmark reports root mean squared forecast errors (RMSFEs) for quarterly annualized real GDP growth; all other statistics are relative to the AR(2) benchmark. For model relative RMSFEs less than 1 (shaded in green), *, **, *** denote rejection of the null hypothesis of equal predictive accuracy of the model and the benchmark at the 10%, 5%, and 1% level, respectively, based on the Diebold and Mariano (1995) test with the small-sample correction of Harvey et al. (1997). GDI is gross domestic income. GDA is the average of GDP and GDI. FSDP is final sales of domestic product, or GDP minus the change in inventories. DFP is domestic final purchases, or FSDP minus net exports. PDFP is private domestic final purchases, or DFP minus government. PCE is personal consumption expenditures. Forecasting methods are described in Section III.

Table C5: Relative RMSFEs for Real GDP Forecasts at Blue Chip Survey Dates, Expansion-Only Sample; Estimation Using Rolling Windows

	1-Step Horizon						2-Step Horizon						3-Step Horizon					
Method	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**
RW	1.26	1.31	1.30	1.18	1.37	1.21	1.44	1.43	1.45	1.36	1.49	1.25	1.57	1.59	1.54	1.51	1.68	1.29
RW-AR(1)	1.05	1.05	1.00	0.97	1.05	0.98	1.06	1.02	0.99	1.05	1.07	0.96	1.11	1.10	1.04	1.14	1.17	1.01
RW-AR(2)	1.08	1.09	1.02	0.97	1.06	0.98	1.08	1.08	1.02	1.09	1.15	0.97	1.09	1.15	1.04	1.23	1.26	1.04
RW-AR(3)	1.10	1.12	1.03	0.99	1.08	1.00	1.12	1.10	1.07	1.17	1.27	1.01	1.17	1.18	1.13	1.34	1.44	1.12
RW-AR(4)	1.13	1.19	1.04	1.00	1.10	1.00	1.17	1.20	1.10	1.19	1.28	1.03	1.22	1.29	1.19	1.38	1.46	1.16
Direct(1)	1.03	1.04	1.03	0.99	1.02	1.05	1.07	1.02	1.07	1.05	1.05	1.04	1.07	1.06	1.04	1.12	1.09	1.05
Direct(2)	1.06	1.05	1.08	1.00	1.02	1.08	1.10	1.06	1.09	1.11	1.08	1.03	1.14	1.08	1.15	1.20	1.16	1.09
Direct(3)	1.08	1.08	1.09	1.00	1.02	1.07	1.14	1.08	1.15	1.14	1.11	1.06	1.17	1.14	1.14	1.20	1.17	1.11
Direct(4)	1.10	1.09	1.11	1.01	1.05	1.08	1.16	1.12	1.14	1.14	1.13	1.05	1.19	1.16	1.17	1.22	1.19	1.11
VAR(1)	1.05	1.05	1.05	1.01	1.04	1.07	1.06	1.06	1.03	1.07	1.08	1.03	1.10	1.10	1.03	1.11	1.13	1.04
VAR(2)	1.07	1.07	1.12	1.00	1.01	1.10	1.06	1.06	1.11	1.09	1.06	1.06	1.07	1.07	1.07	1.19	1.16	1.12
VAR(3)	1.09	1.09	1.13	1.01	1.02	1.09	1.10	1.10	1.14	1.15	1.10	1.07	1.13	1.13	1.15	1.27	1.24	1.18
VAR(4)	1.13	1.14	1.19	1.03	1.04	1.12	1.16	1.16	1.22	1.18	1.08	1.09	1.15	1.15	1.26	1.31	1.24	1.21
AR(2) Benchmark	1.87						1.90						1.80					

Notes: The AR(2) benchmark reports root mean squared forecast errors (RMSFEs) for quarterly annualized real GDP growth; all other statistics are relative to the AR(2) benchmark. For model relative RMSFEs less than 1 (shaded in green), *, **, *** denote rejection of the null hypothesis of equal predictive accuracy of the model and the benchmark at the 10%, 5%, and 1% level, respectively, based on the Diebold and Mariano (1995) test with the small-sample correction of Harvey et al. (1997). GDI is gross domestic income. GDA is the average of GDP and GDI. FSDP is final sales of domestic product, or GDP minus the change in inventories. DFP is domestic final purchases, or FSDP minus net exports. PDFP is private domestic final purchases, or DFP minus government. PCE is personal consumption expenditures. Forecasting methods are described in Section III.

TABLE C6: RELATIVE RMSFES FOR REAL GDP FORECASTS AT BLUE CHIP SURVEY DATES, RECESSION-ONLY SAMPLE; ESTIMATION USING ROLLING WINDOWS

	1-Step Horizon						2-Step Horizon						3-Step Horizon					
Method	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**	**GDA**	**GDI**	**FSDP**	**DFP**	**PDFP**	**PCE**
RW	0.74	0.70	0.80	0.69	0.65	0.58	0.91	0.90	1.03	0.80	0.82	0.76	0.80	0.79	0.92***	0.76	0.76	0.86
RW-AR(1)	0.80	0.74	1.07	0.85	0.81	0.98	0.95	0.93	1.03	0.96	0.97	1.04	0.94	0.93	0.99	0.99	1.00	1.05
RW-AR(2)	0.76	0.72	1.08	0.88	0.84	0.88	0.92	0.89	0.99	0.89	0.90	0.92	0.94	0.91	0.98	0.93	0.94	1.02
RW-AR(3)	0.72	0.72	1.11	0.90	0.87	0.89	0.88	0.88	0.98	0.86	0.87	0.90	0.91	0.90	0.96	0.88	0.89	0.95
RW-AR(4)	0.71	0.68	1.09	0.86	0.84	0.87	0.87	0.84	0.96	0.83	0.86	0.89	0.89	0.86	0.95	0.85	0.88	0.94
Direct(1)	0.87	0.79	0.95	0.80	0.73	0.74	0.93	0.94*	0.99	0.88	0.85	0.85	0.94	0.93	0.96	0.91	0.90	0.93
Direct(2)	0.83	0.76	0.96	0.79	0.70	0.67	0.94	0.93	0.98	0.87	0.86	0.85	0.93	0.96	0.95	0.90	0.89	0.94
Direct(3)	0.83	0.75	0.97	0.80	0.71	0.68	0.92	0.94**	0.97	0.87	0.85	0.85	0.93	0.94	0.96	0.91	0.89	0.94
Direct(4)	0.82	0.75	0.96	0.80	0.70	0.64	0.91*	0.90*	0.97	0.87	0.85	0.84	0.96	0.98	0.98	0.91	0.90	0.94
VAR(1)	0.82	0.82	0.96	0.79	0.71	0.75	0.95	0.95	1.01	0.93	0.90	0.94	0.93	0.93	0.98	0.94	0.90	0.96
VAR(2)	0.78	0.78	1.00	0.79	0.68	0.69	0.93	0.93	0.97	0.86	0.85	0.85	0.94	0.94	0.96	0.87	0.88	0.90
VAR(3)	0.78	0.78	1.02	0.80	0.68	0.70	0.91	0.91	0.99	0.84	0.82	0.85	0.94	0.94	0.99	0.88	0.87	0.92
VAR(4)	0.80	0.80	0.99	0.81	0.67	0.67	0.97	0.97	0.95	0.87	0.85	0.87	1.02	1.02	0.95	0.90	0.90	0.96
AR(2) Benchmark	3.72						4.34						4.61					

Notes: The AR(2) benchmark reports root mean squared forecast errors (RMSFEs) for quarterly annualized real GDP growth; all other statistics are relative to the AR(2) benchmark. For model relative RMSFEs less than 1 (shaded in green), *, **, *** denote rejection of the null hypothesis of equal predictive accuracy of the model and the benchmark at the 10%, 5%, and 1% level, respectively, based on the Diebold and Mariano (1995) test with the small-sample correction of Harvey et al. (1997). GDI is gross domestic income. GDA is the average of GDP and GDI. FSDP is final sales of domestic product, or GDP minus the change in inventories. DFP is domestic final purchases, or FSDP minus net exports. PDFP is private domestic final purchases, or DFP minus government. PCE is personal consumption expenditures. Forecasting methods are described in Section III.