A Sufficient Statistics Approach for Macro Policy Evaluation*

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

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

[(b)] Universitat Pompeu Fabra, BSE and CREI

May 22, 2023

Abstract

The evaluation of macroeconomic policy decisions has traditionally relied on the formulation of a specific economic model. In this work, we show that two statistics are sufficient to detect, often even correct, non-optimal policies, i.e., policies that do not minimize the loss function. The two sufficient statistics are (i) forecasts for the policy objectives conditional on the policy choice, and (ii) the impulse responses of the policy objectives to policy shocks. Both statistics can be estimated without relying on a specific structural economic model. We illustrate the method by studying US monetary policy decisions.

JEL classification: C14, C32, E32, E52.

Keywords: Optimal policy, Impulse responses, Forecasting.

*First draft: March 2020. We thank Adrien Auclert, Davide Debortoli, Martin Eichenbaum, Stefano Eusepi, Jordi Galí, Alain Guay, Ethan Ilzetzki, Sylvain Leduc, Huiyu Li, Pascal Michaillat, Emi Nakamura, Robert Ulbricht, Nicolas Petrosky-Nadeau, Adam Shapiro, Jón Steinsson and seminar participants at Bank of England, Berkeley, Boston College, Brown, CREI, McGill, 2020 NBER Summer Institute (ME), Ohio State, Ottawa, UQAM, UT Austin, Cleveland Fed, St Louis Fed and San Francisco Fed for helpful comments. We graciously thank Refet Gürkaynak for sharing his data on high-frequency monetary surprises. Mesters acknowledges financial support from the European Research Council via Starting Grant 101041145 "POLICYMETRICS" and the Spanish Ministry of Economy and Competitiveness through the Severo Ochoa Programme for Centres of Excellence in R&D (CEX2019-000915-S). The views expressed in this paper are the sole responsibility of the authors and do not necessarily reflect the views of the Federal Reserve Bank of San Francisco or the Federal Reserve System.

1 Introduction

Despite impressive recent progress in structural macro modeling, policy makers often resort to heuristics to decide on policy; combining insights from different models and relying heavily on judgment calls and instincts.[1] This practical approach has benefits in terms of robustness to model mis-specification, but a major downside is that it can be difficult to identify the most appropriate course of policy. Without a specific economic model, how can a policy maker be confident that a policy decision is appropriate? For instance, how to determine that the magnitude and timing of a fiscal package is well calibrated, or that the monetary stance is appropriate, e.g., in a "Goldilocks zone" that best balances inflation and unemployment.

In this paper, we show that it is not necessary to know the full structure of the economy to evaluate or even decide on macroeconomic policy. We consider a policy maker facing a time t decision problem —how to set the policy path today given the state of the economy—, and we show that two statistics are sufficient to detect and correct non-optimal decisions, i.e., policies that do not minimize the loss function.

Our approach is based on the gradient of the loss function with respect to policy shocks. Although little studied in the literature, this gradient has two key attractive properties: (i) it is informative about the optimality of a policy decision, and (ii) it is relatively easy to compute, depending only on two well-known statistics.

First, for a large class of models the gradient with respect to policy shocks must be zero under an optimal policy. From this necessary condition it follows that this gradient is sufficient to *evaluate* a policy decision, i.e., to detect a non-optimal policy. Moreover, for a broad class of linear dynamic models setting the gradient to zero is necessary and sufficient to characterize the optimal allocation. In other words, for that class of models, the gradient is also sufficient to *compute* the optimal policy.

Second, for a large class of loss functions the gradient with respect to policy shocks is given by the (weighted) product of two simple statistics (i) the forecasts for the policy objectives conditional on the policy choice, and (ii) the effects of policy shocks on the policy objectives. These two statistics are already central and well understood concepts for policy makers (e.g., Orphanides, 2019). Our contribution is to show that these two statistics alone can be used to rigorously evaluate and even set policy.

Intuitively, the first sufficient statistic —the forecasts for the objectives— serves to capture the state of the economy at time t —the characteristics of the time t decision problem— and to define a scenario under a baseline policy rule. The second sufficient statistic —the impulse responses to policy shocks— serves to explore whether deviating from that rule can produce a lower loss. At an optimal policy, the gradient —a weighted product of the two

[1] See e.g., Svensson (2003), Mishkin (2010) and Blinder (2020).

statistics— should be zero, and forecasts and impulse responses should be orthogonal: it should not be possible to use the impulse responses to adjust the forecasts and lower the loss function. This orthogonality condition forms the basis of our sufficient statistics approach to policy evaluation.

Importantly, the two sufficient statistics can be estimated using reduced form econometric models. To construct the forecast —the first sufficient statistic—, one only needs to compute the best linear prediction for the policy objectives under the baseline rule, and for that purpose we can draw on a variety of methods developed by the forecasting literature (e.g., Elliot and Timmermann, 2016). Similarly, the impulse responses to policy shocks —the second sufficient statistic— can be estimated from reduced form models in combination with identification restrictions or instrumental variables (e.g., Ramey, 2016). Naturally, the estimation of these sufficient statistics does require some assumptions: it requires correctly specifying a reduced form econometric model, the validity of identifying restrictions and satisfying regularity conditions required for inference.

To formally implement our sufficient statistics approach, we do not work with the gradient but instead with its rescaled version: the *Optimal Policy Perturbation* (OPP). Like the gradient, the OPP is entirely determined by the two sufficient statistics, but it also has a direct economic interpretation. The OPP is the adjustment to the baseline policy scenario that exactly corrects an optimization failure when the loss function is quadratic and the (unspecified) underlying model is linear. In other words, for linear models the OPP allows to compute the optimal policy from sufficient statistics alone.

We then generalize the OPP statistic to take into account a number of considerations that are of importance for the actual implementation of the sufficient statistics approach.

First, in a dynamic setting, the policy stance depends not only on the current level of the policy instrument but also on its entire expected path. As a result, computing the optimal policy path from the OPP requires the identification of policy news shocks —shocks to the expected future level of the policy instrument— at all possible horizons. While this data requirement is unlikely to be met in practice, we show how a *subset OPP* statistic, which only uses a subset of all possible policy shocks, can be used to evaluate and improve (though not fully correct) non-optimal policy decisions. For instance, if only contemporaneous policy shocks can be credibly identified, the corresponding subset OPP will only evaluate and improve the short-end of the policy path.

Second, policy makers often face constraints on their policy choice, coming for instance from physical constraints such as the zero-lower bound, or from prior commitments. We thus generalize the OPP statistic to incorporate constraints on the policy maker's problem. Intuitively, instead of working with the gradient of the loss function, the idea is to work with the gradient of the Lagrangian that incorporates the constraints. The *constrained OPP* is then defined as the adjustment to the policy instruments, which sets the Lagrangian to zero and makes the policy choice optimal given the constraints.

Third, policy decisions are often conducted sequentially, creating the possibility of dynamic inconsistency in policy decisions: a policy path that is optimal as of time $t-1$ may not be optimal when viewed as a decision problem at time t (e.g. Kydland and Prescott, 1977). We show that the OPP can be adjusted to suppress time inconsistent adjustments, with no additional information requirement beyond the two sufficient statistics. Intuitively, the key is to constrain the time t OPP such that the policy maker continues to satisfy the optimality conditions of the time $t-1$ problem.

In the main text, we derive all the properties of the OPP under the class of linear models. In the web-appendix, we show how the OPP statistic (and its variants) can be used to evaluate policy decisions in non-linear models. Intuitively, the gradient captures a necessary condition of optimality: at the optimum there should not exist *any* rule adjustment (including the OPP) that can lower the loss, so that the OPP should always be zero at an optimal policy. However, in non-linear models, an OPP adjustment may not always improve policy, as the gradient need not be a sufficient condition of optimality. The key non-linearity that breaks the OPP policy improvement property is when the coefficients of the policy block affect the coefficients of the non-policy block. Intuitively, an OPP adjustment will generally change a policy maker's reaction to the state of the economy —it will change the policy rule—. If the coefficients of the policy rule affect the economy above and beyond their effects on the policy instrument, there is no longer any guarantee that an OPP adjustment will improve policy, simply because the OPP does not take into account that nonlinear feedback.

To illustrate our sufficient statistics approach to macro policy evaluation, we study US monetary policy decisions over 1990-2022. We start from the Fed's dual inflation–unemployment mandate, and we estimate/recover the sufficient statistics underlying the OPP. We estimate impulse responses using high-frequency monetary surprises as instrumental variables (e.g. Eberly, Stock and Wright, 2020), and we use as forecasts the FOMC Survey of Economic Projections —the policy makers' own forecasts—.

While the contemporaneous fed funds rate has not been set exactly at its optimal level since 1990, the adjustments suggested by the OPP are (in absolute value) overall small, averaging only 25 basis points over the full sample. There are however some noteworthy instances of non-optimal policies. For instance, we find evidence that the Fed should have lowered the fed funds rate faster in the early stage of the Great Recession (when the zero lower bound was not yet binding). We can also reject the optimality of unconventional monetary policy operations in the middle of the Great recession, with the OPP calling for a more aggressive use of unconventional policy measures —LSAP or QE— to lower the slope of the yield curve, a conclusion echoing that of Eberly, Stock and Wright (2020).

In our final exercise, we provide a real time analysis of monetary policy over 2019-2022, paying particular attention to fed funds rate policy in the early stage of the COVID recovery. We find that monetary policy was appropriate in the first half of 2021 despite surging inflation, *once* we take into account the Sept. 2020 FOMC commitment to delay liftoff until the labor market recovery was complete. However, we do find that the Fed should have raised the fed funds rate at its 11/01/2021 meeting, almost 5 months earlier than the actual lift-off date.

The remainder of this paper is organized as follows. We continue the introduction by relating our sufficient statistics OPP approach to existing approaches in the literature. In the next section we provide a simple example that informally explains how we can evaluate and improve macro policy using sufficient statistics. Section 3 formally introduces the general environment and Section 4 presents the OPP statistic and its extensions. Section 5 discusses the implementation of our approach, notably the estimation of the two sufficient statistics. In Section 6 we apply our methodology to empirically study monetary policy decisions in the US. Section 7 concludes and provides potential avenues for further research.

Relation to literature

In the wake of the Lucas (1976) critique, the macroeconomic literature has built elaborate micro-founded models in order to study optimal policy problems.[2] In this paper, we show how an optimal policy problem —how to set the policy path today given the state of the economy?— can be recast as an econometric problem, in fact as two separate econometric tasks —forecasting and impulse response estimation—. Thus, our paper uncovers an important but so far overlooked link between the forecasting literature (e.g., Elliot and Timmermann, 2016), the structural impulse response literature (e.g. Ramey, 2016) and the optimal policy literature.

Related to our paper but with a different focus, McKay and Wolf (2022) considerably expand the scope of impulse responses for counter-factual policy rule analysis. They show that in a general family of linearized structural macroeconomic models the impulse responses to policy news shocks are sufficient to construct policy rule counter-factuals that are robust to the Lucas critique. Our finding that the gradient of the loss function with respect to policy shocks is sufficient to characterize the optimal policy relies on the same insight, though the OPP requires an additional statistic beyond impulse responses —the forecast— to capture the nature of the time t decision problem.

Our sufficient statistics approach to macro policy evaluation naturally shares important similarities with the sufficient statistic approach that originated in public finance (e.g.

[2]See e.g., Chari, Christiano and Kehoe (1994); Woodford (2010); Michaillat and Saez (2019).

Chetty, 2009). Both methods exploit the fact that the welfare consequences of a policy can be derived from high-level elasticities, allowing for policy evaluation without making parametric assumptions or estimating the structural primitives of fully specified models. One feature specific to our macro focus is that we postulate a loss function at the macro level, consistent with the fact that the loss function is often determined by political factors or by statutory requirement. For instance, it is the US Congress that mandates the Federal Reserve to seek stable inflation and full employment. That said, our approach can equally be applied to problems with micro-founded loss functions.

Finally, our sufficient statistics approach can be seen as a key input in the context of forecast-targeting rules (e.g., Svensson and Woodford, 2005; Woodford, 2013). Different from a Taylor (1993)-type instrument rule, a forecast targeting rule specifies that the policy path must be adjusted such that the forecasts for the policy objectives satisfy the first-order conditions of optimality, as implied by a specific structural model. Outside of a specific structural model however these first-order conditions were typically unknown, leaving policy makers with imprecise forecast targeting criteria (e.g., Svensson, 1999). The sufficient statistics approach fills this gap, because the OPP captures precisely the first-order conditions of optimality, without being tied to any particular structural model.

2 A simple example

Before formally describing our general framework, we first present a simple example to illustrate how two key statistics are sufficient to evaluate and improve macro policies. The example is based on Galí (2015, Section 5.1.1), which discusses the optimal policy problem under discretion in the baseline New Keynesian model.

Consider a central bank with loss function

$$\mathcal{L}_t = \frac{1}{2}(\pi_t^2 + x_t^2) \ , \tag{1}$$

with π_t the inflation gap and x_t the output gap. The central bank has only one instrument: the nominal interest rate i_t.

The log-linearized baseline New-Keynesian model is defined by a Phillips curve and an intertemporal (IS) curve given by

$$\pi_t = \mathbb{E}_t \pi_{t+1} + \kappa x_t + \xi_t \ , \tag{2}$$

$$x_t = \mathbb{E}_t x_{t+1} - \frac{1}{\sigma}(i_t - \mathbb{E}_t \pi_{t+1}) \ , \tag{3}$$

with ξ_t an iid cost-push shock. We posit $\kappa\sigma > 1$.

In this paper, we are interested in evaluating policy decisions. To that effect, consider a policy decision that is determined by a policy rule of the form

$$i_t = \phi \pi_t + \epsilon_t \ , \tag{4}$$

with $\phi > 1$ to guarantee a unique equilibrium and ϵ_t a policy shock. Given rule (4), a policy choice is then a pair (ϕ, ϵ_t).

Solving the model and expressing the endogenous variables $Y_t = (\pi_t, x_t)'$ and i_t as functions of the exogenous shocks, we get

$$Y_t = \Gamma_y \xi_t + \mathcal{R}_y \epsilon_t \qquad \text{and} \qquad i_t = \Gamma_i \xi_t + \mathcal{R}_i \epsilon_t \ , \tag{5}$$

with

$$\mathcal{R}_y = \begin{bmatrix} \frac{-\kappa/\sigma}{1+\kappa\phi/\sigma} \\ \frac{-1/\sigma}{1+\kappa\phi/\sigma} \end{bmatrix} \ , \quad \Gamma_y = \begin{bmatrix} \frac{1}{1+\kappa\phi/\sigma} \\ \frac{-\phi/\sigma}{1+\kappa\phi/\sigma} \end{bmatrix} \ , \quad \mathcal{R}_i = \frac{1}{1+\kappa\phi/\sigma} \quad \text{and} \quad \Gamma_i = \frac{\phi}{1+\kappa\phi/\sigma} \ ,$$

where $\mathcal{R}_y$ captures the impulse responses of the policy objectives Y_t to the policy shocks ϵ_t, while Γ_y captures the impulse response to a ξ_t shock. Similarly, Γ_i and $\mathcal{R}_i$ capture the effect of these shocks on the policy rate.

The planner's problem

The optimal allocation can be characterized by minimizing the loss function with respect to π_t, x_t and i_t subject to the Phillips curve and (IS) curve constraints. This gives the well-known optimal targeting rule

$$x_t = -\kappa \pi_t \ , \tag{6}$$

which can can be implemented by an instrument rule of the form (4) with $\phi^{\text{opt}} = \kappa\sigma$ and $\epsilon_t = 0$ (e.g., Galí, 2015). Combining (6) and (5), we can compute the optimal policy rate from

$$i_t^{\text{opt}} = \Gamma_i^{\text{opt}} \xi_t \ , \qquad \text{with} \qquad \Gamma_i^{\text{opt}} = \frac{\kappa\sigma}{1+\kappa^2} \ .$$

A limitation of this approach to characterize the optimal policy is that it requires the full underlying model, that is the exact specification and coefficients of the Phillips and (IS) curves. As we discussed in the introduction, this information requirement can be hard to meet in practice.

An alternative characterization of the optimal policy

We will now see that there is another way to characterize the optimal targeting rule and the optimal policy; an approach that does not require knowing the details of the model, in this case the Phillips curve or the (IS) curve.

The idea is to start from some initial policy rule and then study whether modifying that rule can lower the loss function. Specifically, consider some baseline policy choice given by the pair (ϕ^0, ϵ_t^0). This policy choices implies the allocation Y_t^0 and i_t^0, and the impulse responses under ϕ^0 are denoted by with a 0 superscript, i.e., $\mathcal{R}_y^0$, $\mathcal{R}_i^0$, etc... The idea is then to modify that baseline policy rule with some adjustment, or perturbation, δ_t, i.e.

$$i_t = \phi^0 \pi_t + \epsilon_t^0 + \delta_t \ . \tag{7}$$

Proceeding as with our derivation of (5), the model solution becomes

$$Y_t = \underbrace{\Gamma_y^0 \xi_t + \mathcal{R}_y^0 \epsilon_t}_{Y_t^0} + \mathcal{R}_y^0 \delta_t \qquad \text{and} \qquad i_t = \underbrace{\Gamma_i^0 \xi_t + \mathcal{R}_i \epsilon_t^0}_{i_t^0} + \mathcal{R}_i \delta_t \ , \tag{8}$$

These expressions, akin to laws of motion for Y_t and i_t following a δ_t rule adjustment, show that the effects of a change δ_t in the policy rule can be computed from $\mathcal{R}_y$ and $\mathcal{R}_i$: the impulse response to a policy shock.

We can now establish a key result: the optimal targeting rule can be derived by setting the gradient of the loss function with respect to a perturbation δ_t to zero. Formally, we have the equivalence

$$i_t^0 = i_t^{\text{opt}} \qquad \Longleftrightarrow \qquad \left.\frac{\partial \mathcal{L}_t}{\partial \delta_t}\right|_{\delta_t = 0} = \mathcal{R}_y^{0\prime} Y_t^0 = 0 \ , \tag{9}$$

where $\mathcal{R}_y^{0\prime} Y_t^0 = 0$ *is* the optimal targeting rule.

To prove the result, use (8) to compute the gradient

$$\begin{aligned} \left.\frac{\partial \mathcal{L}_t}{\partial \delta_t}\right|_{\delta_t = 0} &= \mathcal{R}_y^{0\prime} Y_t^0 \\ &= \frac{1}{\sigma + \kappa \phi^0} (-\kappa, -1) \left(\pi_t^0, x_t^0\right)' \\ &= \frac{-1}{\sigma + \kappa \phi^0} \left(\kappa \pi_t^0 + x_t^0\right) \ . \end{aligned} \tag{10}$$

The term in parenthesis is zero under the optimal targeting rule (6); $x_t^0 = -\kappa \pi_t^0$, which establishes the equivalence.

Intuitively, under an optimal policy there should not exist any rule adjustment δ_t that can lower the loss function. But since the effect of a rule adjustment δ_t is the same as the effect of a policy shock of the same size (equation (8)), this also means that it should *not* be possible to use the impulse responses $\mathcal{R}_y^0$ to better stabilize the forecasts Y_t^0: impulse responses and forecasts should be orthogonal, i.e., $\mathcal{R}_y^{0'} Y_t^0 = 0$.

It is also easy to see from (8) that the gradient with respect to a perturbation δ_t is the same as the gradient with respect to a policy shock ϵ_t, and we can talk interchangeably of the gradient with respect to a rule perturbation or to a policy shock. The reason is again that the effect of an adjustment to the policy rule is the same as the effect of a policy shock of the same size. However, ϵ_t and δ_t are different objects: ϵ_t is an exogenous variable (a policy shock) while δ_t is a choice variable.

The equivalence (9) states that setting the gradient of the loss function with respect to δ_t to zero is necessary and sufficient to characterize the optimal policy and optimal targeting rule. We will now discuss two specific applications that exploit this equivalence: First, the $\Longrightarrow$ relation implies that $\mathcal{R}_y^{0'} Y_t^0 = 0$ forms a testable condition to evaluate whether a given policy decision is optimal. Second, the $\Longleftarrow$ relation can be exploited to compute the optimal policy starting from any initial policy decision that implies a unique equilibrium.

Policy evaluation with sufficient statistics

We first illustrate how the two statistics $\mathcal{R}_y^0$ and Y_t^0 can be used to evaluate a policy decision.

Consider a policy maker following the rule (4) and proposing the policy i_t^0 given by the pair (ϕ^0, ϵ_t^0). The policy can be non-optimal for two reasons: $\phi^0 \neq \phi^{\text{opt}}$ or $\epsilon_t^0 \neq 0$.

For i_t^0 to be optimal, we just saw that the gradient of the loss function evaluated at i_t^0 must be zero, i.e., that

$$\mathcal{R}_y^{0'} Y_t^0 = 0 \ , \tag{11}$$

where $\mathcal{R}_y^0$ is the effect of policy shocks under ϕ^0 and Y_t^0 is the allocation under i_t^0. Equation (11) forms a testable condition to evaluate policy decisions: if $\mathcal{R}_y^{0'} Y_t^0 \neq 0$, we can conclude that i_t^0 is not optimal.

Since the statistics $\mathcal{R}_y^0$ and Y_t^0 will typically need to be estimated, the gradient will be computed with uncertainty, and our evaluation of the optimality of a policy choice will resemble a hypothesis test: a statement that the policy is not optimal for some confidence level.

Optimal policy with sufficient statistics

Second, the sufficient condition of optimality —the $\Longleftarrow$ relation in (9)— can be exploited to *compute* the optimal policy i_t^{opt} using two statistics: (i) Y_t^0 —the allocation under i_t^0—, and (ii) $\left\{\mathcal{R}_y^0, \mathcal{R}_i^0\right\}$ —the impulse responses to policy shocks—.

The idea is to use the "law of motion" for Y_t following a δ_t perturbation —$Y_t = Y_t^0 + \mathcal{R}_y^0\delta_t$— to find the adjustment that will best lower the loss function. Specifically, this consists in finding a δ_t^* that satisfies

$$\delta_t^* = \underset{\delta_t}{\text{argmin}} \ \mathcal{L}_t \qquad \text{s.t.} \qquad Y_t = Y_t^0 + \mathcal{R}_y^0\delta_t \ .$$

A closed-form solution for δ_t^* is

$$\delta_t^* = -(\mathcal{R}_y^{0\prime}\mathcal{R}_y^0)^{-1}\mathcal{R}_y^{0\prime}Y_t^0 \ . \tag{12}$$

and we can compute i_t^{opt} from[3]

$$i_t^{\text{opt}} = i_t^0 + \mathcal{R}_i^0\delta_t^* \ . \tag{13}$$

The statistic δ_t^* is what we call the Optimal Policy Perturbation (OPP). Clearly, the OPP has the same property as the gradient — $\delta_t^* \neq 0$ implies $i_t^0 \neq i_t^{\text{opt}}$—. Morover, the OPP δ_t^* allows to compute i_t^{opt} from some arbitrary initial policy choice i_t^0.

To understand *how* the OPP modifies i_t^0 to get to the optimal policy, we can combine (5) and (12) to decompose the OPP as the sum of two terms with

$$\delta_t^* = \frac{1}{\mathcal{R}_i^0}\underbrace{(\Gamma_i^{\text{opt}} - \Gamma_i^0)}_{\text{non-optimal rule}}\xi_t \qquad - \qquad \underbrace{\epsilon_t^0}_{\text{exogeneous mistake}} \qquad . \tag{14}$$

The decomposition shows how the OPP corrects the two possibles sources of optimization failure: (i) an exogenous policy mistake ($\epsilon_t^0 \neq 0$), or (ii) a non-optimal reaction to the cost-push shock ($\Gamma_i^0 \neq \Gamma_i^{\text{opt}}$). In other words, the OPP adjustment consists in (i) removing the effect of the exogenous policy mistake ϵ_t^0, and (ii) changing the policy rule by having the policy maker reacts optimally to the cost-push shock (changing from Γ_i^0 to Γ_i^{opt}). Importantly, the decomposition also highlights that while the OPP can detect and fully correct a non-optimal policy, it does not discriminate between the two sources of optimization failure, and it is not informative about whether a non-optimal policy is due to a non-optimal re-

[3]We can verify that (13) indeed holds in this example and that $i_t^0 + \mathcal{R}_i\delta_t^*$ is the optimal policy:

$$\begin{aligned}
i_t^0 + \mathcal{R}_i^0\delta_t^* &= \mathcal{R}_i^0\epsilon_t^0 + \Gamma_i^0\xi_t - \mathcal{R}_i^0\left(\mathcal{R}_y^{0\prime}\mathcal{R}_y^0\right)^{-1}\mathcal{R}_y^{0\prime}Y_t^0 \\
&= \mathcal{R}_i^0\epsilon_t^0 + \Gamma_i^0\xi_t - \mathcal{R}_i^0\left(\mathcal{R}_y^{0\prime}\mathcal{R}_y^0\right)^{-1}\mathcal{R}_y^{0\prime}(\mathcal{R}_y^0\epsilon_t^0 + \Gamma_y^0\xi_t) \\
&= \left(\Gamma_i^0 - \mathcal{R}_i^0\left(\mathcal{R}_y^{0\prime}\mathcal{R}_y^0\right)^{-1}\mathcal{R}_y^{0\prime}\Gamma_y^0\right)\xi_t \\
&= \frac{\kappa\sigma}{1+\kappa^2}\xi_t = \Gamma_i^{\text{opt}}\xi_t = i_t^{\text{opt}} \ .
\end{aligned}$$

where the last line uses the explicit expressions for the impulse responses from (5) to simplify the expression.

action function or to an exogenous policy mistake. Doing so would require more structural structural assumption, for instance specifying a policy rule.

3 General framework

We now generalize the OPP approach for a general dynamic macro environment that includes a large class of macro models encountered in the literature but without committing to a particular one. We first formalize the policy problem, the environment and our objectives.

The policy problem

Consider a policy maker at time t who aims to stabilize the expected path of the policy objectives $\mathbf{Y}_t = (y_t', y_{t+1}', \ldots)'$, with $y_t = (y_{1,t}, \ldots, y_{M_y,t})'$ such that there are M_y objectives for each horizon. The policy maker can form expectations about the future paths of $\mathbf{Y}_t$, based on the time t information set $\mathcal{F}_t$. We denote the expectation operator by $\mathbb{E}_t(\cdot) = \mathbb{E}(\cdot|\mathcal{F}_t)$.

The objective of the policy maker is to minimize the expected loss function as of time t

$$\mathcal{L}_t = \frac{1}{2}\mathbb{E}_t \mathbf{Y}_t' \mathcal{W} \mathbf{Y}_t \ , \tag{15}$$

where $\mathcal{W} = \text{diag}(\beta \otimes \lambda)$ denotes a diagonal map of preferences with $\lambda = (\lambda_1, \ldots, \lambda_{M_y})'$ capturing the weights on the different variables and $\beta = (\beta_0, \beta_1, \ldots)'$ the discount factors for the different horizons. While we consider a quadratic loss function in the baseline treatment, the web-appendix extends our approach to arbitrary convex loss functions.

To minimize the loss function the policy maker can set M_p policy instruments at time t, denoted by $p_t = (p_{1,t}, \ldots, p_{M_p,t})'$. In addition, the policy maker can set the expected policy path, that is the time-t expected values for p_{t+1}, $p_{t+2}, \ldots$ and so on. We denote by $\mathbb{E}_t \mathbf{P}_t = \mathbb{E}_t(p_t', p_{t+1}', \ldots)'$ the corresponding expected future policy path as a function of the time-t information set.

Environment

We consider a linear environment which can be justified for small fluctuations around a steady-state.[4] A generic model for the non-policy block of the economy at time t is given by

$$\begin{cases} \mathcal{A}_{yy}\mathbb{E}_t\mathbf{Y}_t - \mathcal{A}_{yw}\mathbb{E}_t\mathbf{W}_t - \mathcal{A}_{yp}\mathbb{E}_t\mathbf{P}_t &= \mathcal{B}_{yx}\mathbf{X}_{-t} + \mathcal{B}_{y\xi}\Xi_t \\ \mathcal{A}_{ww}\mathbb{E}_t\mathbf{W}_t - \mathcal{A}_{wy}\mathbb{E}_t\mathbf{Y}_t - \mathcal{A}_{wp}\mathbb{E}_t\mathbf{P}_t &= \mathcal{B}_{wx}\mathbf{X}_{-t} + \mathcal{B}_{w\xi}\Xi_t \end{cases} , \tag{16}$$

[4]The web-appendix discusses nonlinear extensions.

where $\mathbf{W}_t = (w_t', w_{t+1}', \ldots)'$ is a path for additional endogenous variables, the vector $\mathbf{X}_{-t} = (y_{t-1}', w_{t-1}', p_{t-1}', y_{t-2}', \ldots)'$ captures the initial conditions as summarized by the history of the variables y_t, w_t and p_t, and $\Xi_t = (\xi_t', \xi_{t,t+1}', \xi_{t,t+2}', \ldots)'$ denotes the path of the structural shocks to the economy. Specifically, ξ_t is the time t vector of structural (non-policy) shocks, while the shocks $\xi_{t,t+h}$, for $h = 1, 2, \ldots$, are news shocks: information revealed at time t about shocks that realize at time $t+h$. The vector Ξ_t thus includes all shocks that are released at time t. We normalize the news shocks to be mutually uncorrelated with mean zero and unit variance. The linear maps $\mathcal{A}_{..}$ and $\mathcal{B}_{..}$ are conformable and we define the time t information set in terms of the pre-determined inputs as $\mathcal{F}_t = \{\mathbf{X}_{-t}, \Xi_t\}$.

This model is general and accommodates a large class models found in the literature, not only standard New-Keynesian (NK) models (e.g., Smets and Wouters, 2007), but also modern heterogeneous agents NK models (Auclert et al., 2021).

Optimal policy

The optimal policy can be characterized by considering a planner who chooses the paths $\mathbf{Y}_t, \mathbf{W}_t$ and $\mathbf{P}_t$ in order to minimize the loss function, i.e.,

$$\min_{\mathbf{Y}_t, \mathbf{W}_t, \mathbf{P}_t} \mathcal{L}_t \qquad \text{s.t.} \qquad (16)\ . \tag{17}$$

Denote by $\mathbb{E}_t \mathbf{P}_t^{\text{opt}}$ the corresponding optimal policy path. Note that the problem defines the entire optimal policy path as a function of the information available at time t. For clarity of exposition, we will make the following assumption.

Assumption 1. *The optimal policy* $\mathbb{E}_t \mathbf{P}_t^{\text{opt}}$ *is unique.*

The assumption is not essential, and our results continue to hold when replacing $\mathbb{E}_t \mathbf{P}_t^{\text{opt}}$ with a set of optimal policies for which each element of the set solves (17).

Policy rule

We consider a generic model for the policy block with

$$\mathcal{A}_{pp} \mathbb{E}_t \mathbf{P}_t - \mathcal{A}_{py} \mathbb{E}_t \mathbf{Y}_t - \mathcal{A}_{pw} \mathbb{E}_t \mathbf{W}_t = \mathcal{B}_{px} \mathbf{X}_{-t} + \mathcal{B}_{p\xi} \Xi_t + \boldsymbol{\epsilon}_t\ , \tag{18}$$

where $\boldsymbol{\epsilon}_t = (\epsilon_t', \epsilon_{t,t+1}', \epsilon_{t,t+2}', \ldots)'$ is the path of policy news shocks. Specifically, the vector $\epsilon_t = (\epsilon_{1,t}, \ldots, \epsilon_{M_p,t})'$ includes the contemporaneous policy shocks to the M_p policy instruments and $\epsilon_{t,t+h}$ are policy news shocks: information revealed at time t about policy shocks that realize at time $t+h$. The policy news shocks $\boldsymbol{\epsilon}_t$ are mean zero with unit variance and uncorrelated with the initial conditions $\mathbf{X}_{-t}$ and the other non-policy shocks Ξ_t.

We collect all parameters of the policy rule (18) in $\phi = \{\mathcal{A}_{pp}, \mathcal{A}_{py}, \mathcal{A}_{pw}, \mathcal{B}_{px}, \mathcal{B}_{p\xi}\}$. A policy choice is then defined by a pair $(\phi, \boldsymbol{\epsilon}_t)$ consisting of the rule parameters ϕ and the policy news shocks $\boldsymbol{\epsilon}_t$.[5]

Consider some baseline policy choice $(\phi^0, \boldsymbol{\epsilon}_t^0)$ and denote by $\mathbb{E}_t\mathbf{P}_t^0$ and $\mathbb{E}_t\mathbf{Y}_t^0$ the associated baseline paths for the policy instruments and policy objectives. For now, that baseline choice is arbitrary except for the following assumption

Assumption 2. *The rule ϕ^0 underlying the baseline policy path $\mathbb{E}_t\mathbf{P}_t^0$ leads to a unique and determinate equilibrium.*

This assumption is necessary for impulse responses and forecasts —our two sufficient statistics— to be well defined. The following lemma formalizes this.

Lemma 1. *Given the generic model* (16) *and the policy rule* (18)*, under $(\phi^0, \boldsymbol{\epsilon}_t^0)$ with ϕ^0 satisfying Assumption 2, we have*

$$\begin{array}{rcl} \mathbb{E}_t\mathbf{Y}_t^0 & = & \Gamma_y^0\mathbf{S}_t + \mathcal{R}_y^0\boldsymbol{\epsilon}_t^0 \\ \mathbb{E}_t\mathbf{P}_t^0 & = & \Gamma_p^0\mathbf{S}_t + \mathcal{R}_p^0\boldsymbol{\epsilon}_t^0 \end{array}, \tag{19}$$

with $\mathbf{S}_t = (\mathbf{X}_{-t}', \Xi_t')'$ and $\mathbb{E}(\boldsymbol{\epsilon}_t^0\mathbf{S}_t') = 0$.

Lemma 1 characterizes the model solution, expressing the endogenous variables $\mathbb{E}_t\mathbf{Y}_t^0$ and $\mathbb{E}_t\mathbf{P}_t^0$ as functions of the state of the economy $\mathbf{S}_t = (\mathbf{X}_{-t}', \Xi_t')'$, which captures initial conditions $\mathbf{X}_{-t}$ and the non-policy shocks Ξ_t, and the policy shocks $\boldsymbol{\epsilon}_t^0$. Note that $\mathbb{E}_t\mathbf{Y}_t^0$ and $\mathbb{E}_t\mathbf{P}_t^0$ are the oracle forecasts as of time t, that is the expectations for $\mathbf{Y}_t$ and $\mathbf{P}_t$ conditional on the information set $\mathcal{F}_t$ and the policy choices $(\phi^0, \boldsymbol{\epsilon}_t^0)$.

In addition, Lemma 1 defines the impulse responses of $\mathbb{E}_t\mathbf{Y}_t^0$ and $\mathbb{E}_t\mathbf{P}_t^0$ to policy and non-policy shocks. We have that $\mathcal{R}_y^0$ captures the impulse responses of the objectives to policy news shocks at different horizons —from horizon-0 (ϵ_t) to any horizon $h > 0$ ($\epsilon_{t,t+h}$)— under the rule ϕ^0. Similarly, $\mathcal{R}_p^0$ captures the effects of these shocks on the expected policy path. The impulse responses to the state of the economy $\mathbf{S}_t$ are denoted by Γ_y and Γ_p, but these statistics will not play an important role in our approach.

4 Optimal policy with sufficient statistics

In this section we show how, for the generic class of models captured by (16), we can use sufficient statistics to characterize the optimal policy.

[5] While it may seem surprising to allow for non-zero policy shocks —why would anyone propose a policy path involving (necessarily non-optimal) exogenous factors—, this is important in a policy evaluation context, as we do not want to a priori rule out the possibility of exogeneous policy mistakes.

As in the simple example of Section 2, our approach consists in starting form some baseline policy choice $(\phi^0, \boldsymbol{\epsilon}_t^0)$ and modifying that baseline policy rule ϕ^0 as follows

$$\mathcal{A}_{pp}^0 \mathbb{E}_t \mathbf{P}_t - \mathcal{A}_{py}^0 \mathbb{E}_t \mathbf{Y}_t - \mathcal{A}_{pw}^0 \mathbb{E}_t \mathbf{W}_t = \mathcal{B}_{px}^0 \mathbf{X}_{-t} + \mathcal{B}_{p\xi}^0 \boldsymbol{\Xi}_t + \boldsymbol{\epsilon}_t^0 + \boldsymbol{\delta}_t, \tag{20}$$

where $\boldsymbol{\delta}_t = (\delta'_{t,t}, \delta'_{t,t+1}, \ldots)'$ is a path of adjustments to the policy rule equations. To stress that we view $\boldsymbol{\delta}_t$ as a choice parameter we write $\mathbb{E}_t \mathbf{P}_t(\boldsymbol{\delta}_t)$ and $\mathbb{E}_t \mathbf{Y}_t(\boldsymbol{\delta}_t)$ to denote the expected policy path and expected allocation under the *modified* policy rule (20).

The following lemma establishes how $\boldsymbol{\delta}_t$ affects the initial allocation, that is the equilibrium under the policy choice $(\phi^0, \boldsymbol{\epsilon}_t^0)$.

Lemma 2. *Given the generic model* (16) *and the modified policy rule* (20)*, under* $(\phi^0, \boldsymbol{\epsilon}_t^0)$ *with* ϕ^0 *satisfying Assumption 2, we have*

$$\begin{array}{rcl} \mathbb{E}_t \mathbf{Y}_t(\boldsymbol{\delta}_t) & = & \mathbb{E}_t \mathbf{Y}_t^0 + \mathcal{R}_y^0 \boldsymbol{\delta}_t \\ \mathbb{E}_t \mathbf{P}_t(\boldsymbol{\delta}_t) & = & \mathbb{E}_t \mathbf{P}_t^0 + \mathcal{R}_p^0 \boldsymbol{\delta}_t \end{array}, \tag{21}$$

where $\mathbb{E}_t \mathbf{P}_t^0$ *and* $\mathbb{E}_t \mathbf{Y}_t^0$ *are given in Lemma 1.*

In other words, the adjustment $\boldsymbol{\delta}_t$ has a linear effect on the baseline paths $\mathbb{E}_t \mathbf{Y}_t^0$ and $\mathbb{E}_t \mathbf{Y}_t^0$, and that effect is given by the impulse responses to policy news shocks. The lemma effectively mimics equation (8) from the simple example. and it will allow us to study how a rule perturbation $\boldsymbol{\delta}_t$ modifies the baseline allocation. The important difference is that $\boldsymbol{\delta}_t$ adjusts the entire expected policy path as there is one component in $\boldsymbol{\delta}_t$ for each horizon of the policy path.

With Lemma 2, we can now characterize the optimal policy as defined in (17) using sufficient statistics.

Proposition 1. *Given the generic model* (16) *and the modified policy rule* (20)*, under* $(\phi^0, \boldsymbol{\epsilon}_t^0)$ *with* ϕ^0 *satisfying Assumption 2, we have under Assumption 1 that*

$$\mathbb{E}_t \mathbf{P}_t^0 = \mathbb{E}_t \mathbf{P}_t^{\text{opt}} \qquad \Longleftrightarrow \qquad \nabla_{\boldsymbol{\delta}_t} \mathcal{L}_t(\boldsymbol{\delta}_t)\big|_{\boldsymbol{\delta}_t = 0} = \mathcal{R}_y^{0\prime} \mathcal{W} \mathbb{E}_t \mathbf{Y}_t^0 = \mathbf{0}\ , \tag{22}$$

where $\mathcal{L}_t(\boldsymbol{\delta}_t) = \frac{1}{2} \mathbb{E}_t \mathbf{Y}_t(\boldsymbol{\delta}_t)' \mathcal{W} \mathbf{Y}_t(\boldsymbol{\delta}_t)$.

The proposition characterizes the optimal policy in terms of the gradient of the loss function with respect to $\boldsymbol{\delta}_t$. Intuitively, if $\mathbb{E}_t \mathbf{P}_t^0 = \mathbb{E}_t \mathbf{P}_t^{\text{opt}}$ there is no adjustment $\boldsymbol{\delta}_t$ that can lower the loss function and the gradient of the loss function evaluated at $\boldsymbol{\delta}_t = \mathbf{0}$ should be equal to zero. As in the simple example (cf. equation (9)) the first order condition $\mathcal{R}_y^{0\prime} \mathcal{W} \mathbb{E}_t \mathbf{Y}_t^0 = \mathbf{0}$ is the optimal targeting rule, implying that the policy maker should choose ϕ such that this condition holds. The benefit of the equivalence (22) is that it allows us to characterize the optimal policy from only two statistics: the forecasts $\mathbb{E}_t\mathbf{Y}_t^0$ and the impulse responses $\mathcal{R}_y^0$.

4.1 The OPP statistic

While the gradient fully characterizes the optimal policy, it will be useful to work with its rescaled version: the *Optimal Policy Perturbation* (OPP). Specifically, the idea of the OPP is to find the "best" adjustment $\boldsymbol{\delta}_t$ to the baseline rule ϕ^0, that is find the $\boldsymbol{\delta}_t$ that solves the problem

$$\boldsymbol{\delta}_t^* = \underset{\boldsymbol{\delta}_t}{\operatorname{argmin}} \ \mathcal{L}_t(\boldsymbol{\delta}_t) \qquad \text{s.t.} \qquad \mathbb{E}_t\mathbf{Y}_t(\boldsymbol{\delta}_t) = \mathbb{E}_t\mathbf{Y}_t^0 + \mathcal{R}_y^0\boldsymbol{\delta}_t \ , \tag{23}$$

where $\mathcal{L}_t(\boldsymbol{\delta}_t)$ is defined in Proposition 1. This perturbed policy problem has a closed form solution given by

$$\boldsymbol{\delta}_t^* = -(\mathcal{R}_y^{0\prime}\mathcal{W}\mathcal{R}_y^0)^{-1}\mathcal{R}_y^{0\prime}\mathcal{W}\mathbb{E}_t\mathbf{Y}_t^0 \ . \tag{24}$$

The perturbation $\boldsymbol{\delta}_t^*$ is the *Optimal Policy Perturbation* (OPP), generalizing the version introduced for the simple example, see equation (12). The only differences are the dimensions and the weighting of the policy objectives.[6]

Note how the OPP formula looks like the formula of a weighted least squares regression. In fact, the OPP uses $\mathcal{R}_y^0$ —the impulse responses to policy news shocks— in order to best stabilize the policy objectives, i.e., minimize the (weighted) sum-of-squares of $\mathbb{E}_t\mathbf{Y}_t(\boldsymbol{\delta}_t)$, the expected paths for the policy objectives after the policy adjustment. This is nothing but a regression of $\mathbb{E}_t\mathbf{Y}_t^0$ (the forecast before the adjustment) on $-\mathcal{R}_y^0$. The minus sign is present because the goal is not to best fit $\mathbb{E}_t\mathbf{Y}_t^0$, but instead to best "undo" movements in $\mathbb{E}_t\mathbf{Y}_t^0$.

We can now state the two key properties of the OPP.

Proposition 2. *Given the generic model* (16) *and the augmented policy rule* (20)*, under* $(\phi^0, \boldsymbol{\epsilon}_t^0)$ *with* ϕ^0 *satisfying Assumption 2, we have under Assumption 1 that*

1. $\mathbb{E}_t\mathbf{P}_t^0 = \mathbb{E}_t\mathbf{P}_t^{\text{opt}} \iff \boldsymbol{\delta}_t^* = \mathbf{0}$

2. $\mathbb{E}_t\mathbf{P}_t^{\text{opt}} = \mathbb{E}_t\mathbf{P}_t^0 + \mathcal{R}_p^0\boldsymbol{\delta}_t^*$.

First, mimicking Proposition 1 we have that if and only if the OPP is zero the policy of interest is equal to the optimal policy. From that property, we can *evaluate* policy decisions: if the OPP is non-zero, we will conclude that the policy path $\mathbb{E}_t\mathbf{P}_t^0$ is not optimal. Second, we

[6] Throughout we assume that the inverse $(\mathcal{R}_y^{0\prime}\mathcal{W}\mathcal{R}_y^0)^{-1}$ exists. If this is not the case this implies that the effects of the policy instruments are linearly dependent and we can remove one of the instruments from the analysis and simply proceed with the reduced set of instruments for which the invertibility requirement holds.

can use the OPP to *construct* the optimal policy path $\mathbb{E}_t \mathbf{P}_t^{\text{opt}}$ from some arbitrary baseline policy choice that implies a unique equilibrium.

To see how the OPP constructs the optimal policy by adjusting the baseline choice, assume that there exists a rule that underlies $\mathbb{E}_t \mathbf{P}_t^{\text{opt}}$, say ϕ^{opt}, and that leads to a unique equilibrium. Then, together with Assumption 2, we can write both policy paths $\mathbb{E}_t \mathbf{P}_t^0$ and $\mathbb{E}_t \mathbf{P}_t^{\text{opt}}$ as functions of (i) the initial conditions and (ii) the current and expected future shocks. Using Lemma 1, the two policy paths are given by

$$\mathbb{E}_t \mathbf{P}_t^{\text{opt}} = \Gamma_p^{\text{opt}} \mathbf{S}_t \qquad \text{and} \qquad \mathbb{E}_t \mathbf{P}_t^0 = \Gamma_p^0 \mathbf{S}_t + \mathcal{R}_p^0 \boldsymbol{\epsilon}_t^0 \ , \tag{25}$$

where the map Γ_p^{opt} captures the policy maker's optimal response to the state of the economy $\mathbf{S}_t = (\mathbf{X}_{-t}', \boldsymbol{\Xi}_t')'$.

Combining these expressions, we can decompose the OPP as

$$\mathcal{R}_p^0 \boldsymbol{\delta}_t^* = \left(\Gamma_p^{\text{opt}} - \Gamma_p^0\right) \mathbf{S}_t - \mathcal{R}_p^0 \boldsymbol{\epsilon}_t^0 \ . \tag{26}$$

Starting from a baseline choice $(\phi^0, \boldsymbol{\epsilon}_t^0)$, the OPP statistic corrects the two factors behind a non-optimal policy: (i) it changes how the policy maker responds to the state of the economy and makes the reaction function optimal $(\Gamma_p^{\text{opt}} - \Gamma_p^0)$, and (ii) it suppresses the policy mistakes, i.e., non-zero exogenous policy shocks $(-\boldsymbol{\epsilon}_t^0)$.

Discussion

At this point, it is useful to intuitively discuss the properties of the OPP implied by Proposition 2, and more generally the benefits as well as the limits of our sufficient statistics approach to macro policy evaluation.

Econometric-based optimal policy Through the OPP, Proposition 2 states that two statistics ($\mathbb{E}_t \mathbf{Y}_t^0$ and $\mathcal{R}_y^0$) are necessary and sufficient to fully characterize the optimal policy. In fact, we can view the OPP as using the statistics $\mathbb{E}_t \mathbf{Y}_t^0$ and $\mathcal{R}_y^0$ to characterize the optimal policy in two stages: in a first stage, $\mathbb{E}_t \mathbf{Y}_t^0$ serves to capture the state of the economy at time t —the characteristics of the time-t decision problem— and to define a baseline policy scenario under the rule ϕ^0. In a second stage the causal estimate $\mathcal{R}_y^0$ is used to find the deviation from that baseline scenario that produces the lowest loss.

In other words, the OPP can be seen as splitting the optimal policy problem into two separate *econometric* tasks: (i) a forecasting task —approximating $\mathbb{E}_t \mathbf{Y}_t^0$—, and (ii) a causal inference task —estimating $\mathcal{R}_y^0$—.

This econometric interpretation helps understand two attractive properties of the sufficient statistics approach.

First, the sufficient statistics approach has a lower information requirement than a structural model based approach, because reduced form econometric models can be used to compute forecasts and impulse responses. To construct a forecast for $\mathbf{Y}_t^0$, we only need a sample period when the rule ϕ^0 was in effect in order to be able to compute the best linear prediction for $\mathbf{Y}_t^0$. This is a much lower information requirement than the traditional model based approach, which requires specifying the entire model structure, i.e., the entire maps $\mathcal{A}, \mathcal{B}$ in (16). Similarly, the impulse responses to policy shocks can be estimated from a reduced form model, e.g. a VAR or local projection, in combination with identification restrictions or instrumental variables. We will discuss the estimation of the two sufficient statistics in detail in section 5.

Second, the sufficient statistics approach offers a simple two-step algorithm to *compute* the optimal policy path from reduced form models alone: In a first step, compute a set of baseline forecasts $\mathbb{E}_t\mathbf{P}_t^0$ and $\mathbb{E}_t\mathbf{Y}_t^0$ under *some* baseline rule ϕ^0, and in a second step use the impulse responses (under that same rule) to compute the OPP. The optimal policy path is then given by $\mathbb{E}_t\mathbf{P}_t^{\text{opt}} = \mathbb{E}_t\mathbf{P}_t^0 + \mathcal{R}_p^0\boldsymbol{\delta}_t^*$. In this algorithm, the baseline policy choice is an artifact used to capture the characteristics of the time-t problem —a tool to compute the optimal policy—, and any baseline policy choice can do as long as Assumption 2 is verified.

Robustness to the Lucas critique It may seem surprising that a reduced form approach is able to characterize the optimal policy, i.e., that our approach is robust to the Lucas critique. Indeed, a worry could be that the OPP is based on impulse responses to policy shocks —surprise deviations from a prevailing policy rule—, which need not be informative about the effects of alternative policy rules. Lemma 2 shows that this worry is unfounded: for linear models like (16) the effect of an adjustment $\boldsymbol{\delta}_t$ to the policy rule can be computed from the impulse responses to policy shocks alone (equation (21)). Intuitively, this property stems from the independence of the policy block and the non-policy block: a policy choice $(\phi^0, \boldsymbol{\epsilon}_t^0)$ affects the allocation *only* through its effects on the expected policy path. As a result, the impulse responses to shocks to the expected policy path are sufficient, in combination with $\mathbb{E}_t\mathbf{P}_t^0$, to construct the optimal policy path at time t. This property echoes the recent findings of McKay and Wolf (2022) that in a large family of linearized structural macro models the impulse responses to policy news shocks are sufficient to construct arbitrary policy rule counter-factuals that are robust to the Lucas critique.

Limits of the sufficient statistics approach We note two limitations of the sufficient statistics approach.

First, the lower information requirement of reduced form models is not without costs.

Depending on the particular application, there may not exist enough empirical evidence to estimate the impulse responses to *all* policy news shocks. In that case, the sufficient statistics approach can be used to evaluate and improve a baseline policy choice, but it can no longer be used to compute the optimal policy path. We will discuss this point in the next section.

Second, the environment in Section 3 is defined by a quadratic loss function (15) and a dynamic linear model (16). Going beyond this set-up is possible and we briefly discuss it here, leaving a general treatment for the Appendix.

Broadly speaking, the OPP statistic (and its variants) can still be used to *evaluate* policy decisions in non-linear models with convex loss. Intuitively, the gradient continues to capture a necessary condition of optimality: at the optimum there should not exist *any* rule adjustment (including the OPP) the can lower the loss, so that the OPP should always be zero at an optimal policy, and finding a non-zero OPP will indicate that the policy path is non-optimal.

However, in non-linear models, it may not be possible to use the OPP (at least, in its current form) to compute the optimal policy. The reason is that the gradient need not be a sufficient condition of optimality. The key non-linearity that breaks the equivalence result in Proposition 2 is when the coefficients ϕ of the policy rule affect the coefficients of the non-policy block ($\mathcal{A}_{..}$ and $\mathcal{B}_{..}$). In that case, it is no longer true that a policy rule affects the allocation only through the policy path, and the robustness to the Lucas critique (as discussed above) breaks down: the effects of policy news shocks are no longer able to capture the effects of changes in the policy rule ϕ.

4.2 Subset OPP

Computing the OPP requires the entire map $\mathcal{R}_y^0$ of impulse responses to policy shocks. In practice, computing all impulse responses can be infeasible as identifying shocks to every element of the expected policy path can be hard. Moreover, the policy shocks that we do identify in practice need not correspond to a single news shock, but could instead be a linear combination of multiple policy news shocks.

However, we can still use a subset or linear combination of policy shocks to evaluate and improve a baseline policy decision. Specifically, instead of computing the OPP adjustments at *all* horizons of the policy path, we can compute only the policy rule adjustments for the horizons (or linear combinations thereof) for which the effects can be identified.[7]

Let $\boldsymbol{\epsilon}_{a,t}^0$ denote the subset or linear combination of policy shocks that can be identified.

[7]Again, the two step decomposition of the optimal policy problem —forecasting $\mathbb{E}_t\mathbf{Y}_t^0$ and impulse response estimation $\mathcal{R}_y^0$— is crucial for this result. To construct forecasts we only need the best linear prediction for $\mathbf{Y}_t^0$ and the causal map $\mathcal{R}_y^0$ is not needed. Therefore, to evaluate a baseline policy choice we can use a subset of the causal map $\mathcal{R}_y^0$ to see if the baseline policy could be improved in the corresponding direction.

The subset of the causal effects $\mathcal{R}^0_{y,a}$ measures the effect of $\boldsymbol{\epsilon}^0_{a,t}$ on the policy objectives. Similarly $\boldsymbol{\epsilon}^0_{a^\perp,t}$ and $\mathcal{R}^0_{y,a^\perp}$ capture the shocks and the associated impulse responses of the non-identifiable policy shocks.

Let $\boldsymbol{\delta}_t = (\boldsymbol{\delta}_{a,t}, \boldsymbol{\delta}_{a^\perp,t})$ be the corresponding partitions of the policy perturbations. The *subset* OPP is defined by

$$\boldsymbol{\delta}^*_{a,t} = \underset{\boldsymbol{\delta}_{a,t}}{\operatorname{argmin}} \ \mathcal{L}_t(\boldsymbol{\delta}_{a,t}, \mathbf{0}) \qquad \text{s.t.} \qquad \mathbb{E}_t \mathbf{Y}_t(\boldsymbol{\delta}_{a,t}, \mathbf{0}) = \mathbb{E}_t \mathbf{Y}^0_t + \mathcal{R}^0_{y,a} \boldsymbol{\delta}_{a,t} \ , \tag{27}$$

which has the closed form solution

$$\boldsymbol{\delta}^*_{a,t} = -(\mathcal{R}^{0\prime}_{y,a} \mathcal{W} \mathcal{R}^0_{y,a})^{-1} \mathcal{R}^{0\prime}_{y,a} \mathcal{W} \mathbb{E}_t \mathbf{Y}^0_t \ . \tag{28}$$

The subset-OPP statistic has the following properties

Corollary 1. *Given the generic model* (16) *and the augmented policy rule* (20)*, under* $(\phi^0, \boldsymbol{\epsilon}^0_t)$ *with* ϕ^0 *satisfying Assumption 2, we have under Assumption 1 that*

1. $\boldsymbol{\delta}^*_{a,t} \neq 0 \qquad \Longrightarrow \qquad \mathbb{E}_t \mathbf{P}^0_t \neq \mathbb{E}_t \mathbf{P}^{\text{opt}}_t$

2. $\mathcal{L}_t(\boldsymbol{\delta}^*_{a,t}, \mathbf{0}) \leq \mathcal{L}_t(\mathbf{0}, \mathbf{0})$*, i.e. the adjusted path* $\mathbb{E}_t \mathbf{P}^0_t + \mathcal{R}^0_{a,p} \boldsymbol{\delta}^*_{a,t}$ *implies a lower loss than the initial path* $\mathbb{E}_t \mathbf{P}^0_t$*.*

Similar as in Proposition 2, if the subset OPP statistic is non-zero the policy $\mathbb{E}_t \mathbf{P}^0_t$ is non-optimal. Moreover, adjusting the baseline policy with the subset OPP will improve the baseline policy path, though it will generally not give the optimal path $\mathbb{E}_t \mathbf{P}^{\text{opt}}_t$. In other words, the subset OPP allows to compute the best policy path given the sufficient statistics available.

To help understand the intuition behind Corollary 1, it is helpful to restate it in words. Imagine that the empirical literature has only identified some combination of policy news shocks, that is some policy experiments for which we know the effects on both the expected policy path and the expected policy objectives. The subset OPP then consists in using these policy experiment to perturbate the baseline policy path $\mathbb{E}_t \mathbf{P}^0_t$ and search for a more optimal one within the restricted space spanned by these policy experiments. This approach will improve a baseline policy choice, but it need not find the optimal policy

4.3 Constrained OPP

The policy problem that we considered so far —problem (17)— is an unconstrained optimization problem: there are no restrictions on the policy path. In practice however, policy makers may face additional constraints, coming either from physical constraints (e.g., the lower-bound on the policy rate in the context of monetary policy) or from pre-commitments (e.g., a promise to keep the policy rate at zero for some time).

The sufficient statistics approach can easily be extended to incorporate such constraints. In a nutshell, the approach consists in replacing the loss function $\mathcal{L}_t$ with the Lagrangian that incorporates the constraints. The optimal allocation can then be fully characterized using the gradient of the Lagrangian (instead of the gradient of the loss function).

We allow the constraints to be general nonlinear functions of $\mathbb{E}_t\mathbf{Y}_t$ and $\mathbb{E}_t\mathbf{P}_t$ that can be written as[8]

$$C(\mathbb{E}_t\mathbf{Y}_t, \mathbb{E}_t\mathbf{P}_t) \geq \mathbf{c} \ , \tag{29}$$

where $C(\cdot,\cdot) : \mathbb{R}^\infty \times \mathbb{R}^\infty \to \mathbf{R}^{d_c}$ is the known constraint function and $\mathbf{c}$ a vector of constants of length d_c (possibly infinite). As an example, suppose that $\mathbb{E}_t\mathbf{P}_t = \mathbb{E}_t(i_t, i_{t+1}, \ldots)$ is the expected interest rate path of a central bank. We can impose the zero lower bound by setting $C(\mathbb{E}_t\mathbf{Y}_t, \mathbb{E}_t\mathbf{P}_t) = \mathbb{E}_t\mathbf{P}_t$ and $\mathbf{c} = \mathbf{0}$.

To incorporate the constraints into the policy problem we modify the original policy problem (17) to become

$$\min_{\mathbf{Y}_t, \mathbf{W}_t, \mathbf{P}_t} \mathcal{L}_t \qquad \text{s.t.} \qquad (16) \qquad \text{and} \qquad C(\ \mathbb{E}_t\mathbf{Y}_t, \mathbb{E}_t\mathbf{P}_t) \geq \mathbf{c} \ . \tag{30}$$

The optimal solution for $\mathbb{E}_t\mathbf{P}_t$ is denoted by $\mathbb{E}_t\mathbf{P}_t^{\text{opt},c}$, and for simplicity we assume that it is unique (as in Assumption 1).

The question that we study is then whether a baseline policy path $\mathbb{E}_t\mathbf{P}_t^0$ implied by a choice $(\phi^0, \boldsymbol{\epsilon}_t^0)$ is equal to the constrained optimal path and — if not — how it should be adjusted.

Following the same steps as with the unconstrained OPP, we can construct a *constrained OPP* statistic given by

$$\begin{aligned} \boldsymbol{\delta}_t^{c*} = \underset{\boldsymbol{\delta}_t}{\text{argmin}} \ \mathcal{L}_t(\boldsymbol{\delta}_t) \qquad & \text{s.t.} \quad \mathbb{E}_t\mathbf{Y}_t(\boldsymbol{\delta}_t) = \mathbb{E}_t\mathbf{Y}_t^0 + \mathcal{R}_y^0\boldsymbol{\delta}_t \ , \\ & \text{and} \quad C(\mathbb{E}_t \ \mathbf{Y}_t^0 + \mathcal{R}_y^0\boldsymbol{\delta}_t, \mathbb{E}_t\mathbf{P}_t^0 + \mathcal{R}_p^0\boldsymbol{\delta}_t) \geq \mathbf{c} \ . \end{aligned} \tag{31}$$

In contrast to the baseline OPP statistic that was obtained by solving (23), there exists no closed form solution for $\boldsymbol{\delta}_t^{c*}$. Nevertheless, we can easily solve this problem numerically as all inputs are the same as above. A subset constrained OPP can be formulated similarly by minimizing only with respect to $\boldsymbol{\delta}_{a,t}$.

The constrained OPP statistic has the same properties as the OPP, but now with respect to the optimal constrained policy choice defined in (30). Specifically, if and only if $\boldsymbol{\delta}_t^{c*} = 0$

[8] The web-appendix discusses a slightly more general treatment which also allows for constraints on the other endogenous variables $\mathbb{E}_t\mathbf{W}_t$.

we have that $\mathbb{E}_t \mathbf{P}_t^0 = \mathbb{E}_t \mathbf{P}_t^{\text{opt,c}}$. Moreover, we have that $\mathbb{E}_t \mathbf{P}_t^{\text{opt,c}} = \mathbb{E}_t \mathbf{P}_t^0 + \mathcal{R}_y^0 \boldsymbol{\delta}_t^{c*}$, we can use the constrained OPP to compute the optimal constrained policy path from some arbitrary baseline policy choice that implies a unique equilibrium. The web-appendix provides the formal statement and explicit examples.

4.4 Time consistent OPP

So far we have considered the problem of a policy maker making a one time decision about the policy path given the time-t information set.

Here we go one step further and study the sequential nature of policy making that is often encountered in practice: at different time intervals, policy makers convene to decide on their desired policy path.[9] As is well known, such sequential decision making process creates the possibility of dynamic inconsistency: a policy path that is optimal as of time $t-1$ may not be optimal viewed from a time decision problem as of time t (Kydland and Prescott, 1977). In this paper we do not take a stand on which policy problem the policy maker should consider —time-consistent or not— when making sequential decisions, but we provide the tools to evaluate and optimize policy decisions for any of the given problems. With that mind, this section develops a *time consistent OPP* that eliminates dynamic inconsistency.

As preliminary step, we first make dynamic inconsistency more explicit. Using the fact that the OPP fully characterizes the optimal policy (Proposition 1, part 1), we can contrast the time $t-1$ and the time t problems by comparing the OPPs for the time $t-1$ and the time t problems. We have

$$\boldsymbol{\delta}_{t-1}^* = \mathcal{D}^0 \mathbb{E}_{t-1} \mathbf{Y}_{t-1}^0 \quad \text{vs} \quad \boldsymbol{\delta}_t^* = \mathcal{D}^0 \mathbb{E}_t \mathbf{Y}_t^0 \ ,$$

where $\mathcal{D}^0 = -(\mathcal{R}_y^{0\prime} \mathcal{W} \mathcal{R}_y^0)^{-1} \mathcal{R}_y^{0\prime} \mathcal{W}$ can be seen as a weighting map capturing how the policy objectives $y_{t-1}, y_t, y_{t+1}, \ldots$ are weighted in the first order conditions for optimality.

Some simple manipulations give the decomposition

$$\boldsymbol{\delta}_t^* = \boldsymbol{\delta}_{t-1}^* + \underbrace{\mathcal{D}^0 \Delta \mathbb{E}_t \mathbf{Y}_t^0}_{\text{Information update}} + \underbrace{\Delta \mathcal{D}^0 \mathbb{E}_{t-1} \mathbf{Y}_{t-1}^0}_{\text{Preference shift}} \ , \tag{32}$$

where $\Delta \mathbb{E}_t(\cdot) = \mathbb{E}_t(\cdot) - \mathbb{E}_{t-1}(\cdot)$ is the information update operator and the map $\Delta \mathcal{D}^0 = [\mathcal{D}_1^0 - \mathbf{0}, \mathcal{D}_2^0 - \mathcal{D}_1^0, ...]$ is a "pseudo-difference" map with $\mathcal{D}_i^0$ the ith $\infty \times M_y$ block of $\mathcal{D}^0$, i.e. $\mathcal{D}^0 = [\mathcal{D}_1^0, \mathcal{D}_2^0, \ldots]$. There are two differences between the time t and $t-1$ optimization problems: (i) a different information set and (ii) a different objective function and set of instruments.

[9] For instance, the FOMC meets eight times a year to decide on its expected policy path and has some discretion to deviate from earlier announcements.

Consider a policy maker making an optimal decision from the time $t-1$ perspective (such that $\boldsymbol{\delta}^*_{t-1} = 0$). There are two forces that can lead that policy maker to adjust her policy path at time t relative to her initial choice at time $t-1$: (i) new information revealed at time t —the information update $\Delta\mathbb{E}_t$ —, and (ii) a change in the objectives of the policy maker —the preference shift $\Delta\mathcal{D}^0$—. Dynamic inconsistency comes from the preference shift term: even if no new information is revealed between $t-1$ and t, an optimizing policy maker will re-adjust her policy decision, because the time t problem puts different weights on the path $y_{t-1}, y_t, y_{t+1}, \ldots$. This change in weights is captured by $\Delta\mathcal{D}^0$ with two specific changes. First, the time t problem ignores y_{t-1}. Second, the time t problem weighs all the policy objectives as if the problem was started over from time t and lets "bygones be bygones".

Using (32), we can then construct a *time consistent* OPP that eliminates dynamic inconsistency, which we define as

$$\boldsymbol{\delta}^{\mathcal{T}*}_t = \boldsymbol{\delta}^*_t - \Delta\mathcal{D}^0\mathbb{E}_{t-1}\mathbf{Y}^0_{t-1} \ , \tag{33}$$

where the original OPP is adjusted with a "time inconsistency correction factor" given by $\Delta\mathcal{D}^0\mathbb{E}_{t-1}\mathbf{Y}^0_{t-1}$. Intuitively, under a time consistent OPP, an optimizing policy maker will update its policy path between $t-1$ and t *only* if new information is revealed.

Importantly, this correction factor is again entirely determined by our two sufficient statistics, so that no extra information is necessary to implement a time-consistent OPP.[10] Equivalently, we can view this OPP as a special case of the constrained OPP where the constraint ensures that, in the absence of new information, the first order condition of the time t problem is equal to the first order condition of the time $t-1$ problem.[11]

5 Implementation of the OPP

In this section we show how to use our sufficient statistics approach to (i) evaluate policy decisions, and (ii) improve policy. First, we discuss how to estimate the two statistics $\mathbb{E}_t\mathbf{Y}^0_t$ and $\mathcal{R}^0_a$ using reduced form econometric methods, and we then discuss how to use these estimates (along with their corresponding distribution) to evaluate and improve a baseline policy path $\mathbb{E}_t\mathbf{P}^0_t$ using the subset OPP statistic. The web-appendix provides a detailed step-by-step implementation guide and examples.

[10] A subset version of the time consistent OPP can be easily constructed by replacing $\boldsymbol{\delta}^*_t$ by $\boldsymbol{\delta}^*_{a,t}$ — the subset OPP — and $\mathcal{R}^0_y$ by $\mathcal{R}^0_{y,a}$ — the subset of estimable impulse responses —.

[11] Formally, the constraint is given by $C(\mathbb{E}_t\mathbf{Y}_t, \mathbb{E}_t\mathbf{P}_t) = \mathcal{R}'\mathcal{W}\mathbb{E}_{t-1}\mathbb{E}_t\mathbf{Y}_t - \mathcal{R}'\mathcal{W}\mathbb{E}_{t-1}\mathbf{Y}^0_{t-1} = \mathbf{0}$. As a result the time consistent OPP inherits the properties of the constrained OPP (see the web-appendix section S4 for a formal statement).

The baseline policy. So far, we have kept the baseline policy choice (ϕ^0, ϵ_t^0) unspecified, and the OPP properties hold for any arbitrary initial choice (provided ϕ^0 implies a unique equilibrium; cf Assumption 2). To be able to estimate the sufficient statistics, we will require that the baseline rule ϕ^0 was implemented over some sampling period.[12] Within that restriction, different baseline policies are possible. In this section, we focus on policy evaluation on behalf of the policy maker, or the policy maker's staff, so that we treat the policy path $\mathbb{E}_t \mathbf{P}_t^0$ corresponding to (ϕ^0, ϵ_t^0) as chosen by the policy maker and thus known. In the web appendix, we consider alternative baseline policies.

Forecasts. The first task is to construct an estimate $\widehat{\mathbf{Y}}_t^0$ for $\mathbb{E}_t \mathbf{Y}_t^0$, that is the oracle prediction for $\mathbf{Y}_t^0$ conditional on the baseline policy choice $(\phi^0, \boldsymbol{\epsilon}_t^0)$ and the state $\mathbf{S}_t$ capturing the initial conditions $\mathbf{X}_{-t}$ and the macro shocks $\boldsymbol{\Xi}_t$.

We denote by $\mathbf{Z}_t$ a (possibly large) set of variables that can be used to approximate the state $\mathbf{S}_t$. Since we are only interested in prediction (and not in recovering causal effects), we consider the reduced form model

$$\mathbf{Y}_t^0 = \mathbf{B}_{yz}^0 \mathbf{Z}_t + \mathbf{B}_{yp}^0 \mathbb{E}_t \mathbf{P}_t^0 + \mathbf{U}_t^y \ , \tag{34}$$

where the maps $\mathbf{B}_{yz}^0$ and $\mathbf{B}_{yp}^0$ contain the best linear prediction coefficients under the rule ϕ_0.[13] Based on (34) we can estimate the best linear prediction coefficients $\mathbf{B}_{yz}^0$ and $\mathbf{B}_{yp}^0$ using the data sample over which the policy rule ϕ^0 was implemented, and we can then construct the forecast $\widehat{\mathbf{Y}}_t^0 = \widehat{\mathbf{B}}_{yz}^0 \mathbf{Z}_t + \widehat{\mathbf{B}}_{yp}^0 \mathbb{E}_t \mathbf{P}_t^0$ using these estimated coefficients, the path $\mathbb{E}_t \mathbf{P}_t^0$, and the observable predictors $\mathbf{Z}_t$.

Within this general scheme different specific approaches are possible, and we can use methods from the forecasting literature (e.g. Elliot and Timmermann, 2016). For instance, we can impose restrictions on the model coefficients in order to improve forecasting performance, e.g. we can impose a VAR structure (e.g. Banbura, Giannone and Reichlin, 2010), a DSGE structure (e.g. Negro and Schorfheide, 2013), or a dynamic factor model structure (e.g., Stock and Watson, 2002). In addition, we can use shrinkage methods for the estimation of the coefficients (e.g., Stock and Watson, 2012) and we can incorporate judgment (e.g., Lawrence et al., 2006; Manganelli, 2009). More generally, we can combine different models using averaging methods (e.g. Cheng and Hansen, 2015). We provide specific examples in the web-appendix.

[12] For an ex-post policy evaluation, the sample could run throughout the evaluation period. For real time policy improvement, the sample would have to be an in-sample period (before time t) over which the policy rule ϕ^0 was implemented.

[13] These coefficients do not have a causal interpretation, and model (34) cannot be used directly for policy prescriptions as the maps $\mathbf{B}_{yz}^0$ and $\mathbf{B}_{yp}^0$ are only valid for policies under the baseline rule ϕ^0. This is the reason we need a second sufficient statistics; the causal effects of policy shocks.

As a practical matter, policy makers often make their forecasts for $\mathbf{Y}_t^0$ publicly available, and these can then be directly used to compute the OPP and evaluate the expected policy path $\mathbb{E}_t\mathbf{P}_t^0$. For instance, in our empirical work below we will use the FOMC forecasts for inflation and unemployment in order to evaluate monetary policy decisions.

Impulse responses. To estimate impulse responses we can rely on a large macro-econometric literature that estimates impulse responses to policy shocks. The review of Ramey (2016) provides a wealth of identification approaches for recovering the impulse responses and all such methods can be adopted in our setting using either local projections or structural VARs. To give a few examples, one can use zero-, long-run, or inequality restrictions (e.g. Sims, 1980; Blanchard and Quah, 1989; Faust, 1998; Uhlig, 2005), or instrumental variables (e.g. Mertens and Ravn, 2013; Stock and Watson, 2018). A key requirement is that the sample of observations used needs to pertain to the policy regime ϕ^0. We denote the estimated subset of impulse responses, corresponding to $\mathcal{R}_a^0 = (\mathcal{R}_{y,a}^0, \mathcal{R}_{p,a}^0)$, by $\widehat{\mathcal{R}}_a^0$.

Uncertainty assessment. To approximate the distribution of the OPP statistic we need to approximate the distribution of the impulse response estimates $\mathcal{U}_t^{\mathcal{R}_a} = \mathcal{R}_a^0 - \widehat{\mathcal{R}}_a^0$ and the distribution of model uncertainty $\mathcal{U}_t^y = \mathbb{E}_t\mathbf{Y}_t^0 - \widehat{\mathbf{Y}}_t^0$. Note that model uncertainty arises from not being able to perfectly approximate the state of the economy and from parameter estimation uncertainty. In general we denote the approximated distribution by $\widehat{F}$ and we can distinguish between two scenarios for computing $\widehat{F}$.

If a single reduced form model is used to construct both the impulse responses and the forecasts, conventional asymptotic theory or Bayesian methods can be used to obtain an estimate of the joint distribution of $(\mathcal{U}_t^{\mathcal{R}_a}, \mathcal{U}_t^y)$. If model mis-specification is a concern, historical forecast errors can be used to approximate the distribution of $\mathcal{U}_t^y$, see Stock and Watson (2019, Section 15.5).

If forecasts and impulse responses are instead obtained from different models or sources (for instance, if the forecast is the policy maker's published forecast), we will typically have to make the additional assumption that $\mathcal{U}_t^{\mathcal{R}_a}$ is independent of $\mathcal{U}_t^y$.

Subset OPP. After obtaining the approximating distribution $\widehat{F}$, we can compute the distribution of the subset OPP using simulation methods for a given preference map $\mathcal{W}$.

A first application of the subset OPP is for historical policy evaluation: identifying instances when the policy decision deviated from the optimal decision, and if so by how much. Using Corollary 1, we will conclude that a policy $\mathbb{E}_t\mathbf{P}_t^0$ is not optimal whenever the confidence bands of $\boldsymbol{\delta}_{a,t}^*$ exclude zero at any desired level of confidence. The magnitude of the deviation of optimality is given by $\boldsymbol{\delta}_{a,t}^*$.

Another application of the subset OPP is to compute a best policy given the sufficient statistics available, that is to compute the best improvement over the baseline policy. With the distribution of the subset OPP in hand, we can use simulation to compute the distributions of the best policy path $(\mathbb{E}_t\mathbf{P}_t^0 + \mathcal{R}_{a,p}^0\boldsymbol{\delta}_{a,t})$ and the associated paths of policy objectives $(\mathbb{E}_t\mathbf{Y}_t^0 + \mathcal{R}_{a,y}^0\boldsymbol{\delta}_{a,t})$.

6 Illustration: US monetary policy

In this section we illustrate how the OPP statistic can be used in practice to evaluate and improve monetary policy decisions. As a loss function we posit the dual inflation-full employment mandate imposed by the US Congress on the Fed, so that we take as policy objectives

$$y_t = (\pi_t - \pi_t^*, u_t - u_t^*)'$$

with π_t^* and u_t^* the long-run values for inflation and unemployment. The path of objectives is thus $\mathbf{Y}_t = (y_t', y_{t+1}', \ldots, y_{t+H}')'$, where we truncate the paths at a horizon of $H = 5$ years. For the weighting matrix $\mathcal{W}$, we set the discount rate $\beta_h = 1$ for all h and set the preference parameter $\lambda = 1$, consistent with the Fed's balanced approach to its dual mandate (Bernanke, 2015).[14]

Our evaluation period is from 1990 until 2022, and we assume that the Fed's reaction function was stable over this sampling period.

6.1 Recovering the sufficient statistics

Evaluating policy decisions made over 1990-2022 requires (i) conditional forecasts at each decision point over that period ($\mathbb{E}_t\mathbf{Y}_t^0$ and $\mathbb{E}_t\mathbf{P}_t^0$), and (ii) estimates for the impulse responses to monetary shocks over that period ($\mathcal{R}_y^0$ and $\mathcal{R}_p^0$).

Conditional forecasts

As primary source, we exploit the Survey of Economic Projections (SEP), which allows us to get estimates for $\mathbb{E}_t\mathbf{Y}_t^0$ and $\mathbb{E}_t\mathbf{P}_t^0$ over 2007-2022. The SEP is being conducted four times a year (ahead of an FOMC meeting) since October 2007, and it asks FOMC members to report their forecasts for unemployment, inflation (headline and core) and real GDP growth

[14] In the web-appendix, we show results for a range of λ over $[0.2, 2]$. If the user of the sufficient statistic approach is the policy maker (or her staff), we can treat λ (or β) as a preference parameter, i.e., a choice for the policy maker. For a retrospective or external analysis of policy decisions, presenting results for a range of values for λ is useful to understand which alternative values for λ *could* explain some decisions. In the web-appendix, we also propose a conservative (i.e., robust) approach to elicit λ. Specifically, the procedure consists in picking the λ that is least favorable to rejecting that a policy was optimal. In addition, we note that more elaborate loss functions, for instance including a financial stability objective or a motive for smooth policy changes (e.g., Rudebusch, 2001), could be interesting to explore.

over the next 3 or 4 years, depending on the survey date. While the Fed only releases FOMC members' individual forecasts with a five year delay, the median, "central tendency",[15] and the range for members' forecasts are reported after each SEP round. Since April 2009, the SEP also asks for estimates for the long-run values of these variables, and since September 2015 the SEP also asks for FOMC members' forecast for the fed funds rate.

We take the median FOMC forecasts as our estimates $\widehat{\mathbf{Y}}_t^0$ for $\mathbb{E}_t^0\mathbf{Y}_t$. We posit that model uncertainty —the distribution of $\mathcal{U}_t^y = \mathbb{E}_t\mathbf{Y}_t^0 - \widehat{\mathbf{Y}}_t^0$— is normally distributed, and we find for each SEP round the normal distribution that best fit the reported central tendency and range for the SEP forecasts. We take the median FOMC forecast as our estimate for $\mathbb{E}_t\mathbf{P}_t^0$. Since SEP forecasts only go four years out at most, we complement them with estimates for long-run inflation and unemployment, where we posit that these long-run values are reached linearly after 5 years. We use the SEP estimates after 2009, and the Greenbook estimates before 2009. The inflation and unemployment gaps are then defined as deviations of inflation and unemployment from these long run estimates. Finally, since SEP projections are annual, we linearly interpolate them to quarterly frequency in order to combine them with the estimated impulse responses that we detail in the next section.

To obtain forecasts before 2007, we exploit the Monetary Policy Report (MPR) —the predecessor of the SEP—, which is submitted semi-annually to US Congress and includes median, central tendency and range of FOMC members' forecasts. The MPR is more limited in scope than the SEP however, as the forecasts for inflation and unemployment only extend two years out. To complement this information, we use the Fed real time Greenbook estimates for long-run inflation and unemployment, and we posit that these long-run values are reached at a linear convergence rate after 5 years. Clearly, the assumptions needed to evaluate pre-2007 decisions from FOMC forecasts are stronger.

Impulse responses to policy shocks

As monetary shocks, we follow the recent literature (e.g. Eberly, Stock and Wright, 2020) to identify two monetary shocks: (i) shocks to the contemporaneous fed funds rate and (ii) shocks to the slope of the yield curve, which the Fed can affect through forward-guidance or asset purchases (QE). These will allow us to compute a subset OPP vector based on two sets of impulse responses, each corresponding to a specific policy experiment: (i) a *short-rate OPP* assessing the short-end of the fed funds rate path and (ii) a *slope OPP* assessing forward-guidance or asset purchases (QE).[16]

[15] The central tendencies exclude the three highest and three lowest projections for each variable in each year.

[16] In this paper, we proceed as in Eberly, Stock and Wright (2020), and we do not distinguish separately forward-guidance and QE, as there is yet no accepted identification strategy to separate the two instruments. This is an important area for future research as a key lesson from this paper is that further progress on this

To estimate the impulse responses, we follow Kuttner (2001) and Eberly, Stock and Wright (2020), and we use as instrumental variables the monetary policy surprises measured around the FOMC announcements within a 30 minute window. First, we use surprises to the fed funds rate —the difference between the expected fed funds rate (as implied by current-month federal funds futures contracts) and the actual fed funds rate— to identify the effects of a shock to the contemporaneous fed funds rate. Second, we use surprises to the ten-year on-the-run Treasury yield (orthogonalized with respect to surprises to the current fed funds rate) to identify the effects of shocks to the slope of the yield curve.

To estimate the impulse responses to these shocks, we use a Bayesian VAR with inflation, unemployment, the fed funds rate, the 10-year bond-fed funds rate spread and the two monetary policy surprises, which are ordered first. We estimate the reduced form VAR coefficients using Bayesian methods following the default set-up with a Minnesota prior discussed in Canova (2007, Chapter 10). We compute the subset structural impulse responses $\mathcal{R}^0_{y,a}$ and $\mathcal{R}^0_{p,a}$ by (i) computing the impulse responses of all variables to the shocks corresponding to the monetary surprises and (b) rescaling the responses of inflation and unemployment by the contemporaneous responses of the fed funds rate (for the short rate impulse responses) and the spread (for the slope impulse responses). This approach implements SVAR-IV in a convenient way which does not require invertibility of the monetary shocks (e.g. Plagborg-Møller and Wolf, 2021, Corollary 1). The sampling period is 1990-2018.[17]

The subset OPP statistics

Based on our sufficient statistics estimates, we compute the median subset OPP statistics and construct confidence bands as described in Section 5. The subset OPP is a vector with two elements: (i) the short-rate OPP, and (ii) the slope OPP. As constraint on the OPP, we impose a zero lower bound on the fed funds rate following the approach of Section 4.3.

In terms of confidence interval for the OPP, we will report the 90%, 75% and 60% confidence bands. Importantly, we note that the objective of a policy optimality test is different from the traditional objective of hypothesis testing. Specifically, from the perspective of the policy maker it is not clear that high significance is the most interesting/appropriate criteria. Consider the main outcome of the OPP test: "With X% confidence, the baseline policy choice is not optimal". A policy maker particularly averse to making a non-optimal decision may want to change the baseline policy choice at a relatively low X level, say 60% instead of 90%, as she may want to discard a policy that is non-optimal with a 60% probability. A trade-off however is that too low a threshold may lead a policy maker to change policy course

front is of direct relevance for the conduct of monetary policy.

[17] As in Eberly, Stock and Wright (2020), the effect of shocks to the slope of the yield curve is estimated over 2006-2018, the period during which the Fed was actively trying to affect the slope of the yield curve.

too often. While such a decision problem is outside the scope of this paper, it highlights that for the OPP test the classical dichotomy of hypothesis testing (i.e. preference for type 1 vs type 2 errors) really depends on the preference of the policy maker, e.g., making non-optimal decisions vs. changing course frequently.

6.2 A retrospective analysis of US monetary policy

Figure 1 displays the time series for the two elements of the subset OPP —the short-rate OPP and the slope OPP— along with their confidence intervals, as implied by both impulse response estimation uncertainty and model uncertainty —uncertainty around the estimate of $\mathbb{E}_t \mathbf{Y}_t^0$—. Recall that model uncertainty captures uncertainty about the state of the economy and from uncertainty about how this state will affect the economy going forward.

Note first how the uncertainty around the OPP can vary over time. This is due to time-varying uncertainty about the economic outlook. If there is more uncertainty about initial conditions or about the path of the economy going forward —i.e., more model uncertainty—, it is more difficult to establish whether a policy is appropriate or not. In the early stage of the COVID crisis for instance we knew little about the pandemic or about its economic consequences, and uncertainty in the SEP forecasts rose dramatically leading to larger OPP uncertainty.

As a general pattern, note how the two OPPs exhibit some signs of pro-cyclicality, being positive in the late 1990s and 2010s expansions and negative in the early 1990s, early 2000s and (for the slope OPP) the 2007-2008 recessions.[18] We now comment on some of these most striking misses.

Short-rate OPP We start with the short-rate OPP, which evaluates the optimality of the contemporaneous fed funds rate, the traditional tool of monetary policy. While the contemporaneous fed funds rate has not been set exactly at its optimal level since 1990, the optimal adjustment (in absolute value) is overall relatively small averaging only 25 basis points over the full sample. There is however a few interesting cases of non-optimal policy decisions.

In the late 1990s, the short-rate OPP indicates that the fed funds rate was too low by about 0.25ppt. This finding echoes earlier arguments that the Fed may have found itself falling behind the curve in the late 1990s tightening cycle (e.g., Blinder and Reis, 2005).

Another case of suboptimal fed funds rate was on the eve of the Great Recession (when

[18]While this cyclical pattern could be explained with a policy maker always running slightly behind the curve, it could also be a sign of a policy maker intentionally implementing gradual changes in its policy rate, at least during tightening cycles (e.g., Rudebusch, 2001). Incorporating such a smoothing motive with the constrained OPP would be an interesting avenue for future research.

the ZLB was not yet binding). Then, the short-rate OPP indicates that, given the paths for inflation and unemployment expected at the time, the FOMC should have lowered rates faster in early 2008. Afterward, the short-rate OPP was constrained by the zero lower bound.

Over 2016-2019, the OPP is consistently positive, indicating that the fed funds rate is consistently too low. Since inflation was close to target over that period, these misses come from a persistently negative unemployment gap: given a too low unemployment rate, the FOMC should have raised the fed funds rate further.

Most recently, the short-rate OPP shows that the FOMC was too slow in raising the fed funds rate in the face of mounting inflationary pressures during 2021. We will come back to this important point in the next section.

Slope OPP The slope OPP assesses the optimality of QE or forward-guidance. During the Great Recession for instance, the slope OPP indicates that unconventional monetary policy could have been used more aggressively, a conclusion echoing that of Eberly, Stock and Wright (2020). In 2009, the slope OPP drops rapidly to below -1ppt and only slowly revert back to zero. In fact, the slope OPP remains significantly different from zero at high levels of significance over the whole 2009-2013 period. Overall, these results indicate that a more active use of QE or forward-guidance holds considerable promise for improvements in the conduct of policy.

6.3 Improving monetary policy decisions

In this last section, we illustrate how the sufficient statistics approach can be implemented to provide "actionable" policy recommendations. We consider three instructive case studies: (i) April 2008 at the onset of the Great Recession, (ii) April 2010 in the midst of the Great Recession, and (iii) 2021 when the FOMC was confronted with a surge in inflation. While the first two case studies are only pseudo real time exercises (the impulse responses being estimated over 1990-2018), the last exercise is entirely real time.

Fed funds rate policy as of April 2008 April 2008 marks the early stage of the financial crisis; Lehman Brothers was still 6 months away from failing, unemployment was only at 5 percent, and few anticipated the magnitude of the recession that was going to ensue. In fact, the fed funds rate was still at 2.25ppt so the Fed still had room to use conventional policies to stimulate activity.[19] At that meeting, the fed funds rate was lowered by .25ppt to

[19] By the end of 2008 however, unemployment had reached 7.3 percent, and the Fed had dropped the fed funds rate by almost 2ppt (to the zero lower bound) in the span of only three months (September-December) following the failure of Lehman Brothers in September 2008.

2 percent, but it remained at that level until October 2008, i.e., until the collapse of Lehman brothers.

As is clear from the April Tealbook and forecast narratives reported by the FOMC, the central bank was facing two conflicting issues in April 2008: (i) a marked deterioration in the growth outlook due declining housing prices and tensions in the financial market, and (ii) upside risks to inflation coming from "persistent surprises to energy and commodity prices" (Kohn, 2008).

Figure 2 depicts all the information needed to construct the short-rate OPP. The top row (filled dots) reports the median FOMC forecasts for the inflation and unemployment gaps, and the bottom row reports the estimated effect of a 1ppt shock to the contemporaneous fed funds rate.

The two issues of the time —poor economic outlook and inflationary pressures from high energy prices— are visible in the FOMC forecasts in the first row of Figure 2. The median short-rate OPP comes out at -0.30, calling for an additional 25 basis points cuts.[20] The effects of this policy adjustment is depicted by the unfilled dots in the top-panel. We can see that the FOMC could have brought down expected unemployment faster at the cost of a small and delayed increase in inflation. In fact, the effect of monetary policy on inflation is so delayed that the extra inflation would only materialize two years later, i.e., after the commodities-driven burst in inflation has died down.

Slope (QE) policy as of April 2010 It is interesting to contrast the 2008-M4 situation with that of two years later; in 2010-M4. There, the Fed funds rate was stuck at zero but the Fed could have further used unconventional monetary policy to better stabilize the economy. To test this possibility, we can use the slope OPP statistic.

Figure 3 displays the situation in 2010-M4 where the bottom panels show the effects on inflation and unemployment of a 1ppt shock to the slope of the yield curve. The median slope OPP comes at -0.90, calling for for an almost 1ppt decline in the slope of the yield curve. The unfilled dots plot the counter-factual expected paths for the policy objectives after adjusting policy with the slope OPP experiment. The FOMC could have brought down expected unemployment faster in exchange for a small overshoot in expected inflation in 2011.

Fed funds rate liftoff in 2021 As last case study, we consider fed funds rate decisions during 2021, when the Fed faced an unexpectedly strong surge in inflation. This case study serves to illustrate how our sufficient statistics approach can incorporate pre-commitments.

[20] The 75% confidence interval excludes zero, indicating that there is a less than 75 percent chance that the contemporaneous fed funds rate was set optimally.

Indeed, in September 2020 the FOMC statement stipulated that the Committee would not raise the fed funds rate until "`until labor market conditions have reached levels consistent with the Committee's assessment of maximum employment and inflation has risen to 2 percent and is on track to moderately exceed 2 percent for some time.`"

To incorporate this pre-commitment we will compute a constrained OPP, where we impose the constraint that the policy rate remains at the zero lower bound until inflation is expected to lie at least 0.5 ppt above target for 1 year and unemployment is less than 0.5 ppt above its long-run level (as estimated by the FOMC).[21]

To contrast the predictions of the constrained and unconstrained OPP, Figure 4 zooms in on Figure 1 by plotting the short-rate OPP over 2019-2022, showing both the unconstrained OPP and the constrained OPP. Notice how the unconstrained OPP calls for liftoff as early as March 2021, with strong evidence against optimality in March 2021 (with a more than 90 percent chance that the fed funds rate is too low), while the constrained OPP does not call for liftoff until September 2021.

Figures 5 and 6 study these these two decision points (March 2021 and September 2021) in more detail. The top row reports the median SEP forecasts before and after OPP adjustments (similarly to Figure 2), but the bottom panel now reports the expected policy path before and after a short-rate OPP adjustment. This panel is shown to illustrate how the subset OPP can be used in practice to provide policy paths recommendations.

Consider first the case of March 2021 (Figure 5). Inflation was expected to substantially exceed its target for almost two years. Thus, the unconstrained OPP calls for an immediate liftoff, with strong evidence against optimality in March 2021: there is a more than 90 percent chance that the fed funds rate is too low. This conclusion however ignores the September 2020 pre-commitment. For that purpose, we turn to the constrained OPP, and this time we cannot reject that the policy is optimal: the constrained OPP is zero. The reason is that, as of March 2021 unemployment was still substantially above target, so that the Fed's decision to not liftoff is optimal given its Sept. 2020 pre-commitment. Thus, while monetary policy could appear to get behind the curve in March 2021 —not reacting enough the the inflation surge—, the FOMC decision is in fact fully consistent with its Sept. 2020 commitment to delay liftoff until the labor market has recovered.

In contrast, about six months later in November 2021 (Figure 6) the labor market has almost fully recovered while inflation has stayed above target (in fact exceeding the levels expected back in March 2021) and is expected to remain above target for at least a year. Thus, the Sept. 2020 pre-commitment is no longer binding, and the OPP (constrained or

[21]These thresholds are only chosen as means of illustration (the FOMC remained vague in terms of the conditions defining the conditions for liftoff), though the conclusions below are robust to alternative reasonable thresholds.

unconstrained) calls for an immediate rise in the Fed funds rate and a steeper policy path thereafter (green or red line, bottom panel).

7 Conclusion

In this paper, we show that it is not necessary to know the full structure of the economy to evaluate or even decide on macroeconomic policy. We consider a policy maker facing a time t decision problem —how to set the policy path today given the state of the economy—, and we show that two statistics are sufficient to detect and correct non-optimal decisions: (i) the forecasts for the policy objectives conditional on the policy choice, and (ii) the effects of policy shocks on the policy objectives. These two statistics are already central and well understood concepts for policy makers (e.g., Orphanides, 2019). Our contribution is to show that these two statistics alone can be used to rigorously evaluate and even set policy.

The monetary policy setting considered in this paper is only one of many potential applications of a sufficient statistics approach to evaluating macro policy problems. Other fruitful uses include the many areas where macro policy makers must balance difficult trade-offs in complex settings: fiscal policy (e.g., balancing growth considerations with risks to debt sustainability), exchange rate management (balancing monetary independence with exchange rate stability), foreign-reserve management (e.g., balancing the cost of holding reserves with the insurance against sudden stops in capital flows), or even climate change policy (e.g., balancing the costs of climate change with the costs of preventive actions), among others.

References

Auclert, Adrien, Bence Bardóczy, Matthew Rognlie, and Ludwig Straub. 2021. "Using the Sequence-Space Jacobian to Solve and Estimate Heterogeneous-Agent Models." *Econometrica*, 89(5): 2375–2408.

Banbura, Marta, Domenico Giannone, and Lucrezia Reichlin. 2010. "Large Bayesian Vector Auto Regressions." *Journal of Applied Econometrics*, 25(1): 71–92.

Bernanke, Ben S. 2015. "The Taylor Rule: A Benchmark for Monetary Policy?" *Ben Bernanke's Blog*, 28.

Blanchard, Olivier Jean, and Danny Quah. 1989. "The Dynamic Effects of Aggregate Demand and Supply Disturbances." *American Economic Review*, 79(4): 655–673.

Blinder, Alan S. 2020. "What does Jerome Powell know that William McChesney Martin Didn't and What Role Did Academic Research Play in That?" *The Manchester School.*

Blinder, Alan S., and Ricardo Reis. 2005. *Understanding the Greenspan standard.*

Canova, Fabio. 2007. *Methods for Applied Macroeconomic Research.* Princeton University Press.

Chari, Varadarajan V, Lawrence J Christiano, and Patrick J Kehoe. 1994. "Optimal Fiscal Policy in a Business Cycle Model." *Journal of polítical Economy*, 102(4): 617–652.

Cheng, Xu, and Bruce E. Hansen. 2015. "Forecasting with Factor-Augmented Regression: A Frequentist Model Averaging Approach." *Journal of Econometrics*, 186(2): 280–293.

Chetty, Raj. 2009. "Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods." *Annual Review of Economics*, 1(1): 451–488.

Eberly, Janice C., James H. Stock, and Jonathan H. Wright. 2020. "The Federal Reserve's Current Framework for Monetary Policy: A Review and Assessment." *International Journal of Central Banking*, 16(1): 5–71.

Elliot, Graham, and Allan Timmermann. 2016. *Economic Forecasting.* Princeton University Press.

Faust, Jon. 1998. "The Robustness of Identified VAR Conclusions About Money." *Carnegie-Rochester Conference Series on Public Policy*, 49: 207 – 244.

Galí, Jordi. 2015. *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework and its Applications.* Princeton University Press.

Kohn, Donald. 2008. "Survey of Economic Projections." Report to the FOMC.

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.

Kydland, Finn E., and Edward C. Prescott. 1977. "Rules Rather than Discretion: The Inconsistency of Optimal Plans." *Journal of Political Economy*, 85(3): 473–491.

Lawrence, Michael, Paul Goodwin, Marcus O'Connor, and Dilek Önkal. 2006. "Judgmental forecasting: A Review of Progress Over the Last 25 Years." *International Journal of forecasting*, 22(3): 493–518.

Lucas, Robert Jr. 1976. "Econometric Policy Evaluation: A critique." *Carnegie-Rochester Conference Series on Public Policy*, 1(1): 19–46.

Manganelli, Simone. 2009. "Forecasting with Judgment." *Journal of Business & Economic Statistics*, 27(4): 553–563.

McKay, Alisdair, and Christian Wolf. 2022. "What Can Time-Series Regressions Tell Us About Policy Counterfactuals?" Working Paper.

Mertens, Karel, and Morten O. Ravn. 2013. "The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States." *American Economic Review*, 103(4): 1212–1247.

Michaillat, Pascal, and Emmanuel Saez. 2019. "Optimal Public Expenditure with Inefficient Unemployment." *The Review of Economic Studies*, 86(3): 1301–1331.

Mishkin, Frederic S. 2010. "Will Monetary Policy Become More of a Science?" In *The Science and Practice of Monetary Policy Today.* , ed. Wieland V., 81–103. Springer, Berlin.

Negro, Marco Del, and Frank Schorfheide. 2013. "Chapter 2 - DSGE Model-Based Forecasting." In *Handbook of Economic Forecasting.* Vol. 2 of *Handbook of Economic Forecasting*, , ed. Graham Elliott and Allan Timmermann, 57–140. Elsevier.

Orphanides, Athanasios. 2019. "Monetary Policy Strategy and its Communication." Working paper, presented at Jackson Hole Conference August 22, 2019.

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

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

Rudebusch, Glenn D. 2001. "Is the Fed too timid? Monetary policy in an uncertain world." *Review of Economics and Statistics*, 83(2): 203–217.

Sims, Christopher A. 1980. "Macroeconomics and reality." *Econometrica*, 1–48.

Smets, Frank, and Rafael Wouters. 2007. "Shocks and frictions in US business cycles: A Bayesian DSGE approach." *American Economic Review*, 97(3): 586–606.

Stock, James H, and Mark W Watson. 2002. "Macroeconomic Forecasting Using Diffusion Indexes." *Journal of Business & Economic Statistics*, 20(2): 147–162.

Stock, James H, and Mark W Watson. 2012. "Generalized shrinkage methods for forecasting using many predictors." *Journal of Business & Economic Statistics*, 30(4): 481–493.

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.

Stock, James H, and Mark W Watson. 2019. *Introduction to Econometrics, Fourth Edition.* Pearson, New York.

Svensson, Lars E.O. 1999. "Inflation Targeting as a Monetary Policy Rule." *Journal of Monetary Economics*, 43(3): 607 – 654.

Svensson, Lars E.O. 2003. "What Is Wrong with Taylor Rules? Using Judgment in Monetary Policy through Targeting Rules." *Journal of Economic Literature*, 41(2): 426–477.

Svensson, Lars E.O., and Michael Woodford. 2005. "Implementing Optimal Policy through Inflation-Forecast Targeting." In *The Inflation-Targeting Debate. NBER Chapters*, 19–92. National Bureau of Economic Research, Inc.

Taylor, John B. 1993. "Discretion versus Policy Rules in Practice." Vol. 39, 195–214, Elsevier.

Uhlig, Harald. 2005. "What are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure." *Journal of Monetary Economics*, 52(2): 381–419.

Woodford, Michael. 2010. "Optimal Monetary Stabilization Policy." *Handbook of monetary economics*, 3: 723–828.

Woodford, Michael. 2013. "Forecast Targeting as a Monetary Policy Strategy: Policy Rules in Practice." *The Taylor Rule and the Transformation of Monetary Policy*, 185.

Appendix

Proof of Lemma 1. Note that Assumption 2 imposes that the rule ϕ^0 underlying the baseline policy (18) leads to a unique equilibrium. Hence, under this rule we can write the solution as

$$\begin{aligned}
\begin{bmatrix} \mathbb{E}_t \mathbf{Y}_t^0 \\ \mathbb{E}_t \mathbf{W}_t^0 \\ \mathbb{E}_t \mathbf{P}_t^0 \end{bmatrix} &= \begin{bmatrix} \mathcal{A}_{yy} & -\mathcal{A}_{yw} & -\mathcal{A}_{yp} \\ -\mathcal{A}_{wy} & \mathcal{A}_{ww} & -\mathcal{A}_{wp} \\ -\mathcal{A}_{py}^0 & -\mathcal{A}_{pw}^0 & \mathcal{A}_{pp}^0 \end{bmatrix}^{-1} \begin{bmatrix} \mathcal{B}_{yx} & \mathcal{B}_{y\xi} & \mathbf{0} \\ \mathcal{B}_{wx} & \mathcal{B}_{w\xi} & \mathbf{0} \\ \mathcal{B}_{px}^0 & \mathcal{B}_{p\xi}^0 & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{X}_{-t} \\ \mathbf{\Xi}_t \\ \boldsymbol{\epsilon}_t^0 \end{bmatrix} \\
&= \begin{bmatrix} \mathcal{C}_{yx}^0 & \mathcal{C}_{y\xi}^0 & \mathcal{R}_y^0 \\ \mathcal{C}_{wx}^0 & \mathcal{C}_{w\xi}^0 & \mathcal{R}_w^0 \\ \mathcal{C}_{px}^0 & \mathcal{C}_{p\xi}^0 & \mathcal{R}_p^0 \end{bmatrix} \begin{bmatrix} \mathbf{X}_{-t} \\ \mathbf{\Xi}_t \\ \boldsymbol{\epsilon}_t^0 \end{bmatrix} ,
\end{aligned} \tag{35}$$

and we define $\Gamma_y^0 = (\mathcal{C}_{yx}^0, \mathcal{C}_{y\xi}^0)$ and $\Gamma_p^0 = (\mathcal{C}_{px}^0, \mathcal{C}_{p\xi}^0)$. The expressions for the maps, e.g. $\Gamma_y^0, \Gamma_p^0, \mathcal{R}_y^0$ and $\mathcal{R}_p^0$, can be derived explicitly from inverting the map, but we will not require such expressions: existence as imposed by Assumption 2 is sufficient for our purposes. Finally, $\mathbb{E}(\boldsymbol{\epsilon}_t^0 \mathbf{S}_t') = 0$ follows as we assume that the policy news shocks are uncorrelated with the initial conditions $\mathbf{X}_{-t}$ and the non-policy news shocks $\mathbf{\Xi}_t$. □

Proof of Lemma 2. Note that Assumption 2 imposes that the rule ϕ^0 underlying the augmented policy (20) leads to a unique equilibrium. Hence, under this rule we can write the solution as

$$\begin{aligned}
\begin{bmatrix} \mathbb{E}_t \mathbf{Y}_t(\boldsymbol{\delta}_t) \\ \mathbb{E}_t \mathbf{W}_t(\boldsymbol{\delta}_t) \\ \mathbb{E}_t \mathbf{P}_t(\boldsymbol{\delta}_t) \end{bmatrix} &= \begin{bmatrix} \mathcal{A}_{yy} & -\mathcal{A}_{yw} & -\mathcal{A}_{yp} \\ -\mathcal{A}_{wy} & \mathcal{A}_{ww} & -\mathcal{A}_{wp} \\ -\mathcal{A}_{py}^0 & -\mathcal{A}_{pw}^0 & \mathcal{A}_{pp}^0 \end{bmatrix}^{-1} \begin{bmatrix} \mathcal{B}_{yx} & \mathcal{B}_{y\xi} & \mathbf{0} \\ \mathcal{B}_{wx} & \mathcal{B}_{w\xi} & \mathbf{0} \\ \mathcal{B}_{px}^0 & \mathcal{B}_{p\xi}^0 & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{X}_{-t} \\ \mathbf{\Xi}_t \\ \boldsymbol{\epsilon}_t^0 \end{bmatrix} \\
&\quad + \begin{bmatrix} \mathcal{A}_{yy} & -\mathcal{A}_{yw} & -\mathcal{A}_{yp} \\ -\mathcal{A}_{wy} & \mathcal{A}_{ww} & -\mathcal{A}_{wp} \\ -\mathcal{A}_{py}^0 & -\mathcal{A}_{pw}^0 & \mathcal{A}_{pp}^0 \end{bmatrix}^{-1} \begin{bmatrix} \mathcal{B}_{yx} & \mathcal{B}_{y\xi} & \mathbf{0} \\ \mathcal{B}_{wx} & \mathcal{B}_{w\xi} & \mathbf{0} \\ \mathcal{B}_{px}^0 & \mathcal{B}_{p\xi}^0 & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \boldsymbol{\delta}_t \end{bmatrix} \\
&= \begin{bmatrix} \mathcal{C}_{yx}^0 & \mathcal{C}_{y\xi}^0 & \mathcal{R}_y^0 \\ \mathcal{C}_{wx}^0 & \mathcal{C}_{w\xi}^0 & \mathcal{R}_w^0 \\ \mathcal{C}_{px}^0 & \mathcal{C}_{p\xi}^0 & \mathcal{R}_p^0 \end{bmatrix} \begin{bmatrix} \mathbf{X}_{-t} \\ \mathbf{\Xi}_t \\ \boldsymbol{\epsilon}_t^0 \end{bmatrix} + \begin{bmatrix} \mathcal{C}_{yx}^0 & \mathcal{C}_{y\xi}^0 & \mathcal{R}_y^0 \\ \mathcal{C}_{wx}^0 & \mathcal{C}_{w\xi}^0 & \mathcal{R}_w^0 \\ \mathcal{C}_{px}^0 & \mathcal{C}_{p\xi}^0 & \mathcal{R}_p^0 \end{bmatrix} \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \boldsymbol{\delta}_t \end{bmatrix} ,
\end{aligned} \tag{36}$$

and we again let $\Gamma_y^0 = (\mathcal{C}_{yx}^0, \mathcal{C}_{y\xi}^0)$ and $\Gamma_p^0 = (\mathcal{C}_{px}^0, \mathcal{C}_{p\xi}^0)$. Reading of the display (36) together with the definitions of $\mathbb{E}_t \mathbf{Y}_t^0$ and $\mathbb{E}_t \mathbf{P}_t^0$ in Lemma 1 gives the desired result. □

Proof of Proposition 1. We first characterize the optimal policy that is defined as the solution to the planners problem (17), that is

$$\min_{\mathbf{Y}_t, \mathbf{W}_t, \mathbf{P}_t} \mathcal{L}_t \qquad \text{s.t.} \qquad (16) \ . \tag{37}$$

The Lagrange function for this problem is given by

$$\begin{aligned}
\mathsf{L}_t =& \mathbb{E}_t \Big\{ \frac{1}{2} \mathbf{Y}_t' \mathcal{W} \mathbf{Y}_t + \boldsymbol{\mu}_1' (\mathcal{A}_{yy} \mathbf{Y}_t - A_{yw} \mathbf{W}_t - \mathcal{A}_{yp} \mathbf{P}_t - \mathcal{B}_{yx} \mathbf{X}_{-t} - \mathcal{B}_{y\xi} \mathbf{\Xi}_t) \\
&+ \boldsymbol{\mu}_2' (\mathcal{A}_{ww} \mathbf{W}_t - \mathcal{A}_{wy} \mathbf{Y}_t - \mathcal{A}_{wp} \mathbf{P}_t - \mathcal{B}_{wx} \mathbf{X}_{-t} - \mathcal{B}_{w\xi} \mathbf{\Xi}_t) \Big\} \ ,
\end{aligned}$$

where $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ denote the Lagrange multipliers. The first order conditions for $\mathbf{Y}_t, \mathbf{W}_t, \mathbf{P}_t$ are given by

$$\begin{aligned}
\mathbf{0} &= \mathcal{W}\mathbb{E}_t\mathbf{Y}_t + \mathcal{A}'_{yy}\boldsymbol{\mu}_1 - \mathcal{A}'_{wy}\boldsymbol{\mu}_2 \\
\mathbf{0} &= -\mathcal{A}'_{yw}\boldsymbol{\mu}_1 + \mathcal{A}'_{ww}\boldsymbol{\mu}_2 \\
\mathbf{0} &= -\mathcal{A}'_{yp}\boldsymbol{\mu}_1 - \mathcal{A}'_{wp}\boldsymbol{\mu}_2 \ ,
\end{aligned}$$

and from Assumption 1 it follows that this system of equations implies a unique solution $\mathbb{E}_t\mathbf{P}_t^{\text{opt}}$.

Next, we consider the fictitious policy problem of a policy maker considering deviating from the fixed rule $(\phi^0, \boldsymbol{\epsilon}_t^0)$ with some fixed sequence of perturbations $\boldsymbol{\delta}_t$.

$$\min_{\mathbf{Y}_t,\mathbf{W}_t,\mathbf{P}_t,\boldsymbol{\delta}_t} \mathcal{L}_t \qquad \text{s.t.} \qquad (16) \quad \text{and} \quad (18) \ . \tag{38}$$

The Lagrange function for this problem is given by

$$\begin{aligned}
\mathsf{L}_t^f = &\mathbb{E}_t\Big\{\frac{1}{2}\mathbf{Y}_t'\mathcal{W}\mathbf{Y}_t + \boldsymbol{\mu}_1'(\mathcal{A}_{yy}\mathbf{Y}_t - A_{yw}\mathbf{W}_t - \mathcal{A}_{yp}\mathbf{P}_t - \mathcal{B}_{yx}\mathbf{X}_{-t} - \mathcal{B}_{y\xi}\boldsymbol{\Xi}_t) \\
&+\boldsymbol{\mu}_2'(\mathcal{A}_{ww}\mathbf{W}_t - \mathcal{A}_{wy}\mathbf{Y}_t - \mathcal{A}_{wp}\mathbf{P}_t - \mathcal{B}_{wx}\mathbf{X}_{-t} - \mathcal{B}_{w\xi}\boldsymbol{\Xi}_t) \\
&+\boldsymbol{\mu}_3'(\mathcal{A}^0_{pp}\mathbf{P}_t - \mathcal{A}^0_{py}\mathbf{Y}_t - \mathcal{A}^0_{pw}\mathbf{W}_t - \mathcal{B}^0_{px}\mathbf{X}_{-t} - \mathcal{B}^0_{p\xi}\boldsymbol{\Xi}_t - \boldsymbol{\epsilon}_t^0 - \boldsymbol{\delta}_t)\Big\} \ ,
\end{aligned}$$

which leads to the first order conditions for $\mathbf{Y}_t, \mathbf{W}_t, \mathbf{P}_t, \boldsymbol{\delta}_t$ given by

$$\begin{aligned}
\mathbf{0} &= \mathcal{W}\mathbb{E}_t\mathbf{Y}_t + \mathcal{A}'_{yy}\boldsymbol{\mu}_1 - \mathcal{A}'_{wy}\boldsymbol{\mu}_2 - \mathcal{A}^{0'}_{py}\boldsymbol{\mu}_3 \\
\mathbf{0} &= -\mathcal{A}'_{yw}\boldsymbol{\mu}_1 + \mathcal{A}'_{ww}\boldsymbol{\mu}_2 - \mathcal{A}^{0'}_{pw}\boldsymbol{\mu}_3 \\
\mathbf{0} &= -\mathcal{A}'_{yp}\boldsymbol{\mu}_1 - \mathcal{A}'_{wp}\boldsymbol{\mu}_2 + \mathcal{A}^{0'}_{pp}\boldsymbol{\mu}_3 \\
\mathbf{0} &= \boldsymbol{\mu}_3 \ .
\end{aligned}$$

Since $\boldsymbol{\mu}_3 = \mathbf{0}$, it is easy to verify that the first order conditions of the fictitious policy problem (38) are identical to the first order conditions of the planner's policy problem (37). Hence they have the same solution $\mathbb{E}_t\mathbf{P}_t^{\text{opt}}$.

Next, using Lemma (2) we can rewrite the fictitious policy problem (by substituting out the variables $\mathbf{W}_t, \mathbf{P}_t$ in terms of $\boldsymbol{\delta}_t$ and the news shocks and initial conditions) as

$$\min_{\boldsymbol{\delta}_t} \mathcal{L}_t(\boldsymbol{\delta}_t) \qquad \text{s.t.} \qquad \mathbb{E}_t\mathbf{Y}_t = \mathbb{E}_t\mathbf{Y}_t^0 + \mathcal{R}_y\boldsymbol{\delta}_t \ . \tag{39}$$

This recalling that the $\mathbb{E}_t\mathbf{P}_t^0$ corresponds to the case where $\boldsymbol{\delta}_t = \mathbf{0}$ we find the necessary condition for $\mathbb{E}_t\mathbf{P}_t^0$ to be optimal as

$$\nabla_{\boldsymbol{\delta}_t} \ \mathcal{L}_t(\boldsymbol{\delta}_t)|_{\boldsymbol{\delta}_t=\mathbf{0}} = \mathcal{R}_y^{0'}\mathcal{W}\mathbb{E}_t\mathbf{Y}_t^0 = \mathbf{0} \ .$$

This shows that the optimality of $\mathbb{E}_t\mathbf{P}_t^0$ can be characterized by the condition $\mathcal{R}_y^{0'}\mathcal{W}\mathbb{E}_t\mathbf{Y}_t^0 = \mathbf{0}$. Since, $\mathbb{E}_t\mathbf{P}_t^{\text{opt}}$ is unique we have that $\mathbb{E}_t\mathbf{P}_t^{\text{opt}} = \mathbb{E}_t\mathbf{P}_t^0$ if and only if $\mathcal{R}_y^{0'}\mathcal{W}\mathbb{E}_t\mathbf{Y}_t^0 = \mathbf{0}$. □

Proof of Proposition 2. To prove part 1 we note that from Proposition 1 it follows directly that $\mathbb{E}_t \mathbf{P}_t^0 = \mathbb{E}_t \mathbf{P}_t^{\text{opt}}$ if and only if $\boldsymbol{\delta}_t^* = -(\mathcal{R}_y^{0\prime} \mathcal{W} \mathcal{R}_y^0)^{-1} \mathcal{R}_y^{0\prime} \mathcal{W} \mathbb{E}_t \mathbf{Y}_t^0 = 0$.

To prove part 2 note that by Lemma 2 the OPP adjusted allocation can be written as $\mathbb{E}_t \mathbf{Y}_t^0 + \mathcal{R}_y^0 \boldsymbol{\delta}_t^*$. Plugging this into the gradient of problem (39) gives

$$\mathcal{R}_y^{0\prime} \mathcal{W}(\mathbb{E}_t \mathbf{Y}_t^0 + \mathcal{R}_y^0 \boldsymbol{\delta}_t^*) = \mathcal{R}_y^{0\prime} \mathcal{W} \mathbb{E}_t \mathbf{Y}_t^0 - \mathcal{R}_y^{0\prime} \mathcal{W} \mathbb{E}_t \mathbf{Y}_t^0 = 0 \ .$$

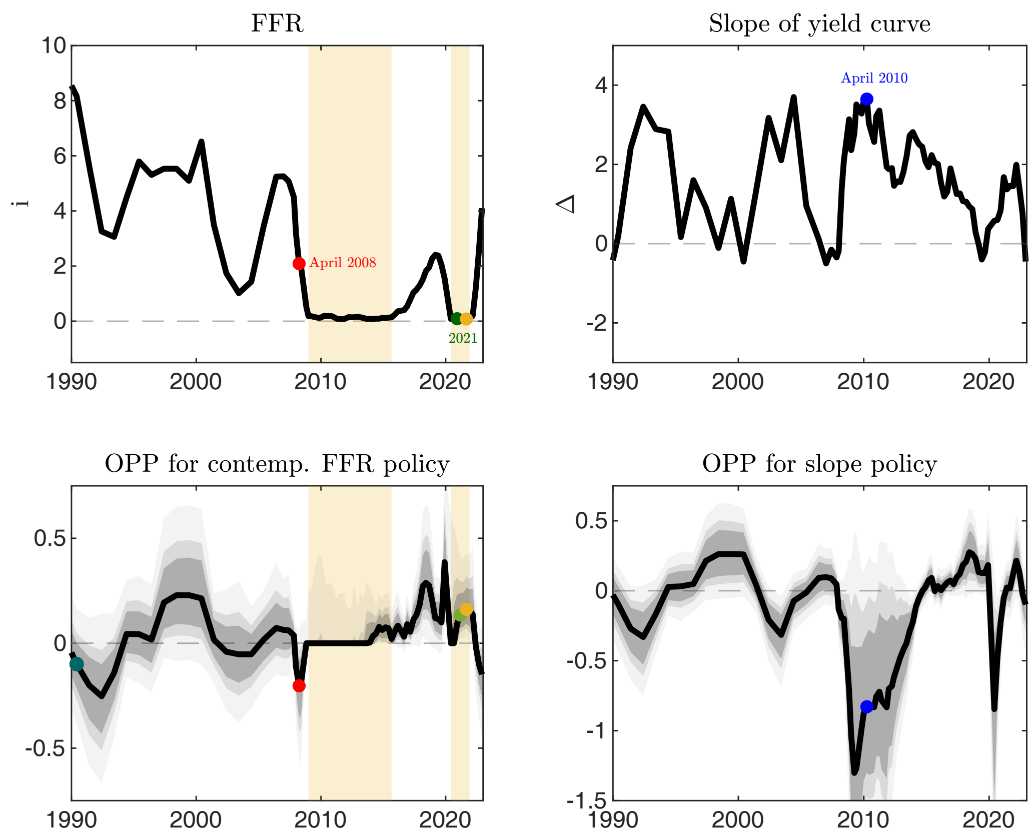
which shows that the OPP adjusted allocation makes the gradient equal to zero and hence $\mathbb{E}_t \mathbf{P}_t^0 + \mathcal{R}_p^0 \boldsymbol{\delta}_t^* = \mathbb{E}_t \mathbf{P}_t^{\text{opt}}$. $\square$

Proof of Corollary 1. Part 1: From Proposition 1 it follows that $\mathcal{R}_y^{0\prime} \mathcal{W} \mathbb{E}_t \mathbf{Y}_t^0 = 0$ if and only if $\mathbb{E}_t \mathbf{P}_t^0 = \mathbb{E}_t \mathbf{P}_t^{\text{opt}}$. Since, $\mathcal{R}_{a,y}^0$ is a subset (or linear combination) of the columns of $\mathcal{R}_y^0$ it follows that $\mathcal{R}_{a,y}^{0\prime} \mathcal{W} \mathbb{E}_t \mathbf{Y}_t^0 \neq 0$ implies that $\mathbb{E}_t \mathbf{P}_t^0 \neq \mathbb{E}_t \mathbf{P}_t^{\text{opt}}$. Part 2: By definition

$$\begin{aligned}
\mathcal{L}_t(\mathbf{0}, \mathbf{0}) &= \frac{1}{2} \mathbb{E}_t \mathbf{Y}_t^{0\prime} \mathcal{W} \mathbf{Y}_t^0 \\
&\geq \frac{1}{2} \mathbb{E}_t (\mathbf{Y}_t^0 + \mathcal{R}_{a,y}^0 \boldsymbol{\delta}_{a,t}^*)' \mathcal{W} (\mathbf{Y}_t^0 + \mathcal{R}_{a,y}^0 \boldsymbol{\delta}_{a,t}^*) \\
&= \mathcal{L}_t(\boldsymbol{\delta}_{a,t}^*, \mathbf{0}) \ .
\end{aligned}$$

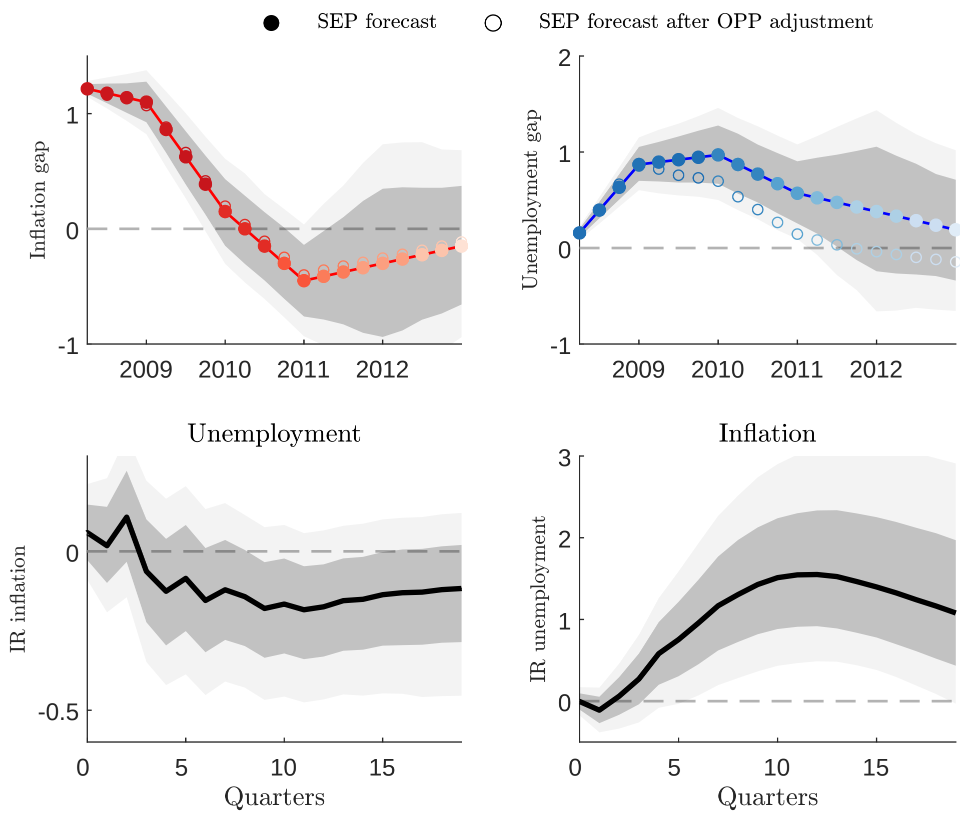
$\square$

Figure 1: A sequence of OPP for Fed monetary policy (1990-2022)



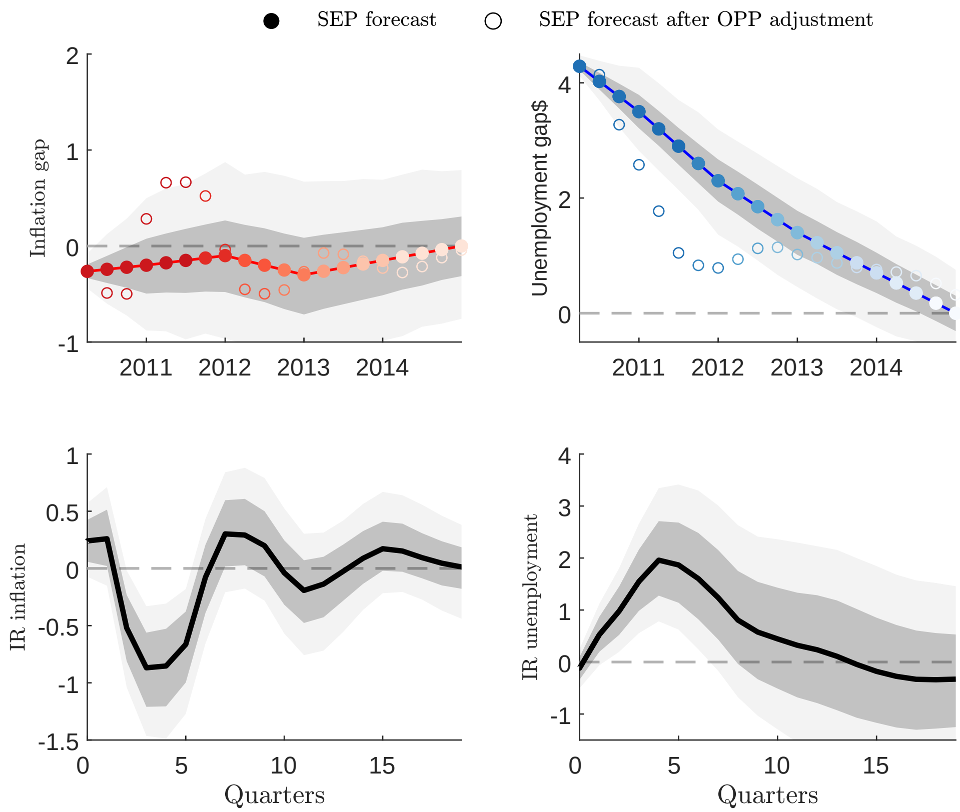
Notes: Top panels: the fed funds rate ("FFR", left-panel) and the difference between the 10-year bond yield and the fed funds rate ("Slope of yield curve", right panel). The yellow shaded areas denote the zero-lower bound (ZLB) periods. Bottom panels: time series for the two elements of the subset OPP: the short-rate OPP (labeled "OPP for contemp. FFR policy", left panel) and the slope OPP (labeled "OPP for slope policy", right panel) over 1990-2022 for a policy maker with a dual inflation–unemployment mandate ($\lambda = 1$). The grey areas capture impulse response and model uncertainty at 60%, 75% and 90% confidence (from darker to lighter shades). The case studies are marked as points: April 2008 (red), April 2010 (blue), March 2021 (green) and November 2021 (yellow).

Figure 2: Fed funds rate policy in April 2008



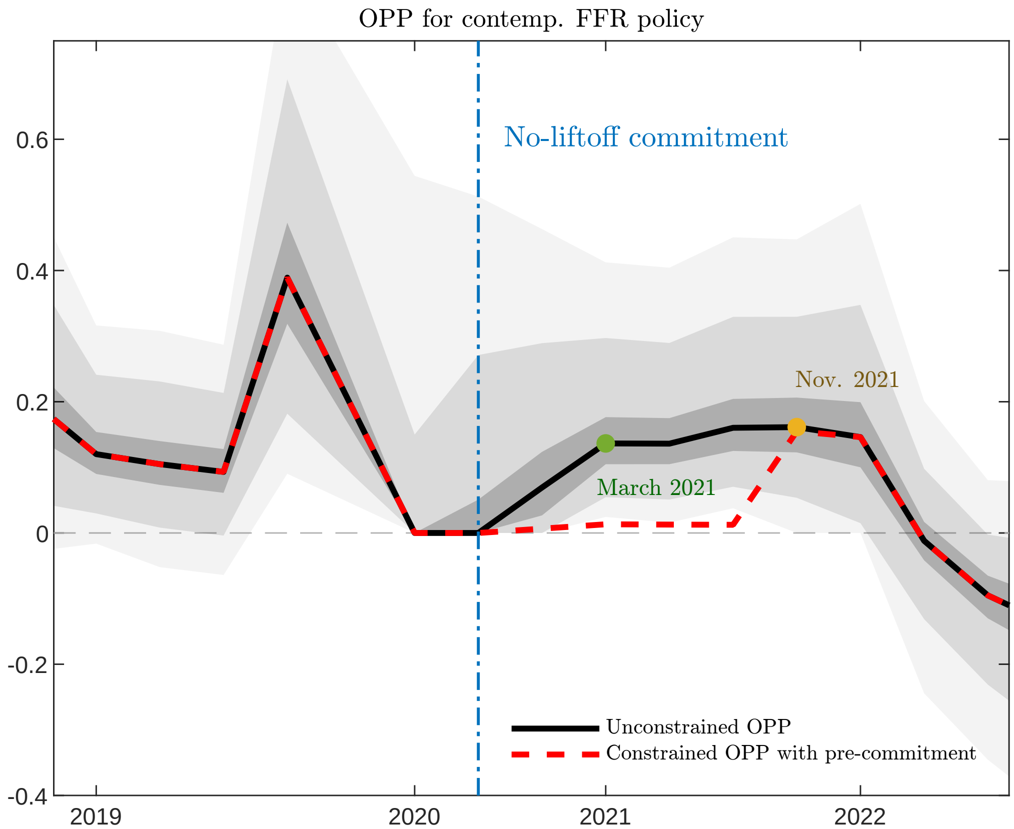
Notes: Top panel: Median SEP forecasts for the inflation and unemployment gaps as of 2008-M4 (in red and blue) along with the 68 and 90 percent confidence bands capturing model uncertainty. Filled circles denote the forecast $\mathbb{E}_t \mathbf{Y}_t^0$, and empty circles denote the forecasts after the short-rate OPP adjustment. Bottom panel: impulse responses of the inflation and unemployment gaps to a fed funds rate shock with 68 and 90 percent confidence intervals.

Figure 3: Slope policy in April 2010



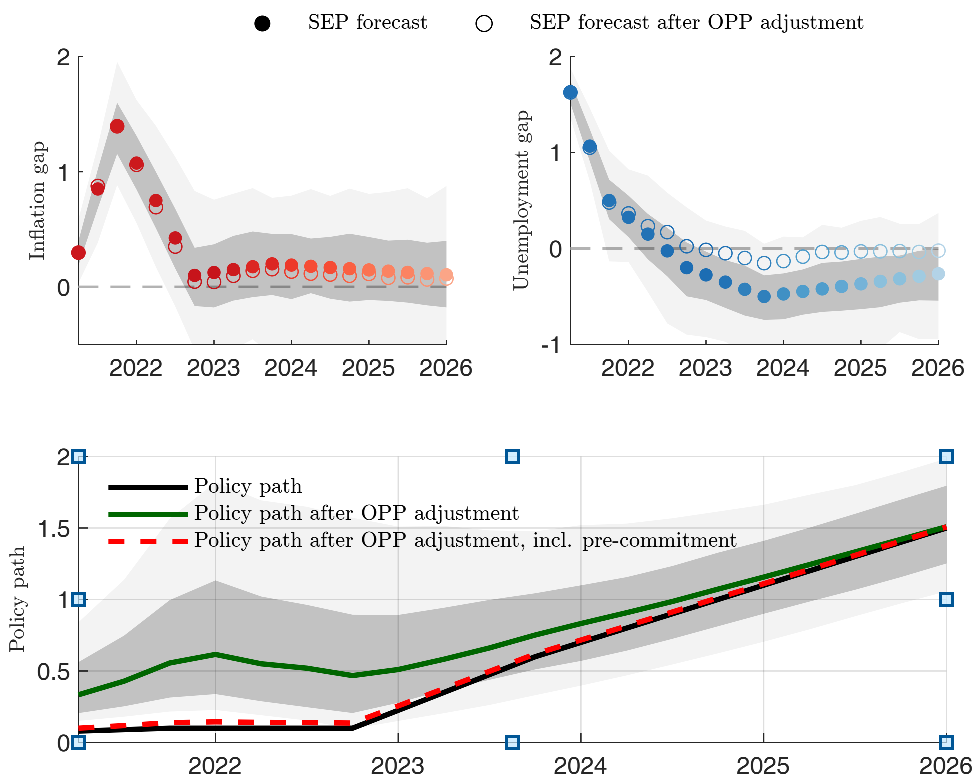
Notes: Top panel: Median SEP forecasts for the inflation and unemployment gaps as of 2010-M4 (in red and blue) along with the 68 and 90 percent confidence bands capturing model uncertainty. Filled circles denote the forecast $\mathbb{E}_t \mathbf{Y}_t^0$, and empty circles denote the forecasts after the slope policy OPP adjustment. Bottom panel: impulse responses of the inflation and unemployment gaps to a slope policy shock with 68 and 90 percent confidence intervals.

Figure 4: A sequence of OPP for Fed funds rate policy (2019-2022)



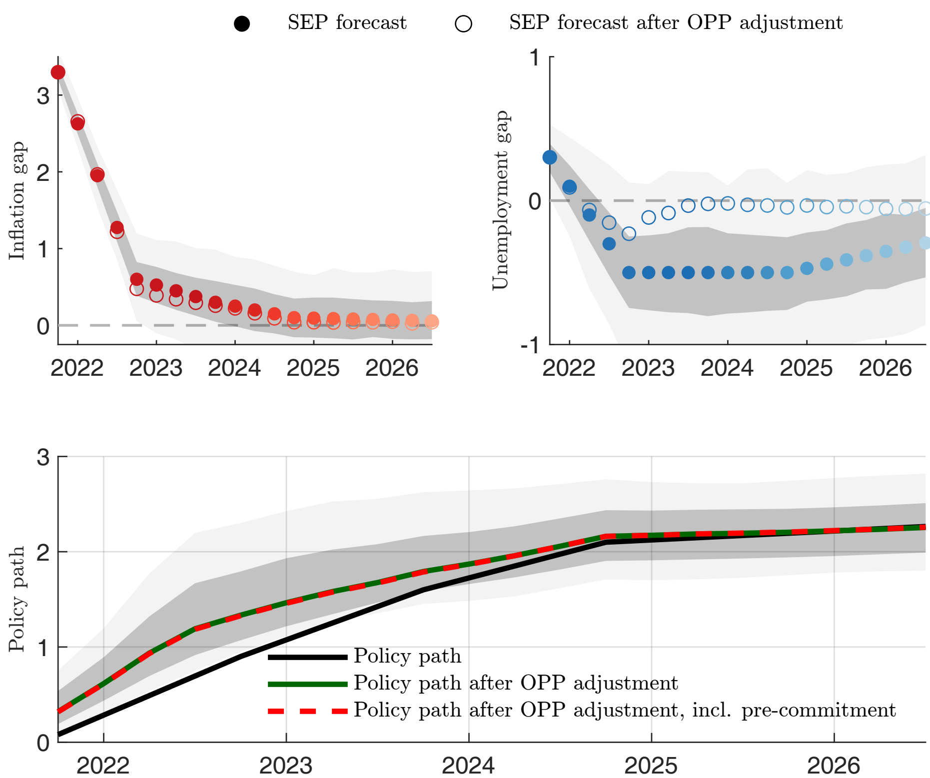
Notes: Time series for the short-rate OPP (labeled "OPP for contemp. FFR policy") over 2019-2022 for a policy maker with a dual inflation–unemployment mandate ($\lambda = 1$). The grey areas capture impulse response and model uncertainty at 60%, 75% and 90% confidence (from darker to lighter shades). The black line denotes the median unconstrained OPP, and the red dashed line denotes the OPP constrained by the Sept. 2020 "no-liftoff commitment" (vertical dashed blue line).

Figure 5: Fed funds rate policy in March 2021



Notes: Top panel: Median SEP forecasts for the inflation and unemployment gaps as of 2021-M3 (in red and blue) along with the 68 and 90 percent confidence bands capturing model uncertainty. Filled circles denote the forecast $\mathbb{E}_t \mathbf{Y}_t^0$, and empty circles denote the forecasts after the short rate unconstrained OPP adjustment. Bottom panel: baseline policy path $\mathbb{E}_t \mathbf{P}_t^0$ (black line), and modified policy path after adjustment by the unconstrained short-rate OPP (green line) along the with 68 and 90 percent confidence bands. The red dashed line denotes the policy path after adjustment by the constrained short-rate OPP.

Figure 6: Fed funds rate policy in November 2021



Notes: Top panel: Median SEP forecasts for the inflation and unemployment gaps as of 2021-M11 (in red and blue) along with the 68 and 90 percent confidence bands capturing model uncertainty. Filled circles denote the forecast $\mathbb{E}_t \mathbf{Y}_t^0$, and empty circles denote the forecasts after the short rate unconstrained OPP adjustment. Bottom panel: baseline policy path $\mathbb{E}_t \mathbf{P}_t^0$ (black line), and modified policy path after adjustment by the unconstrained short-rate OPP (green line) along the with 68 and 90 percent confidence bands. The red dashed line denotes the policy path after adjustment by the constrained short-rate OPP.