Erasmus School of Economics
Master Thesis Quantitative Finance

The Output Gap: A Joint Model of Data Revisions and Real Output Dynamics

Name student: Marissa Berghuis
Student ID number: 405018

Supervisor: Dr. A. Pick
Second assessor: Dr. H.J.W.G. Kole

Date final version: 10 July 2020

ABSTRACT

A known challenge in producing reliable real-time output gap estimates is that initial measurements of real output suffer from data revisions. Therefore, after reviewing several existing output gap estimation methods, this paper introduces an updated estimation method that models data revisions to real output jointly with the dynamics of real output. More specifically, I propose to combine a state-space representation of the dynamics of true output with a state-space representation of uncertain output data. Based on a real-time data set of real U.S. output, the real-time estimates from the joint model considerably outperform those from other models. Furthermore, the real-time estimates from a multivariate joint model, which incorporates economic information in the form of a Phillips curve, are more reliable than those of the univariate joint model—indicating that incorporating such economic information is useful. These results provide a promising basis for models that combine perhaps more complex true output dynamics with those of uncertain output data, and are relevant to academics and practitioners alike.

Keywords: output gap, real-time estimates, data revisions.

Contents

I. Introduction

The output gap—the difference between an economy's actual and potential output—is relevant to many macroeconomic models and policy problems, particularly those of monetary nature. A positive output gap could be a signal that the economy is overheating, requiring tightening monetary policy, whereas a negative output gap reflects slack in the economy (Kara et al., 2007). The ability to give reliable (real-time) estimates of the output gap is therefore of interest to central banks, government institutions, and international organizations alike. For instance, the output gap is often used as a predictor variable in models of inflation. Furthermore, research on monetary policy has suggested that central banks use the Taylor rule, which involves the output gap, when setting interest rates (Álvarez and Gómez-Loscos, 2018). It is not surprising, then, that Rudebusch (2001) and Ehrmann and Smets (2003) find unreliable estimates of the output gap to have a direct impact on optimal monetary policy. In fact, Orphanides (2003) argues that unreliable real-time estimates have caused monetary policymakers to misrepresent the output gap and thereby create an inflationary or disinflationary bias.

One challenge in measuring the output gap is that potential output is unobserved and hence needs to be estimated. The potential output of an economy is generally defined as the highest level of real GDP that can be produced and sustained over the long term without giving rise to inflation (Okun, 1963). While this is unobserved, there are numerous methods for estimating potential output from an actual output series, which can be broadly categorized into univariate and multivariate methods. Univariate methods simply estimate potential output from actual output without incorporating information from other economic variables, such that no assumptions about the structure of the economy have to be made. On the contrary, multivariate methods employ economic theory to incorporate information contained in other variables related to the output gap (St-Amant and Van Norden, 1997).

Univariate methods can be further divided into whether they use filters—such as the Hodrick-Prescott (1997) filter, band-pass filters, or wavelet filtering—or models, such as linear detrending, the Beveridge-Nelson (1981) decomposition using ARIMA models, structural time series models, or Markov switching models (Álvarez and Gómez-Loscos, 2018). Since these methods do not incorporate information from other economic variables, the resulting potential output estimates can be better interpreted as the "trend" of an actual output series. Therefore, they are purely statistical and not consistent with the Okun (1963) definition of potential output. Nevertheless, univariate methods are relatively easy to implement and can be applied as long as output data is available, even when data on other economic variables is not (Blagrave et al., 2015).

The most prevalent multivariate methods are based on Okun's law (Evans, 1989), production functions (e.g. Havik et al., 2014), the Blanchard-Quah (1989) decomposition through supply and demand shocks, the Phillips curve (Kuttner, 1994), the natural rate of interest (Laubach and Williams, 2003), real business cycle models (King et al., 1991), or dynamic stochastic general equilibrium models (Vetlov et al., 2011). The production function approach, which relates output to capital, labor, and total factor productivity, is particularly common in practice and used by institutions such as the European Commission (Havik et al., 2014). While this method adds economic structure to potential output, it involves de-trending the total factor productivity series. The resulting potential output estimates therefore have practically the same properties as those resulting from applying the same (univariate) de-trending method directly to output data (Blagrave et al., 2015). Among academics, dynamic stochastic general equilibrium (DGSE) models gained popularity in recent years. Nonetheless, this approach is difficult to apply in practice and requires substantial modeling skills and implementation time (Blagrave et al., 2015).

Besides having to estimate potential output, another challenge in estimating the output gap is that initial measurements of real output suffer from data revisions. For real-time output gap estimates to be reliable, it is therefore crucial that they do not change much when the underlying output data is revised. It can be expected that real-time estimates are improved when possible future data revisions are taken into account. Therefore, Garratt et al. (2008) and Clements and Galvão (2012) have focused on forecasting post-revision output data through vector-autoregressive (VAR) models of past real-time data. Other researchers take a state-space approach by relating published output data with bias and measurement errors to their unobserved true values (e.g. Jacobs and Van Norden, 2011; Cunningham et al., 2012). After predicting how output data will be revised or estimating the true output values, common estimation methods can be applied to the forecasted post-revision data or estimated true output values in an attempt to improve the real-time estimates of the output gap.

In this paper, I first review several widely used as well as more recently proposed estimation methods for the output gap—which each assume different dynamics for the real output process or take on a different definition of real output—to examine their performance using recent output data. The widely used methods that are reviewed are the univariate Hodrick-Prescott (HP) filter, Beveridge-Nelson (BN) decomposition, and Watson model and the multivariate Kuttner model. The former two are particularly common and often used as benchmark methods in the literature, and the Watson model and Kuttner model (which is a multivariate version of the Watson model in that it adds a Phillips curve) serve as a base for the estimation method proposed in this paper. In addition, I review a method for estimating the output gap that takes into account possible future revisions to the underlying output data through a VAR model of past real-time data.

After reviewing the existing methods, I introduce an updated estimation method for the output gap that models data revisions to real output jointly with the dynamics of real output. More specifically, I propose to combine a state-space representation of the dynamics of true output, such as those by Watson (1986) or Kuttner (1994), with a state-space representation of uncertain output data, such as that by Cunningham et al. (2012). That way, rather than the two-step procedure described above of predicting how output data will be revised and then applying a common estimation method to the forecasted post-revision data, the output gap can immediately be extracted. While Orphanides and Norden (2002) famously find that the unreliability of real-time output gap estimates is mostly due to the parameter instability of the used estimation methods rather than to data revisions, the proposed model may be expected to be less affected by either of these issues. That is, since the parameters of the proposed model are estimated based on a "true" output series that does not get revised much over time, they should be more stable.

To assess the performance of the different estimation methods, I use a data set constructed by Croushore and Stark (2001). This data set consists of quarterly vintages, or records, from Q4 1965 to Q1 2020 of real output of the United States available in real time at quarterly intervals since Q1 1947. Two main performance criteria are adapted: (1) an estimation method should give output gap estimates that are reliable in real time and do not change much when new information becomes available, and (2) real-time estimates should have predictive power over inflation in order to ensure that they have economic content (Camba-Mendez and Rodriguez-Palenzuela, 2003). I find that the real-time estimates from the HP filter and Watson model have relatively poor reliability, whereas the estimates from the BN decomposition almost seem like noise. On the other hand, the real-time estimates from the joint Watson and Kuttner models considerably outperform those from their original counterparts. Also interesting is that the real-time estimates from both the original and the joint Kuttner model are more reliable than those of the corresponding Watson models, indicating that incorporating economic information in the form of a Phillips curve is useful. Similar to the findings by Orphanides and Norden (2002), data revision plays a rather small role in the unreliability of real-time estimates, and most of the revision is attributable to parameter instability of the estimation methods instead. As expected, however, especially the parameters of the joint Kuttner model are indeed more stable than those of its original counterpart, resulting in more reliable real-time estimates. Overall, these results provide a promising basis for models that combine perhaps more complex true output dynamics with those of uncertain output data.

The rest of this paper is organized as follows. Section II presents the data that is used to examine the estimation methods applied in this research. Section III introduces the estimation methods and describes how their performance is assessed. Section IV reports the results of the analyses carried out on the estimation methods and discusses how they relate to the existing literature. Section V concludes and indicates directions for future research.

II. Data

To assess the reliability of estimation methods for the output gap in real time, a data set is needed which contains real-time output data as it was available on any given date in the past. To that end, I use the "Real GNP/GDP (ROUTPUT)" data set constructed by Croushore and Stark (2001), which is available online via the Real-Time Data Research Center of the Federal Reserve Bank of Philadelphia (`https://www.philadelphiafed.org/research-and-data/real-time-center/real-time-data/data-files`).

The data set consists of quarterly vintages, or records, since Q4 1965 of real output of the United States available in real time at quarterly intervals since Q1 1947. In other words, for each quarterly vintage from Q4 1965 to the present, the data set includes historical quarterly data since Q1 1947 on real U.S. output as it was available in published sources in the middle (thus, on the

15th day of the second month) of that quarterly vintage. Real U.S. output is published with a lag of one quarter, such that the first estimate of output in quarter t becomes available in vintage $t+1$. The data set therefore has 217 vintages (equal to the number of quarters between Q4 1965 and Q1 2020) and between 75 and 292 observations per vintage (as more recent vintages have more historical data). Most of the data set is complete, but for a few vintages—all quarters of 1992 and 1996, Q1 1997, Q4 1999, and Q1 2000—the historical data up to around 1960 is missing. Overall, this results in a total of 39,464 observations. Note that for the vintages with missing data, the estimation methods are only applied to the available data.

Figure 1 visualizes how real U.S. output for a given quarter is revised over time. Each line represents a quarter between Q4 1947 (the very bottom line) and Q1 2020 (the very top line). It becomes clear that revisions to initial data releases can be quite large and continue to be made many quarters into the future. In fact, due to new information, output data is almost always revised in the first few quarters after it is initially released, and annual revisions published in July can affect up to the past three years of data. The plot also reveals that revisions are sometimes made to all quarters in unison, for instance in Q4 1999. These are benchmark revisions involving conceptual or methodological changes (Jacobs and Van Norden, 2011).

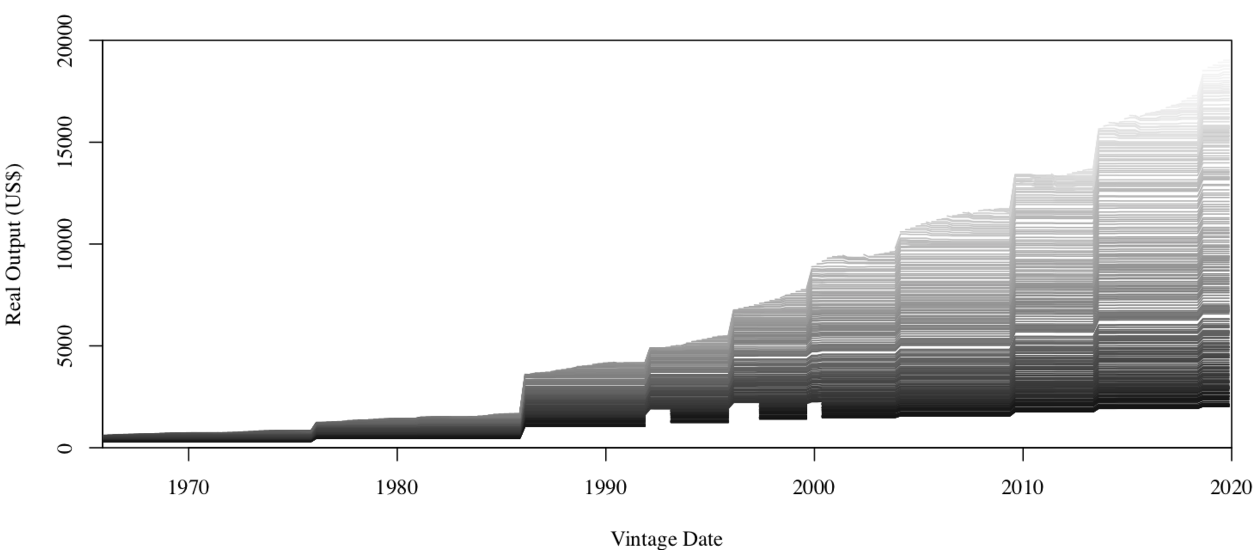

Figure 1. Revisions in Real Output Over Time. A visualization of how real U.S. output for a given quarter is revised over time. Each line represents a quarter between Q4 1947 and Q1 2020.

Since multivariate estimation methods of the output gap also incorporate information from other economic variables, some additional data is needed for some of the methods applied in this paper. More specifically, I calculate the quarterly inflation rate based on the Consumer Price Index (CPI), which is obtained from the same online database as the output data. Following Orphanides and Norden (2002), since CPI data is not revised in a way that output data is, I only use the CPI data from the most recent vintage (Q1 2020).

Table 1 provides summary statistics for the macroeconomic data used in this research. It is clear that real output is positively skewed, in that it experiences exponential growth. For that reason, the estimation methods used in this paper will be applied to the log of real output. The table also reports p-values of augmented Dickey-Fuller (ADF) tests at different lags, which test the null hypothesis that the data series has a unit root and is thus non-stationary. Note that for output data, the ADF tests are performed on the most recent vintage. Since the resulting p-values for (the log of) real output are highly insignificant, the presence of a unit root cannot be rejected. The inflation rate, however, seems to be a stationary process for which up to eight lags may contain predictive information on how the series will change over time.

Table 1. Summary Statistics for Macroeconomic Data.

	Min	Median	Max	ADF(2)	ADF(4)	ADF(8)	ADF(12)
Real output (US$)	306.40	3,050.70	19,219.80	0.912	0.902	0.932	0.917
Log real output	5.72	8.02	9.86	0.666	0.902	0.794	0.930
Inflation rate (%)	-3.42	0.72	4.16	0.010	0.010	0.038	0.360

Minimum, median, and maximum values for the macroeconomic data, and p-values of augmented Dickey-Fuller (ADF) tests at different lags. For output data, the ADF tests are performed on the most recent vintage.

III. Methodology

Measuring the output gap requires an estimate of potential output. Potential output is unobserved, but there are numerous methods for estimating it. This section first presents some existing methods that are reviewed in this paper. Then, an updated estimation method for the output gap is introduced. Lastly, I discuss the methodology behind assessing the performance of the different methods. In what follows, y_t denotes the log of real output at time t, y_t^g denotes its trend (or growth) component, and y_t^c denotes the cyclical component.

A. Existing Estimation Methods for the Output Gap

The existing estimation methods for the output gap that I review in this paper are a selection of widely used as well as more recently proposed methods, which each assume different dynamics for the real output process or take on a different definition of potential output. The widely used methods that are reviewed are the univariate Hodrick-Prescott filter, Beveridge-Nelson decomposition, and Watson model and the multivariate Kuttner model. In addition, I review a method for estimating the output gap that takes into account possible future revisions to the underlying output data through a vector-autoregressive (VAR) model of data vintages.

A.1. Hodrick-Prescott filter

The Hodrick-Prescott (HP) filter is a univariate filtering approach that can be used to decompose the log of real output into the sum of a trend and a cyclical component, such that $y_t = y_t^g + y_t^c$.

This approach assumes that the trend is stochastic and varies smoothly over time (Álvarez and Gómez-Loscos, 2018). The trend is obtained by solving the following minimization problem, which balances the variation in the cyclical component (or the trend's fit to the original series; first term) with the variation in the second difference of the trend component (or the trend's smoothness; second term):

$$\{y_t^g\}_{t=0}^T = \arg\min \sum_{t=1}^T (y_t - y_t^g)^2 + \lambda \sum_{t=1}^T ((y_{t+1}^g - y_t^g) - (y_t^g - y_{t-1}^g))^2, \tag{1}$$

where λ, which has to be set by the user, represents a smoothness parameter that penalizes the variation in the trend. A larger smoothness parameter results in a smoother trend (St-Amant and Van Norden, 1997). As suggested by Hodrick and Prescott (1997), I set $\lambda = 1,600$ for quarterly data.

Despite the fact that the HP filter is simple, widely applicable, and easy to reproduce, it has been criticized by many for suffering from the so-called end-of-sample problem. When decomposing an observation, the HP filter places equal importance on past and future observations, making it two-sided symmetric (Garratt et al., 2008). The filter is therefore only optimal at the observation in the middle of the sample. Towards either end of the sample, the filter becomes more and more one-sided. As a result, Mise et al. (2005) show that while the filter remains unbiased, it becomes inefficient. This poor performance of the HP filter at the end of the sample is a large drawback when estimating the output gap, as this is the part of the sample most relevant for policymakers (St-Amant and Van Norden, 1997).

A.2. Beveridge-Nelson decomposition

The Beveridge-Nelson (BN) decomposition is a univariate model-based approach that can be used to break down the log of real output into additive trend and cyclical components. This approach assumes that the log of real output is non-stationary, which seems reasonable based on the augmented Dickey-Fuller (ADF) tests from Table 1, but that its first differences are stationary. Indeed, an ADF test on the first difference of the log of real output returns highly significant test statistics—suggesting stationarity—up to many lags (p-value = 0.010 even at lag order 12). The BN decomposition starts by modeling these first differences as an autoregressive integrated moving average (ARIMA) process as follows:

$$\begin{aligned}\Delta y_t &= \delta + \frac{\theta_q(L)}{\phi_p(L)}\varepsilon_t \\ &= \delta + \psi(L)\varepsilon_t \\ &= \delta + \psi(1)\varepsilon_t + \tilde{\psi}(L)\varepsilon_t,\end{aligned}$$

where δ is the mean of the process, $\theta_q(L)$ is a lag polynomial of the order q for the moving average, $\phi_p(L)$ is a lag polynomial of the order p for the autoregression, and $\varepsilon_t \sim$ i.i.d. $\mathrm{N}(0, \sigma_\varepsilon^2)$. In this research, $p = 1$ and $q = 2$ lead to the ARIMA model with the lowest Akaike Information Criterion (AIC), based on the output data from the most recent vintage, and will thus be used. The last step, where clearly $\tilde{\psi}(L) = \psi(L) - \psi(1)$, allows the first difference of the log of real output to be decomposed into the first difference of the trend component and the first difference of the cyclical component, such that

$$\Delta y_t^g = \delta + \psi(1)\varepsilon_t$$
$$\Delta y_t^c = \tilde{\psi}(L)\varepsilon_t.$$

Note that the first difference of the trend component follows a random walk with drift. The resulting trend component can be expressed as

$$y_t^g = \frac{\theta_q(1)}{\phi_p(1)} \frac{\phi_p(L)}{\theta_q(L)} y_t, \tag{2}$$

from which it is clear that the trend is a weighted average of past and current, but not future, observations. Beveridge and Nelson (1981) interpret this trend component in terms of the long-run forecast of output, which is quite different from the perspective of the HP filter.

Even though the forecasting perspective of the BN decomposition approach allows the trend to be readily interpreted as potential output (Garratt et al., 2008), there are a few disadvantages of using this approach to estimate the output gap. First, the trend and cycle components are driven by the same shock, so their innovations are perfectly correlated. Furthermore, the variance of the innovations in the trend component may be larger than the innovation in the data, resulting in a too noisy trend. Lastly, it may be that multiple specifications of the ARIMA model fit the data rather well, even though they result in quite different decompositions (Álvarez and Gómez-Loscos, 2018). In this research, however, changing the ARIMA specification is not found to have a large impact on the results and their interpretation.

A.3. Watson model

Watson (1986) models the trend of the log of real output as a random walk with drift and the cycle as an AR(2) process, thereby allowing for persistence in the business cycle, as follows:

$$y_t = y_t^g + y_t^c \tag{3}$$
$$y_t^g = \mu_y + y_{t-1}^g + e_t \tag{4}$$
$$y_t^c = \phi_1 y_{t-1}^c + \phi_2 y_{t-2}^c + u_t, \tag{5}$$

where μ_y is a constant, and $e_t \sim$ i.i.d. $\mathrm{N}(0, \sigma_e^2)$ and $u_t \sim$ i.i.d. $\mathrm{N}(0, \sigma_u^2)$ are uncorrelated. This so-called univariate unobserved components model implies that real output is difference stationary and can be decomposed into an integrated trend component and a stationary cycle component, which seems reasonable according to the ADF tests from Table 1. The parameters of the model—

ϕ_1, ϕ_2, σ_e^2, and σ_u^2—can be estimated by Maximum Likelihood (ML) through the use of a Kalman filter on a state-space representation of the model (see e.g. Harvey, 1990). In order to do so, the model in (3)–(5) is first rewritten in state-space form, such that

$$x_t = Z\alpha_t + \epsilon_t \qquad \text{(observation equation)} \tag{6}$$

$$\alpha_{t+1} = F\alpha_t + R\eta_t \quad \text{(state equation)}, \tag{7}$$

where in this case $x_t = y_t$, $\epsilon_t \sim \mathrm{N}(0, H)$, $\eta_t \sim \mathrm{N}(0, Q)$, and $\alpha_1 \sim \mathrm{N}(\alpha_1, P_1)$ are independent, H is zero, R is the identity matrix of size 4, and the state vector α_t and system matrices F, Q, and Z are defined as

$$\alpha_t = \begin{bmatrix} y_t^g \\ y_t^c \\ y_{t-1}^c \\ \mu_y \end{bmatrix}, \quad F = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & \phi_1 & \phi_2 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad Q = \begin{bmatrix} \sigma_e^2 & 0 & 0 & 0 \\ 0 & \sigma_u^2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \text{and} \quad Z = \begin{bmatrix} 1 & 1 & 0 & 0 \end{bmatrix}.$$

Before applying the Kalman filter recursions to this state-space model, I define a prior distribution of the initial state vector, α_1, according to the methodology by Virmani (2004). More specifically, I take estimates for y_1^g and y_1^c from the output of the HP filter and estimate μ_y by applying OLS to (4) using the trend component from the HP filter. Table 2 presents the prior distribution of the initial state vector for the most recent vintage. Since the state vector includes an observation from one period back, the Kalman filter starts at the second observation. After running the filter as described by Harvey (1990), the model parameters can be estimated through ML by minimizing the negative of the following log-likelihood function:

$$\log L = -\frac{np}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{n}(\log\left|ZP_tZ'\right| + (x_t - Z\alpha_t)'(ZP_tZ')^{-1}(x_t - Z\alpha_t)),$$

where n is the number of observations, and p is the dimension of x_t. In order to support the algorithm in finding suitable parameter estimates, I define initial parameter estimates by performing OLS on (4) and (5) using the growth and trend components from the HP filter, which again follows the methodology by Virmani (2004). The resulting initial model parameters for the most recent vintage are also presented in Table 2. Note that the ϕ's indeed imply that the cycle component is stationary yet highly persistent. Furthermore, in terms of variance, shocks to the cycle are quite a bit larger than shocks to the trend. Lastly, the estimated drift term of 0.0078 for the trend component, similar to that reported by Watson (1986), suggests an average annualized growth rate of 3.14 percent.

A.4. Kuttner model

Kuttner (1994) adds a Phillips curve to the univariate Watson model in order to incorporate more information. The resulting multivariate unobserved components model ensures that potential

Table 2. Initialization of Watson Model Parameters and State Vector Distribution.

	Initial State Vector Distribution			
	y_t^g	y_t^c	y_{t-1}^c	μ_y
α_1	7.6025	0.0121	0.0254	0.0078
P_1	0.4441	0.0002	0.0002	2×10^{-8}
	Initial Model Parameters			
	σ_e	σ_u	ϕ_1	ϕ_2
	0.0026	0.0077	1.1643	-0.3699

Initial values for the parameters and for the state vector distribution of the Watson model for the most recent vintage.

output can be interpreted as the level of output at which inflation is constant, which is consistent with the definition by Okun (1963). The economic theory behind this is that a positive output gap suggests tight product and labor markets, causing inflation to increase in the short-run. On the other hand, when a negative output gap implies slack in product and labor markets, inflation generally decreases in the short-run (Álvarez and Gómez-Loscos, 2018). Therefore, in addition to (3)–(5), Kuttner (1994) includes a Phillips-curve equation that shows a positive relation between the change in the inflation rate, $\Delta\pi_t$, and the lagged output gap, y_{t-1}^c, as follows:

$$\Delta\pi_t = \mu_\pi + \gamma\Delta y_{t-1} + \beta y_{t-1}^c + v_t + \delta_1 v_{t-1} + \delta_2 v_{t-2} + \delta_3 v_{t-3},$$

where μ_π is a constant, and $v_t \sim$ i.i.d. $N(0, \sigma_v^2)$ may be correlated to u_t from (5). As can be seen, the change in the inflation rate is assumed to follow a MA(3) process, and lagged real output growth, Δy_{t-1}, is included to capture its positive correlation with inflation (Kuttner, 1994). Since Kuttner (1994) based this specification on data from 1954 to 1992, it is not necessarily the best fit for the more recent data used in this paper. Table 3 presents estimates of different MA specifications of the inflation data used in this paper, using the growth component from the HP filter for y_{t-1}^c. For each model, the standard error (SE), test statistic from a Ljung-Box Q test on 16 lags of the residuals (Q(16)), and Aikaike Information Criterion (AIC) are reported.

An autocorrelation plot of the change in the inflation rate suggests starting with a model of five lags, as shown in the first row of Table 3. It becomes clear that four lags are significant, which is in fact not consistent with Kuttner (1994)'s specification of only three lags. What is consistent, however, is that only one lag of real output growth is significant. While not reported, various AR specifications were also tested, but the best AR model—two AR lags and one lag of real output growth—has a higher AIC, residual autocorrelation, and standard error (AIC = 591.4076, Q(16) = 33.0488, and SE = 0.6564) than the MA models and is thus disregarded. ARMA specifications were

Table 3. Estimated MA Specifications of Inflation.

	Parameter Estimates									Estimation Statistics		
	μ_π	β	δ_1	δ_2	δ_3	δ_4	δ_5	γ_1	γ_2	SE	$Q(16)$	AIC
(1)	-0.0009	1.4265	-0.6092	-0.1547	0.1644	-0.1955	-0.0551			0.6515	23.1957	593.5774
	(0.0059)	(1.1328)	(0.0593)	(0.0712)	(0.0665)	(0.0737)	(0.0604)					
(2)	-0.0009	1.4546	-0.6233	-0.1298	0.1494	-0.2417				0.6524	23.9882	592.4073
	(0.0061)	(1.1032)	(0.0571)	(0.0650)	(0.0617)	(0.0522)						
(3)	-0.0426	1.4031	-0.6368	-0.1231	0.1739	-0.2246		9.6946	-4.5600	0.6384	23.1106	579.7605
	(0.0207)	(1.2197)	(0.0582)	(0.0657)	(0.0658)	(0.0538)		(4.4855)	(4.6736)			
(4)	-0.0451	0.9782	-0.6465	-0.1221	0.1751	-0.2315		5.5675		0.6421	23.4296	583.1158
	(0.0190)	(1.0922)	(0.0573)	(0.0661)	(0.0646)	(0.0525)		(2.2659)				

Estimates of different MA specifications of the change in the inflation rate, each of the form $\Delta\pi_t = \mu_\pi + \beta y^c_{t-1} + \delta(L)v_t + \gamma(L)\Delta y_{t-1}$. SE refers to the model standard error, Q(16) refers to the test statistic from a Ljung-Box Q test on 16 lags of the residuals, and AIC refers to the Akaike Information Criterion. The values in parentheses denote standard errors.

also not found to improve on the pure MA specifications. I therefore continue with the following MA(4) specification of the Phillips-curve equation as part of the Kuttner model:

$$\Delta\pi_t = \mu_\pi + \gamma\Delta y_{t-1} + \beta y^c_{t-1} + v_t + \delta_1 v_{t-1} + \delta_2 v_{t-2} + \delta_3 v_{t-3} + \delta_4 v_{t-4}, \tag{8}$$

where $v_t \sim$ i.i.d. $\mathrm{N}(0, \sigma^2_v)$ may be correlated to u_t from (5).

Similar to the Watson model, the Kuttner model can be estimated through the use of a Kalman filter on a state-space representation of the model. In (6) and (7), now $x_t = [y_t \;\; \Delta\pi_t]'$, R is the identity matrix of size 12, and the state vector α_t and system matrices F, Q, and Z are defined as

$$\alpha_t = \begin{bmatrix} y^g_t \\ y^c_t \\ y^c_{t-1} \\ v_t \\ v_{t-1} \\ v_{t-2} \\ v_{t-3} \\ v_{t-4} \\ \mu_y \\ \mu_\pi \\ y_{t-1} \\ y_{t-2} \end{bmatrix}, \quad F = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & \phi_1 & \phi_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \quad Q = \begin{bmatrix} \sigma^2_e & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2_u & 0 & \sigma_{uv} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_{uv} & 0 & \sigma^2_v & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\text{and} \quad Z = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \beta & 1 & \delta_1 & \delta_2 & \delta_3 & \delta_4 & 0 & 1 & \gamma & -\gamma \end{bmatrix}.$$

The prior distribution of the initial state vector and the initial model parameter estimates for running the Kalman filter on the most recent vintage are presented in Table 4. Following Virmani (2004), I set the initial values for the MA terms in the state vector to their expectation (zero) and take the rest of the initial values from OLS estimates of (8).

The multivariate Kuttner model has the advantage of incorporating information contained in other variables related to the output gap. In particular, the output gap resulting from the Kuttner model is the one most consistent with observed inflation, given the specification of the real output processes. Nevertheless, incorporating such information only improves the output gap estimates if the underlying economic theory is valid. Furthermore, a drawback that remains with filtering methods, which was already discussed in Section III.A.1, is the end-of-sample problem.

Table 4. Initialization of Kuttner Model Parameters and State Vector Distribution.

Initial State Vector Distribution												
	y_t^g	y_t^c	y_{t-1}^c	v_t	v_{t-1}	v_{t-2}	v_{t-3}	v_{t-4}	μ_y	μ_π	y_t	y_{t-1}
α_1	7.6130	-0.0004	0.0122	0	0	0	0	0	0.0078	-0.0451	7.6146	7.6173
P_1	0.4441	0.0003	0.0003	0.4123	0.4123	0.4123	0.4123	0.4123	2×10^{-8}	0.0004	0.4445	0.4445
Initial Model Parameters												
	β	δ_1	δ_2	δ_3	δ_4	γ	σ_e	σ_u	σ_v	σ_{uv}	ϕ_1	ϕ_2
	0.9782	-0.6465	-0.1221	0.1751	-0.2315	5.5675	0.0026	0.0077	0.6421	0.0006	1.1643	-0.3699

Initial values for the parameters and for the state vector distribution of the Kuttner model for the most recent vintage.

A.5. Vintage-Based VAR Model

Since data on real output is prone to revisions, it can be expected that real-time estimates of the output gap become more reliable when possible future revisions to the underlying output data are taken into account. Therefore, before applying an output gap estimation method to output data, Clements and Galvão (2012) first forecast how this output data will be revised in the future through VAR models of data vintages. In order to do so, they use output growth rates rather than output levels because the former are likely to be more robust to level changes resulting from benchmark revisions. The growth rate observed at t using data from vintage $t+1$ is calculated as $g_t^{t+1} = 400\,(y_t^{t+1} - y_{t-1}^{t+1})$. Assuming that newly released data is revised over the subsequent $k-1$ quarters, after which it remains unrevised, the observations $t-k+1$ through t in vintage $t+1$ can be modeled by a vintage-based VAR model as follows:

$$\mathbf{g}^{t+1} = \boldsymbol{\mu} + \sum_{i=1}^{l} \mathbf{\Gamma_i}\, \mathbf{g}^{t+1-i} + \boldsymbol{\epsilon}^{t+1}, \tag{9}$$

where $\mathbf{g}^{t+1} = \left[g_t^{t+1}\ g_{t-1}^{t+1} \dots\ g_{t-k+1}^{t+1}\right]'$, $\mathbf{g}^{t+1-i} = \left[g_{t-i}^{t+1-i}\ g_{t-1-i}^{t+1-i} \dots\ g_{t-k+1-i}^{t+1-i}\right]'$, and $\boldsymbol{\mu}$ and $\boldsymbol{\epsilon}^{t+1}$ are $k \times 1$ vectors. Following Clements and Galvão (2012), I set the autoregressive order to $l = 1$ and the number of revisions to $k = 14$, such that the three annual revisions are captured regardless of in which quarter the data was first published. Clements and Galvão (2012) also consider other specifications of the VAR model that take into account that the pattern of revisions to a given observation may depend on the quarter in which the observation was first released. However, these models are not found to outperform the model presented in (9) and are thus not considered here.

The parameters of (9) can be estimated by applying OLS to each individual equation, where the first equation models newly released data, the second equation models data that has been revised once, and so on. Starting with vintage Q4 1969, I estimate the parameters for each vintage based on all the vintages preceding it, in the form of an expanding window. I then use the estimated parameters for each vintage $t+1$ to create h-step ahead forecasts for that vintage. The h-step ahead forecast of $\mathbf{g}^{t+1+h}$ is defined as $\mathbf{g}^{t+1+h|t+1} = \begin{bmatrix} g_{t+h}^{t+1+h|t+1} & g_{t+h-1}^{t+1+h|t+1} & \cdots & g_{t+h-k+1}^{t+1+h|t+1} \end{bmatrix}' \equiv E(\mathbf{g}^{t+1+h} \,|\, \mathbf{g}^{t+1}, \mathbf{g}^{t}, \dots)$. An h-step ahead forecast thus consists of a prediction of the newly released g_{t+h}, a prediction of g_{t+h-1} after its first revision, and so forth until the prediction of $g_{t+h-k+1}$ after its final revision. Similar to Clements and Galvão (2012), I calculate h-step ahead forecasts iteratively for $h = 1$ to $h = 27$ for each vintage. The set of the last elements of each of these h-step ahead forecasts form the post-revision forecast for the vintage: $g_{t+2-k}^{t+2|t+1}$, …, $g_{t}^{t+k|t+1}$, $g_{t+1}^{t+k+1|t+1}$, …, $g_{t+28-k}^{t+28|t+1}$. As a result, a post-revision data set is created that not only contains forecasted post-revision values of the last $k-1$ observations in each vintage of the original data set (remember that the values before this are assumed to remain unrevised), but also forecasted post-revision values of the k future observations for each vintage. This method may therefore not only improve real-time estimates of the output gap by taking into consideration data revisions, but possibly also by decreasing the end-of-sample problem.

After calculating the post-revision output growth rates, I convert them back to the log of real output such that a common output gap estimation method can be applied. To best compare the performance of the vintage-based VAR model with the model that is proposed later on in this paper, the estimation method that is applied is the Watson model. Note that the reason why the Kuttner model is not applied is because it would require either predicting the future inflation rate or shortening the sample period.

B. An Updated Estimation Method for the Output Gap

Forecasting post-revision output data through VAR models of data vintages is not the only way to take into account the effect of data revisions when estimating the output gap. There is a large body of literature on modeling revisions to output (and other macroeconomic) data. While various papers have indeed focused on vintage-based VAR models (e.g. Garratt et al., 2008; Clements and Galvão, 2012), such models assume that newly released data is revised over the subsequent $k-1$ quarters and remains unrevised afterwards. In practice, however, data does not necessarily remain unrevised. Therefore, some papers take a state-space approach by relating published output data to their unobserved true values (e.g. Jacobs and Van Norden, 2011; Cunningham et al., 2012). Similar to the vintage-based VAR approach to estimating the output gap, after estimating the true output values through a state-space model, common output gap estimation methods can be applied in an attempt to improve the real-time estimates of the output gap. Hence, both of these approaches require a two-step procedure of (1) predicting post-revision output data and (2) applying an output gap estimation method to this data.

Rather than a two-step procedure, this paper examines the potential of modeling data revisions to real output jointly with the dynamics of real output, so that the output gap can immediately be extracted. To that end, I propose to combine a state-space representation of the dynamics of true output, such as those by Watson (1986) or Kuttner (1994), with a state-space representation of uncertain output data, such as that by Cunningham et al. (2012). More specifically, I add another unobserved component, namely the true value of output, to the state vector of the Watson and Kuttner models. Along the lines of Cunningham et al. (2012), let y_t^* denote the log of the unobserved true value of real output at time t (i.e. the true value of y_t). This true output is assumed to follow the processes in (3)–(5) of the Watson or Kuttner models. Furthermore, if y_t^{t+n} is the estimate of y_t^* in vintage $t+n$ for $n = 1, ..., T-t$, then

$$y_t^{t+n} = y_t^* - \kappa_n + w_t, \tag{10}$$

where κ_n is the bias in estimates of maturity n and $w_t \sim \mathrm{N}(0, \sigma_w^2)$ is the measurement error in the estimate of y_t^*. The reason for allowing the bias to differ with the maturity of the estimate is that the estimate is assumed to improve in each successive vintage, such that the bias decreases as the estimate matures (Cunningham et al., 2012). In fact, bias is measured as

$$\kappa_n = \kappa_1(1+\tau)^{n-1}, \tag{11}$$

where τ represents the rate at which bias decreases as the estimate matures. Outside the scope of this paper but interesting for future research would be to assume that the measurement error w_t is both serially correlated and heteroskedastic. This is because the measurement errors made in a certain vintage may be correlated with each other, and they may be heteroskedastic (decreasing) with respect to maturity for the same reason that bias decreases with maturity (see Cunningham et al., 2012, for further details).

Assuming that true output y_t^* can be modeled by the Watson (1986) dynamics as in (3)–(5), the joint Watson model can be written in the state-space form of (6) and (7), where $x_t = y_t^T$ represents the estimate of true output in the vintage of interest T. Furthermore, $H = \sigma_w^2$, Q is the identity matrix of size 2, and the state vector α_t and system matrices F, R, η_t, and Z are defined as

$$\alpha_t = \begin{bmatrix} y_t^* \\ y_t^{g,*} \\ y_t^{c,*} \\ y_{t-1}^{c,*} \\ \mu_{y^*} \\ \kappa_{T-t} \end{bmatrix}, \quad F = \begin{bmatrix} 0 & 1 & \phi_1 & \phi_2 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & \phi_1 & \phi_2 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{1+\tau} \end{bmatrix}, \quad R = \begin{bmatrix} \sigma_e & \sigma_u \\ \sigma_e & 0 \\ 0 & \sigma_u \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix},$$

$$\eta_t = \begin{bmatrix} e_t \\ u_t \end{bmatrix}, \quad \text{and} \quad Z = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}.$$

The joint Kuttner model follows similarly and can be found in Appendix A. While the joint models—like the original models—only consider data from the vintage of interest T, the parameters related to bias and measurement errors, i.e. those in (10)–(11), are estimated based on previous vintages as I will describe below. This way, experience from past data revisions can be exploited. Note that this requires the assumption that the parameters related to bias and measurement errors are constant across vintages.

Similar to the original models, the joint models can be estimated by ML through the use of a Kalman filter. I set the prior distribution of the initial state vector and the initial model estimates for running the Kalman filter in the same way as for the original models, but based on an estimated true output series rather than the actual output series. In order to estimate the true output series, the expected revisions to each value in the actual output series need to be determined. As indicated in (10), these revisions are assumed to differ with the maturity of the output estimate. Therefore, to determine the expected revisions, I first calculate the revision over the J quarters subsequent to each output estimate in the data set. I set J to one-third of the number of preceding vintages, such that much of the revision is captured while retaining enough output estimates for the revisions to be representative of the entire data set (Cunningham et al., 2012). Next, for each maturity, I calculate the average revision to output estimates of that maturity. Using these average revisions per maturity, I estimate the bias parameters in (11), which turn out to be $\kappa_1 = 1.1442$ and $\tau = -0.0047$ for the most recent vintage. Finally, I estimate the true output series by adding the expected revision, as calculated using the estimated (11), to the actual output series.

C. Assessing the Performance of the Estimation Methods

To assess the performance of the above-mentioned estimation methods, I use two main criteria as proposed by Camba-Mendez and Rodriguez-Palenzuela (2003). First, an estimation method should give output gap estimates that are reliable in real time such that appropriate policy decisions can be made. It is therefore crucial that they do not change much when the underlying output data is revised. The second criteria is that the real-time estimates should have predictive power over inflation in order to ensure that they have economic content.

To determine the extent to which real-time estimates are reliable and do not change much over time, I use a methodology similar to that of Orphanides and Norden (2002). More specifically, I first apply each estimation method to the output series of the most recent vintage (Q1 2020), which results in a series of "final" estimates of the output gap. Next, I create a series of "real-time" estimates by applying the estimation methods to each quarterly vintage from Q4 1969 (because the vintage-based VAR and joint models require an initial period over which to estimate the parameters for the first vintage) and then constructing a new series comprising the first estimate of the output gap for each point in time. Part of the difference between the "real-time" and "final" estimates is due to data revision and part of it is due to parameter revision. To make this distinction, I construct a series of "quasi-real" estimates whereby the estimate in a given quarter is calculated based on only the observations up to that quarter as they were available in vintage Q1 2020. Hence,

- **Total revision:** difference between "final" and "real-time" estimates
- **Data revision:** difference between "quasi-real" and "real-time" estimates
- **Parameter revision:** difference between "final" and "quasi-real" estimates.

Finally, for estimation methods based on Kalman filtering, the "final" estimates are the Kalman-smoothed estimates of the relevant series while the corresponding Kalman-filtered estimates make up another series, namely the "quasi-final" estimates. For these methods, parameter revision is calculated as the difference between the "quasi-final" and "quasi-real" estimates. The difference between the "quasi-final" and "final" estimates indicates the relevance of new information in estimating the output gap given the existing parameters.

After constructing the different series of estimates, I plot them against each other for analysis. Particular focus is on the correlation between the real-time and final estimates of the different methods. Furthermore, following Camba-Mendez and Rodriguez-Palenzuela (2003), to test whether the real-time and final series have the same statistical properties, I perform the Pesaran and Timmermann (1992) test of directional accuracy—which tests whether the two series have equal signs—and an F-test of equal variances. With regard to the various types of revision, I examine their size and persistence in terms of autocorrelation under the different estimation methods. Since the reliability of real-time output gap estimates is presumably most relevant around turning points of the business cycle (Orphanides and Norden, 2002), I also compare the revisions to these estimates in the three months centered about business cycle peaks as dated by the National Bureau of Economic Research (NBER) across the estimation methods.

To assess real-time output gap estimates' predictive power over inflation, a natural inclination would be to use the Phillips-curve equation in (8) to predict the inflation rate. However, since output is published with a lag of one quarter, y_t^c is not yet available at t to be able to predict the inflation rate at $t+1$. Therefore, I adopt a methodology similar to that of Kamada (2005) instead. That is, I calculate the one-step-ahead prediction of the inflation rate as

$$\pi_{t+1} = \pi_{t+1}^e + \zeta y_{t+1}^{c,e} + \omega_{t+1}, \tag{12}$$

where π_{t+1}^e and $y_{t+1}^{c,e}$ are the expected inflation rate and expected output gap, respectively, and $\omega_t \sim$ i.i.d. $\mathrm{N}(0, \sigma_\omega^2)$. The expected inflation rate and output gap are calculated as

$$\pi_{t+1}^e = \frac{\alpha \sum_{i=0}^{3} \pi_{t-i}}{4} + \frac{(1-\alpha)\sum_{i=4}^{7} \pi_{t-i}}{4} \tag{13}$$

$$y_{t+1}^{c,e} = \sum_{i=0}^{m} \psi_i y_{t-i}^c. \tag{14}$$

Hence, the predicted inflation rate depends on both past inflation rates and past output gaps. Note that since output is published with a lag of one quarter, however, I estimate (12) under the restriction that $\psi_0 = 0$. The model parameters are estimated based on the first half of the sample, such that out-of-sample predictions can be made for the second half of the sample. Since output gap estimates differ per estimation method, the parameters of the above model clearly also differ for each estimation method. This also means that for each estimation method, a different number of lags of output gap estimates in (14) is optimal. Given that this optimal number is found to be between two and four lags (where two lags implies that only one output gap estimate is included, as the first lag has coefficient $\psi_0 = 0$), I examine prediction models with two, three, and four lags of output gap estimates.

After generating the one-step-ahead out-of-sample predictions of the inflation rate based on each output gap estimation method, following Camba-Mendez and Rodriguez-Palenzuela (2003), I compare their performance in terms of root mean square error (RMSE) to that of a random walk with drift and to that of an AR model without past output gaps (i.e. with $\zeta = 0$ in (12)). More specifically, for the random walk with drift, I calculate Theil statistics as the RMSE of the prediction model divided by the RMSE of the random walk. A Theil statistic above 1 therefore indicates that the prediction model has lower prediction accuracy than a random walk. For the simple AR model, which is hard to outperform (Camba-Mendez and Rodriguez-Palenzuela, 2003), I calculate the RMSE of the prediction model minus the RMSE of the AR model. Lastly, I perform Diebold and Mariano (1995) tests of equal predictive ability on the prediction model versus the benchmark models.

IV. Results

This section presents the results of the analyses performed on the output gap estimation methods as described in Section III. First, I discuss some general results from the different methods and how they compare to each other. Then, I analyze revisions to real-time estimates from the different methods and the predictive power of these estimates over inflation. Lastly, attention is given to the parameter (in)stability of the different estimation methods. Note that, to make the output gap estimates from different estimation methods comparable, they are expressed as a percentage of potential output.

Before comparing the output gap estimates from the different estimation methods, Figure 7 in Appendix B plots the trend component of the log of real output for the most recent vintage under several methods, zoomed in on different time periods. It appears that the HP filter calculates a rather smooth trend, whereas the BN decomposition more closely follows the actual output series. Furthermore, the potential output resulting from both the Watson and especially the Kuttner model stays relatively stable when actual output goes down during for instance the financial crisis of 2008.

A. Comparison of the Estimation Methods

Figure 2 plots the real-time and final estimates of the output gap from the different estimation methods. The difference between these series of estimates, or the total revision, is also plotted. The shaded periods indicate recessions as defined by the NBER. The resulting plots are comparable

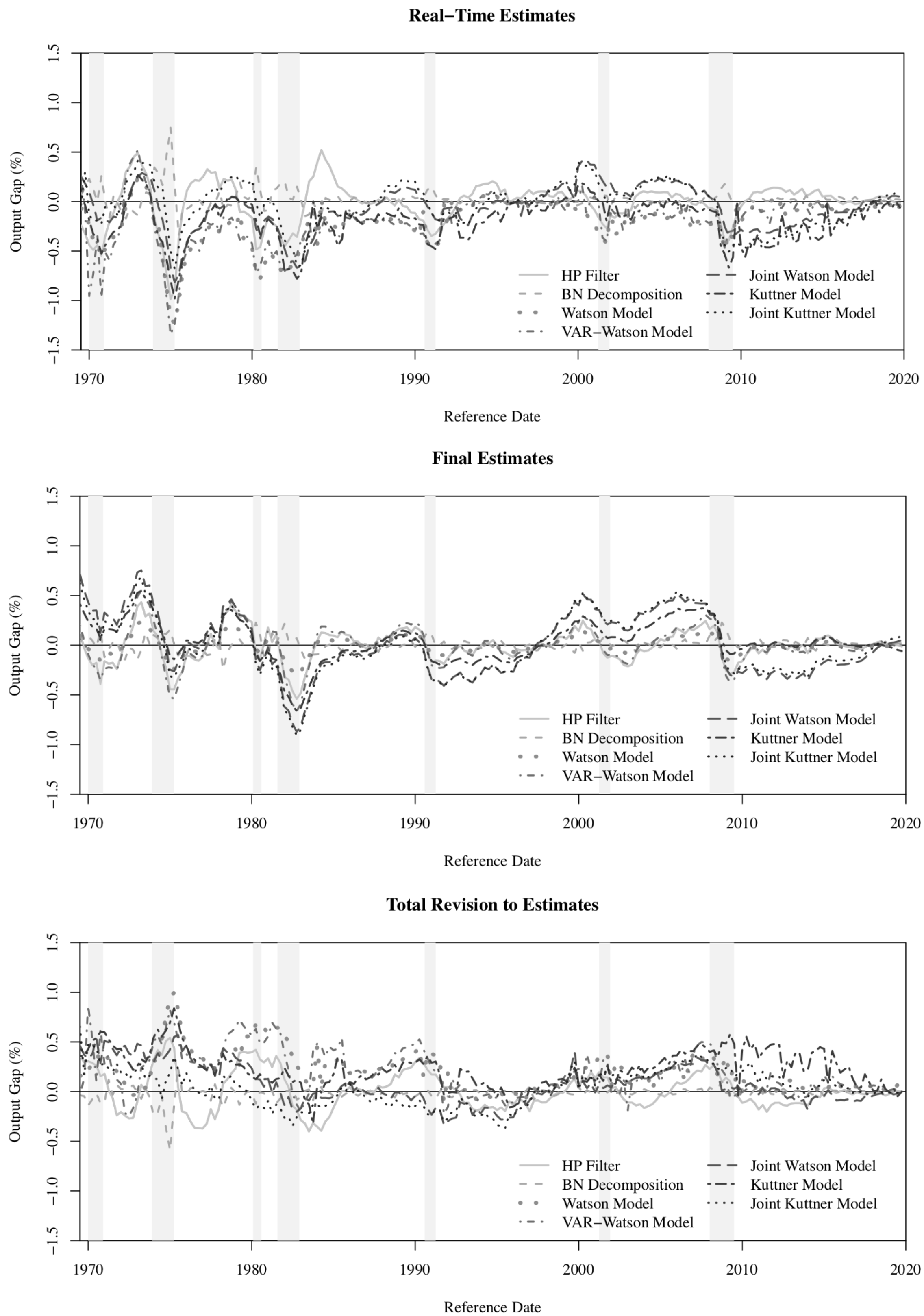

Figure 2. Output Gap Estimates and Their Revisions. Plots of final and real-time estimates of the output gap and their total revisions under the different estimation methods. The shaded periods indicate NBER recessions.

to those by Orphanides and Norden (2002). The estimation methods tend to move upwards or downwards together, although more closely so for the final estimates, and generally indicate a negative output gap around recessions as should be expected. Only the BN decomposition results in rather different measures of the output gap. In fact, this method returns positive output gap estimates around recessions, whereas the estimates during non-recession periods almost seem like

Table 5. Summary Statistics for the Output Gap.

	Mean	SD	Min	Median	Max	COR
HP filter						
Final	-0.0047	0.1635	-0.5408	-0.0038	0.4322	1.0000
Quasi real	-0.0108	0.1711	-0.4540	0.0196	0.4308	0.5481
Real time	-0.0150	0.2226	-0.9854	0.0280	0.5213	0.5379
BN decomposition						
Final	0.0075	0.0636	-0.2428	0.0080	0.2271	1.0000
Quasi real	0.0136	0.0588	-0.2125	0.0131	0.2243	0.9909
Real time	0.0249	0.0985	-0.3439	0.0159	0.7465	0.7226
Watson model						
Final	-0.0006	0.0889	-0.3028	-0.0056	0.2375	1.0000
Quasi final	-0.0224	0.0720	-0.2358	-0.0223	0.1310	0.7517
Quasi real	-0.1986	0.1575	-0.6168	-0.1839	0.1791	0.4356
Real time	-0.2048	0.2276	-1.2318	-0.1548	0.2406	0.4578
VAR-Watson model						
Final	-0.0014	0.1954	-0.6637	-0.0066	0.5486	1.0000
Quasi final	-0.0334	0.1315	-0.4266	-0.0279	0.2536	0.7434
Quasi real	-0.1729	0.1602	-0.6138	-0.1597	0.1792	0.5165
Real time	-0.1943	0.2589	-1.3449	-0.1412	0.5201	0.5091
Joint Watson model						
Final	0.0079	0.3139	-0.8717	-0.0245	0.7540	1.0000
Quasi final	-0.0884	0.2063	-0.6965	-0.1066	0.3607	0.8424
Quasi real	-0.0211	0.1860	-0.5913	0.0039	0.3836	0.7590
Real time	-0.0674	0.2203	-0.8132	-0.0581	0.4266	0.7655
Kuttner model						
Final	0.0410	0.2037	-0.6627	0.0233	0.5469	1.0000
Quasi final	-0.0261	0.1611	-0.4171	-0.0341	0.3514	0.6629
Quasi real	-0.1622	0.1724	-0.7150	-0.1408	0.1983	0.5966
Real time	-0.1929	0.2114	-0.9784	-0.1578	0.2685	0.5939
Joint Kuttner model						
Final	-0.0005	0.3080	-0.9081	-0.0464	0.6782	1.0000
Quasi final	-0.0007	0.2342	-0.6825	0.0080	0.4864	0.9385
Quasi real	-0.0215	0.2067	-0.5688	0.0095	0.3596	0.7353
Real time	-0.0369	0.2302	-0.6816	-0.0056	0.4248	0.8033

Basic summary statistics for estimates of the output gap from the different estimation methods. The final column reports correlation with the final estimate under that method.

noise. While the estimates during recessions are later revised downward, their final estimates are also still positive. The real-time estimates around recessions of the other estimation methods tend to be revised upward. In other words, their initial estimates seem to have been too negative. This can be explained by the end-of-sample problem, as the trend is more responsive to temporary shocks to output at the end of the sample than in the middle of the sample.

Table 5 reports summary statistics for estimates of the output gap from the different estimation methods, including their correlation with the final estimate. The numbers confirm what Figure 2 has already suggested: with a standard deviation between 0.05 and 0.10, the estimates from the BN decomposition are less volatile than those from the other estimation methods. The real-time estimates from the HP filter and Watson model seem to have relatively poor reliability, in that they have a correlation of only 0.54 and 0.46, respectively, with their final estimates. The VAR-Watson model performs better than the original Watson model on that measure, but with a correlation of 0.77, the joint Watson model performs even better. The joint Kuttner model also produces real-times estimates that correlate considerably more with their final estimates than do those of its original counterpart. Further interesting to note is the extremely high correlation of 0.99 between the quasi-real and final estimates from the BN decomposition, which can be readily explained from the fact that the BN decomposition does not use future observations. The quasi-real estimates are constructed from the same vintage as the final estimates, but use only the observations up to the relevant quarter. Since later observations would not be used by the BN decomposition anyway, the quasi-real and final estimates are essentially the same.

To better show the correlation between the different series of estimates from each estimation method, Figure 3 plots the series for each method separately. Apart from the noisy estimates from the BN decomposition, it is clear that the series of estimates from the joint Watson model and especially the joint Kuttner model move most closely together. These are already some promising results for the usefulness of these joint models, but it gets even better when looking at the reliability statistics presented in Table 6. Not only do the real-time estimates from the joint Watson and Kuttner models have a higher correlation with their final estimates than do those from their original counterparts, the proportion of times that they have opposite signs is also much lower (0.16 and 0.12 for the joint Watson and Kuttner models, respectively, versus 0.44 and 0.46 for the original Watson and Kuttner models). While the Pesaran and Timmermann (1992) test rejects for all estimation methods except for the Watson model that the signs of the real-time and final estimates are independent, it does so by far most strongly for the two joint models. The reported NS and NSR are the standard deviation and root mean square, respectively, of the total revision divided by the standard deviation of the final estimate. These statistics serve as proxies for the real-time estimates' noise-to-signal ratio (Orphanides and Norden, 2002) and are relatively low for the two joint models, suggesting a lower noise component. When it comes to the F-test of equal variances, the hypothesis that the real-time and final estimates have equal variances is rejected for all estimation methods except for the original Kuttner model. Interestingly, however, the real-time estimates of the joint models have a lower variance than their final estimates (as indicated by the

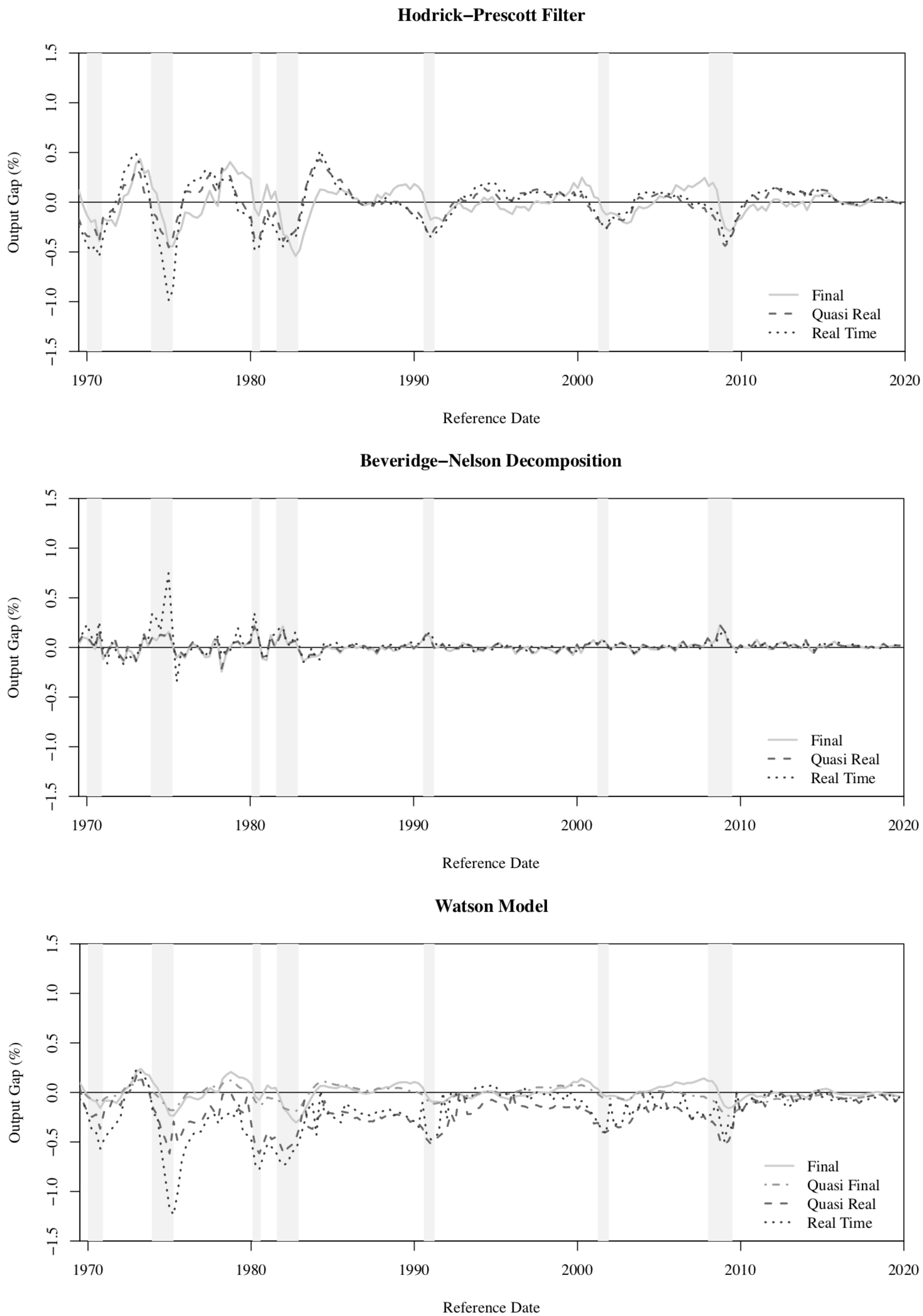

Figure 3. Estimates of the Output Gap by Method. Plots of the different series of estimates of the output gap from the different estimation methods. The shaded periods indicate NBER recessions. Figure continues on Pages 21–22.

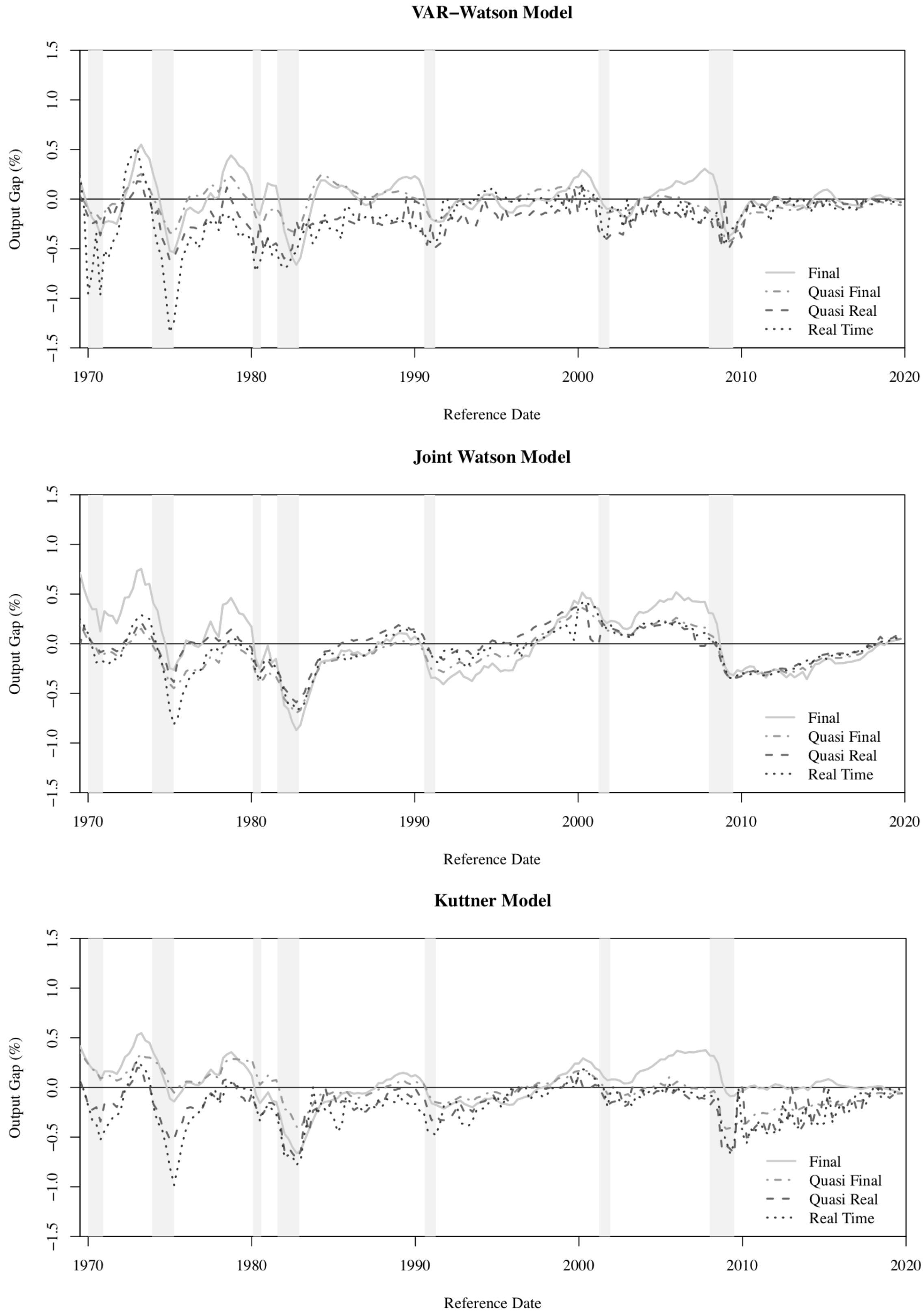

Figure 3. Estimates of the Output Gap by Method. Plots of the different series of estimates of the output gap from the different estimation methods. The shaded periods indicate NBER recessions. Figure continued from Page 20 and continues on Page 22.

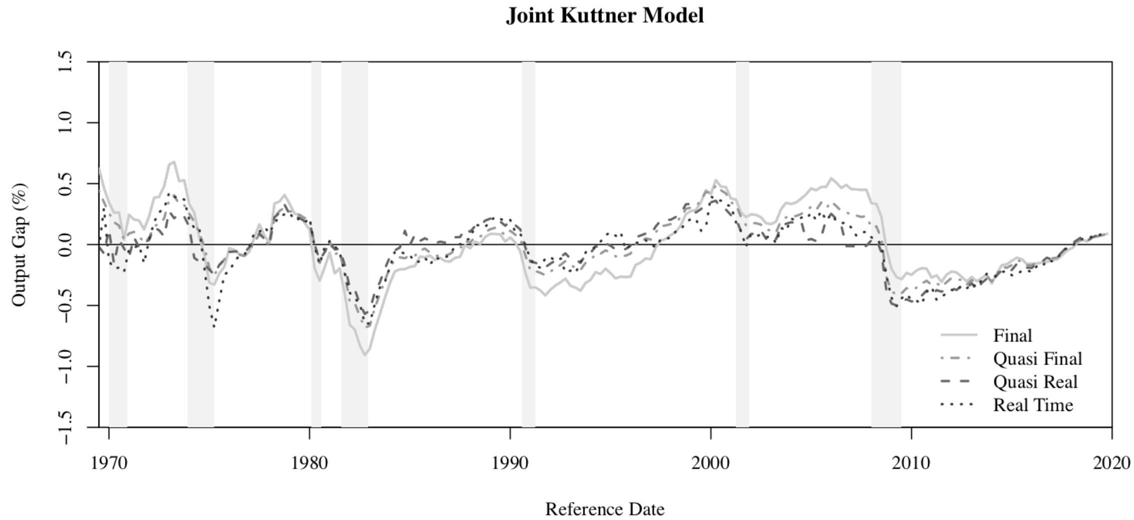

Figure 3. Estimates of the Output Gap by Method. Plots of the different series of estimates of the output gap from the different estimation methods. The shaded periods indicate NBER recessions. Figure continued from Pages 20–21.

F-statistic below 1) while this is the other way around for the other estimation methods. A final observation is that in contrast to the findings by Orphanides and Norden (2002), the estimates from both the original and the joint Kuttner model are more reliable across all reported statistics than those of the corresponding Watson models. This indicates that incorporating economic information in the form of a Phillips curve is useful.

Table 6. Reliability Statistics for the Estimation Methods.

	COR	NS	NSR	OPSIGN	PT	p-value	F	p-value
HP filter	0.5379	1.1783	1.1770	0.3861	3.3460	0.0004	1.8527	0.0000
BN decomposition	0.7226	1.0785	1.1100	0.2376	7.2598	0.0000	2.4037	0.0000
Watson model	0.4578	2.2840	3.2357	0.4406	1.4455	0.0742	6.5619	0.0000
VAR-Watson model	0.5091	1.1858	1.5405	0.4208	2.7465	0.0030	1.7548	0.0001
Joint Watson model	0.7655	0.6465	0.6882	0.1584	9.7767	0.0000	0.4928	0.0000
Kuttner model	0.5939	0.9189	1.4693	0.4554	3.9201	0.0000	1.0771	0.5989
Joint Kuttner model	0.8033	0.5982	0.6083	0.1188	10.8528	0.0000	0.5585	0.0000

Reliability statistics for the different estimation methods. COR is the correlation between the real-time and final estimates. NS is the standard deviation of the total revision divided by that of the final estimate. NSR is the root mean square of the total revision divided by the standard deviation of the final estimate. OPSIGN is the proportion of times that the real-time and final estimates have opposite signs. PT is the test statistic from the Pesaran and Timmermann (1992) test of directional accuracy between the real-time and final estimates. F is the test statistic from a test of equal variances of the real-time and final estimates.

B. Revisions to Real-Time Estimates

Since it is crucial that real-time estimates do not change much when the underlying output data is revised in order for appropriate policy decisions to be made, I analyze the revisions to real-time estimates from the different estimation methods. Table 7 reports summary statistics for the total, data, and parameter revisions to real-time estimates of the output gap from the different estimation

Table 7. Summary Statistics for the Revisions to Real-Time Estimates.

	Mean	SD	RMS	Min	Median	Max	AR
HP filter							
Total revision	0.0103	0.1927	0.1925	-0.4016	-0.0069	0.5423	0.9212
Data revision	0.0043	0.0915	0.0914	-0.2348	-0.0016	0.5314	0.7882
Parameter revision	0.0060	0.1592	0.1589	-0.4105	-0.0165	0.3606	0.9640
BN decomposition							
Total revision	-0.0174	0.0686	0.0706	-0.5906	-0.0098	0.2415	0.3664
Data revision	-0.0112	0.0678	0.0685	-0.5754	-0.0047	0.2426	0.3876
Parameter revision	-0.0061	0.0095	0.0113	-0.0438	-0.0051	0.0306	0.5429
Watson model							
Total revision	0.2042	0.2030	0.2875	-0.1050	0.1857	0.9958	0.8627
Data revision	0.0062	0.1522	0.1520	-0.4108	-0.0034	0.9376	0.6615
Parameter revision	0.1763	0.1388	0.2241	-0.1835	0.1881	0.4788	0.8186
VAR-Watson model							
Total revision	0.1928	0.2317	0.3010	-0.2430	0.1547	0.8632	0.8316
Data revision	0.0214	0.1880	0.1887	-0.4263	0.0000	0.7796	0.5990
Parameter revision	0.1395	0.1476	0.2029	-0.3934	0.1611	0.4696	0.6813
Joint Watson model							
Total revision	0.0753	0.2029	0.2160	-0.3360	0.0296	0.5709	0.9234
Data revision	0.0463	0.1138	0.1226	-0.3593	0.0279	0.5856	0.8181
Parameter revision	-0.0673	0.0894	0.1117	-0.2754	-0.0672	0.2725	0.8565
Kuttner model							
Total Revision	0.2339	0.1872	0.2993	-0.3248	0.2219	0.8373	0.8385
Data revision	0.0308	0.1166	0.1203	-0.2326	0.0172	0.4647	0.4533
Parameter revision	0.1361	0.1429	0.1971	-0.3694	0.1102	0.5657	0.7355
Joint Kuttner model							
Total revision	0.0364	0.1842	0.1873	-0.3678	0.0355	0.6554	0.8761
Data revision	0.0154	0.1019	0.1028	-0.3710	0.0117	0.4333	0.5930
Parameter revision	0.0209	0.1057	0.1074	-0.2250	0.0018	0.4615	0.8332

Basic summary statistics for the total, data, and parameter revisions to real-time estimates of the output gap from the different estimation methods. Total revision is the difference between "final" and "real-time" estimates, data revision is the difference between "quasi-real" and "real-time" estimates, and parameter revision is the difference between "final" (or "quasi-final" for estimation methods based on Kalman filtering) and "quasi-real" estimates. RMS and AR are the root mean square and first-order autocorrelation, respectively, of the revision series.

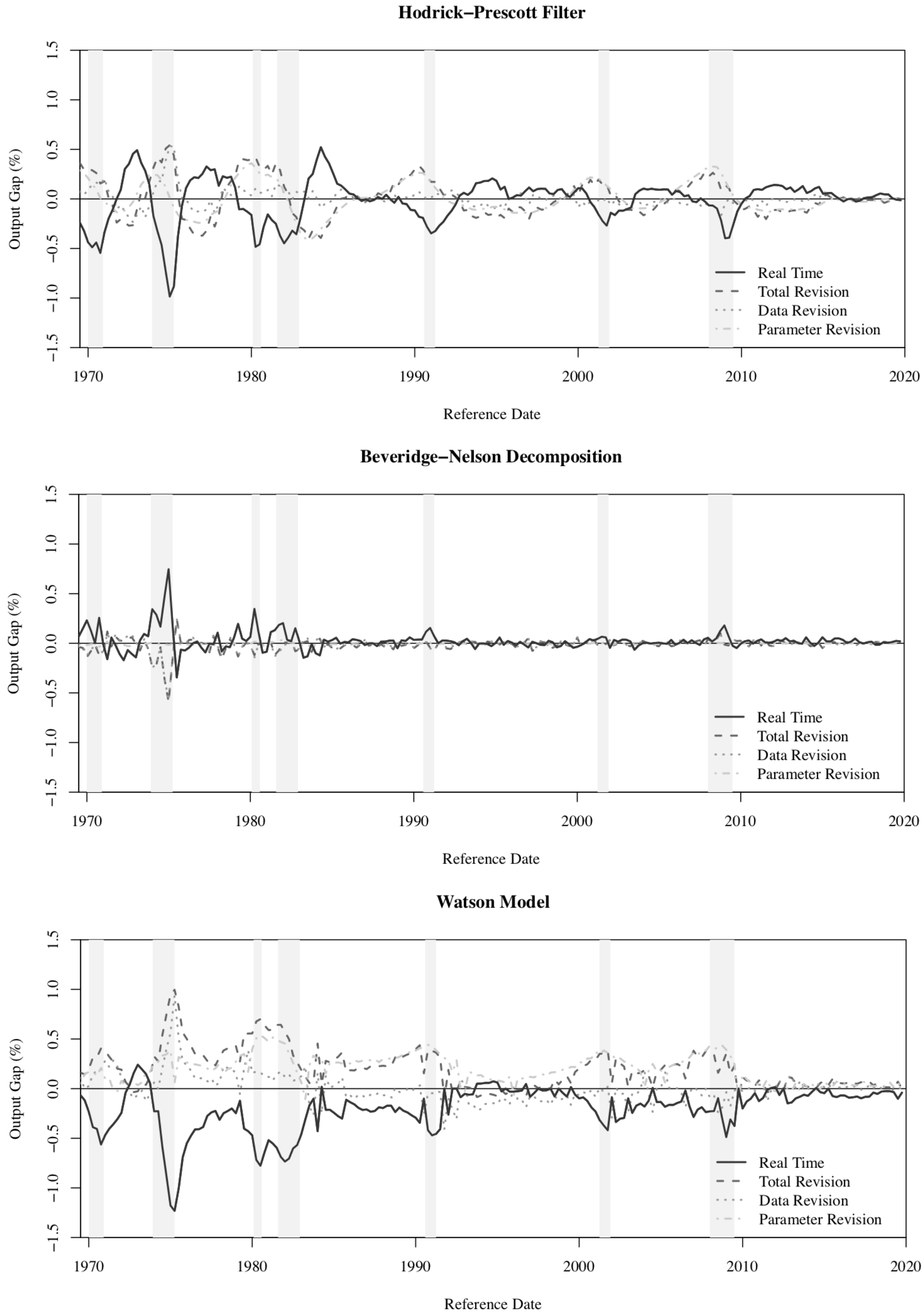

Figure 4. Revisions to Real-Time Estimates by Method. Plots of real-time estimates of the output gap from the different estimation methods and their total, data, and parameter revisions. Figure continues on Pages 25–26.

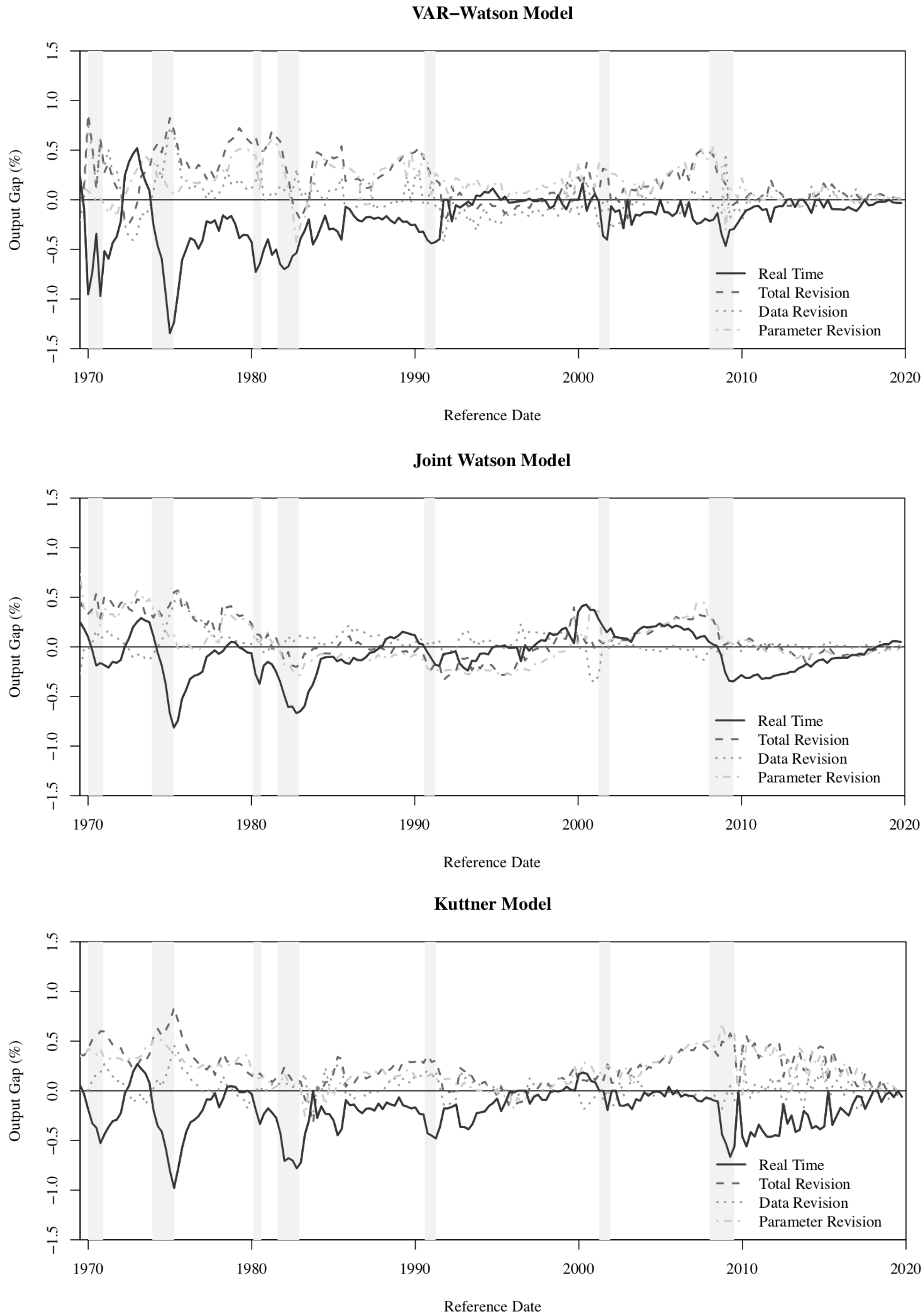

Figure 4. Revisions to Real-Time Estimates by Method. Plots of real-time estimates of the output gap from the different estimation methods and their total, data, and parameter revisions. The shaded periods indicate NBER recessions. Figure continued from Page 24 and continues on Page 26.

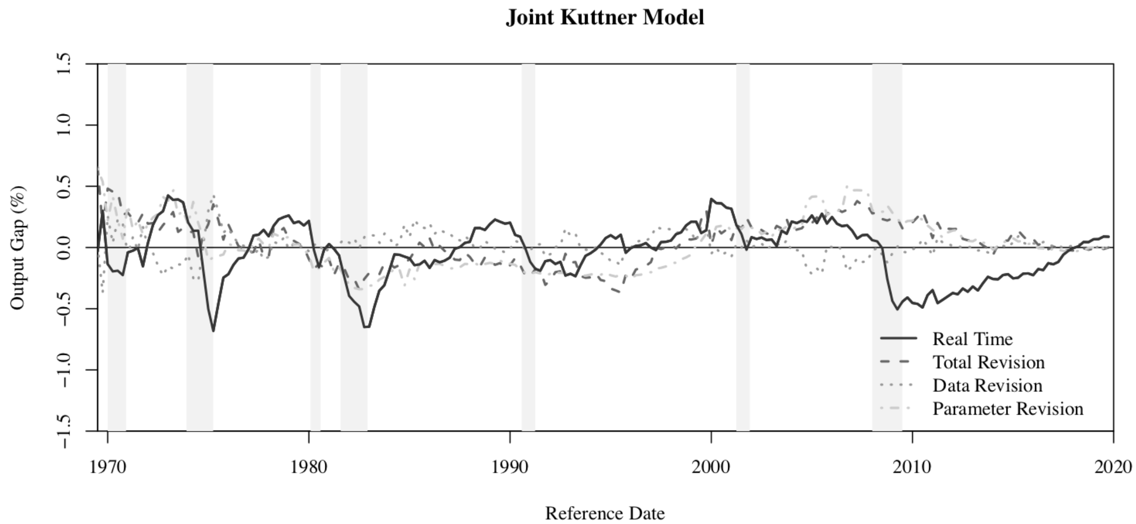

Figure 4. Revisions to Real-Time Estimates by Method. Plots of real-time estimates of the output gap from the different estimation methods and their total, data, and parameter revisions. The shaded periods indicate NBER recessions. Figure continued from Pages 24–25.

methods, including the root mean square (RMS) and first-order autocorrelation (AR), and Figure 4 plots these revisions for each method separately.

On average, except for those of the BN decomposition, the revisions are positive and highly persistent with many of the autocorrelations above 0.80. For the HP filter, VAR-Watson model, and original Watson and Kuttner models, negative output gaps are often revised upward with almost the same magnitude, such that the final estimates are close to zero. The revisions for the joint Watson and Kuttner models show much less of a pattern relative to the real-time estimates. When decomposing the total revision, it becomes clear that only the BN decomposition has a relatively high average data revision of -0.011 percent compared to an average total revision of -0.017 percent. While data revision does play a role for the other estimation methods, similar to the findings by Orphanides and Norden (2002), it is rather small and less variable than the total revision. Instead, most of the revision is attributable to parameter instability of the methods, which will be elaborated on later in this section.

Since the reliability of real-time output gap estimates is presumably most relevant around turning points of the business cycle (Orphanides and Norden, 2002), Table 8 provides the revisions to these estimates in the three months centered about NBER business cycle peaks. Except for those of the BN decomposition, the real-time estimates around peaks of all methods are underestimated (on average more so than the real-time estimates over the entire sample period) and later revised upward. With an average total revision of 0.42 percent around peaks, the VAR-Watson model underestimates the output gap most severely—which may be due to underestimation of post-revision actual output around business cycle peaks. The underestimation by the joint Watson model and especially the joint Kuttner model is least severe, with average total revisions of 0.17 and 0.09 percent, respectively.

Table 8. Summary Statistics for the Revisions to Real-Time Estimates around Peaks.

	Mean	SD	RMS	Min	Median	Max
HP filter	0.2682	0.0990	0.2850	0.0076	0.2905	0.4208
BN decomposition	-0.0340	0.0713	0.0774	-0.2402	-0.0197	0.0458
Watson model	0.3618	0.1909	0.4070	0.0387	0.3453	0.6774
VAR-Watson model	0.4236	0.2344	0.4814	-0.0383	0.4957	0.8632
Joint Watson model	0.1737	0.1759	0.2443	-0.1691	0.2348	0.4616
Kuttner model	0.2802	0.1396	0.3116	0.0875	0.2966	0.5445
Joint Kuttner model	0.0875	0.2428	0.2526	-0.2257	0.1278	0.6554

Basic summary statistics for the total revisions to real-time estimates of the output gap in the three months centered about NBER peaks from the different estimation methods. RMS is the root mean square of the revision series.

C. Predicting Inflation with Real-Time Estimates

To examine whether the real-time estimates from the different estimation methods are not only reliable but also have economic content, I test their predictive power over inflation. Table 9 presents statistics on the accuracy of one-step-ahead out-of-sample predictions of the inflation rate, using varying lags of output gap estimates from the different estimation methods. The RMSE of the prediction models including lags of the output gap is smaller than that of a random walk with drift, as indicated by the Theil statistics below 1. According to the Diebold and Mariano (1995) test, however, the hypothesis that the prediction accuracy is the same as that of a random walk cannot be rejected for any of the models. Nevertheless, regardless of the number of output lags included in the prediction model, those including output gap estimates from the joint Watson and Kuttner models have the lowest Theil statistics (0.82 and 0.83, respectively, for two output lags) and hence the best prediction accuracy.

Table 9. Inflation Prediction Accuracy of Real-Time Output Gap Estimates.

	2 Output Gap Lags				3 Output Gap Lags				4 Output Gap Lags			
	Theil	D-M	RRMSE	D-M	Theil	D-M	RRMSE	D-M	Theil	D-M	RRMSE	D-M
HP filter	0.8597	0.2785	0.0224	0.0452	0.8589	0.2760	0.0218	0.0479	0.8588	0.2757	0.0217	0.0394
BN decomposition	0.8759	0.3383	0.0364	0.0014	0.8689	0.3136	0.0304	0.0008	0.8705	0.3192	0.0317	0.0008
Watson model	0.8366	0.2086	0.0026	0.8762	0.8267	0.1881	-0.0060	0.0021	0.8297	0.2001	-0.0033	0.4570
VAR-Watson model	0.8413	0.2209	0.0066	0.6495	0.8450	0.2286	0.0098	0.5107	0.8310	0.1993	-0.0022	0.2187
Joint Watson model	0.8171	0.1677	-0.0142	0.3910	0.8163	0.1659	-0.0149	0.3653	0.8155	0.1646	-0.0156	0.3378
Kuttner model	0.8617	0.2899	0.0242	0.2866	0.8548	0.2655	0.0182	0.3983	0.8583	0.2771	0.0212	0.3444
Joint Kuttner model	0.8300	0.2006	-0.0031	0.8768	0.8285	0.1966	-0.0043	0.8301	0.8197	0.1773	-0.0120	0.5350

Accuracy of one-step-ahead out-of-sample predictions of the inflation rate, using varying lags of output gap estimates. Theil statistics are the root mean squared error (RMSE) of a prediction model divided by that of a random walk, and RRMSE denotes the relative RMSE of a prediction model minus that of an AR model without lags of the output gap. The D-M columns contain p-values corresponding to the Diebold and Mariano (1995) test of equal predictive ability.

When an AR model without lags of the output gap is used as a benchmark, the prediction models that do include output gap estimates from either of the joint models still outperform. The RMSE of the prediction models including two output lags from the joint Watson model and joint Kuttner model are lower by 0.014 and 0.003, respectively, than the RMSE of the AR benchmark model. This differs from the results of Kamada (2005), who finds that none of the prediction models including output gap estimates outperform a simple AR model. The outperformance reported here is not statistically significant based on the Diebold and Mariano (1995) test, except for that of the prediction model including three output gap lags from the Watson model. Nonetheless, given that a simple AR model is hard to outperform (Camba-Mendez and Rodriguez-Palenzuela, 2003), the results reported here are rather encouraging.

For illustration, Figure 5 plots the one-step-ahead in-sample (before split) and out-of-sample (after split) predictions of the inflation rate using an AR model with two lags of output gap estimates from the joint Watson model. Clearly, the predictions closely follow those of an AR model without lags of the output gap. The predictions based on the other estimation methods are very similar and shown in Figure 8 in Appendix C.

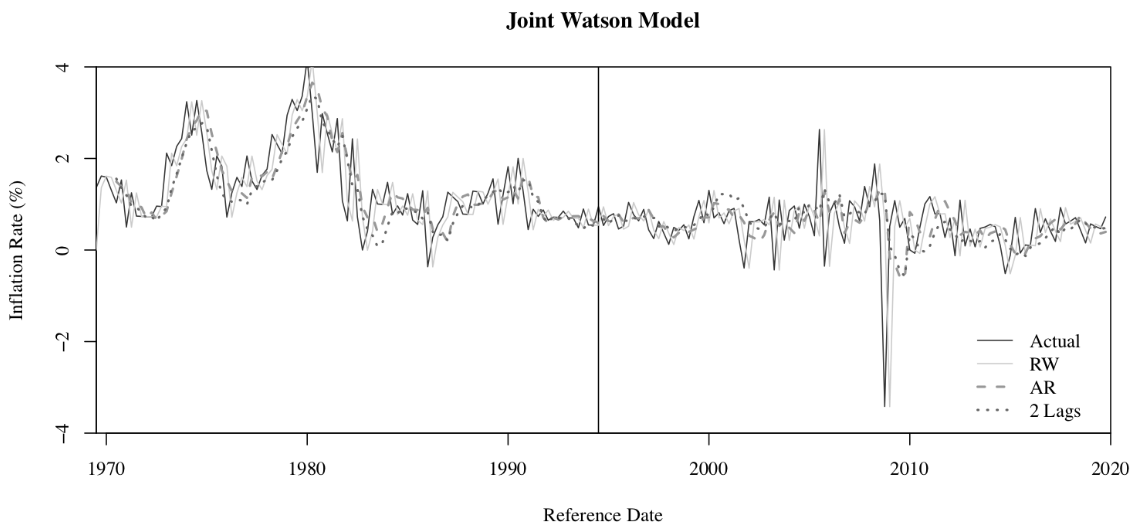

Figure 5. Inflation Predictions Based on the Joint Watson Model. A plot of one-step-ahead in-sample (before split) and out-of-sample (after split) predictions of the inflation rate—using a random walk, an AR model without lags of the output gap, and an AR model with two lags of output gap estimates from the joint Watson model.

D. Parameter Stability of the Estimation Methods

Since Orphanides and Norden (2002) suggest that most of the revision to real-time estimates is attributable to parameter instability of the estimation methods rather than to data revisions, it is interesting to address and compare the parameter stability of the different estimation methods. Some reasons why parameters change are that the underlying data on which they are estimated is revised or that new data is added. To single out the effect of the latter, Figure 6 plots the

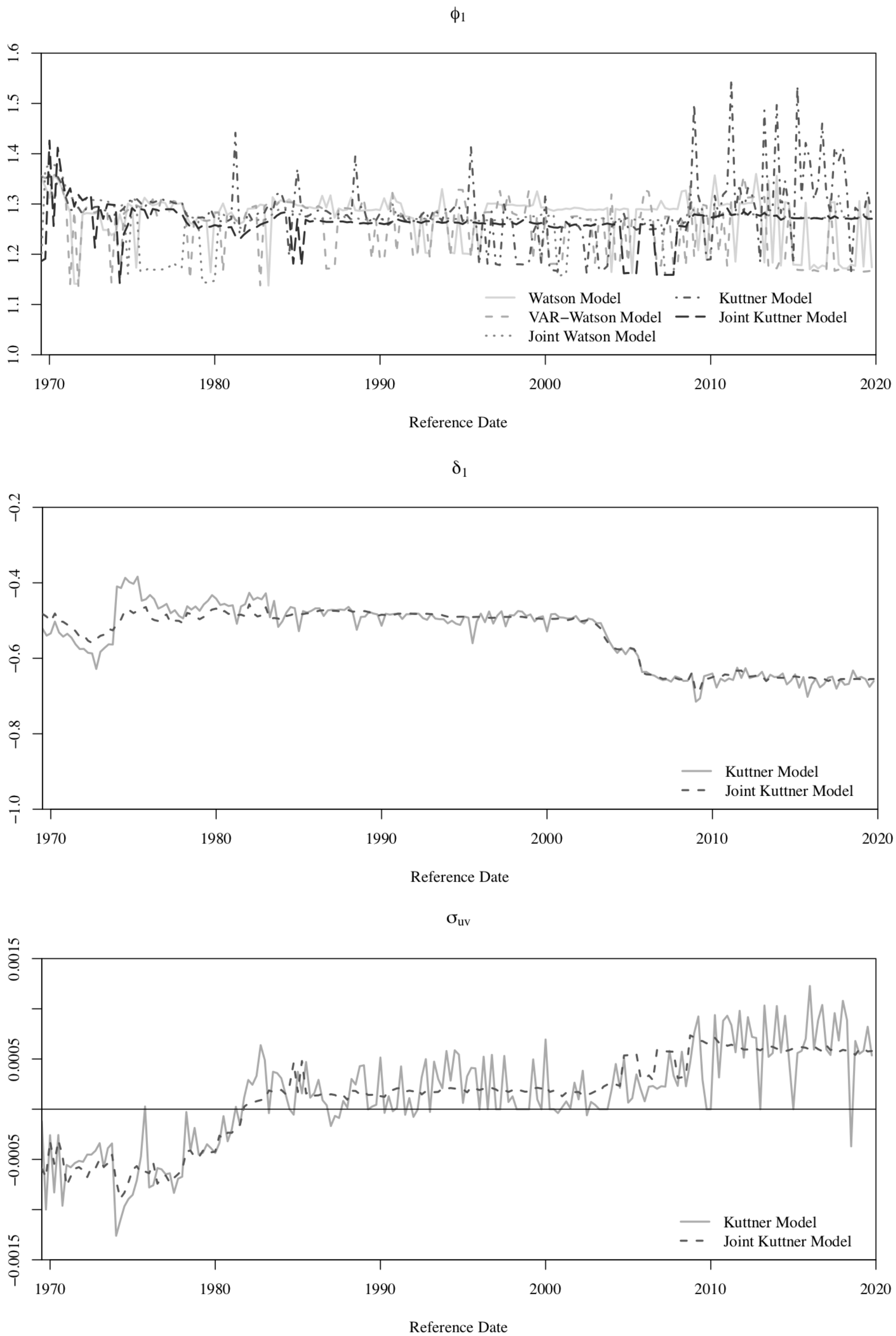

Figure 6. Parameter Estimates Within Final Vintage. Plots of maximum likelihood estimates of some of the parameters of the output gap estimation methods. Parameters for a particular reference date are estimated based on the data up to that date as available in the final vintage. Hence, these are the parameters used to construct the "quasi-real" output gap estimates. Plots of the other parameters are in Appendix D.

maximum likelihood estimates of some of the parameters for each reference date based on the data up to that date as available in the final vintage (plots for the other parameters can be found in Appendix D). Hence, at each successive reference date, one data point is added, but the existing data does not change. Note that these are the same parameters used to construct the "quasi-real" output gap estimates at each reference date.

It is clear that especially the parameters of the joint Kuttner model are more stable than those of its original counterpart. That is, the parameters of the joint Kuttner model follow the same trend as those of the original Kuttner model, but do so in a much smoother fashion. A likely explanation for this finding is that the parameters of the joint model are estimated based on a "true" output series for which, unlike for the actual output series used in the original model, new data added at each successive vintage contains less bias. Table 10 reports the means and standard deviations of the parameters. While the plots are more obvious, the numbers also indicate that the parameters of the joint models generally have a lower standard deviation than those of the original models. For instance, the standard deviation of ϕ_1 is 0.06 under the original Kuttner model but only 0.03 under the joint Kuttner model. For completion, Table 11 and Figure 10 in Appendix E also report the parameter estimates across vintages, which not only change because new data is added but also because the underlying data is revised. Clearly, the parameters are more unstable across vintages than within a vintage, but those of the joint Kuttner model are still less volatile than those of the original Kuttner model.

Table 10. Parameter Estimates Within Final Vintage.

	Watson Model		VAR-Watson Model		Joint Watson Model		Kuttner Model		Joint Kuttner Model	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
ϕ_1	1.2736	0.0484	1.2496	0.0578	1.2664	0.0417	1.2846	0.0621	1.2634	0.0345
ϕ_2	-0.3325	0.0285	-0.3342	0.0285	-0.3437	0.0319	-0.3647	0.0559	-0.3438	0.0317
β							1.1438	0.4593	1.0926	0.4428
δ_1							-0.5390	0.0831	-0.5402	0.0736
δ_2							-0.1773	0.0698	-0.1823	0.0657
δ_3							0.2731	0.0780	0.2894	0.0691
δ_4							-0.3731	0.1163	-0.3758	0.1200
γ							8.8523	1.7731	9.4755	1.5727
σ_{uv}							0.0001	0.0005	0.5847	0.0004

Means and standard deviations of maximum likelihood estimates of the parameters of the output gap estimation methods. Parameters for a particular reference date are estimated based on the data up to that date as available in the final vintage. Hence, these are the parameters used to construct the "quasi-real" output gap estimates.

Even though the parameters of the joint Kuttner model are less volatile than those of the original model, they still change quite a bit when new data is added. Sometimes, this may be because the optimization method gets stuck in a local minimum. Indeed, because of the large number of parameters, the quasi-Newton method used in this research may not adequately isolate the global minimum. Future research may therefore examine whether using techniques like simulated annealing lead to more stable parameters (Planas and Rossi, 2000). Besides this, however, output data series that span different periods often require different parameters or even different model specifications due to changing economic relations. For example, Section III.A.4 already showed that the original Kuttner (1994) specification of the Phillips-curve equation is not the best fit for the more recent data used in this paper. Therefore, a natural next step for future research is to allow the parameters of the proposed estimation methods to vary over time (e.g. Kara et al., 2007).

V. Conclusion

This paper has reviewed several existing estimation methods for the output gap and proposed an updated method that models the dynamics of real output jointly with data revisions to real output. The widely used Hodrick-Prescott (HP) filter is found to produce real-time output gap estimates with relatively poor reliability, which is likely due to the end-of-sample problem, whereas the estimates from the Beveridge-Nelson (BN) decomposition seem to perform quite fine according to the used performance criteria but do not contain much economic information. Defining potential output as the long-run forecast of output, like the BN decomposition does, therefore does not seem to be appropriate. Applying the Watson model to a data set that contains post-revision output data as forecasted by a VAR model of past data vintages, rather than to the original data set, improves real-time output gap estimates. Yet, the joint Watson model proposed in this paper that combines the Watson model with a state-space model that relates published output data to their unobserved true values performs even better. I also find that incorporating economic information in the form of a Phillips curve is useful, as the real-time estimates from both the original and the joint Kuttner model are more reliable than those of the corresponding Watson models.

The outperformance of the joint models proposed in this paper is due not only to the fact that they take into account data revisions, but also to their increased parameter stability. Rather than using the Watson or Kuttner model as a base, the joint model could also be implemented with more complex unobserved-components models of output, such as that by Alichi (2015). The results in this paper certainly provide a promising basis for doing so. Nevertheless, a few limitations and suggestions for future research are due to be mentioned. First of all, the model parameters have been assumed to be constant over time, but changing economic relations often call for different parameters or even different model specifications. Therefore, a natural next step for future research is to allow the parameters to vary over time. Also interesting for future research would be to assume that the measurement error in published output data is both serially correlated and heteroskedastic as suggested by Cunningham et al. (2012). This is because the measurement errors made in a certain vintage may be correlated with each other, and they may be heteroskedastic (decreasing) with respect to maturity. Another way to allow for richer dynamics in the measurement error would be to examine to what extent the measurement error should be seen as "noise", "news", or "spillover" and model it as such, as in Jacobs and Van Norden (2011). Such model enhancements can be expected to improve upon the already promising results in this paper even more.

Appendix A. State-Space Form of Joint Kuttner Model

As proposed in Section III.B, data revisions to real output and the Kuttner (1994) dynamics of real output can be jointly modeled in the state-space form of (6) and (7), where $x_t = \begin{bmatrix} y_t^T & \Delta\pi_t \end{bmatrix}'$, Q is the identity matrix of size 3, and the state vector α_t and system matrices F, R, η_t, Z, and H are defined as

$$\alpha_t = \begin{bmatrix} y_t^* \\ y_t^{g,*} \\ y_t^{c,*} \\ y_{t-1}^{c,*} \\ v_t \\ v_{t-1} \\ v_{t-2} \\ v_{t-3} \\ v_{t-4} \\ \mu_{y^*} \\ \mu_\pi \\ y_{t-1}^* \\ y_{t-2}^* \\ \kappa_{T-t} \end{bmatrix}, \quad F = \begin{bmatrix} 0 & 1 & \phi_1 & \phi_2 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \phi_1 & \phi_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{1+\tau} \end{bmatrix}, \quad R = \begin{bmatrix} \sigma_e & r_{uu} & r_{uv} \\ \sigma_e & 0 & 0 \\ 0 & r_{uu} & r_{uv} \\ 0 & 0 & 0 \\ 0 & r_{uv} & r_{vv} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$\eta_t = \begin{bmatrix} e_t \\ u_t \\ v_t \end{bmatrix}, \quad Z = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & \beta & 1 & \delta_1 & \delta_2 & \delta_3 & \delta_4 & 0 & 1 & \gamma & -\gamma & 0 \end{bmatrix}, \quad \text{and} \quad H = \begin{bmatrix} \sigma_w^2 & 0 \\ 0 & 0 \end{bmatrix}.$$

Note that it holds that (see Hindrayanto et al., 2014):

$$\sigma_u^2 = r_{uu}^2 + r_{uv}^2$$
$$\sigma_v^2 = r_{vv}^2 + r_{uv}^2$$
$$\sigma_{uv} = r_{uu}r_{uv} + r_{vv}r_{uv}.$$

Appendix B. Trend of the Log of Real Output

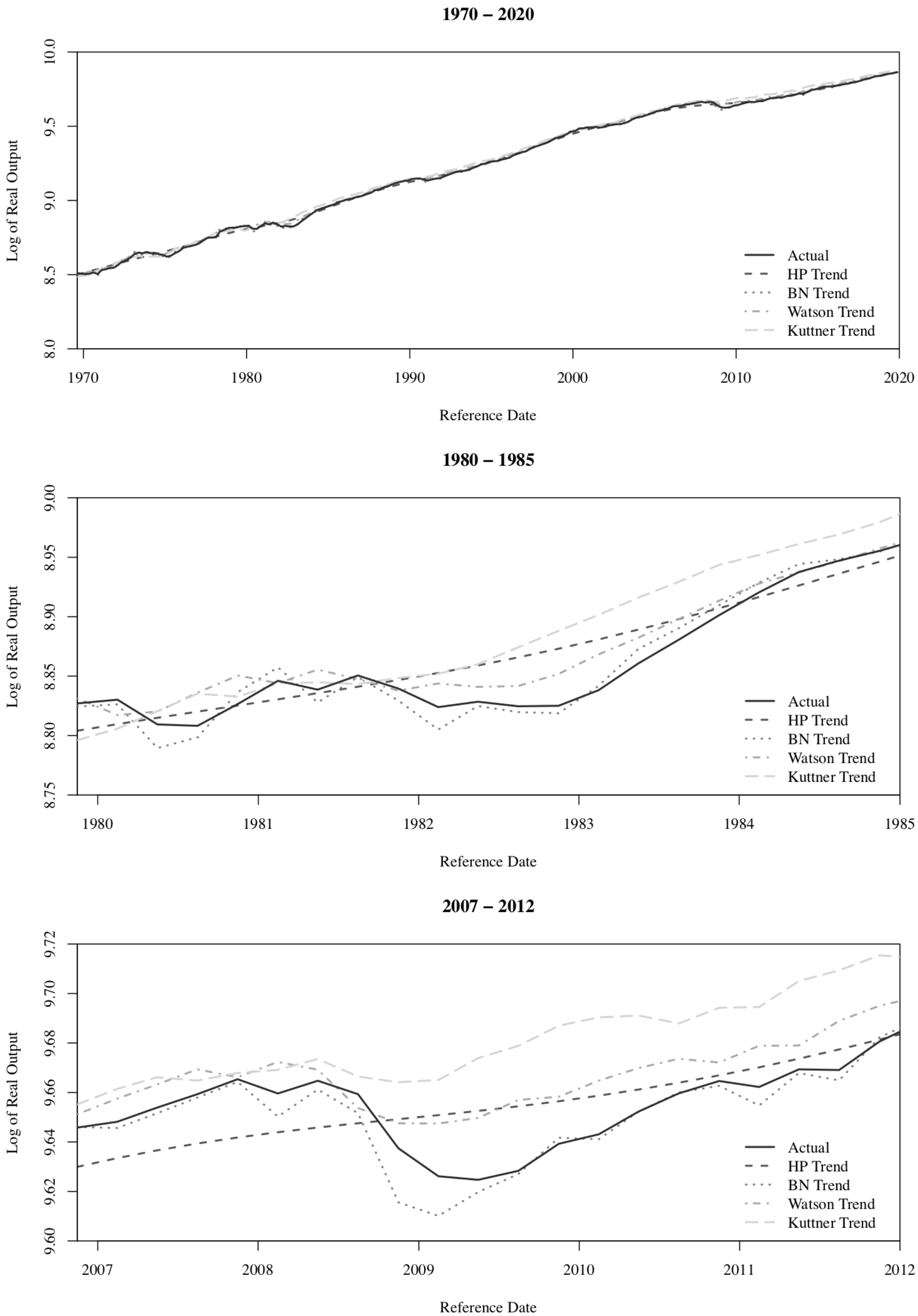

Figure 7. Trend of the Log of Real Output. Plots of the trend component of the log of real output for the most recent vintage under several estimation methods, zoomed in on different time periods.

Appendix C. Inflation Predictions by Estimation Method

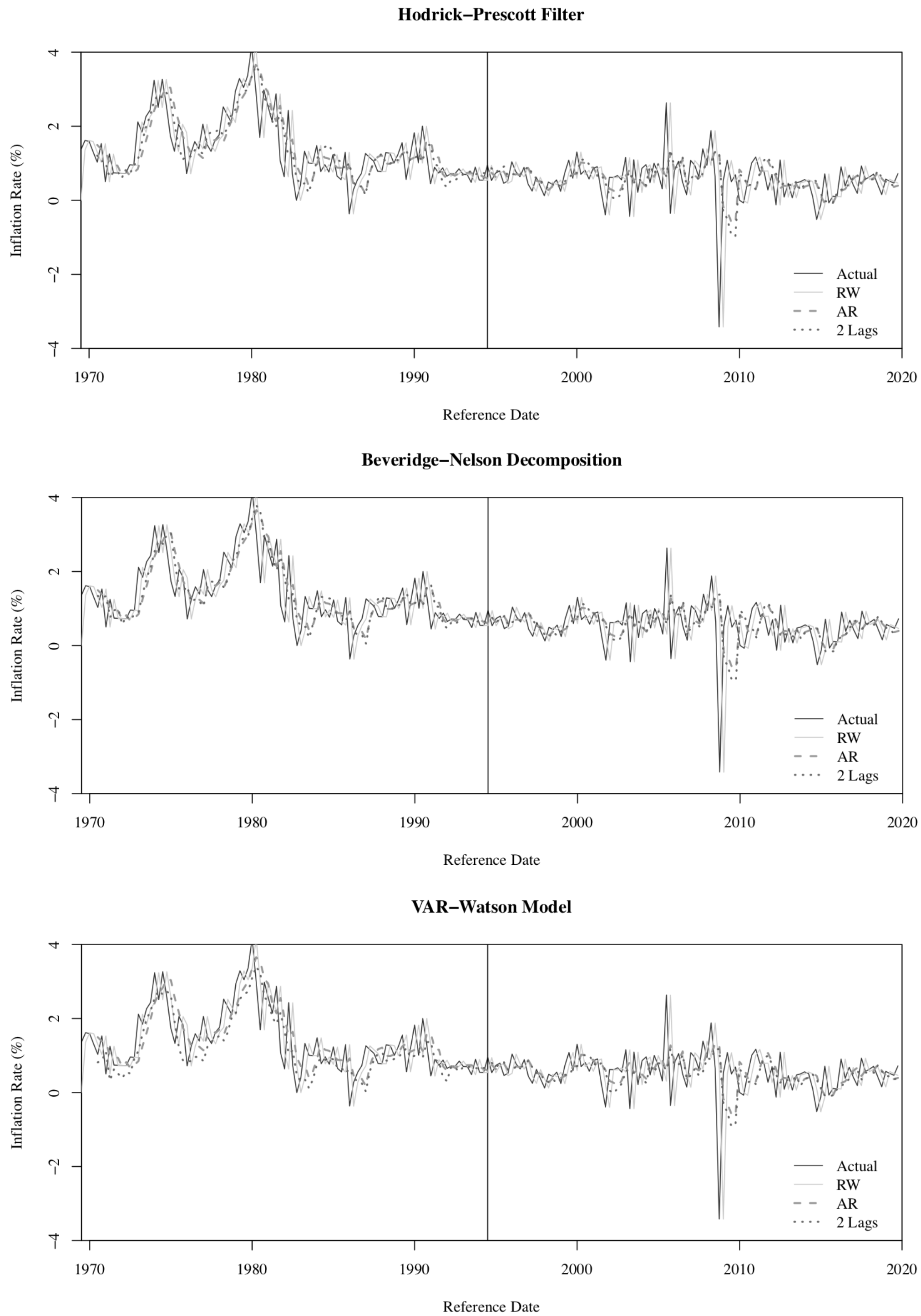

Figure 8. Inflation Predictions by Estimation Method. Plots of one-step-ahead in-sample (before split) and out-of-sample (after split) predictions of the inflation rate—using a random walk, an AR model without lags of the output gap, and an AR model with two lags of output gap estimates from the different estimation methods. Figure continues on Page 35.

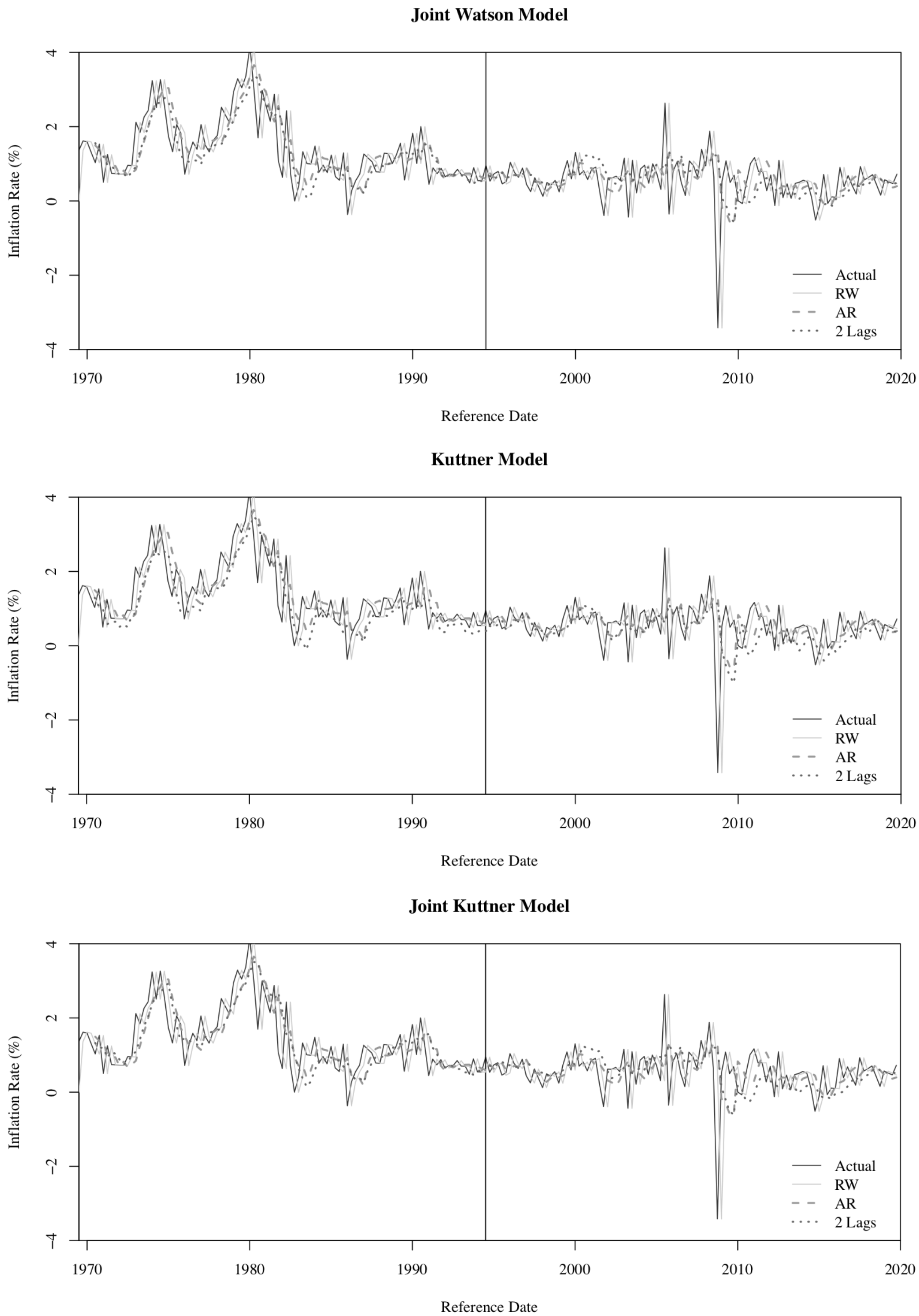

Figure 8. Inflation Predictions by Estimation Method. Plots of one-step-ahead in-sample (before split) and out-of-sample (after split) predictions of the inflation rate—using a random walk, an AR model without lags of the output gap, and an AR model with two lags of output gap estimates from the different estimation methods. Figure continued from Page 34.

Appendix D. Parameter Estimates Within Final Vintage

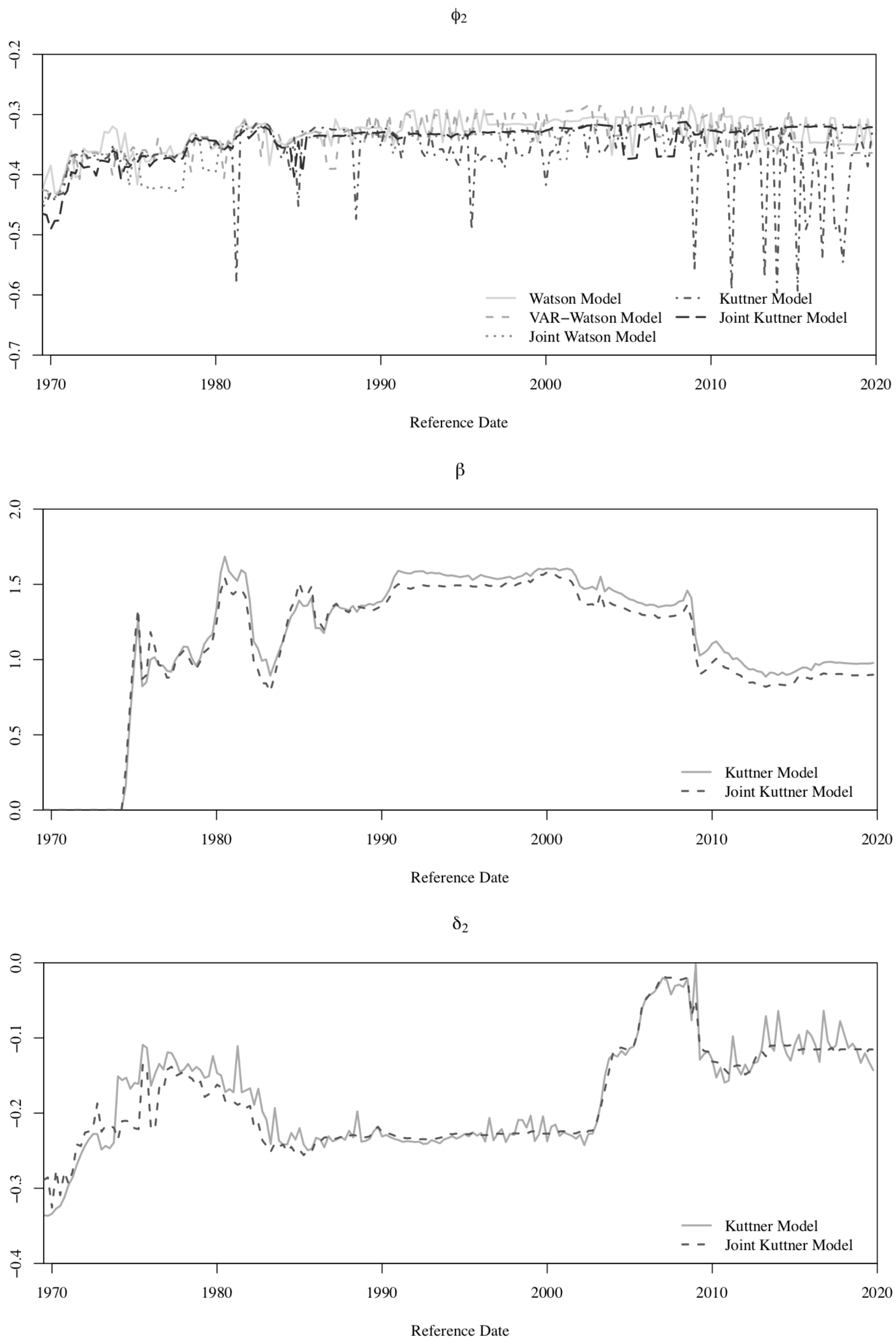

Figure 9. Parameter Estimates Within Final Vintage. Plots of maximum likelihood estimates of the parameters of the output gap estimation methods. Parameters for a particular reference date are estimated based on the data up to that date as available in the final vintage. Hence, these are the parameters used to construct the "quasi-real" output gap estimates. Figure continues on Page 37.

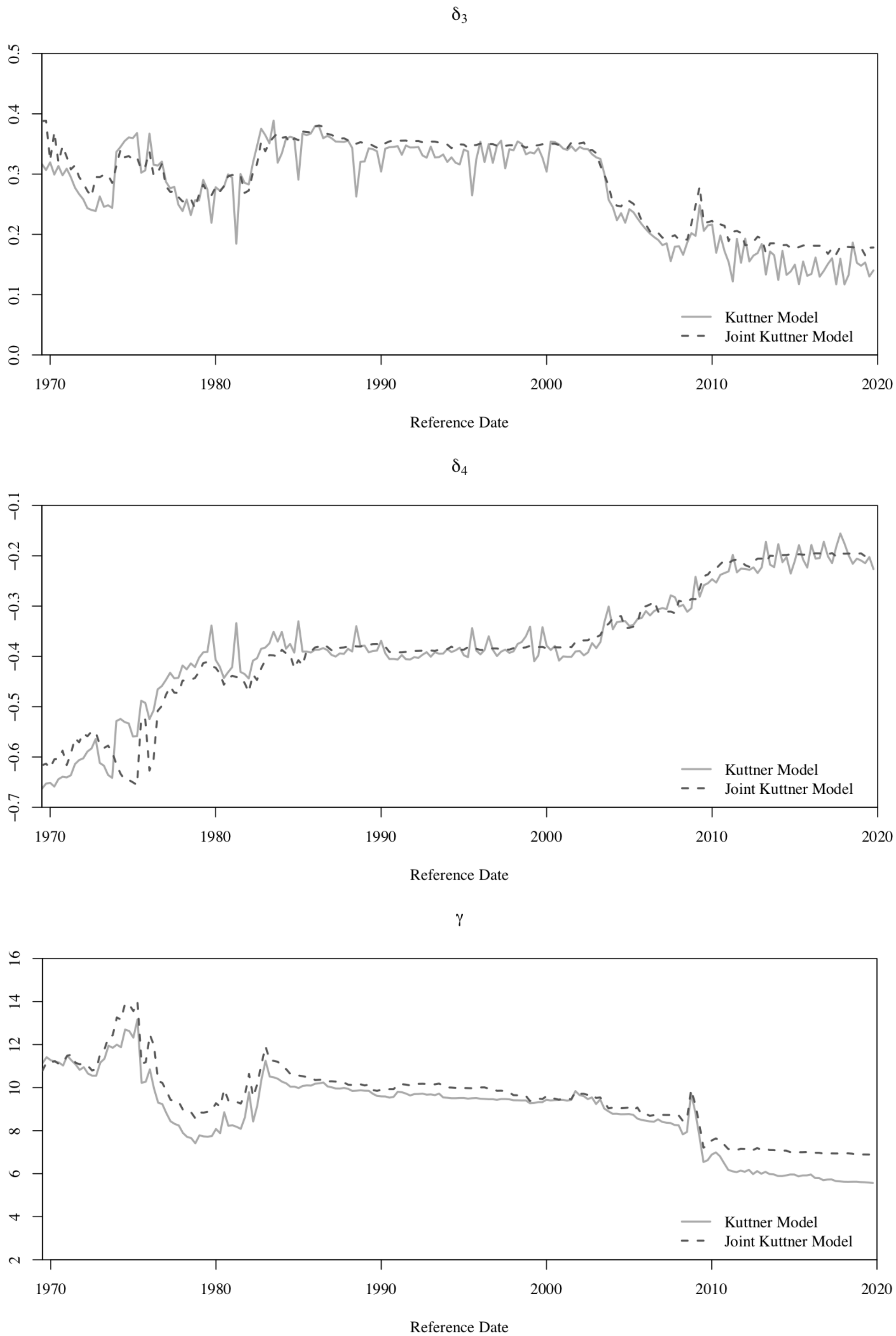

Figure 9. Parameter Estimates Within Final Vintage. Plots of maximum likelihood estimates of the parameters of the output gap estimation methods. Parameters for a particular reference date are estimated based on the data up to that date as available in the final vintage. Hence, these are the parameters used to construct the "quasi-real" output gap estimates. Continued from Page 36.

Appendix E. Parameter Estimates Across Vintages

Table 11. Parameter Estimates Across Vintages.

	Watson Model		VAR-Watson Model		Joint Watson Model		Kuttner Model		Joint Kuttner Model	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
ϕ_1	1.2905	0.0749	1.2803	0.0814	1.2923	0.0628	1.3171	0.0900	1.2871	0.0558
ϕ_2	-0.3525	0.0558	-0.3546	0.0556	-0.3550	0.0532	-0.3912	0.0766	-0.3603	0.0622
β							1.1867	0.7398	1.1180	0.7219
δ_1							-0.5632	0.0771	-0.5631	0.0727
δ_2							-0.1584	0.0628	-0.1614	0.0573
δ_3							0.2517	0.0757	0.2693	0.0655
δ_4							-0.3394	0.1134	-0.3455	0.1075
γ							7.9611	1.6650	8.7508	1.4053
σ_{uv}							0.0002	0.0005	0.5879	0.0004

Means and standard deviations of maximum likelihood estimates of the parameters of the output gap estimation methods. Parameters for a particular vintage date are estimated based on the data available in that vintage. Hence, these are the parameters used to construct the "real-time" output gap estimates.

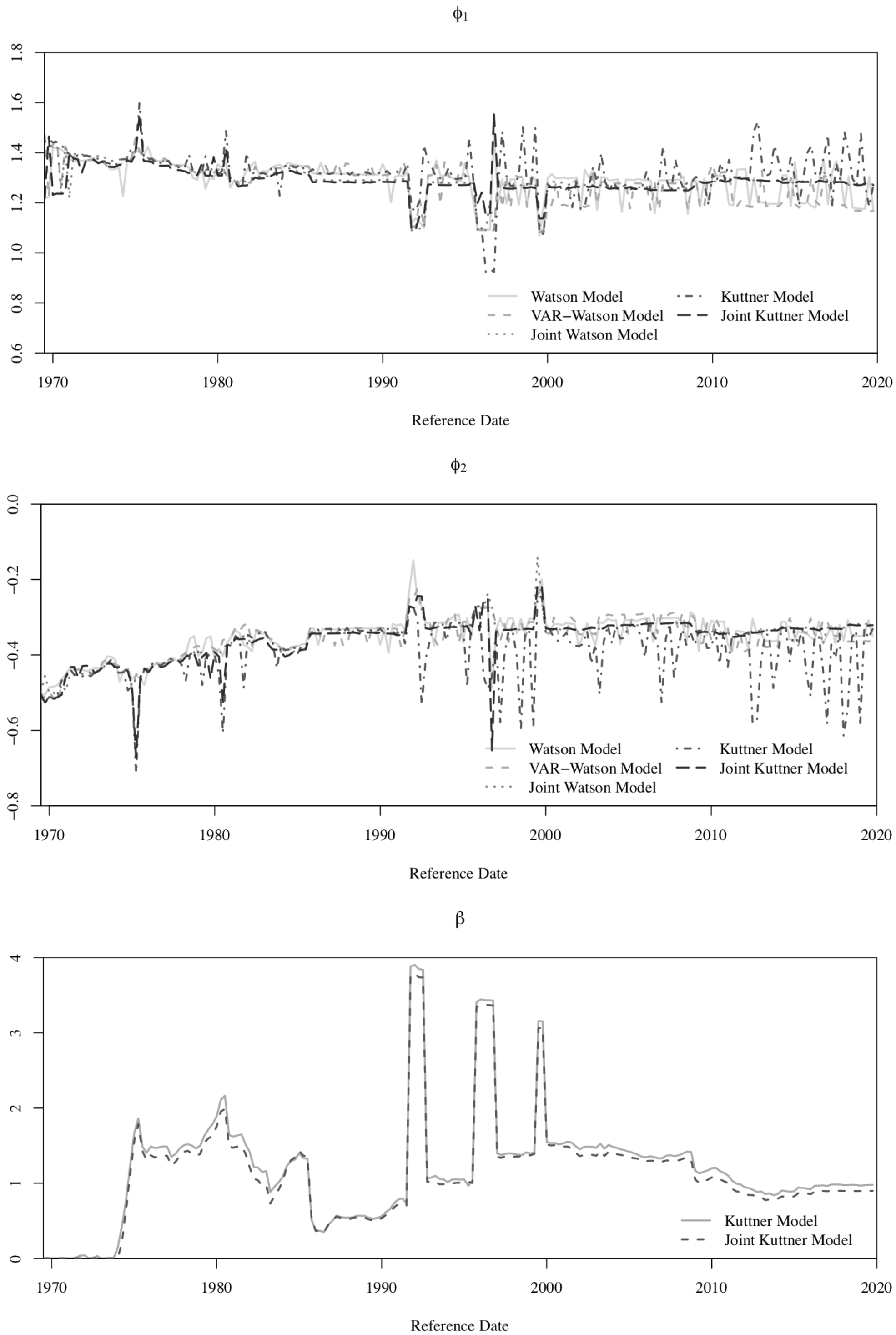

Figure 10. Parameter Estimates Across Vintages. Plots of maximum likelihood estimates of the parameters of the output gap estimation methods. Parameters for a particular vintage date are estimated based on the data available in that vintage. Hence, these are the parameters used to construct the "real-time" output gap estimates. Figure continues on Pages 40–41.

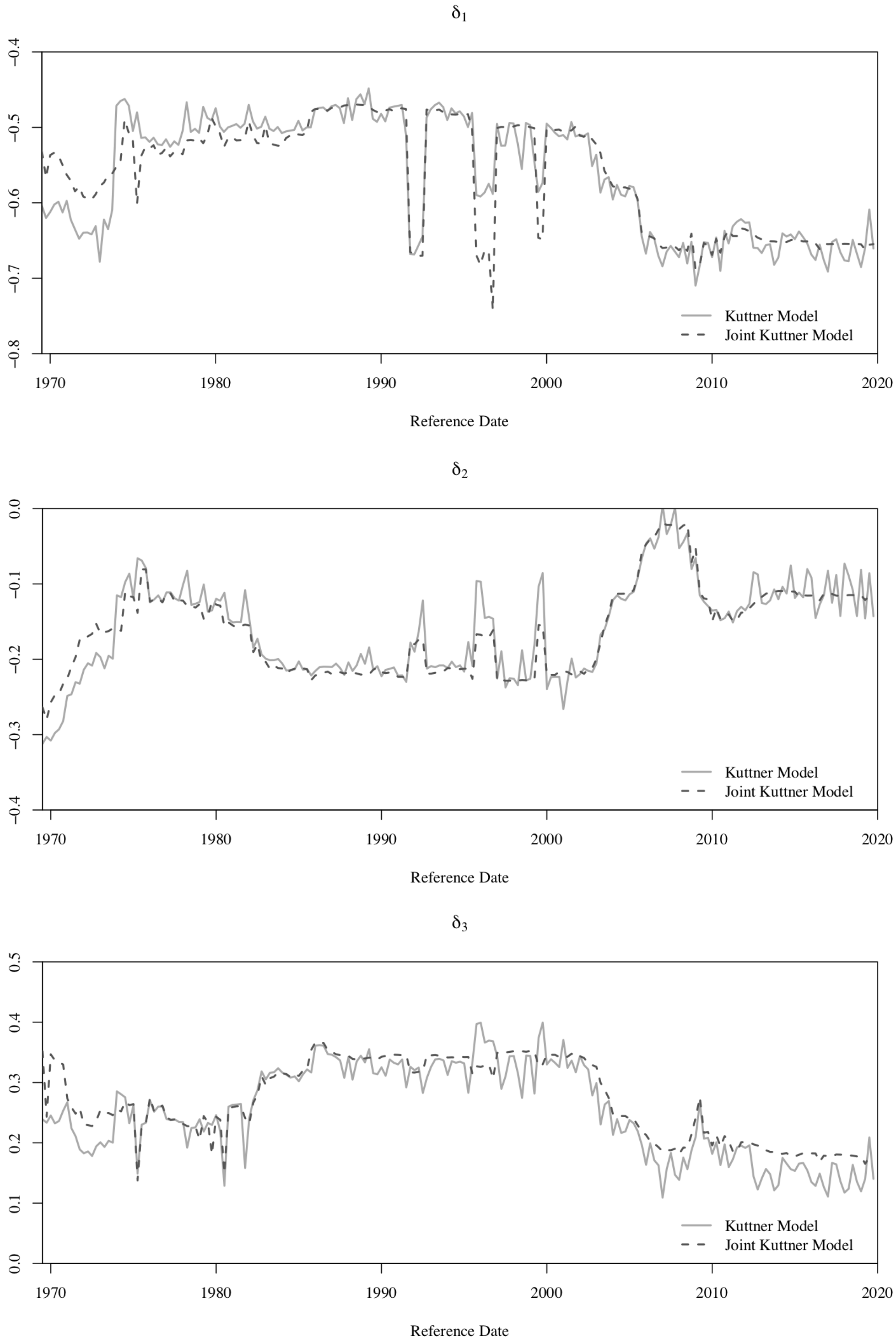

Figure 10. Parameter Estimates Across Vintages. Plots of maximum likelihood estimates of the parameters of the output gap estimation methods. Parameters for a particular vintage date are estimated based on the data available in that vintage. Hence, these are the parameters used to construct the "real-time" output gap estimates. Figure continued from Page 39 and continues on Page 41.

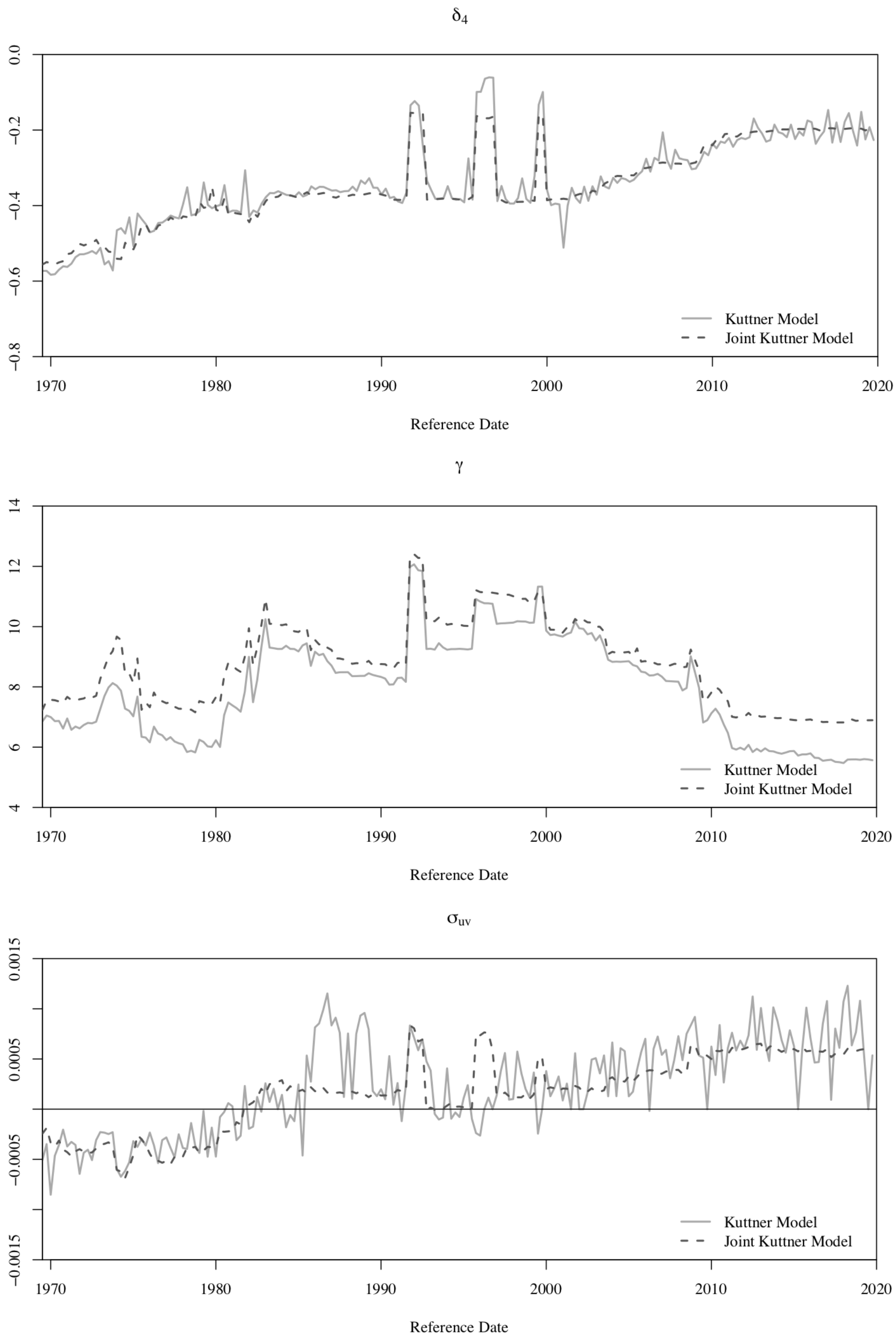

Figure 10. Parameter Estimates Across Vintages. Plots of maximum likelihood estimates of the parameters of the output gap estimation methods. Parameters for a particular vintage date are estimated based on the data available in that vintage. Hence, these are the parameters used to construct the "real-time" output gap estimates. Figure continued from Pages 39–40.

REFERENCES

Alichi, Ali, 2015, A new methodology for estimating the output gap in the United States, Technical Report 15-144.

Álvarez, Luis J., and Ana Gómez-Loscos, 2018, A menu on output gap estimation methods, *Journal of Policy Modeling* 40, 827–850.

Beveridge, Stephen, and Charles 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, 151–174.

Blagrave, Patrick, Roberto Garcia-Saltos, Douglas Laxton, and Fan Zhang, 2015, A simple multivariate filter for estimating potential output, Technical report.

Blanchard, Olivier Jean, and Danny Quah, 1989, The dynamic effects of aggregate demand and supply disturbances, *American Economic Review* 79, 655–673.

Camba-Mendez, Gonzalo, and Diego Rodriguez-Palenzuela, 2003, Assessment criteria for output gap estimates, *Economic Modelling* 20, 529–562.

Clements, Michael P., and Ana Beatriz Galvão, 2012, Improving real-time estimates of output and inflation gaps with multiple-vintage models, *Journal of Business & Economic Statistics* 30, 554–562.

Croushore, Dean, and Tom Stark, 2001, A real-time data set for macroeconomists, *Journal of Econometrics* 105, 111–130.

Cunningham, Alastair, Jana Eklund, Chris Jeffery, George Kapetanios, and Vincent Labhard, 2012, A state space approach to extracting the signal from uncertain data, *Journal of Business & Economic Statistics* 30, 173–180.

Diebold, Francis X., and R. S. Mariano, 1995, Comparing predictive accuracy, *Journal of Business & Economic Statistics* 13, 253–263.

Ehrmann, Michael, and Frank Smets, 2003, Uncertain potential output: implications for monetary policy, *Journal of Economic Dynamics and Control* 27, 1611–1638.

Evans, George W., 1989, Output and unemployment dynamics in the United States: 1950–1985, *Journal of Applied Econometrics* 4, 213–237.

Garratt, Anthony, Kevin Lee, Emi Mise, and Kalvinder Shields, 2008, Real-time representations of the output gap, *Review of Economics and Statistics* 90, 792–804.

Harvey, Andrew C., 1990, *Forecasting, structural time series models and the Kalman filter* (Cambridge University Press).

Havik, Karel, Kieran Mc Morrow, Fabrice Orlandi, Christophe Planas, Rafal Raciborski, Werner Röger, Alessandro Rossi, Anna Thum-Thysen, and Valerie Vandermeulen, 2014, The production function methodology for calculating potential growth rates & output gaps, Technical report.

Hindrayanto, Irma, Jan Jacobs, and Denise Osborn, 2014, On trend-cycle-seasonal interactions, Technical report.

Hodrick, Robert J., and Edward C. Prescott, 1997, Postwar U.S. business cycles: An empirical investigation, *Journal of Money, Credit, and Banking* 29, 1–16.

Jacobs, Jan P.A.M., and Simon Van Norden, 2011, Modeling data revisions: Measurement error and dynamics of "true" values, *Journal of Econometrics* 161, 101–109.

Kamada, Koichiro, 2005, Real-time estimation of the output gap in Japan and its usefulness for inflation forecasting and policymaking, *North American Journal of Economics and Finance* 16, 309–332.

Kara, Hakan, Fethi Öğünç, Ümit Özlale, and Çağri Sarikaya, 2007, Estimating the output gap in a changing economy, *Southern Economic Journal* 269–289.

King, Robert G., Charles I. Plosser, James H. Stock, and Mark W. Watson, 1991, Stochastic trends and economic fluctuations, *American Economic Review* 81, 819–840.

Kuttner, Kenneth N., 1994, Estimating potential output as a latent variable, *Journal of Business & Economic Statistics* 12, 361–368.

Laubach, Thomas, and John C. Williams, 2003, Measuring the natural rate of interest, *Review of Economics and Statistics* 85, 1063–1070.

Mise, Emi, Tae-Hwan Kim, and Paul Newbold, 2005, On suboptimality of the Hodrick–Prescott filter at time series endpoints, *Journal of Macroeconomics* 27, 53–67.

Okun, Arthur M., 1963, *Potential GNP: Its measurement and significance* (Cowles Foundation for Research in Economics, Yale University).

Orphanides, Athanasios, 2003, The quest for prosperity without inflation, *Journal of Monetary Economics* 50, 633–663.

Orphanides, Athanasios, and Simon van Norden, 2002, The unreliability of output-gap estimates in real time, *Review of Economics and Statistics* 84, 569–583.

Pesaran, M. Hashem, and Allan Timmermann, 1992, A simple nonparametric test of predictive performance, *Journal of Business & Economic Statistics* 10, 461–465.

Planas, Christophe, and Alessandro Rossi, 2000, Univariate versus bivariate decomposition and reliability of real-time output gap estimates, Technical report.

Rudebusch, Glenn D., 2001, Is the Fed too timid? monetary policy in an uncertain world, *Review of Economics and Statistics* 83, 203–217.

St-Amant, Pierre, and Simon Van Norden, 1997, Measurement of the output gap: A discussion of recent research at the Bank of Canada, Technical report.

Vetlov, Igor, Tibor Hlédik, Magnus Jonsson, Kucsera Henrik, and Massimiliano Pisani, 2011, Potential output in dsge models, Technical report.

Virmani, Vineet, 2004, Estimating the output gap for the Indian economy: Comparing results from unobserved-components models and the Hodrick-Prescott filter, Unpublished manuscript.

Watson, Mark W., 1986, Univariate detrending methods with stochastic trends, *Journal of Monetary Economics* 18, 49–75.