Forecasting GDP Growth using Disaggregated GDP Revisions

Adam J. Check
University of St. Thomas

Anna K. Nolan
Independent

Tyler C. Schipper *
University of St. Thomas

July 2019

Abstract

This paper investigates the informational content of regular revisions to real GDP growth and its components. We perform a real-time forecasting exercise for the advance estimate of real GDP growth using dynamic regression models that include revisions to GDP and its components. Echoing other work in the literature, we find little evidence that including aggregate GDP growth revisions improves forecast accuracy relative to an AR(1) baseline model; however, models that include revisions to components of GDP improve forecast accuracy. The first revision to consumption is particularly relevant in that every model that includes the revision outperforms the baseline model. Measured by root mean squared forecasting error (RMSFE), improvements are quite sizable, with many models increasing forecasting performance by 5% or more, and with top-performing models forecasting 0.24 percentage points closer to the advance estimate of growth. We use Bayesian model averaging to underscore that our results are driven by the informational content of revisions. The posterior probability of models with the first revision to consumption is significantly higher than our baseline model, despite strong priors that the latter should be the preferred forecasting model.

Keywords: Data revisions, real-time data, forecasting

JEL Classification Numbers: E01, C11, C53, C82

*Correspondence for this paper can be sent to *schi0195@stthomas.edu*. We thank Dean Croushore, Eric Gaus, C. Richard Higgins, and Jermy Piger for reading early drafts, and thank anonymous referees for their helpful feedback on the completed draft. We also benefited from comments at the Midwest Economics Association's Annual Meeting, the Liberal Arts Macro Conference, the Applied Probability and Statistics Workshop at St. Thomas, and a seminar at Drake University. We appreciate Mary Reichardt's editorial suggestions. All remaining errors are our own.

1 Introduction

The revision process for many macroeconomic time series has led to an explosion of work focusing on the use of real-time data. In particular, real-time data has given rise to estimation methods that use multiple vintages of the same time series to improve forecast performance. In light of this important work, we ask whether revisions to GDP growth and its disaggregated components contain information about future advance estimates of GDP growth. Our focus on revisions to disaggregated components is novel. Echoing other work in the literature, we find that revisions to GDP growth have no impact on short-run forecast accuracy. However, we find strong evidence that revisions to certain GDP components, principally consumption, have important information for forecasting economic growth. Further, the improvement in forecast accuracy can be substantial over an AR(1) benchmark model.[1] Measured by root mean squared forecast error (RMSFE), many component-augmented models improve forecast performance by greater than 5%, with the best consumption-augmented model forecasting roughly 0.24 percentage points closer to the advance estimate of GDP growth.

We illustrate the importance of component revisions in two complementary exercises. First, we estimate dynamic regression models using an expanding window via maximum likelihood estimation (MLE) to produce one-period ahead forecasts of advance GDP growth. Consistent with the previous literature, this exercise establishes that aggregate revisions do not contain information that is useful in forecasting advance GDP growth. However, this exercise also establishes that including component revisions *can* improve forecast performance. We highlight the importance of the first revision to consumption as *every* model that contains it outperforms our baseline. Second, we directly address the large-model-space problem that is inherent to our research question by performing a Bayesian model averaging (BMA) exercise. We show that even with strong priors for our baseline AR(1) model, the posterior model probabilities indicate a preference for forecasting models that include the first revision to consumption.

There is a well-developed literature that has attempted to incorporate preliminary data

[1] The use of an AR(1) as a baseline, and ARMA(p,q) models more generally, is quite common (e.g Koop and Potter, 2004; Koenig et al., 2003; Clements and Galvão, 2013a).

into coherent forecasting frameworks.[2] Much of its focus has been on forecasting post-revision releases (e.g Howrey, 1978; Kishor and Koenig, 2012; Clements and Galvão, 2012). Our work more closely aligns with a strain of the literature which forecasts early or advance estimates of macroeconomic time series. We incorporate the guidance of Koenig et al. (2003) by forecasting advance releases of GDP growth using only advance releases in the estimation of our models. Two other studies incorporate information from revisions to forecast advance GDP growth: Clements and Galvão (2013a) and Clements and Galvão (2013b). The former paper concludes that several variants of vintage-based vector autoregression (V-VAR) models show little improvement in forecast accuracy. Clements and Galvão (2013b) suggest an alternative real-time vintage approach to forecasting. They estimate models using lightly revised data while their forecasts are conditioned on the same data as traditional approaches to forecasting that utilize the most recent vintage. Our method contrasts nicely with Clements and Galvão (2013b) in that we also forecast the advance estimate of GDP growth, but our forecasts are conditioned on the same series used for estimation. Our results also share interesting parallels with Garciga and Knotek (2017) who find that forecasting with NIPA aggregates, particularly consumption, improves short-horizon forecasts of GDP growth.

2 Data and Estimation

We use advance estimates of real GDP growth over the period 1991Q4:2017Q1. Our sample begins when the Bureau of Economic Analysis (BEA) switched its preferred measure of production from GNP to GDP. While the two series, fully revised, are quite similar prior to 1991Q4, we focus on GDP to avoid problems with the revision process possibly changing over time. GDP has three regular releases: advance, second, and third. Our advance estimate of real GDP growth is calculated as the percentage change between the new advance estimate of GDP and the third estimate of the previous quarter's GDP, which is the same measure reported by the BEA. Annual and benchmark revisions do not impact our series since each advance estimate of growth is calculated within a data vintage. All of our real-time data is

[2] Croushore (2011) provides an excellent summary of the broader literature on real-time data.

from the Real-Time Data Set for Macroeconomics. A brief description of the construction of our variables is available in Appendix A of the online supplementary material.

We supplement GDP growth data with data on aggregate GDP revisions (henceforth simply aggregate revisions) and revisions to consumption, government spending, investment, imports, and exports (denoted as c, g, i, m, and x respectively). The advance components for quarter t are measured as the percentage change in the expenditure category between the third release for quarter $t-1$ and the advance release for quarter t, to correspond with our calculation of GDP growth. However, we do not use component growth directly, since the goal of our study is to determine if *revisions* contain useful information. In our estimation procedure, we instead include up to two *revisions* for GDP growth or each component's growth. First revisions are the difference between the advance estimate of growth and the second estimate, and second revisions are the difference between the second estimate of growth and the third estimate. This distinction is denoted by a subscript for each revision. For example, the first consumption revision, c_1, is the percentage point difference between consumption growth in the advance release and consumption growth in the second release. In total, we have two aggregate revisions and ten different component revisions that are used to augment our dynamic regression models of GDP growth. Following Koenig et al. (2003), we do all estimation using only the advance estimates for each series.

The BEA creates advance estimates of GDP growth using "partial and preliminary source data as well as trend projections when data are not available" (Fixler et al., 2014). In Euclidean distance, revisions to advance estimates (i.e. our first revision) are 0.65% on average, and the second round of revisions are even smaller at roughly 0.37%. Revisions to components of GDP are similar in that the first revision is on average larger than the second revision.[3] Relevant summary statistics for advance estimates of GDP growth and revisions can be found in Appendix B of the online supplementary material.

Our main goal is to investigate whether the inclusion of revisions in a dynamic linear regression model increases forecast performance relative to a baseline AR(1) model.[4] To do

[3] On average, the revisions were slightly biased — on average the second estimate of GDP was 0.13% above the advance estimate, and the third estimate was roughly 0.02% above the second estimate.

[4] The single auto-regressive term, with no moving average terms, is preferred by AIC with a sample-size correction (AICc). AICc will often suggest more complex ARMA models for all post-war data or fully revised data. Our baseline also *forecasts* advance GDP growth better than other candidate models, notably

so, we consider models spanning all possible combinations of component revisions. With $k = 10$ total types of component revisions, we have $R = 2^{10} = 1,024$ possible models to consider.[5] Let any specific model be denoted by $M_r \in \{M_1, M_2, \cdots, M_{1024}\}$. Under a specific model, M_r, we consider a unique subset of revisions, X^r, with coefficient estimates, μ_r and β_r, and an AR(1) persistence estimate of the errors, ρ_r. Since the values of the errors are unobserved, they need to be estimated via a state-space model. The state-space model with AR(1) errors under model r is given by:

$$y_t = \mu_r + X^r_{t-1}\beta_r + \eta^r_t \tag{1}$$

$$\eta^r_t = \rho_r \eta^r_{t-1} + \varepsilon^r_t \tag{2}$$

where $\varepsilon^r_t \sim N(0, \sigma_r)$. In our baseline model we do not include any revisions, so X^r_{t-1} is empty and equations (1) and (2) represent a standard AR(1) process.[6]

When forecasting, we perform expanding window estimation and produce one-step-ahead forecasts at each step of the expanding window. The estimation process is as follows. First, we focus on one model, M_r, from the set $\{M_1, M_2, \cdots, M_{1024}\}$. This model will have a unique set of regressors, X^r. Given that our test sample starts in 2004Q4, we use the subset of our data from the first observation through 2004Q3 to estimate μ_r, β_r and ρ_r via MLE. We assume these parameters are not time dependent, although their estimated coefficients may change slightly as the estimation window expands. Using these estimates, we form the one-period-ahead out-of-sample point forecast. We then forward the end-date of the training sample by one period, so that it runs through 2004Q4 and repeat the estimation and forecasting process. We continue in this fashion until we reach the end of our sample. We then move to the next model, M_{r+1}, and repeat the process. We measure forecast performance by the one-step-ahead RMSFE, where the error is the difference between actual advance GDP growth and forecasted advance GDP growth.

an AR(2) and an ARMA(1,1), in our preferred test period.

[5] In the case of aggregate revisions, we consider four possible models including two models with one revision (first and second), one with both of the revisions, and one that excludes the revisions entirely.

[6] Formulating our model as an autoregressive distributed lag model (ARDL) with one lag of GDP growth and lagged component revisions as exogenous regressors yields nearly identical results as our state-space model.

All models are initially estimated on the sample period 1991Q4:2004Q3. This period roughly splits our data, so that we can utilize the final 50 observations for our pseudo-out-of-sample exercise (i.e. test sample). Beginning our test sample in 2004Q4 allows us to measure forecast performance during the Great Recession (2008Q1:2009Q2). It also leaves a relatively large amount of time between the start of our sample in 1991Q4 and the beginning of our test data in 2004Q4. While we think that beginning the forecasting exercise in 2004Q4 is sensible, our results are also qualitatively similar under different test samples.

3 Results

3.1 Aggregate Revisions

Table 1 presents the relative forecast performance for the three models that are augmented with aggregate revisions. Using our framework, aggregate revisions do not improve forecast performance, echoing the findings of Clements and Galvão (2013b) using vintage-based VARs. It is important to note that these aggregate revisions are themselves composed of component revisions that may be the result of different estimation processes. This aggregation process may obfuscate important information contained in individual component revisions. This possibility leads to our focus on component revisions in the next section.

Table 1: Forecasting with Aggregate Revisions

Model	Relative RMSFE
First Revision (r_1)	1.021
Second Revision (r_2)	0.994
First and Second Revision ($r_1 + r_2$)	1.005

The first column indicates the covariates that are included in Equation (1). RMSFE is reported relative to the baseline AR(1) model such that values less than 1 indicate that the revisions improve forecast performance.

3.2 Component Revisions

In comparison to aggregate revisions, revisions to GDP components increase forecast performance and therefore appear to contain information that is important for forecasting the advance release of real GDP growth. Table 2 highlights the forecast performance for several models.[7] The best model in terms of forecast performance has RMSFE that is 12.7% below that of the baseline model. This improvement is not limited to a small subset of models. More than half of all component-based models outperformed the baseline model by at least 3.1%. These results are surprising and important on two fronts. First, they indicate that the revisions contain important information about the future path of real GDP growth *beyond* what is captured by past GDP growth. Second, this information is typically hidden through the aggregation process.

Table 2: Forecasting with Component Revisions

Model	RMSFE	ENC-NEW
Component-Augmented Models		
$c_1 + g_1 + i_2 + x_1$ (best)	1.660***	12.029**
c_1	1.752*	6.514**
$c_1 + c_2 + i_1 + i_2 + x_2 + m_1$ (median)	1.842	4.112†
Comparison Models		
AR(1)/Base	1.902	
AR(0)/Mean	2.164	
Naive	1.987	

Astericks on RMSFE indicate statistically significant improvements in forecast accuracy at the 1% (***), 5% (**), and 10% (*) levels using Diebold and Mariano (1995). ENC-NEW reports the encompassing test statistic used in Clark and McCracken (2001) where k is the number of components and $\pi \approx 1$. Astericks denote that the test statistic is significant at the 5%(**) or 10%(*) levels (the critical values reported in Clark and McCracken (2001)). The median model may be statistically signficant but Clark and McCracken (2001) do not report critical values for models with more than four additional parameters. Both test statistics are measured against the AR(1) baseline model.

Given the large set of potential models that we face, there are legitimate concerns about forecast performance being driven by statistical noise. This problem is unavoidable without an *a priori* justification to exclude certain revisions. We tackle this problem in two ways. First, we highlight patterns in the results above that cast doubt on forecast performance being driven by statistical noise. Second, in the following section, we confront the large-model-space problem directly, and in a theoretically rigorous fashion, using Bayesian model

[7] We readily recognize the limitations of the statistical tests in Table 2, especially the DM test in our context (lack of power and nested models). We include them as widely recognized measures of forecast performance. We are comfortable with with these tests given our rigorous full-sample comparison in the following section.

averaging. We also use BMA to address the significance of our results in a full-sample model comparison as suggested in Diebold (2015).

Under the null hypothesis of equal predictive accuracy between the baseline model and a model augmented with component revisions, we might anticipate roughly 50% of models to outperform the baseline. In our preferred test period, 68% of models outperform the baseline, and this percentage varies from 59.9% to 87.9% across shorter test periods.[8] However, nearly 50% of models outperform the baseline model in *every* test period, indicating that the set of outperforming models is quite robust. More importantly, *all* models that contain the first revision to consumption, c_1, outperform the baseline model in our preferred test period, while models that exclude it are about equally likely to outperform or underperform. This is entirely consistent with c_1 containing important information for forecasting, while other components contain little relevant information. Our results are robust to changes in forecast evaluation (MAFE) and censoring outliers in the revision series. Our results indicate that GDP component revisions, particularly consumption, contain meaningful information about the future path of GDP growth. Moreover, the improvements in forecast performance that we document are sizable and robust to changes in the test period for forecast evaluation.

3.3 Bayesian Model Comparison

Above, we have identified a subset of component revision models that outperform the baseline AR(1) model when performance is measured by RMSFE in a pseudo-out-of-sample exercise. We have also shown that the RMSFE of many of these models is statistically superior to those of the baseline AR(1) model when using the Diebold Mariano (DM) test of forecast performance and the encompassing test found in Clark and McCracken (2001). However, Diebold (2015) cautions agains using the DM test to compare forecasting models in pseudo-out-of-sample forecasting exercises . While he concludes that its use for this purpose is "likely fine" (pg. 8), he also recommends the use of statistics that utilize the full sample of data, rather than evaluating the models with a pseudo-out-of-sample exercise with an arbitrarily chosen cut-off date.

Here, we follow his recommendation and use the full-sample Bayesian marginal likelihood

[8] See Appendix C in the online supplementary material for this exercise.

to compare models. The Bayesian marginal likelihood increases as the likelihood function increases, but also includes a built-in punishment for models that have extra parameters. This, combined with prior beliefs across the model space, provides us a theoretically justified way to choose between models when the model space is large. Although this statistic uses all of the data, it can also be viewed as a pseudo-out-of-sample statistic that starts the out-of-sample period at the first observation. In other words, it is a pseudo out-of-sample statistic that uses the *entire* sample of data, rather than starting at an arbitrarily selected cut-off period. This is because the Bayesian marginal likelihood can be decomposed into the product of the one-step ahead forecast densities, starting at the first time period and ending at the last. Finally, it measures model performance more broadly than RMSFE, since it judges models based on their entire *forecast distribution* rather than their *point forecasts*.

To compare models and compute variable inclusion probabilities, we compute the posterior probability of each model. This measure is proportional to the product of the marginal likelihood of the model times the prior probability of the model. We set model priors in favor of the baseline AR(1) model to provide a second layer of protection against data-mining.[9] Our prior belief is that there is a 50% probability that the baseline AR(1) model will produce the best forecasts, with the other 50% probability spread across the other 1,023 models.[10] In Table 3, we also present results under "flat" model priors, in which each of the 1,024 models receives equal prior probability.

Table 3: Posterior Probabilities of Top 5 Models

Model	Prior	Posterior	Posterior (Flat Prior)
c_1	2.5%	56.6%	9.3%
Baseline AR(1)	50.0%	14.4%	0.1%
$c_1 + g_1$	0.3%	9.0%	11.1%
$c_1 + m_1$	0.3%	2.9%	4.3%
$c_1 + i_2$	0.3%	2.0%	0.3%

Under flat model priors, each model received $1/1,024 \approx 0.1\%$ prior probability. The results under the flat model priors are presented in the last column.

[9] The first layer of protection is built into the marginal likelihood since all else equal, models are punished for additional complexity (i.e. more variables) resulting in a smaller posterior probability.

[10] For more precise details, see Appendix D in the online supplementary material.

While we do not definitively select the single "best" forecast model, the models receiving the highest posterior probability are listed in Table 3. Table 3 illustrates that under our baseline model priors, the posterior probability of the AR(1) model falls from 50% to 14.4%, indicating that it is improbable that the AR(1) is the best forecasting model in the set considered.[11] All the models that included component revisions combined received a 50% prior, which rose to a posterior probability of 85.6%, indicating that it is likely that at least one of the models with component revisions outperforms the baseline AR(1) model. This result is likely driven by the importance of the first revision of consumption, which has a posterior inclusion probability of 85%, as seen in Table 4. Although the top rated model differs from that found in the RMSFE exercise above, the broad result remains — the top performing models almost all contain the first revision to consumption.[12]

Table 4: Posterior Probabilities of Component Revisions

Component	Posterior
c_1	85.1%
c_2	2.2%
i_1	2.0%
i_2	3.8%
g_1	12.1%
g_2	1.9%
x_1	2.1%
x_2	3.2%
m_1	5.2%
m_2	1.9%

The prior probability for any given component revision is equivalent to the sum of the model probability of all models that include that revision. Therefore, the prior probability for all components is the same at 10%.

[11] Under flat model priors, the posterior probability of the AR(1) model is largely unchanged from the prior. It is 131st ranked model, and the 130 ranked above it combine to make up roughly 92% of the posterior probability.

[12] In work not appearing in the paper, we also analyzed the forecasting performance using the log-predictive density of the component models vs. the AR(1). All results are qualitatively similar to those obtained via the more standard BMA analysis presented here.

4 Conclusion

This paper conducts a real-time forecasting exercise of the advance estimate of real GDP growth. We compare the forecast performance of dynamic regression models that are augmented with data revisions. Similar to previous work in the literature, we find no increase in forecast performance using aggregate GDP revisions. Surprisingly, component revisions appear to improve forecast performance with nearly 68% of such models outperforming an AR(1) model. The best of these models predicts advance GDP growth about 0.18 percentage points more accurately than our baseline.

Without an *a priori* reason to exclude certain revisions, we consider a large set of alternative models. Because of this, we conduct an extensive investigation into the significance of our results. Component-augmented models outperform our baseline in a variety of test periods. Importantly, we show that *every* model that contains the first revision of consumption outperforms our AR(1) baseline in our preferred test period. This consistency lends credibility to the gains in forecast performance being driven by the information in revisions rather than statistical noise. Finally, we proceed to directly address the inherent large-model-space problem and potential drawbacks associated with using the Diebold-Mariano test statistic in pseudo-out-of-sample model comparisons. To do this, we turn to Bayesian model averaging. It allows us to evaluate forecast performance across our entire sample and evaluates the entire forecast distribution rather than just point forecasts. Despite our strong priors for the AR(1) model, the BMA exercise clearly illustrates a drop in the posterior likelihood of the baseline model and a high likelihood that the best model is one that includes the first revision to consumption.

On the whole, we feel that the results paint a compelling picture that GDP component revisions, particularly the first revision to consumption, improve forecast accuracy. The most direct implication of this finding is that some GDP revisions contain relevant information about future GDP growth beyond what is already captured in an autoregressive process. This initial finding suggests several potential paths for future research. While we have identified the role of consumption, the exact mechanism and the importance of other components could be a fruitful line of inquiry. One avenue is to further investigate the role of forecasting with disaggregated components vs. disaggregated component *revisions*. Given that Garciga and Knotek (2017) document the importance of consumption in forecasting GDP, it is important to better understand the interplay of differing forecasting methods, forecasting aggregates with disaggregated components, and the information contained in revisions. Our results may also have important implications for other strains of the real-time forecasting and the data revisions literature. In particular, they may inform the debate surrounding whether data revisions contain news or noise (e.g Mankiw and Shapiro, 1986) and the efficiency of government data releases (e.g Aruoba, 2008).

Bibliography

Aruoba, S. B. (2008). Data revisions are not well behaved. *Journal of Money Credit and Banking*, 40(2-3):319–340.

Clark, T. E. and McCracken, M. W. (2001). Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics*, 150(1):85–110.

Clements, M. P. and Galvão (2012). Improving real-time estimates of output gaps and inflation trends with multiple-vintage VAR models. *Journal of Business and Economic Statistics*, 30(4):554–562.

Clements, M. P. and Galvão (2013a). Forecasting with vector autoregressive models of data vintages: US output growth and inflation. *International Journal of Forecasting*, 29:698–714.

Clements, M. P. and Galvão (2013b). Real-time forecasting of inflation and output growth with autogressive models in the presence of data revisions. *Journal of Applied Econometrics*, 28:458–477.

Croushore, D. (2011). Frontiers of real-time data analysis. *Journal of Economic Literature*, 49(1):72–100.

Diebold, F. X. (2015). Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of Diebold-Mariano tests. *Journal of Business and Economic Statistics*, 33(1):1–24.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263.

Fixler, D. J., Greenaway-McGrevey, R., and Grimm, B. T. (2011). Revisions to GDP, GDI, and their major components. *Survey of Current Business*, 91:9–31.

Fixler, D. J., Greenaway-McGrevey, R., and Grimm, B. T. (2014). Revisions to GDP, GDI, and their major components. *Survey of Current Business*, 94:1–23.

Garciga, C. and Knotek, E. S. (2017). Forecasting GDP growth with NIPA aggregates. *Federal Reserve Bank of Cleveland Working Paper*, No. 17-08.

Howrey, E. P. (1978). The use of preliminary data in econometric forecasting. *Review of Economics and Statistics*, 6(2):193–200.

Kishor, N. and Koenig, E. (2012). VAR estimation and forecasting when data are subject to revision. *Journal of Business and Economic Statistics*, 30(2):181–190.

Koenig, E. F., Dolmas, S., and Piger, J. (2003). The use and abuse of real-time data in economic forecasting. *Review of Economics and Statistics*, 85(3):618–628.

Koop, G. (2003). *Bayesian Econometrics*, chapter 6. Wiley, Chichester, England.

Koop, G. and Potter, S. (2004). Forecasting in dynamic factor models using Bayesian model averaging. *The Economics Journal*, 7(2):550–565.

Mankiw, G. N. and Shapiro, M. D. (1986). News or noise: An analysis of GNP revisions. *Survey of Current Business*, 66(5):20–25.

U.S. Bureau of Economic Analysis (2015). Why does BEA revise GDP estimates? Retrieved from `https://blog.bea.gov/2015/07/13/why-does-bea-revise-gdp-estimates-2/`.

Young, A. H. (1993). Reliability and accuracy of the quarterly estimates of GDP. *Survey of Current Business*, 73(October):29–43.

Zellner, A. (1958). A statistical analysis of provisional estimates of gross national product and its components, of selected national income components, and of personal saving. *Journal of American Statistical Association*, 53(281):54–65.

Appendix A Data Construction

Our primary series of interest is the advance estimate of real GDP growth. We utilize the series directly from the Real-Time Data Set for Macroeconomists, hosted at the Federal Reserve Bank of Philadelphia. For convenience, we outline some of the critical aspects of constructing the data. Interested readers should consult the extensive documentation that accompanies the data set.[13] The advance estimate of annualized quarter over quarter GDP growth is constructed as:

$$y_t^T = \left[\left(\frac{Y_t^T}{Y_{t-1}^T} \right)^4 - 1 \right] \times 100, \tag{3}$$

where T represents the first vintage that contains an observation of t. For our purposes the difference between T and t will always correspond to about 1 month or the release lag from the end of a quarter until the advance release of real GDP. A more concrete example is shown below:

$$y_{1991Q4}^{1992m1} = \left[\left(\frac{Y_{1991Q4}^{1992m1}}{Y_{1991Q3}^{1992m1}} \right)^4 - 1 \right] \times 100, \tag{4}$$

where y_{1991Q4}^{1992m1} would be the advance estimate of real GDP growth for 1991Q4 based on the vintage released in January 1992. We calculate revisions to GDP as the difference in growth rates between monthly releases for t. Therefore, the first revision r_1 is the difference between growth rates for the first (T) and second ($T+1$) releases of period t GDP growth:

$$r_{1,t} = y_t^{T+1} - y_t^T. \tag{5}$$

The second revision then is:

$$r_{2,t} = y_t^{T+2} - y_t^{T+1}. \tag{6}$$

We omit the vintage superscript and time subscripts from revisions for clarity. It is understood that first revisions can only exist after a second vintage containing t has been released. Its superscript is therefore $T+1$ and the second revision's is $T+2$. Typically level changes in GDP due to definitions and measurement occur in advance releases. Since our growth rates are calculated within vintages and subsequent revisions (i.e. second and third) use the same definitions as the advance release, we do not make any additional changes to our data to account for revisions beyond the advance, second, and third releases.

Our component data for real personal consumption expenditures, real exports of goods and services, real imports of goods and services, and real government consumption and gross investment are also directly from the Philadelphia Federal Reserve's Real-Time Data Set for Macroeconomists. Our data for component revisions is constructed in the same way as for revisions to GDP. For example, let C be the level of consumption in GDP (measured in real terms), then the annualized growth rate of C is:

$$c_t^T = \left[\left(\frac{C_t^T}{C_{t-1}^T} \right)^4 - 1 \right] \times 100. \tag{7}$$

[13] Documentation can be found at `https://www.philadelphiafed.org/research-and-data/real-time-center/real-time-data/`.

The data we use for estimation are the first and second revisions to c, which are again measured as differences in consumption growth:

$$c_{1,t} = c_t^{T+1} - c_t^T. \tag{8}$$

The second revision then is:

$$c_{2,t} = c_t^{T+2} - c_t^{T+1}. \tag{9}$$

As with the aggregate revisions, we drop the vintage superscript and time subscripts in order to refer to revisions generally. We also tolerate the abuse of notion (using c for revisions and growth) in order to present a simpler composition in the text. The notation for government spending, investment, import, and export expenditures follow an identical construction and are designated g, i, m, and x, respectively.

For real gross private domestic investment we use the sum of real non-residential investment, real residential investment, and real change in private inventories. Note that since some of our data occurs after the BEA changed to chain-weighting rather than fixed-price weighting, summing the component elements of real gross private domestic investment will not equal real gross private domestic investment from other sources like the Archival Federal Reserve Economic Database (ALFRED). Given that the two series are highly correlated, we opted to use a consistent data source and rely on the considerable documentation provided by the Philadelphia Federal Reserve's Real-Time Data Set for Macroeconomists.

Appendix B Summary Statistics

Table 5 provides summary statistics for the advance estimates of real GDP growth. It begins in 1991Q4 to coincide with the BEA's change from GNP to GDP as its primary measure of output. Our series spans 102 quarters, with 101 oberseravions of advance GDP growth. We withhold 50 observations for our pseudo-out-of-sample comparisons between models. The single gap in our sample stems from the lack of an advance release of real GDP in 1995Q4.[14] For comparison, the post-war (1947Q2:2017Q1) real GDP growth rate, using the recent September 28th, 2017 vintage, is 3.22% with a standard deviation of 3.91%.

Table 5: Advance Estimates of Real GDP Growth

Period	Dates	N	Mean	Std. Dev.	Gaps
Full Sample	1991Q4 - 2017Q1	101	2.51%	1.95%	1
Estimation Sample	1991Q4 - 2004Q3	51	3.13%	1.74%	1
Test Sample	2004Q3 - 2017Q1	50	1.88%	1.97%	0

We also utilize the differences in growth rates between advance and second estimates (first revision) and second and third estimates (second revision). Summary statistics for these revisions are presented in Table 6. As has been documented by others, the first revision to GDP and GDP growth tends to be much larger than the second revision (roughly 6 times larger in our full sample).[15] However, the difference between first and second revisions is considerably lower in the second half of our sample (27 times larger vs. 2 times larger). There is also a greater variance in first revisions than second revisions. Our revision data has an additional gap (beyond what was presented in Table 5) resulting from a missing second release of GDP growth in 2003Q3.

Table 6: Revisions to Real GDP Growth

	Mean		Std. Dev.		Gaps	
Period	r_1	r_2	r_1	r_2	r_1	r_2
Full Sample	0.129%	0.022%	0.641%	0.370%	2	1
Estimation Sample	0.191%	0.007%	0.591%	0.313%	2	1
Test Sample	0.068%	0.038%	0.688%	0.424%	0	0

r_1 and r_2 refer to the first and second revisions to real GDP growth, respectively.

[14] For greater explanation on these types of data eccentricities see the extensive documentation that accompanies each series in the Real-Time Data Set for Macroeconomists.

[15] Additional information on the revision process for aggregate output can be found in Zellner (1958), Young (1993), Fixler et al. (2011), Croushore (2011), Fixler et al. (2014), and U.S. Bureau of Economic Analysis (2015)

Tables 7 and 8 present summary statistics for the revisions to GDP components. These revisions share many of the characteristics of the revisions to aggregate GDP. Mainly, within each component, first revisions tend to be larger than second revisions, and there is more volatility in first revisions than second revisions. First revisions tend to be positive across each of the components as would be expected given that revisions to overall growth also tend to be positive. Second revisions see more heterogeneity with respect to their sign, but are very small in relative magnitude. In line with the two gaps in aggregate revisions, our first set of component revisions also has two gaps in 1995Q4 and 2003Q3, while our second set of revisions only has a gap in 2003Q3 (the first release in 1995Q3 is missing while the second release is missing for 2003Q3).

Table 7: First Revisions to GDP Components

	Component	Mean	Std. Dev.	Gaps
	Full Sample			
c_1		0.066%	0.369%	2
i_1		0.661%	3.332%	2
g_1		0.119%	0.854%	2
x_1		0.916%	2.591%	2
m_1		0.956%	3.142%	2
	Estimation Sample			
c_1		0.180%	0.364%	2
i_1		0.810%	3.425%	2
g_1		0.253%	1.038%	2
x_1		1.170%	3.047%	2
m_1		1.691%	3.716%	2
	Test Sample			
c_1		-0.049%	0.339%	0
i_1		0.511%	3.263%	0
g_1		-0.015%	0.598%	0
x_1		0.661%	2.037%	0
m_1		0.220%	2.244%	0

c, i, g, x, m refer to the real GDP components of consumption, investment, government spending, exports, and imports, respectively. The first revision is the difference in each component's growth rate between the advance and second estimate. See Equation (8).

Table 8: Second Revisions to GDP Components

Component	Mean	Std. Dev.	Gaps
Full Sample			
c_2	0.014%	0.358%	1
i_2	-0.001%	1.415%	1
g_2	-0.075%	0.475%	1
x_2	0.127%	1.389%	1
m_2	-0.126%	1.486%	1
Estimation Sample			
c_2	0.039%	0.236%	1
i_2	-0.081%	1.433%	1
g_2	-0.181%	0.595%	1
x_2	0.068%	1.424%	1
m_2	-0.215%	1.744%	1
Test Sample			
c_2	-0.012%	0.451%	0
i_2	0.081%	1.405%	0
g_2	0.032%	0.276%	0
x_2	0.187%	1.365%	0
m_2	-0.035%	1.179%	0

c, i, g, x, m refer to the real GDP components of consumption, investment, government spending, exports, and imports, respectively. The second revision is the difference in each component's growth rate between the second and third estimate. See Equation (9).

Appendix C Additional Results

Table 9: Forecasting with Component Revisions

Component	Outperforming Models with component (%)	Performance of models with component (%)
c_1	73.7%	100%
g_1	59.3%	80.5%
i_2	53.4%	72.5%
m_2	52.8%	71.7%
x_1	51.1%	69.3%
x_2	50.1%	68.0%
m_1	46.6%	63.3%
g_2	46.0%	62.5%
c_2	45.6%	61.9%
i_1	43.9%	59.6%

Table 9 includes data on two separate conditional statements. Column two reports the percentage of models that outperform the baseline which include each component. Column three reports the percentage of models with a component that outperform the baseline. There are 695 models (68%) that outperform the baseline model using our preferred test period.

Table 10: Robustness of Forecasts with Component Revisions

	Start of Test Period						
	2004	2005	2006	2007	2008	2009	2010
Baseline Percentile	68.0%	67.6%	66.1%	67.6%	75.4%	59.9%	87.9%
Cumulative % of outperforming models	67.9%	67.4%	65.9%	65.7%	65.7%	48.3%	48.2%

From left to right each column increases the estimation period by one year. Column 2 represents our primary estimation period through 2004Q3 with column 3 advancing the estimation period to 2005Q3. Baseline percentile reports the performance of the baseline model relative to all 1024 models (ranked from lowest to highest RMSFE). Row 2 reports the percentage of all models that outperform the baseline model conditional on *also* outperforming the baseline in all previous test periods.

Appendix D Bayesian Priors

As is well documented in the Bayesian literature, for each model M_r, we can compute the model probability, $p(M_r|Y)$ as:

$$p(M_r|Y) = cp(Y|M_r)p(M_r) \tag{10}$$

where $p(M_r|Y)$ is the probability of model r conditional on the data, Y. This is commonly called the *posterior probability* of model r. Defined in this way, the posterior probability represents the probability that, of the set of models under consideration, model M_r is the *best* model for explaining the data. There are three terms on the right-hand side that together determine this probability: c is a constant, $p(Y|M_r)$ is the marginal likelihood of model r, and $p(M_r)$ is the prior probability assigned to model r. The marginal likelihood includes a reward for model fit and a punishment for the inclusion of irrelevant variables.[16]

D.1 Model Priors

In order to perform Bayesian analysis, we need to set two types of prior beliefs: priors across models and priors on the parameters of each model. Since parsimonious models, like an AR(1), have shown relatively strong forecast performance in many studies of various macroeconomic indicators, we set 50% of our prior belief on the AR(1) model. The other 50% prior probability is spread across all remaining models in a geometrically declining fashion according to model size. That is, we believe with roughly 25% prior probability that a model that includes any single component revision by itself will be the superior forecasting model, with roughly 12.5% probability that a model with any combination of two component revisions will be the superior forecasting model, etc. This is consistent with our belief that parsimonious forecasting models are preferred to more complex ones.

Mathematically, the prior probability of model M_r depends on the number of component revisions, k_r, that are included in that model. For all models with at least one revision included, we set:

$$\begin{aligned} p(M_r|k_r = j) &\propto \frac{0.5^{j+1}}{N_j} \\ N_j &= \binom{10}{j} \\ j &\in \{1, \cdots, 10\}. \end{aligned}$$

For example, the group of models with only one component revision have a collective prior probability of roughly 25%. We divide this prior probability equally across all models that include only one revision. Since there are 10 of these models, each one-component revision model receives a prior probability of roughly 2.5%.

[16] It may be helpful to think of $p(Y|M_r)$ as akin to AIC or BIC, as it includes a built in penalty for large models. In fact, the large sample approximation to $p(Y|M_r)$ under linear regression is $-\frac{1}{2}$ times the Bayesian Information Criterion (BIC).

D.2 Regression Parameter Priors

For each model, we place priors on the regression coefficients, β_r, and the AR(1) term, ρ_r. For the $k \times 1$ vector of regression coefficients, β_r, in model r, we follow convention and set $p(\beta_r) = N(0, V_r)$. Our prior is that each regressor has no effect, with the tightness of that belief controlled by the covariance matrix, V_r. Put another way, our prior is that past revisions to GDP growth do not contain information important to forecasting GDP growth.

Since we are considering a large set of models, we set these priors in an automatic fashion. Our prior belief is that each coefficient is independent, so all non-diagonal elements of V_r are zero. We assume that each coefficient is given as:

$$p(\beta_r^k) = N(0, \sigma_\beta^2)$$
$$\sigma_\beta^2 = 1$$

where β_r^k is the coefficient corresponding to the k^{th} regressor in model r.

For the AR(1) term, ρ_r, we set:

$$p(\rho_r) = N(0.5, \sigma_\rho^2)$$
$$\sigma_\rho^2 = 0.5^2$$

We center the AR(1) coefficient at 0.5 since most economic variables have positive autocorrelation at the first lag. Since the AR(1) coefficient should always lie between -1 and 1, the variance of 0.25 is fairly wide. To estimate the model, we use the method of Koop (2003), and we enforce the restriction that $-1 < \rho_r < 1$ via rejection sampling.