Policy Evaluation with Sufficient Macro Statistics*

— A Primer —

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

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

[(b)] Federal Reserve Bank of New York

November 30, 2025

Abstract

Impulse responses and forecasts are central concepts for policy makers. They are also sufficient statistics to solve many important macroeconomic problems, from policy counterfactuals to policy evaluation, and they offer a promising alternative to the standard structural modeling approach. In this work, we discuss and extend recent progress on the use of these sufficient macro statistics for policy evaluation. We illustrate the methods by evaluating the performance of the ECB over 1999-2023.

JEL classification: C14, C32, E32, E52, N10.

Keywords: macro policy, sufficient statistics, impulse responses, forecasting.

*We thank seminar participants at the RES annual conference (2024) and the ECB for helpful comments. Mesters acknowledges financial support from the 2021 ERC Starting Grant POLICYMETRICS. 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, the Federal Reserve Bank of New York or the Federal Reserve System.

1 Introduction

How should we evaluate the latest policy decision by the Fed or the ECB? How should we compare the performances of the Fed vs the ECB at handling the 2007-2008 crisis? And more generally, how should we evaluate the performance of an elected official in office?

Answering these questions is at the core of good macroeconomic policy making, but until recently there have been surprisingly little robust quantitative methods for doing so. Part of this state of affairs owes to the possibility of model mis-specification. Despite impressive recent progress in structural macro modeling, the underlying economy is so complex that quantitative conclusions drawn from a specific model may be too uncertain to reach any form of consensus.

In recent years, several papers have demonstrated how sufficient statistics can be applied in macroeconomics; for policy counterfactuals, for policy guidance, for policy evaluation and even for the elicitation of policy makers' preferences.[1] This "sufficient macro statistics" approach requires minimal assumptions on the underlying structural economic model, and instead relies on recent advances of econometrics; causal inference and forecasting (e.g., Ramey, 2016; Elliot and Timmermann, 2016; Stock and Watson, 2016).[2]

In this review paper, we summarize some of the main lessons of this recent literature in the context of policy evaluation, highlight important directions for further research, and we illustrate the approach with an evaluation of ECB policy since its inception in 1999.

Premise

What makes a policy maker good or bad? Our starting point is simple: policy makers react to the state of the world by taking actions, i.e. they use their policy instruments to achieve certain goals —the policy objectives—.[3] The policy maker's reaction to the state of the world can be expressed as a rule —a reaction function—, which can be sketched as

$$p_t = \bar{p} + \boldsymbol{\phi}\mathbf{y}_t + \sigma_\varepsilon \varepsilon_t \tag{1}$$

where p_t is the policy maker's instrument at time t, and $\mathbf{y}_t$ is a vector capturing the state of the economy.

In this context, a policy maker is good if her reaction function is "optimal", i.e., minimizes the loss function —a metric capturing deviations from the policy objectives—, and policies

[1]See Barnichon and Mesters (2022, 2023*b*,*a*); Beraja (2023); McKay and Wolf (2022, 2023); Caravello, McKay and Wolf (2024); de Groot et al. (2021); Hebden and Winkler (2021).

[2]We note that these more econometric approaches are not immune to model mis-specification, but the number of functional form assumptions is typically much lower.

[3]Goals can be low inflation and full employment for a central bank, or high economic growth and low inequality for a fiscal policy maker.

can be sub-optimal for three broad reasons:

(a) Systematic over/under reaction to the state of the economy, as determined by the coefficient vector $\boldsymbol{\phi}$. For instance, a "bad" central banker can react too strongly/weakly to inflation.

(b) Erratic deviations from the policy rule, as captured by the variance of policy shocks σ_ε. For instance, a "bad" policy maker makes large and/or frequent random mistakes.

(c) Keeping the long-run value of the policy instrument too high/low, as determined by $\bar{p}$. For instance, a "bad" central banker PM keeps the interest rate systematically too low, because it helps debt financing, a case of "fiscal dominance".

In this review, we study developed economies and focus on (a) and (b) —inappropriate reaction to the state of the economy and occasional mistakes—, so that we can consider a stationary environment, where the long-run values of the policy instruments are consistent with the policy maker's objectives. This excludes cases where the policy instrument is systematically too low/high, e.g. leading to runaway inflation or unsustainable debt dynamics.[4]

The intuition underlying the sufficient statistics approach to policy evaluation is simple. Consider a policy maker, say a central banker, with one instrument, say the policy rate, and whose objective is to set the expected policy path at time t —$\mathbb{E}_t p_{t+h}$ for $h = 0, 1, ...$—, in order to minimize a loss function, for instance the (possibly weighted) sum-of-squares for the expected paths of the inflation and unemployment gaps.[5] Without assuming a structural model, how can we evaluate a proposed policy path $\mathbb{E}_t p_{t+h}$?

The key idea is to use natural experiments that occurred in the past —exogenous policy shocks— in order to gauge whether a perturbation to the proposed policy path could lower the loss. To see that, imagine that you identify a policy shock, which led to a temporary increase in the policy path, as sketched in blue the upper panel of Figure 1. Using this natural experiment, we can modify the proposed policy path $\mathbb{E}_t p_{t+h}$ as in the upper-right panel and study how this affects the expected path of the policy objectives (e.g., inflation and unemployment) and thus the loss. If we can lower the loss by appropriately scaling this policy intervention, this means the proposed policy was not optimal, and we can propose a superior policy path. That said, this natural experiment may not be sufficient to compute the *optimal* policy path, as it only probes the short-end of the policy path: whether raising or lower the policy path over the coming periods can lower the loss. To get a more exhaustive

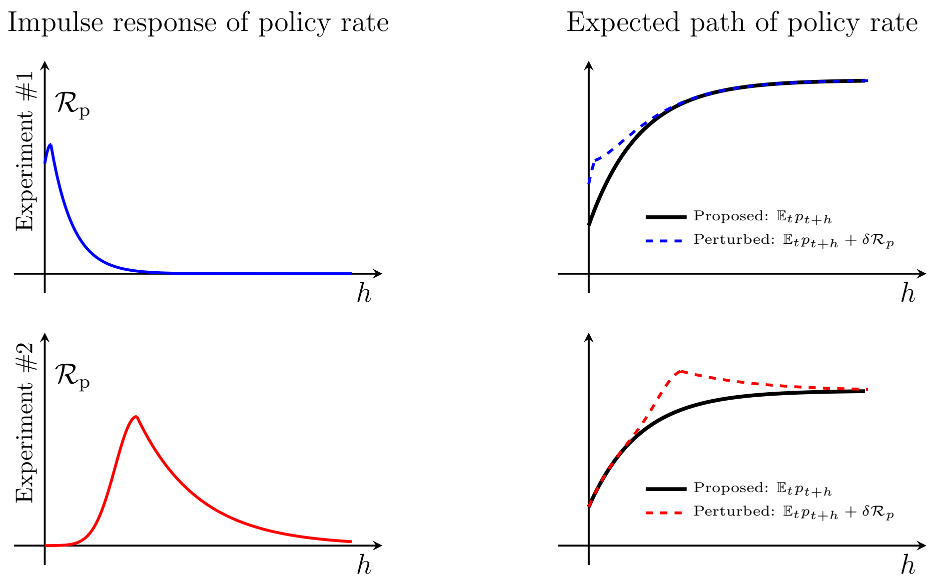
[4]Systematically too low or too high policy instruments (or even drifting policy objectives) is an important source of policy failures in developing countries, and extending our framework to incorporate long-run deviations from targets is an important avenue for future research.

[5]Note that in a static policy problem the policy maker only decides on p_t. In contrast, in a dynamic world the entire expected path $\mathbb{E}_t p_{t+h}$ for $h = 0, 1, ...$ is a policy instrument, which incorporates that expectations about future policies also affect outcomes today.

policy evaluation, we can combine multiple natural experiments. For instance, imagine we can identify two separate policy shocks —one moving the short-end of the policy path and another moving the medium-end as sketched in Figure 1—. In that case, we can span a larger set of policy path counter-factuals and provide more exhaustive policy evaluations and better policy recommendations. In the limit, with sufficiently many natural experiments, we can explore all possible policy counterfactuals and compute the optimal policy which minimizes the loss over all possible policy paths.

In this review we work out this idea in detail and provide a step-by-step implementation guide for constructing various policy evaluation statistics based on the sufficient statistics. We start by placing the approach in the broader context of the macroeconomics literature.

Figure 1: Policy evaluation with Sufficient Macro Statistics



Notes: Left panel: Impulse response of the policy rate to two hypothetical policy shocks: in blue (top row) or in red (bottom row. Right panel: Proposed policy path $\mathbb{E}_t p_{t+h}$ (black line) and perturbed policy path $\mathbb{E}_t p_{t+h} + \delta\mathcal{R}_p$ implied by each policy experiment (in blue or red) and with δ the scaling factor for the policy path perturbation.

Literature review

Jan Tinbergen was the first to explicitly explore the possibility of evaluating macro policy using a statistical model. In 1936 he completed his work on what became the first empirical macroeconomic model which was designed for the Dutch economy and its purpose was to help the Dutch Central Planning Bureau to develop appropriate economic policies, see Dhaene and Barten (1989). Tinbergen (1952) provides an accessible overview of many of these early ideas. The remarks of Theil (1956) are of particular interest as they highlight concerns about uncertainty and the limited control of policy makers which persist today (e.g. Bénassy-Quéré et al., 2018) and form an important element of the sufficient statistics approach. In the first decades after the war the development of macro econometric models flourished. For instance, in the US, Marschak organized a special team at the Cowles Commission for conducting such analysis, see Bodkin, Klein and Marwah (1991) for an extensive discussion.

Lucas (1976) voiced an important criticism of these models: they ignored that agents in the economy typically adjust their behavior when policy decisions are made, as such models for policy evaluation should allow the state of the economy to depend on the actions of the policy maker, not only through the policy instruments but also via the way they shape expectations of agents and perhaps even more structural relationships in the economy. These concerns led to a large literature on structural macro economic modeling of which the New Keynesian theories expounded in Woodford (2003) and Galí (2015), as well as their modern heterogeneous agent counterparts (e.g. Auclert, Rognlie and Straub, 2024), are prime examples. To ensure that the fitted models match the empirical evidence researchers may use impulse response matching, or more general moment matching approaches, to ensure that the model parameters are appropriately set.

All of the aforementioned approaches place a lot of weight on the structural model used. Indeed, if the model is mis-specified the implied policy counterfactuals are generally incorrect. Guided by such concerns and aiming to make more transparent identifying assumptions Sims (1980) and Sims (1982) showed how the more reduced form structural VAR models can enhance transparency and reduce the risk of model mis-specification. In addition, Sims and Zha (1995) show under which conditions the impulse responses to contemporaneous policy shocks, as identified by a structural VAR model, can be used for constructing policy rule counterfactuals. Important works that develop this methodology include Sims and Zha (2006), Bernanke, Gertler and Watson (1997), Leeper and Zha (2003) and Antolín-Díaz, Petrella and Rubio-Ramírez (2021), among many others.

The policy counterfactuals constructed by the Sims and Zha (1995) approach are not robust to the Lucas critique. Intuitively, as a structural VAR model only defines contemporaneous policy shocks, the date $t = 0$ responses to these shocks cannot generally replicate a change in the policy rule coefficients ϕ. Instead to fully match the change in the rule coefficients Sims and Zha (1995) adjust the contemporaneous policy shocks at dates $t = 1, 2, \ldots$. However, it is not obvious that changes in the rule coefficients ϕ at time $t = 0$, yield the outcome as introducing repeated policy shocks for $t = 0, 1, \ldots$, It requires the assumption that the introduction of the shocks does not change the behavior of the agents in the economy, which goes against the Lucas critique. Nonetheless, as argued in e.g. Sims and Zha (1995) and Leeper and Zha (2003) when the shocks are small it would be hard for agents to distinguish between a rule change and a policy surprise so they may not change behavior, see also Kocherlakota (2019). As many regular macro interventions are relatively small, e.g. interest rate changes (Kocherlakota, 2019), SVAR policy evaluation methods remain an important tool for most policy makers.

Broadly speaking, the sufficient statistics approach aims to provide an alternative route for policy analysis that lies between the usage of a fully fledged structural model and the more reduced form SVAR. An early contribution that explores such *semi-structural* route is Beraja (2023) who notes that several linearized models are observationally equivalent under a benchmark policy rule and yield an identical counterfactual equilibrium under an alternative one. Exploiting this counterfactual equivalence allows to reduce the number of restrictions needed in the structural form while retaining robustness to the Lucas critique.

A further reduction in the number of structural restrictions needed can be obtained for a class of structural models where expected policy paths capture *all* effects of policy. For this class of models McKay and Wolf (2023) show that the impulse responses to policy news shocks are sufficient statistics for constructing unconditional policy rule counterfactuals that are robust to the Lucas critique. Intuitively, when all effects of policy are transmitted via the expected policy path, knowing the causal effects of exogenous changes in the policy path *at all horizons* allow to replicate the counterfactual effects of any policy rule that induces a unique equilibrium.

Barnichon and Mesters (2023*b*) evaluate policy decisions for a given time period t and show that for the same class of models optimal policy paths can be characterized by two sufficient statistics: (i) impulse responses to policy news shocks and (ii) forecasts for the macro variables. In this work we show that their results immediately suggest how to construct arbitrary time-t policy path counterfactuals using the same sufficient statistics. de Groot et al. (2021) and Hebden and Winkler (2021) also consider the time-t problem, but use impulse responses from structural models, and focus more on the algorithms needed for computing policy counterfactuals under various constraints.

The key difference between the modern sufficient statistics approach and the SVAR approach of Sims and Zha (1995) lies in the usage of new shocks. Instead of repeatedly introducing a "contemporaneous" policy shocks for dates $t = 0, 1, 2, \ldots$, the sufficient statistics approach replicates the policy counterfactual by using the responses to a sequence of news shocks announced at date $t = 0$ and covering horizons $h = 0, 1, 2, \ldots$. Within the class of models considered, the usage of news shocks, makes the approach robust to the Lucas critique. Importantly, this class covers as special cases the leading dynamic stochastic equilibrium models considered at central banks and other policy institutions. We discuss below in more details for which underlying structural models the approach is robust to the Lucas critique.

Hence to use the sufficient statistics approach for computing policy counterfactuals or optimal policies we require the identification of policy news shocks at ideally *all* horizons. With only a subset of identified shocks the results are often approximations of the desired counterfactuals. To improve the approximation Caravello, McKay and Wolf (2024) propose to supplement the evidence from the subset of identified policy shocks with information from structural macro models. Other ways of extrapolating can rely on smoothness restrictions, e.g. B-splines as in de Boor (2001), or factor structures as in Inoue and Rossi (2021).

While policy counterfactuals and optimal policies are obviously of leading importance in the macro economic toolkit, the sufficient statistics approach can be used for answering other macro questions as well. For instance, Barnichon and Mesters (2023*a*) introduce the distance to minimum loss statistic for comparing policy makers and institutions after their term. Further, Barnichon and Mesters (2022) show how the framework sketched above can be used to learn policy makers preferences through a revealed preference approach.

More broadly the sufficient macro statistics approach draws inspiration from the sufficient statistics approach in public finance (e.g. Chetty, 2009; Kleven, 2020). 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 the loss function is typically a high level assumption, 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, the sufficient macro statistics approach can equally be applied to problems with micro-founded loss functions.

Last, the treatment of uncertainty around the sufficient statistics shares similarities with the robust-control approach to policy making that is outlined in Hansen and Sargent (2008). In particular, parameter and model mis-specification uncertainty are often taken into account when constructing confidence bands around policy recommendations. That said, the decision rules explored by the sufficient macro statistics approach have so far typically focused on minimizing expected loss and have not considered characterizing e.g. minimax optimal policies. This is an important avenue for future research.

Paper outline

The remainder of this paper is organized as follows. In the next section we discuss an illustrative example that relates the sufficient statistics approach to the standard model-based approach to policy evaluation. Section 3 discusses the class of underlying structural models considered. Some useful representations for the equilibrium of the model in terms of the sufficient statistics are shown in Section 4 and the estimation of these statistics is discussed in 5. The methods for time-t and term policy evaluation are discussed Sections 6 and 7, respectively. Section 8 inverts the policy evaluation question and discusses how the sufficient statistics approach can be used to learn the policy maker's preferences. Section 9 illustrates the methods by evaluating the performance of the ECB over 1999-2023. Section 10 concludes.

2 Illustrative example

Before formally describing the general framework, we illustrate the main ideas behind a sufficient statistics approach to macroeconomic policy. In this review, we mainly concentrate on the questions of policy evaluation, and we will focus on two broad questions: (i) real time policy evaluation —how to set or evaluate a macro policy at time t—, and (ii) term policy evaluation —how to evaluate the overall performance of policy makers over their terms. To expound the main ideas underlying the approach and convey the intuition behind the key results, we consider an economy described by the textbook baseline New Keynesian (NK) model (e.g. Galí, 2015). While the sufficient macro statistics approach does *not* rely on specifying a particular model, the NK model helps to convey the inner workings of the approach.

The log-linearized Phillips curve and intertemporal (IS) curve of the baseline New-Keynesian model are given by

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

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

with π_t the inflation gap, x_t the output gap, i_t the nominal interest rate set by the central bank and ξ_t a cost-push shock.[6] The parameters are collected in $\theta = (\kappa, \sigma, \sigma_\xi)'$. We can think of θ as capturing the economic "environment" that is taken as given.

The policy maker sets the interest rate following the rule

$$i_t = \phi_\pi \pi_t + \sigma_\varepsilon \varepsilon_t \ , \tag{4}$$

where ε_t is a policy shock and $\phi = (\phi_\pi, \sigma_\varepsilon)$ is a vector of policy parameters. We impose that the structural shocks ξ_t and ε_t are serially and mutually uncorrelated[7] with mean zero and

[6] In this work, we focus on stationary environments, in which variables evolve around their steady-state. This excludes drifting policy objectives and cases of systematically too low or too high policy instruments; for instance cases of hyper-inflation or unsustainable debt.

[7] This assumption is without loss of generality, and the generic treatment of section 3 accommodates more general (notably serially correlated) exogenous processes.

unit variance, allowing σ_ξ and σ_ε to scale the shocks. The policy rule captures the policy maker's choice set: a reaction coefficient —how the policy maker reacts to the state of the economy—, and an exogenous component —a random mistake—.

Assuming solution unicity ($\phi_\pi > 1$), we can solve the model and express the endogenous variables $Y_t = (\pi_t, x_t)'$ and i_t as functions of the exogenous shocks

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

with

$$\mathcal{R}_y = \sigma_\varepsilon \begin{bmatrix} \frac{-\kappa/\sigma}{1+\kappa\phi_\pi/\sigma} \\ \frac{-1/\sigma}{1+\kappa\phi_\pi/\sigma} \end{bmatrix} \ , \quad \Gamma_y = \sigma_\xi \begin{bmatrix} \frac{1}{1+\kappa\phi_\pi/\sigma} \\ \frac{-\phi_\pi/\sigma}{1+\kappa\phi_\pi/\sigma} \end{bmatrix} \ , \quad \mathcal{R}_i = \sigma_\varepsilon \frac{1}{1+\kappa\phi_\pi/\sigma} \ , \quad \Gamma_i = \sigma_\xi \frac{\phi_\pi}{1+\kappa\phi_\pi/\sigma} \ ,$$

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.

In the context of the introduction and Figure 1, we can think of ε_t as the natural experiment and $\mathcal{R}_i$ captures the consequences of this experiment on the interest rate. Here we only have one experiment, but since the problem is static this will be sufficient for evaluating policy decisions.

Policy counterfactuals with sufficient statistics

Before discussing policy evaluation, we will show how sufficient macro statistics can be used to construct policy rule counterfactuals.

Our starting point is some baseline policy choice given by the pair $(\phi^0, \varepsilon_t^0)$. This policy choices implies a baseline 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. We modify that baseline policy rule with a time-t adjustment, or perturbation, denoted by δ_t, such that

$$i_t = \phi^0 \pi_t + \sigma_\varepsilon^0 \varepsilon_t^0 + \sigma_\varepsilon \delta_t \ . \tag{6}$$

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

$$Y_t = \underbrace{\Gamma_y^0 \xi_t + \mathcal{R}_y^0 \varepsilon_t^0}_{=Y_t^0} + \mathcal{R}_y^0 \delta_t \qquad \text{and} \qquad i_t = \underbrace{\Gamma_i^0 \xi_t + \mathcal{R}_i^0 \varepsilon_t^0}_{=i_t^0} + \mathcal{R}_i^0 \delta_t \ , \tag{7}$$

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 $\mathcal{R}_y^0$ and $\mathcal{R}_i^0$: the impulse responses to a policy shock. The rule adjustment δ_t changes the "conditional forecast" Y_t^0, and the effect is given by the impulse response to policy shocks. In other words, two macro statistics —the impulse responses to policy shocks under the "old" ϕ^0 policy rule (i.e. $\mathcal{R}_y^0$ and $\mathcal{R}_i^0$) and the initial allocation —a baseline conditional forecast Y_t^0— are sufficient to compute the effect of a time t adjustment to the policy rule. We will use this law of motion to search for the optimal policy and evaluate policy decisions.

This counterfactual construction is possible for two reasons: (i) the model is linear and (ii) changing the intercept of the policy rule does not change the coefficients of the Phillips and IS curves (see McKay and Wolf, 2023). This allows to write the effects of the counterfactual policy —moving from i_t^0 to $i_t^0 + \mathcal{R}_i^0\delta_t$— in terms of the old impulse responses $\mathcal{R}_y^0$ and $\mathcal{R}_i^0$.

Importantly, the policy perturbation δ_t can depend on the state of the economy, so that we can use δ_t to compute the effects of specific policy rule counterfactuals, e.g., changing the reaction coefficient from ϕ_π^0 to some ϕ_π^1. To see that, the key is to realize that changing ϕ_π^0 to ϕ_π^1 is equivalent to implementing a perturbation of the form $\delta_t = (i_t^1 - i_t^0)/\mathcal{R}_i^0$. With this adjustment in hand, we can use the law of motion (7) to compute the effects of the counter-factual rule ϕ_π^1: this is $Y_t^1 = Y_t^0 + \mathcal{R}_y^0\delta_t$. As we will see in the general treatment, this approach for counterfactual construction holds much more generally in dynamic forward looking macro models, see Theorem 1 below.

Another important application of δ_t-perturbations consists in adjusting the policy maker's reaction to specific structural *shocks*. Specifically, consider writing the perturbation as

$$\delta_t = \tau_\xi \xi_t + \tau_\varepsilon \varepsilon_t \ ,$$

where $\tau = (\tau_\xi, \tau_\varepsilon)'$ is a vector of *adjustments to shocks*. Plugging in this expression for δ_t in (7) gives

$$Y_t = (\Gamma_y^0 + \mathcal{R}_y^0 \tau_\xi)\xi_t + (\mathcal{R}_y^0 + \mathcal{R}_y^0 \tau_\varepsilon)\varepsilon_t \ . \tag{8}$$

From expression (8), we can see that $\Gamma^0 + \mathcal{R}^0\tau_\xi$ is the impulse response to cost-push shocks *after* the reaction function adjustment τ_ξ. In other words, the adjustment τ_ξ modifies the impulse response to cost-push shocks from Γ^0 to $\Gamma^0 + \mathcal{R}^0\tau_\xi$, which again implies that the "old" impulse responses Γ^0 and $\mathcal{R}^0$ are sufficient macro statistics to compute the effects of these rule counter-factuals. As we will see, this class of rule perturbation will be useful to explore the reasons for sub-optimal policy performances.

Policy evaluation with sufficient statistics

Policy evaluation requires taking a stance on policy objectives and on a performance metric. We posit a quadratic loss function

$$\mathcal{L}_t = Y_t' \mathcal{W} Y_t \ , \tag{9}$$

where $Y_t = (\pi_t, x_t)'$ is the vector of policy objectives (in deviations from their targets) and $\mathcal{W} = \text{diag}(1, \lambda)$ is a weighting matrix, with λ capturing how the policy maker values output gap stabilization over inflation.[8]

A researcher can contemplate two types of policy evaluation questions: (i) real time policy evaluation —this is about evaluating a policy maker's decision at time t—, and (ii) term policy evaluation —this is about evaluating a policy maker's systematic performance over her term—.

Real time policy performance

Our first question is about evaluating (and helping) policy makers facing a time-t decision problem: given a set of initial conditions —the state of the economy today—, what is the optimal policy —the value of the policy instruments— to attain the policy objectives?

The time-t optimal allocation can be characterized by minimizing the time-t loss function with respect to π_t, x_t and i_t subject to the Phillips curve and (IS) curve constraints:

$$\min_{\pi_t, x_t, i_t} \mathcal{L}_t \qquad \text{s.t.} \qquad (2)\text{–}(3) \ . \tag{10}$$

This gives the well known optimal targeting rule $x_t = -\kappa \pi_t / \lambda$, which can be implemented by a policy rule with $\phi_\pi^{\text{opt}} = \kappa\sigma$ and $\sigma_\varepsilon^{\text{opt}} = 0$, and implies $i_t^{\text{opt}} = \frac{\kappa\sigma/\lambda}{1+\kappa^2}\xi_t$ (e.g., Galí, 2015). A limitation of a model based approach however is that it requires the full underlying model, that is the exact specification and coefficients of the Phillips and (IS) curves. This information requirement can be hard to meet in practice.

The sufficient macro statistics approach offers an alternative route that does not require specifying the underlying model equations. The idea is to use the "law of motion" for Y_t following a δ_t rule perturbation —$Y_t = Y_t^0 + \mathcal{R}_y^0 \delta_t$— to find the rule adjustment that minimizes the time-t loss function. Specifically, this consists in finding a δ_t^* that satisfies

$$\delta_t^* = \underset{\delta_t}{\text{argmin}} \ \mathcal{L}_t \qquad \text{s.t.} \qquad Y_t = Y_t^0 + \mathcal{R}_y^0 \delta_t \ . \tag{11}$$

[8]While we consider a quadratic loss function in this review, the method can be extended to arbitrary convex loss functions, see Barnichon and Mesters (2023*b*).

A closed-form solution for δ_t^* is

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

The statistic δ_t^* is what we call the Optimal Policy Perturbation (OPP):[9] an OPP rule adjustment delivers the optimal allocation and the optimal policy is given by $i_t^{\text{opt}} = i_t^0 + \mathcal{R}_i^0\delta_t^*$.[10] This approach for computing the optimal policy only requires the sufficient statistics: the forecast Y_0 and the impulse responses $\mathcal{R}_y^0, \mathcal{R}_i^0$ under the "old" policy rule ϕ^0.

To evaluate a policy maker's decision, one can then compute the distance Δ_t between the loss $\mathcal{L}_t^0$ —the loss under the baseline policy i_t^0 and the optimal policy i_t^{opt}. Some straightforward algebra gives the time-t distance to minimum loss as

$$\begin{aligned}\Delta_t &= \mathcal{L}_t^0 - \mathcal{L}_t^{\text{opt}} \\ &= \delta_t^{*'}\mathcal{R}_y^{0'}\mathcal{W}\mathcal{R}_y^0\delta_t^* \end{aligned} \tag{13}$$

Intuitively, since $\mathcal{R}_y^0\delta_t^*$ gives the counter-factual effect of a δ_t^* rule adjustment on Y_t, the (weighted) sum-of-squares of $\mathcal{R}_y^0\delta_t^*$ gives the "welfare" gains from switching from a sub-optimal policy i_t^0 to the optimal policy i_t^{opt}.

Term policy performance

Our second question is about evaluating policy makers' overall performance, i.e., evaluating policy makers' performance over their whole term in office. To that effect, we consider the unconditional loss function

$$\mathcal{L} = \mathbb{E}\mathcal{L}_t = \mathbb{E}Y_t'\mathcal{W}Y_t \tag{14}$$

$$= \Gamma'\mathcal{W}\Gamma + \mathcal{R}'\mathcal{W}\mathcal{R} \tag{15}$$

[9] Note how the expression for δ_t^* resembles the formula of a weighted least squares regression. In fact, the rule perturbation δ_t^* uses $\mathcal{R}_y^0$ —the impulse responses to policy shocks— in order to best stabilize the policy objectives, i.e., minimize the (weighted) sum-of-squares of Y_t, the allocation after the rule perturbation. This is nothing but a regression of Y_t^0 (the allocation under baseline policy) on $-\mathcal{R}_y^0$. The minus sign is present because the goal is not to best fit Y_t^0, but instead to best "undo" movements in Y_t^0.

[10] We can verify that $i_t^0 + \mathcal{R}_i\delta_t^*$ is the optimal policy:

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

Minimizing the unconditional loss $\mathcal{L}$ is equivalent to minimizing the (weighted) sum-of-squares of the impulse responses of shocks hitting the economy, here Γ and $\mathcal{R}$: an optimal timeless policy is a policy rule that best mutes the effects of shocks. Unlike with the time-t loss function, initial conditions are irrelevant, and this unconditional minimization problem can be seen as a *timeless* perspective of optimal policy; representing the loss of a policy maker appointed at the beginning of time and in place forever.

Specifically, an optimal policy is a policy rule $\phi = (\phi_\pi, \sigma_\varepsilon)$ that solves the problem

$$\min_{\phi_\pi, \sigma_\varepsilon} \mathbb{E}\mathcal{L}_t \qquad \text{s.t.} \qquad (2)\text{–}(3) \; . \tag{16}$$

To evaluate a policy maker over her term, one would like to compute the (unconditional) distance to minimum loss—the distance between the loss under ϕ_0 ($\mathcal{L}^0$) and the minimum loss ($\mathcal{L}^{\text{opt}}$):

$$\Delta = \mathcal{L}^0 - \mathcal{L}^{\text{opt}} \tag{17}$$

Using the model, we can compute the optimal rule (see e.g., Galí, 2015) and show $\phi^{\text{opt}} = (\phi_\pi^{\text{opt}}, \sigma_\varepsilon^{\text{opt}})' = (\kappa\sigma/\lambda, 0)'$, from which we can compute $\mathcal{L}^{\text{opt}}$ and thus Δ.[11]

Again a downside of a model-based approach is that it requires to specify the underlying model, but the sufficient macro statistics approach offers an alternative route. To compute the distance to minimum Δ, one can simply take the unconditional expectations of Δ_t, the time-t distance to minimum loss (Δ_t). Indeed, since $\mathcal{L} = \mathbb{E}\mathcal{L}_t$ we have

$$\begin{aligned} \Delta &= \mathbb{E}\mathcal{L}_t^0 - \mathbb{E}\mathcal{L}_t^{\text{opt}} \\ &= \mathbb{E}(\delta_t^{*\prime} \mathcal{R}_y^{0\prime} \mathcal{W} \mathcal{R}_y^0 \delta_t^*) \end{aligned} \tag{18}$$

This expression shows that we can evaluate a policy maker over her term from a sequence of OPPs corresponding to a sequence of policy decisions. Specifically, given a sample of estimates for the OPP statistics δ_t^* — which in turn requires estimates for the sufficient statistics $\mathcal{R}^0$ and Y_t^0 —, covering the term of the policy maker, we can estimate (18) by its sample average.

Understanding policy performance

One limitation of the distance to minimum loss Δ is that it does not convey why a policy maker delivered a sub-optimal policy performance, whether it was due to large and/or frequent random mistakes, or because of a poor response to the state of the economy. Loosely speaking, Δ does not convey "what went wrong?".

[11] In this example, we have $\mathcal{L}^{\text{opt}} = \Gamma^{\text{opt}\prime}\Gamma^{\text{opt}}$ with $\Gamma^{\text{opt}} = \frac{\sigma_\xi}{1+\kappa^2/\lambda}(1, -\kappa/\lambda)'$.

To answer this question, we can turn to our specialized class of rule perturbations (8), and study the optimality of the policy maker's response to specific structural shocks. The idea is to search for the optimal rule perturbation τ^* to solve

$$\min_{\tau} \mathbb{E} Y_t^{'} \mathcal{W} Y_t \qquad \text{s.t.} \qquad Y_t = \left(\Gamma^0 + \ \mathcal{R}^0 \tau_\xi\right) \xi_t + (\mathcal{R}^0 + \mathcal{R}^0 \tau_\varepsilon) \varepsilon_t$$

Again, there is a closed-form solution with

$$\tau_\xi^* = -(\mathcal{R}^{0'} \mathcal{W} \mathcal{R}^0)^{-1} \mathcal{R}^{0'} \mathcal{W} \Gamma^0 \qquad \text{and} \qquad \tau_\varepsilon^* = -1 \ . \tag{19}$$

The statistic τ^* is what we call the Optimal Reaction Adjustment (ORA). Starting from an initial rule ϕ^0, the ORA makes the rule optimal by optimizing the policy maker's systematic response to each type of structural shocks: (i) it adjusts the reaction to non-policy shocks ξ_t in order to minimize their effects, i.e., to reach Γ^{opt}, and (ii) it cancels monetary mistakes by setting the effect of policy shocks back to zero with $\tau_\varepsilon^* = -1$. In fact, the ORA allows to characterize the optimal policy rule starting from any baseline rule ϕ^0.[12]

Note the similarity between the expressions for δ_t^* and τ^*, as both (13) and (19) are weighted least squares regression formulas. While the OPP δ_t^* is given by a regression of Y_t^0 on $-\mathcal{R}^0$, the ORA τ^* is given by a regression of Γ^0 on $-\mathcal{R}^0$.[13] Intuitively, the ORA is the optimal rule adjustment *conditional* on a specific shock, while the OPP is the optimal rule adjustment to the state of the economy at t (Y_t^0). Since the state of the economy at t is a function of present and past shocks that hit the economy, there is a straightforward connection between ORA and OPP. Here, we have $\delta_t^* = \tau_\xi^* \xi_t + \tau_\epsilon^* \epsilon_t$: the optimal policy perturbation (OPP) is the sum of the optimal reactions (ORA) to each type of shock.

By exploiting this connection, we can be decompose the total DML as

$$\Delta = \Delta_\xi + \Delta_\varepsilon \ , \tag{20}$$

where

$$\begin{aligned} \Delta_\xi &= \tau_\xi^{*'} \mathcal{R}^{0'} \mathcal{W} \mathcal{R}^0 \tau_\xi^* \\ &= \Gamma^{0'} \mathcal{W} \mathcal{R}^0 \left(\mathcal{R}^{0'} \mathcal{W} \mathcal{R}^0\right)^{-1} \mathcal{R}^{0'} \mathcal{W} \Gamma^0 \end{aligned} \qquad \begin{aligned} \Delta_\varepsilon &= \tau_\varepsilon^{*'} \mathcal{R}^{0'} \mathcal{W} \mathcal{R}^0 \tau_\varepsilon^* \\ &= (\mathcal{R}^{0'} \mathcal{W} \mathcal{R}^0)^{-1} \end{aligned}$$

Each element of Δ captures the distance to minimum loss for a specific type of shock. Since each type of shock can be studied separately from the others, we can split the optimal

[12] We can verify that $\tau^* = (\tau_\xi^*, \tau_\varepsilon^*)'$ delivers the optimal reaction function with $\Gamma^0 + \mathcal{R}^0 \tau_\xi{}^* = (I - \mathcal{R}^0 (\mathcal{R}^{0'} \mathcal{R}^0)^{-1} \mathcal{R}^{0'}) \Gamma^0 = \Gamma^{\text{opt}}$.

[13] In both cases, the idea is to use the policy instrument (with effect $\mathcal{R}^0$) to best stabilize the economy, either from a time-t perspective (stabilizing Y_t^0) for the OPP or from a timeless perspective (stabilizing the impulse responses to shocks: Γ^0 or $\mathcal{R}^0$) for the ORA.

policy problem into separate problems, and evaluate policy makers separately for each type of shock. For instance, Δ_ξ captures how well the policy maker responded to cost-push shocks. Expression (20) allows to understand where the policy maker made sub-optimal decisions, whether it mis-reacted to certain types of shocks (a large Δ_ξ) or whether it made frequent or large policy mistakes (a large Δ_ε).

Taking stock

In sum, this example illustrates how we can evaluate policy makers from three sufficient macro statistics: (i) $\mathcal{R}^0$, the impulse responses to policy shocks, (ii) Y_t^0, the state of the economy at time t, which is simply a conditional forecast, and (iii) Γ^0, the impulse responses to non-policy shocks, should one wants to better understand the reasons for sub-optimal policy performances. In the next sections, we will refine and generalize these findings for general linear forward looking macro models.

3 Structural model

A goal of the sufficient macro statistics approach is to impose minimal assumptions on the underlying economic model. In fact, the only structure that we impose is that the data generating process (DGP) belongs to a class of generic macro models, that is: the true underlying DGP is a special case of generic model. Importantly, we will not assume that we can learn which specific model generated the data.[14]

Inspired by Auclert et al. (2021), we adopt a sequence space representation, which is somewhat different from the usual recursive way of writing down dynamic models (e.g. Ljungqvist and Sargent, 2004). In the appendix we describe in more detail the notations and benefits underlying the sequence space representation.

Let $\mathbf{Y}_t = (y_t', y_{t+1}', \ldots)'$ denote the time-t path of the macro variables that populate the economy. Specifically, y_{t+h} is an $M_y \times 1$ vector containing the variables of interest at time $t+h$. The policy path is defined by $\mathbf{P}_t = (p_t', p_{t+1}', \ldots)'$, where p_{t+h} is the $M_p \times 1$ vector of policy instruments available at time $t+h$. For instance, a monetary policy maker decides on the path of the overnight interest rate (and possibly on the path additional non-standard monetary policy actions, such as bond market purchases), while a government

[14]Relaxing this assumption is at the core of the sufficient macro statistics approach. It allows to considerably sidestep/alleviate the many challenges created by the possibility of model mis-specification, as well as the identification challenges that have plagued the estimation of modern structural macro models (e.g. Canova and Sala, 2009; Andrews and Mikusheva, 2015). We will only need to estimate impulse responses implied by the general model and construct forecasts, both of which need not suffer from the general identification problems that arise in structural macro equations/models. Or at least, there exists a plethora of alternative approaches that can be adopted.

decides on paths for spending, taxes and transfers over the coming years (e.g. Alesina, Favero and Giavazzi, 2019). We assume that all variables have been suitably detrended to be stationary.[15]

Our generic linear model for the economy at time t is given by

$$\mathcal{A}_{yy}\mathbb{E}_t\mathbf{Y}_t - \mathcal{A}_{yp}\mathbb{E}_t\mathbf{P}_t = \mathbf{X}_{-t} + \mathcal{B}_{y\xi}\Xi_t \tag{21}$$

$$\mathcal{A}_{pp}\mathbb{E}_t\mathbf{P}_t - \mathcal{A}_{py}\mathbb{E}_t\mathbf{Y}_t = \mathbf{X}_{-t} + \mathcal{B}_{p\xi}\Xi_t + \mathcal{B}_{p\varepsilon}\boldsymbol{\varepsilon}_t \ , \tag{22}$$

where the pre-determined inputs on the right hand side are the time-t paths of news shocks: $\Xi_t = (\xi'_{t,t}, \xi'_{t,t+1}, \xi'_{t,t+2}, \ldots)'$ and $\boldsymbol{\varepsilon}_t = (\varepsilon'_{t,t}, \varepsilon'_{t,t+1}, \varepsilon'_{t,t+2}, \ldots)'$, as well as the path of any time-t initial conditions $\mathbf{X}_{-t}$, which includes the effects of all past shocks.

The vector $\xi_{t-j,t+h}$ includes structural shocks that capture the exogenous information about time period $t+h$ but are released at time $t-j$. Similarly, $\varepsilon_{t-j,t+h}$ is the vector of policy news shocks for period $t+h$ that are released at $t-j$. We assume that all news shocks are mean zero with unit variance and mutually and serially uncorrelated.

The linear maps $\mathcal{A}_{..}$ and $\mathcal{B}_{..}$ are infinite dimensional and conformable such that the multiplications are well defined. The conditional expectation operator is defined as $\mathbb{E}_t(\cdot) = \mathbb{E}(\cdot|\mathcal{F}_t)$, where the time-$t$ information set $\mathcal{F}_t$ is defined in terms of the pre-determined inputs, i.e. $\mathcal{F}_t = \{\mathbf{X}_{-t}, \Xi_t, \boldsymbol{\varepsilon}_t\}$, where $\boldsymbol{\varepsilon}_t$ are policy news shocks that we formally introduce below.

To ease future notation we collect all parameters of the general model (21) in

$$\theta = \{\mathcal{A}_{yy}, \mathcal{A}_{yp}, \mathcal{B}_{y\xi}\} \qquad \text{and} \qquad \phi = \{\mathcal{A}_{pp}, \mathcal{A}_{py}, \mathcal{B}_{p\xi}, \mathcal{B}_{p\varepsilon}\} \ . \tag{23}$$

The parameters θ describe the environment that the policy maker faces and ϕ is the reaction function of the policy maker.

The model (21) is general and accommodates a large class of structural models found in the literature, not only standard New-Keynesian (NK) models (e.g., Smets and Wouters, 2007), but also some modern heterogeneous agents NK models (e.g. Auclert et al., 2021). It is useful to compare the general structure of (21) to that of the simple NK model of the illustrative example. The Phillips and IS curves in equations (2) and (3) can be stacked across horizons $t+h$ for $h = 0, 1, 2, \ldots$, to form the first equation of (21) – the non-policy block – and the interest rate rule (4) can be similarly stacked to form the second equation – the policy block –.

Naturally, there are important structural models that cannot be written in the form of

[15]In this work, we focus on stationary environments, in which variables evolve around their steady-state and where the steady-state coincides with the policy targets Y^*. This excludes drifting policy objectives and cases of systematically too low or too high policy instruments; for instance systematic inflation target misses, cases of hyper-inflation, or cases of unsustainable debt.

(21). First of all, the model is linear and can therefore not accommodate nonlinear structural models. This restriction is largely made for convenience in this review as different types of popular and feasible nonlinearities, such as state dependence and time-varying coefficients, can be easily incorporated (e.g. Barnichon and Mesters, 2023*b*). In addition, identification results for general nonlinear models can be found in the supplementary material of McKay and Wolf (2023).

The second, somewhat implicit restriction in (21) is that the coefficients of the policy rule ϕ do not affect the coefficients in the environment θ. This rules out, e.g., models with learning, where the (some) coefficients θ are gradually updated based on agents learning about the policy rule coefficients ϕ (e.g. Lucas, 1972). Indeed, by the letter of Lucas (1976) this implies that the class of models (21) is not robust to the Lucas critique as those models feature mappings like $\theta(\phi)$, where θ then varies as ϕ is changed. That said, model (21) does allow expectations about current and future variables to change as the policy maker adjusts ϕ, which is the way the Lucas critique is resolved in many macro models.

4 Equilibrium representations and counterfactual policy paths

We will now discuss a number of useful representations for the structural model (21)-(22) that characterize the equilibrium in terms of sufficient macro statistics —impulse responses and forecasts—. A particularly attractive property of these types of representations is that they allow to characterize the equilibrium allocation under *alternative* policy paths (e.g., under an alternative reaction function ϕ) in terms of sufficient statistics alone.

For any specific model in the general class considered we could formulate primitive conditions on the maps $\mathcal{A}_{..}$ and $\mathcal{B}_{..}$ that ensure the existence of a unique and determinate equilibrium. However, since we will generally not be interested in distinguishing among models in the class, we will simply impose a high level assumption on the existence of a *baseline policy rule* vector ϕ^0:

Assumption 1. *There exists a baseline reaction function ϕ^0 under which $\mathcal{A}^0_{pp}$ and $\mathcal{A}_{yy} - \mathcal{A}_{yp}(\mathcal{A}^0_{pp})^{-1}\mathcal{A}^0_{py}$ are invertible maps, i.e. ϕ^0 leads to a unique and determinate equilibrium.*

While the baseline rule could be any rule ensuring a unique equilibrium, in practice it is helpful to think of the baseline rule as a rule that was in place in the recent past, such that a recent sample of data was generated under the baseline rule. This will be necessary to estimate the sufficient macro statistics. In fact, the existence of the baseline rule ensures that we can define impulse responses and forecasts given this rule:

Lemma 1. *Given the generic model* (21)-(22), *under the policy choice* ϕ^0 *that satisfies Assumption 1, we have*

$$\begin{array}{rcl} \mathbb{E}_t\mathbf{Y}_t^0 & = & \Gamma_y^0\mathbf{S}_t + \mathcal{R}_y^0\boldsymbol{\varepsilon}_t \\ \mathbb{E}_t\mathbf{P}_t^0 & = & \Gamma_p^0\mathbf{S}_t + \mathcal{R}_p^0\boldsymbol{\varepsilon}_t \end{array} , \tag{24}$$

where $\mathbf{S}_t = (\boldsymbol{\Xi}_t', \mathbf{X}_{-t}')'$.

Proof. See Barnichon and Mesters (2023*b*). □

The lemma defines the expected paths for the objectives $\mathbf{Y}_t^0$ and the policy path $\mathbf{P}_t^0$ as a function of the state of the economy $\mathbf{S}_t = (\mathbf{X}_{-t}', \boldsymbol{\Xi}_t')'$, which captures initial conditions $\mathbf{X}_{-t}$ and the non-policy news shocks $\boldsymbol{\Xi}_t$, as well as the policy news shocks $\boldsymbol{\varepsilon}_t$. These expected paths, or oracle forecasts, are conditional on the baseline policy rule ϕ^0 and hence we have indexed the outcomes with a superscript ^{0}.

Simultaneously 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. Specifically, we have that $\mathcal{R}_j^0$ captures the impulse responses of $j = y, p$ to policy news shocks at different horizons —from horizon-0 ($\varepsilon_{t,t}$) to any horizon $h > 0$ ($\varepsilon_{t,t+h}$)— under the rule ϕ^0. Similarly, Γ_j^0 captures the impulse responses of $j = y, p$ to the state of the economy $\mathbf{S}_t$.

Policy counterfactuals with sufficient macro statistics

Often we are interested in the outcomes under some alternative policy path $\mathbb{E}_t\mathbf{P}_t^1$ that results from an alternative policy rule ϕ^1. Directly mimicking Lemma 1 would lead to impulse responses $\Gamma_{\cdot\cdot}^1$ and $\mathcal{R}_{\cdot\cdot}^1$ which are defined under the new policy rule ϕ^1. Unfortunately, unless this rule ϕ^1 was used in the past, it is not possible to estimate impulse responses under ϕ^1. To circumvent this we show that there exists a representation of the equilibrium allocation under ϕ^1 in terms of the forecasts and impulse responses *under the baseline rule* ϕ^0. In other words, as long as we can estimate the sufficient statistics under one baseline rule, we can construct any policy counterfactual.

To set this up we first present a useful lemma. Suppose that $\boldsymbol{\delta}_t = (\boldsymbol{\delta}_{0t}', \boldsymbol{\delta}_{1t}', \ldots)'$ with $\boldsymbol{\delta}_{jt} \in \mathbb{R}^{M_p}$. We define the modified policy rule

$$\mathcal{A}_{pp}^0\mathbb{E}_t\mathbf{P}_t - \mathcal{A}_{py}^0\mathbb{E}_t\mathbf{Y}_t = \mathbf{X}_{-t} + \mathcal{B}_{p\xi}^0\boldsymbol{\Xi}_t + \mathcal{B}_{p\varepsilon}^0\boldsymbol{\varepsilon}_t + \mathcal{B}_{p\varepsilon}^0\boldsymbol{\delta}_t , \tag{25}$$

which adjusts the policy rule (22) under ϕ^0 by $\boldsymbol{\delta}_t$, which is rescaled by $\mathcal{B}_{p\varepsilon}^0$ to ensure that the units of the adjustments are the same as the units of the policy shocks. The following lemma characterizes the equilibrium representation under this adjustment.

Lemma 2. *For any* $\boldsymbol{\delta}_t$ *such that either* $\boldsymbol{\delta}_t$ *is deterministic or* $\boldsymbol{\delta}_t$ *admits a representation* $\boldsymbol{\delta}_t = \mathcal{T}_s\mathbf{S}_t + \mathcal{T}_\varepsilon\boldsymbol{\varepsilon}_t$ *for arbitrary fixed maps* $\mathcal{T}_s, \mathcal{T}_\varepsilon$, *given the generic model* (21) *and the modified* *policy rule* (25) *with* ϕ^0 *satisfying assumption 1, we have that*

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

Proof. See Barnichon and Mesters (2023*b*). □

Different policy adjustments can be considered,[16] and Lemma 2 shows that their effects can always be computed from the sufficient statistics defined under the baseline rule ϕ^0. This important lemma can be used in various ways. First and foremost, McKay and Wolf (2023) are interested in the alternative policy rules of the form $\phi^1 = \{\mathcal{A}_{pp}^1, \mathcal{A}_{py}^1, \mathcal{B}_{p\xi}^1, \mathbf{0}\}$ under the assumption that this rule induces a unique equilibrium. To find this counterfactual they show that one needs to choose $\boldsymbol{\delta}_t$ to solve

$$\mathcal{A}_{pp}^1(\mathbb{E}_t\mathbf{P}_t^0 + \mathcal{R}_p^0\boldsymbol{\delta}_t) - \mathcal{A}_{py}^1(\mathbb{E}_t\mathbf{Y}_t^0 + \mathcal{R}_y^0\boldsymbol{\delta}_t) = \mathbf{X}_{-t} + \mathcal{B}_{p\xi}^1\boldsymbol{\Xi}_t \ .$$

Note that since both the baseline and the counterfactual rule are assumed to induce unique equilibria, Lemma 1 implies that the adjustment can be represented as a function of the predetermined inputs, i.e. there exist maps $\mathcal{T}_s$ and $\mathcal{T}_\varepsilon$ such that $\boldsymbol{\delta}_t = \mathcal{T}_s\mathbf{S}_t + \mathcal{T}_\varepsilon\boldsymbol{\varepsilon}_t$ and Lemma 2 can be used to recover the counterfactuals.

Second, Barnichon and Mesters (2023*b*) choose the optimal $\boldsymbol{\delta}_t$ in order to minimize some loss function. This effectively amounts to treating $\boldsymbol{\delta}_t$ as the choice variable in an optimization problem where the optimal $\boldsymbol{\delta}_t$ subsequently becomes a function of $\mathbf{S}_t$ and $\boldsymbol{\varepsilon}_t$. We discuss this usage in detail in Section 6 below.

Next, we show how lemma 2 allows to characterize the allocation under an alternative policy path $\mathbb{E}_t\mathbf{P}_t^1$ as a function of sufficient macro statistics computed under the baseline rule ϕ_0.

Theorem 1. *The counterfactual macro outcome path under the policy path* $\mathbb{E}_t\mathbf{P}_t^1$ *can be computed in two steps:*

1. $\boldsymbol{\delta}_t^{0\to1} = (\mathcal{R}_p^{0\prime}\mathcal{R}_p^0)^{-1}\mathcal{R}_p^{0\prime}(\mathbb{E}_t\mathbf{P}_t^1 - \mathbb{E}_t\mathbf{P}_t^0)$
2. $\mathbb{E}_t\mathbf{Y}_t^1 = \mathbb{E}_t\mathbf{Y}_t^0 + \mathcal{R}_y^0\boldsymbol{\delta}_t^{0\to1}$

Proof. Setting $\mathbb{E}_t\mathbf{P}_t(\boldsymbol{\delta}_t) = \mathbb{E}_t\mathbf{P}_t^1$ and using Lemma 2 to solve for $\boldsymbol{\delta}_t$ gives 1. $\boldsymbol{\delta}_t^{0\to1} = (\mathcal{R}_p^{0\prime}\mathcal{R}_p^0)^{-1}\mathcal{R}_p^{0\prime}(\mathbb{E}_t\mathbf{P}_t^1 - \mathbb{E}_t\mathbf{P}_t^0)$. Using Lemma 2 again with $\boldsymbol{\delta}_t = \boldsymbol{\delta}_t^{0\to1}$ gives 2. □

[16] For instance, $\boldsymbol{\delta}_t$ could come from adjustments to the policy rule coefficients, as the $\mathcal{T}$'s represent adjustment to the policy rule reaction coefficients. In particular, $\mathcal{T}_s\mathbf{S}_t$ represents adjustments to the systematic reaction to non-policy shocks $\boldsymbol{\Xi}_t$ or to the state of the economy $\mathbf{X}'_{-t}$ (the initial conditions). Alternatively, $\boldsymbol{\delta}_t$ could be a deterministic adjustment to the rule; an intercept.

The theorem shows how a policy maker who wishes to explore the consequences of a different policy path can use the baseline forecasts and impulse responses to recover the counterfactual. In step 1 the needed adjustment $\boldsymbol{\delta}_t^{0\to1}$ is recovered from the difference between the baseline path $\mathbb{E}_t\mathbf{P}_t^0$ and the desired path $\mathbb{E}_t\mathbf{P}_t^0$. The notation $\boldsymbol{\delta}_t^{0\to1}$ reflect the change from path 0 to path 1 at time t. In step 2 the effect of this adjustment on the macro outcomes is computed from the baseline forecast $\mathbb{E}_t\mathbf{Y}_t^0$ and the causal effects of policy on the outcomes $\mathcal{R}_y^0$.

Theorem 1 relies on the identification result of McKay and Wolf (2023) but allows to consider policy counterfactuals in terms of policy path counterfactuals, rather than policy rule counterfactuals. This can be helpful in practice, when policy makers are able to articulate the policy path that they are interested, rather the counterfactual rule that this path implies. Indeed, policy makers need not have an explicit formulation of their desired reaction function.

Theorem 1 has a practical limitation however: it requires the identification of all policy news shocks at different horizons, i.e., the estimation of all the columns of $\mathcal{R}_p^0$ and $\mathcal{R}_y^0$. In practice, this may not be possible. For Theorem 1 this implies that not all columns of $\mathcal{R}_p^0$ and $\mathcal{R}_y^0$ can be recovered from the data. Two general solutions exist: (i) fill, or approximate, the missing columns by extrapolating from the known columns or (ii) compute the best approximating policy path using the available evidence. Strategy (i) is pursued in Caravello, McKay and Wolf (2024), who use a collection of structural macro models, which are fitted using impulse response matching based on the available impulse response functions, to perform the extrapolation. Their approach exploits Lemma 2 in order to avoid specifying a policy rule when doing impulse response matching. Alternatively, de Groot et al. (2021) and Hebden and Winkler (2021) use structural models directly to obtain estimates for the impulse responses to all policy shocks.

Strategy (ii) is simple and can be formalized as follows. Let $\mathcal{R}_{a,p}^0$ and $\mathcal{R}_{a,y}^0$ denote the linear combinations of the columns $\mathcal{R}_p^0$ and $\mathcal{R}_y^0$ that correspond to the policy news shocks $\boldsymbol{\varepsilon}_{a,t} = A\boldsymbol{\varepsilon}_t$ which can be identified.

Corollary 1. *The best linear approximation for the macro outcome path under the policy path $\mathbb{E}_t\mathbf{P}_t^1$ can be computed in two steps:*

1. $\boldsymbol{\delta}_{a,t}^{0\to1} = (\mathcal{R}_{a,p}^{0\prime}\mathcal{R}_{a,p}^0)^{-1}\mathcal{R}_{a,p}^{0\prime}(\mathbb{E}_t\mathbf{P}_t^1 - \mathbb{E}_t\mathbf{P}_t^0)$
2. $\mathbb{E}_t\mathbf{Y}_t^{1,a} = \mathbb{E}_t\mathbf{Y}_t^0 + \mathcal{R}_{a,y}^0\boldsymbol{\delta}_{a,t}^{0\to1}$

which is the counterfactual under $\mathbb{E}_t\mathbf{P}_t^0 + \mathcal{R}_{a,p}^0\boldsymbol{\delta}_{a,t}^{0\to1}$.

The result shows that given only $\mathcal{R}_{a,p}^0$ and $\mathcal{R}_{a,y}^0$ the researcher can only compute the counterfactual under $\mathbb{E}_t\mathbf{P}_t^0 + \mathcal{R}_{a,p}^0\boldsymbol{\delta}_{a,t}^{0\to1}$, which is the best approximation to $\mathbb{E}_t\mathbf{P}_t^1$ for which the counterfactual can be identified. The quality of this approximation should be judged on a case-by-case (i.e., counterfactual-by-counterfactual) basis. Sometimes, the approximation $\mathbb{E}_t\mathbf{P}_t^0 + \mathcal{R}_{a,p}^0\boldsymbol{\delta}_{a,t}^{0\to 1}$ is able to capture $\mathbb{E}_t\mathbf{P}_t^1$, in which case the identified causal effects are able to compute the policy counterfactual of interest, sometimes this is not the case.

5 Inference on counterfactual policy paths

We discuss the estimation of the impulse responses and the oracle forecasts under the baseline policy rule. Given these estimates and their distribution we provide an algorithm for evaluating counterfactual policy paths.

5.1 Estimating impulse responses

Theorem 1 reveals that the impulse responses to the policy news shocks, $\mathcal{R}_p^0$ and $\mathcal{R}_y^0$, are of key interest. We will discuss the estimation of these objects noting that similar steps can be taken for estimating the Γ's.[17]

Following the discussion above, we must often satisfy ourselves with identifying some subsets of these shocks on a subset of the outcome variables. To set this up, let $\mathbb{E}_t\mathbf{D}_t^0 = \mathbb{E}_t(y_t^{0'}, \ldots, y_{t+H}^{0'}, p_t^{0'}, \ldots, p_{t+H}^{0'})$ for some finite horizon H, and let $\boldsymbol{\varepsilon}_{a,t}$ denote the subset of policy news shocks that can be identified.

By selecting the appropriate rows from Lemma 1 we obtain

$$\mathbf{D}_s^0 = \mathcal{R}_{a,H}^0\boldsymbol{\varepsilon}_{a,s} + \mathbf{U}_s \ , \qquad s \in \mathfrak{T} \ . \tag{26}$$

where $\mathcal{R}_{a,H}^0$ is a finite dimensional matrix that includes the entries of the maps $\mathcal{R}_{a,y}^0, \mathcal{R}_{a,p}^0$ implied by our choices for $\mathbf{D}_t$ and $\boldsymbol{\varepsilon}_{a,t}$. The time periods used are included in the set $\mathfrak{T}$, nothing that for time-t policy evaluation often $\mathfrak{T} = \{t_0, \ldots, t\}$ will be used whereas for term policy evaluation often the sample is taken as the period over which the policy maker was in office. A key assumption is that during $\mathfrak{T}$ the policy rule ϕ^0 was used. The error term $\mathbf{U}_s$ includes all shocks that are not included in $\boldsymbol{\varepsilon}_{a,s}$ as well as the future errors $\mathbf{D}_s^0 - \mathbb{E}_s\mathbf{D}_s^0$. By construction, since all structural shocks are assumed to be uncorrelated we have that $\mathbb{E}(\boldsymbol{\varepsilon}_{a,s}\mathbf{U}_s') = 0$.

Equation (26) can be viewed as a set of stacked local projections (e.g. Jordà, 2005). The key difficulty for estimating $\mathcal{R}_{a,H}^0$ is that $\boldsymbol{\varepsilon}_{a,s}$ is not observed and therefore an identification strategy is needed. Prominent examples include using zero-, long-run, or inequality restrictions (e.g. Sims, 1980; Blanchard and Quah, 1989; Faust, 1998; Uhlig, 2005), or by using

[17] The Γ's are mainly of interest for evaluating shock specific policy counterfactuals as we will discuss in the next section (e.g. Barnichon and Mesters, 2023*a*).

past exogenous variations as instrumental variables (e.g. Mertens and Ravn, 2013; Stock and Watson, 2018). At the end, pending on preference, any of the identification strategies can be used and often multiple will be needed to identify all shocks of interest, see Ramey (2016) for a broader discussion. After the shocks have been identified conventional econometric methods can be used for estimation and inference (e.g. Stock and Watson, 2016; Kilian and Lütkepohl, 2017).

We note that while the recipe for estimating impulse responses is well known, there can exist disagreement about which impulse response estimates are correct (e.g. Ramey, 2016).[18] In the presence of such disagreement possible solutions include (i) evaluating the policy maker separately for the different impulses responses, highlighting any differences, or (ii) relaxing the non-overlapping identifying assumptions (when known) and working with the identified set of impulse responses, similar as when using sign restrictions.

5.2 Approximating oracle forecasts

The oracle forecasts $\mathbb{E}_t\mathbf{P}_t^0$, $\mathbb{E}_t\mathbf{Y}_t^0$ are defined in Lemma 1 in terms of $\mathbf{S}_t = (\mathbf{X}_{-t}', \Xi_t')'$ and the policy news shocks $\boldsymbol{\varepsilon}_t$. We will discuss two scenarios: (i) the researcher directly downloads the forecasts or (ii) the oracle forecasts need to be approximated by the researcher.

Downloading forecasts

The simplest yet not always feasible way in which a researcher can obtain a baseline forecast is to use the forecasts that are provided by the policy maker. Indeed, several macro policy makers make their forecasts for the policy objectives publicly available and these can then be directly used. Besides using the policy maker's forecasts, the researcher could use also use professional forecasts, such as those from the Survey of Professional Forecasters (SPF) or the Blue Chip forecasts.

It is important to stress that the baseline forecast should not condition on any future values or shock realizations. To make this clear, recall that the time t information set is given by $\mathcal{F}_t = \{\mathbf{X}_{-t}, \Xi_t, \boldsymbol{\varepsilon}_t\}$ and the oracle forecast that we require is $\mathbb{E}_t\mathbf{Y}_t^0 = \mathbb{E}(\mathbf{Y}_t|\mathcal{F}_t^0) = \mathbb{E}(\mathbf{Y}_t|\mathbf{X}_{-t}, \Xi_t, \boldsymbol{\varepsilon}_t^0)$ which is defined by model (21) under the rule ϕ^0. Each element of the information set is known at time t. This forecast should be distinguish from a conditional forecast where one also conditions on future realizations of shocks to pin down e.g. a specific *realized* policy path, i.e. $\mathbb{E}(\mathbf{Y}_t|\mathbf{X}_{-t}, \Xi_t, \boldsymbol{\varepsilon}_t^0, \boldsymbol{\varepsilon}_{t+1}^0, \boldsymbol{\varepsilon}_{t+2}^0, \ldots)$. The latter type of forecasts are often used for scenario analysis in VAR models (Waggoner and Zha, 1999; Antolín-Díaz, Petrella and Rubio-Ramírez, 2021) and are sometimes also published by policy makers.

[18] Indeed opinions about the appropriate identification strategy and estimation method vary, leading to a possibly different estimates.

Approximating forecasts

Next, we discuss some econometric methods that can be used to approximate the oracle forecasts $\mathbb{E}_t\mathbf{D}_t^0 = \mathbb{E}_t(y_t^{0'}, \ldots, y_{t+H}^{0'}, p_t^{0'}, \ldots, p_{t+H}^{0'})$. We can use Lemma 1 to define the equilibrium representation

$$\mathbb{E}_t\mathbf{D}_t^0 = \Gamma_d^0\mathbf{S}_t + \mathcal{R}_d^0\boldsymbol{\varepsilon}_t^0 \ , \tag{27}$$

where Γ_d^0 and $\mathcal{R}_d^0$ collect the needed rows from Γ_y^0, Γ_p^0 and $\mathcal{R}_y^0, \mathcal{R}_p^0$ in order to correctly define $\mathbb{E}_t\mathbf{D}_t^0$ using Lemma 1.

To approximate $\mathbb{E}_t\mathbf{D}_t^0$ we generally need to approximate the state of the economy $\mathbf{S}_t$ and the policy shocks $\boldsymbol{\varepsilon}_t$. In general, we postulate that the researcher approximates these terms by the (possibly large) vector of time-t observable variables $\mathbf{Z}_t$. Note that $\mathbf{Z}_t$ may include (a part of) the expected policy path $\mathbb{E}_t\mathbf{P}_t^0$ when it is observed to the researcher.

The best linear prediction for $\mathbf{D}_t^0$ in terms of $\mathbf{Z}_t$ can be obtained from the forecasting model over the periods $s \in \mathfrak{T}$.

$$\mathbf{D}_s^0 = \mathbf{B}^0\mathbf{Z}_s + \mathbf{V}_s \ , \qquad s \in \mathfrak{T} \ , \tag{28}$$

where $\mathbf{V}_s$ includes the future error $\mathbf{D}_s^0 - \mathbb{E}_t\mathbf{D}_s^0$ as well as the approximation error that stems from replacing $(\mathbf{S}_s, \boldsymbol{\varepsilon}_s)$ by $\mathbf{Z}_s$. The matrix $\mathbf{B}^0$ is defined such that it includes the best linear prediction coefficients, i.e. the ones that minimize the mean-squared-error, and the error $\mathbf{V}_t$ is orthogonal to $\mathbf{Z}_t$ by construction. In contrast, $\mathbf{V}_t$ is not orthogonal to the total time-t information set $\mathcal{F}_t$, such condition would require perfectly observing the state of the economy which seems a major assumption.[19]

Based on model (28) we can estimate the matrix $\mathbf{B}^0$. This matrix may be structured, e.g. sparse, banded etc, and different estimation methods can allow for shrinkage and penalization to improve the model fit. As examples we can think of: (i) penalized regression methods such as Lasso, Ridge and so on, see Kock, Medeiros and Vasconcelos (2020) for implementation details for time series regressions, (ii) a factor augmented regression where $\mathbf{Z}_s$ form a set of common factors and standard regression methods are used to estimate $\mathbf{B}^0$ (e.g. Stock and Watson, 2002; Bai and Ng, 2006), or (iii) the usage of large (Bayesian) vector autoregressive models (e.g. Banbura, Giannone and Reichlin, 2010).

In general, we denote the estimated model parameters by $\widehat{\mathbf{B}}^0$. The resulting forecasts for time period t are given by

$$\widehat{\mathbf{D}}_t^0 = \widehat{\mathbf{B}}^0\mathbf{Z}_t \ . \tag{29}$$

Clearly the objective is to try to make $\widehat{\mathbf{D}}_t^0$ as close as possible to $\mathbb{E}_t\mathbf{D}_t^0$. At the same time,

[19]As such we note that $\mathbf{B}^0$ does not have any causal interpretation. Indeed, in contrast to the impulse responses $\Gamma_{..}^0$ and $\mathcal{R}_{..}^0$, these coefficients merely capture the correlation between the observable predictors and the outcome variables of interest.

it is well known that macro forecasting is hard and in practice mistakes will be made.

5.3 Uncertainty

To evaluate policy decisions we generally need the joint distribution of the sufficient statistics to take into account uncertainty. Since, we allow for different forecasting and impulse response estimators we do not give a detailed treatment for any specific choices. Instead we provide a high level overview for how such joint uncertainty measures can be constructed.

We start by recalling our generic forecasting and impulse response equations:

$$\mathbf{D}_s^0 = \mathbf{B}^0 \mathbf{Z}_s + \mathbf{V}_s \qquad \text{and} \qquad \mathbf{D}_s^0 = \mathcal{R}_{a,H}^0 \boldsymbol{\varepsilon}_{a,s} + \mathbf{U}_s \qquad \text{for all} \quad s \in \mathfrak{T} \ .$$

We are interested in the joint uncertainty around $\widehat{\mathbf{D}}_t^0 - \mathbb{E}_t \mathbf{D}_t^0$ — the forecast mis-specification error — and $\widehat{\mathcal{R}}_{a,H}^0 - \mathcal{R}_{a,H}^0$ — the impulse response estimation error. Note that the impulse response estimates $\widehat{\mathcal{R}}_{a,H}^0$ can correspond to OLS, IV or any other desired estimates.

The joint distribution is denoted by

$$\begin{pmatrix} \widehat{\mathbf{D}}_t^0 - \mathbb{E}_t \mathbf{D}_t^0 \\ \widehat{\mathcal{R}}_{a,H}^0 - \mathcal{R}_{a,H}^0 \end{pmatrix} \overset{a}{\sim} \widehat{F} \ . \tag{30}$$

The distribution $\widehat{F}$ is important in our work as this is the distribution from which we will simulate to compute the distribution of the policy evaluation statistics.

For the forecast mis-specification error note that $\widehat{\mathbf{D}}_t^0 - \mathbb{E}_t \mathbf{D}_t^0 = (\widehat{\mathbf{B}}^0 - \mathbf{B}^0)\mathbf{Z}_s - \mathbb{E}_t \mathbf{V}_t$ which disentangles the error into parameter estimation error $\widehat{\mathbf{B}}^0 - \mathbf{B}^0$ and model mis-specification error $\mathbb{E}_t \mathbf{V}_t = \mathbb{E}(\mathbf{V}_t | \mathcal{F}_t)$. While the former is often easy to account for, handling $\mathbb{E}_t \mathbf{V}_t$ is harder as this is the part of the forecasting model that could have been predicted by the information set $\mathcal{F}_t$, but the researcher did not measure the entire $\mathcal{F}_t$ and as such this component ended up in the error term.

An easy approach for obtaining the distribution of $\widehat{\mathbf{D}}_t^0 - \mathbb{E}_t \mathbf{D}_t^0$ by (i) upper-bounding the variance of $\widehat{\mathbf{D}}_t^0 - \mathbb{E}_t \mathbf{D}_t^0$ by an estimate for the mean squared error of the forecast errors $\widehat{\mathbf{D}}_t^0 - \mathbf{D}_t^0$ in combination with a normality assumption (Scheffe, 1953). The mean squared forecast errors can be estimated using different strategies, most notably pseudo-out-of-sample forecasting see (Stock and Watson, 2019, Section 15.5) for a general discussion. We note that the variance of the forecast errors will upper-bound the variance of the forecast mis-specification error, because forecast errors mix two sources of uncertainty: (i) model mis-specification *and* (ii) future uncertainty. The latter is not needed when evaluating policy decisions as it is outside of the control of the policy maker.

For the impulse response error $\widehat{\mathcal{R}}_{a,H}^0 - \mathcal{R}_{a,H}^0$ conventional methods can be used to approximate the distribution, e.g. asymptotic theory, bootstrap or Bayesian methods (e.g. Kilian and Lütkepohl, 2017).

Finally, note that if the same reduced form model is used for impulse response estimation and forecasting the joint uncertainty can be taking into account. This is easiest when using bootstrap or Bayesian methods. However, when the researcher uses external forecasts, such as those obtained from the policy maker, we will typically have to make the additional assumption that these forecasts errors are independent from the impulse response estimation errors as we will not have any method for recovering the joint distribution.

5.4 Simulating counterfactual policies

We combine the ingredients discussed above and provide an algorithm for approximating policy counterfactuals. Specifically, given the alternative policy path $\mathbb{E}_t\mathbf{P}_t^1$ we use Corollary 1, or Theorem 1, to learn the counterfactual macro outcomes from the sufficient statistics under ϕ^0: the forecasts and impulse responses. When all impulse responses to policy shocks are recovered the algorithm naturally provides a way to conduct inference on exact policy counterfactuals.

Counterfactual computation

0 Obtain the estimates $\widehat{\mathcal{R}}^0_{a,y}, \widehat{\mathcal{R}}^0_{a,p}$, the forecasts $\widehat{\mathbf{Y}}_t, \widehat{\mathbf{P}}_t$ and the distribution $\widehat{F}$

1 Compute by simulation

$$\begin{aligned}\boldsymbol{\delta}^j_{a,t} &= (\mathcal{R}^{j'}_{a,p}\mathcal{R}^j_{a,p})^{-1}\mathcal{R}^{j'}_{a,p}(\mathbb{E}_t\mathbf{P}^1_t - \widehat{\mathbf{P}}^j_t)\\ \widehat{\mathbf{Y}}^{j,a}_t &= \widehat{\mathbf{Y}}^j_t + \mathcal{R}^j_{a,y}\boldsymbol{\delta}^j_{a,t}\end{aligned}$$

where the impulse responses $(\mathcal{R}^j_{a,p}, \mathcal{R}^j_{a,y})$ and forecasts $(\widehat{\mathbf{P}}^j_t, \widehat{\mathbf{Y}}^j_t)$ are simulated from $\widehat{F}$ for $j = 1, \ldots, S_d$.

2 Report the mean counterfactual paths together with the confidence bands obtained from the simulated distributions.

6 Time-t policy evaluation

In this section we show how the representation results from Section 4 can be used to characterize optimal policy paths and the distance to minimum loss at any given point in time. Moreover, we show how to correct such statistics for dynamic inconsistency stemming from decisions in previous period.

We consider a general quadratic loss function

$$\mathcal{L}_t = \frac{1}{2}\mathbb{E}_t\mathbf{Y}_t'\mathcal{W}\mathbf{Y}_t \ , \tag{31}$$

where $\mathcal{W}$ is a diagonal weighting matrix that allows to place more or less importance on different variables and horizons.[20] We note that using a quadratic loss function is convenient but not strictly necessary. The supplementary material of Barnichon and Mesters (2023*b*) works out a sufficient statistics approach for general convex loss functions. Evaluating decisions based on such more general loss functions does require different sufficient statistics; notably the conditional mean forecast is no longer sufficient for describing the state of economy as typically the entire forecast distribution is needed.

For the quadratic loss function (31) and underlying model (21) we define the optimal allocation as the paths $\mathbb{E}_t\mathbf{Y}_t$ and $\mathbb{E}_t\mathbf{P}_t$ that minimize the loss, i.e.,

$$\min_{\mathbf{Y}_t,\mathbf{P}_t} \mathcal{L}_t \qquad \text{s.t.} \qquad (21)\ . \tag{32}$$

We often refer to the problem (32) as the planner's problem and denote the solution(s) to this problem for the policy path by $\mathbb{E}_t\mathbf{P}_t^{\text{opt}}$ and we denote by $\mathcal{L}_t^{\text{opt}}$ the minimum loss that can be achieved under $\mathbb{E}_t\mathbf{P}_t^{\text{opt}}$.

For clarity of exposition, we make the following simplifying assumption.

Assumption 2. *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 (32). While there could be interesting discriminating aspects among different optimal policies, we retain the uniqueness assumption to avoid notational clutter.

6.1 Optimal policy perturbations

To compute or approximate the optimal policy using sufficient statistics we make use of Lemma 2 which shows that adjusting the policy rule by $\boldsymbol{\delta}_t$ changes the equilibrium outcome to $\mathbb{E}_t\mathbf{Y}_t(\boldsymbol{\delta}_t) = \mathbb{E}_t\mathbf{Y}_t^0 + \mathcal{R}_y^0\boldsymbol{\delta}_t$. To use this result let $\boldsymbol{\delta}_{a,t}$ correspond to the subset of $\boldsymbol{\delta}_t$ for which the corresponding policy shocks can be identified, i.e. $\mathcal{R}_y^0\boldsymbol{\delta}_t = \mathcal{R}_{a,y}^0\boldsymbol{\delta}_{a,t} + \mathcal{R}_{-a,y}^0\boldsymbol{\delta}_{-a,t}$, where the subset of impulse responses $\mathcal{R}_{a,y}^0$ can be identified by the researcher. A special case arises when all shocks can be identified and then we consider $\boldsymbol{\delta}_{a,t} = \boldsymbol{\delta}_t$.

[20] We do not take a stand on the origins of the loss function. As such $\mathcal{L}_t$ can be any desired loss function that the policy maker or researcher wants to minimize. This allows for micro founded, e.g. welfare maximizing, loss functions, but also allows for loss functions that simply correspond to mandates imposed on policy makers. For instance, Bernanke (2015) argued that the Fed should consider inflation and unemployment as its target variables and place equal weight on both objectives over a median term horizon, e.g. five years.

We compute the $\boldsymbol{\delta}_{a,t}$-adjustment that minimizes the loss function. We call this specific $\boldsymbol{\delta}_{a,t}$ the (subset) *Optimal Policy Perturbation* (OPP). Specifically, the idea of the OPP is to find the "best" adjustment $\boldsymbol{\delta}_{a,t}$ to the baseline rule ϕ^0 in order to minimize the loss, that is

$$\boldsymbol{\delta}^*_{a,t} = \underset{\boldsymbol{\delta}_{a,t}}{\text{argmin}} \ \mathcal{L}_t(\boldsymbol{\delta}_t) \qquad \text{s.t.} \qquad \mathbb{E}_t \mathbf{Y}_t(\boldsymbol{\delta}_t) = \mathbb{E}_t \mathbf{Y}^0_t + \mathcal{R}^0_{a,y}\boldsymbol{\delta}_{a,t} + \mathcal{R}^0_{-a,y}\boldsymbol{\delta}_{-a,t} \ , \tag{33}$$

where $\mathcal{L}_t(\boldsymbol{\delta}_t) = \frac{1}{2}\mathbb{E}_t \mathbf{Y}_t(\boldsymbol{\delta}_t)'\mathcal{W}\mathbf{Y}_t(\boldsymbol{\delta}_t)$ is the loss function as a function of $\boldsymbol{\delta}_t$. It is easy to see that this adjusted policy problem is linear-quadratic and hence it has a closed form solution given by[21]

$$\boldsymbol{\delta}^*_{a,t} = -(\mathcal{R}^{0'}_{a,y}\mathcal{W}\mathcal{R}^0_{a,y})^{-1}\mathcal{R}^{0'}_{a,y}\mathcal{W}\mathbb{E}_t\mathbf{Y}^0_t \ . \tag{34}$$

We can now state the key properties of the OPP, see also Barnichon and Mesters (2023*b*).

Proposition 1. *Given the generic model* (21) *and the augmented policy rule* (25), ϕ^0 *implying a unique equilibrium, we have under Assumption 2 if all policy shocks can be identified, i.e.* $\boldsymbol{\delta}^*_{a,t} = \boldsymbol{\delta}^*_t$, *that*

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

   In contrast, if only a strict subset of policy shocks can be identified

3. $\boldsymbol{\delta}^*_{a,t} \neq 0 \qquad \Longrightarrow \qquad \mathbb{E}_t\mathbf{P}^0_t \neq \mathbb{E}_t\mathbf{P}^{\text{opt}}_t$
4. $\mathcal{L}_t(\boldsymbol{\delta}^*_{a,t}, \mathbf{0}) \leq \mathcal{L}_t(\mathbf{0}, \mathbf{0})$, *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* $\mathbb{E}_t\mathbf{P}^0_t$.

The first and second part consider the case where all policy shocks can be identified. Here 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 $\mathbb{E}_t\mathbf{P}^0_t$ is not optimal. Second, we can use the OPP to *construct* the optimal policy path $\mathbb{E}_t\mathbf{P}^{\text{opt}}_t$ from some arbitrary baseline policy choice that implies a unique equilibrium.

The third and fourth parts of the proposition consider the case where a strict subset of policy shocks can be identified. Here it holds that if the OPP statistic $\boldsymbol{\delta}^*_{a,t}$ is non-zero the policy $\mathbb{E}_t\mathbf{P}^0_t$ is non-optimal. Moreover, adjusting the baseline policy with the OPP will improve the baseline policy path, though it will generally not give the optimal path $\mathbb{E}_t\mathbf{P}^{\text{opt}}_t$.

[21] It is worth pointing out that throughout we assume that the inverse $(\mathcal{R}^{0'}_{a,y}\mathcal{W}\mathcal{R}^0_{a,y})^{-1}$ 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.

In other words, the OPP allows to compute the best policy path given the sufficient statistics available. Barnichon and Mesters (2023*b*) provide more discussion regarding the properties of the OPP statistics and the associated adjustments.

More generally, while our review focuses on evaluating policy makers, Proposition 1 reveals that the sufficient statistics approach is also able to provide robust policy advice. Indeed, the adjusted policy path is in fact a forecast targeting rule — a prescription for how policy could be set in an optimal way given the loss functions (e.g. Svensson and Woodford, 2005) —. Specifically,

$$\mathbb{E}_t \mathbf{P}_t^{\text{opt}} = \mathbb{E}_t \mathbf{P}_t^0 - \mathcal{R}_p^0 (\mathcal{R}_y^{0\prime} \mathcal{W} \mathcal{R}_y^0)^{-1} \mathcal{R}_y^{0\prime} \mathcal{W} \mathbb{E}_t \mathbf{Y}_t^0 \ ,$$

which highlights that the impulse responses to policy shocks (together with the oracle forecasts) are sufficient statistics for describing an optimal time-t targeting rule (McKay and Wolf, 2023; Barnichon and Mesters, 2023*b*).

Correcting for dynamic inconsistency

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. This ignores that in most macro policy settings policy decisions are made repeatedly. 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). To adjust for this Barnichon and Mesters (2023*b*) introduce a simple correction to the OPP statistic that eliminates dynamic inconsistency.

Specifically, a *time consistent* OPP statistic can be defined as

$$\boldsymbol{\delta}_{a,t}^{\tau *} = \boldsymbol{\delta}_{a,t}^{*} + \Delta \mathcal{D}_a^0 \mathbb{E}_{t-1} \mathbf{Y}_{t-1}^0 \ , \tag{35}$$

where the original OPP is adjusted with a "time inconsistency correction factor" given by $\Delta \mathcal{D}^0 \mathbb{E}_{t-1} \mathbf{Y}_{t-1}^0$ where $\Delta \mathcal{D}_a^0 = \left[\mathcal{D}_{a,1}^0 - \mathbf{0}, \mathcal{D}_{a,2}^0 - \mathcal{D}_{a,1}^0, ...\right]$ is a "pseudo-difference" map with $\mathcal{D}_{a,i}^0$ the ith $d_a \times M_y$ block of $\mathcal{D}_a^0 = -(\mathcal{R}_{a,y}^{0\prime} \mathcal{W} \mathcal{R}_{a,y}^0)^{-1} \mathcal{R}_{a,y}^{0\prime} \mathcal{W}$, i.e. $\mathcal{D}_a^0 = [\mathcal{D}_{a,1}^0, \mathcal{D}_{a,2}^0, \ldots]$. Importantly, the correction factor is again entirely determined by our two sufficient statistics, so that no extra information is necessary to implement a time-consistent OPP.

Intuitively, the correction factor removes any updates in the original OPP that stem from shifting preferences between time $t-1$ and time t. Indeed the difference map captures the difference in weights placed on the $t-1$ objectives $\mathbb{E}_{t-1} \mathbf{Y}_{t-1}^0$ when considering the $\mathcal{L}_{t-1}$ loss and the $\mathcal{L}_t$. A time consistent policies are only allowed to change because of (i) previous mistakes and (ii) changes in the information set; indeed it is easy to show that

$$\boldsymbol{\delta}_{a,t}^{\tau*} = \boldsymbol{\delta}_{a,t-1}^{*} + \mathcal{D}_a^0 \Delta \mathbb{E}_t \mathbf{Y}_t^0 \ ,$$

which writes the time consistent OPP as a function of the past OPP (previous mistakes) and the information update $\Delta \mathbb{E}_t \mathbf{Y}_t^0 = \mathbb{E}_t \mathbf{Y}_t^0 - \mathbb{E}_{t-1} \mathbf{Y}_t^0$. Clearly, for any sequence of periods such corrections can be repeatedly applied to ensure that no dynamic inconsistencies arise among any periods.

It is useful to note that the time-consistent OPP does not lead us to the optimal policy as it was defined in (32). Instead, the planner's problem that has $\boldsymbol{\delta}_{a,t}^{\tau*}$ as optimal adjustment includes a time-consistency restriction:

$$\min_{\mathbf{Y}_t, \mathbf{P}_t} \mathcal{L}_t \qquad \text{s.t.} \qquad (21) \quad \text{and} \quad \mathcal{R}_y^{0\prime} \mathcal{W} \mathbb{E}_{t-1} \mathbf{Y}_t - \mathcal{R}_y^{0\prime} \mathcal{W} \mathbb{E}_{t-1} \mathbf{Y}_{t-1}^0 = 0 \ . \tag{36}$$

The constraint imposes that the first order conditions of optimization problem at time t evaluated given the time $t-1$ information (i.e. $\mathcal{R}_y^{0\prime} \mathcal{W} \mathbb{E}_{t-1} \mathbf{Y}_t$) are set equal to the first order conditions from the time $t-1$ policy problem $\mathcal{R}_y^{0\prime} \mathcal{W} \mathbb{E}_{t-1} \mathbf{Y}_{t-1}^0$. Respecting this constraint imposes that all changes in the OPPs between times $t-1$ and t are due to changes in the information set, i.e. moving from $\mathbb{E}_{t-1}$ to $\mathbb{E}_t$. We denote the minimal time consistent loss as defined by (36) by $\mathcal{L}_t^{\tau,\text{opt}}$.

6.2 Time-t distance to minimum loss

The OPP statistic tells us how far the policy maker is from the optimal policy. Clearly, this is one possible metric for evaluating and comparing policy makers. However, we often want to evaluate policy makers or policy institutions based on the loss that could have been avoided by choosing a more optimal policy action. For this we define the Distance to Minimum Loss statistic for time t (DML-t) as

$$\Delta_t = \mathcal{L}_t^0 - \mathcal{L}_t^{\text{opt}} \ , \tag{37}$$

where $\mathcal{L}_t^0$ is the loss under the baseline policy choice ϕ^0 and $\mathcal{L}_t^{\text{opt}}$ is the loss under the optimal policy as defined in (32). If the entire optimal policy perturbation can be recovered from the sufficient statistics, i.e. if all policy news shocks can be identified, we can compute the DML-t using

$$\Delta_t = \boldsymbol{\delta}_t^{\tau*\prime} \mathcal{R}_y^{0\prime} \mathcal{W} \mathcal{R}_y^0 \boldsymbol{\delta}_t^{\tau*} \ , \tag{38}$$

which expresses the DML-t in terms of the OPP and impulse responses under the baseline policy choice.

In practice, typically not all policy news shocks can be identified and we instead compute the distance to minimum loss that we can empirical identify. That is

$$\Delta_{a,t} = \mathcal{L}_t^0 - \mathcal{L}_t(\boldsymbol{\delta}_{a,t}^*, 0) \ ,$$

where $\mathcal{L}_t(\boldsymbol{\delta}_t) = \frac{1}{2}\mathbb{E}_t \mathbf{Y}_t(\boldsymbol{\delta}_t)'\mathcal{W}\mathbf{Y}_t(\boldsymbol{\delta}_t)$ is the loss function which is here evaluated at the subset optimal policy choice $\boldsymbol{\delta}_{a,t}^*$. The difference is that $\mathcal{L}_t(\boldsymbol{\delta}_{a,t}^*, 0)$ is not the exact optimal policy, but the best approximation thereof that can be obtained using the empirical evidence. This subset of the distance to minimum loss can be computed from

$$\Delta_{a,t} = \boldsymbol{\delta}_{a,t}^{\tau *\prime} \mathcal{R}_{a,y}^{0\prime} \mathcal{W} \mathcal{R}_{a,y}^0 \boldsymbol{\delta}_{a,t}^{\tau *} \ . \tag{39}$$

In the case where the optimal policy is defined to be time consistent as in (36) we have that

$$\begin{aligned}\Delta_{a,t}^{\tau} =& \mathcal{L}_t^0 - \mathcal{L}_t^{\tau,\text{opt}} \\ =& -\boldsymbol{\delta}_{a,t}^{\tau *\prime} \mathcal{R}_y^{0\prime} \mathcal{W} \mathcal{R}_y^0 \boldsymbol{\delta}_{a,t}^{\tau *} - 2\boldsymbol{\delta}_{a,t}^{\tau *\prime} \mathcal{R}_y^{0\prime} \mathcal{W} \mathbb{E}_t \mathbf{Y}_t^0 \ ,\end{aligned}$$

which incorporates the corrections for dynamic inconsistency. As the definition of the time consistent OPP includes the correction we need to take this into account when computing the time consistent DML-t.

6.3 Implementation OPP and DML statistics

Next, we formalize the implementation of the OPP and DML statistics. Broadly speaking we can use any of the forecasting and impulse response estimation methods discussed in Section 5 to obtain the approximating distribution $\widehat{F}$ of $(\widehat{\mathcal{R}}_{a,y}^0 - \mathcal{R}_{a,y}^0, \widehat{\mathcal{R}}_{a,p}^0 - \mathcal{R}_{a,p}^0, \widehat{\mathbf{Y}}_t - \mathbb{E}_t \mathbf{Y}_t^0, \widehat{\mathbf{P}}_t - \mathbb{E}_t \mathbf{P}_t^0)$. Subsequently, we can simulate from this distribution to obtain the distribution of the OPP adjustment and compute the other statistics.

The following algorithm describes the procedure in detail.

OPP and DML computation

0 Obtain the estimates $\widehat{\mathcal{R}}^0_{a,y}, \widehat{\mathcal{R}}^0_{a,p}$, the forecasts $\widehat{\mathbf{Y}}_t, \widehat{\mathbf{P}}_t$ and the distribution $\widehat{F}$

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

$$\boldsymbol{\delta}^j_{a,t} = -(\mathcal{R}^{j\prime}_{a,y}\mathcal{W}\mathcal{R}^j_{a,y})^{-1}\mathcal{R}^{j\prime}_{a,y}\mathcal{W}\widehat{\mathbf{Y}}^j_t$$
$$\Delta^j_{a,t} = \boldsymbol{\delta}^{j\prime}_{a,t}\mathcal{R}^{j\prime}_{a,y}\mathcal{W}\mathcal{R}^j_{a,y}\boldsymbol{\delta}^j_{a,t}$$

or

$$\boldsymbol{\delta}^{\tau,j}_{a,t} = \boldsymbol{\delta}^j_{a,t} - \Delta\mathcal{D}^j_a\widehat{\mathbf{Y}}^j_{t-1}$$
$$\Delta^{\tau,j}_{a,t} = -\boldsymbol{\delta}^{\tau,j\prime}_{a,t}\mathcal{R}^{j\prime}_{a,y}\mathcal{W}\mathcal{R}^j_{a,y}\boldsymbol{\delta}^{\tau,j}_{a,t} - 2\boldsymbol{\delta}^{\tau,j\prime}_{a,t}\mathcal{R}^{j\prime}_{a,y}\mathcal{W}\widehat{\mathbf{Y}}^j_t$$

where the impulse responses $(\mathcal{R}^j_{a,p}, \mathcal{R}^j_{a,y})$ and forecasts $(\widehat{\mathbf{P}}^j_t, \widehat{\mathbf{Y}}^j_t)$ are simulated from $\widehat{F}$ for $j = 1, \ldots, S_d$.

2 For each draw compute the adjusted paths

$$\widehat{\mathbf{W}}^{a,j}_t = \widehat{\mathbf{W}}^j_t + \mathcal{R}^j_{a,w}\boldsymbol{\delta}^j_{a,t}\ , \qquad \mathbf{W} = \mathbf{P}, \mathbf{Y}$$

or

$$\widehat{\mathbf{W}}^{\tau a,j}_t = \widehat{\mathbf{W}}^j_t + \mathcal{R}^j_{a,w}\boldsymbol{\delta}^{\tau,j}_{a,t}\ , \qquad \mathbf{W} = \mathbf{P}, \mathbf{Y}$$

3 Report the mean OPP and DML statistics, and the adjusted policy paths together with the confidence bands obtained from the simulated distributions.

7 Term policy evaluation

We consider the evaluation of a policy maker over her term. The policy maker's problem is to choose a policy rule $\phi = \{\mathcal{A}_{pp}, \mathcal{A}_{py}, \mathcal{B}_{p\xi}, \mathcal{B}_{p\varepsilon}\}$ that minimizes the unconditional loss $\mathcal{L}(\phi, \theta) = \mathbb{E}\mathcal{L}_t$. That is choose a policy rule from the set

$$\Phi^{\text{opt}} = \{\phi \in \Phi : \phi \in \underset{\phi}{\text{argmin}}\, \mathcal{L}(\phi, \theta) \quad \text{s.t.} \quad (21) \text{ and } (22)\}\ . \tag{40}$$

Broadly speaking, choosing an optimal rule for an unconditional objective function corresponds to the timeless perspective for optimal policy making, see Woodford (2003) and

Giannoni and Woodford (2004) for more discussion.

7.1 Distance to minimum loss

In this context, we measure the policy makers performance by considering the (unconditional) distance to minimum loss

$$\Delta = \mathcal{L}^0 - \mathcal{L}^{\text{opt}} \ , \tag{41}$$

where $\mathcal{L}^0 = \mathcal{L}(\phi^0, \theta)$ and $\mathcal{L}^{\text{opt}} = \mathcal{L}(\phi^{\text{opt}}, \theta)$ with $\phi^{\text{opt}} \in \Phi^{\text{opt}}$. This is the difference in loss that results from the choosing a possibly sub-optimal reaction function ϕ^0. In contrast, the time-t distance to minimum loss in (37) is a function of the time-t information set and mixes policy mistakes $\boldsymbol{\varepsilon}_t$ with a sub-optimal policy rule.

To characterize the distance to optimality in terms of sufficient statistics we consider the specific rule adjustment

$$\boldsymbol{\delta}^b_{a,t} = \mathcal{T}_{\xi,b}\boldsymbol{\Xi}_{b,t} + \mathcal{T}_{\varepsilon,a}\boldsymbol{\varepsilon}_{a,t} \ ,$$

where $\mathcal{T}_{\xi,b}$ is an adjustment to the reaction to non-policy shocks $\boldsymbol{\Xi}_{b,t}$ and $\mathcal{T}_{\varepsilon,a}$ is an adjustment to the reaction to policy shocks $\boldsymbol{\varepsilon}_{a,t}$, see footnote 16. We note that for computing the DML it is not necessary that the shocks $\boldsymbol{\Xi}_{b,t}$ and $\boldsymbol{\varepsilon}_{a,t}$ are structural, they may be reduced form shocks or forecast errors. The only consequence is that no shock specific evaluation will be possible, but this may not be the objective, we may be interested in an overall evaluation (see Corollary 2 below for details). To avoid introducing additional reduced form impulse responses we continue with the specific rule adjustment $\boldsymbol{\delta}^b_{a,t}$.

Under this adjustment Lemma 2 gives

$$\mathbb{E}_t\mathbf{Y}_t(\boldsymbol{\delta}^b_{a,t}) = (\Gamma^0_{y,\xi_b} + \mathcal{R}^0_{a,y}\mathcal{T}_{\xi,b})\boldsymbol{\Xi}_{b,t} + (\mathcal{R}^0_{a,y} + \mathcal{R}^0_{a,y}\mathcal{T}_{\varepsilon,a})\boldsymbol{\varepsilon}_{a,t} + N_t \ ,$$

where N_t includes all shocks that are not adjusted by the perturbation[22] and Γ^0_{y,ξ_b} are the impulse responses of $\mathbf{Y}_t$ to shocks $\boldsymbol{\Xi}_{b,t}$. The result shows that the $\mathcal{T}_{ab} = [\mathcal{T}_{\xi,b}, \mathcal{T}_{\varepsilon,a}]$ adjustments to the reaction coefficients change the impulse responses from Γ^0_{y,ξ_b} to $\Gamma^0_{y,\xi_b} + \mathcal{R}^0_{a,y}\mathcal{T}_{\xi,b}$ and from $\mathcal{R}^0_{a,y}$ to $\mathcal{R}^0_{a,y} + \mathcal{R}^0_{a,y}\mathcal{T}_{\varepsilon,a}$. Crucially the adjusted impulse responses are a function of the impulse responses under ϕ^0, i.e., we can compute the new impulse responses using the methods from Section 5.

From these "law of motions" for the impulse responses, we can compute the *optimal* $\boldsymbol{\delta}^b_{a,t}$ adjustment to the policy rule, i.e. the $\mathcal{T}_{ab}$ adjustments that minimize the loss function $\mathcal{L}(\phi, \theta)$. When all shocks can be identified, this allows to compute the optimal loss $\mathcal{L}^{\text{opt}}$ and thus the distance to minimum loss Δ. The following proposition summarizes the results.

[22]Formally, $N_t = \Gamma^0_{y,x}\mathbf{X}_{-t} + \Gamma^0_{y,\xi_{-b}}\boldsymbol{\Xi}_{-b,t} + \mathcal{R}^0_{-a,y}\boldsymbol{\varepsilon}_{-a,t}$.

Proposition 2. *Given the generic model* (21) *and the augmented policy rule* (25), ϕ^0 *implying a unique equilibrium, we have under Assumption 2 if all policy and non-policy shocks can be identified we have that*

1. $\Delta = \Delta_\xi + \Delta_\varepsilon$ *with*

$$\Delta_\xi = \text{Tr}\left(\Gamma^{0\prime}_{y,\xi}\mathcal{W}\mathcal{R}^0_y(\mathcal{R}^{0\prime}_y\mathcal{W}\mathcal{R}^0_y)^{-1}\mathcal{R}^{0\prime}_y\mathcal{W}\Gamma^0_{y,\xi}\right) \quad \textit{and} \quad \Delta_\varepsilon = \text{Tr}\left(\mathcal{R}^{0\prime}_y\mathcal{W}\mathcal{R}^0_y\right)$$

If only a strict subset of policy and non-policy shocks can be identified we have

2. $\Delta_{ab} = \mathcal{L}^0 - \min_{\mathcal{T}_{ab}} \frac{1}{2}\mathbb{E}(\mathbf{Y}_t(\boldsymbol{\delta}^b_{a,t})'\mathcal{W}\mathbf{Y}_t(\boldsymbol{\delta}^b_{a,t})) = \Delta_{\xi,ab} + \Delta_{\varepsilon,aa}$ *with*

$$\Delta_{\xi,ab} = \text{Tr}\left(\Gamma^{0\prime}_{y,\xi_b}\mathcal{W}\mathcal{R}^0_{a,y}(\mathcal{R}^{0\prime}_{a,y}\mathcal{W}\mathcal{R}^0_{a,y})^{-1}\mathcal{R}^{0\prime}_{a,y}\mathcal{W}\Gamma^0_{y,\xi_b}\right) \quad \textit{and} \quad \Delta_{\varepsilon,aa} = \text{Tr}\left(\mathcal{R}^{0\prime}_{a,y}\mathcal{W}\mathcal{R}^0_{a,y}\right) .$$

The first part of proposition 2 makes an important point: a policy maker follows an optimal rule if and only if she responds optimally to each new set of shocks that she faces during her term. This result implies that we can characterize the distance to minimum loss using sufficient macro statistics, and is at the core of the sufficient statistics approach to policy evaluation.

The second part of proposition 2 shows that if only a subset of shocks can be identified we can recover Δ_{ab}; a part of the total distance to minimum loss Δ. A proof is given in Barnichon and Mesters (2023*a*) who also provide bounds on the share of Δ that is captured by Δ_{ab}.

Moreover, the proposition shows that we can decompose the distance to minimum loss into different interpretable components. First, $\Delta_{\xi,ab}$ captures the sub-optimal reaction to the non-policy shocks. We can write

$$\Delta_{\xi,ab} = \sum_{j \in b} \mathcal{T}'_{a,j}\mathcal{R}^{0\prime}_{a,y}\mathcal{W}\mathcal{R}^0_{a,y}\mathcal{T}_{a,j} \quad \text{with} \quad \mathcal{T}_{a,j} = -(\mathcal{R}^{0\prime}_{a,y}\mathcal{W}\mathcal{R}^0_{a,y})^{-1}\mathcal{R}^{0\prime}_{a,y}\mathcal{W}\Gamma^0_{y,j} ,$$

which shows that each increment $\mathcal{T}'_{a,j}\mathcal{R}^{0\prime}_{a,y}\mathcal{W}\mathcal{R}^0_{a,y}\mathcal{T}_{a,j}$ captures the loss that is due to the non-optimal response to shocks of type $\boldsymbol{\Xi}_{b_j,t}$ when setting the policy instrument corresponding to the shocks $\boldsymbol{\varepsilon}_{a,t}$. The statistic $\mathcal{T}_{a,j}$ is the Optimal Reaction Function (ORA) statistic that was introduced in Barnichon and Mesters (2023*a*); it captures how the systematic policy response to the shock $\boldsymbol{\Xi}_{b_j,t}$ should be adjusted to minimize the loss function. The other component $\Delta_{\varepsilon,aa}$ captures the loss that could have been avoided by not making policy mistakes.

7.2 Recovering the DML from time-t statistics

The distance to minimum loss can also be computed using the time-t sufficient statistics. For this we define the first difference of the OPP statistic (34) as

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

Intuitively, $\boldsymbol{\delta}_{a,t}^{\Delta}$ captures the sub-optimal response of the policy maker to the forecast revisions.

Corollary 2. *Given the generic model* (21) *with Assumptions 1-2, given the augmented policy rule* (25)*, if all policy shocks can be identified we have that*

1. $\Delta = \mathbb{E}(\boldsymbol{\delta}_t^{\Delta\prime}\mathcal{R}_y^{0\prime}\mathcal{W}\mathcal{R}_y^0\boldsymbol{\delta}_t^{\Delta})$

If not all policy *shocks can be identified we have*

2. $\Delta_a = \mathbb{E}(\boldsymbol{\delta}_{a,t}^{\Delta\prime}\mathcal{R}_{a,y}^{0\prime}\mathcal{W}\mathcal{R}_{a,y}^0\boldsymbol{\delta}_{a,t}^{\Delta})$ *where* $\Delta_a = \sum_b \Delta_{ab}$

Proof. Note that

$$\begin{aligned}\Delta &= \mathbb{E}(\Delta\mathbb{E}_t\mathbf{Y}_t'\mathcal{W}\mathcal{R}_y^0(\mathcal{R}_y^{0\prime}\mathcal{W}\mathcal{R}_y)^{-1}\mathcal{R}_y^{0\prime}\mathcal{W}\Delta\mathbb{E}_t\mathbf{Y}_t) \\ &= \mathrm{Tr}(\mathcal{W}\mathcal{R}_y^0(\mathcal{R}_y^{0\prime}\mathcal{W}\mathcal{R}_y)^{-1}\mathcal{R}_y^{0\prime}\mathcal{W}\mathbb{E}(\Delta\mathbb{E}_t\mathbf{Y}_t\Delta\mathbb{E}_t\mathbf{Y}_t'))\end{aligned}$$

By Lemma 1 we have

$$\Delta\mathbb{E}_t\mathbf{Y}_t = \Gamma_{y,\xi}\boldsymbol{\Xi}_t + \mathcal{R}_y\boldsymbol{\varepsilon}_t \ ,$$

as $\mathbf{X}_{-1}$ is time $t-1$ measurable and hence cancels out. Using that are shocks are normalized to have mean zero and unit variance, we have that

$$\mathbb{E}(\Delta\mathbb{E}_t\mathbf{Y}_t\Delta\mathbb{E}_t\mathbf{Y}_t') = \mathcal{R}_y^{0\prime}\mathcal{R}_y^0 + \Gamma_{y,\xi}^{0\prime}\Gamma_{y,\xi}^0 \ .$$

Substituting this back into the first expression gives $\Delta = \Delta_\xi + \Delta_\varepsilon$ as they are defined in Proposition 2. The proof of the second part follows using similar steps. □

Note that the subset DML $\Delta_a = \sum_b \Delta_{ab}$ is the total distance to minimum loss *for the policy instruments* a, and it is equal to the sum of the subset DMLs Δ_{ab}, the distance to minimum loss for the a policy instruments' reaction to each non-policy shocks (indexed by b).

Corollary 2 is links the time-t perspective (Barnichon and Mesters, 2022) with the timeless perspective Barnichon and Mesters (2023*a*). This result can be found in a slightly different form in Appendix A2 of Caravello, McKay and Wolf (2024). More generally, it provides researchers with a simple way for estimating the DML Δ when the OPP statistics have been computed. Specifically, computing the DML simply amounts to taking a weighted sum of squares of the differenced OPP statistics over a policy maker's term. For a policy maker starting in t_0 and ending in $t_0 + J$ this would be

$$\widetilde{\Delta}_a = \frac{1}{J} \sum_{j=t_0}^{t_0+J} \boldsymbol{\delta}_{a,j}^{\Delta\prime} \mathcal{R}_{a,y}^{0\prime} \mathcal{W} \mathcal{R}_{a,y}^{0} \boldsymbol{\delta}_{a,j}^{\Delta} \ ,$$

where the simulation methods of Section 6 can be used to obtain estimates for $\boldsymbol{\delta}_{a,j}^{\Delta}$. This approach avoids the need to identify all structural non-policy shocks as was first shown in Caravello, McKay and Wolf (2024).

8 Learning policy maker's preferences

The previous two sections described methods for evaluating policy decisions given a set of objectives that were summarized in a loss function. Who defines the objectives was not formally imposed and different perspectives can be taken. For instance, if the objective is to evaluate a policy maker based on her preferences, the loss function should correspond to the interests of the policy maker, i.e. the variables and weights must align with those of the policy maker. However, this is not necessary when the loss function is merely viewed as an outcome metric that quantifies a research interest. For example, Blinder and Watson (2016) consider GDP growth to evaluate US presidents, while it conceivable that US presidents have other objectives, like inequality for example, focusing on growth remains an outcome of interest. It merely changes the perspective from 'evaluating the policy maker based on her objectives' to 'evaluating the policy maker based on the objectives of the researcher', both are interesting.

Moving forward, this discussion brings to light an alternative usage of the sufficient statistics approach, where the goal is to learn the objectives of the policy maker (Barnichon and Mesters, 2022). Indeed, postulating that policy makers aim to optimize *some* loss function, we can use the methods of the previous sections to search for the loss function that aligns best with the observed sequence of policy decisions, i.e. the loss function that gives the lowest average Distance to Minimum Loss (DML).

To sketch the approach, we assume that the policy maker uses a quadratic loss function that falls in the class (31). This starting point is not necessary as more general loss functions can be used to describe the class, but it avoids having to introduce more notation.[23] Further,

[23] See Killian and Manganelli (2008) for an example of an a-symmetric loss function, where preference parameters for up- and downside risk are also included.

we parametrize the weighting matrix as

$$\mathcal{W} = \mathcal{W}(\lambda) = \text{diag}(\beta \otimes \lambda) ,$$

where $\beta = (\beta_0, \beta_1, \ldots)'$ weighs the different horizons and $\lambda = (\lambda_1, \ldots, \lambda_{m_y})'$ the different macro objectives. For convenience we hold β fixed and search for the λ-weights that best describe the policy maker's preferences.

The weights that best rationalize the decisions of the policy maker minimize the DML:

$$\lambda^* = \underset{\lambda}{\text{argmin}}\, \Delta(\lambda) \qquad \Delta(\lambda) = \mathbb{E}(\boldsymbol{\delta}_t^{\Delta'}(\lambda)\mathcal{R}_y^{0'}\mathcal{W}(\lambda)\mathcal{R}_y^0\boldsymbol{\delta}_t^{\Delta}(\lambda)) , \tag{42}$$

where the differenced OPP statistic is also a function of λ:

$$\boldsymbol{\delta}_t^{\Delta}(\lambda) = -(\mathcal{R}_{a,y}^{0'}\mathcal{W}(\lambda)\mathcal{R}_{a,y})^{-1}\mathcal{R}_{a,y}^{0'}\mathcal{W}(\lambda)\Delta\mathbb{E}_t\mathbf{Y}_t .$$

An estimate for λ^* can be easily obtained by setting up a grid for λ and computing the DML for each value on the grid using the methods outlined above.

The resulting λ^* is the revealed preference of the policy maker, or the least favorable weight vector from the perspective of a researcher who wishes to reject that a policy was option. Interestingly, the idea of backing out the weights that are most consistent with an optimizing agent's objective function has, in a different framework, been considered by Hansen and Singleton (1982). Also, in public finance where interest is in, e.g., measuring preferences for income redistribution from observed tax plans, policy makers' preferences are often measured by finding the parameters of the loss function that best describe observed behavior (e.g., Bourguignon and Spadaro, 2012; Jacobs, Jongen and Zoutman, 2017; Hendren, 2020).[24,25]

9 Monetary policy in the Euro area

The European Central Bank (ECB) has been making policy decisions in the Euro area since the adoption of the euro in 1999. The first twenty years of ECB's history have been extensively reviewed in Hartman and Smets (2018). In particular, Hartman and Smets (2018) evaluate past ECB policy decisions through the lens of estimated Taylor rules capturing the ECB's reaction function. Throughout this section we will revisit some of their findings through the lens of the sufficient statistics approach.

[24] There the approach is referred to as the inverse optimal-tax method.

[25] Also, in the forecasting literature a similar idea has been used to rationalize forecasts (Elliott, Timmermann and Komunjer, 2005; Elliott, Komunjer and Timmermann, 2008).

We start by outlining the set-up. First, we consider the Euro area as the unit of analysis. This means that we consider aggregate Euro area variables as the variables in $\mathbf{Y}_t$ and the model that generated these variables is assumed to be of the form (21).[26] The policy instrument that we consider is the short term interest rate i_t, and the expected policy path is $\mathbb{E}_t\mathbf{P}_t = \mathbb{E}_t(i_t, i_{t+1}, \ldots)'$ and the policy shocks are the contemporaneous and news shocks collected in $\boldsymbol{\varepsilon}_t$. This view imply that we only considered policies of the ECB that affect the economy through the expected interest rate path, see also Eberly, Stock and Wright (2020).[27]

9.1 Estimating the sufficient statistics

Impulse responses to monetary shocks

To evaluate ECB policy decisions, we will rely on the impulse responses to a single policy shock. This implies that we only use one linear combination of the columns of $\mathcal{R}_y^0$ and $\mathcal{R}_p^0$ for evaluation. The impulse responses imply that our policy assessment will be focused on the short-end of the policy path; roughly over the next year. Our evaluation will be silent about the optimality of the medium- to longer-end of the policy path.

To identify the policy shock of interest we consider a mixed frequency Bayesian vector autoregressive model that includes (HICP, year-on-year, monthly), real GDP growth (quarter-on-quarter, quarterly), short term interest rate (EONIA rate extended with the Euro Short Term Rate (€STR, monthly), commodity price index (monthly) and the spread between the short term interest rate and the ten year German yield (monthly).[28] We identify the monetary policy shock of interest by ordering the short term interest rate last and using recursive zero restrictions (e.g. Kilian and Lütkepohl, 2017, Section 8.2).[29]

We estimate the VAR using data from January 2002 until December 2019. We impose a conventional Minnesota prior on the reduced form coefficients (e.g. Canova, 2007) and use $p = 24$ lags to avoid biases from short lag lengths. The results are shown in Figure 1 below.[30] We find that the effect on inflation is significantly negative and persistent for at least 10 quarters. In contrast, the effect on real GDP growth fades out after approximately four quarters.

[26] The alternative is to use the individual countries as the unit of analysis. See the recent evaluation of Einarsson (2024) based on sufficient macro statistics at the country level.

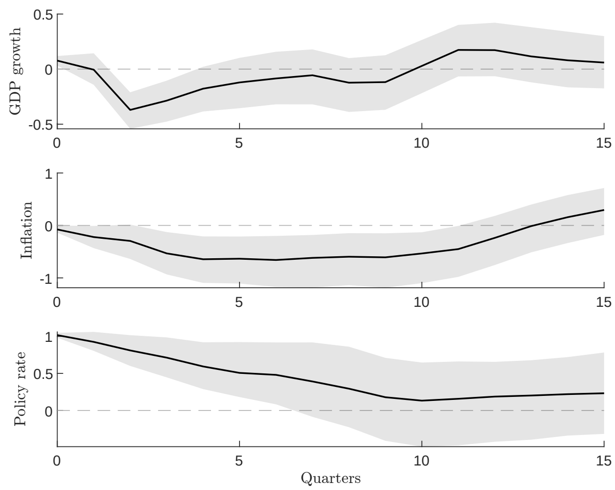
[27] The ECB policy toolkit is not restricted to the policy rate path, and balance sheet operations have also been used since 2007. We leave the evaluation of such policies for future work.

[28] The HICP and commodity price series are obtained from the ECB data portal and the other series are from the Fred database.

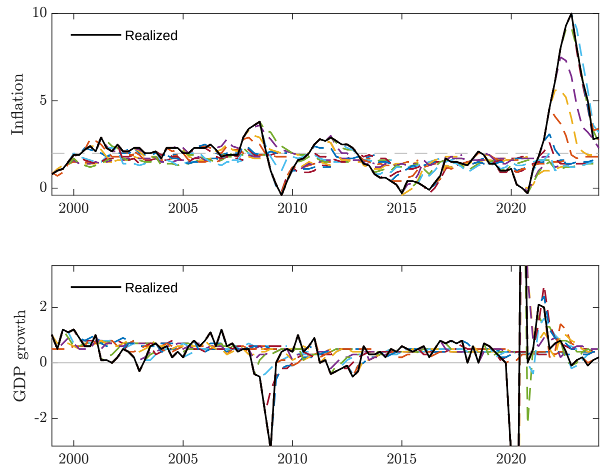
[29] Alternative identification schemes are possible. For instance, high frequency identified monetary surprises (e.g. Altavilla et al., 2019; Odendahl et al., 2024).

[30] Since the forecasts that we use below are only on a quarterly level we average the resulting impulse responses to the quarterly frequency.

Figure 2: The sufficient macro statistics



(a) Impulse responses to policy shock



(b) ECB macroeconomic forecasts

Notes: (a) Inflation is the y-to-y change in the Harmonised Index of Consumer Prices (HICP), and the policy rate is the Euro Overnight Index Average (EONIA). (b) Dashed lines are the ECB forecasts for each quarter.

The oracle forecasts

To approximate the oracle forecasts for interest rates, inflation, growth, unemployment and possibly other variables, several routes can be followed. Here we rely on the macro economic projections provided by the ECB. The forecasts are available from 1999 onward and are published four times a year (in March, June, September and December). We include all forecasts up to December 2023. Each forecasts is around 10 quarters into the future with slight variation across the reporting periods.

The forecasts are shown in the right-panel of Figure 2 together with realized inflation and GDP growth. It is easy to visually confirm that indeed the forecasts mean revert quickly, yet at the same time their accuracy is comparable to conventional time series model forecasts (Kontogeorgos and Lambrias, 2019). Unfortunately, the ECB does not provide any measure of model uncertainty, e.g. stemming from parameter estimates or other specification choices, in their publications. As an alternative to the ECB forecasts we also use the VAR model, as considered above, to construct forecasts. The details are discussed in the appendix.

9.2 Time-t ECB Policy evaluation

We evaluate the ECB policy decisions based on the loss function

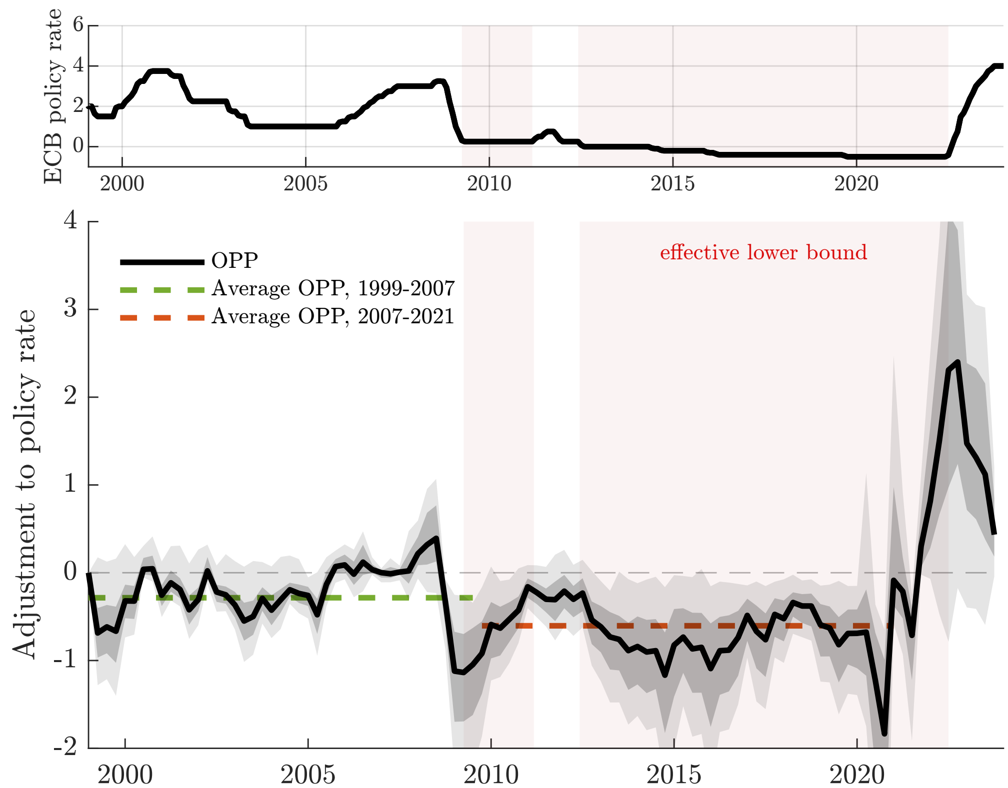
$$\mathcal{L}_t = \mathbb{E}_t \sum_{j=0}^{H} (\pi_{t+j} - \pi^*)^2 + \lambda (x_{t+j} - x^*)^2 \ , \tag{43}$$

where the inflation target π^* is set to 2% and potential output is set to $x^* = 0.6$, which is the average value of GDP growth over the 1999-2006 period, and the preference parameter $\lambda = 1$ with $H = 16$ quarters. In the appendix we explore the changes that occur when adding an interest rate smoothing objective to the loss function. Using the algorithm described above we compute the distribution of the OPP statistic for each quarter over the 1999-2023 period. The average OPP statistic is shown in Figure 3 together with the 67 and 95% confidence bands reflecting impulse response estimation uncertainty.[31]

With the caveat that our evaluation is restricted to the short-end of the policy path, we find that the ECB interest rate policy has been largely optimal during the early years of the ECB, though the policy rate was set slightly too high on average with an average OPP of about $-.25$ ppt. This reflects an inflation rate running persistently below 2 percent over that period. Starting in 2006, with inflation running slightly above 2 percent, the OPP makes a mild case for tightening, before plunging swiftly into negative territories with the dramatic collapse in GDP growth caused by the financial crisis. An interesting finding is that the OPP calls for stronger interest rate cuts in the early stage of the Great Recession, when the zero lower bound was *not yet* binding. In the first meeting of 2009, when the ECB deposit rate was still at 1.65 ppt, the OPP calls for a full 1 percentage point cut, a cut that the ECB will ultimately implement but only progressively and with a 6 months delay. Arguably, this earlier reaction could have attenuated some of the effects of the financial shock. This finding is analog to what was found for the Fed in Barnichon and Mesters (2023*b*). After 2010, the ECB entered a prolonged period (2009-2021) where conventional monetary policy was constrained by the zero/effective lower-bound. Not surprisingly, the average OPP over 2009-2021 is $-.6$ ppt, lower than the $-.25$ppt average over 1999-2008; this captures the fact that the lower bound on the policy rate did constrain monetary policy in the Euro area. Interestingly, while the constraint on ECB interest policy is about half a percentage .5ppt, and similar to the ZLB constraint on US monetary policy (Barnichon and Mesters, 2023*b*), the duration of the constraint is much longer. While US monetary policy was only restricted over 5 years (2009-2014), Euro area monetary policy was constrained for more than 10 years (2009-2021). While this reflects the effect of European debt crisis that

[31]Since we do not have a measure of model uncertainty for the ECB forecasts, the confidence bands are only based on IRF uncertainty.

Figure 3: OPP statistics for Euro area monetary policy



Notes: Top panel: the ECB Deposit Facility Rate. bottom panel: Adjustment to contemporaneous ECB policy rate as implied by the baseline OPP (thick line). Shaded areas report the 67 and 95% confidence bands. The green (or left) dashed line depicts the average OPP over 1999-2007, and the red (or right) dashed line depicts the average OPP over 2007-2021.

affected southern European countries over 2010-2014, this can also point to a less resilient Euro economy/more sclerotic the labor market (implying slower rebound from troughs) or to a lower value for the natural interest rate r^* in the Euro area than in the US. [32].

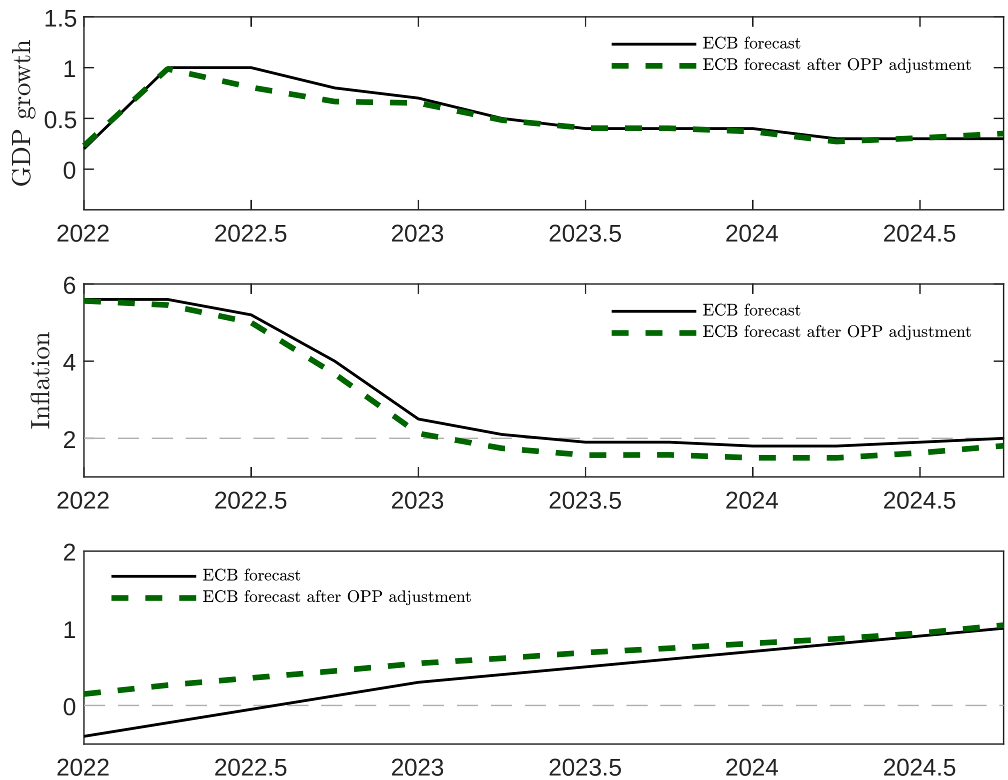
During the COVID recovery, inflation surged and the OPP calls for large increases in the policy rate, as much as 2 percentage points in 2022, a time when the ECB stayed put with the policy rate stuck at the lower bound. That said, a number of caveat are important. First, our loss function ignores any type of interest rate smoothing motive, which would penalize large changes in the policy rate. Another caveat is that of forward guidance: in its July 2021 monetary policy decision Press Release, the ECB stated that the "Governing Council expects the key ECB interest rates to remain at their present or lower levels *until it sees inflation reaching two per cent well ahead of the end of its projection horizon and durably for the rest of the projection horizon*". Further, the policy statement added that "This may also imply a transitory period in which inflation is moderately above target". With such a

[32] Ceteris paribus, a lower r^* makes the ZLB constraint more likely to bind (e.g., Le Bihan et al., 2019).

promise in place, it made sense for the ECB to delay its lift-off in the face of above-2 percent inflation. All that said, with headline inflation as high as 10 percent (substantially higher than US headline inflation), the case for a faster monetary reaction is hard to dismiss.

To better illustrate the sub-optimal delayed reaction of the ECB, we zoom in on the first quarter of 2022 policy decision and Figure reports the ECB forecasts as of February 3, 2022, along with the OPP adjusted paths. At the time the short term interest rates were fixed at the zero lower bound and the ECB did not start raising rates until July 2022. By having an earlier lift-off (and raising the policy rate from -0.5 to about $+0.25$, the ECB could have brought down inflation faster, reaching its inflation target about 6 months earlier (in expectation). The cost would have been lower GDP growth in 2022, by about 0.25 ppt.

Figure 4: OPP adjusted paths 2022-Q1

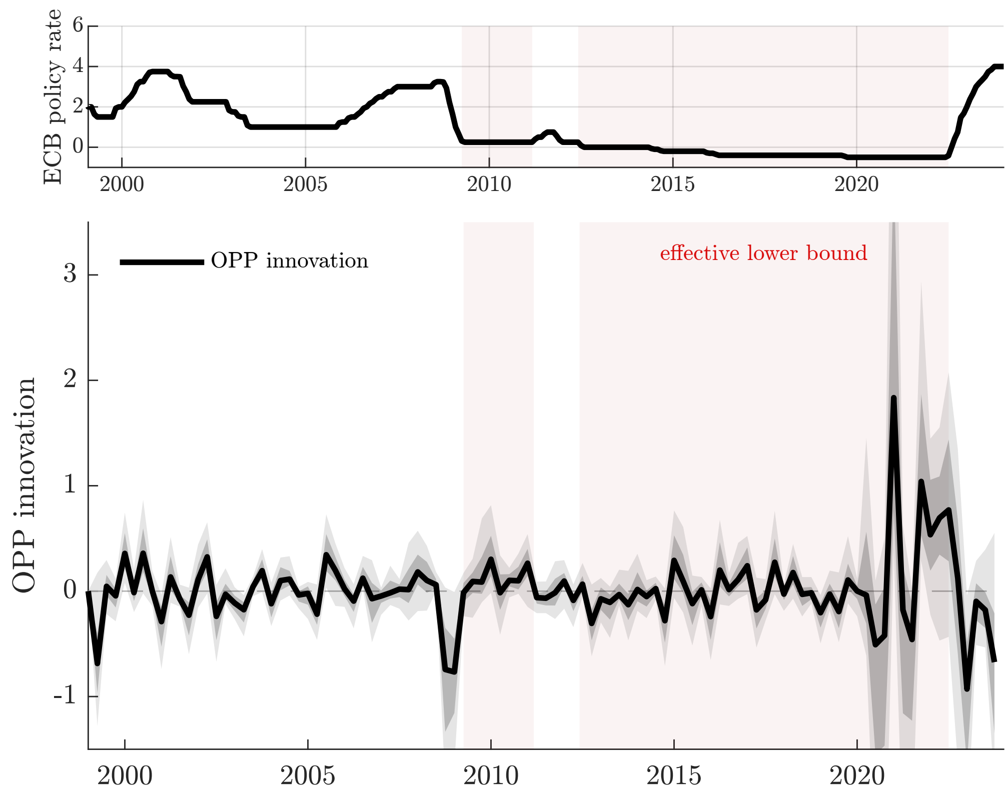


Notes: Plain lines: Expected paths for GDP growth, Inflation (HICP) and the ECB policy rate (Marginal Lending Facility Rate). Dashed green lines: Corresponding expected path after OPP-adjustment of the policy rule.

9.3 Overall (term) ECB policy evaluation

As last exercise, we can evaluate the overall ECB performance over 1999-2023 based on the timeless perspective, and we compute $\Delta_a = \mathbb{E}\delta_{a,t}^{\Delta\prime}\left(\mathcal{R}'\mathcal{W}\mathcal{R}\right)\delta_{a,t}^{\Delta}$. Figure 5 plots the OPP innovation series $\delta^{\Delta}_{a,t}$, and we can clearly see the two main sub-optimal decision dates in ECB's short history: in 2009 when the ECB did not lower interest rates enough in the face of deteriorating forecasts and mounting risks to unemployment, and in 2022 when the ECB did not raise interest rates in the face of rising inflation forecasts. In both cases, the ECB did ultimately react to these shocks, but the reaction came too late according to our sufficient macro statistics.

Figure 5: OPP innovations for Euro area monetary policy



Notes: Top panel: the ECB Deposit Facility Rate. bottom panel: Innovation to the OPP coming from new information (thick line). Shaded areas report the 67 and 95% confidence bands.

In units of loss function, these "policy misses" represent $\Delta_a = 0.2$ units of foregone welfare. To get a better sense of a 0.2 welfare loss, we can convert Δ_a in in terms of "inflation equivalent variation", similarly to the concept of consumption equivalent (CE) variation in welfare analysis.[33] The idea is to look for a "variation" $\Delta\pi$ such that $\mathbb{E}\mathcal{L}_t(\Delta\pi) = \mathbb{E}\mathcal{L}^0_t - \Delta_a$ where $\mathcal{L}_t(\Delta\pi) = \sum_{j=1}^{H} \left[(\pi_{t+j} - \pi^*)(1 - \Delta\pi]\right)^2 + \lambda(x_{t+j} - x^*)^2$. The inflation equivalent variation $\Delta\pi$ is the percentage reduction in the inflation gap over the next H periods that would generate the same welfare gains as Δ_a. To a first-order in $\Delta\pi$, we get $\mathcal{L}_t(\Delta\pi) =$

[33]CE is the amount of consumption —here the lower inflation gap— that an agent would require to be indifferent between staying in the economy with the baseline policy and the policy under the alternative (here, OPP-improved) policy.

$\mathcal{L}_t^0(1-2\Delta\pi)$, such that

$$\Delta\pi \simeq \Delta_a/2.$$

For the ECB over 1999-2023, we found $\Delta_a = 0.2$, such that the welfare gain of a superior ECB policy represents a 10 percent lower (in absolute value) inflation gap for 4 years, or more tellingly a 40 percent lower inflation gap for one year.

10 Conclusion

In this paper we unified the results from a number of recent studies that evaluate macro policy decisions using sufficient macro statistics. First, we disentangled policy evaluation into two separate tasks: time-t policy evaluation and term policy evaluation. The first task is typically performed repeatedly and in real time by policy makers, and the tools that we outlined help the policy maker to correctly calibrate the policy path given the information available at time t. The term policy evaluation results help the policy maker to ex-post evaluate the appropriateness of policy, and allows to highlight inefficiencies in the reaction functions.

These evaluation results are based on a number of representation, or identification, results that allow to represent counterfactual policy decisions in terms of forecasts and impulse responses under some baseline rule. While we have only displayed these results for linear models, representations for general nonlinear models can be found in McKay and Wolf (2023). Also, simple extensions, such as state dependence and time-varying parameters, are discussed in Barnichon and Mesters (2023*b*). That said, handling more complex nonlinearities in practice remains an open topic.

A practical limitation of the sufficient statistics approach is that it requires the identification of all policy shocks at all horizons of the policy path. For most empirical settings this requirement is too strong as only a few shocks can be empirically identified. To improve on this in future work, the identification of policy shocks — most notable at the long end of the policy paths — should take center stage (Caravello, McKay and Wolf, 2024).

A second and less highlighted practical limitation concerns medium- to long-term forecasts, say between 2 to 5 years, which are notoriously difficult (Farmer, Nakamura and Steinsson, 2024). While there exists a large econometric literature that develops macro economic forecasts, most of the work focuses on improving short run (<1 year) forecasts. Unfortunately, the responses of macro variables to policy changes typically take more time to materialize, implying that correctly calibrating policy paths requires accurate medium term forecasts. Improving medium-to long-run forecasting performances is an important task for future research.

Our empirical application focuses on monetary policy as one illustration of how a sufficient statistics framework can be applied to macroeconomic policy evaluation. However, the approach has broader relevance across a wide range of policy areas where decision-makers grapple with complex trade-offs. For example, in fiscal policy, it can aid in navigating the tension between promoting economic growth and maintaining debt sustainability. In the realm of exchange rate policy, it helps address the conflict between preserving monetary policy autonomy and ensuring currency stability. It also has applications in managing foreign reserves, where authorities must weigh the expense of holding reserves against their role as a safeguard against sudden capital outflows. Likewise, in the context of climate policy, this method can support efforts to balance the long-term damages of climate change with the immediate costs of mitigation efforts.

References

Alesina, Alberto, Carlo Favero, and Francesco Giavazzi. 2019. *Austerity: When It Works and When It Doesn't.* Princeton University Press.

Altavilla, Carlo, Luca Brugnolini, Refet S. Gürkaynak, Roberto Motto, and Giuseppe Ragusa. 2019. "Measuring euro area monetary policy." *Journal of Monetary Economics*, 108: 162–179.

Andrews, Isaiah, and Anna Mikusheva. 2015. "Maximum likelihood inference in weakly identified dynamic stochastic general equilibrium models." *Quantitative Economics*, 6(1): 123–152.

Antolín-Díaz, Juan, Ivan Petrella, and Juan F. Rubio-Ramírez. 2021. "Structural scenario analysis with SVARs." *Journal of Monetary Economics*, 117: 798–815.

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.

Auclert, Adrien, Matthew Rognlie, and Ludwig Straub. 2024. "Fiscal and Monetary Policy with Heterogeneous Agents." *Annual Review of Economics*, 17. forthcoming.

Bai, Jushan, and Serena Ng. 2006. "Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions." *Econometrica*, 74(4): 1133–1150.

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

Barnichon, Regis, and Geert Mesters. 2022. "Measuring Fiscal Discipline: A Revealed Preference Approach." Working paper.

Barnichon, Regis, and Geert Mesters. 2023*a*. "Evaluating Policy Institutions –150 Years of US Monetary Policy–." Working Paper.

Barnichon, Regis, and Geert Mesters. 2023*b*. "A Sufficient Statistics Approach for Macroeconomic Policy." *American Economic Review*, 113(11): 2809–45.

Bénassy-Quéré, Agnès, Benoit Coeuré, Pierre Jacquet, and Jean Pisani-Ferry. 2018. *Economic Policy: Theory and Practice. Second edition.* Oxford University Press.

Beraja, Martin. 2023. "A Semistructural Methodology for Policy Counterfactuals." *Journal of Political Economy*, 131(1): 190–201.

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

Bernanke, Ben S, Mark Gertler, and Mark Watson. 1997. "Systematic monetary policy and the effects of oil price shocks." *Brookings papers on economic activity*, 1997(1): 91–157.

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., and Mark W. Watson. 2016. "Presidents and the US Economy: An Econometric Exploration." *American Economic Review*, 106(4): 1015–45.

Bodkin, R. G., L. R. Klein, and K. Marwah. 1991. *A History of Macroeconometric Model-Building.* Edward Elgar, Aldershot.

Bourguignon, F., and A. Spadaro. 2012. "Tax-benefit revealed social preferences." *Journal of Economic Inequality*, 10: 75–108.

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

Canova, Fabio, and Luca Sala. 2009. "Back to square one: Identification issues in DSGE models." *Journal of Monetary Economics*, 56(4): 431–449.

Caravello, Tom, Alisdair McKay, and Christian Wolf. 2024. "Evaluating Policy Counterfactuals: A "VAR-Plus" Approach."

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

de Boor, C. 2001. *A Practical Guide to Splines.* New York, NY, USA:Springer.

de Groot, Oliver, Falk Mazelis, Roberto Motto, and Annukka Ristiniem. 2021. "A toolkit for computing Constrained Optimal Policy Projections (COPPs)." ECB working paper: No 2555.

Dhaene, Geert, and Anton P. Barten. 1989. "When it all began: The 1936 Tinbergen model revisited." *Economic Modelling*, 6(2): 203–219.

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.

Einarsson, Bjarni G. 2024. "Testing optimal monetary policy in a currency union." Central bank of Iceland: working paper 96.

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

Elliott, Graham, Allan Timmermann, and Ivana Komunjer. 2005. "Estimation and Testing of Forecast Rationality under Flexible Loss." *The Review of Economic Studies*, 72(4): 1107–1125.

Elliott, Graham, Ivana Komunjer, and Allan Timmermann. 2008. "Biases in Macroeconomic Forecasts: Irrationality or Asymmetric Loss?" *Journal of the European Economic Association*, 6(1): 122–157.

Farmer, Leland E, Emi Nakamura, and Jón Steinsson. 2024. "Learning about the long run." *Journal of Political Economy*, 132(10): 3334–3377.

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.

Giannoni, Marc, and Michael Woodford. 2004. "Optimal Inflation-Targeting Rules." *The Inflation-Targeting Debate*. University of Chicago Press.

Hansen, Lars Peter, and Kenneth J Singleton. 1982. "Generalized instrumental variables estimation of nonlinear rational expectations models." *Econometrica*, 1269–1286.

Hansen, Lars Peter, and Thomas J. Sargent. 2008. *Robustness.* Princeton University Press.

Hartman, Philipp, and Frank Smets. 2018. "The European Central Bank's Monetary Policy during Its First 20 Years." *Brookings Papers on Economic Activity*, 49(2 (Fall)): 1–146.

Hebden, James, and Fabian Winkler. 2021. "Impulse-Based Computation of Policy Counterfactuals." Finance and Economics Discussion Series 2021-042. Washington: Board of Governors of the Federal Reserve System.

Hendren, Nathaniel. 2020. "Measuring economic efficiency using inverse-optimum weights." *Journal of Public Economics*, 187: 104198.

Inoue, Atsushi, and Barbara Rossi. 2021. "A new approach to measuring economic policy shocks, with an application to conventional and unconventional monetary policy." *Quantitative Economics*, 12(4): 1085–1138.

Jacobs, Bas, Egbert L.W. Jongen, and Floris T. Zoutman. 2017. "Revealed social preferences of Dutch political parties." *Journal of Public Economics*, 156: 81–100.

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

Kilian, Lutz, and Helmut Lütkepohl. 2017. *Structural Vector Autoregressive Analysis.* Cambridge University Press.

Killian, Lutz, and Simone Manganelli. 2008. "The Central Banker as a Risk Manager: Estimating the Federal Reserve's Preferences under Greenspan." *Journal of Money, Credit and Banking*, 40(6): 1103–1129.

Kleven, Henrik J. 2020. "Sufficient Statistics Revisited." *Annual Review of Economics.* forthcoming.

Kocherlakota, Narayana R. 2019. "Practical policy evaluation." *Journal of Monetary Economics*, 102: 29 – 45.

Kock, Anders Bredahl, Marcelo Medeiros, and Gabriel Vasconcelos. 2020. "Penalized Time Series Regression." *Macroeconomic Forecasting in the Era of Big Data: Theory and Practice*, , ed. Peter Fuleky, 193–228. Cham:Springer International Publishing.

Kontogeorgos, Georgios, and Kyriacos Lambrias. 2019. "An analysis of the Eurosystem/ECB projections." European Central Bank Working Paper Series 2291.

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.

Le Bihan, Herve, Jordi Galí, Philippe Andrade, and Julien Matheron. 2019. "The optimal inflation target and the natural rate of interest." *Brookings Papers on Economic Activity*, , (Fall 2019 Edition).

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

Ljungqvist, Lars, and Thomas J. Sargent. 2004. *Recursive Macroeconomic Theory.* Cambridge, Massachusetts:MIT Press.

Lucas, Robert E. 1972. "Expectations and the neutrality of money." *Journal of Economic Theory*, 4(2): 103–124.

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

McKay, Alisdair, and Christian Wolf. 2022. "Optimal policy rules in hank."

McKay, Alisdair, and Christian Wolf. 2023. "What Can Time-Series Regressions Tell Us About Policy Counterfactuals?" *Econometrica*. forthcoming.

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.

Odendahl, Florens, Maria Sole Pagliari, Adrian Penalver, Barbara Rossi, and Giulia Sestieri. 2024. "Euro area monetary policy effects. Does the shape of the yield curve matter?" *Journal of Monetary Economics*, 147: 103617.

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

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

Scheffe, Henry. 1953. "A Method for Judging all Contrasts in the Analysis of Variance." *Biometrika*, 40(1/2): 87–104.

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

Sims, Christopher A. 1982. "Policy Analysis with Econometric Models." *Brookings Papers on Economic Activity*, 1982(1): 107–164.

Sims, Christopher A., and Tao Zha. 1995. "Does monetary policy generate recessions?"

Sims, Christopher A., and Tao Zha. 2006. "Were There Regime Switches in U.S. Monetary Policy?" *American Economic Review*, 96(1): 54–81.

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. 2016. "Chapter 8 - Dynamic Factor Models, Factor-Augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics." In . Vol. 2 of *Handbook of Macroeconomics*, , ed. John B. Taylor and Harald Uhlig, 415 – 525. Elsevier.

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

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

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.

Theil, H. 1956. "On the Theory of Economic Policy." *The American Economic Review*, 46(2): 360–366.

Tinbergen, Jan. 1952. *On the Theory of Economic Policy.* North-Holland Publishing Company, Amsterdam.

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.

Waggoner, Daniel F., and Tao Zha. 1999. "Conditional Forecasts in Dynamic Multivariate Models." *The Review of Economics and Statistics*, 81(4): 639–651.

Woodford, Michael. 2003. *Interest and Prices: Foundations of a Theory of Monetary Policy.* Princeton University Press.

Appendix

Sequence space representation and news shocks

In this section, we clarify our use of a sequence space representation and the role played by news shocks. These elements are important for modern policy evaluation methods, yet they are typically not covered in standard macroeconometric coursework. As we will see, the sequence space representation has important benefits in terms of clarity —loosely speaking, turning a dynamic problem into a seemingly static one—.

Sequence space notation

To help understand the sequence space representation and associated notations, we consider a simple example:

$$y_t = \phi y_{t-1} + \upsilon_t \qquad \upsilon_t \overset{iid}{\sim} (0, \sigma^2) \ ,$$

which is the conventional recursive formulation for an AR(1) model with iid errors. The sequence space representation of this model stacks all current and future outcomes in an infinite vector, i.e. $\mathbf{Y}_t = (y_t, y_{t+1}, y_{t+2} \ldots)'$ and represents the model as

$$\mathbf{\Phi} \mathbf{Y}_t = \boldsymbol{\upsilon}_t \ ,$$

where

$$\mathbf{\Phi} = \begin{bmatrix} 1 & 0 & 0 & 0 & \ldots \\ -\phi & 1 & 0 & 0 & \ldots \\ 0 & -\phi & 1 & 0 & \ddots \\ 0 & 0 & -\phi & 1 & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{\upsilon}_t = \begin{bmatrix} \upsilon_t \\ \upsilon_{t+1} \\ \upsilon_{t+2} \\ \upsilon_{t+3} \\ \vdots \end{bmatrix} .$$

Note that for simplicity we have set $y_{t-1} = 0$, which is not necessary and will be avoided below by introducing initial conditions. The representation $\mathbf{\Phi} \mathbf{Y}_t = \boldsymbol{\upsilon}_t$ has two key benefits: (i) it looks static facilitating easy manipulation[34] and (ii) changes in $\mathbf{\Phi}$ directly document how the entire path of $y_t, y_{t+1}, \ldots$ changes.

In the formulation above the sequence space is written under perfect foresight, i.e. the future shocks are considered observable. While for some exercises this representation is sufficient and convenient, at times we are interested in the model given the information available at time t. Think of a policy maker at time t who is interested in forecasting the path of y. Such policy maker only has information $\mathcal{F}_t = \{\upsilon_t, \upsilon_{t-1}, \ldots\}$.

To define the model given $\mathcal{F}_t$ let $\mathbb{E}_t(\cdot) = \mathbb{E}(\cdot|\mathcal{F}_t)$ be the conditional expectation operator. We have

$$\mathbf{\Phi} \mathbb{E}_t \mathbf{Y}_t = \mathbb{E}_t \boldsymbol{\upsilon}_t \ ,$$

where $\mathbb{E}_t \boldsymbol{\upsilon}_t = (\mathbb{E}_t \upsilon_t, \mathbb{E}_t \upsilon_{t+1}, \ldots)' = (\upsilon_t, 0, \ldots)'$ as the shocks in this example are iid.

[34] Off course we are ignoring many subtle and important aspects of manipulating infinite dimensional maps, but for most of our purposes this will not harm us.

News shocks

In many macro policy settings the exogenous information that is recovered, e.g. using a narrative approach, does not necessarily pertain to the contemporaneous value of the policy instruments. Quite often such exogenous information is also about future values of the policy instruments. Clearly, such information about future policy can affect how agents act today and therefore the release is relevant for policy makers, we refer to these exogenous movements as news shocks.

Concrete examples include a central bank that announces to keep the interest rate low for the coming years, or governments who make plans for spending, taxes and transfers for the coming four years. In each case the exogenous components of such plans can be regarded as news shocks pertaining to the different horizons of the plan. As we will see below such news shocks provide important information that can be used to evaluate policy decisions.

To introduce news shocks in an easy way we build on the previous AR(1) example. Consider

$$y_t = \phi y_{t-1} + \upsilon_t \ , \qquad \upsilon_t = \tilde{\upsilon}_{t-1,t} + \tilde{\upsilon}_{t,t} \ ,$$

where $\tilde{\upsilon}_{t-1,t}$ is the news shock that was released at $t-1$ and contains news about υ_t in time period t, and $\upsilon_{t,t}$ is the contemporaneous shock: released at t about t. We assume that all $\tilde{\upsilon}_{j,t}$ are mutually and serially uncorrelated. The time-t information set $\mathcal{F}_t$ now includes all shocks that are released prior or at time t, i.e. $\mathcal{F}_t = \{\tilde{\upsilon}_{j,k}, j \leq t, k \geq j\}$.

The sequence space representation evaluated given $\mathcal{F}_t$ takes the same general form as above

$$\mathbf{\Phi}\mathbb{E}_t\mathbf{Y}_t = \mathbb{E}_t\boldsymbol{\upsilon}_t$$

but now

$$\mathbb{E}_t\boldsymbol{\upsilon}_t = \begin{bmatrix} \mathbb{E}_t\upsilon_t \\ \mathbb{E}_t\upsilon_{t+1} \\ \mathbb{E}_t\upsilon_{t+2} \\ \mathbb{E}_t\upsilon_{t+3} \\ \vdots \end{bmatrix} = \underbrace{\begin{bmatrix} \tilde{\upsilon}_{t,t} \\ \tilde{\upsilon}_{t,t+1} \\ 0 \\ 0 \\ \vdots \end{bmatrix}}_{\text{time-}t\text{ news shock}} + \underbrace{\begin{bmatrix} \tilde{\upsilon}_{t-1,t} \\ 0 \\ 0 \\ 0 \\ \vdots \end{bmatrix}}_{\text{old news}} .$$

Now there are two shocks that are released at time t: news about time t — $\tilde{\upsilon}_{t,t}$ — and news about $t+1$ — $\tilde{\upsilon}_{t,t+1}$ —. This example, can be generalized by considering

$$y_t = \phi y_{t-1} + \upsilon_t \ , \qquad \upsilon_t = \sum_{j=0}^{\infty} \tilde{\upsilon}_{t-j,t} \ ,$$

where these is now an entire sequence of news shocks. We have

$$\mathbb{E}_t\boldsymbol{\upsilon}_t = \begin{bmatrix} \tilde{\upsilon}_{t,t} \\ \tilde{\upsilon}_{t,t+1} \\ \tilde{\upsilon}_{t,t+2} \\ \tilde{\upsilon}_{t,t+3} \\ \vdots \end{bmatrix} + \begin{bmatrix} \sum_{j=1}^{\infty} \tilde{\upsilon}_{t-j,t} \\ \sum_{j=2}^{\infty} \tilde{\upsilon}_{t+1-j,t} \\ \sum_{j=3}^{\infty} \tilde{\upsilon}_{t+2-j,t} \\ \sum_{j=4}^{\infty} \tilde{\upsilon}_{t+3-j,t} \\ \vdots \end{bmatrix} = \tilde{\boldsymbol{\upsilon}}_t + \mathbf{X}_{t-1} \ ,$$

where $\tilde{\boldsymbol{v}}_t = (\tilde{v}_{t,t}, \tilde{v}_{t,t+1}, \ldots)'$ is the path of time t news shocks and $\mathbf{X}_{t-1}$ captures initial conditions.

What are we identifying?

Having defined the news shocks we clarify how the existing empirically identified macro shocks can be conceptually relate to the news shocks. The short answer is that in most empirical settings we will not know exactly which combination of news shocks is being empirically identified, and the best we can say is that some combination of news shocks is being captured. The good news is that this is fine for most policy evaluation exercises. To make this clear consider the following examples.

Suppose that $\mathbb{E}_t(y_t, y_{t+1}, \ldots)$ is the expected interest rate path and $\tilde{\boldsymbol{v}}_t = (\tilde{v}_{t,t}, \tilde{v}_{t,t+1}, \ldots)'$ is the path of monetary policy news shocks that are announced at time t. In practice, we often use proxies for such shocks that are obtained by measuring changes in asset prices in short windows around press conferences of central banks (e.g. Kuttner, 2001). For the sake of the argument, suppose that these high frequency identified measures are exactly correct, i.e. exogenous and not contaminated by measurement error.

Suppose that the asset used is the short term interest rate, does this make the identified shock the contemporaneous shock, i.e. $\tilde{v}_{t,t}$? Not necessarily, if the press conference only announces changes in future interest rates then the measured high frequency change in the short term interest rate is driven by some $\tilde{v}_{t,t+h}$. In fact, in most cases we will not be sure which specific news shocks are responsible for the change as the press conference could be about many horizons of monetary policy.

Similarly, consider the government spending shocks identified by Ramey and Zubairy (2018). The recovered series contains news about military defense spending that is obtained from news paper articles. Similarly as above, the horizon to which the news pertains is often not clear from the articles and therefore we cannot label the shocks as being a specific $\tilde{v}_{t,t+h}$.

In general, close inspection of the main identifying strategies for macro shocks reveals that it is often not possible to determine to which particular horizon the identified shock pertains. We will therefore postulate that the identified shocks are some linear combinations of the theoretically defined news shocks, i.e. we identify some subset

$$\tilde{\boldsymbol{v}}_{a,t} = A\tilde{\boldsymbol{v}}_t$$

where A is a weighting matrix. In the ideal scenario we would like to span the entire path of news shocks and A is some invertible map, yet in practice we often will only have access to a few news shocks.

Additional empirical results

Including forecast uncertainty

As an alternative to the ECB forecasts we also use the VAR model, as considered above for the estimation of the impulse responses, to construct forecasts. This allows us to account for the uncertainty in the forecast when constructing the OPP/DML confidence bands. Specifically, we use the VAR model repeatedly estimated over increasing windows starting from 2007-M1 until 2022-M12. The pre-2007 period is used for parameter estimation. For each period we sample from the forecast distribution to compute OPP statistic.

The left panel of Figure 5 shows the resulting OPP sequence. We keep the same dates on the x-axis for comparison with Figure 2. Overall, the average of the OPP sequence is similar to what we found using the ECB forecasts. However, it is worth noting that the confidence bands are substantially larger as here we take into account the forecasting uncertainty from the VAR. This highlights an important practical point: for most periods it was hard to know with high probability that the policy was non-optimal.

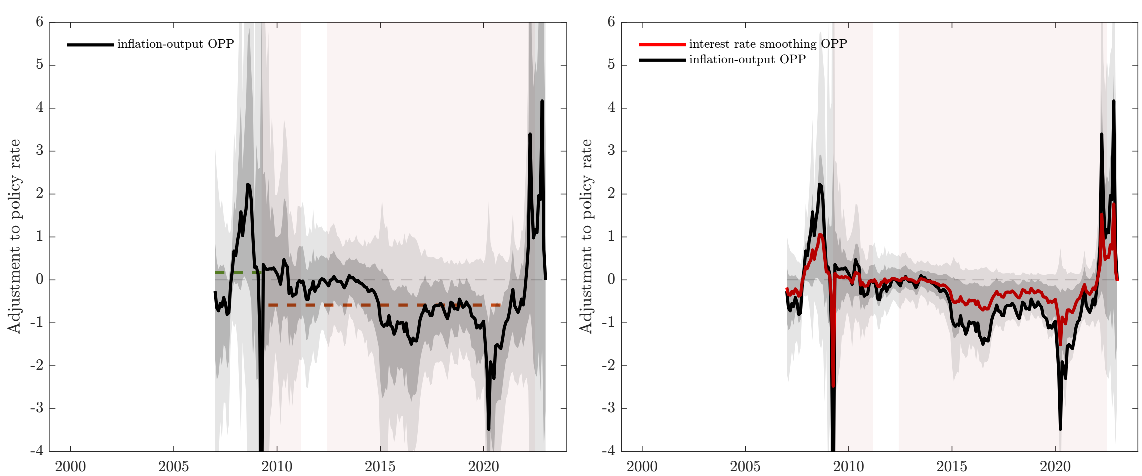
Interest rate smoothing

As discussed in Section 8 in detail, selecting the preferences parameters for the loss function is a non-trivial task. Different perspectives can be taken pending on the research objectives. Here we briefly revisit the our Euro area application and explore the influence of interest rate smoothing on the policy adjustments. In our baseline specification we relied on the conventional inflation-output loss function (43) and placed equal weight on both objectives. To include interest rate smoothing as an objective we now consider

$$\mathcal{L}_t = \mathbb{E}_t \sum_{j=0}^{H} (\pi_{t+j} - \pi^*)^2 + \lambda_1 (x_{t+j} - x^*)^2 + \lambda_2 (i_{t+j} - i_{t+j-1})^2 \ , \tag{44}$$

where the last term captures the desire for avoiding large changes in the interest rate. For comparison purposes we place equal weight on all objectives. The results are shown in the right panel of Figure 6. We find that in general the OPP indicates less large deviations from optimality when interest rate smoothing is taken into account.

Figure 6: OPP statistics for Euro area monetary policy - VAR based



Notes: Left panel: Adjustment to contemporaneous ECB policy rate as implied by the baseline OPP (thick line) computed using VAR forecasts using the conventional inflation-output loss function (43). Right panel: Adjustment to contemporaneous ECB policy rate as implied by the baseline OPP (thick line) computed using VAR forecasts using the including interest rate smoothing in the loss function (44). Shaded areas report the 67 and 95% confidence bands. The green dashed line depicts the average OPP over 1999-2007, and the red dashed line depicts the average OPP over 2009-2021.