Finance and Economics Discussion Series
Divisions of Research & Statistics and Monetary Affairs
Federal Reserve Board, Washington, D.C.

Time-varying uncertainty of the Federal Reserve's output gap estimate

Travis J. Berge

2020-012

Please cite this paper as:
Berge, Travis J. (2021). "Time-varying uncertainty of the Federal Reserve's output gap estimate," Finance and Economics Discussion Series 2020-012r1. Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/FEDS.2020.012r1.

NOTE: Staff working papers in the Finance and Economics Discussion Series (FEDS) are preliminary materials circulated to stimulate discussion and critical comment. The analysis and conclusions set forth are those of the authors and do not indicate concurrence by other members of the research staff or the Board of Governors. References in publications to the Finance and Economics Discussion Series (other than acknowledgement) should be cleared with the author(s) to protect the tentative character of these papers.

April 2021

Time-varying uncertainty of the Federal Reserve's output gap estimate *

Abstract

A factor stochastic volatility model estimates the common component to estimates of the output gap produced by the staff of the Federal Reserve, its time-varying volatility, and time-varying, horizon-specific forecast uncertainty. Output gap estimates are very uncertain, even well after the fact, especially at business cycle turning points. However, the common component of the output gap estimates is clearly procyclical, and innovations to the common factor produce persistent positive effects on economic activity. Output gaps estimated by the Congressional Budget Office have very similar properties. Increased macroeconomic uncertainty, as measured by the common factor's volatility, leads to persistent negative responses in economic variables.

- *Keywords*: Output gap estimates; unobserved variables; real-time data; factor model; stochastic volatility; macroeconomic uncertainty.
- *JEL Codes*: E32, C53.

Travis J. Berge
Board of Governors of the Federal Reserve
20^{th} and C St, NW
Washington, DC 20551
e-mail: travis.j.berge@frb.gov

*I thank Gianni Amisano, Kurt Lewis, Jeremy Rudd, two anonymous referees, and the editor for helpful feedback and suggestions. All errors are my own. The views expressed herein are my own and should not be interpreted as reflecting the opinions of the Federal Reserve Board or the Federal Reserve system.

1 Introduction

The output gap is a critical input to policymaking. As a practical matter, however, estimating the output gap is difficult. Output gap estimates are unstable in real time, and even ex-post, considerable uncertainty remains because neither actual output nor its potential level can be measured with complete accuracy.

This paper aims to understand with what degree of accuracy the output gap can be estimated in real time. Measuring this uncertainty is not easy. Estimates of Gross Domestic Product are revised, sometimes substantially so, for many years after their initial release. And because the output gap is never observed, uncertainty bands based on ex-post forecast errors cannot be computed. Many prominent output gap estimates, such as those produced by the staff of the Federal Reserve, are judgmental, so there is no formal statistical model from which to derive measures of uncertainty. At the same time, uncertainty has become central to the communication and decision making processes at modern central banks. Orphanides (2019) has argued that clear central bank communication about uncertainty may improve economic outcomes.

Orphanides & van Norden (2002) are the first to document the instability of real-time output gap estimates. Their conclusions are pessimistic, and find that revisions to output gap estimates are as variable as the output gap itself.[1] Edge & Rudd (2016) find the output gap estimates from the Federal Reserve to be more stable. Further, and in contrast to Orphanides & van Norden (2005), Edge & Rudd find that real-time output gap estimates from the Board staff do not worsen inflation forecasts, relative to ex-post output gap estimates.

This paper adds to our understanding of real-time output gap estimation in several ways. Because policy is set in a forward-looking manner, it is the path of the output gap that is most relevant. However, the previous literature has focused on the stability of output gap *nowcasts*, the most recent value of the gap. Further, the previous literature assumed a

[1] Since Orphanides & van Norden (2002), a number of papers have developed models aimed at reducing output gap uncertainty, for example, Garratt, Lee, Mise & Shields (2008), Fleischman & Roberts (2011), Garratt, Mitchell & Vahey (2014), Aastveit & Trovik (2014), and Kamber, Morley & Wong (2018).

particular estimate of the output gap is "true," so that deviations from the truth could be used to evaluate output gap estimates.[2] Instead, I analyze the stability of the output gap at different forecast horizons. I further explicitly recognize that the output gap is never observed, by developing a state-space framework that treats real-time estimates as proxies of an unobserved output gap. Since output gap stability has changed over time, the common factor and forecast horizon specific measurement errors are endowed with time-varying volatility.[3] Changing output gap uncertainty may have first-order implications for the conduct of monetary policy, for example, Orphanides (1998) argues that policy ought to be less reactive to the output gap when it is especially uncertain, since excessive interest rate movements may be destabilizing.

The framework also provides a new way to evaluate output gap estimates. Output gap estimates have typically been evaluated according to their ability to predict price inflation. However, it has proven difficult to differentiate output gaps using this criterion given that inflation is difficult to forecast in general (Stock & Watson (2007, 2008)). For this reason, I evaluate the response of several economic variables to unexpected changes in the factor common to the output gap estimates. While price inflation is one of these variables, the judgment is more broad, since it asks the question "How is the information contained within this set of estimates linked to economic outcomes?"

The volatility of the common factor is a proxy for macroeconomic uncertainty. A recent literature explores how to measure uncertainty and its relationship with macroeconomic outcomes. Bloom (2009) uses stock market implied volatilities to proxy macroeconomic uncertainty, whereas Baker, Bloom & Davis (2016) measure uncertainty using media references to policy uncertainty. An alternative approach is to measure uncertainty as the variability common to the unforecastable component of a number of economic or financial variables, for

[2] The use of ex-post forecast errors to measure uncertainty has been studied in Knüppel (2014), Knüppel (2018), Clark, McCracken & Mertens (2018), and Reifschneider & Tulip (2017). This is also the approach of Orphanides & van Norden (2002), Edge & Rudd (2016), and Champagne, Poulin-Bellisle & Sekkel (2018).

[3] A sizable literature applies factor stochastic volatility models to macroeconomic and financial data. Early examples include Pitt & Shephard (1999), Aguilar & West (2000) and Chib, Nardari & Shephard (2006).

example, Jurado, Ludvigson & Ng (2015), Scotti (2016), and Jo & Sekkel (2019). However, it is difficult to isolate the effect of macroeconomic uncertainty: Caldara, Fuentes-Albero, Gilchrist & Zakrajšek (2016), Carriero, Clark & Marcellino (2018*a*) and Carriero, Clark & Marcellino (2018*b*) all provide different approaches for measuring uncertainty and its impact on the macroeconomy.

To summarize, Tealbook output gap estimates are very uncertain; for example, the standard deviation of the measurement error of the output gap in the previous quarter is roughly one percentage point. Nevertheless, the factor common to Tealbook output gap estimates is well identified and procyclical. Further, output, employment, wages, stock prices and the policy rate respond strongly to the factor's innovations in a manner broadly consistent with an aggregate demand shock. Increased macroeconomic uncertainty, as measured by the factor, is harmful to economic activity, consistent with the recent macroeconomic uncertainty literature such as Bloom (2009) and Jurado et al. (2015).

The paper is organized as follows. The next section introduces output gap estimates from the Federal Reserve Board staff and nonparametric measures of output gap uncertainty. Section 3 presents the econometric model. Sections 4 and 5 explore the implications of the model, focusing on output gap estimates produced by the staff of the Federal Reserve Board. The final section concludes.

2 The stability of real-time output gap estimates

Prior to each meeting of the Federal Open Market Committee (FOMC), the staff of the Federal Reserve Board prepare a comprehensive projection of the U.S. economy known as the Tealbook.[4] Since the FOMC typically meets eight times per year, there are usually two Tealbooks produced each quarter. Tealbook projections include forecasts for production, the labor market, prices, wages, financial markets, potential output, and the output gap. Impor-

[4] Prior to 2010, the projection was known as the Greenbook; I use the term Tealbook throughout to refer to this projection.

tantly, although aspects of the forecast are informed by econometric models, the Tealbook is judgmental (Reifschneider, Stockton & Wilcox, 1997; Mishkin, 2007).

To introduce notation, throughout the paper superscripts refer to vintages and subscripts to time periods. Following Orphanides & van Norden (2002) and Edge & Rudd (2016), I will use quarterly real-time vintages, using the first Tealbook projection from each quarter. Therefore, $\hat{y}^v_{v+h-1}$ denotes an estimate of the output gap at horizon h, produced at vintage v. The fixed "-1" reflects the typical notation in the literature, where the nowcast quarter ($h = 0$) is the previous calendar quarter. Edge & Rudd (2016) collect a long time series of Tealbook output gap nowcasts. I add to their dataset additional Tealbook vintages from the Federal Reserve Bank of Philadelphia.[5] These new data include real-time estimates of the path for the output gap. Beginning with the August 1987 vintage, the data contains estimates of the output gap's recent history and a forecast for the next several quarters. Since May 1996, the Tealbook contained an estimate of the output gap since approximately 1960, as well as a forecast. Because Tealbook data are made public with a five-year lag, the final Tealbook forecast in the data is from October 2014.

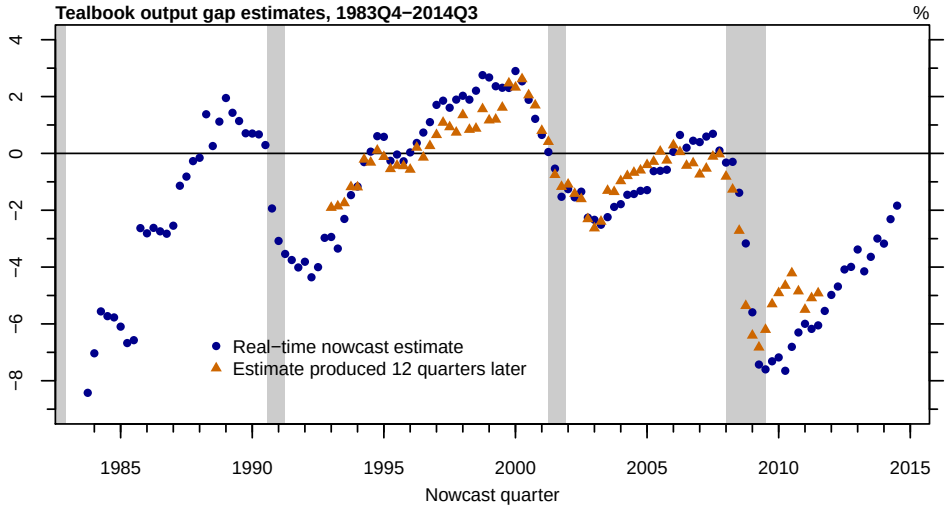
I begin by evaluating output gap revisions at different forecast horizons. For any horizon h, the revision from vintage v until an alternative vintage, $\tilde{v}$, is $\rho^v_{t+h-1} = \hat{y}^{\tilde{v}}_{v+h-1} - \hat{y}^v_{v+h-1}$. The sequence $\{\rho^v_{t+h-1}\}^V_{v=1}$ defines a distribution of revisions at horizon h. Since Tealbook output gap estimates usually extend far back in history, h can run from large negative numbers, representing backcasts, to the number of quarters forecast, typically 8 to 12 quarters ahead. Orphanides & van Norden (2002) and Edge & Rudd (2016) set $\tilde{v}$ to the most recent vintage in their sample. This approach has the drawback that some output gap estimates will have undergone more revisions than others, and implicitly assumes the most recent estimate is "true." Instead, I set $\tilde{v} = v + i$ where i is a fixed number of quarters, producing an easily interpretable distribution of revisions.

Figure 1 shows Tealbook output gap nowcasts alongside the estimate of the same quarter

[5] Appendix A1.1 contains additional details.

produced 12 quarters later. The nowcasts begin with the January 1984 Tealbook estimate of the output gap in 1983Q4 and end with the October 2014 Tealbook estimate of 2014Q3. Due to the data limitations described above, the sequence of backcasts begins later, 1993Q1, and ends with the October 2014 Tealbook estimate of 2011Q3. Table 1 shows the mean, standard deviation, minimum, maximum, and root mean-squared error of the revisions at different forecast horizons setting $i = 12$ quarters. The table shows two noise-signal ratios, defined as the ratio of either the standard deviation or the RMSE of the revisions to the standard deviation of the output gap estimates from vintages $\tilde{v}$.

Figure 1: Tealbook output gap estimates, real-time nowcast and 12 quarters later.



Notes: Tealbook output gap nowcasts (h=0) alongside estimate of same quarter produced 12 quarters hence. Sample period 1983Q4–2014Q3. Shaded regions denote NBER-defined recessions. See text for details.

The mean nowcast revision is about zero, with a standard deviation of about a percentage point. The noise-signal ratio is similar to that in Edge & Rudd (2016), notable because their dataset ends in 2006, whereas the statistics here include the Great Recession. In fact, it is clear from figure 1 that Tealbook nowcasts of the output gap reacted quickly at the onset of the recession, turning negative coincident to the NBER business cycle peak. The estimates from 12 quarters later reach a less negative trough earlier than the real-time estimates, and suggest that the economy recovered more quickly in the immediate aftermath of the recovery than was believed in real time.

Table 1: Summary statistics of revisions to Tealbook output gap estimate.

	Forecast horizon, quarters						
	-16	-12	-8	-4	0	4	8
N. obs	62	62	63	72	75	75	60
Mean	.09	.16	.27	.13	-.02	-.09	-.45
Std dev	.52	.57	.70	.68	.97	1.63	2.35
Min	-.70	-.90	-.95	-1.16	-2.19	-6.15	-8.41
Max	1.30	1.51	1.62	2.15	3.00	3.13	2.90
RMSE	.53	.59	.74	.69	.97	1.63	2.37
Noise-signal ratio							
Using SD	.23	.25	.31	.31	.44	.73	1.05
Using RMSE	.24	.26	.33	.31	.43	.73	1.06

Notes: Summary statistics of output gap revisions at selected forecast horizon. Revisions calculated as $\hat{y}^{\tilde{v}}_{v+h-1} - \hat{y}^{v}_{v+h-1}$ and setting $\tilde{v} = v + 12$ quarters. Tealbooks range from January 1984–October 2014 with missing observations as described in text. See text for details.

Output gap backcasts revise less than nowcasts, which are less volatile than forecasts. The standard deviation of revisions to the 16-quarter backcast is about .5 percentage point, roughly half that of the nowcast. Forecast revisions are much larger. The standard deviation of revisions to the eight quarter ahead forecast is more than two percentage points, and the noise-signal ratio is above one, indicating that output gap revisions are more volatile than the gap itself. (The noise-signal ratios remain below one until the seven quarter ahead forecast.) The mean revision tends to be negative for the forecasts, indicating that they were too optimistic on average. Backcasts instead tend to have a small positive mean revision. Lastly, the revisions to output gap forecasts are asymmetric. Four quarters ahead, the largest negative output gap revision is negative six percentage points, nearly two times the largest upward revision at that horizon.[6] The asymmetry is more pronounced eight quarters ahead.

To put the Tealbook revision properties into context, in the appendix I update the findings from Orphanides & van Norden (2002) and Edge & Rudd (2016). Table A4 in appendix A2 compares the Tealbook's revisions to the trend-cycle estimates considered Orphanides &

[6] The largest negative revision at the four quarter horizon is for the output gap in 2009Q2. The July 2008 Tealbook projection of the output gap in 2009Q2 was -1.5 percent; in 2011Q3, the Tealbook estimate of that output gap was -7.7 percent. The largest positive revision at the four quarter horizon occurs in 2011Q2: the August 2010 projection of that quarter's output gap was -7.3 percent, compared to the July 2013 estimate of -4.2 percent.

van Norden (2002), and from Mueller & Watson (2017) and Hamilton (2018). The output gap estimates from the univariate models usually have larger revisions than the Tealbook estimates. The exception is the Hamilton (2018) filter, which produces output gap estimates that have ex-post revisions comparable to, or even smaller than, the Tealbook. Unsurprisingly, for most models, noise-signal ratios become smaller when the estimate is further in the past. The noise-signal ratios reported in table A4 are somewhat smaller than the ones found in Orphanides & van Norden (2002), likely reflecting the difference in sample periods.

3 Time-varying uncertainty of the Federal Reserve's output gap

I now introduce a factor stochastic volatility (FSV) model. We observe multiple estimates of the output gap in each quarter. Let $i = -B, ..., H$, where B denotes a maximum backcast horizon and H a maximum forecast horizon. We observe $N{=}B{+}H{+}1$ estimates of the output gap in period t, y_t. These estimates have a factor structure:

$$\hat{y}_t = \lambda f_t + e_t \tag{1}$$

where the vector $\hat{y}_t = (\hat{y}_t^{t-B}, \hat{y}_t^{t-B+1}, ..., \hat{y}_t^{t+H})'$ is $N \times 1$, $\lambda = [\lambda_{-B}, \lambda_{-B+1}, ..., \lambda_H]'$ is a vector of factor loadings, f_t is a latent scalar, and e_t a vector of idiosyncratic errors. As an identifying assumption, the first element of λ, the loading on the most distant backcast, is set to one.

The factor and errors are heteroskedastic stationary independent autoregressive processes:

$$f_t = \phi_0 + \sum_{i=1}^{P} \phi_i f_{t-i} + e^{\frac{1}{2}h_{ft}}\varepsilon_{ft}; \quad \varepsilon_{ft} \sim N(0,1) \tag{2}$$

$$e_{it} = \sum_{j=1}^{Q} \rho_j e_{it-j} + e^{\frac{1}{2}h_{it}}\varepsilon_{it}, \quad i = -B, -B+1, ..., H; \quad \varepsilon_{it} \sim N(0,1). \tag{3}$$

The constant term ϕ_0 in equation (2) allows the latent factor to have a non-zero uncondi-

tional mean. That measurement errors are ex-ante zero mean embeds an assumption that $\hat{y}_t$ contains unbiased estimates of the output gap. This assumption is justified given that Tealbook projections of important macroeconomic variables such as real GDP growth, the unemployment rate, and inflation appear to be unbiased.[7] Table 1 further suggests that a null hypothesis of unbiasedness in the Tealbook output gap projection would not be rejected based on its revisions. Autocorrelated measurement errors allow for what Jacobs & van Norden (2011) denote spillover effects, the possibility that a revision to the estimate of period t from vintage v would also carry information about the estimates of nearby periods. In practice, revisions to slow-moving macroeconomic trends such as structural productivity would produce this kind of spillover. Finally, the log volatilities follow independent SV processes:

$$h_{it} = h_{it-1} + \eta_{it} \tag{4}$$

where $\eta_{it} \sim N(0, \sigma_i^2)$ and is independent of ε_{ft} and ε_{it}.

Several estimands are of interest. The first is the latent common factor. The factor summarizes the information content of the estimates contained in $\hat{y}$, while setting λ_{-B} to one scales the factor to the most distant backcast. The standard deviations of the idiosyncratic errors measure the precision of the output gap estimates at each forecast horizon, while the standard deviation of the common factor is a measure of the volatility of the business cycle, a proxy for macroeconomic uncertainty. Ex-ante, we may expect a secular reduction in the volatility of the common factor and the measurement errors. We may further expect these volatilities to vary across the business cycle. The FSV specification is flexible enough to capture both behaviors. Lastly, while we expect backcasts to be more precise than forecasts, the model structure is ex-ante agnostic about the relative variances of the elements of e_t.

[7] See: Messina, Sinclair & Stekler (2015); Reifschneider & Tulip (2017); and Berge, Chang & Sinha (2019).

3.1 Estimation

The baseline FSV model is estimated using Tealbook output gap estimates, setting B to 12 and H to eight quarters, chosen to reflect the most relevant period to policymakers. Thus, the common factor is extracted from at maximum 21 estimates produced over a roughly five-year period. Due to data limitations, estimation begins in 1984. The final Tealbook in the sample is from October 2014, and therefore the dataset consists of 123 output gap nowcasts (1984Q1–2014Q3), and the full sample period is 1984Q1–2016Q3. The lag length of the autoregressive process for the factor is set to two, while the measurement errors follow AR(1) processes.

The model can be cast as a conditionally Gaussian state space model, which is estimated with Bayesian Markov Chain Monte Carlo methods. Details are left to the appendix, but the sampler consists of grouping the parameters and latent variables into blocks and repeatedly drawing from each block's conditional posterior distribution. Following a burn-in period of 10,000 draws, I collect 2,000 draws from the posterior distribution using a thinning factor of 10. Priors are standard and not tightly specified; see appendix A5. Convergence diagnostics are reported in appendix A6.

4 Tealbook model results

4.1 Full sample estimates

The top panel of figure 2 reports the common factor to the Tealbook output gap estimates.[8] The factor is very clearly pro-cyclical. There are three local peaks in the sample period, each of which predates an NBER-defined business cycle peak by four to six quarters. The recoveries from both the three NBER-defined recessions are slow, and the factor is above zero for an extended period of time only during the late 1990s. The latent factor bottoms out well after the 1991 and 2001 NBER business cycle troughs. Following the 2001 recession, the

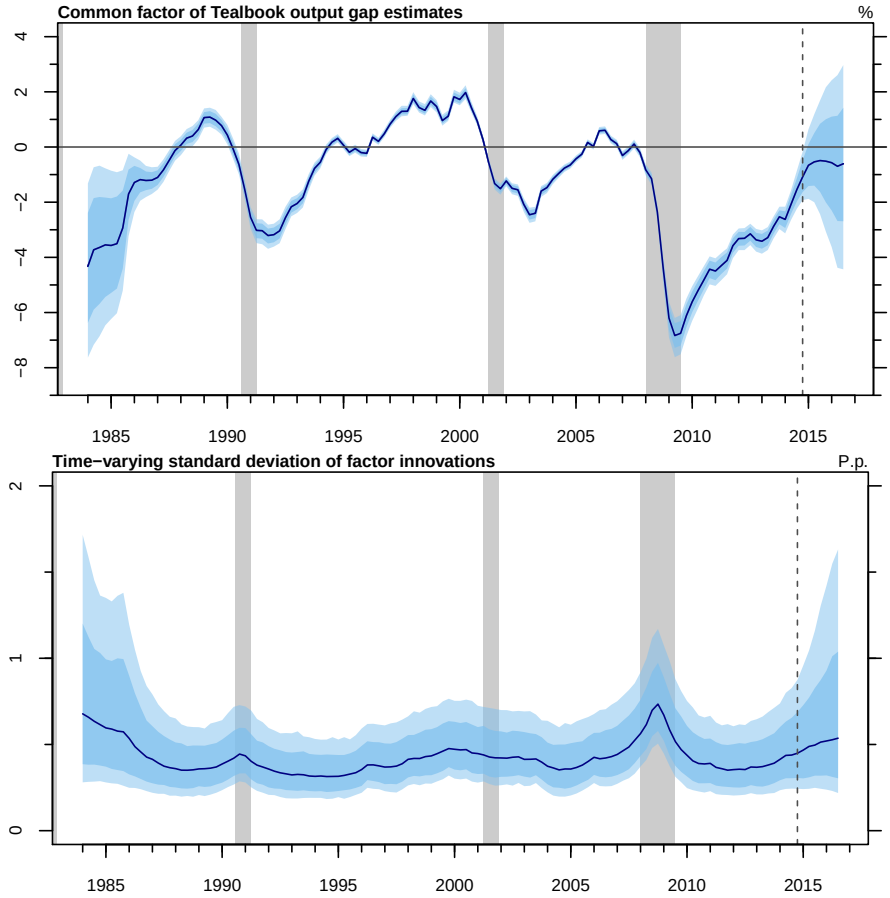
[8] The quarterly and four-quarter changes in the common factor are shown in the appendix; figure A5.

output gap estimate reaches a minimum only in 2003, nearly 2 years following the NBER's declaration of the business cycle trough, and turns positive only in the second half of 2005. Another local trough roughly coincides with the end of the Great Recession, at a level that implies that output was nearly seven percentage points below its potential level. Uncertainty about the common factor is larger at the beginning and end of the sample, reflecting a paucity of data, and at turning points. The standard deviation of innovations to the common factor is plotted in the bottom panel. There is a hint of a decline in volatility from 1984 through the early 1990s. Thereafter, volatility is fairly stable, albeit with a modest increase in the late 1990s—concurrent to the productivity boom of that period—and a noticeable increase during the Great Recession.

Figure 3 shows three measures of uncertainty for the output gap estimates. The top panel plots the log volatilities of selected measurement errors. Prior to 1996, there are very few backcasts and hence they are very uncertain. Thereafter, the backcasts have the lowest uncertainty. The eight-quarter backcast achieves its minimum value in the mid-2000s. After that, the uncertainty rises, especially in the recovery following the Great Recession, reflecting the difficulty in understanding the effect of the recession on potential output. As expected, forecasts are more uncertain, and forecast uncertainty is counter-cyclical. Interestingly, nowcast uncertainty is not especially counter-cyclical. Although nowcast uncertainty increases around the time of the Great Recession, it remains relatively elevated through the remainder of the sample.

The middle panel shows the share of the variance of each horizon's output gap estimate due to the common factor. From the mid 1990s and through the Great Recession, nearly all of the backcast uncertainty derives from the common factor itself. The share of the nowcast uncertainty from the common factor trends higher from the late 1980s and peaking in 2009, indicating an improvement in its precision, consistent with Edge & Rudd (2016). In contrast, the common factor accounts for at most roughly one-quarter of the four and eight quarter ahead forecast uncertainty. The final panel shows a noise-signal ratio, defined the ratio of

Figure 2: Common factor to Tealbook output gap estimates and its volatility.

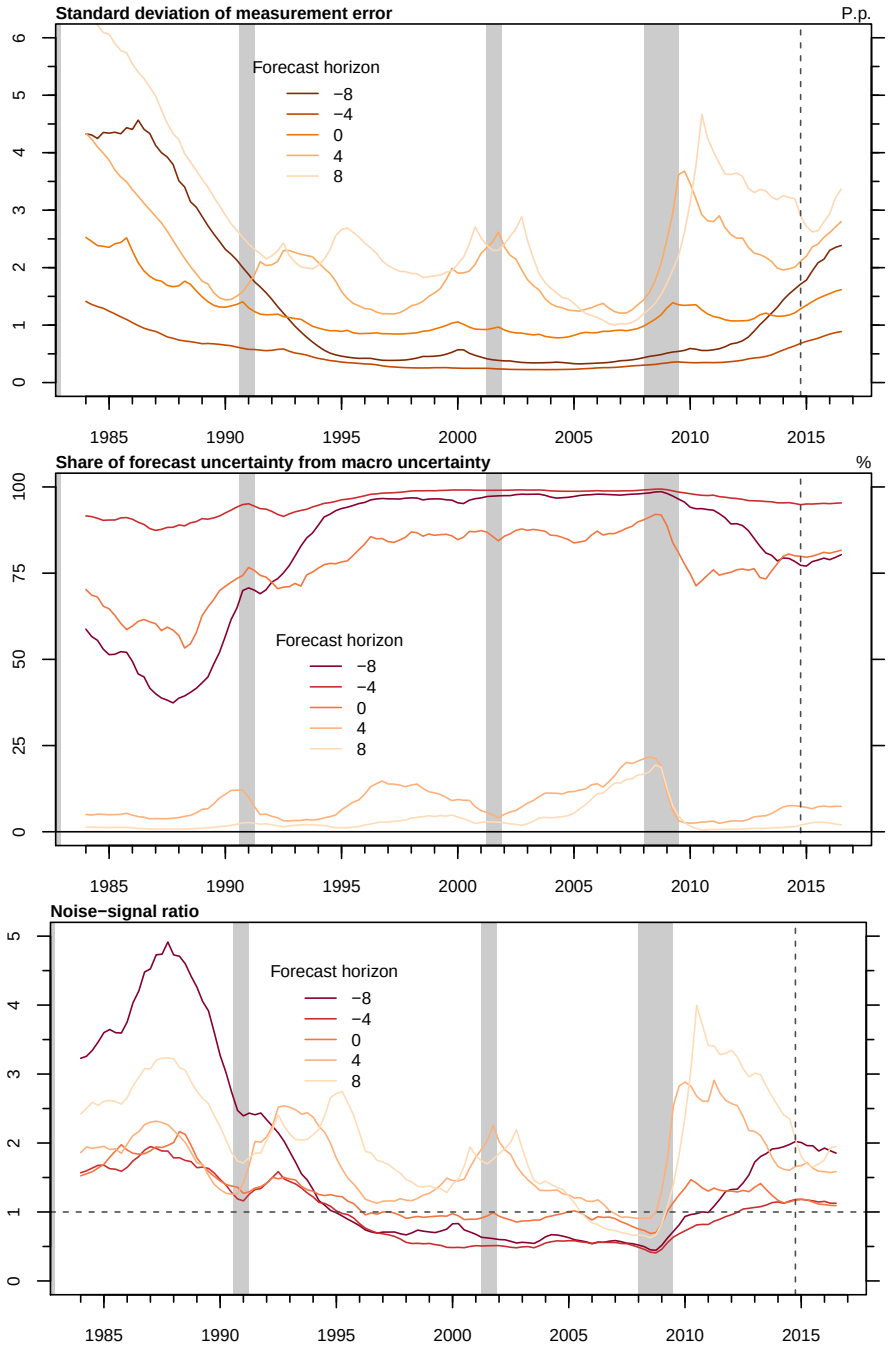


Notes: Sample period is 1984Q1–2016Q3 with most recent Tealbook forecast produced in October 2014, vertical dashed line. Navy lines show the posterior median estimate. The dark and light shaded areas are the 70% and 90% posterior credible sets. Gray shaded regions denote NBER recessions. See text for details.

the standard deviation of measurement error innovations to that of the factor. The overall message of the figure is the same: backcasts of the output gap have noise-signal ratios below one except at the ends of the sample. The noise-signal ratio of the nowcast trends lower through the Great Recession, and is below one roughly between 1996 and 2009. Output gap forecasts are very uncertain, especially at turning points.

The relative imprecision of forecasts is also reflected in the factor loadings. Figure 4 shows that the common factor relies mostly on backcasts. Recent backcasts receive a somewhat larger weight, possibly reflecting that more effort is given to tracking economic activity in recent quarters than in the more distant past. Beyond the nowcast quarter, the factor

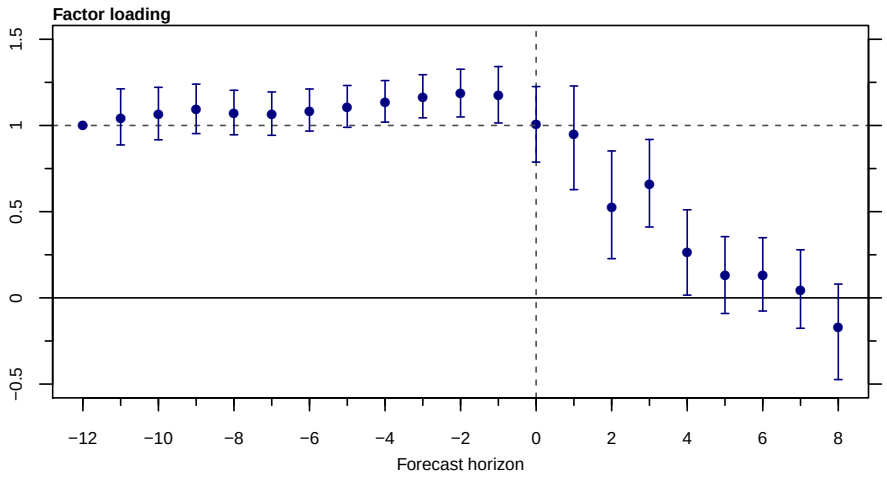
Figure 3: Tealbook output gap estimate uncertainty.



Notes: Sample period is 1984Q1–2016Q3 with most recent Tealbook forecast produced in October 2014, vertical dashed line. Top panel: standard deviation of measurement errors ($\sqrt{e^{h_{it}}/(1-\rho_i^2)}$). Middle panel: share of variance explained by common factor's uncertainty ($\lambda_i^2 var(f_t)/var(\hat{y}_{it}) \times 100$). Bottom panel: ratio of the standard deviation of measurement error innovation to that of the common factor ($e^{\frac{1}{2}h_{it}}/e^{\frac{1}{2}h_{ft}}$). Gray shaded regions denote NBER recessions. See text for details.

loadings become smaller: the loading on the two quarter-ahead forecast is just .6 while the 90 percent credible interval for output gap projections beyond 5 quarters ahead all include zero.

Figure 4: Posterior estimates of factor loadings.



Notes: Point is posterior median estimate, whiskers show 90 percent credible interval. Factor loading at h=-12 set to unity. See text for details.

Table 2 gives posterior distributions of other selected parameters. The common factor is persistent and most of the posterior distribution of the unconditional mean of the factor, μ, is negative. Measurement errors tend to be positively autocorrelated, suggesting strong spillover effects. Estimates of the variability of the log-volatilities are generally quite similar.

To highlight the practical implications of the model, figure 5 shows the real-time estimate of the common factor in October 2014 alongside that Tealbook estimate. Uncertainty about the common factor is smallest for previous quarters: the 90 percent credible interval for the four-quarter backcast of the output gap is about .8 percentage point. The nowcast has a 90 percent credible interval of 1.4 percentage points (-2.3 to -.9 percent), and uncertainty increases dramatically looking ahead. The bottom panel plots the Tealbook output gap estimate along with the model-implied, forecast horizon specific 70 and 90 percent credible intervals. The uncertainty surrounding any particular Tealbook estimate is noticeably larger than for the common factor, the Tealbook nowcast has a credible interval of about 4

Table 2: Posterior distribution for selected model parameters.

	Posterior distribution				
Parameter	5%	15%	50%	85%	95%
ϕ_0	-.06	-.05	-.02	.01	.03
ϕ_1	1.33	1.39	1.46	1.54	1.58
ϕ_2	-.63	-.59	-.51	-.43	-.38
μ	-1.21	-.93	-.43	.21	.73
ρ (h=-8)	.41	.49	.66	.78	.84
ρ (h=-4)	-.06	.07	.23	.39	.48
ρ (h=0)	.74	.79	.87	.92	.95
ρ (h=4)	.85	.88	.91	.95	.97
ρ (h=8)	.84	.87	.92	.96	.98
σ_f	.26	.31	.39	.52	.60
$\sigma_{h=-8}$	.31	.36	.46	.60	.70
$\sigma_{h=-4}$	.26	.30	.38	.49	.58
$\sigma_{h=0}$	.25	.29	.36	.48	.55
$\sigma_{h=4}$	.30	.35	.44	.56	.65
$\sigma_{h=8}$	.33	.38	.48	.64	.75

Notes: Posterior distribution for selected model parameters. Parameter μ is the unconditional mean of the common factor, $\phi_0/(1-\phi_1-\phi_2)$. See text for details.

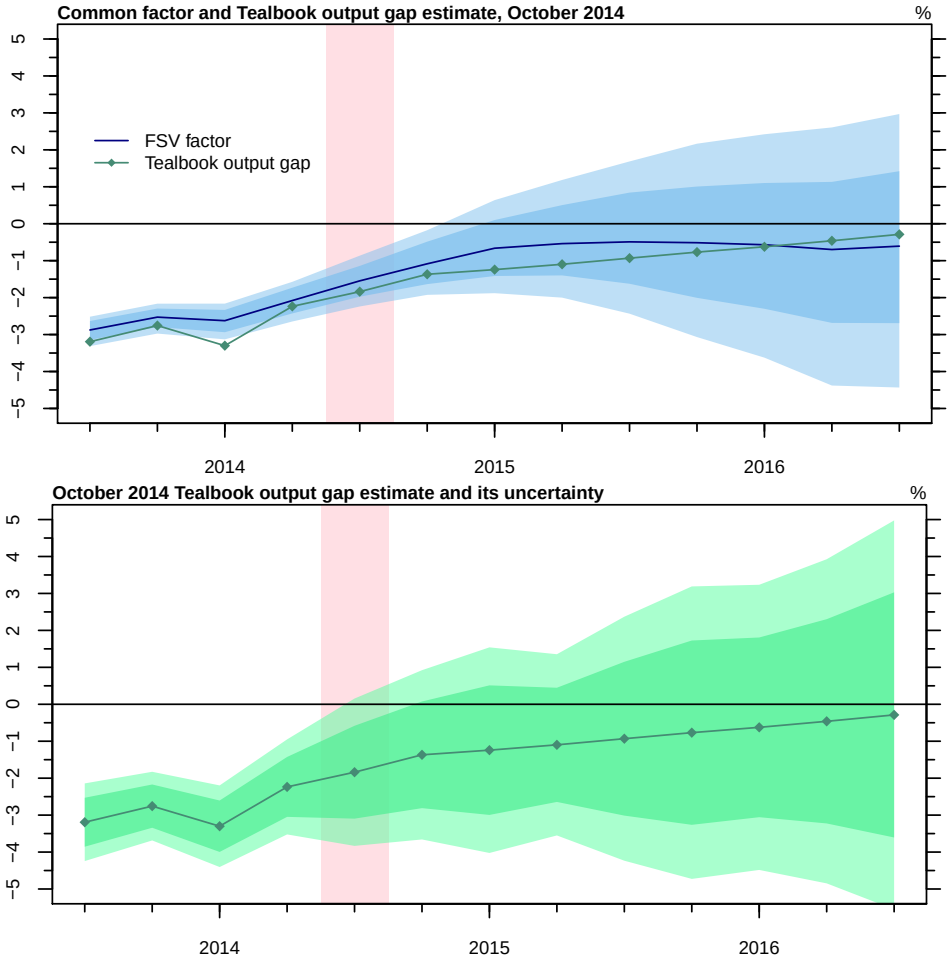
percentage points.

4.2 Forward-looking estimates

In this section, I produce two alternative model estimates intended to minimize the influence of the backward-looking nature of the model. First, I estimate the model using estimates that are purely forward-looking, setting B=0 and H=8. As above, this is a full-sample estimate, but sheds light on whether the identification of the common factor relies on backcasts. Secondly, I estimate a sequence of FSV models using an expanding window. In each period, I re-estimate the model and produce the common factor in the nowcast quarter. The sequence of factor nowcasts are fully real time. As in the baseline, I set B=12 and the out-of-sample estimation begins in 1998Q1.

The results are in figure 6. The forward-looking FSV estimates lag somewhat the baseline,

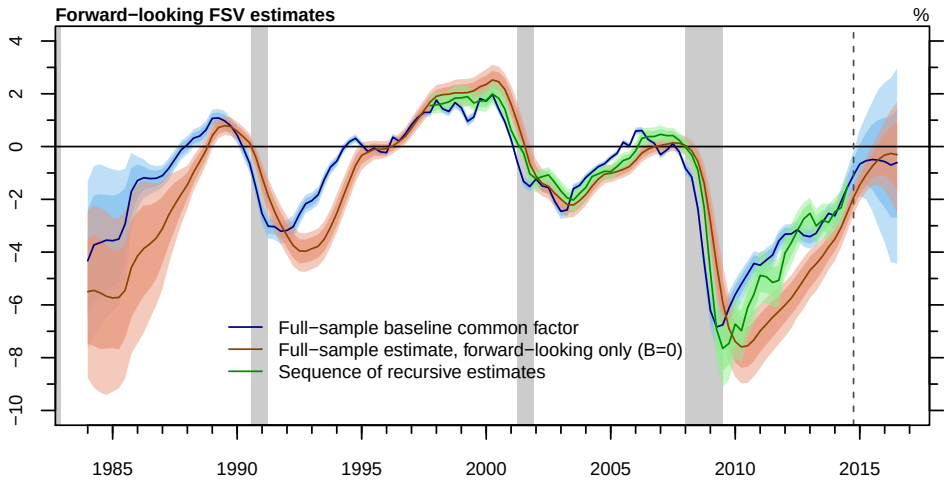
Figure 5: Model estimates in October 2014.



Notes: Top panel: model estimate of common factor and its uncertainty alongside October 2014 Tealbook output gap estimate. Bottom panel: Tealbook output gap estimate and its model-implied uncertainty bands. Pink shaded region denotes the nowcast quarter, 2014Q3. The dark and light shaded areas are 70% and 90% posterior credible sets. See text for details.

especially at business cycle troughs. The local troughs of the baseline factor tend to be smaller, and the factor subsequently recovers more quickly, relative to the forward-looking estimates. The recursive nowcasts are a bit more volatile than the series estimated with a full sample, especially after 2008. On the whole, however, the three estimated output gaps closely follow one another; the correlation coefficients across the three series exceed 90 percent. That the forward-looking factors are similar to the in-sample estimate suggests that backcasts are not a determining influence on the model. It also highlights that although uncertain, real-time Tealbook output gap forecasts contain valuable information.

Figure 6: Baseline and forward-looking FSV estimates.



Notes: Sample period is 1984Q1–2016Q3 with most recent Tealbook forecast produced in October 2014, vertical dashed line. Dark and light shaded areas are 70 and 90 percent credible intervals. Gray shaded areas denote NBER defined recessions. See text for details.

5 What is the output gap?

5.1 The Tealbook output gap's relationship to other macroeconomic shocks

I first explore how innovations to the common factor relate to other macroeconomic shocks. To the extent that macroeconomic shocks affect output but are not explicitly parsed by the Tealbook projection, these innovations may ultimately reveal themselves via the latent estimate. In contrast, if the Tealbook measures a genuine cyclical component to output, than factor innovations should be uncorrelated to other true and identified macroeconomic shocks. As in Stock & Watson (2012), I consider a number of economic shocks. The first is a utilization-adjusted TFP shock (Fernald 2012). I calculate an oil price shock following Hamilton (2003), examine the fiscal policy news shock of Ramey (2011), and extend the Romer & Romer (2004) measure of monetary policy shocks. I consider several proxies of uncertainty: the policy uncertainty index of Baker et al. (2016); the realized volatility of the S&P 500; and the excess bond premium of Gilchrist & Zakrajsek (2012). See appendix A1.3 for details.

Table 3 shows that the factor innovation is correlated with other economic shocks, suggesting that Tealbook output gap estimates are an amalgamation of many different phenomena. Factor innovations are positively correlated to TFP shocks, highlighting the difficulty in differentiating between changes in potential output and the output gap. There is a modest positive correlation to Romer & Romer (2004) monetary policy shocks, which is surprising since Romer & Romer purge intended monetary policy changes of expectations as summarized by Tealbook projections of GDP, inflation, and unemployment. Factor innovations are negatively correlated to the measures of uncertainty and net oil price increases. The uncertainty shock from the FSV model—innovations to the common factor's volatility—is strongly positively correlated to oil price increases. Model-based uncertainty is negatively correlated to productivity and the Baker, Bloom & Davis policy uncertainty measure.

Table 3: Correlation of FSV innovations with other macroeconomic shocks.

Macroeconomic shock	Common factor innovation	Uncertainty innovation
Productivity (Fernald, 2012)	.48	-.25
	[.00]	[.02]
Net oil increase (Hamilton, 2003)	-.13	.29
	[.08]	[.00]
Fiscal policy news (Ramey, 2011)	.02	.08
	[.40]	[.19]
Monetary policy (Romer-Romer, 2004)	.13	-.18
	[.02]	[.13]
Policy uncertainty (BBD, 2016)	-.30	-.34
	[.01]	[.01]
Realized volatility of S&P 500	-.39	-.02
	[.03]	[.40]
Excess bond premium	-.50	-.10
	[.00]	[.15]

Notes: Table reports pairwise correlation coefficient between Tealbook-based FSV factor innovations and each macroeconomic shock. Sample period is 1984Q1–2014Q4. Bracketed value is the p-value from a univariate regression, calculated using standard errors robust to heteroskedasticity and autocorrelation. Macroeconomic shocks are described in appendix A1.3. See text for details.

5.2 Response of other macroeconomic variables

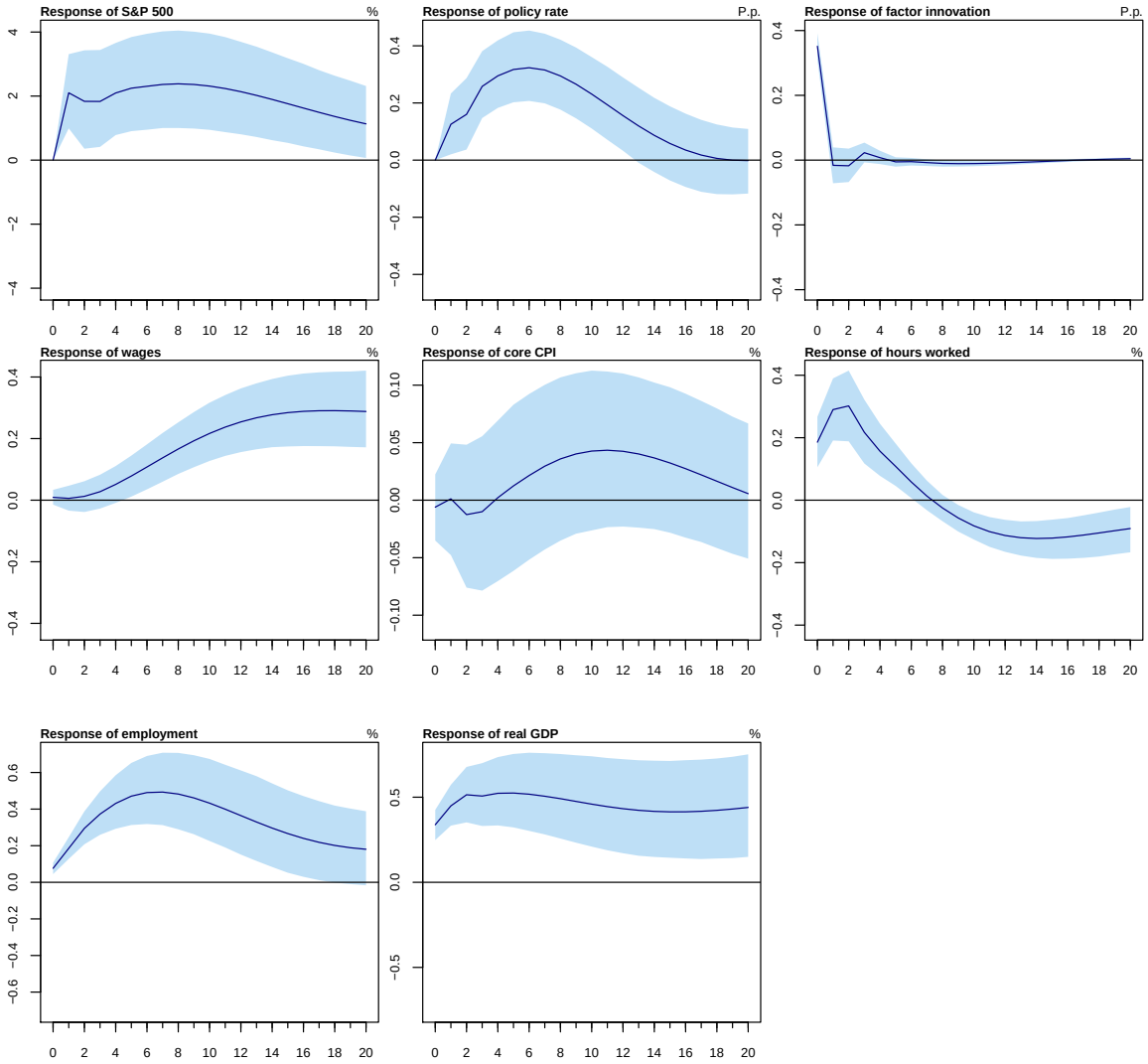
I next consider the response of economic variables to innovations to the common factor. I do so through the lens of a recursively identified vector autoregression model. The variables in estimation order are: the S&P 500 stock market index, the Wu & Xia (2016) shadow policy rate, the factor innovation, average hourly earnings, the core consumer price index, hours worked, employment, and real GDP. The ordering allows economic variables to respond contemporaneously to factor innovations, conditional on changes in equity markets and the policy rate. The eight variable VAR is estimated over the period 1984Q1–2014Q4, and is denoted VAR-8.[9]

The solid lines in figure 7 show the responses to a one standard deviation (.35 pp) shock to the common factor. The shaded areas represent 90 percent confidence intervals. The responses are broadly consistent with an aggregate demand shock. Focusing on the real variables, hours worked, employment, and output all increase. The response of real GDP to the innovation is about one-to-one contemporaneously; thereafter, output responds in a persistent, hump-shaped manner. After five years, output settles at a permanently higher level, consistent with the positive correlation between common factor and TFP innovations. Employment responds similarly, while hours worked overshoot, consistent with the existence of hiring frictions in the labor market (Bloom 2009). The response of wages is unambiguously positive but takes about a year to manifest. Core prices have a muted and delayed response. These delayed responses are consistent with sticky prices and wages as in the New Keynesian literature. Finally, stock market prices and the policy rate increase.

Table 4 reports the forecast error variance decomposition for wage and price inflation, employment, and real GDP. Factor innovations explain large portions of forecast error variance for real activity, namely, 15-25 percent of the error variance in employment and 20-30 percent of real output. The innovations explain a sizable forecast error variance of wages,

[9] Data sources are reported in appendix A1. The VAR lag length is two quarters, as determined via the AIC. Several alternative VAR specifications can be found in appendix A4. The time-series of the factor innovations is shown in figure A6 in the appendix.

Figure 7: Response to Tealbook factor innovation.



Notes: Impulse response from eight variable VAR; factor innovation ordered third. Estimation sample 1984Q1–2014Q4. Figures show response to one standard deviation innovation to common factor (.35 percentage point). Shaded area denotes 90 percent confidence bands from Kilian (1998) bootstrap. See text for details.

but only at longer horizons: at the four quarter horizon, less than one percent of the error variance is explained, whereas 20 quarters hence that number increases to more than 20 percent. The error variance of inflation is hardly explained by the common factor.

Employment innovations are similar to the common factor. Both explain roughly 20 percent of the forecast error variance of employment and real GDP at longer horizons, although the factor innovation explains more of the forecast error variance of wages. Real

Table 4: Forecast error variance decomposition for selected macroeconomic variables.

	Quarters ahead				
	4	8	12	16	20
Fraction explained by factor innovation (%)					
Wages	.4	6.2	14.1	19.0	21.0
Prices	.1	.8	2.4	3.6	4.0
Employment	24.8	22.3	19.1	17.1	16.4
Real GDP	33.6	25.4	21.4	19.5	18.7
Fraction explained by employment (%)					
Wages	.1	2.4	6.9	10.4	12.3
Prices	1.0	1.5	1.2	1.0	1.0
Employment	33.4	24.4	20.9	19.0	18.3
Real GDP	15.5	18.0	17.5	16.9	16.6
Fraction explained by real GDP (%)					
Wages	.1	.8	2.3	3.8	4.9
Prices	.4	.2	.5	1.1	1.6
Employment	.5	.8	.7	.6	.6
Real GDP	12.4	7.7	6.2	5.4	5.0
Fraction explained by policy rate (%)					
Wages	18.5	29.7	28.7	24.2	20.4
Prices	0.6	2.0	2.9	2.9	2.6
Employment	2.5	1.1	.7	.7	.6
Real GDP	1.8	.7	.6	.5	.5

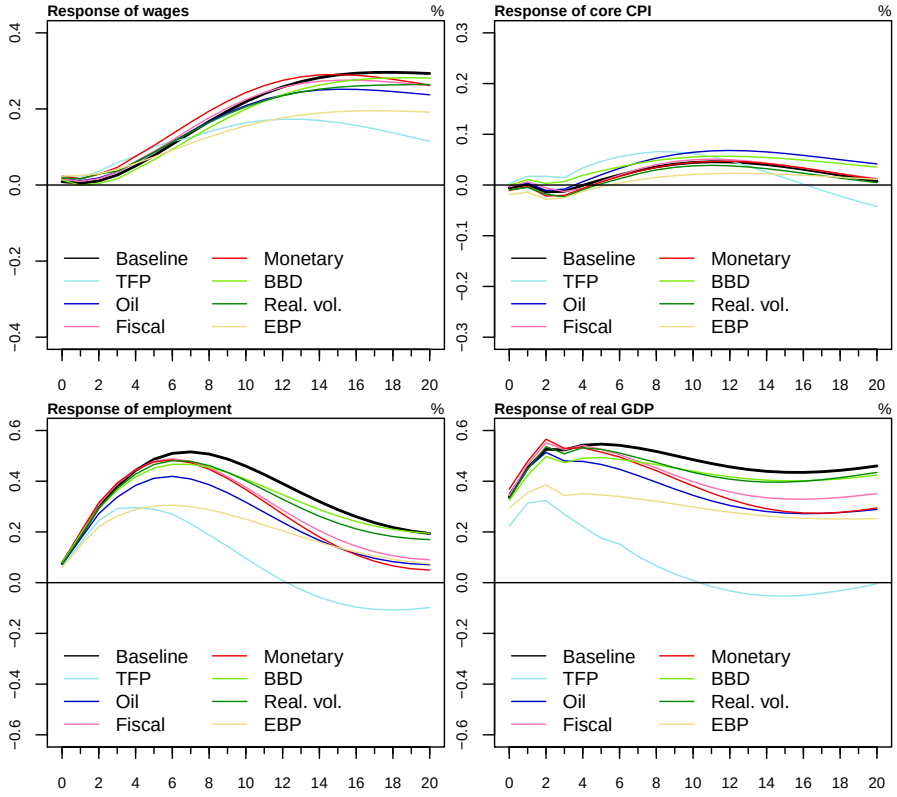
Notes: See text for details.

GDP explains relatively little of the forecast error variance of the variables. Finally, the policy rate explains a substantial portion of the forecast error variance of wages. However, the policy rate does not explain price inflation, employment or real GDP.

Since factor innovations are correlated with other macroeconomic shocks, I reestimate the VAR presented above, but add each of the other macroeconomic shocks considered in table 3 one-by-one. The other shock is ordered above the factor innovation in the VAR. Conditional on the other macroeconomic shocks, the responses of wages, employment, and output to the factor innovation are somewhat smaller. The TFP shock especially reduces the employment and output responses. On the whole, however, figure 8 indicates that the VAR is robust to the inclusion of these macroeconomic shocks.

The appendix contains a number of robustness checks that further examine the VAR.

Figure 8: Response of selected variables conditioning on other economic shocks.



Notes: Impulse response from nine variable VAR. Order of VAR is: S&P 500; Wu-Xia policy rate; other macroeconomic shock; factor innovation; wages; prices; hours worked; employment; real GDP. Sample period 1984Q1–2014Q4. Figures show response to one standard deviation innovation to common factor (.35 percentage point). See text for details.

- Figure A7 shows impulse response functions from the recursively estimated VAR with the factor innovation ordered first. The responses are somewhat larger in magnitude but otherwise very similar to the baseline.

- Figure A8 shows impulse response functions from the recursively estimated VAR with the factor innovation ordered last. The responses of other macroeconomic variables are smaller in magnitude, but again show a qualitatively similar pattern as the baseline.

- In order to see whether the Great Recession drives the macroeconomic dynamics in the VAR-8, figure A9 estimates the baseline VAR using only pre-Great Recession data (1984–2006). The impulse responses are very similar to the baseline.

- I show the response of the baseline VAR-8 to innovations to real GDP and employment in figures A10 and A11. Qualitatively, an innovation to employment produces similar impulse responses as the factor innovation, while the innovation to real GDP generally produces more muted responses.

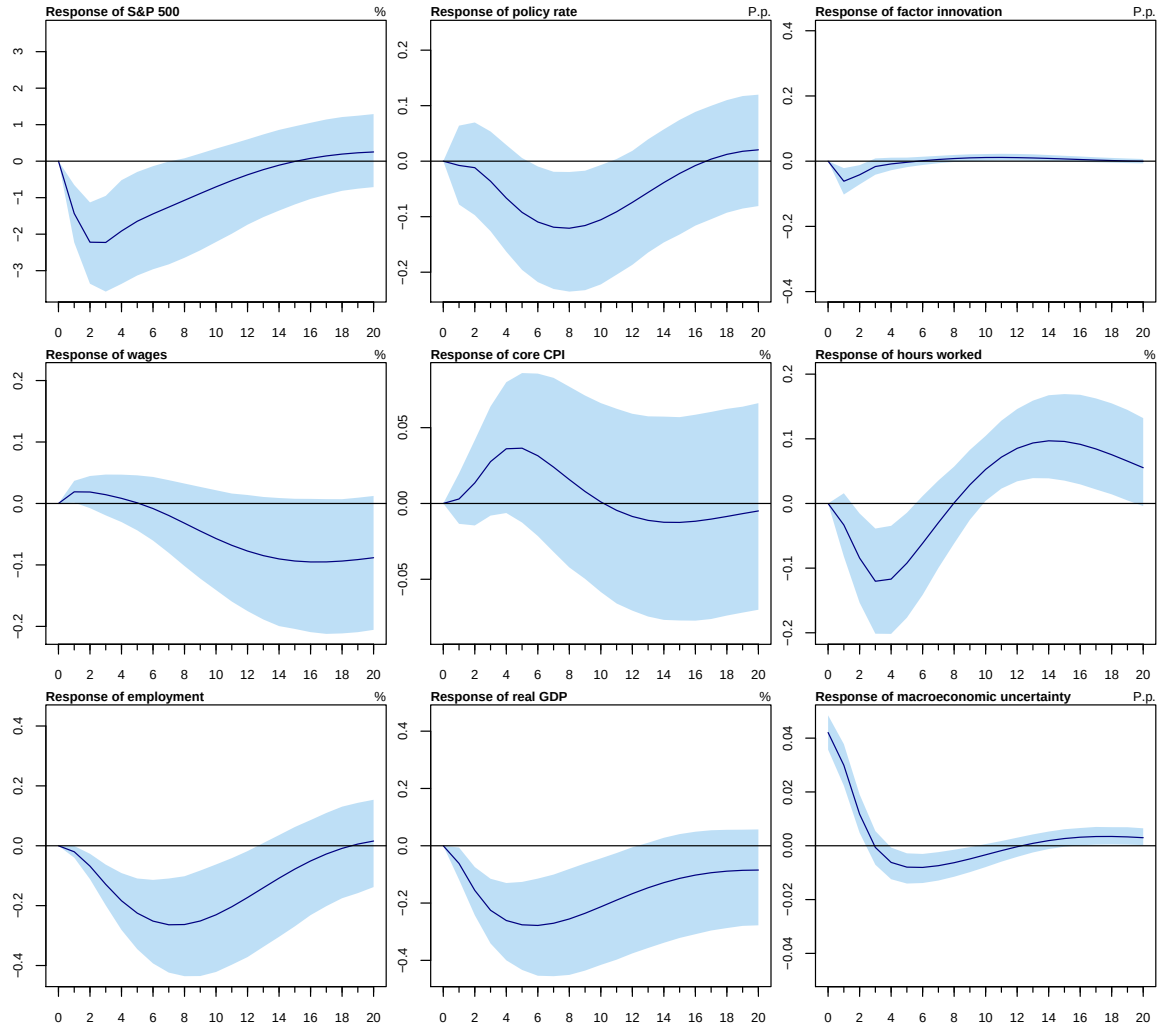
- Total CPI is somewhat more responsive to the factor innovation than the core price index, and the impulse responses of the other variables are essentially unchanged (figure A12).

5.3 Response to economic uncertainty

The FSV framework also provides a measure of macroeconomic uncertainty, namely, the innovation to the log-volatility of the common factor. I augment the VAR-8 with this measure of uncertainty in order to examine the link between macroeconomic uncertainty and economic outcomes.

The results are shown in figure 9. The final panel shows that a one standard deviation uncertainty shock dies out over the course of about one year. Output and employment decline with a nadir six to eight quarters after the shock. Employment and hours worked both decline initially, with hours worked eventually turning positive. Wages have a delayed and negative response, whereas core prices show a positive albeit uncertain response. Turning to the financial variables, the response of the S&P price index is negative and significant. And, as before, the policy rate responds to changes in the common factor as would be prescribed by a Taylor rule. Overall, the responses of these variables are qualitatively similar or perhaps somewhat larger than the estimated responses reported in, for example, Bloom (2009), Jurado et al. (2015), and Jo & Sekkel (2019).

Figure 9: Response to macroeconomic uncertainty.



Notes: Impulse responses to macroeconomic uncertainty as measured by the innovation to the common factor's volatility. Estimated over period 1984Q1–2014Q4. Figures show response to one standard deviation innovation. Responses plotted in order of VAR estimation (uncertainty shock is ordered last). Shaded area denotes 90 percent confidence bands from Kilian (1998) bootstrap. See text for details.

5.4 Comparison to other output gap estimates

The baseline FSV model is estimated using output gap estimates produced by the staff of the Federal Reserve Board. This section compares the baseline results to FSV models estimated with alternative output gap estimates. The first is from the Congressional Budget Office (CBO), an output gap estimate commonly used in policy discussions. I compile quarterly

real-time vintages of the CBO output gap by comparing real-time vintages of GDP to the most recently produced CBO estimate of potential output; see appendix A3 for details. The vintages begin in 1991Q1 and continue through 2019Q4. I also produce real-time vintage estimates of the output gap from two univariate trend-cycle decompositions, the HP filter and the Hamilton (2018) filter. The HP filter is widely used, while Hamilton's filter was the best performing univariate model considered in section two. For these two models, quarterly real-time vintage output gap estimates are produced using vintage data from 1984Q1 to 2019Q4.

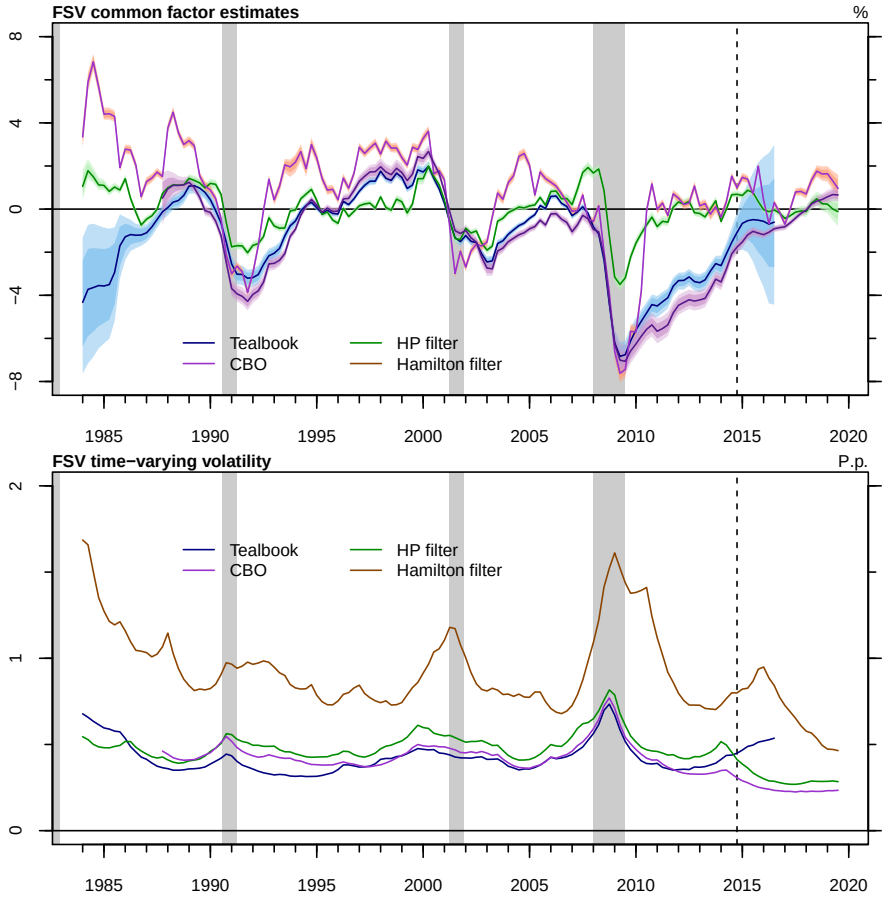
The four factor estimates are shown in figure 10. Although the CBO factor tends to run below that of the Tealbook, the common factors to the Tealbook and CBO are very similar, with a contemporaneous correlation exceeding 95 percent. The HP and Hamilton factors are more similar to each other, although the amplitude of the Hamilton-filter common factor is larger. Most importantly, the Tealbook- and CBO-based factors recover more slowly following recessions than the other two. For example, following the 2008-2009 recession, the CBO factor doesn't return to zero until 2018; the HP and Hamilton factors return to zero by 2011. The bottom panel of the figure shows the time-varying volatilities. The CBO, Tealbook and HP filter factors have broadly similar volatilities, while the Hamilton common factor is more volatile.

Figure 11 compares the VAR-8 results. Each VAR is estimated from 1987Q4–2014Q4, the period common to the four FSV models, and show the response of other macroeconomic variables to a factor innovation of .35 percentage point, as in the baseline VAR.[10]

Innovations to the Tealbook- and CBO-based common factors produce very similar responses to the other economic variables. The impulse responses from the Tealbook and CBO factor innovations are stronger than those from the Hamilton or HP filter-based models. As was the case in the baseline VAR, the impulse responses broadly show an aggregate demand shock, with both nominal and real variables increasing following the innovation. The policy

[10] Results for the full sample of CBO factor innovations are available in appendix A3.

Figure 10: Comparison of factor estimates, 1984–2019.

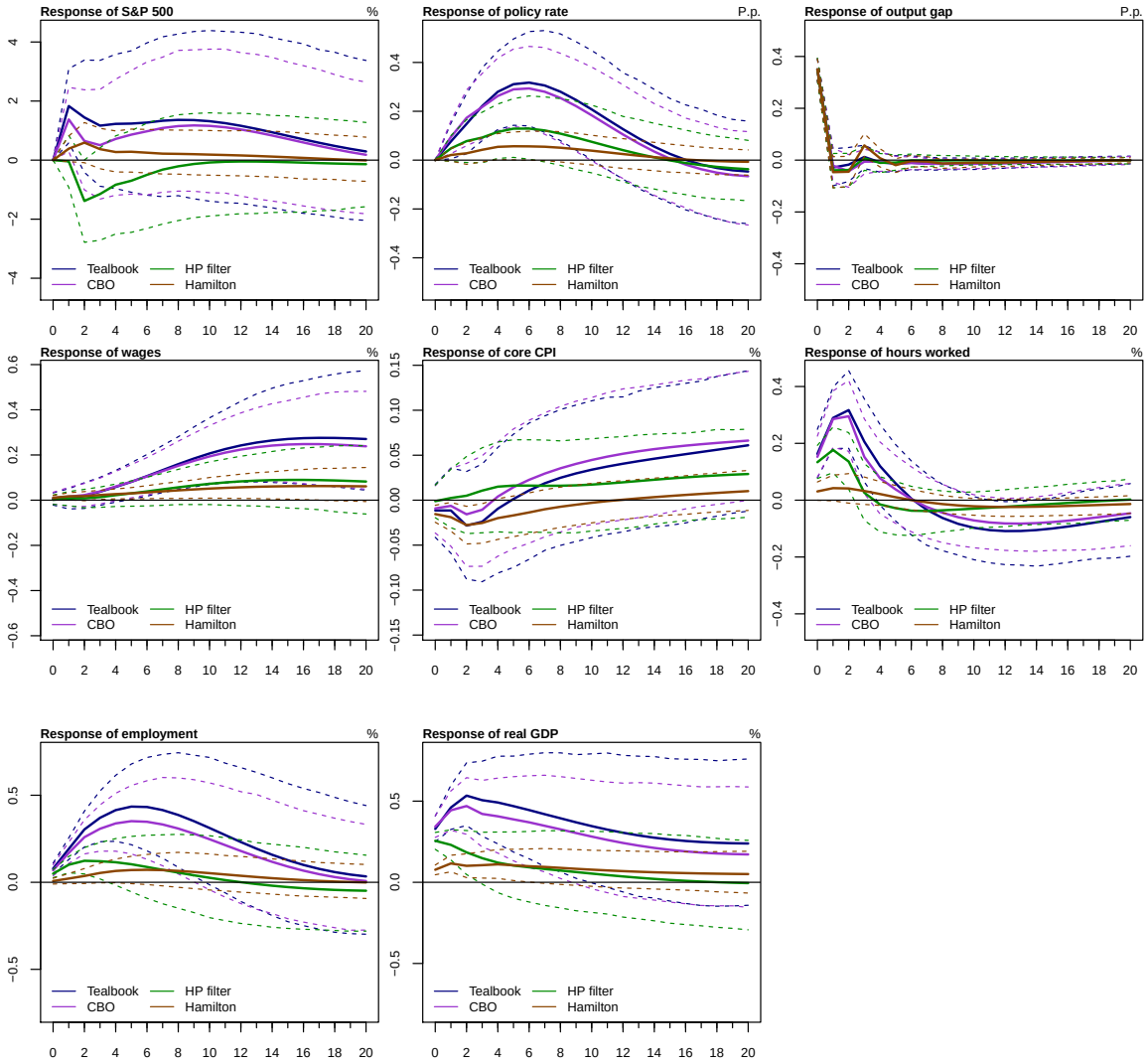


Notes: Top panel shows estimated latent factor from vintage output gap estimates from Tealbook, CBO, HP and Hamilton filter. Dark and light shaded areas are 70 and 90 percent credible intervals. Bottom panel shows the standard deviation of the common factor innovations. Gray shaded areas denote NBER recessions. Vertical dashed line indicates 2014Q4, the final Tealbook estimate. See text for details.

rate, wages, employment, and real GDP respond positively and in a significant manner to each of the innovations, and most strongly to the innovations from the Tealbook and CBO factors. Notably, core prices do not strongly respond to any factor innovation, reflecting the weak relationship between cyclical economic activity and price inflation in recent decades. The Hamilton filter-based factor innovations produce the weakest responses to the other variables. Oddly, the stock price index moves somewhat lower following innovations to the HP filter factor innovation.

The Tealbook and CBO output gap estimates incorporate a much broader range of eco-

Figure 11: Impulse response to innovation of FSV common factor, 1987Q4–2014Q4.



Notes: Figure shows the impulse responses from the quarterly eight-variable VAR containing the common factor innovation derived from either the FSV model using Tealbook, CBO, HP, or Hamilton filter output gap estimates. Sample period is 1987Q4–2014Q4. Dashed lines denote 90 percent confidence bands from from Kilian (1998) bootstrap. See text for details.

nomic information than do the univariate detrending methods. The CBO estimate of potential output is based on a Solow growth accounting model, carefully tracking labor inputs (Shackleton 2018) to different sectors of the economy. The Tealbook potential estimate is judgmental, combining estimates of potential output from growth accounting models, trend-cycle decompositions, especially those containing a Phillips curve relationship, and DSGE models (Mishkin 2007). That these more comprehensive approaches produce output gaps

that are more cyclically sensitive (in the sense that economic variables are respond strongly to the factor innovations) is consistent with other economic variables carrying information relevant to the output gap (Fleischmann & Roberts, 2011; Morley & Wong, 2020). Alternatively, it may also be that explicitly modeling slowly trending variables helps to calibrate the amount of signal taken from a given surprise to estimates of real output, see Kamber et al. (2018).

6 Discussion

This paper is an extensive empirical analysis of the Federal Reserve's output gap estimate. Output gap estimates are very uncertain, even when estimating periods well in the past. Nevertheless, Tealbook output gap estimates contain information that can be aggregated into a clear and pro-cyclical latent factor. Further, the estimates measure a genuine macroeconomic phenomenon. Innovations to the latent factor tend to be followed by increases in wages, employment, output, equity prices, and the policy rate. The output gap estimate from the Congressional Budget Office has similar desirable properties. I also find evidence that the common factor's volatility is a useful proxy for macroeconomic uncertainty, and macroeconomic uncertainty shocks lead notable declines in economic activity.

At the very least, the results strongly support the idea that the Tealbook estimates identify a genuine component to real activity. The framework outlined could be used to explore many other interesting applications, for example, the behavior of other latent economic variables such as potential output, productivity growth, and the neutral rates for unemployment and interest rates.

References

Aastveit, K. A. & Trovik, T. (2014), 'Estimating the output gap in real time: A factor model approach', *The Quarterly Review of Economics and Finance* **54**(2), 180–193.

Aguilar, O. & West, M. (2000), 'Bayesian dynamic factor models and portfolio allocation', *Journal of Business & Economic Statistics* **18**(3), 338–357.

Baker, S. R., Bloom, N. & Davis, S. J. (2016), 'Measuring Economic Policy Uncertainty', *The Quarterly Journal of Economics* **131**(4), 1593–1636.

Basu, S., Fernald, J. & Kimball, M. (2006), 'Are Technology Improvements Contractionary?', *American Economic Review* **96**(5), 1418–48.

Baumeister, C. & Peersman, G. (2013), 'Time-varying effects of oil supply shocks on the US economy', *American Economic Journal: Macroeconomics* **5**(4), 1–28.

Berge, T. J., Chang, A. C. & Sinha, N. R. (2019), Evaluating the Conditionality of Judgmental Forecasts, Finance and economics discussion series, Divisions of Research & Statistics and Monetary Affairs Federal Reserve Board, Washington, D.C.

Bloom, N. (2009), 'The Impact of Uncertainty Shocks', *Econometrica* **77**(3), 623–685.

Caldara, D., Fuentes-Albero, C., Gilchrist, S. & Zakrajšek, E. (2016), 'The macroeconomic impact of financial and uncertainty shocks', *European Economic Review* **88**(C), 185–207.

Carriero, A., Clark, T. E. & Marcellino, M. (2018*a*), Endogenous Uncertainty, Working Papers (Old Series) 1805, Federal Reserve Bank of Cleveland.

Carriero, A., Clark, T. E. & Marcellino, M. (2018*b*), 'Measuring Uncertainty and Its Impact on the Economy', *The Review of Economics and Statistics* **100**(5), 799–815.

Carter, C. K. & Kohn, R. (1994), 'On Gibbs sampling for state space models', *Biometrika* **81**(3), 541.

Champagne, J., Poulin-Bellisle, G. & Sekkel, R. (2018), 'The real-time properties of the bank of canada's staff output gap estimates', *Journal of Money, Credit and Banking* **50**(6), 1167–1188.

Chib, S., Nardari, F. & Shephard, N. (2006), 'Analysis of high dimensional multivariate stochastic volatility models', *Journal of Econometrics* **134**(2), 341 – 371.

Clark, T. E., McCracken, M. W. & Mertens, E. (2018), Modeling time-varying uncertainty of multiple horizon forecast errors, Working Paper 17-15R, Federal Reserve Bank of Cleveland.

Durbin, J. & Koopman, S. (2002), 'A simple and efficient simulation smoother for state space time series analysis', *Biometrika* **89**(3), 603–615.

Edge, R. M. & Rudd, J. B. (2016), 'Real-Time Properties of the Federal Reserve's Output Gap', *The Review of Economics and Statistics* **98**(4), 785–791.

Fernald, J. G. (2012), A quarterly, utilization-adjusted series on total factor productivity, Working Paper Series 2012-19, Federal Reserve Bank of San Francisco.

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

Garratt, A., Lee, K., Mise, E. & Shields, K. (2008), 'Real-Time Representations of the Output Gap', *The Review of Economics and Statistics* **90**(4), 792–804.

Garratt, A., Mitchell, J. & Vahey, S. (2014), 'Measuring Output Gap Nowcast Uncertainty', *International Journal of Forecasting* **30**, 268–279.

Geweke, J. (1992), Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments, *in* J. M. Bernardo, J. O. Berger, A. P. Dawid & A. F. M. Smith, eds, 'Bayesian Statistics', Vol. 4, Oxford University Press, pp. 169–193.

Gilchrist, S. & Zakrajsek, E. (2012), 'Credit Spreads and Business Cycle Fluctuations', *American Economic Review* **102**(4), 1692–1720.

Hamilton, J. D. (2003), 'What is an oil shock?', *Journal of Econometrics* **113**, 363–398.

Hamilton, J. D. (2018), 'Why you should never use the Hodrick-Prescott filter', *Review of Economics & Statistics* **100**(5), 831–843.

Jacobs, J. & van Norden, S. (2011), 'Modeling data revisions: Measurement error and dynamics of "true" values', *Journal of Econometrics* **161**(2), 101–109.

Jarocinski, M. (2015), 'A note on implementing the Durbin and Koopman simulation smoother', *Computational Statistics & Data Analysis* **91**(C), 103.

Jo, S. & Sekkel, R. (2019), 'Macroeconomic uncertainty through the lens of professional forecasters', *Journal of Business & Economic Statistics* **37**(3), 436–446.

Jurado, K., Ludvigson, S. C. & Ng, S. (2015), 'Measuring uncertainty', *American Economic Review* **105**(3), 1177–1216.

Kamber, G., Morley, J. & Wong, B. (2018), 'Intuitive and Reliable Estimates of the Output Gap from a Beveridge-Nelson Filter', *Review of Economics & Statistics* **100**(3), 550–566.

Kilian, L. (1998), 'Small-Sample Confidence Intervals For Impulse Response Functions', *The Review of Economics and Statistics* **80**(2), 218–230.

Kim, S., Shephard, N. & Chib, S. (1998), 'Stochastic volatility: Likelihood inference and comparison with arch models', *The Review of Economic Studies* **65**(3), 361–393.

Knüppel, M. (2014), 'Efficient estimation of forecast uncertainty based on recent forecast errors', *International Journal of Forecasting* **30**(2), 257–267.

Knüppel, M. (2018), 'Forecast-error-based estimation of forecast uncertainty when the horizon is increased', *International Journal of Forecasting* **34**(1), 105–116.

Messina, J. D., Sinclair, T. & Stekler, H. (2015), 'What can we learn from revisions to the greenbook forecasts?', *Journal of Macroeconomics* **45**(C), 54–62.

Mishkin, F. S. (2007), Estimating potential output. Speech at the Conference on Price Measurement for Monetary Policy, Federal Reserve Bank of Dallas, Dallas, Texas. Available at `https://www.federalreserve.gov/newsevents/speech/mishkin20070524a.htm`.

Morley, J. & Wong, B. (2020), 'Estimating and accounting for the output gap with large bayesian vector autoregressions', *Journal of Applied Econometrics* **35**(1), 1–18.

Mueller, U. & Watson, M. (2017), 'Low-frequency econometrics', *Advances in Economics and Econometrics: Eleventh World Congress of the Econometric Society* **2**, 53–94.

Omori, Y., Chib, S., Nakajima, J. & Shephard, N. (2007), 'Stochastic volatility with leverage: fast and efficient likelihood inference', *Journal of Econometrics* **140**, 425–449.

Orphanides, A. (1998), Monetary policy evaluation with noisy information, Finance and Economics Discussion Series 1998-50, Federal Reserve Board.

Orphanides, A. (2019), Monetary policy strategy and its communication, *in* 'Challenges for Monetary Policy', Jackson Hole Economic Policy Symposium, Federal Reserve Bank of Kansas City.

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

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

Pitt, M. K. & Shephard, N. (1999), 'Time-varying covariances: A factor stochastic volatility approach', *Bayesian Statistics* **6**, 547–570.

Primiceri, G. E. (2005), 'Time varying structural vector autoregressions and monetary policy', *Review of Economic Studies* **72**, 821–852.

Ramey, V. A. (2011), 'Identifying Government Spending Shocks: It's All in the Timing', *Quarterly Journal of Economics* **126**, 1–50.

Reifschneider, D. L., Stockton, D. J. & Wilcox, D. W. (1997), 'Econometric models and the monetary policy process', *Carnegie-Rochester Conference Series on Public Policy* **47**(1), 1–37.

Reifschneider, D. & Tulip, P. (2017), Gauging the Uncertainty of the Economic Outlook Using Historical Forecasting Errors: The Federal Reserve's Approach, Finance and Economics Discussion Series Divisions of Research & Statistics and Monetary Affairs 020, Federal Reserve Board.

Romer, C. D. & Romer, D. H. (2004), 'A new measure of monetary shocks: Derivation and implications', *American Economic Review* **94**(4), 1055–1084.

Scotti, C. (2016), 'Surprise and uncertainty indexes: Real-time aggregation of real-activity macro-surprises', *Journal of Monetary Economics* **82**(C), 1–19.

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

Stock, J. H. & Watson, M. W. (2007), 'Why has u.s. inflation become harder to forecast?', *Journal of Money, Credit and Banking* **39**(1).

Stock, J. H. & Watson, M. W. (2012), 'Disentangling the Channels of the 2007-09 Recession', *Brookings Papers on Economic Activity* **43**(1 (Spring)), 81–156.

Stock, J. & Watson, M. W. (2008), Phillips curve inflation forecasts, Working Paper 14322, NBER.

Wu, J. C. & Xia, F. D. (2016), 'Measuring the Macroeconomic Impact of Monetary Policy at the Zero Lower Bound', *Journal of Money, Credit & Banking* **48**(2-3), 253–291.

Appendix

A1 Data sources

A1.1 Tealbook output gap estimates

I have combined two data sources of Tealbook output gap estimates.

- Early output gap nowcasts were provided by Orphanides (1998). They are available in the supplemental material of Edge & Rudd (2016) at `https://www.mitpressjournals.org/doi/abs/1.1162/REST_a_00555`.
- Tealbook output gap estimates are available from the website of the Federal Reserve Bank of Philadelphia, `https://www.philadelphiafed.org/research-and-data/real-time-cente greenbook-data/gap-and-financial-data-set`.

Table A1: Summary of available Tealbook vintage output gap estimates.

First available *nowcast* vintage	January 1951
First vintage with near-term path	August 1987
First vintage containing complete path	May 1996
Most recent vintage	December 2014

A1.2 Real-time vintage GDP estimates

Real-time real and nominal GDP data is from the Federal Reserve Bank of Philadelphia.

- I use quarterly real-time estimates. Each quarterly vintage is the estimate available in the middle month of each quarter. One vintage, 1996Q1, was not available in real-time as a government shutdown delayed the release of NIPA data.
- Vintage estimates of real GDP are used to compute output gaps using trend-cycle decompositions. Vintages run from 1984Q1 to 2019Q4, a total of 144 vintages, and for each vintage, I retain the estimate of real GDP starting from 1960Q1.

- Vintage estimates of nominal GDP are used to compute output gaps implied by the CBO's estimate of nominal potential output, see appendix A3 for details.

- The data are available at: `https://www.philadelphiafed.org/research-and-data/real-time-center/real-time-data`.

A1.3 Macroeconomic shocks

The macroeconomic shocks described in section 5 are defined in table A2. The text below provides additional details.

Total factor productivity. Innovations to productivity are measured using the series provided by Fernald (2012), which is a utilization-adjusted estimate of the measure developed in Basu, Fernald & Kimball (2006).

Oil price shock. I compute the net oil price increase of Hamilton (2003), constructed over a 12-quarter horizon. The net oil increase is calculated as the percentage by which the oil price, measured as the quarterly producer price index for oil, exceeds the previous peak over the past 3 years.

Fiscal policy shock. I use the Ramey (2011) military spending news shock as a proxy for fiscal policy spending shocks. The series ends in 2013Q3, and is available from Prof. Ramey's website, at: `https://econweb.ucsd.edu/~vramey/research.html#govt`.

Monetary policy shock. I extend the Romer & Romer (2004) measure of monetary policy shock. Specifically, I estimate the regression of Romer & Romer over non zero-lower bound period, 1969–2008, and using the regression specification as described in Romer & Romer (2004).[1] To extend the series to the period with a zero-lower bound, I proxy the intended change in the federal funds rate with the realized change to the Wu & Xia (2016) shadow rate, and proxy the lagged value of the federal funds rate with the lagged value of the shadow rate. With these proxies, one can compute the fitted values of the regression using

[1] I obtain Greenbook forecasts from the website of the Federal Reserve Bank of Philadelphia, `https://www.philadelphiafed.org/research-and-data/real-time-center/greenbook-data/`.

the regression coefficients estimated using the pre-zero lower bound period, and the monetary policy shock for the zero-lower bound period is the residual. To obtain the quarterly series, I sum all monetary policy shocks within a quarter.

Policy uncertainty. Policy uncertainty is taken from Baker et al. (2016), a measure of policy uncertainty based on news media references to uncertainty. Quarterly values are obtained by summing monthly values.

Measures of financial uncertainty. I use two measures of financial uncertainty. The first is realized volatility, constructed from daily returns of the S&P 500 index. I also consider the excess bond premium from Gilchrist & Zakrajsek (2012), a bond premium that has been adjusted for predictable default risk. The data can be found at: `http://people.bu.edu/sgilchri/Data/`.

Table A2: Data sources for macroeconomic shocks.

Shock	Period available	Citation	Source
Productivity	1984Q1–2014Q4	Fernald (2012)	`http://www.johnfernald.net/TFP`
Oil price	1984Q1–2014Q4	Hamilton (2003)	Author's calculations
Fiscal policy	1984Q1–2013Q4	Ramey (2011)	Prof. Ramey's website
Monetary policy	1984Q1–2014Q4	RR (2004)	Author's calculations
Policy uncertainty	1984Q1–2014Q4	BBD (2016)	`https://www.policyuncertainty.com`
Financial uncertainty			
Realized volatility	1984Q1–2014Q4	Yahoo	`https://finance.yahoo.com/`
Ex. bond premium	1984Q1–2014Q4	GZ (2012)	Prof. Gilchrist's webpage

Notes: RR (2004) is Romer & Romer (2004); BBD (2016) is Baker et al. (2016); GZ (2012) is Gilchrist & Zakrajsek (2012). See text for details.

A1.4 Data used in VAR

The primary source for the data used in the VAR is FRED, with the exceptions of the S&P 500 index and the Wu & Xia (2016) shadow rate, which are taken from Yahoo and the Federal Reserve Bank of Atlanta, respectively.

Table A3: Data sources for macroeconomic data used in VAR.

Variable	Transformation	Period available	Source
S&P 500	Log	1950Q1–2019Q1	Yahoo (ˆGSPC)
Wu-Xia shadow rate	–	1960Q1–2019Q4	FRB-Atlanta
Real consumption	Log	1947Q1–2019Q4	FRED (PCECC96)
Real private domestic investment	Log	1947Q1–2019Q4	FRED (GPDIC1)
Real GDP	Log	1947Q1–2019Q4	FRED (GDPC1)
GDP deflator	Log	1947Q1–2019Q4	FRED (GDPDEF)
Average hourly earnings	Log	1964Q1–2019Q4	FRED (AHETPI)
Core CPI	Log	1957Q1–2019Q4	FRED (CPILFESL)
Average weekly hours, man.	Log	1939Q1–2019Q4	FRED (AWHMAN)
Total nonfarm payrolls	Log	1939Q1–2019Q4	FRED (PAYEMS)
Industrial production	Log	1919Q1–2019Q4	FRED (INDPRO)

Notes: Wu-Xia shadow rate available from Federal Reserve Bank of Atlanta: `https://www.frbatlanta.org/cqer/research/shadow_rate.aspx`. Mnemonics in parentheses denote Yahoo or FRED mnemonic.

A2 Comparison to results from Orphanides & van Norden (2002)

Here I reconsider the output gap estimates from Orphanides & van Norden (2002): a linear trend; a linear trend with break; a quadratic trend; and the HP filter.[2] I add two more recent detrending methods from Mueller & Watson (2017) and Hamilton (2018). The models are estimated on quarterly real-time vintages of 100 times the log of real GDP. The estimation period begins in 1960Q1 and continues through the vintage-specific nowcast quarter. Since these simple models are not dynamic, only backcasts and nowcasts can be considered. As before, ex-post revisions are calculated using vintage estimates 12 quarters apart.

Table A4 displays the noise-signal ratios for these filters. The Tealbook estimates tend to be more stable than the estimates from these simple models, with the exception of the Hamilton (2018) method, which produces output gaps with small noise-signal ratios. Unsurprisingly, for most models, noise-signal ratios become smaller as the forecast horizon moves further into the past. For example, the HP filter has a noise-signal ratio above one for the nowcasts but just .2 for estimates 16 quarters in the past. This pattern of noise-signal ratio by forecast horizon is typical, although some models have larger noise-signal ratios on average than others. Even the four-quarter backcasts from many of the univariate models have large noise-signal ratios: when the noise-signal ratio is calculated using the RMSE of the revision, the quadratic trend and Mueller-Watson estimates both have noise-signal ratios above .8. In all, the noise-signal ratios presented here are smaller than the ones found in Orphanides & van Norden (2002) and Edge & Rudd (2016), likely reflecting to the different sample periods used.

[2] For the broken trend model, I follow Edge & Rudd (2016) and assume there are two breaks, first in 1973 and again in 1997. I allow the breaks to enter the models after a three year delay.

Table A4: Noise-signal ratios of output gap estimates from univariate detrending methods.

	Forecast horizon				
	-16	-12	-8	-4	0
Linear trend					
Using SD	.25	.26	.25	.26	.34
Using RMSE	.44	.44	.44	.42	.44
Broken linear trend					
Using SD	.30	.30	.29	.28	.35
Using RMSE	.52	.54	.54	.52	.53
Quadratic trend					
Using SD	.59	.68	.74	.75	.76
Using RMSE	.59	.68	.76	.81	.89
HP					
Using SD	.22	.23	.31	.62	1.11
Using RMSE	.22	.24	.31	.62	1.11
Hamilton					
Using SD	.24	.24	.24	.25	.31
Using RMSE	.24	.25	.27	.30	.38
Mueller-Watson					
Using SD	.68	.63	.51	.82	1.12
Using RMSE	.68	.71	.77	.83	1.92
Memo: Tealbook					
Using SD	.23	.25	.31	.31	.44
Using RMSE	.24	.26	.33	.31	.43

Notes: Table presents noise-signal ratios of output gap revisions at selected forecast horizon. Revisions calculated setting $\tilde{v} = v + i$ and $i = 12$ quarters. Sample period is 1984–2014. See text for details.

A3 Application to CBO output gap estimates

In this section, I evaluate the output gap estimates implied by the CBO's estimate of potential output. Real-time vintages for CBO estimates of potential are scarce relative to those from the Federal Reserve, although a more recent vintage history is available since the estimates of potential are released to the public contemporaneously.

A3.1 Data source

I compile real-time vintages of CBO's estimate of the output gap by comparing real-time estimates of nominal potential GDP to real-time vintages of nominal GDP.[3] The CBO typically updates its estimate of potential output twice per year. I compute quarterly real-time vintages of the output gap by comparing vintage estimates of nominal GDP to the most recently available CBO estimate of potential output.[4] Notice that this implies that there are no output gap forecasts, because although the CBO estimate of potential extends into the future, realized output does not. Thus, when possible, I compute output gap forecasts from the CBO using their projections for nominal GDP growth from their 10-year economic projections. Real-time vintages of these economic projections are more limited, beginning only in August 2011. The CBO vintage data is summarized in table A5.

A3.2 CBO output gap estimate revisions

Figure A1 and table A6 and replicate the results described in section 2 of the papaer for the Tealbook output gap estimates. Specifically, the table provides summary statistics of

[3] Real-time vintages of nominal GDP are obtained from the Federal Reserve Bank of Philadelphia; see appendix A1.2 for details. I choose to compute the output gap using nominal potential output and nominal GDP in order to sidestep the issue of changes in the base year of the price index used to compute real GDP. There is a lag between the time that real GDP is re-based to a new chain index and the CBO estimate of real potential output is re-based, since the CBO estimate is updated only twice per year.

[4] In 2013, there was only one estimate of potential output released by the CBO, produced in February of that year. To account for a change in the level of nominal GDP due to the addition of intellectual property products in the July 2013 BEA annual revision, I created a pseudo vintage of potential output for August 2013 by re-basing the February 2013 estimate of potential to a level consistent with the updated definition of GDP.

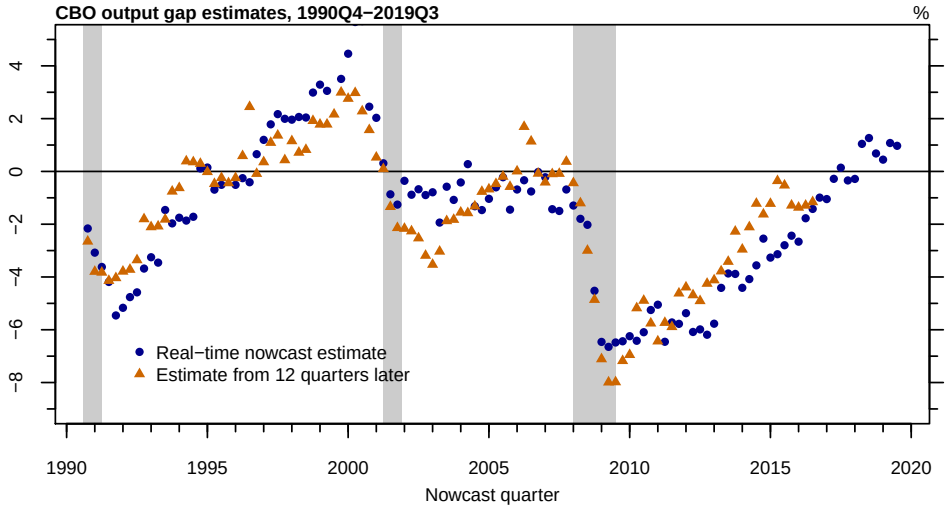
Table A5: Summary of CBO vintages of potential output.

First vintage of potential output	February 1991
First vintage containing projection of GDP	August 2011
Most recent vintage	November 2019

Notes: CBO vintage data available at `https://www.cbo.gov/data/budget-economic-data\#6`. See text for details.

revisions to the CBO output gap estimates across vintages, comparing estimates 12 quarters apart. Notice, however, because vintages of the CBO's 10-year projection begin only in August 2011, the number of forecasts available is quite small and limited to the post-2011 period.

Figure A1: CBO output gap estimates, real-time nowcast and 12 quarters later.



Notes: CBO output gap nowcasts (h=0) alongside estimate of same quarter produced 12 quarters hence. Sample period 1990Q4–2019Q3. Shaded regions denote NBER-defined recessions. See text for details.

With the caveat that they cover slightly different sample periods, and that the underlying estimate of potential output from the CBO is updated less frequently than the estimate of potential output from the Tealbook, the magnitude of the revisions to the CBO output gap estimate are a bit larger than those from the Tealbook. For example, the standard deviation of revisions to the CBO output gap nowcast is about 40 percent larger than the standard deviation of Tealbook nowcast revisions. Backcasts are similarly more volatile, and while

the summary statistics for forecasts are quite different, CBO forecasts are limited to the post-2011 period. With the exception of the forecasts, the noise-signal ratios are somewhat larger for the CBO estimate than the Tealbook. The CBO estimate is more stable than most of the estimates produced by univariate models, table A4.

Table A6: Summary statistics of revisions to CBO output gap estimate.

	Forecast horizon, quarters						
	-16	-12	-8	-4	0	4	8
N. obs	104	104	104	104	103	22	22
Mean	.28	.32	.38	.23	.09	1.29	1.11
Std dev	.90	.92	.99	1.12	1.38	0.98	1.16
Min	-2.30	-2.79	-2.97	-3.91	-3.87	.17	-0.48
Max	2.84	2.48	2.18	2.07	2.86	3.55	3.53
RMSE	.94	.97	1.06	1.14	1.38	1.61	1.58
Noise-signal ratio							
Using SD	.35	.36	.39	.44	.54	.39	.45
Using RMSE	.37	.38	.42	.45	.54	.63	.62

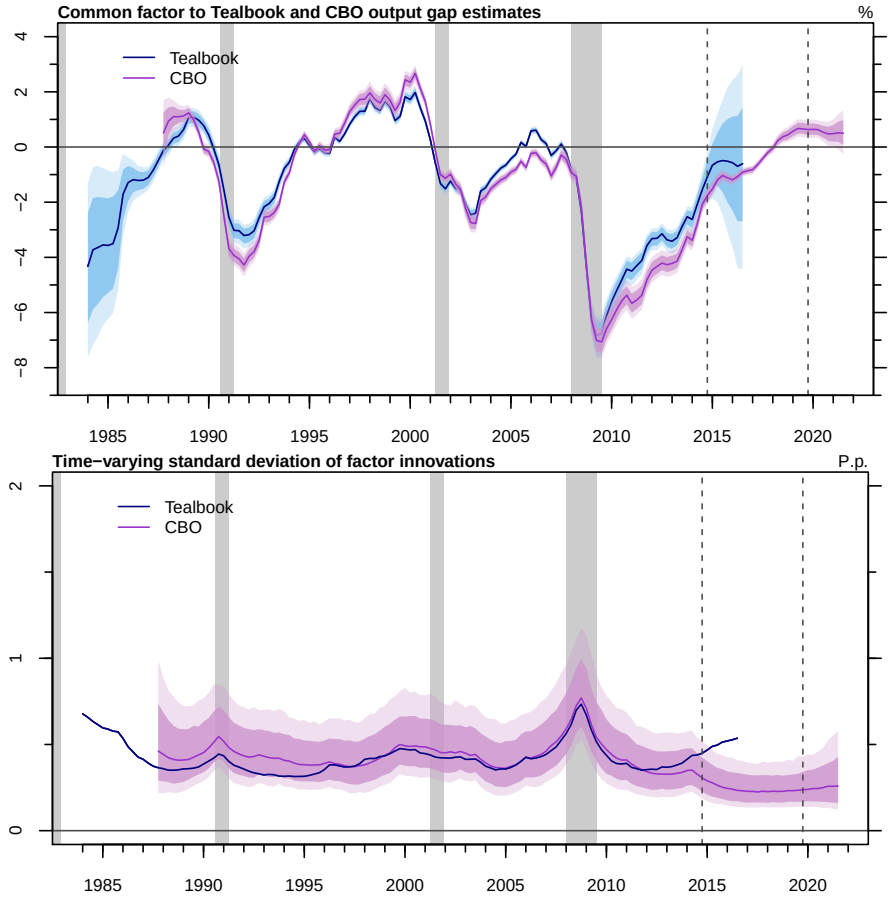
Notes: Summary statistics of output gap revisions at selected forecast horizon. Revisions calculated setting $\tilde{v} = v + 12$ quarters. Vintages range from February 1991–November 2019, missing observations as described in text. See text for details.

A3.3 FSV model results

The common factor and the standard deviation to its revisions are shown in figure A2. Because the vintage data begin in 1991Q1 and end in 2019Q4, the full sample period covers 1987Q4–2021Q3, using B=12 and H=8 as before. For reference, the figure includes the same objects from the baseline Tealbook FSV model.

As expected, the common factor is pro-cyclical, and indeed it is quite close to the factor from the baseline estimation. As with the Tealbook-based factor, the CBO factor implies quite slow recoveries from the three recessions in the sample; indeed, the CBO-based factor never comes above zero between the 2001-2 and 2007-8 recessions. Following the Great Recession, the factor turns positive only in 2018Q2. The two estimates of the output gap also have very similar time-variation in the volatility of their innovations, as shown in the bottom panel.

Figure A2: Common factor to CBO output gap estimates and its volatility.

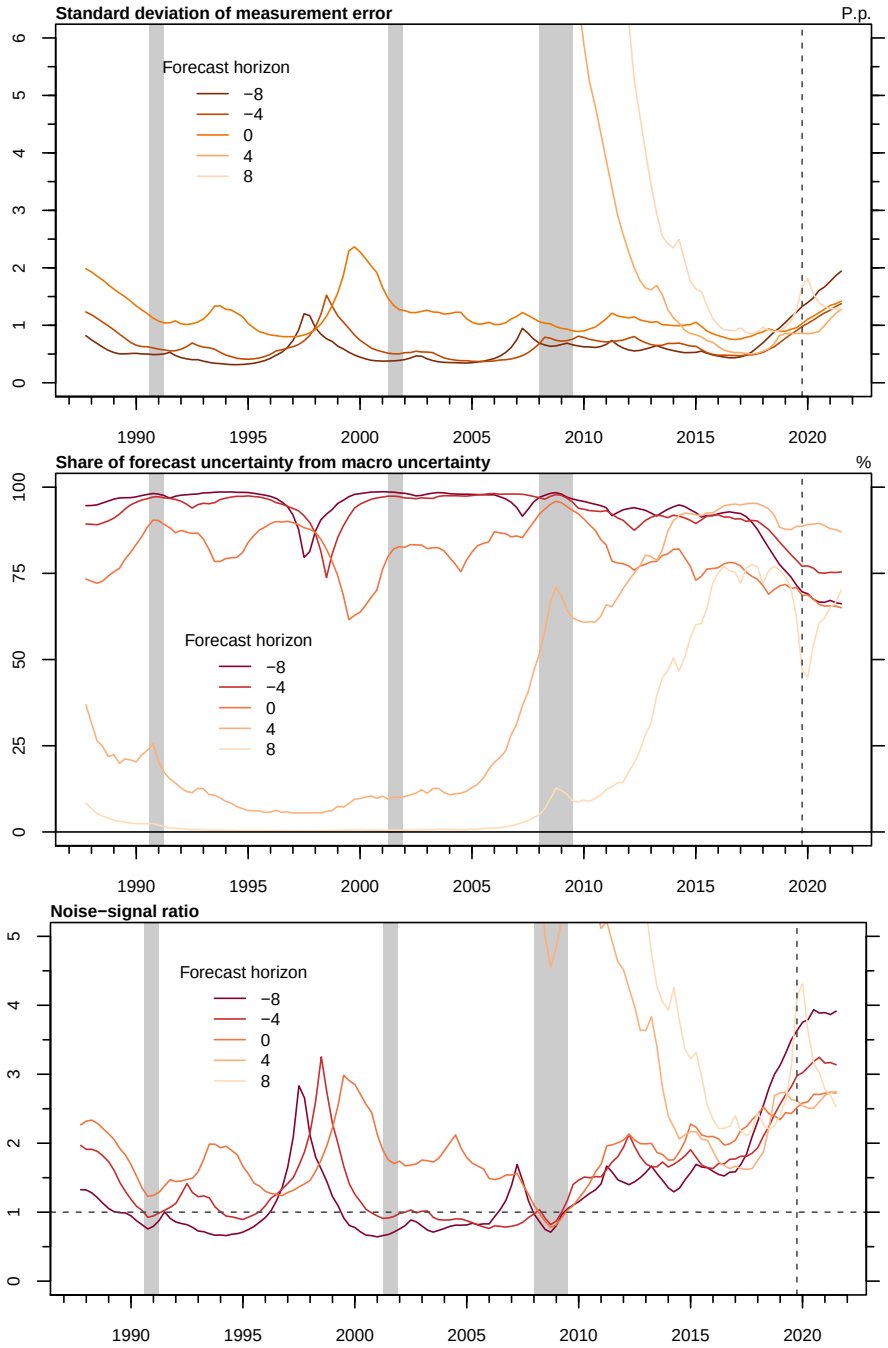


Notes: FSV estimates of common factor and its time-varying volatility, Tealbook and CBO output gap estimates. FSV models estimated separately. Tealbook sample period is 1984Q1–2016Q3; CBO sample period 1987Q3–2021Q3. Vertical dashed lines denote final available vintage of Tealbook and CBO estimates. Solid lines show the posterior median estimate. The dark and light shaded areas are the 70% and 90% posterior credible sets (credible sets of volatility of Tealbook factor innovations suppressed in the interest of legibility.) Gray shaded regions denote NBER recessions. See text for details.

Figure A3 displays the measures of uncertainty for the CBO output gap estimates. For easy comparison to the equivalent figures from the Tealbook estimates, the vertical axes are the same as figure 3. Overall, the nowcast and backcast output gap estimates from the CBO are a bit less precisely estimated than those from the Tealbook. For example, the standard deviation of the CBO output gap nowcast is usually higher than that from the Tealbook nowcast, the average value for the CBO in the period 1987-2014 is 1.1, compared to 0.9 for the Tealbook. Similarly, the horizon-specific noise-signal ratios for the CBO output gap

nowcasts tend to be higher than those from the Tealbook-based FSV model.

Figure A3: CBO output gap estimate uncertainty.

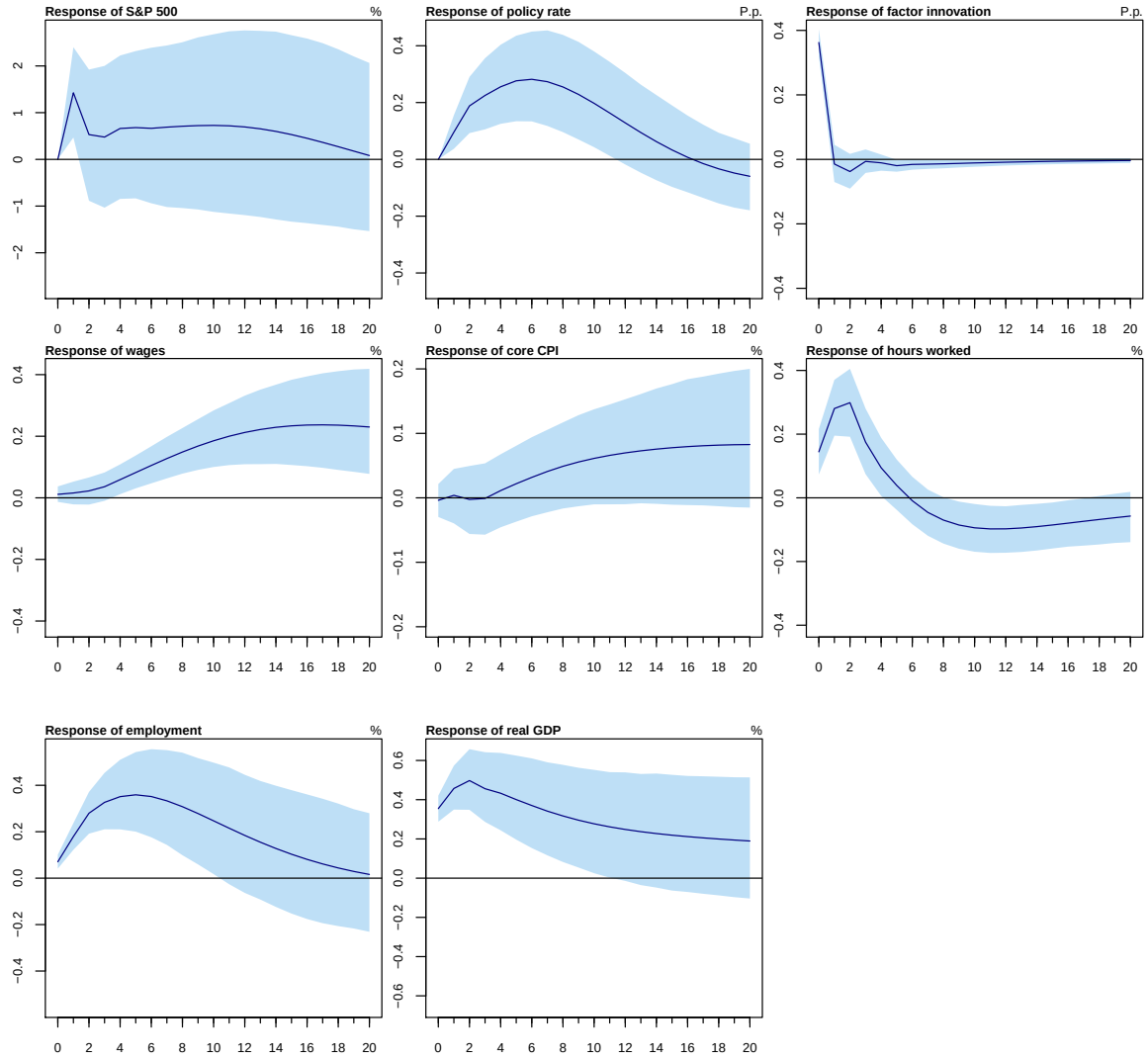


Notes: Sample period is 1987Q4–2021Q3, with most recent CBO vintage produced in November 2019, vertical dashed line. Top panel: standard deviation of measurement errors ($\sqrt{e^{h_{it}}/(1-\rho_i^2)}$). Middle panel: share of variance explained by common factor's uncertainty ($\lambda_i^2 var(f_t)/var(\hat{y}_{it}) \times 100$). Bottom panel: ratio of the standard deviation of measurement error innovation to that of the common factor ($e^{\frac{1}{2}h_{it}}/e^{\frac{1}{2}h_{ft}}$). Gray shaded regions denote NBER recessions. See text for details.

Finally, figure A4 shows the VAR exercise using the innovations to the common factor to the CBO output gap estimates. The impulse responses are very similar to the responses to innovations to the Tealbook-based factor. The response of stock prices, hours worked, employment, and real GDP are somewhat smaller than the responses recovered using the Tealbook common factor. Interestingly, however, core prices respond somewhat more strongly to the CBO-implied factor innovation than to the Tealbook-based innovations. Core prices show a weak but positive response to the CBO-implied common factor.

In sum, while the datasets cover slightly different periods, both the CBO and Tealbook output gap estimates are on the whole quite similar. The common factors extracted from each set of data are very similar. Although the uncertainty about Tealbook output gap estimates is somewhat smaller than the uncertainty of CBO estimates, the two sets of estimates appear to have very similar characteristics, in the sense that the responses of other macroeconomic variables to factor innovations are very similar.

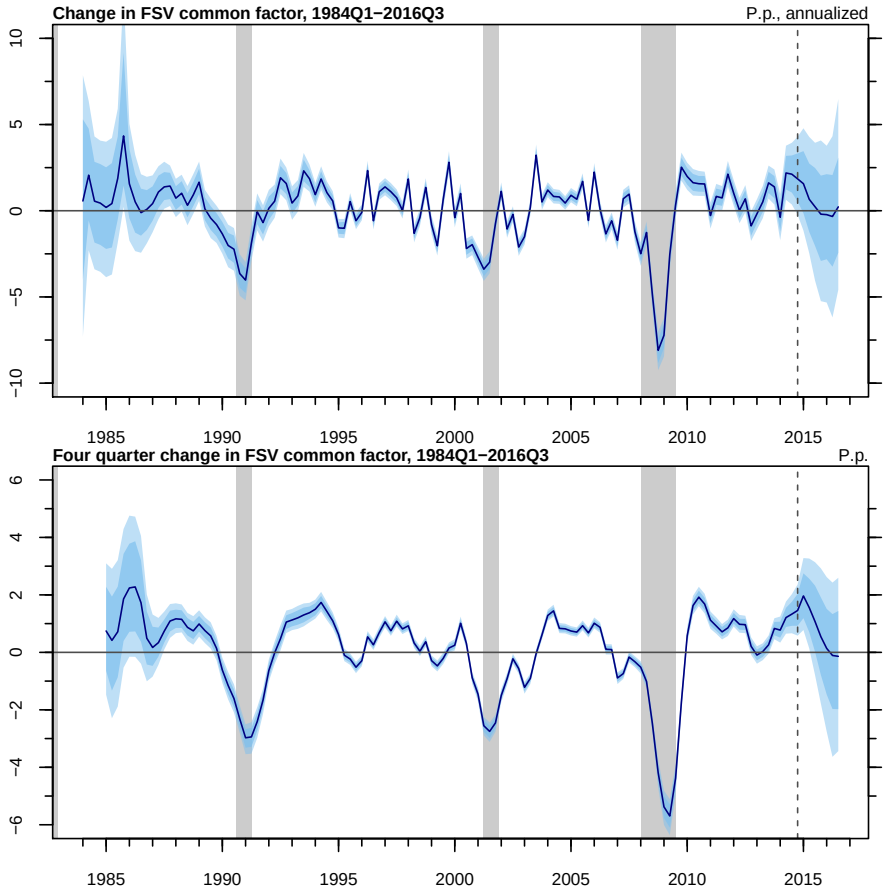
Figure A4: Response to CBO factor innovation, VAR-8.



Notes: Impulse response from eight variable VAR. Sample period 1987Q4–2019Q4. Figures show response to a one standard deviation innovation to common factor (roughly .35 point). Shaded area denotes 90 percent bootstrapped confidence interval. See text for details.

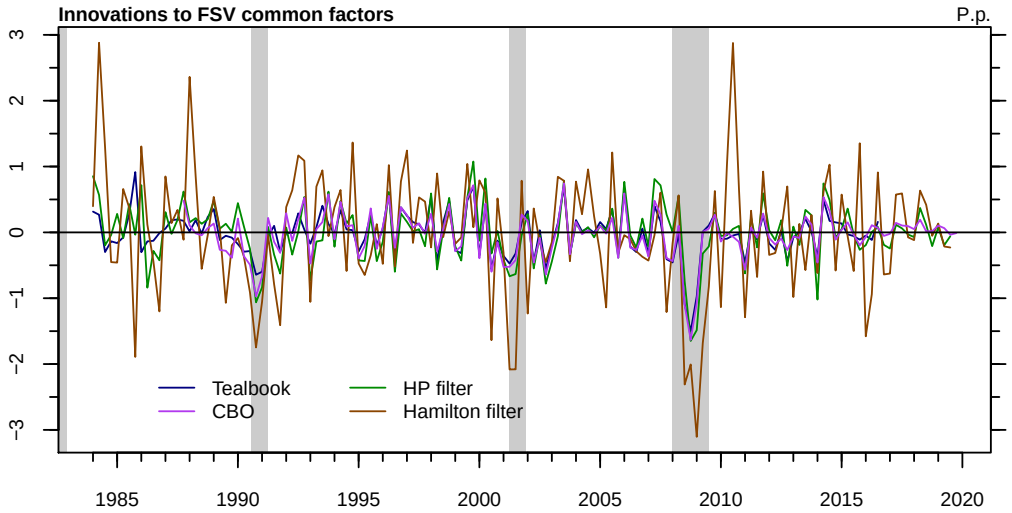
A4 Additional results

Figure A5: Change in Tealbook common factor.



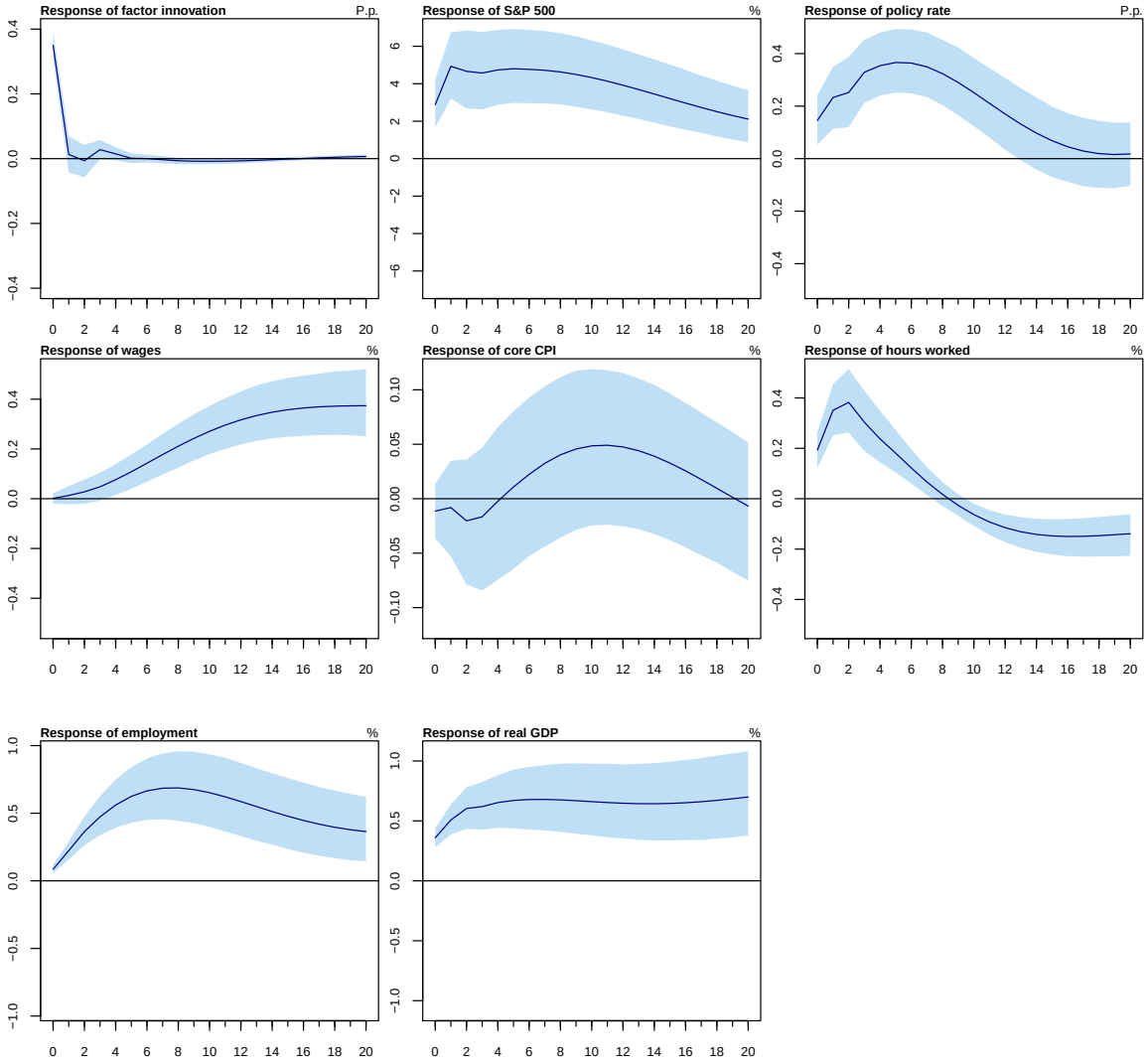
Notes: Top figure: quarterly change in FSV common factor, annualized. Bottom panel: four-quarter change of common factor. Dashed vertical line indicates October 2014 Tealbook, the final Tealbook in the sample. Dark and light blue shaded areas are 70% and 90% posterior credible sets. Gray shaded areas denote NBER-defined recessions.

Figure A6: Innovation to common factor estimates.



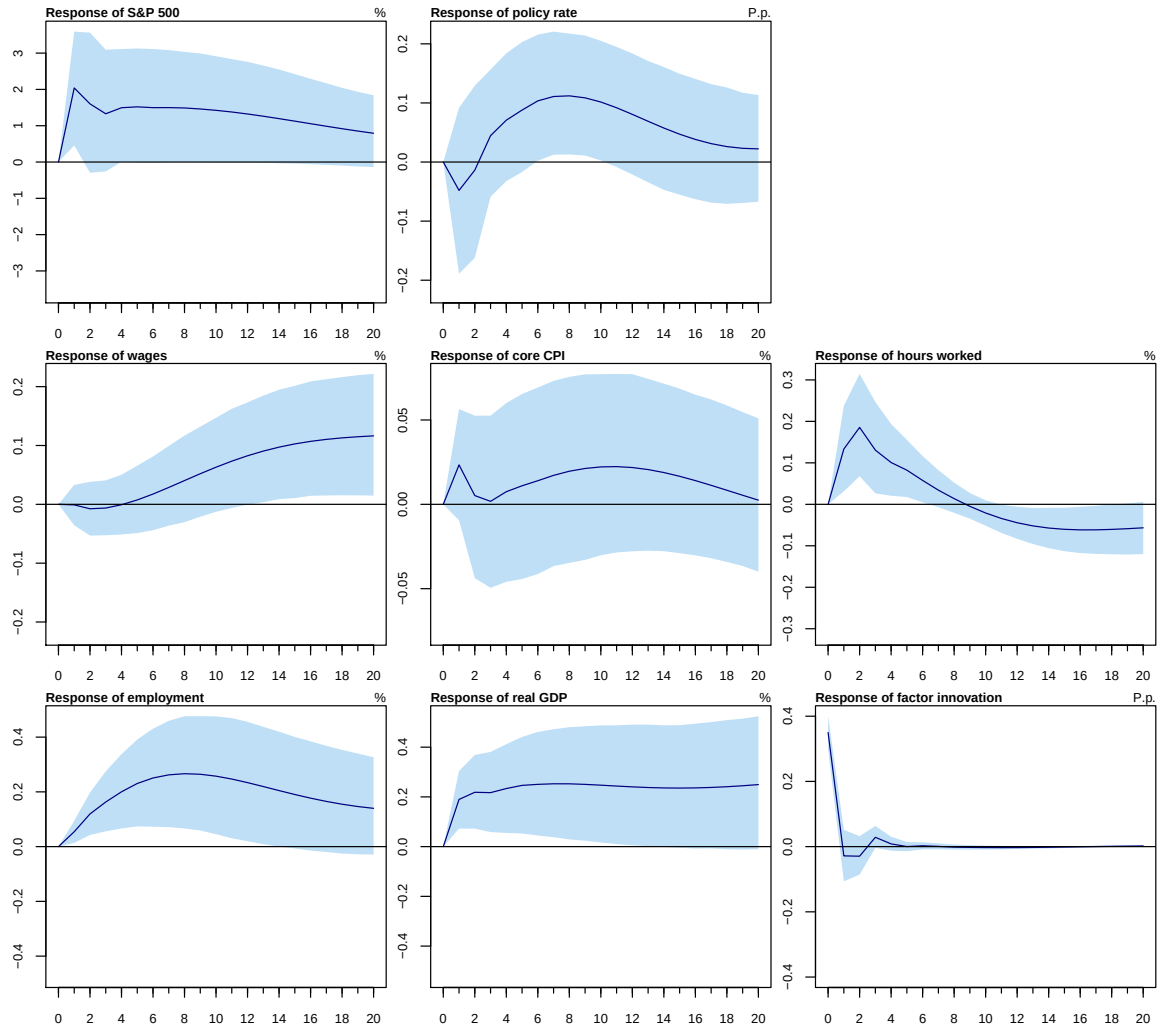
Notes: Figure shows model-implied innovations to common factors from Tealbook-, HP filter-, and Hamilton filter-based FSV models. Sample period 1984Q1–2019Q4. See text for details.

Figure A7: Response to Tealbook common factor innovation, VAR-8, innovation ordered first.



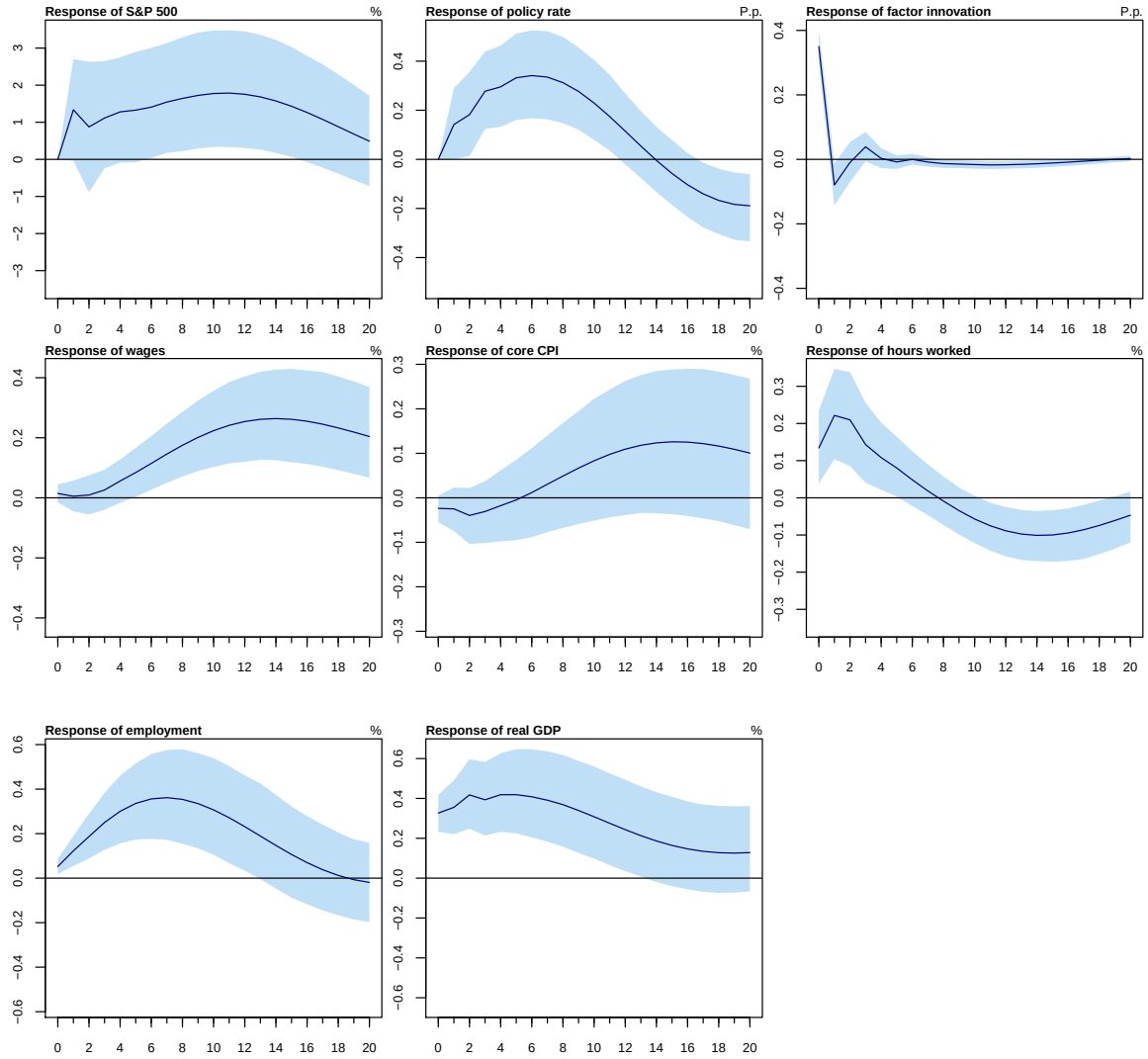
Notes: Impulse response from eight variable VAR; factor innovation ordered first. Sample period 1984Q1–2014Q4. Shaded area denotes 90 percent bootstrapped confidence interval. See text for details.

Figure A8: Response to Tealbook common factor innovation, VAR-8, innovation ordered last.



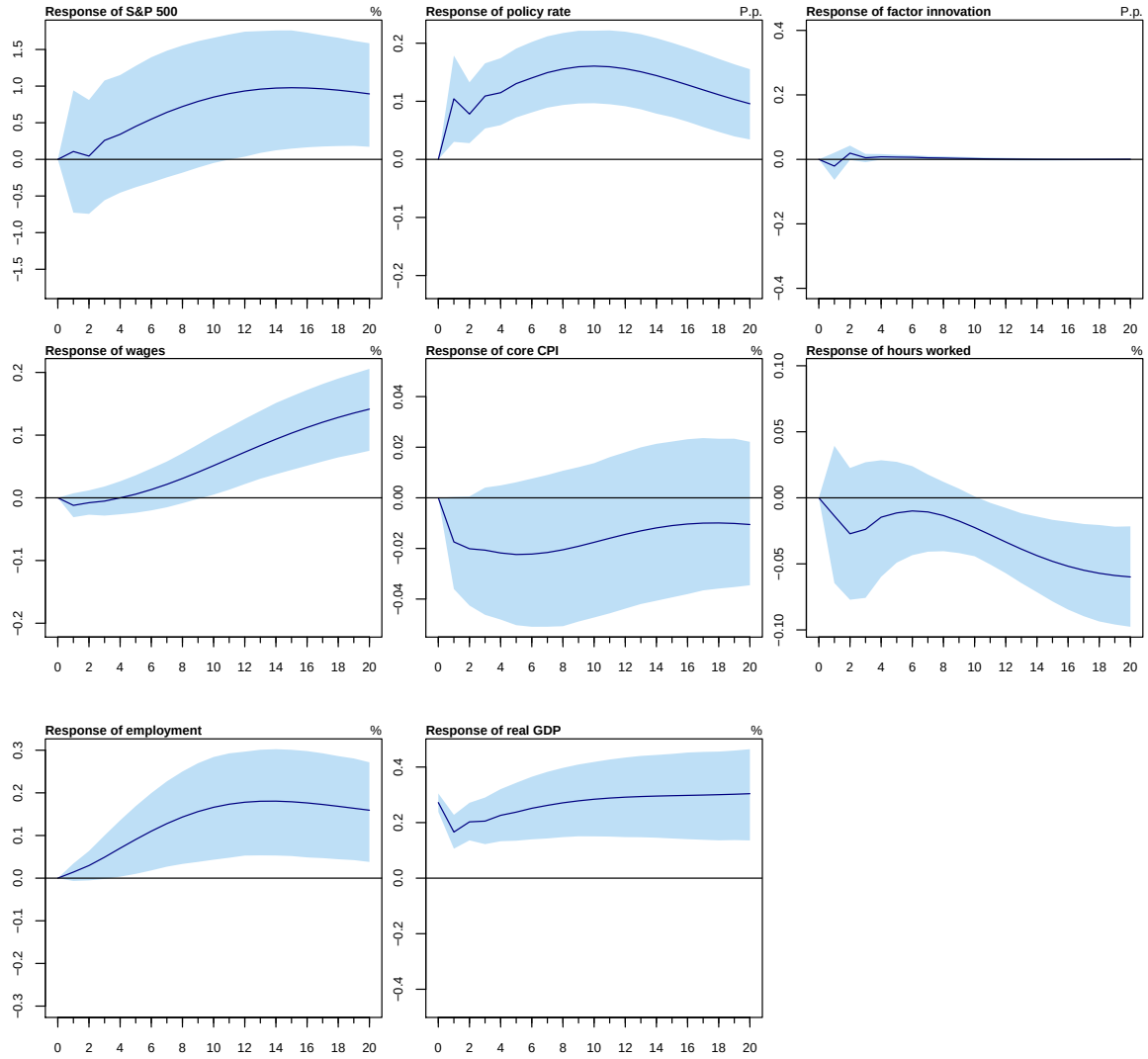
Notes: Impulse response from eight variable VAR, factor innovation ordered last. Sample period 1984Q1–2014Q4. Shaded area denotes 90 percent bootstrapped confidence interval. See text for details.

Figure A9: Response to Tealbook common factor innovation, VAR-8, short sample period.



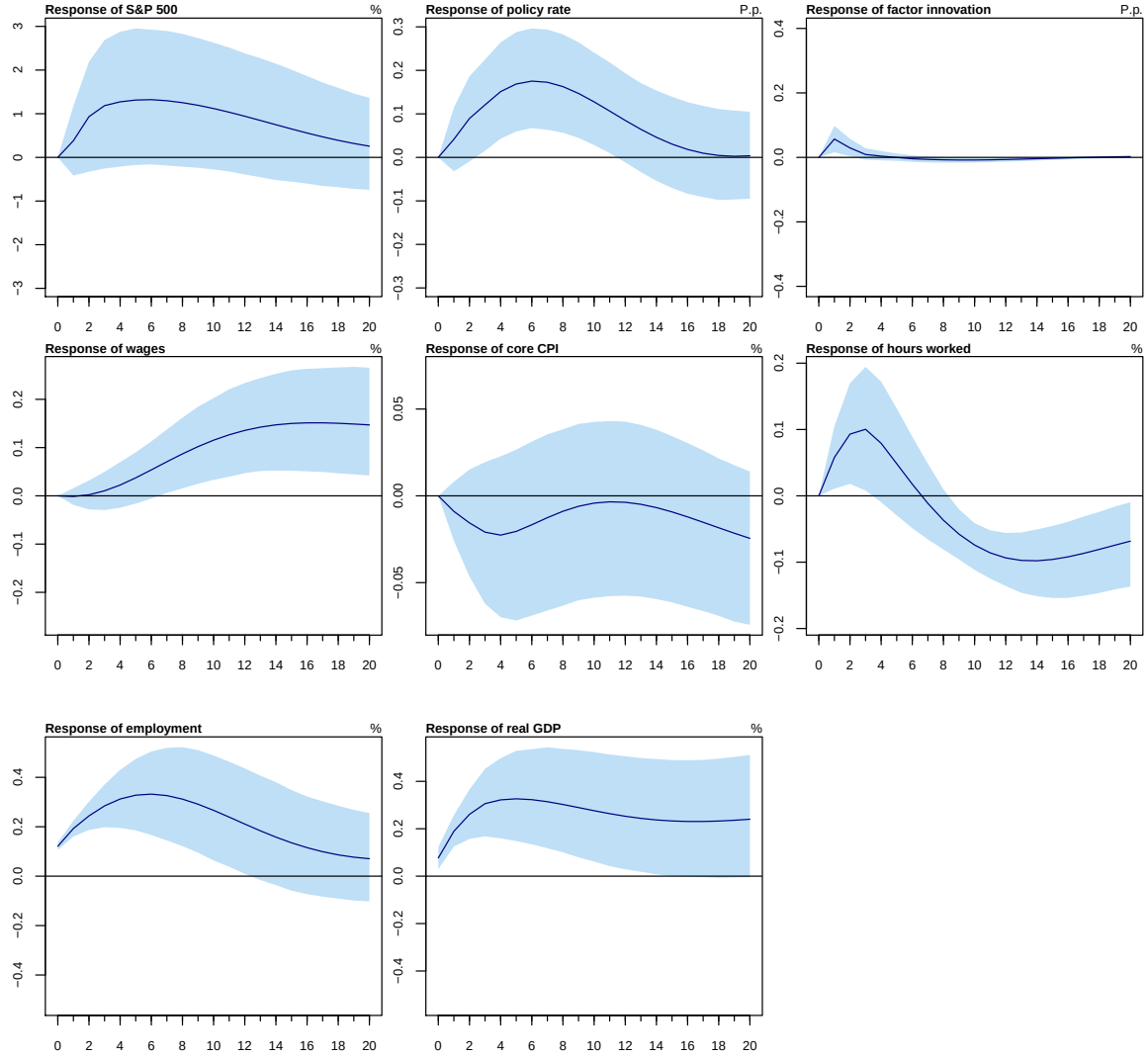
Notes: Impulse response from eight variable VAR. Sample period 1984Q1–2006Q4. Figures show response to factor innovation of 0.35 point. Shaded area denotes 90 percent bootstrapped confidence interval. See text for details.

Figure A10: Response to GDP innovation, VAR-8.



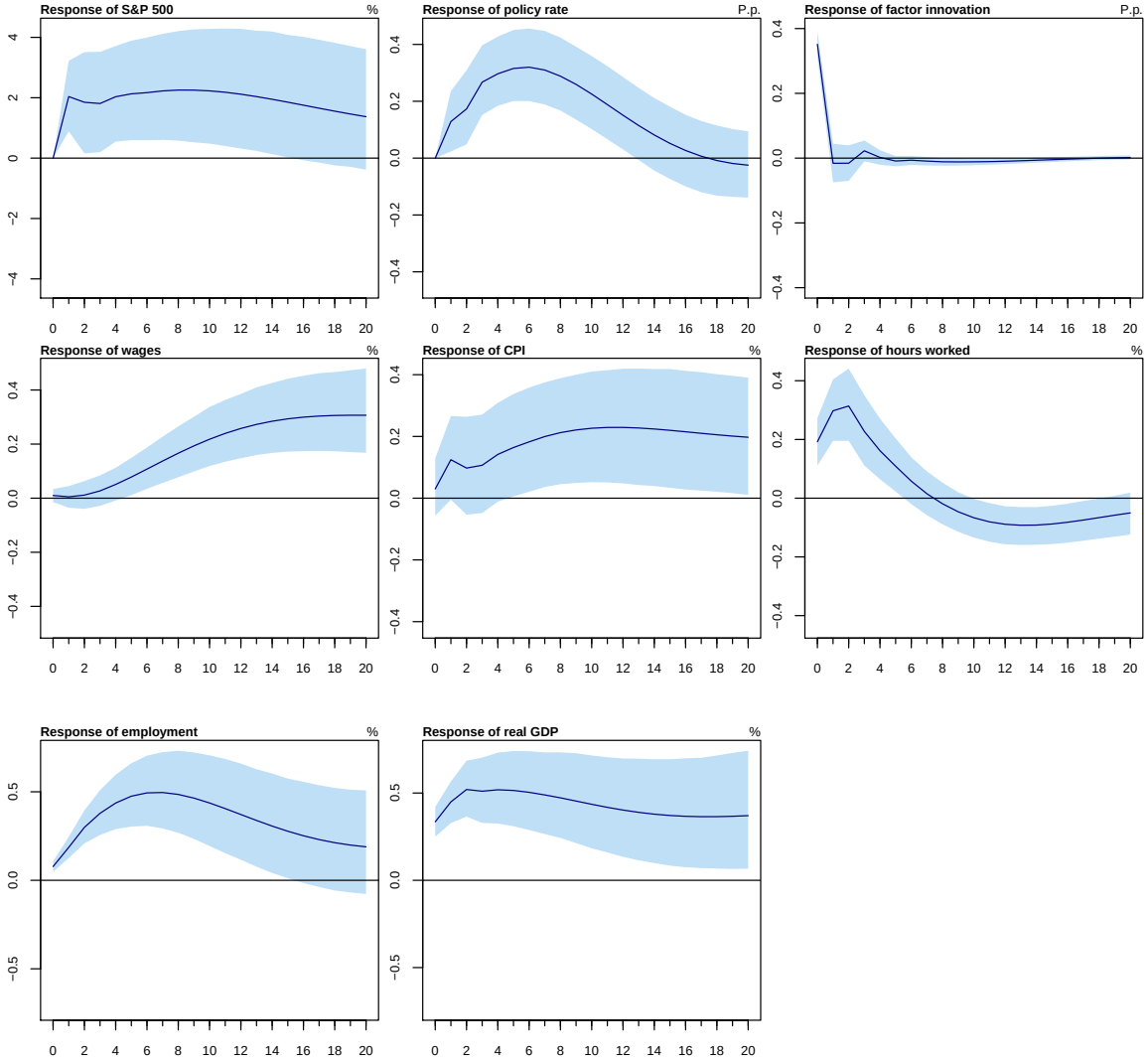
Notes: Impulse response from eight variable VAR. Sample period 1984Q1–2014Q4. Figures show response to one standard deviation innovation to real GDP. Shaded area denotes 90 percent bootstrapped confidence interval. See text for details.

Figure A11: Response to employment innovation, VAR-8.



Notes: Impulse response from eight variable VAR. Sample period 1984Q1–2014Q4. Figures show response to one standard deviation innovation to employment. Shaded area denotes 90 percent bootstrapped confidence interval. See text for details.

Figure A12: Response to Tealbook factor innovation, VAR-8, total CPI.



Notes: Impulse response from eight variable VAR, uses total CPI to measure inflation. Sample period 1984Q1–2014Q4. Shaded area denotes 90 percent bootstrapped confidence interval. See text for details.

A5 Description of MCMC algorithm

The MCMC algorithm for the estimation of the joint posterior distribution of an FSV model follows Pitt & Shephard (1999). Divide the parameters into five blocks: the volatility states for $t = 1, ..., T$ and all i; the variance of the volatilities (the σ^2's); the factor loadings; the parameters for the autoregressive processes; and the time-series of the state vector. I document the algorithm below.

Log volatilities. Focusing on the common factor, we have the measurement equation and its corresponding random walk transition equation:

$$f_t = \phi_0 + \sum_{i=1}^{P} \phi_i f_{t-i} + e^{\frac{1}{2}h_{ft}}\varepsilon_t$$

$$h_{ft} = h_{ft-1} + \eta_t$$

where $\varepsilon_t \sim N(0,1)$ and $\eta_t \sim N(0, \sigma_f^2)$. Kim, Shephard & Chib (1998) and Omori, Chib, Nakajima & Shephard (2007) provide a routine for drawing the time-series of h_{ft} conditional on the parameters and a draw of the states. The routine involves the forward-filter backwards-sampling approach of Carter & Kohn (1994) and approximating a log χ^2 distributed error term ($\log \eta_t^2$) with a mixture of Gaussians. The log-volatilities of the horizon-specific measurement errors are drawn analogously. All log volatilities are initialized using independent draws from $h_{i0} \sim N(1, 10)$.

Variance of log volatilities. Conditional on the log-volatilities and given an inverse-Gamma prior, each σ_i^2 is drawn from an inverse Gamma posterior:

$$\sigma_i^2 \sim IG(\frac{v_1}{2}, \frac{\delta_1}{2}) \tag{1}$$

with $v_1 = v_0 + T - 1$ and $\delta_1 = \delta_0 + \sum_t (h_{it} - h_{it-1})^2$.

Factor loadings. Conditional on the data and other parameters, factor loadings are drawn from a Bayesian regression of output gap estimates on the factor with autocorrelated and

heteroskedastic errors. The measurement equation is:

$$\hat{y}_{it} = \lambda_i f_t + e_{it}$$

with $e_{it} = \rho_i e_{it-1} + e^{\frac{1}{2}h_{it}}\varepsilon_{it}$. First, I filter both sides of the equation with $1 - \rho_i$L, and then divide each side by $e^{\frac{1}{2}h_{it}}$. Because my prior is Gaussian, the posterior of λ_i is drawn from a Gaussian distribution with known mean and variance.

Autoregressive coefficients of the latent factor and measurement errors. To draw the autoregressive parameters of the factor and the measurement errors, I again condition on the state vector so that f_t and e_{it} are treated as observed. The transition equations become standard Bayesian regression problems with heterskedastic errors, and are analyzed using steps analogous to those for the factor loadings. I discard draws that lie outside the unit circle.

State vector. Conditional on all model parameters, I draw the state vector using the simulation smoother of Durbin & Koopman (2002); see also Jarocinski (2015). Draws of the state vector are initialized using a random draw from a multivariate Gaussian distribution, N(μ_0, $10 \times I$). The mean of the common factor is centered at the 1983Q4 value from the most recent vintage output gap available; all other elements of μ_0 are centered at zero.

After a burn-in period of 10,000 draws, I save 2,000 draws from the joint posterior distribution as described above and using a thinning factor of 10.

A5.1 Prior distributions

Prior distributions are conjugate but diffuse.

- The prior for each λ_i $N(1, 10)$.The choice of large variance is intended to represent notable uncertainty around the factor loadings. To initialize the sampler, the factor loadings are all set to 1. Since the factor and loadings are not fully identified, I set the loading of the most distant output gap estimate (i.e., $\hat{y}_{t|t+12}$) to one.

- The prior of the autoregressive coefficients for the true output gap, Φ is multivariate Normal but limited to the stationary region of the parameter-space: $\Phi \sim N(\mu_\Phi, \Sigma_\Phi)_\Phi$, with $\mu_\phi = [0, \frac{3}{2}, -1]'$ and Σ_Φ =diag(10,2,2).

- The prior of the autoregressive coefficient for measurement errors, ρ_i is normal and stationary: $\rho_i \sim N(0, 2)_{|\rho|<1}$.

- The prior for the variability of the volatilities is inverse gamma: σ_f^2 and $\sigma_i^2 \sim IG(\frac{v_0}{2}, \frac{\delta_0}{2})$. I set both v_0 and δ_0 to one, producing an extremely flat prior density.

Table A7: Prior distribution and implied percentiles for parameters of model.

Parameter	Distribution	Percentiles				
		5%	15%	50%	85%	95%
λ_i	N(1,10)	-4.20	-2.28	1.00	4.28	6.20
ϕ_0	N(0,10)	-5.20	-3.28	.00	3.14	5.20
ϕ_1	N(1.5, 2)	-.83	..03	1.50	2.97	3.83
ϕ_2	N(-1, 2)	-3.33	-2.47	-1.00	.47	1.33
ρ_i	N(0,2)	-2.33	-1.47	.00	1.47	2.33
σ_f^2	IG(½,½)	.26	.48	2.20	27.96	254.31
σ_i^2	IG(½,½)	.26	.48	2.20	27.96	254.31

Notes: Table reports prior distributions and values of selected quantiles. Priors of λ_i, ρ_i, and σ_i^2 are symmetric across all $i = -B, -B+1, ..., H$.

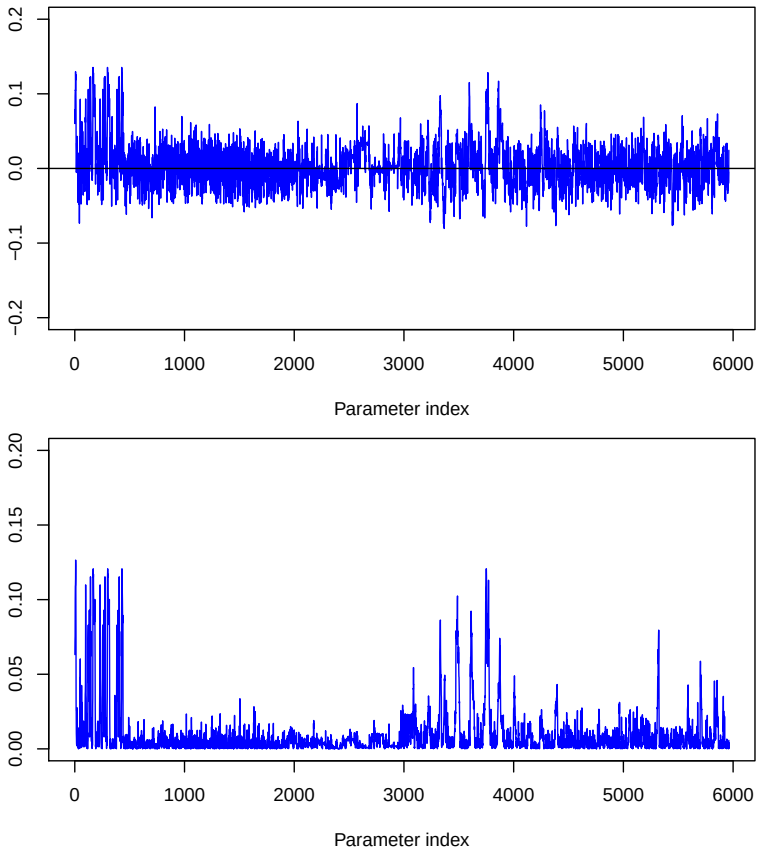
A6 Convergence diagnostics

Following Primiceri (2005) and Baumeister & Peersman (2013), figure A13 provides evidence regarding the convergence and efficiency of the Gibbs sampler. Specifically, for each draw of the Gibbs sampler, I collect the parameters in the following order: stochastic components of λ, (elements 1–20); Φ (elements 21–23); ρ (elements 24–44); the diagonal elements of Σ (elements 45–66); draws of $\{S\}_{t=1}^T$ (elements 60–3079); and lastly, the draws of $\{h_t\}_{t=1}^T$ for each h_i (elements 3080–5961). The top panel displays the 20^{th} order autocorrelation across

draws of the sampler. The bottom panel plots, for each estimand, the inverse of the relative numerical efficiency measure (RNE) of Geweke (1992).

The two panels indicate that the autocorrelation across draws is modest for all elements of the model. Most autocorrelations are very close to zero; almost all of the autocorrelations are smaller in absolute value than .05. Similarly, the inefficiency factors are all far below the upper bound of 20 suggested by Primiceri (2005). In sum, the evidence suggests that the sampler has indeed converged to the ergodic distribution.

Figure A13: Convergence diagnostics for the Gibbs sampler.



Notes: Top panel shows 20th autocorrelation across draws, by parameter. Bottom panel shows the inverse of the relative numerical efficiency measure of Geweke (1992), by parameter.