Federal Reserve Bank of Cleveland Working Paper Series

A Unified Framework to Estimate Macroeconomic Stars

Saeed Zaman

Working Paper No. 21-23

October 2021

Suggested citation: Zaman, Saeed. 2021. "A Unified Framework to Estimate Macroeconomic Stars." Federal Reserve Bank of Cleveland, Working Paper No. 21-23. https://doi.org/10.26509/frbc-wp-202123.

Federal Reserve Bank of Cleveland Working Paper Series
ISSN: 2573-7953

Working papers of the Federal Reserve Bank of Cleveland are preliminary materials circulated to stimulate discussion and critical comment on research in progress. They may not have been subjectto the formal editorial review accorded official Federal Reserve Bank of Cleveland publications.

See more working papers at: www.clevelandfed.org/research. Subscribe to email alerts to be notified when a new working paper is posted at: www.clevelandfed.org/subscribe.

A Unified Framework to Estimate Macroeconomic Stars

Saeed Zaman*
Federal Reserve Bank of Cleveland, USA
University of Strathclyde, UK

This version: October 10, 2021

Abstract

We develop a flexible semi-structural time-series model to estimate jointly several macroeconomic *"stars"* — i.e., unobserved long-run equilibrium levels of output (and growth rate of output), the unemployment rate, the real rate of interest, productivity growth, the price inflation, and wage inflation. The ingredients of the model are in part motivated by economic theory and in part by the empirical features necessitated by the changing economic environment. Following the recent literature on inflation and interest rate modeling, we explicitly model the links between long-run survey expectations and stars to improve the stars' econometric estimation. Our approach permits time variation in the relationships between various components, including time variation in error variances. To tractably estimate the large multivariate model, we use a recently developed precision sampler that relies on Bayesian methods. The by-products of this approach are the time-varying estimates of the wage and price Phillips curves, and the pass-through between prices and wages, both of which provide new insights into these empirical relationships' instability in US data. Generally, the contours of the stars echo those documented elsewhere in the literature – estimated using smaller models – but at times the estimates of stars are different, and these differences can matter for policy. Furthermore, our estimates of the stars are among the most precise. Lastly, we document the competitive real-time forecasting properties of the model and, separately, the usefulness of stars' estimates if they were used as steady-state values in external models.

JEL classification: C5, E4, E31, E24, O4

Keywords: state-space model, Bayesian analysis, time-varying parameters, natural rates, survey expectations, COVID-19 pandemic

---

I am extremely grateful to my advisors Gary Koop and Julia Darby for the valuable guidance and feedback throughout the research process. I am indebted to Joshua Chan, Todd Clark, Ana Galvão, Edward Knotek, Kurt Lunsford, Elmar Mertens, James Mitchell, Aubrey Poon, Ellis Tallman, Willem Van Zandweghe, Randal Verbrugge, and Ben Wong for helpful comments on the first draft. This research has benefited from discussions with Fabio Canova, Siddhartha Chib, Olivier Coibion, Sharada Davidson, Jesus Fernandez-Villaverde, Paolo Gelain, Yuriy Gorodnichenko, Stuart McIntyre, Eric Sims, and Ping Wu. I also thank participants at the Society for Computational Economics 27th International Conference 2021, NBER-NSF SBIES 2021, European Seminar on Bayesian Econometrics 2021, and the Cleveland Fed Brown Bag. The views expressed herein are those of the author and do not necessarily represent the views of the Federal Reserve Bank of Cleveland or the Federal Reserve System.

*Research Department, Federal Reserve Bank of Cleveland, Ohio, USA; and Department of Economics, University of Strathclyde, Glasgow, United Kingdom; email: saeed.zaman@clev.frb.org

1. Introduction

The estimates of long-run equilibrium levels of macroeconomic variables (often denoted with the "star" symbol) are of central importance in macroeconomics. These long-run levels are thought to reflect the fundamental structure of the economy in the absence of shocks. Hence, they are used as reference points, and deviations from these long-run levels reflect cyclical and idiosyncratic fluctuations. These stationary cyclical fluctuations typically serve as the source data for macroeconomic models (e.g., Smets and Wouters, 2003). However, effective identification of cyclical fluctuations in any macroeconomic aggregate requires knowledge of its long-run equilibrium.

In this paper, we estimate jointly seven macroeconomic stars of broader interest to macroeconomists and policymakers: the level of potential output (gdp-star), the growth rate of potential output (g-star), the long-run equilibrium levels of the unemployment rate (u-star), the real short-term interest rate (r-star), labor productivity growth (p-star), price inflation (pi-star), and the nominal wage inflation (w-star).[1] The assumption that a long-run equilibrium exists implies that in the long run, the economy is growing at potential, price inflation is growing at its trend rate, the unemployment rate has no cyclical pressure and only reflects structural factors, nominal wages grow at a rate equal to the sum of labor productivity growth and price inflation, and the real interest rate reflects the rate consistent with output growing at potential and stable inflation.[2]

In practice, determining the values of these stars is difficult because stars and some of their determinants are unobserved (see Weber, Lemke, and Worms, 2007). To infer the estimates of the stars, economists apply a range of econometric methods to observable historical data.[3] The multivariate unobserved components (UC) models, which are statistical models that use economic theory to frame the empirical specification, have been shown to provide reasonable estimates of the stars (e.g., Kuttner, 1994; Laubach and Williams, 2003; Basistha and Nelson, 2007; Chan, Koop, and Potter, 2016). Hence, they are the dominant methods for obtaining estimates of the stars. However, with few exceptions, the popular multivariate UC models that estimate time-varying stars focus on a small number of observables, often just two or three, and minimal structure (e.g., Laubach and Williams, 2003). Studies that entertain more variables have abstracted from important empirical features such as time-varying parameters and stochastic volatility (e.g., Hasenzagl, Pellegrino, Reichlin, and Ricco, 2020; Del Negro et al.,

[1]The subset of these stars, p-star, g-star, u-star, and r-star, reflect the fundamental structural features of the economy, whereas others, pi-star and w-star, are thought to be influenced by central banks and monetary policy.

[2]The literature has referred to the concept of long-run equilibrium using different terminologies, such as "natural," "neutral," "trend," "steady-state," and "long-run." There are subtle differences among them, but they can be interpreted as the same for the purpose of this paper. In some studies, especially those using dynamic stochastic general equilibrium (DSGE) models, the concept of the natural rate refers to medium-horizon equilibrium, and in these same models, the concept of steady state is used to refer to the long-run equilibrium.

[3]The methods range from statistical univariate filters (e.g., Hodrick and Prescott, 1997; Ashley and Verbrugge, 2008) to multivariate models, including semi-structural time-series models (e.g., Pescatori and Turunen, 2016; Morley and Wong, 2020), and fully structural DSGE models (e.g., Del Negro, Giannone, Giannoni, and Tambalotti, 2017.

2017; Fleischman and Roberts, 2011). A priori, one would expect a framework based on greater information that explicitly permits (contemporaneous) interactions between stars and between cyclical components, and a richer structure to provide more reliable estimates of the objects of interest (e.g., stars) than frameworks that ignore them.

Accordingly, in this paper, we take on the challenge of jointly estimating several macroeconomic stars simultaneously, including g-star (and gdp-star), u-star, r-star, p-star, pi-star, and w-star, using a semi-structural time-series model. For each star, we formulate a rich structure whose elements are guided by past research and informed by economic theory. For example, econometric estimation of r-star is informed by various sources: the investment-savings (IS) equation, the Taylor-type rule, the equation linking g-star and r-star, and the equation relating r-star to survey expectations. We allow for time variation in important macroeconomic relationships and error variances. Fernández-Villaverde and Rubio-Ramírez (2010), Koop and Korobilis (2010), and Carriero, Clark, and Marcellino (2019), among many others, highlight the importance of allowing for stochastic volatility in macroeconomic models.[4] Incorporating these empirical features should better distinguish between cyclical fluctuations and lower-frequency movements in the macroeconomic aggregates considered in this paper.

We extend the Chan, Clark, and Koop (2018) – henceforth CCK – approach of using long-run survey expectations to improve pi-star precision to other macroeconomic stars. Specifically, for each macroeconomic variable of interest, we explicitly model the link between the unobserved "star" and the expectations about the star contained in the Blue Chip survey of economic forecasters (or as reported by the Congressional Budget Office [CBO] when the survey estimate is not available).[5] In a high-dimensional model like ours, the use of long-run survey expectations, which are direct measures of stars, could help anchor model-based estimates of stars to reasonable values (especially in times of heightened uncertainty) and potentially improve the precision of the estimates. We estimate the feature-rich UC model using Bayesian techniques, specifically the efficient sampling techniques developed in Chan, Koop, and Potter (2013) and the precision sampler proposed in Chan and Jeliazkov (2009).

All in all, the combination of time-varying parameters, SV, joint modeling of multiple stars, implementation of an expanded structure, and allowing for a direct connection between stars and long-term survey expectations is what differentiates our UC model from those in the existing literature. Many popular UC models could be viewed as special cases of our larger UC model, which facilitates model comparison. We note that among the stars considered, w-star has received less attention in the literature. Our UC model's ability to provide real-time estimates of w-star and its model-based decomposition into its determinants p-star and pi-star, as implied

[4]To the best of our knowledge, features such as time-varying parameters and stochastic volatility (SV) have not been implemented in a multivariate UC framework beyond a system consisting of at most four variables (e.g., Berger, Everaert, and Vierke, 2016). Stock and Watson (2016) implement a UC model with SV (but not time-varying parameters) to estimate pi-star using 18 disaggregated components of aggregate inflation.

[5]The long-run survey expectations can be thought of as a hybrid forecast because it combines judgment based on a range of information and forecast derived from a range of modeling approaches. Our use of such a hybrid forecast implicitly serves as an additional channel through which the issue of omitted variable bias is mitigated.

by economic theory, is a novel contribution. This specific decomposition is useful to monetary policymakers, who often refer to developments in nominal wages to support their forecasts and related discussions on price inflation and employment.

Our results indicate that there are payoffs to modeling stars jointly using a larger multivariate UC model. The metric of Bayesian model comparison generally favors our larger UC model over smaller-scale UC models. The model yields credible estimates of stars and the output gap. For example, the output gap estimate is similar to the CBO estimate based on a production function approach. Generally, the contours of stars echo those documented elsewhere in the literature but at times are different, and these differences can matter for policy. For example, let's consider pi-star. From 2000 to 2010, our UC model has an estimate of the pi-star stable at or close to 2%, whereas pi-star from a popular univariate model (of Stock and Watson, 2007) displays notable fluctuations around 2%, and the bivariate model of CCK indicates a stable pi-star about a few tenths below 2%. These differences matter for central banks tasked with inflation targeting. Compared to some of the popular UC models and the smaller-scale restricted variants of our larger UC model, the precision estimates of the stars and the output gap from our UC model are among the most precise, where precision is measured as the width of 90% credible intervals. The model's reliance on long-term survey expectations data is the key reason for this improved precision.[6] Survey expectations have played a crucial role in guiding the model-based assessment of stars during the COVID-19 pandemic, a period of heightened uncertainty. The accuracy of our UC model's real-time point and density forecasts rivals and, in some cases, outperforms hard-to-beat benchmarks, including small-scale UC models.

We also demonstrate the usefulness of our estimated stars as terminal values for external models. Previous research shows that forecasting models, such as steady-state vector autoregressive (VAR) models, often improve their forecast accuracy by using external information about steady states informed by long-run survey expectations (e.g., Wright, 2013). Using a real-time forecasting comparison, we show that if we were to inform steady states in a VAR with the stars from our UC model, gains in forecast accuracy for some of the variables would be achieved compared to the standard approach relying on survey expectations. Hence, our framework provides a potential source for obtaining the stars' estimates in real time. An advantage of our framework compared to surveys is that it provides estimates of stars (steady states) for variables not covered by surveys (e.g., w-star) and offers both point and uncertainty estimates.

We summarize three additional findings. First, we find that the empirical evidence in the link between r-star and g-star, as implied by theory, is weak (consistent with Hamilton, Harris, Hatzius, and West, 2016; Lunsford and West, 2019), but by bringing survey expectations into the model, the link becomes stronger (providing support to Laubach and Williams, 2016). Second, our results indicate economically and statistically significant evidence of time variation in the model parameters capturing macroeconomic relationships and strong support for SV's inclusion

[6] Our precision estimates are on a par with recent studies highlighting the improved precision of stars derived from their approaches (e.g., r-star by Del Negro et al., 2017; u-star by Crump, Eusepi, Giannoni, and Şahin, 2019; pi-star by CCK).

in the model equations. It lends support to the popular narratives of: "The price Phillips curve has weakened over time," "The wage Phillips curve is alive," and "There is weakening in the procyclicality of labor productivity." Third, a comparison between final and real-time estimates of the stars indicates that their broad movements have generally tracked each other closely. We view this as a valuable finding because it suggests that we have made some progress in mitigating the well-known difficulties associated with the real-time estimation of the stars.

In recent years, several papers have provided estimates of the stars using UC models with more indicators and or an expanded structure. For example, Johannsen and Mertens (2021) [henceforth JM], Pescatori and Turunen (2016), Del Negro et al. (2017), Brand and Mazelis (2019), González-Astudillo and Laforte (2020), among others, have examined the roles of additional determinants in explaining r-star.[7] None of these studies feature time-varying parameters, and only JM allows for SV, but their model size is significantly smaller than ours. Chan, Koop, and Potter (2016) [henceforth CKP] illustrate the value of modeling u-star and pi-star as bounded random walk processes in a bivariate Phillips curve. More recently, using fixed-parameter UC models, Crump et al. (2019) estimate u-star by combining a range of labor market indicators across demographic groups and survey expectations of inflation, and Hasenzagl et al. (2020) jointly estimate pi-star, u-star, gdp-star (and output gap).[8]

The paper is organized as follows. The next section describes in detail the econometric model and its variants. Section 3 describes the data and estimation. Section 4 presents and discusses in detail the estimates of stars and other model parameters. Section 5 reports the real-time forecasting results and a discussion comparing real-time and final estimates of stars. Section 6 illustrates the ability of the model to handle the COVID-19 pandemic data. Section 7 concludes. This paper has a supplementary online appendix that lists detailed Bayesian estimation steps and some additional results.

2. Empirical Macro Model and Variants

The ingredients of our macroeconomic econometric model are guided both by economic theory and by empirical considerations – namely, features that previous research has demonstrated to be empirically relevant. These features include stochastic volatility and time-varying parameters, which in turn imply time-varying predictability. Collectively, these empirical features permit modeling changing macroeconomic relationships in a flexible way.

We represent our empirical model using six sets of equations, which we denote blocks. These

[7] Pescatori and Turunen (2016) enrich the underlying structure to estimate r-star. In particular, to extract a reliable estimate of the output gap, they bring additional information from the Congressional Budget Office's (CBO) estimate of the output gap by treating it as a noisy measure of the "true" output gap.

[8] Morley and Wong (2020) and Chan (2019) propose an alternative modeling framework based on VARs to estimate the long-run equilibrium values. The advantage of the VAR-based framework is the ability to handle larger amounts of information conveniently and flexibly compared to UC models. On the other hand, the advantage of UC modeling, as emphasized by CKP, is the availability of the direct estimates of stars, which in the case presented here proves quite convenient to allow for direct modeling of the relationships between various stars.

six blocks, which allow for *contemporaneous interactions* between them, characterize the joint dynamics of the unemployment rate, output growth, labor productivity growth, price inflation, nominal wage inflation, nominal interest rate, and corresponding stars. To be sure, the model assumes that all innovations are uncorrelated both serially and across equations. However, we emphasize that any assumed current period correlations between the cyclical components and or between stars are directly modeled via the model equations that define the contemporaneous relationships between the components (e.g., the cyclical output gap at time t with the cyclical unemployment gap at t; r-star and g-star).

Before we describe the model, we provide some necessary background information, including the econometric definition of the star and the usefulness of long-run survey expectations in the estimation of stars.

2.1. ***The econometric notion of a long-run equilibrium***

Following CCK, Mertens (2016), Lee and Nelson (2007), and Morley (2002), among many, this paper defines the long-run equilibrium (or star) of a particular macroeconomic series as its infinite-horizon forecast conditional on the current information set. This definition of a star is consistent with the notion of Beveridge-Nelson trend decomposition, and an extensive literature has adopted this approach to estimate stars. Equivalently, as commonly defined in the trend estimation literature, the infinite-horizon forecast could be viewed as an estimate of trend conditional on the current information set (e.g., CCK, Garcia and Poon, 2018, Mertens, 2016, Lee and Nelson, 2007, Morley, 2002). As discussed in Mertens (2016), among others, different information sets would likely yield different estimates of the infinite-horizon forecast (or trend). Mertens showed that including survey projections of long-term inflation (hereafter long-run survey forecasts) in the information set led to more precise and forward-looking estimates of trend inflation.

The link between the infinite-horizon forecast and the underlying trend is described well by the unobserved components (UC) model (see Laubach and Williams, 2003; Lee and Nelson, 2007; Mertens, 2016; CCK). In a UC model, a series (Y_t) is typically represented as the sum of a nonstationary trend component Y_t^*, which is assumed to evolve slowly and a stationary cycle Y_t^c, whose infinite-horizon conditional expectation is assumed to be zero. Accordingly,

$$Y_t = Y_t^* + Y_t^c. \tag{1}$$

The trend component Y_t^* is interpreted as the limiting forecast of the series (conditional on the information set I_t) as the forecast horizon tends to infinity,

$$\lim_{j\to\infty} \mathbb{E}[Y_{t+j}|I_t] = Y_t^*. \tag{2}$$

Differencing the above equation yields,

$$Y_t^* = Y_{t-1}^* + \lim_{j\to\infty} \mathbb{E}[Y_{t+j}|I_t] - \lim_{j\to\infty} \mathbb{E}[Y_{t+j}|I_{t-1}] = Y_{t-1}^* + e_t, \;\; e_t \sim N(0, \sigma_e^2). \tag{3}$$

which suggests a random walk process for the trend Y_t^*.[9] It also suggests a stationary, ergodic mean-zero process for Y_t^c.

Intuitively, the above set of assumptions implies that once the effects of the shocks have fully played out, the macroeconomic series of interest, Y_t, gravitates to its underlying trend level, Y_t^*.

As discussed in CCK, various statistical and econometric models could fit within the above-specified decomposition. This paper formulates a specific unobserved components time-series model and its variants.

2.2. ***The role of survey expectations***

As discussed in the introduction, an important contribution of this paper is to provide a direct role for long-run survey expectations in refining the stars' estimates. Specifically, we follow the approach of CCK (and Pescatori and Turunen, 2016). These papers explicitly estimate an equation linking the observed measure of a long-run forecast obtained from external sources (survey in the case of CCK and CBO projection of the output gap in Pescatori and Turunen to an unobserved object of interest. We extend their approach to the macroeconomic variables considered in this paper.

Several papers have documented an essential role of long-run survey (and institutional) forecasts in helping refine the econometric estimation of model parameters, including the latent components (e.g., pi-star: Kozicki and Tinsley, 2012; Mertens, 2016; Mertens and Nason, 2020; CCK; gdp-star: Pescatori and Turunen, 2016; r-star: Del Negro et al., 2017). Specifically, Mertens and Nason (2020), CCK, Mertens (2016), and Kozicki and Tinsley (2012), in using different methodologies (in combining survey data with model forecasts) to estimate the trend in US inflation, show that long-run survey forecasts of inflation deliver crucial additional information (beyond the recent inflation history) in refining trend estimates and improving model fit. In a similar vein, Pescatori and Turunen (2016) document the usefulness of the CBO's projection of the potential output gap in improving their model's output gap precision. It is this particular literature that motivates us to consider long-run survey forecasts in our large-scale econometric model.

The advantage of survey (and institutional) forecasts stems from the fact that they could be viewed as hybrid forecasts, i.e., a combination of judgment and forecasts derived from various modeling approaches. The fact that human judgment enters into survey expectations is an

[9] The commonly adopted assumption of modeling Y_t^* as a random walk is partly due to consensus among macroeconomists that the factors driving the long-run equilibrium levels are perceived to be quite persistent (e.g., Lee and Nelson, 2007). In practice, the assumption of a (driftless) random walk has generally worked quite well, in that it has provided reasonable estimates of the stars (e.g., Clark, 1987; Kuttner, 1994; Laubach and Williams, 2003; Mertens, 2016).

important reason for their success. As discussed by Kozicki and Tinsley (2012) and others, the good forecasting properties are partly because survey participants have at their disposal a wide range of indicators, including central bank communications, and information about changes in the tax laws, etc. The patterns gleaned from this large information set can help shape opinions, including any perceived structural change, which can immediately influence expectations about the long run.

The usefulness of this hybrid forecast measure is appealing from both theoretical and practical perspectives. From a theoretical standpoint, the notion of such a type of hybrid forecast is well-aligned with the New-Keynesian view that emphasizes the importance of allowing for interlinkages between the current state of the economic variables and their expectations about the future in determining the relevant stars (see Weber, Lemke, and Worms, 2007).

From a practical point of view, such a hybrid measure helps limit the number of variables (i.e., information) that need to be brought into the model. Put differently, explicitly utilizing long-run survey expectations through an equation linking these expectations to the corresponding model's latent objects could be thought of as a shortcut to enrich the necessary information set used in model estimation. Furthermore, in high-dimensional models, such as the one developed in this paper, the use of long-run survey projections, which are targeted and direct measures of stars, could help anchor model-based estimates of stars to reasonable values and have the potential to improve precision of the estimates.

Accordingly, in this paper, with the exceptions of nominal wage inflation and labor productivity, for each of the remaining four variables, we model a direct link between long-run survey projections (or the long-run CBO projections in the years for which survey projections are unavailable) and the corresponding star using the following econometric equations:[10]

$$Z_t^j = C_t^j + \beta^j j_t^* + \varepsilon_t^{zj}, \;\; \varepsilon_t^{zj} \sim N(0, \sigma_{zj}^2), \;\; j = \pi, u, g, r \tag{4}$$

$$C_t^j = C_{t-1}^j + \varepsilon_t^{cj}, \;\; \varepsilon_t^{cj} \sim N(0, \sigma_{cj}^2), \;\; j = \pi, u, g, r \tag{5}$$

where π refers to price inflation, u refers to the unemployment rate, g refers to real GDP growth, r refers to the real interest rate, Z_t^j refers to the long-run survey forecast corresponding to the variable j, and j_t^* is the unobserved j star.

C_t^j is a time-varying intercept assumed to evolve as an RW process to possibly capture the permanent wedge between the survey estimate and the model-based star. This wedge can arise for several reasons, including the fact that star is assumed to be the infinite-horizon forecast, whereas the survey forecast refers to the average forecast for the five-year period starting seven years into the future in the case of BC and the ten-year-ahead forecast in the case of the SPF (for price inflation).

[10] For the long-run inflation forecast, we use the Survey of Professional Forecasters (SPF)/PTR, and for the long-run forecasts of the other three variables, we use the Blue Chip (BC) survey.

The above set of equations defines a simplistic and flexible relationship between the long-run survey expectations and the star.[11]

2.3. ***Unemployment block***

The long-run equilibrium level of unemployment (u-star) is the unemployment rate that prevails when output is growing at potential, and the economy adds jobs so as to maintain the full-employment level. This nonzero equilibrium level of unemployment is primarily the result of labor market imperfections caused by the frictional and structural aspects of the labor markets.

As discussed in Crump et al. (2019), two approaches are commonly used to estimate u-star. The first approach applies UC modeling to detailed labor market data (such as job vacancies, firms' recruiting intensity, demographic changes, flows into and out of unemployment) to extract respective trends. These trends are used to construct implied estimates of u-star (e.g., Daly, Hobijn, Şahin, and Valletta, 2012; Davis, Faberman, and Haltiwanger, 2013; Barnichon and Mesters, 2018; Tasci, 2012). The second approach uses a combination of information from prices (and or nominal wages, survey expectations) and the estimated Phillips curve relationship between price inflation and the aggregate unemployment rate to back out the estimate of u-star.[12] Various modeling techniques ranging from parsimonious UC modeling to structural DSGE models are applied to estimate the Phillips curve relationship (e.g., Staiger, Stock, and Watson, 1996; Orphanides and Williams, 2003; Lee and Nelson, 2007; Galí, 2011; CKP).

Following CKP, we posit that the observed unemployment rate is decomposed into a (bounded) RW trend component (u-star) and a stationary cyclical component. The cyclical component is modeled as an AR(2) process. The use of a parsimonious (time-invariant) AR2 process to identify the cyclical component of the unemployment rate is a commonly used assumption, in our case motivated by a recent string of empirical studies, e.g., Lee and Nelson (2007), Galí (2011), Stella and Stock (2015), CKP, Tallman and Zaman (2017), and Galí and Gambetti (2019), who all document reasonable patterns of the cyclical unemployment component (i.e., movements in this component correlate quite well with the NBER business cycle).[13]

More generally, the assumption of an AR2 to model the cyclical component of macroeconomic variables has a long tradition, going back to at least Clark (1987). However, because we are also modeling the output gap (i.e., the level of real GDP minus the level of potential real GDP), we depart from the previous literature by augmenting the AR2 unemployment gap with

[11] Note that we adopt a relatively less flexible relationship between survey forecasts and stars than CCK.

[12] As mentioned in Crump et al. (2019), one of the criticisms of this approach is that it will be affected by the breakdown of the Phillips curve relationship post-2007. However, by allowing time variation in the coefficients capturing the price and wage Phillips curve relationships, as we do, our approach should face less of a problem. In addition, as illustrated in Del Negro, Giannoni, and Schorfheide (2015) and Clark and Doh (2014) including information from long-run survey expectations of inflation (as we do) should further help capture the inflation behavior in the post-2007 period.

[13] CKP explore the empirical importance of allowing for time variation in the parameters of an AR2 process, and find that the data prefer the time-invariant AR2 process, hence validating the widely used assumption of a simple AR2 process.

the output gap (denoted ogap) as an additional explanatory variable. Sinclair (2009), Grant and Chan (2017a), and Berger, Everaert, and Vierke (2016), among several others, document the empirical importance of jointly modeling the unemployment rate gap and the output gap.

Similarly, as shown later, we add information from the unemployment gap when modeling the output gap. We find that the joint modeling of the output gap and the unemployment gap is empirically and economically useful (confirming previous research findings, e.g., Fleischman and Roberts, 2011). In addition, joint modeling of both the output gap and the unemployment gap allows us to estimate the strength of the relationship between the two cyclical components, popularly known as Okun's law. In equation (7), the coefficient ϕ_u captures the contemporaneous relationship between the output gap and the cyclical unemployment rate gap. The estimate, $\frac{1-\rho_1^u-\rho_2^u}{\phi^u}$, could be interpreted as the Okun's law coefficient.

We note that when jointly modeling output and the unemployment rate, most researchers assume a common cyclical component between the two. However, in light of the empirical evidence that cyclical unemployment displays more persistence than the output gap (e.g., Berger, Everaert, and Vierke (2016)), we model two separate cycles linked to each other via the Okun's law relationship. As shown in Berger, Everaert, and Vierke (2016), in a specification that entertains two separate cycles (cyclical unemployment and the output gap), the data support a time-invariant parameter describing the Okun's law relationship. In contrast, a specification with a common cyclical component favored a time-varying Okun's law relationship (adding support to Knotek II, 2007).

In a nutshell, the inference of Berger, Everaert, and Vierke (2016) suggests that time variation in the parameter capturing Okun's law reflects the sluggish response of the cyclical unemployment rate to movements in the output gap. Once they allowed for a sluggish response of the cyclical unemployment rate by adding persistence, via a one-period lag of the cyclical unemployment rate, evidence of a time-varying Okun coefficient disappears. We found similar evidence. Hence, we adopt the modeling of two separate cycles and a time-invariant Okun's law relationship for our baseline setup; this has the added advantage of requiring estimation of significantly fewer parameters.[14]

With the exception of CKP, most of the literature models u-star as a driftless RW. The use of an unrestricted RW process has empirically been shown to work well, but CKP show that modeling u-star as a bounded RW process is even better. They use bounds because, by construction, the unemployment rate is a bounded variable, which implies that the long-run equilibrium in the labor market would restrict the movements in u-star within a bounded interval.[15] Accordingly, we model u-star as a bounded RW, where the bounds' values are fixed

[14]Bayesian model comparison assessment slightly preferred the approach of two separate cycles with a time-invariant Okun's law compared to a common cycle with a time-varying Okun's law parameter. Also, we note that Sinclair (2009) and Grant and Chan (2017b) model two separate cycles, linked through a time-invariant parameter.

[15]CKP argue that economic forces that govern the movements in u-star are slow-moving and those forces would not cause the unemployment rate to fall to levels close to zero or to levels that are higher than the previous peaks caused by recessions.

at 3.5% (lower bound) and 7.5% (upper bound).[16] [17]

$$U_t = U_t^* + U_t^c \tag{6}$$

$$U_t - U_t^* = \rho_1^u(U_{t-1} - U_{t-1}^*) + \rho_2^u(U_{t-2} - U_{t-2}^*) + \phi^u ogap_t + \varepsilon_t^u, \;\; \varepsilon_t^u \sim N(0, \sigma_u^2) \tag{7}$$

where, $\rho_1^u + \rho_2^u < 1$, $\rho_2^u - \rho_1^u < 1$, and $|\rho_2^u| < 1$

$$U_t^* = U_{t-1}^* + \varepsilon_t^{u*}, \;\; \varepsilon_t^{u*} \sim TN(a_u - U_{t-1}^*, b_u - U_{t-1}^*; 0, \sigma_{u*}^2) \tag{8}$$

where the notation $TN(a, b; \mu, \sigma^2)$ refers to normal distribution with mean μ and variance σ^2 but truncated in the interval (a, b).

Lastly, the equation relating long-run survey projections of the unemployment rate, Z^u to U^*,

$$Z_t^u = C_t^u + \beta^u U_t^* + \varepsilon_t^{zu}, \;\; \varepsilon_t^{zu} \sim N(0, \sigma_{zu}^2) \tag{9}$$

$$C_t^u = C_{t-1}^u + \varepsilon_t^{cu}, \;\; \varepsilon_t^{cu} \sim N(0, \sigma_{cu}^2) \tag{10}$$

2.4. ***Output block***

We are interested in both the potential output (i.e., gdp^*) and the growth rate in potential output (i.e., g^*). To feasibly estimate both of these latent processes, we follow the commonly adopted approach, which decomposes the level of aggregate output into the level of potential output and a cyclical component (output gap), where the cyclical component is defined as the deviation of the observed aggregate output level from potential output. This simple decomposition has a long tradition going back to Harvey (1985) and Clark (1987).

$$gdp_t = gdp_t^* + ogap_t \tag{11}$$

where $gdp \equiv log(GDP)$ and gdp^* refers to potential output, which is unobserved.

A common approach to model gdp^* is to assume a random walk with a time-varying drift term (i.e., the local-level trend process), where the time-varying drift term (interpreted as g^*) is assumed to follow a random walk process (to allow for a stochastic g^*). More recently, Grant and Chan (2017b) show that a model where gdp^* is assumed to follow a second-order Markov

[16] These values are informed by estimating the CKP model over our estimation sample, and are close to values reported in CKP based on their estimation sample. As a further check, most estimates of the u-star reported in the commonly cited literature fall within the bounds we use in this paper.

[17] Mertens (2014), Berger, Everaert, and Vierke (2016), Stella and Stock (2015), and Tallman and Zaman (2017) provide evidence in support of SV in the cyclical components of the unemployment rate and the output gap. Accordingly, in a future revision of the paper, we will allow for SV in the cyclical components of the unemployment rate (i.e., equation 7) and the output gap (i.e., equation 14).

process fits the data better compared to the model where gdp^* is assumed to be a random walk with time-varying drift. Hence, we follow Grant and Chan (2017b) and model gdp^* as

$$gdp_t^* = 2gdp_{t-1}^* - gdp_{t-2}^* + \varepsilon_t^{gdp*}, \;\; \varepsilon_t^{gdp*} \sim N(0, \sigma_{gdp*}^2) \tag{12}$$

Which can be re-written as

$$\triangle gdp_t^* = \triangle gdp_{t-1}^* + \varepsilon_t^{gdp*}$$

If we define $g_t^* \equiv \triangle gdp_t^*$, where $\triangle$ is the first difference operator, then,

$$g_t^* = g_{t-1}^* + \varepsilon_t^{gdp*} \tag{13}$$

An advantage of modeling g^* as a second-order Markov process compared to an RW with time-varying drift is that it requires estimating a single shock parameter (σ_{gdp*}^2), as opposed to two for the latter (one for the shock to gdp^* and the other for the shock to the time-varying drift, aka g^*). This modeling assumption implies that all permanent shocks to output are attributed as shocks to g^*.[18]

The cyclical component, ogap, is assumed to be a stationary AR(2) process augmented with additional explanatory variables: the real interest rate gap and the unemployment gap,

$$ogap_t = \rho_1^g(ogap_{t-1}) + \rho_2^g(ogap_{t-2}) + a^r(r_t^L - r_t^*) + \lambda^g(U_t - U_t^*) + \varepsilon_t^{ogap} \tag{14}$$

where, $\varepsilon_t^{ogap} \sim N(0, \sigma_{ogap}^2)$, $\rho_1^g + \rho_2^g < 1$, $\rho_2^g - \rho_1^g < 1$, and $|\rho_2^g| < 1$

Equation (14) could be interpreted as defining an IS-curve (as in LW and subsequent papers modeling r-star) that allows feedback from the real interest rate gap to the output gap (i.e., the real interest rate gap responds to economic slack). The IS equation is inspired by LW but with two modifications. First, instead of using the interest-rate gap based on the short-term real rate of interest, we use the long-term real interest rate (as in González-Astudillo and Laforte, 2020). Specifically, the long-term real interest rate, r^L, is constructed as the difference between the nominal yield on a 10-year Treasury bond and the 10-year inflation expectations (i.e., the PTR series for PCE inflation).[19]

[18] Note: the second-order Markov process for gdp* could be thought of as a limiting case of the process. Assuming gdp* is an RW with time-varying drift term, where the variance of the shock to a gdp* goes to zero, and the shock to the time-varying drift term (g-star) is the only relevant driver governing the evolution of gdp* and g-star (see Grant and Chan, 2017b).

[19] We assume a $termpremium^* = 0$ for the sake of simplicity. In a future revision of the paper, we will model $termpremium^*$ as the average of the differential between the long-term interest rate (i.e., 10-year Treasury bond) and the federal funds rate, similar to (Johannsen and Mertens, 2021). We also experimented with an alternative specification, in which the interest rate gap is constructed as the difference between the short-term federal funds rate and the first lag of four-quarter trailing PCE inflation, similar in spirit to LW. Based on model fit (log marginal likelihood), this specification was slightly inferior compared to the Base specification. It is worth noting

In theoretical models, the long-term interest rate influences household consumption decisions and business investment decisions. Second, to improve the econometric estimation of the output gap, we enrich the IS equation by bringing in information from the unemployment gap (from the unemployment block) as an explanatory variable.[20] This latter addition is motivated by the approach taken in a long list of papers (e.g.,Morley and Wong, 2020; Grant and Chan, 2017a; Fleischman and Roberts, 2011; Sinclair, 2009; Clark, 1987) that demonstrate the usefulness of the unemployment rate in improving the econometric estimation of the output gap. As mentioned earlier, in the equation for the unemployment gap, we add the output gap to improve the former's estimation. The coefficient λ^g captures the contemporaneous relationship between the output gap and the unemployment gap. The parameter a^r relates the output gap to the real interest rate gap.

We note that innovations ε^2_{gdp*} and ε^2_{ogap} are uncorrelated. In an important contribution, Morley, Nelson, and Zivot (2003), who assume a deterministic g-star, show that this assumption matters for estimating potential output. However, Grant and Chan (2017a) show that in their specification, once a stochastic g-star is allowed for, the correlation between ε^2_{gdp*} and ε^2_{ogap} goes to zero. They also show that the model without correlation performs comparably to the model with correlated innovations based on Bayesian model comparison. Accordingly, to keep estimation tractable, we assume uncorrelated innovations.

The equation linking survey projections of the potential growth rate, Z^g to g^* [21]

$$Z^g_t = C^g_t + \beta^g * 4 * g^*_t + \varepsilon^{zg}_t,\ \ \varepsilon^{zg}_t \sim N(0, \sigma^2_{zg}) \tag{15}$$

$$C^g_t = C^g_{t-1} + \varepsilon^{cg}_t,\ \ \varepsilon^{cg}_t \sim N(0, \sigma^2_{cg}) \tag{16}$$

2.5. ***Productivity block***

Since the publication of Adam Smith's influential book, The Wealth of Nations, it is widely acknowledged that long-run productivity growth is the primary determinant of long-run changes in living standards. Therefore, estimates of the long-run level of (labor) productivity growth

that had the longer history of long-term inflation expectations data been available at the time of the writing, LW would have constructed the interest rate gap using the long-term interest rate (see page 1064 in LW).

[20]Model fit, the precision metric for u-star and the output gap, and the plausibility of the estimates of output gap strongly support the joint modeling of the output gap and the unemployment gap. We note that LW estimated an alternative specification in which they added information from the labor market (hours worked) and found that doing so improved the precision of the estimated output gap; however, that improved precision did not spill over to the r-star estimate (the focus of their paper) and therefore in their baseline specification they omit the labor market variable.

[21]We could potentially bring in additional information from the CBO's projection of the level of potential GDP to improve the econometric estimation of the level of potential real GDP. However, as shown in the results section, the implied estimates of the output gap from our multivariate framework (with or without survey data) are remarkably similar to the CBO's estimates suggesting, the limited value of bringing in additional information from CBO projections.

(p-star) have received considerable discussion in the past decades and are of great interest. Labor productivity is defined here as output per hour worked.

Furthermore, the estimates of p-star are an important input into policymaking, as they are used to gauge the appropriateness of the monetary policy stance. P-star is important for monetary policy because standard macroeconomic models tightly connect p-star to the long-run level of the real interest rate (i.e., r-star). In these models, a lower level of p-star implies a lower level of r-star, and a higher level of p-star implies a higher r-star (see Lunsford, 2017). However, based on post-1960 data, Lunsford found no statistical evidence supporting the link between p-star and r-star.

Several papers have endeavored to estimate the long-run level of productivity growth using various statistical and econometric models. A subset of those papers have documented support in favor of a regime-switching framework to model long-run labor productivity growth (e.g., Kahn and Rich, 2007). Also, to extract more precise and timely estimates of p-star, various authors (e.g., Kahn and Rich, 2007; Roberts, 2001) have proposed using additional variables alongside labor productivity (e.g., real compensation, real consumption, and average hours worked). On the other hand, Edge, Laubach, and Williams (2007) show that estimates of long-run productivity growth obtained from a simple trend-cycle univariate model solved with the Kalman filter do an adequate job of mirroring the long-run projections of productivity growth reported in the SPF and institutional forecasts (e.g., CBO).[22]

On closer inspection, the ability of the Kalman filter to echo the predictions of the professional forecasters is not surprising. Productivity growth is a notoriously volatile series and is subject to extreme revisions from one vintage to another. So, distinguishing highly persistent fluctuations from truly permanent changes is a difficult job for professionals and models alike. Jacobs and van Norden (2016) discuss in detail some of these challenges when working with productivity data.

The findings in Lunsford (2017), Edge, Laubach, and Williams (2007), and Jacobs and van Norden (2016) motivate the formulation of a parsimonious structure for the productivity block relative to other blocks of the model. In particular, we abstract from explicit modeling of direct links between p-star and r-star and between p-star and g-star.[23] But along other dimensions, our formulation is richer than that used in the cited literature, as we allow for time-varying parameters, including stochastic volatility. Specifically, the productivity gap, which is defined as (nonfarm) labor productivity growth[24] (quarterly annualized) less p-star, is modeled

[22] It is worth emphasizing that, in contrast to the regime-switching model (as used in Kahn and Rich, 2007), which allows for deterministic values of p-star, a random walk assumption for p-star (as in Edge, Laubach, and Williams, 2007) allows for the possibility that p-star may be changing (slowly) in every period.

[23] We implemented a model specification that allowed for a direct link between p-star and g-star. Doing so notably reduces the model fit to GDP, the unemployment rate, and the interest rate data. And it lowers the precision of the stars' estimates and the other model parameter estimates. The complete set of results corresponding to the model specification allowing for a link between p-star and g-star are available from the author.

[24] As discussed in Kahn and Rich (2007), the focus outside of the farm sector is primarily on avoiding short-term transitory volatility in the farm sector that is heavily driven by weather and other nontechnological factors.

as a function of a one-quarter lag in the productivity gap and the contemporaneous cyclical unemployment gap.

$$P_t - P_t^* = \rho^p(P_{t-1} - P_{t-1}^*) + \lambda_t^p(U_t - U_t^*) + \varepsilon_t^p,\ \ \varepsilon_t^p \sim N(0, e^{h_t^p}) \tag{17}$$

where, $|\rho^p| < 1$

The inclusion of the cyclical unemployment gap helps tease out movements in productivity associated with the business cycle. The growth in labor productivity (and more generally aggregate productivity) has been shown to be procyclical to some degree (e.g., Roberts, 2001; it typically increases sharply at the onset of recoveries and falls during recessions. However, empirical evidence on the strength and the direction of the cyclical relationship is mixed. This mixed evidence stems from the use of different estimation samples and or cyclical indicators (employment-based or output-based). For instance, Galí and van Rens (2021), using split sample estimation, illustrate empirically the significant weakening in the correlation between labor productivity and employment, especially post-1984. They find that the relationship has become countercyclical in the past three decades when using employment as the cyclical indicator.[25] But it is slightly procyclical when using output as the cyclical indicator. This latter finding motivates our alternative specification that replaces cyclical unemployment with the output gap.

$$P_t - P_t^* = \rho^p(P_{t-1} - P_{t-1}^*) + \phi_t^p(ogap_t) + \varepsilon_t^p,\ \ \varepsilon_t^p \sim N(0, e^{h_t^p}) \tag{17b}$$

Galí and van Rens (2021) find weakening in the correlation between labor productivity and the cyclical indicator, which motivates time variation in the coefficients λ^p and ϕ^p.[26]

$$\lambda_t^p = \lambda_{t-1}^p + \varepsilon_t^{\lambda p},\ \ \varepsilon_t^{\lambda p} \sim N(0, \sigma_{\lambda p}^2) \tag{18}$$

$$\phi_t^p = \phi_{t-1}^p + \varepsilon_t^{\phi p},\ \ \varepsilon_t^{\phi p} \sim N(0, \sigma_{\phi p}^2) \tag{18b}$$

The variance of the error term ε_t^p is allowed to change over time.[27] Allowing for the time variation in the cyclical relationship and the error term allows the model to better discriminate the cyclical movements and idiosyncratic movements in productivity from those associated with shifts in p-star.

[25] Galí and van Rens (2021) using a structural macro model attribute the weakening procyclicality of labor productivity in part to the increased flexibility of the US labor market post-1984, which has enabled firms to make adjustments at the extensive margin quickly and easily in response to shocks.

[26] Fernald and Wang (2016) document weakening in the cyclicality of productivity at the industry level, suggesting that the results of Galí and van Rens (2021) are not due to changes in industry composition.

[27] Carriero, Clark, and Marcellino (2019) and Tallman and Zaman (2020) document the superior forecast accuracy of variables including labor productivity from Bayesian VARs with SV compared to VARs without SV, suggesting the usefulness of allowing for SV. We also find empirical support for the inclusion of SV.

The SV process is defined as a driftless random walk in the log-variance.

$$h_t^p = h_{t-1}^p + \varepsilon_t^{hp}, \ \ \varepsilon_t^{hp} \sim N(0, \sigma_{hp}^2) \tag{19}$$

P-star is modeled as a driftless random walk component, and the variance of the shocks to this component is assumed to be constant. Modeling p-star this way allows it to capture both unobserved and observed factors that are thought to be persistent but hard to measure. In particular, one factor is developments in fiscal policy; for example, high levels of government debt in the longer term tend to crowd out private investment, thereby reducing longer-term productivity growth.

$$P_t^* = P_{t-1}^* + \varepsilon_t^{p*}, \ \ \varepsilon_t^{p*} \sim N(0, \sigma_{p*}^2) \tag{20}$$

Economic theory posits that the long-run nominal wage inflation equals the sum of long-run productivity growth and long-run price inflation. As discussed later in the wage inflation block, this theoretical restriction defines the law of motion for w-star and constitutes an additional channel influencing the dynamics of p-star.

2.6. ***Price inflation block***

We use price inflation as measured by the personal consumption expenditures (PCE) price index, the inflation measure that the Federal Reserve targets. Our formulation for the price inflation block closely follows CKP and CCK, combining elements from both of these papers. Specifically, as in CKP, the stationary component, the inflation gap (defined as the deviation of inflation from pi-star),[28] is modeled as a function of the one-quarter lagged inflation gap, unemployment gap, and an error term, whose variance is allowed to vary over time.

The coefficient, ρ^π on the lagged inflation gap, which captures persistence in inflation dynamics (i.e., persistence in the deviation of inflation from pi-star), is allowed to vary over time.[29]

$$\pi_t - \pi_t^* = \rho_t^\pi(\pi_{t-1} - \pi_{t-1}^*) + \lambda_t^\pi(U_t - U_t^*) + \varepsilon_t^\pi, \ \ \varepsilon_t^\pi \sim N(0, e^{h_t^\pi}) \tag{21}$$

$$\rho_t^\pi = \rho_{t-1}^\pi + \varepsilon_t^{\rho\pi}, \ \ \varepsilon_t^{\rho\pi} \sim TN(0 - \rho_{t-1}^\pi, 1 - \rho_{t-1}^\pi; 0, \sigma_{\rho\pi}^2) \tag{22}$$

The innovations to the AR(1) coefficient, ρ^π are truncated so that $0 < \rho_t^\pi < 1$, ensuring that the inflation gap (in equation 21) is stationary at each point in time t.

[28] Modeling inflation in gap form, where the gap is defined as the difference between inflation and a slowly moving trend, was popularized by Cogley and Sbordone (2008) (and Cogley, Primiceri, and Sargent, 2010), and since then has been a widely used approach to modeling inflation in macroeconomic models for policy and forecasting (e.g., Faust and Wright, 2013).

[29] Chan, Koop, and Potter (2013), CKP, and CCK have found strong empirical support for the time-variation in the coefficient of inflation gap. Our results reinforce the empirical importance of allowing for time-variation in this coefficient.

$$\lambda_t^\pi = \lambda_{t-1}^\pi + \varepsilon_t^{\lambda\pi}, \;\; \varepsilon_t^{\lambda\pi} \sim TN(-1 - \lambda_{t-1}^\pi, 0 - \lambda_{t-1}^\pi; 0, \sigma_{\lambda\pi}^2) \tag{23}$$

λ^π is the slope of the price Phillips curve and is constrained in the interval (-1,0).

The parameter λ estimates the price Phillips curve relationship (i.e., the relationship between the inflation gap and the unemployment gap at business cycle frequency). There is ample empirical evidence in support of a time-varying price Phillips curve (e.g., Stella and Stock (2015); CKP; Del Negro, Lenza, Primiceri, and Tambalotti (2020))), hence our choice of allowing for time-variation in the parameter λ^π.[30]

$$h_t^\pi = h_{t-1}^\pi + \varepsilon_t^{h\pi}, \;\; \varepsilon_t^{h\pi} \sim N(0, \sigma_{h\pi}^2) \tag{24}$$

The SV process is defined as a random walk in the log-variance.

Both the theoretical and empirical literature emphasize the usefulness of the signal from fluctuations in labor costs for inflation dynamics. Post-Keynesian theory posits that excess wage inflation over labor productivity gains puts upward pressure on price inflation, i.e., causality runs from labor costs to price inflation. In comparison, the neo-classical theory suggests that causality runs in the opposite direction, from price inflation to nominal wage inflation. The empirical evidence in the US data is inconclusive in that there is no clear evidence on the direction of causality. If anything, the evidence suggests they co-move together (see Knotek II and Zaman, 2014, and references therein).

Given the empirical evidence of co-movement, we explore an alternative specification in which we allow for a connection between two cyclical inflation components, the nominal wage inflation gap and the price inflation gap, by adding the nominal wage inflation gap as an explanatory variable in the equation describing the price inflation gap. The parameter γ^π captures the strength of the relationship between the two cyclical inflation measures. The expression $\frac{\gamma^\pi}{1-\rho^\pi}$ can be interpreted as the pass-through from cyclical wage inflation to cyclical price inflation.[31]

$$\pi_t - \pi_t^* = \rho_t^\pi(\pi_{t-1} - \pi_{t-1}^*) + \lambda_t^\pi(U_t - U_t^*) + \gamma^\pi(W_t - W_t^*) + \varepsilon_t^\pi, \;\; \varepsilon_t^\pi \sim N(0, e^{h_t^\pi}) \tag{21b}$$

Similarly, as shown later, we add the price inflation gap to the equation describing the nominal wage inflation gap.

[30] See Del Negro et al. (2020) for a comprehensive literature review of the instability of the Phillips curve in the US data, and similarly Banbura and Bobeica (2020) for the euro area data. We note that the literature documenting empirical evidence in support of a nonlinear Phillips curve (e.g., Ashley and Verbrugge, 2020) supports time variation in the parameter λ^π.

[31] We explored the possibility of allowing for time variation in γ^π but the estimation ran into difficulties; hence, we resorted to time-invariant γ^π.

Pi-star is modeled as a driftless random walk component, and the variance of the shocks to this component is assumed to be constant (as in CKP). This latter assumption of homoscedastic errors is in contrast to Stock and Watson (2007), Mertens (2016), and several others. Our choice not to incorporate SV into shocks to pi-star is made to keep the estimation manageable and maintain consistency with our modeling assumptions for the stars.[32]

$$\pi_t^* = \pi_{t-1}^* + \varepsilon_t^{\pi*}, \;\; \varepsilon_t^{\pi*} \sim N(0, \sigma_{\pi*}^2) \tag{25}$$

Following CCK (and others, such as Mertens and Nason (2020)), to improve pi-star's econometric estimation, we bring in information from the long-run survey expectations (of PCE inflation). An important empirical finding of CCK is that long-run survey expectations of inflation are a biased measure of the underlying trend inflation, at sometimes. Hence, simply equating pi-star with the long-run survey expectations or assuming that survey expectations are an unbiased measure of pi-star or calibrating econometric estimates of pi-star to surveys (as is commonly done) may not be a reasonable strategy.

Accordingly, we add an equation linking long-run survey expectations of inflation, Z^{π}, to pi-star, where the intercept, C_t^{π}, is time-varying to capture a possibly time-varying differential between the two.[33]

$$Z_t^{\pi} = C_t^{\pi} + \beta^{\pi}\pi_t^* + \varepsilon_t^{z\pi}, \;\; \varepsilon_t^{z\pi} \sim N(0, \sigma_{z\pi}^2) \tag{26}$$

$$C_t^{\pi} = C_{t-1}^{\pi} + \varepsilon_t^{c\pi}, \;\; \varepsilon_t^{c\pi} \sim N(0, \sigma_{c\pi}^2) \tag{27}$$

Our model specification modifies the baseline model of CCK in five important ways. First, we allow a time-varying Phillips curve relationship by adding the cyclical unemployment component (similar to CKP).[34] Second, we explore a model specification that allows for a link between the nominal wage inflation gap and the price inflation gap (capturing the evolving pass-through from labor costs to price inflation). Third, we adopt a more simplistic approach to modeling the link between survey expectations and pi-star. Fourth, the variance of the shocks to the pi-star process does not entertain SV. Fifth, pi-star is restricted to satisfy the long-run restriction informed by theory (see equation 28).

[32] Allowing SV in the inflation gap component and not in the trend component is not without precedent. Besides CKP, Chan (2013) is a recent paper modeling SV only in the measurement equation (i.e., cyclical/transitory component). Berger, Everaert, and Vierke (2016) find support for SV in the inflation gap component but weak evidence for SV in the trend component. Our preliminary results indicate similar findings: that adding SV to the pi-star equation neither helps nor hurts the model fit.

[33] Our formulation is flexible but less so than the one adopted by CCK. In addition to time variation in the intercept, CCK add time-variation in the slope coefficient, and moving average (MA) in the error term. For PCE inflation, CCK did not find the addition of the MA in the error term useful. We think our relatively simplistic equation suffices in its aim of influencing econometric estimation of pi-star through survey expectations.

[34] CCK explored an alternative specification in which they add cyclical unemployment but as an exogenous variable (constructed using the CBO estimate of the natural rate). CCK found that adding cyclical unemployment slightly worsened the fit of their model.

2.7. ***Wage inflation block***

The long-run equilibrium level of nominal wage inflation (w-star) is the nominal wage growth rate consistent with its fundamentals – p-star and pi-star. As noted earlier, in the long run, economic theory posits that the nominal wage inflation equals the sum of the long-run growth rate of labor productivity and the long-run level of price inflation. In other words, in the long run, labor productivity growth is the only fundamental driver of real wages; therefore, price inflation and nominal wage inflation have to adjust relative to each other to maintain the fundamental relationship. In our setup, we impose this relationship to define w-star.

$$W_t^* = \pi_t^* + P_t^* + \varepsilon_t^{w*}, \;\; \varepsilon_t^{w*} \sim N(0, \sigma_{w*}^2) \tag{28}$$

The above equation implies that W^* is approximately equal to sum of $\pi_t^* + P_t^*$.

In order to assess empirical support for the theoretical restriction defined by equation (28), we explore an alternative specification that models W^* as an RW process,

$$W_t^* = W_{t-1}^* + \varepsilon_t^{w*}, \;\; \varepsilon_t^{w*} \sim N(0, \sigma_{w*}^2) \tag{28b}$$

Equation (29) relates the nominal wage inflation gap – defined as the difference between the nominal wage inflation and w-star – to its one-quarter lagged gap, the cyclical unemployment gap, and the price inflation gap.[35] The variance of the error term, ε_t^w, is allowed to vary over time. The latter feature is motivated by findings in Tallman and Zaman (2020) and Peneva and Rudd (2017), who document a better fit of the VAR with SV to nominal wage data compared to the VAR without SV.

$$W_t - W_t^* = \rho_t^w(W_{t-1} - W_{t-1}^*) + \lambda_t^w(U_t - U_t^*) + \kappa_t^w(\pi_t - \pi_t^*) + \varepsilon_t^w, \;\; \varepsilon_t^w \sim N(0, e^{h_t^w}) \tag{29}$$

The SV process is defined as a random walk in the log-variance.

$$h_t^w = h_{t-1}^w + \varepsilon_t^{hw}, \;\; \varepsilon_t^{hw} \sim N(0, \sigma_{hw}^2) \tag{30}$$

The findings in Knotek II and Zaman (2014) motivate the inclusion of a one-quarter lagged nominal wage inflation gap, with time variation in the parameter ρ^w; the latter quantifies the persistence in wage inflation dynamics.

[35] Following the literature on modeling price inflation in the gap, Knotek II and Zaman (2014) apply a similar transformation to modeling nominal wage inflation for the US. In particular, they construct the nominal wage inflation gap as nominal wage inflation less pi-star, where pi-star is survey expectations. They note that the competitive forecasting properties of their model is due to modeling in gaps. Following Knotek II and Zaman (2014), Bobeica, Ciccarelli, and Vansteenkiste (2019) construct a nominal wage inflation gap for the euro area data and find empirical support for the gap specification.

$$\rho_t^w = \rho_{t-1}^w + \varepsilon_t^{\rho w},\;\; \varepsilon_t^{\rho w} \sim TN(0 - \rho_{t-1}^w, 1 - \rho_{t-1}^w; 0, \sigma_{\rho w}^2) \tag{31}$$

The innovations to the AR(1) coefficient, ρ^w, are truncated so that $0 < \rho_t^w < 1$, to ensure that the wage gap (in equation 29) is stationary at each point in time t.

The parameter λ^w in equation (29) measures the strength of the cyclical relationship between the nominal wage gap and labor market slack (aka the wage Phillips curve). Many studies, both theoretical (e.g., Galí, 2011) and empirical (e.g., Knotek II and Zaman, 2014; Peneva and Rudd, 2017; Galí and Gambetti, 2019), have documented strong support for the existence of a wage Phillips curve in the US data. These studies have also demonstrated the instability of the wage Phillips curve, motivating the need for time-variation in the parameter λ^w.[36]

$$\lambda_t^w = \lambda_{t-1}^w + \varepsilon_t^{\lambda w},\;\; \varepsilon_t^{\lambda w} \sim TN(-1 - \lambda_{t-1}^w, 0 - \lambda_{t-1}^w; 0, \sigma_{\lambda w}^2) \tag{32}$$

where λ^w is the slope of the wage Phillips curve and is constrained in the interval (-1,0).

As discussed earlier in the price inflation block, both theory and empirical evidence point to the connection between price inflation and nominal wage inflation. The standard fully structural models describing the New Keynesian Phillips curve posit a tight relationship between price and wage inflation via the channel of current and expected future marginal costs. In these models, price inflation today is a function of expected price inflation and expected future marginal costs, where marginal costs are generally linked to wages. Knotek II and Zaman (2014) provide empirical evidence of the connection between nominal wage and price inflation. In particular, they show no clear evidence of one Granger-causing the other; instead, both wage and price inflation generally tend to move together. This reasoning would suggest the importance of modeling the direct relationship between wage inflation and price inflation. Hence, this motivates the inclusion of the price inflation gap in the measurement equation (29).

Several studies document a significant weakening in the empirical link between price inflation and nominal wage inflation since the 1980s (e.g., Peneva and Rudd, 2017; Knotek II and Zaman, 2014), motivating time variation in the parameter κ^w.[37] The expression $\frac{\kappa_t^w}{1-\rho_t^w}$ could be interpreted as an estimate of the pass-through from price inflation to wage inflation.

$$\kappa_t^w = \kappa_{t-1}^w + \varepsilon_t^{\kappa w},\;\; \varepsilon_t^{\kappa w} \sim N(0, \sigma_{\kappa w}^2) \tag{33}$$

To assess empirical support for the inclusion of the price inflation gap in the equation describing the nominal wage gap, we estimate an alternative model that replaces equation (29)

[36] Literature has posited various explanations for the instability of the wage Phillips curve, including downward nominal wage rigidities, where the degree of rigidity varies with the phase of the business cycle (see Daly and Hobijn, 2014).

[37] Bobeica, Ciccarelli, and Vansteenkiste (2019) find that in the euro area the link between labor compensation and price inflation continues to remain strong post-1980.

with the following:

$$W_t - W_t^* = \rho_t^w (W_{t-1} - W_{t-1}^*) + \lambda_t^w (U_t - U_t^*) + \varepsilon_t^w, \;\; \varepsilon_t^w \sim N(0, e^{h_t^w}) \tag{29b}$$

2.8. ***Interest rate block***

We close the model with the interest rate block characterizing the interest rate dynamics and the law of motion for r-star (the long-run equilibrium real short-term interest rate).

Our first equation of the block brings information from the nominal short-term interest rate via a Taylor-type rule (TR) to aid in identifying r-star. Past research has shown the TR's usefulness in characterizing the monetary policy reaction function over the past four decades. Specifically, this equation characterizes the dynamics of the short-term nominal interest rate gap, where the gap is the difference between the nominal short-term interest rate i, and the long-run level of the nominal neutral rate of interest, i-star. (i-star = pi-star + r-star). When modeling the nominal short-term interest rate, especially in a framework like ours, one must account for the effective lower bound (ELB) period.

The recent literature provides at least two options for handling the ELB. The first is to explicitly but separately model the observed short-term nominal rate, which cannot go below zero, and the "shadow interest rate," which is a hypothetical unobserved and unbounded counterpart. Wu and Xia (2016) popularized the concept of the shadow interest rate, and JM and González-Astudillo and Laforte (2020) are two recent approaches well suited for inclusion in UC models. The second approach is to treat the estimate of the "shadow rate" obtained from Wu and Xia (2016) as the measure of the short-term nominal interest rate in measurement equations such as the TR (e.g., Pescatori and Turunen, 2016).[38]

Given our model's size and complexity, we adopt the latter approach, which is simpler though not perfect. Using a direct measure of the nominal shadow rate allows us to capture both conventional and unconventional monetary policy effects when the (observed) nominal federal funds rate is constrained at the ELB.[39]

Equation (34) relates the nominal interest rate gap (based on the shadow federal funds rate) to its one-period lag interest rate gap, the current quarter inflation gap (i.e., the deviation of inflation from pi-star), and the unemployment rate gap (i.e., the deviation of the unemployment rate from u-star). This equation roughly characterizes the monetary policy reaction function as defined by Taylor (2001).[40] There is a broad consensus that policy adjustments outside of

[38] The estimates from Wu and Xia (2016) are publicly available and regularly updated. Treating the shadow rate as the measure of the short-term nominal rate in place of the federal funds rate is commonly done, and often academic papers report results indicating robustness to the use of Wu and Xia's shadow rate (e.g., Beyer and Wieland, 2019; Lewis and Vazquez-Grande, 2019)

[39] The nominal *shadow* federal funds rate is identical to the nominal federal funds rate when the effective lower bound is not binding.

[40] It is worth emphasizing that we denote this equation as a "Taylor-type rule" and not an exact Taylor-rule because in our equation, pi-star refers to the estimate of trend inflation, which may or may not be equal to central bank's long-run inflation goal.

cyclical turning points are made very gradually. Hence, this motivates the inclusion of the lagged interest rate gap term.

Chan and Eisenstat (2018b,a) and JM document strong empirical support for constant parameters in the Taylor rule equation while allowing for stochastic volatility in the errors. Accordingly, we allow for SV in the interest rate equation. JM, González-Astudillo and Laforte (2020), and Brand and Mazelis (2019) document the usefulness of adding the TR equation to identify r-star. The latter two do not entertain SV, which JM has found to be empirically important. As discussed later, we also found that adding the TR equation improves the precision of the r-star estimates significantly, and the data strongly favor allowing for SV in the error process.

$$i_t - \pi_t^* - r_t^* = \rho^i(i_{t-1} - \pi_{t-1}^* - r_{t-1}^*) + \lambda^i(U_t - U_t^*) + \kappa^i(\pi_t - \pi_t^*) + \varepsilon_t^i, \;\; \varepsilon_t^i \sim N(0, e^{h_t^i}) \tag{34}$$

where ρ^i is truncated so that $0 < \rho^i < 1$.

$$h_t^i = h_{t-1}^i + \varepsilon_t^{hi}, \;\; \varepsilon_t^{hi} \sim N(0, \sigma_{hi}^2) \tag{35}$$

The SV process is defined as a random walk in the log-variance.

Our second equation motivated by LW heeds the economic theory suggesting the role of various real factors in influencing movements in r-star. These factors include long-run output growth (and long-run productivity growth), trend labor force growth (reflecting shifts in demographics and net migration), taxation structure, government expenditure shifts, and shifts in liquidity preferences (e.g., Del Negro et al., 2017; Bullard, 2018. Accordingly, equation (36) expresses r-star as a linear function of g-star and a “catch-all” component D. In our baseline specification, both g-star and D follow random walk processes similar to LW (and many other papers). The RW assumption for D is an appropriate one, given that our focus is the long-run r-star that should, in principle, be influenced over time by permanent shifts in aggregate supply and demand (Laubach and Williams, 2016).[41]

$$r_t^* = \zeta g_t^* + D_t. \tag{36}$$

$$D_t = D_{t-1} + \varepsilon_t^d, \;\; \varepsilon_t^d \sim N(0, \sigma_d^2) \tag{37}$$

In the more recent literature on long-run r-star, modeling r-star as a random walk process has performed better empirically than model specifications relying on the link between g-star

[41] Researchers have also explored an AR process for component D, which would be consistent if the interest is in medium-term r-star (see Lewis and Vazquez-Grande, 2019), as this would allow r-star to be influenced by the transitory shocks to aggregate demand (via the AR process) and permanent shocks to aggregate supply (via the RW process for g-star). In studies focused on the long-run notion of r-star, such as LW, Laubach and Williams (2016), Clark and Kozicki (2005), and Kiley (2020), specification based on the RW assumption has been shown to be empirically favored by the data compared to AR assumption.

and r-star.[42] Accordingly, we explore an additional specification that models r-star simply as the RW process (similar to g-star, p-star, pi-star, u-star). The RW assumption for r-star implies that we are agnostic about the underlying unobserved forces driving r-star but we acknowledge that those forces reflect persistent structural shifts in aggregate demand and supply that ought to have a bearing on r-star.

$$r_t^* = r_{t-1}^* + \varepsilon_t^{r*}, \;\; \varepsilon_t^{r*} \sim N(0, \sigma_{r*}^2) \tag{36b}$$

Lastly, the equations linking the implied estimate of the long-run survey expectations of the real short-term interest rate, Z^r, to r^* are defined as:[43]

$$Z_t^r = C_t^r + \beta^r r_t^* + \varepsilon_t^{zr}, \;\; \varepsilon_t^{zr} \sim N(0, \sigma_{zr}^2) \tag{38}$$

$$C_t^r = C_{t-1}^r + \varepsilon_t^{cr}, \;\; \varepsilon_t^{cr} \sim N(0, \sigma_{cr}^2) \tag{39}$$

All in all, information from six sources and/or elements informs the econometric identification of r-star. These sources include: an IS equation (14)); a TR equation (34), which allows for SV; an equation linking r-star to survey expectations; the shadow rate; and an equation relating r-star to g-star. As we show shortly, all of these sources play a role in improving r-star's precision. To reiterate, in our framework, we use information from both short-term interest rates (via a TR equation) and long-term interest rates (via an IS equation) to inform the estimation of r-star.[44]

2.9. ***Base model and its variants***

The equations (6), (7)...(39) define our baseline model formulation (denoted *Base*). Figure 1 provides a visual representation of our Base model. And section A1.a. of the online appendix lists all of the equations for the Base model for easy reference. To assess the usefulness of survey information in the econometric estimation of our multivariate UC model, we also estimate a variant of the baseline model that excludes the equations linking long-run survey expectations to stars (i.e., excluding equations 9, 10, 15, 16, 26, 27, 38, and 39). We denote the latter specification as *Base-NoSurv*. The model specifications Base and Base-NoSurv constitute our two main model specifications. To assess the empirical support for numerous additional features (informed by theory and past empirical research) embedded in our modeling framework, we

[42] See Kiley, 2020; González-Astudillo and Laforte, 2020; Orphanides and Williams, 2003; JM.

[43] The r-star survey estimates are not direct estimates; instead, they are inferred from the Blue Chip survey's long-run estimates of the GDP deflator and short-term interest rates using the long-run Fisher equation. The survey expectations for r-star goes back to 1983.Q1. Please refer to the online appendix A9 for details on the procedure to back-cast estimates all the way back to 1959.

[44] Del Negro et al. (2017), JM, Bauer and Rudebusch (2020), and González-Astudillo and Laforte (2020) are recent studies that have highlighted the usefulness of exploiting information from both short-term and long-term interest rates in the identification of r-star.

formulate several additional model specifications, each of which is a restricted variant of the Base. For instance, to assess the empirical support of the theoretical restriction defined by equation 28 (which defines w-star as the sum of pi-star and p-star), we estimate a variant of the baseline model that replaces equation 28 with a random walk assumption for w-star as defined by equation 28b. We denote this specification as *Base-W*RW*.

Similarly, to assess the empirical support for the theoretical restriction defined by equation 36 (the link between g-star and r-star), we estimate a model specification that replaces equation 36 with a random walk assumption for r-star as defined by the equation 36b. Table 1 reports the description of model specifications that are formulated to assess various features of importance. We ran a few additional specifications to explore the role of priors in influencing r-star, and for brevity purposes they are relegated to the online appendix. To keep the length of the paper manageable, we report selected results from the auxiliary model specifications in the main part of the paper with additional results included in the online appendix.

3. Data and Bayesian Estimation

3.1. ***Data***

We estimate the empirical model using the following quarterly data: (1) the unemployment rate; (2) real GDP growth; (3) nonfarm labor productivity growth; (4) the inflation rate in personal consumption expenditures (PCE) price index; (5) average hourly earnings (AHE) of production and nonsupervisory workers (total private industries);[45] (6) the federal funds rate; (7) nominal yield on the 10-year Treasury bond; (8) shadow federal funds rate from Wu and Xia (2016); (9) Blue Chip[46] (real-time) long-run projections of three-month Treasury bill, real output growth, the unemployment rate, and GDP deflator inflation; (10) long-run inflation expectations of PCE inflation (PTR series). We also collect the real-time long-run CBO projections of real output growth, the level of real potential output, and the natural rate of unemployment. For forecast evaluation exercises, the real-time data vintages of real GDP growth, PCE inflation, the unemployment rate, AHE, and nonfarm labor productivity spanning 1998Q1 through 2019Q4 are downloaded from the ALFRED database maintained by the St. Louis Fed and the real-time database maintained by the Federal Reserve Bank of Philadelphia. For the data series labeled (1) through (7), which comprises our core data set, we collect two vintages of revised data: 2020Q2 and 2020Q4 vintages, respectively. We use data starting in 1959Q4 through 2019Q4 from the 2020Q2 vintage (which includes the third estimate of 2019Q4) as a featured sample for this paper. To show the implications of the COVID-19 data for our model estimates, we estimate our model(s) using the 2020Q4 vintage, which has data spanning 1959Q4 through 2020Q3. The vintages corresponding to the revised data are downloaded from

[45]Average hourly earnings (AHE) of production and nonsupervisory workers in total private industries goes back to 1964Q1. From 1959Q4 through 1963Q4, we use the AHE of production and nonsupervisory workers in goods-producing industries. We splice them together.

[46]Blue Chip Economic Indicators data published by Wolters Kluwer Legal and Regulatory Solutions U.S.

Haver Analytics.

3.2. ***Bayesian estimation***

We use Bayesian estimation methods to fit our Base model and its variants. The use of inequality restrictions on latent parameters in our model(s) setup leads to a nonlinear state-space model, which renders estimation using standard Kalman filter methods infeasible. Accordingly, we implement our Markov chain Monte Carlo (MCMC) posterior sampler based on computational methods developed in Chan, Koop, and Potter (2013) and CKP, who use the band and sparse matrix algorithms detailed in Chan and Jeliazkov (2009). Specifically, the MCMC posterior sampler is a significantly scaled-up version of the sampler employed by CKP. The CKP posterior sample developed for a relatively smaller-scale nonlinear state-space model is carefully extended to accommodate the additional structure and numerous features of our model(s). Since the computational methods used in this paper are based on CKP, we relegate the specific details of the sampler to the online appendix A1.

In a methodological sense, this paper's novelty is in assembling the existing sampling algorithms based on the fast band and sparse matrix routines to solve a large nonlinear and a high-dimensional UC model. We found that the use of inequality restrictions, such as bounds on the u-star and other parameters, is crucial to estimate the model, especially in the Base-NoSurv model. Intuitively, features such as truncated distributions that we implement for some of the time-varying parameters, e.g., the Phillips curve (price and wage), persistence, and bounds on u-star facilitate estimation by guiding the estimation procedure to the credible regions of the parameter space.

For each model, we simulate 1 million posterior draws from the MCMC posterior sampler. We then discard the first 500,000 draws, and of the remaining, we keep every 100th draw. Accordingly, all of the reported results for the Base model and its variants are based on 5000 retained draws.

As emphasized by Chan (2017), the MCMC algorithm is considered efficient if the draws it produces have low autocorrelation – they are autocorrelated by construction – and the time it takes to sample a given number of posterior draws is reasonable. In the appendix, we report the efficiency diagnostics of our MCMC algorithm. Those diagnostics, which include inefficiency factors and convergence metrics, indicate good convergence properties (and low autocorrelation) of our sampler for both Base and Base-NoSurv models.[47]

Bayesian model comparison is based on the marginal likelihood metric. In computing marginal likelihood for various models, we use the approach proposed by CCK, which decomposes the marginal density of the data (e.g., inflation) into the product of predictive likelihoods;

[47] Regarding computational time, given the high dimensionality of our model(s) and the number of posterior simulations we require, the speed is quite fast (in our assessment). When applied to the Base model, the MCMC algorithm, which is implemented in Matlab, takes about 350 seconds to generate 10,000 posterior draws using a laptop computer with an Intel(R) Xeon(R) E-2176M CPU @ 2.70 GHz processor. To generate 1 million posterior draws, it takes less than 10 hours.

see appendix A1.d for details.[48] An important advantage of the CCK approach is that it allows us to separately compute marginal data density for each variable of interest: inflation, nominal wages, interest rate, real GDP, the unemployment rate, and labor productivity. The variable-specific marginal densities prove useful because it allows for deeper insights into the source of the deficiencies, which helps differentiate models at a more granular level.

We note that our prior settings are similar to those used in CKP, CCK, and González-Astudillo and Laforte (2020). As discussed in CCK, UC models with several unobserved variables, such as the one developed in this paper, require informative priors. That said, our prior settings for most variables are only slightly informative. The use of inequality restrictions on some parameters such as the Phillips curve, persistence, and bounds on u-star could be viewed as additional sources of information that eliminate the need for tight priors, something also noted by CKP. For the parameters on which there is strong agreement in the empirical literature on their values, such as the Taylor-rule equation parameters, we use relatively tight priors, such that prior distributions are centered on prior means with small variance. In model comparison exercises, the priors are kept the same for the common parameters across models. We also perform some prior sensitivity analysis reported in appendix A2.

4. Full Sample Estimation Results

This section sequentially discusses results for each of the six blocks, particularly the full-sample estimates of stars. Here, we briefly highlight some of the noteworthy findings. Comparing estimates of the stars between model specification Base and Base-NoSurv, a clear pattern emerges. The precision of the estimates, as measured by the width of the 90% credible intervals, indicates that the Base model, which explicitly accounts for the links between stars and survey data, yields more precise estimates of the stars than Base-NoSurv. Although the broader contours seen in the estimates of stars from our two main model specifications, Base and Base-NoSurv, are comparable, at times, the differences can be notable, especially in the case of r-star. In the case of g-star (and, in turn, the output gap), our two model specifications yield very similar trajectories as implied by the posterior mean estimates. This result suggests that the data are very informative about g-star, and not so much about r-star, confirming the finding of Kiley (2020).

Furthermore, the Bayesian model comparison indicates marginally higher support in the data for Base over Base-NoSurv, as shown in Table 3. The breakdown of the marginal data density by variables suggests that the Base model's improved fit over Base-NoSurv is due to its improved fit to inflation and nominal interest rate data. But that improved fit is offset mainly by worsening fit to the unemployment and wage data. The table also reports the marginal data density for two additional Base model variants for which the Bayesian model comparison indicates a comparable fit to the data. Overall, we find that bringing in additional information

[48] Alternatively, one could use the cross-entropy (Kullback-Leibler divergence) approach proposed by Chan and Eisenstat (2018b,a). We leave this for future extensions.

from surveys – by directly modeling the connection to the stars – leads to more reasonable values, provides more precise estimates, and marginally improves the model fit.

4.1. ***Estimation results for u-star***

Figure 2 plots the evolution of u-star (and its uncertainty) covering the period 1960 through 2019. Panel (a) plots the posterior estimates from the Base model and panel (b) from the Base-NoSurv model. Also plotted are the corresponding 90% credible intervals. Both models imply a smooth evolution of u-star. In the past six decades, the (posterior mean of) u-star has fluctuated between 4.4% and 5.7%, peaking in the early 1980s and early 2010s, and troughs in the late 1990s and at the end of our sample period. The contours of u-star from both models are generally similar (as can also be seen conveniently in panel d); however, the level of u-star can differ notably in some periods. From 1960 through the late 1970s, u-star has gradually increased (from 5.4% to 5.7% in Base and 5.0% to 5.6% in Base-NoSurv). But since the mid 1980s through the late 1990s, u-star has steadily drifted lower (to 4.5%). This downward trend in the later period is also documented in the u-star literature based on job-flows data, which attributes the decline in u-star to declining trends in job-separation and job-finding rates (e.g., Crump et al., 2019; Tasci, 2012).

From early 2000 through early 2010, u-star has trended higher, with a sharp pickup during the Great Recession period. Since 2010, u-star has steadily drifted lower. By the end of 2019, Base has u-star at 4.4% (with a 90% interval covering 3.7% to 5.2%) and Base-NoSurv at 4.5% (with a 90% interval covering 3.7% to 5.4%). As shown in Figure 3, at the end of 2019, the unemployment gap implied by both models is negative, i.e., the unemployment rate is below the estimated u-star.

The use of survey information in the Base model mainly contributes to the difference in the levels of u-star across the two models. To facilitate comparison, panel (a) also plots u-star from the survey. As is evident from the plot, u-star from the survey displays more pronounced shifts in u-star than the model-based estimates. However, due to a strong estimated relationship between the survey u-star and Base u-star (i.e., posterior mean of $\beta^u = 0.988$), the Base estimate of u-star reflects the contours in survey u-star. As can be seen in panel (e), which plots the precision (measured as the width of the 90% intervals), taking survey information onboard improves the precision of u-star notably (comparing Base and Base-NoSurv).

Surprisingly, based on the Bayesian model comparison, the Base model has an inferior fit to the unemployment data compared to Base-NoSurv. As shown later, this result contrasts with the results for pi-star, r-star, and g-star, for which survey information helps improve the model fit or at least does not worsen the fit.[49]

[49] Wright (2013) and Tallman and Zaman (2020) use long-run survey information in an attempt to improve VAR model forecasts. In both cases, survey information did not improve the accuracy of unemployment rate forecasts, although they found survey forecasts useful in improving accuracy for a range of other macroeconomic variables.

Sensitivity of u-star to modeling assumptions including information set

Panel (c) plots additional estimates of u-star (posterior mean) from the variants of the Base model to highlight the sensitivity of u-star to modeling assumptions and the informational aspect of joint modeling. The plot denoted Base-NoBoundU* represents the Base model variant that eliminates the bound on the random walk process describing u-star. Doing so has a trivial effect on the estimates of u-star, the precision of u-star, and model fit. Comparing between panels (a) and (c), the posterior mean estimate of u-star is quite similar across Base and Base-NoBoundU*. Similarly, there is little change in the u-star estimate's precision across the two models, with Base only marginally better in the latter part of the sample (as shown in panel e). Not surprisingly, the Bayesian model comparison suggests equal support for both Base and Base-NoBoundU*.

We highlight two noteworthy comments in regard to the implementation of bounds on u-star. First, the trivial difference in the estimates between Base and Base-NoBoundU* arises because the bounds defined on u-star are wide. Put differently, the values of the bound we have set are not binding on the Base model. Second, we find that using bounds on u-star is extremely important in the Base-NoSurv, as it helps keep the estimation tractable. In other words, the advantages of using bounds on the random walk processes that were stressed in CKP were in full display in the estimation of Base-NoSurv. Hence, we prefer to keep bounds on u-star in our main models.

At the time of writing this paper, the world was hit with a COVID-19 shock, an extreme and unprecedented global health shock along various dimensions, leading several analysts to call it a "once-in-a-lifetime upheaval." As we will show in Section 6, the implementation of bounds on u-star is part of the story in preventing our models from blowing up in response to COVID-19 data.

The other model variants plotted in panel (c) are all nested specifications of the Base model: the Bivariate model of GDP and the unemployment rate (a Base model that excludes survey information and everything else except the equations describing the dynamics of GDP and the unemployment rate); Bivariate+Surv, which is bivariate but adds survey data for GDP and unemployment; and CKP Adjusted, which is a bivariate model of inflation and the unemployment rate as in CKP but with no bounds on pi-star. For visual reasons (to limit the number of plots), u-star from Base in panel (c) is not shown. However, for the sake of discussion, we could treat the plot representing Base-NoBoundU* as the estimate for the Base model, since they are identical to each other (as discussed in the preceding paragraph).

These plots show that different model specifications could provide very different signals about the level of long-run unemployment, indicating the sensitivity of u-star to modeling assumptions. As evident from the figure, the small-scale model specifications indicate u-star range-bound between 5.3% and 6.5% over the sample. In contrast, the Base specification has u-star fluctuating over a broader range. A model specification that infers the estimate of u-star from inflation and unemployment data only, i.e., the price Phillips curve (CKP Adj. model), has a higher trajectory of u-star compared to Base. The story is similar in the case of the model specification that infers the estimate of u-star from GDP and the unemployment data only, i.e., the Okun's law relationship (Bivariate model), though the trajectory of u-star is lower than implied by the CKP-Adj model. Once the Bivariate model is augmented with survey data for GDP and unemployment, the trajectory is revised higher to resemble CKP-Adj, but with contours similar to Base (because of the survey data).

Panel (d) compares the u-star estimates from the main model specifications with the CBO estimate of the long-run unemployment rate. Interestingly, except for the 2000-2007 period, the CBO u-star's contour is similar to our model-based estimates, though the level of CBO u-star estimate is significantly higher from 1960 through the early 2000s. From 2000 to 2007, both Base and Base-NoSurv indicate a steadily rising u-star, whereas the CBO has u-star trending lower. At the onset of the Great Recession and through the early phase of the economic recovery, all three have u-star continuing to move higher. Whereas Base and Base-NoSurv peak in late 2010 at 5.5% and 5.2%, respectively, the CBO has u-star peaking in late 2011 at 5.8%. Since then, u-star has steadily moved lower, with the pace of decline quite similar across CBO and Base. CBO has u-star at the end of 2019 at 4.4%, identical to Base and just a tenth shy of Base-NoSurv.

Precision of u-star

Panel (e) plots the precision of u-star estimates for Base, Base-NoSurv, Base-NoBoundU*, and Bivariate+Surv. We make several observations. First, comparing between Base and Bivariate+Surv, both of these model specifications use information from the surveys and the Okun's relationship. However, Base relies on greater information and additional structure (e.g., wage Phillips curve, price Phillips curve, cyclical productivity, monetary policy stance via the Taylor-type policy rule) to infer the u-star compared to Bivariate+Surv, hence the more improved precision of the resulting u-star. The latter reasoning contributed to the very different estimate of u-star from the Base than Bivariate+Surv discussed earlier (and shown in panel c). And the model comparison indicates a significantly higher fit of the Base model to the unemployment data compared to the Bivariate+Surv.

Second, comparing with Base and Base-NoSurv, additional information from survey forecasts improves the precision of u-star, but this improved precision does not translate into improved model fit, which instead worsens somewhat (as shown in Table 4).

Panel (f) shows the precision of u-star for Base-NoSurv, Bivariate (GDP and unemployment), and CKP-Adj (which is a bivariate model of price inflation and the unemployment rate). As indicated earlier, u-star from Base-NoSurv is inferred from a broader information set and structure than the other two small-scale models. Accordingly, the Base-NoSurv estimate of u-star is, for the most part, more precise and the model comparison indicates a substantially higher fit to the unemployment data than the other two. The plots also show that u-star inferred from the Okun's law relationship (i.e., Bivariate model) is less precise than u-star inferred from the price Phillips curve (i.e., CKP-Adj model). In contrast, the Bayesian model comparison lends support to the Bivariate model over the CKP-Adj model.

The results also provide evidence that adding survey data to the Bivariate model (Bivariate+Surv) further improves both the precision of u-star (comparing Bivariate and Bivariate-Surv in panels e and f) and the fit to the unemployment data (Bivariate: -56.5 vs. Bivariate+Surv: -46.5, as shown in Table 4). This latter finding of improved fit from adding survey data is interesting because in the case of Base, adding survey data worsens the model fit (Base: -24.6 vs. Base-NoSurv: -21.7). Importantly, it suggests that survey forecasts of u-star are likely useful in the case of parsimonious models but of limited use for models that are already utilizing various sources of information to infer u-star.

Cyclical unemployment

Figure 3 presents the posterior mean estimate of the unemployment gap (i.e., the cyclical component of the unemployment rate) and the corresponding 90% credible intervals. The top panel plots the estimates from the Base model, and the bottom panel from the Base-NoSurv model. A visual inspection indicates that the movements in cyclical unemployment correspond quite well with the NBER's business cycle dating. For instance, cyclical unemployment falls in economic expansions and rises during recessions. Both models show a significant spike in the cyclical unemployment rate in the 1982-83 and 2007-09 recessions, and a sharper recovery following the 1982-83 recession but a more gradual recovery following the Great Recession. The figure also highlights that both models produce similar estimates of cyclical unemployment. Comparing estimates of (smoothly evolving) u-star in the previous figure to estimates of the cyclical unemployment rate indicates that fluctuations in the observed unemployment rate are attributed mainly to cyclical unemployment.

4.2. ***Estimation results for g-star and the output gap***

Panels (a) and (b) in Figure 4 plot the g-star estimates from several sources, and panel (c) plots the corresponding precision. As is evident, g-star estimates from all sources shown indicate a steady decline throughout the sample, except a temporary rise in the late 1990s, which the literature has attributed to the technology boom. According to the posterior mean estimates of g-star from our Base and Base-NoSurv models, the growth rate of potential output has continuously drifted lower from an annualized rate of close to 4.5% in early 1960 to 1.4% by the end of 2019. The estimate of g-star fell to 1.2% in 2012 and remained there through 2015 and then began very slowly to move up. The story is generally similar based on the inference from simpler (nested) specifications of the Base model: (1) the univariate model (which is Grant and Chan, 2017b model); (2) the bivariate model of real GDP and the unemployment rate; and (3) the bivariate model augmented to include survey data for g-star and u-star (denoted Bivariate+Surv).

This continuous reduction in the growth rate of potential output has been extensively documented elsewhere (e.g., Berger, Everaert, and Vierke, 2016; Grant and Chan, 2017b; Coibion, Gorodnichenko, and Ulate, 2018) and in particular the decline since 2009 has been of great concern among policymakers. Several other researchers, including Summers (2014); Eggertsson, Mehrotra, Singh, and Summers (2016); Pescatori and Turunen (2016); Laubach and Williams (2016); Antolin-Diaz, Drechsel, and Petrella (2017); Holston, Laubach, and Williams (2017), have also documented the secular decline in g-star over the past two decades.

Not surprisingly, the precision of the g-star estimates (and of the output gap) displays patterns that align well with intuition. For instance, model specifications that incorporate survey expectations (i.e., Base and Bivariate + Surv) yield more precise estimates than specifications that ignore survey data. As discussed earlier, previous researchers have shown the unemployment data to be the most critical indicator for the estimation of g-star and the output gap, and the plots provide evidence to that effect. For example, the model specification Bivariate, which builds on the univariate GDP model by adding the unemployment rate, yields a substantial improvement in the precision of the g-star estimate. Interestingly, the g-star estimate's precision from Bivariate is about the same as the Base-NoSurv, suggesting that unemployment is the most crucial variable influencing estimation of g-star. However, as we show shortly, model comparison results indicate support for Base-NoSurv over Bivariate in the output gap case.

Output gap estimates

Next, we examine the estimates of the output gap. In Figure 4, panels (d) and (e) plot the output gap estimates from the same sources as in the case of g-star, and panel (f) plots the corresponding precision. Overall, the estimates of the output gap from Base and Base-NoSurv are quite similar. They accord well with the NBER recession dates. Furthermore, as noted by Morley and Piger (2012) and Johannsen and Mertens (2021), our model-based estimates provide evidence of asymmetry in that recessions are shorter in duration but deeper than expansions in the US. It is instructive to highlight that estimates imply a more negative output gap (of -10.5%; posterior mean) during the 1981-82 recession compared to the Great Recession period (-7%) when output fell more dramatically. At a first pass, this may seem odd. But a closer inspection reveals that in comparison to the 1981-82 recession, during the Great Recession, g-star fell significantly (as can be seen in panel a), resulting in a smaller negative output gap; in contrast, during the 1981-82 recession, g-star is estimated to have remained stable.

Bivariate models are consistent with a similar story, but there are notable differences in the estimates implied from a univariate model. For instance, the latter model suggests a less dramatic fall in the output gap in 1973-74, 1981-82, and 2007-09 recessions. In the mild 2001 recession, the univariate model estimates a positive output gap, although less positive than before the recession. Also, as of 2019, according to this model, the output gap remains negative, which is in sharp contrast to other models and the consensus view (e.g., the CBO output gap).

The precision estimates and the Bayesian model comparison indicate the inferior quality of the univariate model's output gap estimate. Table 5 reports the assessment of model fit to the GDP data for the various model specifications discussed in this section. As in the case of g-star, the output gap estimate from the univariate model is subject to a great deal of uncertainty. The bivariate model, which brings additional information from the unemployment rate, helps improve the precision significantly. This improved precision is also reflected in the bivariate model's substantially improved fit to GDP data compared to the univariate model (Bivariate: -280.4 vs. Univariate: -296.5). In our subjective assessment, the estimate of the bivariate model's output gap is more reasonable because its trajectory aligns well with Base and Base-NoSurv, and as shown shortly, with the consensus view.

Further improvements in precision are realized by bringing in additional information from the surveys (comparing Bivariate vs. Bivariate+Surv); however, the model fit, which reflects uncertainty about other model parameters in addition to g-star and the output gap, is little changed – in fact, it slightly deteriorates. The usefulness of survey data in improving the quality of the g-star and output gap estimates is also apparent comparing precision between Base and Base-NoSurv; however, model fit is nearly similar. Contributing to comparable model fit between the two models is that from 1990 onward, the precision of the output gap across the two models is identical. Unlike, in the case of g-star, the Base-NoSurv yields significantly more precise estimates of the output gap than Bivariate, suggesting the usefulness of the additional information and structure embedded in Base-NoSurv, hence, on net the slightly better fit of the Base-NoSurv model to GDP data compared to Bivariate (Base-NoSurv: -279.1 vs. Bivariate: -280.4)

Posterior parameter estimates for the output block

Next, we discuss the Base model's parameter estimates of the output block that drive the dynamics of g-star and the output gap. The posterior mean estimates of parameters ρ_1^g and ρ_2^g indicate a high degree of persistence ($\rho_1^g + \rho_2^g = 0.74$) and suggest a hump-shaped response of the output gap to shocks (as $\rho_1^g > 1$). These parameters are precisely estimated as evidenced by tight posterior credible intervals. The posterior estimate of parameter λ^g (the coefficient on the unemployment gap in the output gap equation) is negative and highly significant statistically. The estimated posterior mean of λ^g is -0.46 (with 90% interval -0.58 to -0.34). Similarly, the parameter ϕ^u (the coefficient on the output gap in the unemployment equation), discussed earlier, is also negative and highly significant statistically. Together, these estimates indicate a strong Okun's law relationship in the data. The implied posterior mean estimate of the Okun's law coefficient, $\frac{(1-\rho_1^u-\rho_2^u)}{\phi_u}$ is -2.1, with 90% credible intervals spanning -2.3 to -1.8. This estimated coefficient is strikingly identical to the conventional estimate often discussed in macroeconomic textbooks. Therefore, not surprisingly, both the estimated output gap and the unemployment gap (shown earlier) reveal similar cyclical dynamics. For instance, according to both cyclical measures, the 1981-82 recession is estimated to have been deeper than the Great Recession.

The parameter a^r, which relates the output gap to the real rate gap (characterizing the IS relation), is negative and much smaller than the prior mean. The estimated posterior mean of a^r is -0.07 (with 90% interval -0.14 to -0.00).

The posterior mean estimate of $E(\sigma^2_{gdp*})$, the variance parameter of the innovations to the process governing the evolution of g-star (and gdp-star), is nearly identical across the Base and Base-NoSurv models: 0.02^2. For comparison, the prior mean $E(\sigma^2_{gdp*})$ is 0.01^2. Similarly, for the output gap, the posterior mean estimate of $E(\sigma^2_{ogap})$, the variance parameter of the innovations to the IS equation, is also identical across the two models: 0.72^2 (compared with a prior of 1^2).[50] The estimation results suggest that the data are quite informative in influencing the dynamics of both the output gap and g-star, confirming Kiley (2020).

Model-based estimates of the output gap vs. the CBO and others

Figure 5 presents estimates from outside sources, the CBO and the LW model, to gauge how our model-based output gap estimates compare to measures from other sources. Also plotted are estimates from Base and Base-NoSurv in panel (a) and from the Bivariate and Bivariate+Surv models in panel (b) to facilitate comparison with the outside estimates. A few observations immediately stand out. First, the output gap estimate implied from the LW model is notably different over most of the sample period. In particular, during the Great Recession period, the output gap from LW turned slightly negative, while other estimates implied larger negative gaps. The slight negative gap in the LW model is the result of the LW model estimating a dramatic fall in potential output, in line with the collapse in actual output.

Second, both the Base and Base-NoSurv models produce estimates of the output gap generally similar to the CBO estimate. This close similarity is notable because the CBO approach is based upon an entirely different methodology, a production function approach (see Shackleton (2018)).[51]. Note that the Bivariate models of real GDP and the unemployment rate (with and without survey), which are nested specifications of our bigger Base model, produce estimates of the output gap broadly similar, though not identical, to our two main model specifications. Interestingly, in periods when the output gap estimates from our Base (and Base-NoSurv) differ from the CBO estimates, the Bivariate models' estimates are identical to the CBO's. And in periods when the Bivariate models differ from CBO, the Base (and Base-NoSurv) estimates are identical to CBO's.

The estimation results suggest an important takeaway: that joint modeling of real GDP and the unemployment rate is the key to obtaining credible output gap estimates. Morley and Wong (2020), who estimate the output gap using a large BVAR, also found that the unemployment rate is the most crucial indicator for the output gap. (In Section 6, we compare our model

[50]Note, our prior setting implying a high ratio of $\frac{\sigma^2_{ogap}}{\sigma^2_{gdp*}}$ is consistent with the high noise-to-signal ratio suggested in Kamber, Morley, and Wong (2018). Specifically, they recommend fixing the noise-to-signal ratio to a high value to obtain large and more persistent cycles when estimating output using Morley, Nelson, and Zivot (2003) UC model with maximum likelihood estimation.

[51]We note that in recent years, more and more papers using UC models, which jointly model real GDP growth and the unemployment rate, yield estimates of the output gap similar to the CBO output gap (e.g., Johannsen and Mertens, 2021; Berger, Everaert, and Vierke, 2016; Kiley, 2020; González-Astudillo and Laforte, 2020)

estimates with additional estimates, including Morley and Wong's). Recently, Banbura and Bobeica (2020) use a range of small-scale UC models to estimate the output gap and similarly find that the unemployment rate is the most valuable indicator.

Our paper's result indicating a close resemblance of our models' output gap estimates to the CBO's output gap provides evidence supporting the common practice of using output gap estimates from the CBO as an exogenous variable in empirical macroeconomic models (e.g., JM; Stock and Watson, 2020). We view this result as a useful contribution to the applied macroeconomics literature.

On the one hand, the fact that model-based estimates of the output gap bear a strong resemblance to institutional forecasts, i.e., CBO, is encouraging and lends credibility to our model(s). However, on the other hand, in light of the evidence reported in Coibion, Gorodnichenko, and Ulate (2018), the strong resemblance to outside estimates is an unfortunate outcome. We say this for the following reason. Coibion, Gorodnichenko, and Ulate examine estimates of potential output taken from a variety of model-based and external sources, including the CBO and survey forecasts, and based on a range of shock measures, they find that (in real time) the estimates of potential output are unable to distinguish between transitory and permanent shocks effectively. Put differently, they find that their estimates of potential output respond "gradually and similarly" to both supply shocks and demand shocks that drive cyclical fluctuations in real GDP. This is unfortunate, since, by definition, potential output (and g-star) should only adjust in response to permanent shocks.

Coibion, Gorodnichenko, and Ulate (2018) in their conclusion postulate whether a framework that jointly estimates the dynamics of potential output with other relevant stars (as theory implies) would better distinguish between permanent and transitory components and hence lead to more credible estimates of potential output. Unfortunately, our model estimates, based on both full sample and real time (shown in Section 5), suggest otherwise.

4.3. ***Estimation results for p-star***

Figure 6 presents posterior estimates of p-star and other parameters of the productivity block. Panels (a), (b), and (c) present p-star estimates from the Base, Base-NoSurv, and Base-W*RW models, respectively. Both the posterior mean and the 90% credible intervals are shown. Also plotted is the actual labor productivity series. A visual inspection of the actual series indicates the unusually high volatility of the quarterly productivity data. In addition, this series is subject to a high-degree of revisions in subsequent data vintages, suggesting the extreme difficulties of its measurement in real time (see Jacobs and van Norden, 2016). Perhaps, a quote from the former chair of the Federal Reserve, Alan Greenspan (courtesy of Jacobs and van Norden, 2016), would be instructive to reflect a general sentiment about the productivity data:

"The productivity numbers are very rough estimates because we are measuring a whole set of production outputs from one set of data and a whole set of labor inputs from a different set.

That they come out even remotely measuring actual labor productivity is open to question..." (Transcript: Meeting of the Federal Open Market Committee, March 25, 1998, p. 96)

Not surprisingly, researchers have emphasized that these difficulties of extreme volatility, extensive revisions, and real-time measurement issues with productivity data complicate its trend-cycle decomposition (e.g., Edge, Laubach, and Williams, 2007; Kahn and Rich, 2007). Our model-based estimates reflect these challenges. For instance, the estimate of the parameter ρ^p, reported in Table 2, indicates close to zero persistence in the labor productivity data, defined as the difference between the growth rate in labor productivity and p-star. Similarly, our estimation indicates that the labor productivity data have very little influence on the estimate of p-star. Put differently, the data are so volatile to allow for a meaningful identification of trend in the productivity data. The posterior mean of $E(\sigma^2_{p*})$, the variance of the shock process for p-star, is essentially the same as the prior mean.[52] As a result, the degree of time variation in p-star is primarily influenced by the prior setting. So conditional on our prior belief, which allows p-star to evolve slowly from one quarter to the next, we find considerable evidence of gradual time variation in p-star over the post-war sample. The evidence of time variation is economically significant and is consistent with the findings of Roberts (2001), Benati (2007), Edge, Laubach, and Williams (2007), and Fernald (2007).

Comparing across the top three panels in the figure, the paths of p-star (posterior mean estimates) from Base and Base-NoSur are lower than Base-W*RW. The latter model removes the restriction that the long-run w-star grows at a rate equal to the sum of pi-star and p-star. So removing this restriction eliminates the direct influence on p-star from wages and prices and helps raise the level of p-star. As evident from the Bayesian model comparison reported in Table 6, the elimination of this restriction improves the fit of the model to the productivity data but reduces the overall model fit to other data, particularly interest rates – via the changes in pi-star and u-star in the Taylor-type rule equation. That said, all three models generally indicate similar broad patterns in p-star.

After averaging between 2% and 3% in the 1960s, the models indicate that p-star experienced a sharp deceleration in the 1970s through mid-1980s, mirroring the dramatic fall in productivity growth. Both Base and Base-NoSurv estimates show p-star trending lower from 2.4% (2.3%) in early 1970 to 0.5% by mid-1980, whereas Base-W*RW has it falling close to 1.5%, with wide 90% credible intervals that range from 0.4% to 2.3%. From there on through to the late 1990s, p-star increased sharply, at a pace roughly equivalent to its deceleration in prior periods, to reach a level of 2.0 to 2.4% by 1999. The literature attributes part of this acceleration in the latter half of the 1990s to the information technology boom. Roberts (2001), Edge, Laubach, and Williams (2007), and Benati (2007) document estimates of trend productivity generally similar to the p-star implied from the Base-W*RW model.[53]

[52]We tried different values for the prior mean on this parameter, and found that the posterior moves with the prior.

[53]Edge, Laubach, and Williams (2007), who collect real-time estimates of long-run productivity from vari-

In the 2000s, the models have p-star gradually declining to a level close to 1.0%-1.2% by 2010. It remained close to that level through most of the past decade, but since 2018, it has steadily increased. At the end of our sample, all three models estimate the posterior mean of p-star at, or close to, 1.5%. As we show in appendix A12, these estimates of p-star are consistent with the narrative implied by the two-regime Markov-switching model of Kahn and Rich (2007), an influential contribution to the trend productivity literature.

The uncertainty around the posterior mean estimates of p-star is large. Panel (d) quantifies this uncertainty by reporting the width of the 90% credible intervals corresponding to all three models. The plots provide evidence that the theoretical restriction (defined by eq. 28) contributes to a substantially improved precision of p-star (just as it does for pi-star and w-star), as is evident in the Base and Base-NoSurv plots lying below the Base-W*RW. Interestingly, the estimate of p-star from the Base is more precise than Base-NoSurv in 1960 through the mid-1980s, even though we do not utilize survey-based long-run expectations of productivity in the estimation of our model. Based on the Bayesian model comparison, the Base model's fit to the productivity data is marginally better than Base-NoSurv. The improved precision and better fit of the Base model to the productivity data suggest there are important spillover effects in the estimation from survey forecasts of other stars.

Cyclical dynamics of labor productivity

Panel (e) plots the estimate of the parameter λ^p, which relates cyclical unemployment to the productivity gap, from all three models. The plots indicate a high level of uncertainty around the estimates of λ^p. The 90% credible intervals are wide, such that they include both positive and negative values, complicating reliable inference. Going just by the posterior mean estimate, the evidence suggests a countercyclical behavior of labor productivity, with this relationship weakening over time.

The model specification Base-P*CycOutputgap, which relates productivity to the output gap instead of the unemployment gap (through parameter ϕ^p), provides generally similar inference about the cyclical nature of labor productivity. The plot for the time-varying parameter ϕ^p is relegated to appendix A13 to conserve space. The credible intervals are wide and include both positive and negative values. Based on the posterior estimate of parameter ϕ^p, productivity is procyclical in the 1960s and post-2010, but it is either acyclical or countercyclical in other periods. Overall, the empirical evidence generally corroborates the evidence presented in Gali and van Rens (2020). The model comparison results indicate a slightly inferior fit of the Base-P*CycOutputgap model compared to Base (-608.1 vs. -606.9 in the case of productivity data and -1773.8 vs. -1771.7 for the overall fit).

The importance of SV in the productivity equation

ous sources, including historical Economic Reports of the President, document a similar pattern in the trend productivity estimates.

An essential element of our decomposition of labor productivity into the trend, cyclical, and idiosyncratic components, which others have abstracted from, is that we permit a time-varying variance of the idiosyncratic component. Empirically, our results show this to be an important extension, as shown in the SV plots included in panel (f) and the model comparison results reported in Table 6. The plots indicate statistically significant evidence of time variation in the volatility of the idiosyncratic component. The model comparison further provides evidence supporting SV inclusion, as the Base-NoSV is the worst performing and has a significantly worse fit to the productivity data compared to Base.

We also explored the possibility that the SV may be soaking up the variation in productivity, which otherwise would have been attributed to the cyclical component of productivity. The model specification Base-NoSV, which shuts down SV in the idiosyncratic component of productivity (and other model equations), yields estimates of λ^p similar to Base, suggesting that SV is not contributing to the ambiguous result on the cyclicality of labor productivity.

4.4. ***Estimation results for*** π***-star***

Panels (a) and (b) of Figure 7 plot the posterior mean estimates of pi-star along with the 90% credible intervals from the Base and Base-NoSurv model specifications, respectively. Panel (c) plots the corresponding precision estimates, defined as the width of the 90% intervals. A quick visual inspection shows that the pi-star from the Base specification is significantly more precise than Base-NoSurv, as evidenced by narrower credible bands and the precision plot corresponding to Base lying below the Base-NoSurv plot. Based on the marginal likelihood criteria, the fit of the inflation equation to the data in the case of Base is substantially greater than Base-NoSurv (as reported in Table 8). The fit of the overall Base model to the data is also higher than Base-NoSurv. Our finding that adding survey expectations improves both the model fit and the precision of pi-star is consistent with CCK.

The broad contours reflected in the posterior mean that pi-stars from the two models are similar to those documented elsewhere in the literature (e.g., CCK). For instance, pi-star was low in the 1960s, high in the 1970s, fell sharply in the 1980s, continued a steady deceleration in the 1990s, fluctuated in a narrow range between 2.0% and 2.5% in the 2000s, and has been below 2% since 2012. This general pattern is consistent with the widely held view. Focusing on the specifics, unlike some papers (e.g., Stock and Watson, 2007; Mertens, 2016), which show two peaks in pi-star, one in the mid-1970s, and another in the early 1980s, our model-based estimates (both with and without survey data) do not show the earlier peak (similar to CCK). Relatedly, in those same papers, pi-star is estimated to peak at a level of 10% or higher; in contrast, the mean estimate of pi-star in our model specifications peaks at a lower level (similar to CCK and Mertens, 2016 – in his model specification that is augmented with survey data).[54]

[54]Since both Stock and Watson (2007) and Mertens (2016) endow the RW process governing pi-star with SV, whereas we do not, this difference in the modeling assumption may explain the difference in the pi-star estimates around the Great Inflation period. However, CCK, who also allow SV in the pi-star process, yield pi-star estimates generally similar to ours, suggesting that the SV assumption is likely not the answer.

Comparing estimates from Base and Base-NoSurv specifications, the level of pi-star is similar in the 1960s but starting in early 1970, pi-star from the Base specification sharply accelerates, to peak at 6% in early 1980, while the estimates of pi-star from the Base-NoSurv specification also accelerate but peak at a lower level of 4.5%. As shown, uncertainty about pi-star increases sharply in early 1980, with the Base-NoSurv estimates experiencing a much more dramatic rise. It is the case that uncertainty around pi-star (as measured by the width of the 90% credible intervals) inferred from the Base-NoSurv is higher compared to the Base throughout the estimation sample. But in early 1980, the differential in uncertainty is twice as large, as can be seen comparing the dotted and solid plots in panel (c).

A similar rise in model-based estimates of pi-star uncertainty in the late 1970s through early 1980 (known as the Great Inflation period) has been noted elsewhere (e.g., Mertens, 2016). A subset of the literature attributes the rise in pi-star uncertainty to the un-anchoring of inflation expectations during the Great Inflation period. Beginning in early 1980 and continuing through early 2000, both models have (posterior mean of) pi-star steadily declining to 2%. Between 2000 and 2012, whereas in the case of Base, pi-star is flat at 2%, in Base-NoSurv, it is stable at a slightly higher level of 2.3%. Since 2012, pi-star has slowly moved down to reach 1.5% (in Base) and 1.4% (in Base-NoSurv).

Inflation gap persistence

Panel (d) shows the posterior mean estimates of parameter ρ^{π} for Base (solid line) and Base-NoSurv (dotted line). Also plotted are the 90% credible intervals from the Base model. As is evident, the uncertainty around the posterior mean is high. That said, both models indicate quite similar estimates of persistence in the deviations of inflation from pi-star. There is also strong evidence of time variation in inflation gap persistence. For example, gap persistence was low (0.25 to 0.35) in early 1960, but from there on began to increase steadily, reaching close to 0.85 by early 1970, and remained at that level through the early 1980s. Subsequently, the persistence had declined steadily to 0.4 by the early 1990s. From the late 1990s to the early 2000s, persistence fell further to 0.3, and it remains at that level.

Price Phillips curve

Panel (e) plots the posterior mean estimate of parameter λ^{π}, which is the slope of the price Phillips curve. As before, the plots from the main model specifications are shown. (The 90% credible intervals are from the Base model.) The plots indicate that both model specifications yield very similar estimates of the parameter linking the inflation gap to the unemployment gap. The estimates also indicate strong evidence of time variation in the slope of the Phillips curve. For example, both models estimate a steeper Phillips curve in the 1960s that subsequently weakens (becomes less negative) over time through 2010. From there on, it slowly begins to become steeper (more negative), ending 2019 at -0.23, which is still weak, historically speaking, and is surrounded by wide intervals spanning -0.05 to -0.52. This pattern is consistent with the familiar narrative (also documented in several other papers) that the *"Phillips curve has weakened over time."*

SV in price inflation equation

Panel (f) plots the volatility estimates (i.e., the standard deviation of the shocks to the inflation gap, $e^{h_t^\pi}$) corresponding to the two model specifications. The figure shows that the posterior mean estimates are quite similar. The estimates imply high volatility during the period of the Great Inflation that fell subsequently. Inflation volatility increased sharply again during the Great Recession but has trended lower since then. By 2019, inflation volatility had returned to the low levels of the early 2000s but remains shy of the historic lows of the mid-1960s and mid-1990s. The contours of the volatility in the inflation gap are generally similar to those reported in CCK and Chan, Koop, and Potter (2013). Bayesian model comparison indicates strong evidence supporting the inclusion of SV in the price inflation equation.

Link between survey and pi-star

In Figure A6 (online appendix), panels (a) and (b) plot the posterior estimates of the coefficients C^π and β^π, which provide a sense of the estimated relationship between survey forecast and pi-star. The prior assumes an unbiased relationship between survey forecasts and pi-star (i.e., prior mean $\beta^\pi = 1$ and at all periods $C_t^\pi = 0$). The model estimation yields a posterior mean estimate of 0.99 for β^π, with 90% credible intervals spanning 0.91 to 1.07 (also reported in Table 2). The posterior mean of C_t^π displays considerable time variation; however, the 90% credible intervals, for the most part, include zero. Taken together, the posterior estimates of β^π and C_t^π imply that the survey forecast, on average, is a somewhat biased measure of pi-star. This latter result confirms the findings of CCK.

In the online appendix A11.a, we include the results and a discussion comparing pi-star estimates from the Base model to external models: CCK, CKP, and UCSV. The estimates indicate that the CCK model generates the most precise pi-star, followed by the Base model, CKP, and UCSV. The latter model yields volatile and erratic estimates of pi-star (and precision).

Summary: pi-star

All told, we summarize the analysis of the inflation block as follows. First, the Base model and its variants indicate contours of pi-star that corroborate the narrative documented elsewhere in the literature. However, in some periods, pi-star estimates can differ notably across models, and as emphasized in CCK, these differences can have important implications for monetary policy. Second, comparing across Base and Base-NoSurv specifications, and comparing the Base specification with outside models, strongly suggest the usefulness of survey forecasts in improving the econometric estimation of pi-star (i.e., survey forecast information yields sensible estimates of pi-star and improved precision), hence, the corroborating evidence in CCK, Mertens, 2016, and Nason and Smith (2021). Third, we find evidence in support of incorporating time variation in the price Phillips curve, specifically, suggesting, in a broad sense, a weakening of the relationship between inflation and labor market slack since the 1960s, something also documented by several others (e.g., CKP; Stella and Stock, 2015; Del Negro et al., 2020).

Fourth, as found by CKP, CCK, Mertens and Nason (2020), among others, we find evidence supporting time variation in the persistence of the inflation gap. Fifth, as noted by many, including Stock and Watson (2007) and Clark and Doh (2014), we find strong evidence of SV in the innovations of the inflation equation. Sixth, as in CCK, we find evidence that the survey forecast of PCE inflation is a biased measure of pi-star. Lastly, the Base model's improved fit to inflation data compared to some of its variants provides evidence supporting the long-run theoretical restriction defined by equation (28), which imposes w-star = pi-star + p-star.

4.5. ***Estimation results for W-star***

In modeling w-star, a novel feature of our framework is the decomposition of w-star into its fundamental components, pi-star and p-star. Figure 8 presents posterior estimates of w-star along with the decomposition. The first row in the figure plots estimates from the Base model, and the second row plots estimates from the Base-NoSurv. The third row plots p-star estimates from other model variants alongside Base and Base-NoSurv models, and also presents precision estimates of w-star.

The estimates imply that w-star increased steadily in the 1970s and peaked in the early 1980s. This increasing w-star reflected an upward drift in pi-star that more than offset the downward drift in p-star, as evidenced by the widening in the shaded area representing pi-star and the slight narrowing of the shaded area representing p-star. The Base model implies that w-star peaked at 7.3% in the early 1980s, while Base-NoSurv has w-star peaking at 6%. Both models have w-star sharply drifting lower through much of the 1980s, to reach near 4% by early 1990. From there on, the path of w-star across the two models is very similar and indicates a gradual slowing to 2.5% by the end of 2017. W-star moved up to 3.0% by the end of 2018, only to fall back to 2.8% in 2019.

Not surprisingly, w-star is more precisely estimated in the Base model than in the Base-NoSurv model, as shown in panel (f). The considerable uncertainty around w-star, implied by the Base-NoSurv model during the 1970s, is mostly driven by pi-star. As noted earlier, the pi-star estimate from the Base-NoSurv was highly imprecise during the 1970s. Interestingly, despite the inferior precision of the Base-NoSurv compared to Base, the Bayesian model comparison suggests a better fit of the Base-NoSurv model to the nominal wage data than Base (see Table 7). However, the overall fit of the Base-NoSurv to the data (which include data beyond the nominal wage) is slightly worse than Base.

Sensitivity of w-star to modeling assumptions

Panel (e) plots estimates of w-star from two additional Base model variants: Base-W*RW and Base-NoPT. The Base-W*RW model eliminates the long-run restriction that w-star is the sum of p-star and pi-star (on average) and instead models w-star as an RW process. Interestingly, the path of w-star implied by Base-W*RW is similar to Base-NoSurv through the mid-1980s. From there on, w-star from Base-W*RW is below the Base-NoSurv through early 2010. Since then, it is identical to Base and Base-NoSurv. Although in the first half of the estimation sample, w-star from Base-W*RW is less precisely estimated than Base, in the second half of the sample, it is more precise. According to the Bayesian model comparison, the fit of the Base-W*RW to the nominal wage data is inferior to both Base-NoSurv and Base. However, compared to Base, the degree of inferiority is only slight. The overall fit of the Base-W*RW model to the data is substantially worse than that achieved by either the Base or the Base-NoSurv model.

The Base-NoPT model eliminates the pass-through from prices, i.e., it removes the price inflation gap from the equation describing the wage inflation gap. In other words, the direct link between the cyclical components of prices and nominal wages is eliminated, but the connection between the permanent components pi-star and w-star remains. Doing so has notable implications for the estimate of w-star and the model's fit. As shown in panel (e), w-star implied from Base-NoPT is higher than that implied by the other models through the first half of the sample. While the Base model has w-star peaking at a little above 7% in the early 1980s, the Base-NoPT model implies a higher peak of 8.2%. The acceleration in w-star implied by the Base-NoPT model during the 1970s is much stronger than that implied by the Base model estimates. This stronger path of w-star is associated with more precise estimates of w-star in the 1970s compared to the Base and other models, as can be seen in panel (f). However, according to the Bayesian model comparison reported in Table 7, removing the connection between the cyclical components negatively impacts the model fit, as evidenced by the substantially inferior fit of the Base-NoPT model to data compared to the Base model.

Wage persistence, wage Phillips curve, pass-through from prices, and SV in the wage equation

Figure 9 presents the time-varying posterior estimates of the parameters describing the persistence in the nominal wage gap, the wage Phillips curve, the short-run pass-through from prices to wages, and the stochastic volatility of the shocks to the nominal wage equation. Estimates from three models, Base, Base-NoSurv, and Base-W*RW, are presented. A quick visual inspection indicates both statistically and economically significant evidence of time variation in these parameters that capture important empirical relationships.

Wage gap persistence

Panel (a) plots the posterior estimates of the parameter ρ^w, capturing the persistence in the nominal wage inflation gap. Also included are the 90% credible intervals from the Base model. The posterior mean estimates for the three models indicate increasing persistence in the wage inflation gap beginning in early 1960 and peaking in mid-1980. The Base model shows a notably higher peak at 0.45 compared to 0.39 and 0.33 for Base-NoSurv and Base-W*RW, respectively. The higher peak suggests that the Base model attributes a higher share of fluctuations in cyclical nominal wages to the persistence component than that implied by the other two models. From the mid-1980s to the early 1990s, the persistence steadily declines but after that increases through the mid-2000s; from there on through the early 2010s, the estimated persistence in the nominal wage gap falls to levels seen in the mid-1970s. Since then, it has been slowly increasing. It is worth noting that the credible intervals around the posterior mean are wide, suggesting high uncertainty in the inference about the estimated persistence.

Wage Phillips curve

In Figure 9, panel (b) plots the estimates of parameter λ^w, which captures the wage Phillips curve relationship in the US data. The plot provides strong evidence supporting the existence of the wage Phillips curve in the post-war data. Notably, the plot also offers convincing evidence of time variation in this relationship. According to the posterior mean estimate, from the early 1960s through mid-1970s, our models imply that the strength of the wage Phillips curve is at a moderate level, but from there on through the mid-1980s, the wage Phillips curve steepened sharply in Base and Base-NoSurv (less steep in Base-W*RW). By the mid-1980s, in the Base model, the posterior mean of the wage Phillips curve parameter is estimated to be -0.5, with 90% credible intervals ranging from -0.23 to -0.75. From there on, it gradually flattened until the mid-2000s, but soon afterward, it began to flatten more rapidly until early 2010. In 2010, all three models estimate the posterior mean of the Phillips curve parameter at -0.2. The prevalence of the downward wage rigidities during the Great Recession is among the primary explanations for the flattening wage Phillips curve (see Daly and Hobijn, 2014).

From 2010 onward, with an improving economy, the estimated wage Phillips curve has steadily steepened. Our empirical evidence on the wage Phillips curve is consistent with the findings of Knotek II and Zaman (2014), Peneva and Rudd (2017), and Galí and Gambetti (2019), all of whom document strong support for the continuing existence of a wage Phillips curve in the US data.

Pass-through from prices

Panel (c) plots the estimates of the short-run pass-through from prices to wages defined as $\frac{\kappa_t^w}{1-\rho_t^w}$. The posterior estimates of pass-through indicate a weakening relationship between cyclical nominal wage inflation and cyclical price inflation over the estimation sample, confirming the evidence presented in Peneva and Rudd (2017) and Knotek II and Zaman (2014). The relationship between the two was strong in the 1970s to the mid-1980s, but since then, it has gradually weakened such that it has been nonexistent (i.e., the pass-through is estimated to be zero) for the past decade. This period of the breakdown in the relationship between the two cyclical components has coincided with a period of low and stable price inflation. The literature has offered various explanations for this breakdown in the relationship, including an improved anchoring of inflation expectations (Peneva and Rudd, 2017) and an amplification of downward wage rigidities during low levels of price inflation (Daly and Hobijn, 2014).

Our empirical finding also supports the results of Bobeica, Ciccarelli, and Vansteenkiste (2019), who using euro area and US data, show that the link between labor compensation and price inflation in the short run importantly depends on the prevailing inflation regime. They find that a high inflation regime is associated with a tighter connection between labor costs and price inflation and a low inflation regime is associated with a weaker link. All three of our models imply similar inference. Both Base-NoSurv and Base-W*RW indicate a stronger pass-through than the Base model in the 1970s and 1980s, periods associated with high inflation in the US. Comparing between panels (a) and (c) suggests that during that period, Base attributes more of the increase in nominal wage inflation to an increase in persistence than to the pass-through from price inflation, hence the less strong pass-through estimate than seen in the Base-NoSurv and Base-W*RW models.

SV in the wage equation

Panel (d) plots the time-varying standard deviation of the innovations to the nominal wage inflation gap. The plot indicates the importance of allowing for SV in the equation for nominal wage inflation. All three models show very similar posterior mean estimates of SV. The estimates suggest that volatility in the early 1960s was high but then subsequently fell, rising again in the late 1960s and early 1970s. From there on through the mid-1970s, it fell steeply. It increased in the early 1980s, but less sharply than in the previous decades. From the mid-1980s through the early 1990s, volatility moderated to low levels, and since then, it remains flat at that low level. The Bayesian model comparison indicates further evidence supporting the inclusion of SV in the wage equation, as evidenced by the substantially inferior fit of the Base-NoSV model to the nominal wage inflation data compared to the Base and Base-NoSurv models (Base-NoSV: -344.3 vs. Base: -277.5; see Table 7).

4.6. ***Estimation results for r-star***

Figure 10 presents r-star and the "catch-all" component D estimates for our two main model specifications. The top row of the figure plots the estimates of r-star (panel a) and component D (panel b) from the Base model. Also included in panel (a) are the survey expectations of r-star (which enters our Base model). As can be seen, the contours of (posterior mean) r-star from the Base model track the survey estimate; this suggests that survey data play an influential role in guiding the model's assessment of r-star. The posterior mean estimate from the Base model shows r-star staying relatively flat at 3.5% in the 1960s, and then slowly trending down through the 1970s, to reach 3% by early 1980. Thereafter, it fluctuates in a range between 2.8% and 3.5% until the beginning of 2000. From there on, it steadily declines, reaching 1.1% at the end 2019.

Panel (b) plots the estimate of component D, whose dynamics are shaped by the survey expectations data and by information from the Taylor rule and IS equations. As can be seen in the figure, component D is imprecisely estimated. According to the posterior mean estimate, in the 1960s, component D exerts slight upward pressure on r-star that is mostly offset by downward pressure coming from g-star (via equation 36), helping keep r-star relatively flat. After that, with D remaining flat through 2000, developments in g-star shape the trajectory of r-star. From 2000 onward, all forces (as captured through the model structure) work in the same direction to push r-star steadily downward. The estimated link between r-star and g-star is of moderate strength (posterior mean of parameter $m = \frac{\zeta}{4} = 0.701$); see Table 2); therefore, movements in g-star play an influential role in driving r-star.

Moving on to the Base-NoSurv model specification, in panel (c), the mean estimate shows r-star rising from 2% in early 1960 to 3.5% through early 1980 and then remaining stable through early 2000. This trajectory is similar to that in González-Astudillo and Laforte (2020), who also use the Taylor rule and information from long-term interest rates in their estimation. From 2000 onward, r-star steadily declines, reaching 1.4% at the end of 2019. The trajectory of r-star from 2000 onward is similar to that from the Base model. It is worth noting that our models' indication of a secular decline in r-star beginning in 2000 is also documented elsewhere in the literature tackling r-star (the exception being JM).[55] However, the extent of decline varies considerably across studies. The literature offers various explanations for this secular decline in r-star, including a trend decline in g-star (e.g., Laubach and Williams (2016)); rising premiums for convenience yield, i.e., increased demand for safe and liquid Treasury bonds (see Del Negro et al., 2017; Bullard, 2018); and excess global savings (Pescatori and Turunen, 2016).

The uncertainty around the r-star estimate from the Base-NoSurv model is substantially higher than that in Base, as can be seen by comparing panel (a) and panel (c) of Figure 10, and also shown in panel (b) of Figure 11. The increased uncertainty in r-star comes from component D, which is imprecisely estimated without the survey data. Furthermore, based on the marginal likelihood criteria, the Base model is favored over the Base-NoSurv model (see Table 9).

Without the survey information about r-star, the estimated link between g-star and r-star is significantly weaker (posterior mean of $m = \frac{\zeta}{4} = 0.390$; see Table 2), which is consistent with the evidence documented in Hamilton et al. (2016) and Lunsford and West (2019). Therefore, the movements in component D significantly dominate the contours of r-star in the Base-NoSurv model. Both the IS curve and Taylor-rule equations shape the evolution of component D. The hump-shaped patterns in both D and r-star reflect the trends in real long-term interest rates (informed from the IS equation) and short-term interest rates (from the Taylor-rule equation). It is interesting to note that the r-star estimates from JM (and González-Astudillo and Laforte,

[55]JM document that r-star from their preferred specification (which allows for SV in the TR equation) is generally flat over their sample, spanning 1960 through 2018. However, in an alternative specification, which does not permit SV, the r-star estimate exhibits a decline in r-star similar to that documented elsewhere. In our examination, r-star's trajectory is little changed comparing between the Base specification and Base without SV (i.e., Base-NoSV).

2020) – who use the Taylor-rule equation and information from both short- and long-term interest rates – also exhibit hump-shaped behavior (though, in the case of JM, it is only slight). As we show shortly, the prior setting on the shock process for r-star (in our Base and Base-NoSurv cases, component D) plays an essential role in shaping the contours of r-star.

As discussed earlier, the estimates of g-star from both the Base and Base-NoSurv models are quite similar. Therefore, the primary source of the differential in the r-star estimates between Base and Base-NoSurv is the quantitatively weaker relation estimated between r-star and g-star in Base-NoSurv than Base.

Assessment of policy stance

Panel (e) of Figure 10 provides an assessment of the stance of monetary policy. Following Pescatori and Turunen (2016), we gauge monetary policy's stance as the deviation of the short-term nominal interest rate from the long-run nominal neutral rate of interest (defined as the sum of r* and pi*) – this is the interest rate gap from the Taylor-rule equation. A positive interest rate gap characterizes a restrictive monetary policy stance, and a negative interest rate gap implies a stimulative stance. The solid line corresponds to the policy stance inferred from the Base model and the dashed line to that inferred from the Base-NoSurv model. Even with notable differences in the estimates of r-star across the two models, the assessment of the policy stance is remarkably similar throughout the sample. So why are they so similar?

The answer lies in the differences in pi-star estimates across two model specifications, as shown in Figure 7 (panels a and b). In other words, the differences between r-star estimates across the two models are compensated (i.e., offset) by the differences between pi-star, such that the assessment about the stance of monetary policy across the two models is strikingly similar. For instance, in the 1960s, the r-star estimate from the Base is on average 1.34 percentage points (ppts) higher than that of Base-NoSurv. However, over the same period, the pi-star estimate from the Base model is 0.54 ppt lower than that of the Base-NoSurv model, which reduces the difference between their associated assessments of policy stance.

According to our model(s) estimates, the policy stance appeared to be slightly restrictive before the Great Recession, but at the onset of the Great Recession, the policy stance immediately turned accommodative. Since then, it has remained very accommodative (reflecting the effects of unconventional monetary policy). After peaking in late 2015, the degree of accommodation has gradually declined (i.e., the interest rate gap has become less negative), such that, by the end of 2019, it has edged closer to the neutral threshold.

A closer inspection of the figure reveals an interesting insight. Since 1990, both the degree and duration of policy accommodation in response to the recession have been more significant than in the previous recession. For instance, the monetary policy stance was more accommodative in terms of both level and duration following the 2001 recession than following the 1990-1991 recession. Similarly, during and following the Great Recession, the policy stance, in terms of level and duration, was more significant than following the 2001 recession. Broadly speaking, our model-implied assessment of the stance of monetary policy mirrors that in the assessment documented in Brand and Mazelis (2019).

Random walk assumption for r vs. Base (and Base-NoSurv)*

As is commonly done when estimating stars, a random walk assumption for r-star is a popular choice (e.g., Kiley, 2020; JM). Accordingly, we explore this particular modeling choice's fit to the data and empirical properties by replacing the equation linking r-star to g-star and the RW component D with an equation that assumes a random walk process for r-star in our main model specifications (Base and Base-NoSurv). We denote these specifications, Base-R*RW and Base-NoSurv-R*RW. By adopting an RW assumption for r-star, the models will be unable to uncover any specific causes of movements in estimates of r-star.

Figure 11, panel (a) plots the posterior mean r-star estimates from the Base-R*RW and Base-NoSurv-R*RW models. To facilitate comparison, also shown are estimates from Base and Base-NoSurv models. Panel (b) plots the corresponding r-star precision estimates (defined as the width of the 90% credible intervals). A few observations immediately stand out. First, the estimated r-stars from model specifications with the RW assumption, although exhibiting broadly similar contours, are higher than those obtained from the specifications that impose a relationship between r-star and g-star. Second, from 2000 and onward, r-star estimates from the specifications with the assumed link between r-star and g-star experience a more stark decline than those from the specifications with the RW assumption. For instance, by late 2019, the estimates of r-star from both the Base-R*RW and Base-NoSurv-R*RW models settle at close to 2%, whereas those from the Base and Base-NoSurv models fall further to a range of 1.2 to 1.4%. This differential is mostly explained by the lack of direct downward pressure from g-star in the specifications with the RW assumption.

Third, bringing in information from surveys improves the precision of the r-star estimates substantially, irrespective of whether r-star is modeled simply as an RW process or as a combination of an RW component and a component linking r-star to g-star. Fourth, the specifications that allow for a link between r-star and g-star are more precise than those that do not. The Base specification generates the most precise r-star estimates relative to those obtained from the other three specifications.

Based on the model comparison metric reported in Table 9, the model indicating the most precise r-star does not necessarily rank as the best fitting model. The fit of the Base specification is slightly inferior to that of both the Base-R*RW specification and the Base-NoSurv-R*RW, but not by very much. This evidence nicely illustrates that the marginal likelihood metric has a built-in penalty that increases as the model complexity goes up. When we compare between the Base-NoSurv-R*RW and the Base-R*RW specifications, the addition of survey data increases the model's fit to the interest rate data slightly from -215.58 to -214.04 (but reduces overall model fit marginally from -1769.3 to -1770.8). In comparison, moving from the Base-NoSurv to the Base specification, the fit to the interest rate data increases, from -221.98 to -216.4 (and the overall model fit from -1772.8 to -1771.7).

Taken together, the evidence described above suggests that the RW assumption for r-star is a viable option, but bringing in information from surveys helps improve precision of the estimates of r-star substantially. Another result worth highlighting from the model comparison exercise concerns the marginal value of survey information in the estimation of r-star. In a model specification that imposes a relationship between r-star and g-star (e.g., Base; Base-NoSurv), adding survey information is crucial in making the model a competitive alternative. By bringing in survey information, which has a high correlation between g-star and r-star survey expectations (0.8), the estimated link between g-star and r-star in the model becomes stronger ($m = \frac{\zeta}{4} = 0.701$ in Base vs. $m = \frac{\zeta}{4} = 0.390$ in Base-NoSurv).

Given the evidence on model comparison and the precision of r-star estimates, our preferred choice is the Base model because its fit to the data is competitive (just marginally inferior to the specification with the RW assumption). And it produces the most precise r-star estimate compared to the alternative specifications. An additional factor that influences our choice is the model's ability to provide an explanation for movements in r-star, which is made possible because of its direct link to g-star.

In a supplementary set of exercises, we illustrate the usefulness of the Taylor-type rule equation and the equation linking r-star to survey expectations for identifying r-star. The addition of the Taylor-type rule equation turns out to be crucial to yielding plausible and precise estimates of r-star (see appendix A10.d).

We also explored the role of data versus priors in determining r-star, and in the interest of brevity, the results of the exercises are relegated to the online appendix (see A10.a). Here, we briefly mention that the results indicate that in the case of the Base-NoSurv model, the prior views determine the shape of the posterior for r-star, confirming Kiley (2020). However, in the Base model, which links r-star to survey expectations, the data do have some influence. In the case of the Base-NoSurv model, we also find that if we loosen the prior on the variance of the r-star process (to values similar to those in Kiley), then the data begin to shape the posterior. But this comes at the cost of a worsening fit to the interest rate data.

Overall, we find that when it comes to r-star, specification choices matter a lot (something also highlighted by Clark and Kozicki (2005), Beyer and Wieland (2019), and Kiley (2020)). The model specifications that include survey expectations yield estimates that are both reasonable and the most precise. In our examinations, the Base specification, which allows for a link between g-star and r-star, is equally preferred by the data to the variant of the Base specification that assumes r-star as an RW. Since 2000, the best fitting model specifications indicate a steady decline in r-star, similar to results documented elsewhere in the literature.

5. Real-time Estimates and Forecasting

In this section, we perform two real-time, out-of-sample forecasting exercises. In the first exercise, we compare the real-time forecasting performance of our two main models, Base and Base-NoSurv. We evaluate both the point and density forecast accuracy for real GDP growth, PCE inflation, the unemployment rate, nominal wage inflation, labor productivity growth, and the shadow federal funds rate. We show that the Base model is more accurate on average compared to Base-NoSurv for all variables of interest except the unemployment rate. We also document our Base model's superior forecasting properties relative to "hard to beat" benchmarks, including some of the recently proposed UC models for inflation forecasting. By-products of our real-time forecasting exercise are the real-time estimates of the stars from 1999 through 2019. We compare these real-time estimates to the final (smoothed) estimates – based on the entire sample spanning 1959Q4 through 2019Q4.

In the second forecasting exercise, we illustrate the efficacy of the stars' estimates produced from our models by demonstrating their usefulness in forecasting with external models (e.g., steady-state VARs). We find that the quality of our estimates of the stars from the Base model is generally competitive with the survey estimates, which are commonly used as proxies for stars in VAR forecasting models. For brevity purposes, we relegate the discussion of this latter exercise to the online appendix A5.

Table 10 presents the results from comparing the out-of-sample forecasting performance (both point and density) between the two models over the forecast evaluation sample spanning 1999Q1 through 2019Q4. The forecast evaluation is based on real-time data vintages and uses a recursively expanding estimation window, where each recursive run uses an additional quarterly data point in the estimation sample.[56] The forecast accuracy (point and density) is computed from one-quarter ahead to 20 quarters out. Partly due to our focus on the medium-term horizon and partly in the interest of space, we report accuracy metrics for 4, 8, 12, 16, and 20 quarters ahead. We evaluate the forecast accuracy using real-time data; specifically, we treat the "actual" as the *third* release of a given quarterly estimate.[57] For instance, in the case of real GDP, the third estimate for 2018Q4 corresponds to the GDP data available in late 2019Q1. The point forecast accuracy is assessed using the root mean squared error (RMSE) metric, and the density forecast accuracy is evaluated using the log predictive score (LPS). The statistical significance of the point forecast accuracy is gauged using the Diebold-Mariano and West test and, in the case of the density forecast, accuracy is based on the likelihood-ratio test of Amisano and Giacomini

[56]Going back in time means that we are using relatively fewer observations to estimate model(s). As is commonly done when performing real-time forecasting using multivariate UC models, we need to tighten priors on the shocks' variances driving the latent components (see, for instance, Barbarino, Berge, Chen, and Stella, 2020). Accordingly, we devise a systematic approach to adjusting the prior on the scale parameters of the inverse gamma distributions defining the variances of the stars. We multiply the scale parameter with the $factor = (\frac{2T}{N} - 1) * (\frac{T}{N+5(N-T)})$, where N is the total sample size from 1959Q4 through 2019Q4, and T refers to the number of data points in a given data vintage. At the end of the sample, the $factor = 1$ because $T = N$.

[57]Results are qualitatively similar if we instead use the revised data (2020Q1 vintage data) as the actual values in the forecast evaluation exercises. The results are available on request from the author.

(2007).

The top panel of the table reports the results corresponding to the point forecast accuracy, while the lower panel reports results for the density forecast accuracy. The numbers reported in the table correspond to relative RMSE –RMSE Base relative to RMSE of Base-NoSurv – in point forecast comparison, and relative mean LPS – LPS Base minus LPS Base-NoSurv – in the case of density forecast comparison. Hence, numbers less than one in the top panel suggest that the point forecast accuracy of the Base forecast is more accurate on average, and positive numbers in the bottom panel suggest that the density forecast accuracy of the Base forecast is more accurate than that of the Base-NoSurv forecast.

As is evident by the numbers reported in the table, except for the point forecast accuracy of the unemployment rate, the evidence generally favors the Base model as more accurate than Base-NoSurv. The evidence in support of the Base model is strongest for PCE inflation and labor productivity growth. In the case of the unemployment rate, as was the case with the in-sample model fit (see Table 3), the Base-NoSurv model outperforms the Base model in point forecast accuracy. However, over the forecast evaluation sample, the uncertainty around the point forecast of the unemployment rate implied by the Base-NoSurv model was higher than that of the Base model (on average). This higher uncertainty contributed to inferior density forecasts from the Base-NoSurv model compared to the Base model, as is evident by the positive numbers in the row corresponding to the unemployment rate. It is generally the case across all variables, except real GDP growth, that density forecasts from the Base model are more accurate than those of the Base-NoSurv. In the case of real GDP growth, even though, for the most part, the point forecast accuracy between the two models is similar on average, the density forecasts from the Base-NoSurv model are more accurate than those from the Base model.

We also compared the forecasting performance of our Base model to the outside benchmark models, which the forecasting literature has shown to be useful forecasting devices. Specifically, we compare the accuracy of the inflation forecasts from our Base model to the following three models, UCSV of Stock and Watson (2007) [UCSV], Chan, Koop, and Potter (2016) [CKP], and Chan, Clark, and Koop (2018) [CCK]. We compare the accuracy of the unemployment rate forecasts from our Base model to the Chan, Koop, and Potter (2016) model, and the accuracy of the nominal wage inflation from the Base model to the UCSV model applied to the nominal wage inflation – motivated by Knotek II (2015). Our forecasting results provide strong evidence in support of our Base model's competitive forecasting properties. For the sake of brevity, the results are included in the online appendix A4 table A3.

Real-time versus final estimates

Up to this point, we only examined the smoothed estimates of the stars inferred using all the sample data, i.e., from 1959Q4 through 2019Q4, which we denote here as final estimates. As discussed in CKP and Clark and Kozicki (2005), the examination of final estimates is beneficial for "historical analysis," such as the evaluation of past policy. But for real-time analysis, such as forecasting and policymaking, real-time estimates at time t – estimates based on data and model estimation through time t (instead of through 1:T) – are the relevant measures. In estimating the stars, a voluminous number of papers have documented the typical pattern of notable differences between real-time and final estimates; e.g., see Clark and Kozicki (2005) and Beyer and Wieland (2019) for r-star and Tasci (2019) for u-star.[58]

Relatedly, several researchers have attributed the inability to precisely know the location of these stars in real time to past policy mistakes; see Powell (2018) and references therein. The documented differences between the real-time and final estimates, which at times could be dramatic, and the recognition of these differences by policymakers have been the primary reason limiting the usefulness of real-time estimates of the stars in policy discussions in recent years (see Powell, 2018). Hence, there is a strong preference for methods that can provide more credible inferences about stars in real time.[59]

Comparing real-time and final estimates of the stars from our Base model suggests that we have made some progress in mitigating the difficulties in previous real-time estimation of the stars. However, in the Base-NoSurv model, there is less success in mitigating this issue. We believe a big reason for this lack of success in the latter case is that we estimate a very high-dimensional model with a lot less data (as will be the case when stopping estimation at earlier periods). An artifact of this is that it requires the imposition of very tight priors in earlier periods than when estimating with more recent periods, which, in turn, affects the posterior estimates of model parameters and the stars. This latter fact mechanically contributes to more considerable observed differences between real-time and final estimates in the first half of the sample period analyzed. In the case of the Base model, the use of survey information helps anchor the estimates to more reasonable values even in the face of tight priors.

Figure 12 plots the real-time posterior mean estimates of the stars from 1999Q1 to 2019Q4. Also plotted are the corresponding final (posterior mean) smoothed estimates and the 68% and 90% credible bands, respectively. The real-time estimates are the end-sample posterior mean (of the smoothed) estimates at any given period. For example, the 1999Q1 estimate corresponds to estimating the model(s) from 1959Q4 to 1999Q1; similarly, the 1999Q2 estimate corresponds to estimating the models from 1959Q4 to 1999Q2. As can be seen, for the most part, the real-time estimates of the stars generally remain within the credible intervals, especially the 68% credible sets implied based on full-sample information. The exception is g-star, and in turn, the output gap; the latter is shown in Figure A3 of the online appendix for brevity. In the case of g-star, the real-time estimate, which before 2007 is not aware that the Great Recession (GR) is about to

[58]Both revisions to past data and the accrual of additional data could contribute to the observed differences between the real-time and final estimates. The estimation with additional data has been found by many to be the primary factor causing revisions to historical estimates of the stars and contributing to divergence between the real-time and final estimates (see Tasci, 2019; Clark and Kozicki, 2005).

[59]The issue of imprecision in the estimation of stars is an important one. It has been long recognized that considerable uncertainty surrounds the estimated stars complicating reliable inference (e.g., Cross, Darby, and Ireland, 2008). Despite the stars' imprecision, they continue to be used as inputs into policymaking and for other purposes (see Williams, 2018; for r-star). After all, as discussed in Mester (2018), uncertainty about the stars is just one source of uncertainty among many that confront policymakers.

occur, is relatively upbeat, like the survey estimate (at 2.9% vs. 2.25% for the final). Between 2007 and 2011, the real-time estimate gradually decelerates to 2.5%, but from there on through 2015, it declines very rapidly, catching up to the final estimate.

Interestingly, the BC survey (which is real-time data) fell by only two-tenths over this period, but between 2015 and 2017, it dropped another three-tenths to 2.0% and remained at that level through the end of 2019. Whereas beyond 2017, the survey estimate remained flat, the real-time model estimate gradually moved back up, partly in response to stronger realizations in real GDP growth than the implied g-star. The more gradual decline in the real-time g-star during the GR implies a larger (negative) output gap estimate than is indicated by the final estimate.

In the case of u-star, interestingly, in the first half of the forecast evaluation sample, the real-time estimate generally fell outside the 68% credible intervals but inside the 90% intervals. In contrast, in the second half, it remained inside the 68% intervals and tracked quite closely the posterior mean of the final estimate. In the case of p-star, the real-time posterior mean estimate largely remained inside the 68% intervals, and post-2007, the implied inference on p-star is consistent with the final estimate, as is evident by the similar contours of the real-time and final estimates. It is worth noting that both the real-time and the final estimates of p-star are notably lower than the survey estimate of p-star; recall that the survey estimate of p-star is not used in the estimation of the Base model.

In the case of pi-star, the real-time estimate closely tracked the final estimate from 1999 through 2004, and from there on through 2016, it averaged three-tenths higher. From 2017 through 2019, the real-time estimate once again closely tracked the final estimate. The contours of the real-time and final estimates are similar, and the real-time estimate remained within the 90% credible intervals throughout the sample period analyzed. Consistent with a generally higher real-time estimate of p-star and pi-star than the final estimate, the real-time estimate of w-star is higher than the final. Just as in the case of p-star and pi-star, the contours of the real-time and final estimates of w-star are similar, and the real-time estimate of w-star remained within the 90% credible intervals. Also plotted are the implied survey estimates of w-star, constructed by adding p-star and pi-star survey estimates. Arguably, these implied estimates are implausibly high given that the actual realizations of nominal wage inflation in the post-2007 period were mostly below these implied estimates. The results of a forecasting exercise presented in the online appendix further confirm the implausibility of the implied survey estimate.

In the case of r-star, the contours of the real-time and final estimates are remarkably similar. Between 2000 and 2005, and post-2014, the real-time estimate closely tracked the final estimate. From 2006 through 2013, the real-time estimate averaged 35 basis points higher than the final estimate. In our assessment, the magnitude of the gap between real-time and final estimates is relatively small compared to the uncertainty estimate around the posterior mean and estimates of uncertainty typically reported in papers estimating r-star, e.g., Clark and Kozicki (2005), Laubach and Williams (2016), and Lubik and Matthes (2015). Furthermore, the real-time estimate of r-star remained within the 68% credible intervals throughout the sample period considered. The width of the estimated 68% (and 90%) intervals from the Base model has been remarkably stable between 0.9 and 1.2 percent (1.5 and 2.0 percent) in the last 25 years. For reference, the typical estimates of 90% bands from popular models such as LW and Lubik and Matthes (2015) have a width averaging more than 4% and 3.5%, respectively.[60] Given that the 68% and 90% credible intervals are significantly narrower compared to typical estimates reported elsewhere in the literature, we view the evidence of our real-time r-star remaining inside the estimated credible intervals as encouraging.

Overall, the real-time estimates of stars and forecast evaluation based on the past 20 years of data provide empirical evidence supporting the competitive forecasting and real-time properties of the Base model. Unfortunately, the high dimensionality of our models and the limited availability of real-time data on nominal wage inflation prevent the evaluation of the models' forecasting properties over a more extended historical sample.

6. The Implications of the COVID-19 Pandemic For Stars

At the time of writing this paper, the global economy is in the midst of an ongoing global pandemic crisis (GPC), which has continued to inflict significant disruption on economic activity both in the US and globally. The GPC, which started in early 2020, contributed to extreme movements in many US macroeconomic indicators, including those used in this paper. For instance, US real GDP growth in quarterly annualized terms declined from -5% in Q1 to -31% in Q2, the deepest contraction in the post-war data (COVID-19 recession). And in Q3, growth rebounded to +33%, a record increase in the post-war data. These extreme movements, which were several standard deviations away from their historical averages, contributed to the breakdown of many conventional time-series models, especially the time-invariant VAR models estimated with monthly data; see Lenza and Primiceri (2020) and Carriero, Clark, Marcellino, and Mertens (2021).

Up to this point, our analysis has focused on the pre-GPC data. In light of recent work documenting the difficulties of the standard time-series models handling pandemic data, we are naturally curious to see how our two main models respond to the COVID-19 GPC data. What do data during the pandemic imply for estimates of stars, or, in other words, what are the estimated long-run consequences of the GPC data? As we illustrate shortly, the short answer to the first question is that both models, Base and Base-NoSurv, handled the pandemic data well. But the Base model, in our judgment, did a better job than the Base-NoSurv.

The answer to the second question is a bit complicated. But, briefly, the models indicate that, at the moment, the long-run consequences of the economic developments related to the pandemic are highly uncertain. We view this latter characterization as a reasonable one. Let's

[60] We note that recent approaches to model r-star such as JM and Del Negro et al. (2017) also generate precise intervals similar to ours, with JM marginally less precise and Del Negro and all measurably more precise than ours.

take long-term productivity growth. Based on past research, pandemics are associated with a notable hit to long-term productivity growth; think of disruptions to schools and universities, the loss of skilled and knowledgeable workers, and dramatic shifts in demand and supply that necessitate costly and time-consuming adjustments.[61] On the other hand, current research argues that recent technological advances and the unique nature of the GPC – which has raised prospects of recurrent pandemics – will lead to an acceleration in automation (see Leduc and Liu, 2020) and, in turn, boost long-term productivity. The net effect of these opposing factors on p-star is uncertain, and our model-based estimates of p-star reflect this via increased assessment of uncertainty around p-star, as shown in online appendix A8.

We believe that the rich features of our models helped position them, especially the Base model, to handle the pandemic data quite well.[62] In the interests of brevity, we refer the reader to online appendix A8 for the results comparing the pre-pandemic and pandemic periods, and results comparing estimates from the Base model to outside sources, including the CBO and Morley and Wong (2020).

7. Conclusion

This paper takes up the challenge of developing a large-scale UC model to jointly estimate the dynamics of inflation, nominal wages, labor productivity, the unemployment rate, real GDP, interest rates, and their respective survey expectations to back out estimates of the long-run counterparts of these variables. These long-run counterparts include potential output (gdp-star), the growth rate of potential output (g-star), the equilibrium levels of the unemployment rate (u-star), the real short-term interest rate (r-star), price inflation (pi-star), labor productivity growth (p-star), and nominal wage inflation (w-star). The structure of our UC model is guided by economic theory and past empirical research, which has highlighted strong evidence of changing macroeconomic relationships and allows for stochastic volatility in the shocks to cyclical components of a range of macroeconomic indicators. Accordingly, our model structure permits time-varying parameters and stochastic volatility in the main model equations.

An essential feature of our model structure is the explicit role of long-run survey forecasts in possibly informing the econometric estimation of the stars. Like other researchers, we find that adding information from surveys is an important element in yielding reasonable estimates of the stars. Incorporating a rich set of empirical features leads to a very flexible but heavily parameterized model. To feasibly estimate this model (and its variants), we use Bayesian techniques, specifically the efficient sampling techniques developed in Chan, Koop, and Potter (2013), and the precision sampler proposed in Chan and Jeliazkov (2009).

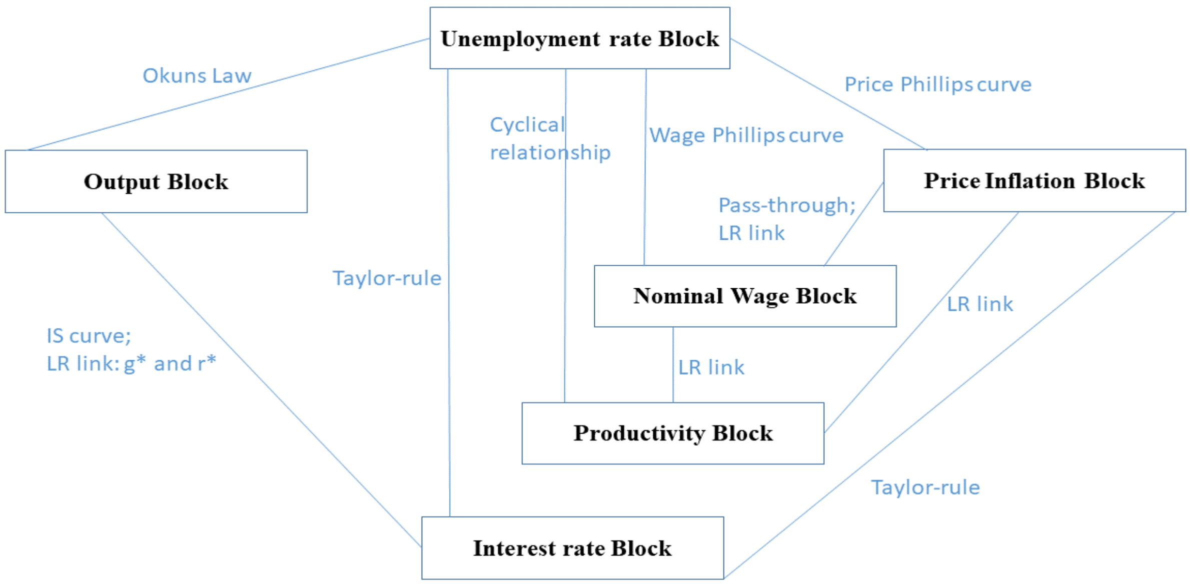
[61] Aucejo, French, Ugalde Araya, and Zafar (2020) document the negative impact of COVID-19 on students' academic performance.

[62] The rich features include: (1) modeling the changing economic relationships via the implementation of time-varying parameters; (2) allowing for the changing variance of the innovations to various equations (i.e., SV); (3) imposing bounds on some of the random walk processes; (4) joint modeling of the output gap and unemployment gap in particular; and (5) the use of survey forecasts;

We explore the empirical relevance of various features incorporated in our baseline model by estimating several variants of the baseline model. The Bayesian model comparison results provide strong support to model features informed by past research and confirm findings documented elsewhere. For instance, we find that allowing for SV in the model equations is very important. Similarly, we find economically and statistically significant evidence of a time-varying price Phillips curve, wage Phillips curve, the evolving cyclicality of labor productivity, a changing pass-through relationship between wages and prices, and evolving persistence in price inflation and wage inflation gaps. Given the richness of our model, we document an expansive set of empirical results that we hope will prove helpful for both applied and theoretical macroeconomists alike.

Our estimates of the stars generally echo the contours of stars documented elsewhere in the literature – estimated using smaller-scale UC models – but at times the estimates of the stars are different, and these differences can matter for policy. We show significant improvements in the precision of the stars' estimates by bringing in additional information from survey expectations. The resulting precision estimates are on a par with recent studies highlighting the improved precision of stars derived from their approaches. We also show that our baseline model held up well when including the COVID-19 pandemic data. The rich set of features endowed in our UC model helped handle the pandemic data without any difficulties. Lastly, we document the competitive real-time forecasting properties of both our main model and, separately, the estimates of the stars, if they were to be used as steady-state values in external models.

Fig. 1. Visual Overview of Interactions Between Blocks



Notes: The solid lines represent a contemporaneous relationship between the element(s) of the blocks. LR link denotes long-run relationship, i.e., link between stars.

Table 1: Description of Model Specifications

Panel (a): Main Models

Model	**Description**
Base	Full model as formulated in equations (6) through (39)
Base-NoSurv	Model as defined by baseline but excluding equations linking surveys to stars (i.e., excluding 9, 10, 15, 16, 26, 27, 38 and 39). Aim: assess the usefulness of survey data

Panel (b): Auxiliary Models

Base-W*RW	Base but replaces eq. (28) with an RW assumption for W^* as defined in eq. (28b). Aim: assess the support for the theoretical restriction $W^* = \pi^* + P^*$
Base-PT-Wages-to-Prices	Base but replaces eq. (21) with eq. (21b) . Aim: assess the usefulness of allowing for pass-through from wages to prices (for reference, Base allows for pass-through from prices to wages)
Base-NoPT	Base but replaces eq. (29) with eq. (29b). Aim: assess the usefulness of inclusion of wage inflation in the price equation AND the inclusion of price inflation in the wage equation.
Base-NoSurv-NoPT	Base w/o surveys and replaces eq. (29) with eq. (29b). Aim: usefulness of inclusion of wage inflation in the price equation AND the inclusion of price inflation in the wage equation in a Base-NoSurv.
Base-P*CycOutputGap	Base but replaces eq. (17) with eq. (17b). Aim: assess empirical link between output gap and productivity gap and whether data support inclusion of output gap instead of unemployment gap.
Base-NoR*Surv	Base but excludes survey expectations of r^*; i.e., removes eq. (38) and eq. (39). Aim: assess specifically the marginal value of survey expectations of r^*
Base-NoR*Surv-NoTRule	Base but excludes (1) survey expectations of r^*, i.e., removes eq. (38) and eq. (39); and (2) policy rule eq. (34). Aim: assess the marginal value of the policy rule equation to the model.
Base-R*RW	Base but replaces eq.(36) and eq.(37) with a RW assumption for r^* as in eq. (36b). Aim: assess the support for the theoretical link between r^* and g^*.
Base-NoSurv-R*RW	Base-NoSurv but replaces eq.(36) and eq.(37) with an RW for r^* as in eq. (36b). Aim: assess support for the link between r^* and g^* in a spec. w/o survey data.
Base-R*AR	Base but replaces eq. (37) with AR assumption for D. Aim: support of r^* defined as comb. of permanent and transitory components.
Base-NoSurv-R*AR	Base-NoSurv but replaces eq. (37) with AR assumption for D Aim: support of r^* defined as comb. of permanent and transitory components.
Base-NoSurv-R*TightPrior	Base specification but tighter prior for std. of the shock to process D. Aim: assess the impact on r^* estimate of using a tighter prior for D.
Base-NoSV	Base specification but with no SV in any of the measurement equations. Aim: assess the importance of stochastic volatility in shock variances.
Base-NoBoundU*	Base specification without the bounds on the U^* process in eq. (8). Aim: assess empirical support for imposing bounds on U^*.

Table 2: Parameter Estimates

Parameter	Parameter description	Posterior estimates					
		Base			Base-NoSurv		
		Mean	5%	95%	Mean	5%	95%
a^r	Coefficient on interest-rate gap	-0.069	-0.141	0.002	-0.054	-0.130	0.015
$\rho_1^g + \rho_2^g$	Persistence in output gap	0.738	0.678	0.797	0.732	0.669	0.794
ρ_1^u	Lag 1 coefficient on UR gap	1.283	1.238	1.327	1.277	1.232	1.322
ρ_2^u	Lag 2 coefficient on UR gap	-0.504	-0.542	-0.466	-0.503	-0.540	-0.466
$\rho_1^u + \rho_2^u$	Persistence in UR gap	0.778	0.736	0.821	0.774	0.733	0.814
ρ^p	Persistence in productivity gap	-0.005	-0.118	0.110	-0.001	-0.119	0.115
$m = \frac{\zeta}{4}$	Relationship between r* and g*	0.701	0.621	0.784	0.390	0.270	0.560
ρ^i	Persistence in interest-rate gap	0.879	0.838	0.917	0.860	0.816	0.904
λ^i	Interest rate sensitivity to UR gap	-0.229	-0.278	-0.180	-0.243	-0.293	-0.194
κ^i	Interest rate sensitivity to inflation	0.061	0.014	0.107	0.085	0.035	0.134
λ^g	Output gap response to UR gap	-0.457	-0.580	-0.339	-0.454	-0.575	-0.333
ϕ^u	UR gap response to Output gap	-0.108	-0.130	-0.086	-0.118	-0.140	-0.096
$\frac{(1-\rho_1^u-\rho_2^u)}{\phi_u}$	Implied Okun's Law	-2.054	-2.327	-1.803	-1.929	-2.179	-1.699
β^g	Link between g* and survey	0.929	0.774	1.082	—	—	—
β^u	Link between u* and survey	0.988	0.919	1.055	—	—	—
β^r	Link between r* and survey	1.018	0.923	1.117	—	—	—
β^π	Link between π* and survey	0.991	0.912	1.070	—	—	—
$\sigma^2_{\pi^*}$	Variance of the shocks to π^*	0.121^2	0.100^2	0.141^2	0.127^2	0.084^2	0.182^2
$\sigma^2_{p^*}$	Variance of the shocks to p^*	0.145^2	0.111^2	0.183^2	0.141^2	0.109^2	0.176^2
$\sigma^2_{u^*}$	Variance of the shocks to u^*	0.075^2	0.064^2	0.089^2	0.084^2	0.071^2	0.100^2
$\sigma^2_{gdp^*}$	Variance of the shocks to gdp^*	0.018^2	0.014^2	0.021^2	0.021^2	0.016^2	0.026^2
σ^2_d	Variance of the shocks to d	0.093^2	0.077^2	0.110^2	0.114^2	0.084^2	0.148^2
$\sigma^2_{w^*}$	Variance of the shocks to w^*	0.158^2	0.112^2	0.215^2	0.158^2	0.111^2	0.220^2
σ^2_{ogap}	Variance of the shocks to Output gap	0.725^2	0.672^2	0.781^2	0.723^2	0.669^2	0.780^2
σ^2_u	Variance of the shocks to UR gap	0.268^2	0.248^2	0.289^2	0.265^2	0.245^2	0.286^2
σ^2_{hp}	Var. of the Volatility – Productivity eq.	0.274^2	0.219^2	0.336^2	0.273^2	0.220^2	0.335^2
σ^2_h	Var. of the Volatility – Price Inf. eq.	0.297^2	0.237^2	0.364^2	0.298^2	0.238^2	0.365^2
σ^2_{hw}	Var. of the Volatility – Wage Inf. eq.	0.301^2	0.237^2	0.374^2	0.303^2	0.239^2	0.373^2
σ^2_{hi}	Var. of the Volatility – Interest rate eq.	0.377^2	0.293^2	0.469^2	0.369^2	0.287^2	0.454^2
$\sigma^2_{\lambda\pi}$	Var. of the shocks to TVP λ^π	0.041^2	0.032^2	0.052^2	0.041^2	0.032^2	0.051^2
$\sigma^2_{\lambda w}$	Var. of the shocks to TVP λ^w	0.041^2	0.032^2	0.051^2	0.040^2	0.032^2	0.050^2
$\sigma^2_{\lambda p}$	Var. of the shocks to TVP λ^p	0.044^2	0.034^2	0.057^2	0.045^2	0.034^2	0.057^2
$\sigma^2_{\kappa w}$	Var. of the shocks to TVP κ^w, PT	0.041^2	0.032^2	0.052^2	0.041^2	0.032^2	0.052^2
$\sigma^2_{\rho w}$	Var. of the shocks to TVP ρ^w	0.042^2	0.033^2	0.053^2	0.042^2	0.032^2	0.052^2
$\sigma^2_{\rho\pi}$	Var. of the shocks to TVP ρ^π	0.048^2	0.037^2	0.062^2	0.047^2	0.036^2	0.060^2

Table 3: Bayesian Model Comparison: Main Models and Selected Variants

	Base	**Base-NoSurv**	Base-R*RW	Base-NoSurv-R*RW
MDD of Inflation	-366.6	-368.6	-366.8	-370.1
MDD of Productivity	-606.9	-607.1	-607.5	-607.4
MDD of Nominal Wage	-277.5	-274.3	-278.0	-274.9
MDD of Unemployment	-24.6	-21.7	-24.7	-21.8
MDD of Interest rate	-216.4	-222.0	-214.0	-215.6
MDD of GDP	-279.7	-279.1	-279.8	-279.5
MDD	-1771.7	-1772.8	-1770.8	-1769.3

Table 4: Model Comparison: Variants Focused on Unemployment Rate (UR)

	Base	Base-NoSurv	Base-NoBoundU*	Bivariate	Bivariate+Surv	CKP-Adj
MDD of UR	-24.6	-21.7	-25.0	-56.5	-46.5	-61.9

Table 5: Model Comparison: Variants Focused on GDP

	Base	Base-NoSurv	Univariate	Bivariate	Bivariate+Surv
MDD of GDP	-279.7	-279.1	-296.5	-280.4	-281.9

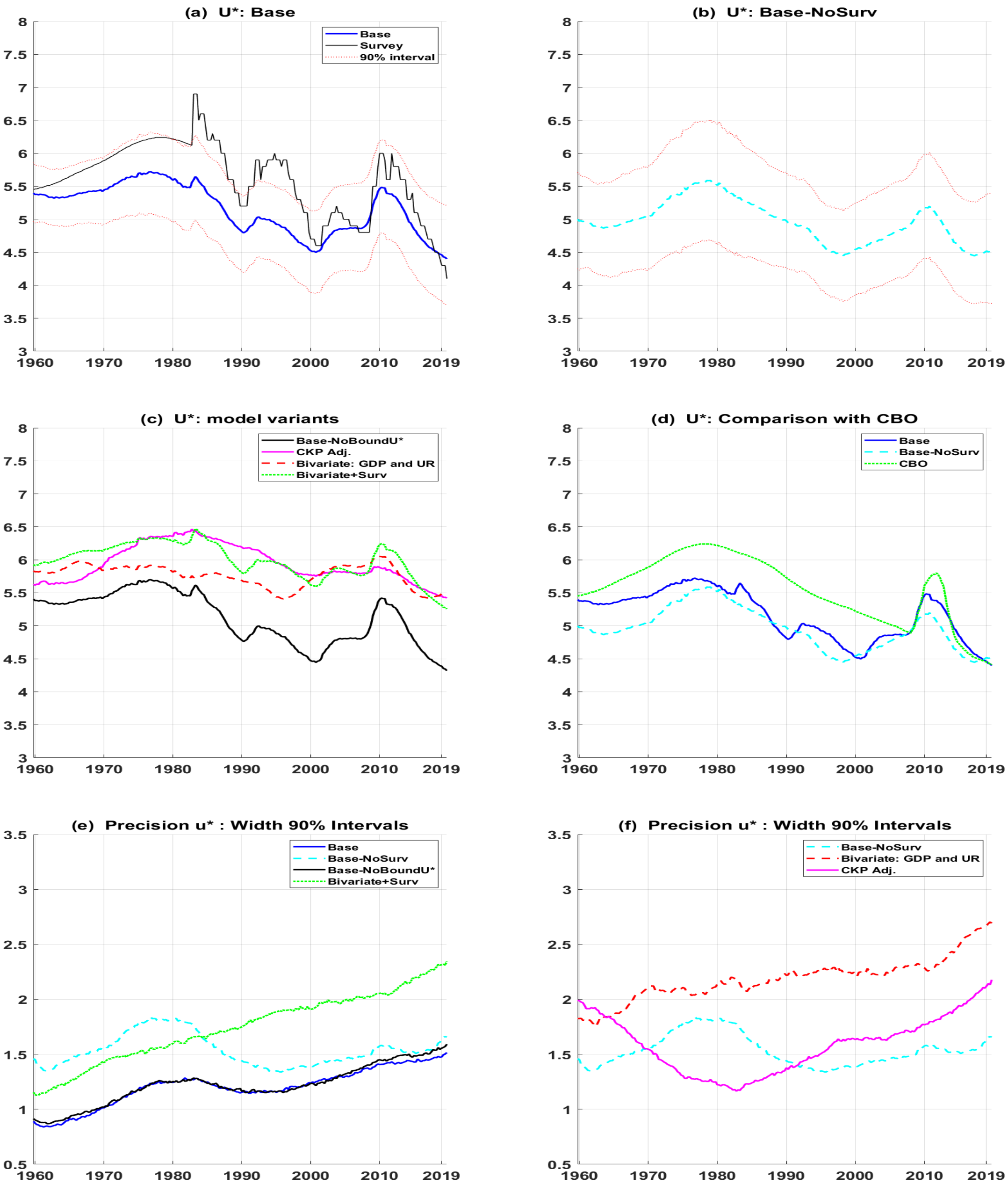
Table 6: Model Comparison: Variants Focused on Labor Productivity

	Base	Base-NoSurv	Base-W*RW	Base-P*CycOutputGap	Base-NoSV
MDD of Productivity	-606.9	-607.1	-603.5	-608.1	-633.4
MDD of model	-1771.7	-1772.8	-1790.8	-1773.8	-2024.3

Table 7: Model Comparison: Variants Focused on Nominal Wages

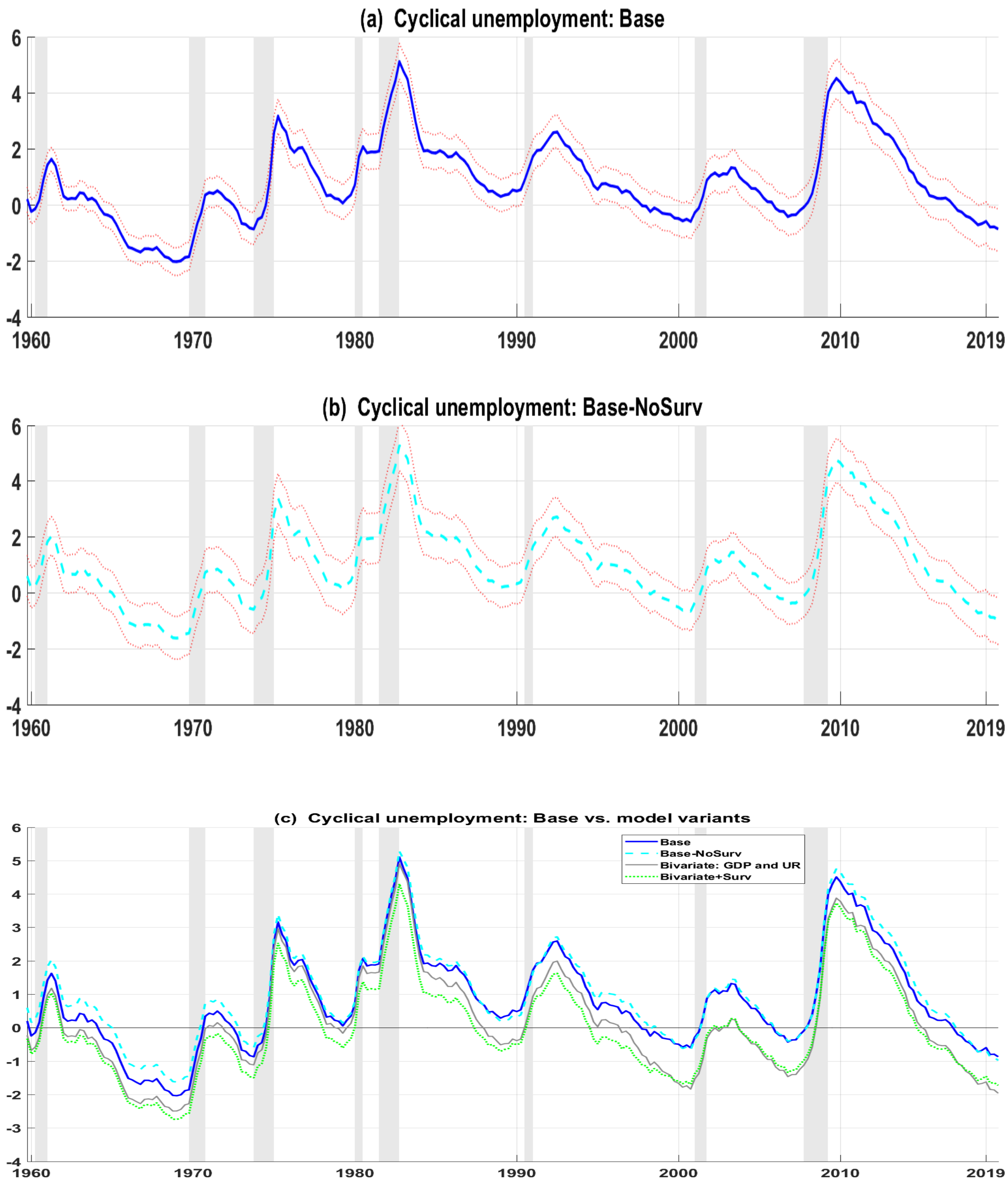
	Base	Base-NoSurv	Base-W*RW	Base-NoPT	Base-NoSV
MDD of Nominal wages	-277.5	-274.3	-277.7	-281.5	-344.3
MDD of model	-1771.7	-1772.8	-1790.8	-1776.3	-2024.3

Fig. 2. Full Sample Estimates for Unemployment Rate Block



Notes: The posterior estimates are based on the full sample (from 1959Q4 through 2019Q4). In panel (e), CKP Adj. refers to the bivariate model of inflation and the unemployment rate as in CKP but with no bounds on pi-star.

Fig. 3. Full Sample Estimates for Cyclical Unemployment



Notes: The posterior estimates are based on the full sample (from 1959Q4 through 2019Q4). The shaded areas represent NBER recession dates.

Fig. 4. Full Sample Estimates for Output Block

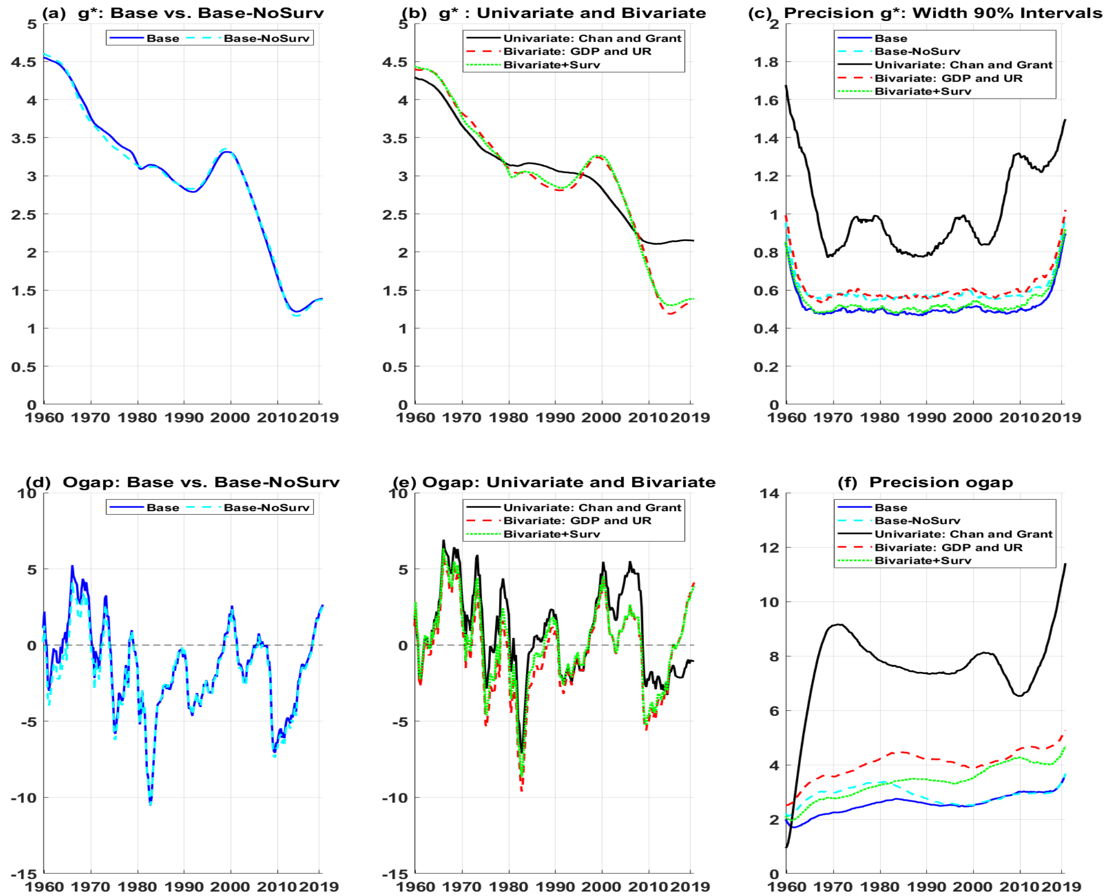


Fig. 5. Model-based Output Gap vs. CBO

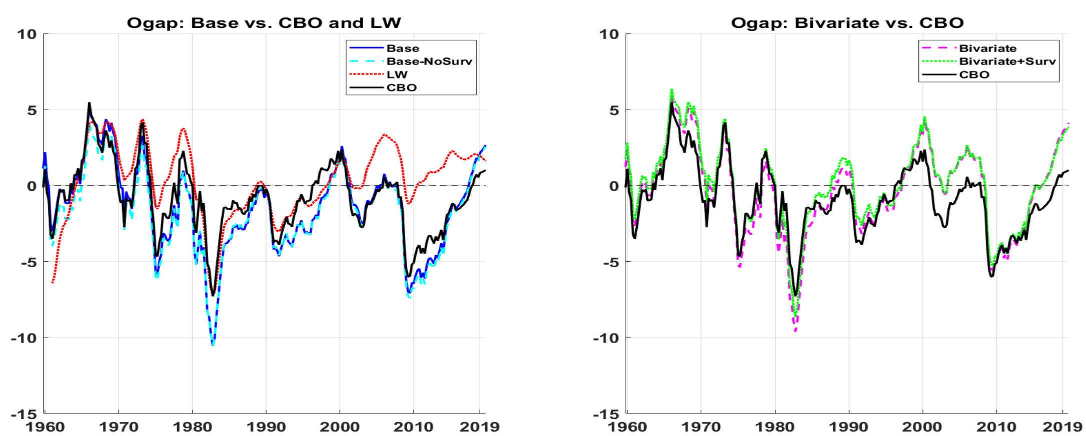
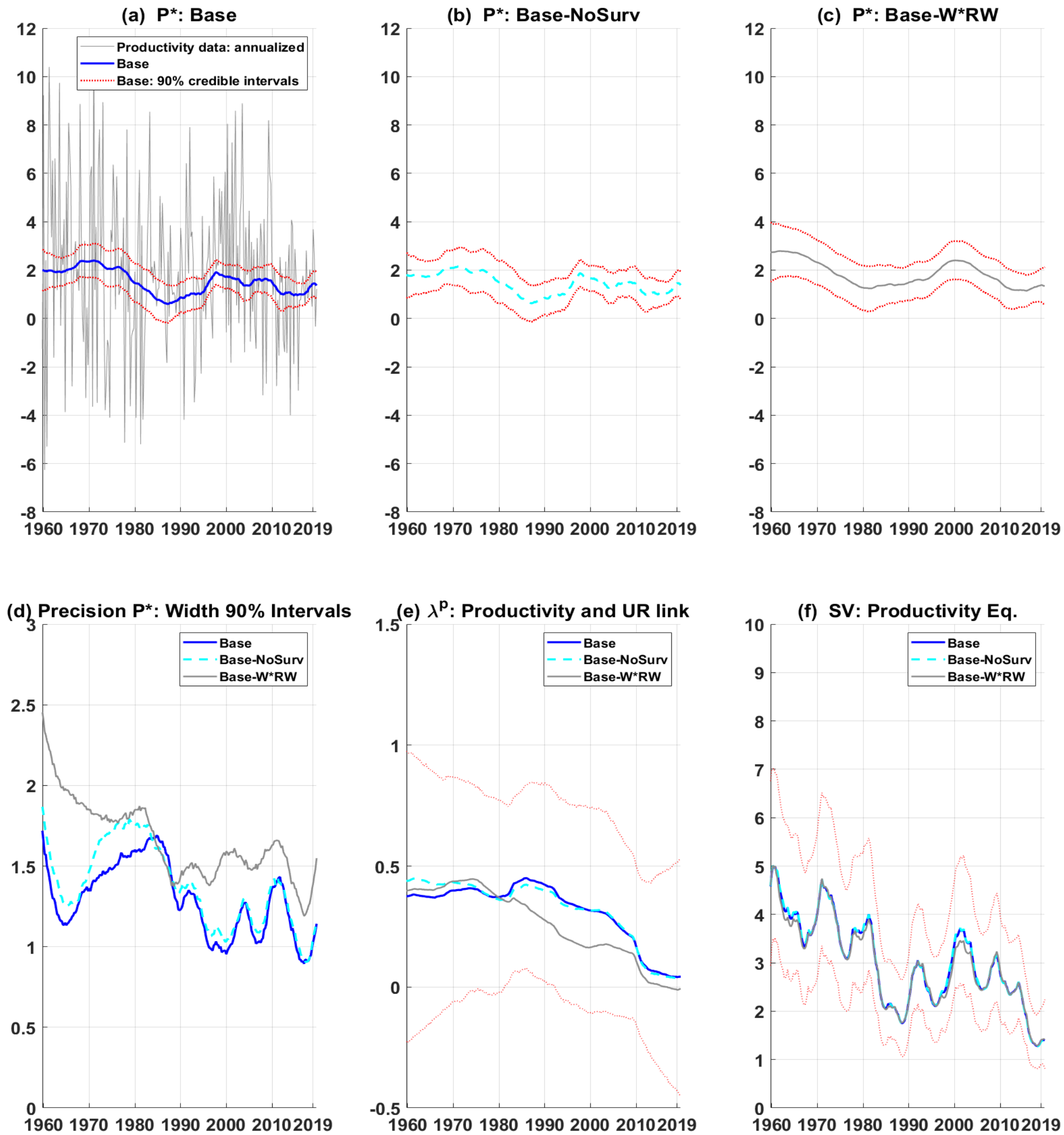
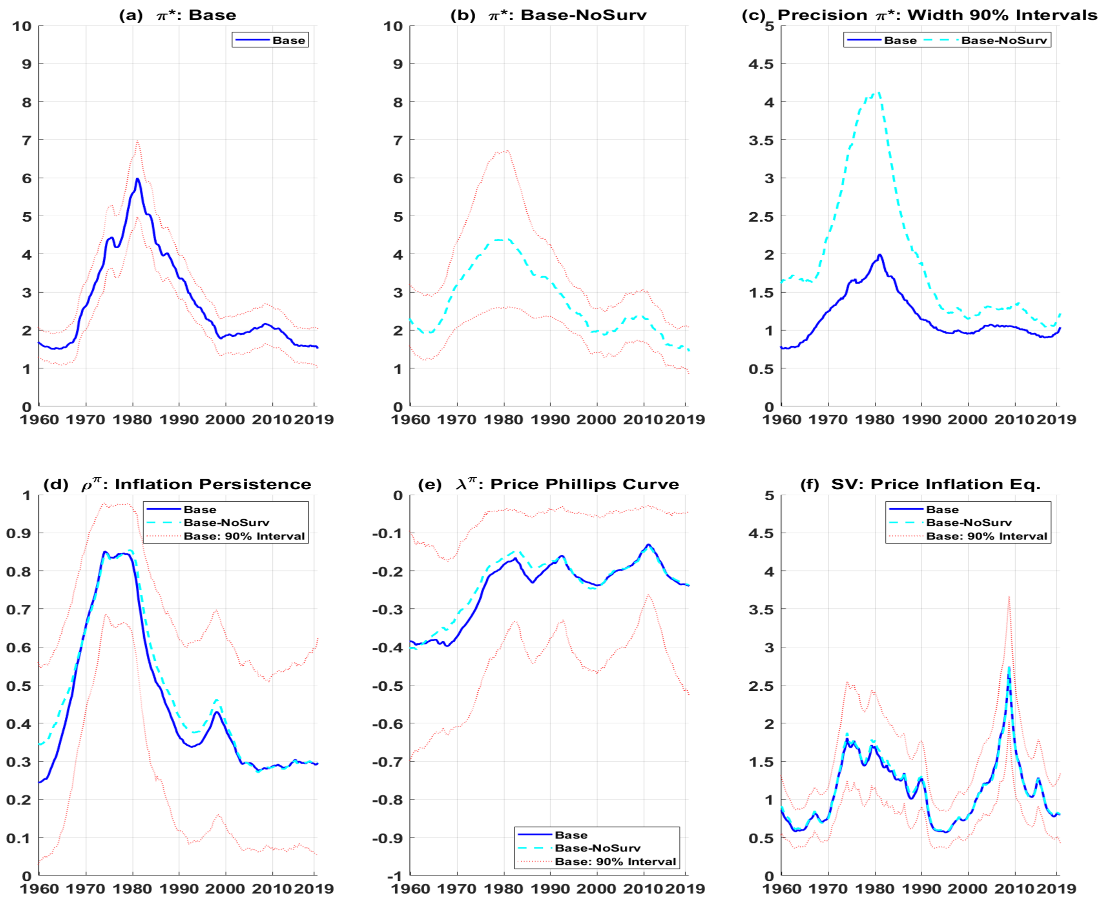


Fig. 6. Full Sample Estimates for Productivity Block



Notes: The posterior estimates are based on the full sample (from 1959Q4 through 2019Q4).

Fig. 7. Full Sample Estimates for Price Inflation Block

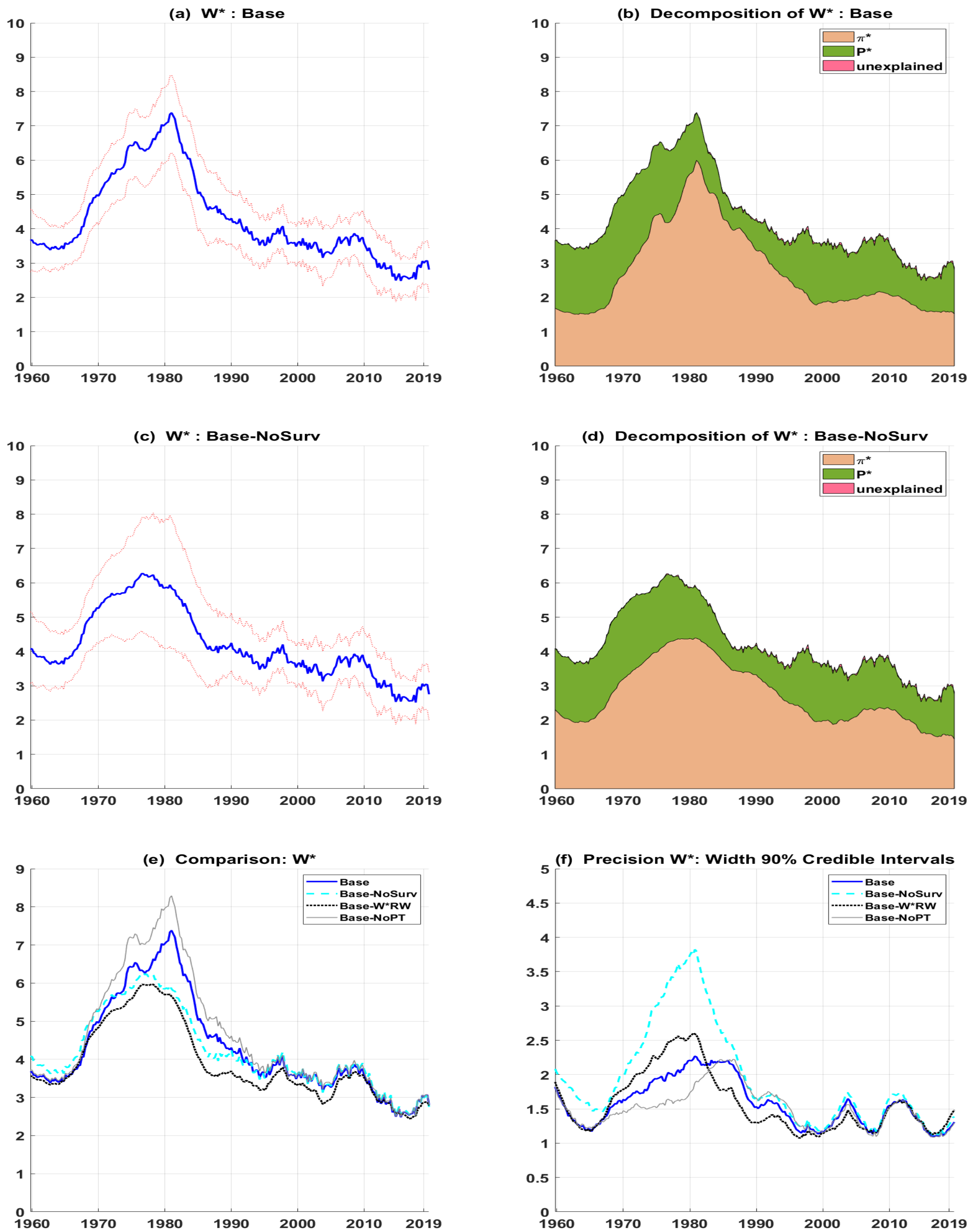


Notes: The posterior estimates are based on the full sample (from 1959Q4 through 2019Q4).

Table 8: Model Variants for Price Inflation

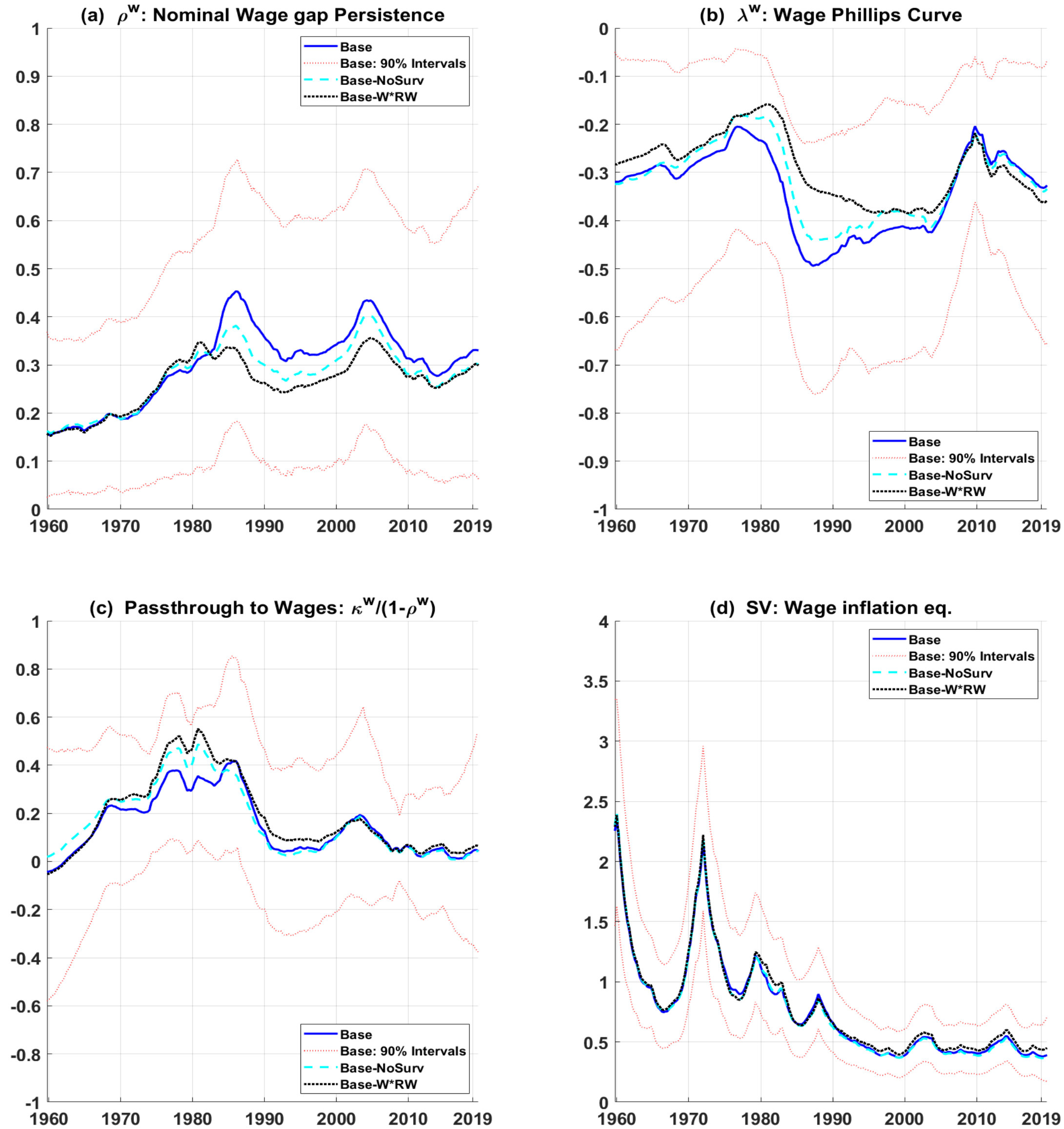
Model variant	MDD of Price inflation	MDD Model
Base	-366.6	-1771.7
Base-NoSurv	-368.6	-1772.8
Base-W*RW	-367.0	-1790.8
Base-PT-Wages-to-Prices	-366.7	-1771.4
Base-NoPT	-365.2	-1776.3
Base-NoSV	-413.9	-2024.3

Fig. 8. Full Sample Estimates for Nominal Wage Block



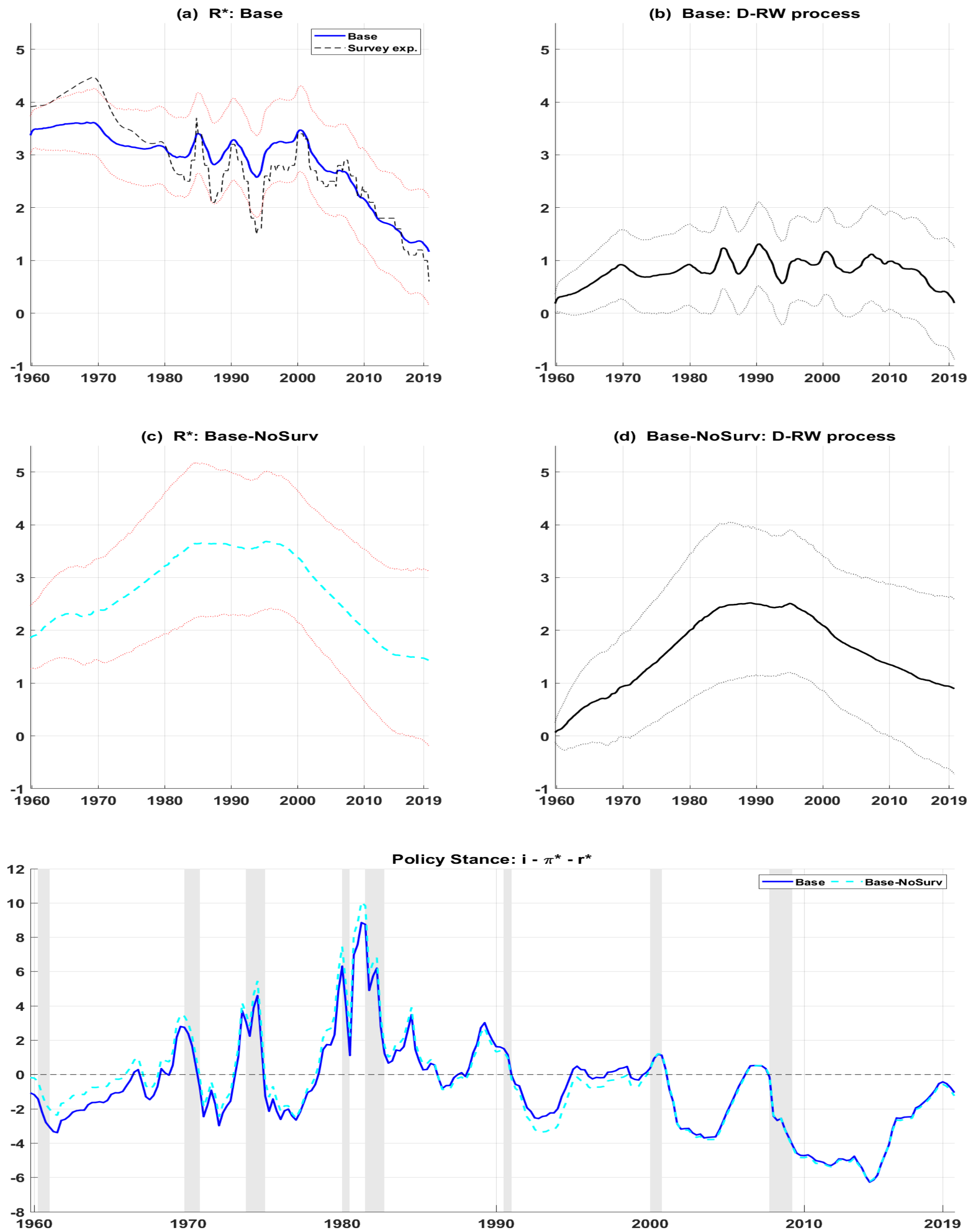
Notes: The posterior estimates are based on the full sample (from 1959Q4 through 2019Q4).

Fig. 9. Wage Persistence, Wage Phillips Curve, Pass-through from Prices, and SV



Notes: The posterior estimates are based on the full sample (from 1959Q4 through 2019Q4).

Fig. 10. Full Sample Estimates for Interest Rate Block



Notes: The posterior estimates are based on the full sample (from 1959Q4 through 2019Q4). In the top panel, the plot labeled "survey exp." is an implied estimate: inferred from the Blue Chip survey long-run estimates of the GDP deflator and short-term interest rates (3-month Treasury bill) using the long-run Fisher equation. Specifically, the long-run forecast of 3-month Treasury bill less the long-run forecast of GDP deflator. To this differential, we add +0.3 to reflect the average differential between the federal funds rate and the 3-month Treasury bill.

Fig. 11. More Estimates for Interest Rate Block

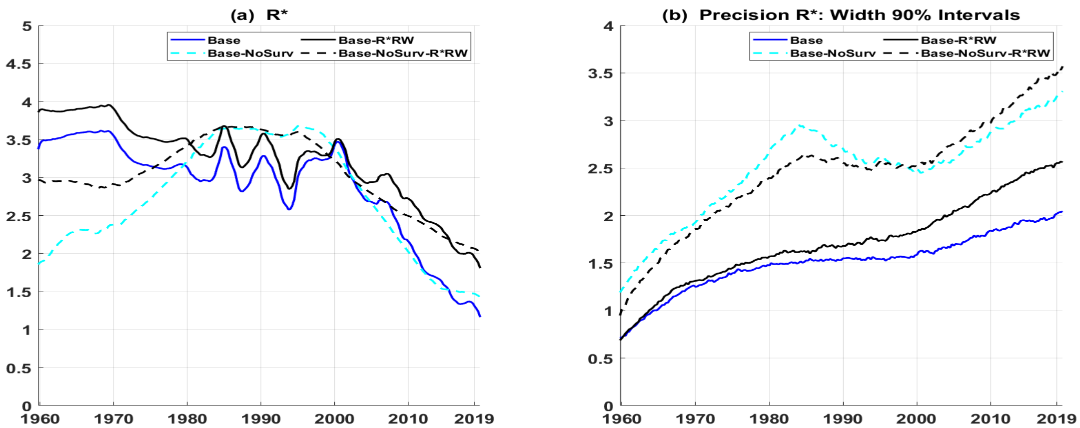
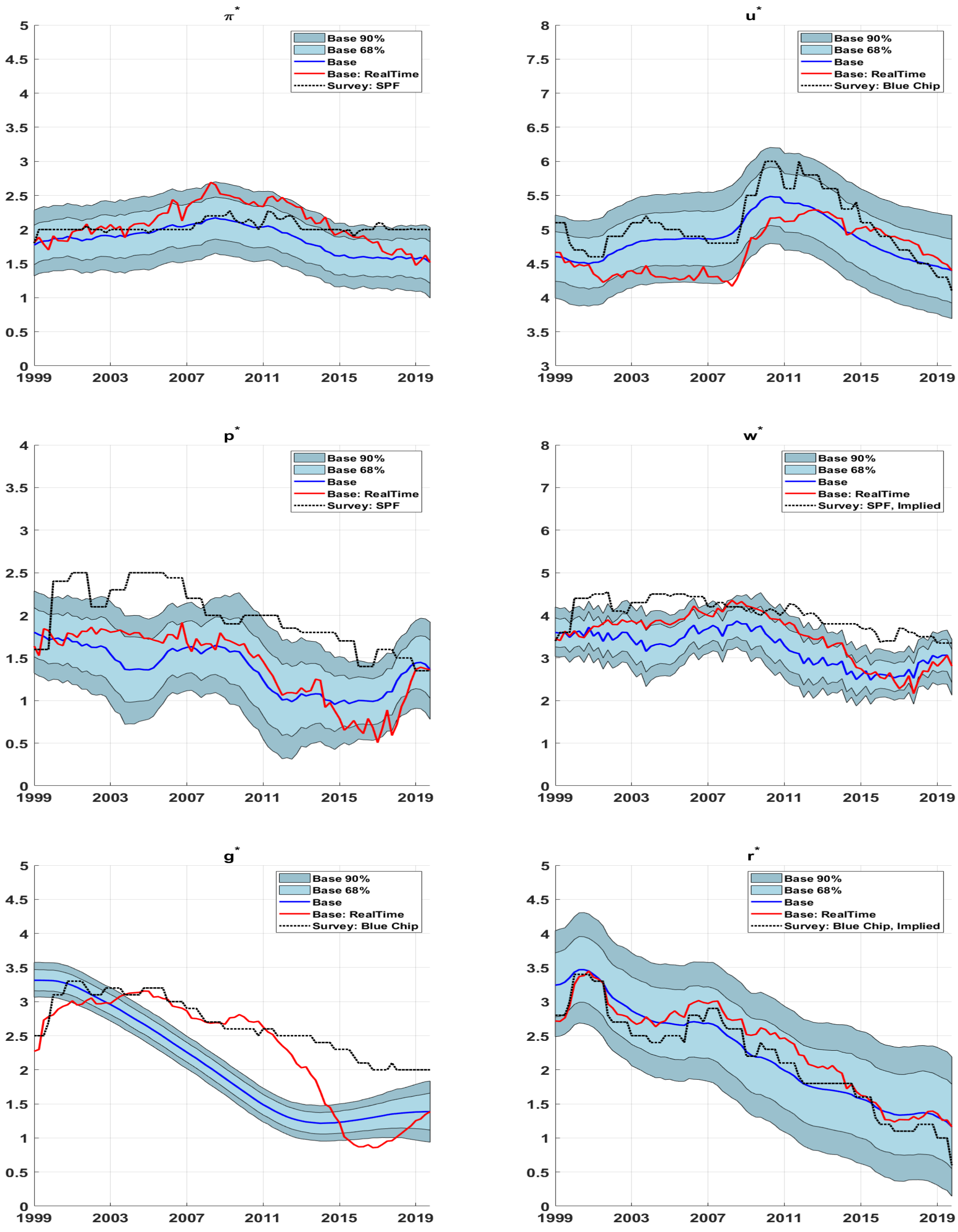


Table 9: Model variants for Interest Rate

Model variant	MDD of Interest rate	MDD Model
Base	-216.4	-1771.7
Base-NoSurv	-222.0	-1772.8
Base-R*RW	-214.0	-1770.8
Base-NoSurv-R*RW	-215.6	-1769.3
Base-NoSurv-R*TightPrior	-232.0	-1773.6

Fig. 12. Real-time Recursive Estimates of Stars: Base Model



Notes: The plots labeled Base are posterior estimates reflecting information in the full sample (from 1959Q4 through 2019Q4). The plots labeled Base:RealTime are posterior estimates reflecting information available at a given point in time (i.e., truly real time).

Table 10: Real-Time Forecasting Accuracy: Base vs. Base-NoSurv

Panel A: Point Forecast Accuracy (Recursive evaluation: 1999.Q1-2019.Q4)

Relative RMSE: RMSE **Base** / RMSE **BaseNoSurv**

	h=4Q	h=8Q	h=12Q	h=16Q	h=20Q
Real GDP	1.07*	1.03	1.01	1.00	0.99
PCE Inflation	0.96	0.94*	0.91*	0.89	0.88*
Productivity	0.98*	0.98	0.97	0.98*	0.97
Nominal Wage (AHE)	0.98	1.01	1.05	1.05	1.02
Unemployment Rate	1.07*	1.05*	1.03*	1.02*	1.00
Shadow FFR	0.98	1.00	1.00	0.99	0.98

Panel B: Density Forecast Accuracy (Recursive evaluation: 1999.Q1-2019.Q4)

Relative Log Predictive Score (LPS): LPS Base - LPS BaseNoSurv

	h=4Q	h=8Q	h=12Q	h=16Q	h=20Q
Real GDP	-0.003*	-0.002*	-0.002*	-0.001*	-0.001*
PCE Inflation	0.018*	0.020*	0.025*	0.027*	0.026*
Productivity	0.001	0.002*	0.002*	0.002*	0.003*
Nominal Wage (AHE)	0.017*	0.016*	0.018*	0.018*	0.019*
Unemployment Rate	-0.003	0.005	0.009*	0.012*	0.013*
Shadow FFR	0.020*	0.007*	0.003	0.000	0.000

Notes: The top panel compares the point forecast accuracy of the Base model with the Base-NoSurv model. Numbers less than 1 indicate that the Base model is more accurate on average. The bottom panel reports the corresponding density forecast accuracy performance. A positive value (for the relative mean predictive log score) suggests that the Base model is on average more accurate. The log predictive scores are computed using parametric normal approximation. The table reports statistical significance based on the likelihood-ratio test of Amisano and Giacomini (2007) for the density forecast accuracy, and based on the Diebold-Mariano and West test (with the lag $h-1$ truncation parameter of the HAC variance estimator) for the point forecast accuracy. The test statistics for the likelihood-ratio test use a two-sided t-test. In the case of the Diebold-Mariano and West test, the test statistics use two-sided standard normal critical values for horizons less than or equal to 8 quarters, and two-sided t-statistics for horizons greater than 8 quarters. *up to 10% significance level.

References

Amisano, Gianni and Raffaella Giacomini (2007). "Comparing density forecasts via weighted likelihood ratio tests." *Journal of Business & Economic Statistics*, 25(2), pp. 177–190. doi:10.1198/073500106000000332.

Antolin-Diaz, Juan, Thomas Drechsel, and Ivan Petrella (2017). "Tracking the slowdown in long-run GDP growth." *The Review of Economics and Statistics*, 99(2), pp. 343–356. doi:10.1162/REST_a_00646.

Ashley, Richard and Randal J. Verbrugge (2008). "Frequency dependence in regression model coefficients: An alternative approach for modeling nonlinear dynamic relationships in time series." *Econometric Reviews*, 28(1-3), pp. 4–20. doi:10.1080/07474930802387753.

Ashley, Richard and Randal J. Verbrugge (2020). "Finding a stable Phillips curve relationship: A persistence-dependent regression model." Working Paper 19-09R, Federal Reserve Bank of Cleveland. doi:10.26509/frbc-wp-201909r.

Aucejo, Esteban M., Jacob French, Maria Paola Ugalde Araya, and Basit Zafar (2020). "The impact of COVID-19 on student experiences and expectations: Evidence from a survey." *Journal of Public Economics*, 191, p. 104,271. doi:10.1016/j.jpubeco.2020.104271.

Banbura, Marta and Elena Bobeica (2020). "Does the Phillips curve help to forecast Euro area inflation?" Working Paper 2471, European Central Bank. doi:10.2866/61538.

Barbarino, Alessandro, Travis J. Berge, Han Chen, and Andrea Stella (2020). "Which output gap estimates are stable in real time and why?" Technical Report 2020-101, Board of Governors of the Federal Reserve System. doi:10.17016/FEDS.2020.102.

Barnichon, Regis and Geert Mesters (2018). "On the demographic adjustment of unemployment." *The Review of Economics and Statistics*, 100(2), pp. 219–231. doi:10.1162/REST_a_00688.

Basistha, Arabinda and Charles R. Nelson (2007). "New measures of the output gap based on the forward-looking new Keynesian Phillips curve." *Journal of Monetary Economics*, 54(2), pp. 498–511. doi:10.1016/j.jmoneco.2005.07.021.

Bauer, Michael D. and Glenn D. Rudebusch (2020). "Interest rates under falling stars." *American Economic Review*, 110(5), pp. 1316–1354. doi:10.1257/aer.20171822.

Benati, Luca (2007). "Drift and breaks in labor productivity." *Journal of Economic Dynamics and Control*, 31(8), pp. 2847–2877. doi:10.1016/j.jedc.2006.11.004.

Berger, Tino, Gerdie Everaert, and Hauke Vierke (2016). "Testing for time variation in an unobserved components model for the U.S. economy." *Journal of Economic Dynamics and Control*, 69, pp. 179–208. doi:10.1016/j.jedc.2016.05.017.

Beyer, Robert C. M. and Volker Wieland (2019). "Instability, imprecision and inconsistent use of equilibrium real interest rate estimates." *Journal of International Money and Finance*, 94, pp. 1–14. doi:10.1016/j.jimonfin.2019.01.005.

Bobeica, Elena, Matteo Ciccarelli, and Isabel Vansteenkiste (2019). "The link between labor cost and price inflation in the Euro area." Working Paper 2235, European Central Bank.

doi:10.2866/894414.

Brand, Claus and Falk Mazelis (2019). “Taylor-rule consistent estimates of the natural rate of interest.” Working Paper 2257, European Central Bank. doi:10.2866/44361.

Bullard, James B. (2018). “R-star wars: the phantom menace.” *Business Economics*, 53(2), pp. 60–65. doi:10.1057/s11369-018-0077-1. URL https://doi.org/10.1057%2Fs11369-018-0077-1.

Carriero, Andrea, Todd E. Clark, and Massimiliano Marcellino (2019). “Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors.” *Journal of Econometrics*, 212(1), pp. 137–154. doi:10.1016/j.jeconom.2019.04.024.

Carriero, Andrea, Todd E. Clark, Massimiliano Marcellino, and Elmar Mertens (2021). “Addressing COVID-19 outliers in BVARs with stochastic volatility.” Working Paper 21-02, Federal Reserve Bank of Cleveland. doi:10.26509/frbc-wp-202102.

Chan, Joshua C. C. (2013). “Moving average stochastic volatility models with application to inflation forecast.” *Journal of Econometrics*, 176(2), pp. 162–172. doi:10.1016/j.jeconom.2013.05.003.

Chan, Joshua C. C. (2017). “The stochastic volatility in mean model with time-varying parameters: An application to inflation modeling.” *Journal of Business & Economic Statistics*, 35(1), pp. 17–28. doi:10.1080/07350015.2015.1052459.

Chan, Joshua C. C. (2019). “Large hybrid time-varying parameter VARs.” Technical Report 2019-77, Centre for Applied Macroeconomic Analysis, Crawford School of Public Policy, The Australian National University.

Chan, Joshua C. C., Todd E. Clark, and Gary Koop (2018). “A new model of inflation, trend inflation, and long-run inflation expectations.” *Journal of Money, Credit and Banking*, 50(1), pp. 5–53. doi:10.1111/jmcb.12452.

Chan, Joshua C. C. and Eric Eisenstat (2018a). “Bayesian model comparison for time-varying parameter VARs with stochastic volatility.” *Journal of Applied Econometrics*, 33(4), pp. 509–532. doi:10.1002/jae.2617.

Chan, Joshua C. C. and Eric Eisenstat (2018b). “Comparing hybrid time-varying parameter VARs.” *Economics Letters*, 171, pp. 1–5. doi:10.1016/j.econlet.2018.06.031.

Chan, Joshua C. C. and Ivan Jeliazkov (2009). “Efficient simulation and integrated likelihood estimation in state space models.” *International Journal of Mathematical Modelling and Numerical Optimisation*, 1(1/2), p. 101. doi:10.1504/IJMMNO.2009.030090.

Chan, Joshua C. C., Gary Koop, and Simon M. Potter (2013). “A new model of trend inflation.” *Journal of Business & Economic Statistics*, 31(1), pp. 94–106. doi:10.1080/07350015.2012.741549.

Chan, Joshua C. C., Gary Koop, and Simon M. Potter (2016). “A bounded model of time variation in trend inflation, Nairu and the Phillips curve.” *Journal of Applied Econometrics*, 31(3), pp. 551–565. doi:10.1002/jae.2442.

Clark, Peter K. (1987). “The cyclical component of U.S. economic activity.” *The Quarterly Journal of Economics*, 102(4), p. 797. doi:10.2307/1884282.

Clark, Todd E. and Taeyoung Doh (2014). "Evaluating alternative models of trend inflation." *International Journal of Forecasting*, 30(3), pp. 426–448. doi:10.1016/j.ijforecast.2013.11.005.

Clark, Todd E. and Sharon Kozicki (2005). "Estimating equilibrium real interest rates in real time." *The North American Journal of Economics and Finance*, 16(3), pp. 395–413. doi:10.1016/j.najef.2005.04.002.

Cogley, Timothy, Giorgio E. Primiceri, and Thomas J. Sargent (2010). "Inflation-gap persistence in the US." *American Economic Journal: Macroeconomics*, 2(1), pp. 43–69. doi:10.1257/mac.2.1.43.

Cogley, Timothy and Argia M. Sbordone (2008). "Trend inflation, indexation, and inflation persistence in the new keynesian phillips curve." *American Economic Review*, 98(5), pp. 2101–2126. doi:10.1257/aer.98.5.2101. URL https://doi.org/10.1257%2Faer.98.5.2101.

Coibion, Olivier, Yuriy Gorodnichenko, and Mauricio Ulate (2018). "The cyclical sensitivity in estimates of potential output." *Brookings Papers on Economic Activity*, 2018(2), pp. 343–441. doi:10.1353/eca.2018.0020.

Cross, Rod, Julia Darby, and Jonathan Ireland (2008). "Uncertainties surrounding natural rate estimates in the G7." In Reinhard Neck, Christian Richter, and Peter Mooslechner, editors, *Quantitative Economic Policy*, volume 20, pp. 337–363. Springer Berlin Heidelberg, Berlin, Heidelberg. doi:10.1007/978-3-540-74684-3_16.

Crump, Richard, Stefano Eusepi, Marc Giannoni, and Ayşegül Şahin (2019). "A unified approach to measuring u*." Working Paper 25930, National Bureau of Economic Research. doi:10.3386/w25930.

Daly, Mary C. and Bart Hobijn (2014). "Downward nominal wage rigidities bend the Phillips curve." *Journal of Money, Credit and Banking*, 46(S2), pp. 51–93. doi:10.1111/jmcb.12152.

Daly, Mary C., Bart Hobijn, Ayşegül Şahin, and Robert G. Valletta (2012). "A search and matching approach to labor markets: Did the natural rate of unemployment rise?" *Journal of Economic Perspectives*, 26(3), pp. 3–26. doi:10.1257/jep.26.3.3.

Davis, Steven J., R. Jason Faberman, and John C. Haltiwanger (2013). "The establishment-level behavior of vacancies and hiring." *The Quarterly Journal of Economics*, 128(2), pp. 581–622. doi:10.1093/qje/qjt002.

Del Negro, Marco, Domenico Giannone, Marc P. Giannoni, and Andrea Tambalotti (2017). "Safety, liquidity, and the natural rate of interest." *Brookings Papers on Economic Activity*, 2017(1), pp. 235–316. doi:10.1353/eca.2017.0003.

Del Negro, Marco, Marc P. Giannoni, and Frank Schorfheide (2015). "Inflation in the Great Recession and new Keynesian models." *American Economic Journal: Macroeconomics*, 7(1), pp. 168–196. doi:10.1257/mac.20140097.

Del Negro, Marco, Michele Lenza, Giorgio Primiceri, and Andrea Tambalotti (2020). "What's up with the Phillips curve?" Working Paper 27003, National Bureau of Economic Research. doi:10.3386/w27003.

Edge, Rochelle M., Thomas Laubach, and John C. Williams (2007). "Learning and shifts in long-run productivity growth." *Journal of Monetary Economics*, 54(8), pp. 2421–2438. doi:10.1016/j.jmoneco.2007.01.003.

Eggertsson, Gauti B., Neil R. Mehrotra, Sanjay R. Singh, and Lawrence H. Summers (2016). "A contagious malady? Open economy dimensions of secular stagnation." *IMF Economic Review*, 64(4), pp. 581–634. doi:10.1057/s41308-016-0019-8.

Faust, Jon and Jonathan H. Wright (2013). "Forecasting inflation." In *Handbook of Economic Forecasting*, volume 2, pp. 2–56. Elsevier. doi:10.1016/B978-0-444-53683-9.00001-3.

Fernald, John G. (2007). "Trend breaks, long-run restrictions, and contractionary technology improvements." *Journal of Monetary Economics*, 54(8), pp. 2467–2485. doi:10.1016/j.jmoneco.2007.06.031.

Fernald, John G. and J. Christina Wang (2016). "Why has the cyclicality of productivity changed? What does it mean?" *Annual Review of Economics*, 8(1), pp. 465–496. doi:10.1146/annurev-economics-080315-015018.

Fernández-Villaverde, Jesús and Juan Rubio-Ramírez (2010). "Macroeconomics and volatility: Data, models, and estimation." Working Paper 16618, National Bureau of Economic Research. doi:10.3386/w16618.

Fleischman, Charles A. and John M. Roberts (2011). "From many series, one cycle: Improved estimates of the business cycle from a multivariate unobserved components model." Technical Report 2011-46, Board of Governors of the Federal Reserve System.

Galí, Jordi (2011). "The return of the wage Phillips curve." *Journal of the European Economic Association*, 9(3), pp. 436–461. doi:10.1111/j.1542-4774.2011.01023.x.

Galí, Jordi and Luca Gambetti (2019). "Has the U.S. wage Phillips curve flattened? A semi-structural exploration." Working Paper 25476, National Bureau of Economic Research. doi:10.3386/w25476.

Galí, Jordi and Thijs van Rens (2021). "The vanishing procyclicality of labour productivity." *The Economic Journal*, 131(633), pp. 302–326. doi:10.1093/ej/ueaa065.

Garcia, Juan Angel and Aubrey Poon (2018). "Trend inflation and inflation compensation." Working Paper 154, International Monetary Fund. doi:10.5089/9781484362402.001.

González-Astudillo, Manuel and Jean-Philippe Laforte (2020). "Estimates of R* consistent with a supply-side structure and a monetary policy rule for the U.S. economy." Technical Report 2020-085, Board of Governors of the Federal Reserve System. doi:10.17016/FEDS.2020.085.

Grant, Angelia L. and Joshua C. C. Chan (2017a). "A Bayesian model comparison for trend-cycle decompositions of output." *Journal of Money, Credit and Banking*, 49(2-3), pp. 525–552. doi:10.1111/jmcb.12388.

Grant, Angelia L. and Joshua C. C. Chan (2017b). "Reconciling output gaps: Unobserved components model and Hodrick–Prescott filter." *Journal of Economic Dynamics and Control*, 75, pp. 114–121. doi:10.1016/j.jedc.2016.12.004.

Hamilton, James D., Ethan S. Harris, Jan Hatzius, and Kenneth D. West (2016). "The equilibrium real funds rate: Past, present, and future." *IMF Economic Review*, 64(4), pp. 660–707.

doi:10.1057/s41308-016-0015-z.

Harvey, Andrew C. (1985). "Trends and cycles in macroeconomic time series." *Journal of Business & Economic Statistics*, 3(3), pp. 216–227. doi:10.2307/1391592.

Hasenzagl, Thomas, Filippo Pellegrino, Lucrezia Reichlin, and Giovanni Ricco (2020). "A model of the Fed's view on inflation." *The Review of Economics and Statistics*, pp. 1–45. doi:10.1162/rest_a_00974.

Hodrick, Robert J. and Edward C. Prescott (1997). "Postwar U.S. business cycles: An empirical investigation." *Journal of Money, Credit and Banking*, 29(1), p. 1. doi:10.2307/2953682.

Holston, Kathryn, Thomas Laubach, and John C. Williams (2017). "Measuring the natural rate of interest: International trends and determinants." *Journal of International Economics*, 108, pp. S59–S75. doi:10.1016/j.jinteco.2017.01.004.

Jacobs, Jan P.A.M. and Simon van Norden (2016). "Why are initial estimates of productivity growth so unreliable?" *Journal of Macroeconomics*, 47, pp. 200–213. doi:10.1016/j.jmacro.2015.11.004.

Johannsen, Benjamin K. and Elmar Mertens (2021). "A time-series model of interest rates with the effective lower bound." *Journal of Money, Credit and Banking*, 53(5), pp. 1005–1046. doi:10.1111/jmcb.12771.

Kahn, James A. and Robert W. Rich (2007). "Tracking the new economy: Using growth theory to detect changes in trend productivity." *Journal of Monetary Economics*, 54(6), pp. 1670–1701. doi:10.1016/j.jmoneco.2006.07.008.

Kamber, Güneş, James Morley, and Benjamin Wong (2018). "Intuitive and reliable estimates of the output gap from a Beveridge-Nelson filter." *The Review of Economics and Statistics*, 100(3), pp. 550–566. doi:10.1162/rest_a_00691.

Kiley, Michael T. (2020). "What can the data tell us about the equilibrium real interest rate?" *International Journal of Central Banking*, 16(3), pp. 181–209.

Knotek II, Edward S. (2007). "How useful is Okun's law?" *Economic Review*, 92(Q IV), pp. 73–103.

Knotek II, Edward S. (2015). "Difficulties forecasting wage growth." *Economic Trends*. URL https://www.clevelandfed.org/newsroom-and-events/publications/economic-trends/2015-economic-trends/et-20150508-difficulties-forecasting-wage-growth.aspx.

Knotek II, Edward S. and Saeed Zaman (2014). "On the relationships between wages, prices, and economic activity." *Economic Commentary (Federal Reserve Bank of Cleveland)*, pp. 1–6. doi:10.26509/frbc-ec-201414.

Koop, Gary and Dimitris Korobilis (2010). "Bayesian multivariate time series methods for empirical macroeconomics." *Foundations and Trends(R) in Econometrics*, 3(4), pp. 267—-358. doi:10.1561/0800000013.

Kozicki, Sharon and P. A. Tinsley (2012). "Effective use of survey information in estimating the evolution of expected inflation." *Journal of Money, Credit and Banking*, 44(1), pp. 145–169. doi:10.1111/j.1538-4616.2011.00471.x.

Kuttner, Kenneth N. (1994). "Estimating potential output as a latent variable." *Journal of Business & Economic Statistics*, 12(3), pp. 361–368. doi:10.1080/07350015.1994.10524551.

Laubach, Thomas and John C. Williams (2003). "Measuring the natural rate of interest." *Review of Economics and Statistics*, 85(4), pp. 1063–1070. doi:10.1162/003465303772815934.

Laubach, Thomas and John C. Williams (2016). "Measuring the natural rate of interest redux." *Business Economics*, 51(2), pp. 57–67. doi:10.1057/be.2016.23.

Leduc, Sylvain and Zheng Liu (2020). "Can pandemic-induced job uncertainty stimulate automation?" Working Paper 2020-19, Federal Reserve Bank of San Francisco. doi:10.24148/wp2020-19.

Lee, Jaejoon and Charles R. Nelson (2007). "Expectation horizon and the Phillips curve: The solution to an empirical puzzle." *Journal of Applied Econometrics*, 22(1), pp. 161–178. doi:10.1002/jae.928.

Lenza, Michele and Giorgio Primiceri (2020). "How to estimate a VAR after march 2020." Working Paper 27771, National Bureau of Economic Research. doi:10.3386/w27771. URL https://doi.org/10.3386%2Fw27771.

Lewis, Kurt F. and Francisco Vazquez-Grande (2019). "Measuring the natural rate of interest: A note on transitory shocks." *Journal of Applied Econometrics*, 34(3), pp. 425–436. doi:10.1002/jae.2671.

Lubik, Thomas A. and Christian Matthes (2015). "Calculating the natural rate of interest: A comparison of two alternative approaches." *Richmond Fed Economic Brief*, (Oct).

Lunsford, Kurt G. (2017). "Productivity growth and real interest rates in the long run." *Economic Commentary (Federal Reserve Bank of Cleveland)*, pp. 1–6. doi:10.26509/frbc-ec-201720.

Lunsford, Kurt G. and Kenneth D. West (2019). "Some evidence on secular drivers of US safe real rates." *American Economic Journal: Macroeconomics*, 11(4), pp. 113–139. doi:10.1257/mac.20180005.

Mertens, Elmar (2014). "On the reliability of output gap estimates in real time." URL https://www.elmarmertens.com/research/workingpapers#h.icmo08w5r6zr.

Mertens, Elmar (2016). "Measuring the level and uncertainty of trend inflation." *Review of Economics and Statistics*, 98(5), pp. 950–967. doi:10.1162/REST_a_00549.

Mertens, Elmar and James M. Nason (2020). "Inflation and professional forecast dynamics: An evaluation of stickiness, persistence, and volatility." *Quantitative Economics*, 11(4), pp. 1485–1520. doi:10.3982/QE980.

Mester, Loretta J. (2018). "Demographics and their implications for the economy and policy." *Cato Journal*, 38(2), pp. 399–413.

Morley, James C. (2002). "A state–space approach to calculating the Beveridge–Nelson decomposition." *Economics Letters*, 75(1), pp. 123–127. doi:10.1016/S0165-1765(01)00581-X.

Morley, James C., Charles R. Nelson, and Eric Zivot (2003). "Why are the Beveridge-Nelson and unobserved-components decompositions of GDP so different?" *Review of Economics and*

Statistics, 85(2), pp. 235–243. doi:10.1162/003465303765299765.

Morley, James C. and Jeremy Piger (2012). "The asymmetric business cycle." *Review of Economics and Statistics*, 94(1), pp. 208–221. doi:10.1162/REST_a_00169.

Morley, James C. and Benjamin Wong (2020). "Estimating and accounting for the output gap with large Bayesian vector autoregressions." *Journal of Applied Econometrics*, 35(1), pp. 1–18. doi:10.1002/jae.2733.

Nason, James M. and Gregor W. Smith (2021). "Measuring the slowly evolving trend in US inflation with professional forecasts." *Journal of Applied Econometrics*, 36(1), pp. 1–17. doi:10.1002/jae.2784.

Orphanides, Athanasios and John C. Williams (2003). "Robust monetary policy rules with unknown natural rates." Working Paper 2003-01, Federal Reserve Bank of San Francisco. doi:10.24148/wp2003-01.

Peneva, Ekaterina V. and Jeremy B. Rudd (2017). "The passthrough of labor costs to price inflation." *Journal of Money, Credit and Banking*, 49(8), pp. 1777–1802. doi:10.1111/jmcb.12449.

Pescatori, Andrea and Jarkko Turunen (2016). "Lower for longer: Neutral rate in the U.S." *IMF Economic Review*, 64(4), pp. 708–731. doi:10.1057/s41308-016-0017-x.

Powell, Jerome H. (2018). "Monetary policy in a changing economy." In *Changing Market Structure and Implications for Monetary Policy*, volume 2018. Federal Reserve Bank of Kansas City, Jackson Hole, WY. URL https://www.kansascityfed.org/documents/6973/OpeningRemarks_JH2018.pdf.

Roberts, John M. (2001). "Estimates of the productivity trend using time-varying parameter techniques." *The BE Journal of Macroeconomics*, 1(1), p. 153460051,014. doi:10.2202/1534-6005.1014.

Shackleton, Robert (2018). "Estimating and projecting potential output using CBO's forecasting growth model." Working Paper 2018-03, Congressional Budget Office.

Sinclair, Tara M. (2009). "The relationships between permanent and transitory movements in U.S. output and the unemployment rate." *Journal of Money, Credit and Banking*, 41(2-3), pp. 529–542. doi:10.1111/j.1538-4616.2009.00220.x.

Smets, Frank and Raf Wouters (2003). "An estimated dynamic stochastic general equilibrium model of the Euro area." *Journal of the European Economic Association*, 1(5), pp. 1123–1175. doi:10.1162/154247603770383415.

Staiger, Douglas, James H. Stock, and Mark Watson (1996). "How precise are estimates of the natural rate of unemployment?" Working Paper 5477, National Bureau of Economic Research. doi:10.3386/w5477.

Stella, Andrea and James H. Stock (2015). "A state-dependent model for inflation forecasting." In Siem Jan Koopman and Neil Shephard, editors, *Unobserved Components and Time Series Econometrics*, pp. 14–29. Oxford University Press. doi:10.1093/acprof:oso/9780199683666.003.0003.

Stock, James H. and Mark W. Watson (2007). "Why has u.s. inflation become harder to forecast?" *Journal of Money, Credit and Banking*, 39, pp. 3–33. doi:10.1111/j.1538-4616.2007.00014.x.

Stock, James H. and Mark W. Watson (2016). "Core inflation and trend inflation." *Review of Economics and Statistics*, 98(4), pp. 770–784. doi:10.1162/REST_a_00608.

Stock, James H. and Mark W. Watson (2020). "Slack and cyclically sensitive inflation." *Journal of Money, Credit and Banking*, 52(S2), pp. 393–428. doi:10.1111/jmcb.12757.

Summers, Lawrence H. (2014). "U.s. economic prospects: Secular stagnation, hysteresis, and the zero lower bound." *Business Economics*, 49(2), pp. 65–73. doi:10.1057/be.2014.13.

Tallman, Ellis W. and Saeed Zaman (2017). "Forecasting inflation: Phillips curve effects on services price measures." *International Journal of Forecasting*, 33(2), pp. 442–457. doi:10.1016/j.ijforecast.2016.10.004.

Tallman, Ellis W. and Saeed Zaman (2020). "Combining survey long-run forecasts and nowcasts with BVAR forecasts using relative entropy." *International Journal of Forecasting*, 36(2), pp. 373–398. doi:10.1016/j.ijforecast.2019.04.024.

Tasci, Murat (2012). "The ins and outs of unemployment in the long run: Unemployment flows and the natural rate." Working Paper 12-24, Federal Reserve Bank of Cleveland. doi:10.26509/frbc-wp-201224.

Tasci, Murat (2019). "Challenges with estimating U star in real time." *Economic Commentary (Federal Reserve Bank of Cleveland)*. doi:10.26509/frbc-ec-201918.

Taylor, John B. (2001). "A historical analysis of monetary policy rules." In John B. Taylor, editor, *Monetary Policy Rules*, volume 31 of *Studies in Business Cycles*, pp. 319–347. University of Chicago Press. URL http://www.nber.org/chapters/c7419.

Weber, Axel A., Wolfgang Lemke, and Andreas Worms (2007). "How useful is the concept of the natural real rate of interest for monetary policy?" *Cambridge Journal of Economics*, 32(1), pp. 49–63. doi:10.1093/cje/bem018.

Williams, John C. (2018). "The future fortunes of R-star: Are they really rising?" *FRBSF Economic Letter*, 2018(13).

Wright, Jonathan H. (2013). "Evaluating real-time VAR forecasts with an informative democratic prior." *Journal of Applied Econometrics*, 28(5), pp. 762–776. doi:10.1002/jae.2268.

Wu, Jing Cynthia and Fan Dora Xia (2016). "Measuring the macroeconomic impact of monetary policy at the zero lower bound." *Journal of Money, Credit and Banking*, 48(2-3), pp. 253–291. doi:10.1111/jmcb.12300.