



# Estimating U.S. Output Growth with Vintage Data in a State-Space Framework

Richard G. Anderson and Charles S. Gascon

This study uses a state-space model to estimate the “true” unobserved measure of total output in the U.S. economy. The analysis uses the entire history (i.e., all vintages) of selected real-time data series to compute revisions and corresponding statistics for those series. The revision statistics, along with the most recent data vintage, are used in a state-space model to extract filtered estimates of the “true” series. Under certain assumptions, Monte Carlo simulations suggest this framework can improve published estimates by as much as 30 percent, lasting an average of 11 periods. Real-time experiments using a measure of real gross domestic product show improvement closer to 10 percent, lasting for 1 to 2 quarters. (JEL C10, C53, E01)

Federal Reserve Bank of St. Louis *Review*, July/August 2009, 91(4), pp. 349-69.

**S**tatistical agencies face a tradeoff between accuracy and timely reporting of macroeconomic data. As a result, agencies release their best estimates of the “true” unobserved series in the proceeding month, quarter, or year with some measurement error.<sup>1</sup> As agencies collect more information, they revise their estimates, and the data are said to be more “mature.” As the reported data mature, the estimates, on average, are assumed to converge toward the “true” unobserved values. This study examines a methodology in which the “true” value of an economic variable is latent in the sense of the state vector in a state-space model. In doing so, we use recent modeling suggestions by

Jacobs and van Norden (2006) and Cunningham et al. (2007) regarding relationships among real-time data, measurement error as a heteroskedastic stochastic process, and the latent, “true” data for an economic variable of interest.

The importance of *potential output growth* in policymaking motivates our study. Forward-looking macroeconomic models suggest that the predicted future path of the output gap should be important to policymakers. To the extent that policymakers are concerned with a Federal Reserve-style “dual mandate,” an output gap equal to 1 percent of potential output may be quite alarming if projections suggest it will continue, but relatively innocuous if the gap is expected to shrink rapidly during the next few quarters. Recent studies on inflation forecasting conclude that the output gap, when measured in real-time using vintage data, has little predictive power for inflation (e.g., Orphanides and van Norden, 2005; and Stock and Watson, 2007). It is also important to study the real-time measurement of potential output because policymakers occasionally face

---

<sup>1</sup> In Appendix B we address the philosophical question of why an econometrician might believe s/he can improve published data—after all, statisticians who produce data have access to the same historical data used by econometricians and, hence, should create models using the same understanding of the revision process that econometricians use. Over long periods, benchmarks and redefinitions muddy the analysis. But, it is an act of hubris to assert that any simple statistical model can produce consistently more-accurate near-term data than are produced by the specialists constructing the published data. Hubris aside, we have written this paper regardless.

---

Richard G. Anderson is a vice president and economist and Charles S. Gascon is a senior research associate at the Federal Reserve Bank of St. Louis.

© 2009, The Federal Reserve Bank of St. Louis. The views expressed in this article are those of the author(s) and do not necessarily reflect the views of the Federal Reserve System, the Board of Governors, or the regional Federal Reserve Banks. Articles may be reprinted, reproduced, published, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are included. Abstracts, synopses, and other derivative works may be made only with prior written permission of the Federal Reserve Bank of St. Louis.

possible changes/breaks in the underlying growth trend of productivity and, hence, potential output.

Our objective in this study is not to assess inflation-forecasting models, although that has been a major use of potential output measures; rather, it is to estimate the “true” value of real output for use in the construction of trend-like measures of potential output. One of the larger recent studies in this vein, albeit focused on inflation prediction, is by Orphanides and van Norden (2005). The study considers, as predictive variables for inflation, both a wide range of output gap measures (which differ with respect to data vintage and the trend estimator) and lagged values of real output growth. Their conclusion regarding output gap models as predictors of inflation is straightforward—the output gap does not reliably predict inflation, although the differences in forecast performance between output-gap and output-growth models are not statistically significant:

[O]ur analysis suggests that a practitioner could do well by simply taking into account the information contained in real output growth without attempting to measure the level of the output gap. This model was consistently among the best performers, particularly over the post-1983 forecast sample. (p. 597)

Motivated by these findings, this article models the true (unobserved) output measure of real output and the implications of such for estimators of a real output trend. To so do, we explore the measurement error and subsequent data-revision process for real gross domestic product (RGDP).

## LITERATURE REVIEW

Early studies of real-time data focused on the sensitivity of certain statistics to data vintage (Howrey, 1978 and 1984; Croushore and Stark, 2001; Diebold and Rudebusch, 1991; and Orphanides and van Norden, 2002 and 2005). Later research posed the problem more formally as a signal-extraction problem (Kishor and Koenig, 2005; Aruoba, Diebold, and Scotti, 2008; and Aruoba, 2008). Both approaches emphasized the sensitivity of statistical inferences, including measures of the forecasting power of the output gap.

Recent analyses have focused on “the possibility that the sequence of vintages released over time may contain useful information with which to interpret the most recent vintage of data and to anticipate future outcomes” (Garratt et al., 2008, p. 792). Such a possibility was discussed by Howrey (1978) but only recently has become the centerpiece of certain studies.

A long literature has addressed the use of real-time data, starting with Howrey’s 1978 paper on forecasting with preliminary data and including Croushore and Stark’s (2000) release of a vintage economic dataset at the Philadelphia Fed. This literature, until recently, has focused on three main issues: (i) embedding an estimate of the data revision process into forecasting models, (ii) assessing the sensitivity of statistical inferences in macroeconomic data to data vintage, and (iii) checking the forecastability of revisions, in the context of Mankiw and Shapiro’s (1986) classic discussion of “news vs. noise.”<sup>2</sup>

Some authors have argued there are policy implications of such issues. Croushore (2007) argues that revisions to published personal consumption expenditures (PCE) inflation rates are forecastable, at least from August to August of the following year, and identifies an upward bias to revisions, indicating that initial estimates consistently are too low. He suggests that policymakers should “account for” this bias and predictability in setting monetary policy. Kozicki (2004) analyzes vintages of the output gap, employment gap, and inflation data and finds that revised data and real-time data suggest differing policy actions. Kozicki suggests that policymakers should place greater emphasis on more-certain data and be less aggressive in response to changes in data subject to large revisions. Previously, Orphanides and van Norden (2002 and 2005) argued that failure to appreciate the difference between real-time and final data risks serious policy errors.

<sup>2</sup> Our analysis is silent on the discussion of “news vs. noise” in real-time data analysis—“news” meaning that the statistical agency publishes efficient estimates using all available information, “noise” meaning there is measurement error unrelated to the true value. These are not mutually exclusive; both conditions may not hold. News implies revisions have mean zero, noise does not. Empirical results suggest that noise dominates the data-generating process.

Recently, data availability has encouraged researchers to explore a methodology in which they estimate the “true” values and measurement errors as a latent state vector. In such studies, the revisions are modeled as a statistical process, emphasizing the “maturity” of each observation, rather than the vintage of the time series. These models permit forecasts of data that are to be released, as well as “backcasts” of data already published. The methodology may be applied to individual observations, as well as various trend estimators, such as those considered by Orphanides and van Norden (2005).<sup>3</sup> Recent work includes Jacobs and van Norden (2006); Cunningham et al. (2007); Aruoba (2008); Aruoba, Diebold, and Scotti (2008); Garratt, Koop, and Vahey (2008); and Garratt et al. (2008).

This recent literature traces its beginning to Jacobs and van Norden (2006). They argue that previous state-space models built on a transition process for the vintage data plus a set of forecasting equations do not allow adequately rich dynamics in the data-revision process:

Our formulation of the state-space model is novel in that it defines the measured series as a set of various vintage estimates for a given point in time, rather than a set of estimates from the same vintage. We find this leads to a more parsimonious state-space representation and a cleaner distinction between various aspects of measurement error. It also allows us to augment the model of published data with forecasts in a straightforward way. (p. 3)

In this spirit, we note the differences between using state-space models as estimators of unobserved components such as trends (perhaps across various vintages of real-time data) and as estimators of “true” underlying data. In the former, each datum within a time series of a particular vintage is implicitly assumed to be equally accurately measured; the trend (usually, a time-varying direction vector) is extracted without explicit concern for measurement error, except so far as the robustness of the extracted trend may be explored across vintages. In the latter, each datum within a time

series is assumed to have an amount of measurement error that is inversely correlated with the maturity of the datum. Interpreted loosely, the information content of an older datum for a given activity date is asserted to contain more information about the “true” value of that datum than a recent datum for the same activity date.

This modeling framework has been applied by staffs at the Bank of England and the European Central Bank (Cunningham et al., 2007). Their model differs somewhat from that of Jacobs and van Norden and focuses more attention on modeling the measurement-error process, including potential bias and heteroskedasticity, but the underlying philosophy is similar. Our research applies the Jacobs and van Norden framework to U.S. data on quarterly GDP from the Federal Reserve Bank of St. Louis real-time Archival Federal Reserve Economic Data (ALFRED) database.

## STATISTICAL FRAMEWORK

The rich modeling framework proposed by Cunningham et al. (2007) allows serial correlation in measurement errors, nonzero correlation between the state of the economy and measurement errors, and maturity-dependent heteroskedasticity in measurement errors. As a consequence of the richness of the statistical specification and the number of dimensions to the data, the estimation is divided into two parts. First, all available data vintages are used to estimate selected parameters governing measurement error bias and variance. Second, the most recently published release is used to estimate the state-space model.

The modeling setup is as follows.<sup>4</sup> Let the data-generating process for the true (unobserved) variable of interest,  $y_t, t=1, \dots, T$ , be a simple autoregressive (AR( $q$ )) process:

$$(1) \quad A(L)y_t = \varepsilon_t,$$

where the polynomial is defined in the usual manner and the stationary disturbance is spherical

<sup>3</sup> These trend estimators are discussed in Orphanides and van Norden (2005, Appendix A).

<sup>4</sup> The model follows Jacobs and van Norden (2006) and Cunningham et al. (2007).

## UNDERSTANDING REAL-TIME DATA

The table presents a stylized real-time dataset. The columns denote the release date, or vintage, of the data. The rows denote the activity date, or observation date, of the data. Economic data are normally released with a one-period lag; that is, data for January are reported in February. Therefore, the release date,  $v$ , lags the activity date,  $t$ , by one period.

Each element in the dataset is reported with a subscript identifying the activity date and a **(bold)** superscript identifying the maturity,  $j$ . Data of constant maturity are reported along each diagonal.

### Stylized Real-Time Dataset

Period		Vintage ( $v$ )						
		$v = 2$	$v = 3$	$v = 4$	...	$v = T-1$	$v = T$	$v = T+1$
Activity date ( $t$ )	$t = 1$	$y_1^1$	$y_1^2$	$y_1^3$	...	$y_1^{T-2}$	$y_1^{T-1}$	$y_1^T$
	$t = 2$		$y_2^1$	$y_2^2$	...	$y_2^{T-3}$	$y_2^{T-2}$	$y_2^{T-1}$
	$t = 3$			$y_3^1$	...	$y_3^{T-4}$	$y_3^{T-3}$	$y_3^{T-2}$
	$\vdots$				$\ddots$	$\vdots$	$\vdots$	$\vdots$
	$t = T-2$					$y_{T-2}^1$	$y_{T-2}^2$	$y_{T-2}^3$
	$t = T-1$						$y_{T-1}^1$	$y_{T-1}^2$
	$t = T$							$y_T^1$

(homoskedastic),  $E(\varepsilon_t) = 0$ ,  $V(\varepsilon_t) = E(\varepsilon_t \varepsilon_t') = \sigma_\varepsilon^2 I$ . Trends (deterministic or stochastic) and structural breaks, including regime shifts, are explicitly ruled out (and perhaps have been handled by prefiltering the data).

### Measurement-Error Model

Let the data published by the statistical agency be denoted

$$y_t^j, \quad t = 1, \dots, T; \quad j = 1, \dots, J,$$

where  $t$  is an activity date and  $j$  is the maturity of the data for that activity date. (See the boxed insert.) We assume that initial publication of data for  $t$  occurs in period  $t+1$ , so that  $j \geq 1$ . Period  $T$  is the final revision date for data published in period  $T+1$ . We assume the published data are decomposable as

$$(2) \quad y_t^j \equiv y_t + c^j + v_t^j,$$

where  $y_t$  denotes the true “unobserved” value,  $c^j$  denotes a bias in published data of vintage  $j$ , and  $v_t^j$  is a measurement error.

Previous studies have suggested that data releases tend to be biased estimates of the later releases. Let  $c^j$  denote the bias of data at maturity  $j$ , such that  $c^1$  is the bias for initially published data. We assume the bias is independent of vintage and solely a function of maturity,  $j$ , and that the bias decays according to the rule

$$(3) \quad c^j = c^1 (1 + \lambda)^{j-1}, \quad -1 \leq \lambda \leq 0.$$

We assume the measurement error,  $v_t^j$ , follows a simple AR( $q$ ) process:

$$(4) \quad B(L)v_t^j = \eta_t^j,$$

where  $E(\eta_t^j) = 0$ . The measurement-error variance is assumed heteroskedastic in maturity and decays toward zero:

$$V(\eta_t^j) = E\left(\eta_t^j (\eta_t^j)'\right) = \Sigma_{\eta}^j,$$

with main diagonal

$$(5) \quad \sigma_{\eta^j}^2 = \sigma_{\eta^1}^2 (1 + \delta)^{j-1}, \quad -1 \leq \delta \leq 0.$$

It is necessary to designate a maturity at which data are assumed “fully mature.” Here, we denote the horizon as  $N$  and refer to it below as the “revision horizon.” For RGDP data, we set  $N = 20$  (5 years of quarterly data). This choice, to some extent, is arbitrary, and hence it is useful to examine the robustness of results to the value chosen. Our choices are guided by visual examination of the revised time series and discussed further in a later section.

### STATE-SPACE MODEL

The measurement equation of the state-space model has as its dependent variable a vector of the most recent release of data,

$$y_t^j = [y_1^T, y_2^{T-1}, \dots, y_{T-1}^2, y_T^1]^T.$$

The superscript  $j$  denotes the vintage of data, measuring  $y_t$  on activity date  $t$ , which is available at vintage  $T+1$ . Note that the maturities of the elements of

$$[y_1^T, y_2^{T-1}, \dots, y_{T-1}^2, y_T^1]^T$$

differ—some elements may be the 10th or 20th release of data for a specific activity date, while the last element is the initial release of data for activity period  $T$ . The measurement equation equates this vector to the sum of a vector of maturity-related measurement biases,  $c^j$ ; the unknown true value,  $y_t$ ; and a measurement error,  $v_t^T$ :

$$(6) \quad y_t^j = c^j + [1 \quad 1] \begin{bmatrix} y_t \\ v_t^T \end{bmatrix} + 0.$$

The transition equation for the state vector is

$$(7) \quad \begin{bmatrix} y_t \\ v_t^T \end{bmatrix} = \begin{bmatrix} \mu \\ 0 \end{bmatrix} + \begin{bmatrix} \alpha & 0 \\ 0 & \beta \end{bmatrix} \begin{bmatrix} y_{t-1} \\ v_{t-1}^T \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ \eta_t^T \end{bmatrix}$$

with disturbance covariance matrix

$$(8) \quad \begin{bmatrix} \sigma_{\varepsilon^T}^2 & \sigma_{\varepsilon, \eta^{(j,t)}}^2 \\ \sigma_{\varepsilon, \eta^{(j,t)}}^2 & \sigma_{\eta^{(j,t)}}^2 \end{bmatrix}.$$

Note that the variance of  $\eta_t^T$ , denoted  $\sigma_{\eta^j}^2$ , is a function of  $t$  through its dependence on  $j$ , the maturity of each datum, reflecting the assumed heteroskedasticity in the measurement-error process.<sup>5</sup> Similarly, the covariance between the shocks to the variable of interest and the measurement error,  $\sigma_{\varepsilon \eta^{(j,t)}}^2 = \rho_{\varepsilon, \eta} \sigma_{\varepsilon} \sigma_{\eta^j}$ , is a function of maturity,  $j$ .

Cunningham et al. (2007) make the interesting suggestion that the measurement equation may be augmented with auxiliary variables  $y_t^Z$ :

$$(9) \quad \begin{bmatrix} y_t^T \\ y_t^Z \end{bmatrix} = \begin{bmatrix} c^j \\ c^Z \end{bmatrix} + \begin{bmatrix} I & 0 & I & 0 \\ \Lambda^Z & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y_{t-q+1} \\ v_t \\ \vdots \\ v_{t-p+1} \end{bmatrix} + \begin{bmatrix} 0 \\ v_t^Z \end{bmatrix}.$$

Candidate variables include surveys and/or private-sector measures/forecasts, asserting that private-sector agents already have solved their own variants of the signal-extraction problem. At this time, our model omits the use of auxiliary data.

The estimation is partitioned into two parts. Assuming the measurement equation has been augmented with an auxiliary variable and allowing for AR(1) processes in the transition equation, the parameters to be estimated in the state-space model are

$$\Phi_1 = (\alpha_1, \sigma_{v^Z}^2, \mu, \sigma_{\varepsilon}^2, c^Z, \Lambda^Z),$$

conditional on estimated parameters for the measurement error’s data-generating process,

$$\Phi_2 = (\sigma_{\eta^1}^2, \delta, \beta_1, \rho_{\varepsilon, \eta}, c^1, \lambda).$$

The estimation of  $\Phi_2$  proceeds assuming that successive revisions to each datum are well-

<sup>5</sup> State-space models with deterministically time-dependent variances are discussed by Durbin and Koopman (2001, pp. 172-74) and Kim and Nelson (1999).

behaved, in the statistical sense that the revisions may be used for estimation.<sup>6</sup> Let  $W$  denote a matrix with  $J$  rows, in which each row is regarded as a vector of revisions to data of maturity  $j$ . The number of columns is  $T-N$ , that is, the number of published data vectors minus the revision horizon. A general expression for a representative row in the revision matrix  $W$  is

$$(10) \quad W(j, \cdot) = y_t^{j+N} - y_t^j, \quad j+N < T, \quad 1 \leq t \leq T.$$

Consider  $j = 1$  and  $N = 20$ . In this case, the numbers in the first row of  $W$  are

$$W(1, \cdot) = y_t^{1+20} - y_t^1, \quad 1+N < T.$$

Similarly, consider  $j = 12$  and  $N = 20$ :

$$W(12, \cdot) = y_t^{12+20} - y_t^{12}, \quad 12+N < T.$$

Clearly,  $W$  has  $J$  rows and  $T-N$  columns.

Consider an estimator for the bias process,

$$(11) \quad c^j = c^1(1+\lambda)^{j-1}, \quad -1 \leq \lambda \leq 0, \quad 1 \leq j \leq J.$$

The row means of  $W$  provide sample measures of  $c^j$ . The parameters  $c^1$ , the mean revision of the initial release, and  $\lambda$ , the revision decay rate, are estimated via generalized method of moments (GMM) subject to the constraint  $-1 \leq \lambda \leq 0$ .

Next, we need an estimator for  $\rho_{\varepsilon, \eta}$  as part of  $\sigma_{\varepsilon \eta}^2 = \rho_{\varepsilon, \eta} \sigma_{\varepsilon} \sigma_{\eta}^T$ . Cunningham et al. (2007) propose an estimator based on an approximation to  $\rho_{\varepsilon, \eta}$ , designated  $\rho_{y, v}^*$ , calculated as the mean (across the  $J$  maturities) of the  $J$  correlations between the  $j$ th rows of  $W$  and the corresponding vector of published data at maturity  $j+N$ ; that is, by the construction of  $W$ ,  $N+j < T-t$ .

Finally, estimators are required for  $\sigma_{\eta^1}^2$ ,  $\delta$ , and  $\beta_1$  (assuming an AR(1) process in  $v$ ). A sample estimate of the variance-covariance is obtained as  $J^{-1}WW'$ . The analytical covariance matrix for the first-order case is

$$(12) \quad V = \left( \frac{\sigma_{\eta^1}^2}{1-(1+\delta)\beta_1^2} \right) \times \begin{pmatrix} 1 & (1+\delta)\beta_1 & \cdots & (1+\delta)^{J-1}\beta_1^{J-1} \\ (1+\delta)\beta_1 & (1+\delta) & \cdots & (1+\delta)^{J-1}\beta_1^{J-2} \\ \vdots & \vdots & \ddots & \vdots \\ (1+\delta)^{J-1}\beta_1^{J-1} & (1+\delta)^{J-1}\beta_1^{J-2} & \cdots & (1+\delta)^{J-1} \end{pmatrix},$$

and we estimate  $\sigma_{\eta^1}^2$ ,  $\delta$ , and  $\beta_1$  via GMM by minimizing

$$(13) \quad \left( \text{vec} V - \text{vec} \hat{V} \right)' \left( \text{vec} V - \text{vec} \hat{V} \right).$$

Cunningham et al. (2007) suggest methods to obtain covariance matrices for higher lag orders.

## SIMULATION RESULTS

We conduct Monte Carlo simulations to explore the ability of the state-space framework to extract a “true” series from a “published” series that has been contaminated with measurement error.

The simulations evaluate the ability of the model’s state-space vector  $[\hat{y}_t, \hat{v}_t]$  to track the vector of true values,  $y_t$ , relative to the tracking ability of the vector of most recently “published” values,  $y_t^T$ . For each parameterization,  $T = 100$  and we calculate 1,000 replications.

The specification of the experiment is as follows:

- The “true” data:

$$y_t = \alpha y_{t-1} + \varepsilon_t, \quad t = 2, \dots, T, \quad \varepsilon_t \sim N(0, 1), \quad y_1 = \varepsilon_1 \quad (\text{that is, } y_0 = 0).$$

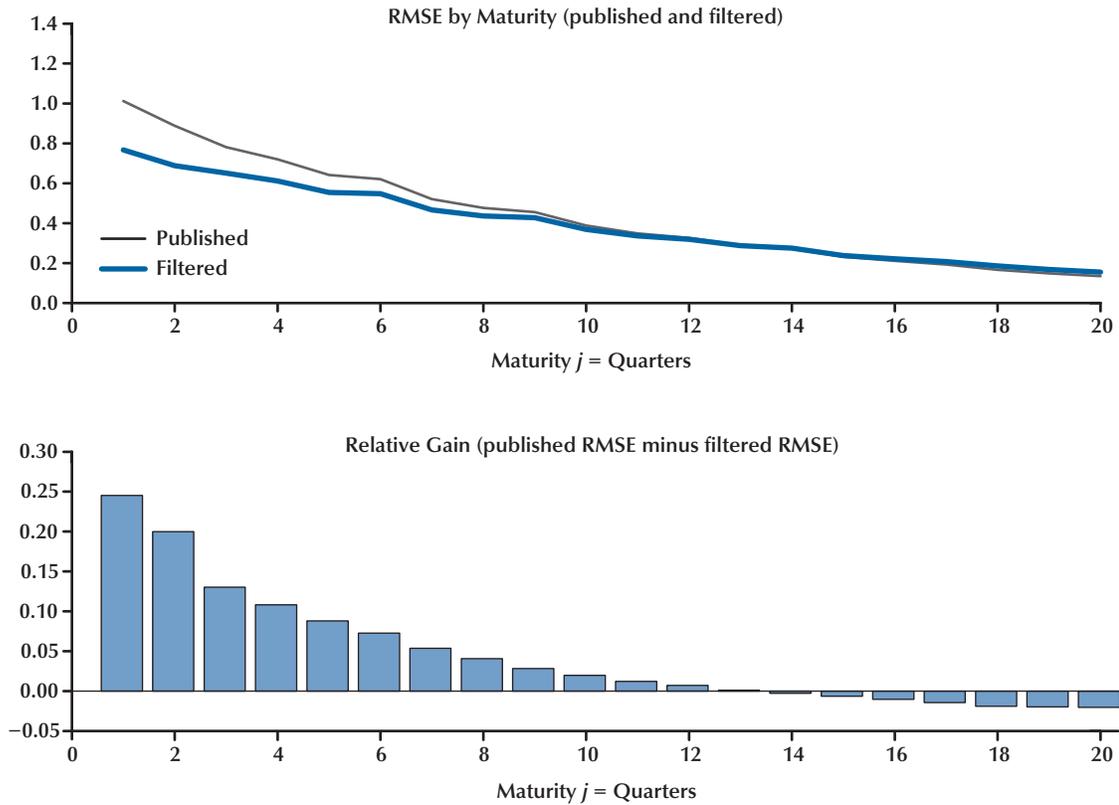
- The “published” data:

$$y_t^j = y_t + v_t^j, \quad t = 1, \dots, T,$$

where the superscripts  $t$  and  $j$  denote, respectively, the activity date and maturity of the most recently published data.

- The measurement error:

<sup>6</sup> Hereafter, this exercise is conditional on the revision horizon  $N$ .

**Figure 1****RMSE by Maturity,  $j$ , Model-Simulation Parameters:  $\alpha = 0.60$ ,  $\beta = 0.10$ ,  $\rho_{\varepsilon\eta} = -0.50$** 

$$v_t^j = \beta v_{t-1}^j + \eta_t^j, \quad t = 2, \dots, T; \quad \eta_t^j \sim N(0, \sigma_{\eta_t^j}^2);$$

$$\sigma_{\eta_1^j}^2 = 1; \quad \sigma_{\eta_t^j}^2 = \sigma_{\eta_1^j}^2 (1 + \delta)^{j-1},$$

$$v_1^j = \eta_1^j \text{ (that is, } v_0^j = 0) \text{ and with } \delta = -0.06.$$

- The covariance between the state of the economy and the measurement error:

$$\text{cov}(\varepsilon_t, \eta_t^j) = \rho_{\varepsilon, \eta} \sigma_{\varepsilon} \sigma_{\eta_t^j}.$$

Figure 1 shows root mean square forecast errors (RMSEs) of the “published” values,  $y_t - y_t^T$ , and the filtered values,  $y_t - \hat{y}_t$ , for one parameterization. The top panel shows the RMSE at each maturity; the bottom panel shows the difference between the filtered RMSEs and published RMSEs. The figure indicates that the filtered values are

better estimates of the true values for the first 13 maturities, after which time the filtered values cease to provide an advantage.

Table 1 provides corresponding results.<sup>7</sup> The first three columns report varying parameterizations. The fourth and fifth columns report the improvement due to the state-space filter for data maturities 1 and 10, respectively. At maturity 1, the RMSEs of filtered estimates are approximately

<sup>7</sup> In some respects, the results presented in Table 1 are similar to those in Cunningham et al. (2007), while others are puzzlingly different. The comparable table in Cunningham et al. (2007, Table C) displays a sharp deterioration of the model’s advantage as the covariance shock to the economy and measurement error decreases. Additionally, the gains reported are much larger and more persistent; in some cases, the filtered RMSEs are close to 50 percent of the published RMSEs at maturity 1 and remain so at maturity 9. The gains from filtering last at least 18 but sometimes over 100 periods. At this time we have not resolved these discrepancies; but we thank the authors for graciously providing their simulation code.

**Table 1**  
**Measurement Accuracy Improvement Due to the State-Space Filter**

Parameterization			Improvement due to state-space filter		
$\rho_{\varepsilon,\eta}$	$\alpha$	$\beta$	RMSE <sub>filtered</sub> /RMSE <sub>published</sub> At maturity 1	At maturity 10	Earliest maturity at which RMSE <sub>published</sub> < RMSE <sub>filtered</sub>
0.5	0.1	0.1	0.7179	1.1338	8
0.5	0.1	0.6	0.6489	1.0762	9
0.5	0.6	0.1	0.7500	1.1437	8
0.5	0.6	0.6	0.7344	1.1581	7
0	0.1	0.1	0.7179	1.0357	10
0	0.1	0.6	0.6479	1.0277	9
0	0.6	0.1	0.7493	1.0226	10
0	0.6	0.6	0.7337	1.0478	9
-0.5	0.1	0.1	0.7177	0.9378	15
-0.5	0.1	0.6	0.6537	0.9377	20
-0.5	0.6	0.1	0.7578	0.9494	14
-0.5	0.6	0.6	0.7331	0.9431	14

70 percent of the RMSEs obtained when using the most recently “published” data. The values range from 64 to 75 percent improvement depending on the parameterization. The sixth column reports the earliest maturity at which the filtered values cease to provide an advantage; these values range from 8 to 20 periods, with an average of 11 periods.

Our simulations suggest that the state-space framework may promise significant gains in measurement accuracy for recently released data if actual data are well behaved and tend to follow a low-order AR process. Previous studies suggest this might be reasonable for RGDP.

## EMPIRICAL MODEL

Our empirical work examines vintage data of the annualized growth rate of quarterly RGDP constructed with data from the Federal Reserve Bank of St. Louis ALFRED database, specifically, nominal GDP and the implicit price deflator (GDPDEF) data.<sup>8</sup> The construction of RGDP accounts for changes in the base year of GDPDEF, as to maintain the correct interactions between the base year and subsequent vintages. Thus, in

this vintage RGDP matrix, the most recently published data vector matches the data available from the Bureau of Economic Analysis (BEA). The specifics of the process are described in Appendix B.

Estimation proceeds in two steps: First, we estimate the parameters of the measurement-error process,

$$\Phi_2 = (\sigma_{\eta^2}, \delta, \beta_1, \rho_{\varepsilon,\eta}, c^1, \lambda).$$

Second, conditional on these parameters, we estimate the parameters of the state-space model (omitting any auxiliary data),

$$\Phi_1 = (\alpha_1, \sigma_{v,z}^2, \mu, \sigma_{\varepsilon}^2).$$

Cunningham et al. (2007) note one reason to use this two-step procedure is that identification conditions may fail if all parameters were estimated together.<sup>9</sup> Moreover, the framework set

<sup>8</sup> The adoption of a chain-weighted price index in the middle of the sample adds an additional dynamic to the RGDP revision process. It would be ideal to use only post chain-weighted data; however the sample size is not sufficient for estimation.

<sup>9</sup> Cunningham et al. (2007) do not explore the satisfaction and/or violation of the relevant conditions; neither have we, although so doing seems a worthwhile task, to say the least.

**Table 2**  
**Estimated Revision-Bias Parameters**

RGDP	Estimate	Lower bound	Upper bound
$c^1$	0.5793	0.3317	0.8269
$\lambda$	-0.2828	-0.4515	-0.1141

NOTE: Upper and lower bounds represent 95 percent confidence intervals (CIs).

**Table 3**  
**Estimated  $\hat{v}$  Parameters**

RGDP	Estimate	Lower bound	Upper bound
Initial variance, $\sigma_\eta^2$	2.7002	2.3988	3.0017
Variance decay, $\delta$	-0.0786	-0.0931	-0.0641
First-order serial correlation, $\beta_1$	0.2004	0.1451	0.2557
Correlation with mature data, $\rho_{yv}^*$	0.3181	0.1399	0.4963

NOTE: Upper and lower bounds represent 95 percent CIs.

forth requires only the most recently published vintage of data, joint estimation of the parameters would require inputting the entire history of revisions into the model.

### Estimation of $\Phi_2$ Parameters

As noted previously, the first step of the estimation is to choose the revision horizon. Here it is  $N = 20$  (5 years). Our choice is explored in Appendix A. We input the  $W$  matrix (produced by equation (10)) into equation (11) to estimate values for the mean revision of the initial release,  $c^1$ , and the revision decay parameter,  $\lambda$ . For robustness purposes, Figure A2 shows the estimated and actual values of  $c^j$  at different horizons.

Table 2 reports the parameters estimated via GMM. The mean revision to the initial release is statistically different from zero. The initial release to RGDP is, on average, 0.57 percentage points lower than RGDP reported five years later. The revision decay parameter describes the rate at which revisions decay as the data mature: At revision maturity 2, RGDP is estimated to be 28 percent lower than the initial revision.

The next step is to calculate the correlation between the measurement error and the “true” unobserved state of the economy,  $\rho_{\varepsilon,\eta}$ . Because we do not observe the measurement error or the “true” state, we continue to use revisions to data of maturity  $j$  as a proxy of the measurement error. We assume data reported with maturity  $j+N$  are good estimates of the “true” state of the economy. The estimated mean correlation between the revisions of maturity  $j$  and the reported values at  $j+N$  are used as an estimate for  $\rho_{\varepsilon,\eta}$  denoted  $\rho_{y,v}^*$ ; that is,

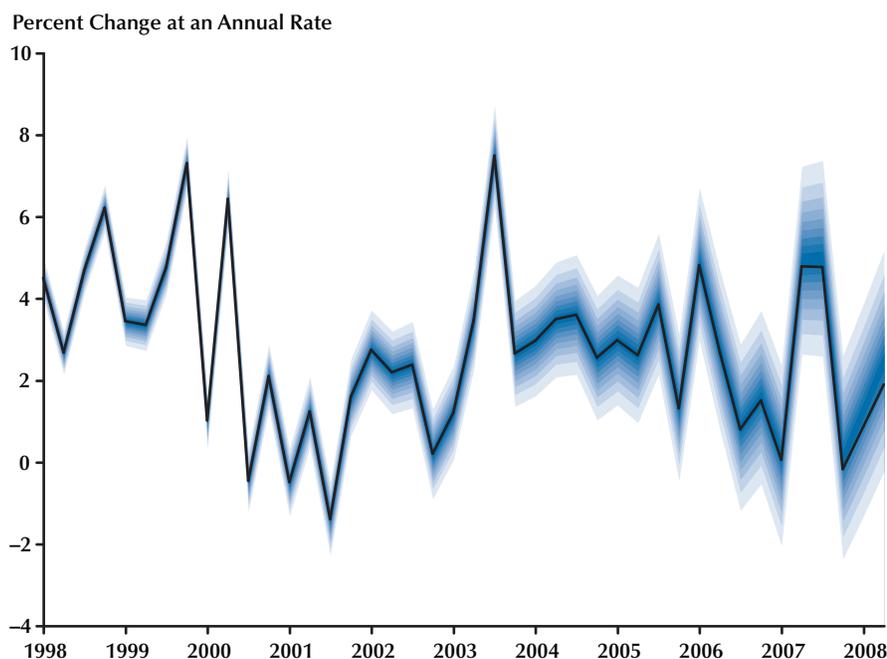
$$\rho_{\varepsilon,\eta} \approx \rho_{y,v}^* = \frac{1}{J} \sum_{j=1}^J \rho_{y^{j+N}, W(j,\cdot)}$$

The last row of Table 3 reports estimates for  $\rho_{yv}^*$ . The correlation between revisions to the data and the estimated “true” state are positive; although it is not reported, the correlation is positive for all  $j$ . Appendix Figure A3 explores the choice of the revision horizon: The values of  $\rho_{yv}^*$  stabilize for sufficiently large revision horizons.

The final set of first-stage estimates—the serial correlation between revisions, the initial

## Figure 2

### Actual and Filtered RGDP Growth (vintage 07/31/2008)



NOTE: The fan chart depicts the probability of various outcomes for RGDP after the data are “fully revised” (i.e., data reported in 5 years). Fully revised data are expected to fall within the fan chart 90 percent of the time. Each pair of shaded regions indicates an additional 10 percent CI.

variance of revisions, and the variance decay rate,  $\beta_1, \sigma_{\varepsilon_1}^2, \delta$ , respectively—are derived from the variance-covariance matrix of  $W$ , denoted  $\hat{V}$ .

The parameters are estimated per equations (12) and (13). Table 3 reports the estimated parameters for our preferred  $N$ . All estimates are significantly different from zero. Notice that the first order–serial correlation in the revisions is positive. The initial variance is markedly higher from that assumed in the simulation. However, the variance decay parameter is  $-0.07$ , which is close to the  $-0.06$  value used in the simulation.

### Estimation of $\Phi_1$ Parameters

Using the parameters estimated in the previous section, the vector of recently published data is put into the state-space model.<sup>10</sup> The parameter driving the state-space model’s covariance matrix,

$\sigma_{\varepsilon}^2$ , the variance of the shock to the  $AR(q)$  data-generating process for the “true” data, is estimated to be 3.55 (0.55); for U.K. investment data, Cunningham et al. (2007) report an error variance of 3.22 (0.67). Estimation results are shown in Figure 2. The solid black line is the most recently published data, the darkest band is the mean filtered value, and the outermost band is the 90 percent confidence interval (CI). As the variance of the revisions decay, so does the CI.

The RGDP growth rate at the most recent data point, 2008:Q2, was initially published as 1.89 percent on July 31, 2008. The estimated value is

<sup>10</sup> Estimation of the model is problematic. Although the data-generating processes for both the “true” data and the measurement error are initially asserted to be  $AR(q)$ , in the model the  $AR(q)$  parameters are not identified. The parameters are also omitted from estimation in Cunningham et al. (2007). Absent promising findings in the next section, far more estimation is necessary before confidence may be placed in such results.

2.47 percent, with 90 percent CIs between 5.17 percent and 10.23 percent. As of the February 27, 2009, release, 2008:Q2 RGDP was reported as 2.83 percent.

This state-space modeling framework shares features with multivariate stochastic volatility models. Harvey, Ruiz, and Shephard (1994) introduce such a model as an alternative to generalized autoregressive conditional heteroskedasticity (GARCH) models for high-frequency data. However, their model's problem differs from the current model. In their model, "multivariate" refers to four countries' exchange rates, modeled together, rather than 20 or more maturities of a single activity date. Their problem is similar to the current model, though, to the extent that the stochastic variance is assumed to follow an AR(1) process. This line of econometrics deserves further investigation.

## REAL-TIME MODEL EVALUATION

Using real-time RGDP data series to evaluate model accuracy follows closely the methodology of our simulation exercise. The main restriction with the actual data is that we do not observe the true values of each datum. We proxy the true values from data that have become "fully mature" at time  $t+N$  (where  $N=20$ ). Our real-time sample is restricted to those vintages of data with 10 years of data preceding (to estimate the parameters in  $\Phi_2$ ) and 5 years of data following (to evaluate the forecast) the vintages of interest. This exercise uses the data range 1985:Q4–2003:Q2 and vintages between  $v_0 = 01/30/2002$  and  $v_k = 07/31/2003$  as the most recently published data. This does not limit our ability to make real-time forecasts with the current data; however, it does inhibit us from testing the forecasting performance for 5 years.

We estimate the model for vintages  $v_0, v_1, \dots, v_k$  independently, keeping the number of observations, and maturities, fixed across vintages. For each successive release, we omit the oldest datum and add the most recent. This process corresponds nicely with the idea of running  $k$  iterations of the model simulation. The metric used to evaluate the model performance is the ratio of the RMSE

using the filtered datum as a predictor of the "mature" datum relative to the RMSE using the published datum as a predictor of the "mature" datum.

As described previously, the first stage in the real-time forecasting exercise is to estimate the parameters in  $\Phi_2$ . Given a revision horizon of  $N=20$  (5 years) for RGDP, at least 10 years of vintage data are used to estimate the  $W$  matrix and corresponding parameters.<sup>11</sup> For each successive vintage, we update the dataset (i.e., add a column to the  $W$  matrix) and reestimate the parameters in  $\Phi_2$ .

The stability of each parameter is assessed in Figure 3. The horizontal axis reports the vintage. 01/30/02 indicates that data available only on or before January 30, 2002, were used in the estimation of the parameters, thus the final column of the revision matrix  $W$  contains revisions to data released 20 quarters prior. The latest data points are identical to those reported in Tables 2 and 3.

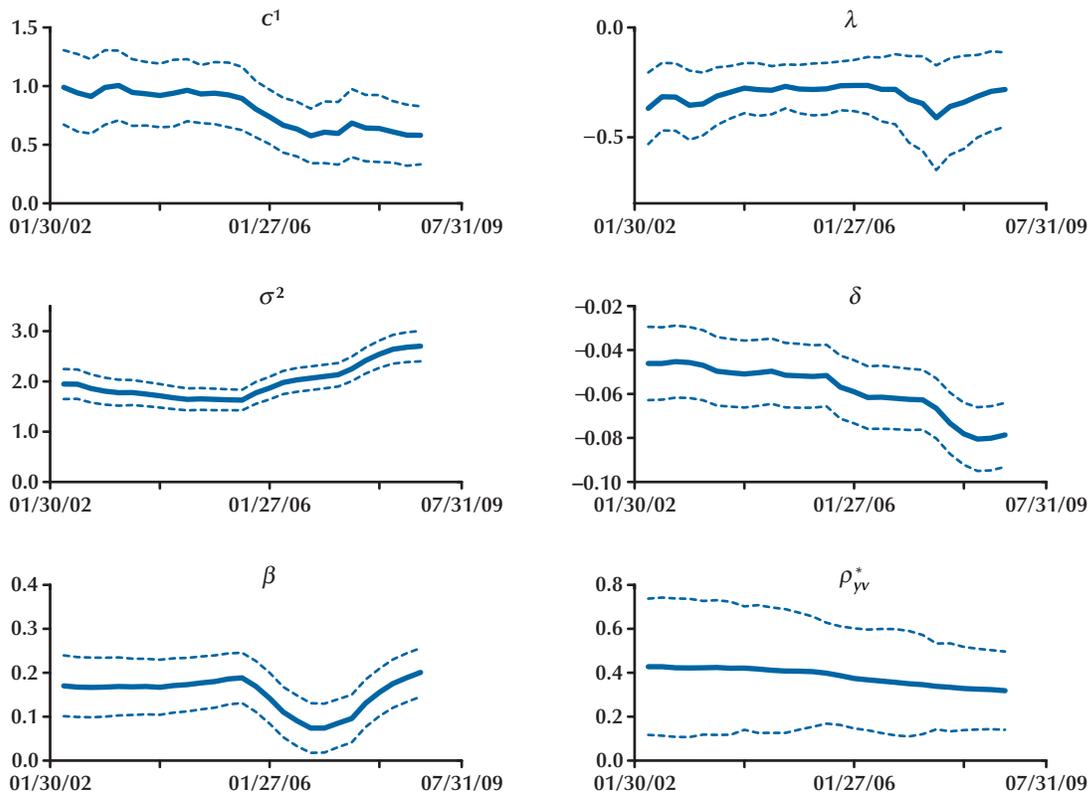
The mean revision of the initial release (top-left panel of Figure 3) steadily decreases as the real-time sample includes more recent data, indicating some improvement in the BEA's ability to report less-biased estimates of the "true" values. Conversely, the initial variance of the measurement error (left-middle panel) increases, indicating increased uncertainty around the initial estimates. The decay parameters corresponding to the initial mean and variance of the revisions are reported by  $\lambda$  and  $\delta$ , respectively.

The parameters in Figure 3 are displayed for all available vintages; however, the real-time forecasting exercise uses only the vintages for which the fully mature data are available. As noted earlier, this reduces the real-time sample to only the first seven vintages of data. For example, the mature values for January 30, 2002, data are reported after 20 quarters, on January 31, 2007.

The top panel of Figure 4 plots the RMSEs of the published data and of the filtered values. The bars in the bottom panel measure the difference

<sup>11</sup> We require that the  $W$  matrix have at least as many vintages (or columns) as it has maturities (or rows). The calculation in equation (10) requires at least  $N$  vintages of data as well as the observed "true" values. In other words, for every vintage,  $v$ , we must also observe the data at  $v+N$ .

**Figure 3**  
**Estimated  $\Phi_2$  Parameters in Real Time**



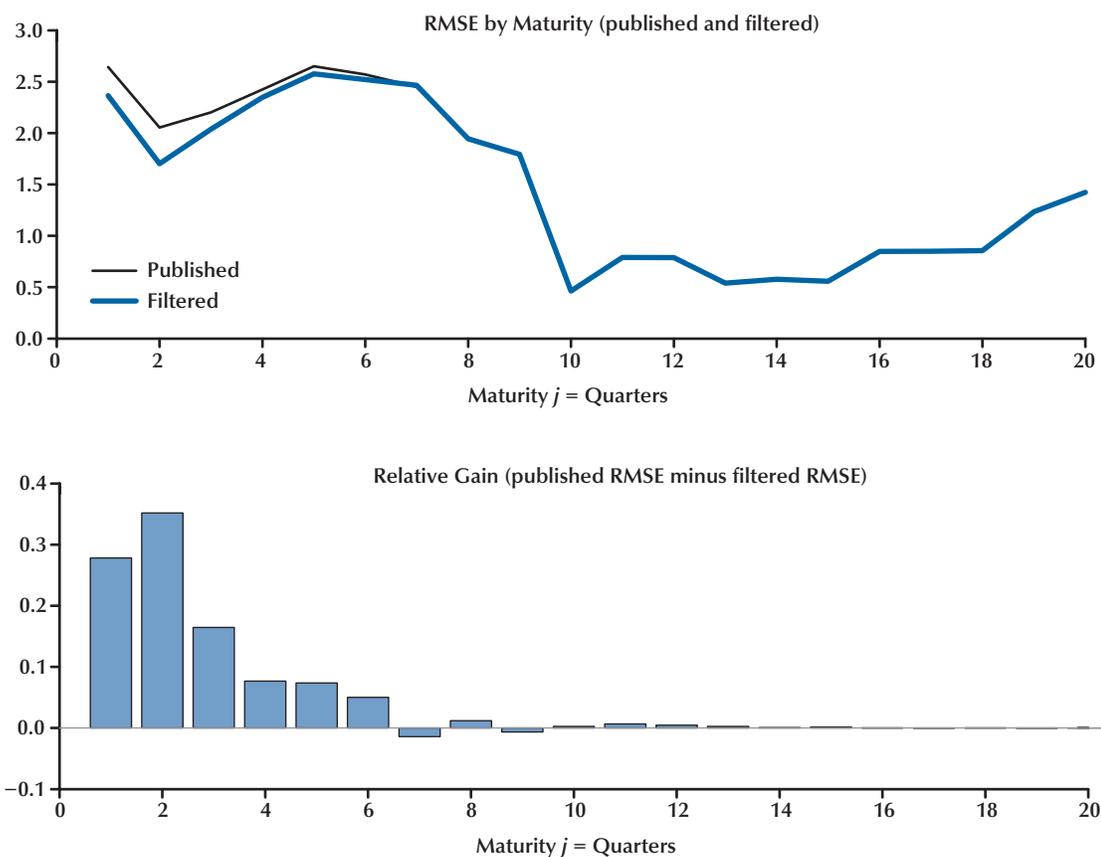
between the two series.<sup>12</sup> In some ways, the results are similar to those simulated in Figure 1: The filtered values tend to be superior estimates for the first 6 quarters and then diminish. Unlike in the simulation, however, the transition between quarters is not particularly smooth. In the top panel of Figure 4, the RMSEs of both series actually increase for data maturities 3 through 6 and steeply decline thereafter. According to the bars in the bottom panel of Figure 3, for maturity 2, the data from the filtered values show greater improvement than those from the initial release.

For further examination, Table 4 reports the improvement due to the state-space filter for the

seven vintages. The first two columns report the vintages of the published data and the fully mature data, respectively. The remaining five columns report the improvement due to the state-space filter for data of varying maturities. The bottom row of the table reports the average across the seven vintages. During the first year, the average RMSE of the filtered values is 87 percent of the average RMSE of the published data. The filtered values most improve the data published July 31, 2003: The RMSE of the filtered values is 48 percent of the RMSE of the published data.<sup>13</sup> For data 2 years old, there is only modest improvement: The RMSE of the filtered values is 97 percent of

<sup>12</sup> The following results should be interpreted with some caution; they are constrained by only seven consecutive releases of data and additional data points may drastically alter the results.

<sup>13</sup> This outlier is driven by a particularly inaccurate initial release of 2.37 percent for 2003:Q1, the filtered value was 3.74 percent, and the value on January 31, 2007, was 3.46 percent.

**Figure 4****Real-Time Model Performance, RGDP****Table 4****Real-Time Model Performance**

Release of published data	Release of fully mature data	Improvement due to state-space filter ( $RMSE_{\text{filtered}}/RMSE_{\text{published}}$ )				
		Maturities 1-4 (Year 1)	Maturities 5-8 (Year 2)	Maturities 9-12 (Year 3)	Maturities 13-16 (Year 4)	Maturities 17-20 (Year 5)
1/30/2002	1/31/2007	0.9387	0.8561	1.0018	1.0018	1.0002
4/26/2002	4/27/2007	0.8933	0.9988	0.9885	0.9957	1.0002
7/31/2002	7/27/2007	0.9056	0.9601	1.0179	0.9942	1.0003
10/31/2002	10/31/2007	0.9184	0.9980	1.0128	0.9977	1.0000
1/30/2003	1/30/2008	0.9809	0.9725	1.0083	0.9985	1.0000
4/25/2003	4/30/2008	0.9900	0.9815	1.0254	0.9979	0.9988
7/31/2003	7/31/2008	0.4852	1.0039	1.0032	0.9997	0.9996
Average		0.8731	0.9673	1.0083	0.9979	0.9999

the RMSE of the published data. Thus, after data have been revised for 2 years, there is no apparent gain from the filtered values.

In addition to the improved estimates of the true values, as measured by the RMSEs, the filtered values also provide CIs, which are not provided by data releases. The CIs indicate the extent to which incoming data are likely to be revised, providing some assessment as to how much weight to assign to each datum. Of the 504 mature data points observed, approximately 83 percent fell within the 90 percent CI and approximately 41 percent within the 50 percent CI. These numbers seem reasonable, as the sample consists of seven consecutive vintages—meaning any given outlier could be repeated up to seven times within our sample.

## CONCLUSION

A long line of papers has explored methods to pool vintages of economic data, seeking to extract the true (or, at least strongest) underlying signal for a variable of interest. A recent, and likely fruitful, path is to introduce a cohort-style analysis that examines the revisions as a function of the age of the data and estimates the “true” unobservable values via a state-space framework. Here, we have begun the application of such techniques to U.S. data, specifically using a measure of RGDP. The framework is equally applicable to quarterly or monthly data, although we have not yet considered the case of mixed frequencies (including when monthly observations are published quarterly, such as for GDP).

Monte Carlo experiments suggest that, for a wide range of parameter values in AR data-generating processes, the framework explored here may be able to extract estimates of recent values of economic variables and reduce uncertainty by as much as 30 percent. Obviously, empirical application of such techniques introduces statistical challenges when pooling data across cohorts. The “revision horizon,” to a large extent, is an arbitrary selection, and robustness experiments are required. Further, if underlying unobserved true data are to be recovered as a state

vector, issues regarding the lack of statistical identification require further exploration. It appears, however, even with these caveats, the modeling framework does provide estimators for two important variances—the variance of the empirical measurement error embedded in each published datum and the variance of the data-generating process of the true underlying economic variable.

Real-time experiments, albeit with limited data, suggest that uncertainty in RGDP estimates appear to be reduced by close to 10 percent at early maturities. In addition, CIs extracted from the model provide information unattainable from data releases alone. Both, perhaps, will assist economists and policymakers, by providing a set of “revision CIs” around releases of incoming data. One limitation of the application of the methodology is the large amount of data required to produce and evaluate estimates in real time. Nonfarm payroll employment data, with monthly revisions and a long release history, is a good candidate for the application of this methodology.

## REFERENCES

- Aruoba, S. Borağan. “Data Revisions Are Not Well-Behaved.” *Journal of Money, Credit and Banking*, March-April 2008, 40(2-3), pp. 319-40.
- Aruoba, S. Borağan; Diebold, Francis X. and Scotti, Chiara. “Real-Time Measurement of Business Conditions.” Unpublished manuscript, April 2007; 2008 version: Working Paper 08-19, Federal Reserve Bank of Philadelphia; [www.philadelphiafed.org/research-and-data/publications/working-papers/2008/wp08-19.pdf](http://www.philadelphiafed.org/research-and-data/publications/working-papers/2008/wp08-19.pdf).
- Boskin, Michael J. “Getting the 21st Century GDP Right: Progress and Challenges.” *American Economic Review*, May 2000, 90(2), *AEA Papers and Proceedings*, pp. 247-52.
- Croushore, Dean. “Revisions to PCE Inflation Measures: Implications for Monetary Policy.” Unpublished manuscript, University of Richmond, November 2007.

- Croushore, Dean and Stark, Tom. "A Funny Thing Happened on the Way to the Data Bank: A Real-Time Data Set for Macroeconomists." Federal Reserve Bank of Philadelphia *Business Review*, September/October 2000, pp. 15-27; [www.phil.frb.org/research-and-data/publications/business-review/2000/september-october/brso00dc.pdf](http://www.phil.frb.org/research-and-data/publications/business-review/2000/september-october/brso00dc.pdf).
- Croushore, Dean and Stark, Tom. "Data Revisions and the Identification of Monetary Policy Shocks." *Journal of Econometrics*, November 2001, 105(1), pp. 111-30.
- Cunningham, Alastair; Eklund, Jana; Jeffrey, Chris; Kapetanios, George and Labhard, Vincent. "A State Space Approach to Extracting the Signal from Uncertain Data." Working Paper 336, Bank of England, November 2007; [www.bankofengland.co.uk/publications/workingpapers/wp336.pdf](http://www.bankofengland.co.uk/publications/workingpapers/wp336.pdf).
- Diebold, Francis X. and Rudebusch, Glenn D. "Forecasting Output with the Composite Leading Index: A Real-Time Analysis." *Journal of the American Statistical Association*, September 1991, 86(415), pp. 603-10.
- Durbin, James and Koopman, Siem J. *Time Series Analysis by State Space Methods*. Oxford: Oxford University Press, 2001.
- Faust, Jon; Rogers, John H. and Wright, Jonathan. "News and Noise in G-7 GDP Announcements." *Journal of Money, Credit, and Banking*, June 2008, 37(3), pp. 403-19.
- Garratt, Anthony; Koop, Gary and Vahey, Shaun P. "Forecasting Substantial Data Revisions in the Presence of Model Uncertainty." *Economic Journal*, July 2008, 118(530), pp. 1128-44.
- Garratt, Anthony; Lee, Kevin; Mise, Emi and Shields, Kalvinder. "Real Time Representations of the Output Gap." *Review of Economics and Statistics*, November 2008, 90(4), pp. 792-804.
- Grimm, Bruce T. and Weadock, Teresa L. "Gross Domestic Product: Revisions and Source Data." *Survey of Current Business*, February 2008, 86(8), pp. 11-15.
- Harvey, Andrew; Ruiz, Esther and Shephard, Neil. "Multivariate Stochastic Variance Models." *Review of Economic Studies*, April 1994, 60(2), pp. 247-64.
- Howrey, E. Philip. "The Use of Preliminary Data in Econometric Forecasting." *Review of Economics and Statistics*, May 1978, 60(2), pp. 193-200.
- Howrey, E. Philip. "Data Revision, Reconstruction, and Prediction: An Application to Inventory Investment." *Review of Economics and Statistics*, August 1978, 66(3), pp. 386-93.
- Jacobs, Jan P.A.M. and van Norden, Simon. "Modeling Data Revisions: Measurement Error and Dynamics of 'True' Values." CCSO Working Paper 2006/07, CCSO Centre for Economic Research, December 2006; [www.eco.rug.nl/ccso/quarterly/200607.pdf](http://www.eco.rug.nl/ccso/quarterly/200607.pdf).
- Kim, Chang-Jin and Nelson, Charles R. *State-Space Models with Regime Switching*. Cambridge, MA: MIT Press, 1999.
- Kishor, N. Kundan and Koenig, Evan F. "VAR Estimation and Forecasting When Data Are Subject to Revision." Working Paper 0501, Federal Reserve Bank of Dallas, February 2005; <http://dallasfed.org/research/papers/2005/wp0501.pdf>.
- Kozicki, Sharon. "How Do Data Revisions Affect the Evaluation and Conduct of Monetary Policy?" Federal Reserve Bank of Kansas City *Economic Review*, First Quarter 2004, pp. 5-37.
- Landerfeld, J. Steven; Seskin, Eugene P. and Fraumeni, Barbara M. "Taking the Pulse of the Economy: Measuring GDP." *Journal of Economic Perspectives*, Spring 2008, 22(2), pp. 193-216.
- Mankiw, N. Gregory and Shapiro, Matthew D. "News or Noise? An Analysis of GNP Revisions." NBER Working Paper No. 1939, National Bureau of Economic Research, June 1986; [www.nber.org/papers/w1939.pdf?new\\_window=1](http://www.nber.org/papers/w1939.pdf?new_window=1).
- Orphanides, Athanasios and van Norden, Simon. "The Unreliability of Output-Gap Estimates in Real Time." *Review of Economics and Statistics*, November 2002, 84(4), pp. 569-83.

**Anderson and Gascon**

Orphanides, Athanasios and van Norden, Simon. "The Reliability of Inflation Forecasts Based on Output Gap Estimates in Real Time." *Journal of Money, Credit, and Banking*, June 2005, 37(3), pp. 583-601.

Sargent, Thomas. "Two Models of Measurements and the Investment Accelerator." *Journal of Political Economy*, April 1989, 97(2), pp. 251-87.

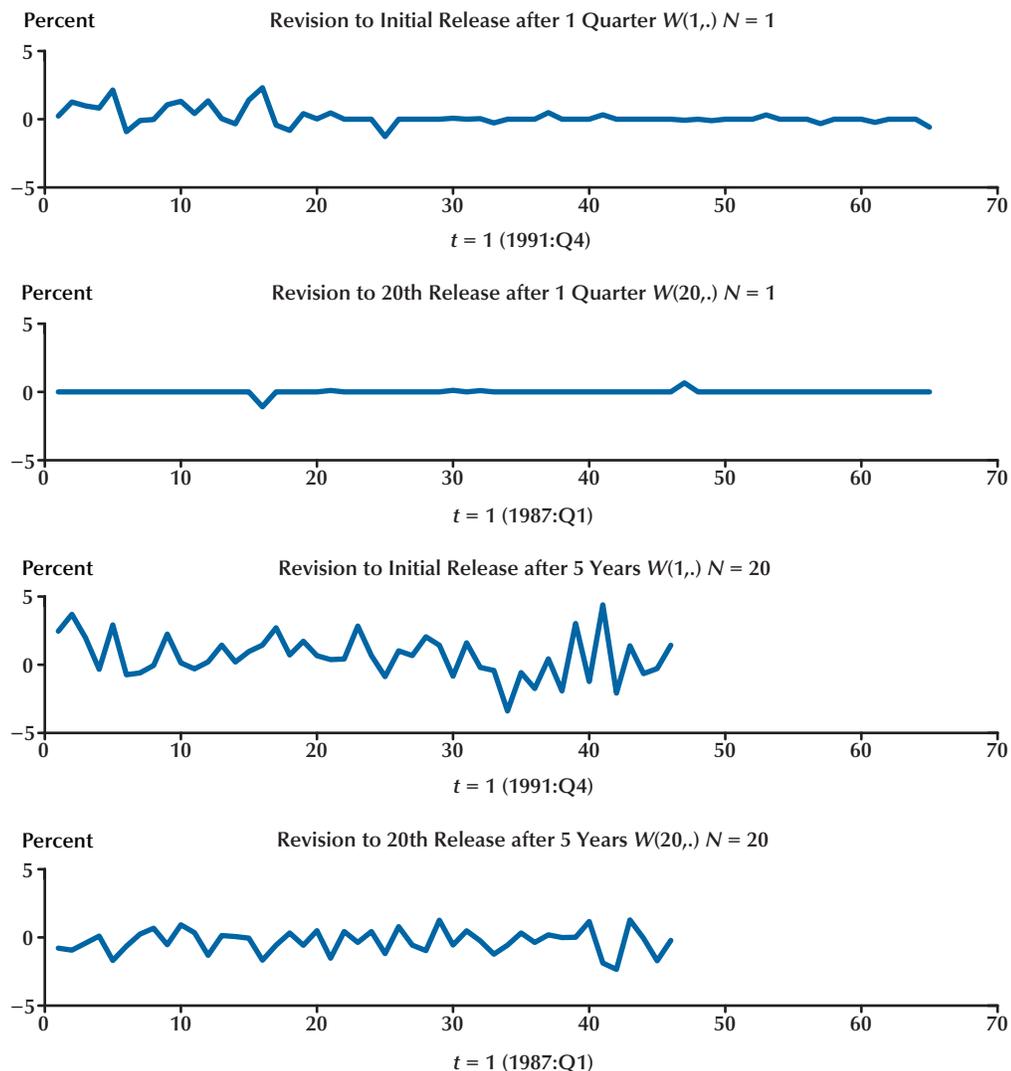
Stock, James H. and Watson, Mark W. "Why Has U.S. Inflation Become Harder to Forecast?" *Journal of Money, Credit, and Banking*, January 2007, 39(1), pp. 3-33.

## APPENDIX A: SELECTION OF REVISION HORIZON ( $N$ )

Figure A1 shows the rows of the  $W$  matrix produced by equation (10) for RGDP over time at two different maturities,  $j$ , and horizons,  $N$ . The top panel shows the revision to the initial release after 1 quarter; the second panel shows the revision to the 20th release after 1 quarter. As expected, the revisions to the 20th release after 1 quarter tend to be zero. When the horizon is extended from 1 quarter to 5 years (bottom two panels), the 20th release does exhibit revision. Figure A2 shows the estimated and actual revision process of data subject to a revision horizon  $N = 1, 5, 10$ , and 20. Figure A3 shows the correlation between revisions and the “true” data subject to a revision horizon.

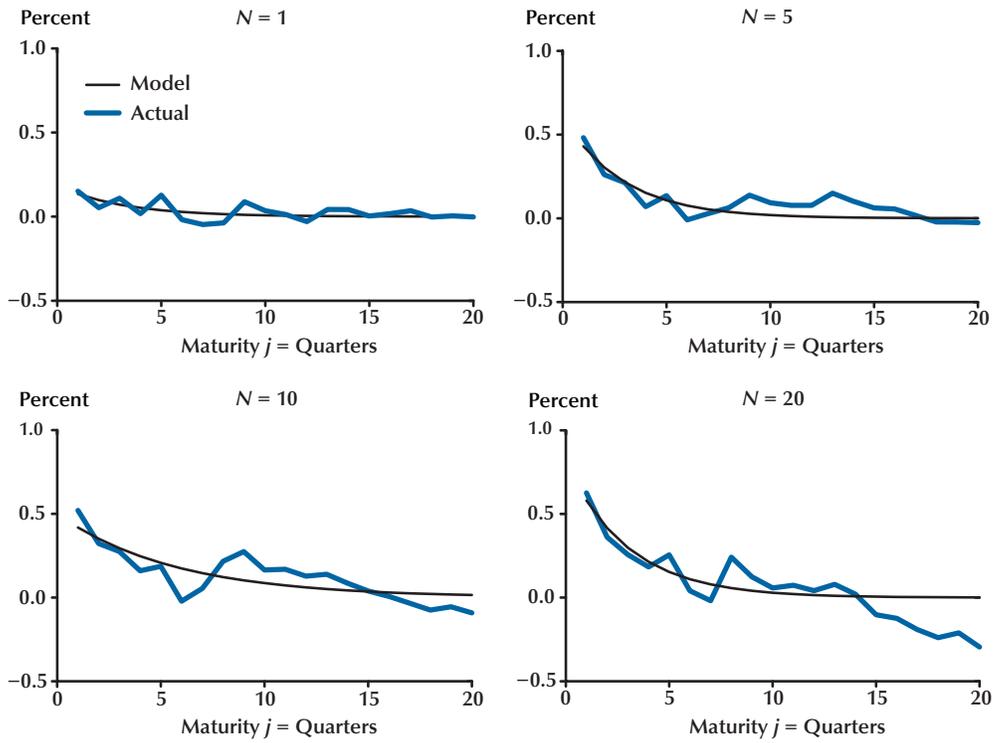
### Figure A1

#### Revisions to Initial Release and 5-Year-Old RGDP Data at Different Horizons



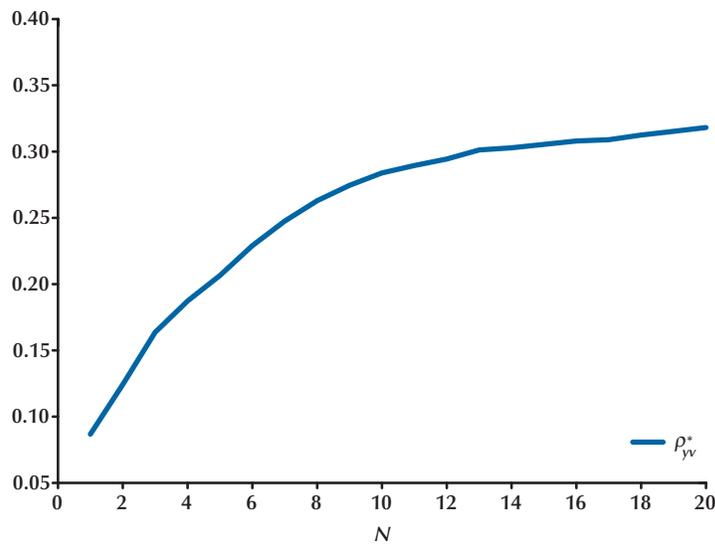
### Figure A2

Mean Revisions to RGDP by Maturity,  $j$ , at Revision Horizon  $N$



### Figure A3

Correlation of Revisions Between Maturity,  $j$ , and Published Estimates at Maturity,  $j + N$



## APPENDIX B: ABOUT THE DATA

### *How the Real GDP Data Are Created*

The Federal Reserve Bank of St. Louis ALFRED database allows researchers to retrieve vintage versions of economic data that were available on specific dates in history. Most data are available in real and nominal terms. If a researcher is interested in one vintage of data, the real series may be suitable; however, in our case we are interested in all vintages of real GDP. As reported in ALFRED, the unit of measure on the real series changes by vintage. For example, between December 4, 1991, and January 18, 1996, real GDP is reported in billions of 1987 dollars, whereas between January 19, 1996, and October 28, 1999, the series is reported in billions of chained 1992 dollars. Due to changes in the deflator, it is not suitable to obtain the real series from ALFRED and simply calculate the revisions. As an alternative, the nominal GDP (GDPA), GDP, and implicit price deflator (GDPDEF) series are used to create a vintage real GDP series.<sup>14</sup>

As with real GDP, the unit of measure of GDPDEF changes across vintages. Therefore, before deflating GDPA, GDPDEF must be reindexed. The data available in ALFRED are “as reported,” meaning the base year varies from 1987 = 100 for vintages before January 18, 1996, to 2000 = 100 for vintages after December 12, 2003. Further complicating the issue, the data released in the base years (1985, 1992, 1996, and 2000) are also subject to revision; therefore the indexing of GDPDEF can also change between vintages within the same base year. Because we are interested in revisions to the data resulting from new information, and not simply changes in the base year, we reindex all GDPDEF data to a constant base year. To match the new series to the most recently reported data, we choose to index all of the data by setting 2000 = 100 in the July 31, 2008, release. We denote the new deflator series DEFL.

The real GDP series are constructed by multiplying each date and vintage of the GDPA by the corresponding date and vintage of DEFL. After deflating the data, annualized growth rates of each vintage are calculated, and we denote the resulting series RGDP.

Because the models are not well suited for mixed-frequency data,<sup>15</sup> we elect to use only the data vintages in which a new advance estimate is released. Consistent with our dataset, the first maturity ( $n = 1$ ) in national income and product accounts (NIPA) data is the advance estimate. In the NIPA data from ALFRED, the preliminary estimate would be the second maturity; however, we omit this vintage, as well as the final estimate. We label the fourth release, which is released at the same time as the subsequent quarter’s advance estimate, as the second maturity ( $n = 2$ ).

Table B1 presents a stylized real-time dataset after the preliminary and final vintages have been removed from the data. The columns denote the data vintages; the rows denote the dates of the observations. For descriptive purposes, each element in the dataset is reported with a superscript identifying the maturity,  $j$ , of the observation.

The analysis in this paper hinges on the value chosen for the maturity horizon, or “look-ahead distance,” denoted  $J$ . The value of  $J$  is the assumed horizon at which the data are assumed to be true, in that no further revisions to the data will occur. This paper is absent a discussion about the appropriate horizon. Our visual inspection of the data, summarized in Appendix A, and data limitations lead us to set a 5-year horizon ( $J = 20$ ) for GDP and RGDP. For robustness purposes, in Figures A1, A2, and A3 all parameters in  $\Phi_2$  are reported for alternative values for  $J$ .

<sup>14</sup> The GDPDEF is chosen over the preferred chain-type price index (GDPCTPI) when available. The oldest vintage for GDPDEF is December 4, 1991, whereas the oldest vintage for GDPCTPI is January 19, 1996.

<sup>15</sup> The BEA releases the quarterly GDP series at a monthly frequency: The first release is the advance, the second the preliminary, and the third the final release.

**Table B1****Real-Time Dataset: Annualized Growth Rate of Real GDP**

Activity date ( $t$ )	Vintage ( $v$ )						
	1999:Q3	1999:Q4	2000:Q1	...	2008:Q1	2008:Q2	2008:Q3
1999:Q2	5.6 <sup>1</sup>	4.7 <sup>2</sup>	4.7 <sup>3</sup>	...	6.3 <sup>35</sup>	6.3 <sup>36</sup>	6.3 <sup>37</sup>
1999:Q3		6.8 <sup>1</sup>	7.9 <sup>2</sup>	...	7.7 <sup>34</sup>	7.7 <sup>35</sup>	7.7 <sup>36</sup>
1999:Q4			10.1 <sup>1</sup>	...	11.0 <sup>33</sup>	11.0 <sup>34</sup>	11.0 <sup>35</sup>
⋮				⋮	⋮	⋮	⋮
2007:Q4					5.8 <sup>1</sup>	5.5 <sup>2</sup>	4.9 <sup>3</sup>
2008:Q1						5.9 <sup>1</sup>	6.10 <sup>2</sup>
2008:Q2							4.16 <sup>1</sup>

NOTE: Superscripts denote maturity,  $j$ . Following the notation in the paper,  $y_t^j$  denotes  $y$  at time  $t$  at maturity,  $j$ , or  $y_{2007:Q4}^3$  equals 4.9 percent.

**Why Are NIPA Data Revised?**

Clearly, revisions to NIPA data are not caused by statisticians at the BEA finding computational errors and fixing them. Two main causes of such revisions to NIPA data are that over shorter horizons new data become available (thus prompting revisions) and over longer horizons methodology changes. Statisticians and economists at the BEA are well aware of these problems and over time have made significant updates to the data collection and publication process. At the same time, this paper assumes that by mining the data and revision process we can more accurately predict the true values of a series of interest. We make this assumption not because of any inadequacy of the BEA's work, but rather because of the complexity of the task.

Short-term data revisions are largely a result of the tradeoff faced by the BEA. On one hand, there is pressure for timely releases of information; on the other hand, there is an assumption that the data released accurately measure the underlying variable of interest. Because of the desire for timely estimates, the BEA releases their first, or "advance," estimate with only 75 percent of data for the past quarter (Landefeld et al., 2008). The estimates of the true value are revised as more data become available. Table B2 outlines the four data types used to construct the GDP series as well as the total share of each for 2003:Q3, as reported Grimm and Weadock (2008) in the *Survey of Current Business*. Trend-based data are imputed data; complete data are data that have been reported for the quarter for all three months of the quarter; monthly trend-based data include two months of data and imputed-data for the third month of the quarter; and revised data are simply revised estimates of the complete data. Notice that the advance estimate ( $n = 1$ ) does not contain any revised data and less than half of the data is complete, whereas over three-fourths of the data in the final release ( $n \approx 2$ ) is complete or has been revised. At the time of the annual revision,<sup>16</sup> over 90 percent of the data is complete or has been revised. Detailed information on the data sources, revision process, and methodology used to create the NIPA data are provided by Landefeld et al. (2008).

<sup>16</sup> The maturity of these data is a function of  $t$ . For Q1 data the annual revision will occur at  $n \approx 4$ ; for Q2 data at  $n \approx 3$ ; for Q3 data at  $n \approx 2$ ; and Q4 data will not be subject to an annual revision until the next year, or  $n \approx 5$ .

**Table B2****Data Sources for Short-Term Revisions to GDP (percent)**

Data source	Share of 2003:Q3 GDP	
	Advance estimate	Final estimate
Trend-based data	25.1	20.9
Monthly and trend-based data	29.7	1.2
Complete data	45.3	8.4
Revised data	—	69.5

SOURCE: Grimm and Weadock (2008).

In addition to problems caused by the lack of data available, challenges exist in regard to quantifying the actions of economic agents, such as the growth in the service sector, identifying new products as they enter the economy, and quality improvements for existing products (see Boskin, 2000). Because of the large scale of these problems, the BEA normally addresses these issues of definitions and methodology in 5-year “benchmark” revisions. In forecasting the true values of GDP, we make no assumptions about the changes these revisions make. The inability of the model to forecast changes that occur during benchmark revisions is a shortcoming of our work as well as that of other scholars in this field.

