Monitoring the Economy in Real Time: Trends and Gaps in Real Activity and Prices

Thomas Hasenzagl[1], Filippo Pellegrino[2], Lucrezia Reichlin[3], and Giovanni Ricco[4]

[1] *University of Minnesota and Federal Reserve Bank of Minneapolis*
[2] *London School of Economics and Political Science*
[3] *London Business School, Now-Casting Economics, and CEPR*
[4] *University of Warwick, OFCE-SciencesPo, and CEPR*

17th January 2022

Abstract

A mixed-frequency semi-structural model is used for estimating unobservable quantities such as the output gap, the Phillips curve and the NAIRU in real time. We consider two specifications: in one the output gap is observed as the official CBO measure, in the other is unobserved and derived via minimal theory-based restrictions. We find that the CBO model implies a smoother trend output but the second model better captures the business cycle dynamics of nominal and real variables. The methodology offers both a framework for evaluating official estimates of unobserved quantities of economic interest and tracking them in real time.

Keywords: real-time forecasting, output gap, Phillips curve, semi-structural models, Bayesian estimation.
JEL Classification: C11, C32, C53, E31, E32, E52.

We are grateful to the 13th International Conference on Computational and Financial Econometrics, European Seminar on Bayesian Econometrics 2021, and Université de Montréal seminar participants for helpful comments and suggestions. The opinions in this paper are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Minneapolis.

Introduction

Forming a view on the state of the economy in real time faces two difficulties: the publication delays of relevant data, which requires nowcasting the variables of interest on the basis of incomplete information, and the measurement error of long-term trends, when such estimates are needed to assess the size of the output gap and other structural unobserved components (e.g. see discussions in Orphanides and van Norden, 2002 and Coibion et al., 2017).

A large literature on nowcasting has proposed reduced form methods to deal with the first problem, in the case of observed variables typically transformed to stationarity (see Giannone et al., 2008 for a seminal contribution). Addressing the second problem requires economic theory-based restrictions and possibly the use of judgement – i.e. subjective beliefs – to model unobserved components and the long-run behaviour of variables that cannot be fully inferred from data[1].

Addressing this second problem cannot be avoided in many relevant situations. However, trends and cycles can be extracted from observable series in multiple ways, many of which can deliver plausible and alternative estimates bearing different implications. In fact, measures of the output gaps are important, for example, to gauge the monetary policy stance or to assess fiscal rules compliance.[2] The output gap is also a cornerstone in the analysis of the business cycle. It is correlated to the cyclical components of unemployment via the Okun's law and to inflation via the Phillips curve. Both are key measures that inform real-time fiscal and monetary policy decisions. If the output gap is zero, an economy is said to grow at potential. The unemployment rate consistent with that potential level of output, is the non-accelerating inflation rate of unemployment (NAIRU). Similarly, the underlying trend inflation is the long-run inflation rate that would prevail in the absence of cyclical factors moving economic activity off potential output.

Misinterpreting the size of these key unobservable quantities can be costly. For example, as it has been shown by Orphanides (2001), the misinterpretation of the size of output gap has been the source of damaging policy mistakes.

This paper offers two main contributions for the analysis of the output gap and other structural components. The first one is methodological. We build on Hasenzagl et al. (forthcoming) to develop a semi-structural multivariate time series

[1] Trends are intrinsically long-run concepts for which finite sample data have little or no information as discussed, for example, in Sims, 2000 in the contest of VARs.

[2] This is, in particular, true in the assessment of the compliance of countries to the fiscal rules of the European Union. Trend-cycle decompositions of economic variables are more generally used to assess macroeconomic imbalances, as is for example the case of the Basel credit gap (see Basel Committee for Banking Supervision, 2010, and discussion in Hasenzagl et al., 2020).

models which can handle an extensive dataset of macroeconomic variables with rich dynamics, non-stationary trends, mixed-frequency data, publication delays and asynchronous releases. This model can produce sequences of nowcasts of observables and unobservables trends and cycles, therefore delivering real-time estimates of the output gap, natural rate of unemployment, potential output, the Okun's law, trend and cyclical components of inflation. It is therefore a useful tool for policy analysis. The framework we are proposing can be easily generalised for different applications when the reliability of official measures of unobserved quantities needs independent appraisal.

In particular, we consider two versions of our model:

1. An *undisciplined* model that provides estimates of a number of structural components for the variables included – such as the output gap, the Phillips curve, the Okun's law, the underlying trend inflation, equilibrium unemployment – using minimal restrictions based on economic theory and diffuse priors.

2. A *tracking* model that directly incorporates the CBO measure of the output gap as an observables, but identifies the other structural quantities of interest using the same assumptions of the undisciplined model, as well as the same diffuse priors.

The conjecture that we want to explicitly test is that the first model is superior, and in particular at business cycle frequency, due to the large amount of information and prior knowledge summarised by the CBO output gap.

More generally, the rational for considering the second model is that the CBO output gap measure is the standard reference for both the academic literature and policy analysis, and it is based on a long experience of using data for measuring the slack of the US economy. Naturally, considering it as an input in the multivariate model will have implications for the estimates of trends, the Phillips curve and the Okun's law. Differences in the estimates of those quantities across models can be used to judgementally evaluate the validity of the CBO view versus the statistical view.

Both models deliver nowcasts of the output gap and the other structural components. While in the undisciplined model, the output gap is informed the economic theory-based restrictions only, in the tracking version, the CBO estimate is considered an input. Our methodology allows to use multivariate information to produce nowcasts of those missing values thereby allowing to track it in real-time.

The second contribution is an appraisal of the official CBO measures of output gap and potential. Our analysis shows that the CBO output gap reads the US economy as having been constantly below potential since the 2001 recession which is explained by an implicit view on the output potential being unaffected by events since then. Conversely, the undisciplined model estimates an almost symmetrical output gap fluctuating around a potential whose slope has declined since 2001.

This has implications for the shape of the Phillips curve. Both models provide a decomposition of inflation in terms of an underlying inflation rate – i.e. the trend pinned by expectations –, oil price-driven fluctuations and a Phillips curve component connected to the output gap. While in both models oil shocks dominate and cloud the Phillips curve component, the model with the CBO measure finds on average a smaller role for the latter although the contrary is true since 2008.

The two views of the US economy are to a large extent observationally equivalent. In both models the estimates of the structural components are informed by the data and the multivariate restrictions that requires cyclical components to connect real, labour market and nominal variables via the Okun's law and the Phillips curve. However, while in the undisciplined model the estimate of the relative variability of the trend and the cycle of GDP is determined solely by the data and the diffuse priors (which we set as close to be uninformative), in the tracking model the CBO's gap provides an additional observable that informs the trend-cycle decomposition. Although, which model is used is a matter of judgement – on the plausibility of the different estimates and on their different implications of the various structural quantities across models –, we can use a forecast evaluation as a cross-validation criterion. Again one would expect the model incorporating the CBO's measure to be superiors at business cycle frequencies.

Surprisingly, the forecast evaluation shows that the undisciplined model does better at forecasting at the business cycle frequency although it does not offer an advantage on the very short term forecast. This is true for GDP, labor market variables and inflation. Our results point to a cautionary appraisal of official measures and suggest that the CBO implicit view of a very smooth potential output may lead to a misinterpretation of the cyclical components and hence of the medium term inflationary pressures.

As a byproduct, the tracking model can produce a monthly series for the CBO output gap and its nowcasts. A real time evaluation of the CBO output gap nowcast during the COVID sample, shows that the estimates are quite stable and revisions remarkably small. So, if one is prepared to believe in the CBO view of the output gap, our framework can be used to obtain a timely indicator of such gap.

Related Literature. Our work builds on the tradition of structural time series models (see Harvey, 1985) that models economic variables in terms of common and idiosyncratic cycles and trends with economic interpretations. In particular, our approach builds on Hasenzagl et al. (forthcoming). However, the focus of this manuscript is the real-time tracking of the output gap and other structural components, instead of the study of inflation dynamics. To this aim, we employ mixed-frequency methods that allow us to combine data at monthly and quarterly frequencies. We also show how to incorporate institutional measures of structural components in a flexible manner. A paper with a similar approach but a different aim is Jarociński and Lenza (2018) who employ a variety of Bayesian trend-cycle methods to estimate the output gap in the Euro Area from fully revised, quarterly data.

Since we perform real-time nowcasting and forecasting exercises, this paper also connects to the real-time forecasting literature, as in Giannone et al. (2008) and in Aruoba et al. (2009), and in particular it connects to area that focusses on providing real-time estimates of the output gap. For example, Garratt et al. (2008) adopts a cointegrated VAR framework to model real-time measures of the output gap and observed time series. More recently, Del Negro et al. (2017) proposed a BVAR informed by structural relationships to study the natural interest rate. In the case of Cimadomo et al. (2020) a similar BVAR model is used to nowcast and forecast US GDP growth by incorporating mixed-frequency data. Close in spirit to our work are also Aastveit and Trovik (2014), Aastveit and Trovik (2014) and Barigozzi and Luciani (2020) who propose the use of dynamic factor models. The latter develops a large factor model with non-stationary components that can provide measurements of the output gap and it is employed in a pseudo real-time forecasting exercise. In contrast to those papers, this work adopts a modelling approach that builds on the trend-cycle tradition, and offers a transparent method to model cyclical and structural components and measure them in real-time.

Finally, our paper connects to the literature that studies the real-time properties of structural quantities (see Watson, 2007). This literature has often observed that estimates of the output gap and other structural concepts are method-specific and largely unstable and unreliable across data vintages at different points in time. For example, Orphanides and van Norden (2002) show that revisions of the output gap for the United States are of the same order of magnitude as the estimated gap itself. Coibion et al. (2018) document the puzzling cyclical behaviour of revisions to measures of potential output for the U.S. and other countries. This unreliability of structural measures is indeed a serious issue for economic stabilisation policy, which requires reliable estimates of the output gap in real-time, when policy decisions are made. Against that backdrop, we show that our estimates are stable and hence reliable.

The paper starts with a discussion of our methodology in Section 1. We then introduce the two models. Section 2 discuss in-sample and out-of-sample results while in Section 3 we provide a real time analysis of the Covid period. The last section concludes. The Online Appendix provides details on the Bayesian estimation of the model, and additional results for all of the models discussed in the paper.

1 A Mixed-Frequency Trend-Cycle Framework

In its general form, our empirical framework provides a stylised representation of a number of key macroeconomic variables that are measured at different frequency from statistical offices (see Table 1, for details). In particular, it describes the joint dynamics of real activity (real output), the labour market (employment, unemployment rate), prices and inflation (headline inflation and oil price), and economic expectations (professional forecasts of inflation and output, and consumers' expectations of inflation). These variables are modelled in terms of common cycles and trends that are meant to capture structural components, their dynamics, plus a number of variable-specific components that absorb idiosyncratic dynamics and measurement errors.[3]

In particular, the general state-space representation of the models we consider is of the form

$$\begin{pmatrix} cycle_t^{cbo} \\ y_t \\ F_t^{spf} y_{t+12} \\ u_t \\ e_t \\ oil_t \\ \pi_t \\ F_t^{spf} \pi_{t+12} \\ F_t^{uom} \pi_{t+12} \end{pmatrix} = \underbrace{\begin{pmatrix} \sum_{j=0}^{2} L^j & 0 \\ \sum_{j=0}^{2} L^j & 0 \\ \sum_{j=0}^{3} \gamma_{3,j} L^j & 0 \\ \sum_{j=0}^{3} \gamma_{4,j} L^j & 0 \\ \sum_{j=0}^{3} \gamma_{5,j} L^j & 0 \\ 0 & 1 \\ \sum_{j=0}^{3} \gamma_{7,j} L^j & \delta_7 \\ \sum_{j=0}^{3} \gamma_{8,j} L^j & \delta_8 \\ \sum_{j=0}^{3} \gamma_{9,j} L^j & \delta_9 \end{pmatrix} \begin{pmatrix} \psi_t^{gap} \\ \psi_t^{epc} \end{pmatrix} + \begin{pmatrix} \sum_{j=0}^{2} L^j \psi_{1,t} \\ \sum_{j=0}^{2} L^j \psi_{1,t} \\ \psi_{3,t} \\ \psi_{4,t} \\ \psi_{5,t} \\ \psi_{6,t} \\ \psi_{7,t} \\ \psi_{8,t} \\ \psi_{9,t} \end{pmatrix}}_{\text{Common \& Idiosyncratic Cycles}} + \underbrace{\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \sum_{j=0}^{2} L^j & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 3 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \tau_t^y \\ \mu_t^{spf,y} \\ \tau_t^u \\ \tau_t^e \\ \tau_t^{oil} \\ \tau_t^{\pi} \\ \mu_t^{spf,\pi} \\ \mu_t^{uom,\pi} \end{pmatrix}}_{\text{Trends \& Biases}}, \quad (1)$$

where L is the lag operator.

We consider two versions of this model:

[3] This decomposition can be seen, in finite sample, as a convenient disentangling of permanent/more persistent and stationary/less persistent components. We model the more persistent components as stochastic trends (random walks) and the less persistent ones as stationary ARMA processes. This allows us to connect the first group to the long-run behaviour of the variables and hence long-term expectations, and the latter as mean-reverting features compatible with business cycle fluctuations.

Table 1: US data and common components

Variable name	Label at time t	Frequency	Loads on *BC*	*EP*	*GDP trend*	*Trend* π
CBO: cycle of real GDP*	gap_t^{cbo}	Q	✓			
Real GDP	y_t	Q	✓		✓	
SPF: expected real GDP	$F_t^y \pi_{t+12}$	Q	✓		✓	
Unemployment rate	u_t	M	✓			
Employment	e_t	M	✓			
WTI spot oil price	oil_t	M		✓		
CPI	π_t	M	✓	✓		✓
SPF: expected inflation	$F_t^{spf} \pi_{t+12}$	Q	✓	✓		✓
UoM: expected inflation	$F_t^{uom} \pi_{t+12}$	M	✓	✓		✓

Notes: Data used in the trend-cycle model. All data is in levels, except for CPI which is in YoY (%). 'UoM: expected inflation' is the University of Michigan, 12-months ahead expected inflation. 'SPF: expected inflation' is the Survey of Professional Forecasters, 4-quarters ahead expected inflation rate. Data includes observations from Jan-1985 to Dec-2019. (*) Used in the tracking model only.

1. A 'tracking' model, that is exactly of the form in Equation 1, and that incorporates the Congressional Budget Office measure of the cycle of GDP, proxying for an 'observed' quarterly measurement of the output gap;

2. An 'undisciplined' model where the first equation, for $cycle_t^{cbo}$, does not appear and the GDP cycle is therefore only an 'unobserved' component that the model has to estimate. Both models incorporate assumptions derived from economic theory that allow us to interpret their components in terms of structural relationships.

Assumption 1 (Output potential). The potential output, τ_t^y, is the non-stationary stochastic trend determining the long-run common dynamics of real GDP and expected real GDP.

These assumptions allows us to model potential output as a stochastic trend with drift that is common to realised and expected output. A similar understanding underpins the modelling of trend inflation as a driftless stochastic trend, common across different measures of inflation and inflation expectations.

Assumption 2 (Trend inflation). Trend inflation, τ_t^π, is a non-stationary stochastic trend common between inflation and inflation expectations and captures long-term inflation dynamics.

It is worth observing that trend inflation and trend output are defined, in the spirit of Beveridge and Nelson (1981), to correspond to the long-run forecast for the respective variables, which would implies for a generic rational forecast that

$$\lim_{h\to\infty} F_t[y_{t+h}] = \lim_{h\to\infty} \tau^y_{t+h} \ , \quad (2)$$

$$\lim_{h\to\infty} F_t[\pi_{t+h}] = \tau^\pi_t \ . \quad (3)$$

In a similar manner we also define trends for labour market variables and oil, for which we do not incorporate expectations in the model. All of these stochastic trends, excluding the one for employment, are driftless.

Assumption 3 (Non-accelerating inflation rate of unemployment). The trend unemployment/non-accelerating inflation rate of unemployment (NAIRU), τ^u_t, is defined as the non-stationary stochastic trend defining the long-run behaviour of the unemployment rate. Similarly, trend employment is the non-stationary stochastic trend defining the long-run behaviour of employment.

The cyclical components are modelled as stationary stochastic cycles, under the following set of assumptions.

Assumption 4 (Output gap). The output gap ψ^{gap}_t is a stationary stochastic common component that captures co-movement between the real variables, labor market variables, inflation, and inflation expectations. It informs the price gap via the Phillips curve, and the unemployment gap via Okun's law. Both relationships are modelled as moving averages of output gap realisations during the previous three months.

It is worth observing that the all variables, except for real GDP and the CBO's output cycle, are connected to the output gap with a lag polynomial. This is to allow the model to nest, under parametric restrictions, the case of rational expectations, as discussed in Hasenzagl et al. (forthcoming).

We also consider a second common cyclical component, which we call the "energy price cycle". This component captures volatility in prices due to energy prices shocks, due possibly to tension in commodity markets or global events.

Assumption 5 (Energy price cycle). The energy price cycle ψ^{epc}_t is a stationary stochastic common cyclical component connecting oil prices, inflation, and inflation expectations.

This common cycle captures the standard oil price component that is included in measures of headline inflation but also an expectations-driven component that affects headline inflation (see Hasenzagl et al., forthcoming, for a discussion). In fact, as discussed in Coibion and Gorodnichenko (2015), large oil price shocks may have larger effects on inflation expectations than shocks to other prices and thus cause expectations-driven fluctuations in inflation.

Assumption 6 (Idiosyncratic stationary components). All variables have an idiosyncratic stationary component, $\psi_{i,t}$, which absorbs different sources of idiosyncratic dynamics such as idiosyncratic shocks, non-classic measurement error, differences in definitions, and other sources of noise.

The idiosyncratic stationary components are important for creating a 'wedge' in the empirical model, that will absorb different, small dynamic components that are not part of the multivariate dynamics and thus could distort the empirical estimates.

When modelling expectation we also allow for persistent 'bias' components.

Assumption 7 (Bias in Expectations). Agent's expectations can deviate from a rational forecast due to time-varying bias – respectively $\mu_t^{spf,y}$, $\mu_t^{spf,\pi}$ for the professional forecasters' and $\mu_t^{uom,\pi}$ for consumers' expectations. The bias terms are modelled as stochastic random walk components.

These biases are useful for capturing persistent deviations from trends in observed real-world forecasts. While, in the general form of the model, we allow for these biases to affect both consumers' and professionals' expectations, empirically only consumers' expectations exhibit persistent biases (see Coibion and Gorodnichenko, 2015, for a discussion). In line with this observation we set $\mu_t^{spf,\pi}$ to zero (SPF forecast for inflation are 'on trend' at all times), and we only allow the constant $\mu^{spf,y}$ to account for measurement differences in the expected output trend, possibly due to measurement and aggregation issues. We discuss issues related to mixed-frequency aggregation in the next section.

Finally, when modelling the dynamics that govern the evolution of the unobserved components over time we follow the general approach of Harvey (1985).

Assumption 8 (State dynamics). The stationary cycles are modelled as ARMA(2,1) stochastic processes, where the coefficients are restricted to have defined periodicity. The trends are random walks. Specifically, potential output and trend employment are random walks with drift, and all the remaining trends are driftless random walks. All of the processes have mutually orthogonal stochastic innovations.

A ARMA(2,1) process displays pseudo-cyclical behaviour and can be written as

$$\begin{aligned} \widehat{\psi}_t &= \rho\cos(\lambda)\widehat{\psi}_{t-1} + \rho\sin(\lambda)\widehat{\psi}^*_{t-1} + v_t \ , \\ \widehat{\psi}^*_t &= -\rho\sin(\lambda)\widehat{\psi}_{t-1} + \rho\cos(\lambda)\widehat{\psi}^*_{t-1} + v^*_t \ , \end{aligned} \tag{4}$$

where the parameters $0 \leq \lambda \leq \pi$ and $0 \leq \rho \leq 1$ can be interpreted, respectively, as the frequency of the cycle and the damping factor on the amplitude while $\widehat{\psi}^*_t$ is an auxiliary cycle and v_t and v^*_t are uncorrelated white noise disturbances (see Harvey, 1990).[4] The disturbances make the cycle stochastic rather than deterministic and, if $\rho < 1$, the process is stationary.

1.1 Mixed-frequency Aggregation

Our dataset is a panel of incomplete monthly time series. Indeed, data is sampled at monthly frequency and quarterly indicators are considered as monthly variables observed every three months. More specifically, inflation and labour markets are released at monthly frequency by their respective statistical offices, while national account statistics on real GDP and the survey of expectations for professional forecasters are (generally) released quarterly. The consumer survey we employ is monthly. Bridging these variables together at the higher release frequency (i.e. monthly) allows us to have timely estimates of the business cycle and develop a more granular view on inflation.

This modelling choice also allows us to disaggregate our quarterly data into monthly figures. Indeed, we employ a set of restrictions similar to those proposed in Mariano and Murasawa (2003), which are compatible with quarterly data in levels.

Assumption 9 (**Aggregation rules for the CBO cycle and real GDP).** For the CBO cycle and real GDP, we link observed quarterly numbers to latent monthly figures (denoted with the use of a tilde) via

$$\begin{aligned} &cycle^{cbo}_t = (1 + L + L^2)\ \widetilde{cycle}^{cbo}_t , \\ &y_t = (1 + L + L^2)\ \tilde{y}_t, \end{aligned}$$

[4]It is straightforward to show that the model can be rewritten as

$$(1 - 2\rho\cos(\lambda)L + \rho^2L^2)\widehat{\psi}_t = (1 - \rho\cos(\lambda)L)v_t + (\rho\sin(\lambda)L)v^*_t \ .$$

Hence, under the restriction $\sigma^2_v = 0$, the solution of the model is an AR(2), otherwise an ARMA(2,1). The intuition for the use of the auxiliary cycle is closely related to the standard multivariate AR(1) representation of univariate AR(p) processes.

where

$$\widetilde{cycle}_t^{cbo} = \tilde{y}_t - \tau_t^y, \tag{5}$$

$$\tilde{y}_t = \psi_t^{gap} + \psi_{1,t} + \tau_t^y, \tag{6}$$

for any t. This is standard and used in a range of articles including Giannone et al. (2008) and Bańbura and Modugno (2014).

Assumption 10 (**Aggregation rules for expectational data).** We treat the professional expectations differently, depending on whether they refer to real GDP or inflation, since the very nature of these forecasts is distinct. Indeed, at any t,

$$F_t^{spf} y_{t+12} = F_t^{spf} \tilde{y}_{t+12} + F_t^{spf} \tilde{y}_{t+11} + F_t^{spf} \tilde{y}_{t+10}$$

and $F_t^{spf} \pi_{t+12}$ is simply considered as a end-of-month one-year ahead forecast for year-on-year inflation, released at quarterly frequency.[5] Hence, we only need to enforce a mixed-frequency aggregation rule for the real GDP professional expectations. Not knowing the exact prediction rule followed by professional forecasters, we enforce an aggregation rule on their long-term view. This is achieved building on Equation 6 and letting the trend of $F_t^{spf} y_{t+12}$ be $(1 + L + L^2)\tau_{t+12}^y$. Since the latter is a random walk with drift, it follows that this moving average is equivalent to $3\tau_t^y$ plus a time-invariant drift. We let the model free to estimate the drift itself and the loadings, the loadings on the output gap component and a stationary cycle to account for complex dynamics. The time-invariant drift is denote as $\mu^{spf,y} \equiv \mu_t^{spf,y}$ for every t.

1.2 Bayesian Estimation

For the estimation, each row of the observation equation is divided by the standard deviation of the first-differenced data.[6] Hence, jointly with the dynamics of the unobserved components, the model is written in linear state-space form and estimated with an Adaptive Metropolis-Within-Gibbs algorithm (see subsection B.1).

To allow for incomplete data, we employ a Kalman filter compatible with missing observations (see, for instance, Shumway and Stoffer, 1982) and reconstruct the data on the basis of the information set available at each point in time. We further employ

[5]This follows from the fact that inflation itself is monthly.

[6]As showed in Hasenzagl et al. (forthcoming) this gives a better mixing in the Metropolis algorithm.

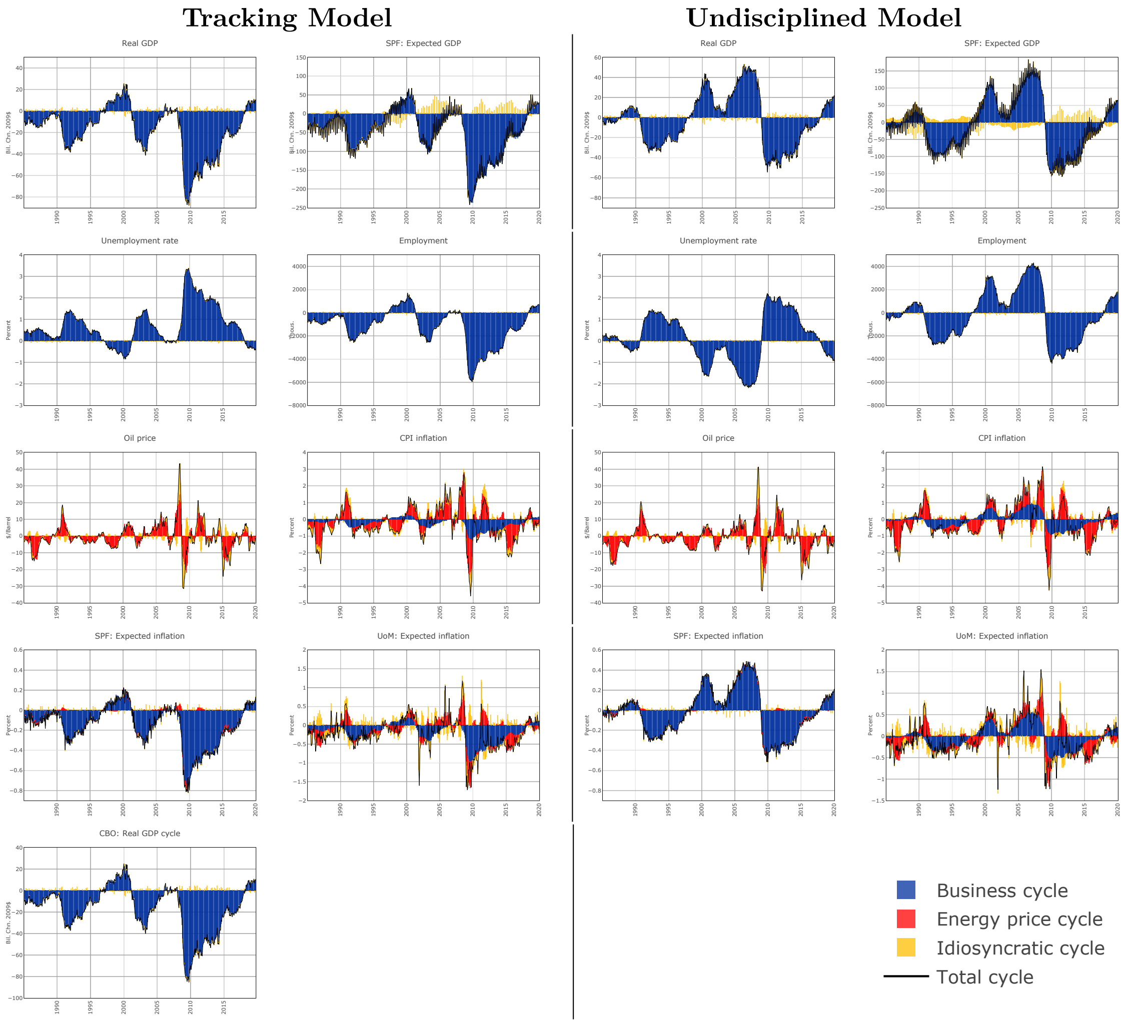

Figure 1: Historical decomposition of the stationary components of all the variables in the tracking model (left) and the undisciplined model (right).

the simulation smoother in Durbin and Koopman (2002) with the Jarociński (2015)'s modification to condition our estimates of cycles and trends on the full sample.

2 Trends and Gaps in the US Economy

Figure 1 reports the cyclical components of all variables and for the two models. The results for the tracking model are in the first two columns and the results for the undisciplined model are in the last two. These are in-sample results from January 1985 to December 2019 using fully revised data. The decomposition shows three components: business cycle, energy price cycle and a residual idiosyncratic component reflecting stationary measurement error.

The historical decompositions show interesting commonalities and differences. The tracking model – in line with the assessment of the CBO onto which it is geared – estimates an output gap that shows large contractions in the cyclical component of real GDP after the early 1990s recession – i.e. the dot-com bubble, and the Great Recession (see Figure 1). This contrasts with the assessment of the undisciplined model, which shows larger expansions in the late 1990s and early 2000s that culminate into the dot-com crisis and the Great Recession. Importantly, both models explain almost the entire deviation of output from its trend in terms of the output gap, with a less-than-marginal role played by the idiosyncratic stationary component of output. Thus, the different assessments of the cyclical component in the two models reflects the different evaluations of potential output.

The different output gap estimates are reflected in the cycles of employment and unemployment which are linked to the output gap by Okun's law and the decomposition of inflation.

However, the latter is dominated by the oil component which is unaffected by the output-trend decomposition. As a consequence, cyclical inflation is more similar across the two models than the employment and unemployment cycles. This similar assessment of the price gap is an interesting robust feature across models. It shows that the models are able to use the freedom allowed by the Phillips curve, modelled as a lag polynomial of the output gap, to fit a similar cyclical component of prices. Indeed, the slope of the 'reduced form' Phillips curve is very similar in the two models[7] although the shape is quite different.

Estimates of the trends are reported in Figure 2 where they are plotted against their associated observable variables. The output gap is the difference between output and its trend. It can be easily seen that the undisciplined model fits an output gap which fluctuates almost symmetrically around the trend as it would be the case in a standard Neoclassical or New Keynesian macroeconomic model. Conversely, in the model informed by the CBO, potential output is above GDP most of the time and recessions appears as shortfalls against this higher level. Clearly, the undisciplined model attributes a larger part of the output variance to the trend, interpreting the slowdown since 2001 and especially that since 2008, as a change to potential output rather than as cyclical fluctuations. As a consequence, the estimate of the NAIRU in the second half of the sample is higher. Not surprisingly, trend inflation is almost identical across the two models.

[7]The estimate of the slope of the Phillips curve obtained by regressing the business cycle component of CPI on the business cycle component of unemployment is -0.43 for the undisciplined model and -0.36 for the tracking.

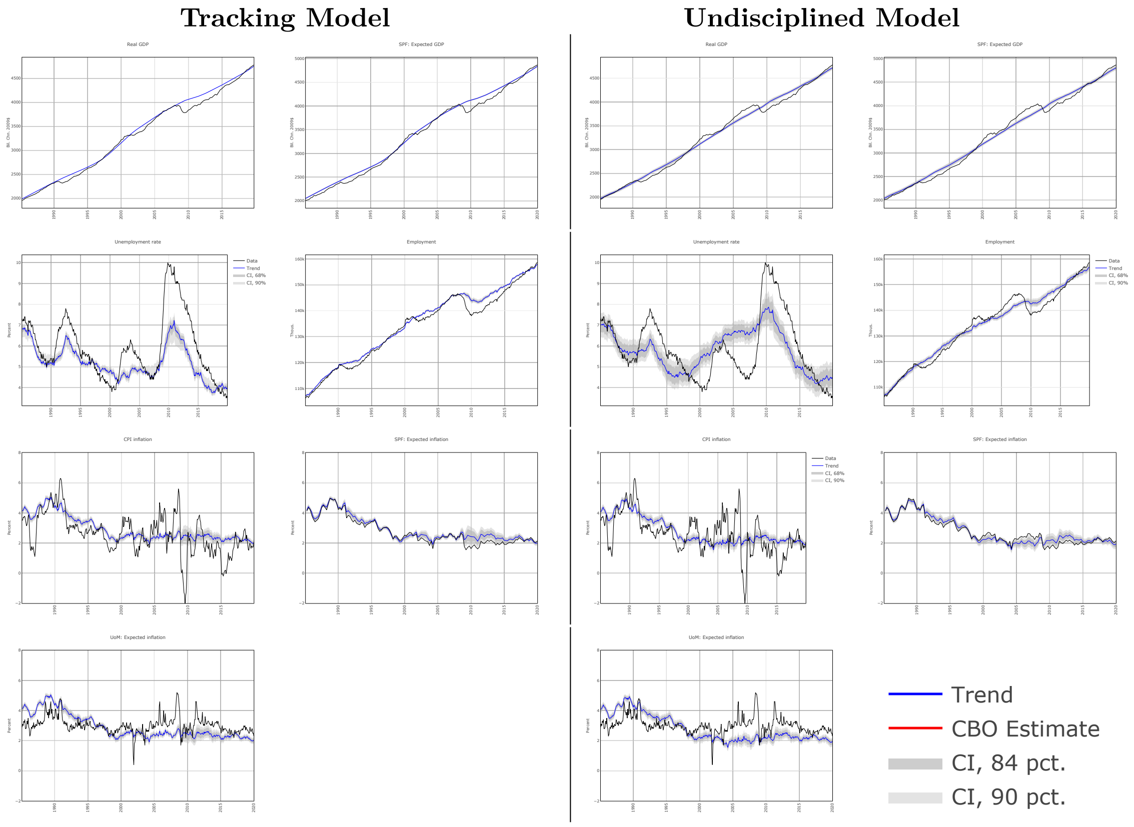

Figure 2: Trends for all the variables in the tracking model (left) and the undisciplined model (right), along with 84% 90% posterior coverage bands and The charts also report the CBO estimates for output potential and NAIRU.

To summarise, what explains the differences between the two models? By anchoring the output gap estimates to the official gap measure, the tracking model naturally embeds the CBO's view on the output trend. This does not happen in the case of the undisciplined model, which estimates an output trend supporting roughly symmetric fluctuations in the sample of interest. Ultimately, the divergence in the assessment provided by the two models pertains to the 'long run' view that the two models have about economic trends.

Notice that trends are intrinsically long-run concepts for which finite sample data have little or no information, and are therefore informed implicitly or explicitly by the beliefs incorporated in the model (see discussion in Sims, 2000, for example). In the case of the tracking model beliefs about the long-run enter via the CBO's cyclical measure, while for undisciplined model is only informed by the economic theory-based restrictions. In both models, the statistical priors are relatively uninformative so as to guarantee a fairly flexible potential output and symmetric fluctuations via the Gaussian likelihood function. It is important to stress that the data are compatible with both models and the two assessments are both plausible.

Whether to adopt one model or the other is ultimately a matter of judgement. Our multivariate framework is designed so as to inform that judgement by evaluating the implications of the two views for variables other than output.

3 A Real-Time Evaluation in Times of COVID

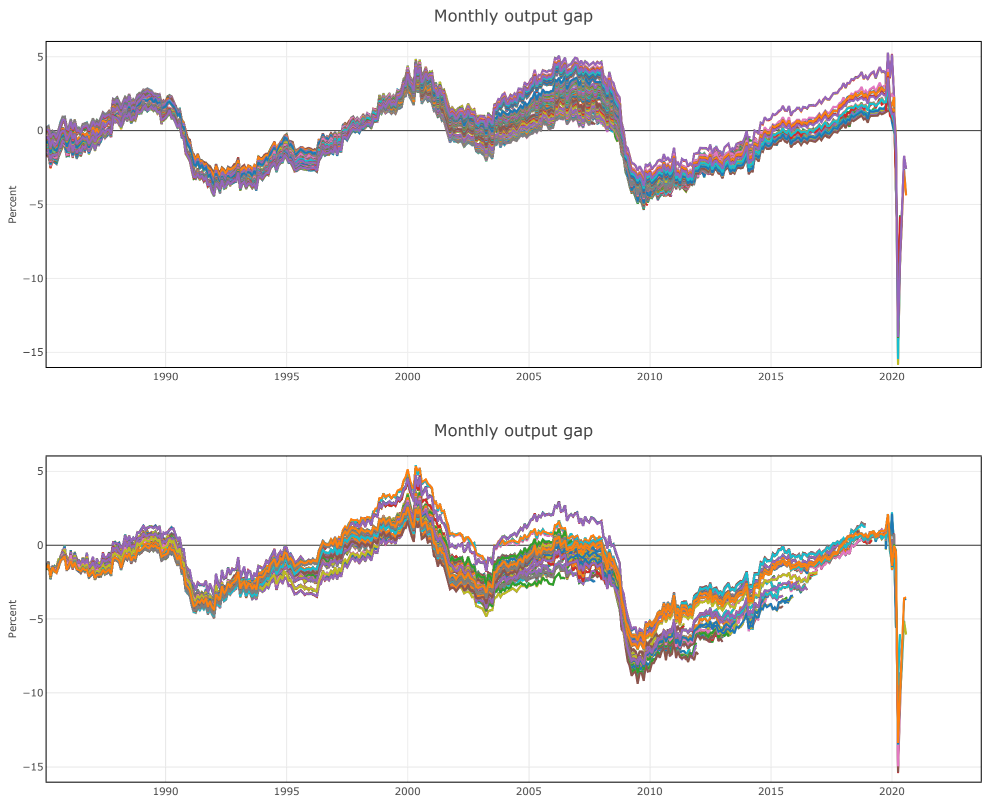

Figure 3: The chart reports the real-time estimates of the output gap from the undisciplined (top) and the tracking model (bottom). The out-of-sample evaluation starts in January 2005 and ends in September 2020.

We now turn to a real-time forecasting exercise, which can be see as akin to a cross-validation exercise of the two models. To this end, we construct a set of real-time data vintages from Archival FRED (2021) and Survey of Professional Forecasters (2021) starting on January 1, 2005 and using the prior 20 years as our pre-sample. Then we iterate over the real time release calendar of the variables in the model and update our estimates of the trends and gaps at each new data release. For each release, we also project the trends and gaps forward and use them to forecast the variables in the model. To avoid overfitting, we re-estimate all model parameters at the first release of each year, and then keep them fixed for the remainder of the year.

The output gap estimates in Figure 3 are defined as the ratio of the sum of the business cycle component and the idiosyncratic cycle to the trend. For the tracking model the sum of the business cycle and idiosyncratic components is equal to the output gap as estimated by the CBO. In the pre-sample period, prior to 2005, the different lines reflect instability of in-sample estimates while from 2005 they are affected by both, instability of estimates and revision errors. In the tracking model, revisions of the output gap can only be due to revisions of the CBO estimates as reported by the CBO or the model's revisions of the estimated trend.

To understand the relative stability of output gap measures across models we compute two statistics. For each reference period we compute the standard deviation of the output gap and potential output across all vintages. We then compute our first statistics by averaging these standard deviations across all reference periods. For the undisciplined model, this statisitic is 0.54 for the output gap and 6.79 for potential output compared to 0.61 and 8.33 for the tracking model. The second statistic is the average of the maximum of the absolute value of revisions for each reference month. For the undisciplined model, the second statisitic is 0.91 for the output gap and 16.38 for potential output compared to 1.37 and 15.09 for the tracking model.

The results suggest that, on average, the gap is more stable in the undisciplined model than in the tracking model. The same is true for potential output although the difference across models is smaller. This suggests that most of the difference in variability of the gap measure across models is due to the larger revisions of the output gap by the CBO with respect to the revisions of the undisciplined model. The statistics described here and the same statisitcs computed for the in-sample period ending at the end of 2004 are also reported in table 2 in the Online Appendix.

The COVID pandemic period also provides a good illustration of the flexibility of the framework. During that period the estimate of potential output (but for a small uptick) seems to track developments smoothly despite the enormous size of the economic shock and its unprecedented nature.

A comparative assessment of the out-of-sample performances of the two models is provided by the average mean squared error of forecasts up to three years ahead (in Figure 4). The tracking model outperforms the undisciplined model up to six quarters ahead for the SPF GDP survey expectations. However, at business cycle frequency (horizon beyond a year), where the structural decomposition matters more, the undisciplined model outperforms the tracking for real output, employment, unemployment rate, inflation, and inflation surveys. This result point to the fact that the undisciplined model is better at pinning down the business cycles

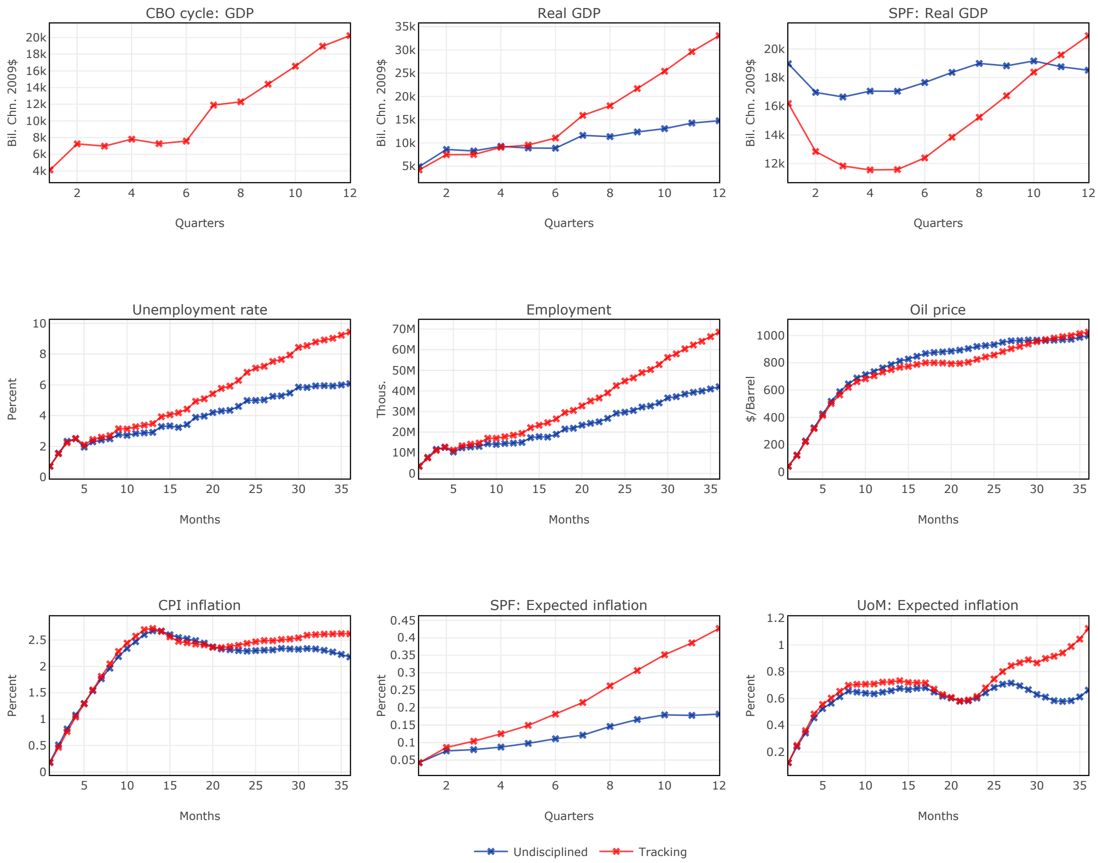

Figure 4: The chart reports the average mean squared error of the undisciplined model (in blue) and the tracking model (in red). The out-of-sample evaluation starts in January 2005 and ends in September 2020.

components. This is a surprising results indicating the limited value of the CBO measure for the model in the cyclical assessment of the economy.

This result sheds some light on the debate of the size of the Phillips curve since whether the Phillips curve is on average flat or steep depends on the estimate of the size of the output gap. Based on our view of the gap, the Phillips curve is on average relatively steep (the slope is -0.42 in the undisciplined model and -0.36 in the tracking model). The model with the steeper curve does better at forecasting inflation.

Notice that for both models the Phillips curve is relatively steeper than some of the estimates in the literature. This is explained by the fact that the models separate an expectation-driven energy price cycle, that is a component of oil prices which is independent from local real economic conditions, from the output gap component. The effects of energy prices shocks can be seen as confounding factors in other studies, reducing the estimated correlation between the slack in the economy and price pressures.

4 Concluding Comments

We propose a mixed-frequency semi-structural time series model which can be used to track observable and unobservable quantities such as the output gap, the Phillips curve and the NAIRU, in real time. The model has many attractive features: it provides a framework for real time analysis for non-stationary data without relying on de-trending methods, and it can be used to identified unobserved components of economic interest by using minimal economic theory-based restrictions and diffuse statistical priors. These features makes it suitable to combine structural analysis, nowcasting and forecasting.

The multivariate nature of the model can also help understanding the implications of using externally constructed measures of variables of interest and therefore judging their reliability. We show that using as an example the CBO measure of the output gap. We do so by considering two versions of the model: one in which the CBO output gap is treated as an observable input variable and one in which the output gap is derived only be using minimal economic and statistical restrictions. The two models are to some extent observationally equivalent but, in the first, the CBO gap constraints the estimates of the trend, the Phillips curve and other quantities of interest. The estimation results show that, in particular, it implies a flatter Phillips curve and a stiffer output potential than the unconstrained model. The differences are larger in the period after the 2001 dot-com bubble. Since 2001 the undisciplined model estimates a slowdown of potential growth and interprets the 2008 recessions to have had some permanent negative effect on potential output growth. The tracking model, on the other hand, implies a smoother trend therefore reading the large negative shocks of that period as having been entirely cyclical.

We compare the two models using two criteria: out-of sample performance and stability of the output gap in real time. This can be seen as a cross validation approach. We find that the unconstrained version of the model – by allowing the GDP trend to absorb more of the output volatility and decline after the 2001 – performs better in forecasting inflation at business cycle frequency. We interpret this as indicating that a model that allows for large shocks to have very persistent effect on output is a better gauge of the Phillips curve dynamics. Moreover, the unconstrained output gap is on average less revised than the CBO own estimates.

The proposed framework can be easily applied to other empirical problems. For example, extensions of the model could be usefully apply to evaluate institutional measures of the natural rate of interest, the credit gap and others.

References

Aastveit, Knut Are and Torres Trovik, "Estimating the output gap in real time: A factor model approach," *The Quarterly Review of Economics and Finance*, 2014, *54* (2), 180–193.

Archival FRED, 2021.

Aruoba, S. Borağan, Francis X. Diebold, and Chiara Scotti, "Real-Time Measurement of Business Conditions," *Journal of Business & Economic Statistics*, 2009, *27* (4), 417–427.

Bańbura, Marta and Michele Modugno, "Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data," *Journal of Applied Econometrics*, 2014, *29* (1), 133–160.

Barigozzi, Matteo and Matteo Luciani, "Measuring the Output Gap Using Large Datasets," SSRN December 2020.

Basel Committee for Banking Supervision, *Guidance for national authorities operating the countercyclical capital buffer* December 2010.

Beveridge, Stephen and Charles R Nelson, "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*, 1981, *7* (2), 151–174.

Cimadomo, Jacopo, Domenico Giannone, Michele Lenza, Andrej Sokol, and Francesca Monti, "Nowcasting with large Bayesian vector autoregressions," Working Paper Series 2453, European Central Bank August 2020.

Coibion, Olivier and Yuriy Gorodnichenko, "Is the Phillips Curve Alive and Well after All? Inflation Expectations and the Missing Disinflation," *American Economic Journal: Macroeconomics*, January 2015, *7* (1), 197–232.

_ , _ , and Mauricio Ulate, "The Cyclical Sensitivity in Estimates of Potential Output," NBER Working Papers 23580, National Bureau of Economic Research, Inc July 2017.

_ , _ , and _ , "The Cyclical Sensitivity in Estimates of Potential Output," *Brookings Papers on Economic Activity*, 2018, *49* (2 (Fall)), 343–441.

Del Negro, Marco, Domenico Giannone, Marc P. Giannoni, and Andrea Tambalotti, "Safety, Liquidity, and the Natural Rate of Interest," *Brookings Papers on Economic Activity*, 2017, *48* (1 (Spring), 235–316.

Durbin, James and Siem Jan Koopman, "A simple and efficient simulation smoother for state space time series analysis," *Biometrika*, 2002, pp. 603–615.

_ and _ , *Time series analysis by state space methods*, Vol. 38, OUP Oxford, 2012.

Garratt, Anthony, Kevin Lee, Emi Mise, and Kalvinder Shields, "Real-Time Representations of the Output Gap," *The Review of Economics and Statistics*, November 2008, *90* (4), 792–804.

Giannone, Domenico, Lucrezia Reichlin, and David Small, "Nowcasting: The real-time informational content of macroeconomic data," *Journal of Monetary Economics*, May 2008, *55* (4), 665–676.

Haario, Heikki, Eero Saksman, and Johanna Tamminen, "Componentwise adaptation for high dimensional MCMC," *Computational Statistics*, 2005, *20* (2), 265–273.

Harvey, Andrew C, "Trends and cycles in macroeconomic time series," *Journal of Business & Economic Statistics*, 1985, *3* (3), 216–227.

_ , *Forecasting, structural time series models and the Kalman filter*, Cambridge university press, 1990.

Hasenzagl, Thomas, Filippo Pellegrino, Lucrezia Reichlin, and Giovanni Ricco, "A Model of the Fed's View on Inflation," *Review of Economics and Statistics*, forthcoming.

_ , Lucrezia Reichlin, and Giovanni Ricco, "Financial Variables as Predictors of Real Growth Vulnerability," CEPR Discussion Papers 14322, C.E.P.R. Discussion Papers January 2020.

Jarociński, Marek, "A note on implementing the Durbin and Koopman simulation smoother," *Computational Statistics & Data Analysis*, 2015, *91*, 1–3.

Jarociński, Marek and Michele Lenza, "An Inflation?Predicting Measure of the Output Gap in the Euro Area," *Journal of Money, Credit and Banking*, September 2018, *50* (6), 1189–1224.

Koopman, Siem J and James Durbin, "Fast filtering and smoothing for multivariate state space models," *Journal of Time Series Analysis*, 2000, *21* (3), 281–296.

Mariano, Roberto S and Yasutomo Murasawa, "A new coincident index of business cycles based on monthly and quarterly series," *Journal of applied Econometrics*, 2003, *18* (4), 427–443.

Orphanides, Athanasios, "Monetary Policy Rules Based on Real-Time Data," *American Economic Review*, September 2001, *91* (4), 964–985.

_ **and Simon van Norden**, "The Unreliability of Output-Gap Estimates in Real Time," *The Review of Economics and Statistics*, November 2002, *84* (4), 569–583.

Shumway, Robert H and David S Stoffer, "An approach to time series smoothing and forecasting using the EM algorithm," *Journal of time series analysis*, 1982, *3* (4), 253–264.

Sims, Christopher A., "Using a likelihood perspective to sharpen econometric discourse: Three examples," *Journal of Econometrics*, April 2000, *95* (2), 443–462.

Survey of Professional Forecasters, 2021.

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

Appendix A Data

A.1 GDP SPF

The Survey of Professional Forecasters includes expectations for real GDP in levels and growth rates. We decided not to use the official release for the expectation of real GDP in levels, because it is not adjusted for changes in the basis year, data revisions and in the seasonal adjustment mechanism.

Instead, we computed the one-year ahead SPF expectation for the growth rates and we used it jointly with the latest vintage of data available for real GDP to compute an adjusted prediction for the levels.

Appendix B Adaptive Metropolis-Within-Gibbs

B.1 Algorithm

The estimation algorithm is an improved version of the Metropolis-Within-Gibbs in Hasenzagl et al. (forthcoming) that employs the Single Component Adaptive Metropolis proposed in Haario et al. (2005).

This hybrid algorithm is structured in two blocks: (1) a Single Component Adaptive Metropolis (Haario et al., 2005) step for the estimation of the state-space parameters, (2) a Gibbs sampler (Koopman and Durbin, 2000; Jarociński, 2015) to draw the unobserved states conditional on the model parameters. Since we have non-stationary unobserved states, we use the Kalman filter with exact diffuse initial conditions (Koopman and Durbin, 2000; Durbin and Koopman, 2012) to compute the log-likelihood of the model. Finally, we used the priors in Hasenzagl et al. (forthcoming).

Algorithm: Adaptive Metropolis-Within-Gibbs

Initialisation

Let $\mathscr{K} := \{1, \ldots, n_k\}$ and denote as $\mathbf{P}(\mathscr{K})$ a function that returns a random permutation of $\mathscr{K}$ (uniformly taken from the full set of permutations of $\mathscr{K}$). Let also $\boldsymbol{\theta}_0$ be a n_k dimensional vector corresponding to the initial value for the Metropolis parameters. This vector is associated to a high posterior mass.

Single component adaptive metropolis

let $m = 1$

for $j = 1, \ldots, 10000$

let $\mathbf{S}_j = \mathbf{P}(\mathscr{K})$

for each k in $\mathbf{S}_j$

1. *Adaptation:* Update the standard deviation of the proposal distribution

$$\sigma_{k,j} = \begin{cases} 1 & \text{if } j \leq 10, \\ \exp\Big(\alpha_{k,j-1} - 0.44\Big)\sigma_{k,j-1} & \text{otherwise,} \end{cases}$$

where $\alpha_{k,j-1}$ is the acceptance rate for the iteration $j-1$, for the parameter at position $S_{k,j}$. Besides, 44% is the standard target acceptance rate for single component Metropolis algorithms.

2. *New candidate:* Generate a candidate vector of parameters $\boldsymbol{\theta}^*_m$ such that

$$\theta^*_{l,m} = \begin{cases} \theta_{l,m-1} & \text{if } l \neq k, \\ \underline{\theta} \overset{iid}{\sim} \mathcal{N}\Big(\theta_{l,m-1}, \sigma_{k,j}\Big) & \text{otherwise,} \end{cases}$$

   for $l = 1, \ldots, n_k$.

3. *Accept-reject:* Set

$$\boldsymbol{\theta}_m = \begin{cases} \boldsymbol{\theta}^*_m & \text{accept with probability } \eta_m, \\ \boldsymbol{\theta}_{m-1} & \text{reject with probability } 1 - \eta_m, \end{cases}$$

   where

$$\eta_m := \min\left(1, \frac{p\Big[\mathbf{Y} \mid \mathbf{f}(\boldsymbol{\theta}^*_m)^{-1}\Big]\, p\Big[\mathbf{f}(\boldsymbol{\theta}^*_m)^{-1}\Big]\, J(\boldsymbol{\theta}^*_m)}{p\Big[\mathbf{Y} \mid f(\boldsymbol{\theta}_{m-1})^{-1}\Big]\, p\Big[\mathbf{f}(\boldsymbol{\theta}_{m-1})^{-1}\Big]\, J\Big[\boldsymbol{\theta}_{m-1}\Big]}\right),$$

   $\mathbf{f}$ and J are defined below.

4. *Increase counter:* Increase m by one.

Gibbs sampling

For $j > 5000$ (burn-in period), use the univariate approach for multivariate time series of Koopman and Durbin (2000) to the simulation smoother proposed in Durbin and Koopman (2002) to sample the unobserved states, conditional on the parameters. In doing so, we follow the refinement proposed in Jarociński (2015).

Burn-in period

Discard the output of the first $j = 1, \ldots, 5000$ iterations.

Jacobian

As in Hasenzagl et al. (forthcoming) most parameters are bounded in their support (e.g. the variance parameters must be larger than zero). In order to deal with this complexity, this manuscript transforms the bounded parameters ($\boldsymbol{\Theta}$) so that the support of the transformed parameters ($\boldsymbol{\theta}$) is unbounded. Indeed, the Adaptive Metropolis-Within-Gibbs draws the model parameters in the unbounded space. At a generic iteration j, the following transformations have been applied to a generic parameter i with a Normal, Inverse-Gamma or Uniform prior:

$$\begin{aligned}
\theta_{i,j}^{N} &= \Theta_{i,j}^{N} \\
\theta_{i,j}^{IG} &= \ln(\Theta_{i,j}^{IG} - a_i) \\
\theta_{i,j}^{U} &= \ln\left(\frac{\Theta_{i,j}^{U} - a_i}{b_i - \Theta_{i,j}^{U}}\right),
\end{aligned}$$

where a_i and b_i are the lower and the upper bounds for the i-th parameter. These transformations are functions $f(\Theta) = \theta$, with inverses $f(\theta)^{-1} = \Theta$ given by:

$$\begin{aligned}
\Theta_{i,j}^{N} &= \theta_{i,j}^{N} \\
\Theta_{i,j}^{IG} &= \exp(\theta_{i,j}^{IG}) + a_i \\
\Theta_{i,j}^{U} &= \frac{a_i + b_i \exp(\theta_{i,j}^{U})}{1 + \exp(\theta_{i,j}^{U})}.
\end{aligned}$$

These transformations must be taken into account when evaluating the natural logarithm of the prior densities by adding the Jacobians of the transformations of the variables:

$$\begin{aligned}
\ln\left(\frac{d\Theta_{i,j}^{N}}{d\theta_{i,j}^{N}}\right) &= 0 \\
\ln\left(\frac{d\Theta_{i,j}^{IG}}{d\theta_{i,j}^{IG}}\right) &= \theta_{i,j}^{IG} \\
\ln\left(\frac{d\Theta_{i,j}^{U}}{d\theta_{i,j}^{U}}\right) &= \ln(b_i - a_i) + \theta_{i,j}^{U} - 2\ln(1 + \exp(\theta_{i,j}^{U})).
\end{aligned}$$

Appendix C Additional Real-Time Results

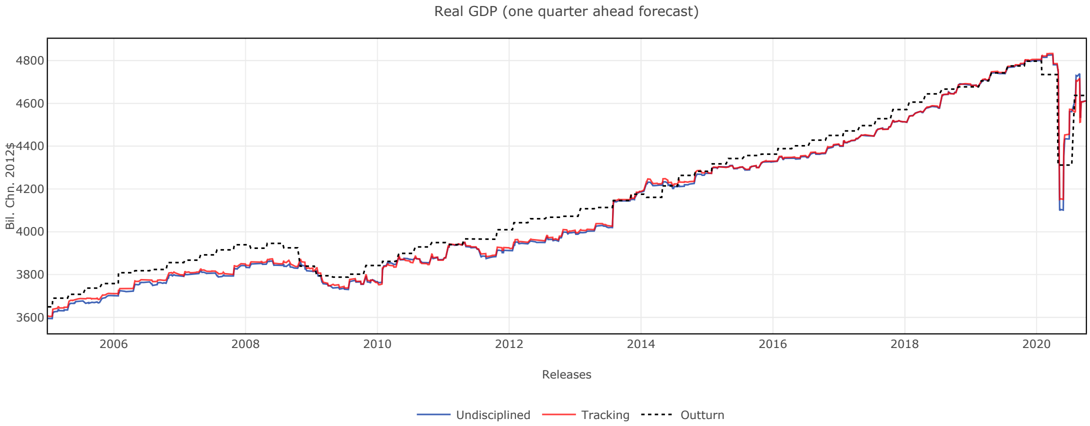

Figure 5: The chart reports the one quarter ahead, real time forecasts of Real GDP from the two models and compares them to the outturn. The out-of-sample evaluation starts in Jan-2005 and ends in Sept-2020.

Table 2: The first two rows of this table report the standard deviation of the output gap and potential output computed across vintages for each reference month and then averaged across reference months. The last two columns report the maximum absolute value of revisions computed for each reference month and then averaged across reference months.

	Output Gap		Potential Output	
	Undisciplined	Tracking	Undisciplined	Tracking
Mean of std dev	0.54	0.61	6.79	8.33
Mean of std dev (until 2005)	0.5	0.5	5.06	7.28
Mean of max revision	0.91	1.37	16.38	15.09
Mean of max revision (until 2005)	0.46	1.13	9.10	10.93

Figure 6: The chart reports the one month ahead, real time forecasts of inflation from the two models and compares them to the outturn. The out-of-sample evaluation starts in Jan-2005 and ends in Sept-2020.

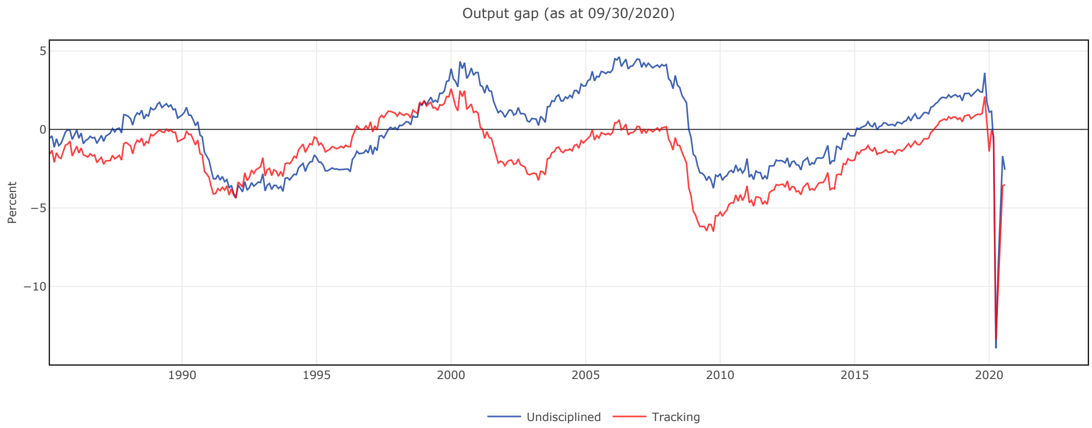

Figure 7: The chart compares the output gap estimates from the two models computed using the final (09/30/2020) data vintage from the out-of-sample forecasting exercise.