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

September 19, 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 over the past 150 years; 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.

JEL classification: C14, C32, E32, E52.

Keywords: optimal policy, policy evaluation, reaction function, monetary history.

*Preliminary. We thank our discussant Frank Smets as well as Ryan Chahrour, Sylvain Leduc, Adam Shapiro and seminar participants at Paris School of Economics, Mannheim, UC Berkeley, INSPER, PUC Rio, FGV, Bank of Italy, ECB, Riksbank and the International Francqui Chair Symposium in Brussels for helpful comments. Mesters acknowledge support from the Spanish Ministry of Economy and Competitiveness, 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.[1]

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 their 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 different ways each policy maker *reacted* to the same of sequence shocks, i.e., from the different reaction functions.

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, or even war shocks. By comparing how well different policy makers performed in response to such common shocks, we can approach the ideal empirical setting

[1] See Fair (1978) for an early discussion of these points.

sketched above: assessing and comparing performance from the different ways each policy makers 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 counter-factual: how a different reaction would have affected the economy. That counter-factual 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* on 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. Moreover, ORAs are portable moments in the sense of Nakamura and Steinsson (2018): when two policy makers were exposed to common policy shocks and non-policy shocks, we can 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).

Using the ORA the evaluation and comparison of policy makers thus reduces to estimating 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 interesting questions regarding the conduct of monetary policy. To name a few, (i) did the founding of the Federal Reserve in 1913 led to superior macro outcomes than during the Gold standard period (e.g., Bordo and Kydland, 1995)? And if so, by how much? (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? (iii) To what extent is the stable and low inflation environment of 2000s versus the 1970s the outcome of good policy or simply good luck (e.g., Clarida, Galí and Gertler, 2000; Gali and Gambetti, 2009*a*)?

To assess and compare monetary policy performance across historical periods, we evaluate how monetary policy responded to five types of non-policy shocks that are common across periods: (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.

Over these historical periods US monetary policy was confronted with different sequences of shocks and possibly very different economic environments, but our ORA-based policy evaluation allows to compare policy performance over these four periods while by-passing the many confounding factors that have plagued previous comparisons. The (still substantial) empirical challenge is to consistently estimate the impulse responses to common monetary and non-policy shocks over each sub-period. Fortunately, we can leverage on a large empirical literature on structural shocks identification, and we will use as much as possible the state of the art in each setting: Hamilton (2003) for energy price shocks, Ramey and Zubairy (2018) for government spending shocks, Leduc, Sill and Stark (2007) for inflation expectation shocks, Gali (1999) for productivity shocks and Reinhart and Rogoff (2009) for banking panics. As monetary shocks common across periods, we identify shocks to the contemporaneous policy rate. For the post WWII periods, we rely on Romer and Romer (2004*b*) and Gürkaynak, Sack and Swanson (2005), and for the early Fed years we use the Friedman and Schwartz (1963) narrative dates extended by Romer and Romer (1989). Identification is more challenging (and less developed) for the pre WWII periods, and we propose a new identification strategy for monetary shocks for the Gold Standard period. Specifically, we exploit the specificity of the Gold Standard, in that the monetary base depends on the amount of gold in circulation, and we use unanticipated large Gold mine discoveries (discoveries that led to Gold rushes) as an instrument for movements in the monetary base and thereby the short-term rate.

Evaluating and comparing policy makers require 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.[2] Given that loss function, our results point to overall improvements in the conduct of monetary policy. For instance, in response to bank runs, we do find that the policy response is substantially better after 1990 than during the Gold Standard or the early Fed period. During the Early Fed period the monetary response was much too passive in the face of financial disturbances; not lowering the discount rate enough, in fact running a contractionary monetary policy. In contrast, the monetary response was much closer to optimal in the post Volcker period, though the zero-lower bound did constrain partially the Fed's response.

That said, improvements in the conduct of monetary policy have not been monotonic, and the post WWII period saw the worst performance in response to inflationary shocks: energy price shocks, inflation expectation shocks and TFP shocks. For all these shocks, the Fed reaction was much too weak over 1951-1984, particularly in response to inflation expectation shocks. This result echoes a large literature on the Fed's failure to satisfy the Taylor principle in the 1970s, but also goes further by allowing to compute the distance to the optimal reaction —by how much more the Fed should have responded to supply shocks—. In addition, we find that the Fed's excess passivity carries to aggregate demand shocks. For instance, the Fed's interest rate reaction was too weak in the face of the military buildup shocks of the Vietnam war.

Last, we find that performance is universally superior during the post Volcker period: the distances to the optimal reaction coefficients are smallest (and non-significantly different from zero) for all types of shocks that we considered. The only (mild) exception is the reaction to the financial shock for which we find some evidence that the zero-lower bound may have partially constrained the Fed's response, though the distance to an optimal reaction is still substantially smaller than during the early Fed period.

Related literature

Perhaps surprisingly, the literature has produced few methods for evaluating and comparing policy makers over time or over space.

[2]Importantly, our approach could accommodate other loss functions, for instance different loss functions across time periods, or even micro-founded welfare-based loss functions.

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. This approach amounts to specifying a structural model, calculating what would have been the optimal policy based on the model and comparing the loss under such optimal policy to the loss under the implemented policy. This general approach has been used within the context of other structural models, notably New Keynesian models (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). By assessing policy performance in response to *specific* non-policy shocks, we can break the optimal policy assessment problem into smaller parts, which do not require a fully-fledged structural model, and this allows us to greatly reduce the risk of model mis-specification as impulse responses can be estimated with *reduced form* econometric methods that are arguably more robust to model mis-specification. That said, compared to a model-based assessment, our approach will only evaluate performance from the reaction to the identified non-policy shocks that could be identified, a possible subset of all non-policy shocks. Similarly, a comparison of policy makers will be only based on the subset of identified non-policy shocks that were common across policy makers.

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.

In the context of monetary policy, the literature has studied the performance of the Fed pre- and post-Volcker;[3] specifically by assessing whether the Taylor principle —a central bank should react more than one-to-one to inflation movements— was satisfied.[4] However, beyond that Taylor principle, that literature can say little about the optimality of the reaction function, whether the Fed was reacting too much or too little after 1984. Overall, that approach can only provide a coarse evaluation of reaction functions.

Closer to our approach, Galı, López-Salido and Vallés (2003) and Blanchard and Galí (2007) study the effect of technology shocks or oil shocks to assess the performance of the Fed pre- and post-Volcker. Different from our approach however, their assessment of good

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

[4]Based on Taylor rule estimates, these papers found that the parameters of the Taylor rule shifted around 1984 and that the Fed responded more vigorously to inflation variations after 1984, though this conclusion has not gone unchallenged (e.g., Orphanides, 2003).

monetary policy focuses on the response of the real interest rate to a technology shock and on its distance to a specific New-Keynesian model. In contrast, our approach allows to evaluate the distance to optimality for the Fed's reaction to technology or oil price shocks without relying on any specific structural model. Instead, impulse responses are shown to be sufficient statistics for a broad class of structural models.

A less structural literature has proposed ways to study policy rule counter-factuals (e.g., Sims and Zha, 2006; Bernanke et al., 1997; Leeper and Zha, 2003), though those approaches are not fully robust to the Lucas critique. Instead, our approach builds on recent work that shows how robustness to the Lucas critique is possible in a large class of macroeconomic models (McKay and Wolf, 2023). The key underlying assumption in these works is that the underlying (unknown) model has is structured such that the coefficients of the non-policy block are independent of the coefficients of the policy block. The present paper builds on these insights to evaluate policy institutions with minimal assumptions on the underlying economic structure.

Last, our paper relates to the sufficient statistics approach to macro policy evaluation expounded in Barnichon and Mesters (2023). The key difference is that they focus on a different policy problem: 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.

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.[5]

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:

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

The vectors $\Gamma(\phi)$ and $\mathcal{R}(\phi)$ capture the impulse responses of the policy objectives to the structural shocks ξ_t and ϵ_t. Note that $\Gamma(\phi)$ and $\mathcal{R}(\phi)$ also depend on other parameters besides ϕ but we omit this from the notation for now.

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

$$\mathcal{L}_t = \frac{1}{2}(\pi_t^2 + x_t^2) \ . \tag{5}$$

More general forward looking loss functions are considered in the general treatment.

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 irrespective of the parameter values in (1)-(2) due to the inclusion of ξ_t

[5] 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.

in (3) (e.g. Galí, 2015, page 133).

Reaction function evaluation

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

Let $\phi^0 = (\phi^0_\pi, \phi^0_\xi) \in \Phi$ denote the central bank's chosen reaction function. 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

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

where $\Gamma \equiv \Gamma(\phi^0)$ and $\mathcal{R} \equiv \mathcal{R}(\phi^0)$ 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 $\Gamma + \mathcal{R}\tau$ 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 to $\Gamma + \mathcal{R}\tau$, and the impulse response $\mathcal{R}$ contains all the information needed to compute the effect of any 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

$$\begin{aligned}\tau^* &= \underset{\tau}{\operatorname{argmin}} \ \mathbb{E}\mathcal{L}_t \qquad \text{s.t.} \qquad Y_t = (\Gamma + \mathcal{R}\tau)\xi_t + \mathcal{R}\epsilon_t \\ &= \underset{\tau}{\operatorname{argmin}} \ \sigma^2_\xi (\Gamma + \mathcal{R}\tau)'(\Gamma + \mathcal{R}\tau) \ ,\end{aligned} \tag{8}$$

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

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

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[6]

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

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 (Γ and $\mathcal{R}$) 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 and $\mathcal{R}$ 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 —the impulse responses to the cost-push shock— is orthogonal to $\mathcal{R}$ —the impulse responses to the policy shock—, the ORA τ^* is zero and the reaction coefficient ϕ_ξ is optimal. Intuitively, the impulse response to cost-push shocks (Γ) 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 the impulse response to monetary shocks ($\mathcal{R}$) captures what the policy maker could have done to counteract these shocks —how adjusting the reaction coefficient ϕ_ξ by τ could have better stabilized the impulse response Γ—. If Γ and $\mathcal{R}$ are orthogonal, there is nothing more that the policy maker could have done to stabilize Γ.[7] Conversely, if the reaction coefficient ϕ_ξ^0 is 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 the projection of Γ on $-\mathcal{R}$: the goal of the ORA is to use the impulse responses to a monetary shock in order to best

[6] To see this, compute

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

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 Galí (2015, Page 133, equation (10)).

[7] Optimal policy making does not mean that the policy maker succeeded at perfectly canceling the effects of cost-push shocks (making Γ perfectly flat at zero). It means that a different reaction would not have done any better.

stabilize the impulse response to the non-policy shock. This is equivalent to best fitting the vector Γ with the vector $-\mathcal{R}$.

Third, τ^* measures the distance to an optimal reaction function in one specific direction — the systematic policy response to cost-push shocks—, but this is sufficient to characterize the entire optimal reaction function, as $(\phi_\pi^0, \phi_\xi^0 + \tau^*) \in \Phi^{\text{opt}}$. Intuitively this result holds, because cost-push shocks are the only non-policy shocks, so that optimally responding to cost-push shocks is isomorphic to using an optimal reaction function. 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.[8] Conversely, focusing on a subset of these non-policy shocks will 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 key insight underlying our approach, is that the effects of any reaction function are *encoded* in the impulse responses Γ and $\mathcal{R}$, 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_1^0 and ϕ_2^0, respectively, and let the economic environment that they faced be captured by the parameter vectors θ_1 and θ_2, respectively, which include all coefficients in the Phillips and IS curves.

For each policy maker we compute the ORA statistic:

$$\tau_j^* = -(\mathcal{R}^{j\prime}\mathcal{R}^j)^{-1}\mathcal{R}^{j\prime}\Gamma^j \qquad \text{for} \qquad j = 1, 2 \ ,$$

where $\mathcal{R}^j \equiv \mathcal{R}(\phi_j^0, \theta_j)$ and $\Gamma^j \equiv \Gamma(\phi_j^0, \theta_j)$.

Since the ORA measures the distance to the optimal reaction to cost-push shocks, we

[8]Note that the ORA is not focused on evaluating the reaction coefficient to endogenous variables (for instance, assessing the optimality of the reaction coefficient ϕ_π). Evaluating the optimality of ϕ_π is much more difficult without a fully specified structural model, but an insight underlying our approach is that this is also not necessary: to characterize the optimal reaction function, it is sufficient to assess the systematic reaction to all non-policy shocks affecting the economy, see Section 4.

can use the ORA to rank policy makers. For instance, we would rank policy maker 1 above policy maker 2 if $|\tau_1^*| < |\tau_2^*|$.

The key insight is that while environments can be different across policy makers (and thus the optimal reaction function), the ORA statistics τ_1^* and τ_2^* measure the same quantity: the distance to the optimal reaction of the same policy instrument (here the contemporaneous interest rate) to the same non-policy shock (here a cost-push shock). By evaluating the performance of different policy makers in response to the same type of shock, we are able to evaluate and compare policy makers (or more generally policy institutions) who were facing different initial conditions, different shocks and different economic environments.

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 generic 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. For convenience we work under perfect forecast and postulate that the paths for the endogenous variables are determined by the generic model

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

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. While we wrote the model under perfect foresight it is useful to think about ξ_0 and ϵ_0 as the contemporaneous policy shocks, whereas ξ_t and ϵ_t for $t \geq 1$ are the sums of all news shocks arriving from time zero onward, e.g. $\xi_t = \sum_{j=0}^{t} \nu_{t,j}$, with $\nu_{t,j}$ the news arriving at time j about period t.

We normalize all shocks, that is 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). It is useful to 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 shocks satisfy this condition.

The structural maps $\mathcal{A}_{..}$ and $\mathcal{B}_{..}$ are conformable and may depend on underlying structural parameters. We conveniently split them in two parts: the economic environment $\theta = \{\mathcal{A}_{yy}, \mathcal{A}_{yp}, \mathcal{B}_{yx}, \mathcal{B}_{y\xi}\}$ which the policy maker takes as given, and the reaction function $\phi = \{\mathcal{A}_{pp}, \mathcal{A}_{py}, \mathcal{B}_{px}, \mathcal{B}_{p\xi}\}$, 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

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

Overall, model (11) is general and allows future policy decisions to affect current and future outcomes. Many structural models found in the literature can be written in this form; prominent examples include New Keynesian models and more modern heterogeneous agents models, see McKay and Wolf (2023) for a more in dept discussion. We could include initial conditions, stemming from before period zero, but since we will project on the non-policy shocks anyway we avoid such additional notation.

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_t^* for some time periods $t = 0, 1, 2, \ldots$. For ease of notation we will set the targets to zero.[9]

We measure performance using the unconditional loss function

$$\mathcal{L} = \frac{1}{2}\mathbb{E}\mathbf{Y}'\mathcal{W}\mathbf{Y} \ , \tag{12}$$

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.

Importantly, the loss (12) is the researcher's evaluation criterion for scoring policy maker performance, and it may or may not correspond to the preferences of the policy maker herself or of society.

[9] Alternatively we can think of y_t as defined in deviation from the desired targets, either way this restriction is only imposed for notation convenience.

The actions of the policy maker are summarized by the reaction function ϕ. We define a reaction function to be optimal (from the perspective of the researcher) if it minimizes the loss function (12) subject to an economy where $\boldsymbol{\epsilon} = \mathbf{0}$. This imposes that we are interested in evaluation criteria that are based on systematic performance. An alternative would be to evaluate policy makers based on their idiosyncratic mistakes, which is not the objective of our methodology.

Formally, the set of optimal reaction functions is given by

$$\Phi^{\text{opt}} = \left\{ \phi : \phi \in \underset{\phi \in \Phi}{\operatorname{argmin}} \, \mathcal{L} \quad \text{s.t} \quad (11) \text{ with } \boldsymbol{\epsilon} = \mathbf{0} \right\} \ . \tag{13}$$

The definition implies that we only consider optimal reaction functions that lie in Φ, i.e. the set of reaction functions which imply a unique equilibrium. Moreover, we emphasize that optimality is defined with respect to the coefficients of the model (11) and not with respect to the coefficients of any specific underlying structural model.[10] The reason for this is that model (11) is generic in the sense that it does not impose any a priori restrictions on the actions of the policy maker.

4 Measuring reaction function optimality

We propose to evaluate and rank policy makers (or institutions) by measuring the distance between their reaction function, denoted by ϕ^0, and the set of optimal reaction functions Φ^{opt} in (13). 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.

To start, it is useful to note that 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

$$\mathbf{Y} = \Gamma(\phi)\boldsymbol{\Xi} + \mathcal{R}(\phi)\boldsymbol{\epsilon} \ , \tag{14}$$

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. It is useful to realize the similarity between (14) and (4) which was obtained for the illustrative New Keynesian model. The static NK example is a special case with only contemporaneous shocks.

[10] For instance, the NK model from the simple example can be written in the form of (11) and we could define optimality with respect to the coefficients of such underlying model. It is clear that this would compromise the generality of our approach.

Clearly, the maps $\Gamma(\phi)$ and $\mathcal{R}(\phi)$ in (14) 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. The precise mapping from the model coefficients $\mathcal{A}_{..}$ and $\mathcal{B}_{..}$ to $\Gamma(\phi)$ and $\mathcal{R}(\phi)$ is provided in the appendix, but we will not require knowledge of this mapping.

4.1 Optimal reaction adjustments

Consider a policy maker with reaction function $\phi^0 \in \Phi$ which is unknown to the researcher. 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 the augmenting the policy rule under ϕ^0 as follows

$$\mathcal{A}^0_{pp}\mathbf{P} - \mathcal{A}^0_{py}\mathbf{Y} = (\mathcal{B}^0_{p\xi} + \mathcal{T})\mathbf{\Xi} + \boldsymbol{\epsilon} \ , \tag{15}$$

where $\mathcal{T}$ adjusts the response to the non-policy shocks. Given that $\phi^0 \in \Phi$ we can follow the same steps that led to (14) and compute the equilibrium. Specifically, we can combine this augmented policy rule with the non-policy equation in model (11) to obtain the equilibrium representation

$$\mathbf{Y} = (\Gamma + \mathcal{R}\mathcal{T})\mathbf{\Xi} + \mathcal{R}\boldsymbol{\epsilon} \ , \tag{16}$$

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

The equilibrium effect of $\mathcal{T}$ is found to be equal to $\mathcal{R}\mathcal{T}\mathbf{\Xi}$ on the non-policy variables and is proportional to the effect of the policy shocks as captured by $\mathcal{R}$. In other words, as in the simple example, it is possible to compute the effect of a different policy rule — changing the reaction to the non-policy shocks $\mathbf{\Xi}$— from the impulse responses to policy shocks. Such counter-factual analysis is fully robust to the Lucas critique provided that in the (unspecified) underlying economic model, the coefficients of the macro block in (11), i.e. θ, are invariant to changes in the coefficients of the policy rule, i.e. ϕ, see McKay and Wolf (2023); Barnichon and Mesters (2023).

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

$$\mathcal{T}^* = \underset{\mathcal{T}}{\operatorname{argmin}}\, \mathcal{L} \qquad \text{s.t.} \quad \mathbf{Y} = (\Gamma + \mathcal{R}\mathcal{T})\mathbf{\Xi} + \mathcal{R}\boldsymbol{\epsilon} \ , \tag{17}$$

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

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

$$\mathcal{T}^* = -(\mathcal{R}'\mathcal{W}\mathcal{R})^{-1}\mathcal{R}'\mathcal{W}\Gamma \ , \tag{18}$$

which exists provided that the inverse exists. The expression shows that the ORA is equal to the projection of the selected non-policy impulse responses $\mathcal{W}^{1/2}\Gamma$ on the identifiable and selected policy impulse responses $\mathcal{W}^{1/2}\mathcal{R}$. Recall that the weighting matrix $\mathcal{W}$ is merely a selection tool used to select the non-policy variables that are of interest to the researcher.

We summarize the result in the following proposition for which a more formal proof is given in the appendix.

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 shows that fully characterizing the optimal reaction function requires identifying all the different non-policy shocks as well as all policy news shocks. The policy maker's choice for the maps $\mathcal{A}_{pp}$ and $\mathcal{A}_{py}$ is irrelevant — as long as the invertibility requirement for $\mathcal{A}$ is satisfied—, all that is relevant for the researcher can be measured through $\mathcal{B}_{p\xi}$. The result mimics the finding in equation (10) for the baseline NK model for a broad class of macro models.

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. Unfortunately in practice we will not be able to identify all shocks, hence compromising the computation of $\mathcal{T}^*$. Therefore in this section we discuss a policy rule evaluation statistic that requires only a subset of the impulse responses.

To set this up, let $\boldsymbol{\epsilon}_a$ denote any subset or linear combination of $\boldsymbol{\epsilon}$ which can be identified. Similarly, let $\mathbf{X}_b$ denote a linear combination of $\boldsymbol{\Xi}$. We proceed to derive a subset version of the ORA statistic.

We consider the augmented subset of the policy rule

$$\mathcal{A}^0_{p_a p}\mathbf{P} - \mathcal{A}^0_{p_a y}Y = (\mathcal{B}^0_{p_a \xi_b} + \mathcal{T}_{ab})\boldsymbol{\Xi}_b + \mathcal{B}^0_{p_a \xi_{-b}}\boldsymbol{\Xi}_{-b} + \boldsymbol{\epsilon}_a \ , \tag{19}$$

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 policy equations, 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.

$$\mathcal{T}^*_{ab} = \underset{\mathcal{T}_{ab}}{\operatorname{argmin}}\, \mathcal{L} \qquad \text{s.t.} \quad \mathbf{Y} = (\Gamma_b + \mathcal{R}_a\mathcal{T}_{ab})\boldsymbol{\Xi}_b + \Gamma_{-b}\boldsymbol{\Xi}_{-b} + \mathcal{R}\boldsymbol{\epsilon} \ , \tag{20}$$

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

$$\mathcal{T}_{ab}^* = -(\mathcal{R}_a'\mathcal{W}\mathcal{R}_a)^{-1}\mathcal{R}_a'\mathcal{W}\Gamma_b \ , \tag{21}$$

which exists provided that the inverse exists.

Proposition 2. *Given the generic model* (11)*, with* Φ *non-empty, let* $\phi_{ab}^* = \{\mathcal{A}_{pp}^0, \mathcal{A}_{py}^0, \mathcal{B}_{p_a\xi_b}^0 + \mathcal{T}_{ab}^*, \mathcal{B}_{-p_a-\xi_b}^0\}$*, we have that* $\mathcal{L}(\phi_{ab}^*) \leq \mathcal{L}(\phi^0)$ *for all* $\phi^0 \in \Phi$*.*

The result is of great practical relevance as it shows that researchers never have to recover the entire causal maps Γ^0 and $\mathcal{R}^0$ to evaluate the reaction function. For instance, consider a researcher interested in evaluating how a central bank is adjusting its contemporaneous policy rate in reaction to contemporaneous oil price shocks. For a loss function involving the inflation and unemployment gaps, the only requirements are to estimate two sets of impulse responses for inflation and unemployment: the impulse responses to a contemporaneous policy shock and the impulse response to an oil price shock.

The subset ORA statistic has an intuitive two-step interpretation. In a first step we project $\mathbf{Y}$ on the specific non-policy shocks of interest, recalling that all policy shocks are normalized to have unit variance we get

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

This step effectively isolates our object of interest —the response to the *specific* non-policy shocks $\mathbf{\Xi}_b$—, and removes the confounding effects of the other shocks $\mathbf{\Xi}_{-b}$ and $\boldsymbol{\epsilon}_t$ (and the not-modeled initial conditions). In the second step we then solve the policy problem in the projected space, i.e. we solve

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

which defines the ORA statistic. In the web-appendix we discuss a few additional interpretations for the ORA statistic.

4.3 Comparing policy institutions with ORAs

With the ORA and its properties established 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. Our comparisons are based on evaluating policy makers on their use of the same policy instruments for offsetting the same non-policy shocks. As such we may generally compare policy makers from the same institution across different time periods or policy maker from different but comparable institutions from different countries.

Suppose that there are p policy makers that the researcher aims to compare. Each policy maker faces 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 $j = 1, \ldots, p$. Following the notation defined above, let ϕ_j^0 denote the chosen reaction function of policy maker j. While here we treat the parameters as fixed within the term of each policy maker, an extension with time-varying parameters is discussed in the appendix.

To compare policy makers using the ORA, the idea is to compare well they each responded to same non-policy shocks. To do so, the key requirement is that we identify the same policy and non-policy shocks across different policy makers. That is, $\boldsymbol{\epsilon}_a$ and $\boldsymbol{\Xi}_b$ must pertain to the same subset or linear combination of news shocks across policy makers. Going back to our example of a researcher assessing a central bank's reaction to contemporaneous oil shocks, the ORA can allow us to compare different central bankers by comparing how well the contemporaneous policy rate responded to a contemporaneous oil shock. This will require estimating the impulses responses to the same policy shocks and the same oil shock.[11]

The ORA statistics for each policy maker are given by

$$\mathcal{T}_{ab}^{j*} = -(\mathcal{R}_a^{j\prime} \mathcal{W} \mathcal{R}_a^j)^{-1} \mathcal{R}_a^{j\prime} \mathcal{W} \Gamma_b^j \ , \qquad j = 1, \ldots, p \ .$$

We recall that $\mathcal{R}_a^j$ and Γ_b^j are the impulse responses of the objectives with respect to the policy and non-policy shocks computed under the reaction function ϕ_j^0 and given the economic environment θ_j.

The weighting by $\mathcal{W}$ implements the preferences of the researcher over the different objectives or ranking criteria. As such if the researcher has no further preferences over the types of shocks we may simply aggregate the entries of $\mathcal{T}_{ab}^{j*}$ to rank the policy makers, i.e.

$$t_{ab}^{j*} = \|\mathcal{T}_{ab}^{j*}\| \ , \tag{22}$$

where any desired norm $\|\cdot\|$ can be used. We rank policy makers based on t_{ab}^{j*}, for $j = 1, \ldots, p$, where the smallest value corresponds to the best performing policy maker. For interpretation purposes it is generally useful to present the ranking separately for each combination of instrument and non-policy shocks as each ranking is informative about a specific dimension of policy.

[11] This requirement is no different from any other study that compares impulse responses across time or across countries (e.g., Galí and Gambetti, 2009*b*): the same shocks needs to be identified to ensure that the same objects are compared.

4.4 Computing ORA statistics

We discuss the computation of the optimal reaction function adjustments using observational data for a single policy maker. When computing ORAs for different policy makers the recipe of this section can be obviously repeated, or it may be desirable to jointly compute the ORAs.

The starting point for computing the ORA statistic in practice is the equilibrium representation (14) under ϕ^0. Indeed, considering

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

we can note that 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 $\mathbf{\Xi}_b$. 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

$$\mathbf{Y}_t^w = \Gamma_b^w \mathbf{\Xi}_{b,t} + \mathcal{R}_a^w \boldsymbol{\epsilon}_{a,t} + \mathbf{V}_t^w \ , \qquad t = t_s, \ldots, t_e, \tag{23}$$

where $\mathbf{\Xi}_{b,t}$ and $\boldsymbol{\epsilon}_{a,t}$ are the linear combinations 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 (23) 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 using standard local projection methods. 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, we recall from Plagborg-Møller and Wolf (2021) that in population local projections and structural VARs estimate the same impulse responses; therefore all SVAR methods discussed in Kilian and Lütkepohl (2017), for instance, 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 both frequentist (asymptotic and bootstrap) and Bayesian estimators.

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. We note that if the distribution F is normal we can use the delta method to analytically compute the distribution of $\mathcal{T}_{ab}^*$, but we generally recommend using bootstrap or Bayesian methods.

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.[12]

That said, the ORA statistics can be used as building blocks for computing various policy counterfactuals. The first counterfactuals of interest are the adjusted non-policy impulse responses $\Gamma_b^w + \mathcal{R}_a^w\mathcal{T}_{ab}^*$, 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 average effects of the non-policy shocks.

Using the identified structural shocks we can also compute ORA-based historical decompositions which directly quantify how observed variables would be different after ORA. Specifically, given the identified non-policy shocks $\mathbf{\Xi}_{b,t}$, we compute

$$\Delta\mathbf{Y}_t^w = \mathcal{R}_a^w\mathcal{T}_{ab}^*\mathbf{\Xi}_{b,t} \qquad \text{and} \qquad \Delta\mathbf{P}_t^w = \mathcal{R}_{p,a}^w\mathcal{T}_{ab}^*\mathbf{\Xi}_{b,t} \ , \quad \text{for} \quad t = t_s, \ldots, t_e \ , \tag{24}$$

where $\mathcal{R}_{p,a}^w$ are the impulse responses of the policy instruments to the subset of policy news shock. Subsequently we can sum the changes to the same variables, that are caused by shocks from different time periods, to get an overall measure of the consequences of the

[12] 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 changes need not change one-for-one in changes in the policy rate.

sub-optimal reaction.[13] Also, we can compute change in the loss from

$$\Delta \mathcal{L}_t = (\Delta \mathbf{Y}_t^w)'(\Delta \mathbf{Y}_t^w) \ , \tag{25}$$

which when summed over the policy maker's term, or over different shocks, can quantify the overall economic consequences of the ORA adjustments.

We stress that the magnitudes of these counterfactuals $\Delta \mathbf{Y}_t^w$ and $\Delta \mathbf{P}_t^w$ cannot be interpreted as magnitudes of "policy failures" across periods, as these magnitudes are not comparable across periods. The reason is that if the economic environments are different across periods (as is most likely the case), a given Optimal Rule Adjustment can have different effects on the endogenous variables: the effects can can be amplified or reduced depending on the economic environment and on 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.

That said the counterfactuals are economically interesting as they allow to assess the costs of the sub-optimal reaction. Alternatively put, they can be used to measure the good/bad luck of a policy maker. For instance, consider two policy makers with identical ORAs, if (25) implies much larger losses for one of the policy makers, we can conclude that this policy maker was less lucky in that the same deviation from optimality had larger consequences under her term.[14]

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 1876-2020 period using quarterly data. 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 1990-2020.

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. A Gold Standard monetary regime is now generally considered a sub-optimal regime with excessive fluctuations in inflation and unem-

[13] To give a concrete example, let $\mathbf{Y}_t^w = (\pi_t, \ldots, \pi_{t+H})'$ such that in each period t the researcher is interested in controlling inflation over H horizons. It follows that π_t will enter in $\mathbf{Y}_t^w, \mathbf{Y}_{t-1}^w, \ldots, \mathbf{Y}_{t-H}^w$ and the shocks $\boldsymbol{\Xi}_{b,t}, \ldots, \boldsymbol{\Xi}_{b,t-H}$ all imply a different counterfactual for π_t. We then sum the changes that the different shocks imply on π_t.

[14] This is a different interpretation of the good luck/bad luck hypothesis, which stipulates that a period of macroeconomic stability could be the lucky result of small shocks, see e.g., Galí and Gambetti (2009*b*).

ployment (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*).[15] The post Volcker period starts in 1984 —the beginning of the so-called Great Moderation period— and ends right before the pandemic.

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

$$\mathcal{L} = \frac{1}{2}\mathbb{E}\sum_{h=0}^{H}\beta^h(\pi_{t+h}^2 + \lambda u_{t+h}^2) \ , \tag{26}$$

where π_t denotes inflation, u_t the unemployment rate, β the discount factor and λ the preference parameter. Our baseline choice for the loss function sets $\beta = \lambda = 1$ and considers $H = 20$ quarters.[16] Inflation is measured as year-on-year inflation based on the output deflator from Balke and Gordon (1986). 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). The top panel in Figure 1 shows the time series for inflation and unemployment.

5.1 Naive approach

To provide a benchmark for our results, we first evaluate the Fed based on realized outcomes for inflation and unemployment, as shown in Figure 1. In terms of first moments (first two rows of 1), the Early Fed period fares very poorly because of high average unemployment (driven by the Great Depression), while the post WWII period fares poorly because of high average inflation (driven by the Great Inflation of the 1970s). Instead, the Gold Standard period appears (perhaps surprisingly given common wisdom) as the most successful monetary regime, even better than the post Volcker period often referred to as the Great Moderation. In terms of second moments however (last two rows of 1), the Pre Fed period now also fares poorly out with very high inflation volatility, at least much higher than after World War II,

[15] 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. In April 1942, at the request of the Department of the Treasury, the Fed formally committed to maintaining a low interest-rate peg on short-term Treasury bills and also implicitly capped the rate on long-term Treasury bonds. The goal of the peg was to stabilize the securities market and allow the federal government to engage in cheaper debt financing of World War II. This system lasted until the 1951 Treasury-Fed accord separated government debt management from monetary policy (Romero, 2013).

[16] The robustness of our findings with respect to these choices is assessed in the web-appendix. We stress that the choice for the loss function is merely an evaluation criteria in our context what the true loss function of the different Fed chairs was is irrelevant from our perspective.

though again the worse period is the early Fed with very high volatility in both inflation and unemployment. The post-Volcker period shows the most stable inflation by far, though the volatility of unemployment is comparable to that found in the Pre Fed and post WWII periods. Overall, the Post Volcker period appears as the most successful one, followed by the Post WWII and the Pre Fed period roughly on a par, and with the Early Fed period displaying the worse outcomes by far.

Unfortunately, these unconditional realized outcomes cannot be used to assess monetary policy performance. While they could be due to poor monetary policy —an inadequate reaction function—, many co-founding factors outside the Fed control could also explain those results. For instance, the poor realizations in terms of inflation and unemployment 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 across time, we thus 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 $y_t = (\xi_t, \pi_t, u_t, \epsilon_t, p_t, w_t')'$, where w_t denotes additional control variables. We have

$$A_0 y_t = A_1 y_{t-1} + \ldots + A_p y_{t-p} + e_t \ , \tag{27}$$

where $A_0, \ldots, A_p$ are the coefficient matrices and e_t captures the structural shocks. ϵ_t is chosen as the conventional contemporaneous monetary policy shock, and ξ_t can correspond to shock to energy prices, financial intermediation, productivity, government spending and inflation expectations, see the discussion below.

We order the non-policy shock, but order the monetary shock after unemployment and inflation (and before the federal funds rate), for "exogeneity insurance" as in Romer and Romer (2004*b*).

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). Importantly, while we rely on a structural VAR for estimating the impulse responses, other estimation methods could have been used, for instance local projections (Jordà, 2005; Stock and Watson, 2018).

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 is important to ensure comparability of the shocks across periods. With the draws of the parameters from the posterior density we compute the impulse responses $\mathcal{R}_0$ and Γ_0, and the subset ORA statistic $\mathcal{T}_0^*$ using (20). Besides reporting Γ_0 we also report the adjusted $\Gamma_0 + \mathcal{R}_0\mathcal{T}_0^*$ to assess how the impulse response to the non-policy shocks could have been adjusted.

5.3 Shock identification

For each of our periods, we identify a contemporaneous 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 exgenous shocks, we need to identify contemporaneous shocks to the policy rate, that is shocks $\epsilon_{t,t}$. We consider two main approaches for identifying such shocks. As our baseline we use the state of the art in the literature, 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 use surprises in fed funds futures prices around FOMC announcement as proxies for monetary shocks. Since forward-guidance was used extensively during that period (at least after 2007), time t monetary shocks could a priori mix different monetary news shocks: shocks to the contemporaneous policy rate (our object of interest $\epsilon_{t,t}$), news shocks about future monetary policy ($\epsilon_{t,t+h}, h > 0$), as well as news shocks to the contemporaneous policy rate that were announced before time t ($\epsilon_{t,t-j}, j > 0$). Fortunately, fed funds futures allow us to isolate contemporaneous shocks ($\epsilon_{t,t}$). First, to eliminate news shocks about future monetary policy ($\epsilon_{t,t+h}, h > 0$), we use surprises to fed funds futures at a short horizon, here 3-months ahead fed funds futures (FF4). With quarterly data, this ensures that the identified policy surprises do not include news shocks to the future path of policy. Second, since fed funds futures at time t are based on the time t information set, they already includes news shocks that were announced before time t ($\epsilon_{t,t-j}, j > 0$). As a result, surprises to FF4 allow us to isolate surprises to the contemporaneous policy rate.[17]

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 consider that these monetary shocks capture solely contemporaneous policy shocks ($\epsilon_{t,t}$) and not news shocks to policy.

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". The Friedman and Schwartz (1963) dates are also modeled using dummies. Again, since the concept of forward guidance in policy did not exist, we consider that these monetary shocks capture solely contemporaneous policy shocks ($\epsilon_{t,t}$) and not news shocks to policy. In addition, the narrative accounts underlying these dates do not point to any elements of forward guidance: all dates capture shocks to monetary policy within the quarter (Romer and Romer, 1989).

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

[17] In other words, we have $p_{t,t} = \mathbb{E}_t p_t + \epsilon_{t,t}$ and HFI policy surprises measure $\epsilon_{t,t} = p_{t,t} - \mathbb{E}_t p_t$.

the 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.[18] 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 where the mine output reached its peak production—. Figure 2 plots gold production along with our identified dates: the peak of the Comstock Lode mine in Nevada in 1877-Q1, the discovery of Gold in Alaska in 1896-Q3 (which led to the Alaska gold rush), the discovery of large Gold mines in Nevada in 1902-Q1 (which led to the Nevada gold rush), the Nevada gold mine maximum in 1910-Q1.[19] We code the Gold shocks as one when a new mine was discovered and minus one when the peak was reached.

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.[20] 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 it 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, lowers inflation after four quarters and raises unemployment after four quarter. Other than that, the responses are unconstrained. [21] As we will see the results

[18]Given the unpredictability of the amount of gold available in any given region (either at the onset of a gold rush or at its zenith), we can consider these events as unanticipated.

[19]Another important date is the Alaska gold mine maximum in 1915, though it is not useful in this context since the strict gold standard era stops in 1913 with the founding of the Fed.

[20]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.

[21]One drawback of this approach is that a VAR sign-restriction approach need isolate solely contemporaneous monetary shocks. Since the VAR controls for the information set only through observed macro variables, the VAR residuals —and thus our resulted identified monetary shocks— may mix contemporaneous shocks ($\epsilon_{t,t}$) with news shocks revealed before time t ($\epsilon_{t,t-j}, j > 0$) and 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.

from the sign restricted identification scheme are remarkably in line with the results obtained with our baseline identification of monetary shocks.

5.3.2 Non-policy shocks

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 in the spirit of Romer and Romer (2017)'s scaling of their financial distress index.[22]

Government spending shocks For government spending shocks we consider 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 1890-2020.

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 has been responding to innovations to inflation expectations, with the clear example being the de-anchoring of inflation expectations in the 1970s. To do so, we aim to identify inflation expectation shocks, meant to capture threats to the anchoring of inflation expectations.

[22] 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).

As measure of inflation expectations, we rely on the Livingston survey that has been continuously run over 1946-2019.[23] Of interest for us, the Livingston survey 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.[24]

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

We split our results into two parts. First, we discuss the ORA statistics and compare the Fed over time. Second, we zoom in on the specific sub-periods and assess the economic magnitudes of the improvements in the reaction function.

ORA-based assessments of monetary policy over 1879-2019

Table 2 shows the baseline ORA statistics computed using our baseline monetary policy shocks and baseline non-policy shocks. The key benefit of the ORA is that the distance to the optimal reaction coefficient is comparable across periods —capturing how the policy maker should have adjusted its reaction to the same non-policy shock—, so that elements within the same columns are comparable and allow to compare policy performance (with respect to a specific non-policy shock) across periods. Moreover, while the magnitude of the reaction function adjustment is difficult to interpret economically,[25] the sign is easy to

[23]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).

[24]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.

[25]Recall that the ORA is a reaction coefficient adjustment. 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 in order to minimize the loss function., i.e., to best stabilize the impulse responses to that non-policy shock. The ORA is difficult to interpret economically in terms of impulse responses because that coefficient adjustment will also affect the equilibrium allocation indirectly through the feedback terms in the policy rule.

understand: a negative sign implies that the policy response was too weak (i.e., too passive) in the face of a given type of shock.

Overall, we find strong improvements in the conduct of monetary policy, but *only* in the last 30 years, i.e., after Volcker dis-inflation program. Before that, we repeatedly estimate large and significant ORAs to non-policy shocks, particularly before World War II, but also after 1951 with monetary policy being generally too passive in the face of both aggregate demand type shocks and aggregate supply type shocks.

We first describe the general results, contrasting performance across the different periods, before turning to the more specific results in the next section.

During the Gold Standard period, monetary policy is (unsurprisingly) too passive in the face of adverse shocks: whether these are bank runs, military build ups or technological progress: Faced with these adverse non-policy shocks, a monetary authority should have lowered its discount rate.[26]

Interestingly, this excess passivity continues to hold (in fact worsens) after the creation of the Fed. We find that the Early Fed not only responds much too little in the face of bank runs —a result echoing previous findings in the literature (e.g., Friedman and Schwartz, 1963; Hamilton, 1987)—, but also in the face of government spending shocks —letting government spending shocks excessively affect unemployment and inflation— or inflation expectation shocks, in particular not reacting enough to the negative inflation expectation shocks of 1931-1932.

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. Table 2 confirms and generalizes this assessment to most aggregate shocks: the reaction function is too weak in the face of increasing government spending shocks, energy prices and especially inflation expectations. This latter result extends the findings of Leduc, Sill and Stark (2007) that the largest deviations from the Taylor principle —the failure to raise the real rate in the face of rising inflation— were in terms of the response to inflation expectation shocks. Importantly however, the ORA allows to go much further by measuring not just whether the Taylor principle holds or not but also by how much the policy rule should have been adjusted in order to minimize the loss function. We will come back to these important points in the next section.

Last, the post Volcker shows improvements in monetary policy across all dimensions, with ORA statistics much smaller than in the earlier regimes, and not statistically significant, including the reaction to financial shocks.

[26]Interestingly, the response to energy price shocks is appropriate because energy price shocks only increased inflation for a brief period; too fast for monetary policy (and its transmission lags) to have time to materially prevent the inflation increase, see Figure S9 in the Appendix.

In the appendix, we show robustness to our identification of monetary shocks. The results are remarkably consistent with our baseline estimates, with ORAs generally of similar magnitudes and same levels of statistical significance.[27] The only notably exception is the response to financial shocks in the post Volcker period, where the deviation from optimality is small but negative and significantly different from zero. The presence of the zero lower bound could thus have limited somewhat the Fed's ability to best react to the 2007-2008 financial crisis, though again the deviation from optimality is substantially smaller than in earlier periods, in line with our conclusion of substantial improvements in the conduct of monetary policy.

ORA-based adjustments to monetary policy over 1879-2019

The ORAs reported in Table 2 capture how much the Fed should have adjusted the reaction coefficient of the contemporaneous policy rate to a one standard-deviation non-policy shock in order to minimize the unconditional loss function (26). While the ORAs are useful to compare reaction functions across periods, the magnitudes can be hard to interpret. Thus, we will show how the ORA affects the impulse responses to non-policy shocks, effectively showing how these ORA adjustments translate into different policy path responses to non-policy shocks and "improved" (i.e., more stable) impulse responses of inflation and unemployment.

The appendix display all the ORA-adjusted impulse responses, but we will discuss the most interesting ones in the main text.

Responding to financial shocksThat's it, let's submit ... NOW For the early Fed period, we estimate a particularly large ORA statistic in response to financial shocks, and Figure 4 displays the corresponding impulse responses underlying this ORA estimated over 1913-1940. For comparison, Figure 5 displays the same impulse responses but estimatd over 1990-2019. In both figures, the top row shows the impulse responses of inflation, unemployment and the interest rate to a monetary policy shock.[28] The bottom rows show the responses of the same variables to a financial shock: for both periods, inflation contracts whereas unemployment increases, though the inflation response is more muted for the post-Volcker period, consistent with the anchoring of inflation expectations.

The ORA adjusted impulse responses $\Gamma_b + \mathcal{R}_a \mathcal{T}^*_{ab}$ in the bottom rows of Figures 4 and 5 (dashed green lines) show how adjusting the reaction coefficient to financial shocks would have changed the impulse responses of inflation, unemployment and the policy rate. For

[27]The online appendix presents an additional robustness exercise: robustness to the definition of the different monetary periods.

[28]Note that the magnitude of the response of inflation is very different across the periods; the responses being much small in the post Volcker period, consistent with the anchoring of inflation expectations in the recent period (e.g., Gürkaynak, Levin and Swanson, 2010).

the early Fed period, we find that in response to an adverse shock, the Fed *raised* the discount rate. In other words, the Fed was not only too passive, but in fact was following a contractionary policy. Combined with the decline in inflation caused by the financial shock, this means that the real policy rate increased substantially following financial shocks. This finding echoes an earlier literature on the monetary factors behind the Great Depression (e.g., Friedman and Schwartz, 1963; Hamilton, 1987). Instead, the ORA calls for lowering the discount rate. Since this lower policy rate also mutes the inflation decline, this means that the real policy rate now declines in response to a financial shock: an expansionary monetary policy. In contrast, in the post Volcker period, the ORA calls for little adjustment to the path of the fed funds rate.

Responding to energy price and inflation expectation shocks Figures 6 and 7 plot impulse responses estimated over the post WWII period. The impulse response to monetary shocks are in line with earlier evidence (e.g., Coibion, 2012). In response to an energy price shock or an inflation expectation shock, inflation rises progressively, while the policy rate response is relatively mild. In fact, in both cases the real interest rate declines following an energy price shock or inflation expectation shock. In other words, the Taylor principle is not satisfied, a finding echoing an earlier literature on the performance of the Fed during the 1970s (Clarida, Galí and Gertler, 2000; Leduc, Sill and Stark, 2007). However, the ORA goes further and allows us to compute how the policy rate should have responded to these shocks, as displayed by the dashed green lines. The ORA adjustment restores the Taylor principle and the real rate rises following both shocks. Most striking is the large response of the policy rate to an inflation expectation shock (lower-right panel, Figure 7), which contrasts with the absence of response before ORA adjustment.

Responding to government spending and TFP shocks While earlier studies of Fed performance have generally focused on the (weak) response of the Fed to financial distress in the early 1930s or to the (weak) response of the Fed to inflationary pressures during the 1970s, our approach allows to study and compare performance in response to many other shocks. Figures 8 plots the impulse responses to negative government spending shocks in the Early Fed period. We can see that the Fed's excessive passivity extends to government spending shocks as well. In fact, instead of lowering the discount rate to tame the rise in unemployment following a decline in government spending, the Fed was actually raising the discount rate, worsening the negative unemployment effect of the adverse shock to public spending (green line, lower right panel).

In a similar vein, Figure S16 plots the impulse responses to negative TFP shocks for the post WWII period. Again, the Fed response is too weak, leading to excessive inflation. The

ORA adjustment restores a contractionary monetary policy by raising the real interest rate following an adverse TFP shock (green line, lower right panel).

Counterfactual historical policy scenarios

With the ORA in hand, we can also create counter-factual historical policy scenarios using (24) and (25). We use the median ORA estimates to compute the counterfactuals. Figure 10 plots the effects of the sum of the ORA adjustments (across the different non-policy shocks) on the paths of the policy rate, the inflation rate and the unemployment rate. These counter-factuals are interesting from a historical perspective, highlighting moments in time where the consequences of the sub-optimal reactions.

In the pre Fed period, there were substantial deviations from an optimal reaction coefficients, calling 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 (Figure 11, left column). That said, over the per Fed period, our identified non-policy shocks explain only a small share of the total variance of inflation and unemployment over 1879-1912, so that the ORA corrections only have a moderate effect on the behavior of inflation and unemployment over that period.

In the early Fed period, the ORA calls for large adjustments. First, the large increase in military spending over 1918 is responsible for part of the inflation outburst in 1919-1920, and the ORA calls for an almost 1ppt higher discount rate to tame that increase. Second, and most strikingly, the ORA calls for large adjustments in the early stage of the Great Depression (1931-1932). In response to the bank runs and the negative inflation expectation shocks of 1931 (Figure 11), the ORA cancels the discount rate increases 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 subsequently lowers the discount rate all the way to almost (but still above) zero in 1932. This would have avoided about 10 percentage points in unemployment —as much as half of the rise in unemployment over 1930-1932— as well as the deflation. The re-inflation shock following Roosevelt's election would have then been countered by a higher discount rate.[29]

Turning to the post world war II period, the ORA calls for substantially higher fed funds rate throughout the 1960-1980 period. While the ORA does call for more tightening (about 1/2 ppt) in the face of large government programs related to US space program in the early 60s and the Vietnam war in the second half of the 60s, the largest adjustments by far occur

[29] Of course, had the US avoided the deflation and the large unemployment run-up, such a re-inflation shock would not have happened, for instance because Roosevelt may not even have been elected. This counter-factual exercise is based on an ex-post realization of shocks.

during the 70s (as much as 6 ppt higher fed funds rate) in response to the oil price shocks of 1974 and 1979 and the inflation expectation shocks occurring during that period. With such strong response, as much as about 5 ppt of inflation could have been avoided (at the cost of relatively mild extra unemployment), see Figure 10. In fact, one might argue that the "trend" inflation of the 1970s could have been avoided.

In contrast, during the post Volcker period, the ORA policy rate adjustments are small and never significantly different from zero.

6 Conclusion

In this paper, we showed that it is possible to evaluate and compare policy makers based on the distance-to-optimality of their reaction function coefficients to well-chosen non-policy shocks. We introduced ORA statistics to measure the distance and showed that these could be computed from two sufficient statistics: (i) the impulse responses of the macro 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.

Intuitively, our approach consists in projecting the policy objectives on the space spanned by non-policy shocks and then studying the optimal policy problem in that sub-space. Thanks to the projection, it is possible to evaluate the optimality of a *specific* reaction function coefficient: how "well" the policy maker or institution reacted to a specific non-policy shock over a given period. The idea is then to compare policy makers by comparing their distance to the optimal reaction coefficient to the same non-policy shock. Since computing the distance to the optimal reaction coefficient only requires impulse response estimates, it becomes possible to evaluate and compare policy makers or policy institutions across time or even space.

While this paper studied the performance of US monetary policy over the past 150 years, the methodology could be applied to many other important questions, such as comparing the performance of different central banks over the same period, e.g., the Fed vs the ECB during the Great Recession, or comparing the performances of policy makers over time, e.g., democrats vs republicans (Blinder and Watson, 2016).

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 to Macro Policy Evaluation.” *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, 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.

Coibion, Olivier. 2012. "Are the Effects of Monetary Policy Shocks Big or Small?" *American Economic Journal: Macroeconomics*, 4: 1–32.

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

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.

Gali, Jordi, and Luca Gambetti. 2009*a*. "On the Sources of the Great Moderation." *American Economic Journal: Macroeconomics*, 1(1): 26–57.

Galí, Jordi, and Luca Gambetti. 2009*b*. "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, Andrew Levin, and Eric Swanson. 2010. "Does Inflation Targeting Anchor Long-Run Inflation Expectations? Evidence from the US, UK and Sweden." *Journal of the European Economic Association*, 8(6): 1208–1242.

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.

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.

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.

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.

Nakamura, Emi, and Jón Steinsson. 2018. "Identification in Macroeconomics." *Journal of Economic Perspectives*, 32(3): 59–86.

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.

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.

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.

Romero, Jessie. 2013. "Treasury-Fed Accord." *Federal Reserve History.*

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

We first discuss how the general model (11) can be written as (14). Define

$$\mathcal{A} = \begin{bmatrix} \mathcal{A}_{yy} & \mathcal{A}_{yp} \\ \mathcal{A}_{py} & \mathcal{A}_{pp} \end{bmatrix} \;,\; \mathcal{B}_\xi = \begin{bmatrix} \mathcal{B}_{y\xi} \\ \mathcal{B}_{p\xi} \end{bmatrix} \;,\; \mathbf{J} = \begin{bmatrix} \mathbf{0} \\ \mathbf{I} \end{bmatrix} \text{ and } \mathbf{Z} = \begin{bmatrix} \mathbf{Y} \\ \mathbf{P} \end{bmatrix} \;. \tag{28}$$

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 $\mathcal{D} = (\mathcal{A}_{yy} - \mathcal{A}_{yp}\mathcal{A}_{pp}^{-1}\mathcal{A}_{yp})^{-1}$. Recalling that $\phi = \{\mathcal{A}_{pp}, \mathcal{A}_{py}, \mathcal{B}_{p\xi}\}$ is useful to note that

$$\Gamma(\phi) = \Gamma(\{\mathcal{A}_{pp}, \mathcal{A}_{py}, \mathbf{0}\}) + \mathcal{R}(\phi)\mathcal{B}_{p\xi} \;. \tag{29}$$

Proof of Proposition 1. The proof proceeds in two steps: (a) we show the equivalence for $\{\min_\phi \mathcal{L} \text{ s.t. (11) with } \boldsymbol{\epsilon} = \mathbf{0}\} = \{\min_{\mathcal{B}_{p\xi}} \mathcal{L} \text{ s.t. (11) with } \boldsymbol{\epsilon} = \mathbf{0} \;, \mathcal{A}_{pp} = \mathcal{A}_{pp}^0,\ \mathcal{A}_{py} = \mathcal{A}_{py}^0\}$ and (b) we show that the value for $\mathcal{B}_{p\xi}$ that solves the second problem is $\mathcal{B}_{p\xi}^0 + \mathcal{T}^*$. 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

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

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 $\mathcal{R}^0 = \mathcal{D}^0\mathcal{A}_{yp}^0(\mathcal{A}_{pp}^0)^{-1}$ and if $\mathcal{B}_{p\xi}^0 = \mathbf{0}$ we have that $\Gamma^0 = \Gamma(\{\mathcal{A}_{pp}^0, \mathcal{A}_{py}^0, \mathbf{0}\}) = \mathcal{D}^0\mathcal{B}_{y\xi}$. 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 $\mathcal{B}_{p\xi}^0 \neq \mathbf{0}$, using (29) we have $\mathcal{B}_{p\xi}^0 + \mathcal{T}^* = \mathcal{B}_{p\xi}^0 - (\mathcal{R}^{0'}\mathcal{W}\mathcal{R}^0)^{-1}\mathcal{R}^{0'}\mathcal{W}\Gamma^0 = \mathcal{B}_{p\xi}^0 - (\mathcal{R}^{0'}\mathcal{W}\mathcal{R}^0)^{-1}\mathcal{R}^{0'}\mathcal{W}\Gamma(\{\mathcal{A}_{pp}^0, \mathcal{A}_{py}^0, \mathbf{0}\}) - (\mathcal{R}^{0'}\mathcal{W}\mathcal{R}^0)^{-1}\mathcal{R}^{0'}\mathcal{W}\mathcal{R}^0\mathcal{B}_{p\xi}^0 = \mathcal{B}_{p\xi}^*$. □

Proof of Proposition 2. We have that

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

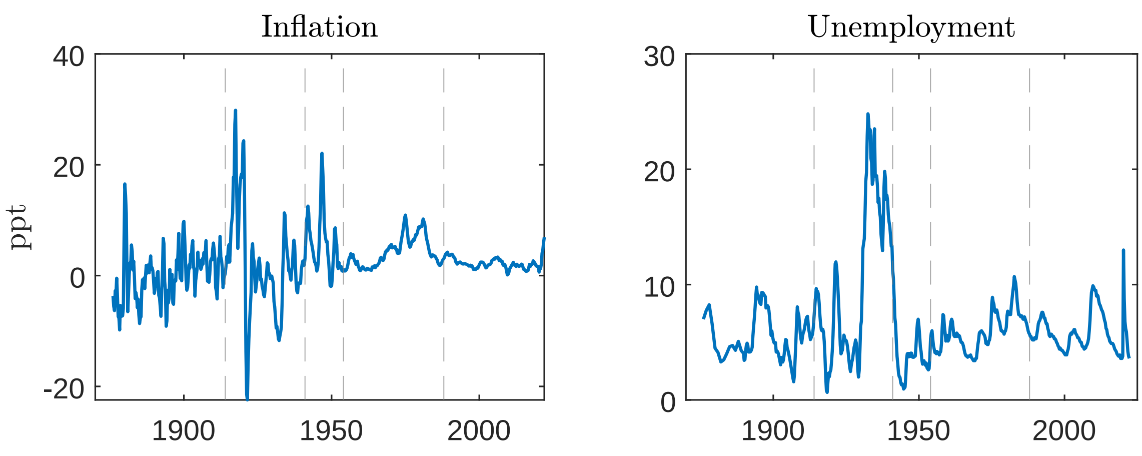
□

Table 1: Realized Outcomes

	Pre Fed 1879-1912	Early Fed 1913-1941	Post WWII 1951-1984	Post Volcker 1990-2019
$\overline{\pi}$	0.4	1.9	4.3	2.0
$\overline{u}$	5.3	10.2	5.6	5.9
Var(π)	19.4	90.1	7.7	0.5
Var(u)	3.5	48.6	3.2	2.6

Notes: $\overline{\pi}$ and $\overline{u}$ denote the sample means of inflation and unemployment, and Var(π) and Var(u) denote the sample variances of inflation and unemployment, as computed over the different periods.

Figure 1: Inflation and unemployment, 1876–2020



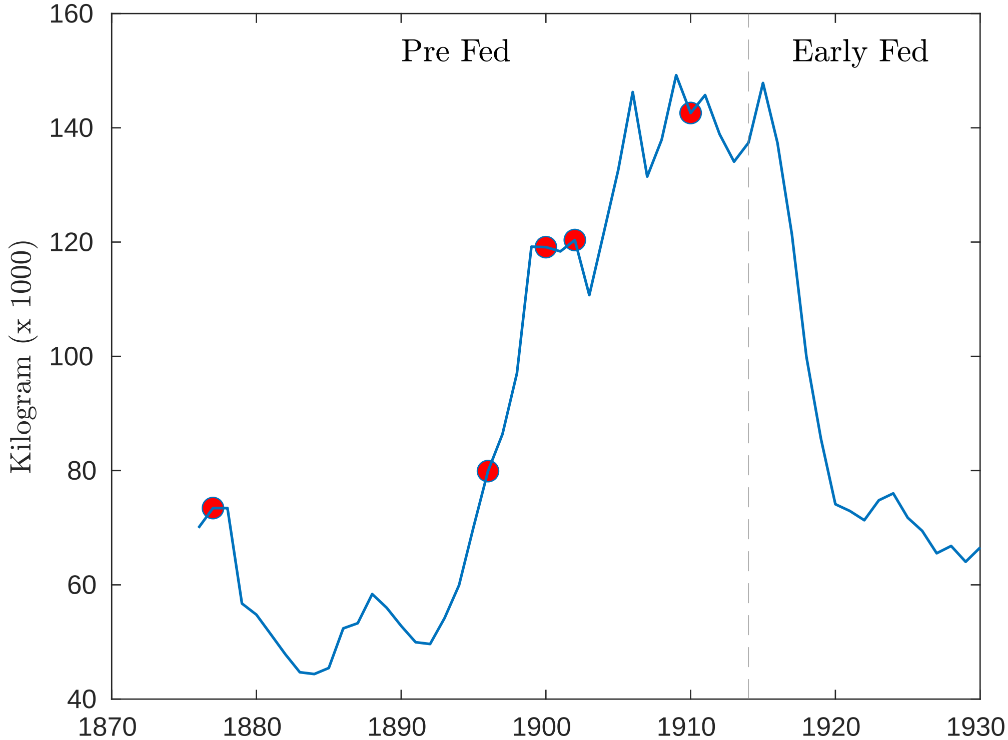
Notes: Year-on-year inflation based on the output deflator and the unemployment rate. The vertical lines highlight the different periods: pre-Fed 1879-1912, early-Fed 1913-1941, post-WWII 1951-1984 and post-Volcker 1990-2020.

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 $\pi\uparrow$
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)
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)
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)
Post Volcker 1990−2019	**−0.1** (−0.5,0.5)	**0.1** (−0.7,1.0)	**0.2** (−0.5,1.1)	**−0.1** (−0.4,0.4)	**−0.1** (−0.6,0.2)

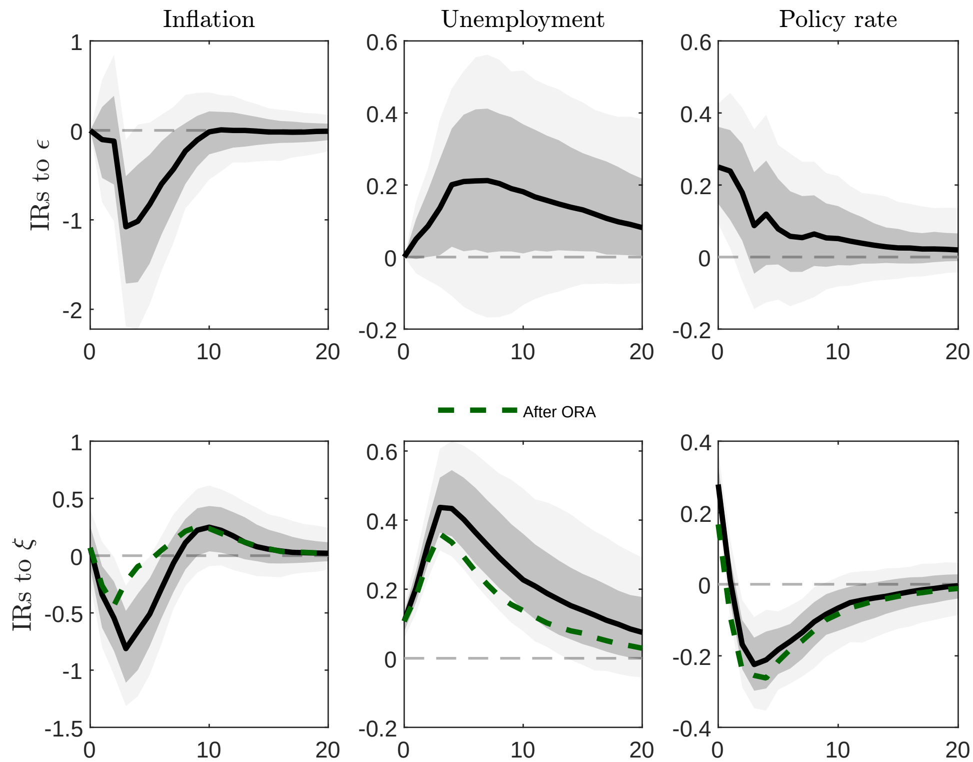
Notes: Median ORA statistics together with 68% credible sets. The monetary policy shocks are identified as described in main text: using gold rush discovery in the pre-Fed period, Romer and Romer (1989)'s Friedman-Schwartz dates in the early Fed period, Romer and Romer (2004) monetary shocks for the post WWII period and high-frequency surprises in the post Volcker period. The financial shocks are bank panics or innovations to the BAA-AAA spread, the government spending shocks are from (Ramey and Zubairy, 2018, G), TFP shocks from (Gali, 1999, TFP), 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 (Post WWII and Post Volcker periods). For the Pre Fed period the TFP, G and Energy ORAs are computed over the 1890-1913 period.

Figure 2: US Gold Production, 1876–1930



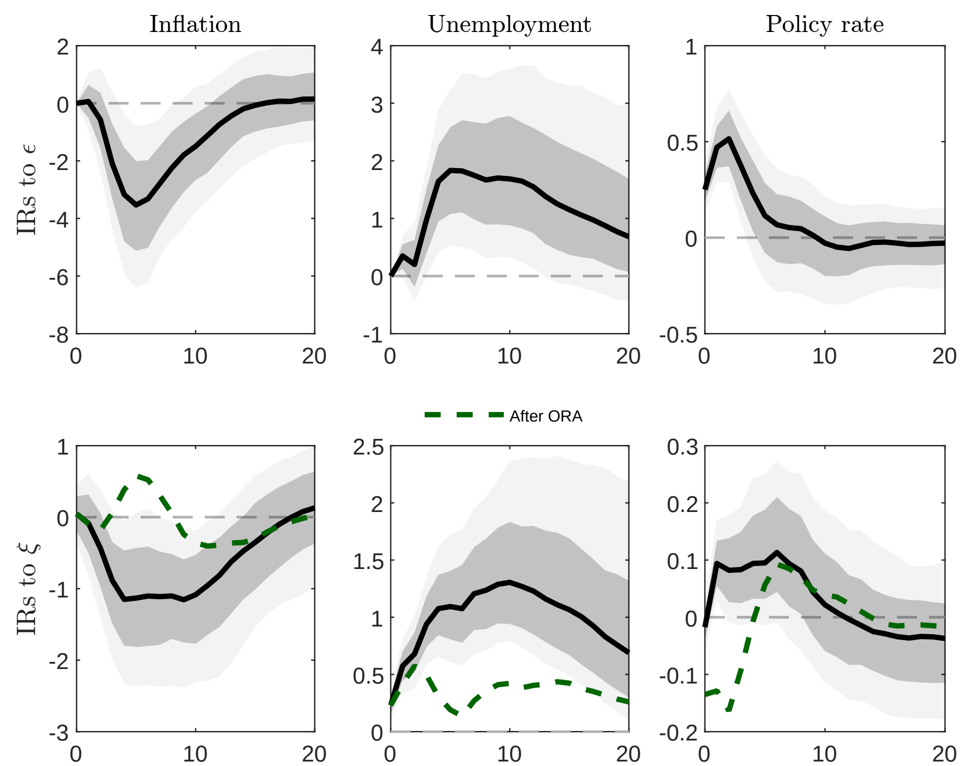
Notes: We show US gold production in kilograms (x1000). The red dots correspond to the peak in the Comstock lode mine in Nevada (1877-Q1), the discovery of Gold in Alaska (1896-Q3), the discovery of Gold mines in Nevada in (1902-Q1), the Nevada gold mine maximum (1910-Q1).

Figure 3: Pre Fed, 1879-1912, Reaction to Financial shocks



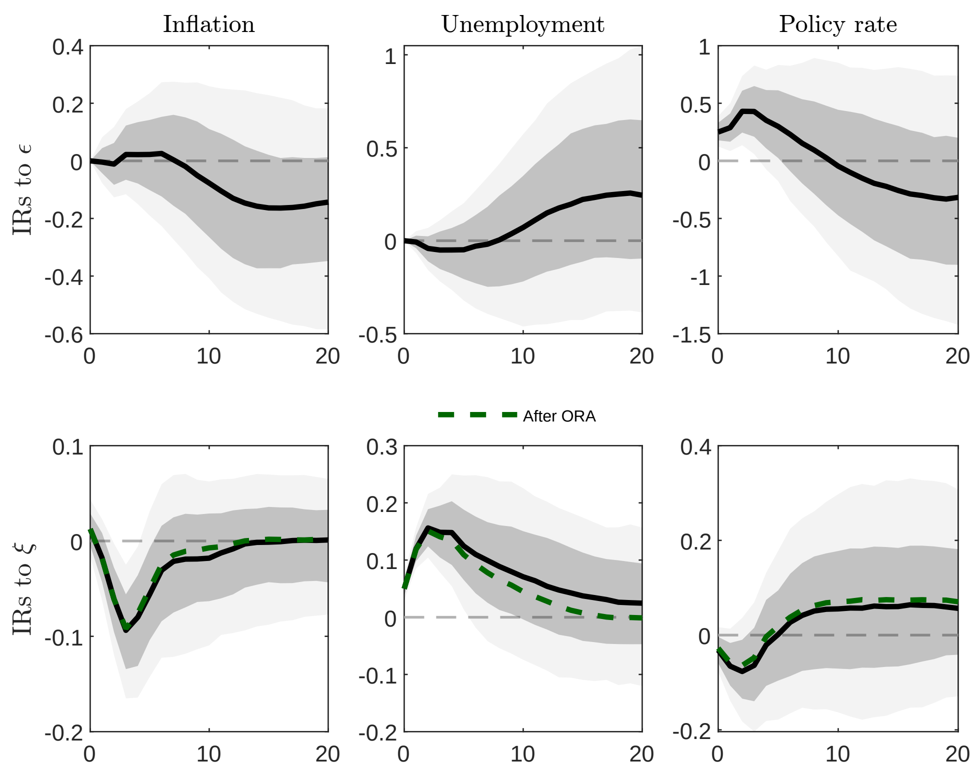
Notes: The top (resp.) row shows the median responses (thick line) of inflation, unemployment and the 3-months treasury bill (the "policy" rate) to a monetary policy shock ϵ (resp. financial shock ξ). The bottom row shows the median responses (thick line) of inflation, unemployment and the policy rate to a bank panic. 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 4: Early Fed, 1913-1941, Reaction to Financial shocks



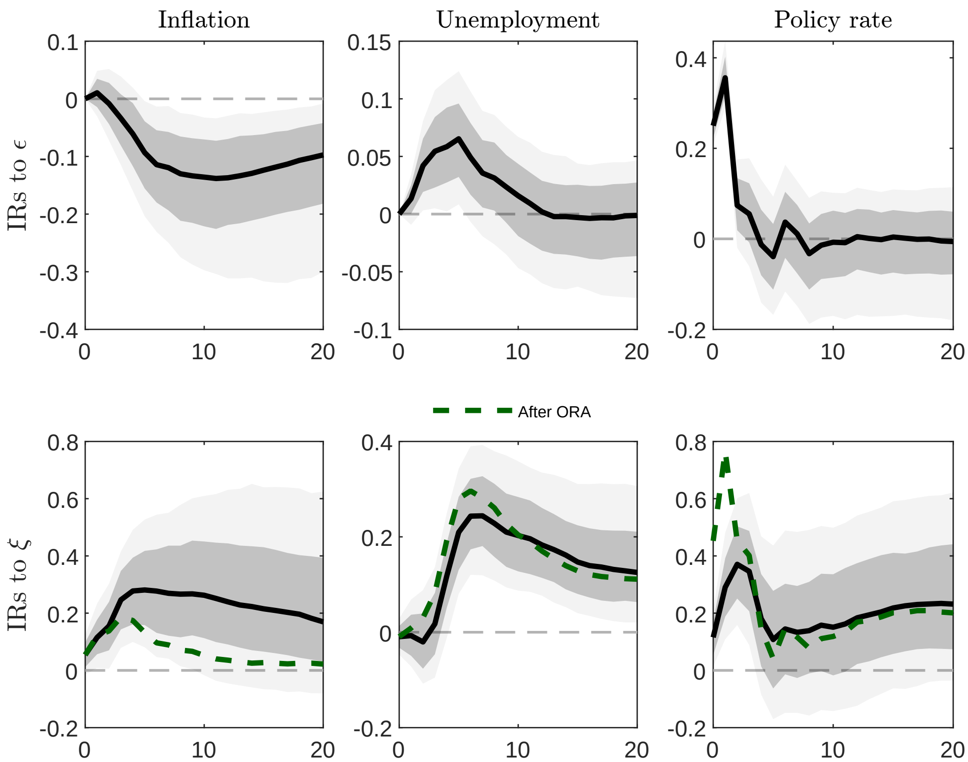
Notes: The top (resp.) row shows the median responses (thick line) of inflation, unemployment and the Fed's discount rate to a monetary policy shock ϵ (resp. financial shock ξ). The dotted green lines show the ORA adjusted impulse responses $\Gamma_0^0 + \mathcal{R}_0^0\tau_0^*$. The 95% and 67% credible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

Figure 5: Post Volcker Fed, 1990-2019, Reaction to Financial shocks



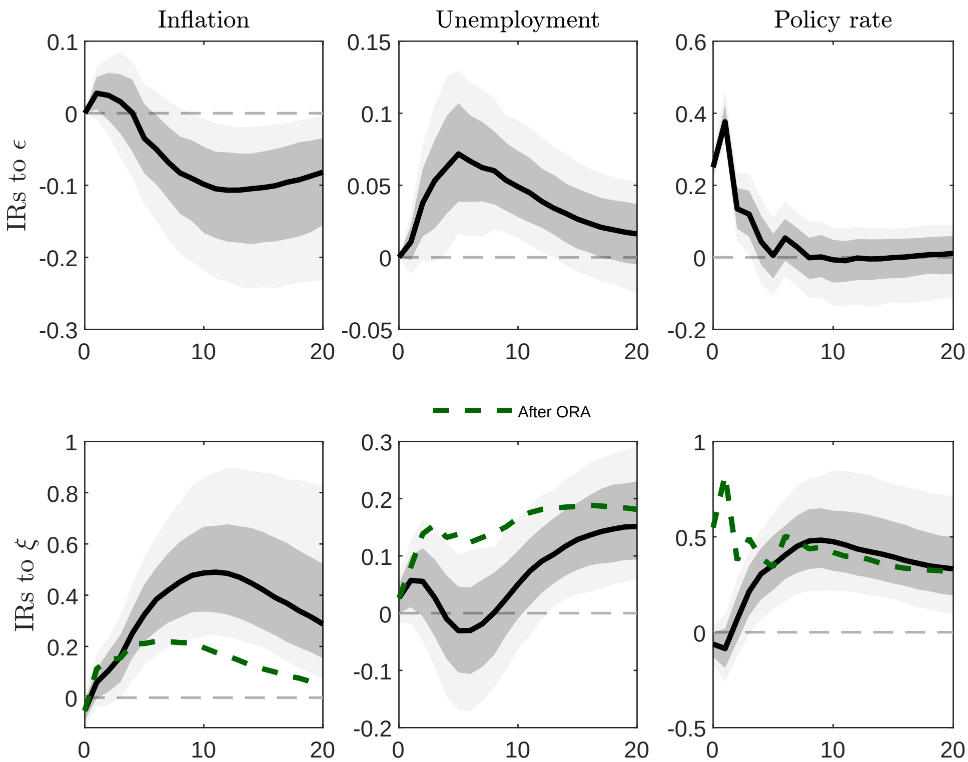
Notes: The top (resp.) row shows the median responses (thick line) of inflation, unemployment and the fed funds rate to a monetary policy shock (resp. financial shock). The dotted green lines show the ORA adjusted impulse responses $\Gamma_0^0 + \mathcal{R}_0^0 \tau_0^*$. The 95% and 67% credible sets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

Figure 6: Post WWII Fed, 1951-1984, Reaction to Energy shocks



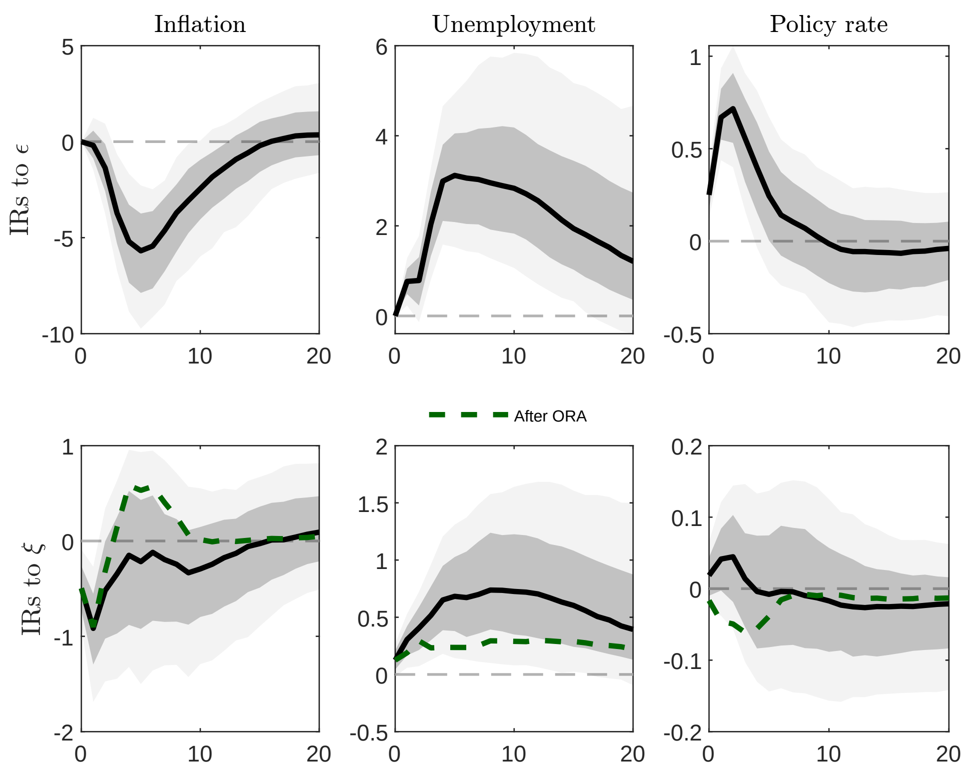
Notes: The top (resp. bottom) row shows the median responses (thick line) of inflation, unemployment and the fed funds rate to a monetary policy shock (resp. energy price shock). The dotted green lines show the ORA adjusted impulse responses $\Gamma_0^0 + \mathcal{R}_0^0\tau_0^*$. The 95% and 67% credible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

Figure 7: Post WWII Fed, 1951-1984, Reaction to π^e shocks



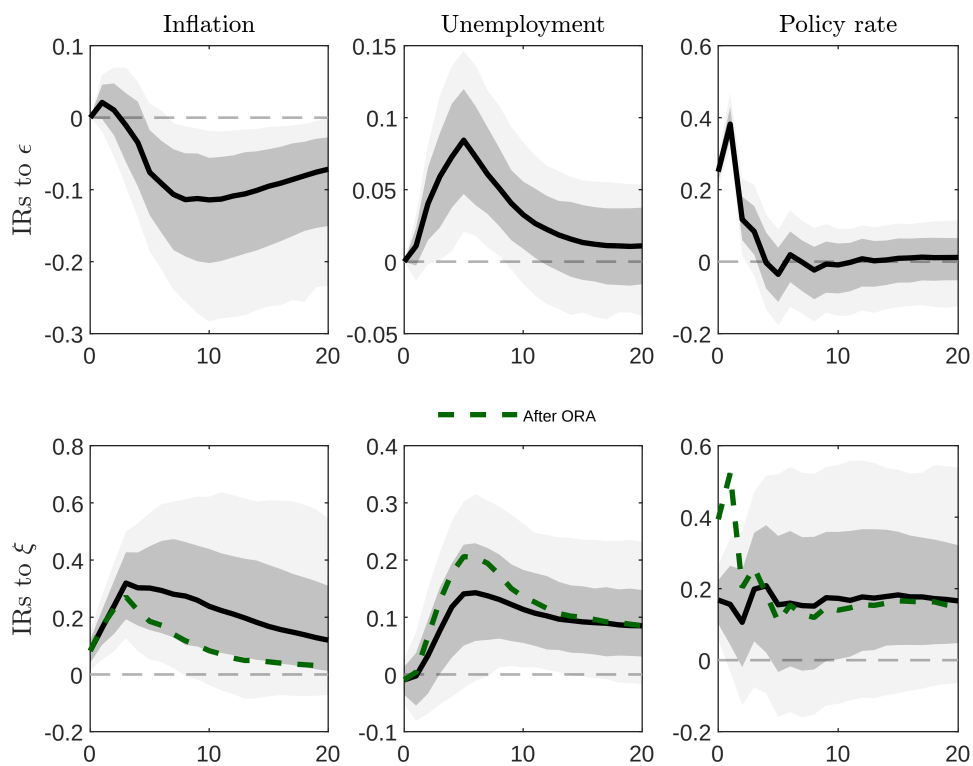
Notes: The top (resp. bottom) row shows the median responses (thick line) of inflation, unemployment and the fed funds rate to a monetary policy shock (resp. inflation expectations shock). The dotted green lines show the ORA adjusted impulse responses $\Gamma_0^0 + \mathcal{R}_0^0 \tau_0^*$. The 95%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

Figure 8: Early Fed, 1913-1941, Reaction to G shocks



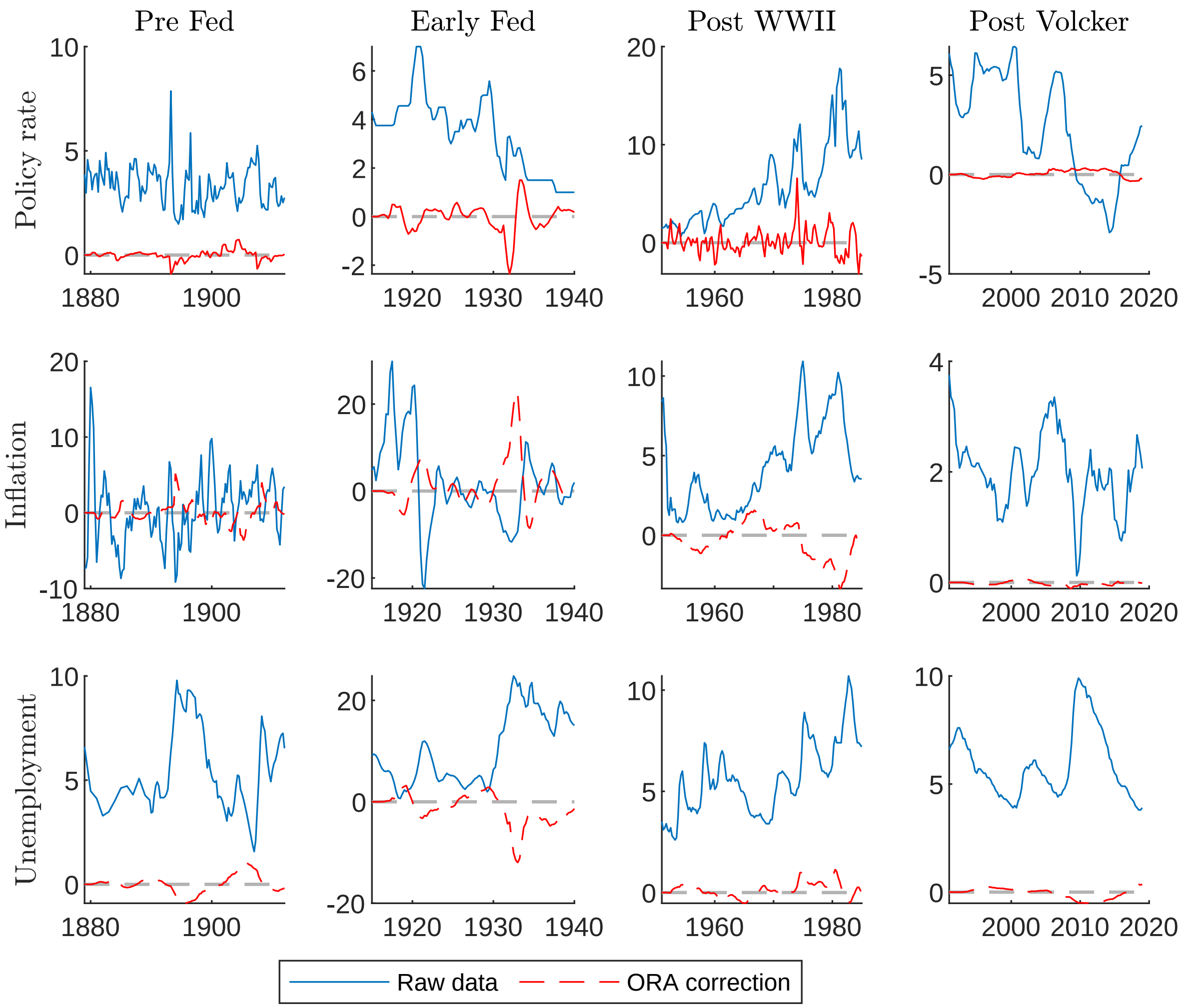
Notes: The top (resp. bottom) row shows the median responses (thick line) of inflation, unemployment and the fed funds rate to a monetary policy shock (resp. government spending shock). The dotted green lines show the ORA adjusted impulse responses $\Gamma_0^0 + \mathcal{R}_0^0 \tau_0^*$. The 95% and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

Figure 9: Post WWII Fed, 1951-1984, Reaction to G shocks



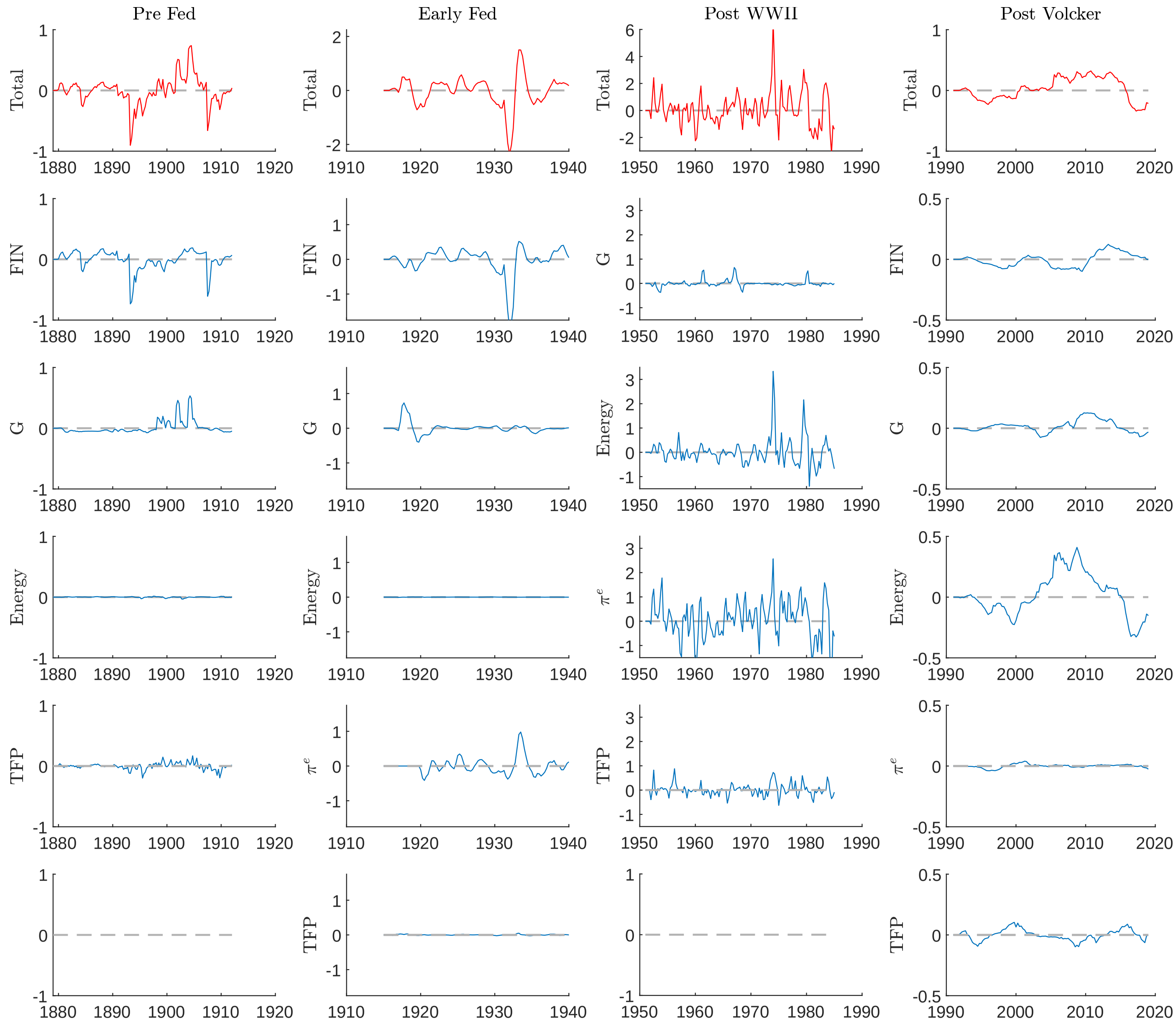
Notes: The top (resp. bottom) row shows the median responses (thick line) of inflation, unemployment and the fed funds rate to a monetary policy shock (resp. government spending shock). The dotted green lines show the ORA adjusted impulse responses $\Gamma_0^0 + \mathcal{R}_0^0 \tau_0^*$. The 95% and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

Figure 10: ORA corrections 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 correction ("ORA correction", dashed red line) over each period, calculated following (24). The middle and bottom rows show the same information but for the inflation rate and the unemployment rate.

Figure 11: ORA DECOMPOSITION



Notes: Each column decomposes median ORA adjustment to the policy rate in the contribution of each non-policy shock: financial (FIN), government spending (G), Energy price (Energy), inflation expectations (π^e) and TFP for the four periods: Pre Fed, EArly Fed, Post WWII and Post Volcker. The total ORA corrections to the contemporaneous policy rate (and given by (24)) are depicted in the top row.