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. (2020). "Time-varying uncertainty of the Federal Reserve's output gap estimate," Finance and Economics Discussion Series 2020-012. Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/FEDS.2020.012.

January 2020

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

Abstract

What is the output gap and when do we know it? A factor stochastic volatility model estimates the common component to forecasts of the output gap produced by the staff of the Federal Reserve, its time-varying volatility, and time-varying, horizon-specific forecast uncertainty. The common factor to these forecasts is highly procyclical and unexpected increases to the common factor are associated with persistent responses in other key macroeconomic variables. The estimates of this common factor, however, are very uncertain, even well after the fact. Output gap uncertainty increases at business cycle turning points. Lastly, increased macroeconomic uncertainty, as measured by the output gap's time-varying volatility, produces pronounced negative responses to other macroeconomic variables.

- *Keywords*: Output gap; 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

---

*The views expressed herein are my own and do not reflect those of the Governors 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 a very difficult task. Empirical estimates of the output gap are unstable in real time, and even ex-post, considerable uncertainty remains because the truth is never observed.

This paper aims to understand with what degree of accuracy the output gap can be estimated in real time. Measuring the uncertainty that surrounds output gap estimates is not a trivial task. In the United States, estimates of Gross Domestic Product and its price deflator are revised, sometimes substantially so, for many years after their initial release. The true output gap is never observed, so that uncertainty bands based on ex-post forecast errors cannot be computed. And since many prominent output gap estimates, including those produced by the staff of the Federal Reserve, are produced judgmentally, 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. Understanding and communicating macroeconomic uncertainty may be a key step towards improving economic outcomes (Orphanides 2019).

A related goal of this paper is to provide a new framework to discriminate between various output gap estimates. The previous literature discriminates between output gap estimates by evaluating their correlation with future price inflation. However, since inflation is extremely difficult to forecast in general (Stock & Watson (2007, 2008)), it has proven impracticable to differentiate output gap estimates using this criterion. I instead evaluate the link between changes in forecast-implied output gaps and the subsequent behavior of several macroeconomic variables. Although price inflation is one of these variables, the judgment of output gaps estimates is broader, since the analysis asks the question: "How is the object implied by a set of forecasts linked to other economic phenomena?"

Orphanides & van Norden (2002) are the first to document the behavior of real-time output gap estimates by evaluating how much these estimates revise over time. Their conclusions about the utility of output gap estimates in real time are pessimistic: in the models and sample they consider, ex-post revisions to the output gap are as variable as the estimated gap itself.[1] Edge &

[1] Since Orphanides & van Norden (2002), a number of papers attempt to ameliorate the uncertainty of output gap estimates, for example, Garratt, Lee, Mise & Shields (2008) Garratt, Mitchell & Vahey (2014) and Aastveit & Trovik (2014).

Rudd (2016) evaluate output gap estimates of the staff of the Federal Reserve and find them to be much more stable. Further, and in contrast to Orphanides & van Norden (2005), Edge & Rudd find that the real time output gap estimates from the Board staff do not worsen inflation forecasts relative to ex-post estimates.

This paper contributes to the literature on real time output gap estimation in several ways. First, the previous literature had to assume a particular estimate of the output gap is "true," so that deviations from the truth can be used to evaluate output gap forecasts.[2] The approach here instead explicitly recognizes that the output gap is never observed. Perhaps more importantly, the bulk of the previous literature focuses on understanding the uncertainty of real-time output gap nowcasts, estimates of the output gap in the most recent quarter. However, policy is best set in a forward-looking manner, so it is not only the current quarter's output gap that is relevant for policymakers, but rather its expected path. The method developed here produces estimates of the uncertainty surrounding the entire path of an output gap forecast.

To do so, I develop a state-space framework that treats different vintage estimates of economic variables as mismeasured proxies of some true value that is never observed. I endow both the latent variable and the forecast errors with time-varying uncertainty.[3] Time-varying forecast uncertainty is critically important since it plays a central role in understanding two key macroeconomic events of the past several decades: the Great Moderation and the Great Recession. Output gap stability has changed dramatically over time: Edge & Rudd estimate that the volatility of revisions to the Federal Reserve's output gap estimate from 1994Q1–2006Q4 is less than half the volatility from 1980Q1–1992Q4. These changes output gap uncertainty may have first-order implications for the conduct of monetary policy. Orphanides (1998) argues that policy ought to be less reactive to an estimated output gap in periods when it is especially uncertain, since excessive interest rate movements are destabilizing.

The model also provides an estimate of the uncertainty about the output gap itself, a proxy for macroeconomic uncertainty. A recent literature explores how to measure uncertainty, and its

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

[3] A sizable literature applies factor stochastic volatility models to macroeconomic and financial data. Early examples include, for example, Pitt & Shephard (1999), Aguilar & West (2000) and Chib, Nardari & Shephard (2006). More recent applications of factor models with stochastic volatility applied to macroeconomic data include Del Negro & Otrok (2008), Carriero, Clark & Marcellino (2018*b*), and Jo & Sekkel (2019).

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

Finally, the framework produces estimates of macroeconomic innovations that are themselves of interest because they can be used to understand what the output gap measures. When innovations to the output gap implied by the Board staff estimates are included into an auxiliary vector autoregression, I find that the response of output, employment and wages to the output gap are sizable and persistent. Inflation responds to an output gap innovation, but in a much more muted manner. Similarly, when innovations to the output gap's volatility are included in a VAR, I find macroeconomic uncertainty generates notable negative effects on the macroeconomy, consistent with the recent macroeconomic uncertainty literature such as Bloom (2009) and Jurado et al. (2015). In contrast, output gap innovations implied by the HP filter or Hamilton (2018) filter are associated with responses in other macroeconomic variables that are less comprehensible. For example, following a positive innovation from an HP-filter output gap, equity prices, wages, employment, and output all decline.

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

2 The stability of real-time output gap estimates

2.1 Real-time Federal Reserve 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 twice per quarter, there are eight Tealbooks produced per year. Tealbook projections include forecasts for production, the labor market, prices, wages, financial markets, potential output, and the output gap. Importantly, while some aspects of the forecast are informed by statistical and econometric models, the Tealbook is ultimately judgmental in nature (Reifschneider, Stockton & Wilcox, 1997; Mishkin, 2007).

Edge & Rudd (2016) collect a long time series of output gap nowcasts. To this data, I am able to append more complete vintage Tealbook output gap estimates using data available from the Federal Reserve Bank of Philadelphia.[5] The vintage estimates have the benefit that they contain not only the output gap nowcast, but estimates of the *path* of the output gap. Since May 1996, Tealbook output gap estimates include the path of the output gap since approximately 1960. Because the Tealbook is made public with at least a five-year lag, the most recent Tealbook vintage in the sample is the December 2013 projection; see table A1.

To introduce notation, throughout the paper superscripts refer to vintages while subscripts to time periods. Thus, $\hat{y}^{v}_{t+h-1}$ denotes the output gap forecast of period $t+h-1$ produced in vintage v.[6] For any forecast horizon h, the revision from vintage v until an alternative vintage, $\tilde{v}$, is $\rho^{\tilde{v},v}_{t+h-1} = \hat{y}^{\tilde{v}}_{t+h-1} - \hat{y}^{v}_{t+h-1}$, where $v \in t$. The sequence $\{\rho^{\tilde{v},v}_{t+h-1}\}^{T}_{t=1}$ defines an empirical 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, 8-12 quarters ahead.

Orphanides & van Norden (2002) and Edge & Rudd (2016) set $\tilde{v} = V$, the most recent vintage. This approach has the drawback that some output gap estimates will have undergone more revisions than others. It also implicitly assumes that the most recent estimate is "true." I choose instead

[4] Prior to 2010, the Board staff projection was known as the Greenbook. I use the term Tealbook throughout to refer to the Federal Reserve projection.

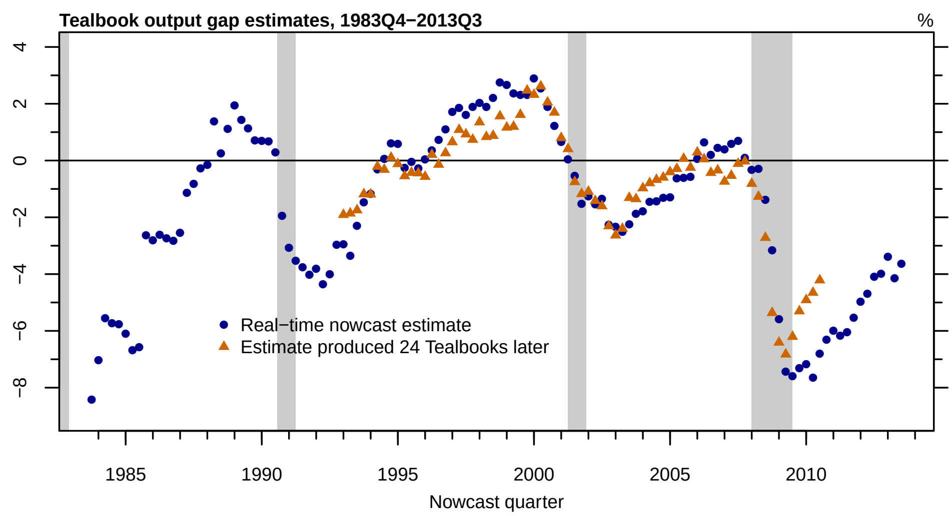
[5] Appendix A1.1 contains details of the source data for the output gap estimates.

[6] The fixed '-1' reflects the typical notation in the literature, where the nowcast quarter is defined as the previous calendar quarter.

to hold fixed the window between the real-time and final estimate by setting $\tilde{v} = v + i$, where i is an integer that denotes a number of quarters. I choose $i = 12$ quarters (24 Tealbooks), so that for the nowcast estimates, three years will have passed between the nowcast estimate and the "final" estimate. This choice sidesteps having to assume a particular vintage is true and produces an easily interpreted distribution of revisions.

Figure 1 shows the Tealbook output gap nowcasts alongside the estimate of the same quarter produced 12 quarters later. The nowcast estimates begin with the January 1984 Tealbook estimate of the output gap in 1983Q4 and end with the October 2013 Tealbook estimate of 2013Q3. Due to the data limitations described above, the sequence of backcasts begins later (1993Q1) and ends with the October 2013 Tealbook estimate of 2010Q3.

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 from 12 quarters (24 Tealbooks) hence. Sample period 1983Q4–2013Q3. Each quarter's nowcast is from the first Tealbook produced in the subsequent quarter. Shaded regions denote NBER-defined recessions. See text for details.

Table 1 shows the mean, standard deviation, minimum, maximum, and root mean-squared error of the revisions at different forecast horizons. 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 estimate from vintages $\tilde{v}$. The nowcast column, $h = 0$, corresponds to the findings from Edge & Rudd (2016) using a different sample period. At the nowcast horizon, the standard deviation of the revision of output gap estimates three years apart is a percentage point. The noise-signal ratio is similar to that in Edge & Rudd, which is notable given that their dataset ended in 2006Q4 whereas the statistics here include the Great Recession. Indeed, it is clear from figure 1 that Tealbook nowcasts of the output gap reacted quite quickly during the recession, turning negative coincident to the NBER business cycle peak. The estimates from 24 Tealbooks later are broadly similar, but reach a less negative trough earlier than the real-time estimates.

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

	Forecast horizon						
	-16	-12	-8	-4	0	4	8
Mean	.01	.11	.21	.10	-.08	-.17	-.51
Std dev	.45	.53	.67	.68	.96	1.64	2.39
Min	-.70	-.90	-.95	-1.16	-2.19	-6.15	-8.41
Max	1.09	1.29	1.62	2.15	3.00	3.13	2.90
RMSE	.44	.54	.70	.69	.96	1.64	2.42
Noise-signal ratio							
Using SD	.22	.26	.33	.33	.47	.79	1.15
Using RMSE	.21	.26	.34	.33	.46	.79	1.17
N.obs	58	58	59	68	71	71	57

Notes: Table presents summary statistics of output gap revisions. Revisions calculated setting $\tilde{v}$ to be equal to $v + 12$ quarters (24 Tealbook vintages). Tealbooks range from January 1984–October 2013, missing estimates as described in text. See text for details.

Tealbook output gap backcasts revise less than nowcasts, which in turn are less volatile than forecasts. The standard deviation of revisions to the 16 quarter backcast—an estimate of the output gap 4 years prior—is about .5 percentage point, half the standard deviation of the nowcast horizon. Going forward, forecast revisions become much larger. For the 8-quarter ahead forecast, the signal-noise ratio is above 1, indicating that the volatility of output gap revisions is greater than the volatility of the output gap itself. (The signal-noise ratios remain below 1 until the 6 quarter-ahead forecast.) The mean revision tends to be negative for the forecasts, indicating that the original forecast was too optimistic, although given the relatively large standard deviation, one cannot claim that this bias is meaningful. Backcast revisions have a mean of about zero.

Lastly, revisions to output gap forecasts are asymmetric. For the eight quarter-ahead forecast, the largest revision is negative, -8.4 percent, a revision to the forecast of the output gap in 2009Q2.[7]

[7] The August 2007 Tealbook projection of the output gap in 2009Q2 was .2 percent; the subsequent August 2010 Tealbook estimate of the output gap in 2009Q2 was -8.2 percent. The largest positive output gap revision is 3.6 percent. (The value is not shown in table 1 because the forecast horizon is $h = 1$.) The August 2010 projection of the output gap in 2010Q3 was -7.7 percent, compared to the August 2013 estimate of just -4.1 percent.

The asymmetry is also seen in the four quarter-ahead forecasts. However, the nowcasts and backcast revisions appear about symmetric.

2.2 Comparison to statistical output gap estimates

The nowcast horizon of table 1 is essentially an update of Edge & Rudd (2016), but the results for other forecast horizons are novel. To place these results into better context, 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.[8] I add to this list two more recent detrending methods, Mueller & Watson (2017) and Hamilton (2018). The models are estimated on quarterly real-time vintages of log real GDP, with estimation always beginning in 1960Q1.[9] Since these detrending methods are not dynamic, they can estimate an output gap only through the nowcast quarter. As before, revisions are calculated using vintage estimates 12 quarters apart.

Table 2 displays the noise-signal ratios for the univariate filters at select forecast horizons. Unsurprisingly, for most of the models, noise-signal ratios tend to become smaller when the estimate is further in the past. For example, for the HP filter, whereas the nowcast output gap estimates have noise-signal ratios above 1, the output gap estimate four quarters back has noise-signal ratio of about .7, eight quarters back have noise-signal ratio of about .4, and 16 quarters back .25. This pattern of noise-signal ratio by forecast horizon is typical, although some models have larger signal-noise ratios on average than others. However 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 linear model, broken linear model, quadratic trend, and Mueller-Watson output gap estimates all have noise-signal ratios approaching or above one. The noise-signal ratios presented here are smaller than the ones found in Orphanides & van Norden (2002), likely owing to the different sample period. Finally, output gap revisions from the Hamilton (2018) method are among the smallest noise-signal ratios, at all forecast horizons.

[8] 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 these breaks to enter the models after a three year delay.

[9] Real-time GDP data are described in appendix A1.2.

Table 2: Noise-signal ratios of output gap estimates from univariate detrending methods.

	Forecast horizon				
	-16	-12	-8	-4	0
Linear trend					
Using SD	.56	.57	.57	.52	.54
Using RMSE	1.04	1.08	1.10	1.06	1.02
Broken linear trend					
Using SD	.64	.63	.62	.58	.58
Using RMSE	1.13	1.19	1.21	1.15	1.08
Quadratic trend					
Using SD	.87	1.01	1.13	1.17	1.22
Using RMSE	.86	1.01	1.14	1.22	1.33
HP					
Using SD	.24	.30	.39	.68	1.15
Using RMSE	.24	.30	.40	.68	1.15
Hamilton					
Using SD	.36	.35	.34	.32	.33
Using RMSE	.36	.36	.36	.37	.42
Mueller-Watson					
Using SD	.65	.75	.75	.89	1.29
Using RMSE	.65	.77	.87	.92	1.68
Memo: Tealbook					
Using SD	.22	.26	.33	.33	.47
Using RMSE	.21	.26	.34	.33	.46

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 1984–2013. See text for details.

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

I now introduce a factor stochastic volatility (FSV) model that produces time-varying estimates of output gap uncertainty, as well as time-varying uncertainty about each forecast horizon estimate thereof.

As above, we observe estimates of the output gap in period t across vintages v. Let $i = -B, ..., H$, where B denotes a maximum backcast horizon and H a maximum forecast horizon. Then for each quarter, collect the $N{=}B{+}H{+}1$ estimates of the output gap in t into the vector $\hat{y}_t = (\hat{y}_t^{v-B}, \hat{y}_t^{v-B+1}, ..., \hat{y}_t^{v+H})'$. The elements of $\hat{y}_t$ are sequential in vintage, with the longest backcast vintage ordered first.

The output gap estimates have a factor structure:

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

where $\hat{y}_t$ is the $N \times 1$ vector defined above, $\lambda = [\lambda_{-B}, \lambda_{-B+1}, ..., \lambda_H]'$ is a vector of factor loadings, f_t is a latent scalar common to all vintage estimates, and e_t is 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 elements of λ have the interpretation of weights in a combined forecast of f.

The common factor and idiosyncratic errors are stationary independent autoregressive processes:

$$\begin{aligned} f_t &= \phi_0 + \sum_{i=1}^{P} \phi_i f_{t-i} + u_{ft} && (2) \\ e_{it} &= \sum_{j=1}^{Q} \rho_j e_{it-j} + u_{it}. && (3) \end{aligned}$$

The constant term ϕ_0 in equation (2) is included so that the latent estimate can have a non-zero unconditional mean. That measurement errors are *ex-ante* zero mean embeds an assumption that $\hat{y}_t$ contains unbiased estimates of f. This assumption is necessary since one cannot identify a constant in both equations (2) and (3). However, the assumption is justified given that Tealbook projections of other important macroeconomic variables, such as real GDP growth, the unemployment rate, and inflation appear to be unbiased (Messina, Sinclair & Stekler, 2015; Reifschneider & Tulip, 2017; Berge, Chang & Sinha, 2019). Table 1 also indicates that one could not reject a null hypothesis of unbiasedness based on output gap revisions.

Autocorrelated measurement errors allow for what Jacobs & van Norden (2011) denote spillover effects, the possibility that revisions to the output gap estimate of period t from vintage v also carries information about the output gap in surrounding periods. For example, information may arrive at $v+j$ causing a revision of the estimate $\hat{y}_t^{v+j}$. This information may also lead to revisions in nearby estimates, such as $\hat{y}_{t-1}^{v+j}$ or $\hat{y}_{t+1}^{v+j}$. In practice, revisions to macroeconomic trends such as structural productivity would produce spillovers.

Finally, I assume that u_{ft} and $u_{et} = (u_{e1t}, u_{e2t}, ..., u_{eVt})'$ are conditionally independent and

Gaussian with time-varying volatility:

$$\begin{pmatrix} u_{ft} \\ u_{et} \end{pmatrix} \sim N\left(0, \begin{bmatrix} e^{h_{ft}} & 0 \\ 0 & \Omega_t \end{bmatrix}\right) \tag{4}$$

The matrix Ω_t is diagonal with entries $\{e^{h_{-Bt}}, e^{h_{-B+1t}}, ..., e^{h_{Ht}}\}$. The log of the common and vintage-specific volatilities follow independent random-walk processes:

$$h_{it} = h_{it-1} + \sigma_i \eta_{it} \tag{5}$$

where each σ_i determines the variability of the volatilities, and η_{it} are standard normal innovations. The time-invariant matrix with entries σ_i^2 on its diagonal is Σ.

Several estimands are of interest. The first is the latent factor. I will refer to this object as the 'FSV,' 'latent,' or 'measured' output gap. The factor is not a "true" output gap, instead, it is the output gap implied by the estimates $\hat{y}$. Different models of y will imply a different latent gap. Other objects of interest are the time-varying standard deviations of the factor and measurement errors. The volatility of the common factor is the volatility of the business cycle, a measure of macroeconomic uncertainty. The standard deviations of the idiosyncratic errors capture the time-varying uncertainty around horizon-specific output gap estimates. Time-variation allows for the possibility for the measurements of f to be more or less accurate through time. Although we expect backcasts to be less uncertain than forecasts, the model structure here is *ex-ante* agnostic about the relative variances of the elements of e.

3.1 Estimation

As a baseline, I use Tealbook output gap estimates as inputs. Because of data limitations, the baseline estimation begins in 1984.[10] The dataset therefore consists of 127 output gap *nowcasts* (1984Q1–2013Q3), but has many fewer output gap estimates at the other forecast horizons. I set B to 12 and H to 8 quarters, chosen to reflect the period of interest for policymakers. Thus, each quarter's output gap is informed by at maximum 21 Tealbook estimates, although data limitations imply that there are many missing observations at the start and end of the sample. The lag length

[10] Additional results in section 5.4 provide a Tealbook-based output gap estimate since 1962.

of ϕ is set to 2, while the measurement errors follow AR(1) processes.

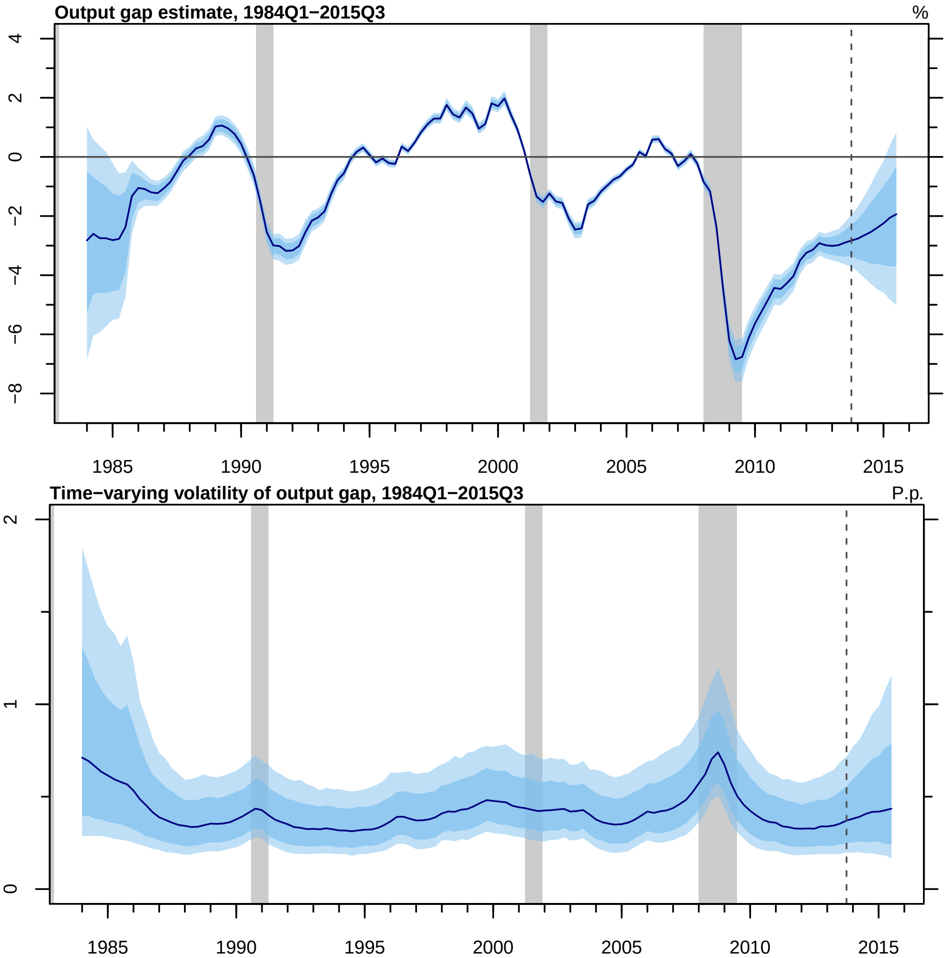
The model can be cast as a conditionally Gaussian state-space model, which is estimated with Bayesian Markov Chain Monte Carlo methods. Details are left to the appendix, but the sampler consists of grouping the parameters and latent variables into blocks and repeatedly drawing from each block's known conditional posterior distribution. The sampler consists of the following steps: 1) Draw the time-series of the log-volatilities, h^T; 2) draw the autoregressive coefficients of the latent common factor, ϕ; 3) draw the autoregressive coefficients of the measurement errors, ρ; 4) draw the variance of innovations to the log-volatilities, Σ; and, 5) draw the state vector. Following a burn-in period of 10,000 draws, I collect 2,000 draws from the posterior distribution using a thinning factor of 10. My prior distributions are standard and not tightly specified; see Appendix A2. Convergence diagnostics are reported in Appendix A3.

4 Results

4.1 The Tealbook output gap and its uncertainty

The top panel of figure 2 reports the posterior median estimate of the latent output gap within its 70 and 90 percent credible intervals. The estimated output gap is very clearly pro-cyclical. There are three local peaks during the sample period, each of which predates an official NBER-defined recession by four to six quarters. Only one timespan during this period appears to have had a notably positive output gap, the late 1990s. The recovery from the 2001 recession is slow: consistent with the "jobless recovery," the output gap estimate reaches a minimum in 2003, nearly 2 years following the NBER's declaration of the business cycle trough. The gap turns positive only in 2006-2007 before the onset of the Great Recession, during which output bottoms out at 7 percentage points below its potential level. The credible intervals measure the uncertainty about the Tealbook's modal output gap estimate in any given period. That the credible intervals widen at both ends of the sample reflects a relative paucity of data during these periods. The bottom panel shows the time-varying standard deviation of innovations to the gap, the volatility of the cyclical component of macroeconomic activity. The model hints at a decline in output gap volatility from 1984 through the early 1990s, although because the decline is modest. Thereafter, output gap volatility is fairly stable, albeit with a noticeable increase during the Great Recession.

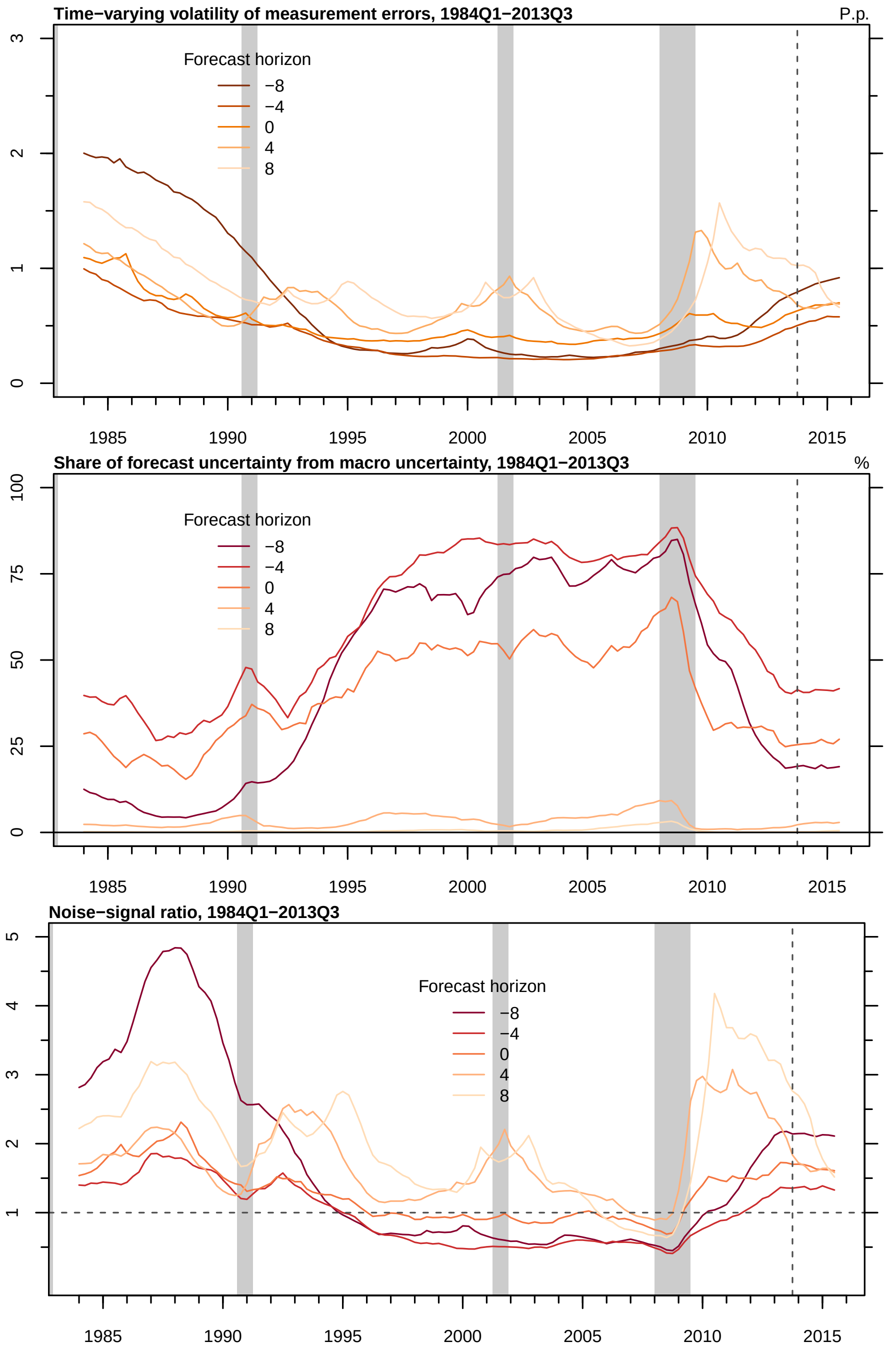
Figure 2: Estimated output gap and its uncertainty.



Notes: Most recent Tealbook produced in October 2013, vertical dashed line. Navy lines show the posterior median estimate. The dark and light blue shaded areas are the 70% and 90% posterior credible sets. Gray shaded regions denote NBER recessions. See text for details.

Figure 3 shows three different summaries of the uncertainty that surrounds any given output gap estimate. The top panel plots the standard deviation of the error for selected forecast horizons. Prior to 1996, there are very few backcasts of the output gap available, and hence a relatively high degree of uncertainty. Thereafter, the backcasts have the lowest uncertainty, as would be expected. The eight-quarter backcast achieves its minimum value in the early 2000s, about .3 percentage point. After that, the backcast uncertainty rises somewhat, especially in the recovery following the Great Recession. This increase likely reflects the difficulty during the post-recession period knowing how much of the decline in output was owed to structural rather than cyclical factors, as well as the fact that we do not yet have backcasts for these periods. As expected, output gap forecasts

Figure 3: Estimated uncertainty for horizon-specific measurement error.



Notes: Top panel of the figure shows estimated time-varying standard deviation of horizon-specific measurement errors, selected horizons ($e^{h_{it}/2}$). Middle panel shows the share of the variance explained by the macroeconomic uncertainty for forecast horizon $(\lambda_i^2 e^{h_{ft}}/(\lambda_i^2 e^{h_{ft}} + e^{h_{it}}) \times 100)$. Bottom panel shows the noise-signal ratio, defined as ratio of standard deviation of the measurement error to that of the FSV output gap. Gray shaded regions denote NBER recessions. Most recent Tealbook is October 2013, vertical dashed line. Each colored line denotes posterior median estimate. See text for details.

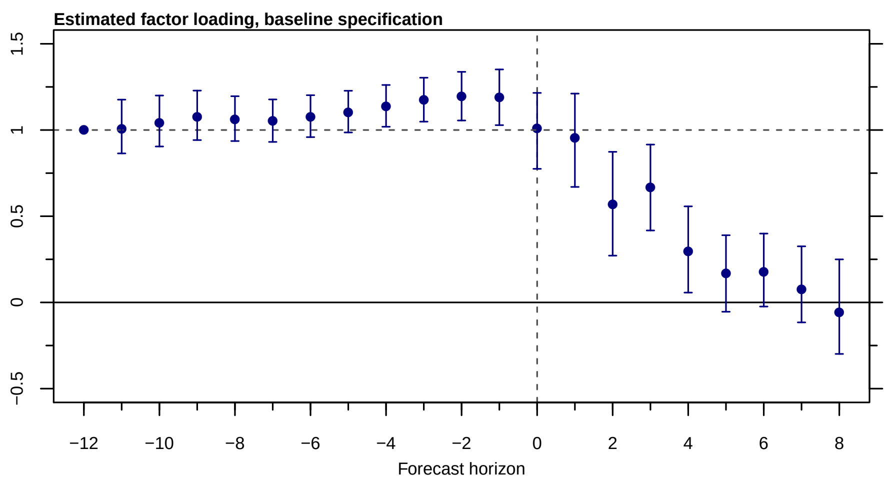
are more uncertain than backcasts. The uncertainty around forecasts is counter-cyclical, indicating that output gap forecasts are especially uncertain during downturns. The uncertainty of nowcast output gap estimates is not as cyclical as the forecasts. The uncertainty of the nowcast increases in the mid- to late-1990s, a period when, in real time, it was uncertain what effect technology was having on productivity and therefore potential output. Nowcast uncertainty increases around the time of the Great Recession, and remains relatively elevated through 2013. During this time, the standard deviation of the nowcast output gap measurement error is estimated to be roughly .7 percentage point.

The middle panel of the figure shows a decomposition of the variance for each measurement error into two parts: the variance of the common factor—a measure of macroeconomic uncertainty and the irreducible portion of the forecast variance—and the variance of the forecast itself.The panel plots the share of uncertainty due to macroeconomic uncertainty. For backcasts, beginning in the mid 1990s and lasting through the period of the Great Recession, roughly 75 percent of the uncertainty of Tealbook backcasts derives from the output gap itself, and is therefore irreducible. The share of the uncertainty of the output gap nowcast driven by macroeconomic uncertainty trends higher for most of the sample, indicating that the Tealbook output gap nowcast became more accurate over time. The share of the nowcast variance explained by macroeconomic uncertainty peaks at roughly 65 percent at the time of the Great Recession. Thereafter, the share nowcast uncertainty from macroeconomic uncertainty falls, reflecting the difficulty in understanding to what degree the decline in output during the Great Recession represented a change in the economy's productive capacity. As for the forecasts, the bulk of the uncertainty derives from the measurement error themselves.

The final panel of the figure shows the noise-signal ratio from the model that is analogous to those calculated by Orphanides & van Norden (2002) and Edge & Rudd (2016). It plots the ratio of the standard deviation of each forecast horizon measurement error to the standard deviation of the FSV output gap. The overall message of the figure is the same: backcasts of the output gap obtain noise-signal ratios well below one. The noise-signal ratio for nowcasts is slightly higher, around one until the Great Recession. The forecasts are very uncertain, with noise-signal ratios well above one.

That output gap forecasts are more uncertain than backcasts is also reflected in the factor loadings. Figure 4 reports the posterior median draw for each factor loading with its 90 percent credible interval. The common factor weighs most heavily on recent backcasts. That recent backcasts receive a somewhat larger weight than backcasts further into the past may reflect that more effort is placed into understanding economic activity in recent quarters than the more distant past in Tealbook projections. The common factor loads less heavily onto output gap forecasts. The factor loading for the two quarter-ahead forecast is smaller, .6, while the 90 percent credible interval for factor loadings on output gap projections beyond 5 quarters ahead all include zero.

Figure 4: Posterior estimates of factor loadings.



Notes: Posterior estimate of factor loadings, baseline model. Point is posterior median estimate, whisker bars denote 90 percent credible interval. Factor loading at $h = -12$ set to unity (horizontal dashed line). Vertical dashed line indicates nowcast horizon (h=0). See text for details.

Table 3 shows posterior distributions of other model parameters. The common component is estimated to be quite persistent, and most of the posterior distribution of the unconditional mean of the FSV output gap, μ, is negative. The measurement errors tend to be positively autocorrelated—indicating a given vintage Tealbook output gap estimate may deviate from the FSV gap in a persistent way, and suggesting strong spillover effects vintage-to-vintage. Lastly, the estimates of σ for various horizons are by and large similar.

To understand the practical implications of the model, figure 5 shows the estimate of the common factor at the time of the October 2013 Tealbook forecast alongside the Tealbook estimate itself. Considerable uncertainty surrounds the FSV output gap. The 90 percent credible interval of the output gap in 2013Q3, the nowcast quarter, is nearly $1\frac{1}{2}$ percentage point, ranging from

Table 3: Posterior distribution for selected model parameters.

	Posterior distribution				
Parameter	5%	15%	50%	85%	95%
ϕ_0	-.07	-.05	-.02	.00	.03
ϕ_1	1.34	1.39	1.47	1.54	1.58
ϕ_2	-.62	-.59	-.52	-.44	-.39
μ	-1.41	-1.17	-.77	-.33	.01
ρ (h=-8)	.35	.45	.61	.76	.83
ρ (h=-4)	-.05	.07	.24	.40	.49
ρ (h=0)	.75	.80	.87	.93	.96
ρ (h=4)	.84	.87	.91	.95	.97
ρ (h=8)	.82	.86	.91	.95	.97
σ_f	.27	.31	.41	.53	.61
$\sigma_{h=-8}$	.31	.36	.46	.62	.73
$\sigma_{h=-4}$	.27	.30	.39	.51	.59
$\sigma_{h=0}$	.25	.29	.38	.48	.57
$\sigma_{h=4}$	.30	.35	.45	.59	.68
$\sigma_{h=8}$	.33	.39	.50	.65	.75

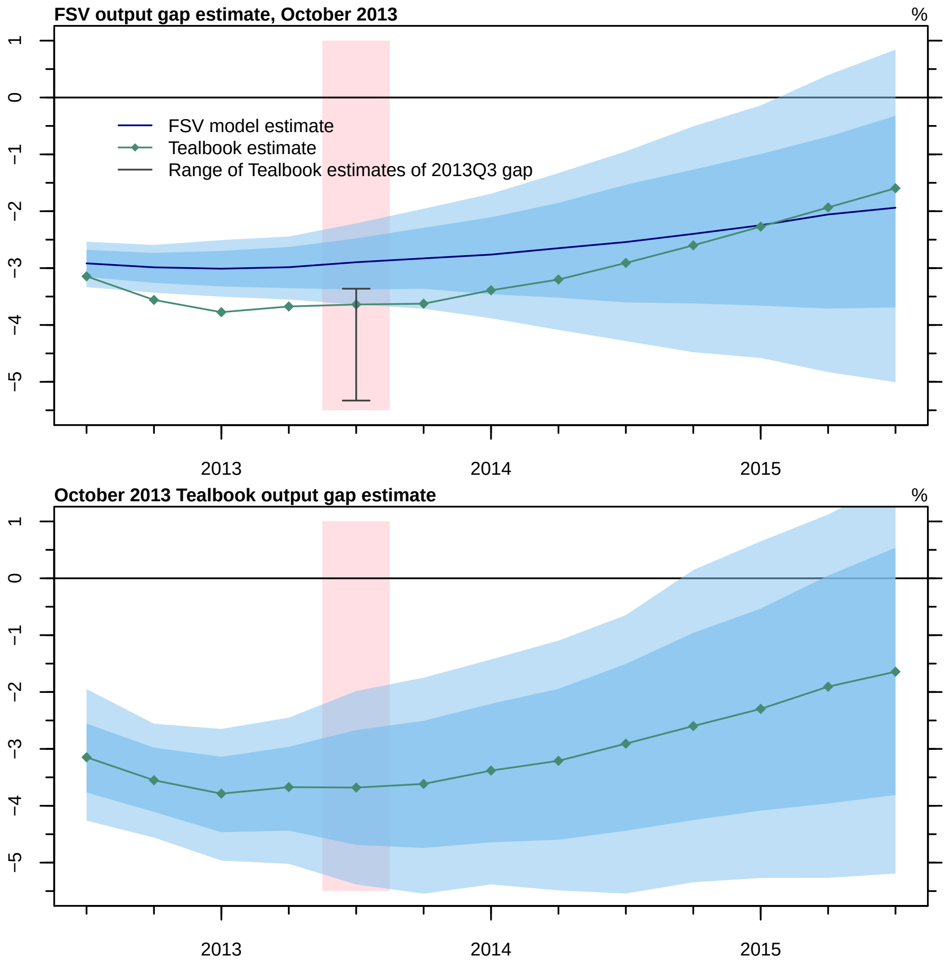
Notes: Posterior distribution for selected model parameters. μ defined as unconditional mean of autoregressive process of common factor, $\mu = \phi_0/(1-\phi_1-\phi_2)$. See text for details.

-3.6 to -2.2. Uncertainty about previous quarters is smaller: the 90 percent credible interval for the 4-quarter backcast of the output gap is about ¾ percentage point. Looking ahead, uncertainty increases dramatically. The width of the 90 percent interval for the eight quarter ahead projection has increased to 5.5 percentage point. The gray whisker bars denote the range of output gap estimates for 2013Q3 produced in Tealbooks up to that point—the data that are entering the model for 2013Q3. The FSV output gap is higher than any Tealbook output gap estimate for 2013Q3 through that point in time.

The bottom panel plots the Tealbook output gap estimate along with it's model-implied 70 and 90 percent credible intervals.[11] The uncertainty surrounding the staff forecast is quite large. The credible set for the Tealbook output gap estimate is noticeably larger than that for the common factor. The nowcast estimate of the bottom panel has a 90 percent credible interval that is about

[11] Specifically, to produce these uncertainty bands, I simulate the measurement error processes from h=-12,-11,...,8. The relevant quantiles are calculated from the simulated measurement errors, which are then added to the Tealbook projection at each forecast horizon.

Figure 5: October 2013 output gap estimates.



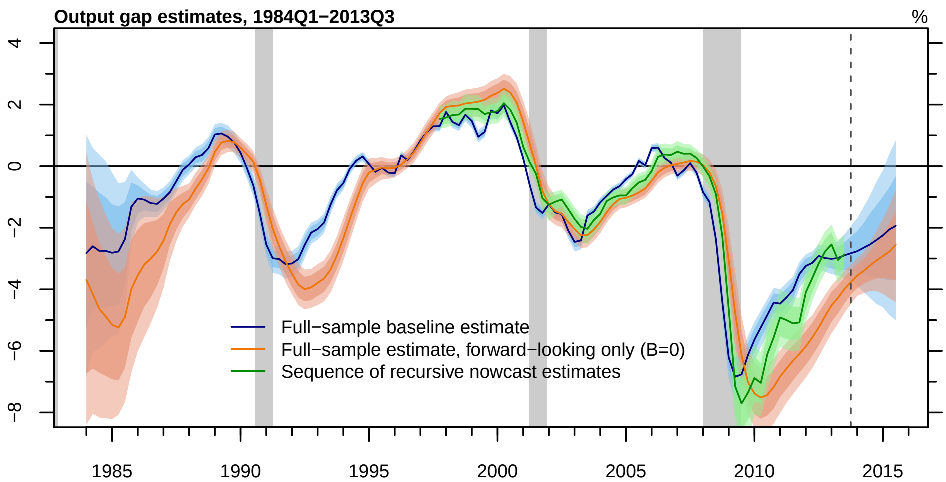
Notes: Top panel shows the model estimate of common factor alongside October 2013 Tealbook output gap projection. Pink shaded region denotes the nowcast quarter, 2013Q3. The dark and light blue shaded areas are 70% and 90% posterior credible sets. Gray whisker bar denotes the range of Tealbook forecasts for the output gap in 2013Q3 produced in preceding Tealbook forecasts. Bottom panel plots the uncertainty bands implied for Tealbook estimate. See text for details.

3.4 percentage point. Going forward, the credible interval widens to 4³/4 percentage point and 7 percentage point at 4 and 8 quarters ahead, respectively.

4.2 Forward-looking estimates

Although Tealbook forecasts are real time in nature, the FSV is estimated over a historical dataset. In this section, I produce two alternative output gap estimates intended to minimize the influence of the backward-looking nature of the model. First, I re-estimate the model but use output gap estimates that are purely forward-looking, setting $B = 0$ and $H = 8$. As above, this is a full-sample estimate, but is intended to measure how much of the output gap estimate is derived from the Tealbook output gap backcasts. Secondly, I estimate a sequence of FSV models using an expanding window (using B =-12 and H =8). Each period, I fully re-estimate the model to produce a truly real-time output gap nowcast. The out-of-sample estimation begins in 1998Q1.[12]

Figure 6: Baseline and alternative output gap estimates.



Notes: Blue line is the baseline model. The orange line the common factor from the model using only forward-looking Tealbook output gap estimates estimated using the full sample. The green line is the sequence of output gap *nowcasts* estimated recursively in real-time (sequence of models have the same values of B and H as the baseline). Dark and light shaded areas are 70 and 90 percent credible intervals. Gray shaded areas denote NBER defined recessions. See text for details.

The result of these exercises is summarized in figure 6. The FSV output gap from the forecast-only model lags behind the baseline model. The largest divergences occur following business cycle turning points, especially the business cycle troughs in the early 1990s and following the Great Recession. In both cases, ex-post, the decline in the output gap from the baseline model is smaller, and subsequently recovers more quickly relative to the estimate from only forward-looking output gap estimates. That the forecast-only model lags the baseline highlights the difficulty in forecasting the output gap, especially at turning points.

The sequence of real-time output gap nowcasts, which begin only in the fourth quarter of 1997, are more volatile than the series estimated with a full sample. The nowcast estimates also appear to lag the full-sample estimate of the output gap, although by less than the model based on output gap

[12] The start-date of the out-of-sample period implies nearly 15 years of output gap nowcasts for the first estimation. However, the first Tealbook estimate that contains both backcasts and nowcasts is from August 1996, meaning that the start date implies only seven Tealbooks of backcasts and nowcasts for the first estimation period.

forecasts. On the whole, however, the three estimated output gaps closely follow one another, the correlation coefficients across the three series all exceed .9. The two forward-looking measures are largely consistent with the full-sample estimate of the output gap, suggesting that model estimates of the output gap—albeit based on the Tealbook estimates thereof—may be helpful in real time.

5 What is the output gap?

The structure of the model presented above provides estimates of the innovations to the FSV output gap, derived from a particular sequence of output gap estimates. In this section I evaluate whether these innovations produce reasonable responses in other macroeconomic variables. Because the FSV model also provides innovations to the output gap's uncertainty, the model can also shed light on the macroeconomic response to uncertainty.

5.1 Relationship with other macroeconomic shocks

I first explore how output gap innovations as derived from the Tealbook-based FSV model relate to other identified macroeconomic shocks. To the extent that there exist macroeconomic shocks that are not explicitly parsed by the Tealbook, then realizations of these shocks may be revealed themselves as innovations to the latent Tealbook output gap estimate.

I consider a number of macroeconomic shocks, as in Stock & Watson (2012). The first identified macroeconomic shock is the utilization-adjusted TFP shock from Fernald (2012). I calculate an oil price shock, following Hamilton (2003), and extend the Romer & Romer (2004) measure of monetary policy shocks. I consider several proxies for uncertainty: the macroeconomic policy uncertainty of Baker et al. (2016); realized volatility of the S&P 500 stock market index; the innovation to the VXO index, as implied by a fitted AR(2) model; and the excess bond premium of Gilchrist & Zakrajsek (2012). Details of the macroeconomic shocks are in Appendix A1.3.

Table 4 shows that there is considerable correlation between the output gap innovation implied by the Tealbook and other macroeconomic shocks, suggesting that the Board staff's output gap is an amalgamation of many different macroeconomic shocks. Output gap innovations are positively correlated to TFP shocks, highlighting the difficulty in differentiating between changes in potential output and the output gap. Output gap innovations are highly negatively correlated to the various uncertainty measures and positively correlated to Romer & Romer (2004) monetary policy shocks. This result is surprising given that the Romer & Romer method purges intended changes to monetary policy of macroeconomic expectations, as contained in the Tealbook projections of GDP, inflation and the unemployment rate. The correlation to the net oil price increase is modest and negative. The table also reports that the uncertainty shock from the FSV model is less correlated with the macroeconomic shocks considered in the table. The uncertainty shock is negatively correlated with productivity shocks and positively correlated with oil price increases, but is not closely correlated to the other macroeconomic innovations.

Table 4: Correlation of FSV innovations with other macroeconomic shocks.

Macroeconomic shock	Output gap innovation	Uncertainty innovation
Productivity (Fernald, 2012)	.48	-.19
	[.00]	[.01]
Net oil increase (Hamilton, 2003)	-.15	.30
	[.10]	[.04]
Monetary policy (Romer-Romer, 2004)	.25	-.18
	[.03]	[.23]
Policy uncertainty (BBD, 2016)	-.37	-.19
	[.03]	[.11]
Realized volatility of S&P 500	-.41	.09
	[.03]	[.38]
VXO innovation	-.34	.09
	[.06]	[.28]
Excess bond premium	-.53	.05
	[.00]	[.16]

Notes: Table gives pairwise correlation coefficient between innovation to FSV innovation and each identified macroeconomic shock. Sample period is 1984Q1–2013Q4. Bracketed value is the p-value from a univariate regression of the output gap innovation onto each macroeconomic shock, calculated using standard errors robust to heteroskedasticity and autocorrelation. Macroeconomic shocks described in Appendix A1.3. See text for details.

To summarize, table 4 suggests that the judgmental output gap estimates of the Federal Reserve Board staff is in fact a blend of macroeconomic shocks. The output gap of the Federal Reserve Board appears to incorporate a wide variety of shocks to macroeconomic activity.

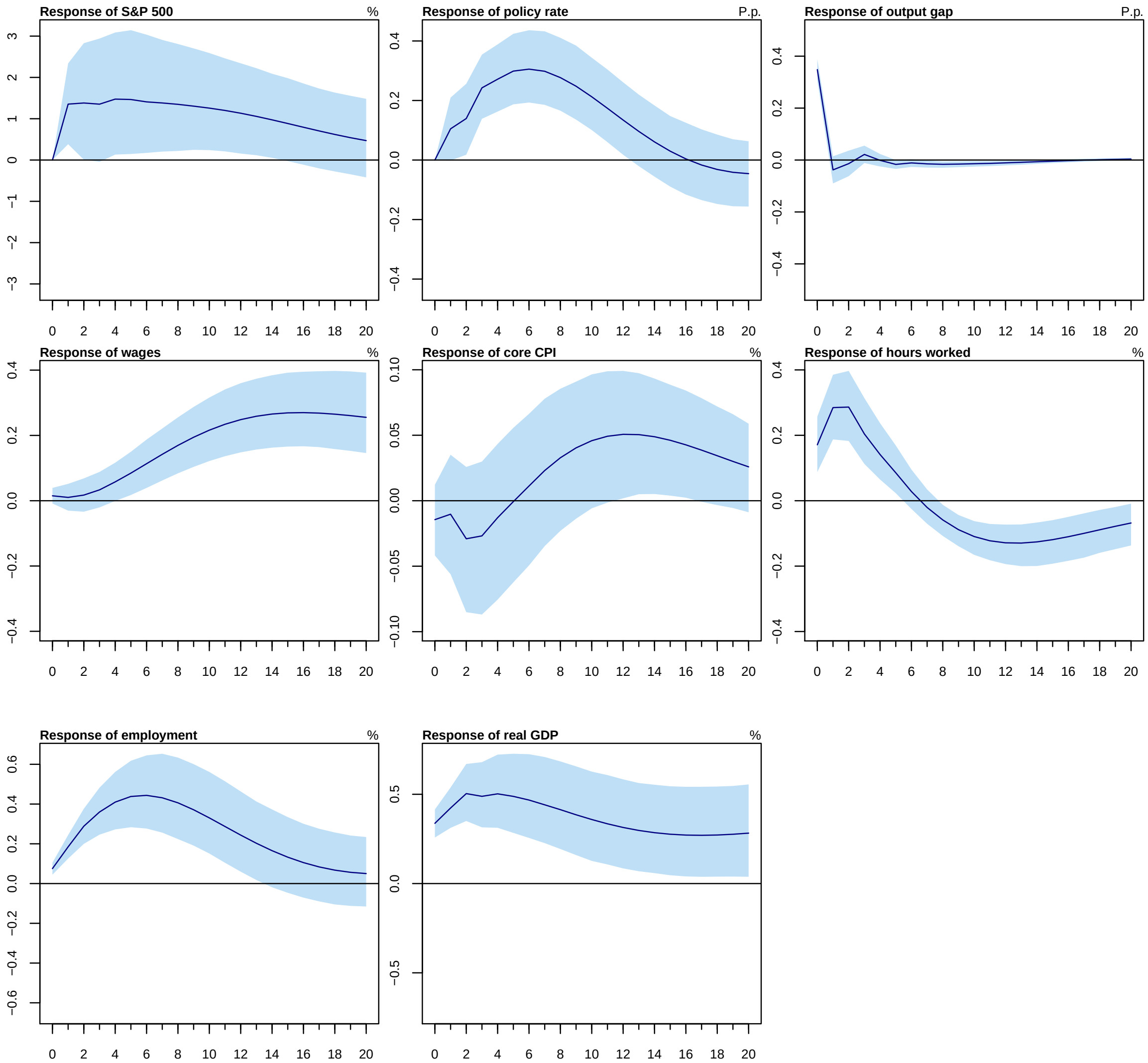
5.2 Response of other macroeconomic variables

Given that the output gap appears to be comprised of a variety of different macroeconomic shocks, I now consider how macroeconomic variables respond to an output gap innovation through the lens of a recursively identified vector autoregression model. The variables in estimation order are: the S&P500 stock market index, the Wu & Xia (2016) shadow policy rate, the output gap innovation, average hourly earnings, the core consumer price index, hours, employment, and real GDP. The ordering allows real economic variables to respond contemporaneously to output gap innovations, but ensures that any changes in equity markets or the policy rate are controlled for when calculating the impact of output gap innovations. I denote this eight variable VAR as VAR-8. The VAR is estimated over the period 1984Q1–2013Q4.[13]

The solid lines in figure 7 show the impulse responses to a one standard deviation (.35 pp) shock to the output gap. The shaded areas represent 90 percent confidence intervals. The responses are consistent with what may be denoted an aggregate demand shock. Both real and nominal variables increase in a persistent manner. Focusing on the real side, hours worked, employment, and output all respond contemporaneously to an output gap innovation. The response of real GDP to an output gap 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 output gap and TFP innovations. The response of employment is qualitatively similar to output, whereas hours worked exhibit overshooting. That hours overshoot may reflect hiring frictions in the labor market as in, e.g., (Bloom 2009). The nominal variables in the VAR respond positively to an output gap innovation but with some delay, consistent with sticky prices and wages as is standard in the New Keynesian literature. The response of wages is unambiguously positive but takes about a year to manifest. Price inflation also responds positively to an output gap innovation, but in a muted manner and with notable uncertainty. Finally, stock market prices and the policy rate increase following a positive innovation to the output gap. The policy rate tightens as economic activity increases and price and wage pressures build. Interestingly, the response of the policy rate to the output gap is very close to the one-to-one response prescribed by the Taylor (1999) rule.

[13] Data sources are reported in appendix A1. Several alternative VAR specifications are provided in the appendix.

Figure 7: Response to output gap innovation, VAR-8.



Notes: Impulse response from quarterly eight variable VAR; output gap innovation ordered third. Estimated over period 1984Q1–2013Q4. Figures show response to one standard deviation innovation to output gap (sample period average). Shaded area denotes 90 percent confidence bands from Kilian (1998)'s bootstrap-after-bootstrap. See text for details.

The top panel of table 5 reports the forecast error variance decomposition for wage and price inflation, employment, and real GDP. Output gap shocks are associated with large fractions of forecast error variance for real activity. Shocks to the output gap, for example, are associated with 15-25 percent of the forecast error variance in employment, and between roughy 20-30 percent of real output. Output gap shocks explain sizable forecast error variance of wages, but only longer horizons: at the four quarter horizon, output gap is associated with less than 1 percent of the error variance whereas 20 quarters hence that number increases to more than 20 percent. In contrast, the error variance of inflation is hardly explained by the output gap.

Table 5: Forecast error variance decomposition for selected macroeconomic variables.

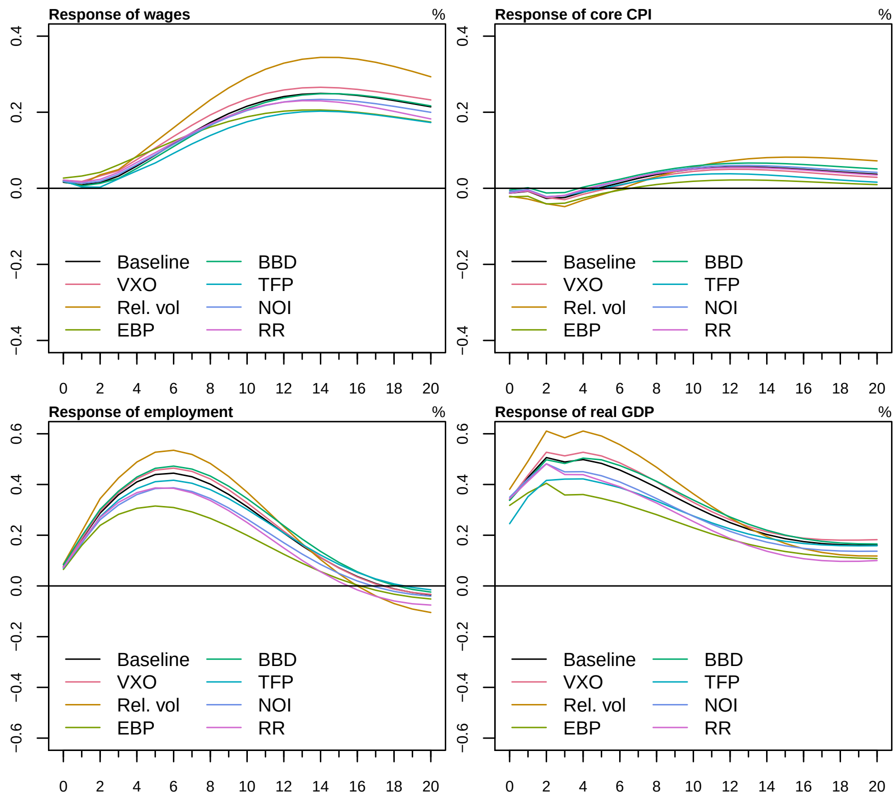
	Quarters ahead				
	4	8	12	16	20
Fraction explained by output gap (%)					
Wages	0.7	6.8	14.9	20.0	22.3
Prices	0.6	0.4	1.1	1.8	2.1
Employment	24.3	21.8	19.0	17.1	16.4
Real GDP	32.0	25.2	21.6	19.5	18.5
Fraction explained by employment (%)					
Wages	0.1	2.6	7.2	10.9	13.0
Prices	0.8	1.2	0.9	0.8	0.7
Employment	32.3	23.6	21.1	19.7	19.0
Real GDP	13.2	16.4	16.5	16.2	16.0
Fraction explained by real GDP (%)					
Wages	0.2	0.8	2.4	3.8	4.9
Prices	0.2	0.3	1.1	2.3	3.3
Employment	0.6	1.0	0.9	0.8	0.8
Real GDP	12.9	8.0	6.4	5.5	5.0
Fraction explained by policy rate (%)					
Wages	19.0	30.0	28.0	22.7	18.4
Prices	3.8	6.8	8.4	8.0	7.1
Employment	2.7	1.1	0.9	1.4	1.7
Real GDP	1.5	0.7	1.1	1.3	1.2

Notes: See text for details.

To put these results in perspective, the remainder of the table reports the error variance attributed to employment, real GDP, and the policy rate. The output gap is similar to employment innovations in terms of the error variance explained. Both the output gap and employment explain roughly 20 percent of the error variance of employment and real GDP at longer horizons, although the output gap explains a larger fraction of the error variance of wages than does employment. Real GDP explains relatively little of the error variance of the other variables, especially beyond a couple of years. Finally, the fraction explained by shocks to the real economy are quantitatively much more important than those to the policy rate for core CPI, employment and real GDP.

Since output gap innovations are correlated with other macroeconomic shocks, I next explore the whether the results from the VAR-8 are robust to the inclusion of the other identified macroeconomic shocks. Specifically, I reestimate the VAR presented above, but add each of the other macroeconomic shocks considered in table 4 to the VAR. The other macroeconomic shock is always ordered above the output gap shock in the VAR, so that the responses to output gap innovations always condition on the other macroeconomic innovation. Figure 8 shows that the results are robust to the inclusion of these alternative shocks.

Figure 8: Response of selected variables to output gap innovation, VAR-9.



Notes: Impulse response from nine variable VAR. Order of VAR is: S&P 500; Wu-Xia policy rate; other macroeconomic shock; output gap shock; wages; prices; hours worked; employment; real GDP. Sample period 1984Q1–2013Q4. Figures show response to one standard deviation innovation to output gap. See text for details.

I have also examined the robustness of the VAR by performing a number of robustness checks, which can be found in the appendix.

- Figure A1 shows the impulse response functions from the recursively estimated VAR with the output gap innovation ordered first. The results are very similar to the baseline.
- Figure A2 shows the impulse response functions from the recursively estimated VAR with the output gap innovation ordered last. The responses of other macroeconomic variables are less pronounced, but they 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 A3 estimates the same VAR but using only the pre-recession period (1984–2006). The impulse responses are very similar to the baseline, suggesting that the Great Recession is not responsible for the estimated impulse response functions.

- For completeness, I show the response of the baseline VAR-8 to innovations to real GDP and employment, (figures A4 and A5). The qualitative pattern of the responses of other macroeconomic variables are similar to output gap innovation, with some important differences. Most importantly, the output gap innovation does not respond strongly to innovations to either real output or employment. Qualitatively, an innovation to employment produces similar impulse responses as the output gap innovation, while the innovation to real GDP generally produces more muted responses.

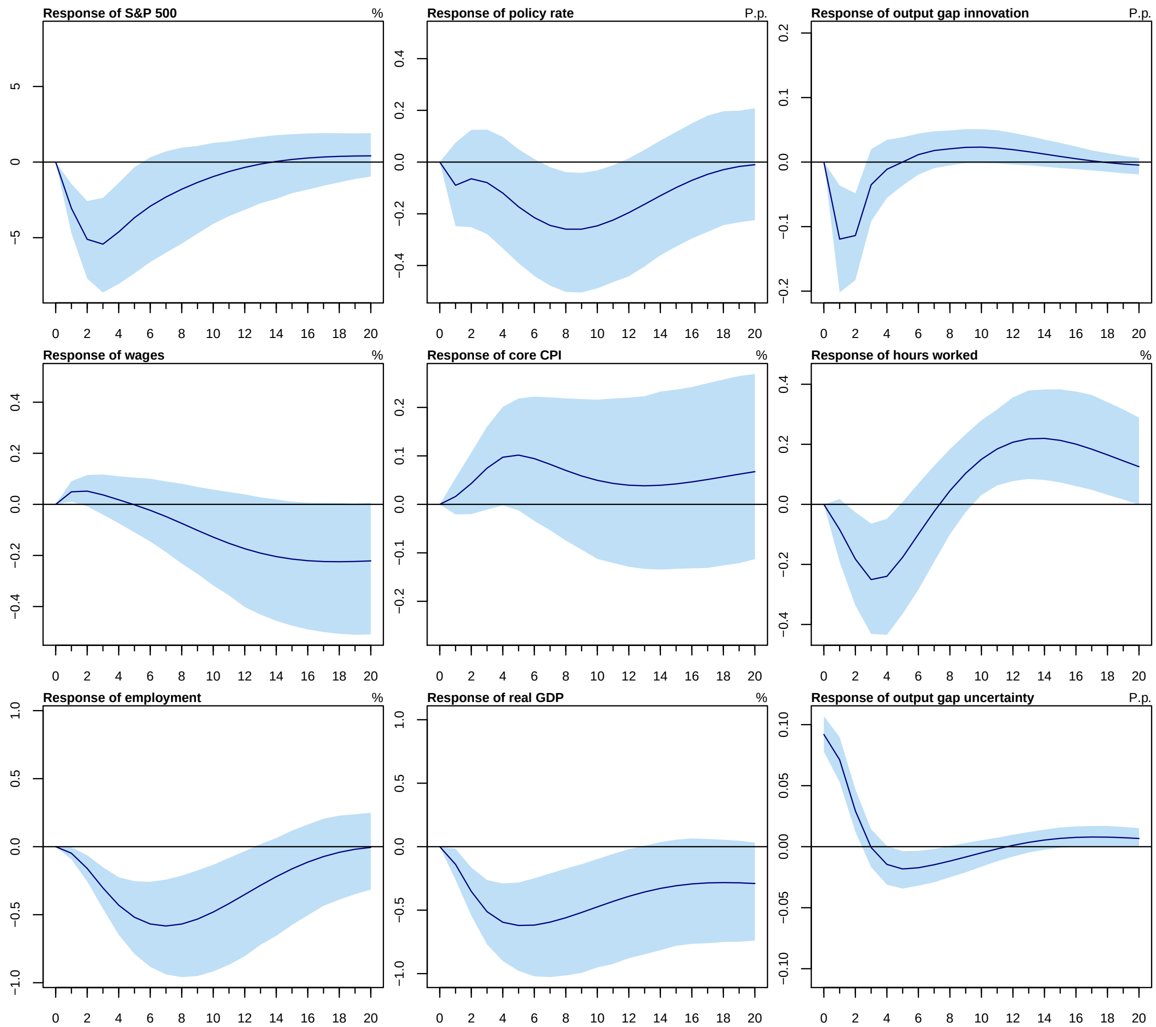
- The responses of the macroeconomic variables to an output gap innovation are nearly identical to the baseline VAR when core CPI is replaced with total CPI (figure A6).

5.3 Response to macroeconomic uncertainty

The FSV framework also provides a measure of macroeconomic uncertainty, namely, the innovation to the log-volatility of the FSV output gap, h_{ft}. The measure is of particular interest given that the innovations to the volatility of the output gap are orthogonal to the shocks to the output gap itself. To be conservative, I place the output gap uncertainty shock last in the VAR.

The results from the VAR are shown in figure 9. The final panel shows that a one standard deviation uncertainty shock gradually dies out over the course of about one year. As shown in the third panel, the output gap responds negatively to this volatility shock. Given the output gap responds in a negative fashion, the remainder of the VAR is analogous to the baseline VAR. Real output and employment decline and obtain a local nadir six to eight quarters after the shock. Within the labor market, employment and hours worked both decline initially, with hours worked eventually turning positive. Wages respond negatively with a delayed response, whereas core prices show a positive but uncertain response. Turning to the financial variables, the response of the S&P price index is negative and significant. And, as before, the cumulative change in the policy rate is nearly one-to-one with changes in the estimated output gap. Overall, the responses of these variables are qualitatively similar or perhaps somewhat larger than the estimated responses reported in the macro uncertainty literature, for example, Bloom (2009), Jurado et al. (2015), and Jo & Sekkel (2019).

Figure 9: Response to innovation to output gap volatility.



Notes: Impulse response to innovation to output gap volatility. Estimated over period 1984Q1–2013Q4. Figures show response to one standard deviation innovation to output gap log volatility (sample period average). Responses plotted in order of VAR estimation. Shaded area denotes 90 percent confidence bands from Kilian (1998)'s bootstrap-after-bootstrap. See text for details.

5.4 Are Tealbook output gaps different?

The VAR results above are based on an FSV model that takes as primitive the output gap estimates produced by the Federal Reserve Board staff. To understand how those Tealbook estimates compare to other output gap estimates, in this section I estimate the FSV model on real-time vintage output gap estimates from univariate detrending methods. Specifically, I estimate the model presented in 3 using output gap estimates from an HP filter and the Hamilton (2018) filter. I produce real-time output gap estimates from the HP filter and Hamilton filter models since 1962Q4.[14] For comparison, I also re-estimate the model based on Tealbook projections over this extended period.

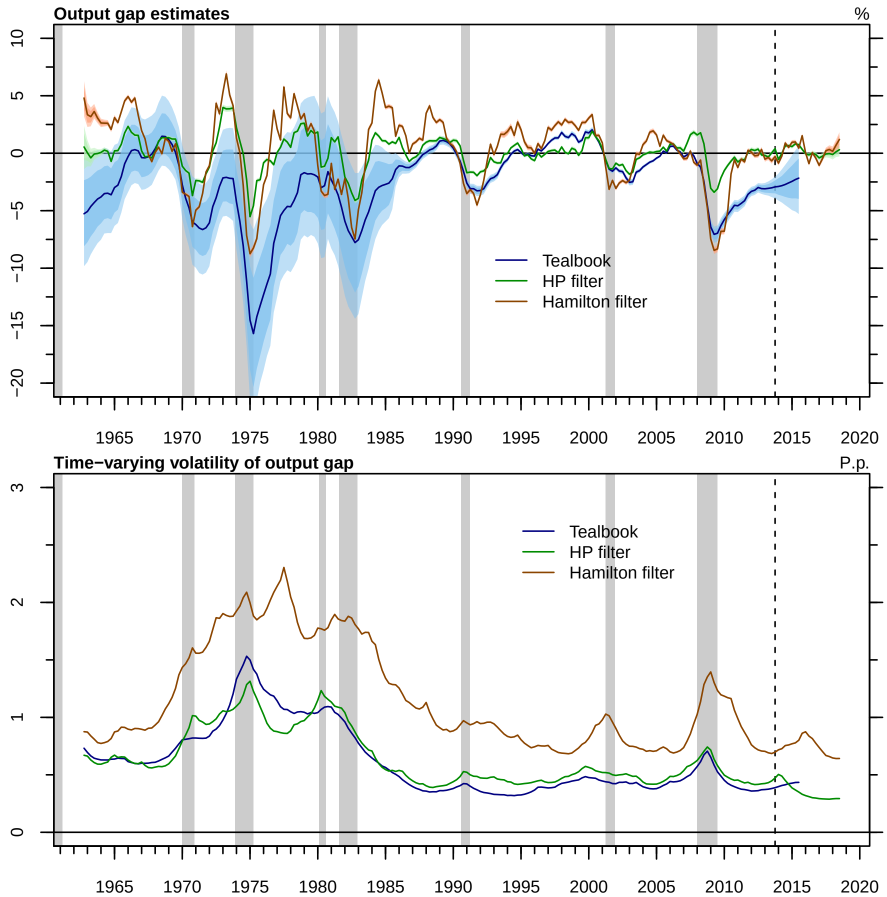
The three estimates are shown in figure 10. The estimates of the common factors derived from the HP-FSV and Hamilton-FSV models are quite similar to one another, with a contemporaneous correlation of about .8. As before, the two series have quite narrow credible intervals. In contrast, the Tealbook-derived FSV model has quite wide credible sets, especially early in the estimation period when there is only one measurement for each quarter's output gap estimate.

There are two important differences between the Tealbook-derived FSV output gap and those derived from the HP and Hamilton filters. First, the Tealbook output gap is negative for all of the period from 1970 until the late 1980s, whereas the two other filters produce output gaps that oscillate around zero in this period. The other notable difference is that the Tealbook-FSV output gap estimate recovers much more slowly following recessions than the HP or Hamilton filter-derived estimates. The bottom panel of the figure shows the estimated volatilities. The HP filter and Tealbook output gaps produce very similar output gap volatility estimates, whereas the Hamilton filter output gaps are somewhat more volatile.

In order to discern among these three output gap estimates, figure 11 presents the VAR-8 results using for the three output gaps shown in figure 10, alongside the baseline VAR-8 results explored above. The figure shows important differences in the response of other macroeconomic variables to the output gap innovations. First, when estimated over over the entire sample, responses to the Tealbook-based output gap changes, likely reflecting both changes in the Board staff output gap estimate itself and changes to the macroeconomy more broadly. Most importantly, the response of price inflation to an output gap innovation is much different when the model is estimated using the longer sample: core inflation shows a delayed but very strong response to the output gap innovation. The response of hours, employment, and output to an output gap shock are all positive upon impact for the full-sample model, but then exhibit overshooting.

[14] The initial real-time vintage for real GDP growth is 1965Q4, which contains estimates of real GDP from 1947Q1–1965Q3. Since the maximum backcast horizon I consider is 12 quarters, the earliest quarter with non-missing gap estimates is 1962Q4.

Figure 10: Comparison of output gap estimates, 1962Q4–2018Q4.

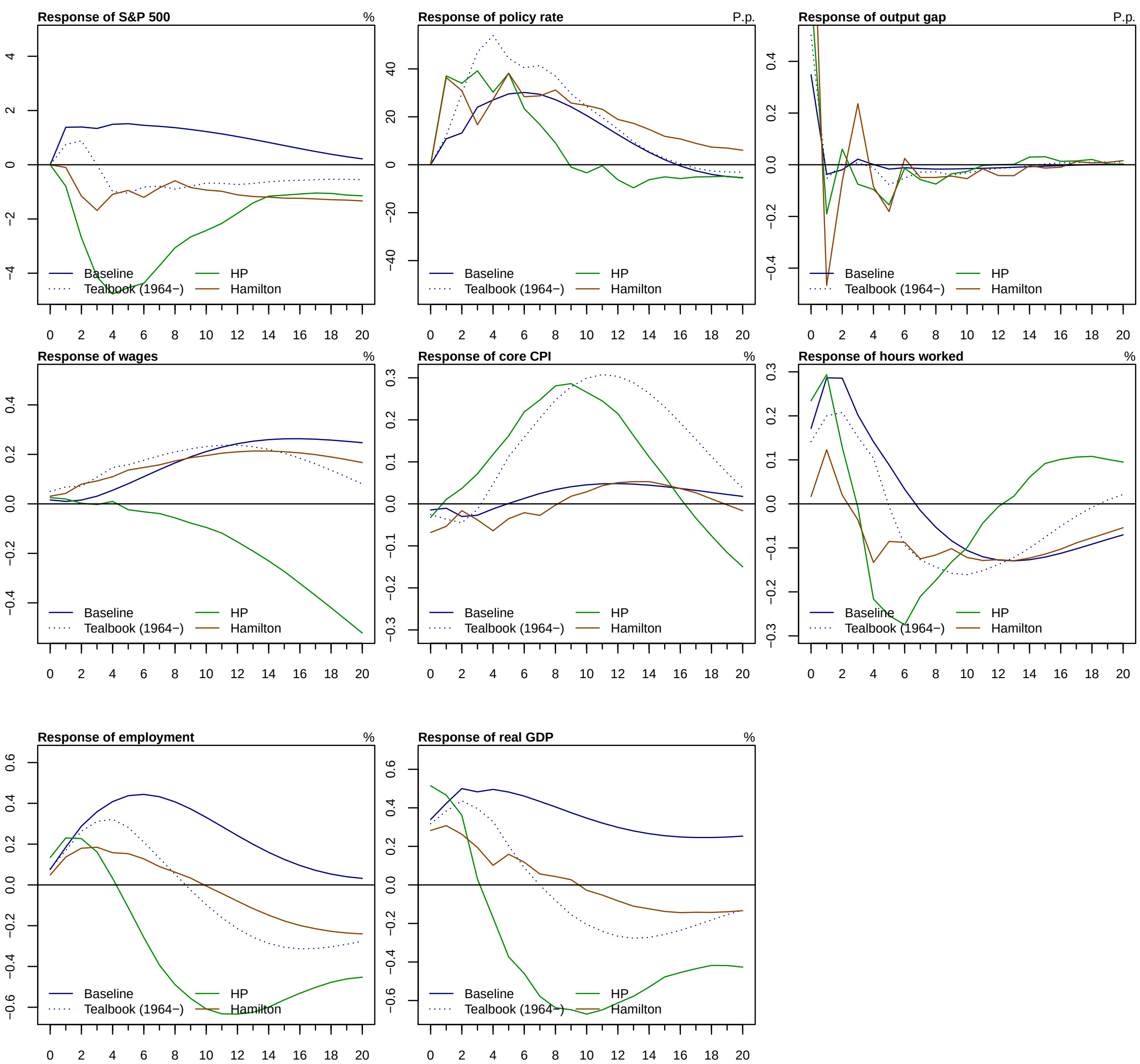


Notes: Top panel shows estimated latent factor from vintage estimates of Tealbook, HP filter and Hamilton filter. Dark and light shaded areas are 70 and 90 percent credible intervals. Bottom panel shows posterior median estimate of standard deviation of estimated common factor (credible sets suppressed in interest of legibility.) Gray shaded areas denote NBER defined recessions. Vertical dashed line indicates 2013Q4, the final Tealbook estimate. See text for details.

The responses of the innovations derived from the HP or Hamilton FSV models clearly differ from the Tealbook FSV innovations. Focusing first on the Hamilton-FSV model, real output and employment both show a positive and fairly persistent response to an output gap innovation, although the response is muted compared to the Tealbook-based models. The response of wages and prices to a Hamilton output gap innovation is very similar to that from the Tealbook. However, hours worked displays, on net, a negative response to an output gap innovation. Also counterintuitively, the S&P 500 falls in response to a positive Hamilton output gap innovation. The responses of macroeconomic indicators to output gaps as derived from the HP filter are also puzzling. Hours, employment and output respond positively on impact, but turn sharply negative thereafter. Perhaps the most puzzling response to the HP FSV output gap innovation is the response of wages, which decline following a positive shock to the output gap, even though prices show a positive, hump-shaped response.

In all, it is difficult to characterize the responses of the macroeconomic variables following the output gap shocks when derived from the HP or Hamilton filters. These puzzling results suggest that the output gaps derived from the Tealbook projections are more desirable as a measure of aggregate demand than those derived from the univariate filters.

Figure 11: Comparison of response to different output gap innovations.



Notes: Figure shows a comparison of the impulse response from the quarterly eight-variable VAR. Each VAR differs only in the estimated output gap used: Tealbook (post-1984 sample or full sample), HP filter, or Hamilton filter-based output gap estimate from FSV model. Figures show response to one standard deviation output gap innovation. See text for details.

6 Discussion

This paper is an extensive empirical analysis of the Federal Reserve's output gap estimate. First, I show that output gap estimates are quite unstable, and provide forecast horizon specific measures of that instability. I then develop a model that treats successive output gap estimates as mismeasured estimates of a true but latent output gap. By endowing both the latent output gap and the forecast errors thereof with stochastic volatility, I measure the uncertainty that surrounds the output gap, and the time-varying uncertainty of horizon-specific output gap forecasts.

The FSV model using historic Tealbook output gap estimates produces a clearly identified output gap estimate, although the estimates themselves are quite uncertain. The 70 percent credible interval surrounding the Tealbook output gap nowcast at the end of 2013 is roughly 1.5 percentage points wide. Forecasts are even more uncertain: the eight quarter ahead Tealbook output gap forecast at the end of 2013 has a 70 percent credible region of more than 4 percentage points. Nevertheless, the results indicate that the Tealbook output gap estimate measures a real macroeconomic phenomenon. Innovations to the latent output gap are associated with sizable increases in wages, employment, and output. Monetary policy responds endogenously to output gap innovations. I also find evidence that macroeconomic uncertainty shocks—defined as shocks to the variance of the true latent output gap—are associated with adverse outcomes in the macroeconomy.

Because the common factor estimated from the FSV model is dependent on the output gap vintage estimates that are taken as primitive by the model, I am able to compare output gap estimates across models. Whereas I find that the macroeconomic responses are generally quite sensible, the output gaps from the HP and Hamilton filters produce macroeconomic responses that are counterintuitive, indicating that the output gap measured by those models in real-time is gravely mismeasured.

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.

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

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

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.

Del Negro, M. & Otrok, C. (2008), Dynamic factor models with time-varying parameters: measuring changes in international business cycles, Staff Reports 326, Federal Reserve Bank of New York.

Del Negro, M. & Primiceri, G. E. (2015), 'Time varying structural vector autoregressions and monetary policy: A corrigendum', *Review of Economic Studies* **82**, 1342–1345.

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.

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.

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`.

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

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.

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.

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.

Taylor, J. B. (1999), A Historical Analysis of Monetary Policy Rules, *in* J. B. Taylor, ed., 'Monetary Policy Rules', University of Chicago Press, pp. 319–347.

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 now 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-center/greenbook-data/gap-and-financial-data-set`.

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

First available *nowcast* vintage	January 1951
First vintage with near-term path	August 1987
First vintage containing complete path	May 1996
Most recent output gap vintage	December 2013

A1.2 Real-time real GDP estimates

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

- I use quarterly real-time estimates. Each quarterly vintage is the estimate of real GDP available in the middle month of each quarter. One vintage, 1996Q1, produces missing values for the nowcasts from statistical models, as a government shutdown delayed the release of NIPA data for 1995Q4.
- The data is 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 (three year) horizon. The net oil increase is calculated as the percentage amount by which the oil price, measured as the producer price index for oil, in a quarter exceeds the previous peak over the past 3 years.

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).[15] 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 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 three measures of financial uncertainty. The first is a measure of quarterly realized volatility constructed from daily returns of the S&P 500 index. The second is the innovation to an AR(2) model fit to the VXO index, available from 1985. Finally, I consider the excess bond premium from Gilchrist & Zakrajsek (2012), a bond premium that has been adjusted for predictable default risk.

[15] 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/`.

Table A2: Data sources for macroeconomic shocks.

Shock	Period available	Citation	Source
Productivity	1984Q1–2013Q4	Fernald (2012)	`http://www.johnfernald.net/TFP`
Oil price	1984Q1–2013Q4	Hamilton (2003)	Author's calculations
Monetary policy	1984Q1–2013Q4	RR (2004)	Author's calculations
Policy uncertainty	1984Q1–2013Q4	BBD (2016)	`https://www.policyuncertainty.com`
Financial uncertainty			
Realized vol.	1984Q1–2013Q4	Yahoo	`https://finance.yahoo.com/`
VXO	1986Q3–2013Q4	CBOE	FRED (VXOCLS)
Ex. bond premium	1984Q1–2013Q4	GZ (2012)	`http://people.bu.edu/sgilchri/Data/`

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

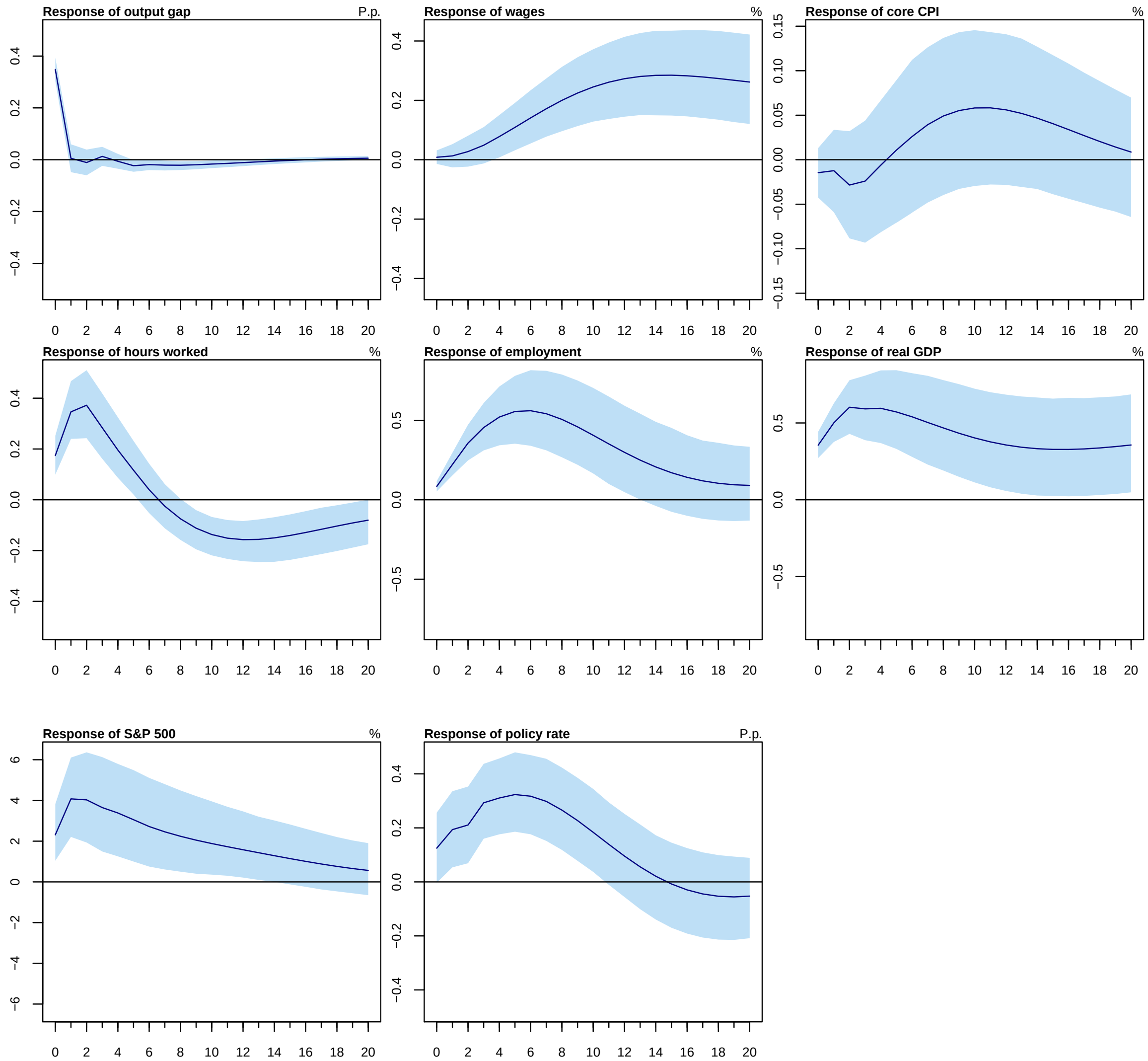
The data used in the VAR exercise is well known. The primary source for the data is FRED, with the exceptions of the S&P 500 index and the Wu & Xia (2016) shadow rate, which are taken from Yahoo and the Federal Reserve Bank of Atlanta, respectively.

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

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

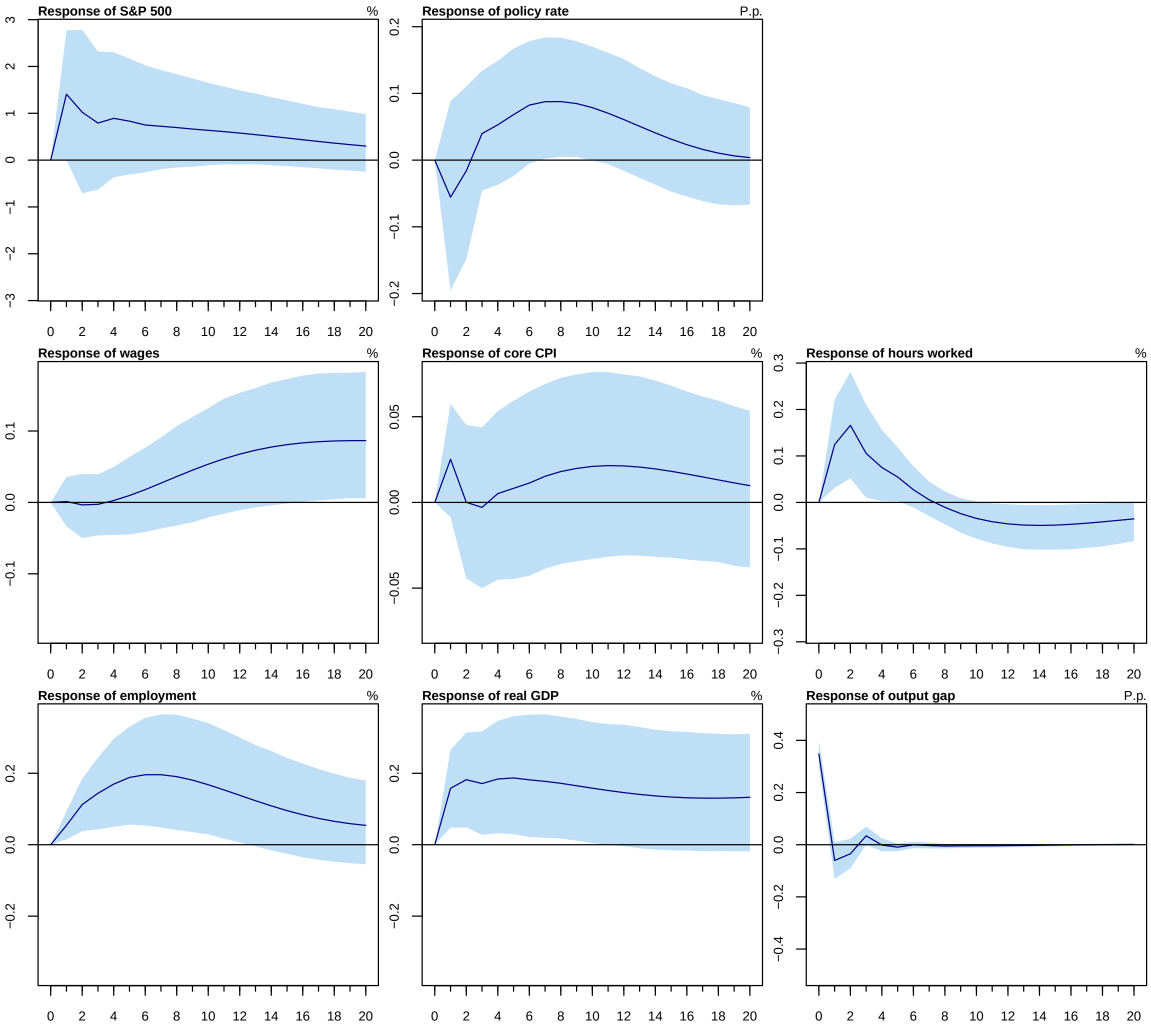
Notes: Wu-Xia shadow rate available from Federal Reserve Bank of Atlanta: `https://www.frbatlanta.org/cqer/research/shadow_rate.aspx`. Mnemonics in parentheses denote Yahoo or FRED mnemonic.

Figure A1: Response to output gap innovation, VAR-8, output gap innovation ordered first.



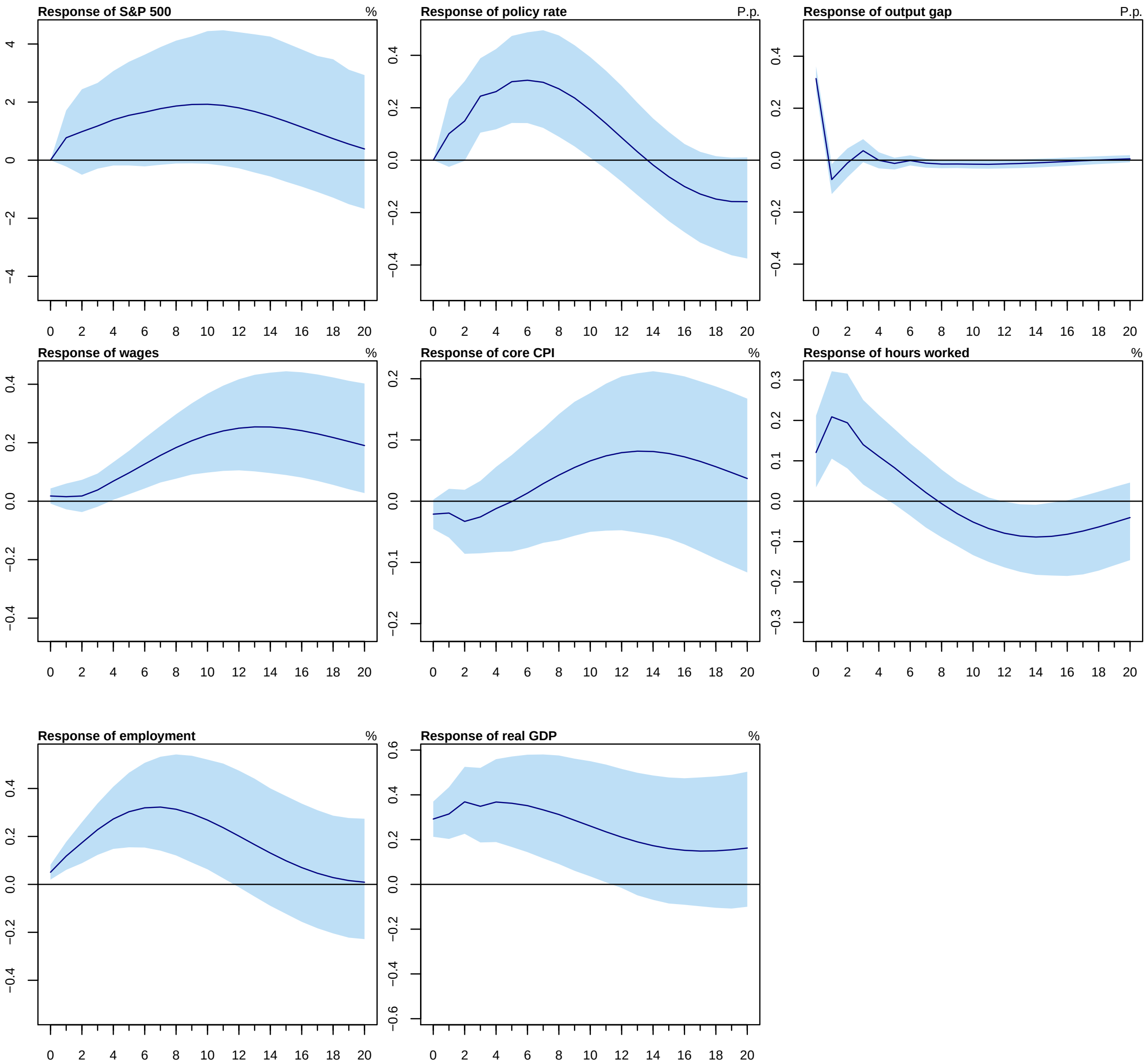
Notes: Impulse response from eight variable VAR; output gap innovation ordered first. Sample period 1984Q1–2013Q4. Figures show response to one standard deviation innovation to output gap (sample period average). Shaded area denotes 90 percent bootstrapped confidence interval. See text for details.

Figure A2: Response to output gap innovation, VAR-8, output gap innovation ordered last.



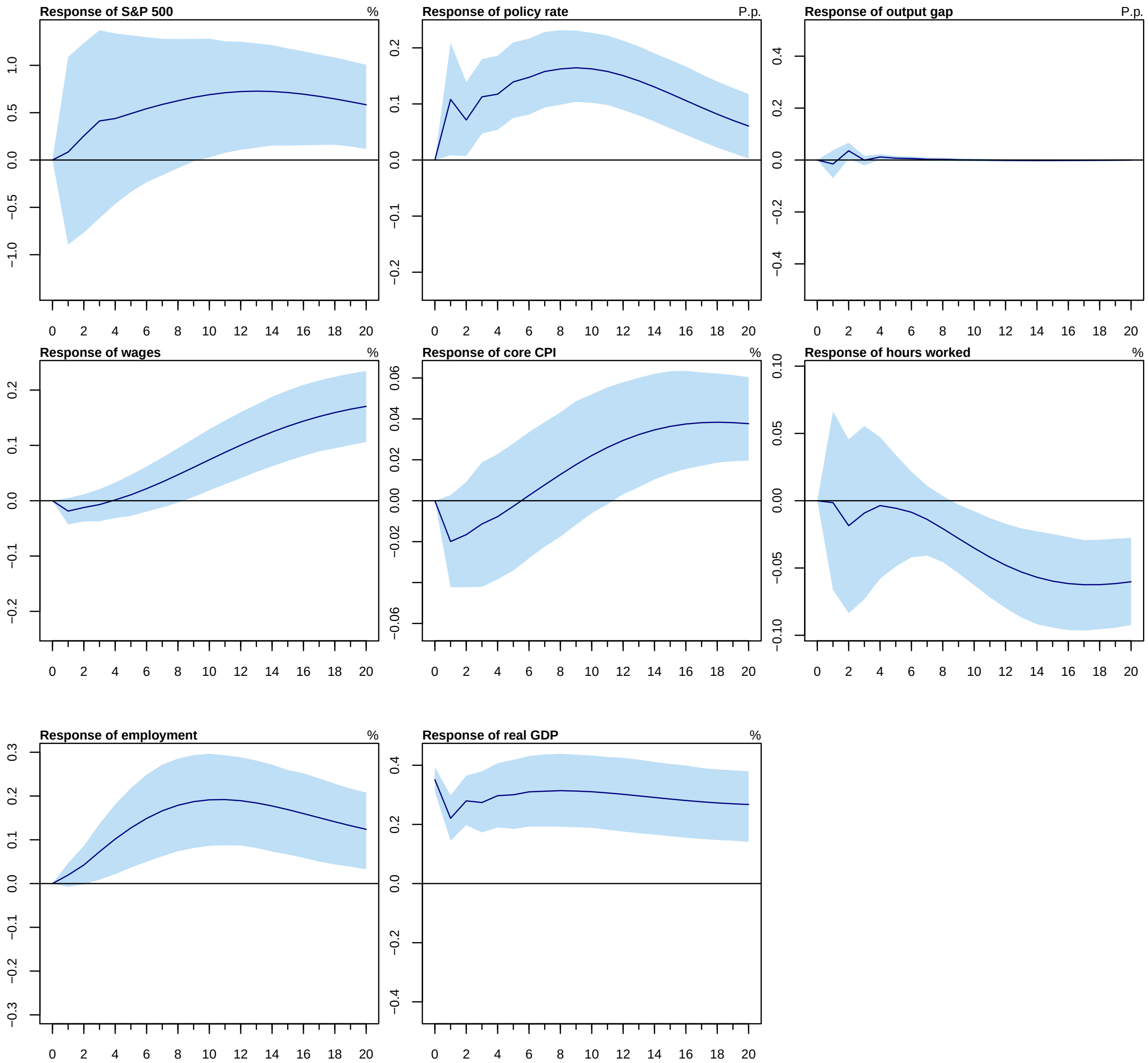
Notes: Impulse response from eight variable VAR; output gap innovation ordered last. Sample period 1984Q1–2013Q4. Figures show response to one standard deviation innovation to output gap (sample period average). Shaded area denotes 90 percent bootstrapped confidence interval. See text for details.

Figure A3: Response to output gap innovation, VAR-8, short sample period.



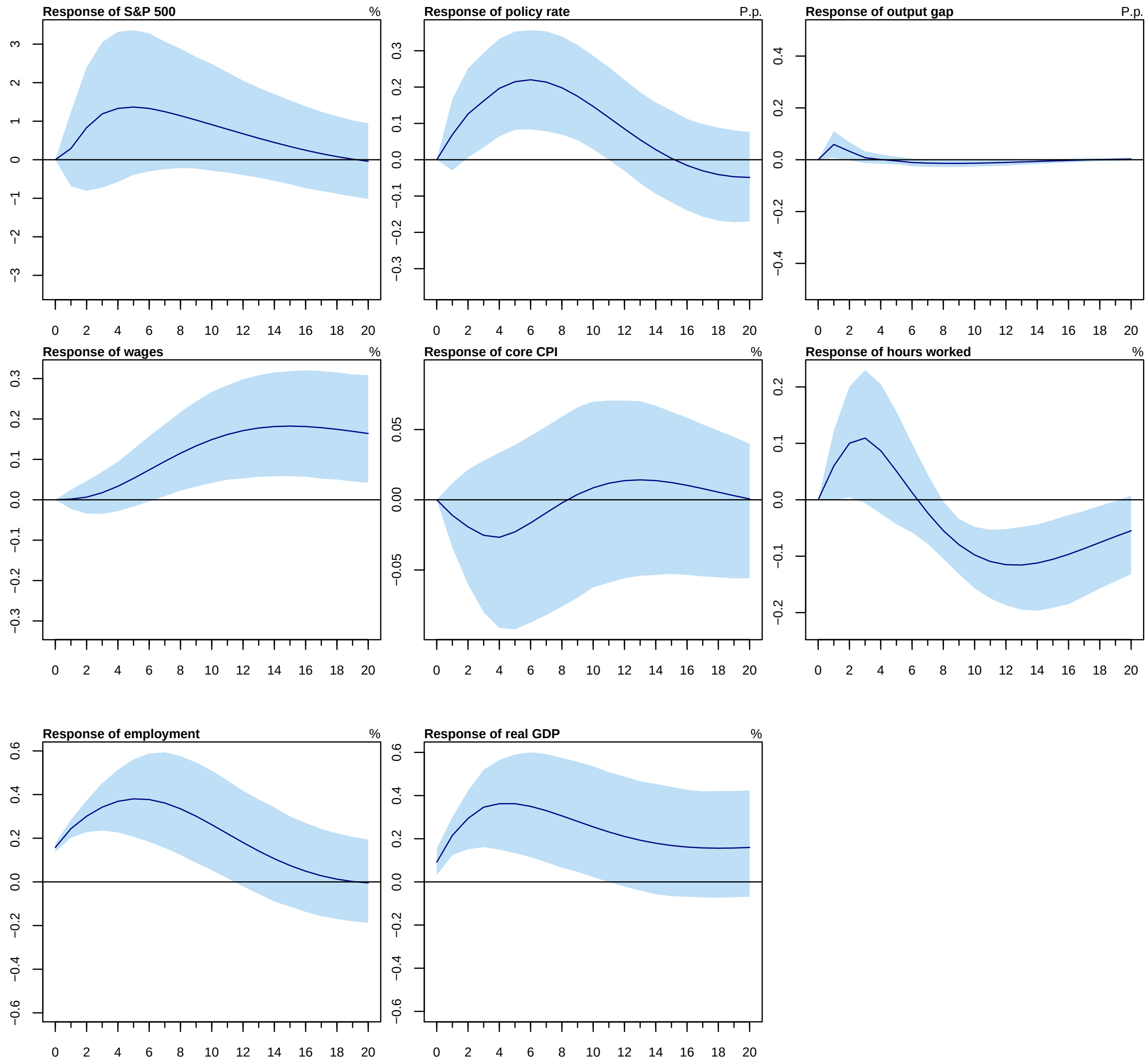
Notes: Impulse response from eight variable VAR. Sample period 1984Q1–2006Q4. Figures show response to one standard deviation innovation to output gap. Shaded area denotes 90 percent bootstrapped confidence interval. See text for details.

Figure A4: Response to GDP innovation, VAR-8.



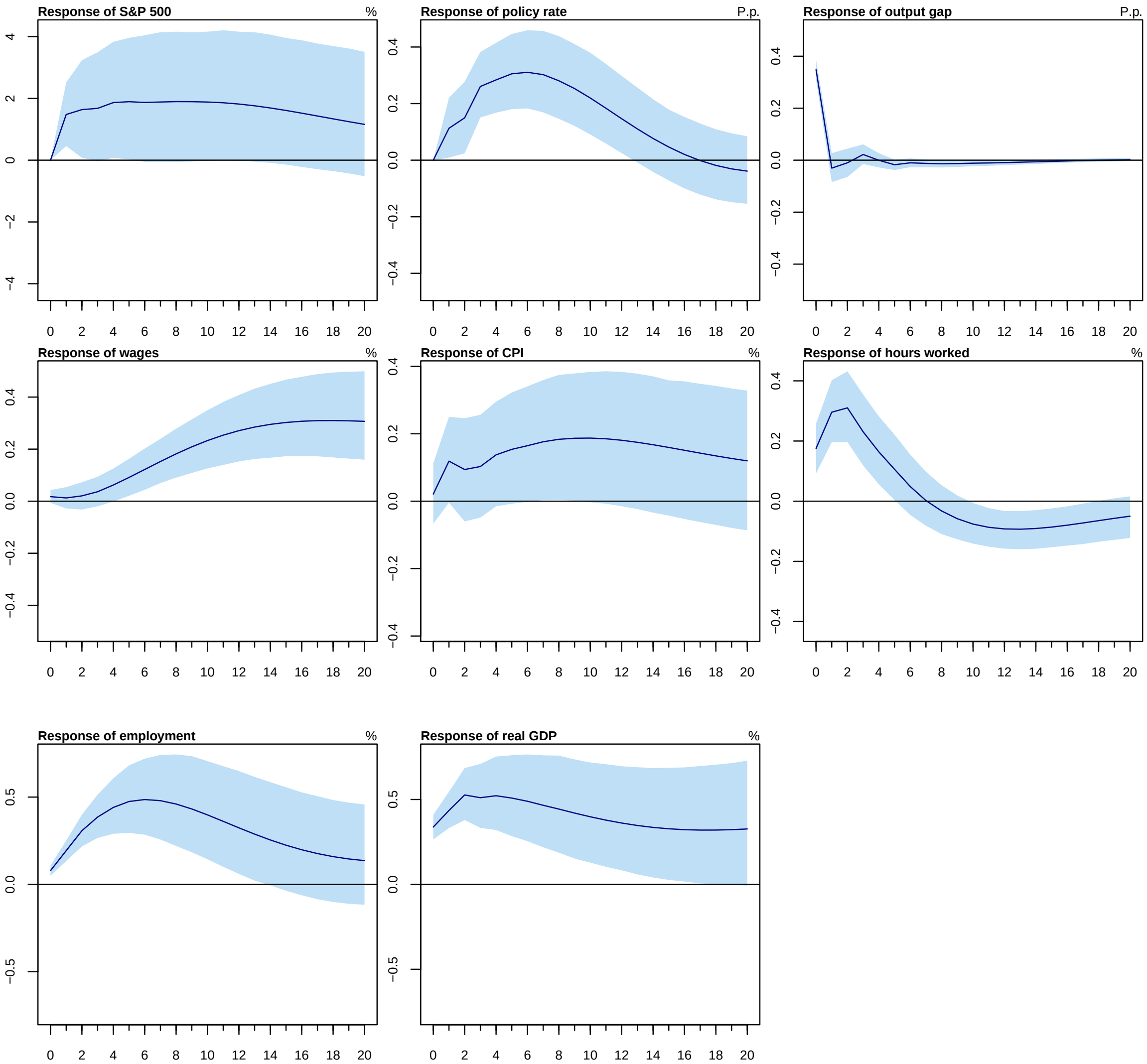
Notes: Impulse response from eight variable VAR. Sample period 1984Q1–2013Q4. Figures show response to one standard deviation innovation to real GDP. Shaded area denotes 90 percent bootstrapped confidence interval. See text for details.

Figure A5: Response to employment innovation, VAR-8.



Notes: Impulse response from eight variable VAR. Sample period 1984Q1–2013Q4. Figures show response to one standard deviation innovation to employment. Shaded area denotes 90 percent bootstrapped confidence interval. See text for details.

Figure A6: Response to output gap innovation, VAR-8, total CPI.



Notes: Impulse response from eight variable VAR, uses total CPI to measure inflation. Sample period 1984Q1–2013Q4. Figures show response to one standard deviation innovation to output gap (sample period average). Shaded area denotes 90 percent bootstrapped confidence interval. See text for details.

A2 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 four blocks: the parameters for the autoregressive processes; the variance of the volatilities (the σ's); the time-series of the state vector; and, the volatility states ($h_{y^*,t}$ and and $h_{i,t}$) for $T = 1, ..., T$ and all i. I describe the algorithm below.

Log volatilities, h_{it}. Conditional on all parameters, both u_{ft} and $u_{et|i}$ can be treated as observable. Focusing on u_{ft}, we have the measurement equation and its corresponding transition equation:

$$\begin{aligned} \phi(L)f_t &= e^{h_{it}/2} \\ h_{ft} &= h_{ft-1} + \log \eta_t^2 \end{aligned}$$

Kim, Shephard & Chib (1998) and Chib, Omori, Nakajima & Shephard (2007) provide a routine for drawing the time-series of h_{ft} involving the forward-filter backwards-sampling approach of Carter & Kohn (1994), and approximating the χ^2 distributed term (η^2) with a mixture of Gaussians. See as well Primiceri (2005) and Del Negro & Primiceri (2015). The log-volatilities of various horizon-specific measurement errors are drawn in an analogous manner.

Factor loadings, λ. Conditional on the data and other parameters, factor loadings are drawn from Bayesian regression of output gap estimates on the factor with known heteroskedastic error structures. Moreover, because all correlations are by definition captured by the factor, one can draw each of the n factor loadings separately:

$$\lambda_i \sim N(\mu_{\lambda_i}, V_{\lambda_i}),$$

with $V_{\lambda_i} = (V_{0\lambda_i}^{-1} + \sum_t f_t/h_{it})^{-1}$ and $\mu_{\lambda_i} = V_{\lambda_i}(V_{0\lambda_i}^{-1}\mu_{0\lambda_i} + \sum_t f_t \times \hat{y}_{t|t+i}/h_{it})$.

Autoregressive coefficients of the latent factor, Φ. Conditional on the state vector and the history of log-volatilities, the autoregressive parameters of the latent common factor can be drawn from standard Bayesian regressions with known heteroskedastic errors. Defining $\tilde{y} = f_t/h_{ft}$

and $\tilde{x}_t = [1/h_{ft}, f_{t-1}/h_{ft-1}, f_{t-2}/h_{ft-2}]'$, then:

$$\phi \sim N(\mu_\phi, V_\phi),$$

with $V_\phi = (V_{0\phi}^{-1} + \tilde{x}'\tilde{x})^{-1}$ and $\mu_\phi = V_\phi(V_{0\phi}^{-1}\mu_{0\phi} + \tilde{x}'\tilde{y})$.

Autoregressive coefficients of measurement errors, ρ_i. Conditional on the state vector and the history of log-volatilities, the autoregressive parameters of the measurement errors can be drawn from standard Bayesian regressions with known heteroskedastic errors. Define $y_t^* = e_{it}/h_{it}$ and $x_t^* = e_{it-1}/h_{it-1}$. Then:

$$\rho_i \sim N(\mu_{\rho_i}, V_{\rho_i}),$$

with $V_{\rho_i} = (V_{0\rho_i}^{-1} + x^{*'}x^*)^{-1}$ and $\mu_{\rho_i} = V_{\rho_i}(V_{0\rho_i}^{-1}\mu_{0\rho_i} + x^{*'}y^*)$.

Variance of log volatilities, σ_i^2. Since we model each stochastic volatility to follow a unit-root process, σ_i^2 is drawn from

$$\sigma\ IG(\frac{v_1}{2}, \frac{\delta_1}{2}) \tag{1}$$

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

State vector. Conditional on draws of the model parameters I draw the state vector using the simulation smoother Durbin & Koopman (2002); see also Jarocinski (2015).

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 1.

A2.1 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: $\rho_i \sim N(0,5)_{|\rho|<1}$.

- 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. Each draw, all stochastic log-volatilities are initialized using independent draws from $h_{i0} \sim N(1, 10)$.

Table A4: 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, 5)	-3.68	-2.32	.00	2.32	3.68
σ_i^2	IG(½,½)	.26	.48	2.20	27.96	254.31

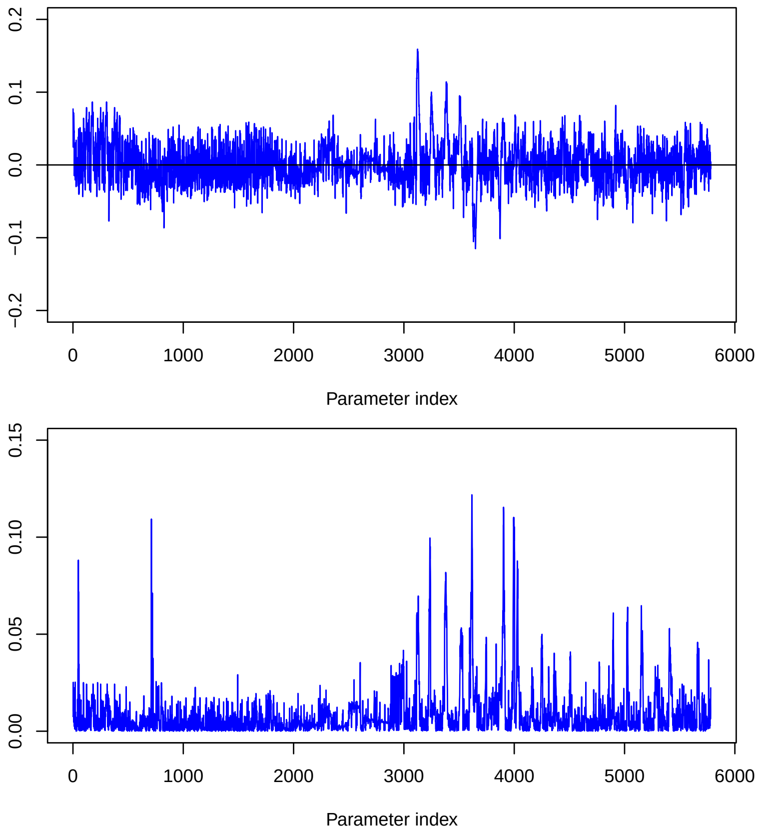
Notes: Table reports prior distributions and values of chosen percentiles. Prior for σ_i^2 is specified as $IG(\frac{v_0}{2}, \frac{\delta_0}{2})$, with v_0 and δ_0 both set to 1 for all i. Priors for λ_i and ρ_i are the same across all i.

A3 Convergence diagnostics

Following Primiceri (2005) and Baumeister & Peersman (2013), figure A7 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: estimated elements of λ, (elements 1–20); the draws of Φ (elements 21–24); the draws of ρ (elements 25–36); the diagonal elements of Σ (elements 37–59); draws of $\{S\}_{t=1}^T$ (elements 60–2921); and lastly, the draws of $\{h_t\}_{t=1}^T$ for each h_i (elements 2922–5781). The top panel displays the 20^{th} order autocorrelation across draws of the sampler, for each parameter in the model. The bottom panel plots, for each estimate listed above, the inverse of the relative numerical efficiency measure (RNE) introduced by 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 an ergodic distribution.

Figure A7: 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.