Web-appendix for:

Testing Macroeconomic Policies with Sufficient Statistics

Regis Barnichon[(a)] *and* *Geert Mesters*[(b)]

[(a)] Federal Reserve Bank of San Francisco and CEPR

[(b)] Universitat Pompeu Fabra, Barcelona GSE and VU Amsterdam

September 1, 2021

Abstract

In this web-appendix, we provide the following additional results:

S0: Clarification: Why the OPP is not a policy recommendation?

S1: Details for OPP inference

S2: Extension for arbitrary convex loss functions

S3: General implementation for real time policy evaluation

S4: Additional results for the empirical study

References to lemmas, equations, etc..., which start with a "S" are references to this document. References, which consist of only a number refer to the main text.

S0 Why the OPP is not a policy recommendation?

In this section we briefly clarify why the OPP statistic cannot, without further assumptions, be used as a policy recommendation. For convenience we consider the set-up from our simple example in Section 2:

$$\begin{aligned} Y &= \mathcal{R}(\phi)p + \Gamma(\phi)\xi \\ p &= \phi' Y + \epsilon \ , \end{aligned}$$

where the coefficients ϕ capture the response of the policy maker to $Y = (\pi, x)'$. Suppose that the policy maker proposes the policy choice $p^0 = \phi^{(0)\prime} Y^0$, i.e., uses the reaction function $\phi^{(0)}$, and then uses the policy optimality test for determining whether this policy choice is optimal, i.e. she computes

$$\delta^* = -(\mathcal{R}(\phi^{(0)})'\mathcal{R}(\phi^{(0)}))^{-1}\mathcal{R}(\phi^{(0)})'Y^0 \ .$$

If $\delta^* \neq 0$, the gradient of the loss function is different from zero and the policy choice is not optimal. Thus, provided that $\mathcal{R}(\phi^{(0)})$ is known or estimable (from a period with a stable regime with $\phi^{(0)}$), the OPP statistic can be used to detect a non-optimal policy choice. This is the point of the policy optimality test proposed in this paper. In this context, it could be tempting to use the OPP to correct the optimization failure, i.e., to use the OPP statistic as a policy prescription. Indeed (ignoring uncertainty for convenience) Proposition 1-part 2 states that the policy choice

$$p^1 = p^0 + \delta^*$$

would minimize the loss function *if* the reaction function remained fixed at $\phi^{(0)}$, i.e., *if* we adjusted p^0 by a policy shock of size $\epsilon = \delta^*$. The problem however is that the OPP statistic is not a policy shock: δ^* is a function of Y^0, and it is not orthogonal to the information set. Setting $p^1 = p^0 + \delta^*$ implies that

$$\begin{aligned} p^1 &= \phi^{(0)\prime} Y^0 - (\mathcal{R}(\phi^{(0)})'\mathcal{R}(\phi^{(0)}))^{-1}\mathcal{R}(\phi^{(0)})'Y^0 \\ &= (\phi^{(0)} - (\mathcal{R}(\phi^{(0)})'\mathcal{R}(\phi^{(0)}))^{-1}\mathcal{R}(\phi^{(0)}))'Y^0 \\ &= \phi^{(1)\prime} Y^0 \end{aligned}$$

Hence, adjusting p^0 by the OPP amounts to changing the coefficient of the rule to $\phi^{(1)} = \phi^{(0)} - (\mathcal{R}(\phi^{(0)})'\mathcal{R}(\phi^{(0)}))^{-1}\mathcal{R}(\phi^{(0)})$. However, changing the coefficient ϕ of the policy rule also changes the structure of the economy, as both $\mathcal{R}$ and Γ are functions of ϕ. Since we do not know the functional forms of $\mathcal{R}(\cdot)$ and $\Gamma(\cdot)$, i.e. since we do not know the underlying economic model, it is not possible to say whether the policy choice p_1 would ultimately lower the loss function.

Answering this question requires more structural assumptions that we do not make in this paper. To give an example, consider a framework similar to that of Leeper and Zha (2003) where there are only two regimes, the causal effects are given by

$$\mathcal{R}(\phi) = \begin{cases} \mathcal{R}^1 & \text{if} \quad \phi_1 \leq \phi^T \\ \mathcal{R}^2 & \text{if} \quad \phi_1 > \phi^T \end{cases} ,$$

where ϕ^T is some threshold that defines the regime shift based on (say) the first coefficient of the reaction function. In this setting, we can revisit the question of what happens when $\delta^* \neq 0$ and the policy maker considers $p^1 = p^0 + \delta^*$. If the new reaction coefficient $\phi^{(1)}$ does not change the causal effect $\mathcal{R}$, say because $\phi_1^{(0)}$ and $\phi_1^{(1)}$ are both $\leq \phi^T$ (or both $> \phi^T$), there is no regime change and implementing the policy p^1 instead of p^0 will lower the loss function. However, if $\phi^{(1)}$ implies a regime change and the causal effect changes from (say) $\mathcal{R}^1$ to $\mathcal{R}^2$, there is no guarantee that implementing p^1 will lower the loss function, and p^1 is no longer a valid policy recommendation.

S1 Details for OPP inference

In this section we provide the econometric details for Section 7. In particular, we provide the details for (i) dynamic causal effect estimation, (ii) estimation of forecast misspecification uncertainty, (iii) OPP implementation, (iv) preference estimation and (v) testing the reaction function.

S1.1 Inference for dynamic causal effects

We discuss a specific local projection estimator for the dynamic causal effects that allows us to obtain equation (25) under standard assumptions. We discuss the estimator for the subset dynamic causal effects $\mathcal{R}_a^0$ as it include the full causal effects estimator as a special case when $\mathcal{R}_a^0 = \mathcal{R}^0$. The subset causal effects are needed for the subset OPP statistic discussed in Section 4 which is adopted in the empirical Section 8.

In general, we impose three types of primitive assumptions: (i) a constant policy regime assumption, (ii) an identification assumption (e.g. existence of valid instruments in our case) and (iii) a set of standard regularity conditions.

First, we require that the economy was in a constant regime over the sampling period $s = t_0, \dots, t$ and we let n denote the number of time periods.

Assumption S1. *For periods $s = t_0, \dots, t$ the reaction function was given by $\phi^0 \in \Phi$ and the economy can be represented by*

$$\mathcal{Y}_s^0 = A(L; \phi^0) e_s \ , \qquad A(L; \phi^0) = \sum_{j=0}^{\infty} A_j(\phi^0) L^j , \tag{S1}$$

where $\mathcal{Y}_s^0 = (y_s^{0\prime}, x_s^{0\prime}, p_s^{0\prime})'$ is $N \times 1$, with $N = M_y + M_x + K$, $e_s = (\xi_s', \epsilon_s^{0'})'$ is the $N_e \times 1$ vector of uncorrelated structural shocks with mean zero and unit variance, and $A(L; \phi^0)$ has absolutely summable coefficients and $A_0^{p\epsilon}(\phi^0)$ has full rank.[1]

The assumption imposes that for periods $s = t_0, \dots, t$ the policy regime was stable, i.e. ϕ^0 was fixed, and the endogenous variables in the economy can be written as a linear combination of current and lagged structural shocks. Such assumptions are commonly imposed in the macro-econometric literature on the estimation of impulse responses (e.g. Plagborg-Møller and Wolf, 2021). It can be relaxed at the expense of more assumptions, for instance by modeling the time-variation in the causal effects (see e.g. Primiceri, 2005; Paul, 2019).

In practice the following trade-off will arise. To accommodate that the causal effects pertain to a stable regime, i.e. Assumption S1 holds, it is attractive to rely on a short sampling period to estimate the causal effects. The unfortunate consequence is that this will generally increase the variance of the estimates, and ultimately lower the power of the policy optimality test. Therefore a careful assessment of the stability of the policy regime is important.

Similar as in the main text, Assumption S1 allows us to rewrite the model for $Y_s^0 = (y_s^{0'}, \dots, y_{s+H}^{0'})'$ in terms of the policy choices $p_s^0 = (p_{a,s}^{0'}, p_{a^\perp,s}^{0'})'$, where $p_{a,s}^0$ is the subset of policy choices of interest.

$$Y_s^0 = \mathcal{R}_a^0 p_{a,s} + \underbrace{\mathcal{R}_{a^\perp}^0 p_{a^\perp,s} + \Gamma_s(\phi^0) + F_{s+1}(\phi^0)}_{=v_s} \ , \tag{S2}$$

where $\mathcal{R}_a^0$ are the $M_y(H+1) \times K_a$ causal effects of interest, $\Gamma_s(\phi^0)$ is a function of the current and lagged structural shocks excluding ϵ_s^0 and satisfies $\mathbb{E}_s \Gamma_s(\phi^0) = \Gamma_s(\phi^0)$ and $F_{s+1}(\phi^0)$ includes all future shocks, i.e. $\mathbb{E}_s F_{s+1}(\phi^0) = 0$. For convenience we have defined v_s as the error term that summarizes all shocks that the researcher does not have access to.

Second, we postulate that the researcher has access to a sequence of instrumental variables

[1] Recall that $A_0^{p\epsilon}(\phi^0)$ is defined as the $K \times K$ lower right block of $A_0(\phi^0)$.

$\{z_s\}$, where z_s has dimension $L \times 1$, with $L \geq K_a$, and z_s correlates only with the policy choice $p^0_{a,s}$.

Assumption S2. *The instrumental variables z_s satisfy*

1. $\mathbb{E}(z_s v_s') = 0$ *for all* s
2. $\frac{1}{n}\sum_{s=t_0}^{t} \mathbb{E}(z_s p^{0\prime}_{a,s})$ *has uniformly full column rank.*

The first part of the assumption imposes that the instruments are exogenous whereas the second part imposes that they are relevant, i.e. correlated with $p^0_{a,s}$. We restrict our exposition to the case of strong instruments (Assumption S2 part 2.). Such assumption may be too strong for some applications. In such cases confidence regions for $\mathcal{R}^0_a$ should be constructed using weak instrument robust methods, see Andrews, Stock and Sun (2019) for a comprehensive review. Attractive instruments are the policy shocks to $p^0_{a,s}$, i.e. $\epsilon^0_{a,s}$ or proxies for such shocks. In our empirical study we follow this route and rely on high frequency identified monetary surprises as instruments for the policy rate and the slope of the yield curve.

We stack $\mathcal{R}_a$ in the vector $r_a = \text{vec}(\mathcal{R}_a)$, and we use the instruments to define the following moment estimator for the $K_a M_y(H+1) \times 1$ vector $r^0_a \equiv \text{vec}(\mathcal{R}^0_a)$.

$$\hat{r}_a = (Q_a'\hat{D}Q_a)^{-1}Q_a'\hat{D}Z'Y^0 \qquad \text{and} \qquad \hat{r}_a = \text{vec}(\widehat{\mathcal{R}}_a) \ , \tag{S3}$$

where $Q_a = Z'P^0_a$, $P^0_a = (P^{0\prime}_{a,t_0}, \ldots, P^{0\prime}_{a,t})'$, with $P^0_{a,s} = p^{0\prime}_{a,s} \otimes I_{M(H+1)}$, $Z = (Z'_{t_0}, \ldots, Z'_t)'$, with $Z_s = z_s' \otimes I_{M(H+1)}$ and $Y^0 = (Y^{0\prime}_{t_0}, \ldots, Y^{0\prime}_t)'$. For the weighting matrix $\hat{D}$ different choices can be considered including $\hat{D} = (n^{-1}Z'Z)^{-1}$ which leads to the two-stage least squares estimator. We not that the estimator (S3) is analog to the standard LP-IV estimator, with the only difference that we estimate all causal effects at once and not equation-by-equation as is commonly done (e.g. Jordà, 2005). The only reason for joint estimation is that equation (25) requires the distribution of all dynamic causal effects which is easier to obtain in this way.

Third, we require a set of standard regularity conditions – defined in terms of dependence and moment assumptions – that ensure the applicability of a law of large numbers and a central limit theorem.

Assumption S3.

1. $\{(z_s', p^{0\prime}_{a,s}, v_s')\}$ *is an α-mixing sequence with mixing coefficients of size* $-a/(a-2)$, *for* $a > 2$;
2. $\mathbb{E}|z_{i,s}v_{j,s}|^a < \Delta < \infty$ *and* $\mathbb{E}|z_{i,s}p^0_{j,a,s}|^{(a/2)+\rho} < \Delta < \infty$ *for all* i, j, s *and some* $\rho > 0$;

3. $V = \mathrm{Var}(n^{-1/2}\sum_{s=t_0}^{t} Z_s v_s)$ *is uniformly positive definite and there exists* $\hat{V}$*, symmetric and positive definite, such that* $\hat{V} - V \xrightarrow{p} 0$*;*

4. $\hat{D} - D \xrightarrow{p} 0$ *where* $D = O(1)$ *and is symmetric and uniformly positive definite.*

The stated regularity conditions are standard and allow for heterogeneity and dependence, see White (2000), Theorem 5.23, for more discussion.

With assumptions S1-S3 in place we formalize the limiting distribution for $\hat{r}_a$ in the following proposition.

Proposition A1. *Given assumptions S1-S3 we have that*

$$\Omega_a^{-1/2}\sqrt{n}(\hat{r}_a - r_a^0) \xrightarrow{d} N(0, I) \qquad \text{and} \qquad \widehat{\Omega}_a - \Omega_a \xrightarrow{p} 0 \ ,$$

where

$$\widehat{\Omega}_a \equiv (Q_a'\hat{D}Q_a/n)^{-1}Q_a'\hat{D}\hat{V}\hat{D}Q_a(Q_a'\hat{D}Q_a/n)^{-1} \ .$$

This asymptotic approximation $\sqrt{n}(\hat{r}_a - r_a^0) \overset{a}{\sim} N(0, \widehat{\Omega}_a)$ is used below to construct confidence bounds around the subset OPP statistic which avoids that we reject optimality because of estimation error in the causal effects. Due to possible serial correlation in the error term we suggest to use a heteroskedasticity and serial correlation robust estimator for $\hat{V}$, e.g., Lazarus et al. (2018).

Finally, we note that estimator (S3) is only one of many estimators that can be considered for estimating dynamic causal effects. Alternative options are discussed in Ramey (2016) and Stock and Watson (2016), but invariably they will require assumptions that are similar to assumptions S1-S3 to obtain an asymptotic distribution as in Proposition A1.

S1.2 Model misspecification uncertainty

As discussed in the main text, to compute confidence bands for the OPP statistic we need an estimate for the distribution of $\mathbb{E}_t Y_t^0 - \widehat{Y}_t$.

To do so, at least two possibilities exist. First, a researcher can approximate the distribution of $\mathbb{E}_t Y_t^0 - \widehat{Y}_t$ by the distribution of the historical forecast errors $\{Y_s - \widehat{Y}_s\}_{s=t^0}^{t}$. Using this sequence one can estimate the historical bias and variance and use these to *upper-bound* the distribution of model misspecification error using a normality assumption. Alternatively, one could rely on the policy makers' self assessment of model uncertainty.

Here we provide some additional detail for the first route where historical forecast errors are used to assess model misspecification uncertainty. Recall that historical misspecification errors $\{\mathbb{E}_s Y_s^0 - \widehat{Y}_s\}_{s=t_0}^{t}$ are not observable, and we cannot exploit such sequence to predict the distribution of $\mathbb{E}_t Y_t^0 - \widehat{Y}_t$. To see that, note that we have

$$\underbrace{Y_t - \widehat{Y}_t}_{\text{forecast error}} = \underbrace{Y_t - \mathbb{E}_t Y_t^0}_{\text{future error}} + \underbrace{\mathbb{E}_t Y_t^0 - \widehat{Y}_t}_{\text{misspecification error}} .$$

Thus, forecast errors mix two sources of uncertainty: (i) misspecification, i.e., model uncertainty, *and* (ii) future uncertainty. Unfortunately, the two sources of forecast error — misspecification and future uncertainty— are indistinguishable, since only forecast errors are observable. As a result, using the variance of forecast errors will upper-bound the variance of mis-specification uncertainty.

Using this conservative approach, we can simply compute the mean and variance of the forecast errors and use these sample moments in combination with a normality assumption to approximate the distribution of misspecification errors.

A more refined approach (outside the scope of this paper) would be to explicitly model the misspecification errors by exploiting the observation that the future errors should be orthogonal to the time t information set. For instance, under the assumption that the information set can be described by a small number of principal components based on a large panel of macro time series (e.g. Stock and Watson, 2016), we may consider

$$Y_t - \widehat{Y}_t = B f_t + \eta_t \ ,$$

where f_t denote the macro factors. The model misspecification uncertainty is then captured by the distribution of $B f_t$, which can be approximated using standard methods.

S1.3 Confidence interval for the OPP

To construct a confidence interval for the subset OPP, we use the distributions for $\hat{r}_a = \text{vec}(\widehat{\mathcal{R}}_a)$ and $\mathbb{E}_t Y_t^0$ to approximate the distribution of $\delta_{a,t}^* = -(\mathcal{R}_a^{0\prime} \mathcal{W} \mathcal{R}_a^0)^{-1} \mathcal{R}_a^{0\prime} \mathcal{W} \mathbb{E}_t Y_t^0$ for any given preference matrix $\mathcal{W}$. The algorithm in the box below summarizes the details.

In practice, we report the mean OPP estimate $\widehat{\delta}_{a,t}$ and the level α confidence interval $\left[\delta_{a,t}^{(\alpha S_d)}, \delta_{a,t}^{((1-\alpha) S_d)}\right]$. The mean OPP estimate for $S_d \to \infty$ can also be analytically computed using

$$\widehat{\delta}_{a,t} \quad = \quad -(\widehat{\mathcal{R}}_a' \mathcal{W} \widehat{\mathcal{R}}_a + \widehat{\Gamma}_a)^{-1} \widehat{\mathcal{R}}_a' \mathcal{W} \widehat{Y}_{t|t} \ , \tag{S4}$$

where $\widehat{\Gamma}_a = n^{-1} \sum_{i=1}^{M_y(H+1)} \sum_{j=1}^{M_y(H+1)} \mathcal{W}_{ii} \mathcal{W}_{jj} \widehat{\Omega}_{a,(i,j)}$, with $\mathcal{W}_{ii}$ the ith diagonal element of $\mathcal{W}$, and $\widehat{\Omega}_{a,(i,j)}$ denotes the (i,j) block of $\widehat{\Omega}_a$ that is of dimension $K_a \times K_a$.

Implementation of the policy optimality test

1 Obtain the approximations $r_a^0 \overset{a}{\sim} N(\hat{r}_a, n^{-1}\widehat{\Omega}_a)$ and $\mathbb{E}_t Y_t^0 - \widehat{Y}_t \overset{a}{\sim} F_{Y_t^0}$.

2 Compute for a given matrix $\mathcal{W}$ by simulation

$$\delta_{a,t}^j = -(\mathcal{R}_a^{j\prime}\mathcal{W}\mathcal{R}_a^j)^{-1}\mathcal{R}_a^{j\prime}\mathcal{W}\widehat{Y}_t^j \ , \qquad \text{for} \quad j = 1, \ldots, S_d \ ,$$

using independent draws from

$$\text{vec}(\mathcal{R}_a^j) = r^j \sim N(\hat{r}_a, \widehat{\Omega}_a) \ , \qquad \widehat{Y}_t^j = \widehat{Y}_t + U^j \ , \qquad U^j \sim F_{Y_t^0} \ .$$

3 Report for some level of confidence $\alpha \in (0, 1)$

$$\widehat{\delta}_{a,t} = \frac{1}{S_d}\sum_{j=1}^{S_d} \delta_{a,t}^j \qquad \text{and} \qquad \left[\delta_{a,t}^{(\alpha S_d)}, \delta_{a,t}^{((1-\alpha)S_d)}\right] \ ,$$

where $\delta_{a,t}^{(k)}$ denotes the (element wise) kth largest draw of $\{\delta_{a,t}^j, j = 1, \ldots, S_d\}$.

4 If $0 \notin \left[\delta_{a,t}^{(\alpha S_d)}, \delta_{a,t}^{((1-\alpha)S_d)}\right]$ reject that $p_{a,t}^0$ is optimal.

S1.4 Inference for preference parameters

In section 7 we outlined an approach for estimating robust preference parameters $\mathcal{W} = \text{diag}(\beta \otimes \lambda)$. Here we provide the necessary details for implementation and a consistency result. To briefly recap, we model the weights $\omega = \beta \otimes \lambda$ as a function of the $d_\theta \times 1$ parameter vector θ, e.g. $\omega = \omega(\theta)$, and θ is estimated by numerically solving

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \widehat{S}(\theta) \ , \qquad \widehat{S}(\theta) = \left\| \hat{D}_\theta^{1/2} \frac{1}{n} \sum_{s=t_0}^{t} \widehat{\delta}_{a,s}(\theta) \right\|^2 \ ,$$

where in contrast to the main text we allow for weighting by the $K_a \times K_a$ positive definite matrix $\hat{D}_\theta$. Further, Θ is the parameter space of θ and the mean OPP estimate is given by

$$\widehat{\delta}_{a,s}(\theta) = -(\widehat{\mathcal{R}}_a'\mathcal{W}(\theta)\widehat{\mathcal{R}}_a + \widehat{\Gamma}_a)^{-1}\widehat{\mathcal{R}}_a'\mathcal{W}(\theta)\widehat{Y}_s \ , \qquad \mathcal{W}(\theta) = \text{diag}(\omega(\theta))$$

where $\widehat{\Gamma}_a = n^{-1}\sum_{i=1}^{M(H+1)}\sum_{j=1}^{M(H+1)} \mathcal{W}_{ii}(\theta)\mathcal{W}_{jj}(\theta)\widehat{\Omega}_{a,(i,j)}$, with $\mathcal{W}_{ii}(\theta)$ the ith diagonal element of $\mathcal{W}(\theta)$, and $\widehat{\Omega}_{a,(i,j)}$ denotes the (i, j) block of $\widehat{\Omega}_a$ (defined in Proposition A1), which is of dimension $K_a \times K_a$.

The population counterpart of $\widehat{S}(\theta)$ is defined as

$$S(\theta) = \lim_{n\to\infty} \left\| D_\theta^{1/2} \frac{1}{n} \sum_{s=t_0}^{t} \mathbb{E}(\tilde{\delta}_{a,s}(\theta)) \right\|^2 \qquad \tilde{\delta}_{a,s}(\theta) = (\mathcal{R}_a^{0\prime} \mathcal{W}(\theta) \mathcal{R}_a^0)^{-1} \mathcal{R}_a^{0\prime} \mathcal{W}(\theta) \widehat{Y}_s \ , \qquad \text{(S5)}$$

where D_θ is positive definite and defined below as the probability limit of $\hat{D}_\theta$. Note that $\mathbb{E}(\tilde{\delta}_{a,s}(\theta))$ is not necessarily equal to $\mathbb{E}(\delta^*_{a,s}(\theta))$ as the forecasts of the policy maker might be biased.

We define the pseudo-true preference parameters θ_0 as the minimizer of $S(\theta)$, e.g.

$$\theta_0 = \arg\min_{\theta\in\Theta} S(\theta) \ .$$

The pseudo-true preference parameters θ_0 can be interpreted as follows. Under the assumption that the policy maker constructed unbiased forecasts and was acting optimally on average we have that $\mathbb{E}(\tilde{\delta}_{a,s}(\theta_0)) = \mathbb{E}\left(\delta^*_{a,s}(\theta_0)\right) = 0$ and θ_0 are the corresponding true preference parameters which are identified by minimizing $S(\theta)$. If the policy maker was not acting optimally on average we have that $\mathbb{E}\left(\tilde{\delta}_{a,s}(\theta_0)\right) \neq 0$, and the parameters θ_0 correspond to the parameter values that bring the policy maker as close as possible to optimality as θ_0 minimizes the objective function. In other words, θ_0 denotes the worst case preference parameters for a researcher that aims to reject optimality.

Next, we impose a set of mild regularity conditions based on which we will be able to justify that $\hat{\theta}$ converges to θ_0 in probability as $n \to \infty$.

Assumption S4. *We assume that*

1. *The parameter space* Θ *is compact subset of* $\mathbb{R}^{d_\theta}$
2. *The weights* $\omega(\theta) = (\omega_1(\theta), \ldots, \omega_{M(H+1)}(\theta))'$ *are a continuous function of* θ *and satisfy* $\min_i \omega_i(\theta) \geq c_{\min} > 0$ *and* $\max_i \omega_i(\theta) \leq c_{\max} < \infty$ *for all* $\theta \in \Theta$
3. θ_0 *is the unique minimizer of* $S(\theta)$ *defined in* (S5)
4. *The sequence of forecasts* $\{\widehat{Y}_s\}$ *are* α*-mixing of size* $r/(r-1)$, $r > 1$, *such that* $\mathbb{E}|\widehat{Y}_{i,s}|^{r+\delta} < \Delta < \infty$ *for some* $\delta > 0$ *and all* i, s
5. $\hat{D}_\theta - D_\theta \xrightarrow{p} 0$ *where* $D_\theta = O(1)$ *and is symmetric and uniformly positive definite.*

Part 1 imposes compactness of the parameter space, which simplifies the proof and can be relaxed similarly as in Newey and McFadden (1994), section 2.6. Part 2 ensures that the weights are non-degenerate uniformly over the parameter space. Part 3 imposes that θ_0

is the unique minimizer of the population objective function. This condition can be easily verified for the chosen parameterization of $\omega(\theta)$. Part 4 restricts the dependence in the forecasts and imposes a mild moment assumption. Recall that the forecasts are constructed for $Y_s = [y_{m,s+h} - y^*_{m,s+h}]_{m=1,\ldots,M,h=0,\ldots,H}$ and an inspection of $\widehat{Y}_s$ for our empirical application suggests that the dependence in this series is mild. Part 5 imposes the conditions on the weighting matrix $\hat{D}_\theta$ and its population counterpart D_θ.

These assumptions imply the following consistency result.

Proposition A2. *Given assumptions S1, S2, S3 and S4, we have that*

$$\hat{\theta} \xrightarrow{p} \theta_0 \ ,$$

where θ_0 is defined in (S5)*.*

The proposition implies that that for large samples $\hat{\theta}$ is close to θ_0, the preference parameter vector that is least favorable to rejecting optimality.

S1.5 Inference for systematic optimization failures

In this section we discuss the details for the implementation of the test for the reaction function.

Recall, from Proposition 2 that under an optimal reaction function we should have that $\mathbb{E}(\delta^*_{a,s}\Xi'_s) = 0$, where $\Xi_s = (\xi'_s, \xi'_{s-1}, \epsilon'_{s-1}, \ldots)'$ is the set of past and present structural shocks (except the contemporaneous policy shock ϵ^0_t). We consider testing a subset of these moment conditions as in practice researchers will typically not have access to proxies for all structural shocks. Let $\Xi_{c,s}$ denote the $L_\xi \times 1$ vector of structural shocks of interest and let $\Xi_{c^\perp,s}$ include all other shocks.

To implement the test we note that given Assumption S1 the OPP statistic $\delta^*_{a,s}$ can be written as a linear combination of the structural shocks, see also equation (15). With this observation in mind we may decompose our estimated OPP as follows

$$\begin{aligned} \widehat{\delta}_{a,s} &= \delta^*_{a,s} + \tilde{\eta}_s \\ &= B_{a,\Xi_c}\Xi_{c,s} + \underbrace{B_{a,\Xi_{c^\perp}}\Xi_{c^\perp,s} + B_{a,\epsilon}\epsilon^0_s + \tilde{\eta}_s}_{=\breve{\eta}_s} \ , \end{aligned} \tag{S6}$$

where the term $\tilde{\eta}_s = \widehat{\delta}_{a,s} - \delta^*_{a,s}$ captures the measurement error in the OPP estimate. The term $\breve{\eta}_s$ gathers all the components that have no direct interest for our purpose.

Next, we impose the existence of at least one optimal reaction function.

Assumption S5. *There exists a non-empty set* $\Phi^{\text{opt}} \subset \Phi$ *such that for any* $\phi^{\text{opt}} \equiv \phi \in \Phi^{\text{opt}}$ *the policy choice* $p_s = \phi^{\text{opt}} w_s$ *minimizes* $\mathcal{L}_s$, *for* $s = t_0, \ldots, t$, *i.e.*

$$\mathcal{L}_s\left(\phi^{\text{opt}}, 0\right) \leq \mathcal{L}_s\left(\tilde{\phi}, \epsilon_s\right), \qquad \forall\ \tilde{\phi} \notin \Phi^{\text{opt}},\ \epsilon_s \neq 0\ ,$$

where $\mathcal{L}_s\left(\phi, \epsilon_s\right) = \mathbb{E}_s Y_s^{'} \mathcal{W} Y_s$, *with* $p_s = \phi w_s + \epsilon_s$.

The assumption is identical to part 2 of Assumption 1 in the main text. We only restate it here for convenience. The assumption implies that if $\phi^0 \in \Phi^{\text{opt}}$, i.e. the chosen reaction function is optimal, we have that $B_{a,\Xi_c}\Xi_{c,s}$ and $B_{a,\Xi_{c^\perp}}\Xi_{c^\perp,s}$ in (S6) are equal to zero as the optimal reaction function cancels the effects of the non-policy structural shocks on the gradient of the loss function.

If we observed $\Xi_{c,s}$ we could simply run an OLS regression to test whether Ξ_c has an effect on the subset OPP.[2] However, as noted in Stock and Watson (2018) structural shocks are rarely completely observable, and while proxies for the structural shocks are often available, they typically contain measurement error. Therefore to test the reaction function, we adopt an instrumental variable approach in the spirit of Stock and Watson (2018).

Specifically, suppose that the structural shocks of interest $\Xi_{c,s}$ have a non-zero effect on the variables w_s, which are taken as any subset of $\{y_s^0, x_s^0\}_{s \leq t}$ as defined in Assumption S1. Further, suppose that there exist proxies or instruments z_s^ϕ that are correlated with $\Xi_{c,s}$, but not with the other structural shocks $\Xi_{c^\perp,s}$ nor the measurement error $\tilde{\eta}_s$. Formally we make the following assumption.

Assumption S6. *We have that*

1. *there exits a* $L_\xi \times 1$ *vector of variables* w_s *(subset of* $\{y_s^0, x_s^0\}_{s \leq t}$*) such that* $\mathbb{E}(w_s \Xi_{c,s}')$ *has full column rank.*
2. *there exists a* $L_\phi \times 1$ *vector* z_s^ϕ, *with* $L_\phi \geq L_\xi$ *such that* $\mathbb{E}(z_s^\phi \check{\eta}_s') = 0$ *for all* s *and* $\frac{1}{n}\sum_{s=t_0}^{t} \mathbb{E}(z_s^\delta w_s')$ *has uniformly full column rank.*

To give an example, in our empirical application in the main text we postulated that w_s (in our case, the inflation rate) depends on $\Xi_{c,s}$: the current and past oil and productivity shocks. The proxies z_s^ϕ that we use are the oil supply shock estimates of Baumeister and Hamilton (2019) and the productivity shock estimates of Basu, Fernald and Kimball (2006).

Without loss of generality we normalize $\Xi_{c,s}$ such that it has a unit effect on w_s (see e.g. Stock and Watson, 2018, page 923), and we consider the following regression

$$\widehat{\delta}_{a,s} = B_{a,\Xi_c} w_s + \eta_s\ , \tag{S7}$$

[2] This follows as structural shocks are serially and mutually uncorrelated. Also assuming that the measurement error $\tilde{\eta}_s$ is uncorrelated with the structural shocks $\Xi_{c,s}$.

where $\eta_s = \check{\eta}_s + B_{a,\Xi_c}(\Xi_{c,s} - w_s)$.

To address the endogeneity problem in (S7) —w_s is correlated with η_s—, we consider the following moment estimator for $b_{a,\Xi_c} = \text{vec}(B_{a,\Xi_c})$

$$\hat{b}_{a,\Xi_c} = (Q^{\phi\prime}\hat{D}^{\phi}Q^{\phi})^{-1}Q^{\phi\prime}\hat{D}^{\phi}d_a \ , \tag{S8}$$

where $Q^{\phi} = Z^{\phi\prime}W$, $W = (W_{t_0}', \ldots, W_t')'$, with $W_s = w_s' \otimes I_{K_\xi}$, $Z^{\phi} = (Z_{t_0}^{\phi\prime}, \ldots, Z_t^{\phi\prime})'$, with $Z_s^{\phi} = z_s^{\phi\prime} \otimes I_{L_\xi}$ and $d_a = (\widehat{\delta}_{a,t_0}', \ldots, \widehat{\delta}_{a,t}')'$. For the weighting matrix $\hat{D}^{\phi}$, we set $\hat{D}^{\phi} = (n^{-1}Z^{\phi\prime}Z^{\phi})^{-1}$, which leads to the two-stage least squares estimator.

With the instrumental variable estimate for B_{a,Ξ_c} we can construct a standard Wald test of joint significance for the B_{a,Ξ_c} elements. If we can reject that the B_{a,Ξ_c} elements are jointly zero, we can reject the null hypothesis that the policy maker's reaction function ϕ^0 is optimal.

To implement this we require standard regularity conditions similar to those stated in Assumption S3.

Assumption S7. *We assume that*

1. $\{(z_s^{\phi\prime}, w_s', \eta_s')\}$ *is an α-mixing sequence with mixing coefficients of size* $-a_\phi/(a_\phi - 2)$, *for* $a_\phi > 2$;
2. $\mathbb{E}|z_{i,s}^{\phi}\eta_{j,s}|^{r_\phi} < \Delta_\phi < \infty$ *and* $\mathbb{E}|z_{i,s}^{\phi}w_{j,s}|^{r_\phi/2+\rho_\phi} < \Delta_\phi < \infty$ *for all* i, j, s *and some* $\rho_\phi > 0$;
3. $V^{\phi} = \text{Var}(n^{-1/2}\sum_{s=t_0}^{t} Z_s^{\phi}\eta_s)$ *is uniformly positive definite and there exists* $\hat{V}^{\phi}$, *symmetric and positive definite, such that* $\hat{V}^{\phi} - V^{\phi} \xrightarrow{p} 0$;
4. $\hat{D}^{\phi} - D^{\phi} \xrightarrow{p} 0$ *where* $D^{\phi} = O(1)$ *and is symmetric and uniformly positive definite.*

The following proposition formalizes the detection of systematic optimization failures using the OPP.

Proposition A3. *Given Assumptions S1, S5, S6 and S7 we have that*

$$\textit{if} \quad n\hat{b}_{a,\Xi_c}'\widehat{\text{Var}}(\hat{b}_{a,\Xi_c})^{-1}\hat{b}_{a,\Xi_c} > \chi^2_{K_aL_\xi,1-\alpha} \qquad \textit{we reject } H_0 : \phi^0 \in \Phi^{\text{opt}}$$

with confidence level α. *Here* $\hat{b}_{a,\Xi_c}$ *is defined in* (S8), $\chi^2_{K_aL_\xi,1-\alpha}$ *is the* $1-\alpha$ *critical value of the* χ^2*-distribution with* K_aL_ξ *degrees of freedom and*

$$\widehat{\text{Var}}(\hat{b}_{a,\Xi_c}) \equiv (Q^{\phi\prime}\hat{D}^{\phi}Q^{\phi}/n)^{-1}Q^{\phi\prime}\hat{D}^{\phi}\hat{V}^{\phi}\hat{D}^{\phi}Q^{\phi}(Q^{\phi\prime}\hat{D}^{\phi}Q^{\phi}/n)^{-1} \ ,$$

with $\hat{V}^{\phi}$ *any consistent estimate for the asymptotic variance* $V^{\phi} = \text{Var}(n^{-1/2}\sum_{s=t_0}^{t} Z_s^{\phi}\eta_s)$, *i.e.* $\hat{V}^{\phi} - V^{\phi} \xrightarrow{p} 0$.

The proposition formalizes the use of the Wald statistic for testing the reaction function. Specifically, if the Wald statistic $n\hat{b}'_{a,\Xi_c}\widehat{\text{Var}}(\hat{b}_{a,\Xi_c})^{-1}\hat{b}_{a,\Xi_c}$ exceeds the critical value we can reject the null hypothesis that the reaction function ϕ^0 is optimal.

S2 Considering arbitrary convex loss functions

In the main text we restricted ourselves to quadratic loss functions when testing the optimality of a given policy choice. In this section we show that the main idea – exploiting the gradient of the loss function to evaluate optimality– continues to apply for essentially any convex loss function. The only difference is that the evaluation of the gradient will require the full forecast densities instead of only the mean forecasts.

To show this, let $\mathcal{L}_t(Y_t;\theta)$ denote a loss function which is convex with respect to Y_t and may depend on preference parameters denoted by θ. The quadratic loss function (6) in the main text is a special case. The policy maker's problem can be summarized by

$$\mathbb{E}_t\mathcal{L}_t(Y_t;\theta)\ , \qquad Y_t = \mathcal{R}(\phi)p_t + \Gamma_t(\phi) + F_{t+1}(\phi)\ ,$$

where the generic model for the economy is based on Assumption 1, see also equation (10). Again the terms $\mathcal{R}(\phi)$, $\Gamma_t(\phi)$ and $F_{t+1}(\phi)$ can be expressed in terms of the polynomial $A(L;\phi)$ and the structural shocks.

To evaluate whether a given policy choice p_t^0 minimizes $\mathbb{E}_t\mathcal{L}_t(Y_t;\theta)$ we rely on the same gradient statistic as in our simple example from Section 2. The gradient evaluated at p_t^0 is given by

$$\nabla_{p_t}\mathbb{E}_t\mathcal{L}_t(Y_t;\theta)|_{p_t^0} = \mathcal{R}^{0'} \times \nabla_{Y_t}\mathbb{E}_t\mathcal{L}_t(Y_t;\theta)|_{p_t^0}\ .$$

Given that $\mathcal{L}_t(Y_t;\theta)$ is convex with respect to p_t we have that if $\nabla_{p_t}\mathbb{E}_t\mathcal{L}_t(Y_t;\theta)|_{p_t^0} \neq 0$ the policy choice p_t^0 is not optimal.

To evaluate the gradient we need to compute the derivative $\nabla_{Y_t}\mathbb{E}_t\mathcal{L}_t(Y_t;\theta)|_{p_t^0}$. Under a quadratic loss $\mathcal{L}_t = Y_t'\mathcal{W}Y_t$, this expression simplifies to $\nabla_{Y_t}\mathbb{E}_t\mathcal{L}_t(Y_t;\theta)|_{p_t^0} = 2\mathcal{W}\mathbb{E}_tY_t^0$ as in the main text, but for a general convex loss we have

$$\nabla_{Y_t}\mathbb{E}_t\mathcal{L}_t(Y_t;\theta)|_{p_t^0} = \int_{Y_t^0}\nabla_{Y_t}\mathcal{L}_t(Y_t^0;\theta)p(Y_t^0|\mathcal{F}_t)\mathrm{d}Y_t^0\ , \tag{S9}$$

where $p(Y_t^0|\mathcal{F}_t)$ is the forecast density under the proposed policy choice p_t^0. Thus, provided the forecast density is available, we can construct the OPP statistic and OPP-based tests as in the main text. The only difference is that there is no closed form expression for the gradient, and numerical integration methods will be necessary.

S3 Real time policy evaluation

In this section, we describe how one could implement the policy optimality test inversion of Section 8.2 using more general loss functions. As illustration, we will use the loss function of section 8.1:

$$\mathcal{L}_t = \mathbb{E}_t \left\|\Pi_t\right\|^2 + \lambda \mathbb{E}_t \left\|U_t\right\|^2 \ , \tag{S10}$$

with $\Pi_t = (\pi_t - \pi^*, \ldots, \pi_{t+H} - \pi^*)'$ the vector of inflation gaps and $U_t = (u_t - u_t^*, \ldots, u_{t+H} - u_{t+H}^*)'$ the vector of unemployment gaps. The discount rate is implicitly set to $\beta_h = 1$ for all h, and we take a horizon of $H = 5$ years.

We again consider the FOMC as of April 2008. The exercise is thus identical to the one in the main text bar one difference. With an arbitrary loss function, the vectors $\mathbb{E}_t Y_t^0$ and $\mathcal{R}_{\mathsf{i}}$ can be of large dimensions. For instance, using the loss function (S10), the vector of objectives Y_t has dimension $2(H+1)\times 1$, and the dynamic causal effect $\mathcal{R}_{\mathsf{i}}$ has dimension $2(H+1)\times 1$. That's a total of $4(H+1)$ parameters, which represents 84 parameters using quarterly data and $H = 5$ years. For the policy optimality test inversion to be informative for policy makers, the parameter space must thus be shrunk such that the number of parameters to be considered remains manageable, with each parameter corresponding to the most important features of the forecast or causal effect.

To do so, we first summarize the forecast $\mathbb{E}_t Y_t^0$ with a simple functional form, here a Gaussian basis function as in Barnichon and Matthes (2018) though other functional forms are possible. Specifically, we summarize the expected path of some policy objective y_t with

$$y_{t+h} - y^* = a e^{-\frac{(h-b)^2}{c^2}} . \tag{S11}$$

In this example, the Gaussian basis function offers an attractive dimension-reduction tool, reducing the number of parameters per forecast from 20 (using quarterly data and $H = 5$ years) to only 3. In addition, the three coefficients a, b and c have a direct economic interpretation in terms of features of the forecast: parameter a is the peak of the deviation of y from its target y^*, parameter b is the timing of this peak, and parameter c captures the persistence of the deviation from target, as $\tau = c\sqrt{\ln 2}$ is the amount of time required for the target deviation to return to 50% of its maximum value.

Next, to consider a range of forecasts that are representative of the different views at the FOMC in April 2008, we vary the forecasts along two dimensions (i) the severity of the expected rise in unemployment, and (ii) the duration of the expected excessive inflation.[3]

Specifically, having summarized each FOMC member's forecast with coefficients (a_x, b_x, c_x)

[3] Clearly, other approaches are possible. The important point is that the number of parameters in the inverted policy optimality test must be small enough to be manageable and informative to policy makers.

with $x = \pi$ or u, we vary the coefficients a_u and c_π to summarize the range of forecasts considered by the FOMC.

Figure S1 plots the SEP economic projections of the 17 FOMC members as well as the different forecasts considered in our exercise, summarized here with the expected unemployment rate and inflation rate in 2009q4.

Next, as in the main text, we study how the optimality of the policy decision varies with the causal effects of the policy rate. Again, because the impulse responses (IRs) are high dimensional objects, we shrink their dimensions with Gaussian basis functions as in (S11).

To consider a range of inflation-unemployment trade-offs as in the main text, we vary the peak of the impulse response of inflation —the coefficient a_π— holding the impulse response of unemployment constant.[4] To express these different trade-offs in a simple unit, we can report the Phillips multiplier

$$\mathcal{P} = \frac{\sum_{j=0}^{h} \mathcal{R}_j^\pi}{\sum_{j=0}^{h} \mathcal{R}_j^u} \ . \tag{S12}$$

The Phillips multiplier is simply the ratio of the cumulative IR of inflation over the cumulative IR of unemployment, analogously to the government spending multiplier (see King and Watson, 1994; Ramey and Zubairy, 2018; Barnichon and Mesters, 2021).

Heatmaps and optimality regions can then be constructed exactly as in the main text, where we plot the OPP statistic and policy optimality test p-values as we vary λ and $\mathcal{P}$. The results are displayed in Figures S2 and S3. The conclusions are very similar to the main text. Only a low λ and a low Phillips multiplier (bottom-right panel) can justify the policy stance of most FOMC members in April 2008.

S4 Additional results for the empirical study

In this section we discuss additional results for our empirical study on testing US monetary policy decisions. These results are complementary to those presented in Section 8 of the main text. In particular, the different subsections discuss the following aspects.

1. Understanding the policy perturbations
2. Sensitivity to the discount rate β
3. Sensitivity to the preference parameter λ
4. Alternative dynamic causal effect estimates: SVAR inference

[4] It is only the *relative* movement of inflation with respect to unemployment that matters for the optimality of the policy choice.

5. Testing the stability of the macro environment
6. Using alternative forecasts: Greenbook

S4.1 Impulse responses to the policy perturbations

In our empirical study we use perturbations to the policy rate and to the slope of the yield curve. Clearly, this is a subset of all possible policy perturbations, and it is of interest to understand the policy experiments that we are considering.

To this extent, Figure S4 plots the full set of impulse responses to (i) innovations to the fed funds rate, and (ii) innovations to the slope of yield curve. Compared to the figures in the main text, Figure S4 also reports the path of the fed funds rate and the path of the slope of the yield curve. We can see that both policy experiments correspond to somewhat persistent changes to the policy instrument, similar to earlier estimates of impulse responses to monetary shocks (e.g., Barnichon and Matthes, 2018).

S4.2 The discount rate β

In our main specification we fixed the discount rate $\beta_j = 1$, for all $j = 0, \ldots, H$, when we constructed the preference matrix $\mathcal{W} = \text{diag}(\beta \otimes \lambda)$. In this section we investigate the sensitivity of the policy optimality test with respect to this assumption. In particular, we model $\beta_j = \beta^j$ and we vary the constant β by considering $\beta = 0.8, 0.85, \ldots, 1$. The results for the policy optimality test are shown in Figure S5. The panels correspond to the bottom panels of Figure 5 in the main text. The different colors show the mean OPP estimate for different choices of β. Reasonable choices for β have little effect on the OPP statistic. If anything, decreasing β enlarges the OPP statistic, i.e. moves it further away from zero.

S4.3 The preference parameter λ

In the main text we used the least favorable value $\hat{\lambda} = 0.6$ to compute the OPP-based tests. In this section we compute the OPPs for different choices of λ between $[0.2, 1]$. The results for different choices of λ are shown in Figure S6. We find that the OPP for the Fed Funds rate is not sensitive to the choice for λ. In fact, all of our main findings hold for all choices of λ and the differences are often small.

S4.4 Alternative dynamic causal effect estimates

In the main text we used LP-IV type estimates for the dynamic causal effects, see equation (S3). In this section we estimate the causal effects using the SVAR-IV methodology (e.g.

Montiel Olea, Stock and Watson, 2020). The instrumental variables remain identical: high frequency surprises to the fed funds rate and to the slope of the yield curve.

To compute the SVAR-IV dynamic causal effects we use the same specification as for LP-IV and consider a SVAR with unemployment, inflation, the spread between the 10 year and short term interest rates, the short term interest rate and the excess bond yield measure of Gilchrist and Zakrajšek (2012). This implies that we use the same control variables as we used for the LP-IV specification.

The impulse response estimates are shown in Figure S7. We find that the patterns are very close. There are two differences: (i) for the fed funds rate shock the response of inflation is lower for the SVAR-IV method and (ii) for the slope shock the response of unemployment is lower for the SVAR-IV method.

Next, we use the SVAR-IV estimates to compute the OPP statistics. The results are shown in Figure S8. The differences between LP-based and SVAR-based OPPs for the fed funds rate are very small. If anything, the SVAR-based OPP for the fed funds rate is more often significantly different from zero, notably around 1999-2000 when the OPP is significantly positive. For the SVAR-based OPPs for the slope of the yield curve, the SVAR-based OPP is somewhat muted, still significant at the 68% level but not anymore at the 95% level. This is caused by the fact that SVAR-IV estimates a slightly smaller impulse response of unemployment in response to a slope shock.

S4.5 Testing the stability of the dynamic causal effects

The retrospective study of past policy decisions is based on the assumption that the dynamic causal effects are stable for some time period (so that we can estimate $\mathcal{R}_a^0$). To verify whether this was the case for our empirical study we consider testing the stability of the dynamic causal effect estimates using the structural change tests for linear models with endogenous variables proposed in Hall, Han and Boldea (2012). We implement the tests using the wild fixed-regressor bootstrap as this allows for heteroskedasticity and an unstable reduced form, see Boldea, Cornea-Madeira and Hall (2019).

The specifications that we consider are given by

$$y_{t+h} = \mathcal{R}_{x,h}^0 x_t + \gamma^y w_t + u_t \qquad y = \pi, u \qquad x = i, s \;, \tag{S13}$$

where s denotes the slope of the yield curve and i the policy rate. The fed funds rate i_t is instrumented by the monetary policy surprises to the fed funds rate measured around the FOMC announcements within a 30 minute window (e.g. Kuttner, 2001). The slope of the yield curve is instrumented by the surprises to the ten-year Treasury yield (orthogonalized with respect to surprises to the current fed funds rate). Note that in this way the 2SLS estimator for $\mathcal{R}^0_{x,h}$ corresponds to the moment estimator (S3) implemented with $\hat{D} = (n^{-1}Z'Z)^{-1}$ as the weighting matrix. The only difference is that we conduct the break tests equation-by-equation to study whether there exists breaks at different horizons.

We are interested in testing whether $\mathcal{R}^0_{x,h}$ is stable over the sampling period. Specifically, we test the hypothesis $H_0 : m = 0$ against $H_1 : m = 1$, where m is the number of structural breaks in $\mathcal{R}^0_{x,h}$. The test statistic used is the sup-Wald test, which under homoskedasticity corresponds to the sup-F test of Andrews (1993). We follow Boldea, Cornea-Madeira and Hall (2019) and implement the test using the wild fixed-regressor bootstrap. We refer to their paper for the details.

The resulting bootstrap p-values are reported in Table S1. We find no evidence of parameter instability as we can never reject the null of constant $\mathcal{R}^0_{x,h}$ whether for the causal effect of a perturbation to the fed funds rate or to the slope of the yield curve. This holds for both inflation and unemployment and for all the horizons considered.

S4.6 Alternative forecasts: Greenbook

In the main text we relied on the SEP forecasts. In this section we consider the Greenbook forecasts as an alternative. One benefit of the Greenbook forecasts is that they are published more frequently, which yields more testable periods for the OPP. However, the Greenbook is published with a delay so the period 2016-2020 is unavailable.

The results are shown in Figure S9. We find that the results are very similar when compared to the SEP forecasts. If anything, the confidence bands for the OPPs based on Greenbook forecasts are slightly smaller.

References

Andrews, Donald W. K. 1993. "Tests for Parameter Instability and Structural Change With Unknown Change Point." *Econometrica*, 61(4): 821–856.

Andrews, Isaiah, James Stock, and Liyang Sun. 2019. "Weak Instruments in IV Regression: Theory and Practice." *Annual Review of Economics*. forthcoming.

Barnichon, Regis, and Christian Matthes. 2018. "Functional Approximation of Impulse Responses." *Journal of Monetary Economics*.

Barnichon, Regis, and Geert Mesters. 2021. "The Phillips Multiplier." *Journal of Monetary Economics*, 117: 689–705.

Basu, Susanto, John G. Fernald, and Miles S. Kimball. 2006. "Are Technology Improvements Contractionary?" *American Economic Review*, 96(5): 1418–1448.

Baumeister, Christiane, and James D. Hamilton. 2019. "Structural Interpretation of Vector Autoregressions with Incomplete Identification: Revisiting the Role of Oil Supply and Demand Shocks." *American Economic Review*, 109(5): 1873–1910.

Boldea, Otilia, Adriana Cornea-Madeira, and Alastair R. Hall. 2019. "Bootstrapping structural change tests." *Journal of Econometrics*, 213(2): 359–397.

Gilchrist, Simon, and Egon Zakrajšek. 2012. "Credit Spreads and Business Cycle Fluctuations." *American Economic Review*, 102(4): 1692–1720.

Hall, Alastair R., Sanggohn Han, and Otilia Boldea. 2012. "Inference regarding multiple structural changes in linear models with endogenous regressors." *Journal of Econometrics*, 170(2): 281–302.

Jordà, Oscar. 2005. "Estimation and Inference of Impulse Responses by Local Projections." *The American Economic Review*, 95: 161–182.

King, Robert G., and Mark W. Watson. 1994. "The post-war U.S. phillips curve: a revisionist econometric history." *Carnegie-Rochester Conference Series on Public Policy*, 41: 157–219.

Kuttner, Kenneth N. 2001. "Monetary policy surprises and interest rates: Evidence from the Fed funds futures market." *Journal of Monetary Economics*, 47(3): 523–544.

Lazarus, Eben, Daniel J. Lewis, James H. Stock, and Mark W. Watson. 2018. "HAR Inference: Recommendations for Practice." *Journal of Business & Economic Statistics*, 36(4): 541–559.

Leeper, Eric M., and Tao Zha. 2003. “Modest policy interventions.” *Journal of Monetary Economics*, 50(8): 1673–1700.

Montiel Olea, José L., James H. Stock, and Mark W. Watson. 2020. “Inference in Structural Vector Autoregressions identified with an external instrument.” *Journal of Econometrics*. forthcoming.

Newey, Whitney K., and Daniel McFadden. 1994. “Chapter 36 Large sample estimation and hypothesis testing.” In . Vol. 4 of *Handbook of Econometrics*, 2111 – 2245. Elsevier.

Paul, Pascal. 2019. “The Time-Varying Effect of Monetary Policy on Asset Prices.” *The Review of Economics and Statistics*. Forthcoming.

Plagborg-Møller, Mikkel, and Christian K. Wolf. 2021. “Local Projections and VARs Estimate the Same Impulse Responses.” *Econometrica*, 89(2): 955–980.

Primiceri, Giorgio. E. 2005. “Time Varying Structural Vector Autoregressions and Monetary Policy.” *The Review of Economic Studies*, 72: 821–852.

Ramey, Valerie. 2016. “Macroeconomic Shocks and Their Propagation.” In *Handbook of Macroeconomics*. , ed. J. B. Taylor and H. Uhlig. Amsterdam, North Holland:Elsevier.

Ramey, Valerie A., and Sarah Zubairy. 2018. “Government Spending Multipliers in Good Times and in Bad: Evidence from U.S. Historical Data.” *Journal of Political Economy*, 126.

Stock, James H., and Mark W. Watson. 2016. “Chapter 8 - Dynamic Factor Models, Factor-Augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics.” In . Vol. 2 of *Handbook of Macroeconomics*, , ed. John B. Taylor and Harald Uhlig, 415 – 525. Elsevier.

Stock, James H., and Mark W. Watson. 2018. “Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments.” *The Economic Journal*, 128(610): 917–948.

White, Halbert L. Jr. 2000. *Asymptotic Theory for Econometrians - 2nd edition.* San Diego, California:Academic Press.

Proofs

Proof of Proposition A1. The assumptions S1 (implying equation (S2)), S2 and S3 correspond exactly to assumptions (i)-(v) in Theorem 5.23 of White (2000). Hence, the proof of White (2000) Theorem 5.23 applies. □

Proof of Proposition A2. We verify the conditions of Theorem 2.1 in Newey and McFadden (1994). Parts (i), (ii) and (iii) follow directly from assumption S4 part 1, 2 and 3, noting that the continuity of $S(\theta)$ wrt θ follows as $S(\theta)$ is a continuous function of $\delta^*_{a,s}$, which in its turn is a continuous function of $\omega(\theta)$. It remains to verify part (iv) which requires

$$\sup_{\theta\in\Theta}\left|\widehat{S}(\theta)-S(\theta)\right| \xrightarrow{p} 0$$

Note first that, by assumption S4 part 5, we have that there exists a finite constant $C_\theta > 0$ such that $\|D_\theta\| \leq C_\theta$. Also, we have that $\sup_{\theta\in\Theta} S(\theta) < \infty$ as

$$\begin{aligned}
S(\theta) &= \left\| D^{1/2}(\mathcal{R}^{0\prime}_a \mathcal{W}(\theta)\mathcal{R}^0_a)^{-1}\mathcal{R}^{0\prime}_a \mathcal{W}(\theta) \lim_{n\to\infty}\frac{1}{n}\sum_{s=t_0}^{t}\mathbb{E}\widehat{Y}_u \right\|^2 \\
&\leq C_\theta \Delta^2 \left\| (\mathcal{R}^{0\prime}_a \mathcal{W}(\theta)\mathcal{R}^0_a)^{-1}\mathcal{R}^{0\prime}_a \mathcal{W}(\theta)\iota \right\|^2 \\
&\leq C_\theta \Delta^2 c_{\min}^{-2} c_{\max}^2 M^2 (H+1)^2 \|(\mathcal{R}^{0\prime}_a \mathcal{R}^0_a)^{-1}\|^2 \|\mathcal{R}^0_a\|^2 < \infty
\end{aligned}$$

as $\mathbb{E}|\widehat{Y}_s| < \Delta < \infty$ by Assumption S4 part 4, the weights are uniformly bounded by Assumption S4 part 2, and $\mathcal{R}^0_a$ has full column rank. Next,

$$\widehat{S}(\theta) - S(\theta) = (\widehat{m}_n(\theta) - m(\theta))'(\widehat{m}_n(\theta) - m(\theta)) + 2(\widehat{m}_n(\theta) - m(\theta))'m(\theta)$$

where $\widehat{m}_n(\theta) = \hat{D}_\theta^{1/2}\frac{1}{n}\sum_{s=t_0}^{t}\widehat{\delta}_{a,s}(\theta)$ and $m(\theta) = D_\theta^{1/2}\lim_{n\to\infty}\frac{1}{n}\sum_{s=t_0}^{t}\mathbb{E}\tilde{\delta}_{a,s}(\theta)$. Since, $\|m(\theta)\|^2 =$

$S(\theta) < \infty$ it remains to study

$$
\begin{aligned}
\widehat{m}_n(\theta) - m(\theta) &= \hat{D}_\theta^{1/2} \frac{1}{n} \sum_{s=t_0}^{t} \widehat{\delta}_{a,s}(\theta) - D_\theta^{1/2} \frac{1}{n} \sum_{s=t_0}^{t} \mathbb{E}\tilde{\delta}_{a,s} \\
&= \hat{D}_\theta^{1/2} \frac{1}{n} \sum_{s=t_0}^{t} (\widehat{\delta}_{a,s}(\theta) - \mathbb{E}\tilde{\delta}_{a,s}) + (\hat{D}_\theta^{1/2} - D_\theta^{1/2}) \frac{1}{n} \sum_{s=t_0}^{t} \mathbb{E}\tilde{\delta}_{a,s} \\
&= \hat{D}_\theta^{1/2} \frac{1}{n} \sum_{s=t_0}^{t} \widehat{\delta}_{a,s}(\theta) - \mathbb{E}\tilde{\delta}_{a,s} + o_p(1) \\
&= -\hat{D}_\theta^{1/2} (\widehat{\mathcal{R}}_a' \mathcal{W}(\theta) \widehat{\mathcal{R}}_a + \widehat{\beta}_a)^{-1} \widehat{\mathcal{R}}_a' \mathcal{W}(\theta) \frac{1}{n} \sum_{s=t_0}^{t} (\widehat{Y}_s - \mathbb{E}\widehat{Y}_s) \\
&- [\hat{D}_\theta^{1/2} (\widehat{\mathcal{R}}_a' \mathcal{W}(\theta) \widehat{\mathcal{R}}_a + \widehat{\beta}_a)^{-1} \widehat{\mathcal{R}}_a' \mathcal{W}(\theta) - D_\theta^{1/2} (\mathcal{R}_a^{0\prime} \mathcal{W}(\theta) \mathcal{R}_a^0)^{-1} \mathcal{R}_a^{0\prime} \mathcal{W}(\theta)] \frac{1}{n} \sum_{s=t_0}^{t} \mathbb{E}\widehat{Y}_s
\end{aligned}
$$

We note that by Proposition A1 we have $\widehat{\mathcal{R}}_a \xrightarrow{p} \mathcal{R}_a^0$ and $\widehat{\beta}_a = o_p(1)$. Assumption S4 part 5 states $\hat{D}_\theta - D_\theta = o_p(1)$. Given assumption S4 part 4, White (2000), Corollary 3.48, implies that $\frac{1}{n}\sum_{s=t_0}^{t}(\widehat{Y}_s - \mathbb{E}\widehat{Y}_s) \xrightarrow{p} 0$. Since, $\left\| D_\theta^{1/2} (\mathcal{R}_a^{0\prime} \mathcal{W}(\theta) \mathcal{R}_a^0)^{-1} \mathcal{R}_a^{0\prime} \mathcal{W}(\theta) \right\|^2 < \infty$ for all $\theta \in \Theta$ (see above) and $\frac{1}{n}\sum_{s=t_0}^{t} \mathbb{E}\widehat{Y}_s < \infty$, the continuous mapping theorem implies $\frac{1}{n}\sum_{s=t_0}^{t} \widehat{\delta}_{a,s}(\theta) - \mathbb{E}\tilde{\delta}_{a,s} \xrightarrow{p} 0$ for all $\theta \in \Theta$. Hence, $|\widehat{S}(\theta) - S(\theta)| \leq \|\widehat{m}_n(\theta) - m(\theta)\|^2 + 2\|\widehat{m}_n(\theta) - m(\theta)\|^2 \|m(\theta)\|^2 \xrightarrow{p} 0$.

□

Proof of Proposition A3. Equation (S7) under Assumptions S6 and S7 corresponds to exactly to assumptions (i)-(v) in Theorem 5.23 of White (2000). The theorem implies that

$$
\widehat{\mathrm{Var}}(\hat{b}_{a,\Xi_c})^{-1/2} \sqrt{n} (\hat{b}_{a,\Xi_c} - b_{a,\Xi_c}) \xrightarrow{d} N(0, I_{K_a L_\xi}) \ .
$$

By Assumptions S1 and S5 and Proposition 2 we have that $b_{a,\Xi_c} = 0$ under H_0 and thus

$$
n \hat{b}_{a,\Xi_c}' \widehat{\mathrm{Var}}(\hat{b}_{a,\Xi_c})^{-1} \hat{b}_{a,\Xi_c} \xrightarrow{d} \chi^2_{K_a L_\xi}
$$

under H_0. Hence, we reject $H_0 : \phi^0 \in \Phi^{\mathrm{opt}}$ for any level of confidence α when we have that $n \hat{b}_{a,\Xi_c}' \widehat{\mathrm{Var}}(\hat{b}_{a,\Xi_c})^{-1} \hat{b}_{a,\Xi_c} > \chi^2_{K_a L_\xi, 1-\alpha}$. □

Table S1: Structural Break Tests for $\mathcal{R}^0_{x,h}$

FFR: 1990-2007

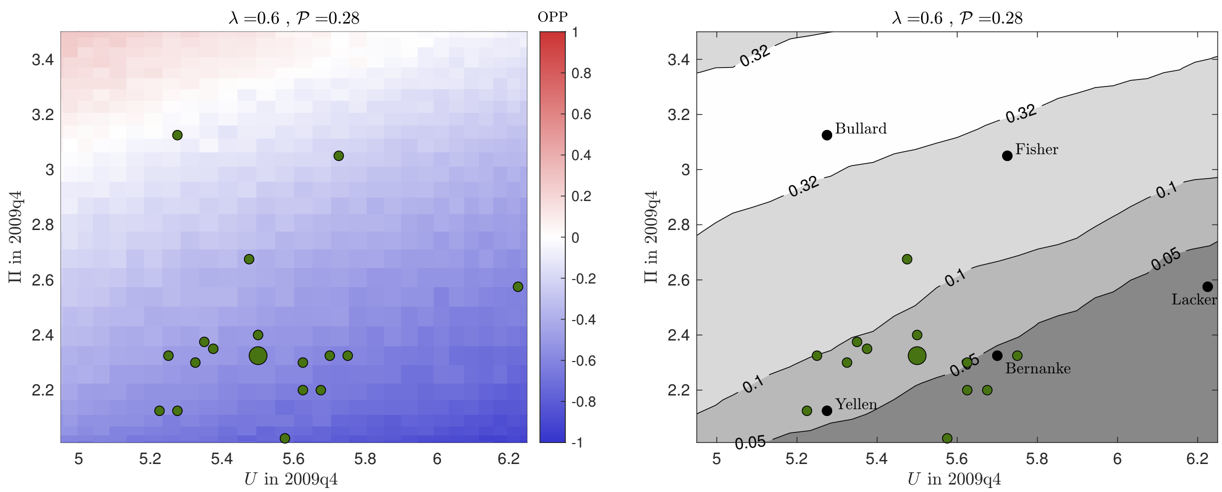
h	π	u
0	0.49	0.57
5	0.43	0.10
10	0.58	0.27
15	0.44	0.29
20	0.60	0.62

Slope: 2008-2018

h	π	u
0	0.71	0.62
5	0.72	0.36
10	0.96	0.47
15	0.81	0.34
20	0.50	0.79

Notes: We report the fixed-regressor bootstrap p-values for testing the hypothesis $H_0 : m = 0$ vs $H_1 : m = 1$, where m is the number of breaks in the causal effect $\mathcal{R}^0_{x,h}$ of an innovation to x (the fed funds rate or the slope of the yield curve) on inflation (π) or unemployment (u) after $h = 0, 5, 10, 15, 20$ quarters. In the top panel, the fed funds rate is instrumented with the high frequency monetary policy surprises of Kuttner (2001). In the bottom panel the slope of the yield curve is instrumented with surprises to the ten-year Treasury yield (orthogonalized with respect to surprises to the current fed funds rate). The bootstrap was implemented following Boldea, Cornea-Madeira and Hall (2019).

Figure S1: The OPP statistic across the economic outlook, April 2008

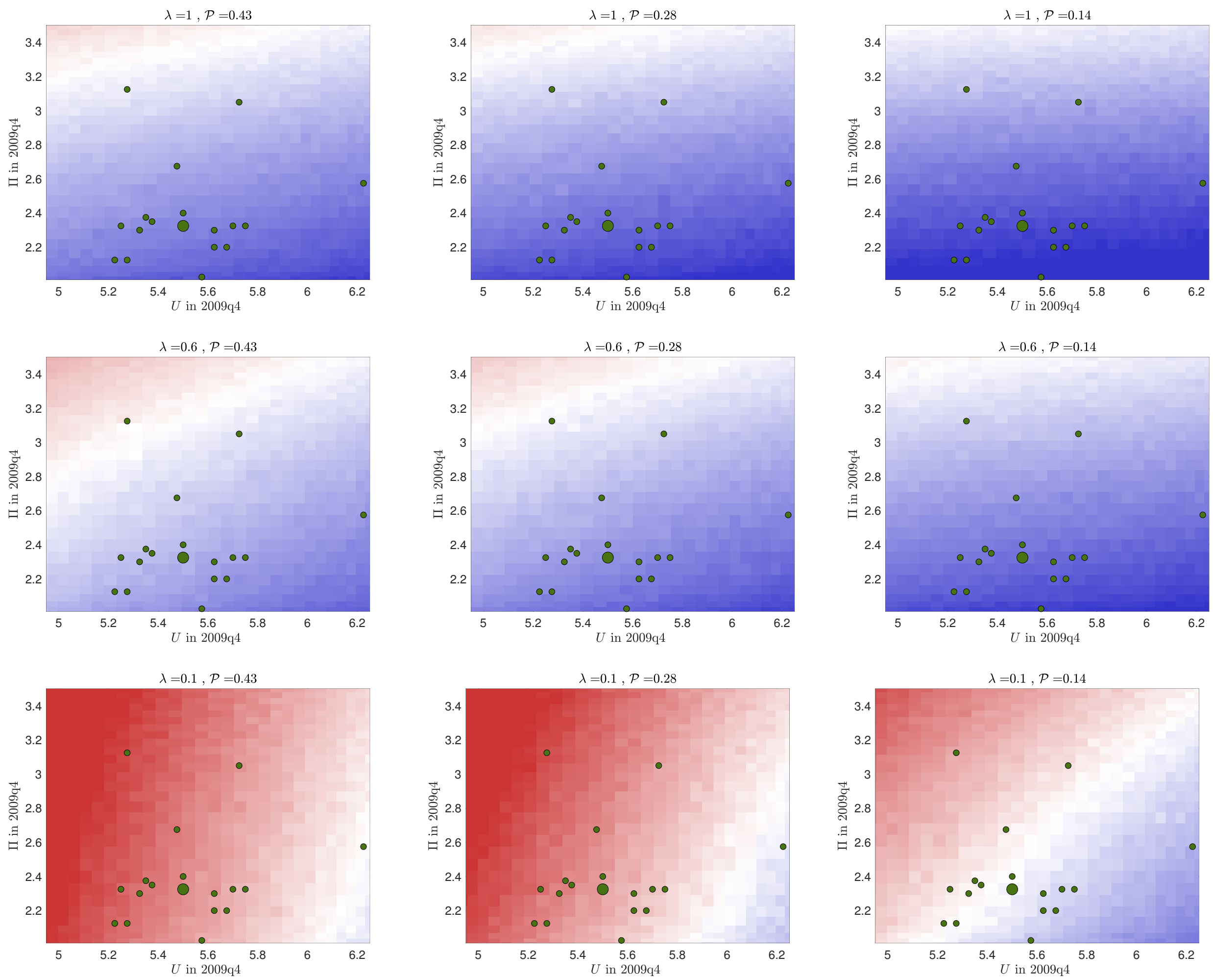


(a) OPP statistic

(b) policy optimality test

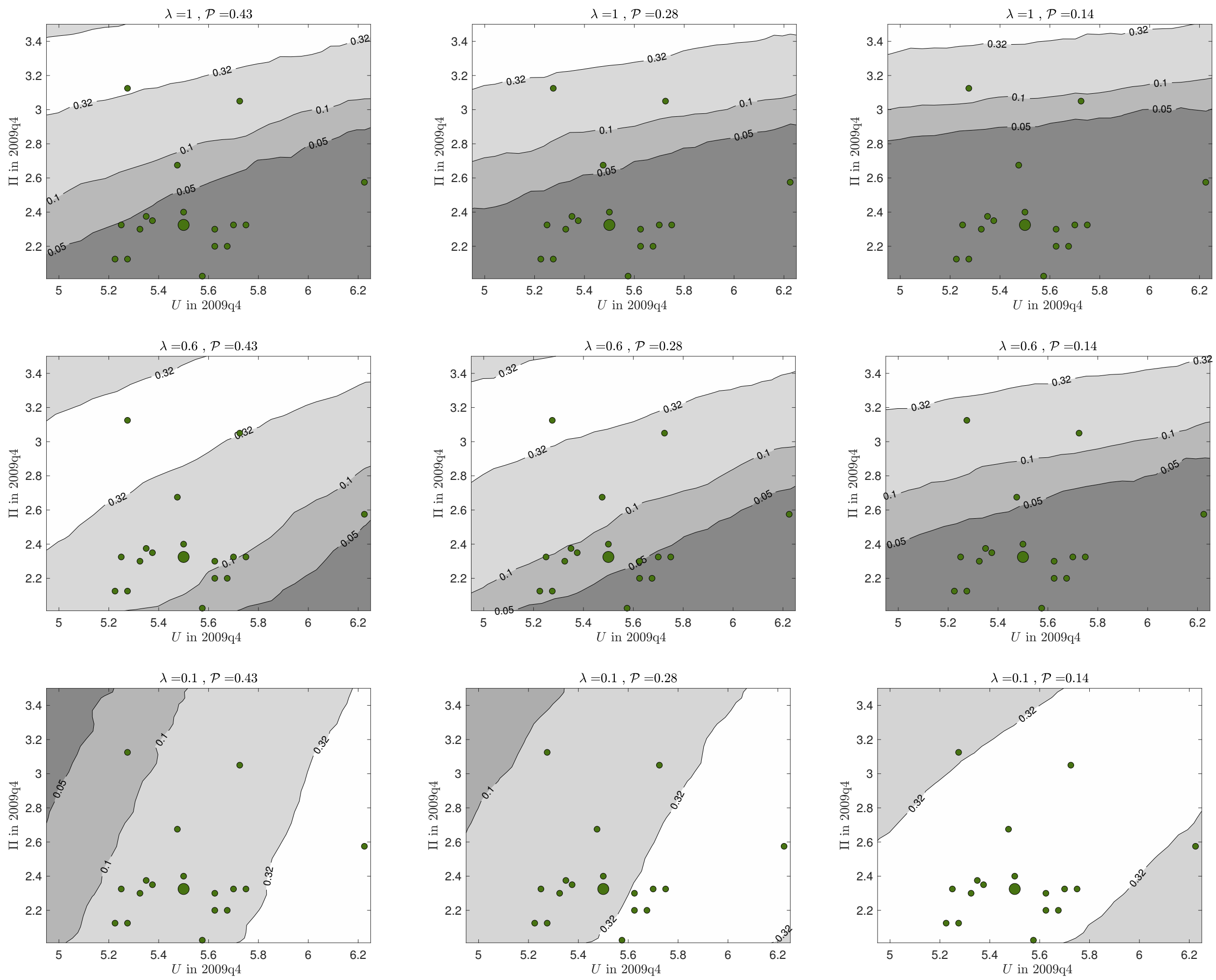
Notes: Panel (a): The heatmap depicts the magnitude of the OPP for different economic outlooks (unemployment in 2009-Q4 on the x-axis and inflation in 2009-Q4 on the y-axis) with positive OPP values in red (the OPP statistic calling for tighter policy) and negative OPP values in blue (the OPP statistic calling for looser policy). The green dots mark the conditional forecasts of the different FOMC members and the large green dot marks the FOMC median forecast. Panel (b): The black lines are the contours for p-values of the policy optimality test, denoting probabilities of an optimal policy of .32, .1 and .05. The shaded grey regions depict the corresponding confidence sets (lighter tones denote higher p-values). The dots mark the conditional forecasts of the different FOMC members and the large dot marks the FOMC median forecast. For dots inside the dark-grey area, the policy optimality test can reject the null that the policy is optimal at a 5% confidence level: there is a less than 5% probability that the policy choice is optimal.

Figure S2: The OPP statistic across the sufficient statistics, April 2008



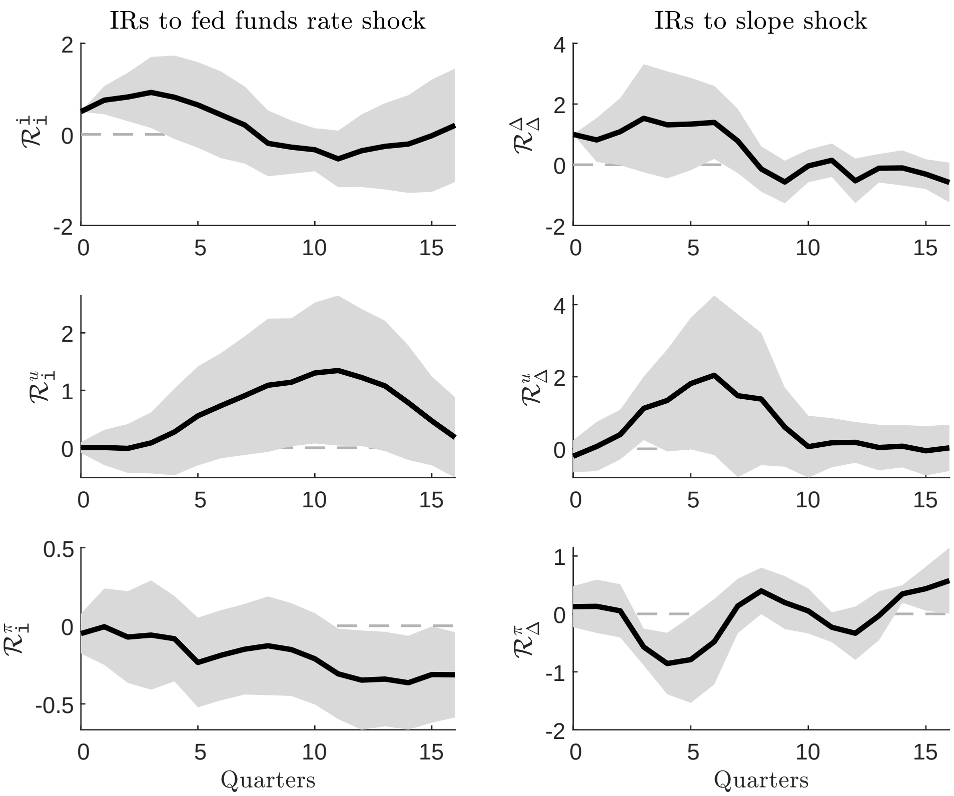
Notes: The heatmaps depict the magnitude of the OPP statistic for different values of the sufficient statistics: (i) the economic outlook (unemployment in 2009-Q4 on the x-axis and inflation in 2009-Q4 on the y-axis), (ii) the weight on unemployment (λ), and (iii) the Phillips multiplier ($\mathcal{P}$). Positive OPP values appear in red (the OPP thought experiment "calling" for tighter policy) and negative OPP values appear in blue (the OPP thought experiment "calling" for looser policy). The dots mark the conditional forecasts of the different FOMC members and the large dot marks the FOMC median forecast.

Figure S3: Inverting the policy optimality test, April 2008



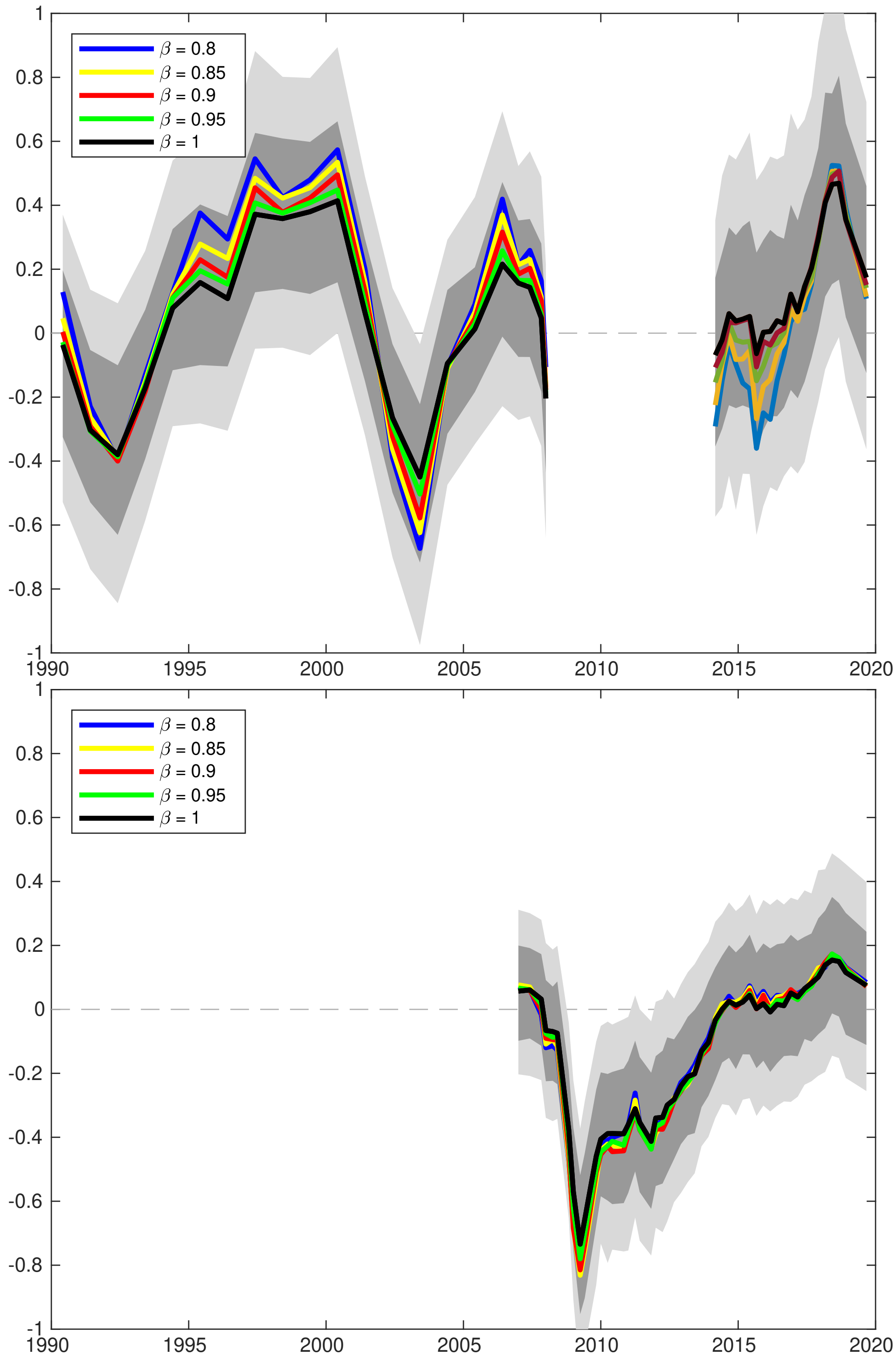
Notes: The panels show the optimality confidence sets for different values of the sufficient statistics: (i) the economic outlook (unemployment in 2009-Q4 on the x-axis and inflation in 2009-Q4 on the y-axis), (ii) the weight on unemployment (λ), and (iii) the Phillips multiplier ($\mathcal{P}$). The black lines are the contours for p-values of the policy optimality test, denoting probabilities of an optimal policy of .32, .1 and .05. The shaded grey regions depict the corresponding confidence sets (lighter tones denote higher p-values). The dots mark the conditional forecasts of the different FOMC members and the large dot marks the FOMC median forecast. For dots inside the white area, the policy optimality test cannot reject the null that the policy is optimal at the 32% confidence level: there is a more than 32% probability that the policy choice is optimal.

Figure S4: Impulse responses to policy innovations



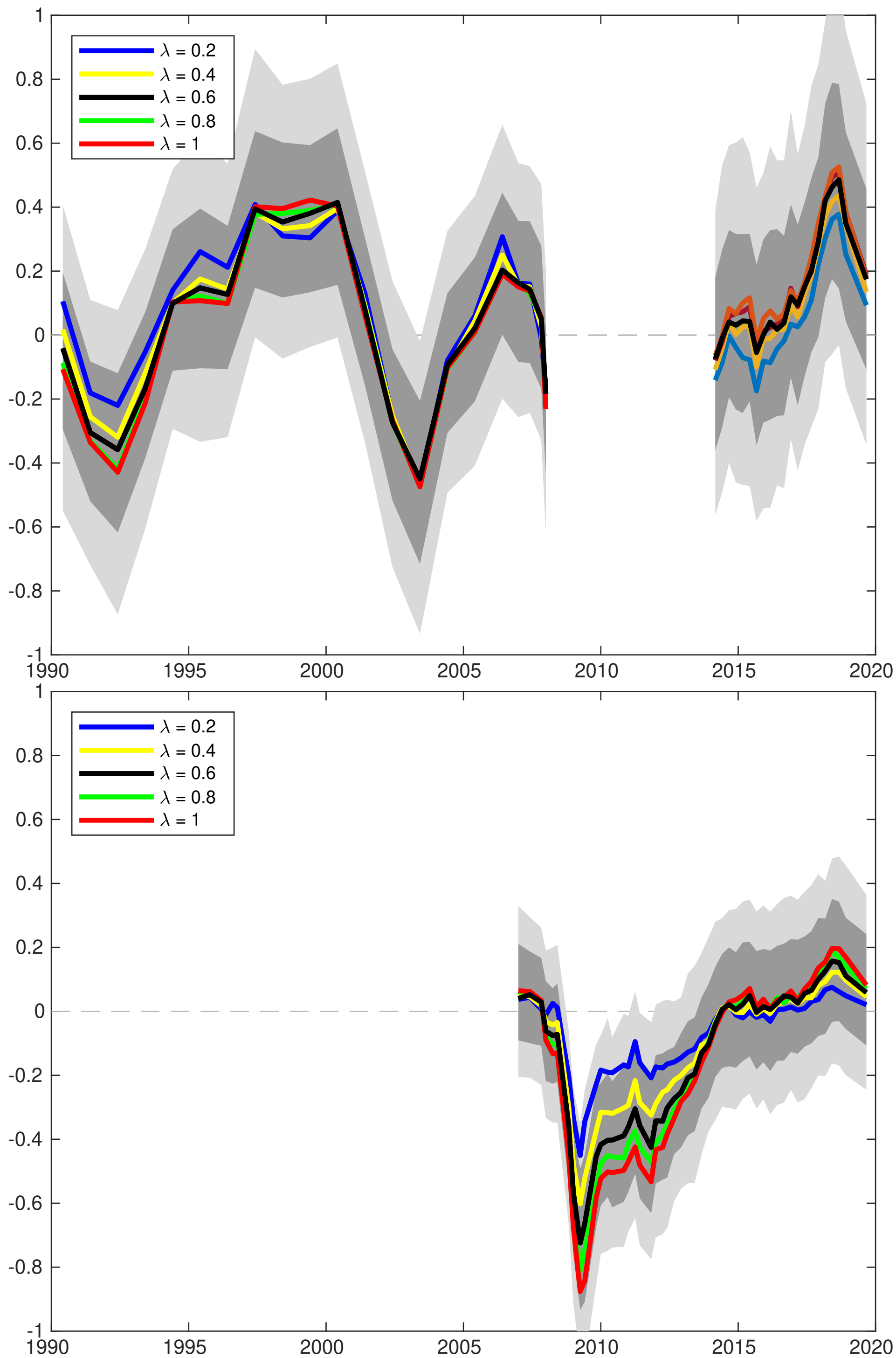
Notes: Left panel: impulse responses (IRs) of the fed funds rate, inflation and unemployment gaps to a fed funds rate shock. Right panel: impulse responses (IRs) of the slope of the yield curve, inflation and unemployment gaps to a slope policy shock. Shaded bands denote the 95 percent confidence intervals.

Figure S5: OPP FOR DIFFERENT β



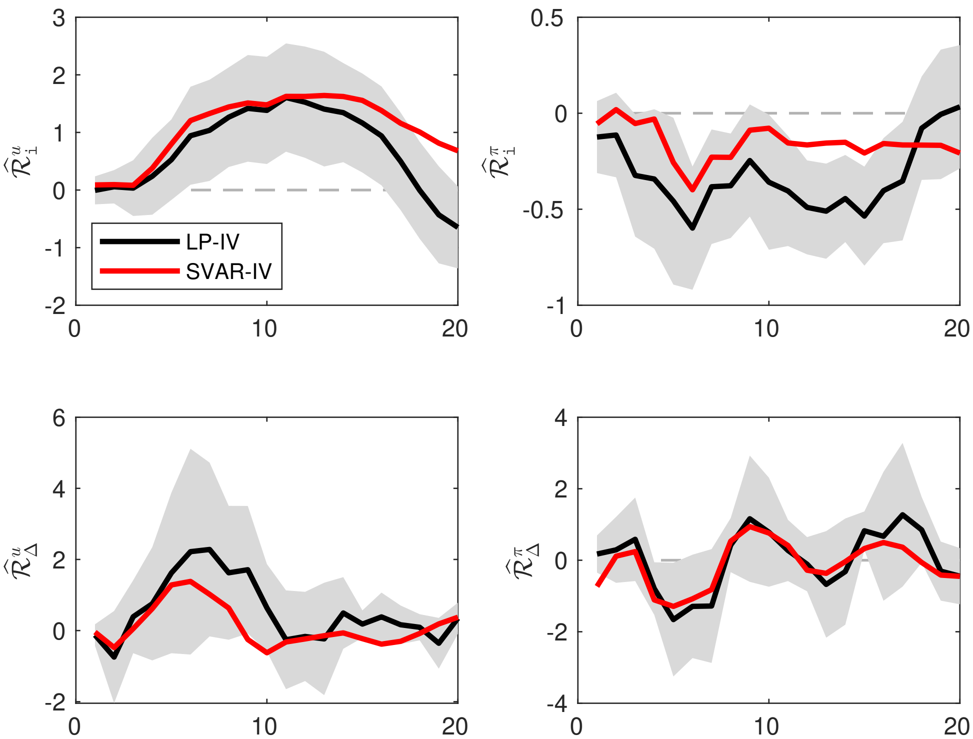
Notes: Top panel: fed funds rate OPP over 1990-2019. The black line corresponds to the OPP with $\beta = 1$ and the colored lines correspond to the OPP with β's between 0.8 and 1. Bottom panel: slope OPP over 2008-2019. The black line corresponds to the OPP with $\beta = 1$ and the colored lines correspond to the OPP with β's between 0.8 and 1. The grey areas capture impulse response and mis-specification uncertainty at respectively 68% (darker shade) and 95% (lighter shade) confidence.

Figure S6: OPP FOR DIFFERENT λ



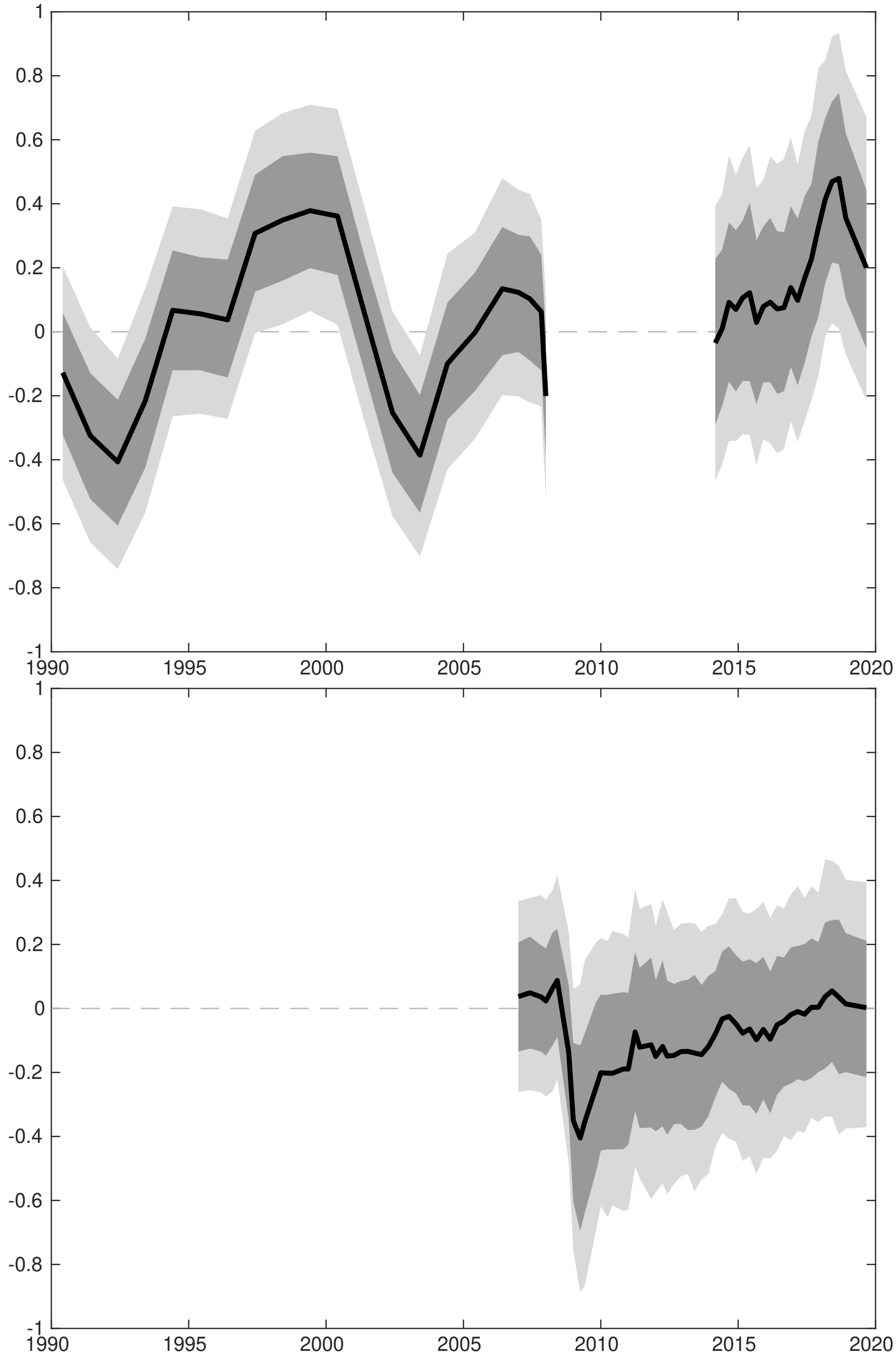
Notes: Top panel: fed funds rate OPP over 1990-2019. The black line corresponds to the OPP with $\lambda = 0.6$ and the colored lines correspond to the OPP with λ's between 0.2 and 1.2. Bottom panel: slope OPP over 2008-2019. The black line corresponds to the OPP with $\lambda = 0.6$ and the colored lines correspond to the OPP with λ's between 0.2 and 1.2. The grey areas capture impulse response and mis-specification uncertainty at respectively 68% (darker shade) and 95% (lighter shade) confidence.

Figure S7: LP-IV vs SVAR-IV



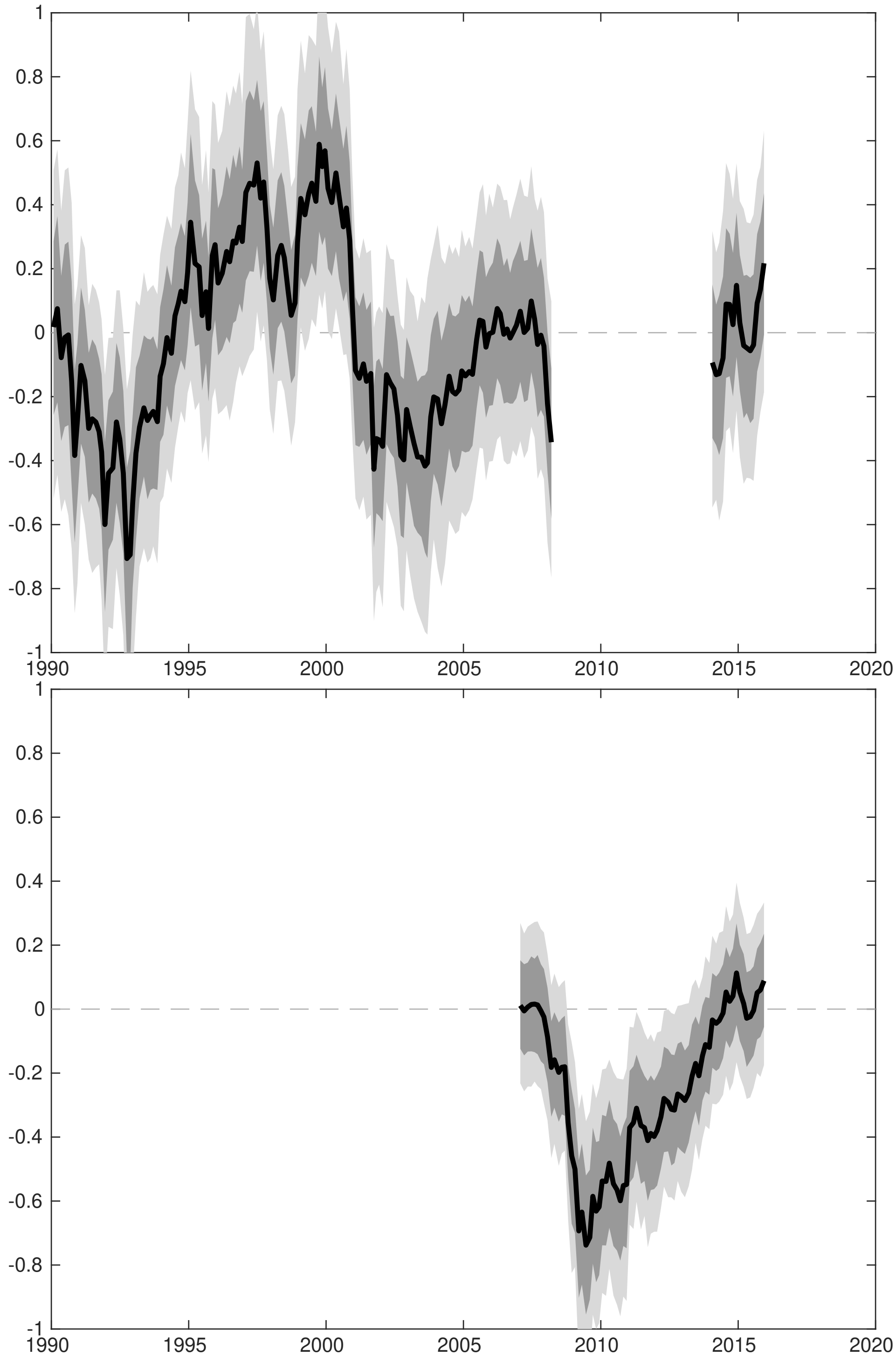
Notes: The panels compare the LP-IV and SVAR-IV estimates of the impulse responses of inflation (left column) or unemployment (right column) to an innovation to the fed funds rate (top row) or to the slope of the yield curve (bottom row).

Figure S8: Using SVAR-IV to estimate $\mathcal{R}_a^0$



Notes: OPP sequences over 1990-2019. Top panel: fed funds rate OPP estimated using SVAR-IV estimates for $\mathcal{R}_a^0$. Bottom panel: slope OPP estimated using SVAR-IV estimates for $\mathcal{R}_a^0$.

Figure S9: Using the Greenbook forecasts to measure $\mathbb{E}_t Y_t^0$



Notes: OPP sequences over 1990-2019. Top panel: fed funds rate OPP estimated using the Greenbook forecasts for $\mathbb{E}_t Y_t^0$. Bottom panel: slope OPP estimated using the Greenbook forecasts for $\mathbb{E}_t Y_t^0$. The grey areas capture impulse response and mis-specification uncertainty at respectively 68% (darker shade) and 95% (lighter shade) confidence.