On the Reliability of Output-Gap Estimates in Real Time*

Elmar Mertens†
Federal Reserve Board

November 2, 2014

*I would like to thank Domenico Giannone, Mike Kiley, Dominique Ladiray, David Lopez-Salido, Simon van Norden (my discussant), Jeremy Rudd, John Roberts, Tara Sinclair as well as conference participants at the NBER-NSF Time Series Conference 2013, the 9th Annual CIRANO-CIREQ Workshop on Data Revisions in Macroeconomic Forecasting and Policy, the Society for Nonlinear Dynamics and Econometrics Conference 2014, the Annual Conference of the Canadian Economics Association 2014, the Joint Statistical Meetings 2014, the first Conference of the Society for Economic Measurement, and seminar participants at Boston College and the Federal Reserve Board for highly useful conversations and comments about this project.
Updates to this paper will be posted at www.elmarmertens.com/research/workingpapers/MertensOutputgapSV.pdf.

†For correspondence: Elmar Mertens, Board of Governors of the Federal Reserve System, Washington D.C. 20551. email `elmar.mertens@frb.gov`. Tel.: +(202) 452 2916. The views in this paper do not necessarily represent the views of the Federal Reserve Board, or any other person in the Federal Reserve System or the Federal Open Market Committee. Any errors or omissions should be regarded as solely those of the author.

On the Reliability of Output-Gap Estimates in Real Time

Abstract

Real-time estimates of the Output Gap — defined as the cyclical component of GDP — have previously been shown to be unreliable, since they are subject to large revisions when new data comes in. However, this result has so far only been derived for constant parameter models.

This paper uses statistical models where the volatility of shocks to trend and cycle can vary over time. In this case, output gap estimates derived from data vintages going back to the 1970s are much closer to "final" estimates derived from all available sample data. The final estimates not only fall mostly within the credible intervals generated by the real-time data. When generated from a model with stochastic volatility, these credible sets are also tighter, at least over low-volatility periods.

JEL Classification: C32, C53, E32

Keywords: Output Gap, Trend Cycle Decomposition, Real-time Data, Stochastic Volatility, Kalman Filter, MCMC, Particle Filter

1. INTRODUCTION

An important task in business cycle analysis is to decompose macroeconomic fluctuations into trend and cycle. The ability to distinguish between the actual level of economic activity and its trend level is not only useful for forecasting future GDP, but potentially also a useful measure of economic slack with relevance for understanding inflation. The notion of "trend" level adopted in this paper follows the empirical literature on trend-cycle decompositions, where the trend level of real GDP is defined as the long-term forecast for real GDP derived from a statistical model — this is similar though not identical to the detrending procedures of Beveridge and Nelson (1981).[1] In analysis of this kind, the relative volatilities of trend and cycle have an important influence on the estimates: When the volatility of the trend is, say, large relative to that of the cycle, swings in the data will largely be explained by trend changes, resulting in relatively small output gap estimate, and vice versa. For example, the recent recession clearly saw a big drop in the level output. Whether a trend-cycle decomposition explains this largely with a fall in trend GDP or sees the recent downturn rather as a cyclical phenomenon largely depends on the model's assessment of the relative variance between shocks to trend and cycle.

In an important couple of papers, Orphanides and van Norden (2002; 2005) have documented that real-time estimates of the output gap — defined as the deviation between actual output and trend — are subject to substantial revisions when new data become available. Apart from issues related to data revisions in official statistics, Orphanides and van Norden identify the end-point problem of statistical filters as the main cause of these revisions.

Orphanides and van Norden (2002) considered a wide range of univariate and bivariate statistical models, summarizing what has been current practice and state-of-the-art at the time of their writing. Notably, all models studied in their work are characterized by constant parameters, including the assumption of constant variances. Since then, advances in computational methods and the spread of simulation-based Bayesian methods lead to substantial developments in the usability

[1] Alternatively, the "potential" level of GDP could also be defined in the context of specific structural models like a New Keynesian DSGE model, see for example Woodford (2003), Galì (2003), Smets and Wouters (2007), Edge et al. (2008) and Justiniano et al. (2011).

of models with time-varying parameters. At the same time, the recognition has grown that postwar data for the U.S. (and other economies) has been characterized by distinct changes of various sorts. As a prominent example, it is now commonplace to refer the period of the mid-1980s to 2007 as the "Great Moderation" (Bernanke, 2004) since the volatility of many macroeconomic aggregates was markedly lower over that period, compared to the 1970s and early 1980s, and likely also when compared to the recent years characterized by the deep recession of 2007-2009. These changes in the overall volatility of GDP and other macroeconomic variables may also have been accompanied by changes in the relative variances of trend and cycle, which are crucial for determining the weights that a statistical filter attaches to the likelihood that a given change in the data reflects a movement in trend or cycle. The estimates presented here suggest that this has indeed been the case, thus causing a substantial improvement in output gap estimates obtained from a model that accounts for such variance changes.

This paper investigates the implications from these well-document changes in macroeconomic volatility when the output gap is modeled with stochastic volatility. To keep things simple and illustrate the potential interest of the technique implemented, a univariate model, using only data on real GDP, and a bivariate in real GDP and unemployment are considered. In its constant-parameter version, the univariate model for real GDP corresponds to a model used by Harvey and Jaeger (1993) and Clark (1987) and was one of the better (though still unreliably) performing models in the study by Orphanides and van Norden (2002). The bivariate model embeds an Okun's law assumption by which real GDP and unemployment variables share a common cyclical factor. Both models jointly estimate the decomposition of GDP (and unemployment) into trend and cycle together with estimates of stochastic volatility and other model parameters.

The models are estimated for all quarterly data vintages since 1970, which are available from the Real-Time Data Research Center at the Federal Reserve Bank of Philadelphia. A key finding is that real-time estimates of the output gap from the stochastic-volatility versions of each model are closer to "final" estimates derived from all available sample data. The final estimates not only fall mostly within the confidence sets generated by the real-time data, these confidence sets are also tighter, when estimated from a model with stochastic volatility — at least over periods with low volatility. When variances are assumed to be constant, model estimates display large revisions, similar to what has been reported before in the literature. Compared to the univariate Harvey-Clark model, output gap revisions are smaller in my bivariate model, in part because of the inclusion of unemployment data.[2] In either case, by changing the size of economic shocks over time, my stochastic volatility models can simultaneously account for small output gaps during the Great Moderation and larger swings during the Great Recession of the recent years as well as in the 1970s.

The model estimates also deliver a fairly smooth estimate of trend GDP and a virtually flat estimate of the trend in unemployment since the 1960s. These trend levels correspond to the model's forecast of the long-run levels of GDP and unemployment. These trend levels are not necessarily identical to notions of natural rates of unemployment or GDP — a property that the model shares with other statistical decompositions. By identifying trend levels from long-run forecasts of the underlying variables, the specification of the model does not take a stand on whether the resulting gap measures reflect variations in aggregate demand, but rather whether these variations are estimated to have long-run effects on the economy or not.[3]

Of related interest, Edge and Rudd (2012) have recently studied the quality of judgmental output gap estimates, produced by the Federal Reserve staff, and found a considerable reduction in the size of ex-post revision over recent decades, in contrast to statistical measures (derived from constant-parameter models). A potential hypothesis explaining their results and mine, would be to suggest that adjustments due to time-varying volatility play an important role in judgmental estimates of the output gap, and I plan to investigate this hypothesis further.

[2]Barnes et al. (2013) find that Okun's Law errors predict future GDP revisions. As the bivariate model used here jointly estimates Okun's Law as well as processes for trend unemployment and GDP, the model can condition on this information.

[3]The trend concept used here corresponds to the ones underlying the identification strategies of Shapiro and Watson (1988) and Blanchard and Quah (1989) for identifying shocks to "supply" (or "technology") from shocks to the permanent component of real output. However, under a broader set of assumptions about the economic environment, such permanent effects could also be induced by shock to aggregate demand, arising for example from fiscal policy Galì (1999). Recent evidence in favor of such effects from shocks to aggregate demand on the trend component of unemployment is discussed by Reifschneider et al. (2013).

The remainder of this paper is organized as follows: Section 2 lays out the trend concept and estimation strategy used throughout the paper. This section also describes the potential role of stochastic volatility in reducing estimation uncertainty, and how this might also reduce the size of revisions between “filtered” and “smoothed” estimates of the output gap. Section 3 applies these ideas to the univariate model of Clark (1987) and Harvey (1985) and largely confirms the findings of Orphanides and van Norden (2002) about the unreliability of output gap estimates in real time; the section then also studies a version of the model with stochastic volatility which considerably decreases the size of output gap revisions. Section 4 describes my preferred bivariate model and analyzes output gap revisions from the model with and without stochastic volatility. Section 5 provides a forecast evaluation. Section 6 gives an overview of the related literature and Section 7 concludes this paper.

2. TREND CONCEPT AND REAL-TIME MEASURES

Similar to Orphanides and van Norden (2002) and other statistical approaches, this paper considers unobserved component models that decompose the level of real GDP (y_t) into the sum a of a trend component ($\bar{y}_t$), which is non-stationary, and a stationary component ($\tilde{y}_t$) which will be referred to as the “output gap”.[4] (Section 6 surveys other, more structural definitions that are also used elsewhere in the literature.) The different models used in this paper specify different dynamics for the output gap and for the relationship between output and the unemployment rate, but throughout the trend component will be modeled as a random walk with a time-varying drift (μ_{t-1}). The time-varying drift serves to capture low-frequency changes in average GDP growth of the U.S. economy over the postwar period and is supposed to follow a random walk.[5]

[4]Henceforth, the terms “output” and “GDP” shall always refer to real quantities, expressed in logs and denoted by y_t, unless noted otherwise.

[5]Average growth rates in real GDP have mostly been trending downwards in U.S. data over the last five decades or so. These low-frequency changes have for example been noted by Blanchard and Fischer (1989, Chapter 1), Gordon (2012) and Stock and Watson (2012).

$$y_t = \bar{y}_t + \tilde{y}_t \qquad \tilde{y}_t \sim I(0) \qquad E(\tilde{y}_t) = 0 \tag{1}$$

$$\bar{y}_t = \bar{y}_{t-1} + \mu_{t-1} + \bar{\varepsilon}_t^y \qquad \bar{\varepsilon}_t^y \sim N(0, \bar{\sigma}_{y,t}^2) \tag{2}$$

$$\mu_t = \mu_{t-1} + \eta_t^\mu \qquad \eta_t^\mu \sim N(0, \phi_\mu) \tag{3}$$

The correlation between shocks to the trend level ($\bar{\varepsilon}_t^y$) and the output gap will be specified later (and differently so for different models). Throughout, the shock to the drift rate (η_t^μ) is assumed to be iid and uncorrelated with other model components. The trend in output will thus be subject to two disturbances, $\bar{\varepsilon}_t^y$ and η_t^μ. When considering potential changes in the variances shocks to trend and cycle, I will focus on stochastic volatility in the shocks to the trend level $\bar{\varepsilon}_t^y$ and shocks to the output gap $\tilde{y}_t$. In order to limit the number of time-varying parameters and reflecting the belief that μ_t should capture slow moving changes in average growth rates, the variance of η_t^μ will instead be assumed to be constant.

The defining property of the output gap is its assumed stationarity (at least in means), which implies that long-term forecasts of the output gap converge towards zero.[6] The trend component $\bar{y}_t$ can thus also be expressed in a way that is consistent with the forecast-based trend definition of Beveridge and Nelson (1981), whereby the trend is identified as the limit of a sequence of long-term forecasts for the level of GDP (adjusted for average trend growth) with ever-increasing forecast horizons:

$$\bar{y}_t = \lim_{k\to\infty} E\left(y_{t+k} - k \cdot \mu_t | \Omega_t\right) \qquad \Omega_t = \{\mu_t, \bar{y}_t, \ldots\}. \tag{4}$$

While (4) has been motivated by the stationarity of the output gap, the martingale property of the conditional expectation in (4) is also consistent with the assumed random walk behavior for the

[6]More generally, stationarity implies that long-term forecasts tend towards a constant value, which has been normalized to zero here.

growth-adjusted trend in (2). Importantly, the forecasts in (4) are conditioned on an information set that spans the trend components defined in (2) and (3). The distinction matters here, since this paper is concerned with the revisions in estimates of $\bar{y}_t$ resulting from the use of different information sets which typically do not span the true trend.[7]

Revisions in quasi-real time

Typically, an econometrician observes only a subset of variables, whose history is denoted by Z^t.[8] Estimates of unobserved components at time t (like trend level and output gap) that are conditioned on Z^t will be called "filtered" or "quasi-real-time" estimates, denoted $E(\tilde{y}_t|Z^t) = \tilde{y}_{t|t}$ (in the case of the output gap, and analogously for the trend components).[9] Estimates of $\tilde{y}_t$ that are conditioned also on subsequent observations of the data are called "smoothed" estimates, denoted $\tilde{y}_{t|t+k}$ ($k > 0$). Estimates conditioned on all available data Z^T ($T \geq t$) are commonly called "final estimates" and I will follow this practice — however it should be understood that these estimates are "final" only with respect to the data set available at the time of the study. Since the true output gap $\tilde{y}_t$ can typically never be estimated with near-certainty even when T is much larger than t, these "final" estimates remain subject to revisions as further data becomes available. (Sections 3 and 4 document the robustness of the results in this paper to different choices of T.)

Orphanides and van Norden (2002) documented that the revisions in output gap estimates, obtained from comparing filtered and final estimates are large and this paper corroborates their

[7] While the trend concept used in this paper, $\bar{y}_t$ in (4), is certainly consistent with the trend concept of Beveridge and Nelson (1981), it should be noted that conditioning the trend definition on observable data would be somewhat closer in spirit to the procedures described by Beveridge and Nelson in their original paper. In this case, the Beveridge-Nelson trend is conditioned on Z^t, and thus identical to the econometrician's filtered estimate of $\bar{y}_t$

$$\tau_t \equiv \lim_{k\to\infty} E\left(y_{t+k} - k\cdot\mu_t|Z^t\right) = E(\bar{y}_t|Z^t)$$

However, since the alternative trend measure τ_t is explicitly defined with respect to the econometrician's information set at time t, the issue of revisions triggered by the arrival of subsequent data, which is the central focus of this paper, is moot (Harvey, 1989, Chapter 6).

[8] The history $Z^t \equiv \{Z_t, Z_{t-1}, Z_{t-2}, \ldots\}$ collects all past observations of the vector of variables Z_t. The univariate model described in Section 3 considers the case where Z_t consists only of the level of GDP while in the bivariate model of Section 4 Z_t contains the level of GDP and the unemployment rate.

[9] The expression "real-time estimate" will be introduced later in the context of the data vintages available in real time.

finding with an updated data set applied to one of their univariate models (see Section 3).[10] For the sake of the discussion, suppose estimation uncertainty is measured by variance. The variability in revisions $\tilde{y}_{t|t} - \tilde{y}_{t|T}$ provides then a lower bound for output gap uncertainty in (quasi-)real-time, since the latter is the sum of the variability of revisions and the residual uncertainty around the final estimates:[11]

$$\text{Var}\left(\tilde{y}_t - \tilde{y}_{t|t}\right) = \text{Var}\left(\tilde{y}_{t|T} - \tilde{y}_{t|t}\right) + \text{Var}\left(\tilde{y}_t - \tilde{y}_{t|T}\right) \tag{5}$$

This variance decomposition helps also to preview the central results of this paper. Similar to Creal et al. (2010), Section 4 below finds that estimates of uncertainty about the output gap tend to be smaller when derived from models with stochastic volatility, suggesting a lower value on the left-hand side of (5). Moreover, as reported below, output gap revisions derived from stochastic volatility models are typically smaller (in absolute value) than those obtained from a constant-parameter model, suggesting a lower value for the first term on the right-hand side of (5).

Data vintages and real-time revisions

The above discussion focused on the distinction between "final" and "quasi-real-time" estimates, $y_{t|t}$ and $y_{t|T}$, that arises from the availability of additional data as the conditioning set grows from Z^t to Z^T. In practice, forecasters and decisionmakers are also faced with the availability of new readings for past values of data revising previously released numbers. Over the last twenty years, a growing literature has well documented the consequences of revisions to real-time data for forecasting and inference in macroeconomic time series as surveyed by Croushore and Stark (2001) and Croushore (2006). In the case of real GDP, data revisions often stem from the ongoing efforts of statistical agencies (like the Bureau of Economic Analysis, "BEA", in the U.S.) in collecting underlying source data. But revisions are also caused by definitional changes that arise

[10] Orphanides and van Norden (2002) also found that revisions tend to be even larger in absolute value when they are derived from comparing final estimates against "real-time" estimates that are conditioned only on the earliest available data vintages as will be discussed shortly below.

[11] Equation (5) is an application of the law of total variance and thus holds as long as expectations satisfy the law of iterated expectations, such that future revisions are unforecastable given current information: $E(y_{t|T} - y_{t|t}|Z^t) = 0$.

from the BEA's regular revisions of the national account system.[12] In the case of the unemployent rate, revisions occur only in publications of the seasonally adjusted series and are solely due to the relatively minor revisions in the seasonal adjustments as the underlying source data from the Bureau of Labor Statistics' (BLS) household survey does not get revised.[13]

For this study, I use quarterly records of vintage data for real GDP/GNP and the unemployment rate that were compiled for the Real-Time Data Set for Macroeconomists (RTDSM) published by the Federal Reserve Bank of Philadelphia.[14] The quarterly vintages provide time series for real GDP/GNP and the unemployment rate reflecting available data at the middle of each quarter since the mid-1960s. This paper uses the vintages from 1970:Q1 through 2014:Q1. From each vintage, data is used from 1960:Q1 onwards. Monthly data for unemployment rate is converted into quarterly averages, matching the quarterly data for GDP/GNP. All model estimates are thus conditioned on samples of quarterly data, with each sample starting in 1960:Q1, but differing in the variables included, the length of the sample, and the vintage from which the data has been taken.

To capture these real-time issue, it is useful to augment the previously used notation by indexing observations not only by the calendar date t to which a value pertains, but also by the date of the data vintage v, when a particular value was recorded. In sum, histories of observations recorded at a vintage date v for a calendar date t are henceforth denoted $Z^{t,v} = \{Z_t^v, Z_{t-1}^v \ldots, Z_0^v\}$. Typically,

[12]Apart from the BEA's annual revisions, that occur every July, wider-ranging changes can also be introduced as part of "benchmark" or "comprehensive" revisions, which are typically released every five years. For example, with the benchmark revision of December 1991, BEA began reporting GDP instead of GNP as the headline measure and during the last comprehensive revision of the national accounts in 2013, the definition of GDP was broadened to include also certain types of R&D expenditures, which increased the level of GDP by about 3%. In each case, the BEA also revised its historical estimates of GDP according to its latest definitions and procedures.

[13]As stated in the documentation of the Federal Reserve Bank of Philadelphia on their vintages of unemployment rate data (`http://www.philadelphiafed.org/research-and-data/real-time-center/real-time-data/data-files/RUC/specific_documentation_RUC.pdf`):

> "On the basis of our experience in collecting these data, revisions to the rate of unemployment are small and confined to changes in seasonal factors. In general, the revisions are on the order of 0.1 to 0.2 percentage points and affect only a few of the months in each of the past few years."

[14]The data set and further documentation are available at `http://www.philadelphiafed.org/research-and-data/real-time-center/real-time-data/` and discussed further by Croushore and Stark (2001). Data for real GDP/GNP is labeled *ROUTPUT* and the unemployment rate data is labeled *RUC* in the RTDSM. As explained there, *ROUTPUT* provides GNP data through the last quarterly vintage of 1991 and then switches to GDP data, reflecting the changes introduced by the BEA's benchmark revision of December 1991.

each vintage provides data through the end of the previous quarter.[15] In the data set being used here, calendar time $t = 1, \ldots, T$ ranges from 1960:Q1 through 2013Q:4 and the vintage index $v = 1, \ldots, V$ goes from 1970:Q1 to 2014:Q1. Following the terminology of Orphanides and van Norden (2002), three different types of output gap estimates will be compared:

Final estimates are conditioned on the latest available data and are thus defined by $E(\tilde{y}_t|Z^{T,V})$.

Quasi-real-time estimates are conditioned on the final data vintage, V, as well, but limited to observations through t. Formally, they are thus expressed as $E(\tilde{y}_t|Z^{t,V})$. The distinction between "final" and "quasi-real-time" estimates corresponds also to the difference between "smoothed" and "filtered" estimates found elsewhere in the (non-linear) filtering literature, that abstracts from real-time data issues.[16]

Real-time estimates of the output gap at t are derived from the first vintage for which observations through t are available. Denoting this vintage by $v^*(t)$, real-time estimates are defined by $E(\tilde{y}_t|Z^{t,v^*(t)})$.[17]

Real-time revisions are then computed as the difference between real-time and final estimates, and analogously for quasi-real-time revisions. Quasi-real-time revisions result from the availability of additional observations in a given vintage, and thus for given definitions and methods for measuring GDP/GNP; thus they merely reflect properties of the statistical filtering procedures used to construct "filtered" and "smoothed" estimates. In contrast, real-time revisions arise not only from such statistical issues, but also from revisions to the underlying vintage data, including changes in data definitions.[18] Arguably, the size of real-time revisions is the most relevant measure for

[15]In the data set used here, the only exception from this rules is the 1996:Q1 vintage, which provides only GDP data through 1995:Q3 (thus merely updating the values provided by 1995:Q4 vintage, without adding an observation for the value of GDP in 1996:Q4).

[16]In the case of a purely linear model, like one of the constant-parameter versions estimated here, final and quasi-real-time estimates do not correspond to "filtered" and "smoothed" estimates of a Kalman filter, since they also reflect re-estimated of parameters.

[17]Since, the first vintage observation for a given date t becomes typically available with a lag of one quarter, we have $v^*(t) \approx t + 1$. Note further that $V = v^*(T)$.

[18]For example consider the difference between a real-time estimate of the output gap derived from a vintage prior 1991, and thus based on GNP data, with the final estimate derived from the 2014:Q1 vintage and thus based on the

real-world forecasters and decisionmakers. Analyzing both real-time and quasi-real-time revisions helps however to better understand the source of revisions. As in Orphanides and van Norden (2002) I find that the bulk of revisions — measured by root-mean-squared error or "RMSE" — shows up already in the quasi-real-time revisions, suggesting that revisions are largely caused by the so-called endpoint problem in the statistical filtering procedures and this result also carries over to revisions computed from stochastic volatility models. However, as will be seen further below as well, for some episodes there are notable revisions in real time, but not in quasi-real time.

3. REAL-TIME REVISIONS IN A UNIVARIATE MODEL

This section considers output gap estimates derived from a univariate model, which is a simple application of the trend-cycle decomposition described by equations (1)– (3) above. The constant-parameter version of this model has been extensively used elsewhere in the literature, starting with Harvey (1985), Clark (1987), and Watson (1986). Amongst the wide range of models considered by Orphanides and van Norden (2002), this "Harvey-Clark" model also generated smaller output gap revisions than others.[19] The model is applied here to an updated set of vintage data, extending roughly ten years beyond what was available at the time of Orphanides and van Norden (2002). While they used maximum-likelihood techniques, model estimates are generated here from Bayesian MCMC simulations, whose posterior distribution reflects the joint uncertainty about states (like trend and gap) and model parameters. By and large the results confirm the original findings of Orphanides and van Norden (2002); in particular the large absolute size of output gap revisions relative to the average absolute size of the estimated gaps, which is also reflected in a large uncertainty around these estimates. But then, this section also considers a version of the Harvey-Clark model, where shocks to trend and cycle have stochastic volatility (SV), which establishes the basic result of this paper: Output gap revisions in the SV model are smaller (in absolute

present definitions of GDP. While this is certainly an interesting comparison, to be meaningful it involves however also the implicit assumption that differences between past readings of GNP and current readings of historical GDP mostly pertain to differences in their trend components.

[19]This model is referred to as the Harvey-Clark model, since Harvey (1985) and Clark (1987) let the drift rate be time-varying as in (3) whereas Watson (1986) assumes a constant drift rate in the trend, $\mu_t = \mu$.

size) than in the constant-parameter version of the model and, at times of low volatility, output gap uncertainty in the SV model drops.

As a simple example of the trend-cycle decomposition laid out in equations (1)–(3) above, consider the case where the output gap follows an AR(2) process,

$$a(L)\tilde{y}_t = \tilde{\varepsilon}^y_t \qquad a(L) = 1 + a_1 L + a_2 L^2 \qquad \tilde{\varepsilon}^y_t \sim N(0, \tilde{\sigma}_{y,t}) \tag{6}$$

and where the conditioning set consists solely of output data, $Z_t = y_t$. The assumed stationarity of $\tilde{y}_t$ requires that all roots of $a(L)$ lie outside the unit circle, which will be enforced in the model's estimation by rejection sampling.[20] As in most of the other papers cited above, shocks to the trend level ($\bar{\varepsilon}^y_t$) and the cycle ($\tilde{\varepsilon}^y_t$) are assumed to be mutually uncorrelated in the univariate model. As surveyed in Section 6, this assumption has been found to be crucial for the size, persistence and overall pattern of output gaps estimates from such univariate models. The assumption will be relaxed for the bivariate model studied in Section 4. Keeping the orthogonality assumption for the original Harvey-Clark model makes my results directly comparable to those reported by Orphanides and van Norden (2002).

Constant Parameters

The top left panel of Figure 1 displays final estimates as well as real-time and quasi real-time estimates of the output gap derived from the univariate Harvey-Clark model with constant parameters. By and large real-time and quasi real-time estimates are very similar and the discussion will focus on the comparison between real-time and final estimates.[21] Output gap revisions, shown in the bottom left panel of Figure 1, are large; ranging from almost -3 percent to about $5\,^1\!/_2$ percent with a root mean squared error (RMSE) a little higher than $2\,^3\!/_4$ percent.[22]

[Figure 1 about here.]

[20]The same will be applied also in the bivariate model of Section 4.

[21]An important difference between real-time and quasi-real-time estimates of the output gap shown in Figure 1 pertains however to the mid-1970s where real-time estimates suggest a much more deeply negative output gap than quasi-real-time or even final estimates. This revision in the output gap is mostly prompted by revisions in GDP data released during the second half of the 1970s (and not obviously caused by a benchmark revision).

[22]All values are also tabulated in Table 1 below.

Strikingly, real-time estimates of the output gap are rarely positive, averaging at about $-1\,^1\!/_4$ percent, while the final estimates are more evenly balanced around zero (the assumed population mean of the output gap).[23] In consequence, output gap revisions are on average positive, which also accounts for about two-thirds of their RMSE. A similar pattern can also be seen in the results Orphanides and van Norden (2002). In their study, the downward bias of real-time estimates is even more pronounced for models that do not account for drift in the average growth rate of GDP, as it is done in the Harvey-Clark model, which also points to the underlying cause for this bias: As widely discussed for example by Blanchard and Fischer (1989, Chapter 1), Gordon (2012) and Stock and Watson (2012), average growth rates for real GDP have mostly decreased over the last five decades. Models that assume a constant average growth rate will thus over-estimate trend growth over the most recent history (and thus underestimate the output gap). In the context of the Harvey-Clark model, changes in trend growth are captured by μ_t in equation (3) above, which ameliorates the bias. But still, given the full sample evidence — including the last recession and its aftermath — final estimates of $\mu_{t|T}$ are typically below real-time estimates $\mu_{t|t}$, in particular real-time estimates made towards the end of the productivity boom of the late 1990s.

Orphanides and van Norden (2002) not only found that real-time estimates of the output gap are unreliable, because of the subsequent revisions, but they also documented substantial uncertainty surrounding output gap estimates. This result is also confirmed by the posterior distributions generated by the MCMC estimation used in this paper. These posteriors account jointly for the uncertainty about unknown model parameters — like the persistence of the output gap in (6) or the different shock variances — but also for the uncertainty about the latent states of the model ($\bar{y}_t$, $\tilde{y}_t$, μ_t), causing non-normality and asymmetry in the posterior distribution.[24] As shown in Figure 2, the 90% credible sets around univariate estimates of the output gap are wide, straddling four to

[23]The finding of mostly negative real-time estimates pertains, of course, only to the endpoints $\tilde{y}_{t|t}$ of each vintage estimate. Estimates of $\sum_j y_{t-j|t}$ within each vintage are then again closer to zero.

[24]Absent parameter uncertainty, the posterior distribution for estimates of $\tilde{y}_t$ would be normal and thus symmetric. In fact, the constant-parameter version of the model also satisfies the conditions for steady state convergence of the Kalman filter. In this case, the posterior distribution for estimates of $\tilde{y}_{t|T}$ for t sufficiently away from the sample's endpoints $t = 0$ and $t = T$ have constant variance as well. The variations in the width of the credible sets seen in the right-most panels of Figure 2 stems from the non-normality of the posterior arising from parameter uncertainty.

eight percentage points (and more).[25] Measured in real time, the 90% credible sets cover a value of zero for the output gap most of the time from the 1970s until 2008. In fact, only since 2009 have the 90% credible sets for real-time estimats of the output gap been persistently different from zero.

[Figure 2 about here.]

Clearly, as argued by Orphanides and van Norden (2002), real-time estimates of the output gap generated by the univariate Harvey-Clark model should not be appealing for decision makers, who must act in real time, because of the large revisions in these estimates and the considerable uncertainty around them. However, these estimates condition on a single economic variable, real GDP, which is subject to considerable revisions. Furthermore, the model ignores persuasive evidence suggesting significant changes in the average size of economic shocks over time for the U.S. (and elsewhere in the world). The remainder of this section describes a version of the Harvey-Clark model with stochastic volatility. Foreshadowing the main result of this paper, output gap estimates generated by the Harvey-Clark model with stochastic volatility are smaller than in the constant-parameter version of the model.

Stochastic Volatility

For the stochastic-volatility ("SV") version of the Harvey-Clark model, I assume that the volatilities of shocks to the output gap ($\tilde{\varepsilon}_t^y$) and the trend level ($\bar{\varepsilon}_t^y$) follow independent random walks in their logs:

$$\tilde{\varepsilon}_t^y \sim N(0, \tilde{\sigma}_{y,t}^2) \qquad \log(\tilde{\sigma}_{y,t}^2) = \log(\tilde{\sigma}_{y,t-1}^2) + z_t \qquad z_t \sim N(0, \phi_{\tilde{\sigma}}) \tag{7}$$

and analogously for shocks to the trend level $\bar{\varepsilon}_t \sim N(0, \bar{\sigma}_t^2)$. This random walk specification for stochastic volatility is commonly used in time-series macro-econometrics, see for example Stock and Watson (2007), Justiniano and Primiceri (2008), Cogley et al. (2010) to name but a few. The random walk specification (in logs) implies unbounded variances in population, which is of course

[25] Throughout this paper, all credible sets for estimates of $\tilde{y}_t$ (or any other variable) are pointwise. Pointwise credible set do not account for the joint uncertainty surrounding a particular trajectory of $\tilde{y}_t$. Sims and Zha (1999) discuss the inherent multi-dimensionality of such objects and the resulting difficulties in their representation.

unappealing. But as demonstrated by the aforementioned papers, this does not cause unbounded estimates, when conditioning on actual data. Furthermore, within the class of linear processes, the random walk can also more flexibly handle structural breaks than, say, a stationary AR(1) specification, since the latter imposes mean reversion in volatility to some fixed long-run value. In a recently published study, Clark and Ravazzolo (2014) compare different SV specifications and find the random walk to dominate alternative specifications. (Having said that, Section 5 will discuss how density forecasts at longer horizons deteriorate due to the lack of mean reversion in the random walk specification.) As in Stock and Watson (2007) and Stella and Stock (2013), the variance of shocks to the SV processes, $\phi_{\bar{\sigma}}$ and $\phi_{\tilde{\sigma}}$ are fixed at 0.2, a value that generates relatively smooth changes in stochastic volatility. The choice to fix these parameter values also reflects some difficulty in reliably estimating these parameters across all vintages.

The top left and bottom left panels of Figure 1 report final, quasi- and real-time estimates of output gap estimates generated by the univariate SV model and the corresponding revisions in these estimates. As in the constant-parameter case, real-time and quasi-real-time estimates are fairly similar and the discussion will concentrate on real-time and final estimates. Revisions from real-time to final estimates are still sizable, but noticeably smaller than in the constant-parameter case; with root mean squared errors reduced by about a third. The bulk of this reduction reflects a considerable decline in the average size of the revisions towards zero, while the volatility of the revisions has only modestly decreased.

Stochastic volatility vs. constant parameters

As discussed before, "final" estimates are never really final, but rather the "last available" estimates obtained from the most recent data. It is thus instructive to see how the RMSE in revisions generated from different models may have changed over time. For example, would a study like Orphanides and van Norden (2002), conducted 10 years ago also have found a reduction in revisions' RMSEs when comparing the constant-parameter version of the Harvey-Clark model against the variant with stochastic volatility?

[Figure 3 about here.]

Figure 3 displays RMSE of revisions in real time and quasi-real time for both versions of the univariate model, when different vintages are treated as “final”. For quasi-real time revisions, these RMSE are computed as

$$RMSE(\tau) = \sqrt{\frac{1}{\tau - \tau_0 + 1} \sum_{t=\tau_0}^{\tau} \left(\tilde{y}_{t|\tau} - \tilde{y}_{t|t}\right)^2} \quad (8)$$

where the last observation of each vintage is denoted by τ and $\tau_0 = $ 1960:Q1 denotes the last observation of the first vintage. For real-time revisions, the same applies using $E\left(\tilde{y}_t|Z^{t,v^*(\tau)}\right)$ in place of $\tilde{y}_{t|\tau}$ and $E\left(\tilde{y}_t|Z^{t,v^*(t)}\right)$ in place of $\tilde{y}_{t|t}$. $RMSE(\tau)$ is shown for values of τ ranging from 1960:Q1 to 2013:Q4.[26] For quasi-real time revisions, the RMSE generated by the univariate SV model are consistently below those from the constant-parameter version of the model, typically by almost half a percentage point since the mid-1980s and even a little more before then. Considering real-time revisions, both versions of the univariate model yield fairly similar RMSE for most of the “final” vintages considered, in particular over the period from the mid 1980s until 2008. Output gap revisions from the SV model however display considerably smaller RMSE — by about one percentage point — in the wake of the big recessions in the mid 1970s, early 1980s and since 2009. Clearly, the benefits of using a model with stochastic volatility are better borne out at times of changing volatilities and would have been much less visible about ten years ago, when Orphanides and van Norden (2002) conducted their study in a sample strongly influenced by the Great Moderation period.

Equation (5) above decomposes real-time uncertainty about the output gap (measured by variance) into the variability of revisions and the residual variance of the true output gap around the final estimates. If variability is a reliable proxy for the uncertainty embedded in the posterior distributions estimated here, it stands to reason whether the decline of variability in output gap revisions is also mirrored in estimates of uncertainty around final and real-time uncertainty which are documented in Figure 2. The right-most panels of the figure compare the width of the 90% credible sets

[26]Notice that by construction we have $RMSE(\tau_0) = 0$. The values plotted for $\tau = T =$ 2013:Q4 correspond to the full-sample RMSE quoted above, that are also listed in Table 1.

of estimates generated by the univariate SV model against its counterpart with constant parameters. Creal et al. (2010) estimate a multivariate business cycle model for the U.S. using data from the 1950s to 2007 and find that the inclusion of stochastic volatility in the specification of shocks to the model's cyclical component — akin to the output gap concept used here — reduces uncertainty in their estimates of the business cycle. In the real-time parlance introduced in Section 2, the estimates reported by Creal et al. correspond to final estimates per 2007:Q4.[27] Figure 2 confirms, but also refines their finding.

Measured by the width of the 90% credible sets, output gap uncertainty generated by the SV model — be it in real-time or for final estimates per 2014:Q1 — is lower than in the constant-parameter case for estimates of the output gap during most of the Great Moderation period (1985 – 2007).[28] Overall, uncertainty from the SV model displays considerable variation over time. These variations are more pronounced when uncertainty is measured in real time, but also in the case of the final estimates, the width of the 90% credible sets varies from almost 8 percent to about 5 percent during the mid 1990s and early 2000s. Not surprisingly, uncertainty about the state of the economy in the SV model is smallest, when the volatility of economic shocks are low.[29] While uncertainty in the SV model is substantially lower than in the constant-parameter case over the Great Moderation period, it is about as high or even slightly higher during the high-volatility times of the 1970s as well as since 2009.

[Figure 4 about here.]

As shown in Figure 4, estimates of stochastic volatility in shocks to trend and cycle have fallen during the early 1980s. Trend volatility has since 2000 ramped up only slightly, and spiked to its highest reading at the height of the last recession in 2009, while gap volatility has regularly increased with every NBER recession and then fallen again. This characterization applies both to the final estimates — shown in the upper panels of the figure — and real-time estimates shown in

[27] This discussion refers in particular to the results shown in Figure 8 of Creal et al. (2010).

[28] A similar pattern also emerges when measuring the width of credible sets with coverage rates other than 90%, like the inter-quartile range for example.

[29] This pattern has also been found by Watson (2007), who computed standard errors for one-sided band-pass estimates of the output gap for different subsamples.

the middle panels. Final estimates of the volatility in shocks to trend and cycle from the constant-parameter model are shown in the figure as well; as to be expected, these estimates roughly average over the values of their time-varying counterparts in the SV model. The somewhat improved behavior of output gap revisions in the SV model stems from the time-varying volatilities shown in Figure 4, which give rise to time-varying Kalman gains when filtering the latent states of the model. Volatility estimates are also subject to revisions — see the middle panels of Figure 4 — but to the extent that these revisions preserve the time-variation in Kalman gains estimated from previous vintages, they prompt only small revisions in estimates of the output gap. The constant-parameter model is, of course, also subject to revisions in its volatility estimates as the model gets re-estimated with every new vintage; as shown in the lower panels of Figure 4. But, in the constant-parameter model, the Kalman gain converges quickly to a steady-state value, and changes in the volatility estimates shown here, prompt the constant-parameter model to reapply this revised steady state gain over the entire history of data, thus causing larger revisions in estimates of the output gap.

4. A BIVARIATE MODEL IN REAL GDP AND THE UNEMPLOYMENT RATE

This section describes a bivariate extension of the univariate Harvey-Clark model analyzed above. The extension beyond the univariate case is motivated by at least two considerations: First, as documented for example by Fleischman and Roberts (2011) the use of multivariate data, in particular in conjunction with common factor restrictions, can considerably improve the reliability of output gap estimates. Second, while stochastic volatility (or time-varying parameters in general) might be appealing in order to better capture salient features of the data, it also increases the number of latent variables in the model for a given set of observations. The resulting risk of overfitting output gap estimates in sample should potentially be alleviated, when common factor restrictions can be exploited in multivariate data.

The bivariate model augments the model with a trend-cycle decomposition for the unemployment rate and links the resulting gaps in unemployment and output via an Okun's law relationship.

As before, a constant-parameter version of the model is compared to a version with stochastic volatility. In fact, given the multivariate nature of the model, different versions of stochastic volatility as well as a version that allows for time-variation in the model's Okun's law coefficients will be discussed. After some further description of the model,

The unemployment trend identified by the model is conceptually identical to the "NAIRU" estimates of Staiger et al. (1997), Gordon (1997), Laubach (2001), Reifschneider et al. (2013), Stella and Stock (2013) and Watson (2014), to name but a few. However it should be stressed, that — at least in my model, which purely conditions on real variables — there is no presumption that this long-term forecast for unemployment is necessarily related to the concept of a "non-accelerating inflation rate of unemployment."

To the extent that the unemployment rate is informative about the output gap — and a wide body of research, of course, suggests it is — an immediate benefit of incorporating unemployment data in the model should be the largely negligible revisions in unemployment data witnessed across different vintages.[30]

Another candidate variable with potential information about the output gap is inflation, as used for example by Gerlach and Smets (1997) and Kuttner (1992), whose models were also included by Orphanides and van Norden (2002), or Laubach and Williams (2003) and Fleischman and Roberts (2011). However, as recently surveyed by Stock and Watson (2009), empirical linkages between inflation and slack seem to have evolved, and might be less clear than once thought, for example due to non-linearities in the dependence of inflation on economic slack as reported by Stock and Watson (2010) or Stella and Stock (2013). For the purpose of this paper, I have thus chosen not to condition my output gap estimates on inflation data, and provide instead estimates that reflect only information contained in real data.

[30] As remarked before, these revisions seem to result solely from revisions in seasonal adjustments, while the underlying source data from the BLS's household survey is not subject to revisions.

Constant-parameter model

The bivariate model retains the trend-cycle decomposition for output from the Harvey-Clark model, specifically equations (1)–(3) and the AR(2) specification for the output gap in (6) and augments the model with an observation equation for the unemployment rate (u_t). Similar to output, the unemployment rate is split into a stochastic trend ($\bar{u}_t$), with no drift, and a stationary gap component ($\tilde{u}_t$).

$$u_t = \bar{u}_t + \tilde{u}_t \qquad \tilde{u}_t \sim I(0) \tag{9}$$

$$\bar{u}_t = \bar{u}_{t-1} + \bar{\varepsilon}^u_t \tag{10}$$

The dynamics of the unemployment gap are linked to output gap via an Okun's law relationship, which represents the unemployment gap as a MA(2) of the output gap plus a serially uncorrelated error. Effectively, this is a dynamic factor model where the dynamics in both gaps are described by a single factor ($\tilde{y}_t$):

$$\tilde{u}_t = b_0\ \tilde{y}_t + b_1\ \tilde{y}_{t-1} + e^u_t \qquad e^u_t \sim N(0, \phi_u) \tag{11}$$

and $\tilde{y}_t$ is still described by the AR(2) in (6) above.[31]

As highlighted before, estimation of the univariate model in Section 3 assumed that shocks to trend and cycle are uncorrelated. In the bivariate model the assumption is now relaxed and the variance-covariance matrix of the shocks to trend levels and common cycle is unrestricted, such

[31] Throughout this paper, maintain the assumption of an AR(2) for the output gap and a MA(2) dependence of the unemployment gap on the output gap. In particular the AR(2) specification of the output gap has been widely adopted dating back at least toWatson (1986). I have also experimented with different lag choices, but without much substantial effect, except for signs of overfitting when using lags larger than two.

that $\mathbf{\Sigma}$ in

$$\begin{bmatrix} \bar{\varepsilon}_t^u \\ \bar{\varepsilon}_t^y \\ \tilde{\varepsilon}_t^y \end{bmatrix} \sim N\left(\mathbf{0}, \mathbf{\Sigma}\right) \tag{12}$$

can be any positive definite matrix. As before, shocks to trend growth in real GDP, η_t^μ in (3), are assumed to be orthogonal to all other shocks in the model. Likewise, the Okun's Law error e_t^u in (11) is assumed to be independent from all other shocks.

Stochastic Volatility in the bivariate model

Two versions of embedding stochastic volatility in the bivariate model are considered here. Both versions feature time variation in the variance-covariance matrix of shocks to the common cycle and the two trend levels as well as the variance of Okun's Law errors:

$$\begin{bmatrix} \bar{\varepsilon}_t^u \\ \bar{\varepsilon}_t^y \\ \tilde{\varepsilon}_t^y \end{bmatrix} \sim N\left(\mathbf{0}, \mathbf{\Sigma}_t\right) \qquad\qquad e_t^u \sim N(0, \phi_{u,t}) \tag{13}$$

Both versions differ in the extent to which the coefficients of $\mathbf{\Sigma}_t$ co-move over time. The less restricted version, abbreviated as "SV", features three independent sources of stochastic volatility, one for each Choleski factor of $\mathbf{\Sigma}_t$:

$$\mathbf{\Sigma}_t = \boldsymbol{Q}\boldsymbol{S}_t\boldsymbol{Q}' = \begin{bmatrix} 1 & 0 & 0 \\ q_{21} & 1 & 0 \\ q_{31} & q_{32} & 1 \end{bmatrix} \begin{bmatrix} \sigma_{1,t}^2 & 0 & 0 \\ 0 & \sigma_{2,t}^2 & 0 \\ 0 & 0 & \sigma_{3,t}^2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ q_{21} & 1 & 0 \\ q_{31} & q_{32} & 1 \end{bmatrix}' \tag{14}$$

In a more restricted version of the model, which will be referred to as the "scale SV" model, a single time-varying parameter accounts for all variations in $\mathbf{\Sigma}$ over time.

$$\mathbf{\Sigma_t} = \sigma_t^2 \cdot \mathbf{\Omega} = \sigma_t^2 \cdot \begin{bmatrix} 1 & \omega_{12} & \omega_{13} \\ \omega_{12} & \omega_{22} & \omega_{23} \\ \omega_{13} & \omega_{23} & \omega_{33} \end{bmatrix} \tag{15}$$

Carriero et al. (2012) propose to restrict VARs with stochastic volatility shocks as in the scale SV model and find considerable improvements in out-of-sample forecasts. In both cases, the logs of $\phi_{u,t}$ as well as $\sigma_{i,t}^2$ $\forall i \in 1, 2, 3$ and σ_t^2, respectively, follow independent random walks with fixed innovation variances as in the univariate SV model, see (7) above.[32]

Stochastic volatility in the Okun's law error (e_t^u) serves to capture variations in the goodness of fit in the in (11) assumed dependence of the unemployment gap on the output gap.[33] Time-variation in real-time estimates of the Okun's law coefficients b_0 and b_1 — including an alternative version of the model with explicit time-variation in these coefficients — will be discussed in a future update of this draft.

[Figure 5 about here.]

Stochastic volatility vs. constant parameters

For all three versions of the bivariate model, "constant parameters", "Scale SV", and "SV," Figure 5 compares final and (quasi-)real-time estimates of the output gap. In general, output gap uncertainty in the bivariate model is noticeably smaller than in the univariate model, as is evident from the tightness of the 90% credible sets shown in Figure 5 compared to the unvariate results seen in Figure 2. Similar to the univariate model, uncertainty is more variable in the stochastic-volatility versions of the model and the credible sets are tighter for output gaps in the Great Moderation

[32] As before the volatility of shocks to the log-variances is fixed at 0.2, except for the case of shocks to the log-variance of Okun's law errors which has been set to 0.1.

[33] Results not shown here indicate that models with stochastic volatility in shocks to trends and cycle, but no time-varying parameters in Okun's law, typically perform worse — in terms of output gap revisions and out-of-sample forecasts than what is presented here.

period, when volatility is estimated to be lower. (The underlying volatility patterns are similar to those shown in Figure 4 for the univariate model.)

[Table 1 about here.]

Summary statistics for output gap revisions in real time and quasi-real time are reported in Table 1. Consistent with evidence from Fleischman and Roberts (2011), real-time and quasi-real-time revisions track final estimates much more closely in the bivariate model, when compared against the univariate estimates (be they computed with constant parameters or stochastic volatility). Initial estimates are much stronger correlated with final estimates than in the univariate case — the correlation coefficients exceed 0.95 instead of about 0.7. Revision RMSE of the bivariate estimates are in the order of one percent, instead of two percent and more, which corresponds to less than half of the volatility of the final output gap estimates.

[Figure 6 about here.]

As in the univariate case, the inclusion of stochastic volatility decreases revision RMSE. In particular for the "Scale SV" model, this is achieved by shrinking the average revision towards zero, such that the RMSE correspond to the revisions' volatility. Figure 6 displays the RMSE of revisions computed for different "final" vintages. Similar to the univariate case, the improvements reported in Table 1 are in part boosted by revisions accrued since the deep recession that ended in 2009, while over much of the Great Moderation period the RMSE were sometimes at par (and for the quasi-real time revisions at times even higher for the SV models). Naturally, the benefits of the (Scale) SV models are stronger in the wake of the larger changes in macroeconomic volatility seen in the early 1980s and after 2007. Most strikingly, the highly parsimonious Scale-SV model, which adds only two time-varying parameters to the bivariate model, displays consistent and sizable improvements over the entire sample of real-time vintages.

5. FORECAST EVALUATION

The output gap concept employed throughout this paper relies on a statistical decomposition of real GDP (and unemployment) into a permanent and a transitory component. As discussed in Section 2, trend estimates are equivalent to long-term forecasts. While forecasts at the infinite horizon can never be compared against data, this can be done for shorter horizons. This section evaluates out-of-sample forecasts, generated in real time and quasi-real time, respectively, from each of the models analyzed so far. Forecasts are computed for three variables: quarterly output growth $\Delta y_{t+h|t}$, cumulative output growth $(y_{t+h|t} - y_t)/h$ and — in the case of the bivariate models — the unemployment rate $u_{t+h|t}$.[34] For each variable, forecasts are made at different horizons ranging from one quarter to four years based on each model's predictive density obtained from every vintage since 1970:Q1.

The predictive densities are evaluated with two metrics: First, the RMSE computed from the difference between the mean of each predictive density and subsequent realizations of the forecast variable. Second, the log-predictive density score as in Geweke and Amisano (2010). The significance of differences between these metrics is assessed with the Diebold and Mariano (1995) test.[35] Similar forecast metrics were for example also considered by Clark and Ravazzolo (2014) who compared the forecasting performance of autoregressive time-series models with different specification for stochastic volatility and found broadly similar results.

Given the conditionally linear structure of each model, first and second moments of the predictive densities can be computed analytically from the MCMC draws of each model as described by Cogley et al. (2013). As in Adolfson et al. (2007), Christoffel et al. (2010) and Clark and Ravazzolo (2014) the contribution of a forecast made for x_{t+h} at t to the log-predictive score is then

[34] All growth rates are expressed in annualized terms.

[35] Since the models with time-varying parameters are nested by their constant-parameter counterparts, p-values are computed for one-sided tests. Reflecting the over-lapping forecast horizons, denoted by h, test statistics are computed using the standard errors of Newey and West (1987) with $h + 1$.

computed via a normal approximation from the first two moments of the predictive density

$$s_t(x^O_{t+h}) = -\frac{1}{2}\left[\log\left(2\cdot\pi\right) + \log\left\{\text{Var}\left(x_{t+h}|Z^{t,v^*(t)}\right)\right\} + \frac{\left(x_{t+h} - E(x_{t+h}|Z^{t,v^*(t)})\right)^2}{\text{Var}\left(x_{t+h}|Z^{t,v^*(t)}\right)}\right] \qquad (16)$$

where x^O_{t+h} denotes the observed realization. Higher scores indicate better performance at density forecasting; for $h = 1$, the sum of scores also corresponds to a log-likelihood. Predictive scores are averaged across t for two samples, the full sample of vintage forecasts from 1970:Q1 to 2014:Q1 and for a Great Moderation sample (1985:Q1-2007:Q4). RMSE are reported for the two samples as well. Both in terms of RMSE and predictive scores, results are broadly similar across both samples.

Realizations for real-time forecasts of GDP are measured by second-release data provided by the RTDSM of the Federal Reserve Bank of Philadelphia described in Section 2. This is a conventional choice for real-time forecast analysis, which avoids undue influence of subsequent definitional changes in GDP on the forecast evaluation, while ensuring some reduction in uncertainty about the "final" data. Realizations for unemployment forecasts as well as quasi-real time forecasts of GDP are evaluated against final data.[36]

[Table 2 about here.]

[Table 3 about here.]

By and large, forecasts for output growth from models with and without stochastic volatility are surprisingly similar, as can be seen from Table 2. If anything, near-term forecasts of the SV models display little higher density scores, in particular over the Great Moderation period. And while some of the differences are marginally significant, they seem economically fairly small, at the order of a percent in RMSE. Also, for unemployment forecasts, the quality of point forecasts are largely similar, as shown in Table 3, and the predictive scores of the SV models are slightly

[36]Quasi-real time scores are computed by replacing $Z^{t,v^*(t)}$ in (16) by $Z^{t,V}$.

lower than in the constant-parameter benchmark.[37] Considering forecasts made from quasi-real time vintages, their performance is even more similar as shown on Appendix A.

[Figure 7 about here.]

The broad similarity in forecast performance can also be seen in Figure 7. Comparing one-quarter ahead forecasts for output growth generated from the bivariate model with constant parameters and its "SV" counterpart, the figure shows that forecasts error are virtually identical over the entire forecast sample from 1970:Q1 – 2014:Q1. The bottom panel of the figure also displays variations in forecast uncertainty, measured by the variance of each predictive density. The sole source of variation in the constant parameter model are re-estimated parameters, whereas variation in forecast uncertainty of the SV model reflects both re-estimation of parameters, as well as the model's endpoint estimates of the different stochastic volatilities, generated by each vintage, as well as their anticipated further evolution out of sample.

Given the sizable reduction in output gap revisions achieved by the stochastic volatility models, it might seem striking that their forecast ability seems to be only at par with their constant-parameter cousins. However, it should be noted that output gap revisions are both a function of quality in real-time estimates which reflect the same endpoint estimates underlying the out-of-sample forecasts analyzed here and in-sample estimates of the output gap derived from future vintage estimates. Taken together, the forecast evidence presented in this section, and the results on output gap revisions of Sections 3 and 4, suggest that the stochastic volatility are less prone to revise output gap estimates backwards, while the forecast evidence cannot clearly indicate which real-time estimates of the output gap might be more preferable.

[Figure 8 about here.]

The most striking differences in forecast performance, visible in Tables 2 and 3, is the deterioration in predictive densities of the stochastic volatility models at longer horizons. Closer analysis

[37] Corresponding tables for quasi-real time forecasts, as well as forecasts of cumulative output growth can be found in Appendix A.

suggests that this is due to the random-walk specification for the processes of log-variances, as in (7). At longer horizons this assumption leads to undue extrapolation of transitory increases in realized forecast errors, as can be seen in Figure 8.[38]

6. RELATED LITERATURE

My paper revisits the concerns raised by Orphanides and van Norden (2002) about the reliability of real-time estimates of the output gap derived from statistical models. I extend their work by considering unobserved component models that allow for time-varying parameters, in particular stochastic volatility. As such, this paper is not only related to the broad literature on output gap measurement, but also draws from work on real-time data and simulation-based Bayesian estimation of models with stochastic volatility and time-varying parameters.

There is a vast literature on the measurement and estimation of the output gap. As most recently pointed out by Kiley (2013), different papers employ different definitions of the "output gap," which typically fall in one of the following three categories:

1. Statistical decompositions, that measure deviations from the stochastic trend in GDP.

2. The "production function" approach, which estimates deviations of output from "the level consistent with current technologies and normal utilization of capital and labor" that is for example used by the Congressional Budget Office (2001).

3. More structural approaches that measure deviations of output from a hypothetical benchmark level, that is characterized by the absence of certain frictions (like sticky prices) and which is typically defined in the context of a DSGE model, see for example Woodford (2003), Galì (2003), Smets and Wouters (2007), Edge et al. (2008) and Justiniano et al. (2013).

This paper squarely belongs to the first category, which measures the output gap as the cyclical

[38]A similar pattern can also be discerned from the results of Clark and Ravazzolo (2014) — who otherwise conclude that the random walk specification even dominates other alternatives — and it is even more compounded when forecasting variables in levels — like cumulative output growth over multiple quarters or the unemployment rate — and by extending the sample to include the first recession of the 1970s.

component of real GDP in a statistical trend cycle definition. In this context, the stochastic trend is supposed to capture the long-term behavior of the data, whereas the cyclical component has no permanent effects on the data, see for example Harvey (1989, Chapter 6) or Blanchard and Fischer (1989, Chapter 1).[39]

My paper adopts the typical convention of modeling the trend as a random walk (with time-varying drift).[40] As noted — amongst others — by Harvey (1989) and Oh et al. (2008), there is a tight connection between models that specify a random walk behavior for the trend component and the decomposition of Beveridge and Nelson (1981). The Beveridge-Nelson decomposition explicitly defines the stochastic trend as the long-term forecast of the data (adjusted for the expected future rate of trend growth) and thus follows a martingale process. Since the Beveridge-Nelson trend is defined as a conditional expectation, a subtle but crucial role is played by what is assumed about the underlying conditioning set. When the Beveridge-Nelson trend at time t is defined with respect to the econometrician's information set at t, the distinction between filtering and smoothing becomes moot and there would be no distinction between a real-time and "final" measures of the output gap. While this restriction is particularly binding in the case of the ARIMA models that were the focus of the original work by Beveridge and Nelson (1981), it is less of a concern in the context of the unobserved component models used in this paper, as long as the information set underlying the Beveridge-Nelson trend is understood to be broader than the econometrician's.

However, a more important difference between estimates of ARIMA-based Beveridge-Nelson trends and unobserved component has been pointed out by Watson (1986). At least in the case of univariate models, the former produce highly volatile trend estimates, that explain most of the variations in the data, whereas the latter yield much smoother trend estimates and thus more substantial variations in the corresponding gap estimates. As shown by Morley et al. (2003), these differences are caused by the often-imposed orthogonality restrictions on the shocks to trend and

[39]For an alternative class of trend-cycle decompositions based on properties defined in the frequency domain see for example Harvey and Trimbur (2003) or King and Rebelo (1993) and Baxter and King (1999) or the concept of diffusion trends of Lippi and Reichlin (1994).

[40]Evidence in favor of the existence of a stochastic trend component in U.S. data for real GDP goes back to the work of Nelson and Plosser (1982).

cycle in unobserved component models. The unobserved component models used by Orphanides and van Norden (2002) were for example also subject to these restrictions and for the sake of comparability, I have retained the same restrictions in my estimates of the univariate model discussed above. However, the estimates for my bivariate model — both in the cases of constant and time-varying parameters — are not subject to any such orthogonality restrictions. As shown above, even the constant-parameter estimates of my bivariate model with unrestricted correlations between shocks to trend(s) and cycle imply sizable variations in the output gap.[41]. More generally, a set of other studies found sizable output gap estimates from unobserved component models that do not impose a correlation of zero between shocks to trend and cycle once they introduced deviations from the normal-linear framework into the analysis; like explicit asymmetries in the business cycle (Beaudry and Koop, 1993; Piger et al., 2005; Sinclair, 2009a) or fat tailed shocks to the rate of future trend growth (Perron and Wada, 2009).

Orphanides and van Norden (2002) study output gap revisions estimated from a variety of univariate and bivariate models. The constant-parameter version of the univariate model discussed above is part of this set and goes back to the work of Clark (1987) and Harvey (1985), who incorporate a time-varying rate of trend growth into the model studied by Watson (1986). The bivariate models studied by Orphanides and van Norden (2002) link the output gap to inflation via (reduced form) Phillips Curve specifications (Kuttner, 1992; Gerlach and Smets, 1997). A more recent example in this spirit has for example been analyzed by Stella and Stock (2013). As discussed by these authors, the relationship between inflation and economic slack may itself be fraught by instability, see also Stock and Watson (2009). As a result, I have chosen to condition my multivariate purely on measures of real activity, thus yielding estimates of the output that are not affected by potential concerns about the appropriate specification of the Phillips Curve. More generally, the use of multivariate information for trend cycle decompositions has been widely recognized, including the recent work by Creal et al. (2010), Fleischman and Roberts (2011), Planas et al. (2013)

[41]However, Sinclair (2009b) confirms the original results of Morley et al. (2003) within a bivariate model (with constant parameters) that uses output and unemployment. However, her model differs from mine by specifying separate AR processes for the gaps in output and unemployment, whereas my model imposes a dynamic factor with a common cyclical factor.

or Borio et al. (2014).

Orphanides and van Norden (2002) still had to rely on maximum likelihood methods, which proved particularly cumbersome for the computation of uncertainty measures that incorporate the joint uncertainty about parameters and latent variables in their state space models. Subsequent increases in computing power have, of course, made simulation-based Bayesian methods, like the ones employed in this paper, much more feasible. Recent applications of Bayesian methods to the estimation of trend-cycle decompositions include Planas et al. (2008), who consider the Phillips-curve augmented model of Kuttner (1992), Harvey et al. (2007), who employ the generalized filters of Harvey and Trimbur (2003), and Creal et al. (2010), who incorporate stochastic volatility into the Harvey-Trimbur models.[42] The inclusion of stochastic volatility into the specification of business cycle models is, of course, well-motivated by the apparent decrease in the variability of U.S. macroeconomic aggregates witnessed during the 1980s, as documented for example by Perez-Quiros and McConnell (2000), Blanchard and Simon (2001), Stock and Watson (2003), Chauvet and Potter (2001), Kim and Nelson (1999). With respect to my paper, a particularly noteworthy result from Creal et al. (2010) is their finding of tighter credible sets around the output gap estimates obtained from their model with stochastic volatility, which foreshadows also some of my results. A key contribution of my paper is to document that these tighter confidence intervals also carry over into smaller revisions between real-time and "final" estimates.

7. CONCLUSION

The results of Orphanides and van Norden (2005) caution against the use of real-time estimates of the output gap derived from statistical models with constant parameters, since they are unreliable indicators of "final" estimates, derived from subsequent data releases. This paper confirms their finding for models with constant parameters. But this paper also shows that revisions are much smaller when using a model with stochastic volatility.

[42]Considering business cycle models more broadly, Kim and Nelson (1998) regime switching model is an early example of the use of Bayesian simulation methods in this context.

In line with previous research indicating profound changes in the volatility of macroeconomic data over time, this paper finds substantial changes in the variability of shocks to trend and cycle, which also affect their relative importance. In response to a change in the volatility of incoming data, a model with stochastic volatility can adjust its estimates of current volatility, without necessarily changing its estimates of the historical variability of shocks, thus prompting smaller revisions in past output gap estimates. This allows real-time estimates from models with stochastic volatility to be more reliable indicators of final estimates. Output gap uncertainty derived from models with stochastic volatility is also time-varying. When volatility is estimated to be low, output gap uncertainty is smaller, than in models with constant parameters.

These result obtain already in a simple univariate model, albeit still at somewhat unsatisfactory average levels of revisions and uncertainty in output gap estimates. But the result is also confirmed more persuasively in a bivariate model, using real GDP and the unemployment rate exploiting a common factor structure inspired by Okun's law.

Strikingly, the forecasting performance of the constant-parameter models studied here and their counterparts with stochastic volatility is not all too different when considering out-of-sample forecasts at horizons up to four years ahead. These forecasts reflect the same endpoint estimates of each model's latent state variables as the real-time estimates of the output gap. The measured reductions in output gap revisions thus rather reflect the non-linear revision process of the models with stochastic volatility, rather than an apparently superior quality in their endpoint estimates of the output gap.

APPENDIX

A. ADDITIONAL RESULTS

This appendix provides background tables and figures not included in the main text.

[Figure 9 about here.]

[Figure 10 about here.]

[Table 4 about here.]

[Table 5 about here.]

[Table 6 about here.]

[Table 7 about here.]

B. MCMC ALGORITHM

The model is estimated using Bayesian MCMC methods and all results are based on the simulated posterior distribution for all latent variables and parameters. The estimation relies mainly on Gibbs sampling plus a Metropolis Hastings step for the stochastic volatilities. Convergence is assessed by the scale reduction test of Gelman et al. (2003), comparing the outcomes of six chains starting from disperse initial conditions that were drawn from the respective prior for each parameter and the latent variables. As in Cogley and Sargent (2005) and Cogley et al. (2010), the AR of the cyclical factor is estimated with rejection sampling to ensure stability. The baseline results use four lags, instead of the two-lag specification used for the univariate model by Watson (1986), Harvey (1985), Clark (1987). The qualitative results seem however largely unchanged compared to a specification with two lags. The number of lags in the moving average relationship between the unemployment gap and the cyclical factor has been set identical to the number of lags in the cyclical factor's AR.

REFERENCES

Adolfson, M., J. Linde, and M. Villani (2007). Forecasting Performance of an Open Economy DSGE Model. *Econometric Reviews 26*(2-4), 289–328.

Barnes, M. L., F. Gumbau-Brisa, and G. P. Olivei (2013). Do real-time Okun's law errors predict GDP data revisions? Working Papers 13-3, Federal Reserve Bank of Boston.

Baxter, M. and R. G. King (1999). Measuring business cycles: Approximate band-pass filters for economic time series. *The Review of Economics and Statistics 81*(4), 575–593.

Beaudry, P. and G. Koop (1993, April). Do recessions permanently change output? *Journal of Monetary Economics 31*(2), 149–163.

Bernanke, B. S. (2004, February). The Great Moderation. Remarks by Governor Ben S. Bernanke at the meetings of the Eastern Economic Association, Washington D.C.

Beveridge, S. and C. R. Nelson (1981). A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the "business cycle". *Journal of Monetary Economics 7*(2), 151–174.

Blanchard, O. and J. Simon (2001). The long and large decline in U.S. output volatility. *Brookings Papers on Economic Activity 32*(1), 135–174.

Blanchard, O. J. and S. Fischer (1989). *Lectures on Macroeconomics*. Cambridge, MA: The MIT Press.

Blanchard, O. J. and D. Quah (1989, September). The dynamic effects of aggregate demand and supply disturbances. *The American Economic Review 79*(4), 655–673.

Borio, C., P. Disyatat, and M. Juselius (2014, February). A parsimonious approach to incorporating economic information in measures of potential output. BIS Working Papers 442, Bank for International Settlements.

Carriero, A., T. E. Clark, and M. Marcellino (2012). Common drifting volatility in large bayesian vars. Working Paper 1206, Federal Reserve Bank of Cleveland.

Chauvet, M. and S. Potter (2001, Special I). Recent changes in the us business cycle. *Manchester School 69*(5), 481–508.

Christoffel, K., A. Warne, and G. Coenen (2010, May). Forecasting with DSGE models. Working Paper Series 1185, European Central Bank.

Clark, P. K. (1987, November). The cyclical component of u.s. economic activity. *The Quarterly Journal of Economics 102*(4), 797–814.

Clark, T. E. and F. Ravazzolo (2014). Macroeconomic forecasting performance under alternative specifications of time-varying volatility. *Journal of Applied Econometrics n/a*(n/a), n/a–n/a.

Cogley, T., G. E. Primiceri, and T. J. Sargent (2010, January). Inflation-gap persistence in the U.S. *American Economic Journal: Macroeconomics 2*(1), 43–69.

Cogley, T. and T. J. Sargent (2005, April). Drift and volatilities: Monetary policies and outcomes in the post WWII U.S. *Review of Economic Dynamics 8*(2), 262–302.

Cogley, T., T. J. Sargent, and P. Surico (2013, January). Price-level uncertainty and stability in the UK. *mimeo*.

Congressional Budget Office (2001). CBO's method for estimating potential output: An update.

Creal, D., S. J. Koopman, and E. Zivot (2010, June/July). Extracting a robust us business cycle using a time-varying multivariate model-based bandpass filter. *Journal of Applied Econometrics 25*(4), 695–719.

Croushore, D. (2006). Forecasting with real-time macroeconomic data. In G. Elliott, C. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 1, Chapter 17, pp. 961–982. Elsevier.

Croushore, D. and T. Stark (2001, November). A real-time data set for macroeconomists. *Journal of Econometrics 105*(1), 111–130.

Diebold, F. X. and R. S. Mariano (1995, July). Comparing predictive accuracy. *Journal of Business & Economic Statistics 13*(3), 253–63.

Edge, R. M., M. T. Kiley, and J.-P. Laforte (2008, August). Natural rate measures in an estimated DSGE model of the U.S. economy. *Journal of Economic Dynamics and Control 32*(8), 2512–2535.

Edge, R. M. and J. B. Rudd (2012). Real-time properties of the Federal Reserve's output gap. Finance and Economics Discussion Series 2012-86, Board of Governors of the Federal Reserve System (U.S.).

Fleischman, C. A. and J. M. Roberts (2011). From many series, one cycle: improved estimates of the business cycle from a multivariate unobserved components model. Finance and Economics Discussion Series 2011-46, Board of Governors of the Federal Reserve System (U.S.).

Galì, J. (1999). Technology, employment, and the business cycle: Do technology shocks explain aggregate fluctuations? *The American Economic Review 89*(1), 249–271.

Galì, J. (2003). New perspectives on monetary policy, inflation, and the business cycle. In M. Dewatripont, L. P. Hansen, and S. J. Turnovsky (Eds.), *Advances in Economics and Econometrics: Theory and Applications*, Eighth World Congress of the Econometric Society, Chapter 6, pp. 151–197. Cambridge, UK: Cambridge University Press.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2003). *Bayesian Data Analysis* (2nd ed.). Chapman and Hall.

Gerlach, S. and F. Smets (1997, January). Output gaps and inflation: Unobservable components estimates for the G-7 countries. *mimeo* Bank for International Settlements.

Geweke, J. and G. Amisano (2010). Comparing and evaluating bayesian predictive distributions of asset returns. *International Journal of Forecasting 26*(2), 216 – 230. Special Issue: Bayesian Forecasting in Economics.

Gordon, R. J. (1997, Winter). The Time-Varying NAIRU and Its Implications for Economic Policy. *Journal of Economic Perspectives 11*(1), 11–32.

Gordon, R. J. (2012, August). Is U.S. economic growth over? faltering innovation confronts the six headwinds. NBER Working Papers 18315, National Bureau of Economic Research, Inc.

Harvey, A. and A. Jaeger (1993). Detrending, stylized facts and the business cycle. *Journal of Applied Econometrics 8*(3), 231–47.

Harvey, A. C. (1985, June). Trends and cycles in macroeconomic time series. *Journal of Business & Economic Statistics 3*(3), 216–27.

Harvey, A. C. (1989). *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press.

Harvey, A. C. and T. M. Trimbur (2003, May). General Model-Based Filters for Extracting Cycles and Trends in Economic Time Series. *The Review of Economics and Statistics 85*(2), 244–255.

Harvey, A. C., T. M. Trimbur, and H. K. Van Dijk (2007, October). Trends and cycles in economic time series: A bayesian approach. *Journal of Econometrics 140*(2), 618–649.

Justiniano, A. and G. E. Primiceri (2008, June). The time-varying volatility of macroeconomic fluctuations. *The American Economic Review 98*(3), 604–41.

Justiniano, A., G. E. Primiceri, and A. Tambalotti (2011, May). Is there a trade-off between inflation and output stabilization? NBER Working Papers 17071, National Bureau of Economic Research, Inc.

Justiniano, A., G. E. Primiceri, and A. Tambalotti (2013, April). Is there a trade-off between inflation and output stabilization? *American Economic Journal: Macroeconomics 5*(2), 1–31.

Kiley, M. T. (2013). Output gaps. *Journal of Macroeconomics 37*, 1–18.

Kim, C.-J. and C. R. Nelson (1998, May). Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime switching. *The Review of Economics and Statistics 80*(2), 188–201.

Kim, C.-J. and C. R. Nelson (1999, November). Has The U.S. Economy Become More Stable? A Bayesian Approach Based On A Markov-Switching Model Of The Business Cycle. *The Review of Economics and Statistics 81*(4), 608–616.

King, R. G. and S. T. Rebelo (1993). Low frequency filtering and real business cycles. *Journal of Economic Dynamics and Control 17*(1-2), 207–231.

Kuttner, K. (1992, January). Monetary policy with uncertain estimates of potential output. *Economic Perspectives 16*, 2–15.

Laubach, T. (2001, May). Measuring the NAIRU: Evidence from seven economies. *The Review of Economics and Statistics 83*(2), 218–231.

Laubach, T. and J. C. Williams (2003, November). Measuring the natural rate of interest. *The Review of Economics and Statistics 85*(4), 1063–1070.

Lippi, M. and L. Reichlin (1994, January). Diffusion of Technical Change and the Decomposition of Output into Trend and Cycle. *Review of Economic Studies 61*(1), 19–30.

Morley, J. C., C. R. Nelson, and E. Zivot (2003, May). Why are the Beveridge-Nelson and unobserved-components decompositions of GDP so different? *The Review of Economics and Statistics 85*(2), 235–243.

Nelson, C. R. and C. I. Plosser (1982). Trends and random walks in macroeconmic time series : Some evidence and implications. *Journal of Monetary Economics 10*(2), 139–162.

Newey, W. and K. West (1987). A simple positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*, 703–708.

Oh, K. H., E. Zivot, and D. Creal (2008, October). The relationship between the beveridge-nelson decomposition and other permanent-transitory decompositions that are popular in economics. *Journal of Econometrics 146*(2), 207–219.

Orphanides, A. and S. van Norden (2002, November). The unreliability of output-gap estimates in real time. *The Review of Economics and Statistics 84*(4), 569–583.

Orphanides, A. and S. van Norden (2005, June). The reliability of inflation forecasts based on output gap estimates in real time. *Journal of Money, Credit and Banking 37*(3), 583–601.

Perez-Quiros, G. and M. M. McConnell (2000, December). Output fluctuations in the united states: What has changed since the early 1980's? *American Economic Review 90*(5), 1464–1476.

Perron, P. and T. Wada (2009, September). Let's take a break: Trends and cycles in us real gdp. *Journal of Monetary Economics 56*(6), 749–765.

Piger, J., J. Morley, and C.-J. Kim (2005). Nonlinearity and the permanent effects of recessions. *Journal of Applied Econometrics 20*(2), 291–309.

Planas, C., W. Roeger, and A. Rossi (2013). The information content of capacity utilization for detrending total factor productivity. *Journal of Economic Dynamics and Control 37*(3), 577–590.

Planas, C., A. Rossi, and G. Fiorentini (2008, January). Bayesian analysis of the output gap. *Journal of Business & Economic Statistics 26*, 18–32.

Reifschneider, D., W. Wascher, and D. Wilcox (2013). Aggregate supply in the united states: recent developments and implications for the conduct of monetary policy. Finance and Economics Discussion Series 2013-77, Board of Governors of the Federal Reserve System (U.S.).

Shapiro, M. D. and M. W. Watson (1988). Sources of business cycle fluctuations. In S. Fischer (Ed.), *NBER Macroeconomics Annual 1988*, pp. 111–148. Cambridge (MA): The MIT Press.

Sims, C. A. and T. Zha (1999). Error bands for impulse responses. *Econometrica 67*(5), 1113–1156.

Sinclair, T. M. (2009a, December). Asymmetry in the business cycle: Friedman's plucking model with correlated innovations. *Studies in Nonlinear Dynamics & Econometrics 14*(1), 1–31.

Sinclair, T. M. (2009b, 03). The relationships between permanent and transitory movements in u.s. output and the unemployment rate. *Journal of Money, Credit and Banking 41*(2-3), 529–542.

Smets, F. and R. Wouters (2007, June). Shocks and frictions in U.S. business cycles: A Bayesian DSGE approach. *The American Economic Review 97*(3), 586–606.

Staiger, D. O., J. H. Stock, and M. W. Watson (1997). How precise are estimates of the natural rate of unemployment? In *Reducing Inflation: Motivation and Strategy*, NBER Chapters, Chapter 5, pp. 195–246. National Bureau of Economic Research, Inc.

Stella, A. and J. H. Stock (2013). A state-dependent model for inflation forecasting. *forthcoming* in *Unobserved Components and Time Series Econometrics*, Oxford University Press.

Stock, J. H. and M. W. Watson (2003). Has the business cycle changed and why? In *NBER Macroeconomics Annual 2002, Volume 17*, NBER Chapters, pp. 159–230. National Bureau of Economic Research, Inc.

Stock, J. H. and M. W. Watson (2007, February). Why has U.S. inflation become harder to forecast? *Journal of Money, Credit and Banking 39*(S1), 3–33.

Stock, J. H. and M. W. Watson (2009, September). Phillips curve inflation forecasts. In J. Fuhrer, Y. K. Kodrzycki, J. S. Little, and G. P. Olivei (Eds.), *Understanding Inflation and the Implications for Monetary Policy — A Phillips Curve Retrospective*, Chapter 3, pp. 99–184. Cambridge, MA: MIT Press.

Stock, J. H. and M. W. Watson (2010, October). Modeling inflation after the crisis. NBER Working Papers 16488, National Bureau of Economic Research, Inc.

Stock, J. H. and M. W. Watson (2012, May). Disentangling the channels of the 2007-2009 recession. NBER Working Papers 18094, National Bureau of Economic Research, Inc.

Watson, M. W. (1986, July). Univariate detrending methods with stochastic trends. *Journal of Monetary Economics 18*(1), 49–75.

Watson, M. W. (2007, Spring). How accurate are real-time estimates of output trends and gaps? *Economic Quarterly 93*(2), 143–161.

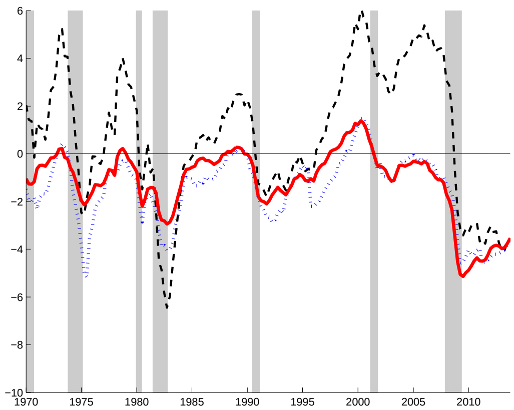
Watson, M. W. (2014, May). Inflation Persistence, the NAIRU, and the Great Recession. *American Economic Review 104*(5), 31–36.

Woodford, M. (2003). *Interest and Prices*. Princeton, NJ: Princeton University Press.

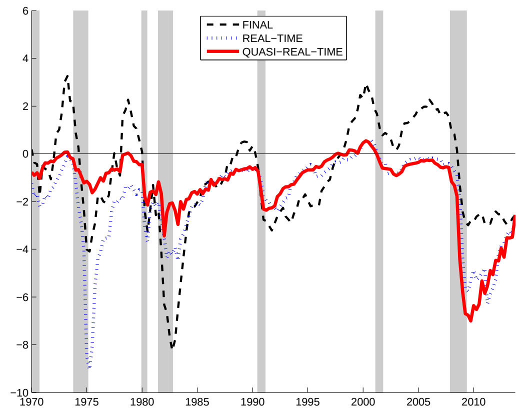
Figure 1: Output Gap Estimates and Revisions in the Univariate Model

Output Gap Estimates

(a) Constant parameters

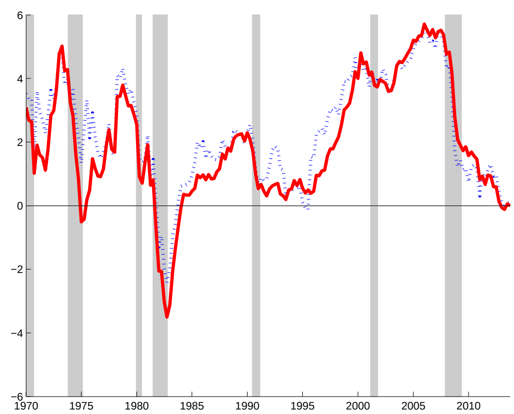


(b) SV



Revisions

(c) Constant parameters



(d) SV

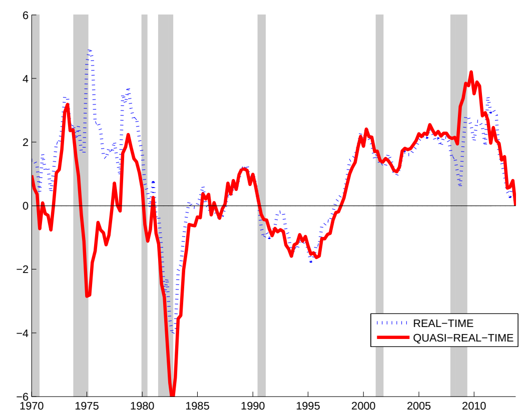
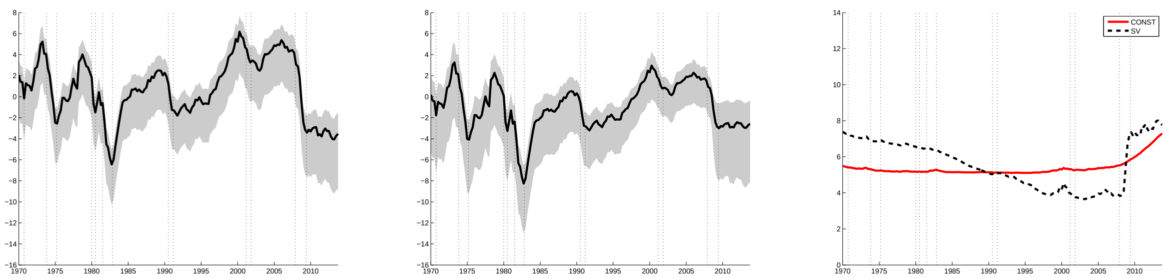


Figure 2: Uncertainty in Univariate Output Gap Estimates

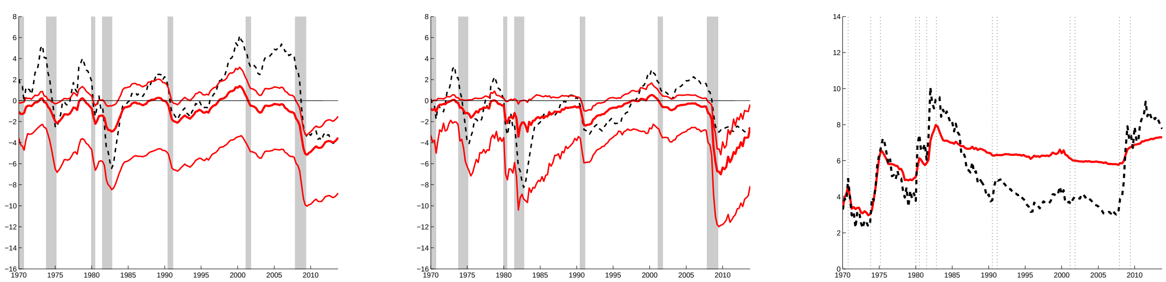
Final Estimates

(a) Constant parameters (b) SV (c) Width of credible sets



Quasi Real-time Estimates

(d) Constant parameters (e) SV (f) Width of credible sets



Real-time Estimates

(g) Constant parameters (h) SV (i) Width of credible sets

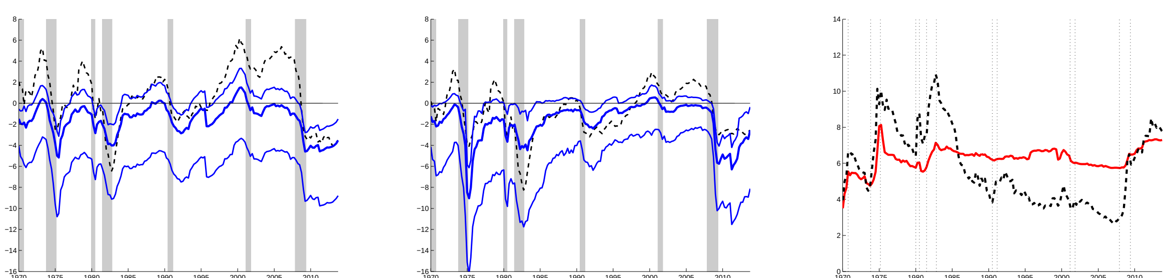
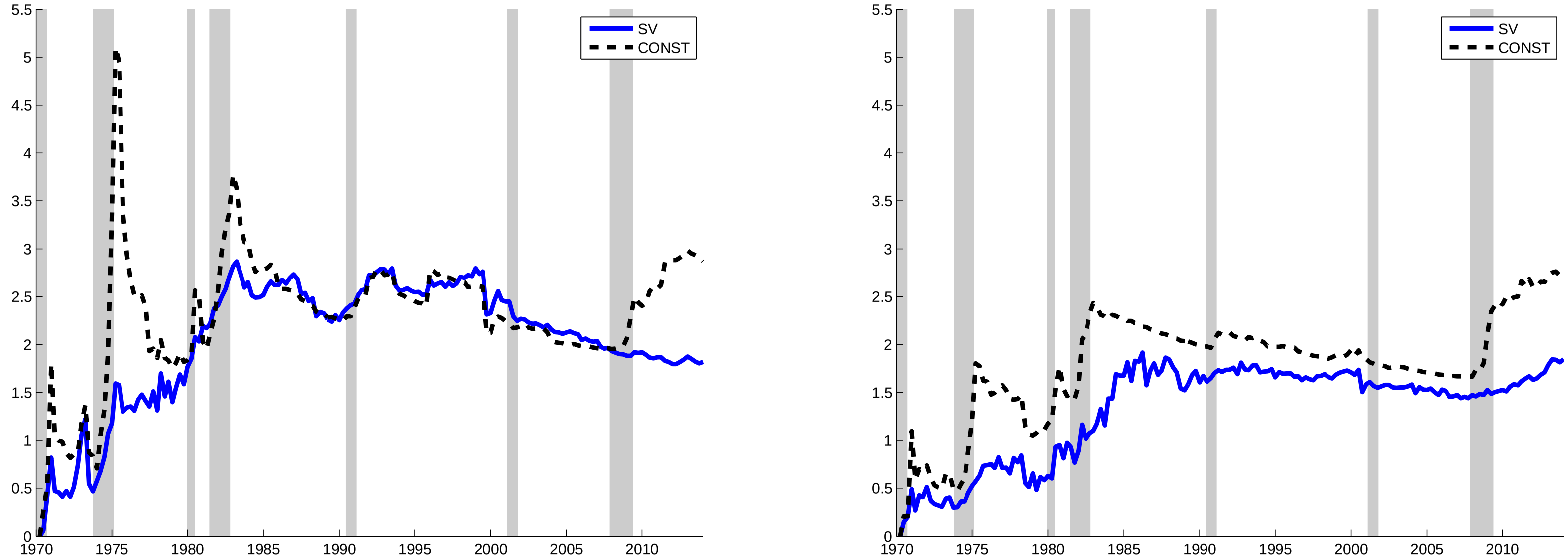


Figure 3: RMSE of Univariate Output Gap Revisions for Different "Final" Vintages

(a) Real time

(b) Quasi-real time



Note: Root mean squared errors (RMSE) are computed for quasi-real-time estimates as

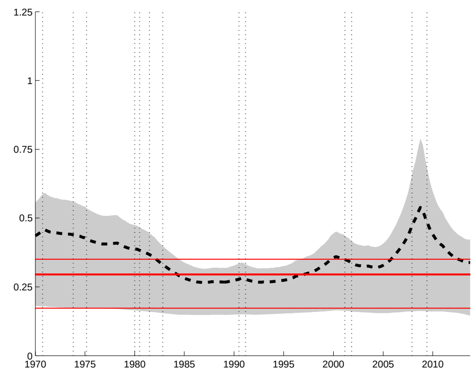
$$RMSE(\tau) = \sqrt{\frac{1}{\tau - \tau_0 + 1} \sum_{t=\tau_0}^{\tau} \left(\tilde{y}_{t|\tau} - \tilde{y}_{t|t}\right)^2}$$

with values of τ ranging from $\tau_0 = $ 1960:Q1 to $T = $ 2013:Q3. For real-time revisions, RMSE are computed using $E\left(\tilde{y}_t | Z^{t,v^*(\tau)}\right)$ in place of $\tilde{y}_{t|\tau}$ and $E\left(\tilde{y}_t | Z^{t,v^*(t)}\right)$ in place of $\tilde{y}_{t|t}$.

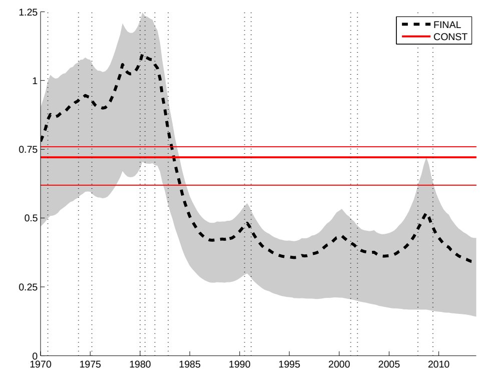
Figure 4: Volatility of Trend and Cycle in the Univariate Model

Final Estimates: SV vs. constant-parameter model

(a) Trend Vol.

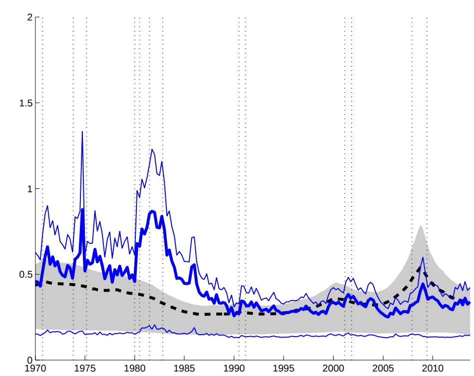


(b) Gap Vol.

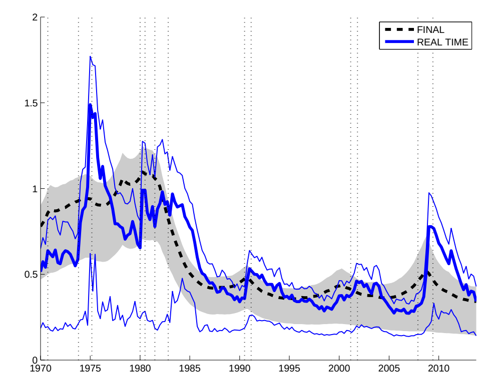


Final vs. Real-time Estimates in the SV model

(c) Trend Vol,

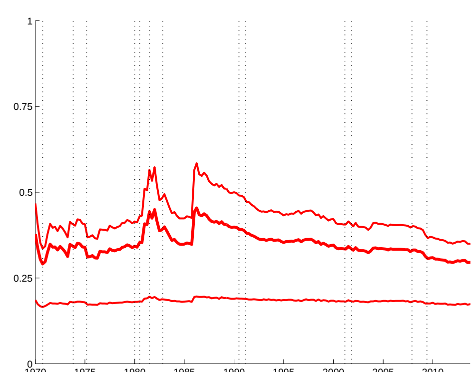


(d) Gap SV: final vs. real-time



Real-time Estimates from the constant-parameter model

(e) Trend Vol.

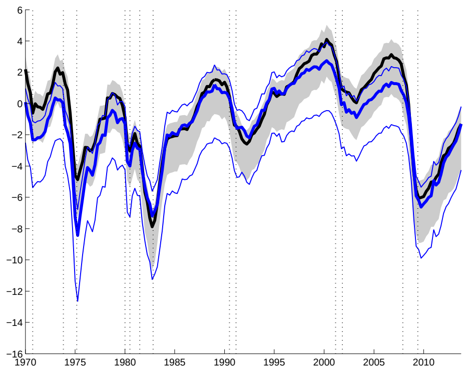


(f) Gap Vol.

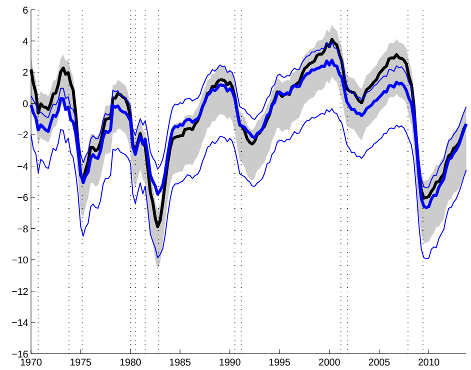
Figure 5: Bivariate Output Gap Estimates

Constant-parameter model

(a) Real-time vs. Final

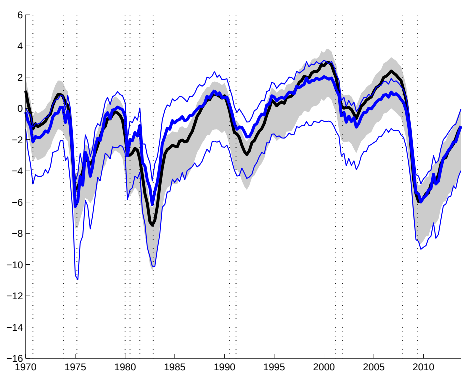


(b) Quasi-real-time vs. Final

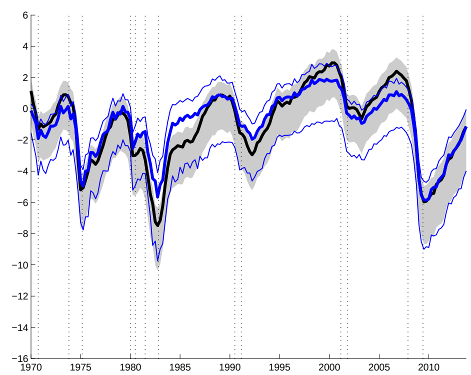


Scale SV model

(c) Real-time vs. Final

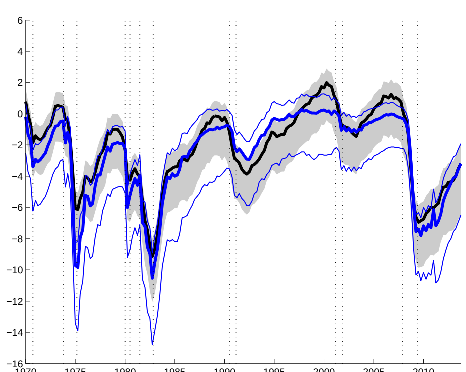


(d) Quasi-real-time vs. Final

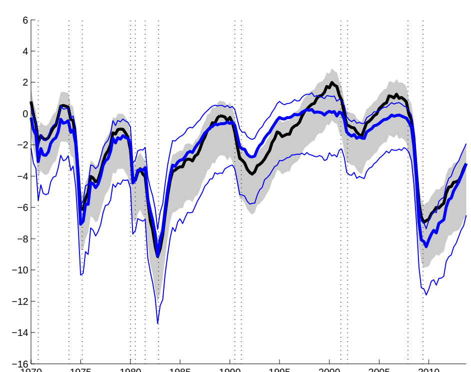


SV model

(e) Real-time vs. Final



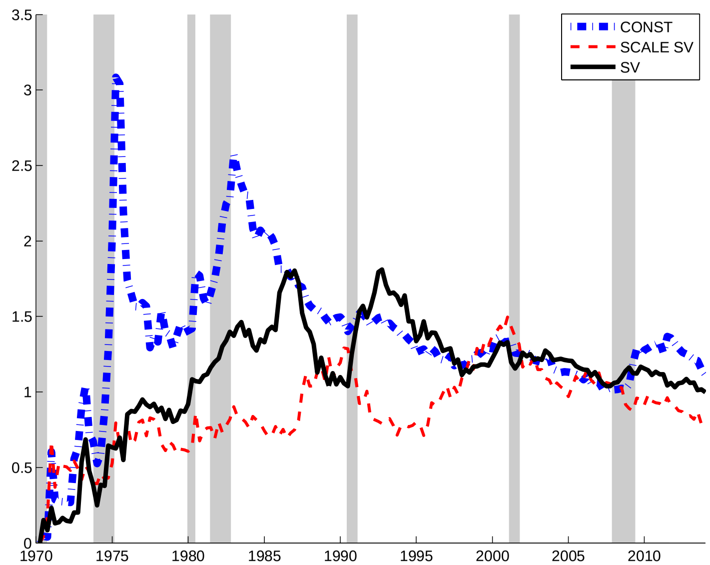
(f) Quasi-real-time vs. Final



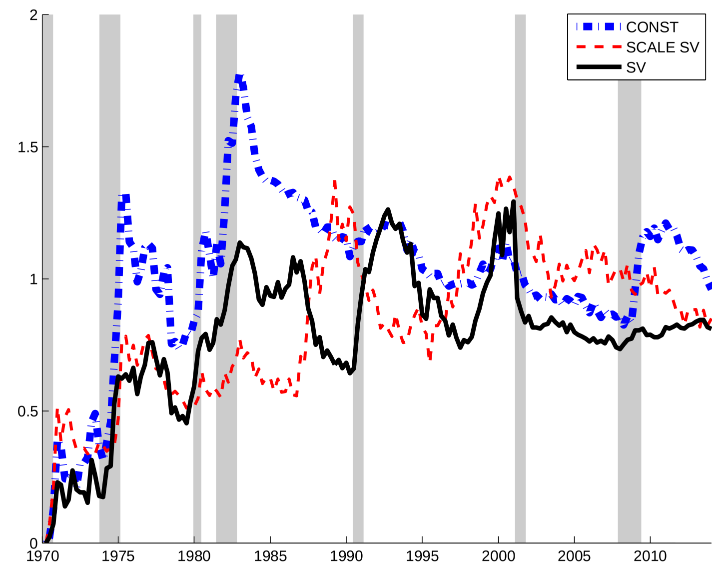
Note: Final estimates in black with gray-shaded 90% credible sets. (Quasi-)Real-time estimates are depicted by blue solid lines.

Figure 6: RMSE of Bivariate Output Gap Revisions for Different “Final” Vintages

(a) Real time



(b) Quasi-real time

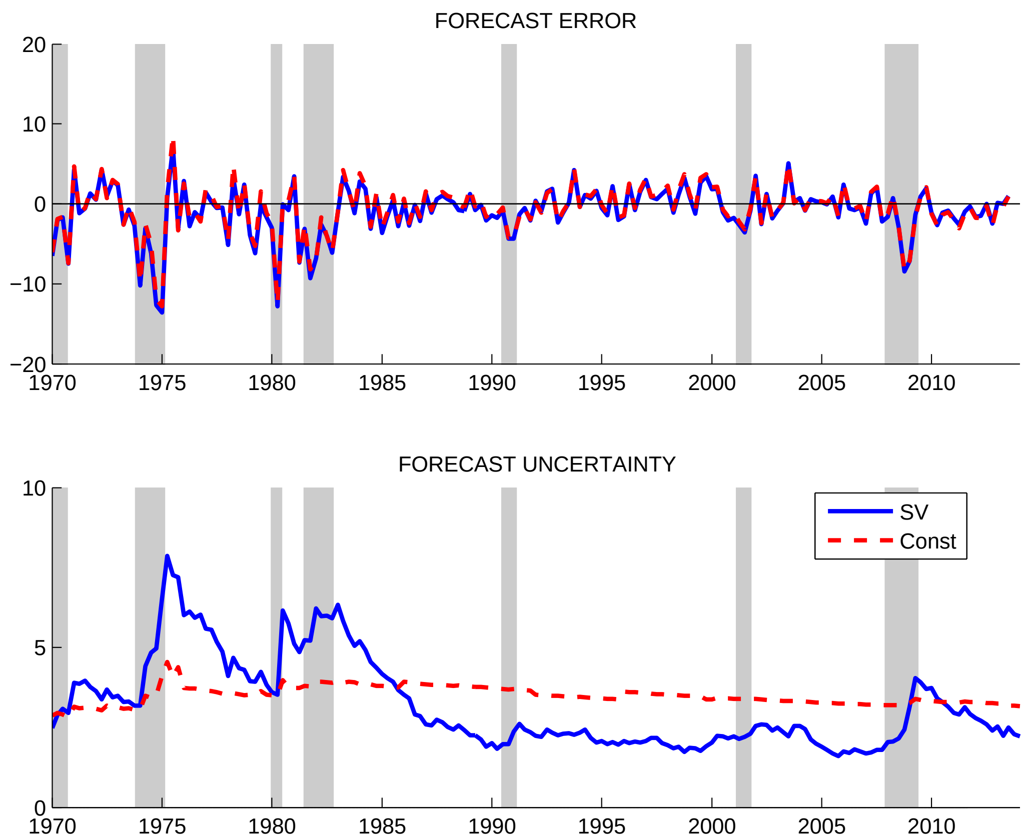


Note: Root mean squared errors (RMSE) are computed for quasi-real-time estimates as

$$RMSE(\tau) = \sqrt{\frac{1}{\tau - \tau_0 + 1} \sum_{t=\tau_0}^{\tau} \left(\tilde{y}_{t|\tau} - \tilde{y}_{t|t}\right)^2}$$

with values of τ ranging from $\tau_0 = $ 1960:Q1 to $T = $ 2013:Q3. For real-time revisions, RMSE are computed using $E\left(\tilde{y}_t|Z^{t,v^*(\tau)}\right)$ in place of $\tilde{y}_{t|\tau}$ and $E\left(\tilde{y}_t|Z^{t,v^*(t)}\right)$ in place of $\tilde{y}_{t|t}$.

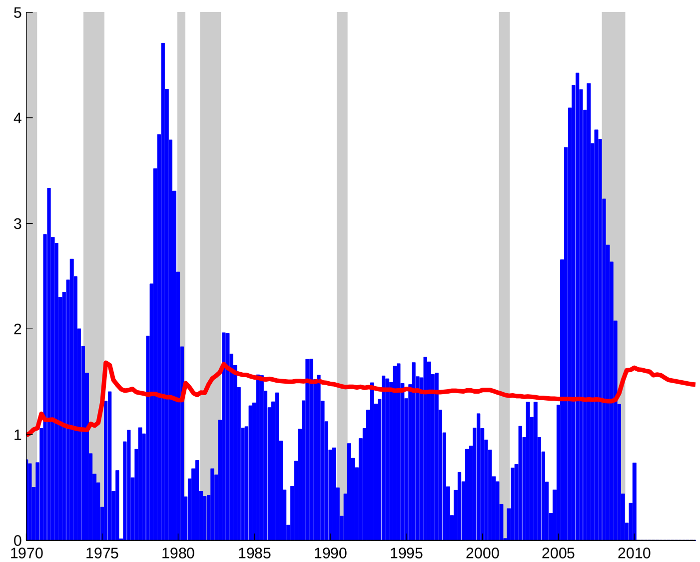
Figure 7: Forecast Errors and Uncertainty in Quarterly GDP Growth



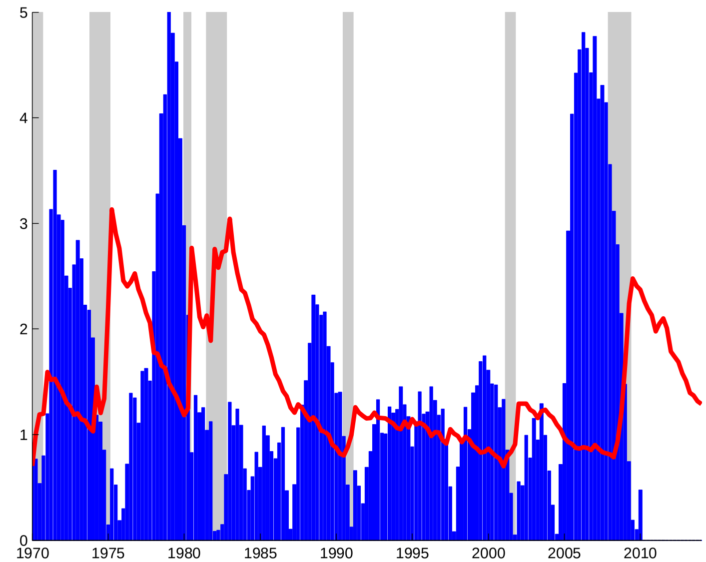
Note: The top panel displays errors in one-quarter ahead forecasts for quarterly output growth generated in real time by the constant-parameter version (red dashed) of the bivariate model as well as its "SV" counterpart (blue). The bottom panel depicts the corresponding forecast uncertainty, measured by the standard deviation of the predictive density. In both panels, the dates on the horizontal axis denote the vintage on which the forecasts are conditioned. Since the forecast errors depend also in the availability of subsequent data for the realizations, the forecast errors are not observable at the these vintage dates.

Figure 8: Absolute Forecast Errors and Uncertainty in Unemployment 4-years Ahead

(a) Constant parameters



(b) SV

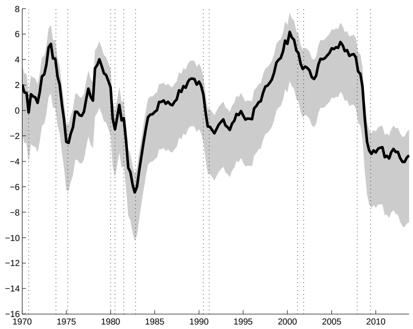


Note: Both panels display absolute errors and forecast uncertainty computed from the predictive density of the bivariate model for the unemployment rate four-years ahead. Results from the constant-parameter model are shown in the top panel, results from the SV version of the model are in the bottom panel. Forecast uncertainty is measured by the standard deviation of the predictive density. In both panels, the dates on the horizontal axis denote the vintage on which the forecasts are conditioned. Notice that the forecast errors can only be observed about four years after those dates.

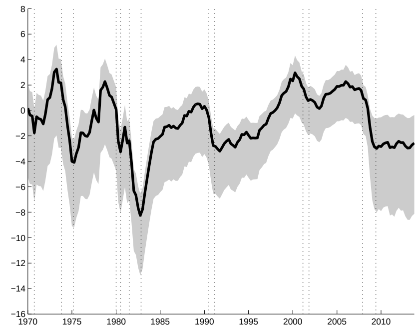
Figure 9: Posterior Distributions of Univariate Output Gap Estimates

Final Estimates

(a) Constant parameters

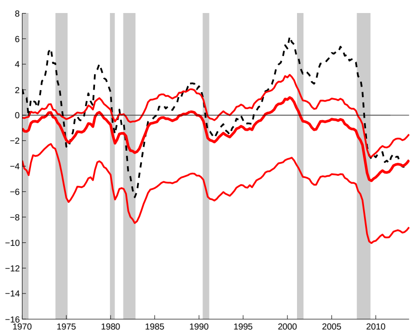


(b) SV

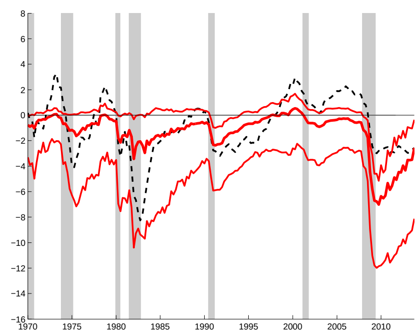


Quasi-real-time vs. Final Estimates

(c) Constant parameters

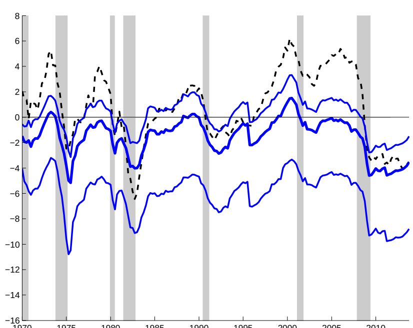


(d) SV



Real-time vs. Final Estimates

(e) Constant parameters



(f) SV

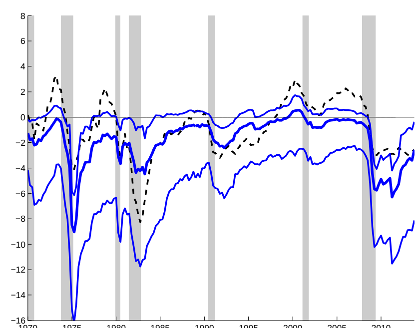
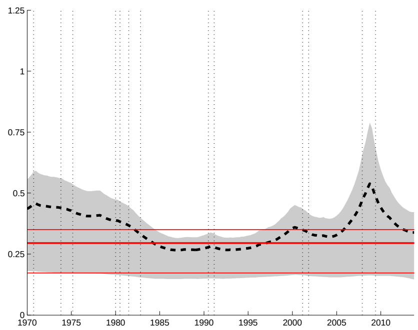


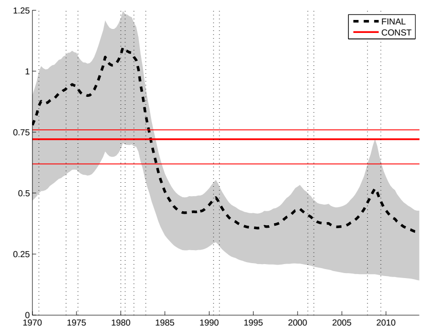
Figure 10: Volatility of Trend and Cycle in the Univariate Model (Quasi-real time)

Final Estimates: SV vs. constant-parameter model

(a) Trend Vol.

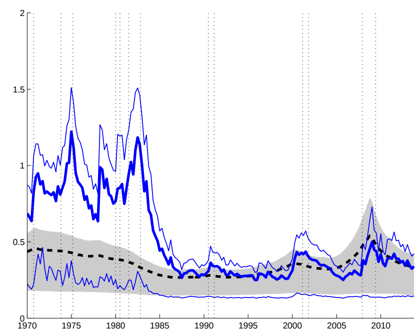


(b) Gap Vol.

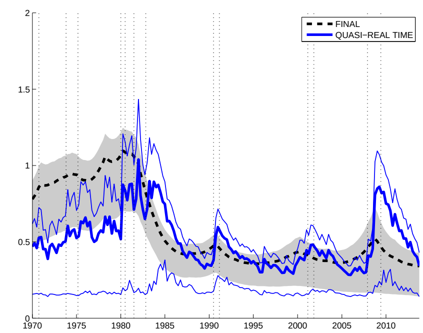


Final vs. Quasi-real-time Estimates in the SV model

(c) Trend Vol,

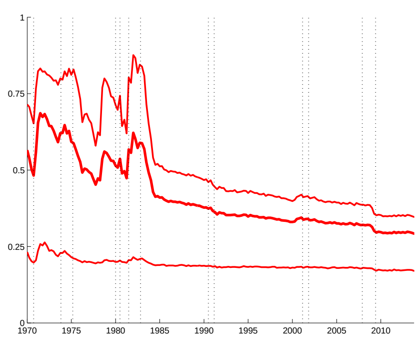


(d) Gap SV: final vs. real-time



Quasi-real-time Estimates from the constant-parameter model

(e) Trend Vol.



(f) Gap Vol.

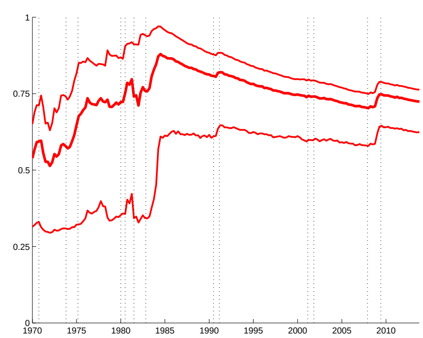


Table 1: Summary Statistics of Output Gap Revisions

Models	RMSE	VOL	AVG	MIN	MAX	RR	COR
Panel A: Real time							
const. (univariate)	2.87	1.68	2.33	−2.42	5.57	1.01	0.89
SV (univariate)	1.82	1.56	0.94	−4.04	4.95	0.82	0.71
const	1.12	0.84	0.74	−0.91	3.57	0.42	0.95
scale SV	0.82	0.82	−0.04	−2.18	1.48	0.35	0.94
SV	1.00	0.86	0.51	−1.26	3.72	0.41	0.94
Panel B: Quasi-real time							
const. (univariate)	2.71	1.88	1.95	−3.50	5.71	0.96	0.81
SV (univariate)	1.84	1.79	0.45	−6.19	4.20	0.84	0.60
const.	0.96	0.82	0.50	−2.09	2.29	0.36	0.96
scale SV	0.85	0.84	−0.12	−2.62	1.44	0.36	0.95
SV	0.81	0.76	0.27	−1.34	1.92	0.33	0.95

Note: RMSE, VOL, AVG, MIN and MAX are the root mean squared error, volatility, average, minimum and maximum values of the revisions. RR denotes the ratio of the RMSE relative to the volatility of the final output gap estimates and COR denotes the correlation between (quasi-)real-time and final estimates.

Table 2: Real-time Forecasts of Quarterly Output Growth

	Root mean-squared errors					log-predictive scores				
Models	h= 1 qtr	h= 4 qtrs	h= 8 qtrs	h=12 qtrs	h=16 qtrs	h= 1 qtr	h= 4 qtrs	h= 8 qtrs	h=12 qtrs	h=16 qtrs
From 1970:Q1 to 2014:Q1										
Univariate										
Const	3.17	3.46	3.38	3.39	3.42	−2.58	−2.69	−2.67	−2.68	−2.69
SV	1.03*	1.01**	1.02***	1.01**	1.01*	0.10**	0.05	−0.01	0.02	0.06
Bivariate										
Const	3.17	3.48	3.43	3.43	3.46	−2.58	−2.70	−2.70	−2.70	−2.71
SV scale	1.03*	0.99*	1.01*	1.00	1.00	0.01	0.03	−0.02	−0.03	−0.03
SV	1.02*	1.02**	1.02**	1.02**	1.01**	0.10***	0.04	−0.02	0.03	0.06
From 1985:Q1 to 2007:Q4										
Univariate										
Const	1.80	1.88	2.34	2.33	2.35	−2.32	−2.35	−2.44	−2.43	−2.43
SV	1.02**	1.03***	1.02*	1.01*	1.01	0.22***	0.17***	−0.01	0.02	0.06
Bivariate										
Const	1.84	1.95	2.34	2.33	2.36	−2.31	−2.35	−2.43	−2.42	−2.43
SV scale	1.02**	1.01	1.01	1.01	1.01	0.03***	0.02	−0.02	−0.05**	−0.06**
SV	1.02	1.02	1.02	1.01	1.01	0.22***	0.16***	−0.02	0.01	0.06

Note: For each group of models (univariate and bivariate), the first line reports RMSE and log-predictive scores, respectively, of the constant-parameter version of the model. Subsequent lines list relative RMSE and the difference in the log-predictive score with respect to the constant-parameter model. Relative RMSE are computed with the constant-parameter model in the denominator, and differences in log-predictive scores for a given model against the constant-parameter benchmark are computed such that higher values indicate better performance than in the benchmark. All growth rates are expressed as annualized differences in logs.

Table 3: Real-time Forecasts for the Unemployment Rate

	Root mean-squared errors					log-predictive scores				
Models	h= 1 qtr	h= 4 qtrs	h= 8 qtrs	h=12 qtrs	h=16 qtrs	h= 1 qtr	h= 4 qtrs	h= 8 qtrs	h=12 qtrs	h=16 qtrs
From 1970:Q1 to 2014:Q1										
Bivariate										
Const	0.29	0.92	1.45	1.71	1.82	−0.17	−1.42	−2.00	−2.18	−2.20
SV scale	0.98	1.02	1.00	1.01*	1.02**	0.08**	−0.02	−0.05	−0.06	−0.06
SV	0.99	1.05*	1.05*	1.07**	1.08**	0.01	−0.51*	−0.93*	−0.96**	−0.75*
From 1985:Q1 to 2007:Q4										
Bivariate										
Const	0.18	0.54	1.16	1.57	1.74	0.08	−0.92	−1.60	−1.96	−2.07
SV scale	0.98	1.00	1.01*	1.02**	1.03**	0.14***	0.02	−0.08*	−0.14*	−0.13*
SV	0.97	1.00	1.03	1.05	1.07	0.16**	−0.10	−0.85*	−1.12*	−0.98*

Note: For each group of models (univariate and bivariate), the first line reports RMSE and log-predictive scores, respectively, of the constant-parameter version of the model. Subsequent lines list relative RMSE and the difference in the log-predictive score with respect to the constant-parameter model. Relative RMSE are computed with the constant-parameter model in the denominator, and differences in log-predictive scores for a given model against the constant-parameter benchmark are computed such that higher values indicate better performance than in the benchmark.

Table 4: Real-time Forecasts of Cumulative Output Growth

	Root mean-squared errors					log-predictive scores				
Models	h= 1 qtr	h= 4 qtrs	h= 8 qtrs	h=12 qtrs	h=16 qtrs	h= 1 qtr	h= 4 qtrs	h= 8 qtrs	h=12 qtrs	h=16 qtrs
					From 1970:Q1 to 2014:Q1					
Univariate										
Const	3.17	2.46	1.94	1.57	1.35	−2.58	−2.35	−2.20	−1.99	−1.84
SV	1.03*	1.03*	1.04**	1.07***	1.09***	0.10**	−0.03	−0.21	−0.29*	−0.31**
Bivariate										
Const	3.17	2.46	1.98	1.65	1.43	−2.58	−2.40	−2.34	−2.21	−2.10
SV scale	1.03*	1.00	1.01	1.02	1.02*	0.01	0.10	0.15*	0.14	0.18
SV	1.02*	1.03**	1.06**	1.08**	1.10**	0.10***	−0.06	−0.20	−0.23	−0.20
					From 1985:Q1 to 2007:Q4					
Univariate										
Const	1.80	1.22	1.20	1.06	0.96	−2.32	−1.83	−1.65	−1.51	−1.40
SV	1.02**	1.06***	1.06**	1.08**	1.08**	0.22***	0.09	−0.31	−0.43*	−0.46**
Bivariate										
Const	1.84	1.31	1.25	1.10	1.00	−2.31	−1.85	−1.66	−1.52	−1.42
SV scale	1.02**	1.02	1.02	1.03	1.04	0.03***	0.03	−0.03	−0.06**	−0.07**
SV	1.02	1.02	1.04	1.06	1.07	0.22***	0.06	−0.30	−0.40*	−0.43*

Note: For each group of models (univariate and bivariate), the first line reports RMSE and log-predictive scores, respectively, of the constant-parameter version of the model. Subsequent lines list relative RMSE and the difference in the log-predictive score with respect to the constant-parameter model. Relative RMSE are computed with the constant-parameter model in the denominator, and differences in log-predictive scores for a given model against the constant-parameter benchmark are computed such that higher values indicate better performance than in the benchmark. All growth rates are expressed as annualized differences in logs.

Table 5: Quasi-real-time Forecasts of Quarterly Output Growth

	Root mean-squared errors					log-predictive scores				
Models	h= 1 qtr	h= 4 qtrs	h= 8 qtrs	h=12 qtrs	h=16 qtrs	h= 1 qtr	h= 4 qtrs	h= 8 qtrs	h=12 qtrs	h=16 qtrs
From 1970:Q1 to 2014:Q1										
Univariate										
Const	3.29	3.43	3.34	3.32	3.29	−2.61	−2.65	−2.63	−2.63	−2.63
SV	1.01**	1.00	1.01*	1.01*	1.01*	0.07**	−0.02	−0.06	−0.08	−0.03
Bivariate										
Const	3.23	3.38	3.32	3.33	3.31	−2.59	−2.64	−2.63	−2.64	−2.64
SV scale	1.00	1.00	1.00	1.00	1.00	0.03	0.01	−0.02	−0.04	−0.06
SV	1.00	1.01**	1.01**	1.01***	1.01***	0.08**	−0.03	−0.04	−0.07	−0.02
From 1985:Q1 to 2007:Q4										
Univariate										
Const	1.92	2.13	2.66	2.64	2.65	−2.33	−2.39	−2.50	−2.49	−2.49
SV	1.01	1.01	1.01	1.01	1.00	0.19***	0.11*	−0.07	−0.09	0.02
Bivariate										
Const	1.92	2.09	2.59	2.60	2.64	−2.32	−2.37	−2.48	−2.48	−2.48
SV scale	1.02**	1.01	1.00	1.01	1.00	0.05***	0.03***	0.01	−0.02	−0.02
SV	1.00	1.02	1.02**	1.02***	1.01***	0.18***	0.10*	−0.05	−0.09	0.01

Note: For each group of models (univariate and bivariate), the first line reports RMSE and log-predictive scores, respectively, of the constant-parameter version of the model. Subsequent lines list relative RMSE and the difference in the log-predictive score with respect to the constant-parameter model. Relative RMSE are computed with the constant-parameter model in the denominator, and differences in log-predictive scores for a given model against the constant-parameter benchmark are computed such that higher values indicate better performance than in the benchmark. All growth rates are expressed as annualized differences in logs.

Table 6: Quasi-real-time Forecasts of Cumulative Output Growth

	Root mean-squared errors					log-predictive scores				
Models	h= 1 qtr	h= 4 qtrs	h= 8 qtrs	h=12 qtrs	h=16 qtrs	h= 1 qtr	h= 4 qtrs	h= 8 qtrs	h=12 qtrs	h=16 qtrs
From 1970:Q1 to 2014:Q1										
Univariate										
Const	3.29	2.34	1.87	1.57	1.37	−2.61	−2.32	−2.16	−2.02	−1.91
SV	1.01**	1.01*	1.02*	1.03*	1.03*	0.07**	−0.06	−0.31	−0.47*	−0.53*
Bivariate										
Const	3.23	2.26	1.80	1.53	1.34	−2.59	−2.28	−2.14	−2.03	−1.93
SV scale	1.00	1.00	1.00	1.01	1.02	0.03	0.09	0.11	0.09	0.09
SV	1.00	1.03**	1.05***	1.07***	1.09***	0.08**	−0.08	−0.28	−0.37	−0.42
From 1985:Q1 to 2007:Q4										
Univariate										
Const	1.92	1.35	1.43	1.34	1.25	−2.33	−1.88	−1.82	−1.79	−1.77
SV	1.01	1.02	1.02	1.02	1.02	0.19***	0.03	−0.49*	−0.77*	−0.87*
Bivariate										
Const	1.92	1.32	1.34	1.24	1.16	−2.32	−1.85	−1.74	−1.68	−1.65
SV scale	1.02**	1.02*	1.01	1.02	1.02	0.05***	0.04*	−0.04	−0.08**	−0.07***
SV	1.00	1.03	1.06*	1.07**	1.08***	0.18***	0.01	−0.45*	−0.66*	−0.76*

Note: For each group of models (univariate and bivariate), the first line reports RMSE and log-predictive scores, respectively, of the constant-parameter version of the model. Subsequent lines list relative RMSE and the difference in the log-predictive score with respect to the constant-parameter model. Relative RMSE are computed with the constant-parameter model in the denominator, and differences in log-predictive scores for a given model against the constant-parameter benchmark are computed such that higher values indicate better performance than in the benchmark. All growth rates are expressed as annualized differences in logs.

Table 7: Quasi-real-time Forecasts for the Unemployment Rate

	Root mean-squared errors					log-predictive scores				
Models	h= 1 qtr	h= 4 qtrs	h= 8 qtrs	h=12 qtrs	h=16 qtrs	h= 1 qtr	h= 4 qtrs	h= 8 qtrs	h=12 qtrs	h=16 qtrs
From 1970:Q1 to 2014:Q1										
Bivariate										
Const	0.29	0.92	1.45	1.70	1.82	−0.18	−1.44	−2.00	−2.15	−2.18
SV scale	0.87***	1.00	1.00	1.01	1.02**	0.19***	0.07	0.03	−0.00	−0.01
SV	0.88***	1.03	1.04*	1.06**	1.08**	0.17***	−0.35*	−0.75*	−0.83*	−0.67*
From 1985:Q1 to 2007:Q4										
Bivariate										
Const	0.18	0.54	1.15	1.57	1.74	0.07	−0.92	−1.59	−1.95	−2.07
SV scale	0.87**	1.00	1.02**	1.03***	1.03***	0.26***	0.06***	−0.07	−0.13*	−0.12*
SV	0.84***	0.98	1.03	1.06	1.08	0.33***	0.01	−0.70	−1.00*	−0.91*

Note: For each group of models (univariate and bivariate), the first line reports RMSE and log-predictive scores, respectively, of the constant-parameter version of the model. Subsequent lines list relative RMSE and the difference in the log-predictive score with respect to the constant-parameter model. Relative RMSE are computed with the constant-parameter model in the denominator, and differences in log-predictive scores for a given model against the constant-parameter benchmark are computed such that higher values indicate better performance than in the benchmark.