Barcelona School of Economics

WORKING PAPERS

Evaluating Policy Institutions -150 Years of US Monetary Policy-

BSE Working Paper 1410| October 2023

Régis Barnichon, Geert Mesters

bse.eu/research

Evaluating Policy Institutions*

–150 Years of US Monetary Policy–

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

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

[(b)] Universitat Pompeu Fabra, Barcelona School of Economics and CREI

October 27, 2023

Abstract

How should we evaluate and compare the performances of policy institutions? We propose to evaluate institutions based on their reaction function, i.e., on how well they reacted to the different shocks that hit the economy. We show that reaction function evaluation is possible with only two sufficient statistics (i) the impulse responses of the policy objectives to non-policy shocks, which capture what an institution *did* on average to counteract these shocks, and (ii) the impulse responses of the policy objectives to policy shocks, which capture what an institution *could have* done to counteract the shocks. A regression of (i) on (ii) —a regression in impulse response space— allows to compute the distance to the optimal reaction function, and thereby evaluate and rank institutions. We use our methodology to evaluate US monetary policy; from the Gold standard period, the early Fed years and the Great Depression, to the post World War II period, and the post-Volcker regime. We find no material improvement in the reaction function over the first 100 years, and it is only in the last 30 years that we estimate large and uniform improvements in the conduct of monetary policy.

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

Keywords: optimal policy, reaction function, structural shocks, impulse responses, monetary history.

*We thank our discussant Frank Smets as well as Ryan Chahrour, John Friedman, Toru Kitagawa, Sylvain Leduc, Karel Mertens, Emi Nakamura, Adam Shapiro, Jón Steinsson, Christian Wolf and seminar participants at CEMFI, Paris School of Economics, Mannheim, UC Berkeley, Brown, the NBER Summer Institute (ME, FEM), INSPER, PUC Rio, FGV, Bank of Italy, ECB, Riksbank and the International Francqui Chair Symposium in Brussels for helpful comments. Mesters acknowledges financial support from the Severo Ochoa Programme for Centres of Excellence in R&D (Barcelona School of Economics CEX2019-000915-S), funded by MCIN/AEI/10.13039/50110001103 and 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 or the Federal Reserve System.

1 Introduction

How should we evaluate and compare the performance of policy institutions? How should we evaluate and compare policy makers after their term in office? These questions are of central importance to the good functioning of democratic and accountable institutions, but there is little consensus on a method for evaluating and comparing performance.

A naive approach could consist in measuring performance based on realized macroeconomic outcomes. For instance, we could assess a central banker based on average inflation and unemployment outcomes over her term. Unfortunately, that approach suffers from three types of confounding factors: (i) different policy makers may face different initial conditions, e.g. a central banker can inherit a strong or weak economy from her predecessor, (ii) different policy makers may face different economic disturbances, e.g., a central banker may experience a financial crisis or an energy price shock that will affect her ability to stabilize inflation and unemployment, and (iii) different policy makers may live in different economic environments, e.g., a steeper or flatter Phillips curve will affect a central banker's ability to control inflation.

This triplet of confounding factors coming from different initial conditions, different disturbances and different economic environments severely limits our ability to evaluate policy makers based on realized outcomes.

To make progress it is instructive to consider an ideal, yet infeasible, approach for comparing policy makers: an experimental approach. Consider setting up a laboratory, in which different policy makers are given the same mandate —minimizing a loss function involving some policy objectives— and are subjected to the same initial conditions and the same economic environment. The different policy makers are then exposed to the same sequence of shocks, and they each make decisions that aim to achieve the mandate. Afterward, we can compare performance from the realized losses and conclude which policy maker performed better. Such comparison would be on equal grounds as the only source of variation would come from the policy makers' reaction functions, i.e., from the different ways each policy maker *reacted* to the same of sequence shocks.

In this paper, we propose an empirical method that aims to mimic this ideal "reaction function comparison" experiment while making minimal structural assumptions on the underlying economic model and the underlying policy rule. Our approach exploits a simple idea: while different policy makers are never exposed to the same *sequences* of non-policy shocks, they are often exposed to the same *types* of shocks; for instance energy price shocks, financial shocks, war shocks, or even pandemic shocks. By comparing how well different policy makers performed in response to such common shocks, we can approach the ideal empirical setting sketched above: assessing and comparing performance from the different ways each policy maker reacted to the same types of shocks.

Geometrically speaking, our strategy amounts to projecting realized macroeconomic outcomes on a space spanned by well chosen non-policy shocks (common to the policy makers under comparison) and to study policy performance in that space. In fact, in that subspace policy evaluation reduces to a simple optimization problem that only involves two well-known (and estimable) sufficient statistics: (i) the impulse responses of the policy objectives to non-policy shocks, and (ii) the same impulse responses but to policy shocks.

The first set of impulse responses —the impulse responses to a specific non-policy shock— capture the average effects of that non-policy shock under the policy maker's reaction function and allow to compute a *conditional* loss; a loss conditional on that non-policy shock. For instance, with a quadratic loss function the conditional loss is simply the sum-of-squares of that impulse response. While it is tempting to assess and compare performance based on that impulse response alone, this is not enough since other factors beyond a policy maker's reaction function could generate a lower conditional loss, i.e., a more stable impulse response. For instance, a different economic environment could make the economy more stable independently of the policy makers' reaction function. To assess how well a policy maker reacted to that specific non-policy shock, we need to know the outcome of a policy rule counterfactual: how a different reaction would have affected the economy. That counterfactual can be recovered by the second set of impulse responses —the impulse responses to policy shocks—, which allow to compute how a different reaction function would have affected the conditional loss —what the policy maker could have done to counteract the non-policy shock—.

We show that for a large class of models and quadratic loss functions the *distance to the optimal reaction*, or Optimal Reaction Adjustment (ORA), can be computed from a simple regression in "impulse response space": a regression of the impulse responses to the non-policy shock on the impulse responses to policy shocks.

The ORA measures by how much more or less a policy maker should have responded to a given non-policy shock, and it provides a direct measure of policy performance conditional a specific type of non-policy shock. Overall policy performance can then be assessed by measuring the ORAs for different types of non-policy shocks.

While environments can be different across policy makers, the ORA statistic "controls" for the economic environment, capturing the distance to an optimal reaction function *given* the economic environment. We can thus use the ORAs to compare policy makers or policy institutions across time (say the Fed in 1930s vs the Fed in the 2000s) or across space (say the Fed vs the ECB).

With the ORA depending only on impulse responses to shocks, the evaluation and comparison of policy makers reduces to a well-known econometric task: the estimation of structural impulse responses, and this realization opens a number of important avenues for policy evaluation, as one can draw on a large macro-econometric literature to evaluate policy institutions. See e.g., Ramey (2016) for a recent discussion of structural shock identification and Stock and Watson (2016), Kilian and Lütkepohl (2017) and Li, Plagborg-Møller and Wolf (2022) for recent work on impulse response estimation methods.

We then apply our methodology to study the performance of US monetary policy over the past 150 years. Our method allows us to address and revisit many important questions regarding the conduct of monetary policy over the past 150 years: (i) Did the founding of the Federal Reserve in 1913 led to superior macro outcomes than during the passive Gold standard period (e.g., Bordo and Kydland, 1995)? Or did the founding of Fed led to worse performance? (ii) While many people would agree that monetary policy was superior during the 2007-2009 financial crisis than during the 1929-1933 financial crisis (e.g., Wheelock et al., 2010), can we confirm and quantify this improvement? In other words, did Bernanke fulfill his promise to Milton Friedman when he said that the Fed "won't do it again", i.e., won't repeat the mistakes of the Great Depression (Bernanke, 2002)? (iii) More generally, did monetary policy improve since the Great Depression? Is the Great Moderation post Volcker a sign of good policy or simply the outcome of good luck? (e.g., Clarida, Galí and Gertler, 2000; Galí and Gambetti, 2009)?

To assess and compare monetary policy performance across historical periods, we evaluate how monetary policy responded to five types of non-policy shocks: (i) financial shocks, (ii) government spending shocks, (iii) energy price shocks, (iv) inflation expectation shocks and (v) productivity shocks, and we evaluate US monetary policy over four distinct periods: (a) 1879-1912 covering the Gold standard period until the founding of the Federal Reserve, (b) 1913-1941 covering the early Fed years to the US entering World War II, (c) 1954-1984 covering the post World War II period until the beginning of the Great Moderation, and (d) 1990-2019 covering the Great Moderation period, the financial crisis and up to the COVID crisis. In each case, we leverage on a large empirical literature on structural shocks identification to identify banking panics (Reinhart and Rogoff, 2009), energy price shocks (Hamilton, 2003), government spending shocks (Ramey and Zubairy, 2018), TFP shocks (Gali, 1999), inflation expectation shocks (Leduc, Sill and Stark, 2007) and monetary shocks (Friedman and Schwartz, 1963; Romer and Romer, 1989, 2004*b*; Gürkaynak, Sack and Swanson, 2005). The identification of monetary shocks is more challenging (and less developed) for the Gold Standard period, and we propose a new identification strategy based on large gold mine discoveries.

Evaluating and comparing policy makers requires to take a stand on a set of objectives, i.e., on a loss function. In our empirical application, we consider a quadratic loss function with equal weights on inflation and unemployment.[1] Given that loss function, our main

[1]Our approach can accommodate other loss functions, for instance different loss functions across time

results are as follows: (i) we estimate large and uniform improvements in the conduct of monetary policy, but *only* in the last 30 years, (ii) we cannot reject that the Fed's reaction to recent financial shocks (notably the 2007-2008 financial crisis) was appropriate, in contrast to the "highly" sub-optimal reaction of the Fed during the Great Depression, (iii) despite much larger realized losses in the 1920s-1930s, the performance of the early Fed is no worse than the performance of the passive Gold Standard, and (iv) the Fed's reaction function during the 1960s-1970s is almost as sub-optimal as the reaction function of the early Fed.

Related literature

An early contribution is Fair (1978) who highlights the distortions stemming from different initial conditions and economic environments. His solution was to adopt optimal control methods to compare policy makers through the lens of a fully specified model. Modern versions of this approach include (e.g. Galı, López-Salido and Vallés, 2003; Gali and Gertler, 2007; Blanchard and Galí, 2007). Unfortunately, specifying the correct model for (i) the policy rule and (ii) the macroeconomic non-policy block is a very difficult task (e.g., Svensson, 2003; Mishkin, 2010). A less structural approach has studied monetary performance through the lens of estimated policy rules —requiring only the specification of a policy rule—.[2] In particular, a number of studies compared the Fed in the pre- and post-Volcker periods by assessing whether the Taylor principle was satisfied. However, beyond the Taylor principle, that approach can say little about the optimality of reaction functions, and thus can only provide a coarse evaluation of reaction functions.

In the context of fiscal policy Blinder and Watson (2016) improve on the naive approach of policy evaluation —measuring performance based on unconditional realized outcomes— by *projecting out* specific macro shocks, i.e., by trying to control for good (or bad) luck. In contrast, our approach *projects on* the space spanned by specific non-policy shocks and study performance in that space: comparing policy makers by studying how well they reacted to the same type of shock.

Closer to our work, the literature has proposed reduced-form methods to study policy rule counterfactuals (e.g., Sims and Zha, 2006; Bernanke et al., 1997; Leeper and Zha, 2003), though these approaches are not fully robust to the Lucas critique. Instead, our approach builds on recent work showing that robustness to the Lucas critique is possible in a large class of macroeconomic models (McKay and Wolf, 2023). When the coefficients of the non-policy block are independent of the coefficients of the policy block, it is possible to reproduce any policy rule counterfactual with an appropriate combination of policy news shocks at

periods, or even micro-founded welfare-based loss functions.

[2]See Judd and Rudebusch (1998); Taylor (1999); Clarida, Galí and Gertler (2000); Orphanides (2003); Boivin (2005); Coibion and Gorodnichenko (2011) for policy rules estimates.

different horizons. The information requirement for such impulse-based rule counterfactuals is typically very large however —requiring the identification of all policy news shocks—, and the counterfactual effects can often only be approximated. Our work exploits a little studied yet attractive class of policy rule counterfactuals —counterfactual reactions to non-policy shocks—, which are simple to compute exactly; requiring only one policy shock at a time.

Last, our paper relates to the sufficient statistics approach for macroeconomic policy proposed in Barnichon and Mesters (2023). Different from our focus on reaction function evaluation, Barnichon and Mesters (2023) focus on the time t optimal policy problem —how to set the policy *path* today given the state of the economy—, instead of the unconditional policy problem that we consider here —how to set up the policy *rule* to minimize the unconditional loss—. Barnichon and Mesters (2023) show that the characterization of the time t optimal policy path can be reduced to the estimation of two sufficient statistics (i) forecasts for the policy objectives conditional on some baseline policy choice, (ii) the impulse responses of the policy objectives to policy shocks. However, these two statistics are not sufficient to evaluate the optimality of the underlying policy rule. The present paper shows that a sufficient statistics approach to rule evaluation is possible, but it requires a different set of statistics, and notably additional identifying restrictions: the identification of (at least some) non-policy shocks.

Our historical evaluation of monetary policy relates to monumental narrative studies of monetary policy, from Friedman and Schwartz (1963) seminal work to the more recent work of Meltzer (2003; 2009*a*; 2009*b*). Our study builds on this narrative evidence in that much our shock identification draws on the narrative identification approach pioneered by Friedman and Schwartz (1963) and Romer and Romer (1989). While our historical study is necessarily less thorough than these historical accounts, we show that it is possible to use narrative *qualitative* accounts to make objective and *quantitative* statements about historical policy performance.

The remainder of this paper is organized as follows. The next section illustrates our method for a simple New Keynesian model. Section 3 presents the general environment. Section 4 provide the results for evaluating and ranking policy makers. The results from the empirical study for monetary policy are discussed in Section 5. Section 6 concludes.

2 Illustrative example

Before formally describing our general framework, we first illustrate how it is possible to evaluate and compare policy makers' reaction functions without having access to the underlying economic model nor the policy rule. To describe the economy, we take a baseline New Keynesian (NK) model, which allows us to highlight the main mechanisms of our approach and relate to the broad NK literature (e.g. Galí, 2015).

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

$$\pi_t = \mathbb{E}_t\pi_{t+1} + \kappa x_t + \xi_t \ , \tag{1}$$
$$x_t = \mathbb{E}_t x_{t+1} - \frac{1}{\sigma}(i_t - \mathrm{E}_t\pi_{t+1}) \ , \tag{2}$$

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.

The policy maker sets the interest rate by responding to the economy according to

$$i_t = \phi_\pi \pi_t + \phi_\xi \xi_t + \epsilon_t \ , \tag{3}$$

where $\phi = (\phi_\pi, \phi_\xi)$ is a vector of reaction coefficients —for short, the "reaction function"—, which captures the systematic response of the central bank, and ϵ_t is a policy shock. We impose that the structural shocks are serially and mutually uncorrelated.[3]

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

$$Y_t = \Gamma\xi_t + \mathcal{R}\epsilon_t \ , \quad \text{with} \quad \Gamma = \begin{bmatrix} \frac{1-\kappa\phi_\xi/\sigma}{1+\kappa\phi_\pi/\sigma} \\ \frac{-\phi_\pi/\sigma-\phi_\xi/\sigma}{1+\kappa\phi_\pi/\sigma} \end{bmatrix} \ , \quad \mathcal{R} = \begin{bmatrix} \frac{-\kappa/\sigma}{1+\kappa\phi_\pi/\sigma} \\ \frac{-1/\sigma}{1+\kappa\phi_\pi/\sigma} \end{bmatrix} \ . \tag{4}$$

The vectors Γ and $\mathcal{R}$ capture the impulse responses of the policy objectives to the structural shocks ξ_t and ϵ_t. Note that Γ and $\mathcal{R}$ depend on both ϕ —the policy rule coefficients— as well as the parameters of the macro block (1)-(2).

In this example we will measure the performance of the central bank using the loss function

$$\mathcal{L}_t = \frac{1}{2}Y_t^{'}Y_t \ . \tag{5}$$

Given this loss function, an optimal reaction function is defined as any $\phi = (\phi_\pi, \phi_\xi)$ that minimizes the expected loss given the underlying structure of the economy, i.e., given equations (1)-(2). Formally, let $\Phi = \{\phi \in \mathbb{R}^2 : \phi_\pi > 1\}$, the set of optimal reaction functions is given by

$$\Phi^{\text{opt}} = \left\{ \phi : \phi \in \underset{\phi\in\Phi}{\operatorname{argmin}}\, \mathbb{E}\mathcal{L}_t \quad \text{s.t.} \quad (1)-(3) \text{ with } \epsilon_t = 0 \right\} \ ,$$

which is non-empty (e.g., Galí, 2015, page 133).

[3] In the web appendix, we show that this assumption is without loss of generality, as our approach can be re-written to accommodate more general (notably serially correlated) exogenous processes for ξ_t and ϵ_t.

Reaction function evaluation

We will now illustrate how the impulse responses $\mathcal{R}$ and Γ are sufficient statistics to evaluate a policy maker's reaction function.

Let $\phi^0 = (\phi^0_\pi, \phi^0_\xi) \in \Phi$ denote the reaction function chosen by the central bank. To evaluate ϕ^0, we consider a thought experiment where ϕ^0_ξ —the reaction coefficient to the cost-push shock— is adjusted by some amount τ. The adjusted policy rule becomes

$$i_t = \phi^0_\pi \pi_t + (\phi^0_\xi + \tau)\xi_t + \epsilon_t \ . \tag{6}$$

Following the same steps that led to (4), we can solve the model under that modified policy rule and express the endogenous variables as a function of exogenous shocks to get

$$Y_t = (\Gamma^0 + \mathcal{R}^0\tau)\xi_t + \mathcal{R}\epsilon_t \ , \tag{7}$$

where $\Gamma^0 \equiv \Gamma(\phi^0)$ and $\mathcal{R}^0 \equiv \mathcal{R}(\phi^0)$ denote the impulse responses to the structural shocks under the rule ϕ^0 and are defined as in (4).

From expression (7), we can see that $\Gamma^0 + \mathcal{R}^0\tau$ is the impulse response to cost-push shocks *after* the reaction function adjustment τ. In other words, the adjustment τ modifies the impulse response to cost-push shocks from Γ^0 to $\Gamma^0 + \mathcal{R}^0\tau$, which means that the impulse response $\mathcal{R}^0$ contains all the information needed to compute the effect of an adjustment to the rule coefficient ϕ_ξ. This insight, which holds more generally in a large class of dynamic models (see Section 4) is at the heart of our sufficient statistics approach to evaluating reaction function from structural impulse responses.

To evaluate the reaction function, the idea is then to compute whether it is possible to adjust ϕ^0_ξ and lower the loss function. Mathematically, we will look for a τ^* that can best lower the loss function, that is

$$\begin{aligned}\tau^* &= \underset{\tau}{\operatorname{argmin}} \ \mathbb{E}\mathcal{L}_t \qquad \text{s.t.} \qquad Y_t = (\Gamma^0 + \mathcal{R}^0\tau)\xi_t + \mathcal{R}\epsilon_t \\ &= \underset{\tau}{\operatorname{argmin}} \ (\Gamma^0 + \mathcal{R}^0\tau)'(\Gamma^0 + \mathcal{R}^0\tau) \ ,\end{aligned} \tag{8}$$

where the second equality uses that the structural shocks have mean zero and are uncorrelated. A closed form solution for τ^* is given by

$$\tau^* = -(\mathcal{R}^{0'}\mathcal{R}^0)^{-1}\mathcal{R}^{0'}\Gamma^0 \ . \tag{9}$$

We refer to the statistic τ^* as the *Optimal Reaction Adjustment*, or ORA, as it measures how much more (or less) the policy maker should have responded to the cost-push shock in order to minimize the loss function. Specifically, τ^* has the property that[4]

$$(\phi_\pi^0, \phi_\xi^0 + \tau^*) \in \Phi^{\text{opt}} \ . \tag{10}$$

In words, adjusting the reaction function ϕ^0 by τ^* makes the reaction function optimal.

A number of points are worth noting.

First, to evaluate a reaction function it is not necessary to know the model nor the policy rule, as the impulse responses to policy and non-policy shocks ($\mathcal{R}^0$ and Γ^0) are sufficient to evaluate a reaction function. If the reaction function ϕ^0 was optimal, there should not exist *any* alternative reaction to ξ_t that can reduce loss, and the optimal adjustment τ^* —a function of the impulse responses Γ^0 and $\mathcal{R}^0$ alone— should be zero. Further, because a τ^* adjustment makes the reaction function optimal, τ^* is a measure of policy performance, as it measures the distance to the optimal reaction coefficient ϕ_ξ^*.

Second, the formula for the ORA has a geometric interpretation. If Γ^0 is orthogonal to $\mathcal{R}^0$, the ORA τ^* is zero and the reaction coefficient ϕ_ξ is optimal. Intuitively, Γ^0 (the impulse response to a cost-push shock) captures what the policy maker *did* on average to counteract cost-push shocks —how cost-push shocks affected the economy under the prevailing policy rule—, while $\mathcal{R}^0$ (the impulse response to a monetary shock) captures what the policy maker *could have done* to counteract cost-push shocks —how adjusting the reaction coefficient ϕ_ξ by τ could have better stabilized the effect of cost-push shocks by transforming Γ^0 into $\Gamma^0 + \tau\mathcal{R}^0$—. If Γ^0 and $\mathcal{R}^0$ are orthogonal, there is nothing more that the policy maker could have done to stabilize Γ^0: the reaction function was optimal. Conversely, if the reaction coefficient ϕ_ξ^0 was not optimal, a *regression in impulse response space* —regressing one impulse response on another— can determine the optimal reaction to cost-push shocks. Indeed, it is easy to see that the ORA τ^* is the coefficient of a regression of Γ^0 on $-\mathcal{R}^0$: the goal of the ORA is to use the impulse responses to a monetary shock in order to best stabilize the impulse response to the cost-push shock. This is equivalent to best fitting the vector Γ^0 with the vector $-\mathcal{R}^0$.

Third, note how the ORA τ^* assesses the reaction function in one specific "direction" —the systematic policy response ϕ_ξ to cost-push shocks—. The ORA is not focused on

[4]To see this, compute

$$\begin{aligned}
\phi_\xi^0 + \tau^* &= \phi_\xi^0 - (\mathcal{R}^{0'}\mathcal{R}^0)^{-1}\mathcal{R}^{0'}\Gamma^0 \\
&= \phi_\xi^0 - \frac{-\kappa/\sigma(1-\kappa\phi_\xi^0/\sigma) - 1/\sigma(-\phi_\pi^0/\sigma - \phi_\xi^0/\sigma)}{\kappa^2/\sigma^2 + 1/\sigma^2} \\
&= \frac{\kappa/\sigma - \phi_\pi^0/\sigma^2}{\kappa^2/\sigma^2 + 1/\sigma^2} = \frac{\kappa\sigma - \phi_\pi^0}{\kappa^2+1} \ ,
\end{aligned}$$

and the adjusted reaction function is optimal as $(\phi_\pi^0, \frac{\kappa\sigma-\phi_\pi^0}{\kappa^2+1}) \in \Phi^{\text{opt}}$, see e.g., Galí (2015, p133, eq. (10)).

evaluating the reaction coefficient to endogenous variables, such as the optimality of the reaction coefficient ϕ_π. Evaluating the reaction to endogenous variables is generally more difficult, but an insight underlying our approach is that this is also not necessary. In this example cost-push shocks are the only non-policy shocks, and optimally responding to cost-push shocks is sufficient to characterize the *entire* optimal reaction function: $(\phi_\pi^0, \phi_\xi^0 + \tau^*) \in \Phi^{\text{opt}}$. In the general treatment of Section 4 where we allow for arbitrary many types of non-policy shocks, fully characterizing the optimal reaction function will require the impulse responses to all the different non-policy shocks. When this is not possible, focusing on a subset of these non-policy shocks will still allow to assess optimality in specific "directions": how well a policy maker responded to specific disturbances.

Finally, it may seem surprising to be able to assess a reaction function without specifying or estimating any policy rule. This reason this is possible, and the insight underlying our approach, is that the effects of any reaction function are *encoded* in the impulse responses Γ^0 and $\mathcal{R}^0$, see (4) with Γ and $\mathcal{R}$ depending on ϕ_π and ϕ_ξ. Thus, even if we do not know the specific form of some past policy rule, that reaction function left a footprint on the effects of policy and non-policy shocks, and that footprint is sufficient to evaluate the reaction function. This is the essence of our sufficient statistics approach.

Comparing reaction functions

The ORA statistic can be used to compare the reaction functions of different policy makers, i.e., to compare the performances of policy makers after their term. To avoid excessive notation at this stage, consider comparing two policy makers that used reaction functions ϕ^0 and ϕ^1, respectively, and let the (possibly different) economic environment that they faced be captured by the parameter vectors θ^0 and θ^1, respectively, which include all coefficients in the Phillips and IS curves.

For each policy maker j we compute the ORA statistic:

$$\tau_j^* = -(\mathcal{R}^{j\prime}\mathcal{R}^j)^{-1}\mathcal{R}^{j\prime}\Gamma^j \qquad \text{for} \qquad j = 0, 1 \ ,$$

where $\mathcal{R}^j \equiv \mathcal{R}(\phi^j, \theta^j)$ and $\Gamma^j \equiv \Gamma(\phi^j, \theta^j)$; making the dependence on θ^j explicit, see (4).

Since the ORA measures the distance to the optimal reaction *given* the economic environment, we can use the ORA to rank policy makers who served in different environments. For instance, if $|\tau_1^*| < |\tau_0^*|$, we would rank policy maker 1 above policy maker 0: faced with the same exogenous disturbance, policy maker 1 reacted better than policy maker 0. Naturally, for this reaction function comparison to work, the researcher must be able to identify the same types of non-policy shocks across policy makers.

The key insight is that while environments can be different across policy makers (and thus the optimal reaction function can be different), the ORA statistics "control" for the economic environment: τ_0^* and τ_1^* measure how well each policy maker reacted to a unit cost-push shock *given* their economic environments θ^0 and θ^1 (respectively). By comparing these distances, we can compare policy makers (or more generally policy institutions) who were facing different initial conditions, different shock realizations and different economic environments during their term in office.

In sum, this example illustrates how we can (i) evaluate and (ii) compare policy makers based on their reaction function without specifying an explicit reaction function nor a specific structural macro model. Instead, the only requirement is to estimate two sufficient statistics: the impulse responses Γ and $\mathcal{R}$ over a policy maker's term. The next sections show that these findings continue to hold for a general linear macro model and discuss the econometric implementation.

3 Environment

We describe a general stationary macro environment for a single policy maker (or institution) who faces an infinite horizon economy. To describe the economy we distinguish between two types of observable variables: policy instruments $p_t \in \mathbb{R}^{M_p}$ and non-policy variables $y_t \in \mathbb{R}^{M_y}$. The policy instruments are different from the other variables as they are under the direct control of the policy maker.

To describe a forward looking economy we use a sequence space representation (e.g., Auclert et al., 2021). Let $\mathbf{P} = (p_0', p_1', \ldots)'$ and $\mathbf{Y} = (y_0', y_1', \ldots)'$ denote the paths for the policy instruments and non-policy variables. Working under perfect foresight, we consider a generic model for the paths of the endogenous variables

$$\begin{array}{rcl} \mathcal{A}_{yy}\mathbf{Y} - \mathcal{A}_{yp}\mathbf{P} & = & \mathcal{B}_{y\xi}\boldsymbol{\Xi} \\ \mathcal{A}_{pp}\mathbf{P} - \mathcal{A}_{py}\mathbf{Y} & = & \mathcal{B}_{p\xi}\boldsymbol{\Xi} + \boldsymbol{\epsilon} \end{array} , \tag{11}$$

where $\boldsymbol{\epsilon} = (\epsilon_0', \epsilon_1', \ldots)'$ and $\boldsymbol{\Xi} = (\xi_0', \xi_1', \cdots)'$ are sequences of policy and non-policy shocks, respectively. The first equation captures the non-policy block of the economy, while the second equation captures the policy rule.

We normalize all elements of $\boldsymbol{\Xi}$ and $\boldsymbol{\epsilon}$ to have mean zero and unit variance. Also, we assume that they are serially and mutually uncorrelated, consistent with the common definition of structural shocks (e.g. Bernanke, 1986; Ramey, 2016).[5] The structural maps

[5]Note that if the elements of $\boldsymbol{\Xi}$ or $\boldsymbol{\epsilon}$ are not serially uncorrelated it is always possible to redefine $\mathcal{B}_{y\xi}, \mathcal{B}_{p\xi}$ and $\mathcal{A}_{pp}$ such that the equation residuals —the shocks— are uncorrelated. For example if var$(\boldsymbol{\Xi}) = \Sigma$, then redefine $\mathcal{B}_{y\xi}\boldsymbol{\Xi} = \tilde{\mathcal{B}}_{y\xi}\tilde{\boldsymbol{\Xi}}$ with $\tilde{\mathcal{B}}_{y\xi} = \mathcal{B}_{y\xi}\Sigma^{1/2}$ and $\tilde{\boldsymbol{\Xi}} = \Sigma^{-1/2}\boldsymbol{\Xi}$ such that $\tilde{\boldsymbol{\Xi}}$ is serially uncorrelated. The same can be done for $\mathcal{B}_{p\xi}\boldsymbol{\Xi}$.

$\mathcal{A}_{..}$ and $\mathcal{B}_{..}$ are conformable and may depend on underlying structural parameters. We conveniently split them in two parts: the economic environment $\theta = \{\mathcal{A}_{yy}, \mathcal{A}_{yp}, \mathcal{B}_{y\xi}\}$ which the policy maker takes as given, and the reaction function $\phi = \{\mathcal{A}_{pp}, \mathcal{A}_{py}, \mathcal{B}_{p\xi}\}$, which is under the control of the policy maker. We impose that ϕ and θ are independent in the sense that $\partial\theta_i/\partial\phi_j = 0$ for all entries i, j, i.e. changing the reaction function does not directly change the coefficients θ and all effects of ϕ on $\mathbf{Y}$ go via the policy path $\mathbf{P}$.

We denote by Φ the set of all reaction functions ϕ for which the model (11) implies a unique equilibrium, that is all ϕ for which

$$\mathcal{A} = \begin{pmatrix} \mathcal{A}_{yy} & \mathcal{A}_{yp} \\ \mathcal{A}_{py} & \mathcal{A}_{pp} \end{pmatrix} \qquad \text{is invertible.}$$

Many structural models found in the literature can be written in the form of (11); prominent examples include New Keynesian models and heterogeneous agents models, see McKay and Wolf (2023) for a more in depth discussion.

For any $\phi \in \Phi$ we can write the expected path of the non-policy variables as a linear function of the policy and non-policy shocks:

Lemma 1. *Given the generic model* (11) *with* $\phi \in \Phi$*, we have*

$$\mathbf{Y} = \Gamma(\phi)\boldsymbol{\Xi} + \mathcal{R}(\phi)\boldsymbol{\epsilon} \ . \tag{12}$$

The maps $\Gamma(\phi)$ and $\mathcal{R}(\phi)$ capture the causal effects of the structural shocks $\boldsymbol{\Xi}$ and $\boldsymbol{\epsilon}$ on the non-policy variables. Note the similarity between (12) and (4), as the illustrative static NK model is a special case with only contemporaneous shocks. Clearly, the maps $\Gamma(\phi)$ and $\mathcal{R}(\phi)$ in (12) also depend on the environment as summarized by θ, but since θ is not under the control of the policy maker we omit this from the notation.

Lemma 1 implies that the identification of the impulse responses requires observing some part of the *future* shocks in $\boldsymbol{\Xi}$ and $\boldsymbol{\epsilon}$. Our perfect foresight notation masks this requirement, but it is useful to clarify that in practice this requires the identification of news shocks. To see this, note that we can decompose ξ_t and ϵ_t as[6]

$$\xi_t = \sum_{j=0}^{t} \underbrace{\mathbb{E}_j \xi_t - \mathbb{E}_{j-1}\xi_t}_{\xi_{t,j}} \qquad \text{and} \qquad \epsilon_t = \sum_{j=0}^{t} \underbrace{\mathbb{E}_j \epsilon_t - \mathbb{E}_{j-1}\epsilon_t}_{\epsilon_{t,j}} \ , \tag{13}$$

where $\mathbb{E}_j(\cdot) = \mathbb{E}(\cdot|\mathcal{F}_j)$, with $\mathcal{F}_j$ the information set available at time j. The increment $\xi_{t,j} \equiv \mathbb{E}_j\xi_t - \mathbb{E}_{j-1}\xi_t$ is the component of ξ_t that is released at time $j \leq t$. In other words $\xi_{t,j}$

[6]As is common in the optimal policy literature, we impose $\mathbb{E}_{-1}\xi_t = 0$ and $\mathbb{E}_{-1}\epsilon_t = 0$, for all $t = 0, 1, \ldots$. Alternatively, one could let the sums run from $-\infty$ until t.

is a news shock released at $j \leq t$, and (13) decomposes the shock ξ_t —a shock realized at time t— as a sum of news shocks $\xi_{t,j}$ revealed all the way until time t with $\xi_t = \sum_{j=0}^{t} \xi_{t,j}$. Similarly for $\epsilon_{t,j}$. By construction the news shocks are serially uncorrelated.

Thus, to identify the impulse responses in (12), we require observing proxies for the news shocks in $\boldsymbol{\xi}_0 = (\xi_{0,0}, \xi_{1,0}, \xi_{2,0}, \ldots)'$ and $\boldsymbol{\epsilon}_0 = (\epsilon_{0,0}, \epsilon_{1,0}, \epsilon_{2,0}, \ldots)'$.[7] For notational convenience we drop the zero subscript and work under perfect foresight.

Evaluation criteria

We consider a researcher who is interested in evaluating a policy maker based on her success at stabilizing some subset of the non-policy variables y_t around some desired targets y^* for some time periods $t = 0, 1, 2, \ldots$. For ease of notation we will set the targets to zero, though we could also think of y_t as defined in deviation from the desired targets. In general, we will see that the target values y^* are not needed to rank/assess reaction functions.

We measure performance using the unconditional loss function

$$\mathcal{L} = \frac{1}{2}\mathbb{E}\mathbf{Y}'\mathcal{W}\mathbf{Y} \ , \tag{14}$$

where $\mathcal{W}$ is a diagonal matrix, with non-negative entries, which selects and weights the specific variables and horizons that are of interest to the researcher. The loss (14) is the researcher's evaluation criterion for scoring policy maker performance —an input into our framework—, and it may or may not correspond to the preferences of society or the policy maker.

The actions of the policy maker are summarized by the reaction function ϕ. We define a reaction function to be optimal if it minimizes the loss function (14) when $\boldsymbol{\epsilon} = \mathbf{0}$.[8] Formally, the set of optimal reaction functions is given by

$$\Phi^{\text{opt}} = \left\{ \phi : \phi \in \underset{\phi \in \Phi}{\operatorname{argmin}}\, \mathcal{L} \quad \text{s.t} \quad (11) \text{ with } \boldsymbol{\epsilon} = \mathbf{0} \right\} \ . \tag{15}$$

The definition implies that we only consider optimal reaction functions that lie in Φ; the set of reaction functions which imply a unique equilibrium.

[7] Note that in practice our approach will not require the identification of all news shocks. In fact, it will not even require all news shocks to even exist; our methodology can be applied using as little as one policy and one non-policy shock.

[8] In other words, we will be evaluating policy makers based on how well they reacted to exogenous disturbances, in line with the thought experiment discussed in the introduction. An alternative (which we do not pursue here) would be to evaluate policy makers based on their idiosyncratic exogenous mistakes; i.e., based on the policy shocks.

4 Measuring reaction function optimality

We propose to evaluate a policy maker after her term by measuring the distance between her reaction function, denoted by ϕ^0, and the set of optimal reaction functions Φ^{opt}. We postulate that each policy maker faces an economy that can be represented by the generic model (11) where the parameters θ and ϕ may vary across policy makers. We first develop the methodology for evaluating the reaction function of a single policy maker in population. Subsequently we formalize how the methodology can be used to rank the performance of multiple policy makers.

Following the same steps as the simple example of Section 2, we propose to measure the distance between ϕ^0 and Φ^{opt} by considering a thought experiment where we adjust the policy maker's reaction coefficients for non-policy shocks. Specifically, consider augmenting the policy rule under ϕ^0 as follows

$$\mathcal{A}^0_{pp}\mathbf{P} - \mathcal{A}^0_{py}\mathbf{Y} = (\mathcal{B}^0_{p\xi} + \mathcal{T})\mathbf{\Xi} + \boldsymbol{\epsilon} \ , \tag{16}$$

where $\mathcal{T}$ adjusts the response to the non-policy shocks.[9] Each element of $\mathcal{T}$ corresponds to a different rule counterfactual, in which we modify how one element of the policy path responds to one of the non-policy shocks.

The following lemma establishes how a rule adjustment $\mathcal{T}$ affects the equilibrium allocation

Lemma 2. *Consider the generic model* (11) *with* $\phi^0 \in \Phi$ *and the modified policy rule* (16)*.*

1. *We have*

$$\mathbf{Y} = (\Gamma^0 + \mathcal{R}^0\mathcal{T})\mathbf{\Xi} + \mathcal{R}^0\boldsymbol{\epsilon} \ , \tag{17}$$

where $\Gamma^0 \equiv \Gamma(\phi^0)$ *and* $\mathcal{R}^0 \equiv \mathcal{R}(\phi^0)$*.*

2. *Given an element* τ_{ij} *of* $\mathcal{T}$*, we have*

$$\frac{\partial \Gamma^0_j}{\partial \tau_{ij}} = \mathcal{R}^0_i \ , \qquad \frac{\partial \Gamma^0_k}{\partial \tau_{ij}} = 0 \qquad \textit{for} \qquad i, j, k = 0, 1, \ldots \ , \ k \neq j \ . \tag{18}$$

where Γ^0_j *and* $\mathcal{R}^0_i$ *are (respectively) the* j*th and* i*th columns of* Γ^0 *and* $\mathcal{R}^0$*.*

[9] To help understand the elements of $\mathcal{T}$ in this sequence-space representation, imagine that there is only one policy instrument and one type of non-policy shock: an oil price shock. The upper-left element of $\mathcal{T}$ (τ_{00}) is an adjustment to the contemporaneous response of the policy instrument to a contemporaneous oil shock. The element τ_{01} is an adjustment to the contemporaneous response of the policy instrument to a news shock announced today but affecting oil prices next period, the element τ_{10} is an adjustment to the response of the policy instrument next period to a contemporaneous oil shock, and so on.

The reaction adjustment $\mathcal{T}$ affects the equilibrium by changing the impulse responses to non-policy shocks from Γ^0 to $\Gamma^0 + \mathcal{R}^0\mathcal{T}$, so that knowledge of the impulse response matrix $\mathcal{R}^0$ is sufficient to compute the policy rule counterfactuals embedded in the $\mathcal{T}$ adjustments. This property echoes the general result of McKay and Wolf (2023), who show that it is possible to reproduce any policy rule counterfactual with an appropriate combination of policy news shocks at different horizons.[10] The information requirement for such impulse-based rule counterfactuals is typically very large however —requiring the identification of all policy news shocks—, and the counterfactual effects can only be approximated, often without bounds on the approximation error (McKay and Wolf, 2023). Lemma 2 shows that there exists a class of rule counterfactuals —counterfactual reactions to non-policy shocks—, which can be constructed *exactly* with modest information requirements. Indeed, as shown by (18) the rule counterfactual embedded in τ_{ij} leaves all impulse responses unaffected except for one —the impulse response Γ_j^0— and the effect of that rule adjustment on Γ_j^0 is linear and given by the impulse response $\mathcal{R}_i^0$. Constructing the counterfactual thus only requires the identification of one type of policy shock at a time. This property will be key to be able to assess and compare policy makers in practice.

4.1 The ORA statistic

The *Optimal Reaction Adjustment* (ORA) is defined as the $\mathcal{T}$ that minimizes the loss function.

$$\mathcal{T}^* = \underset{\mathcal{T}}{\operatorname{argmin}}\, \mathcal{L} \qquad \text{s.t.} \quad \mathbf{Y} = (\Gamma^0 + \mathcal{R}^0\mathcal{T})\boldsymbol{\Xi} + \mathcal{R}^0\boldsymbol{\epsilon}\ , \tag{19}$$

The ORA determines how the reaction coefficients in front of the non-policy shocks $\boldsymbol{\Xi}$ should have been adjusted to minimize the loss.

Since the setting is linear-quadratic a closed form solution for $\mathcal{T}^*$ is given by

$$\mathcal{T}^* = -(\mathcal{R}^{0'}\mathcal{W}\mathcal{R}^0)^{-1}\mathcal{R}^{0'}\mathcal{W}\Gamma^0\ , \tag{20}$$

which exists provided that the inverse exists. The expression shows that the ORA is equal to the (weighted) least-square regression of the selected non-policy impulse responses Γ^0 on the selected policy impulse responses $\mathcal{R}^0$.[11]

We have the following result:

[10] The derived counterfactual is robust to the Lucas critique provided that the coefficients of the macro block (here θ) are invariant to changes in the coefficients of the policy rule (here, ϕ). This property holds in most modern macro models as in our generic model (11).

[11] The weighting matrix $\mathcal{W}$ is merely a selection tool used to select the non-policy variables that are of interest to the researcher.

Proposition 1. *Given the generic model* (11)*, with* Φ *non-empty, we have that* $\phi^* \in \Phi^{\text{opt}}$ *where* $\phi^* = \{\mathcal{A}^0_{pp}, \mathcal{A}^0_{py}, \mathcal{B}^0_{p\xi} + \mathcal{T}^*\}$.

The proposition states that it is possible to fully characterize the optimal reaction function from $\mathcal{B}_{p\xi}$ alone, and it is not necessary to optimize with respect to the maps $\mathcal{A}_{pp}$ and $\mathcal{A}_{py}$ (as long as the invertibility requirement for $\mathcal{A}$ is satisfied). In other words, the class of policy rule counterfactuals embodied in (16) —counterfactual reactions to non-policy shocks— is sufficient to fully characterize the optimal policy rule, and our focus on that sub-class of rule counterfactuals is without loss of generality for policy rule evaluation.

4.2 Subset optimal reaction adjustments

So far we showed that the optimal reaction function can be recovered from the impulse responses to policy and non-policy shocks. In practice however we may not be able to identify all shocks, and hence the impulse responses to all policy and non-policy shocks.

Fortunately, the class of rule counterfactuals that we consider—counterfactual reactions to non-policy shocks— allows us to split the optimal policy rule problem into orthogonal problems that can be solved separately; each sub-problem focusing on the optimal reaction of one element of the policy path in response to one type of non-policy shock. In this section we provide a policy rule evaluation statistic that requires only a *subset* of all impulse responses. This property will allow to operationalize our approach with limited information requirements.

To set this up, let $\boldsymbol{\epsilon}_a$ denote any subset of $\boldsymbol{\epsilon}$ which can be identified. Similarly, let $\boldsymbol{\Xi}_b$ denote a subset of $\boldsymbol{\Xi}$. Our subset approach consists in adjusting a subset of the coefficients of the policy rule with

$$\mathcal{A}^0_{p_a p}\mathbf{P} - \mathcal{A}^0_{p_a y}\mathbf{Y} = (\mathcal{B}^0_{p_a \xi_b} + \mathcal{T}_{ab})\boldsymbol{\Xi}_b + \mathcal{B}^0_{p_a \xi_{-b}}\boldsymbol{\Xi}_{-b} + \boldsymbol{\epsilon}_a \ , \tag{21}$$

where $\mathcal{T}_{ab}$ adjusts the ϕ^0 response to the non-policy shocks $\boldsymbol{\Xi}_b$ and $\boldsymbol{\Xi}_{-b}$ denotes all other non-policy shocks. Note that all other equations of the policy block, i.e. those corresponding to $\boldsymbol{\epsilon}_{-a}$, are unchanged and only the equations corresponding to $\boldsymbol{\epsilon}_a$ are adjusted by $\mathcal{T}_{ab}$.

Following the same steps as above we can define the subset ORA as the $\mathcal{T}_{ab}$ that minimizes the expected loss function.

$$\mathcal{T}^*_{ab} = \underset{\mathcal{T}_{ab}}{\operatorname{argmin}}\, \mathcal{L} \qquad \text{s.t.} \quad \mathbf{Y} = (\Gamma^0_b + \mathcal{R}^0_a \mathcal{T}_{ab})\boldsymbol{\Xi}_b + \Gamma^0_{-b}\boldsymbol{\Xi}_{-b} + \mathcal{R}^0\boldsymbol{\epsilon} \ , \tag{22}$$

The ORA determines how the reaction coefficients in front of the non-policy shocks $\boldsymbol{\Xi}_b$ should have been adjusted to minimize the unconditional loss. A closed form solution for the subset

ORA is given by

$$\mathcal{T}_{ab}^{*} = -(\mathcal{R}_a^{0\prime}\mathcal{W}\mathcal{R}_a^0)^{-1}\mathcal{R}_a^{0\prime}\mathcal{W}\Gamma_b^0 \ , \tag{23}$$

which exists provided that the inverse exists. The subset ORA has the following properties:

Proposition 2. *Consider the generic model* (11) *with* Φ *non-empty.*

1. *We have*

$$\mathcal{B}_{p_a\xi_b}^0 + \mathcal{T}_{ab}^{*} = \underset{\mathcal{B}_{p_a\xi_b}}{\operatorname{argmin}}\, \mathbb{E}\mathcal{L} \ .$$

2. *For* $\phi_{ab}^{*} = \{\mathcal{A}_{pp}^0, \mathcal{A}_{py}^0, \mathcal{B}_{p_a\xi_b}^0 + \mathcal{T}_{ab}^{*}, \mathcal{B}_{-p_a-\xi_b}^0\}$*, we have that* $\mathcal{L}(\phi_{ab}^{*}) \leq \mathcal{L}(\phi^0)$ *for all* $\phi^0 \in \Phi$.

The first part of Proposition 2 states that it is possible to split the optimal policy rule problem into smaller orthogonal problems, which can be solved exactly from a subset of identified impulse responses. For instance, consider a researcher interested in evaluating how a central bank is using its contemporaneous policy rate in reaction to contemporaneous oil price shocks. If we denote by τ_{ij} the corresponding rule adjustment, the only requirements to compute τ_{ij}^{*} (and thereby compute the optimal reaction to oil price shocks) are two sets of impulse responses: the impulse responses to a contemporaneous policy shock and the impulse response to an oil price shock.

Geometrically, each rule counterfactual embedded in τ_{ij} amounts to projecting the optimal policy problem on a space spanned by one type of non-policy shocks. Since non-policy shocks are orthogonal, the projection spaces corresponding to different non-policy shocks are all orthogonal, meaning that our approach effectively splits the optimal policy problem into orthogonal problems. The two important realizations underlying our approach are (i) each sub-problem is much easier to solve —requiring the identification of only one policy and non-policy shock at a time— (Proposition 2, part 1), and (ii) the entire policy problem can be reconstructed through this approach: in the limit where we can identify all shocks, the projection steps span the entire optimal policy problem (Proposition 1).[12]

[12] Specifically, the ORA has a two-step geometric interpretation. In a first step we project $\mathbf{Y}$ on specific non-policy shocks, which gives under ϕ^0

$$\mathbb{E}[\mathbf{Y}\boldsymbol{\Xi}_b'] = \Gamma_b^0 + \mathcal{R}_a^0\mathcal{T}_{ab} \ .$$

This step effectively isolates the economy's response to a subset of non-policy shocks, thereby removing the confounding effects of the other shocks ($\boldsymbol{\Xi}_{-b}$ and $\boldsymbol{\epsilon}_t$). In the second step we solve the policy problem in that projected space —in "impulse response space"—, i.e. we solve

$$\mathcal{T}_{ab}^{*} = \underset{\mathcal{T}_{ab}}{\operatorname{argmin}}\ (\Gamma_b^0 + \mathcal{R}_a^0\mathcal{T}_{ab})'\mathcal{W}(\Gamma_b^0 + \mathcal{R}_a^0\mathcal{T}_{ab}) \ ,$$

which effectively finds a rule adjustment that best stabilizes the impulse responses to the non-policy shocks $\boldsymbol{\Xi}_b$. In the web-appendix we discuss a few additional interpretations for the ORA statistic.

The second part of Proposition 2 states that subset ORA adjustment will improve the policy rule, but it may not deliver the optimal rule, as the subset ORA only assesses the policy rule in specific directions. Loosely speaking, the larger the number of directions —the larger are the "subsets" a and b—, and the more "exhaustive" the evaluation will be. In the limiting case where all policy and non-policy shocks can be identified, the ORA adjustment delivers the optimal reaction function (Proposition 1 applies) and the policy evaluation is "exhaustive".

4.3 Comparing policy institutions with ORAs

Having established the ORA's properties, we now discuss how the ORA can be used to compare policy institutions or policy makers. As examples we can think of evaluating different central banks chairs based on their ability to control inflation and output gaps, or different presidents of a country based on their ability to keep output close to potential. More generally, we can compare policy makers from the same institution across different time periods or policy maker from comparable institutions in different countries.

Suppose that there are p policy makers that the researcher aims to compare. Each policy maker operates an economy that can be described by the general model (11), but the parameters θ and ϕ that govern the model may vary across policy makers, say θ_j and ϕ_j, for $j = 0, \ldots, p$, where ϕ_j denote the reaction function chosen by policy maker j.[13]

Since the ORA measures the distance to an optimal reaction function, we can use the ORA to compare policy makers. Effectively, this will amount to comparing policy makers from the way they responded to common shocks that hit during their term, in line with the thought experiment sketched in the introduction. Going back to our oil shock example, the idea will be to compare central bankers from the way they each used their contemporaneous policy rate in response to oil shocks. This will require estimating, for each policy maker, the impulses responses to shocks to the contemporaneous policy rate and impulse responses to oil shocks.

In general, the subset ORA statistics for each policy maker are given by:[14]

$$\mathcal{T}_{ab}^{j*} = -(\mathcal{R}_a^{j'} \mathcal{W} \mathcal{R}_a^j)^{-1} \mathcal{R}_a^{j'} \mathcal{W} \Gamma_b^j \ , \qquad j = 0, \ldots, p \ ,$$

where $\mathcal{R}_a^j$ and Γ_b^j are the impulse responses with respect to the policy and non-policy shocks under the reaction function ϕ_j and the economic environment θ_j.

[13]While we treat the parameters as fixed within the term of each policy maker, an extension with time-varying parameters can be easily constructed at the expense of more notation.

[14]The weighting by $\mathcal{W}$ implements the preferences of the researcher over the different objectives or ranking criteria.

Note that the impulse responses $\mathcal{R}_a^j = \mathcal{R}_a(\phi^j, \theta^j)$ and $\Gamma_b^j = \Gamma_b(\phi^j, \theta^j)$ depend on the policy rule parameters and the economic environment, which can and will vary across policy makers. The point of the ORA is to precisely take this variation into account, computing each policy maker's optimal reaction function *given* the economic environment. That said, to ensure that the ORAs are comparable we need to ensure that the impulse responses correspond to the same policy and non-policy shocks, i.e. $\boldsymbol{\epsilon}_a$ and $\boldsymbol{\Xi}_b$ must be of the "type" a and b across policy makers: the ORAs must measure the optimality of the *same* policy instrument in response to the *same* exogenous disturbance. This requirement is important, yet it is no different from numerous earlier works on time-varying impulse responses (e.g. Cogley and Sargent, 2005; Primiceri, 2005) or country-specific impulse responses (e.g. Ilzetzki, Mendoza and Végh, 2013). All such studies rely on the assumption that it is possible to identify the same shocks across time or space.

Note that each element of $\mathcal{T}_{ab}^{j*}$ is informative about a specific dimension of policy. For instance, the first element of the ORA could compare how well policy makers used their policy rate following an oil shock, the second element how well policy makers used their policy rate following a financial shock, etc... To obtain a summary ranking, we can aggregate the entries of $\mathcal{T}_{ab}^{j*}$, i.e.

$$t_{ab}^{j*} = \|\mathcal{T}_{ab}^{j*}\| \ , \tag{24}$$

where any desired norm $\|\cdot\|$ can be used. We can then rank policy makers based on t_{ab}^{j*}, for $j = 0, \ldots, p$, where the smallest value corresponds to the best performing policy maker.

4.4 Computing ORA statistics

An attractive feature of the ORA is that it can be readily computed from standard econometric methods. The sufficient statistics underlying the ORA —impulse responses to structural shocks— are well studied statistics, and we can draw on a large macro-econometric literature precisely devoted to the estimation of these statistics, from the identification of structural shocks (e.g., Ramey, 2016) to the estimation of impulse responses (e.g., Li, Plagborg-Møller and Wolf, 2022).

To make this clear, consider the equilibrium representation (12) under some rule ϕ

$$\mathbf{Y} = \Gamma_b \boldsymbol{\Xi}_b + \Gamma_{-b} \boldsymbol{\Xi}_{-b} + \mathcal{R}_a \boldsymbol{\epsilon}_a + \mathcal{R}_{-a} \boldsymbol{\epsilon}_{-a} \ ,$$

where the entries of $\mathcal{R}_a$ and Γ_b are equal to projection of the variables $\mathbf{Y}$ on the subset shocks $\boldsymbol{\epsilon}_a$ or $\boldsymbol{\Xi}_b$. For convenience we assume that the researcher is interested in a finite number of variables such that $\mathcal{W}$ has a finite number of non-zero diagonal elements and we let $\mathbf{Y}^w$ be the finite collection of selected elements of $\mathcal{W}^{1/2}\mathbf{Y}$. Further, let $\mathcal{R}_a^w$ and Γ_b^w denote the subset causal effects corresponding to the selected rows of $\mathcal{W}^{1/2}\mathcal{R}_a$ and $\mathcal{W}^{1/2}\Gamma_b$.

To compute the subset impulse responses we rely on a sample of realizations of the outcome variables $\mathbf{Y}^w$ during the policy makers term, i.e. $\{\mathbf{Y}_t^w, t = t_s, \ldots, t_e\}$ with t_s the starting period and t_e the ending period. The subset causal effects can be estimated by considering

$$\mathbf{Y}_t^w = \Gamma_b^w \boldsymbol{\Xi}_{b,t} + \mathcal{R}_a^w \boldsymbol{\epsilon}_{a,t} + \mathbf{V}_t^w \ , \qquad t = t_s, \ldots, t_e, \tag{25}$$

where $\boldsymbol{\Xi}_{b,t}$ and $\boldsymbol{\epsilon}_{a,t}$ are the subset of news shocks that are realized at time t and $\mathbf{V}_t^w$ includes all other structural shocks, both policy and non-policy inputs that are not included in the selections a and b, respectively, as well as initial conditions and future errors.

We can recognize (25) as a system of stacked local projections (Jordà, 2005). This implies that given (i) an appropriate identification strategy and (ii) an accompanying estimation method, we can estimate the impulse responses $\mathcal{R}_a^w$ and Γ_b^w. Any identification strategy — short run, long run, sign, external instruments, etc — can be used, based on which an appropriate estimation method — OLS or IV, with or without shrinkage, etc — can be selected, see Ramey (2016) and Stock and Watson (2018) for different options. Moreover, since local projections and structural VARs estimate the same impulse responses in population (Plagborg-Møller and Wolf, 2021), SVAR methods (e.g., Kilian and Lütkepohl, 2017) can also be adopted for estimating the impulse responses Γ_b^w and $\mathcal{R}_a^w$. Given such estimates we compute the ORA noting that $\mathcal{T}_{ab}^* = -(\mathcal{R}_a'\mathcal{W}\mathcal{R}_a)^{-1}\mathcal{R}_a'\mathcal{W}\Gamma_b = -(\mathcal{R}_a^{w\prime}\mathcal{R}_a^w)^{-1}\mathcal{R}_a^{w\prime}\Gamma_b^w$.

Here we will not discuss any specific approach but instead directly postulate that the researcher is able to obtain estimates, say $\widehat{\mathcal{R}}_a^w$ and $\widehat{\Gamma}_b^w$, of which the distribution can be approximated by

$$\text{vec}\left(\begin{bmatrix} \widehat{\mathcal{R}}_a^w \\ \widehat{\Gamma}_b^w \end{bmatrix} - \begin{bmatrix} \mathcal{R}_a^w \\ \Gamma_b^w \end{bmatrix} \right) \overset{a}{\sim} F \ ,$$

where F is some known distribution function that can be estimated consistently by $\widehat{F}$. Such approximation can be obtained for many impulse response estimators using either frequentist (asymptotic and bootstrap) or Bayesian methods.

Using the approximating distribution $\widehat{F}$, we can simulate draws for $\mathcal{R}_a^w$ and Γ_b^w, and compute $\mathcal{T}_{ab}^* = -(\mathcal{R}_a^{w\prime}\mathcal{R}_a^w)^{-1}\mathcal{R}_a^{w\prime}\Gamma_b^w$ for each draw. Given the sequence of draws we can construct a confidence set for $\mathcal{T}_{ab}^*$, or any of its individual entries at any desired level of confidence.

4.5 ORA-based counterfactuals

The ORA statistic measures directly how the reaction to the identified non-policy shocks should be adjusted. The key benefit is that this metric is comparable across policy makers. The price to pay for such invariance is that the statistic does not have a simple economic interpretation in terms of percentage points adjustments to the policy instrument or improvements in the loss function.[15]

That said, the ORA statistics can be used for computing such counterfactuals.

First, one can compute the adjusted (i.e., "ORA-improved") impulse responses to non-policy shocks —$\Gamma_b^w + \mathcal{R}_a^w \mathcal{T}_{ab}^*$—, which measure how the average responses to the different non-policy shock could have been adjusted. In practice, we recommend to report both Γ_b^w and $\Gamma_b^w + \mathcal{R}_a^w \mathcal{T}_{ab}^*$ to highlight how the ORA would have changed the dynamic effects of the non-policy shocks.

Second, one can compute ORA-adjusted historical paths for the variables of interest. Using series of identified non-policy shocks, we can use the ORA-adjusted impulse responses to quantify how much of the historical variation in $\mathbf{Y}_t^w$ could have been avoided with a different (i.e., ORA-improved) reaction function, or equivalently how much of the variation in $\mathbf{Y}_t^w$ was due to a sub-optimal reaction function and thus "unnecessary".[16] Specifically, given the identified non-policy shocks $\boldsymbol{\Xi}_{b,t}$, we can compute

$$\Delta \mathbf{Y}_t^w = \mathcal{R}_a^w \mathcal{T}_{ab}^* \boldsymbol{\Xi}_{b,t} \qquad \text{and} \qquad \Delta \mathbf{P}_t^w = \mathcal{R}_{p,a}^w \mathcal{T}_{ab}^* \boldsymbol{\Xi}_{b,t} \ , \quad \text{for} \quad t = t_s, \dots, t_e \ , \tag{26}$$

where $\Delta \mathbf{Y}_t^w$ and $\Delta \mathbf{P}_t^w$ are the ORA historical adjustments to the policy objectives and the policy instruments (respectively), and where $\mathcal{R}_{p,a}^w$ are the impulse responses of the policy instruments to the subset of policy news shocks.

With these ORA-adjusted paths in hand, we can also compute how much of the realized loss could have been avoided with a different reaction function:

$$\Delta \mathcal{L}_t = (\Delta \mathbf{Y}_t^w)'(\Delta \mathbf{Y}_t^w) \ . \tag{27}$$

We stress that the magnitudes of these counterfactuals $\Delta \mathbf{Y}_t^w$, $\Delta \mathbf{P}_t^w$ and $\Delta \mathcal{L}_t$ cannot be used to compare policy makers across periods. The reason is that if the economic environments are different across periods (as is most likely the case), a given Optimal Reaction Adjustment can have different effects on the endogenous variables: a given optimization failure may have smaller or large effects on welfare depending on the economic environment as well as the other parameters of the policy rule. In other words, while the ORAs are comparable across periods —depending *only* on how well the policy maker reacted to a specific non-policy shock—, the counterfactuals $\Delta \mathbf{Y}_t^w$ and $\Delta \mathbf{P}_t^w$ are not, because they are affected by other factors outside the policy maker's control.

[15]The ORA is an adjustment to the policy rule coefficients in front of non-policy shocks, but since the policy rule also includes responses the endogenous variables (and thus feedback loops), the ORA adjustment will generally not translate into a one-for-one change in the policy rate.

[16]If we only have a subset of all non-policy shocks, this exercise will provide a lower bound on the additional variation caused by a sub-optimal reaction function.

5 Evaluating US monetary policy, 1879-2019

In this section we use our methodology to evaluate the conduct of monetary policy in the US over the 1879-2019 period. We consider four distinct periods: (i) the Gold Standard period 1879-1912 before the creation of the Federal Reserve, (ii) the early Fed years 1913-1941, (iii) the post World War II period 1954-1984 and (iv) the post-Volcker period 1990-2019.

During the Gold Standard period, there was no active monetary policy (the Federal Reserve did not exist yet), and we use this period as a benchmark to see what a fictional policy institution could have done in this period. The Gold Standard monetary regime is now generally considered a sub-optimal regime with excessive fluctuations in inflation and unemployment (e.g. Friedman and Schwartz, 1963). In that context, this passive monetary policy period is instructive as a benchmark against which we can compare later Fed performances. The early Fed period starts with the founding of the Fed in 1913 and ends with the US entering the second world war. The post-war period starts in 1951 with the Fed regaining some independence after the Treasury-Fed accord (e.g. Romer and Romer, 2004*a*).[17] The post Volcker period covers the Great Moderation period and ends right before the pandemic.

We evaluate the Fed as a policy institution based on the loss function

$$\mathcal{L} = \frac{1}{2}\mathbb{E}\sum_{h=0}^{H}\beta^h(\pi_{t+h}^2 + \lambda u_{t+h}^2) \ , \tag{28}$$

where π_t denotes the inflation gap, u_t the unemployment rate gap, β the discount factor and λ the preference parameter. While the targets π^* and u^* are irrelevant to rank/assess reaction functions,[18] we posit that $\pi^* = 2$ and $u^* = 5$ in order to compute realized losses in the naive approach that we describe next.

Our baseline choice for the loss function sets $\beta = \lambda = 1$, and we take $H = 30$ quarters, a horizon large enough to ensure that the impulse responses have time to mean-revert. The data are quarterly, inflation is measured as year-on-year inflation based on the output deflator from Balke and Gordon (1986), and the unemployment rate before 1948 is taken from the NBER Macrohistory database over 1929-1948 and extended back to 1876 by interpolating the annual series from Weir (1992) and Vernon (1994).

[17]We exclude the period covering World War II until the Treasury-Fed accord of 1951, as the Fed was financing the war effort and had no independence.

[18]The ORA only depends on impulse responses, which are path deviations following an innovation, and as such do not depend on the constant terms in $\mathbf{Y}$.

5.1 Naive approach

To provide a benchmark for our results, we first take a naive approach where we evaluate the Fed based on realized outcomes for inflation and unemployment, as shown in Figure 1. Table 1 report realized losses for inflation and unemployment ($\mathcal{L}_x = \sum_{j=t_s}^{t_e} x_j^2$ for $x = \pi, u$) as well as the total realized loss ($\mathcal{L}_\pi + \mathcal{L}_u$).

The Early Fed period comes out as the worse period by far, with losses almost an order of magnitude larger than the other period. This is driven by the Great Depression; not only the large increase in unemployment but also the large movements in inflation, from the high inflation of the early 20s to the large deflation of the early 30s. In comparison, the passive Gold Standard period appears much more successful, suffering only from high inflation volatility. In fact, losses during the Gold Standard period are of similar magnitudes to the losses realized during the Post World War II, being on a par in terms of unemployment losses. The only period with clear superior outcomes is the Post Volcker Period, also referred to as the Great Moderation, with both stable inflation and unemployment and thus low losses throughout.

A naive interpretation of these macroeconomic outcomes could suggest that (i) monetary policy was superior during the Post Volcker period, and (ii) the founding of the Fed in 1913 caused worse outcomes than the passive Gold Standard. Unfortunately, we cannot make such causal claims, as many co-founding factors outside the Fed control could explain these macroeconomic outcomes. For instance, the poor inflation and unemployment realizations over 1913-1941 could have been caused by bad luck (an unfortunate sequence of shocks), adverse initial conditions or by a difficult economic environment. Similarly, the good performance of the economy in the Post Volcker period could be the outcome of good luck instead of good policy.

To assess policy performance we instead turn to the ORA methodology proposed in this paper.

5.2 Econometric implementation for ORA

To evaluate policy performance, we will assess how well the monetary authorities adjusted the contemporaneous policy rate in response to five separate non-policy shocks: financial shocks, government spending shocks, energy price shocks, inflation expectation shocks and TFP shocks.

This requires identifying six structural shocks: (i) shocks to the contemporaneous policy rule —the traditional monetary shock—, and (ii) the five non-policy shocks listed above, as we describe below.

To estimate the corresponding impulse responses, we rely on a Bayesian structural vector autoregressive model (SVAR) that includes a proxy for the policy shock, the non-policy shock, the outcome variables π_t and u_t, the growth rate of the monetary base, the policy rate, as well as possibly additional control variables w_t. During the 1879-1912 Gold Standard period where there is no policy institution, we take the 3-months treasury rate as the "policy rate" that a fictitious central bank could have controlled. For the 1913-1941 early Fed period, we use the fed discount rate as the policy rate. To capture the policy stance during the post WWII periods, we use the fed funds rate as the policy rate. The specific additional variables w_t and instruments z_t are discussed in detail below. The historical monetary data are taken from Balke and Gordon (1986).

The SVAR is specified for $y_t = (z_t^{\xi}, \pi_t, u_t, z_t^{\epsilon}, p_t, w_t')'$, where z_t^{ξ} is an instrument (or proxy) for the contemporaneous non-policy shock, z_t^{ϵ} is an instrument for the conventional contemporaneous monetary policy shock and w_t denotes additional control variables. We order the non-policy proxy first. As in Romer and Romer (2004*b*), we order the monetary proxy after unemployment and inflation (and before the federal funds rate), imposing the additional restriction that monetary policy does not affect inflation and unemployment within the period. We have

$$A_0 y_t = A_1 y_{t-1} + \ldots + A_p y_{t-p} + e_t \ , \tag{29}$$

where $A_0, \ldots, A_p$ are the coefficient matrices.

We estimate the reduced form of the SVAR model using standard Bayesian methods, which shrink the reduced form VAR coefficients using a Minnesota style prior. The prior variance hyper-parameters follow the recommendations in Canova (2007).

We normalize all shocks such that they have unit variance which can be implemented in practice by computing the conventional one standard deviation impulse responses. This scaling ensures comparability of the shocks across periods.

With the draws of the parameters from the posterior density we can compute the impulse responses to a policy shock ϵ_t (denoted by $\mathcal{R}_0^0$) as the ratio of the response of $y_{t+h} = (\pi_{t+h}, u_{t+h})$ over the response of p_t to the shock corresponding to z_t^{ϵ}. The impulse responses to non-policy shock ξ_t are denoted by Γ_0^0. We report the subset ORA statistic $\tau_0^* = -(\mathcal{R}_0^{0\prime}\mathcal{W}\mathcal{R}_0^0)^{-1}\mathcal{R}_0^{0\prime}\Gamma_0^0$ and the ORA adjusted impulse responses $\Gamma_0^* \equiv \Gamma_0^0 + \mathcal{R}_0^0\tau_0^*$.

5.3 Shock identification

For each period, we identify a monetary policy shock and five non-policy shocks: financial shocks, government spending shocks, energy price shocks, inflation expectation shocks and TFP shocks.

5.3.1 Monetary policy shocks

Since we want to compare policy makers based on their *contemporaneous* policy response to exogenous shocks, we need to identify contemporaneous shocks to the policy rate, that is shocks $\epsilon_{t,t}$. We consider two approaches for identifying such shocks. As our baseline we use the state of the art in the literature for each period, and as robustness we use a sign restriction identification.

Post Volcker regime For the Post Volcker period we use the high-frequency identification (HFI) approach, pioneered by Kuttner (2001) and Gürkaynak, Sack and Swanson (2005), and we use surprises in fed funds futures prices around FOMC announcement as proxies for monetary shocks. To isolate innovations to the contemporaneous policy rate, we use surprises to fed funds futures at a short horizon, here 3-months ahead fed funds futures (FF4), which (with quarterly data) ensures that the identified shock does not include news shocks to the future path of policy. While innovations to the contemporaneous policy rate could a priori include anticipated news shocks —forward-guidance was used extensively after 2007—, fed funds futures as of time t are based on the time t information set and thus already includes news shocks that were announced before time t. As a result, HFI surprises to FF4 fed isolate contemporaneous shocks to the policy rate (i.e., our object of interest $\epsilon_{t,t}$).

Post World War II regime For the Post World War II period we use the Romer and Romer (2004*b*) identified monetary policy shocks as instruments. Since there was no use of forward guidance before 1990 —Fed policymakers' views on the future policy path was "closely guarded" before 1990 (Rudebusch and Williams, 2008)—, we can consider that the Romer and Romer (2004*b*) monetary shocks capture solely contemporaneous policy shocks ($\epsilon_{t,t}$) and not news shocks to policy.[19]

Early Fed regime During the Early Fed period we use the Friedman and Schwartz (1963) dates extended by Romer and Romer (1989) as instruments to identify monetary policy shocks. We include five episodes —1920Q1, 1931Q3, 1933Q1, 1937Q1 and 1941Q3— where movements in money were "unusual given economic developments" (Romer and Romer, 1989). In the words of Romer and Romer (1989), these "unusual movements arose, in Friedman and Schwartz's view, from a conjunction of economic events, monetary institutions and the doctrines and beliefs of the time and of particular individuals determining policy". Since the concept of forward guidance in policy did not exist, we consider that the Friedman

[19]Technically speaking, the Romer and Romer (2004*b*) approach identifies the monetary policy shock $\epsilon_t = \sum_{j=0}^{t} \epsilon_{t,j}$. Without forward guidance, we have $\epsilon_{t,j} = 0$ for $j < t$ such that $\epsilon_t = \epsilon_{t,t}$.

and Schwartz's dates capture solely contemporaneous policy shocks ($\epsilon_{t,t}$) and not news shocks to policy.[20].

Pre Fed regime For the Pre Fed Gold Standard period, there is no clear baseline identification approach to identify monetary shocks, and we propose a new approach that exploits a unique feature of the Gold Standard. Under a Gold Standard, the monetary base depends on the amount of gold in circulation, which can itself vary for exogenous reasons related to the random nature of gold discoveries or development of new extraction techniques (e.g., Barsky and De Long, 1991). As such, we use unanticipated large gold mine discoveries (discoveries that led to gold rushes) as an instrument for movements in the monetary base. To the extent that the timing of the discovery is unrelated to the state of the business cycle, gold mine discovery will be a valid instrument. Mirroring Gold discovery, we will also use peak mine extraction —the moment when one of these large mines reached peak production—. The appendix provides more details on the construction of our instrument.

Alternative identification scheme One limitation of using the "state of the art" identification scheme in each period is that we rely on a different methodology to identify $\epsilon_{t,t}$ over each period. Since each methodology has different strengths and weaknesses, this could affect the results and the ORA comparison across periods.[21] To guard ourselves against this possibility, we will also use an identification of monetary shock that is consistent across regimes, which will ensure that the monetary shocks are identified in the exact same way across regimes. Specifically, we use sign restrictions, another popular method to identify monetary shocks (e.g., Uhlig, 2005). This approach has the benefit that the same identification scheme can be implemented over the entire sampling period. With the VAR including inflation, unemployment, the policy rate and the growth rate of the monetary base, we impose the following sign restrictions: a positive monetary shock raises the short-term rate in impact, lowers money growth on impact, and lowers inflation and raises unemployment after a year. Other than that, the responses are unconstrained.[22]

[20] The narrative accounts underlying these dates support this view, as all dates refer to changes in monetary variables within the quarter (Romer and Romer, 1989)

[21] For instance, exogeneity and relevance may differ across instrumental variables, see e.g., Barnichon and Mesters (2020) for a discussion of the different strengths and limits of the Romer and Romer (2004*b*) and the Gürkaynak, Sack and Swanson (2005) shock proxies.

[22] One potential drawback of the sign-restriction approach is that the identified monetary shocks may not isolate contemporaneous monetary shocks $\epsilon_{t,t}$. Since the VAR uses a limited set of observed macro variables to control for agents' information set, the VAR residuals —and thus our resulted identified monetary shocks— may mix contemporaneous shocks ($\epsilon_{t,t}$) with news shocks revealed before time t ($\epsilon_{t,t-j}, j > 0$) but not entirely captured by the VAR. While this is unlikely to be a problem before 1990 (see earlier discussions), it could be one in the post Volcker period where forward guidance was actively used. As robustness check, we thus expanded the VAR information set by adding SPF forecasts for the 3-month treasury bill rates to control for news shocks revealed before time t. Results were very similar.

5.3.2 Non-policy shocks

We now describe the identification of our five types of non-policy shocks. We again rely on standard identification methods in the literature, and we assume that these identification schemes identify the same shocks across periods.[23]

Financial shocks As financial shocks we use narratively identified bank panics. Each included panic was triggered by either a run on a particular trust fund or by foreign developments. The dates for the banking panics are taken from Reinhart and Rogoff (2009), Schularick and Taylor (2012) and Romer and Romer (2017). To capture the severity of the bank run, each non-zero entry is rescaled by the change in the BAA-AAA spread at the time of the run, similar to the re-scaling of Bernanke et al. (1997) and Romer and Romer (2017).[24]

Government spending shocks For government spending shocks we use the news shocks to defense spending as constructed in Ramey and Zubairy (2018).

Productivity shocks To identify productivity shocks we use the identification scheme of Gali (1999) and Barnichon (2010): we estimate bi-variate VARs with log output per hour and unemployment over each policy regime, and we impose long-run identifying restrictions, specifically that only productivity shocks can have permanent effects on productivity. The quarterly time series for output per hour is taken from Petrosky-Nadeau and Zhang (2021) and starts in 1890.

Energy shocks To identify energy shocks, we extend the approach of Hamilton (1996) and Hamilton (2003) by identifying energy shocks as instances when energy price rises above its 3-year maximum or falls below its 3-year minimum. Since coal was the primary US energy source until World War II and oil only became the pre-dominant energy source after World War II, we measure energy price prices from the wholesale price index for fuel and lighting, available over 1890-2019.

Inflation expectation shocks An important feature of a successful central bank is the anchoring of inflation expectations. In this context, we aim to measure how well the Fed

[23]The same assumption is implicit in earlier work exploiting the same five non-policy shocks, see Jordà, Schularick and Taylor (2013); Romer and Romer (2017), Ramey and Zubairy (2018), Blanchard and Galí (2007), Leduc, Sill and Stark (2007) and Galí and Gambetti (2009).

[24]Using bank runs as 0-1 dummies does not change conclusions drastically though it makes the estimates a bit less precise. Since the time series for AAA yields only start in 1919, we backcasted AAA yields before 1919 with yields on 10-year maturity government bonds from the Macro History database (Jordà et al., 2019).

has been responding to innovations to inflation expectations —a clear example being the de-anchoring of inflation expectations in the 1970s (Reis, 2021). To do so, we aim to identify inflation expectation shocks .

As measure of inflation expectations, we rely on the Livingston survey that has been continuously run over 1946-2019,[25] and includes a question about 8-months ahead inflation expectations. Prior to World War II, there are no systematic inflation expectation survey, so we instead rely on Cecchetti (1992)'s measure of 6-months ahead inflation expectations for the Early Fed period.[26]

To identify innovations to inflation expectations, we proceed similarly to Leduc, Sill and Stark (2007) and project inflation expectations on a set of controls that include past values of inflation expectation, inflation, unemployment, lags of the 3-month and 10-year treasury rates. In addition, we also project on current and past values of the other identified non-policy shocks: financial, government spending, energy price and TFP. The idea of this exercise is to capture movements in inflation expectations that cannot be explained by the other shocks, i.e., that go above and beyond the typical effect of the non-policy shocks on inflation expectations.

5.4 Results

Table 2 shows the baseline ORA statistics computed over the four periods for our five non-policy shocks. Recall that the ORA is an adjustment to the coefficient $\mathcal{B}_{p\xi}$ in the policy rule,[27] so that a negative ORA indicates that the policy rate was too high, either because the policy rate increased too much or because it did not decline enough following a non-policy shock.

In the main text, we focus on the main lessons of our exercise, leaving a more in-depth presentation of our results for the appendix. Our main results are as follows: (i) we estimate large and uniform improvements in the conduct of monetary policy, but *only* in the last 30 years, (ii) we cannot reject that the Fed's reaction to recent financial shocks (notably the 2007-2008 financial crisis) was appropriate, in contrast to the "highly" sub-optimal reaction of the Fed during the Great Depression, (iii) despite much larger realized losses in the 1920s-1930s, the performance of the early Fed is no worse than the performance of the passive Gold

[25] The Livingston survey is conducted with a pool of professional forecasters from non-financial businesses, investment banking firms, commercial banks, academic institutions, government, and insurance companies, see Leduc, Sill and Stark (2007).

[26] Cecchetti (1992)'s measure of inflation expectations relies on Mishkin (1981)'s insight that the ex-ante real interest rate can be recovered from a projection of the ex-post real interest rate on the time t information set. The difference between the ex-ante and ex-post real interest rate provides a measure of inflation expectations.

[27] For instance, an ORA of 0.5 means that in response to a 1 standard deviation non-policy shock, the reaction coefficient should have been 0.5 point larger.

Standard, and (iv) the Fed reaction function during the 1960s-1970s is almost as sub-optimal as the reaction function of the 1920s-1930s Fed, though the nature of the main non-policy shocks is different.

Improved policy in the Post Volcker period

Overall, we estimate strong improvements in the conduct of monetary policy, but *only* in the last 30 years, i.e., roughly after Volcker's dis-inflation program.

For the first 100 years of our study, we find *no* material improvement in the reaction function, with similar deviations from optimality over the first three periods. Comparing the rows of Table 2 for the three periods before Volcker, we can see ORAs of similar magnitudes with the average absolute ORA —a summary measure of performance per period (Table 2, right column)— hovering around 0.6 for 100 years.

It is only in the last 30 years that we estimate superior performances. In the post Volcker period, the ORAs are substantially smaller (and non-significant) than in the other periods, with an average absolute ORA of 0.2. In fact, the Post Volcker ORA statistics are smaller across *all* non-policy shocks, meaning that policy performance improves in all dimensions, from the responses to supply-type shocks like energy price shocks and TFP shocks to the responses to demand-type shocks like government spending shocks and financial shocks.[28] We will now focus in more details on the reaction to financial shocks, contrasting the Post Volcker Fed with the Early Fed of the 1920s-1930s.

Responding to financial shocks

In a 2002 speech in honor of Milton Friedman 90th birthday, (then) Fed governor Bernanke famously said: "Regarding the Great Depression. You're right, we did it. We're very sorry. But thanks to you, we won't do it again." (Bernanke, 2002). In an irony of history, the speech was made a full five years before the 2007-2008 financial crisis; a crisis that saw an unprecedented Fed response (see e.g., Bernanke, 2013) *with* Bernanke as Fed chairman.

Our results strikingly confirm Bernanke's quote, both his historical claim as well as his prophecy: the "poor" reaction function of the early Fed led to massive welfare losses, while the "good" reaction function of the Post Volcker Fed ensured little welfare losses coming from a sub-optimal reaction function.

To see this, we can first contrast the financial ORAs —the ORAs for financial shocks— estimated for the Early Fed period and for the Post Volcker period. With $\tau^* = -1.2$

[28]Importantly, the non-significance of the Post Volcker ORAs is *not* due to imprecisely estimated impulse responses. As we show in the Appendix, the Post Volcker impulse responses are estimated with reasonable precisions and the point estimates are sensible. The ORAs are small, *because* the impulse responses to non-policy shocks are (almost) orthogonal to the impulse responses to policy shocks.

(statistically significant), the Fed reaction to banking panics was too tight —a result echoing previous findings in the literature (e.g., Friedman and Schwartz, 1963; Hamilton, 1987)—. In contrast, the estimated ORA for the post Volcker Fed is four times smaller with $\tau^* = -0.3$ and not statistically significant, indicating that the post Volcker Fed period reacted much more appropriately and pointing to large improvements in the Fed's reaction to financial shocks.[29]

To better appreciate this improvement, Figures 2 and 3 display the impulse responses underlying the financial ORAs estimated for 1913-1941 and 1990-2019. The top rows show the impulse responses of inflation, unemployment and the interest rate to a monetary policy shock, while the bottom rows show the responses of the same variables to a financial shock.[30]

For the Early Fed period, notice how the Fed *raised* the discount rate in response to financial shocks. Combined with the decline in inflation caused by the financial shock, this means that the real policy rate increased substantially and monetary policy was contractionary, confirming earlier work on the monetary factors behind the Great Depression (e.g., Friedman and Schwartz, 1963; Hamilton, 1987). The ORA corrects this sub-optimal reaction function and turns the table on monetary policy by running an expansionary policy.[31] To see that, Figure 2 (dashed green line) reports the ORA adjusted impulse responses —$\Gamma_0^* = \Gamma_0^0 + \mathcal{R}_0^0 \tau_0^*$—, which depict how the ORA adjustment translates into different policy path responses to non-policy shocks and "improved" (i.e., more stable) impulse responses of inflation and unemployment. The ORA leads to a major adjustment to the policy path —the policy rate now goes down substantially on impact—, and the paths of inflation and unemployment are consequently much more stable. In contrast, for the Post Volcker period (Figure 3) the policy rate goes down following a financial shock (black line, lower-right panel), and the ORA only slightly adjusts the response of the policy rate (green line), leading to modest adjustments to the responses of inflation and unemployment.

As we saw in Section 4.5, we can also use the ORAs to quantify how much of the historical variation in inflation or unemployment was "unnecessary", being the outcome of a suboptimal reaction function. Figure 5 depicts the results of this exercise, showing the historical ORA adjustments to the policy rate ($\Delta \mathbf{P}_t$) and to inflation and unemployment

[29] That said, a point estimate at -0.3 indicates that the Fed should have lowered the fed funds rate more in response to financial shocks (according to the posterior mean). This could indicate that the presence of the zero lower bound may have limited somewhat the Fed's ability to best react to the 2007-2008 financial crisis.

[30] For both periods, a higher policy rate raises unemployment and lowers inflation, while a financial shock lowers inflation and raises unemployment. That said, the inflation response is more muted in the post-Volcker period, consistent with the anchoring of inflation expectations post Volcker or more generally with different economies across historical periods.

[31] Recall that the ORA is the outcome of a regression of the responses to a non-policy shock on the responses to a policy shock: a regression of the inflation and unemployment impulse responses in the bottom row on the corresponding impulse responses in the top row.

($\Delta \mathbf{Y}_t$) over each period. The corresponding gains in welfare (27) —the welfare losses that could have been avoided with a different reaction function— are listed in Table 3.

Figure 5 (second column) shows that the suboptimal reaction function of the early Fed translated into large welfare losses. For instance, during the Great Depression the ORA adjustment term amounts to as much as 10 ppt of unemployment and 15 ppt of inflation.[32] In units of welfare losses, this represents 35 points extra welfare loss due to sub-optimal policy (Table 3). In contrast, the welfare losses that can be attributed to the Post Volcker Fed are small. The Post Volcker ORA adjustment terms are small except in the early phase of the Great Recession, where the ORA calls for an additional 0.5 ppt drop in the fed funds rate in 2009, an adjustment that would have avoided about 0.5 ppt of unemployment (at some mild inflation cost). Overall, this represents only 0.7 point extra welfare loss due to sub-optimal policy (Table 3), an order of magnitude smaller than the welfare loss attributed to the early Fed.

The early Fed vs the passive Gold Standard

In contrast to the suggestive evidence of the naive approach (Table 1), the passive Gold Standard is *not* markedly superior to the early Fed. In other words, the founding of the Fed did not deteriorate performance relative to the passive monetary regime of the Gold Standard. Instead, performances were just as "bad" before and after the founding of the Fed.

Comparing the ORA before and after the founding of the Fed (Table 2, first two rows), we observe similar deviations from optimality. During the passive Gold Standard, monetary policy is (unsurprisingly) too passive in the face of adverse shocks: be it bank runs or military buildups.[33] Comparing ORAs across two the periods, we can see that (i) the excessive passivity simply continued after the founding of the Fed —the ORAs are similar across the two periods—, and (ii) the excessive passivity of the early Fed is not limited to financial distress, and it also extends to other shocks, here government spending shocks.[34]

[32] The ORA would have erased the discount rate hikes observed in 1931 —hikes often been blamed for turning the initial recession caused by the 1929 stock-market crash into a full blown depression (e.g., Hamilton, 1987)— and ultimately lowered the discount rate all the way to almost (but still above) zero in 1932. This would have avoided as much as 10 percentage points in unemployment —as much as half of the rise in unemployment over 1930-1932— as well as the deflation.

[33] To give a few noteworthy "misses" of the passive Gold Standard, the ORAs call for lower interest rates (about 3/4 ppt) in the aftermaths of the 1893 and 1907 bank runs, as well as higher interest rates in response to higher military spending following the war against Spain in 1898, and the navy build-up of 1902-1904. See Figure 5 and the Appendix for more details, notably the impulse responses underlying the ORAs estimated for 1879-1913.

[34] In particular, we find that the Fed's delayed reaction to the large increase in military spending in 1917 is responsible for some of the inflation outburst of 1919-1920 (see also Romer, 1992). See the Appendix for more details.

The Great Inflation

US monetary policy during the 1970s has generally been considered poor (e.g., Romer and Romer, 2004*a*), in particular not responding more than one-to-one with changes in inflation (Clarida, Galí and Gertler, 2000) and violating the so-called Taylor principle. However, beyond that Taylor principle, it has been difficult to quantify how "poor" monetary policy had been. The ORAs displayed in Table 2 can help address this limitation.

Overall, Fed performance during the 60s-70s is on a par with the poor performance of the early Fed, with ORAs of similar magnitudes, though the nature of the underlying shocks is different. Post World War II, the Fed reaction was too weak following all the different supply-type shocks that we identified: energy price shocks, TFP shocks as well as inflation expectation shocks. In fact, the reaction to inflation expectation shocks over the 60s-70s displays the largest deviation from optimality over the entire 150 year of monetary history with $\tau^* = 1.2$, even slightly larger (in absolute value) than the Fed's poor reaction to bank runs during the Great Depression.

To better appreciate these sub-optimal reactions, Figure 4 plots the impulse responses underlying the ORAs for inflation expectation shocks (similar results hold for energy or TFP shocks, see the appendix). In response to an inflation expectation shock, inflation rises progressively, but the policy rate does not respond, leading to negative real interest rates and further increasing inflation. The (large) ORA adjustment restores the Taylor principle: after the ORA, the policy rate rises strongly following an inflation expectation shock (lower-right panel, Figure 4) and stems the rise in inflation (at the cost of higher unemployment).

We can again use these counterfactuals to assess how much of the realized welfare losses were caused by these sub-optimal reactions, i.e., could have been avoided with a different reaction function. Figure 5 (third column) shows large ORA adjustment terms with substantial adjustments to the fed funds rate in response to the oil price shocks and inflation expectation shocks of the 1970s. For instance, about 5 ppt of inflation could have been avoided by 1980 (at the cost of extra unemployment). Overall, this represents about 3 points of welfare loss that can be attributed to the Fed, almost as large as the entire loss experienced during the Post Volcker period.

Robustness and caveats

In the appendix, we show robustness to our identification of monetary shocks, and we also consider robustness to the definition of the different monetary periods. Overall, our results are consistent with our baseline estimates, with ORAs of similar magnitudes and levels of statistical significance.

As final comments, we note two important caveats. First, our analysis takes as starting point a loss function with equal weights on inflation and unemployment and studies whether different reaction functions could have achieved lower losses, thereby attributing some of the variation in inflation and unemployment to sub-optimal policies. Different loss functions *might* justify past reaction functions, and it is not our objective to argue in favor of one loss function vs another.

Second, we do not take a stand on the reasons for past sub-optimal reaction functions. A better understanding of the functioning of the economy (Friedman and Schwartz, 1963), better and more timely data (Romer, 1986; Orphanides, 2001), better forecasting (Dominguez, Fair and Shapiro, 1988) and better causal inference methods (Romer and Romer, 1989) could all be part of the improvements in policy over the last 30 years. Parsing out these different reasons is an important question for future research.

6 Conclusion

In this paper, we propose to evaluate makers based on how well they reacted to the exogenous shocks that they faced during their term. We show how such a reaction function evaluation is possible with minimal assumptions on the underlying structural economic model. We introduce a new statistic, the ORA, which measures the distance to the optimal reaction function and can be computed from two sets of sufficient statistics: (i) the impulse responses of the policy objectives to non-policy shocks, and (ii) the same impulse responses to policy shocks. Importantly, explicit knowledge of the policy maker's reaction function is not necessary, because the effect of an (unspecified) reaction function is already encoded in the impulse responses to shocks, which are estimable.

We apply this methodology to evaluate US monetary policy over the past 150 years; from the Gold standard period to the post-Volcker regime. We find no material improvement in the reaction function over the first 100 years, and it is only in the last 30 years that we estimate large and uniform improvements in the conduct of monetary policy.

Going forward, the methodology could be applied to many other important evaluation questions; not only in the context of monetary policy (e.g., comparing central banks such as the Fed vs the ECB during the Great Recession), but also in the context of fiscal policy (e.g., comparing the performance of US presidents, Blinder and Watson, 2016), health policy (e.g., comparing governments' policy responses to COVID), or climate change mitigation policy. We leave these questions for future research.

References

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

Balke, N., and R. J. Gordon. 1986. "Appendix B Historical Data." *The American Business Cycle: Continuity and Change*. National Bureau of Economic Research.

Barnichon, Regis. 2010. "Productivity and unemployment over the business cycle." *Journal of Monetary Economics*, 57(8): 1013–1025.

Barnichon, Regis, and Geert Mesters. 2020. "Identifying Modern Macro Equations using Old Shocks." *Quarterly Journal of Economics*, 135: 2255–2298.

Barnichon, Regis, and Geert Mesters. 2023. "A Sufficient Statistics Approach for Macroeconomic Policy." *American Economic Review*. forthcoming.

Barsky, Robert B, and J Bradford De Long. 1991. "Forecasting pre-World War I inflation: the Fisher effect and the gold standard." *The Quarterly Journal of Economics*, 106(3): 815–836.

Bernanke, Ben S. 1986. "Alternative explanations of the money-income correlation." *Carnegie-Rochester Conference Series on Public Policy*, 25: 49 – 99.

Bernanke, Ben S. 2002. "On Milton Friedman's ninetieth birthday."

Bernanke, Ben S. 2013. *The Federal Reserve and the financial crisis.* Princeton University Press.

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

Blanchard, Olivier, and Jordi Galí. 2007. "The Macroeconomic Effects of Oil Price Shocks: Why Are the 2000s so Different from the 1970s?" In *International Dimensions of Monetary Policy.* 373–421. National Bureau of Economic Research, Inc.

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

Boivin, Jean. 2005. "Has US monetary policy changed? Evidence from drifting coefficients and real-time data."

Bordo, Michael D, and Finn E Kydland. 1995. "The gold standard as a rule: An essay in exploration." *Explorations in Economic History*, 32(4): 423–464.

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

Cecchetti, Stephen G. 1992. "Prices During the Great Depression: Was the Deflation of 1930-1932 Really Unanticipated?" *The American Economic Review*, 141–156.

Clarida, Richard, Jordi Galí, and Mark Gertler. 2000. "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory." *The Quarterly Journal of Economics*, 115(1): 147–180.

Cogley, Timothy, and Thomas J. Sargent. 2005. "Drifts and volatilities: monetary policies and outcomes in the post WWII US." *Review of Economic Dynamics*, 8(2): 262 – 302. Monetary Policy and Learning.

Coibion, Olivier, and Yuriy Gorodnichenko. 2011. "Monetary policy, trend inflation, and the great moderation: An alternative interpretation." *American Economic Review*, 101(1): 341–370.

Dominguez, Kathryn M, Ray C Fair, and Matthew D Shapiro. 1988. "Forecasting the depression: Harvard versus Yale." *The American Economic Review*, 595–612.

Fair, Ray C. 1978. "The Use of Optimal Control Techniques to Measure Economic Performance." *International Economic Review*, 19(2): 289–309.

Friedman, Milton, and Anna Jacobson Schwartz. 1963. *A monetary history of the United States, 1867-1960.* Princeton University Press.

Gali, Jordi. 1999. "Technology, employment, and the business cycle: do technology shocks explain aggregate fluctuations?" *American economic review*, 89(1): 249–271.

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

Galí, Jordi, and Luca Gambetti. 2009. "On the sources of the great moderation." *American Economic Journal: Macroeconomics*, 1(1): 26–57.

Gali, Jordi, and Mark Gertler. 2007. "Macroeconomic Modeling for Monetary Policy Evaluation." *Journal of Economic Perspectives*, 21(4): 25–46.

Galı, Jordi, J David López-Salido, and Javier Vallés. 2003. "Technology shocks and monetary policy: assessing the Fed's performance." *Journal of Monetary Economics*, 50(4): 723–743.

Gürkaynak, Refet, S., Brian Sack, and Eric Swanson. 2005. "The Sensitivity of Long-Term Interest Rates to Economic News: Evidence and Implications for Macroeconomic Models." *American Economic Review*, 95: 425–436.

Hamilton, James D. 1987. "Monetary factors in the great depression." *Journal of Monetary Economics*, 19(2): 145–169.

Hamilton, James D. 1996. "This is what happened to the oil price-macroeconomy relationship." *Journal of Monetary Economics*, 38(2): 215–220.

Hamilton, James D. 2003. "What is an oil shock?" *Journal of Econometrics*, 113(2): 363–398.

Ilzetzki, Ethan, Enrique G Mendoza, and Carlos A Végh. 2013. "How big (small?) are fiscal multipliers?" *Journal of monetary economics*, 60(2): 239–254.

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

Jordà, Òscar, Katharina Knoll, Dmitry Kuvshinov, Moritz Schularick, and Alan M Taylor. 2019. "The rate of return on everything, 1870–2015." *The Quarterly Journal of Economics*, 134(3): 1225–1298.

Jordà, Òscar, Moritz Schularick, and Alan M Taylor. 2013. "When credit bites back." *Journal of Money, Credit and Banking*, 45(s2): 3–28.

Judd, John P, and Glenn D Rudebusch. 1998. "Taylor's Rule and the Fed: 1970-1997." *Economic Review-Federal Reserve Bank of San Francisco*, 3–16.

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

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.

Leduc, Sylvain, Keith Sill, and Tom Stark. 2007. "Self-fulfilling expectations and the inflation of the 1970s: Evidence from the Livingston Survey." *Journal of Monetary economics*, 54(2): 433–459.

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

Li, Dake, Mikkel Plagborg-Møller, and Christian K. Wolf. 2022. "Local Projections vs. VARs: Lessons From Thousands of DGPs." Working paper.

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

Meltzer, Allan H. 2003. *A history of the Federal Reserve, Volume 1: 1913-1951.* University of Chicago Press.

Meltzer, Allan H. 2009*a*. "A History of the Federal Reserve, Volume 2, Books 1, 1951-1969."

Meltzer, Allan H. 2009*b*. "A History of the Federal Reserve, Volume 2, Books 2, 1970-1986."

Mishkin, Frederic S. 1981. "The real interest rate: An empirical investigation." Vol. 15, 151–200, Elsevier.

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

Orphanides, Athanasios. 2001. "Monetary policy rules based on real-time data." *American Economic Review*, 91(4): 964–985.

Orphanides, Athanasios. 2003. "Historical monetary policy analysis and the Taylor rule." *Journal of monetary economics*, 50(5): 983–1022.

Petrosky-Nadeau, Nicolas, and Lu Zhang. 2021. "Unemployment crises." *Journal of Monetary Economics*, 117: 335–353.

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

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

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

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

Reinhart, Carmen M, and Kenneth S Rogoff. 2009. "This time is different." In *This Time Is Different.* princeton university press.

Reis, Ricardo. 2021. "Losing the Inflation Anchor." *Brookings papers on economic activity*, 2021(2): 307–379.

Romer, Christina D. 1986. "Is the Stabilization of the Postwar Economy a Figment of the Data?" *The American Economic Review*, 76(3): 314–334.

Romer, Christina D. 1992. "What ended the great depression?" *The Journal of Economic History*, 52(4): 757–784.

Romer, Christina D, and David H Romer. 1989. "Does monetary policy matter? A new test in the spirit of Friedman and Schwartz." *NBER macroeconomics annual*, 4: 121–170.

Romer, Christina D, and David H Romer. 2004*a*. "Choosing the Federal Reserve chair: lessons from history." *Journal of Economic Perspectives*, 18(1): 129–162.

Romer, Christina D., and David H. Romer. 2004*b*. "A New Measure of Monetary Shocks: Derivation and Implications." *American Economic Review*, 94: 1055–1084.

Romer, Christina D., and David H. Romer. 2017. "New Evidence on the Aftermath of Financial Crises in Advanced Countries." *American Economic Review*, 107(10): 3072–3118.

Rudebusch, Glenn D, and John C Williams. 2008. "Revealing the secrets of the temple: The value of publishing central bank interest rate projections." In *Asset Prices and Monetary Policy*. 247–289. University of Chicago Press.

Schularick, Moritz, and Alan M Taylor. 2012. "Credit booms gone bust: Monetary policy, leverage cycles, and financial crises, 1870-2008." *American Economic Review*, 102(2): 1029–61.

Sims, Christopher A, and Tao Zha. 2006. "Does monetary policy generate recessions?" *Macroeconomic Dynamics*, 10(2): 231–272.

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.

Svensson, Lars EO. 2003. "What is wrong with Taylor rules? Using judgment in monetary policy through targeting rules." *Journal of Economic Literature*, 41(2): 426–477.

Taylor, John B. 1999. "A historical analysis of monetary policy rules." In *Monetary policy rules*. 319–348. University of Chicago Press.

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.

Vernon, J.R. 1994. "Unemployment rates in postbellum America: 1869–1899." *Journal of Macroeconomics*, 16(4): 701–714.

Weir, David R. 1992. "A century of US unemployment, 1890-1990: revised estimates and evidence for stabilization." *Research in Economic History*, 14(1): 301–46.

Wheelock, David C, et al. 2010. "Lessons learned? Comparing the Federal Reserve's responses to the crises of 1929-1933 and 2007-2009." *Federal Reserve Bank of St. Louis Review*, 92(Mar): 89–108.

Appendix A: Details and Proofs

Proof of Lemma 1. Define

$$\mathcal{A} = \begin{bmatrix} \mathcal{A}_{yy} & \mathcal{A}_{yp} \\ \mathcal{A}_{py} & \mathcal{A}_{pp} \end{bmatrix} \ , \ \mathcal{B}_{\xi} = \begin{bmatrix} \mathcal{B}_{y\xi} \\ \mathcal{B}_{p\xi} \end{bmatrix} \ , \ \mathbf{J} = \begin{bmatrix} \mathbf{0} \\ \mathbf{I} \end{bmatrix} \text{ and } \mathbf{Z} = \begin{bmatrix} \mathbf{Y} \\ \mathbf{P} \end{bmatrix} \ . \tag{30}$$

The model (11) is equivalent to

$$\mathcal{A}\mathbf{Z} = \mathcal{B}_{\xi}\boldsymbol{\Xi} + \mathbf{J}\boldsymbol{\epsilon} \ .$$

For any $\phi \in \Phi$ we have that there exists unique equilibrium representation. This implies that $\mathcal{A}$ is invertible and we obtain

$$\mathbf{Z} = \underbrace{\mathcal{A}^{-1}\mathcal{B}_{\xi}}_{=\mathcal{D}_1}\boldsymbol{\Xi} + \underbrace{\mathcal{A}^{-1}\mathbf{J}}_{=\mathcal{D}_2}\boldsymbol{\epsilon} \ .$$

The block structure of $\mathcal{D}_1$ and $\mathcal{D}_2$ is given by

$$\mathcal{D}_1 = \begin{bmatrix} \Gamma(\phi) \\ \Gamma_p(\phi) \end{bmatrix} \quad \text{and} \quad \mathcal{D}_1 = \begin{bmatrix} \mathcal{R}(\phi) \\ \mathcal{R}_p(\phi) \end{bmatrix} \ ,$$

where the maps $\Gamma(\phi)$ and $\mathcal{R}(\phi)$ appear in the first position as they capture the effects of the shocks on $\mathbf{Y}$. The other maps capture the effects of the shocks on $\mathbf{P}$. Explicit expression can be obtained by by noting that $\mathcal{A}$ being invertible implies that $\mathcal{A}_{pp}$ and $\mathcal{A}_{yy} - \mathcal{A}_{yp}\mathcal{A}_{pp}^{-1}\mathcal{A}_{py}$ are invertible as $\mathcal{A}_{yy}$ is generally not invertible. We have

$$\Gamma(\phi) = \mathcal{D}(\mathcal{B}_{y\xi} + \mathcal{A}_{yp}\mathcal{A}_{pp}^{-1}\mathcal{B}_{p\xi}) \qquad \text{and} \qquad \mathcal{R}(\phi) = \mathcal{D}\mathcal{A}_{yp}\mathcal{A}_{pp}^{-1} \ ,$$

with $\mathcal{D} = (\mathcal{A}_{yy} - \mathcal{A}_{yp}\mathcal{A}_{pp}^{-1}\mathcal{A}_{yp})^{-1}$. □

Proof of Lemma 2. Given some $\phi \in \Phi$ we can follow the same steps as the proof of Lemma 1 but using an augmented policy rule

$$\mathcal{A}_{pp}\mathbf{P} - \mathcal{A}_{py}\mathbf{Y} = (\mathcal{B}_{p\xi} + \mathcal{T})\boldsymbol{\Xi} + \boldsymbol{\epsilon} \ ,$$

and we obtain the equilibrium representation

$$\mathbf{Y} = (\Gamma(\phi) + \mathcal{R}(\phi)\mathcal{T})\boldsymbol{\Xi} + \mathcal{R}(\phi)\boldsymbol{\epsilon} \ , \tag{31}$$

where

$$\Gamma(\phi) = \mathcal{D}(\mathcal{B}_{y\xi} + \mathcal{A}_{yp}\mathcal{A}_{pp}^{-1}\mathcal{B}_{p\xi}) \qquad \text{and} \qquad \mathcal{R}(\phi) = \mathcal{D}\mathcal{A}_{yp}\mathcal{A}_{pp}^{-1} \ ,$$

with $\mathcal{D} = (\mathcal{A}_{yy} - \mathcal{A}_{yp}\mathcal{A}_{pp}^{-1}\mathcal{A}_{yp})^{-1}$. We obtain the first part of Lemma 2 for $\phi = \phi_0$.

Further, recalling that $\theta = \{\mathcal{A}_{pp}, \mathcal{A}_{py}, \mathcal{B}_{p\xi}\}$ we have

$$\Gamma(\phi) = \Gamma(\{\mathcal{A}_{pp}, \mathcal{A}_{py}, \mathbf{0}\}) + \mathcal{R}(\phi)\mathcal{B}_{p\xi} \ , \tag{32}$$

from which the second part of Lemma 2 follows directly by adjusting the rule coefficients $\mathcal{B}_{p\xi}$ to $\mathcal{B}_{p\xi} + \mathcal{T}$. □

Proof of Proposition 1. The proof proceeds in two steps: (a) we show the equivalence for

$\{\min_\phi \mathcal{L}$ s.t. (11) with $\boldsymbol{\epsilon}=\mathbf{0}\} = \{\min_{\mathcal{B}_{p\xi}} \mathcal{L}$ s.t. (11) with $\boldsymbol{\epsilon}=\mathbf{0}$, $\mathcal{A}_{pp}=\mathcal{A}^0_{pp},\ \mathcal{A}_{py}=\mathcal{A}^0_{py}\}$ and (b) we show that the value for $\mathcal{B}_{p\xi}$ that solves the second problem is $\mathcal{B}^0_{p\xi}+\mathcal{T}^*$. To show (a) we note that under $\boldsymbol{\epsilon}=\mathbf{0}$ we have that $\mathbf{Y}$ can be written as

$$\mathbf{Y} = \mathcal{D}\mathcal{B}_{y\xi}\boldsymbol{\Xi} + \mathcal{D}\mathcal{A}_{yp}\mathcal{A}_{pp}^{-1}\mathcal{B}_{p\xi}\boldsymbol{\Xi} = \Gamma(\phi)\boldsymbol{\Xi}$$

Using that the entries of $\boldsymbol{\Xi}$ have mean zero, unit variance and are uncorrelated we have that

$$\mathcal{L} = \frac{1}{2}\mathbb{E}(\mathbf{Y}'\mathcal{W}\mathbf{Y}) = \mathrm{Tr}((\mathcal{B}_{y\xi}+\mathcal{A}_{yp}\mathcal{A}_{pp}^{-1}\mathcal{B}_{p\xi})'\mathcal{D}'\mathcal{W}\mathcal{D}(\mathcal{B}_{y\xi}+\mathcal{A}_{yp}\mathcal{A}_{pp}^{-1}\mathcal{B}_{p\xi}))\ .$$

The derivative maps of $\mathcal{L}$ with respect to $\phi=\{\mathcal{A}_{pp},\mathcal{A}_{py},\mathcal{B}_{p\xi}\}$ are given by

$$\begin{aligned}\mathcal{A}_{pp}^{-1'}\mathcal{A}_{yp}'\mathcal{D}'\mathcal{W}\mathcal{D}(\mathcal{B}_{y\xi}+\mathcal{A}_{yp}\mathcal{A}_{pp}^{-1}\mathcal{B}_{p\xi})\mathcal{B}_{p\xi}'\mathcal{A}_{pp}^{-1'}+&\\ \mathcal{A}_{pp}^{-1'}\mathcal{A}_{yp}'\mathcal{D}'\mathcal{W}\mathcal{D}(\mathcal{B}_{y\xi}+\mathcal{A}_{yp}\mathcal{A}_{pp}^{-1}\mathcal{B}_{p\xi})(\mathcal{B}_{y\xi}+\mathcal{A}_{yp}\mathcal{A}_{pp}^{-1}\mathcal{B}_{p\xi})'\mathcal{D}'\mathcal{A}_{py}'\mathcal{A}_{pp}^{-1'} &= \mathbf{0}\\ \mathcal{A}_{pp}^{-1'}\mathcal{A}_{yp}'\mathcal{D}'\mathcal{W}\mathcal{D}(\mathcal{B}_{y\xi}+\mathcal{A}_{yp}\mathcal{A}_{pp}^{-1}\mathcal{B}_{p\xi})(\mathcal{B}_{y\xi}+\mathcal{A}_{yp}\mathcal{A}_{pp}^{-1}\mathcal{B}_{p\xi})'\mathcal{D}' &= \mathbf{0}\\ \mathcal{A}_{pp}^{-1'}\mathcal{A}_{yp}'\mathcal{D}'\mathcal{W}\mathcal{D}(\mathcal{B}_{y\xi}+\mathcal{A}_{yp}\mathcal{A}_{pp}^{-1}\mathcal{B}_{p\xi}) &= \mathbf{0}\end{aligned}$$

The last equation gives the derivative map with respect to $\mathcal{B}_{p\xi}$. Solving this expression for $\mathcal{B}_{p\xi}$ yields

$$\mathcal{B}^*_{p\xi} = -[\mathcal{A}_{pp}^{-1'}\mathcal{A}_{yp}'\mathcal{D}'\mathcal{W}\mathcal{D}\mathcal{A}_{yp}\mathcal{A}_{pp}^{-1}]^{-1}\mathcal{A}_{pp}^{-1'}\mathcal{A}_{yp}'\mathcal{D}'\mathcal{W}\mathcal{D}\mathcal{B}_{y\xi}\ .$$

Further, it is easy to see that if the last equation holds then the first two equations also hold. This holds regardless of $\mathcal{A}_{pp}$ and $\mathcal{A}_{py}$ as long as the invertibility conditions above are satisfied.

To show part (b), note that $\mathcal{R}^0 = \mathcal{D}^0\mathcal{A}^0_{yp}(\mathcal{A}^0_{pp})^{-1}$ and if $\mathcal{B}^0_{p\xi}=\mathbf{0}$ we have that $\Gamma^0 = \Gamma(\{\mathcal{A}^0_{pp},\mathcal{A}^0_{py},\mathbf{0}\}) = \mathcal{D}^0\mathcal{B}_{y\xi}$. This implies that $\mathcal{B}^*_{p\xi} = \mathcal{T}^* = -(\mathcal{R}^{0'}\mathcal{W}\mathcal{R}^0)^{-1}\mathcal{R}^{0'}\mathcal{W}\Gamma^0$ and the proof is complete. Now suppose that $\mathcal{B}^0_{p\xi}\neq\mathbf{0}$, using (32) we have $\mathcal{B}^0_{p\xi}+\mathcal{T}^* = \mathcal{B}^0_{p\xi} - (\mathcal{R}^{0'}\mathcal{W}\mathcal{R}^0)^{-1}\mathcal{R}^{0'}\mathcal{W}\Gamma^0 = \mathcal{B}^0_{p\xi} - (\mathcal{R}^{0'}\mathcal{W}\mathcal{R}^0)^{-1}\mathcal{R}^{0'}\mathcal{W}\Gamma(\{\mathcal{A}^0_{pp},\mathcal{A}^0_{py},\mathbf{0}\}) - (\mathcal{R}^{0'}\mathcal{W}\mathcal{R}^0)^{-1}\mathcal{R}^{0'}\mathcal{W}\mathcal{R}^0\mathcal{B}^0_{p\xi} = \mathcal{B}^*_{p\xi}$. □

Proof of Proposition 2. From

$$\mathbb{E}\mathcal{L} = (\Gamma^0_b+\mathcal{R}^0_a\mathcal{T}_{ab})'(\Gamma^0_b+\mathcal{R}^0_a\mathcal{T}_{ab}) + \text{terms independent of } \mathcal{T}_{ab}\ ,$$

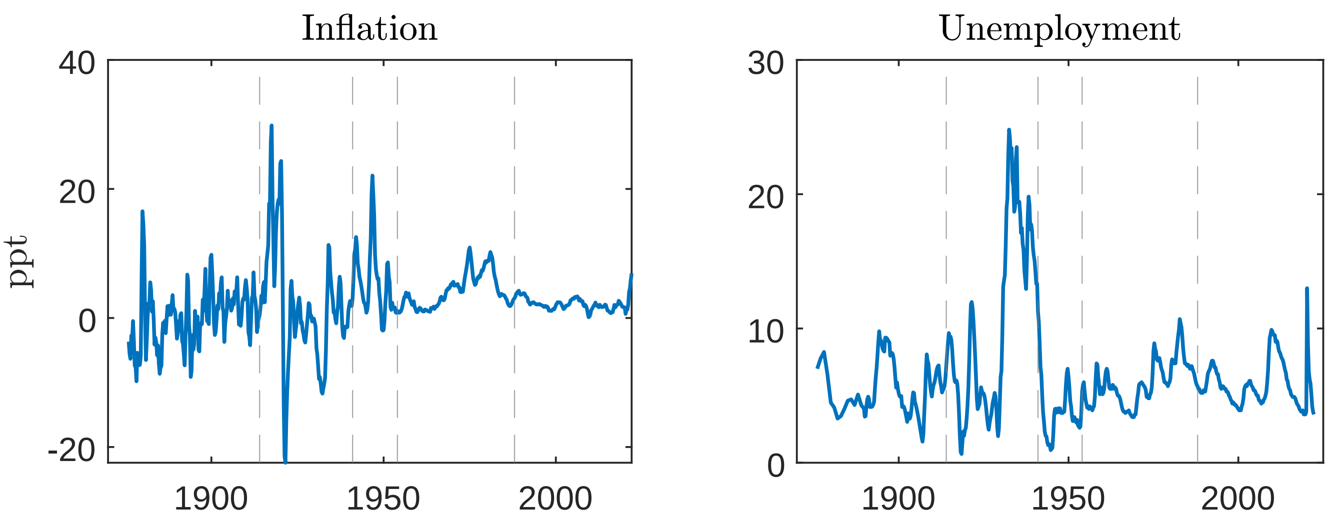
and the definition of $\mathcal{T}^*_{ab}$, we have that $\frac{\partial\mathbb{E}\mathcal{L}}{\partial\mathcal{T}_{ab}}\Big|_{\mathcal{T}^*_{ab}} = 0$, which establishes the first part since the optimization problem is convex.

We have that

$$\begin{aligned}\mathcal{L}(\phi^0) &= \mathcal{L}(\mathcal{A}^0_{pp},\mathcal{A}^0_{py},\mathcal{B}^0_{px},\mathcal{B}^0_{p_a\xi_b}+\mathcal{T}_{ab},\mathcal{B}^0_{-p_a-\xi_b})\big|_{\mathcal{T}_{ab}=0}\\ &\geq \min_{\mathcal{T}_{ab}}\mathcal{L}(\mathcal{A}^0_{pp},\mathcal{A}^0_{py},\mathcal{B}^0_{px},\mathcal{B}^0_{p_a\xi_b}+\mathcal{T}_{ab},\mathcal{B}^0_{-p_a-\xi_b})\\ &= \mathbb{E}\mathcal{L}_t(\phi^*)\ ,\end{aligned}$$

which establishes the second part. □

Figure 1: Inflation and unemployment, 1879–2019



Notes: Year-on-year inflation (GDP deflator) and the unemployment rate. The vertical lines highlight the different periods: Pre Fed 1879-1912, Early Fed 1913-1941, Post WWII 1951-1984 and Post Volcker 1990-2019.

Table 1: Realized losses

	Pre Fed 1879-1912	Early Fed 1913-1941	Post WWII 1951-1984	Post Volcker 1990-2019
$\mathcal{L}_\pi$	24.3	83.0	11.9	0.7
$\mathcal{L}_u$	3.7	70.0	3.6	3.0
$\mathcal{L}$	28.1	153.1	15.5	3.8

Notes: Realized losses for inflation ($\mathcal{L}_\pi$), unemployment $\mathcal{L}_u$ and total ($\mathcal{L}_\pi + \mathcal{L}_u$) for the different periods.

Table 2: ORA statistics for US monetary policy

Non-policy shock Shock sign convention	Bank panics $u \uparrow$	G $u \uparrow$	Energy $\pi \uparrow$	π^e $\pi \uparrow$	TFP $\pi \uparrow$	Average |ORA|
Pre Fed 1879−1912	**−0.9*** (−1.5,−0.3)	**−0.6*** (−1.3,0)	**−0.1** (−0.5,0.4)	—	**0.6** (−0.2,1.1)	**0.6**
Early Fed 1913−1941	**−1.2*** (−1.9,−0.8)	**−0.5*** (−0.9,−0.1)	**0.0** (−0.3,0.3)	**0.7*** (0.3,1.0)	**0.1** (−0.2,0.5)	**0.5**
Post WWII 1951−1984	—	**−0.2** (−0.8,0.3)	**0.8*** (0.1,1.4)	**1.2*** (0.6,1.8)	**0.5** (−0.2,1.2)	**0.7**
Post Volcker 1990−2019	**−0.3** (−0.8,0.2)	**0.1** (−0.4,0.6)	**−0.2** (−0.8,0.7)	**−0.1** (−0.4,0.3)	**−0.3** (−0.7,0.1)	**0.2**

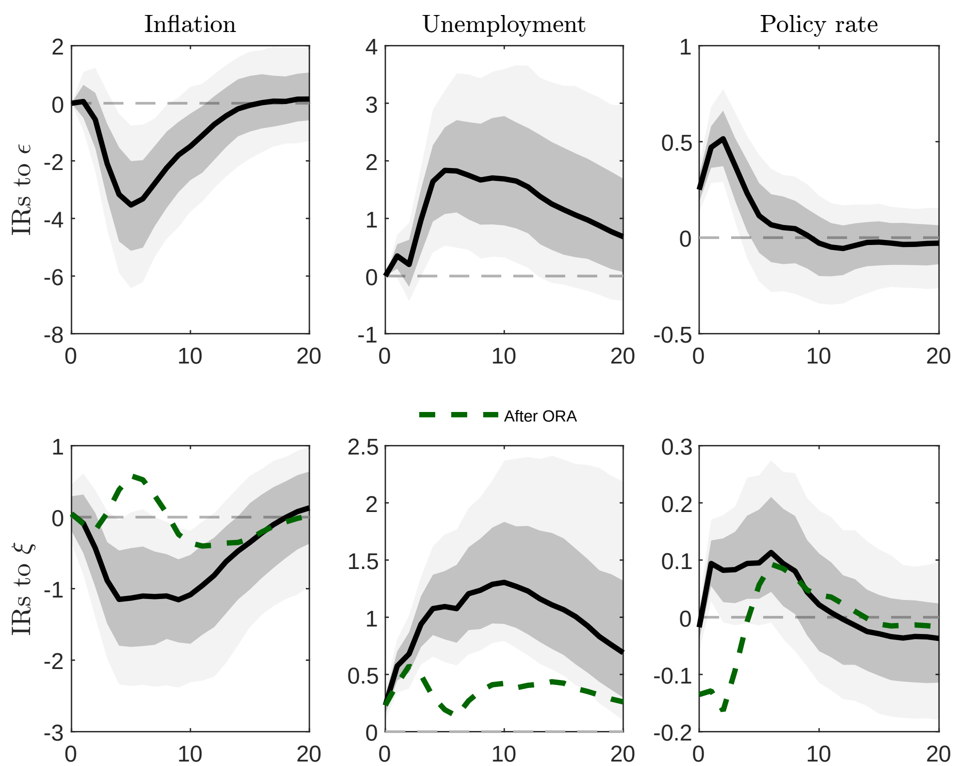
Notes: Median ORA statistics together with 68% credible sets. The monetary policy shocks are identified as described in the main text: using gold rush discoveries in the pre-Fed period, Romer and Romer (1989)'s Friedman-Schwartz dates in the early Fed period, Romer and Romer (2004) monetary shocks for the post WWII period and high-frequency surprises in the post Volcker period. The financial shocks are bank panics from Reinhart and Rogoff (2009), the government spending shocks (G) are from Ramey and Zubairy (2018), TFP shocks from Gali (1999), energy shocks are computed using the peak-over-threshold approach of Hamilton (1996), and inflation expectation shocks (π^e) are innovations to inflation expectations as measured from Cecchetti (1992) for Early Fed period and from the Livingston survey after 1946. For the Pre Fed period the TFP, G and Energy ORAs are computed over the 1890-1912 period. The right column ("Average |ORA|") reports the average absolute ORAs estimated for each period.

Table 3: The effects of ORA adjustments on the loss function

	Pre Fed 1879-1912	Early Fed 1913-1941	Post WWII 1951-1984	Post Volcker 1990-2019
$\Delta\mathcal{L}_\pi$	-2.3	-16.3	-3.3	0.0
$\Delta\mathcal{L}_u$	-1.2	-18.8	0.7	-0.6
$\Delta\mathcal{L}$	-3.4	-35.2	-2.7	-0.6

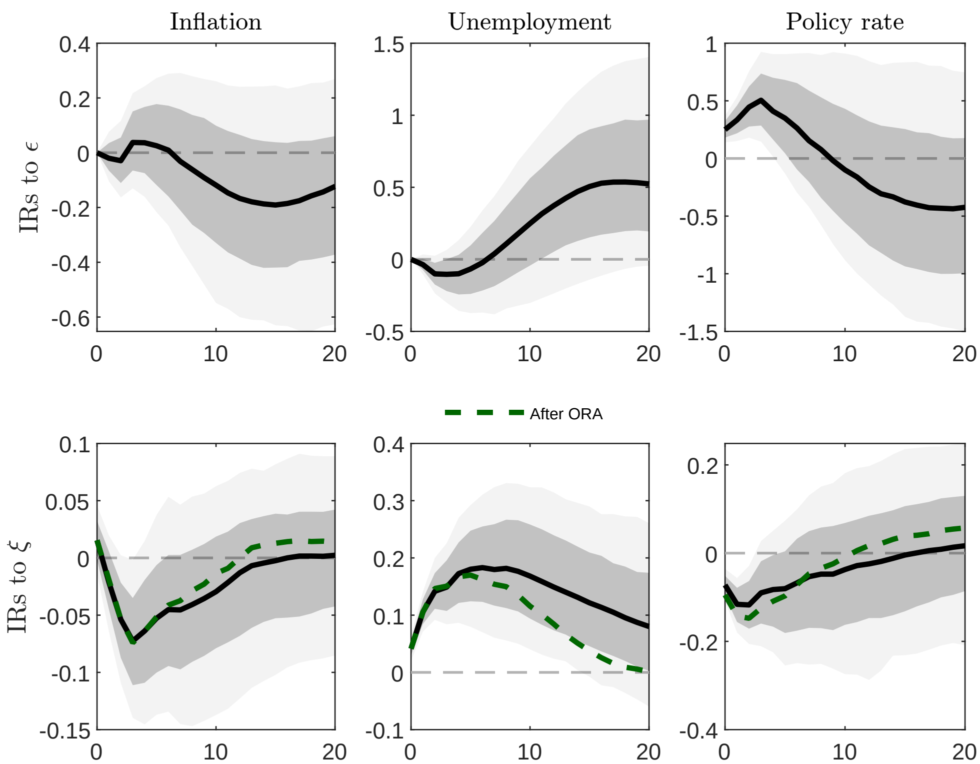
Notes: Effects of ORA adjsutments on realized losses for inflation ($\Delta\mathcal{L}_\pi$), unemployment $\Delta\mathcal{L}_u$ and total ($\Delta\mathcal{L} = \Delta\mathcal{L}_\pi + \Delta\mathcal{L}_u$) for the different periods, as computed from (27).

Figure 2: EARLY FED, 1913-1941, REACTION TO FINANCIAL SHOCKS



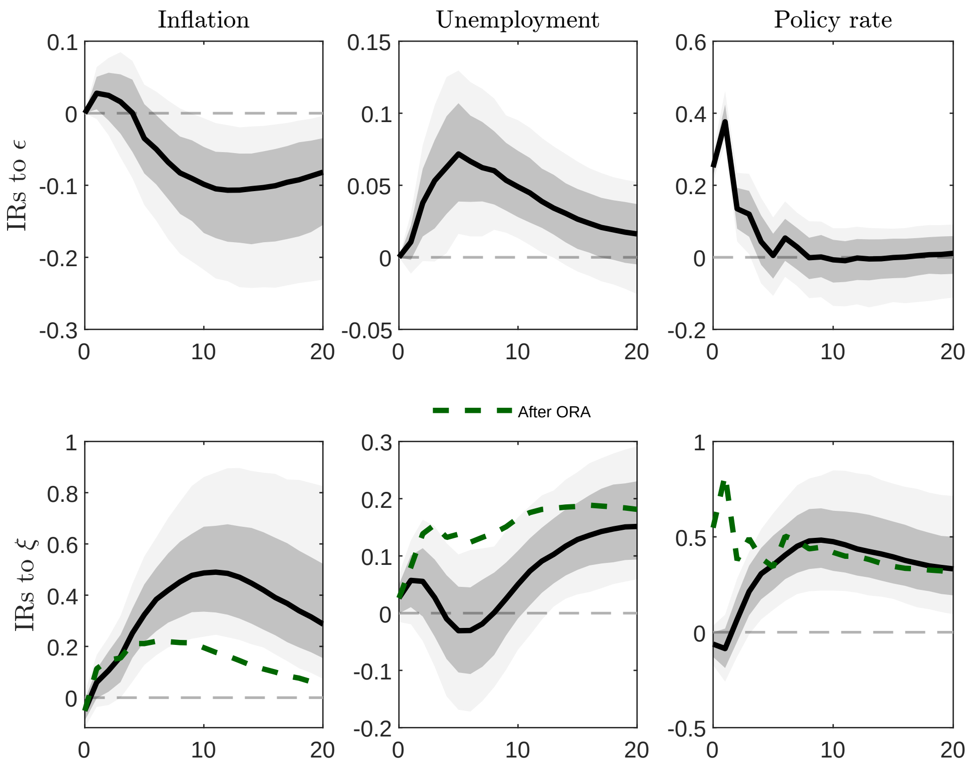
Notes: The top (resp. bottom) row shows the median responses (thick line) of inflation, unemployment and the Fed's discount rate to a monetary policy shock ϵ (resp. financial shock ξ). The dotted green lines show the ORA adjusted impulse responses: $\Gamma_0^0 + \mathcal{R}_0^0 \tau_0^*$. The 95% and 67% credible sets are plotted as dark and light shaded areas, respectively.

Figure 3: Post Volcker Fed, 1990-2019, Reaction to Financial shocks



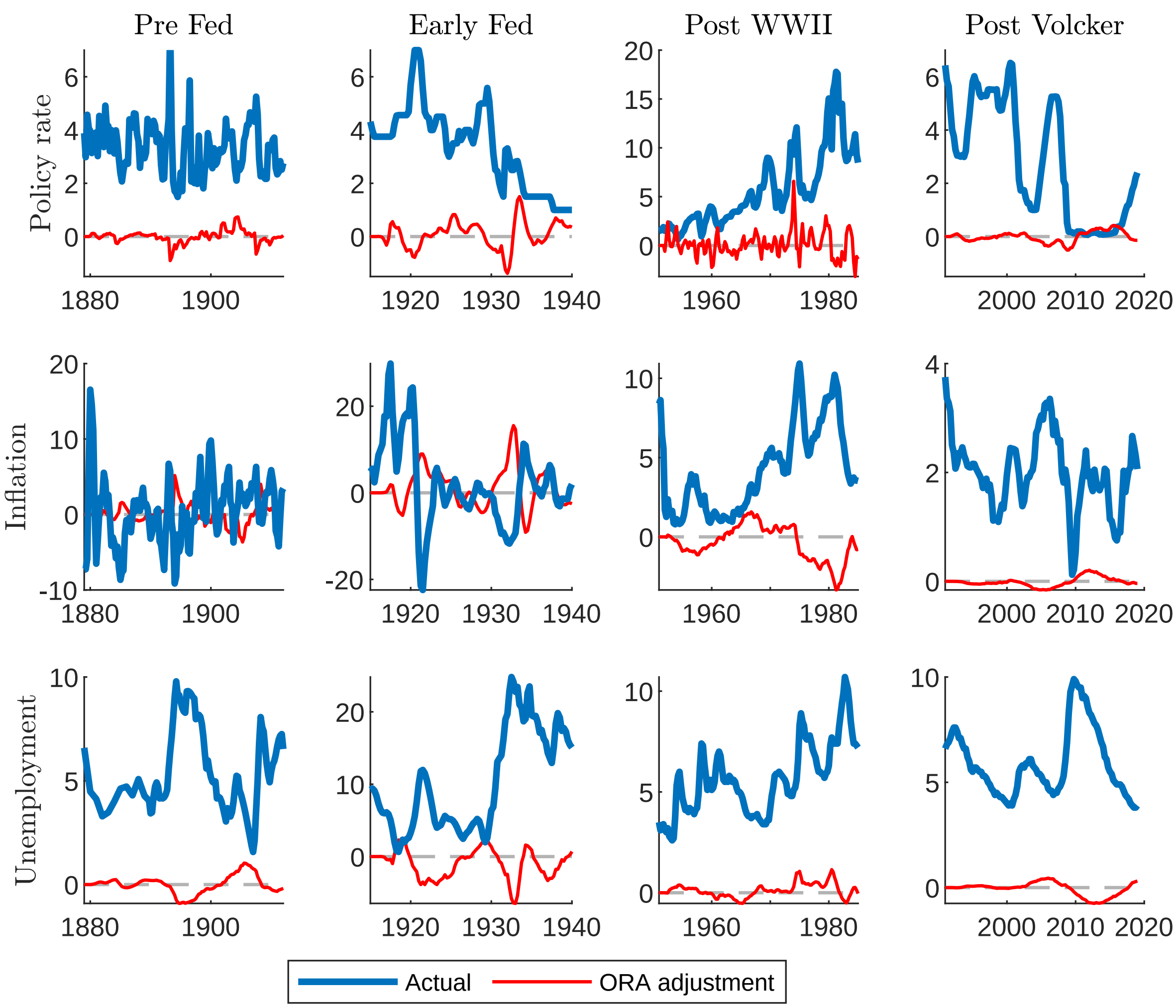
Notes: The top (resp. bottom) row shows the median responses (thick line) of inflation, unemployment and the fed funds rate to a monetary policy shock (resp. financial shock). The dotted green lines show the ORA adjusted impulse responses: $\Gamma_0^0 + \mathcal{R}_0^0 \tau_0^*$. The 95% and 68% credible sets are plotted as dark and light shaded areas, respectively.

Figure 4: Post WWII Fed, 1951-1984, Reaction to π^e shocks



Notes: The top (resp. bottom) row shows the median responses (thick line) of inflation, unemployment and the fed funds rate to a monetary policy shock (resp. inflation expectations shock). The dotted green lines show the ORA adjusted impulse responses: $\Gamma_0^0 + \mathcal{R}_0^0 \tau_0^*$. The 95% and 68% credible sets are plotted as dark and light shaded areas, respectively.

Figure 5: ORA adjustments over 1879-2019



Notes: The top row shows the policy rate ("raw data", blue plain line) along with the adjustment to the contemporaneous policy rate implied by the median ORA adjustment ("ORA adjustment", red line) over each period, calculated following (26). The middle and bottom rows show the same information but for inflation and unemployment. In each panel, the sum of the blue line ("raw data") and the red line ("ORA adjustment") gives the counter-factual ORA-adjusted time series for the policy rate, inflation and unemployment (respectively).